Probabilidades e Estatística

OBJECTIVO - *Inferir* sobre a população com base em informação amostral:

- extrair de forma eficiente a informação da amostra referente à população em estudo;

$\Rightarrow$ *Inferência*;

- fornecer medidas que avaliam a qualidade das inferências.

Inclui teoria e técnicas usadas no dia-a-dia:

- análise descritiva de dados,
- inquéritos (estudos de mercado/opinião, eleições),
- inferências para avaliar e comparar a eficácia de procedimentos (ensaios experimentais e tratamento clínicos),
- optimização (na qualidade do fabrico de produtos, consumo),
- previsão (ambiente e economia),
- decisão (planeamento, processos jurídicos).

1 Análise Exploratória de Dados

É essencial estarem bem definidas: a **população** (real ou conceptual) e/ou **variável(eis)** em estudo, bem como o **objectivo** do trabalho.

Classificação das variáveis

- qualitativa nominal ou categórica (não ordenável)
- qualitativa ordinal
- quantitativa discreta
- quantitativa contínua ($\subset \mathbb{R}$)

Com base numa ***amostra*** - conjunto de dados $(x_1, \ldots, x_n)$ - ou, no caso bivariado ainda com uma segunda amostra $(y_1, \ldots, y_n)$, pretende obter-se:

Medidas descritivas numéricas

- localização: *média* ($\bar{x} = \sum_{i=1}^{n} x_i/n$), *mediana* (valor central da amostra ordenada), *quantil* (q_α, $\alpha \in (0,1)$, com quartis $Q_i = q_{i/4}$ $i = 1, 2, 3$)

$$\text{Tipo 1: } q_\alpha = x[\lceil n\alpha \rceil], \qquad \text{Tipo 2: } q_\alpha = \begin{cases} \frac{x[n\alpha]+x[n\alpha+1]}{2} & , n\alpha \in \mathbb{N} \\ x[\lceil n\alpha \rceil] & , n\alpha \notin \mathbb{N} \end{cases}$$

sendo, $\lfloor a \rfloor \leq a \leq \lceil a \rceil$ para qualquer $a \in \mathbb{R}$ e $x[1] \leq \cdots \leq x[n]$ a amostra ordenada; *"outliers"* (moderados com barreiras $Q_i \pm 1.5 \times IQR$, $i = 1, 3$)

- dispersão: *amplitudes amostral* e *interquartil* ($IQR = q_{0.75} - q_{0.25}$), *variância* ($s^2$)

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1} = \sum_{i=1}^{n} \frac{x_i^2}{n-1} - \frac{n}{n-1}\bar{x}^2,$$

  desvio padrão (s) e coeficiente de variação ($s/\bar{x}$)

- bidimensional: *covariância* e *correlação*, respectivamente,

$$cov(x,y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \sum_{i=1}^{n} \frac{x_i y_i}{n-1} - \frac{n}{n-1}\bar{x}\bar{y} \quad \text{e} \quad cor(x,y) = \frac{cov(x,y)}{s_x s_y}$$

(Parâmetros populacionais: média μ, variância σ^2, correlação ρ - geralmente desconhecidos e estimam-se usando as quantidades anteriores)

Descrição gráfica

- histograma, diagrama de barras e queijo
- diagrama de caule-e-folhas
- caixa-de-bigodes
- diagrama de dispersão

Critérios a seguir: existem uma série de regras básicas estabelecidas mas por vezes diferentes, nomeadamente dependendo do *software* estatístico. Por exemplo na construção das classes do histograma, consideram-se intervalos disjuntos e exaustivos, entre 5 a 20, e a altura de cada retângulo do histograma é proporcional à fração da frequência da classe (no total das medições ou observações).

TPC: Instalar o *software estatístico R*; Exercs 1.,2. e 3.
Nota. Existe muita informação sobre o R na web, menciono um documento bastante resumido de muitos à escolha: *Holland (2020) A short R tutorial*, e o livro gratuito também com a matéria de PE, Pinheiro, Cunha, Carvajal e Gomes (2009, Elsevier) Estatística Básica, A arte de trabalhar com dados. Sobre o ggplot ficam as sugestões: *ggplot2 Tutorial (2017)* de Gert Janssenswillen, *Data Visualization with ggplot2 (2015)* de Jihui Lee e, *R Charts by R Coder*.

2 Probabilidade

Definição 2.1 (**Experiência Aleatória**). *Processo segundo o qual obtêm-se a observação imprevisível, ou resultado aleatório.*

Definição 2.2 (**Espaço de Resultados Ω**). *Conjunto de todos os resultados possíveis.*

É importante distinguir se Ω é: discreto (finito, infinito numerável e.g. $\mathbb{N}$) ou contínuo (infinito não numerável e.g. $\mathbb{R}$).

Definição 2.3 (**Acontecimento ou Evento $A \subset \Omega$**). *Qualquer subconjunto de Ω.*

Casos particulares:

- *acontecimento elementar* - formado por um único elemento de Ω;

- *acontecimentos mutuamente exclusivos, incompatíveis ou disjuntos* - não têm elementos comuns i.e. a sua interseção é o conjunto vazio.

Seja $A \in \mathcal{A}$ qualquer acontecimento mensurável de Ω (i.e. pertencente a uma σ-álgebra, classe adequada de conjuntos), mas abreviaremos o formalismo daqui em diante para $A \subset \Omega$.

Definição 2.4 (Probabilidade, Axiomática de Kolmogorov). $P : \mathcal{A} \rightarrow [0, 1]$ *é uma função de conjuntos:*

1. $P(\Omega) = 1$,
2. $P(A) \geq 0$, $\forall A \subset \Omega$,
3. $A_1, A_2, \ldots \subset \Omega$ *disjuntos dois a dois,*

$$P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Nota: A axiomática comporta a definição de Laplace ((nº casos favoráveis)/(nº casos possíveis) com Ω com todos os resultados equiprováveis) e a interpretação frequêncista ($\lim_{n \to \infty}$ (nº observações favoráveis)/(nº provas) com n provas "iguais").

TPC: Rever propriedades dos conjuntos e suas operações e.g. em *Conjuntos (St.Aubyn, Figueiredo, Loura, Ribeiro e Viegas 2004)*.

Propriedades

1. $P(\emptyset) = 0$
2. $P(\overline{A}) = 1 - P(A)$
3. $B \subset A \implies P(B) \leq P(A)$
4. $P(A \backslash B) = P(A \cap \overline{B}) = P(A) - P(A \cap B)$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
6. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap c)$

TPC: Provar as propriedades.

Definição 2.5 (Probabilidade condicionada). *Sejam* $A, B \subset \Omega$ *e* $P(B) > 0$,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

NOTA: Ao condicionar estabelece-se um novo espaço de resultados, no qual todas as propriedades anteriores são válidas.

Teorema 2.1 (Lei da Probabilidade Composta)**.** *Sejam* $A_1, A_2, \ldots, A_n \subset \Omega$ *e* $P(A_1 \cap A_2 \cap \ldots \cap A_{n-1}) > 0$, *tem-se*

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

Definição 2.6 (**Acontecimentos independentes**)**.**

- $A, B \subset \Omega$ *dizem-se independentes sse*

$$P(A \cap B) = P(A) \times P(B).$$

- $A_1, A_2, \ldots, A_n \subset \Omega$ *dizem-se independentes sse*

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \times \ldots \times P(A_n), \ldots, P(A_i \cap A_j) = P(A_i) \times P(A_j) \ \forall i \neq j.$$

Propriedades Se $A, B \subset \Omega$ independentes, então

1. $\bar{A}, B$ são independentes,
2. $A, \bar{B}$ são independentes,
3. $\bar{A}, \bar{B}$ são independentes.

TPC: Provar as propriedades.

EXERCÍCIO.

1. Considere A e B tais que: $P(A) = 1/4$, $P(B) = 1/3$, $P(A \cap B) = 0$.
   (a) Poderão ser A e B incompatíveis?
   (b) São A e B independentes?

Definição 2.7 (**Partição de Ω**)**.** *Coleção de subconjuntos de Ω, $A_1, A_2, \ldots, A_n$ tais que $A_i \cap A_j = \emptyset \ \forall i \neq j$, e $\cup_{i=1}^{n} A_i = \Omega$.*

Teorema 2.2 (Lei da Probabilidade Total)**.** *Sendo $B \subset \Omega$ e $\{A_i\}_{i=1}^{n}$ uma partição de Ω cada com probabilidade não nula,*

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

Teorema 2.3 (Teorema de Bayes)**.** *Sendo $B \subset \Omega$, $\{A_i\}_{i=1}^{n}$ uma partição de Ω, todos com probabilidade não nula,*

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}, \qquad i = 1, \ldots n.$$

3 Variáveis aleatórias discretas

Abreviadamente, uma variável aleatória (v.a.) é uma função real a representar o resultado de uma experiência aleatória, $X : \Omega \to \Omega_X \subset \mathbb{R}$ (formalmente uma v.a. é uma função real $X : \Omega \to \mathbb{R}$ tal que, $X^{\leftarrow}(B) \in \mathcal{A}$, para todo o conjunto Boreliano de $\mathbb{R}$).

A variável diz-se discreta se Ω_X for finito ou infinito numerável com:

Definição 3.1 (Função de probabilidade). *$P(X = x)$, $x \in \mathbb{R}$:*

$$P(X = x) \geq 0,\ \forall x \in \mathbb{R}, \qquad e \qquad \sum_{x \in \Omega_X} P(X = x) = 1.$$

Definição 3.2 (Parâmetros e medidas de localização e dispersão).

1. *Valor médio ou esperança matemática:*

$$\mu = E[X] = \sum_{x \in \Omega_X} xP(X = x), \qquad E[\phi(X)] = \sum_{x \in \Omega_X} \phi(x)P(X = x)$$

2. *Quantil de ordem p, χ_p: $P(X \leq \chi_p) \geq p$ e $P(X \geq \chi_p) \geq 1 - p$*

3. *Moda, $m_0 : P(X = m_0) = \max_{x \in \mathbb{R}} P(X = x)$*

4. *Variância:*

$$\sigma^2 = var(X) = E[(X - \mu)^2] = \sum_{x \in \Omega_X} (x - \mu)^2 P(X = x) = E[X^2] - (E[X])^2$$

5. *Desvio padrão: $\sigma = \sqrt{\sigma^2}$*

6. *Coeficiente de variação (adimensional): $\sigma/|\mu|$*

Propriedade 3.1. *[Propriedades do valor médio e variância]*
Sendo a, b constantes reais e X uma variável aleatória,

1. $E[a] = a$

2. $E[aX + b] = aE[X] + b$

3. $var(X) \geq 0$

4. $var(a) = 0$

5. $var[aX + b] = a^2 var(X)$

TPC: Provar as propriedades.

Definição 3.3 (V.a. Uniforme discreta $\{1, \ldots, n\}$). *$X \sim Unif\{1, \ldots, n\}$, $\Omega_X = \{1, \ldots, n\}$, reflete uma população com n elementos, todos igualmente prováveis:*

$$P(X = x) = 1/n, \quad x = 1, \ldots, n, \qquad E[X] = \frac{n+1}{2},\ Var(X) = \frac{n^2 - 1}{12}.$$

EXERCÍCIOS:

1. Para as v.a.'s discretas $Unifome(1, 3)$ e $Uniforme(1, 4)$:
   a) Explicite a represente graficamente as funções de probabilidade.
   b) Calcule as medidas de localização e dispersão.

2. Calcular as medidas de localização e dispersão para as v.a.'s discretas:
   a) $X : P(X = x) = 1/10, 2/10, 7/10,\ x = 0, 1, 2,$
   b) $Y : P(Y = y) = 7/10, 2/10, 1/10,\ y = 0, 1, 2.$

Definição 3.4 (V.a. Bernoulli). *$X \sim Ber(p)$, $\Omega_X = \{0, 1\}$:*

$$P(X = x) = p^x(1 - p)^{1-x} \qquad x = 0, 1, \qquad E[X] = p,\ Var(X) = p(1 - p).$$

Definição 3.5 (V.a. Binomial (n, p)). *$X \sim Bin(n, p)$, $\Omega_X = \{0, 1, \ldots, n\}$, representa o número de sucessos em n repetições independentes da prova de Bernoulli (p):*

$$P(X = x) = C^n_x p^x(1 - p)^{n-x},\ x = 0, 1, \ldots, n;\ E[X] = np,\ Var(X) = np(1 - p).$$

Propriedade 3.2. *$X \sim Bin(n, p)$ se e só se $n - X \sim Bin(n, 1 - p)$.*

EXERCÍCIOS 3.1(c), 3.2.

Definição 3.6 (V.a. Geométrica). *$X \sim Geom(p)$, $\Omega_X = \mathbb{N}_1$, representa o número de repetições independentes da prova de Bernoulli (p) até obter o primeiro sucesso:*

$$P(X = x) = p(1 - p)^{x-1} \qquad x = 1, 2, \ldots;\ E[X] = \frac{1}{p},\ Var(X) = \frac{1 - p}{p^2}.$$

TPC: Exercícios de testes sobre Binomial e Geométrica:
Semestre 1, 2017/2018 - 18/11/2017 - 9:00 Grupo 1
Semestre 2, 2018/2019 - 4/5/2019 - 9:00 Grupo 1

Definição 3.7 (V.a. Poisson). *$X \sim Poisson(\lambda)$, $\Omega_X = \mathbb{N}_0$, representa frequentemente contagens:*

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!},\ x = 0, 1, 2, \ldots, \qquad E[X] = Var(X) = \lambda.$$

TPC: Exercícios de testes com a Poisson:
Semestre 2, 2016/2017 - 06/05/2017 - 11:00 Grupo 1,
Semestre 1, 2012/2013 - 01/02/2013 - 11:30 GI1.3

Definição 3.8 (Função de distribuição cumulativa). *$F_X(x)$, $x \in \mathbb{R}$:*

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i),\ \forall x \in \mathbb{R}.$$

EXERCÍCIO: Calcular a f.d. para a v.a. $Unifome(1, 3)$.

Propriedade 3.3. *Propriedades da função de distribuição cumulativa $F_X(x) = P(X \leq x)$:*

1. $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$;
2. *É uma função monótona não decrescente: se $x_1 > x_2$, $F(x_1) \geq F(x_2)$;*
3. *É uma função descontínua (em escada) sendo os pontos de descontinuidade os resultados possíveis da experiência aleatória; é contínua à direita, $F(x^+) = \lim_{t\downarrow x} F(t) = F(x)$;*
4. $P(X = x) = F_X(x) - F_X(x^-)$, *sendo* $F_X(x^-) = \lim_{t\uparrow x} F(t) = P(X < x)$;
5. $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$.

4 V.a.s contínuas

Designaremos uma variável aleatória (v.a.) contínua X se tem *função de distribuição* (f.d. cumulativa) $F_X(x) = P(X \leq x)$ contínua $\forall x \in \mathbb{R}$ (de forma mais rigorosa necessitaremos que F seja diferenciável a menos de um número finito de pontos em qualquer intervalo finito).

Definição 4.1 (Função densidade de probabilidade). *Denota-se por $f_X(x)$, $x \in \mathbb{R}$:*

$$f_X(x) \geq 0,\ \forall x \in \mathbb{R}, \qquad e \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1.$$

Propriedade 4.1. *Relações entre as funções de densidade e distribuição:*

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du, \quad e \quad f_X(x) = \frac{dF_X(x)}{dx}; \qquad P(X = x) = 0.$$

Propriedade 4.2. *$F_X(x)$ verifica (além das propriedades especificadas na Propriedade 3.3), sendo uma função contínua i.e. $F_X(x^-) = F_X(x) = F_X(x^+)$,*

$$\begin{aligned} P(a < X \leq b) &= P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) \\ &= \int_a^b f_X(x)\,dx = F_X(b) - F_X(a). \end{aligned}$$

Definição 4.2 (Parâmetros e medidas de localização e dispersão).

1. *Valor médio ou esperança matemática:*

$$\mu = E[X] = \int_{-\infty}^{\infty} x\, f_X(x)\,dx, \qquad E[\phi(X)] = \int_{-\infty}^{\infty} \phi(x)\, f_X(x)\,dx$$

2. *Quantil de ordem p, χ_p: $F_X(\chi_p) = p$*
3. *Moda, $m_0 : f_X(m_0) = \max_{x\in\mathbb{R}} f_X(x)$*
4. *Variância:*

$$\sigma^2 = var(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2\, f_X(x)\,dx = E[X^2] - (E[X])^2$$

5. *Desvio padrão:* $\sigma = \sqrt{\sigma^2}$

6. *Coeficiente de variação (adimensional):* $\sigma/|\mu|$

NOTA: A Propriedade 3.1 aplica-se igualmente no caso das v.a.'s contínuas.

Definição 4.3 (V.a. Uniforme contínua).
$X \sim Unif[a,b]$: $\Omega_X = [a,b]$ com os resultados equiprováveis:

$$f_X(x) = \frac{1}{b-a},\ a \leq x \leq b; \quad F_X(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x < b, \\ 1, & x \geq b. \end{cases} \quad E[X] = \frac{b+a}{2},\ Var(X) = \frac{(b-a)^2}{12}.$$

EXERCÍCIOS: 1. Considere uma variável aleatória a tomar valores no intervalo $[0,1]$ todos igualmente prováveis.

a) Indique o espaço de resultados associado e a função densidade de probabilidade e, represente-a graficamente.

b) Determine a função de distribuição de X e represente-a graficamente.

c) Calcule a mediana, o valor esperado e a variância de X.

d) Repita as alíneas anteriores considerando os intervalos $[0,1/2]$ e $[0,2]$.

2. Semestre 2, 2021/2022, 06/07/2022, Exame 2 Pergunta 3.

Definição 4.4 (V.a. Exponencial). *$X \sim Exp(\lambda)$: $\Omega_X = \mathbb{R}_+$ comum na modelação do "tempo de vida" de componentes eletrónicas e "tempo entre chegadas":*

$$f_X(x) = \lambda e^{-\lambda x},\ x \geq 0; \quad F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0, \end{cases} \quad E[X] = \frac{1}{\lambda}, \quad var(X) = \frac{1}{\lambda^2}.$$

Propriedade 4.3. *Falta de memória da v.a. exponencial:*

$$P(X > t + x | X > t) = P(X > x), \quad \forall x, t > 0.$$

EXERCÍCIO: 4.3

TPC: Semestre 1, 2016/2017, 19/11/2016, 9h Grupo I.2.

Definição 4.5 (V.a. Gaussiana ou Normal). *$X \sim N(\mu, \sigma^2)$: $\Omega_X = \mathbb{R}$ tem f.d.p. simétrica, com grande impacto muito devido ao Teorema do Limite Central (TLC):*

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; x \in \mathbb{R}, \qquad E[X] = \mu \quad var(X) = \sigma^2, \\ F_X(x) &= P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt. \end{aligned}$$

Propriedades a salientar:

Propriedade 4.4. *Fecho para transformações lineares:*

$$X \sim N(\mu, \sigma^2), a, b \in \mathbb{R} \quad \Longrightarrow \quad Y = aX + b \sim N(a\mu + b, a^2\sigma^2).$$

consequentemente,

$$X \sim N(\mu, \sigma^2) \quad \Longrightarrow \quad Z = \frac{X - \mu}{\sigma} \sim N(0, 1); \quad \Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Propriedade 4.5. *Cálculo da probabilidade com recurso à normal reduzida:*

$$P(a < X \leq b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right),$$

$$\Phi(-z) = 1 - \Phi(z), \qquad \forall z \in \mathbb{R}.$$

EXERCÍCIOS: 4.9 e Semestre 2, 2017/2018, 5/5/2018, 11h) - Grupo II.1.a)b) mas também podem fazer todo o Grupo I para revisão.

5 Distribuições conjuntas de probabilidade e complementos

Seja (X, Y) um par aleatório com espaço de resultados $\Omega_{X,Y} (\subset \mathbb{R}^2)$.

5.1 Pares aleatórios discretos

$\Omega_{X,Y}$ é um conjunto de valores finito ou infinito numerável.

Definição 5.1 (Funções conjuntas de probabilidade e distribuição cumulativa).

$$P(X = x, Y = y) = \begin{cases} > 0, & (x, y) \in \Omega_{X,Y} \\ 0, & c.c., \end{cases}, \qquad \sum_{(x,y) \in \Omega_{X,Y}} P(X = x, Y = y) = 1;$$

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \sum_{x_i \leq x, y_i \leq y} P(X = x_i, Y = y_i)$$

Definição 5.2 (Funções de probabilidade marginais).

$$P(X = x) = \sum_{\Omega_Y} P(X = x, Y = y), \quad P(X \leq x) = \sum_{x_i \leq x} \sum_{\Omega_Y} P(X = x_i, Y = y), \quad x \in R,$$

Definição 5.3 (Função de probabilidade condicional). *F.p. de X condicional a $Y = y$,*

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \quad x \in \mathbb{R}, \; y : P(Y = y) > 0.$$

Definição 5.4 (V.a.'s independentes). *Duas v.a.'s X e Y são independentes se*

$$P(X = x, Y = y) = P(X = x) \times P(Y = y), \quad \forall\, (x, y) \in \mathbb{R}^2.$$

5.2 Momentos e Correlação

Definição 5.5 (Valor esperado, Momentos marginais e condicionados). *Seja ψ uma função real,*

$$E[\psi(X,Y)] = \sum_{\Omega_{X,Y}} \psi(x,y)P(X=x,Y=y);$$

$$\begin{aligned} E[\psi(X)] &= \sum_{\Omega_X} \psi(x)P(X=x), \qquad \mu_X = E[X] = \sum_{\Omega_X} xP(X=x), \\ \sigma^2 &= Var(X) = E[(X-\mu_X)^2] = E[X^2] - \mu_X^2; \end{aligned}$$

$$E[X|Y=y] = \sum_{\Omega_X} xP(X=x|Y=y), \qquad E[\psi(X)|Y=y] = \sum_{\Omega_X} \psi(x)P(X=x|Y=y).$$

Muitas propriedades são naturalmente extensíveis do caso univariado, como por exemplo a propriedade da linearidade $E[X+Y] = E[X] + E[Y]$.

Definição 5.6 (Covariância). *Mede a associação linear entre X e Y, permitindo obter informação sobre a dependência entre as variáveis aleatórias,*

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X\mu_Y,$$

$Cov > 0$ indica uma associação positiva entre as variáveis, enquanto que $Cov < 0$ indica uma associação negativa entre as variáveis.

Definição 5.7 (Correlação). *Medida adimensional que mede a associação linear entre X e Y,*

$$-1 \leq Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \leq 1$$

Propriedade 5.1. *Propriedades da variância, covariância e correlação:*

1. $Var(aX+bY) = a^2var(X) + b^2Var(Y) + 2ab\,Cov(X,Y)$
2. *Se X, Y independentes:* $Cov(X,Y) = E[XY] - \mu_X\mu_Y = 0$ *e,*
   $Var(X+Y) = Var(X) + Var(Y)$
3. $Cov(X,Y) = Cov(Y,X)$
4. $Cov(X,X) = Var(X)$
5. $Cov(aX+b,Y) = a\,Cov(X,Y)$
6. $Cov(X+Z,Y) = Cov(X,Y) + Cov(Z,Y)$
7. *Se $Y = aX+b$: $Corr(X,Y) = -1$ se $a < 0$, e $Corr(X,Y) = 1$ se $a > 0$.*

EXERCÍCIO: Considere o par aleatório discreto (X,Y) com função de probabilidade conjunta dada por ($Y = X^2$):

$Y\backslash X$	-1	0	1
0	0	1/3	0
1	1/3	0	1/3

1. Diga, justificando, se X e Y são variáveis aleatórias independentes.
2. Calcule $Cov(X,Y)$ e comente o resultado.

EXERCÍCIO: 5.1

TPC: Semestre 2, 2018/2019, 04/05/2019, 9h Grupo II.2.
Semestre 2, 2017/2018, 05/05/2018, 11h Grupo II.2.

5.3 Pares aleatórios contínuos

No caso contínuo $\Omega_{X,Y}$ é um conjunto de valores infinito não numerável. As fórmulas anteriores adaptam-se, substituindo as somas por integrais e as funções de probabilidade por funções de densidade.

EXERCÍCIO: 5.13

TPC: Semestre 2, 2020/2021, 15/5/2021, 9h Pergunta 4.
Semestre 2, 2017/2018 - 05/05/2018 - 9:00 Grupo II.1

5.4 Combinações lineares de v.a.s normais independentes

$$\sum_{i=1}^{n} c_i\, X_i \ ,\ c_i \in \mathbb{R}, \quad \text{com casos particulares:} \quad \sum_{i=1}^{n} X_i \quad \text{e} \quad \bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}.$$

Teorema 5.1. *Uma combinação linear de v.a.'s normais é uma v.a. normal.*

Corolário 5.1. *Se $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, independentes, então*

$$\sum_{i=1}^{n} c_i\, X_i \sim N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right).$$

Consequentemente, se $X_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$, independentes e identicamente distribuídas:

$$\sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right) \qquad e \qquad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Exercício: 5.16.
TPC: Semestre 1, 2017/2018 - 18/11/2017 - 9:00 Grupo 2

5.5 Combinações lineares de v.a.s binomiais independentes

$$X_i \sim Bin(n_i, p), \text{ independentes } i = 1, \ldots, n \quad \Longrightarrow \quad \sum_{i=1}^{n} X_i \sim Bin\left(\sum_{i=1}^{n} n_i, p\right)$$

Exercícios: Teste - Semestre 1, 2019/2020, 16/11/2019, 9h, GI.2.

5.6 Combinações lineares de v.a.s Poisson independentes

$$X_i \sim Pois(\lambda_i), \text{ independentes } i = 1, \ldots, n \quad \Longrightarrow \quad \sum_{i=1}^{n} X_i \sim Pois\left(\sum_{i=1}^{n} \lambda_i\right)$$

Exercício P5 (mas recomendo fazerem tudo como revisão) do teste:
Semestre 2, 2020/2021, 09/07/2021, 11h30

5.7 TLC (variáveis aleatórias independentes e identicamente distribuídas)

Teorema 5.2 (TEOREMA do LIMITE CENTRAL). *Sendo X_i v.a.'s independentes, $E[X_i] = \mu$, $var(X_i) = \sigma^2 < \infty$, $i = 1, \ldots, n$, então*

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leq z\right) = \lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i/n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \phi(z).$$

Consequentemente, tem-se para a distribuição limite da soma e média,

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \sim^{aprox} N(0,1) \quad \Leftrightarrow \quad \sum_{i=1}^{n} X_i \sim^{aprox} N\left(n\mu, n\sigma^2\right)$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim^{aprox} N(0,1) \quad \Leftrightarrow \quad \bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} \sim^{aprox} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Visualização do TLC através de simulações no R: ver páginas 158-160 em
Pinheiro, Cunha, Carvajal e Gomes (2009, Elsevier) Estatística Básica, A ARTE DE TRABALHAR COM DADOS

Exercícios de testes: Semestre 2, 2017/2018, 05/05/2018, 9h Grupo 2
Semestre 2, 2016/2017 - 06/05/2017 - 11:00 Grupo 2
Semestre 2, 2018/2019 - 4/5/2019 - 9:00 Grupo 2

Propriedade 5.2. *Processo de Poisson: seja $X_t \sim Poisson(\lambda t)$, para cada t, uma variável aleatória a representar o número de ocorrências no intervalo de tempo $(0, t]$, $t > 0$, i.e.*

$$X_t \sim Poisson(\lambda t) \qquad E[X_t] = Var(X_t) = \lambda t, \qquad t > 0.$$

Verifica-se,

a) O número de ocorrências em intervalos de tempo disjuntos são independentes;

b) O tempo entre ocorrências sucessivas (seja Y) tem distribuição exponêncial de parâmetro λ, i.e. $E[Y] = 1/\lambda$.

EXERCÍCIOS: 3.11, 4.11

TPC: Semestre 2, 2017/2018 - 05/05/2018 - 9:00 Grupo 1
Semestre 2, 2017/2018 - 23/07/2018 - 9:00 Grupo 1

6 Estimação Pontual

De uma forma geral, interessa estudar uma determinada população representada por uma variável aleatória (v.a.) X que segue um modelo probabilístico conhecido, por exemplo,

$$Bin(n,p), \quad Poisson(\lambda), \quad N(\mu,\sigma), \quad Exponencial(\lambda) \quad \ldots$$

Alguns dos parâmetros populacionais mais comuns são:

$\mu = E(X)$ média populacional

$\sigma^2 = Var(X)$ - variância populacional

$p = P(sucesso)$ - probabilidade de sucesso em populações Bernoulli ou Binomial

$\lambda = E(X)$ ou $\lambda = 1/E(X)$ numa população Poisson ou Exponencial respectivamente.

Definição 6.1 (Amostra Aleatória). *Uma amostra aleatória de dimensão n proveniente de uma população X é um conjunto de n variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), representada por:*

$$(X_1, X_2, \ldots, X_n).$$

No que se segue θ representa um parâmetro desconhecido da população X.

Definição 6.2 (Estimador). *Um estimador $\hat{\theta}_n$ é uma função exclusiva da amostra aleatória e portanto independente de parâmetros desconhecidos. Note-se que é uma variável aleatória.*

Duas questões essenciais: **(6.1)** qualidade e **(6.2)** determinação de estimadores.

6.1 Qualidade dos estimadores

Duas propriedades desejáveis são:

a) ***Centralidade***: $E[\hat{\theta}_n] = \theta$ e,

b) ***Variância mínima***: $\min Var(\hat{\theta}_n)$.

Dois estimadores a salientar:

Estimador Média Amostral

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

É um estimador centrado para a média populacional μ. Quando concretizada a amostra aleatória - representada por $(x_1, \ldots, x_n)$ - obtém-se a *estimativa* $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$, i.e. um valor concreto para μ que quase certamente não será o verdadeiro valor de μ mas que deverá aproximar-se deste pelas boas propriedades do estimador (em particular prova-se que tende para μ quando $n \to \infty$).

Estimador Variância Amostral (também designado variância amostral corrigida)

$$S_{n-1}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2}{n-1} - \frac{n}{n-1}\bar{X}^2.$$

É um estimador comum para a variância populacional σ^2. A divisão por $(n-1)$ garante a centralidade.

Definição 6.3 (Erro Quadrático Médio (EQM)). *O EQM mede em simultâneo a variância e o quadrado do enviesamento (i.e. quanto, em média, o estimador se desvia do parâmetro a estimar):*

$$EQM(\hat{\theta}) = E[(\hat{\theta}-\theta)^2] = Var(\hat{\theta}) + (E[\hat{\theta}]-\theta)^2.$$

Definição 6.4 (Eficiência de um estimador $\hat{\theta}_1$ relativamente a outro estimador $\hat{\theta}_2$). *A eficiência compara os estimadores pelos seus EQM através do quociente,*

$$\frac{EQM(\hat{\theta}_2)}{EQM(\hat{\theta}_1)},$$

se maior que 1, $\hat{\theta}_1$ diz-se mais eficiente; se menor que 1, $\hat{\theta}_2$ diz-se mais eficiente.

6.2 Método da Máxima Verosimilhança

O método consiste em maximizar a função de verosimilhança L, que se traduz na função de probabilidade ou de densidade da amostra aleatória,

$$L(\theta|x_1,\dots,x_n) = \begin{cases} P(X_1=x_1,\dots,X_n=x_n|\theta), & \text{se } X \text{ é uma v.a. discreta} \\ f_{X_1,\dots,X_n}(x_1,\dots,x_n|\theta), & \text{se } X \text{ é uma v.a. contínua.} \end{cases}$$

Sob certas condições de regularidade da função de probabilidade ou de densidade, o estimador da máxima verosimilhança $\hat{\theta}_{MV}$ obtém-se da resolução de (por exemplo, no caso discreto):

$$\frac{d}{d\theta}\Big(\log L(\theta|x_1,\dots,x_n)\Big) = \frac{d}{d\theta}\Big(\log\prod_{i=1}^n P_\theta(X_i=x_i)\Big) = \frac{d}{d\theta}\left(\sum_{i=1}^n \log P_\theta(X_i=x_i)\right) = 0$$

$$\frac{d^2}{d\theta^2}\left(\log\prod_{i=1}^n P_\theta(X_i=x_i)\right)\Bigg|_{\theta=\hat{\theta}_{MV}} < 0.$$

Propriedade 6.1 (Propriedade da Invariância).

$$\widehat{h(\theta)}_{MV} = h\left(\hat{\theta}_{MV}\right)$$

para e.g. h função injectiva.

Exercícios: 6.2
Exercícios de testes: Semestre 1, 2021/2022, 12/02/2022, 10h30 Pergunta 6.
Semestre 2, 2018/2019 - 12/06/2019 - 9h Grupo I

7 Intervalos de confiança (IC)

O método da variável fulcral permite obter, com base numa amostra aleatória $\underline{X} = (X_1, X_2, \dots, X_n)$, um IC aleatório, correspondente a uma probabilidade $1-\alpha$ (usualmente 0.95, 0.99 e 0.9), para um parâmetro desconhecido θ (e.g. μ, σ^2, p):

$$[I(\underline{X}), S(\underline{X})] \qquad \text{tal que} \qquad P\left(I(\underline{X}) \le \theta \le S(\underline{X})\right) = 1-\alpha.$$

1. Identificar a variável (aleatória!) fulcral, *função do estimador $T(\underline{X})$ e do parâmetro desconhecido θ, com distribuição conhecida*:

$$V(T(\underline{X}), \theta);$$

2. escolher α (valores mais comuns: 0.05, 0.01 ou 0.1) tal que $(1-\alpha) \times 100\%$ corresponderá ao nível de confiança do IC (respectivamente, 95%, 99% e 90%);

3. obter os quantis de probabilidade $\alpha/2$ e $1-\alpha/2$,

$$\begin{cases} \chi_{\alpha/2} \ : \ P(V < \chi_{\alpha/2}) = \alpha/2 \\ \chi_{1-\alpha/2} \ : \ P(V > \chi_{1-\alpha/2}) = \alpha/2; \end{cases}$$

4. obter os limites inferior e superior do IC aleatório, $I(\underline{X})$ e $S(\underline{X})$, resolvendo a dupla desigualdade,

$$P\left(\chi_{\alpha/2} \le V(T(\underline{X}), \theta) \le \chi_{1-\alpha/2}\right) = P\left(I(\underline{X}) \le \theta \le S(\underline{X})\right) = 1-\alpha;$$

5. concretizar o IC com base na amostra dada, obtendo-se $[i(\underline{x}), s(\underline{x})]$.

Dois exemplos:

1. **IC para μ com σ^2 conhecido a 95%**
Seja X com distribuição normal de média μ desconhecida mas variância σ^2 conhecida. Pela variável fulcral,

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1), \tag{1}$$

conseguem-se obter os quantis $F^{-1}_{N(0,1)}(0.025) = \Phi^{-1}(0.025) = -1.96$ e $F^{-1}_{N(0,1)}(0.975) = \Phi^{-1}(0.975) = 1.96$ tais que,

$$P\left(-1.96 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le 1.96\right) = P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95.$$

Isto é, pela resolução da dupla desigualdade, obtém-se o IC aleatório a 95% para μ,

$$\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

No caso de população arbitrária com n elevado, o argumento anterior repete-se 'aproximadamente' justificado pelo Teorema do Limite Central.

2. **IC para μ com σ^2 desconhecido e, IC para σ^2, ambos a 95%**
Seja X com distribuição normal de média μ e variância σ^2 ambos desconhecidos. Prova-se que

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)} \qquad \text{e} \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)},$$

sendo $t_{(n-1)}$ a distribuição $t-$Student com $n-1$ graus de liberdade e, $\chi^2_{(n-1)}$ a distribuição Qui-quadrado com $n-1$ graus de liberdade. Escolhendo estas variáveis para variáveis fulcrais, os intervalos de confiança para os parâmetros μ e σ^2 obtêm-se de forma semelhante ao caso anterior:

$$P\left(\chi_{\alpha/2} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq \chi_{1-\alpha/2}\right) = 0.95 \quad \text{ou} \quad P\left(\chi_{\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_{1-\alpha/2}\right) = 0.95.$$

obtendo-se, pela resolução da dupla desigualdade, os ICs aleatórios a 95% para μ e σ^2 respectivamente,

$$\left[\bar{X} - F^{-1}_{t_{n-1}}(0.975)\frac{S}{\sqrt{n}}, \bar{X} + F^{-1}_{t_{n-1}}(0.975)\frac{S}{\sqrt{n}}\right] \quad \text{e} \quad \left[\frac{(n-1)S^2}{F^{-1}_{\chi_{n-1}}(0.975)}, \frac{(n-1)S^2}{F^{-1}_{\chi_{n-1}}(0.025)}\right].$$

No caso de população arbitrária com n elevado, o IC anterior para μ ainda é válido, aproximadamente pelo Teorema do Limite Central, substituindo-se $t_{n-1}(0.975)$ por $\Phi^{-1}(0.975)$.

Duas características importantes:
A **amplitude do IC** reflete a precisão do IC e portanto da estimação.
O **erro de estimação**, ou semi-amplitude do IC, também é uma medida usual para avaliar a precisão da estimação.