EXAMEN DE ESTADÍSTICA (DESCRIPTIVA Y REGRESIÓN)

1º Óptica Modelo A **15 de junio de 2018**

Duración: 1 hora.

(5 pts.) 1. Una persona se sometió a un programa de ejercicios para disminuir el índice de masa corporal (IMC). A continuación se indican el número de semanas que llevaba haciendo ejercicio y el IMC correspondiente.

Tiempo ejercicio	4,0	8	12,0	16	20,0	24,0
IMC	25,4	24	22,8	22	21,6	21,4

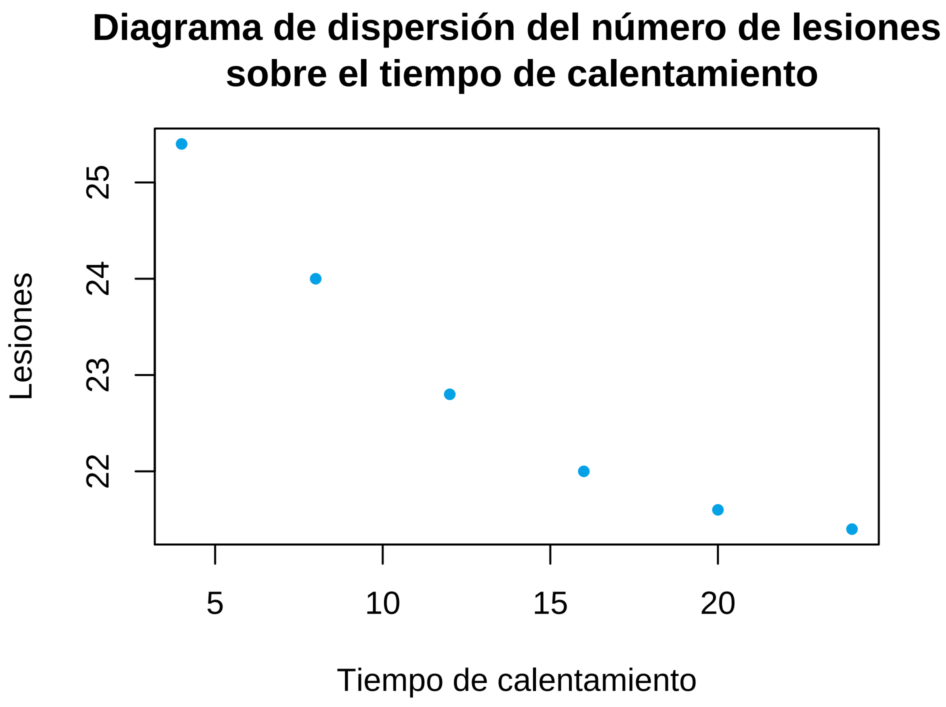
Se pide:

a) Calcular el IMC al cabo de 30 semanas de hacer ejercicio empleando los modelos de regresión lineal, logarítmico y exponencial.

b) ¿Cuál de las tres predicciones es más fiable? Razonar la respuesta.

Usar las siguientes sumas para los cálculos (X tiempo de ejercicio e Y IMC):
$\sum x_i = 84$, $\sum \log(x_i) = 14{,}897$, $\sum y_j = 137{,}2$, $\sum \log(y_j) = 18{,}7667$,
$\sum x_i^2 = 1456$, $\sum \log(x_i)^2 = 39{,}1823$, $\sum y_j^2 = 3149{,}52$, $\sum \log(y_j)^2 = 58{,}7206$,
$\sum x_i y_j = 1864{,}8$, $\sum x_i \log(y_j) = 260{,}3165$, $\sum \log(x_i) y_j = 335{,}4895$, $\sum \log(x_i)\log(y_j) = 46{,}3734$.

Solución



a)

b) $\bar{x} = 14$ min, $s_x^2 = 46{,}6667$ min^2.
$\overline{\log(x)} = 2{,}4828$ log(min), $s_{\log(x)}^2 = 0{,}3659$ log(min)^2.
$\bar{y} = 22{,}8667$ lesiones, $s_y^2 = 2{,}0356$ lesiones^2.
$\overline{\log(y)} = 3{,}1278$ log(lesiones), $s_{\log(y)}^2 = 0{,}0038$ log(lesiones)^2.
$s_{x\log(y)} = -0{,}4029$, $s_{\log(x)y} = -0{,}8593$
Coeficiente de determinanción exponencial: $r^2 = 0{,}9274$
Coeficiente de determinación logarítmico: $r^2 = 0{,}9913$
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: $y = e^{3{,}2486+-0{,}0086x}$.
Predicción: $y(20) = 21{,}6712$ lesiones.

c) El modelo logarítmico es mejor para predecir el tiempo de calentamiento en función del número de lesiones.
Modelo de regresión logarítmico: $x = 28{,}6971 + -2{,}3483\log(y)$.
Predicción: $x(10) = 23{,}29$ 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.

(5 pts.) 2. Se ha medido el tiempo de espera para que un medicamento A haga efecto en un grupo de 150 pacientes, obteniéndose los siguientes resultados:

Tiempo	Pacientes
$(0, 5]$	5
$(5, 10]$	15
$(10, 15]$	32
$(15, 20]$	36
$(20, 30]$	42
$(30, 60]$	20

Se pide:

a) ¿Existen datos atípicos en la muestra? Justificar la respuesta.

b) ¿Cuál fue el tiempo de espera mínimo para el 20 % de los pacientes que más esperaron?

c) ¿Cuál fue el tiempo medio de espera? ¿Es muy representativa esta medida? Justificar la respuesta

d) Se ha propuesto utilizar otro medicamento B cuyos tiempos de espera son Y=5+0.5X. (Y= Tiempo de espera con el medicamento B, X= Tiempo de espera con el medicamento A) ¿Cuál será el tiempo medio de espera del medicamento B? ¿Es una medida más o menos representativa que el del medicamento A? Justificar la respuesta.

e) Si tomamos otro grupo de pacientes tratados con el medicamento A cuya media fue 18 min con una desviación típica de 15 min, ¿en cuál de los dos grupos sería más alto un tiempo de espera de 25 min?

Usar las siguientes sumas para los cálculos: $\sum x_i = 3105$ min, $\sum x_i^2 = 83650$ min^2, $\sum(x_i - \bar{x})^3 = 206851{,}65$ min^3 y $\sum(x_i - \bar{x})^4 = 8140374{,}96$ min^4.

Solución
