

Maestría en Computo Estadístico
Inferencia Estadística
Tarea 4

5 de octubre de 2020

Enrique Santibáñez Cortés

Repositorio de Git: Tarea 4, IE.

Escriba de manera concisa y clara sus resultados, justificando los pasos necesarios. Serán descontados puntos de los ejercicios mal escritos y que contenga ecuaciones sin una estructura gramatical adecuada. Las conclusiones deben escribirse en el contexto del problema. Todos los programas y simulaciones tienen que realizarse en R.

1. Sea X una v.a continua cuya función de distribución F es estrictamente creciente. El Teorema de la Transformación Integral nos dice que $Y = F(X)$ tiene distribución Uniforme(0, 1).

a) Sea $U \sim \text{Uniforme}(0, 1)$ y $X' = F^{-1}(U)$. Muestre que $X' \sim F$.

RESPUESTA

Como U tiene una distribución $\text{Unif}(0, 1)$, esto implica que $F_U(u) = u$ (por definición de la función acumulada de una distribución uniforme). Entonces tenemos que

$$F_{X'}(x') = \mathbb{P}(X' \leq x') = \mathbb{P}(F^{-1}(U) \leq x') = \mathbb{P}(U \leq F(x')) = F_U(F(x')) = F(x').$$

Por lo tanto, $X' \sim F$.

- b) Escriba un programa que tome variables aleatorias Uniforme(0, 1) y que las utilice para generar variables aleatorias de una distribución $\text{Exp}(\beta)$. El problema debe recibir el tamaño de la muestra que se desea generar, denotado por m , y al parámetro β . Deberá regresar una muestra de tamaño m .

RESPUESTA

Ocupando el teorema de la Transformación Integral, podemos generar números aleatorios de “cualquier” distribución deseada a partir de una distribución Uniforme(0,1), utilizando lo demostrado en el inciso a). Para ello calculemos la inversa de una distribución exponencial:

$$\begin{aligned} F(x) = 1 - e^{-x\beta} &\Rightarrow x = 1 - e^{-y\beta} \\ x - 1 &= -e^{-y\beta} \\ -y\beta &= \log(1 - x) \\ y &= -\frac{\log(1 - x)}{\beta} \end{aligned}$$

Entonces con lo anterior tenemos que la generación de los números aleatorios es de la siguiente forma:

```
random_exp <- function(m, beta){  
  return(-log(1-runif(m))/beta)  
}
```

- c) Simule $m = 100$ muestras $\text{Exp}(1/2)$. Con esta muestra contruya un QQ plot exponencial y una gráfica que compare el histograma de la muestra con la función de densidad de $\text{Exp}(1/2)$. Comente.

RESPUESTA

Usamos la función anterior para simular las 100 muestras de $\text{Exp}(1/2)$:

```
set.seed(19970808)  
simulacion_exp_12 <- random_exp(100, 1/2)  
head(simulacion_exp_12)
```

```
## [1] 0.2529386 0.6660432 0.4792715 0.8914632 1.7647800 0.5549034
```

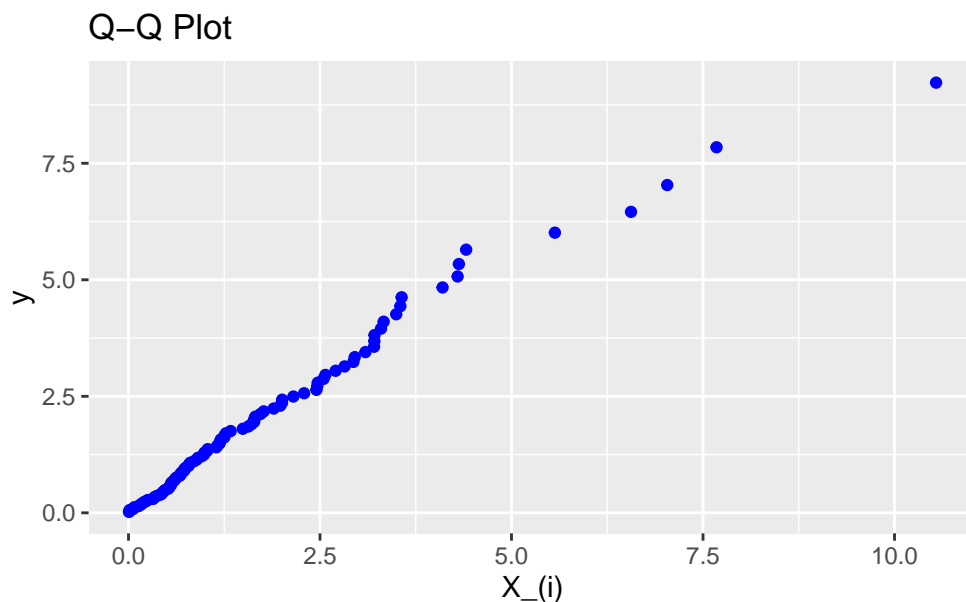
Contruimos un QQ plot con la muestra, para ello gráficamos

$$X_{(i)} \text{ vs } -2 \log \left(1 - \frac{i}{n+1} \right).$$

```
library(tidyverse)
datos<-data.frame(x_i=sort(simulacion_exp_12), i=seq(1:100)) %>%
  mutate(y=-2*log(1-i/(101)))
head(datos,3)
```

```
##           x_i i           y
## 1 0.006044715 1 0.01990066
## 2 0.012608229 2 0.04000133
## 3 0.014985241 3 0.06030608
```

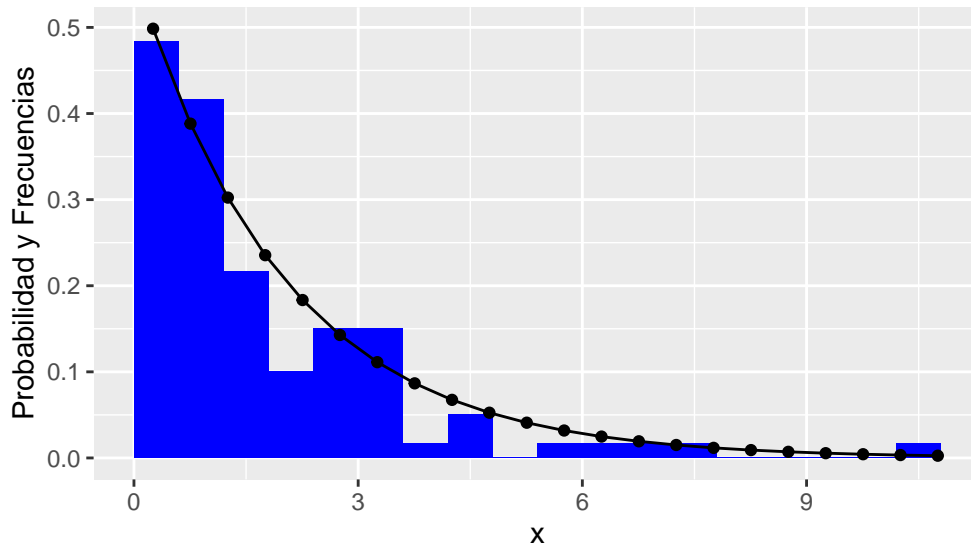
```
ggplot(data=datos, aes(x=x_i, y=y))+
  geom_point(colour="blue")+
  labs(x="X_(i)", title = "Q-Q Plot")
```



El QQ-plot nos indica una clara relación entre los cuantiles de la muestra y los cuantiles teóricos, por lo que observamos un claro ajuste satisfactorio. Posiblemente en las colas es dónde no se ajusta bien, pero eso es debido al tamaño de muestra pequeño que simulamos.

```
exp_teoric <- data.frame(y=dexp(x=seq(min(datos$x_i),max(datos$x_i),0.5),rate=1/2), x=seq(min(datos$x_i),max(datos$x_i),0.5)))
ggplot(data=datos, aes(x_i, y=..density..))+
  geom_histogram(fill="blue", breaks=seq(0,11,0.6))+
  geom_point(data=exp_teoric, aes(x,y))+
  geom_line(data=exp_teoric, aes(x,y))+
  labs(y="Probabilidad y Frecuencias", x="x", title="CDF y histograma exponencial.")
```

CDF y histograma exponencial.



Por lo tanto, con lo anterior validamos que efectivamente se generan números aleatorios de una exponencial (o cualquier otra) apartir de una Uniforme. Ya que podemos observar en el histograma y la CDF una gran similitud. ■.

2. Sea A el triángulo de vértices $(0, 0)$, $(0, 1)$, $(1, 0)$ y suponga que X, Y tiene una densidad conjunta uniforme en el triángulo. (a) Halle las distribuciones marginales de X, Y y $Z = X + Y$. (b) ¿Son X y Y independientes? ¿Por qué?

RESPUESTA

Por como esta descrita la densidad conjunta de X, Y podemos escribirla como

$$f(x, y) = \begin{cases} 2, & 0 \leq x \leq 1, x \leq y \leq 1 \\ 0, & \text{otro caso.} \end{cases}.$$

Validamos que efectivamente cumpla las condiciones de densidad conjunta, observemos que se cumple que $f(x, y) \geq 0, \forall (x, y)$. y además que integra 1:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = \int_0^1 \int_x^1 2 dy dx = \int_0^1 2(1 - x) dx = 2x - \frac{2x^2}{2} \Big|_0^1 = 1.$$

La definición de distribuciones marginales:

Definición: 1 Las distribuciones marginales de v.a continuas X y Y , denotadas por $f_X(x)$ y $f_Y(y)$, se definen como

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Entonces **a) las marginales son:**

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_x^1 2 dy = y|_x^1 = 2(1 - x).$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_y^1 2 dx = x|_y^1 = 2(1 - y),$$

es decir,

$$f_X(x) = \begin{cases} 2(1 - x) & 0 \leq x \leq 1 \\ 0 & \text{en otro caso.} \end{cases} \quad y \quad f_Y(y) = \begin{cases} 2(1 - y) & 0 \leq y \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

La marginal de $Z = X + Y$ se entrega en la próxima tarea.

Ocupando la siguiente definición (vista en clase):

Definición: 2 *Dos variables aleatorias continuas se dice que son **independientes** si su distribución conjunta se puede expresar como el producto de sus marginales.*

$$f(x, y) = f_X(x)f_Y(y).$$

Lo anterior es equivalente a decir que si la distribución conjunta no se puede expresar como el producto de sus marginales entonces no es independientes (equivalencia lógica). Entonces observemos que:

$$f(x, y) = 2 \neq 2(1 - x) \cdot 2(1 - y) = f_X(x)f_Y(y).$$

Por lo tanto, **b) podemos concluir que X y Y no son independientes por que el producto de sus marginales no es la conjunta.** ■.

3. Halle la densidad condicional de $X|Y = y$ si (X, Y) tiene densidad conjunta

$$f_{X,Y}(x, y) = \frac{1}{y} \exp\left(-\frac{x}{y} - y\right) \text{ para } x, y > 0.$$

También calcule $\mathbb{E}(X|Y = y)$.

RESPUESTA

Recordemos la definición de distribución condicional:

Definición: 3 *Las distribuciones condicionales de $X|Y = y$ y $Y|X = x$ denotadas por $f_{X|y}(x)$ y $f_{Y|x}(y)$ se definen de la siguiente forma. Para cada valor de y*

$$f_{X|y}(x) = \frac{f(x, y)}{f_Y(y)}.$$

Es una función de densidad en X . Sus valores pueden depender (en el sentido de función matemática, de y , pero como este valor es dado, no representa un factor aleatorio). Análogamente, para cada valor x

$$f_{Y|x}(y) = \frac{f(x, y)}{f_X(x)}.$$

Entonces calculemos la función de densidad marginal de Y en este problema es:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_0^{\infty} \frac{1}{y} \exp\left(-\frac{x}{y} - y\right) dx = \frac{1}{y} \int_0^{\infty} \exp\left(-\frac{x}{y} - y\right) dx$$

ocupando un cambio de variable $m = -\frac{x}{y} - y \Rightarrow dm = -\frac{1}{y} dx$, entonces

$$\begin{aligned} f_Y(y) &= \frac{1}{y} \int_0^{\infty} \exp\left(-\frac{x}{y} - y\right) dx = \frac{1}{y} \int_0^{\infty} \exp(m) (-y dm) = -\exp(m) = -\exp\left(-\frac{x}{y} - y\right) \Big|_{x=0}^{\infty} \\ &= \exp(-y). \end{aligned}$$

Es decir, la marginal de Y es:

$$f_Y(y) = \exp(-y), \quad y > 0.$$

Entonces la distribución conjunta de $X|Y = y$ es:

$$f_{X|Y}(x) = \frac{f(x, y)}{f_Y(y)} = \frac{\frac{1}{y} \exp\left(-\frac{x}{y} - y\right)}{\exp(-y)} = \frac{1}{y} \exp\left(\frac{x}{y} - y + y\right) = \frac{1}{y} \exp\left(\frac{x}{y}\right).$$

Por lo tanto:

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x) dx = \int_0^{\infty} \frac{x}{y} \exp\left(\frac{x}{y}\right) dx.$$

Si observamos, es la esperanza de una variable que se distribuye como una exponencial con parámetro $\frac{1}{y}$. Entonces, como sabemos que la esperanza de una variable poisson con parámetro λ es $\frac{1}{\lambda}$. Entonces **podemos concluir que esperanza condicional es:**

$$\mathbb{E}(X|Y = y) = y. \quad \blacksquare.$$

4. Sea $Y \sim \exp(\theta)$ y dado $Y = y$, X tiene distribución de Poisson de media y . Encuentre la ley de X .

RESPUESTA

Veamos que la densidad de Y es

$$f_Y(y) = \theta e^{-\theta y}, \quad y > 0.$$

Y podemos ver que la densidad condicional de X dado $Y = y$ es

$$f_{X|Y}(x) = \frac{y^x e^{-y}}{x!}, \quad x > 0.$$

Entonces usando la ley de probabilidad total tenemos que

$$f_X(x) = \int_0^{\infty} f_{X|Y}(x) f_Y(y) dy = \int_0^{\infty} \frac{y^x e^{-y}}{x!} \theta e^{-\theta y} dy = \theta \int_0^{\infty} \frac{y^x e^{-(1+\theta)y}}{x!} dy$$

haciendo un cambio de variable tenemos que $m = (1 + \theta)y \Rightarrow dm = (1 + \theta)$, entonces

$$\theta \int_0^{\infty} \frac{(m/(1 + \theta))^x e^{-m}}{x!} dm = \frac{\theta}{(1 + \theta)^{x+1}} \int_0^{\infty} \frac{m^x e^{-m}}{x!} dy = \frac{\theta}{(1 + \theta)^{x+1}}.$$

Por lo tanto, tenemos que

$$f_X(x) = \frac{\theta}{(1 + \theta)^{x+1}}, \quad x = 0, 1, \dots$$

Si hacemos un cambio de variable $p = \frac{1}{1+\theta}$, tenemos que lo anterior es igual a

$$f_X(x) = \frac{\theta}{(1 + \theta)^{x+1}} = (1 - p)p^x.$$

Entonces podemos observar que $X \sim Geo(p)$. \blacksquare .

5. Sea (X, Y) un vector aleatorio con la siguiente densidad

$$f(x, y) = \begin{cases} \pi^{-1} & x^2 + y^2 \leq 1 \\ 0 & x^2 + y^2 > 1 \end{cases}$$

Demuestre X y Y no están correlacionadas, pero que no son independientes.

RESPUESTA

Observemos que X y Y pueden tomar cualquier valor en el intervalo $[-1, 1]$, por lo que podemos decir que son variables continuas. Entonces ocupando la definición de densidades marginales 1, las densidades marginales son:

$$f_Y(y) = \int_{-\sqrt{y^2-1}}^{\sqrt{y^2-1}} \pi^{-1} dx = x\pi^{-1} \Big|_{-\sqrt{y^2-1}}^{\sqrt{y^2-1}} = 2\sqrt{y^2-1}\pi^{-1}, \quad -1 \leq y \leq 1.$$

$$f_X(x) = \int_{-\sqrt{x^2-1}}^{\sqrt{x^2-1}} \pi^{-1} dy = y\pi^{-1} \Big|_{-\sqrt{x^2-1}}^{\sqrt{x^2-1}} = 2\sqrt{x^2-1}\pi^{-1}, \quad -1 \leq x \leq 1,$$

es decir,

$$f_X(x) = \begin{cases} 2\sqrt{x^2-1}\pi^{-1} & -1 \leq x \leq 1. \\ 0 & \text{en otra caso.} \end{cases}, \text{ y } f_Y(y) = \begin{cases} 2\sqrt{y^2-1}\pi^{-1} & -1 \leq y \leq 1. \\ 0 & \text{en otra caso.} \end{cases}$$

Ocupando el teorema 2 podemos observar que

$$f(x, y) = \pi^{-1} \neq 2\sqrt{x^2-1}\pi^{-1}2\sqrt{y^2-1}\pi^{-1} = f_X(x)f_Y(y).$$

Por lo que podemos concluir que X y Y no son independientes.

Ahora calculemos las siguientes esperanzas $\mathbb{E}[X], \mathbb{E}[Y], \mathbb{E}[XY]$:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-1}^1 2x\sqrt{x^2-1}\pi^{-1} dx$$

ocupando un cambio de variable $m = x^2 - 1 \Rightarrow dm = 2x dx$, entonces

$$\int 2x\sqrt{x^2-1}\pi^{-1} dx = \pi^{-1} \int \sqrt{m} dm = \pi^{-1} \frac{2m^{3/2}}{3} = \frac{2(x^2-1)^{3/2}\pi^{-1}}{3}.$$

Por lo que

$$E[X] = \frac{2(x^2-1)^{3/2}\pi^{-1}}{3} \Big|_{-1}^1 = \frac{2(1^2-1)^{3/2}\pi^{-1}}{3} - \frac{2(1^2-1)^{3/2}\pi^{-1}}{3} = 0.$$

Ahora,

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{-1}^1 2y\sqrt{y^2-1}\pi^{-1} dy$$

ocupando un cambio de variable $m = y^2 - 1 \Rightarrow dm = 2y dy$, entonces

$$\int 2y\sqrt{y^2-1}\pi^{-1} dy = \pi^{-1} \int \sqrt{m} dm = \pi^{-1} \frac{2m^{3/2}}{3} = \frac{2(y^2-1)^{3/2}\pi^{-1}}{3}.$$

Por lo que

$$E[Y] = \frac{2(y^2-1)^{3/2}\pi^{-1}}{3} \Big|_{-1}^1 = \frac{2(1^2-1)^{3/2}\pi^{-1}}{3} - \frac{2(1^2-1)^{3/2}\pi^{-1}}{3} = 0.$$

Y por último:

$$\begin{aligned}\mathbb{E}[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x,y)dxdy = \int_{-1}^1 \int_{-\sqrt{x^2-1}}^{\sqrt{x^2-1}} xy\pi^{-1}dydx = \int_{-1}^1 x\pi^{-1} \left(\int_{-\sqrt{x^2-1}}^{\sqrt{x^2-1}} y \right) dydx \\ &= \int_{-1}^1 x\pi^{-1} \left(\frac{y^2}{2} \Big|_{-\sqrt{x^2-1}}^{\sqrt{x^2-1}} \right) dx = \int_{-1}^1 x\pi^{-1} \left(\frac{\sqrt{x^2-1}^2 - (-\sqrt{x^2-1})^2}{2} \right) dx \int_{-1}^1 x\pi^{-1} (0) dx = 0.\end{aligned}$$

Entonces ocupando los resultados anteriores tenemos que

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0 - 0 = 0.$$

Lo anterior implica que $\rho = 0$. **Y por lo tanto, podemos concluir que las X y Y no están correlacionadas.** ■.

Este es un claro ejemplo de que si $Cov(X, Y) = 0$, no necesariamente es cierto que las variables sean independientes, simplemente dice que no hay dependencia lineal, pero puede ser otro tipo de dependencia, que en el sentido contrario si es cierto.

6. Se entrega en la próxima tarea.

7. Cargue en R al conjunto de datos “Maíz.csv”, el cual contiene el precio mensual de la tonelada de maíz y el precio de la tonelada de tortillas en USD. En este ejercicio tendrá que estimar los coeficientes de una regresión lineal simple.

a) Calcule de forma explícita la estimación de los coeficientes via mínimos cuadrados y ajuste la regresión correspondiente. Concluya.

RESPUESTA

Para estimar los coeficientes de mínimos cuadrados tenemos que

$$Y = \beta_0 + \beta_1 X + \epsilon$$

entonces, los estimadores de los coeficientes son de la forma:

$$\hat{\beta}_1 = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sum x_i^2 - (\sum x_i)^2} \quad \text{y} \quad \hat{\beta}_0 = \frac{N \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{N \sum x_i^2 - (\sum x_i)^2} = \bar{y} - \hat{\beta}_1 * \bar{x}$$

Creemos una función para estimar los coeficientes por el método de mínimos cuadrados:

```
min_cuadrados_reg <- function(x,y){
  N<-length(x)
  sum_xy <- sum(x*y)
  sum_x <- sum(x)
  sum_y <- sum(y)
  sum_xx <-sum(x**2)

  hat_beta1 <- (N*sum_xy-sum_x*sum_y)/(N*sum_xx-(sum_x)**2)
  hat_beta0 <- mean(y)-hat_beta1*mean(x)
  list(beta0=hat_beta0, beta1=hat_beta1)
}
```

Cargamos los datos de maíz,

```
maiz <- read.csv("Maiz.csv", col.names = c("p_tonelada_maiz", "p_tonelada_tortilla"))
head(maiz, 3) # Mostamos los primeros registros.
```

```
##   p_tonelada_maiz p_tonelada_tortilla
## 1      138.9115      748.7452
## 2      144.9533      755.2285
## 3      122.7757      739.5067
```

Suponemos que podemos explicar el precio de la tonelada de tortilla apartir del precio de la tonelada de maíz, entonces ingresando nuestros datos a la función que escribimos tenemos que

```
min_cuadrados_reg(maiz$p_tonelada_maiz, maiz$p_tonelada_tortilla)
```

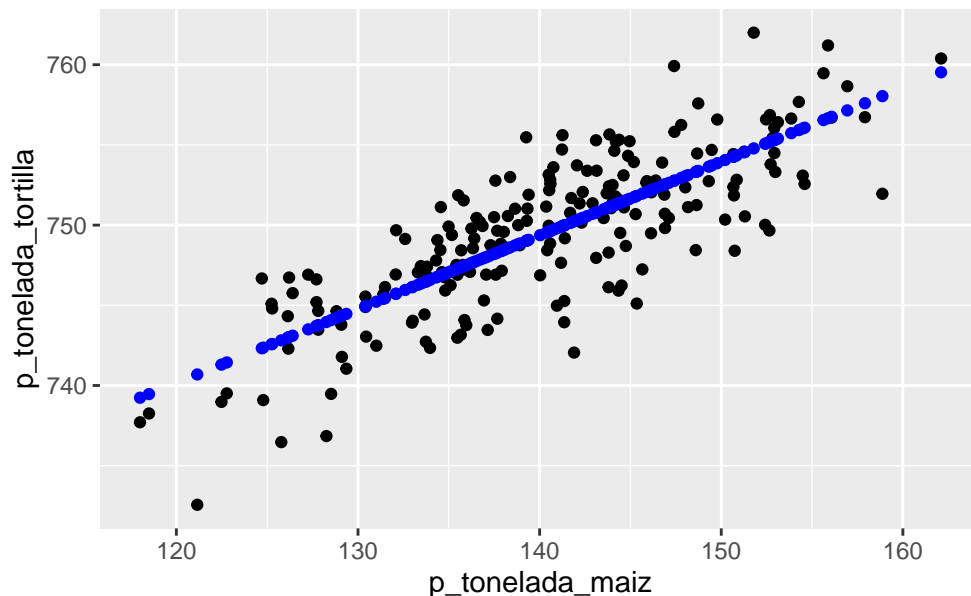
```
## $beta0
## [1] 684.9545
##
## $beta1
## [1] 0.4600343
```

Es decir,

$$\hat{Y}_{p_tone_tortilla} = 684,9545 + 0,4600343X_{p_tone_maiz}.$$

Graficamos los reales contra los estimados:

```
maiz <- maiz %>%
  mutate(y_hat_min = 684.9545+0.4600343*p_tonelada_maiz)
ggplot(data=maiz, aes(x=p_tonelada_maiz, y=p_tonelada_tortilla))+
  geom_point()+
  geom_point(aes(y=y_hat_min), color="blue")
```

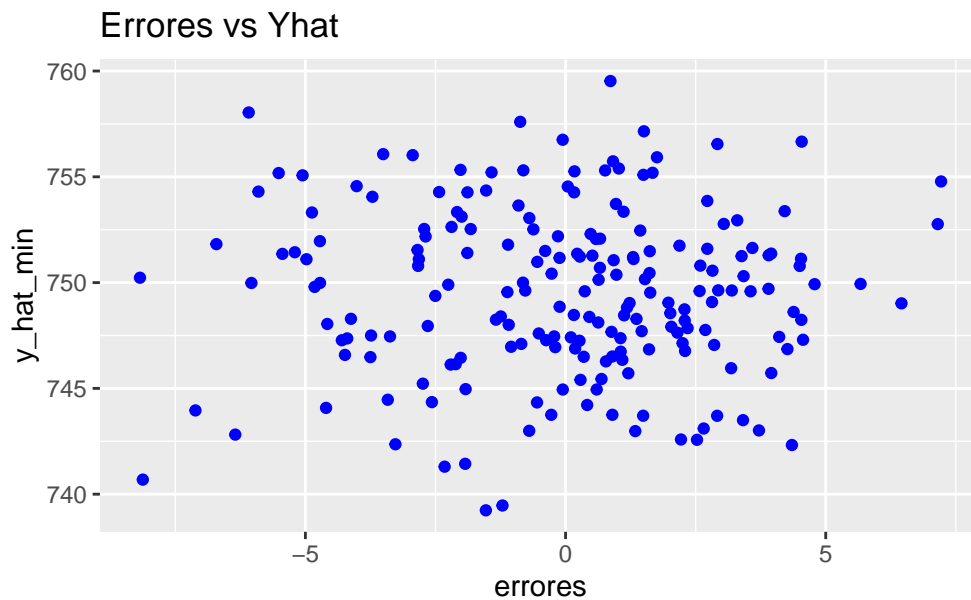


Para decir mas al respecto, realicemos un análisis de residuales con el objetivo de validar los supuestos que tenemos:

- los errores tiene varianza constante.
- se distribuyen como una normal.

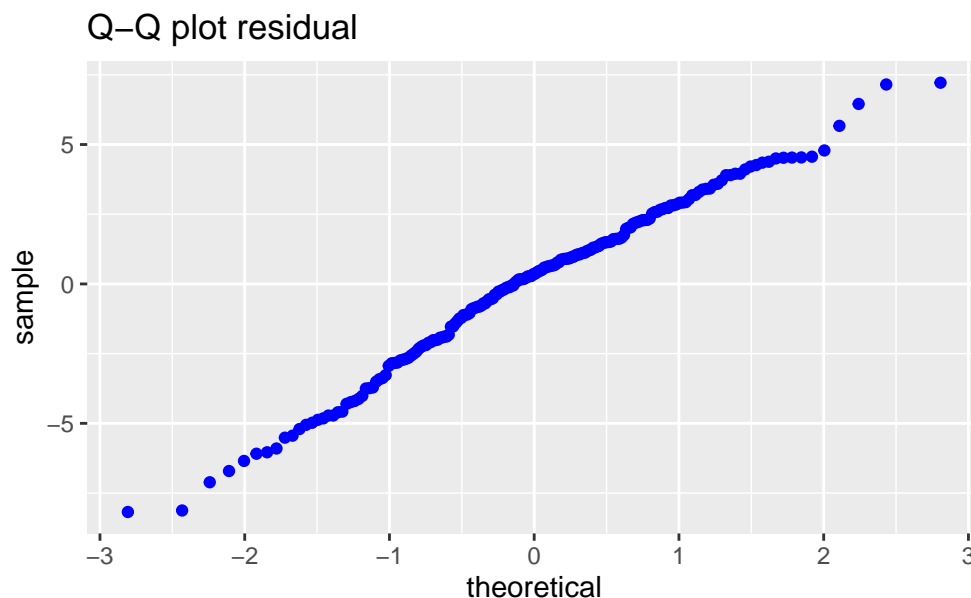
Para validar que tiene varianza constante grafiquemos los errores contra los predichos, si observamos un patrón de dispersión no aleatorio esto indicaría que la varianza de los residuos es mayor para ciertos valores predichos por el modelo, por lo que no cumpliría el supuesto de varianza constante.

```
maiz <- maiz %>%  
  mutate(errores = p_tonelada_tortilla - y_hat_min)  
  
ggplot(data=maiz, aes(x=errores, y=y_hat_min)) +  
  geom_point(color="blue") +  
  labs(title = "Errores vs Yhat")
```



Ahora validemos que los errores se distribuyen como una normal, para realizemos un gráfico QQ-plot.

```
ggplot(data=maiz, aes(sample=errores)) +  
  geom_qq(color="blue") +  
  labs(title = "Q-Q plot residual")
```



Por lo tanto, analizando los gráficos observamos que existe un patrón aleatorio de los errores vs predichos

por que indica que la varianza es constante y además si observamos el gráfico QQ-plot podemos concluir que si se distribuye como una normal. Por lo que concluimos que es un buen ajuste este modelo. \

Entonces una conclusión de los datos que puede ser útil es decir que cada que suba una unidad el precio de la tonelada de maiz subirá 0.46 unidades la tonelada de tortilla.

- b) Calcule de forma explícita la estimación de los coeficientes via regresión no-paramétrica tipo kernel (ver Nadaraya, E. A. (1964). “On Estimating Regression”. Theory of Probability and its Applications. 9 (1): 141–2. doi:10.1137/1109020) y ajuste la regresión correspondiente. Concluya.

RESPUESTA

Tenemos que aproximación de la curva de regresión de $Y : (\bar{y}(x))$ a partir de los datos, se tiene los estadísticos

$$\bar{y}(x) = \frac{\sum_{i=1}^N y_i K\left(\frac{x-x_1}{h(n)}\right)}{\sum_{i=1}^N K\left(\frac{x-x_1}{h(n)}\right)}$$

donde $K(X)$ es una función de densidad que satisface las condiciones:

- $K(x) < C < \infty$
- $\lim_{x \rightarrow \pm\infty} |xK(x)| = 0$.

Entonces podemos usar un kernel gaussiano, ya que cumple con las dos condiciones. Ahora procedemos a realizar la función y a graficar los resultados

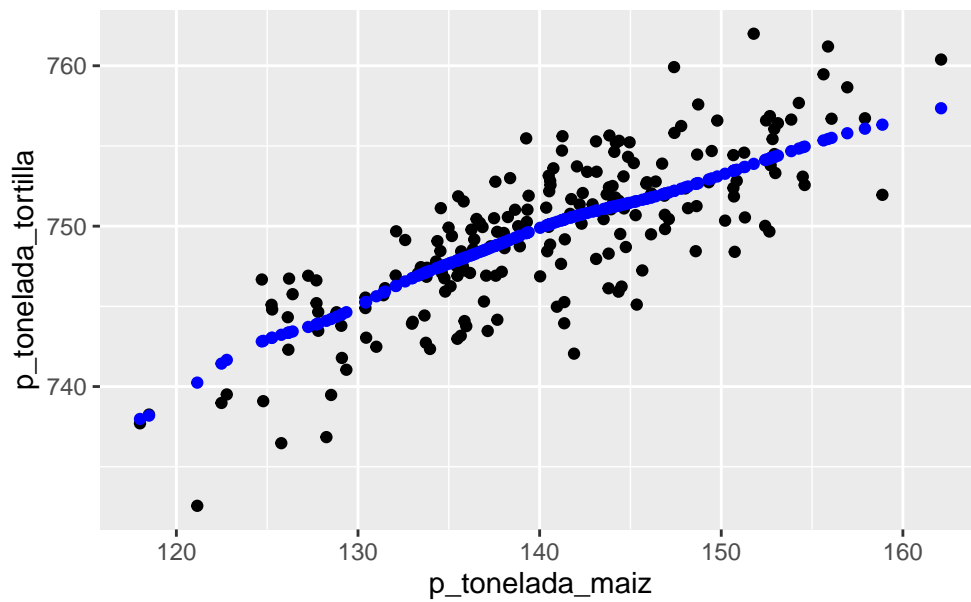
```
kernel_gaussiano<-function(x, x_i, h, n){ # funcion del kernel gaussiano
  dnorm((x-x_i)/(h*n))
}

y_stimate <- function(x, h, Dx, Dy){#funcion para estimar usando un kernel,
  # en este caso el gaussiano
  n <- length(Dx) # tamaño de los datos
  k <- kernel_gaussiano(x, Dx, h, n) #vector de kernels
  k2 <- Dy*k
  f <- sum(k2)/sum(k)
  f
}

maiz$y_hat <- sapply(maiz$p_tonelada_maiz, y_stimate, 0.015, maiz$p_tonelada_maiz, maiz$p_tonelad
head(maiz)

##   p_tonelada_maiz p_tonelada_tortilla y_hat_min   errores   y_hat
## 1      138.9115      748.7452  748.8586 -0.1133289 749.4015
## 2      144.9533      755.2285  751.6380  3.5905089 751.4512
## 3      122.7757      739.5067  741.4355 -1.9288257 741.6617
## 4      152.4164      750.0170  755.0713 -5.0542405 754.1267
## 5      133.0140      744.0338  746.1455 -2.1116647 746.7691
## 6      148.6396      751.2484  753.3338 -2.0853993 752.6506

ggplot(data=maiz, aes(x=p_tonelada_maiz, y=p_tonelada_tortilla))+
  geom_point()+
  geom_point(aes(x=p_tonelada_maiz, y=y_hat), color="blue")
```



Observamos un “buen” ajuste. Debido a que desconozco que supuestos (a excepción de la función $K(x)$) tiene este modelo no puedo validar con más detalle los resultados. Algo que considero un poco complicado es determinar h para este modelo, no encontré literatura en donde lo implementen (si tienen alguna estaría chido leerla).

c) Compare ambos resultados. ¿Qué diferencias observa?

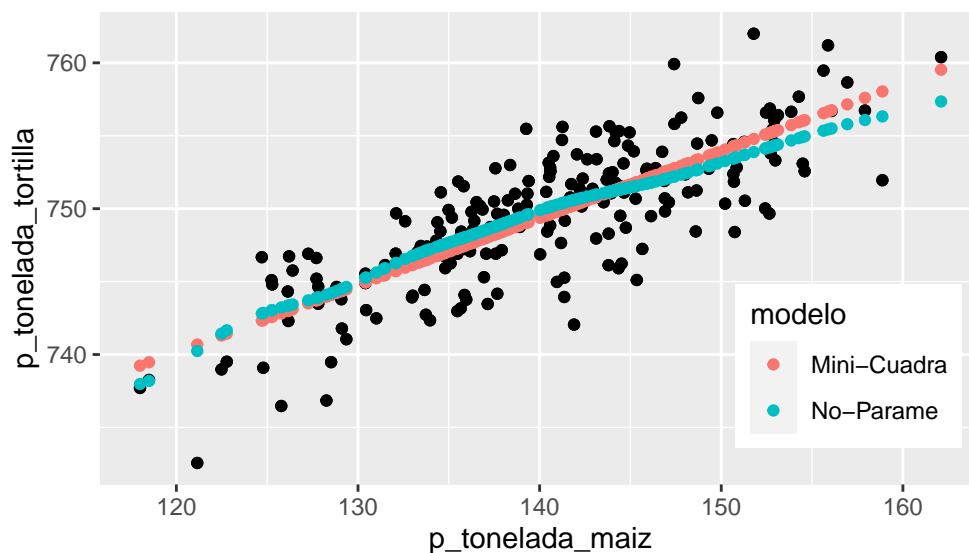
RESPUESTA

Veamos las dos estimaciones de ambos modelos.

```
maiz <-maiz %>%
  select(p_tonelada_maiz, p_tonelada_tortilla, y_hat_min, y_hat) %>%
  gather(key="modelo", value = "y", 3:4) %>%
  mutate(modelo =ifelse(modelo=="y_hat_min", "Mini-Cuadra", "No-Parame") )

ggplot(data=maiz, aes(x=p_tonelada_maiz, y=p_tonelada_tortilla))+
  geom_point()+
  geom_point(aes(x=p_tonelada_maiz, y=y, color=modelo))+
  labs(title="Comparación de los estimaciones.")+
  theme(legend.position = c(0.85, 0.25))
```

Comparación de los estimaciones.



Comparando las estimaciones ambas a mi parecer son muy buenas, observamos mayor diferencias en los valores extremos de los datos pero no es demasiada. Desde mi perspectiva elegiría el modelo de regresión por su simplicidad y además tengo más herramientas para validar su rendimiento, en cambio con el modelo no-paramétrico desde que función kernel o la amplitud (h) a escoger no supe como validar que los que yo elegí son buenos o no. Pero igual considero que eso puede ser una ventaja sobre la regresión, ya que como puedes utilizar una gran variedad de funciones kernel (obviamente que cumplan con los supuestos) y h 's por lo que existirá una gran diversidad de ajustes distintos. ■

8. Supongamos que se desea estudiar un número, n , grande de muestras de sangre para tratar de detectar una enfermedad rara en cada una de las muestras. Si cada muestra es analizada individualmente, se requieren n pruebas. Supongamos que cada muestra se divide a la mitad y que de cada muestra se selecciona una de las mitades, las que se usan para crear una mezcla de todas ellas (i.e. se juntan en una sola muestra). Si la prueba es suficientemente sensitiva, esta mezcla puede examinarse y, si tal prueba sale negativa, no se requieren hacer más pruebas y solo se necesita una (no hay enfermos en la muestra). Si la prueba de la mezcla es positiva, cada una de las mitades reservadas de las muestras (las mitades no mezcladas) pueden ser analizadas individualmente. En dicho caso, se necesitan un total de $n + 1$. Así, si la enfermedad es rara, es posible que se puedan lograr algunos ahorros a través de este método de analizar en grupo. Generalicemos el esquema anterior y supongamos que las n muestras primero se agrupan en m grupos de tamaño k (muestras); es decir, $n = mk$. Cada grupo se prueba con el método de grupo que se explico en el párrafo anterior; si un grupo sale positivo, cada individuo en el grupo es analizado individualmente. Si X_i es el número de pruebas que se hacen en el i -ésimo grupo, el número total de pruebas que se hacen es $N = \sum_{i=1}^m X_i$. Sea p la probabilidad de que un individuo cualquiera salga negativo en la prueba. Calcule la esperanza de N . ¿Por qué es importante que la enfermedad sea rara? Hint: Lea cuidadosamente el ejercicio. Note que la prueba de una muestra resulta en una v.a. *Bernoulli*(p).

RESPUESTA

Veamos como se distribuye X_i , solo puede tomar dos valores 1 o $k + 1$. La probabilidad de que X_i sea 1 es p^k , ya que es la probabilidad de que todos los k individuos en el grupo i salgan negativos y sabiendo que p es la probabilidad de que un individuo cualquiera salga negativo, y como la probabilidad de que un individuo salga negativo es independiente a que otro individuo salga negativo se concluye que es p^k . Ahora, La probabilidad de que X_i sea $k + 1$ es $1 - p^k$, ya que es la probabilidad de que salgan algún positivo en el grupo i pero es equivalente al complemento de que todos salgan negativos. Por lo tanto, tenemos que la

función de densidad de X_i es:

$$f(x_i) = \begin{cases} p^k & x_i = 1 \\ 1 - p^k & x_i = k + 1. \end{cases}$$

Ahora, por como esta definidos los X_i suponemos que son independientes. Entonces tenemos que

$$\begin{aligned} \mathbb{E}[N] &= \mathbb{E}\left[\sum_{i=1}^m X_i\right] = \sum_{i=1}^m \mathbb{E}[X_i] = \sum_{i=1}^m 1 \cdot p^k + (1 - p^k) * (k + 1) = \sum_{i=1}^m p^k + k + 1 - kp^k - p^k \\ &= \sum_{i=1}^m (k + 1 - kp^k) = m(k + 1 - kp^k). \end{aligned}$$

Si la enfermedad es rara eso implica que p^k sea pequeña, por lo que en teoría $p^k > 1 - p^k$, es decir, sería más probable hacer una prueba que $k + 1$ pruebas. Entonces en términos económicos y de tiempos, la metodología expuesta es muy eficaz y más óptima que analizar siempre todas las muestras por separados. ■.

Honours problems

1. Sea X una v.a continua con segundo momento finito. Demuestra que la mediana $M(X)$ de X satisface

$$|M(X) - E(X)| \leq (Var(X))^{1/2}.$$

Hint: Use los honour problems de la tarea anterior.

RESPUESTA

Recordemos el siguiente teorema (visto en clase).

Teorema: 1 (*Desigualdad de Jensen*) Si X es una v.a con primer momento finito y g es convexa, entonces

$$E(g(X)) \geq g(E(X)).$$

Si definimos a $g(x) = (|x - c|)^2 \Leftrightarrow g(x) = (x - c)^2$, observemos que:

$$g'(x) = 2x - 2c \text{ y } g''(x) = 2.$$

Entonces como $g''(x) > 0$ podemos decir que $g(x)$ es convexa. Ocupando el teorema 1 tenemos que

$$\mathbb{E}(|X - c|^2) \geq |c - \mathbb{E}(X)|^2. \quad (1)$$

Observemos que cuando $c = \mathbb{E}(X)$

$$E(|x - \mathbb{E}(X)|^2) = Var(X).$$

Ocupando lo demostrado en la tarea 3, podemos decir que cuando $c = M(X)$ se minimiza $\mathbb{E}(|X - c|^2)$, por lo que se cumple que

$$Var(X) = \mathbb{E}(|x - \mathbb{E}(X)|^2) \geq \mathbb{E}(|X - M(X)|^2)$$

Y por lo tanto, ocupando 1:

$$Var(X) = \mathbb{E}(|x - \mathbb{E}|^2) \geq \mathbb{E}(|X - M(X)|^2) \geq (|M(X) - \mathbb{E}(X)|^2) \Rightarrow$$

$$(Var(X))^{1/2} \geq |M(X) - \mathbb{E}(X)| \quad \blacksquare.$$

Ejercicios de las notas:

1. Se quiere verificar la programación del tiempo que debe durar la luz verde en un semáforo que se encuentra en una cierta intersección, con vuelta a la izquierda. Como no se tiene información al respecto, se envía a un estudiante graduado a hacer observaciones sobre el número de carros X y el número de camiones Y que llegan en un ciclo (entre verde y verde); y con esto se construye la tabla de distribuciones conjunta. Se plantean una serie de preguntas y con ello confirmas qué tan eficiente fue el ciclo planeado. Los resultados fueron los siguientes

	x			
	$p(x, y)$	0	1	2
y	0	.025	.015	.010
	1	.050	.030	.020
	2	.125	.075	.050
	3	.150	.090	.060
	4	.100	.060	.040
	5	.050	.030	.020

- 1.1. Verifica que es una tabla válida de probabilidades conjuntas.

RESPUESTA

Observemos que todas las probabilidades son menores a 1 y mayores a 0, por lo que si cumple que $p(x, y) > 0$. Ahora validemos que la suma de todas las probabilidades conjuntas sumas 1:

$$\begin{aligned} \sum_{i=0}^5 \sum_{j=0}^2 p(X = j, Y = i) &= 0,025 + 0,050 + 0,125 + 0,150 + 0,100 + 0,050 + 0,015 + 0,030 + 0,075 + \\ &\quad 0,090 + 0,060 + 0,030 + 0,010 + 0,020 + 0,050 + 0,060 + 0,040 + 0,020 \\ &= 1 \end{aligned}$$

Por lo tanto si cumple los dos supuestos para ser una probabilidad conjunta.

- 1.2. Calcula las distribuciones marginales para carros y camiones.

RESPUESTA

Por definición 1 tenemos que:

$$p(x) = \sum_{i=0}^5 p(x, Y = i) \Rightarrow \begin{cases} 0,025 + 0,050 + 0,125 + 0,150 + 0,100 + 0,050 = \mathbf{0.5} & \text{si } x = 0 \\ 0,015 + 0,030 + 0,075 + 0,090 + 0,060 + 0,030 = \mathbf{0.3} & \text{si } x = 1 \\ 0,010 + 0,020 + 0,050 + 0,060 + 0,040 + 0,020 = \mathbf{0.2} & \text{si } x = 2 \end{cases} \quad y$$

$$p(y) = \sum_{j=0}^2 p(X = j, y) \Rightarrow \begin{cases} 0,025 + 0,015 + 0,010 = \mathbf{0.05} & \text{si } y = 0 \\ 0,050 + 0,030 + 0,020 = \mathbf{0.10} & \text{si } y = 1 \\ 0,125 + 0,075 + 0,050 = \mathbf{0.25} & \text{si } y = 2 \\ 0,150 + 0,090 + 0,060 = \mathbf{0.30} & \text{si } y = 3 \\ 0,100 + 0,060 + 0,040 = \mathbf{0.20} & \text{si } y = 4 \\ 0,050 + 0,030 + 0,020 = \mathbf{0.10} & \text{si } y = 5 \end{cases}$$

1.3. ¿Cuál es la probabilidad de que se tengan exactamente un carro y un camión en un ciclo dado?

RESPUESTA

Utilizando la probabilidad conjunta sería:

$$p(X = 1, Y = 1) = 0,030.$$

Observemos que para este caso en específico se cumple que

$$p(X = 1, Y = 1) = 0,030 = (0,10) \cdot (0,30) = p(X = 1)p(Y = 1).$$

1.4. Supongamos que la vuelta a la izquierda tiene una capacidad máxima de 5 carros y un camión es equivalente a 3 carros. ¿Cuál es la probabilidad de que se sature la línea en un ciclo dado?

RESPUESTA

Como la capacidad máxima es de 5 carros, la línea se satura cuando sobrepasan esta capacidad, entonces encontremos los valores para los cuales se cumple que:

$$\begin{aligned} p(X + 3Y > 5) &\Leftrightarrow 1 - p(X + 3Y \leq 5) \\ &= 1 - \sum_{i=0}^1 \sum_{j=0}^2 p(X = j, Y = i) \\ &= 1 - (0,05 + 0,10) \\ &= 1 - (0,15) = 0,85. \end{aligned}$$

Entonces, si observamos observamos una probabilidad muy alta de que se saturé la línea en un ciclo dado. Por lo que una buena sugerencia es reducir el tiempo del semáforo.

1.5. ¿Cuál es la probabilidad de que se tenga exactamente un carro, dado que ya se tiene un camión en un ciclo predeterminado y se sabe que no habrá ningún otro camión más? Contesta esta misma pregunta para $Y = 0, 1, 2, 3, 4, 5$. \end{enumerate}

RESPUESTA

Ocupando la definición condicional 3 es sencillo calcular las probabilidades solicitadas:

$$p(x = 1|y = 0) = \frac{p(x = 1, y = 0)}{p(y = 0)} = \frac{0,015}{0,05} = 0,30 = p(x = 1)$$

$$p(x = 1|y = 1) = \frac{p(x = 1, y = 1)}{p(y = 1)} = \frac{0,030}{0,10} = 0,30 = p(x = 1)$$

$$p(x = 1|y = 2) = \frac{p(x = 1, y = 2)}{p(y = 2)} = \frac{0,075}{0,25} = 0,30 = p(x = 1)$$

$$p(x = 1|y = 3) = \frac{p(x = 1, y = 3)}{p(y = 3)} = \frac{0,090}{0,30} = 0,30 = p(x = 1)$$

$$p(x = 1|y = 4) = \frac{p(x = 1, y = 4)}{p(y = 4)} = \frac{0,060}{0,20} = 0,30 = p(x = 1)$$

$$p(x = 1|y = 5) = \frac{p(x = 1, y = 5)}{p(y = 5)} = \frac{0,030}{0,10} = 0,30 = p(x = 1)$$

Observamos que no para $X = 1$ la probabilidad

$$p(X = 1|Y = i) = p(X = 1).$$

Es decir, tal vez exista ‘independencia’ entre X y Y . Falta validarlo para los demás valores de X si se cumple lo mismo o también ocupar 2 para comprobarlo. ■.

2. Calcula la regresión de Y_1 sobre $Y_2 = y_2 = 1$ en el ejemplo. Tendría ésta un significado valorable? explica tanto como puedas.

RESPUESTA

Recordemos que Y_1 representa el tiempo que tarda un cliente desde que entra al local hasta que deja la ventanilla de servicio, y Y_2 el tiempo que dura en la fila hasta que llega a la ventanilla de servicio, y además que las funciones de probabilidades conjuntas y marginales son:

$$f_{Y_1, Y_2}(y_1, y_2) = \begin{cases} e^{-y_1} & 0 \leq y_2 \leq y_1 \leq \infty \\ 0 & \text{en otra parte} \end{cases}, \quad f_{Y_1}(y_1) = \begin{cases} y_1 e^{-y_1} & 0 \leq y_1 \leq \infty \\ 0 & \text{en otra parte} \end{cases} \quad y$$

$$f_{Y_2}(y_2) = \begin{cases} e^{-y_2} & 0 \leq y_2 \leq \infty \\ 0 & \text{en otra parte} \end{cases}.$$

Entonces ocupando lo anterior podemos calcular la función de probabilidad condicional $Y_1|Y_2 = y_2$:

$$f_{Y_1|y_2}(y_1) = \frac{f_{Y_1, Y_2}(y_1, y_2)}{f_{Y_2}(y_2)} = \frac{e^{-y_1}}{e^{-y_2}} = e^{-y_1+y_2} \quad 0 \leq y_2 \leq y_1 \leq \infty$$

Por lo que, la regresión de Y_1 sobre $Y_2 = y_2 = 1$ es

$$\mathbb{E}[Y_1|Y_2 = y_2] = \int_{-\infty}^{\infty} f_{Y_1|y_2}(y_1) y_1 dy_1 = \int_{y_2}^{\infty} e^{-y_1+y_2} y_1 dy_1 = e^{y_2} \int_{y_2}^{\infty} e^{-y_1} y_1 dy_1,$$

integrando por partes tenemos que $u = y_1 \Rightarrow du = dy_1$, $v = \int e^{-y_1} \Rightarrow -e^{-y_1}$ entonces

$$\int e^{-y_1} y_1 dy_1 = -y_1 e^{-y_1} + \int e^{-y_1} = -y_1 e^{-y_1} - e^{-y_1}$$

Por lo que podemos concluir que

$$\begin{aligned} \mathbb{E}[Y_1|Y_2 = y_2] &= e^{y_2} \left[\frac{-y_1}{e^{y_1}} - \frac{1}{e^{y_1}} \right]_{y_1=y_2}^{\infty} = e^{y_2} \left[\lim_{y_1 \rightarrow \infty} \left(\frac{-y_1}{e^{y_1}} - \frac{1}{e^{y_1}} \right) + \frac{y_2}{e^{y_2}} + \frac{1}{e^{y_2}} \right] \\ &= e^{y_2} \left[\frac{y_2}{e^{y_2}} + \frac{1}{e^{y_2}} \right] = y_2 + 1 \end{aligned} \quad (2)$$

Por lo que,

$$\mathbb{E}[Y_1|Y_2 = 1] = 2.$$

Si analizamos la regresión teórica encontrada 2 observamos una función lineal en función de y_2 para explicar y_1 , la interpretación de este hecho es que el tiempo total que tarda un cliente desde que entra al local hasta que deja la ventanilla de servicio se puede estimar con el tiempo que tardar en la fila hasta que llega a la ventanilla de servicio más un minuto. Recordando que el tiempo total es igual al tiempo que tarda en la fila más el tiempo que tarda en la ventanilla de servicio, por como esta definida la regresión 2 podemos inferir que el tiempo de la ventanilla de servicio es constante (igual a 1) y esto nos dice demasiado, nos dice que por cada cliente te tardarás un minuto en atenderlo en la ventanilla de servicio. ■.

3. En un área determinada, cierto material es seleccionado al azar y pesado en dos tiempos distintos. Sea W = peso del material (verdadero) y X_1 , X_2 las dos mediciones hechas. Uno puede pensar estas mediciones como

$$X_1 = W + e_1$$

$$X_2 = W + e_2$$

, donde e_1 y e_2 son los errores de medición. Supongamos que e_1 y e_2 son independientes entre sí e independientes de W y que $V(e_1) = V(e_2) = \sigma_e^2$. Expresar a ρ , el coeficiente de correlación entre X_1 y X_2 , en términos de σ_W^2 (la varianza del peso verdadero) y σ_e^2 ; interpreta el resultado.

RESPUESTA

Recordemos que el coeficiente de correlación entre dos variables X y Y esta definido como

$$\rho = \frac{Cov(X, Y)}{\sigma_x \sigma_y},$$

donde

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Entonces procedemos a calcular las siguientes esperanzas para determinar el coeficiente de correlación entre X_1 y X_2 :

$$\begin{aligned} \mathbb{E}[X_1 X_2] &= \mathbb{E}[(W + \epsilon_1)(W + \epsilon_2)] = \mathbb{E}[W^2 + W\epsilon_1 + W\epsilon_2 + \epsilon_1\epsilon_2] \\ &= \mathbb{E}[W^2] + \mathbb{E}[W\epsilon_1] + \mathbb{E}[W\epsilon_2] + \mathbb{E}[\epsilon_1\epsilon_2] \\ &= \mathbb{E}[W^2] + \mathbb{E}[W]\mathbb{E}[\epsilon_1] + \mathbb{E}[W]\mathbb{E}[\epsilon_2] + \mathbb{E}[\epsilon_1]\mathbb{E}[\epsilon_2] \\ &= \sigma_W^2 + \mu_W^2 + \mu_W\mu_{\epsilon_1} + \mu_W\mu_{\epsilon_2} + \mu_{\epsilon_1}\mu_{\epsilon_2}. \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X_1]\mathbb{E}[X_2] &= \mathbb{E}[W + \epsilon_1]\mathbb{E}[W + \epsilon_2] = (\mathbb{E}[W] + \mathbb{E}[\epsilon_1])(\mathbb{E}[W] + \mathbb{E}[\epsilon_2]) \\ &= (\mu_W + \mu_{\epsilon_1})(\mu_W + \mu_{\epsilon_2}) = \mu_W^2 + \mu_W\mu_{\epsilon_1} + \mu_W\mu_{\epsilon_2} + \mu_{\epsilon_1}\mu_{\epsilon_2}. \end{aligned}$$

Entonces tenemos que

$$\begin{aligned} Cov(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \sigma_W^2 + \mu_W^2 + \mu_W\mu_{\epsilon_1} + \mu_W\mu_{\epsilon_2} + \mu_{\epsilon_1}\mu_{\epsilon_2} - (\mu_W^2 + \mu_W\mu_{\epsilon_1} + \mu_W\mu_{\epsilon_2} + \mu_{\epsilon_1}\mu_{\epsilon_2}) \\ &= \sigma_W^2. \end{aligned}$$

Ahora calculemos la varianza de X_1 y X_2 :

$$\begin{aligned} Var(X_1) &= Var(W + \epsilon_1) = Var(W) + Var(\epsilon_1) = \sigma_W^2 + \sigma_{\epsilon_1}^2 \\ Var(X_2) &= Var(W + \epsilon_2) = Var(W) + Var(\epsilon_2) = \sigma_W^2 + \sigma_{\epsilon_2}^2. \end{aligned}$$

Y esto implica que el coeficiente de correlación se pueda expresar como:

$$\rho = \frac{Cov(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} = \frac{\sigma_W^2}{\sqrt{\sigma_W^2 + \sigma_{\epsilon_1}^2}\sqrt{\sigma_W^2 + \sigma_{\epsilon_2}^2}}.$$

La relación lineal entre estas dos variables depende de la varianza del peso verdadero y de la varianza de los errores de medición, es decir, la relación de las dos mediciones en dos tiempos distintos se puede explicar por la varianza que tengan los errores y la varianza del peso del material. ■.

4. Para $t = (t_1, t_2)$, la función generatriz de momentos conjunta del vector $W = (X, Y)$ está dada por

$$\begin{aligned} M_{X,Y}(t_1, t_2) &= \mathbb{E}(e^{t'W}) = \mathbb{E}(e^{t_1x+t_2y}) \\ &= \exp \left\{ t_1\mu_x + t_2\mu_y + \frac{1}{2}(\sigma_x^2 t_1^2 + 2\rho\sigma_x\sigma_y t_1 t_2 + \sigma_y^2 t_2^2) \right\} \end{aligned}$$

Demostrar la igualdad anterior.

RESPUESTA

Tenemos que

$$\begin{aligned} M_{X,Y}(t_1, t_2) &= \mathbb{E}(e^{t'W}) = \mathbb{E}(e^{t_1x+t_2y}) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1x+t_2y} f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1x+t_2y} f_{Y|x}(y) f(x) dx dy \quad \text{definición de dist. conjunta} \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f_{Y|x}(y) e^{t_2y} dy \right\} e^{t_1x} f_X(x) dx \quad \text{paso algebraico} \end{aligned}$$

Ahora como sabemos

$$(Y|X = x) \sim N\left(\mu_{Y|X} = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x), \sigma_{Y|X}^2 = \sigma_y^2(1 - \rho^2)\right)$$

Entonces la integral es la generadora de momentos de normal normal y como sabemos que la generadora de momentos de una $Z \sim N(\mu_z, \sigma_z^2)$ es igual a

$$M_Z(t) = e^{\mu_z t + \frac{1}{2}\sigma_z^2 t^2}.$$

Sabiendo eso, tenemos entonces que

$$\begin{aligned} M_{X,Y}(t_1, t_2) &= \int_{-\infty}^{\infty} e^{t_1x} f_X(x) \left\{ \exp \left[\frac{1}{2}\sigma_2^2 t_2^2 (1 - \rho^2) + t_2 \left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1) \right) \right] \right\} \\ &= \exp \left[\frac{1}{2}\sigma_2^2 t_2^2 (1 - \rho^2) + t_2 \mu_2 - t_2 \rho \frac{\sigma_2}{\sigma_1} \mu_1 \right] \int_{-\infty}^{\infty} e^{((\rho\sigma_2/\sigma_1)t_2 + t_1)x} f_X(x) dx \end{aligned}$$

Observemos que la última integral se puede resolver facilmente utilizando la definición de generadora de momentos para una variable normal (como la última integral):

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{((\rho\sigma_2/\sigma_1)t_2+t_1)x} f_X(x)dx &= \int_{-\infty}^{\infty} e^{tx} f_X(x)dx, \text{ donde } t = (\rho\sigma_2/\sigma_1)t_2 + t_1. \\
&= \mathbb{E}[e^{tx}] = \exp\left[\mu_1 t + \frac{1}{2}\sigma_1^2 t^2\right] \\
&= \exp\left[\mu_1((\rho\sigma_2/\sigma_1)t_2 + t_1) + \frac{1}{2}\sigma_1^2((\rho\sigma_2/\sigma_1)t_2 + t_1)^2\right] \\
&= \exp\left[\mu_1((\rho\sigma_2/\sigma_1)t_2 + t_1) + \frac{1}{2}\sigma_1^2((\rho\sigma_2/\sigma_1)^2 t_2^2 + 2(\rho\sigma_2/\sigma_1)t_2 t_1 + t_1^2)\right] \\
&= \exp\left[\mu_1((\rho\sigma_2/\sigma_1)t_2 + t_1) + \frac{\sigma_1^2(\rho\sigma_2/\sigma_1)^2 t_2^2}{2} + \sigma_1 \rho \sigma_2 t_2 t_1 + \frac{\sigma_1^2 t_1^2}{2}\right] \\
&= \exp\left[\frac{\mu_1 \rho \sigma_2 t_2}{\sigma_1} + \mu_1 t_1 + \frac{\rho^2 \sigma_2^2 t_2^2}{2} + \sigma_1 \rho \sigma_2 t_2 t_1 + \frac{\sigma_1^2 t_1^2}{2}\right]
\end{aligned}$$

Sustituyendo lo anterior podemos concluir que:

$$\begin{aligned}
M_{X,Y}(t_1, t_2) &= \exp\left[\frac{1}{2}\sigma_2^2 t_2^2(1 - \rho^2) + t_2 \mu_2 - t_2 \rho \frac{\sigma_2}{\sigma_1} \mu_1\right] \cdot \exp\left[\frac{\mu_1 \rho \sigma_2 t_2}{\sigma_1} + \mu_1 t_1 + \frac{\rho^2 \sigma_2^2 t_2^2}{2} + \sigma_1 \rho \sigma_2 t_2 t_1 + \frac{\sigma_1^2 t_1^2}{2}\right] \\
&= \exp\left[\frac{1}{2}\sigma_2^2 t_2^2(1 - \rho^2) + t_2 \mu_2 - t_2 \rho \frac{\sigma_2}{\sigma_1} \mu_1 + \frac{\mu_1 \rho \sigma_2 t_2}{\sigma_1} + \mu_1 t_1 + \frac{\rho^2 \sigma_2^2 t_2^2}{2} + \sigma_1 \rho \sigma_2 t_2 t_1 + \frac{\sigma_1^2 t_1^2}{2}\right] \\
&= \exp\left[\mu_1 t_1 + \mu_2 t_2 + \frac{\sigma_1^2 t_1^2 + \sigma_2^2 t_2^2 + 2\rho \sigma_1 t_1 t_2}{2}\right]. \quad \blacksquare.
\end{aligned}$$

5. Sean X y Y v.a. Mostrar que $\mathbb{E}(Y|X)$ es el mejor predictor de Y entre todas las funciones de X .

RESPUESTA

Para mostrar que es el mejor predictor de Y probaremos que $\mathbb{E}(Y|X) = f(X)$ cumple

$$\min_{f(X)} \mathbb{E}[(Y - f(X))^2].$$

Para mostrarlo veamos que:

$$\begin{aligned}
\mathbb{E}[(Y - f(X))^2] &= \mathbb{E}[(Y - f(X) + \mathbb{E}(Y|X) - \mathbb{E}(Y|X))^2] \\
&= \mathbb{E}[((Y - \mathbb{E}(Y|X)) + (\mathbb{E}(Y|X) - f(X)))^2] \\
&= \mathbb{E}[(Y - \mathbb{E}(Y|X))^2 + (\mathbb{E}(Y|X) - f(X))^2 + 2(Y - \mathbb{E}(Y|X))(\mathbb{E}(Y|X) - f(X))] \\
&= \mathbb{E}[(Y - \mathbb{E}(Y|X))^2] + \mathbb{E}[(\mathbb{E}(Y|X) - f(X))^2] + 2\mathbb{E}[(Y - \mathbb{E}(Y|X))(\mathbb{E}(Y|X) - f(X))],
\end{aligned}$$

ahora ocupando que:

$$\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}[Y].$$

Esto implica que

$$\begin{aligned}
2\mathbb{E}[(Y - \mathbb{E}(Y|X))(\mathbb{E}(Y|X) - f(X))] &= 2(\mathbb{E}[(Y - \mathbb{E}(Y|X))])\mathbb{E}[(\mathbb{E}(Y|X) - f(X))] \\
&= 2(\mathbb{E}[Y] - \mathbb{E}[\mathbb{E}(Y|X)])\mathbb{E}[(\mathbb{E}(Y|X) - f(X))] \\
&= 2(\mathbb{E}[Y] - \mathbb{E}[Y])\mathbb{E}[(\mathbb{E}(Y|X) - f(X))] = 0.
\end{aligned}$$

Por lo que implica que

$$\mathbb{E}[(Y - f(X))^2] = \mathbb{E}[(Y - \mathbb{E}(Y|X))^2] + \mathbb{E}[(\mathbb{E}(Y|X) - f(X))^2],$$

ahora por como esta definido (la esperanza de algo elevado al cuadrado) $\mathbb{E}[(\mathbb{E}(Y|X) - f(X))^2]$ podemos decir que es mayor o igual a cero para cualquier $f(X)$. Por lo que,

$$\mathbb{E}[(Y - f(X))^2] = \mathbb{E}[(Y - \mathbb{E}(Y|X))^2] + \mathbb{E}[(\mathbb{E}(Y|X) - f(X))^2] \geq \mathbb{E}[(Y - \mathbb{E}(Y|X))^2].$$

Es decir, tiene un mínimo en $\mathbb{E}[(Y - \mathbb{E}(Y|X))^2]$, el cual se alcanza cuando $f(X) = \mathbb{E}(Y|X)$, o visto de otra forma $f(X) = \mathbb{E}(Y|X)$ minimiza

$$\min_{f(X)} \mathbb{E}[(Y - f(X))^2].$$

Y por lo tanto, podemos concluir que $\mathbb{E}(Y|X)$ es el mejor predictor de Y entre todas las funciones de X .
■.