UFV- CCE - DET

EST 105 - 1^a avaliação - 2^0 semestre de 2017 - 2/set/17

Nome: ______________________________________ **Matrícula:**____________________

Assinatura: ______________________________. Favor apresentar documento com foto.

- São 4 questões e formulário em páginas numeradas de 1 a 8, total de 30 pontos, FAVOR CONFERIR ANTES DE INICIAR.

- ATENÇÃO: Assinale (X) em qual turma está matriculado (sua nota será divulgada no sistema SAPIENS).

	TURMA	HORÁRIO	SALA	PROFESSOR
( )	T1	2ª 10-12 5ª 08-10	PVB310	Moysés
( )	T2	2ª 16-18 5ª 14-16	PVB310	Carol
( )	T3	2ª 08-10 4ª 10-12	PVB109	Paulo
( )	T5	3ª 16-18 6ª 14-16	PVB310	Camila
( )	T6	2ª 14-16 4ª 16-18	PVB107	Carol/Moysés
( )	T7	4ª 08-10 6ª 10-12	PVB206	CHOS - coordenador
( )	T8	2ª 18:30-20:10 4ª 20:30-22:10	PVB210	Eduardo
( )	T9	3ª 10-12 PVB300 6ª 08-10	PVB307	Cecon
( )	T20 = EST 085 T1 2ª14-16 PVA284 T2 2ª18:30-20:10 PVA388 Leísa			

- Interpretar corretamente as questões é parte da avaliação, portanto não é permitido questionamentos durante a prova !

- É OBRIGATÓRIO APRESENTAR OS CÁLCULOS organizadamente, para ter direito à revisão.

- PODE UTILIZAR A CALCULADORA, porém mostre os valores utilizados na fórmula.

- BOA SORTE e BOA PROVA !!!.

FORMULÁRIO

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad \text{ou} \quad \overline{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i} \qquad Md_X = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} \quad \text{ou} \quad Md_X = X_{\left(\frac{n+1}{2}\right)}$$

$$\overline{X}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}} \quad \text{ou} \quad \overline{X}_H = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \frac{f_i}{X_i}} \qquad \overline{X}_G = \sqrt[n]{\prod_{i=1}^{n} X_i} \quad \text{ou} \quad \overline{X}_G = \sqrt[\sum f_i]{\prod_{i=1}^{k} X_i^{f_i}}$$

$$SQD_X = \sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n} \quad \text{ou} \quad SQD_X = \sum_{i=1}^{k} f_i X_i^2 - \frac{\left(\sum_{i=1}^{k} f_i X_i\right)^2}{\sum_{i=1}^{k} f_i}$$

$$S_X^2 = \frac{SQD_X}{n-1} \quad \text{ou} \quad S_X^2 = \frac{SQD_X}{\sum_{i=1}^{k} f_i - 1}$$

$$S_X = \sqrt{S_X^2} \quad S(\overline{X}) = \frac{S_X}{\sqrt{n}} \quad CV_X(\%) = \frac{S_X}{\overline{X}} 100\%$$

$$\widehat{\rho}_{XY} = r_{XY} = \frac{SPD_{XY}}{\sqrt{SQD_X \ SQD_Y}} \qquad SPD_{XY} = \sum_{i=1}^{n} X_i Y_i - \frac{\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n}$$

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad \widehat{\varepsilon}_i = Y_i - \widehat{Y}_i$$

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i = b_0 + b_1 X_i \qquad b_1 = \widehat{\beta}_1 = \frac{SPD_{XY}}{SQD_X} = r_{XY} \frac{S_Y}{S_X} \qquad b_0 = \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}$$

$$r^2(\%) = \frac{SQ\text{regressão}}{SQ\text{total}} 100\%$$

$$SQ\text{regressão} = \widehat{\beta}_1^2 SQD_X = \widehat{\beta}_1 SPD_{XY} = (SPD_{XY})^2 / SQD_X \qquad SQ\text{total} = SQD_Y$$

1.(5 pontos) Dado que,

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} \quad \text{e} \quad \sum_{k=1}^{n} k^3 = \left[\frac{n(n+1)}{2}\right]^2$$

utilize as propriedades de somatório e calcule: $\displaystyle\sum_{k=3}^{40} (k-1)^3$.

Solução: Sabemos que

$$\begin{aligned}
(a-b)^3 &= (a-b)^2(a-b) = \left(a^2 - 2ab + b^2\right)(a-b) \\
&= \left(a^2 - 2ab + b^2\right)(a-b) = a^3 - 2a^2b + ab^2 - a^2b + 2ab^2 - b^3 \\
&= a^3 - 3a^2b + 3ab^2 - b^3
\end{aligned}$$

Em nosso problema $a = k$ e $b = 1$, assim

$$(k-1)^3 = k^3 - 3\cdot k^2 \cdot 1 + 3 \cdot k \cdot 1^2 - 1^3 = k^3 - 3k^2 + 3k - 1$$

Modo 1:

Fazendo-se $k - 1 = t$ temos:

- se $k = 3$ então $t = 2$;
- se $k = 40$ então $t = 39$;

$$\begin{aligned}
\sum_{k=3}^{40} (k-1)^3 &= \sum_{t=2}^{39} t^3 = \left(\sum_{t=1}^{1} t^3\right) - \left(\sum_{t=1}^{1} t^3\right) + \sum_{t=2}^{39} t^3 = \sum_{t=1}^{39} t^3 - \left[\frac{1\,(1+1)}{2}\right]^2 \\
&= \left[\frac{39\,(39+1)}{2}\right]^2 - 1 = (780)^2 - 1 \\
&= 608400 - 1 \\
&= 608399
\end{aligned}$$

Modo 2:

$$\begin{aligned}
\sum_{k=3}^{40} (k-1)^3 &= \sum_{k=3}^{40} \left(k^3 - 3k^2 + 3k - 1\right) = \sum_{k=3}^{40} k^3 - 3\sum_{k=3}^{40} k^2 + 3\sum_{k=3}^{40} k - \sum_{k=3}^{40} 1 \\
&= \sum_{k=3}^{40} k^3 + \sum_{k=1}^{2} k^3 - \sum_{k=1}^{2} k^3 - 3\left(\sum_{k=3}^{40} k^2 + \sum_{k=1}^{2} k^2 - \sum_{k=1}^{2} k^2\right) \\
&+ 3\left(\sum_{k=3}^{40} k + \sum_{k=1}^{2} k - \sum_{k=1}^{2} k\right) - 38 \\
&= \sum_{k=1}^{40} k^3 - \sum_{k=1}^{2} k^3 - 3\left(\sum_{k=1}^{40} k^2 - \sum_{k=1}^{2} k^2\right) + 3\left(\sum_{k=1}^{40} k - \sum_{k=1}^{2} k\right) - 38 \\
&= \left[\frac{40\,(40+1)}{2}\right]^2 - \left[\frac{2\,(2+1)}{2}\right]^2 - 3\left(\frac{40\,(40+1)\,(2\times 40+1)}{6}\right. \\
&- \left.\frac{2\,(2+1)\,(2\times 2+1)}{6}\right) + 3\left(\frac{40\,(40+1)}{2} - \frac{2\,(2+1)}{2}\right) - 38 \\
&= (672400 - 9) - 3\,(22140 - 5) + 3\,(820 - 3) - 38 \\
&= 672391 - 66405 + 2451 - 38 \\
&= 608399
\end{aligned}$$

2.(5 pontos) Dado que,

$$\sum_{\substack{k=1\\k\neq 3,6}}^{20} Z_k = 10, \qquad \sum_{j=3}^{15} Y_j = 15, \qquad \sum_{j=3}^{15} Y_j^2 = 25, \qquad \sum_{i=1}^{10} X_i = 20, \qquad \sum_{i=1}^{10} X_i^2 = 50,$$

utilize as propriedades de somatório e calcule:

$$\sum_{\substack{k=1\\k\neq 3,6}}^{20} \sum_{j=3}^{15} \sum_{i=1}^{10} \left[(X_i - Y_j)^2 + Z_k\right].$$

Solução: Temos que:

$$\begin{aligned} NT_i &= 10 - 1 + 1 - 1 = 10 \\ NT_j &= 15 - 3 + 1 - 3 = 13 \\ NT_k &= 20 - 1 + 1 - 1 = 18 \end{aligned}$$

Seja $A = \sum_{\substack{k=1\\k\neq 3,6}}^{20} \sum_{j=3}^{15} \sum_{i=1}^{10} \left[(X_i - Y_j)^2 + Z_k\right]$.

$$\begin{aligned} A &= \sum_k \sum_j \sum_i \left[X_i^2 - 2X_iY_j + Y_j^2 + Z_k\right] \\ &= \sum_k \sum_j \sum_i X_i^2 - 2\sum_k \sum_j \sum_i X_iY_j + \sum_k \sum_j \sum_i Y_j^2 + \sum_k \sum_j \sum_i Z_k \\ &= \sum_k 1 \sum_j 1 \sum_i X_i^2 - 2\sum_k 1 \sum_j Y_j \sum_i X_i + \sum_k 1 \sum_j Y_j^2 \sum_i 1 + \sum_k Z_k \sum_j 1 \sum_i 1 \\ &= NT_kNT_j \sum_i X_i^2 - 2NT_k \sum_j Y_j \sum_i X_i + NT_k \sum_j Y_j^2 NT_i + \sum_k Z_k NT_j NT_i \\ &= 18 \cdot 13 \cdot 50 - 2 \cdot 18 \cdot 15 \cdot 20 + 18 \cdot 25 \cdot 10 + 10 \cdot 13 \cdot 10 \\ &= 11700 - 10800 + 4500 + 1300 \\ &= 6700 \end{aligned}$$

3.(10 pontos) A tabela abaixo contém dados relativos a apartamentos de 2 quartos disponíveis para venda em uma imobiliária de Viçosa.

Apartamento (i)	1	2	3	4	5	6
Área (m^2)	69	75	84	59	97	72
Preço (R\$ $\times$ 1000)	400	420	480	520	557	589

Admita X = Área e Y = Preço, pede-se:

a.(1 pt) Utilize sua calculadora para informar as seguintes somas:

$$\sum X = 456, \sum Y = 2966, \sum X^2 = 35516, \sum Y^2 = 1494370, \sum XY = 226537$$

b.(2 pts) Informe as áreas média e mediana dos apartamentos à venda.

$$\bar{X}_A = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{456}{6} = 76$$

$$Md_A = \frac{X_{(3)} + X_{(4)}}{2} = \frac{72 + 75}{2} = \frac{147}{2} = 73,5$$

c.(1 pt) Qual é a média geométrica dos preços dos apartamentos?

$$\bar{Y}_G = \sqrt[n]{\prod_{i=1}^{n} Y_i} = \sqrt[6]{400 \cdot 420 \cdot 480 \cdot 520 \cdot 557 \cdot 589} = 489,50$$

d.(2 pts) Qual das amostras é mais homogênea (área ou preço)? Justifique sua resposta.

$$\begin{aligned}
s_A^2 &= \frac{35516 - \frac{(456)^2}{6}}{6-1} = \frac{860}{5} = 172 \\
s_A &= \sqrt{s_A^2} = \sqrt{172} = 13,11 \\
s_P^2 &= \frac{1494370 - \frac{(2966)^2}{6}}{6-1} = \frac{28177,33}{5} = 5635,42 \\
s_P &= \sqrt{s_P^2} = \sqrt{5635,42} = 75,07 \\
CV_A &= \frac{13,11}{76} \cdot 100\% = 17,25\% \\
CV_P &= \frac{75,07}{494,33} \cdot 100\% = 15,18\%
\end{aligned}$$

A amostra de preços é mais homogênea pois possui menor coeficiente de variação.

e.(2 pts) Qual é o erro padrão da média das áreas dos apartamentos à venda?

$$s\left(\bar{X}_A\right) = \frac{13,11}{\sqrt{6}} = 5,3521$$

f.(2 pts) Calcule o coeficiente de correlação amostral entre a área e o preço dos apartamentos.

$$SDQ_A = 860$$

$$SDQ_P = 28177,33$$

$$SPD_{A,P} = 226537 - \frac{456 \cdot 2966}{6} = 1121$$

$$r_{A,P} = \frac{1121}{\sqrt{860 \cdot 28177,33}} = \frac{1121}{4922,65} \cong 0,2277$$

4.(10 pontos) Investigou-se, via regressão linear simples (RLS), como a idade dos consumidores (X, em anos) poderia influenciar a intenção de compra de produtos de uma determinada linha. Para tal, n = 8 indivíduos foram apresentados a fotografias que ilustravam alguns desses produtos. As notas de preferência (Y), definidas em uma escala contínua de 1 a 6, foram atribuídas a cada imagem, conforme apresentado na tabela a seguir:

i	1	2	3	4	5	6	7	8
X_i	18	19	20	23	24	25	26	27
Y_i	5,4	5,6	5,3	4,7	3,9	4,0	3,7	3,3

Pede-se:

a.(4 pts) A equação de regressão linear simples (RLS) ajustada e a interpretação da estimativa do coeficiente da regressão.

Temos que

$$n = 8; \quad \sum_{i=1}^{n} X_i = 182; \quad \sum_{i=1}^{n} X_i^2 = 4220; \quad \sum_{i=1}^{n} Y_i = 35,9; \quad \sum_{i=1}^{n} Y_i^2 = 166,49; \quad \sum_{i=1}^{n} X_iY_i = 796,6.$$

$$\begin{aligned} SPD_{XY} &= 796,6 - \frac{(182)\,(35,9)}{8} = -20,125 \\ SQD_X &= 4220 - \frac{(182)^2}{8} = 79,5 \\ SQD_Y &= 166,49 - \frac{(35,9)^2}{8} = 5,3888 \end{aligned}$$

$$\widehat{\beta}_1 = \frac{\text{SPD}_{XY}}{\text{SQD}_X} = \frac{-20,125}{79,5} = -0,2531$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{X} = \frac{35,9}{8} - (-0,2531) \times \frac{182}{8} = 4,4875 + 5,758 = 10,2455$$

$\boxed{\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i = 10,2455 - 0,2531 X_i}$.

Olhando-se direto na calculadora obteremos: $\boxed{\widehat{Y}_i = 10,2465 - 0,2531X}$.

b.(2 pts) A nota de preferência média estimada e o desvio da regressão para um consumidor com 20 anos;

Para $X_3 = 20 \Rightarrow \widehat{Y}_3 =?$

$$\widehat{Y}_3 = 10,2455 - 0,2531 \times 20 = 5,1835.$$

O desvio da regressão é $\widehat{\epsilon}_3 = Y_3 - \widehat{Y}_3 = 5,3 - 5,1835 = 0,1165$.

Olhando-se direto na calculadora obteremos: $\widehat{\epsilon}_3 = Y_3 - \widehat{Y}_3 = 5,3 - 5,1845 = 0,1155$.

c.(2 pts) Estime a nota de preferência média para um consumidor com idade de 10 anos. Comente sobre esse resultado.

Para $X_i = 10 \Rightarrow \widehat{Y}_i = ?$

$$\widehat{Y}_i = 10,2455 - 0,2531 \times 10 = 7,7145.$$

Olhando-se direto na calculadora obteremos:

$$\widehat{Y}_i = 10,2465 - 0,2531 \times 10 = 7,7155.$$

Essa estimativa não é confiável, pois trata-se de uma extrapolação.

d.(2 pts) Calcule e interprete o coeficiente de determinação do modelo de RLS ajustado.

Temos que

$$\begin{aligned} SQ_{\text{Regressão}} &= \frac{(SPD_{XY})^2}{SQD_X} = \frac{(-20,125)^2}{79,5} = 5,0945, \\ SQ_{\text{Total}} &= SQD_Y = 5,3888. \end{aligned}$$

$$\begin{aligned} r^2(\%) &= \frac{SQ_{\text{Regressão}}}{SQ_{\text{Total}}} \times 100\% = \frac{5,0945}{5,3888} \times 100\% \\ &= 94,54\%. \end{aligned}$$

O coeficiente de determinação $r^2(\%)$ foi de 94,54%, desta forma, o percentual da variabilidade observada do índice de concentração sérica de testosterona , explicado pela regressão linear simples, nos valores do peso corporal dos animais é 94,54%.

Olhando-se direto na calculadora obteremos:

$$\boxed{r^2(\%) = r_{XY}^2 \times 100\% = (-0,9723)^2 \times 100\% = 94,54\%}$$