EXAMEN DE ESTADÍSTICA (DESCRIPTIVA Y REGRESIÓN)

2º Fisioterapia — Modelo A — **31 de mayo de 2018**

Duración: 1 hora y media.

(5 pts.) 1. Las edades de una muestra de pacientes que acuden a una clínica de fisioterapia son las siguientes:

25, 30, 44, 44, 51, 51, 53, 56, 57, 58, 58, 58, 59, 59, 61, 63, 63, 63, 66, 68, 70, 71, 72, 74, 82, 85

Se pide:

a) Calcular los cuartiles.

b) Dibujar el diagrama de cajas e identificar los datos atípicos (no agrupar los datos en intervalos).

c) Considerando los grupos de los menores y mayores de 65 años, ¿en cuál de ellos es más representativa la media?

d) ¿Qué distribución es menos simétrica, la de los menores o la de los mayores de 65 años?

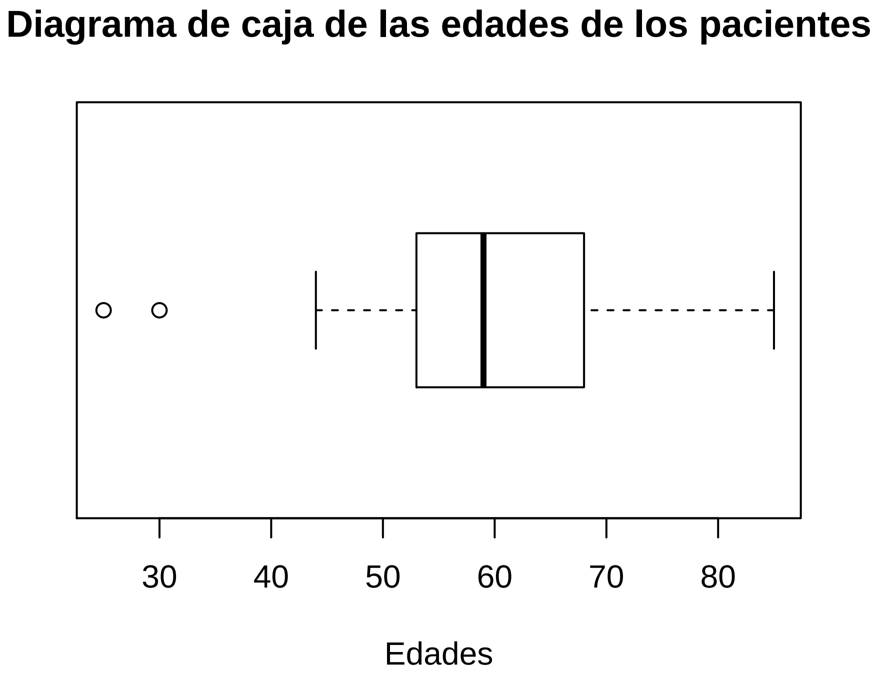
e) ¿Qué edad es relativamente mayor con respecto a su grupo, 50 años en el grupo de los menores o 75 en el de los mayores?

Usar las siguientes sumas para los cálculos.
Menores de 65: $\sum x_i = 953$ años, $\sum x_i^2 = 52475$ años^2, $\sum(x_i - \bar{x})^3 = -30846{,}51$ años^3 y $\sum(x_i - \bar{x})^4 = 939658{,}83$ años^4.
Mayores de 65: $\sum x_i = 588$ años, $\sum x_i^2 = 43530$ años^2, $\sum(x_i - \bar{x})^3 = 1485$ años^3 y $\sum(x_i - \bar{x})^4 = 26983{,}5$ años^4.

Solución

a) $C_1 = 53$ años, $C_2 = 59$ años and $C_3 = 68$ años.

b) Hay 2 datos atípicos: 25, 30.



c) Sea x las edades de los pacientes menores de 65 e y las de los mayores de 65.
$\bar{x} = 52{,}9444$ años, $s_x^2 = 112{,}1636$ años^2, $s_x = 10{,}5907$ años and $cv_x = 0{,}2$.
$\bar{y} = 73{,}5$ años, $s_y^2 = 39$ años^2, $s_y = 6{,}245$ años y $cv_y = 0{,}085$.
La media es más representativa en los pacientes mayores de 65 ya que su coeficiente de variación es menor.

d) $g_{1x} = -1{,}4426$ y $g_{1y} = 0{,}7621$, de manera que la distribución de edades de los pacientes menores de 65 es menos simétrica.

e) Las puntuaciones típicas son $z_x(50) = -0{,}278$ y $z_y(72) = -0{,}2402$, de manera que 50 años es relativamente menor en el grupo de los pacientes menores de 65.

(5 pts.) 2. La siguiente tabla recoge el número de lesiones en un equipo durante una temporada y el número medio de minutos diarios de calentamiento que hacen sus jugadores.

Tiempo calentamiento	15	35	22	28	21	18	25	30	23	20
Lesiones	42	2	16	6	17	29	10	3	12	20

Se pide:

a) Dibujar el diagrama de dispersión.

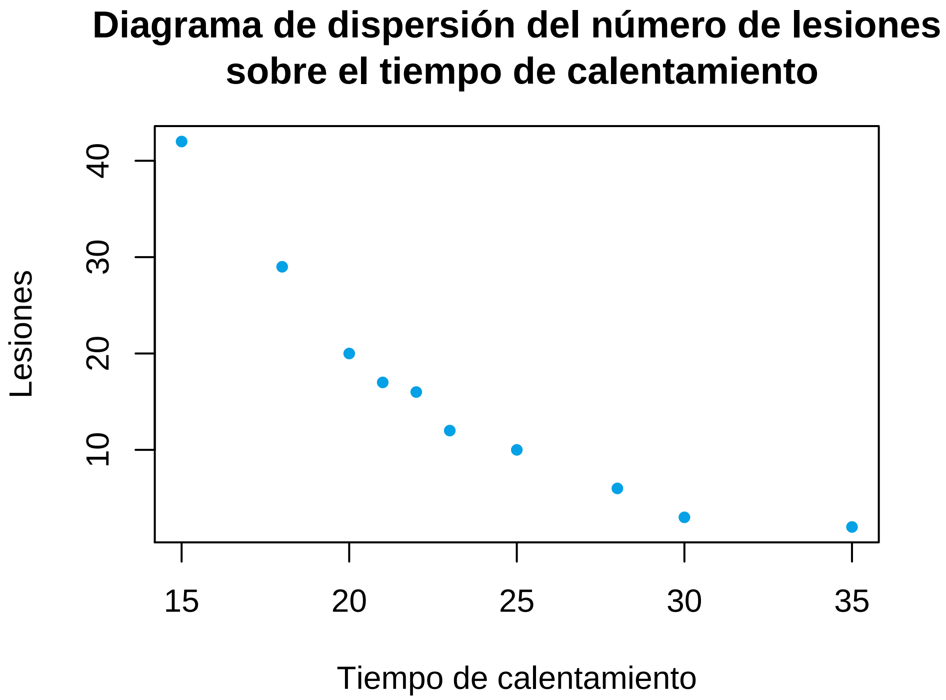
b) ¿Qué modelo de regresión es más apropiado para predecir el número de lesiones en función del tiempo de calentamiento, el logarítmico o el exponencial? Utilizar dicho modelo para predecir el número de lesiones esperado para 20 minutos de calentamiento diarios.

c) ¿Qué modelo de regresión es más apropiado para predecir el tiempo de calentamiento en función del número de lesiones, el logarítmico o el exponencial? Utilizar dicho modelo para predecir el mínimo tiempo de calentamiento diario necesario para no tener más de 10 lesiones en la temporada.

d) ¿Son fiables estas predicciones? ¿Cuál de ellas es más fiable?

Usar las siguientes sumas para los cálculos (X tiempo de calentamiento e Y número de lesiones): $\sum x_i = 237$, $\sum \log(x_i) = 31{,}3728$, $\sum y_j = 157$, $\sum \log(y_j) = 24{,}0775$, $\sum x_i^2 = 5937$, $\sum \log(x_i)^2 = 98{,}9906$, $\sum y_j^2 = 3843$, $\sum \log(y_j)^2 = 66{,}3721$, $\sum x_i y_j = 3115$, $\sum x_i \log(y_j) = 519{,}1907$, $\sum \log(x_i) y_j = 465{,}8093$, $\sum \log(x_i)\log(y_j) = 73{,}3995$.

Solución



a)

b) $\bar{x} = 23{,}7$ min, $s_x^2 = 32{,}01$ min^2.
$\overline{\log(x)} = 3{,}1373$ log(min), $s^2_{\log(x)} = 0{,}0565$ log(min)^2.
$\bar{y} = 15{,}7$ lesiones, $s_y^2 = 137{,}81$ lesiones^2.
$\overline{\log(y)} = 2{,}4078$ log(lesiones), $s^2_{\log(y)} = 0{,}8399$ log(lesiones)^2.
$s_{x\log(y)} = -5{,}1446$, $s_{\log(x)y} = -2{,}6744$
Coeficiente de determinanción exponencial: $r^2 = 0{,}9844$
Coeficiente de determinación logarítmico: $r^2 = 0{,}9185$
Por tanto, el modelo de regresión exponencial es mejor para predecir el número de lesiones en función del tiempo de calentamiento.
Modelo de regresión exponencial: $y = e^{6{,}2168+-0{,}1607x}$.
Predicción: $y(20) = 20{,}1341$ lesiones.

c) El modelo logarítmico es mejor para predecir el tiempo de calentamiento en función del número de lesiones.
Modelo de regresión logarítmico: $x = 164{,}1851 + -47{,}3292\log(y)$.
Predicción: $x(10) = 55{,}2056$ min.

d) De acuerdo al coeficiente de determinación ambas predicciones son muy fiables pero la última lo es menos ya que es para un valor que no está incluido en el rango de valores de la muestra.