UFV- CCE - DET

EST 105 - 1^a avaliação - 2^0 semestre de 2014 - 6/set/14

Nome: ______________________________ **Matrícula:** ____________________

Assinatura: ________________________. Favor apresentar documento com foto.

- São 5 questões e formulário em páginas numeradas de 1 a 8, total de 30 pontos, FAVOR CONFERIR ANTES DE INICIAR.
- ATENÇÃO: informe a seguir em qual turma está matriculado (sua nota será divulgada no sistema SAPIENS).

```
--------------------------------------------------------------
TURMA         HORÁRIO   SALA    PROFESSOR
--------------------------------------------------------------
T20: EST 085 T1 3=18:30-20:10 PVA126 - Monitor II - Gabi Nunes
T20: EST 085 T3 5=16:00-18:00 PVA361 - Monitor II
--------------------------------------------------------------
T1: 2=10-12 e 5=8-10    PVB310 - Ana Carolina
--------------------------------------------------------------
T2: 2=16-18 e 5=14-16  PVB310 - CHOS (coordenador)
--------------------------------------------------------------
T5: 3=16-18 e 6=14-16  PVB310 - Ana Carolina e Moysés
--------------------------------------------------------------
T6: 2=14-16 e 4=16-18  PVB107 - Fernando
--------------------------------------------------------------
T7: 4=8-10 e 6=10-12   PVB206 - Moysés
--------------------------------------------------------------
T8: 3=18:30-20:10 e 5=20:30-22:10 PVB306 - Paulo Emiliano
--------------------------------------------------------------
```

- 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.
- NOTA ZERO se mostrar a resposta correta e não apresentar os valores utilizados na fórmula.
- BOA SORTE e BOA PROVA !!!.

FORMULÁRIO

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

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

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

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

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

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

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

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

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

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

1.(5 pontos) Dado que,

$$\sum_{i=5}^{23} Y_i = 40, \quad \sum_{i=5}^{23} Y_i^2 = 65, \quad \sum_{k=1}^{10} Z_k = 30 \quad \text{e} \quad \sum_{k=1}^{10} Z_k^2 = 60,$$

utilize as propriedades de somatório e calcule,

$$\sum_{i=5}^{23}\sum_{k=1}^{10}\left[(Y_i-3)\,(Z_k+1)^2\right].$$

Solução: Temos que:

$$\begin{aligned} NT_i &= 23-5+1-0=19 \\ NT_i &= 10-1+1-0=10 \end{aligned}$$

$$\begin{aligned}
\sum_{i=5}^{23}\sum_{k=1}^{10}\left[(Y_i-3)\,(Z_k+1)^2\right] &= \sum_{i=5}^{23}\sum_{k=1}^{10}(Y_i-3)\left(Z_k^2+2Z_k+1\right) \\
&= \sum_{i=5}^{23}\sum_{k=1}^{10}\left(Y_iZ_k^2+2Y_iZ_k+Y_i-3Z_k^2-6Z_k-3\right) \\
&= \sum_{i=5}^{23}\sum_{k=1}^{10}Y_iZ_k^2+2\sum_{i=5}^{23}\sum_{k=1}^{10}Y_iZ_k+\sum_{i=5}^{23}\sum_{k=1}^{10}Y_i-3\sum_{i=5}^{23}\sum_{k=1}^{10}Z_k^2-6\sum_{i=5}^{23}\sum_{k=1}^{10}Z_k-\sum_{i=5}^{23}\sum_{k=1}^{10}3 \\
&= \sum_{i=5}^{23}Y_i\sum_{k=1}^{10}Z_k^2+2\sum_{i=5}^{23}Y_i\sum_{k=1}^{10}Z_k+10\sum_{i=5}^{23}Y_i-3\times 19\sum_{k=1}^{10}Z_k^2-6\times 19\sum_{k=1}^{10}Z_k-3\times 19\times 10 \\
&= 40\times 60+2\times 40\times 30+10\times 40-57\times 60-114\times 30-570 \\
&= 2400+2400+400-3420-3420-570 \\
&= 5200-7410 \\
&= -2210
\end{aligned}$$

2.(4 pontos) Dado que,

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

utilize as propriedades de somatório e calcule,

$$\sum_{k=1}^{20} (k-2)^3$$

Solução: modo 1

$$\begin{aligned}
\sum_{k=1}^{20} (k-2)^3 &= \sum_{k=1}^{20} (k-2)(k-2)^2 = \sum_{k=1}^{20} (k-2)\big(k^2-4k+4\big) \\
&= \sum_{k=1}^{20} \big(k^3-6k^2+12k-8\big) \\
&= \sum_{k=1}^{20} k^3 - 6\sum_{k=1}^{20} k^2 + 12\sum_{k=1}^{20} k - \sum_{k=1}^{20} 8 \\
&= \left(\frac{20\times 21}{2}\right)^2 - 6\times\left(\frac{20\times 21\times 41}{6}\right) + 12\times\left(\frac{20\times 21}{2}\right) - 8\times(20) \\
&= 44100 - 17220 + 2520 - 160 \\
&= 29240
\end{aligned}$$

Solução: modo 2

Fazendo-se $k-2=t$ temos:

- se $k=1$ então $t=-1$;
- se $k=20$ então $t=18$;

$$\begin{aligned}
\sum_{k=1}^{20} (k-2)^3 &= \sum_{t=-1}^{18} t^3 = (-1)^3 + 0^3 + \sum_{t=1}^{18} t^3 = -1 + \left[\frac{18\,(18+1)}{2}\right]^2 \\
&= -1 + (171)^2 = -1 + 29241 \\
&= 29240
\end{aligned}$$

3.(5 pontos) Nos itens a seguir assinale (V) se estiver inteiramente correto ou assinale (F) caso contrário e **indique e corrija onde estiver errado** (1 ponto cada item),

a.(**F**) Para duas amostras A e B, foram calculados os erros padrão das médias e obtidas respectivamente as seguintes estimativas $1,20$ e $2,5$. Desta forma, conclui-se que a média da amostra B é mais precisa ou é uma estimativa associada a uma maior precisão.

Para duas amostras A e B, foram calculados os erros padrão das médias e obtidas respectivamente as seguintes estimativas $S(\bar{X}_A) = 1,20$ e $S(\bar{X}_B) = 2,5$. Desta forma, conclui-se que a média da amostra A é mais precisa, ou seja, é uma estimativa associada a uma maior precisão, pois $S(\bar{X}_A) < S(\bar{X}_B)$.

b.(**F**) Estima-se que a temperatura média para Viçosa nos próximos dias seja igual a 23^0C com variância 9 $^0C^2$. Se a temperatura fosse medida em graus Fahrenheit (^0F), a temperatura média seria igual a 73,4 ^0F com variância 61,16 $^0F^2$. (Dica: Temperatura em ^0F = 32 + 1,8 × Temperatura ^0C).

Estima-se que a temperatura média para Viçosa nos próximos dias seja igual a 23^0C com variância 9 $^0C^2$. Se a temperatura fosse medida em graus Fahrenheit (^0F), a temperatura média seria igual a 73,4 ^0F com variância 29,16 $^0F^2$.

$$\bar{F} = \frac{\sum_{i=1}^{n} F_i}{n} = \frac{\sum_{i=1}^{n}(32 + C_i)}{n} = \frac{1}{n}\left(32n + 1,8\frac{\sum_{i=1}^{n} C_i}{n}\right) = 32 + 1,8 \times 23 = 73,4$$

$$\begin{aligned} S_F^2 &= \frac{\sum_{i=1}^{n}\left(F_i - \bar{F}\right)^2}{n-1} = \frac{\sum_{i=1}^{n}[32 + 1,8C_i - (32 + 41,4)]^2}{n-1} = \frac{1,8^2\sum_{i=1}^{n}\left(C_i - \bar{C}\right)^2}{n-1} \\ &= 1,8^2 S_{C_i}^2 = 29,16 \end{aligned}$$

c.(**F**) O coeficiente de variação é um valor adimensional que expressa o valor do desvio padrão da amostra como um percentual do valor da média aritmética da amostra. Quanto maior é o valor do coeficiente de variação, mais homogênea é a amostra.

O coeficiente de variação é um valor adimensional que expressa o valor do desvio padrão da amostra como um percentual do valor da média aritmética da amostra. Quanto maior é o valor do coeficiente de variação, menos homogênea é a amostra.

d.(**F**) A amplitude total da amostra $\{5,\ 12,\ 25,\ 40,\ 2,\ 51\}$ é igual a 39.

$$AT = X_{(n)} - X_{(1)} = 51 - 2 = 49$$

e.(**F**) O coeficiente de determinação, associado ao modelo de regressão linear simples (RLS) ajustado: $\hat{Y} = b_0 + b_1\ X$, expressa o percentual da variabilidade observada nos valores X que foi *explicada* pela RLS nos valores de Y.

O coeficiente de determinação, associado ao modelo de regressão linear simples (RLS) ajustado: $\hat{Y} = b_0 + b_1\ X$, expressa o percentual da variabilidade observada nos valores Y que foi *explicada* pela RLS nos valores de X

4.(10 pontos) A tabela seguinte mostra o valor da área construída em **dezenas de metros quadrados** ($X_i = m^2 \times 10$) e os respectivos valores do aluguel mensal em centenas de reais ($Y_i =$ R\$ $\times$ 100), para seis imóveis avaliados, $i = 1, \ldots, 6$.

i	1	2	3	4	5	6
X_i	6	7,5	8,5	9	10	12
Y_i	3,5	5	5,3	6	7,5	9,5

Pede-se:

a.(2 pts) Utilize sua calculadora para informar as seguintes somas:

$$\sum_{i=1}^{n} X_i = 53 \quad \sum_{i=1}^{n} X_i^2 = 489,5 \quad \sum_{i=1}^{n} Y_i = 36,8 \quad \sum_{i=1}^{n} Y_i^2 = 247,84 \quad \sum_{i=1}^{n} X_iY_i = 346,55$$

b.(2 pts) Apresente o modelo de regressão linear simples (RLS) que permite estimar o valor médio do aluguel mensal como uma função da área construída (aproxime a terceira casa após a vírgula, por ex: 0,9235=0,924 ou 4,1007=4,101).

Temos que

$$\begin{aligned} SQD_X &= 489,5 - \frac{(53)^2}{6} = 21,333 \\ SQD_Y &= 247,84 - \frac{(36,8)^2}{6} = 22,133 \\ SPD_{XY} &= 346,55 - \frac{53 \times 36,8}{6} = 21,483 \end{aligned}$$

Assim,

$$\widehat{\beta}_1 = \frac{SPD_{XY}}{SQD_X} = \frac{21,483}{21,333} = 1,007$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1\bar{Y} = 6,133 - 1,007 \times 8,833 = -2,762$$

Logo

$$\widehat{Y}_i = -2,762 + 1,007X_i$$

c.(2 pts) Com base no modelo de RLS ajustado: (**c1.**) estime o valor médio do aluguel mensal para um apartamento com 85m^2 de área construída e (**c2.**) apresente o desvio da regressão associado a esta estimativa.

c1) Para $X_3 = 8,5 \Rightarrow \widehat{Y}_3 =?$

$$\widehat{Y}_3 = -2,762 + 1,007 \times 8,5 = 5,798$$

c2) Temos que $\widehat{\varepsilon}_3 = Y_3 - \widehat{Y}_3 = 5,3 - 5,798 = -0,498$, isto é R\$49,8 são superestimados.

d.(2 pts) Para um aumento de **cinco metros ao quadrado** (5m^2) na área construída, qual é o aumento estimado no valor médio do aluguel? Justifique sua resposta. Temos que :

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i,$$

e para um aumento de 5m^2 temos

$$\widehat{Y}_i^* = \widehat{\beta}_0 + \widehat{\beta}_1 (X_i + 0,5) \qquad (X_i \text{ em } 10\text{m}^2)$$

$$\begin{aligned} \widehat{Y}_i^* - \widehat{Y}_i &= \widehat{\beta}_0 + \widehat{\beta}_1 (X_i + 0,5) - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_i\right) = 0,5\widehat{\beta}_1 \\ &= 0,5 \times 1,007 = 0,5035 \qquad \text{(centenas de reais)} \end{aligned}$$

ou seja, para um aumento de cinco metros ao quadrado (5m^2) na área construída, o aumento estimado no valor médio do aluguel é de R$50,35.

e.(2 pts) Apresente a estimativa do coeficiente de correlação linear simples.

$$r_{XY} = \frac{SPD_{XY}}{\sqrt{SQD_X SQD_Y}} = \frac{21,483}{\sqrt{21,333 \times 22,133}} = 0,9887$$

(Este valor não é percentual).

O coeficiente de determinação é

$$r^2(\%) = (r_{XY})^2 \times 100\% = 97,73\%$$

Se o aluno apresentou somente o coeficiente de determinação a nota será 0,5 pontos; Se apresentou o coeficiente de determinação e o coeficiente de correlação, porém não especificou qual é o coeficiente de determinação, a nota será 0,5 pontos.

5.(6 pontos) Segundo o IBGE (disponível em: http://www.ibge.gov.br) as estimativas de população, com data de referência em 1 de julho de 2014, o Brasil conta com 202,7 milhões de habitantes distribuídos pelos 5.570 municípios que compõem as 27 Unidades da Federação. A tabela a seguir mostra a população em 2014 (milhões de habitantes) dos 10 municípios mais populosos da Federação.

ORDEM	UF	MUNICÍPIO	POPULAÇÃO	ORDEM	UF	MUNICÍPIO	POPULAÇÃO
1^0	SP	São Paulo	11.89	6^0	MG	Belo Horizonte	2.49
2^0	RJ	Rio de Janeiro	6.45	7^0	AM	Manaus	2.02
3^0	BA	Salvador	2.90	8^0	PR	Curitiba	1.86
4^0	DF	Brasília	2.85	9^0	PE	Recife	1.61
5^0	CE	Fortaleza	2.57	10^0	RS	Porto Alegre	1.47

Considere os valores de população destas 10 cidades, pede-se:

a.(2 pts) a média e a mediana.

$$\overline{X} = \frac{\sum\limits_{i=1}^{10} X_i}{10} = \frac{11,89 + 6,45 + \cdots + 1,47}{10} = \frac{36,11}{10} = 3,611$$

Temos que $n = 10$ é par logo

$$Md_X = \frac{X_{(5)} + X_{(6)}}{2} = \frac{2,49 + 2,57}{2} = \frac{5,06}{2} = 2,53.$$

b.(2 pts) O coeficiente de variação.

$$S_X = \sqrt{\frac{1}{10-1}\left(224,6051 - \frac{(36,11)^2}{10}\right)} = 3,2354$$

$$CV_X\,(\%) = \frac{S_X}{\bar{X}} \times 100\% = \frac{3,2354}{3,611} \times 100\% = 89,6\%$$

c.(2 pts) Explique porquê a média não é uma boa medida descritiva de posição para esta amostra.

Sem os valores discrepantes $11,89$ e $6,45$ a média seria $\bar{X}^* = \dfrac{17,77}{8} = 2,2213.$

O valor médio 3,611 é muito maior que o valor mediano $2,53$ e, também é superior a 8 dos 10 elementos da amostra. Apenas dois valores discrepantes elevam a média de $2,2213$ para $3,611$.