MAT-042, 2° Semestre 2013 **Pauta de Corrección**
Certamen 1. Octubre 28, 2013
Tiempo: 90 Minutos Felipe Osorio, Patricio Videla.

1.a. Los números sorteados en la lotería de Maryland tienen un promedio de 518.96 y una desviación estándar de 291.7 unidades. Desde el boxplot y el coeficiente de asimetría $\gamma_1 = -0.0921$ se aprecia que los datos son simétricos y que su curtosis es negativa ($\gamma_2 = -1.2093$) indicando que la distribución es achatada (platicúrtica). Además, la información provista por el histograma indica que los datos siguen un comportamiento bastante uniforme.

1.b Puede ser notado que

$$(x_i - x_j)^2 = ((x_i - \overline{x}) - (x_j - \overline{x}))^2 = (x_i - \overline{x})^2 + (x_j - \overline{x})^2 - 2(x_i - \overline{x})(x_j - \overline{x}),$$

de este modo,

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2 &= \sum_{i=1}^n \sum_{j=1}^n \{(x_i - \overline{x})^2 + (x_j - \overline{x})^2 - 2(x_i - \overline{x})(x_j - \overline{x})\} \\ &= n\sum_{i=1}^n (x_i - \overline{x})^2 + n\sum_{j=1}^n (x_j - \overline{x})^2 - 2\sum_{i=1}^n (x_i - \overline{x}) \sum_{j=1}^n (x_j - \overline{x}).\end{aligned}$$

Sabemos que $\sum_{i=1}^n (x_i - \overline{x}) = 0$ (análogamente para $\sum_{j=1}^n (x_j - \overline{x}) = 0$), luego

$$\sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2 = 2n \sum_{i=1}^n (x_i - \overline{x})^2.$$

Finalmente, tenemos

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x})^2 = \frac{1}{2n(n-1)} \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2.$$

2.a. Desde el conjunto de datos $(x_1, y_1), \dots, (x_6, y_6)$, podemos obtener:

$$\sum_{i=1}^n x_i = 0, \quad \sum_{i=1}^n x_i^2 = 130, \quad \sum_{i=1}^n y_i = 0.41, \quad \sum_{i=1}^n y_i^2 = 10.5009, \quad \sum_{i=1}^n x_i y_i = -10.59,$$

con $n = 6$. De este modo,

$$\begin{aligned}\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) &= \sum_{i=1}^n x_i y_i - n\,\overline{x}\,\overline{y} = -10.59, \\ \sum_{i=1}^n (x_i - \overline{x})^2 &= \sum_{i=1}^n x_i^2 - n\,\overline{x}^2 = 130.\end{aligned}$$

Luego, tenemos que la estimación de los coeficientes de regresión es dada por:

$$\begin{aligned}&\widehat{\beta} = \frac{\operatorname{cov}(\boldsymbol{x}, \boldsymbol{y})}{\operatorname{var}(\boldsymbol{x})} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = -\frac{10.59}{130} = -0.08146, \\ &\widehat{\alpha} = \overline{y} - \widehat{\beta}\,\overline{x} = 0.06833.\end{aligned}$$

2.b. El vector de valores predichos está dado por:

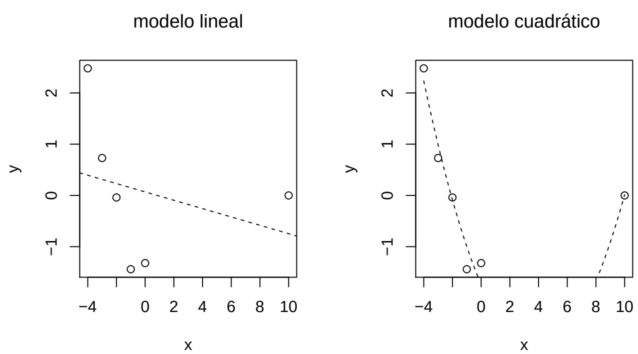
$$\widehat{\boldsymbol{y}} = (2.0858, 0.4173, -0.2713, -1.5898, -1.3883, 0.7463)^\top,$$

luego

$$r^2 = \{\operatorname{cor}(\boldsymbol{y}, \widehat{\boldsymbol{y}})\}^2 = \frac{\{\operatorname{cov}(\boldsymbol{y}, \widehat{\boldsymbol{y}})\}^2}{\operatorname{var}(\boldsymbol{y})\operatorname{var}(\widehat{\boldsymbol{y}})} = 0.08237.$$

De este modo, un 8.24% de la variabilidad de la respuesta es explicada por el modelo.

2.c. Comparando el valor de r^2 con el gráfico de los datos y el modelo ajustado se aprecia que el ajuste es pobre. Un ajuste bastante mejor puede ser obtenido considerando, por ejemplo, un modelo cuadrático.



3.a. Se pide calcular la probabilidad:

$$P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1 \cup A_2) = 0.12 + 0.07 - 0.13 = 0.06.$$

3.b. En este caso, se requiere calcular:

$$P(A_1 \cap A_2 \cap A_3^c) = P(A_1 \cap A_2) - P(A_1 \cap A_2 \cap A_3) = 0.06 - 0.01 = 0.05.$$

3.c. Se solicita calcular la probabilidad siguiente:

$$\begin{aligned} P\{(A_1 \cup A_2 \cup A_3) \cup (A_1 \cap A_2) \cup (A_1 \cap A_3) \cup (A_2 \cap A_3)\} &= 1 - P(A_1 \cap A_2 \cap A_3) \\ &= 1 - 0.01 = 0.99. \end{aligned}$$

3.d. Se debe calcular:

$$\begin{aligned} P(A_1^c \cap A_2^c \cap A_3^c) &= P\{(A_1 \cup A_2 \cup A_3)^c\} = 1 - P(A_1 \cup A_2 \cup A_3) \\ &= 1 - (0.12 + 0.07 + 0.05 - 0.06 - 0.03 - 0.02 + 0.01) \\ &= 1 - 0.14 = 0.86. \end{aligned}$$

Observación: Desde la información dada en el enunciado, sigue que:

$$\begin{aligned} P(A_1 \cap A_3) &= 0.12 + 0.05 - 0.14 = 0.03, \\ P(A_2 \cap A_3) &= 0.07 + 0.05 - 0.10 = 0.02. \end{aligned}$$

4. En efecto, tenemos que

(a) $P(A) = \sum_{j=1}^{k} a_j P_j(A) \geq 0$, pues $P_j(A) \geq 0$, para todo $j = 1, \ldots, k$.

(b) $P(\Omega) = \sum_{j=1}^{k} a_j P_j(\Omega) = \sum_{j=1}^{k} a_j = 1$,

(c) Sea $A_1, A_2, \ldots, A_n$ colección de eventos disjuntos, entonces

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{j=1}^{k} a_j P_j\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{j=1}^{k} a_j \sum_{i=1}^{n} P_j(A_i) = \sum_{i=1}^{n}\sum_{j=1}^{k} a_j P_j(A_i) = \sum_{i=1}^{n} P(A_i).$$

Luego, sigue que P es medida de probabilidad.