Estatística

Comparação de 2 proporções:

$H_0 = p_1 - p_2 = \Delta p$.

Uso da normal c/ média $p_1 - p_2$ e variância

$$\sigma^2(p_1' - p_2') = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$ e com p_1 e p_2 estimados por p_1' e p_2'.

$$Z_{calc} = \frac{(p_1' - p_2') - \Delta p}{\sqrt{\frac{p_1'(1-p_1')}{n_1} + \frac{p_2'(1-p_2')}{n_2}}}$$

Exercício 11:

Teste p/ 2 proporções:

$H_0: p_1 = p_2$ 3º caso

$H_1: p_1 \neq p_2$ $\alpha = 5\%$

$n_1 = 80$ $n_2 = 50$

$p_1' = \frac{f_1}{n_1} = \frac{32}{80} = 0{,}4$ $p_2' = \frac{f_2}{n_2} = \frac{26}{50} = 0{,}52$

$$Z_{calc} = \frac{(0{,}4 - 0{,}52) - 0}{\sqrt{\frac{0{,}4(1-0{,}4)}{80} + \frac{0{,}52(1-0{,}52)}{50}}}$$

$Z_{calc} = -1{,}34$

Se $|Z_{calc}| > Z_{\alpha/2}$, rejeito H_0

$Z_{\alpha/2} = Z_{2{,}5\%} = 1{,}96$

Como $|-1{,}34| < 1{,}96$, aceito H_0

Resposta: SIM.

Exercício 12:

Teste p/ μ c/ σ desconhecido

1º caso

$H_0: \mu = \mu_0$ $\rightarrow$ $H_0: \mu = 45000$

$H_1: \mu < \mu_0$ $H_1: \mu < 45000$

$\mu_0 = 45000$

$\bar{X} = 44.175$ $\alpha = 5\%$

$S_x = 3000$ $n = 16$

$$t_{calc} = \frac{\bar{X} - \mu_0}{S_x/\sqrt{n}} = \frac{44.175 - 45.000}{3000/\sqrt{16}}$$

$t_{calc} = -1{,}10$

Se $t_{calc} < t_{n-1,\alpha}$, rejeito H_0

Como $-1{,}10 > -1{,}753$, aceito H_0

Resposta: NÃO. Aceita-se o lote.

Exercício 13:

Teste p/ σ^2

2º caso:

$H_0: \sigma^2 = \sigma_0^2 \longrightarrow H_0: \sigma^2 = 25$

$H_1: \sigma^2 < \sigma^2 \qquad H_1: \sigma^2 < 25$

$\sigma_0^2 = 25$

$n = 10 \qquad \alpha = 5\%$

$S_x^2 = 12{,}4$

$$\chi^2_{calc} = \frac{(n-1).S_x^2}{\sigma_0^2} = \frac{(10-1).12{,}4}{25}$$

$$\chi^2_{calc} = 4{,}964$$

Se $\chi^2_{calc} < \chi^2_{n-1,1-\alpha}$, rejeita H_0

$\chi^2_{9,95\%} = 3{,}325$

Como $4{,}964 > 3{,}325$, aceita H_0

Resposta: NÃO

Exercício 19:

Teste p/ proporção:

$H_0: p = p_0 \longrightarrow H_0: p = 0{,}50$

$H_1: p \neq p_0 \qquad H_1: p \neq 0{,}50$

$p_0 = 0{,}50$

$n = 100 \qquad \alpha = 5\%$

$$p' = \frac{f'}{n} = \frac{59}{100} = 0{,}59$$

$$Z_{calc} = \frac{p' - p_0}{\sqrt{p_0(1-p_0)/n}}$$

$$Z_{calc} = \frac{0{,}59 - 0{,}50}{\sqrt{0{,}5(1-0{,}5)/100}}$$

$$Z_{calc} = 1{,}80$$

Se $|Z_{calc}| > Z_{\alpha/2}$, rejeita H_0

$Z_{2,5\%} = 1{,}96$

Como $1{,}80 < 1{,}96$, aceita H_0

Resp: Não, não há viés a 5%.

Testes de Aderência

* Testes de aderência pelo χ^2

H_0: a distribuição é do tal tipo

H_1: " " não " " " "

$$\chi^2_v = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

$E_i = n.p_i$

onde p_i é a probabilidade de se obter um valor da variável na classe considerada e n é o número de elementos da amostra.

$v = k - 1 - m$

k = nº de classes condição: $E_i \geq 5$

m = nº de parâmetros

Slide:

$$\chi^2_v = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{k} \frac{O_i^2}{E_i} - n$$

- χ^2_v = estatística do teste com v graus de liberdade
- O_i = frequência observada de uma determinada classe ou valor da var.
- E_i = freq. esperada, segundo o modelo testado, dessa classe ou valor
- $n = \sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i$ = número de elementos da amostra

Exercício 01:

Teste de aderência pelo χ^2

H_0: a distrib. do número de defeitos / unidade é do tipo Poisson

H_1: " " " " " " " não " " " "

Distrib. Poisson:

$$P(X = k) = \frac{\mu^k . e^{-\mu}}{k!} \quad (k = 0, 1, 2, \ldots)$$

Calcular $\bar{x}$ p/ estimar μ:

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{155}{100} = 1{,}55$$

Usando $\mu = 1{,}55$, calcular os valores $p_i = P(X=i)$, $i = 0,1,2,\dots,7$

$p_0 = \frac{\mu^k \cdot e^{-\mu}}{k!} = \frac{1{,}55^0 \cdot e^{-1{,}55}}{0!} = e^{-1{,}55} = 0{,}212$ $\qquad E_0 = n \cdot p_0 = 100 \cdot 0{,}212 = 21{,}2$

$p_1 = \frac{\mu^k \cdot e^{-\mu}}{k!} = \frac{1{,}55^1 \cdot e^{-1{,}55}}{1!} = 0{,}329$ $\qquad E_1 = n \cdot p_1 \quad 100 \cdot 0{,}329 = 32{,}9$

$p_2 = \frac{\mu^k \cdot e^{-\mu}}{k!} = \frac{(1{,}55)^2 \cdot e^{-1{,}55}}{2!} = 0{,}255$ $\qquad E_2 = n \cdot p_2 = 100 \cdot 0{,}255 = 25{,}5$

$\vdots$ $\qquad \vdots$

$p_7 = \dots$

x_i	$f_i = O_i$	$x_i f_i$	p_i	E_i	$O_i - E_i$	$(O_i - E_i)^2/E_i$
0	25	0	0,212	21,2	3,8	0,681
1	35	35	0,329	32,9	2,1	0,134
2	18	36	0,255	25,5	-7,5	2,206
3	13	39	0,132	13,2	-0,2	0,003
4	4	16	0,051	5,1		
5	2	10	0,016	1,6	1,8	0,950
6	2	12	0,004	0,4		
7	1	7	0,001	0,1		
	100	155		100		3,474

9 $\qquad$ 7,2

devido ao método utilizado, valores de $E_i < 5$ devem ser agrupados em classes adjacentes para não distorcer o resultado.

$\chi^2_r = 3{,}474$

$\chi^2_{crítico} \rightarrow r = k - 1 - m$

k = nº de classes consideradas = 5

m = nº de parâmetros estimados = 1

$r = 5 - 1 - 1 = 3$

$\chi^2_{3,5\%} = 7{,}815$

Se $\chi^2_r > \chi^2_{crítico}$, rejeito H_0

como $3{,}474 < 7{,}815$, aceito H_0

Pode-se afirmar que a distribuição é do tipo Poisson.

26/10/15

Método de Kolmogorov-Smirnov

H_0: a distrib. é de tal tipo

H_1: " " não " " " "

$d = \max |f(x) - G(x)|$

Se d for > o valor crítico p/ α e n_1, rejeita-se H_0

$G(x_i)$ à esquerda: $(i-1)/n$

$G(x_i)$ à direita: i/n

Exercício 2:

$n = 10$

amostra ordenada: 25,4; 27,0; 27,3; 27,8; 28,0;
29,2; 29,5; 30,2; 30,6; 33,5

H_0: a distrib. é normal

H_1: " " não " "

$\mu = 30$; $\sigma = 2$; $\alpha = 5\%$

$$z = \frac{x - \mu}{\sigma}$$

$G(x_i)$ à esquerda: $(i-1)/n$

$G(x_i)$ à direita: i/n

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{25{,}4 - 30}{2} = -2{,}30$$

$G(x_1)$ à esq. $= (1-1)/10 = 0{,}0$

$G(x_1)$ à dir. $= 1/10 = 0{,}1$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{27{,}0 - 30}{2} = -1{,}50$$

$G(x_2)$ à esq. $= (2-1)/10 = 0{,}1$

$G(x_2)$ à dir. $= 2/10 = 0{,}2$

$$z_3 = \frac{x_3 - \mu}{\sigma} = \frac{27{,}3 - 30}{2} = -1{,}35$$

$G(x_3)$ à esq. $= (3-1)/10 = 0{,}2$

$G(x_3)$ à dir $= 3/10 = 0{,}3$

$\vdots$

z_{10}

$\vdots$

$G(x_{10})$

x_i	z_i	$f(x_i$	$G(x_i)$ esq.	$G(x_i)$ dir	$\lvert f(x_i) - G(x_i) \rvert$ esq.	$\lvert f(x_i) - G(x_i) \rvert$ dir.
25,4	-2,30	0,0167	0,0	0,1	0,0167	0,0833
27,0	-1,50	0,0668	0,1	0,2	0,0332	0,1332
27	-1,35	0,0885	0,2	0,3	0,1115	0,2115
27,8	-1,10	0,1357	0,3	0,4	0,1643	0,2643
28,0	-1,00	0,1587	0,4	0,5	0,2413	0,3413
29,2	-0,40	0,3446	0,5	0,6	0,1554	0,2554
29,5	-0,25	0,4013	0,6	0,7	0,1987	0,2987
30,2	0,10	0,5398	0,7	0,8	0,1602	0,2602
30,6	0,30	0,6179	0,8	0,9	0,1821	0,2821
33,5	1,75	0,9599	0,9	1,0	0,0599	0,1599

$d = \max \lvert f(x_i) - G(x_i) \rvert = 0{,}3413$

$n = 10$, $\alpha = 5\%$ } Tabela ks-smi valor crítico $\doteq 0{,}409$

Se $d >$ valor crítico, rejeita-se H_0

Como $0{,}3413 < 0{,}409$, aceita-se H_0

Pode-se afirmar que a distribuição seja normal c/ $\mu = 30$ e $\sigma = 2{,}0$ a 5% de significância.

Verificação Gráfica da Aderência

Exercício 03:

Dados ordenados:

83, 92, 94, 96, 97, 101, 106, 112

Cálculo dos percentis: $\frac{50(2i-1)}{n}$

1º percentil: $50(2.1-1)/8 = 6{,}25\%$

2º " : $50(2.2-1)/8 = 18{,}75\%$

3º " : $50(2.3-1)/8 = 31{,}25\%$

4º " : $50(2.4-1)/8 = 43{,}75\%$

5º " : $50(2.5-1)/8 = 56{,}25\%$

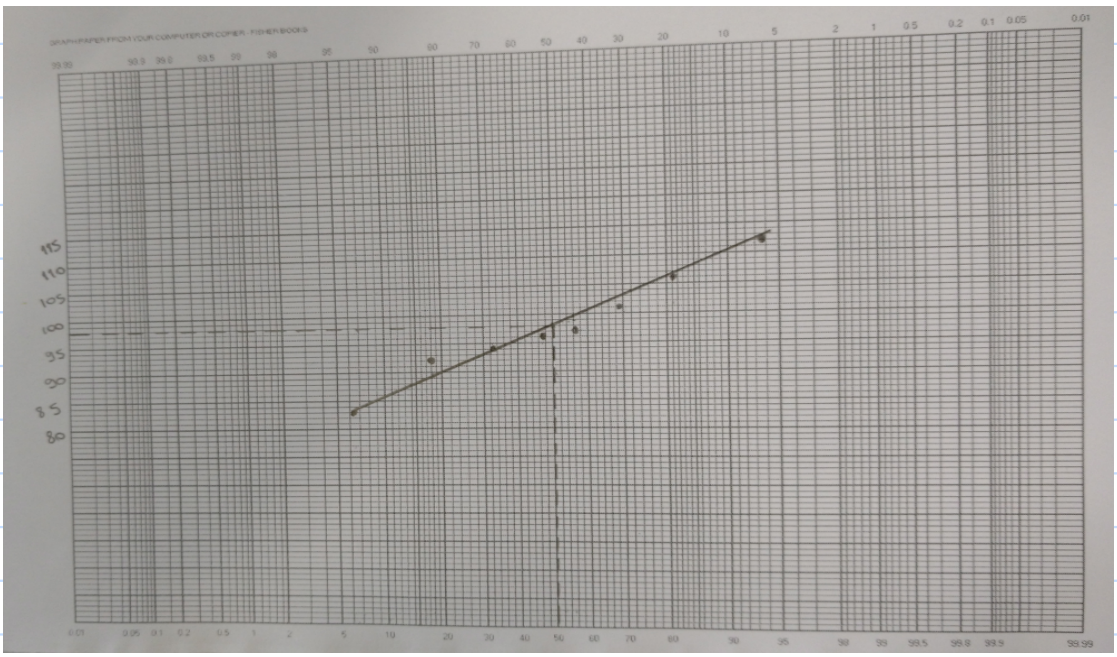
6º " : $50(2.6-1)/8 = 68{,}75\%$

7º " : $50(2.7-1)/8 = 81{,}25\%$

8º " : $50(2.8-1)/8 = 93{,}75\%$

Lançar no papel de prob. normal

eixo X:	6,25	18,75	31,25	43,75	56,25	68,75	81,25	93,75
eixo Y:	83	92	94	96	97	101	106	112



A distribuição se ajusta a uma normal.

p/ estimar μ obter ponto de 50%

$\mu = 98$

p/ estimar σ obter pontos de -1σ e $+1\sigma$:

tabela normal c/ valor de 1,0 encontro 34,13%.

ponto de -1σ: 50% − 34,13% = 15,87%

ponto de $+1\sigma$: 50% + 34,13% = 84,13%

$\mu + 1\sigma$: 106,5 $\quad \mu = 97$

$\mu - 1\sigma$: 87,5 $\quad \sigma = 9{,}5$

Calculados:

$\bar{X} = 97{,}625$

$S_x^2 = 8{,}863$

Teste de Independência

2 ou mais variáveis qualitativas de interesse, a representação tabular das frequências observadas pode ser feita através de uma tabela de contingência

H_0: as variáveis são independentes

H_1: não independentes (há algum grau de associação)

$$X_r^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

r: nº de linhas

s: nº de colunas

O_{ij}: freq. observada

E_{ij}: freq. esperada $= n.p_{ij}$

n: nº de elementos da amostra

r: nº de graus de liberdade $= (r-1)(s-1)$

p_{ij} é a probabilidade de ocorrer uma observação na célula considerada.

Havendo independência entre as variáveis (conforme H_0)

$$X_r^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$p_{ij} = p_i . p_j$ ← valores populacionais

Como não temos p_i e p_j, é necessário estimá-los pelo valor da amostra (p_i' e p_j')

$$p_i' = \frac{f_i}{n} \qquad p_j' = \frac{f_j}{n}$$

Então, $E_{ij} = n.p_{ij} = n.p_i.p_j = n.p_i'.p_j' = n.\frac{f_i}{n}.\frac{f_j}{n} \Rightarrow \boxed{E_{ij} = \frac{f_i.f_j}{n}}$

Se $X_r^2 > X_{r,\alpha}^2$; rejeita H_0

Ex 4: p. 141 Costa Neto

Realizar teste de independência para dados abaixo, com nível de significância de 1%.

	favorável	desf.	indif.	total
♂	33	12	15	60
♀	7	20	13	40
total	40	32	28	100

$r = 2$ (linhas)

$s = 3$ (colunas)

$r = (2-1)(3-1) = 2$

calcular o valor esperado em cada uma das posições

$$E_{ij} = \frac{f_i.f_j}{n}$$

$$E_{11} = \frac{60.40}{100} = 24$$

linha coluna

$$E_{12} = \frac{60.32}{100} = 19{,}2$$

$$E_{13} = \frac{60.28}{100} = 16{,}8$$

$$E_{21} = \frac{40.40}{100} = 16{,}0$$

$$E_{22} = \frac{40.32}{100} = 12{,}8$$

$$E_{23} = \frac{40.28}{100} = 11{,}2$$

$$X_r^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

O_{ij}	E_{ij}	$O_{ij} - E_{ij}$	$(O_{ij} - E_{ij})^2 / E_{ij}$
33	24	9	3,375
12	19,2	-7,2	2,7
15	16,8	-1,8	0,193
7	16	-9	5,063
20	12,8	7,2	4,050
13	11,2	1,8	0,289

$\chi_r^2 = 15,670$

Assim, H_0: variáveis (homem / mulher) independentes
H_1: variáveis não são independentes

$\chi^2_{r,\alpha} = \chi^2_{2,1\%} = 9,210$

$\chi^2_r > \chi^2_{r,\alpha} \Rightarrow 15,670 > 9,210 \Rightarrow$ rejeito $H_0 \Rightarrow$ as variáveis não são independentes.

ANOVA 1

Análise de Variância - uma classificação e amostras de mesmo tamanho

k amostras de tamanho n retiradas de k populações com médias μ_i $(i = 1, 2, \dots, k)$.

05/11/15

- H_0: $\mu_1 = \mu_2 = \dots = \mu_k$
- x_{ij}: $(i = 1, 2, \dots, k;\ j = 1, 2, \dots, n)$ é o j-ésimo valor da i-ésima amostra de n elementos.
- $T_i = \sum_j x_{ij}$: soma dos valores da i-ésima amostra
- $Q_i = \sum_j x_{ij}^2$: soma dos quadrados dos valores da i-ésima amostra
- $T = \sum_i T_i$: soma total dos valores
- $Q = \sum_i Q_i$: soma total dos quadrados dos valores

* SQT : Soma dos Quadrados Total

Estimativa válida de σ^2 se H_0 for verdadeira (todas as populações são normais c/ mesma média e mesma variância)

$$S_T^2 = \frac{\sum_i \sum_j (x_{ij} - \bar{\bar{x}})^2}{nk - 1} \quad \text{ou} \quad S_T^2 = \frac{\sum_i \sum_j x_{ij}^2 - \left[\left(\sum_i \sum_j x_{ij}\right)^2 / nk\right]}{nk - 1}$$

Substituindo Q e T:

$$S_T^2 = \frac{Q - T^2/}{nk -}$$

→ nº de graus de liberdade

Ao numerador de S_T^2 dá-se o nome de SQT

$$\boxed{SQT = Q - T^2/nk}$$

$$S_T^2 = \frac{SQT}{nk-1}$$

* SQE: Soma dos Quadrados entre amostras

$$S_E^2 = n \cdot \sum_i (\bar{x}_i - \bar{\bar{x}})^2 = \frac{n}{k-1} \cdot \left[\sum_i \bar{x}_i^2 - \frac{\left(\sum_i \bar{x}_i\right)^2}{k} \right]$$

Substituindo $\bar{\bar{x}}_i$, $\bar{x}_i$, T_i

$$S_E^2 = \frac{T_i^2/n - T^2/nk}{k-1}$$ → nº de G.L.

Ao numerador dá-se o nome de SQE:

$$\boxed{SQE = T_i^2/n - T^2/nk}$$

* SQR: Soma dos Quadrados Residuais

$$SQT = SQE + SQR \Rightarrow SQR = SQT - SQE$$

$$SQR = Q - \frac{T^2}{nk} - \sum_i T_i^2/n + T^2/nk \Rightarrow SQR = Q - \sum_i T_i^2/n$$

$$S_R^2 = \frac{SQR}{k(n-1)}$$

Subst. Ho pela hipótese que S_E^2 e S_R^2 estimem a variância σ^2

Teste p/ ANOVA:

$$\boxed{f_{calc} = \frac{S_E^2}{S_R^2}}$$ Ho: $\mu_1 = \mu_2 = \dots = \mu_k$

$f_\alpha = f_{(k-1);\, k(n-1),\alpha}$

Se $f_{calc} > f_\alpha$ rejeita-se Ho

Exemplo 1:

	manhã		Tarde		noite	
	x_{1j}	x_{2j}^2	x_{2j}	x_{2j}^2	x_{3j}	x_{3j}^2
	4	16	2	4	3	9
	5	25	4	16	6	36
	5	25	3	9	6	36
	4	16	7	49	4	16
	8	64	5	25	5	25
	4	16	4	16	4	16
	3	9	2	4	6	36
	7	49	5	25	6	36
T_i	40	↓	32	↓	40	↓
Q_i		220		148		220
T_i^2	1600		1024		1600	

$T = \sum T_i = 112$

$Q = \sum Q_i = 578$

$\sum T_i^2 = 4224$

$n = 8$

$k = 3$

$SQE = \sum T_i^2/n - T^2/nk$

$= 4224/8 - (112)^2/24 = 5{,}33$

$SQR = Q - \sum T_i^2/n$

$= 578 - (112)^2/8 = 50$

$SQT = Q - T^2/nk$

$= 578 - (112)^2/24 = 55{,}33$

$$S_E^2 = \frac{SQE}{k-1} = \frac{5{,}33}{2} = 2{,}665$$

$$S_R^2 = \frac{SQR}{k(n-1)} = \frac{50}{3 \cdot 7} = 2{,}381$$

$$f_{calc} = \frac{S_E^2}{S_R^2} = \frac{2{,}665}{2{,}381} = 1{,}119$$

$$f_\alpha = f_{(k-1);\, k(n-1);\, \alpha} = f_{(3-1);\, 3(8-1);\, 5\%}$$

$$f_{2;21;\,5\%} = 3{,}47$$

∴ como $f_{calc} < f_\alpha$, aceito H_0, não há diferença entre as médias

Análise de Variância - uma classificação com amostras de tamanhos diferentes

k amostras de tamanhos diferentes n_i ($i = 1, 2, \dots, k$) retiradas de k populações com médias μ_i

$$S_E^2 = \frac{\sum_i T_i^2/n_i - T^2/\sum_i n_i}{k-1} = \frac{SQE}{k-1}$$

$$S_R^2 = \frac{Q - \sum_i T_i^2/n_i}{\sum_i n_i - k} = \frac{SQR}{\sum_i n_i - k}$$

$$\boxed{f_{calc} = \frac{S_E^2}{S_R^2} \qquad f_\alpha = f_{(k-1);\, \Sigma n_i - k;\, \alpha}}$$

Se $f_{calc} > f_\alpha$ rejeito H_0

$$SQE = \sum_i T_i^2/n_i - T^2/\sum_i n_i$$

$$SQR = Q - \sum_i T_i^2/n_i$$

$$SQT = SQE + SQR$$

$$SQT = Q - T^2/\sum n_i$$

Exercício 2:

	A		B		C	
	x_{1j}	x_{2j}^2	x_{2j}	x_{2j}^2	x_{3j}	x_{3j}^2
	15	225	19	361	29	841
	18	324	20	400	27	729
	18	324	16	256	23	529
	20	400	21	441	20	400
	19	361	19	361	21	441
	24	576	15	225		
	19	361	9	81		
			13	169		
			19	361		
T_i	133		151		120	
Q_i		2571		2655		2940
T_i^2	17.689		22.801		14400	

$T = \sum T_i = 404$

$Q = \sum Q_i = 8166$

$k = 3 \qquad n_2 = 9$

$n_1 = 7 \qquad n_3 = 5$

$$SQE = \sum T_i^2/n_i - T^2/\sum n_i$$

$$= \frac{17689}{7} + \frac{22801}{9} + \frac{14400}{5} - \frac{(404)^2}{7+9+5}$$

$$SQE = 168{,}25$$

$$SQR = Q - \sum T_i/n_i$$

$$= 8166 - \frac{17689}{7} - \frac{22801}{9} - \frac{14400}{5}$$

$$SQR = 222{,}56$$

$$S_E^2 = \frac{SQE}{k-1} = \frac{168,25}{3-1} = 84,13$$

$$S_R^2 = \frac{SQR}{\sum n_i - k} = \frac{222,56}{7+9+5-3} = 12,53$$

$$f_{calc} = \frac{S_E^2}{S_R^2} = \frac{84,13}{12,53} = 6,71$$

$f_\alpha = f_{(k-1);(\sum n_i - k);\alpha} = f_{(3-1);(7+9+5-3);\alpha}$

p/ $\alpha = 1\% \rightarrow f_{2,18,1\%} = 6,01$

p/ $\alpha = 5\% \rightarrow f_{2,18,5\%} = 3,55$

$H_0: \mu_A = \mu_B = \mu_C$

Se $f_{calc} \geq f_\alpha$; rejeita-se H_0

Como $6,71 > 6,01$ e $6,71 > 3,55$; rejeita-se H_0 p/ $\alpha = 1\%$ e $\alpha = 5\%$; ou seja: sim, existe diferença.

ANOVA 2

Análise de Variância com 2 Classificações

Ex: Checar no por turma e por engenharia

- Elementos observados segundo 2 critérios, constituindo 2 classificações cruzadas
- Total de n.k observações, sendo n colunas e k linhas

2 classificações $\begin{cases} \rightarrow k \text{ amostras de } n \text{ elementos} \\ \rightarrow n \text{ amostras de } k \text{ elementos} \end{cases}$

$H_{01}: \mu_1 = \mu_2 = \dots = \mu_k$

$H_{02}: \mu_1 = \mu_2 = \dots = \mu_n$

T_i = soma dos valores da i-ésima linha $= \sum_{j=1}^{n} X_{ij}$

Q_i = soma dos quadrados dos valores da i-ésima linha $= \sum_{j=1}^{n} X_{ij}^2$

T_j = soma dos valores da j-ésima coluna $= \sum_{i=1}^{k} X_{ij}$

Q_j = soma dos quadrados dos valores da j-ésima coluna $= \sum_{i=1}^{k} X_{ij}^2$

$T = \sum_{i=1}^{k} T_i = \sum_{j=1}^{n} T_j$ = soma dos valores

$Q = \sum_{i=1}^{k} Q_i = \sum_{j=1}^{n} Q_j$ = soma total dos quadrados

$\bar{X}_i = \frac{T_i}{n}$ = média da i-ésima linha

$\bar{X}_j = \frac{T_j}{k}$ média da j-ésima coluna

$\bar{\bar{X}} = \frac{T}{nk}$ = média de todos os valores

SQT = soma dos quadrados total = $Q - T/nk$

$$S_T^2 = \frac{\sum\sum (X_{ij} - \bar{\bar{X}})^2}{(nk-1)} = \frac{\sum\sum X_{ij}^2 - [(\sum\sum X_{ij})^2/nk]}{nk-1} = \frac{Q - T^2/nk}{nk-1} = \boxed{\frac{SQT}{nk-1}}$$

SQL: soma dos quadrados das linhas $= \sum_{i=1}^{k} \frac{T_i^2}{n} - \frac{T^2}{nk}$

$$S_i^2 = n \cdot \frac{\sum (\bar{X}_i - \bar{\bar{X}})^2}{k-1} = \frac{\sum T_i^2/n - T^2/nk}{k-1} = \boxed{\frac{SQL}{k-1}}$$

SQC: soma dos quadrados das colunas $= \sum_{j=1}^{n} \frac{T_j^2}{k} - \frac{T^2}{nk}$

$$S_j^2 = \frac{\sum (\bar{X}_j - \bar{\bar{X}})^2}{n-1} = \frac{\sum T_j^2/n - T^2/nk}{n-1} = \boxed{\frac{SQC}{n-1}}$$

SQR: soma dos quadrados residuais $= Q - \sum T_i^2/n - \sum T_j^2/k + T^2/nk$

$$S_R^2 = \boxed{\frac{SQR}{(k-1)(n-1)}}$$

$SQT = SQL + SQC + SQR$

→ Teste para média das linhas:

$H_{01}: \mu_1 = \mu_2 = \dots = \mu_k$ $\quad f_L = \frac{S_i^2}{S_R^2}$ $\quad f_\alpha = f_{(k-1);(n-1)(k-1);\alpha}$

se $f_L > f_\alpha$ rejeita Ho

→ Teste para média das colunas:

$H_{02}: \mu_1 = \mu_2 = \dots = \mu_n$ $\quad F_C = \frac{S_j^2}{S_R^2}$ $\quad f_\alpha = f_{(n-1);(n-1)(k-1);\alpha}$

se $f_C > f_\alpha$ rejeita Ho

Exercício 3:

Cinco modelos de moto com 3 diferentes pneus.
Durabilidades estão na tabela a seguir.
Com significância de 5%, existem diferenças entre modelos de motos e pneus?

	f_1	f_2	f_3	T_i	T_i^2
A	54	57	55	166	27556
B	32	40	36	108	11664
C	38	42	38	118	13924
D	46	45	44	135	18225
E	50	53	51	154	23716
T_j	220	237	224		
T_j^2	48400	56169	50176		

$\begin{cases} k = 5 \\ n = 3 \end{cases}$

$\begin{cases} \alpha = 5\% \\ H_{01}: \mu_A = \mu_B = \dots = \mu_E \\ H_{02}: \mu_1 = \mu_2 = \mu_3 \end{cases}$

Quadrados:

				Q_i
	2916	3249	3025	9190
	1024	1600	1296	3920
	1444	1764	1444	4632
	2116	2025	1936	6077
	2500	2809	2601	7910
Q_j	10000	11447	10302	

- $T = \sum T_i = \sum T_j = 681$
- $Q = \sum Q_i = \sum Q_j = 31799$
- $\sum T_i^2 = 95085$
- $\sum T_j^2 = 154795$

- $SQT = Q - T^2/nk = 31799 - 681^2/15 = 831{,}6$
- $SQL = \sum T_i^2/n - T^2/nk = 95085/3 - 681^2/15 = 777{,}6$
- $SQC = \sum T_j^2/k - T^2/nk = 154795/5 - 681^2/15 = 31{,}6$
- $SQR = 831{,}6 - 777{,}6 - 31{,}6 = 22{,}4$

Finalmente:

$S_L^2 = \frac{SQL}{k-1} = \frac{777{,}6}{4} = 194{,}4$ $\quad S_C^2 = \frac{SQC}{n-1} = \frac{31{,}6}{2} = 15{,}8$ $\quad S_R^2 = \frac{22{,}4}{4 \cdot 2} = 2{,}80$

$f_C = \frac{S_C^2}{S_R^2} = \frac{15{,}8}{2{,}8} = 5{,}64$ $\quad f_L = \frac{S_L^2}{S_R^2} = \frac{194{,}4}{2{,}8} = 69{,}43$

* Teste para médias das linhas (modelos):

Se $f_L > f_\alpha$ rejeito H_0

$f_\alpha = F_{(k-1);(n-1)(k-1);\alpha} = F_{4;8;5\%} = 3{,}84$

Como $69{,}43 > 3{,}84$; rejeito H_{01} ⇒ existem diferenças entre os modelos de motos

* Teste para médias das colunas (pneus):

Se $f_C > f_\alpha$; rejeito H_0

$f_\alpha = F_{(n-1);(n-1)(k-1);\alpha} = F_{2;8;5\%} = 4{,}46$

Como $5{,}64 > 4{,}46$; rejeito H_{02} ⇒ existem diferenças entre os fornecedores de pneus.

ANOVA 3 12/11/15

Teste de análise de variância para duas classificações com repetição de amostragem.

Ex 024 (pg. 163 Costa Neto)

Observados tempos em segundos gastos por funcionários (4) p/ realizar 3 métodos ≠s. Cada operário montou 2 peças por método, verifique se há diferenças significativas entre métodos e/ou operários (c/ 5%).

		operários						
		A x x²	B x x²	C x x²	D x x²	T_i	Q_i	T_i^2
métodos	I	54 -1 1 52 -3 9	46 -9 81 47 -8 64	56 0 0 54 -1 1	51 -4 16 60 5 25	-21	197	441
	II	59 4 16 57 2 4	61 6 0 55 0 0	59 4 16 61 6 36	56 1 1 57 2 4	25	113	625
	III	59 4 16 62 7 49	63 8 64 58 3 9	63 8 64 61 6 36	59 4 16 60 5 25	45	279	2025
	T_j	13	0	23	13	49		3091
	Q_j	95	254		87		589	
	T_j^2	169	0	529	169	867		

x Obtidos subtraindo 55 de todos os valores

$r = 2$

$n = 4$ colunas

$k = 3$ linhas

$T = \sum T_i = \sum T_j = 49$

$Q = \sum Q_i = \sum Q_j = 589$

$\sum T_i^2 = 3091$

$\sum T_j^2 = 867$

$\sum\sum T_{ij}^2 = (-4 + (-17)^2 + (-1)^2 + (1)^2 + (6)^2 + (6)^2 + (10)^2 + (3^2 + (11)^2 + (11)^2 + (14)^2 + (9)^2 = 1007$

variação entre linhas: $SQL = \frac{\sum T_i^2}{nr} - \frac{T^2}{nkr} = \frac{3091}{4.2} - \frac{49^2}{4.3.2} = 286,34$

entre colunas: $SQC = \frac{\sum T_j^2}{kr} - \frac{T^2}{nkr} = \frac{867}{3.2} - \frac{49^2}{4.3.2} = 44,96$

total: $SQT = Q - \frac{T^2}{nkr} = 589 - \frac{49^2}{4.3.2} = 488,96$

entre tratamentos: $SQTr = \frac{\sum\sum T_{ij}^2}{r} - \frac{T^2}{n.k.r} = \frac{1007}{2} - \frac{49^2}{4.3.2} = 403,46$

interação: $SQI = SQTr - SQL - SQC$

$= 403,46 - 286,34 - 44,46 = 72,66$

residual: $SQR = SQT - SQTr = 488,96 - 403,46 = 85,50$

$S_L^2 = \frac{SQL}{k-1} = \frac{236,34}{3-1} = 143,17$

$S_{TR}^2 = \frac{SQTR}{(nk-1)} = \frac{403,46}{(4.3-1)} = 36,68$

$S_C^2 = \frac{SQC}{n-1} = \frac{44,46}{4-1} = 14,82$

$S_R^2 = \frac{SQR}{nk(r-1)} = \frac{85,50}{12.1} = 7,13$

$S_I^2 = \frac{SQI}{(k-1).(n-1)} = \frac{72,66}{(3-1)(4-1)} = 12,4$

$f_I = \frac{S_I^2}{S_R^2} = \frac{12,11}{7,13} = 1,70$

$f_{crit} = f_{(k-1)(n-1);\, nk(r-1);\, \alpha}$

Se $F_I > F_\alpha$ existe interação

$f_{(3-1)(4-1);\ 4.3(2-1);\ 5\%} = f_{6;12;5\%} = 3{,}00$

$1{,}70 < 3{,}00$ então não existe interação

→ Se houver interação:

$f_L = \frac{S_L^2}{S_I^2}$ $\quad f_C = \frac{S_C^2}{S_I^2}$

→ Se não houver interação:

$f_L = \frac{S_L^2}{S_R^2}$ $\quad f_C = \frac{S_C^2}{S_R^2}$

* S_R^2 é recalculado incorporando a interação

Como não existe interação, SQI é anexada ao SQR

$SQR = 85{,}50 + 72{,}66 = 158{,}16$

$G.L. = 12 + 6 = 18$

$\Rightarrow S_R^2 = \frac{158{,}16}{18} = 8{,}79$

$f_L = \frac{S_L^2}{S_R^2} = \frac{143{,}17}{8{,}79} = 16{,}29$

$f_C = \frac{S_C^2}{S_R^2} = \frac{14{,}82}{8{,}79} = 1{,}69$

- $f_{2;18;5\%} = 3{,}55$
- $f_{3;18;5\%} = 3{,}16$

$H_{01}: \mu_I = \mu_{II} = \mu_{III}$

$H_{02}: \mu_A = \mu_B = \mu_C = \mu_D$

Se $f_{calc} > f_\alpha$: rejeita-se H_0

∴ como $f_L = 16{,}29 > 3{,}55$: rejeita-se H_{01},
ou seja, métodos são ≠s

∴ como $f_C = 1{,}69 < 3{,}16$: aceita-se H_{02},
ou seja, operários são =s

Comparações múltiplas:

Exercício 07 (Dados: Exemplo 1)

Concreto A: $\bar{X}_1 = 72{,}0$	ANOVA:	$k = 3$	$\alpha = 5\%$
" B: $\bar{X}_2 = 66{,}8$	$n = 5$	$S_R^2 = 10{,}73$	$r = k(n-1)$
" C: $\bar{X}_3 = 75{,}0$	$\lvert\bar{X}_1 - \bar{X}_2\rvert = 5{,}2$	$\lvert\bar{X}_1 - \bar{X}_3\rvert = 3{,}0$	$\lvert\bar{X}_2 - \bar{X}_3\rvert = 8{,}2$

* Método de Tukey:

$$\lvert\bar{X}_1 - \bar{X}_m\rvert > q_{k;r;\alpha} \cdot \sqrt{\frac{S_R^2}{n}}$$

$q_{k,r,\alpha} = q_{3,12,5\%} = 3{,}77$ (pág. Costa Neto) $\qquad 3{,}77 \cdot \sqrt{\frac{10{,}73}{5}} = 5{,}52$

como apenas $\lvert\bar{X}_2 - \bar{X}_3\rvert = 8{,}2 > 5{,}52$, existe diferença somente entre os concretos B e C.

* Método de Scheffé

$$|\bar{x}_i - \bar{x}_m| > \sqrt{S_R^2 \cdot \frac{2(k-1)}{n} \cdot f_{k-1;\, k(n-1);\, \alpha}}$$

$f_{k-1;\, k(n-1);\, \alpha} = f_{3-1;\, 3(5-1);\, 5\%} = f_{2,12,5\%} = 3{,}89$

$\sqrt{10{,}73 \cdot \frac{2 \cdot (3-1)}{5} \cdot 3{,}89} = 5{,}78$

como apenas $|\bar{x}_2 - \bar{x}_3| = 8{,}2 > 5{,}78$ existe diferença somente entre concretos B e C a 5%

Como no exemplo $5{,}78 > 5{,}52$, o método de Tukey é mais poderoso, mais crítico, ele pegaria diferenças menores.

Exercício 04: → só a resposta FAZER depois

(ANOVA c/ 2 classif. s/ repetição)

fonte de variação	graus de liberdade	quadrado médio	Fcalc	Fα	Resultado
SQL = 2921,96	7	S_L^2 = 417,42	F_L = 4,64	2,19	Rejeita-se H_{01}
SQC = 31,58	2	S_C^2 = 15,79	F_C = 0,18	2,73	Rejeita-se H_{02}
SQR = 1260,92	14	S_R^2 = 90,3			
SQT = 4213,96	23				

Exercício 5 (ANOVA c/ 2 classif. s/ repetição)

fonte de variação	G.L.	quadrado médio	Fcalc	Fα	Resultado
SQL = 214,92	3	S_L^2 = 71,64	F_L = 7,09	4,76	Rejeita-se H_{01}
SQC = 7,17	2	S_C^2 = 3,59	F_C = 0,35	5,14	Aceita-se H_{02}
SQR = 60,83	6	S_R^2 = 10,14			
SQT = 282,92	11				

Exercício 6: (ANOVA c/ 2 classif c/ repetição)

fonte de variação	graus de liberdade	quadrado médio	Fcalc	Fα
SQL = 6,85	1	S_L^2 = 6,85		
SQC = 0,62	1	S_C^2 = 0,62		
SQI = 0,17	1	S_I^2 = 0,17	F_I = 0,16	4,49 (5%)
SQTr = 7,64	3	S_{Tr}^2 = 2,55		
SQR = 16,57	16	S_R^2 = 1,04		
SQT = 24,21				

como $F_I < F_\alpha$ não existe interação

$S_R^2 = 0{,}98$

$F_L = 6{,}99$

$F_C = 0{,}63$

$F_\alpha = 4{,}45$

Rejeita H_{01}

Aceita H_{02}

Correlação

16/11/15

→ Duas variáveis quantitativas de interesse

→ Coeficiente de correlação

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} \qquad -1 \leq r \leq +1$$

- $S_{xy} = \sum x_i y_i - \frac{\sum x_i \cdot \sum y_i}{n}$
- $S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$
- $S_{xy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$

$r > 0$: correlação positiva
$r < 0$: correlação negativa
$|r| = 1$: correlação perfeita
$r = 0$: ausência de correlação

Regressão

→ Reta estimativa de regressão

$$\hat{y} = a + bx$$

- b = coeficiente angular ou coeficiente de regressão (representa o declive da reta)
- a = coeficiente linear (representa o intercepto da reta com o eixo vertical).

$$\boxed{b = \frac{S_{xy}}{S_{xx}}} \qquad \boxed{a = \bar{y} - b\bar{x}}$$

Exercício 1:

Pessoas (x)	Dias (y)	x^2	y^2	xy
6	4,62	36	21,34	27,72
6	4,5	36	20,25	27
6	4,43	36	19,62	26,58
6	4,81	36	23,14	28,86
7	4,12	49	16,97	28,84
7	3,88	49	15,05	27,16
7	4,01	49	16,08	28,07
7	3,67	49	13,47	25,69
8	3,21	64	10,30	25,68
8	3,05	64	9,30	24,4
8	3,10	64	9,99	25,28
8	3,30	64	10,89	26,4
9	2,86	81	8,18	25,74
9	2,53	81	6,40	22,77
9	2,71	81	7,34	24,39
9	2,62	81	6,86	23,58
10	1,83	100	3,35	18,3
10	2,02	100	4,08	20,2
10	2,24	100	5,02	22,4
10	1,95	100	3,80	19,5
160	65,52	1320	231,45	498,56

$$S_{xy} = \sum x \cdot y - \frac{(\sum x \cdot \sum y)}{n} = 498,56 - \frac{(160 \cdot 65,52)}{20} \Rightarrow S_{xy} = -25,6$$

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 1320 - \frac{160^2}{20} \Rightarrow S_{xx} = 40$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 231,45 - \frac{(65,52)^2}{20} \Rightarrow S_{yy} = 16,81$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{-25{,}6}{\sqrt{40 \cdot 16{,}81}} \Rightarrow r = -0{,}9872$$

Correlação forte negativa, coef. de correlação é negativo e próximo de 1 em módulo.

$$b = \frac{S_{xy}}{S_{xx}} = \frac{-25{,}6}{40} \Rightarrow b = -0{,}64$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{160}{20} \Rightarrow \bar{x} = 8$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{65{,}52}{20} \Rightarrow \bar{y} = 3{,}28$$

$$a = \bar{y} - b\bar{x} = 3{,}28 - 8 \cdot (-0{,}64)$$

$$a = 8{,}4$$

$$\hat{y} = a + bx = 8{,}4 - 0{,}64x$$

p/ $y = 4{,}5$ dias

$$0{,}64x = 8{,}4 - 4{,}5 \Rightarrow x = 6{,}09$$

7 recursos

Teste p/ coef. de correlação

$H_0: \rho = 0$

$H_1: \rho \neq 0 \rightarrow \rho < 0$ ou $\rho > 0$

onde ρ é o coef. de correlação populacional

$$t_{calc} = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$

$$t_{crítico} = t_{n-2;\,\alpha \text{ ou } \alpha/2}$$

$H_1: \rho < 0$ ou $\rho > 0$ $\quad$ $H_1: \rho \neq 0$

se $t_{calc} > t_{crítico}$; rejeito H_0, ou seja, existe correlação (positiva ou negativa)

Exercício 2:

Altura: x

Peso: y

$z = x - 170$

$w = y - 75$

z	w	z^2	w^2	zw
4	-2	16	4	-8
-9	-9	81	81	81
0	-11	0	121	0
10	19	100	361	190
12	4	144	16	48
-6	-3	36	9	18
-14	-13	196	169	182
-2	-11	4	121	22
6	15	36	225	90
5	6	25	36	30
6	-5	638	1143	653

$$S_{xy} = \sum x_i y_i - \frac{(\sum x_i \cdot \sum y_i)}{n}$$

$$= 653 - \frac{(6 \cdot (-5))}{10} = 656$$

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

$$= 638 - \frac{(6)^2}{10} = 634{,}4$$

$$S_{yy} = S_{ww} = \sum w^2 - \frac{(\sum w)^2}{n}$$

$$= 1143 - \frac{(-5)^2}{10} = 1140{,}5$$

$$r = \frac{S_{xy}}{\sqrt{S_x \; S_{yy}}} = \frac{656}{\sqrt{634{,}4 \cdot 1140{,}5}} \Rightarrow r = 0{,}7712$$

$H_0: \rho = 0$

$H_1: \rho > 0$ (correlação positiva)

$\alpha = 5\%$

$$t_{calc} = r \cdot \sqrt{\frac{n-2}{1-r^2}} = 0{,}7712 \cdot \sqrt{\frac{10-2}{1-(0{,}7712)^2}}$$

$t_{calc} = 3{,}41$

$t_{critico} = t_{n-2;\alpha} = t_{8;5\%} = 1{,}86$

Como $t_{calc} > t_{crit.}$; rejeita-se H_0, ou seja, existe correlação positiva

17/11/2015

Variação Residual: $S_R^2 = \dfrac{S_{YY} - b\,S_{XY}}{n-2}$

Variância de b: $S^2(b) = \dfrac{S_R^2}{S_{XX}}$

Variância de a: $S^2(a) = \dfrac{S_R^2 \cdot \sum x_i^2}{n \cdot S_{XX}}$

Intervalo de confiança p/ coef. angular da reta (b) (*)

$$\beta = b \pm t_{n-2;\alpha/2} \cdot S(b)$$

Intervalo de confiança p/ a constante α (*)

$$\alpha = a \pm t_{n-2;\alpha/2} \cdot S(a)$$

Intervalo p/ a reta Teórica ($\alpha + \beta x_i$) ou valor médio de y:

$$\hat{y}' \pm t_{n-2;\alpha/2} \cdot \sqrt{S_R^2\left(\frac{1}{n} + \frac{(x' - \bar{x})^2}{S_{XX}}\right)}$$

Intervalo de confiança p/ previsão $y_i = \alpha + \beta x_i + \varepsilon_i$:

$$\hat{y}' \pm t_{n-2;\alpha/2} \cdot S_R \sqrt{1 + \frac{1}{n} + \frac{(x' - \bar{x})^2}{S_{XX}}}$$

Exercício 03:

x	y	x^2	y^2	$x \cdot y$
1	0,5	1	0,25	0,5
2	0,6	4	0,36	1,2
3	0,9	9	0,81	2,7
4	0,8	16	0,64	3,2
5	1,2	25	1,44	6,0
6	1,5	36	2,25	9,0
7	1,7	49	2,89	11,9
8	2,0	64	4,00	16,0
36	9,20	204	12,64	50,50

$$S_{XY} = \sum x.y - \frac{(\sum x . \sum y)}{n}$$

$$S_{XY} = 50{,}50 - \frac{(36 . 9{,}2)}{8} = 9{,}1$$

$$S_{XX} = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$S_{XX} = 204 - \frac{(36)^2}{8} = 42$$

$$S_{YY} = \sum y^2 - \frac{(\sum y)^2}{n}$$

$$S_{YY} = 12{,}64 - \frac{(9{,}2)^2}{8} = 2{,}06$$

$$\bar{x} = \frac{\sum x}{n} = \frac{36}{8} = 4{,}5 \qquad \bar{y} = \frac{\sum y}{n} = \frac{9{,}20}{8} = 1{,}15$$

$$b = \frac{S_{xy}}{S_{xx}} = \frac{9{,}1}{42} = 0{,}2167 \qquad a = \bar{y} - b\bar{x} = 1{,}15 - 0{,}2167 \cdot 4{,}5 = 0{,}17$$

$$\hat{y} = a + bx \Rightarrow \hat{y} = 0{,}1799 + 0{,}2167x$$

IC p/ α e β: $S_R^2 = \dfrac{S_{YY} - bS_{XY}}{n-2} = \dfrac{2{,}06 - 0{,}2167 \cdot 9{,}9}{8-2} = 0{,}0142$

$$S^2(b) = \frac{S_R^2}{S_{xx}} = \frac{0{,}0142}{42} = 0{,}00035$$

$$S^2(a) = \frac{S_R^2 \cdot \sum x^2}{n \cdot S_{xx}} = \frac{0{,}0142 \cdot 209}{8 \cdot 42} = 0{,}008925$$

* $\beta = b \pm t_{n-2;\,\alpha/2} \cdot S(b)$

$$0{,}2167 \pm \underset{\substack{\shortparallel \\ 2{,}447}}{t_{8-2;\,2{,}5\%}} \cdot \sqrt{0{,}00035} = 0{,}2167 \pm 0{,}0458$$

$$P(0{,}1709 \leq \beta \leq 0{,}2625) = 0{,}95$$

* $\alpha = a \pm t_{n-2;\,\alpha/2} \cdot S(a)$

$$0{,}1799 \pm \underset{\substack{\shortparallel \\ 2{,}447}}{t_{8-2;\,2{,}5\%}} \cdot \sqrt{0{,}008925} = 0{,}1799 \pm 0{,}2312$$

$$P(-0{,}0563 \leq \alpha \leq 0{,}4061) = 0{,}95$$

* $y' = 0{,}1799 + 0{,}2167 \cdot x'$

p/ $x = 10$:

$$\hat{y}' = 0{,}1799 + 0{,}2167 \cdot 10 = 2{,}3419$$

* Intervalo p/ valor médio de y:

$$\hat{y}' \pm t_{n-2;\,\alpha/2} \cdot \sqrt{S_R^2\left(\frac{1}{n} + \frac{(x' - \bar{x})^2}{S_{xx}}\right)}$$

$$\hat{y}' = 2{,}3419 \pm 0{,}2728 \qquad P(2{,}0691 \leq \alpha + 10\beta \leq 2{,}6147) = 0{,}95$$

* Intervalo p/ a previsão:

$$\hat{y}' \pm t_{n-2;\,\alpha/2} \cdot S_R \sqrt{1 + \frac{1}{n} + \frac{(x' - \bar{x})^2}{S_{xx}}}$$

$$2{,}3419 \pm t_{6;\,2{,}5\%} \cdot 0{,}1212 \sqrt{1 + \frac{1}{8} + \frac{(10 - 4{,}5)^2}{42}}$$

$$2{,}3419 \pm 0{,}4029$$

$$P(1{,}9390 \leq y' \leq 2{,}7448) = 0{,}95$$

* Teste de hipótese p/ β:

$$t_{calc} = \frac{b - \beta_0}{S_R/\sqrt{S_{xx}}} = \frac{b - \beta_0}{S(b)}$$

$t_{crítico} = t_{n-2;\,\alpha}$ ou $\alpha/2$

* Teste de hipótese p/ α:

$$t_{calc} = \frac{a - \alpha_0}{S(a)}$$

$t_{crítico} = t_{n-2;\,\alpha}$ ou $\alpha/2$

Exemplo 2 a)

$H_0: \beta = 0{,}1$

$H_1: \beta > 0{,}1$

$S_R^2 = 0{,}0197$

$S_{xx} = 42$

$S^2(b) = 0{,}00035$

$b = 0{,}2167$

$n = 8$

$\alpha = 5\%$

$$t_{calc} = \frac{b - \beta_0}{S(b)}$$

$$t_{calc} = \frac{0{,}2167 - 0{,}10}{\sqrt{0{,}00035}}$$

$t_{calc} = 6{,}24$

$t_{crítico} = t_{n-2;\,\alpha} = t_{6;\,5\%} = 1{,}943$

Como $t_{calc} > t_{crit.}$, rejeita-se H_0; ou seja pode-se afirmar que a inclinação da reta é > 10% a 5% de significância

Exemplo 2b:

$H_0: \alpha = 0$

$H_1: \alpha \neq 0$

$a = 0{,}1799$

$\alpha_0 = 0$

$S^2(a) = 0{,}0089$

$$t_{calc} = \frac{a - \alpha_0}{S(a)} = \frac{0{,}1799 - 0}{\sqrt{0{,}0089}} = 1{,}854$$

$t_{crit.} = t_{n-2;\,\alpha/2} = t_{6;\,2{,}5\%}$

$t_{crit} = 2{,}447$

Como $t_{calc} < t_{crit.}$, aceito H_0; pode-se afirmar que a reta passa pela origem.

19/11/2015

ANOVA aplicada à regressão:

$SQT = S_{YY}$

$SQE = b^2 . S_{xx}$

$SQR = SQT - SQE = S_{YY} - b^2 . S_{xx}$

- Variância total: $S_Y^2 = S_{YY}/(n-1)$
- Variância residual: $S_R^2 = (S_{YY} - b^2 . {}_{xx})/(n-2)$
- Variância da regressão: $S_E^2 = b^2 . S_{xx}$

$$f_{calc} = \frac{S_E^2}{S_R^2} = \frac{b^2 . S_{xx}}{S_R^2}$$

$f_{crit} = f_{1;\,n-2;\,\alpha}$

↳ sempre 1

Se $f_{calc} > f_{crit}$, existe regressão

Coeficiente de determinação: (*)

$$R^2 = \frac{SQE}{SQT} = \frac{SQT - SQR}{SQT}$$

Regressão Polinomial

$$\hat{y}_P = a + b(x - \bar{x}) + c(x - \bar{x})^2$$

$$\begin{cases} \sum y_i = na + c.\sum (x_i - \bar{x})^2 \\ \sum (x_i - \bar{x}).y_i = b.\sum (x - \bar{x})^2 \\ \sum (x_i - \bar{x})^2.y_i = a.\sum (x_i - \bar{x})^2 + c.\sum (x_i - \bar{x})^4 \end{cases}$$

Exercício 4:

x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})y_i$	$(x_i - \bar{x})^2 y_i$	$(x_i - \bar{x})^4$
1	0,5	-3,5	12,25	-1,75	6,125	150,0625
2	0,6	-2,5	6,25	-1,50	3,750	39,0625
3	0,9	-1,5	2,25	-1,35	2,025	5,0625
4	0,8	-0,5	0,25	-0,40	0,200	0,0625
5	1,2	0,5	0,25	0,60	0,300	0,0625
6	1,5	1,5	2,25	2,25	3,375	5,0625
7	1,7	2,5	6,25	6,25	10,625	39,0625
8	2,0	3,5	12,25	12,25	24,500	150,0625
36	9,20		42	9,10	50,9	388,5

* $\sum (x_i - \bar{x}).y_i = b.\sum (x_i - \bar{x})^2$

$9,10 = b.42$

$b = 0,2167$

* $\sum (x_i - \bar{x})^2 y_i = a.\sum (x_i - \bar{x})^2 + c\sum (x_i - \bar{x})^4$

$50,9 = a.42 + c.388,5$ (I)

* $\sum y_i = n.a + c.\sum (x_i - \bar{x})^2$

$9,20 = 8.a + c.42$ (II)

De (I) e (II) $\Rightarrow$ $a = 1,2686$ e $c = 0,0155$

- Regressão Polinomial:

$$\hat{y}_P = a + b(x - \bar{x}) + c(x - \bar{x})^2$$

$$\hat{y}_P = 1,0686 + 0,2167(x - 9,5) + 0,0155(x - 9,5)^2$$

$$\downarrow$$

$$\hat{y}_P = 0,408 + 0,0772x + 0,0155x^2$$

Análise de Melhoria:

H_0: não houve melhoria no ajuste

H_1: houve melhoria

Se $F_{calc} > F_{crit.}$; rejeita-se H_0; houve melhoria significativa do ajuste

fontes de variação:

- da melhoria: $\sum (\hat{y}_{Pi} - \hat{y}_i)$
- residual da parábola: $\sum (y_i - \hat{y}_{Pi})^2$
- residual da reta: $S_{YY} - b^2 S_{XX}$

Quadrados Médios:

- da melhoria: $\Sigma(\hat{y}_{Pi} - \hat{y}_i)^2$
- residual da parábola: $S_P^2 = \dfrac{\Sigma(y_i - \hat{y}_{Pi})^2}{n-3}$

$$f_{calc} = \frac{\Sigma(\hat{y}_{Pi} - \hat{y}_i)^2}{S_P^2} \qquad f_{crit.} = f_{1,\,n-3,\,\alpha}$$

(*)

Exercício 05:

x_i	y_i	$\hat{y}_{Pi}$	$y_i - \hat{y}_{Pi}$	$(y_i - \hat{y}_{Pi})^2$
1	0,5	0,501	-0,001	0,000001
2	0,6	0,624	-0,024	0,000576
3	0,9	0,779	0,121	0,014641
4	0,8	0,965	-0,165	0,027225
5	1,2	1,181	0,019	0,000361
6	1,5	1,429	0,071	0,005041
7	1,7	1,708	-0,008	0,000064
8	2,0	2,018	-0,018	0,000324

fonte de variação residual da parábola:

$\Sigma(y_i - \hat{y}_{Pi})^2 = 0{,}048233$

p/ reta:

$S_{YY} = 2{,}06$

$S_{XX} = 42$

$b = 0{,}2167$

$-b^2 S_{XX} = 2{,}06 - (0{,}2167)^2 \cdot 42 = 0{,}0877$

fonte de variação p/ reta

→ fonte de variação da melhoria:

reta - parábola $= 0{,}0877 - 0{,}048233 = 0{,}0395$

→ quadrado médio melhoria $= 0{,}0395$

→ quadrado médio residual da parábola:

$$S_P^2 = \frac{\Sigma(y_i - \hat{y}_{Pi})^2}{n-3} = \frac{0{,}048233}{8-3} = 0{,}0096$$

- $f_{calc} = \dfrac{\Sigma(\hat{y}_{Pi} - \hat{y}_i)^2}{S_P^2} = \dfrac{0{,}0395}{0{,}0096} = 4{,}11$
- $f_{crit} = f_{1;\,n-3;\,\alpha} = f_{1;\,5;\,5\%} = 6{,}6$

∴ Como $f_{calc} < f_{crit}$, aceita-se H_0, ou seja, não houve melhoria.
Usamos a reta p/ representar o fenômeno.

Exercício 06:

$(y_i - \hat{y}_{Pi})^2$ Quadrática	$(y_i - y_{ei})^2$ Cúbica
2,958	0,757
0,922	0,096
0,044	0,176
0,436	0,048
5,103	2,016
1,000	0,005
0,137	0,096
3,423	1,877
0,397	1,44
0,423	0,109
14,9369	6,621

y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1,5	-18,15	329,4
3	-16,65	277,2
6	-13,65	186,3
10	-9,65	93,12
15	-4,65	21,62
20	0,35	0,123
25	5,35	28,62
30	10,35	107,1
40	20,35	414,1
46	26,35	694,3
196,5		2152,025

↳ $\bar{y} = 19,65$

- $SQR_q = 14,9369$
- $SQR_c = 6,621$
- $SQT = \sum (y_i - \bar{y})^2 = S_{yy} = 2152,025$

FONTES DE VARIAÇÃO	SQ	GL	QM	f_{calc}
MELHORIA	8,3159	1	8,3159	7,5359
CÚBICA	6,621	$n-4=6$	1,1035	
QUADRÁTICA	14,9369	$n-3=7$		

$f_{crit} = f_{1;n-4;\alpha} = f_{1;6;5\%} = 5,99$

∴ Como $f_{calc} > f_{crit.}$, rejeito H_0, ou seja houve melhoria, a cúbica é melhor que a quadrática.