$1a$. Prova - Estat para CC

Renato Assunção - DCC, UFMG

Março de 2014

1. Considere um modelo de regressão linear em que a matriz de desenho $\mathbf{X}$ é de dimensão $n \times 1$ com apenas uma única coluna. Esta coluna é a coluna de 1's. Isto é,

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

   - Explique geometricamente o que é o espaço $\mathcal{M}(\mathbf{X})$ neste modelo particular.
   - Obtenha a matriz $\mathbf{H}$ de projeção ortogonal de $\mathbf{Y}$ no espaço $\mathcal{M}(\mathbf{X})$.
   - Obtenha o vetor projetado $\hat{\mathbf{Y}} = \mathbf{HY}$.
   - Qual a expressão de de $\hat{\boldsymbol{\beta}}$?
   - O que é o vetor de resíduos $\mathbf{r}$ neste caso?

2. Considere o modelo de regressão usual com matriz de desenho $\mathbf{X}$ $n \times p$ cuja primeira coluna é o vetor $\mathbf{1}$. Mostre que a soma dos resíduos $\sum_i r_i$ é igual a zero.

3. Mostre que o vetor de resíduos $\mathbf{r}$ é ortogonal ao vetor ajustado $\hat{\mathbf{Y}}$ e conclua que eles são vetores aleatórios independentes. Assuma o modelo de regressão usual com matriz de desenho $\mathbf{X}$ $n \times p$ cuja primeira coluna é o vetor $\mathbf{1}$.

4. Seja $X_1, \ldots X_n$ v.a.'s i.i.d. $N(0,1)$. Defina $Y_1 = X_1$ e $Y_i = X_i - X_{i-1}$ para $i = 2, \ldots, n$. Encontre a distribuição do vetor $\mathbf{Y} = (Y_1, \ldots, Y_n)'$.

5. Se $X_1, \ldots, X_n$ são v.a.'s tais que $\mathbb{V}(X_1) = \sigma^2$ e satisfazendo $X_{i+1} = \rho X_i$ onde $\rho \in (0,1)$ é uma constante e $i = 2, \ldots, n$. Encontre a matriz de covariância $\mathbb{V}(\mathbf{X})$ do vetor $\mathbf{X}$.

6. No modelo abaixo,
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
onde a matriz de desenho $\mathbf{X}$ tem dimensão $n \times p$, o vetor $\boldsymbol{\varepsilon}$ tem distribuição normal multivariada com vetor esperado $\boldsymbol{\mu} = (0, \ldots, 0)'$ e matriz de covariância igual a matriz diagonal com elementos
$$\text{diag}(\mathbb{V}(\boldsymbol{\varepsilon}) = \sigma^2(c_1, c_2, \ldots, c_n)$$
onde $c_i > 0$ são constantes CONHECIDAS. Assim, os erros não possuem variância constante.

   Ignorando a matriz de covariância diferente da usual, aplica-se a fórmula matricial para obter o estimador de mínimos quadrados de $\boldsymbol{\beta}$. Mostre que este estimador é não viciado para estimar $\boldsymbol{\beta}$. EXTRA BONUS: Obtenha a matriz de covariância do estimador $\hat{\boldsymbol{\beta}}$.

7. Suponha que $\mathbf{Y} = (Y_1, Y_2)$ seja uma vetor aleatório gaussiano com valor esperado $\boldsymbol{\mu} = (10, 15)$ e matriz de covariância
$$\mathbb{V}(\mathbf{Y}) = \begin{bmatrix} 4 & 1.6 \\ 1.6 & 1 \end{bmatrix}$$
Dois pontos que fazem da amostra são $\mathbf{y}_1 = (12, 15)$ e $\mathbf{y}_2 = (10, 17)$, ambos a uma distância euclidiana de 2 unidades de $\boldsymbol{\mu}$. Qual deles está a uma distância estatística maior de $\boldsymbol{\mu}$?

8. Se colocarmos mais atributos na matriz $\mathbf{X}$ (sempre mantendo as colunas lineamente independentes), podemos garantir que o R^2 vai sempre aumentar. Explique porque isto acontece.

9. São observadas n v.a.'s $Y_1, \ldots, Y_n$ que supostamente seguem uma distribuição Poisson com parâmetro λ. Uma maostra de $n = 30$ elementos é obtida e $\bar{y} = 1.633$. Com este valor, calcula-se o valor esperado apresentado na tabela abaixo. Desej-ase usar o teste qui-quadrado para avaliar a qualidade do ajsute de uma distribuição Poisson a estes dados.

Contagem	0	1	2	3	4	≥ 5
Observado	7	8	8	4	2	1
(Esperado)	5.86	9.57	7.81	4.25	1.74	0.77

- Explique como os valores esperados foram obtidos.
- As duas últimas entradas da tabela possuem valor esperado muito abaixo de 5. O que deve ser feito com estas categorias para realizar o teste qui-quadrado?
- Explique como é calculado o valor da estatística de teste X^2. Deixe as contasa indicadas.
- Suponha que a estatística qui-quadrado tenha resultado no valor $X^2 =??$. Obtenha um valor aproximado para o p-valor a partir da tabela abaixo. Para alguns valores de probabilidade α, ela fornece o quantil q definido como sendo $\mathbb{P}(S \leq q) = \alpha$.

α	0.01	0.05	0.10	0.90	0.95	0.99
q	0.30	0.71	1.06	7.78	9.49	13.20