etrônico

Aul

Passo Estratégico de Mat. Financeira p/ CFC 2019.1 (Bacharel em Ciências Contábeis) - Pós-edital

Professor: Equipe Rafael Barbosa, Rafael Barbosa

1 - Introdução

Fala, pessoal! Tudo tranquilo com vocês?

Vamos ver mais um relatório do Passo Estratégico de Estatística para **Auditor Fiscal do ISS Campinas.**

Falaremos hoje sobre os seguintes temas: *Correlação e Regressão.*

Esse é um assunto muito cobrado em provas, portanto, atenção total a esse relatório.

Boa revisão!

2 - Análise Estatística

2.1 - Análise Estatística: Últimos 5 anos – VUNESP

Considerando as provas objetivas dos últimos 5 anos da VUNESP:

Tabela 01

ASSUNTO	Qtde de concursos que previram a disciplina Estatística	Qtde de concursos que previram o assunto no edital	% de incidência do assunto no edital da disciplina
Correlação e Regressão	66	30	45,45%

Tabela 02

ASSUNTO	Qtde de concursos que previram o assunto no edital	Qtde de concursos que efetivamente cobraram o assunto em prova	% de incidência do assunto nas provas da banca
Correlação e Regressão	30	3	10%

Tabela 03

ASSUNTO	Total de questões das provas de Estatística	Total de questões em que o assunto foi abordado	% de incidência do assunto no total de questões da disciplina
Correlação e Regressão	178	7	3,93%

Assunto: Correlação e Regressão
Tabela 1: de todos os editais da VUNESP(amostra) que trouxeram a Estatística, em **45,45%** dos casos havia a cobrança do assunto.
Tabela 2: quando o edital pedia o assunto no conteúdo programático da disciplina, o mesmo foi cobrado nas respectivas provas em **10%** dos casos.
Tabela 3: de todas as questões de Estatística da VUNESP (amostra) nos últimos 5 anos, o assunto foi cobrado em **3,93%** do total de questões.

2.2 - Conclusão da Análise Estatística

Veremos neste relatório dois assuntos com uma boa incidência em provas. Pois, como vocês podem perceber, foram explorados em 3,93% da amostra. Assim, você deve estudar esses temas na sua revisão, pois é provável que caia algo na sua prova constante desse relatório. Dê atenção especial a esta aula, ela pode fazer a diferença na sua aprovação.

Procure resolver muitas questões sobre regressão, pois somente com a prática você conseguirá resolver uma questão desse assunto com rapidez na sua prova.

Bons estudos.

3 - Análise das Questões

Correlação e Regressão.

1. **VUNESP - Especialista em Regulação e Fiscalização de Serviços Públicos I (ARSESP)/Econômico Financeiro/2018**

Usando dados amostrais para estudar a correlação entre preço x da gasolina (em reais) e movimento y de vendas semanais (em litros) em postos de combustíveis de certa região, um grupo de pesquisadores verificou existir correlação linear entre as duas variáveis. A reta de regressão y = bx + a estabelecida no estudo tem coeficiente angular – 4,50 e coeficiente linear 15 500 (valores aproximados). Suponha que o preço R$ 4,00 por litro pertença ao intervalo de preços verificados na pesquisa. Usando a reta de regressão para uma estimativa do movimento de vendas, e considerando uma unidade de venda (ou posto) com preço da gasolina de R$ 4,00 por litro, então o movimento semanal de vendas (em litros) estimado nesse posto será de

a) 15 594,50.

b) 15 518.

c) 15 482.

d) 15 598,30.

e) 15 000.

Comentários:

A questão cobra conceitos sobre regressão linear simples.

Na regressão linear simples, quando observamos de forma gráfica duas variáveis cujo valor de correlação linear seja alto (igual a 1), a relação entre ambas se dará por uma reta (linear). Dessa forma, ela pode ser definida pela equação da reta, em que y=αx+β.

Assim, tendo os parâmetros da reta, para um dado valor de x, poderemos obter y.

Sabemos que a reta de regressão calculada passa pelos pontos médios das variáveis.

Em uma função do tipo y = ax + b, damos os seguintes nomes aos números reais a e b:

a = Coeficiente Angular: Determina a inclinação da reta (gráfico) gerada pela equação y = ax + b

b = Coeficiente Linear: Determina o ponto de encontro da reta com o eixo das ordenadas (eixo y).

De acordo com o enunciado, temos que: a = -4,5 e b = 15500, então nossa função será:

y=−4,5+15500.

Também temos a informação que o preço R$ 4,00 por litro pertença ao intervalo de preços verificados na pesquisa, ou seja, x = 4 pertence a essa reta. Agora, basta substituirmos x po 4 na função acima.

$$y = -4{,}5 \cdot 4 + 15500$$

$$y = -18 + 15500$$

$$y = 15482\ litros\ de\ gasolina.$$

Portanto, usando a reta de regressão para uma estimativa do movimento de vendas, e considerando uma unidade de venda (ou posto) com preço da gasolina de R$ 4,00 por litro, então o movimento semanal de vendas (em litros) estimado nesse posto será de 15.482 litros de gasolina.

Gabarito: C

---

2. VUNESP - Estatístico (TJ SP)/Judiciário/2015

Leia o texto para responder à questão.

Para se fazer um estudo sobre a relação da precipitação pluviométrica (x) em mm e a produção de frutas (y) em toneladas em certa região, coletaram-se dados durante 8 anos. Na tabela seguinte, registraram-se os valores obtidos de x e de y, respectivamente, precipitação média mensal e produção média mensal em cada ano. Na tabela, estão também os valores resultantes de alguns processamentos dos dados, incluídos valores obtidos com a reta de regressão y = a + bx.

Ano	Precipitação pluviométrica (x mm)	Produção de frutas (y ton.)	Projeção (ton) (y = a + bx)	Variação Explicada	Variação não explicada	Variação Total
1	101,00	25,00	25,12	4,02	0,01	4,51
2	103,00	26,00	25,61	2,29	0,15	1,26
3	106,00	27,00	26,35	0,60	0,42	0,02
4	109,00	27,00	27,09	0,00	0,01	0,02
5	110,00	26,00	27,34	0,04	1,78	1,26
6	112,00	28,00	27,83	0,49	0,03	0,77
7	115,00	28,00	28,57	2,08	0,32	0,77
8	117,00	30,00	29,06	3,74	0,89	8,26
Totais		**217,00**		**13,26**	**3,61**	**16,87**

Ao se montar a tabela ANOVA para esse caso, obtiveram-se os seguintes valores:

Fonte de Variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	1	13,26		
Resíduo	6	3,61		
TOTAL	7	16,87		

No teste de hipótese ao nível de 5% de significância para verificar a linearidade (H_0: $\beta = 0$ contra H_1: $\beta \neq 0$), o valor crítico de F para se rejeitar H_0 é, aproximadamente,

a) 19,4.

b) 8,85.

c) 5,99.

d) 3,47.

e) 3,01.

Comentários:

A questão cobra conceitos sobre regressão.

Sabemos que a distribuição F possui um grau de liberdade associado ao numerador e outro ao denominador.

No numerador, temos o GL referente ao modelo de regressão (1 grau de liberdade).

Já, no denominador, temos o GL referente aos resíduos (6 graus de liberdade).

Consultado a tabela da distribuição F, para 1 grau de liberdade no numerador e 6 no denominador, temos F=5,99. Vejamos:

Tabela da distribuição F.

Corpo da tabela da valores de fc tais que P(F>fc)=0,05

Valores aproximados com 2 casas após a vírgula

		Graus de liberdade do numerador									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
GL denominador	**1**	161,45	199,5	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88
	2	18,51	19	19,16	19,25	19,3	19,33	19,35	19,37	19,38	19,4
	3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6	5,96
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
	6	**5,99**	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,1	4,06
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,5	3,44	3,39	3,35
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
	10	4,96	4,1	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
	11	4,84	3,98	3,59	3,36	3,2	3,09	3,01	2,95	2,9	2,85
	12	4,75	3,89	3,49	3,26	3,11	3	2,91	2,85	2,8	2,75
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
	14	4,6	3,74	3,34	3,11	2,96	2,85	2,76	2,7	2,65	2,6
	15	4,54	3,68	3,29	3,06	2,9	2,79	2,71	2,64	2,59	2,54
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
	17	4,45	3,59	3,2	2,96	2,81	2,7	2,61	2,55	2,49	2,45
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
	19	4,38	3,52	3,13	2,9	2,74	2,63	2,54	2,48	2,42	2,38
	20	4,35	3,49	3,1	2,87	2,71	2,6	2,51	2,45	2,39	2,35
	21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
	22	4,3	3,44	3,05	2,82	2,66	2,55	2,46	2,4	2,34	2,3
	23	4,28	3,42	3,03	2,8	2,64	2,53	2,44	2,37	2,32	2,27

24	4,26	3,4	3,01	2,78	2,62	2,51	2,42	2,36	2,3	2,25
25	4,24	3,39	2,99	2,76	2,6	2,49	2,4	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,2
28	4,2	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,7	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
60	4	3,15	2,76	2,53	2,37	2,25	2,17	2,1	2,04	1,99
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91

Portanto, no teste de hipótese ao nível de 5% de significância, para verificar a linearidade (H0: β = 0 contra H1: β ≠ 0), o valor crítico de F para se rejeitar H0 é, aproximadamente, 5,99.

Gabarito: C

---

3. VUNESP - Estatístico (TJ SP)/Judiciário/2015

Leia o texto para responder à questão.

Para se fazer um estudo sobre a relação da precipitação pluviométrica (x) em mm e a produção de frutas (y) em toneladas em certa região, coletaram-se dados durante 8 anos. Na tabela seguinte, registraram-se os valores obtidos de x e de y, respectivamente, precipitação média mensal e produção média mensal em cada ano. Na tabela, estão também os valores resultantes de alguns processamentos dos dados, incluídos valores obtidos com a reta de regressão y = a + bx.

Ano	Precipitação pluviométrica (x mm)	Produção de frutas (y ton.)	Projeção (ton) (y = a + bx)	Variação Explicada	Variação não explicada	Variação Total
1	101,00	25,00	25,12	4,02	0,01	4,51
2	103,00	26,00	25,61	2,29	0,15	1,26
3	106,00	27,00	26,35	0,60	0,42	0,02
4	109,00	27,00	27,09	0,00	0,01	0,02
5	110,00	26,00	27,34	0,04	1,78	1,26
6	112,00	28,00	27,83	0,49	0,03	0,77
7	115,00	28,00	28,57	2,08	0,32	0,77
8	117,00	30,00	29,06	3,74	0,89	8,26
Totais		**217,00**		**13,26**	**3,61**	**16,87**

Ao se montar a tabela ANOVA para esse caso, obtiveram-se os seguintes valores:

Fonte de Variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	1	13,26		
Resíduo	6	3,61		
TOTAL	7	16,87		

Ao complementar a tabela e calcular o valor de F, encontra-se o valor aproximado de

a) 13.

b) 22.

c) 30.

d) 38.

e) 42.

Comentários:

Mais uma questão cobrando conceitos sobre regressão.

Sabemos que para calcular cada quadrado médio, devemos dividir a soma de quadrados pelo número de graus de liberdade:

Fonte de Variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	1	13,26	13,26÷1=13,26	
Resíduo	6	3,61	3,61÷6	
TOTAL	7	16,87		

A estatística F é dada pela divisão entre os quadrados médios:

$$F = \frac{13{,}26}{\frac{3{,}61}{6}}$$

$$F = \frac{13{,}26 \text{ x } 6}{3{,}61}$$

$$F = \frac{79{,}56}{3{,}61} = 22{,}04$$

Portanto, ao complementar a tabela e calcular o valor de F, encontra-se o valor aproximado de 22.

Gabarito: B

4. VUNESP - Estatístico (TJ SP)/Judiciário/2015

Leia o texto para responder à questão.

Para se fazer um estudo sobre a relação da precipitação pluviométrica (x) em mm e a produção de frutas (y) em toneladas em certa região, coletaram-se dados durante 8 anos. Na tabela seguinte, registraram-se os valores obtidos de x e de y, respectivamente, precipitação média mensal e produção média mensal em cada ano. Na tabela, estão também os valores resultantes de alguns processamentos dos dados, incluídos valores obtidos com a reta de regressão y = a + bx.

Ano	Precipitação pluviométrica (x mm)	Produção de frutas (y ton.)	Projeção (ton) (y = a + bx)	Variação Explicada	Variação não explicada	Variação Total
1	101,00	25,00	25,12	4,02	0,01	4,51
2	103,00	26,00	25,61	2,29	0,15	1,26
3	106,00	27,00	26,35	0,60	0,42	0,02
4	109,00	27,00	27,09	0,00	0,01	0,02
5	110,00	26,00	27,34	0,04	1,78	1,26
6	112,00	28,00	27,83	0,49	0,03	0,77
7	115,00	28,00	28,57	2,08	0,32	0,77
8	117,00	30,00	29,06	3,74	0,89	8,26
Totais		**217,00**		**13,26**	**3,61**	**16,87**

Os dados da tabela permitem ainda concluir que p por cento das variações da produção de frutas podem ser explicadas pela variação da precipitação pluviométrica. O valor mais próximo de p, para esse caso, é

a) 22%

b) 38%

c) 62%

d) 70%

e) 78%

Comentários:

A questão cobra conceitos sobre regressão.

Nessa questão, a VUNESP apresentou um enunciado muito longo, com várias informações, e depois fez uma pergunta muito simples.

Ela quer saber a relação entre a variação explicada e a total. Para isso, basta dividirmos uma pela outra e encontraremos o valor mais próximo de p.

$$\frac{13,26}{16,87} = 0,7860 \; ou \; 78,60\%$$

Dessa forma, nosso gabarito é a letra E.

Gabarito: E

---

5. VUNESP - Estatístico (TJ SP)/Judiciário/2015

Leia o texto para responder à questão.

Para se fazer um estudo sobre a relação da precipitação pluviométrica (x) em mm e a produção de frutas (y) em toneladas em certa região, coletaram-se dados durante 8 anos. Na tabela seguinte, registraram-se os valores obtidos de x e de y, respectivamente, precipitação média mensal e produção média mensal em cada ano. Na tabela, estão também os valores resultantes de alguns processamentos dos dados, incluídos valores obtidos com a reta de regressão y = a + bx.

▪ Ano	Precipitação pluviométrica (x mm)	Produção de frutas (y ton.)	Projeção (ton) (y = a + bx)	Variação Explicada	Variação não explicada	Variação Total
1	101,00	25,00	25,12	4,02	0,01	4,51
2	103,00	26,00	25,61	2,29	0,15	1,26
3	106,00	27,00	26,35	0,60	0,42	0,02
4	109,00	27,00	27,09	0,00	0,01	0,02
5	110,00	26,00	27,34	0,04	1,78	1,26
6	112,00	28,00	27,83	0,49	0,03	0,77
7	115,00	28,00	28,57	2,08	0,32	0,77
8	117,00	30,00	29,06	3,74	0,89	8,26
Totais		**217,00**		**13,26**	**3,61**	**16,87**

Define-se como coeficiente de determinação r^2 a relação $\frac{\sum(\text{Variação Explicada})}{\sum(\text{Variação Total})}$ e, a partir desse valor, é possível calcular o coeficiente de correlação, valor que mede a "força" da relação entre as variáveis estudadas. Considerando os dados da tabela, assinale a alternativa cujo valor é o que mais se aproxima do coeficiente de correlação para o caso.

a) 0,57.

b) 0,64.

c) 0,71.

d) 0,75.

e) 0,90.

Comentários:

A questão aborda conceitos sobre regressão.

Primeiramente, precisamos calcular o coeficiente de determinação. Para isso, basta aplicarmos a fórmula apresentada pela questão.

$$\frac{\sum(\text{Variação Explicada})}{\sum(\text{Variação Total})}$$

$$\frac{13{,}26}{16{,}87} = 0{,}7860$$

Sabemos que o coeficiente de correlação é a raiz quadrada do coeficiente de determinação.

$$r = \sqrt{0{,}7860}$$

Aqui devemos lembrar que números no intervalo entre 0 e 1, tirar a raiz quadrada eleva o resultado. Assim, a raiz de 0,7860 é um pouco maior que 0,7860. Ou seja, por eliminação, podemos marcar a alternativa E.

Gabarito: E

---

6. UNESP - Economista (DESENVOLVE)/2014

O resultado da estimação de uma regressão simples foi = 2 – 0,8x, sendo o coeficiente de determinação R2 = 0,81. O coeficiente de correlação entre as variáveis x e y é:

a) 0,81.

b) –0,81.

c) –0,8.

d) –0,9.

e) 0,8.

Comentários:

A questão cobra conceitos sobre regressão.

Na questão anterior, vimos que o coeficiente de correlação é a raiz quadrada do coeficiente de determinação:

$$r = \pm\sqrt{0{,}81}$$
$$r = \pm 0{,}9$$

Analisando a reta da regressão, temos um coeficiente angula negativo (-0,8). Ou seja, temos uma relação inversa entre x e y, ou seja, quando uma aumenta a outra tende a diminuir. Portanto, a correlação será negativa. Assim, utilizaremos somente a resposta – 0,90.

$$r = -0{,}9$$

Portanto, o coeficiente de correlação entre as variáveis x e y é -0,90.

Gabarito: D

---

7. FCC - Auditor Fiscal de Tributos I (São Luís)/Abrangência Geral/2018

Analisando um gráfico de dispersão referente a 10 pares de observações (t, Y t) com t = 1, 2, 3, ... , 10, optou-se por utilizar o modelo linear Y t = α + βt + ε t com o objetivo de se prever a variável Y, que representa o faturamento anual de uma empresa em milhões de reais, no ano (2007 + t). Os parâmetros α e β são desconhecidos e εt é o erro aleatório com as respectivas hipóteses do modelo de regressão linear simples. As estimativas de α e β (a e b, respectivamente) foram obtidas por meio do método dos mínimos quadrados com base nos dados dos 10 pares de observações citados. Se a = 2 e a soma dos faturamentos dos 10 dados observados foi de 64 milhões de reais, então, pela equação da reta obtida, a previsão do faturamento para 2020 é, em milhões de reais, de

a) 11,6

b) 15,0

c) 13,2

d) 12,4

e) 14,4

Comentários:

A questão cobra conceitos sobre regressão.

Precisamos encontrar o valor da previsão de faturamento de uma empresa para o ano de 2020.

A reta calculada passa pelo ponto $(\bar{t},\bar{Y})$

Assim:

$$\bar{Y} = a + b \times \bar{t}$$

Para calcular a média do faturamento, basta somar todos os faturamentos anuais e dividir por 10, já que são 10 anos considerados.

$$\bar{Y} = \frac{\sum Y}{10} = \frac{64}{10} = 6{,}4$$

Já a média de t, é obtida através da soma dos valores 1 + 2 +3 ... + 10 dividido por 10. Vejamos:

$$\bar{t} = \frac{1+2+3+4+5+6+7+8+9+10}{10} = \frac{55}{10} = 5{,}5$$

Com esses dados em mãos, podemos voltar a reta calculada anteriormente:

$$\bar{Y} = \ a + b \times \bar{t}$$

Substituindo os valores, temos:

$$6{,}4 = \ 2 + b \times 5{,}5$$

$$b = \frac{4{,}4}{5{,}5} = 0{,}80$$

Portanto, nossa reta é dada por:

$$Y = 2 + 0{,}8t$$

Como em 2020, teremos t=13. A previsão de faturamento será de:

$$Y = 2 + 0{,}8 \times 13$$

$$Y = 2 + 10{,}4 = 12{,}4$$

Portanto, a previsão do faturamento dessa empresa para 2020 será de 12,4 milhões.

Gabarito: D

---

8. FCC - Auditor Público Externo (TCE-RS)/Administração Pública ou de Empresas/2018

Utilizando o método da regressão linear, por mínimos quadrados, obteve-se a equação da reta estimada Tˆ = 20 + 0,8 t correspondente a uma série de tempo referente às vendas, em 1.000 unidades, de um produto no ano t. Esta equação foi obtida com base nas observações das vendas nos 12 primeiros anos, isto é, para t = 1, 2, 3, ... ,12.

A soma das vendas observadas, em 1.000 unidades, nesses 12 primeiros anos, foi

a) 252,6

b) 280,0

c) 302,4

d) 292,8

e) 336,0

Comentários:

A questão cobra conceitos sobre regressão linear simples.

Na regressão linear simples, quando observamos de forma gráfica duas variáveis cujo valor de correlação linear seja alto (igual a 1), a relação entre ambas se dará por uma reta (linear). Dessa forma, ela pode ser definida pela equação da reta, em que y=αx+β.

Assim, tendo os parâmetros da reta, para um dado valor de x, poderemos obter y.

Sabemos que a reta de regressão calculada passa pelos pontos médios das variáveis.

Primeiramente, devemos calcular o ponto médio de t, que assume valores inteiros de 1 a 12:

$$\bar{t} = \frac{1+2+3+4+5+6+7+8+9+10+11+12}{10} = \frac{78}{12} = 6{,}5$$

Dessa forma, temos que para t=6,5, a reta nos dá a média de produtos vendidos, em 1.000 unidades.

$$\bar{T} = 20 + 0{,}8 \times 6{,}5 = 25{,}2$$

Ou seja, foram vendidos, em média, 25,2 mil unidades por ano. Então, para 12 anos, o total vendido será dado pela multiplicação de 25,2 por 12.

$$25{,}2 \times 12 = 302{,}4\ mil\ unidades$$

Portanto, a soma das vendas observadas, em 1.000 unidades, nesses 12 primeiros anos, foi igual a 302,4 mil unidades.

Gabarito: C

---

9. FCC - Auditor Fiscal da Receita Estadual (SEF SC)/Auditoria e Fiscalização/2018

A tabela a seguir indica o valor y do salário, em número de salários mínimos (**SM)** e os respectivos tempos de serviço, em **anos,x**, de 5 funcionários de uma empresa:

x (anos)	2	3	5	3	2
y (SM)	3	4	7	4	2

Suponha que valha a relação: $y_i=\alpha+\beta x_i+\varepsilon_i$, em que i representa a i-ésima observação, α e β são parâmetros desconhecidos e ε_i é o erro aleatório com as hipóteses para a regressão linear simples. Se as estimativas de αα e β forem obtidas pelo método de mínimos quadrados por meio dessas 5 observações, a previsão de salário para um funcionário com 4 anos de serviço será, em **SM**, igual a

a) 6,1

b) 5,2

c) 6,0

d) 5,5

e) 5,8

Comentários:

Primeiro precisamos calcular a média de X e Y.

Média de X = $\frac{2+3+5+3+2}{5}$ = 3

Média de Y = $\frac{3+4+7+4+2}{5}$ = 4

Agora vamos ao cálculo da estimativa do coeficiente angular:

Valor de X	Valor de Y	(X−$\bar{X}$)	Y−$\bar{Y}^-$	(X−$\bar{X})^2$	(X−$\bar{X}$)×(Y−$\bar{Y}$)
2	3	-1	-1	1	1
3	4	0	0	0	0
5	7	2	3	4	6
3	4	0	0	0	0
2	2	-1	-2	1	2
Total				6	9

$$\boldsymbol{b} = \frac{\sum(\boldsymbol{X}-\bar{X}) \times (\boldsymbol{Y}-\bar{Y})}{\boldsymbol{2}}$$

$$b = \frac{9}{6} = 1{,}5$$

A estimativa do coeficiente linear fica:

$$\boldsymbol{a} = \bar{Y} - \boldsymbol{b}\bar{X}$$

a=4 − 1,5×3

a= −0,5

A reta calculada fica:

Y=a+bX

Y=−0,5+1,5X

Como X=4:

Y=−0,5+1,5×4=5,5

Gabarito: D

10. FCC - Analista Judiciário (TRT 13ª Região)/Apoio Especializado/Estatística/2014

Suponha que a quantidade consumida (Y) de determinado produto por uma família depende do preço do produto (X_2) e da renda da família (X_3). Consultando, aleatoriamente, 10 famílias e considerando Y_i como sendo o número de unidades consumidas do produto pela família i (i = 1,2, 3, ... ,10), X_{2i} como sendo o preço unitário (em reais) pago pela família i e X_{3i} como sendo a renda anual (em 1.000 reais) da família i, adotou-se o seguinte modelo linear $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$ para prever Y, em que ε_i é o erro aleatório com as respectivas hipóteses do modelo de regressão linear múltipla. Utilizando o método dos mínimos quadrados, obteve-se as estimativas dos parâmetros desconhecidos β_1, β_2 e β_3, com base nas informações apresentadas pelas 10 famílias. Pelo quadro de análise de variância verifica-se que a variação residual corresponde a 17,5% da variação total. Então, o valor da estatística F (F calculado) utilizado para verificar a existência da regressão, a um determinado nível de significância, é igual a

a) 15,00.

b) 12,50.

c) 14,25.

d) 10,00.

e) 16,50.

Comentários:

A questão cobra conceitos sobre o teste F.

NoTeste F, podemos utilizar os valores do Quadrado médio de SQM e SQE para a sua obtenção, onde temos:

$$F = \frac{QM\ de\ SQM}{QM\ de\ SQE}$$

FIQUE **ATENTO!**

O teste F é definido como **a razão entre duas variáveis independentes do qui-quadrado que são divididos por seus respectivos graus de liberdade**. O teste F segue a distribuição F da Snedecor.

O teste F geralmente é empregado para se verificar a **igualdade de duas variâncias populacionais**. Caso um pesquisador deseje testar se duas amostras independentes foram retiradas de uma população normal com a mesma variabilidade, então ele geralmente emprega o teste F.

A estatística F é dada por

$$F = \frac{\frac{SQE}{glSQE}}{\frac{SQR}{glSQR}}$$

Sendo que: SQE é a Soma de Quadrados Explicados; SQR é a Soma de Quadrado dos Resíduos; e **gl** são os graus de liberdade.

Pelo enunciado da questão, temos que a variação residual (SQR) corresponde a 17,5% da variação Total (SQT). Dado que: SQT=SQE+SQR; e SQE corresponde a 82,5% da SQT:

$$F = \frac{\frac{0{,}825\text{SQT}}{2}}{\frac{0{,}1750SQT}{7}}$$

Nessa questão, a SQT possui 9 graus de liberdade (n−1=10−1=9) e a SQR, 7 graus de liberdade (n−k=10−3=7). Assim, o grau de liberdade da SQE corresponde à diferença entre os graus de liberdade de SQT e SQR, ou seja: 9−7=2.

Por fim, basta isolarmos SQT e calcular:

Isolando a SQT:

$$F = \frac{\frac{0{,}825}{2}}{\frac{0{,}1750}{7}}$$

$$F = \frac{5{,}775}{0{,}35}$$

$$F = 0{,}1650$$

$$F = 0{,}1650 \times 100 = 16{,}50$$

Portanto, o valor da estatística F (F calculado) utilizado para verificar a existência da regressão, a um determinado nível de significância, é igual a 16,50.

Gabarito: E

---

11. FCC - Auditor Fiscal da Receita Estadual (SEFAZ RJ)/2014

Considere o modelo yi=α+βxi+εi,i=1,2,3,... onde:

I. yi e xi representam, respectivamente, o tempo de reação a certo estímulo, em segundos, e a idade, em anos, do indivíduo i.

II. α e β representam os parâmetros desconhecidos do modelo.

III. εi representa o erro aleatório com as respectivas hipóteses para a regressão linear simples.

IV. As estimativas de α e β foram obtidas pelo método de mínimos quadrados por meio de 10 observações, utilizando-se as seguintes informações:

$$\sum_{i=1}^{10} y_i = 1020; \sum_{i=1}^{10} x_i = 300; \sum_{i=1}^{10} x_i y_i = 40200; \sum_{i=1}^{10} x_i^2 = 13000; \sum_{i=1}^{10} y_i^2 = 128.000$$

$$\bar{x} = 30;\ \bar{y} = 102;\ \bar{x}^2 = 900;\ \bar{y}^2 = 10404.$$

Nessas condições, a soma de quadrados residuais do modelo é igual a

a) 810

b) 515

c) 920

d) 460

e) 785

Comentários:

Mais uma questão cobrando conceitos de regressão linear.

Para responder essa questão, primeiramente, precisamos calcular o b, que é o estimador de β:

$$b = \frac{\sum xy - n \times \bar{x} \times \bar{y}}{\sum x^2 - n \times \bar{x}^2}$$

$$b = \frac{40.200 - 10 \times 30 \times 102}{13.000 - 10 \times 900} = \frac{9.600}{4.000} = 2{,}4$$

Agora vamos calcular a soma dos quadrados total:

$SQT = \sum y_2 - n \times \bar{y}^2 = 128.000 - 10 \times 10.404 = 23.960$

Já a soma de quadrados do modelo de regressão será:

$$SQM = b \times (\sum xy - n \times \bar{x} \times \bar{y}) = 2{,}4 \times 9.600 = 23.040$$

Finalmente, calculamos a soma de quadrados dos resíduos:

$$SQR = SQT - SQM = 23.960 - 23.040 = 920$$

Gabarito: C

---

12. FCC - Auditor Fiscal da Fazenda Estadual (SEFAZ PI)/2015

O modelo Yt = α + βt + εt, t = 1, 2, 3, ..., foi considerado para prever o lucro de uma companhia no ano (2007 + t).

Sabe-se que:

Yt representa o lucro, em milhões de reais no ano t;

α e β são parâmetros desconhecidos;

εt é o correspondente erro aleatório, com as respectivas hipóteses da regressão linear;

as estimativas de α e β foram obtidas pelo método de mínimos quadrados, considerando-se as observações Yt no período de 6 anos (2008 a 2013).

Os dados relativos às observações são:

$$\sum_{t=1}^{6} t = 21$$

$$\sum_{t=1}^{6} t^2 = 91$$

$$\sum_{t=1}^{6} tY = 140$$

$$\sum_{t=1}^{6} Y_t = 36$$

Nessas condições, a previsão de mínimos quadrados para o lucro da companhia, em milhões de reais, no ano de 2014, é igual a

a) 7,55

b) 8,15

c) 7,90

d) 8,80

e) 9,50

Comentários:

Mais uma questão cobrando conceitos sobre regressão linear simples.

O primeiro passo será calcular as médias:

$$\bar{t} = \frac{\sum t}{n} = \frac{21}{6} = 3{,}5$$

$$\bar{Y} = \frac{\sum t}{n} = \frac{36}{6} = 6$$

Sabemos que b é a estimativa de β. Então:

$$b = \frac{\sum tY - n\bar{t}\bar{Y}}{\sum t2 - nt^{-2}}$$

$$b = \frac{140 - 6 \text{ x } 3{,}5 \text{ x } 6}{91 - 6 \text{ x } 3{,}5^{-2}}$$

$$b = \frac{140 - 126}{91 \text{ x } 73{,}5}$$

$$b = \frac{14}{17{,}5} = 0{,}80$$

Seja a a estimativa de α.

$$a = \bar{Y} - b\bar{t}$$

$$a = 6 - 0{,}8\, x\, 3{,}5$$

$$a = 3{,}2$$

Assim, para t=7, (sétimo ano da série, que é o ano de 2014), temos:

$$Y = 3{,}2 + 0{,}8 \times 7$$

$$Y = 8{,}8$$

Portanto, a previsão de mínimos quadrados para o lucro da companhia, em milhões de reais, no ano de 2014, é igual a 8,80.

Gabarito: D

---

13. FGV - Analista de Controle Interno (PrefRecife)/Finanças Públicas/2014

Numa regressão linear simples, obteve-se um coeficiente de correlação igual a 0,78. O coeficiente de determinação é aproximadamente igual a

a) 0,36.

b) 0,48.

c) 0,50.

d) 0,61.

e) 0,69.

Comentários:

Sabemos que o coeficiente de determinação é o quadrado do coeficiente de correlação:

Assim, temos:

$$R^2=0,78^2=0,6084$$

Gabarito: D

---

14. FGV - Analista Judiciário (TJ RO)/Estatístico/2015

Para fins da análise dos resíduos de uma regressão múltipla, apurou-se a seguinte tabela de decomposição amostral.

Adicionalmente são fornecidos os valores tabelados:

$F_{24|4|12,67\%}=0,5$

$F_{24|4|73,59\%}=2$

Onde,

$F_{n|m|\alpha}$= Valor da F - Snedecor para uma probabilidade acumulada inferior, com n graus de liberdade no numerador e m graus no denominador.

Então os valores de X, Y, Z e W são, respectivamente, iguais a:

a) 24.000, 28, 1/2 e 87,33%;

b) 24.000, 28, 2 e 12,67%;

c) 40.000, 20, 2 e 12,67%;

d) 24.000, 28, 1/2 e 73,59%;

e) 24.000, 28, 2 e 26,41%.

Comentários:

Para respondermos esta questão podemos encontrar primeiro o Y, somando os graus de liberdade da equação e dos resíduos.

$$Y = 4 + 24 = 28$$

Para sabermos o valor de X, será necessário somar os quadrados daequação e dos resíduos, juntas.

$$8.000+X=32.000$$

$$X=24.000$$

Agora vamos encontrar Z, partindo da estatística F que é obtida pela relação entre os quadrados médios. E cada quadrado médio é resultado da divisão entre a soma de quadrados e o número de graus de liberdade.

Quadrado médio do modelo de regressão (QMM):

QMM=8.000/4=2.000

Quadrado médio dos resíduos (QMR):

QMR=X24=24.000/ 24=1.000

Assim, Z será:

Z=2.000/ 1.000=2

Por fim, precisamos encontrar o valor de W:

Observe que a questão inverteu os graus de liberdade. Foram dados escores da distribuição F para 24 graus de liberdade no numerador e 4 no denominado. Porém, precisamos do contrário do que foi dado no enunciado. Assim, quando invertemos os graus de liberdade (usando 4 graus para o numerador e 24 para o denominador), o escore também inverte. Assim, ficaremos com:

Probabilidade acumulada	Dados do enunciado		Inversão		Nova probabilidade acumulada
12,67%	Graus de liberdade: 24 e 4	Escore: 0,5	Graus de liberdade: 4 e 24	Escore: 1/0,5 = 2	100% - 12,67% = 87,33%
73,59%	Graus de liberdade: 24 e 4	Escore: 2	Graus de liberdade: 4 e 24	Escore: 1/2 = 0,5	100% - 73,59% = 26,41%

O nosso escore calculado (2) corresponde à probabilidade acumulada de 87,33%.

Assim, o p-valor será a chance de valores maiores ou iguais ao escore 2.

P-valor = 100%−87,33%=12,67%

Portanto, gabarito letra B.

Gabarito: B

---

15. FGV - Analista Judiciário (TJ BA)/Apoio Especializado/Estatística/2015

Modelo da questão 61:

Um modelo de regressão é proposto para explicar o nível de criminalidade, considerando o grau de instrução e a classe de renda como variáveis explicativas. Formalmente,

$\ln IC_i=\alpha+\beta \ln GI_i+\gamma \ln CR_i+\varepsilon_i$

Onde IC_i, GI_i e CR_i são o índice de criminalidade, o grau de instrução e a classe de renda da localidade i, respectivamente.

Além disso, ε_i é o termo aleatório e $\ln$ representa o logaritmo neperiano. Os resultados da estimação foram os seguintes:

Parâmetro	Estimativa	Erro-Padrão	|t-Student|	p-valor
α	2.32	0.76	3.05	0.63%
β	-1.07	0.48	-2.23	3.74%
γ	-0.85	0.59	-1.44	16.51%

Sobre o modelo de regressão apresentado na questão 61, que busca explicar o comportamento do índice de criminalidade, foram realizados alguns cálculos a partir das estimativas dos parâmetros para que se pudesse realizar também uma análise da variância, cujos resultados estão na tabela abaixo:

Fonte	S. Quadrados	G. Liberdade	Q. Médio	F-Snedecor	p-valor
Equação	600	2	300	6,7	0,27%
Resíduos	900	20	45		
Total	1500	22			

Com base nos números acima é correto concluir que:

a) a regressão deve ser rejeitada, pois apenas 40% da variação total da criminalidade podem ser explicadas através dela;

b) o modelo como um todo é significativo, ao nível de 2% de significância, explicando uma parte da criminalidade;

c) o desvio-padrão dos resíduos é igual a 45;

d) a regressão apresenta um razoável grau de aderência tendo uma estatística R2 igual a 60%;

e) o coeficiente de correlação entre o índice de criminalidade e o conjunto das variáveis explicativas é de 63,24%.

Comentários:

A única alternativa correta desta questão é a letra B, Para um modelo como um todo é significativo, ao nível de 2% de significância, explicando uma parte da criminalidade.
Assim, para este nível, temos:

- p-valor = 0,27%
- nível de significância = 2%

Conclusão: Sendo o nível de significância é **maior** que o p-valor, devemos rejeitar a hipótese de nulidade dos coeficientes e concluir que **o modelo é significativo**.

Agora vamos ver os erros das outras alternativas:

A: Mesmo a variância total sendo de 40%, não é possível concluirmos que a regressão deve ser rejeitada, pois a questão não disse qual o critério de decisão.

C: Sabemos que o quadrado médio dos resíduos vale 45. Assim, para acharmos o desvio padrão será necessário calcularmos a raiz quadrada.

D: R^2 é igual a 0,40 ou 40%.

E: Podemos calcular o coeficiente de correlação por meio da raiz quadrada do coeficiente de determinação. Isso daria origem a um "coeficiente de correlação múltiplo".
Vamos tirar a raiz de 0,40:

$$\sqrt{0{,}40} = 63{,}24\%$$

Observe que o valor é igual ao exposta na letra E. Porém, não temos os logaritmos e este coeficiente não se refere diretamente à relação entre o índice de criminalidade e ao conjunto das demais variáveis e sim aos seus logaritmos.
Gabarito: B

16. FGV - Técnico Superior Especializado (DPE RJ)/Estatística/2014

Através de um estudo para fins comparativos, entre o perfil dos cidadãos que procuram a Defensoria Pública e a natureza dos seus problemas ou dificuldades levantadas, foram obtidos, considerando-se o total de processos, os seguintes percentuais:

Atributos	Ano 1	Ano 2	Ano 3	Ano 4
Percentual de Mulheres	50	60	60	70
Percentual de Causas de Família	10	20	30	20

Então, é possível afirmar que

a) exceto pelo primeiro ano, as mulheres respondem pela maior parte das causas de família.

b) a maior parte das causas de família são geradas a partir de atendimentos às mulheres.

c) o coeficiente de correlação entre os percentuais levantados é de 0,8.

d) o coeficiente de correlação entre os percentuais levantados é de 0,5.

e) a estabilidade do percentual de mulheres, entre o 2º e 3º ano, por estar acompanhada de uma elevação das causas de família demonstra que a relação existe mas é fraca.

Comentários:

Vamos analisar cada alternativa da questão:

a) exceto pelo primeiro ano, as mulheres respondem pela maior parte das causas de família.

Errada:Não temos esta informação na tabela, pois as variáveis "Percentual de Causas de Família" e "Percentual de Mulheres" referem-se ao total de processos por ano atendidos pela Defensoria Pública.

b) a maior parte das causas de família são geradas a partir de atendimentos às mulheres.

Errada: Com os dados fornecidos pela questão, também não é possível ter esta conclusão.

c) o coeficiente de correlação entre os percentuais levantados é de 0,8.

Errada: Calculando o coeficiente de correlação de Pearson para as variáveis X e Y :

Atributos	Ano 1	Ano 2	Ano 3	Ano 4	Média
Percentual de Mulheres (X)(X)	50	60	60	70	60
Percentual de Causas de Família (Y)(Y)	10	20	30	20	20

Assim, temos que:

E(X)=60

E(Y)=20

$$r(X,Y) = \frac{cov(X,Y)}{\sqrt{\boldsymbol{var\ (X).var\ (Y)}}}$$

Vamos calcular a var(X) e var(Y):

$$\mathbf{var(X)=E[(x-E(X))^2]}$$

$$\mathbf{var(X)=E[(50-60)^2+(60-60)^2+(60-60)^2+(70-60)^2]}$$

$$\mathbf{var(X)=E(200)=200}$$

$$\mathbf{var(Y)=E[(y-E(Y))^2]}$$

$$var(Y)=E[(10-20)^2+(20-20)^2+(30-20)^2+(20-20)^2]$$

$$var(Y)=E(200)=200$$

Calculando a cov(X,Y):

$$cov(X,Y)=E\{[x-E(X)]\times[y-E(Y)]\}$$

$$cov(X,Y)=E\{[50-60]\times[10-20]+[60-60]\times[20-20]+[60-60]\times[30-20]+[70-60]\times[20-20]\}$$

$$cov(X,Y)=E(100)=100$$

Agora aplicamos os valores de var(X) , var(Y)e cov(X,Y) na fórmula da correlação:

$$\mathbf{r(X,Y)} = \frac{\mathbf{100}}{\sqrt{\mathbf{200\,.\,200}}}$$

$$\mathbf{r(X,Y)} = \frac{\mathbf{100}}{\mathbf{200^2}} = \frac{\mathbf{100}}{\mathbf{200}} = \mathbf{0,5}$$

d) o coeficiente de correlação entre os percentuais levantados é de 0,5.

Correta: conforme cálculo na alternativa C.

e) a estabilidade do percentual de mulheres, entre o 2º e 3º ano, por estar acompanhada de uma elevação das causas de família demonstra que a relação existe mas é fraca.

Errada:A tabela não indica qual é o percentual de mulheres que responde pelas causas de família.

Gabarito: D

---

17. FGV - Analista Judiciário (TJ AL)/Apoio Especializado/Estatística/2018

Após estimado um Modelo de Regressão Múltipla e obtidas as estimativas dos parâmetros, o passo seguinte é a análise da variância, através das somas de quadrados. A propósito estão disponíveis as seguintes informações:

SQE = soma de quadrados da equação = 2.400

SQR = soma de quadrados dos resíduos = 1.600

Tamanho da amostra n = 41

Número de regressores = 8

$P(F_{8,32}>3)=0,9874$

Assim sendo, é correto afirmar que:

a) o R^2 do modelo estimado é igual a 40%;

b) a estatística F-Snedecor observada é igual a 6;

c) a variância estimada dos resíduos é igual a 40;

d) ao nível de significância de 98%, o modelo é rejeitado;

e) o valor do R^2 ajustado é igual a 0,55.

Comentários:

Como a questão não apresenta o valor para o SQR, precisamos encontra-lo através da fórmula:

SQR = SQT - SQE

1600 = SQT - 2400

1600 + 2400 = SQT

4000 = SQT

Agora temos os seguintes dados:

R^2 = SQE/SQT → 2400/4000 = 0,60

Para calcularmos a estatística F, devemos utilizar a fórmula abaixo:

$$F = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-k-1}}$$

$$F = \frac{\frac{0,6}{8}}{\frac{1-0,6}{41-8-1}}$$

$$F = \frac{0,075}{\frac{0,40}{32}}$$

$$F = \frac{0,075}{0,125} = 6$$

Podemos ver que a alternativa correta é a letra B: a estatística F-Snedecor observada é igual a 6.

Gabarito: B

---

18. FGV - Analista Judiciário (TJ RO)/Estatístico/2015

Para explicar o estoque total de processos acumulados nas varas de justiça (Y) a cada ano, foi proposto um modelo de regressão linear simples baseado no número de servidores disponíveis, representado por X. Depois de extraída uma amostra com n = 100 foram obtidos os seguintes resultados.

$$\sum_{i=1}^{100} Y_i.(X_i - \overline{X}) = -4800,\ Var(Y) = 625,$$

$$Var(X) = 16,\ \overline{Y} = 140 \text{ e } \overline{X} = 20$$

Supondo válido o modelo e significativos seus parâmetros, com os dados acima é correto afirmar que:

a) a cada servidor adicional lotado nas varas, o número total de processos acumulados por ano se reduz em cinco unidades;

b) o número médio de processos que chegam às varas é de 140 por ano;

c) se três novos servidores forem admitidos nas varas, o número médio de processos acumulados irá cair em nove unidades;

d) o valor do coeficiente de determinação da regressão proposta e estimada é igual a 0,48;

e) a estimativa do coeficiente linear do modelo é igual a 90.

Comentários:

Vamos calcular o coeficiente angular da reta de regressão, sabendo que o numerador da fórmula do coeficiente vale - 4.800.

Agora precisamos calcular o denominador:

$$\sum(X-\overline{X})^2$$

$$= n \times Var(X)$$

$$= 100 \times 16$$

$$= 1.600$$

Assim, temos que:

$$b = -4.800/\ 1.600 = -3$$

Com o coeficiente angular calculado, podemos concluir que a alternativa correta é a letra C.

Pois se o coeficiente angular é de -3, temos que a cada servidor adicional, o número médio de processos acumulados cais 3 unidades.

Se forem 3 novos servidores, o número de processos cai 9 unidades (3 para cada servidor adicional).

Gabarito: C

---

4 – CHECKLIST DE ESTUDO

1. Relembrar os conceitos básicos sobre Regressão.
2. Revisar Regressão Linear Simples:
   a. Conceitos iniciais e cálculo das estimativas dos parâmetros;
   b. Análise de variância da regressão linear simples. Coeficiente de determinação. Estatística F.;
   c. Inferência sobre as estimativas dos parâmetros. Teste t.
3. Regressão Linear Múltipla.
4. Outros tópicos sobre regressão (Estatística).

5 – PONTOS DE DESTAQUE

PONTO #1: REGRESSÃO: CONCEITUAÇÃO

Uma forma simples de entendermos a Regressão é perceber que esta busca trazer uma forma de compreendermos como se dá a **relação entre duas variáveis** e, a partir do dado de uma destas variáveis, obter o valor da outra.

ESCLARECENDO

Vamos exemplificar: digamos que nós tivéssemos coletado o valor médio dos imóveis a partir do centro de uma cidade. Depois, perfazendo círculos concêntricos com raios aumentando a cada 10 quilômetros (imagine uma imagem semelhante a um alvo), começássemos a tomar nota do valor médio dos imóveis de cada área/distância durante os próximos 100 quilômetros. Ao final, teríamos uma coletânea que poderia ser similar a tabela abaixo:

Valor em mil reais	Distância do Centro em Km
650	0
600	10
550	20
500	30
450	40
400	50
350	60
300	70
250	80
200	90

De posse dessa tabela, podemos tirar algumas conclusões:

- Primeira conclusão - quanto mais nos distanciamos do centro, mais o valor do imóvel diminui.
- Segunda conclusão - se fizéssemos a correlação linear dos dados desse exemplo, veríamos que esta é igual a 1.Isto é, indica que a relação entre as duas variáveis pode ser representada por uma reta (a cada 10 km de distância, temos 50 mil reais a menos no valor do imóvel).

Em resumo para qualquer valor de distância até o centro, poderíamos obter um determinado valor para o imóvel na mesma proporção. Esse exemplo é hipotético, pois raramente teremos uma correlação linear igual a 1. No entanto, quanto mais próximo desse valor está a regressão, mais confiável ela é.

RESUMINDO

CONCEITO

Em outras palavras, a **regressão** é uma forma de se obter o valor de uma certa variável Y (**dependente** ou regredida), que geralmente é medida ou inferida, a partir de dados de uma outra variável X (**independente** ou regressora), geralmente conhecida e fixa, como a distância apresentada no nosso exemplo.

Assim, a variável Y sempre tem que depender da variável X e não contrário. **Percebam que a correlação linear nos ajuda a entendermos a dependência entre as variáveis.** Pelo exemplo hipotético, sabemos que a distância em relação ao centro é fundamental para sabermos o valor do imóvel. No entanto, no mundo real, poderíamos dizer que existem outros elementos, ou variáveis, que influenciam na composição do valor do imóvel. Veremos como se dá isso mais adiante.

De forma geral, a regressão ou os modelos de regressão abrangem um campo da estatística que busca identificar a relação entre duas ou mais variáveis visando a estimação de uma destas variáveis.

PONTO #2: REGRESSÃO LINEAR SIMPLES: CONCEITOS INICIAIS

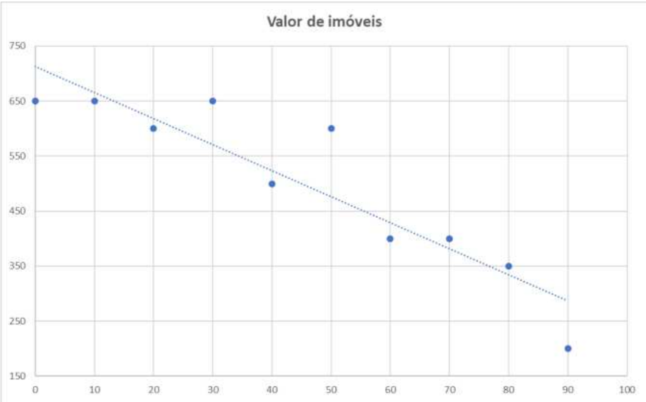
Como foi dito no tópico anterior, quando se observa de forma gráfica duas variáveis cujo valor de correlação linear seja alto (igual a 1), a relação entre ambas se dará por uma reta (linear).Dessa forma, ela pode ser definida pela equação da reta, em que $y = \alpha x + \beta$.

Assim, tendo os parâmetros da reta, para um dado valor de x, poderemos obter y. Resta agora identificarmos como obtermos os parâmetros α e β.Primeiro, vamos ver uma outra tabela (hipotética) com novos valores de imóveis:

Valor em mil reais	Distância do Centro em Km
650	0
650	10
600	20
650	30
500	40
600	50
400	60
400	70
350	80
200	90

Nessa segunda tabela, vemos que os valores já não estão seguindo uma tendência tão clara de redução de valor a cada 10 quilômetros.Percebemos inclusive valores que são mais altos em distâncias maiores (quebrando a lógica do "quanto mais distante, mais barato fica").

Se quiséssemos correlacionar esses dados em um gráfico e tentássemos traçar uma reta, resultaria em uma figura similar a seguinte:



Essa hipotética reta que passa por entre os pontos seria o exemplo de um **modelo de regressão**. No entanto, vemos que alguns pontos estão próximos da reta e outros mais afastados. De modo geral, um bom modelo de regressão visa traçar uma reta em que os **desvios médios sejam os menores possíveis**.

Em uma reta cuja equação seja definida por $y_i = \alpha x_i + \beta$, sabemos que α é o coeficiente angular e β a interseção. E, para calcular cada um dos parâmetros, seguimos as seguintes equações:

$$\alpha = \frac{\sum[(x_i - \bar{X}) \times (y_i - \bar{Y})]}{\sum(x_i - \bar{X})^2}$$

$$\beta = \bar{Y} - \alpha\bar{X}$$

Vejam que os parâmetros são obtidos a partir do próprio conjunto de valores, e basicamente giram em função dos desvios, seja em função de X ou em função de Y. Percebam, também, que o denominador de α nada mais é do que o somatório dos mínimos quadrados.

Existem outras formulações para o cálculo do parâmetroα como:

$$\alpha = \frac{\sum xy}{\sum x^2}$$

Vejamos um exemplo de questão:

CESPE - Auditor Federal de Controle Externo/Apoio Técnico e Administrativo/Planejamento e Gestão/2008

Uma agência de desenvolvimento urbano divulgou os dados apresentados na tabela a seguir, acerca dos números de imóveis ofertados (X) e vendidos (Y) em determinado município, nos anos de 2005 a 2007.

ano	número de imóveis	
	ofertados (X)	vendidos (Y)
2005	1.500	100
2006	1.750	400
2007	2.000	700

Correio Braziliense, 29/4/2008, p. 17 (com adaptações).

Considerando as informações do texto, julgue os itens subsequentes.
A estimativa do valor do coeficiente α da reta de regressão Y=αX, em que Y representa o número esperado de imóveis vendidos para uma quantidade X de imóveis ofertados, é superior a 0,23 e inferior a 0,26.
Certo ou Errado?

Para calcular α, temos:

$$\alpha = \frac{\sum xy}{\sum x^2} \rightarrow \alpha = \frac{(1500 \times 10^2) + (1750 \times 4 \times 10^2) + (2000 \times 7 \times 10^2)}{1500^2 + 1750^2 + 2000^2}$$

$$\alpha = \frac{1500 \times 10^2 + 7000 \times 10^2 + 14000 \times 10^2}{93125 \ \times 10^2} = \frac{22500}{93125} \cong 0{,}24$$

PONTO #3: ANÁLISE DE VARIÂNCIA DA REGRESSÃO LINEAR SIMPLES

Para encontrarmos os valores que seguem uma reta de regressão, nós utilizamos a equação geral abaixo:

$$y = \alpha x + \beta$$

No entanto, para se obter o correto valor de y em uma reta, nós teríamos que somar nessa equação o erro (e_i) que está associado à própria regressão.

$$y_i = \alpha x_i + \beta + e_i$$

Assim, sendo y_i o valor observado e y o valor estimado pela reta de regressão, poderíamos identificar o erro a partir da seguinte equação:

$$e_i = y_i - y$$

Sem muitos preâmbulos, com base nas equações que nós já vimos até agora, poderíamos dizer que:

$$\sum (y_i - \bar{y})^2 = \sum {e_i}^2 + \sum (\hat{y} - \bar{y})^2$$

E cada uma das partes desta equação assume alguns papéis que são por vezes requisitadas em questões de concurso, vejamos:

$$\sum (y_i - \bar{y})^2 \rightarrow Soma\ dos\ Quadrados\ Total\ (SQT)$$

$$\sum(\hat{y}-\bar{y})^2 \rightarrow Soma\ dos\ Quadrados\ do\ Modelo\ da\ Regressão\ (SQM)$$

$$\sum e_i{}^2 \rightarrow Soma\ dos\ Quadrados\ do\ Erro\ (SQE)$$

QUADRADOS MÉDIOS

Para calcularmos os valores médios (Quadrados Médios) de cada uma dessas parcelas, devemos dividi-las por seus respectivos **graus de liberdade**, quais são:

$$GL\ de\ SQT \rightarrow n-1$$

$$GL\ de\ SQM \rightarrow 1$$

$$GL\ de\ SQE \rightarrow n-2$$

Obtendo os quadrados médios:

$$Quadrado\ médio\ de\ SQT \rightarrow \frac{SQT}{n-1}$$

$$Quadrado\ médio\ de\ SQM \rightarrow \frac{SQM}{1}$$

$$Quadrado\ médio\ de\ SQE \rightarrow \frac{SQE}{n-2}$$

Mas professor, onde posso usar essas informações? Calma, jovem, vamos ver como calculamos o TESTE F.

TESTE F

No Teste F, podemos utilizar os valores do Quadrado médio de SQM e SQE para a sua obtenção, onde temos:

$$F = \frac{QM\ de\ SQM}{QM\ de\ SQE}$$

FIQUE ATENTO!

O teste F é definido como **a razão entre duas variáveis independentes do qui-quadrado que são divididos por seus respectivos graus de liberdade**. O teste F segue a distribuição F da Snedecor.

O teste F geralmente é empregado para se verificar a **igualdade de duas variâncias populacionais**. Caso um pesquisador deseje testar se duas amostras independentes foram retiradas de uma população normal com a mesma variabilidade, então ele geralmente emprega o teste F.

Atenção! Este teste também pode ser utilizado para determinar se duas estimativas independentes das variâncias populacionais são homogêneas.

No caso de análises de regressão, o Teste F é empregado para se verificar a aderência do modelo, isto é, o quão explicativo é o modelo testado. No qual a rejeição de H0 representa que “ao menos uma variável explicativa contribui significativamente para o modelo”.

COEFICIENTE DE DETERMINAÇÃO

Ainda se utilizando das partes sublinhadas neste tópico, podemos também obter o coeficiente de determinação, em que:

$$r^2 = \frac{SQM}{SQT}$$

O r² representa a proporção da variância na variável dependente, que é predita da variável independente, variando de 0 a 1. Onde: quanto mais perto de 1, mais explicativo é o modelo.

PONTO #4: INFERÊNCIA SOBRE AS ESTIMATIVAS DOS PARÂMETROS. TESTE T

Ainda quando analisamos a regressão linear, podemos obter algumas inferências sobre as estimativas dos parâmetros.

Primeiramente, podemos obter a Variância de $\bar{Y}$, em que:

$$Var(\bar{Y}) = \frac{\sigma^2}{n}$$

Percebam que nesse caso dividimos a variância pela quantidade de elementos. Já a variância de β pode ser obtida da seguinte forma:

$$Var(\beta) = \frac{\sigma^2}{\sum x_i^2}$$

Nesse caso, temos a variância dividida pelo somatório dos quadrados dos elementos.

Com respeito à variância de α, temos a seguinte relação:

$$Var(\alpha) = \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right)\sigma^2$$

Para a obtenção das covariâncias, temos que:

$$Cov(\bar{X}, \beta) = 0$$

E:

$$Cov(\alpha, \beta) = -\frac{\bar{X}\sigma^2}{\sum x_i^2}$$

O TESTE T

Quando se pretende se **verificar a significância de uma regressão**, isto é, a adequação do modelo de predição, podem ser realizados testes de hipóteses para determinar se o valor de α é igual a Zero. Um desses teste é o teste T, em que:

$$t_{amostra} = \frac{\alpha}{\sqrt{\frac{QM\ de\ SQE}{SQE}}}$$

Sendo que:

$$H_0: \alpha = 0$$

$$H_1: \alpha \neq 0$$

Para este teste, quando analisamos regressão, rejeita-se H_0, se:

$$t_{amostra} < \ -t_{\frac{n}{2}}(n-2)\ ou\ t_{amostra} > t_{\frac{n}{2}}(n-2)$$

PONTO #5: REGRESSÃO LINEAR MÚLTIPLA

Quando foi falado de regressão linear simples, vimos a relação entre uma variável x com uma variável y, onde:

$$y = \alpha x + \beta$$

Para a regressão múltipla, a ideia é a de que podemos correlacionar uma série de variáveis x_1, x_2, x_3 ..., ditas independentes, que influenciam uma determinada variável y.

No exemplo que demos dos valores dos imóveis em função da distância do centro, poderíamos supor que outras variáveis numéricas como número de quartos e banheiros nos imóveis, presença de hospitais e/ou área de lazer perto, e outros elementos ajudariam na composição do valor dos imóveis.

Assim, a regressão linear múltipla busca identificar a relação que se dá entre essas variáveis x e a variável y. Dessa forma, a equação geral pode ser dada por:

$$y = a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \ldots . + a_nx_n + \ \beta + e$$

Para a obtenção dos parâmetros a, podemos calcular utilizando o recurso das matrizes da seguinte forma:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ . \\ y_n \end{bmatrix} = \begin{bmatrix} X_{11} X_{21} X_{31} \cdots X_{k1} \\ X_{12} X_{22} X_{32} \cdots X_{k2} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ . \quad . \quad . \quad \cdots \quad . \\ X_{1n} X_{2n} X_{3n} \cdots X_{kn} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ . \\ a_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ . \\ e_n \end{bmatrix}$$

Sendo:

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = (X^T X)^{-1} X^T Y$$

PONTO #6: OUTROS TÓPICOS SOBRE REGRESSÃO (ESTATÍSTICA)

Os cálculos de variância para a regressão múltipla são similares aos da regressão simples, em que:

$$Soma\ de\ Quadrados\ dos\ Erros = \sum (y_i - \hat{y}_i)^2$$

$$Soma\ de\ Quadrados\ Totais = \sum (y_i - \bar{y})^2$$

$$Soma\ de\ Quadrados\ do\ Modelo = \sum (\hat{y} - \bar{y})^2$$

A relação de graus de liberdade para os somatórios pode ser sintetizada em:

Somatório	Graus de liberdade
SQT	$n - 1$
SQM	$k - 1$,em que k representa o número de parâmetros.
SQR	$n - k$

Tanto o teste F quanto o R^2 se apresentam da mesma forma:

$$F = \frac{QM\ de\ SQM}{QM\ de\ SQE}$$

$$r^2 = \frac{SQM}{SQT}$$

6 - Considerações Finais

Chegamos ao final desse nosso primeiro relatório do Passo Estratégico para **Auditor Fiscal do ISS Campinas.**

É preciso entender que estamos diante de assuntos de muita importância para a sua prova. Por isso, prestem bastante atenção nesses assuntos e não deixem de revisar esses pontos.

As questões trazidas neste relatório servem apenas como exemplo, por isso encorajamos que vocês arregacem as mangas e pratiquem bastante. Fazer o máximo de questões possível vai aproximar vocês da excelência.

Por hoje é só!

Perseverança e bons estudos!

7- Lista das Questões

1. **VUNESP - Especialista em Regulação e Fiscalização de Serviços Públicos I (ARSESP)/Econômico Financeiro/2018**

Usando dados amostrais para estudar a correlação entre preço x da gasolina (em reais) e movimento y de vendas semanais (em litros) em postos de combustíveis de certa região, um grupo de pesquisadores verificou existir correlação linear entre as duas variáveis. A reta de regressão y = bx + a estabelecida no estudo tem coeficiente angular – 4,50 e coeficiente linear 15 500 (valores aproximados). Suponha que o preço R$ 4,00 por litro pertença ao intervalo de preços verificados na pesquisa. Usando a reta de regressão para uma estimativa do movimento de vendas, e considerando uma unidade de venda (ou posto) com preço da gasolina de R$ 4,00 por litro, então o movimento semanal de vendas (em litros) estimado nesse posto será de

a) 15 594,50.

b) 15 518.

c) 15 482.

d) 15 598,30.

e) 15 000.

2. **VUNESP - Estatístico (TJ SP)/Judiciário/2015**

Leia o texto para responder à questão.

Para se fazer um estudo sobre a relação da precipitação pluviométrica (x) em mm e a produção de frutas (y) em toneladas em certa região, coletaram-se dados durante 8 anos. Na tabela seguinte, registraram-se os valores obtidos de x e de y, respectivamente, precipitação média mensal e produção média mensal em cada ano. Na tabela, estão também os valores resultantes de alguns processamentos dos dados, incluídos valores obtidos com a reta de regressão y = a + bx.

Ano	Precipitação pluviométrica (x mm)	Produção de frutas (y ton.)	Projeção (ton) (y = a + bx)	Variação Explicada	Variação não explicada	Variação Total
1	101,00	25,00	25,12	4,02	0,01	4,51
2	103,00	26,00	25,61	2,29	0,15	1,26
3	106,00	27,00	26,35	0,60	0,42	0,02
4	109,00	27,00	27,09	0,00	0,01	0,02
5	110,00	26,00	27,34	0,04	1,78	1,26
6	112,00	28,00	27,83	0,49	0,03	0,77
7	115,00	28,00	28,57	2,08	0,32	0,77
8	117,00	30,00	29,06	3,74	0,89	8,26
Totais		**217,00**		**13,26**	**3,61**	**16,87**

Ao se montar a tabela ANOVA para esse caso, obtiveram-se os seguintes valores:

Fonte de Variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	1	13,26		
Resíduo	6	3,61		
TOTAL	7	16,87		

No teste de hipótese ao nível de 5% de significância para verificar a linearidade (H_0: $\beta = 0$ contra H_1: $\beta \neq 0$), o valor crítico de F para se rejeitar H_0 é, aproximadamente,

a) 19,4.

b) 8,85.

c) 5,99.

d) 3,47.

e) 3,01.

3. VUNESP - Estatístico (TJ SP)/Judiciário/2015

Leia o texto para responder à questão.

Para se fazer um estudo sobre a relação da precipitação pluviométrica (x) em mm e a produção de frutas (y) em toneladas em certa região, coletaram-se dados durante 8 anos. Na tabela seguinte, registraram-se os valores obtidos de x e de y, respectivamente, precipitação média mensal e produção média mensal em cada ano. Na tabela, estão também os valores resultantes de alguns processamentos dos dados, incluídos valores obtidos com a reta de regressão y = a + bx.

Ano	Precipitação pluviométrica (x mm)	Produção de frutas (y ton.)	Projeção (ton) (y = a + bx)	Variação Explicada	Variação não explicada	Variação Total
1	101,00	25,00	25,12	4,02	0,01	4,51
2	103,00	26,00	25,61	2,29	0,15	1,26
3	106,00	27,00	26,35	0,60	0,42	0,02
4	109,00	27,00	27,09	0,00	0,01	0,02
5	110,00	26,00	27,34	0,04	1,78	1,26
6	112,00	28,00	27,83	0,49	0,03	0,77
7	115,00	28,00	28,57	2,08	0,32	0,77
8	117,00	30,00	29,06	3,74	0,89	8,26
Totais		**217,00**		**13,26**	**3,61**	**16,87**

Ao se montar a tabela ANOVA para esse caso, obtiveram-se os seguintes valores:

Fonte de Variação	Graus de liberdade	Soma de quadrados	Quadrado médio	F
Regressão	1	13,26		
Resíduo	6	3,61		
TOTAL	7	16,87		

Ao complementar a tabela e calcular o valor de F, encontra-se o valor aproximado de

a) 13.

b) 22.

c) 30.

d) 38.

e) 42.

4. VUNESP - Estatístico (TJ SP)/Judiciário/2015

Leia o texto para responder à questão.

Para se fazer um estudo sobre a relação da precipitação pluviométrica (x) em mm e a produção de frutas (y) em toneladas em certa região, coletaram-se dados durante 8 anos. Na tabela seguinte, registraram-se os valores obtidos de x e de y, respectivamente, precipitação média mensal e produção média mensal em cada ano. Na tabela, estão também os valores resultantes de alguns processamentos dos dados, incluídos valores obtidos com a reta de regressão y = a + bx.

Ano	Precipitação pluviométrica (x mm)	Produção de frutas (y ton.)	Projeção (ton) (y = a + bx)	Variação Explicada	Variação não explicada	Variação Total
1	101,00	25,00	25,12	4,02	0,01	4,51
2	103,00	26,00	25,61	2,29	0,15	1,26
3	106,00	27,00	26,35	0,60	0,42	0,02
4	109,00	27,00	27,09	0,00	0,01	0,02
5	110,00	26,00	27,34	0,04	1,78	1,26
6	112,00	28,00	27,83	0,49	0,03	0,77
7	115,00	28,00	28,57	2,08	0,32	0,77
8	117,00	30,00	29,06	3,74	0,89	8,26
Totais		**217,00**		**13,26**	**3,61**	**16,87**

Os dados da tabela permitem ainda concluir que p por cento das variações da produção de frutas podem ser explicadas pela variação da precipitação pluviométrica. O valor mais próximo de p, para esse caso, é

a) 22%

b) 38%

c) 62%

d) 70%

e) 78%

5. VUNESP - Estatístico (TJ SP)/Judiciário/2015

Leia o texto para responder à questão.

Para se fazer um estudo sobre a relação da precipitação pluviométrica (x) em mm e a produção de frutas (y) em toneladas em certa região, coletaram-se dados durante 8 anos. Na tabela seguinte, registraram-se os valores obtidos de x e de y, respectivamente, precipitação média mensal e produção média mensal em cada ano. Na tabela, estão também os valores resultantes de alguns processamentos dos dados, incluídos valores obtidos com a reta de regressão y = a + bx.

Ano	Precipitação pluviométrica (x mm)	Produção de frutas (y ton.)	Projeção (ton) (y = a + bx)	Variação Explicada	Variação não explicada	Variação Total
1	101,00	25,00	25,12	4,02	0,01	4,51
2	103,00	26,00	25,61	2,29	0,15	1,26
3	106,00	27,00	26,35	0,60	0,42	0,02
4	109,00	27,00	27,09	0,00	0,01	0,02
5	110,00	26,00	27,34	0,04	1,78	1,26
6	112,00	28,00	27,83	0,49	0,03	0,77
7	115,00	28,00	28,57	2,08	0,32	0,77
8	117,00	30,00	29,06	3,74	0,89	8,26
Totais		**217,00**		**13,26**	**3,61**	**16,87**

Define-se como coeficiente de determinação r^2 a relação $\dfrac{\sum(\text{Variação Explicada})}{\sum(\text{Variação Total})}$ e, a partir desse valor, é possível calcular o coeficiente de correlação, valor que mede a "força" da relação entre as variáveis estudadas. Considerando os dados da tabela, assinale a alternativa cujo valor é o que mais se aproxima do coeficiente de correlação para o caso.

a) 0,57.

b) 0,64.

c) 0,71.

d) 0,75.

e) 0,90.

6. UNESP - Economista (DESENVOLVE)/2014

O resultado da estimação de uma regressão simples foi = 2 – 0,8x, sendo o coeficiente de determinação R2 = 0,81. O coeficiente de correlação entre as variáveis x e y é:

a) 0,81.

b) –0,81.

c) –0,8.

d) –0,9.

e) 0,8.

7. FCC - Auditor Fiscal de Tributos I (São Luís)/Abrangência Geral/2018

Analisando um gráfico de dispersão referente a 10 pares de observações (t, Y t) com t = 1, 2, 3, ... , 10, optou-se por utilizar o modelo linear Y t = α + βt + ε t com o objetivo de se prever a variável Y, que representa o faturamento anual de uma empresa em milhões de reais, no ano (2007 + t). Os parâmetros α e β são desconhecidos e εt é o erro aleatório com as respectivas hipóteses do modelo de regressão linear simples. As estimativas de α e β (a e b, respectivamente) foram obtidas por meio do método dos mínimos quadrados com base nos dados dos 10 pares de observações citados. Se a = 2 e a soma dos faturamentos dos 10 dados observados foi de 64 milhões de reais, então, pela equação da reta obtida, a previsão do faturamento para 2020 é, em milhões de reais, de

a) 11,6

b) 15,0

c) 13,2

d) 12,4

e) 14,4

8. FCC - Auditor Público Externo (TCE-RS)/Administração Pública ou de Empresas/2018

Utilizando o método da regressão linear, por mínimos quadrados, obteve-se a equação da reta estimada Tˆ = 20 + 0,8 t correspondente a uma série de tempo referente às vendas, em 1.000 unidades, de um produto no ano t. Esta equação foi obtida com base nas observações das vendas nos 12 primeiros anos, isto é, para t = 1, 2, 3, ... ,12.

A soma das vendas observadas, em 1.000 unidades, nesses 12 primeiros anos, foi

a) 252,6

b) 280,0

c) 302,4

d) 292,8

e) 336,0

9. FCC - Auditor Fiscal da Receita Estadual (SEF SC)/Auditoria e Fiscalização/2018

A tabela a seguir indica o valor y do salário, em número de salários mínimos (**SM)** e os respectivos tempos de serviço, em **anos**,**x**, de 5 funcionários de uma empresa:

x (anos)	2	3	5	3	2
y (SM)	3	4	7	4	2

Suponha que valha a relação: $y_i=\alpha+\beta x_i+\varepsilon_i$, em que i representa a i-ésimaobservação, α e β são parâmetros desconhecidos e ε_i é o erro aleatório com as hipóteses para a regressão linear simples. Se as estimativas de $\alpha\alpha$ e β forem obtidas pelo método de mínimos quadrados por meio dessas 5 observações, a previsão de salário para um funcionário com 4 anos de serviço será, em **SM**, igual a

a) 6,1

b) 5,2

c) 6,0

d) 5,5

e) 5,8

10. FCC - Analista Judiciário (TRT 13ª Região)/Apoio Especializado/Estatística/2014

Suponha que a quantidade consumida (Y) de determinado produto por uma família depende do preço do produto (X 2) e da renda da família (X 3). Consultando, aleatoriamente, 10 famílias e considerando Yi como sendo o número de unidades consumidas do produto pela família i (i = 1,2, 3, ... ,10), X 2i como sendo o preço unitário (em reais) pago pela família i e X3i como sendo a renda anual (em 1.000 reais) da família i, adotou-se o seguinte modelo linear $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$ para prever Y, em que εi é o erro aleatório com as respectivas hipóteses do modelo de regressão linear múltipla. Utilizando o método dos mínimos quadrados, obteve-se as estimativas dos parâmetros desconhecidos β_1, β_2 e β_3, com base nas informações apresentadas pelas 10 famílias. Pelo quadro de análise de variância verifica-se que a variação residual corresponde a 17,5% da variação total. Então, o valor da estatística F (F calculado) utilizado para verificar a existência da regressão, a um determinado nível de significância, é igual a

a) 15,00.

b) 12,50.

c) 14,25.

d) 10,00.

e) 16,50.

11. FCC - Auditor Fiscal da Receita Estadual (SEFAZ RJ)/2014

Considere o modelo yi=α+βxi+εi,i=1,2,3,... onde:

I. yi e xi representam, respectivamente, o tempo de reação a certo estímulo, em segundos, e a idade, em anos, do indivíduo i.

II. α e β representam os parâmetros desconhecidos do modelo.

III. εi representa o erro aleatório com as respectivas hipóteses para a regressão linear simples.

IV. As estimativas de α e β foram obtidas pelo método de mínimos quadrados por meio de 10 observações, utilizando-se as seguintes informações:

$$\sum_{i=1}^{10} y_i = 1020;\ \sum_{i=1}^{10} x_i = 300;\ \sum_{i=1}^{10} x_i\, y_i = 40200;\ \sum_{i=1}^{10} x_i^2 = 13000;\ \sum_{i=1}^{10} y_i^2 = 128.000$$

$$\bar{x} = 30;\ \bar{y} = 102;\ \bar{x}^2 = 900;\ \bar{y}^2 = 10404.$$

Nessas condições, a soma de quadrados residuais do modelo é igual a

a) 810

b) 515

c) 920

d) 460

e) 785

12. FCC - Auditor Fiscal da Fazenda Estadual (SEFAZ PI)/2015

O modelo Yt = α + βt + εt, t = 1, 2, 3, ..., foi considerado para prever o lucro de uma companhia no ano (2007 + t).

Sabe-se que:

Yt representa o lucro, em milhões de reais no ano t;

α e β são parâmetros desconhecidos;

εt é o correspondente erro aleatório, com as respectivas hipóteses da regressão linear;

as estimativas de α e β foram obtidas pelo método de mínimos quadrados, considerando-se as observações Yt no período de 6 anos (2008 a 2013).

Os dados relativos às observações são:

$$\sum_{t=1}^{6} t = 21$$

$$\sum_{t=1}^{6} t^2 = 91$$

$$\sum_{t=1}^{6} tY = 140$$

$$\sum_{t=1}^{6} Y_t = 36$$

Nessas condições, a previsão de mínimos quadrados para o lucro da companhia, em milhões de reais, no ano de 2014, é igual a

a) 7,55

b) 8,15

c) 7,90

d) 8,80

e) 9,50

13. FGV - Analista de Controle Interno (PrefRecife)/Finanças Públicas/2014

Numa regressão linear simples, obteve-se um coeficiente de correlação igual a 0,78. O coeficiente de determinação é aproximadamente igual a

a) 0,36.

b) 0,48.

c) 0,50.

d) 0,61.

e) 0,69.

14. FGV - Analista Judiciário (TJ RO)/Estatístico/2015

Para fins da análise dos resíduos de uma regressão múltipla, apurou-se a seguinte tabela de decomposição amostral.

Adicionalmente são fornecidos os valores tabelados:

$F_{24|4|12,67\%}=0,5$

$F_{24|4|73,59\%}=2$

Onde,

$F_{n|m|\alpha}$= Valor da F - Snedecor para uma probabilidade acumulada inferior, com n graus de liberdade no numerador e m graus no denominador.

Então os valores de X, Y, Z e W são, respectivamente, iguais a:

a) 24.000, 28, 1/2 e 87,33%;

b) 24.000, 28, 2 e 12,67%;

c) 40.000, 20, 2 e 12,67%;

d) 24.000, 28, 1/2 e 73,59%;

e) 24.000, 28, 2 e 26,41%.

15. FGV - Analista Judiciário (TJ BA)/Apoio Especializado/Estatística/2015

Modelo da questão 61:

Um modelo de regressão é proposto para explicar o nível de criminalidade, considerando o grau de instrução e a classe de renda como variáveis explicativas. Formalmente,

$\ln IC_i=\alpha+\beta \ln GI_i+\gamma \ln CR_i+\varepsilon_i$

Onde IC_i, GI_i e CR_i são o índice de criminalidade, o grau de instrução e a classe de renda da localidade i, respectivamente.

Além disso, ε_i é o termo aleatório e $\ln$ representa o logaritmo neperiano. Os resultados da estimação foram os seguintes:

Parâmetro	Estimativa	Erro-Padrão	|t-Student|	p-valor
α	2.32	0.76	3.05	0.63%

β	-1.07	0.48	-2.23	3.74%
γ	-0.85	0.59	-1.44	16.51%

Sobre o modelo de regressão apresentado na questão 61, que busca explicar o comportamento do índice de criminalidade, foram realizados alguns cálculos a partir das estimativas dos parâmetros para que se pudesse realizar também uma análise da variância, cujos resultados estão na tabela abaixo:

Fonte	S. Quadrados	G. Liberdade	Q. Médio	F-Snedecor	p-valor
Equação	600	2	300	6,7	0,27%
Resíduos	900	20	45		
Total	1500	22			

Com base nos números acima é correto concluir que:

a) a regressão deve ser rejeitada, pois apenas 40% da variação total da criminalidade podem ser explicadas através dela;

b) o modelo como um todo é significativo, ao nível de 2% de significância, explicando uma parte da criminalidade;

c) o desvio-padrão dos resíduos é igual a 45;

d) a regressão apresenta um razoável grau de aderência tendo uma estatística R2 igual a 60%;

e) o coeficiente de correlação entre o índice de criminalidade e o conjunto das variáveis explicativas é de 63,24%.

16. FGV - Técnico Superior Especializado (DPE RJ)/Estatística/2014

Através de um estudo para fins comparativos, entre o perfil dos cidadãos que procuram a Defensoria Pública e a natureza dos seus problemas ou dificuldades levantadas, foram obtidos, considerando-se o total de processos, os seguintes percentuais:

Atributos	Ano 1	Ano 2	Ano 3	Ano 4
Percentual de Mulheres	50	60	60	70
Percentual de Causas de Família	10	20	30	20

Então, é possível afirmar que

a) exceto pelo primeiro ano, as mulheres respondem pela maior parte das causas de família.

b) a maior parte das causas de família são geradas a partir de atendimentos às mulheres.

c) o coeficiente de correlação entre os percentuais levantados é de 0,8.

d) o coeficiente de correlação entre os percentuais levantados é de 0,5.

e) a estabilidade do percentual de mulheres, entre o 2º e 3º ano, por estar acompanhada de uma elevação das causas de família demonstra que a relação existe mas é fraca.

17. FGV - Analista Judiciário (TJ AL)/Apoio Especializado/Estatística/2018

Após estimado um Modelo de Regressão Múltipla e obtidas as estimativas dos parâmetros, o passo seguinte é a análise da variância, através das somas de quadrados. A propósito estão disponíveis as seguintes informações:

SQE = soma de quadrados da equação = 2.400

SQR = soma de quadrados dos resíduos = 1.600

Tamanho da amostra n = 41

Número de regressores = 8

P(F8,32>3)=0,9874

Assim sendo, é correto afirmar que:

a) o R^2 do modelo estimado é igual a 40%;

b) a estatística F-Snedecor observada é igual a 6;

c) a variância estimada dos resíduos é igual a 40;

d) ao nível de significância de 98%, o modelo é rejeitado;

e) o valor do R^2 ajustado é igual a 0,55.

18. FGV - Analista Judiciário (TJ RO)/Estatístico/2015

Para explicar o estoque total de processos acumulados nas varas de justiça (Y) a cada ano, foi proposto um modelo de regressão linear simples baseado no número de servidores disponíveis, representado por X. Depois de extraída uma amostra com n = 100 foram obtidos os seguintes resultados.

$$\sum_{i=1}^{100} Y_i.(X_i - \overline{X}) = -4800,\ Var(Y) = 625,$$

$$Var(X) = 16,\ \overline{Y} = 140 \text{ e } \overline{X} = 20$$

Supondo válido o modelo e significativos seus parâmetros, com os dados acima é correto afirmar que:

a) a cada servidor adicional lotado nas varas, o número total de processos acumulados por ano se reduz em cinco unidades;

b) o número médio de processos que chegam às varas é de 140 por ano;

c) se três novos servidores forem admitidos nas varas, o número médio de processos acumulados irá cair em nove unidades;

d) o valor do coeficiente de determinação da regressão proposta e estimada é igual a 0,48;

e) a estimativa do coeficiente linear do modelo é igual a 90.

8 - GABARITO

1) C
2) C
3) B
4) C
5) E
6) D
7) D
8) C
9) D
10) E
11) C
12) D
13) D
14) B
15) B
16) D
17) B
18) C

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.