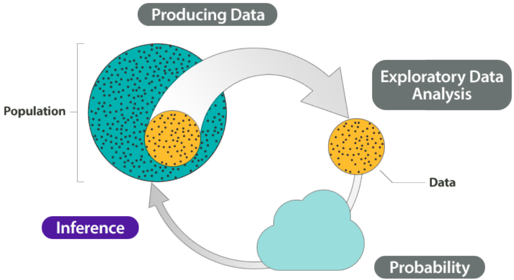
Basic Statistics: A Crash Course

Mestrado em Análise e Engenharia de Big Data
Mestrado em Matemática e Aplicações



Fonte: https://courses.lumenlearning.com

Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa

2019/20

Algumas notas introdutórias

- Este é um (mini)curso sobre conceitos básicos em Estatística pensado e desenhado para dar suporte a cursos avançados nas áreas de Ciência dos Dados/Big Data/Matemática e Aplicações

- Não dispensa trabalho, pesquisa e estudo autónomo!

- Tem como objetivo principal recordar conhecimentos iniciais básicos abordados nos 1^{os} ciclos de formação (licenciaturas), incluindo os seguintes tópicos

  1. Variáveis e vetores aleatórios
  2. Distribuições de probabilidade. Principais distribuições contínuas
  3. Inferência estatística: estimação de parâmetros e testes de hipóteses

- Atenção: Fica, em particular, de fora desta revisão toda a área de **Estatística Descritiva e Análise Exploratória de Dados**, de grande importância e que deixamos a cargo do aluno trabalhar de forma autónoma!

Variáveis e vetores aleatórios

Notação

1. Constantes/valores fixos: Letras minúsculas a itálico
   Exemplos: a, c, ℓ, x, y, z

2. Variáveis: Letras minúsculas
   Exemplos: x, y, z

3. Parâmetros (população): Letras gregas minúsculas
   Exemplos: λ, μ, ρ, σ

4. Vetores (de constantes, variáveis ou parâmetros): Letras minúsculas a negrito
   Exemplos:

   $\boldsymbol{a}$ $(\boldsymbol{a}'=(a_1,...,a_n))$, $\mathbf{x}$ $(\mathbf{x}' = (x_1, ..., x_n))$, $\boldsymbol{\mu}$ $(\boldsymbol{\mu}' = (\mu_1, ..., \mu_n))$

- Em muitas situações, a análise dos resultados de uma dada experiência aleatória passa pela descrição numérica dos resultados dessa experiência aleatória.

- Porem, o espaço de resultados Ω não é um conjunto numérico $\Rightarrow$ Descrição dos resultados através de valores assumidos por uma **variável** no decurso de uma experiência aleatória.

- Em resumo, pretende-se passar de Ω para $\mathbb{R}$ (ou mais geralmente para $\mathbb{R}^n$)

- Designa-se por **váriável aleatória** a função x de Ω para $\mathbb{R}$

$$x : \Omega \rightarrow \mathbb{R}$$

- Exemplo simples:
  - ▷ Experiência aleatória: Lançamento de uma moeda
  - ▷ Espaço de resultados $\Omega = \{\text{cara}, \text{coroa}\}$
  - ▷ Variável:

$$x = \begin{cases} 1 & \text{se o resultado é cara} \\ 0 & \text{se o resultado é coroa} \end{cases}$$

Variáveis aleatórias

- A lei de probabilidades ou distribuição de uma v.a. x pode ser descrita através de diversas funções: *função de distribuição*, *função massa de probabilidade* ou *função densidade de probabilidade*

- **Função de distribuição**: $F(x) = P(x \leq x)$ ($x \in \mathbb{R}$; $0 \leq F(x) \leq 1$) (sendo $F(x)$ diferenciável)

- **Função massa/densidade de probabilidade**:

v.a. discretas *Função massa de probabilidade*:
$f(x) = P(x = x)$ verificando as condições $f(x) \geq 0, \forall x$ e $\sum_{x \in \mathbb{K}} f(x) = 1$.

v.a. contínuas *Função densidade de probabilidade*:
$f(x) = \frac{d}{dx} F(x) \Leftrightarrow F(x) = \int_{-\infty}^{x} f(t)dt$, verificando as condições
$f(x) \geq 0, \forall x$ e $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Exemplo

Seja $p\,(0 < p < 1)$ a probabilidade de um certo acontecimento numa experiência aleatória e x a v.a. que representa a ocorrência do acontecimento, sendo 0 se ele não ocorre (insucesso) e 1 se ele ocorre (sucesso). A variável em causa é portanto discreta, assumindo os valores 0 e 1 com probabilidade $1-p$ e p, respetivamente:

$$x = \left\{ \begin{array}{cc} 0 & 1 \\ 1-p & p \end{array} \right.$$

sendo a sua distribuição de probabilidade (fmp) definida por

$$f(x) = P(x = x) = p^x(1-p)^{1-x}\ (x = 0, 1)$$

É fácil verificar que

- $p^x(1-p)^{1-x} \geq 0, \forall x$
- $\sum_{x=0}^{1} p^x(1-p)^{(1-x)} = 1$

A função de distribuição é dada por

$$F(x) = P(x \leq x) = \left\{ \begin{array}{ll} 0 & x < 0, \\ 1-p & 0 \leq x < 1, \\ 1 & x \geq 1 \end{array} \right.$$

Esta v.a. é designada por **variável aleatória de Bernoulli**.

Exemplo

Suponhamos que sobre um dado segmento de recta (a, b) se escolhe um ponto ao acaso, i.e., tal que a probabilidade de escolha seja independente da posição. A densidade de probabilidade deve ser então considerada constante, i.e.,

$$f(x) = \begin{cases} c & a \leq x \leq b \\ 0 & x < a \text{ ou } x > b \end{cases}$$

Para que $f(x)$ seja fdp deve ser não negativa e verificar

$$\int_a^b f(x)dx = 1 \Rightarrow c = \frac{1}{b-a}$$

A função de distribuição é dada por

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases}$$

Esta v.a. é designada por **variável aleatória uniforme**.

- Dada uma v.a. x chama-se *valor médio*, valor esperado ou média e representa-se por $E(x)$, μ_x ou μ a quantidade definida por,

  - no caso discreto,

  $$\mu = E(x) = \sum_{i=1}^{n} x_i p_i$$

  sendo x uma v.a. com distribuição (x_i, p_i) $(i = 1, \ldots, n)$

  - no caso contínuo,

  $$\mu = E(x) = \int_{-\infty}^{+\infty} x f(x) dx$$

  sendo x uma v.a. com densidade $f(x)$

- Dada uma v.a. x chama-se *variância* e representa-se por $Var(x)$, σ_x^2 ou σ^2 a quantidade definida por

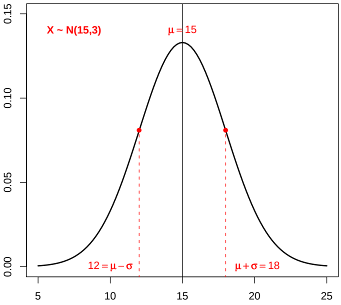
$$\sigma^2 = Var(x) = E[(x - \mu)^2] = E(x^2) - \mu^2$$

Principais distribuições contínuas

- A distribuição normal/Gaussiana é o modelo contínuo mais popular e frequentemente usado, facto que justifica a designação comum de modelo normal
- Uma variável aleatória contínua, x diz-se ter distribuição Gaussiana ou normal com valor médio μ e desvio-padrão σ, i.e., $x \frown N(\mu, \sigma)$, se a sua *fdp* corresponde a

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \; -\infty < x < +\infty, \; -\infty < \mu < +\infty, \; 0 < \sigma < +\infty$$

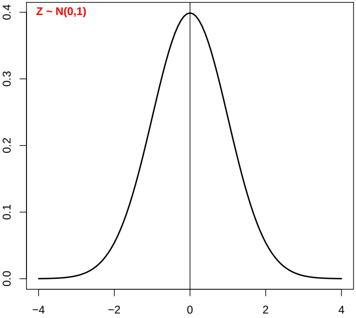
- Função densidade de probabilidade:



- Curva unimodal, sendo a moda $x = \mu$ (maximizante da fdp)
- Simétrica em relação ao eixo $x = \mu$
- Pontos de inflexão em $x = \mu \pm \sigma$
- Eixo das abcissas como assimptota

- Uma v.a. z diz-se ser uma Gaussiana padrão, normal padrão ou normal reduzida, isto é, uma Gaussiana de parâmetros $\mu = 0$ (centrada em zero) e $\sigma = 1$ (com escala unitária), $z \frown N(0,1)$, se a sua fdp for

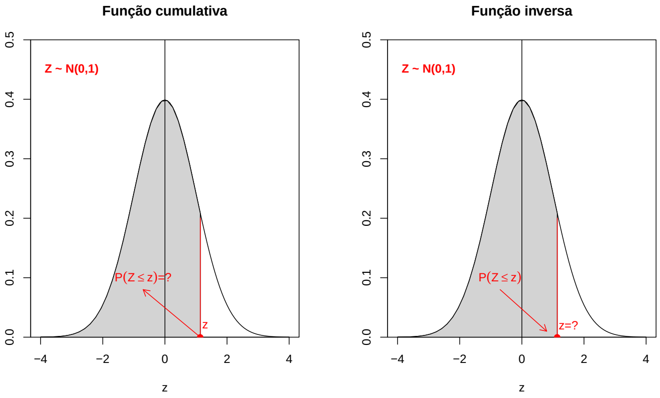
$$z \frown N(0,1) \Rightarrow f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \ -\infty < z < +\infty$$



- Padronização da Gaussiana: $x \frown N(\mu, \sigma) \Rightarrow z = \frac{x-\mu}{\sigma} \frown N(0,1)$

Função de distribuição normal padrão

- Função de distribuição da variável normal padrão:
  - Função de distribuição (Função cumulativa)$\longrightarrow \Phi(z) = P(z \leq z)$
  - Inversa da função de distribuição (Função inversa)$\longrightarrow \Phi^{-1}(p) = z_p$ $(0 < p < 1)$ (quantis da Guassiana). Genericamente z_p representa o quantil de probabilidade p (ou percentil $p \times 100$) da distribuição normal padrão, tal que $P(z < z_p) = p$



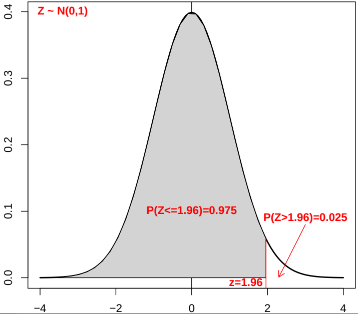
- Note-se que porque a curva é simétrica em relação a $z = 0$, $\Phi(-z) = 1 - \Phi(z)$

Exemplo

$\Phi(1.96) = P(z \leq 1.96) = 0.975$
$z_{0.975} = \Phi^{-1}(0.975) = 1.96$, denota o quantil de probabilidade 0.975 da normal padrão

Código R:

```
> round(pnorm(1.96),3)
[1] 0.975
> round(qnorm(0.975),3)
[1] 1.96
```



Distribuição Normal

Alguns teoremas relevantes:

- Seja $x \frown N(\mu, \sigma)$ e $y = a \pm bx$, então $y \frown N(a \pm b\mu, |b|\sigma)$

- Sejam $x_1 \frown N(\mu_1, \sigma_1)$ e $x_2 \frown N(\mu_2, \sigma_2)$ independentes, então

$$(x_1 \pm x_2) \frown N\left(\mu_1 \pm \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

- Generalizando $x_i \frown N(\mu_i, \sigma_i)\,(i = 1 \ldots n)$ independentes, então

$$\sum_i x_i \frown N\left(\sum_i \mu_i, \sqrt{\sum_i \sigma_i^2}\right)$$

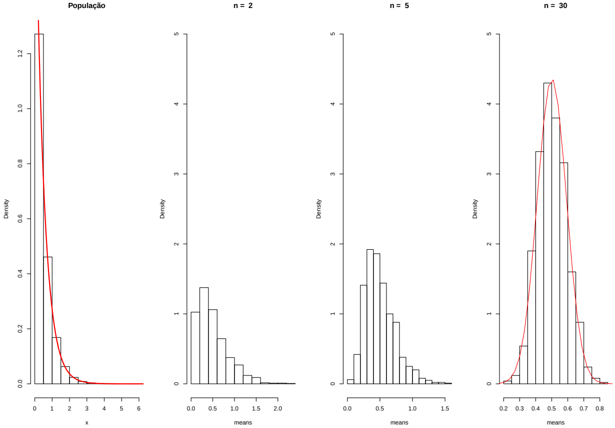
- Sejam $x_i \frown N(\mu, \sigma)$ n v.a iid, então

$$S_n = \sum_i x_i \frown N(n\mu, \sigma\sqrt{n})$$

Distribuição Normal

- **Teorema Limite Central**: Sejam $\{x_1, x_2, ..., x_n\}$ variáveis aleatórias independentes e identicamente distribuídas com valor médio μ e variância finita σ^2 e $S_n = \sum_{i=1}^{n}$ então para $n \to \infty$ tem-se

$$S_n \overset{a}{\sim} N\left(n\mu, \sigma\sqrt{n}\right)$$



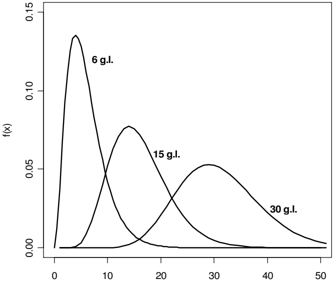
Distribuição Qui-quadrado

- Para n inteiro positivo a v.a. x $(0 \leq x < +\infty)$ com *fdp*

$$f(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)}\, x^{n/2-1}\, e^{-y/2} & x > 0 \\ 0 & \text{c.c.} \end{cases}$$

com $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$, $t > 0$, diz-se ter distribuição qui-quadrado com n graus de liberdade, i.e., $x \frown \chi^2_n$

- Função densidade de probabilidade:



Função de distribuição Qui-quadrado

- Função de distribuição da variável $x \frown \chi^2_n$
  - Função de distribuição (Função cumulativa)$\longrightarrow P(x \leq \chi^2_n)$
  - Inversa da função de distribuição (Função inversa) $\longrightarrow \chi^2_{n,p}$ (quantil de probabilidade p da distribuição χ^2_n)

Exemplo

Considere-se a v.a. $x \frown \chi^2_{10}$
$P(x \leq 5) = 0.109$
$\chi^2_{10,0.90} = 15.987$, denota o quantil de probabilidade 0.90 da distribuição χ^2_{10}

Código R:

```
> round(pchisq(5,10),3)
[1] 0.109
> round(qchisq(0.90,10),3)
[1] 15.987
```

Alguns teoremas relevantes:

- Sejam $x_i \frown \chi^2_{(n_i)}$ $(i = 1 ... p)$ independentes, então

$$\sum_{i=1}^{p} x_i \frown \chi^2_{(\sum n_i)}$$

- Seja $z \frown N(0, 1)$ então $z^2 \frown \chi^2_1$

- Seja $z_i \frown N(0, 1)$ $(i = 1, ..., n)$ então

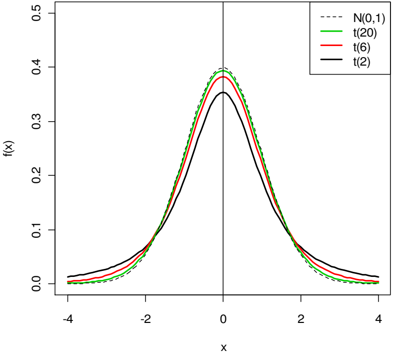
$$\sum_{i=1}^{n} z_i^2 \frown \chi^2_n$$

Distribuição t-Student

- Sejam $z \frown N(0,1)$ e $y \frown \chi^2_n$, então

$$x = \frac{z}{\sqrt{y/n}} \frown t_n$$

- Uma variável aleatória x $(-\infty < x < +\infty)$ cuja distribuição é bem modelada por uma distribuição t-Student, representa-se por $x \frown t_n$ onde n é o número de graus de liberdade (g.l.).
- Função densidade de probabilidade:



Função de distribuição t-Student

- Função de distribuição da variável $x \frown t_n$
  - Função de distribuição (Função cumulativa) $\longrightarrow P(x \leq t_n)$
  - Inversa da função de distribuição (Função inversa) $\longrightarrow t_{n,p}$ (quantil de probabilidade p da distribuição t_n)

Exemplo

Considere-se a v.a. $x \frown t_5$
$P(x \leq 1) = 0.818$
$t_{5,0.90} = 1.476$, denota o quantil de probabilidade 0.90 da distribuição t_5

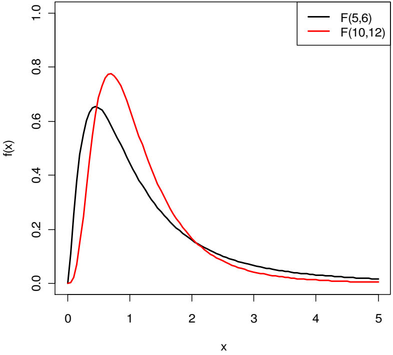
Código R:

```
> round(pt(1,5),3)
[1] 0.818
> round(qt(0.90,5),3)
[1] 1.476
```

- Uma variável aleatória x $(0 \leq x < +\infty)$ tem distribuição F-Snedecor com n (numerador) e d (denominador) graus de liberdade, e representa-se por $x \frown F(n, d)$, se a variável aleatória for tal que

$$x = \frac{y/n}{w/d}$$

onde $y \frown \chi^2_n$ e $w \frown \chi^2_d$ são v.a. independentes.
- Função densidade de probabilidade:



Função de distribuição F-Snedecor

- Função de distribuição da variável $x \frown F_{n,d}$
  - Função de distribuição (Função cumulativa)$\longrightarrow P(x \leq F_{(n,d)})$
  - Inversa da função de distribuição (Função inversa) $\longrightarrow F_{(n,d),p}$ (quantil de probabilidade p da distribuição $F_{(n,d)}$)

Exemplo

Considere-se a v.a. $x \frown F_{(5,8)}$
$P(x \leq 1) = 0.525$
$f_{(5,8),0.90} = 2.726$, denota o quantil de probabilidade 0.90 da distribuição $F_{(5,8)}$

Código R:

```
> round(pf(1,5,8),3)
[1] 0.525
> round(qf(0.90,5,8),3)
[1] 2.726
```

Inferência estatística

- $(x_1, x_2, ..., x_n)$ diz-se uma **amostra aleatória** de dimensão n se as variáveis são mutuamente independentes e têm a mesma função de distribuição $F(x)$ (cada variável x_i $(i = 1, ..., n)$ tem a mesma fdp/fmp $f(x)$)

- Ou seja, $x_1, x_2, ..., x_n$ são variáveis independentes e identicamente distribuídas, *iid*

- Note-se que uma a.a., tal como definida, não é mais que o *vetor aleatório* $\mathbf{x} = (x_1, x_2, ..., x_n)$

- Atendendo à independência entre as variáveis numa amostra aleatória, a *distribuição de probabilidade conjunta* do vetor $(x_1, ..., x_n)$ é definida pela função

$$f_{x_1,...,x_n}(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i)$$

onde $f(x_i)$ $(i = 1, ..., n)$ representa a função densidade de probabilidade (fdp) da população quando esta é contínua e a função massa de probabilidade (fmp) quando é discreta.

- Na maioria das situações a fdp/fmp da população é membro de uma *família paramérica*, isto é, a forma funcional da função depende de um parâmetro (ou de um vetor de parâmetros) θ, tal que

$$f(x_1, ..., x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

- Eem Estatística usamos os valores observados de uma amostra aleatória (conhecidos) para estimar o(s) parâmetro(s) que define(m) a população (desconhecidos)

- Assim, para um dado ponto amostral fixo, observado/conhecido, define-se a *função verosimilhança de uma amostra*

$$L(\theta|x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

Exemplo

Seja $(x_1, x_2, ..., x_n)$ uma a.a. da população Bernoulli(p), $p \in [0, 1]$ com fmp

$$f(x) = P(x = x) = p^x(1 - p)^{1-x} \; (x = 0, 1)$$

A **fdp conjunta** é dada por

$$f(\boldsymbol{x}) = f(x_1, ..., x_n) = \prod_{i=1}^{n} P(x = x_i) = \prod_{i=1}^{n} p^x(1 - p)^{1-x} = p^{\sum x_i}(1 - p)^{n-\sum x_i}$$

que fixando p, depende apenas de $\boldsymbol{x} = (x_1, ..., x_n)$.

A **função verosimilhança** é portanto dada por

$$L(p) = p^{\sum x_i}(1 - p)^{n-\sum x_i}$$

que fixando $(x_1, ..., x_n)$ depende apenas de p.

- Em regra o objetivo último da análise estatística é inferir sobre a população de estudo com base numa amostra aleatória

- O desconhecimento sobre a população pode ser total, i.e., nada se sabe sobre a sua distribuição de probabilidade

- O desconhecimento sobre a população pode ser apenas parcial, i.e., é conhecida a forma funcional da função de distribuição da população mas não se conhecem os valores do(s) parâmetro(s) que caracterizam a distribuição

- Supondo conhecida a forma geral da distribuição de probabilidade da população, a sua completa especificação implica a *estimação* do(s) parâmetro(s) θ que lhe está(ão) associado(s)

- Incluída no âmbito da inferência estatística, a **estimação de parâmetros** tem por objectivo estimar parâmetros (valores desconhecidos). Pode subdividir-se em:

  1. *Estimação pontual* - Estimação dos parâmetros através de 1 único valor (ponto), representação do verdadeiro valor do parâmetro;

  2. *Estimação intervalar* - Estimação dos parâmetros através de um intervalo de valores que, com uma certa confiança, incluirá o o verdadeiro valor do parâmetro - **Intervalos de Confiança**.

Estimador

Um estimador pontual de um parâmetro é uma **função** da amostra $(x_1, x_2, ..., x_n)$. Qualquer estatística é um estimador pontual.

Estimativa

Valor assumido pelo estimador numa amostra concreta $(x_1, ..., x_n)$.

- Existem vários métodos para encontrar estimadores. Um dos mais utilizados é o **Método da Máxima Verosimilhança**.

- Permite, em regra, obter estimadores com melhores propriedades

- O estimador de máxima verosimilhança (EMV) de θ (para simplificar vamos admitir que existe apenas 1 parâmetro desconhecido), $\hat{\theta}$, será o valor que maximiza a função verosimilhança face a uma amostra concreta.

- Uma Estimativa de Máxima Verosimilhança (EMV) é o valor do parâmetro para o qual a amostra observada é mais plausível/verosímil.

Exemplo

Suponha que numa linha de produção é registado se as unidades produzidas, de forma independente, se apresentam (1) com defeito / (0) sem defeito. Selecionadas aleatoriamente 5 unidades obtiveram-se os seguintes resultados: $\boldsymbol{x} = (1, 0, 1, 0, 0)$.

Considerando

$$\mathrm{P}(x = 1) = p \text{ e } \mathrm{P}(x = 0) = 1 - p$$

Qual a probabilidade de observar a amostra mencionada?

$$\begin{aligned} P(\mathbf{x}) &= P(x_1 = 1) \times P(x_2 = 0) \times P(x_3 = 1) \times P(x_4 = 0) \times P(x_5 = 0) \\ &= p(1-p)p(1-p)(1-p) \\ &= p^2(1-p)^3 \\ &= (\textit{se por exemplo } p = 0.4) \\ &= 0.4^2 \times 0.6^3 = 0.03456 \end{aligned}$$

Método da Máxima Verosimilhança (continuação)

Esta função, dada a amostra $x = (1, 0, 1, 0, 0)$, depende apenas do valor de p.

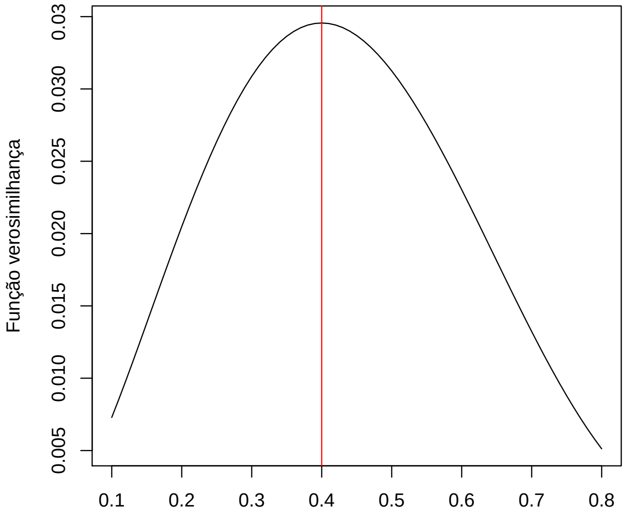
Assim, dada podemos definir a função verosimilhança dada por

$L(p|x = (1, 0, 1, 0, 0)) = p^2(1 - p)^3$

Qual o valor de p para o qual a amostra observada será mais plausível/verosímel?

	p	$L(p\|\underline{X})$
1	0.05	0.0021
2	0.15	0.0138
3	0.25	0.0264
4	0.35	0.0336
5	0.45	0.0337
6	0.55	0.0276
7	0.65	0.0181
8	0.75	0.0088
9	0.85	0.0024
10	0.95	0.0001

Método da Máxima Verosimilhança (continuação)



Método da Máxima Verosimilhança (continuação)

- Formalmente, o **Estimador de Máxima Verosimilhança** (EMV) para θ (dada uma amostra $\boldsymbol{x}$) corresponde ao **máximo (absoluto) da função** $L(\theta)$

- Assim, o EMV $\hat{\theta}$ será aquele que maximizar a função $L(\theta)$, sendo dado por

$$\frac{d}{d\theta} L(\theta|\boldsymbol{x}) = 0$$

- Note-se que a condição anterior é necessária mas não é suficiente $\Rightarrow$ Segunda derivada negativa!

- Os zeros da 1ª derivada só localizam extremos no interior do domínio da função (estudo do comportamento da função nos limites do domínio) e, por outro lado, podem ser máximos relativos ou absolutos! Queremos encontrar o máximo absoluto! $\Rightarrow$ Estudo dos limites da função nas fronteiras do domínio!

- Na maioria das situações é mais fácil trabalhar com $\log L(\theta|\boldsymbol{x})$. Determinar os extremantes de $\log L(\theta|\boldsymbol{x})$ equivale a determinar os extremantes de $L(\theta|\boldsymbol{x})$, pois a função log é estritamente crescente em $[0, +\infty[$

Exemplo

Voltando ao exemplo anterior, definiu-se uma v.a. de Bernoulli(p), $p \in [0,1]$ com **função verosimilhança** dada por

$$L(p) = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

que fixando $(x_1, ..., x_n)$ depende apenas de p. Assim, a **função log verosimilhança** é dada por

$$L^* = \log L(p) = \sum x_i \log(p) + (n - \sum x_i)\log(1-p)$$

sendo a sua 1ª derivada

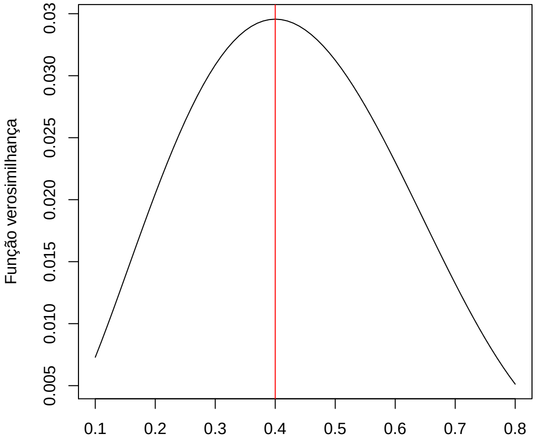
$$\frac{d}{dp}L(p|\boldsymbol{x}) = \frac{\sum x_i}{p} + \frac{\sum x_i - n}{1-p}$$

Igualando a zero

$$\frac{d}{dp}L(p|\boldsymbol{x}) = 0 \Rightarrow \hat{p} = \frac{\sum x_i}{n}$$

(Note-se que $\frac{d^2}{dp}L(p|\boldsymbol{x}) < 0$ e que $\lim\limits_{p\to 0} L^* = \lim\limits_{p\to 1} L^* = -\infty$, logo tem-se um máximo global, ou seja, $\hat{p} = \frac{\sum x_i}{n} = \bar{x}$ é o EMV de p)

Aplicando ao exemplo anterior, obtem-se $\hat{p} = \frac{\sum x_i}{n} = \frac{2}{5} = 0.4$



- Estimação pontual: um só valor para θ

- Estimação intervalar: intervalo de valores para θ

Estimador intervalar e estimativa intervalar

Seja $\mathbf{x} = (x_1, ..., x_n)$ uma a.a. recolhida de uma população com f.d. $F(x|\theta)$.

Sejam $L(\mathbf{x}) = L(x_1, ..., x_n)$ e $U(\mathbf{x}) = U(x_1, ..., x_n)$ duas estatísticas tais que $L(\mathbf{x}) < U(\mathbf{x})$ e $P(L(\mathbf{x}) \leq \theta \leq U(\mathbf{x})) = 1 - \alpha$, em que $1 - \alpha$ não depende de θ.

Então:

- o intervalo $]L(\mathbf{x}), U(\mathbf{x})[$ é um **Estimador Intervalar** designado por **Intervalo de confiança** para θ a $(1 - \alpha) \times 100\%$

- Para uma amostra concreta $\mathbf{x} = \boldsymbol{x}$, então $]L(\boldsymbol{x}), U(\boldsymbol{x})[$ representa uma **estimativa intervalar**

- $1 - \alpha$ designa-se por **Coeficiente de confiança**

- Método da **variável fulcral** (ou variável pivot):
  - Uma estatística função da amostra e de θ, cuja distribuição não depende de θ, designa-se por variável fulcral
  - Se $Q(\mathbf{x}, \theta)$ é uma variável fulcral então é possível encontrar dois valores a e b tal que
  $$P(a < Q(\mathbf{x}, \theta) < b) = 1 - \alpha$$
  - Manipulando algebricamente, obtem-se $P(h(\mathbf{x}, a) < \theta < h(\mathbf{x}, b)) = 1 - \alpha$
- Um estimador intervalar de um parâmetro θ desconhecido é portanto um intervalo aleatório que, com uma certa probabilidade de cobertura $1 - \alpha$, contem/captura o parâmetro θ
- Seja $(x_1, ..., x_n)$ v.a. iid da população $N(\mu, \sigma)$. Então têm as seguintes v.a. fulcrais:
  - $\frac{\bar{x} - \mu}{s/\sqrt{n}} \frown t_{(n-1)}$
  - $\frac{(n-1)s^2}{\sigma^2} \frown \chi^2_{n-1}$

- Genericamente, um **testes de hipóteses** é um conjunto de procedimentos estatísticos que visam determinar se certas afirmações (hipóteses), feitas sobre uma população (ou mais do que uma) são ou não suportadas pelos dados de uma amostra concreta.

- Elementos de um teste de hipóteses:

1. Hipóteses: Hipótese nula (H_0) e Hipótese alternativa (H_1)

2. Estatística do teste

3. Definição da região de rejeição (ou região crítica) e decisão sobre H_0

Hipóteses estatísticas

- Hipótese estatística é toda a proposição feita acerca da(s) população(ções) em estudo.
- Num teste de hipóteses têm-se 2 hipóteses: uma que acreditamos ser verdadeira — **Hipótese Nula** — e a sua contrária — **Hipótese Alternativa**.
- O objectivo final dum teste estatístico é decidir acerca da plausibilidade das hipóteses formuladas.
- Na formulação das hipóteses, deve ter-se em conta que:

  1. O teste é constituído por duas hipóteses complementares, a hipótese nula, H_0 e a hipótese alternativa, H_1;
  2. H_0 é admitida como verdadeira ao longo do procedimento (sob H_0);
  3. Existindo evidências contra H_0 diz-se que se rejeita a favor de H_1;
  4. As hipóteses nula e alternativa são sempre complementares e no seu conjunto estabelecem o universo de possibilidades.

- Depois de formuladas as hipóteses, coloca-se a questão "*Como decidir acerca da plausibilidade da hipótese colocada?*"

- O processo de decisão vai basear-se numa estatística — *estatística do teste* (genericamente $T \equiv T(\mathbf{x})$)— para a qual se conhece a respectiva distribuição amostral, função dos dados e que, sob H_0, não depende de parâmetros desconhecidos.

- Esta estatística, que é calculada admitindo que H_0 é verdadeira e tendo por os valores observados, vai permitir testar a plausibilidade das hipótese.

Exemplo

Suponha-se um teste ao valor médio — Hipóteses: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$
Neste caso, o parâmetro em análise é o valor médio. Recorde que, numa população onde $x \frown N(\mu, \sigma)$ (σ desconhecido), a estatística $\bar{x}$ (estimador de μ) tem distribuição t_{n-1}:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \frown t_{n-1}$$

Logo, atendendo ao valor médio fixado em H_0, i.e. para $\mu = \mu_0$, e para uma amostra concreta é possível calcular a estatística $\frac{\bar{x}-\mu}{s/\sqrt{n}}$.

Região crítica (ou região de rejeição)

- A decisão sobre a plausibilidade das hipóteses formuladas vai basear-se no conhecimento da distribuição de probabilidade da estatística de teste, sob H_0.
- Os valores mais extremos da estatística são aqueles cuja probabilidade de ocorrência é menor e, portanto, correspondem aos valores para os quais se considera improvável a validade da hipótese nula. Constituem por isso a **região crítica** ou região de rejeição
- Considere-se $H_0 : \theta \leq \theta_0$ vs. $H_1 :: \theta > \theta_0$. A hipótese H_0 é rejeitada sse $T > c_\alpha$, sendo este um teste de nível α, com $P(T > c_\alpha) = \alpha$ sob H_0
- Considere-se $H_0 : \theta \geq \theta_0$ vs. $H_1 :: \theta < \theta_0$. A hipótese H_0 é rejeitada sse $T < c_\alpha$, sendo este um teste de nível α, com $P(T < c_\alpha) = \alpha$ sob H_0
- Considere-se $H_0 : \theta = \theta_0$ vs. $H_1 :: \theta \neq \theta_0$. A hipótese H_0 é rejeitada sse $|T| > c_{\alpha/2}$, sendo este um teste de nível α, com $P(|T| > c_{\alpha/2}) = \alpha/2$ sob H_0

Exemplo

Nma amostra de 16 indivíduos com glaucoma de ângulo aberto registaram-se as idades dos pacientes:

62 62 68 48 51 60 51 57 57 41 62 50 53 34 62 61

a) Podemos concluir que a idade média da população a partir da qual a amostra é 60 anos? ($\alpha = 0.01$).
b) Calcule o IC da idade média desta população a 95%.

Código R:

```
> idade<-c(62, 62, 68, 48, 51, 60, 51, 57, 57, 41, 62, 50, 53, 34, 62, 61)
> t.test(x = idade,mu = 60)

        One Sample t-test

data:  idade
t = -2.2822, df = 15, p-value = 0.03749
alternative hypothesis: true mean is not equal to 60
95 percent confidence interval:
 50.20944 59.66556
sample estimates:
mean of x
  54.9375
```

Valor-p ou p-value

- Como vimos o nível de significância é uma característica fundamental de um teste, porque representa a probabilidade máxima de tomar a decisão errada de rejeitar a hipótese nula quando ela é verdadeira.

- Naturalmente, o nível de significância α deve ser pequeno, em geral $\alpha = 0.01, 0.05, 0.10$

- Contudo, a forma de determinar o valor para o nível de significância é um tanto ou quanto arbitrária...

- Para obstar a esta indefinição relativamente ao valor a fixar para o nível de significância há uma outra forma de apresentar o resultado de um teste: o **valor** - p

- Assim, com $p \in [0, 1]$, sendo $T(\mathbf{x})$ uma estatística apropriada ao teste (distribuição amostral conhecida)

  Teste unilateral $p = P[T(\mathbf{x}) > |t(\boldsymbol{x})|]$
  Teste bilateral $p = 2 \times P[T(\mathbf{x}) > |t(\boldsymbol{x})|]$

- Assim, para pequenos valores de p a hipótese nula deve ser rejeitada (i.e., em regra, para $p \leq \alpha$)

Pegando no exemplo anterior, admitindo $x \frown N(\mu, \sigma)$ (σ desconhecido):

Hipóteses: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

Sob H_0,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \frown t_{n-1}$$

Para a amostra concreta obteve-se $t_{obs} = -2.2822$.

Assim, sob H_0, o valor-p será dado por:

$$p = 2 \times P(t > 2.2822), \quad com\ t \frown t_{15}$$

Assim, $p = 2 \times 0.0187 = 0.0374$