**Universidade Federal do Ceará**

**Departamento de Estatística e Matemática Aplicada**

**Prof. : Juvêncio S. Nobre**

**CC0290 - Modelos de Regressão I - 2022.2**

**Lista de exercícios # 1.5**

**Distribuição: 19/09/2022**

**1.** Considere o seguinte modelo:

$$y_i = \beta_0 + \beta_1 x_i^2 + e_i, \;\; i = 1, \ldots, n.$$

i) Podemos ajustar esse modelo por meio da teoria de MRLS? Quais suposições são necessárias?

ii) Interprete os parâmetros do modelo.

iii) Obtenha os EMQ de forma tradicional e especifique sua distribuição exata. Faça suposições apropriadas.

iv) Reescreva o modelo acima na forma matricial.

v) Usando o que já foi visto em sala, desenvolva IC's exatos para $\beta_0$, $\beta_1$, $\beta_0 + \beta_1 x_i^2$ e um intervalo de predição para $Y_x$ ao nível de confiança $(1-\alpha)\%$.

**Sugestão:** Construa IC para um $\mathbf{c}^\top \boldsymbol{\beta}$.

vi) Determine um estimador não viesado para a variância da fonte de variação.

vii) Faça as suposições necessárias para determinar as distribuições da SQT, SQReg e SQRes.

viii) Adicionalmente, encontre a distribuição da estatística F comumente utilizada para testar $\mathcal{H}_0 : \beta_1 = 0$.

**2.** Considere uma amostra aleatória com 6 observações e que o interesse é modelar o custo (em R$) em função da quantidade de peças produzidas. Admite-se que as variáveis $x$ e $y$ estão relacionadas de acordo com o MRLS: $y_i = \beta_0 + \beta_1 x_i + e_i$, e que as pressuposições são válidas (inclusive de normalidade). Usando o software `R` obtemos os seguintes resultados:

```
> y=c(2042,2301,2421,2518,2606,2718)
> x=c(105,130,141,159,160,172)
```

```
> summary(lm(y~x))
Call:
lm(formula = y ~ x)

Residuals:
      1       2       3       4       5       6
 -7.151   8.063  20.797 -57.730  20.519  15.501

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1025.2469    90.1187   11.38 0.000340 ***
x              9.7515     0.6163   15.82 9.33e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.7 on 4 degrees of freedom
Multiple R-Squared: 0.9843,     Adjusted R-squared: 0.9803
F-statistic: 250.3 on 1 and 4 DF,  p-value: 9.326e-05

> anova(lm(y~x))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1 284275  284275  250.31 9.326e-05 ***
Residuals  4   4543    1136
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Determine:

i) A reta de regressão estimada pelo método de mínimos quadrados e interprete as respectivas estimativas.

ii) O coeficiente de determinação e verifique se é estatisticamente diferente de zero, através do teste F, ao nível de 5%.

iii) O resultado do teste de hipóteses: $\mathcal{H}_0 : \beta_1 = 0$ vs. $\mathcal{H}_1 : \beta_1 \neq 0$, ao nível de 5%.

iv) Se com base no valor do coeficiente de determinação é possível **afirmar** que o modelo está bem ajustado? Justifique.

v) Admita que as peças são produzidas por encomendas, i.e., são todas vendidas. Determine o número mínimo de peças produzidas para o qual o lucro é positivo (encomendas com número menor de peças não serão aceitas), sabendo que cada peça é vendida por R$ 18,00.

**3.** Considere um MRLS. Faça as suposições adequadas e determine um intervalo de confiança **exato** para $\mathbb{E}[y_i|x_i] - \mathbb{E}[y_j|x_j]$, com $x_i \neq x_j$.

**4.** Considere o modelo

$$y_i = \beta x_i^3 + e_i, \ \ \mathrm{i} = 1, \ldots, n,$$

em que os $e_1, \ldots, e_n$ representam uma sequência de variáveis aleatórias independentes normais com média zero e variância $\sigma^2$ e $x_i$ representam constantes reais, $i = 1, \ldots, n$.

i) Determine um vetor de estatísticas suficientes (não trivial) para o modelo.

ii) Obtenha os estimadores de máxima verossimilhança dos parâmetros do modelo acima. Mostre que os parâmetros são ortogonais.

iii) Mostre que o EMV de $\beta$ é uma combinação linear das observações e encontre sua distribuição amostral.

iv) Utilizando o método da quantidade pivotal, obtenha o intervalo de comprimento mínimo para $\beta$ ao nível de $(1-\alpha)\%$ de confiança.

**5.** Seja $x$ a quantidade de certo produto, em milhares de unidades, e $y$ o respectivo custo total de produção em milhares de dólares (1000 USD). Admite-se que o custo marginal seja constante, i.e., o MRLS pode ser utilizado para modelar a relação entre as variáveis. É dada a seguinte amostra de 10 pares de valores:

Pode-se verificar que $\sum_{i=1}^{10} x_i = 55$, $\sum_{i=1}^{10} y_i = 205$, $\sum_{i=1}^{10} x_i^2 = 385$, $\sum_{i=1}^{10} y_i^2 = 4965$ e $\sum_{i=1}^{10} x_i y_i = 1375$.

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 7 | 11 | 15 | 14 | 18 | 21 | 23 | 30 | 32 | 34 |

i) Estime a função de custo total e interprete as estimativas obtidas.

ii) Teste, ao nível de significância de 1%, a hipótese de que o custo marginal é nulo.

iii) Determine o intervalo de 95% de confiança para o valor do custo fixo.

iv) Calcule o coeficiente de determinação da regressão. Interprete o resultado.

v) Se o produtor vende em regime de competição perfeita ao preço de USD 3,50 por unidade, quantas unidades ele deve produzir para que sua renda líquida seja de USD 2.000,00?

vi) Determine a estimativa do custo médio total para $x = 10$ e o respectivo intervalo de confiança, ao nível de confiança de 95%. Interprete o resultado.

**6.** Considere o seguinte modelo

$$y_i = \beta_0 x_i^{\beta_1} e_i, \quad i = 1, ..., n.$$

Faça suposições convenientes a respeito da fonte de variação e obtenha o modelo de regressão ajustado, por intermédio da metodologia do MRLS, com base na seguinte amostra:

| $x$ | 1 | 1 | 100 | 100 | 10.000 | 10.000 |
|---|---|---|---|---|---|---|
| $y$ | 1 | 10 | 1.000 | 1.000 | 100 | 1.000 |

**7.** Imagine que o gerente de uma lanchonete tem o seguinte registro diário com as temperaturas máximas e também com o número de pedidos de chá gelado:

Tabela 1: Registros diários de temperatura máxima ($^\circ$) e do número de pedidos de chás gelados de uma determinada lanchonete.

| Dia | Temperatura máxima | Número de pedidos |
|---|---:|---:|
| 1 | 29 | 77 |
| 2 | 28 | 62 |
| 3 | 34 | 93 |
| 4 | 31 | 84 |
| 5 | 29 | 64 |
| $\vdots$ | $\vdots$ | $\vdots$ |

O objetivo é utilizar um modelo para determinar o estoque que a loja deve ter em função da previsão de temperatura máxima. Discuta a respeito da utilização de um MRLS baseado na suposição de normalidade para o fenômeno acima. Caso julgue a suposição de normalidade restritiva demais ou irrealista para modelar o fenômeno de interesse, qual modelo probabilístico poderia ser uma alternativa ao invés da Normal? Seria necessário algum tipo de transformação para fazer uso de método de mínimos quadrados? Caso afirmativo, especifique e justifique.