Teoria da Regressão

$3^{\underline{a}}$ *Lista de Exercícios*

1. Considere dois conjuntos de dados

$$(u_1, x_1), \ldots, (u_n, x_n) \quad \text{e} \quad (v_1, z_1), \ldots, (v_n, z_n)$$

e os dois modelos

$$u_i = \alpha_1 + \beta_1 x_i + \eta_i \quad \text{e} \quad v_i = \alpha_2 + \beta_2 z_i + \delta_i$$

para $i = 1, \ldots, n$ em que $x_i's$ e $z_i's$ são variáveis explicativas e todos $\eta_i's$ e $\delta_i's$ são independentemente distribuidas com médias zeros e variância comun σ^2. Escreva essas duas equações como um único modelo de regressão linear $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ caracterizando $\mathbf{Y}$,$\mathbf{X}$, $\boldsymbol{\beta}$ e $\boldsymbol{\epsilon}$.

2. Suponha que nós precisamos comparar dois efeitos de duas drogas cada uma administrada a n sujeitos. O modelo para o efeito da primeira droga é

$$y_{1i} = \beta_0 + \beta_1 x_{1i} + \epsilon_{1i}$$

enquanto para a segunda droga é dado por

$$y_{2i} = \beta_0 + \beta_2 x_{2i} + \epsilon_{2i}$$

para $i = 1, \ldots, n$. Assumimos que todas as observações são independentes e para cada i ambos ϵ_{1i} e ϵ_{2i} são normalmente distribuidos com média zero e variância σ^2.

(a) Obtenha os estimadores de mínimos quadrados para $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ e sua matriz de covariância.

(b) Estime σ^2

(c) Escreva a estatística de teste para testar a hipótese $\beta_1 = \beta_2$ versus a alternativa que $\beta_1 \neq \beta_2$

3. Considere que 2 modelos $\mathbf{Y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}_1$ e $\mathbf{Y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2$ em que os $\mathbf{X}_i's$ são matrizes $n_i \times p$. Suponha que $\boldsymbol{\epsilon}_i \sim N(0, \sigma_i^2\mathbf{I})$ em que $i = 1, 2$ e que $\boldsymbol{\epsilon}_1$ e $\boldsymbol{\epsilon}_2$ são independentes.

(a) Assumindo que $\sigma_i's$ são conhecidos, obtenha um teste para hipóteses $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

(b) Assumindo que $\sigma_1 = \sigma_2$ mas desconhecidos, derive um teste para hipóteses $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

4. Seja

$$\begin{aligned} Y_1 &= \theta_1 + \theta_2 + \epsilon_1 \\ Y_2 &= 2\theta_2 + \epsilon_2 \\ Y_3 &= -\theta_1 + \theta_2 + \epsilon_3 \end{aligned}$$

em que $\epsilon_i (i = 1, 2, 3)$ são independentes $N(0, \sigma^2)$. Derive a estatística F para testar a hipótese $H_0 : \theta_1 = 2\theta_2$.

5. Suponha $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ em que $E(\boldsymbol{\epsilon}) = 0$, $Cov(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}_n$, a matriz $\mathbf{X}$ de dimensão $n \times k$ tem posto $k \leq n$, e $\boldsymbol{\beta}$ é um k-vetor de parâmetros de regressão e $\mathbf{Y} = (y_1, \ldots, y_n)^t$. Suponha que nós estamos interessado em predizer a $(n+1)-$ésima observação y_{n+1} em $x_{n+1} = (x_{n+1,1}, \ldots, x_{n+1,k})^T$; i.e, $y_{n+1} = x'_{n+1}\boldsymbol{\beta} + \epsilon_{n+1}$ tem a mesma distribuição que os outros $\epsilon'_i s$ e é independentes deles. O preditor baseado nos estimadores de mínimos quadrados de β é dado por $\hat{y}_{n+1} = x'_{n+1}\hat{\boldsymbol{\beta}}$, onde $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

(a) Mostre que $\hat{y}_{n+1}$ é uma função linear dos $(y_1, \ldots, y_n)$ tal que $E(\hat{y}_{n+1} - y_{n+1}) = 0$.

(b) Suponha que $\tilde{y}_{n+1} = \mathbf{a}'\mathbf{Y}$ é um outro preditor de y_{n+1} tal que $E(\tilde{y}_{n+1} - y_{n+1}) = 0$. Mostre que $\mathbf{a}$ deve satisfazer $\mathbf{a}'\mathbf{X} = x'_{n+1}$.

(c) Encontre $var(\hat{y}_{n+1})$ e $var(\tilde{y}_{n+1})$.

(d) Mostre que $var(\hat{y}_{n+1}) \leq var(\tilde{y}_{n+1})$

6. Em 48 estados americanos foram registradas as seguintes variáveis taxa(taxa de combustível no estado), licença(proporção de motorista licenciados), renda (renda per-capita), estradas (ajuda federal para as estradas) e consumo (consumo de combustível por habitante). O interesse é explicar o consumo de cmbustível pelas variáveis taxa, licença, renda e estradas.

(a) Construa os gráficos de dispersão convinientes.

(b) Ajuste um modelo normal linear para explicar o consumo contra as demais variáveis selecionando as variáveis explicativas que contribuem significamente para o modelo. Interprete os diferentes parâmetros do modelo.

(c) Estime os parâmetros do modelo e apresente os respectivos erros padrões.

(d) Avalie a qualidade do ajuste do modelo.

(e) Construa o intervalo de confiança para consumo médio e vitórias segundo um conjunto de valores das explicativas a sua escolha.

(f) Apresente um conclusão que evite o jargão estatístico.

ME	9.00	0.525	3571	1976	541
VT	9.00	0.580	3865	1586	561
RI	8.00	0.544	4399	431	410
NY	8.00	0.451	5319	11868	344
PA	8.00	0.529	4447	8577	464
IN	8.00	0.530	4391	5939	580
MI	7.00	0.574	4817	6930	525
MN	7.00	0.608	4332	8159	566
MO	7.00	0.572	4206	8508	603
SD	7.00	0.724	4716	5915	865
KS	7.00	0.663	4593	7834	649
MD	9.00	0.511	4897	2449	464
WV	8.50	0.551	4574	2619	460
SC	8.00	0.548	3448	5399	577
FL	8.00	0.563	4188	5975	574
TN	7.00	0.518	3640	6905	571
MS	8.00	0.578	3063	6524	577
LA	8.00	0.487	3528	3495	487
TX	5.00	0.566	4045	17782	640
ID	8.50	0.663	3635	3274	648
CO	7.00	0.626	4449	4639	587
AZ	7.00	0.603	4300	3635	632
NV	6.00	0.672	5215	2302	782
OR	7.00	0.623	4296	4083	610
NH	9.00	0.572	4092	1250	524
MA	7.50	0.529	4870	2351	414
CT	10.00	0.571	5342	1333	457
NJ	8.00	0.553	5126	2138	467
OH	7.00	0.552	4512	8507	498
IL	7.50	0.525	5126	14186	471
WI	7.00	0.545	4207	6580	508
IA	7.00	0.586	4318	10340	635
ND	7.00	0.540	3718	4725	714
NE	8.50	0.677	4341	6010	640
DE	8.00	0.602	4983	602	540
VA	9.00	0.517	4258	4686	547
NC	9.00	0.544	3721	4746	566
GA	7.50	0.579	3846	9061	631
KY	9.00	0.493	3601	4650	534
AL	7.00	0.513	3333	6594	554
AR	7.50	0.547	3357	4121	628

OK	6.58	0.629	3802	7834	644
MT	7.00	0.586	3897	6385	704
WY	7.00	0.672	4345	3905	968
MN	7.00	0.563	3656	3985	699
UT	7.00	0.508	3745	2611	591
WA	9.00	0.571	4476	3942	510
CA	7.00	0.593	5002	9794	524

Estes dados podem ser encontrados em http://www.ime.usp.br/∼giapaula/reg2.dat.