CE-057: Introdução aos modelos lineares
Aulas Práticas

Primeiro Semestre de 2002

Última atualização: 13 de setembro de 2002

Aula Prática 01

Nesta aula é feita uma apresentação geral da parte prática do curso.

É apresentado o programa computacional R no qual as aulas práticas são baseadas. Recursos associados ao programa e disponíveis na *web* são comentados. Uma sessão demonstrativa ilustra os aspectos básicos, capacidades e manuseio do programa.

O programa R é gratuito e de codigo aberto e a página oficial do projeto está em:
http://www.r-project.org.
Há também um espelho (*mirror*) brasileiro da área de *downloads* do programa no Departamento de Estatística da UFPR:
http://www.est.ufpr.br/R.

Aula Prática 02

O objetivo desta aula é que os alunos se familiarizem com aspectos básicos do programa R.

Os alunos devem acessar e seguir os passos do tutorial disponível em: http://www.est.ufpr.br/Rtutorial

Aula Prática 03

1. Considere as seguintes matrizes,

$$A = \begin{bmatrix} 2 & -3 & 2 \\ 4 & 5 & 7 \end{bmatrix}; B = \begin{bmatrix} 5 & 8 \\ 3 & 1 \end{bmatrix}; C = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}; D = \begin{bmatrix} 5 & 9 & 4 \\ 3 & 2 & 5 \end{bmatrix};$$

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}; y = \begin{bmatrix} 12 \\ 18 \\ 15 \\ 10 \\ 24 \\ 22 \\ 22 \\ 19 \end{bmatrix}$$

   1.1 Mostre os comandos para definir estas matrizes no programa R.

   1.2 Obtenha os resultados a seguir “na mão”.

      (a) A’; X’, y’;
      (b) A + D; (A + B)’;
      (c) (AD)’;
      (d) BA; X’X; X’y;
      (e) I_4;
      (f) IX’X; X’XI;
      (g) $5I_5$;
      (h) $B \otimes A$; $A \otimes B$; $C \otimes B$;

   1.3 Agora obtenha novamente os resultados usando o programa R. Mostre os comandos utilizados.

2. Ainda considerando as matrizes do exemplo anterior, mostre os comandos do R para obter:

   2.1 a soma de todos os elementos da matriz **B**.

   2.2 a soma dos elementos de cada linha da matriz **A**.

   2.3 a soma dos elementos de cada coluna da matriz **A**.

2.4 a soma dos quadrados dos elementos do vetor **y**

2.5 o vetor **y1** obtido a partir do vetor **y**:

$$y1 = \begin{bmatrix} 18 \\ 15 \\ 10 \\ 24 \end{bmatrix}$$

2.6 o vetor **y2** obtido a partir do vetor **y**:

$$y2 = \begin{bmatrix} 12 \\ 15 \\ 10 \\ 22 \\ 22 \\ 19 \end{bmatrix}$$

2.7 a sub-matrix **A1** obtida a partir da matrix **A**:

$$A1 = \begin{bmatrix} -3 & 2 \\ 5 & 7 \end{bmatrix}$$

2.8 a sub-matrix **A2** obtida a partir da matrix **A**:

$$A2 = \begin{bmatrix} 2 & 2 \\ 4 & 7 \end{bmatrix}$$

DICAS:

- as funções `rep(), seq()` e `cbind()` podem ser utilizadas para gerar objetos com padrões de repetição como por exemplo a matrix **X** acima. Tente gerar esta matrix obtendo estes comandos.
- a fuņcao `apply()` pode ser utilizada para efetuar operações em linhas e/ou colunas de uma matrix. Use esta função no exercício acima.

3. O conceito de *array* generaliza a idéia de matrix. Enquanto em uma matrix os elementos são organizados em duas dimensões (linhas e colunas), em um array os elementos podem ser organizados em um número arbitrário de dimensões.

   No R um array é definido utilizando a função `array()`.

3.1 Defina um array com o comando a seguir e inspecione o objeto certificando-se que voce entendeu como arrays são criados.

```
ar1 <- array(1:24, dim=c(3,4,2))
ar1
```

3.2 Inspecione a "help" (digite `help(array)` da função array, rode e inspecione os exemplos lá contidos.

3.3 Veja agora um exemplo de dados já incluido no R no formato de array. Para "carregar" e visualizar os dados digite:

```
data(Titanic)
Titanic
```

Para maiores informações sobre estes dados digite:

```
help(Titanic)
```

Agora responda às seguintes perguntas, mostrando os comandos do R utilizados:

(a) quantas pessoas havia no total?

(b) quantas pessoas havia na tripulação (*crew*)?

(c) quantas crianças sobreviveram?

(d) qual a proporção (em %) entre pessoas do sexo masculino e feminino entre os passageiros da primeira classe?

Aula Prática 04

Para os exercícios desta aula prática considere as matrizes definidas a seguir. Recomenda-se que os exercícios sejam resolvidos manualmente e depois utilizando o programa computacional R.

$$A = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 4 & 3 \\ -2 & 1 & 5 \end{bmatrix}; B = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}; C = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix};$$

$$D = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 7 & 12 \end{bmatrix}; E = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}; F = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}; K = 2$$

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

1. Usando as matrizes A e B, verifique as seguintes relações:

   (a) TR[A±B]=TR[A] ± TR[B];

   (b) TR[A]=TR[A'];

   (c) TR[AB]=TR[BA];

   (d) TR[A^{-1}BA]=TR[B];

   (e) TR[AA']= $\sum\limits_{i,j} a_{il}^2$.

2. Obtenha:

   (a) A F.E.C. de **A**, **B**, **C** e **D**;

   (b) O rank de **A**, **B**, **C**, **D**, X e X'X ;

(c) $\mathbf{A}^{-1}$, $\mathbf{B}^{-1}$, $\mathbf{C}^{-1}$, $\mathbf{D}^{-1}$ e $(\text{X'X})^{-1}$ se existirem;

(d) Mostre que as matrizes acima que possuem inversa satisfazem a condição AA^{-1}=A^{-1}A=I.

3. Verifique as seguintes propriedades de matriz inversa:

   (a) $(\mathbf{EE}^{-1})$=$\mathbf{I}$;

   (b) $(\mathbf{E}^{-1})^{-1}$=$\mathbf{E}$;

   (c) Se $\exists\ E^{-1}$ e $\exists\ F^{-1}$ então $(\text{EF})^{-1} = F^{-1}E^{-1}$;

   (d) $(E^{-1})' = (E')^{-1}$

   (e) $(EK)^{-1} = (KE)^{-1} = \frac{1}{K}E^{-1}$

4. Dado o seguinte sistema de equações lineares

$$\left\{ \begin{array}{rcrcl} 2x_1 & + & x_2 & = & 3 \\ 2x_1 & + & 3x_2 & = & 5 \end{array} \right.$$

   (a) Escreva o sistema na forma matricial $Ax = b$;

   (b) Determine a solução $x = A^{-1}b$;

   (c) Pré multiplique ambos os membros de $Ax = b$ por A';

   (d) Determine a solução do novo sistema através de $x = (A'A)^{-1}A'b$

Aula Prática 05

1. Dados,

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 0 & 3 \\ 1 & 1 & 0 \end{bmatrix} ; b = \begin{bmatrix} 9 \\ 11 \\ 3 \end{bmatrix}$$

   Encontre:

   (a) r[A];

   (b) r[A⋮b];

   (c) A^{-1};

   (d) Encontre a solução do sistema Ax=b;

   (e) Discuta sua consistência.

2. Dados:

$$C = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 0 \\ 4 & 3 & 3 \end{bmatrix} ; d = \begin{bmatrix} 6 \\ 4 \\ 10 \end{bmatrix}$$

   Refaça os itens do exercício anterior.

3. Dada a matriz

$$X = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

   (a) Encontre a inversa de Moore-Penrose X^+ (veja **help** para o comando `ginv` do R);

   Nota: Para usar a função `ginv` voce deve carregar o pacote `MASS` (digite `require(MASS)`).

   (b) Verifique suas propriedades;

      i. $XX^+X=X$

ii. $X^+XX^+=X^+$

iii. $XX^+=(XX^+)'=X^{+}{}'X'$

iv. $X^+X=(X^+X)'=X'X^{+}{}'$

4. Dada a matriz

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Utilizando o algoritmo de Searle (1971), obtenha uma inversa condicional de X.

Aula Prática 06

Dada a matriz,

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

1. Encontre duas inversas de mínimos quadrados para X.

2. Considere o seguinte sistema de equações lineares $X\theta = y$,

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 3 \\ 4 \\ 6 \\ 7 \\ 8 \\ 9 \\ 8 \\ 10 \end{bmatrix}$$

   (a) Obtenha $X'X\theta = X'y$ (Sistema de Equações Normais);

   (b) Determine:

      i. $\theta_1^\circ = X^+ y$;

      ii. $\theta_2^\circ = (X'X)^+ X'y$;

      iii. $\theta_3^\circ = (X'X)^- X'y$;

      iv. $\theta_4^\circ = X^l y$;

   (c) Qual vetor solução θ_i° que corresponde ao erro de menor norma?

Aula Prática 07

Dado o sistema Ax=b,

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 5 \\ 3 \end{bmatrix};$$

1. Soluções aproximadas

   (a) Defina um vetor x^q solução qualquer para o sistema Ax=b inconsistente.

   (b) Encontre a melhor solução aproximada x^b (BAS) de Ax=b inconsistente.

   (c) Encontre a soma de quadrados do erro para as soluções anteriores.

2. Soluções de mínimos quadrados.

   (a) Encontre uma solução de mínimos quadrados (x°) para A'Ax=A'b.

   (b) Mostre que (x°) também é solução de mínimos quadrados do sistema Ax=b.

Aula Prática 08

Escrevendo Funções No R

O objetivo desta prática é uma rápida introdução ao uso de *funções* no R.

Funções nada mais são que um grupo de comandos para executar uma determinada tarefa que são agrupados de modo a tornar a execução mais simples e eficiente.

O formato básico de uma função é o seguinte:

```
NOMEDAFUNÇÃO <- function(ARGUMENTOS){
  COMMANDOS
  \ldots
  return(RESULTADO)
}
```

EXEMPLO 1: Uma função para calcular a cubo de um número.

```
cubo <- function(x){
  cubo <- x^3
  return(cubo)
}
cubo(2)
cubo(5)
cubo(1:5)
```

Nesta função x é o *argumento* e a função retorna o objeto `cubo`. Uma versão simplificada desta função é:

```
cubo <- function(x){x^3}
cubo(2)
cubo(5)
cubo(1:5)
```

pois o comando `return` não é obrigatório e em sua ausência a função retorna o objeto definido na última linha.

EXEMPLO 2: Intervalo de confiança para média.

Considere o seguinte conjunto de dados.

```
alt <- c(165, 164, 185, 188, 196, 183, 152, 175, 190, 162,
         162, 185, 162, 193, 167)
```

O intervalo de confiança para a média pode ser obtido da seguite forma:

```
alt.m <- mean(alt)
alt.se <- sqrt(var(alt))/sqrt(15)
tval <- qt(c(0.025, 0.975), 14)
tval
alt.ic <- c(alt.m + tval * alt.se)
alt.ic
```

Agora, vamos escrever e testar uma função para calcular o intervalo de confiança:

```
ic.mean <-function(x, conf){
  n <- length(x)
  alt.m <- mean(x)
  alt.se <- sqrt(var(x))/sqrt(n)
  tval <- qt(c((1-conf)/2, 1-((1-conf)/2)), n-1)
  alt.ic <- c(alt.m + tval * alt.se)
  return(alt.ic)
}
```

e para usar a função:

```
ic.mean(alt, conf = 0.95)
ic.mean(alt, conf = 0.99)
ic.mean(alt, conf = 0.90)
```

Vamos agora *sofisticar* um pouco a função adicionando um teste para checar o valor de nível de confiança e incluindo nomes no vetor resposta.

```
ic.mean <- function(x, conf){
  if(conf < 0 | conf > 1)
    stop("argumento conf deve ser um numero no intervalo [0,1]")
  n <- length(x)
  alt.m <- mean(x)
  alt.se <- sqrt(var(x))/sqrt(n)
  tval <- qt(c((1-conf)/2, 1-((1-conf)/2)), n-1)
```

```
  alt.ic <- c(alt.m + tval * alt.se)
  names(alt.ic) <- c("lim.inf", "lim.sup")
  return(alt.ic)
}

ic.mean(alt, conf = 0.95)
ic.mean(alt, conf = 95)
```

1. Amplie a função acima para que retorne também o intervalo de confiança para a variância.

2. escreva uma função para calcular o intervalo de confiança para uma proporção.

3. escreva uma função para a qual a entrada seja um sistema e a saída contenha:

   - a consistência do sistema;
   - o rank da matrix do sistema;
   - a solução, se existir;
   - uma solução aproximada de mínimos quadrados, se não hover solução exata;
   - os vetores de valores preditos e erros
   - a soma de quadrados do erro.

4. Teste a função que voce escreveu com os exercícios das aulas anteriores:

   - aula 4, ex. 4
   - aula 5, ex. 1 e 2
   - aula 6, ex. 2
   - aula 7

Aula Prática 09

Considere o conjunto de dados abaixo com os quais pretende-se ajustar um modelo de regressão linear simples da forma $Y = X\beta + \varepsilon$, onde ε é vetor variáveis aleatórias normais independentes e homocedásticas,

x	y
$4,5$	$87,1$
$5,1$	$93,1$
$4,8$	$89,8$
$4,4$	$91,4$
$5,9$	$99,5$
$4,7$	$92,1$
$5,1$	$95,5$
$5,2$	$99,3$
$4,9$	$93,4$
$5,1$	$94,4$

1. Calcule as quantidades indicadas a seguir,

   (a) Determine as matrizes/vetores X, Y, $X'X$, $X'Y$ e $(X'X)^{-1}$;

   (b) Encontre o vetor solução $\mathbf{b}$, para os coeficientes do modelo de regressão linear simples;

   (c) Encontre a matriz de projeção H e verifique sua propriedade de idempotência;

   (d) Encontre o vetor de valores estimados $\hat{Y}$;

   (e) Encontre o vetor de erros ($\hat{e}$), a soma de quadrados dos erros (SQE) e o quadrado médio do erro (QME);

   (f) Encontre $\hat{Y}$ e $\hat{e}$ utilizando a matriz H;

   (g) Encontre a matriz de variância e covariância estimada dos erros ou resíduos;

   (h) Monte o intervalo de confiança (95%) para o coeficiente de inclinação da reta.

2. Escreva uma função para fazer uma análise de regressão retornando os resultados relevantes calculados no exercício anterior.

Aula Prática 10

Considere o conjunto de dados abaixo com os quais pretende-se ajustar um modelo de regressão linear simples da forma $Y = X\beta + \varepsilon$, onde ε é vetor variáveis aleatórias normais independentes e homocedásticas,

x	y
$4,5$	$87,1$
$5,1$	$93,1$
$4,8$	$89,8$
$4,4$	$91,4$
$5,9$	$99,5$
$4,7$	$92,1$
$5,1$	$95,5$
$5,2$	$99,3$
$4,9$	$93,4$
$5,1$	$94,4$

1. Calcule as quantidades indicadas a seguir,

   (a) Determine as matrizes/vetores $Y'JY$, $Y'Y$, $b'X'Y$, $(X'X)^{-1}$, onde J é uma matriz de 1's e b é uma solução de mínimos quadrados;

   (b) Encontre a SQTotal, SQErro e a SQRegressão, para o modelo de regressão linear simples;

   (c) Construa o quadro da Análise de Variância da Regressão;

   (d) Encontre a matriz de variância e covariância estimada de b;

   (e) Encontre o valor médio estimado para uma observação qualquer;

   (f) Encontre a variância estimada para esta observação;

2. Escreva uma função para fazer uma análise de regressão retornando os resultados relevantes calculados no exercício anterior.

Aula Prática 11

Considere o conjunto de dados abaixo com os quais pretende-se ajustar um modelo de regressão linear múltipla da forma $Y = X\beta + \varepsilon$, onde ε é vetor variáveis aleatórias normais, independentes e homocedásticas,

Clique para ver e/ou copiar a arquivo aula11dados.txt com o conjunto de dados.

$x1$	$x2$	$x3$	y
19.5	43.1	29.1	11.9
24.7	49.8	28.2	22.8
30.7	51.9	37.0	18.7
29.8	54.3	31.1	20.1
19.1	42.2	30.9	12.9
25.6	53.9	23.7	21.7
31.4	58.5	27.6	27.1
27.9	52.1	30.6	25.4
22.1	49.9	23.2	21.3
25.5	53.5	24.8	19.3
31.1	56.6	30.0	25.4
30.4	56.7	28.3	27.2
18.7	46.5	23.0	11.7
19.7	44.2	28.6	17.8
14.6	42.7	21.3	12.8
29.5	54.4	30.1	23.9
27.7	55.3	25.7	22.6
30.2	58.6	24.6	25.4
22.7	48.2	27.1	14.8
25.2	51.0	27.5	21.1

1. Considerando os dados acima:

   (a) Faça a análise de regressão, considerando um modelo de regressão linear múltipla;

   (b) Encontre a matriz de correlação para as variáveis preditoras do modelo;

   (c) Discuta sobre a correlação entre estas variáveis;

   (d) Encontre a estimativa de σ^2 do modelo;

(e) Estime a matriz de variância e covariância de $\hat{\beta}$;

(f) Faça a padronização das variáveis preditoras X e da resposta Y;

(g) Faça a análise de regressão com as variáveis transformadas;

(h) Obtenha os coeficientes de regressão das variáveis originais a partir dos coeficientes da regressão com as variáveis transformadas;

(i) Encontre a inversa da matriz de correlação entre as variáveis preditoras (r_{xx}^{-1})

(j) Encontre a estimativa de $(\sigma')^2$ no modelo transformado;

(k) Estime a matriz de variância e covariância de $\hat{\beta}'$;

(l) Obtenha e interprete os VIF para cada variável preditora;

(m) Calcule o determinante de $X'X$ não transformado. Qual seu valor? O que isso implica?

2. Escreva uma função que retorne os resultados relevantes calculados no exercício anterior.

Aula Prática 12

1. Simulando dados de acordo com o modelo de regressão linear simples.

   (a) defina um modelo de regressão linear simples e simule dados deste modelo
   (dica: use a função `rnorm()` para gerar amostras da distribuição normal),

   (b) faça um gráfico dos dados,

   (c) use os dados para estimar os parâmetros do modelo usando a sua função escrita em aulas passadas,

   (d) compare suas estimativas com os verdadeiros valoes dos parâmetros e discuta as diferenças,

   (e) faça um gráfico de resíduos contra valores preditos,

   (f) estime novamente os parâmetros agora utilizando a função `lm()` do R,

   (g) utilize a função `plot()` no objeto gerado pela função `lm()` e discuta os gráficos produzidos.

2. Simulando dados de acordo com o modelo de regressão linear múltipla

   (a) simule dados de acordo com o modelo de regressão linear múltipla,

   (b) obtenha as estimativas dos parâmetros,

   (c) simule novamente agora de um modelo com multicolinearidade,

   (d) tente encontrar as estimativas e discuta os resultados.

Aula Prática 13

Considere o conjunto de dados a seguir, proveniente de um experimento instalado em delineamento completamente casualizado.

Rep.	Trat. 1	Trat. 2	Trat. 3
1	5	6	9
2	4	7	8
3	3	8	10
4	4	-	-

Considerando o modelo $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ usualmente adotado para este tipo de experimento:

1. determine as matrizes do sistema de equações normais;

2. determine uma solução para o sistema de equações normais;

3. obtenha os **projetores ortogonais** $P = XX^+$ e $P_1 = \frac{1}{n}J_n$ onde J_n é uma matriz quadrada com todos elementos iguais a 1 e de dimensão igual ao número de dados;

4. encontre a SQTotal, SQParâmetros , SQMédia, SQTratamentos e a SQErro.

5. construa o quadro da análise de variância.

Solução dos Exercícios

O arquivo ce057pratica.R contém os comandos do R para resolver exercícios propostos nas aulas práticas.