

## Formulário

### Variáveis Aleatórias (v.a.'a):

v.a.'s inteiras	v.a.'s contínuas
<b>Função massa de probabilidade (f.m.p.)</b> $f(x) = P(X = x)$ $\sum_x f(x) = 1$	<b>Função densidade de probabilidade (f.d.p.)</b> $f(x) \neq P(X = x)$ área total sob a curva = 1
<b>Função distribuição (f.d.)</b> $F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$ $P(a < X \leq b) = \sum_{x=a+1}^b f(x) = F(b) - F(a)$	<b>Função distribuição (f.d.)</b> $F(x) = P(X \leq x) = P(X < x) =$ área à esquerda de $x$ $P(a < X \leq b) = P(a < X < b) = P(a \leq X < b) =$ $= P(a \leq X \leq b) = F(b) - F(a)$

**Propriedades Gerais:** **1-** Se  $F(x_p) = p$ , então  $x_p$  é o quantil de ordem  $p$  ( $0 < p < 1$ ). **2-** Valor esperado ou média da população -  $\mu = E[X]$  **3-**  $E[aX + b] = aE[X] + b \forall a, b \in \mathbb{R}$ . **4-** Sejam  $X$  e  $Y$  v.a.'s quaisquer,  $E[X + Y] = E[X] + E[Y]$ . **5-** Variância da população -  $\sigma^2 = Var[X]$ . **6-**  $Var[aX + b] = a^2 Var[X] \forall a, b \in \mathbb{R}$ . **7-** Sejam  $X$  e  $Y$  duas v.a. independentes,  $Var[X + Y] = Var[X] + Var[Y]$ .

### Distribuições:

**Bernoulli,  $B(p)$ :**  $f(x) = p^x(1-p)^{1-x}$ ,  $x = 0, 1$ .  $E[X] = p$ ,  $Var[X] = p(1-p)$ .

**Binomial,  $B(n, p)$ :**  $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ ,  $x = 0, 1, 2, \dots, n$ .  $E[X] = np$ ,  $Var[X] = np(1-p)$ .

**Poisson,  $P(\lambda)$ :**  $f(x) = e^{-\lambda} \lambda^x / x!$ ,  $x = 0, 1, 2, \dots$ .  $E[X] = \lambda$ ,  $Var[X] = \lambda$ .

**Normal,  $N(\mu, \sigma)$ :**  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ ,  $x \in \mathbb{R}$ ,  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ .  $E[X] = \mu$ ,  $Var[X] = \sigma^2$ .

**Teorema do Limite Central:** Sejam  $X_1, X_2, \dots, X_n$  v.a.'s i.i.d. com média  $\mu$  e variância  $\sigma^2$ .

$$\frac{(\sum_{i=1}^n X_i) - n\mu}{\sigma\sqrt{n}} \xrightarrow{n \rightarrow \infty} Z \sim N(0, 1) \quad \Leftrightarrow \quad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{n \rightarrow \infty} Z \sim N(0, 1).$$

Em particular: Se  $X \sim B(n, p) \Rightarrow \frac{X - np}{\sqrt{np(1-p)}} \overset{\circ}{\sim} N(0, 1)$ ,  $n \rightarrow \infty$ ; Se  $X \sim P(\lambda) \Rightarrow \frac{X - \lambda}{\sqrt{\lambda}} \overset{\circ}{\sim} N(0, 1)$ ,  $\lambda \rightarrow \infty$ .

**Características amostrais:** Numa amostra aleatória  $(X_1, \dots, X_n)$

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

**Distribuições de amostragem:** Sejam  $(X_1, X_2, \dots, X_n)$  e  $(Y_1, Y_2, \dots, Y_m)$  duas amostras aleatórias de populações Normais com médias  $\mu_X$ ,  $\mu_Y$  e variâncias  $\sigma_X^2$ ,  $\sigma_Y^2$  respectivamente. Seja  $S_D$  o desvio padrão da amostra de diferenças  $X_i - Y_i$  quando as amostras estão emparelhadas, e defina-se  $S_p = \sqrt{\frac{\frac{1}{n} + \frac{1}{m}}{n+m-2} \sqrt{(n-1)S_X^2 + (m-1)S_Y^2}}$  quando as amostras são independentes.

- Estatísticas envolvendo a média amostral

$$\frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n}} \sim N(0, 1);$$

$$\text{IC para } \mu_X : \bar{X} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}};$$

$$\frac{\bar{X} - \mu_X}{S_{cX}/\sqrt{n}} \sim t_{n-1};$$

$$\text{IC para } \mu_X : \bar{X} \pm t_{1-\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}};$$

Se as amostras forem independentes

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p} \sim t_{n+m-2};$$

$$\text{IC para } \mu_X - \mu_Y : \bar{X} - \bar{Y} \pm t_{1-\frac{\alpha}{2}, n+m-2} S_p$$

Se as amostras forem emparelhadas

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_{cD}/\sqrt{n}} \sim t_{n-1}$$

$$\text{IC para } \mu_X - \mu_Y : \bar{X} - \bar{Y} \pm t_{1-\frac{\alpha}{2}, n-1} S_{cD}/\sqrt{n}.$$

Se a distribuição não for Normal, pelo TLC,

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

$$\text{IC para } \mu_X : \bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}};$$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \overset{\circ}{\sim} N(0, 1);$$

$$\text{IC para } \mu_X : \bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}};$$

- Estatísticas envolvendo a variância amostral corrigida

$$\frac{(n-1)S_X^2}{\sigma_X^2} \sim \chi_{(n-1)}^2; \quad \text{IC para } \sigma_X^2 : \left( \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}, n-1}^2}, \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}, n-1}^2} \right)$$

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n-1, m-1}; \quad \text{IC para } \frac{\sigma_X^2}{\sigma_Y^2} : \left( \frac{S_X^2}{S_Y^2} f_{\frac{\alpha}{2}, m-1, n-1}, \frac{S_X^2}{S_Y^2} f_{1-\frac{\alpha}{2}, m-1, n-1} \right)$$

$$\text{com } f_{1-\alpha, \nu, \omega} = \frac{1}{f_{\alpha, \omega, \nu}}$$

**Proporções:**  $\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \overset{\circ}{\sim} N(0, 1) \quad \hat{p} = \frac{X}{n}; \quad \text{IC para } p : = \hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$

**ANOVA:** Comparação de médias (ou medianas) de  $g$  grupos.

ANOVA paramétrica: Todos os grupos devem ter distribuição Normal com a mesma variância. Os grupos devem ter dimensão  $n$ .

$$SS_T = \sum_{i=1}^g \sum_{j=1}^n (Y_{ij} - \bar{Y}_{..})^2 \quad SS_G = n \sum_{i=1}^g (\bar{Y}_i - \bar{Y}_{..})^2 \quad SS_E = \sum_{i=1}^g \sum_{j=1}^n (Y_{ij} - \bar{Y}_i)^2 \quad SS_T = SS_G + SS_E$$

Fonte de Variação	Soma de quadrados	de graus de liberdade	Média dos quadrados	$F_0$	$p - value$
Entre Grupos	$SS_G$	$g - 1$	$MS_G = \frac{SS_G}{g-1}$	$\frac{MS_G}{MS_E}$	(·)
Dentro dos grupos	$SS_E$	$g(n - 1)$	$MS_E = \frac{SS_E}{g(n-1)}$		
Total	$SS_T$	$gn - 1$			

Efeitos aleatórios:  $\hat{\sigma}_\tau^2 = (MS_G - MS_E)/n$

ANOVA não-paramétrica: teste de Kruskal-Wallis.

**Comparações múltiplas (pares de médias)** Consideram-se todas as comparações de pares de médias envolvidas na ANOVA. Método de **Bonferroni** e de **Tukey** solucionam a problemática associada ao nível de significância do conjunto de comparações ( $m$ ). **Método de Bonferroni** - reduz o tamanho individual para que o tamanho total seja o desejado:  $\alpha = m\alpha_m$  onde  $\alpha_m$  é o tamanho de cada comparação individual.

**Regressão linear:** Modelo:  $Y_i = b_0 + b_1x + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma), \quad iid.$

Estimadores dos parâmetros e suas propriedades:  $T_i = \frac{\hat{b}_i - b_i}{\hat{\sigma}_{b_i}} \sim t_{n-2}.$

Tabela de regressão (contém os testes individuais  $H_0 : b_i = 0 \quad vs \quad H_1 : b_i \neq 0$ )

Coeficiente	Coeficientes não-estandardizados		Coeficientes estandardizados		$p - value$
	$b$	Erro padrão	$\beta$	$t$	
Constante	$\hat{b}_0$	$\hat{\sigma}_{b_0}$		$t_{0obs}$	(·)
Declive	$\hat{b}_1$	$\hat{\sigma}_{b_1}$	$\hat{\beta}_1$	$t_{1obs}$	(·)

Avaliação da qualidade e validação dos pressupostos da regressão:

Avaliação: Diagramas de dispersão; Coeficiente de determinação  $R^2$  - mede a proporção de variabilidade de  $Y$  explicada por  $x$ ; Teste ao declive (significado da regressão).

Validação: os resíduos ( $e_i$ ) devem ser Normais com variância constante e independentes: QQ-plot ; Gráfico de resíduos versus valores preditos  $\hat{Y}_i$  ou observados  $Y_i$ .

**Testes de hipóteses:** Num teste de hipóteses a hipótese nula deve ser sempre simples.

$p$ -value do teste: é a probabilidade de observar um valor da estatística de teste tanto ou mais afastado que o valor observado na amostra, assumindo que  $H_0$  é verdadeira.

Num teste de tamanho  $\alpha$  rejeita-se a hipótese nula (de igualdade) quando  $p$ -value  $\leq \alpha$ .

Para transformar um  $p$ -value bilateral em unilateral divide-se por dois desde que a(s) amostra(s) aponte(m) no sentido da hipótese alternativa. Caso contrário calcula-se  $(1-p\text{-value}/2)$ .

Há 3 procedimentos para realizar um teste de hipóteses:

1. Cálculo da região crítica
2. Através do  $p$ -value.
3. Através de intervalos de confiança (válido apenas para testes bilaterais). Neste caso rejeita-se  $H_0$  se o valor em teste não pertencer ao IC construído para o parâmetro.