Revisão de Estatística

Este manual destina-se a servir de uma revisão dos elementos básicos da estatística. A cobertura e apresentação do material introduz assuntos que serão importantes nos entendimentos e nas implementações de conceitos de finanças em que o curso está focado. É uma boa idéia espanar o pó dos seus velhos livros de estatística ou fazer uma viagem até a biblioteca para ler com atenção os livros textos introdutórios da estatística. Os conceitos destacados nesta revisão são importantes ferramentas para este curso.

A estatística é o estudo das variáveis aleatórias. Os retornos dos investimentos são variáveis aleatórias. Portanto, a estatística é um componente crucial do campo de investimentos

À medida que você progredir pelo manual, você encontrará alguns exercícios práticos que você esperava fazer. As soluções são fornecidas no final do manual.

A. Variáveis aleatórias

Em geral, uma variável para a qual os valores não são conhecidos até que um experimento seja realizado é chamada a variável aleatória. Também, o retorno sobre uma ação particular stock é uma variável aleatória desde que o retorno sobre a ação durante qualquer período particular não seja conhecido com certeza.

Exemplo:

Considere o resultado de um lançamento de dados. Existem seis resultados possíveis (1, 2, 3, 4, 5, 6). O resultado é uma variável aleatória, pois o valor real não é conhecido antes de se lançar o dado. Se ele for um dado leal, cada resultado tem uma probabilidade de acontecer igual a 1/6.

Seja a variável aleatória *X*, o resultado do lançamento de um dado. *X* é então um exemplo de uma *variável aleatória discreta,* pois ela pode assumir somente 6 valores.

Exercício 1: Escreva os resultados possíveis de lançamento de uma moeda e a probabilidade associada a cada lançamento. (As soluções dos exercícios práticos estão no final do manual).

Exercício 2*:* Suponha que você lance dois dados. Você está interessado na soma dos dados. Escreva os resultados possíveis (onde cada resultado é a soma dos valores do lançamento de dois dados) e suas probabilidades associadas. (Sugestão: o resultado de lançar um dado é independente do resultado de lançar o outro dado. Daí, a probabilidade de obter um par particular de valores é $\frac{1}{6} x \frac{1}{6} = \frac{1}{36}$).

No caso de uma variável aleatória discreta, a *função de densidade de probabilidade* (*fdp*) é definida como um gráfico (ou tabela) que mostra todos os valores que a variável aleatória pode assumir com suas probabilidades associadas.

Para a variável aleatória *X* que é o resultado do lançamento de um dado, a função de densidade de probabilidade está mostrada no gráfico abaixo.

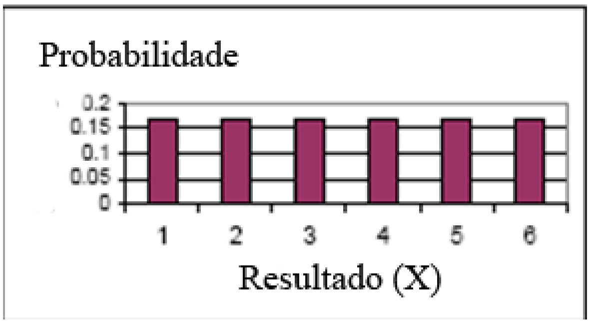


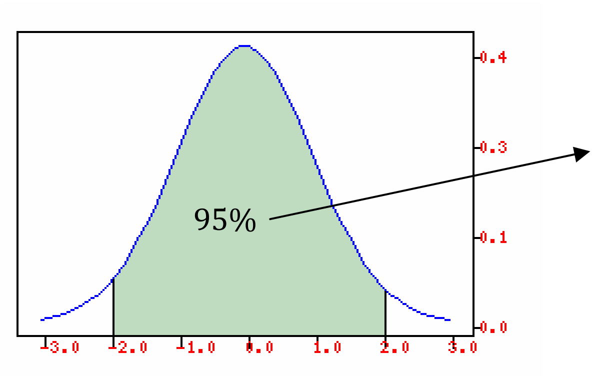
Figura 1. Função densidade de probabilidade de *X*.

A probabilidade de cada resultado é 0,16, neste caso.

Bertolo, L.A.
COC - Catanduva

Exercício 3: Calcular a probabilidade de se obter uma soma menor do que 4 quando se lançarem dois dados.

Uma *variável aleatória contínua* é aquela que pode assumir qualquer valor real (não apenas números inteiros) num intervalou da linha de números reais. Para exemplo, seja *Y* o retorno de um lote de ações da IBM. *Y* pode teoricamente assumir valores de –100% (devido a responsabilidade limitada, você não pode perder mais do que seu investimento original) até valores positivos muito grandes, e é claro qualquer valor entre esses. A probabilidade que *Y* assume qualquer valor único é zero. Isto é porque uma variável aleatória contínua pode assumir um número infinito de valores e a chance de qualquer valor ocorrer é zero. Variáveis aleatórias contínuas podem seguir várias fdp. Algumas fdp mais comuns são a distribuição normal, distribuição exponencial, e a distribuição uniforme.



Existe 95% de estar dentro de 2σ da média (centro)

Figura 2. Função densidade de probabilidade da distribuição normal padrão (média 0 e variância 1)

A distribuição normal é uma das mais importantes distribuições. Se *X* é uma variável aleatória normal com média μ e variância σ^2 (simbolizada como $X \sim N(\mu, \sigma^2)$), então sua função densidade de probabilidade é dada pela seguinte expressão:

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

A média μ e a variância σ^2 são parâmetros desta distribuição e determinam sua localização e dispersão respectivamente. O intervalo de valores que uma variável aleatória normal pode assumir é $-\infty$ a $+\infty$. A função de densidade de probabilidade normal é uma curva simétrica na forma de sino centrada no valor médio *a*. Quanto maior para a variância b^2, mais disperso são os valores possíveis. Duas distribuições tendo a mesma variância com diferentes médias, elas terão formas idênticas, mas serão localizadas em pontos diferentes sobre o eixo x.

Devido à simetria da distribuição normal (Figura 2), metade da probabilidade (50%) está associada com valores à esquerda da média, e metade com valores à direita da média. Isto é, a probabilidade de observar um resultado que seja menor do que *a* é a soma das probabilidades de observar todos os valores menores que *a* – que é o mesmo que a área sob o *fdp* à esquerda de *a*. Para distribuições normais, 68% dos resultados possíveis cairão entre +/-1 desvio padrão da média. Isto é, 68% da área sob a *fdp*, ficam entre a+b e a-b. Também, 95% dos resultados possíveis cairão entre +/-2 desvios padrões da média (pintada na Figura 2), e 99% de todos os resultados ficarão entre +/-3 desvios padrões da média.

Uma variável aleatória *normal padrão* tem uma função de densidade de probabilidade normal com média 0 e variância 1.

Uma das razões para que a distribuição normal seja uma distribuição importante é devido ao *Teorema do Limite Central*. O teorema do limite central diz que o valor médio de N variáveis aleatórias independentes de qualquer função de densidade de probabilidade (enquanto ela tiver uma média e uma variância e N sendo suficientemente grande) terá aproximadamente uma distribuição normal padrão após subtrair sua média e dividir pelo seu desvio padrão. Isto é, todas as distribuições convergem assintóticamente para a distribuição normal.

B. Momentos de Variáveis aleatórias

Os primeiros quatro *momentos* de uma variável aleatória são, respectivamente, a média, variância, simetria e curtose.

Média: A *média* de uma variável aleatória *X* é o valor médio da variável aleatória em um número infinito de repetições do experimento. A média dá uma medida do centro ou localização dos dados de uma variável aleatória. A média é também referida como o valor esperado da variável aleatória, a qual é denotada por E(*X*). Note também que a média é também denotada por μ.

Para uma variável aleatória discreta, o valor esperado é a média ponderada dos valores da variável aleatória com os pesos sendo a probabilidade anexada a cada valor.

$$E(X) = \sum_{i=1}^{N} P_i x_i$$

onde, *X* é uma variável aleatória discreta
N é o número de resultados possíveis
P_i é a probabilidade do resultado *i*
x_i é o valor de *X* quando o resultado *i* ocorre

Exemplo:

Qual é o valor esperado do resultado de lançamento de um dado?

$$E(X) = \frac{1}{6}x1 + \frac{1}{6}x2 + \frac{1}{6}x3 + \frac{1}{6}x4 + \frac{1}{6}x5 + \frac{1}{6}x6 = 3{,}5$$

Quão certo você está de obter este valor esperado?

De forma interessante, como você já deve ter notado o valor esperado nunca será observado neste caso.

Exemplo:

Vamos dar a seguinte informação sobre duas ações e as várias condições do tempo que podem ocorrer.

Estado do tempo	Probabilidade	Retorno sobre a Ação do Amusement Park (*A*)	Retorno sobre a Ação Ski Resort (*S*)
Extremamente Frio	10%	-15%	35%
Frio	30%	-5%	15%
Médio	40%	10%	5%
Quente	20%	30%	-5%

Quais são os retornos esperados de cada ação?

$E(A) = (-15\% \times 0{,}1) + (-5\% \times 0{,}3) + (10\% \times 0{,}4) + (30\% \times 0{,}2) = 7\%$

$E(S) = (35\% \times 0{,}1) + (15\% \times 0{,}3) + (5\% \times 0{,}4) + (-5\% \times 0{,}2) = 9\%$

Variância: A *variância* de uma variável aleatória dá uma idéia da dispersão dos valores possíveis da variável aleatória.

A variância de uma variável aleatória discreta *X* é definida como:

$$Var(X) = \sigma^2 = \sum_{i=1}^{N} P_i [x_i - E(X)]^2$$

Bertolo, L.A.
COC - Catanduva

Isto pode ser visualizado como o valor esperado de $(x_i - E(X))^2$ onde :

X é uma variável aleatória discreta
N é o número de resultados possíveis
P_i é a probabilidade de resultado i
X_i é o valor de X quando o resultado i ocorrer
$E(X)$ é a média de X

O *desvio padrão* de uma variável aleatória, denotado por σ é a raiz quadrada da variância. Precisa-se calcular E(X) primeiro, o qual já foi encontrado.

Exemplo:
Seja X o valor observado no lançamento de um dado.

Desvio Padrão (X) = $\sigma_x = \sqrt{Variância(X)}$ = 1,71

Exemplo Amusement Park / Ski Resort:

$$Var(A) = 0,1\,x(-15-7)^2 + 0,3\,x(-5-7)^2 + 0,4\,x(10-7)^2 + 0,2\,x(30-7)^2 = 201\%^2 = 0,0201$$

Desvio Padrão(A) = $\sigma_A = \sqrt{201\%^2}$ = 14,18%

$$Var(S) = 0,1\,x(35-9)^2 + 0,3\,x(15-9)^2 + 0,4\,x(5-9)^2 + 0,2\,x(-5-9)^2 = 124\%^2 = 0,0124$$

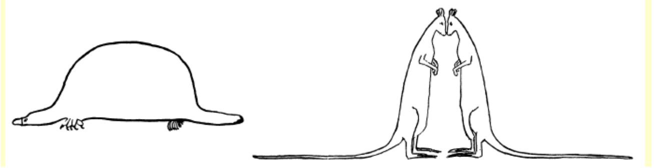
Desvio Padrão(S) = $\sigma_S = \sqrt{124\%^2}$ = 11,14%

Distorção: Freqüentemente você quer saber se os seus dados exibem uma forma simétrica. A *Distorção* mede o grau da assimetria da função de densidade de probabilidade. Quando os dados são simétricos, a média e a mediana são as mesmas. Para uma distribuição assimétrica que é chamada também de uma distribuição distorcida, a média e a mediana são diferentes. A distorção mede a tendência de uma distribuição alongar numa particular direção. Se a cauda alongada da função de densidade de probabilidade está do lado esquerdo da média, a densidade é dita ser distorcida para a esquerda (ou negativamente distorcida) e se a cauda alongada da função de densidade de probabilidade está do lado direito da média, a densidade é dita ser distorcida para a direita (ou positivamente distorcida).

Curtose: *Curtose* mede grau de saliência (achatamento) de uma densidade próxima ao seu centro. Uma distribuição é leptocúrtica (tem excesso de curtose) se a função de densidade de probabilidade é mais aguçada no seu centro e tem uma cauda mais longa que uma distribuição normal padrão. Uma distribuição platicúrtica (tem menos curtose) é mais achatada ao redor do centro e tem cauda mais curta do que uma distribuição normal padrão[1].

Covariância: *Covariância* é uma medida de como duas variáveis aleatórias movem uma com a outra. A medida para duas variáveis aleatórias discretas é dada por:

[1] Eu costumo pensar nos significados destas palavras assim, a primeira figura representa a platicúrtica e a segunda a leptocúrtica



$$Cov(X,Y) = \sum_{i=1}^{N} P_i\big(x_i - E(X)\big)\big(y_i - E(Y)\big)$$

Se duas variáveis aleatórias, X e Y, têm uma covariância positiva, significa que as duas variáveis estão geralmente movendo-se na mesma direção (relativa às suas médias respectivas). Isto é, na média, quando X estiver acima da sua média, Y está acima da sua média e quando X estiver abaixo de sua média, Y está abaixo de sua média.

Similarmente, se duas variáveis aleatórias, X e Y, têm uma covariância negativa, significa que, na média, as duas variáveis estão se movendo em direções opostas (relativas às suas médias respectivas). Isto é, na média, quando X está acima de sua média, Y está abaixo de sua média e quando X está abaixo de sua média, Y está acima de sua média.

Uma covariância zero implica que não existe uma associação nem positiva e nem negativa entre as variáveis aleatórias, i.e, não há associação entre X estar acima ou abaixo de sua média e Y estar acima ou abaixo de sua média.

Nota: Uma olhada na definição da covariância revela que Cov(X,X) = Var(X).

Exemplo Amusement Park / Ski Resort:

Cov(A,S) = 0,1x(-15 - 7)(35 - 9) + 0,3 (-5 - 7)(15 - 9) + 0,4 (10 - 7)(5 - 9) + 0,2 (30 - 7)(-5 - 9)
= $-148\%^2$ = -0,0148

<u>Correlação</u>: A magnitude da covariância é difícil de interpretar porque ela depende das unidades de medida das variáveis aleatórias e da dispersão (desvio padrão) das duas variáveis aleatórias. Para fornecer uma medida útil do grau para o qual duas variáveis movem-se juntas, a covariância medida é escalada como mostrado abaixo para calcular a *correlação* entre duas variáveis aleatórias.

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \rho_{X,Y}$$

<u>Exemplo</u>:

Assuma Cov(X,Y) = $200\%^2$ e Cov(X,Z) = $200\%^2$. Se σ_X = $20\%^2$ σ_Y = 20% e σ_Z = 25%, então a Corr(X,Y) e Corr(X,Z) serão diferentes, embora suas covariâncias sejam iguais. A covariância entre X e Z is ligeiramente menos aguçada que aquela entre X e Y, porque Z tem um intervalo mais largo de valores prováveis (um maior desvio padrão). Escalando as covariâncias pelo apropriado desvios padrões revela isto, $\rho_{X,Y} =$ 0,5 e $\rho_{X,Z} =$ 0,4.

A correlação mede o grau (ou intensidade) da covariância entre duas variáveis aleatórias e está sempre entre –1,0 e +1,0.

Corr(X,Y) = +1 implica que *X e Y* são perfeitamente *linearmente* correlacionados positivamente. Isto é, *X* e *Y* diferem somente por algum múltiplo e/ou constante. Especificamente, $Y = aX + b$ onde $a>0$ e b são constantes. Neste caso, conhecer o valor de *X* revelará *exatamente* o valor de *Y*.

Corr(X,Y) = -1 implica que X e Y estão perfeitamente *linearmente* correlacionados negativamente. Especificamente, $Y = aX + b$ onde $a<0$ e b são constantes. Novamente, conhecer o valor de *X* revelará *exatamente* o valor de *Y*.

<u>Exemplo</u> Amusement Park / Ski Resort:

$$Corr(A,S) = \frac{Cov(A,S)}{\sigma_A \sigma_S} = \frac{-148\%^2}{(14{,}18\%)(11{,}14\%)} = -0{,}937$$

Alguns resultados comumente usados:

Seja $Z = aX + bY$ onde X e Y são variáveis aleatórias e a e b são constantes.

Então,

$$E(Z) = E(aX + bY) = aE(X) + bE(Y)$$

$$Var(Z) = Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$$

Seja $W = cX + dY$, onde c e d são constantes.

Então,

$$Cov(Z,W) = Cov(aX + bY, cX + dY) = acVar(X) + (ad+bc)Cov(X,Y) + bdVar(Y)$$

Exercício 4: Qual é a covariância e a correlação entre uma constante e qualquer variável aleatória?

C. Estatísticas da Amostra

Uma variável aleatória é descrita pela sua função de densidade de probabilidade. Momentos de uma variável aleatória caracterizam a fdp. Na maioria dos casos, nós não conhecemos a *população*, ou verdadeiro, valores dos momentos. A única informação que temos é uma amostra extraída da população. Os momentos da amostra dão uma *estimativa* dos valores dos momentos da população. Note que os momentos da amostra não necessariamente se igualam aos momentos da população. (Eles são iguais assintoticamente, i.e, se temos um número infinito de extrações da variável X).

Média da Amostra: A *média da amostra* é a média aritmética das observações da amostra. Vamos dizer que você tenha n observações sobre a variável X denominadas $x_1, x_2, x_3, \ldots, x_n$. A média da amostra da variável X, denotada por $\bar{X}$, é calculada como segue.

$$\bar{X} = \sum_{i=1}^{n} \frac{x_i}{n}$$

Exercício 5: Considere-se lançando uma moeda 10 vezes. Seja H (cara) = 1 e T (coroa) = 0. O valor esperado do experimento é 5, mas não existe garantia de que este será o resultado.
A média da amostra nem sempre é igual a da população, ou na verdade, a média.

Variância da Amostra: Vamos dizer que você tenha n observações sobre a variável X denominada $x_1, x_2, x_3, \ldots, x_n$. Para computar a *variância da amostra* de X, use $\bar{X}$ no lugar da média da população. Para obter uma estimativa *imparcial* da variância da população, a medida da variância da amostra é definida como

$$s_X^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{X})^2}{n-1}$$

Nota: O termo n-1 é usado em vez de n para tornar s^2 uma estimativa imparcial da população variância σ^2. s^2 como definido acima é uma estimativa imparcial da população σ^2 desde que $E(s^2) = \sigma^2$.

O desvio padrão da amostra é medido pelo s.

Covariância da Amostra: Vamos dizer que você tenha n observações sobre a variável Y denominada $y_1, y_2, y_3, \ldots, y_n$ e você tem n correspondendo observações sobre a variável X denominada $x_1, x_2, x_3, \ldots, x_n$. Suponha que você queira estimar a covariância entre variável Y *e* variável X.

A medida da *covariância da amostra* é definida como

$$s_{X,Y} = \sum_{i=1}^{n} \frac{(y_i - \bar{Y})(x_i - \bar{X})}{n-1}$$

<u>Correlação da Amostra</u>: A *correlação da amostra* entre y e x é definida como

$$r_{X,Y} = \frac{s_{X,Y}}{s_X s_Y}$$

<u>Exercício 6</u>: Acesse os dados do arquivo Excel para esta tarefa no curso da web page. O arquivo dá os retornos mensais de várias ações durante um período de 10 anos de Janeiro de 1994 a Dezembro de 2003. Calcule a média das amostras, os desvios padrões da amostra, a variância das amostras, as covariâncias da amostra, e as correlações da amostra para os retornos das ações.

D. Inferências Estatísticas

O propósito da inferência estatística é usar uma amostra para obter informações sobre a população da qual a amostra foi extraída. Assumindo que a distribuição de probabilidade da população (p.ex.. a distribuição normal), podemos usar as características da amostra para tomar decisões relativas à população maior.

Exemplo: Suponha que você queira estimar a taxa de defeitos dos novos brinquedos que sua companhia construiu. Digamos que um lote de produção consiste de 100.000 itens. Este é a população de interesse. Para propósitos de teste, entretanto, você encontrará que vale à pena examinar somente uma *amostra randômica*, digamos 150 brinquedos desta população, pois seria muito caro, ou praticamente impossível, testar todos os 100.000 brinquedos. Dada a taxa de defeitos observada na amostra randômica, você pode fazer inferências sobre a taxa de defeitos da população inteira.

<u>Testando Hipóteses</u>

Ao planejar o seu teste, você precisa formular suas hipóteses de pesquisa. Por exemplo, você pode estar interessado em saber se a expectativa de vida dos homens nos US é maior do que 65 anos. Por conseguinte, você terá duas hipóteses competidoras. A *hipótese nula* (H_0) para este estudo é $\mu = 65$. As *hipóteses alternativas* (H_a) são as declarações que se opõem à hipótese nula. Neste caso $\mu > 65$ são as hipóteses alternativas. Uma vez nossas identificadas as hipóteses, a tarefa é determinar qual destas duas hipóteses competidoras é suportada melhor pelos dados da amostra.

Devido a sua amostra não consistir da população completa, sempre há a possibilidade de extrair uma conclusão incorreta quando inferir o valor de um parâmetro da população de um parâmetro da amostra. Quando se testam hipóteses, existem dois tipos de erros possíveis.

Erro do Tipo I: Um erro do Tipo I ocorre se você rejeitar H_0 quando de fato H_0 é verdadeiro.

Erro do Tipo II: Um erro do Tipo II ocorre se você falhar em rejeitar H_0 quando de fato H_0 é falso.

Certamente gostaríamos de eliminar qualquer chance de cometer qualquer tipo de, mas isto não é possível. Para um dado tamanho da amostra, não se pode controlar ambas as chances de um erro do Tipo I e uma de erro do Tipo II. Na prática, um pesquisador geralmente escolhe controlar o erro do Tipo I. O *nível de significância* (também conhecido como o *tamanho)* de um teste é a probabilidade de um erro do Tipo I, e o nível de significância é geralmente predeterminado. Por exemplo, se o nível de significância é selecionado como 5%, então uma hipótese nula verdadeira tem somente uma chance de 5% chance de ser rejeitada falsamente. Em outras palavras, se rejeitarmos H_0, nós estamos totalmente confiantes de que ela é a decisão correta.

Bertolo, L.A.
COC - Catanduva

O *poder* de um teste é igual a um menos a probabilidade do erro do Tipo II. Isto dá a probabilidade de que você corretamente rejeitará H_0 quando H_0 é falso.

Exemplo:

Você lança uma moeda dez vezes e obtém 8 caras (H) e 2 coroas (T). Você decidirá que a probabilidade de obter uma cara é mais do que ½?

H_0: p=0.5
H_A: p>0.5

Assumindo o número de caras que você observou é distribuído binomialmente, a probabilidade de obter 8H e 2T = $\frac{10!}{8!x2!}x0{,}5^8x0{,}5^2 = 0{,}0439$

Portanto, se seguirmos o padrão comum de configurar o nível de significância a 5%, então rejeitaremos H_0 neste case (muito embora o H_0 seja verdadeiro).

Na prática, os parâmetros da população são desconhecidos, e as rejeições de H_0 podem ocorrer que são incorretas. Nós, entretanto, não sabemos que eles são incorretos é claro. Insucesso em rejeitar H_0 pode também ser incorreto.

DE QUE MAMEIRA VOCÊ PODE IMAGINAR A DISTINÇÃO ENTRE DECISÕES INCORRETAS E CORRETAS?

distribuição *t*: Usualmente, não conhecemos a variância da população, e temos que usar a variância da amostra como a sua estimativa. Se $x_1, x_2, x_3, \ldots, x_n$ é uma amostra aleatória de uma distribuição normal com média μ e variância σ^2, a variável aleatória

$$t = \frac{\bar{X} - \mu}{desv\, pad(\bar{X})} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

segue uma distribuição *t* com n-1 graus de liberdade denotada por t_{n-1}. A forma da distribuição *t* é simétrica e muito semelhante à distribuição normal padrão. É menos aguçada e tem cauda mais grossa do que uma distribuição normal padrão.

Exemplo:

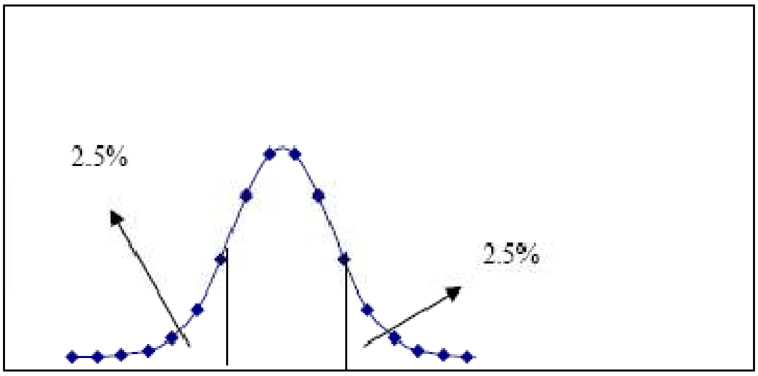
A empresa que você trabalha está considerando a recompra de algumas de suas ações em circulação. Ela acredita que a recompra é vista como uma boa notícia no mercado. Os administradores exigem que você examine como o mercado reagiu historicamente à recompra de ações.

Assim suas hipóteses são: $H_0 : \mu = 0$

$H_A : \mu \neq 0$

onde μ é a média do mercado ajustada diariamente do retorno de todas as ações anunciadas que foram recompradas calculada subtraindo o retorno de mercado do dia anunciado do retorno para a empresa *i* no dia anunciado. (Mais tarde teremos algumas preocupações acerca do mérito de tal Exercício e apontaremos vários refinamentos. VOCÊ PODE IMAGINAR QUAL PREOCUPAÇÃO ESTAMOS TENDO?)

Este é um exemplo de um teste bicaudal. Rejeitamos H_0 se o teste estatístico t computado estiver na região de rejeição determinada pelos graus de liberdade e o nível de significância do teste (probabilidade de erro do tipo I). Vamos apresentar o nível de significância a 5%. Portanto, a região de rejeição é definida de modo que a probabilidade do teste estatístico t fracassar na região pela chance (quando H_0 é verdadeiro) para 5%.



$-t_{crítico}$ $t_{crítico}$

Figura 3. Região de rejeição da distribuição *t*

Suponha que você seja capaz de colher uma amostra de 100 empresas durante os cinco últimos anos passados que anunciaram recompra de ações. O teste estatístico $\frac{\bar{X}-0}{\frac{s}{\sqrt{100}}}$ segue uma distribuição t_{99} se H_0 é verdadeiro *e* se os retornos das ações forem distribuídos normalmente. Como $\bar{X}$ é uma estimativa imparcial da média, espera-se que o valor de $\bar{X}$ seja perto de 0 se H_0 é verdadeiro, o que resultará num pequeno teste estatístico. Se H_A é verdadeiro, espera-se que $\bar{X}$ desvie substancialmente de 0, que resultaria num teste estatístico t que é grande (em valor absoluto). Rejeitamos H_0 se o teste estatístico estiver suficientemente longe o bastante de zero para cair na região de rejeição.

O valor $t_{crítico}$ de um teste bicaudal com 99 graus de liberdade e um nível de significância de 5% é 1,98. Usando as observações da amostra, é encontrado que $\bar{X}$ é 2,1% e S=5,6%. Os valores correspondentes observados da estatística t é 3,75 o qual está numa região de rejeição. Portanto, somos capazes de rejeitar as hipóteses que a reação média ao anúncio da recompra de ações é zero. Dizemos que a reação média é estatisticamente diferente de zero e positiva na média.

Em vez de ter que encontrar o $t_{crítico}$, cada vez que você realizar um teste, os pacotes estatísticos existentes (o EXCEL está incluído) fornecem os *valores-p* para você. Um valor-p é o menor nível de significância para o qual o teste estatístico t da amostra levará à rejeição de H_0. O valor-p de 3,75 com 99 graus de liberdade num teste bicaudal é 0.0298% (encontrado usando a função TDIST do EXCEL). Assim mesmo se inicialmente selecionarmos um nível de significância de 0,03% para nosso teste, ainda rejeitaríamos H_0.

E. Advertência: Bisbilhotando Dados (Escavando)

Num teste de hipótese clássico, quando se seleciona um nível de significância (digamos de 5%), estamos fixando a probabilidade de um erro Tipo I ser 5%. Isto é, uma probabilidade do teste estatístico t fracassar na região de rejeição (por chance) quando H_0 é verdadeiro é 5%. Portanto, se conduzirmos nosso experimento 100 vezes, esperamos rejeitar a hipótese nula verdadeira 5 vezes puramente pela chance somente.

Considere o seguinte cenário. Assuma 100 pesquisadores usando 100 respectivas variáveis para tentar explicar os retornos do S&P 500 index. Suponha que para 5 destas variáveis a hipótese nula de que β seja zero é rejeitada. Em outras palavras, os pesquisadores concluíram que estas 5 variáveis tenham poder de previsão para retornos S&P 500.

PODEMOS CONCLUIR QUE OS RESULTADOS OBTIDOS SÃO ESTATISTICAMENTE SIGNIFICANTES?

"Finanças is rife com dados extraídos de forma errada. David J. Leinweber, diretor administrativo da First Quadrant Corp. em Pasadena, Calif., que administra $20 bilhões de ativos, gosta de ilustrar o problema com "Stupid Data-Miner Tricks. Por exemplo, ele filtrou o CD-ROM das Nações Unidas e

Bertolo, L.A.
COC - Catanduva

descobriu que historicamente, a única melhor predição do Standard & Poor's 500-stock index foi a produção de manteiga em Bangladesh."

-Peter Coy, *Business Week*, June 16, 1997, page 40

F. Regressão Linear

Antes de rever a mecânica da regressão linear, vamos introduzir três amplas categorias de regressão, as quais são caracterizadas pela natureza dos dados que estão sendo estudados. Note que, em geral, uma regressão linear é uma ferramenta que aproxima a relação linear entre as variáveis. Estas relações não são determinísticas apesar disto. Isto é, não podemos capturar cada elemento de influência. Portanto, contaremos explicitamente com a natureza estocástica do modelo (adicionando um termo de erro).

1. Regressão Série de Tempo

A *série de tempo* é um conjunto de observações extraídas de uma entidade em diferentes instantes de tempo. Por exemplo, os retornos das ações da IBM por mês de 1960 a 1999 é uma série temporal. Podemos querer saber se estes retornos mensais estão relacionados a certas variáveis. A regressão série de tempo dos retornos mensais das ações da IBM pelas taxas mensais de inflação, para o exemplo, estima a relação série de tempo entre retornos das ações da IBM e a inflação. Em outras palavras, as variações nos retornos das ações da IBM durante o tempo estão relacionadas às variações na inflação durante o tempo? Poderíamos encontrar que quando a inflação cresce, o retorno da IBM decresce.

2. Regressão de Corte Transversal (Cross-sectional)

Um *corte transversal* é um conjunto de observações extraídas de um número de diferentes entidades num único instante de tempo. Por exemplo, os retornos de todas as ações negociadas na NYSE no mês de Junho de 1999 é um corte transversal. Uma regressão de corte transversal verifica se as diferenças nos retornos através destas ações estão relacionadas a variáveis particulares. Poderíamos encontrar talvez que quanto menor a ação, maior o retorno.

3. Regressão Painel dados

Painel de dados é um conjunto de observações que é ambas a série de tempo e o corte transversal. Portanto, podemos olhar para um painel de dados como dados de corte transversal durante um período de tempo.

Regressão Simples

Suponhamos que estamos interessados em examinar se X pode explicar Y. Em outras palavras, conhecer o valor de X nos ajudará predizer o valor de Y? Se uma relação linear entre X *e* Y existe, podemos então escrever a seguinte equação.

$$Y = \alpha + \beta X$$

Como sabemos que esta relação não é perfeita, adicionamos um termo de erro à equação para refletir a imperfeição (a natureza estocástica da relação).

$$Y = \alpha + \beta X + \varepsilon$$

onde $E(\varepsilon) = 0$.

Para estimar esta relação, desde que α e β são desconhecidos, coletamos uma amostra de observações e assumimos que para cada observação i

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Isto é chamado de *regressão Y sobre (uma constante e) X* onde Y é a variável dependente (ou explicada) e X é a variável independente (ou explicativa).

Um método para estimar α e β é chamado *Ordinary Least Squares* (OLS) e ele escolhe o $\widehat{\alpha}$ e o $\hat{\beta}$ que melhor ajustam os dados de modo que a soma dos desvios quadráticos para cada observação ($\varepsilon_i^{\,2}$)seja o mínimo.

As estimativas de α e β são

$\widehat{\alpha} = y - \hat{\beta}\, x$

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{Cov_{X,Y}}{Var_X}$$

As estimativas OLS são *imparciais* significando $E(b_1) = \beta_1$ e $E(b_2) = \beta_2$.

O critério de estimação (a soma dos erros quadráticos) está estreitamente relacionado a uma medida do ajuste conhecido as R^2. Esta medida quantifica a habilidade de *X* explicar *Y* referindo o quanto da variação em *Y* pode ser explicado pela variação em *X*.

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{\beta}x_i - \hat{\beta}\bar{x})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = \frac{Soma\ dos\ Quadrados\ Explicada}{Soma\ dos\ Quadrados\ Total}$$

onde $0 \leq R^2 \leq 1$.

Note que Soma dos Quadrados Totais = Soma dos Quadrados Explicados + Soma dos Quadrados Residuais.

Quando $R^2 = 0$, o modelo pode explicar a variação em *Y* não tão bem mas apenas usando $\bar{y}$ como uma estimativa para cada um y_i. Isto é porque, de fato, $\hat{\beta} = 0$ neste caso. Quando $R^2 = 1$, todas as observações de y_i e x_i caírem sobre uma linha perfeita, e não há residuais. Neste caso, *X* explica perfeitamente *Y*.

Uma vez tendo nossas estimativas de α e β, para empregar quaisquer inferências estatísticas destas estimativas, devemos modelar como o termo erro randômico é distribuído. Vamos assumir $\varepsilon \sim N(0, \sigma^2)$. Para testar hipóteses sobre α e β, precisamos estimar as variâncias de $\widehat{\alpha}$ e $\hat{\beta}$.

$$s_{\widehat{\alpha}}^2 = s^2\left[\frac{1}{N} + \frac{\bar{x}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right]$$

$$s_{\widehat{\beta}}^2 = \frac{s^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

onde

$$s^2 = \frac{\sum_{i=1}^{N}\varepsilon_i^{\,2}}{(N-2)}$$

é uma estimativa imparcial de σ^2.

A variável aleatória $t = \frac{(\hat{\beta} - \beta)}{s_{\widehat{\beta}}} \sim t_{(N-2)}$.

O teste de hipóteses é então conduzido como descrito na Seção D. A *estatística t* para testar hipóteses sobre α é calculada similarmente.

Algumas descobertas gerais:

1. Quanto maio o σ^2 (onde $\varepsilon \sim N(0,\sigma^2)$, maiores as variâncias dos estimadores de $\widehat{\alpha}$ e $\hat{\beta}$.
2. Quanto maior a variância da amostra de *X*, mais precisos os estimadores de $\widehat{\alpha}$ e $\hat{\beta}$.

3. Um aumento no tamanho da amostra geralmente leva a um aumento na precisão dos estimadores de $\hat{\alpha}$ e $\hat{\beta}$.

<u>Interpretação dos Coeficientes das Regressões *Ordinary Least Squares* (OLS)</u>

Suponha que procuramos identificar os determinantes dos retornos das ações. Assuma que você rode uma regressão crosssectional dos retornos das ações (em %) de um número de ações por um ano sobre uma constante e seus respectivos índices contábeis-pelo-mercado (B/M). Vamos dizer que sua equação estimada torna-se

$$r_i = 7{,}60 + 2{,}80\ B/M_i + \varepsilon_i$$

Assuma que ambos o intercepto e os coeficientes inclinação são estatisticamente significativamente diferentes de zero (sua estatística-t excede o valor $t_{critical}$). Estes coeficientes podem ser interpretados como segue.

<u>Intercepto:</u> Superficialmente, uma ação com um índice B/M zero é estimada ter um retorno mensal de 7,6%. Além disto, o intercepto captures os efeitos médios *diferentes de zero* que não são relacionados a B/M. Recorde que E(ε) = 0; assim enquanto o termo erro também captures efeitos não relacionados a B/M, estes efeitos tem um impacto zero na média.

<u>Inclinação:</u> Todas as ações começam com um retorno de 7,6% (o valor do intercepto). A inclinação então estima a sensibilidade de um retorno de ação pelo índice B/M da ação. Para um acréscimo unitário no índice B/M de uma ação, o retorno anual cresce por 2,8%.

Para prever o retorno da ação para o próximo ano, precisaríamos estimar o índice B/M, digamos 1,13. Daí então o retorno previsto é

$$r_i = 7{,}6 + 2{,}8\,(1{,}13) = 10{,}76\%$$

<u>Exercício 7</u>: Usando os dados para esta tarefa (nas web-page), faça a regressão dos retornos de cada ação numa constante e os retornos S&P 500 de Janeiro de 1994 a Dezembro de 2003. Teste respectivamente que o intercepto e a inclinação são cada um diferentes de zero no nível de significância 5%. Discuta quão bem o modelo explica os retornos de cada uma das ações.

G. Variáveis Dummy

As variáveis Dummy são usadas na análise de regressão para examinar se classes de observações relacionam diferentemente às variáveis independentes. Por exemplo, na regressão dos retornos das ações sobre os índices B/M, deveríamos considerar que a relação entre B/M e retornos das ações é diferente para pequenas empresas daquelas das grandes empresas. Defina a variável dummy SMALL para ser

$$\mathrm{SMALL_i} = \begin{cases} 1 & \text{se a capitalização do mercado } \leq \$\ 1 \text{ bilhão} \\ 0 & \text{caso contrário} \end{cases}$$

Poderíamos então rodar a seguinte regressão

$$r_i = \alpha + \beta_1 B/M_i + \beta_2 (B/M_i \times SMALL_i) + \varepsilon_i$$

onde (B/M × SMALL) é chamado de termo de interação.

Para as grandes empresas, a interpretação da regressão permanece a mesma que antes desde que β_2 se iguala a zero para estas empresas. Para as pequenas empresas, a inclinação da relação entre retornos e B/M não pode mais ser medido por apenas β_1. A inclinação para pequenas empresas é $(\beta_1 + \beta_2)$. Se β_2 é estatisticamente diferente de zero (usando um teste-t) então concluímos que os retornos das ações de pequenas empresas tem uma relação diferente com B/M daquela dos retornos das grandes empresas.

Exercício 8: Suponha que o intercepto da regressão dos retornos das ações sobre B/M e (B/M×SMALL) seja 7,6%, β_1 seja 1,7%, e β_2 seja 2,4%. Preveja os retornos das ações de uma empresa cuja capitalização de mercado seja $50 bilhões e cujo índice B/M seja 0,95 e de uma empresa cuja capitalização de mercado seja $950 milhões e cujo índice B/M seja 1,75.

Soluções dos Exercícios Práticos

1. Dois resultados possíveis: cara ou coroa. 50% probabilidade para cada um.

2.

Soma	Modos de obter	Prob
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	6	6/36
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36

3. Usando a função densidade precedente, a probabilidade de obter menos do que 4 é encontrada por acumulação das probabilidades das somas abaixo de 4, i.e. 2 e 3.

$$\Pr(2) + \Pr(3) = \frac{1}{36} + \frac{2}{36} = \frac{3}{36}$$

4. Cov(a,X) = Corr(a,X) = 0. Como a é uma constante, ela sempre emprega o mesmo valor sem desvio (a despeito do valor de X).

5. (Exercício imaginação)

6.

	Média	Desv. Pad.	Variância
Dell	5,04%	15,68%	0,0246 = 246%^{2}
Eastman Kodak	0,21%	8,86%	0,0079
GE	1,48%	7,12%	0,0051
VW Market Index	0,94%	4,67%	0,0022

Matriz de Covariância:

Variância da Dell difere ligeiramente da acima por 120/199 porque a função COV no Excel assumes que ela tenha a população, não uma amostra

0,0244			
0,0029	0,0078		
0,0029	0,0017	0,0050	
0,0037	0,0014	0,0022	0,0022

Matriz de Correlação:

1,0000			
0,2103	1,0000		
0,2608	0,2757	1,0000	
0,5122	0,3356	0,6623	1,0000

Dell tem uma correlação maior com o Mercado que a Eastman does indicando que os movimentos do Mercado explicam a maior parte dos movimentos nos preços da Dell.

7. Regressão dos retornos da Dell sobre os retornos da S&P 500.

RESUMO DOS RESULTADOS

Estatística de regressão	
R múltiplo	0,510062
R-Quadrado	0,260163
R-quadrado ajustado	0,253893
Erro padrão	0,135455
Observações	120

ANOVA

	gl	*SQ*	*MQ*	*F*	*F de significação*
Regressão	1	0,761346	0,761346	41,49456	2,67E-09
Resíduo	118	2,165075	0,018348		
Total	119	2,926422			

	Coeficientes	*Erro padrão*	*Stat t*	*valor-P*	*95%Infer.*	*95%Super.*	*Inferior 95,0%*	*Superior 95,0%*
Interseção	0,035796	0,01257	2,847681	0,005196	0,010904	0,060689	0,010904	0,060689
Variável X1	1,752193	0,272011	6,441627	2,67E-09	1,213538	2,290848	1,213538	2,290848

Os resultados da regressão para a Dell indicam que a Dell é muito sensível aos movimentos da S&P 500. Um aumento de 1% nos preços S&P 500 tipicamente geram um aumento de 1,75% nos preços da Dell. O teste estatístico t testando a hipótese nula de que a inclinação é zero nesta regressão é 6,44, que é altamente significativo. O valor-p interpreta a estatística-t de 6,44 para nós e mostra que a probabilidade de ter a inclinação de 1,75, dado que a verdadeira inclinação é zero, é minúscula a 0,000000267%. Estamos, portanto, muito confiantes de que os retornos da Dell estão relacionados aos retornos da S&P 500.

Agora estamos confiantes de que há uma relação entre Dell e o S&P 500, como se pode descrever do informativo o retorno S&P 500 está para os retornos da Dell? O R-Quadrado indica que 26% da variação nos retornos da Dell durante o período de amostragem é explicada pelo retorno S&P 500. O restante 74% da variação não é explicado pelo S&P 500. Dado que estamos explicando os retornos, os quais são altamente variáveis, o R-Quadrado de 26% é relativamente alto, como você verá mais tarde na aula.

8. $E(r_i) = 7{,}6 + 1{,}7(0{,}95) + 2{,}4(0) = 9{,}215\%$

$E(r_j) = 7{,}6 + 1{,}7(1{,}75) + 2{,}4(1{,}75) = 14{,}775\%$