FORMULÁRIO DE ESTATÍSTICA – MAEG

- **ESTATÍSTICA DESCRITIVA (**amostra com n observações)

- **Variância** $s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$;

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

- **Covariância** $s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$

- **Coeficiente de correlação** $r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{s_{xy}}{s_x\, s_y}, \quad -1 \le r \le 1$

- **PROBABILIDADE: MOMENTOS E PARÂMETROS**

- **Momentos:** $\mu'_k = E(X^k)$; $\mu_k = E[(X-\mu)^k]$
- **Variância:** $\sigma^2 = \mathrm{Var}(X) = E(X-\mu)^2 = E(X^2) - \mu^2$
- **Momentos de ordem *r*+*s*:** $\mu'_{rs} = E(X^r Y^s)$; $\mu_{rs} = E\{(X-\mu_X)^r (Y-\mu_Y)^s\}$
- **Covariância:** $\mathrm{Cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right] = E(XY) - E(X)E(Y)$
- $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\mathrm{Cov}(X,Y)$
- **Coeficiente de assimetria:** $\gamma_1 = E(X-\mu)^3 / \sigma^3$
- **Coeficiente de Kurtosis:** $\beta_2 = E(X-\mu)^4 / \sigma^4$
- **Coeficiente de correlação** $\rho = \frac{E\left[(X-\mu_X)(Y-\mu_Y)\right]}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}} = \frac{\mathrm{cov}(X,Y)}{\sigma_X\, \sigma_Y}, \quad -1 \le \rho \le 1$
- **Funções geradoras de momentos:** $M_X(s) = E(e^{sX})$ se o V.E. existir numa viz. de 0
- **Funções geradoras de probabilidades:** $P_X(s) = E(s^X)$
- **Momentos condicionados**

  $E(X) = E(E(X \mid Y))$ $\quad\quad$ $\mathrm{Var}(X) = E(\mathrm{Var}(X \mid Y)) + \mathrm{Var}(E(X \mid Y))$

- **Funções Gama e Beta**

  A função Gama está definida para $x > 0, x \in \Re$

  $\Gamma(x) = (x-1)\Gamma(x-1)$ **para** $x > 1$

  $\Gamma(1) = 1$ **e** $\Gamma(1/2) = \sqrt{\pi}$

  $\Gamma(x) = (x-1)!$ **para** $x \in N$

  $\beta(m,n) = \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)}$ **para** $m > 0, n > 0, m \in \Re, n \in \Re$

- **DISTRIBUIÇÕES TEÓRICAS**
  - **BERNOULLI** $X \sim B(1,\theta)$

    $f(x\,|\,\theta)=\theta^x(1-\theta)^{1-x}, \quad x=0,1 \quad (0<\theta<1)$

    $E(X)=\theta$; $\mathrm{Var}(X)=\theta(1-\theta)$; $M(s)=(1-\theta)+\theta\, e^s$; $\Im(\theta)=1/\left[\theta(1-\theta)\right]$
  - **BINOMIAL** $X \sim B(k,\theta)$

    $f(x\,|\,\theta)=\binom{k}{x}\theta^x(1-\theta)^{k-x},\ x=0,1,2,...,k \quad (0<\theta<1)$

    $E(X)=k\theta$; $\mathrm{Var}(X)=k\theta(1-\theta)$; $M(s)=\left[(1-\theta)+\theta\, e^s\right]^k$; $\Im(\theta)=k/\left[\theta(1-\theta)\right]$

    Propriedades: $X \sim B(k;\theta) \Leftrightarrow (k-X) \sim B(k;1-\theta)$

    $Y_1 \sim B(n_1;\theta,),\ Y_2 \sim B(n_2;\theta), Y_1$ e Y_2 indep. $\Rightarrow Y=Y_1+Y_2 \sim B(n_1+n_2,\theta)$
  - **POISSON** $X \sim \mathrm{Po}(\lambda)$

    $f(x\,|\,\lambda)=\dfrac{e^{-\lambda}\lambda^x}{x!},\ x=0,1,2,...,\ (\lambda>0)$

    $E(X)=\lambda$; $\mathrm{Var}(X)=\lambda$; $M(s)=\exp\{\lambda(e^s-1)\}$; $\Im(\lambda)=1/\lambda$

    Propriedades: $X_i \sim \mathrm{Po}(\lambda_i)$ e X_i indep. $\Rightarrow \sum_{i=1}^k X_i \sim \mathrm{Po}(\sum_{i=1}^k \lambda_i)$

    $X \sim b(n,\theta)$, θ pequeno, n grande $\Rightarrow X \overset{a}{\sim} \mathrm{Po}(n\theta)$
  - **UNIFORME (CONTÍNUA)** $X \sim U(\alpha,\beta)$

    $f(x\,|\,\alpha,\beta)=\dfrac{1}{\beta-\alpha}, \quad \alpha<x<\beta \qquad E(X)=\dfrac{\alpha+\beta}{2}$; $\mathrm{Var}(X)=\dfrac{(\beta-\alpha)^2}{12}$;
  - **NORMAL** $X \sim N(\mu,\sigma)$

    $f(x\,|\,\mu,\sigma)=\dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\dfrac{1}{2\sigma^2}(x-\mu)^2\right\}, \quad x\in\Re \quad (\mu\in\Re,\sigma^2>0)$

    $E(X)=\mu$; $\mathrm{Var}(X)=\sigma^2$; $M(s)=\exp\left\{\mu s+\dfrac{\sigma^2 s^2}{2}\right\}$;

    $\Im(\mu)=1/\sigma^2$ (σ^2 conhecido); $\Im(\sigma^2)=1/(2\sigma^4)$ (μ conhecido)

    Propriedades: A normal (0,1) é simétrica em torno de 0

    $\phi(x)=\phi(-x), \quad \Phi(x)=1-\Phi(-x)$

    $X_i \sim N(\mu_i,\sigma_i^2) \Rightarrow Y=\sum_{i=1}^n \alpha_i X_i \sim N\left(\sum_{i=1}^n \alpha_i\mu_i,\sigma_Y^2\right)$

    X_i indep $\rightarrow \sigma_Y^2=\sum_{i=1}^n \alpha_i^2\sigma_i^2$

    $X_i \sim N(\mu,\sigma^2)$ e X_i indep. $\Rightarrow \overline{X}=\dfrac{1}{n}\sum_{i=1}^n X_i \sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$
  - **EXPONENCIAL** $X \sim \mathrm{Ex}(\lambda)$ ou $X \sim G(1,\lambda)$

    $f(x\,|\,\lambda)=\lambda e^{-\lambda x} \quad x>0 \quad (\lambda>0) \qquad F(x\,|\,\lambda)=1-e^{-\lambda x} \quad x>0 \quad (\lambda>0)$

    $E(X)=\dfrac{1}{\lambda}$; $\mathrm{Var}(X)=\dfrac{1}{\lambda^2}$; $M(s)=\dfrac{\lambda}{\lambda-s}, \quad s<\lambda$; $\Im(\lambda)=1/\lambda^2$

    Propriedades: $X_i \sim \mathrm{Ex}(\lambda)$ e X_i indep $(i=1,2,...,k) \Rightarrow Y=\min_i X_i \sim \mathrm{Ex}(k\lambda)$

    ver as propriedades da distribuição gama

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

$$f(x \mid \alpha, \lambda) = \frac{\lambda^{\alpha} e^{-\lambda x} x^{\alpha-1}}{\Gamma(\alpha)}, \qquad x > 0, \ (\alpha, \lambda > 0)$$

$$E(X) = \frac{\alpha}{\lambda}; \mathrm{Var}(X) = \frac{\alpha}{\lambda^2}; M(s) = \left(\frac{\lambda}{\lambda - s}\right)^{\alpha}, s < \lambda; \ \Im(\lambda) = \alpha / \lambda^2 \ (\alpha \text{ conhecido})$$

Propriedade:

$X_i \sim G(\alpha_i; \lambda)$ e X_i indep $(i = 1,2,...,n) \Rightarrow X = \sum_{i=1}^{n} X_i \sim \mathrm{G}\left(\sum_{i=1}^{n} \alpha_i; \lambda\right)$

Função gama: $\Gamma(\alpha) = \int_0^{+\infty} e^{-x} x^{\alpha-1} dx \qquad (\alpha > 0)$

Propriedades da função gama: $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1), \alpha > 1; \ \Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi}$

- **QUI-QUADRADO** $X \sim \chi^2_{(n)}$

$$f(x \mid n) = \frac{e^{-\frac{x}{2}} x^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma(n/2)}, \quad x > 0, \quad n > 0$$

$$E(X) = n; \ \mathrm{Var}(X) = 2n; \ M(s) = (1 - 2s)^{-\frac{n}{2}}, s < \frac{1}{2}$$

Propriedades:

$X \sim \chi^2(n) \Leftrightarrow X \sim \mathrm{G}\left(\frac{n}{2}; \frac{1}{2}\right)$

$\sqrt{2\chi^2(n)} - \sqrt{2n-1} \overset{a}{\sim} N(0,1)$ ou $\sqrt{2\chi^2(n)} - \sqrt{2n} \overset{a}{\sim} N(0,1)$

$X \sim \mathrm{G}(n; \lambda) \Leftrightarrow 2\lambda X \sim \chi^2(2n)$

$X \sim N(0,1) \Rightarrow Y = X^2 \sim \chi^2(1)$

X_i indep., $i = 1,2,...,n$

$X_i \sim \chi^2(n_i) \Rightarrow \sum_{i=1}^{n} X_i \sim \chi^2(r)$ com $r = \sum_{i=1}^{n} n_i$

$X_i \sim N(0,1) \Rightarrow \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$

- ***t*-“STUDENT”** $T \sim t(n)$

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

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

Propriedades: simetria em torno de 0

$\lim_{n \to \infty} f(t) = \phi(t)$

$T \sim t(n) \Rightarrow T^2 \sim F(1, n)$

- ***F*-SNEDCOR** $F \sim F(m,n)$

  $U \sim \chi^2(m)$, $V \sim \chi^2(n)$ e U,V indep. $\Rightarrow F = \dfrac{U/m}{V/n} \sim F(m,n)$

  $E(X) = \dfrac{n}{n-2}$ $(n > 2)$; $\text{Var}(X) = \dfrac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ $(n > 4)$

  Propriedades: $X \sim F(m,n) \Leftrightarrow \dfrac{1}{X} \sim F(n,m)$

- **DISTRIBUIÇÕES POR AMOSTRAGEM**
  - **<u>Distribuições do mínimo e do máximo</u>**

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

    Estatística de ordem $v \rightarrow$ $G_V(y) = \sum_{j=v}^{n} \binom{n}{j} [F(y)]^j \, [1 - F(y)]^{n-j}$

    $$g_V(y) = \frac{n!}{(v-1)!\,(n-v)!} [F(y)]^{v-1} \, [1 - F(y)]^{n-v} \, f(y)$$

    Estatísticas de ordem u e v ($u < v$)

    $$g_{U,V}(y,z) = \frac{n!}{(u-1)!\,(v-u-1)!\,(n-v)!} [F(y)]^{u-1} [F(z) - F(y)]^{v-u-1} [1 - F(z)]^{n-v} f(y) \, f(z), \quad y < z$$

  - **<u>Média e variância amostrais</u>**

    $E(\overline{X}) = \mu$; $\text{Var}(\overline{X}) = \dfrac{\sigma^2}{n}$; $E(S^2) = \dfrac{n-1}{n}\sigma^2$; $E(S'^2) = \sigma^2$ $\text{var}(S'^2) = \dfrac{1}{n}\left(\mu_4 - \dfrac{n-3}{n-1}\mu_2^2\right)$

  - **<u>Teorema do limite central e corolários</u>**

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

    Binomial: $X_i \sim B(1;\theta)$ e X_i indep. $\Rightarrow \dfrac{\sum_{i=1}^{n} X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \overset{a}{\sim} N(0,1)$

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

    Poisson: $X \sim \text{Po}(\lambda)$ e $\lambda \rightarrow +\infty \Rightarrow \dfrac{X - \lambda}{\sqrt{\lambda}} \overset{a}{\sim} N(0,1)$

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

- **distribuições por amostragem – 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	Variâncias conhecidas $\dfrac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} \sim N(0,1)$	Variâncias desconhecidas mas iguais $\dfrac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}\sqrt{\dfrac{(m-1)S_1'^2+(n-1)S_2'^2}{m+n-2}}} \sim t(m+n-2)$
	Variâncias desconhecidas $Z=\dfrac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1'^2}{m}+\dfrac{S_2'^2}{n}}} \sim t(r)$	onde r é a parte inteira de $\dfrac{\left(\dfrac{s_1'^2}{m}+\dfrac{s_2'^2}{n}\right)^2}{\dfrac{1}{m-1}\left(\dfrac{s_1'^2}{m}\right)^2+\dfrac{1}{n-1}\left(\dfrac{s_2'^2}{n}\right)^2}$
Variância	$\dfrac{nS^2}{\sigma^2}=\dfrac{(n-1)S'^2}{\sigma^2} \sim \chi^2(n-1)$	
Quociente de variâncias	$\dfrac{S_1'^2}{S_2'^2}\dfrac{\sigma_2^2}{\sigma_1^2} \sim F(m-1,n-1)$ ou	$\dfrac{S_2'^2}{S_1'^2}\dfrac{\sigma_1^2}{\sigma_2^2} \sim F(n-1,m-1)$

- **distribuições por amostragem – 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	$\dfrac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{m}+\dfrac{\sigma_2^2}{n}}} \overset{a}{\sim} N(0,1)$	$\dfrac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_1'^2}{m}+\dfrac{S_2'^2}{n}}} \overset{a}{\sim} N(0,1)$

- **INFERÊNCIA PARAMÉTRICA**

  - Distribuição assimptótica dos estimadores de MV $\rightarrow$ $\sqrt{n\Im_X(\theta)}(\hat{\theta}-\theta) \overset{a}{\sim} n(0,1)$

    podendo utilizar-se também $\sqrt{n\Im_X(\hat{\theta})}(\hat{\theta}-\theta) \overset{a}{\sim} n(0,1)$

  - Critério da factorização - $T(X_1,...,X_n)$ estatística suficiente para θ

    $$\Leftrightarrow f(x_1,...,x_n\,|\,\theta)=h(x_1,...,x_n)\times g\left[T(x_1,...,x_n)\,|\,\theta\right]$$

  - Teorema de Lehmann e Scheffé – Estatísticas suficientes mínimas

    $\dfrac{f(x_1,x_2,...,x_n;\theta)}{f(y_1,y_2,...,y_n;\theta)}$ constante em θ se e só se $T(x_1,x_2,...,x_n)=T(y_1,y_2,...,y_n)$

  - T estimador de θ

$$EQM(T) = \text{var}(T) + (E(T) - \theta)^2$$

Cond. Nec. e Suf. de consistênca em m.q.: $\lim_{n\to\infty} E(T) = \theta$ e $\lim_{n\to\infty} \text{var}(T) = 0$

- **Quantidade de informação de Fisher**

$$\Im_X(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\left(\ln f(X;\theta)\right)\right)^2\right] = -E\left[\left(\frac{\partial^2}{\partial\theta^2}\left(\ln f(X;\theta)\right)\right)\right]$$

$$\Im_{X_1,X_2,\ldots,X_n}(\theta) = n\Im_X(\theta) = nE\left[\left(\frac{\partial}{\partial\theta}\left(\ln f(X;\theta)\right)\right)^2\right] = -nE\left[\left(\frac{\partial^2}{\partial\theta^2}\left(\ln f(X;\theta)\right)\right)\right]$$

- **Desigualdade de Frechet-Cramer-Rao (FCR)**

O universo verifica as condições de regularidade

$\forall T$ estimador centrado para $\tau(\theta)$, $\text{var}(T) \geq \dfrac{\left(\tau'(\theta)\right)^2}{n\Im_X(\theta)}$.

- **Lema de Neyman-Pearson**

W – região de rejeição

- $\dfrac{L(\theta_1)}{L(\theta_0)} \geq c$ para $(x_1, x_2, \ldots, x_n) \in W$
- $\dfrac{L(\theta_0)}{L(\theta_1)} \geq c$ para $(x_1, x_2, \ldots, x_n) \notin W$
- $\Pr\left((X_1, X_2, \ldots, X_n) \in W \mid H_0\right) = \alpha$

- **RVM e teorema de Karlin-Rubin**

W – região de rejeição

- Razão de Verosimilhanças Monótona – Diz-se que a família $\{f(x\mid\theta) : \theta \in \Theta\}$, sendo Θ um intervalo aberto de $\Re$, tem RVM na estatística T se e só se $\dfrac{L(\theta_2)}{L(\theta_1)}$, $\theta_2 > \theta_1$, é função monótona de $t = T(x_1, x_2, \cdots, x_n)$.
- Teorema de Karlin-Rubin – Se $(X_1, X_2, \cdots, X_n)$ é amostra casual simples de população com função densidade (probabilidade) com RVM crescente na estatística *T*, então, para testar $H_0: \theta \leq \theta_0$ contra $H_1: \theta > \theta_0$, os testes com região de rejeição do tipo $T(x_1, x_2, \cdots, x_n) > t_0$ onde t_0 é determinado em função da dimensão desejada, são UMP.

	Crescente	Decrescente
$H_0: \theta \leq \theta_0$	$> t_0$	$< t_0$
$H_0: \theta \geq \theta_0$	$< t_0$	$> t_0$

- **ANOVA**
- Um critério

$$\text{SQT} = \text{SQD} + \text{SQE} \qquad \text{SQT} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}_{\circ\circ}\right)^2$$

$$\text{SQD} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}_{i\circ}\right)^2 \qquad \text{SQE} = \sum_{i=1}^{m} n_i\left(\overline{X}_{i\circ} - \overline{X}_{\circ\circ}\right)^2.$$

$$\overline{X}_{i\circ} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \quad (i = 1,2,\ldots,m) \qquad \overline{X}_{\circ\circ} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} X_{ij}}{\sum_{i=1}^{m} n_i} = \frac{\sum_{i=1}^{m} n_i\,\overline{X}_{i\circ}}{n}$$

$$\frac{SQE/(m-1)}{SQD/(n-m)} \sim F(m-1, n-m)$$

- Dois critérios (1 obs por célula)

SQT = SQL + SQC + SQR

SQT = $\sum_{i=1}^{m}\sum_{j=1}^{n}\left(X_{ij} - \overline{X}_{\circ\circ}\right)^2$ SQL = $\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\overline{X}_{i\circ} - \overline{X}_{\circ\circ}\right)^2$,

SQC = $\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\overline{X}_{\circ j} - \overline{X}_{\circ\circ}\right)^2$ SQR = $\sum_{i=1}^{m}\sum_{j=1}^{n}\left(X_{ij} - \overline{X}_{i\circ} - \overline{X}_{\circ j} + \overline{X}_{\circ\circ}\right)^2$

.

$$\overline{X}_{i\circ} = \frac{1}{n}\sum_{j=1}^{n} X_{ij}, \quad \overline{X}_{\circ j} = \frac{1}{m}\sum_{i=1}^{m} X_{ij}$$

$$\overline{X}_{\circ\circ} = \frac{1}{m}\sum_{i=1}^{m}\overline{X}_{i\circ} = \frac{1}{n}\sum_{j=1}^{n}\overline{X}_{\circ j} = \frac{1}{m\,n}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}$$

$$\frac{SQL/(m-1)}{SQR/\left((m-1)(n-1)\right)} \sim F(m-1, (m-1)(n-1))$$

$$\frac{SQC/(n-1)}{SQR/\left((m-1)(n-1)\right)} \sim F(n-1, (m-1)(n-1))$$

- **INFERÊNCIA NÃO PARAMÉTRICA**
- **<u>Teste do χ^2 à bondade do ajustamento</u>**

$$Q = \sum_{j=1}^{m}\frac{(N_j - n\,p_{oj})^2}{n\,p_{oj}} \overset{a}{\sim} \chi^2(m-1-k)$$

(m - nº de classes; N_j - freq. da classe j na amostra; k - nº parâmetros a estimar)

- **<u>Testes de Kolmogorov-Smirnov</u>**

$$D_n^+ = \max\left(\max_{i=1,2,\ldots n}\left[i/n - F_0(X_{(i)});0\right]\right)$$

$$D_n^- = \max\left(\max_{i=1,2,\ldots n}\left[F_0(X_{(i)}) - (i-1)/n;0\right]\right)$$

$$D_n = \max(D_n^+, D_n^-)$$

- **<u>Tabelas de contingência – independência e homogeneidade</u>**

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

(r- nº de classes de A; s- nº de classes de B; N_{ij} - freq. da classe i,j; $fe_{ij} = \frac{N_{i\circ}N_{\circ j}}{n}$ - frequência esperada da classe i, j)

Medidas de Associação

Coeficiente de contingência de Pearson: $C = \sqrt{Q/(Q+n)}$

Coeficiente de Tschuprow: $T = \sqrt{\dfrac{Q}{n\sqrt{(r-1)(s-1)}}}$

Coeficiente de Cramer: $V = \sqrt{\dfrac{Q}{n(q-1)}}$ $\qquad q = \min(r,s)$

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

(r- nº de classes consideradas; s- nº de sub-populações; N_{ij} - freq. da classe i da sub-população j na amostra; $fe_{ij} = N_{.j}\,\hat{p}_{i|j}$ - frequência esperada da classe i da sub-população j)

- <u>**Teste do sinal**</u>

  Recorrer à distribuição binomial

- <u>**Teste de Wilcoxon**</u>

  $S = \sum_{i=1}^{n} S_i \qquad\qquad T = \sum_{i=1}^{n} i\,S_i$

  Pequenas amostras $\rightarrow$ Tabelas

  Grandes amostras $\rightarrow$ $T^* = \dfrac{T - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}} \overset{\circ}{\sim} n(0;1)$