

Exámenes de Bioestadística



Alfredo Sánchez Alberca
asalber@ceu.es
<https://aprendeconalf.es>

Tabla de contenidos

Prefacio	3
I. Farmacia y Biotecnología	4
1. 2013-01-24 Examen de Farmacia y Biotecnología	5
2. 2014-01-27 Examen de Farmacia y Biotecnología	7
3. 2014-11-24 Examen de Farmacia y Biotecnología	9
4. 2015-01-13 Examen de Farmacia y Biotecnología	13
5. 2016-11-28 Examen de Farmacia y Biotecnología	15
6. 2017-01-10 Examen de Farmacia y Biotecnología	18
7. 2017-11-27 Examen de Farmacia y Biotecnología	24
8. 2018-01-19 Examen de Farmacia y Biotecnología	27
9. 2018-12-17 Examen de Farmacia y Biotecnología	31
10.2019-12-16 Examen de Farmacia y Biotecnología	37
11.2020-10-26 Examen de Farmacia y Biotecnología	42
12.2021-01-18 Examen de Farmacia y Biotecnología	47
II. Fisioterapia	53
13.2018-04-09 Examen de Fisioterapia	54
14.2018-05-31 Examen de Fisioterapia	58
15.2018-06-18 Examen de Fisioterapia	64

III. Óptica	70
16.2017-12-18 Examen de Óptica	71
17.2018-10-24 Examen de Óptica	75
18.2020-11-13 Examen de Óptica	79

Prefacio

Colección de exámenes resueltos de Bioestadística de los grados en Farmacia, Biotecnología, Fisioterapia y Óptica y Optometría de la Universidad CEU San Pablo.

Los exámenes están agrupados por titulación y ordenados cronológicamente. Cada ejercicio va acompañado de su solución, que puede desplegarse pulsando sobre el bloque *Solución*. En algunos ejercicios la solución se ha grabado en vídeo y aparece en el bloque *Resolución*.

Parte I.

Farmacia y Biotecnología

1. 2013-01-24 Examen de Farmacia y Biotecnología

Grados: Farmacia y Biotecnología

Fecha: 24 de Enero de 2013

Ejercicio 1.1. En un grupo de personas sometidas a una anestesia general se ha medido la dosis de sustancia anestésica recibida X en mg y el tiempo que estuvieron dormidas Y en horas. Las frecuencias observadas aparecen en la siguiente tabla:

$X \backslash Y$	$[1, 2)$	$[2, 3)$	$[3, 4)$	n_x
$(20, 30]$	14	10	0	24
$(30, 40]$	12	26	7	45
$(40, 50]$	2	12	17	31
n_y	28	48	24	100

Se pide:

1. ¿En qué variable es más representativa la media? Justificar la respuesta
2. ¿Por encima de cuánto tiempo estarán dormidas el 10% de las personas que reciben una dosis entre 30 y 40 mg?
3. ¿En qué variable hay más asimetría? Justificar la respuesta.
4. Según el modelo de regresión lineal, ¿cuánta sustancia anestésica será necesaria para dormir a alguien durante al menos dos horas? ¿Es fiable la predicción? Justificar la respuesta.

Resolución

<https://www.youtube.com/watch?v=q5j2ryj0oCQ>

Ejercicio 1.2. Se sometió a una persona a unas sesiones de entrenamiento para el manejo de una máquina de análisis químicos y se valoró la destreza en el manejo en diversas ocasiones, valorandola en una escala de 0 a 100. Los resultados obtenidos aparecen en la siguiente tabla

Sesiones	2	5	7	10	12	16
Destreza	15	40	62	86	92	95

Se pide:

1. Calcular la destreza alcanzada al cabo de 8 sesiones empleando el modelo logarítmico.
2. Calcular el número de sesiones necesarias para alcanzar una destreza de 80 empleando el modelo exponencial.
3. Justificar razonadamente cuál de las predicciones anteriores es más fiable.

Resolución

<https://www.youtube.com/watch?v=Jx8R4fTFjoE>

Ejercicio 1.3. Al realizar un estudio de regresión lineal de dos variables X e Y , se sabe que las rectas de regresión se cortan en el punto $(5, 15)$, que el coeficiente de correlación lineal es -0.85 y que la pendiente de la recta de regresión de X sobre Y es el doble que la de la recta de Y sobre X . Se pide:

1. Calcular las ecuaciones de las rectas de regresión de Y sobre X y de X sobre Y .
2. ¿Qué porcentaje de la variabilidad de Y queda explicado por el modelo lineal?

Resolución

<https://www.youtube.com/watch?v=XKrRifxAfDg>

2. 2014-01-27 Examen de Farmacia y Biotecnología

Grados: Farmacia y Biotecnología

Fecha: 27 de Enero de 2014

Ejercicio 2.1. Se realiza un estudio para determinar la efectividad de un medicamento para controlar la hipertensión a 180 pacientes hipertensos. Para ello se les suministra una cantidad determinada del mismo obteniéndose los siguientes resultados

Dosis (mg)	n_i	f_i	N_i	F_i
(100, 400]	15			0.2167
(400, 700]				
(700, 800]	36			
(800, 900]		0.3333		
(900, 1000]				

Se pide:

1. Completar la tabla.
2. ¿Cuál ha sido la dosis media de medicamento administrado? ¿Es representativa?
3. Si se considera que a partir de una administración de 725mg hay que hacer un seguimiento para controlar posibles hipotensiones, ¿qué porcentaje de pacientes necesitan ese seguimiento?
4. ¿Cuál fue la cantidad mínima de medicamento suministrada al 60% de los pacientes más medicados?
5. ¿Cuál fue la cantidad más habitual de medicamento administrado?
6. Si a un paciente se le suministró una cantidad de medicamento de 725 mg y a otro una cantidad tipificada de 0.95, ¿A cuál se le administró una cantidad mayor? Justificar la respuesta.
7. Calcular el coeficiente de asimetría e interpretarlo.
8. Dibujar el diagrama de cajas e interpretarlo.
9. Si se cambia de medicamento por otro cuya cantidad a administrar viene dada en función del anterior medicamento mediante la relación $Y = 100 + 0.7X$, siendo X la cantidad de medicamento original e Y la cantidad de medicamento nuevo, ¿cuál será la media de la cantidad administrada del nuevo medicamento? ¿Es más representativa que en el medicamento original? Justificar la respuesta.

Nota: Para facilitar los cálculos se dan las siguientes sumas:

$$\begin{array}{ll} \sum x_i n_i = 137700 & \sum x_i^2 n_i = 112410000 \\ \sum (x_i - \bar{x})^3 n_i = -1965735000 & \sum (x_i - \bar{x})^4 n_i = 1162291162500 \end{array}$$

Ejercicio 2.2. En un estudio se ha medido el calor liberado en una reacción química en distintos instantes desde el comienzo de la reacción, obteniendo los siguientes datos:

Tiempo en minutos	2.5	3.7	4.1	5.3	6.2
Calor en calorías	15.9	44.5	65.6	206.5	498.7

Se pide:

1. Calcular el modelo de regresión lineal del calor sobre el tiempo. Según este modelo ¿Cuándo cambiarán las calorías por cada minuto que pase?
2. Calcular el modelo de regresión exponencial del calor sobre el tiempo.
3. Utilizando el mejor de los dos modelos anteriores, predecir el calor generado a los 5 minutos de la reacción. ¿Es fiable la predicción? Justificar la respuesta.

 Resolución

<https://www.youtube.com/watch?v=IbyZwmlcv8Y>

3. 2014-11-24 Examen de Farmacia y Biotecnología

Grados: Farmacia y Biotecnología

Fecha: 24 de Noviembre de 2014

Ejercicio 3.1. El siguiente diagrama refleja la distribución acumulada del número de partes de asistencia médica que han dado durante un año una muestra de 100 asegurados.

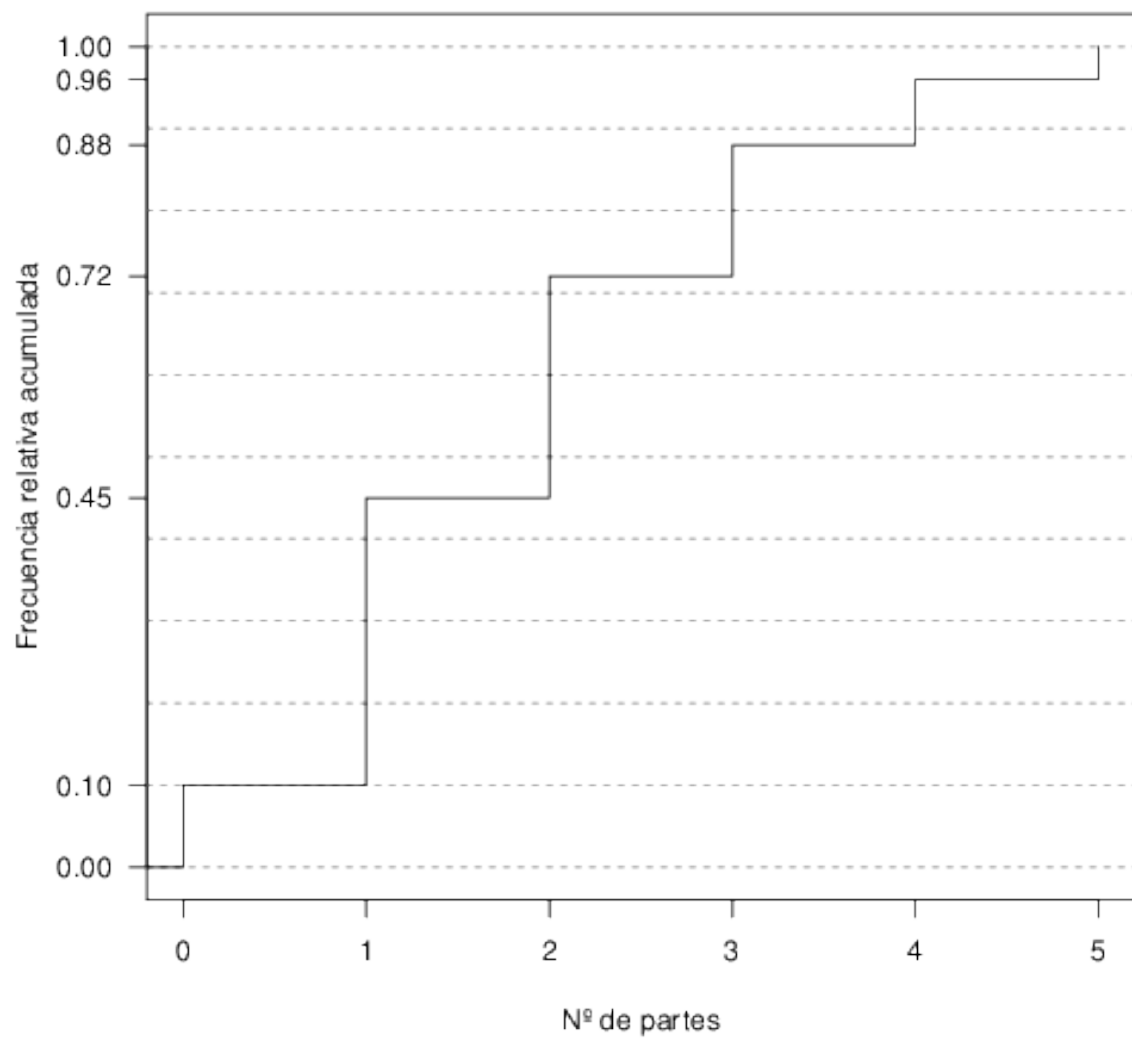


Figura 3.1.: Polígono de frecuencias relativas acumuladas

Dibujar el diagrama de cajas y ver si existen datos atípicos en la muestra.

💡 Resolución

<https://www.youtube.com/watch?v=-de4qdVapDY>

Ejercicio 3.2. En un centro de salud, la ocupación del aparcamiento durante la última semana fue registrada mediante la siguiente tabla:

x	n_i	$x_i n_i$	$x_i^2 n_i$	$(x_i - \bar{x})^3 n_i$	$(x_i - \bar{x})^4 n_i$
(0, 40]	84	1680	33600	-12155062.50	638140781.25
(40, 80]	185	11100	666000	-361328.13	4516601.56
(80, 120]	72	7200	720000	1497375.00	41177812.50
(120, 160]	40	5600	784000	12301875.00	830376562.50
(160, 200]	19	3420	615600	23603640.63	2537391367.19
$\sum$	400	29000	2819200	24886500.00	4051603125.00

Se pide:

1. Calcular el tiempo por encima del cual han estacionado el 60% de los vehículos. ¿Qué porcentaje de vehículos ha estacionado por encima de 100 minutos?
2. Calcular la media del tiempo de estacionamiento. ¿Es representativa la media?
3. Calcular el coeficiente de asimetría y de apuntamiento de la distribución del tiempo de estacionamiento e interpretarlos.
4. Sabiendo que la empresa responsable del aparcamiento cobra 3 céntimos por minuto de aparcamiento más una cantidad fija de 20 céntimos, calcular el precio medio del aparcamiento y la desviación típica del mismo.

Resolución

<https://www.youtube.com/watch?v=8XcjtIqB83U>

Ejercicio 3.3. Una botánica comenzó a tratar una planta de soja con cierta sustancia estimulante del crecimiento. Para ver cómo evolucionaba el peso en función del tiempo, hizo 8 medidas del peso en distintos instantes, obteniendo los siguientes resultados

$$\begin{array}{llll}
 \sum t = 250 & \sum p = 2369 & \sum \ln(t) = 25.94 & \sum \ln(p) = 27.17 \\
 \sum t^2 = 10358 & \sum p^2 = 2848799 & \sum (\ln t)^2 = 87.70 & \sum (\ln p)^2 = 139.52 \\
 \sum tp = 133399 & \sum \ln(t)p = 9527.34 & \sum t \ln(p) = 1193.26 & \sum \ln(t) \ln(p) = 100.60
 \end{array}$$

Se pide:

1. Estudiar si el modelo logarítmico es más fiable que el lineal para predecir los días de tratamiento en función el peso de la planta. Utilizar el mejor modelo para estimar los días de tratamiento necesarios para que la planta pese 500 gr.
2. Considerando el resultado del apartado anterior, ¿cuál sería el mejor modelo de regresión para estimar el peso de la planta a los 30 días de tratamiento? Hacer dicha estimación.

 Resolución

<https://www.youtube.com/watch?v=T4hu8kOIF28>

Ejercicio 3.4. La recta de regresión de una variable Y sobre otra X tiene ecuación $y + \frac{1}{4}x - 1 = 0$, mientras que la recta de regresión de X sobre Y , tiene ecuación $y + \frac{3}{4}x - 2 = 0$. Se pide:

1. Calcular la media de X y la de Y .
2. Calcular el coeficiente de correlación lineal e interpretarlo.

 Resolución

<https://www.youtube.com/watch?v=z-CxzqBf9t4>

4. 2015-01-13 Examen de Farmacia y Biotecnología

Grados: Farmacia y Biotecnología

Fecha: 13 de Enero de 2015

Ejercicio 4.1. Un laboratorio farmacéutico está probando un nuevo medicamento y se están estudiando los efectos secundarios que provoca. Se sabe que la probabilidad de que el medicamento provoque mareos es de 0.20 y de que provoque náuseas es de 0.50. También se sabe que la probabilidad de que provoque solo mareos es 0.15.

1. ¿Que probabilidad hay de que el medicamento provoque algún efecto secundario?
2. ¿Que probabilidad hay de que el medicamento provoque los dos efectos secundarios?
3. ¿Que probabilidad hay de que el medicamento provoque solo náuseas?
4. ¿Que probabilidad hay de que el medicamento provoque náuseas si ha provocado mareos?
5. ¿Son independientes los efectos secundarios?

Resolución

<https://www.youtube.com/watch?v=5On725uKqBU>

Ejercicio 4.2. Se sabe que en personas con infección urinaria el número medio de bacterias por mm^3 de orina es 5, mientras que en personas sanas la media es de 2 bacterias por mm^3 . Se pide:

1. Calcular la probabilidad de que en una muestra de medio mm^3 de orina de un individuo con infección haya alguna bacteria.
2. Calcular la probabilidad de que en una muestra de dos mm^3 de orina de un individuo sano haya menos de 3 bacterias.
3. Si un test diagnóstico para detectar la infección urinaria da positivo cuando en un mm^3 de orina hay más de 6 bacterias, ¿cuál es la sensibilidad del test? ¿Y cuál es su especificidad?
4. Si la prevalencia de la infección urinaria en la población es del 5%, ¿cuál es el valor predictivo positivo del test diagnóstico del apartado anterior? ¿Y su valor predictivo negativo?

5. Si se toman 5 muestras de un mm^3 de una persona con infección de orina, ¿cuál es la probabilidad de que se produzca algún falso negativo?

Nota: La sensibilidad de un test es el porcentaje de resultados positivos en personas enfermas, la especificidad es el porcentaje de resultados negativos en personas sanas, la prevalencia es el porcentaje de personas enfermas en la población, el valor predictivo positivo es el porcentaje de personas enfermas entre las que han dado positivo en el test, y el valor predictivo negativo es el porcentaje de personas sanas entre las que han dado negativo en el test.

 Resolución

<https://www.youtube.com/watch?v=FmKjZjTREYU>

Ejercicio 4.3. En un estudio realizado a 720 niños de 8 años se observó que 324 tenían un peso superior a 27.2 kg y que 216 tenían un peso entre 24.6 y 27.2 kg. Suponiendo que el peso de los niños de 8 años sigue una distribución normal, se pide:

1. Calcular la media y la desviación típica del peso de los niños de 8 años.
2. ¿Cuántos niños de 8 años tendrán un peso comprendido entre 24 y 28 kg?
3. Si un niño de 8 años pesa 28.5 kg, ¿cuánto debe adelgazar para situarse por debajo del percentil 60 del peso?

 Resolución

<https://www.youtube.com/watch?v=wgIbPz3wTbk>

5. 2016-11-28 Examen de Farmacia y Biotecnología

Grados: Farmacia y Biotecnología Fecha: 28 de Noviembre de 2016

Ejercicio 5.1. La siguiente tabla contiene la distribución de las puntuaciones obtenidas por una muestra de alumnos de medicina que se presentaron al examen del MIR.

x	n_i	$x_i n_i$	$x_i^2 n_i$	$(x_i - \bar{x})^3 n_i$	$(x_i - \bar{x})^4 n_i$
(0, 40]	84	1680	33600	-12155062.50	638140781.25
(40, 80]	185	11100	666000	-361328.13	4516601.56
(80, 120]	72	7200	720000	1497375.00	41177812.50
(120, 160]	40	5600	784000	12301875.00	830376562.50
(160, 200]	19	3420	615600	23603640.63	2537391367.19
Σ	400	29000	2819200	24886500.00	4051603125.00

1. Calcular el rango intercuartílico de las puntuaciones e interpretarlo. ¿Hay datos atípicos en la muestra?
2. Si la nota de corte para aprobar el examen es 150, ¿qué porcentaje de alumnos aprobó el examen?
3. Estudiar la representatividad de la media.
4. Según la asimetría y el apuntamiento de la muestra, ¿se puede suponer que proviene de una población normal? Justificarlo con el cálculo de estadísticos de forma.
5. Calcular la puntuación típica que le correspondería a un alumno con una puntuación de 150 puntos.

Solución

1. $C_1 = 43.48$ puntos, $C_3 = 97.78$ puntos y $RI = 54.3$ puntos.
Vallas: $V_1 = -37.97$ puntos y $V_2 = 179.23$ puntos. Por tanto, existen datos atípicos.
2. $F_{150} = 0.925$, de manera que el porcentaje de estudiantes que aprobaron el examen fue 7.5%.
3. $\bar{x} = 72.5$ puntos, $s^2 = 1791.75$ puntos², $s = 42.3291$ puntos, $cv = 0.5838$.
Como el coeficiente de variación es mayor que 0.5 pero no demasiado, existe

una variabilidad moderada y la representatividad de la media es también moderada.

4. $g_1 = 0.8203$, de manera que la distribución es asimétrica hacia la izquierda. $g_2 = 0.1551$, de manera que la distribución es un poco más apuntada de lo normal (leptocúrtica). Como tanto g_1 como g_2 están entre -2 y 2 podemos asumir que la muestra proviene de una población normal.
5. $z(150) = 1.83$.

Ejercicio 5.2. La siguiente tabla refleja la evolución del Producto Interior Bruto (PIB) per capita (en miles de euros) y la mortalidad infantil (niños por cada mil habitantes) de una serie de años.

Año	1993	1994	1995	1996	1997	1998	1999	2000
PIB	17	17	18	18	19	20	21	22
Mortalidad	6	5.6	5.2	4.9	4.6	4.3	4.1	4

1. Estimar el PIB correspondiente a una mortalidad de 3.8 según el modelo lineal. ¿Qué modelo de regresión expresa mejor la relación entre el PIB y la mortalidad, el lineal o el exponencial?
2. Si el PIB del año 2001 fue de 23, ¿cuál es la mortalidad esperada ese año según el modelo de regresión exponencial?
3. Considerando los modelos lineales del PIB sobre la mortalidad y de la mortalidad sobre el PIB, ¿cuál de los dos modelos es más fiable?

Utilizar las siguientes sumas ($X=\text{PIB}$ y $Y=\text{Mortalidad}$) para los cálculos: $\sum x_i = 152$, $\sum \log(x_i) = 23.5229$, $\sum y_j = 38.7$, $\sum \log(y_j) = 12.5344$, $\sum x_i^2 = 2912$, $\sum \log(x_i)^2 = 69.2305$, $\sum y_j^2 = 190.87$, $\sum \log(y_j)^2 = 19.7912$, $\sum x_i y_j = 726.5$, $\sum x_i \log(y_j) = 236.3256$, $\sum \log(x_i) y_j = 113.3308$, $\sum \log(x_i) \log(y_j) = 36.76$.

Solución

1. Modelo de regresión lineal del PIB sobre la mortalidad infantil: $\bar{x} = 19 \cdot 10^3 \text{€}$, $s_x^2 = 3 \cdot 10^6 \text{€}^2$.
 $\bar{y} = 4.8375$ niños por cada mil, $s_y^2 = 0.4573$ (niños por cada mil) 2 .
 $s_{xy} = -1.1 \cdot 10^3 \text{€}$ niños por cada mil.
Recta de regresión del PIB sobre la mortalidad infantil: $x = 30.6351 + -2.4052y$.
 $x(3.8) = 21.4954$.
2. $\log(x) = 2.9404 \log(10^3 \text{€})$, $s_{\log(x)}^2 = 0.0081 \log(10^3 \text{€})^2$.
 $s_{\log(x)y} = -0.0577 \log(10^3 \text{€}) \cdot \text{niños por cada mil}$.
Coeficiente de determinación lineal del PIB sobre la mortalidad infantil $r^2 = 0.8819$.

Coefficiente de determinación exponencial del PIB sobre la mortalidad infantil $r^2 = 0.9002$.

Por tanto, el modelo exponencial explica mejor la relación entre el PIB y la mortalidad infantil ya que su coeficiente de determinación es mayor.

3. $\log(\bar{y}) = 1.5668 \log(\text{niños por cada mil})$, $s_{\log(y)}^2 = 0.019 \log(\text{niños por cada mil})^2$.

$s_{x \log(y)} = -0.2284 \cdot 10^3 \text{€} \log(\text{niños por cada mil})$.

Modelo de regresión exponencial de la mortalidad infantil sobre el PIB:

$$y = e^{3.0135 + 0.0761x}.$$

$y(23) = 3.5332\$$. niños por cada mil.

4. La fiabilidad de ambos modelos es la misma ya que tienen el mismo coeficiente de determinación.

Ejercicio 5.3. Sabiendo que las rectas de regresión correspondientes a dos variables X e Y se cortan en el punto $(2, 3)$ y que la predicción que da la recta de regresión para $x = 3$ es $y = 1$, ¿cuánto cambiará Y según el modelo lineal por cada unidad que aumente X ? Si el coeficiente de correlación lineal es -0.8 , ¿cuánto cambiará X según el modelo lineal por cada unidad que aumente Y ?

Solución

$$\bar{x} = 2 \text{ and } \bar{y} = 3.$$

$b_{yx} = -2$, de modo que Y decrece 2 unidades cuando X crece una unidad.

$b_{xy} = -0.32$, de modo que X decrece 0.32 unidades cuando Y crece una unidad.

6. 2017-01-10 Examen de Farmacia y Biotecnología

Grados: Farmacia, Biotecnología Fecha: 10 de enero de 2017

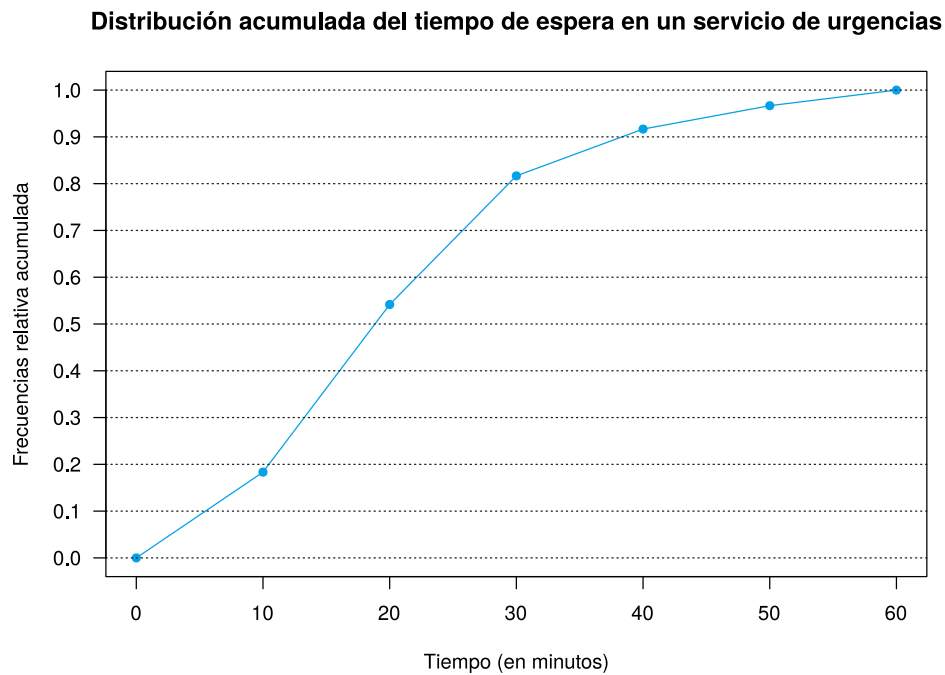
Ejercicio 6.1. La siguiente tabla recoge la distribución de frecuencias del tiempo de espera en un servicio de urgencias de una muestra de pacientes.

Tiempo	Pacientes
(0, 10]	22
(10, 20]	43
(20, 30]	33
(30, 40]	12
(40, 50]	6
(50, 60]	4

Se pide:

1. Dibujar el polígono de frecuencias relativas acumuladas del tiempo de espera.
2. Calcular la mediana el tiempo de espera e interpretarla.
3. ¿Qué porcentaje de pacientes han tenido que esperar más de 38 minutos?

 Solución



- 1.
2. $Me = 18.89$ min.
3. El 10% de los pacientes han tenido que esperar más de 18 minutos.

Ejercicio 6.2. En dos poblaciones de mujeres A y B se ha tomado una muestra y se ha medido el número de embarazos de cada mujer durante su vida fértil obteniendo los siguientes resultados:

A	2	3	4	4	3	2	6	1	5	3	4	4	3	2	5	0
B	1	0	2	1	0	2	0	3	0	1	0	2	5	1	1	1

Se pide:

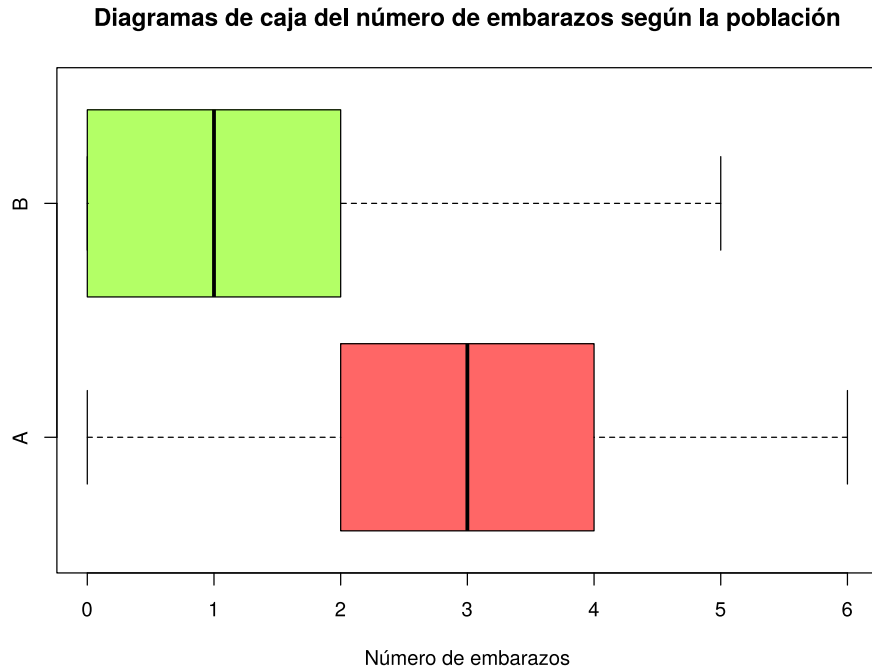
1. Construir los diagramas de caja de ambas muestras y compararlos.
2. ¿En qué muestra es más representativa la media? Justificar la respuesta.
3. Calcular el coeficiente de asimetría de ambas distribuciones. ¿Qué distribución es más asimétrica?
4. ¿Qué número de embarazos es relativamente mayor, 5 embarazos en la población A o 3 en la B?

Utilizar las siguientes sumas para los cálculos:

$$\sum a_i = 51, \sum a_i^2 = 199, \sum (a_i - \bar{a})^3 = -11.6016, \sum (a_i - \bar{a})^4 = 217.9954,$$

$$\sum b_i = 20, \sum b_i^2 = 52, \sum (b_i - \bar{b})^3 = 49.5, \sum (b_i - \bar{b})^4 = 220.3125.$$

Solución



- 1.
2. $\bar{a} = 3.1875$ embarazos, $s_a^2 = 2.2773$ embarazos², $s_a = 1.5091$ embarazos, $cv_a = 0.4734$.
 $\bar{b} = 1.25$ embarazos, $s_b^2 = 1.6875$ embarazos², $s_b = 1.299$ embarazos, $cv_b = 1.0392$.
 Como el coeficiente de variación de A es menor que el coeficiente de variación de B , la media de la población A es más representativa que la media de la población B .
3. $g_{1,a} = -0.211$ y $g_{1,b} = 1.4113$, de modo que la distribución de B es más asimétrica que la distribución de A .
4. $z_a(5) = 1.2011$ y $z_b(3) = 1.3472$, de modo que 3 embarazos en la población B es relativamente mayor que 5 embarazos en la población A .

Ejercicio 6.3. En un estudio se ha medido la reducción en el nivel de colesterol de un

grupo de personas hipertensas tras un programa de ejercicios. Los resultados aparecen en la siguiente tabla.

1. ¿Qué modelo de regresión explica mejor la reducción de colesterol en función de los minutos de ejercicio, el lineal o el exponencial? Justificar la respuesta.
2. Según el modelo de regresión lineal, ¿cuánto disminuirá el colesterol por cada minuto más de ejercicio?
3. Según el modelo logarítmico, ¿cuántos minutos de ejercicio se necesitan para reducir el colesterol 100 mg/dl? ¿Es fiable la predicción? Justificar la respuesta.

Utilizar las siguientes sumas para los cálculos (X =Minutos de ejercicio e Y =Reducción de colesterol):

$$\begin{aligned}\sum x_i &= 2148, \sum \log(x_i) = 53.0559, \sum y_j = 199, \sum \log(y_j) = 27.1766, \\ \sum x_i^2 &= 507082, \sum \log(x_i)^2 = 282.9578, \sum y_j^2 = 5779, \sum \log(y_j)^2 = 80.035, \\ \sum x_i y_j &= 50750, \sum x_i \log(y_j) = 6359.0468, \sum \log(x_i) y_j = 1097.978, \sum \log(x_i) \log(y_j) = 147.0682.\end{aligned}$$

Solución

1. Modelo de regresión lineal de la reducción del colesterol sobre el tiempo de ejercicio:

$$\bar{x} = 214.8 \text{ min}, s_x^2 = 4569.16 \text{ min}^2.$$

$$\bar{y} = 19.9 \text{ mg/dl}, s_y^2 = 181.89 \text{ (mg/dl)}^2.$$

$$s_{xy} = 800.48 \text{ min mg/dl}.$$

$$r^2 = 0.771.$$

Modelo de regresión exponencial de la reducción de colesterol sobre el tiempo de ejercicio: $\log(\bar{y}) = 2.7177 \log(\text{mg/dl})$, $s_{\log(y)}^2 = 0.6178 \log(\text{mg/dl})^2$.

$$s_{x \log(y)} = 52.1504 \text{ min log(mg/dl)}.$$

$$r^2 = 0.9635.$$

Por tanto, el modelo de regresión exponencial es mejor ya que su coeficiente de determinación es mayor.

2. Recta de regresión de la reducción del colesterol sobre el tiempo de ejercicio:
 $y = -17.7312 + 0.1752x$.

Por cada minuto más de ejercicio la reducción del colesterol aumenta 0.1752 mg/dl.

3. Modelo de regresión logarítmico del tiempo de ejercicio sobre la reducción del colesterol: $x = -14.6075 + 84.4135 \log(y)$.

$$x(100) = 374.131.$$

A pesar de que el coeficiente de determinación está muy cerca de 1, la estimación no es muy fiable porque 100 mg/dl está bastante lejos del rango de valores de la muestra.

Ejercicio 6.4. En el servicio de emergencias de un municipio se sabe que por término medio se producen 6 avisos cada día. Sabiendo que el servicio está organizado en 3 turnos diarios de 8 horas, se pide:

1. Calcular la probabilidad de que en un turno se produzcan más de 3 avisos.
2. Calcular la probabilidad de que en alguno de los tres turnos de un día no se produzca ningún aviso.

 Solución

1. Llamando X al número de avisos en un turno de 8 horas, $X \sim P(6)$ y $P(X > 3) = 0.1429$.
2. Llamando Y al número de turnos sin avisos, $Y \sim B(3, 0.1353)$ y $P(Y > 0) = 0.3535$.

Ejercicio 6.5. Para detectar una enfermedad con una prevalencia del 10% se dispone de un test diagnóstico con una sensibilidad del 95% y una especificidad del 85%. Se pide:

1. Calcular los valores predictivos positivo y negativo del test e interpretarlos. ¿Se trata de un test más útil para detectar la enfermedad o para descartarla?
2. ¿Cuál debería la especificidad del test para que el valor predictivo positivo fuera del 80%?

 Solución

1. $VPP = P(D|+) = 0.413$ y $VPN = P(\bar{D}|-) = 0.9935$.
2. La especificidad debería ser 97.37%.

Ejercicio 6.6. Se ha medido la presión arterial sistólica a 8000 individuos de una población se ha observado que 2254 tiene más de 130 mmHg y 3126 tienen entre 110 y 130 mmHg. Suponiendo que la presión arterial sistólica sigue una distribución de probabilidad normal, se pide:

1. Calcular la media y la desviación típica.
2. Si se consideran hipertensas las personas con una presión arterial superior a 140 mmHg, ¿cuántas personas hipertensas hay en la población?
3. Si una analítica sanguínea marca como anormales el 5% de los individuos con menor presión arterial y el 5% con mayor presión arterial, ¿entre qué presiones arteriales debe estar la presión de un individuo para que la analítica lo considere normal?

 Solución

1. Llamando X a la presión arterial, $X \sim N(118.723, 19.5221)$.
2. $P(X > 140) = 0.1379$ y por tanto existen 1103.0473 personas con hipertensión.
3. La presión es normal en el intervalo $(86.612, 150.8341)$.

Ejercicio 6.7. En una asignatura se hacen dos exámenes parciales a lo largo del curso. El primer parcial lo aprobaron el 60% de los alumnos y el segundo parcial lo aprobaron el 68%. Del grupo de alumnos que aprobaron el primer parcial, el 80% de ellos aprobaron el segundo. Si se elige un alumno al azar, calcular:

1. Probabilidad de que no haya aprobado ningún parcial.
2. Probabilidad de que haya aprobado el primer parcial si no ha aprobado el segundo.

 Solución

Llamando E_1 al evento consistente en aprobar el primer examen y E_2 al evento consistente en aprobar el segundo examen:

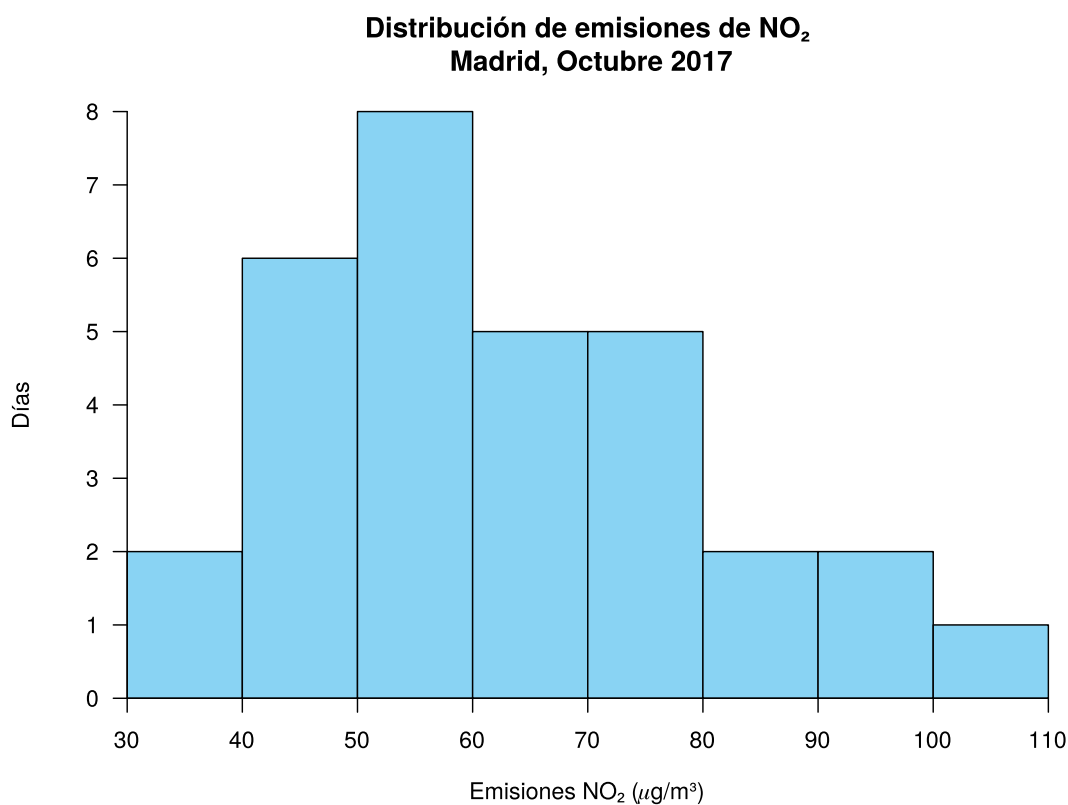
1. $P(\overline{E}_1 \cap \overline{E}_2) = 0.2$.
2. $P(E_1 | \overline{E}_2) = 0.375$.

7. 2017-11-27 Examen de Farmacia y Biotecnología

Titulación: Grado en Farmacia y Biotecnología

Fecha: 27 de Noviembre de 2017

Ejercicio 7.1. El siguiente diagrama muestra la distribución de emisiones de NO (g/m^3) en Madrid en los días de octubre de 2017.



1. La normativa europea sobre calidad del aire establece que el valor medio mensual no debe exceder de 40 g/m^3 . ¿Se ha cumplido la norma en el mes de Octubre? ¿Es este un valor representativo de las mediciones tomadas en octubre?

2. El Ayuntamiento de Madrid ha decidido que se establecerán restricciones de velocidad en los accesos los días en los que se superen los 72 g/m^3 y que además de estas restricciones se establecerán también restricciones al aparcamiento los días que se superen los 92 g/m^3 . ¿Qué porcentaje de días de octubre se establecieron solo restricciones de velocidad en los accesos?
3. De acuerdo con esta muestra de datos tomada durante el mes de octubre, ¿puede establecerse por la forma de la distribución de la muestra que la distribución de las emisiones en toda la ciudad sigue una distribución normal?
4. Además del nivel de NO, el Ayuntamiento también controla los niveles de SO, y se sabe que el nivel medio de esta sustancia durante el mes de octubre fue de 2.85 g/m^3 con una desviación típica de 0.42 g/m^3 . Si un día hubo un nivel de NO de 46 y un nivel de SO de 2.24, ¿cuál de las dos sustancias tenía niveles más altos en referencia a sus mediciones?
5. Si el índice de calidad del aire (ICA) puede estimarse multiplicando el nivel de NO por 0.95 y sumándole una cantidad fija de 30. ¿Cuál fue el índice medio de la calidad del aire en Madrid el mes de octubre? ¿Es un valor más o menos representativo que el nivel de emisiones medio de NO?
6. ¿Existen días atípicos en las emisiones de NO del mes de octubre? Justificar la respuesta.

Utilizar las siguientes sumas para los cálculos: $\sum x_i = 1945 \text{ g/m}^3$, $\sum x_i^2 = 131575 (\text{g/m}^3)^2$, $\sum (x_i - \bar{x})^3 = 93995.838 (\text{g/m}^3)^3$ y $\sum (x_i - \bar{x})^4 = 7766271.021 (\text{g/m}^3)^4$.

Solución

1. $\bar{x} = 62.7419 \text{ g/m}^3$, de manera que no se cumple el requisito.
 $s^2 = 307.8044 (\text{g/m}^3)^2$, $s = 17.5444 \text{ g/m}^3$, $cv = 0.2796$.
 Como el coeficiente de variación es menor que 0.3 hay poca variabilidad en los datos y la media es bastante representativa.
2. $F(72) = 0.7097$ y $F(92) = 0.9161$, de manera que el porcentaje de días únicamente con restricciones de velocidad es 20.64%.
3. $g_1 = 0.5615$ y $g_2 = -0.3558$. Como ambos están entre -2 y 2, se puede asumir que la distribución de emisiones es normal.
4. NO : $z(46) = -0.9543$.
 SO : $z(2.24) = -1.4524$.
 Así pues, las emisiones de NO son relativamente mayores.
5. Sea $y = 0.95x + 30$ el ICA.
 $\bar{y} = 89.6048$, $s_y = 16.6671$, $cv = 0.186$. Como el coeficiente de variación es menor, la media del ICA es más representativa.
6. $Q_1 = 49.5816 \text{ g/m}^3$, $Q_3 = 74.0093 \text{ g/m}^3$ y $IQR = 24.4277 \text{ g/m}^3$. Vallas:
 $v_1 = 12.94 \text{ g/m}^3$ y $v_2 = 110.65 \text{ g/m}^3$. Por tanto, no hay datos atípicos.

Ejercicio 7.2. La siguiente tabla muestra las tasas de incidencia de gripe por cada 100.000 habitantes registradas al cabo de un número de días desde el comienzo de el estudio.

Días	1	5	8	12	20	26	38	44
Tasa de gripe	60	66	71	80	106	132	194	235

Se pide:

1. Calcular la tasa de incidencia de gripe a los 50 días desde el comienzo del estudio mediante un modelo de regresión lineal.
2. ¿Cuánto varía la tasa de incidencia de gripe cada día según el modelo lineal?
3. Calcular la tasa de incidencia de gripe a los 50 días desde el comienzo del estudio mediante un modelo de regresión exponencial.
4. ¿Cuál de las predicciones anteriores es más fiable? Razonar la respuesta.

Utilizar las siguientes sumas para los cálculos (X =Días e Y =Tasa de gripe):
 $\sum x_i = 154$, $\sum \log(x_i) = 19.8494$, $\sum y_j = 944$, $\sum \log(y_j) = 37.2024$, $\sum x_i^2 = 4690$,
 $\sum \log(x_i)^2 = 60.2309$, $\sum y_j^2 = 140918$, $\sum \log(y_j)^2 = 174.8363$, $\sum x_i y_j = 25182$,
 $\sum \log(x_i) y_j = 2795.2484$, $\sum x_i \log(y_j) = 772.3504$, $\sum \log(x_i) \log(y_j) = 96.1974$.

Solución

1. Modelo lineal de la tasa de gripe sobre los días:
 $\bar{x} = 19.25$ días, $s_x^2 = 215.6875$ días².
 $\bar{y} = 118$ personas, $s_y^2 = 3690.75$ personas².
 $s_{xy} = 876.25$ días personas.
Recta de regresión de la tasa de gripe sobre los días: $y = 39.7951 + 4.0626x$.
 $y(50) = 242.9247$.
2. 4.0626 personas por día.
3. $\overline{\log(y)} = 4.6503 \log(\text{personas})$, $s_{\log(y)}^2 = 0.2293 \log(\text{personas})^2$.
 $s_{x \log(y)} = 7.0255$ días $\log(\text{personas})$.
Modelo exponencial de la tasa de gripe sobre los días: $y = e^{4.0233+0.0326x}$.
 $y(50) = 284.8357$.
4. Coeficiente de determinación lineal de la tasa de gripe sobre los días $r^2 = 0.9645$. Coeficiente de determinación exponencial de la tasa de gripe sobre los días $r^2 = 0.9982$.
Así pues, el modelo exponencial explica un poco mejor la evolución de la tasa de gripe con respecto a los días.

8. 2018-01-19 Examen de Farmacia y Biotecnología

Grados: Farmacia y Biotecnología

Fecha: 19 de Enero de 2018

Ejercicio 8.1. Se realizó un estudio en pacientes mayores de 60 años de la relación entre la edad de los pacientes X y el número de veces que han acudido a consulta médica en el último año Y , obteniéndose los siguientes resultados:

Edad	62	65	71	79	83	88	90	95
Consultas	2	3	4	6	6	8	10	14

Se pide:

1. Calcular el número de veces que acudirá a consulta médica un paciente de 70 años según el modelo de regresión lineal.
2. Lo mismo pero empleando el modelo de regresión exponencial.
3. Razonar cuál de las dos predicciones es más fiable.
4. Teniendo en cuenta que la ecuación del modelo potencial es $Y = aX^b$, explicar qué transformaciones hay que aplicar a las variables X e Y para convertirlo en un modelo lineal.

Usar las siguientes sumas para los cálculos: $\sum x_i = 633$, $\sum \log(x_i) = 34.8835$, $\sum y_j = 53$, $\sum \log(y_j) = 13.7827$, $\sum x_i^2 = 51109$, $\sum \log(x_i)^2 = 152.28$, $\sum y_j^2 = 461$, $\sum \log(y_j)^2 = 26.6206$, $\sum x_i y_j = 4509$, $\sum x_i \log(y_j) = 1144.0108$, $\sum \log(x_i) y_j = 235.1289$, $\sum \log(x_i) \log(y_j) = 60.7921$.

Solución

1. Modelo lineal de Consultas sobre Edad:
 $\bar{x} = 79.125$ años, $s_x^2 = 127.8594$ años².
 $\bar{y} = 6.625$ consultas, $s_y^2 = 13.7344$ consultas².
 $s_{xy} = 39.4219$ años consultas.
Recta de regresión de Consultas sobre Edad: $y = -17.771 + 0.3083x$.
 $y(70) = 3.8116$ consultas.
2. $\log(y) = 1.7228 \log(\text{consultas})$, $s_{\log(y)}^2 = 0.3594 \log(\text{consultas})^2$.

$$s_x \log(y) = 6.6823 \text{ años } \log(\text{consultas}).$$

Modelo exponencial de consultas sobre Edad: $y = e^{-2.4124+0.0523x}$.

$$y(70) = 3.4762 \text{ consultas.}$$

3. Coeficiente de determinación lineal de Consultas sobre Edad $r^2 = 0.885$.
Coeficiente de determinación exponencial de Consultas sobre Edad $r^2 = 0.9716$.
Por tanto, el modelo exponencial explica un poco mejor el número de consultas médicas con respecto a la edad.
4. Hay que aplicar la transformación logarítmica tanto a las Consultas como a la Edad: $\log(Y) = \log(aX^b) \Rightarrow \log(Y) = \log(a) + \log(X^b) = \log(a) + b \log(X) = a' + b \log(X)$.

Ejercicio 8.2. Se ha medido la concentración de polen en granos/m³ de una determinada planta en el centro de una ciudad a lo largo de un año y se han obtenido los siguientes resultados:

Nivel de polen	Nº de días
0 – 300	51
300 – 500	60
500 – 600	79
600 – 800	91
800 – 1000	60
1000 – 1300	24

1. Si el 75% de los días se considera que los niveles de polen no han sido excesivos. ¿A partir de que nivel de polen se ha considerado un nivel excesivo?
2. Se ha establecido una alerta naranja cuando el nivel de polen se sitúa entre 575 y 850. ¿En cuántos días del año se produjo dicha alerta naranja?
3. ¿Hay datos atípicos en la muestra?
4. Sabemos que hay otra planta que tiene un ciclo de polinización muy parecido al de la muestra, con unos niveles de polen que se pueden calcular de la forma $Y = 0.5X - 100$. (Y = niveles de la otra planta y X =niveles de la planta de la muestra) ¿Cuál sería la media de nivel de polen para esta otra planta? ¿Es un valor más o menos representativo que la de los niveles de la planta de la muestra recogida?
5. ¿Se podría considerar que el nivel de polen en general sigue una distribución normal?

Usar las siguientes sumas para los cálculos: $\sum x_i = 220400$ granos/m³, $\sum x_i^2 = 159575000$ (granos/m³)², $\sum (x_i - \bar{x})^3 = 261917220.867$ (granos/m³)³ y $\sum (x_i - \bar{x})^4 = 4872705679772.61$ (granos/m³)⁴.

 Solución

1. $P_{75} = 784.0417$ granos/m³.
2. $F(575) = 0.4664$ y $F(860) = 0.8192$, por lo que la frecuencia del número de días con alerta es 0.3528 que corresponde a 128.77 días.
3. $Q_1 = 434.1849$ granos/m³, $Q_3 = 784.0417$ granos/m³ y $RI = 349.8568$ granos/m³. Vallas: $v_1 = -90.6001$ granos/m³ y $v_2 = 1308.8269$ granos/m³. Como todos los valores caen dentro de las vallas no hay datos atípicos.
4. $\bar{x} = 603.8356$ granos/m³, $s_x^2 = 72574.3291$ (granos/m³)², $s_x = 269.3962$ granos/m³ y $cv_x = 0.4461$
 $\bar{y} = 201.9178$ granos/m³, $s_y = 134.6981$ granos/m³ y $cv_y = 0.6671$.
La media de X es más representativa que la media de Y ya que $cv_x < cv_y$.
5. $g_1 = 0.0367$ y $g_2 = -0.4654$. Como ambos están entre -2 y 2, se puede asumir que el nivel de polen se distribuye normalmente.

Ejercicio 8.3. Los niveles de polen de gramíneas registrados en Madrid durante el año 2017 se distribuyeron de forma normal y tuvieron una media de 90. Si en 42 días del año 2017 se superó los 120 granos de polen/m³ de aire, se pide:

1. Calcular la desviación típica del nivel de polen de gramíneas en 2017.
Nota: Si no se sabe calcular tomar una desviación típica de 20 granos/m³ para los demás apartados.
2. ¿Durante cuantos días no se llegaron a 50 granos de polen/m³ de aire?
3. Si el 20% de los días hubo un nivel excesivamente alto de polen y hubo que avisar a la población, ¿a partir de qué nivel de polen se produjeron estos avisos?

 Solución

Sea X el nivel de polen en Madrid en 2017. $X \sim N(90, \sigma)$.

1. $\sigma = 25$ granos/m³.
2. $P(X \leq 50) = 0.0548$ que corresponde a 20.0017 días.
3. $P_{80} = 111.0405$ granos/m³.

Ejercicio 8.4. Se han ensayado dos tipos de medicamentos para reducir los niveles de colesterol y se ha observado que el medicamento A es efectivo en el 75% de los casos y el medicamento B es efectivo en el 85% de los casos. Sin embargo, hay un 5% de casos en los que ninguno de los medicamentos es efectivo.

1. ¿En qué porcentaje de casos sería efectivo exclusivamente el medicamento A ?
2. Si en un paciente se ha comprobado que el medicamento A es efectivo, ¿qué probabilidad hay de que también lo sea el medicamento B ?

3. Si en un paciente se ha comprobado que el medicamento B no es efectivo, ¿qué probabilidad hay de que el medicamento A si lo sea?
4. ¿Es independiente la efectividad de estos medicamentos?

 Solución

1. $P(A \cap \bar{B}) = 0.1$, es decir, un 10%.
2. $P(B|A) = 0.8667$.
3. $P(A|\bar{B}) = 0.6667$.
4. $P(B|A) \neq P(B)$, de manera que los sucesos son dependientes.

Ejercicio 8.5. El número medio de nacimientos que se producen en una semana en un hospital es 14. Se pide:

1. Calcular la probabilidad de que un día se produzcan más de 2 nacimientos.
2. Calcular la probabilidad de que una semana haya más de un día en que no se produzcan nacimientos.

 Solución

1. Sea X el número de nacimientos en un día. $X \sim P(2)$ y $P(X > 2) = 0.3233$.
2. Sea Y el número de días sin nacimientos en una semana. $Y \sim B(7, 0.1353)$ y $P(Y > 1) = 0.2427$.

Ejercicio 8.6. Se está trabajando en el diseño de un test para detectar una enfermedad y para ensayarlo se dispone de una muestra de 250 individuos de los cuales 50 presentan la enfermedad y 200 son individuos sanos. Si se pretende que el test tenga un valor predictivo positivo de 0.7 y un valor predictivo negativo de 0.9, al realizar los ensayos sobre la muestra:

1. ¿Cuántos de los individuos sanos deberían dar positivo en el test?
2. ¿Cuántos de los individuos enfermos deberían dar negativo en el test?
3. ¿Qué probabilidad hay de que un individuo que ha dado dos veces positivo en el test sufra la enfermedad?

 Solución

Sea E el suceso correspondiente a tener la enfermedad.

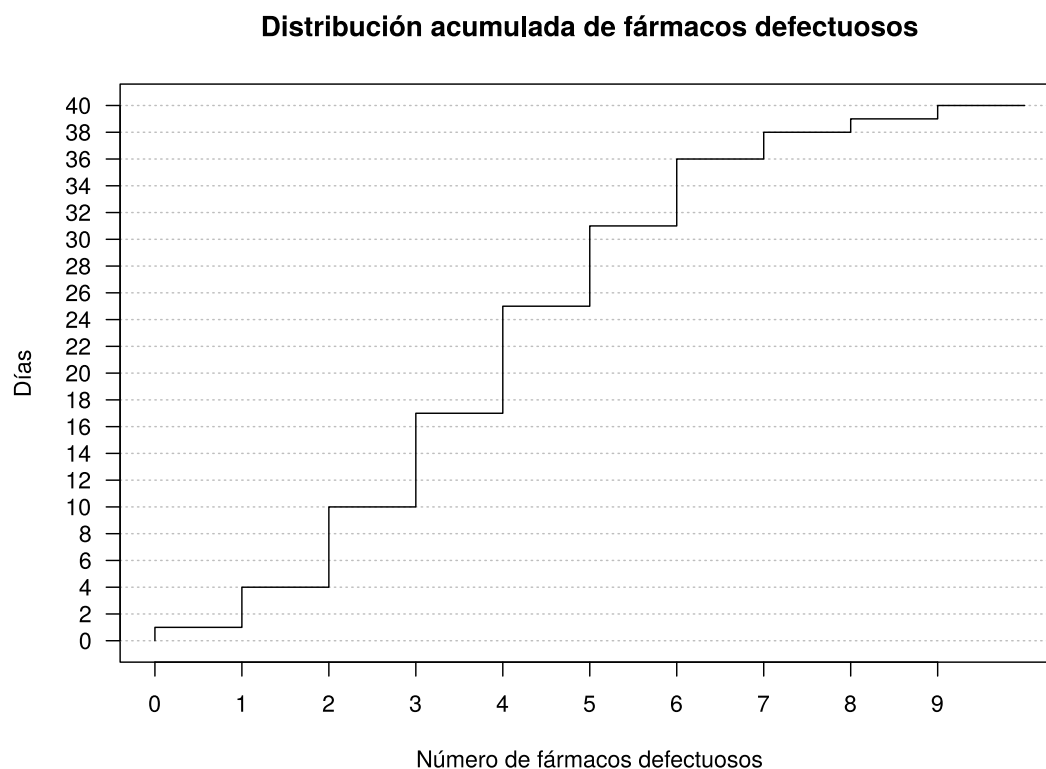
1. $P(+|\bar{E}) = 0.0625 \Rightarrow 12.5$ personas.
2. $P(-|E) = 0.4165 \Rightarrow 20.825$ personas.
3. $P(E|+ \cap +) = 0.9561$.

9. 2018-12-17 Examen de Farmacia y Biotecnología

Grados: Farmacia y Biotecnología

Fecha: 17 de Diciembre de 2018

Ejercicio 9.1. El siguiente gráfico representa la distribución acumulada del número de fármacos defectuosos diarios producidos por una máquina en una muestra de 40 fármacos.



1. Construir la tabla de frecuencias del número de fármacos defectuosos.
2. Dibujar el diagrama de caja y bigotes del número de fármacos defectuosos.
3. Estudiar la asimetría de la distribución del número de fármacos defectuosos.

4. Si el número de fármacos defectuosos producidos por una segunda máquina sigue la ecuación $y = 3x + 2$, donde x e y son el número de fármacos defectuosos producidos con la primera y la segunda máquina respectivamente, ¿en qué máquina es más representativa la media del número de fármacos defectuosos?
5. ¿Qué número de fármacos defectuosos es relativamente menor, 3 fármacos en la primera máquina o 9 fármacos en la segunda?

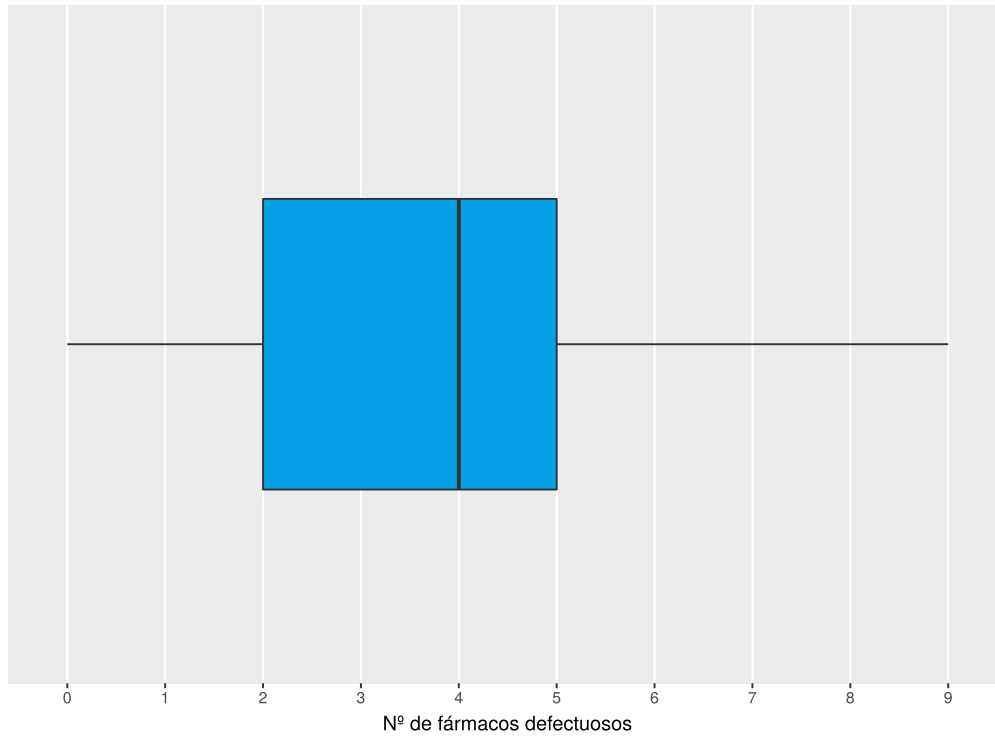
Solución

1.

Fármacos defectuosos	n_i	f_i	N_i	F_i
0	1	0.025	1	0.025
1	3	0.075	4	0.100
2	6	0.150	10	0.250
3	7	0.175	17	0.425
4	8	0.200	25	0.625
5	6	0.150	31	0.775
6	5	0.125	36	0.900
7	2	0.050	38	0.950
8	1	0.025	39	0.975
9	1	0.025	40	1.000

2.

Diagrama de caja del número diario de fármacos defectuosos



3. $\bar{x} = 3.975$ fármacos, $s_x = 1.9936$ fármacos y $g_1 = 0.3184$. Por tanto, la distribución es un poco asimétrica hacia la derecha.
4. $cv_x = 0.5015$, $\bar{y} = 13.925$ fármacos, $s_y = 5.9808$ fármacos y $cv_y = 0.4295$. Así pues, la media de y es más representativa que la media de x ya que su coeficiente de variación es menor.
5. $z_x = -0.4891$ y $z_y = -0.8235$, por tanto, 9 fármacos defectuosos en la máquina y es relativamente menor.

Ejercicio 9.2. Un laboratorio farmacéutico fabrica dos modelos de tensiómetro, uno de brazo y otro de muñeca. Para comprobar la precisión de los mismos se ha realizado un control de calidad tomando la tensión a 20 pacientes y se han obtenido los siguientes resultados:

$$\begin{aligned} \sum x_i &= 265.4 \text{ mmHg}, \sum y_i = 262.5 \text{ mmHg}, \sum z_i = 262.4 \text{ mmHg}, \\ \sum x_i^2 &= 3701.14 \text{ mmHg}^2, \sum y_i^2 = 3629.41 \text{ mmHg}^2, \sum z_i^2 = 3615.38 \text{ mmHg}^2, \\ \sum x_i y_i &= 3658.28 \text{ mmHg}^2, \sum x_i z_i = 3655.95 \text{ mmHg}^2, \sum y_i z_i = 3613.97 \text{ mmHg}^2. \end{aligned}$$

Donde X es la tensión con el tensiómetro de brazo, Y con el tensiómetro de muñeca y Z la tensión real. Se pide:

1. ¿Qué tensiómetro predice mejor la tensión real?

2. Si un paciente tiene una tensión real de 13.5 mmHg, ¿qué se espera que marque el tensiómetro de brazo?

 Solución

1. Presión arterial con el tensiómetro de brazo: $\bar{x} = 13.27$ mmHg, $s_x^2 = 8.9641$ mmHg².
Presión arterial con el tensiómetro de muñeca: $\bar{y} = 13.125$ mmHg, $s_y^2 = 9.2049$ mmHg².
Presión arterial real: $\bar{z} = 13.12$ mmHg, $s_z^2 = 8.6346$ mmHg².
 $s_{xz} = 8.6951$ mmHg², $s_{yz} = 8.4985$ mmHg², $r_{xz}^2 = 0.9768$ y $r_{yz}^2 = 0.9087$.
Así pues, el tensiómetro de brazo predice mejor la presión arterial real con un modelo de regresión lineal ya que el coeficiente de determinación lineal es mayor.
2. Recta de regresión de X sobre Z : $x = 0.0581 + 1.007z$.
Predicción: $x(13.5) = 13.6527$ mmHg.

Ejercicio 9.3. La recta de regresión de Y sobre X es $y = 1.2x - 0.6$. Se pide:

1. Indicar razonadamente cuáles de las siguientes rectas no pueden ser la recta de regresión de X sobre Y .
 - a. $x = 0.9y - 0.6$
 - b. $x = -0.7y + 0.4$
 - c. $x = 0.8y - 0.7$
 - d. $x = -0.6y - 0.5$
 - e. $x = 0.4y - 0.6$
 - f. $x = -0.5y + 0.9$
2. Entre las que puedan ser indicar razonadamente con cuál de ellas serían más fiables las predicciones.

 Solución

1. (b), (d) y (f) no son posibles porque tienen pendiente negativa, y (a) no es posible porque el coeficiente de determinación lineal es mayor que 1.
2. (c) da las mejores predicciones porque su coeficiente de determinación es mayor.

Ejercicio 9.4. En un estudio epidemiológico se tomó una muestra de 400 personas con cáncer de pulmón y otra de 1200 sin cáncer de pulmón. Entre las personas con cáncer de pulmón se observó que había 180 fumadoras, mientras que entre las personas sin cáncer de pulmón había 1140 no fumadoras. Se pide:

1. Calcular el riesgo relativo de desarrollar cáncer al fumar e interpretarlo.
2. Calcular el odds ratio de desarrollar cáncer al fumar e interpretarlo.

Solución

Sea C el evento correspondiente a tener cáncer de pulmón.

1. $RR(C) = 9$. Esto quiere decir que la probabilidad de tener cancer de pulmón al fumar es 9 veces mayor que sin fumar.
2. $OR(C) = 15.5455$. Como es positivo existe una asociación directa entre fumar y tener cáncer de pulmón. El odds de tener cáncer de pulmón si se fuma es más the 15 veces mayor que si no se fuma.

Ejercicio 9.5. Se quiere desarrollar un test diagnóstico que sirva para descartar una enfermedad cuando el test es negativo (valor predictivo negativo) con al menos una probabilidad del 90% para aplicarlo en una población en la que se sabe que hay un 15% de individuos que presentan la enfermedad. Si la sensibilidad del test se ha establecido en un 80%,

1. ¿Qué especificidad mínima tendría que tener el test?
2. ¿Cuál sería en ese caso su probabilidad de diagnóstico correcto?
3. ¿Cuál sería la probabilidad de descartar la enfermedad en un paciente que se le ha aplicado dos veces el test y ha dado negativo en ambos casos?

Solución

Sea E el evento correspondiente a tener la enfermedad y $+$ y $-$ los eventos correspondientes a tener un resultado positivo o negativo en el test diagnóstico respectivamente.

1. Especificidad mínima $P(-|\bar{E}) = 0.3176$.
2. $P(VP) + P(VN) = P(E \cap +) + P(\bar{E} \cap -) = 0.12 + 0.27 = 0.39$.
3. $P(\bar{E} | -_1 \cap -_2) = 0.9346$.

Ejercicio 9.6. Se sabe que en una ciudad una de cada 20 personas, por término medio, tiene sangre del tipo AB .

1. Si se eligen al azar doscientos donantes de sangre, ¿cuál es la probabilidad de que al menos 5 tengan sangre de tipo AB ?
2. Si se eligen al azar 10 personas, ¿cuál es la probabilidad de que haya más de 8 con grupo sanguíneo distinto de AB ?

 Solución

1. Sea X el número de donantes con grupo sanguíneo AB en una muestra de 200 donantes de sangre. Entonces, $X \sim B(200, 1/20) \approx P(10)$, y $P(X \geq 5) = 0.9707$.
2. Sea Y el número de donantes con grupo sanguíneo distinto de AB en una muestra de 10 donantes. Entonces, $Y \sim B(10, 19/20)$, y $P(Y > 8) = 0.9139$.

Ejercicio 9.7. En una asignatura hay 230 alumnos, de los cuales 150 son chicas y 80 chicos. Se sabe que las notas siguen distribuciones normales con la misma desviación típica en las chicas y en los chicos. Si hay 120 chicas y 56 chicos que tienen una nota superior a 5 y 36 chicos tienen una nota entre 5 y 7. Se pide:

1. Las medias y las desviaciones típicas de las notas de chicas y chicos.
2. ¿Cuántas chicas tendrán una nota entre 4.5 y 8?
3. ¿Por encima de qué nota estarán el 10% de las chicas?

 Solución

Sea X la nota de un chico aleatorio del curso e Y la nota de una chica aleatoria del curso. Entonces, $X \sim N(\mu_x, \sigma)$ y $Y \sim N(\mu_y, \sigma)$.

1. $\mu_x = 5.87$, $\mu_y = 6.41$ y $\sigma = 1.68$.
2. $P(4.5 \leq Y \leq 8) = 0.7018$, es decir, 105.27 chicas.
3. $P_{90} = 8.8$.

10. 2019-12-16 Examen de Farmacia y Biotecnología

Grados: Farmacia y Biotecnología

Fecha: 16 de Diciembre de 2019

Ejercicio 10.1. Se ha analizado en 50 casos el tiempo en minutos que, después de una operación, un paciente ha tardado en eliminar la anestesia, obteniéndose el siguiente resultado:

Tiempo	Pacientes
10 – 30	2
30 – 45	11
45 – 60	18
60 – 90	9
90 – 120	8
120 – 180	2

Se pide:

1. ¿Presenta la muestra algún dato atípico?
2. ¿Es la media un valor representativo de la muestra? ¿Cuánto vale?
3. Si un protocolo de postoperatorio contempla monitorizar al 15% de los pacientes que más tardan en eliminar la anestesia. ¿A partir de que tiempo se debe monitorizar a un paciente según la muestra?
4. Si se suministra un fármaco antagonista del anestésico, se sabe que el tiempo de eliminación de la anestesia disminuye en un 25%. ¿Cómo afectará esta disminución a la representatividad de la nueva media?
5. Si se sabe que el tiempo de eliminación de otro tipo de anestesia B tiene media 50 minutos y desviación típica 15 minutos, ¿qué tiempo de eliminación es relativamente mayor, 70 minutos con este tipo de anestesia o 60 minutos con el tipo de anestesia B ?

Usar las siguientes sumas para los cálculos:

$$\sum x_i n_i = 3212.5 \text{ min}, \quad \sum x_i^2 n_i = 249706.25 \text{ min}^2,$$

$$\sum (x_i - \bar{x})^3 n_i = 1400531.25 \text{ min}^3 \text{ y}$$

$$\sum (x_i - \bar{x})^4 n_i = 143958437.7 \text{ min}^4.$$

 Solución

1. $C_1 = 44.3182$, $C_3 = 81.6667$, $RI = 37.3485$, $v_1 = -11.7045$ y $v_2 = 137.6894$.
Puesto que la última clase contiene valores por encima de la valla superior, podría haber datos atípicos.
2. $\bar{x} = 64.25$ min, $s^2 = 866.0625$ min², $s = 29.4289$ min y $cv = 0.458$. Por tanto, la representatividad de la media es moderada.
3. $P_{85} = 99.375$ min.
4. Aplicando la transformación lineal $y = 0.75x$, $\bar{y} = 48.1875$ min, $s_y = 22.0717$ min y $cv = 0.458$. Por tanto, la representatividad de la media es la misma.
5. Puntuación típica para la primera anestesia: $z(70) = 0.1954$.
Puntuación típica para la anestesia B : $z(60) = 0.6667$.
Por tanto, 60 min es relativamente mayor con la anestesia B .

Ejercicio 10.2. La siguiente tabla contiene las notas de un grupo de 10 alumnos de matemáticas de farmacia en tres exámenes parciales.

Parcial 1(X)	Parcial 2(Y)	Parcial 3(Z)
5.5	3.2	5.0
7.5	6.5	2.0
2.5	4.0	1.0
6.0	4.0	6.0
8.0	7.5	6.0
4.0	3.5	1.0
7.0	5.5	4.0
9.5	10.0	9.0
10.0	9.5	8.0
1.0	3.0	0.5

Se pide:

1. ¿Cuáles son las dos notas que mejor se correlacionan linealmente?
2. Utilizando modelos lineales, ¿cuáles serían las notas estimadas en los parciales 2 y 3 de un alumno que obtuvo un 6.5 en el parcial 1?

Usar las siguientes sumas para los cálculos:

$$\begin{aligned} \sum x_i &= 61, \sum y_i = 56.7, \sum z_i = 42.5, \\ \sum x_i^2 &= 449, \sum y_i^2 = 382.49, \sum z_i^2 = 264.25, \\ \sum x_i y_i &= 405.85, \sum x_i z_i = 327, \sum y_i z_i = 295. \end{aligned}$$

 Solución

1. $\bar{x} = 6.1$, $s_x^2 = 7.69$, $\bar{y} = 5.67$, $s_y^2 = 6.1001$, $\bar{z} = 4.25$, $s_z^2 = 8.3625$, $s_{xy} = 5.998$, $s_{xz} = 6.775$, $s_{yz} = 5.4025$, $r_{xy}^2 = 0.7669$, $r_{xz}^2 = 0.7138$ y $r_{yz}^2 = 0.5722$. Por tanto, las dos variables más correlacionadas linealmente son X e Y , ya que su coeficiente de determinación es mayor.
2. Recta de regresión de Y sobre X : $y = 0.9122 + 0.78x$ y $y(6.5) = 5.982$.
Recta de regresión de Z sobre X : $z = -1.1242 + 0.881x$ y $z(6.5) = 4.6024$.

Ejercicio 10.3. Para ver si existe algún tipo de asociación entre la osteoporosis y el sexo se ha tomado una muestra aleatoria de personas entre 65 y 70 años y se ha observado el sexo y cuántos presentaban osteoporosis. Los resultados se reflejan en la siguiente tabla.

	Osteoporosis	No osteoporosis
Mujeres	480	2320
Hombres	255	1505

Se pide:

1. Calcular la prevalencia de la osteoporosis en la población.
2. Calcular el riesgo relativo de presentar osteoporosis de las mujeres con respecto a los hombres e interpretarlo.
3. Calcular el odds ratio de presentar osteoporosis de las mujeres con respecto a los hombres e interpretarlo.
4. ¿Cuál de las dos medidas de asociación es más apropiada para estudiar la asociación entre la osteoporosis y el sexo? Justificar la respuesta.

 Solución

Sea E el evento consistente en tener osteoporosis.

1. Prevalencia: $P(E) = 0.1612$.
2. $RR(E) = 1.1832$. Por tanto, el riesgo de sufrir osteoporosis en mujeres es mayor que en hombres, pero no mucho. No existe una asociación fuerte entre la osteoporosis y el sexo.
3. $OR(E) = 1.2211$. Por tanto, el odds de sufrir en mujeres es mayor que en hombres, pero no mucho.
4. Puesto que es posible calcular la prevalencia de la osteoporosis, ambas medidas pueden calcularse, pero el riesgo relativo es más fácil de interpretar.

Ejercicio 10.4. La probabilidad de contraer la gripe en dos ciudades A y B del mismo tamaño es del 14% y del 8% respectivamente. Se pide:

1. Calcular la probabilidad de que en una muestra aleatoria de 10 individuos de la ciudad A haya más de 2 que contraigan la gripe.
2. Calcular la probabilidad de que en una muestra aleatoria de 50 individuos de la ciudad B haya más de 2 y menos de 5 que contraigan la gripe.
3. Calcular la probabilidad de que en una muestra aleatoria de 8 individuos tomados de ambas ciudades haya 2 que contraigan la gripe.
4. Suponiendo que contraer la gripe en ambas ciudades son sucesos independientes, calcular la probabilidad de que en una muestra de 5 personas que han estado en las dos ciudades haya alguna que contraiga la gripe.

Solución

1. Sea X el número de personas que contraen la gripe en una muestra de 10 personas de la población A , entonces $X \sim B(10, 0.14)$ y $P(X > 2) = 0.1545$.
2. Sea Y el número de personas que contraen la gripe en una muestra de 50 personas de la población B , entonces $Y \sim B(50, 0.08) \approx P(4)$ y $P(2 < Y < 5) = 0.3907$.
3. Sea Z el número de personas que contraen la gripe en una muestra de 8 personas de las poblaciones A y B , entonces $Z \sim B(8, 0.11)$ y $P(Z = 2) = 0.1684$.
4. Sea U el número de personas que contraen la gripe en una muestra de 5 personas que han vivido en ambas ciudades, entonces $U \sim B(5, 0.2088)$ y $P(U > 0) = 0.69$.

Ejercicio 10.5. En un estudio sobre el nivel de colesterol de los habitantes de una población se midió el nivel de colesterol de 10000 hombres y 10000 mujeres, obteniéndose que 3420 hombres y 1234 mujeres tenían un nivel de colesterol superior a 230 mg/dl, y que 4936 hombres tenían entre 210 y 230 mg/dl. Suponiendo que los niveles de colesterol en los hombres y en las mujeres siguen distribuciones normales con la misma desviación típica, calcular:

1. Las medias y la desviación típica de las distribuciones del nivel de colesterol en hombres y mujeres.
Nota: Si no se saben calcular las medias y desviación típica, tomar 215 mg/dl y 220 mg/dl como las medias de mujeres y hombres respectivamente, y 10 mg/dl como la desviación típica, para los próximos apartados.
2. El porcentaje de hombres cuyo nivel de colesterol estará entre 200 y 240 mg/dl.
3. El rango intercuartílico del nivel de colesterol en las mujeres.

Solución

1. Sean X e Y los niveles de colesterol en hombres y mujeres respectivamente, entonces $X \sim N(224.1164, 14.4556)$ e $Y \sim N(213.2581, 14.4556)$.

2. $P(200 \leq X \leq 240) = 0.8164$. 3. $RI = 19.5003$ mg/dl.

11. 2020-10-26 Examen de Farmacia y Biotecnología

Grados: Farmacia y Biotecnología

Fecha: 26 de octubre de 2020

Ejercicio 11.1. La siguiente tabla recoge el número de pacientes diarios que ingresaron en un hospital durante el mes de septiembre.

Pacientes	Frecuencia
(10, 14]	6
(14, 18]	10
(18, 22]	7
(22, 26]	6
(26, 30]	1

Se pide:

1. Estudiar la dispersión del 50% de los datos centrales.
2. Calcular la media y estudiar la dispersión con respecto a ella.
3. Estudiar la normalidad de los datos.
4. Se sabe que en mismo hospital durante el mes de abril la media fue 35 pacientes y la varianza 40 pacientes². ¿En qué mes hubo más variabilidad relativa?
5. ¿Qué número de ingresos es relativamente mayor, 20 ingresos en septiembre o 40 en abril?

Usar las siguientes sumas para los cálculos:

$\sum x_i n_i = 544$ pacientes, $\sum x_i^2 n_i = 10464$ pacientes², $\sum (x_i - \bar{x})^3 n_i = 736.14$ pacientes³ y $\sum (x_i - \bar{x})^4 n_i = 25367.44$ pacientes⁴.

Solución

1. $Q_1 = 16$ pacientes, $Q_3 = 20$ pacientes y $RI = 4$ pacientes. Por tanto, la dispersión central es pequeña.

2. $\bar{x} = 18.1333$ pacientes, $s^2 = 19.9822$ pacientes², $s = 4.4701$ pacientes y $cv = 0.2465$. Por tanto, la dispersión con respecto a la media es pequeña y la media representa bien.
3. $g_1 = 0.2747$ y $g_2 = -0.2346$. Como los coeficientes de asimetría y apuntamiento están entre -2 y 2, podemos asumir que la muestra proviene de una población normal.
4. Sea Y el número de pacientes diarios hospitalizados durante el mes de abril. Entonces, $cv_y = 0.8779$. Como el coeficiente de variación del mes de abril es mayor que el de septiembre, la dispersión relativa es mayor en abril.
5. Septiembre: $z(20) = -2.7143$.
Abril: $z(40) = -83.9682$.
Así pues, 40 pacientes hospitalizados en abril es relativamente mayor que 20 pacientes hospitalizados en septiembre ya que su puntuación típica es mayor.

Ejercicio 11.2. El siguiente diagrama muestra la distribución de notas en tres asignaturas distintas.

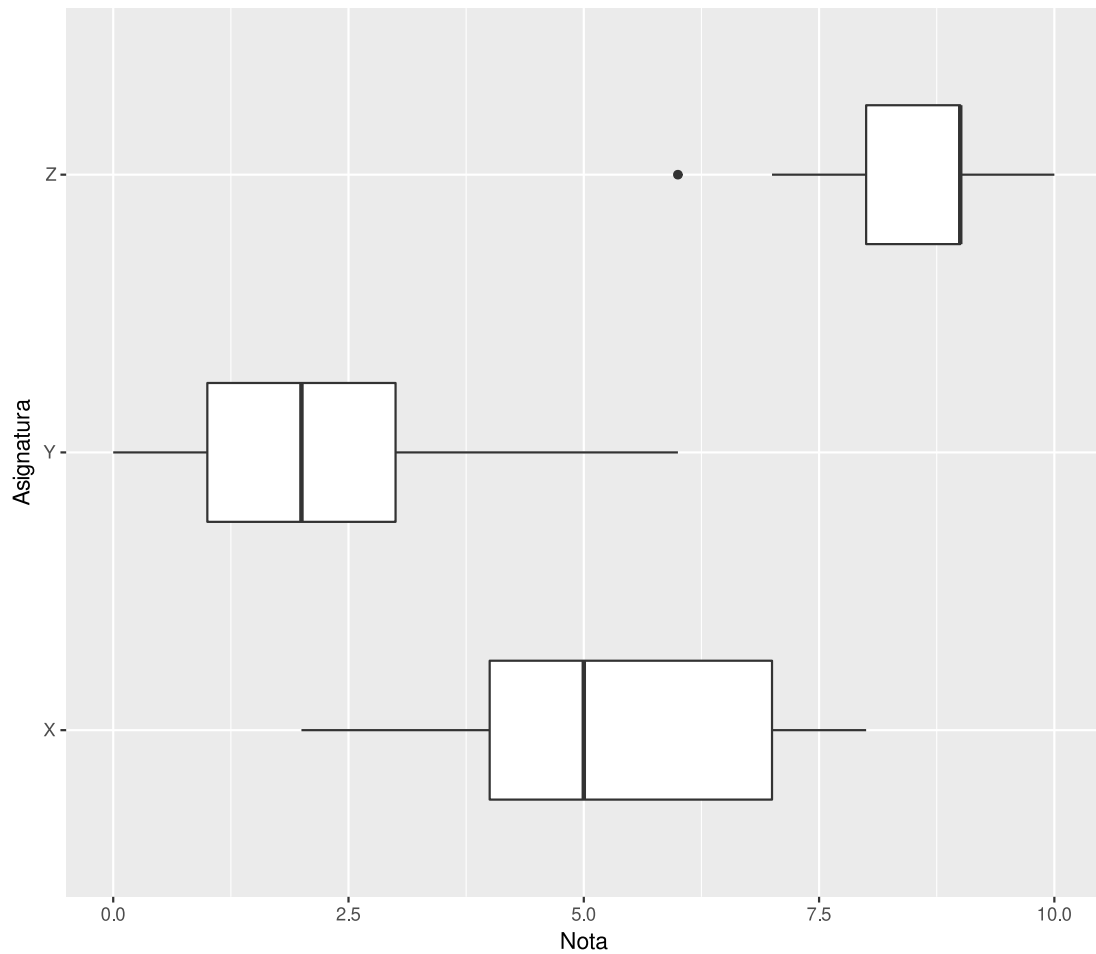


Figura 11.1.: Diagrama de caja de notas

1. ¿Qué asignatura es más difícil?
2. ¿En qué asignatura hay más variabilidad central de los datos?
3. ¿En qué asignaturas hay datos atípicos?
4. ¿Qué asignatura tiene una distribución más asimétrica?

💡 Solución

1. La asignatura Y ya que sus puntuaciones son menores (la caja y los bigotes están más desplazados hacia la izquierda).
2. La asignatura X porque la anchura de la caja es mayor.
3. La asignatura Z porque hay una nota que está fuera de los bigotes.

- La asignatura Z porque la distancia desde el primer cuartil hasta la mediana (lado izquierdo de la caja) es mayor que la distancia desde la mediana al tercer cuartil (lado derecho de la caja).

Ejercicio 11.3. Se quiere estudiar si la estatura de los hijos depende de la estatura de los padres y para ello se ha tomado una muestra de 10 familias con un hijo mayor de 20 años y se ha medido la estatura del padre (X), de la madre (Y) y del hijo (Z) en centímetros, obteniendo los siguientes resultados:

$$\begin{aligned}\sum x_i &= 1774 \text{ cm}, \sum y_i = 1630 \text{ cm}, \sum z_i = 1795 \text{ cm}, \\ \sum x_i^2 &= 315300 \text{ cm}^2, \sum y_i^2 = 266150 \text{ cm}^2, \sum z_i^2 = 322737 \text{ cm}^2, \\ \sum x_i y_i &= 289364 \text{ cm}^2, \sum x_i z_i = 318958 \text{ cm}^2, \sum y_i z_i = 292757 \text{ cm}^2.\end{aligned}$$

Se pide:

- ¿De qué estatura depende más linealmente la estatura del hijo, de la estatura del padre o de la madre?
- Utilizando el mejor modelo lineal, predecir la estatura de un hijo cuyo padre mide 181 cm y cuya madre mide 163 cm.
- ¿Cuánto aumentará la estatura del hijo por cada centímetro que aumente la estatura del padre? ¿Y de la madre?
- ¿Cómo afectaría a la fiabilidad de los modelos que las estaturas se hubiesen medido en pulgadas? (Una pulgada son 2.54 cm).

Solución

- $\bar{x} = 177.4 \text{ cm}$, $s_x^2 = 59.24 \text{ cm}^2$,
 $\bar{y} = 163 \text{ cm}$, $s_y^2 = 46 \text{ cm}^2$,
 $\bar{z} = 179.5 \text{ cm}$, $s_z^2 = 53.45 \text{ cm}^2$,
 $s_{xz} = 69.8861 \text{ cm}^2$ y $s_{yz} = 17.2 \text{ cm}^2$.
 $r_{xz}^2 = 0.9273$ y $r_{yz}^2 = 0.1203$, de manera que la estatura de los hijos depende linealmente más de la estatura del padre ya que su coeficiente de determinación es mayor.
- Recta de regresión de Z sobre X : $z = -29.7808 + 1.1797x$ y $z(181) = 183.747 \text{ cm}$.
- La estatura del hijo aumentará 1.1797 cm por cada cm que aumente la estatura del padre y 0.3739 cm por cada cm que aumente la estatura de la madre.
- La fiabilidad sería la misma, ya que al aplicar la misma transformación lineal a X y Z , las varianzas quedan multiplicadas por el cuadrado de la pendiente

y la covarianza también queda multiplicada por el cuadrado de la pendiente.

12. 2021-01-18 Examen de Farmacia y Biotecnología

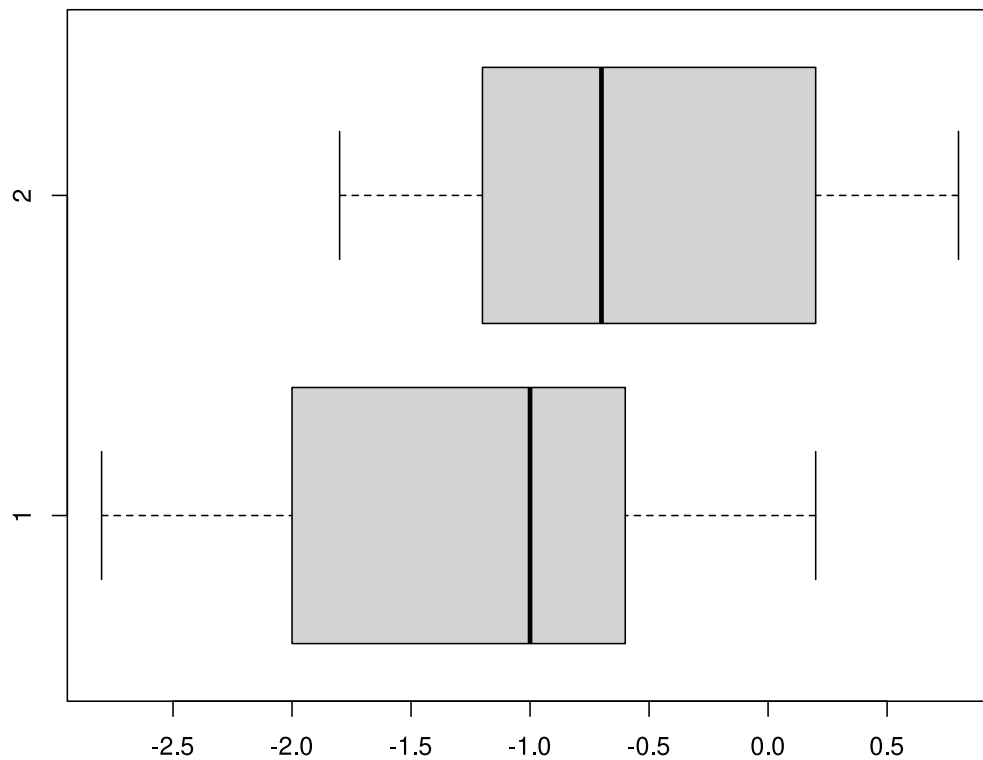
Grados: Farmacia y Biotecnología

Fecha: 18 de enero de 2021

Ejercicio 12.1. La siguiente tabla muestra las diferencias de notas entre las notas de bachillerato y las del examen de selectividad en los centros públicos (X) y privados (Y) de una región:

Centros públicos	-1.2	-0.7	-0.4	-0.9	-1.6	0.5	0.2	-1.8	0.8
Centros privados	-2.1	-0.5	-0.7	-1.9	0.2	-2.8	-1		

1. ¿Cuál de los siguientes diagramas de cajas corresponde a cada variable? Comparar la dispersión central de las dos variables según los diagramas de caja. ¿En qué variable es menor la mediana de las diferencias de notas?



2. ¿En qué centros es más representativa la media de la diferencia de notas, en los públicos o en los privados?
3. ¿En qué centros la distribución de la diferencia de notas es más simétrica?
4. ¿En qué centros la distribución de la diferencia de notas es más apuntada?
5. ¿Qué diferencia es relativamente menor, -0.5 puntos en un centro público o -1 en un centro privado?

Usar las siguientes sumas para los cálculos:

Public: $\sum x_i = -5.1$, $\sum x_i^2 = 9.63$, $\sum (x_i - \bar{x})^3 = 0.95$ y $\sum (x_i - \bar{x})^4 = 8.76$. Private: $\sum y_i = -8.8$, $\sum y_i^2 = 17.64$, $\sum (y_i - \bar{y})^3 = -0.82$ y $\sum (y_i - \bar{y})^4 = 11.28$.

Solución

1. El diagrama de cajas 1 corresponde a los centros privados y el 2 a los centros públicos. La dispersión central (anchura de las cajas) es similar en ambas variables. La mediana es menor en los centros privados.
2. Centros públicos: $\bar{x} = -0.5667$, $s^2 = 0.7489$, $s = 0.8654$ y $cv = 1.5271$. Centros privados: $\bar{y} = -1.2571$, $s^2 = 0.9396$, $s = 0.9693$ y $cv = 0.7711$. Por tanto, la media de las diferencias de notas es más representativa en los centros privados.
3. $g_{1x} = 0.1626$ y $g_{1y} = -0.1285$. Por tanto, la distribución de las diferencias de notas en los centros privados es más simétrica ya que su coeficiente de asimetría está más cerca de 0.
4. $g_{2x} = -1.2651$ y $g_{2y} = -1.1748$. Así pues, la distribución de las diferencias de notas en los centros privados es más apuntada ya que su coeficiente de apuntamiento es mayor.
5. Centro público: $z(-0.5) = 0.077$. Centro privado: $z(-1) = 0.2653$. Por tanto, una diferencia de notas de -0.5 puntos en centros públicos es relativamente menor que una diferencia de -1 puntos en centros privados.

Ejercicio 12.2. Un auditor ha estudiado la relación entre el salario y el número de ausencias de los celadores de un hospital. La tabla siguiente muestra los salarios en miles de euros (X) y el número medio de ausencias anuales con esos salarios (Y).

Salario	20.0	22.5	25	27.5	30.0	32.5	35.0	37.5	40.0
Ausencias	2.3	2.0	2	1.8	2.2	1.5	1.2	1.3	0.6

1. Calcular la recta de regresión que explique las ausencias en función del salario.
2. ¿Cuál es el número de ausencias esperado de un celador con un salario de 29000€?
¿Es esta predicción fiable?
3. ¿Cuánto aumentará o disminuirá el número de ausencias por cada incremento de 1000€ en el salario?

Usar las siguientes sumas para los cálculos:

$$\begin{aligned} \sum x_i &= 270 \cdot 10^3 \text{€}, \quad \sum y_i = 14.9 \text{ ausencias}, \\ \sum x_i^2 &= 8475 (10^3 \text{€})^2, \quad \sum y_i^2 = 27.11 \text{ ausencias}^2, \\ \sum x_i y_i &= 420 \cdot 10^3 \text{€ ausencias}. \end{aligned}$$

 Solución

1. $\bar{x} = 30 \cdot 10^3 \text{€}$, $s_x^2 = 41.6667 (10^3 \text{€})^2$,
 $\bar{y} = 1.6556$ ausencias, $s_y^2 = 0.2714$ ausencias²,
 $s_{xy} = -3 \cdot 10^3 \text{€}$ ausencias
Recta de regresión de las ausencias sobre el salario: $y = 3.8156 - 0.072x$.
2. $y(29) = 1.7276$ ausencias
 $r^2 = 0.796$, de modo que el modelo lineal se ajusta bien ya que el coeficiente de determinación no está lejos de 1, pero el tamaño muestral es demasiado pequeño para que las predicciones sean fiables.
3. El número de ausencias disminuirá 0.072 por cada incremento de 1000€ en el salario.

Ejercicio 12.3. En un estudio de regresión se sabe que la recta de regresión de Y sobre X es $y + 2x - 10 = 0$ y la recta de regresión de X sobre Y es $y + 3x - 14 = 0$.

1. Calcular las medias de X e Y .
2. Calcular el coeficiente de correlación lineal e interpretarlo.

 Solución

1. $\bar{x} = 4$ y $\bar{y} = 2$.
2. $r = -0.8165$. El coeficiente de correlación lineal está cerca de -1 lo que significa que existe una relación lineal fuerte e inversa entre X e Y .

Ejercicio 12.4. Un test para diagnosticar el cáncer de próstata produce un 1% de falsos positivos y un 0.2% de falsos negativos. Se sabe también que una población 1 cada 400 hombres sufre este tipo de cáncer.

1. Calcular la sensibilidad y la especificidad del test.
2. Si un hombre tiene un resultado positivo en el test, ¿cuál es la probabilidad de que tenga cáncer de próstata?
3. Calcular e interpretar el valor negativo predictivo del test.
4. ¿Es este teste mejor para detectar o para descartar el cáncer de próstata?
5. Para ver si existe asociación entre el cáncer de próstata y la práctica del deporte, se tomó una muestra de 1000 hombres, de los cuales 700 practicaban deporte, y se observó que había 2 hombres con cáncer de próstata en el grupo de los que practicaban deporte y 3 hombres con cáncer de próstata en el grupo de los que no.

practicaban deporte. Calcular el riesgo relativo y el odds ratio de sufrir cáncer de próstata cuando se practica deporte.

Solución

Sea C el suceso correspondiente a sufrir cáncer de próstata y $+$ y $-$ los sucesos consistentes en tener un resultado positivo y negativo en el test respectivamente.

1. La sensibilidad es $P(+|D) = 0.2$ y la especificidad $P(-|\bar{D}) = 0.99$.
2. El valor predictivo positivo es $P(D|+) = 0.0476$.
3. El valor predictivo negativo es $P(\bar{D}|-) = 0.998$.
4. Como el valor predictivo negativo es mayor que el valor predictivo positivo, el test es mejor para descartar la enfermedad que para confirmarla. De hecho el test no permite detectar la enfermedad ya que el valor predictivo positivo es menor que 0.5.
5. $RR(D) = 0.2857$ y $OR(D) = 0.2837$. Por tanto, existe una asociación entre la práctica del deporte y el cáncer de próstata, de manera que la probabilidad de sufrir cáncer de próstata cuando un hombre practica deporte es casi un cuarto de la probabilidad de sufrirlo cuando no se practica deporte, y con el odds ocurre algo similar.

Resolución

<https://www.youtube.com/watch?v=pWbsQeCsKY4>

Ejercicio 12.5. La probabilidad de que un hijo de una madre con el gen del daltonismo y un padre sin el gen del daltonismo sea un varón daltónico es 0.25.

1. Si esta pareja tiene 5 hijos, ¿cuál es la probabilidad de que a lo sumo 2 sean varones daltónicos?
2. Si esta pareja tiene 5 hijos, y el sexo de los hijos es equiprobable, ¿cuál es la probabilidad de que 3 o más sean mujeres?
3. Si se toma una muestra aleatoria de 10000 hombres de una población en la que hay un varón daltónico por cada 5000 hombres, ¿cuál es la probabilidad de que haya más de 3 varones daltónicos?

 Solución

1. Sea X el número de hijos varones daltónicos en una muestra de 5 hijos de la pareja. Entonces $X \sim B(5, 0.25)$ y $P(X \leq 2) = 0.8965$.
2. Sea Y el número de mujeres en una muestra de 5 hijos de la pareja. Entonces $Y \sim B(5, 0.5)$ y $P(Y \geq 3) = 0.5$.
3. Sea Z el número de varones daltónicos en una muestra de 1000 hombres de la población. Entonces $Z \sim B(10000, 2e - 04) \approx P(2)$ y $P(Z > 3) = 0.1429$.

 Resolución

<https://www.youtube.com/watch?v=4Si6aj1fM0k>

Ejercicio 12.6. La capacidad craneal de los primates sigue una distribución normal de media 1200 cm^3 y desviación típica 140 cm^3 .

1. Calcular la probabilidad de que la capacidad craneal de un primate sea mayor de 1400 cm^3 .
2. Calcular la probabilidad de que la capacidad craneal de un primate sea exactamente 1400 cm^3 .
3. Calcular la capacidad craneal por encima de la cual estarán el 20% of primates.
4. Calcular el rango intercuartílico de la capacidad craneal de los primates e interpretarlo.

 Solución

Sea X la capacidad craneal de los primates. Entonces $X \sim N(1200, 140)$.

1. $P(X > 1400) = 0.0766$.
2. $P(X = 1400) = 0$.
3. $P_{80} = 1317.827 \text{ cm}^3$.
4. $Q_1 = 1105.5714 \text{ cm}^3$, $Q_3 = 1294.4286 \text{ cm}^3$ y $IQR = 188.8571 \text{ cm}^3$. Por tanto, el 50% central de los datos está concentrado en un intervalo de amplitud 188.8571 cm^3 , que es poca dispersión.

 Resolución

https://www.youtube.com/watch?v=hwJ9ZQ_JDBs

Parte II.

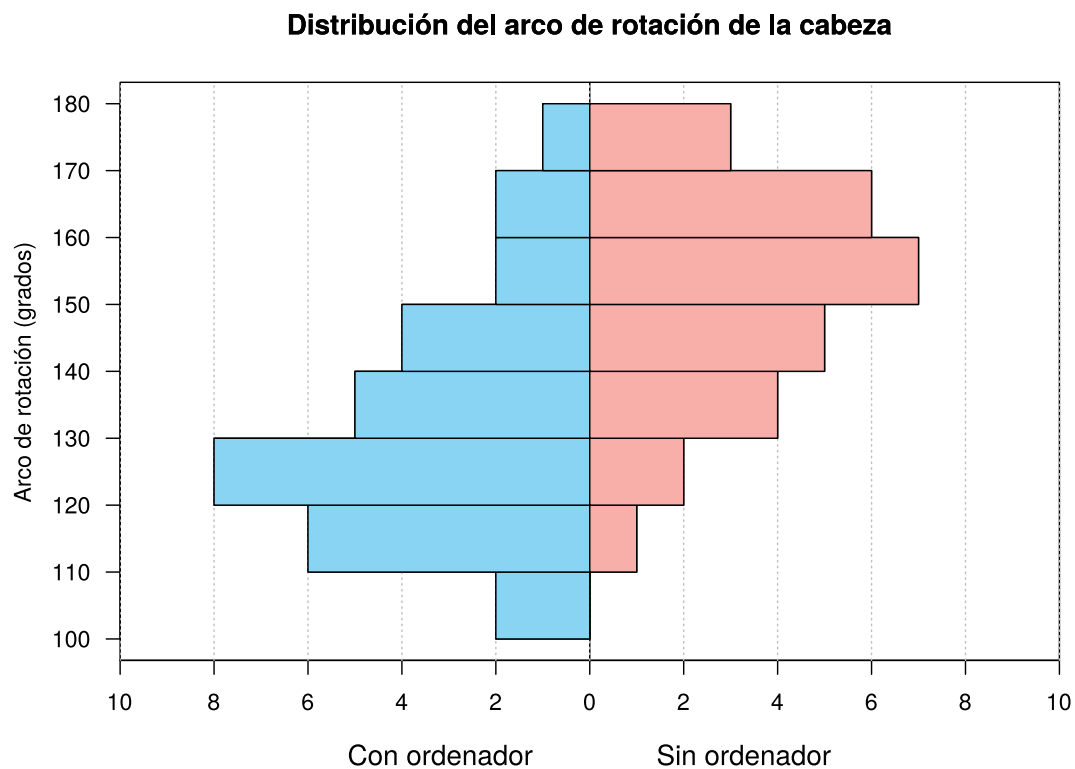
Fisioterapia

13. 2018-04-09 Examen de Fisioterapia

Grados: Fisioterapia

Fecha: 9 de abril de 2018

Ejercicio 13.1. En un estudio se ha medido el arco de rotación de la cabeza en personas que trabajan habitualmente con ordenador y en personas que no. Los resultados aparecen en el siguiente histograma.



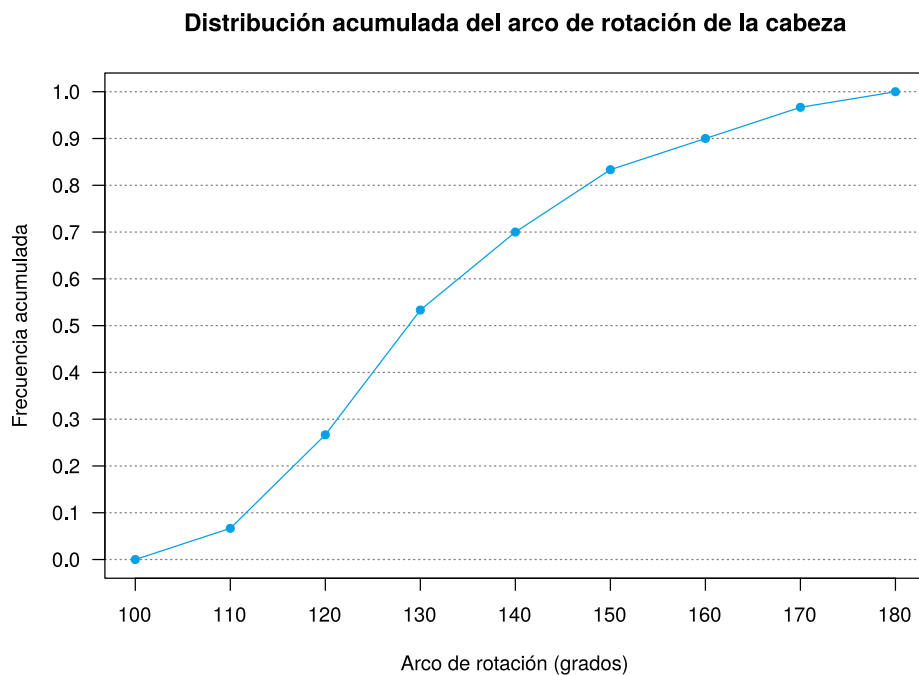
Se pide:

1. Dibujar el polígono de frecuencias relativas acumuladas del arco de rotación de la cabeza en personas que trabajan con ordenador.

2. Si se considera que una persona con un arco de rotación menor o igual de 115 grados tiene movilidad reducida, ¿qué porcentaje personas que trabajan con ordenador tienen movilidad reducida?
3. ¿En qué distribución es más representativa la media del arco de rotación de la cabeza, en la de las personas que trabajan con ordenador o en las que no?
4. Calcular la media global del arco de rotación de la cabeza.
5. ¿Qué distribución es más asimétrica, la de las personas que trabajan con ordenador o la de las que no?
6. ¿Qué valor del arco de rotación de la cabeza es relativamente mayor, 150 grados en las personas que trabajan con ordenador o 170 grados en las que no?

Usar las siguientes sumas para los cálculos. Con ordenador: $\sum x_i = 3970$ grados, $\sum x_i^2 = 534750$ grados², $\sum (x_i - \bar{x})^3 = 103662.2222$ grados³ y $\sum (x_i - \bar{x})^4 = 7903715.5556$ grados⁴. Sin ordenador: $\sum x_i = 4230$ grados, $\sum x_i^2 = 645900$ grados², $\sum (x_i - \bar{x})^3 = -42359.6939$ grados³ y $\sum (x_i - \bar{x})^4 = 4101700.5284$ grados⁴.

Solución



- 1.
2. $F(115) = 0.1667 \rightarrow 16.67\%$ de los que trabajan con ordenador tienen movilidad reducida.
3. Con ordenador: $\bar{x} = 132.3333$ grados, $s_x^2 = 312.8889$ grados², $s_x = 17.6887$ grados y $cv_x = 0.1337$

Sin ordenador: $\bar{x} = 151.0714$ grados, $s_x^2 = 245.2806$ grados², $s_x = 15.6614$ grados y $cv_x = 0.1037$

La media de los que trabajan sin ordenador es más representativa que la de los que trabajan con ordenador ya que su coeficiente de variación es menor.

4. $\bar{x} = 141.3793$.
5. Con ordenador $g_1 = 0.6243$ y sin ordenador $g_1 = -0.3938$. Por tanto, la distribución de los que trabajan con ordenador es más asimétrica.
6. Puntuaciones típicas: $z(150) = 0.9988$ y $z(170) = 1.2086$. Por tanto, un arco de rotación de 150 grados en los que trabajan con ordenador es relativamente menor que un arco de rotación de 170 en los que trabajan sin ordenador.

Ejercicio 13.2. La concentración de un fármaco en sangre C , medida en mg/dl, viene dada en función del tiempo t , en horas, según se recoge en la siguiente tabla:

Tiempo	2	3	4	5	6	7	8
Concentración	25	36	48	64	86	114	168

Se pide:

1. ¿Qué modelo de regresión te parece más fiable para predecir la concentración que habría a las 4.8 horas, el lineal o el exponencial?
2. Realiza dicha predicción con el mejor modelo de los dos anteriores.

Usar las siguientes sumas para los cálculos:

$\sum x_i = 35$, $\sum \log(x_i) = 10.6046$, $\sum y_j = 541$, $\sum \log(y_j) = 29.147$,
 $\sum x_i^2 = 203$, $\sum \log(x_i)^2 = 17.5205$, $\sum y_j^2 = 56937$, $\sum \log(y_j)^2 = 124.0131$,
 $\sum x_i y_j = 3328$, $\sum x_i \log(y_j) = 154.3387$, $\sum \log(x_i) y_j = 951.6961$, $\sum \log(x_i) \log(y_j) = 46.0805$.

Solución

1. Modelo lineal de Concentración sobre Tiempo:

$\bar{x} = 5$ horas, $s_x^2 = 4$ horas².

$\bar{y} = 77.2857$ mg/dl, $s_y^2 = 2160.7755$ (mg/dl)².

$s_{xy} = 89$ horas mg/dl.

Coeficiente de determinación lineal de Concentración sobre Tiempo
 $r^2 = 0.9165$.

Modelo exponencial de Concentración sobre tiempo:

$\overline{\log(y)} = 4.1639 \log(\text{mg/dl})$, $s_{\log(y)}^2 = 0.3785 \log(\text{mg/dl})^2$.

$s_{x \log(y)} = 1.2291$ horas $\log(\text{mg/dl})$.

Coeficiente de determinación exponencial de Concentración sobre Tiempo

$$r^2 = 0.9979.$$

Por tanto, el modelo exponencial explica mejor la relación entre la concentración y el tiempo ya que su coeficiente de determinación es mayor.

2. Modelo exponencial de Concentración sobre Tiempo: $y = e^{2.6275+0.3073x}$.
 $y(4.8) = 60.4853$ mg/dl.

14. 2018-05-31 Examen de Fisioterapia

Grados: Fisioterapia

Fecha: 31 de mayo de 2018

Ejercicio 14.1. 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:

1. Calcular los cuartiles.
2. Dibujar el diagrama de cajas e identificar los datos atípicos (no agrupar los datos en intervalos).
3. Considerando los grupos de los menores y mayores de 65 años, ¿en cuál de ellos es más representativa la media?
4. ¿Qué distribución es menos simétrica, la de los menores o la de los mayores de 65 años?
5. ¿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.

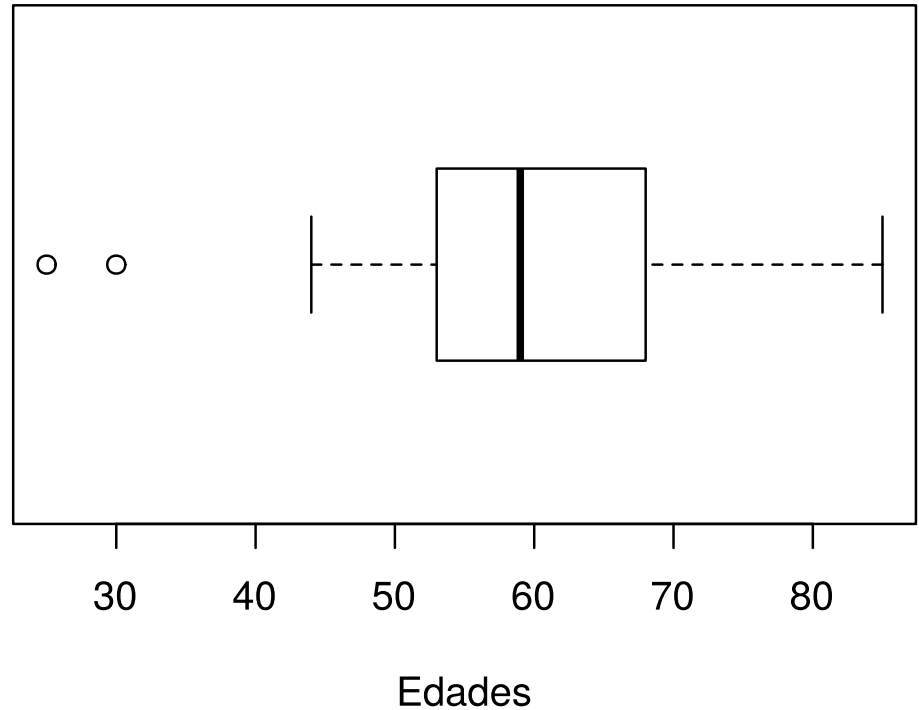
Menores de 65: $\sum x_i = 953$ años, $\sum x_i^2 = 52475$ años², $\sum (x_i - \bar{x})^3 = -30846.51$ años³ y $\sum (x_i - \bar{x})^4 = 939658.83$ años⁴.

Mayores de 65: $\sum x_i = 588$ años, $\sum x_i^2 = 43530$ años², $\sum (x_i - \bar{x})^3 = 1485$ años³ y $\sum (x_i - \bar{x})^4 = 26983.5$ años⁴.

Solución

1. $C_1 = 53$ años, $C_2 = 59$ años and $C_3 = 68$ años.
2. Existen dos datos atípicos: 25 y 30.

Diagrama de caja de las edades de los pacientes



3. Sea X la edad de los pacientes menores de 65 años e Y la de los mayores.
 $\bar{x} = 52.9444$ años, $s_x^2 = 112.1636$ años², $s_x = 10.5907$ años y $cv_x = 0.2$.
 $\bar{y} = 73.5$ años, $s_y^2 = 39$ años², $s_y = 6.245$ años y $cv_y = 0.085$.
 La media es más representativa en los pacientes mayores de 65 ya que el coeficiente de variación es menor.
4. $g_{1x} = -1.4426$ y $g_{1y} = 0.7621$, de manera que la distribución menos simétrica es la de los menores de 65 años.
5. Las puntuaciones típicas son $z_x(50) = -0.278$ y $z_y(72) = -0.2402$, de manera que una edad de 50 años es relativamente menor en el grupo de los menores de 65 años.

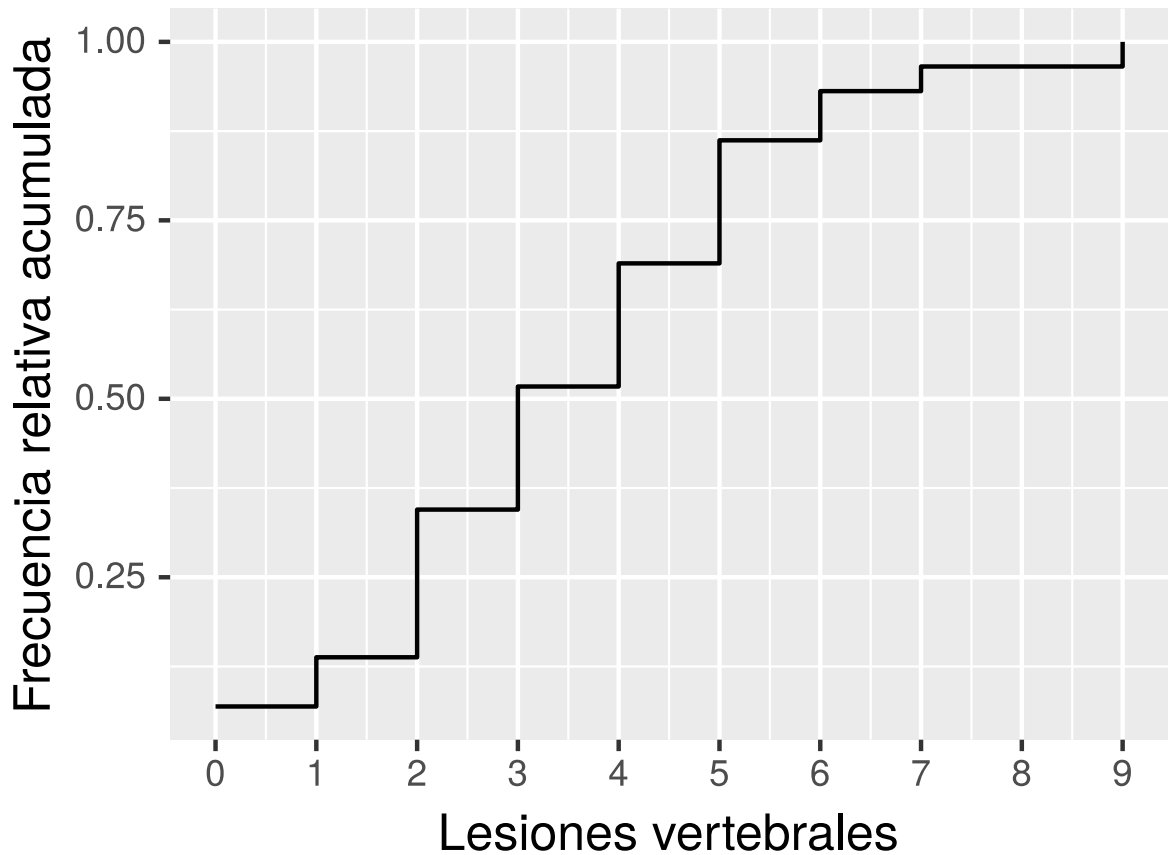
Ejercicio 14.2. La siguiente tabla recoge el número de lesiones en un equipo durante una temporada y el número medio de minutos diarios de calentamiento que hacen sus jugadores.

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:

1. Dibujar el diagrama de dispersión.
2. ¿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.
3. ¿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.
4. ¿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): $\sum x_i = 237$, $\sum \log(x_i) = 31.3728$, $\sum y_j = 157$, $\sum \log(y_j) = 24.0775$, $\sum x_i^2 = 5937$, $\sum \log(x_i)^2 = 98.9906$, $\sum y_j^2 = 3843$, $\sum \log(y_j)^2 = 66.3721$, $\sum x_i y_j = 3115$, $\sum x_i \log(y_j) = 519.1907$, $\sum \log(x_i) y_j = 465.8093$, $\sum \log(x_i) \log(y_j) = 73.3995$.



- 1.
2. $\bar{x} = 23.7 \text{ min}$, $s_x^2 = 32.01 \text{ min}^2$.
 $\overline{\log(x)} = 3.1373 \log(\text{min})$, $s_{\log(x)}^2 = 0.0565 \log(\text{min})^2$.
 $\bar{y} = 15.7 \text{ lesiones}$, $s_y^2 = 137.81 \text{ lesiones}^2$.
 $\overline{\log(y)} = 2.4078 \log(\text{lesiones})$, $s_{\log(y)}^2 = 0.8399 \log(\text{lesiones})^2$.
 $s_{x \log(y)} = -5.1446$, $s_{\log(x) y} = -2.6744$
 Coeficiente de determinación 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: $y(20) = 20.1341 \text{ lesiones}$.
3. 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.

4. 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.

Ejercicio 14.3. Se recurre a cierta técnica con ultrasonidos en el proceso de diagnosis de una enfermedad. Su sensibilidad es del 91% y su especificidad del 98%. Sabiendo que la prevalencia de dicha enfermedad es del 20%, se pide:

1. Si a un individuo se le aplica la técnica y el resultado es positivo, ¿cuál es la probabilidad de que sufra esta enfermedad?
2. Si el resultado fuese negativo, ¿cuál sería la probabilidad de que no tuviera la enfermedad?
3. ¿La técnica es más fiable para confirmar la enfermedad o para descartarla? Justificar la respuesta.
4. Calcular la probabilidad de obtener un diagnóstico acertado con esta técnica.

Solución

Sea E el suceso consistente en tener la enfermedad y $+$ y $-$ los sucesos correspondientes a obtener un resultado positivo y negativo respectivamente en el test.

1. $VPP = 0.9192$.
2. $VPN = 0.9776$.
3. Es más fiable para descartar la enfermedad ya que el valor predictivo negativo es mayor que el valor predictivo positivo.
4. $P(E \cap +) + P(\bar{E} \cap -) = 0.966$.

Ejercicio 14.4. Se sabe que la longitud del fémur de un feto a las 25 semanas de embarazo sigue una distribución normal de media 44 mm y desviación típica 2 mm. Se pide:

1. Calcular la probabilidad de que, tomando un feto de 25 semanas al azar, el fémur mida más de 46 mm.
2. Calcular la probabilidad de que, tomando un feto de 25 semanas al azar, el fémur mida entre 46 y 49 mm.
3. Determina un intervalo (a, b) centrado en la media, que contenga el 80% de los valores de la longitud del fémur de fetos de 25 semanas.

 Solución

Sea $X \sim N(44, 2)$ la longitud del fémur de un feto a las 25 semanas de embarazo.

1. $P(X > 46) = 0.1587$.
2. $P(46 < X < 49) = 0.1524$.
3. El intervalo centrado en 44 que contiene 80% de las longitudes del fémur de fetos de 25 semanas es $(41.4369, 46.5631)$.

Ejercicio 14.5. La probabilidad de que una lesión A se reproduzca es $4/5$, la de que se reproduzca otra lesión B es $1/2$, y la de que ninguna se reproduzca $1/20$. Hallar la probabilidad de que:

1. Al menos una se reproduzca.
2. Sólo se reproduzca la lesión B .
3. Se reproduzca la lesión B si se ha reproducido la A .
4. Se reproduzca la lesión B si no se reproduce la lesión A .

 Solución

1. $P(A \cup B) = 19/20$.
2. $P(B \cap \bar{A}) = 3/20$.
3. $P(B/A) = 7/16$.
4. $P(B/\bar{A}) = 3/4$.

Ejercicio 14.6. Una clínica de fisioterapia abre 6 horas al día y se sabe que el número medio de pacientes por día que llegan a la clínica es 12. Se pide:

1. Calcular la probabilidad de que lleguen más de 4 pacientes en 1 horas.
2. Si la clínica tiene 4 fisioterapeutas y cada uno puede atender a un paciente por hora, ¿cuál es la probabilidad de que un día cualquiera haya alguna hora en la que algún paciente no pueda ser atendido? ¿Cuántos empleados debería haber para asegurarse de que esta probabilidad es menor del 10%?

 Solución

1. Sea X el número de pacientes que llegan en 1 horas. $X \sim P(2)$ y $P(X > 4) = 0.0527$.
2. Sea Y el número de horas en un día en las que algún paciente no puede ser atendido. $Y \sim B(6, 0.0527)$ y $P(Y > 0) = 0.2771$.
Se necesitan 5 empleados para que esta probabilidad sea menor del 10%, ya que $P(X > 5) = 0.0527$ y $P(Y > 0) = 0.0954$, siendo ahora $Y \sim B(6, 0.0166)$.

15. 2018-06-18 Examen de Fisioterapia

Grados: Fisioterapia

Fecha: 18 de junio de 2019

Ejercicio 15.1. En un estudio sobre la efectividad de un programa de prevención de riesgos laborales en oficios que requieren estar sentados muchas horas, se tomó una muestra aleatoria de individuos entre 40 y 50 años que pasaban más de 5 horas sentados y se observó si habían seguido las recomendaciones del programa de prevención o no y el número de lesiones vertebrales transcurridos 10 años. Los resultados obtenidos aparecen en la siguiente tabla.

Con programa de prevención	1	3	2	4	4	0	2	4	2	2	5	2	3	2	0
Sin programa de prevención	6	3	1	3	7	6	5	5	9	5	5	4	4	3	

Se pide:

1. Dibujar el polígono de frecuencias relativas acumuladas de la muestra global.
2. Según el rango intercuartílico, ¿en qué muestra hay una mayor dispersión central del número de lesiones vertebrales, en la de los que siguieron el programa de prevención o en la de los que no?
3. ¿En qué muestra hay una mayor dispersión relativa del número de lesiones vertebrales, en la de los que siguieron el programa de prevención o en la de los que no?
4. ¿Qué muestra tienen un apuntamiento más normal del número de lesiones vertebrales, la de los que siguieron el programa de prevención o en la de los que no?
5. ¿Qué número de lesiones vertebrales es relativamente mayor, 2 lesiones siguiendo el programa de prevención o 4 sin seguirlo?

Usar las siguientes sumas para los cálculos:

Siguiendo el programa de prevención: $\sum x_i = 36$ lesiones, $\sum x_i^2 = 116$ lesiones², $\sum (x_i - \bar{x})^3 = -0.48$ lesiones³ y $\sum (x_i - \bar{x})^4 = 135.97$ lesiones⁴.

No siguiendo el programa de prevención: $\sum y_i = 66$ lesiones, $\sum y_i^2 = 362$ lesiones², $\sum (y_i - \bar{y})^3 = 27.92$ lesiones³ y $\sum (y_i - \bar{y})^4 = 586.9$ lesiones⁴.

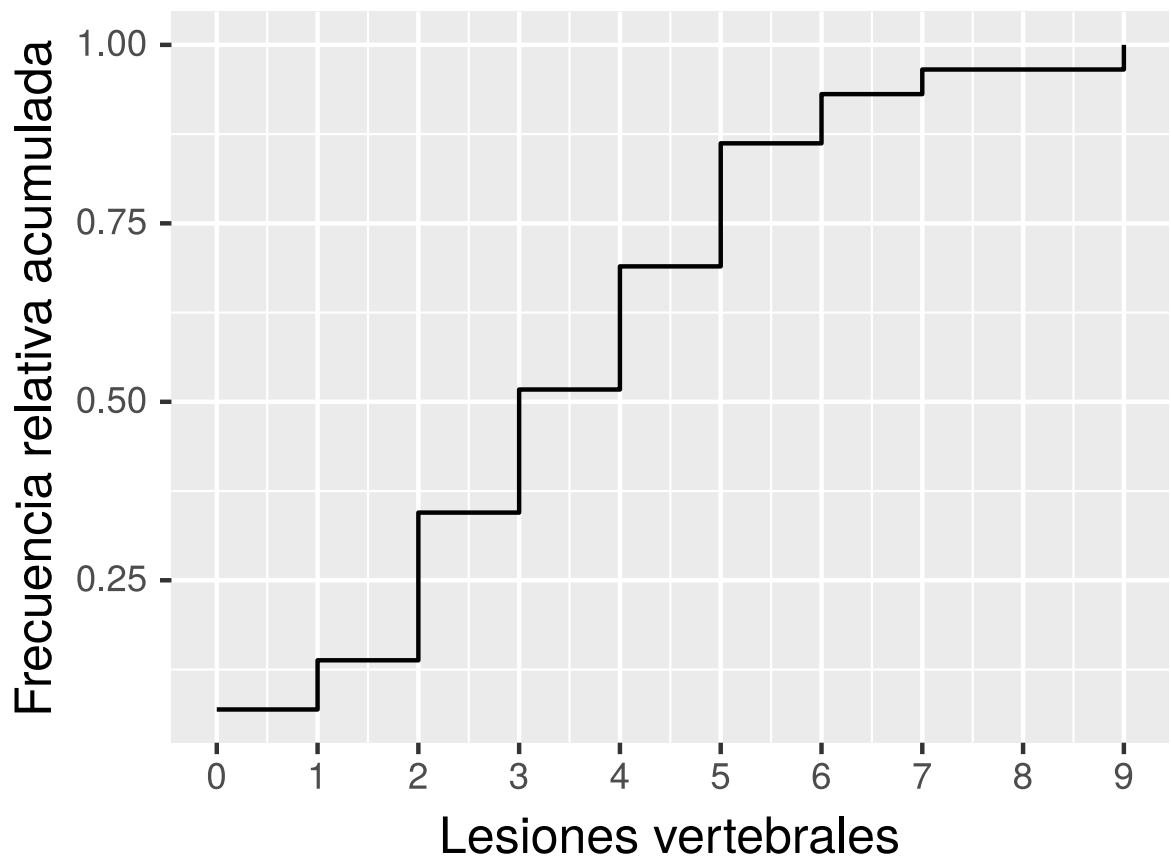


Figura 15.1.: Polígono de frecuencias relativas acumuladas de lesiones vertebrales

- 1.
2. Con programa de prevención: $C_1 = 2$ lesiones, $C_3 = 4$ lesiones, $RI = 2$ lesiones.
Sin programa de prevención: $C_1 = 3$ lesiones, $C_3 = 6$ lesiones, $RI = 3$ lesiones.
La muestra que no siguió el programa de prevención tiene una dispersión central mayor ya que su rango intercuartílico es mayor.
3. Con programa de prevención: $\bar{x} = 2.4$ lesiones, $s^2 = 1.9733$ lesiones², $s = 1.4048$ lesiones and $cv = 0.5853$.
Sin programa de prevención: $\bar{y} = 4.7143$ lesiones, $s^2 = 3.6327$ lesiones², $s = 1.906$ lesiones and $cv = 0.4043$.

La muestra que siguió el programa de prevención tiene una dispersión relativa con respecto a la media mayor ya que su coeficiente de variación es mayor.

4. Con programa de prevención: $g_2 = -0.6722$.
Sin programa de prevención: $g_2 = 0.1768$.
Así pues, la muestra que no siguió el programa de prevención tiene un apuntamiento más normal ya que el coeficiente de apuntamiento está más próximo a 0.
5. Con programa de prevención: $z(2) = -0.2847$.
Sin programa de prevención: $z(4) = -0.3748$.
Así pues, 4 lesiones sin seguir el programa de prevención es relativamente menor que 2 lesiones siguiendo el programa ya que su puntuación típica es menor.

Ejercicio 15.2. El precio de cierto relajante muscular evolucionó entre 2015 y 2019 como indica la siguiente tabla:

Año	2015	2016	2017	2018	2019
Precio (€)	1.40	1.60	1.92	2.30	2.91

1. ¿Qué modelo de regresión es mejor para predecir el precio del función del año, el lineal o el exponencial?
2. Utilizar el mejor de los dos modelos anteriores para predecir el precio del medicamento en 2020.

Solución

1. $\bar{x} = 2017$ años, $s_x^2 = 2$ años².
 $\bar{y} = 2.026$ €, $s_y^2 = 0.2882$ €².
 $\overline{\log(y)} = 0.672 \log(\text{€})$, $s_{\log(y)}^2 = 0.0673 \log(\text{€})^2$.
 $s_{xy} = 0.744$ años·€, $s_{x \log(y)} = 0.3653$ años·log(€).
Coeficiente de determinación lineal: $r^2 = 0.9603$.
Coeficiente de determinación exponencial: $r^2 = 0.9909$.
Así pues, el modelo de regresión exponencial es mejor para predecir el precio ya que su coeficiente de determinación es mayor.
2. Modelo de regresión exponencial: $y = e^{-367.6861+0.1826x}$.
Predicción: $y(2020) = 3.3867$ €.

Ejercicio 15.3. En un problema de regresión lineal entre dos variables X e Y se conoce $\bar{x} = 3$, $s_x^2 = 2$, $s_y^2 = 10.8$ y la ecuación de la recta de regresión de Y sobre X es $y = 90.9 - 2.3x$.

1. Calcular la media de Y .
2. Calcular e interpretar el coeficiente de correlación lineal.

 Solución

1. $\bar{y} = 84$.
2. $r = -0.9898$.

Ejercicio 15.4. En un estudio sobre la efectividad de un programa de prevención de riesgos laborales en oficios que requieren estar sentados muchas horas, se tomó una muestra aleatoria 500 de individuos entre 40 y 50 años sin lesiones vertebrales que pasaban más de 5 horas sentados. La mitad de los individuos siguieron un programa de prevención de riesgos laborales mientras que los demás no. Transcurridos 5 años el número de personas que desarrollaron alguna lesión vertebral en el grupo de los que siguieron el programa de prevención fue de 12, mientras que en el otro grupo fue de 32. En los siguientes 5 años hubo otras 21 personas que desarrollaron alguna lesión vertebral en el grupo de los que siguieron el programa de prevención, mientras que en el otro grupo fue de 48.

1. Calcular la incidencia acumulada de lesiones vertebrales en la muestra total a los 5 y a los 10 años.
2. Calcular el riesgo absoluto de desarrollar alguna lesión vertebral a los 10 años en el grupo de los que siguieron el programa de prevención y en el de los que no.
3. Calcular el riesgo relativo de desarrollar alguna lesión vertebral a los 10 años de los que siguieron el programa de prevención de riesgos con respecto a los que no e interpretarlo.
4. Calcular el odds ratio de desarrollar alguna lesión vertebral a los 10 años de los que siguieron el programa de prevención de riesgos con respecto a los que no e interpretarlo.
5. ¿Cuál de las dos medidas anteriores, riesgo relativo u odds ratio, tiene más sentido en este estudio? Justificar la respuesta.

 Solución

Sea E el evento consistente en sufrir una lesión vertebral.

1. Incidencia acumulada después de 5 años: $R(E) = 0.088$.
Incidencia acumulada después de 10 años: $R(E) = 0.226$.
2. Riesgo en el grupo tratamiento: $R_T(E) = 0.132$.
Riesgo en el grupo control: $R_C(E) = 0.32$.
3. $RR(D) = 0.4125$. Por tanto, el riesgo de sufrir una lesión vertebral es menos de la mitad si se sigue el programa de prevención.

4. $OR(D) = 0.3232$. Por tanto, el odd de sufrir una lesión vertebral es menos de un tercio si se sigue el programa de prevención.
5. Puesto que se trata de un estudio prospectivo se puede estimar la prevalencia de D y ambos estadísticos son válidos, pero el riesgo relativo es más fácil de interpretar.

Ejercicio 15.5. En la siguiente tabla se muestran los resultados de un estudio para evaluar la utilidad de una tira reactiva para el diagnóstico de infección urinaria:

Resultado	Con infección	Sin infección
Positivo	60	80
Negativo	10	200

1. Calcular la sensibilidad y la especificidad del test.
2. Calcular los valores predictivos positivo y negativo del test. ¿El test es mejor para confirmar la enfermedad o para descartarla?
3. Si a partir de un estudio de prevalencia efectuado previamente conociéramos que la verdadera prevalencia de la infección urinaria en la población es del 2%, ¿cómo se verían afectados los valores predictivos?

Solución

Sea E el suceso consistente en tener la infección urinaria y $+$ y $-$ los sucesos correspondientes a obtener un resultado positivo y negativo respectivamente en el test.

1. Sensibilidad = 0.8571 y Especificidad = 0.7143.
2. $VPP = 0.4286$ y $VPN = 0.9524$. Puesto que $VPP < VPN$ el test es mejor para descartar la enfermedad.
3. $VPP = 0.0577$ y $VPN = 0.9959$. El valor predictivo positivo decrece mucho mientras el valor predictivo negativo aumenta un poco.

Ejercicio 15.6. El tiempo de recuperación tras un tipo de lesión sigue una distribución normal con varianza 64 días. Se sabe además que el 10% de las personas con esta lesión tardan en curarse más de 80 días.

1. ¿Cuál es el tiempo esperado de curación para esta lesión?
2. ¿Qué porcentaje de individuos tardará en curarse entre 60 y 75 días?
3. Si se toma una muestra aleatoria de 12 individuos con esta lesión, ¿cuál es la probabilidad de que haya entre 9 y 11 individuos, ambos incluidos, que tarden menos de 80 días en curarse?

4. Si se toma una muestra aleatoria de 500 individuos con esta lesión, ¿cuál es la probabilidad de haya menos de 4 por encima del percentil 99 del tiempo de curación?

 Solución

Sea X el tiempo requerido para recuperarse de la lesión. Entonces $X \sim N(\mu, 8)$.

1. $\mu = 69.7476$ días.
2. $P(60 < X < 75) = 0.6327$.
3. Sea Y el número de individuos con la lesión que requieren más de 80 días para recuperarse en una muestra aleatoria de 12 individuos. Entonces $Y \sim B(12, 0.9)$ y $P(9 \leq Y \leq 11) = 0.6919$.
4. Sea Z be el número de individuos con la lesión que requieren un tiempo de recuperación por encima del percentil 99 en una muestra aleatoria de 500 individuos. Entonces $Z \sim B(500, 0.01) \approx P(5)$ y $P(Z \leq 4) = 0.265$.

Parte III.

Óptica

16. 2017-12-18 Examen de Óptica

Titulación: Grado en Óptica

Fecha: 18 de Diciembre de 2017

Ejercicio 16.1. En un examen de estadística al que se han presentado 66 alumnos se ha contado el número de exámenes finalizados cada media hora, obteniendo el siguiente polígono:

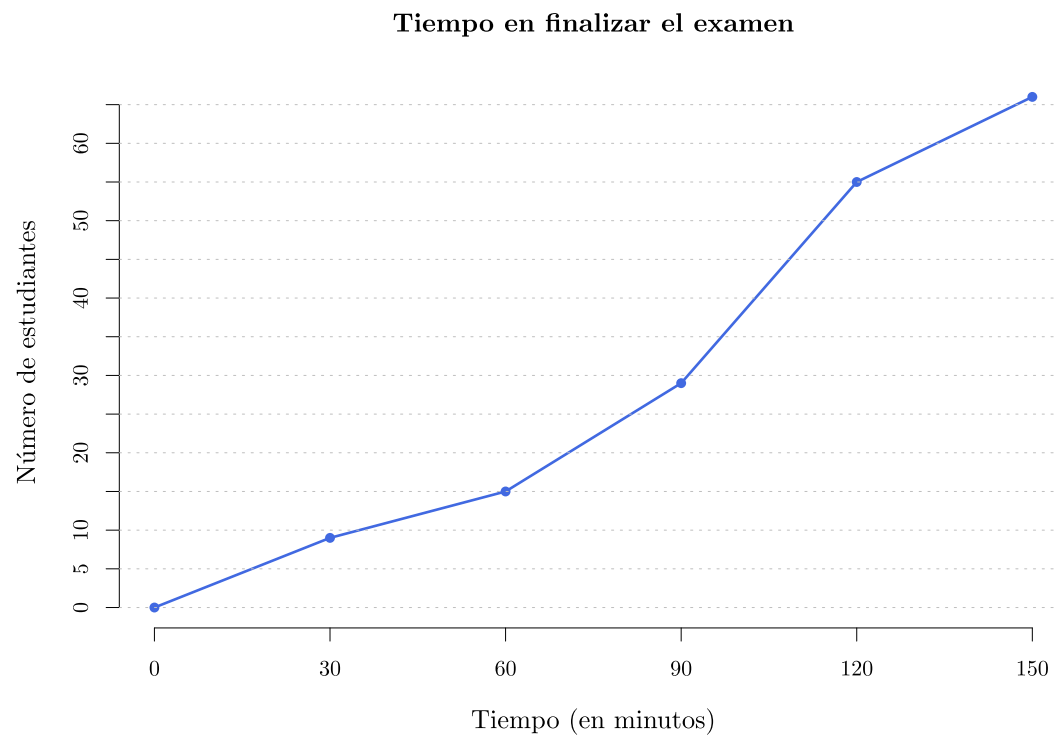


Figura 16.1.: polígono de frecuencias acumuladas del tiempo de finalización de un examen

Se pide:

1. Construir la tabla de frecuencias del tiempo de finalización del examen.

- ¿Cuánto tiempo tiene que pasar para que hayan finalizado el examen la mitad de los alumnos?
- ¿Qué porcentaje de alumnos habrá terminado a los 100 minutos de examen?
- ¿Cuál es el tiempo de duración del examen que mejor representa a los estudiantes de la muestra? ¿Es un valor representativo?
- Según su simetría y apuntamiento, ¿se puede afirmar que la distribución del tiempo de duración del examen es normal?
- Se sabe que en otro examen diferente el tiempo medio de finalización es de 100 minutos y la desviación típica es 20 minutos. Un alumno tardó 110 minutos en el examen de Estadística y 122 minutos en este otro examen. ¿Qué examen finalizó antes en términos relativos?

Utilizar las siguientes sumas para los cálculos: $\sum x_i = 5670$ min, $\sum x_i^2 = 580050$ min², $\sum (x_i - \bar{x})^3 = -2155537.1901$ min³ y $\sum (x_i - \bar{x})^4 = 311877385.4245$ min⁴.

Solución

1.

X	x_i	n_i	f_i	N_i	Fi
0 – 30	15	9	0.1364	9	0.1364
30 – 60	45	6	0.0909	15	0.2273
60 – 90	75	14	0.2121	29	0.4394
90 – 120	105	26	0.3939	55	0.8333
120 – 150	135	11	0.1667	66	1

2. $Me = 94.6154$ min.

3. $F(100) = 0.5708 \Rightarrow 57.08\%$ de los estudiantes.

4. $\bar{x} = 85.9091$ min, $s^2 = 1408.2645$ min², $s = 37.5268$ min y $cv = 0.4368$, luego la dispersión relativa con respecto a la media es moderada y la representatividad de la media también.

5. $g_1 = -0.618$, de manera que la distribución es asimétrica hacia la izquierda. $g_2 = -0.6173$, de manera que la distribución es menos apuntada de lo normal (leptocúrtica). Como tanto g_1 como g_2 están entre -2 y 2 podemos asumir que la muestra proviene de una población normal.

6. Examen de Estadística: $z(110) = 0.642$.

Otro examen: $z(122) = 1.1$.

Así pues, terminó antes el examen de Estadística.

Ejercicio 16.2. En un estudio sobre la presbicia se tomó una muestra de pacientes y se midió la edad y la distancia mínima a la que eran capaces de leer una frase en cm. Los resultados se muestran en la siguiente tabla:

Edad	25	46	76	32	18	43	40	51	68	54	47	37
Distancia mínima de enfoque	15	28	51	23	14	28	25	36	47	42	31	29

Se pide:

1. Calcular la recta de regresión de la distancia mínima de enfoque con respecto a la edad. ¿Cuánto aumenta la distancia mínima de enfoque por cada año que pasa?
2. Calcular el modelo de regresión logarítmico de la distancia mínima de enfoque con respecto a la edad.
3. Utilizar el mejor modelo de regresión de los dos anteriores para predecir la distancia mínima de enfoque a los 65 años. ¿Es fiable esta predicción?
4. Si la presbicia se diagnostica cuando la distancia mínima de enfoque es de 35 cm, ¿a qué edad aparecerá la presbicia?

Utilizar las siguientes sumas para los cálculos (X = Edad e Y = Distancia):

$$\begin{aligned} \sum x_i &= 537, \sum \log(x_i) = 44.7858, \sum y_j = 369, \sum \log(y_j) = 40.2703, \sum x_i^2 = 27033, \\ \sum \log(x_i)^2 &= 168.9328, \sum y_j^2 = 12815, \sum \log(y_j)^2 = 136.9234, \sum x_i y_j = 18561, \\ \sum x_i \log(y_j) &= 1872.0711, \sum \log(x_i) y_j = 1425.5363, \sum \log(x_i) \log(y_j) = 152.0296. \end{aligned}$$

Solución

1. Modelo de regresión de la distancia mínima de enfoque sobre la edad:
 $\bar{x} = 44.75$ años, $s_x^2 = 250.1875$ años².
 $\bar{y} = 30.75$ cm, $s_y^2 = 122.3542$ cm².
 $s_{xy} = 170.6875$ años cm.
Recta de regresión de la distancia mínima de enfoque sobre la edad: $y = 0.2198 + 0.6822x$.
Por cada año que pasa la distancia mínima de enfoque aumenta 0.6822 cm.
2. $\overline{\log(x)} = 3.7322 \log(\text{años})$, $s_{\log(x)}^2 = 0.1488 \log(\text{años})^2$.
 $s_{\log(x)y} = 4.031 \log(\text{años})$ cm.
Modelo de regresión logarítmico de la distancia mínima de enfoque sobre la edad: $y = -70.3708 + 27.0945 \log(x)$.
3. Coeficiente de determinación lineal: $r^2 = 0.9517$.
Coeficiente de determinación logarítmico: $r^2 = 0.8926$.
Por tanto, el modelo lineal explica mejor la relación entre la distancia mínima de enfoque y la edad, ya que su coeficiente de determinación es mayor.
 $y(65) = 44.5653$ cm.
4. Recta de regresión de la edad sobre la distancia mínima de enfoque: $x = 1.8529 + 1.395y$.

$$x(35) = 50.6789 \text{ años.}$$

17. 2018-10-24 Examen de Óptica

Titulación: Grado en Óptica

Fecha: 24 de Octubre de 2018

Ejercicio 17.1. En un estudio sobre la presbicia se tomó una muestra de pacientes y se midió la edad y la distancia mínima a la que eran capaces de leer una frase en cm. Los datos se muestran en la siguiente tabla.

Edad	7	12	15	15	18	21	25	28	32	35	43	46	48	51	54	57	60	66	72	92
Distancia	13	14	12	14	13	14	13	14	16	13	18	19	22	22	26	25	27	28	29	36

Utilizar las siguientes sumas para los cálculos (X = Edad e Y = Distancia mínima de enfoque):

Menores de 40

$\sum x_i = 208$ años, $\sum x_i^2 = 5066$ años², $\sum (x_i - \bar{x})^3 = 993.84$ años³, $\sum (x_i - \bar{x})^4 = 103981.55$ años⁴,

$\sum y_j = 136$ cm, $\sum y_j^2 = 1860$ cm², $\sum (y_j - \bar{y})^3 = 9.12$ cm³, $\sum (y_j - \bar{y})^4 = 40.35$ cm⁴,
 $\sum x_i y_j = 2861$ años·cm.

Mayores de 40

$\sum x_i = 589$ años, $\sum x_i^2 = 36639$ años², $\sum (x_i - \bar{x})^3 = 30793.08$ años³, $\sum (x_i - \bar{x})^4 = 1342559.42$ años⁴,

$\sum y_j = 252$ cm, $\sum y_j^2 = 6604$ cm², $\sum (y_j - \bar{y})^3 = 665.76$ cm³, $\sum (y_j - \bar{y})^4 = 18260.51$ cm⁴,

$\sum x_i y_j = 15523$ años·cm.

Se pide:

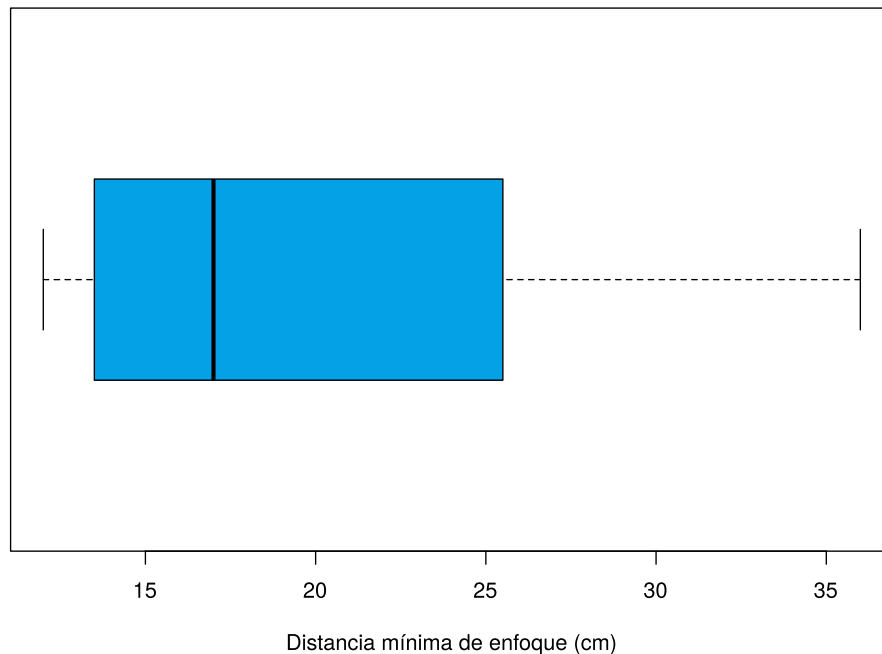
1. Dibujar el diagrama de caja y bigotes de la distancia mínima de enfoque. ¿Existen datos atípicos?
2. ¿En qué distribución de la distancia mínima de enfoque es más representativa la media, en la de menores o en la de mayores de 40 años?
3. ¿Qué distribución de la distancia mínima de enfoque es más asimétrica, la de los menores o la de los mayores de 40 años?
4. ¿Qué distancia mínima de enfoque es relativamente menor, una distancia de 12 cm en los menores de 40 años, o una distancia de 30 cm en los mayores de 40 años?

5. Dibujar el diagrama de dispersión de la edad y la distancia mínima de enfoque. Según del diagrama, ¿existe relación lineal entre la distancia mínima de enfoque y la edad?
6. ¿En qué grupo existe una relación lineal más fuerte entre la distancia mínima de enfoque y la edad, en los menores o en los mayores de 40 años?
7. Según la recta de regresión, ¿cuánto aumenta la distancia mínima de enfoque por cada año que pasa en el grupo de los mayores de 40 años?
8. A qué edad se espera tener una distancia mínima de enfoque de 32 cm en el grupo de los mayores de 40 años? ¿Es fiable esta predicción?

💡 Solución

1.

Diagrama de caja de la distribución de la distancia mínima de enfoque



No hay datos atípicos.

2. Menores de 40: $\bar{y} = 13.6$ cm, $s_y^2 = 1.04$ cm², $s_y = 1.0198$ cm y $cv_y = 0.075$.
Mayores de 40: $\bar{y} = 25.2$ cm, $s_y^2 = 25.36$ cm², $s_y = 5.0359$ cm y $cv_y = 0.1998$.
Así pues, la media es más representativa en los menores de 40 años ya que su coeficiente de variación es menor.
3. Menores de 40: $g_1 = 0.86$
Mayores de 40: $g_1 = 0.52$

Por tanto, la distribución de los menores de 40 es más asimétrica ya que el coeficiente de asimetría está más lejos de 0.

4. Menores de 40: $z(12) = -1.57$.

Mayores de 40: $z(30) = 0.95$.

Así pues, una distancia de 12 cm en menores de 40 es relativamente menor.

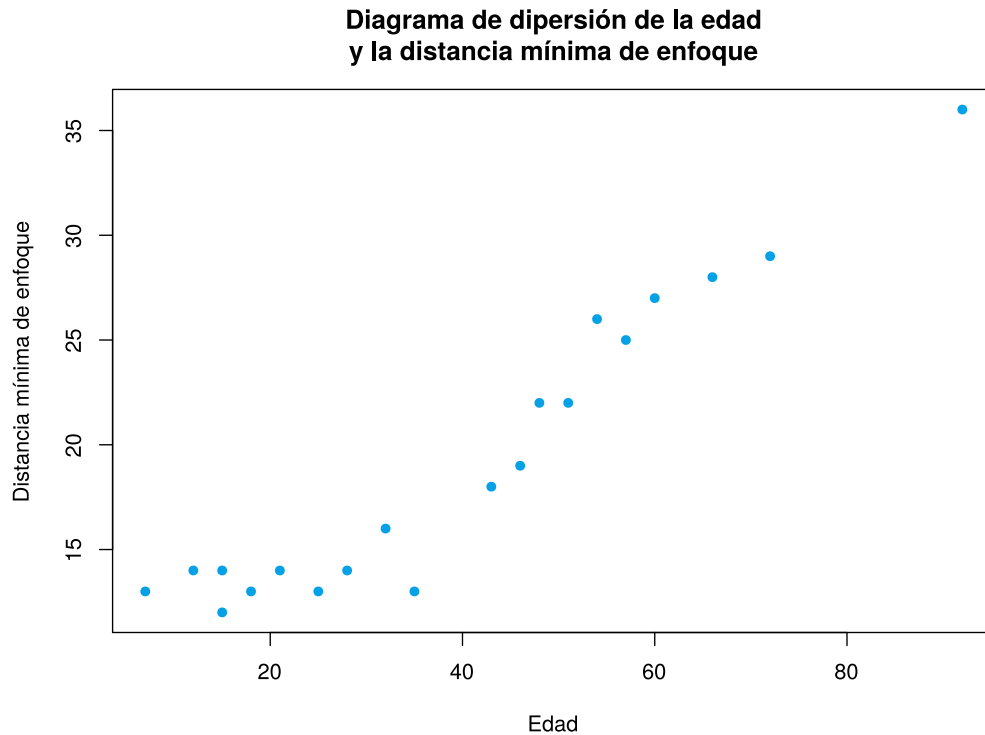


Figura 17.1.: Diagrama de dispersión de la edad y la distancia mínima de enfoque

5.

Se observan claramente dos tendencias, una para los menores de 40 y la otra para los mayores. En el caso de los menores no parece haber una relación fuerte entre la distancia mínima de enfoque y la edad, mientras que en el caso de los mayores sí parece que la hay y además es lineal.

6. Menores de 40: $\bar{x} = 20.8$ años, $s^2 = 73.96$, $s_{xy} = 3.22$ y $r^2 = 0.13$.

Mayores de 40: $\bar{x} = 58.9$ años, $s^2 = 194.69$, $s_{xy} = 68.02$ y $r^2 = 0.94$.

Por tanto, la relación lineal es más fuerte en los mayores de 40 ya que el coeficiente de determinación es mayor.

7. Recta de regresión de Y sobre X en los mayores de 40: $y = 4.6218 + 0.3494x$. Así pues, por cada año que pasa la distancia mínima de enfoque aumenta 0.3494 cm.
8. Recta de regresión de X sobre Y en los mayores de 40: $x = -8.6909 + 2.6822y$. $X(32) = 77.14$ años. Según el coeficiente de determinación la predicción es muy fiable, aunque el tamaño muestral no es muy grande y eso resta un poco de fiabilidad.

18. 2020-11-13 Examen de Óptica

Titulación: Grado en Óptica

Fecha: 13 de Noviembre de 2020

Ejercicio 18.1. En una muestra de familias se ha medido la estatura del padre (X), de la madre (Y) y de un hijo (Z) en centímetros, obteniendo los siguientes resultados:

Estatura padre	Estatura madre	Estatura hijo
175	164	177
182	175	180
190	165	193
165	160	172
172	155	173
183	172	188
187	160	185
174	151	177
168	165	168
178	163	182

Se pide:

1. ¿En qué grupo es más representativa la media, en el de padres o en el de madres?
2. ¿Hay alguna estatura atípica entre los hijos?
3. Según su forma, ¿puede provenir la muestra de madres de una distribución normal?
4. ¿Quién es más alto dentro de su grupo, una madre con una altura de 165 cm o un hijo con una altura de 178 cm?
5. ¿Cómo afectaría a la representatividad de las medias que las alturas se midiesen en metros en vez de en centímetros?

Usar las siguientes sumas para los cálculos:

Estatura padre: $\sum x_i = 1774$ cm, $\sum x_i^2 = 315300$ cm², $\sum (x_i - \bar{x})^3 = 210.48$ cm³ y $\sum (x_i - \bar{x})^4 = 67596.27$ cm⁴.

Estatura madre: $\sum y_i = 1630$ cm, $\sum y_i^2 = 266150$ cm², $\sum (y_i - \bar{y})^3 = 180$ cm³ y $\sum (y_i - \bar{y})^4 = 52324$ cm⁴.

Estatura hijo : $\sum z_i = 1795$ cm, $\sum z_i^2 = 322737$ cm², $\sum (z_i - \bar{z})^3 = 1008$ cm³ y $\sum (z_i - \bar{z})^4 = 61906.62$ cm⁴.

Solución

1. Padres: $\bar{x} = 177.4$ cm, $s^2 = 59.24$ cm², $s = 7.6968$ cm y $cv = 0.0434$.
Madres: $\bar{y} = 163$ cm, $s^2 = 46$ cm², $s = 6.7823$ cm y $cv = 0.0416$.
La estatura media es un poco más representativa en el grupo de las madres.
2. Las vallas en la muestra de hijos son $f_1 = 155$ cm y $f_2 = 203$ cm por lo que no hay estaturas atípicas entre los hijos.
3. $g_{1y} = 0.0577$ y $g_{2y} = -0.5272$. Como el coeficiente de asimetría y el de apuntamiento están dentro del intervalo de -2 a 2, podemos asumir que la muestra de estaturas de madres proviene de una población normal.
4. Puntuación típica madres: $z(165) = 0.2949$.
Puntuación típica hijos: $z(178) = -0.2052$.
Así pues, una madre de 165 cm es relativamente más alta que un hijo de 178 cm.
5. La representatividad de las medias no cambiaría ya que tanto las medias como las desviaciones típicas estarían divididas por 100.

Ejercicio 18.2. Uno de los parámetros que se suele utilizar para diagnosticar el glaucoma de ángulo abierto es la distancia mínima al borde de la abertura de la membrana de Bruch (X) de la retina, pero se sabe que esta medida depende de la edad del paciente (Y) en años y del área de la abertura de esta membrana por la que pasa el nervio óptico (Z). En un estudio se ha medido en 1000 pacientes estas variables obteniendo los siguientes resultados:

$$\begin{aligned} \sum x_i &= 346337.03 \text{ } \mu\text{m}, \sum y_i = 47212.1 \text{ años}, \sum z_i = 2002.384 \text{ mm}^2, \\ \sum x_i^2 &= 123828243.48 \text{ } \mu\text{m}^2, \sum y_i^2 = 2601264.99 \text{ años}^2, \sum z_i^2 = 4175.89 \text{ mm}^4, \\ \sum x_i y_i &= 15855138.59 \text{ } \mu\text{m}\cdot\text{años}, \sum x_i z_i = 686623.65 \text{ } \mu\text{m}\cdot\text{mm}^2, \sum y_i z_i = 94144.37 \\ &\text{años}\cdot\text{mm}^2. \end{aligned}$$

Se pide:

1. Calcular las rectas de regresión de la distancia mínima al borde de la abertura de la membrana de Bruch sobre la edad, y de la distancia mínima al borde de la abertura de la membrana de Bruch sobre el área de la abertura de la membrana.
2. ¿Cuánto aumenta o disminuye la distancia mínima al borde de la abertura de la membrana de Bruch por cada año más del paciente?
3. ¿Qué porcentaje de la variabilidad de la distancia mínima al borde de la abertura de la membrana de Bruch explica cada uno de los modelos lineales anteriores?

4. Utilizando el mejor de los modelos lineales anteriores predecir la distancia mínima al borde de la abertura de la membrana de Bruch de un paciente de 60 años con un área de la abertura de la membrana de 2 mm^2 .

 Solución

- $\bar{x} = 346.337 \text{ } \mu\text{m}$, $s_x^2 = 3878.9051 \text{ } \mu\text{m}^2$,
 $\bar{y} = 47.2121 \text{ años}$, $s_y^2 = 372.2826 \text{ años}^2$,
 $\bar{z} = 2.0024 \text{ mm}^2$, $s_z^2 = 0.1664 \text{ mm}^4$,
 $s_{xy} = -496.1599 \text{ } \mu\text{m} \cdot \text{años}$ y $s_{xz} = -6.8761 \text{ } \mu\text{m} \cdot \text{mm}^2$.
Recta de regresión de X sobre Y : $x = 409.259 + -1.3328y$.
Recta de regresión de X sobre Z : $x = 429.1056 + -41.335z$.
- La distancia mínima al borde de la abertura de la membrana de Bruch disminuye $1.3328 \text{ } \mu\text{m}$ por cada año más del paciente.
- $r_{xy}^2 = 0.1705$, de manera que la recta de regresión de X sobre Y explica el 17.05% de la variabilidad de la distancia mínima al borde de la abertura de la membrana de Bruch, y $r_{xz}^2 = 0.0733$, de manera que la recta de regresión de X sobre Z explica el 7.33% de la variabilidad de la distancia mínima al borde de la abertura de la membrana de Bruch.
- $x(60) = 329.2939 \text{ } \mu\text{m}$.