RISK SIMULATOR
Manual do Usuário

Johnathan Mun, Ph.D., MBA, MS, CFC, CRM, FRM, MIFC

RISK SIMULATOR 2012

Materiais baseados em publicações protegidas por direito autoral do Dr. Johnathan Mun, fundador e CEO, Real Options Valuation, Inc.

Escrito por Dr. Johnathan Mun.

Escrito, concebido e publicado nos Estados Unidos da América.

Para comprar cópias adicionais deste documento, entre em contato com a Real Options Valuation, Inc. por este endereço de email:

Admin@RealOptionsValuation.com ou visite o site www.realoptionsvaluation.com

SUMÁRIO

1. INTRODUÇÃO

Bem-vindo ao software RISK SIMULATOR

O **Risk Simulator** é um software de previsão, otimização e simulação Monte Carlo. O software foi escrito na linguagem Microsoft .NET C# e funciona com o Excel como um suplemento. Este software também é compatível e usado com frequência com os softwares, Real Options Super Lattice Solver (SLS) e Employee Stock Options Valuation Toolkit (ESOV), desenvolvidos pela Real Options Valuation, Inc. Observe que embora este manual do usuário pretenda ser completo, ele não é de modo algum um substituto do DVD de treinamento, dos cursos de treinamento presenciais e dos livros escritos pelo criador do software (por exemplo, os livros do Dr. Johnathan Mun, *Real Options Analysis*, 2ª edição, Wiley Finance 2005; *Modeling Risk*, 2ª edição*: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization*, 2ª edição, Wiley 2010; e *Valuing Employee Stock Options (2004 FAS 123R),* Wiley Finance, 2004). Visite nosso site www.realoptionsvaluation.com para obter mais informações sobre esses itens.

O software ***Risk Simulator*** contém os seguintes módulos:

- Simulação Monte Carlo (executa simulações paramétricas e não paramétricas de 42 distribuições de probabilidade com diferentes perfis de simulação, simulações truncadas e correlacionadas, distribuições personalizáveis, simulações controladas de precisão e erro e muitos outros algoritmos)
- Previsão (executa ARIMA, Box-Jenkins, regressão múltipla, extrapolação não linear, processos estocásticos e análise de série temporal)
- Otimização sob incerteza (executa otimizações usando variáveis inteiras discretas e contínuas para otimização de portfólios e projetos com ou sem simulação)
- Ferramentas analíticas e de modelagem (executam análises tornado, aranha e de sensibilidade, bem como simulação de bootstrap, teste de hipóteses, ajuste da distribuição etc.)

O software ***Real Options SLS*** é usado para calcular opções simples e complexas e permite criar modelos de opções personalizáveis. O software ***Real Options SLS*** contém os seguintes módulos:

- Single Asset SLS (para resolver opções de abandono, escolha, contração, adiamento e expansão, além de resolver opções personalizadas)
- Multiple Asset and Multiple Phase SLS (para resolver opções sequenciais de várias fases, opções com vários ativos e fases subjacentes, combinações de várias fases sequenciais com abandono, escolha, contração, adiamento, expansão e troca de opções; também pode ser usado para resolver opções personalizadas)
- Multinomial SLS (para resolver opções trinomiais de reversão à média, opções quadrinomiais de difusão com salto e opções pentanomiais de arco-íris)
- Funções suplementares do Excel (para resolver todas as opções acima além de modelos de contrato fechado e opções personalizadas em um ambiente baseado no Excel)

Requisitos e procedimentos de instalação

Para instalar o software, siga as instruções na tela. Os requisitos mínimos do software são:

- Processador Pentium IV ou mais recente (recomenda-se dual core).
- Windows XP, ou Vista ou Windows 7.
- Microsoft Excel XP, 2003, 2007, 2010 ou posterior.
- Microsoft .NET Framework 2.0 ou 3.0 ou posterior.
- 500 MB de espaço livre.
- 2 GB de RAM no mínimo (recomenda-se de 2 a 4 GB).
- Direitos de administrador para instalar o software.

A maioria dos novos computadores vêm com o Microsoft .NET Framework 2.0/3.0 já pré-instalado. No entanto, se durante a instalação, o Risk Simulator exibir uma mensagem de erro sobre o .NET Framework, saia da instalação. Instale o software .NET Framework relevante incluído no CD (escolha o seu idioma). Conclua a instalação do .NET, reinicie o computador e reinstale o Risk Simulator.

Há um arquivo de licença de avaliação padrão válido por 10 dias que acompanha o software. Para obter uma licença corporativa completa, entre em contato com a Real Options Valuation, Inc. por email para admin@realoptionsvaluation.com, pelo telefone +1 (925) 271-4438 ou visite o nosso site www.realoptionsvaluation.com. Visite o site e clique em DOWNLOAD para obter a versão mais recente do software ou clique no link FAQ para obter informações atualizadas sobre problemas e correções relativos ao licenciamento ou à instalação.

Licenciamento

Se você instalou o software e comprou uma licença completa para usá-lo, você precisa enviar um email para nós com a ID do seu hardware para que possamos gerar um arquivo de licença para você. Siga estas instruções:

Para **Windows XP** executando Excel XP, Excel 2003, Excel 2007, ou Excel 2010:

- No Excel, clique em ***Risk Simulator* | *Licença***. Anote a ID do HARDWARE alfanumérica de 11-20 dígitos e a envie por email para: admin@realoptionsvaluation.com. (você também pode selecionar a ID do hardware e copiar com o botão direito do mouse ou clicar no link do email da ID do hardware). Quando recebermos essa ID, enviaremos por email uma licença permanente para você. De posse desse arquivo de licença, basta salvá-lo no disco rígido, iniciar o Excel, clicar em ***Risk Simulator* | *Licença***, clicar em ***Instalar licença*** e selecionar o novo arquivo de licença. Reinicie o Excel para concluir a operação. O processo demora menos de um minuto e você estará totalmente licenciado para executar o software.

Para **Windows Vista/Windows 7** executando Excel XP, Excel 2003, Excel 2007, ou Excel 2010:

- Inicie o Excel 2007/2010 no Windows Vista ou Windows 7, vá para a guia de menus do Risk Simulator, clique no ícone ***Licença*** ou em ***Risk Simulator│Licença*** e copie sua ID do hardware de 11-20 dígitos alfanuméricos e envie-a por email (também é possível selecionar a ID do hardware, copiar com o botão direito do mouse ou clicar no link de email da ID do hardware) para: admin@realoptionsvaluation.com. Quando recebermos essa ID, enviaremos por email uma licença permanente para você. De posse desse arquivo de licença, basta salvá-lo no disco rígido. Inicie o Excel, clique em ***Risk Simulator │ Licença*** ou clique no ícone ***Licença*** , clique em ***Instalar licença*** e selecione o novo arquivo de licença. Reinicie o Excel para concluir a operação. O processo demora menos de um minuto e você estará totalmente licenciado para executar o software.

Depois de concluída a instalação, inicie o Microsoft Excel. Se a licença tiver sido instalada com êxito, você verá um item adicional "*Risk Simulator*" na barra de menus do Excel XP/2003 ou no grupo SUPLEMENTO do Excel 2007/2010, e uma nova barra de ícones no Excel, como mostra a Figura 1.1. Além disso, será exibida uma tela inicial como mostra a Figura 1.2, indicando que o software está funcionando e está carregado no Excel. A Figura 1.3 também mostra a barra de ferramentas do Risk Simulator. Se esses itens existirem no Excel, você estará pronto para começar a usar o software. As próximas seções fornecem instruções passo a passo sobre como usar o software.

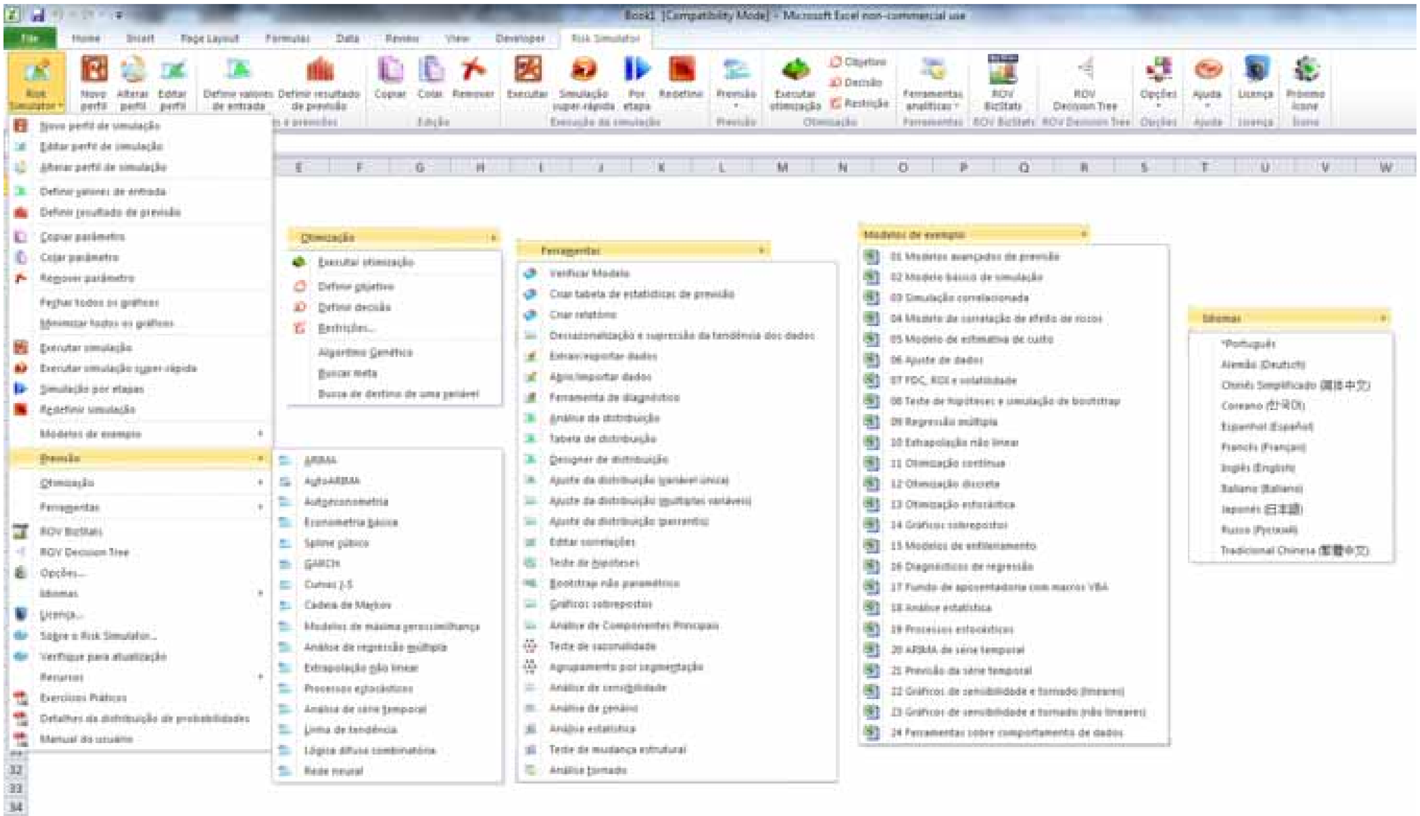


Figura 1.1– Menu do Risk Simulator e barra de ícones do Excel 2007/2010

Figura 1.2 – Tela inicial do Risk Simulator

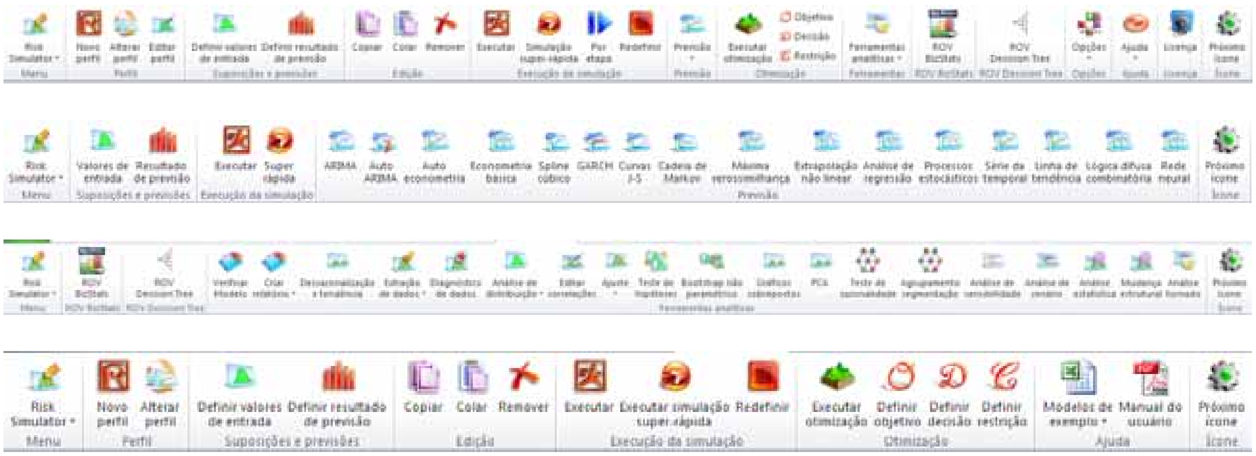


Figura 1.3 – Barras de ferramentas de ícones do Risk Simulator no Excel 2007/2010

NOVIDADES DA VERSÃO 2011/2012

Uma lista abrangente das funcionalidades do Risk Simulator

Veja a seguir uma lista dos principais recursos do Risk Simulator. Os itens realçados indicam os últimos acréscimos à versão 2011/2012.

Recursos gerais

1. Disponível em 11 idiomas: inglês, francês, alemão, italiano, japonês, coreano, português, espanhol, russo, chinês simplificado e chinês tradicional.
2. ROV Decision Tree (Árvore de Decisão) é usado para criar valor e avaliar modelos de árvore de decisão. As seguintes metodologias e análises avançadas adicionais também estão incluídas:
   - Modelos de Árvore de Decisão
   - Simulação de risco (Monte Carlo)
   - Análise de Sensibilidade
   - Análise de Cenário
   - Análise Bayesiana (probabilidade conjunta e a posteriori)
   - Valor Esperado da Informação
   - MINIMAX
   - MAXIMIN
   - Perfis de Risco
3. Células comentadas: ative ou desative os comentários de células e decida se deseja mostrá-los em todos os valores de entrada, resultados de previsão e variáveis de decisão.
4. Modelos de exemplo detalhados: 24 modelos de exemplo no Risk Simulator e mais de 300 modelos no Modeling Toolkit.
5. Relatórios detalhados: todas as análises vêm com relatórios detalhados.
6. Manual do usuário detalhado: manual do usuário passo a passo.
7. Licenciamento flexível: capacidade de ativar ou desativar determinadas funcionalidades para permitir a personalização da experiência de análise de risco. Por exemplo, se você só está interessado nas ferramentas de previsão do Risk Simulator, pode obter uma licença especial que ativa apenas as ferramentas de previsão, mas desativa os outros módulos. Assim, economiza-se custos com o software.
8. Requisitos flexíveis: funciona no Windows 7, Vista e XP; integração com o Excel 2010, 2007, 2003; funciona em máquinas virtuais que executem sistemas operacionais MAC.
9. Cores e gráficos completamente personalizáveis: inclinação, 3D, cor, tipo de gráfico e muito mais.

10. Exercícios práticos: guia detalhado passo a passo para a execução do Risk Simulator, incluindo guias sobre a interpretação dos resultados.
11. Copiar e colar várias células: permite que suposições, variáveis de decisão e previsões sejam copiadas e coladas.
12. Criação de perfis: permite que vários perfis sejam criados em um único modelo, ou seja, é possível criar, duplicar, editar e executar vários cenários de modelos de simulação em um único modelo.
13. Ícones revisados no Excel 2007/2010: uma barra de ferramentas de ícones totalmente redesenhada, mais intuitiva e amigável. Há quatro conjuntos de ícones que se ajustam à maioria das resoluções de tela (1280 x 760 ou melhor).
14. Menus de atalho: acesse todas as ferramentas e menus do Risk Simulator clicando com o botão direito do mouse.
15. Integração com softwares da ROV: funciona bem com outros softwares da ROV, incluindo o Real Options SLS, o Modeling Toolkit, o Basel Toolkit, o ROV Compiler, o ROV Extractor and Evaluator, o ROV Modeler, o ROV Valuator, o ROV Optimizer, o ROV Dashboard, o ESO Valuation Toolkit, entre outros.
16. Funções RS no Excel: funções RS para configurar suposições e previsões e suporte a cliques com o botão direito do mouse no Excel.
17. Solução de problemas: esta ferramenta permite reabilitar o software, verificar os requisitos do sistema, obter a ID do hardware, entre outros.
18. Turbo Speed Analysis: este novo recurso executa previsões e outras ferramentas de análise em velocidades altíssimas (aprimorado na versão 5.2). A análise e os resultados permanecem os mesmos, mas agora são calculados e geram relatórios muito rapidamente.
19. Recursos da Web, estudos de caso e vídeos: baixe gratuitamente modelos, vídeos de introdução, estudos de caso, whitepapers e outros materiais no nosso site.

Módulo de Simulação

20. Seis geradores de números aleatórios: gerador subtrativo avançado ROV, gerador de embaralhamento aleatório subtrativo, gerador de embaralhador de longo período, gerador de embaralhamento aleatório portátil, gerador rápido de hexadecimais IEEE, gerador portátil mínimo básico.
21. Dois métodos de amostragem: Monte Carlo e hipercubo latino.
22. Três cópulas de correlação: aplicação de cópula normal, cópula T e cópula quase-normal para simulações correlacionadas.
23. Quarenta e duas distribuições de probabilidade: arco seno, Bernoulli, beta, beta 3, beta 4, binomial, Cauchy, qui-quadrada, cosseno, personalizada, uniforme discreta, logarítmica dupla, Erlang, exponencial, exponencial 2, distribuição F, gama, geométrica, máximo de Gumbel, mínimo de Gumbel, hipergeométrica, Laplace, logística, lognormal (aritmética) e lognormal (log), lognormal 3 (aritmética) e lognormal 3 (log), binomial negativa, normal, parabólica, Pareto, Pascal, Pearson V,

Pearson VI, PERT, Poisson, potência, potência 3, Rayleigh, T e T2, triangular, uniforme, Weibull, Weibull 3.

24. Parâmetros alternativos: uso de percentis como uma alternativa à inserção de parâmetros.
25. Distribuição não paramétrica personalizada: crie suas próprias distribuições, executando simulações históricas e aplicando o método Delphi.
26. Truncamento de distribuição: habilitação de limite de dados.
27. Funções do Excel: defina suposições e previsões usando funções no Excel
28. Simulação multidimensional: simulação de parâmetros de entrada incertos.
29. Controle de precisão: determina se o número de tentativas de simulação executadas é suficiente.
30. Simulação super-rápida: executa 100.000 tentativas em poucos segundos.

Módulo de Previsão

31. ARIMA: modelos ARIMA ou média móvel integrada autorregressiva (P,D,Q).
32. AutoARIMA: executa as combinações mais comuns do ARIMA para encontrar o modelo de melhor ajuste.
33. Autoeconometria: executa milhares de combinações de modelos e permutações para obter o modelo de melhor ajuste para os dados existentes (linear, não linear, interação, defasagem, diferenças, taxa, diferença).
34. Econometria básica: modelos de regressão interativa, linear/não linear e econométricos.
35. Spline cúbico: interpolação e extrapolação não lineares.
36. GARCH: projeções de volatilidade usando modelos autorregressivos à heteroscedasticidade condicional generalizada: GARCH, GARCH-M, TGARCH, TGARCH-M, EGARCH, EGARCH-T, GJR-GARCH e GJR-TGARCH.
37. Curva J: curvas J exponenciais.
38. Variáveis dependentes limitadas: logit, probit e tobit.
39. Cadeias de Markov: dois elementos concorrentes ao longo do tempo e previsões de participação no mercado.
40. Regressão múltipla: regressão normal linear e não linear, com metodologias stepwise (avanço, regresso, correlação, avanço-regresso).
41. Extrapolação não linear: previsão de série temporal não linear.
42. Curva S: curvas S logísticas.
43. Análise de série temporal: oito modelos de decomposição de séries temporais para a previsão de níveis, tendências e sazonalidades.
44. Linhas de tendência: previsão e ajuste usando média linear, não linear, polinomial, potência, logarítmica, exponencial e média móvel com melhor ajuste.
45. Previsão de lógica difusa combinatória
46. Previsão de rede neural (linear, logística, tangente hiperbólica, cosseno com tangente hiperbólica)

Modelo de Otimização

47. Otimização linear: otimização multifásica e otimização linear geral.
48. Otimização não linear: resultados detalhados, incluindo matrizes Hessianas, funções de LaGrange e mais.
49. Otimização estática: execuções rápidas para obter otimizações contínuas, integradas e binárias.
50. Otimização dinâmica: simulação com otimização.
51. Otimização estocástica: critérios quadráticos, tangenciais, centrais, de avanço, de convergência.
52. Fronteira eficiente: combinações de otimizações estocásticas e dinâmicas em fronteiras eficientes multivariadas.
53. Algoritmos genéticos: usados em vários problemas de otimização.
54. Otimização multifásica: testes de otimização local ou global que permitem um melhor controle sobre como a otimização é executada e aumentam a precisão e a dependência dos resultados.
55. Percentis e médias condicionais: estatísticas adicionais para otimização estocástica, incluindo percentis e médias condicionais, que são críticas no cálculo de valor condicional em medidas de risco.
56. Algoritmo de pesquisa: algoritmos de pesquisa simples, rápidos e eficientes para variáveis de decisão simples básicas e aplicações de busca de metas.
57. Simulação super-rápida em otimização dinâmica e estocástica: executa simulações super-rápidas integradas à otimização.

Módulo de Ferramentas analíticas

58. Verificar modelo: testa os erros mais comuns no seu modelo.
59. Editor de correlação: permite que grandes matrizes de correlação sejam inseridas e editadas diretamente.
60. Criar relatório: geração automática de um relatório de suposições e previsões em um modelo.
61. Criar relatório estatístico: gera um relatório comparativo de todas as estatísticas de previsão.
62. Diagnóstico de dados: executa testes de heteroscedasticidade, micronumerosidade, observações discrepantes, não linearidade, autocorrelação, normalidade, esfericidade, não estacionariedade, multicolinearidade e correlações.
63. Extração e exportação de dados: extrai dados para o Excel ou para arquivos de texto simples e arquivos do Risk Simulator, executa relatórios estatísticos e relatórios de resultado de previsão.
64. Abrir e importar dados: recupera os resultados de execução de simulações anteriores.
65. Supressão de tendência e dessazonalização: dessazonaliza e suprime tendências nos dados.

66. Análise da distribuição: calcula o FDP, o FDA e o IFDA exatos de todas as 42 distribuições e gera tabelas de probabilidade.
67. Designer de distribuição: crie suas próprias distribuições personalizadas.
68. Ajuste da distribuição (múltiplo): executa múltiplas variáveis simultaneamente, levando em conta as correlações e a significância das correlações.
69. Ajuste da distribuição (simples): testes de Kolmogorov-Smirnov e qui-quadrado em distribuições contínuas, completos com relatórios e suposições de distribuição.
70. Teste de hipóteses: testa se duas previsões são estatisticamente semelhantes ou diferentes.
71. Bootstrap não paramétrico: simulação das estatísticas para obter a precisão e a exatidão dos resultados.
72. Gráficos sobrepostos: gráficos sobrepostos de suposições e previsões totalmente personalizáveis (tipos de gráfico FDA, FDP, 2D/3D).
73. Análise de Componentes Principais: testa as variáveis do melhor indicador e as maneiras de reduzir a matriz de dados.
74. Análise de cenário: milhares de cenários estáticos bidimensionais.
75. Teste de sazonalidade: testa várias defasagens de sazonalidade.
76. Agrupamento por segmentação: agrupa dados em combinações estatísticas para segmentá-los.
77. Análise de sensibilidade: sensibilidade dinâmica (análise simultânea).
78. Teste de quebra estrutural: testa se os dados da série temporal têm quebras estruturais estatísticas.
79. Análise tornado: perturbação estática de sensibilidades, análises aranha e tornado, bem como tabelas de cenários.

Módulo de Estatísticas e Estatísticas de negócios

80. Ajuste da distribuição (percentil): uso de percentis e otimização para encontrar a distribuição de melhor ajuste.
81. Tabelas e gráficos das distribuições de probabilidade: executa 45 distribuições de probabilidade, seus quatro momentos, o FDA, o IFDA, o FDP, os gráficos, os vários gráficos sobrepostos de distribuição, além de gerar tabelas de distribuição de probabilidade.
82. Análise estatística: estatísticas descritivas, ajuste da distribuição, histogramas, gráficos, extrapolação não linear, teste de normalidade, estimativa de parâmetros estocásticos, previsão de série temporal, projeções da linha de tendência etc.
83. ROV BIZSTATS: mais de 130 modelos analíticos e estatísticas comerciais:

Absolute Values, ANOVA: Randomized Blocks Multiple Treatments, ANOVA: Single Factor Multiple Treatments, ANOVA: Two Way Analysis, ARIMA, Auto ARIMA, Autocorrelation and Partial Autocorrelation, Autoeconometrics (Detailed), Autoeconometrics (Quick), Average, Combinatorial Fuzzy Logic Forecasting, Control Chart: C, Control Chart: NP, Control Chart: P, Control Chart: R, Control

Chart: U, Control Chart: X, Control Chart: XMR, Correlation, Correlation (Linear, Nonlinear), Count, Covariance, Cubic Spline, Custom Econometric Model, Data Descriptive Statistics, Deseasonalize, Difference, Distributional Fitting, Exponential J Curve, GARCH, Heteroskedasticity, Lag, Lead, Limited Dependent Variables (Logit), Limited Dependent Variables (Probit), Limited Dependent Variables (Tobit), Linear Interpolation, Linear Regression, LN, Log, Logistic S Curve, Markov Chain, Max, Median, Min, Mode, Neural Network, Nonlinear Regression, Nonparametric: Chi-Square Goodness of Fit, Nonparametric: Chi-Square Independence, Nonparametric: Chi-Square Population Variance, Nonparametric: Friedman's Test, Nonparametric: Kruskal-Wallis Test, Nonparametric: Lilliefors Test, Nonparametric: Runs Test, Nonparametric: Wilcoxon Signed-Rank (One Var), Nonparametric: Wilcoxon Signed-Rank (Two Var), Parametric: One Variable (T) Mean, Parametric: One Variable (Z) Mean, Parametric: One Variable (Z) Proportion, Parametric: Two Variable (F) Variances, Parametric: Two Variable (T) Dependent Means, Parametric: Two Variable (T) Independent Equal Variance, Parametric: Two Variable (T) Independent Unequal Variance, Parametric: Two Variable (Z) Independent Means, Parametric: Two Variable (Z) Independent Proportions, Power, Principal Component Analysis, Rank Ascending, Rank Descending, Relative LN Returns, Relative Returns, Seasonality, Segmentation Clustering, Semi-Standard Deviation (Lower), Semi-Standard Deviation (Upper), Standard 2D Area, Standard 2D Bar, Standard 2D Line, Standard 2D Point, Standard 2D Scatter, Standard 3D Area, Standard 3D Bar, Standard 3D Line, Standard 3D Point, Standard 3D Scatter, Standard Deviation (Population), Standard Deviation (Sample), Stepwise Regression (Backward), Stepwise Regression (Correlation), Stepwise Regression (Forward), Stepwise Regression (Forward-Backward), Stochastic Processes (Exponential Brownian Motion), Stochastic Processes (Geometric Brownian Motion), Stochastic Processes (Jump Diffusion), Stochastic Processes (Mean Reversion with Jump Diffusion), Stochastic Processes (Mean Reversion), Structural Break, Sum, Time-Series Analysis (Auto), Time-Series Analysis (Double Exponential Smoothing), Time-Series Analysis (Double Moving Average), Time-Series Analysis (Holt-Winter's Additive), Time-Series Analysis (Holt-Winter's Multiplicative), Time-Series Analysis (Seasonal Additive), Time-Series Analysis (Seasonal Multiplicative), Time-Series Analysis (Single Exponential Smoothing), Time-Series Analysis (Single Moving Average), Trend Line (Difference Detrended), Trend Line (Exponential Detrended), Trend Line (Exponential), Trend Line (Linear Detrended), Trend Line (Linear), Trend Line (Logarithmic Detrended), Trend Line (Logarithmic), Trend Line (Moving Average Detrended), Trend Line (Moving Average), Trend Line (Polynomial Detrended), Trend Line (Polynomial), Trend Line (Power Detrended), Trend Line (Power), Trend Line (Rate Detrended), Trend Line (Static Mean Detrended), Trend Line (Static Median Detrended), Variance (Population), Variance (Sample), Volatility: EGARCH, Volatility: EGARCH-T, Volatility: GARCH, Volatility: GARCH-M, Volatility: GJR GARCH, Volatility: GJR TGARCH, Volatility: Log Returns Approach, Volatility: TGARCH, Volatility: TGARCH-M, Yield Curve (Bliss), and Yield Curve (Nelson-Siegel).

2. SIMULAÇÃO MONTE CARLO

A simulação Monte Carlo, assim chamada devido ao famoso cassino no principado de Mônaco, é uma metodologia muito poderosa. Para o exercício de uma profissão, a simulação abre as portas para a resolução de problemas difíceis e complexos, ainda que práticos, com grande facilidade. A simulação Monte Carlo cria futuros artificiais gerando até milhões de caminhos de amostras de resultados e observa suas características predominantes. Para os analistas em uma empresa, cursar disciplinas de matemática avançada no nível de pós-graduação simplesmente não é lógico nem prático. Um analista brilhante usaria todas as ferramentas disponíveis ao seu dispor para obter a mesma resposta da maneira mais fácil e prática possível. Em todo caso, quando modelada corretamente, a simulação Monte Carlo fornece respostas semelhantes aos métodos matematicamente mais elegantes. Assim, o que é a simulação Monte Carlo e como ela funciona?

O que é a simulação Monte Carlo?

A simulação Monte Carlo na sua forma mais simples é um gerador de números aleatórios útil para previsões, estimativas e análises de risco. Uma simulação calcula numerosos cenários de um modelo escolhendo valores repetidamente de uma *distribuição de probabilidade* predefinida pelo usuário para as variáveis incertas e usando esses valores no modelo. Todos esses cenários produzem resultados associados em um modelo, em que cada cenário pode ter uma *previsão*. As previsões são eventos (normalmente com fórmulas ou funções) que você define como resultados importantes do modelo. Em geral, são eventos como totais, lucro líquido ou despesas brutas.

De maneira simplificada, pense na simulação Monte Carlo como escolher bolas de golfe de uma grande cesta repetidamente com reposições. O tamanho e a forma da cesta depende dos ***valores de entrada*** da distribuição (por exemplo, uma distribuição normal com uma média 100 e um desvio padrão 10, comparada com uma distribuição uniforme ou triangular) em que algumas cestas são mais fundas ou mais simétricas do que outras, permitindo que certas bolas sejam tiradas com mais frequência do que outras. O número de bolas tiradas repetidamente depende do número de ***tentativas*** simuladas. Para um modelo grande com várias suposições relacionadas, imagine o modelo grande como uma cesta muito grande, com muitas cestas menores dentro. Cada cesta menor tem seu próprio conjunto de bolas que estão se movimentando. Algumas vezes essas cestas menores estão juntas umas das outras (se houver uma ***correlação*** entre as variáveis) e as bolas de golfe estão se movimentando em pares enquanto outras se movimentam independentemente umas das outras. As bolas que são retiradas de cada vez dessas interações dentro do modelo (a cesta maior) são tabuladas e registradas, fornecendo o ***resultado de saída da previsão*** da simulação.

Introdução ao Risk Simulator

Visão geral de alto nível do software

O software *Risk Simulator* tem vários usos diferentes, incluindo simulação Monte Carlo, previsão, otimização e análise de risco.

- O *módulo Simulação* permite executar simulações em modelos existentes do Excel, gerar e extrair previsões de simulação (distribuição de resultados), executar o ajuste da distribuição (encontrar automaticamente a distribuição estatística mais adequada), calcular correlações (manter relações entre variáveis aleatórias simuladas), identificar sensibilidades (criando gráficos tornado e de sensibilidade), testar hipóteses estatísticas (encontrando diferenças estatísticas entre pares de previsões), executar uma simulação de *bootstrap* (testando a robustez das estatísticas do resultado) e executar simulações personalizadas e não paramétricas (simulações usando dados históricos sem especificar distribuições ou parâmetros de previsão para fazer previsões sem dados ou aplicar previsões de especialistas).
- O *módulo Previsão* pode ser usado para gerar previsões de séries temporais automáticas (com ou sem sazonalidade e tendência), regressões multivariadas (modelando relações entre variáveis), extrapolações não lineares (ajuste de curva), processos estocásticos (caminhos aleatórios, reversões à média, difusão com salto e processos mistos), ARIMA Box-Jenkins (previsões econométricas), AutoARIMA, econometria básica e autoeconometria (modelagem de relações e geração de previsões), curvas J exponenciais, curvas S logísticas, modelos GARCH e suas variações (modelagem e previsão da volatilidade), modelos de máxima verossimilhança para variáveis dependentes limitadas (modelos logit, tobit e probit), cadeias de Markov, linhas de tendência, curvas spline, entre outros.
- O *módulo Otimização* é usado para otimizar múltiplas variáveis de decisão sujeitas a restrições para maximizar ou minimizar um objetivo. Ele pode ser executado como uma otimização estática, dinâmica ou estocástica sob incerteza junto com a simulação Monte Carlo, ou como uma otimização estocástica com simulações super-rápidas. O software pode lidar com otimizações lineares e não lineares com variáveis binárias, inteiras e contínuas, e também gerar fronteiras eficientes de Markowitz.
- O *módulo Ferramentas analíticas* permite executar agrupamento por segmentação, teste de hipóteses, testes estatísticos de dados brutos, diagnóstico de dados de suposições de previsões técnicas (por exemplo, heteroscedasticidade, multicolinearidade e semelhantes), análises de sensibilidade e de cenário, análise de gráficos sobrepostos, gráficos aranha, gráficos tornado e muitas outras ferramentas poderosas.
- O Real Options Super Lattice Solver é um software autônomo que complementa o Risk Simulator e é usado para resolução de problemas de opções reais simples ou complexos.

As próximas seções apresentam informações básicas sobre o uso do módulo Simulação do Risk Simulator. Para obter informações mais detalhadas sobre as aplicações dos outros módulos, consulte outros capítulos. Para acompanhar, verifique se você tem o Risk Simulator instalado no computador antes de continuar. ***É recomendável que primeiro você assista aos vídeos de introdução disponíveis na Web (www.realoptionsvaluation.com/risksimulator.html) ou tente fazer os exercícios passo a passo no final deste capítulo antes de revisar o texto deste capítulo.*** Os vídeos e os exercícios são ótimas maneiras de preparar o usuário imediatamente, ao passo que este capítulo se concentra mais na teoria e em explicações detalhadas das propriedades de simulação.

Executar uma simulação Monte Carlo

Em geral, para executar uma simulação no seu modelo Excel existente, estas etapas devem ser executadas:

1. **Iniciar um novo perfil de simulação ou abrir um perfil existente**
2. **Definir valores de entrada nas células relevantes**
3. **Definir resultados de previsão nas células relevantes**
4. **Executar a simulação**
5. **Interpretar os resultados**

Se desejar praticar, abra o arquivo de exemplo intitulado ***Modelo básico de simulação*** e siga os exemplos abaixo para criar uma simulação. O arquivo de exemplo pode ser encontrado em ***Iniciar | Real Options Valuation | Risk Simulator | Exemplos*** ou acessado diretamente em ***Risk Simulator | Modelos de exemplo.***

1. Iniciar um novo perfil de simulação

Para iniciar uma nova simulação, primeiro é necessário criar um perfil de simulação. Um perfil de simulação contém um conjunto completo de instruções sobre como executar uma simulação, por exemplo, todas as suposições, previsões, preferências de execução etc. O uso de perfis facilita a criação de vários cenários de simulação. Ou seja, usando o mesmo modelo, é possível criar vários perfis, cada um com propriedades e requisitos de simulação específicos. A mesma pessoa pode criar diferentes cenários de teste usando suposições e entradas de distribuição diferente ou várias pessoas podem testar suposições e entradas no mesmo modelo.

- **Inicie o Excel** e crie um novo modelo ou abra um existente (você pode usar o exemplo do Modelo básico de simulação para acompanhar)
- Clique em ***Risk Simulator | Novo perfil de simulação***
- Especifique um título para a simulação, além de todas as informações pertinentes (Figura 2.1)

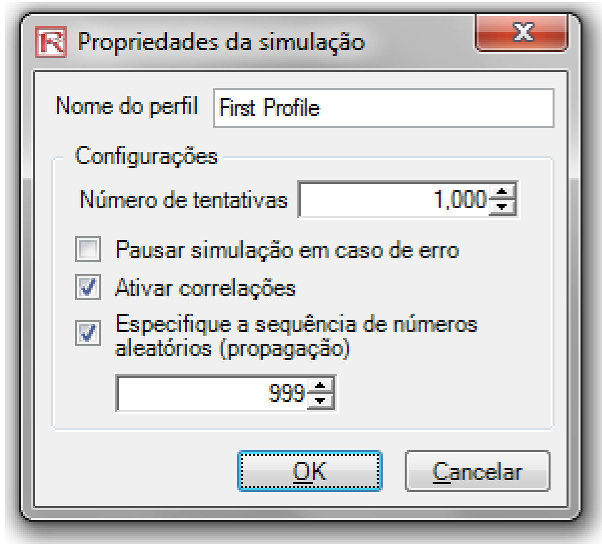


Figura 2.1 – Novo perfil de simulação

- **Título**: A especificação do título de uma simulação permite criar vários perfis de simulação em um único modelo do Excel. Isso significa que agora você pode salvar diferentes perfis de cenário de simulação no mesmo modelo sem ter que excluir as suposições existentes e alterá-las sempre que um novo cenário de simulação for necessário. É possível alterar o nome do perfil mais tarde (***Risk Simulator | Editar perfil***).
- **Número de tentativas**: Insira o número de tentativas de simulação necessário. Ou seja, executar 1.000 tentativas significa que serão geradas 1.000 iterações diferentes de resultados de acordo com os valores de entrada. Você pode alterar esse número se desejar, mas só são aceitos inteiros positivos. O número padrão de execuções é 1.000 tentativas. Você pode usar o controle de precisão e erro para ajudar a determinar automaticamente quantas tentativas de simulação devem ser executadas (consulte a seção sobre controle de precisão e erro para obter detalhes).
- **Pausar simulação em caso de erro**: Se selecionado, a simulação será interrompida sempre que um erro for encontrado no modelo do Excel. Ou seja, se o seu modelo encontrar um erro de cálculo (por exemplo, alguns valores de entrada gerados em uma tentativa de simulação podem resultar em um erro de divisão por zero em uma das células da planilha), a simulação será interrompida. Isso é importante para ajudar a auditar o modelo do Excel a fim de verificar se ele não contém erros de cálculo. No entanto, se você tiver certeza de que o modelo funciona, não será necessário selecionar essa preferência.
- **Ativar correlações**: Se selecionado, as correlações entre pares de valores de entrada serão calculadas. Caso contrário, as correlações serão todas definidas como zero e a simulação será executada supondo que não há correlações entre os valores de entrada. Por exemplo, a aplicação de correlações renderá resultados mais precisos se elas existirem de fato e tenderá a resultar em uma confiança de previsão mais baixa, se houver correlações negativas. Após ativar as correlações, você poderá definir os coeficientes de correlação relevantes em cada suposição gerada (consulte a seção sobre correlações para obter mais detalhes).

- **Especifique a sequência de números aleatórios (propagação)**: A simulação por definição apresentará resultados um pouco diferentes toda vez que uma simulação for executada. Isso ocorre por causa da rotina de geração de números aleatórios na simulação Monte Carlo e é um fato teórico em todos os geradores de números aleatórios. No entanto, ao criar apresentações, é possível que você precise dos mesmos resultados (principalmente quando o relatório apresentado exibe um conjunto de resultados que também devem ser exibidos durante uma apresentação ao vivo, ou quando você compartilha modelos com outras pessoas e deseja que os mesmos resultados sejam obtidos sempre). Nesses casos, selecione essa preferência e insira um número de propagação inicial. O número de propagação pode ser qualquer inteiro positivo. Usando o mesmo valor de propagação inicial, o mesmo número de tentativas e os mesmos valores de entrada, a simulação sempre apresentará a mesma sequência de números aleatórios, garantindo o mesmo conjunto final de resultados.

Observe que depois de criado um novo perfil de simulação, você poderá modificar essas seleções. Para fazer isso, verifique se o perfil ativo atual é aquele que você deseja modificar; caso contrário, clique em ***Risk Simulator | Alterar perfil de simulação***, selecione o perfil que deseja alterar e clique em ***OK*** (a Figura 2.2 mostra um exemplo com vários perfis e como ativar um perfil selecionado). Em seguida, clique em ***Risk Simulator | Editar perfil de simulação*** e faça as alterações necessárias. Também é possível duplicar ou renomear um perfil existente. Ao criar vários perfis no mesmo modelo do Excel, dê um nome exclusivo a cada perfil para poder diferenciá-los depois. Além disso, esses perfis são armazenados dentro de setores ocultos do arquivo *.xls do Excel e você não precisa salvar nenhum arquivo adicional. Os perfis e os respectivos conteúdos (suposições, previsões etc.) são automaticamente salvos quando você salva o arquivo do Excel. Por fim, o último perfil ativo quando você sai do Excel e salva o arquivo será aberto na próxima vez que o arquivo do Excel for acessado.

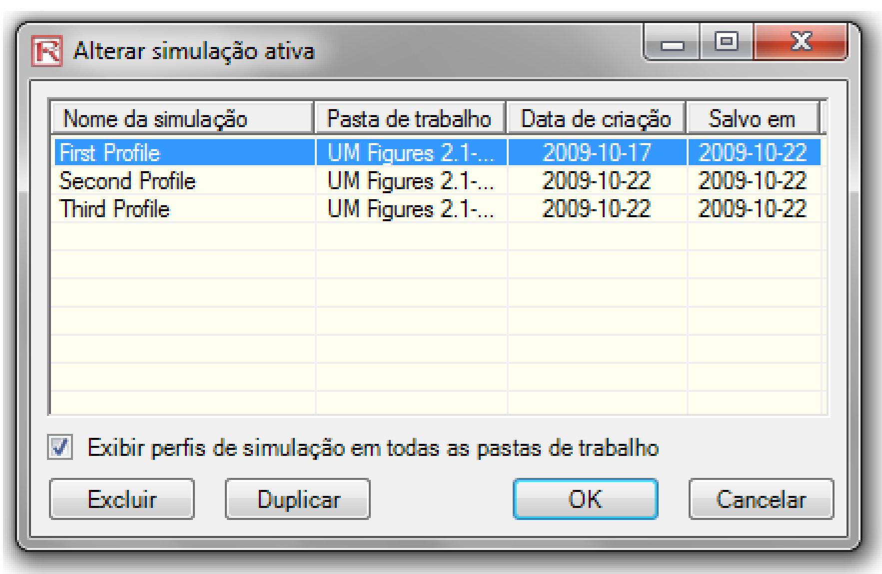


Figura 2.2 – Alterar simulação ativa

2. Definir valores de entrada

A próxima etapa é definir os valores de entrada em seu modelo. Observe que as suposições podem ser atribuídas somente a células que não contêm equações ou funções, por exemplo, valores numéricos inseridos que sejam entradas em um modelo, enquanto os resultados de previsão podem ser atribuídos somente a células que contêm equações e funções, por exemplo, os resultados de um modelo. Lembre-se de que as suposições e previsões não podem ser definidas a menos que já exista um perfil de simulação. Para definir novos valores de entrada no modelo, faça o seguinte:

- Verifique se existe um perfil de simulação, abra um perfil existente ou inicie um novo perfil (***Risk Simulator | Novo perfil de simulação***)
- Selecione a célula na qual deseja definir uma suposição (por exemplo, a célula ***G8*** no exemplo do modelo básico de simulação)
- Clique em ***Risk Simulator | Definir valores de entrada*** ou clique no ícone Definir valores de entrada na barra de ferramentas de ícones do Risk Simulator
- Selecione a distribuição desejada e insira os parâmetros de distribuição relevantes (por exemplo, distribuição ***Triangular*** com ***1,5; 2,0; 2,5*** como os valores mínimo, máximo e mais provável) e clique em ***OK*** para inserir os valores de entrada no modelo (Figura 2.3)

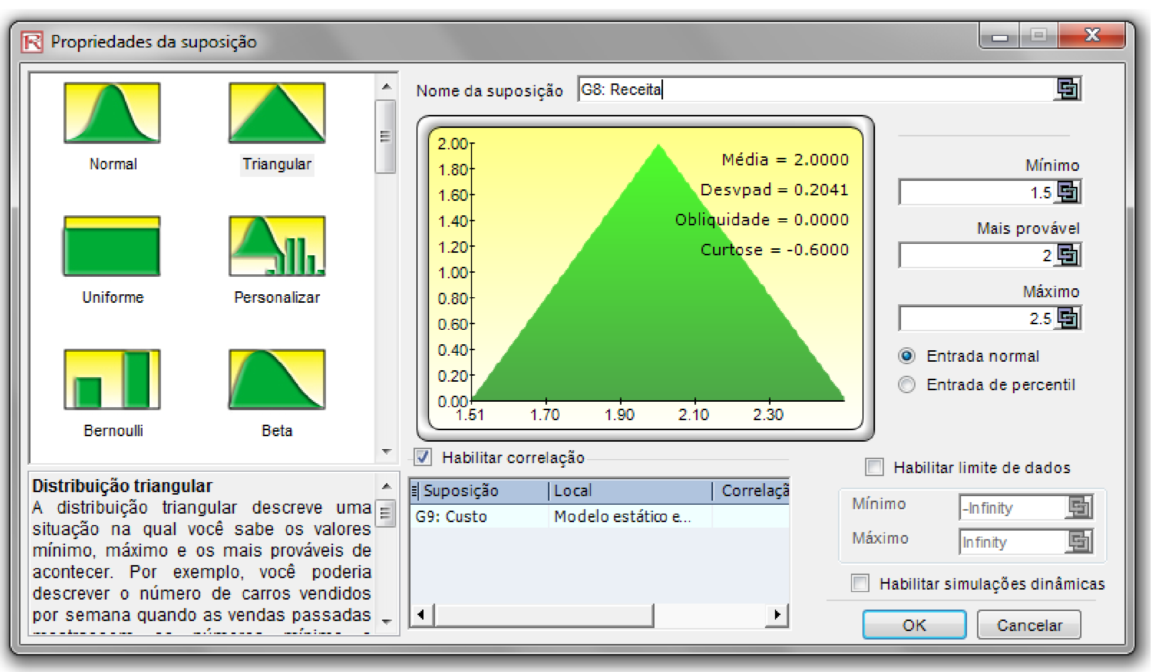


Figura 2.3 – Definir um valor de entrada

Observe que você também pode definir suposições selecionando a célula na qual deseja definir a suposição e ***clicando com o botão direito do mouse***, acesse o menu de atalho do Risk Simulator para definir valores de entrada. Além disso, para usuários experientes, é possível definir valores de entrada usando as ***Funções RS*** do Risk Simulator: selecione a célula desejada e, no Excel, clique em *Inserir*, *Função* e selecione *Todas as Categorias* e role para baixo até a lista de funções *RS* (não é recomendável usar as funções RS se você não é um usuário experiente). Para os exemplos a seguir, siga as instruções básicas sobre como acessar menus e ícones.

Observe que em Propriedades da suposição, há várias áreas que merecem ser mencionadas. A Figura 2.4 mostra as diferentes áreas:

- **Nome da suposição:** Esta é uma área opcional que permite inserir nomes exclusivos para as suposições com o intuito de ajudar a rastrear o que cada uma delas representa. Uma boa prática de modelagem é usar nomes curtos, mas precisos para as suposições.
- **Galeria de distribuições:** Esta área à esquerda mostra todas as distribuições diferentes disponíveis no software. Para alterar as exibições, clique com o botão direito do mouse na galeria e selecione ícones grandes, ícones pequenos ou lista. Há dezenas de distribuições disponíveis.
- **Parâmetros de entrada:** Dependendo da distribuição selecionada, são mostrados os parâmetros obrigatórios relevantes. Você pode inserir os parâmetros diretamente ou vinculá-los a células específicas na sua planilha. A inclusão de parâmetros por codificação ou digitação é útil quando se supõe que os parâmetros da suposição não se alteram. A vinculação a células da planilha é útil quando os parâmetros de entrada precisam ser visíveis ou podem ser alterados (clique no ícone de vínculo para vincular um parâmetro de entrada a uma célula da planilha).
- **Habilitar limite de dados:** Em geral, este recurso não é usado pela maioria dos analistas, mas existe para truncar as suposições de distribuição. Por exemplo, se uma distribuição normal for selecionada, os limites teóricos estarão entre o infinito negativo e o infinito positivo. No entanto, na prática, a variável simulada existe apenas dentro de uma faixa menor que pode ser inserida para truncar a distribuição de maneira apropriada.
- **Correlações:** É possível atribuir correlações de paridade a valores de entrada. Se forem necessárias suposições, marque a preferência *Ativar correlações* clicando em ***Risk Simulator* | *Editar perfil de simulação***. Consulte a seção sobre correlações mais adiante neste capítulo para obter mais detalhes sobre a atribuição de correlações e seus efeitos em um modelo. Observe que você pode truncar uma distribuição ou correlacioná-la a outra suposição, mas não as duas coisas.
- **Descrições resumidas:** Elas existem para cada uma das distribuições da galeria. As descrições resumidas explicam quando uma determinada distribuição é usada, bem como os parâmetros de entrada obrigatórios. Consulte a seção *Noções básicas sobre distribuições de probabilidade na simulação Monte Carlo* para obter detalhes sobre cada tipo de distribuição disponível no software.
- **Entrada normal e Entrada de percentil:** Estas opções permitem executar um teste rápido de *due diligence* nos valores de entrada. Por exemplo, se estiver definindo uma distribuição normal com algumas entradas de média e desvio padrão, você poderá clicar na entrada de percentil para ver quais são os 10º e 90º percentis correspondentes.
- **Habilitar simulações dinâmicas:** Por padrão, esta opção não é selecionada, mas para executar uma simulação multidimensional (por exemplo, se você vincular os parâmetros de entrada da suposição a outra célula que seja uma suposição, serão simuladas as entradas ou a simulação),

portanto, lembre-se de selecionar esta opção. A simulação dinâmica só funciona se as entradas estão vinculadas a outros valores de entrada alteráveis.

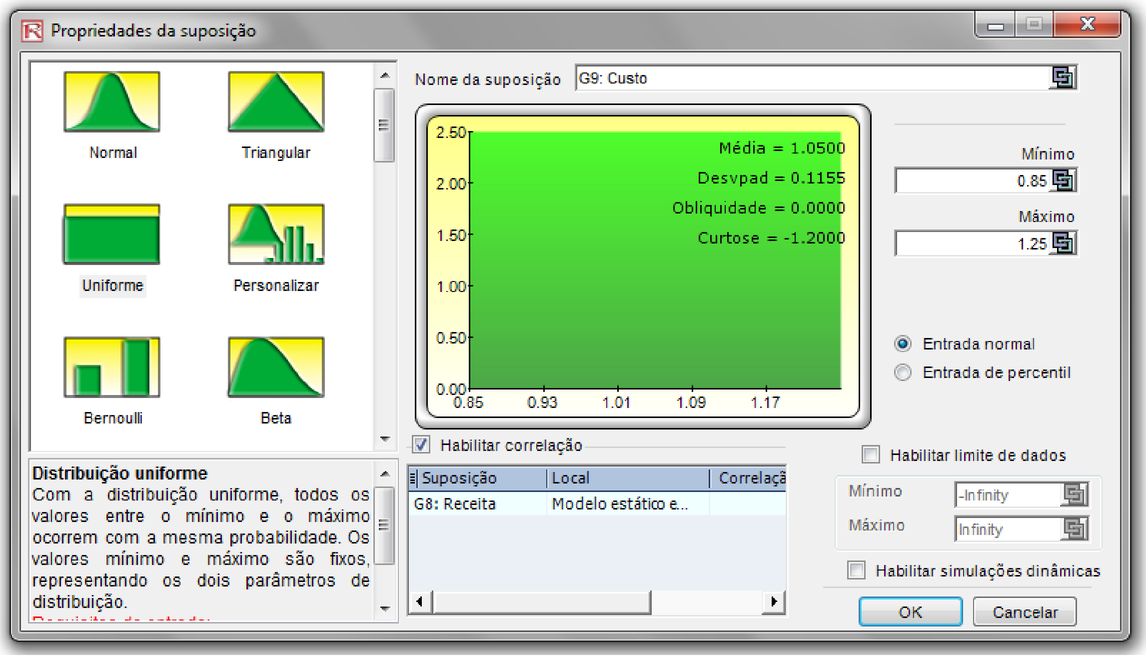


Figura 2.4 – Propriedades da suposição

Nota: Se você estiver acompanhando com o exemplo, continue definindo outra suposição na célula ***G9***. Desta vez, use a distribuição ***Uniforme*** com os valores mínimo ***0,85*** e máximo ***1,25***. Em seguida, defina os resultados de previsão na próxima etapa.

3. Definir resultados de previsão

A próxima etapa é definir os resultados de previsão no modelo. As previsões só podem ser definidas em células de resultado com equações ou funções. O processo de definição de previsão é descrito a seguir:

- Selecione a célula na qual deseja definir uma suposição (por exemplo, a célula ***G10*** no exemplo do modelo básico de simulação)
- Clique em ***Risk Simulator*** e selecione ***Definir resultado de previsão*** ou clique no ícone de definição de resultado de previsão na barra de ferramentas de ícones do Risk Simulator (Figura 1.3)
- Insira as informações relevantes e clique em ***OK***

Observe que você também pode definir resultados de previsão selecionando a célula na qual deseja definir a suposição e ***clicando com o botão direito do mouse***, acesse o menu de atalho do Risk Simulator para definir resultados de previsão.

A Figura 2.5 ilustra as propriedades da previsão.

- **Nome da previsão:** Especifique o nome da célula de previsão. Isso é importante porque, quando você tem um modelo grande com várias células de previsão, nomeá-las individualmente permite acessar os resultados rapidamente. Não subestime a importância dessa etapa simples. Uma boa prática de modelagem é usar nomes curtos, mas precisos para as suposições.
- **Precisão de previsão:** Em vez de confiar em uma estimativa aleatória de quantas tentativas devem ser executadas na simulação, você pode configurar controles de precisão e erro. Quando uma combinação de erro-precisão for obtida, a simulação será pausada e você será informado da precisão obtida, automatizando o processo de determinado do número de tentativas da simulação que não requer palpites sobre o número de tentativas necessárias. Releia a seção sobre controle de erro e precisão para obter detalhes mais específicos.
- **Mostrar janela de previsão:** Permite ao usuário mostrar ou não uma determinada janela de previsão. O padrão é sempre mostrar um gráfico de previsão.

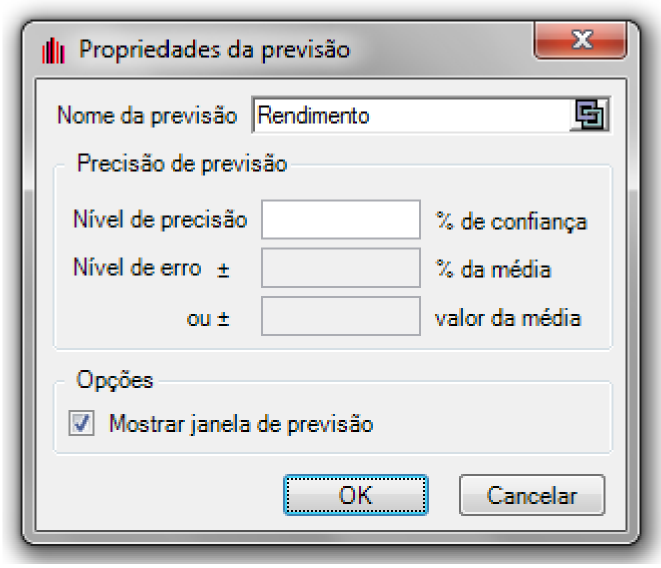


Figura 2.5 – Definir resultado de previsão

4. Executar a simulação

Se tudo estiver correto, clique em ***Risk Simulator | Executar simulação*** ou clique no ícone ***Executar*** na barra de ferramentas do Risk Simulator e a simulação continuará. Você também pode redefinir uma simulação depois que ela for executada para executá-la novamente (***Risk Simulator | Redefinir simulação*** ou o ícone Redefinir simulação na barra de ferramentas) ou pausá-la durante uma execução. Além disso, a função *etapa* (***Risk Simulator | Simulação por etapas*** ou o ícone Simulação por etapas na barra de ferramentas) permite simular uma única tentativa de cada vez, o que é útil para ensinar outras pessoas sobre simulação (por exemplo, você pode mostrar que em cada tentativa, todos os valores nas células de suposição são substituídos e o modelo inteiro é recalculado a cada vez). Você também pode acessar o menu Executar simulação clicando com o botão direito do mouse em qualquer lugar no modelo e selecionando Executar simulação.

O Risk Simulator também permite executar a simulação em uma velocidade extremamente rápida, chamada super-rápida. Para fazer isso, clique em ***Risk Simulator*** **|** ***Executar simulação super-rápida*** ou use o ícone Super-rápido. Observe a diferença de velocidade da simulação super-rápida quando executada. Para praticar, clique em ***Redefinir simulação*** e ***Editar perfil de simulação***, altere ***Número de tentativas*** para ***100.000*** e clique em ***Executar simulação super-rápida***. A execução demora apenas alguns segundos. No entanto, lembre-se de que a simulação super-rápida não será executada se o modelo tiver erros, código VBA (*Visual Basic for Applications*) ou vínculos para aplicativos ou fontes de dados externas. Nessas situações, você será notificado e a simulação será executada em velocidade normal. As simulações em velocidade normal sempre são executadas, mesmo que contenham erros, código VBA ou vínculos externos.

5. Interpretação dos resultados de previsão

A última etapa da simulação Monte Carlo é interpretar os gráficos de previsão resultantes. As Figuras 2.6 a 13 mostram gráficos de previsão e as estatísticas correspondentes geradas após a execução da simulação. Os itens a seguir geralmente são importantes para interpretar os resultados de uma simulação:

- **Gráfico de previsão:** O gráfico de previsão mostrado na Figura 2.6 é um histograma de probabilidade que mostra a contagem da frequência dos valores que ocorrem no número total de tentativas simuladas. A barra vertical mostra a frequência com que um valor*x* específico ocorre no total de tentativas, enquanto a frequência cumulativa (linha suave) mostra as probabilidades totais de todos os valores que ocorrem na previsão em *x* e abaixo dele.
- **Estatísticas de previsão:** As estatísticas de previsão mostradas na Figura 2.7 resumem a distribuição dos valores de previsão nos quatro momentos de uma distribuição. Consulte a seção *Noções básicas sobre estatísticas de previsão* para obter mais detalhes sobre o significado de algumas dessas estatísticas. Para alternar entre as guias Histograma e Estatísticas, pressione a barra de espaço.

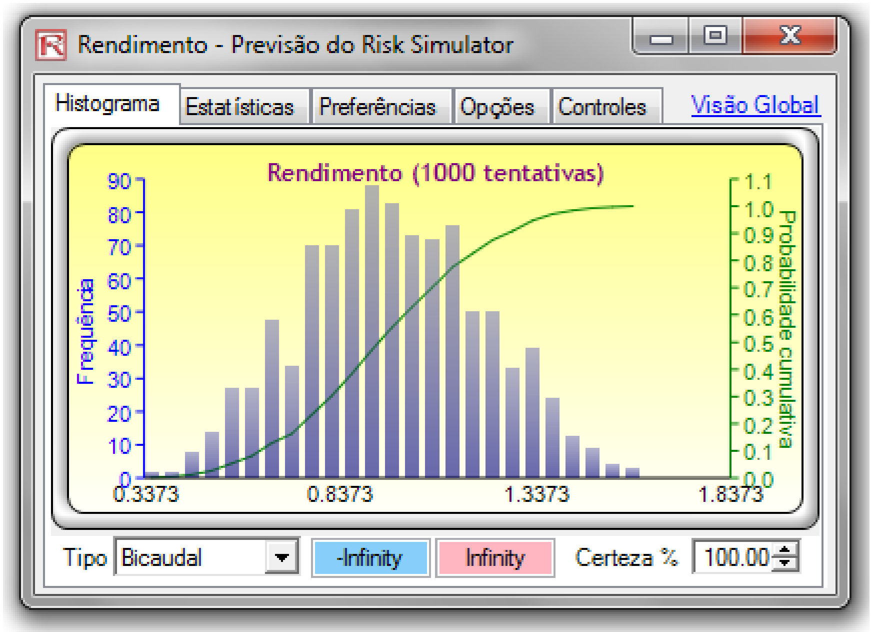
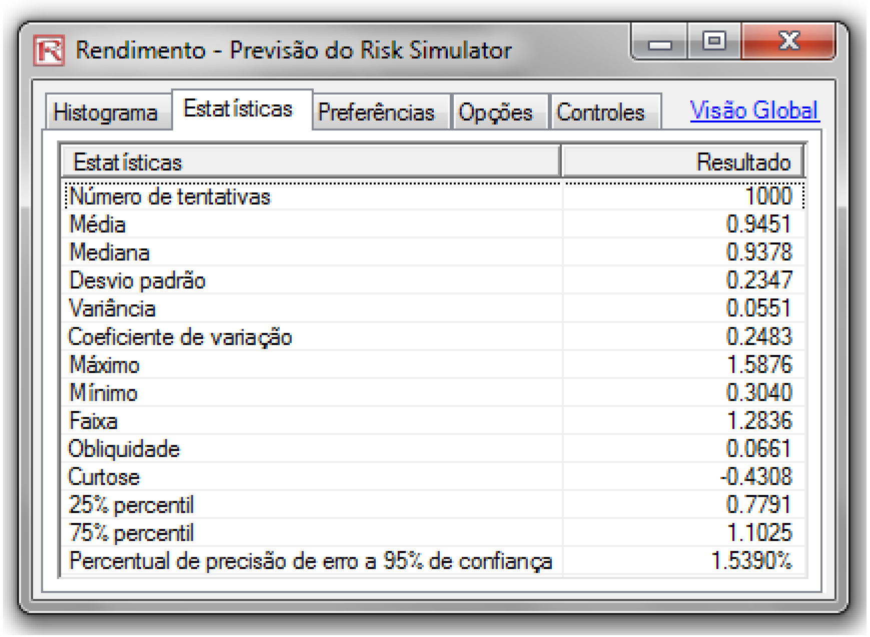


Figura 2.6 – Gráfico de previsão



Estatísticas	Resultado
Número de tentativas	1000
Média	0.9451
Mediana	0.9378
Desvio padrão	0.2347
Variância	0.0551
Coeficiente de variação	0.2483
Máximo	1.5876
Mínimo	0.3040
Faixa	1.2836
Obliquidade	0.0661
Curtose	-0.4308
25% percentil	0.7791
75% percentil	1.1025
Percentual de precisão de erro a 95% de confiança	1.5390%

Figura 2.7 – Estatísticas de previsão

- **Preferências:** A guia Preferências no gráfico de previsão permite alterar a aparência dos gráficos. Por exemplo, se ***Sempre visível*** estiver selecionado, os gráficos de previsão ficarão sempre visíveis, independentemente de qualquer outro software que esteja sendo executado no computador. ***Resolução do histograma*** permite alterar o número de compartimentos do histograma, de 5 a até 100 compartimentos. Além disso, a seção ***Atualização de dados*** permite controlar a velocidade de execução da simulação e a frequência em que o gráfico de previsão é atualizado. Ou seja, se você deseja que o gráfico de previsão seja atualizado em quase todas as tentativas, isso diminuirá a velocidade da simulação, porque é alocada mais memória para a atualização do gráfico do que para a execução da simulação. Isso é meramente uma preferência do usuário e não altera os resultados da simulação, apenas a velocidade com a qual ela será concluída. Para aumentar a velocidade da simulação, minimize o Excel enquanto a simulação estiver sendo executada. Isso reduz a quantidade de memória necessária para atualizar visivelmente a planilha do Excel e libera memória para executar a simulação. As opções ***Limpar tudo*** e ***Minimizar tudo*** controlam todos os gráficos de previsão abertos.

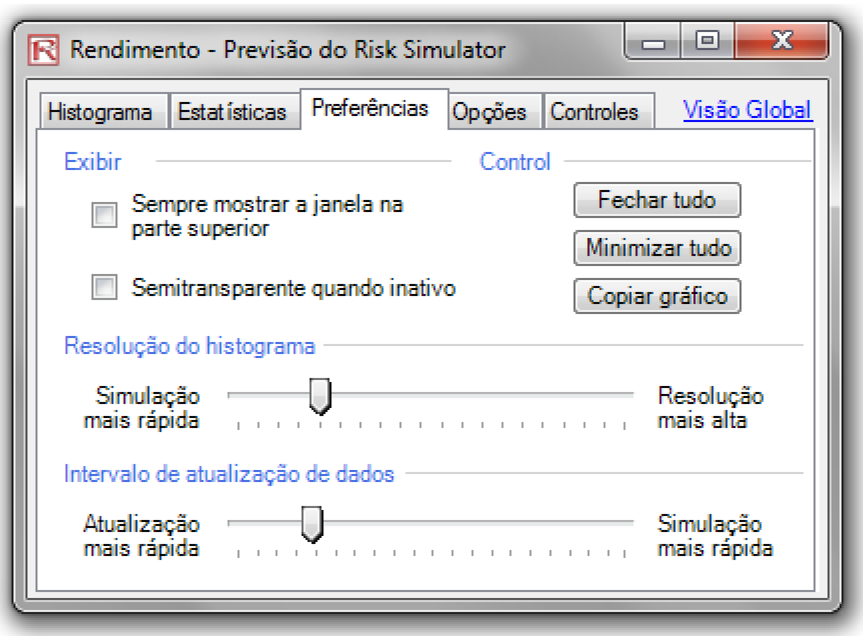


Figura 2.8 – Preferências do gráfico de previsão

- **Opções:** Esta opção do gráfico de previsão permite mostrar todos os dados de previsão ou filtrar valores que estejam dentro de um intervalo específico escolhido por você ou dentro de um desvio padrão de sua escolha. Além disso, o nível de precisão pode ser definido aqui para a previsão específica a fim de mostrar os níveis de erro na exibição das estatísticas. Consulte a seção sobre controle de erro e precisão para obter mais detalhes. *Mostrar estas estatísticas* é uma preferência do usuário se for necessário exibir as linhas de média, mediana, primeiro e quarto quartis (25º e 75º percentis) no gráfico de previsão.
- **Controles**: Esta guia tem todas as funcionalidades que permitem alterar o tipo, a cor, o tamanho, o zoom, a inclinação, 3D e outras características do gráfico de previsão. Ela também fornece gráficos sobrepostos (FDP, FDA) e executa ajuste da distribuição nos dados de previsão (consulte a seção Ajuste de dados para obter mais detalhes sobre essa metodologia).

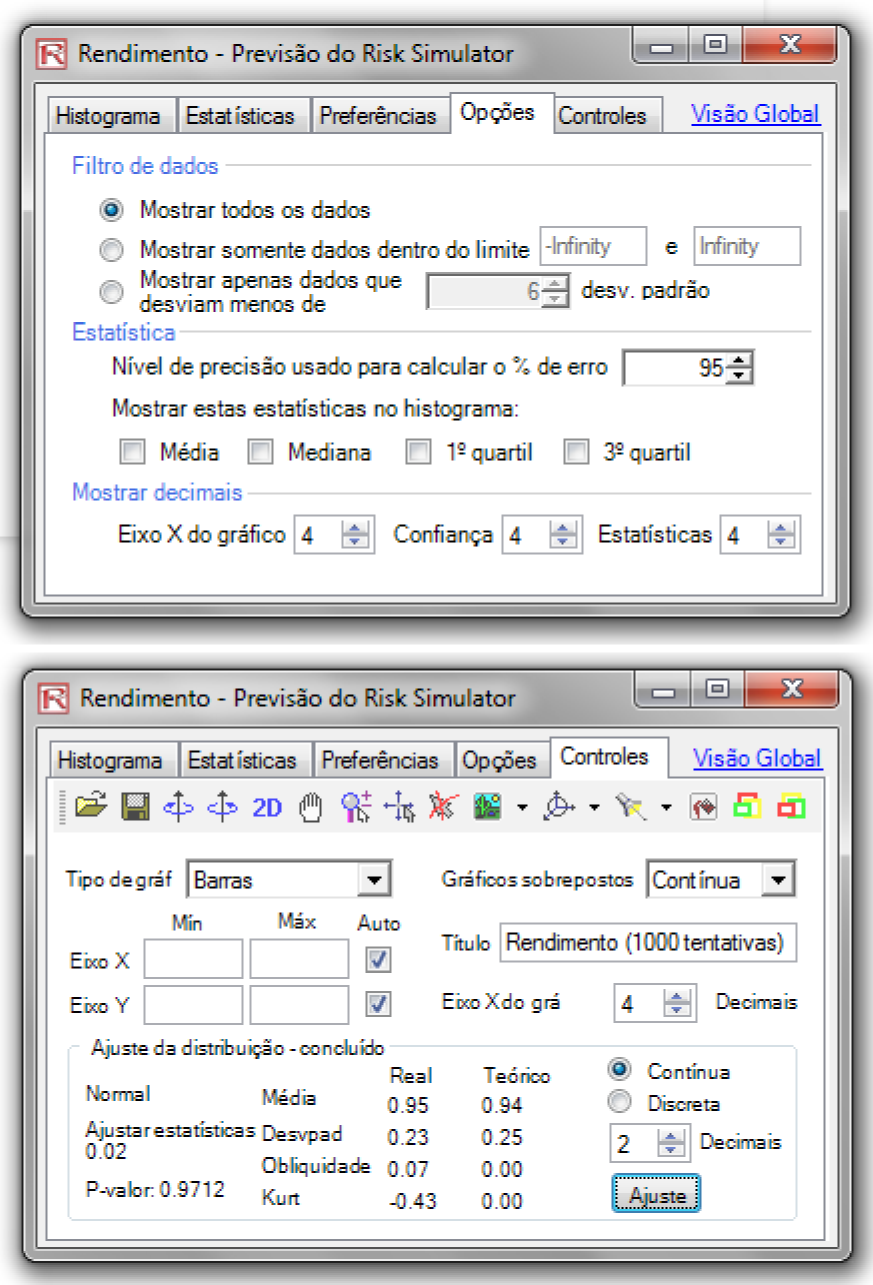


Figura 2.9 – Opções e controles do gráfico de previsão

Usar gráficos de previsão e intervalos de confiança

Nos gráficos de previsão, você pode determinar a probabilidade de ocorrência, denominada intervalos de confiança. Ou seja, dados dois valores, quais são as chances de o resultado estar entre um desses dois valores? A Figura 2.10 ilustra que há uma probabilidade de 90% de que o resultado final (nesse caso, o nível de renda) estará entre $0,5565 e $1,3423. Para obter o intervalo de confiança bicaudal, primeiro selecione o tipo ***Bicaudal***, insira o valor de certeza desejado (por exemplo, ***90***) e pressione ***TAB***. Os dois valores calculados que correspondem ao valor de certeza serão exibidos. Nesse exemplo, há uma probabilidade de 5% de que a renda esteja abaixo de $0,5565 e outros 5% de probabilidade de que a renda esteja acima de $1,3423. Ou seja, o intervalo de confiança bicaudal é um intervalo simétrico centrado na mediana ou no valor do 50º percentil. Dessa forma, ambas as caudas terão a mesma probabilidade.

Figura 2.10 – Intervalo de confiança bicaudal do gráfico de previsão

Como alternativa, pode-se calcular uma probabilidade unicaudal. A Figura 2.11 mostra a seleção da cauda esquerda com 95% de confiança (por exemplo, escolha o tipo ***Cauda esquerda ≤***, insira ***95*** como o nível de certeza e pressione a tecla ***TAB***). Isso significa que há uma probabilidade de 95% de o rendimento estar abaixo de $1,3423 ou uma probabilidade de 5% de que a renda esteja acima de $1,3423, correspondendo perfeitamente aos resultados vistos na Figura 2.10.

Figura 2.11 – Intervalo de confiança unicaudal do gráfico de previsão

Além de avaliar qual é o intervalo de confiança (ou seja, dado um nível de probabilidade e encontrando os valores de renda relevantes), você pode determinar a probabilidade de uma renda específica. Por exemplo, qual é a probabilidade de a renda ser menor ou igual a $1? Para fazer isso, selecione o tipo de probabilidade ***Cauda esquerda*** ≤, insira ***1*** na caixa de entrada de valor e pressione a tecla ***TAB***. A certeza correspondente será calculada. Nesse caso há uma probabilidade de 59,70% de a renda ser menor ou igual a $1.

Figura 2.12 – Avaliação da probabilidade do gráfico de previsão

Para garantir a integridade, selecione o tipo de probabilidade ***Cauda direita >***, insira o valor ***1*** na caixa de entrada de valor e pressione a tecla ***TAB***. A probabilidade resultante indicará a probabilidade de cauda direita depois do valor 1, ou seja, a probabilidade de a renda exceder $1. Nesse caso, percebe-se que há uma probabilidade de 40,30% de a renda exceder $1. A soma de 59,70% e 40,30% é, evidentemente, 100%: a probabilidade total abaixo da curva.

Figura 2.13 – Avaliação da probabilidade do gráfico de previsão

DICAS:

- Para redimensionar a janela de previsão, clique no canto inferior direito da janela e arraste. É recomendável redefinir a simulação atual sempre antes de executar uma simulação novamente. Para fazer isso, selecione ***Risk Simulator | Redefinir simulação***.
- Pressione a tecla ***TAB*** para atualizar o gráfico e os resultados quando inserir os valores de certeza ou valores de cauda direita e esquerda.
- Você também pode pressionar a ***barra de espaço*** repetidamente para alternar entre as guias Histograma, Estatísticas, Preferências, Opções e Controle.
- Além disso, se você clicar em ***Risk Simulator | Opções***, terá várias opções diferentes do Risk Simulator, incluindo permitir que o Risk Simulator seja iniciado sempre com o Excel ou somente quando você quiser (vá para ***Iniciar | Programas | Real Options Valuation | Risk Simulator | Risk Simulator***), alterar as ***cores das células*** de suposições e previsões, bem como ativar e desativar os ***comentários nas células*** , que permitem ver quais células são valores de entrada e quais são resultados de previsão e seus respectivos parâmetros e nomes de entrada. Dedique alguns momentos a experimentar os resultados e os vários atrativos do gráfico de previsão, especialmente a guia ***Controles***.

Controle de correlação e precisão

Noções básicas de correlações

O coeficiente de correlação é uma medida da força e da direção da relação entre duas variáveis, e aceita qualquer valor entre –1,0 e +1,0. Ou seja, o coeficiente de correlação pode ser decomposto em seu sinal (relação positiva ou negativa entre duas variáveis) e a magnitude ou força da relação (quanto mais alto o valor absoluto do coeficiente de correlação, mais forte a relação).

O coeficiente de correlação pode ser calculado de várias maneiras. A primeira abordagem é calcular manualmente a correlação *r* de duas variáveis *x* e *y* usando:

$$r_{x,y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}}$$

A segunda abordagem é usar a função *CORREL* do Excel. Por exemplo, se os 10 pontos de dados para *x* e *y* forem listados nas células A1:B10, então a função do Excel a ser usada será *CORREL (A1:A10, B1:B10)*.

A terceira abordagem é executar a *ferramenta Multiajuste* do *Risk Simulator*e a matriz de correlação resultante será calculada e exibida.

É importante observar que a correlação não significa causalidade. Duas variáveis aleatórias completamente não relacionadas podem exibir alguma correlação, mas isso não significa qualquer relação de causa entre as duas (por exemplo, a atividade solar e eventos no mercado de ações são correlacionadas, mas não existe nenhuma relação causal entre as duas).

Há dois tipos gerais de correlação: correlação paramétrica e não paramétrica. O coeficiente de correlação de Pearson é a medida de correlação mais comum. Normalmente se refere a ele simplesmente como coeficiente de correlação. No entanto, a correlação de Pearson é uma medida paramétrica, o que significa que requer que as duas variáveis correlacionadas tenham uma distribuição normal subjacente e que a relação entre as variáveis seja linear. Quando essas duas condições são violadas, que é frequentemente o caso na simulação Monte Carlo, os contrapontos não paramétricos se tornam mais importantes. A correlação de ranking de Spearman e Tau de Kendall são as duas alternativas. A correlação de Spearman é normalmente a mais usada e a mais apropriada quando aplicada no contexto da simulação Monte Carlo — não há dependência ou linearidade em distribuições normais, o que quer dizer que correlações entre variáveis diferentes com distribuições diferentes podem ser aplicadas. Para calcular a correlação de Spearman, primeiro classifique todos os valores das variáveis *x* e *y* e depois aplique o cálculo da correlação de Pearson.

No caso do Risk Simulator, a correlação usada é a correlação de ranking de classificação de Spearman não paramétrica. No entanto, para simplificar o processo de simulação e ser consistente com a função de correlação do Excel, as entradas de correlação obrigatórias são as do coeficiente de correlação de Pearson. O Risk Simulator aplicará seus próprios algoritmos para convertê-las na correlação de ranking de Spearman, simplificando o processo. Porém, para simplificar a interface do usuário, os usuários podem inserir a correlação de Pearson de produto e momento mais comum (por exemplo, as calculadas usando a função *CORREL* do Excel), enquanto nos códigos matemáticos, essas correlações simples são convertidas em correlações baseadas no ranking de Spearman para simulações de distribuição.

Aplicação de correlações no Risk Simulator

No Risk Simulator, as correlações podem ser aplicadas de diversas maneiras:

- Ao definir suposições (***Risk Simulator*** **|** ***Definir valores de entrada***), insira as correlações na grade da matriz de correlação na Galeria de distribuições.
- Com os dados existentes, execute a ferramenta Multiajuste (***Risk Simulator*** **|** ***Ferramentas*** **|** ***Ajuste da distribuição*** **|** ***Múltiplas variáveis***) para executar o ajuste da distribuição e obter a matriz de correlação entre variáveis de paridade. Se existir um perfil de simulação, as suposições ajustadas conterão automaticamente os valores relevantes de correlação.
- Para suposições existentes, clique em ***Risk Simulator*** **|** ***Ferramentas*** **|** ***Editar correlações*** para inserir as correlações de paridade de todas as suposições diretamente na interface.

Observe que a matriz de correlação deve ser definida positiva. Ou seja, a correlação deve ser matematicamente válida. Por exemplo, suponha que você esteja tentando correlacionar três variáveis: notas de alunos de pós-graduação em um determinado ano, o número de cervejas que eles consomem por semana e o número de horas que estudam por semana. Alguém pode supor que exista a seguinte correlação:

Notas e cerveja: – *Quanto mais eles bebem, mais baixas as notas (não aparecem para as provas)*
Notas e estudo: + *Quanto mais eles estudam, mais altas as notas*
Cerveja e estudo: – *Quanto mais eles bebem, menos estudam (bebem e vão a festas o tempo todo)*

No entanto, se você inserir uma correlação negativa entre notas e estudo, e supondo que os coeficientes de correlação têm muita magnitude, a matriz de correlação será definida não positiva. Isso desafiaria a lógica, os requisitos de correlação e a matemática da matriz. No entanto, alguns coeficientes pequenos podem funcionar ainda que com a lógica incorreta. Quando uma matriz de correlação não positiva ou incorreta for inserida, o Risk Simulator informará automaticamente e se oferecerá para ajustar essas correlações para algo que seja definido semipositivo e ainda manter a estrutura geral da relação de correlação (os mesmos sinais e as mesmas forças relativas).

Efeitos da correlação na simulação Monte Carlo

Embora os cálculos necessários para correlacionar variáveis em uma simulação sejam complexos, os efeitos resultantes são muito claros. A Figura 2.14 mostra um modelo de correlação simples (Modelo de efeitos de correlação, na pasta de exemplo). O cálculo da receita é simplesmente o preço multiplicado pela quantidade. O mesmo modelo é aplicado novamente para ausência de correlações, correlação positiva (+0,8) e correlação negativa (–0,8) entre preço e quantidade.

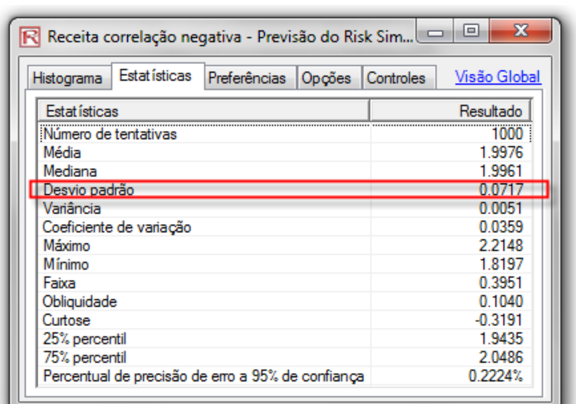
Modelo de correlação

	Sem correlação	Correlação positiva	Correlação negativa
Preço	$2.00	$2.00	$2.00
Quantidade	1.00	1.00	1.00
Receita	$2.00	$2.00	$2.00

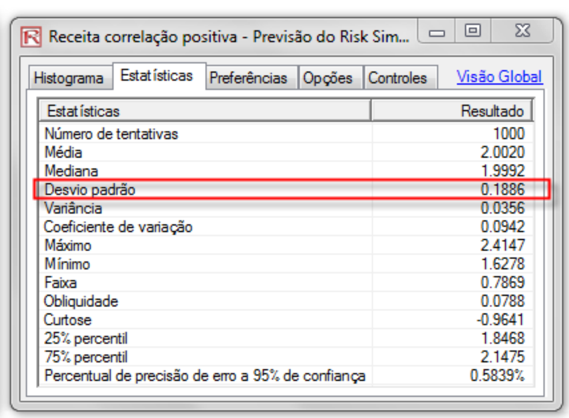
Para replicar este modelo, use estas suposições:
Os preços são definidos como distribuições triangulares (1,8; 2,0; 2,2), enquanto que as quantidades são definidas como distribuições uniformes (0,9; 1,1) com correlações definidas em 0,0, +0,8; -0,8 em 1.000 tentativas com valor de propagação de 123456.

Figura 2.14 – Modelo de correlação simples

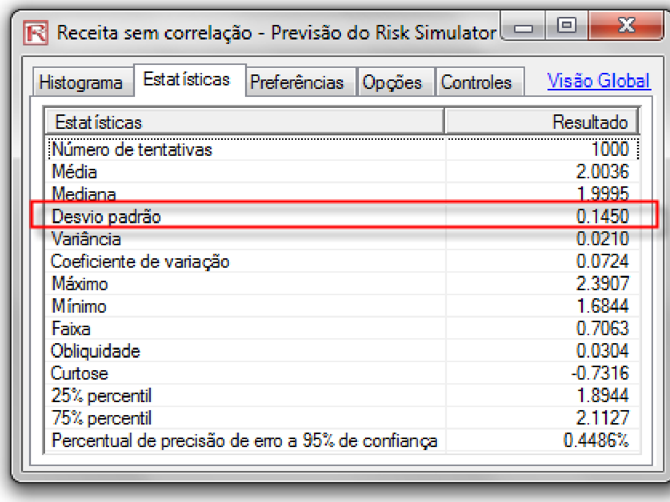
As estatísticas resultantes são mostradas na Figura 2.15. Observe que o desvio padrão do modelo sem correlações é 0,1450, comparado a 0,1886 para a correlação positiva e 0,0717 para a correlação negativa. Ou seja, para modelos simples, correlações negativas tendem a reduzir o *spread* médio da distribuição e criar uma distribuição de previsão mais concentrada se comparada a correlações positivas com *spreads* médios maiores. De qualquer forma, a média permanece relativamente estável. Isso significa que as correlações não afetam muito o valor esperado de projetos, mas podem reduzir ou aumentar o risco de um projeto.



Estatísticas	Resultado
Número de tentativas	1000
Média	1.9976
Mediana	1.9961
Desvio padrão	0.0717
Variância	0.0051
Coeficiente de variação	0.0359
Máximo	2.2148
Mínimo	1.8197
Faixa	0.3951
Obliquidade	0.1040
Curtose	-0.3191
25% percentil	1.9435
75% percentil	2.0486
Percentual de precisão de erro a 95% de confiança	0.2224%



Estatísticas	Resultado
Número de tentativas	1000
Média	2.0020
Mediana	1.9992
Desvio padrão	0.1886
Variância	0.0356
Coeficiente de variação	0.0942
Máximo	2.4147
Mínimo	1.6278
Faixa	0.7869
Obliquidade	0.0788
Curtose	-0.9641
25% percentil	1.8468
75% percentil	2.1475
Percentual de precisão de erro a 95% de confiança	0.5839%



Receita sem correlação - Previsão do Risk Simulator

Histograma | Estatísticas | Preferências | Opções | Controles | Visão Global

Estatísticas	Resultado
Número de tentativas	1000
Média	2.0036
Mediana	1.9995
Desvio padrão	0.1450
Variância	0.0210
Coeficiente de variação	0.0724
Máximo	2.3907
Mínimo	1.6844
Faixa	0.7063
Obliquidade	0.0304
Curtose	-0.7316
25% percentil	1.8944
75% percentil	2.1127
Percentual de precisão de erro a 95% de confiança	0.4486%

Figura 2.15 – Resultados de correlação

A Figura 2.16 ilustra os resultados depois de executar uma simulação, extrair os dados brutos das suposições e calcular as correlações entre as variáveis. A figura mostra que as suposições de entrada são recuperadas na simulação. Ou seja, você insere as correlações +0,8 e –0,8 e os valores simulados resultantes têm as mesmas correlações.

Estes são os valores brutos extraídos da simulação. Em seguida, eles são correlacionados para verificar se realmente as correlações inseridas nas suposições seriam aquelas efetivamente modeladas. O coeficiente da correlação de Pearson é uma correlação paramétrica linear e os resultados indicam que as correlações inseridas (+0,80 e -0,80) realmente são aquelas entre as variáveis. Consulte "Modeling Risk", de Dr. Johnathan Mun (Wiley 2006) para obter mais detalhes.

Correlação positiva de preço	Correlação positiva de quantidade		Correlação negativa de preço	Correlação negativa de quantidade	
1.95	0.91		1.89	1.06	
1.92	0.95		1.98	1.05	
2.02	1.04	Correlação de Pearson:	1.89	1.09	Correlação de Pearson:
2.04	1.03		1.88	1.04	
1.89	0.91	**0.80**	1.96	0.93	**-0.80**
1.98	1.05		2.02	0.93	
2.05	1.03		2.00	1.02	
1.87	0.91		1.86	1.04	
1.84	0.91		1.96	1.02	
2.06	1.03		1.90	1.02	

Figura 2.16 – Correlações recuperadas

Controle de precisão e erro

Uma ferramenta muito eficiente na simulação Monte Carlo é o controle de precisão. Por exemplo, qual é o número de tentativas, suficiente para executar em um modelo complexo? O controle de precisão elimina a adivinhação para estimar o número relevante de tentativas, permitindo que a simulação pare quando o nível de precisão pré-especificado for atingido.

A funcionalidade de controle de precisão permite definir o nível de precisão que você deseja que a previsão tenha. De modo geral, quanto mais tentativas forem calculadas, menor o intervalo de confiança e mais precisas as estatísticas. O recurso de controle de precisão do Risk Simulator usa a característica dos intervalos de confiança para determinar quando uma precisão específica de uma estatística foi atingida. Para cada previsão, você pode definir o intervalo de confiança específico para o nível de precisão.

Não confunda três termos muito diferentes: **erro**, **precisão** e **confiança**. Apesar de serem aparentemente semelhantes, os conceitos são muito diferentes um do outro. A explicação vem a seguir. Suponha que você é um fabricante de tortilhas que está interessado em descobrir quantas tortilhas quebradas existem, em média, em uma caixa de 100 tortilhas. Uma maneira de fazer isso é coletar uma amostra de caixas pré-empacotadas de 100 tortilhas, abri-las e contar quantas estão quebradas. Você fabrica 1 milhão de caixas por dia (essa é a sua *população*), mas você abre aleatoriamente 10 caixas (esse é o tamanho da sua *amostra*, também conhecido como o número de *tentativas* em uma simulação). O número de tortilhas quebradas em cada caixa é o seguinte: 24, 22, 4, 15, 33, 32, 4, 1, 45 e 2. O número médio calculado de tortilhas quebradas é 18,2. Com base nessas 10 amostras ou tentativas, a média é 18,2 unidades, enquanto com base na amostra, o intervalo de confiança de 80 por cento está entre 2 e 33 unidades (ou seja, 80 por cento do tempo, o número de tortilhas quebradas está entre 2 e 33 *com base nesse tamanho da amostra ou no número de tentativas executadas*). No entanto, como ter certeza de que 18,2 é a média correta? Seriam 10 tentativas o suficiente para estabelecer isso? O intervalo de confiança entre 2 e 33 é muito amplo e muito variável. Suponha que você queira um valor de média mais preciso, no qual o erro é ±2 tortilhas 90 por cento do tempo — isso significa que se você abrir *cada uma das* 1 milhão de caixas fabricadas em um dia, 900.000 delas terão em média ±2 tortilhas quebradas. Quantas caixas de tortilhas adicionais seriam necessárias na amostra (ou tentativas executadas) para obter esse nível de precisão? Aqui, as 2 tortilhas são o nível de erro enquanto os 90 por cento são o nível de precisão. Se um número suficiente de tentativas for executado, o intervalo de confiança de 90 por cento será idêntico ao nível de precisão de 90 por cento, no qual uma medida mais precisa da média é obtida de forma que em 90 por cento do tempo, o erro e, portanto, a confiança, será ±2 tortilhas. Como exemplo, digamos que a média seja 20 unidades, então o intervalo de confiança de 90 por cento estará entre 18 e 22 unidades, ou seja, o intervalo é preciso 90 por cento do tempo, em que ao abrir todas as 1 milhão de caixas, 900.000 delas terão entre 18 e 22 tortilhas quebradas. O número de tentativas necessárias para atingir essa precisão é baseado na equação de erro de amostragem $\bar{x} \pm Z\frac{s}{\sqrt{n}}$, onde $Z\frac{s}{\sqrt{n}}$ é o erro de 2 tortilhas, $\bar{x}$ é a média da amostra, Z é a pontuação Znormal-padrão obtida do nível de precisão de 90 por cento, s é o desvio padrão da amostra e n é o número de tentativas necessárias para atingir esse nível de erro com a precisão especificada. As Figuras 2.17 e 2.18 ilustram como o controle de precisão pode ser executado em várias previsões simuladas no Risk Simulator. Com esse recurso, o usuário não precisa decidir quantas tentativas executar em uma simulação e elimina todas as possibilidades de adivinhação. A Figura 2.17 ilustra o gráfico de previsão com um nível de previsão definido como 95%. Esse valor pode ser alterado e será refletido na guia Estatísticas como mostra a Figura 2.18.

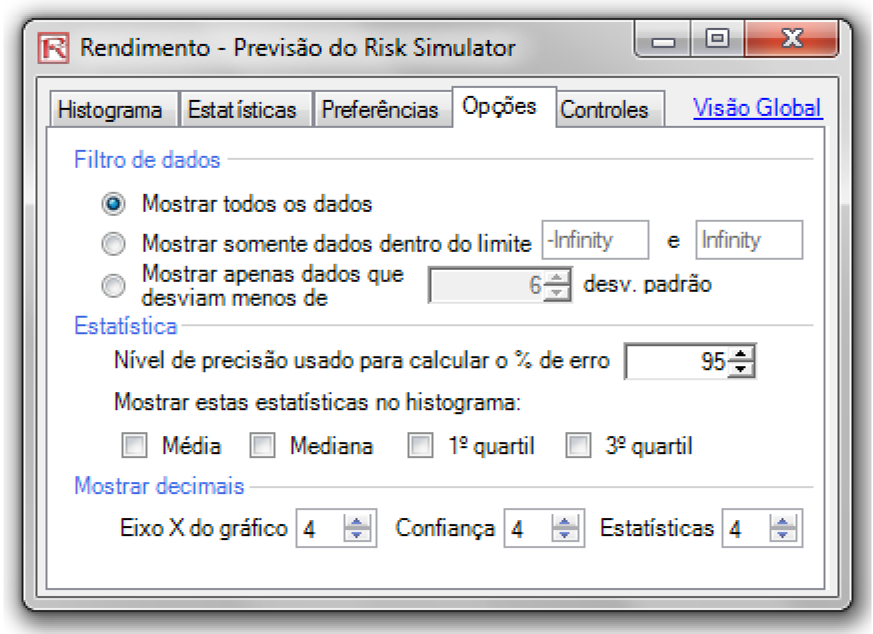


Figura 2.17 – Configuração do nível de precisão da previsão

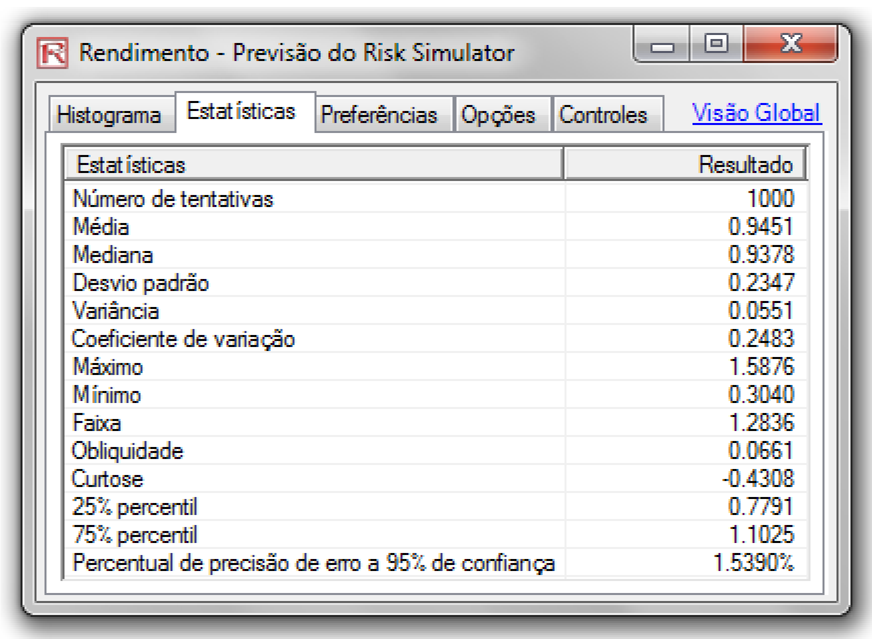


Figura 2.18 – Cálculo do erro

Noções básicas sobre as estatísticas de previsão

A maioria das distribuições pode ser definida em até quatro momentos. O primeiro momento descreve sua localização ou tendência central (retorno esperado), o segundo momento descreve sua largura ou *spread* (riscos), o terceiro momento descreve sua obliquidade direcional (eventos mais prováveis) e o quarto momento descreve os picos ou a espessura das caudas (perdas ou ganhos catastróficos). Todos os quatro momentos devem ser calculados na prática e interpretados para fornecer uma visão mais abrangente do projeto analisado. O Risk Simulator fornece os resultados de todos os quatro momentos em sua exibição *Estatísticas* nos gráficos de previsão.

Medida do centro da distribuição: primeiro momento

O primeiro momento de uma distribuição mede a taxa de retorno esperada de um projeto específico. Ele mede a localização dos cenários do projeto e os possíveis resultados em média. As estatísticas comuns para o primeiro momento incluem a média, a mediana (centro de uma distribuição) e a moda (valor que ocorre mais normalmente). A Figura 2.19 ilustra o primeiro momento. Nesse caso, o primeiro momento dessa distribuição é medido pelo valor médio (μ).

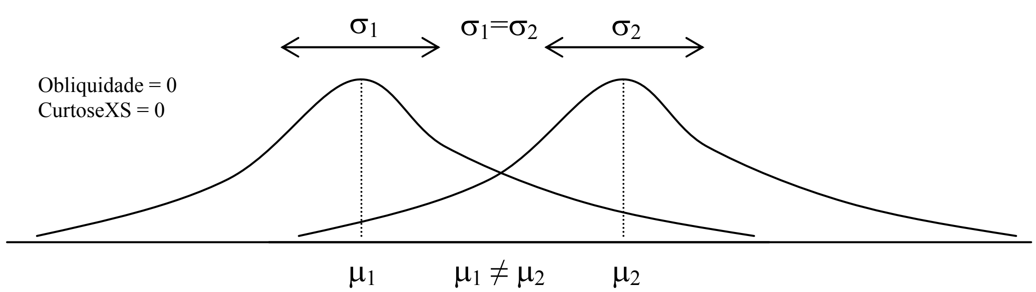


Figura 2.19 – Primeiro momento

Medida do spread da distribuição: segundo momento

O segundo momento mede o *spread* de uma distribuição, que é uma medida de risco. O *spread* ou a largura de uma distribuição mede a variabilidade de uma variável, ou seja, o potencial que a variável tem de cair em regiões diferentes da distribuição, em outras palavras, os cenários potenciais de resultados. A Figura 2.20 ilustra duas distribuições com primeiros momentos idênticos (médias idênticas), mas segundos momentos ou riscos muito diferentes. A Figura 2.21 evidencia isso. Por exemplo, suponha que existem duas ações e os movimentos da primeira ação (linha mais escura) com a menor flutuação são comparados com os movimentos da segunda ação (linha pontilhada) com uma flutuação de preço muito maior. Nitidamente um investidor veria a ação com a maior flutuação como sendo mais arriscada, pois os resultados da ação mais arriscada são relativamente mais imprevisíveis do que os da ação menos arriscada. O eixo vertical na Figura 2.21 mede os preços das ações, assim, a ação mais arriscada tem uma faixa mais ampla de resultados potenciais. Essa faixa é representada como a largura de uma distribuição (o eixo horizontal) na Figura 2.20, no qual a distribuição mais ampla representa o ativo mais arriscado. Assim, a largura ou o *spread* de uma distribuição mede os riscos de uma variável.

Observe que na Figura 2.20, as duas distribuições têm um primeiro momento ou tendências centrais idênticas, mas nitidamente as distribuições são muito diferentes. Essa diferença na largura da distribuição é mensurável. Matemática e estatisticamente, a largura ou o risco de uma variável pode ser medido por meio de várias estatísticas diferentes, incluindo a faixa, o desvio padrão (σ), a variância, o coeficiente de variação e os percentis.

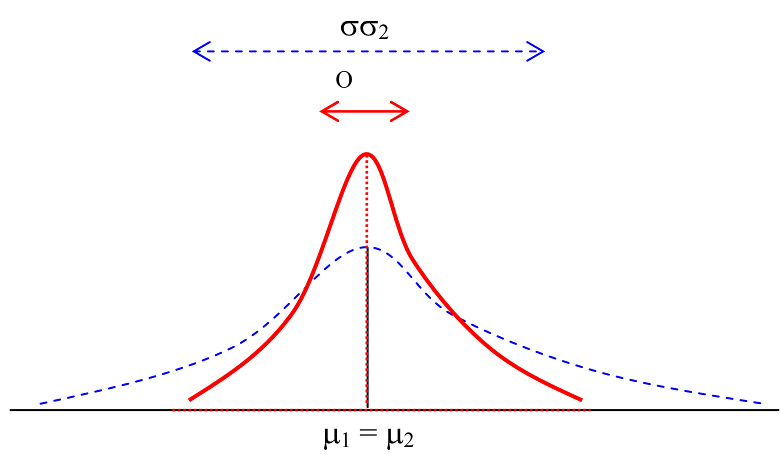


Figura 2.20 – Segundo momento

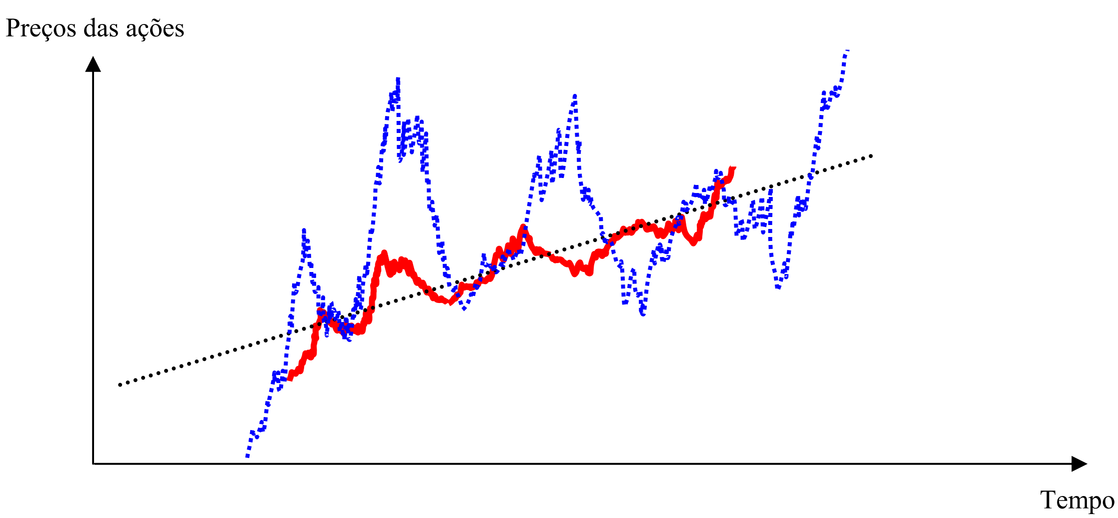


Figura 2.21 – Flutuações no preço das ações

Medida da obliquidade da distribuição: terceiro momento

O terceiro momento mede a obliquidade de uma distribuição, ou seja, como a distribuição é arrastada para um lado ou para o outro. A Figura 2.22 ilustra uma obliquidade negativa ou para a esquerda (a cauda da distribuição aponta para a esquerda) e a Figura 2.23 mostra uma obliquidade positiva ou para a direita (a cauda da distribuição aponta para a direita). A média é sempre inclinada na direção da cauda da distribuição enquanto a mediana permanece constante. Outra maneira de analisar isso é que a média se move, mas o desvio padrão, a variância ou a largura pode ainda permanecer constante. Se o terceiro momento não for considerado, analisando apenas os retornos esperados (por exemplo, a mediana ou a média) e o risco (desvio padrão), um projeto positivamente inclinado poderia ser incorretamente escolhido! Por exemplo, se o eixo horizontal representa a receita líquida de um projeto, poderia se preferir uma distribuição com obliquidade negativa, ou inclinada para esquerda, já que há uma probabilidade mais alta de maiores retornos (Figura 2.22) se comparada a uma probabilidade mais alta de retornos de nível mais baixo (Figura 2.23). Portanto, em uma distribuição inclinada, a mediana é uma medida melhor dos retornos, já que as medianas das Figuras 2.22 e 2.23 são idênticas, os riscos são idênticos e, portanto, um projeto com uma distribuição negativamente inclinada do lucro líquido é uma melhor opção. A desconsideração da obliquidade de distribuição de um projeto pode ocasionar a escolha do projeto errado. Por exemplo, dois projetos podem ter o primeiro e o segundo momentos idênticos, ou seja, ambos podem ter perfis de retorno e risco idênticos, mas suas obliquidades de distribuição podem ser bastante diferentes.

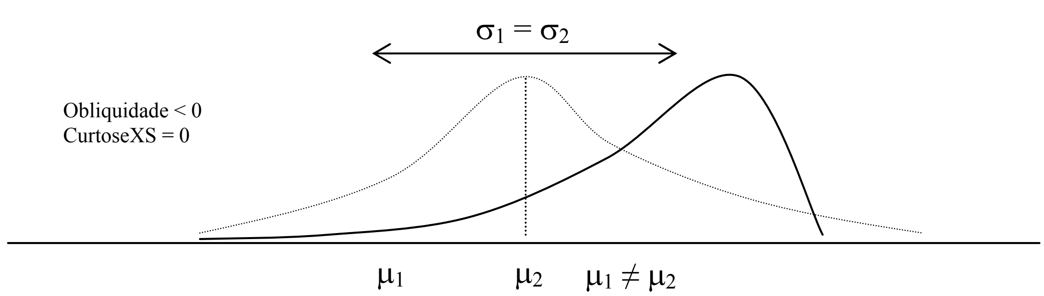


Figura 2.22 – Terceiro momento (obliquidade para a esquerda)

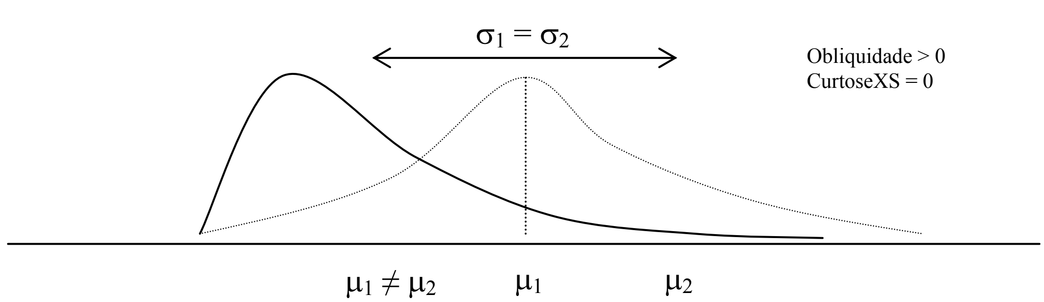


Figura 2.23 – Terceiro momento (obliquidade para a direita)

Medida dos eventos de cauda catastróficos em uma distribuição: quarto momento

O quarto momento, ou curtose, mede os picos de uma distribuição. A Figura 2.24 ilustra esse efeito. O plano de fundo (denotado pela linha pontilhada) é uma distribuição normal com curtose 3,0 ou curtose excessiva (CurtoseXS) 0.0. Os resultados do Risk Simulator mostram o valor de CurtoseXS, usando 0 como o nível normal de curtose, o que significa que CurtoseXS negativa indica caudas mais achatadas (distribuições platicúrticas como a distribuição uniforme), enquanto valores positivos indicam caudas mais pesadas (distribuições leptocúrticas como a distribuição T de Student ou a lognormal). A distribuição descrita pela linha em negrito tem uma curtose excessiva mais alta, logo a área abaixo da curva é mais espessa nas caudas com área menor no centro. Essa condição tem impactos importantes sobre a análise de risco. Por exemplo, nas duas distribuições na Figura 2.24, os primeiros três momentos (média, desvio padrão e obliquidade) podem ser idênticos, mas o quarto momento (curtose) é diferente. Essa condição significa que, apesar de os retornos e riscos serem idênticos, as probabilidades de eventos extremos e catastróficos (grandes perdas ou ganhos potenciais) ocorrerem são mais altas para uma distribuição de curtose mais alta (por exemplo, retornos do mercado acionário são leptocúrticos ou têm curtose alta). Ignorar a curtose de um projeto pode ser prejudicial. Geralmente um valor mais alto de curtose excessiva indica que a os riscos negativos são mais altos, por exemplo, o valor em risco (VaR) de um projeto pode ser significativo.

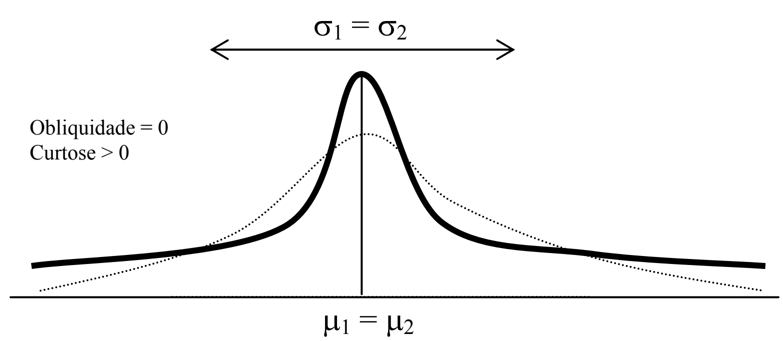


Figura 2.24 – Quarto momento

Funções dos momentos

Você já se perguntou por que essas estatísticas de risco são chamadas de momentos? No jargão matemático, momento significa elevado à potência de algum valor. Em outras palavras, o terceiro momento significa que em uma equação, três é mais provavelmente a maior potência. Na verdade, as equações abaixo ilustram as funções e aplicações matemáticas de alguns momentos para uma estatística de amostra. Por exemplo, observe que a potência mais alta para a média do primeiro momento é um, o desvio padrão do segundo momento é dois, a obliquidade do terceiro momento é três e a potência mais alta para o quarto momento é quatro.
Primeiro momento: média aritmética ou média simples (amostra)

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$ A função equivalente do Excel é MÉDIA

Segundo momento: desvio padrão (amostra)

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$ A função equivalente do Excel é DESVPAD para o desvio padrão de uma amostra

A função equivalente do Excel é DESVPADP para o desvio padrão de uma população

Terceiro momento: obliquidade

$$obliquidade = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \frac{(x_i - \overline{x})^3}{s}$$ A função equivalente do Excel é INCLINAÇÃO

Quarto momento: curtose

$$curtose = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \frac{(x_i - \overline{x})^4}{s} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

A função equivalente do Excel CURT

Noções básicas sobre distribuições de probabilidade na simulação Monte Carlo

Esta seção demonstra a força da simulação Monte Carlo, mas antes de começar a simulação é necessário entender o conceito de distribuições de probabilidade. Para começar a entender probabilidade, considere este exemplo: você quer analisar a distribuição de salários sem isenções em um departamento de uma grande empresa. Primeiro, você coleta dados brutos — neste caso, os salários de cada funcionário não-isento no departamento. Depois, organiza os dados em um formato válido e plota os dados como uma distribuição de frequência em um gráfico. Para criar uma distribuição de frequência, divida os salários em intervalos de grupos e liste esses intervalos no eixo horizontal do gráfico. Em seguida, liste o número ou a frequência dos funcionários em cada intervalo no eixo vertical do gráfico. Agora, é possível ver a distribuição de salários sem isenções no departamento.

O gráfico na Figura 2.25 revela que a maioria dos funcionário (aproximadamente 60 de um total de 180) ganham entre $7,00 e $9,00 por hora.

Figura 2.25 – Histograma de frequência I

Você pode representar esses dados como uma distribuição de probabilidade. Uma distribuição de probabilidade mostra o número de funcionários em cada intervalo como uma fração do número total de funcionários. Para criar uma distribuição de probabilidade, divida o número de funcionários em cada intervalo pelo número total de funcionários e liste os resultados no eixo vertical do gráfico.

O gráfico na Figura 2.26 mostra o número de funcionários em cada grupo de salário como uma fração do total de funcionários. Você pode estimar a probabilidade de que um funcionário escolhido ao acaso do grupo total ganhe um salário pertencente a um determinado intervalo. Por exemplo, supondo que as condições eram as mesmas no momento em que foi criada a amostra, a probabilidade é de 0,33 (um em três) de que um funcionário escolhido ao acaso do total de funcionários ganhe entre $8,00 e $8,50 por hora.

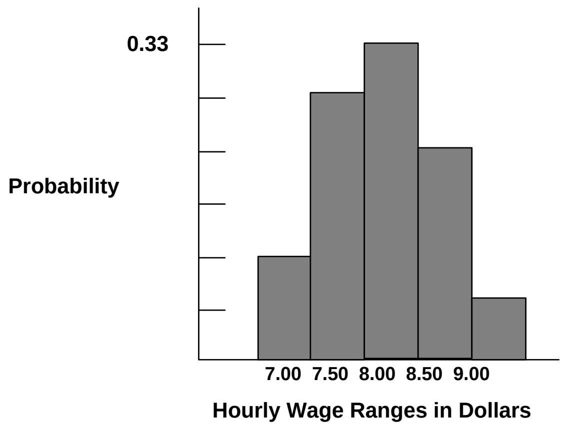


Figura 2.26 – Histograma de frequência II

As distribuições de probabilidade podem ser discretas ou contínuas. *As distribuições de probabilidade discretas* descrevem valores distintos, geralmente inteiros, sem valores intermediários e são mostradas como uma série de barras verticais. Uma distribuição discreta, por exemplo, poderia descrever o número de caras em quatro rodadas de cara ou coroa usando uma moeda como 0, 1, 2, 3 ou 4. As *distribuições contínuas* são abstrações matemáticas, porque supõem a existência de todos os valores intermediários possíveis entre dois números. Ou seja, uma distribuição contínua presume que há um número infinito de valores entre dois pontos na distribuição. No entanto, em muitas situações, você pode usar uma distribuição contínua para fazer a aproximação de uma distribuição discreta, ainda que o modelo contínuo não descreva a situação de maneira exata.

Seleção da distribuição de probabilidade apropriada

A plotagem dos dados é um guia para selecionar uma distribuição de probabilidade. As seguintes etapas fornecem outro processo para selecionar distribuições de probabilidade que melhor descrevem as variáveis incertas nas suas planilhas.

Para selecionar a distribuição de probabilidade correta, use estas etapas:

- Observe a variável em questão. Liste tudo o que você sabe sobre as condições que envolvem a variável. Pode ser possível coletar informações valiosas sobre a variável incerta a partir dos dados históricos. Se não houver dados históricos disponíveis, use seu próprio julgamento, com base na sua experiência, listando tudo o que você sabe sobre a variável incerta.
- Revise as descrições de distribuições de probabilidade.
- Selecione a distribuição que caracteriza a variável. Uma distribuição caracteriza uma variável quando as condições da distribuição correspondem àquelas da variável.

Simulação Monte Carlo

A simulação Monte Carlo na sua forma mais simples é um gerador de números aleatórios útil para previsões, estimativas e análises de risco. Uma simulação calcula numerosos cenários de um modelo escolhendo valores repetidamente de uma *distribuição de probabilidade* predefinida pelo usuário para as variáveis incertas e usando esses valores no modelo. Todos esses cenários produzem resultados associados em um modelo, em que cada cenário pode ter uma *previsão*. As previsões são eventos (normalmente com fórmulas ou funções) que você define como resultados importantes do modelo. Em geral, são eventos como totais, lucro líquido ou despesas brutas.

De maneira simplificada, pense na simulação Monte Carlo como escolher bolas de golfe de uma grande cesta repetidamente com reposições. O tamanho e a forma da cesta depende das ***suposições*** da distribuição (por exemplo, uma distribuição normal com uma média 100 e um desvio padrão 10, comparada com uma distribuição uniforme ou triangular) em que algumas cestas são mais fundas ou mais simétricas do que outras, permitindo que certas bolas sejam tiradas com mais frequência do que outras. O número de bolas tiradas repetidamente depende do número de ***tentativas*** simuladas. Para um modelo grande com várias suposições relacionadas, imagine o modelo grande como uma cesta muito grande, com muitas cestas menores dentro. Cada cesta menor tem seu próprio conjunto de bolas que estão se movimentando. Algumas vezes essas cestas menores estão juntas umas das outras (se houver uma ***correlação*** entre as variáveis) e as bolas de golfe estão se movimentando em pares enquanto outras se movimentam independentemente umas das outras. As bolas que são retiradas de cada vez dessas interações dentro do modelo (a cesta maior) são tabuladas e registradas, fornecendo o ***resultado de previsão*** da simulação. Com a simulação Monte Carlo, o Risk Simulator gera valores aleatórios para cada distribuição de probabilidade de suposição que são totalmente independentes. Em outras palavras, o valor aleatório selecionado para uma tentativa não afeta o próximo valor aleatório gerado. Use a amostragem Monte Carlo quando quiser simular cenários hipotéticos verossímeis para seu modelo de planilha.

Distribuições discretas

Veja a seguir uma lista detalhada dos diferentes tipos de distribuições de probabilidade que podem ser usadas na simulação Monte Carlo. A listagem está incluída neste apêndice apenas para referência.

Distribuição Bernoulli ou distribuição sim/não

A distribuição Bernoulli é uma distribuição discreta com dois resultados: cara ou coroa, êxito ou fracasso, 0(zero) ou 1(um). A distribuição Bernoulli é a distribuição binomial com uma tentativa e pode ser usada para simular condições do tipo sim/não ou êxito/fracasso. Essa distribuição é o fundamento de outras distribuições mais complexas. Por exemplo:

- Distribuição binomial: distribuição Bernoulli com um número mais alto de tentativas totais *n* que calcula a probabilidade de *x* êxitos no número total de tentativas.
- Distribuição geométrica: distribuição Bernoulli com um número mais alto de tentativas que calcula o número de fracassos necessários antes do primeiro sucesso.
- Distribuição binomial negativa: distribuição Bernoulli com um número mais alto de tentativas que calcula o número de fracassos antes do *X*ésimo sucesso.

Os construtos matemáticos da distribuição Bernoulli são os seguintes:

$$P(n) = \begin{cases} 1-p & \text{para } x=0 \\ p & \text{para } x=1 \end{cases}$$

ou

$$P(n) = p^x (1-p)^{1-x}$$

$$\text{média} = p$$

$$\text{desvio padrão} = \sqrt{p(1-p)}$$

$$\text{obliquidade} = \frac{1-2p}{\sqrt{p(1-p)}}$$

$$\text{curtose excessiva} = \frac{6p^2 - 6p + 1}{p(1-p)}$$

Probabilidade de êxito (p) é o único parâmetro de distribuição. Além disso, é importante observar que há apenas uma tentativa na distribuição Bernoulli e o valor simulado resultante será 0 ou 1.
Requisitos de entrada:
Probabilidade de êxito > 0 e < 1 (ou seja, $0{,}0001 \leq p \leq 0{,}9999$)

Distribuição binomial

A distribuição binomial descreve quantas vezes um determinado evento ocorre em um número fixo de tentativas, como o número de caras em 10 rodadas de cara ou coroa ou o número de itens defeituosos entre 50 itens escolhidos

Condições

As três condições subjacentes à distribuição binomial são:

- Para cada tentativa, apenas dois resultados são possíveis e mutuamente exclusivos.
- As tentativas são independentes: o que acontece na primeira tentativa não afeta a próxima tentativa.
- A probabilidade de um evento ocorrer permanece a mesma de uma tentativa para outra.

Os construtos matemáticos da distribuição binomial são os seguintes:

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)} \quad \text{para } n > 0;\ x = 0,\ 1,\ 2,\ \ldots n;\ \text{e } 0 < p < 1$$

$$\text{média} = np$$

$$\text{desvio padrão } = \sqrt{np(1-p)}$$

$$\text{obliquidade} = \frac{1-2p}{\sqrt{np(1-p)}}$$

$$\text{curtose excessiva} = \frac{6p^2 - 6p + 1}{np(1-p)}$$

A probabilidade de êxito (*p*) e o número inteiro do total de tentativas (*n*) são os parâmetros de distribuição. O número de tentativas bem-sucedidas é indicado por *x*. É importante observar que a probabilidade de êxito (*p*) de 0 ou 1 é uma condição trivial e não requer simulação, portanto, não é permitida no software.

Requisitos de entrada:

Probabilidade de êxito > 0 e < 1 (ou seja, $0{,}0001 \leq p \leq 0{,}9999$)

Número de tentativas ≥ 1 ou inteiros positivos e ≤ 1000 (para tentativas maiores, use a distribuição normal com a média binomial calculada relevante e o desvio padrão como os parâmetros da distribuição normal).

Distribuição binomial negativa

A distribuição binomial negativa é útil para modelar a distribuição do número de tentativas adicionais necessárias além do número de ocorrências bem-sucedidas necessárias (*R*). Por exemplo, para fechar um total de 10 oportunidades de vendas, quantas chamadas de vendas extras você precisaria fazer além de 10 chamadas, dada alguma probabilidade de êxito em cada chamada? O eixo x mostra o número de chamadas adicionais necessárias ou o número de chamadas fracassadas. O número de tentativas não é fixo, as tentativas continuam até o R-ésimo êxito e a probabilidade de sucesso é a mesma em cada tentativa. A probabilidade de êxito (*p*) e o número de êxitos necessários (*R*) são os parâmetros de distribuição. É essencialmente uma *superdistribuição* das distribuições binomial e geométrica. A distribuição mostra as probabilidades de cada número de tentativas excessivas de *R* em produzir o êxito necessário *R*.

Condições

As três condições subjacentes à distribuição binomial negativa são:

- O número de tentativas não é fixo.
- As tentativas continuam até o *r*-ésimo êxito.
- A probabilidade de êxito é a mesma de uma tentativa para outra.

Os construtos matemáticos da distribuição binomial negativa são os seguintes:

$$P(x)=\frac{(x+r-1)!}{(r-1)!x!}p^r(1-p)^x \quad \text{para } x=r,\ \mathrm{r}+1,\ ...;\ \text{e } 0<p<1$$

$$\text{média } =\frac{r(1-p)}{p}$$

$$\text{desvio padrão } =\sqrt{\frac{r(1-p)}{p^2}}$$

$$\text{obliquidade} = \frac{2-p}{\sqrt{r(1-p)}}$$

$$\text{curtose excessiva} = \frac{p^2-6p+6}{r(1-p)}$$

A probabilidade de êxito (*p*) e o número de êxitos necessários (*R*) são os parâmetros de distribuição.

Requisitos de entrada:

Êxitos necessários devem ser inteiros positivos > 0 e < 8000.

Probabilidade de êxito > 0 e < 1 (ou seja, 0,0001 ≤ *p* ≤ 0,9999). É importante observar que a probabilidade de êxito (*p*) de 0 ou 1 é uma condição trivial e não requer simulação, portanto, não é permitida no software.

Distribuição geométrica

A distribuição geométrica descreve o número de tentativas até a primeira ocorrência bem-sucedida, como o número de vezes que você precisa girar uma roleta antes de ganhar.

Condições

As três condições subjacentes à distribuição geométrica são:

- O número de tentativas não é fixo.
- As tentativas continuam até o primeiro êxito.
- A probabilidade de êxito é a mesma de uma tentativa para outra.

Os construtos matemáticos da distribuição geométrica são os seguintes:

$$P(x)=p(1-p)^{x-1} \quad \text{para } 0<p<1 \text{ e } x=1,\ 2,\ ...,\ n$$

$$\text{média} = \frac{1}{p} - 1$$

$$\text{desvio padrão} = \sqrt{\frac{1-p}{p^2}}$$

$$\text{obliquidade} = \frac{2-p}{\sqrt{1-p}}$$

$$\text{curtose excessiva} = \frac{p^2 - 6p + 6}{1-p}$$

Probabilidade de êxito (p) é o único parâmetro de distribuição. O número de tentativas bem-sucedidas simuladas é indicado por x, que só pode conter inteiros positivos.

Requisitos de entrada:

Probabilidade de êxito > 0 e < 1 (ou seja, $0{,}0001 \leq p \leq 0{,}9999$). É importante observar que a probabilidade de êxito (p) de 0 ou 1 é uma condição trivial e não requer simulação, portanto, não é permitida no software.

Distribuição hipergeométrica

A distribuição hipergeométrica é semelhante à distribuição binomial, já que ambas descrevem quantas vezes um determinado evento ocorre dado um certo número fixo de tentativas. A diferença é que as tentativas da distribuição binomial são independentes, enquanto as tentativas da distribuição hipergeométrica mudam a probabilidade de cada tentativa subsequente e são chamadas "tentativas sem substituição". Por exemplo, suponha que uma caixa de peças fabricadas contenha algumas peças defeituosas. Você escolhe uma peça da caixa, descobre que é defeituosa e a remove da caixa. Se você escolher outra peça da caixa, a probabilidade de ela ser defeituosa é um pouco menor do que para a primeira peça, porque você retirou uma peça defeituosa. Se você tivesse substituído a peça defeituosa, as probabilidades teriam permanecido as mesmas e o processo teria satisfeito as condições de uma distribuição binomial.

Condições

As condições subjacentes à distribuição hipergeométrica são:

- O número total de itens ou elementos (o tamanho da população) é um número fixo, uma população finita. O tamanho da população deve ser menor ou igual a 1.750.
- O tamanho da amostra (o número de tentativas) representa uma parte da população.
- A probabilidade de êxito inicial conhecida da população muda depois de cada tentativa.

Os construtos matemáticos da distribuição hipergeométrica são os seguintes:

$$P(x) = \frac{\dfrac{(N_x)!}{x!(N_x - x)!} \dfrac{(N - N_x)!}{(n-x)!(N - N_x - n + x)!}}{\dfrac{N!}{n!(N-n)!}} \quad \text{para } x = Max(n - (N - N_x), 0), \; ..., Min(n, N_x)$$

média = $\frac{N_x n}{N}$

desvio padrão = $\sqrt{\frac{(N - N_x)N_x n(N-n)}{N^2(N-1)}}$

obliquidade = $\sqrt{\frac{N-1}{(N-N_x)N_x n(N-n)}}$

curtose excessiva = função complexa

O número de itens na população ou o tamanho da população (*N*), as tentativas amostradas ou o tamanho da amostra (*n*) e o número de itens da população que obtiveram êxito ou os êxitos da população (N_x) são os parâmetros de distribuição. O número de tentativas bem-sucedidas é indicado por*x*.

Requisitos de entrada:
Tamanho da população ≥ 2 e inteiro
Tamanho da amostra > 0 e inteiro
Êxitos da população > 0 e inteiro
Tamanho da população > êxitos da população
Tamanho da amostra < êxitos da população
Tamanho da população < 1750

Distribuição de Pascal

A distribuição de Pascal é útil para modelar a distribuição do número total de tentativas necessárias para obter o número de ocorrências bem-sucedidas necessárias. Por exemplo, para fechar um total de 10 oportunidades de vendas, quantas chamadas de vendas no total seriam necessárias, dada alguma probabilidade de êxito em cada chamada? O eixo x mostra o número total de chamadas necessárias, que inclui chamadas bem-sucedidas e malsucedidas. O número de tentativas não é fixo, as tentativas continuam até o R-ésimo êxito e a probabilidade de sucesso é a mesma em cada tentativa. A distribuição de Pascal está relacionada à distribuição binomial negativa. A distribuição binomial negativa calcula o número de eventos necessários além do número de êxitos necessários dada alguma probabilidade (ou seja, o total de falhas), enquanto a distribuição de Pascal calcula o número total de eventos necessários (ou seja, a soma de êxitos e falhas) para alcançar os êxitos necessários dada alguma probabilidade. Êxitos necessários e probabilidade são os parâmetros de distribuição.

Os construtos matemáticos da distribuição Pascal são os seguintes:

$$f(x)=\begin{cases}\dfrac{(x-1)!}{(x-s)!(s-1)!}p^S(1-p)^{X-S} \text{ for all } x \geq s \\ 0 \text{ otherwise}\end{cases}$$

$$F(x)=\begin{cases}\sum_{x=1}^{k}\dfrac{(x-1)!}{(x-s)!(s-1)!}p^S(1-p)^{X-S} \text{ for all } x \ \geq \ s \\ 0 \text{ otherwise}\end{cases}$$

$$\text{média} = \frac{s}{p}$$

$$\text{desvio padrão} \ = \sqrt{s(1-p)p^2}$$

$$\text{obliquidade} = \frac{2-p}{\sqrt{r(1-p)}}$$

$$\text{curtose excessiva} = \frac{p^2-6p+6}{r(1-p)}$$

Requisitos de entrada:

Êxitos necessários deve ser maior que 0 e ser um inteiro

0 ≤ Probabilidade ≤ 1

Distribuição Poisson

A distribuição Poisson descreve o número de vezes que um evento ocorre em um dado intervalo, como o número de telefonemas por minuto ou o número de erros por página em um documento.

Condições

As três condições subjacentes à distribuição Poisson são:

- O número de ocorrências possíveis em qualquer intervalo é ilimitado.
- As ocorrências são independentes. O número de ocorrências em um intervalo não afeta o número de ocorrências em outros intervalos.
- O número médio de ocorrências deve permanecer inalterado de um intervalo para outro.

Os construtos matemáticos da distribuição Poisson são os seguintes:

$$P(x)=\frac{e^{-\lambda}\lambda^x}{x!} \quad \text{para } x \text{ e } \lambda > 0$$

$$\text{média} \ = \ \lambda$$

$$\text{desvio padrão} = \sqrt{\lambda}$$

$$\text{obliquidade} = \frac{1}{\sqrt{\lambda}}$$

curtose excessiva = $\frac{1}{\lambda}$

A taxa ou o lambda (λ) é o único parâmetro de distribuição.

Requisitos de entrada:

Taxa > 0 e ≤ 1000 (ou seja, 0,0001 ≤ taxa ≤ 1000)

Distribuição uniforme discreta

A distribuição uniforme discreta também é conhecida como a distribuição de *resultados equiprováveis*, na qual a distribuição tem um conjunto de *N* elementos e cada elemento tem a mesma probabilidade. Essa distribuição está relacionada à distribuição uniforme, mas seus elementos são discretos e não contínuos. Os construtos matemáticos da distribuição binomial são os seguintes:

$$P(x) = \frac{1}{N}$$

média = $\frac{N+1}{2}$ valor classificado

desvio padrão = $\sqrt{\frac{(N-1)(N+1)}{12}}$ valor classificado

obliquidade = 0 (ou seja, a distribuição é perfeitamente simétrica)

curtose excessiva = $\frac{-6(N^2+1)}{5(N-1)(N+1)}$ valor classificado

Requisitos de entrada:

Mínimo < máximo, ambos devem ser inteiros (inteiros negativos e zero são permitidos)

Distribuições contínuas

Distribuição de arco seno

A distribuição de arco seno tem o formato de U e é um caso especial da distribuição beta no qual o formato e a escala são iguais a 0,5. Valores próximos ao mínimo e ao máximo têm alta probabilidade de ocorrência enquanto valores entre esses dois extremos têm probabilidades muito baixas de ocorrência. Mínimo e máximo são os parâmetros de distribuição.

Os construtos matemáticos da distribuição arco seno são os seguintes:

$$f(x) = \begin{cases} \dfrac{1}{\pi\sqrt{x(1-x)}} & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{2}{\pi}\sin^{-1}(\sqrt{x}) & for\ 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

$$\text{média} = \frac{Min + Max}{2}$$

$$\text{desvio padrão} = \sqrt{\frac{(Max - Min)^2}{8}}$$

obliquidade = 0

curtose excessiva = 1.5

Requisitos de entrada:

Mínimo < Máximo

Distribuição beta

A distribuição beta é muito flexível e geralmente é usada para representar variabilidade em uma faixa fixa. Uma das aplicações mais importantes da distribuição beta é seu uso como uma distribuição conjugada do parâmetro de uma distribuição Bernoulli. Nessa aplicação, a distribuição beta é usada para representar a incerteza na probabilidade de ocorrência de um evento. Também é usada para descrever dados empíricos e prever o comportamento aleatório de porcentagens e frações, já que a faixa de resultados fica normalmente entre 0 e 1.

A importância da distribuição beta está na grande variedade de formas que ela pode assumir quando você varia dois parâmetros: alfa e beta. Se os parâmetros forem iguais, a distribuição será simétrica. Se um parâmetro for 1 e o outro for maior do que 1, a distribuição será no formato J. Se alfa for menor do que beta, diz-se que a distribuição é positivamente oblíqua (a maioria dos valores está próxima do valor mínimo). Se alfa for maior do que beta, diz-se que a distribuição é negativamente oblíqua (a maioria dos valores está próxima do valor máximo).

Os construtos matemáticos da distribuição beta são os seguintes:

$$f(x) = \frac{(x)^{(\alpha-1)}(1-x)^{(\beta-1)}}{\left[\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right]} \quad \text{parar } \alpha > 0; \beta > 0; x > 0$$

$$\text{média} = \frac{\alpha}{\alpha + \beta}$$

$$\text{desvio padrão} = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)}}$$

$$\text{obliquidade} = \frac{2(\beta-\alpha)\sqrt{1+\alpha+\beta}}{(2+\alpha+\beta)\sqrt{\alpha\beta}}$$

$$\text{curtose excessiva} = \frac{3(\alpha+\beta+1)[\alpha\beta(\alpha+\beta-6)+2(\alpha+\beta)^2]}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)} - 3$$

Alfa (α) e beta (β) são os dois parâmetros de forma da distribuição e Γ é a função gama.

Condições

As duas condições subjacentes à distribuição beta são:

- A variável incerta é um valor aleatório entre 0 e um valor positivo.
- O formato da distribuição pode ser especificado usando dois valores positivos.

Requisitos de entrada:

Alfa e beta > 0, podendo ser qualquer valor positivo

Distribuição beta multiplicativa deslocada

A distribuição beta é muito flexível e geralmente é usada para representar variabilidade em uma faixa fixa. Ela é usada para descrever dados empíricos e prever o comportamento aleatório de porcentagens e frações, já que a faixa dos resultados em geral está entre 0 e 1. A vantagem da distribuição beta é a grande variedade de formas que pode assumir quando os dois parâmetros são variados, alfa e beta. A distribuição beta multiplicativa deslocada é obtida pela multiplicação da distribuição beta por um fator e pelo deslocamento dos resultados de acordo com um parâmetro de local para permitir que a faixa de resultados seja expandida além de seus limites naturais de 0 e 1 com um ponto de partida diferente de 0. Alfa, beta, local e fator são os parâmetros de entrada.

Requisitos de entrada:

Alfa > 0

Beta > 0

Local pode ser qualquer número positivo ou negativo incluindo zero

Fator > 0

Distribuição Cauchy, Lorentz ou Breit-Wigner

A distribuição Cauchy, também chamada de distribuição de Lorentz ou de Breit-Wigner, é uma distribuição contínua que descreve comportamento ressonante. Ela também descreve a distribuição de distâncias horizontais nas quais um segmento de linha inclinado a um ângulo aleatório corta o eixo x.

Os construtos matemáticos da distribuição Cauchy ou Lorentz são os seguintes:

$$f(x) = \frac{1}{\pi}\frac{\gamma/2}{(x-m)^2+\gamma^2/4}$$

A distribuição Cauchy é um caso especial em que não há momentos teóricos (média, desvio padrão, obliquidade e curtose) já que todos são indefinidos.

Localização da moda (α) e escala (β) são os únicos parâmetros nesta distribuição. O parâmetro de localização especifica o pico ou a moda da distribuição, enquanto o parâmetro de escala especifica a metade da largura na metade do máximo da distribuição. Além disso, a média e a variância de uma distribuição de Cauchy ou Lorentz são indefinidas.

Além disso, a distribuição de Cauchy é a distribuição T de Student com apenas um grau de liberdade. Essa distribuição também é construída tomando-se o quociente de duas distribuições normais padrão (distribuições normais com uma média zero e uma variação de um) que são independentes uma da outra.

Requisitos de entrada:

O local alfa pode ser qualquer valor

Beta da escala > 0, podendo ser qualquer valor positivo

Distribuição de cosseno

A distribuição de cosseno parece uma distribuição logística na qual o valor da mediana entre o mínimo e o máximo possui a moda ou o pico mais alto, tendo a probabilidade máxima de ocorrência, enquanto as caudas extremas próximas aos valores mínimo e máximo têm probabilidades menores. Mínimo e máximo são os parâmetros de distribuição.

Os construtos matemáticos da distribuição cosseno são os seguintes:

$$f(x) = \begin{cases} \frac{1}{2b}\cos\left[\frac{x-a}{b}\right] & \textit{for } \min \le x \le \max \\ 0 & \textit{otherwise} \end{cases}$$

$$\textit{where } a = \frac{\min + \max}{2} \textit{ and } b = \frac{\max - \min}{\pi}$$

$$F(x) = \begin{cases} \frac{1}{2}\left[1 + \sin\left(\frac{x-a}{b}\right)\right] & \textit{for } \min \le x \le \max \\ \mathit{1} & \textit{for } x > \mathit{max} \end{cases}$$

$$\text{média} = \frac{Min + Max}{2}$$

$$\text{desvio padrão} = \sqrt{\frac{(Max - Min)^2(\pi^2 - 8)}{4\pi^2}}$$

obliquidade = 0

$$\text{curtose excessiva} = \frac{6(90-\pi^4)}{5(\pi^2-6)^2}$$

Requisitos de entrada:

Mínimo < Máximo

Distribuição de Erlang

A distribuição de Erlang é igual à distribuição gama, com o requisito de que o parâmetro alfa ou o parâmetro de formato deve ser um inteiro positivo. Um exemplo de aplicação da distribuição de Erlang é calibrar a taxa de transição de elementos através de um sistema de compartimentos. Tais sistemas são amplamente usados em biologia e ecologia (por exemplo, em epidemiologia, um indivíduo pode progredir a uma taxa exponencial de um estado saudável passando a portador de uma doença e continuar exponencialmente de portador a infectado). Alfa (também conhecido como formato) e beta (também conhecido como escala) são os parâmetros de distribuição.

Os construtos matemáticos da distribuição Erlang são os seguintes:

$$f(x) = \begin{cases} \dfrac{\left(\dfrac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta}}{\beta(\alpha-1)} & \mathit{for}\ x \geq 0 \\ 0 & \mathit{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-x/\beta} \displaystyle\sum_{i=0}^{\alpha-1} \frac{(x/\beta)^i}{i!} & \mathit{for}\ x \geq 0 \\ 0 & \mathit{otherwise} \end{cases}$$

$\text{média} = \alpha\beta$

$\text{desvio padrão} = \sqrt{\alpha\beta^2}$

$\text{obliquidade} = \dfrac{2}{\sqrt{\alpha}}$

$\text{curtose excessiva} = \dfrac{6}{\alpha} - 3$

Requisitos de entrada:

Alfa (formato) deve ser maior que 0 e ser um inteiro

Beta (escala) > 0

Distribuição exponencial

A distribuição exponencial é amplamente usada para descrever eventos que se repetem em momentos aleatórios, como o intervalo de tempo entre eventos como falhas em equipamentos eletrônicos ou entre entradas em um estande de serviços. Ela está relacionada à distribuição Poisson, que descreve o número de ocorrências de um evento em um determinado intervalo de tempo. Uma característica importante da distribuição exponencial é a sua propriedade de "não memorização", o que significa que a vida útil futura de determinado objeto tem a mesma distribuição, independentemente do tempo de sua existência. Em outras palavras, o tempo não tem efeito algum em resultados futuros.

Os construtos matemáticos da distribuição exponencial são os seguintes:

$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0; \lambda > 0$

média = $\frac{1}{\lambda}$

desvio padrão = $\frac{1}{\lambda}$

obliquidade = 2 (este valor se aplica a todas as entradas de taxa de sucesso λ)

curtose excessiva = 6 (este valor se aplica a todas as entradas de taxa de sucesso λ)

A taxa de êxito (λ) é o único parâmetro de distribuição. O número de tentativas bem-sucedidas é indicado por *x*.

Condições

A condição subjacente à distribuição exponencial é:

- A distribuição exponencial descreve a quantidade de tempo entre as ocorrências.

Requisitos de entrada:

Taxa > 0

Distribuição exponencial deslocada

A distribuição exponencial é amplamente usada para descrever eventos que se repetem em momentos aleatórios, como o intervalo de tempo entre eventos como falhas em equipamentos eletrônicos ou entre entradas em um estande de serviços. Ela está relacionada à distribuição Poisson, que descreve o número de ocorrências de um evento em um determinado intervalo de tempo. Uma característica importante da distribuição exponencial é a sua propriedade de não memorização, o que significa que a vida útil futura de determinado objeto tem a mesma distribuição, independentemente do tempo de sua existência. Em outras palavras, o tempo não tem efeito algum em resultados futuros. A taxa de êxito (λ) é o único parâmetro de distribuição.

Requisitos de entrada:

Lambda da taxa > 0

Local pode ser qualquer número positivo ou negativo incluindo zero

Distribuição de valor extremo ou distribuição Gumbel

A distribuição de valor extremo (tipo 1) normalmente é usada para descrever o maior valor de uma resposta em um período de tempo, por exemplo, em fluxos de inundações, precipitações e terremotos. Outras aplicações incluem a força necessária para quebrar um material, projetos de construções e cargas e tolerâncias de aeronaves. A distribuição de valor extremo também é conhecida como distribuição Gumbel.

Os construtos matemáticos da distribuição de valor extremo são os seguintes:

$f(x) = \frac{1}{\beta} z e^{-Z}$ onde $z = e^{\frac{x-\alpha}{\beta}}$ para $\beta > 0$; e qualquer valor de x e α

média = $\alpha + 0.577215\beta$

desvio padrão = $\sqrt{\frac{1}{6}\pi^2\beta^2}$

obliquidade = $\frac{12\sqrt{6}(1.2020569)}{\pi^3} = 1.13955$ (isso se aplica a todos os valores de moda e escala)

curtose excessiva = 5,4 (isso se aplica a todos os valores de moda e escala)

Moda (α) e escala (β) são os parâmetros de distribuição.

Cálculo de parâmetros

Há dois parâmetros padrão para a distribuição de valor extremo: moda e escala. O parâmetro moda é o valor de maior probabilidade para a variável (o ponto mais alto na distribuição da probabilidade). Depois de selecionar o parâmetro de moda, você pode estimar o parâmetro de escala. O parâmetro de escala é um número maior que 0. Quanto maior o parâmetro de escala, maior a variância.

Requisitos de entrada:

Alfa da moda pode ser qualquer valor

Beta da escala > 0

Distribuição F ou distribuição de Fisher-Snedecor

A distribuição F, também conhecida como a distribuição de Fisher-Snedecor, é outra distribuição contínua usada mais frequentemente para o teste de hipóteses. Especificamente, ela é usada para testar a diferença estatística entre duas variâncias na análise de testes de variância e testes de índice de verossimilhança. A distribuição F com o grau de liberdade do numerador *n* e grau de liberdade do denominador *m* está relacionada à distribuição qui-quadrada pois:

$$\frac{\chi_n^2 / n}{\chi_m^2 / m} \overset{d}{\sim} F_{n,m}$$

média = $\dfrac{m}{m-2}$

desvio padrão = $\dfrac{2m^2(m+n-2)}{n(m-2)^2(m-4)}$ para todos os $m > 4$

obliquidade = $\dfrac{2(m+2n-2)}{m-6}\sqrt{\dfrac{2(m-4)}{n(m+n-2)}}$

curtose excessiva = $\dfrac{12(-16+20m-8m^2+m^3+44n-32mn+5m^2n-22n^2+5mn^2}{n(m-6)(m-8)(n+m-2)}$

O grau de liberdade do numerador *n* e o grau de liberdade do denominador *m* são os únicos parâmetros de distribuição.
Requisitos de entrada:
Graus de liberdade do numerador e do denominador > 0 e inteiros

Distribuição gama (distribuição Erlang)

A distribuição gama se aplica a uma grande variedade de quantidades físicas e está relacionada a outras distribuições: lognormal, exponencial, Pascal, Erlang, Poisson e qui-quadrada. Ela é usada em processos meteorológicos para representar a concentração de poluentes e as quantidades de precipitação. A distribuição gama também é usada para medir o tempo entre as ocorrências de eventos, quando o processo do evento não é completamente aleatório. Outras aplicações da distribuição gama incluem controle de estoques, teoria econômica e a teoria de riscos de seguro.
Condições
A distribuição gama é usada com mais frequência como a distribuição da quantidade de tempo até a *r*-ésima ocorrência de um evento em um processo Poisson. Quando usada dessa maneira, as três condições subjacentes à distribuição gama são:

- O número de ocorrências possíveis em qualquer unidade de medida não está limitado a um número fixo.
- As ocorrências são independentes. O número de ocorrências em uma unidade de medida não afeta o número de ocorrências em outras unidades.

- O número médio de ocorrências deve permanecer inalterado de uma unidade para outra.

Os construtos matemáticos da distribuição gama são os seguintes:

$$f(x) = \frac{\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta} \quad \text{com qualquer valor de } \alpha > 0 \text{ e } \beta > 0$$

média = α

desvio padrão = $\sqrt{\alpha\beta^2}$

obliquidade = $\frac{2}{\sqrt{\alpha}}$

curtose excessiva = $\frac{6}{\alpha}$

O alfa do parâmetro de forma (α) e o beta do parâmetro de escala (β) são os parâmetros de distribuição e Γ é a função gama.

Quando o parâmetro alfa é um inteiro positivo, a distribuição gama é chamada de distribuição de Erlang, usada para prever tempos de espera em sistemas de enfileiramento, nos quais a distribuição de Erlang é a soma de variáveis aleatórias independentes e distribuídas de forma idêntica, cada uma possuindo uma distribuição exponencial de não memorização. Definindo n como o número dessas variáveis aleatórias, o construto matemático da distribuição de Erlang é:

$$f(x) = \frac{x^{n-1}e^{-x}}{(n-1)!}$$, para todos os $x > 0$ e todos os inteiros positivos de n

Requisitos de entrada:

Beta da escala > 0, podendo ser qualquer valor positivo

Alfa da forma ≥ 0,05 e qualquer valor positivo

A localização pode ser qualquer valor

Distribuição de Laplace

A distribuição de Laplace também é chamada de distribuição exponencial dupla, pois pode ser construída com duas distribuições exponenciais (com um parâmetro de local adicional) combinadas, criando um pico incomum no meio. A função densidade de probabilidade da distribuição de Laplace lembra a distribuição normal. No entanto, enquanto a distribuição normal é expressa em termos da diferença quadrática da média, a densidade de Laplace é expressa em termos da diferença absoluta da média, tornando as caudas da distribuição de Laplace mais achatadas que as da distribuição normal. Quando o parâmetro de local é definido como zero, a variável aleatória da distribuição de Laplace é distribuída exponencialmente com um inverso do parâmetro de escala. Alfa (também conhecido como local) e beta (também conhecido como escala) são os parâmetros de distribuição.

Os construtos matemáticos da distribuição Laplace são os seguintes:

$$f(x)=\frac{1}{2\beta}\exp\left(-\frac{|x-\alpha|}{\beta}\right)$$

$$F(x)=\begin{cases}\frac{1}{2}\exp\left[\frac{x-\alpha}{\beta}\right] \text{ when } x<\alpha \\ 1-\frac{1}{2}\exp\left[-\frac{x-\alpha}{\beta}\right] \text{ when } x\geq\alpha\end{cases}$$

$$\text{média}=\alpha$$

$$\text{desvio padrão } = 1.4142\beta$$

obliquidade = 0

curtose excessiva = 3

Requisitos de entrada:

Alfa (local) pode assumir valores positivos ou negativos, incluindo zero

Beta (escala) > 0

Distribuição logarítmica dupla

A distribuição logarítmica dupla é semelhante à distribuição Cauchy, na qual a tendência central tem um pico e tem a densidade de probabilidade de valor máximo, mas diminui mais rapidamente quanto mais longe está do centro, criando uma distribuição simétrica com um pico extremo entre os valores mínimo e máximo. Mínimo e máximo são os parâmetros de distribuição.

Os construtos matemáticos da distribuição logarítmica dupla são os seguintes:

$$f(x)=\begin{cases}\frac{-1}{2b}\ln\left(\frac{|x-a|}{b}\right) & \textit{for} \min\leq x\leq\max \\ 0 & \textit{otherwise}\end{cases}$$

$$\textit{where } a=\frac{\min+\max}{2} \textit{ and } b=\frac{\max-\min}{2}$$

$$F(x)=\begin{cases}\frac{1}{2}-\left(\frac{|x-a|}{2b}\right)\left[1-\ln\left(\frac{|x-a|}{b}\right)\right] & \textit{for} \min\leq x\leq a \\ \frac{1}{2}+\left(\frac{|x-a|}{2b}\right)\left[1-\ln\left(\frac{|x-a|}{b}\right)\right] & \textit{for } a\leq x\leq\max\end{cases}$$

$$\text{média} = \frac{Min + Max}{2}$$

$$\text{desvio padrão} = \sqrt{\frac{(Max - Min)^2}{36}}$$

obliquidade = 0

Requisitos de entrada:

Mínimo < Máximo

Distribuição logística

A distribuição logística normalmente é usada para descrever crescimento, ou seja, o tamanho de uma população expresso como uma função de uma variável de tempo. Ela também pode ser usada para descrever reações químicas e a taxa de crescimento de uma população ou um indivíduo.

Os construtos matemáticos da distribuição logística são os seguintes:

$$f(x) = \frac{e^{\frac{\alpha - x}{\beta}}}{\beta\left[1 + e^{\frac{\alpha - x}{\beta}}\right]^2} \quad \text{para qualquer valor de } \alpha \text{ e } \mu$$

$$\text{média } = \alpha$$

$$\text{desvio padrão } = \sqrt{\frac{1}{3}\pi^2\beta^2}$$

obliquidade = 0 (este valor se aplica a todas as entradas de média e escala)

curtose excessiva = 1,2 (este valor se aplica a todas as entradas de média e escala)

Média (α) e escala (β) são os parâmetros de distribuição.

Cálculo de parâmetros

Há dois parâmetros padrão para a distribuição logística: média e escala. O parâmetro de média é o valor médio e, para esta distribuição, é igual à moda, porque é uma distribuição simétrica. Depois de selecionar o parâmetro de média, você pode estimar o parâmetro de escala. O parâmetro de escala é um número maior que 0. Quanto maior o parâmetro de escala, maior a variância.

Requisitos de entrada:

Beta da escala > 0, podendo ser qualquer valor positivo

Alfa da média pode ser qualquer valor

Distribuição lognormal

A distribuição lognormal é amplamente usada em situações nas quais os valores são positivamente inclinados, por exemplo, em análises financeiras para avaliação de títulos ou avaliações de propriedades, nas quais os valores não podem ficar abaixo de zero.

Os preços de ações normalmente são positivamente inclinados em vez de distribuídos normalmente (simetricamente). Os preços de ações apresentam essa tendência pois não podem cair abaixo do limite mais baixo, que é zero, mas podem aumentar até atingir qualquer preço sem limite. Do mesmo modo, os preços de imóveis demonstram obliquidade positiva pois não podem tornar-se negativos.

Condições

As três condições subjacentes à distribuição lognormal são:

- As variáveis incertas podem crescer sem limites mas não podem ficar abaixo de zero.
- A variável incerta tem uma obliquidade positiva, com a maioria dos valores perto do limite inferior.
- O logaritmo natural da variável incerta produz uma distribuição normal.

Geralmente, se o coeficiente de variabilidade é maior que 30 por cento, use uma distribuição lognormal. Caso contrário, use a distribuição normal.

Os construtos matemáticos da distribuição lognormal são os seguintes:

$$f(x) = \frac{1}{x\sqrt{2\pi}\ln(\sigma)} e^{-\frac{[\ln(x)-\ln(\mu)]^2}{2[\ln(\sigma)]^2}} \quad \text{para } x > 0; \mu > 0 \text{ e } \sigma > 0$$

$$\text{média } = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$\text{desvio padrão} = \sqrt{\exp\left(\sigma^2 + 2\mu\right)\left[\exp\left(\sigma^2\right) - 1\right]}$$

$$\text{obliquidade} = \left[\sqrt{\exp\left(\sigma^2\right) - 1}\right](2 + \exp(\sigma^2))$$

$$\text{curtose excessiva} = \exp\left(4\sigma^2\right) + 2\exp\left(3\sigma^2\right) + 3\exp\left(2\sigma^2\right) - 6$$

Média (μ) e desvio padrão (σ) são os parâmetros de distribuição.

Requisitos de entrada:

Média e desvio padrão > 0, podendo ser qualquer valor positivo

Conjuntos de parâmetros lognormais

Por padrão, a distribuição lognormal usa a média aritmética e o desvio padrão. Para aplicações nas quais dados históricos estão disponíveis, é mais apropriado usar a média logarítmica e o desvio padrão ou a média geométrica e o desvio padrão.

Distribuição lognormal deslocada

A distribuição lognormal é amplamente usada em situações nas quais os valores são positivamente inclinados, por exemplo, em análises financeiras para avaliação de títulos ou avaliações de propriedades, nas quais os valores não podem ficar abaixo de zero. Os preços de ações normalmente são positivamente inclinados em vez de distribuídos normalmente (simetricamente). Os preços de ações apresentam essa tendência pois não podem cair abaixo do limite mais baixo, que é zero, mas podem aumentar até atingir qualquer preço sem limite. Em contrapartida, a distribuição lognormal deslocada é igual à distribuição lognormal, mas é deslocada de maneira que o valor resultante possa assumir valores negativos. Média, desvio padrão e deslocamento são os parâmetros de distribuição.

Requisitos de entrada:

Média > 0

Desvio padrão > 0

Deslocamento pode ser um valor positivo ou negativo incluindo zero

Distribuição normal

A distribuição normal é a distribuição mais importante em teoria de probabilidade porque descreve muitos fenômenos naturais, como o QI ou a altura das pessoas. Pessoas responsáveis por tomar decisões podem usar a distribuição normal para descrever variáveis incertas, como a taxa de inflação ou o preço futuro da gasolina.

Condições

As três condições subjacentes à distribuição normal são:

- Algum valor da variável incerta é o mais provável (a média da distribuição).
- A variável incerta pode tanto estar acima da média como abaixo dela (simétrica em relação à média).
- A variável incerta provavelmente estará mais próxima da média do que mais afastada dela.

Os construtos matemáticos da distribuição normal são os seguintes:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{para todos os valores de } x \text{ e } \mu\text{; enquanto } \sigma > 0$$

$$\text{média} = \mu$$

$$\text{desvio padrão} = \sigma$$

obliquidade = 0 (isso se aplica a todas as entradas de média e desvio padrão)

curtose excessiva = 0 (isso se aplica a todas as entradas de média e desvio padrão)

Média (μ) e desvio padrão (σ) são os parâmetros de distribuição.

Requisitos de entrada:

Desvio padrão > 0, podendo ser qualquer valor positivo

A média pode assumir qualquer valor

Distribuição parabólica

A distribuição parabólica é um caso especial da distribuição de beta quando Forma = Escala = 2. Os valores fecham ao mínimo e máximo tem probabilidades baixas de ocorrência ao passo que valores entre estes dois extremos têm probabilidades mais altas ou ocorrência. O mínimo e máximo são os parâmetros de distributional.

Os construtos matemáticos da distribuição parabólica são os seguintes:

$$f(x) = \frac{(x)^{(\alpha-1)}(1-x)^{(\beta-1)}}{\left[\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right]} \quad \text{for } \alpha > 0; \beta > 0; x > 0$$

Where the functional form above is for a Beta distribution, and for a Parabolic function, we set Alpha = Beta = 2 and a shift of location in Minimum, with a multiplicative factor of (Maximum – Minimum).

$$\text{média} = \frac{Min + Max}{2}$$

$$\text{desvio padrão} = \sqrt{\frac{(Max - Min)^2}{20}}$$

obliquidade = 0

curtose excessiva = –0.8571

Requisitos de entrada:

Mínimo < Máximo

Distribuição Pareto

A distribuição Pareto é amplamente usada na investigação de distribuições associadas a fenômenos empíricos, como o tamanho da população de cidades, a ocorrência de recursos naturais, o tamanho de empresas, renda pessoal, flutuações no preço de ações e agrupamento de erros em circuitos de comunicação.

Os construtos matemáticos da distribuição Pareto são os seguintes:

$$f(x) = \frac{\beta L^{\beta}}{x^{(1+\beta)}} \quad \text{para } x > L$$

$$\text{média} = \frac{\beta L}{\beta - 1}$$

$$\text{desvio padrão} = \sqrt{\frac{\beta L^2}{(\beta-1)^2(\beta-2)}}$$

obliquidade = $\sqrt{\frac{\beta-2}{\beta}}\left[\frac{2(\beta+1)}{\beta-3}\right]$

curtose excessiva = $\frac{6(\beta^3+\beta^2-6\beta-2)}{\beta(\beta-3)(\beta-4)}$

Forma (α) e local (β) são os parâmetros de distribuição.

Cálculo de parâmetros

Há dois parâmetros padrão para a distribuição Pareto: localização e formato. O parâmetro de localização é o limite inferior da variável. Depois de selecionar o parâmetro de localização, você pode estimar o parâmetro de formato. O parâmetro de formato é um número maior que 0 e geralmente maior que 1. Quanto maior o parâmetro de formato, menor a variação e mais grossa a cauda direita da distribuição.

Requisitos de entrada:

Localização > 0, podendo ser qualquer valor positivo

Formato ≥ 0,05

Distribuição de Pearson V

A distribuição de Pearson V está relacionada à distribuição gama invertida, na qual ela é a recíproca da variável distribuída de acordo com a distribuição gama. A distribuição de Pearson V também é usada para modelar atrasos de tempo nos quais haja quase certeza de algum atraso mínimo e o atraso máximo não tem limites, por exemplo, atraso na chegada de serviços de emergência e o tempo para consertar uma máquina. Alfa (também conhecido como formato) e beta (também conhecido como escala) são os parâmetros de distribuição.

Os construtos matemáticos da distribuição Pearson V são os seguintes:

$$f(x)=\frac{x^{-(\alpha+1)}e^{-\beta/x}}{\beta^{-\alpha}\Gamma(\alpha)}$$

$$F(x)=\frac{\Gamma(\alpha,\beta/x)}{\Gamma(\alpha)}$$

$$\text{média}=\frac{\beta}{\alpha-1}$$

$$\text{desvio padrão}=\sqrt{\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}}$$

$$\text{obliquidade}=\frac{4\sqrt{\alpha-2}}{\alpha-3}$$

$$\text{curtose excessiva}=\frac{30\alpha-66}{(\alpha-3)(\alpha-4)}-3$$

Requisitos de entrada:

Alfa (formato) > 0

Beta (escala) > 0

Distribuição de Pearson VI

A distribuição de Pearson VI está relacionada à distribuição gama, na qual ela é a função racional de duas variáveis distribuídas de acordo com duas distribuições gama. Alfa 1 (também conhecido como formato 1), alfa 2 (também conhecido como formato 2) e beta (também conhecido como escala) são os parâmetros de distribuição.

Os construtos matemáticos da distribuição Pearson VI são os seguintes:

$$f(x) = \frac{(x/\beta)^{\alpha_1 - 1}}{\beta \, \mathrm{B}(\alpha_1, \alpha_2)[1 + (x/\beta)]^{\alpha_1 + \alpha_2}}$$

$$F(x) = F_B\left(\frac{x}{x+\beta}\right)$$

$$\text{média} = \frac{\beta\alpha_1}{\alpha_2 - 1}$$

$$\text{desvio padrão} = \sqrt{\frac{\beta^2 \alpha_1(\alpha_1 + \alpha_2 - 1)}{(\alpha_2 - 1)^2(\alpha_2 - 2)}}$$

$$\text{obliquidade} = 2\sqrt{\frac{\alpha_2 - 2}{\alpha_1(\alpha_1 + \alpha_2 - 1)}}\left[\frac{2\alpha_1 + \alpha_2 - 1}{\alpha_2 - 3}\right]$$

$$\text{curtose excessiva} = \frac{3(\alpha_2 - 2)}{(\alpha_2 - 3)(\alpha_2 - 4)}\left[\frac{2(\alpha_2 - 1)^2}{\alpha_1(\alpha_1 + \alpha_2 - 1)} + (\alpha_2 + 5)\right] - 3$$

Requisitos de entrada:

Alfa 1 (formato 1) > 0

Alfa 2 (formato 2) > 0

Beta (escala) > 0

Distribuição PERT

A distribuição PERT é amplamente usada em gerenciamento de projetos e programas para definir os cenários de pior caso, caso nominal e melhor caso para o prazo de conclusão do projeto. Ela está relacionada às distribuições beta e triangular. A distribuição PERT pode ser usada para identificar riscos em projetos e modelos de custo com base na probabilidade de cumprir metas e objetivos em qualquer número de componentes de projeto usando os valores mínimo, mais provável e máximo, mas foi projetada para gerar uma distribuição que se assemelhe mais a distribuições de probabilidade realistas. A distribuição PERT pode fornecer um ajuste próximo à distribuição normal ou lognormal. Assim como a distribuição triangular, a distribuição PERT dá mais ênfase ao valor "mais provável" do que às estimativas mínima e máxima. No entanto, diferentemente da distribuição triangular, a distribuição PERT constrói uma curva suavizada, que dá progressivamente mais ênfase a valores próximos ao valor mais provável, em detrimento de valores mais próximos das bordas. Na prática, isso significa que "confiamos" na estimativa para o valor mais provável e acreditamos que, mesmo se ele não for exatamente preciso (e as estimativas raramente o são), temos uma expectativa de que o valor resultante será próximo daquela estimativa. Supondo que muitos fenômenos reais são distribuídos normalmente, a vantagem da distribuição PERT é que ela produz uma curva com forma semelhante à da curva normal, sem conhecer os parâmetros precisos da curva normal relacionada. Mínimo, mais provável e máximo são os parâmetros de distribuição.

Os construtos matemáticos da distribuição PERT são os seguintes:

$$f(x) = \frac{(x-\min)^{A1-1}(\max-x)^{A2-1}}{B(A1,A2)(\max-\min)^{A1+A2-1}}$$

$$\textit{where } A1 = 6\left[\frac{\frac{\min+4(\text{likely})+\max}{6}-\min}{\max-\min}\right] \textit{ and } A2 = 6\left[\frac{\max-\frac{\min+4(\text{likely})+\max}{6}}{\max-\min}\right]$$

$$\textit{and } B \textit{ is the Beta function}$$

$$\text{média} = \frac{Min+4Mode+Max}{6}$$

$$\text{desvio padrão} = \sqrt{\frac{(\mu-Min)(Max-\mu)}{7}}$$

$$\text{obliquidade} = \sqrt{\frac{7}{(\mu-Min)(Max-\mu)}}\left(\frac{Min+Max-2\mu}{4}\right)$$

Requisitos de entrada:

Mínimo ≤ Mais provável ≤ Máximo, podendo ser positivos, negativos ou zero

Distribuição potencial

A distribuição potencial está relacionada à distribuição exponencial uma vez que a probabilidade de resultados pequenos é grande, mas diminui exponencialmente à medida que o valor do resultado aumenta. Alfa (também conhecido como formato) é o único parâmetro de distribuição.

Os construtos matemáticos da distribuição potencial são os seguintes:

$$f(x) = \alpha x^{\alpha-1}$$

$$F(x) = x^{\alpha}$$

$$\text{média} = \frac{\alpha}{1+\alpha}$$

$$\text{desvio padrão} = \sqrt{\frac{\alpha}{(1+\alpha)^2(2+\alpha)}}$$

$$\text{obliquidade} = \sqrt{\frac{\alpha+2}{\alpha}}\left(\frac{2(\alpha-1)}{\alpha+3}\right)$$

Requisitos de entrada:

Alfa (formato) > 0

Distribuição potencial multiplicativa deslocada

A distribuição potencial está relacionada à distribuição exponencial uma vez que a probabilidade de resultados pequenos é grande, mas diminui exponencialmente à medida que o valor do resultado aumenta. Alfa (também conhecido como formato) é o único parâmetro de distribuição.

Requisitos de entrada:

Alfa (formato) > 0

Local pode ser qualquer número positivo ou negativo incluindo zero

Fator > 0

Distribuição qui-quadrada

A distribuição qui-quadrada é uma distribuição de probabilidade usada predominantemente em testes de hipóteses e está relacionada à distribuição gama e à distribuição normal padrão. Por exemplo, as somas de distribuições normais independentes são distribuídas como um qui-quadrado (χ^2) com k graus de liberdade.

$$Z_1^2 + Z_2^2 + \ldots + Z_k^2 \overset{d}{\sim} \chi_k^2$$

Os construtos matemáticos da distribuição qui-quadrada são os seguintes:

$$f(x) = \frac{0.5^{-k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2} \text{ para todo } x > 0$$

média = k

desvio padrão = $\sqrt{2k}$

obliquidade = $2\sqrt{\frac{2}{k}}$

curtose excessiva = $\frac{12}{k}$

Γ é a função gama. O grau de liberdade k é o único parâmetro de distribuição.
A distribuição qui-quadrada também pode ser modela usando uma distribuição gama definindo:

parâmetro de forma = $\frac{k}{2}$ e escala = $2S^2$, onde S é a escala.

Requisitos de entrada:
Graus de liberdade > 1, devendo ser um inteiro < 300

Distribuição t de Student

A distribuição t de Student é a mais amplamente usada em teste de hipóteses. Essa distribuição é usada para estimar a média de uma população normalmente distribuída quando o tamanho da amostra é pequeno, sendo usada para testar a significância estatística da diferença entre duas amostras ou intervalos de confiança para amostras pequenas.
Os construtos matemáticos da distribuição t são os seguintes:

$$f(t) = \frac{\Gamma[(r+1)/2]}{\sqrt{r\pi}\,\Gamma[r/2]}(1 + t^2/r)^{-(r+1)/2}$$

média = 0 (isso se aplica a todos os graus de liberdade r exceto se a distribuição for deslocada para outra localização central diferente de zero)

desvio padrão = $\sqrt{\frac{r}{r-2}}$

obliquidade = 0 (se aplica a todos os graus de liberdade r)

curtose excessiva = $\dfrac{6}{r-4}$ for all $r > 4$

onde $t = \dfrac{x-\bar{x}}{s}$ e Γ é a função gama.

Os graus de liberdade r são o único parâmetro de distribuição. A distribuição t está relacionada à distribuição F, como mostrado a seguir: o quadrado de um valor de t com r graus de liberdade é distribuído como F com 1 e r graus de liberdade. O formato geral da função de densidade da probabilidade da distribuição t lembra o formato de sino de uma variável normalmente distribuída com média 0 e variância 1, exceto que essa é um pouco mais baixa e mais larga e é leptocúrtica (caudas gordas nas extremidades e um pico no centro). À medida que o grau de liberdade aumenta (digamos, acima de 30), a distribuição t se aproxima da distribuição normal com média 0 e variância 1.
Requisitos de entrada:
Grau de liberdade ≥ 1, devendo ser um inteiro

Distribuição triangular

A distribuição triangular descreve uma situação na qual você sabe os valores mínimo, máximo e os mais prováveis (MP) de acontecer. Por exemplo, você poderia descrever o número de carros vendidos por semana quando as vendas passadas mostrassem os números mínimo e máximo e o número normal de carros vendidos.
Condições
As três condições subjacentes à distribuição triangular são:
- O número mínimo de itens é fixo.
- O número máximo de itens é fixo.
- O número mais provável de itens encontra-se entre os valores mínimo e máximo, formando uma distribuição de formato triangular, que mostra que os valores perto do mínimo e do máximo têm menor chance de ocorrência do que aqueles perto do valor mais provável (MP).

Os construtos matemáticos da distribuição triangular são os seguintes:

$$f(x) = \begin{cases} \dfrac{2(x-Min)}{(Max-Min)(MP-min)} & \text{para } Min < x < \text{Mais provavel} \\ \dfrac{2(Max-x)}{(Max-Min)(Max-MP)} & \text{para Mais provavel} < x < \text{Max} \end{cases}$$

média = $\dfrac{1}{3}(Min+MP+Max)$

desvio padrão = $\sqrt{\dfrac{1}{18}(Min^2+MP^2+Max^2-Min\,Max-Min\,MP-Max\,MP)}$

obliquidade = $\dfrac{\sqrt{2}(Min+Max-2MP)(2Min-Max-MP)(Min-2Max+MP)}{5(Min^2+Max^2+MP^2-MinMax-MinMP-MaxMP)^{3/2}}$

curtose excessiva = –0,6 (isso se aplica a todas as entradas de *Mín*, *Máx* e *Mais provável*)

O valor mínimo (*Mín*), o valor mais provável (*MP*) e o valor máximo (*Máx*) são os parâmetros de distribuição.
Requisitos de entrada:
Mín ≤ Mais provável ≤ Máx, podendo assumir qualquer valor
No entanto, Mín < Máx, podendo assumir qualquer valor

Distribuição uniforme

Com a distribuição uniforme, todos os valores entre o mínimo e o máximo ocorrem com a mesma probabilidade.
Condições
As três condições subjacentes à distribuição uniforme são:

- O valor mínimo é fixo.
- O valor máximo é fixo.
- Todos os valores entre o mínimo e o máximo ocorrem com a mesma probabilidade.

Os construtos matemáticos da distribuição uniforme são os seguintes:

$$f(x) = \frac{1}{Max - Min} \quad \text{para todos os valores de forma que } Min < Max$$

$$\text{média} = \frac{Min + Max}{2}$$

$$\text{desvio padrão} = \sqrt{\frac{(Max - Min)^2}{12}}$$

obliquidade = 0 (isso se aplica a todas as entradas de *Mín* e *Máx*)
curtose excessiva = –1,2 (isso se aplica a todas as entradas de *Mín* e *Máx*)

Valor máximo (*Máx*) e valor mínimo (*Mín*) são os parâmetros de distribuição.
Requisitos de entrada:
Mín < Máx, podendo assumir qualquer valor

Distribuição de Weibull (distribuição de Rayleigh)

A distribuição Weibull descreve os dados resultantes de testes de vida útil e fatiga. Normalmente ela é usada para descrever o tempo de falha em estudos de confiabilidade, como a força necessária para quebrar um material em testes de confiabilidade e controle de qualidade. As distribuições Weibull também são usadas para representar várias quantidades físicas, como velocidade do vento. A distribuição Weibull é uma família de distribuições que podem assumir as propriedades de muitas outras distribuições. Por exemplo, dependendo do parâmetro de forma que você definir, a distribuição Weibull pode ser usada para modelar as distribuições exponencial e Rayleigh, entre outras. A distribuição Weibull é muito flexível. Quando o parâmetro de forma Weibull é igual a 1,0, a distribuição Weibull é idêntica à distribuição exponencial. O parâmetro de localização Weibull permite configurar uma distribuição exponencial para iniciar em uma localização qualquer diferente de 0,0. Quando o parâmetro de forma é menor do que 1,0, a distribuição Weibull se torna uma curva com forte inclinação decrescente. Um fabricante pode achar esse efeito útil para descrever falhas de peças durante um período de processamento. Os construtos matemáticos da distribuição Weibull são os seguintes:

$$f(x) = \frac{\alpha}{\beta}\left[\frac{x}{\beta}\right]^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$$

$$\text{média } = \beta\Gamma(1+\alpha^{-1})$$

$$\text{desvio padrão } = \beta^2\left[\Gamma(1+2\alpha^{-1}) - \Gamma^2(1+\alpha^{-1})\right]$$

$$\text{obliquidade} = \frac{2\Gamma^3(1+\beta^{-1}) - 3\Gamma(1+\beta^{-1})\Gamma(1+2\beta^{-1}) + \Gamma(1+3\beta^{-1})}{\left[\Gamma(1+2\beta^{-1}) - \Gamma^2(1+\beta^{-1})\right]^{3/2}}$$

curtose excessiva =

$$\frac{-6\Gamma^4(1+\beta^{-1}) + 12\Gamma^2(1+\beta^{-1})\Gamma(1+2\beta^{-1}) - 3\Gamma^2(1+2\beta^{-1}) - 4\Gamma(1+\beta^{-1})\Gamma(1+3\beta^{-1}) + \Gamma(1+4\beta^{-1})}{\left[\Gamma(1+2\beta^{-1}) - \Gamma^2(1+\beta^{-1})\right]^2}$$

Formato (α) e escala de localização central (β) são os parâmetros de distribuição e Γ é a função gama.
Requisitos de entrada:
Alfa da forma $\geq$ 0,05
Beta da escala > 0, podendo ser qualquer valor positivo

Distribuições Weibull e Rayleigh multiplicativa deslocada

A distribuição Weibull descreve os dados resultantes de testes de vida útil e fatiga. Normalmente ela é usada para descrever o tempo de falha em estudos de confiabilidade, como a força necessária para quebrar um material em testes de confiabilidade e controle de qualidade. As distribuições Weibull também são usadas para representar várias quantidades físicas, como velocidade do vento. A distribuição Weibull é uma família de distribuições que podem assumir as propriedades de muitas outras distribuições. Por exemplo, dependendo do parâmetro de forma que você definir, a distribuição Weibull pode ser usada para modelar as distribuições exponencial e Rayleigh, entre outras. A distribuição Weibull é muito flexível. Quando o parâmetro de forma Weibull é igual a 1,0, a distribuição Weibull é idêntica à distribuição exponencial. A escala de localização central ou o parâmetro beta Weibull permite configurar uma distribuição exponencial para iniciar em uma localização qualquer diferente de 0,0. Quando o parâmetro de forma é menor do que 1,0, a distribuição Weibull se torna uma curva com forte inclinação decrescente. Um fabricante pode achar esse efeito útil para descrever falhas de peças durante um período de processamento. Forma (α) e escala (β) são os parâmetros de distribuição.

Requisitos de entrada:

Alfa da forma ≥ 0,05

Escala de localização central ou beta > 0, podendo ser qualquer valor positivo

Local pode ser qualquer número positivo ou negativo incluindo zero

Fator > 0

3. PREVISÃO

A previsão é a ação de prever o futuro, seja ela baseada em dados históricos ou apenas uma especulação acerca do futuro quando não há histórico. Quando há dados históricos, uma abordagem quantitativa ou estatística é melhor, mas se eles não existem, então uma abordagem qualitativa ou arbitrária é, potencialmente, o único recurso. A Figura 3.1 lista as metodologias mais comuns para previsões.

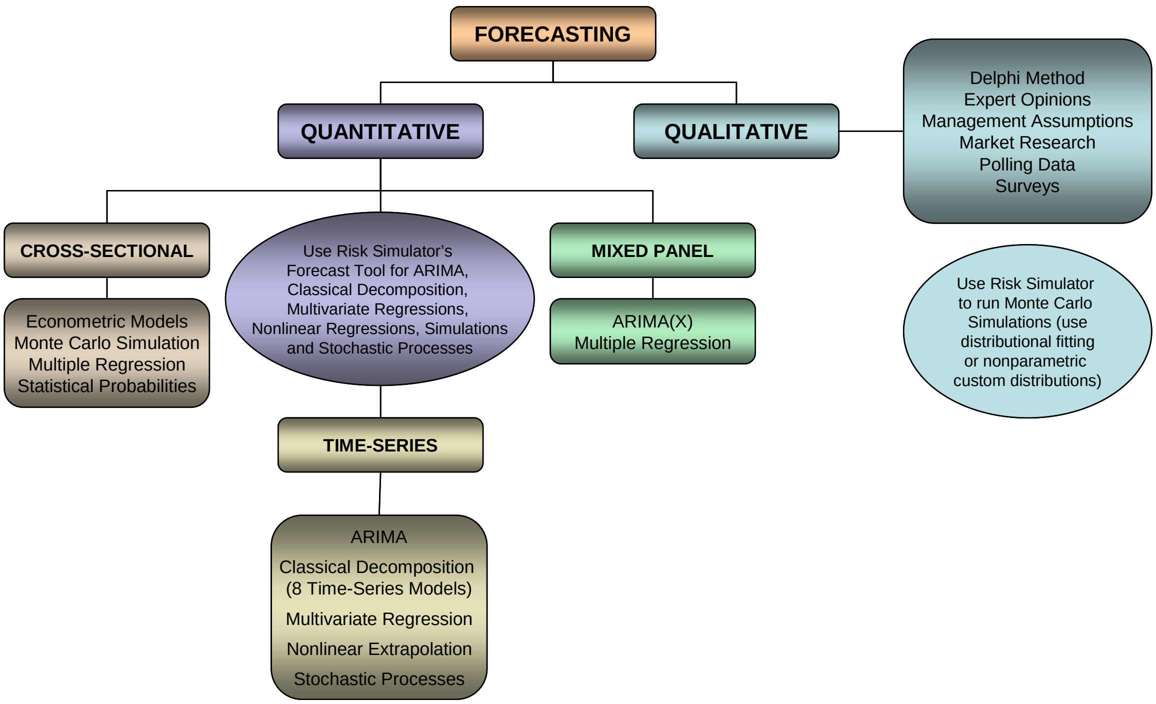


Figura 3.1 – Métodos de previsão

Diferentes tipos de técnicas de previsão

Geralmente, as previsões podem ser divididas em quantitativas e qualitativas. As previsões qualitativas são usadas quando há poucos ou nenhum dado histórico, contemporâneo ou comparativo. Há vários métodos qualitativos, como a abordagem Delphi ou de opinião de especialista (uma previsão de criação de consenso por peritos na área, especialistas de marketing ou membros da equipe interna), suposições de gerenciamento (metas de taxa de crescimento definidas pela alta administração), bem como pesquisas de mercado, dados externos ou votações e pesquisas (dados obtidos por meio de terceiros, índices setoriais e da indústria ou de pesquisas de mercado ativas). Essas estimativas podem ser estimativas de ponto único (um consenso médio) ou um conjunto de valores de previsão (uma distribuição de previsões). A última pode ser inserida no *Risk Simulator* como uma distribuição personalizada e as previsões resultantes podem ser simuladas. Ou seja, uma simulação não paramétrica usando os pontos de dados estimados como a distribuição.

Na parte quantitativa da previsão, os dados disponíveis ou dados que precisam ser previstos podem ser divididos séries temporais (valores que têm um elemento temporal, como receitas de anos diferentes, índices de inflação, taxas de juros, participação no mercado, taxas de falha etc.), dados cruzados (valores que independem do tempo, como a média nacional das notas dos universitários do segundo ano em um determinado ano, dados os níveis da pontuação de cada aluno no vestibular, seu QI e número de bebidas alcoólicas consumidas por semana) ou um painel misto (mistura entre séries temporais e dados de painel, por exemplo: previsão das vendas para os próximos 10 anos dados o orçamento das despesas de marketing e as projeções de participação no mercado, isso significa que os dados de vendas são séries temporais, mas há variáveis exógenas, como despesas de marketing e participação no mercado, que ajudam a modelar previsões).

O software *Risk Simulator* fornece ao usuário diversas metodologias de previsão:

1. ARIMA (modelo de média móvel integrada autorregressiva)
2. AutoARIMA
3. Econometria básica
4. Autoeconometria
5. Lógica difusa combinatória
6. Distribuições personalizadas
7. GARCH (modelo autorregressivo à heteroscedasticidade condicional generalizado)
8. Curvas J
9. Cadeias de Markov
10. Máxima verossimilhança
11. Rede neural
12. Regressão multivariada
13. Extrapolação não linear
14. Curvas S
15. Spline cúbico
16. Previsão estocástica
17. Análise da série temporal
18. Linhas de tendência

Os detalhes analíticos de cada método de previsão não fazem parte do escopo deste manual. Para obter mais detalhes, consulta *Modeling Risk*, 2ª edição: *Applying Monte Carlo Simulation, Real Options Analysis, Stochastic Forecasting and Portfolio Optimization*, de Dr. Johnathan Mun (Wiley Finance 2010), que também criou o software the Risk Simulator. Apesar disso, listamos a seguir algumas abordagens mais comuns. Todas as outras abordagens de previsão podem ser aplicadas sem dificuldades com o Risk Simulator.

A seguir é fornecida uma rápida revisão de cada metodologia e diversos exemplos simples para começar a usar o software. Descrições mais detalhadas e modelos de exemplos de cada técnica são encontradas ao longo deste capítulo e do próximo.

- ARIMA

Média móvel integrada autorregressiva (ARIMA, também conhecida como ARIMA Box-Jenkins) é uma técnica de modelagem econométrica avançada. ARIMA observa dados históricos de séries temporais e executa rotinas de otimização de retroajuste para explicar a autocorrelação histórica (a relação de um valor em oposição a outro no tempo), a estabilidade dos dados para corrigir as características não estacionárias dos dados. Este modelo de previsão aprende com o tempo corrigindo seus erros de previsão. Normalmente, é necessário ter conhecimento avançado em econometria para construir bons modelos de previsão usando esta abordagem.

- AutoARIMA

O modelo AutoARIMA automatiza algumas das modelagens ARIMA tradicionais testando automaticamente várias permutações de especificações do modelo e retorna o modelo mais apropriado. A execução de AutoARIMA é semelhante à execução das previsões ARIMA normais. A diferença é que as entradas P, D e Q não são obrigatórias e diferentes combinações desses valores são automaticamente executados e comparados.

- Econometria básica

A econometria se refere a um ramo de técnicas de previsão, modelagem e análise de negócios para modelar o comportamento ou prever certas variáveis de negócios, economia, finanças, física, fabricação, operações e outras. A execução de modelos de Econometria básica é semelhante à execução da análise de regressão normal, exceto que as variáveis independentes e dependentes podem ser modificadas antes da execução de uma regressão.

- Autoeconometria básica

Semelhante à econometria básica, mas executa automaticamente milhares de variáveis lineares, não lineares, interativas, defasadas e mistas nos dados para determinar o modelo econométrico mais apropriado que descreve o comportamento da variável dependente, útil para modelar os efeitos das variáveis e para prever resultados futuros, sem exigir que o analista seja um econometrista especializado.

- Lógica difusa combinatória

O termo "lógica difusa" é derivado da teoria de conjuntos difusos, ou Fuzzy, para lidar com o raciocínio que seja aproximado e não exato. Essa lógica se contrapõe à "lógica clásica", na qual os conjuntos binários têm lógica binária. As variáveis de lógica difusa podem ter um valor verdadeiro que varie entre 0 e 1 e não seja restrito aos dois valores verdadeiros da lógica proposicional clássica. Esse esquema de ponderação difuso é usado com método combinatório para produzir resultados de previsão de série temporal no Risk Simulator.

- Distribuições personalizadas

Usando o Risk Simulator, é possível coletar opiniões de especialistas e gerar uma distribuição personalizada. Essa técnica de previsão é bastante útil quando o conjunto de dados é pequeno ou o melhor ajuste é ruim quando aplicado a uma rotina de ajuste da distribuição.

- GARCH

O modelo de heteroscedasticidade condicional autorregressiva (GARCH) é usado para modelar níveis de volatilidade históricos e prever a volatilidade futura de um instrumento negociável, por exemplo, preços de ações, preços de commodities, preços do petróleo etc. O conjunto de dados deve ser uma série temporal de níveis de preços brutos. Primeiro, GARCH converte os preços em retornos relativos e executa uma otimização interna para ajustar os dados históricos a uma estrutura a termo de volatilidade de reversão à média, supondo que a volatilidade é natureza heteroscedástica (se altera com o tempo de acordo com algumas características econométricas). Diversas variações dessa metodologia estão disponíveis no Risk Simulator, incluindo EGARCH, EGARCH-T, GARCH-M, GJR-GARCH, GJR-GARCH-T, IGARCH e T-GARCH.

- Curva J

A curva J ou curva de crescimento exponencial na qual o crescimento do próximo período depende do nível do período atual e o aumento é exponencial. Isso significa que ao longo do tempo, os valores aumentarão significativamente de um período para o outro. Normalmente, esse modelo é usado para prever o crescimento biológico e reações químicas ao longo do tempo.

- Cadeias de Markov

Uma cadeia de Markov existe quando a probabilidade de um estado futuro depende de um estado anterior e quando unidos formam uma cadeia se reverte a um nível de estado estável no longo prazo. Essa abordagem é usada para prever a participação de mercado de dois concorrentes. Os valores de entrada obrigatórios são a probabilidade inicial de um cliente na primeira loja (o primeiro estado) retornar à mesma loja no próximo período, comparada com a probabilidade de mudar para uma loja concorrente no próximo estado.

- Máxima verossimilhança em *logit*, *tobit* e *probit*

A estimativa de máxima verossimilhança (MLE) é usada para prever a probabilidade de que algo ocorra, dado algumas variáveis independentes. Por exemplo, usa-se a MLE para prever se uma linha de crédito ou débito ficará inadimplente, dado as características do devedor (30 anos de idade, solteiro, salário de $100.000 por ano e dívida total no cartão de crédito de $10.000); ou a probabilidade de um paciente desenvolver câncer de pulmão, se for um homem entre 50 e 60 anos que fuma 5 maços de cigarros por mês etc. Nessas circunstâncias, a variável dependente é limitada (ou seja, limitada a ser binária 1 e 0 para inadimplência/morrer e adimplência/viver ou limitada a valores inteiros como 1, 2, 3 etc.) e o resultado desejado do modelo é prever a probabilidade de ocorrência de um evento. A análise de regressão tradicional não funciona nessas situações, pois a probabilidade prevista normalmente é menor que zero ou maior que um e muitas das suposições de regressão obrigatórias são violadas, como a independência e a normalidade dos erros e os erros serão grandes.

- Rede neural

O termo “rede neural” é tradicionalmente usado para se referir a uma rede ou um circuito de neurônios biológicos, ao passo que o uso moderno do termo se refere às redes neurais artificiais que compreendem neurônios artificiais ou nós recriados em um ambiente de software. Essa metodologia tenta imitar os neurônios ou o cérebro humano no modo de pensar e identificar padrões, em nosso caso, para identificar padrões com a finalidade de prever dados de série temporal.

- Regressão multivariada

A regressão multivariada é usada para modelar a estrutura de relações e características de uma determinada variável dependente já que ela depende de outras variáveis exógenas independentes. Usando a relação modelada, podemos prever os valores futuros da variável dependente. Também é possível determinar a exatidão e o melhor ajuste para o modelo. Os modelos lineares e não lineares podem ser ajustados na análise de regressão múltipla.

- Extrapolação não linear

Supõe-se que a estrutura fundamental dos dados a serem previstos seja não linear ao longo do tempo. Por exemplo, um conjunto de dados como 1, 4, 9, 16, 25 é considerado como não linear, pois os pontos de dados são de uma função quadrada.

- Curvas S

A curva S, ou curva de crescimento logístico, começa como uma curva J, com taxas de crescimento exponenciais. Com o tempo, o ambiente torna-se saturado (por exemplo, saturação do mercado, competição, superlotação), o crescimento diminui e o valor da previsão por fim termina em um nível máximo de saturação. Esse modelo é muito usado para prever participação de mercado ou o crescimento das vendas de um novo produto, desde introdução no mercado até a maturidade e o declínio, a dinâmica de populações e outros fenômenos naturais.

- Spline cúbico

Às vezes há valores ausentes em um conjunto de dados de uma série temporal. Por exemplo, pode-se ter as taxas de juros dos anos 1 a 3, seguidas pelas dos anos 5 a 8 e 10. As curvas *spline* podem ser usadas para interpolar os valores de taxa de juros dos anos ausentes com base nos dados existentes. As curvas *spline* também podem ser usadas para prever ou extrapolar valores de períodos de tempo futuros além do período de tempo dos dados disponíveis. Os dados podem ser lineares ou não lineares.

- Previsão de processo estocástico

Às vezes as variáveis são estocásticas e não podem ser prontamente previstas usando meios tradicionais. Essas variáveis são chamadas de estocásticas. Porém, a maioria dos fenômenos financeiros, econômicos e naturais (por exemplo, o movimento das moléculas no ar) segue uma lei ou relacionamento matemático conhecido. Apesar de os valores resultantes serem incertos, a estrutura matemática fundamental é conhecida e pode ser simulada usando a simulação de risco Monte Carlo. Os processos para os quais o *Risk Simulator* oferece suporte incluem o movimento de caminho aleatório, reversão à média, difusão com salto e processos mistos, úteis para prever variáveis de séries temporais não estáticas.

- Decomposição e análise de série temporal

Em dados de série temporal bem comportados, entre os quais se incluem as receitas de vendas e as estruturas de custos de grandes corporações, os valores tendem a ter até três elementos, um valor de base, uma tendência e uma sazonalidade. A análise de série temporal usa os dados históricos e os decompõe nesses três elementos, recompondo-os em previsões do futuro. Em outras palavras, esse método de previsão, bem como outros já descritos, primeiro executa um retroajuste (backcast) dos dados históricos antes de fornecer estimativas de valores futuros (previsões).

- Linhas de tendência

As linhas de tendência podem ser usadas para determinar se um conjunto de dados de série temporal segue qualquer tendência considerável. As tendências podem ser lineares ou não lineares (como exponencial, logarítmica, média móvel, potência ou polinomial).

Execução da ferramenta Previsão no Risk Simulator

Em geral, para criar previsões, algumas etapas rápidas são necessárias:

- Inicie o Excel e entre ou abra os dados históricos
- Selecione os dados, clique em *Simulação* e selecione *Previsão*
- Selecione as seções relevantes (ARIMA, Regressão multivariada, Exploração não linear, Previsão estocástica, Análise de série temporal) e insira os valores relevantes

A Figura 3.2 mostra a ferramenta *Previsão* e as várias metodologias.

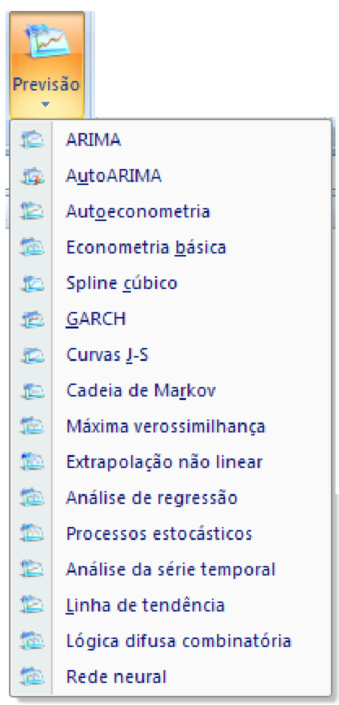


Figura 3.2 – Métodos de previsão do Risk Simulator

A seguir é fornecida uma rápida revisão de cada metodologia e diversos exemplos simples para começar a usar o software. O arquivo de exemplo pode ser encontrado em ***Iniciar | Real Options Valuation | Risk Simulator | Exemplos*** ou acessado diretamente em ***Risk Simulator | Modelos de exemplo.***

Análise da série temporal

Teoria:

A Figura 3.3 lista os oito modelos de série temporal mais comuns separados por sazonalidade e tendência. Por exemplo, se a variável de dados não possuir tendência ou sazonalidade, um modelo de média móvel simples ou de suavização exponencial simples seria suficiente. No entanto, se a sazonalidade existe, mas nenhuma tendência discernível está presente, um modelo sazonal aditivo ou multiplicativo seria melhor etc.

	No Seasonality	With Seasonality
No Trend	*Single Moving Average*	*Seasonal Additive*
	Single Exponential Smoothing	*Seasonal Multiplicative*
With Trend	*Double Moving Average*	*Holt-Winter's Additive*
	Double Exponential Smoothing	*Holt-Winter's Multiplicative*

Figura 3.3 – Os oito métodos de série temporal mais comuns

Procedimento:

- Inicie o Excel e abra os dados históricos, se necessário. O exemplo abaixo usa o arquivo ***Previsão da série temporal*** na pasta de exemplos
- Selecione os dados históricos (os dados devem ser listados em uma única coluna)
- Selecione ***Risk Simulator | Previsão | Análise da série temporal***
- Escolha o modelo a ser aplicado, insira as suposições relevantes e clique em OK

Receita de vendas histórica

Ano	Trimestre	Período	Vendas
2004	1	1	$684.20
2004	2	2	$584.10
2004	3	3	$765.40
2004	4	4	$892.30
2005	1	5	$885.40
2005	2	6	$677.00
2005	3	7	$1,006.60
2005	4	8	$1,122.10
2006	1	9	$1,163.40
2006	2	10	$993.20
2006	3	11	$1,312.50
2006	4	12	$1,545.30
2007	1	13	$1,596.20
2007	2	14	$1,260.40
2007	3	15	$1,735.20
2007	4	16	$2,029.70
2008	1	17	$2,107.80
2008	2	18	$1,650.30
2008	3	19	$2,304.40
2008	4	20	$2,639.40

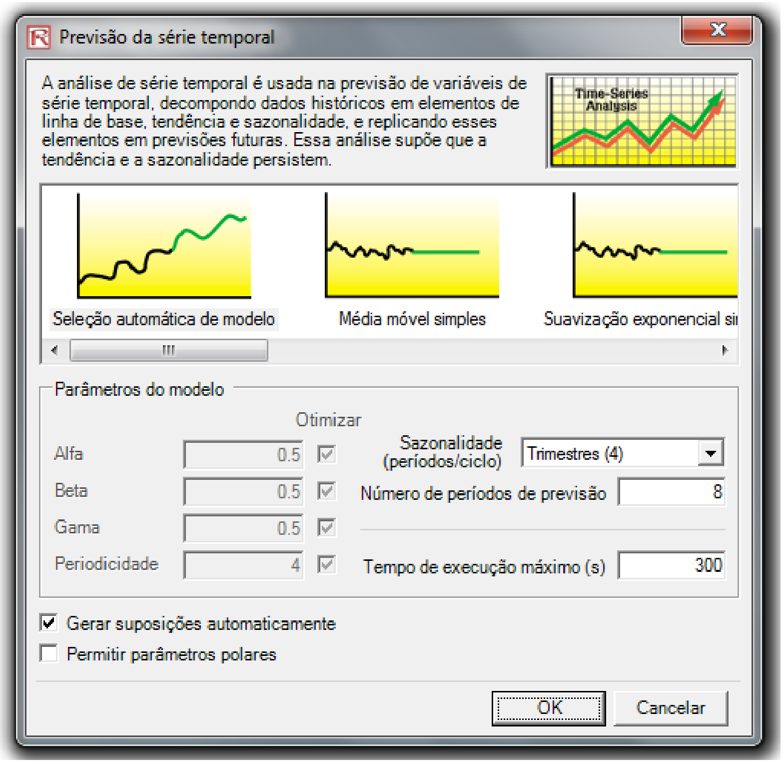


Figura 3.4 – Análise da série temporal

Interpretação dos resultados:

A Figura 3.5 ilustra os resultados gerados usando a ferramenta *Previsão*. O modelo usado foi o multiplicativo de Holt-Winters. Observe que na Figura 3.5, o gráfico de ajuste de modelo e previsão indica que a tendência e a sazonalidade são demonstradas pelo modelo multiplicativo de Holt-Winters. O relatório de análise da série temporal fornece os parâmetros alfa, beta e gama otimizados, as medidas de erro, os dados ajustados, os valores previstos e o gráfico de previsão ajustado. Os parâmetros são somente para referência. Alfa captura os efeitos de memorização das alterações do nível básico ao longo do tempo, beta é o parâmetro de tendência que mede o ponto forte da tendência e gama mede o ponto forte da sazonalidade dos dados históricos. A análise decompõe os dados históricos nesses três elementos e depois os recompõe para prever o futuro. Os dados ajustados ilustram os dados históricos, bem como os dados ajustados que usam o modelo recomposto e mostram a proximidade das previsões no passado (uma técnica chamada backcasting). Os valores previstos são estimativas ou suposições de ponto único (se a opção de geração de suposição automática for escolhida e um perfil de simulação existir). O gráfico ilustra os valores históricos, ajustados e previstos. O gráfico é uma ferramenta visual e de comunicação poderosa para ver a qualidade do modelo de previsão.

Notas:

Este módulo de análise de série temporal contém os oito modelos de séries temporais vistos na Figura 3.3. Você pode selecionar o modelo específico para executar com base nos critérios de tendência e sazonalidade ou escolher Seleção automática de modelo para iterar automaticamente todos os oito métodos, otimizar os parâmetros e encontrar o modelo de melhor ajuste para os dados. Como alternativa, se você escolher um dos oito modelos, poderá desmarcar as caixas de seleção *Otimizar* e inserir seus próprios parâmetros alfa, beta e gama. Consulte o livro do Dr. Johnathan Mun, *Modeling Risk*, 2ª edição*: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* (Wiley, 2010) para obter mais detalhes sobre as especificações técnicas desses parâmetros. Além disso, é necessário inserir os períodos de sazonalidade relevantes se você escolher a seleção de modelo automática ou um dos modelos sazonais. O valor de sazonalidade deve ser um inteiro positivo, por exemplo, se os dados são trimestrais, insira 4 como o número de temporadas ou ciclos por ano, ou insira 12 se os dados forem mensais. Em seguida, insira o número de períodos a serem previstos. Esse valor também deve ser um inteiro positivo. O tempo de execução máximo é definido como 300. Normalmente, nenhuma alteração é necessária. No entanto, durante uma previsão com uma quantidade significativa de dados históricos, a análise pode ser um pouco mais demorada e se o tempo de processamento exceder o tempo de execução, o processo será encerrado. Você também pode definir que a previsão gere suposições automaticamente. Ou seja, em vez de estimativas de ponto único, as previsões serão suposições. Finalmente, a opção de parâmetros polares permite otimizar os parâmetros alfa, beta e gama para incluir zero e um. Alguns softwares de previsão permitem esses parâmetros polares, outros, não. O Risk Simulator permite que você escolha qual deseja usar. Normalmente, não há necessidade de usar parâmetros polares.

Multiplicativa de Holt-Winter

Estatísticas resumidas

Alfa, beta, gama	REQM	Alfa, beta, gama	REQM
0.00, 0.00, 0.00	914.824	0.00, 0.00, 0.00	914.824
0.10, 0.10, 0.10	415.322	0.10, 0.10, 0.10	415.322
0.20, 0.20, 0.20	187.202	0.20, 0.20, 0.20	187.202
0.30, 0.30, 0.30	118.795	0.30, 0.30, 0.30	118.795
0.40, 0.40, 0.40	101.794	0.40, 0.40, 0.40	101.794
0.50, 0.50, 0.50	102.143		

A análise foi executada com alfa = 0.2429, beta = 1.0000, gama = 0.7797 e sazonalidade = 4

Resumo da análise de série temporal

Quando há sazonalidade e tendência, são necessários modelos mais avançados para decompor os dados em seus elementos básicos: um nível de caso de base (L) ponderado pelo parâmetro alfa; um componente de tendência (b) ponderado pelo parâmetro beta; e um componente de sazonalidade (S) ponderado pelo parâmetro gama. Existem vários métodos, mas os mais comuns são os métodos de sazonalidade aditiva de Holt-Winters e de sazonalidade multiplicativa de Holt-Winters. No modelo de sazonalidade aditiva de Holt-Winter, o nível de caso de base, a sazonalidade e a tendência somados para obter o ajuste de previsão.

O teste de melhor ajuste para a previsão da média móvel usa a raiz do erro quadrático médio (REQM). O REQM calcula a raiz quadrada dos desvios quadrados da média dos valores ajustados versus os pontos de dados reais.

O erro quadrático médio (EQM) é uma medida de erro absoluto que eleva os erros ao quadrado (a diferença entre os dados históricos reais e os dados ajustados à previsão previstos pelo modelo) para evitar que os erros positivos e negativos cancelem um aos outros. Essa medida também tende a exagerar erros grandes, pois eleva seu peso ao quadrado e, assim, atribui um peso maior do que a erros pequenos, o que pode ajudar na comparação de modelos de séries temporais diferentes. A REQM é a raiz quadrada do EQM e é a medida de erro mais popular, também conhecida como função perda quadrática. A REQM pode ser definida como a média dos valores absolutos dos erros de previsão e é muito apropriada quando o custo desses erros é proporcional ao seu tamanho absoluto. A REQM é usada como o critério para seleção do modelo de séries temporais mais adequado.

O erro médio percentual absoluto (MAPE) é uma estatística de erro relativo medida como uma média do erro percentual dos pontos de dados históricos e é mais adequada quando o custo do erro de previsão está relacionado mais intimamente ao erro percentual do que ao tamanho numérico do erro. Por último, uma medida associada é a estatística U-Theil, que mede a ingenuidade da previsão do modelo. Ou seja, se a estatística U-Theil for menor do que 1,0, então o método de previsão usado fornecerá uma estimativa que é estatisticamente melhor do que um palpite.

Período	Real	Ajuste de previsão
1	684.20	
2	584.10	
3	765.40	
4	892.30	
5	885.40	684.20
6	677.00	667.55
7	1006.60	935.45
8	1122.10	1198.09
9	1163.40	1112.48
10	993.20	887.95
11	1312.50	1348.38
12	1545.30	1546.53
13	1596.20	1572.44
14	1260.40	1299.20
15	1735.20	1704.77
16	2029.70	1976.23
17	2107.80	2026.01
18	1650.30	1637.28
19	2304.40	2245.93
20	2639.40	2643.09
Previsão21		2713.69
Previsão22		2114.79
Previsão23		2900.42
Previsão24		3293.81
Previsão25		3346.55
Previsão26		2580.81
Previsão27		3506.19
Previsão28		3947.61

Medidas de erro	
REQM	71.8132
EQM	5157.1348
MAD	53.4071
MAPE	4.50%
U-Theil	0.3054

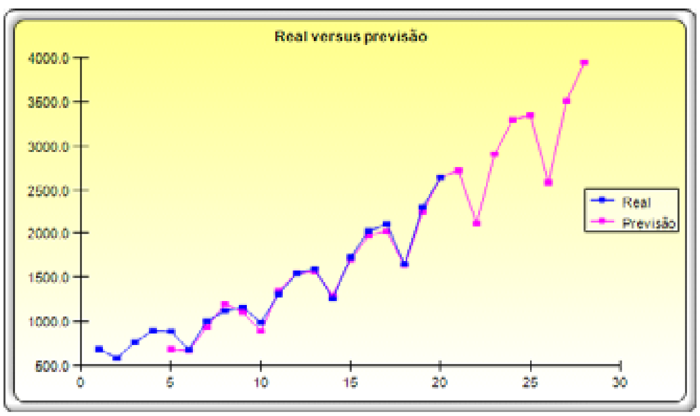


Figura 3.5 – Exemplo de relatório de previsão de Holt-Winters

Regressão multivariada

Teoria:

Supõe-se que o usuário tenha conhecimentos suficientes sobre os fundamentos de análise de regressão. A equação de regressão linear bivariada assume a forma de $Y = \beta_0 + \beta_1 X + \varepsilon$ onde β_0 é a intercepção, β_1 é a inclinação e ε é o termo de erro. Ela é bivariada porque tem apenas duas variáveis, uma variável dependente *Y* e uma variável independente *X*, onde *X* também é conhecido como o regressor (às vezes uma regressão bivariada também é conhecida como uma regressão univariada, pois há apenas uma variável independente *X*). A variável dependente recebe esse nome por *depender* da variável independente, por exemplo, a receita de vendas depende dos custos com marketing com propaganda e promoção de um produto, ou seja, as vendas são a variável dependente e os custos com marketing são a variável independente. Um exemplo de regressão bivariada é a simples inserção da linha de melhor ajuste em um conjunto de pontos de dados em um plano bidimensional, como visto no painel esquerdo da Figura 3.6. Em outros casos, é possível executar uma regressão multivariada, na qual haja vários ou *n* números de variáveis *X* independentes, nas quais a equação de regressão geral terá agora a forma de $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 ... + \beta_n X_n + \varepsilon$. Nesse caso, a melhor linha de ajuste estará em um plano de *n + 1* dimensões.

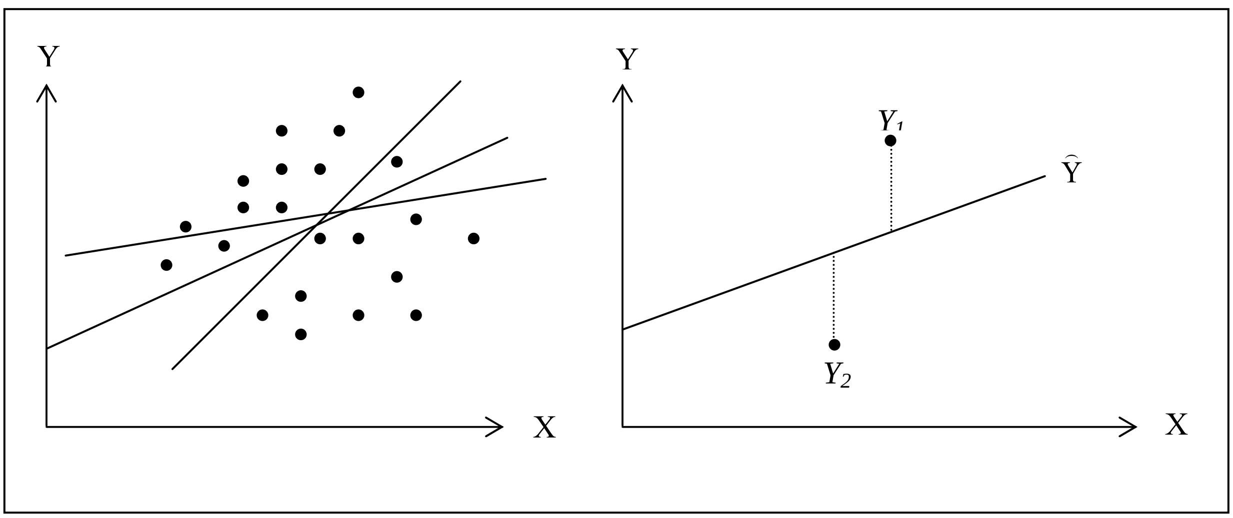


Figura 3.6 – Regressão bivariada

No entanto, ajustar uma linha através de um conjunto de pontos de dados em um diagrama de dispersão, como na Figura 3.6, pode resultar em diversas linhas possíveis. A linha de melhor ajuste é definida como a única linha que minimiza o total de erros verticais, ou seja, a soma das distâncias absolutas entre os pontos de dados reais (Y_i) e a linha estimada ($\hat{Y}$), como mostrado no painel da direita da Figura 3.6. Para encontrar a melhor linha de ajuste que minimiza os erros, é necessária uma abordagem mais sofisticada: a análise de regressão. Portanto, a análise de regressão encontra a única linha de melhor ajuste exigindo que os erros totais sejam minimizados ou calculando;

$$Min\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

onde apenas uma linha única minimiza essa soma dos erros elevados ao quadrado. Os erros (distância vertical entre os dados reais e a linha prevista) são elevados ao quadrado para evitar que os erros negativos cancelem os erros positivos. A solução desse problema de minimização com relação à inclinação e à intercepção requer primeiro o cálculo da derivada primeira e defini-la como zero:

$$\frac{d}{d\beta_0}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = 0 \text{ e } \frac{d}{d\beta_1}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = 0$$

o que resulta nas equações quadráticas dos mínimos da regressão bivariada:

$$\beta_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\sum_{i=1}^{n}X_iY_i - \frac{\sum_{i=1}^{n}X_i\sum_{i=1}^{n}Y_i}{n}}{\sum_{i=1}^{n}X_i^2 - \frac{\left(\sum_{i=1}^{n}X_i\right)^2}{n}}$$

$$\beta_0 = \overline{Y} - \beta_1\overline{X}$$

Para a regressão multivariada, a analogia é ampliada para explicar as múltiplas variáveis independentes, em que $Y_i = \beta_1 + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \varepsilon_i$ e as inclinações estimadas são calculadas por:

$$\hat{\beta}_2 = \frac{\sum Y_i X_{2,i}\sum X_{3,i}^2 - \sum Y_i X_{3,i}\sum X_{2,i}X_{3,i}}{\sum X_{2,i}^2\sum X_{3,i}^2 - \left(\sum X_{2,i}X_{3,i}\right)^2}$$

$$\hat{\beta}_3 = \frac{\sum Y_i X_{3,i}\sum X_{2,i}^2 - \sum Y_i X_{2,i}\sum X_{2,i}X_{3,i}}{\sum X_{2,i}^2\sum X_{3,i}^2 - \left(\sum X_{2,i}X_{3,i}\right)^2}$$

A execução de regressões multivariadas requer muita atenção na configuração e interpretação dos resultados. Por exemplo, é necessário ter boa compreensão do modelo econométrico, por exemplo, identificar armadilhas de regressão como quebras estruturais, multicolinearidade, heteroscedasticidade, autocorrelação, testes de especificação, não linearidades etc., antes que um modelo apropriado possa ser construído. Consulte *Modeling Risk*, 2ª edição*: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* (Wiley, 2010) do Dr. Johnathan Mun, para uma análise mais detalhada e uma discussão sobre a regressão multivariada, bem como para identificar as armadilhas de regressão.

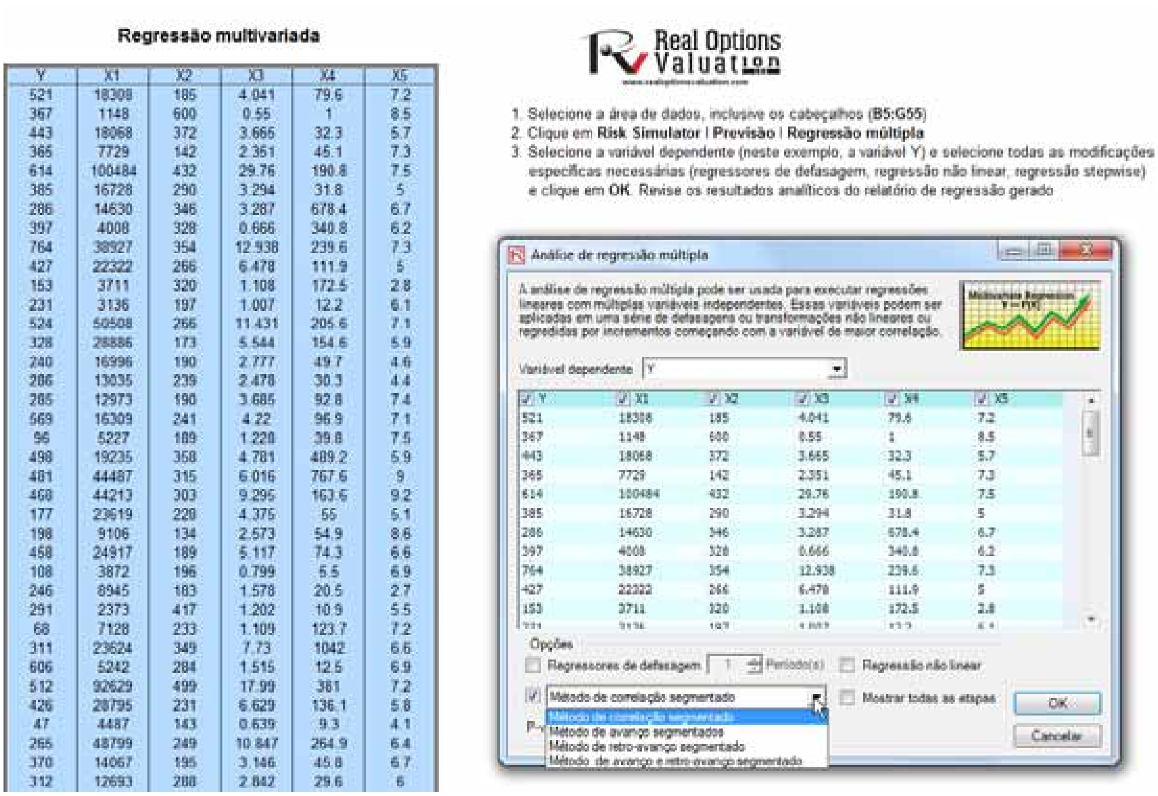
Procedimento:

- Inicie o Excel e abra os dados históricos, se necessário. A ilustração a seguir usa o arquivo ***Regressão múltipla*** na pasta de exemplos
- Verifique se os dados estão dispostos em colunas, selecione toda a área de dados, incluindo o nome da variável, e selecione ***Risk Simulator | Previsão | Regressão múltipla***

- Selcione a variável dependente e marque as opções relevantes (defasagem, regressão stepwise , regressão não linear etc.) e clique em OK

Interpretação dos resultados:

A Figura 3.8 ilustra um exemplo de relatório dos resultados de regressão multivariada. O relatório contém todos os resultados da regressão, a análise dos resultados de variância, o gráfico ajustado e os resultados do teste de hipóteses. Os detalhes técnicos da interpretação desses resultados não fazem parte do escopo deste manual. Consulte *Modeling Risk*, 2ª edição*: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization* (Wiley, 2010) do Dr. Johnathan Mun, para uma análise mais detalhada e uma discussão sobre a regressão multivariada e a interpretação dos relatórios de regressão.



Regressão multivariada

Y	X1	X2	X3	X4	X5
521	18308	185	4.041	79.6	7.2
367	1148	600	0.55	1	8.5
443	18068	372	3.665	32.3	5.7
365	7729	142	2.351	45.1	7.3
614	100484	432	29.76	190.8	7.5
385	16728	290	3.294	31.8	5
286	14630	346	3.287	678.4	6.7
397	4008	328	0.666	340.8	6.2
764	38927	354	12.938	239.6	7.3
427	22322	266	6.478	111.9	5
153	3711	320	1.108	172.5	2.8
231	3136	197	1.007	12.2	6.1
524	50508	266	11.431	205.6	7.1
328	28886	173	5.544	154.6	5.9
240	16996	190	2.777	49.7	4.6
286	13035	239	2.478	30.3	4.4
285	12973	190	3.685	92.8	7.4
569	16309	241	4.22	96.9	7.1
96	5227	189	1.228	39.8	7.5
498	19235	358	4.781	489.2	5.9
481	44487	315	6.016	767.6	9
468	44213	303	9.295	163.6	9.2
177	23619	220	4.375	55	5.1
198	9106	134	2.573	54.9	8.6
458	24917	189	5.117	74.3	6.6
108	3872	196	0.799	5.5	6.9
246	8945	183	1.578	20.5	2.7
291	2373	417	1.202	10.9	5.5
68	7128	233	1.109	123.7	7.2
311	23624	349	7.73	1042	6.6
606	5242	284	1.515	12.5	6.9
512	92629	499	17.99	381	7.2
426	28795	231	6.629	136.1	5.8
47	4487	143	0.639	9.3	4.1
265	48799	249	10.847	264.9	6.4
370	14067	195	3.146	45.8	6.7
312	12693	288	2.842	29.6	6

Figura 3.7 – Execução de uma regressão multivariada

Relatório da análise de regressão

Estatísticas de regressão

R2 (coeficiente de determinação)	0.3272
R2 ajustado	0.2508
R múltiplo (coeficiente de correlação múltipla)	0.5720
Erro padrão das estimativas (SEy)	149.6720
Número de observações	50

R2, ou coeficiente de determinação, indica que 0.33 da variação da variável dependente que pode ser explicado e contabilizado pelas variáveis independentes nessa análise de regressão. No entanto, em uma regressão múltipla, o R2 ajustado considera a existência de outras variáveis independentes ou regressores e ajusta esse valor de R2 para obter uma visão mais precisa da força de explicação da regressão. Assim, apenas 0.25 da variação da variável dependente pode ser explicado pelos regressores.

O coeficiente de correlação múltipla (R múltiplo) mede a correlação entre a variável dependente real (Y) e a estimada ou ajustada (Y) com base na equação de regressão. Também é a raiz quadrada do coeficiente de determinação (R2).

O erro padrão das estimativas (SEy) descreve a dispersão dos pontos de dados acima e abaixo da linha ou do plano de regressão. Esse valor é usado como parte do cálculo para obter o intervalo de confiança das estimativas depois.

Resultados da regressão

	Interceptação	X1	X2	X3	X4	X5
Coeficientes	57.9555	-0.0035	0.4644	25.2377	-0.0086	16.5579
Erro padrão	108.7901	0.0035	0.2535	14.1172	0.1016	14.7996
Estatística t	0.5327	-1.0066	1.8316	1.7877	-0.0843	1.1188
P-valor	0.5969	**0.3197**	**0.0738**	**0.0807**	**0.9332**	**0.2693**
5% mais baixos	-161.2966	-0.0106	-0.0466	-3.2137	-0.2132	-13.2687
95% mais altos	277.2076	0.0036	0.9753	53.6891	0.1961	46.3845

Graus de liberdade	
Graus de liberdade para regressão	5
Graus de liberdade para residual	44
Total de graus de liberdade	49

Teste de hipóteses	
Estatística t crítica (99% de confiança com df igual a 44)	2.6923
Estatística t crítica (95% de confiança com df igual a 44)	2.0154
Estatística t crítica (90% de confiança com df igual a 44)	1.6802

Os coeficientes fornecem a interceptação e a inclinação estimadas da regressão. Por exemplo, os coeficientes são estimativas dos verdadeiros valores da população b nesta equação de regressão: Y = b0 + b1X1 + b2X2 + ... + bnXn. O erro padrão mede a precisão dos coeficientes previstos, e as estatísticas t são os índices de cada coeficiente previsto para seu erro padrão.

A estatística t é usada em testes de hipóteses, nos quais é possível definir a hipótese nula (Ho) de tal maneira que a média real do coeficiente = 0, e a hipótese alternativa (Ha) de tal maneira que a média real do coeficiente seja diferente de 0. Um teste t é executado e a estatística t calculada é comparada aos valores críticos nos graus de liberdade relevantes para o residual. O teste t é muito importante, pois calcula se cada um dos coeficientes é estatisticamente significativo na presença de outros regressores. Isso significa que o teste t verifica estatisticamente se um regressor ou uma variável independente deve permanecer na regressão ou ser descartada.

O coeficiente é estatisticamente significativo se a sua estatística t calculada excede a estatística t crítica nos graus de liberdade relevantes (df). Os três principais níveis de confiança usados para testar a significância são 90%, 95% e 99%. Se a estatística t de um coeficiente exceder o nível crítico, ela será considerada estatisticamente relevante. Como alternativa, o p-valor calcula a probabilidade de ocorrência de cada estatística t, o que significa que quanto menor o p-valor, mais significativo é o coeficiente. Os níveis de significância do p-valor costumam ser 0,01, 0,05 e 0,10, o que corresponde aos níveis de confiança 99%, 95% e 90%.

Os coeficientes cujos p-valores são realçados em azul indicam que são estatisticamente significativos no nível de confiança de 90% ou no nível alfa de 0,10, enquanto aqueles cujos p-valores são realçados em vermelho indicam que não são estatisticamente relevantes em qualquer outro nível alfa.

Análise de variância

	Somas dos quadrados	Média dos quadrados	Estatística F	P-valor
Regressão	479388.49	95877.70	4.28	0.0029
Residual	985675.19	22401.71		
Total	1465063.68			

Teste de hipóteses

Estatística F crítica (99% de confiança com df igual a 5 e 44)	3.4651
Estatística F crítica (95% de confiança com df igual a 5 e 44)	2.4270
Estatística F crítica (90% de confiança com df igual a 5 e 44)	1.9828

A tabela de análise de variância (ANOVA) fornece um teste F da significância estatística geral do modelo de regressão. Em vez de analisar regressores individuais como no teste t, o teste F analisa as propriedades estatísticas de todos os coeficientes estimados. A estatística F é calculada como a proporção da média dos quadrados da regressão em relação à média dos quadrados de residuais. O numerador mede quanto da regressão é explicado, enquanto o denominador mede quanto não é explicado. Assim, quanto maior a estatística F, mais significante é o modelo. O p-valor correspondente é calculado para testar a hipótese nula (Ho), segundo a qual todos os coeficientes são simultaneamente iguais a zero, em contraste com a hipótese alternativa (Ha) de que eles são simultaneamente diferentes de zero, indicando um modelo de regressão geral significante. Se o p-valor for menor do que as significâncias alfa 0,01, 0,05 ou 0,10, a regressão será significante. A mesma abordagem pode ser aplicada à estatística F comparando a estatística F calculada aos valores F críticos em vários níveis de significância.

Previsão

Período	Real (Y)	Previsão (F)	Erro (E)
1	521.0000	299.5124	221.4876
2	367.0000	487.1243	(120.1243)
3	443.0000	353.2789	89.7211
4	365.0000	276.3296	88.6704
5	614.0000	776.1336	(162.1336)
6	385.0000	298.9993	86.0007
7	286.0000	354.8718	(68.8718)
8	397.0000	312.6155	84.3845
9	764.0000	529.7550	234.2450
10	427.0000	347.7034	79.2966
11	153.0000	266.2526	(113.2526)
12	231.0000	264.6375	(33.6375)
13	524.0000	406.8009	117.1991
14	328.0000	272.2226	55.7774
15	240.0000	231.7882	8.2118
16	286.0000	257.8862	28.1138
17	285.0000	314.9521	(29.9521)
18	569.0000	335.3140	233.6860
19	96.0000	282.0356	(186.0356)
20	498.0000	370.2062	127.7938
21	481.0000	340.8742	140.1258
22	468.0000	427.5118	40.4882
23	177.0000	274.5298	(97.5298)
24	198.0000	294.7795	(96.7795)
25	458.0000	295.2180	162.7820
26	108.0000	269.6195	(161.6195)
27	246.0000	195.5955	50.4045

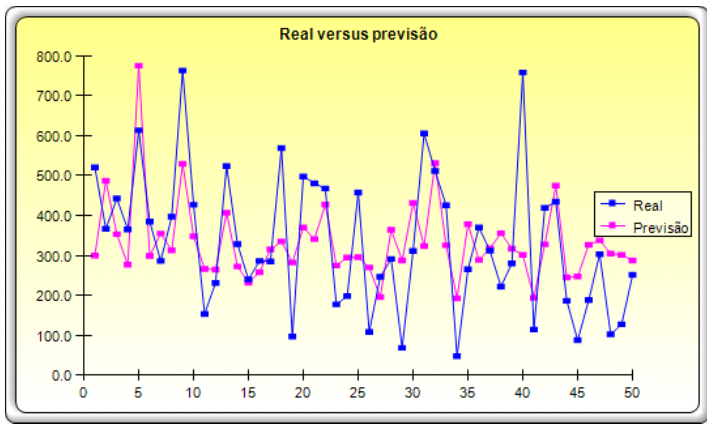


Figura 3.8 – Resultados de regressão multivariada

Previsão estocástica

Teoria:

Um processo estocástico nada mais é do que uma equação definida matematicamente que pode criar uma série de resultados ao longo do tempo, os quais não são de natureza determinista. Ou seja, uma equação ou um processo que não segue uma regra simples discernível, como preço aumentará *X* por cento todo ano ou a receita aumentará pelo fator de *X* mais *Y* por cento. Um processo estocástico é, por definição, não determinístico e uma pessoa pode inserir números em uma equação de processo estocástico e obter resultados diferentes a cada vez. Por exemplo, a trajetória do preço de uma ação é de natureza estocástica e não é possível prevê-la com certeza. No entanto, a evolução do preço ao longo do tempo se insere em um processo que gera esses preços. O processo é fixo e pré-determinado, mas os resultados não são. Assim, por simulação estocástica, criamos várias trajetórias de preços, obtemos amostras estatísticas dessas simulações e fazemos inferências quanto às trajetórias potenciais que o preço real pode tomar, dada a natureza e os parâmetros do processo estocástico, usados para gerar a série temporal. Três processos estocásticos básicos estão incluídos na ferramenta *Previsão do Risk Simulator*, incluindo o movimento browniano geométrico ou caminho aleatório, que é o processo mais comum e usado devido à sua simplicidade e variadas aplicações. Os dois outros processos estocásticos são; o processo de reversão à média e processo de difusão com salto.

O interessante sobre a simulação de processo estocástico é que os dados históricos não são necessariamente obrigatórios. Ou seja, o modelo não precisa ajustar conjuntos de dados históricos. Basta calcular os retornos e a volatilidade esperados dos dados históricos, estimá-los usando dados externos comparáveis ou fazer suposições sobre esses valores. Consulte *Modeling Risk*, 2ª edição*: Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization,* 2ª edição (Wiley 2010) de Dr. Johnathan Mun, para obter mais detalhes sobre como cada um dos valores de entrada é calculado, por exemplo, taxa de reversão à média, probabilidades de salto, volatilidade etc.

Procedimento:

- Para iniciar o módulo, selecione ***Risk Simulator | Previsão | Processos estocásticos***
- Selecione o processo desejado, insira os valores obrigatórios, clique em Atualizar gráfico algumas vezes para verificar se o processo está se comportando da maneira esperada e clique em OK (Figura 3.9)

Interpretação dos resultados:

A Figura 3.10 mostra os resultados de um processo estocástico de exemplo. O gráfico mostra um conjunto de exemplo das iterações enquanto o relatório explica o básico sobre processos estocásticos. Além disso, são fornecidos os valores de previsão (média e desvio padrão) para cada período de tempo. Usando esses valores, você pode decidir qual período de tempo é relevante para a análise e definir suposições com base nesses valores de média e desvio padrão usando a distribuição normal. Essas suposições podem então ser simuladas no seu próprio modelo personalizado.

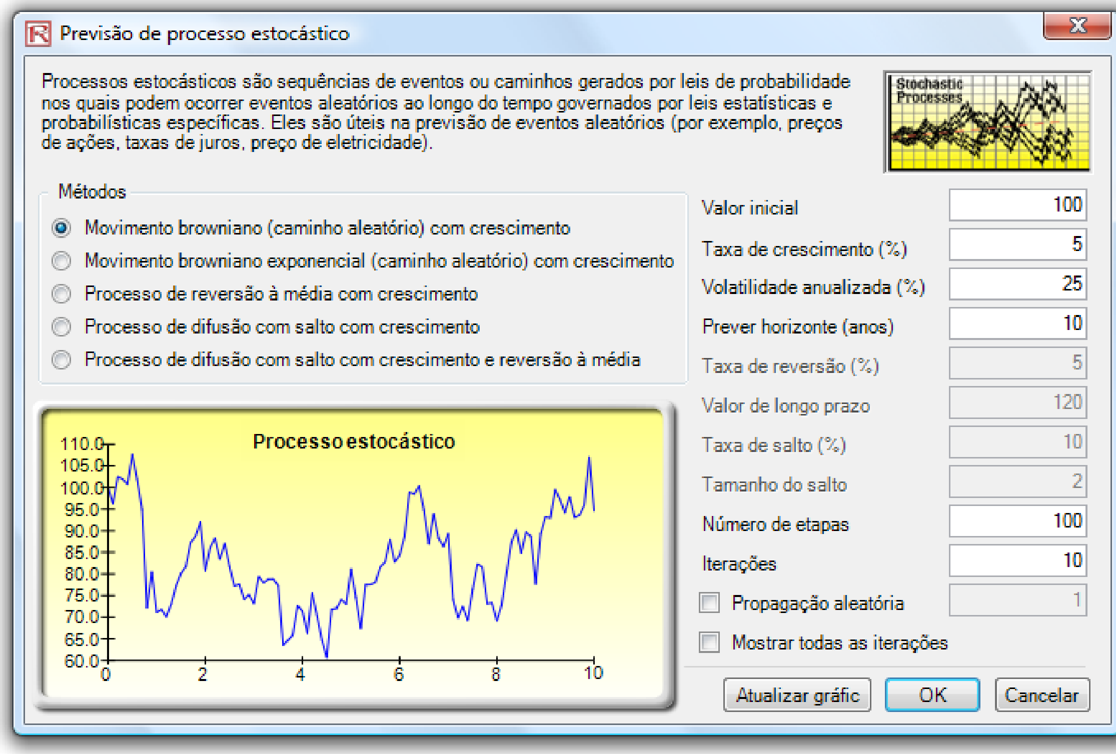


Figura 3.9 – Previsão de processo estocástico

Previsão de processo estocástico

Resumo estatístico

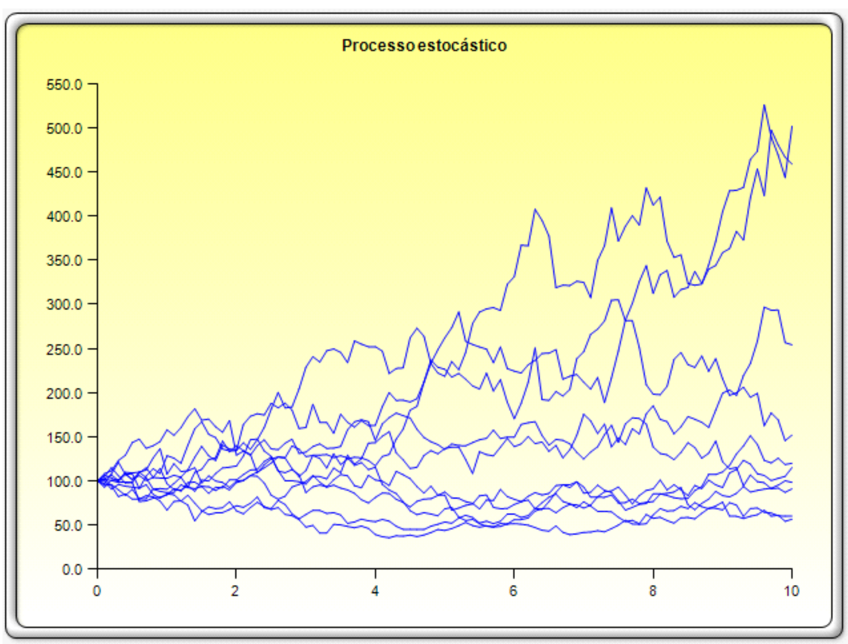
Um processo estocástico é uma sequência de eventos ou caminhos gerados pelas leis de probabilidade. Ou seja, eventos aleatórios podem ocorrer ao longo do tempo, mas são regidos por leis estatísticas probabilísticas específicas. Os principais processos estocásticos incluem caminho aleatório ou movimento browniano, reversão à média e difusão com salto. Esses processos podem ser usados para prever diversas variáveis que parecem seguir tendências aleatórias, mas ainda são restringidos pelas leis probabilísticas.

O processo de caminho aleatório ou movimento browniano pode ser usado na previsão de preços de ações, preços de commodities e outros dados de série temporal estocásticos, conhecida uma taxa de crescimento e uma volatilidade ao redor do caminho de crescimento. Os processos de reversão à média podem ser usados para reduzir as flutuações do processo de caminho aleatório permitindo que o caminho tenha por meta um valor de longo prazo, tornando-o útil na previsão de variáveis de série temporal que possuem uma taxa de longo prazo, como taxas de juros e de inflação, que são taxas de longo prazo usadas por órgãos reguladores ou pelo mercado). O processo de difusão com salto é útil para prever dados de série temporal quando a variável pode, ocasionalmente, apresentar saltos aleatórios, como os preços do petróleo ou da eletricidade (choques de evento exógeno discreto podem fazer preços subirem ou descerem drasticamente). Finalmente, esses três processos estocásticos podem ser combinados conforme a necessidade.

Os resultados à direita indicam a média e o desvio padrão de todas as iterações geradas a cada etapa. Se a opção Mostrar todas as iterações for selecionada, cada caminho de iteração será mostrado em uma planilha separada. O gráfico gerado abaixo mostra um conjunto de amostras dos caminhos de iteração.

Processo estocástico: Movimento browniano (caminho aleatório) com crescimento

Valor inicial	100	Etapas	100.00	Taxa de salto	N/A
Taxa de crescimento	5.00%	Iterações	10.00	Tamanho do salto	N/A
Volatilidade	25.00%	Taxa de reversão	N/A	Propagação aleatória	183197059
Horizonte	10	Valor de longo prazo	N/A		



Tempo	Média	Desvpad
0.0000	100.00	0.00
0.1000	100.88	5.03
0.2000	102.49	7.80
0.3000	101.59	12.59
0.4000	101.96	13.38
0.5000	102.07	16.89
0.6000	99.95	21.30
0.7000	100.80	18.23
0.8000	101.44	19.93
0.9000	102.38	23.57
1.0000	100.56	26.00
1.1000	102.81	23.14
1.2000	102.56	24.50
1.3000	106.97	29.55
1.4000	109.82	38.26
1.5000	110.34	36.91
1.6000	109.09	34.70
1.7000	108.04	32.11
1.8000	110.77	32.30
1.9000	111.25	31.49
2.0000	109.07	26.33
2.1000	114.23	32.22
2.2000	118.18	31.60
2.3000	120.34	31.49
2.4000	120.68	35.01
2.5000	122.35	41.37
2.6000	122.11	43.21
2.7000	119.15	43.01
2.8000	119.25	43.29
2.9000	114.91	44.99
3.0000	115.17	51.98
3.1000	124.39	56.12
3.2000	120.20	54.13
3.3000	118.60	57.79
3.4000	122.20	55.07
3.5000	122.53	56.45
3.6000	123.28	55.32
3.7000	122.94	62.09
3.8000	123.91	61.88
3.9000	124.09	61.30
4.0000	121.04	60.85
4.1000	124.76	63.41
4.2000	123.78	62.22
4.3000	123.51	62.76
4.4000	121.58	64.71
4.5000	123.88	73.86
4.6000	123.71	76.64

Figura 3.10 – Resultado de previsão estocástica

Extrapolação não linear

Teoria:

A extrapolação envolve a criação de projeções estatísticas usando tendências históricas projetadas para um período de tempo específico no futuro. Usa-se somente para previsões de séries temporais. Para dados de seção transversal ou combinados (série temporal com dados de seção transversal), a regressão multivariada é mais apropriada. Esta metodologia é útil quando não se espera alterações significativas, ou seja, espera-se que os fatores causais permaneçam constantes ou quando os fatores causais de uma situação não são claramente entendidos. Isso também desestimula a inserção de vieses pessoais no processo. A extrapolação é bastante confiável, relativamente simples e barata. No entanto, a extrapolação, que supõe que tendências recentes e históricas terão continuidade, produz grandes erros de previsão em caso de descontinuidades dentro do período de tempo projetado. Isto é, a extrapolação pura da série temporal supõe que tudo o que se conhece esteja contido nos valores históricos da série que está sendo prevista. Se supormos que o comportamento passado é um bom indicador do comportamento futuro, a extrapolação será atraente. Assim, ela é uma abordagem útil quando só se necessita de muitas previsões de curto prazo.

Esta metodologia estima a função *f(x)* para qualquer valor de *x* arbitrário, interpolando uma curva não linear suavizada em todos os valores de *x*, e usando essa curva suavizada, extrapola os valores futuros de *x* além do conjunto de dados histórico. O método emprega a forma da função polinomial ou a forma funcional racional (uma proporção de dois polinômios). Em geral, uma forma funcional polinomial é suficiente para dados bem comportados, entretanto, as formas funcionais racionais podem ser às vezes mais exatas (principalmente com funções polares, ou seja, funções com denominadores que se aproximam de zero).

Procedimento:

- Inicie o Excel e abra os dados históricos, se necessário. A ilustração abaixo usa o arquivo ***Extrapolação não linear*** na pasta de exemplos
- Escolha os dados da série temporal e selecione ***Risk Simulator | Previsão | Extrapolação não linear***
- Selecione o tipo de extrapolação (seleção automática, função polinomial ou função racional), insira o número do período de previsão desejado (Figura 3.11) e clique em OK

Interpretação dos resultados:

O relatório de resultados mostrado na Figura 3.12 mostra os valores de previsão extrapolados, as medidas de erro e a representação gráfica dos resultados de extrapolação. As medidas de erro devem ser usadas para verificar a validade da previsão e são especialmente importantes quando usadas para comparar a qualidade de previsão e a precisão da extrapolação versus a análise da série temporal.

Notas:

Quando os dados históricos são suaves e seguem curvas e padrões não lineares, a extrapolação é melhor do que a análise de série temporal. No entanto, quando padrões da dados seguem ciclos sazonais e uma tendência, a análise de série temporal fornece resultados melhores.

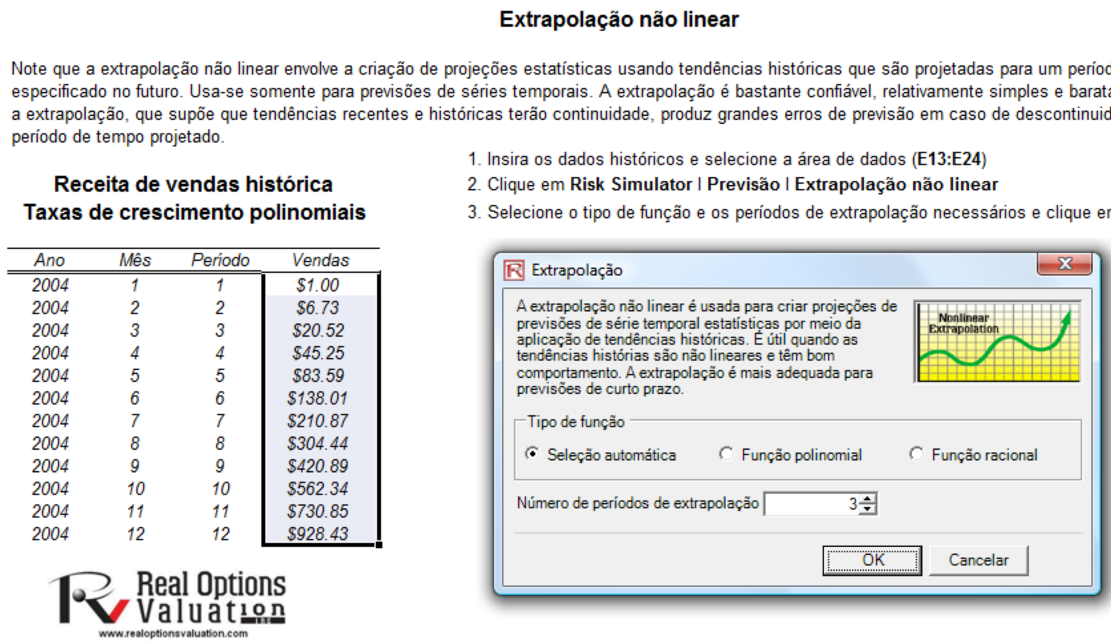


Figura 3.11 – Execução de uma extrapolação não linear

Extrapolação não linear

Resumo estatístico

A extrapolação envolve a criação de projeções estatísticas usando tendências históricas projetadas para um período de tempo específico no futuro. Usa-se somente para previsões de séries temporais. Par dados de seção transversão ou combinados (série temporal com dados de seção transversal), a regressão multivariada é mais apropriada. Esta metodologia é útil quando não se espera alterações significativas, ou seja, espera-se que os fatores causais permaneçam constantes ou quando os fatores causais de uma situação não são claramente entendidos. Isso também desestimula a inserção de vieses pessoais no processo. A extrapolação é bastante confiável, relativamente simples e barata. No entanto, a extrapolação, que supõe que tendências recentes e históricas terão continuidade, produz grandes erros de previsão em caso de descontinuidades dentro do período de tempo projetado. Isto é, a extrapolação pura da série temporal supõe que tudo o que se conhece esteja contido nos valores históricos da série que está sendo prevista. Se supormos que o comportamento passado é um bom indicador do comportamento futuro, a extrapolação será atraente. Assim, ela é uma abordagem útil quando só se necessita de muitas previsões de curto prazo.

Esta metodologia estima a função f(x) para qualquer valor de x arbitrário, interpolando uma curva não linear suavizada em todos os valores de x, e usando essa curva suavizada, extrapola os valores futuros de x além do conjunto de dados histórico. O método emprega a forma da função polinomial ou a forma funcional racional (uma proporção de dois polinômios). Em geral, uma forma funcional polinomial é suficiente para dados bem comportados, entretanto, as formas funcionais racionais podem ser às vezes mais exatas (principalmente com funções polares, ou seja, funções com denominadores que se aproximam de zero).

Período	Real	Ajuste de previsão
1	1.00	
2	6.73	1.00
3	20.52	-1.42
4	45.25	99.82
5	83.59	55.92
6	138.01	136.71
7	210.87	211.96
8	304.44	304.43
9	420.89	420.89
10	562.34	562.34
11	730.85	730.85
12	928.43	928.43
Previsão 13		1157.03
Previsão 14		1418.57
Previsão 15		1714.95

Medidas de erro	
REQM	19.6799
EQM	387.2974
MAD	10.2095
MAPE	31.56%
U-Theil	1.1210

Tipo de função: Racional

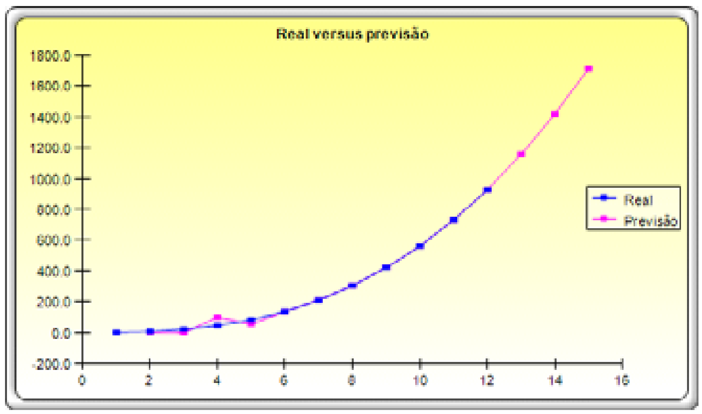


Figura 3.12 – Resultados de extrapolação não linear

Série temporal avançada Box-Jenkins ARIMA

Teoria:

Uma ferramenta de previsão de série temporal avançada muito poderosa é ARIMA ou *média móvel integrada autorregressiva*. A previsão ARIMA agrupa três ferramentas separadas em um modelo abrangente. O primeiro segmento da ferramenta é o termo autorregressivo ou “AR", que corresponde ao número do valor defasado do residual no modelo de previsão incondicional. Essencialmente, o modelo captura a variação histórica de dados reais para um modelo de previsão e usa essa variação ou residual para criar um modelo de previsão melhor. O segundo segmento da ferramenta é a ordem de integração ou o termo “I”. Esse termo de integração corresponde ao número da série temporal diferenciada a ser prevista. Esse elemento explica qualquer taxa de crescimento não linear existente nos dados. O terceiro termo é a média móvel ou “MA", que é essencialmente a média móvel de erros de previsão defasados. Pela incorporação desses erros de previsão defasados, o modelo aprende a partir de seus erros de previsão e os corrige por meio do cálculo de média móvel. O modelo ARIMA segue a metodologia de Box-Jenkins com cada termo representando etapas na construção do modelo até que apenas ruídos aleatórios permaneçam. Além disso, o modelo ARIMA usa técnicas de correlação para gerar previsões. É possível usar ARIMA para formar padrões que podem não estar visíveis em dados "plotados". Além disso, os modelos ARIMA podem ser combinados com variáveis exógenas, mas essas variáveis devem ter pontos de dados suficientes para cobrir o número adicional de períodos a serem previstos. Finalmente, devido à complexidade dos modelos, este módulo pode levar mais tempo para ser executado.

Há muitas razões pelas quais um modelo ARIMA é superior à análise de série temporal comum e a regressões multivariadas. O senso comum na análise de série temporal e na regressão multivariada é que os residuais de erro estão correlacionados aos seus próprios valores defasados. Essa correlação serial viola a suposição padrão da teoria de regressão de que perturbações não estão correlacionadas a outras perturbações. Os principais problemas associados à correlação serial são:

- A análise de regressão e a análise de série temporal básica não são mais eficientes entre os diferentes estimadores lineares. No entanto, como os residuais de erro podem ajudar a prever os residuais de erros atuais, podemos usar essa informação para formar uma previsão melhor da variável dependente usando ARIMA.
- Os erros padrão calculados usando a fórmula de regressão e série temporal não são corretos e geralmente não são fiéis e se há um conjunto de variáveis dependentes defasadas definidas como regressores, as estimativas de regressão são tendenciosas e inconsistentes, mas podem ser corrigidas com ARIMA.

Os modelos de média móvel integrada autorregressiva ou ARIMA (p,d,q) são extensões do modelo AR que usam três componentes para modelar a correlação serial nos dados de séries temporais. O primeiro componente é o período autorregressivo (AR). O modelo AR(p) usa as defasagens p da série temporal na equação. Um modelo AR(p) tem esta forma: $y_t = a_1 y_{t-1} + ... + a_p y_{t-p} + e_t$. O segundo componente é o termo de ordem de integração (d). Cada ordem de integração corresponde à diferenciação da série temporal. I(1) significa a diferenciação dos dados uma vez. I(d) significa a diferenciação dos dados d vezes. O terceiro componente é o período da média móvel (MA). O modelo MA(q) usa as defasagens q dos erros de previsão para aperfeiçoar a previsão. Um modelo MA(q) tem a forma: $y_t = e_t + b_1 e_{t-1} + ... + b_q e_{t-q}$. Por fim, um modelo ARMA(p,q) possui a forma combinada: $y_t = a_1 y_{t-1} + ... + a_p y_{t-p} + e_t + b_1 e_{t-1} + ... + b_q e_{t-q}$.

Procedimento:

- Inicie o Excel e insira os dados ou abra uma planilha com dados históricos para serem previstos (a ilustração mostrada a seguir usa o arquivo de exemplo ***ARIMA de série temporal***)
- Escolha os dados da série temporal e selecione ***Risk Simulator | Previsão | ARIMA***
- Insira os parâmetros P, D e Q relevantes (apenas inteiros positivos) e o número do período de previsão desejado e clique em ***OK***

Nota sobre ARIMA e AUTOARIMA:

- Para ARIMA e AutoARIMA, você pode modelar e prever períodos futuros usando apenas a variável dependente (Y), ou seja, a ***variável de série temporal*** sozinha ou pode adicionar variáveis exógenas (X_1, X_2,…, X_n) como em uma análise de regressão em que é possível ter múltiplas variáveis independentes. Você pode executar quantos períodos de previsão desejar se usar apenas a variável da série temporal (Y). No entanto, se você adicionar variáveis exógenas (X), observe que seus períodos de previsão estarão limitados aos períodos de dados das variáveis exógenas menos o períodos de dados das variáveis de séries temporais. Por exemplo, você poderá prever apenas até 5 períodos se tiver dados históricos de séries temporais de 100 períodos e apenas se tiver variáveis exógenas de 105 períodos (100 períodos históricos para corresponder à variável de série temporal e 5 períodos futuros adicionais de variáveis exógenas independentes para prever a variável dependente da série temporal).

Interpretação dos resultados:

Na interpretação dos resultados do modelo ARIMA, a maioria das especificações é idêntica à análise de regressão multivariada (consulte *Modeling Risk*, 2ª edição do Dr. Johnathan Mun para obter mais detalhes técnicos sobre a interpretação dos modelos de análise de regressão multivariada e ARIMA). Há, entretanto, diversos conjuntos adicionais de resultados específicos à análise ARIMA, como observado na Figura 3.14. A primeira é a adição do critério de informação de Akaike (AIC) e do critério de Schwarz (SC), que são normalmente usados na seleção e na identificação do modelo ARIMA. Ou seja, o AIC e o SC são usados para determinar se um modelo particular com um conjunto de parâmetros específicos p, d e q é um bom ajuste estatístico. O SC impõe uma penalidade maior para coeficientes adicionais do que o AIC, mas em geral o modelo com os menores valores de AIC e SC devem ser escolhidos. Finalmente, um conjunto de resultados adicional chamado de estatísticas de autocorrelação (AC) e autocorrelação parcial (PAC) é fornecido no relatório ARIMA.

Por exemplo, se a autocorrelação AC(1) é diferente de zero, as séries são correlacionadas em série de primeira ordem. Se a AC for extinta mais ou menos geometricamente com uma defasagem crescente, isso implica que a série segue um processo autorregressivo de ordem inferior. Se a AC cair a zero depois de um pequeno número de defasagens, isso implica que a série segue um processo de média móvel de ordem inferior. Em contraste, PAC mede a correlação de valores que estão k períodos distantes, depois de remover a correlação das defasagens intermédias. Se for possível capturar o padrão de autocorrelação por uma autorregressão de ordem menor do que k, então a autocorrelação parcial na defasagem k será próxima a zero. As estatísticas Q de Ljung-Box e seus p-valores na defasagem k também são fornecidas, nas quais a hipótese nula sendo testada é tal que não há autocorrelação até a ordem k. As linhas pontilhadas na “plotagem” das autocorrelações são os limites aproximados dos dois erros padrão. Se a autocorrelação estiver dentro desses limites, não será significativamente diferente de zero no nível de significância de 5%, aproximadamente. Encontrar o modelo ARIMA certo requer prática e experiência. AC, PAC, SC e AIC são ferramentas de diagnóstico, muito úteis para ajudar a identificar a especificação de modelo correta.

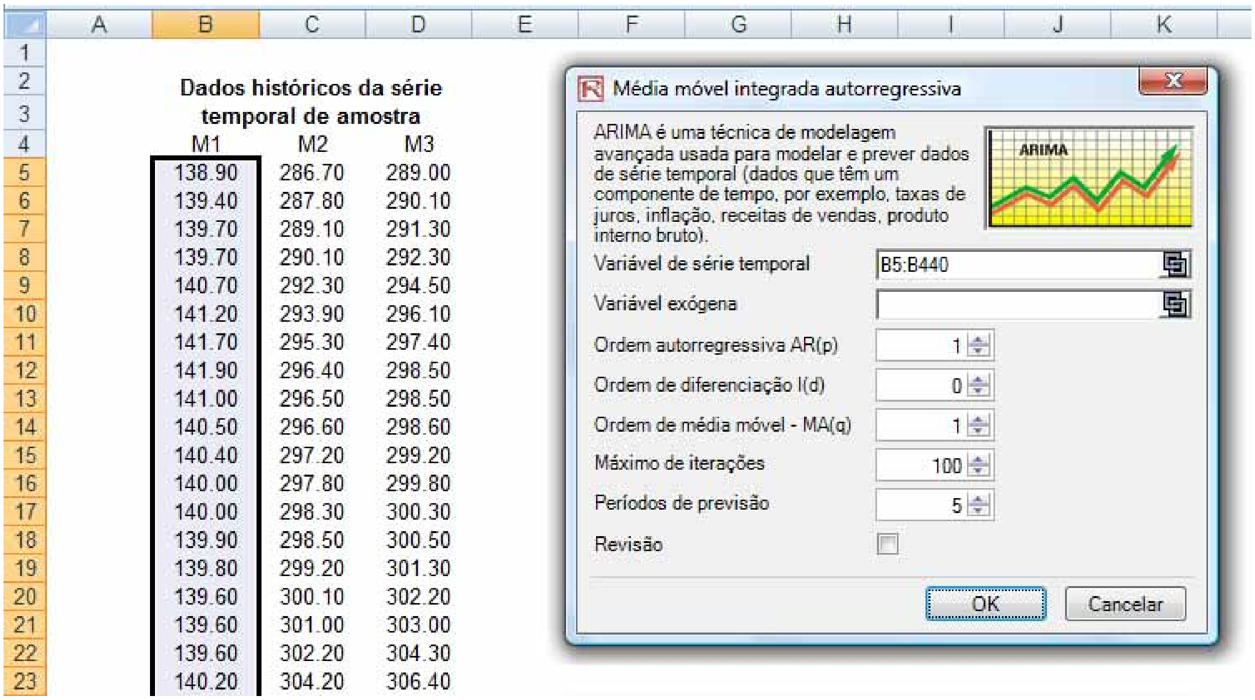


Figura 3.13 Ferramenta de previsão ARIMA Box Jenkins

ARIMA (modelo de média móvel integrada autorregressiva)

Estatísticas de regressão

R2 (coeficiente de determinação)	0.9999	Critério de informação Akaike (AIC)	4.6213
R2 ajustado	0.9999	Critério de Schwarz (SC)	4.6632
R múltiplo (coeficiente de correlação múltipla)	1.0000	Probabilidade de log	-1005.13
Erro padrão das estimativas (SEy)	297.52	Estatística de Durbin-Watson (DW)	1.8588
Número de observações	435	Número de iterações	5

Os modelos de média móvel integrada autorregressiva ou ARIMA (p,d,q) são extensões do modelo AR que usam três componentes para modelar a correlação serial nos dados de séries temporais. O primeiro componente é o período autorregressivo (AR). O modelo AR(p) usa as defasagens p da série temporal na equação. Um modelo AR(p) tem esta forma: y(t)=a(1)*y(t-1)+...+a(p)*y(t-p)+e(t). O segundo componente é o período da ordem (d) de integração. Cada ordem de integração corresponde à diferenciação da série temporal. I(1) significa a diferenciação dos dados uma vez. I(d) significa a diferenciação dos dados d vezes. O terceiro componente é o período da média móvel (MA). O modelo MA(q) usa as defasagens q dos erros de previsão para aperfeiçoar a previsão. Um modelo MA(q) tem esta forma: y(t)=e(t)+b(1)*e(t-1)+...+b(q)*e(t-q). Por último, um modelo ARMA(p,q) tem esta forma combinada: y(t)=a(1)*y(t-1)+...+a(p)*y(t-p)+e(t)+b(1)*e(t-1)+...+b(q)*e(t-q).

R2, ou coeficiente de determinação, indica a variação percentual da variável dependente que pode ser explicada e contabilizada pelas variáveis independentes nessa análise de regressão. No entanto, em uma regressão múltipla, o R2 ajustado considera a existência de outras variáveis independentes ou regressores e ajusta esse valor de R2 para obter uma visão mais precisa da força de explicação da regressão. No entanto, em determinadas circunstâncias ARIMA (por exemplo, com modelos não convergentes), o R2 tende a não ser confiável.

O coeficiente de correlação múltipla (R múltiplo) mede a correlação entre a variável dependente real (Y) e a estimada ou ajustada (Y) com base na equação de regressão. Essa correlação também é a raiz quadrada do coeficiente de determinação (R2).

O erro padrão das estimativas (SEy) descreve a dispersão dos pontos de dados acima e abaixo da linha ou do plano de regressão. Esse valor é usado como parte do cálculo para obter o intervalo de confiança das estimativas depois.

O AIC e o SC são frequentemente usados na seleção de modelos. O SC impõe uma penalidade maior para coeficientes adicionais. Geralmente, o usuário deve selecionar um modelo com o valor mais baixo de AIC e SC.

A estatística de Durbin-Watson mede a correlação serial nos residuais. Em geral, DW menor do que 2 implica uma correlação serial positiva.

Resultados da regressão

	Interceptação	AR(1)	MA(1)
Coeficientes	-0.0626	1.0055	0.4936
Erro padrão	0.3108	0.0006	0.0420
Estatística t	-0.2013	1691.1373	11.7633
P-valor	0.8406	**0.0000**	**0.0000**
5% mais baixos	0.4498	1.0065	0.5628
95% mais altos	-0.5749	1.0046	0.4244

Graus de liberdade		**Teste de hipóteses**	
Graus de liberdade para regressão	2	Estatística t crítica (99% de confiança com df igual a 432)	2.5873
Graus de liberdade para residual	432	Estatística t crítica (95% de confiança com df igual a 432)	1.9655
Total de graus de liberdade	434	Estatística t crítica (90% de confiança com df igual a 432)	1.6484

Os coeficientes fornecem a interceptação e a inclinação estimadas da regressão. Por exemplo, os coeficientes são estimativas dos verdadeiros valores da população b nesta equação de regressão: Y = b0 + b1X1 + b2X2 + ... + bnXn. O erro padrão mede a precisão dos coeficientes previstos, e as estatísticas t são os índices de cada coeficiente previsto para seu erro padrão.

A estatística t é usada em testes de hipóteses, nos quais é possível definir a hipótese nula (Ho) de tal maneira que a média real do coeficiente = 0, e a hipótese alternativa (Ha) de tal maneira que a média real do coeficiente seja diferente de 0. Um teste t é executado e a estatística t calculada é comparada aos valores críticos nos graus de liberdade relevantes para o residual. O teste t é muito importante, pois calcula se cada um dos coeficientes é estatisticamente significativo na presença de outros regressores. Isso significa que o teste t verifica estatisticamente se um regressor ou uma variável independente deve permanecer na regressão ou ser descartada.

O coeficiente é estatisticamente significativo se a sua estatística t calculada excede a estatística t crítica nos graus de liberdade relevantes (df). Os três principais níveis de confiança usados para testar a significância são 90%, 95% e 99%. Se a estatística t de um coeficiente exceder o nível crítico, ela será considerada estatisticamente relevante. Como alternativa, o p-valor calcula a probabilidade de ocorrência de cada estatística t, o que significa que quanto menor o p-valor, mais significativo é o coeficiente. Os níveis de significância do p-valor costumam ser 0,01, 0,05 e 0,10, o que corresponde aos níveis de confiança 99%, 95% e 90%.

Os coeficientes cujos p-valores são realçados em azul indicam que são estatisticamente significativos no nível de confiança de 90% ou no nível alfa de 0,10, enquanto aqueles cujos p-valores são realçados em vermelho indicam que não são estatisticamente relevantes em qualquer outro nível alfa.

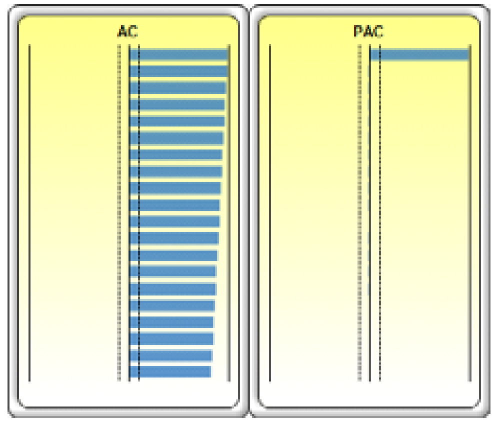
Análise de variância

	Somas dos quadrados	Média dos quadrados	Estatística F	P-valor	Teste de hipóteses	
Regressão	38415447.53	19207723.76	3171851.1	**0.0000**	Estatística F crítica (99% de confiança com df igual a 2 e 4	4.6546
Residual	2616.05	6.06			Estatística F crítica (95% de confiança com df igual a 2 e 4	3.0166
Total	38418063.58				Estatística F crítica (90% de confiança com df igual a 2 e 4	2.3149

A tabela de análise de variância (ANOVA) fornece um teste F da significância estatística geral do modelo de regressão. Em vez de analisar regressores individuais como no teste t, o teste F analisa as propriedades estatísticas de todos os coeficientes estimados. A estatística F é calculada como a proporção da média dos quadrados da regressão em relação à média dos quadrados de residuais. O numerador mede quanto da regressão é explicado, enquanto o denominador mede quanto não é explicado. Assim, quanto maior a estatística F, mais significante é o modelo. O p-valor correspondente é calculado para testar a hipótese nula (Ho), segundo a qual todos os coeficientes são simultaneamente iguais a zero, em contraste com a hipótese alternativa (Ha) de que eles são simultaneamente diferentes de zero, indicando um modelo de regressão geral significante. Se o p-valor for menor do que as significâncias alfa 0,01, 0,05 ou 0,10, a regressão será significante. A mesma abordagem pode ser aplicada à estatística F comparando a estatística F calculada aos valores F críticos em vários níveis de significância.

Autocorrelação

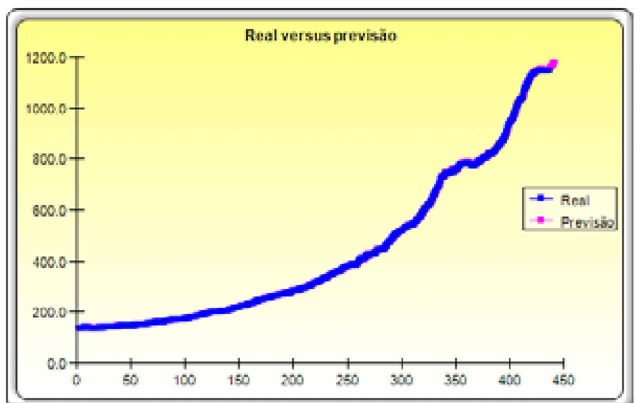
Defasagem de tempo	AC	PAC	Limite inferior	.imite superior	Estatística Q	Prob
1	0.9921	0.9921	(0.0958)	0.0958	431.1216	-
2	0.9841	(0.0105)	(0.0958)	0.0958	856.3037	-
3	0.9760	(0.0109)	(0.0958)	0.0958	1,275.4818	-
4	0.9678	(0.0142)	(0.0958)	0.0958	1,688.5499	-
5	0.9594	(0.0098)	(0.0958)	0.0958	2,095.4625	-
6	0.9509	(0.0113)	(0.0958)	0.0958	2,496.1572	-
7	0.9423	(0.0124)	(0.0958)	0.0958	2,890.5594	-
8	0.9336	(0.0147)	(0.0958)	0.0958	3,278.5669	-
9	0.9247	(0.0121)	(0.0958)	0.0958	3,660.1152	-
10	0.9156	(0.0139)	(0.0958)	0.0958	4,035.1192	-
11	0.9066	(0.0049)	(0.0958)	0.0958	4,403.6117	-
12	0.8975	(0.0068)	(0.0958)	0.0958	4,765.6032	-
13	0.8883	(0.0097)	(0.0958)	0.0958	5,121.0697	-
14	0.8791	(0.0087)	(0.0958)	0.0958	5,470.0032	-
15	0.8698	(0.0064)	(0.0958)	0.0958	5,812.4256	-
16	0.8605	(0.0056)	(0.0958)	0.0958	6,148.3694	-
17	0.8512	(0.0062)	(0.0958)	0.0958	6,477.8620	-
18	0.8419	(0.0038)	(0.0958)	0.0958	6,800.9622	-
19	0.8326	(0.0003)	(0.0958)	0.0958	7,117.7709	-
20	0.8235	0.0002	(0.0958)	0.0958	7,428.3952	-



Se a autocorrelação AC(1) é diferente de zero, as séries são correlacionadas em série de primeira ordem. Se a AC(k) for extinta mais ou menos geometricamente com uma defasagem crescente, isso implica que a série segue um processo autorregressivo de ordem inferior. Se a AC(k) cair a zero depois de um pequeno número de defasagens, isso implica que a série segue um processo de média móvel de ordem inferior. A correlação parcial PAC(k) mede a correlação dos valores que estão separados por k períodos, depois de remover a correlação das defasagens intermédias. Se for possível capturar o padrão de autocorrelação por uma autorregressão de ordem menor do que k, então a autocorrelação parcial na defasagem k será próxima a zero. As estatísticas Q Ljung-Box e seus p-valores na defasagem k têm a hipótese nula de que não há autocorrelação até a ordem k. As linhas pontilhadas nos gráficos da autocorrelação são os dois limites de erro padrão aproximados. Se a autocorrelação estiver dentro desses limites, não será significativamente diferente de zero no nível de significância de 5% (aproximadamente).

Previsão

Período	Real (Y)	Previsão (F)	Erro (E)
2	139.4000	139.6056	(0.2056)
3	139.7000	140.0069	(0.3069)
4	139.7000	140.2586	(0.5586)
5	140.7000	140.1343	0.5657
6	141.2000	141.6948	(0.4948)
7	141.7000	141.6741	0.0259
8	141.9000	142.4339	(0.5339)
9	141.0000	142.3587	(1.3587)
10	140.5000	141.0466	(0.5466)
11	140.4000	140.9447	(0.5447)
12	140.0000	140.8451	(0.8451)
13	140.0000	140.2946	(0.2946)
14	139.9000	140.5663	(0.6663)
15	139.8000	140.2823	(0.4823)
16	139.6000	140.2726	(0.6726)
17	139.6000	139.9775	(0.3775)
18	139.6000	140.1232	(0.5231)
19	140.2000	140.0513	0.1487
20	141.3000	140.9862	0.3138
21	141.2000	142.1738	(0.9738)
22	140.9000	141.4377	(0.5377)
23	140.9000	141.3513	(0.4513)
24	140.7000	141.3939	(0.6939)
25	141.1000	141.0731	0.0270
26	141.6000	141.8311	(0.2311)
27	141.9000	142.2065	(0.3065)
28	142.1000	142.4709	(0.3709)
29	142.7000	142.6402	0.0598
30	142.9000	143.4561	(0.5561)
31	142.9000	143.3532	(0.4532)
32	143.5000	143.4040	0.0960
33	143.8000	144.2784	(0.4784)
34	144.1000	144.2966	(0.1966)
35	144.8000	144.7374	0.0626
36	145.2000	145.5692	(0.3692)
37	145.2000	145.7582	(0.5582)
38	145.7000	145.6649	0.0351
39	146.0000	146.4605	(0.4605)
40	146.4000	146.5176	(0.1176)
41	146.8000	147.0891	(0.2891)



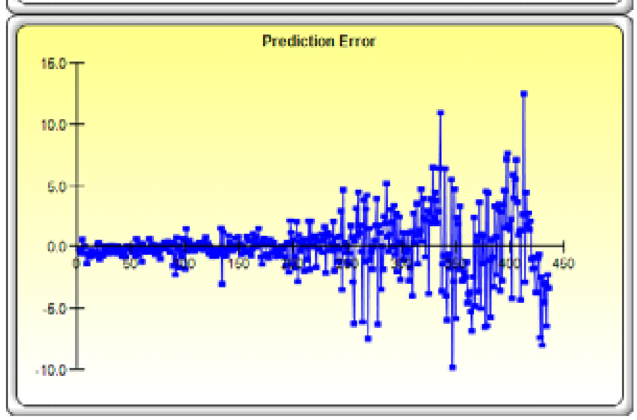


Figura 3.13 Relatório de previsão ARIMA Box Jenkins

AUTOARIMA (série temporal avançada Box-Jenkins ARIMA)

Teoria:

Esta ferramenta fornece análises idênticas ao módulo ARIMA, mas a diferença é que o módulo AutoARIMA automatiza algumas das modelagens ARIMA tradicionais por meio do teste automático de várias permutações de especificações do modelo e retorna o modelo de melhor ajuste. A execução de AutoARIMA é semelhante à execução das previsões ARIMA normais. A diferença é que as entradas P, D e Q não são obrigatórias e diferentes combinações desses valores são automaticamente executados e comparados.

Procedimento:

- Inicie o Excel e insira os dados ou abra uma planilha com dados históricos a serem previstos (a Figura 3.14 mostra o arquivo de exemplo ***Modelos avançados de previsão*** disponível no menu ***Exemplos*** do Risk Simulator)
- Na planilha ***AutoARIMA***, selecione ***Risk Simulator | Previsão | AUTOARIMA.*** Você também pode acessar o método usando a faixa de ícones Previsão ou clicando com o botão direito do mouse em qualquer lugar no modelo e selecionando o menu de atalho de previsão.
- Clique no ícone de vínculo; vá para os dados de séries temporais; insira o número de períodos de previsão desejados e clique em ***OK***

Nota sobre ARIMA e AUTOARIMA:

- Para ARIMA e AutoARIMA, você pode modelar e prever períodos futuros usando apenas a variável dependente (*Y*), ou seja, a ***variável de série temporal*** sozinha ou pode adicionar variáveis exógenas (X_1, X_2,…, X_n) como em uma análise de regressão em que é possível ter múltiplas variáveis independentes. Você pode executar quantos períodos de previsão desejar se usar apenas a variável da série temporal (*Y*). No entanto, se você adicionar variáveis exógenas (*X*), observe que seus períodos de previsão estarão limitados aos períodos de dados das variáveis exógenas menos os períodos de dados das variáveis de séries temporais. Por exemplo, você poderá prever apenas até 5 períodos se tiver dados históricos de séries temporais de 100 períodos e apenas se tiver variáveis exógenas de 105 períodos (100 períodos históricos para corresponder à variável de série temporal e 5 períodos futuros adicionais de variáveis exógenas independentes para prever a variável dependente da série temporal).

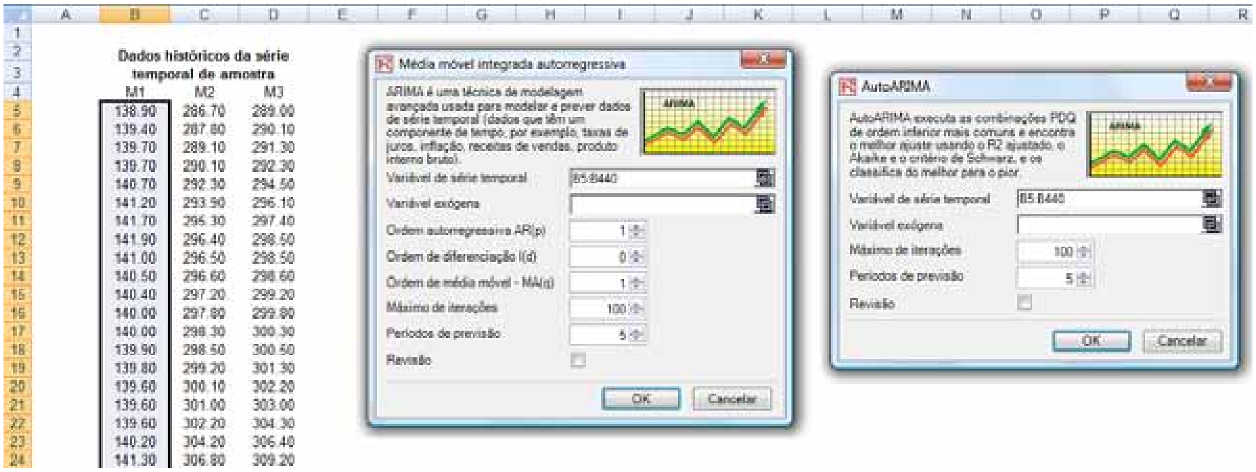


Figura 3.14 Módulo AUTOARIMA

Econometria básica

Teoria:

A econometria se refere a um ramo de técnicas de previsão, modelagem e análise de negócios para modelar o comportamento ou prever certas variáveis de negócios ou economia. A execução de modelos de Econometria básica é semelhante à execução da análise de regressão normal, exceto que as variáveis independentes e dependentes podem ser modificadas antes da execução de uma regressão. O relatório gerado é o mesmo mostrado anteriormente na seção Regressão múltipla e a interpretação é idêntica àquelas descritas anteriormente.

Procedimento:

- Inicie o Excel e insira os dados ou abra uma planilha com dados históricos a serem previstos (a Figura 3.15 mostra o arquivo de exemplo ***Modelos avançados de previsão*** disponível no menu ***Exemplos*** do Risk Simulator)
- Selecione os dados na planilha ***Econometria básica*** e selecione ***Risk Simulator | Previsão | Econometria básica***
- Insira as variáveis dependentes e independentes desejadas (consulte a Figura 3.15 para obter exemplos) e clique em ***OK*** para executar o modelo e o relatório, ou clique em ***Mostrar resultados*** para ver os resultados antes de gerar o relatório, caso precise fazer alterações no modelo

Conjunto de dados econométricos básicos

Y	X1	X2	X3	X4	X5
521	18308	185	4.041	79.6	7.2
367	1148	600	0.55	1	8.5
443	18068	372	3.665	32.3	5.7
365	7729	142	2.351	45.1	7.3
614	100484	432	29.76	190.8	7.5
385	16728	290	3.294	31.8	5
286	14630	346	3.287	678.4	6.7
397	4008	328	0.666	340.8	6.2
764	38927	354	12.938	239.6	7.3
427	22322	266	6.478	111.9	5
153	3711	320	1.108	172.5	2.8
231	3136	197	1.007	12.2	6.1
524	50508	266	11.431	205.6	7.1
328	28886	173	5.544	154.6	5.9
240	16996	190	2.777	49.7	4.6
286	13035	239	2.478	30.3	4.4
285	12973	190	3.685	92.8	7.4
569	16309	241	4.22	96.9	7.1
96	5227	189	1.228	39.8	7.5
498	19235	358	4.781	489.2	5.9
481	44487	315	6.016	767.6	9
468	44213	303	9.295	163.6	9.2
177	23619	228	4.375	55	5.1
198	9106	134	2.573	54.9	8.6
458	24917	189	5.117	74.3	6.6
108	3872	196	0.799	5.5	6.9
246	8945	183	1.578	20.5	2.7
291	2373	417	1.202	10.9	5.5
68	7128	233	1.109	123.7	7.2
311	23624	349	7.73	1042	6.6
606	5242	284	1.515	12.5	6.9
512	92629	499	17.99	381	7.2
426	28795	231	6.629	136.1	5.8
47	4487	143	0.639	9.3	4.1

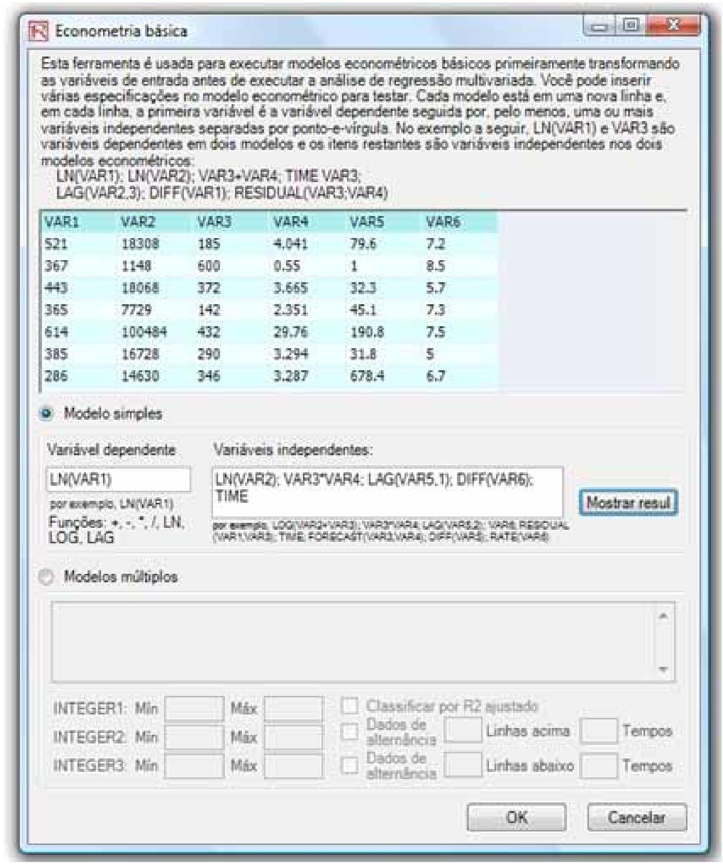


Figura 3.15 Módulo Econometria básica

Nota:

- Consulte o Capítulo 9 para obter mais detalhes sobre a interpretação de resultados de regressão e também os resultados da análise de econometria básica.
- Para executar um modelo econométrico, selecione os dados (B5:G55), incluindo os cabeçalhos, e clique em ***Risk Simulator* | *Previsão* | *Econometria básica***. Você pode digitar as variáveis e suas modificações para as variáveis dependentes e independentes (Figura 8.15). Observe que apenas uma variável é permitida como variável dependente (*Y*), enquanto pode-se usar múltiplas variáveis na seção Variáveis independentes (*X*), separadas por um ponto-e-vírgula";" e funções matemáticas básicas podem ser usadas (por exemplo, LN, LOG, LAG, +, -, /, *, TEMPO, RESIDUAL, DIFERENÇA). Clique em ***Mostrar resultados*** para visualizar o modelo calculado e clique em ***OK*** para gerar o relatório do modelo econométrico.
- Você também pode gerar automaticamente vários modelos inserindo um modelo de exemplo e usando a variável predefinida ***"INTEGER(N)"***, bem como ***deslocar dados*** para cima ou para baixo nas linhas. Por exemplo, se você usar a variável *LAG(VAR1, INTEGER1)* e definir *INTEGER1* como entre *MIN* = 1 e *MAX* = 3, então os três modelos a seguir serão executados: *LAG(VAR1,1)*, então *LAG(VAR1,2)* e por último *LAG(VAR1,3)*. Além disso, às vezes pode ser necessário testar se os dados da série temporal têm alterações estruturais ou se o comportamento

do modelo é consistente ao longo do tempo, alterando os dados e executando o mesmo modelo. Por exemplo, se você tem 100 meses de dados listados cronologicamente, pode deslocá-los 3 meses por vez por 10 vezes (isto é, o modelo será executado nos meses 1 a 100, 4 a 100, 7 a 100 etc.). Usando a seção ***Modelos múltiplos*** em Econometria básica, você pode executar centenas de modelos simplesmente inserindo uma equação de modelo único se usar estas variáveis inteiras predefinidas e métodos de deslocamento.

Previsões de curvas JS

Teoria:

A curva J, ou curva de crescimento exponencial, é aquela na qual o crescimento do próximo período depende do nível do período atual e o aumento é exponencial. Isso significa que ao longo do tempo, os valores aumentarão significativamente de um período para o outro. Normalmente, esse modelo é usado para prever o crescimento biológico e reações químicas ao longo do tempo.

Procedimento:

- Inicie o Excel e selecione ***Risk Simulator | Previsão | Curvas JS***.
- Selecione o tipo de curva J ou S, insira as suposições de entrada obrigatórias (consulte as Figuras 3.16 e 3.17 para obter exemplos) e clique em ***OK*** para executar o modelo e o relatório.

A curva S ou curva de crescimento logístico começa como uma curva J, com taxas de crescimento exponenciais. Com o tempo, o ambiente torna-se saturado (por exemplo, saturação do mercado, competição, superlotação), o crescimento diminui e o valor da previsão por fim termina em um nível máximo de saturação. Esse modelo é muito usado para prever participação de mercado ou o crescimento das vendas de um novo produto, desde introdução no mercado até a maturidade e o declínio, a dinâmica de populações, o crescimento de culturas de bactérias e outras variáveis de ocorrência natural. A Figura 3.17 ilustra um exemplo de curva S.

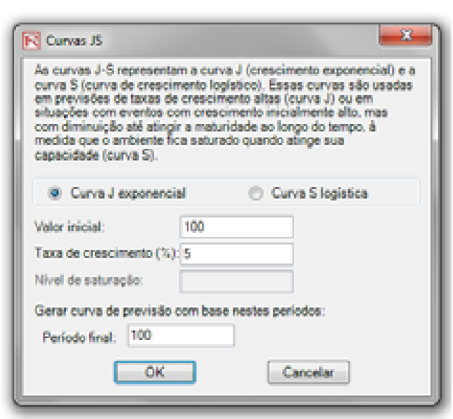


Figura 3.16 Previsão de curva J

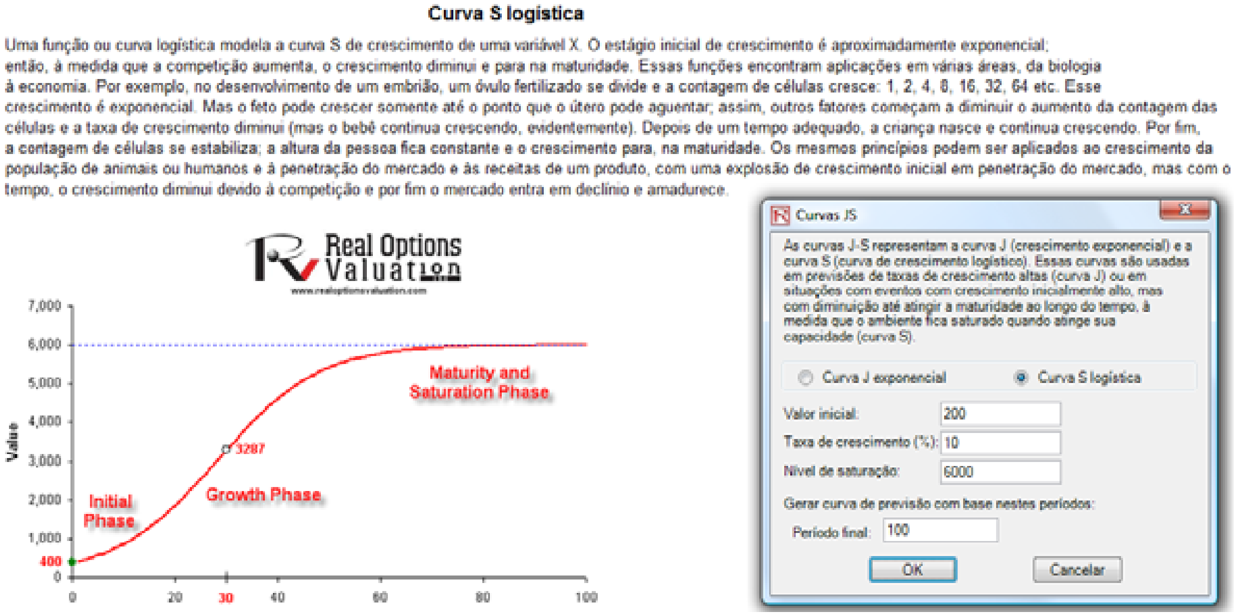


Figura 3.17 Previsão de curva S

Previsões de volatilidade GARCH

Teoria:

O modelo de heteroscedasticidade condicional autorregressiva (GARCH) é usado para modelar níveis de volatilidade históricos e prever a volatilidade futura de um instrumento negociável, por exemplo, preços de ações, preços de commodities, preços do petróleo etc. O conjunto de dados deve ser uma série temporal de níveis de preços brutos. Primeiro, GARCH converte os preços em retornos relativos e executa uma otimização interna para ajustar os dados históricos a uma estrutura a termo de volatilidade de reversão à média, supondo que a volatilidade é natureza heteroscedástica (se altera com o tempo de acordo com algumas características econométricas). As especificidades teóricas de um modelo GARCH não estão dentro do escopo deste manual. Para obter mais detalhes sobre modelos GARCH, consulte "*Advanced Analytical Models*", do Dr. Johnathan Mun (Wiley 2008).

Procedimento:

- Inicie o Excel, abra o arquivo de exemplo ***Modelos avançados de previsão***, vá para a planilha ***GARCH*** e selecione ***Risk Simulator | Previsão | GARCH***.
- Clique no ícone de vínculo, selecione *Localização dos dados*, insira as suposições de entrada obrigatórias (consulte a Figura 3.18) e clique em ***OK*** para executar o modelo e o relatório.

Nota: A situação de previsão de volatilidade típica requer P = 1, Q = 1, Periodicidade = número de períodos por ano (12 para dados mensais, 52 para dados semanais, 252 ou 365 para dados diários), Base = mínimo de 1 até o valor de periodicidade e Períodos de previsão = número de previsões de volatilidade anualizadas que você deseja obter. Há vários modelos GARCH disponíveis no Risk Simulator, incluindo

EGARCH, EGARCH-T, GARCH-M, GJR-GARCH, GJR-GARCH-T, IGARCH e T-GARCH. Consulte Modeling Risk, segunda edição (Wiley 2010), sobre modelagem GARCH, para obter mais detalhes sobre a utilidade de cada especificação.

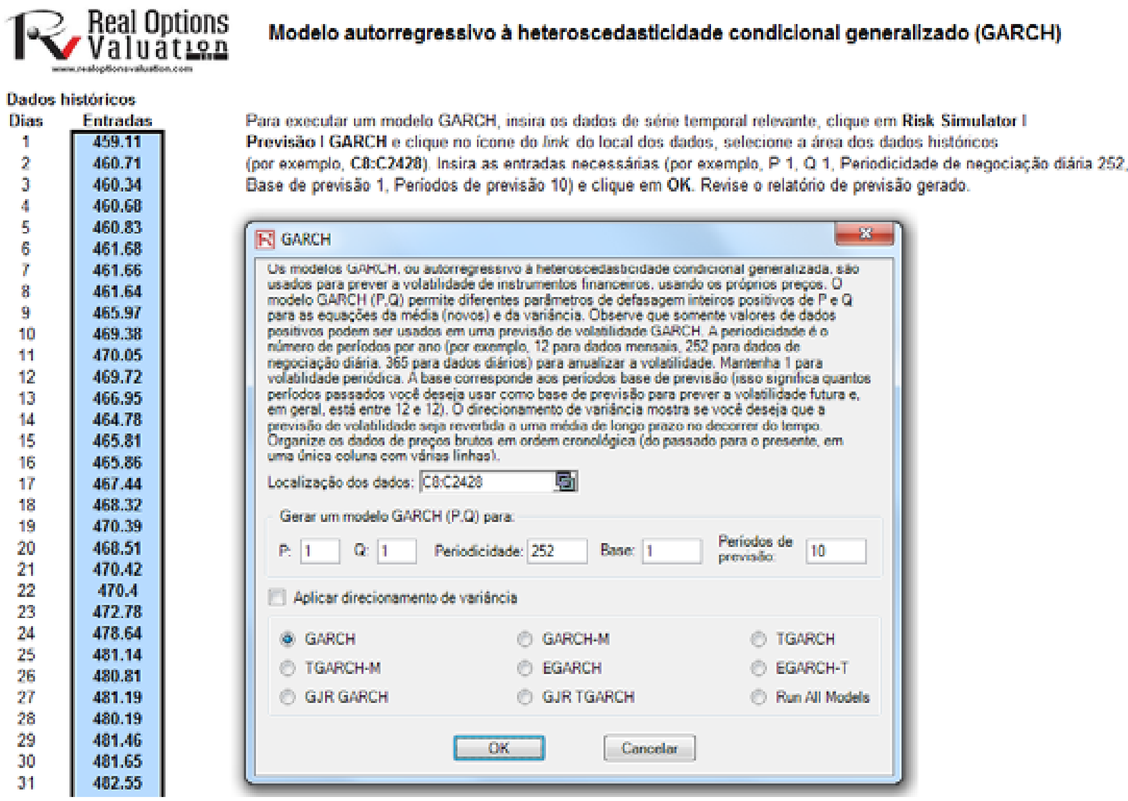


Figura 3.18 Previsão de volatilidade GARCH

	$z_t \sim \text{Normal}$	$z_t \sim \text{T}$
GARCH-M	$y_t = c + \lambda\sigma_t^2 + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$	$y_t = c + \lambda\sigma_t^2 + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$
GARCH-M	$y_t = c + \lambda\sigma_t + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$	$y_t = c + \lambda\sigma_t + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$

GARCH-M	$y_t = c + \lambda \ln(\sigma_t^2) + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$	$y_t = c + \lambda \ln(\sigma_t^2) + \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$
GARCH	$y_t = x_t\gamma + \varepsilon_t$ $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$	$y_t = \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$
EGARCH	$y_t = \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\ln\left(\sigma_t^2\right) = \omega + \beta \cdot \ln\left(\sigma_{t-1}^2\right) +$ $\alpha\left[\left\|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right\| - E(\|\varepsilon_t\|)\right] + r\frac{\varepsilon_{t-1}}{\sigma_{t-1}}$ $E(\|\varepsilon_t\|) = \sqrt{\frac{2}{\pi}}$	$y_t = \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\ln\left(\sigma_t^2\right) = \omega + \beta \cdot \ln\left(\sigma_{t-1}^2\right) +$ $\alpha\left[\left\|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right\| - E(\|\varepsilon_t\|)\right] + r\frac{\varepsilon_{t-1}}{\sigma_{t-1}}$ $E(\|\varepsilon_t\|) = \frac{2\sqrt{\nu-2}\,\Gamma((\nu+1)/2)}{(\nu-1)\Gamma(\nu/2)\sqrt{\pi}}$
GJR-GARCH	$y_t = \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 +$ $r\varepsilon_{t-1}^2 d_{t-1} + \beta\sigma_{t-1}^2$ $d_{t-1} = \begin{cases} 1 & \text{if } \varepsilon_{t-1} < 0 \\ 0 & \text{otherwise} \end{cases}$	$y_t = \varepsilon_t$ $\varepsilon_t = \sigma_t z_t$ $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 +$ $r\varepsilon_{t-1}^2 d_{t-1} + \beta\sigma_{t-1}^2$ $d_{t-1} = \begin{cases} 1 & \text{if } \varepsilon_{t-1} < 0 \\ 0 & \text{otherwise} \end{cases}$

Cadeias de Markov

Teoria:

Uma cadeia de Markov existe quando a probabilidade de um estado futuro depende de um estado anterior e quando unidos formam uma cadeia se reverte a um nível de estado estável no longo prazo. Essa abordagem é usada para prever a participação de mercado de dois concorrentes. Os valores de entrada obrigatórios são a probabilidade inicial de um cliente na primeira loja (o primeiro estado) retornar à mesma loja no próximo período, comparada com a probabilidade de mudar para uma loja concorrente no próximo estado.

Procedimento:

- Inicie o Excel e selecione ***Risk Simulator | Previsão | Cadeia de Markov***.
- Insira as suposições de entrada obrigatórias (consulte a Figura 3.19 para obter um exemplo) e clique em ***OK*** para executar o modelo e o relatório.

Nota:

Configure as duas probabilidades como 10% e execute novamente a cadeia de Markov para ver os efeitos da mudança de comportamentos claramente no gráfico resultante.

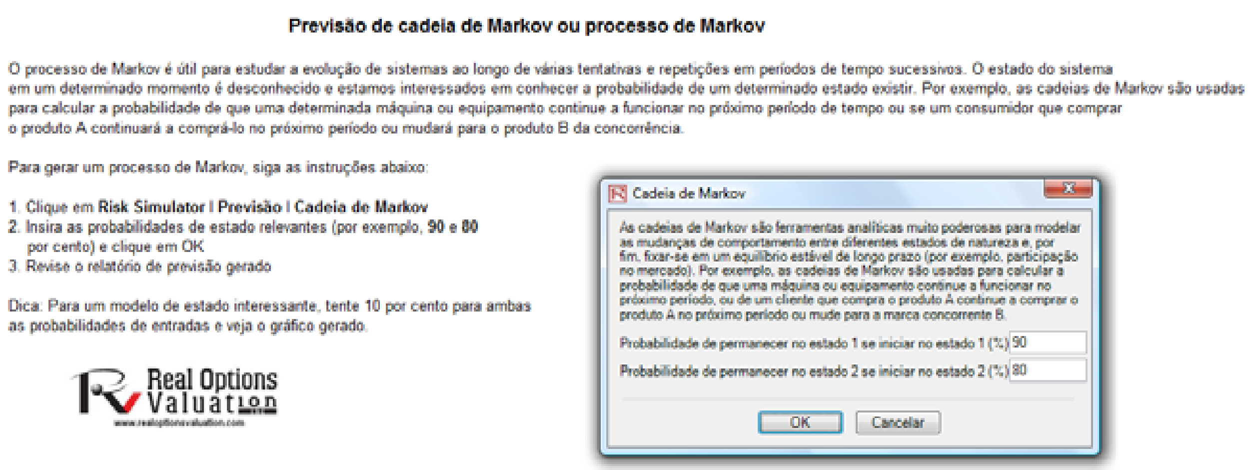


Figura 3.19 Cadeias de Markov (troca de regimes)

Modelos de Máxima Verossimilhança (MLE) em logit, probit e tobit

Teoria:

As variáveis dependentes limitadas descrevem a situação na qual a variável dependente contém dados limitados em escopo e faixa, como uma resposta binária (0 ou 1), dados truncados, ordenados ou censurados. Por exemplo, conhecido um conjunto de variáveis independentes (por exemplo, idade, renda, nível de educação de titulares de cartão de crédito ou mutuários da casa própria), podemos modelar a probabilidade de inadimplência usando o MLE (estimador de máxima verossimilhança). A resposta ou variável dependente Y é binária, ou seja, só pode ter dois resultados possíveis, que serão indicados como 1 e 0 (por exemplo, Y pode representar a presença/ausência de uma determinada condição, adimplência/inadimplência em empréstimos anteriores, sucesso/falha de alguns dispositivos, respostas sim/não em uma pesquisa etc.). Há ainda um vetor de regressores X da variável independente, que supostamente influenciam o resultado Y. Uma abordagem típica de regressão dos quadrados mínimos é inválida porque os erros de regressão são heteroscedásticos e anormais e as estimativas de probabilidade esperadas resultantes retornarão valores absurdos maiores do que 1 ou menores do que 0. A análise MLE trata esses problemas usando uma rotina de otimização iterativa para maximizar uma função de probabilidade de log quando as variáveis dependentes forem limitadas.

Uma regressão logit ou logística é usada para prever a probabilidade de ocorrência de um evento ao ajustar os dados em uma curva logística. É um modelo linear generalizado usado para regressão binomial e, como muitas formas de análise de regressão, usa diversas variáveis de previsão que podem ser numéricas ou categóricas. O MLE aplicado em uma análise logística multivariada binária é usado para modelar variáveis dependentes que determinam a probabilidade esperada de êxito de pertencer a um determinado grupo. Os coeficientes estimados para o modelo logit são as proporções de chances logarítmicas e não podem ser interpretados diretamente como probabilidades. É necessário antes um cálculo rápido e a abordagem é simples.

Especificamente, o modelo logit é definido como Y estimado = LN[Pi/(1–Pi)] ou, alternativamente, Pi = EXP(Y estimado)/(1+EXP(Y estimado)), e os coeficientes βi são proporções de chances de log, sendo assim, ao usar o antilog ou EXP(βi), obtém-se as proporções de chances de Pi/(1–Pi). Isso significa que com um aumento de uma unidade em βi, a proporção de chances de log aumentam de acordo com esse valor. Finalmente, a taxa de alteração na probabilidade dP/dX = βiPi(1–Pi). O erro padrão mede a precisão dos coeficientes previstos, enquanto as estatísticas t são as razões entre cada coeficiente previsto e seu erro padrão e são usadas no teste de hipóteses de regressão típica da significância de cada parâmetro estimado. Para estimar a probabilidade de êxito de pertencer a um determinado grupo (por exemplo, prever se um fumante desenvolverá doenças pulmonares conhecida a quantidade de cigarros fumados por ano), simplesmente calcule o valor de Y estimado usando os coeficientes de MLE. Por exemplo, se o modelo é Y = 1,1 + 0,005 (cigarros), então, uma pessoa que fume 100 maços por ano tem um Y estimado de 1,1 + 0,005(100) = 1,6. Em seguida, calcule o antilog inverso da proporção de chances fazendo:

EXP(Y estimado)/[1 + EXP(Y estimado)] = EXP(1,6)/(1+ EXP(1,6)) = 0,8320. Assim, essa pessoa tem 83,20% de chance de desenvolver complicações pulmonares durante sua vida.

Um modelo probit (também conhecido como um modelo normit) é uma especificação alternativa comum para um modelo de resposta binária que emprega uma função probit estimada usando o estimador de máxima verossimilhança e a abordagem é chamada de regressão probit. Os modelos de regressão logística e probit tendem a produzir previsões muito semelhantes, nas quais as estimativas dos parâmetros em uma regressão logística tendem a ser de 1,6 a 1,8 vezes maiores do que em um modelo probit correspondente. A escolha de usar probit ou logit é inteiramente discricionária e a principal distinção é que a distribuição logística possui uma curtose maior (caudas mais largas) para explicar valores extremos. Por exemplo, suponha que a compra de um imóvel seja a decisão a ser modelada e essa variável de resposta seja binária (compra ou não do imóvel) e dependa de uma série de variáveis Xi independentes como renda, idade etc., de tal forma que Ii = β0 + β1X1 +...+ βnXn, em que quanto maior for o valor de Ii, maior será a a probabilidade de compra do imóvel. Para cada família há um limite I* crítico, que se excedido, significa que o imóvel será comprado; caso contrário, ele não será comprado, e a probabilidade do resultado (P) é considerada como distribuída normalmente, para que Pi = CDF(I), usando uma CDF (função de distribuição acumulada) normal padrão. Portanto, use os coeficientes estimados exatamente como os de um modelo de regressão e, usando o valor Y estimado, aplique uma distribuição normal padrão (você pode usar a função DIST.NORMP do Excel ou a ferramenta Análise de distribuição do Risk Simulator, selecionando a distribuição normal e definindo a média como 0 e o desvio padrão como 1). Finalmente, para obter uma unidade de medida de probabilidade ou probit, defina Ii + 5 (isso ocorre porque toda vez que a probabilidade Pi < 0,5, o Ii estimado é negativo, devido ao fato de a distribuição normal ser simétrica próxima de uma média zero).

O modelo tobit (tobit censurado) é um método de modelagem econométrica e biométrica usado para descrever a relação entre uma variável dependente não negativa Yi e uma ou mais variáveis independentes Xi. Um modelo tobit é um modelo econométrico em que a variável dependente é censurada, porque os valores abaixo de zero não são observados. O modelo tobit supõe que há uma variável não observável latente Y*. Essa variável é linearmente dependente das variáveis Xi por meio de um vetor de coeficientes βi que determina as interrelações. Além disso, há uma condição de erro Ui normalmente distribuída para capturar influências aleatórias nessa relação. A variável observável Yi é definida como sendo igual às variáveis latentes sempre que essas são acima de zero. Caso contrário, supõe-se Yi igual a zero. Ou seja, Yi = Y* se Y* > 0 e Yi = 0 se Y* = 0. Se o parâmetro da relação βi for estimado usando uma regressão de quadrados mínimos normal de Yi observada em Xi, os estimadores da regressão resultante serão inconsistentes e apresentarão coeficientes de inclinação polarizada em ordem decrescente e uma interceptação polarizada em ordem crescente. Apenas o MLE seria consistente para um modelo tobit. Nesse modelo, há uma estatística auxiliar chamada sigma, que é equivalente ao erro padrão da estimativa em uma regressão dos quadrados mínimos típica padrão e os coeficientes estimados são usados da mesma maneira que uma análise de regressão.

Procedimento:

- Inicie o Excel, abra o arquivo de exemplo ***Modelos avançados de previsão***, vá para a planilha ***MLE***, selecione o conjunto de dados incluindo os cabeçalhos e clique em ***Risk Simulator | Previsão | Máxima verossimilhança***.
- Na lista suspensa, selecione a variável dependente (consulte a Figura 3.20) e clique em ***OK*** para executar o modelo e o relatório.

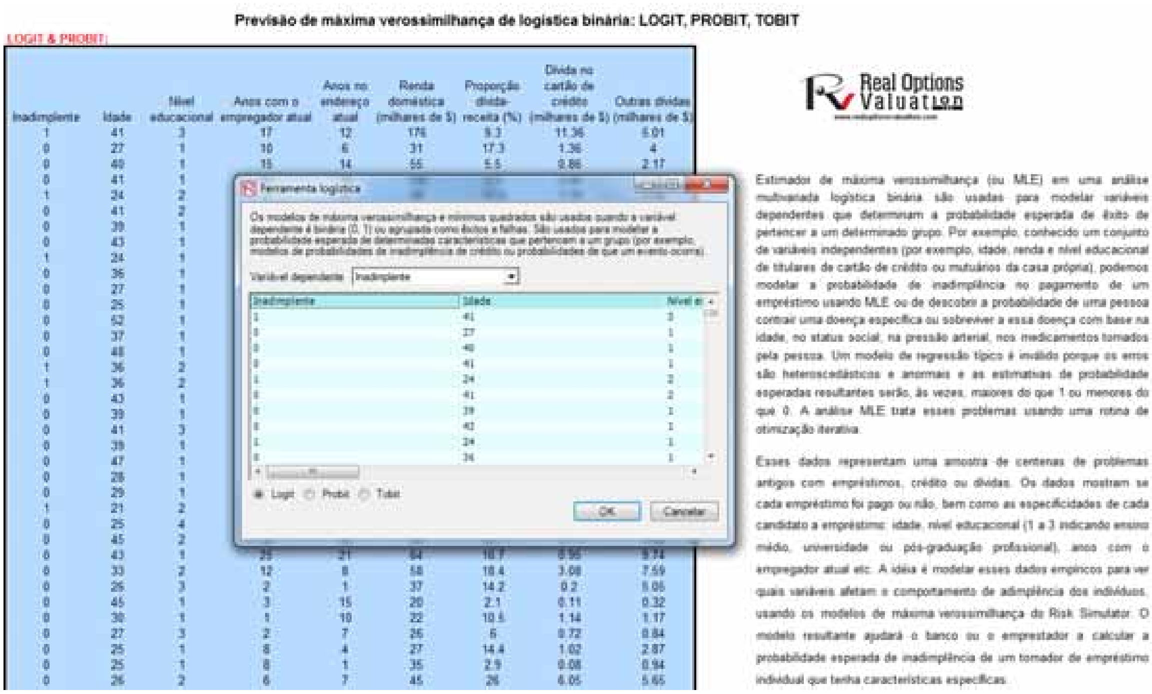


Figura 3.20 Módulo de máxima verossimilhança

Spline (interpolação e extrapolação de spline cúbico)

Teoria:

Às vezes há valores ausentes em um conjunto de dados de uma série temporal. Por exemplo, pode-se ter as taxas de juros dos anos 1 a 3, seguidas pelas dos anos 5 a 8 e 10. As curvas spline podem ser usadas para interpolar os valores de taxa de juros dos anos ausentes com base nos dados existentes. As curvas *spline* também podem ser usadas para prever ou extrapolar valores de períodos de tempo futuros além do período de tempo dos dados disponíveis. Os dados podem ser lineares ou não lineares. A Figura 3.21 ilustra como um *spline* cúbico é executado e a Figura 3.22 mostra o relatório de previsão resultante deste módulo. Os valores de *X* conhecidos representam os valores no eixo X de um gráfico (no nosso exemplo, os anos das taxas de juros conhecidas e, geralmente, o eixo *x* apresenta os valores previamente conhecidos, como o tempo ou o número de anos) e os valores conhecidos de *Y* representam os valores no eixo *y* (no nosso caso, as taxas de juros conhecidas). A variável do eixo *y* geralmente é aquela na qual você deseja interpolar valores ausentes ou extrapolar os valores para o futuro.

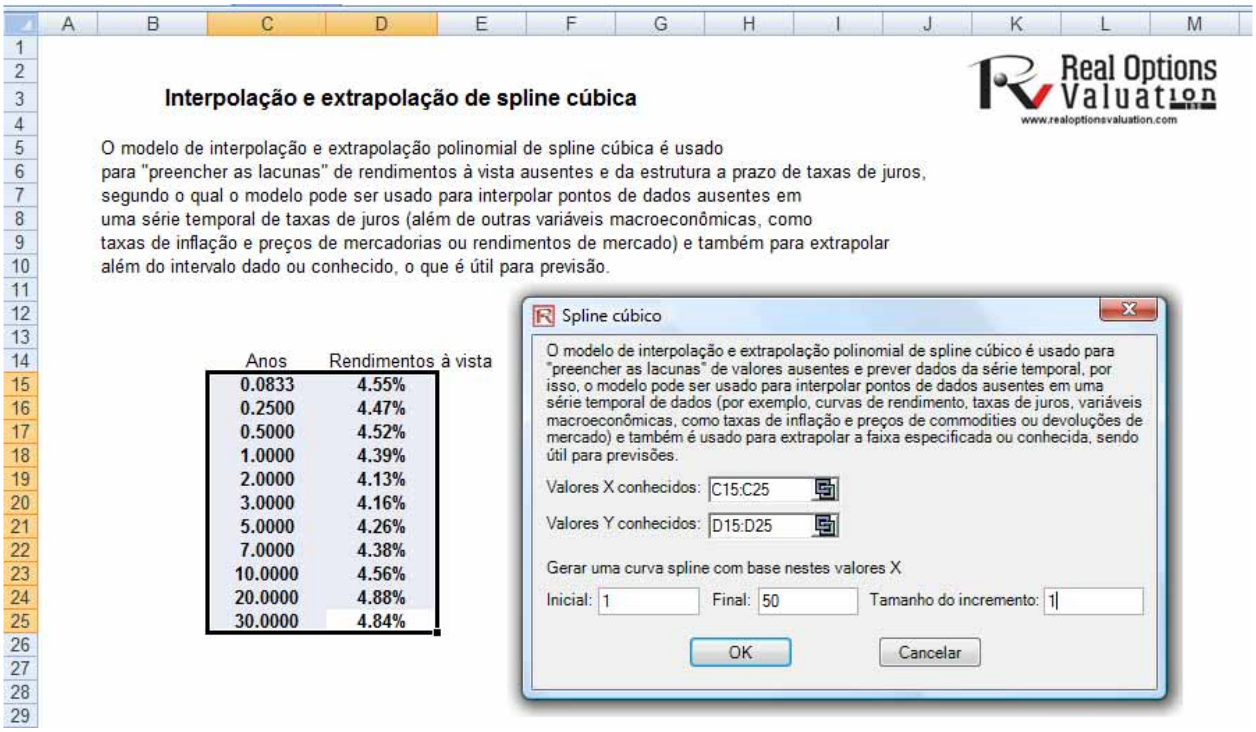


Figura 3.21 Módulo de *spline* cúbico

Procedimento:

- Inicie o Excel, abra o arquivo de exemplo ***Modelos avançados de previsão***, vá para a planilha ***Spline cúbico***, selecione o conjunto de dados excluindo os cabeçalhos e clique em ***Risk Simulator | Previsão | Spline cúbico***.
- A localização dos dados será inserida automaticamente na interface do usuário se você selecionar os dados primeiro, mas também é possível clicar manualmente no ícone de vínculo e vincular os valores de *X conhecidos* e de *Y conhecidos* (consulte a Figura 3.21 para obter um exemplo), inserir os valores obrigatórios *Inicial* e *Final* para extrapolar e interpolar, bem como o *Tamanho do incremento* obrigatório entre esses valores inicial e final. Clique em ***OK*** para executar o modelo e o relatório (consulte a Figura 3.22).

Previsões de spline cúbico

O modelo de interpolação e extrapolação polinomial de spline cúbico é usado para "preencher as lacunas" de valores ausentes e prever dados da série temporal, por isso, o modelo pode ser usado para interpolar pontos de dados ausentes em uma série temporal de dados (por exemplo, curvas de rendimento, taxas de juros, variáveis macroeconômicas, como taxas de inflação e preços de commodities ou devoluções de mercado) e também é usado para extrapolar a faixa especificada ou

Resultados da interpolação e extrapolação de spline

X	Y ajustado	Notas
1.0	4.39%	Interpolar
2.0	4.13%	Interpolar
3.0	4.16%	Interpolar
4.0	4.22%	Interpolar
5.0	4.26%	Interpolar
6.0	4.32%	Interpolar
7.0	4.38%	Interpolar
8.0	4.44%	Interpolar
9.0	4.50%	Interpolar
10.0	4.56%	Interpolar
11.0	4.61%	Interpolar
12.0	4.66%	Interpolar
13.0	4.70%	Interpolar
14.0	4.74%	Interpolar
15.0	4.77%	Interpolar
16.0	4.80%	Interpolar
17.0	4.83%	Interpolar
18.0	4.85%	Interpolar
19.0	4.87%	Interpolar
20.0	4.88%	Interpolar
21.0	4.89%	Interpolar
22.0	4.89%	Interpolar
23.0	4.89%	Interpolar
24.0	4.89%	Interpolar
25.0	4.89%	Interpolar
26.0	4.88%	Interpolar
27.0	4.87%	Interpolar
28.0	4.86%	Interpolar
29.0	4.85%	Interpolar
30.0	4.84%	Interpolar
31.0	4.83%	Extrapolar
32.0	4.82%	Extrapolar
33.0	4.81%	Extrapolar
34.0	4.80%	Extrapolar
35.0	4.79%	Extrapolar

Estas são as entradas de valores conhecidas no modelo de interpolação e extrapolação de spline cúbico:

Observação	X conhecido	Y conhecido
1	0.0833	4.55%
2	0.2500	4.47%
3	0.5000	4.52%
4	1.0000	4.39%
5	2.0000	4.13%
6	3.0000	4.16%
7	5.0000	4.26%
8	7.0000	4.38%
9	10.0000	4.56%
10	20.0000	4.88%
11	30.0000	4.84%

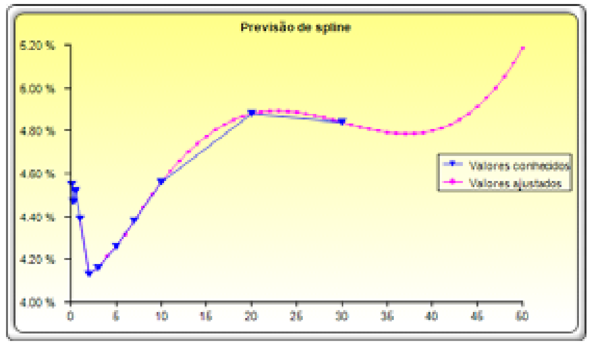


Figura 3.22 Resultados de previsão de *spline*

4. OTIMIZAÇÃO

Esta seção analisa o processo e as metodologias de otimização mais detalhadamente, pois isso está relacionado ao uso do Risk Simulator. Essas metodologias incluem o uso de otimização de valores inteiros, contínuos versus discretos, bem como otimizações estáticas, dinâmicas e estocásticas.

Metodologias de otimização

Há muitos algoritmos para executar otimização e há vários procedimentos diferentes quando a otimização é usada junto com a simulação Monte Carlo. No Risk Simulator, há três procedimentos de otimização e tipos de otimização distintos, além de diferentes tipos de variáveis de decisão. Por exemplo, o Risk Simulator pode lidar com ***Variáveis de decisão contínuas*** (1,2535, 0,2215 etc.), ***Variáveis de decisão inteiras*** (como 1, 2, 3, 4 ou 1,5, 2,5, 3,5 etc.), ***Variáveis de decisão binárias*** (1 e 0 para decisões de continuar ou não) e ***Variáveis de decisão mistas*** (variáveis inteiras e contínuas). Além disso, o Risk Simulator pode lidar com ***otimizações lineares*** (como, quando o objetivo e as restrições são funções e equações lineares) e com ***otimizações não lineares*** (como, quando os objetivos e as restrições são uma mistura de funções e equações lineares e não lineares).

No que diz respeito ao processo de otimização, o Risk Simulator pode ser usado para executar uma ***otimização discreta***, ou seja, uma otimização que é executada em um modelo discreto ou estático, no qual nenhuma simulação é executada. Em outras palavras, todas as entradas no modelo são estáticas e não se alteram. Esse tipo de otimização se aplica quando se supõe que o modelo seja conhecido e não há nenhuma incerteza. Além disso, uma otimização discreta pode ser executada primeiro, para determinar o portfólio ótimo e a alocação ótima de variáveis de decisão correspondentes, antes de aplicar procedimentos de otimização mais avançados. Por exemplo, antes de executar um problema de otimização estocástica, uma otimização discreta é executada para determinar se existe solução para o problema de otimização antes que uma análise mais longa seja realizada.

Em seguida, a ***otimização dinâmica*** é aplicada quando a simulação Monte Carlo é usada com a otimização. Outro nome usado para esse procedimento é ***otimização de simulação***. Ou seja, uma simulação é executada antes, os resultados da simulação são aplicados no modelo do Excel e, em seguida, aplica-se uma otimização aos valores simulados. Em outras palavras, uma simulação é executada por N tentativas, em seguida, executa-se um processo de otimização por M iterações até que se obtenha resultados ótimos, ou seja, encontrado um conjunto impraticável. Ou seja, usando o módulo de otimização do Risk Simulator, você pode escolher quais estatísticas de suposição e previsão devem ser usadas e substituídas no modelo após a execução da simulação. Depois, essas estatísticas de previsão podem ser aplicadas no processo de otimização. Essa abordagem é útil quando há um modelo grande, com a interação de muitas suposições e previsões, e quando algumas estatísticas de previsão são necessárias na otimização. Por exemplo, se o desvio padrão de uma suposição ou previsão é obrigatório no modelo de otimização (por exemplo, para calcular a razão de Sharpe na alocação de ativos e problemas de otimização, nos quais temos a média dividida pelo desvio padrão do portfólio), então esta abordagem deve ser usada.

O processo de ***otimização estocástica***, ao contrário, é semelhante ao procedimento de otimização dinâmica, com exceção de que todo o processo de otimização dinâmica é repetido *T* vezes. Isto é, executa-se uma simulação com *N* tentativas e, em seguida, uma otimização com *M* iterações para obter os resultados ótimos. Depois, o processo é replicado *T* vezes. O resultado será um gráfico de previsão para cada variável de decisão com *T* valores. Em outras palavras, uma simulação é executada e a previsão ou as estatísticas de suposição são usadas no modelo de otimização, para encontrar a alocação ideal das variáveis de decisão. Em seguida, outra simulação é executada, gerando estatísticas de previsão diferentes e esses valores atualizados são otimizados e assim por diante. Assim, as variáveis de decisão final terão cada uma seu próprio gráfico de previsão indicando a faixa ótima das variáveis de decisão ótimas. Por exemplo, em vez de obter estimativas de um único ponto no processo de otimização dinâmica, agora você pode obter uma distribuição das variáveis de decisão e, assim, uma faixa de valores ótimos para cada variável de decisão, também conhecido como otimização estocástica.

Por último, um procedimento de otimização de fronteira eficiente se aplica aos conceitos de incrementos marginais e precificação sombreada na otimização. Ou seja, o que aconteceria com os resultados da otimização, se uma das restrições fosse um pouco relaxada? Por exemplo, se a restrição de orçamento estiver definida como $1 milhão. O que aconteceria com o resultado do portfólio e as decisões ótimas se a restrição fosse agora $1,5 ou $2 milhões etc. Esse é o conceito de fronteiras eficientes de Markowitz em investimentos financeiros, em que se o desvio padrão do portfólio pode aumentar um pouco, quais serão os retornos adicionais gerados pelo portfólio? Esse processo é semelhante ao de otimização dinâmica, com a exceção de que a *uma* das restrições pode se alterar e, a cada alteração, o processo de simulação e otimização é executado. Esse processo pode ser melhor aplicado manualmente usando o Risk Simulator. Ou seja, execute uma otimização dinâmica ou estocástica e execute novamente outra otimização com uma restrição, repetindo esse processo, várias vezes. Esse processo manual é importante porque ao alterar a restrição, o analista pode determinar se os resultados são semelhantes ou diferentes e, portanto, se deve ser feita uma análise adicional, ou determinar quanto deve ser o aumento marginal da restrição para obter uma alteração significativa no objetivo e nas variáveis de decisão.

Deve-se ressaltar um item. Há outros softwares que supostamente executam a otimização estocástica, mas, de fato, não o fazem. Por exemplo, depois que uma simulação é executada, *uma* iteração do processo de otimização é gerada e, depois, outra simulação é executada e a *segunda* iteração de otimização é gerada e assim por diante. Um desperdício de tempo e recursos. Ou seja, em otimização, o modelo é submetido a um rigoroso conjunto de algoritmos, no qual várias iterações (de até milhares de iterações) são necessárias para obter os resultados ótimos. Assim, gerar *uma* iteração de cada vez é um desperdício de tempo e recursos. O mesmo portfólio pode ser solucionado usando o Risk Simulator em menos de um minuto, comparado a várias horas usando essa abordagem retrógrada. Tal abordagem de otimização-simulação também renderá resultados ruins, além de não ser uma abordagem de otimização estocástica. Cuidado com tais metodologias ao aplicar otimização aos seus modelos.

Veja a seguir dois exemplos de problemas de otimização. Um, usa variáveis de decisão contínuas, enquanto o outro, usa variáveis de decisão inteiras discretas. Nos dois modelos, você pode aplicar otimização discreta, dinâmica, estocástica ou ainda as fronteiras eficientes com precificação sombreada. Qualquer uma dessas abordagens pode ser usada nestes dois exemplos. Portanto, para simplificar, somente a configuração do modelo serão ilustrada e cabe ao usuário decidir qual processo de otimização executar. Além disso, o modelo contínuo usa a abordagem de otimização não linear (porque o risco do portfólio calculado é uma função não linear e o objetivo é uma função não linear dos retornos do portfólio divididos pelos riscos do portfólio), enquanto o segundo exemplo de uma otimização inteira é um modelo de otimização linear (o objetivo e todas as restrições são lineares). Consequentemente, os dois exemplos englobam todos os procedimentos já mencionados.

Otimização com variáveis de decisão contínuas

A Figura 4.1 ilustra o exemplo de modelo de otimização contínua. O exemplo usa o arquivo ***Otimização contínua*** disponível no menu Iniciar em ***Iniciar | Real Options Valuation | Risk Simulator | Exemplos*** ou diretamente por meio de ***Risk Simulator | Modelos de exemplo.*** Neste exemplo, há 10 classes de ativos distintas (diferentes tipos de fundos mútuos, ações ou ativos) cuja finalidade é alocar de maneira mais eficiente e eficaz os investimentos do portfólio, de tal forma, que seja obtida melhor relação custo-benefício. Ou seja, gerar os melhores retornos possíveis do portfólio, conhecidos os riscos inerentes a cada classe de ativos. Para entender o conceito de otimização, é necessário analisar o modelo de exemplo mais atentamente para ver a melhor maneira de usar o processo de otimização. O modelo mostra que cada uma das 10 classes de ativos tem seu próprio conjunto de retornos e de volatilidades anualizados. Essas medidas de risco e retorno são valores anualizados, de forma que possam ser comparadas consistentemente em todas as classes de ativos. Os retornos são calculados usando a média geométrica dos retornos relativos, enquanto os riscos são calculados usando a abordagem logarítmica relativa de retornos de ações. Consulte o Apêndice neste capítulo para obter mais detalhes sobre o cálculo da volatilidade anualizada e dos retornos anualizados de uma ação ou classe de ativos.

MODELO DE OTIMIZAÇÃO DE ALOCAÇÃO DE ATIVO

Descrição da classe de ativo	Retornos anualizados	Risco de volatilidade	Pesos alocados	Alocação mínima obrigatória	Alocação máxima obrigatória	Taxa de retorno sobre o risco	Ranking dos retornos (alto-baixo)	Ranking do risco (alto-baixo)	Ranking do retorno sobre risco (alto-baixo)	Ranking da alocação (alto-baixo)
Classe de ativo 1	10.54%	12.36%	10.00%	5.00%	35.00%	0.8524	9	2	7	1
Classe de ativo 2	11.25%	16.23%	10.00%	5.00%	35.00%	0.6929	7	8	10	1
Classe de ativo 3	11.84%	15.64%	10.00%	5.00%	35.00%	0.7570	6	7	9	1
Classe de ativo 4	10.64%	12.35%	10.00%	5.00%	35.00%	0.8615	8	1	5	1
Classe de ativo 5	13.25%	13.28%	10.00%	5.00%	35.00%	0.9977	5	4	2	1
Classe de ativo 6	14.21%	14.39%	10.00%	5.00%	35.00%	0.9875	3	6	3	1
Classe de ativo 7	15.53%	14.25%	10.00%	5.00%	35.00%	1.0898	1	5	1	1
Classe de ativo 8	14.95%	16.44%	10.00%	5.00%	35.00%	0.9094	2	9	4	1
Classe de ativo 9	14.16%	16.50%	10.00%	5.00%	35.00%	0.8584	4	10	6	1
Classe de ativo 10	10.06%	12.50%	10.00%	5.00%	35.00%	0.8045	10	3	8	1
Total do portfólio	*12.6419%*	*4.58%*	*100.00%*							
Taxa de retorno sobre o risco	*2.7596*									

Figura 4.1 Modelo de otimização contínua

Os pesos alocados na coluna E incluem as variáveis de decisão, as quais são as variáveis que precisam ser ajustadas e testadas para que o peso total se limite a 100% (célula E17). Geralmente, para iniciar a otimização, essas células são definidas com um valor uniforme, neste caso, as células E6 a E15 são definidas como 10% cada. Além disso, cada variável de decisão pode ter restrições específicas na faixa permitida. Neste exemplo, as alocações inferior e superior permitidas são 5% e 35%, como mostram as colunas F e G. Isso significa que cada classe de ativos pode ter seus próprios limites de alocação. Em seguida, a coluna H mostra a taxa de retorno sobre o risco, que é simplesmente o percentual de retorno dividido pelo percentual de risco; quanto maior esse valor, maior o retorno. O modelo restante mostra os rankings individuais de classe de ativos por retornos, risco, taxa de retorno sobre o risco e alocação. Em outras palavras, os rankings mostram resumidamente qual classe de ativos tem o menor risco ou o maior retorno etc.

Os retornos totais do portfólio na célula C17 são *SOMARPRODUTO(C6:C15, E6:E15)*, ou seja, a soma dos pesos alocados multiplicados pelos retornos anualizados de cada classe de ativos. Em outras palavras, temos $R_P = \omega_A R_A + \omega_B R_B + \omega_C R_C + \omega_D R_D$, onde R_P é o retorno no portfólio, $R_{A,B,C,D}$ são os retornos individuais nos projetos e $\omega_{A,B,C,D}$ são os respectivos pesos ou alocação de capital em cada projeto.

Além disso, o risco diversificado do portfólio na célula D17 é calculado por $\sigma_P = \sqrt{\sum_{i=1}^{i} \omega_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=1}^{m} 2\omega_i \omega_j \rho_{i,j} \sigma_i \sigma_j}$. Aqui, $\rho_{i,j}$ são as respectivas correlações cruzadas entre as classes de ativos. Dessa forma, se a correlação cruzada for negativa, haverá efeitos de diversificação de riscos e o risco do portfólio diminuirá. Contudo, para simplificar os cálculos, supomos zero correlações entre as classes de ativos por meio deste cálculo de risco de portfólio, mas supomos que elas existam ao aplicar a simulação nos retornos, como será visto mais adiante. Portanto, em vez de aplicar correlações estáticas entre esses diferentes retornos de ativos, aplicamos as correlações nas próprias suposições de simulação, criando uma relação mais dinâmica entre os valores de retorno simulados.

Por último, calcula-se a taxa de retorno sobre o risco, ou razão de Sharpe, para o portfólio. Esse valor é visto na célula C18 e representa o objetivo a ser maximizado neste exercício de otimização. Para resumir, temos as seguintes especificações neste modelo de exemplo:

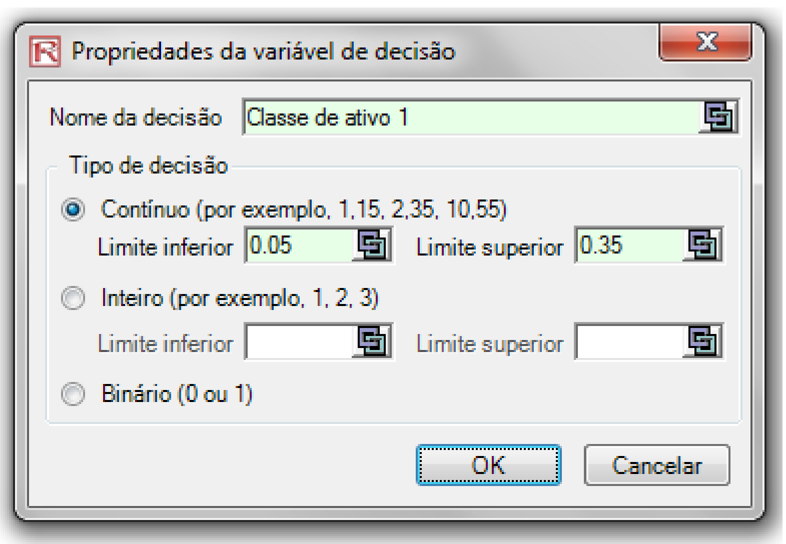
Objetivo:	*Maximizar a taxa de retorno sobre o risco (C18)*
Variáveis de decisão:	*Pesos alocados (E6:E15)*
Restrições das variáveis de decisão:	*Mínimo e máximo necessários (F6:G15)*
Restrições:	*A soma total dos pesos alocados é 100% (E17)*

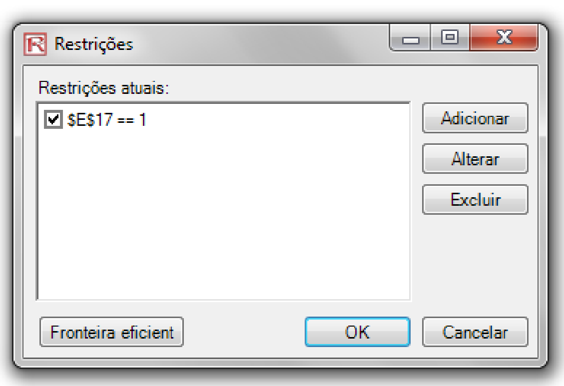
Procedimento:

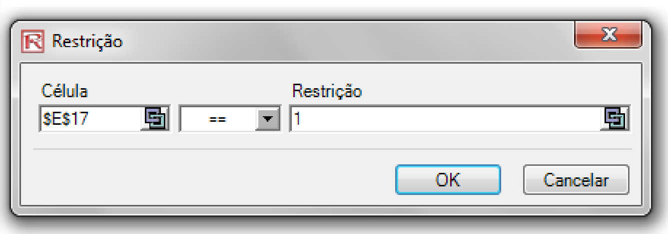
- Abra o arquivo de exemplo e inicie um novo perfil clicando em ***Risk Simulator* | *Novo perfil*** e nomeie-o.
- A primeira etapa na otimização é definir as variáveis de decisão. Selecione a célula E6 e defina a primeira variável de decisão(***Risk Simulator* | *Otimização* | *Definir decisão***) e clique no ícone de vínculo para selecionar o nome da célula (B6), bem como os valores de limite inferior e superior das células F6 e G6. Em seguida, usando a função de cópia do Risk Simulator, copie a variável de decisão da célula E6 e cole-a nas células restantes em E7 e E15.
- A segunda etapa na otimização é definir a restrição. A única restrição aqui é que o total de alocação no portfólio some 100%. Assim, clique em ***Risk Simulator* | *Otimização* | *Restrições…*** e selecione ***ADICIONAR*** para adicionar uma nova restrição. Em seguida, selecione a célula E17 e a iguale (=) a 100%. Clique em OK quando concluir.

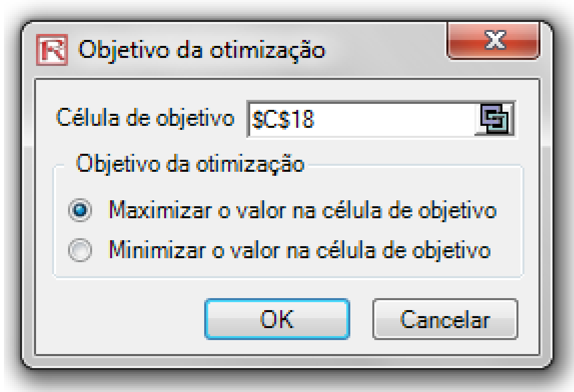
- A etapa final na otimização é definir a função do objetivo e iniciar a otimização. Para isso, selecione a célula de objetivo C18, clique em ***Risk Simulator | Otimização | Executar otimização*** e escolha a otimização desejada (estática, dinâmica ou estocástica). Para iniciar, selecione ***Otimização estática.*** Verifique se a célula de objetivo está definida como C18 e selecione ***Maximizar***. Agora você pode revisar as variáveis de decisão e restrições, se necessário, ou clique em OK para executar a otimização estática.
- Quando a otimização estiver concluída, será possível selecionar ***Reverter*** para reverter aos valores originais das variáveis de decisão e do objetivo. Alternativamente, você poderá selecionar ***Substituir*** para aplicar as variáveis de decisão otimizadas. Geralmente, Substituir é escolhido depois que a otimização está concluída.

A Figura 4.2 mostra as capturas de tela das etapas do procedimento acima. Você pode adicionar suposições de simulação nos retornos e no risco do modelo (colunas C e D) e aplicar a otimização dinâmica e estocástica para mais prática.









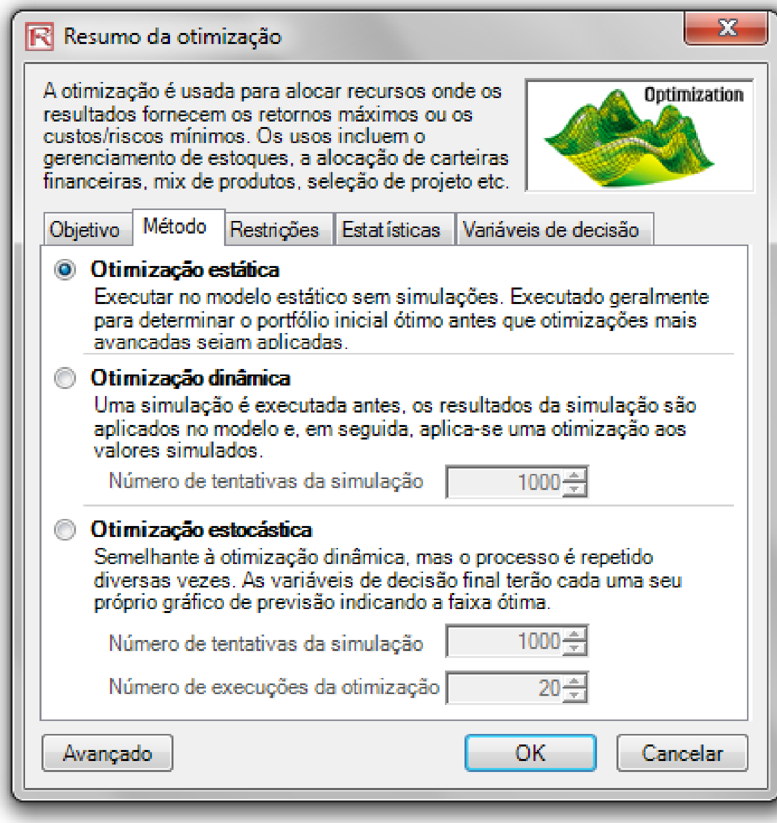


Figura 4.2 Execução de otimização contínua no Risk Simulator

Interpretação dos resultados:

Os resultados finais da otimização são mostrados na Figura 4.3, em que a alocação de ativos ótima para o portfólio é vista nas células E6:E15. Ou seja, sabendo-se que as restrições de cada ativo flutuam entre 5% e 35% e que a soma da alocação deve ser igual a 100%, a alocação que maximiza a taxa de retorno sobre o risco é vista na Figura 4.3. Alguns pontos importantes devem ser observados ao revisar os resultados e os procedimentos de otimização realizados até agora:

- A forma correta de executar a otimização é maximizar a relação custo-benefício ou a razão de Sharpe conforme fizemos.

- Se em vez disso maximizarmos os retornos totais do portfólio, o resultado de alocação ótima será trivial e não exigirá otimização para ser obtido. Ou seja, simplesmente aloque 5% (o mínimo necessário) para os 8 ativos mais baixos, 35% (o máximo permitido) para o ativo de maior retorno e o restante (25%) para os ativos com os segundos melhores retornos. A otimização não é obrigatória. Contudo, ao alocar o portfólio dessa forma, o risco será bem maior do que quando a taxa de retornos sobre o risco é maximizada, apesar de os retornos do portfólio serem maiores.
- Em contraste, é possível minimizar o risco total do portfólio, mas os retornos agora serão menores.

A Tabela 4.1 ilustra os resultados de três objetivos diferentes otimizados:

Objetivo:	Retornos do portfólio	Risco do portfólio	Taxa de retorno sobre o risco do portfólio
Maximizar taxa de retorno sobre o risco	*12,69%*	*4,52%*	*2,8091*
Maximizar o retorno	*13,97%*	*6,77%*	*2,0636*
Minimizar o risco	*12,38%*	*4,46%*	*2,7754*

Tabela 4.1 Resultados da otimização

Da tabela, a melhor abordagem é maximizar a taxa de retorno sobre o risco, ou seja, para a mesma quantidade de risco, a alocação fornece a maior quantidade de retorno. Por outro lado, para a mesma quantidade de retorno, a alocação fornece a quantidade mais baixa de risco possível. Essa abordagem de melhor relação custo-benefício ou taxa de retorno sobre o risco é a base da fronteira eficiente de Markowitz na teoria moderna de portfólio. Ou seja, se restringirmos os níveis de risco totais do portfólio e sucessivamente o aumentarmos com o tempo, obteremos várias alocações eficientes de portfólio para diferentes características de risco. Assim, diferentes alocações eficientes de portfólio podem ser obtidas para indivíduos diferentes com preferências de risco diferentes.

MODELO DE OTIMIZAÇÃO DE ALOCAÇÃO DE ATIVO

Descrição da classe de ativo	Retornos anualizados	Risco de volatilidade	Pesos alocados	Alocação mínima obrigatória	Alocação máxima obrigatória	Taxa de retorno sobre o risco
Classe de ativo 1	10.54%	12.36%	11.09%	5.00%	35.00%	0.8524
Classe de ativo 2	11.25%	16.23%	6.86%	5.00%	35.00%	0.6929
Classe de ativo 3	11.84%	15.64%	7.78%	5.00%	35.00%	0.7570
Classe de ativo 4	10.64%	12.35%	11.23%	5.00%	35.00%	0.8615
Classe de ativo 5	13.25%	13.28%	12.09%	5.00%	35.00%	0.9977
Classe de ativo 6	14.21%	14.39%	11.04%	5.00%	35.00%	0.9875
Classe de ativo 7	15.53%	14.25%	12.30%	5.00%	35.00%	1.0898
Classe de ativo 8	14.95%	16.44%	8.90%	5.00%	35.00%	0.9094
Classe de ativo 9	14.16%	16.50%	8.37%	5.00%	35.00%	0.8584
Classe de ativo 10	10.06%	12.50%	10.35%	5.00%	35.00%	0.8045
Total do portfólio	***12.6919%***	***4.52%***	***100.00%***			
Taxa de retorno sobre o risco	***2.8091***					

Figura 4.3 Resultados de otimização contínua

Otimização com variáveis inteiras discretas

Às vezes, as variáveis de decisão não são números inteiros contínuos, mas discretos, por exemplo, 0 e 1. Ou seja, podemos usar tal otimização como botões liga/desliga ou decisões do tipo sim/não. A Figura 4.4 ilustra um modelo de seleção de projeto que lista 12 projetos. O exemplo usa o arquivo***Otimização discreta*** disponível no menu Iniciar em ***Iniciar*** | ***Real Options Valuation*** | ***Risk Simulator*** | ***Exemplos*** ou diretamente por meio de ***Risk Simulator*** | ***Modelos de exemplo.*** Como anteriormente, cada projeto tem seus próprios retornos (VPLE para valor líquido atual expandido e e VPL para não expandido; o VPLE é simplesmente o VPL mais quaisquer valores de opções estratégicas reais), custos de implementação, riscos etc. Se necessário, este modelo pode ser modificado para incluir equivalências de período integral necessárias (FTE) e outros recursos de várias funções, podendo-se definir restrições adicionais nesses recursos adicionais. As entradas neste modelo geralmente são vinculadas a outros modelos de planilha. Por exemplo, cada projeto terá seu próprio fluxo de caixa descontado ou modelo de retorno do investimento. O uso é para maximizar a razão de Sharpe do portfólio de acordo com alguma alocação de orçamento. É possível criar muitas outras versões do modelo, por exemplo, maximizando os retornos do portfólio, minimizando os riscos, adicionando restrições nas quais o número total de projetos escolhidos não pode exceder 6 etc. Todos esses itens podem ser executados usando o modelo existente.

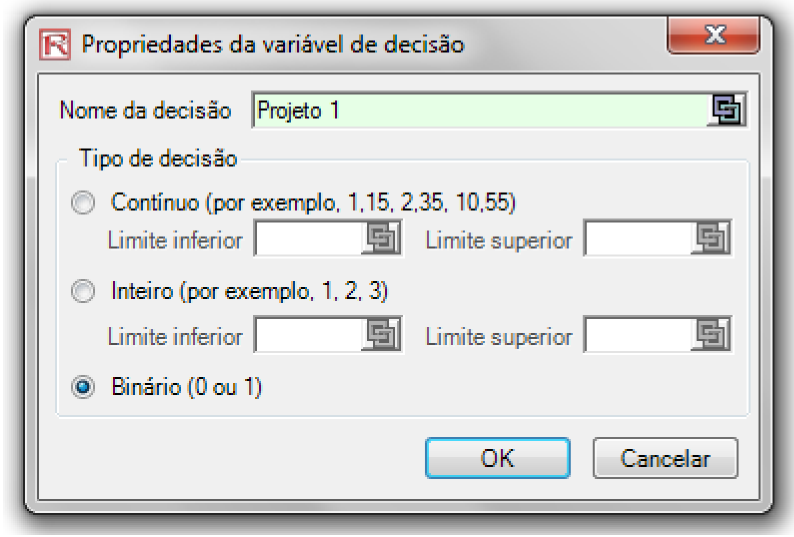
Procedimento:

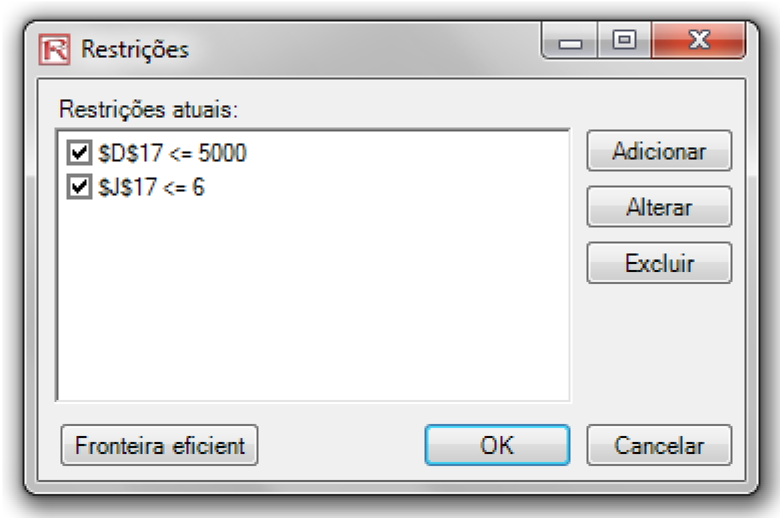
- Abra o arquivo de exemplo e inicie um novo perfil clicando em ***Risk Simulator*** | ***Novo perfil*** e nomeie-o.
- A primeira etapa na otimização é configurar as variáveis de decisão. Para definir a primeira variável, selecione a célula J4 e ***Risk Simulator*** | ***Otimização*** | ***Definir decisão***, clique no ícone de vínculo para escolher o nome da célula (B4) e selecione a variável ***Binária***. Em seguida, usando a função de cópia do Risk Simulator , copie a variável de decisão da célula J4 e cole-a nas células restantes em J5 e J15. Esse é o melhor método caso tenha muitas variáveis de decisão. É possível dar a cada variável um nome exclusivo para posterior identificação.
- A segunda etapa na otimização é definir a restrição. Há duas restrições aqui: a alocação do orçamento total no portfólio deve ser menor que $5.000 e o número total de projetos não deve exceder 6. Assim, clique em ***Risk Simulator*** | ***Otimização*** | ***Restrições…*** e selecione ***ADICIONAR*** para adicionar uma nova restrição. Em seguida, selecione a célula D17 e a faça menor ou igual (<=) a 5000. Repita definindo a célula J17 <= 6.
- A etapa final na otimização é definir a função do objetivo e iniciar a otimização. Para fazer isso, selecione a célula de objetivo C19 e clique em ***Risk Simulator*** | ***Otimização*** | ***Definir objetivo***, execute a otimização usando ***Risk Simulator*** | ***Otimização*** | ***Executar otimização*** e escolha a otimização desejada (estática, dinâmica ou estocástica). Para iniciar, selecione ***Otimização estática.*** Verifique se a célula de objetivo é a razão de Sharpe ou a taxa de retorno sobre o risco do portfólio e selecione ***Maximizar***. Agora você pode revisar as variáveis de decisão e restrições, se necessário, ou clique em OK para executar a otimização estática.

A Figura 4.5 mostra as capturas de tela das etapas do procedimento acima. Você pode adicionar suposições de simulação no VPLE e no risco do modelo (colunas C e E) e aplicar a otimização dinâmica e estocástica para praticar mais.

Projetos	VPLE	Custo	Risco $	Risco %	Taxa de retorno sobre o risco	Índice de probabilidade	Seleção
Projeto 1	$458.00	$1,732.44	$54.96	12.00%	8.33	1.26	1.0000
Projeto 2	$1,954.00	$859.00	$1,914.92	98.00%	1.02	3.27	1.0000
Projeto 3	$1,599.00	$1,845.00	$1,551.03	97.00%	1.03	1.87	1.0000
Projeto 4	$2,251.00	$1,645.00	$1,012.95	45.00%	2.22	2.37	1.0000
Projeto 5	$849.00	$458.00	$925.41	109.00%	0.92	2.85	1.0000
Projeto 6	$758.00	$52.00	$560.92	74.00%	1.35	15.58	1.0000
Projeto 7	$2,845.00	$758.00	$5,633.10	198.00%	0.51	4.75	1.0000
Projeto 8	$1,235.00	$115.00	$926.25	75.00%	1.33	11.74	1.0000
Projeto 9	$1,945.00	$125.00	$2,100.60	108.00%	0.93	16.56	1.0000
Projeto 10	$2,250.00	$458.00	$1,912.50	85.00%	1.18	5.91	1.0000
Projeto 11	$549.00	$45.00	$263.52	48.00%	2.08	13.20	1.0000
Projeto 12	$525.00	$105.00	$309.75	59.00%	1.69	6.00	1.0000
Total	$17,218.00	$8,197.44	$7,007	40.70%			12.00
Meta:	MÁX	< =$5000					<=6
Razão de Sharpe	2.4573						

Figura 4.4 Modelo de otimização inteira discreta


Propriedades da variável de decisão
Nome da decisão
Projeto 1
Tipo de decisão
Contínuo (por exemplo, 1,15, 2,35, 10,55)
Limite inferior
Limite superior
Inteiro (por exemplo, 1, 2, 3)
Limite inferior
Limite superior
Binário (0 ou 1)
OK
Cancelar


Restrições
Restrições atuais:
D17 <= 5000
J17 <= 6
Adicionar
Alterar
Excluir
Fronteira eficient
OK
Cancelar

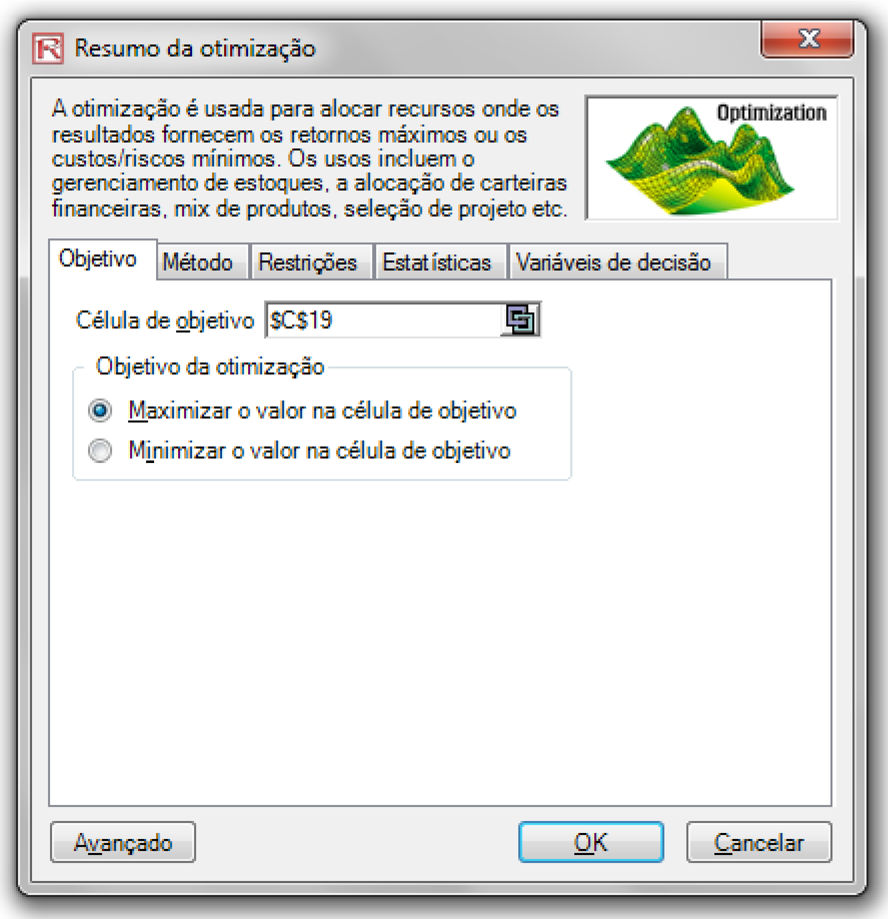


Figura 4.5 Execução de otimização inteira discreta no Risk Simulator

Interpretação dos resultados:

A Figura 4.6 mostra um exemplo de seleção ótima de projetos que maximiza a razão de Sharpe. Em contraste, sempre é possível maximizar as receitas totais, mas como anteriormente, esse é um processo trivial no qual simplesmente se escolhe o projeto de maior retorno e percorre-se a lista até ficar sem dinheiro ou exceder a restrição de orçamento. Isso, teoricamente, renderá projetos indesejáveis, pois os projetos de maior rendimento geralmente apresentam riscos mais altos. Agora, se desejar, você pode replicar a otimização usando uma otimização estocástica ou dinâmica adicionando suposições nos valores de VPLE, custo ou risco.

Projetos	VPLE	Custo	Risco $	Risco %	Taxa de retorno sobre o risco	Índice de probabilidade	Seleção
Projeto 1	$458.00	$1,732.44	$54.96	12.00%	8.33	1.26	1.0000
Projeto 2	$1,954.00	$859.00	$1,914.92	98.00%	1.02	3.27	0.0000
Projeto 3	$1,599.00	$1,845.00	$1,551.03	97.00%	1.03	1.87	0.0000
Projeto 4	$2,251.00	$1,645.00	$1,012.95	45.00%	2.22	2.37	1.0000
Projeto 5	$849.00	$458.00	$925.41	109.00%	0.92	2.85	0.0000
Projeto 6	$758.00	$52.00	$560.92	74.00%	1.35	15.58	1.0000
Projeto 7	$2,845.00	$758.00	$5,633.10	198.00%	0.51	4.75	0.0000
Projeto 8	$1,235.00	$115.00	$926.25	75.00%	1.33	11.74	1.0000
Projeto 9	$1,945.00	$125.00	$2,100.60	108.00%	0.93	16.56	0.0000
Projeto 10	$2,250.00	$458.00	$1,912.50	85.00%	1.18	5.91	0.0000
Projeto 11	$549.00	$45.00	$263.52	48.00%	2.08	13.20	1.0000
Projeto 12	$525.00	$105.00	$309.75	59.00%	1.69	6.00	1.0000
Total	$5,776.00	$3,694.44	$1,539	26.64%			6.00
Meta:	MÁX	< =$5000					<=6
Razão de Sharpe	3.7543						

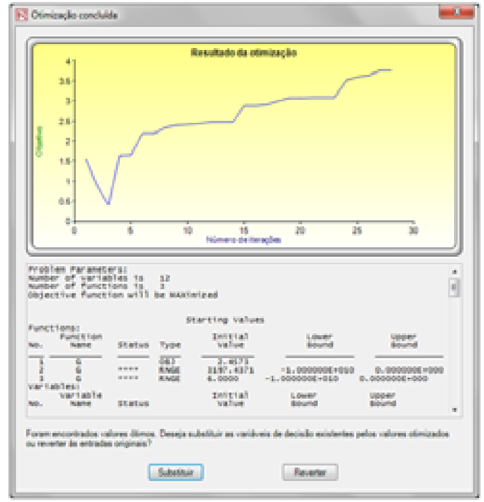


Figura 4.6 Seleção ótima de projetos que maximiza a razão de Sharpe

Para obter mais exemplos práticos de otimização, consulte o estudo de caso no Capítulo 11 sobre *análise de risco integrada*, no livro *Real Options Analysis: Tools and Techniques*, 2ª edição (Wiley Finance, 2005). Esse estudo de caso ilustra como gerar uma fronteira eficiente e como combinar previsão, simulação, otimização e opções reais em um processo analítico integrado.

Fronteira eficiente e configurações de otimização avançada

O segundo gráfico na Figura 4.5 mostra as restrições de otimização. Aqui, se você clicou no botão ***Fronteira eficiente*** *depois* de definir algumas restrições, poderá agora alterá-las. Ou seja, cada uma das restrições pode ser criada para percorrer cada valor máximo e mínimo. Por exemplo, a restrição na célula J17 <= 6 pode ser definida para ser executada entre 4 e 8 (Figura 4.7). Ou seja, cinco otimizações serão executadas, cada uma com as seguintes restrições: J17 <= 4, J17 <= 5, J17 <= 6, J17 <= 7 e J17 <= 8. Os resultados ótimos serão "plotados" como uma fronteira eficiente e o relatório será gerado (Figura 4.8). Especificamente, o processo a seguir ilustra as etapas necessárias para criar uma restrição mutante:

- Em um modelo de otimização, ou seja, um modelo com objetivo, variáveis de decisão e restrições já configuradas, clique em ***Risk Simulator*** | ***Otimização*** | ***Restrições*** e em ***Fronteira eficiente***.
- Selecione a restrição que deseja alterar ou percorrer (como J17) e insira os parâmetros para Mín., Máx. e Tamanho do incremento (Figura 4.7), clique em ***ADICIONAR*** e em ***OK*** duas vezes. Desmarque a restrição D17 <= 5000 antes de executar.
- Execute a otimização normalmente (***Risk Simulator*** | ***Otimização*** | ***Executar otimização***). Você pode escolher estática, dinâmica ou estocástica.
- Os resultados serão mostrados em uma interface do usuário (Figura 4.8). Clique em ***Criar relatório*** para gerar uma planilha de relatório com todos os detalhes das execuções de otimização.

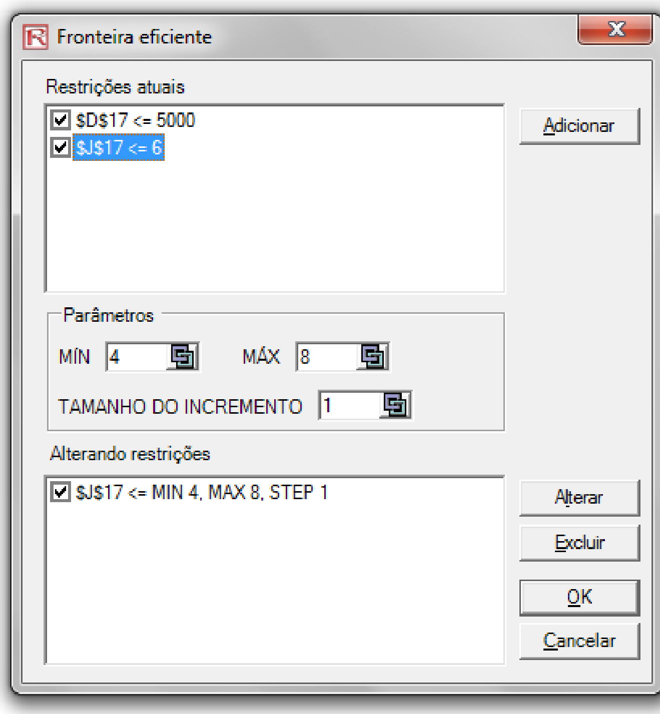


Figura 4.7 Geração de restrições mutantes em uma fronteira eficiente

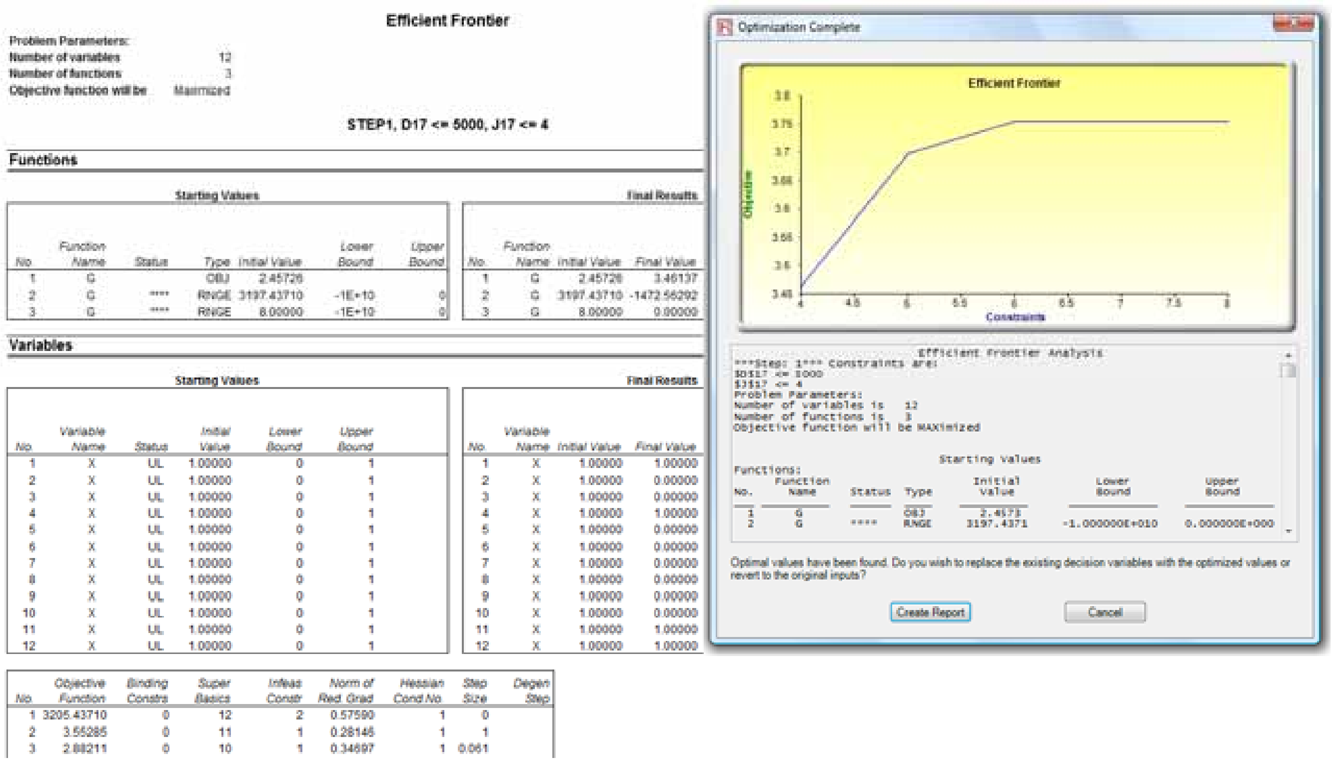


Figura 4.8 Resultados da fronteira eficiente

Otimização estocástica

O próximo exemplo ilustra a aplicação da otimização estocástica usando um modelo de exemplo com quatro classes de ativos, cada uma com características de risco e de retorno diferentes. O objetivo é encontrar a melhor alocação de portfólio, tal que seja maximizada a relação custo-benefício ou a taxa de retorno sobre o risco do portfólio. Ou seja, a meta é alocar 100% de um investimento individual entre várias classes de ativos diferentes (diferentes tipos de fundos mútuos ou estilos de investimento: crescimento, valor, crescimento agressivo, rendimento, global, índice, contrário, momento etc.). Este modelo é diferente de outros pois possui várias suposições de simulação (valores de risco e retorno para cada ativo nas colunas C e D), como mostra a Figura 4.9.

Executa-se uma simulação e, em seguida, a otimização e todo o processo é repetido várias vezes para obter distribuições de cada variável de decisão. Toda a análise pode ser automatizada usando a otimização estocástica. Para executar uma otimização, é preciso primeiro identificar várias especificações importantes no modelo:

Objetivo: *Maximizar a taxa de retorno sobre o risco (C12)*

Variáveis de decisão: *Pesos alocados (E6:E9)*

Restrições das variáveis de decisão: *Mínimo e máximo necessários (F6:G9)*

Restrições: *Total de 100% dos pesos alocados do portfólio (E11 definida como 100%)*

Suposições da simulação: *Valores de retorno e risco (C6:D9)*

O modelo mostra as várias classes de ativos. Cada classe de ativo tem seu próprio conjunto de retornos e volatilidades anualizados. Essas medidas de risco e retorno são valores anualizados, de forma que possam ser comparadas consistentemente em todas as classes de ativos. Os retornos são calculados usando a média geométrica dos retornos relativos, enquanto os riscos são calculados usando a abordagem logarítmica relativa de retornos de ações.

Os pesos alocados na coluna E incluem as variáveis de decisão, as quais são as variáveis que precisam ser ajustadas e testadas para que o peso total se limite a 100% (célula E17). Geralmente, para iniciar a otimização, essas células são definidas com um valor uniforme. Neste caso, as células E6 a E9 são definidas como 25% cada. Além disso, cada variável de decisão pode ter restrições específicas na faixa permitida. Neste exemplo, as alocações inferior e superior permitidas são 10% e 40%, como mostram as colunas F e G. Isso significa que cada classe de ativos pode ter seus próprios limites de alocação.

	A	B	C	D	E	F	G
1							
2							
3							
4		MODELO DE OTIMIZAÇÃO DE ALOCAÇÃO DE ATIVO					
5		Descrição da classe de ativo	Retornos anualizados	Risco de volatilidade	Pesos alocados	Alocação mínima obrigatória	Alocação máxima obrigatória
6		Ativo 1	10.57%	12.38%	25.00%	10.00%	40.00%
7		Ativo 2	11.13%	16.19%	25.00%	10.00%	40.00%
8		Ativo 3	10.51%	15.95%	25.00%	10.00%	40.00%
9		Ativo 4	10.64%	12.45%	25.00%	10.00%	40.00%
10							
11		*Total do portfólio*	*10.7129%*	*7.18%*	*100.00%*		
12		*Taxa de retorno sobre o risco*	*1.4921*				
13							

Figura 4.9: Modelo de alocação de ativo pronto para otimização estocástica

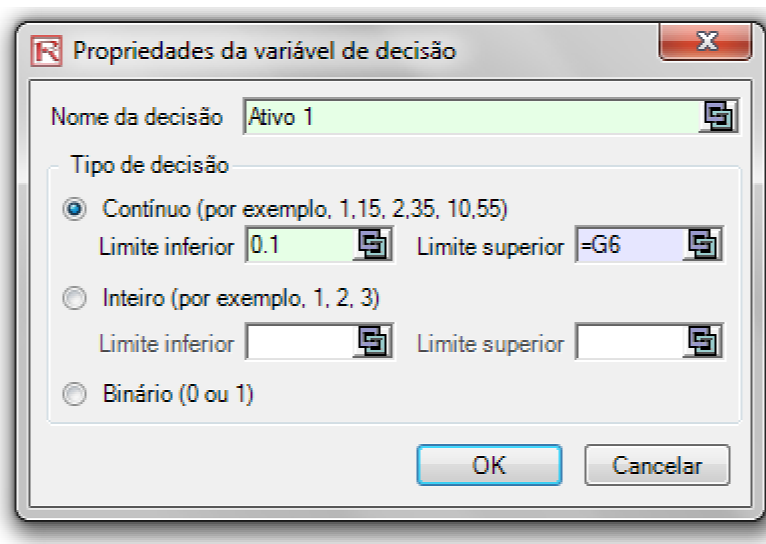
Em seguida, a coluna H mostra a taxa de retorno sobre o risco, que é simplesmente o percentual de retorno dividido pelo percentual de risco; quanto maior esse valor, maior o retorno. As partes restantes do modelo mostram os rankings individuais de classe de ativos por retornos, risco, taxa de retorno sobre o risco e alocação. Em outras palavras, os rankings mostram resumidamente qual classe de ativos tem o menor risco ou o maior retorno etc.

Execução de uma otimização

Para executar o modelo, clique em **Risk Simulator | Otimização | Executar otimização**. Como alternativa, você pode configurar o modelo usando as seguintes etapas.

1. Crie um novo perfil (**Risk Simulator | Novo perfil**).
2. Para otimização estocástica, defina suposições de distribuição sobre o risco e os retornos de cada classe de ativo. Ou seja, selecione a célula **C6** e defina uma suposição (**Risk Simulator | Definir valores de entrada**) e crie sua suposição conforme necessário. Repita para as células **C7** a **D9**.

3. Selecione a célula **E6** e defina a variável de decisão (**Risk Simulator | Otimização | Definir decisão** ou clique no ícone *Definir decisão **D***), torne-a uma **Variável contínua** e vincule o nome, o mínimo e o máximo da variável de decisão necessários às células relevantes (**B6, F6, G6**).
4. Use a função **copiar do Risk Simulator** na célula **E6**, selecione as células **E7 a E9** e use a função **colar do Risk Simulator** (**Risk Simulator | Copiar parâmetro**) e **Risk Simulator | Colar parâmetro** ou use os ícones de cópia e colagem). Não use as funções normais do Excel para copiar e colar.
5. Em seguida, para configurar as restrições da otimização, selecione **Risk Simulator | Otimização | Restrições**, **ADICIONAR** e a célula **E11** e a iguale a **100%** (alocação total e inclua o sinal %).
6. Selecione a célula **C12**, o objetivo a ser maximizado e transforme-a no objetivo: **Risk Simulator | Otimização | Definir objetivo** ou clique no ícone **O**.
7. Para executar a otimização, clique em **Risk Simulator | Otimização | Executar otimização**. Revise as diferentes guias para verificar se todas as entradas necessárias nas etapas 2 e 3 estão corretas. Selecione **Otimização estocástica** e execute-a por 500 tentativas repetidas 20 vezes (a Figura 4.10 ilustra essas etapas de configuração).



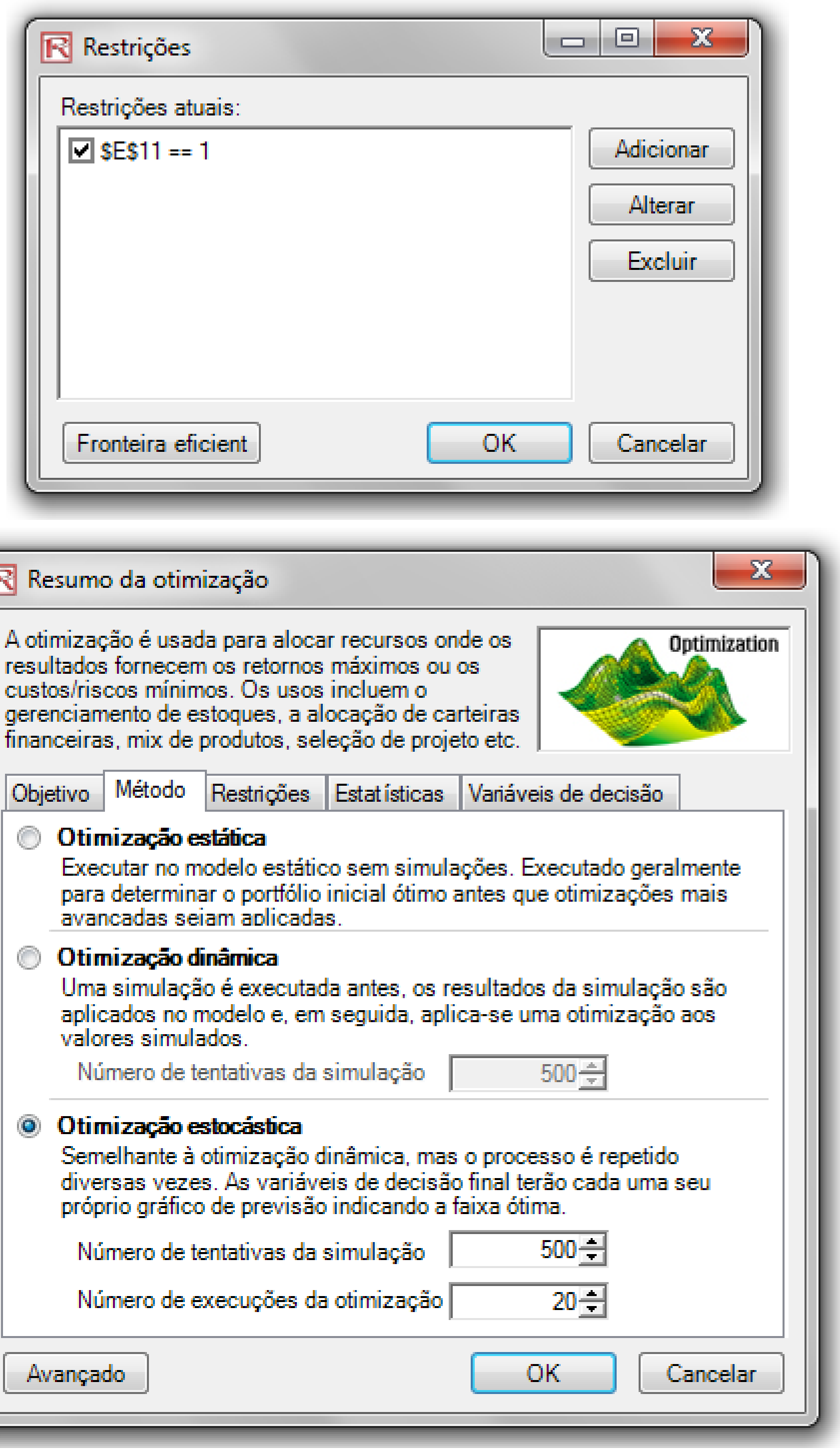


Figura 4.10: Configuração do problema de otimização estocástica

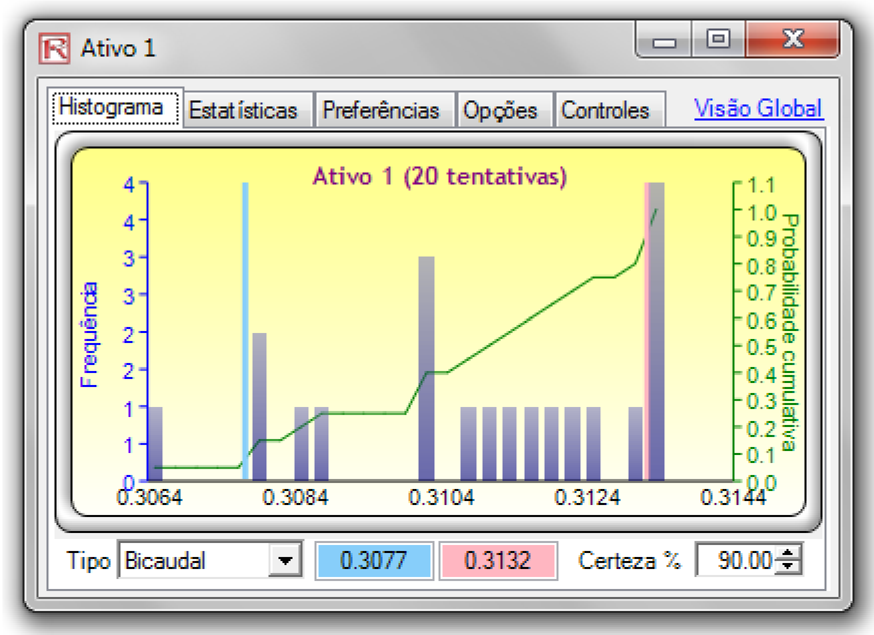
Quando a simulação for concluída, clique em **OK** para gerar um relatório de otimização estocástica detalhado, contendo gráficos de previsão das variáveis de decisão.

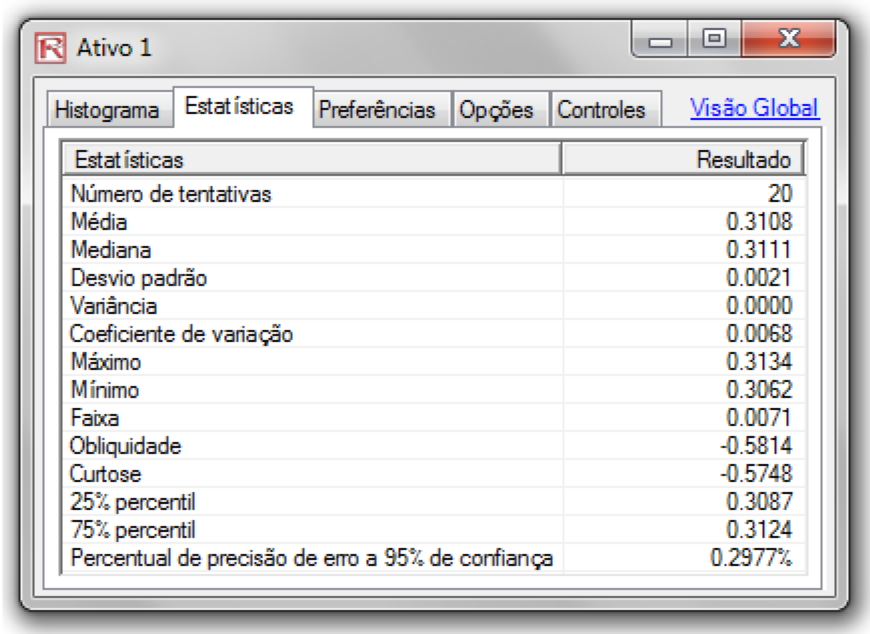
Exibir e interpretar resultados de previsões

A otimização estocástica é realizada quando primeiro é executada uma simulação e, depois, uma otimização. Em seguida, toda a análise é repetida várias vezes. O resultado é uma distribuição de cada variável de decisão, em vez de uma estimativa de único ponto (Figura 4.11). Isso significa que em vez de dizer que você deve investir 31,08% no ativo 1, a decisão ótima é investir entre 30,77% e 31,32% desde que o portfólio total some 100%. Dessa forma, os resultados dão ao gestor ou aos tomadores de decisão a uma faixa de flexibilidade dentro das decisões ótimas ao mesmo tempo em que levam em consideração os riscos e as incertezas das entradas.

Notas:

- **Simulação super-rápida com otimização**. Você também pode executar uma otimização estocástica com a simulação super-rápida. Para fazer isso, redefina a otimização redefinindo todas as quatro variáveis de decisão de volta a 25%. Em seguida, clique em ***Executar otimização*** e no botão ***Avançado*** (Figura 4.10) e marque a caixa de seleção ***Executar simulação super-rápida***. Na interface de execução da otimização, selecione ***Otimização estocástica*** na guia ***Método*** e a defina para ser executada por ***500*** tentativas e ***20*** execuções de otimização, e clique em ***OK***. Essa abordagem integra a simulação super-rápida à otimização. Observe que a otimização estocástica é executada muito mais rapidamente. Agora é possível executá-la de novo com um maior número de tentativas de simulação.
- **Estatísticas de simulação para otimizações estocástica e dinâmica**. Observe que se houver suposições de entrada de simulação no modelo de otimização (ou seja, essas suposições de entrada forem necessárias para executar as rotinas de otimização dinâmica ou estocástica), a guia ***Estatísticas*** estará preenchida na janela ***Executar otimização***. Na lista suspensa, você pode selecionar as estatísticas desejadas, como média, desvio padrão, coeficiente de variação, média condicional e variância condicional, um percentil específico etc. Isso significa que se for executada uma otimização estocástica, uma simulação de milhares de tentativas será executada primeiro, as estatísticas selecionadas serão calculadas e o valor será temporariamente colocado na célula de suposição de simulação e, em seguida, será executa uma otimização com base nessa estatística e todo o processo é repetido várias vezes. Esse método é importante e útil para bancos no cálculo do valor em risco condicional ou *VaR* condicional.





Estatísticas	Resultado
Número de tentativas	20
Média	0.3108
Mediana	0.3111
Desvio padrão	0.0021
Variância	0.0000
Coeficiente de variação	0.0068
Máximo	0.3134
Mínimo	0.3062
Faixa	0.0071
Obliquidade	-0.5814
Curtose	-0.5748
25% percentil	0.3087
75% percentil	0.3124
Percentual de precisão de erro a 95% de confiança	0.2977%

Figura 4.11: Resultados simulados da abordagem de otimização estocástica

5. FERRAMENTAS ANALÍTICAS DE SIMULAÇÃO DE RISCO

Este capítulo trata das ferramentas analíticas do Risk Simulator. Essas ferramentas analíticas são discutidas por meio de exemplos de aplicação do software Risk Simulator, com ilustrações passo-a-passo. Essas ferramentas são valiosas para analistas que trabalham na área de análise de risco. A aplicabilidade de cada ferramenta é discutida em detalhe neste capítulo.

Ferramentas Tornado e Sensibilidade em simulação

Teoria:

Uma das mais poderosas ferramentas de simulação é a análise tornado: ela captura os impactos estáticos de cada variável no resultado do modelo. Ou seja, a ferramenta perturba automaticamente cada variável no modelo por um valor predefinido, captura a flutuação na previsão do modelo ou no resultado final e lista as perturbações resultantes classificadas em ordem decrescente de relevância. As Figuras 5.1 a 5.6 ilustram a aplicação de uma análise tornado. Por exemplo, a Figura 5.1 é um exemplo de modelo de fluxo de caixa descontado, que mostra os valores de entrada. Quais são os fatores de sucesso que mais afetam o resultado do modelo? Ou seja, o que realmente controla o valor presente líquido de $96,63 ou qual variável de entrada causa mais impacto nesse valor?

A ferramenta de gráfico tornado está disponível em ***Risk Simulator | Ferramentas | Análise tornado***. Para seguir o primeiro exemplo, abra o arquivo ***Gráficos de sensibilidade e tornado (lineares)*** na pasta de exemplos. A Figura 5.2 mostra esse modelo de exemplo, no qual a célula G6 contendo o valor presente líquido é escolhida como meta do resultado a ser analisado. Os precedentes da célula de meta no modelo são usados para criar o gráfico tornado. Os precedentes são todas as variáveis de entrada e intermediárias que afetam o resultado do modelo. Por exemplo, se o modelo consiste em $A = B + C$, onde C = D + E, então B, D e E são precedentes de A (C não é um precedente, pois é apenas um valor calculado intermediário). A Figura 5.2 também mostra a faixa de testes de cada variável precedente usada para estimar o alvo do resultado. Se as variáveis precedentes forem entradas simples, a faixa de teste será uma perturbação simples, de acordo com a faixa escolhida (por exemplo, o padrão é ±10%). Cada variável precedente pode ser perturbada em diferentes porcentagens, se necessário. Uma faixa maior é importante, já que a capacidade para testar valores extremos é melhor do que perturbações menores ao redor dos valores esperados. Em algumas circunstâncias, valores extremos podem ter um impacto maior, menor ou não equilibrado (por exemplo, podem ocorrer não linearidades onde houver aumento ou diminuição das economias de escala e imprevistos do escopo para valores maiores ou menores de uma variável) e somente uma faixa mais ampla captura esse impacto não linear.

Modelo de fluxo de caixa descontado

Ano base	*2005*		*Soma dos benefícios líqu*	***$1,896.63***
Taxa de desconto ajustada pelo risco de r	*15.00%*		*Soma dos investimentos*	***$1,800.00***
Taxa de desconto de risco privado	*5.00%*		*Valor presente líquido*	***$96.63***
Taxa anualizada de crescimento de venda	*2.00%*		*Taxa de retorno interna*	*18.80%*
Taxa de erosão dos preços	*5.00%*		*Retorno do investimento*	*5.37%*
Taxa de juros efetiva	*40.00%*			

	2005	*2006*	*2007*	*2008*	*2009*
Preço médio do produto A	$10.00	$9.50	$9.03	$8.57	$8.15
Preço médio do produto B	$12.25	$11.64	$11.06	$10.50	$9.98
Preço médio do produto C	$15.15	$14.39	$13.67	$12.99	$12.34
Quantidade do produto A	50.00	51.00	52.02	53.06	54.12
Quantidade do produto B	35.00	35.70	36.41	37.14	37.89
Quantidade do produto C	20.00	20.40	20.81	21.22	21.65
Receita total	***$1,231.75***	***$1,193.57***	***$1,156.57***	***$1,120.71***	***$1,085.97***
Custo dos bens vendidos	$184.76	$179.03	$173.48	$168.11	$162.90
Lucro bruto	***$1,046.99***	***$1,014.53***	***$983.08***	***$952.60***	***$923.07***
Despesas operacionais	$157.50	$160.65	$163.86	$167.14	$170.48
Custos de vendas, gerais e admin.	$15.75	$16.07	$16.39	$16.71	$17.05
Rendimento operacional (EBITDA)	***$873.74***	***$837.82***	***$802.83***	***$768.75***	***$735.54***
Depreciação	$10.00	$10.00	$10.00	$10.00	$10.00
Amortização	$3.00	$3.00	$3.00	$3.00	$3.00
EBIT	***$860.74***	***$824.82***	***$789.83***	***$755.75***	***$722.54***
Pagamentos de juros	$2.00	$2.00	$2.00	$2.00	$2.00
EBT	***$858.74***	***$822.82***	***$787.83***	***$753.75***	***$720.54***
Impostos	$343.50	$329.13	$315.13	$301.50	$288.22
Rendimento líquido	***$515.24***	***$493.69***	***$472.70***	***$452.25***	***$432.33***
Depreciação	$13.00	$13.00	$13.00	$13.00	$13.00
Alteração no capital de giro líquido	$0.00	$0.00	$0.00	$0.00	$0.00
Despesas de capital	$0.00	$0.00	$0.00	$0.00	$0.00
Fluxo de caixa livre	***$528.24***	***$506.69***	***$485.70***	***$465.25***	***$445.33***
Investimentos	***$1,800.00***				

Figura 5.1: Modelo de exemplo

Procedimento:

- Selecione a única célula de resultado, por exemplo, uma célula com uma função ou equação, em um modelo do Excel, como a célula G6 é selecionada no nosso exemplo
- Selecione ***Risk Simulator | Ferramentas | Análise tornado***
- Revise as precedentes e as renomeie de forma apropriada (renomear as precedentes com nomes mais curtos torna gráficos tornado e aranha visualmente mais agradáveis) e clique em OK

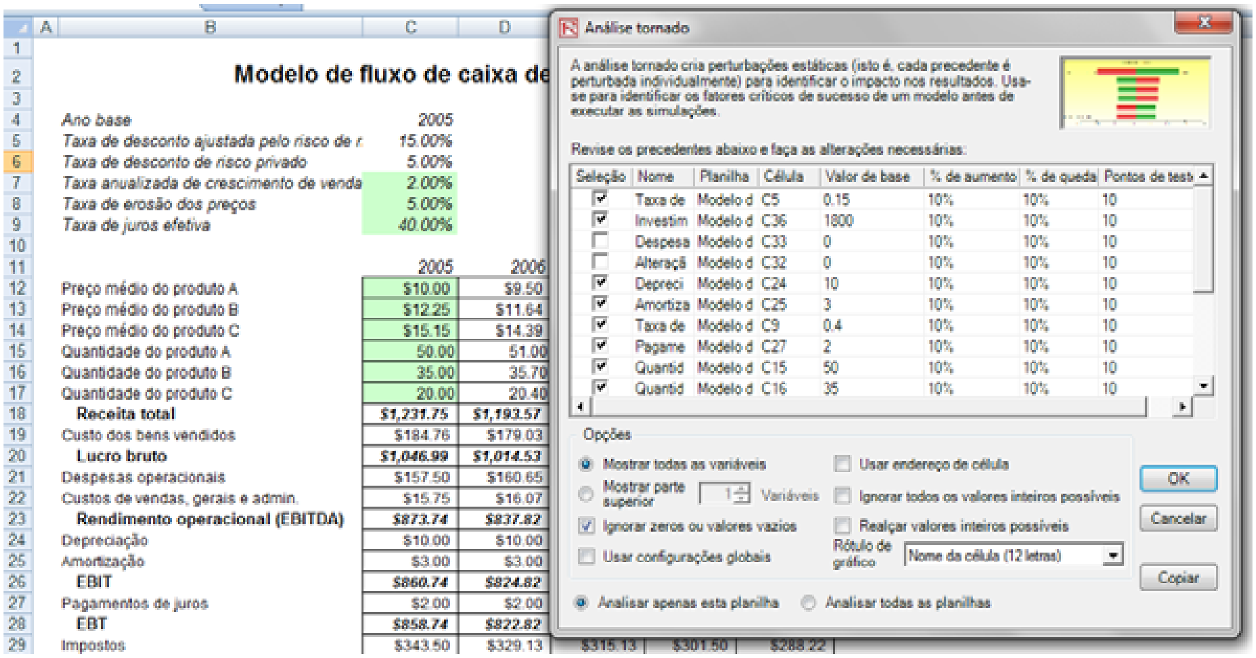


Figura 5.2 – Execução da análise tornado

Interpretação dos resultados:

A Figura 5.3 mostra o relatório de análise tornado resultante, que indica que o investimento de capital exerce o maior impacto sobre o valor presente líquido, seguido pela a alíquota de impostos, o preço de venda médio e a quantidade solicitada das linhas de produto etc. O relatório contém quatro elementos distintos:

- Resumo estatístico listando o procedimento executado.
- A tabela de sensibilidade (Figura 5.4) mostra os valores de base do VPL inicial de 96,63 e como cada entrada é alterada (por exemplo, o investimento é alterado de \$1.800 para \$1.980 no lucro com uma oscilação de +10%, e de \$1.800 para \$1.620 na perda, com uma oscilação de –10%). Os valores de lucro e perda resultantes no VPL são –\$83,37 e \$276,63, com uma alteração total de \$360, tornando-a a variável com maior impacto sobre o VPL. As variáveis precedentes são classificadas do maior ao menor impacto.
- O gráfico aranha (Figura 5.5) ilustra esses efeitos graficamente. O eixo y é o valor alvo de VPL e o eixo x representa a alteração percentual em cada valor precedente (o ponto central é o valor do caso básico, em 96,63 com alteração de 0% do valor básico de cada precedente). Uma linha inclinada positivamente indica uma possível relação ou efeito, enquanto linhas inclinadas negativamente indicam uma relação negativa (por exemplo, se o investimento estiver inclinado negativamente, isso significará que quanto maior for o nível do investimento, menor será o VPL). O valor absoluto da inclinação indica a magnitude do efeito (uma linha muito inclinada indica um impacto maior no eixo y do VPL, dada uma alteração no eixo x precedente).
- O gráfico tornado ilustra isso de outra maneira gráfica, na qual a precedente de maior impacto é listada primeiro. Um eixo x é o valor de VPL com o centro do gráfico sendo a condição do caso básico. As barras verdes no gráfico indicam um efeito positivo e as barras vermelhas indicam um efeito negativo. Por conseguinte, em investimentos, a barra

vermelha no lado direito indica um efeito negativo do investimento em um VPL mais alto. Em outras palavras, o investimento de capital e o VPL são negativamente correlacionados. O contrário também vale para o preço e a quantidade dos produtos A a C (as barras verdes estão no lado direito do gráfico).

Gráficos tornado e aranha

Resumo estatístico

Uma das mais poderosas ferramentas de simulação é o gráfico tornado –– ele captura os impactos estáticos de cada variável no resultado do modelo. Ou seja, a ferramenta perturba automaticamente cada variável precedente no modelo por um valor predefinido especificado pelo usuário, captura a flutuação na previsão do modelo ou no resultado final e lista as perturbações resultantes classificadas em ordem decrescente de relevância. Os precedentes são todas as variáveis de entrada e intermediárias que afetam o resultado do modelo. Por exemplo, se o modelo consiste em A = B + C, onde C = D + E, então B, D e E são precedentes de A (C não é um precedente, pois é apenas um valor calculado intermediário). A faixa e o número de valores perturbados são especificados pelo usuário e podem ser definidos para testar valores extremos em vez de perturbações menores em torno dos valores esperados. Em algumas circunstâncias, valores extremos podem ter um impacto maior, menor ou não equilibrado (por exemplo, podem ocorrer não linearidades onde houver aumento ou diminuição das economias de escala e imprevistos do escopo para valores maiores ou menores de uma variável) e somente uma faixa mais ampla captura esse impacto não linear.

Um gráfico tornado lista todas as entradas que controlam o modelo, começando pela variável de entrada que exerce o maior efeito nos resultados. O gráfico é obtido pela perturbação de cada entrada precedente em uma faixa consistente (por exemplo, ±10% do caso básico) de cada vez e pela comparação dos resultados com o caso básico. Um gráfico aranha se assemelha a uma aranha, com um corpo central e muitas patas salientes. A linha positivamente inclinada indica uma relação positiva e uma linha negativamente inclinada indica uma relação negativa. Além disso, os gráficos aranha podem ser usados para mostrar relações lineares e não lineares. Os gráficos tornado e aranha ajudam a identificar os fatores críticos de sucesso de uma célula de resultado, para identificar as entradas que devem ser simuladas. As variáveis críticas identificadas como incertas são aquelas que devem ser simuladas. Não desperdice tempo simulando variáveis que não são incertas ou que exercem pouco impacto nos resultados.

Resultado

	Valor de base: 96.6261638553219			Alterações da entrada		
Célula precedente	Queda do resultado	Aumento do resultado	Faixa efetiva	Queda da entrada	Aumento da entrada	Valor do caso de base
C36: Investimentos	276.62616	-83.373836	360.00	$1,620.00	$1,980.00	$1,800.00
C9: Taxa de juros efetiva	219.72693	-26.474599	246.20	36.00%	44.00%	40.00%
C12: Preço médio do produto A	3.4255424	189.82679	186.40	$9.00	$11.00	$10.00
C13: Preço médio do produto B	16.706631	176.5457	159.84	$11.03	$13.48	$12.25
C15: Quantidade do produto A	23.177498	170.07483	146.90	45.00	55.00	50.00
C16: Quantidade do produto B	30.533	162.71933	132.19	31.50	38.50	35.00
C14: Preço médio do produto C	40.146587	153.10574	112.96	$13.64	$16.67	$15.15
C17: Quantidade do produto C	48.047369	145.20496	97.16	18.00	22.00	20.00
C5: Taxa de desconto ajustada pelo risco de	138.23913	57.029841	81.21	13.50%	16.50%	15.00%
C8: Taxa de erosão dos preços	116.80381	76.640952	40.16	4.50%	5.50%	5.00%
C7: Taxa anualizada de crescimento de ven	90.588354	102.68541	12.10	1.80%	2.20%	2.00%
C24: Depreciação	95.084173	98.168155	3.08	$9.00	$11.00	$10.00
C25: Amortização	96.163566	97.088761	0.93	$2.70	$3.30	$3.00
C27: Pagamentos de juros	97.088761	96.163566	0.93	$1.80	$2.20	$2.00

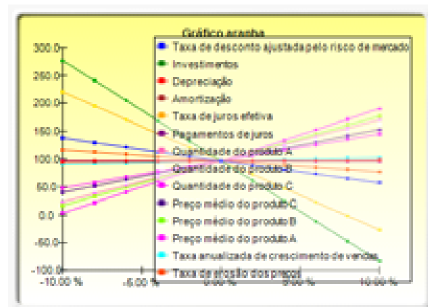


Figura 5.3 – Relatório de análise tornado

Notas:

A análise tornado é uma análise de sensibilidade *estática* aplicada a cada variável de entrada no modelo, ou seja, cada variável é perturbada individualmente e os efeitos resultantes são tabulados. Isso torna a análise tornado um componente-chave a ser executado antes de uma simulação. Uma das primeiras etapas na análise de risco é capturar e identificar os fatores de maior impacto no modelo. A próxima etapa é identificar quais desses fatores de impacto são incertos. Esses fatores de impacto incertos são fatores importantes para o sucesso de um projeto, dos quais dependem os resultados do modelo. Essas variáveis são as que devem ser simuladas. Não desperdice tempo simulando variáveis que não são incertas ou que exercem pouco impacto nos resultados. Os gráficos tornado auxiliam na identificação desses fatores de sucesso com rapidez e facilidade. Seguindo este exemplo, é possível que o preço e a quantidade devam ser simulados, supondo que os investimentos necessários e a alíquota de impostos efetiva sejam conhecidos antecipadamente e inalteráveis.

	Valor de base: 96.6261638553219			Alterações da entrada		
Célula precedente	Queda do resultado	Aumento do resultado	Faixa efetiva	Queda da entrada	Aumento da entrada	Valor do caso de base
C36: Investimentos	276.62616	-83.373836	360.00	$1,620.00	$1,980.00	$1,800.00
C9: Taxa de juros efetiva	219.72693	-26.474599	246.20	36.00%	44.00%	40.00%
C12: Preço médio do produto A	3.4255424	189.82679	186.40	$9.00	$11.00	$10.00
C13: Preço médio do produto B	16.706631	176.5457	159.84	$11.03	$13.48	$12.25
C15: Quantidade do produto A	23.177498	170.07483	146.90	45.00	55.00	50.00
C16: Quantidade do produto B	30.533	162.71933	132.19	31.50	38.50	35.00
C14: Preço médio do produto C	40.146587	153.10574	112.96	$13.64	$16.67	$15.15
C17: Quantidade do produto C	48.047369	145.20496	97.16	18.00	22.00	20.00
C5: Taxa de desconto ajustada pelo risco d	138.23913	57.029841	81.21	13.50%	16.50%	15.00%
C8: Taxa de erosão dos preços	116.80381	76.640952	40.16	4.50%	5.50%	5.00%
C7: Taxa anualizada de crescimento de ven	90.588354	102.68541	12.10	1.80%	2.20%	2.00%
C24: Depreciação	95.084173	98.168155	3.08	$9.00	$11.00	$10.00
C25: Amortização	96.163566	97.088761	0.93	$2.70	$3.30	$3.00
C27: Pagamentos de juros	97.088761	96.163566	0.93	$1.80	$2.20	$2.00

Figura 5.4 – Tabela de sensibilidade

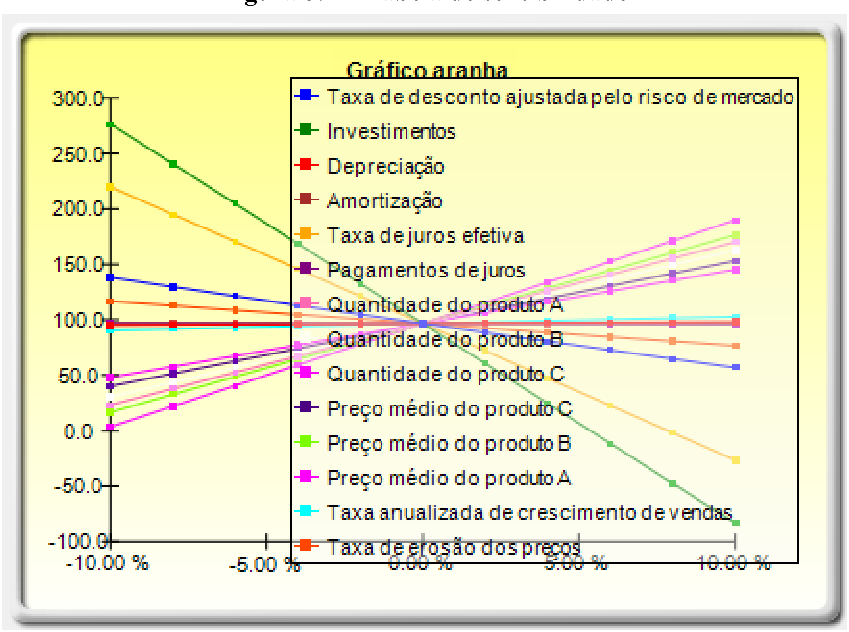


Figura 5.5 – Gráfico aranha

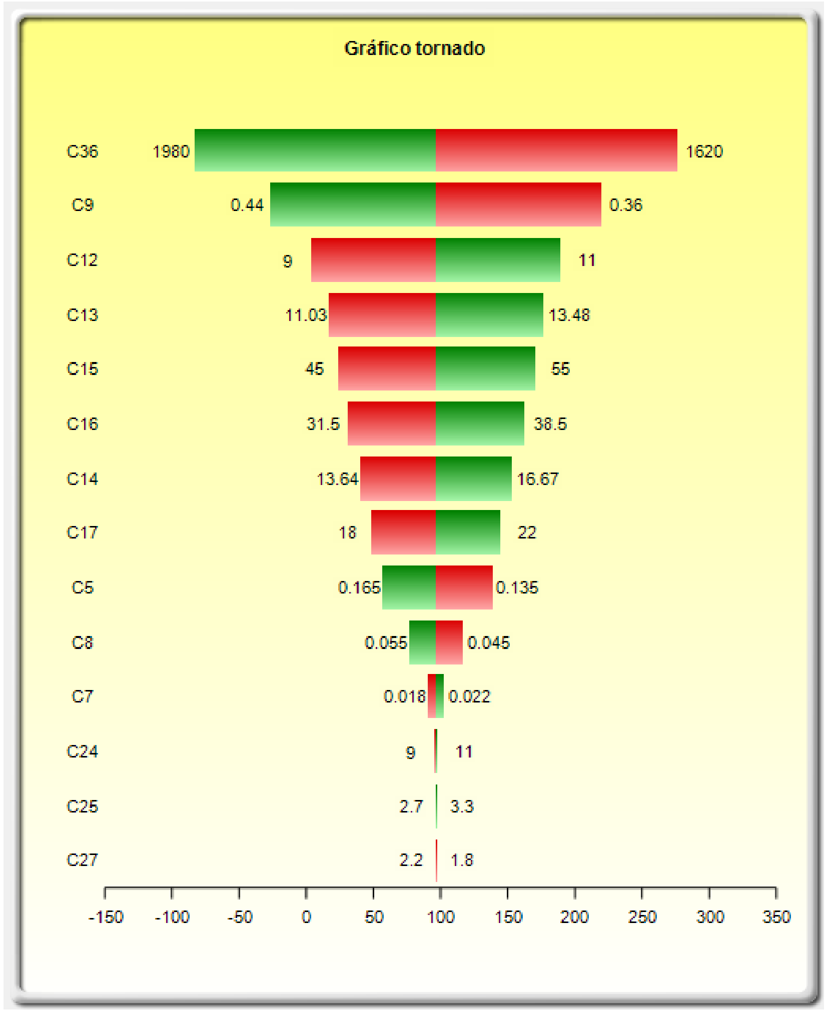


Figura 5.6 – Gráfico tornado

Embora o gráfico tornado seja de mais fácil compreensão, o gráfico "aranha" é importante para determinar se existem não linearidades no modelo. Por exemplo, a Figura 5.7 mostra outro gráfico "aranha" no qual as não linearidades são bastante evidentes (as linhas no gráfico não são retas, mas curvas). O modelo usado é o ***Gráficos de sensibilidade e tornado (não lineares)***, que usa o modelo de precificação de opção de Black-Scholes como um exemplo. Tais não linearidades não podem ser verificadas em um gráfico tornado e podem ser informações importantes no modelo ou fornecer detalhes importantes sobre a dinâmica do modelo para as pessoas responsáveis pela tomada de decisões.

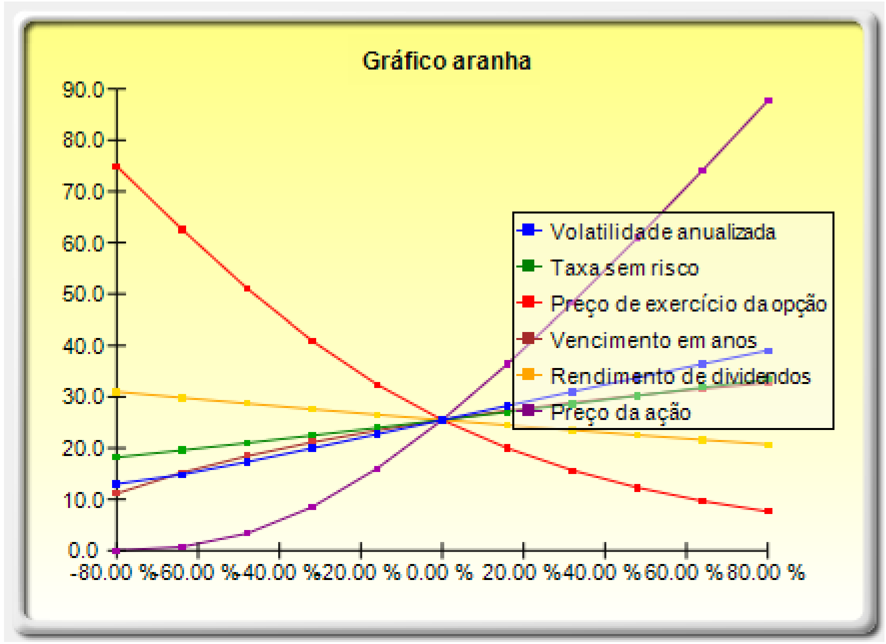


Figura 5.7 – Gráfico aranha não linear

Notas adicionais sobre tornado:

A Figura 5.2 mostra a interface da ferramenta de análise tornado. Observe que existem alguns novos aprimoramentos a partir do Risk Simulator versão 4 e posterior. Veja a seguir algumas dicas para executar a análise tornado e detalhes adicionais sobre esses novos aprimoramentos:

- A análise tornado nunca deve ser executada apenas uma vez. Ela é considerada uma ferramenta de diagnóstico de modelo, o que significa que deve ser executada várias vezes no mesmo modelo. Por exemplo, em um modelo grande, o tornado pode ser executado na primeira vez com todas as configurações padrão e mostrando todas as precedentes (selecione **Mostrar todas as variáveis**). Isso pode resultar em um relatório grande e gráficos tornado longos (e potencialmente confusos). Porém, isso fornece um ótimo ponto de partida para determinar a quantidade de precedentes consideradas fatores críticos de sucesso (por exemplo, o gráfico tornado pode mostrar que as primeiras 5 variáveis têm alto impacto sobre o resultado, ao passo que as 200 variáveis restantes exercem pouco ou nenhum impacto). Nesse caso, uma segunda análise tornado é executada com menos variáveis (selecione **Mostrar as 10 variáveis principais** se as primeiras 5 forem críticas, o que cria bons relatórios e gráfico tornado, que mostra um contraste entre os fatores-chave e os fatores menos críticos, ou seja, você nunca deve mostrar um gráfico tornado com apenas as variáveis-chave sem mostrar algumas variáveis menos críticas, para que haja assim um contraste sobre os efeito no resultado). Por fim, os pontos de teste padrão podem ser aumentados de ±10% para algum valor maior para testar não linearidades (o gráfico “aranha” mostrará linhas não lineares e os gráficos tornado serão inclinados para um dos lados se os efeitos dos precedentes forem não lineares).
- A opção **Usar endereço de célula** é uma boa idéia se o seu modelo for grande, o que permite identificar a localização (o nome da planilha e o endereço da célula) de uma

célula precedente. Se ela não for selecionada, o software aplicará sua própria *lógica fuzzy* para tentar determinar o nome de cada variável precedente (às vezes os nomes podem parecer confusos em um modelo grande com variáveis repetidas ou podem ser muito extensos, dando uma aparência desagradável ao gráfico tornado).

- As opções **Analisar esta planilha** e **Analisar todas as planilhas** permitem controlar se as precedentes devem fazer parte apenas da planilha atual ou se todas as planilhas devem ser incluídas na mesma pasta de trabalho. Essa opção é útil quando você quer apenas analisar um resultado com base nos valores da planilha atual em vez de executar uma pesquisa global de todas as precedentes vinculadas entre várias planilhas na mesma pasta de trabalho.
- A opção **Usar configurações globais** é útil quando você tem um modelo grande e deseja testar todas as precedentes em ±50%, em vez do padrão de 10%. Em vez de alterar os valores de teste de cada precedente, um por vez, você pode selecionar essa opção, alterar a configuração e *clicar em qualquer lugar* na interface do usuário, para alterar toda a lista de precedentes. Desmarcar essa opção permite controlar a alteração dos pontos de teste de uma precedente por vez.
- A opção **Ignorar zeros ou valores vazios** é uma opção ativada por padrão, na qual células precedentes com valor zero ou vazio não serão executadas no tornado. Essa é a configuração típica.
- A opção **Realçar valores inteiros possíveis** permite identificar rapidamente todas as células precedentes possíveis que têm, no momento, entradas de números inteiros. Às vezes isso pode ser importante se o seu modelo usar desvios (por exemplo, funções como SE uma célula for 1 algo acontecerá e SE a célula contiver valor 0 outra coisa acontecerá, ou inteiros como 1, 2, 3 etc. que você não deseja testar). Por exemplo, ±10% do valor igual a 1 de um desvio de sinalizador retornará os valores de teste 0,9 e 1,1, ambos valores de entrada irrelevantes e incorretos no modelo, o que pode ser interpretado pelo Excel como erro da função. Quando selecionada, essa opção destaca rapidamente possíveis áreas problemáticas para a análise tornado. Você pode determinar quais precedentes ativar ou desativar manualmente ou pode usar **Ignorar valores inteiros possíveis** para desativar todos simultaneamente.

Análise de sensibilidade

Teoria:

Uma função relacionada é a análise de sensibilidade. Enquanto a análise tornado (gráficos tornado e aranha) aplica perturbações estáticas *antes* que uma simulação seja executada, a análise de sensibilidade aplica perturbações dinâmicas criadas *após* a execução da simulação. Os gráficos tornado e aranha são resultados de perturbações estáticas, o que significa que cada variável de suposição ou precedente é perturbada com um valor predefinido, uma por vez, e as flutuações nos resultados são tabuladas. Em contraste, os gráficos de sensibilidade são os resultados de perturbações dinâmicas já que várias suposições são perturbadas simultaneamente e suas interações no modelo e as correlações entre variáveis são capturadas nas flutuações dos resultados. Os gráficos tornado identificam quais variáveis controlam mais os resultados e que são, assim, adequadas para a simulação, enquanto os gráficos de sensibilidade identificam o impacto nos resultados quando múltiplas variáveis que interagem são simuladas no modelo. Esse efeito é claramente ilustrado na Figura 5.8. Observe que a classificação dos fatores de sucesso são semelhantes ao gráfico tornado dos exemplos anteriores. No entanto, quando são acrescentadas correlações entre as suposições, a Figura 5.9 mostra uma imagem bem diferente. Observe, por exemplo, que a erosão de preços causou pouco impacto no VPL, mas quando algumas das suposições de entrada estão correlacionadas, a interação existente entre essas variáveis correlacionadas faz com que a erosão de preços tenha um impacto maior.

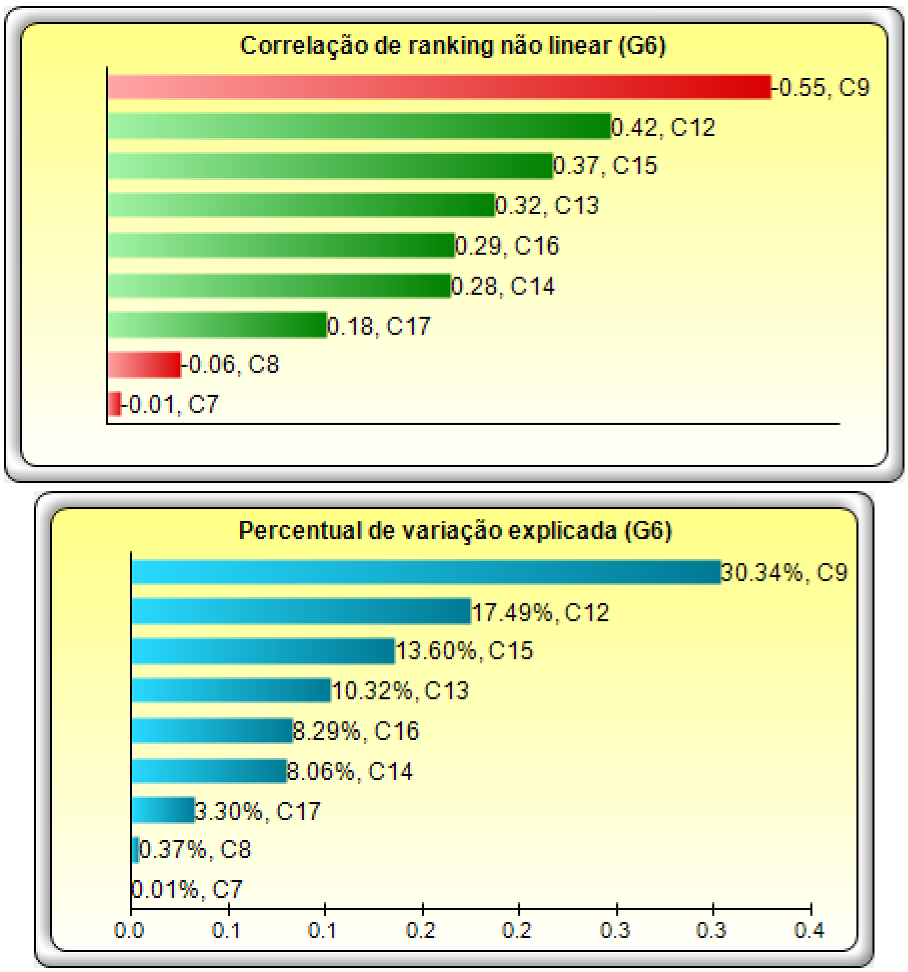


Figura 5.8 – Gráfico de sensibilidade sem correlações

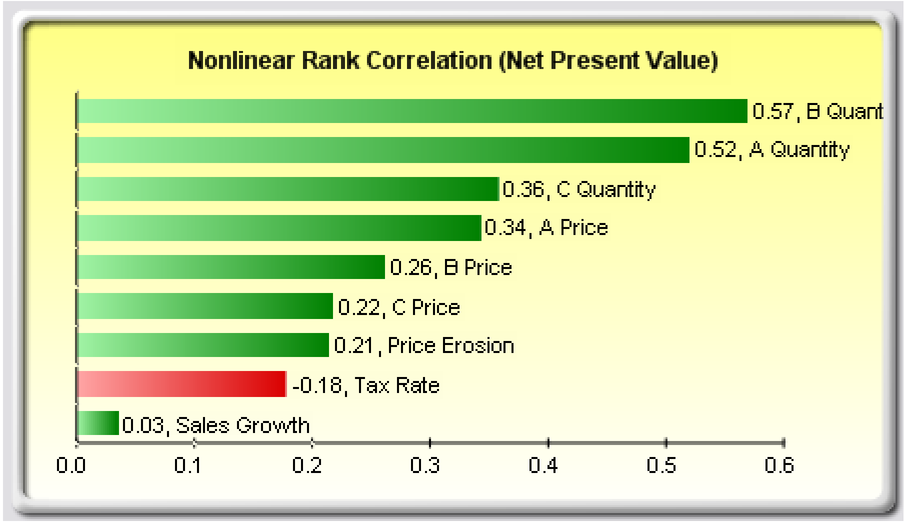


Figura 5.9 – Gráfico de sensibilidade com correlações

Procedimento:

- Abra ou crie um modelo, defina suposições e previsões e execute a simulação (o exemplo em questão usa o arquivo *Gráficos de sensibilidade e tornado (lineares)*
- Selecione ***Risk Simulator* | *Ferramentas* | *Análise de sensibilidade***
- Selecione a previsão que deseja analisar e clique em OK (Figura 5.10)

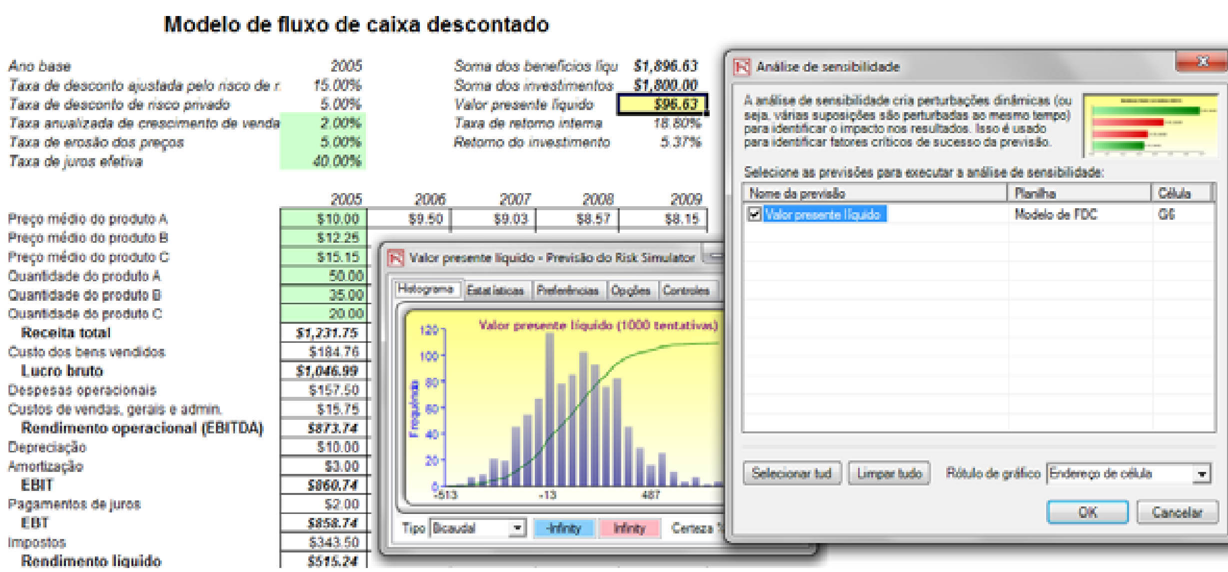


Figura 5.10 – Execução da análise de sensibilidade

Interpretação dos resultados:

Os resultados da análise de sensibilidade compõe um relatório e dois gráficos-chave. O primeiro é um gráfico de correlação de classificação não linear (Figura 5.11) que classifica os pares de correlação suposição-previsão do mais alto ao mais baixo. Essas correlações são não lineares e não paramétricas, ou seja, elas não têm requisitos de distribuição (por exemplo, uma suposição com uma distribuição Weibull pode ser comparada a outra com uma distribuição beta). Os resultados desse gráfico são semelhantes àqueles da análise tornado descrita anteriormente, porém, sem o valor de investimento de capital, que decidimos ser um valor conhecido e, portanto, não foi simulado, com uma exceção especial. A alíquota de impostos taxa de juros foi relegada a uma posição muito mais baixa no gráfico da análise de sensibilidade (Figura 5.11), se comparada ao gráfico tornado (Figura 5.6). A razão disso é que a alíquota de impostos, por si só, tem um impacto significativo, mas depois que outras variáveis interajam no modelo, a alíquota aparenta ter um efeito dominante menor. Isso acontece porque a alíquota de impostos tem uma distribuição menor enquanto as alíquotas de impostos históricas tendem a não flutuar muito, e também porque a alíquota de impostos é um valor percentual fixo da renda antes dos impostos, na qual outras variáveis precedentes têm maior efeito). Esse exemplo prova que a execução de uma análise de sensibilidade após uma simulação é importante para verificar se há interações no modelo e se os efeitos de determinadas variáveis ainda persistem. O segundo gráfico (Figura 5.12) ilustra a variação percentual explicada. Ou seja, o quanto da variação das flutuações na previsão pode ser explicado por cada uma das suposições, após explicar todas as interações entre as variáveis? Observe que a soma de todas as variações explicadas é, em geral, próxima de 100% (às vezes há outros elementos que causam impacto no modelo, mas não podem ser capturados diretamente aqui), e se houver correlação, a soma poderá, às vezes, exceder 100% (devido aos efeitos da interação que são cumulativos).

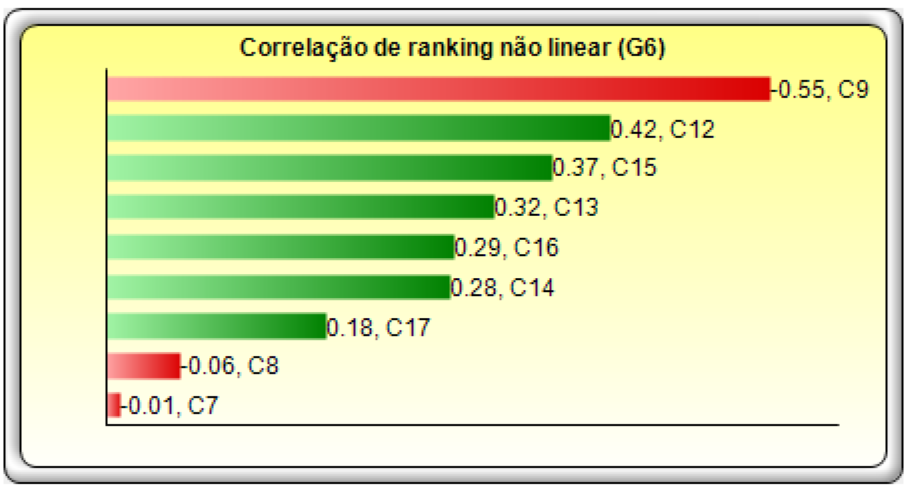


Figura 5.11 – Gráfico de correlação de ranking

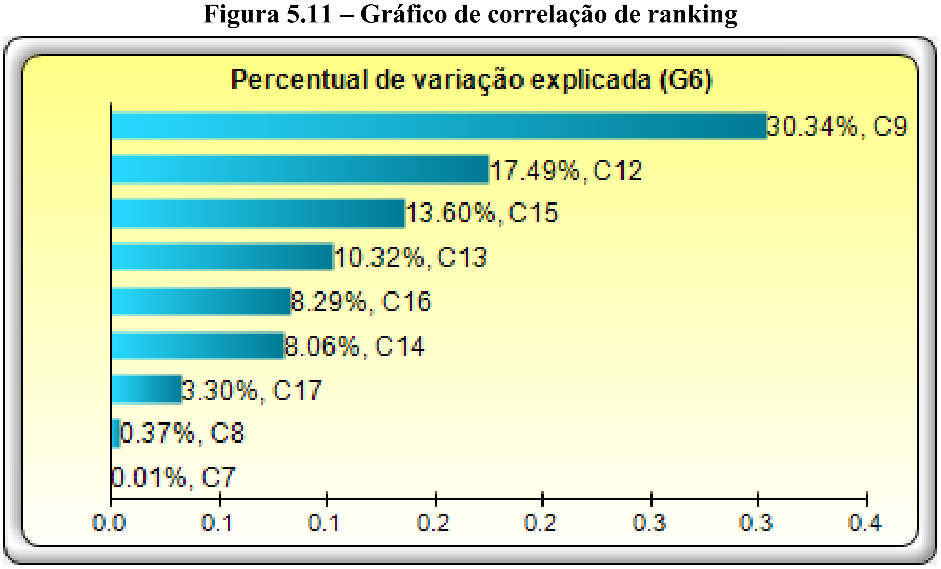


Figura 5.12 – Gráfico de contribuição à variância

Notas:

A análise tornado é executada antes de uma simulação, enquanto a análise de sensibilidade é executada após a simulação. Os gráficos aranha na análise tornado podem considerar não linearidades, enquanto os gráficos de correlação de ranking na análise de sensibilidade podem explicar condições não lineares e de distribuição livre.

Ajuste da distribuição: variável única e múltiplas variáveis

Teoria:

Uma outra ferramenta de simulação poderosa é o ajuste da distribuição. Ou seja, qual distribuição um analista usa para uma determinada variável de entrada em um modelo? Quais são os parâmetros de distribuição relevantes? Se não existirem dados de históricos, o analista deverá então fazer suposições sobre as variáveis em questão. Uma abordagem é usar o método Delphi quando um grupo de especialistas está incumbido de estimar o comportamento de cada variável. Por exemplo, um grupo de engenheiros mecânicos pode ser responsável por avaliar as possibilidades extremas do diâmetro de uma bobina por meio de experiências rigorosas ou palpites. Esses valores podem ser usados como parâmetros de entrada da variável (por exemplo, a distribuição uniforme com valores extremos entre 0,5 e 1,2). Quando não for possível testar (por exemplo, participação no mercado e taxa de crescimento da renda), a gerência ainda poderá fazer estimativas dos resultados possíveis e fornecer os cenários de melhor caso, caso mais provável e o pior caso.

No entanto, se dados de históricos confiáveis estiverem disponíveis, o ajuste da distribuição poderá ser realizado. Supondo que os padrões históricos se mantêm e que a história tende a se repetir, os dados históricos podem então ser usados para encontrar a distribuição de melhor ajuste com seus parâmetros relevantes para definir melhor as variáveis a serem simuladas. As Figuras 5.13 a 5.15 ilustram um exemplo de ajuste da distribuição. A ilustração usa o arquivo ***Ajuste de dados*** na pasta de exemplos.

Procedimento:

- Abra uma planilha com os dados existentes para ajuste
- Selecione os dados que deseja ajustar (os dados devem estar em uma única coluna com várias linhas)
- Selecione ***Risk Simulator | Ferramentas | Ajuste da distribuição (variável única)***
- Selecione as distribuições específicas que deseja ajustar ou mantenha o padrão, no qual todas as distribuições são selecionadas, e clique em OK (Figura 5.13)
- Analise os resultados do ajuste, escolha a distribuição relevante desejada e clique em OK (Figura 5.14)

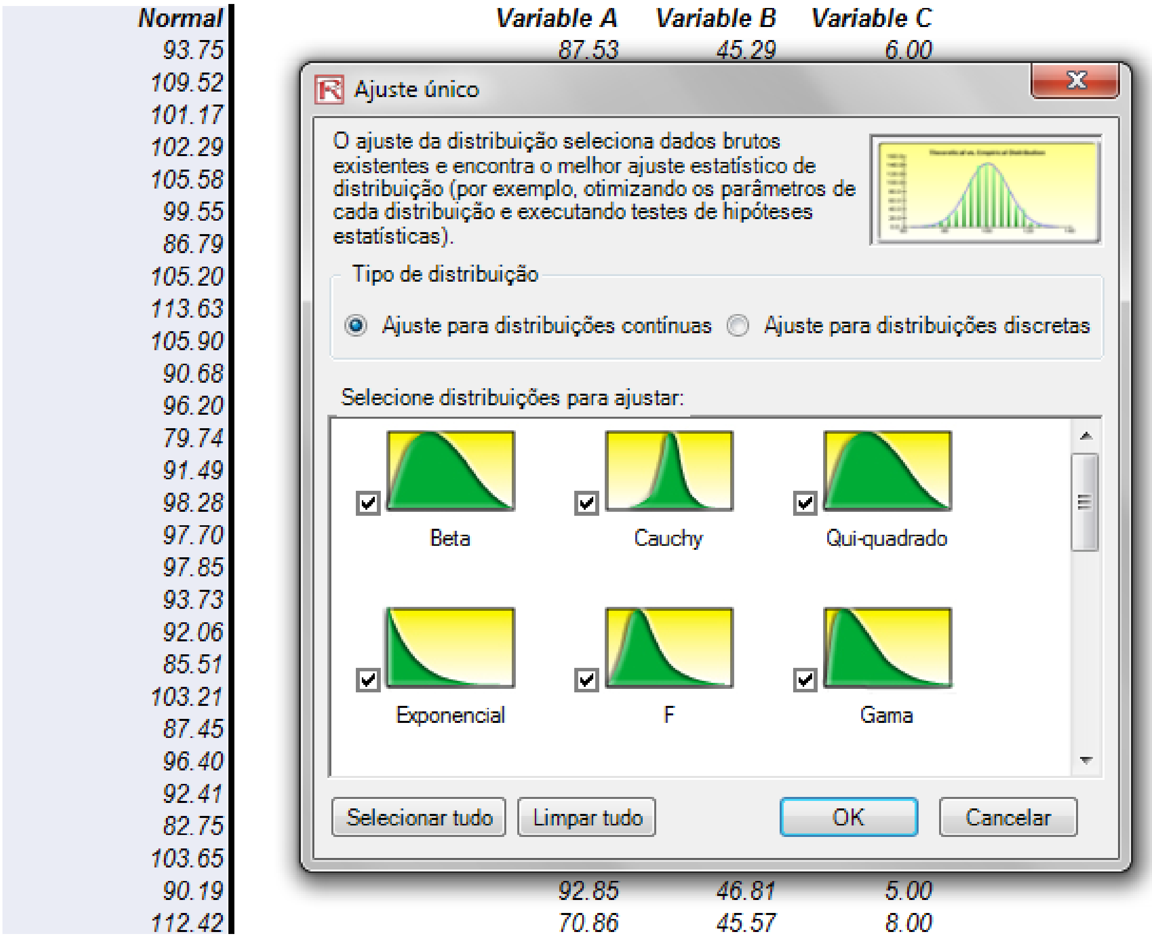
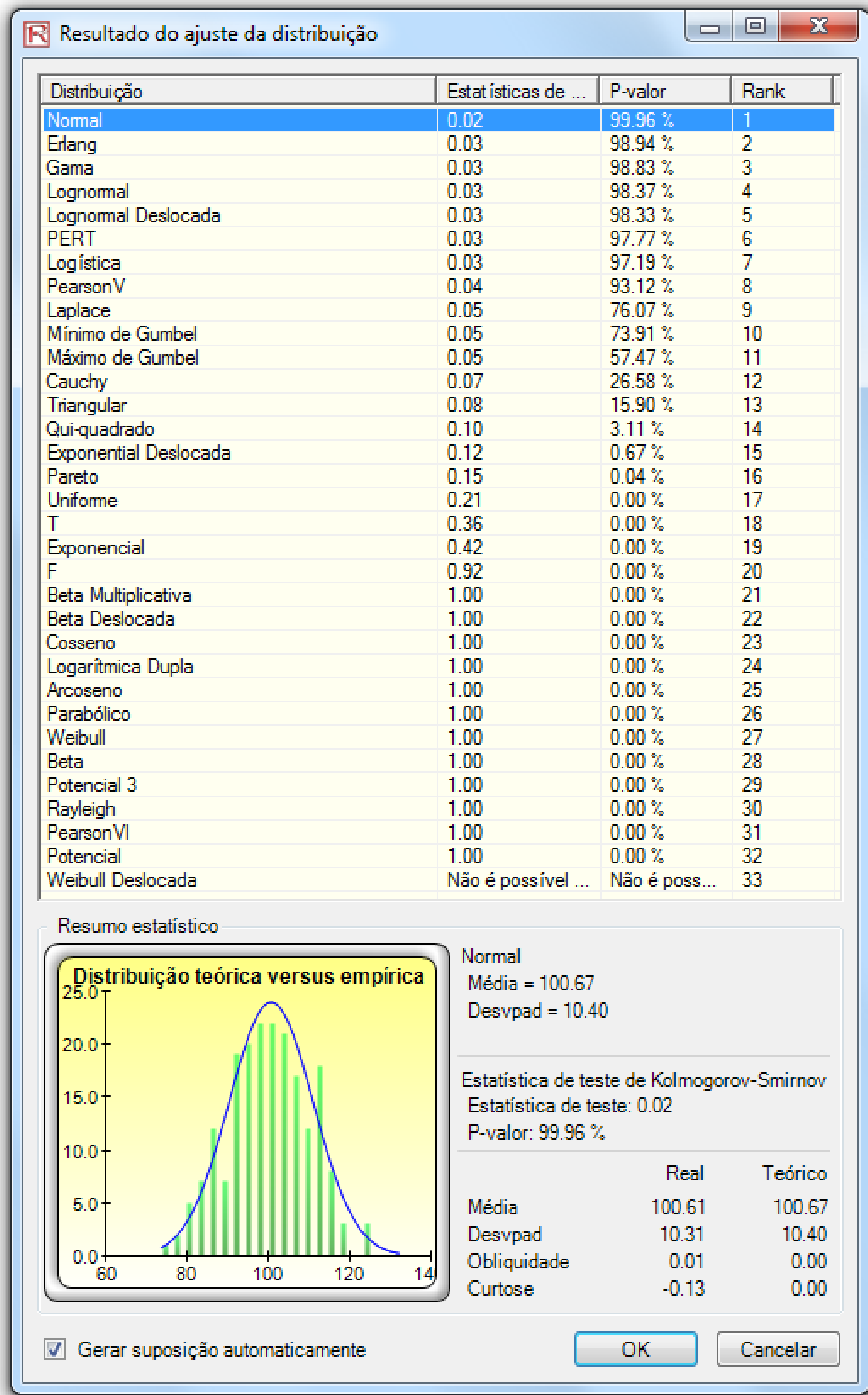


Figura 5.13 – Ajuste da distribuição de variável única

Interpretação dos resultados:

A hipótese nula sendo testada é tal que a distribuição ajustada é a mesma distribuição da população de origem dos dados da amostra. Portanto, se um p-valor calculado for menor que um nível alfa crítico (normalmente 0,10 ou 0,05), a distribuição será a incorreta. Inversamente, quanto maior o p-valor, melhor a distribuição se ajustará aos dados. É possível pensar em um p-valor como uma *porcentagem explicada*, ou seja, se o p-valor for 0,9727 (Figura 5.14), então a configuração de uma distribuição normal com uma média de 99,28 e um desvio padrão de 10,17 explicará cerca de 97,27% da variação nos dados, indicando um bom ajuste. Os dois resultados (Figura 5.14) e o relatório (Figura 5.15) mostram a estatística do teste, o p-valor, as estatísticas teóricas (baseadas na distribuição selecionada), as estatísticas empíricas (baseadas nos dados brutos), os dados originais (para manter um registro dos dados usados) e a suposição com os parâmetros de distribuição relevantes (por exemplo, se você selecionou a opção para gerar automaticamente suposições e se um perfil de simulação já existe). Os resultados também classificam as distribuições selecionadas e como elas ajustam os dados.



Distribuição	Estatísticas de ...	P-valor	Rank
Normal	0.02	99.96 %	1
Erlang	0.03	98.94 %	2
Gama	0.03	98.83 %	3
Lognormal	0.03	98.37 %	4
Lognormal Deslocada	0.03	98.33 %	5
PERT	0.03	97.77 %	6
Logística	0.03	97.19 %	7
PearsonV	0.04	93.12 %	8
Laplace	0.05	76.07 %	9
Mínimo de Gumbel	0.05	73.91 %	10
Máximo de Gumbel	0.05	57.47 %	11
Cauchy	0.07	26.58 %	12
Triangular	0.08	15.90 %	13
Qui-quadrado	0.10	3.11 %	14
Exponential Deslocada	0.12	0.67 %	15
Pareto	0.15	0.04 %	16
Uniforme	0.21	0.00 %	17
T	0.36	0.00 %	18
Exponencial	0.42	0.00 %	19
F	0.92	0.00 %	20
Beta Multiplicativa	1.00	0.00 %	21
Beta Deslocada	1.00	0.00 %	22
Cosseno	1.00	0.00 %	23
Logarítmica Dupla	1.00	0.00 %	24
Arcoseno	1.00	0.00 %	25
Parabólico	1.00	0.00 %	26
Weibull	1.00	0.00 %	27
Beta	1.00	0.00 %	28
Potencial 3	1.00	0.00 %	29
Rayleigh	1.00	0.00 %	30
PearsonVI	1.00	0.00 %	31
Potencial	1.00	0.00 %	32
Weibull Deslocada	Não é possível ...	Não é poss...	33

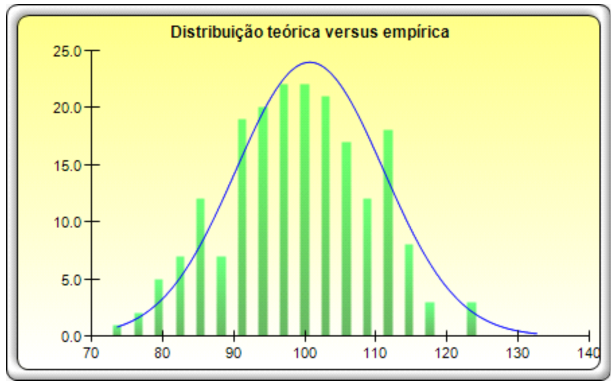
	Real	Teórico
Média	100.61	100.67
Desvpad	10.31	10.40
Obliquidade	0.01	0.00
Curtose	-0.13	0.00

Figura 5.14: Resultado do ajuste da distribuição

Ajuste da distribuição de variável única

Resumo estatístico

Suposição ajustada		100.61
Distribuição ajustada		**Normal**
Média		100.67
Sigma		10.40
Estatística de Kolmogorov-Smirnov		0.02
P-valor para estatística de teste		0.9996
	Real	Teórico
Média	100.61	100.67
Desvio padrão	10.31	10.40
Obliquidade	0.01	0.00
Curtose excessiva	-0.13	0.00



Dados ajustados originais

73.53	78.21	78.52	79.50	79.72	79.74	81.56	82.08	82.68	82.75	83.34	83.64	84.09
84.66	85.00	85.35	85.51	86.04	86.79	86.82	86.91	87.02	87.03	87.45	87.53	87.66
88.05	88.45	88.51	89.95	90.19	90.54	90.68	90.96	91.25	91.49	91.56	91.94	92.06
92.36	92.41	92.45	92.70	92.80	92.84	93.21	93.26	93.48	93.73	93.75	93.77	93.82
94.00	94.15	94.51	94.57	94.64	94.69	94.95	95.57	95.62	95.71	95.78	95.83	95.97
96.20	96.24	96.40	96.43	96.47	96.81	96.88	97.00	97.07	97.21	97.23	97.48	97.70
97.77	97.85	98.15	98.17	98.24	98.28	98.32	98.33	98.35	98.65	99.03	99.27	99.46
99.47	99.55	99.73	99.96	100.08	100.24	100.36	100.42	100.44	100.48	100.49	100.83	101.17
101.28	101.34	101.45	101.46	101.55	101.73	101.74	101.81	102.29	102.55	102.58	102.60	102.70
103.17	103.21	103.22	103.32	103.34	103.45	103.65	103.66	103.72	103.81	103.90	103.99	104.46
104.57	104.76	105.20	105.44	105.50	105.52	105.58	105.66	105.87	105.90	105.90	106.29	106.35
106.59	107.01	107.68	107.70	107.93	108.17	108.20	108.34	108.42	108.43	108.49	108.70	109.15
109.22	109.35	109.52	109.75	110.04	110.16	110.25	110.54	111.05	111.06	111.44	111.76	111.90
111.95	112.07	112.19	112.29	112.32	112.42	112.48	112.85	112.92	113.50	113.59	113.63	113.70
114.13	114.14	114.21	114.91	114.95	115.40	115.58	115.66	116.58	116.98	117.60	118.67	119.24
119.52	124.14	124.16	124.39	132.30								

Figura 5.15: Relatório do ajuste da distribuição

Para ajustar múltiplas variáveis, o processo é muito semelhante ao ajuste das variáveis individuais. No entanto, os dados devem ser organizados em colunas (por exemplo, cada variável é organizada em uma coluna) e todas as variáveis são ajustadas, uma por vez.

Procedimento:

- Abra uma planilha com os dados existentes para ajuste
- Selecione os dados que deseja ajustar (os dados devem estar em várias colunas com várias linhas)
- Selecione ***Risk Simulator | Ferramentas | Ajuste da distribuição (múltiplas variáveis)***
- Revise os dados, escolha os tipos de distribuição relevantes desejados e clique em OK

Notas:

Observe que os métodos de ranking estatístico usados nas rotinas de ajuste da distribuição são os testes qui-quadrado e de Kolmogorov-Smirnov. O primeiro é usado para testar distribuições discretas e o último, para distribuições contínuas. Em resumo, usa-se um teste de hipóteses com uma rotina de otimização interna para encontrar os parâmetros que melhor se ajustam a cada distribuição testada. Os resultados são classificados do melhor para o pior ajuste.

Simulação de bootstrap

Teoria:

A simulação de *bootstrap* é uma técnica simples que estima a confiabilidade ou precisão das estatísticas da previsão ou de outra amostra de dados brutos. Essencialmente, a simulação de *bootstrap* é usada no teste de hipóteses. Métodos clássicos usados no passado contavam com fórmulas matemáticas para descrever a precisão das estatísticas da amostra. Esses métodos supõem que a distribuição de uma estatística da amostra usa uma distribuição normal, tornando o cálculo do erro padrão ou do intervalo de confiança da estatística relativamente fácil. No entanto, quando a distribuição amostral de uma estatística não for distribuída normalmente ou encontrada com facilidade, esses métodos clássicos se tornarão difíceis para uso ou inválidos. Em contraste, o *bootstrap* analisa as estatísticas de amostra de forma empírica, tomando amostras repetidamente dos dados e criando distribuições das estatísticas diferentes para cada amostragem.

Procedimento:

- Execute uma simulação
- Selecione ***Risk Simulator | Ferramentas | Bootstrap não paramétrico***
- Selecione apenas uma previsão para realizar *bootstrap*, selecione as estatísticas nas quais deseja realizar *bootstrap*, insira o número de tentativas de *bootstrap* e clique em OK (Figura 5.16)

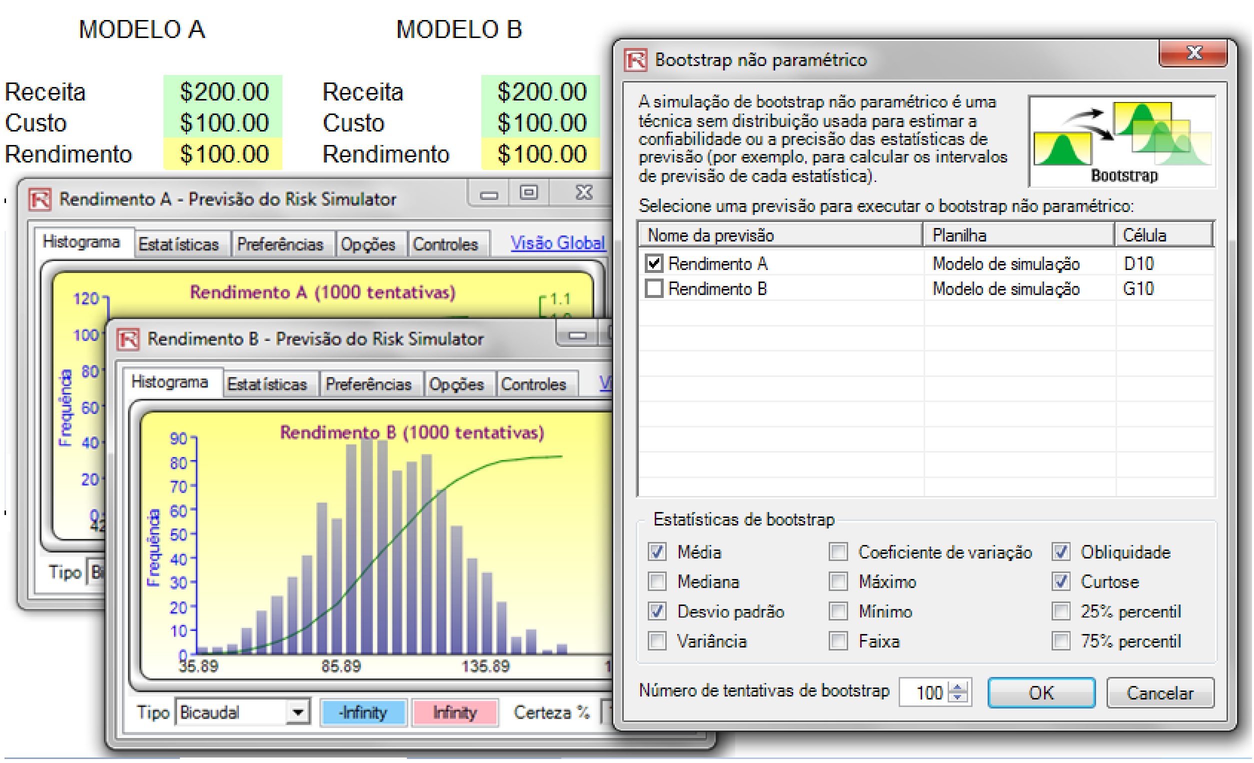


Figura 5.16 – Simulação de *bootstrap* não paramétrico

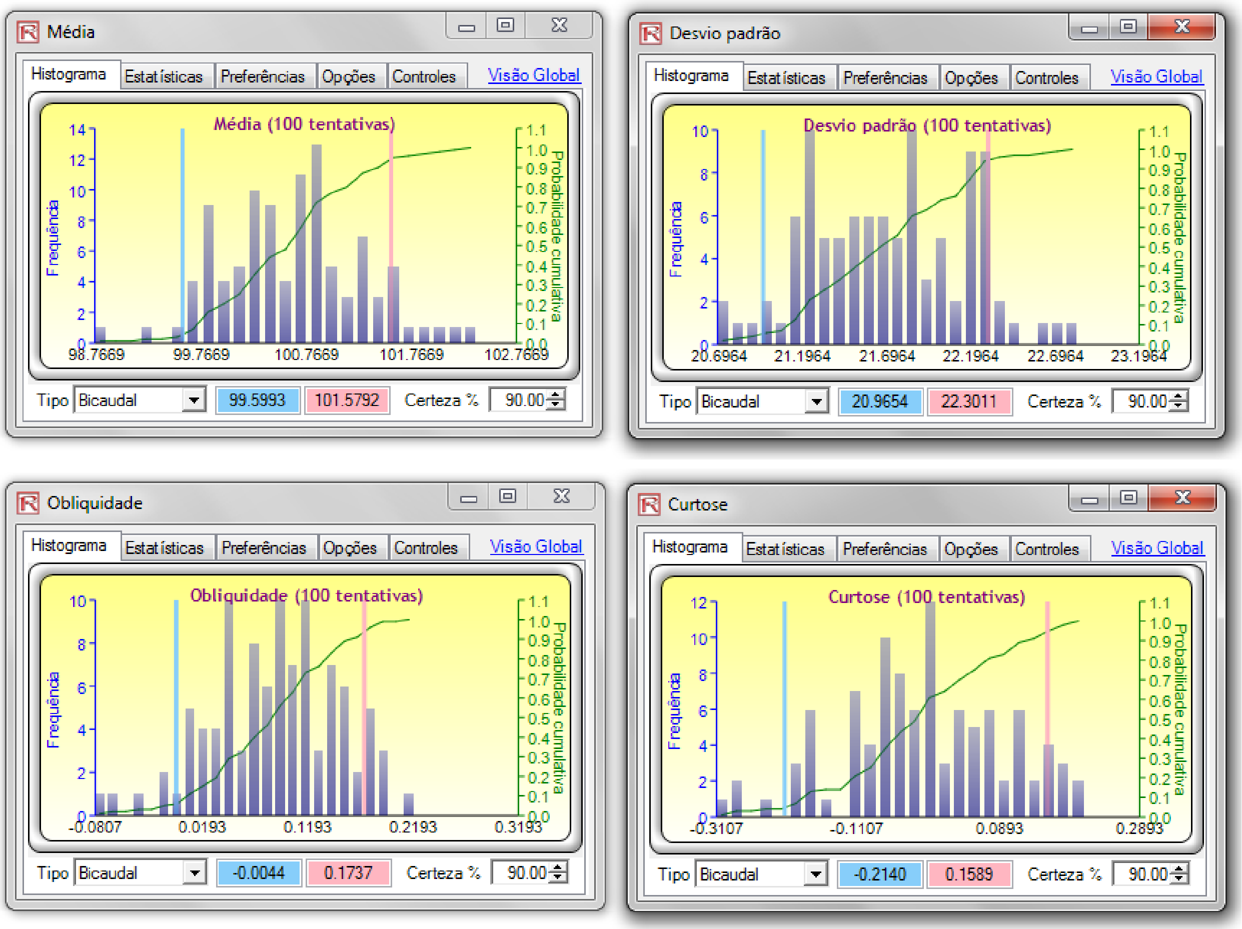


Figura 5.17 – Resultados da simulação de bootstrap

Interpretação dos resultados:

Em essência, a simulação de *bootstrap* não paramétrico pode ser definida como uma *simulação baseada em uma simulação*. Após executar uma simulação, as estatísticas resultantes são exibidas, mas a precisão de tais estatísticas e sua significância estatística podem ser às vezes questionáveis. Por exemplo, se a estatística de obliquidade de uma execução de simulação for –0,10, essa distribuição será de fato negativamente inclinada ou o valor ligeiramente negativo pode ser atribuído a uma chance aleatória? E quanto a –0,15 ou –0,20 etc. Ou seja, qual deve ser a distância para que a distribuição seja considerada negativamente inclinada? A mesma pergunta pode ser aplicada a todas as outras estatísticas. Uma distribuição é estatisticamente idêntica a outra com relação a algumas estatísticas calculadas ou são significativamente diferentes? A Figura 5.17 mostra alguns resultados do *bootstrap* de exemplo. Por exemplo, a estatística de obliquidade de 90 por cento de confiança está entre –0,0189 e 0,0952, de tal forma que o valor 0 está nessa faixa de confiança, indicando que com 90 por cento de confiança, a obliquidade dessa previsão não é estatisticamente muito diferente de zero ou que a distribuição pode ser considerada simétrica e não oblíqua. Inversamente, se o valor 0 está fora da faixa de confiança, o contrário será verdadeiro: a distribuição será oblíqua (inclinada positivamente se a estatística de previsão for positiva e inclinada negativamente se a estatística da previsão for negativa).

Notas:

O termo *bootstrap* significa usar a distribuição das estatísticas propriamente ditas para analisar a precisão delas. Uma simulação não paramétrica significa simplesmente escolher bolas de golfe de forma aleatória em uma grande cesta com reposição, na qual cada bola é baseada em um ponto de dados históricos. Suponha que há 365 bolas de golfe em uma cesta (representando 365 pontos de dados históricos). Imagine que o valor de cada bola de golfe escolhida por acaso está gravado em um grande quadro branco. Os resultados das 365 bolas escolhidas com reposição estão gravados na primeira coluna do quadro, com 365 linhas de números. As estatísticas relevantes (média, mediana, desvio padrão etc.) são calculadas sobre essas 365 linhas. O processo é repetido, por exemplo, cinco mil vezes. O quadro branco será agora preenchido com 365 linhas e 5.000 colunas. Assim, 5.000 conjuntos de estatísticas (ou seja, 5.000 médias, 5.000 medianas, 5.000 desvios padrões etc.) são tabulados e suas distribuições são mostradas. Em seguida, as *estatísticas das estatísticas* relevantes são tabuladas e a partir desses resultados é possível verificar o nível de confiança em que as estatísticas simuladas se encontram. Em outras palavras, em uma simples simulação de 10.000 tentativas, a média de previsão resultante é estimada em $5,00. O analista pode confiar nesses resultados? O *bootstrap* permite que o usuário verifique o intervalo de confiança da estatística média calculada, indicando a distribuição das estatísticas. Por fim, os resultados de *bootstrap* são importantes porque, de acordo com a *Lei dos números grandes* e o *Teorema do limite central* em estatísticas, a média das médias da amostra significa um estimador não-polarizado e se aproxima da média da população verdadeira quando o tamanho da amostra aumenta.

Teste de hipóteses

Teoria:

Um teste de hipóteses é executado durante o teste das médias e das variâncias de duas distribuições para determinar se elas são estatisticamente idênticas ou diferentes uma da outra. Ou seja, para ver se as diferenças entre as médias e as variâncias de duas previsões diferentes ocorridas são baseadas em chances aleatórias ou se são de fato estatisticamente diferentes uma da outra.

Procedimento:

- Execute uma simulação
- Selecione ***Risk Simulator | Ferramentas | Teste de hipóteses***
- Selecione apenas *duas* previsões para testar por vez, selecione o tipo de teste de hipóteses que deseja executar e clique em OK (Figura 5.18)

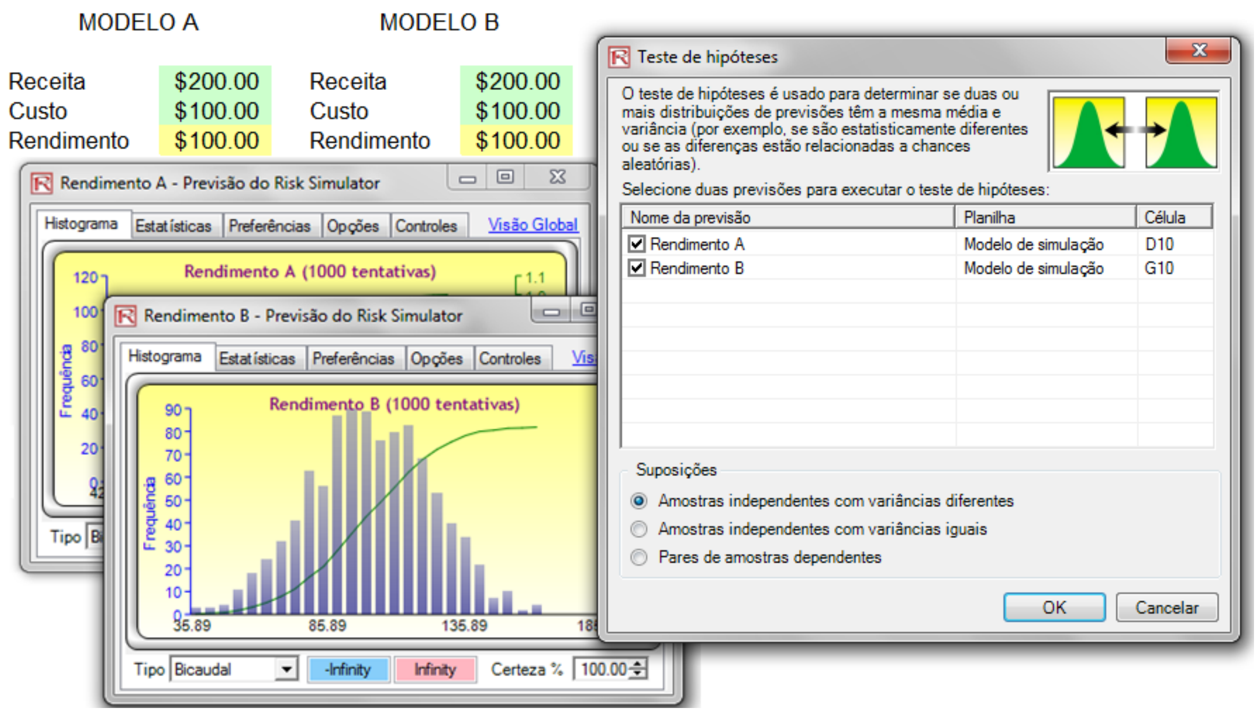


Figura 5.18 – Teste de hipóteses

Interpretação do relatório:

Um teste de hipóteses bicaudal é executado na hipótese nula H_o, de forma que as médias da população das duas variáveis sejam estatisticamente idênticas. A hipótese alternativa (H_a) é que as médias da população sejam diferentes entre si. Se os p-valores calculados são menores ou iguais a 0,01, 0,05 ou 0,10, isso significa que a hipótese nula é rejeitada, o que quer dizer que as médias da previsão são estatisticamente muito diferentes nos níveis de significância 1%, 5% e 10%. Se a hipótese nula não é rejeitada quando os p-valores são altos, as médias das duas distribuições de previsão são estatisticamente semelhantes. A mesma análise é executada em variâncias de duas previsões ao mesmo tempo usando o teste-F de paridade. Se os p-valores são baixos, as variâncias (e os desvios padrão) são estatisticamente diferentes; caso contrário, para p-valores altos, as variâncias são estatisticamente idênticas entre si.

Teste de hipóteses sobre as médias e as variâncias de duas previsões

Resumo estatístico

Um teste de hipóteses é executado durante o teste das médias e das variâncias de duas distribuições para determinar se elas são estatisticamente idênticas ou diferentes uma da outra. Ou seja, para ver se as diferenças entre duas médias e duas variâncias ocorridas são baseadas em chances aleatórios ou se são de fato diferentes uma da outra. O teste-t de duas variáveis de variâncias diferentes (espera-se que a variância da população da previsão 1 seja diferente da variância da população da previsão 2) é apropriado quando as distribuições da previsão são de populações diferentes (por exemplo, dados coletados de dois locais geográficos diferentes, duas unidades de negócios distintas e assim por diante). O teste-t de duas variáveis de variâncias iguais (espera-se que a variância da população da previsão 1 seja igual à variância da população da previsão 2) é apropriado quando as distribuições da previsão são de populações semelhantes (por exemplo, dados coletados de dois motores diferentes com especificações parecidas e assim por diante). O teste-t de duas variáveis dependentes emparelhadas é apropriado quando as distribuições de previsão são de populações semelhantes (por exemplo, dados coletados do mesmo grupo de clientes, mas em ocasiões diferentes e assim por diante).

Um teste de hipóteses bicaudal é executado na hipótese nula Ho, de forma que as médias da população das duas variáveis sejam estatisticamente idênticas. A hipótese alternativa é que as médias da população sejam diferentes entre si. Se os p-valores calculados são menores ou iguais a 0,01, 0,05 ou 0,10, isso significa que a hipótese é rejeitada, o que quer dizer que as médias da previsão são estatisticamente muito diferentes nos níveis de significância 1%, 5% e 10%. Se a hipótese nula não é rejeitada quando os p-valores são altos, as médias das duas distribuições de previsão são estatisticamente semelhantes. A mesma análise é executada em variâncias de duas previsões ao mesmo tempo usando o teste-F de paridade. Se os p-valores são baixos, as variâncias (e os desvios padrão) são estatisticamente diferentes; caso contrário, para p-valores altos, as variâncias são estatisticamente idênticas entre si.

Resultado

Suposição do teste de hipóteses:	Variâncias diferentes:
Estatística t calculada:	1.38424
P-valor para estatística t:	0.16644
Estatística F calculada:	1.089657
P-valor para estatística F:	0.174974

Figura 5.19 – Resultados do teste de hipóteses

Notas:

O teste-t de duas variáveis de variâncias diferentes (espera-se que a variância da população da previsão 1 seja diferente da variância da população da previsão 2) é apropriado quando as distribuições da previsão são de populações diferentes (por exemplo, dados coletados de dois locais geográficos diferentes, duas unidades de negócios distintas e assim por diante). O teste-t de duas variáveis de variâncias iguais (espera-se que a variância da população da previsão 1 seja igual à variância da população da previsão 2) é apropriado quando as distribuições da previsão são de populações semelhantes (por exemplo, dados coletados de dois motores diferentes com especificações parecidas e assim por diante). O teste t de duas variáveis dependentes emparelhadas é apropriado quando as distribuições de previsão são da mesma população (por exemplo, dados coletados do mesmo grupo de clientes, mas em ocasiões diferentes e assim por diante).

Extrair dados e salvar resultados da simulação

Os dados brutos de uma simulação podem ser extraídos com facilidade usando a rotina de *Extração de dados* do Risk Simulator. É possível extrair suposições e previsões, mas antes é necessário executar uma simulação. Os dados extraídos poderão então ser usados para uma variedade de outras análises.

Procedimento:

- Abra ou crie um modelo, defina suposições e previsões e execute a simulação
- Selecione ***Risk Simulator | Ferramentas | Extração de dados***
- Selecione as suposições e/ou previsões das quais deseja extrair os dados e clique em OK

Os dados podem ser extraídos em vários formatos:

- Dados brutos em uma nova planilha, na qual os valores simulados (suposições e previsões) podem ser salvos ou analisados posteriormente caso necessário
- Arquivo de texto simples, que permitie exportar os dados para outro software de análise de dados
- Arquivo do Risk Simulator em que os resultados (suposições e previsões) possam ser recuperados posteriormente selecionando ***Risk Simulator* | *Ferramentas* | *Abrir/importar dados***.

A opção mais comum é a terceira opção. Ela salva os resultados simulados como um arquivo *.risksim, do qual os resultados podem ser recuperados. Não é necessário executar novamente uma simulação toda vez. A Figura 5.21 mostra a caixa de diálogo para extrair ou exportar e salvar os resultados da simulação.

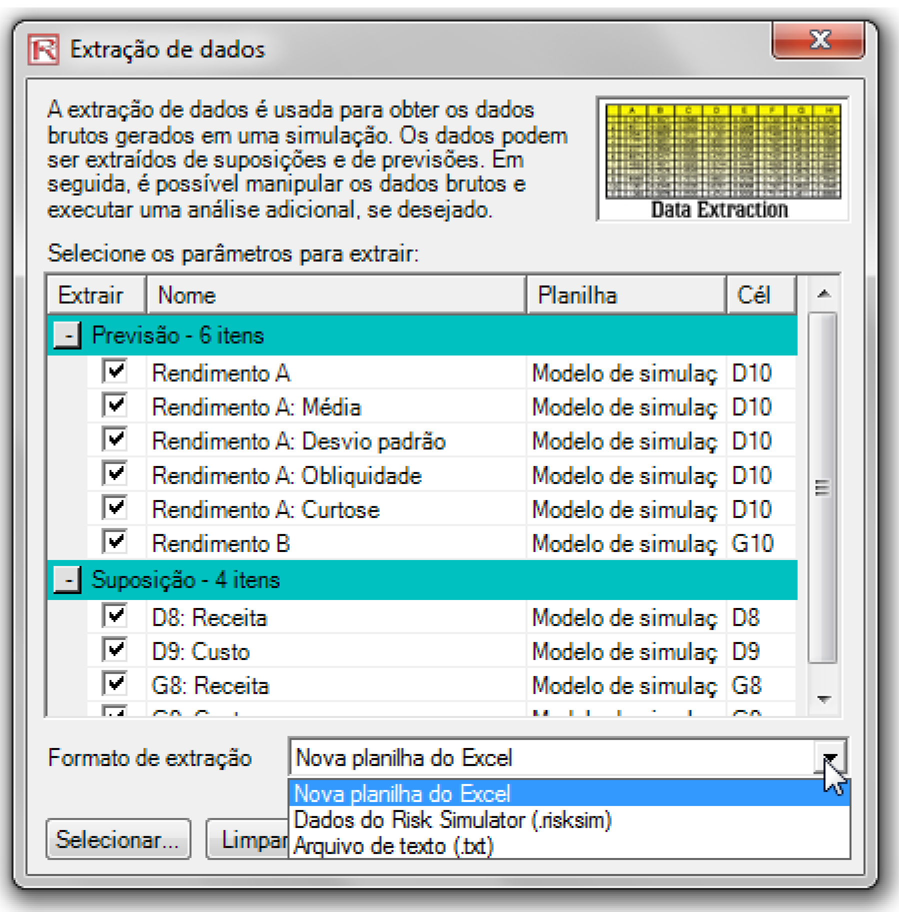


Figura 5.21 – Relatório de simulação da amostra

Criar relatório

Depois de executar uma simulação, você pode gerar um relatório das suposições, previsões e dos resultados obtidos durante a execução da simulação.

Procedimento:

- Abra ou crie um modelo, defina suposições e previsões e execute a simulação
- Selecione ***Risk Simulator* | *Criar relatório***

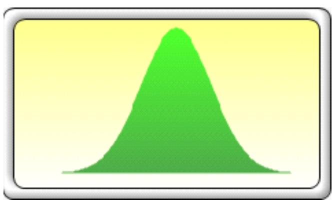
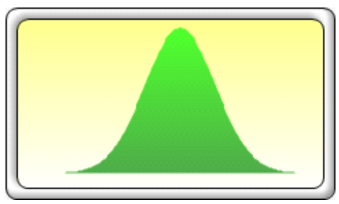
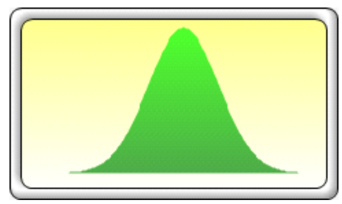
Simulação - Modelo de simulação

Geral

Número de tentativas	1000
Parar a simulação em cas	Não
Propagação aleatória	123456
Habilitar correlações	Sim

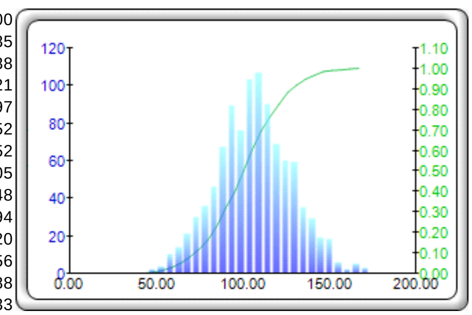
Suposições

Nome	**D8: Receita**	**Nome**	**D9: Custo**	**Nome**	**G8: Receita**
Habilitado	Sim	Habilitado	Sim	Habilitado	Sim
Célula	D8	Célula	D9	Célula	G8
Simulação dinâmica	Não	Simulação dinâmica	Não	Simulação dinâmica	Não
Faixa		**Faixa**		**Faixa**	
Mínimo	-Infinity	Mínimo	-Infinity	Mínimo	-Infinity
Máximo	Infinity	Máximo	Infinity	Máximo	Infinity
Distribuição	Normal	**Distribuição**	Normal	**Distribuição**	Normal
Média	200	Média	100	Média	200
Desvio padrão	20	Desvio padrão	10	Desvio padrão	20

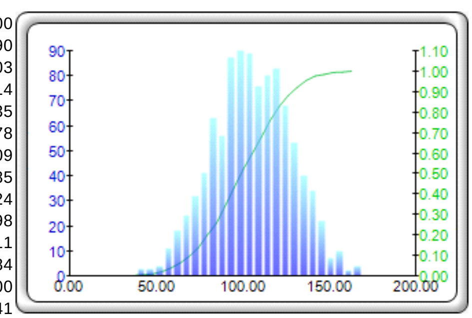




Previsões

Nome	**Rendimento A**	Número de pontos de dad	1000
Habilitado	Sim	Média	100.5485
Célula	D10	Mediana	100.4388
		Desvio padrão	21.6421
Precisão de previsão		Variância	468.3797
Nível de precisão	---	Coeficiente de variação	0.2152
Nível de erro	---	Máximo	167.0852
		Mínimo	38.8005
		Faixa	128.2848
		Obliquidade	0.0794
		Curtose	-0.0220
		25% percentil	86.4456
		75% percentil	114.7888
		Precisão de erro em 95%	0.0133



Nome	**Rendimento B**	Número de pontos de dad	1000
Habilitado	Sim	Média	99.1790
Célula	G10	Mediana	99.1303
		Desvio padrão	22.5914
Precisão de previsão		Variância	510.3735
Nível de precisão	---	Coeficiente de variação	0.2278
Nível de erro	---	Máximo	162.7509
		Mínimo	32.5085
		Faixa	130.2424
		Obliquidade	-0.0498
		Curtose	-0.2711
		25% percentil	84.0684
		75% percentil	114.8800
		Precisão de erro em 95%	0.0141



Matriz de correlação

	D8: Receita	D9: Custo	G8: Receita	G9: Custo
D8: Receita	1.00			
D9: Custo	0.00	1.00		
G8: Receita	0.00	0.00	1.00	
G9: Custo	0.00	0.00	0.00	1.00

Figura 5.21 – Relatório de simulação da amostra

Ferramenta de diagnóstico de previsão e regressão

Esta ferramenta analítica avançada do Risk Simulator é usada para determinar as propriedades econométricas dos dados. O diagnóstico inclui a verificação de heteroscedasticidade, não linearidade, observações discrepantes, erros de especificação, micronumerosidade, propriedades estacionárias e estocásticas, normalidade e esfericidade dos erros e multicolinearidade. Cada teste é descrito mais detalhadamente em seus respectivos relatórios no modelo.

Procedimento:

- Abra o modelo de exemplo (**Risk Simulator | Exemplos | Diagnóstico de regressão**) e, na planilha *Dados da série temporal*, selecione os dados incluindo os nomes das variáveis (células **C5:H55**).
- Clique em **Risk Simulator | Ferramentas | Ferramenta de diagnóstico**.
- Verifique os dados e, no menu suspenso, selecione *Variável dependente Y*. Clique em **OK** quando terminar (Figura 5.22).

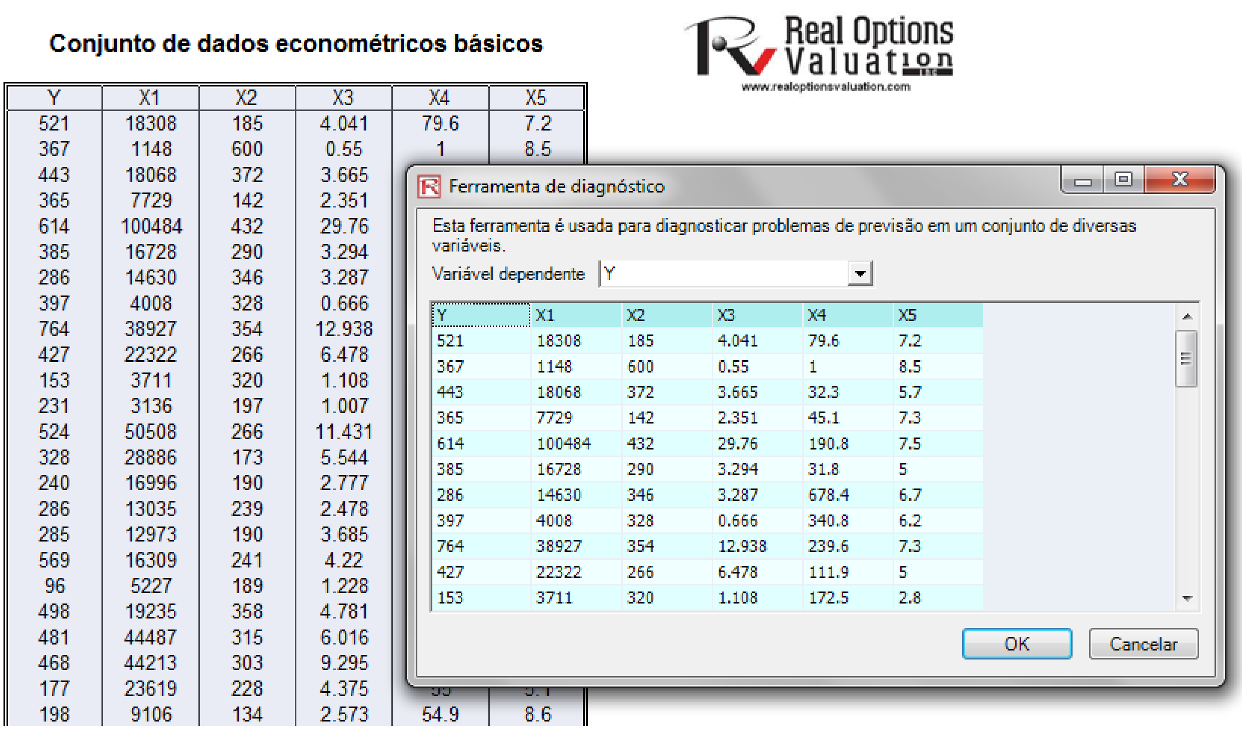


Figura 5.22 – Execução da ferramenta de diagnóstico de dados

Uma violação comum na análise de previsão e regressão é a heteroscedasticidade, ou seja, o aumento da variância dos erros ao longo do tempo (consulte a Figura 5.23 para obter os resultados do teste usando a ferramenta de diagnóstico). Visualmente, a largura das flutuações verticais dos dados aumenta ou se dispersa com o tempo e, normalmente, o coeficiente de determinação (coeficiente de R2) cai significantemente quando há heteroscedasticidade. Se a variância da variável dependente não for constante, a variância do erro não será constante. A menos que a heteroscedasticidade da variável dependente seja pronunciada, seu efeito não será grave: as estimativas dos mínimos quadrados continuaram a ser não polarizadas e as estimativas da inclinação e interceptação serão distribuídas normalmente, caso os erros sejam distribuídos normalmente, ou pelo menos distribuídas assintoticamente (à medida que o número de pontos de dados aumenta) caso os erros sejam normalmente distribuídos. A estimativa para a variância da inclinação e a variância geral será imprecisa, mas não haverá probabilidade de o erro ser substancial se os valores da variável independente forem simétricos em relação à sua média.

Se o número de dados for pequeno (micronumerosidade), pode ser difícil detectar violações na suposição. Em amostras pequenas, é difícil detectar violações de suposição, como não normalidade ou heteroscedasticidade de variâncias, até mesmo quando estão presentes. Com um número pequeno de pontos de dados, a regressão linear oferece menos proteção contra a violação de suposições. Com poucos pontos de dados, pode ser difícil determinar o nível de correspondência entre a linha ajustada e os dados ou se uma função não linear seria mais apropriada. Mesmo que nenhuma das suposições de teste seja violada, uma regressão linear com pequeno número de dados pode não ter capacidade suficiente para detectar uma diferença significativa entre a inclinação e zero, mesmo que a inclinação seja diferente de zero. A capacidade depende do erro residual, da variação observada na variável independente, da significância selecionada no nível alfa do teste e do número de pontos de dados. A capacidade diminui enquanto a variância residual aumenta, diminui enquanto o nível de significância é diminuído (por exemplo, à medida que o teste fica mais rigoroso), aumenta enquanto a variação na variável independente observada aumenta e aumenta à medida que o número de pontos de dados aumenta.

Os valores podem não ser igualmente distribuídos devido à presença de observações discrepantes. Essas observações discrepantes são valores anômalos nos dados. As discrepâncias podem ter uma forte influência sobre a inclinação e a interceptação ajustadas, dando um ajuste fraco à massa de pontos de dados. As discrepâncias tendem a aumentar a estimativa da variância residual, reduzindo a chance de rejeição da hipótese nula, por exemplo, criando um elevado número de erros de previsão. Elas podem ocorrer devido aos erros de registro, que podem ser corrigíveis, ou porque os valores da variável dependente não são todos retirados da mesma população. Discrepâncias aparentes podem ocorrer também porque os valores da variável dependente são da mesma população anormal. No entanto, um ponto pode ser um valor incomum, na variável independente ou na dependente, sem necessariamente ser uma observação discrepante no gráfico da dispersão. Na análise de regressão, a linha ajustada pode ser altamente sensível às observações discrepantes. Em outras palavras, a regressão de quadrados mínimos não é resistente às discrepâncias, assim, a estimativa de inclinação ajustada também não é. Um ponto distante verticalmente de outros pontos pode fazer com que a linha ajustada passe próximo a ele, em vez de seguir a tendência linear geral do resto dos dados, principalmente se o ponto for relativamente distante na direção horizontal do centro dos dados.

No entanto, é necessário ter muito cuidado ao decidir se as observações discrepantes devem ser removidas. Apesar de, na maioria dos casos, os resultados da regressão parecerem melhores quando as discrepâncias são removidas, é necessário haver antes uma justificativa. Por exemplo, se algo estiver regredindo o desempenho no retorno das ações de uma determinada firma, as discrepâncias causadas por quedas no mercado de ações devem ser incluídas, pois elas não são discrepâncias , mas fatos inevitáveis no ciclo de negócios. Antecipar essas discrepâncias e usar a equação de regressão para prever o fundo de aposentadoria de uma pessoa, com base nas ações, renderá, na melhor das hipóteses, resultados incorretos. Em contrapartida, suponha que as discrepâncias são causadas por uma única condição comercial não recorrente (por exemplo, fusão e aquisição) e não se prevê a repetição de tais alterações estruturais, assim, essas observações discrepantes devem ser removidas e os dados devem passar por uma limpeza antes da execução de uma análise de regressão. A análise aqui apenas identifica as discrepâncias, cabendo ao usuário determinar se elas devem permanecer, ou serem excluídas.

Às vezes, uma relação não linear entre as variáveis dependente e independente é mais apropriada do que uma relação linear. Nesses casos, executar uma regressão linear não será uma boa opção. Se o modelo linear não for a forma correta, as estimativas de inclinação e a interceptação e os valores ajustados na regressão linear serão polarizados, e as estimativas de inclinação e interceptação ajustadas não serão significativas. Acima de uma faixa restrita de variáveis dependentes ou independentes, os modelos não lineares podem ser bem aproximados por modelos lineares (isso é a base da interpolação linear), mas para uma previsão precisa é necessário selecionar um modelo apropriado para os dados. Uma transformação não linear deve ser aplicada primeiro aos dados antes de executar uma regressão. Uma abordagem simples é tomar o logaritmo natural da variável independente (outras abordagens incluem tomar a raiz quadrada ou elevar a variável independente à segunda ou à terceira potência) e executar uma regressão ou uma previsão usando os dados transformados não linearmente.

Resultados do diagnóstico

	Heteroscedasticidade		**Micronumerosidade**	**Observações discrepantes**			**Não linearidade**	
Variável	P-valor do teste W	Resultado do teste de hipóteses	Resultado da aproximação	Limite inferior natural	Limite superior natural	Número de discrepâncias potenciais	P-valor do teste não linear	Resultado do teste de hipóteses
Y			nenhum problema	-7.86	671.70	2		
X1	0.2543	Homoskedastic	nenhum problema	-21377.95	64713.03	3	0.2458	linear
X2	0.3371	Homoskedastic	nenhum problema	77.47	445.93	2	0.0335	nonlinear
X3	0.3649	Homoskedastic	nenhum problema	-5.77	15.69	3	0.0305	nonlinear
X4	0.3066	Homoskedastic	nenhum problema	-295.96	628.21	4	0.9298	linear
X5	0.2495	Homoskedastic	nenhum problema	3.35	9.38	3	0.2727	linear

Figura 5.23 – Resultados do teste de observações discrepantes, heteroscedasticidade, micronumerosidade e não linearidade

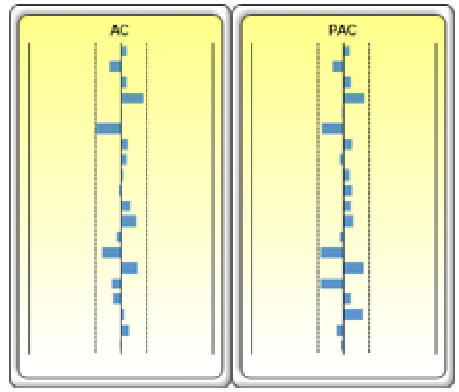
Outro problema comum na previsão de dados de série temporal é se os valores da variável independente são realmente independentes um do outro ou se são dependentes. Os valores de variável dependente coletados em uma série temporal podem estar autocorrelacionados. Para valores de variável dependente, serialmente correlacionados, as estimativas da inclinação e da interceptação não serão polarizadas, mas as estimativas da previsão e das variâncias não serão confiáveis e, assim, a validade de determinados testes de melhor ajuste estatístico será imperfeita. Por exemplo, as taxas de juros, taxas de inflação, vendas, receitas e outros dados de série temporal são tipicamente autocorrelacionados, já que o valor no período atual está relacionado a um valor no período anterior e assim por diante (a taxa de inflação de março está relacionada ao nível nível de fevereiro, que por sua vez, está relacionado ao nível de janeiro etc.). Se tais relações óbvias forem ignoradas, serão geradas previsões polarizadas e menos precisas. Nesses casos, um modelo de regressão autocorrelacionada ou um modelo ARIMA pode ser mais indicado (**Risk Simulator | Previsão | ARIMA**). Por fim, as funções de autocorrelação de uma série estacionária tendem a decair lentamente (consulte o relatório estacionário no modelo).

Se a autocorrelação AC(*1*) é diferente de zero, as séries são correlacionadas em série de primeira ordem. Se a AC(*k*) for extinta mais ou menos geometricamente com uma defasagem crescente, isso implica que a série segue um processo autorregressivo de ordem inferior. Se a AC(*k*) cair a zero depois de um pequeno número de defasagens, isso implica que a série segue um processo de média móvel de ordem inferior. A correlação parcial PAC(*k*) mede a correlação dos valores que estão separados por *k* períodos, depois de remover a correlação das defasagens intermédias. Se for possível capturar o padrão de autocorrelação por uma autorregressão de ordem menor do que *k*, então a autocorrelação parcial na defasagem k será próxima a zero. As estatísticas Q Ljung-Box e seus p-valores na defasagem *k* têm a hipótese nula de que não há autocorrelação até a ordem k. As linhas pontilhadas nos gráficos da autocorrelação são os dois limites de erro padrão aproximados. Se a autocorrelação estiver dentro desses limites, não será significativamente diferente de zero no nível de significância de 5%.

A autocorrelação mede a relação com o passado da variável dependente *Y* com ela mesma. As defasagens de distribuição, ao contrário, são relações de defasagem de tempo entre a variável dependente *Y* e as diferentes variáveis independentes *X*. Nos EUA, por exemplo, o movimento e a direção das taxas de hipoteca tendem a seguir a taxa do mercado interbancário, mas com uma defasagem de tempo (em geral de 1 a 3 meses). Às vezes, a defasagem de tempo obedece ciclos e sazonalidade (por exemplo, as vendas de sorvete tendem a atingir o pico durante os meses de verão e, sendo assim, estão relacionadas às vendas do verão passado, 12 meses atrás). A análise da defasagem da distribuição (Figura 5.24) mostra como a variável dependente está relacionada a cada uma das variáveis independentes em várias defasagens de tempo, quando todas as defasagens são consideradas simultaneamente, para determinar quais defasagens são estatisticamente significantes e devem ser consideradas.

Autocorrelação

Defasagem de tempo	AC	PAC	Limite inferior	Limite superior	Estatística Q	Prob
1	0.0580	0.0580	-0.2828	0.2828	0.1786	0.6726
2	-0.1213	-0.1251	-0.2828	0.2828	0.9754	0.6140
3	0.0590	0.0756	-0.2828	0.2828	1.1679	0.7607
4	0.2423	0.2232	-0.2828	0.2828	4.4865	0.3442
5	0.0067	-0.0078	-0.2828	0.2828	4.4890	0.4814
6	-0.2654	-0.2345	-0.2828	0.2828	8.6516	0.1941
7	0.0814	0.0939	-0.2828	0.2828	9.0524	0.2489
8	0.0634	-0.0442	-0.2828	0.2828	9.3012	0.3175
9	0.0204	0.0673	-0.2828	0.2828	9.3276	0.4076
10	-0.0190	0.0865	-0.2828	0.2828	9.3512	0.4991
11	0.1035	0.0790	-0.2828	0.2828	10.0648	0.5246
12	0.1658	0.0978	-0.2828	0.2828	11.9466	0.4500
13	-0.0524	-0.0430	-0.2828	0.2828	12.1394	0.5162
14	-0.2050	-0.2523	-0.2828	0.2828	15.1738	0.3664
15	0.1782	0.2089	-0.2828	0.2828	17.5315	0.2881
16	-0.1022	-0.2591	-0.2828	0.2828	18.3296	0.3050
17	-0.0861	0.0808	-0.2828	0.2828	18.9141	0.3335
18	0.0418	0.1987	-0.2828	0.2828	19.0559	0.3884
19	0.0869	-0.0821	-0.2828	0.2828	19.6894	0.4135
20	-0.0091	-0.0269	-0.2828	0.2828	19.6966	0.4770



Defasagens de distribuição

P-valores de períodos de defasagem de distribuição de cada variável independente

Variable	1	2	3	4	5	6	7	8	9	10	11	12
X1	0.8467	0.2045	0.3336	0.9105	0.9757	0.1020	0.9205	0.1267	0.5431	0.9110	0.7495	0.4016
X2	0.6077	0.9900	0.8422	0.2851	0.0638	0.0032	0.8007	0.1551	0.4823	0.1126	0.0519	0.4383
X3	0.7394	0.2396	0.2741	0.8372	0.9808	0.0464	0.8355	0.0545	0.6828	0.7354	0.5093	0.3500
X4	0.0061	0.6739	0.7932	0.7719	0.6748	0.8627	0.5586	0.9046	0.5726	0.6304	0.4812	0.5707
X5	0.1591	0.2032	0.4123	0.5599	0.6416	0.3447	0.9190	0.9740	0.5185	0.2856	0.1489	0.7794

Figura 5.24 – Resultados de autocorrelação e defasagem de distribuição

Outra exigência ao executar um modelo de regressão é a suposição de normalidade e a esfericidade do período de erro. Se a suposição de normalidade for violada ou observações discrepantes estiverem presentes, o teste de melhor ajuste da regressão linear pode não ser o mais poderoso ou informativo disponível, o que poderia significar a diferença entre detectar ou não um ajuste linear. Se os erros não forem independentes e não estiverem normalmente distribuídos, eles poderão indicar que os dados podem estar autocorrelacionados ou apresentar não linearidades ou outros erros mais destrutivos. A independência dos erros também pode ser detectada nos testes de heteroscedasticidade (Figura 5.25).

O teste de normalidade realizado nos erros é um teste não paramétrico que não faz suposições sobre a forma específica da população da qual a amostra é retirada, permitindo que conjuntos de dados amostrais menores sejam analisados. Esse teste avalia a hipótese nula de erros amostrais terem sido extraídos de uma população distribuída normalmente em comparação a uma hipótese alternativa de que a amostra de dados não é distribuída normalmente. Se a estatística D calculada for maior ou igual aos valores críticos D em vários valores de significância, rejeite a hipótese nula e aceite a hipótese alternativa (os erros não são distribuídos normalmente). Caso contrário, se a estatística D for menor que o valor crítico D, não rejeite a hipótese nula (os erros são distribuídos normalmente). Esse teste se baseia em duas frequências cumulativas: uma derivada do conjunto de dados da amostra e a outra, da distribuição teórica baseada na média e no desvio padrão dos dados da amostra.

Resultado do teste

Media dos erros de regressão	0.00
Desvio padrão dos erros	141.83
Estatística D	0.1036
D crítico em 1%	0.1138
D crítico em 5%	0.1225
D crítico em 10%	0.1458

Hipótese nula: Os erros estão distribuídos normalmente.

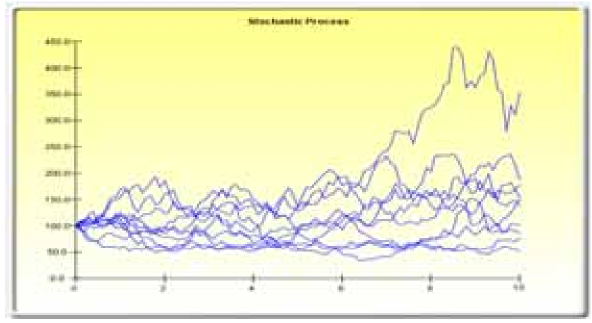
Conclusão: Os erros sãonormalmente distribuídos ao nível alfa de 1%.

Erros	Frequência relativa	Observado	Esperado	O-E
-219.04	0.02	0.02	0.0612	-0.0412
-202.53	0.02	0.04	0.0766	-0.0366
-186.04	0.02	0.06	0.0948	-0.0348
-174.17	0.02	0.08	0.1097	-0.0297
-162.13	0.02	0.10	0.1265	-0.0265
-161.62	0.02	0.12	0.1272	-0.0072
-160.39	0.02	0.14	0.1291	0.0109
-145.40	0.02	0.16	0.1526	0.0074
-138.92	0.02	0.18	0.1637	0.0163
-133.81	0.02	0.20	0.1727	0.0273
-120.76	0.02	0.22	0.1973	0.0227
-120.12	0.02	0.24	0.1985	0.0415
-113.25	0.02	0.26	0.2123	0.0477
-113.12	0.02	0.28	0.2125	0.0675
-97.53	0.02	0.30	0.2458	0.0542
-96.78	0.02	0.32	0.2475	0.0725
-79.46	0.02	0.34	0.2876	0.0524
-73.50	0.02	0.36	0.3021	0.0579
-68.87	0.02	0.38	0.3136	0.0664

Figura 5.25 – Teste para normalidade de erros

Ocasionalmente, certos tipos de dados de série temporal não podem ser modelados usando outros métodos exceto o processo estocástico, pois os eventos subjacentes são estocásticos por natureza. Por exemplo, não é possível prever e modelar, de forma adequada, preços de ações, taxas de juros, o preço do petróleo e o preço de outras commodities usando um modelo de regressão simples, porque essas variáveis são muito incertas e voláteis e não seguem uma regra estática predefinida de comportamento; em outras palavras, o processo não é estacionário. A propriedade estacionária é verificada aqui usando o teste de execução enquanto outra dica visual é encontrada no relatório de autocorrelação (o ACF tende a decair lentamente). Um processo estocástico é uma sequência de eventos ou caminhos gerados pelas leis de probabilidade. Ou seja, eventos aleatórios podem ocorrer ao longo do tempo, mas são regidos por leis estatísticas probabilísticas específicas. Os principais processos estocásticos incluem caminho aleatório ou movimento browniano, reversão à média e difusão com salto. Esses processos podem ser usados para prever diversas variáveis que parecem seguir tendências aleatórias, mas são restringidos pelas leis probabilísticas. A equação de geração do processo é conhecida antecipadamente, mas os resultados reais gerados são desconhecidos (Figura 5.26).

O processo de caminho aleatório ou movimento browniano pode ser usado na previsão de preços de ações, preços de commodities e outros dados de série temporal estocásticos, aos quais foram dados uma taxa de crescimento e uma volatilidade ao redor do caminho de crescimento. Os processos de reversão à média podem ser usados para reduzir as flutuações do processo de caminho aleatório permitindo que o caminho tenha por meta um valor de longo prazo, tornando-o útil na previsão de variáveis de série temporal que possuem uma taxa de longo prazo, como taxas de juros e de inflação, que são taxas de longo prazo usadas por órgãos reguladores ou pelo mercado). O processo de difusão com salto é útil para prever dados de série temporal quando a variável pode, ocasionalmente, apresentar saltos aleatórios, como os preços do petróleo ou da eletricidade (choques de evento exógeno discreto podem fazer preços subirem ou descerem drasticamente). É possível também fazer a mistura e a correspondência desses processos, quando necessário.



Resumo estatístico

Os parâmetros a seguir são estimados para um processo estocástico, dados os dados fornecidos. Cabe a você determinar se a probabilidade de ajuste (semelhante a um cálculo de melhor ajuste) é suficiente para garantir o uso de uma previsão de processo estocástico e, se assim for, se é um modelo de caminho aleatório, de reversão à média, de difusão com salto ou de combinações deles. Ao escolher o modelo de processo estocástico correto, você precisa confiar em suas experiências passadas e as expectativas econômicas e financeiras a priori que melhor representam o conjunto de dados subjacentes. Esses parâmetros podem ser inseridos em uma previsão de processo estocástico (**Risk Simulator** I **Previsão** I **Processos estocásticos**).

Periódico

Taxa de crescimento	*-1.48%*	*Taxa de reversão*	*283.89%*	*Taxa de salto*	0.204082
Volatilidade	*88.84%*	*Valor de longo prazo*	*327.72*	*Tamanho do salto*	*237.89*

Probabilidade de ajuste do modelo estocástico: 46.48%
um alto ajuste significa que um modelo estocástico é melhor que os modelos convencionais.

Execuções	*20*	*Normal padrão*	*-1.7321*
Positivo	*25*	*P-valor (unicaudal)*	*0.0416*
Negativo	*25*	*P-valor (bicaudal)*	*0.0833*
Execução esperada	*26*		

Um p-valor baixo (abaixo de 0,10; 0,05; 0,01) significa que a sequência não é aleatória e, portanto, apresenta problemas estacionários e um modelo ARIMA pode ser mais apropriado. Inversamente, p-valores mais altos indicam aleatoriedade e modelos de processo estocástico podem ser apropriados.

Figura 5.26 – Estimativa de parâmetro do processo estocástico

É necessário ter atenção neste ponto. A calibração dos parâmetros estocásticos mostra todos os parâmetros para todos os processos e não distingue qual processo é melhor, pior ou mais apropriado para uso. Essa determinação cabe ao usuário. Por exemplo, se há uma taxa de reversão de 283%, provavelmente um processo de reversão à média será inapropriado ou dada uma elevada taxa de salto de, digamos, 100%, isso significa que um processo de difusão com salto provavelmente não é apropriado e assim por diante. Além disso, a análise não pode determinar o que é a variável e qual é a fonte dos dados. Por exemplo, se os dados brutos são dos preços de ações históricos, são preços históricos de eletricidade, taxas de inflação ou movimento molecular de partículas subatômicas etc. Apenas o usuário poderia saber e, assim, usando um conhecimento e teoria prévios, seria capaz de escolher o processo correto para o uso (por exemplo, os preços de ações tendem a seguir o caminho aleatório do movimento browniano, ao passo que as taxas de inflação seguem um processo de reversão à média ou talvez um processo de difusão com salto seja mais apropriado caso você esteja prevendo o preço da eletricidade).

Existe multicolinearidade quando há uma relação linear entre as variáveis independentes. Quando isso acontece, a equação de regressão não pode ser estimada. Em situações de colinearidade próxima, a equação de regressão estimada será polarizada e fornecerá resultados imprecisos. Essa situação ocorre quando uma abordagem de regressão *stepwise* é usada, na qual as variáveis independentes significantes estatisticamente serão deixadas de fora da combinação de regressão muito antes do esperado, o que resulta em uma equação de regressão que não é eficiente nem precisa. Um teste rápido da presença de multicolinearidade em uma equação de regressão múltipla é que o valor de R2 é relativamente alto enquanto as estatísticas t são relativamente baixas.

Outro teste rápido é criar uma matriz de correlação entre as variáveis independentes. Uma correlação cruzada alta indica um potencial para autocorrelação. A regra geral é que a uma correlação com valor absoluto maior do que 0,75 indica forte multicolinearidade. Outro teste para a multicolinearidade é usar o fator de inflação de variância (VIF), obtido ao regressar cada variável independente a todas as outras variáveis independentes, obtendo o valor de R2 e calculando o VIF. Um VIF superior a 2,0 pode ser considerado uma forte multicolinearidade. Um VIF superior a 10,0 indica multicolinearidade destrutiva (Figura 5.27).

Matriz de correlação

CORRELAÇÃO	X2	X3	X4	X5
X1	0.333	0.959	0.242	0.237
X2	1.000	0.349	0.319	0.120
X3		1.000	0.196	0.227
X4			1.000	0.290
				1.000

Fator de inflação de variância

FIV	X2	X3	X4	X5
X1	1.12	**12.46**	1.06	1.06
X2	N/A	1.14	1.11	1.01
X3		N/A	1.04	1.05
X4			N/A	1.09
				N/A

Figura 5.27 – Erros de multicolinearidade

A matriz de correlação lista as correlações do momento do produto de Pearson (ou normalmente chamado de R de Pearson) entre pares de variáveis. O coeficiente de correlação varia entre –1,0 e + 1,0, inclusive. O sinal indica a direção de associação entre as variáveis enquanto o coeficiente indica a magnitude ou força da associação. O R de Pearson mede uma relação linear e é menos eficiente para medir relações não lineares. Para testar se as correlações são significantes, um teste de hipóteses bicaudal é executado e os p-valores resultantes são listados acima. Os p-valores menores que 0,10, 0,05 e 0,01 estão destacados em azul para indicar significância estatística. Em outras palavras, um p-valor para um par de correlação menor do que um valor de significância determinado é estatisticamente muito diferente de zero, indicando que há significância em uma relação linear entre as duas variáveis.

O coeficiente de correlação do momento do produto de Pearson (R) entre duas variáveis (x e y) está relacionado à medida de covariância (cov) onde $R_{x,y} = \frac{COV_{x,y}}{s_x s_y}$. O benefício ao dividir a covariância pelo produto do desvio padrão das duas variáveis (s) é que o coeficiente de correlação resultante fica limitado a um valor entre –1,0 e +1,0, inclusive. Isso torna a correlação uma boa medida relativa para fazer comparações entre variáveis diferentes (principalmente com magnitudes e unidades diferentes). A correlação não paramétrica baseada no ranking de Spearman também é incluída abaixo. O R de Spearman está relacionado ao R de Pearson, pois os dados primeiro são classificados e depois correlacionados. As correlações de ranking fornecem uma estimativa melhor da relação entre duas variáveis, quando uma ou as duas são não lineares.

É necessário destacar que uma correlação significante não sugere uma causa. As associações entre variáveis não sugerem de maneira alguma que a alteração de uma variável causa a alteração em outra. Quando duas variáveis que estão se movendo de forma independente uma da outra, mas em um caminho relacionado, elas podem estar correlacionadas, mas sua relação pode ser espúria (por exemplo, uma correlação entre manchas solares e o mercado de ações pode ser forte, mas alguém pode supor que não há causalidade e que esse relacionamento é puramente espúrio).

Ferramenta Análise estatística

Outra ferramenta muito poderosa do Risk Simulator é a ferramenta Análise estatística, que determina as propriedades estatísticas dos dados. Os diagnósticos executados incluem a verificação de várias propriedades estatísticas, desde as estatísticas de descrição básica, testes e até calibragem de propriedades estocásticas dos dados.

Procedimento:

- Abra o modelo de exemplo (**Risk Simulator | Exemplos | Análise estatística**) e, na planilha *Dados*, selecione os dados incluindo os nomes das variáveis (células **C5:E55**).
- Clique em **Risk Simulator | Ferramentas | Análise estatística** (Figura 5.28).
- Verifique o *tipo de dados*, se os dados selecionados são de uma variável única ou de múltiplas variáveis organizadas em linhas. No exemplo, supôs-se que as áreas de dados selecionadas são de múltiplas variáveis. Clique em **OK** quando terminar.
- *Escolha os testes estatísticos* que deseja realizar. A sugestão (e definição padrão) é escolher todos os testes. Clique em **OK** quando terminar (Figura 5.29).

Analise os relatórios gerados para entender melhor os testes estatísticos realizados (as Figuras 5.30 a 5.33 mostram exemplos de relatórios).

X1	X2	X3
18308	185	4.041
1148	600	0.55
18068	372	3.665
7729	142	2.351
100484	432	29.76
16728	290	3.294
14630	346	3.287
4008	328	0.666
38927	354	12.938
22322	266	6.478
3711	320	1.108
3136	197	1.007
50508	266	11.431
28886	173	5.544
16996	190	2.777
13035	239	2.478
12973	190	3.685
16309	241	4.22
5227	189	1.228
19235	358	4.781
44487	315	6.016
44213	303	9.295
23619	228	4.375
9106	134	2.573

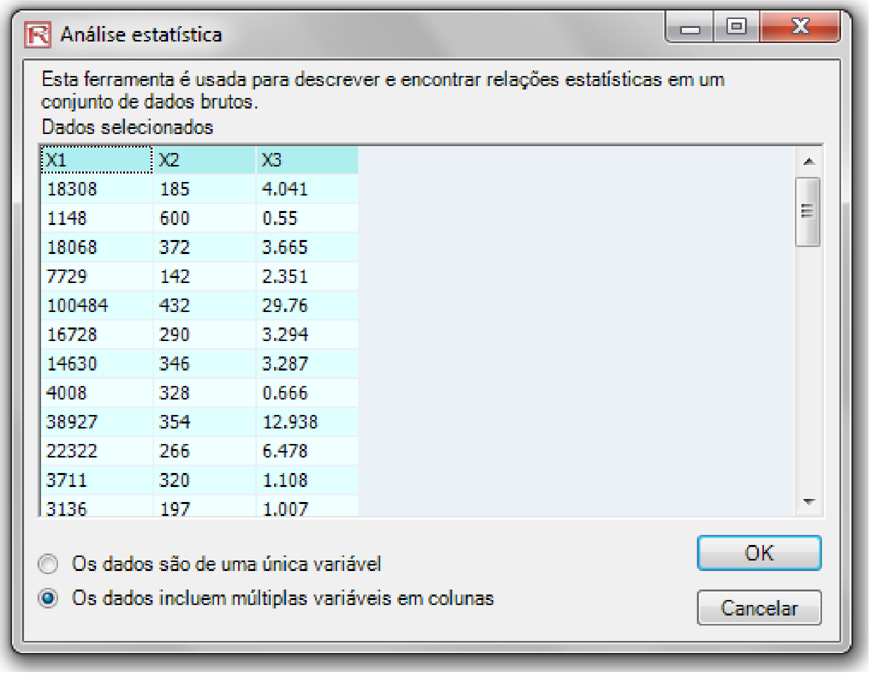


Figura 5.28 – Execução da ferramenta Análise estatística

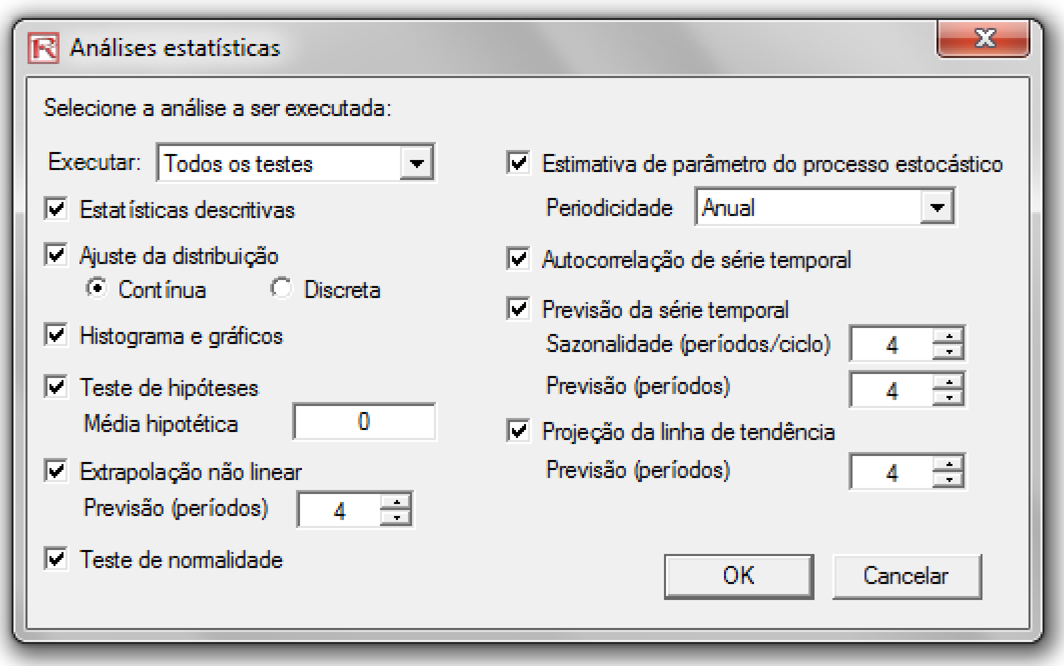


Figura 5.29 – Testes estatísticos

Estatísticas descritivas

Análise de estatísticas

Quase todas as distribuições podem ser descritas em quatro momentos (algumas distribuições exigem um momento, enquanto outras exigem dois e assim por diante). As estatísticas descritivas capturam quantitativamente esses momentos. O primeiro momento descreve a localização de uma distribuição (isto é, a média, a mediana e a moda) e é interpretado como o valor esperado, os retornos esperados ou o valor médio das ocorrências.

A média aritmética calcula a média de todas as ocorrências somando todos os pontos de dados e dividindo-os pelo número de pontos. A média geométrica é calculada obtendo a raiz da potência dos produtos de todos os pontos de dados e exige que todos sejam positivos. A média geométrica é mais precisa para porcentagens ou taxas que flutuam significativamente. Por exemplo, você pode usar a média geométrica para calcular a taxa média de crescimento, considerando-se juros compostos com taxas variáveis. A média truncada calcula a média aritmética do conjunto de dados depois que as observações discrepantes são removidas. Como as médias tendem a distorções significativas quando há observações discrepantes, a média truncada reduz tais erros em distribuições oblíquas.

O erro padrão da média calcula o erro que envolve a média da amostra. Quanto maior o tamanho da amostra, menor o erro, de maneira que para um tamanho de amostra infinitamente grande, o erro se aproxima de zero, indicando que o parâmetro da população foi estimado. Devido a erros de amostragem, o intervalo de confiança de 95% para a média é fornecido. De acordo com uma análise dos pontos de dados da amostra, a média real da população deve estar entre os intervalos inferior e superior da média.

A mediana é o ponto de dados no qual 50% de todos os pontos de dados se encontram acima desse valor e 50% abaixo dele. Entre as três estatísticas de primeiro momento, a mediana é a menos suscetível a observações discrepantes. Em uma distribuição simétrica, a mediana é igual à média aritmética. Há uma distribuição oblíqua quando a mediana se distancia da média. A moda mede o ponto de dados mais frequente.

O mínimo é o menor valor no conjunto de dados enquanto o máximo é o maior valor. O intervalo é a diferença entre os valores máximo e mínimo.

O segundo momento mede o spread ou a largura de uma distribuição e é frequentemente descrito por meio de medidas, como desvios padrão, variações, quartis e intervalos interquartílicos. O desvio padrão indica o desvio médio de todos os pontos de dados em relação à sua média. É uma medida popular, pois está associada ao risco (os desvios padrão mais altos significam uma distribuição mais ampla, risco mais alto ou dispersão mais ampla dos pontos de dados em torno da média) e suas unidades são idênticas ao conjunto de dados original. O desvio padrão da amostra é diferente do desvio padrão da população, pois o primeiro usa uma correção de grau de liberdade para contabilizar amostras pequenas. Além disso, os intervalos de confiança inferior e superior são fornecidos e o desvio padrão da população real se encontra nesse intervalo. Se o seu conjunto de dados cobre cada elemento da população, use o desvio padrão da população. As duas medidas de variância são simplesmente os valores quadrados dos desvios padrão.

O coeficiente de variabilidade é o desvio padrão da amostra dividido pela média da amostra, sendo uma medida de dispersão sem unidade que pode ser comparada em distribuições diferentes (agora você pode comparar distribuições de valores denominados em milhões de dólares com um valor em bilhões de dólares, ou metros e quilogramas etc.). O primeiro quartil mede o 25º percentil dos pontos de dados quando organizados do menor para o maior valor. O terceiro quartil é o valor do ponto de dados do 75º percentil. Ocasionalmente os quartis são usados como os intervalos superior e inferior de uma distribuição, pois truncam o conjunto de dados para ignorar discrepâncias. O intervalo interquartílico é a diferença entre o terceiro e o primeiro quartis, e é muito usado para medir a largura do centro de uma distribuição.

A obliquidade é o terceiro momento em uma distribuição. Ela caracteriza o grau de assimetria de uma distribuição em torno de sua média. Uma obliquidade positiva indica uma distribuição com uma cauda assimétrica que se estende em direção a valores mais positivos. Uma obliquidade negativa indica uma distribuição com uma cauda assimétrica que se estende em direção a valores mais negativos.

A curtose caracteriza o afunilamento ou o achatamento relativo de uma distribuição comparada à distribuição normal. Esse é o quarto momento em uma distribuição. Um valor de curtose positivo indica uma distribuição que relativamente afunilada. Uma curtose negativa indica uma distribuição relativamente achatada. A curtose aqui medida foi centrada em zero (outras medidas de curtose são centradas em torno de 3,0). Como ambas são igualmente válidas, a centralização em zero torna a interpretação mais simples. Uma curtose altamente positiva indica uma distribuição afunilada em torno de seu centro e caudas leptocúrticas ou gordas. Isso indica uma probabilidade maior de eventos extremos (por exemplo, eventos catastróficos, ataques terroristas, quebras do mercado de ações) do que é previsto em uma distribuição normal.

Estatísticas resumidas

Estatísticas	*X1*		
Observações	*50.0000*	*Desvio padrão (amostra)*	*21906.1026*
Média aritmética	*21667.5400*	*Desvio padrão (população)*	*21685.9352*
Média geométrica	*13627.8235*	*Intervalo de confiança inferior para desvio padrão*	*18826.9470*
Média truncada	*19292.9348*	*Intervalo de confiança superior para desvio padrão*	*26325.0585*
Erro padrão da média aritmética	*3097.9907*	*Variância (amostra)*	*479877330.9473*
Intervalo de confiança inferior para média	*15471.5585*	*Variância (população)*	*470279784.3284*
Intervalo de confiança superior para média	*27863.5215*	*Coeficiente de variabilidade*	*1.0110*
Mediana	*15469.5000*	*Primeiro quartil (Q1)*	*5242.0000*
Mínimo	*1148.0000*	*Terceiro quartil (Q3)*	*24917.0000*
Máximo	*100484.0000*	*Intervalo interquartílico*	*19675.0000*
Faixa	*99336.0000*	*Obliquidade*	*1.9773*
		Curtose	*4.1820*

Figura 5.30 – Relatório da ferramenta Análise estatística da amostra

Teste de hipóteses (teste t na média da população de uma variável)

Resumo estatístico

Estatísticas do conjunto de dados:		***Estatísticas calculadas:***	
Observações	*50*	*Estatística t*	*6.9941*
Média da amostra	*21667.54*	*P-valor (cauda direita)*	***0.0000***
Desvio padrão da amostra	*21906.10*	*P-valor (cauda esquerc*	***1.0000***
		P-valor (bicaudal)	***0.0000***
Estatísticas fornecidas pelo usuário:			
		Hipótese nula (Ho):	*μ = Média hipotética*
Média hipotética	*0.00*	*Hipótese alternativa (Ha):*	*μ < > Média hipotética*

Notas: "<>" indica "maior que" para cauda direita, "menor que" para cauda esquerda ou "diferente de" para testes de hipóteses bicaudais.

Figura 5.31 – Relatório da ferramenta Análise estatística da amostra (teste de hipóteses de uma variável)

Teste para normalidade

O teste de normalidade é um tipo de teste não paramétrico que não faz suposições sobre a forma específica da população da qual a amostra é retirada, permitindo que conjuntos de dados amostrais menores sejam analisados. Esse teste avalia a hipótese nula de a amostra de dados ter sido retirada de uma população distribuída normalmente em comparação a uma hipótese alternativa de que a amostra de dados não é distribuída normalmente. Se o p-valor calculado for menor ou igual ao valor de significância de alfa, rejeite a hipótese nula e aceite a hipótese alternativa. Caso contrário, se o p-valor for maior que o valor de significância alfa, não rejeite a hipótese nula. Esse teste se baseia em duas frequências cumulativas: uma derivada do conjunto de dados da amostra e a outra, da distribuição teórica baseada na média e no desvio padrão dos dados da amostra. Uma alternativa a esse teste é o teste qui-quadrado de normalidade. O teste qui-quadrado requer mais pontos de dados para que seja executado se comparado ao teste de normalidade aqui usado.

Resultado do teste

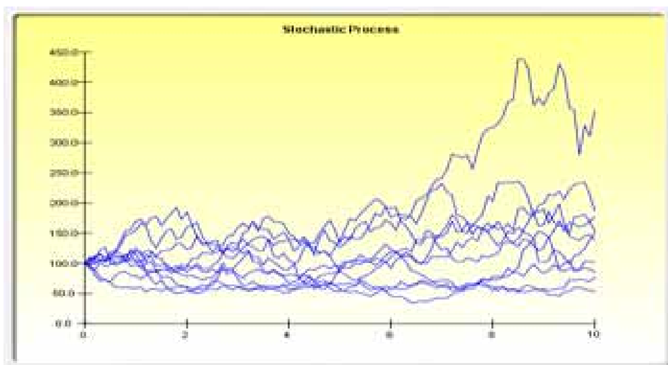
Média dos dados	*21667.54*
Desvio padrão	*21906.10*
Estatística D	*0.2042*
D crítico em 1%	*0.1138*
D crítico em 5%	*0.1225*
D crítico em 10%	*0.1458*

Hipótese nula: os dados estão distribuídos normalmente.

Conclusão: A amostra de dados é não normalmente distribuída.

Dados	*Frequência relativa*	*Observado*	*Esperado*	*O-E*
1148.00	*0.02*	*0.02*	*0.1745*	*-0.1545*
1641.00	*0.02*	*0.04*	*0.1803*	*-0.1403*
2373.00	*0.02*	*0.06*	*0.1892*	*-0.1292*
3136.00	*0.02*	*0.08*	*0.1988*	*-0.1188*
3680.00	*0.02*	*0.10*	*0.2058*	*-0.1058*
3711.00	*0.02*	*0.12*	*0.2062*	*-0.0862*
3872.00	*0.02*	*0.14*	*0.2083*	*-0.0683*
4008.00	*0.02*	*0.16*	*0.2101*	*-0.0501*
4487.00	*0.02*	*0.18*	*0.2164*	*-0.0364*

Figura 5.32 – Relatório da ferramenta Análise estatística da amostra (teste de normalidade)



Resumo estatístico

Os parâmetros a seguir são estimados para um processo estocástico, considerando-se os dados fornecidos. Cabe a você determinar se a probabilidade de ajuste (semelhante a um cálculo de melhor ajuste) é suficiente para garantir o uso de uma previsão de processo estocástico e, se assim for, se é um modelo de caminho aleatório, de reversão à média, de difusão com salto ou de combinações deles. Ao escolher o modelo de processo estocástico correto, você precisa confiar em suas experiências passadas e as expectativas econômicas e financeiras a priori que melhor representam o conjunto de dados subjacentes. Esses parâmetros podem ser inseridos em uma previsão de processo estocástico (**Risk Simulator** I **Previsão** I **Processos estocásticos**).

(Anualizado)

*Taxa de crescimento**	*-4.80%*	*Taxa de reversão***	*N/A*	*Taxa de salto***	*14.29%*
*Volatilidade**	*152.56%*	*Valor de longo prazo***	*21772.52*	*Tamanho do salto***	*32960.71*

Probabilidade de ajuste do modelo estocástico: *54.62%*

*Os valores são anualizados

**Os valores são periódicos

Figura 5.33 – Relatório da ferramenta Análise estatística da amostra (estimativa do parâmetro estocástico)

Ferramenta Análise da distribuição

Esta é uma ferramenta de probabilidade estatística do Risk Simulator que é muito útil em diversas situações e pode ser usada para calcular a função densidade de probabilidade (FDP), também chamada de função de probabilidade de massa (PMF) para distribuições discretas (usaremos esses termos de maneira intercambiável), na qual dadas algumas distribuições e seus parâmetros, pode-se determinar a probabilidade da ocorrência considerados um resultado *x*. Além disso, a função distribuição acumulada (FDA) também pode ser calculada, que é a soma dos valores PDF até esse valor *x*. Por fim, a função de distribuição acumulada inversa (IFDA) é usada para calcular o valor *x*, dada a probabilidade cumulativa de ocorrência.

Essa ferramenta está disponível em **Risk Simulator | Ferramentas | Análise da distribuição**. Como exemplo, a Figura 5.34 mostra o cálculo de uma distribuição binomial (por exemplo, uma distribuição com dois resultados, como jogar uma moeda, em que o resultado é cara ou coroa, com probabilidades prescritas de cara e de coroa). Suponha que uma moeda tenha sido jogada duas vezes e o resultado cara tenha sido definido como êxito. Usamos a distribuição binomial com tentativas = 2 (a moeda é jogada duas vezes) e a probabilidade = 0,50 (a probabilidade de êxito, de o resultado ser cara). Selecionando a FDP e configurando a faixa de valores de *x* como 0 a 2, com um incremento de 1 (isso significa que estamos solicitando os valores 0, 1, 2 para *x*), as probabilidades resultantes são fornecidas na tabela e em um formato gráfico, além dos quatro momentos teóricos da distribuição. Enquanto os resultados do jogo são cara-cara, coroa-coroa, cara-coroa e coroa-cara, a probabilidade exata de não obter caras é 25%, uma cara é 50% e duas caras é 25%.

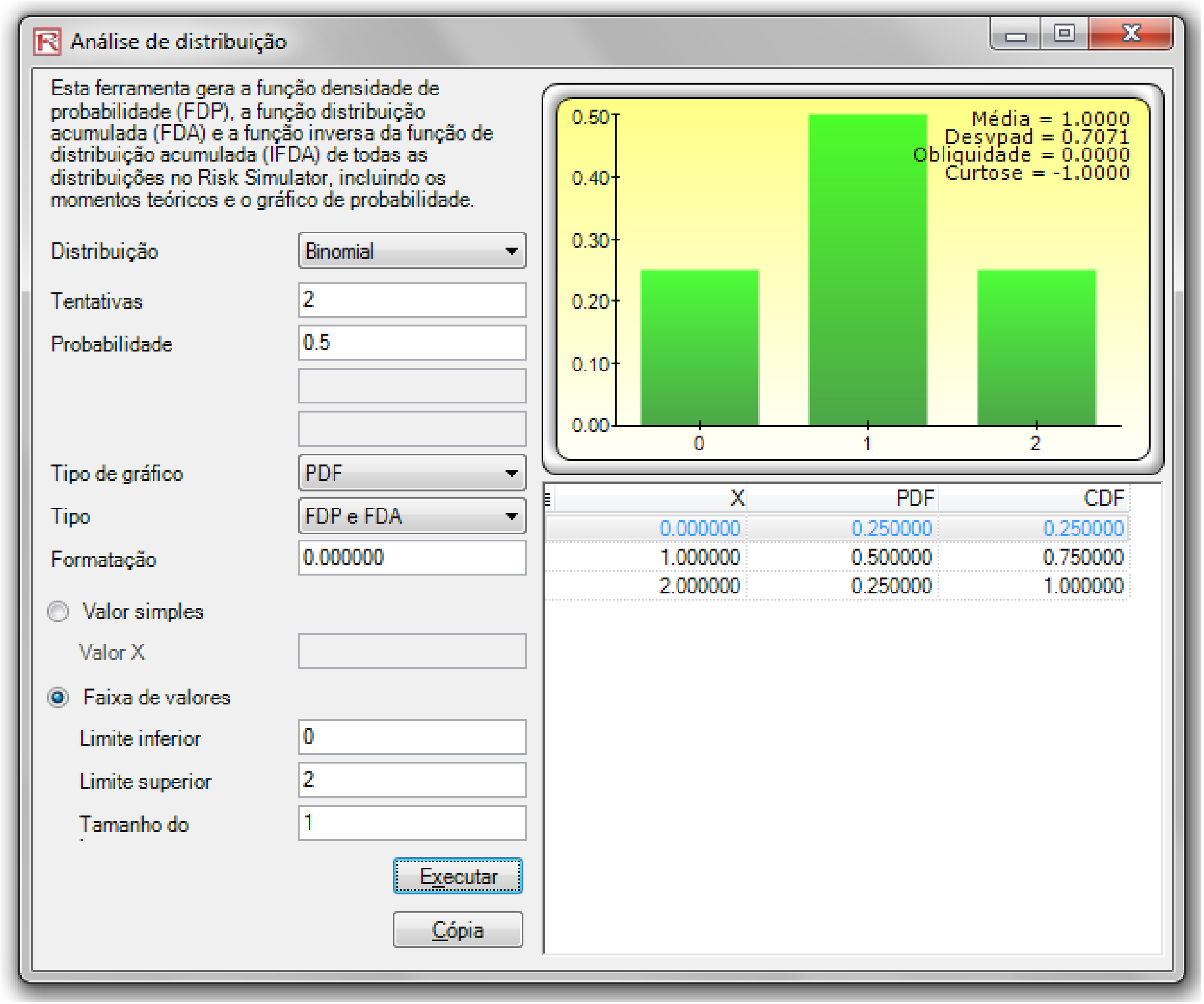


Figura 5.34 – Ferramenta Análise da distribuição (distribuição binomial com 2 tentativas)

Da mesma forma, podemos obter as probabilidades exatas de jogar a moeda, digamos 20 vezes, como na Figura 5.35. Os resultados são apresentados em uma tabela e em um gráfico.

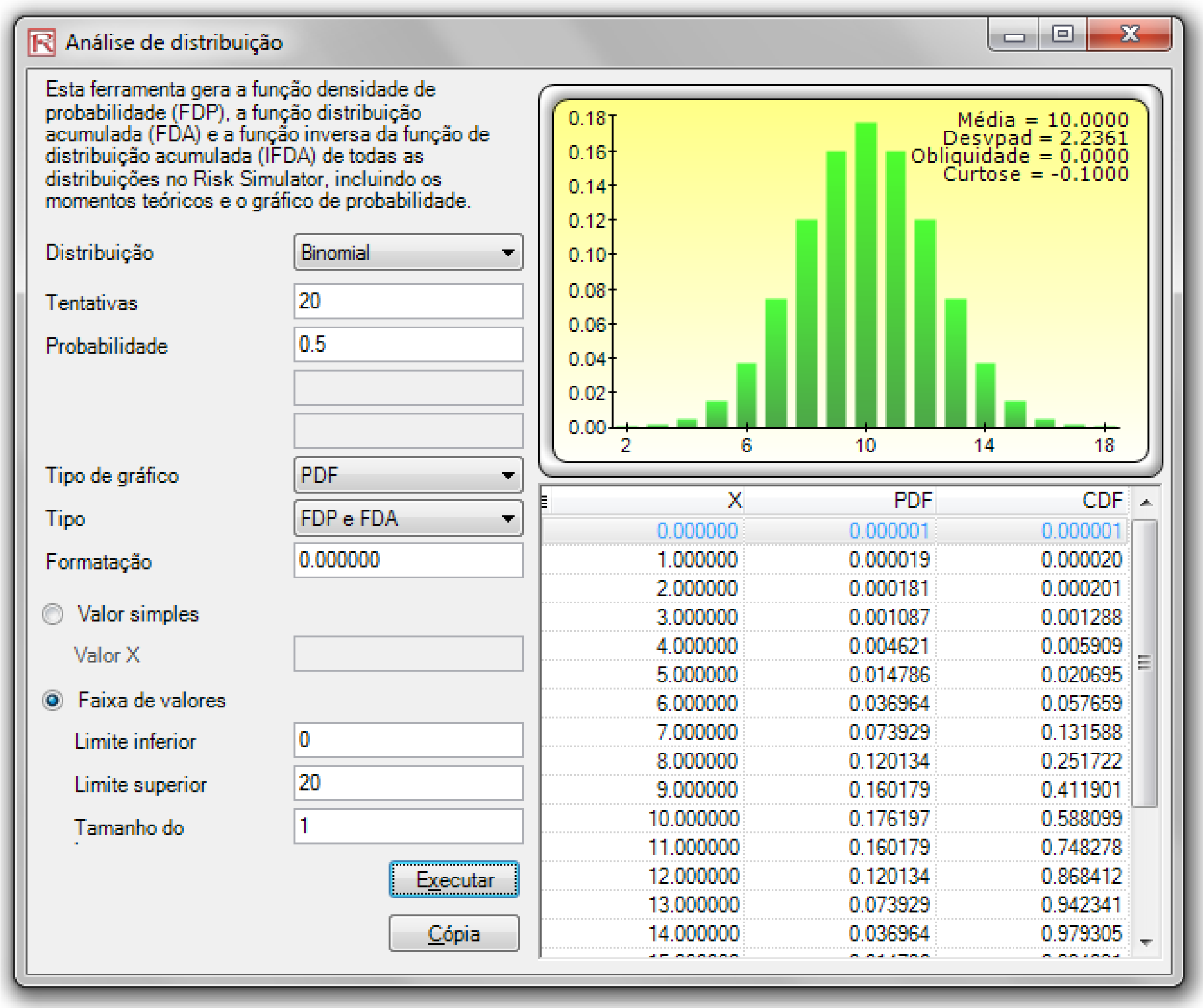


Figura 5.35 – Ferramenta Análise da distribuição (distribuição binomial com 20 tentativas)

A Figura 5.36 mostra a mesma distribuição binomial, mas agora a FDA é calculado. A FDA é simplesmente a soma dos valores de FDP até o ponto *x*. Por exemplo, na Figura 5.35, observamos que as probabilidades de 0, 1 e 2 são 0,000001, 0,000019 e 0,000181, cuja soma é 0,000201, que é o valor de FDA em*x* = 2 na Figura 5.36. Enquanto FDP calcula as probabilidades de obter-se exatamente 2 caras, FDA calcula a probabilidade de obter não mais do que 2 caras ou até 2 caras (ou probabilidades de 0, 1 e 2 caras). O complemento (por exemplo, 1 – 0,00021 obtém 0,999799 ou 99,9799%) é a probabilidade de se obter pelo menos 3 caras ou mais.

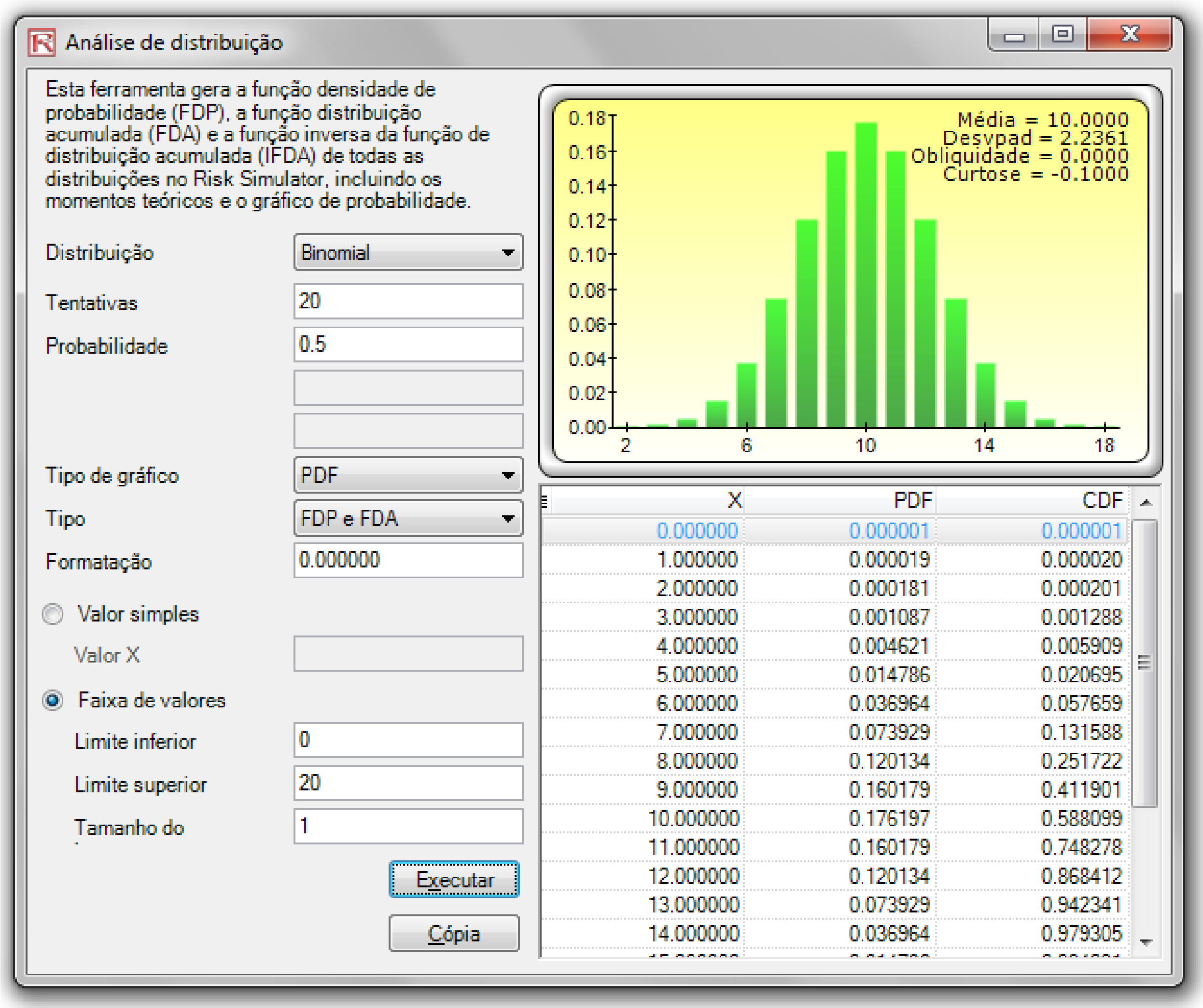


Figura 5.36 – Ferramenta Análise da distribuição (FDA da distribuição binomial com 20 tentativas)

Usando a ferramenta Análise da distribuição, distribuições ainda mais avançadas podem ser analisadas no Risk Simulator, como gama, beta, binomial negativa e muitas outras. Como outro exemplo do uso da ferramenta em uma distribuição contínua e da funcionalidade de IFDA, a Figura 5.37 mostra a distribuição normal padrão (distribuição normal com uma média zero e desvio padrão igual a um), na qual aplicamos IFDA para encontrar o valor de *x* que corresponde à probabilidade cumulativa de 97,50% (FDA). Ou seja, uma FDA unicaudal de 97,50% é equivalente a um intervalo de confiança bicaudal de 95% (há uma probabilidade de 2,50% na cauda direita e 2,50% na cauda esquerda, restando 95% no centro ou na área do intervalo de confiança, que equivale a uma área de 97,50% para uma cauda). O resultado é o conhecido escore Z de 1,96. Assim, usando a ferramenta Análise da distribuição, pode-se obter rápida e facilmente os escores padronizados para outras distribuições, bem como as probabilidades cumulativas e exatas de outras distribuições.

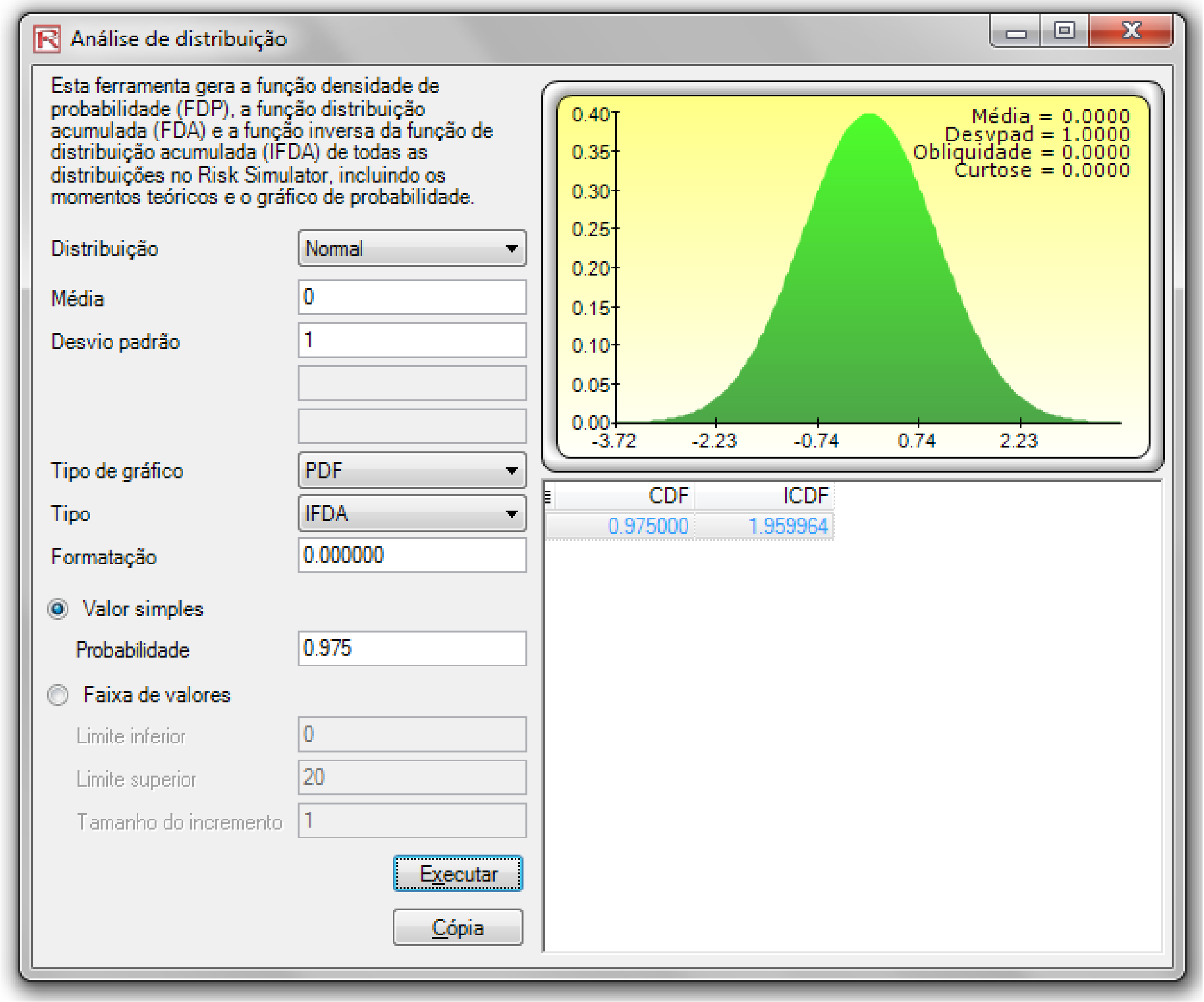


Figura 5.37 – Ferramenta Análise da distribuição (IFDA e escore Z da distribuição normal)

Ferramenta Análise de cenário

A ferramenta Análise de cenário do Risk Simulator permite executar vários cenários rapidamente e sem esforço alterando um ou dois parâmetros de entrada para determinar o resultado de uma variável. A Figura 5.38 ilustra como essa ferramenta trabalha no modelo de exemplo de fluxo de caixa descontado (modelo 7 na pasta de modelos de exemplo do Risk Simulator). Nesse exemplo, a célula G6 (valor presente líquido) é selecionada como o resultado de interesse, enquanto as células C9 (alíquota de impostos efetiva) e C12 (preço do produto) são selecionadas como entradas para causar perturbação. É possível definir os valores inicial e final a serem testados, além do tamanho do incremento ou o número de etapas que devem ser executadas entre os valores inicial e final. O resultado é uma tabela de análise de cenário (Figura 5.39), na qual os cabeçalhos de linha e de coluna são as duas variáveis de entrada e o corpo a tabela mostra os valores presentes líquidos.

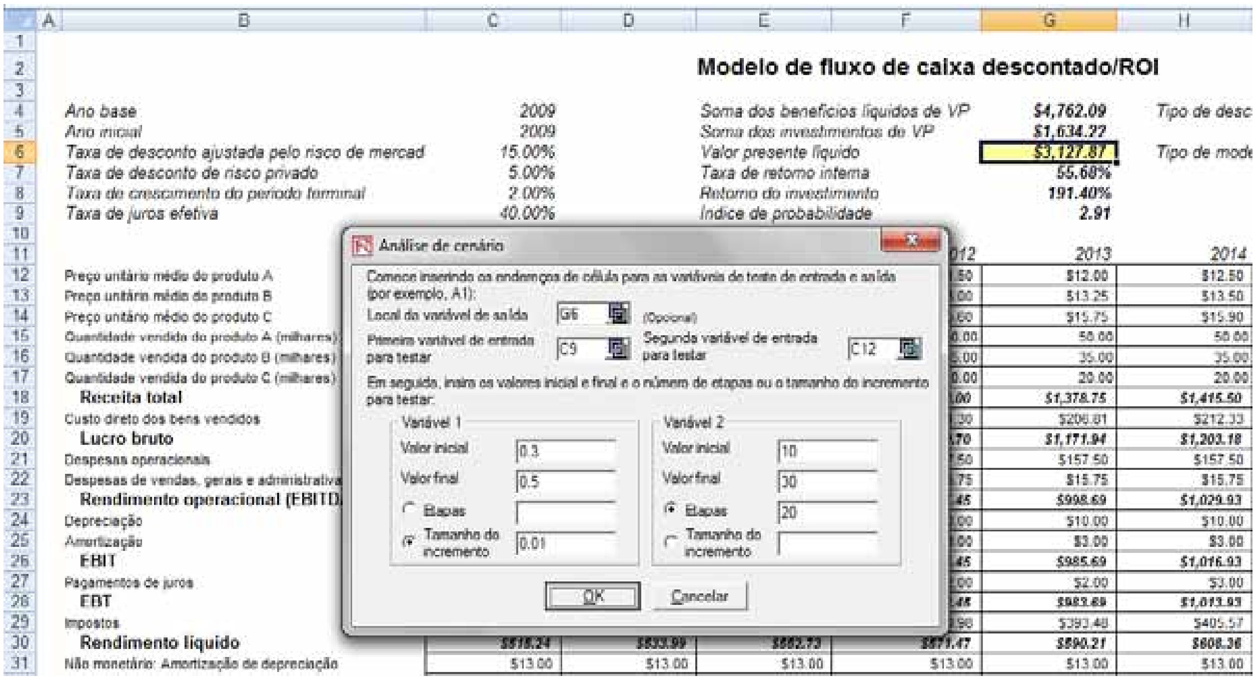


Figura 5.38 – Ferramenta Análise de cenário

TABELA DE ANÁLISE DE CENÁRIO

Variável de saída:	G6	**Valor inicial do caso básico:**	$3,127.87								
Variável de coluna:	C12	**Mín:**	10	**Máx:**	30	**Etapas:**	20	**Incremento:**	---	**Valor inicial do caso básico:**	$10.00
Variável de linha:	C9	**Mín:**	0.3	**Máx:**	0.5	**Etapas:**	---	**Incremento:**	0.01	**Valor inicial do caso básico:**	40.00%

	$10.00	**$11.00**	**$12.00**	**$13.00**	**$14.00**	**$15.00**	**$16.00**	**$17.00**	**$18.00**	**$19.00**
30.00%	$3,904.83	$4,134.43	$4,364.04	$4,593.64	$4,823.24	$5,052.84	$5,282.44	$5,512.04	$5,741.64	$5,971.24
31.00%	$3,827.14	$4,053.46	$4,279.78	$4,506.10	$4,732.42	$4,958.74	$5,185.06	$5,411.39	$5,637.71	$5,864.03
32.00%	$3,749.44	$3,972.48	$4,195.52	$4,418.56	$4,641.61	$4,864.65	$5,087.69	$5,310.73	$5,533.77	$5,756.81
33.00%	$3,671.75	$3,891.51	$4,111.27	$4,331.03	$4,550.79	$4,770.55	$4,990.31	$5,210.07	$5,429.83	$5,649.60
34.00%	$3,594.05	$3,810.53	$4,027.01	$4,243.49	$4,459.97	$4,676.45	$4,892.94	$5,109.42	$5,325.90	$5,542.38
35.00%	$3,516.35	$3,729.55	$3,942.76	$4,155.96	$4,369.16	$4,582.36	$4,795.56	$5,008.76	$5,221.96	$5,435.16
36.00%	$3,438.66	$3,648.58	$3,858.50	$4,068.42	$4,278.34	$4,488.26	$4,698.18	$4,908.10	$5,118.03	$5,327.95
37.00%	$3,360.96	$3,567.60	$3,774.24	$3,980.88	$4,187.53	$4,394.17	$4,600.81	$4,807.45	$5,014.09	$5,220.73
38.00%	$3,283.27	$3,486.63	$3,689.99	$3,893.35	$4,096.71	$4,300.07	$4,503.43	$4,706.79	$4,910.15	$5,113.51
39.00%	$3,205.57	$3,405.65	$3,605.73	$3,805.81	$4,005.89	$4,205.97	$4,406.06	$4,606.14	$4,806.22	$5,006.30
40.00%	$3,127.87	$3,324.67	$3,521.48	$3,718.28	$3,915.08	$4,111.88	$4,308.68	$4,505.48	$4,702.28	$4,899.08
41.00%	$3,050.18	$3,243.70	$3,437.22	$3,630.74	$3,824.26	$4,017.78	$4,211.30	$4,404.82	$4,598.35	$4,791.87
42.00%	$2,972.48	$3,162.72	$3,352.96	$3,543.20	$3,733.45	$3,923.69	$4,113.93	$4,304.17	$4,494.41	$4,684.65
43.00%	$2,894.79	$3,081.75	$3,268.71	$3,455.67	$3,642.63	$3,829.59	$4,016.55	$4,203.51	$4,390.47	$4,577.43
44.00%	$2,817.09	$3,000.77	$3,184.45	$3,368.13	$3,551.81	$3,735.49	$3,919.18	$4,102.86	$4,286.54	$4,470.22
45.00%	$2,739.39	$2,919.79	$3,100.20	$3,280.60	$3,461.00	$3,641.40	$3,821.80	$4,002.20	$4,182.60	$4,363.00
46.00%	$2,661.70	$2,838.82	$3,015.94	$3,193.06	$3,370.18	$3,547.30	$3,724.42	$3,901.54	$4,078.66	$4,255.79
47.00%	$2,584.00	$2,757.84	$2,931.68	$3,105.52	$3,279.37	$3,453.21	$3,627.05	$3,800.89	$3,974.73	$4,148.57
48.00%	$2,506.31	$2,676.87	$2,847.43	$3,017.99	$3,188.55	$3,359.11	$3,529.67	$3,700.23	$3,870.79	$4,041.35
49.00%	$2,428.61	$2,595.89	$2,763.17	$2,930.45	$3,097.73	$3,265.01	$3,432.29	$3,599.58	$3,766.86	$3,934.14
50.00%	$2,350.91	$2,514.91	$2,678.92	$2,842.92	$3,006.92	$3,170.92	$3,334.92	$3,498.92	$3,662.92	$3,826.92

	$20.00	**$21.00**	**$22.00**	**$23.00**	**$24.00**	**$25.00**	**$26.00**	**$27.00**	**$28.00**	**$29.00**	**$30.00**
30.00%	$6,200.85	$6,430.45	$6,660.05	$6,889.65	$7,119.25	$7,348.85	$7,578.45	$7,808.05	$8,037.65	$8,267.26	$8,496.86
31.00%	$6,090.35	$6,316.67	$6,542.99	$6,769.31	$6,995.63	$7,221.96	$7,448.28	$7,674.60	$7,900.92	$8,127.24	$8,353.56
32.00%	$5,979.85	$6,202.89	$6,425.94	$6,648.98	$6,872.02	$7,095.06	$7,318.10	$7,541.14	$7,764.18	$7,987.22	$8,210.26
33.00%	$5,869.36	$6,089.12	$6,308.88	$6,528.64	$6,748.40	$6,968.16	$7,187.92	$7,407.68	$7,627.45	$7,847.21	$8,066.97
34.00%	$5,758.86	$5,975.34	$6,191.82	$6,408.30	$6,624.79	$6,841.27	$7,057.75	$7,274.23	$7,490.71	$7,707.19	$7,923.67
35.00%	$5,648.36	$5,861.57	$6,074.77	$6,287.97	$6,501.17	$6,714.37	$6,927.57	$7,140.77	$7,353.97	$7,567.17	$7,780.38
36.00%	$5,537.87	$5,747.79	$5,957.71	$6,167.63	$6,377.55	$6,587.47	$6,797.39	$7,007.32	$7,217.24	$7,427.16	$7,637.08
37.00%	$5,427.37	$5,634.01	$5,840.65	$6,047.30	$6,253.94	$6,460.58	$6,667.22	$6,873.86	$7,080.50	$7,287.14	$7,493.78
38.00%	$5,316.88	$5,520.24	$5,723.60	$5,926.96	$6,130.32	$6,333.68	$6,537.04	$6,740.40	$6,943.76	$7,147.13	$7,350.49
39.00%	$5,206.38	$5,406.46	$5,606.54	$5,806.62	$6,006.70	$6,206.79	$6,406.87	$6,606.95	$6,807.03	$7,007.11	$7,207.19
40.00%	$5,095.88	$5,292.68	$5,489.49	$5,686.29	$5,883.09	$6,079.89	$6,276.69	$6,473.49	$6,670.29	$6,867.09	$7,063.89
41.00%	$4,985.39	$5,178.91	$5,372.43	$5,565.95	$5,759.47	$5,952.99	$6,146.51	$6,340.03	$6,533.56	$6,727.08	$6,920.60
42.00%	$4,874.89	$5,065.13	$5,255.37	$5,445.61	$5,635.86	$5,826.10	$6,016.34	$6,206.58	$6,396.82	$6,587.06	$6,777.30
43.00%	$4,764.40	$4,951.36	$5,138.32	$5,325.28	$5,512.24	$5,699.20	$5,886.16	$6,073.12	$6,260.08	$6,447.04	$6,634.01
44.00%	$4,653.90	$4,837.58	$5,021.26	$5,204.94	$5,388.62	$5,572.30	$5,755.98	$5,939.67	$6,123.35	$6,307.03	$6,490.71
45.00%	$4,543.40	$4,723.80	$4,904.20	$5,084.61	$5,265.01	$5,445.41	$5,625.81	$5,806.21	$5,986.61	$6,167.01	$6,347.41
46.00%	$4,432.91	$4,610.03	$4,787.15	$4,964.27	$5,141.39	$5,318.51	$5,495.63	$5,672.75	$5,849.87	$6,027.00	$6,204.12
47.00%	$4,322.41	$4,496.25	$4,670.09	$4,843.93	$5,017.77	$5,191.62	$5,365.46	$5,539.30	$5,713.14	$5,886.98	$6,060.82
48.00%	$4,211.91	$4,382.48	$4,553.04	$4,723.60	$4,894.16	$5,064.72	$5,235.28	$5,405.84	$5,576.40	$5,746.96	$5,917.52
49.00%	$4,101.42	$4,268.70	$4,435.98	$4,603.26	$4,770.54	$4,937.82	$5,105.10	$5,272.38	$5,439.67	$5,606.95	$5,774.23
50.00%	$3,990.92	$4,154.92	$4,318.92	$4,482.92	$4,646.93	$4,810.93	$4,974.93	$5,138.93	$5,302.93	$5,466.93	$5,630.93

Figura 5.39 – Tabela de análise de cenário

Ferramenta Agrupamento por segmentação

Uma última técnica analítica de interesse é o agrupamento por segmentação. A Figura 6.25 ilustra um conjunto de dados de exemplo. Para selecionar os dados e executar a ferramenta, clique em Risk Simulator | Ferramentas | Agrupamento por segmentação. A Figura 5.40 mostra um exemplo da segmentação de dois grupos. Ou seja, tomando o conjunto de dados original, executamos alguns logaritmos internos (uma combinação ou agrupamento hierárquico de médias k e outros métodos de momentos para encontrar os grupos de ajustamento ou combinações estatísticas naturais de melhor ajuste) para dividir ou segmentar estatisticamente o conjunto de dados original em dois grupos. Você pode ver os membros dos dois grupos na Figura 5.40. É claro que você pode segmentar esse conjunto de dados em quantos grupos desejar. Essa técnica é valiosa em uma variedade de situações, incluindo marketing (segmentação do mercado de clientes em vários grupos de gerenciamento de relacionamento de clientes etc.), ciências naturais, engenharia, entre outras.

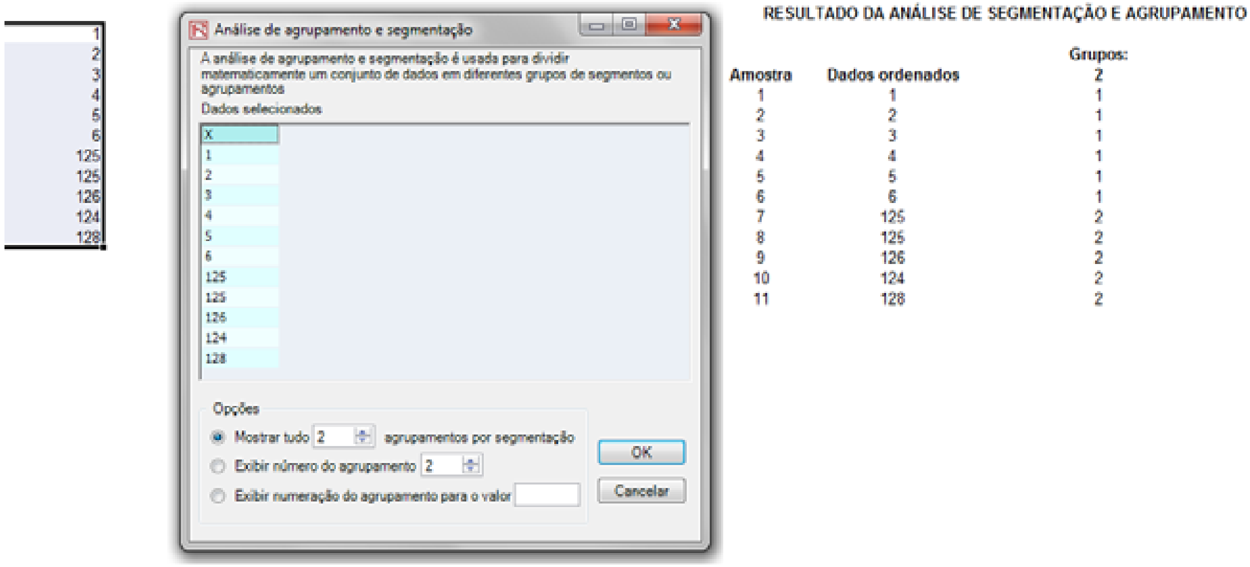


Figura 5.40 – Ferramenta Agrupamento por segmentação e resultados

Novas ferramentas do Risk Simulator 2011/2012

Métodos de geração de número aleatório, Monte Carlo versus hipercubo latino e cópula de correlação

A partir da versão 2011/2012, há seis geradores de número aleatório, três cópulas de correlação e dois métodos de amostragem de simulação disponíveis (Figura 5.41). Essas preferências são configuradas em ***Risk Simulator | Opções***.

O gerador de número aleatório (RNG) é a parte principal de qualquer software de simulação. Com base no número aleatório gerado, diferentes distribuições matemáticas podem ser construídas. O método padrão é a metodologia patenteada ROV Risk Simulator, que fornece os melhores e mais sólidos números aleatórios. Há suporte para seis geradores de número aleatório e, em geral, o método padrão do ROV Risk Simulator e o método de embaralhamento aleatório subtrativo avançado são as duas abordagens de uso recomendadas. Não aplique os outros métodos, a menos que o seu modelo ou sua análise necessite especificamente deles e, mesmo nesse caso, recomendamos testar os resultados com as duas abordagens indicadas. Quanto mais baixa a posição na lista de RNGs, mais simples é o algoritmo e mais rápido ele é executado. De maneira oposta, quanto mais alta a posição na lista, mais sólidos são os resultados.

Na seção Correlações, há três métodos: cópula normal, cópula T e cópula quasi-normal. Esses métodos usam técnicas de integração matemáticas e, quando há dúvidas, a cópula normal fornece os resultados mais seguros e conservadores. A cópula t fornece valores extremos nas caudas das distribuições simuladas, enquanto a cópula quasi-normal retorna valores que estão entre esses valores.

Na seção Métodos de simulação, há os métodos de simulação Monte Carlo (MCS) e amostragem por hipercubo latino (LHS). Observe que as cópulas e as outras funções multivariadas ***não*** são compatíveis com o LHS. Isso acontece porque o LHS pode ser aplicado a uma única variável, mas não a uma distribuição em conjunto. Na realidade, o LHS possui impacto muito limitado na precisão do resultado do modelo quanto mais distribuições houver em um modelo, uma vez que o LHS só se aplica a distribuições individualmente. A vantagem do LHS também será comprometida se uma pessoa não concluir o número de amostras determinadas no início, isto é, se a simulação for interrompida após o início. O LHS também aplica uma carga pesada em um modelo de simulação com um grande número de entradas porque precisa gerar e organizar amostras de cada distribuição antes de executar a primeira amostra de uma distribuição. Isso pode causar uma longa demora na execução de um modelo grande e fornecer pouca precisão adicional. Por último, o LHS é mais adequado quando as distribuições são bem comportadas e simétricas e não possuem correlações. Apesar disso, o LHS é uma abordagem poderosa que rende uma distribuição de amostragem uniforme, na qual o MCS pode às vezes gerar distribuições irregulares (ocasionalmente, os dados da amostra podem estar concentrados em uma área da distribuição) em comparação a uma distribuição de amostragem mais uniforme (todas as partes da distribuição serão amostradas) quando o LHS é aplicado.

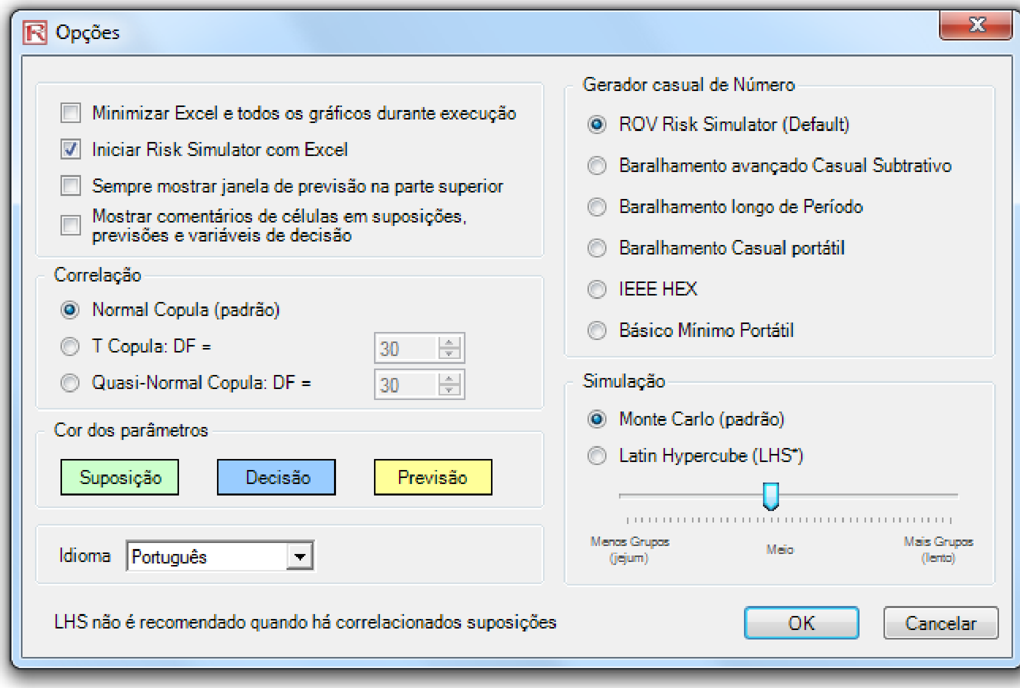


Figura 5.41 – Opções do Risk Simulator

Supressão de tendência de dados, dessazonalização de dados e teste de sazonalidade

A ferramenta de supressão de tendência e dessazonalização de dados do Risk Simulator permite remover qualquer componente de sazonalidade ou tendência nos seus dados. Este processo permite mostrar somente as alterações absolutas no valor de período a período, permitindo a identificação de possíveis padrões cíclicos nos dados da série temporal. A dessazonalização e a supressão de tendência removem ciclos de crescimento, tendência, variação, curva e outros ciclos sazonais que possam afetar os dados da série temporal, deixando o comportamento verdadeiramente estrutural dos dados ao longo do tempo.

Os períodos de sazonalidade representam quantos períodos teriam de decorrer antes que o ciclo se repita (por exemplo, 24 horas em um dia, 12 meses em um ano, 4 trimestres em um ano, 60 minutos em uma hora etc.), mas às vezes existem outros períodos sazonais não inteiramente evidentes mediante a mera observação dos dados ou da variável. Esse teste de sazonalidade observa os dados da série temporal para determinar a periodicidade da sazonalidade de melhor ajuste para os dados. Agora, usando essa sazonalidade, é possível ajustar os efeitos de sazonalidade usando a ferramenta Dessazonalização dos dados como mostrado acima ou usando a ferramenta Análise de série temporal para fornecer uma previsão melhor.

Procedimento para supressão de tendência e dessazonalização:

- Selecione os dados que deseja analisar (por exemplo, B9:B28) e clique em ***Risk Simulator | Ferramentas | Supressão de tendência e dessazonalização dos dados***
- ***Selecione Dessazonalização dos dados*** e/ou ***Supressão de tendência,*** selecione os modelos de supressão de tendência que deseja executar, insira as ordens relevantes (por exemplo, ordem polinomial, ordem de média móvel, ordem de diferença e ordem de razão) e clique em OK
- Revise os dois relatórios gerados para obter mais detalhes sobre a metodologia, a aplicação e os gráficos de resultados e os dados sem tendência/sazonalidade

Procedimento para o teste de sazonalidade:

- Selecione os dados que deseja analisar (por exemplo, B9:B28) e clique em ***Risk Simulator | Ferramentas | Teste de Sazonalidade de Dados***
- Selecione os dados que deseja analisar (por exemplo, B9:B28) e clique em ***Risk Simulator | Ferramentas | Dados Teste de sazonalidad***e
- Insira o período de sazonalidade máximo para o teste. Ou seja, se você inserir 6, o Risk Simulator testará os seguintes períodos de sazonalidade: 1, 2, 3, 4, 5, 6. O período 1 certamente não implica nenhuma sazonalidade nos dados
- Revise o relatório gerado para obter mais detalhes sobre a metodologia, a aplicação e os gráficos resultantes e os resultados do teste de sazonalidade. A melhor periodicidade de sazonalidade é listada primeiro (classificada pela menor medida de erro de REQM) e todas as medidas de erro relevantes são incluídas para comparação: a raiz do erro quadrático médio (REQM), o erro quadrático médio (EQM), o desvio absoluto médio (MAD) e o erro médio percentual absoluto (MAPE)

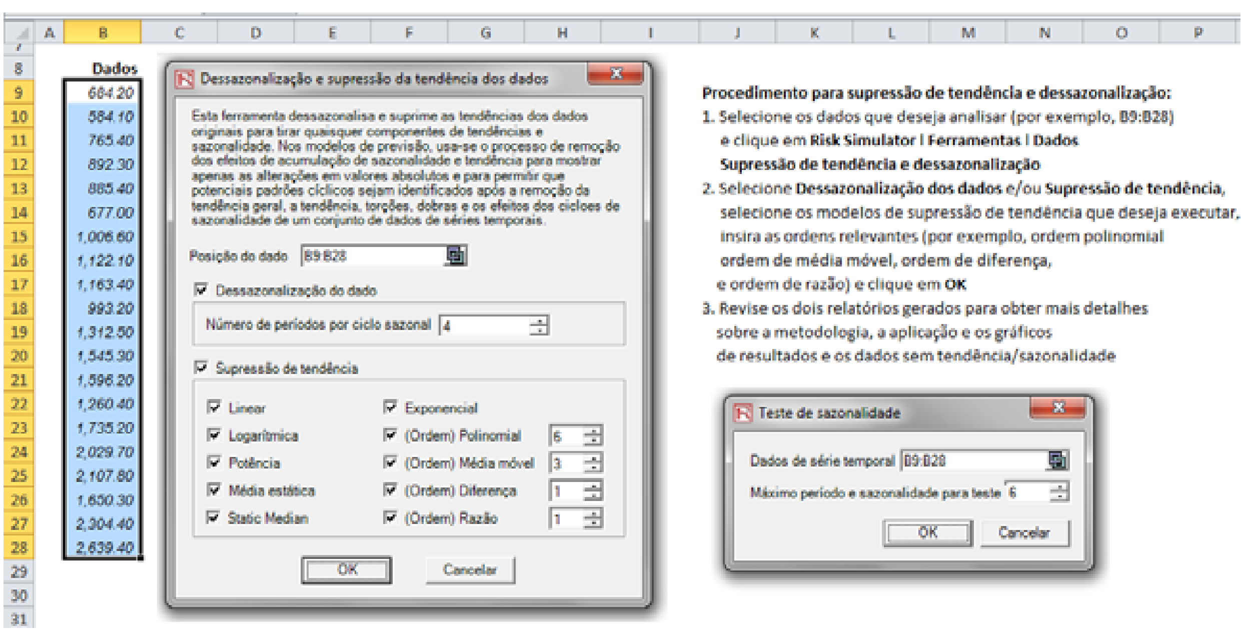


Figura 5.42 – Dados de supressão de tendência e dessazonalização

Análise de componentes principais

A análise de componentes principais (ACP) é uma maneira de identificar padrões em um conjunto de dados e de reformular os dados, como forma de destacar as suas semelhanças e diferenças. É muito difícil encontrar padrões de dados em dimensões elevadas, quando existem múltiplas variáveis. Também é muito difícil representar e interpretar gráficos tridimensionais. Quando os padrões nos dados são encontrados, eles podem ser compactados e o número de dimensões é reduzido. Essa redução das dimensões dos dados não significa muito na redução da perda de informações. Em vez disso, os níveis similares de informação, podem, agora, ser obtidos com um número menor de variáveis.

Procedimento:

- Selecione os dados a serem analisados (por exemplo, B11:K30), clique em ***Risk Simulator | Ferramentas | Análise de componentes principais*** e clique em OK
- Revise o relatório gerado com os resultados calculados

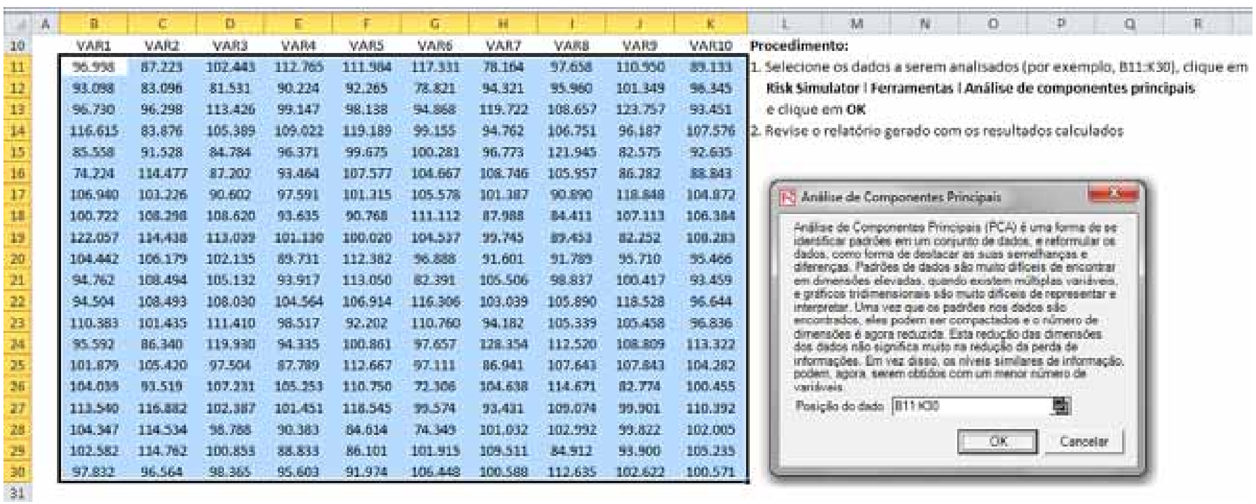


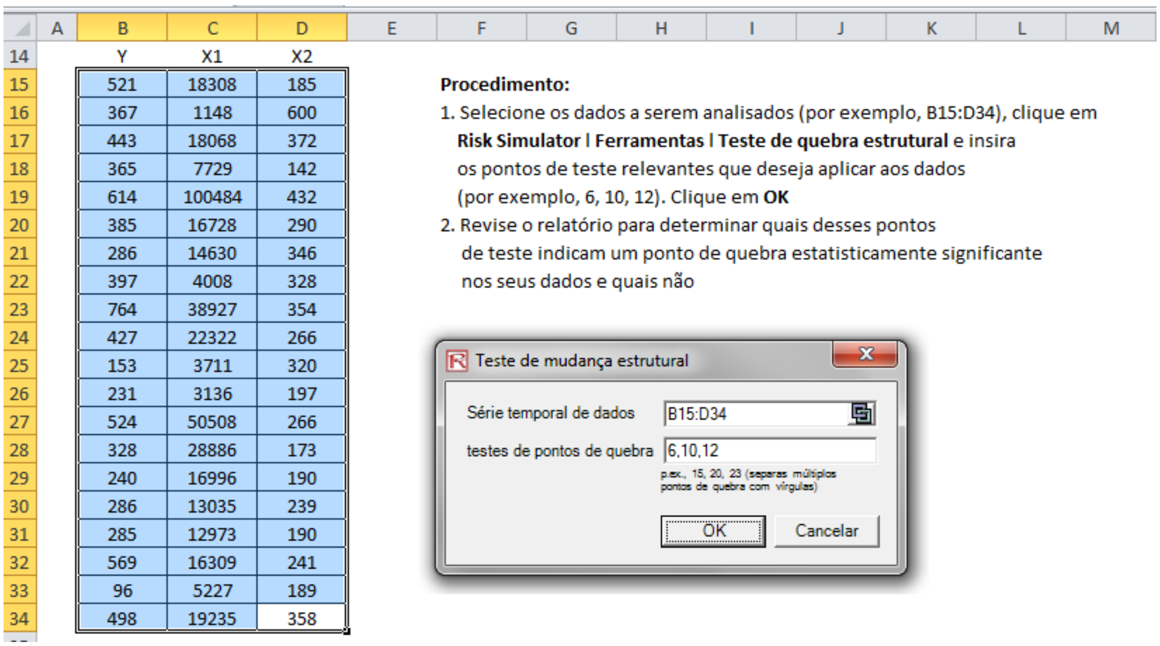
Figura 5.43 – Análise de componentes principais

Série temporal de dados da quebra estrutural

A quebra estrutural testa se os coeficientes em diferentes conjuntos de dados são iguais. E esse teste é mais comumente usado na análise de séries temporais para testar a presença de uma quebra estrutural. A série temporal pode ser dividida em dois subgrupos e cada subgrupo é testado contra o outro e no conjunto de dados completo, para determinar estatisticamente se realmente existe uma pausa a partir de um determinado período de tempo. O teste de quebra estrutural é usado com frequência para determinar se as variáveis independentes têm diferentes impactos em diferentes subgrupos da população, como para testar se uma nova campanha de marketing, uma atividade, um grande evento, uma aquisição ou uma alienação, por exemplo, têm um impacto sobre os dados de séries temporais. Suponha que dado um conjunto com 100 pontos de dados de uma série temporal, pode-se definir pontos de interrupção diferentes para testar, por exemplo, os pontos de dados 10, 30 e 51 (isso significa que três testes de quebra estrutural serão realizados nestes conjuntos de dados: os pontos de dados de 1 a 9 são comparados com os de 10 a 100; os pontos de dados de 30 a 100 são comparados com os de 1 a 29, e os de 1 a 50 são comparados com os de 51 a 100, para ver se de fato, no início dos dados de ponto 10, 30 e 51, há uma quebra na estrutura subjacente).

Procedimento:

- Selecione os dados a serem analisados (por exemplo, B15:D34), clique em ***Risk Simulator | Ferramentas | Teste de quebra estrutural*** e insira os pontos de teste relevantes que deseja aplicar aos dados (por exemplo, 6, 10, 12). Clique em OK
- Revise o relatório para determinar quais desses pontos de teste indicam um ponto de quebra estatisticamente significante nos seus dados e quais não



	Y	X1	X2
15	521	18308	185
16	367	1148	600
17	443	18068	372
18	365	7729	142
19	614	100484	432
20	385	16728	290
21	286	14630	346
22	397	4008	328
23	764	38927	354
24	427	22322	266
25	153	3711	320
26	231	3136	197
27	524	50508	266
28	328	28886	173
29	240	16996	190
30	286	13035	239
31	285	12973	190
32	569	16309	241
33	96	5227	189
34	498	19235	358

Figura 5.44 – Análise de quebra estrutural

Previsões de linha de tendência

As linhas de tendência podem ser usadas para determinar se um conjunto de dados de série temporal segue qualquer tendência considerável (Figura 5.45). As tendências podem ser lineares ou não lineares (como exponencial, logarítmica, média móvel, potência ou polinomial).

Procedimento:

- Selecione os dados que deseja analisar e clique em ***Risk Simulator | Previsão | Linha de tendência***, selecione as linhas de tendência relevantes que você deseja aplicar aos dados (por exemplo, selecione todos os métodos por padrão), insira o número de períodos para a previsão (por exemplo, 6 períodos) e clique em *OK*
- Revise o relatório para determinar quais dessas linhas de tendência de teste fornecem o melhor ajuste e a melhor previsão para os seus dados

Receita de vendas histórica

Ano	Trimestre	Período	Vendas
2006	1	1	$684.20
2006	2	2	$584.10
2006	3	3	$765.40
2006	4	4	$892.30
2007	1	5	$885.40
2007	2	6	$677.00
2007	3	7	$1,006.60
2007	4	8	$1,122.10
2008	1	9	$1,163.40
2008	2	10	$993.20
2008	3	11	$1,312.50
2008	4	12	$1,545.30
2009	1	13	$1,596.20
2009	2	14	$1,260.40
2009	3	15	$1,735.20
2009	4	16	$2,029.70
2010	1	17	$2,107.80
2010	2	18	$1,650.30
2010	3	19	$2,304.40
2010	4	20	$2,639.40

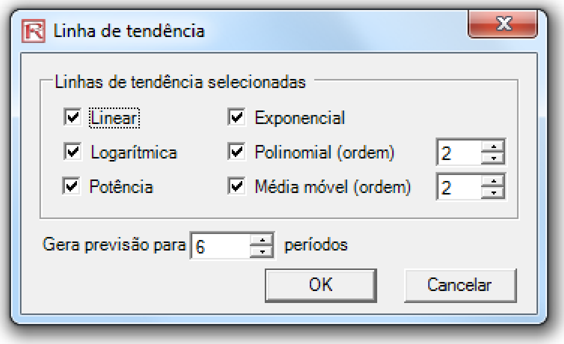


Figura 5.45 – Previsões de linha de tendência

Ferramenta de verificação de modelo

Depois que um modelo é criado e as premissas e as previsões são definidas, você pode executar a simulação normalmente ou a ferramenta de verificação de modelo (Figura 5.46) para testar se o modelo foi configurado corretamente. Como alternativa, se o modelo não for executado e você suspeitar que algumas configurações estejam incorretas, execute a ferramenta em ***Risk Simulator | Ferramentas | Verificar modelo*** para identificar onde pode haver problemas com o seu modelo. Observe que essa ferramenta verifica os problemas de modelo mais comuns e também problemas nas premissas e previsões do Risk Simulator. Ela não é de maneira nenhuma abrangente o suficiente para testar todos os tipos de problemas. É responsabilidade do desenvolvedor do modelo certificar-se de que o modelo funcione corretamente.

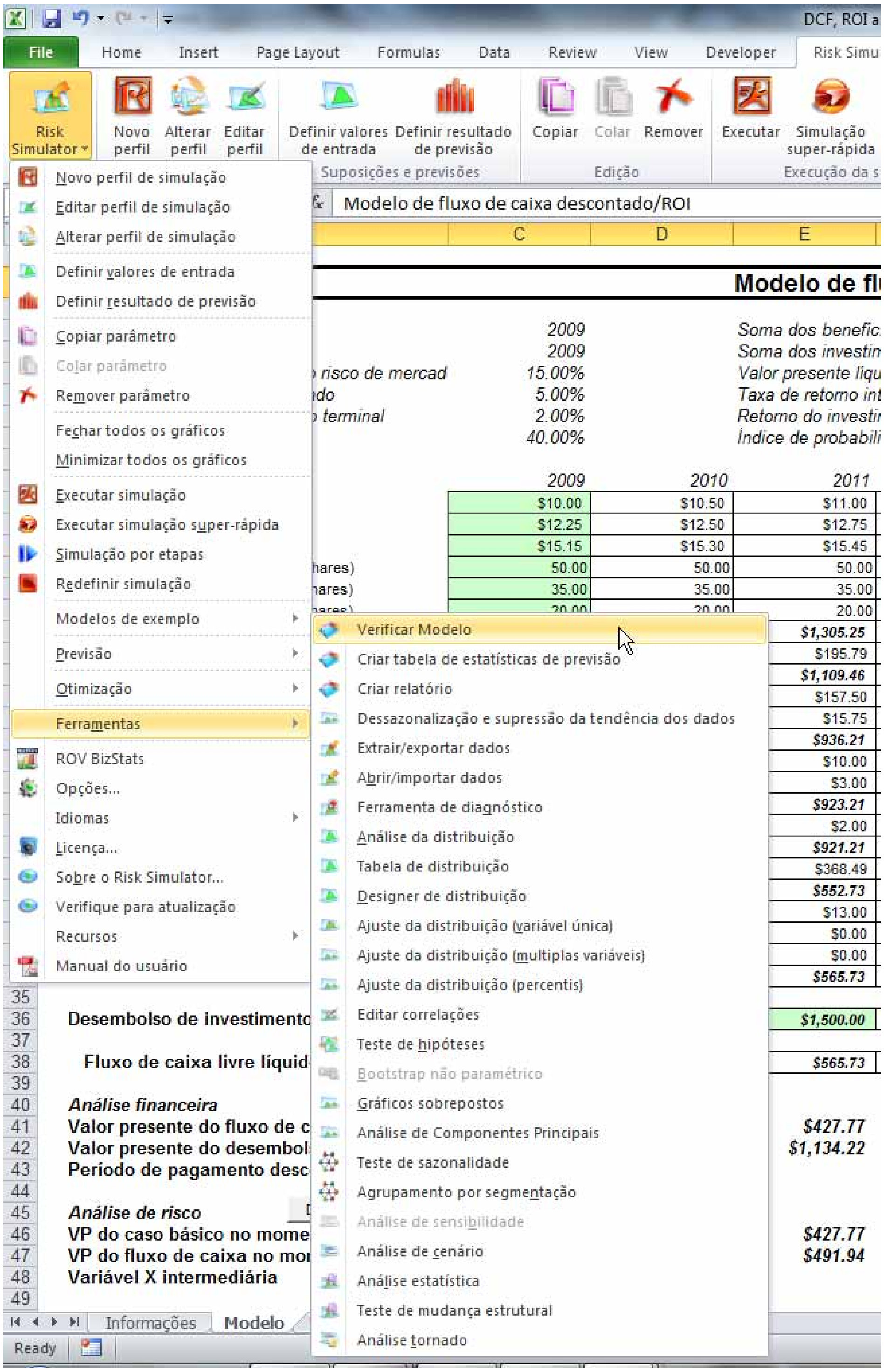


Figura 5.46 – Ferramenta de verificação de modelo

Ferramenta de ajuste da distribuição (percentil)

A ferramenta de ajuste da distribuição (percentil), mostrada na Figura 5.47, é outra maneira de ajustar as distribuições de probabilidade. Há diversas ferramentas relacionadas e cada uma apresenta usos e vantagens específicos:

- Ajuste da distribuição (percentis) – usando o método alternativo de entrada (percentis e combinações de primeiro/segundo momentos), você pode encontrar os parâmetros de distribuição de melhor ajuste sem a necessidade de dados brutos. Esse método é adequado quando não houver dados suficientes, quando somente percentis e momentos estiverem disponíveis ou como um meio de recuperar a distribuição integralmente com apenas dois ou três pontos de dados, mas o tipo de distribuição precisar ser suposto ou conhecido.
- Ajuste da distribuição (variável única) — usando métodos estatísticos para ajustar seus dados brutos a todas as 42 distribuições a fim de encontrar o melhor ajuste da distribuição e os parâmetros de entrada. São necessários vários pontos de dados para um bom ajuste. O tipo de distribuição pode não ser conhecido com antecedência.
- Ajuste da distribuição (múltiplas variáveis) — usa métodos estatísticos para ajustar seus dados brutos em diversas variáveis ao mesmo tempo, usando os mesmos algoritmos como a única variável de ajuste, mas incorpora uma matriz de correlação de paridade entre as variáveis. São necessários vários pontos de dados para um bom ajuste. O tipo de distribuição pode não ser conhecido com antecedência.
- Distribuição personalizada (definir suposição) — usa técnicas de reamostragem não paramétrica para gerar uma distribuição personalizada com os dados brutos existentes e para simular a distribuição com base nessa distribuição empírica. São necessários poucos pontos de dados e o tipo de distribuição não é conhecido com antecedência.

Procedimento:

- Clique em ***Risk Simulator | Ferramentas | Ajuste da distribuição (percentis)***, escolha a distribuição de probabilidade e os tipos de valores que você deseja usar, insira os parâmetros e clique em ***Executar*** para obter os resultados. Revise os resultados de R-quadrado ajustados e compare os resultados dos ajustes teóricos com os empíricos para determinar se a sua distribuição é um bom ajuste.

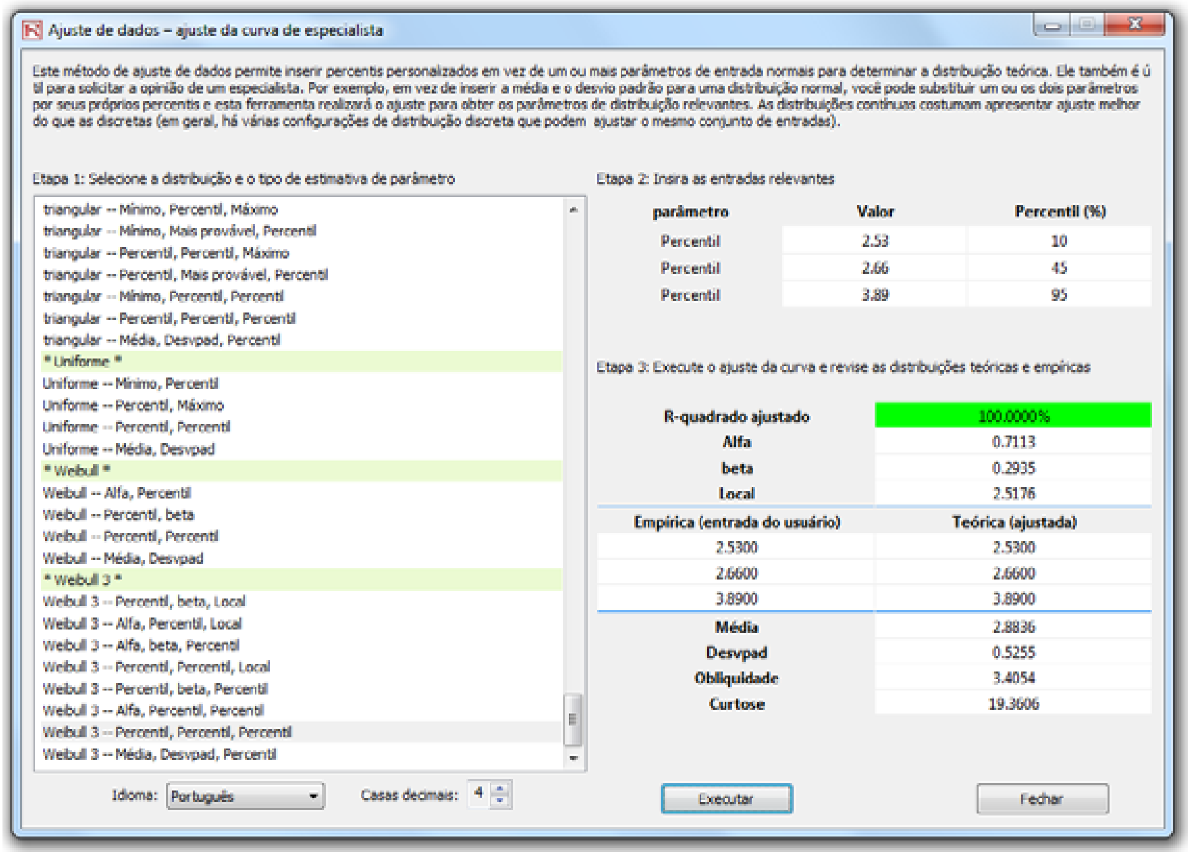


Figura 5.47 – Ferramenta de ajuste da distribuição percentil

Gráficos e tabelas de distribuição: Ferramenta de distribuição de probabilidade

Esta nova ferramenta de distribuição de probabilidade é um módulo muito rápido e poderoso, usado para gerar gráficos e tabelas de distribuição (Figuras 5.48 a 5.51). Observe que existem três ferramentas semelhantes no Risk Simulator, mas cada uma possui funções muito distintas:

- Análise da distribuição — usada para calcular rapidamente o FDP, a FDA e a IFDA das 42 distribuições de probabilidade disponíveis no Risk Simulator, e para retornar uma *tabela de probabilidades* desses valores.
- Tabelas e gráficos de distribuição — esta é a ferramenta de distribuição de probabilidade descrita neste manual e usada para comparar *parâmetros diferentes da mesma distribuição* (por exemplo, as formas e os valores de FDP, FDA e IFDA de uma distribuição Weibull com alfa e beta iguais a [2, 2], [3, 5] e [3,5, 8] e sobrepô-los um ao outro).
- Gráficos sobrepostos: usada para comparar *distribuições diferentes* (valores de entrada teóricos e previsões de saída empíricas simuladas) e sobrepô-las uma à outra para uma comparação visual.

Procedimento:

- Execute o ROV BizStats em ***Risk Simulator | Tabelas e gráficos de distribuição***, clique no botão ***Aplicar entradas globais*** para carregar um conjunto de exemplo de parâmetros de entrada ou insira seus próprios valores e clique em ***Executar*** para calcular os resultados. Os quatro momentos resultantes e a FDA, a IFDA e a FDP são calculados para cada uma das 45 distribuições de probabilidade (Figura 5.48).

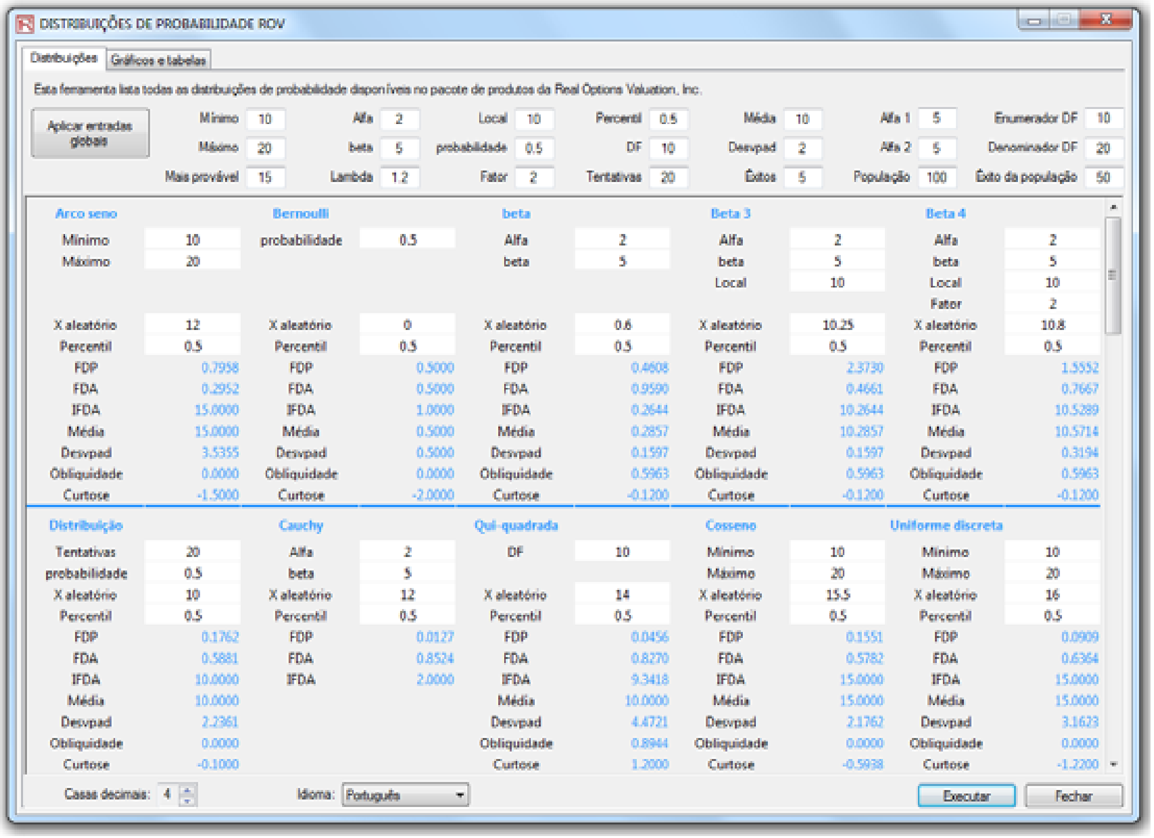


Figura 5.48 – Ferramenta de distribuição da probabilidade (45 distribuições de probabilidade)

- Clique na guia ***Gráficos e tabelas*** (Figura 5.49), selecione uma distribuição [**A**] (por exemplo, Arco seno), escolha se deseja executar a FDA, a IFDA ou a FDP [**B**], insira as entradas relevantes e clique em ***Executar gráfico*** ou em ***Executar tabela*** [**C**]. Você pode alternar entre as guias Gráfico e Tabela para ver os resultados. Também é possível experimentar alguns ícones do gráfico [**E**] para ver os efeitos no gráfico.
- Também é possível alterar dois parâmetros [**H**] para gerar vários gráficos e tabelas de distribuição inserindo a entrada ***De/Para/Incremento*** ou usar as entradas ***Personalizar*** e pressionar ***Executar***. Por exemplo, como ilustrado na Figura 5.50, execute a distribuição beta e selecione FDP [**G**], selecione o alfa e o beta a ser alterados [**H**] usando as entradas Personalizar [**I**] e insira os parâmetros de entrada relevantes: 2;5;5 para alfa e 5;3;5 para beta [**J**] e clique em ***Executar gráfico***. Serão geradas três distribuições beta [**K**]: beta (2,5), beta (5,3) e beta (5,5) [**L**]. Examine as diversas configurações de tipos de gráficos, linhas de

grade, idioma e casas decimais [**M**] e execute novamente a distribuição usando valores simulados teóricos e empíricos [**N**].

- A Figura 5.51 ilustra as tabelas de probabilidade geradas para uma distribuição binomial, na qual seleciona-se que a probabilidade de sucesso e o número de tentativas bem-sucedidas (X variável aleatório) devem variar [**O**] usando a opção ***De/Para/Incremento***. Tente replicar o cálculo como mostrado e clique na guia Tabela [**P**] para ver os resultados da função de densidade da probabilidade criados. Neste exemplo, a distribuição binomial com um conjunto de entrada inicial de tentativas = 20, probabilidade de sucesso = 0,5 e número de tentativas bem-sucedidas X = 10, onde a probabilidade de sucesso pode variar de 0, 0,25, … até 0,50 e é mostrada como a variável de linha, e o número de tentativas bem-sucedidas também pode variar de 0, 1, 2, … até 8 e é mostrado como a variável de coluna. A FDP é escolhida e, portanto, os resultados na tabela mostram a probabilidade de que o dado evento aconteça. Por exemplo, a probabilidade de obter exatamente dois sucessos em 20 tentativas, nas quais cada tentativa tem uma chance de 25% de sucesso é uma probabilidade de 0,0669 ou 6,69%.

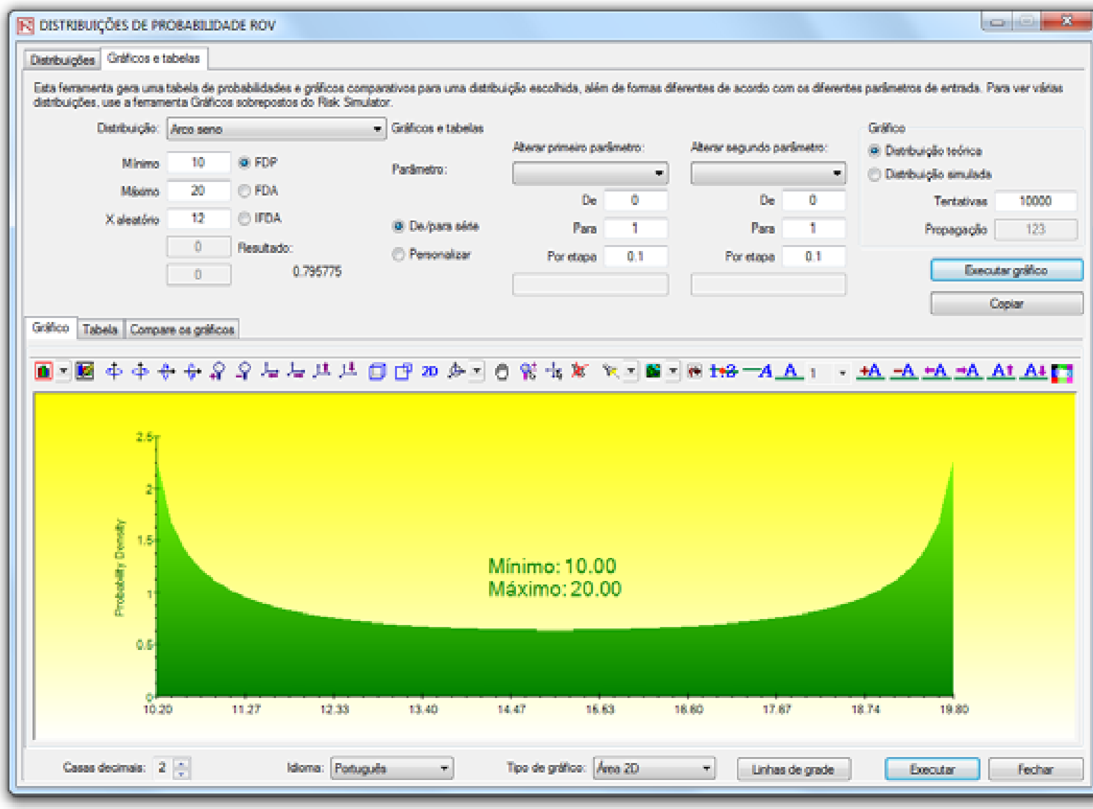


Figura 5.49 – Distribuição de probabilidade ROV (gráficos FDP e FDA)

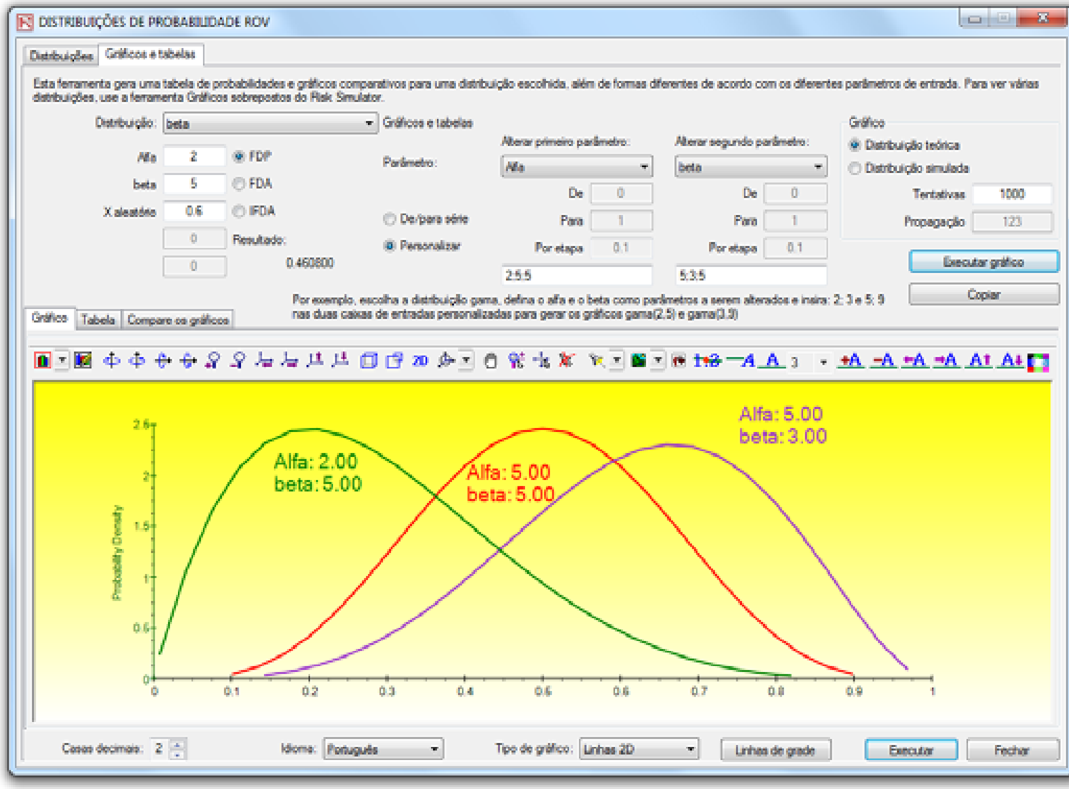
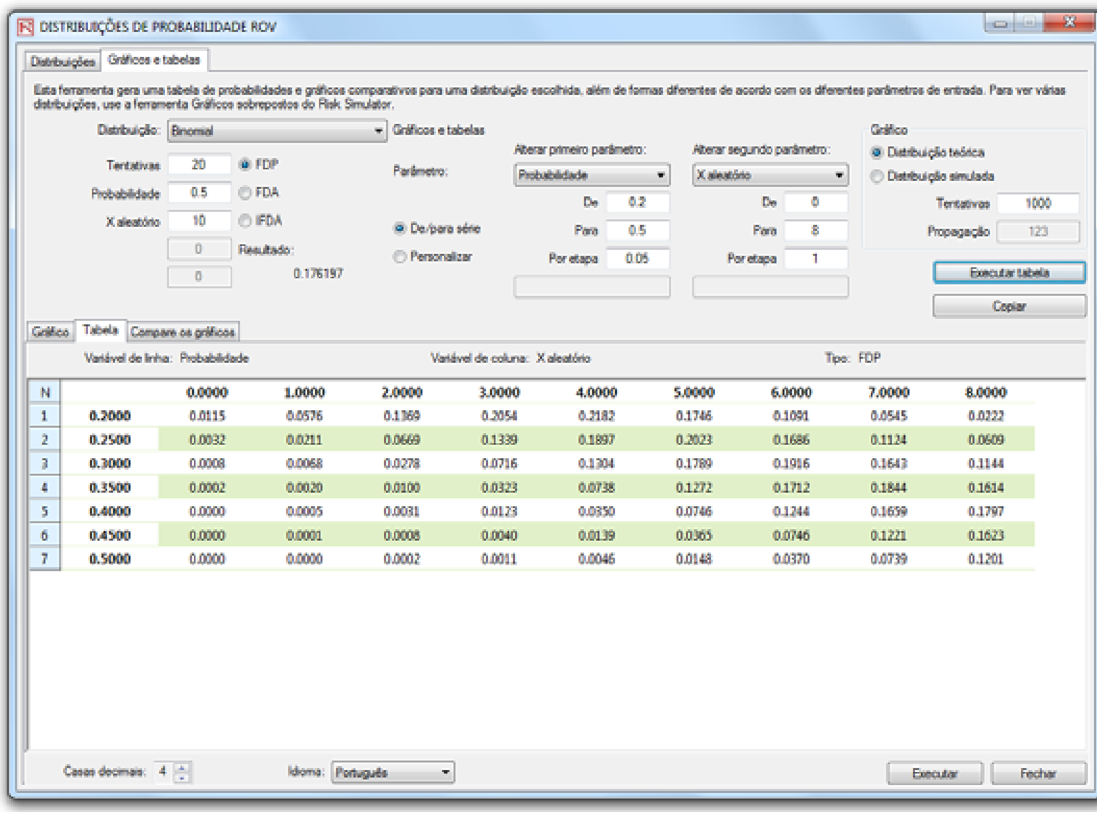


Figura 5.50 – Distribuição de probabilidade ROV (vários gráficos sobrepostos)



N		0.0000	1.0000	2.0000	3.0000	4.0000	5.0000	6.0000	7.0000	8.0000
1	0.2000	0.0115	0.0576	0.1369	0.2054	0.2182	0.1746	0.1091	0.0545	0.0222
2	0.2500	0.0032	0.0211	0.0669	0.1339	0.1897	0.2023	0.1686	0.1124	0.0609
3	0.3000	0.0008	0.0068	0.0278	0.0716	0.1304	0.1789	0.1916	0.1643	0.1144
4	0.3500	0.0002	0.0020	0.0100	0.0323	0.0738	0.1272	0.1712	0.1844	0.1614
5	0.4000	0.0000	0.0005	0.0031	0.0123	0.0350	0.0746	0.1244	0.1659	0.1797
6	0.4500	0.0000	0.0001	0.0008	0.0040	0.0139	0.0365	0.0746	0.1221	0.1623
7	0.5000	0.0000	0.0000	0.0002	0.0011	0.0046	0.0148	0.0370	0.0739	0.1201

Figura 5.51 – Distribuição de probabilidade ROV (tabelas de distribuição)

ROV BizStats

A nova ferramenta ROV BizStats é um módulo muito rápido e poderoso do Risk Simulator usado para executar estatísticas comerciais e modelos analíticos nos dados. Ela inclui mais de 130 estatísticas comerciais e modelos analíticos (Figuras 5.52 a 5.55). Veja a seguir algumas etapas iniciais sobre a execução do módulo e os detalhes sobre cada elemento do software.

Procedimento:

- Execute o ROV BizStats em ***Risk Simulator | ROV BizStats***, clique em ***Exemplo*** para carregar dados de exemplo e o perfil de modelo [**A**]. Alternativamente, você pode inserir seus dados ou copiar e colar na grade de dados [**D**] (Figura 5.52). É possível adicionar suas próprias notas ou nomes de variáveis na primeira linha Notas [**C**].
- Selecione o modelo relevante [**F**] para ser executado na Etapa 2 e, usando as configurações de entrada dos dados de exemplo [**G**], insira as variáveis relevantes [**H**]. Separe as variáveis para o mesmo parâmetro usando ponto-e-vírgula e use uma nova linha (pressione Enter para criar uma nova linha) para parâmetros diferentes.
- Clique em ***Executar*** [**I**] para calcular os resultados [**J**]. Você pode exibir as estatísticas, os gráficos ou os resultados relevantes da análise nas diversas guias na Etapa 3.
- Se necessário, forneça um nome de modelo para salvar no perfil, na Etapa 4 [**L**]. É possível salvar vários modelos no mesmo perfil. Os modelos existentes podem ser editados ou excluídos [**M**], reordenados [**N**] e todas as alterações podem ser salvas [**O**] em um único perfil com a extensão de nome de arquivo *.bizstats.

Notas:

- O tamanho da grade de dados pode ser definido no menu e a grade pode acomodar até mil colunas de variáveis com um milhão de linhas de dados por variável. O menu também permite que você altere as configurações de idioma e de casas decimais para os seus dados.
- Para começar, é recomendável carregar o arquivo de exemplo [**A**] bastante abrangente que inclui dados e alguns modelos pré-criados [**S**]. Clique duas vezes em cada um desses modelos para executá-los. Os resultados são mostrados na área de relatório [**J**], que algumas vezes pode ser um gráfico ou as estatísticas do modelo [**T/U**]. Usando o arquivo de exemplo, você pode ver como os parâmetros de entrada [**H**] são inseridos de acordo com a descrição do modelo [**G**] e pode continuar a criar seus próprios modelos.
- Clique nos cabeçalhos de variável [**D**] para selecionar uma ou várias variáveis ao mesmo tempo. Clique com o botão direito do mouse para adicionar, excluir, copiar, colar ou exibir [**P**] as variáveis selecionadas.
- Os modelos também podem ser inseridos por um console de comando [**V/W/X**]. Para ver seu funcionamento, clique duas vezes para executar um modelo [**S**] e vá para o console

de comando [**V**]. É possível replicar o modelo ou criar um modelo próprio e clicar em Executar comando [**X**] quando você estiver pronto. Cada linha do console representa um modelo e seus parâmetros relevantes.

- Todo o perfil **.bizstats* (no qual os dados e os vários modelos são criados e salvos) pode ser editado diretamente em XML [**Z**]. Para fazer isso, abra o editor de XML no menu Arquivo. As alterações no perfil podem ser feitas por meio de programação e entrarão em vigor quando o arquivo for salvo.

Dicas:

- Click no cabeçalho da(s) coluna(s) para selecionar a(s) coluna(s) inteira(s) ou variável(eis)e, uma vez selecionado, pressione o botão da direita do seu mouse no cabeçalho para *Auto Ajuste* auto ajuste da coluna, *Cortar*, *Copiar*, *Apagar*, ou *Colar* dados. Você poderá também clicar nos múltiplos cabeçalhos de colunas e selecioná-las e, com o botão da direita do mouse, escolher e opção *Visualizar*, para visualizar os gráficos.
- Caso a tenha um valor que é completamente visualizado, clique sobre a célula e paire com o ponteiro sobre a mesma de forma visualizar o conteúdo completo da célula, ou simplesmente redimensione a mesma para que todo o valor seja visualizado normalmente (ou arraste a borda esquerda no cabeçalho da coluna da célula desejada, para aumentá-la ou dê dois cliques na mesma borda, ou então com o botão da direita acionado escolha Auto Ajuste)
- Use as teclas acima, abaixo, esquerda, direta para mover-se em torno da célula, ou utiliza as teclas *Home* e *End* do seu teclado para movimentos mais amplos nas linhas a serem editadas. Também é possível utilizar a combinação de teclas, tais como: *Ctrl+Home* para pular para o topo da coluna posicionada à esquerda, que tenha alguma célula com algum valor, *Ctrl+End* para pular para o extremo direito inferior da coluna que contenha algum valor registrado. Da mesma forma, acionando o comando *Shift+Up/Down* seleciona uma área específica.
- É possível inserir pequenas observações ou notas para cada variável na linha de Notas. Lembre-se de fazer suas notas curtas e simples.
- Tente usar os diversos ícones de gráfico, no modo Visualizar, para alterar a visão dos mesmos e observar os gráficos (p.ex., Rotação, Deslocamento, Aumento, Colorir, e adicionar Legendas, e outras)
- O botão *Copiar* é utilizado para copiar nas abas *Resultados*, *Gráficos* e *Estatísticas*, no *Passo 3*, após o modelo tiver "rodado". Caso nenhum modelo estiver rodando, então a função de cópia irá apenas copiar uma página em branco.
- O botão Relatório somente será ativado caso existam modelos salvos no *Passo 4*, ou caso existam dados da grade, caso contrário e relatório gerado estará vazio. Você também necessitará que o Excel esteja instalado com a capacidade de extração de dados produção de relatórios dos resultados, e que o PowerPoint da Microsoft esteja disponível para rodar os relatórios Gráficos.
- Caso tenha dúvida sobre como "rodar" um modelo específico ou um método estatístico, inicie no perfil *Exemplo* e reveja como os dados são organizados, no *Passo 1*, ou como os parâmetros são informados no *Passo 2*. Você pode utilizar esses modelos de exemplo para adaptar seus próprios dados e modelos
- O Idioma pode ser alterado no Menu *Idioma*. Observe que atualmente existem 10 idiomas disponíveis para este software, com perspectivas de novos idiomas serem acrescidos em breve. Contudo, alguns resultados, por terem um entendimento universal, ainda serão visualizados em Inglês.
- É possível alterar a forma de como os modelos no *Passo2* são exibidos mudando a lista apresentada em *Ver*. Você pode listar os modelos alfabeticamente, por categoria, e por

requisitos dos dados de entrada – observar que em certos idiomas em Unicode (p.ex., Chinês, Japonês e Coreano), não existe ordenamento alfabético e, desse modo, a primeira opção é inválida.

- O software pode manipular diferentes arranjos regionais para números e pontuação decimal (p.ex., um mil dólares e cinquenta centavos pode ser escrito como 1,000.50 ou 1.000,50, ou 1’000,50, etc.). O arranjo decimal pode ser definido no ROV BizStats no menu ***Dado | Arranjo Decimal***. Contudo, quando houver dúvida, mantenha no ROV BizStat a definição regional coo a Inglesa Americana e o padrão default norte americanos 1,000.50 (esse padrão é comprovadamente operacional e considerado nos exemplos do ROV BizStats).

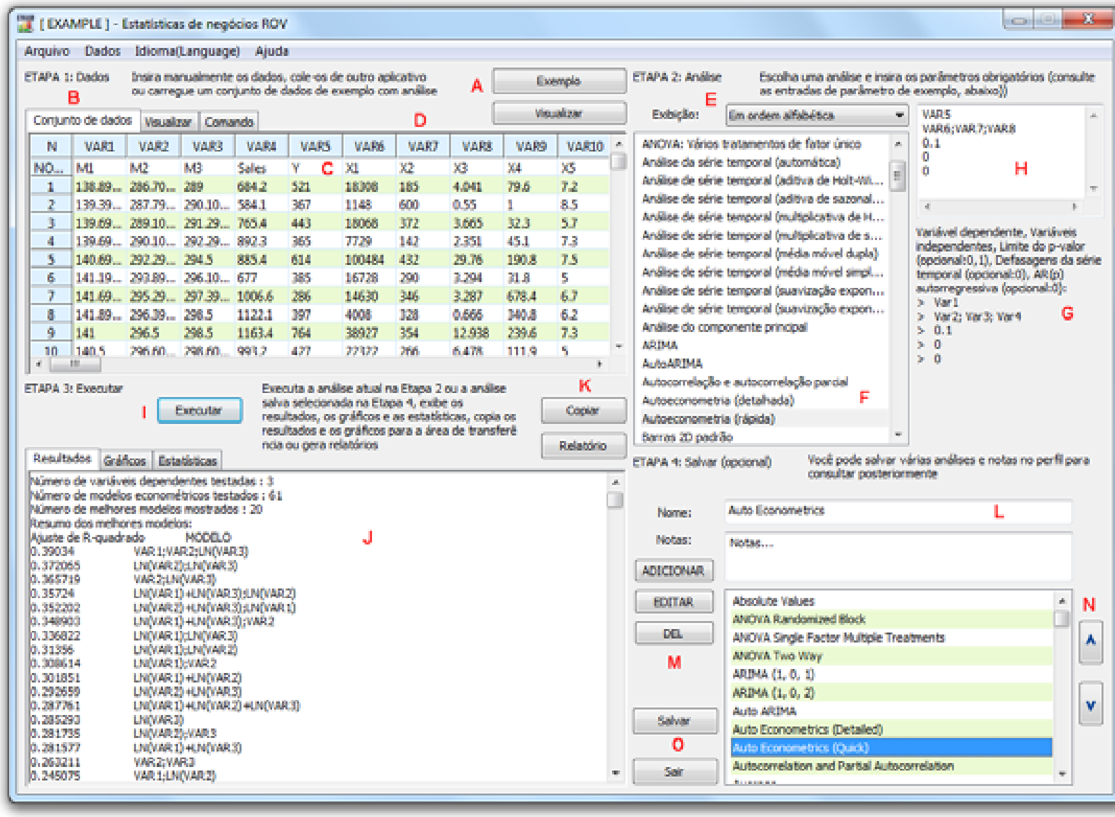


Figura 5.52 – ROV BizStats (análise estatística)

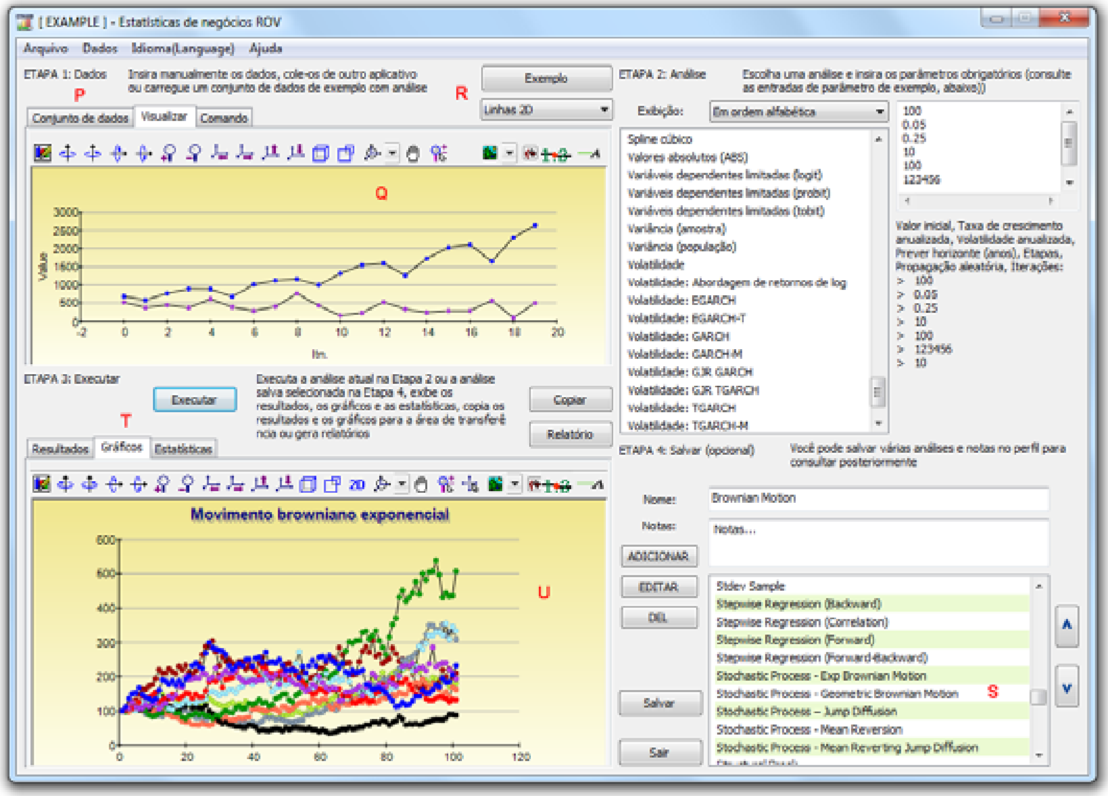


Figura 5.53 – ROV BizStats (visualização de dados e gráficos de resultados)

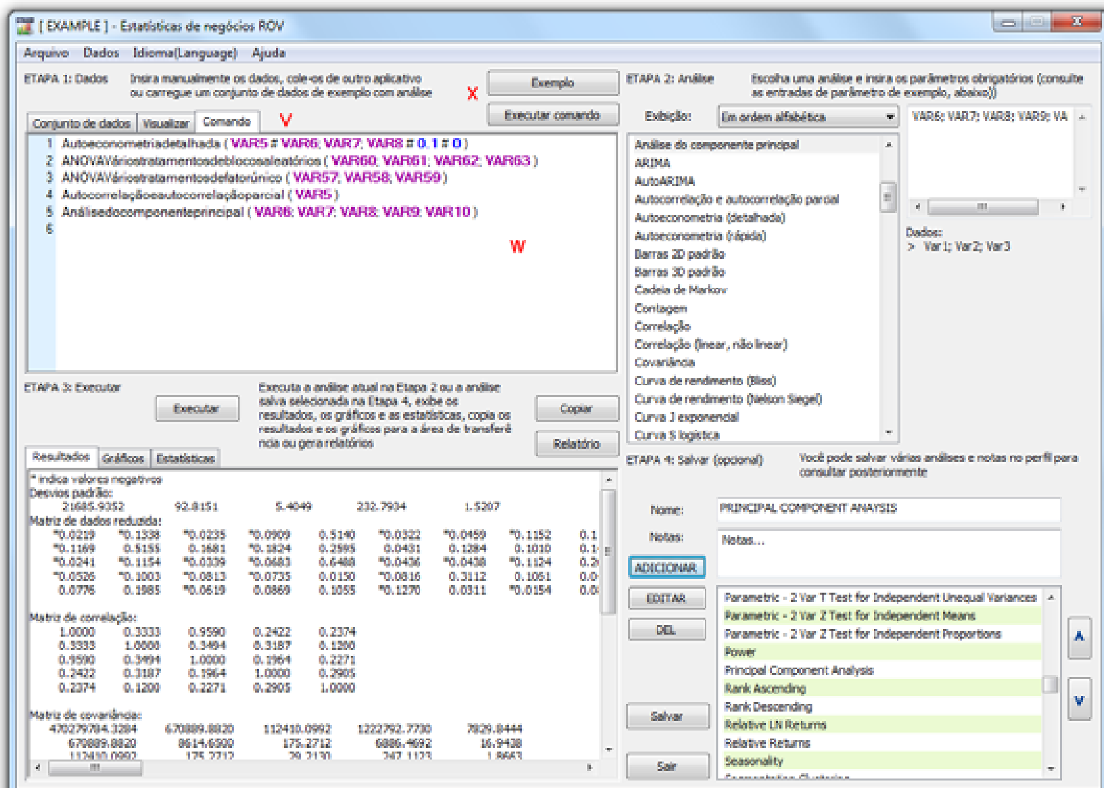


Figura 5.54 – ROV BizStats (console de comando)

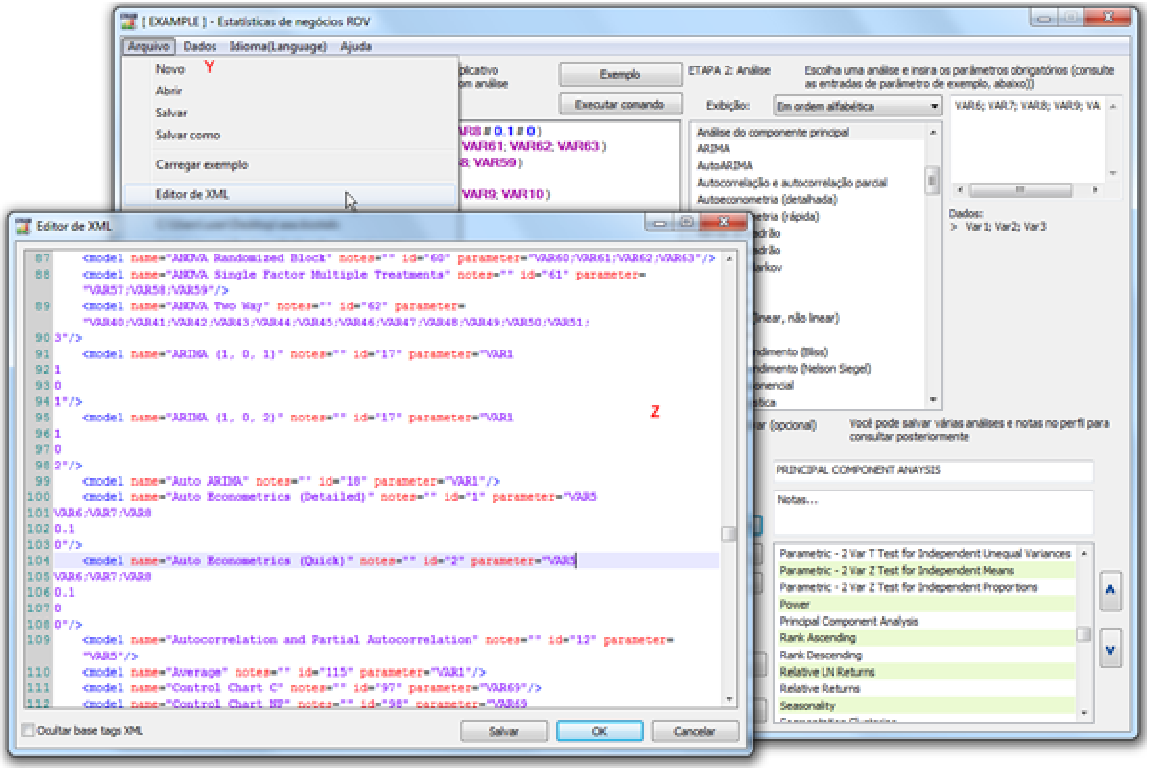


Figura 5.55 – ROV BizStats (editor de XML)

Metodologias de previsão de lógica difusa combinatória e de rede neural

O termo “rede neural” é tradicionalmente usado para se referir a uma rede ou um circuito de neurônios biológicos, ao passo que o uso moderno do termo se refere às redes neurais artificiais que compreendem neurônios artificiais ou nós recriados em um ambiente de software. Essa metodologia tenta imitar os neurônios ou o cérebro humano no modo de pensar e identificar padrões, em nosso caso, para identificar padrões com a finalidade de prever dados de série temporal. Essa metodologia é encontrada no módulo **ROV BizStats** do Risk Simulator, localizado em **Risk Simulator | ROV BizStats | Rede neural**, bem como em **Risk Simulator | Previsão | Rede neural.** A Figura 5.56 mostra a metodologia de previsão de rede neural.

Procedimento

- Clique em **Risk Simulator | Previsão | Rede neural.**
- Para começar, insira manualmente os dados ou cole-os da área de transferência (por exemplo, selecione e copie dados no Excel, inicie a ferramenta e clique no botão *Colar* para colar os dados)
- Selecione se você deseja executar um modelo de rede neural *Linear* ou *Não linear*, insira o número desejado de *Períodos de previsão* (por exemplo, 5), o número de *Camadas* ocultas na rede neural (por exemplo, 3) e o número de *Períodos de teste* (por exemplo, 5).
- Clique em *Executar* para iniciar a análise e revise os gráficos e os resultados calculados. Também é possível *Copiar* os resultados e o gráfico para a área de transferência e colá-los em outro aplicativo.

Observe que o número de camadas ocultas na rede é um parâmetro de entrada e deverá ser calibrado com seus dados. Em geral, quanto mais complicado o padrão de dados, maior o número de camadas ocultas necessárias e mais demorado será o cálculo. Recomenda-se começar com 3 camadas. Os períodos de teste são simplesmente o número de pontos de dados usados na calibração final do modelo de rede neural. Recomenda-se usar como período de teste pelo menos o mesmo número de períodos que se deseja prever.

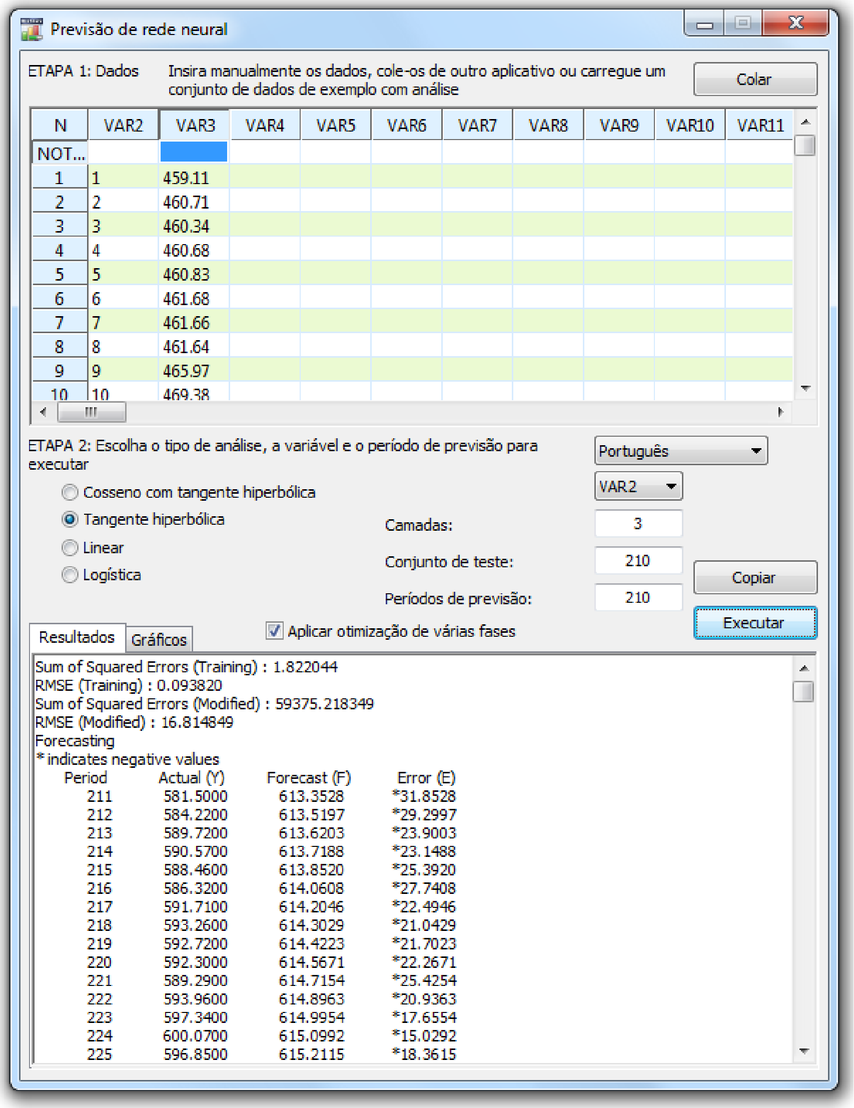


Figura 5.56 – Previsão de rede neural

Em contraste, o termo “lógica difusa” é derivado da teoria de conjuntos difusa para lidar com o raciocínio que seja aproximado e não exato. Essa lógica se contrapõe à “lógica nítida”, na qual os conjuntos binários têm lógica binária. As variáveis de lógica difusa podem ter um valor verdadeiro que varie entre 0 e 1 e não seja restrito aos dois valores verdadeiros da lógica proposicional clássica. Esse esquema de ponderação difuso é usado com método combinatório para produzir resultados de previsão de série temporal no Risk Simulator, como ilustrado na Figura 5.57, e é mais adequado quando aplicado a dados de série temporal com sazonalidade e tendência. Essa metodologia é encontrada no módulo **ROV BizStats** do Risk Simulator, localizado em **Risk Simulator | ROV BizStats | Lógica difusa combinatória**, bem como em **Risk Simulator | Previsão | Lógica difusa combinatória.** A Figura 5.57 mostra a metodologia de previsão de rede neural.

Procedimento

- Clique em **Risk Simulator | Previsão | Lógica difusa combinatória.**

- Para começar, insira manualmente os dados ou cole-os da área de transferência (por exemplo, selecione e copie dados no Excel, inicie a ferramenta e clique no botão *Colar* para colar os dados)
- Na lista suspensa, selecione a variável na qual deseja executar a análise, insira o período de sazonalidade (por exemplo, 4 para dados trimestrais, 12 para dados mensais etc.) e o número desejado de *Períodos de previsão* (por exemplo, 5).
- Clique em *Executar* para iniciar a análise e revise os gráficos e os resultados calculados. Também é possível *Copiar* os resultados e o gráfico para a área de transferência e colá-los em outro aplicativo.

Observe que as técnicas de redes neurais e de lógica difusa ainda não foram estabelecidas como um método válido e confiável no domínio das previsões de negócios, seja no nível estratégico, tático ou operacional. Ainda é necessário fazer muitas pesquisas sobre esses campos de previsão avançada, entretanto, o Risk Simulator fornece noções básicas sobre essas duas técnicas com a finalidade de executar previsões de séries temporais. Recomenda-se que você não use essas técnicas isoladamente, mas em combinação com as outras metodologias de previsão do Risk Simulator a fim de construir modelos mais robustos.

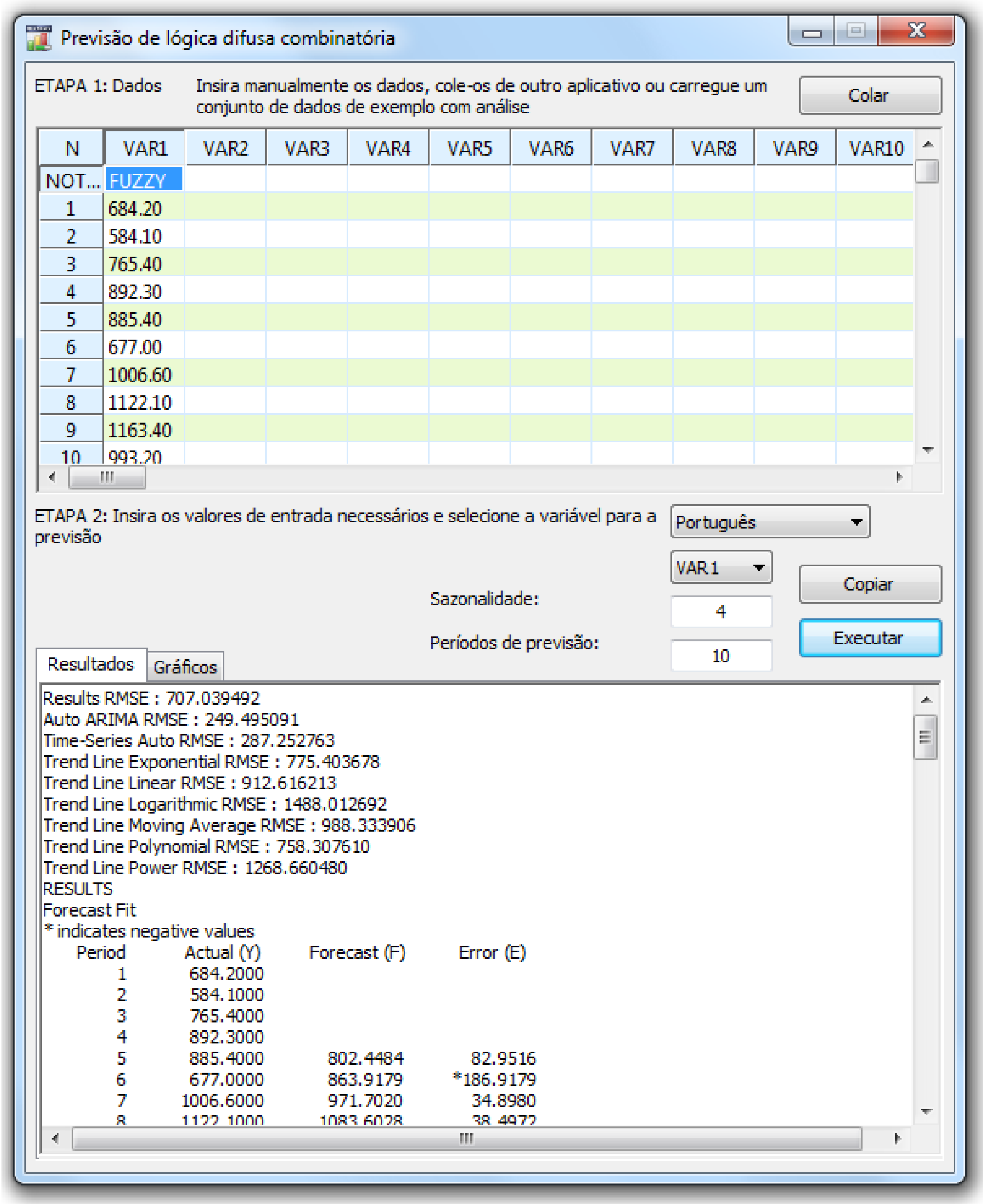


Figura 5.57 – Previsão de série temporal de lógica difusa

Otimizador de Buscar meta

A ferramenta Buscar meta é um algoritmo de pesquisa aplicado para encontrar a solução de uma única variável dentro de um modelo. Se você conhece o resultado que deseja de uma fórmula ou um modelo, mas não tem certeza sobre o valor de entrada de que a fórmula precisa para obter esse resultado, use o recurso **Risk Simulator | Ferramentas | Buscar meta**. Observe que Buscar meta funciona apenas com um valor de entrada variável. Caso deseje aceitar mais de um valor de entrada, use as rotinas de otimização avançada do Risk Simulator. A Figura 5.58 mostra um modelo simples e a aplicação de Buscar meta.

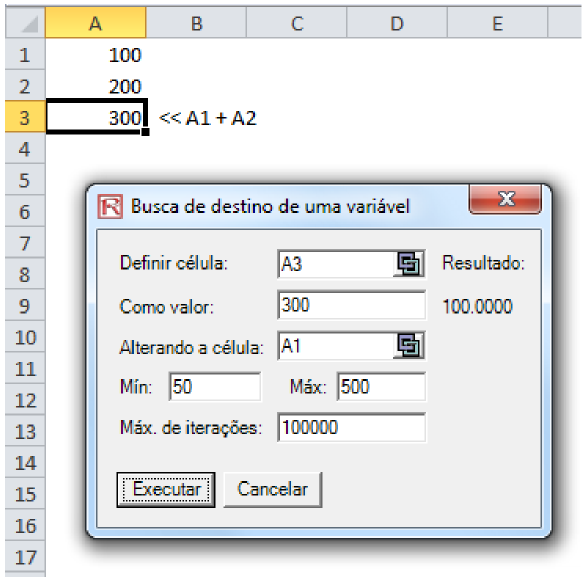


Figura 5.58 – Buscar meta

Otimizador de variável única

A ferramenta Otimizador de variável única é um algoritmo de pesquisa usado para encontrar a solução de uma única variável dentro de um modelo, como a rotina de buscar meta abordada anteriormente. Caso você deseje o máximo ou o mínimo resultado possível de um modelo, mas não tenha certeza quanto ao valor de entrada de que a fórmula precisa para obter esse resultado, use o recurso **Risk Simulator | Ferramentas | Otimizador de variável única** (Figura 5.59). Observe que o Otimizador de variável única é executado muito rapidamente, mas só pode ser usado para encontrar uma entrada variável. Caso deseje aceitar mais de um valor de entrada, use as rotinas de otimização avançada do Risk Simulator. A ferramenta foi incluída no Risk Simulator, porque ocasionalmente pode ser necessário executar um cálculo de otimização rápido para uma única variável de decisão e a ferramenta fornece essa capacidade sem exigir a configuração de um modelo de otimização com perfis, suposições de simulação, variáveis de decisão, objetivos e restrições.

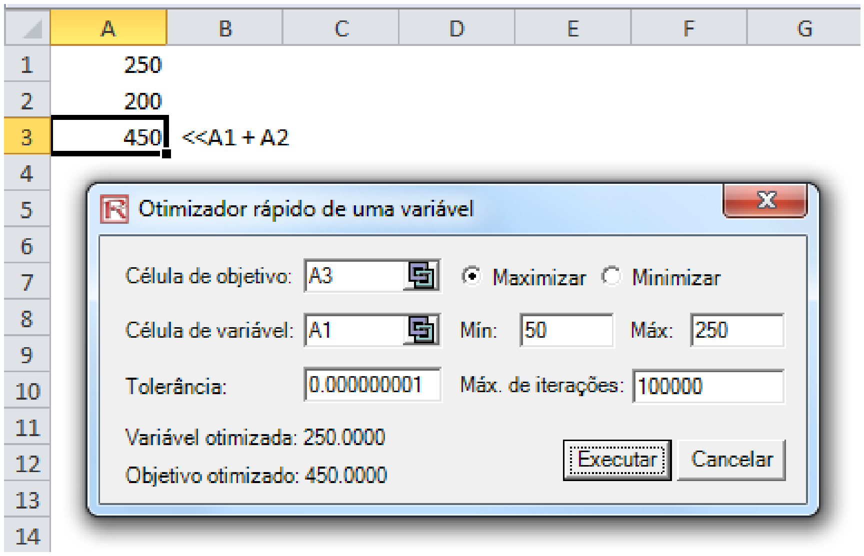


Figura 5.59 – Otimização de variável única

Otimização de algoritmo genético

O algoritmo genético é um algoritmo heurístico de pesquisa que imita o processo da evolução natural. Ele é usado com frequência para gerar soluções úteis para problemas de otimização e de pesquisa. Os algoritmos genéticos pertencem à classe geral de algoritmos evolucionários, que geram soluções para problemas de otimização usando técnicas inspiradas na biologia evolutiva, como hereditariedade, mutação, seleção natural e recombinação.

O algoritmo genético está disponível em **Risk Simulator | Ferramentas | Algoritmo genético** (Figura 5.60). Tome cuidado durante a calibração dos valores de entrada do modelo, pois os resultados são muito sensíveis a esses valores (os valores de entrada padrão são fornecidos como orientação geral para os níveis de entrada mais comuns). Recomenda-se selecionar o Teste de pesquisa de gradiente para obtenção de um conjunto de resultados mais robusto (é possível cancelar a seleção dessa opção para começar e, em seguida, selecioná-la, executar a análise novamente e comparar os resultados).

Nota: em muitos problemas, os algoritmos genéticos podem apresentar a tendência de convergir para um valor ótimo local ou até para pontos arbitrários em vez do valor ótimo global do problema. Isso significa que ele não sabe como sacrificar o ajuste de curto prazo para obter o ajuste de longo prazo. Para problemas de otimização específicos e instâncias de problemas, outros algoritmos de otimização podem encontrar soluções melhores do que os algoritmos genéticos (considerando-se o mesmo prazo de cálculo). Portanto, é recomendável primeiro executar o algoritmo genético e, depois, executá-lo novamente marcando a opção *Aplicar teste de pesquisa de gradiente* (Figura 5.60) para verificar a robustez do modelo. Esse teste de pesquisa de gradiente tenta executar combinações de técnicas de otimização tradicionais com métodos de algoritmo genético para retornar a melhor solução possível. Por fim, a menos que haja uma necessidade teórica específica para usar o algoritmo genético, recomendamos o uso do módulo

Otimização do Risk Simulator para obter resultados mais robustos. Esse módulo permite que você execute rotinas mais avançadas de otimização estocástica e dinâmica com base no risco.

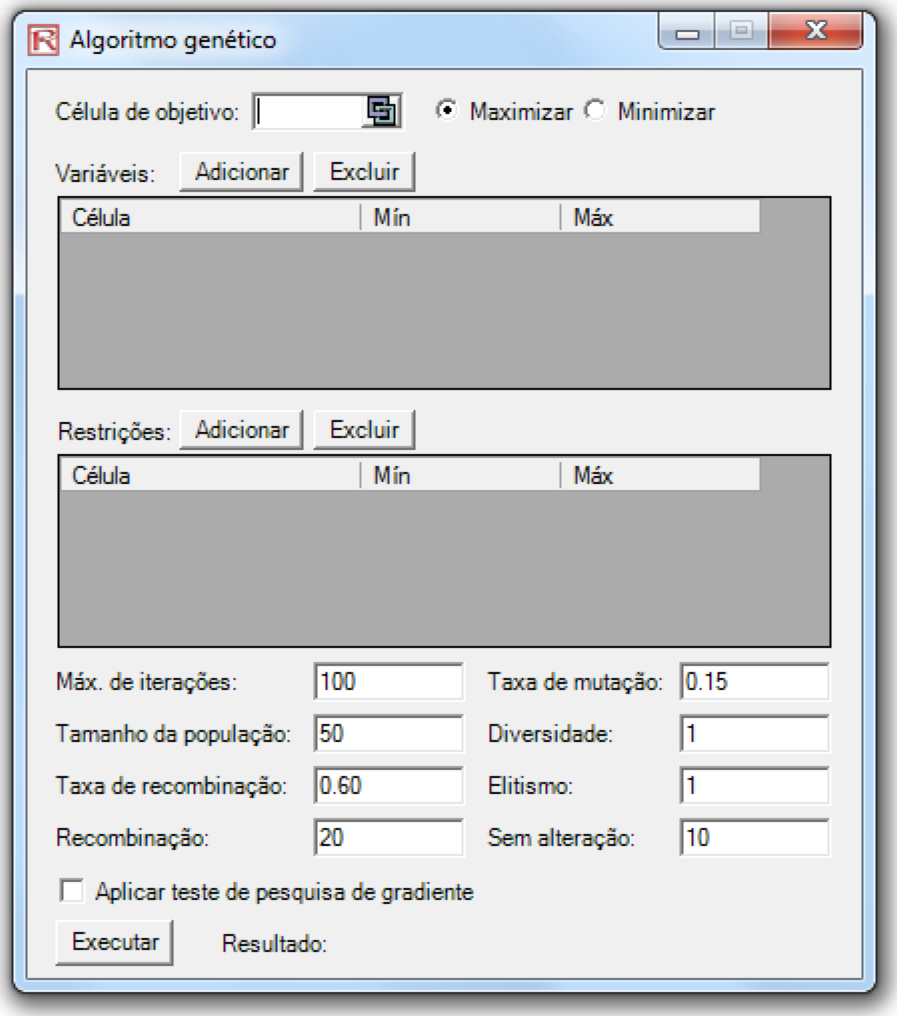


Figura 5.60 – Algoritmo genético

Módulo ROV Decision Tree (Árvore de Decisão)

Árvore de Decisão

ROV Decision Tree (Árvore de Decisão, Figura 5.61) é usado para criar valor e avaliar modelos de árvore de decisão. As seguintes metodologias e análises avançadas adicionais também estão incluídas:

- Modelos de Árvore de Decisão
- Simulação de risco (Monte Carlo)
- Análise de Sensibilidade
- Análise de Cenário
- Análise Bayesiana (probabilidade conjunta e a posteriori)
- Valor Esperado da Informação
- MINIMAX
- MAXIMIN
- Perfis de Risco

A seguir, estão alguns das principais dicas rápidas e procedimentos para a utilização intuitiva dessas ferramentas:

- Existem 11 idiomas disponíveis neste módulo e o idioma atual pode ser alterado através do menu Idiomas.
- *Inserir ou Inserir nó de Opção ou nó Terminal, selecionando primeiro qualquer nó existente e, em seguida, clicando no ícone nó de Opção (quadrado) ou no ícone do nó Terminal (triângulo), ou usar as funções do menu Inserir.*
- Modificar individualmente as propriedades do Nó de Opção ou do Nó Terminal, clicando duas vezes em um nó. Às vezes, quando você clica em um nó, todos os nós filhos subsequentes também são selecionados (isso permite que você mova toda a árvore a partir daquele nó selecionado). Se você deseja selecionar apenas um nó, você pode ter que clicar sobre o fundo branco e clicar novamente no nó para selecioná-lo individualmente. Além disso, você pode mover os nós individuais ou toda a árvore que começou a partir do nó selecionado, dependendo da configuração atual (botão direito do mouse, ou no menu Editar e selecione Mover nós individualmente ou Mover Juntos nós).
- A seguir são apresentadas algumas descrições rápidas das coisas que podem ser personalizadas e configuradas na interface do usuário das propriedades do nó. É mais simples para testar diferentes configurações para cada um dos seguintes para ver seus efeitos na Árvore de Estratégia:
    - *Nome. Nome acima do nó.*
    - *Valor. Valor a ser mostrado abaixo do nó.*
    - *Concatenar com Excel. Liga o valor à uma célula da planilha Excel.*
    - *Notas. Notas podem ser inseridas acima ou abaixo do nó.*
    - *Mostrar no Modelo. Mostrar qualquer combinação de Nome, Valor e Notas.*

  - *Cor Local versus Cor Global. As cores dos nós podem ser trocadas localmente ou globalmente.*
  - *Rótulo dentro da Forma. O texto pode ser colocado dentro do nó (você pode precisar fazer o nó maior para acomodar mais texto).*
  - *Nome Ramo do Evento. O texto pode ser colocado sobre o ramo que conduz ao nó, para indicar o principal evento para este nó.*
  - *Selecione Opções Reais. Um tipo de opção real específica pode ser atribuído ao nó atual. Atribuindo opções reais para nós permite que a ferramenta para gere uma lista de variáveis de entradas necessárias.*
- *Os Elementos Globais são todos personalizáveis, incluindo elementos de fundo da Árvore Estratégia, Linhas de conexão, nós de Opção, nós terminais e caixas de texto. Por exemplo, as seguintes configurações podem ser alteradas para cada um dos elementos:*
  - *Propriedades da Fonte (nome, valor, notas, rótulo, e nomes do evento).*
  - *Tamanho do Nó*
  - *Margens (estilo de linha, grossura e cor)*
  - *Sombra (cores e tudo onde puder ser aplicada sombra)*
  - *Cor Global*
  - *Forma Global*
- O comando Requisitos de Dados para a Modelagem, no menu Editar, abre uma janela ancorada no lado direito da Árvore de Estratégia de tal forma, que quando um nó de Opção ou nodo Terminal for selecionado e qualificado, as propriedades desse nó serão exibidas e podem ser atualizadas diretamente. Este recurso fornece uma alternativa ao duplo-clique em um nó de cada vez.
- *Arquivos de exemplo estão disponíveis no menu Arquivo, para ajudar a você começar a construir Árvores de Estratégia.*
- *O item Proteger Arquivo no menu Arquivo, permite que a Árvore de Estratégia seja criptografada, gerando uma senha criptografada com até 256 bits. Tenha cuidado quando um arquivo está sendo codificado, porque se a senha for perdida, o arquivo não mais poderá ser aberto.*
- *Capturar tela ou imprimir o modelo existente pode ser feito através do menu Arquivo. A capturada de tela pode então ser colada em outros software aplicativos.*
- *Adicionar, Duplicar, Renomear e Excluir uma Árvore de Estratégia pode ser realizada através de botão direito do mouse na guia Árvore de Estratégia ou no menu Editar.*
- *Você pode Mudar Estilos Existentes, ou Gerenciar e Criar Estilos Personalizados de sua Árvore de Estratégia (o que inclui tamanho, forma, esquemas de cores e tamanho da fonte / cor das especificações da Árvore de Estratégia inteira).*
- Inserir nó de Decisão, nó de Incerteza, ou nó Terminal selecionando qualquer nó existente e, em seguida, clicando no ícone de nó de Decisão (quadrado), ícone Incerteza nó (círculo), ou no ícone de nó Terminal (triângulo), ou use as funcionalidades do menu Inserir
- Modificar decisão individual, incerteza, ou propriedades do nó Terminal, clicando duas vezes em um nó. A seguir estão alguns itens adicionais originais no módulo de Árvore

de Decisão que podem ser personalizados e configurados na interface do usuário, referente à propriedade do nó:

  - Nós de Decisão: Sobrescrever personalizado ou Auto Calcular o valor em um nó. A opção de calcular automaticamente é definida como padrão, e quando você clica em Executar em um modelo completo de Árvore de Decisão, os nós de decisão serão atualizados com os resultados.
  - Nós de Incerteza: Nomes de Eventos, Probabilidades, e conjuntos de parâmetros da Simulação. Você somente pode adicionar nomes de eventos de probabilidade, probabilidades e hipóteses de simulação somente após a criação dos ramos de incerteza.
  - Nós Terminal: Entrada manual, Link do Excel, e conjuntos de Pressupostos Simulação. Os Payoffs do evento Terminal podem ser inseridos manualmente ou ligados a uma célula do Excel (por exemplo, se você tiver um modelo Excel grande, que calcula o retorno, é possível vincular o modelo à célula do resultado do modelo Excel) ou pode definir hipóteses de distribuição de probabilidade para executar as simulações.

- *Ver a janela Propriedades do Nó está disponível no menu Editar e as propriedades do nó selecionado serão atualizadas quando um nó for selecionado.*
- *O módulo de Arvore de Decisão também vem com as seguintes análises avançadas:*
  - *Modelagem de Simulação Monte Carlo nas Árvores de Decisão*
  - *Análise Bayesiana para a obtenção de probabilidades a posteriori*
  - *Valor Esperado da Informação Perfeita, Análise MINIMAX e MAXIMIN, Perfis de Risco e Valor da Informação Imperfeita*
  - *Análise de Sensibilidade*
  - *Análise de Cenário*
  - *Análise de Função de Utilidade*

Modelagem da Simulação

Essa ferramenta é executada a simulação de Risco Monte Carlo na Árvore de Decisão (**Figura 5.62**). Esta ferramenta permite que você defina os parâmetros de entrada das distribuições de probabilidade para a execução de simulações. Você tanto pode definir um pressuposto para o nó selecionado ou definir um novo pressuposto e usar este novo pressuposto (ou usar o previamente criado pressupostos) em uma equação numérica ou fórmula. Por exemplo, você pode definir um novo pressuposto chamado Normal (p.ex., distribuição normal com uma média de 100 e desvio padrão de 10) e executar uma simulação na árvore de decisão, ou usar esta hipótese em uma equação como (100 * Normal + 15,25).

Análise Bayesiana

Esta ferramenta de análise bayesiana (**Figura 5.63**) pode ser feita com quaisquer dois eventos incertos, que estão ligados por um caminho. No exemplo à direita, as incertezas A e B estão ligadas, onde, na linha do tempo, o evento A ocorre primeiro e o evento B ocorre após o primeiro.

O evento A é uma "Pesquisa de Mercado" com dois resultados possíveis (Favorável ou Desfavorável). O segundo evento B está associado às "Condições de Mercado", também com dois resultados (Forte e Fraco). Esta ferramenta é usada para calcular as probabilidades Bayesianas: conjunta, marginal, e a posteriori, informado as probabilidades a priori e as probabilidades condicionais; ou, as probabilidades condicionais podem ser calculadas quando você tem as probabilidades a posteriori condicionais. Selecione abaixo a análise desejada e clique em Carregar Exemplo, para ver uma amostra das entradas correspondente à análise selecionada e os resultados na grade à direita, bem como quais resultados são utilizados como insumos na árvore de decisão da figura.

- PASSO 1: Entre com os nomes do primeiro e segundo evento de incerteza e escolha quantos eventos probabilísticos (ou estados da natureza ou resultados) cada evento tem.
- PASSO 2: Entre com os nomes de cada evento de probabilístico ou resultado.
- PASSO 3: Entre com o segundo evento de probabilidade a priori e a probabilidade condicional para cada evento ou resultado. As probabilidades devem somar 100%.

Valor Esperado da Informação Perfeita, Análise Minimax e Maximin, Perfis de Risco e Valor da Informação Imperfeita

Esta ferramenta calcula o Valor Esperado da Informação Perfeita (VEIP), AnáliseMinimax e Maximin, bem como o Perfil de Risco e o Valor da informação imperfeita (Figura DT4). Para começar, insira o número de ramos de decisão ou estratégias em consideração (por exemplo, construir uma grande instalação, média ou pequena instalação) e o número de eventos incertos ou resultados dos estados da natureza (por exemplo, mercado bom, mercado ruim), e digite as recompensas esperadas (Payoffs) em cada cenário.

O Valor Esperado da Informação Perfeita (VEIP), ou seja, supondo que se tenha uma previsão perfeita e se saiba exatamente o que fazer (através de pesquisa de mercado ou outros meios, para melhor discernir os resultados probabilísticos), calcular se há valor adicionado em tais informações (isto é, se a pesquisa de mercado irá adicionar valor), em comparação com estimativas probabilísticas mais ingênuas dos estados da natureza. Para começar, insira o número de ramos de decisão ou de estratégias em consideração (p.ex., construir uma grande instalação, média ou pequena instalação) e o número de eventos incertos ou estados da natureza (p.ex., mercado bom, mercado ruim), e digite as recompensas (Payoffs) esperadas em cada cenário.

Minimax (minimizar o arrependimento máximo) e Maximin (maximizar o retorno ou Payoff mínimo) são duas abordagens alternativas para encontrar o caminho decisão ótima. Estas duas abordagens não são usadas frequentemente, mas ainda fornecem uma visão adicional, no processo de tomada de decisão. Digite o número de alternativas de decisão ou caminhos que existem (p.ex., a construção de uma instalação de grande porte, médio ou pequeno porte), bem como os eventos incerteza ou estados da natureza em cada caminho (p.ex., economia boa versus economia ruim). Em seguida, complete a tabela de Payoffs para os vários cenários e calcule os resultados Minimax e Maximin. Você também pode clicar em Carregar Exemplo para ver um exemplo de cálculo.

Sensibilidade

Análise de Sensibilidade (**Figura 5.65**) sobre as probabilidades de entrada é realizada para se determinar seus impactos sobre os valores obtidos no caminho de decisão. Primeiro selecione um dos Nós de Decisão abaixo, depois a probabilidade de um dos eventos da lista para teste. Se existem múltiplos eventos de incerteza com probabilidades idênticas, elas devem ser analisadas de forma independente ou concomitante.

Os gráficos de sensibilidade mostram os valores dos caminhos de decisão, sob níveis variáveis de probabilidade. Os valores numéricos são mostrados na tabela de resultados. A localização de linhas cruzadas, se for o caso, representa, nos eventos probabilísticos, que um determinado caminho decisão se tornou dominante sobre o outro.

Tabela de Cenários

As tabelas de cenários (**Figura 5.66**) podem ser geradas para determinar os resultados, dadas algumas alterações na entrada. Você pode escolher um ou mais Caminhos de Decisão para analisar (os resultados de cada caminho escolhido será representado como uma tabela e gráfico em separado), e um ou dois Nós de Incerteza ou Terminal, como variáveis de entrada para a tabela de cenários.

- Selecione um ou mais caminhos de Decisão para analisar a partir da lista abaixo.
- Selecione um ou dois Eventos de Incerteza ou Payoffs do Terminal para modelar.
- Decida se você deseja mudar a probabilidade do evento por conta própria ou todos os eventos de probabilidade idêntica de uma vez.
- Entre com o intervalo do cenário de entrada

Funções de Utilidade

Funções de utilidade (**Figura 5.67**), ou U(x), são por vezes usadas no lugar de valores esperados dos Payoffs dos Terminais em uma árvore de decisão. A U(x) pode ser desenvolvida de duas maneiras: usando a experimentação tediosa e detalhanda de cada resultado possível, ou utilizando um método de extrapolação exponencial (usado aqui). Elas podem ser modeladas por um tomador de decisão que é avesso ao risco (as desvantagens são mais desastrosas ou dolorosas do que um potencial crescimento de igual valor), ou do tipo neutro-ao-risco (as vantagens e desvantagens têm atratividades iguais), ou amante do risco (o potencial de upside é mais atraente). Digite os valores mínimo e máximo dos retornos esperados do seu terminal e o número de pontos de dados para calcular a curva de utilidade e a tabela.

Se você tivesse a alternativa entre realizar uma aposta com a chance de 50:50, onde você tanto pode ganhar $X ou perde - $X/2 versus não jogar, recebendo um retorno igual a $0, qual seria o valor de X? Por exemplo, se você está indiferente entre uma aposta onde você pode ganhar $100 ou perder $50, com igual probabilidade, em comparação com não jogar, então o valor de X é de $100. Digite o X na caixa de resultados positivos abaixo. Note-se que quanto maior é X, menos avesso ao risco você é, enquanto um X menor indica que você está mais avesso ao risco.

Entrar com os dados solicitados, selecione o tipo de U(x), e clique em Calcule Utilidade para obter os resultados. Você também pode aplicar os valores de U(x) à árvore de decisão e rexecuta-la, ou reverter à árvore anterior utilizando os valores de retorno (Payoffs) esperados.

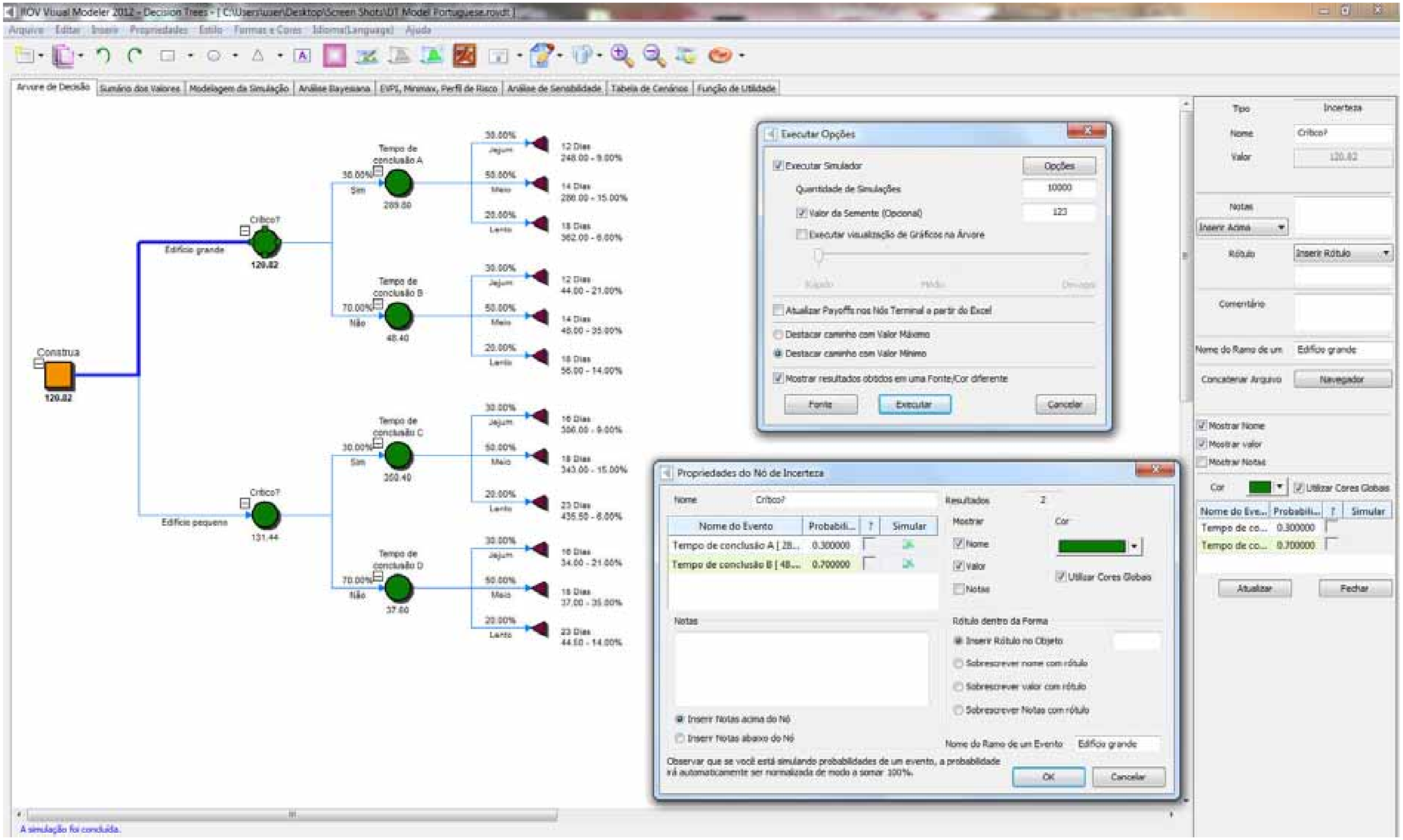


Figura 5.61 – ROV Decision Tree (Arvore de Decisão)

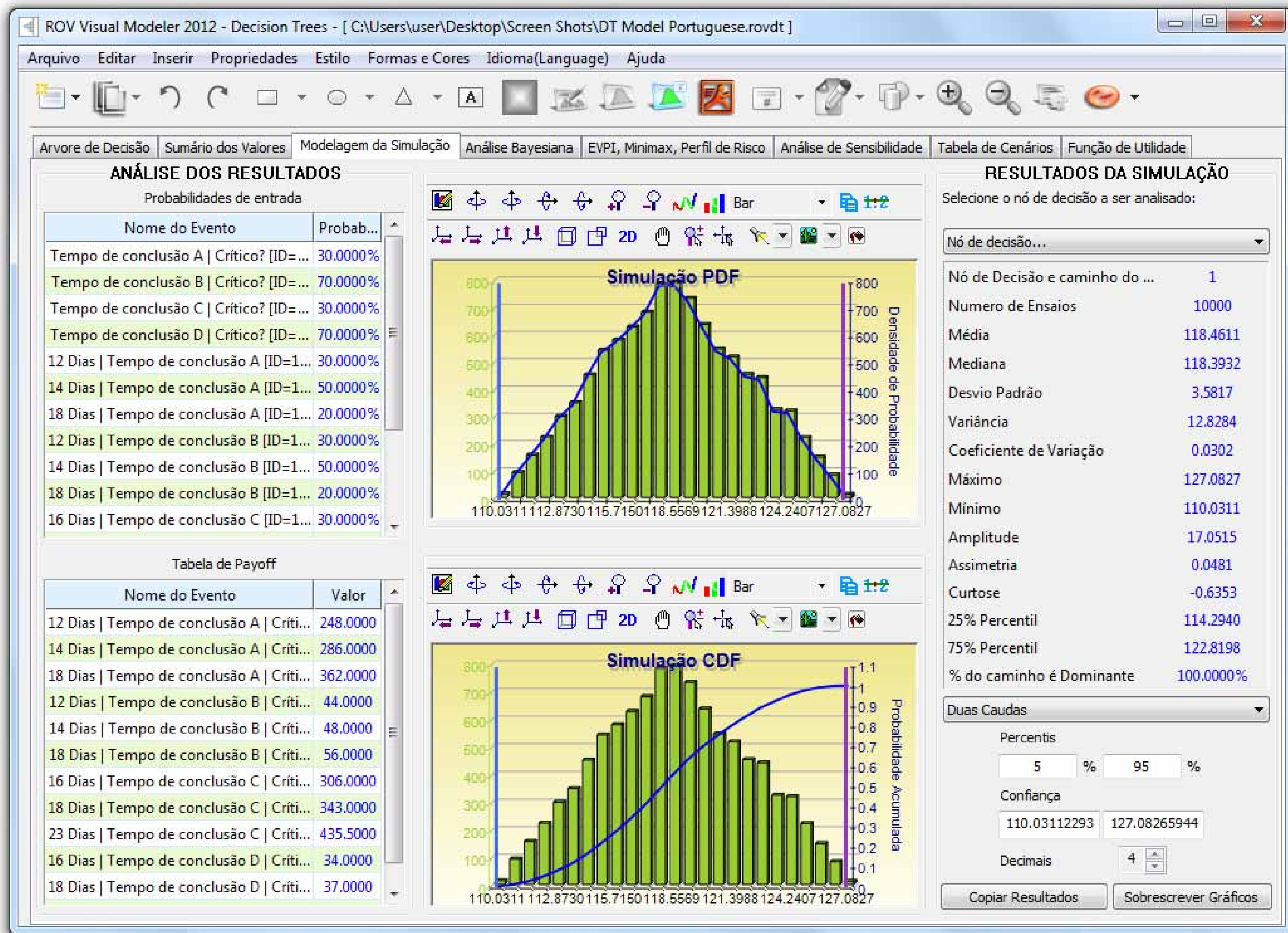


Figura 5.62 – ROV Decision Tree (Resultados da Simulação)

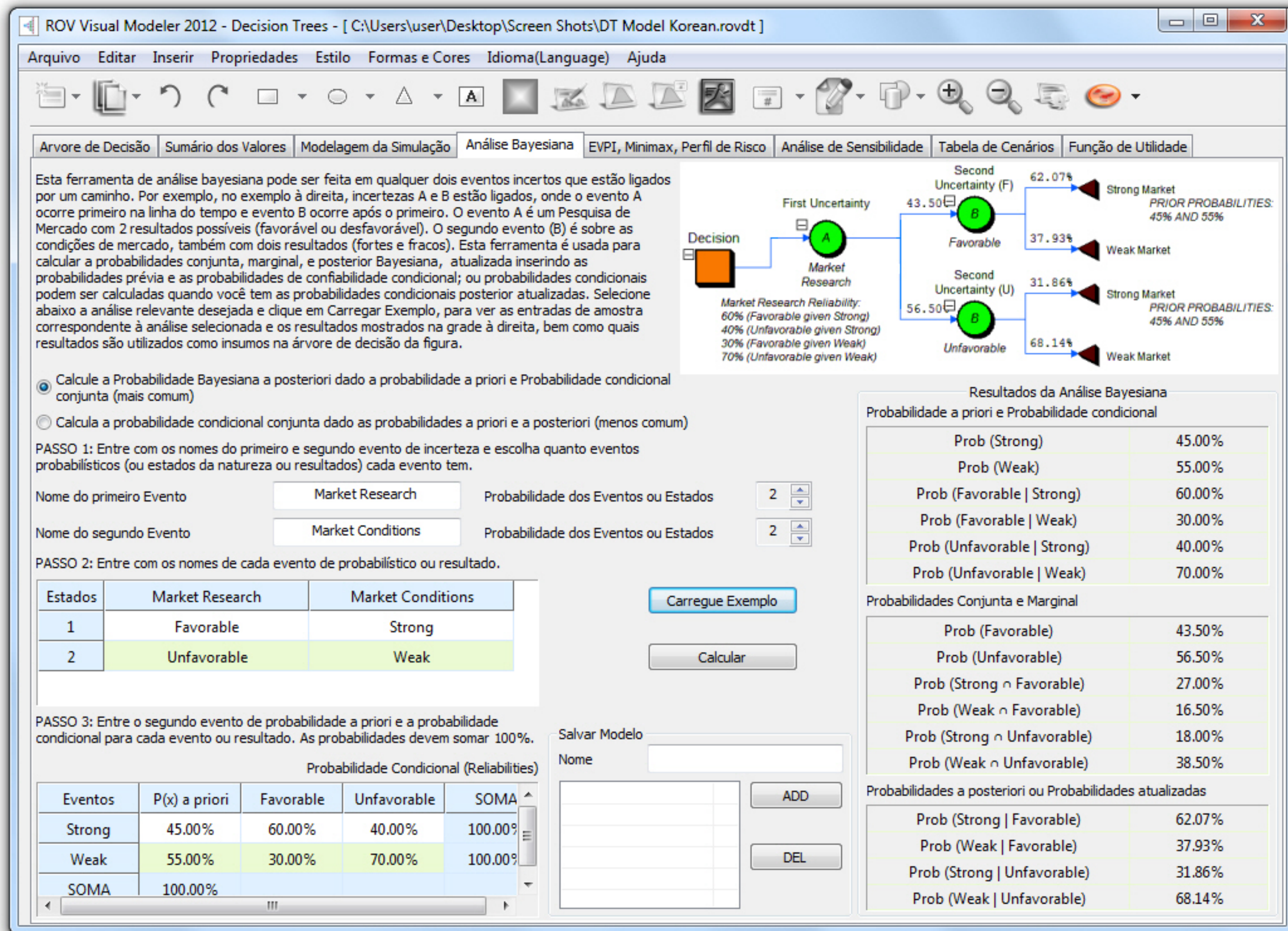


Figura 5.63 – ROV Decision Tree (Análise Bayesiana)

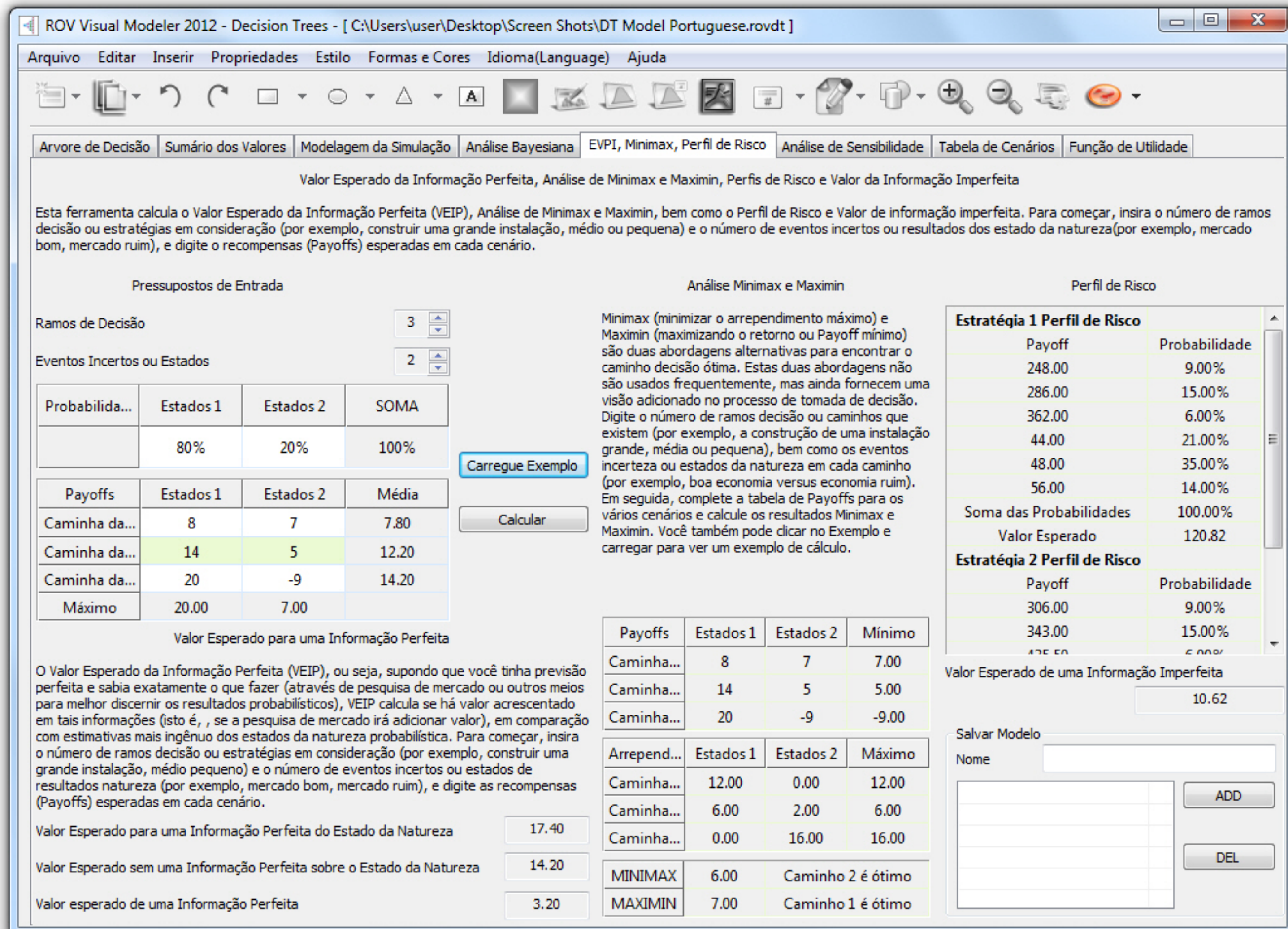


Figura 5.64 – ROV Decision Tree (VEIP, MINIMAX, Perfil de Risco)

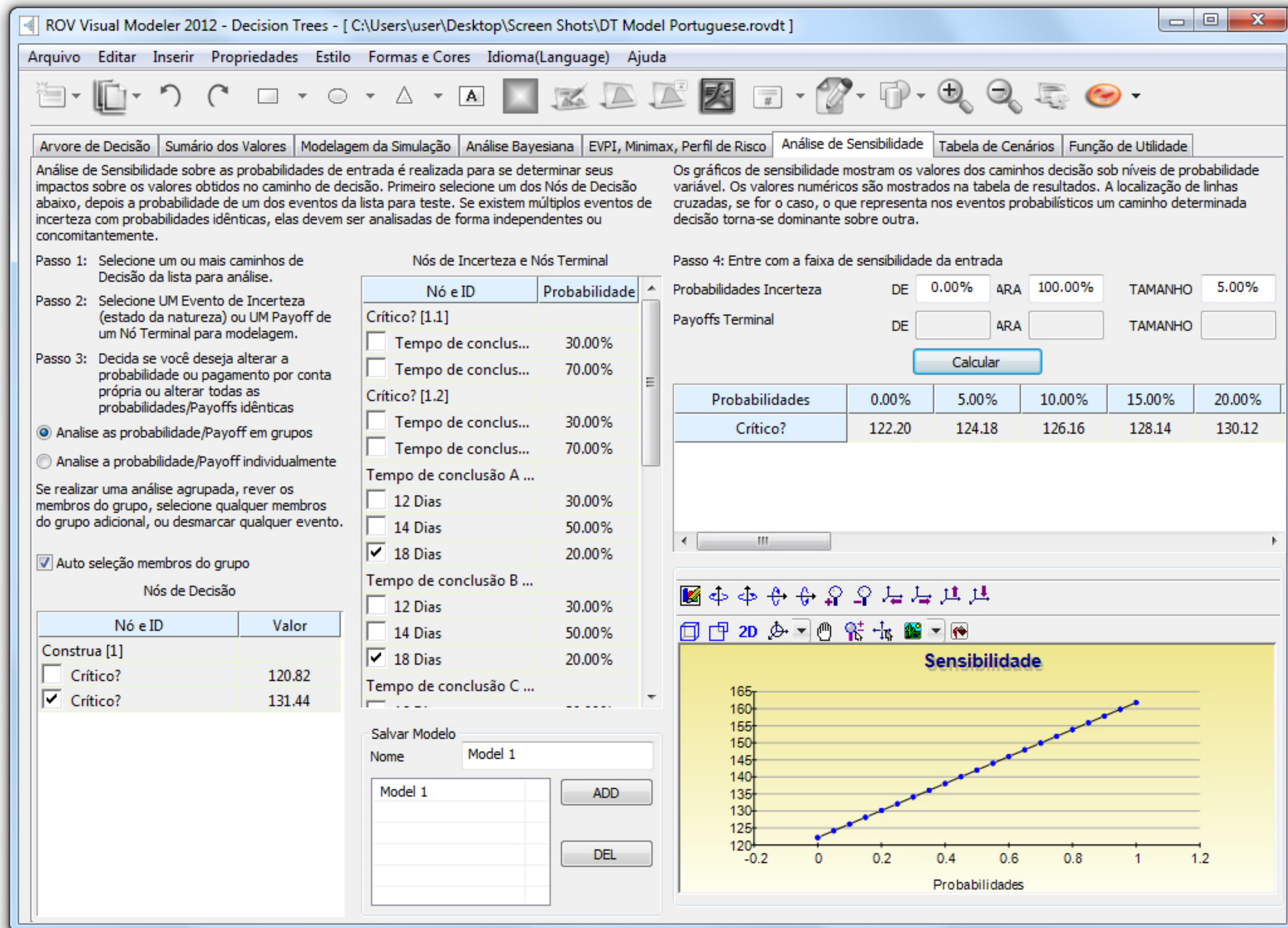


Figura 5.65 – ROV Decision Tree (Análise de Sensibilidade)

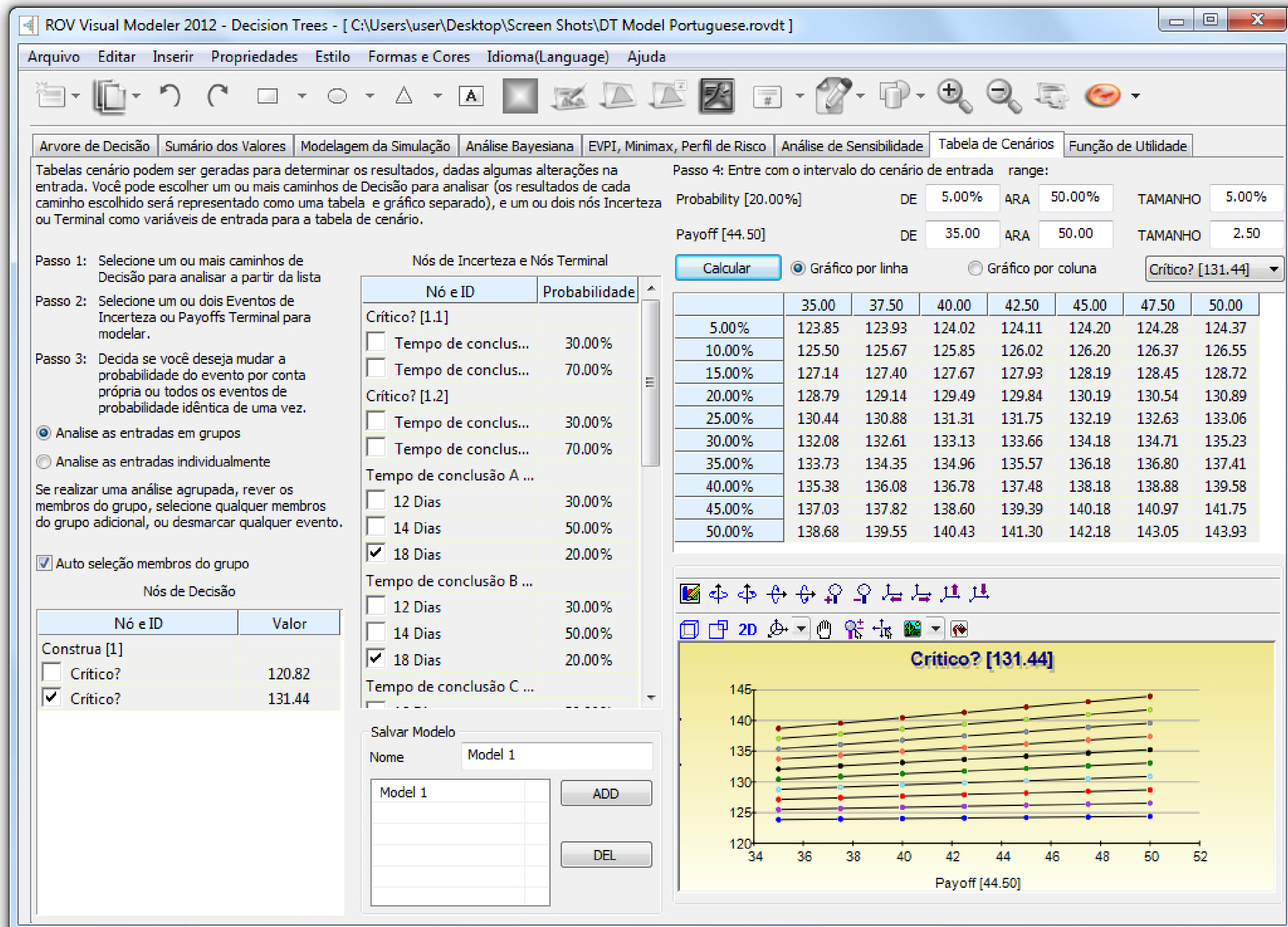


Figura 5.66 – ROV Decision Tree (Tabelas de Cenários)

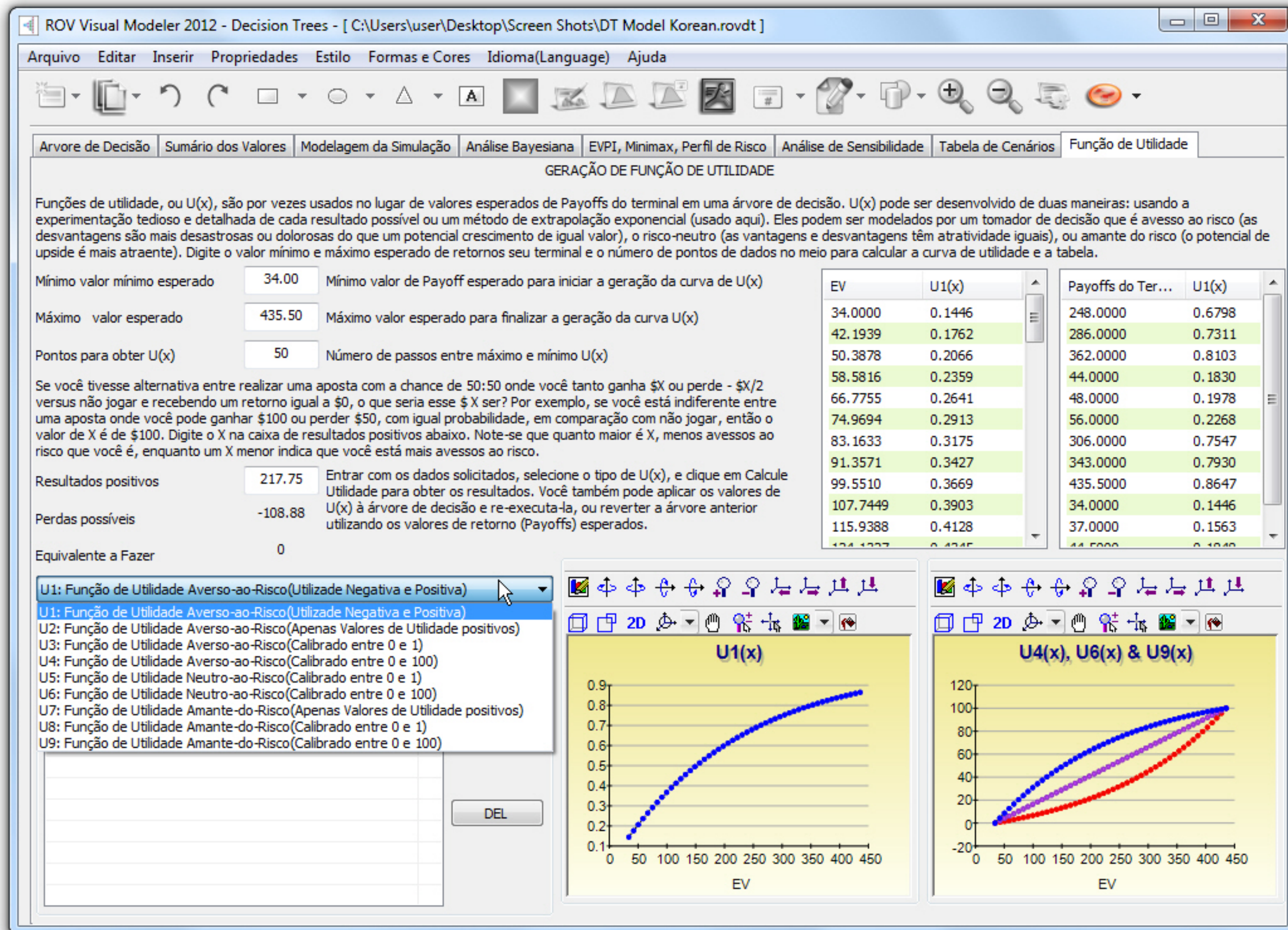


Figura 5.67 – ROV Decision Tree (Funções de Utilidade)

Dicas e técnicas úteis

Veja a seguir algumas dicas e técnicas úteis para usuários avançados do Risk Simulator. Para obter detalhes sobre o uso de ferramentas específicas, consulte as seções relevantes no manual do usuário.

DICAS: suposições (interface Definir valores de entrada)

- Salto rápido: selecione uma distribuição e digite qualquer letra para ir para a primeira distribuição que comece com essa letra (por exemplo, clique em Normal e insira W para ir para a distribuição Weibull).
- Exibições abertas com o botão direito: selecione uma distribuição, clique com o botão direito do mouse e selecione exibições diferentes das distribuições (ícones grandes, ícones pequenos, lista).
- Pressionar Tab para atualizar os gráficos: depois de inserir novos parâmetros de entrada (por exemplo, inserção de uma nova média ou um novo valor de desvio padrão), pressione Tab no teclado ou clique em qualquer outra parte da interface do usuário fora da caixa de entrada para que o gráfico de distribuição seja atualizado automaticamente.
- Inserir correlações: você pode inserir correlações de paridade diretamente aqui (as colunas podem ser redimensionadas se necessário), usar a ferramenta de ajuste da distribuição para calcular automaticamente e inserir todas as correlações de paridade ou, após a configuração de algumas premissas, usar a ferramenta de edição de correlação para inserir sua matriz de correlações.
- Equações em uma célula de suposição: apenas células vazias ou com valores estáticos podem ser definidas como suposições. Contudo, pode haver situações em que uma função ou uma equação seja necessária em uma célula de suposição. Para fazer isso, insira primeiro o valor de entrada na célula e digite a equação ou a função. Quando a simulação está sendo executada, os valores simulados substituirão a função e depois que a simulação é concluída, a função ou a equação é mostrada de novo.

DICAS: copiar e colar

- Copiar e colar usando a tecla *Escape*: quando você seleciona uma célula e usa a função Copiar do Risk Simulator, tudo é copiado para a área de transferência do Windows, incluindo o valor, a equação, a função, a cor, a fonte e o tamanho da célula, bem como as suposições, as previsões e as variáveis de decisão do Risk Simulator. Então, à medida que você aplica a função Colar do Risk Simulator, tem duas opções. A primeira é aplicar diretamente o comando Colar do Risk

Simulator para que todos os valores, cores, fontes, equações, funções e parâmetros da célula sejam colados na nova célula. A segunda opção é clicar primeiro em Escape no teclado e, então, aplicar o comando Colar do Risk Simulator. A tecla Escape informa ao Risk Simulator que você deseja colar apenas a suposição, a previsão ou a variável de decisão e não os valores, as cores, as equações, as funções e as fontes da célula. Pressionar Escape antes de colar permite que você mantenha os valores e os cálculos da célula de destino e cole apenas os parâmetros do Risk Simulator.

- Copiar e colar em várias células: você pode selecionar várias células para copiar e colar (com suposições contíguas e não contíguas).

DICAS: correlações

- Definir suposição: define correlações de paridade usando a caixa de diálogo Definir valores de entrada (ideal para inserção de várias correlações).
- Editar correlações: configura uma matriz de correlações inserindo manualmente ou colando da área de transferência do Windows (ideal para matrizes de correlações grandes e diversas correlações).
- Ajuste da distribuição de múltiplas variáveis: calcula e insere automaticamente correlações de paridade (ideal para a execução de ajuste de múltiplas variáveis, calcular automaticamente correlações e decidir o que constitui uma correlação estatisticamente significante).

DICAS: diagnóstico de dados e análise estatística

- Estimativa do parâmetro estocástico: nos relatórios de diagnóstico de dados e análise estatística, há uma guia sobre estimativas de parâmetro estocástico que estima a volatilidade, o crescimento, a taxa de reversão à média e as taxas de difusão com salto de acordo com dados históricos. Observe que os resultados desse parâmetro são baseados apenas em dados históricos utilizados e que os parâmetros podem mudar ao longo do tempo e dependendo da quantidade de dados históricos ajustados. Além disso, os resultados da análise mostram todos os parâmetros e não determina qual modelo de processo estocástico (por exemplo, movimento browniano, reversão à média, difusão com salto ou processo misto) é o melhor ajuste. Cabe ao usuário fazer essa escolha dependendo da variável da série temporal a ser prevista. A análise não pode determinar qual é o melhor processo, apenas o usuário pode fazer isso (por exemplo, o processo movimento browniano é melhor para a modelagem de preços de ações, mas a análise não pode determinar se os dados históricos analisados são de uma ação ou de outra variável, somente o usuário saberá isso). Por último, uma boa dica é que se determinado parâmetro não está na faixa normal, o processo que precisa desse parâmetro de entrada tem mais probabilidade de não ser o correto (por exemplo, se a taxa de reversão à média é de 110%, provavelmente a reversão à média não é o processo correto e assim por diante).

DICAS: análise da distribuição, gráficos e tabelas de probabilidade

- Análise da distribuição: usada para calcular rapidamente o FDP, a FDA e a IFDA das 42 distribuições de probabilidade disponíveis no Risk Simulator, e para retornar uma *tabela* desses valores.
- Tabelas e gráficos de distribuição: usada para comparar *parâmetros diferentes da mesma distribuição* (por exemplo, as formas e os valores de FDP, FDA e IFDA de uma distribuição Weibull com alfa e beta iguais a [2, 2], [3, 5] e [3,5, 8] e sobrepô-los um ao outro).
- Gráficos sobrepostos: usada para comparar *distribuições diferentes* (valores de entrada teóricos e previsões de saída empíricas simuladas) e sobrepô-las uma à outra para uma comparação visual.

DICAS: fronteira eficiente

- Variáveis de fronteira eficiente: para acessar as variáveis de fronteira, primeiro defina as restrições do modelo antes de configurar as variáveis de fronteira.

DICAS: células de previsão

- Células de previsão sem equação: você pode definir previsões de saída em células sem equações ou valores (basta ignorar a mensagem de aviso), mas lembre-se de que o gráfico de previsão resultante estará vazio. As previsões de saída são tipicamente definidas em células vazias quando há macros para serem calculadas e a célula será continuamente atualizada.

DICAS: gráficos de previsão

- *Tab* versus *barra de espaço*: pressione Tab no teclado para atualizar o gráfico de previsão e obter os percentis e os valores de confiança depois de inserir algumas entradas. Pressione a barra de espaço para alternar entre as várias guias no gráfico de previsão.
- Exibição normal versus exibição global: clique nessas exibições para alternar entre uma interface com guias e uma interface global na qual todos os elementos dos gráficos de previsão estão visíveis ao mesmo tempo.
- Copiar: esta ação copia o gráfico de previsão ou a exibição global na íntegra, dependendo da exibição usada, a normal ou a global.

DICAS: previsão

- Endereço de vínculo da célula: se você selecionar primeiro os dados na planilha e depois executar uma ferramenta de previsão, o endereço da célula dos dados selecionados será inserido automaticamente na interface do usuário; caso contrário, será necessário inserir manualmente o endereço da célula ou usar o ícone de vínculo para vincular ao local de dados relevante.
- REQM da previsão: use como a medida de erro universal em vários modelos de previsão para obter comparações diretas da precisão de cada modelo.

DICAS: previsão: ARIMA

- Períodos de previsão: o número de linhas de dados exógenos deve exceder as linhas de dados da série temporal em pelo menos os períodos de previsão desejados (por exemplo, se você desejar prever cinco períodos no futuro e tiver 100 pontos de dados da série temporal, necessitará ter pelo menos 105 ou mais pontos de dados na variável exógena); caso contrário, execute ARIMA sem a variável exógena para prever quantos períodos desejar sem limitações.

DICAS: previsão: econometria básica

- Separação de variável com ponto-e-vírgula: separe as variáveis independentes usando um ponto-e-vírgula.

DICAS: previsão: logit, probit e tobit

- Requisitos de dados: as variáveis dependentes para executar os modelos logit e probit devem ser somente binárias (0 e 1), enquanto o modelo tobit aceita binários e outros valores decimais numéricos. As variáveis independentes para todos os três modelos podem aceitar qualquer valor numérico.

DICAS: previsão: processos estocásticos

- Entradas de amostra padrão: quando estiver em dúvida, use as entradas padrão como um ponto de partida para desenvolver seu próprio modelo.
- Ferramenta de análise estatística para estimativa de parâmetro: use esta ferramenta para calibrar os parâmetros de entrada nos modelos de processo estocástico estimando-os a partir de seus dados brutos.
- Modelo de processo estocástico: quando a interface do usuário do processo estocástico fica congelada por muito tempo, é provável que as suas entradas estejam incorretas e que o modelo não esteja corretamente especificado (por

exemplo, se a taxa de reversão à média é de 110%, é provável que a reversão à média não seja o processo correto etc.). Tente entradas diferentes ou use um modelo diferente.

DICAS: previsão: linhas de tendência

- Resultados da previsão: role para o fim do relatório para ver os valores previstos.

DICAS: chamadas de função

- Funções RS: são funções para definir valores de entrada e obter funções estatísticas de previsão que você pode usar na planilha do Excel. Para usar estas funções, é necessário primeiro instalar as funções RS (Iniciar, Programas, Real Options Valuation, Risk Simulator, Ferramentas e Instalar funções) e depois executar uma simulação antes de configurar as funções RS no Excel. Consulte o modelo de exemplo 24 para saber como usar as funções.

DICAS: exercícios e vídeos para iniciantes

- Exercícios para iniciantes: há diversos exemplos práticos passo a passo e exercícios de interpretação de resultados disponíveis em Iniciar > Programas > Real Options Valuation > Risk Simulator. Esses exercícios foram desenvolvidos para ajudá-lo a aprender rapidamente a usar o software.
    - Vídeos para iniciantes: estão disponíveis gratuitamente no nosso site em ***www.realoptionsvaluation.com/download.html*** ou ***www.rovdownloads.com/download.html***

DICAS: ID do hardware

- Clique com o botão direito do mouse em Cópia do ID do hardware: na interface do usuário Instalar licença, selecione ou clique duas vezes no HWID para selecionar seu valor, clique com o botão direito do mouse para copiar ou clique no link Enviar ID do hardware por email para gerar um email com o ID do hardware.
- Solução de problemas: execute a Solução de problemas na pasta Iniciar > Programas > Real Options Valuation > Risk Simulator e execute a ferramenta Obter ID do hardware para obter o ID do hardware do seu computador.

DICAS: amostragem por hipercubo latino (LHS) versus simulação Monte Carlo (MCS)

- Correlações: na configuração de correlações de paridade em valores de entrada, é recomendável usar a configuração Monte Carlo no menu Opções do Risk Simulator. A amostragem por hipercubo latino não é compatível com o método de cópula correlacionada para simulação.
- Compartimentos de LHS: um número elevado de compartimentos retarda a simulação, embora forneça um conjunto de resultados de simulação mais uniforme.
- Aleatoriedade: todas as técnicas de simulação aleatórias no menu Opções foram testadas e todas são boas simulações e alcançam os mesmos níveis de aleatoriedade quando um grande número de tentativas é executado.

DICAS: recursos online

- Livros, vídeos para iniciantes, modelos, whitepapers: disponíveis gratuitamente no nosso site em www.realoptionsvaluation.com/download.html ou www.rovdownloads.com/download.html.

DICAS: otimização

- Resultados impossíveis: se a execução da otimização retornar resultados impossíveis, altere as restrições de uma igualdade (=) para uma desigualdade (>= ou <=) e tente novamente. Isso também se aplica quando você está executando uma análise de fronteira eficiente.

DICAS: perfis

- Vários perfis: crie vários perfis e alterne entre eles em um único modelo. Isso permite que você execute cenários em simulações e seja capaz de alterar os parâmetros de entrada ou os tipos de distribuição no modelo para ver os efeitos nos resultados.
- Perfil necessário: suposições, previsões ou variáveis de decisão não poderão ser criadas se não houver um perfil ativo. No entanto, uma vez que você possua um perfil, não precisará criar novos perfis. Se desejar executar um modelo de simulação incluindo suposições ou previsões adicionais, mantenha o mesmo perfil.
- Perfil ativo: o último perfil usado quando você salvar a planilha no Excel será aberto automaticamente na próxima vez que o arquivo do Excel for aberto.
- Vários arquivos do Excel: quando estiver alternando entre vários modelos do

Excel abertos, o perfil ativo será do modelo atual e ativo do Excel.

- Perfis em várias pastas de trabalho: tenha cuidado quando tiver vários arquivos do Excel abertos, pois se apenas um desses arquivos tiver um perfil ativo e você, acidentalmente, alternar para outra planilha e definir suposições e previsões nela, as suposições e as previsões não serão executadas e serão inválidas.
- Exclusão de perfis: você pode clonar e excluir perfis existentes, mas observe que pelo menos um perfil deve existir no arquivo do Excel, caso exclua perfis.
- Local do perfil: os perfis criados por você (contendo as suposições, as previsões, as variáveis de decisão, os objetivos, as restrições etc.) são salvos como uma planilha oculta criptografada. É por isso que quando você salva o arquivo da pasta de trabalho do Excel, o perfil também é automaticamente salvo.

DICAS: menu de atalho e outras teclas de atalho

- Clicar com o botão direito do mouse: para abrir o menu de atalho do Risk Simulator, clique com o botão direito do mouse em uma célula em qualquer lugar no Excel.

DICAS: salvar

- Salvar o arquivo do Excel: salva as configurações do perfil, as suposições, as previsões, as variáveis de decisão e o seu modelo do Excel (incluindo relatórios, gráficos e extrações de dados do Risk Simulator).
- Salvar configurações do gráfico: salva as configurações de gráfico de previsão de maneira que possam ser recuperadas e aplicadas a gráficos de previsões futuros (use os ícones de salvar e abrir nos gráficos de previsão).
- Salvar e extrair dados simulados no Excel: extrai as suposições e as previsões de uma execução simulada, mas o arquivo do Excel ainda precisará ser salvo para preservar os dados para recuperação posterior.
- Salvar gráficos e dados simulados no Risk Simulator: use o Risk Simulator, a extração de dados e salve em um arquivo **.RiskSim* para permitir reabrir o gráfico de previsão ativo e dinâmico com os mesmos dados posteriormente sem ter que executar a simulação de novo.
- Salvar e gerar relatórios: os relatórios de simulação e outros relatórios analíticos são extraídos como planilhas separadas em sua pasta de trabalho e o arquivo inteiro do Excel deverá ser salvo para preservar os dados para recuperação posterior.

DICAS: técnicas de amostragem e simulação

- Gerador de número aleatório: há seis geradores de número aleatório para os quais há suporte (consulte o manual do usuário para obter mais detalhes). Em geral, o método padrão do *ROV Risk Simulator* e o método *Embaralhamento aleatório subtrativo avançado* são as duas abordagens recomendadas. Não aplique os outros métodos a menos que o seu modelo ou sua análise necessite especificamente deles e, mesmo nesse caso, recomendamos testar os resultados com as duas abordagens indicadas.

DICAS: SDK e bibliotecas de DLL

- SDK, DLL e OEM: todas as análises no Risk Simulator podem ser chamadas fora deste software e integradas a qualquer software de propriedade do usuário. Entre em contato com admin@realoptionsvaluation.com para obter detalhes sobre como usar o nosso SDK para acessar os arquivos de análises da biblioteca de vínculos dinâmicos (DLL).

DICAS: iniciar o Risk Simulator com o Excel

- Solução de problemas ROV: execute esta solução de problemas para obter o ID do hardware do seu computador para fins de licenciamento, para verificar as configurações e os pré-requisitos do computador e para reabilitar o Risk Simulator caso ele tenha sido acidentalmente desabilitado.
- Iniciar o Simulator quando o Excel for iniciado: você pode permitir que o Risk Simulator inicie automaticamente quando o Excel for iniciado ou pode iniciá-lo manualmente em Iniciar > Programas > Real Options Valuation > Risk Simulator. Essa preferência pode ser definida no menu Opções do Risk Simulator.

DICAS: simulação super-rápida

- Desenvolvimento de modelo: se você desejar executar seu modelo em uma velocidade super-rápida, recomenda-se que, durante a construção do modelo, sejam executadas algumas simulações super-rápidas para testá-lo e certificar-se de que o produto final executará a simulação super-rápida. Não teste a velocidade super-rápida somente depois que o modelo final estiver concluído para não ter que identificar se há vínculos corrompidos ou funções incompatíveis.
- Velocidade normal: em caso de dúvida, a simulação na velocidade normal sempre funciona.

DICAS: análise tornado

- Análise tornado: esta análise jamais deve ser executada apenas uma vez. Ela é considerada uma ferramenta de diagnóstico de modelo, o que significa que deve ser executada várias vezes no mesmo modelo. Por exemplo, em um modelo grande, o tornado pode ser executado na primeira vez com todas as configurações padrão e mostrando todos os precedentes (selecione Mostrar todas as variáveis). Esta análise única pode resultar em um relatório grande e gráficos tornado muito grandes (e potencialmente confusos). Porém, ela é um ótimo ponto de partida para determinar a quantidade de precedentes considerados fatores críticos de sucesso (por exemplo, o gráfico tornado pode mostrar que as primeiras cinco variáveis têm alto impacto sobre o resultado, ao passo que as 200 variáveis restantes exercem pouco ou nenhum impacto). Nesse caso, uma segunda análise tornado é executada com menos variáveis (selecione Mostrar as 10 variáveis principais se as primeiras cinco forem críticas, o que cria um bom relatório e um gráfico tornado que mostra um contraste entre os fatores-chave e os fatores menos críticos, ou seja, você nunca deve mostrar um gráfico tornado com apenas as variáveis-chave sem mostrar algumas variáveis menos críticas, para que haja assim um contraste sobre os efeitos no resultado).
- Valores padrão: os pontos de teste padrão podem ser aumentados de ±10% para um valor maior a fim de testar não linearidades (o gráfico aranha mostra linhas não lineares e os gráficos tornado são oblíquos para um dos lados se os efeitos dos precedentes forem não lineares).
- Valores zero e inteiros: entradas com valores inteiros ou iguais a zero só devem ser desmarcadas na análise tornado antes da execução. Caso contrário, o percentual de perturbação pode invalidar o modelo (por exemplo, se o modelo usa uma tabela de pesquisa na qual Jan = 1, Fev = 2, Mar = 3 e assim sucessivamente, a perturbação do valor 1 em +/- 10% resulta em 0,9 e 1,1, o que não faz sentido para o modelo).
- Opções de gráfico: tente várias opções de gráfico para encontrar as opções que deve ser ativadas ou desativadas no seu modelo.

DICAS: solução de problemas

- Solução de problemas ROV—execute esta solução de problemas para obter o HWID de seu computador para fins de licenciamento, para ver as configurações e pré-requisitos de seu computador e para reabilitar o Risk Simulator caso ele tenha sido acidentalmente desabilitado.

ÍNDICE REMISSIVO