Formulário de Estatística

Fórmula de *Sturges*	$K = 1 + 3.322 \times \log_{10}(n) = 1 + \log_2(n)$, $K = \text{nº classes},\ n = \text{nº observações}$
Quantis de ordem *p*	$q_p = \begin{cases} x_{(w)}, & w \text{ inteiro} \\ x_{(i)} + (w-i)(x_{(i+1)} - x_{(i)}), & w, \text{ não inteiro} : i < w < i+1 \end{cases}$ $w = p(n+1)$. Se p=1/4, 2/4,3/4 temos os quartis; 1/10,…, 9/10 temos os decis.
Mediana (dados agrupados em classes)	$Med = \widetilde{X} = \mathrm{L}_i + \dfrac{\frac{n}{2} - \mathrm{F}_i}{\mathrm{f}_i}(\mathrm{U}_i - \mathrm{L}_i)$ $n = \text{nº observações}$ $[\mathrm{L}_i, \mathrm{U}_i[$ é a classe mediana (a que contém a observação n/2) f_i, F_i = freq. da classe i e freq. acumulada até à classe $i-1$
Moda (dados agrupados em classes)	$Moda = \mathrm{L}_i + \dfrac{\mathrm{f}_i - \mathrm{f}_{i-1}}{(\mathrm{f}_i - \mathrm{f}_{i-1}) + (\mathrm{f}_i - \mathrm{f}_{i+1})}(\mathrm{U}_i - \mathrm{L}_i)$ classe modal $= [\mathrm{L}_i, \mathrm{U}_i[$ f_i = frequência da classe i
Variância amostral (dados agrupados em classes)	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \dfrac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n.\overline{X}^2\right]$ $S^2 = \dfrac{1}{n-1}\left[\sum_{i=1}^{K} f_i X'^2_i - n.\overline{X}^2\right] - \dfrac{\Delta x^2}{12}$; onde $\overline{X} = \sum_{i=1}^{K} f_i X'_i$
Coeficiente de correlação	$r = \dfrac{\left[\sum_{i=1}^{n} x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}\right]}{\sqrt{\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum x_i)^2}{n}\right]\left[\sum_{i=1}^{n} y_i^2 - \frac{(\sum y_i)^2}{n}\right]}}$
Distribuição Binomial B(n , p)	$\mathrm{P}(\mathrm{X} = k) = C_k^n p^k (1-p)^{n-k}$, $k = 0,1,2,\dots,n$ E(X)= np, Var(X)= $np(1-p)$.
Distribuição Binomial Negativa BN(r , p) ---------------------------- BN(r=1,p) ⇔ G(p)	$\mathrm{P}(\mathrm{X} = k) = C_{r-1}^{k-1} p^r (1-p)^{k-r}$, $k = r, r+1, r+2, \dots$ E(X)= $\frac{r}{p}$, Var(X)= $\frac{r(1-p)}{p^2}$.
Distribuição Hiper-Geométrica H(N, D , n)	$\mathrm{P}(\mathrm{X} = k) = \dfrac{C_k^D C_{n-k}^{N-D}}{C_n^N}$, $k = máx(0, n-N+D), \dots, \min(n, D)$ E(X) $= np$, $\mathrm{Var}(X) = np(1-p)\frac{N-n}{N-1}$, com $p = \frac{D}{N}$.
Distribuição de Poisson P(λ)	$\mathrm{P}(\mathrm{X} = k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$, $k = 0,1,2,\dots; \lambda > 0$ E(X) = Var(X) = λ.

Distribuição Uniforme U(a , b)	$f(x)=\begin{cases}\frac{1}{b-a} & , \ a\le x\le b\\ 0 & , \ \_\_\_\_\_\_\_\end{cases}$ $\quad$ $E(X)=\frac{a+b}{2}$, $Var(X)=\frac{(b-a)^2}{12}$.
Distribuição Normal $N(\mu\ ;\ \sigma^2)$	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right\}$, $\quad -\infty<x,\mu<+\infty$, $\sigma>0$ $E(X)=\mu$, $Var(X)=\sigma^2$.
Distribuição Exponencial E(λ)	$f(x)=\begin{cases}\lambda e^{-\lambda x} & , x>0\\ 0 & , \ \_\_\_\_\_\_\_\end{cases}$ $\quad$ $E(X)=\frac{1}{\lambda}$, $Var(X)=\frac{1}{\lambda^2}$.
IC para a Média (σ^2 conhecida) (população normal)	$\left(\bar{x}-z_{(1-\alpha/2)}\cdot\frac{\sigma}{\sqrt{n}}\ ;\ \bar{x}+z_{(1-\alpha/2)}\cdot\frac{\sigma}{\sqrt{n}}\right)$; $\quad Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$
IC Média (σ^2 desconhecida) (população normal)	$\left(\bar{x}-t_{(1-\alpha/2)}\cdot\frac{s}{\sqrt{n}}\ ;\ \bar{x}+t_{(1-\alpha/2)}\cdot\frac{s}{\sqrt{n}}\right)$; $\quad T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{n-1}$
IC para a Diferença de Médias com variâncias conhecidas. (populações normais)	$(\bar{x}_1-\bar{x}_2)\pm z_{(1-\alpha/2)}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$; $\quad Z_D=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$
IC para a Diferença de Médias com variâncias desconhecidas. (populações normais)	$(\bar{x}_1-\bar{x}_2)\pm t_{(1-\alpha/2)}\cdot s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$, com $\sigma_1^2=\sigma_2^2$; $T_D^*=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_p^2}{n_1}+\frac{S_p^2}{n_2}}}\sim t_{n_1+n_2-2}$ $(\bar{x}_1-\bar{x}_2)\pm t'_{(1-\alpha/2)}\cdot\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$, com $\sigma_1^2\neq\sigma_2^2$; $\quad t'_{(1-\alpha/2)}=\frac{w_1 t_1+w_2 t_2}{w_1+w_2}$; $S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ $\quad w_1=S_1^2/n_1$; $w_2=S_2^2/n_2$; $t_1=t_{(1-\alpha/2,n_1-1)}$; $t_2=t_{(1-\alpha/2,n_2-1)}$
IC para AMOSTRAS GRANDES e populações que não se sabe se são normais:	Nos cinco intervalos acima usa-se sempre a tabela da normal, e não da t-student, para encontrar a constante; o resto mantém-se.
I.C. (aproximado) para a proporção	$\left(\hat{p}-z_{(1-\alpha/2)}\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ;\hat{p}+z_{(1-\alpha/2)}\cdot\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$; $\quad Z=\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\sim N(0,1)$
I.C. (aproximado) para a diferença de proporções	$(\hat{p}_1-\hat{p}_2)\pm z_{(1-\alpha/2)}\cdot\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$; $Z_D=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}\sim N(0,1)$ NOTA:Em testes, na raiz, usar $\hat{p}=(n_1\hat{p}_1+n_2\hat{p}_2)/(n_1+n_2)$ em vez de $\hat{p}_1$, $\hat{p}_2$
IC para a Variância (população normal)	$\left(\frac{(n-1)S^2}{b};\frac{(n-1)S^2}{a}\right)$; $\quad X^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{(n-1)}$ $a=\chi^2_{(\alpha/2;n-1)}$; $b=\chi^2_{(1-\alpha/2;n-1)}$
IC para a Razão de Variâncias $\frac{\sigma_1^2}{\sigma_2^2}$ (populações normais)	$\left(\frac{S_1^2}{b.S_2^2};\frac{S_1^2}{a.S_2^2}\right)$; $\quad F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\frac{\sigma_2^2}{\sigma_1^2}\frac{S_1^2}{S_2^2}\sim F_{(n_1-1,n_2-1)}$ $b=F_{(1-\alpha/2;n_1-1,n_2-1)}$, $a=F_{(\alpha/2;n_1-1,n_2-1)}=1/F_{(1-\alpha/2;n_2-1,n_1-1)}$