FICH

Universidad Nacional del Litoral

Facultad de Ingeniería y Ciencias Hídricas

Estadística

Ingeniería en Informática

Mg. Susana Vanlesberg: Profesor Titular
Dr. Mario Silber: Profesor Adjunto
Dra. Andrea Bergesio: Jefe de Trabajos Prácticos
A.I.A. Juan Pablo Taulamet: Auxiliar de Primera

:: GUÍA 7 ::	
REGRESIÓN Y CORRELACIÓN	
:: RESPUESTAS ::	:: 2014 ::

Ejercicio 1

(a)

Existe una relación lineal con tendencia decreciente.

(b) Llamamos $X =$ categoría, $Y =$ ausencias.

Modelo planteado para las observaciones: $Y_i = \alpha + \beta X_i + \varepsilon_i$ con ε_i independiente de X_i, $E(\varepsilon_i) = 0$ y $V(\varepsilon_i) = \sigma^2$, $i = 1, \ldots, n$. Este modelo implica $E(Y_i|X_i = x_i) = \alpha + \beta x_i, i = 1, \ldots, n$.

Modelo estimado: $\widehat{Y} = 47.348 - 2.9274x$

(c) Coeficiente de determinación: $r^2 = 0.789$ indica que casi el 80 % de la variabilidad de Y es explicada por el modelo propuesto.

(d) $\widehat{Y}_{x=10} = 47.348 - 2.927 * 10 = 18.074$. Este valor sirve tanto para estimar $E(Y|x = 10)$ como para pronosticar Y cuando $x = 10$. En este último caso, el error estándar asociado **estimado** es

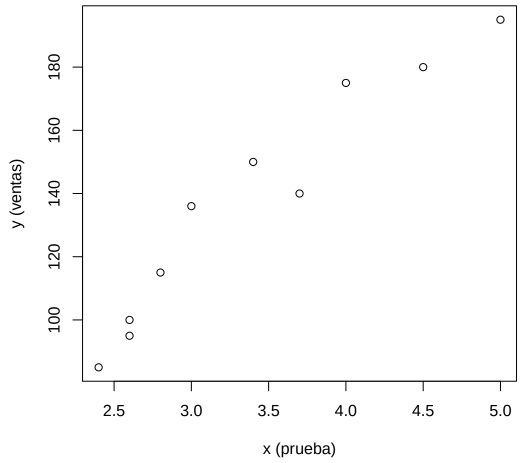
$$\widehat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(\overline{x} - 10)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}},$$

donde $\widehat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{n-2}}$. En este caso $\widehat{\sigma} = 4.368$. Luego el error estándar estimado asociado al valor pronosticado resulta

$$4.368\sqrt{1 + \frac{1}{16} + \frac{(10 - 7.25)^2}{117}} = 4.637.$$

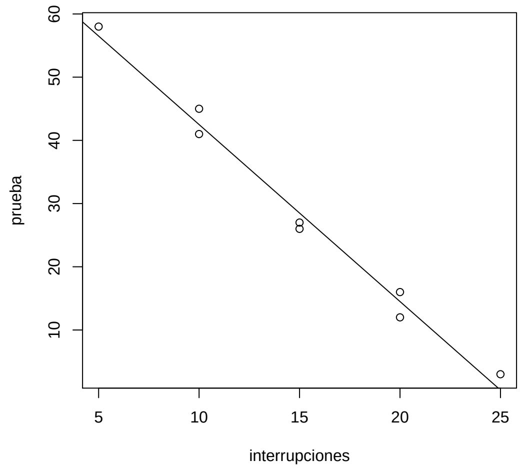
Ejercicio 2

Como hay relación lineal creciente entre las variables estudiadas, el incremento en las ventas por cada punto en la prueba aumenta $\widehat{\beta} = 41.681$ unidades en la ventas.



Ejercicio 3

(a) y (b) Llamamos $X =$ cantidad de interrupciones, e $Y =$ resultado en la prueba.



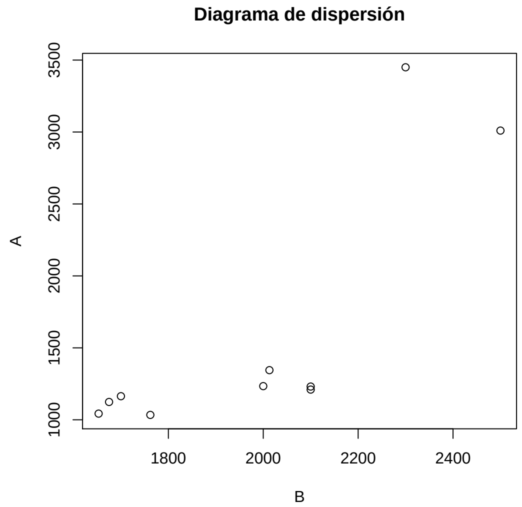
Modelo estimado: $\widehat{Y} = 70.5 - 2.8x$.

(c) $\widehat{Y} = 70.5 - 2.8 * 22 = 8.9$. Es decir, se se lo interrumpe 22 veces, su hostilidad es tan alta como 8.9.

(d) $\widehat{Y} = 70.5 - 2.8 * 35 < 0$, lo que resulta un valor imposible para la prueba. Luego para esta cantidad de interrupciones el modelo propuesto no es apropiado.

Ejercicio 4

(a) A partir de la observación del gráfico de dispersión, proponer una relación lineal no parece la mejor elección. Se propone una transformación para la variable B. El modelo que estimamos



es $Y = \alpha + \beta X^2 + \varepsilon$ donde ε independiente de X, $E(\varepsilon) = 0$, o equivalentemente, $E(Y/X = x) = \alpha + \beta x^2$, donde elegimos $Y = A$ y $X = B$.

Modelo estimado: $\widehat{Y} = -884.5 + 0.00064x^2$.

Observación: Con este modelo, $r^2 = 0.667$. Mientras que si no se realiza una transformación, se consigue $r^2 = 0.634$.

(b) El número esperado de visitas para el sitio A se estima en $\widehat{Y} = -884.5 + 0.000618 * 2500^2 = 2978.62$ cuando el número de visitas para el sitio B es 2500.

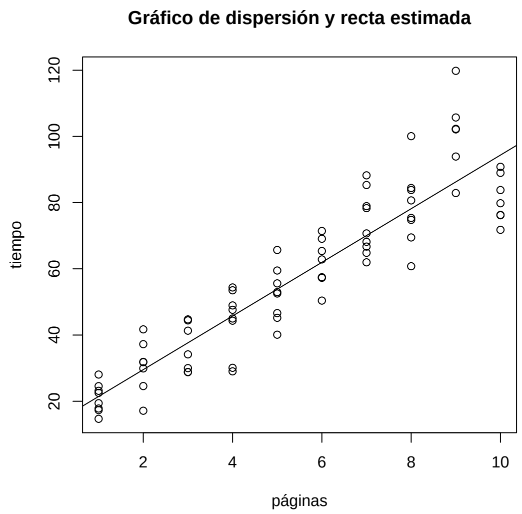
(c) Para la medida del item anterior el error estándar **estimado** asociado resulta

$$\widehat{\sigma}\sqrt{\frac{1}{n} + \frac{(\overline{z} - 2500^2)^2}{\sum_{i=1}^{n}(z_i - \overline{z})^2}},$$

donde $z_i = x_i^2$ y $\widehat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{n-2}}$. En este caso $\widehat{\sigma} = 537.1$. Luego el error estándar estimado correspondinte a la estimación del parámetro $E(Y/X = 2500) = \alpha + \beta 2500^2$ resulta

$$537.1\sqrt{\frac{1}{10} + 0.419} = 387.05.$$

Ejercicio 5



En base al diagrama de dispersión para los datos disponibles se propone el modelo

$$E(Y_i|X_i = x_i) = \alpha + \beta x_i, i = 1, \dots, 75,$$

donde Y_i es el i-ésimo tiempo medido para la cantidad de páginas X_i.

El modelo estimado resulta $\widehat{Y} = 13.345 + 8.102x$, es decir por cada página que se manda a imprimir se estima que el tiempo de impresión aumentará 8.102 unidades. Además es $\widehat{\sigma} = 10.31$.

El coeficiente de correlación $r = 0.914$ lo que indica que la relación lineal es creciente y adecuada. Además $r^2 = 0.835$, esdecir el 83 % aproximadamente de la variablidad observada en los tiempos de impresión es explicada por la relación propuesta con el número de páginas a imprimir.