Aula 6

PROBABILIDADES: VARIÁVEL ALEATÓRIA CONTÍNUA E DISTRIBUIÇÃO NORMAL

META

Estudar o comportamento e aplicação das Variáveis Aleatórias Contínuas, bem como da Distribuição Normal.

OBJETIVOS

Ao final desta aula, o estudante deverá:

Saber utilizar as variáveis aleatórias contínuas e calcular seus principais parâmetros: Esperança Matemática e Variância.

Saber o conceito e propriedades da distribuição Normal e como utilizá-la?

PRÉ-REQUISITO:

Conhecimentos dos teoremas das probabilidades. Também são importantes: Papel, Calculadora ou Computador para realização dos cálculos.

INTRODUÇÃO

Olá! Tudo bem? Vamos dar seqüência ao estudo sobre probabilidades? Para isto é fundamental revisar todos os conceitos de estatística e probabilidades estudados até a aula anterior, para maior facilidade e compreensão deste estudo.

Os temas que serão tratados nesta aula fazem parte de uma seqüência lógica do que foi visto até o momento sobre o estudo das probabilidades.

Primeiro você vai entender o que seja uma variável aleatória contínua e suas principais características. Como se sabe, as variáveis aleatórias contínuas resultam, em geral, de mensurações (medidas) e podem assumir quaisquer valores em um intervalo real, consequentemente o número de valores possíveis de uma variável aleatória contínua é infinito não numerável.

A partir da elaboração da Função Densidade de Probabilidade de uma variável contínua você pode calcular seus principais parâmetros que são: Esperança Matemática (média) e Variância.

Entre as distribuições de variável aleatória contínua a Distribuição Normal ou Distribuição de Gauss é a mais utilizada e com certeza uma das mais importantes distribuições de probabilidades, uma vez que para amostras grandes vários fenômenos estudados convergem para uma curva normal. Também, pelo importante Teorema do Limite Central, se o tamanho da amostra for suficiente grande, a média de uma amostra aleatória terá uma distribuição aproximadamente Normal.

Conforme pode observar a maioria das variáveis econômicas e sociais tendem para normalidade a medida que se aumenta o tamanho da amostra. Apresentaremos uma metodologia detalhada de uso desta distribuição a partir da utilização da distribuição normal padronizada que permite um cálculo mais rápido da probabilidade de determinada variável vir a ocorrer em certo intervalo, isto é: em uma determinada área.

VARIÁVEL ALEATÓRIA CONTÍNUA

Seja X uma variável aleatória. Se o número de valores possíveis de X for infinito não numerável, denomina-se X de Variável Aleatória Contínua. Como se sabe, as variáveis contínuas resultam, em geral, de mensurações (medidas) e podem assumir quaisquer valores em um intervalo real. Assim, comprimento, área, altura, volume, peso, pressão, renda familiar, etc, são alguns exemplos de variáveis contínuas.

Função Densidade de Probabilidade.

Seja X uma variável aleatória contínua. A função f, denominada função densidade de probabilidade (fdp) de X satisfaz as seguintes condições:

a) $f(x) \geq 0$, para todo $x \in R$

b) $\int f(x)\,dx = 1$ - Além disso, para qualquer
a < b em R, $P(a < X < b) = \int_b^a f(x)\,dx$

Observações:

1) $P(X = x_0) = 0$, visto que: $P(X = x_0) = \int_{x_0}^{x_0} f(x)\,dx$

2) $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$, caso X seja uma variável aleatória continua.

3) $f(x)$ não é probabilidade. Somente quando a função for integrada entre dois limites, ela gera uma probabilidade, representando a área sob a curva da função entre dois valores a < b.

Parâmetros de uma Variável Aleatória Continua.

Esperança Matemática (Média): $E(X) = \int_{-\infty}^{\infty} x\, f(x)\,dx$

Variância: - $V(X) = E(X^2) - [E(X)]^2$

Exemplo: Seja X uma variável aleatória contínua com a seguinte função densidade de probabilidade:

$f(x) = 1/2x$ se $0 \leq x \leq 2$

$f(x) = 0$ caso contrário

a) $P(1 \leq X \leq 1{,}5) = \int_1^{1,5} 1/2x\,dx = 1/2\int_1^{1,5} x\,dx = 1/2\left(\frac{1{,}5^2}{2} - \frac{1^2}{2}\right) = 1/4(2{,}25-1)$

$= 1{,}25/4 = 5/16$

b) $E(x) = \int_0^2 x.f\ (x)\ dx = \int_0^2 \frac{1}{2}\ x^2.dx = \frac{1}{2}\int_0^2 x^2.dx = \frac{1}{2}\left(\frac{2^3}{3} - \frac{0^3}{3}\right) = \frac{4}{3}$

c) $V(x) = E(x^2) - [E(x)]^2$

$E(x^2) = \int_0^2 x^2 \frac{1}{2}\ x.dx = \int_0^2 \frac{1}{2}\ x^3 dx = \frac{1}{8}\ (2^4 - 0^4) = 2$

$V(X) = 2 - \left(\frac{4}{3}\right)^2 = 2 - 16/9 = 2/9$

d) $DP = \sqrt{\frac{2}{9}} = \frac{1}{3}\ \sqrt{2} = 0{,}47$

Distribuição Normal ou Distribuição de Gauss

Trata-se de uma das mais importantes distribuições de probabilidades, uma vez que para amostras grandes vários fenômenos estudados convergem para uma curva normal. Também, pelo importante Teorema do Limite Central, se o tamanho da amostra for suficiente grande, a média de uma amostra aleatória terá uma distribuição aproximadamente Normal.

A apresentação gráfica dessas variáveis pode ser feita através de um histograma ou polígono de freqüências. Em grande parte dos casos, esses gráficos têm aspecto similar, sugerindo uma distribuição para os dados em foram de sino, como mostra a figura abaixo:

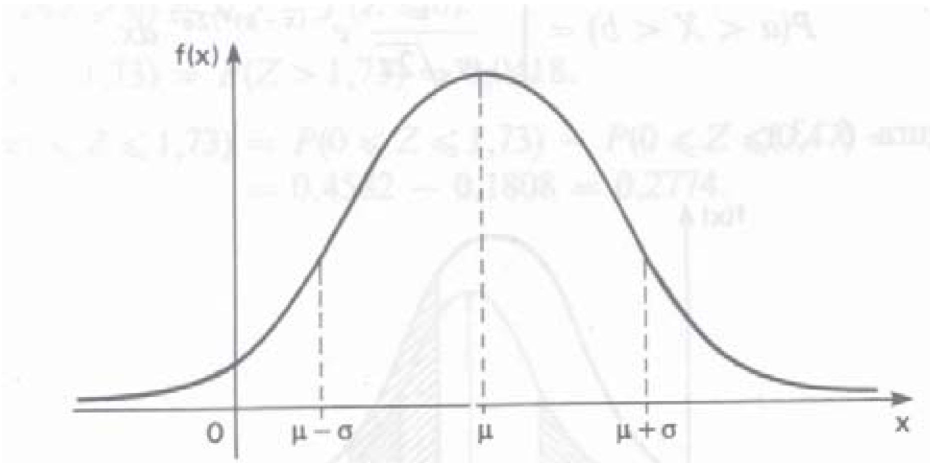
A curva sob o gráfico acima pode ser interpretada como um limite teórico para a descrição dos dados, pois à medida que aumentamos o tamanho da amostra e o intervalo de classe tende para zero as características amostrais se aproximam dos parâmetros populacionais, de tal sorte que podemos supor uma linha totalmente polida, simétrica e mesocúrtica para a população investigada. A essa curva dá-se o nome de curva normal e os dados que aderem a ela seguem a chamada distribuição normal.

A caracterização de distribuições contínuas se dá através de uma função densidade de probabilidade. É através dessas funções, por exemplo, que podemos calcular a probabilidade de uma variável aleatória pertencer a um determinado intervalo.

Dizemos que a Variável Aleatória X tem distribuição normal, com parâmetros μ e σ^2, $-\infty < \mu < +\infty$ e $0 < \sigma^2 < +\infty$, se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$ para todo x $\in$ R, Notação: **X~N(μ , σ^2)**

Os parâmetros μ e σ^2 referem-se respectivamente a média e a variância da variável aleatória X. A figura abaixo ilustra uma particular curva normal, determinada por valores particulares de μ e σ^2.



Principais Características da Distribuição Normal:

a) E(X) = μ
b) Var(X) = σ^2
c) $\mu - \sigma$ e $\mu + \sigma$ são pontos de inflexão de f(x);
d) x = μ é ponto de máximo de f(x), e o valor máximo é $1/\sigma\sqrt{2\pi}$;
e) f(x) é simétrica ao redor de x = μ, isto é: $f(\mu + x) = f(\mu - x)$, para todo $-\infty < x < +\infty$

f) A variável aleatória (x) pode assumir qualquer valor real.
g) O gráfico é uma curva em forma de sino, simétrica em torno da média μ.
h)A área total sob a curva vale 1. Essa área corresponde à probabilidade de a variável aleatória assumir qualquer valor real.
i) Como a curva é simétrica em torno da média, os valores eqüidistantes da média ocorrem com igual probabilidade.

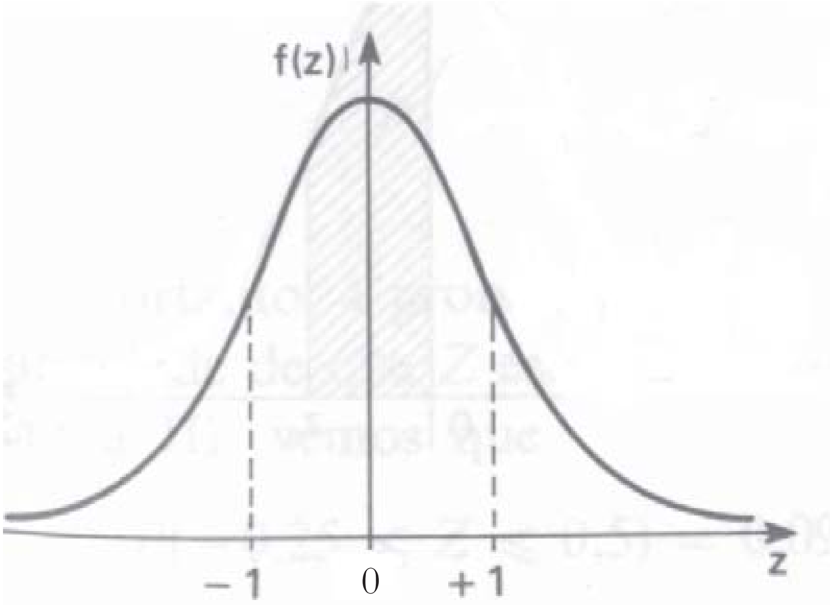
Dificuldades de uso de uma Distribuição Normal:

1. Integração de f (x), pois para o cálculo é necessário o desenvolvimento em séries;
2. Elaboração de uma tabela de probabilidades. f(x) depende de dois parâmetros, μ e σ, acarretando enorme trabalho para tabelar essas probabilidades considerando-se as várias combinações de μ e σ.

Em função dessas duas dificuldades a solução ideal é transformar a variável (Xi) em uma outra variável (z), gerando uma Distribuição Normal Padronizada com parâmetros conhecidos, isto é: media igual a 0 e variância igual a 1. A variável aleatória "z" definida por: z = (Xi - μ)/σ tem uma Distribuição Normal Padronizada N(0;1).

E[z] = E[Xi - μ]/ó = 1/ ó{E[Xi - μ]} = 1/ ó{E[Xi] –E[μ]} = 1/ ó{μ - μ} = 0

Var[z] = Var[Xi - μ]/ó = 1/ ó²{Var[Xi - μ]} = 1/ ó²{Var[Xi]} = ó² / ó² = 1



Suponha, então, que X~N(μ, s^2) e queiramos determinar P(X1 $\leq$ X $\leq$X2) =

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}}.e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

Usando a Distribuição Normal Padronizada esta probabilidade passa a ser encontrada da seguinte forma:

$P(1 \leq X \leq X2) = P\{(Xi - \mu)/\sigma \leq z \leq (Xi - \mu)/\sigma\}$

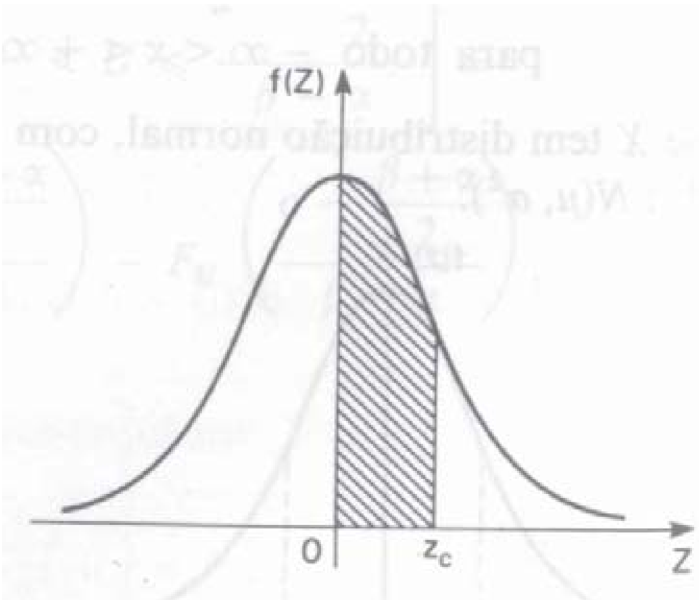
As probabilidades são obtidas a partir da Tabela da Distribuição Normal Padronizada. Esta tabela fornece as probabilidades para valores de "z" sob uma curva normal padrão, probabilidades estas equivalentes aos respectivos valores da variável "X", visto que as áreas ocupadas sob a curva pelas variáveis "X" e "z" são iguais.

Exemplo: Uma mostra de 600 carneiros da raça Corriedale apresenta média de produção de lã igual a 60 kg com desvio padrão de 8 kg. Quantos animais produzem lã: a) Entre 60 e 70 kg. b) Acima de 72 kg.

a) $P(X1 \leq X \leq X2) = P\{(Xi - \mu)/ó \leq z \leq (Xi - \mu)/ó\}$

$P(60 \leq X \leq 70) = P\{(60 – 60)/8 \leq z \leq (70 - 60)/8\} = P\{0 \leq z \leq 1,25\} =$

*0,3944 (ver gráfico) - Número de animais = 0,3944 * 600 = 236,64 = 264 animais*



b) $P(X \geq X2) = P\{ z \geq (Xi - \mu)/ó\}$

$P(X \geq 72) = P\{ z \geq (72 - 60)/8\} = P\{ z \geq 1,5\} = 0,5 – 0,4332 =$

0,0668 = 40 animais

Aproximação da Binomial pela Normal

Uma outra razão da importância da Normal se refere à sua utilização como aproximação para outras distribuições, com no caso da Distribuição Binomial. Esta aproximação quando o número de observações ou tentativas "n" é relativamente grande, sendo, portanto aceitável quando $n \geq 30$ - como também: $np \geq$ e $nq \geq 5$.

Quando se usa a Normal para aproximar probabilidades binomiais **a média e o desvio padrão** são calculados com base nos parâmetros relativos ao valor esperado e a variância do número de sucessos para a distribuição binomial, isto é: $\mu = np$ e $\sigma = \sqrt{npq}$.

Então $E(x) = \sum_{i=1}^{n}(x_i) = np$ e $Var(x) = \sum_{i=1}^{n} Var(x_i) = npq$

Logo, para n suficientemente grande, a variável é:

$$Z = \frac{X - np}{\sqrt{npq}} \cong N(0,1)$$

Para melhorar ainda mais a aproximação, usaremos o recurso da correção de continuidade, como segue:

$$P(X \geq k) \cong P(X \geq k - 0,5)$$

$$P(A \leq X \leq B) \cong P(A - 0,5 \leq X \leq B + 0,5)$$

Ex 01: Lançam-se uma moeda 20 vezes. Qual a probabilidade de se obter de uma a cinco caras, usando: Aproximação da binomial pela normal.

Solução: Seja: X: B(20,1/2),

$\mu = np = 20 * 1/2 = 10$ e $\sigma^2 = npq = 20 * 1/2 * ½ = 5$ e

$\sigma = 2,24$. Portanto: $P(A \leq X \leq B) \cong P(A - 0,5 \leq X \leq B + 0,5)$

$P(1 \leq X \leq 5) \cong P(1 - 0,5 \leq X \leq 5 + 0,5) \cong P(0,5 \leq X \leq 5,5)$

$P(1 \leq X \leq 5) \cong P\{(0,5 - 10)/2,24 \leq X \leq (5,5 - 10)/2,24\} \cong$

$P(-4,24 \leq X \leq -2,01) \cong 0,5 - 0,4778 = 0,0222$

Ex 02: Suponha que uma variável aleatória tenha distribuição binomial com parâmetros n = 10 e p = 1/2 e queiramos calcular $P(Y \geq 7)$.

$\mu = np = 10 * ½ = 5$ e $\sigma^2 = npq = 10 * 1/2 * 1/2 = 2,5$

e $\sigma = 1,58$

Portanto: $P(X \geq 7) \cong P(X \geq 6{,}5) \cong P\{X \geq (6{,}5 - 5)/1{,}58\}) \cong$

$P\{X \geq 0{,}95) = 0{,}1711$

Para um "z" N(0,1). Utilizando a Tabela da binomial, vemos que a probabilidade verdadeira é 0,172.

Observação: Esse ajuste de meia unidade chamado de correção de continuidade é necessário, visto que no processo de aproximação pela normal, um evento discreto tem de ser considerado como existindo em um intervalo contínuo de valores, cujos limites estão em torno do valor de referência ou desse intervalo com afastamento de meia unidade para menos e/ou para mais.

ATIVIDADES

1. Seja X uma variável aleatória contínua com a seguinte função densidade de probabilidade:

$f(X) = 5X^3$ se $0 \leq X \leq 1$

$f(X) = 0$ caso contrário

Calcular: Esperança Matemática, Desvio Padrão e $P(X \leq 0{,}5)$

2. Sabendo que uma variável aleatória continua apresenta a densidade de probabilidade:

$f(X) = 0{,}5\,X$ no intervalo de $\{0 ; 2\}$, determinar:

Esperança Matemática, Desvio Padrão e $P(X > 0{,}5)$.

3. A pintura de um cartaz de propaganda consome certo tempo de execução, que corresponde a uma variável aleatória contínua cuja função densidade é indicada por: $F(X) = 9X^2 - 8X^3$ no intervalo de $\{0 ; 1\}$, determinar:

a) A probabilidade que se gaste menos de meia hora (1/2) para a pintura.

b) A probabilidade para que o tempo gasto se situe entre 1/2 e 3/4 de hora.

c) A Esperança Matemática e o Coeficiente de Variação.

4. Certa peça de reposição para veículos automotores tem duração média de 15.000 km com desvio padrão de 1.000 km, dependendo das condições de uso, e distribuem-se normalmente. Se for oferecida para esta peça uma garantia de 13.000 km, quantas peças deveriam ser substituídas antes de completar-se a garantia? Sabe-se que a venda anual dessas peças é de 23.000 unidades.

5. Uma mostra de 600 carneiros da raça Corriedale apresenta média de produção de lã igual a 60 kg e desvio padrão igual a 8 kg. Quantos animais produzem lã acima de 70 kg?

6. Suponhamos que o peso de uma população de suínos está normalmente distribuído com media = 230 kg e s = 20. Qual a probabilidade de ocorrência de suínos com pesos entre 220 e 280 kg?

7. Dois estudantes foram informados de que alcançaram as variáveis reduzidas (z) de 0,8 e – 0,4, respectivamente, em um exame de múltipla escolha de matemática. Se seus graus foram 88 e 64, respectivamente, determinar: média e desvio padrão dos graus do exame.

8. Certa empresa necessita, para sua linha de montagem, de determinado tipo de arruelas. Estas arruelas devem estar normalmente distribuídas com média de 0,502 mm e desvio padrão de 0,005 mm de diâmetro. A finalidade, para a qual estas arruelas são necessárias, permite uma tolerância máxima da medida média mais ou menos um desvio padrão. Caso isso não ocorra, a linha de montagem rejeita as peças. Qual a quantidade de peças que podem ser rejeitadas em um lote de 10.000 arruelas?

CONCLUSÃO

Com esta aula, você deve ser capaz de entender o que é uma variável aleatória contínua e de que forma ela assume valores em determinado espaço amostral. Vai perceber que diferentemente da variável aleatória discreta que assume valores pontuais, a variável contínua não tem existência real em determinado ponto, como os valores assumidos por esta variável são infinitos, suas probabilidades estão associadas à ocorrência da variável investigada em determinado intervalo de área. A soma contínua destas áreas no intervalo de menos infinito a mais infinito corresponde ao espaço amostral da variável, sendo, portanto igual a um.

Com a elaboração da Função Densidade de Probabilidade de uma variável contínua você pode calcular seus principais parâmetros: A Esperança Matemática (média) e a Variância e com isto estimar o comportamento da maioria das variáveis.

Será capaz de entender que qualquer Distribuição Normal pode ser trabalhada a partir de uma distribuição normal padronizada, que é um artifício utilizado pela transformação da variável "X" pesquisada em uma outra padronizada "z", evitando deste modo cálculos longos e cansativos.

Vai saber utilizar a tabela de probabilidades de áreas de uma distribuição normal padronizada e calcular qualquer ocorrência que esteja associada a variável investigada.

Com esta aula completamos os temas selecionados para a segunda avaliação individual.

Temos uma lista de exercícios para serem resolvidos em grupos de no máximo cinco pessoas ou individual. Com certeza você vai ficar muito satisfeito com os resultados do seu desempenho.

RESUMO

Nesta aula falamos de conceitos muito importantes relativos a variáveis aleatórias contínuas e distribuição normal. A variável continua ela pode assumir qualquer valor em determinado intervalo, isto é: ela pode assumir um número infinito de valores entre dois pontos.

Também estudamos uma metodologia bem detalhada para uso da distribuição normal e pode observar que ela é uma distribuição de aplicação bastante abrangente, ocupando deste modo posição de destaque tanto na estatística teórica como na aplicada, principalmente por esta distribuição está associada à maioria dos estudos que envolvem fenômenos naturais e físicos. Além disto, a distribuição normal também é utilizada como aproximação da binomial, situação esta que ocorre quando o número de observações ou tentativas "n" é relativamente grande, sendo, aceitável quando $n \geq 30$, enquanto: $np \geq 5$ e $nq \geq 5$.

Mas a razão mais importante para uso da distribuição normal é que as distribuições das médias ou das proporções em grandes amostras tendem a ser distribuídas normalmente.

A melhor forma de estudar qualquer Distribuição Normal é através da distribuição normal padronizada, artifício utilizado pela transformação da variável "X" pesquisada em uma outra padronizada "z", gerando uma distribuição com média 0 e variância 1. Esta distribuição normal padronizada com parâmetros conhecidos possui uma tabela de probabilidades áreas destinada a calcular qualquer ocorrência que esteja associada a variável investigada.

A apresentação gráfica da distribuição normal começa a partir de um histograma ou polígono de freqüências e vão tomando o aspecto de uma curva à medida que o intervalo de classe tende para zero, enquanto o tamanho da amostra tende para infinito, em geral este diagrama tem o aspecto similar à distribuição para os dados em forma de sino.

AUTO-AVALIAÇÃO

Sou capaz de trabalhar com variáveis aleatórias contínuas?
Sou capaz de calcular os parâmetros de uma variável aleatória contínua e entender seu significado?
Sou capaz de aplicar o modelo da distribuição de probabilidades Normal a partir do uso da normal padronizada?
Sou capaz de aplicar o modelo da distribuição de probabilidades Normal como aproximação da Binomial?

PRÓXIMA AULA

Amostragem: População e amostra. Tipos de amostragem. Cadastro e amostra piloto. Nível de confiança. Estimativa da média e proporção populacional por ponto e por Intervalo.

REFERÊNCIAS

RODRIGUES, PEDRO CARVALHO. **Bioestatística**. Universidade Federal Fluminense.
FONSECA, JAIRO DA. **Curso de Estatística**. Editora Atlas.
OLIVEIRA, FRANCISCO ESTEVAM MARTINS DE. **Estatística e Probabilidade**. Editora Atlas.
TANAKA. **Elementos de Estatística**. Editora McGraw.Hill.
BARBETTA, PEDRO A. **Estatística Aplicada as Ciências Sociais**. Editora da UFSC.
GÓES, LUIZ A. C**. Estatística I e II**. Editora Saraiva.
DÍAZ, FRANCISCA; LOPES, FRANCISCO JAVIER. **Bioestatística**. Editora Thomson.