Problemas resueltos

Probabilidad y Estadística

Mar Angulo Martínez
mar.angulo@u-tad.com
2021-2022

Examen Parcial 7 diciembre 2022

Problema 1 (3,5 puntos)

Se han tomado cinco muestras con la misma cantidad de glucógeno, se les ha aplicado una cantidad de glucogenasa (en milimoles/litro): X, y se ha medido la velocidad de reacción, Y, (en micromoles/minuto). Se han obtenido los siguientes datos:

x_i	0,2	0,5	1,25	2,1	3,05
y_i	8	10	18	35	60

a) ¿Se puede deducir a partir de estos datos que la velocidad de reacción aumenta o disminuye linealmente con la concentración de glucogenasa?

En caso afirmativo dar una expresión matemática del modelo de ajuste y utilizarlo para predecir la cantidad de glucogenasa aplicada en una reacción cuya velocidad ha sido de 45 micromoles/minuto.

b) Calcular e interpretar el coeficiente de regresión de y sobre x

a) ¿Cuánta variabilidad en la velocidad de reacción no queda explicada por la concentración de glucogenasa? Indicar también qué % representa ese valor

b) Dar una medida de la bondad del ajuste e interpretarla

c) Representar en un boxplot los datos correspondientes a la velocidad de reacción. ¿Qué conclusiones puedes extraer de la gráfica?

Problema 1

Concentración de glucogenasa(xi) (milimoles/litro)	0,2	0,5	1,25	2,1	3,05	**7,1**
Velocidad de reacción(micromol/minuto) (yi)	8	10	18	35	60	**131**
x_i^2	0,04	0,25	1,5625	4,41	9,3025	**15,565**
y_i^2	64	100	324	1.225	3.600	**5.313**
xi.yi	1,6	5	22,5	73,5	183	**285,6**

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$= 7,1/5 = $1{,}42$ milimoles/litro $s_x^2 = \frac{\sum_{i=1}^{n} x_i^2}{n}$ - $\bar{x}^2$= 15,565/5-$(1{,}42)^2$=1,1 → s_x=1,047

$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$= 131/5 = 26,2 micromoles/minuto $s_y^2 = \frac{\sum_{i=1}^{n} y_i^2}{n}$ - $\bar{y}^2$= 5.313/5-$(26{,}2)^2$=376,16 → s_y=19,395

$s_{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n}$ - $\bar{x}\bar{y}$= 285,6/5- $(1{,}42x26{,}2)$= 19,916

- Para saber si realmente existe una relación lineal entre las variables calculamos el valor de r
- Esto da respuesta también el apartado d) del problema

r= $\frac{s_{xy}}{s_x \cdot s_y}$ =$\frac{19{,}916}{1{,}047x19{,}405}$ =0,98

- Interdependencia fuerte y directa: la velocidad de reacción aumenta **linealmente** a medida que aumenta la concentración de glucogenasa
- Los valores observados están muy próximos a los valores teóricos y la recta de regresión pasará muy cerca de la nube de puntos
- El ajuste es muy bueno y las predicciones serán muy fiables. La fiabilidad será de hecho del 96,08% (R^2)
- Las dos rectas de regresión forman un ángulo muy próximo a 0º.

a) El modelo de ajuste que permite predecir la cantidad de glucogenasa (X) a partir de la velocidad de reacción (Y) es la recta de regresión de X sobre Y

- Utilizaremos la recta de regresión de x sobre y (queremos predecir el valor de x)

x - $\bar{x} = \frac{s_{xy}}{s_y^2}(y-\bar{y})$ ⟶ x-1,42=$\frac{19{,}916}{376{,}56}(y-26{,}2)$ ⟶ x-1,42=$0{,}053\ (y-26{,}2)$

y la predicción es x(y=45) x=$1{,}42+0{,}053(45-26{,}2)$=2,416 milimoles/litro

c) ¿Cuánta variabilidad en la velocidad de reacción no queda explicada por la concentración de glucogenasa? Indicar también qué % representa ese valor

- La variabilidad en la velocidad de reacción (y) que NO queda explicada por la concentración de glucogenasa (x) es la varianza residual o varianza no explicada s_y^2.(1- r^2) = *376,16x(1-0,9604)=14,896 unidades de varianza*
  que representan el $r^2 = 3{,}96\%$ *de la variabilidad total del % en la velocidad de reacción.*

b) Coeficiente de regresión de Y sobre X es la pendiente de la recta de regresión de Y sobre X

- $b_{yx} = \frac{s_{xy}}{s_y^2} = \frac{19{,}916}{1{,}1} = 18{,}105\ micromoles\ /minuto$
- Por cada milimol/l más de glucogenasa, la velocidad de reacción aumenta en 18,105 micromoles por minuto

y_i	8	10	18	35	60
n_i	1	1	1	1	1

Obtenemos Q_1=10 Me=18 Q_3=35 Entonces RI=35-10=25 $Lím_{inf} = 10 - 1{,}5x25 < 0$ $Lím_{sup} = 35 + 1{,}5x25 = 72{,}5$

e) Representar en un boxplot los datos correspondientes a la velocidad de reacción. ¿Qué conclusiones puedes extraer de la gráfica?

e) Conclusiones de la gráfica

f) La distribución presenta asimetría hacia la derecha puesto que Q_3-Me>Me-Q_1; esto significa que valores pequeños de la distribución tienen frecuencias muy grandes y algunos valores grandes lo que lleva a incrementar la media. La media es mayor que la mediana.

Además comprobamos que la mayor dispersión corresponde a un valor mayor de concentración (entre 35 y 60 micromoles/ minuto). En general, la dispersión mayor corresponde a los valores grandes de concentración. En los valores más pequeños encontramos muy poca dispersión.

Problema 2

Una agencia de viajes ofrece tres tipos de destinos: regional, nacional e internacional. En general, los porcentajes de ventas son el 30% de viajes regionales, el 50% de nacionales y el 20% de internacionales y las reclamaciones que recibe son del 1% en viajes regionales y nacionales y del 1,5% en viajes internacionales

a) De un total de 10 clientes, calcular la probabilidad de que al menos dos de ellos contraten un destino internacional y no emitan ninguna reclamación
b) Calcular la probabilidad de que la quinta reclamación que recibe la agencia se produzca cuando alcanza los 40 contratos
c) ¿Cuántos viajes contrata en media la agencia hasta que se produce la primera reclamación en un destino internacional?

- Datos del problema:
- p(viaje regional)=p(R)=0,3 p(viaje nacional)=p(N)=0,5 p(viaje internacional)=p(I)=0,2
- Definimos el suceso D≡ $se\ recibe\ reclamación$
- Entonces p(D/R)=0,01 p(D/N)=0,01 p(D/I)=0,015

a) De un total de 10 clientes, calcular la probabilidad de que al menos dos de ellos contraten un destino internacional y no emitan ninguna reclamación

Vemos lo que ocurre para un contrato;
¿cuál es la probabilidad de que sea destino internacional y no haya reclamación?
P(I∩ $\overline{D}$) =p($\overline{D}$/I).p(I)=0,985*0,2=0,197
De un total de 10 clientes, el número de ellos que contratan destino internacional y no reclaman: X ⟶Bin(10;0,017)
P(X≥ $2) = 1 - p(X < 2) = 1 - p(X = 0) - p(X = 1) = 1 - \binom{10}{0} 0{,}197^0 0{,}803^{10} - \binom{10}{1} 0{,}197^1 0{,}803^9$=0,615

b) **Calcular la probabilidad de que la quinta reclamación que recibe la agencia se produzca cuando alcanza los 40 contratos**

- Definimos Y≡ $número\ de\ contratos$ sin $reclamación\ hasta\ que\ se\ produce\ la\ quinta\ reclamación$
- Es una binomial negativa donde p=probabilidad de reclamación=p(D)
- P(D)=p(D/R)*p(R)+p(D/N)*p(N)+p(D/I)*p(I)=0,01*0,3+0,01*0,5+0,015*0,2=0,011

Entonces Y→ $\boldsymbol{BN(5;0,011)}$ y p(Y=35)=$\binom{35+5-1}{35}0{,}989^{35} * 0{,}011^{5} = 8{,}99 * 10^{-6} \approx 0$

c) **¿Cuántos viajes contrata en media la agencia hasta que se produce la primera reclamación en un destino internacional?**

- Definimos

N≡ $número\ de\ contratos$ sin $reclamación\ hasta\ que\ se\ produce\ la\ primera\ reclamación\ en\ un\ destino\ internacional$

- Es una distribución geométrica donde p=p(D/I)=0,015
- Y la EN=$\frac{0{,}985}{0{,}015}$=65,66≈ $66\ viajes$ sin $reclamación\ hasta\ la\ primera\ reclamación\ (en\ destinos\ internacionales)$
- Entonces el número medio de viajes totales contratados es 66+1 =67 viajes o contratos en total.

Problema 3

El número de productos (en cientos) que vende diariamente un centro comercial es una variable aleatoria X que, para cierta constante k, tiene función de densidad

$$f(x)=\begin{cases} k(x-1)(3-x) & si\ 1\leq x\leq 3 \\ 0 & en\ caso\ contrario \end{cases}$$

a) Determinar k para que f(x) sea realmente una función de densidad

b) Obtener la función de distribución de X

c) ¿Cuál es la probabilidad de que a lo largo de una semana completa haya al menos un día en que el centro vende más de 200 productos?

Definimos la variable X≡ ventas diarias en cientos de productos de un centro comercial

- Tenemos que determinar k para que f(x) sea realmente una función de densidad
- $\int_1^3 k(x-1)(3-x)dx = 1$ ⟷ $\frac{4}{3}$k=1 ⟷ k=3/4

b) La función de distribución de X es $F(x)=\begin{cases} 0 & si\ x<1 \\ \int_1^x \frac{3}{4}(t-1)(3-t)dt = \frac{-x^3}{4}+\frac{3x^2}{2}-\frac{9x}{4}+1 & si\ 1\leq x\leq 3 \\ 1 & si\ x>3 \end{cases}$

c) ¿Cuál es la probabilidad de que a lo largo de una semana completa haya al menos un día en que el centro vende más de 200 productos?

- Las ventas a lo largo de una semana completa se definen mediante
- Probabilidad de que un día seleccionado al azar venda más de 200 productos:

$$\int_{2}^{3} \frac{3}{4}(x-1)(3-x)dx = 0{,}5$$

- T$\equiv$ $\textit{número de días en que vende más de 2 cientos de productos} \quad \rightarrow \quad Bin(7;\, 0{,}5)$
- Tenemos que calcular p(T$\geq 1) = 1 - p(T = 0) = 1 - 0{,}5^7$=1-0,0078=0,9922