MAE126 - Noções de Estatística II

Turma Prof. Popov

17 de junho de 2006

Correção da Lista 7

1. Numa pesquisa sobre rendimentos por hora, com assalariados segundo o grau de instrução, obtiveram-se os dados da tabela abaixo (rendimentos expressos como porcentam de salário mínimo). Construa a tabela ANOVA e verifique se existe diferença significativa entre os rendimentos das duas categorias.

Escolaridade	n	$\sum x_i$	$\sum x_i^2$
Fundamental	41	87,7	202,45
Médio	21	71,1	258,79

A hipótese que queremos testar aqui é $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$. Para construir a tabela de ANOVA, comecemos calculando algumas estatísticas que vão nos ser úteis:

$$\overline{y}_1 = \frac{\sum y_{1j}}{n_1} = \frac{87,7}{41} = 2,14 \qquad \overline{y}_2 = \frac{\sum y_{2j}}{n_2} = \frac{71,1}{21} = 3,39$$

$$\overline{y} = \frac{\sum y_{1j} + \sum y_{2j}}{n} = \frac{87,7 + 71,1}{62} = 2,56$$

$$\begin{aligned} S_1^2 &= \frac{1}{n_1 - 1} \sum (y_{1j} - \overline{y}_1)^2 = \frac{1}{n_1 - 1} \left(\sum y_{1j}^2 - \frac{(\sum y_{1j})^2}{n_1} \right) \\ &= \frac{1}{40} \left(202,45 - \frac{87,7^2}{41} \right) = 0,37 \end{aligned}$$

$$\begin{aligned} S_2^2 &= \frac{1}{n_2 - 1} \sum (y_{2j} - \overline{y}_2)^2 = \frac{1}{n_2 - 1} \left(\sum y_{2j}^2 - \frac{(\sum y_{2j})^2}{n_2} \right) \\ &= \frac{1}{20} \left(258,79 - \frac{71,1^2}{21} \right) = 0,90 \end{aligned}$$

Basta agora calcularmos $SQEnt$ e $SQDen$:

$$
\begin{aligned}
SQEnt &= n_1(\overline{y}_1 - \overline{y})^2 + n_2(\overline{y}_2 - \overline{y})^2 \\
&= 41(2,14 - 2,56)^2 + 21(3,39 - 2,56)^2 \\
&= 21,70 \\
SQDen &= (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 \\
&= 40(0,37) + 20(0,90) \\
&= 32,80
\end{aligned}
$$

Podemos assim montar a tabela de ANOVA:

F.V.	g.l.	SQ	QM	F
Entre	1	$SQEnt$	$QMEnt$	$QMEnt/QMDen$
Dentro	$n-2$	$SQDen$	$QMDen$	
Total	$n-1$	$SQTot$	$QMTot$	

Preenchido com as estatísticas que calculamos (notando que $SQTot = SQEnt = SQDen$):

F.V.	g.l.	SQ	QM	F
Entre	1	21,70	21,70	39,45
Dentro	60	32,80	0,55	
Total	61	54,50	0,89	

A região crítica é obtida a partir da tabela F com 1 e 60 graus de liberdade, com 5%, temos:

$$RC = \{f_{obs} \geq 4,00\}$$

Como $f_{obs} = 39,45 \in RC$, rejeitamos a hipótese nula, concluindo que as médias de salários diferem.

2. Quer-se verificar se o tempo médio de espera para o atendimento nas agências do BB depende da cidade. Em 3 cidades brasileiras, foram colhidos os seguintes dados em 27 agências:

   Cidade 1: 12, 15, 22, 17, 18, 11, 21, 19, 15;
   Cidade 2: 17, 19, 25, 23, 29, 24, 18, 31;
   Cidade 3: 23, 24, 29, 33, 19, 22, 29, 28, 31, 35.

   Construa a tabela ANOVA e verifique a hipótese de que as médias não dependem da cidade.

   A nossa hipótese de interesse é $H_0 : \mu_1 = \mu_2 = \mu_3$ contra H_A : ao menos uma das médias difere das outras. Calculemos algumas estatísticas de interesse:

$$\overline{y}_1 = \frac{\sum y_{1j}}{n_1} = \frac{150}{9} = 16,67 \qquad \overline{y}_2 = \frac{\sum y_{2j}}{n_2} = \frac{186}{8} = 23,25$$

$$\overline{y}_3 = \frac{\sum y_{3j}}{n_1} = \frac{273}{10} = 27,30 \qquad \overline{y} = \frac{\sum y_{kj}}{n} = \frac{150+186+273}{27} = 22,56$$

$$\begin{aligned} S_1^2 &= \frac{1}{n_1-1}\sum(y_{1j}-\overline{y}_1)^2 = \frac{1}{n_1-1}\left(\sum y_{1j}^2 - \frac{(\sum y_{1j})^2}{n_1}\right) \\ &= \frac{1}{8}\left(2614 - \frac{150^2}{9}\right) = 14,25 \end{aligned}$$

De forma análoga obtemos: $S_2^2 = 25,93$, $S_3^3 = 26,46$ (com $n_1 = 9, n_2 = 8$ e $n_3 = 10$).

Basta agora calcularmos $SQEnt$ e $SQDen$:

$$\begin{aligned} SQEnt &= n_1(\overline{y}_1-\overline{y})^2 + n_2(\overline{y}_2-\overline{y})^2 + n_3(\overline{y}_3-\overline{y})^2 \\ &= 9(16,67-22,56)^2 + 8(23,25-22,56)^2 + 10(27,30-22,56)^2 \\ &= 540,71 \\ SQDen &= (n_1-1)S_1^2 + (n_2-1)S_2^2 + (n_3-1)S_3^2 \\ &= 8(14,25) + 7(25,93) + 9(26,46) \\ &= 533,65 \end{aligned}$$

Podemos assim montar a tabela de ANOVA:

F.V.	g.l.	SQ	QM	F
Entre	$k-1$	$SQEnt$	$QMEnt$	$QMEnt/QMDen$
Dentro	$n-k$	$SQDen$	$QMDen$	
Total	$n-1$	$SQTot$	$QMTot$	

Preenchido com as estatísticas que calculamos (notando que $SQTot = SQEnt = SQDen$):

F.V.	g.l.	SQ	QM	F
Entre	2	$540,71$	$270,35$	$12,16$
Dentro	24	$533,65$	$22,24$	
Total	26	$1074,36$	$41,32$	

A região crítica é obtida a partir da tabela F com 2 e 24 graus de liberdade, com 5%, temos:

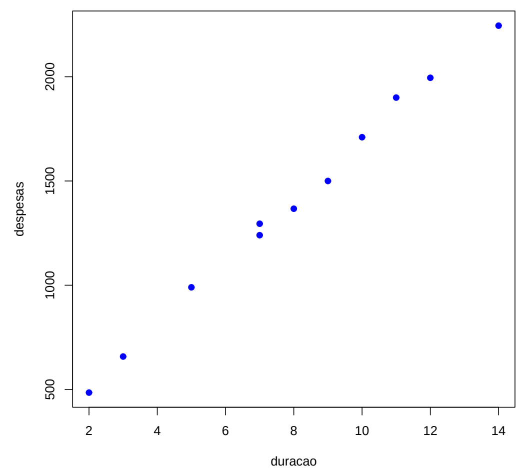
$$RC = \{f_{obs} \geq 3,40\}$$

Como $f_{obs} = 12,16 \in RC$, rejeitamos a hipótese nula, concluindo que as médias do tempos diferem entre as cidades.

3. A tabela abaixo se refere às despesas de 11 casais de classe média alta com as viagens para França (excluindo preço da passagem aérea):

Despesas	Duração da Viagem	Despesas	Duração da Viagem
658	3	2245	14
990	5	1900	11
485	2	1500	9
1295	7	1710	10
1240	7	1995	12
1367	8		

(a) Construa o diagrama de dispersão.



(b) Calcule os coeficientes α e β da reta dos mínimos quadrados. Qual é o significado destes coeficientes?

Queremos ajustar uma reta do tipo:

$$\text{despesas} = \alpha + \beta \text{duracao}$$

Basta utilizarmos as fórmulas:

$$\hat{\beta} = \frac{\sum x_i y_i - n\overline{x}\overline{y}}{\sum x_i^2 - n\overline{x}^2} = \frac{143445 - 11(8)(1398,64)}{842 - 11(8)^2} = 147,57$$

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x} = 1398,64 - 147,57(8) = 218,08$$

Assim:

$$\text{despesas} = 218,08 + 147,57\text{duracao}$$

O coeficiente α quer dizer que para uma viagem com 0 dias, as despesas seriam de 218,08. O coeficiente β significa que a cada dia mais gasto em viagem, implica em $147,57$ em gastos adicionais.

(c) De acordo com (b), qual seria o gasto esperado para uma viagem de 20 dias?

O gasto esperado seria de:

$$\text{despesas} = 218,08 + 147,57(20) = 3169,48$$