Universiadade Federal da Bahia Instituto de Matemática e Estatística Prof. Dr. Gilberto Pereira Sassi

Lista de exercícios – Regressão linear simples.

Em alguns casos desta lista de exercícios, você vai precisar alguma ferramenta computacional como o R, Python e afins.

1. Diabete e obesidade são condições sérias que acometem uma proporção grande da população mundial. Uma forma de medir a quantidade de gordura corporal é monitorar o peso, mas medir precisamente a quantidade de gordura corporal envolve o uso de equipamentos sofisticados de Raio-X. Em vez de medir a quantidade de gordura corporal, poderíamos usar o Índice de Massa Corporal (BMI) para aproximar a quantidade de gordura corporal. Em um estudo com 250 homens na Universidade Brigham Young, o BMI e gordura corporal foi mensurado para cada homem. Algumas informações do experimento estão na Tabela 1.

$$S_x = 6322, 28 \quad S_{x^2} = 162674, 18 \quad S_y = 4757, 90 \quad S_{y^2} = 107679, 27 \quad S_{x \cdot y} = 125471, 10$$

Tabela 1: Algumas informações do experimento.

- (a) Calcule as estimativas para o intercepto a e para a inclinação b.
- (b) Usando a equação do item (a), qual seria a gordura corporal de um homem com BMI = 30.
- (c) Estime σ^2 .
- (d) Calcule a variância de \hat{a} .
- (e) Calcule a variância de b.
- (f) Podemos rejeitar $H_0: b=0$ ao nível de significância $\alpha=5\%$? Calcule o valor-p.
- (g) Construa um intervalo de confiança para o intercepto com coeficiente de confiança $\gamma = 95\%$.
- (h) Construa um intervalo de confiança para a inclinação com coeficiente de confiança $\gamma = 99\%$.
- (i) Encontre o intervalo de confiança para a predição da gordura corporal quando BMI=25 com coeficiente de confiança $\gamma=95\%$.
- (i) Calcule R^2 e interprete.
- (k) Qual o incremento médio na gordura corporal ao aumentarmos em uma unidade o BMI?
- 2. Um estudo tem o objetivo de analisar a força compressiva (x) e a permeabilidade intrínseca (y) de várias misturas e curas de concreto. Algumas informações do experimento estão na Tabela 2.

$$n = 14$$
 $S_y = 572$ $S_{y^2} = 23.530$ $S_x = 43$ $S_{x^2} = 157, 42$ $S_{x \cdot y} = 1697, 80$

Tabela 2: Algumas informações do experimento.

- (a) Calcule as estimativas do intercepto e inclinação.
- (b) Estime σ^2 .
- (c) Usando a equação do item (a), qual seria a permeabilidade de uma mistura com força compressiva x=4,3.
- (d) Rejeitamos $H_0: b=0$ ao nível de significância $\alpha=5\%$? Calcule o valor-p.
- (e) Calcule $\operatorname{Var}(\hat{a})$ e $\operatorname{Var}(\hat{b})$ para este modelo.

- (f) Construa um intervalo de confiança para a e b com coeficiente de confiança $\gamma = 95\%$.
- (g) Qual a permeabilidade média, se x = 2, 5?
- (h) Encontre o intervalo para a predição de y_0 quando x=2,5. Interprete esse intervalo.
- (i) Calcule o R^2 e interprete.
- (j) Qual é o aumento na permeabilidade intrínseca (y) se aumentarmos em uma unidade a força compressiva?
- 3. Métodos de regressão foram usados pra analisar a relação entre temperatura da superfície da pista (x) e a deflexão do pavimento (y). Algumas informações deste estudo estão na Tabela 3.

$$n = 20$$
 $S_y = 12,75$ $S_{y^2} = 8,86$ $S_x = 1478$ $S_{x^2} = 143.215,8$ $S_{xy} = 1083,67$

Tabela 3: Alguns sumários do estudos.

- (a) Estime a e b.
- (b) Estime σ^2 .
- (c) Qual a deflexão média do pavimento se a temperatura da superfície da pista é $x = 85^{\circ}F$.
- (d) Qual a mudança na deflexão do pavimento se a temperatura da superfície da pista aumenta em $1^{\circ}F$?
- (e) Podemos rejeitar $H_0: b=0$ ao nível de significância $\alpha=5\%$? Calcule o valor-p.
- (f) Calcule $Var(\hat{a})$ e $Var(\hat{b})$.
- (g) Construa um intervalo de confiança para o intercepto a e inclinação b com coeficiente de confiança $\gamma = 99\%$.
- (h) Construa um intervalo de confiança para a predição da deflexão do pavimento se a temperatura de superfície da pista é $x=85^{\circ}F$. Use $\gamma=99\%$.
- 4. A Tabela 4 mostra dados sobre a classificação dos jogadores quaterback para a liga de futebol americano NFL em 2008. Acredita-se que a classificação (y) está associada ao número de jardas por tentativa (x).

Jogador	Time	Jardas por tentativa (x)	Pontos de classificação (y)
Philip Rivers	SD	8,39	105,50
Chad Pennington	MIA	7,67	97,40
Kurt Warner	ARI	7,66	96,90
Drew Brees	NO	7,98	96,20
Peyton Manning	IND	7,21	95,00
Aaron Rodgers	GB	7,53	93,80
Matt Schaub	HOU	8,01	92,70
Tony Romo	DAL	7,66	91,40
Jeff Garcia	TB	7,21	90,20
Matt Cassel	NE	7,16	89,40
Matt Ryan	ATL	7,93	87,70
Shaun Hill	SF	7,10	87,50
Seneca Wallace	SEA	6,33	87,00
Eli Manning	NYG	6,76	86,40
Donovan McNabb	PHI	6,86	86,40
Jay Cutler	DEN	7,35	86,00
Trent Edwards	BUF	7,22	85,40
Jake Delhomme	CAR	7,94	84,70
Jason Campbell	WAS	6,41	84,30
David Garrard	JAC	6,77	81,70
Brett Favre	NYJ	6,65	81,00
Joe Flacco	BAL	6,94	80,30
Kerry Collins	TEN	6,45	80,20
Ben Roethlis-berger	PIT	7,04	80,10
Kyle Orton	CHI	6,39	79,60
JaMarcus Russell	OAK	6,58	77,10
Tyler Thigpen	KC	6,21	76,00
Gus Freotte	MIN	7,17	73,70
Dan Orlovsky	DET	6,34	72,60
Marc Bulger	STL	6,18	71,40
Ryan Fitzpatrick	CIN	5,12	70,00
Derek Anderson	CLE	5,71	66,50

Tabela 4: Dados NFL.

- (a) Estime o intercepto e a inclinação em uma regressão linear simples.
- (b) Estime σ^2 .

- (c) Qual seria a classificação do jogador se o número de jardas por tentativa é x = 7, 5.
- (d) Qual o incremente médio na classificação do jogador para cada incremento em uma unidade no número de jardas por tentativa.
- (e) Rejeitamos $H_0: b=0$ ao nível de significância $\alpha=1\%$? Calcule o valor-p.
- (f) Estime $Var(\hat{a})$ e $Var(\hat{b})$.
- (g) Decida entre as hipóteses: $H_0: b=10$ e $H_1; b\neq 10$. Use $\alpha=5\%$. Calcule o valor-p.
- (h) Construa um intervalo de confiança para a e b com coeficiente de confiança $\gamma = 95\%$.
- (i) Qual a classificação média para número de jardas por tentativa é x = 8?
- (j) Construa um intervalo de confiança para a predição da classificação quando o número de jardas por tentativa é x=8?
- (k) Calcule o R^2 e interprete.
- (l) Construa o gráfico de probabilidade normal para os resíduos. Podemos assumir a normalidade na regressão linear simples?
- (m) Desenhe o gráfico de dispersão entre x e o resíduo. Podemos assumir a linearidade entre y e x?
- 5. A Tabela 5 apresenta os dados com o preço de venda e o IPTU para 24 casas.

Preço de venda	IPTU
25,9000	4,9176
29,5000	5,0208
27,9000	4,5429
25,9000	4,5573
29,9000	5,0597
29,9000	3,8910
30,9000	5,8980
28,9000	5,6039
35,9000	5,8282
31,5000	5,3003
31,0000	6,2712
30,9000	5,9592
30,0000	5,0500
36,9000	8,2464
41,9000	6,6969
40,5000	7,7841
43,9000	9,0384
37,5000	5,9894
37,9000	7,5422
44,5000	8,7951
37,9000	6,0831
38,9000	8,3607
36,9000	8,1400
45,8000	9,1416

Tabela 5: Preço e IPTU por metro quadrado de 24 casa em 1000 reais.

- (a) Assuma que a regressão linear simples é adequada para os dados da Tabela 5. Estime os coeficientes da regressão linear simples entre o preço de venda (y) e o iptu (x).
- (b) Qual o preço médio de uma casa se o IPTU é x = 7, 5.
- (c) Análise os resíduos para checar a qualidade do ajuste.
- (d) Rejeitamos $H_0: b=0$ ao nível de significância $\alpha=5\%$? Use o teste-t e calcule o valor-p.
- (e) Rejeitamos $H_0: b=0$ ao nível de significância $\alpha=5\%$? Use ANOVA e calcule o valor-p.
- (f) Estime $Var(\hat{a})$ e $Var(\hat{b})$.
- (g) Teste a hipótese $H_0: a=0$. Use $\alpha=1\%$. Calcule o valor-p.
- (h) Construa um intervalo de confiança para a e b com coeficiente de confiança $\gamma = 95\%$.
- (i) Construa um intervalo de confiança para a predição de y_0 se x=7,5 com coeficiente de confiança $\gamma=99\%$.
- (j) Calcule o R^2 e interprete.
- 6. A Tabela 6 apresenta dados de quilometragem e as cilindradas para veículos da DaimlerChrysler para o modelo de 2005.

Modelo	Cilindrada do motor cm ³	Consumo km/l
300C/SRT-8	3.523,22	13,09
CARAVAN 2WD	3.293,80	13,82
CROSSFIRE ROADSTER	3.211,86	15,05
DAKOTA PICKUP 2WD	3.703,48	11,95
DAKOTA PICKUP 4WD	3.703,48	10,37
DURANGO 2WD	5.702,70	10,25
GRAND CHEROKEE 2WD	3.703,48	12,12
GRAND CHEROKEE 4WD	5.702,70	10,29
LIBERTY/CHEROKEE 2WD	2.425,29	13,94
LIBERTY/CHEROKEE 4WD	3.703,48	11,90
NEON/SRT-4/SX 2.0	1.999,22	17,56
PACIFICA 2WD	3.523,22	12,75
PACIFICA AWD	3.523,22	11,99
PT CRUISER	2.425,29	14,50
RAM 1500 PICKUP 2WD	8.193,53	7,95
RAM 1500 PICKUP 4WD	5.702,70	8,63
SEBRING 4-DR	2.703,87	14,92
STRATUS 4-DR	2.425,29	16,11
TOWN & COUNTRY 2WD	2.425,29	14,37
VIPER CONVERTIBLE	8.193,53	11,01
WRANGLER/TJ $4WD$	2.425,29	11,22

Tabela 6: dados de quilometragem da gasolina

- (a) Ajuste um modelo de regressão linear simples relacionando a quilometragem de gasolina e a cilindrada dos veículos.
- (b) Encontre a estimativa média da quilometragem para um carro com $x = 2458,06cm^3$ cilindradas.
- (c) Teste a significância da regressão ao nível de significância $\alpha = 5\%$. Calcule o valor-p.
- (d) Estime o erro padrão do intercepto e inclinação.
- (e) Decida entre as hipóteses: $H_0: b \ge -0.05$ e $H_1: b < -0.05$. Use $\alpha = 1\%$. Calcule o valor-p.
- (f) Teste as hipóteses $H_0: a=0$ e $H_1: a\neq 0$ usando $\alpha=1\%$. Calcule o valor-p.
- (g) Construa um intervalo de confiança para o intercepto e inclinação com o coeficiente de confiança $\gamma = 95\%$.
- (h) Construa um intervalo de confiança para a predição de y_0 se $x=2458,06cm^3$ com coeficiente de confiança $\gamma=99\%$.
- (i) Qual a variabilidade total da quilometragem é explicada pelas cilindradas?
- (j) Análise os resíduos para checar as suposições da regressão linear simples.
- 7. Na Tabela 7, apresentamos dados sobre a produção de papel em que
 - y: concentração de licor verde (Na_2S) gramas por litro;
 - x: produção da máquina de papel toneladas por dia.

x 825	830	890	895	890	910	915	960	990	1.010	1.012	1.030	1.050
y 40	42	49	46	44	48	46	43	53	52	54	57	58

Tabela 7: Dados sobre uma máquina de produção de papel.

- (a) Ajuste um modelo de regressão linear simples para os dados na Tabela 7.
- (b) Estime σ^2 .
- (c) Construa um diagrama de dispersão para $x \in y$, e desenha a reta ajustada pelo item (a).
- (d) Estime a concentração média de licor verde se a produção da máquina da papel é x=950 toneladas por dia.
- (e) Teste a significância da regressão usando $\alpha = 5\%$. Calcule o valor-p.
- (f) Estime $Var(\hat{a})$ e $Var(\hat{b})$.
- (g) Decida entre as hipóteses $H_0: a=0$ e $H_1: a\neq 0$ usando $\alpha=5\%$. Calcule o valor-p.
- (h) Construa um intervalo de confiança para o intercepto e a inclinação com coeficiente de confiança $\gamma = 95\%$.

- (i) Construa um intervalo de predição para y_0 se x=950 com coeficiente de confiança $\gamma=99\%$.
- (i) Calcule R^2 e interprete.
- (k) Analise o resíduos para checar as suposições da regressão linear simples.
- 8. Um estudo tem o objetivo de analisar a relação entre exposição à ruído e hipertensão. Os dados estão na Tabela 8.

x 60	63	65	70	70	70	80	90	80	80	85	89	90	90	90	90	94	100	100	100
y 1	0	1	2	5	1	4	6	2	3	5	4	6	8	4	5	7	9	7	6

Tabela 8: Dados com exposição ao ruído e hipertensão

- (a) Construa um gráfico de dispersão entre y, aumento na pressão sanguínea em milímetros de mercúrio, e x, pressão sonora em decibéis. Este gráfico indica que podemos usar regressão linear simples?
- (b) Ajuste um modelo de regressão linear simples.
- (c) Estime σ^2 .
- (d) Qual o aumento na pressão média sanguínea associada com a pressão sonora x = 85 decibéis.
- (e) Estime $Var(\hat{a})$ e $Var(\hat{b})$.
- (f) Esta regressão é significativa usando $\alpha = 5\%$. Calcule o valor-p.
- (g) Teste $H_0: a = 0$ e $H_1: a \neq 0$ usando $\alpha = 5\%$. Calcule o valor-p.
- (h) Construa um intervalo de confiança para o intercepto e a inclinação com coeficiente de confiança $\gamma = 95\%$.
- (i) Encontra um intervalo de confiança para predição do aumento na pressão sanguínea se a exposição ao ruído é x=85 decibéis com coeficiente de confiança $\gamma=99\%$.
- (j) Qual a porcentagem da variância total do aumento da pressão sanguínea é devida pela exposição ao ruído?
- (k) Analise o ruído para checar a qualidade do ajuste.
- 9. Um estudo deseja analisar a relação entre o desgaste de uma peça de metal e a viscosidade do óleo usado na lubrificação. Os dados estão na Tabela 9.

x 1,6	9,4	15,5	20,0	22,0	35,5	43,0	40,5	33,0
у 240,0	181,0	193,0	155,0	172,0	110,0	113,0	75,0	94,0

Tabela 9: Dados com y: desgaste da peça de metal com (volume em 10^{-4} milimetros cúbicos), e x: viscosidade do óleo.

- (a) Construa o diagrama de dispersão para as variáveis x e y. É razoável usar regressão linear simples para este conjunto de dados é plausível?
- (b) Estime o intercepto e a inclinação para o modelo de regressão linear simples.
- (c) Estime σ^2 .
- (d) Qual seria o desgaste médio para um óleo com viscosidade x = 30.
- (e) Esta regressão linear é significativa? Use $\alpha = 5\%$. Calcule o valor-p.
- (f) Estime $Var(\hat{a})$ e Var(b).
- (g) Decida entre as hipóteses $H_0: a=0$ e $H_1: a\neq 0$. Use $\alpha=5\%$. Calcule o valor-p.
- (h) Decida entre as hipóteses $H_0: a \le 2500$ e $H_1: a > 2500$. Use $\alpha = 5\%$. Calcule o valor-p.

- (i) Construa um intervalo de confiança para o intercepto e inclinação com coeficiente de confiança $\gamma = 99\%$.
- (j) Construa um intervalo de confiança para a predição do desgaste médio quando a viscosidade é x=30 com coeficiente de confiança $\gamma=95\%$.
- (k) Calcule R^2 e interprete o resultado.
- (l) Analise os resíduos para verificar a qualidade do ajuste da regressão linear simples.
- 10. Um motor de foguete é produzido unindo dois tipos de propelentes: um de ignição e um sustentador. Os engenheiros suspeitam que a resistência ao cisalhamento da liga (y) está associada a idade em semanas da liga do propelente quando o motor é fundido. Os dados do estudo estão na Tabela 10.

Resistência ao cisalhamento (y)	Idade em semanas (x)
2.158,70	15,50
1.678,15	23,75
2.316,00	8,00
2.061,30	17,00
2.207,50	5,00
1.708,30	19,00
1.784,70	24,00
2.575,00	2,50
2.357,90	7,50
2.277,70	11,00
2.165,20	13,00
2.399,55	3,75
1.779,80	25,00
2.336,75	9,75
1.765,30	22,00
2.053,50	18,00
2.414,40	6,00
2.200,50	12,50
2.654,20	2,00
1.753,70	21,50

Tabela 10: Resistência ao cisalhamento (psi) e Idade do propelente quando o motor é fundido.

- (a) Construa o diagrama de dispersão para os dados da Tabela 10. O modelo de regressão linear simples é adequado para esse conjunto de dados?
- (b) Encontre as estimativas do intercepto e da inclinação da regressão linear simples.
- (c) Estime σ^2 .
- (d) Estime a resistência média ao cisalhamento do motor construído com um propelente com 20 semanas.
- (e) Construa um intervalo de confiança para o intercepto e a inclinação com coeficiente de confiança $\gamma = 95\%$.
- (f) Construa um intervalo de confiança para a predição da resistência média ao cisalhamento se o motor foi fundido com um propelente com 20 semanas com coeficiente de confiança $\gamma = 95\%$.
- (g) Calcule o R^2 e interprete o resultado.
- (h) Construa o gráfico de probabilidade normal para os resíduos. Algum ponto está afastado da reta y=x?
- (i) Delete os pontos identificados no item (h), e atualize as estimativas com o intercepto e a inclinação. Atualize R^2 e compare com o valor obtido no item (g). Atualize a estimativa para σ^2 , e compare com a estimativa do item (c).
- 11. Um estudo deseja analisar a microestrutura para pó ultrafino de zircônia parcialmente estabilizada como uma função de temperatura. Os dados estão na Tabela 11.
 - (a) Ajuste um modelo de regressão linear simples.
 - (b) Estime σ^2 .
 - (c) Estime a porosidade média para a temperatura $1400^{\circ}C$.
 - (d) Desenhe o diagrama de dispersão para $x \in y$, e desenhe no mesmo gráfico a reta obtida no item (a).
 - (e) Construa um intervalo de confiança para o intercepto e inclinação com coeficiente de confiança $\gamma=95\%.$

Temperatura ${}^{\circ}C(x)$	Porosidade (%) (y)
1.100,00	30,80
$1.200,\!00$	19,20
1.300,00	6,00
1.100,00	13,50
1.500,00	11,40
1.200,00	7,70
1.300,00	3,60

Tabela 11: Dados para microestrutura para pó ultrafino de zircônia parcialmente estabilizada.

- (f) Estime a porosidade média para a temperatura $1500^{\circ}C$.
- (g) Construa um intervalo de confiança para a predição y_0 quando a temperatura 1500° C com coeficiente de confiança $\gamma = 99\%$.
- 12. Um pesquisador analisou a idade (x) e o comprimento ou tamanho (y) de 27 dugongos ("peixes-bois marinhos"). Os dados estão na Tabela 12.

X	у
1,00	1,80
1,50	1,85
1,50	1,87
1,50	1,77
2,50	2,02
4,00	2,27
5,00	2,15
5,00	2,26
7,00	2,47
8,00	2,19
8,50	2,26
9,00	2,40
9,50	2,39
9,50	2,41
10,00	2,50
12,00	2,32
12,00	2,32
13,00	2,43
13,00	2,47
14,50	2,56
15,50	2,65
15,50	2,47
16,50	2,64
17,00	2,56
22,50	2,70
29,00	2,72
31,50	2,57

Tabela 12: Peso e tamanho de dugongos.

- (a) Encontre as estimativas do intercepto e da inclinação para os dados da Tabela 12.
- (b) Estime σ^2 .
- (c) Análise os resíduos para checar a qualidade do ajuste.
- (d) Construa um intervalo de confiança para o intercepto e a inclinação com coeficiente de confiança $\gamma = 95\%$.
- (e) Qual o tamanho ou comprimento médio para um dugongo com 11 anos de idade.
- (f) Construa um intervalo de confiança para a predição de y_0 com coeficiente de confiança $\gamma = 99\%$.
- (g) Calcule o R^2 e interprete.
- (h) Qual o incremento médio no tamanho (comprimento) ao aumentarmos em um ano o tempo de vida de um dugongo.
- 13. Um pesquisador deseja analisar a relação entre duas variáveis x e y. Algumas informações deste estudo estão na Tabela 13. Complete as informações da Tabela 13.
 - (a) Essa regressão é significativa? Use $\alpha = 5\%$.
 - (b) Temos evidência estatística para rejeitar $H_0: a=0$? Use $\alpha=5\%$.

Equação da reta: $y = 12, 9 + 2, 34x$ e $R^2 = 98, 1\%$.									
	Coeficiente	Desvio padrão do coeficiente	T_0	Valor-p					
Intercepto Inclinação	12,857 2,3445	1,032 0,1150							
		Análise de variância							
Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrados médios	$ F_0 $	Valor-p				
Regressão Resíduos	1 8	912, 43 17, 55	912,43	_	_				
Total	9	929, 98	_	-	-				

Tabela 13: Algumas informações do experimento.

- (c) Qual estimativa para σ^2 .
- (d) Construa um intervalo de confiança para o intercepto e a inclinação com coeficiente de confiança $\gamma = 95\%$.
- 14. Um pesquisador deseja analisar a relação entre duas variáveis x e y. Algumas informações deste estudo estão na Tabela 14. Complete as informações da Tabela 14.

Equação da reta: 1	$y = 26, 8 + 1, 48x e R^2$	$^{2} = 93.7\%$.			
	Coeficiente	Desvio padrão do coeficiente	T_0	Valor-p	
Intercepto Inclinação	26,753 1,4756	2,373 0,1063			
		Análise de variância			
Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrados médios	$ F_0 $	Valor-p
Regressão Resíduos	1	94,8		_	_
Total	15	1500,0	_	-	-

Tabela 14: Algumas informações do experimento.

- (a) Essa regressão é significativa? Use $\alpha=5\%.$
- (b) Temos evidência estatística para rejeitar $H_0: a=0$? Use $\alpha=5\%$.
- (c) Qual estimativa para σ^2 .
- (d) Construa um intervalo de confiança para o intercepto e a inclinação com coeficiente de confiança $\gamma=99\%$.