Distribuições de probabilidade de variáveis aleatórias contínuas

Universidade Estadual de Santa Cruz

Ivan Bezerra Allaman

Destaques

Aplicações das áreas de exatas

Aplicações das áreas biológicas

Distribuição Exponencial

Introdução

- É utilizada frequentemente como modelo para distribuição dos tempos entre a ocorrência de sucessivos eventos tais como:
    - clientes chegando a uma unidade de atendimento;
    - tempo de vida de um componente;

Principais características

A função densidade de probabilidade é expressa pela seguinte função:

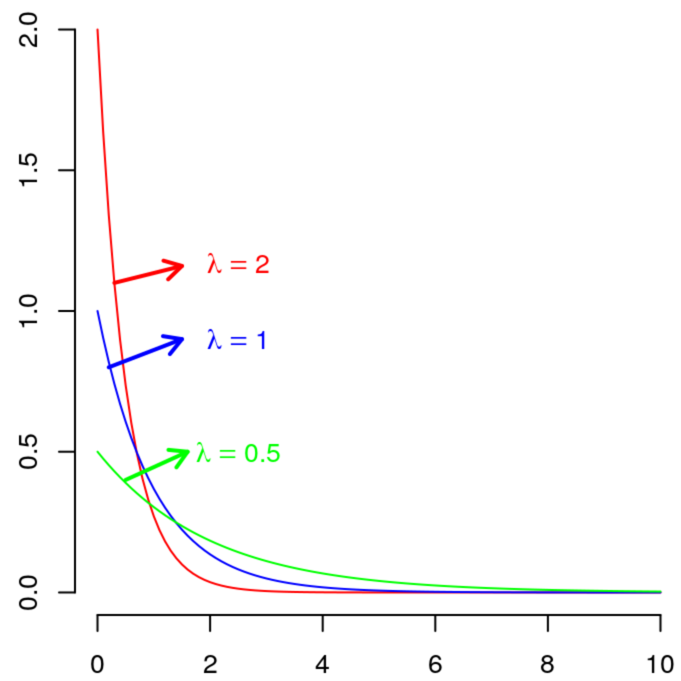
$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \texttt{caso contrário} \end{cases}$$

em que:

- λ é um parâmetro de escala (responsável pela dispersão da distribuição)
- $\frac{1}{\lambda}$ é a média da distribuição
- $\frac{1}{\lambda^2}$ é a variância da distribuição

Entendendo a distribuição

- Alterações no parâmetro λ deixam a curva mais ou menos acentuada.



Ver simulação!

Aplicações

1. Seja X = o tempo entre duas chegadas sucessivas ao guichê de atendimento rápido de um banco local. Se X possui distribuição exponencial com $\lambda = 1$, calcule os itens a seguir:

   a. O tempo esperado entre duas chegadas sucessivas

      Se X possui distribuição exponencial, então o valor esperado é $1/\lambda$. Logo, o valor esperado é igual a $1/1 = 1$.

   b. O desvio padrão do tempo entre chegadas sucessivas

      O variância é $1/\lambda^2$. Logo, o desvio padrão é raiz da variância, e portanto $\sqrt{1/1^2} = 1$.

d. $P(X \leq 4)$

Se X tem distribuição exponencial, então a distribuição está definida para X ≥ 0. Logo, para encontrarmos a probabilidade pedida, basta integrar a função densidade entre os limites 0 e 4.

$$
\begin{aligned}
P(X \leq 4) &= \int_0^4 1 \cdot e^{-1 \cdot x} dx \\
&= \int_0^4 e^{-x} dx \\
&= -e^{-x} \Big|_0^4 \\
&= (-e^{-4} - (-e^{-0})) \\
&= (-0,01832 - (-1)) = 0,9817
\end{aligned}
$$

e. $P(2 \leq X \leq 5)$

$$\begin{aligned}
P(2 \leq X \leq 5) &= \int_2^5 1 \cdot e^{-1 \cdot x} dx \\
&= \int_2^5 e^{-x} dx \\
&= -e^{-x} \Big|_2^5 \\
&= (-e^{-5} - (-e^{-2})) \\
&= (-0,00674 - (-0,13534)) = 0,1286
\end{aligned}$$

Distribuição Normal

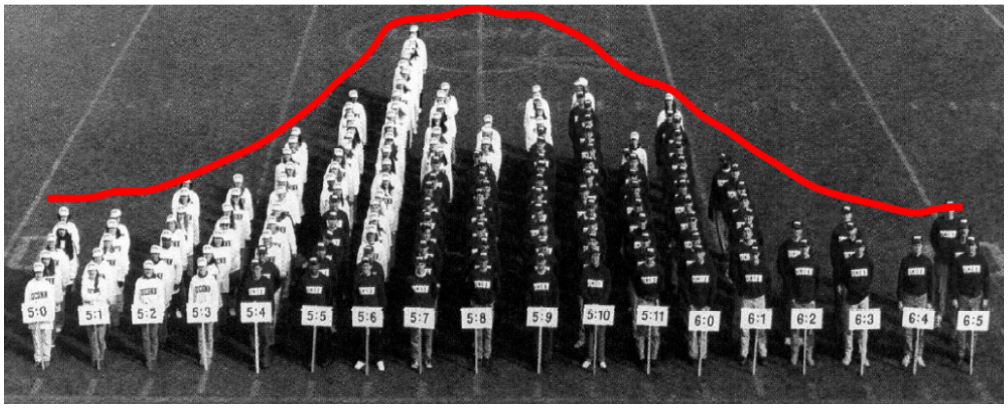
Introdução

- Foi introduzida pela primeira vez por **Abraham de Moivre** em 1733.

- Seu resultado foi estendido por **Laplace**, em seu livro *Analytical Theory of Probabilities* em 1812.

- É uma das mais importantes distribuições de probabilidade, sendo aplicada em inúmeros fenômenos.
- A forma gráfica da distribuição normal lembra um *sino*.

Imaginem que em uma escola distribuíssemos os alunos em intervalos de classes de altura. Teríamos o seguinte formato:



Principais características

A função densidade de probabilidade é expressa pela seguinte função:

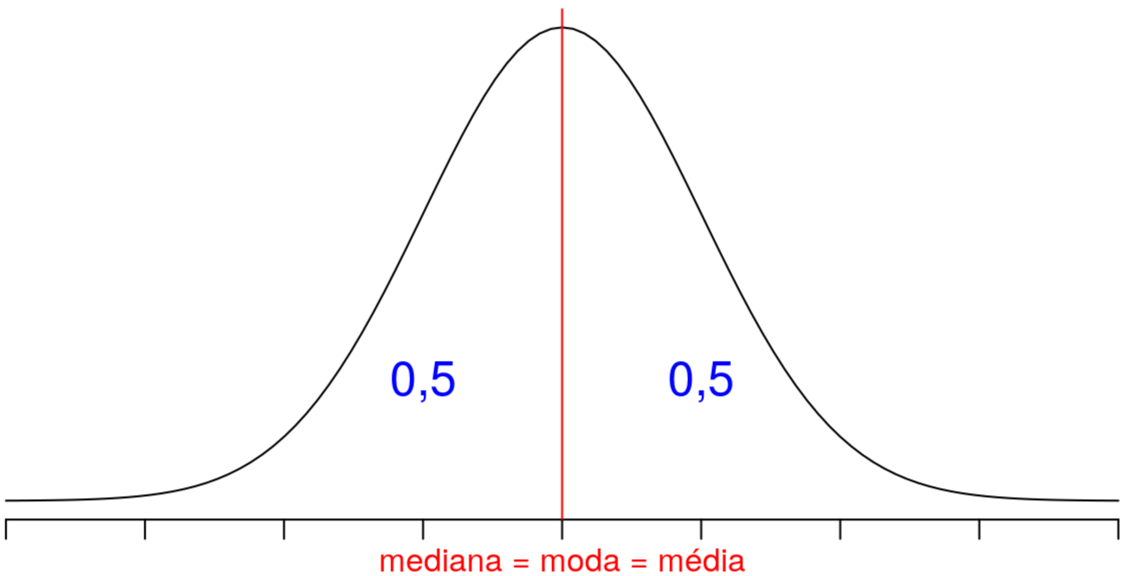
$$f(X = x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

em que:

- μ é o parâmetro de locação (média da população)
- σ é o parâmetro de escala (desvio padrão da população)
- π é uma constante igual a $3,1415\cdots$
- e é o número neperiano igual a $2,718\cdots$
- $\frac{1}{\sigma\sqrt{2\pi}}$ é um fator de escalonamento que faz com que a área sob a curva seja igual a 1

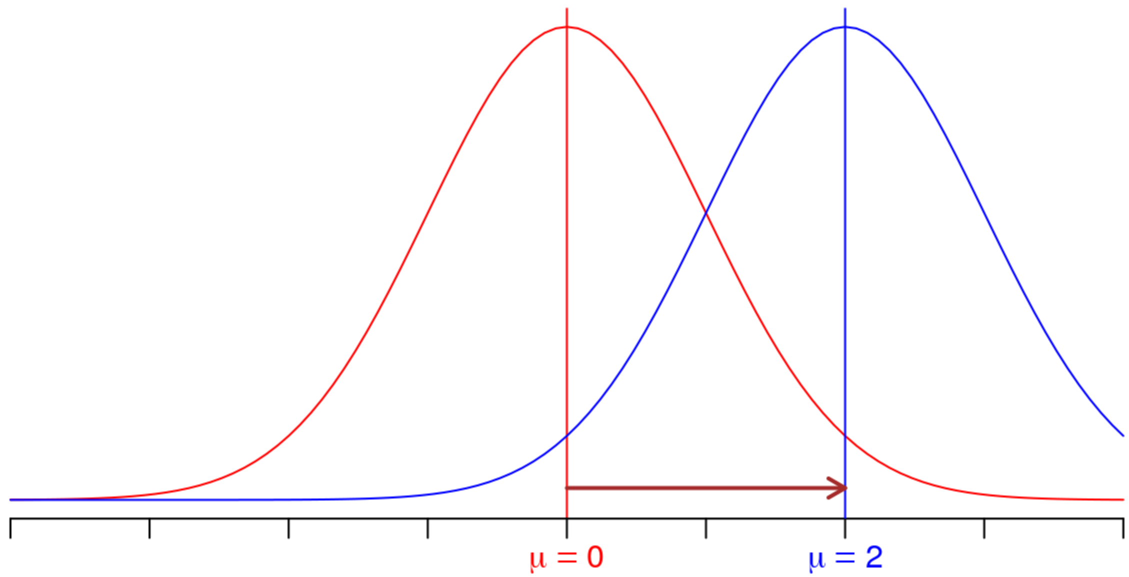
Com relação a curva tem-se que

- é simétrica em relação a média (μ)
- a média (μ), mediana (md) e moda (mo) coincidem

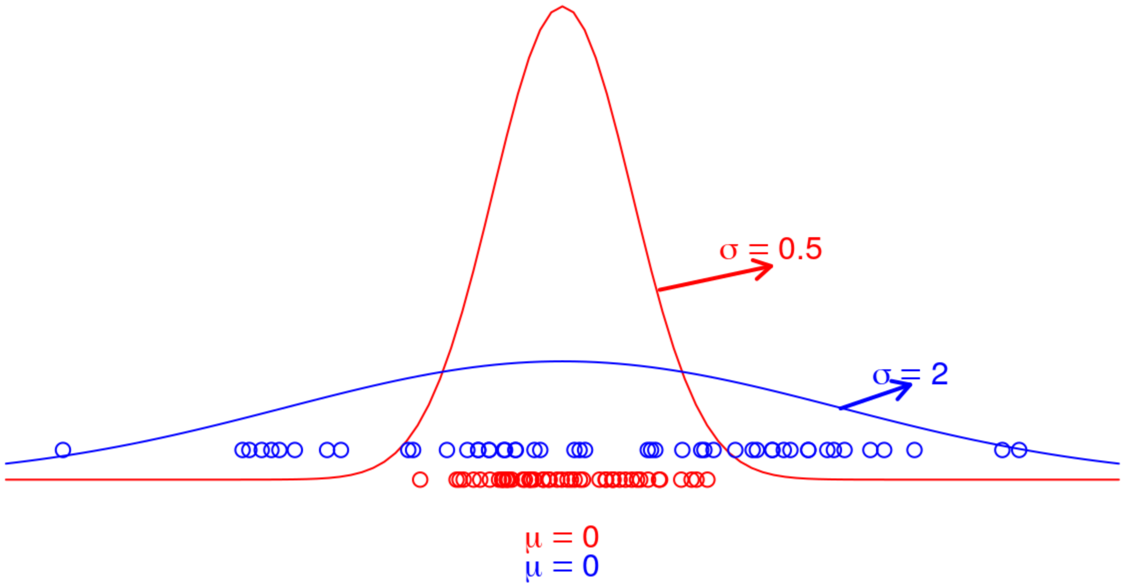


Entendendo a distribuição

- Alterações no valor da média
  - implicam no deslocamento do ponto de máximo ao longo do eixo y, sem alterações na forma básica



- Alterações no valor do desvio padrão
  - implicam em uma maior ou menor afastamento dos dados em torno da média



Distribuição normal padrão

Introdução

- Para calcularmos probabilidades utilizando a função densidade é necessário integramos a função no intervalo requerido. No caso da normal, tal integração apresenta um grau relativo de dificuldade.

- Esses problemas foram solucionados por meio de uma mudança de variável, obtendo-se, assim, a distribuição normal *padronizada* ou *reduzida*, cujo a **média** é igual a **zero** e o **desvio** igual a **um**.

$$Z = \frac{X - \mu}{\sigma}$$

- Logo, a função densidade reduziu-se a:

$$f(Z = z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2}$$

- Deste modo, temos a seguinte equivalência para cálculo de probabilidades:

$$P(X_1 \leq X \leq X_2) = P(Z_1 \leq Z \leq Z_2)$$

- Uma vez as distribuições normal e normal padrão são equivalentes, a proporção de valores caindo dentro de um, dois, três ou quatro desvios padrão da média são:

Ver simulação!

Encontrando probabilidades

- Para calcularmos probabilidades utilizando a normal padrão, veremos como funciona as função `pnorm` do R.
  - Por padrão, a função te fornece a seguinte probabilidade:

$$P(-\infty \leq Y \leq Y_1)$$

  - Utilizando o argumento `lower.tail=FALSE`, a probabilidade fornecida é:

$$P(Y_1 \leq Y \leq +\infty)$$

Ver simulação!

- Uma outra alternativa seria utilizar o aplicativo gratuito para android **Probability distribution.**

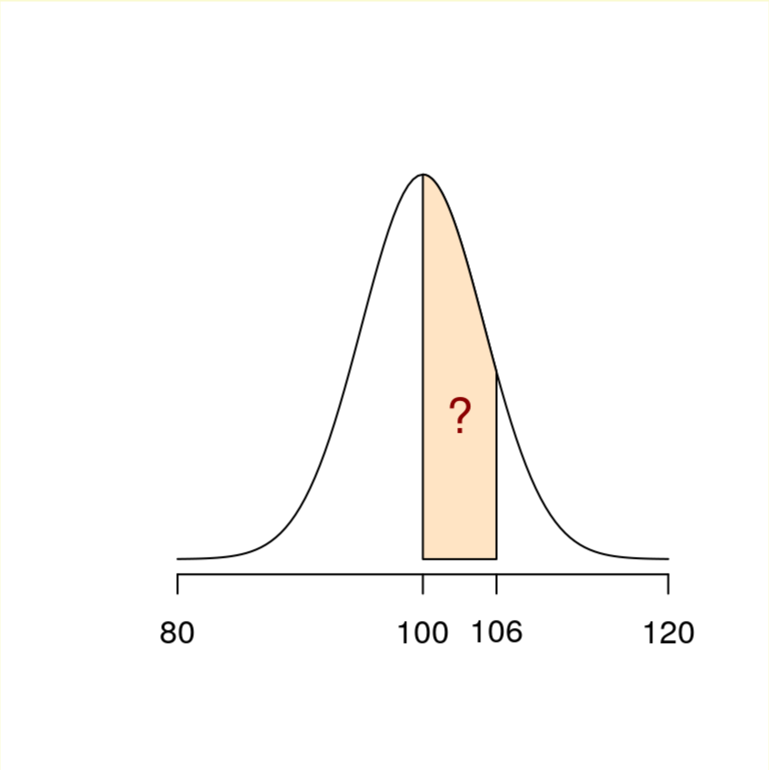
Aplicação

2. Seja $Y \sim N(100, 25)$, determine as seguintes probabilidades:

   a. $P(100 \leq Y \leq 106)$

      - Resolva por partes.
         - $\mu = 100$
         - $\sigma^2 = 25$
         - $\sigma = 5$
      - Sempre desenhe o que se pede
         - $P(100 \leq Y \leq 106)$

a. …



- Temos que transformar os quantis de Y (100 e 106) em Z para encontrarmos a probabilidade. O valor 100 não será preciso transformar pois é o valor da média e já sabemos que vale 0 em Z.

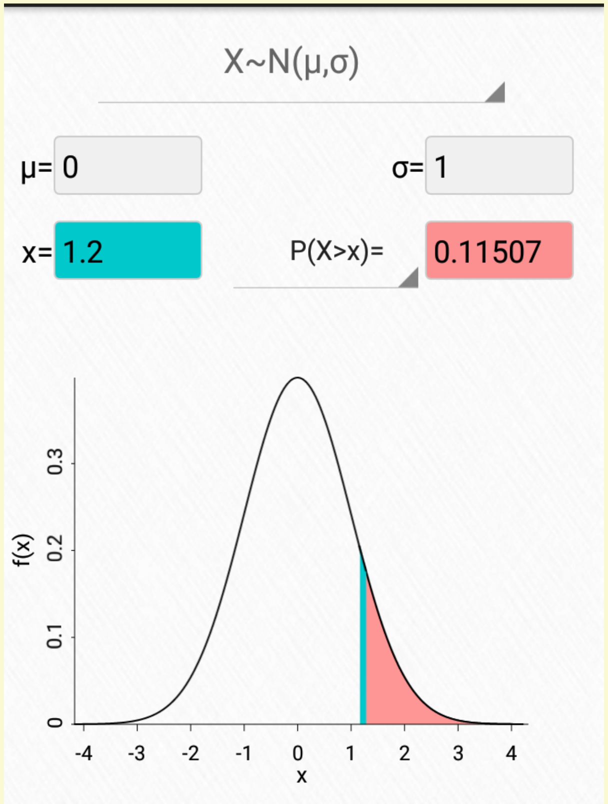
a. …

- Tem-se que para $Y = 106$ o valor de Z é:

$$Z = \frac{106 - 100}{5} = 1,2$$

- Utilizando o aplicativo Probability distribution, vamos calcular quanto é a probabilidade de Z ser maior ou igual a 1,2.

a. …



- Logo tem-se: 0,5 - 0,11507 = 0,38493

b. $P(89 \leq Y \leq 107)$

c. $P(Y \geq 108)$

Distribuição Qui-quadrado (χ^2)

Introdução

- A distribuição χ^2 é utilizado em vários problemas estatísticos, dentre eles podemos destacar:
  - Intervalos de confiança para a variância amostral de uma população normal;
  - Testes de aderência, ou seja, para verificar se os dados provenientes de uma amostra se adere a uma determinada distribuição de probabilidade (como a distribuição normal por exemplo);
  - Testes em tabelas de contingência para avaliar a relação entre duas variáveis qualitativas;
  - Testes de hipóteses para variâncias;

Principais características

- Se considerarmos uma variável aleatória com distribuição normal padrão $Z : N(0, 1)$ e elevarmos ao quadrado tal variável aleatória, então teremos uma noval variável com distribuição de probabilidade chamada **qui-quadrado** com a simbologia χ^2.
- Se somarmos o quadrado de *n* variáveis aleatórias independentes com distribuição normal padrão, também teremos uma variável aleatória qui-quadrado com $\phi = n$ graus de liberdade.

- Se consideramos n variáveis aleatórias normais independentes X_i, com mesma média μ e mesma variância σ^2, então a quantidade

$$\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2}$$

terá distribuição qui-quadrado com $\phi = n$ graus de liberdade.

- Se considerarmos os parâmetros μ e σ^2 desconhecidos e utilizarmos seus estimadores $\bar{x}$ e s^2, então a quantidade

$$\frac{\sum_{i=1}^{n}(X_i - \bar{x})^2}{\sigma^2}$$

terá distribuição qui-quadrado com $\phi = n - 1$ graus de liberdade.

- Como o estimador S^2 é igual a:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{x})^2}{n-1}$$

- Teremos uma variável aleatória χ^2 igual a:

$$\chi^2_{\phi=n-1} = \frac{(n-1)S^2}{\sigma^2}$$

- Este resultado é muito importante na inferência estatística para cálculo de intervalos de confiança e testes de hipóteses a cerca de variâncias populacionais.

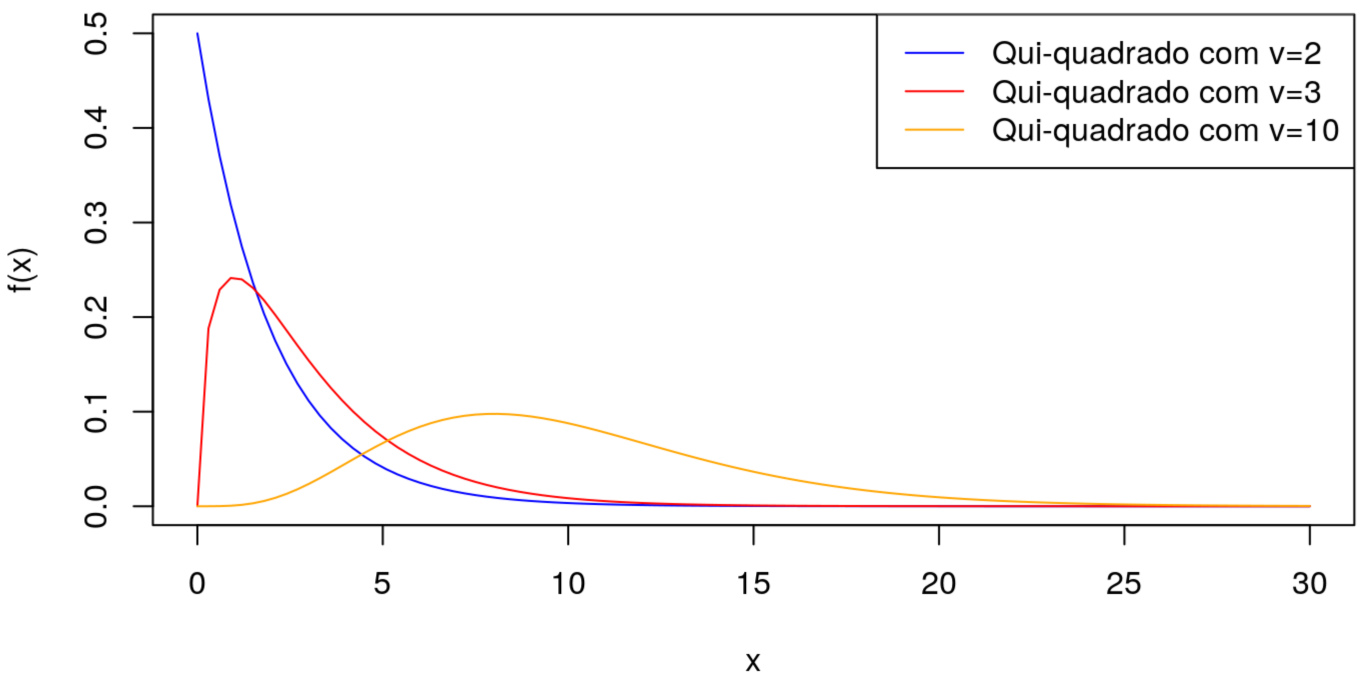
A função densidade de probabilidade é expressa pela seguinte função:

$$f(\chi^2) = \frac{1}{2^{\phi/2}\Gamma(\phi/2)} e^{-\frac{1}{2}\chi^2} (\chi^2)^{\frac{\phi}{2}-1}$$

em que:

- ϕ é o parâmetro de forma é chamado graus de liberdade.

- Abaixo alguns gráficos da distribuição qui-quadrado com diferentes graus de liberdade.



Aplicação

3. Suponha que a variância amostral seja uma medida contínua. Encontre a probabilidade de que em uma amostra aleatória de 25 observações, provenientes de uma população com distribuição normal com $\sigma^2 = 6$, irá ter uma variância amostral S^2:

   a. Maior que 9,1

   b. Entre 3,462 e 10,745

Distribuição F de Snedecor

Introdução

- Ao contrário do que muitos acham, a distribuição **F** não foi concebida por Fisher e sim, por Snedecor. O nome **F** foi em homenagem ao grandioso Fisher.
- Embora Snedecor tenha elaborado a distribuição F em termos matemáticos, Mahalanobis já havia tabulado os percentis da distribuição um tempo antes.
- A distribuição **F** tem diversas aplicações na inferência estatística clássica, dentre elas podemos destacar:
  - Intervalos de confiança para a razão entre duas variâncias;
  - Teste de hipóteses para a razão entre duas variâncias;
  - Análise de variância;

Principais características

- Sejam U e Q variáveis aleatórias independentes com distribuição qui-quadrado com $v_1 > 0$ e $v_2 > 0$ graus de liberdade, respectivamente, então a quantidade

$$\frac{\frac{U}{v_1}}{\frac{Q}{v_2}}$$

terá distribuição F, como parâmetros v_1 e v_2 graus de liberdade.

- Uma outra identidade da distribuição F é a seguinte: Se $U = v_1 S_1^2/\sigma_1^2$ e $Q = v_2 S_2^2/\sigma_2^2$ são variáveis aleatórias independentes com distribuição qui-quadrado conforme já visto na seção anterior, então a quantidade:

$$\frac{U/v_1}{Q/v_2} = \frac{\frac{v_1 S_1^2/\sigma_1^2}{v_1}}{\frac{v_2 S_2^2/\sigma_2^2}{v_2}} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

possui distribuição F se $\sigma_1^2 = \sigma_2^2$, com $v_1 = n_1 - 1$ e $v_2 = n_2 - 1$ graus de liberdade.

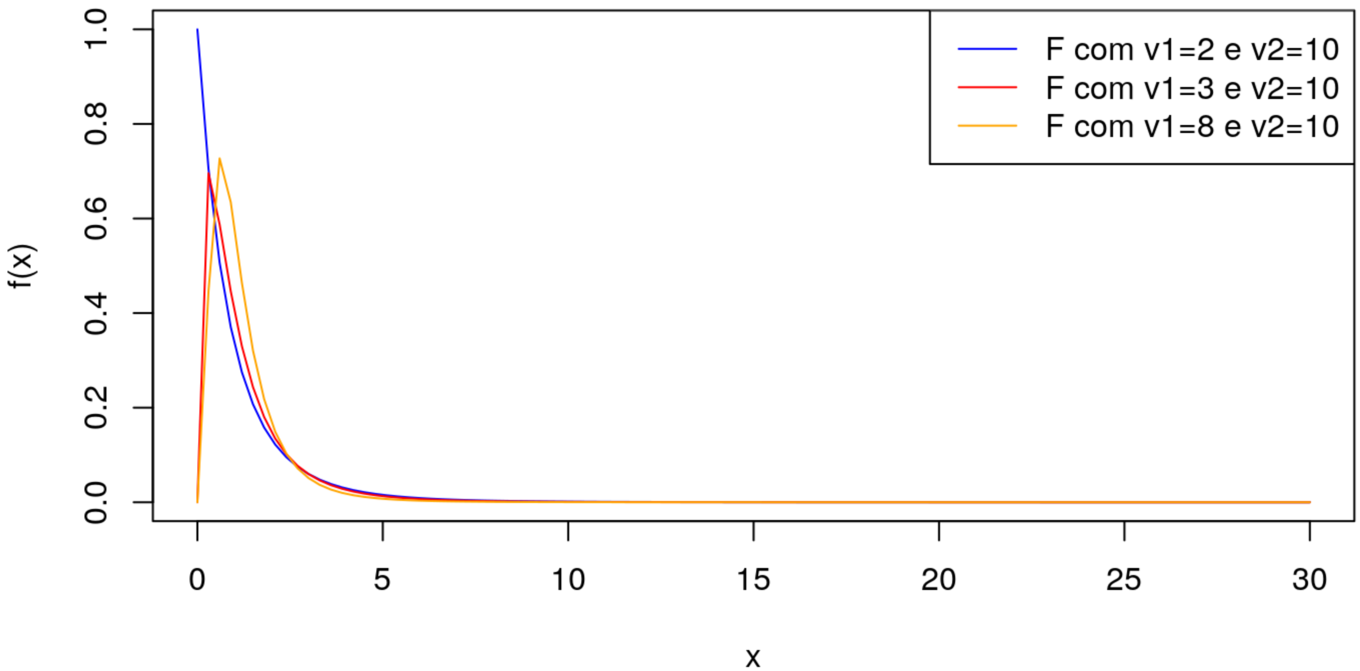
A função densidade de probabilidade é expressa pela seguinte função:

$$f(F) = \frac{\Gamma\left(\frac{\phi_1+\phi_2}{2}\right)}{\Gamma\left(\frac{\phi_1}{2}\right)\Gamma\left(\frac{\phi_2}{2}\right)} \cdot \left(\frac{\phi_1}{\phi_2}\right)^{\phi_1/2} \cdot F^{\phi_1/2-1} \cdot \left(1 + \frac{\phi_1}{\phi_2}F\right)^{-(\phi_1+\phi 2)/2}$$

em que:

- ϕ_1 é o graus de liberdade do numerador.
- ϕ_2 é o graus de liberdade do denominador

- Abaixo alguns gráficos da distribuição F com diferentes graus de liberdade.



Distribuição t de student

Introdução

- Foi elaborada por Willian Gosset em 1908.
- É frequentemente utilizada na inferência estatística quando se desconhece a variância populacional podendo ser aplicada em:
  - Intervalos de confiança para média;
  - Intervalos de confiança para diferença entre duas médias;
  - Testes de hipóteses para média;
  - Testes de hipóteses para diferenças entre duas médias, tanto para amostras independentes quanto para amostras dependentes;

Principais características

- Se **Z** é uma variável aleatória com distribuição normal padrão independente de uma outra variável aleatória **U** com distribuição qui-quadrado com **v** graus de liberdade, então a quantidade

$$\frac{Z}{\sqrt{U/v}}$$

terá distribuição t de student com v graus de liberdade.

- Uma outra identidade interessante é a seguinte: seja **Z** uma variável aleatória normal padrão expressa como

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

e **U** uma variável aleatória qui-quadrado com $v = n - 1$ graus de liberdade expressa como

$$U = \frac{(n-1)S^2}{\sigma^2}$$

então a variável **T** terá distribuição t de student com $n - 1$ graus de liberdade e terá a seguinte forma

$$\frac{\bar{x} - \mu}{S/\sqrt{n}}$$

- A distribuição t de student é simétrica como a distribuição normal e quando *n* tende ao infinito, a distribuição t se assemela a distribuição normal padrão como podemos ver na figura a seguir.