FORMULÁRIO DE ESTATÍSTICA II

VALOR ESPERADO, MOMENTOS E PARÂMETROS

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

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

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

DISTRIBUIÇÕES TEÓRICAS

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

  $f(x\,|\,\theta) = \theta^x (1-\theta)^{1-x},\ x = 0,1$ $\quad E(X) = \theta \quad \mathrm{Var}(X) = \theta(1-\theta)$

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

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

  $E(X) = n\theta$; $\mathrm{Var}(X) = n\theta(1-\theta)$;

  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 $\Rightarrow X_1 + X_2 \sim B(n_1+n_2,\theta)$

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

  $f(x\,|\,\lambda) = \dfrac{e^{-\lambda}\lambda^x}{x!},\ x = 0,1,2,\ldots$; $E(X) = \lambda$; $\mathrm{Var}(X) = \lambda$;

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

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

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

  $E(X) = \mu$; $\mathrm{Var}(X) = \sigma^2$; $\Im(\mu) = 1/\sigma^2$, σ^2 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 $\bar{X} = \dfrac{1}{k}\sum_{i=1}^k X_i \sim N\left(\mu, \dfrac{\sigma^2}{k}\right)$
  - $X_i \sim N(\mu_i,\sigma_i^2)$ (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** $X \sim \text{Ex}(\lambda)$**,** $(\lambda > 0)$; $X \sim \text{Ex}(\lambda) \Leftrightarrow X \sim \text{G}(1, \lambda)$

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

- **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 $n = \sum_{i=1}^{k} n_i$
  - $X_i \sim N(0,1), \ (i = 1, 2, ..., n)$ independentes $\Rightarrow \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$

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

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

  $E(T) = 0$; $\text{Var}(T) = \dfrac{n}{n-2}$ $(n > 2)$;

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

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

AMOSTRAGEM. DISTRIBUIÇÕES POR AMOSTRAGEM

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

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

DISTRIBUIÇÃO DO MÍNIMO E DO MÁXIMO

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

- **POPULAÇÕES NORMAIS**

Média	$\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\dfrac{\bar{X} - \mu}{S'/\sqrt{n}} \sim t(n-1)$
Variância	$\dfrac{nS^2}{\sigma^2} = \dfrac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	

- **GRANDES AMOSTRAS**

Caso geral

Média	$\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0,1)$	$\dfrac{\bar{X} - \mu}{S'/\sqrt{n}} \overset{a}{\sim} N(0,1)$

População de Bernoulli

Proporção	$\dfrac{\bar{X} - \theta}{\sqrt{\dfrac{\theta(1-\theta)}{n}}} \overset{a}{\sim} N(0,1)$	$\dfrac{\bar{X} - \theta}{\sqrt{\dfrac{\bar{X}(1-\bar{X})}{n}}} \overset{a}{\sim} N(0,1)$

- **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**

EMQ (estimadores dos mínimos quadrados)

Propriedades:

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

No modelo com termo independente:

$$y_t=\beta_0+\beta_1x_{1i}+...+\beta_kx_{ki}+u_i,\ i=1,2,\ldots,n$$

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

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

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

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

Testes de restrições lineares sobre os coeficientes de regressão

Caso geral (m restrições lineares): $F=\frac{\left(VR_0-VR_1\right)/m}{VR_1/(n-k-1)}\sim F(m,n-k-1)$

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

VR_1 = Variação residual do modelo sem restrições

Casos particulares

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

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

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 pontual: $y_0 = \beta_0 + \beta_1 x_{01} + \ldots + \beta_k x_{0k} + u_0$;

2. Previsão em Média: $\hat{y}_0 = \beta_0 + \beta_1 x_{01} + \ldots + \beta_k x_{0k}$; $t_{\hat{y}_0} = \dfrac{\hat{y}_0 - y_0}{s_0} \sim t(n-k-1) \quad com \quad s_0 = s/\sqrt{n}$