ME720 - Modelos Lineares Generalizados

Profa. Larissa Avila Matos

1. Um conjunto de dados que trata da corrosão de superfície metálicas foi coletado em 32 superfícies amostrais. Foi considerado um modelo de regressão que relaciona a quantidade de corrosão (y) em função do tempo (x_1), tensão (x_2), pH no momento da corrosão (x_3) e temperatura (x_4). O ajuste do modelo de regressão foi feito no software R, e as saídas numéricas estão listadas abaixo.

```
> fit.model<-lm(y~x1+x2+x3+x4)
> summary(fit.model)
Call:
lm(formula = y ~ x1 + x2 + x3 + x4)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -96.3742    32.2164  -2.991   ????
x1            0.0624     0.0145   ?????   ????
x2            0.0016     0.0021   0.774   ????
x3           -0.2672     0.1757  -1.521   ????
x4            0.3301     ??????   ?????   ????

> anova(fit.model)
Analysis of Variance Table
Terms added sequentially (first to last)
Response: y
          Df  Sum Sq Mean Sq F value Pr(>F)
x1         1  67.850  67.850   ?????  ????
x2         1   1.797   1.797   ?????  ????
x3         1   5.921   5.921   ?????  ????
x4         1  51.438  51.438   ?????  ????
Residuals ?? 140.237   ?????
```

(a) Escreva a forma matricial do modelo de regressão linear múltipla para os dados e indique as dimensões de cada termo.

(b) Indique as principais suposições para os erros na análise de regressão.

(c) Calcule $SQ(\beta_3, \beta_4|\beta_0, \beta_1, \beta_2)$. Qual o significado deste resulatdo?

(d) Construa um intervalo de 95% de confiança para β_1, e use-o para testar a importância do tempo (x_1). Qual é a estimativa de σ^2?

(e) Construa a tabela ANOVA e então use a tabela para construir um teste F para testar o modelo de regressão ($H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$) ao nível de $\alpha = 0.05$.

2. Suponha que o seguinte modelo de regressão linear seja postulado

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon},$$

onde $\mathbf{Y}$ é o vetor $n \times 1$ de variáveis aleatórias, $\mathbf{X}_1$ é uma matrix $n \times p$ conhecida de posto completo, $\boldsymbol{\beta}_1$ é o vetor $p \times 1$ de parâmetros desconhecidos, e $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$ com σ^2 desconecido. Seja $\mathbf{b}_1$ o estimador de mínimos quadrados de $\boldsymbol{\beta}_1$ sob o modelo postulado, e seja $\hat{\mathbf{Y}} = \mathbf{X}_1\mathbf{b}_1$.

Suponha que o verdadeiro modelo é, na verdade,

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2,$$

onde $\mathbf{X}_2$ é uma matrix $n \times q$ de constantes conhecidas, $\boldsymbol{\beta}_2 \neq \mathbf{0}$ é um vetor $q \times 1$ de parâmetros desconhecidos, e $\boldsymbol{\epsilon}_2$ tem a mesma distribuição de $\boldsymbol{\epsilon}$.

(a) Mostre que $\mathbf{b}_1$ é, em geral, um estimador viesado de $\boldsymbol{\beta}_1$. Sob qual condição $\mathbf{b}_1$ é não viesado?

(b) Encontre a matriz de covariância de $\mathbf{b}_1$.

(c) Considere a decomposição de soma de quadrados total de $\mathbf{Y}$ em soma de quadrados do modelo (SQM) e soma de quadrados dos resíduos (SQR) para o modelo postulado. Encontre o valor esperado de SQR. Simplifique a expressão o máximo possível. Sabemos que o estimador usual de σ^2 é dado por $QMR = SQR/(n-p)$. O estimador QMR é não viesado?

(d) Para testar a hipótese $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$, suponha que seja usado o teste F baseado na estatística $F = \frac{SQM/p-1}{SQR/(n-p)}$, o qual assume incorretamente que o modelo postulado é verdadeiro. Quais são as verdadeiras distribuições de SQM e SQR sob H_0? Comente a respeito da validade do teste F.

(e) Seja $\mathbf{b}_1^*$ o estimador de mínimos quadrados de $\boldsymbol{\beta}_1$ sob o modelo verdadeiro assumindo que $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ tem posto completo. Quando $\boldsymbol{\beta}_2 = \mathbf{0}$, compare $\mathbf{b}_1$ e $\mathbf{b}_1^*$ em termos de viés e variância.