

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE RIBEIRÃO PRETO
DEPARTAMENTO DE ADMINISTRAÇÃO

RAD1507 – Estatística Aplicada à Administração I

Lista 10

Nome: _____

Correlação e regressão linear.

Exercícios 14.6 Tópico 14.6 (Refazer) 14.41
 14.49 14.62

14.6 Uma empresa de produtos para consumo deseja medir a eficácia de diferentes meios de propaganda na promoção de seus produtos. Especificamente, a empresa está interessada na eficácia da propaganda no rádio e em jornais (incluindo o custo de cupons de desconto). É selecionada para estudo uma amostra de 22 cidades, com populações aproximadamente iguais, durante um período de teste de um mês. A cada cidade é alocado um nível específico de despesas, tanto para propaganda em rádio quanto para propaganda em jornais. As vendas do produto (em milhares de dólares), bem como os níveis de despesa com os meios de propaganda, durante o mês de teste, foram registradas, com os seguintes resultados, armazenados no arquivo ~~propaganda.xls~~:

Cidade	Vendas (\$1.000)	Propaganda em Rádio (\$1.000)	Propaganda em Jornais (\$1.000)
1	973	0	40
2	1.119	0	40
3	875	25	25
4	625	25	25
5	910	30	30
6	971	30	30
7	931	35	35
8	1.177	35	35

Cidade	Vendas (\$1.000)	Propaganda em Rádio (\$1.000)	Propaganda em Jornais (\$1.000)
9	882	40	25
10	982	40	25
11	1.628	45	45
12	1.577	45	45
13	1.044	50	0
14	914	50	0
15	1.329	55	25
16	1.330	55	25
17	1.405	60	30
18	1.436	60	30
19	1.521	65	35
20	1.741	65	35
21	1.866	70	40
22	1.717	70	40

- Expresse a equação de regressão múltipla.
- Interprete o significado das inclinações, b_1 e b_2 , neste problema.
- Interprete o significado do coeficiente de regressão, b_0 .
- Faça a previsão para a média aritmética das vendas para uma cidade na qual a propaganda em rádio custe \$20.000 e a propaganda em jornais custe \$20.000.
- Construa uma estimativa para o intervalo de confiança de 95% da média aritmética das vendas para cidades nas quais a propaganda em rádio custe \$20.000 e a propaganda em jornais custe \$20.000.
- Construa um intervalo de previsão de 95%, para as vendas de uma cidade individual na qual a propaganda em rádio custe \$20.000 e a propaganda em jornais custe \$20.000.

14.6 UTILIZANDO VARIÁVEIS BINÁRIAS (DUMMY) E TERMOS DE INTERAÇÃO EM MODELOS DE REGRESSÃO

Os modelos de regressão múltipla, discutidos nas Seções 14.1 a 14.5, partiram do pressuposto de que cada uma das variáveis independentes é numérica. No entanto, em algumas situações, pode ser desejável incluir variáveis categóricas como variáveis independentes no modelo de regressão. Por exemplo, na Seção 14.1, você utilizou preço e despesas com promoções para prever as vendas mensais de barras energéticas OmniPower. Além dessas variáveis independentes numéricas, pode ser desejável incluir o efeito do local de exposição das barras nas prateleiras da loja (por exemplo, exposição na ponta do corredor ou exposição fora da ponta do corredor), ao desenvolver um modelo para prever vendas de OmniPower.

A utilização das **variáveis binárias (dummy)** permite que você inclua variáveis independentes categóricas como parte do modelo de regressão. Se uma determinada variável independente categórica possui duas categorias, você precisa, então, de somente uma variável binária (dummy) para representar as duas categorias. Uma variável binária (dummy), X_d , é definida como

$$\begin{aligned} X_d &= 0, \text{ se a observação estiver na categoria 1} \\ X_d &= 1, \text{ se a observação estiver na categoria 2} \end{aligned}$$

Para ilustrar a aplicação de variáveis dummy (binárias) na regressão, considere um modelo para prever o valor de avaliação para uma amostra de 15 casas, com base no tamanho do imóvel (em milhares de pés quadrados) e se a casa possui ou não uma lareira. Para incluir a variável categórica que se refere à presença de uma lareira, a variável binária (dummy), X_2 , é definida como

$$\begin{aligned} X_2 &= 0, \text{ se a casa não possui uma lareira} \\ X_2 &= 1, \text{ se a casa possui uma lareira} \end{aligned}$$

A Tabela 14.5 apresenta os dados, que estão também armazenados no arquivo `casas.xls`. Na última coluna da Tabela 14.5, você pode verificar o modo como os dados categóricos são convertidos em valores numéricos.

$X_1 = \text{Tamanho da Casa}$

Casa	Y = Valor de Avaliação (\$1.000)	(Milhares de Pés Quadrados)	Lareira	$X_2 = \text{Lareira}$
1	84,4	2,00	Sim	1
2	77,4	1,71	Não	0
3	75,7	1,45	Não	0
4	85,9	1,76	Sim	1
5	79,1	1,93	Não	0
6	70,4	1,20	Sim	1
7	75,8	1,55	Sim	1
8	85,9	1,93	Sim	1
9	78,5	1,59	Sim	1
10	79,2	1,50	Sim	1
11	86,7	1,90	Sim	1
12	79,3	1,39	Sim	1
13	74,5	1,54	Não	0
14	83,8	1,89	Sim	1
15	76,8	1,59	Não	0

Pressupondo que a inclinação do valor de avaliação com o tamanho do imóvel seja a mesma, para as casas que possuam e para as que não possuam lareira, o modelo de regressão múltipla é

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

em que

- Y_i = valor de avaliação, em milhares de dólares, para a casa i
- β_0 = intercepto de Y
- X_{1i} = tamanho do imóvel, em milhares de pés quadrados, para a casa i
- β_1 = inclinação do valor de avaliação com o tamanho do imóvel, mantendo-se a presença ou a ausência de uma lareira
- X_{2i} = variável binária, representando a presença ou a ausência de uma lareira para a casa i
- β_2 = efeito incremental decorrente da presença de uma lareira em relação ao valor de avaliação, mantendo-se constante o tamanho do imóvel
- ϵ_i = erro aleatório em Y para a casa i

Encontre estimativas para β_0 , β_1 e β_2 .

14.41 O gerente de marketing de uma grande cadeia de supermercados gostaria de determinar o efeito do espaço disponível em prateleiras e se o produto posicionado na frente ou no fundo do corredor influi nas vendas de ração para animais de estimação. É selecionada uma amostra aleatória de 12 lojas com igual tamanho, com os seguintes resultados (contidos no arquivo `ração.xls`):

Loja	Espaço de Prateleira (Pés)	Localização	Vendas Semanais (Dólares)
1	5	Fundo	160
2	5	Frente	220
3	5	Fundo	140
4	10	Fundo	190
5	10	Fundo	240
6	10	Frente	260
7	15	Fundo	230
8	15	Fundo	270
9	15	Frente	280
10	20	Fundo	260
11	20	Fundo	290
12	20	Frente	310

- a. Expresse a equação de regressão múltipla.
- b. Interprete o significado das inclinações neste problema.
- c. Faça a previsão da média aritmética das vendas semanais de ração para animais de estimação para uma loja com 8 pés de espaço de prateleira posicionado no fundo do corredor. Construa uma estimativa para o intervalo de confiança de 95% e um intervalo de previsão de 95%.
- d. Realize uma análise de resíduos nos resultados e determine a adequação do modelo.
- e. Existe uma relação significativa entre vendas e as duas variáveis independentes (espaço de prateleira e posição no corredor), no nível de significância de 0,05?
- f. No nível de significância de 0,05, determine se cada uma das variáveis independentes oferece uma contribuição ao modelo de regressão. Indique o modelo de regressão mais apropriado para esse conjunto de dados.
- g. Construa estimativas de intervalos de confiança de 95% para a inclinação da população da relação entre vendas e espaço na prateleira e entre vendas e posição no corredor.

14.49 A diretora de um programa de treinamento de uma grande companhia de seguros está avaliando três diferentes métodos de treinamento de corretores. Os três métodos são: tradicional, baseado em CD-ROM e baseado na Internet. Ela divide 30 treinandos em três grupos de 10 designados aleatoriamente. Antes do início do treinamento, é aplicada a cada um dos treinandos uma prova de proficiência, que mede competências

em matemática e informática. Ao final do treinamento, todos os alunos fazem a mesma prova de encerramento do treinamento. Os resultados, armazenados no arquivo `corretores.xls`, são os seguintes:

Prova de Proficiência	Prova de Encerramento do Treinamento	Método
94	14	Tradicional
96	19	Tradicional
98	17	Tradicional
100	38	Tradicional
102	40	Tradicional
105	26	Tradicional
109	41	Tradicional
110	28	Tradicional
111	36	Tradicional
130	66	Tradicional
80	38	Baseado em CD-ROM
84	34	Baseado em CD-ROM
90	43	Baseado em CD-ROM
97	43	Baseado em CD-ROM
97	61	Baseado em CD-ROM
112	63	Baseado em CD-ROM
115	93	Baseado em CD-ROM
118	74	Baseado em CD-ROM
120	76	Baseado em CD-ROM
120	79	Baseado em CD-ROM
92	55	Baseado na Internet
96	53	Baseado na Internet
99	55	Baseado na Internet
101	52	Baseado na Internet
102	35	Baseado na Internet
104	46	Baseado na Internet
107	57	Baseado na Internet
110	55	Baseado na Internet
111	42	Baseado na Internet
118	81	Baseado na Internet

Desenvolva um modelo de regressão múltipla para prever o resultado da prova de encerramento do treinamento, com base no resultado da prova de proficiência e do método de treinamento utilizado.

- a. Expresse a equação de regressão múltipla.
- b. Interprete o significado das inclinações neste problema.
- c. Faça a previsão para a média aritmética do resultado da prova de final de treinamento para um aluno com um resultado de prova de proficiência igual a 100 cujo treinamento tenha sido baseado na Internet.
- d. Realize uma análise de resíduos nos resultados e determine a adequação do modelo.
- e. Existe uma relação significativa entre prova de final de treinamento e as variáveis independentes (prova de proficiência e método de treinamento), no nível de significância de 0,05?
- f. No nível de significância de 0,05, determine se cada uma das variáveis independentes oferece uma contribuição ao modelo de regressão. Indique o modelo de regressão mais apropriado para esse conjunto de dados.
- g. Construa uma estimativa para o intervalo de confiança de 95% da inclinação da população para a relação entre prova de final de treinamento e cada uma das variáveis independentes.
- h. Interprete o significado do coeficiente de determinação múltipla.
- i. Calcule o r^2 ajustado.

14.62 Desenvolva um modelo para prever o valor de avaliação (em milhares de dólares), utilizando o tamanho de imóveis (em milhares de pés quadrados) e a idade dessas casas (em anos), a partir da tabela a seguir (cujos dados estão armazenados no arquivo `casas.xls`):

Imóvel	Valor de Avaliação (\$1.000)	Tamanho do Imóvel (Milhares de Pés Quadrados)	Idade (Anos)
1	184,4	2,00	3,42
2	177,4	1,71	11,50
3	175,7	1,45	8,33
4	185,9	1,76	0,00
5	179,1	1,93	7,42
6	170,4	1,20	32,00
7	175,8	1,55	16,00
8	185,9	1,93	2,00
9	178,5	1,59	1,75
10	179,2	1,50	2,75
11	186,7	1,90	0,00
12	179,3	1,39	0,00
13	174,5	1,54	12,58
14	183,8	1,89	2,75
15	176,8	1,59	7,17

- Expresse a equação de regressão múltipla.
- Interprete o significado das inclinações nessa equação.
- Faça a previsão da média aritmética do valor de avaliação para uma casa que tenha como tamanho 1.750 pés quadrados e 10 anos de construção.
- Realize uma análise de resíduos nos resultados e determine a adequação do modelo.
- Determine se existe uma relação significativa entre o valor de avaliação e as duas variáveis independentes (tamanho e idade), no nível de significância de 0,05.
- Determine o valor- p em (e) e interprete o seu significado.
- Interprete o significado do coeficiente de determinação múltipla neste problema.
- Determine o r^2 ajustado.
- No nível de significância de 0,05, determine se cada uma das variáveis independentes oferece uma contribuição significativa ao modelo de regressão. Indique o modelo de regressão mais apropriado para esse conjunto de dados.

- j.** Determine os valores- p em (i) e interprete seus significados.
- k.** Construa uma estimativa para o intervalo de confiança de 95% para a inclinação da população entre valor de avaliação e tamanho. De que modo a interpretação da inclinação, nesse caso, difere daquela correspondente ao Problema 13.77, nos Problemas de Revisão do Capítulo 13?
- l.** Calcule e interprete os coeficientes de determinação parcial.
- m.** A assessoria da imobiliária declarou que a idade do imóvel não exerce nenhuma influência sobre o valor de avaliação. Com base em suas respostas para (a) até (l), você concorda com essa afirmativa? Explique.