

LISTA DE EXERCÍCIOS DE REGRESSÃO – PROF. HOWARD ROATTI

Exercício 01

É esperado que a massa muscular de uma pessoa diminua com a idade. Para estudar essa relação, uma nutricionista selecionou 18 mulheres, com idade entre 40 e 79 anos, e observou em cada uma delas a idade (X) e a massa muscular (Y).

Massa muscular (Y)	Idade (X)
82.0	71.0
91.0	64.0
100.0	43.0
68.0	67.0
87.0	56.0
73.0	73.0
78.0	68.0
80.0	56.0
65.0	76.0
84.0	65.0
116.0	45.0
76.0	58.0
97.0	45.0
100.0	53.0
105.0	49.0
77.0	78.0
73.0	73.0
78.0	68.0

- (a) Ajuste uma reta de regressão para a relação entre as variáveis Y: massa muscular (dependente) e X: idade (independente).
- (b) Considerando a reta estimada dada no item (a), estime a massa muscular média de mulheres com 50 anos.

Exercício 02

Os dados a seguir correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 25 famílias.

Renda Familiar (X)	Gasto com Alimentação (Y)
3	1,5
5	2,0
10	6,0

10	7,0
20	10,0
20	12,0
20	15,0
30	8,0
40	10,0
50	20,0
60	20,0
70	25,0
70	30,0
80	25,0
100	40,0
100	35,0
100	40,0
120	30,0
120	40,0
140	40,0
150	50,0
180	40,0
180	50,0
200	60,0
200	50,0

- Obtenha a equação de regressão do gasto com alimentação em função da renda familiar.
- Qual o significado prático do valor da inclinação da reta de regressão do item (a)?
- Considerando a reta estimada dada no item (a), estime o gasto de alimentação para as rendas: 55, 90, 250, 300 e 450.

Exercício 4

Uma pesquisa foi realizada com o objetivo de verificar se existe associação entre a falta de sono e a capacidade de as pessoas resolverem problemas simples. Foram testadas 10 pessoas, mantendo-se sem dormir por um determinado número de horas. Após cada um destes períodos, cada pessoa teve de resolver um teste com adições simples, anotando-se então os erros cometidos. Os dados resultantes são os seguintes:

Número de erros	6	8	6	10	8	14	12	14	12	16
Número de horas sem dormir	8	8	12	12	16	16	20	20	24	24

- Obtenha a equação de regressão do número de erros em função do número de horas sem dormir.
- Estime o número de erros para pessoas que ficaram sem dormir o seguinte número de horas: 2, 6, 18, 25, 30.

Exercício 5

O proprietário da *Showtime Movie Theater, Inc.*, gostaria de estimar o faturamento bruto semanal (Y) como função dos gastos com publicidade (TV e J). Dados históricos para uma amostra de seis semanas são apresentados a seguir.

Faturamento Bruto Semanal (Y) US \$ 1.000	Anúncio de Televisão (TV) US \$ 1.000	Anúncio de Jornal (J) US \$ 1.000
96	5.0	1.5
90	2.0	2.0
95	4.0	1.5
92	2.5	2.5
95	3.0	3.3
94	3.5	2.3
94	2.5	4.2
94	3.0	2.5

Foram obtidas três regressões:

(a) $Y = 88.64 + 1.60TV$ $R^2 = 0.65$

(b) $Y = 93.86 - 0.04J$ $R^2 = 0.0004$

(c) $Y = 83.23 + 2.29TV + 1.30J$ $R^2 = 0.92$

1 – Qual você escolheria e porque?

2 - Qual é a estimativa do faturamento bruto para uma semana quando US\$ 3.500 são gastos em anúncio de televisão e US\$1.800 são gastos em anúncio de jornal?

Exercício 6

Uma empresa que vende por correios componentes de computadores pessoais, software e hardware possui um depósito geral para a distribuição dos produtos. Atualmente, a administração se encontra examinando o processo de distribuição deste depósito e está interessado em estudar os fatores que afetam os custos de distribuição do depósito. Atualmente um pequeno cargo de manipulação se adiciona ao pedido, independentemente da quantidade pela que se fizeram. Foram coletados dados correspondentes de 24 meses e respeito aos custos de distribuição de depósito, as vendas e número de pedidos. A continuação apresenta os resultados:

Mês	Y	X_1	X_2	Mês	Y	X_1	X_2
1	52,95	386	4.015	13	62,98	372	3.977
2	71,66	446	3.806	14	72,30	328	4.428
3	85,56	512	5.309	15	58,99	408	3.964
4	63,69	401	4.262	16	79,38	491	4.582
5	72,81	457	4.296	17	94,44	527	5.582
6	68,44	458	4.097	18	59,74	444	3.450
7	52,46	301	3.213	19	90,50	623	5.079
8	70,77	484	4.809	20	93,24	596	5.735
9	82,03	517	5.237	21	69,33	463	4.269
10	74,39	503	4.732	22	53,71	389	3.708
11	70,84	535	4.413	23	98,18	547	5.387
12	54,08	553	2.921	24	66,80	415	4.161

Y: Custo de distribuição (em milhares de dólares); X_1 : Vendas (em milhares de dólares) e X_2 : Número de pedidos.

- (a) Ajuste os dados a um modelo de regressão com duas variáveis regressoras e interprete as estimativas dos parâmetros do modelo.
- (b) Estime o custo de distribuição do depósito mensal da empresa quando as vendas são \$ 400.000 dólares e o número de pedidos é de 4.500.
- (c) Descreva se o seu modelo é relativamente bom de acordo com as métricas.

Exercício 7

Baixe o dataset em [Behavior of the urban traffic of the city of Sao Paulo in Brazil Data Set](#). Crie um modelo de regressão linear múltipla para estimar a porcentagem de lentidão no trânsito. Lembre-se de descrever de acordo com as métricas se seu modelo é relativamente bom.

Exercício 8

Baixe o dataset em [Combined Cycle Power Plant Data Set](#). Crie um modelo de regressão linear múltipla para estimar saída de energia elétrica horária líquida. Lembre-se de descrever de acordo com as métricas se seu modelo é relativamente bom.

Exercício 9

Baixe o dataset em [Revisiting a Concrete Strength regression](#). Crie um modelo de regressão linear múltipla para estimar saída de energia elétrica horária líquida. Lembre-se de descrever de acordo com as métricas se seu modelo é relativamente bom.

Exercício 10

Retorne ao exercícios 7, 8 e 9. Implemente os métodos listados no exemplo a seguir:

```
# Fit a model
fit <- lm(Fertility ~ . , data = swiss)

# Function for Root Mean Squared Error
RMSE <- function(error) { sqrt(mean(error^2)) }
RMSE(fit$residuals)

# If you want, say, MAE, you can do the following:

# Function for Mean Absolute Error
mae <- function(error) { mean(abs(error)) }
mae(fit$residuals)
```

Pesquise a diferença da métrica MAE (Mean Absolute Error), descreva a diferença dela em relação ao RMSE. Para cada uma das atividades anteriores listadas, comente os resultados obtidos pelas métricas RMSE e MAE.