

RAD1408 – Estatística Aplicada à Administração I
Prof. Dr. Evandro Marcos Saidel Ribeiro

RESUMO

ESTATÍSTICA APLICADA À ADM. I

TESTE DE HIPÓTESE: Introdução

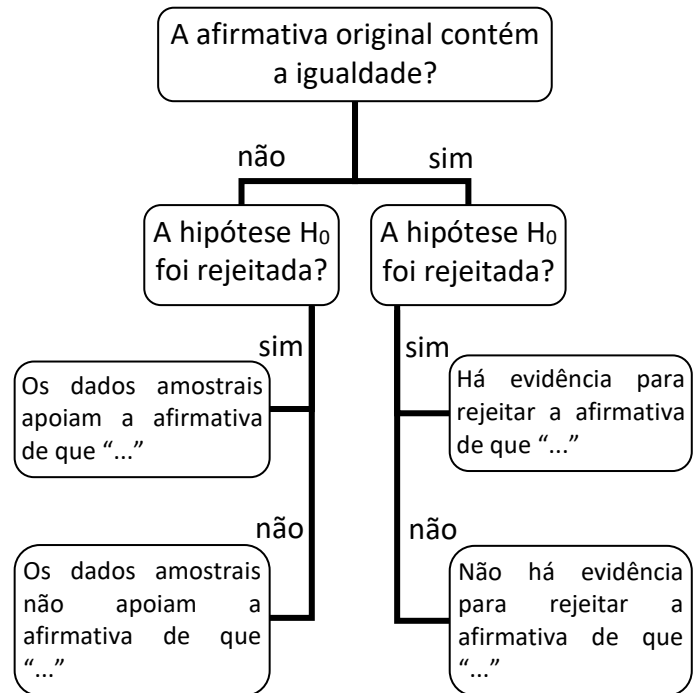
Ter em mente a “Regra do evento raro”:

Se, sob uma dada suposição, a probabilidade de um evento observado particular é excepcionalmente pequena, concluímos que a suposição provavelmente não é correta.

Procedimento para Teste de Hipótese:

- 1 – Escreva a afirmativa original (AO) na forma simbólica.
- 2 – Escreva o oposto da afirmativa original (Oposto) na forma simbólica.
- 3 – Expresse H_0 e H_1 . (H_0 é sempre a igualdade)
- 4 – Selecione o nível de significância α . Verifique o tipo de teste (unilateral, bilateral). Verifique a distribuição a ser utilizada.
- 5 – Calcule a estatística teste.
- 6 – A partir da estatística teste verifique se H_0 é rejeitada ou não, para tanto e utilize um método:
 - (a) Método tradicional: Compare a estatística teste com o valor crítico, se a estatística teste está na região crítica, Rejeite H_0 .
 - (b) Método valor-P: a partir da estatística teste determine o valor-P. Se valor-P $< \alpha$, Rejeite H_0 .
- 7 – Estabeleça a conclusão sobre a Afirmativa Original (fraseado final a seguir).

Fraseado Final para o Teste de Hipótese:



1 TESTE DE HIPÓTESE - uma amostra

Inferência sobre uma proporção: $H_0 : p = p_1$

Estatística teste:
$$z_{teste} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

sendo $q = 1 - p$.

2 TESTE DE HIPÓTESE - duas amostras

Inferências sobre duas médias: Amostras Independentes $H_0 : \mu_1 = \mu_2$

Estatística teste:
$$t_{teste} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

grau de liberdade:

$gl =$ escolha o menor entre $(n_1 - 1)$ e $(n_2 - 1)$

3 ANOVA (ANalysis Of VAriance) - Inferência a partir de mais de duas amostras

Hipóteses: $H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : pelo menos uma das médias é diferente das outras

- ANOVA de um fator -----

- o Cálculos com tamanhos iguais - - - - -

Variância entre amostras = $n s_{\bar{x}}^2$, sendo $s_{\bar{x}}^2$ a variância das médias amostrais.

Variância dentro das amostras = s_p^2 , sendo s_p^2 a média das variâncias amostrais.

Estatística Teste:
$$F_{teste} = \frac{n s_{\bar{x}}^2}{s_p^2}$$

- grau de lib. do numerador: $gl_{Num} = k - 1$

- grau de lib. do denominador: $gl_{Den} = k(n - 1)$

Exemplo para ANOVA de um fator com tamanhos amostrais iguais:

	A	B	C	D	E	F
1						
2		Tam 1	Tam 2	Tam 3		
3		79945,00	65040,00	53685,00		
4		51490,00	67750,00	60975,00		
5		89430,00	69105,00	90785,00		
6		65040,00	81300,00	82655,00		
7	n_i	4	4	4	=CONT.VALORES(D3:D6)	
8	\bar{x}_i	71476,25	70798,75	72025,00	=MÉDIA(D3:D6)	
9	s_i^2	278310790	51867706,25	307797533,3	=VAR(D3:D6)	
10						
11	$s_{\bar{x}}^2$	377303,646	=VAR(B8:FD8)			
12						
13	s_p^2	212658676	=MÉDIA(B9:D9)		$n =$	4
14						
15		Variância entre amostras:		1509214,583	=F13*B11	$= n s_{\bar{x}}^2$
16		Variância dentro da am.:		212658676	=B13	$= s_p^2$
17						
18		Estatística F =	0,0071	=D15/D16	=	$F_{teste} = \frac{n s_{\bar{x}}^2}{s_p^2}$
19						

- o Cálculos com tamanhos amostrais diferentes - - - - -

Estatística Teste:
$$F_{teste} = \frac{\left[\frac{\sum n_i (\bar{x}_i - \bar{\bar{x}})^2}{k - 1} \right]}{\left[\frac{\sum (n_i - 1) s_i^2}{N - k} \right]}$$
, sendo $\bar{\bar{x}}$ a média de todos os valores amostrais e

- grau de liberdade do numerador: $gl_{Num} = k - 1$; grau de liberdade do denominador: $gl_{Den} = N - k$

Exemplo para ANOVA de um fator com tamanhos amostrais diferentes:

	A	B	C	D	E	F	G	H	I	J
4			PL (em R\$ mil) de Empresas							
5			Tam 1	Tam 2	Tam 3		i	$n_i(\bar{x}_i - \bar{x})^2$	$(n_i - 1)s_i^2$	
6			79945	84010	92140		1	4,51E+08	2,89E+08	
7			67750	67750	73170		2	3,01E+07	2,08E+08	
8			56910	69105	90785		3	5,11E+08	7,00E+08	
9			65040	81300	82655		Soma	9,92E+08	1,20E+09	
10			71815	74525	78590		k - 1 =	2		
11					105690		N - k =	14		
12					81300		Variância Entre		4,96E+08	=H9/H10
13		n_i	5	5	7		Variância Dentro		8,55E+07	=I9/H11
14		\bar{x}_i	68292,00	75338,00	86332,86					
15		s_i^2	7,22E+07	5,20E+07	1,17E+08		Estadística teste =		5,80218	=I12/I13
16		N	17	=SOMA(C13:E13)						
17		k	3	=CONT.VALORES(C13:E13)				valor-P =	0,01461	
18		\bar{x}	77792,94	=MÉDIA(C6:E12)						

4 CORRELAÇÃO

Considere pares de dados (y,x)

y	x
y_1	x_1
y_2	x_2
...	...
y_n	x_n

Coefficiente de Correlação Linear de Pearson:

$$r = \frac{SQ_{xy}}{\sqrt{SQ_{xx}} \sqrt{SQ_{yy}}}$$

sendo

$$SQ_{wz} = (\sum wz) - \frac{(\sum w)(\sum z)}{n} \quad e$$

$$SQ_{zz} = (\sum z^2) - \frac{(\sum z)^2}{n}$$

O teste de hipótese:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

$$gl = n - 2$$

$$t_{teste} = r \sqrt{\frac{n-2}{1-r^2}} \quad \text{Ou valor crítico} \quad r_\alpha = \frac{t_\alpha}{\sqrt{t_\alpha^2 + n - 2}}$$

Na HP:

Digite os dados:
y, [enter], x, Σ
r: [g], 2, [xy]

Registros	Valor
R ₁	n
R ₂	Σx
R ₃	Σx^2
R ₄	Σy
R ₅	Σy^2
R ₆	Σxy

No Excel: Ferramentas

- > Análise de Dados
- > Correlação

5 REGRESSÃO LINEAR SIMPLES

- Equação da reta: $\hat{y}_i = b_0 + b_1 x_i$ sendo

$$b_1 = \frac{SQ_{xy}}{SQ_{xx}} \quad e \quad b_0 = \bar{y} - b_1 \bar{x}$$

- Para avaliar o ajuste:

$$SQ_{Tot} = \sum_{i=1}^n (y_i - \bar{y})^2 = (\sum y^2) - \frac{(\sum y)^2}{n}$$

$$SQ_{Reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = b_0 (\sum y) + b_1 (\sum xy) - \frac{(\sum y)^2}{n}$$

$$SQ_{Erro} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = (\sum y^2) - b_0 (\sum y) - b_1 (\sum xy)$$

$$R^2 = 1 - \frac{SQ_{Erro}}{SQ_{Tot}}$$

Significado do R^2 : proporção das variações de y que são explicadas pelas variações de x.

$$\text{Erro Padrão: } S_{xy} = \sqrt{\frac{SQ_{Erro}}{n-2}} \quad S_{b_1} = \frac{S_{xy}}{\sqrt{SQ_{xx}}}$$

$$S_{b_0} = S_{xy} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SQ_{xx}}}$$

Teste de Hipótese

$$H_0 : \beta_i = 0 \quad t_{teste} = \frac{b_i}{S_{b_i}}$$

$$H_1 : \beta_i \neq 0 \quad gl = n - 2$$

No Excel: > Ferramentas > Análise de Dados
> Regressão

Na HP:

$$b_0 : 0, [g], 2$$

$$b_1 : 1, [g], 2, [x y], R \downarrow, [xy], [-]$$

5.1 Intervalo de confiança na regressão

Para obter o erro deve-se calcular h_i em cada valor de x_i .

$$h_i = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{SSX} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2}$$

h_i será utilizado para o cálculo do intervalo de confiança, como visto nas expressões seguintes

5.1.1 Estimativas para média

Intervalo de Confiança para $\mu_{Y|X=X_i}$:

$$\hat{Y} \pm t_{\alpha/2} S_{YX} \sqrt{h_i}$$

5.1 Estimativas para valor

Intervalo de Confiança para $Y_{X=X_i}$:

$$\hat{Y} \pm t_{\alpha/2} S_{YX} \sqrt{1 + h_i}$$

6 MATRIZ DE CORRELAÇÃO

A matriz de correlação é escrita na forma:

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & 1 \end{bmatrix} \quad (1)$$

O cálculo de cada elemento desta matriz pode ser obtido como mostrado na seção 4.

A matriz R é simétrica, ou seja, $r_{ij} = r_{ji}$.

7 MATRIZ DE COVARIÂNCIA (a partir da correlação)

O cálculo de cada elemento da matriz de covariância pode ser feito a partir da matriz de correlação. Assim, a partir da correlação r_{12} , entre duas variáveis x_1 e x_2 , a covariância é dada por:

$$S_{ij} = r_{ij} s_i s_j \quad (2)$$

Na expressão (2) s_i e s_j são os desvios-padrão das variáveis i e j , respectivamente.

8 REGRESSÃO LINEAR MÚLTIPLA

Considere mais de duas variáveis (Ex: y, x_1, x_2)

Y	X ₁	X ₂
Y ₁	X ₁₁	X ₁₂
Y ₂	X ₂₁	X ₂₂
...
Y _n	X _{n1}	X _{n2}

O objetivo é obter o modelo de regressão linear múltipla, considerando k variáveis independentes:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k$$

Obter o modelo significa encontrar os valores de $b_0, b_1, b_2, \dots, b_k$.

Etapas para o estudo de regressão linear múltipla:

A resposta do problema (obtenção de b 's) pode ser dada através de matrizes:

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \boldsymbol{\beta} = (\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{Y}) \quad (3)$$

A seguir os passos para obtenção dos b 's:

- Gráficos de dispersão $y - x_1, y - x_2, \dots, y - x_k$
- A partir da tabela de dados considere as matrizes
- \mathbf{Z} e \mathbf{Y} dadas por:

$$\mathbf{Z} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \quad \text{e} \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

- Obtenha Z' , ou seja, a transposta da matriz Z :

$$Z' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \dots & x_{nk} \end{bmatrix}$$

- Multiplique a matriz Z' pela matriz Z :
 $Z'Z$
- Considere o resultado anterior e inverta obtendo a matriz

$$M = (Z'Z)^{-1} \quad (4)$$

Para inverter a matriz veja: **9** MATRIZ INVERSA.

- Faça a multiplicação das matrizes Z' e Y obtendo a matriz A dada por

$$A = Z'Y \quad (5)$$

- Multiplique as matrizes obtidas em (4) e (5) obtendo assim os valores de b 's definidos em (3), ou seja:

$$\beta = (Z'Z)^{-1}(Z'Y)$$

- Com os valores de b 's escreva a equação linear:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k \quad (6)$$

- A partir da expressão (6) calcule estimativas para cada caso em análise.
- Calcule os resíduos ε_i para cada estimativa:

$$\varepsilon_i = y_i - \hat{y}_i$$

- Faça um gráfico dos resíduos (no eixo-y) como função das estimativas (no eixo-x).

9 MATRIZ INVERSA

A matriz inversa de uma matriz A (se existir) é a matriz A^{-1} tal que $A A^{-1} = A^{-1} A = I$.

Determinação da Matriz Inversa pela matriz Adjunta:

$$A^{-1} = \frac{1}{\det(A)} \text{adj}(A) \quad (7)$$

Exemplo: Para uma matriz **3x3**:

Considere os elementos nomeados de acordo com a expressão (8) abaixo:

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \quad (8)$$

A matriz adjunta de A é obtida pela matriz dos cofatores transposta, que no caso da matriz A da expressão (8) resulta em:

$$\text{adj}(A) = \begin{bmatrix} + \begin{vmatrix} e & f \\ h & i \end{vmatrix} & - \begin{vmatrix} b & c \\ h & i \end{vmatrix} & + \begin{vmatrix} b & c \\ e & f \end{vmatrix} \\ - \begin{vmatrix} d & f \\ g & i \end{vmatrix} & + \begin{vmatrix} a & c \\ g & i \end{vmatrix} & - \begin{vmatrix} a & c \\ d & f \end{vmatrix} \\ + \begin{vmatrix} d & e \\ g & h \end{vmatrix} & - \begin{vmatrix} a & b \\ g & h \end{vmatrix} & + \begin{vmatrix} a & b \\ d & e \end{vmatrix} \end{bmatrix} \quad (9)$$

O determinante de A é obtido por:

$$\det(A) = aei + bfg + dhc - gec - ahf - bdi \quad (10)$$

O resultado de (9) é uma matriz 3x3 o resultado de 10 é um número. Estes resultados são utilizados em (7) para obter a matriz inversa da matriz A .