Examen Parcial. 18 de Marzo de 2009

Estadística I. GDO - ADE

Nombre: Grupo pequeño n^o:

Puntuación sobre 25 puntos. Tiempo de realización: 1 h 15 min
Todas las respuestas finales y los cálculos intermedios deben escribirse en esta misma hoja en los espacios reservados para ello. NO SE PUEDE DESGRAPAR LAS HOJAS.
Para las cuentas en sucio se puede usar la parte de atrás de las hojas.
El resultado final ha de expresarse con 4 decimales.

1. A continuación se presenta una muestra de 10 valores de dos variables estadísticas continuas, X e Y.

X:	3.3	4.8	4.2	3.4	1.9	2.0	2.2	3.9	3.0	2.5
Y:	60	40	52	68	85	80	78	55	62	72

Se ha usado el programa R para obtener un resumen numérico de cada una de estas variables y del análisis de regresión de Y sobre X. Se han obtenido los siguientes resultados:

```
> numSummary(datos[,c("x", "y")], statistics=c("mean", "sd",
+   "quantiles"), quantiles=c(0,.25,.5,.75,1))
    mean        sd   0%     25%    50%     75%  100%   n
x   3.12  0.982966  1.9   2.275   3.15   3.775   4.8  10
y  65.20 14.061768 40.0  56.250  65.00  76.500  85.0  10

> summary(RegModel.1)

Call:
lm(formula = y ~ x, data = datos)

Residuals:
    Min      1Q  Median      3Q     Max
-4.8703 -1.8264 -0.3974  1.5386  6.6973

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  108.627      3.802   28.57 2.44e-09 ***
x            -13.919      1.168  -11.92 2.25e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.443 on 8 degrees of freedom
Multiple R-squared: 0.9467,Adjusted R-squared:  0.94
F-statistic: 142.1 on 1 and 8 DF,  p-value: 2.254e-06
```

a) **(2 pts.)** Indica razonadamente cuál de las variables, X o Y, es más dispersa.

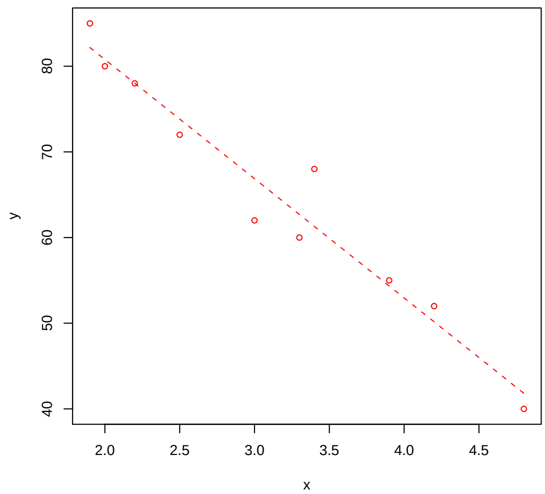
Solución.

$$CV_x = \frac{s_x}{\overline{x}} = \frac{0.982966}{3.12} = 0.3151 \quad CV_x = \frac{s_y}{\overline{y}} = \frac{14.061768}{65.20} = 0.2157.$$

Por lo tanto X es más dispersa al tener mayor coeficiente de variación.

b) **(1.5 pts.)** Representa los datos en un diagrama de dispersión.

Solución.



c) **(2 pts.)** Obtén, indicando cómo lo haces, el coeficiente de correlación e interpreta su significado de acuerdo a la relación lineal entre las dos variables.

Solución. *El coeficiente de correlación es la raíz del coeficiente de determinación. El coeficiente de determinación viene recogido en la salida de R como* Multiple R-squared. *Por tanto:*

$$r_{(x,y)} = \sqrt{r^2_{(x,y)}} = \sqrt{0.9467} = -0.9730.$$

El coeficiente de correlación ha de ser negativo, porque la relación entre ambas variables es claramente negativa (a la vista del diagrama de dispersión y de la ecuación de la recta de regresión), por tanto tomamos la raiz negativa del coeficiente de determinación. Como es un valor muy alto y negativo, podemos decir que existe una relación lineal muy fuerte y negativa entre x e y, es decir, a mayores valores de x, menores valores de y.

d) **(1.5 pts.)** Obtén, indicando cómo lo haces, la ecuación de la recta de regresión de y sobre x, y represéntala en el diagrama realizado anteriormente.

Solución. *Los coeficientes de la recta de regresión vienen recogidos en la salida de R en la columna* Estimate *del apartado* Coefficients *como* Intercept *(ordenada en el origen) y* x *(pendiente).*
La ecuación de la recta de regresión es: $y = -13.919x + 108.627$.

2. En un grupo de alumnos de 2^o de Bachillerato por la rama de Ciencias Sociales de un instituto de Getafe, las expectativas para el año siguiente son: el 60 % quiere estudiar en la Carlos III, el 20 % quiere estudiar en la Rey Juan Carlos y el resto en otras universidades.
Además, se sabe que de entre los que quieren estudiar en la Carlos III, el 40 % quiere estudiar el grado en A.D.E., y ese porcentaje es del 10 % entre los que estudiarán en la Rey Juan Carlos y del 5 % entre los que estudiarán en otras universidades.

a) **(3 pts.)** Calcula la probabilidad que tiene un alumno elegido al azar en ese grupo de estudiar el grado en A.D.E.

Solución. *Por la ley de la probabilidad total:*

$$\begin{aligned} P(ADE) &= P(ADE|UC3M)\cdot P(UC3M) + P(ADE|RJC)\cdot P(RJC) \\ &+ P(ADE|otras)\cdot P(otras) \\ &= 0.4\cdot 0.6 + 0.1\cdot 0.2 + 0.05\cdot 0.2 = 0.2700. \end{aligned}$$

b) **(3 pts.)** Sabiendo que un alumno de ese grupo acaba estudiando el grado en A.D.E., ¿cuál es la probabilidad de que estudie en la Carlos III?

Solución. *Por el teorema de Bayes:*

$$\begin{aligned} P(UC3M|ADE) &= \frac{P(ADE|UC3M)\cdot P(UC3M)}{P(ADE)} \\ &= \frac{0.4\cdot 0.6}{0.27} = 0.8889. \end{aligned}$$

3. Una empresa aseguradora recibe partes de accidentes de vehículos a razón de 3 al día. Además, se sabe que el número de partes que llegan al día sigue una distribución de Poisson.

a) **(3 pts.)** Calcula la probabilidad de que en un día determinado lleguen 3 partes.

Solución. *Sea $X =$"número de partes que llegan al día"; $X \sim \mathcal{P}(3)$.*

$$P(X=3) = \tfrac{e^{-3}\cdot 3^3}{3!} = e^{-3}\cdot 4.5 = 0.2240.$$

b) **(3 pts.)** Calcula la probabilidad de que en dos días consecutivos lleguen más de 2 partes.

Solución. *Sea $Y =$"número de partes que llegan en dos días"; $Y \sim \mathcal{P}(3\cdot 2) = \mathcal{P}(6)$.*

$$\begin{aligned} P(Y>2) &= 1 - P(Y\le 2) = 1 - (P(Y=0) + P(Y=1) + P(Y=2)) \\ &= 1 - \left(\tfrac{e^{-6}\cdot 6^0}{0!} + \tfrac{e^{-6}\cdot 6^1}{1!} + \tfrac{e^{-6}\cdot 6^2}{2!}\right) \\ &= 1 - e^{-6}(1+6+18) = 1 - 0.0620 = 0.9380. \end{aligned}$$

c) **(3 pts.)** Para cada parte que llega, la cuantía que la compañía aseguradora tiene que pagar al asegurado sigue una distribución normal de media 300 euros y desviación típica 400 euros. Calcula la probabilidad de que en un parte determinado haya que pagarle al asegurado más de 1200 euros.

Solución. *Sea $C =$"cuantía que se paga por parte"; $C \sim N(300, 400)$.*

$$\begin{aligned} P(C>1200) &= P\left(\tfrac{C-300}{400} > \tfrac{1000-300}{400}\right) \quad \left\{Z = \tfrac{C-300}{400} \sim N(0,1)\right\} \\ &= P(Z > 2.25) = 0.0122. \end{aligned}$$

d) **(3 pts.)** La probabilidad de que la cuantía a pagar para un parte determinado supere los 300 euros es 0.5. Calcula la probabilidad de que entre los 10 siguientes partes que lleguen haya 5 con una cuantía de pago superior a los 300 euros.

Solución. *Sea $W =$"número de partes, entre los 10 siguientes, con una cuantía de pago superior a los 300 euros."*
Suponiendo que los distintos partes son independientes entre sí, tenemos que $W \sim \mathcal{B}(10, 0.5)$.

$$P(W=5) = \binom{10}{5} 0.5^5 \cdot 0.5^5 = 0.2461.$$