

FORMULÁRIO DE ESTATÍSTICA II

VALOR ESPERADO, MOMENTOS E PARÂMETROS

$$\text{Var}(X) = E(X - \mu)^2 = E(X^2) - \mu^2;$$

$$\text{Cov}(X, Y) = E\{(X - \mu_X)(Y - \mu_Y)\} = E(XY) - E(X)E(Y); \quad \rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$E(aX + bY) = aE(X) + bE(Y) \text{ e } \text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y) \text{ com } a, b \text{ constantes}$$

DISTRIBUIÇÕES TEÓRICAS

- **UNIFORME (DISCRETA)**

$$\text{Caso } f(x) = \frac{1}{n}, \quad x = 1, 2, \dots, n; \quad E(X) = \frac{n+1}{2}; \quad \text{Var}(X) = \frac{n^2 - 1}{12}$$

$$\text{Caso } f(x) = \frac{1}{m+1}, \quad x = 0, 1, 2, \dots, m; \quad E(X) = \frac{m}{2}; \quad \text{Var}(X) = \frac{m(m+2)}{12};$$

- **BINOMIAL** $X \sim B(n; \theta)$, $(0 < \theta < 1)$

$$f(x | \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, 2, \dots, n$$

$$E(X) = n\theta; \quad \text{Var}(X) = n\theta(1 - \theta); \quad \mathfrak{S}(\theta) = \frac{n}{\theta(1 - \theta)}, \quad n \text{ conhecido}$$

Propriedades:

- $X \sim B(n, \theta) \Leftrightarrow (n - X) \sim B(n, 1 - \theta)$
- $X_1 \sim B(n_1; \theta), X_2 \sim B(n_2; \theta), X_1 \text{ e } X_2 \text{ independentes} \Rightarrow X_1 + X_2 \sim B(n_1 + n_2, \theta)$

- **BERNOULLI** $X \sim B(1; \theta)$

- **POISSON** $X \sim \text{Po}(\lambda)$, $(\lambda > 0)$

$$f(x | \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots; \quad E(X) = \lambda; \quad \text{Var}(X) = \lambda; \quad \mathfrak{S}(\lambda) = 1/\lambda;$$

Propriedades:

- $X_1 \sim \text{Po}(\lambda_1), X_2 \sim \text{Po}(\lambda_2), X_1 \text{ e } X_2 \text{ independentes} \Rightarrow X_1 + X_2 \sim \text{Po}(\lambda_1 + \lambda_2)$
- Se $X \sim B(n; \theta)$, com n grande θ pequeno então $X \stackrel{a}{\sim} \text{Po}(n\theta)$

- **UNIFORME (CONTÍNUA)** $X \sim U(\alpha, \beta)$, $(\alpha < \beta)$

$$f(x | \alpha, \beta) = \frac{1}{\beta - \alpha} \quad \alpha < x < \beta; \quad E(X) = \frac{\alpha + \beta}{2}; \quad \text{Var}(X) = \frac{(\beta - \alpha)^2}{12};$$

- **NORMAL** $X \sim N(\mu, \sigma^2)$, $(-\infty < \mu < +\infty, 0 < \sigma < +\infty)$

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

$$E(X) = \mu; \quad \text{Var}(X) = \sigma^2; \quad \mathfrak{S}(\mu) = 1/\sigma^2, \quad \sigma^2 \text{ conhecido}; \quad \mathfrak{S}(\sigma^2) = 1/(2\sigma^4), \quad \mu \text{ conhecido}$$

Propriedades:

- Normal estandardizada $Z = \frac{X - \mu}{\sigma} \sim N(0, 1); \quad \phi(z) = \phi(-z) \text{ e } \Phi(z) = 1 - \Phi(-z)$
- $X_i \sim N(\mu, \sigma^2) (i = 1, 2, \dots, k) \text{ independentes} \Rightarrow Y = \sum_{i=1}^k X_i \sim N(k\mu, k\sigma^2) \text{ e } \bar{X} = \frac{1}{k} \sum_{i=1}^k X_i \sim N\left(\mu, \frac{\sigma^2}{k}\right)$
- $X_i \sim N(\mu_i, \sigma_i^2) (i = 1, 2, \dots, k) \text{ independentes} \Rightarrow \sum_{i=1}^k \alpha_i X_i \sim N(\mu_Y, \sigma_Y^2) \text{ com } \mu_Y = \sum_{i=1}^k \alpha_i \mu_i \text{ e } \sigma_Y^2 = \sum_{i=1}^k \alpha_i^2 \sigma_i^2$

- **EXPONENCIAL** $X \sim \text{Ex}(\lambda)$, ($\lambda > 0$) ; $X \sim \text{Ex}(\lambda) \Leftrightarrow X \sim G(1, \lambda)$

$$f(x | \lambda) = \lambda e^{-\lambda x}, x > 0 \quad E(X) = \frac{1}{\lambda}; \text{Var}(X) = \frac{1}{\lambda^2}; \mathfrak{S}(\lambda) = 1/\lambda^2$$

Propriedades:

- $X_i \sim \text{Ex}(\lambda)$ ($i = 1, 2, \dots, k$) independentes $\Rightarrow \sum_{i=1}^k X_i \sim G(k, \lambda)$ e $\min_i X_i \sim \text{Ex}(k\lambda)$

- **GAMA** $X \sim G(\alpha, \lambda)$, ($\lambda > 0, \alpha > 0$)

$$f(x | \alpha, \lambda) = \frac{\lambda^\alpha e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)}, \quad x > 0; E(X) = \frac{\alpha}{\lambda}; \text{Var}(X) = \frac{\alpha}{\lambda^2}; \mathfrak{S}(\lambda) = \alpha/\lambda^2, \lambda^2 \text{ conhecido}$$

Propriedades:

- $X_i \sim G(\alpha_i; \lambda)$, ($i = 1, 2, \dots, k$) independentes $\Rightarrow \sum_{i=1}^k X_i \sim G\left(\sum_{i=1}^k \alpha_i; \lambda\right)$
- $X \sim G(\alpha, \lambda)$ então $Y = cX \sim G\left(\alpha, \frac{\lambda}{c}\right)$ onde c constante positiva

- **QUI-QUADRADO** $X \sim \chi^2(n)$, (n inteiro positivo).

$$X \sim \chi^2(n) \Leftrightarrow X \sim G(n/2; 1/2); E(X) = n; \text{Var}(X) = 2n$$

Propriedades:

- $X_i \sim \chi^2_{(n_i)}$ ($i = 1, 2, \dots, k$) independentes $\Rightarrow \sum_{i=1}^k X_i \sim \chi^2_{(n)}$ com $n = \sum_{i=1}^k n_i$
- $X \sim G(n; \lambda) \Leftrightarrow 2\lambda X \sim \chi^2(2n)$
- $X_i \sim N(0,1)$, ($i = 1, 2, \dots, n$) independentes $\Rightarrow \sum_{i=1}^n X_i^2 \sim \chi^2(n)$

- **t-“STUDENT”**

$$T = \frac{U}{\sqrt{V/n}} \sim t(n) \text{ com } U \sim N(0,1) \text{ e } V \sim \chi^2(n) \text{ independentes}$$

$$E(T) = 0; \text{Var}(T) = \frac{n}{n-2} \quad (n > 2); \text{ Sendo } T \sim t(n) \Rightarrow \lim_{n \rightarrow \infty} F_T(t | n) = \Phi(t)$$

- **F-SNEDCOR**

$$F = \frac{U/m}{V/n} \sim F(m, n) \text{ com } U \sim \chi^2(m), V \sim \chi^2(n) \text{ (independentes)}$$

$$\text{Propriedades: } X \sim F(m, n) \Rightarrow \frac{1}{X} \sim F(n, m) \quad . \quad T \sim t_{(n)} \Rightarrow T^2 \sim F(1, n)$$

TEOREMA DO LIMITE CENTRAL E COROLÁRIOS

$$\text{TLC: Sendo } X_i \text{ iid com } E(X_i) = \mu \text{ e } \text{Var}(X_i) = \sigma^2 \Rightarrow \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n} \sigma} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\sim} N(0,1)$$

$$\text{Corolário: Sendo } X_i \sim B(1; \theta), \text{ iid} \Rightarrow \frac{\sum_{i=1}^n X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \stackrel{a}{\sim} N(0,1)$$

$$\text{Correcção de continuidade: } P(a \leq X \leq b) \approx \Phi\left(\frac{b + \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right), \text{ com } a \text{ e } b \text{ inteiros}$$

$$\text{Corolário: Sendo } X \sim \text{Po}(\lambda), \text{ quando } \lambda \rightarrow +\infty \Rightarrow \frac{X - \lambda}{\sqrt{\lambda}} \stackrel{a}{\sim} N(0,1)$$

$$\text{Correcção de continuidade: } P(a \leq X \leq b) \approx \Phi\left(\frac{b + \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - \frac{1}{2} - \lambda}{\sqrt{\lambda}}\right), \text{ com } a \text{ e } b \text{ inteiros}$$

AMOSTRAGEM. DISTRIBUIÇÕES POR AMOSTRAGEM

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

$$E(\bar{X}) = \mu \quad ; \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad ; \quad E(S^2) = \frac{n-1}{n} \sigma^2; \quad E(S'^2) = \sigma^2$$

DISTRIBUIÇÃO DO MÍNIMO E DO MÁXIMO

$$G_1(x) = 1 - [1 - F(x)]^n \quad ; \quad G_n(x) = [F(x)]^n$$

• POPULAÇÕES NORMAIS

| | | |
|-----------------------|---|---|
| Média | $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ | $\frac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t(n-1)$ |
| Diferença de médias | $\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$ | $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \sim t(\nu)$ |
| | $T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{m} + \frac{1}{n}}} \cdot \frac{1}{\sqrt{\frac{(m-1)S_1'^2 + (n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$ | onde ν é o maior inteiro contido em r , $r = \frac{\left(\frac{S_1'^2}{m} + \frac{S_2'^2}{n}\right)^2}{\frac{1}{m-1} \left(\frac{S_1'^2}{m}\right)^2 + \frac{1}{n-1} \left(\frac{S_2'^2}{n}\right)^2}$ |
| Variância | $\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$ | |
| Relação de variâncias | $\frac{S_1'^2}{S_2'^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1)$ | |

• GRANDES AMOSTRAS

Caso geral

| | | |
|---------------------|--|--|
| Média | $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\sim} N(0,1)$ | $\frac{\bar{X} - \mu}{S'/\sqrt{n}} \stackrel{a}{\sim} N(0,1)$ |
| Diferença de médias | $\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \stackrel{a}{\sim} N(0,1)$ | $\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{m} + \frac{S_2'^2}{n}}} \stackrel{a}{\sim} N(0,1)$ |

População de Bernoulli

| | | |
|-------------------------|--|--|
| Proporção | $\frac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \stackrel{a}{\sim} N(0,1)$ | $\frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \stackrel{a}{\sim} N(0,1)$ |
| Diferença de proporções | $\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{m} + \frac{\theta_2(1-\theta_2)}{n}}} \stackrel{a}{\sim} N(0,1)$ | $\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\bar{X}_1(1-\bar{X}_1)}{m} + \frac{\bar{X}_2(1-\bar{X}_2)}{n}}} \stackrel{a}{\sim} N(0,1)$ |
| Igualdade de proporções | $\frac{\bar{X}_1 - \bar{X}_2 - (\theta_1 - \theta_2)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \hat{\theta}(1-\hat{\theta})}} \stackrel{a}{\sim} N(0,1)$ onde $\hat{\theta} = \frac{m\bar{X}_1 + n\bar{X}_2}{m+n}$ | |

• **ESTATÍSTICA-TESTE DO χ^2**

Teste de Ajustamento: $Q = \sum_{j=1}^m \frac{(N_j - fe_j)^2}{fe_j} \sim \chi^2(m-1)$

Com estimação de k parâmetros para obter as estimativas $\hat{p}_{\circ j}$: $\chi^2_{(m-k-1)}$

Teste de Independência: $Q = \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{ij} - fe_{ij})^2}{fe_{ij}} \sim \chi^2_{((r-1)(s-1))}$

Teste de Homogeneidade: $Q = \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{ij} - fe_{ij})^2}{fe_{ij}} \sim \chi^2_{((r-1)s)}$

• **MODELO REGRESSÃO LINEAR** $y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + u_t, t = 1, 2, \dots, n.$

EMQ (estimadores dos mínimos quadrados)

| Caso geral | Caso particular: $y_t = \beta_1 + \beta_2 x_t + u_t$ |
|--|---|
| $b = (X^T X)^{-1} X^T Y$ | $b_1 = \bar{y} - b_2 \bar{x} ; \hat{Var}(b_1 X) = \frac{s^2 \sum x_t^2}{n \sum x_t^2 - (\sum x_t)^2}$ |
| $\hat{u}_t = y_t - \hat{y}_t$ | $b_2 = \frac{n \sum x_t y_t - \sum x_t \sum y_t}{n \sum x_t^2 - (\sum x_t)^2};$ |
| $s^2 = \frac{\sum \hat{u}_t^2}{(n-k)}$ | $\hat{Var}(b_2 X) = \frac{ns^2}{n \sum x_t^2 - (\sum x_t)^2}$ |
| $Côv(b X) = s^2 (X^T X)^{-1}$ | $s^2 = \frac{\sum \hat{u}_t^2}{(n-2)}$ |

Propriedades:

$$\sum_{t=1}^n x_{jt} \hat{u}_t = 0 \quad (j = 1, 2, \dots, k); \quad \sum_{t=1}^n \hat{y}_t \hat{u}_t = 0; \quad \sum_{t=1}^n y_t^2 = \sum_{t=1}^n \hat{y}_t^2 + \sum_{t=1}^n \hat{u}_t^2; \quad r_{\hat{y}\hat{y}}^2 = \frac{(\sum_t (y_t - \bar{y}_t)(\hat{y}_t - \bar{\hat{y}}_t))^2}{\sum_t (y_t - \bar{y}_t)^2 \sum_t (\hat{y}_t - \bar{\hat{y}}_t)^2}$$

No modelo com termo independente:

$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + u_t, t = 1, 2, \dots, n$

$VT = VE + VR; \quad VT = \sum_{t=1}^n (y_t - \bar{y})^2; \quad VE = \sum_{t=1}^n (\hat{y}_t - \bar{y})^2; \quad VR = \sum_{t=1}^n \hat{u}_t^2$

$R^2 = \frac{VE}{VT} = 1 - \frac{VR}{VT}; \quad \bar{R}^2 = 1 - \frac{VR/(n-k)}{VT/(n-1)} = 1 - (1-R^2) \frac{n-1}{n-k}.$

Inferência estatística do MRL, com $y_t | X \sim N(x_t, \beta, \sigma^2)$:

• $q = \frac{\sum_{t=1}^n \hat{u}_t^2}{\sigma^2} = \frac{(n-k)s^2}{\sigma^2} \sim \chi^2(n-k)$

• $t_j = \frac{b_j - \beta_j}{s_{b_j}} \sim t(n-k) \quad \text{ou} \quad F_j = \frac{(b_j - \beta_j)^2}{s_{b_j}^2} \sim F(1, n-k)$

Testes de restrições lineares sobre os coeficientes de regressão

Caso geral (m restrições lineares): $F = \frac{(VR_0 - VR_1)/m}{VR_1/(n-k)} \sim F(m, n-k)$

VR_0 = Variação residual do modelo com as m restrições lineares;

VR_1 = Variação residual do modelo sem restrições

Casos particulares

• Uma única restrição ($m=1$): $t_{\hat{\delta}} = \frac{\hat{\delta} - \delta}{s_{\hat{\delta}}} \sim t(n-k)$ em que $s_{\hat{\delta}}$ é o erro padrão de $\hat{\delta} = cb$.

• Nulidade conjunta: $F = \frac{R^2 / (k-1)}{(1-R^2)/(n-k)} = \frac{VE/(k-1)}{VR/(n-k)} \sim F(k-1, n-k)$

• Nulidade de um subconjunto (m)

$$F = \frac{(VR_0 - VR_1)/m}{VR_1/(n-k)} \sim F(m, n-k) \text{ ou } F = \frac{(R^2 - R_0^2)/m}{(1-R^2)/(n-k)} \sim F(m, n-k)$$

VR_1 = Variação residual do modelo sem restrições;

VR_0 = Variação residual do modelo com restrições

R_0 = Coeficiente de determinação do modelo com restrições

COMPLEMENTOS AO MODELO DE REGRESSÃO LINEAR

Permanência de estrutura: $F_{Chow} = \frac{(VR_0 - VR_1)/k}{VR_1/(n-2k)} \sim F(k, n-2k)$

n_1 : N° de observações do modelo original;

$(n - n_1)$: N° de observações adicionais;

VR_0 = Variação residual do modelo com n observações

VR_1 = Soma das variações residuais dos modelos com n_1 e $(n-n_1)$ observações

Previsão

1. Previsão em Média: $\theta = \beta_1 + \beta_2 c_2 + \dots + \beta_k c_k$; $\hat{\theta} = b_1 + b_2 c_2 + \dots + b_k c_k$; $t_{\hat{\theta}} = \frac{\hat{\theta} - \theta}{s \sqrt{c(X^T X)^{-1} c^T}} = \frac{\hat{\theta} - \theta}{s_{\hat{\theta}}} \sim t(n-k)$

2. Previsão pontual: $y_0 = \beta_1 + \beta_2 c_2 + \dots + \beta_k c_k + u_0$; $\hat{y}_0 = b_1 + b_2 c_2 + \dots + b_k c_k$;

$$t_d = \frac{y_0 - \hat{y}_0}{s \sqrt{1 + c(X^T X)^{-1} c^T}} = \frac{d}{s_d} \sim t(n-k) \quad \text{com } s_d = \sqrt{s^2 + s_{\hat{\theta}}^2}$$

Heteroscedasticidade:

$$Cov(b | X) = (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} = (X^T X)^{-1} \sum_{t=1}^n \sigma_t^2 x_t^T x_t (X^T X)^{-1}$$

$$\hat{Cov}(b | X) = (X^T X)^{-1} \sum_{t=1}^n \hat{u}_t^2 x_t^T x_t (X^T X)^{-1}, \quad \hat{u}_t^2 = (y_t - x_t \cdot b)^2.$$

$$\text{Estatística-teste de White: } W = nR^2 \overset{a}{\sim} \chi^2(p-1)$$

R^2 e p - Coeficiente de determinação e n° de regressores da regressão auxiliar, respectivamente.

Inferência sobre β_j : $t_j = \frac{b_j - \beta_j}{s_{b_j}^*} \overset{a}{\sim} N(0,1)$ com $s_{b_j}^*$ = erro-padrão heterocedástico-consistente

Autocorrelação:

$$\text{Estatística-teste de Breusch-Godfrey: } nR^2 \overset{a}{\sim} \chi^2(q),$$

R^2 e q - Coeficiente de determinação da regressão auxiliar e ordem máxima de defasamento, respectivamente.