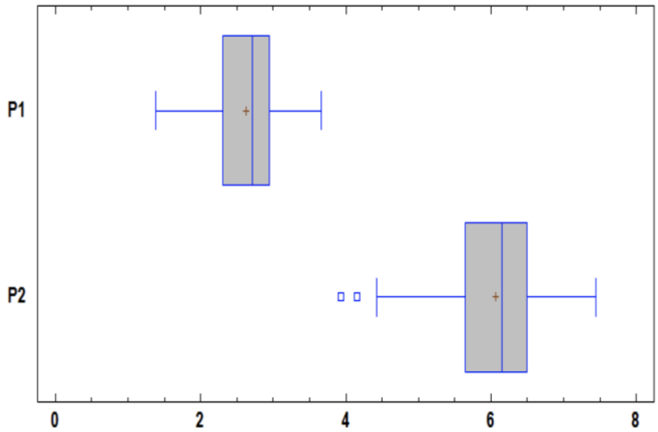
Examen 2

1. Queremos estudiar la relación entre la nota final de nuestros estudiantes (variable “Grade”) y las notas obtenidas en dos actividades relativas a la evaluación continua (variables “P1” y “P2”). Lee detenidamente y responde a las siguientes cuestiones:



a) Teniendo en cuenta los diagramas de cajas para las variables “P1” y “P2” (ver la Figura), ¿Hay algún problema de presencia de atípicos? Si tu respuesta es SI: especifica en qué te basas.

b) Por lo que se puede observar en la Figura, ¿qué actividad (P1 o P2) crees que tiene menor promedio de notas? ¿Por qué?

c) Con objeto de modelizar la relación entre la variable respuesta (“Grade”) y las variables explicativas (“P1” y “P2”) se propone el siguiente modelo. Interpreta el valor del coeficiente de determinación $R^2\ ajustado$.

$$Grade = 1.03 + 0.64 * P1 + 0.19 * P2 + 0.43 * Type$$

$$R^2\ ajustado = 86.7\%$$

SOLUCIÓN:

a) Sí hay presencia de atípicos, en la segunda variable P1, ya que en la parte izquierda del segundo boxplot (el de abajo) se ven dos cuadraditos que representan a los atípicos.

b) Como se observa en la Figura, el boxplot de P1 está más hacia la izquierda, es decir, hacia los valores más pequeños de la variable (mirar el eje de valores horizontal). Miremos la caja del boxplot que es donde se concentra el 50% de datos más centrales de la variable, y donde está la media y también la mediana. Por lo cual, la variable P1 tendrá una media (promedio) menor que la variable P2, lo que significa que las notas de la actividad P1 en nuestra muestra de datos, en promedio parecen ser menores que las notas en la actividad P2.

c) El coeficiente de determinación R^2 en regresión simple es igual a la correlación lineal de Pearson, pero aquí estamos en una regresión lineal múltiple porque tenemos más de 1 variable explicativa. En regresión múltiple el R^2 se tiene que ajustar, y se le llama $R^2\ ajustado$ y tiene una interpretación parecida, porque mientras más alto sea, mejor será el ajuste del modelo, ya que este mide el porcentaje de variabilidad de la variable dependiente, en función de las variables explicativas (o independientes). Por lo cual, mientras mayor sea la habilidad de las variables explicativas de "explicar" a la variable Y, mejor será el modelo. Un $R^2\ ajustado =\ 86.7\%$ es bastante bueno.

2. En el área de Probabilidad, el conjunto de todos los sucesos posibles se denomina:
    a. Probabilidad Conjunta
    b. Espacio Muestral
    c. Conjunto vacío
    d. Probabilidad

SOLUCIÓN:

(b) Espacio Muestral

3. La expresión $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ se cumple sólo cuando:
    a. $P(A \cap B) = 0$
    b. $A \subset B$
    c. $P(A \cap B) \neq 0$
    d. Se cumple en todos los casos
    e. Ninguna es cierta

SOLUCIÓN:

(d) Se cumple en todos los casos

4. El mal funcionamiento de ciertos dispositivos electrónicos de iluminación ornamental puede deberse a su circuito impreso o a su grupo de LEDs.

    Una fábrica produce un 6% de dispositivos con mal funcionamiento. Si hay irregularidades en el circuito impreso, la probabilidad de que el dispositivo funcione mal es del 68%. Un 8% de los circuitos presentan irregularidades. Se ha observado también que un 3% de los grupos de LEDs son defectuosos, y que la probabilidad de que el dispositivo presente mal funcionamiento o sea defectuoso el grupo de LEDs, es del 8%.

    Se pide:
    a) Probabilidad de que el dispositivo tenga un mal funcionamiento y el circuito impreso presente irregularidades.

b) Probabilidad de que el dispositivo tenga mal funcionamiento si el grupo de LEDs es defectuoso.
c) Se toma un dispositivo al azar y resulta que tiene mal funcionamiento. Calcule la probabilidad de que su circuito impreso presente irregularidades.

SOLUCIÓN:

Denotemos:
D: dispositivo con mal funcionamiento
C: irregularidades en el circuito impreso
L: LEDs defectuosos

En el problema nos dan la siguiente información:
$P(D) = 0.06$
$P(D|C) = 0.68$
$P(C) = 0.08$
$P(L) = 0.03$
$P(D \cup L) = 0.08$

a) Nos piden:
$$P(D \cap C) = P(C)P(D|C) = 0.08 * 0.68 = 0.0544$$
b) Nos piden:
$$P(D|L) = \frac{P(D \cap L)}{P(L)}$$

Nos falta hallar la $P(D \cap L)$ que podemos calcular usando la fórmula de la probabilidad de la unión, ya que:
$$P(D \cup L) = P(D) + P(L) - P(D \cap L)$$

Entonces, sustituyendo los datos que tenemos:
$$0.08 = 0.06 + 0.03 - P(D \cap L)$$
$$P(D \cap L) = 0.01$$

Con lo cual, la condicional de arriba es:
$$P(D|L) = \frac{P(D \cap L)}{P(L)} = \frac{0.01}{0.03} = 0.3333$$

c) Usando el Teorema de Bayes:
$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} = \frac{0.68 * 0.08}{0.06} = 0.906$$

5. Supongamos que el tiempo en horas que un estudiante dedica cada semana a estudiar estadística se distribuye según una variable aleatoria con función de densidad dada por la siguiente función:

$$f(x) = ke^{-0.25x} \qquad ,x > 0$$

a) ¿Cuál debe ser el valor de k para que $f(x)$ sea una función de densidad?
b) ¿Cuál es la probabilidad de que a lo largo de una semana el estudiante dedique más de 5 horas a estudiar estadística?
c) Si a lo largo de 8 semanas el estudiante anota el tiempo en horas que dedica semanalmente a estudiar estadística, ¿cuál es la probabilidad de que en exactamente 2 de esas 8 anotaciones indique que ha estudiado estadística durante más de 5 horas semanales?

SOLUCIÓN:

a) Para que $f(x)$ sea una función de densidad, se tiene que cumplir la propiedad de que la integral de menos infinito a infinito de esa función es igual a 1:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Como $f(x)$ solo está definida para los valores $x > 0$, el rango de definición de la integral se puede actualizar:

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{+\infty} ke^{-0.25x}dx = k\left(-\frac{1}{0.25}\right)e^{-0.25x}|_0^\infty$$

Teniendo en cuenta que $e^{-\infty}$ tiende a cero, y $e^0 = 1$, queda:

$$k = 0.25$$

b) Denotemos por X el tiempo en horas que el estudiante dedica cada semana a estudiar estadística.
Nos piden:

$$P(X > 5) = \int_{5}^{+\infty} 0.25e^{-0.25x}dx = e^{-0.25x}|_5^\infty = e^{-0.25\times 5} = 0.2865$$

c) Sea Y la variable aleatoria que cuenta el número de semanas, de un total de 8, en las que el estudiante dedica más de 5 horas semanales a estudiar estadística. Sabemos que Y sigue una distribución Binomial con parámetros $n = 8$ (semanas) y $p = P(X > 5) = 0.2865$. Nos piden $P(Y = 2)$. Utilizando la función de probabilidad de una v.a. $Binomial(n = 8, p = 0.2865)$, se obtiene:

$$P(Y = 2) = \binom{8}{2}(0.2865)^2\,(1 - 0.2865)^6 = \frac{8!}{2!\,6!}(0.2865)^2\,(1 - 0.2865)^6$$
$$= 0.3032$$