MATCP_Formulário_Estatística

1. Distribuições Discretas

Quadro de Distribuição

x_i	x_1	x_2	$\cdots$	x_n	$\cdots$
$P(X=x_i)=p_i$	p_1	p_2	$\cdots$	p_n	

$P_1)\, p_i \geq 0, \quad i=1,2,3,\ldots$

$P_2)\, \sum_i p_i = 1$

Função de Probabilidade	**Função de distribuição**
$f(x)=P(X=x)=\begin{cases} 0\,, & x \neq x_i \\ p_i\,, & x=x_i \end{cases}$	$F(x)=P(X\leq x)=\sum_{x_i \leq x} f(x_i)$
Esperança matemática ou valor médio $\mu=E(X)=\sum_i x_i f(x_i)=\sum_i x_i p_i$	**Variância e desvio padrão** $\sigma^2=V(X)=E[(X-\mu)^2]=E(X^2)-\mu^2$ $=\sum_i x_i^2 f(x_i)-\mu^2 \quad e \quad \sigma=+\sqrt{V(X)}$

1.1 Distribuição Binomial - $X \sim B_i(n,p)$

Função de Probabilidade	**Função de distribuição**
$f(x)=\begin{cases} 0 & , \quad x \neq 0,1,2,\ldots,n \\ C_x^{\ n} p^x q^{n-x}, & x=0,1,2,\ldots,n \end{cases}$ $onde\; q=1-p \; e \quad C_x^{\ n}=\dfrac{n!}{p!(n-p)!}$	$F(x)=P(X\leq x)=\sum_{x_i\leq x} C_{x_i}^{\ n} p^{x_i} q^{n-x_i}$
Esperança matemática ou valor médio $\mu=E(X)=\sum_i x_i f(x_i)=np$	**Variância e desvio padrão** $\sigma^2=V(X)=E[(X-\mu)^2]=npq$ $\sigma=+\sqrt{npq}$

1.2 Distribuição de Poisson - $X \sim P_o(\mu)$

Função de Probabilidade	**Função de distribuição**
$f(x)=\begin{cases} 0 & , \quad x \neq 0,1,2,\ldots,n,\ldots \\ e^{-\mu}\dfrac{\mu^x}{x!}, & x=0,1,2,\ldots,\ldots \end{cases}$	$F(x)=P(X\leq x)=\sum_{x_i\leq x} e^{-\mu}\dfrac{\mu^{x_i}}{x_i!}$
Esperança matemática ou valor médio $\mu=E(X)=\sum_i x_i f(x_i)=\mu$	**Variância** e **desvio padrão** $\sigma^2=V(X)=E[(X-\mu)^2]=\mu$ $\sigma=+\sqrt{\mu}$

2. Distribuições Contínuas

Função densidade de probabilidade	**Função de distribuição**
$f(x)\ é\ f.d.p\ sse \begin{cases} f(x) \geq 0 & x \in \sim \\ \int_{-\infty}^{+\infty} f(x)dx = 1 \end{cases}$ $P(a < X < b) = P(a \leq X \leq b) = \int_a^b f(x)dx \quad (b \geq a)$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$
Esperança matemática ou valor médio $\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx$	**Variância e desvio padrão** $\sigma^2 = V(X) = E[(X-\mu)^2] = E(X^2) - \mu^2$ $= \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2 \quad e \quad \sigma = +\sqrt{V(X)}$

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

Função densidade de probabilidade	**Função de distribuição**
$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2} \quad \forall x \in R$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{y-\mu}{\sigma}\right)^2} dy \approx \Phi\left(\dfrac{x-\mu}{\sigma}\right)$
Esperança matemática ou valor médio $E(X) = \mu$	**Variância e desvio padrão** $V(X) = \sigma^2$
$P(X \leq a) = \Phi\left(\dfrac{a-\mu}{\sigma}\right)$	$P(a < X < b) = \Phi\left(\dfrac{b-\mu}{\sigma}\right) - \Phi\left(\dfrac{a-\mu}{\sigma}\right)$

Aditividade da Distribuição Normal

Se $\{X_1, X_2, ..., X_n\}$ são v.a's independentes tais que $X_i \sim N(\mu_i, \sigma_i^2), i = 1,2,...,n$ então

$$Y = \sum_{i=1}^{n} a_i X_i \sim N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2), \quad a_i \in \sim, i = 1,2,...,n$$

Casos particulares

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

2) $X_1 \sim N(\mu_1, \sigma_1^2)$ e $X_2 \sim N(\mu_2, \sigma_2^2)$ então $X_1 \pm X_2 \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$

3) Se $X_i \sim N(\mu, \sigma^2), i = 1,2,...,n$ então $Y = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$

Teorema do Limite Central

Se $\left\{X_1, X_2, ..., X_n\right\}$ são v.a's independentes tais que $E(X_i) = \mu_i, V(X_i) = \sigma_i^2,\ i = 1,2,3,...,n$ e

então $Y = \sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{D.} N\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right)$

Obs: É considerada uma boa aproximação para $n \geq 30$

Amostragem

Distribuição da média amostral e diferença de médias amostrais	**Distribuição da proporção amostral e diferença de proporções amostrais**
$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim N(\mu; \frac{\sigma^2}{n})$, $n \geq 30$	$\hat{P} = \frac{X}{n} \sim N(p; \frac{pq}{n})$ $n \geq 30$
$\overline{X}_1 - \overline{X}_2 \sim N(\mu_1 - \mu_2; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$, $n_1 \geq 30$ $n_2 \geq 30$	$\hat{P}_1 - \hat{P}_2 \sim N\left(p_1 - p_2; \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$, $n_1 \geq 30$ $n_2 \geq 30$

Média e variância de uma amostra (dados classificados)

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i n_i}{n} \qquad s^2 = \frac{n\sum_{i=1}^{k} x_i^2 n_i - \left(\sum_{i=1}^{k} x_i n_i\right)^2}{n(n-1)}$$

(Para dados não classificados fazer $n_i = 1$)

Intervalos de Confiança

$z_c = \Phi^{-1}(1 - \alpha/2)$

Grau de confiança (%)	99.73	99	98	96	95.45	95	90	80	68.27	50
Coeficiente de confiança z_c	3	2.58	2.33	2.05	2	1.96	1.645	1.28	1	0.675

	σ^2	**Tipo de população(s)**	**Intervalo de confiança**
Média	Conhecida	Normal Ou qualquer ($n \geq 30$)	$\left]\overline{x} - z_c \frac{\sigma}{\sqrt{n}}, \overline{x} + z_c \frac{\sigma}{\sqrt{n}}\right[$
	Desconhecida	Qualquer $(n \geq 30)$	$\left]\overline{x} - z_c \frac{\sigma}{\sqrt{n}}, \overline{x} + z_c \frac{\sigma}{\sqrt{n}}\right[$ $\sigma \approx s$
Proporção	_________	Bernoulli	$\left]\hat{p} - z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right[$ $n \geq 30$
Diferença de médias	Conhecidas	Normais ou quaisquer ($n_1 \geq 30$ e $n_2 \geq 30$)	$\left]\overline{x}_1 - \overline{x}_2 - z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \overline{x}_1 - \overline{x}_2 + z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right[$
	Desconhecidas $\sigma_1^2 \neq \sigma_2^2$	Quaisquer ($n_1 \geq 30$ e $n_2 \geq 30$)	$\left]\overline{x}_1 - \overline{x}_2 - z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \overline{x}_1 - \overline{x}_2 + z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right[$ $\sigma_1^2 \approx s_1^2 \quad \sigma_2^2 \approx s_2^2$
	Desconhecidas $\sigma_1^2 = \sigma_2^2$	Normais ($n_1 \geq 30$ e $n_2 \geq 30$)	$\left]\bar{x}_1 - \bar{x}_2 - z_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{x}_1 - \bar{x}_2 + z_c s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right[$ $s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$
Diferença de proporções	_________	Bernoulli	$\left]\hat{p}_1 - \hat{p}_2 - z_c \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}, \hat{p}_1 - \hat{p}_2 + z_c \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right[$ $n_1 \geq 30$ e $n_2 \geq 30$

Testes de Hipóteses

Valores de z_c

Significância (**α**)	0.01	0.02	0.04	0.05	0.1	0.2	Tipo de teste
$z_c = \Phi^{-1}(1-\alpha)$	2.33	2.05	1.75	1.645	1.28	0.84	Unilateral à direita
$z_c = \Phi^{-1}\left(1-\alpha/2\right)$	2.575	2.33	2.05	1.96	1.645	1.28	Bilateral

Testes de hipóteses para a média de uma população

Teste de média unilateral		Teste de média Bilateral	σ^2	Tipo de população	Estatística de teste
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$ $RC_Z =]-\infty, -z_c[$	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$ $RC_Z =]z_c, +\infty[$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ $RC_Z =]-\infty, -z_c[\cup]z_c, +\infty[$	Conhecida	Normal Ou qualquer $(n \geq 30)$	$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$
			Desconhecida $\sigma \approx s$	Qualquer $(n \geq 30)$	

Testes de hipóteses para diferença de médias de duas populações

Teste de diferença de médias unilateral		Teste de média Bilateral	σ_1^2, σ_2^2	Tipo de populações	Estatística de teste
$H_0: \mu_1 - \mu_2 = k$ $H_1: \mu_1 - \mu_2 < k$ $Rc =]-\infty, -z_c[$	$H_0: \mu_1 - \mu_2 = k$ $H_1: \mu_1 - \mu_2 > k$ $Rc =]z_c, +\infty[$	$H_0: \mu_1 - \mu_2 = k$ $H_1: \mu_1 - \mu_2 \neq k$ $Rc =]-\infty, -z_c[\cup]z_c, +\infty[$	Conhecidas	Normal Ou qualquer $(n \geq 30)$	$Z = \frac{\overline{X}_1 - \overline{X}_2 - k}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
			Desconhecidas $\sigma_1^2 \approx s_1^2, \sigma_2^2 \approx s_2^2$	Qualquer $(n \geq 30)$	

Testes de hipóteses para a proporção de uma população

Teste de proporção unilateral		Teste de proporção Bilateral	Tipo de população	Estatística de teste
$H_0: p = p_0$ $H_1: p < p_0$ $Rc =]-\infty, -z_c[$	$H_0: p = p_0$ $H_1: p > p_0$ $Rc =]z_c, +\infty[$	$H_0: p = p_0$ $H_1: p \neq p_0$ $Rc =]-\infty, -z_c[\cup]z_c, +\infty[$	Bernoulli $n \geq 30$	$Z = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

Testes de hipóteses para a diferença de proporções de duas populações

Teste de diferença de proporções unilateral		Teste de diferença de proporções bilateral	Tipo de populações	Estatística de teste
$H_0: p_1 - p_2 = k$ $H_1: p_1 - p_2 < k$ $Rc = \left]-\infty, -z_c\right[$	$H_0: p_1 - p_2 = k$ $H_1: p_1 - p_2 > k$ $Rc = \left]z_c, +\infty\right[$	$H_0: p_1 - p_2 = k$ $H_1: p_1 - p_2 \neq k$ $RC_Z = \left]-\infty, -z_c\right[\cup \left]z_c, +\infty\right[$	Bernoulli $n_1 \geq 30$ $n_2 \geq 30$	$Z = \frac{\hat{P}_1 - \hat{P}_2 - k}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$ $Z = \frac{\hat{P}_1 - \hat{P}_2 - k}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$ se k=0

Tipos de Erro

Erro	Probabilidade
Tipo I	$P(Erro\ tipo I) = P(rejeitar\ H_0 \mid H_0\ \ verdadeira) = \alpha$
Tipo II	$\beta = P(Erro\ tipo\ II) = P(não\ rejeitar\ H_0 \mid H_0\ \ falsa)$

Potência do teste: $1 - \beta = P(rejeitar\ \ H_0 \mid H_0\ falsa)$

Regressão Linear Simples

Sistema normal de equações	Estimativa dos parâmetros da reta de regressão
$\begin{cases} \hat{a}\cdot n+\hat{b}\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}y_i \\ \hat{a}\sum_{i=1}^{n}x_i+\hat{b}\sum_{i=1}^{n}x^2{}_i=\sum_{i=1}^{n}x_i y_i \end{cases}$	$\begin{cases} \hat{b}=\dfrac{\left(\sum_{i=1}^{n}x_i y_i\right)-n\cdot\overline{x}\cdot\overline{y}}{\sum_{i=1}^{n}x^2{}_i-n\cdot\overline{x}^2} \\ \hat{a}=\overline{y}-\hat{b}\cdot\overline{x} \end{cases}$ ou $\begin{cases} \hat{b}=\dfrac{S_{xy}}{S_{xx}} \\ \hat{a}=\overline{y}-\hat{b}\cdot\overline{x} \end{cases}$

Produtos cruzados	Erros e Variabilidade	Correlação
$s_{xx}=\sum_{i=1}^{n}(x_i-\overline{x})^2=\left(\sum_{i=1}^{n}x_i^2\right)-n\cdot\overline{x}^2$	$SE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=S_{yy}-\hat{b}S_{xy}$	$\hat{\rho}=\hat{b}\sqrt{\dfrac{S_{xx}}{S_{yy}}}$
$S_{yy}=\sum_{i=1}^{n}(y_i-\overline{y})^2=\left(\sum_{i=1}^{n}y_i^2\right)-n\cdot\overline{y}^2$	$SR=\sum_{i=1}^{n}(\hat{y}-\overline{y})^2=\dfrac{S^2{}_{xy}}{S_{xx}}$	$Obs:$
$S_{xy}=\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})=\left(\sum_{i=1}^{n}x_i y_i\right)-n\cdot\overline{x}\cdot\overline{y}$	$ST=\sum_{i=1}^{n}(y_i-\overline{y})^2=S_{yy}$	$SR=\hat{\rho}^2 ST$
	$ST=SE+SR$	$SE=(1-\hat{\rho}^2)ST$

$$\hat{\boldsymbol{\sigma}}_{\varepsilon}{}^2=\frac{1}{n-2}\left(1-\hat{\boldsymbol{\rho}}^2\right)S_{yy}$$

$$Obs:\ \hat{\boldsymbol{\sigma}}_{\varepsilon}{}^2=\frac{SE}{n-2}$$

I.C. Y=a+bX	Estatística	Erro
Parâmetro **a** $\hat{a}\pm\Delta$	$T_a=\dfrac{\hat{a}-a}{\hat{\boldsymbol{\sigma}}_{\varepsilon}\sqrt{\dfrac{1}{n}+\dfrac{\overline{x}^2}{S_{xx}}}}\sim t_{n-2}$	$\Delta=t_{\alpha/2}\hat{\boldsymbol{\sigma}}_{\varepsilon}\sqrt{\dfrac{1}{n}+\dfrac{\overline{x}^2}{S_{xx}}}$
Parâmetro **b** $\hat{b}\pm\Delta$	$T_b=\dfrac{\hat{b}-b}{\hat{\boldsymbol{\sigma}}_{\varepsilon}/\sqrt{S_{xx}}}\sim t_{n-2}$	$\Delta=t_{\alpha/2}\hat{\boldsymbol{\sigma}}\sqrt{\dfrac{1}{S_{xx}}}$
Média de Y_0 : $E(Y_0)$ $\hat{Y}_0\pm\Delta$	$T=\dfrac{\hat{Y}_0-E(Y_0)}{\hat{\boldsymbol{\sigma}}_{\varepsilon}\sqrt{\dfrac{1}{n}+\dfrac{(x_0-\overline{x})^2}{S_{xx}}}}$	$\Delta=t_{\alpha/2}\hat{\boldsymbol{\sigma}}_{\varepsilon}\sqrt{\dfrac{1}{n}+\dfrac{(x_0-\overline{x})^2}{S_{xx}}}$

Previsão de Y $\hat{Y}_0 \pm \Delta$	$T = \dfrac{Y_0 - \hat{Y}_0}{\hat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}} \sim t_{n-2}$	$\Delta = t_{\alpha/2} \hat{\sigma}_\varepsilon \sqrt{1 + \dfrac{1}{n} + \dfrac{(x_0 - \overline{x})^2}{S_{xx}}}$