FORMULÁRIO DE ESTATÍSTICA II

VALOR ESPERADO, MOMENTOS E PARÂMETROS

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

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

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

DISTRIBUIÇÕES TEÓRICAS

- **UNIFORME (DISCRETA)**

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

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

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

$f(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$, $x = 0, 1, 2, ..., n$

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

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

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

$f(x \mid \lambda) = \dfrac{e^{-\lambda}\lambda^x}{x!}$, $x = 0,1,2,...$; $E(X) = \lambda$; $\text{Var}(X) = \lambda$; $\Im(\lambda) = 1/\lambda$;

Propriedades:

- $X_1 \sim \text{Po}(\lambda_1)$, $X_2 \sim \text{Po}(\lambda_2)$, X_1 e X_2 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 $X \overset{a}{\sim} \text{Po}(n\theta)$

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

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

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

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

$E(X) = \mu$; $\text{Var}(X) = \sigma^2$; $\Im(\mu) = 1/\sigma^2$, σ^2 conhecido ; $\Im(\sigma^2) = 1/(2\sigma^4)$, μ conhecido

Propriedades:

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

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

  $f(x \mid \lambda) = \lambda e^{-\lambda x}$, $x > 0$ $E(X) = \dfrac{1}{\lambda}$; $\text{Var}(X) = \dfrac{1}{\lambda^2}$; $\Im(\lambda) = 1/\lambda^2$

  Propriedades:

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

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

  $f(x \mid \alpha, \lambda) = \dfrac{\lambda^{\alpha} e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)}$, $x > 0$; $E(X) = \dfrac{\alpha}{\lambda}$; $\text{Var}(X) = \dfrac{\alpha}{\lambda^2}$; $\Im(\lambda) = \alpha/\lambda^2$, λ^2 conhecido

  Propriedades:

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

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

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

  Propriedades:

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

- ***t*-"STUDENT"**

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

  $E(T) = 0$; $\text{Var}(T) = \dfrac{n}{n-2}$ $(n > 2)$; Sendo $T \sim t(n) \Rightarrow \lim_{n\to\infty} F_T(t \mid n) = \Phi(t)$

- ***F*-SNEDCOR**

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

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

TEOREMA DO LIMITE CENTRAL E COROLÁRIOS

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

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

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

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

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

AMOSTRAGEM. DISTRIBUIÇÕES POR AMOSTRAGEM

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

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

DISTRIBUIÇÃO DO MÍNIMO E DO MÁXIMO

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

- **POPULAÇÕES NORMAIS**

Média	$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\dfrac{\overline{X}-\mu}{S'/\sqrt{n}} \sim t(n-1)$
Diferença de médias	$\dfrac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} \sim N(0,1)$ $T = \dfrac{\dfrac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}}{\sqrt{\dfrac{(m-1)S_1'^2+(n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$	$Z = \dfrac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1'^2}{m}+\dfrac{S_2'^2}{n}}} \overset{a}{\sim} t(\nu)$ onde ν é o maior inteiro contido em *r*, $r = \dfrac{\left(\dfrac{s_1'^2}{m}+\dfrac{s_2'^2}{n}\right)^2}{\dfrac{1}{m-1}\left(\dfrac{s_1'^2}{m}\right)^2+\dfrac{1}{n-1}\left(\dfrac{s_2'^2}{n}\right)^2}$
Variância	$\dfrac{nS^2}{\sigma^2} = \dfrac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	
Relação de variâncias	$\dfrac{S_1'^2}{S_2'^2}\dfrac{\sigma_2^2}{\sigma_1^2} \sim F(m-1,n-1)$	

- **GRANDES AMOSTRAS**

Caso geral

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

População de Bernoulli

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

- **ESTATÍSTICA-TESTE DO χ^2**

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

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

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

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

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

EMQ (estimadores dos mínimos quadrados)

Caso geral	**Caso particular:** $y_t = \beta_1 + \beta_2 x_t + u_t$
$b = (X^T X)^{-1} X^T Y$ $\hat{u}_t = y_t - \hat{y}_t$ $s^2 = \frac{\sum \hat{u}_t^2}{(n-k)}$ $C\hat{o}v(b \mid X) = s^2 (X^T X)^{-1}$	$b_1 = \bar{y} - b_2\bar{x}$ **;** $V\hat{a}r(b_1 \mid X) = \frac{s^2 \sum x_t^2}{n\sum x_t^2 - \left(\sum x_t\right)^2}$ $b_2 = \frac{n\sum x_t y_t - \sum x_t \sum y_t}{n\sum x_t^2 - \left(\sum x_t\right)^2}$; $V\hat{a}r(b_2 \mid X) = \frac{ns^2}{n\sum x_t^2 - \left(\sum x_t\right)^2}$ $s^2 = \frac{\sum \hat{u}_t^2}{(n-2)}$

Propriedades:

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

No modelo com termo independente:

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

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

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

Inferência estatística do MRL, com $y_t \mid X \sim N(x_{t\bullet}\beta, \sigma^2)$:

- $q = \frac{\sum_{t=1}^{n} \hat{u}_t^2}{\sigma^2} = \frac{(n-k)s^2}{\sigma^2} \sim \chi^2(n-k)$
- $t_j = \frac{b_j - \beta_j}{s_{b_j}} \sim t(n-k)$ ou $F_j = \frac{\left(b_j - \beta_j\right)^2}{s_{b_j}^2} \sim F(1, n-k)$

Testes de restrições lineares sobre os coeficientes de regressão

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

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): $t_{\hat{\delta}} = \dfrac{\hat{\delta}-\delta}{s_{\hat{\delta}}} \sim t(n-k)$ em que $s_{\hat{\delta}}$ é o erro padrão de $\hat{\delta} = cb$.

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

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

n_1: Nº de observações do modelo original;

$(n-n_1)$: Nº 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: $\theta = \beta_1 + \beta_2 c_2 + ... + \beta_k c_k$; $\hat{\theta} = b_1 + b_2 c_2 + ... + b_k c_k$; $t_{\hat{\theta}} = \dfrac{\hat{\theta}-\theta}{s\sqrt{c(X^T X)^{-1} c^T}} = \dfrac{\hat{\theta}-\theta}{s_{\hat{\theta}}} \sim t(n-k)$

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

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

Heteroscedasticidade:

$$Cov(b \mid X) = (X^T X)^{-1} X^T \Sigma X (X^T X)^{-1} = (X^T X)^{-1} \sum\nolimits_{t=1}^{n} \sigma_t^2 x_{t\bullet}^T x_{t\bullet} (X^T X)^{-1}$$

$$C\hat{o}v(b \mid X) = (X^T X)^{-1} \sum\nolimits_{t=1}^{n} \hat{u}_t^2 x_{t\bullet}^T x_{t\bullet} (X^T X)^{-1}, \quad \hat{u}_t^2 = (y_t - x_{t\bullet} b)^2 .$$

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

R^2 e p - Coeficiente de determinação e nº de regressores da regressão auxiliar, respectivamente.

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

Autocorrelação:

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 desfasamento, respectivamente.