By @kakashi_copiador

Estratégia
Concursos

Aula 07

Caixa Econômica Federal (CEF) (Técnico Bancário) Passo Estratégico de Probabilidade e Estatística - 2023 (Pré-Edital)

Autor:
Allan Maux Santana

26 de Janeiro de 2023

Índice

CORRELAÇÃO / REGRESSÃO LINEAR

Sumário

O que é mais cobrado dentro do assunto

Assunto	Grau de incidência
ANÁLISE DE REGRESSÃO LINEAR	*72,00%*
CORRELAÇÃO LINEAR	*28,00%*
TOTAL	*100,00%*

ROTEIRO DE REVISÃO E PONTOS DO ASSUNTO QUE MERECEM DESTAQUE

A ideia desta seção é apresentar um roteiro para que você realize uma revisão completa do assunto e, ao mesmo tempo, destacar aspectos do conteúdo que merecem atenção.

Para revisar e ficar bem preparado no assunto, você precisa, basicamente, seguir os passos a seguir:

Correlação

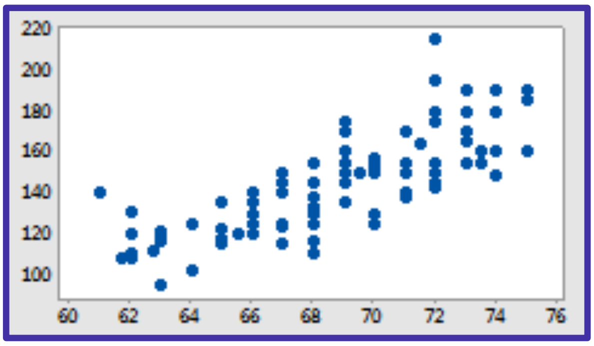
Sabemos que em alguns casos existe a necessidade de se estudar o comportamento entre duas variáveis, por exemplo: ***Altura x Massa***

Será que existe alguma correlação entre elas?

Será que com o aumento da Altura, haverá, necessariamente, o aumento da Massa, ou vice-versa?

Podemos entender um pouco melhor sobre isso, ao representarmos essas variáveis em um ***Gráfico de Dispersão***, vejam a seguir

Altura



Massa

A ***Correlação*** mensura o ***grau*** de ***relacionamento*** entre duas variáveis.

Já a ***Regressão*** determina uma ***equação*** (***função***) matemática que descreve o relacionamento entre essas duas variáveis.

Será que esses pontos podem ser correlacionados através de uma reta (linha)?

Vejam que eles estão bem próximos, mas, apesar de podermos até observar um pouco isso, não poderemos dar essa resposta apenas olhando o gráfico. Utilizaremos a seguinte fórmula:

Coeficiente de Correlação Linear de Pearson ***(r)***:

$$r = \frac{\sum[(X_i - \overline{X}) \cdot (Y_i - \overline{Y})]}{\sqrt{\sum(X_i - \overline{X})^2 \cdot \sum(Y_i - \overline{Y})^2}}$$

$$-1 \leq r \leq 1$$

Lembrem que existe outra forma de calcular o numerador e o denominador de "r":

$$\sum[(X_i - \overline{X}) \cdot (Y_i - \overline{Y})] = \sum(X_i \cdot Y_i) - n \cdot \overline{X} \cdot \overline{Y}$$

$$\sum(X_i - \overline{X})^2 = \sum X_i^2 - n \cdot (\overline{X})^2$$

$$\sum(Y_i - \overline{Y})^2 = \sum Y_i^2 - n \cdot (\overline{Y})^2$$

Correlação ***Positiva*** indica Grandezas que variam no ***mesmo sentido***. Reta Crescente.

Correlação ***Negativa*** indica Grandezas que variam em ***sentidos contrários***. Reta Decrescente.

A ***Correlação Linear*** será ***perfeita***, quando ***r = 1*** ou ***-1***. Isso significa que todos os pontos estão exatamente em cima da reta.

Caso o resultado seja ***nulo***, ***não*** há ***correlação linear***, mas poderá existir uma outra correlação, ok?

Propriedades do Coeficiente de Correlação:

P.1.: Se ***adicionarmos***, ou ***subtrairmos***, constantes às variáveis, o Coeficiente de Correlação ***não*** será ***alterado***.

P.2.: Se ***multiplicarmos***, ou ***dividirmos***, as variáveis por ***constantes***, poderá, ou não, ser alterado.

Constantes de mesmo sinal não alterará o Coeficiente de Correlação. No entanto, se as constantes tiverem sinais contrários, logo o Coeficiente de Correlação terá o sinal simétrico ao inicial.

Regressão Linear

O modelo ***Estatístico*** de uma ***Regressão Linear Simples*** entre X e Y é dado por:

$$\boldsymbol{Y_i = \alpha + \beta \cdot X_i + v_i}$$

$\boldsymbol{v_i}$: Variável Aleatória (Erro ou desvio)

$\boldsymbol{X_i}$: Variável Independente

$\boldsymbol{Y_i}$: Variável Dependente

Pressupostos da Variável Aleatória:
E (v_i) ***= 0 (média dos Erros)***
Var $(v_i) = \sigma^2$ ***(homocedasticia)***
Cov (v_i, v_j) ***= 0 para*** $i \neq j$ ***(erros independentes)***

Vamos agora ao ponto principal do nosso assunto. Lembrem do modelo:

$$Y_i = \alpha + \beta \cdot X_i + v_i$$

O método de ***Mínimos Quadrados*** vai determinar os estimadores de α e β.

Sejam "a" e "b" as estimativas de α e β. Logo, A reta da ***regressão estimada*** é dada por:

$$\hat{Y} = a + b\,X_i$$

O valor do ***Coeficiente Angular (b)*** é dado por:

$$b = \frac{\sum (X_i - \overline{X}) \cdot (Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2}$$

Sabemos, também, que há uma outra fórmula que diminuirá nossos cálculos para determinar "b", vejam:

$$b = \frac{\sum (X_i \cdot Y_i) - n \cdot (\overline{X} \cdot \overline{Y})}{\sum X_i^{\,2} - n \cdot \overline{X}^2}$$

Como a reta passa pelo ponto $(\overline{X}, \overline{Y})$, logo, para determinar o ***Coeficiente Linear "a",*** basta substituir "b", encontrado na fórmula acima, em:

$$\overline{Y} = a + b \cdot \overline{X}$$

Vamos às questões estratégicas para treinar o que vimos, agora.

QUESTÕES ESTRATÉGICAS

Nesta seção, apresentamos e comentamos uma amostra de questões objetivas selecionadas estrategicamente: são questões com nível de dificuldade semelhante ao que você deve esperar para a sua prova e que, em conjunto, abordam os principais pontos do assunto.

A ideia, aqui, não é que você fixe o conteúdo por meio de uma bateria extensa de questões, mas que você faça uma boa revisão global do assunto a partir de, relativamente, poucas questões.

1. *(CEBRASPE (CESPE) - Agente de Polícia Federal/2021)*

Um estudo objetivou avaliar a evolução do número mensal Y de milhares de ocorrências de certo tipo de crime em determinado ano. Com base no método dos mínimos quadrados ordinários, esse estudo apresentou um modelo de regressão linear simples da forma

$$\hat{Y} = 5 - 0,1\ x\ T$$

em que $\hat{Y}$ representa a reta ajustada em função da variável regressora T, tal que 1 ≤ T ≤ 12.

Os erros padrão das estimativas dos coeficientes desse modelo, as razões t e seus respectivos p-valores encontram-se na tabela a seguir.

	Erro Padrão	Razão *t*	*p*-valor
Intercepto	0,584	8,547	0,00
Coeficiente angular	0,064	1,563	0,15

Os desvios padrão amostrais das variáveis y e t foram, respectivamente, 1 e 3,6.

Com base nessas informações, julgue o item a seguir.

Se a média amostral da variável T for igual a 6,5, então a média amostral da variável Y será igual a 4,35 mil ocorrências.

C - Certo

E - Errado

Comentários:

Pessoal, a questão dar a seguinte reta de regressão:

$$\hat{Y} = 5 - 0{,}1 \text{ x } T$$

E pede a média amostra de Y, sabendo que a média amostral da T é 6,5. Aqui basta fazer a substituição do valor de T é já teremos a resposta.

$$\hat{Y} = 5 - 0{,}1 \text{ x } 6{,}5$$

$$\hat{Y} = 5 - 0{,}65$$

$$\hat{Y} = 4{,}35$$

Gabarito: CERTO

2. *(CEBRASPE (CESPE) - Agente de Polícia Federal/2021)*

Um estudo objetivou avaliar a evolução do número mensal Y de milhares de ocorrências de certo tipo de crime em determinado ano. Com base no método dos mínimos quadrados ordinários, esse estudo apresentou um modelo de regressão linear simples da forma

$$\boldsymbol{\hat{Y} = 5 - 0,1\,x\,T}$$

em que $\hat{Y}$ representa a reta ajustada em função da variável regressora T, tal que 1 ≤ T ≤ 12.

Os erros padrão das estimativas dos coeficientes desse modelo, as razões t e seus respectivos p-valores encontram-se na tabela a seguir.

	Erro Padrão	Razão *t*	*p*-valor
Intercepto	0,584	8,547	0,00
Coeficiente angular	0,064	1,563	0,15

Os desvios padrão amostrais das variáveis y e t foram, respectivamente, 1 e 3,6.

Com base nessas informações, julgue o item a seguir.

A correlação entre as variáveis Y e T foi igual a -0,1.

C - Certo

E - Errado

Comentários:

Pessoal, nessa questão é pedida a correlação entre as variáveis Y e T. Dando uma olhada preliminar nela parece uma questão bem trabalhosa de se fazer. Sendo que a banca deu as seguintes informações.

$S_y = 1$

$S_T = 3,6$

E sabemos que a correlação entre duas variáveis é dada por:

$$r_{YT} = \frac{S_{YT}}{S_Y S_T}$$

Onde,

$$S_{YT} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y}).(T_i - \bar{T})}{n-1} = Covariância$$

Veja que na questão anterior foi dada a média de T (6,5) e calculamos o média de Y (4,35). Agora temos que calcular o valor do numerador da covariância. Na questão foi dito que T varia de 1 a 12 (1 ≤ T ≤ 12). Desta forma, montando uma tabela teremos o seguinte:

	T	$Y = 5 - 0,1 \,.\, T$	$(T_i - \bar{T})$	$(Y_i - \bar{Y})$
	1	4,9	-5,5	
	2	4,8	-4,5	
	3	4,7	-3,5	
	4	4,6	-2,5	
	5	4,5	-1,5	
	6	4,4	-0,5	
	7	4,3	0,5	
	8	4,2	1,5	
	9	4,1	2,5	
	10	4,0	3,5	
	11	3,9	4,5	
	12	3,8	5,5	
Soma	78	52,2	0	

Só com o valor do somatório de $(T_i - \bar{T})$ já podemos parar de fazer os cálculos, pois o denominador da covariância dará zero e por consequência a correlação entre Y e T dará zero e não -0,1 como informado na questão

Só para não restar dúvidas, iremos calcular as médias.

$$\bar{T} = \frac{\sum_{i=1}^{n} T_i}{n} = \frac{78}{12} = 6,5$$

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{52,2}{12} = 4,35$$

Gabarito: ERRADO

3. *(IBFC - Supervisor de Pesquisas (IBGE)/Suporte Gerencial/2021)*

Num modelo de regressão linear pelo método dos mínimos quadrados, sabe-se que a inclinação da reta é a a = 3,24 e o intercepto da reta é b = 12,6, então o valor de $\widehat{Y}$ para x = 30 é:

a) 126,8.

b) 136,8.

c) 116,2.

d) 108,2.

e) 109,8.

Comentários:

Pessoal, nessa questão é pedido o valor de $\hat{Y}$ e nos dar as seguintes informações:

X = 30

a = 3,24

b = 12,6

Aqui teríamos que saber a ***equação da regressão linear*** pelo método dos mínimos quadrados.

$$\hat{Y} = b + a.X$$

$$\hat{Y} = 12{,}6 + 3{,}24.30 = 12{,}6 + 97{,}2 = 109{,}8$$

Gabarito: E

4. *(CEBRASPE (CESPE) - Analista Judiciário (TJ PA)/Estatística/2020)*

Texto 7A3-I

O coeficiente de correlação linear de Pearson entre duas variáveis aleatórias discretas X e Y definidas sobre um mesmo espaço amostral é dado por

$$CORR(X,Y) = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}\,\sqrt{n(\sum_{i=1}^{n} y_i^2) - (\sum_{i=1}^{n} y_i)^2}}.$$

Já na reta de melhor ajuste Y = aX + b, determinada pelo método dos mínimos quadrados, os coeficientes são dados por

$$a = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

$$b = \frac{\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i}{n}.$$

Uma forma de avaliar a precisão do modelo consiste em comparar o estimador não viesado da variância residual, obtido das diferenças entre os valores observados e os previstos pelo modelo, $\hat{S}_e = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ *, com o estimador não viesado da variância dos valores observados,* $S_e = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$ *.*

A tabela a seguir apresenta as penas de reclusão (P), em anos, cominadas a um grupo de dez réus, e suas respectivas rendas familiares mensais per capitas (R), em número de salários mínimos, em que a última coluna foi obtida usando a reta ajustada pelo método dos mínimos quadrados.

réu	P	R	$P \times R$	P^2	R^2	$(R-\overline{R})^2$	$(R-\hat{R})^2$
1	14	0,25	3,5	196	0,0625	3,0625	0,054756
2	12	0,5	6	144	0,25	2,25	0,000144
3	10,9	1	10,9	118,81	1	1	0,046311
4	6	1,5	9	36	2,25	0,25	0,25
5	5	1,75	8,75	25	3,0625	0,0625	0,248004
6	3	2	6	9	4	0	0,553536
7	3	2,5	7,5	9	6,25	0,25	0,059536
8	2,3	3	6,9	5,29	9	1	0,0067898
9	1,8	3,5	6,3	3,24	12,25	2,25	0,2101306
10	2	4	8	4	16	4	1,016064
totais	60	20	72,85	550,34	54,125	14,125	2,4452714

Dados:

$1903{,}4^{1/2} = 43{,}63$

$141{,}25^{1/2} = 11{,}88$

Com base no texto 7A3-I, a renda familiar per capita esperada X, em número de salários mínimos, obtida aplicando-se a reta de melhor ajuste aos dados determinada pelo método dos mínimos quadrados para um réu ao qual tenha sido cominada uma pena de 4 anos de reclusão é

a) 2,3 < X < 2,6.

b) 2,1 < X < 2,3.

c) 1,9 < X < 2,1.

d) 1,2 < X < 1,9.

e) 1,0 < X < 1,2.

Comentários:

Pessoal, a primeira coisa a ser feita é calcular os valores de **a** e **b**. Para isso, utilizaremos as seguintes informações:

$n = 10$

$\sum x_i = 60$

$\sum y_i = 20$

$\sum x_i y_i = 72{,}85$

$\sum x_i^2 = 550{,}34$

Agora basta aplicar as fórmulas dadas na questão para calcular os coeficientes:

$$a = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2} = \frac{10 \cdot 72{,}85 - 60 \cdot\ 20}{10 \cdot\ 550{,}34 - 60^2} = \frac{728{,}5 - 1200}{5503{,}4 - 3600} = \frac{-471{,}5}{1903{,}4} = -0{,}247$$

$$b = \frac{\sum y_i - a \sum x_i}{n} = \frac{20 - (-0{,}247).\ 60}{10} = \frac{20 + 14{,}82}{10} = \frac{34{,}82}{10} = 3{,}482$$

Agora aplicamos a equação da reta linear.

$$Y = a \cdot X + b$$

$$Y = -0{,}24 \cdot X + 3{,}44$$

Portanto, a renda familiar per capita esperada para um réu ao qual tenha sido cominada uma pena de 4 anos de reclusão será a seguinte:

$$Y = -0{,}247\ \cdot\ 4 + 3{,}482 = -0{,}988 + 3{,}482 = \mathbf{2{,}494}$$

Gabarito: A

5. ***(CEBRASPE (CESPE) - Analista Judiciário (TJ AM)/Estatística/2019)***

Um estudo considerou um modelo de regressão linear simples na forma y = 0,8x + b + ϵ, em que y é a variável dependente, x representa a variável explicativa do modelo, o coeficiente b denomina-se intercepto e ϵ é um erro aleatório que possui média nula e desvio padrão σ. Sabe-se que a variável y segue a distribuição normal padrão e que o modelo apresenta coeficiente de determinação R^2 igual a 85%.

Com base nessas informações, julgue o item que se segue.

A correlação linear entre as variáveis x e y é superior a 0,9.

C - Certo

E - Errado

Comentários:

Pessoal, nessa questão é dado o coeficiente de determinação e é pedido a correlação linear entre X e Y.

$$R^2 = coeficiente\ de\ determinação = 0{,}85$$

Já a correlação linear é dada pela raiz quadrada do coeficiente de determinação.

$$R = \sqrt{coeficeinte\ de\ determinação} = \sqrt{0{,}85} = 0{,}92$$

Gabarito: CERTO

Q.06 (FGV/ Analista Legislativo (ALERO)/Estatística/2018)

Se b0 e b1 são as estimativas por mínimos quadrados de β0 e β1, respectivamente, então seus valores são dados por

a) $\boldsymbol{b_1 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 / \sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y})\,;\ b_0 = \bar{Y} - b_1\bar{X}}$

b) $\boldsymbol{b_1 = \sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y}) / \sum_{i=1}^{n}(Y_i - \bar{Y})^2\,;\ b_0 = \bar{Y} + b_1\bar{X}}$

c) $\boldsymbol{b_1 = \sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y}) / \sum_{i=1}^{n}(Y_i - \bar{Y})^2\,;\ b_0 = \bar{Y} - b_1\bar{X}}$

d) $\boldsymbol{b_1 = \sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y}) / \sum_{i=1}^{n}(X_i - \bar{X})^2\,;\ b_0 = \bar{Y} - b_1\bar{X}}$

e) $\boldsymbol{b_1 = \sum_{i=1}^{n}(X_i - \bar{X})^2 / \sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y})\,;\ b_0 = \bar{Y} + b_1\bar{X}}$

Comentários:

Pessoal, nessa questão são pedidas as fórmulas dos estimadores de mínimos quadrados. Portanto,

$$b_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$b_0 = \bar{Y} - b_1\bar{X}$$

Resposta letra "D"

Gabarito: D

Q.07 (FGV/AFRE (SEFAZ RJ)/2011)

A tabela abaixo mostra os valores de duas variáveis, X e Y.

X	Y
4	4,5
4	5
3	5
2	5,5

Sabe-se que:

$$\sum X = 13$$

$$\sum Y = 20$$

$$\sum XY = 64$$

$$\sum X^2 = 45$$

$$\left(\sum X\right)^2 = 169$$

O valor de "b" na regressão simples $Y = a + bX$ é

a) 11 /5.

b) -3 /8.

c) -4 /11.

d) -4 /17.

e) -11/65.

Comentários:

Nessa questão temos que encontrar o coeficiente "b". Pela tabela dada na questão sabemos que temos um "n" igual a 4. Com isso e o somatório de X e Y, podemos encontrar as médias.

$$\bar{X} = \frac{\sum X}{n} = \frac{13}{4} = 3{,}25$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{20}{4} = 4$$

Agora, basta calcular "b" através da seguinte expressão:

$$b = \frac{\sum XY - n\,.\,\bar{X}.\,\bar{Y}}{\sum X^2 - n\,.\,(\bar{X})^2}$$

Substituindo os valores temos o seguinte:

$$b = \frac{64 - 4\,.3{,}25\,.\;5}{45 - 4\,.\;(3{,}25)^2} = \frac{64 - 65}{45 - 4\,.10{,}5645} = \frac{-1}{45 - 42{,}25} = \frac{-1}{2{,}75}$$

Colocando de acordo com as alternativas.

$$b = -\frac{100}{275} = -\frac{\mathbf{4}}{\mathbf{11}}$$

Portanto, resposta letra "C"

Gabarito: C

Q.08 (CESGRANRIO / ESCRITURÁRIO/BB /2018)

Uma instituição financeira pretende lançar no mercado um aplicativo para celular. Para isso, deseja relacionar o grau de conhecimento dos clientes com as variáveis: nível de escolaridade e idade.

Uma amostra aleatória de 46 clientes foi selecionada e, posteriormente, aplicou-se o modelo de regressão linear, sendo a variável dependente o grau de conhecimento, em uma escala crescente, e as variáveis independentes (i) o nível de escolaridade, em anos de estudo com aprovação, e (ii) a idade, em anos completos.

Os resultados obtidos para os coeficientes foram:

	Coeficientes	Erro padrão	Estatística t	valor-P
Intersecção	50,7	4,1	12,4	8,5E-16
Nível de escolaridade (anos de estudo com aprovação)	4,0	0,3	12,4	9,1E-16
Idade (anos completos)	-0,6	0,1	-8,4	1,2E-10

O grau de conhecimento esperado de um cliente com 10 anos de estudos com aprovação e com 30 anos de idade completos é

a) 108,7
b) 94,1
c) 54,1
d) 72,7
e) 86,1

Comentários:

Temos um Modelo de Regressão Linear, logo:

VARIÁVEL DEPENDENTE: grau de conhecimento (C), em escala crescente,

VARIÁVEIS INDEPENDENTES: o nível de escolaridade (E), em anos de estudo com aprovação, e a idade (I), em anos completos, portanto:

$$C = 50{,}7 + 4 \cdot E - 0{,}6 \cdot I$$

Logo, como E= 10 anos de estudos com aprovação e com I = 30 anos de idade completos, temos que:

$$C = 50{,}7 + 4 \cdot 10 - 0{,}6 \cdot 30$$

$$C = 72{,}7$$

Gabarito: D

LISTA DE QUESTÕES ESTRATÉGICAS

1. ***(CEBRASPE (CESPE) - Agente de Polícia Federal/2021)***

Um estudo objetivou avaliar a evolução do número mensal Y de milhares de ocorrências de certo tipo de crime em determinado ano. Com base no método dos mínimos quadrados ordinários, esse estudo apresentou um modelo de regressão linear simples da forma

$$\hat{Y} = 5 - 0{,}1\ x\ T$$

em que $\hat{Y}$ representa a reta ajustada em função da variável regressora T, tal que 1 ≤ T ≤ 12.

Os erros padrão das estimativas dos coeficientes desse modelo, as razões t e seus respectivos p-valores encontram-se na tabela a seguir.

	Erro Padrão	Razão *t*	*p*-valor
Intercepto	0,584	8,547	0,00

Coeficiente angular	0,064	1,563	0,15

Os desvios padrão amostrais das variáveis y e t foram, respectivamente, 1 e 3,6.

Com base nessas informações, julgue o item a seguir.

Se a média amostral da variável T for igual a 6,5, então a média amostral da variável Y será igual a 4,35 mil ocorrências.

C - Certo

E - Errado

2. *(CEBRASPE (CESPE) - Agente de Polícia Federal/2021)*

Um estudo objetivou avaliar a evolução do número mensal Y de milhares de ocorrências de certo tipo de crime em determinado ano. Com base no método dos mínimos quadrados ordinários, esse estudo apresentou um modelo de regressão linear simples da forma

$$\widehat{Y} = 5 - 0,1\, x\, T$$

em que $\widehat{Y}$ representa a reta ajustada em função da variável regressora T, tal que 1 ≤ T ≤ 12.

Os erros padrão das estimativas dos coeficientes desse modelo, as razões t e seus respectivos p-valores encontram-se na tabela a seguir.

	Erro Padrão	Razão *t*	*p*-valor
Intercepto	0,584	8,547	0,00
Coeficiente angular	0,064	1,563	0,15

Os desvios padrão amostrais das variáveis y e t foram, respectivamente, 1 e 3,6.

Com base nessas informações, julgue o item a seguir.

A correlação entre as variáveis Y e T foi igual a -0,1.

C - Certo

E - Errado

3. *(IBFC - Supervisor de Pesquisas (IBGE)/Suporte Gerencial/2021)*

Num modelo de regressão linear pelo método dos mínimos quadrados, sabe-se que a inclinação da reta é a a = 3,24 e o intercepto da reta é b = 12,6, então o valor de $\widehat{Y}$ para x = 30 é:

a) 126,8.

b) 136,8.

c) 116,2.

d) 108,2.

e) 109,8.

4. *(CEBRASPE (CESPE) - Analista Judiciário (TJ PA)/Estatística/2020)*

Texto 7A3-I

O coeficiente de correlação linear de Pearson entre duas variáveis aleatórias discretas X e Y definidas sobre um mesmo espaço amostral é dado por

$$CORR(X,Y) = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}\,\sqrt{n(\sum_{i=1}^{n} y_i^2) - (\sum_{i=1}^{n} y_i)^2}}.$$

Já na reta de melhor ajuste Y = aX + b, determinada pelo método dos mínimos quadrados, os coeficientes são dados por

$$a = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

$$b = \frac{\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i}{n}.$$

Uma forma de avaliar a precisão do modelo consiste em comparar o estimador não viesado da variância residual, obtido das diferenças entre os valores observados e os previstos pelo modelo, $\hat{S}_e = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ *, com o estimador não viesado da variância dos valores observados,* $S_e = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ *.*

A tabela a seguir apresenta as penas de reclusão (P), em anos, cominadas a um grupo de dez réus, e suas respectivas rendas familiares mensais per capitas (R), em número de salários mínimos, em que a última coluna foi obtida usando a reta ajustada pelo método dos mínimos quadrados.

réu	P	R	$P \times R$	P^2	R^2	$(R-\bar{R})^2$	$(R-\hat{R})^2$
1	14	0,25	3,5	196	0,0625	3,0625	0,054756
2	12	0,5	6	144	0,25	2,25	0,000144
3	10,9	1	10,9	118,81	1	1	0,046311
4	6	1,5	9	36	2,25	0,25	0,25
5	5	1,75	8,75	25	3,0625	0,0625	0,248004
6	3	2	6	9	4	0	0,553536
7	3	2,5	7,5	9	6,25	0,25	0,059536
8	2,3	3	6,9	5,29	9	1	0,0067898
9	1,8	3,5	6,3	3,24	12,25	2,25	0,2101306
10	2	4	8	4	16	4	1,016064
totais	60	20	72,85	550,34	54,125	14,125	2,4452714

Dados:

$1903{,}4^{1/2} = 43{,}63$

$141{,}25^{1/2} = 11{,}88$

Com base no texto 7A3-I, a renda familiar per capita esperada X, em número de salários mínimos, obtida aplicando-se a reta de melhor ajuste aos dados determinada pelo método dos mínimos quadrados para um réu ao qual tenha sido cominada uma pena de 4 anos de reclusão é

a) $2{,}3 < X < 2{,}6$.

b) $2{,}1 < X < 2{,}3$.

c) $1{,}9 < X < 2{,}1$.

d) $1{,}2 < X < 1{,}9$.

e) $1{,}0 < X < 1{,}2$.

5. *(CEBRASPE (CESPE) - Analista Judiciário (TJ AM)/Estatística/2019)*

Um estudo considerou um modelo de regressão linear simples na forma $y = 0{,}8x + b + \epsilon$, em que y é a variável dependente, x representa a variável explicativa do modelo, o coeficiente b denomina-se intercepto e ϵ é um erro aleatório que possui média nula e desvio padrão σ. Sabe-se que a variável y segue a distribuição normal padrão e que o modelo apresenta coeficiente de determinação R^2 igual a 85%.

Com base nessas informações, julgue o item que se segue.

A correlação linear entre as variáveis x e y é superior a 0,9.

C - Certo

E – Errado

Q.06 (FGV/ Analista Legislativo (ALERO)/Estatística/2018)

Se b_0 e b_1 são as estimativas por mínimos quadrados de $\beta 0$ e $\beta 1$, respectivamente, então seus valores são dados por

a) $b_1 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 / \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})\ ;\ b_0 = \bar{Y} - b_1\bar{X}$

b) $b_1 = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) / \sum_{i=1}^{n}(Y_i - \bar{Y})^2\ ;\ b_0 = \bar{Y} + b_1\bar{X}$

c) $b_1 = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) / \sum_{i=1}^{n}(Y_i - \bar{Y})^2\ ;\ b_0 = \bar{Y} - b_1\bar{X}$

d) $b_1 = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) / \sum_{i=1}^{n}(X_i - \bar{X})^2\ ;\ b_0 = \bar{Y} - b_1\bar{X}$

e) $b_1 = \sum_{i=1}^{n}(X_i - \bar{X})^2 / \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$; $b_0 = \bar{Y} + b_1\bar{X}$

Q.07 (FGV/AFRE (SEFAZ RJ)/2011)

A tabela abaixo mostra os valores de duas variáveis, X e Y.

X	Y
4	4,5
4	5
3	5
2	5,5

Sabe-se que:

$$\sum X = 13$$

$$\sum Y = 20$$

$$\sum XY = 64$$

$$\sum X^2 = 45$$

$$\left(\sum X\right)^2 = 169$$

O valor de "b" na regressão simples $Y = a + bX$ é

a) 11 /5.

b) -3 /8.

c) -4 /11.

d) -4 /17.

e) -11/65.

Q.08 (CESGRANRIO / ESCRITURÁRIO/BB /2018)

Uma instituição financeira pretende lançar no mercado um aplicativo para celular. Para isso, deseja relacionar o grau de conhecimento dos clientes com as variáveis: nível de escolaridade e idade.

Uma amostra aleatória de 46 clientes foi selecionada e, posteriormente, aplicou-se o modelo de regressão linear, sendo a variável dependente o grau de conhecimento, em uma escala crescente, e as variáveis independentes (i) o nível de escolaridade, em anos de estudo com aprovação, e (ii) a idade, em anos completos.

Os resultados obtidos para os coeficientes foram:

	Coeficientes	Erro padrão	Estatística t	valor-P
Intersecção	50,7	4,1	12,4	8,5E-16
Nível de escolaridade (anos de estudo com aprovação)	4,0	0,3	12,4	9,1E-16
Idade (anos completos)	-0,6	0,1	-8,4	1,2E-10

O grau de conhecimento esperado de um cliente com 10 anos de estudos com aprovação e com 30 anos de idade completos é

a) 108,7
b) 94,1
c) 54,1
d) 72,7
e) 86,1

Gabarito

1	2	3	4	5	6	7	8	9	10
C	E	E	A	C	D	C	D	*	*

ESSA LEI TODO MUNDO CONHECE:
PIRATARIA É CRIME.

Mas é sempre bom revisar o porquê e como você pode ser prejudicado com essa prática.

1. Professor investe seu tempo para elaborar os cursos e o site os coloca à venda.

2. Pirata divulga ilicitamente (grupos de rateio), utilizando-se do anonimato, nomes falsos ou laranjas (geralmente o pirata se anuncia como formador de "grupos solidários" de rateio que não visam lucro).

3. Pirata cria alunos fake praticando falsidade ideológica, comprando cursos do site em nome de pessoas aleatórias (usando nome, CPF, endereço e telefone de terceiros sem autorização).

4. Pirata compra, muitas vezes, clonando cartões de crédito (por vezes o sistema anti-fraude não consegue identificar o golpe a tempo).

5. Pirata fere os Termos de Uso, adultera as aulas e retira a identificação dos arquivos PDF (justamente porque a atividade é ilegal e ele não quer que seus fakes sejam identificados).

6. Pirata revende as aulas protegidas por direitos autorais, praticando concorrência desleal e em flagrante desrespeito à Lei de Direitos Autorais (Lei 9.610/98).

7. Concurseiro(a) desinformado participa de rateio, achando que nada disso está acontecendo e esperando se tornar servidor público para exigir o cumprimento das leis.

8. O professor que elaborou o curso não ganha nada, o site não recebe nada, e a pessoa que praticou todos os ilícitos anteriores (pirata) fica com o lucro.

Deixando de lado esse mar de sujeira, aproveitamos para agradecer a todos que adquirem os cursos honestamente e permitem que o site continue existindo.