Maestría en Computo Estadístico
Estadística Multivariada
Tarea 1
2 de marzo de 2021
Enrique Santibáñez Cortés
Repositorio de Git: Tarea 2, EM.

Ejercicio 1. Demuestre la siguiente igualdad:

$$\sum_{i=1}^{n}(\mathbf{x}_i-\mu)(\mathbf{x}_i-\mu)'=\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})'+n(\bar{\mathbf{x}}-\mu)(\bar{\mathbf{x}}-\mu)',$$

donde $\mathbf{x}_i \sum N_p(\mu,\Sigma), \quad i=1,\cdots,n.$

RESPUESTA
Desarrollemos la suma del lado izquierdo pero agreguemos dos cero, sumando y restando $\bar{x}$ de la siguiente forma:

$$\begin{aligned}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)(\mathbf{x}_i-\mu)' &= \sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\mu)(\mathbf{x}_i-\bar{\mathbf{x}}+\bar{\mathbf{x}}-\mu)'\\ &= \sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})'+\sum_{i=1}^{n}(\bar{\mathbf{x}}-\mu)(\mathbf{x}_i-\bar{\mathbf{x}})+\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\bar{\mathbf{x}}-\mu)+\sum_{i=1}^{n}(\bar{\mathbf{x}}-\mu)(\bar{\mathbf{x}}-\mu)'\\ &= \sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})'+\sum_{i=1}^{n}(\bar{\mathbf{x}}-\mu)(\mathbf{x}_i-\bar{\mathbf{x}})+\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\bar{\mathbf{x}}-\mu)+n(\bar{\mathbf{x}}-\mu)(\bar{\mathbf{x}}-\mu)'\end{aligned}$$

Ahora, como sabemos que $\bar{\mathbf{x}}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ entonces veamos a que son equivalentes los terminos de enmedio de la ecuación anterior,

$$\begin{aligned}\sum_{i=1}^{n}(\bar{\mathbf{x}}-\mu)(\mathbf{x}_i-\bar{\mathbf{x}})' &= \sum_{i=1}^{n}(\bar{\mathbf{x}}\mathbf{x}_i'-\mu\mathbf{x}_i'-\bar{\mathbf{x}}\bar{\mathbf{x}}'+\mu\bar{\mathbf{x}}')=n\bar{\mathbf{x}}\bar{\mathbf{x}}'-n\mu\bar{\mathbf{x}}'-n\bar{\mathbf{x}}\bar{\mathbf{x}}'+n\mu\bar{\mathbf{x}}'=0,\\ \sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\bar{\mathbf{x}}-\mu)' &= \sum_{i=1}^{n}(\mathbf{x}_i\bar{\mathbf{x}}'-\bar{\mathbf{x}}\bar{\mathbf{x}}'-\mu\mathbf{x}_i+\bar{\mathbf{x}}\mu)=n\bar{\mathbf{x}}\bar{\mathbf{x}}'-n\bar{\mathbf{x}}\bar{\mathbf{x}}'-n\mu\bar{\mathbf{x}}'+n\mu\bar{\mathbf{x}}=0.\end{aligned}$$

Por lo tanto, podemos concluir que

$$\begin{aligned}\sum_{i=1}^{n}(\mathbf{x}_i-\mu)(\mathbf{x}_i-\mu)' &= \sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})'+\underbrace{\sum_{i=1}^{n}(\bar{\mathbf{x}}-\mu)(\mathbf{x}_i-\bar{\mathbf{x}})'}_{=0}+\underbrace{\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\bar{\mathbf{x}}-\mu)'}_{=0}+n(\bar{\mathbf{x}}-\mu)(\bar{\mathbf{x}}-\mu)'\\ &= \sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})'+n(\bar{\mathbf{x}}-\mu)(\bar{\mathbf{x}}-\mu)' \quad \blacksquare.\end{aligned}$$

Ejercicio 2. Para el resultado:

> **Resultado: 1** *Para una matriz* $\mathbf{B}$ $(p\times p)$*, simétrica y positiva definida y un escalar* $b>0$*, se sigue que*
>
> $$\frac{1}{|\Sigma|^b}e^{-tr(\Sigma^{-1}B)/2}\leq\frac{1}{|B|^b}(2b)^{pb}e^{-pb}.$$
>
> *para toda* Σ *positiva definida de dimensión* $p\times p$*.*

Compruebe que la igualdad se sostiene únicamente para

$$\Sigma = \frac{1}{2b}\mathbf{B}.$$

RESPUESTA

Propiedad: 1 *Toda matriz* $\mathbf{A}$ *positiva definida tiene descomposión raíz cuadrada, y se define como*

$$\mathbf{A^{1/2}} = \sum_{\mathbf{i=1}}^{\mathbf{k}} \sqrt{\lambda}\mathbf{e_i}\mathbf{e_i'} \tag{1}$$

Esta matriz tiene las siguientes propiedades:

- $(\mathbf{A^{1/2}})' = \mathbf{A^{1/2}}$. *Es decir, es simétrica.*
- $A^{1/2}A^{1/2} = A$.
- $(A^{1/2})^{-1}$.
- $A^{1/2}A^{-1/2} = A^{-1/2}A^{1/2} = I$ *y* $A^{-1/2}A^{-1/2} = A^{-1}$.

Ocupando la propiedad 1, tenemos que la matriz $\mathbf{B}$ tiene descomposición raíz cuadrada denotada como $\mathbf{B^{1/2}}$, donde se cumple que $\mathbf{B^{1/2}B^{1/2}} = \mathbf{B}$, $\mathbf{B^{1/2}B^{-1/2}} = \mathbf{I}$ y $\mathbf{B^{-1/2}B^{-1/2}}$. Entonces, podemos observar que

$$tr(\mathbf{\Sigma^{-1}B}) = \mathbf{tr}\left[(\mathbf{\Sigma^{-1}B^{1/2}})\mathbf{B^{1/2}}\right] = \mathbf{tr}\left[\mathbf{B^{1/2}}(\mathbf{\Sigma^{-1}B^{1/2}})\right].$$

Ahora, cuando $\mathbf{\Sigma^{-1}} = \frac{\mathbf{1}}{\mathbf{2b}}\mathbf{B}$ tenemos que

$$B^{1/2}\Sigma^{-1}B^{1/2} = B^{1/2}\left(\frac{1}{2b}B\right)^{-1}B^{1/2} = 2bB^{1/2}B^{-1}B^{1/2} = 2bB^{1/2}B^{-1/2}B^{-1/2}B^{1/2} = 2bI_{p\times p}.$$

Ocupando lo anterior podemos obtener que

$$tr(\mathbf{\Sigma^{-1}B}) = \mathbf{tr}\left[\mathbf{B^{1/2}}(\mathbf{\Sigma^{-1}B^{1/2}})\right] = \mathbf{tr}[(\mathbf{2b})\mathbf{I_{p\times p}}] = \mathbf{2bp}.$$

Ahora, ocupando nuevamente las propiedades 1 tenemos

$$\frac{1}{|\Sigma|} = |\Sigma|^{-1}\frac{|B|}{|B|} = \frac{|B^{1/2}\Sigma^{-1}B^{1/2}|}{|B|} = \frac{|(2b)I_{p\times p}|}{|B|} = \frac{(2b)^p}{|B|}.$$

Por lo tanto, ocupando el resultado 1 considerando $\Sigma = \frac{\mathbf{1}}{\mathbf{2b}}\mathbf{B}$ tenemos que

$$\begin{aligned}
\frac{1}{|\Sigma|^b}e^{-tr(\Sigma^{-1}B)/2} &\leq \frac{1}{|B|^b}(2b)^{pb}e^{-pb} \\
\left(\frac{(2b)^p}{|B|}\right)^b e^{-2bp/2} &\leq \frac{1}{|B|^b}(2b)^{pb}e^{-pb} \\
\frac{1}{|B|^b}(2b)^{pb}e^{-bp} &= \frac{1}{|B|^b}(2b)^{pb}e^{-pb}.
\end{aligned}$$

Por lo tanto, **hemos demostrado que cuando** $\Sigma = \frac{1}{2b}B$ **se cumple la igualdad del resultado 1**. ■.

Ejercicio 3. Justifique el siguiente resultaddo para $p = 2$:

Resultado: 2 *Los contornos*

$$(\mathbf{x}-\mu)\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu)=\mathbf{c^2}$$

forman elipsoides concéntricos centrados en μ *y la longitud de los ejes esta dada por* $\pm c\sqrt{\lambda_i}e_i$*, donde* $\Sigma e_i=\lambda_i e_i$*, para* $i=1,\cdots,p$*.*

RESPUESTA

Para $p=2$, sea $X'AX$ (con $X\neq 0$) definido positivo y A una matriz simétrica. La distancia de X a un punto fijo μ es $(X-\mu)'A(X-\mu)$. Ahora consideremos a $X=\begin{pmatrix}X_1 & X_2\end{pmatrix}$ y $A=\begin{pmatrix}a_{11} & a_{12}\\ a_{12} & a_{22}\end{pmatrix}$. Sea A es simetrica positiva con λ_1 y λ_2 valores propios y e_1,e_2 vectores propios respectivamente, entonces por **el teorema de descomposición espectral tenemos que**

$$A=\lambda_1 e_1 e_1'+\lambda_2 e_2 e_2'.$$

Entonces podemos observar que,

$$\begin{aligned}c^2=X'AX&=X'(\lambda_1 e_1e_1'+\lambda_2e_2e_2')X\\&=\lambda_1X'e_1e_1'X+\lambda_2X'e_2e_2'X\\&=\lambda_1(X'e_1)^2+\lambda_2(X'e_2)^2\end{aligned}$$

Entonces, como sabemos que $X'AX$ representa una distancia cuadradada denotemosla como c, tenemos entonces que

$$c^2=\lambda_1y_1^2+\lambda_2y_2^2\Rightarrow \tag{2}$$

$$1=\left(\frac{\sqrt{\lambda_1}y_1}{c}\right)^2+\left(\frac{\sqrt{\lambda_2}y_2}{c}\right)^2 \tag{3}$$

por lo que podemos ver que tiene la ecuación de una elipse con centro en el origen, donde $y_1=x_1'e_1$ y $y_2=x_2'e_2$. Ahora despejamos a x_1 y x_2 para encontra los valores de los ejes.

$$c=\sqrt{\lambda_1}(x_1e_1')\Rightarrow ce_1=\sqrt{\lambda_1}(x_1e_1'e_1) \tag{4}$$

$$x_1=\frac{c}{\sqrt{\lambda_1}}e_1. \tag{5}$$

Para x_2

$$c=\sqrt{\lambda_2}(x_2e_2')\Rightarrow ce_2=\sqrt{\lambda_2}(x_2e_2'e_2) \tag{6}$$

$$x_2=\frac{c}{\sqrt{\lambda_2}}e_2. \tag{7}$$

Lo anterior, es considerando la distancia de X al origen, pero se puede extender considerando la distancia X a un punto fijo μ. La diferencia rádica en que la elipse formada tendrá ahora centro en μ, basta con ver la ecuación de la elipse (2). Ahora, considerando a $A=\Sigma$, donde Σ es positiva definida llegamos a que los ejes de la elipse estan dados por $\frac{c}{\sqrt{\lambda_1}}e_1$ y $\frac{c}{\sqrt{\lambda_2}}e_2$, donde λ_1 y λ_2 valores propios y e_1,e_2 vectores propios respectivamente.

Entonces, como nosotros estamos interesados en la distancia generaliza $(\mathbf{x}-\mu)\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu)=\mathbf{c^2}$, entonces utilizando que los valores propios y vectores propios de la inversa de una matriz positiva definida están relacionados con los valores propios y vectores propios de la matriz positiva definida de la siguiente forma:

Sea A una matriz positiva definida con λ_1 y λ_2 valores propios y e_1, e_2 vectores propios, entonces A^{-1} tiene $1/\lambda_1$ y $1/\lambda_2$ valores propios y los mismos vectores propios e_1, e_2. Por lo tanto, **podemos concluir que $(\mathbf{x}-\mu)\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu) = \mathbf{c^2}$ forma una elipse con centro en μ y eje de simétricas (ocupando lo anterior y sutituyendo en (4) y (6))**

$$\frac{c}{\sqrt{\frac{1}{\lambda_i}}} e_i = c\sqrt{\lambda_i} e_i \quad \blacksquare.$$

Ejercicio 4. En climas nórdicos, las carreteras debe ser limpiadas de la nieve rápidamente después de una tormenta. Una de las medidas de la severidad de la tormenta es $x_1 =$ *duración en horas*, mientras que la efectividad de la limpieza de la nieve se puede cuantificar por $x_2 =$ *horas de trabajo* para limpiar la nieve. En la tabla inferior se muestran los resultados de 25 incidentes en Wisconsin.

x_1	x_2	x_1	x_2	x_1	x_2
12.5	13.7	9.0	24.4	3.5	26.1
14.5	16.5	6.5	18.2	8.0	14.5
8.0	17.4	10.5	22.0	17.5	42.3
9.0	11.0	10.0	32.5	10.5	17.5
19.5	23.6	4.5	18.7	12.0	21.8
8.0	13.2	7.0	15.8	6.0	10.4
9.0	32.1	8.5	15.6	13.0	25.6
7.0	12.3	6.5	12.0		
7.0	11.8	8.0	12.8		

a) Detecte cualquier posible dato atípico mediante el diagrama de dispersión de las variables originales.

RESPUESTA

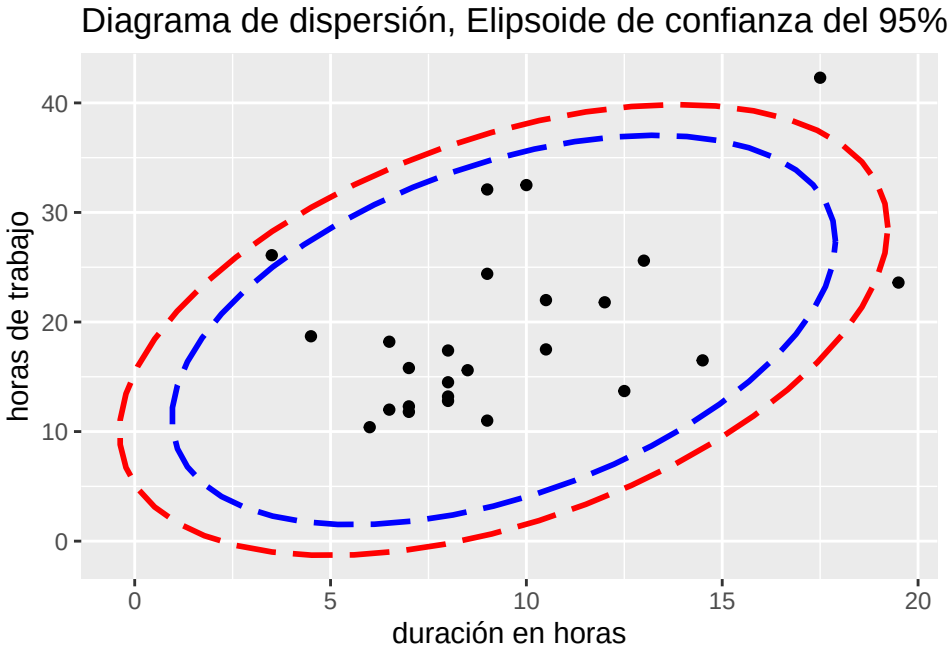
Sabemos que la distancia generalizada tiene una distribución conocida, en concreto, $(\mathbf{x}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu) \sim \chi^{\mathbf{2}}_{\mathbf{p}}$ con p grados de libertad. Entonces el elipsoide sólido de valores de x que satisface

$$(\mathbf{x}-\mu)'\mathbf{\Sigma^{-1}}(\mathbf{x}-\mu) \leq \chi^{\mathbf{2}}_{\mathbf{p}}(\alpha)$$

tiene una probabilidad $1-\alpha$ donde $\chi^2_p(\alpha)$ denota el percentil superior $(100\alpha)\,\%$ de la distribución χ^2_p. Ahora, si asumimos $\alpha = 0{,}05, 0{,}10$ podremos encontrar las elipsoides que se encuentran en la región del elipsoide de confianza del $95\,\%$ y $90\,\%$. Todo lo anterior se sigue cumpliendo si se usa $(\mathbf{x}-\mu)'\mathbf{S^{-1}}(\mathbf{x}-\mu) \sim \chi^{\mathbf{2}}_{\mathbf{p}}$ por el TLC. Ahora, con ayuda de R procedemos a calcular las regiones de confianza:

```
library(tidyverse)
# ingresamos los datos
x1 <- c( 12.5 ,14.5 ,8.0 ,9.0 , 19.5 , 8.0 ,9.0 , 7.0 , 7.0,  9.0, 6.5, 10.5, 10.0, 4.5,
         7.0, 8.5, 6.5, 8.0,  3.5, 8.0, 17.5, 10.5, 12.0, 6.0, 13.0)

x2 <- c(13.7, 16.5, 17.4, 11.0, 23.6, 13.2, 32.1, 12.3, 11.8, 24.4, 18.2, 22.0, 32.5,
        18.7, 15.8, 15.6, 12.0, 12.8, 26.1, 14.5, 42.3, 17.5, 21.8, 10.4,
25.6)

datos_ejer4 <- data.frame(x1, x2)

# graficamos las regiones de confianza y los puntos.
ggplot(datos_ejer4, aes(x1, x2))+
    geom_point() +
```

```
    labs(title="Diagrama de dispersión, Elipsoide de confianza del 95% y 90%",
         x="duración en horas", y="horas de trabajo") +
    stat_ellipse(aes(x=x1, y=x2), type="norm", col="red", size=1, linetype=5,
                 level=.95)+ # región de confianza al 95%
    stat_ellipse(aes(x=x1, y=x2), type="norm", col="blue", size=1, linetype=5,
                 level = .90) # región de confianza al 90%
```



Por lo tanto, **si observamos la regiones de confianza podemos detectar que la región de confianza al** $90\,\%$ **(linea azul) podemos detectar que 3 registros se encuentran fuera de ella, es decir, existen 3 outliers. Ahora si consideramos la región de confianza al** $95\,\%$ **(linea roja) observamos que tenemos solo dos outliers.** Dependiendo de que tan estrictos se quiera ser podemos concluir la existencia de 2 o 3 outlirers.

b) Determine la potencia de la transformación $\hat{\lambda}_1$ que convierte los valores de x_1 aproximadamente a normales. Construya el $Q-Q$ *plot* de las observaciones transformadas.

RESPUESTA

Box y Cox (1964) sugieren un método para encontrar una transformación apropiada a partir de la familia de transformaciones potencia dada por

$$x_i^{(\lambda)} = \begin{cases} \frac{x^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ ln(\lambda) & \lambda = 0 \end{cases} \tag{8}$$

Dada las observaciones $x_1, \cdots, x_n$, la solución de Box−Cox para elegir una potencia apropiada λ, es la solución que maximiza la expresión

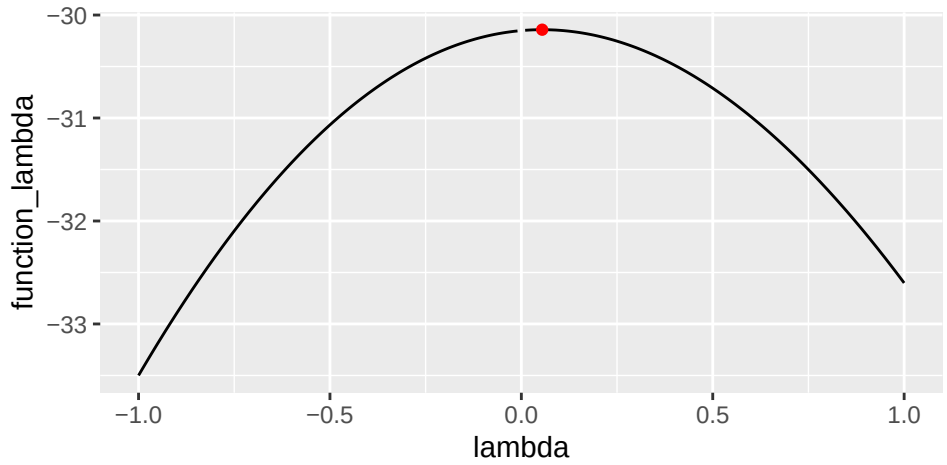
$$t(\lambda) = -\frac{n}{2} ln \left[\frac{1}{n} \sum_{i=1}^{n} \left(x_i^{(\lambda)} - \bar{x}^{(\lambda)} \right) \right] + (\lambda + 1) \sum_{i=1}^{n} ln(x_i),$$

donde

$$\bar{x}^{(\lambda)} = \frac{1}{n} \sum_{i=1}^{n} x_i^{(\lambda)} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i^{\lambda} - 1}{\lambda} \right)$$

Entonces, pasaremos a construir la función $t(\lambda)$ para encontrar de forma númerica el valor de λ que maximice la función anterior (ocupamos la función *optimize* en R.).

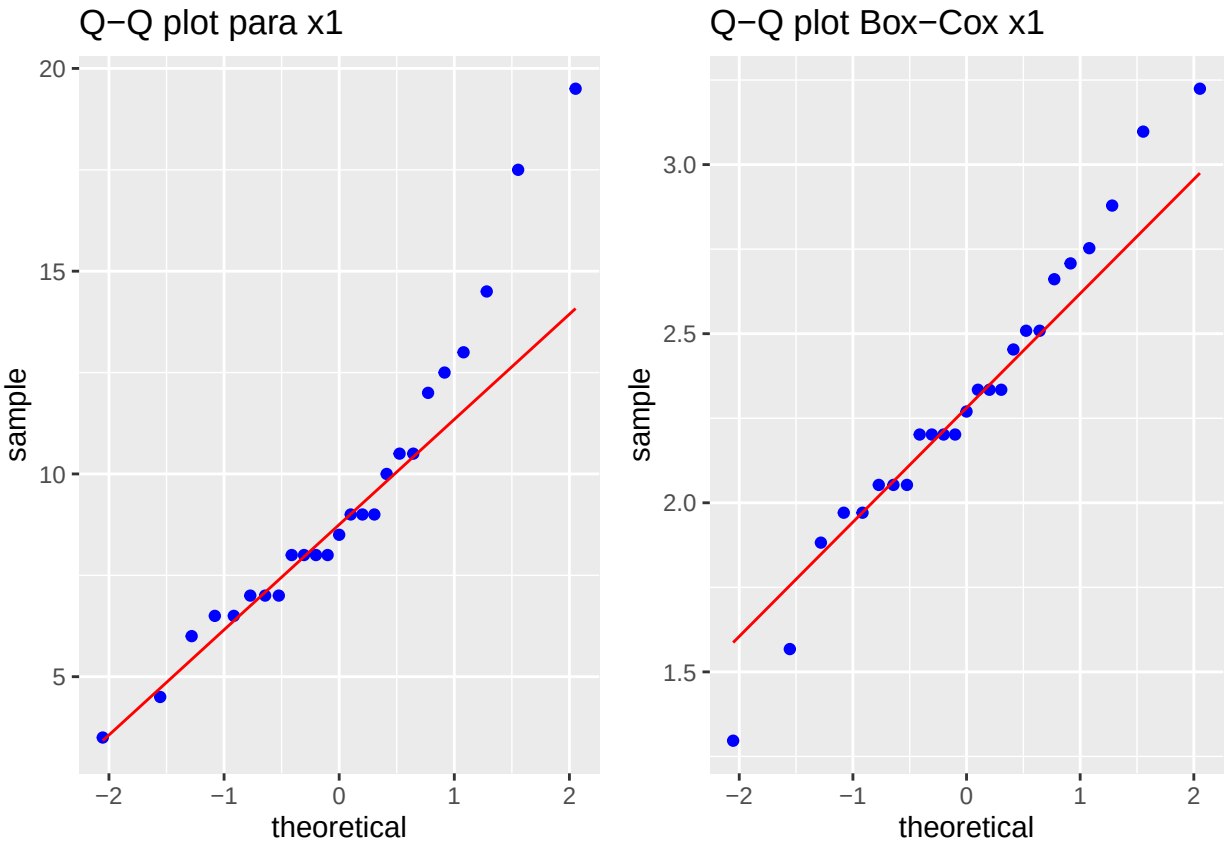
```
x <- datos_ejer4$x1
# función box-cox
t_lambda <- function(par){
    n <- length(x)
    if (par!=0){
        x_lambda <- (x^par -1)/par
    }
    else {x_lambda <- log(par)}
    x_bar_lambda <- mean(x_lambda)

    -(n/2)*log((1/n)*sum((x_lambda-x_bar_lambda)^2))+(par-1)*sum(log(x))
}

# encontramos el maximo de la función.
point_max <- optimize(f=t_lambda, interval = c(-5,5), maximum=TRUE)
point_max <- data.frame(lambda=point_max$maximum, t_lambda_re=point_max$objective)

function_box <- data.frame(lambda=seq(-1,1, 0.01))
function_box$t_lambda_re <- sapply(function_box$lambda, FUN=t_lambda)

ggplot(function_box, aes(lambda, t_lambda_re))+
    geom_line()+
    geom_point(data=point_max, aes(lambda, t_lambda_re), color="red")+
    labs(y="function_lambda")
```



Por lo tanto, **la transformación $\hat{\lambda_1}$ que convierte los valores de x_1 aproximadamente a normales es igual a** 0.0545014. Entonces, realizemos la transformación con este valor de λ y grafiquemos el $Q-Q$ de los datos originales y tranformados.

```
library(gridExtra)
datos_ejer4$x1_trans <- (x^point_max$lambda-1)/point_max$lambda

qq_x1 <- ggplot(data=datos_ejer4, aes(sample=x1))+
    geom_qq(color="blue")+
    stat_qq_line(color="red")+
```

```
labs(title = "Q-Q plot para x1")

qq_x1_tranform <- ggplot(data=datos_ejer4, aes(sample=x1_trans))+
    geom_qq(color="blue")+
    stat_qq_line(color="red")+
labs(title = "Q-Q plot Box-Cox x1")

grid.arrange(qq_x1, qq_x1_tranform, ncol=2) # Ponemos las gráficas juntas.
```



Observando los graficos $Q-Q$ podemos notar que cuando se aplica la transformación de Box−Cox da indicios de que los datos transformados sean normales, mientras que los datos originales no se observa normalidad.

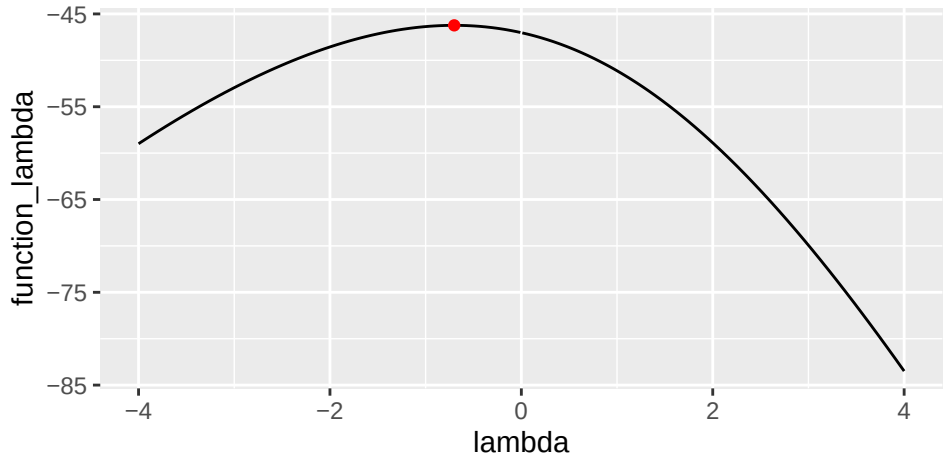
(c) Determine la potencia de la transformación $\hat{\lambda}_2$ que convierte los valores de x_2 aproximadamente a normales. Construya el $Q-Q$ *plot* de las observaciones transformadas.

RESPUESTA

Realicemos la misma metodología explicada en el inciso anterior pero ahora para la variable x_2.

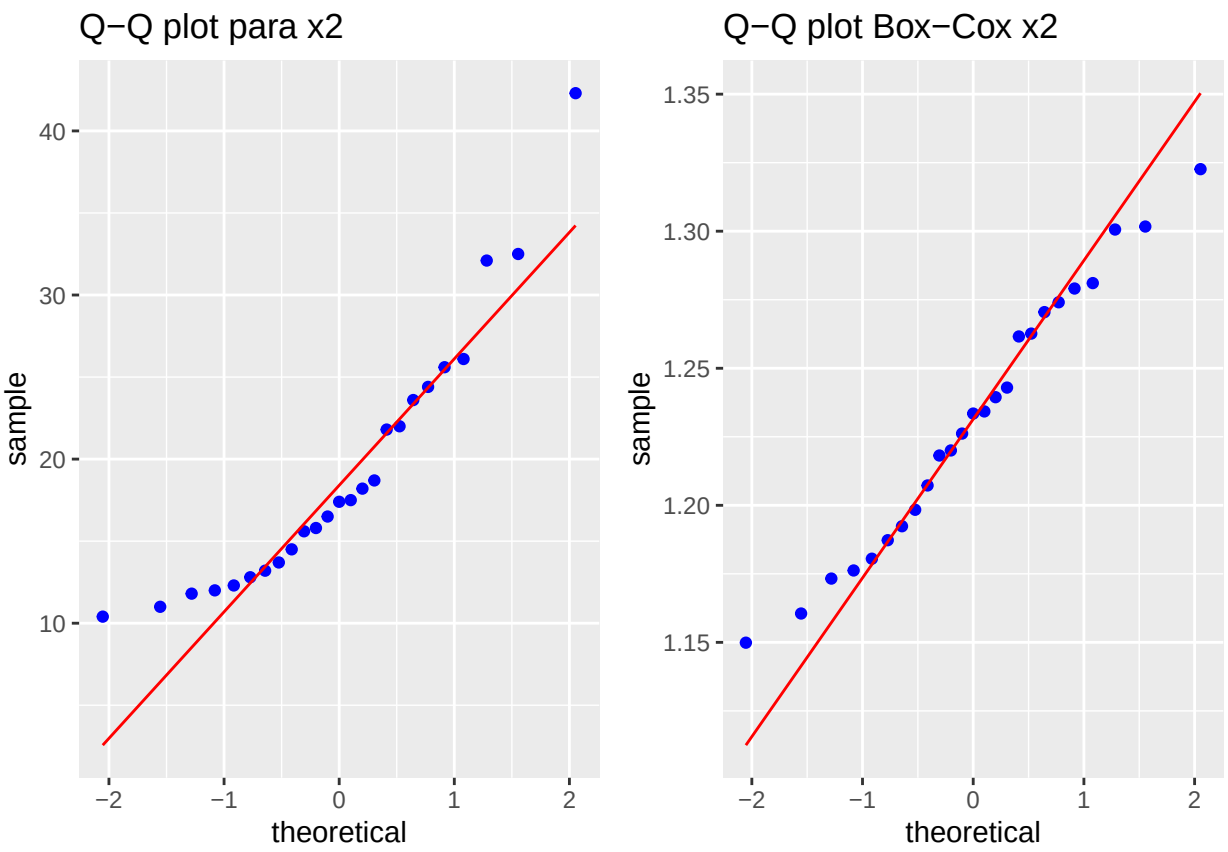
```
x <- datos_ejer4$x2

# encontramos el maximo de la función.
point_max <- optimize(f=t_lambda, interval = c(-5,5), maximum=TRUE)
point_max <- data.frame(lambda=point_max$maximum, t_lambda_re=point_max$objective)

function_box <- data.frame(lambda=seq(-4,4, 0.01))
function_box$t_lambda_re <- sapply(function_box$lambda, FUN=t_lambda)
```

```
ggplot(function_box, aes(lambda, t_lambda_re))+
    geom_line()+
    geom_point(data=point_max, aes(lambda, t_lambda_re), color="red")+
    labs(y="function_lambda")
```



Por lo tanto, **la transformación $\hat{\lambda}_2$ que convierte los valores de x_2 aproximadamente a normales es igual a** -0.7013795. Entonces, realizemos la transformación con este valor de λ y grafiquemos el $Q-Q$ de los datos originales y tranformados.

```
datos_ejer4$x2_trans <- (x^point_max$lambda-1)/point_max$lambda

qq_x2 <- ggplot(data=datos_ejer4, aes(sample=x2))+
    geom_qq(color="blue")+
    stat_qq_line(color="red")+
labs(title = "Q-Q plot para x2")

qq_x2_tranform <- ggplot(data=datos_ejer4, aes(sample=x2_trans))+
    geom_qq(color="blue")+
    stat_qq_line(color="red")+
labs(title = "Q-Q plot Box-Cox x2")

grid.arrange(qq_x2, qq_x2_tranform, ncol=2) # Ponemos las gráficas juntas.
```



Observando los graficos $Q-Q$ podemos notar que cuando se aplica la transformación de Box−Cox da indicios de que los datos transformados sean normales, mientras que los datos originales no se observa normalidad.

(d) Determine la potencia de la transformación que convierte las observaciones bivariadas en aproximadamente normales.

RESPUESTA

Cuando tenemos un conjunto de observaciones multivariadas $\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n}$ se selecciona una transformación de potencia para cada variable. Sean $\lambda_1, \cdots, \lambda_p$ las transformaciones de potencias para las p variables. Cada λ_k seleccionada maximiza

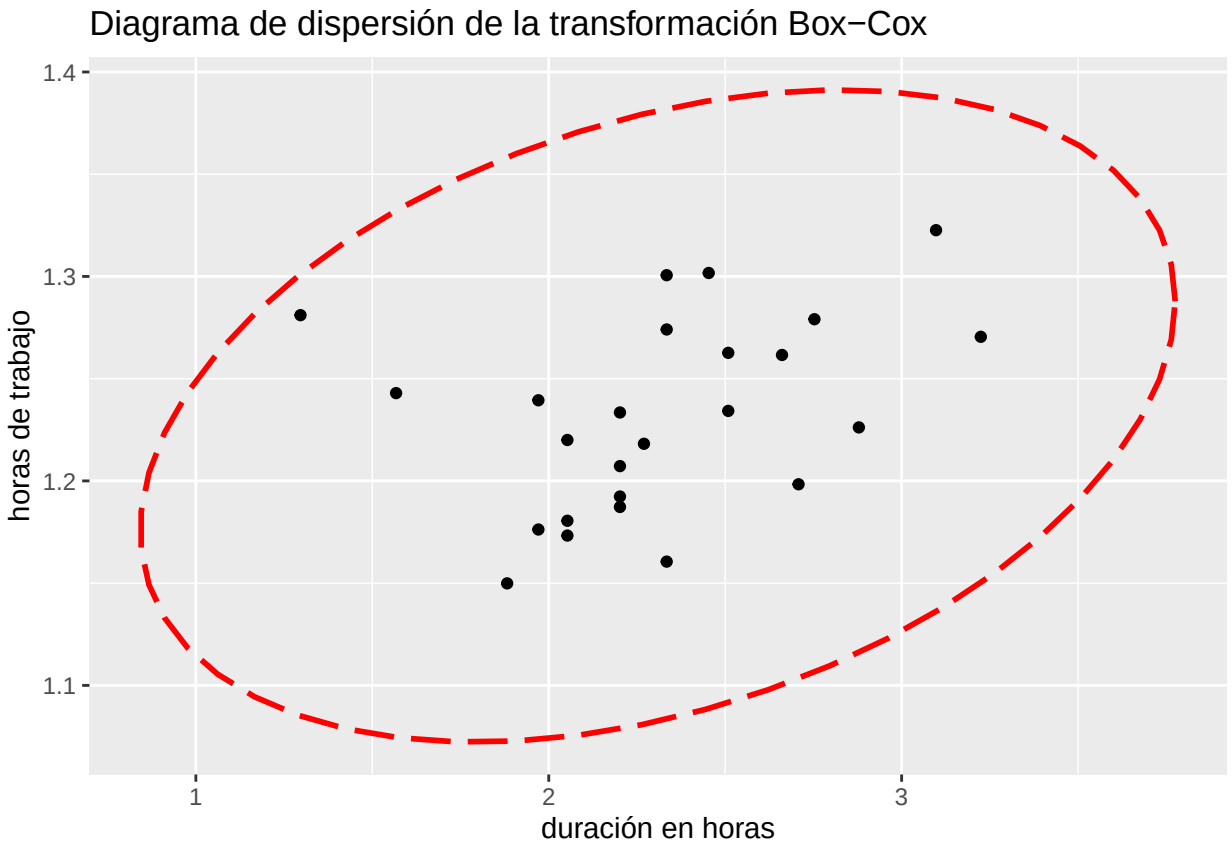
$$t_k(\lambda) = -\frac{n}{2} ln \left[\frac{1}{n} \sum_{i=1}^{n} \left(x_{ik}^{(\lambda)} - \bar{x_k}^{(\lambda)} \right) \right] + (\lambda + 1) \sum_{i=1}^{n} ln(x_{ik}),$$

Este procedimiento es equivalente a transformar cada distribución marginal a una distribución aproximadamente normal. Aunque el hecho de que las marginales sean normales no es sufiente para asegurar que la distribución conjunta sea normal. Sin embargo en aplicaciones práticas esto podría ser suficiente para asegurar normalidad conjunta. \

Por lo anterior, podemos considerar los $\hat{\lambda}_1$ y $\hat{\lambda}_2$ de los incisos anteriores para tener una aproximación de distribución conjunta normal bivariada. Para mostrar eso, grafiquemos el diagrama de disperción de los pares (x_1, x_2), si esta tiene una forma a una elipse entonces apoyaría el supuesto de distribución normal bivariada.

```
ggplot(datos_ejer4, aes(x1_trans, x2_trans))+
    geom_point() +
    labs(title="Diagrama de dispersión de la transformación Box-Cox",
         x="duración en horas", y="horas de trabajo") +
```

```
    stat_ellipse(aes(x=x1_trans, y=x2_trans), type="norm", col="red", size=1, linetype=5,
                 level=0.99)
```



Por lo tanto, **observando la transformación Box−Cox podemos apoyar que tienen una distribución normal bivariada.**

Ejercicio 5. Para p y n fijos, genérese una muestra de tamaño N de una ley $T_2(p, n)$ de Hotelling. Para esto construya una función que tome como entradas los valores de n, p, N , y utilice un generador de números aleatorios gaussianos. Represénte los resultados mediante un histograma, y haga pruebas para diferentes valores de entrada.

RESPUESTA
Si $\mathbf{x} \sim N_p(\mu, \Sigma)$ y $(n-1)\mathbf{S} \sim \mathbf{W_p}(\mathbf{S}|\mathbf{\Sigma})$ la distribución de la variable escalar

$$T^2 = (\mathbf{x} - \mu)' S^{-1} (\mathbf{x} - \mu)$$

se denomina distribución T^2 de Hotelling con p y $n-1$ grados de libertad. Diremos que $T^2 \sim T^2(p, n-1)$. Además, tenemos que si $\bar{\mathbf{x}} \sim N_p(\mu, \Sigma/n)$ entonces la distribución de $n(\bar{\mathbf{x}} - \mu)' S^{-1} (\bar{\mathbf{x}} - \mu)$ es también una T^2 de Hotelling.

Además, como $\mathbf{S}$ converga a $\mathbf{\Sigma}$ cuando n es grande, entonces T^2 converge a la distancia Mahalanobis y por lo tanto la distribución de Hotelling converge a la distribución χ^2_p.

Ahora procedemos a construir la función en R apartir de una muestra multivarida normal, con media de medias aleatorias y Σ matriz diagonal (estamos suponiendo indenpendicia entre las variables) para simplificar los calculos.

```
library(MASS)
muestra_T2<- function (p, n, N){
```

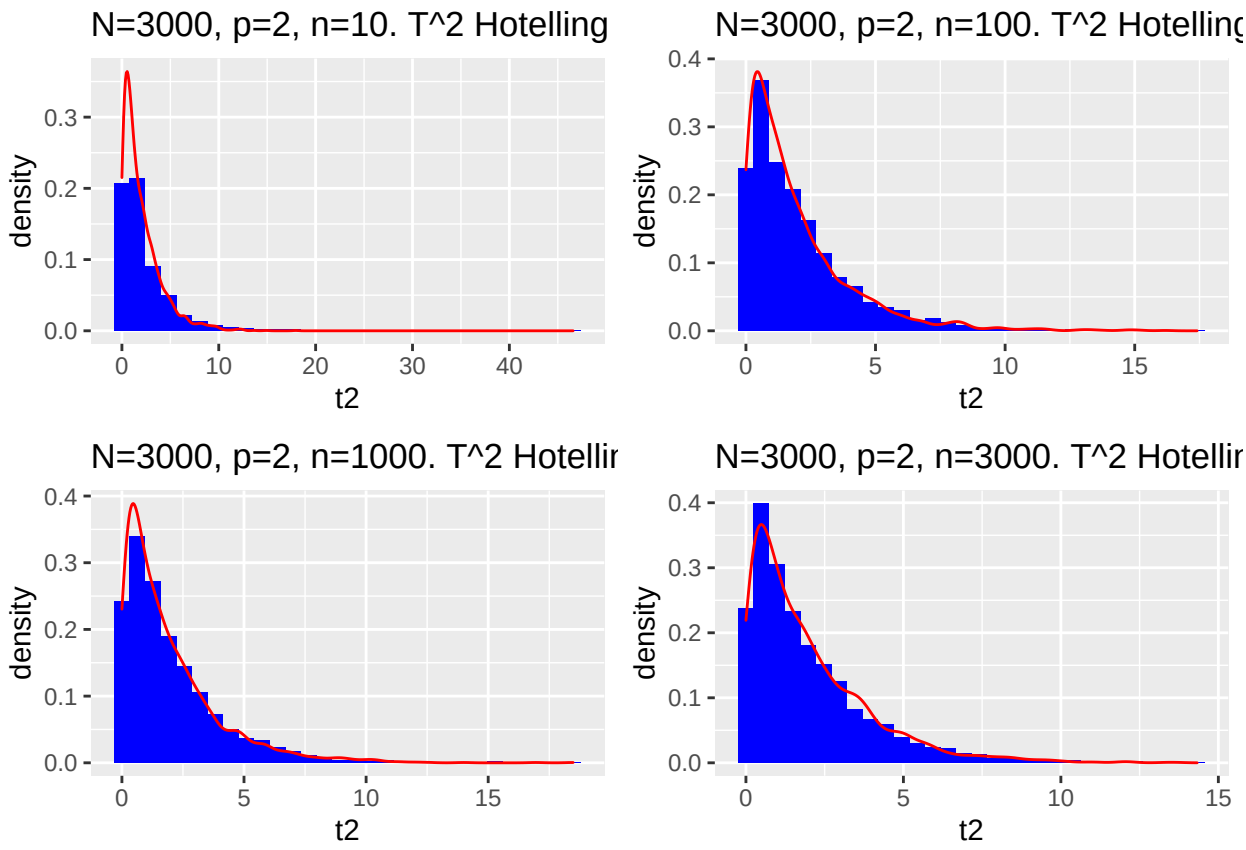
```
    mu <- rnorm(p) # inicializamos el vector de medias.
    sigma <- diag(rexp(p, 0.02),p) # inicializamos el
    random_T2 <- c()
    for (i in 1:N){
        random_guassian <- mvrnorm(n = n+1, mu=mu, Sigma=sigma)
        S <- cov(random_guassian)
        random_T2[i] <- n*(colMeans(random_guassian)-t(mu))%*%solve(S)%*%t(colMeans(random_guassi
    }
    random_T2
}
```

Graficaremos los histogramas para las distintas muestras. No encontre como un paquete o librería que simulará los números aleatorios de una T^2 de Hotelling en R, pero como sabemos que n es grande se aproxima a una distribución χ^2_p entonces las comparaciónes se realizaron con esta distribución. Creamos una función para hacer las gráficas.

```
# función para graficar las comparaciones
grap_t2_chi <- function(p, n, N){
    mue <- data.frame(t2 = muestra_T2(p, n,N), chi=rchisq(N,p))

    ggplot(mue, aes(x=t2))+
    geom_histogram(aes(y = ..density..), fill="blue")+
    geom_density(aes(chi), color="red")+
    labs(title=paste0("N=",N,", p=",p,", n=",n,". T^2 Hotelling y Chi-Square"))
}
```

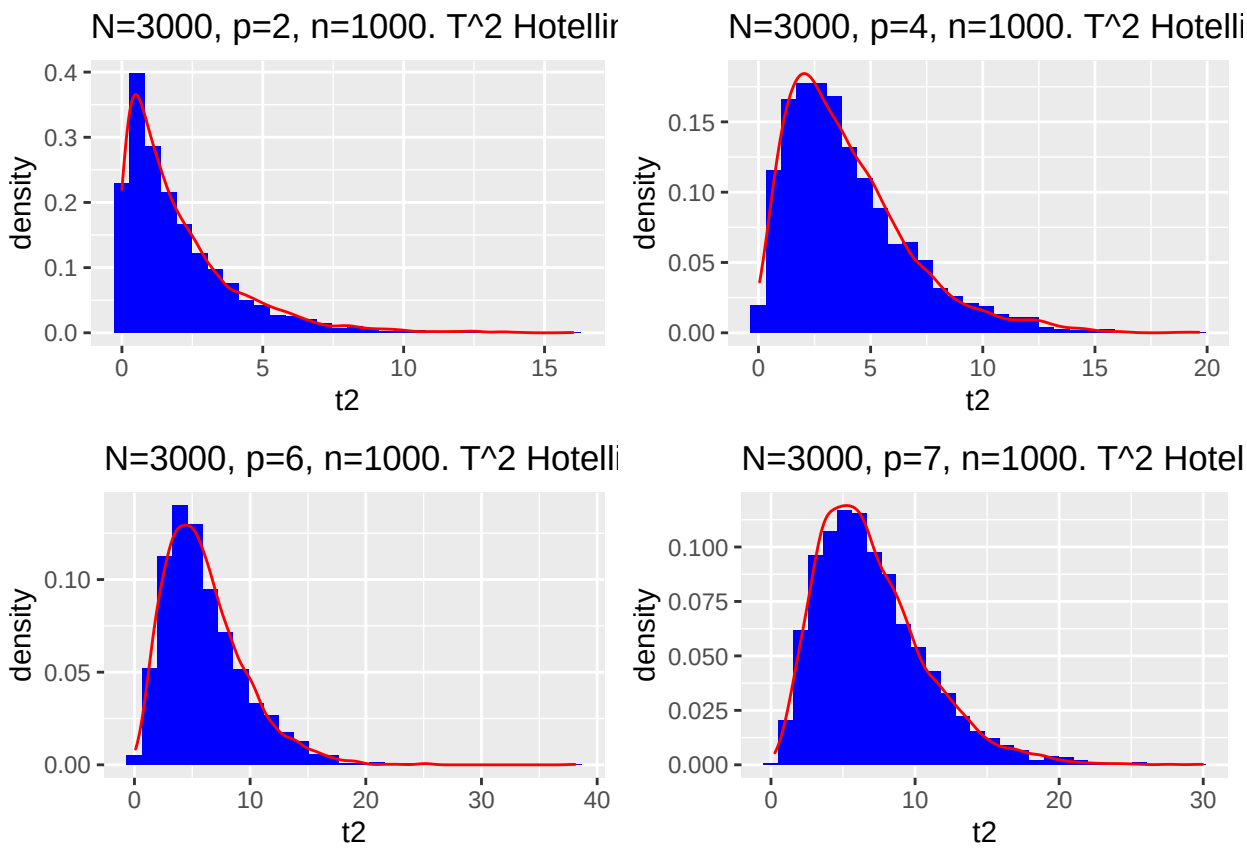
Creamos distintas muestras con distintos parametos. Los histogramas corresponden a la muestra generada de T^2 de Hotelling y la linea roja representa la función de densidad de la χ^2_p. Cuándo $p = 2$ y $N = 3000$ permancen constante tenemos,

```
gra_1 <- grap_t2_chi(p=2, n=10, N=3000)
gra_2 <- grap_t2_chi(p=2, n=100, N=3000)
gra_3 <- grap_t2_chi(p=2, n=1000, N=3000)
gra_4 <- grap_t2_chi(p=2, n=3000, N=3000)

grid.arrange(gra_1, gra_2, gra_3, gra_4, ncol=2, nrow=2) # Ponemos las gráficas juntas.
```



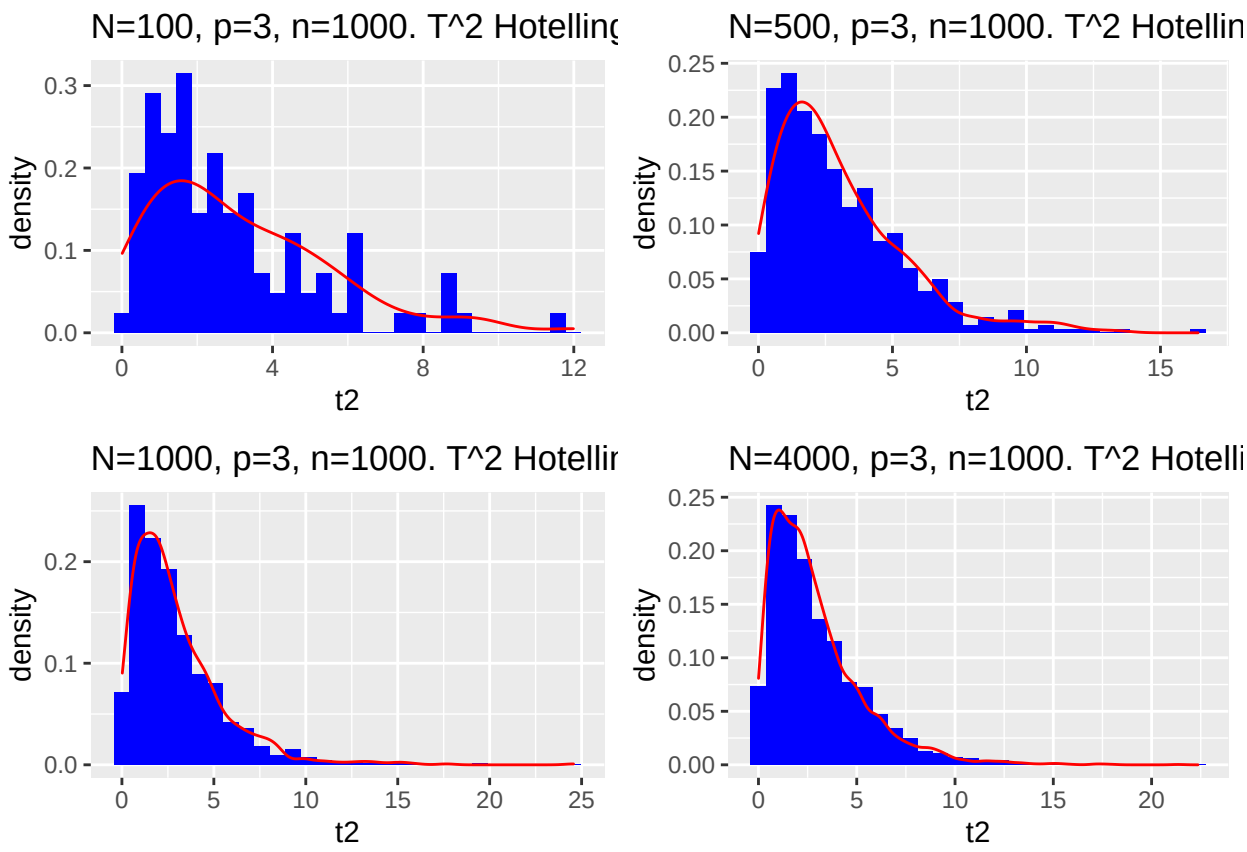
Es decir, cuando el tamaño de n más grande la distribución T^2 se va aproximando más a la distribución χ_2. Ahora veamos que pasa $n = 1000$ y $N = 3000$ permance constante

```
gra_1 <- grap_t2_chi(p=2, n=1000, N=3000)
gra_2 <- grap_t2_chi(p=4, n=1000, N=3000)
gra_3 <- grap_t2_chi(p=6, n=1000, N=3000)
gra_4 <- grap_t2_chi(p=7, n=1000, N=3000)

grid.arrange(gra_1, gra_2, gra_3, gra_4, ncol=2, nrow=2) # Ponemos las gráficas juntas.
```



Cuando p varia observamos como la distribuciones se van haciendo más simétricas. Es decir, podeíamos decir que p influye en la kurtosis de la distribución. Y por último, observemos que pasa cuando $p = 3$ y $n = 1000$ permance constante

```
gra_1 <- grap_t2_chi(p=3, n=1000, N=100)
gra_2 <- grap_t2_chi(p=3, n=1000, N=500)
gra_3 <- grap_t2_chi(p=3, n=1000, N=1000)
gra_4 <- grap_t2_chi(p=3, n=1000, N=4000)

grid.arrange(gra_1, gra_2, gra_3, gra_4, ncol=2, nrow=2) # Ponemos las gráficas juntas.
```



De igual manera, cuando N es grande la distribución T^2 se va aproximando a una distribución χ^2. ■.