EXAMEN DE ESTADÍSTICA (DESCRIPTIVA Y REGRESIÓN)

2º Fisioterapia Modelo A **19 de junio de 2020**

Duración: 1 hora.

(5 pts.) 1. Para ver si la situación de confinamiento debida al COVID19 ha influído en el rendimiento de un curso, se ha contado el número de asignaturas suspensas de cada alumno en el curso actual y en el curso anterior, obteniendo la siguiente tabla:

Asignaturas suspensas	Curso anterior	Curso actual
0	7	8
1	15	12
2	11	8
3	5	7
4	4	3
5	2	2
6	1	2
8	0	1

Se pide:

a) Dibujar los diagramas de cajas del número de asignaturas suspensas en el curso actual y en el anterior y compararlos.

b) ¿Se puede asumir que ambas muestras, la del curso actual y la del anterior, provienen de poblaciones normales?

c) ¿En qué muestra es más representativa la media?

d) ¿Qué número de asignaturas suspensas es mayor, 7 asignaturas en el curso actual, o 6 en el curso anterior?

Usar las siguientes sumas para los cálculos:
Curso anterior: $\sum x_i n_i = 84$, $\sum x_i^2 n_i = 254$, $\sum (x_i - \bar{x})^3 n_i = 122{,}99$ y $\sum (x_i - \bar{x})^4 n_i = 669{,}21$.
Curso actual: $\sum y_i n_i = 91$, $\sum y_i^2 n_i = 341$, $\sum (y_i - \bar{y})^3 n_i = 301{,}16$ y $\sum (y_i - \bar{y})^4 n_i = 2012{,}88$.

Solución

a) Han aprobado el 66.5 % de los estudiantes no trabajadores y el 59 % de los trabajadores.

b) No trabajadores: $\bar{x} = 1{,}8667$, $s^2 = 2{,}16$, $s = 1{,}4697$ y $cv = 0{,}7873$.
Trabajadores: $\bar{y} = 2{,}1163$, $s^2 = 3{,}4516$, $s = 1{,}8578$ y $cv = 0{,}8779$.
La muestra de los alumnos que no trabajan tiene una dispersión relativa con respecto a la media ligeramente mayor ya que su coeficiente de variación es mayor.

c) No trabajadores: $g_1 = 0{,}8609$.
Trabajadores: $g_1 = 1{,}0922$.
Así pues, la muestra de los alumnos que no trabajan es más asimétrica ya que su coeficiente de asimetría está más lejos de 0.

d) No trabajadores: $\bar{y} = 3{,}2067$.
Trabajadores: $\bar{x} = 3{,}5686$.
El coeficiente de asimetría no cambia al ser la pendiente de la transformación positiva.

e) No trabajadores: $z(7) = 3{,}4928$.
Trabajadores: $z(6) = 2{,}0904$.
Así pues, un 7 en la muestra de alumnos que no trabajan es relativamente mayor que un 6 en la muestra de alumnos que trabajan.

(5 pts.) 2. Un estudio intenta poner a punto una nueva técnica de detección de un cierto anticuerpo. Para ello se utiliza un inmunosensor piezoeléctrico, que permite medir el cambio en la señal en Hz al variar la concentracióndel anticuerpo (μg/ml). Se recolectaron los siguientes datos:

Concentración (μg/ml)	5	8	20	35	50	80	110
Señal (Hz)	50	70	100	150	170	190	200

Se pide:

a) Construir el modelo logarítmico del cambio de la señal sobre la concentración de los anticuerpos.

b) Se vio que con una concentración de 100μg/ml el cambio en la señaltiende a estabilizarse. Predecir el valor de la señal correspondiente a tal concentración haciendo uso del modelo logarítmico.

c) Usar el mejor modelo para predecir la concentración de anticuerpos para que haya un cambio en la señal de 120 Hz.

Usar las siguientes sumas para los cálculos (X=Concentración e Y=Señal):
$\sum x_i = 308$ Hz, $\sum \log(x_i) = 23{,}2345$ log(Hz), $\sum y_j = 930$ μg/ml, $\sum \log(y_j) = 33{,}4575$ log(μg/ml), $\sum x_i^2 = 22714$ Hz^2, $\sum \log(x_i)^2 = 85{,}1299$ log(Hz)^2, $\sum y_j^2 = 144900$ μg/ml^2, $\sum \log(y_j)^2 = 161{,}6475$ log(μg/ml)^2,
$\sum x_i y_j = 53760$ Hz$\cdot\mu$g/ml, $\sum x_i \log(y_j) = 1580{,}3905$ Hz$\cdot$ log(μg/ml), $\sum \log(x_i) y_j = 3496{,}6333$ log(Hz)μg/ml,
$\sum \log(x_i)\log(y_j) = 114{,}7297$ log(Hz) log(μg/ml).

Solución

a) $\bar{x} = 44$ Hz, $s_x^2 = 1308{,}8571$ Hz^2.
$\bar{y} = 132{,}8571$ μg/ml, $s_y^2 = 3048{,}9796$ 10^{-4} s^2.
$s_{xy} = 1834{,}2857$ Hz$\cdot 10^{-2}$ s.
$b_{yx} = 1{,}4014$ μg/ml/Hz.
Así pues, el tiempo de respuesta aumenta 1,4014 centésimas de segundo por cada Hz más que se incrementa la dosis.

b) $\overline{\log(y)} = 4{,}7796$ log(μg/ml), $s_{\log(y)}^2 = 0{,}2476$ log(μg/ml)^2.
$s_{x\log(y)} = 15{,}466$ Hz$\cdot$ log(10^{-2} s).
Modelo de regresión exponencial: $y = e^{4{,}2597+0{,}0118x}$.
Predicción: $y(75) = 171{,}734$ μg/ml.
Coeficiente de determinación exponencial: $r^2 = 0{,}7382$
Así pues, el modelo exponencial se ajusta muy bien a la nube de puntos del diagrama de dispersión pero la muestra es muy pequeña para considerar las predicciones fiables.

c) Modelo de regresión logarítmico: $x = -254{,}5891 + 62{,}4711\ln(y)$.
Predicción: $x(100) = 33{,}1008$ Hz.
