Fórmulas Estatística I

- Média amostral: $\bar{X} = \frac{\sum\limits_{i=1}^{n} X_i}{n}$

- Variância amostral: $\hat{\sigma^2} = \frac{\sum\limits_{i=1}^{n}(X_i - \bar{X})^2}{n}$ ou $S^2 = \frac{\sum\limits_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$

- Coeficiente de variação: $CV = (S/\bar{X}).100\%$

- Probabilidade Condicional: $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$

- Esperança: $\mu = E(X) = \sum\limits_{x:f(x)>0} xf(x)$ ou $\int\limits_{-\infty}^{\infty} xf(x)dx$

- Variância: $\sigma^2 = Var(X) = E[(X-\mu)^2] = E(X^2) - \mu^2$

- Propriedade da esperança: $E[g(X)] = \sum\limits_{x:f(x)>0} g(x)f(x)$ ou $\int\limits_{-\infty}^{\infty} g(x)f(x)dx$

- Distribuição Binomial: $X \sim Bin(n,p)$
  $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, 2, ..., n$
  $E(X) = np$ e $Var(X) = np(1-p)$

- Distribuição Hipergeométrica: $X \sim Hiperg(N,n,k)$
  $f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$, $max\{0, n-(N-k)\} \leq x \leq min\{n,k\}$
  $E(X) = \frac{nk}{N}$ e $Var(X) = \frac{N-n}{N-1}.n.\frac{k}{N}\left(1 - \frac{k}{N}\right)$

- Distribuição Binomial Negativa: $X \sim BN(k,p)$
  $f(x) = \binom{x-1}{k-1} p^k q^{x-k}$, $x = k, k+1, k+2, ...$
  $E(X) = \frac{k}{p}$ e $Var(X) = \frac{k(1-p)}{p^2}$

- Distribuição Geométrica: $X \sim Geom(p)$
  $f(x) = p(1-p)^{x-1}$, $x = 1, 2, 3, ...$
  $E(X) = \frac{1}{p}$ e $Var(X) = \frac{1-p}{p^2}$
  notação alternativa: $X \sim Geom(p)$
  $f(x) = p(1-p)^x$, $x = 0, 1, 2, 3, ...$
  $E(X) = \frac{1-p}{p}$ e $Var(X) = \frac{1-p}{p^2}$

- Distribuição de Poisson: $X \sim Poisson(\lambda)$
  $f(x) = \frac{\exp^{-\lambda}(\lambda)^x}{x!}$, $x = 0, 1, 2, ...$
  $E(X) = \lambda$ e $Var(X) = \lambda$

- Distribuição Uniforme contínua: $X \sim Uniforme(A, B)$

  $f(x) = \frac{1}{B-A}, \quad A \leq x \leq B$
  $E(X) = \frac{A+B}{2}$ e $Var(X) = \frac{(B-A)^2}{12}$

- Distribuição Exponencial: $X \sim Exp(\lambda)$

  $f(x) = \lambda \exp^{-\lambda x}, \quad x > 0$
  $E(X) = \frac{1}{\lambda}$ e $Var(X) = \frac{1}{\lambda^2}$

- Distribuição Normal: $X \sim N(\mu, \sigma^2)$

  $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty \leq x \leq \infty$
  $E(X) = \mu$ e $Var(X) = \sigma^2$

- Transformação da Normal para a Normal Reduzida:
  $Z = \frac{X-\mu}{\sigma}$

- Teorema Central do Limite: Sejam $X_1, ..., X_n$ v.a. independentes e identicamente distribuídas, com $E(X_i) = \mu$ e $Var(X_i) = \sigma^2 < \infty$, para $i = 1, ..., n$, então, quando $n \to \infty$:
  $\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$, ou seja: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \approx N(0, 1)$

  Casos particulares e observações:

  - se $X_i \sim Bernoulli(p)$, para $i = 1, ..., n$: $\hat{p} \approx N(p, \frac{p(1-p)}{n})$
  - se $Y \sim Binomial(n, p)$: $Y \approx N(np, np(1-p))$: aproximação da Binomial pela Normal
  - se σ desconhecido, pode-se usar adicionalmente o Teorema de Slutsky e: $\frac{\bar{X}-\mu}{S/\sqrt{n}} \approx N(0, 1)$

- Sejam $X_1, ..., X_n$ v.a. independentes e identicamente distribuídas, com $X_i \sim N(0, 1)$, para $i = 1, ..., n$, então: $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t - Student(n-1)$

- Função de verossimilhança: $L(\theta) = f(x_1, x_2, ..., x_n/\theta) = \prod_{i=1}^{n} f(x_i)$, sendo a última igualdade decorrente da suposição de independência de $X_1, ..., X_n$

- $\hat{\theta}$ é estimador não viciado de θ se: $E(\hat{\theta}) = \theta$, para todo valor de θ.

- $\hat{\theta}$ é estimador consistente de θ se: $\lim_{n\to\infty} E(\hat{\theta}) = \theta$ e $\lim_{n\to\infty} V(\hat{\theta}) = 0$

Tabela 1
Intervalos de confiança frequentes
1. Para apenas um parâmetro

Para a média μ	
Caso	*Intervalo*
1. σ é conhecida e X tem distribuição normal ou o tamanho da amostra n é suficientemente grande	$\bar{X} \pm z_{1-\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
2. σ é desconhecida e X tem distribuição normal	$\bar{X} \pm t_{1-\alpha/2} \dfrac{S}{\sqrt{n}}$ la t tem $n-1$ g.l.
3. σ é desconhecida e o tamanho de amostra n é suficientemente grande	$\bar{X} \pm z_{1-\alpha/2} \dfrac{S}{\sqrt{n}}$
Para a variancia σ^2	
Caso	*Intervalo*
X tem distribuição normal	$\left[\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}}, \dfrac{(n-1)S^2}{\chi^2_{\alpha/2}}\right]$ la χ^2 tem $n-1$ g.l.
Para a proporção p	
Caso	*Intervalo*
O tamanho da amostra n é suficientemente grande	$\bar{p} \pm z_{1-\alpha/2} \sqrt{\dfrac{\bar{p}\bar{q}}{n}}$

2. Para os parâmetros

Para a diferença de médias $\mu_1 - \mu_2$	
Caso	*Intervalo*
1. σ_1 e σ_2 são conhecidas, as amostras são independentes e cada uma das populações tem distribuição normal ou tamanhos de amostra n_i suficientemente grandes	$\bar{X} - \bar{Y} \pm z_{1-\alpha/2} \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$
2. σ_1 e σ_2 são desconhecidas porém iguais, as amostras são independentes e as populações tem distribuição normal	$\bar{X} - \bar{Y} \pm t_{1-\alpha/2} \sqrt{S_p^2 \left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$ Con $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ e a t con $n_1 + n_2 - 2$ g.l.
3. σ_1 e σ_2 são desconhecidas porém diferentes, as amostras são independentes e as populações tem distribuição normal	$\bar{X} - \bar{Y} \pm t_{1-\alpha/2} \sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}$ Os graus de liberdade são dados por v, onde $v = \dfrac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{(S_1^2/n_1)^2}{n_1+1} + \frac{(S_2^2/n_2)^2}{n_2+1}} - 2$
4. σ_1 e σ_2 são desconhecidas, as amostras são independentes e têm tamanhos de amostra suficientemente grandes	$\bar{X} - \bar{Y} \pm z_{1-\alpha/2} \sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}$
Para o quociente de variâncias σ_1^2/σ_2^2	
Caso	*Intervalo*
As amostras são independentes e as populações têm distribuição normal	$\left[\dfrac{S_1^2}{S_2^2 F_{1-\alpha/2}}, \dfrac{S_1^2}{S_2^2 F_{\alpha/2}}\right]$ a F tem $n_1 - 1$ e $n_2 - 1$ g.l.
Para a diferença de proporções $p_1 - p_2$	
Caso	*Intervalo*
As amostras são independentes e com tamanhos suficientemente grandes	$\bar{p}_1 - \bar{p}_2 \pm z_{1-\alpha/2} \sqrt{\dfrac{\bar{p}_1\bar{q}_1}{n_1} + \dfrac{\bar{p}_2\bar{q}_2}{n_2}}$

Tabela 2
Testes de Hipótese frequentes
1. Para apenas um parâmetro

H_1	Caso	Estatística de teste	Rejeitar H_0 se:
$H_0 : \mu = \mu_0$			
$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	1. σ é conhecida e X tem distribuição normal e o tamanho de amostra n é suficientemente grande	$Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$Z > z_{1-\alpha}$ $Z < z_\alpha = -z_{1-\alpha}$ $\|Z\| > z_{1-\alpha/2}$
$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	2. σ é desconhecida e X tem distribuição normal	$T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$ la t tem $n-1$ g.l.	$T > t_{1-\alpha}$ $T < t_\alpha = -t_{1-\alpha}$ $\|T\| > t_{1-\alpha/2}$
$\mu > \mu_0$ $\mu < \mu_0$ $\mu \neq \mu_0$	3. σ é desconhecida e o tamanho de amostra n é suficientemente grande	$Z = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$	$Z > z_{1-\alpha}$ $Z < z_\alpha = -z_{1-\alpha}$ $\|Z\| > z_{1-\alpha/2}$
$H_0 : \sigma^2 = \sigma_0^2$			
$\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$ $\sigma^2 \neq \sigma_0^2$	X tem distribuição normal	$\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$ la χ^2 tem $n-1$ g.l.	$\chi^2 > \chi^2_{1-\alpha}$ $\chi^2 < \chi^2_\alpha$ $\chi^2 < \chi^2_{\alpha/2}$ o $\chi^2 > \chi^2_{1-\alpha/2}$
$H_0 : p = p_0$			
$p > p_0$ $p < p_0$ $p \neq p_0$	O tamanho de amostra n é suficientemente grande	$Z = \dfrac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$	$Z > z_{1-\alpha}$ $Z < z_\alpha = -z_{1-\alpha}$ $\|Z\| > z_{1-\alpha/2}$

2. Sobre os parâmetros

H_1	Caso	Estatística de teste	Rejeitar H_0 se:
$H_0 : \mu_1 = \mu_2$			
$\mu_1 - \mu_2 > 0$ $\mu_1 - \mu_2 < 0$ $\mu_1 - \mu_2 \neq 0$	1. σ_1 e σ_2 são conhecidas, as amostras são independentes e cada uma das populações tem distribuição normal ou os tamanhos de amostra n_i são suficientemente grandes	$Z = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$Z > z_{1-\alpha}$ $Z < z_\alpha = -z_{1-\alpha}$ $\|Z\| > z_{1-\alpha/2}$
$\mu_1 - \mu_2 > 0$ $\mu_1 - \mu_2 < 0$ $\mu_1 - \mu_2 \neq 0$	2. σ_1 e σ_2 são desconhecidas porém iguais, as amostras são independentes e as populações têm distribuição normal	$T = \dfrac{\bar{X} - \bar{Y}}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ Con $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ e a t é com n_1+n_2-2 g.l.	$T > t_{1-\alpha}$ $T < t_\alpha = -t_{1-\alpha}$ $\|T\| > t_{1-\alpha/2}$
$\mu_1 - \mu_2 > 0$ $\mu_1 - \mu_2 < 0$ $\mu_1 - \mu_2 \neq 0$	3. σ_1 e σ_2 são desconhecidas porém diferentes, as amostras são independentes e as populações têm distribuição normal	$T = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ e a t é com v g.l. con $v = \dfrac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{(S_1^2/n_1)^2}{n_1+1} + \frac{(S_2^2/n_2)^2}{n_2+1}} - 2$	$T > t_{1-\alpha}$ $T < t_\alpha = -t_{1-\alpha}$ $\|T\| > t_{1-\alpha/2}$
$\mu_1 - \mu_2 > 0$ $\mu_1 - \mu_2 < 0$ $\mu_1 - \mu_2 \neq 0$	4. σ_1 e σ_2 são desconhecidas, as amostras são independentes e têm tamanhos suficientemente grandes	$Z = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$Z > z_{1-\alpha}$ $Z < z_\alpha = -z_{1-\alpha}$ $\|Z\| > z_{1-\alpha/2}$
$H_0 : \sigma_1^2 = \sigma_2^2$			
$\sigma_1^2 > \sigma_2^2$ $\sigma_1^2 < \sigma_2^2$ $\sigma_1^2 \neq \sigma_2^2$	As amostras são independentes e as populações têm distribuição normal	$F = \dfrac{S_1^2}{S_2^2}$ e a F é com n_1-1 e n_2-1 g.l.	$F > F_{1-\alpha}$ $F < F_\alpha$ $F < F_{\alpha/2}$ o $F > F_{1-\alpha/2}$
$H_0 : p_1 = p_2$			
$p_1 > p_2$ $p_1 < p_2$ $p_1 \neq p_2$	As amostras são independentes e os tamanhos de amostra n_i são suficientemente grandes	$Z = \dfrac{\bar{p}_1 - \bar{p}_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ con $\bar{p} = (n_1\bar{p}_1 + n_2\bar{p}_2)/(n_1+n_2)$	$Z > z_{1-\alpha}$ $Z < z_\alpha = -z_{1-\alpha}$ $\|Z\| > z_{1-\alpha/2}$