

Universidad Simón Bolívar

Departamento de Cómputo científico y Estadística

Estadística para Ingenieros – CO3321

Enero-Marzo 2019

Resultados Laboratorio 2

Preparador: Andy Guevara

Arturo Yeppez. Carnet: 15-11551

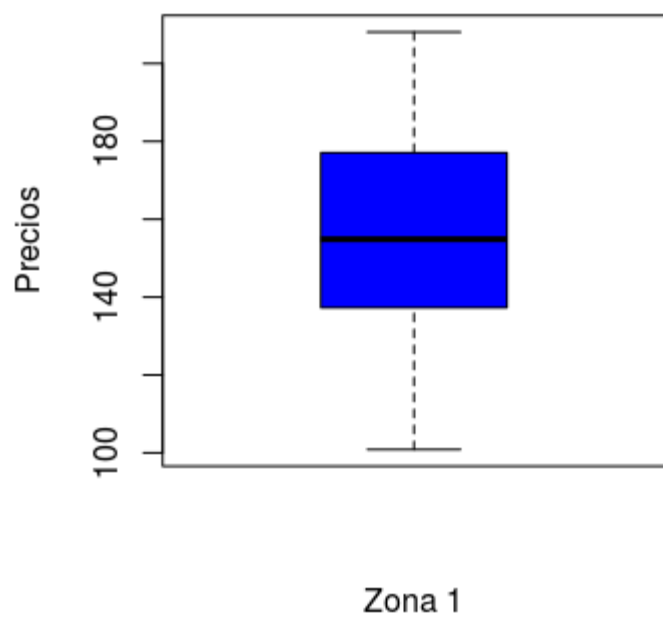
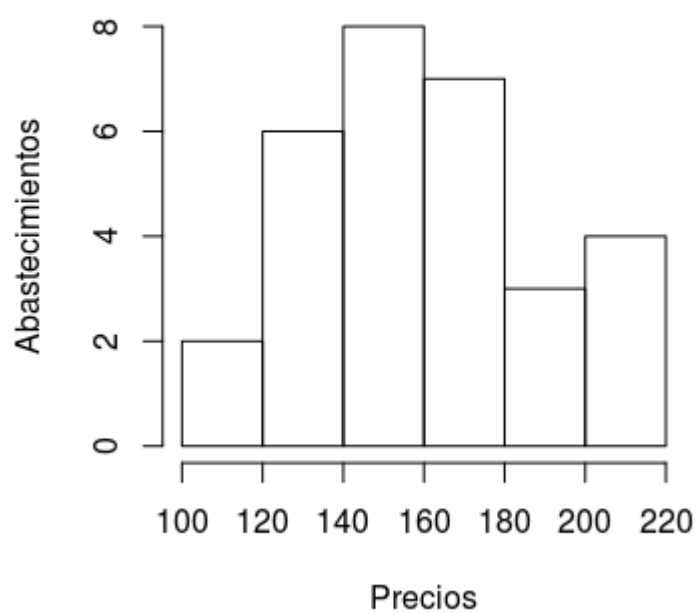
Antonella Requena. Carnet: 15-11196

Dada un estudio de datos en mercados de 4 zonas, que incluye en su recolección de datos los precios en dólares (\$) del kilo de carne y el tipo de carne al que se refiere. Se cargan los datos proporcionados por el archivo `zonas_carne.txt` con el procedimiento `read.table` y se almacenan en la variable `mercado`

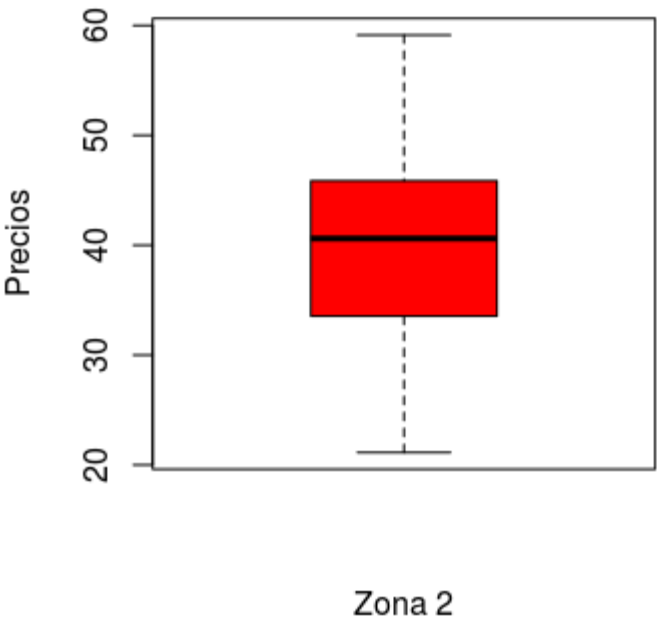
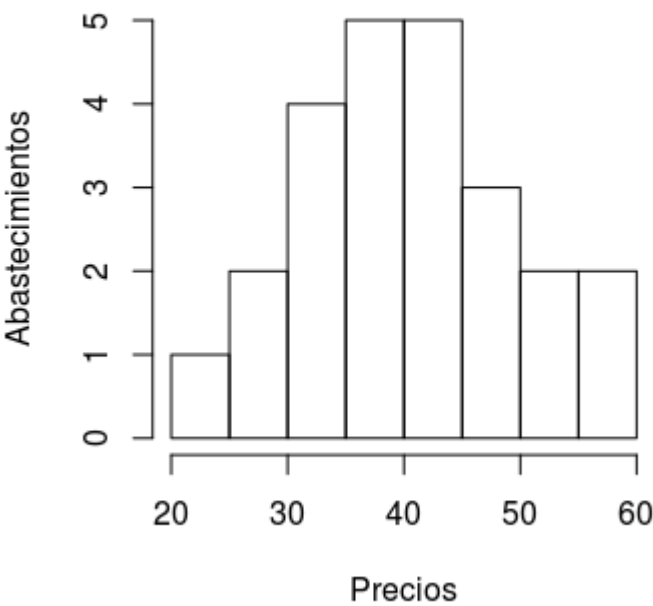
A continuación, se responden las preguntas correspondientes al laboratorio

1. Realice un análisis descriptivo de los precios de las zonas del tipo de carne de guisar. En el estudio considere histograma, boxplot y una tabla con información del mínimo, cuartiles, media y desviación estándar de cada zona.

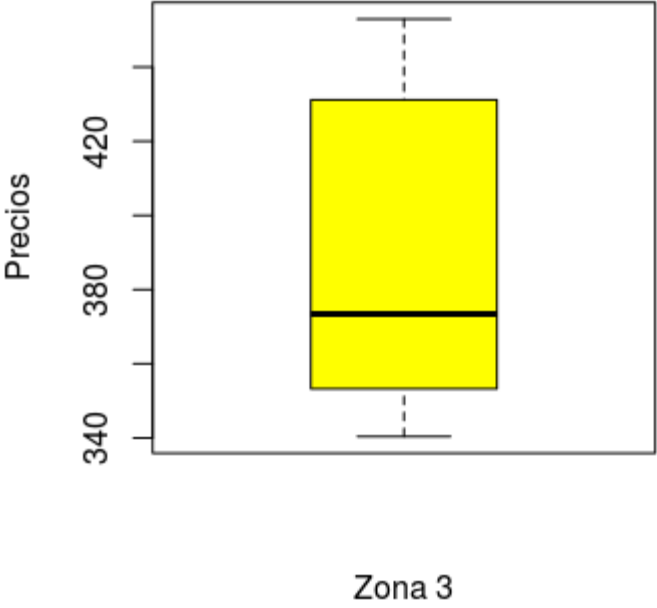
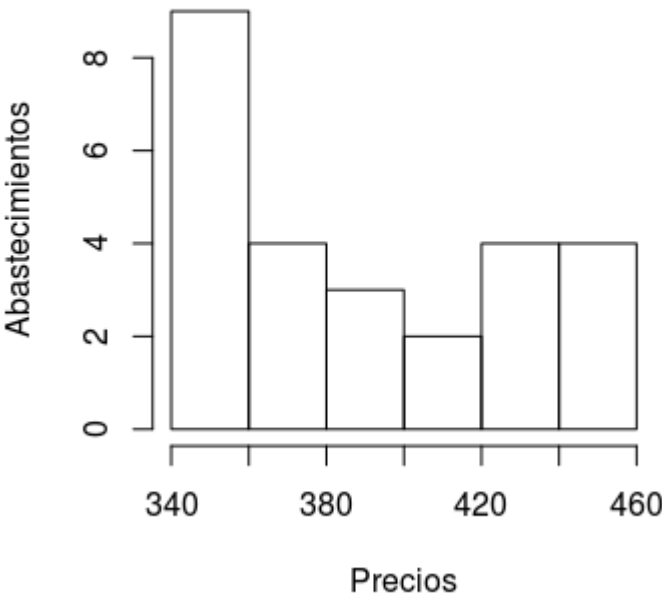
Precios de carne de guisar en Zona 1



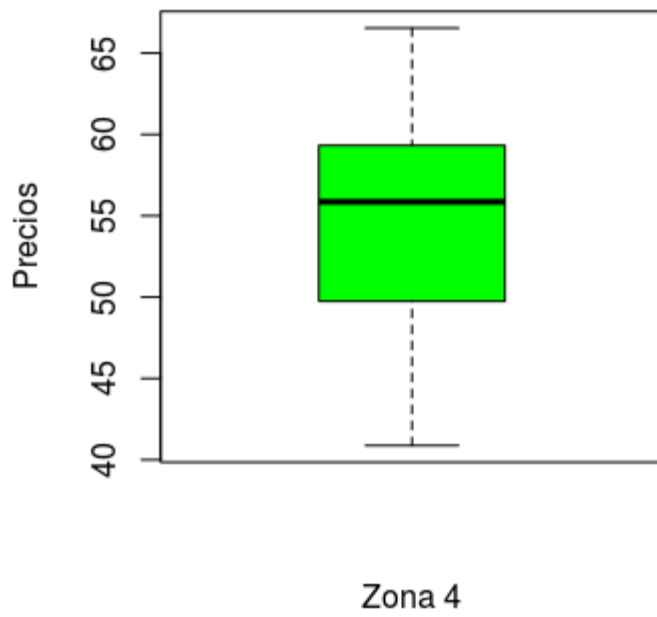
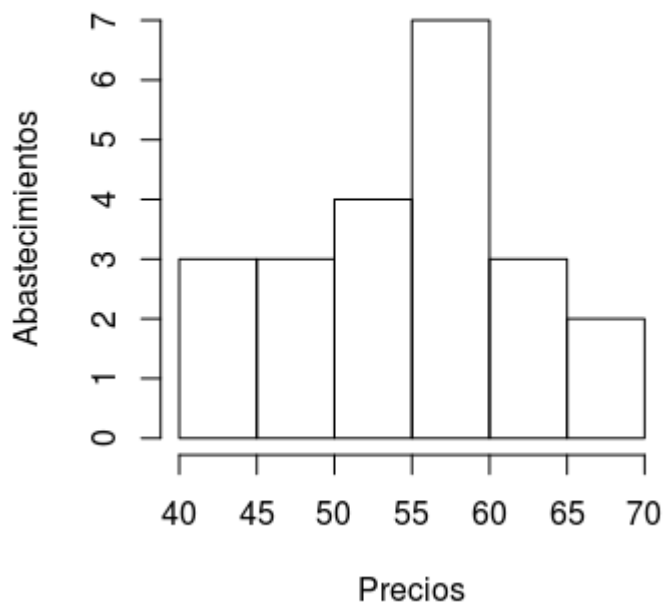
Precios de carne de guisar en Zona 2



Precios de carne de guisar en Zona 3



Precios de carne de guisar en Zona 4



	▲ Minimos ▼	Q1 ▼	Q2 ▼	Q3 ▼	Media ▼	DesviacionEstandar ▼
1	100.86	138.2950	154.870	177.0475	160.07733	29.157271
2	21.13	34.0225	40.605	45.6225	40.48833	10.066123
3	340.40	353.9775	373.425	431.0800	388.81885	40.882623
4	40.88	50.2425	55.860	59.1350	54.57136	7.543072

Basándonos en los histogramas, boxplots, y la tabla podemos concluir que, el mínimo de precio en la zona 1 es 100.86, que el 25% de los abastecimientos tienen los precios de kg de carne entre 100 y 130 dólares (primer cuantil), que el 50% de los abastecimientos lo tienen entre 130 y 150 dolares (segundo cuantil), y que el 75% entre 150 y 180 dolares. La media es 160.07 dolares. La desviación estándar es 29.15.

El mínimo de precio en la zona 2 es 21.13, el 25% de los abastecimientos tienen los precios de kg de carne entre 25 y 35 dólares (primer cuantil), el 50% de los abastecimientos lo tienen entre 35 y 45 dolares (segundo cuantil), y el 75% entre 45 y 48 dolares. La media es 40.4 dolares. La desviación estándar es 10.06.

El mínimo de precio en la zona 3 es 340.40, el 25% de los abastecimientos tienen los precios de kg de carne entre 340 y 345 dólares (primer cuantil), el 50% de los abastecimientos lo tienen entre 345 y 370 dolares (segundo cuantil), y el 75% entre 370 y 450 dolares. La media es 388.81 dolares. La desviación estándar es 40.88.

El mínimo de precio en la zona 4 es 40.88, el 25% de los abastecimientos tienen los precios de kg de carne entre 40 y 50 dólares (primer cuantil), el 50% de los abastecimientos lo tienen entre 50 y 55 dólares (segundo cuantil), y el 75% entre 55 y 58 dolares. La media es 54.57 dólares. La desviación estándar es 7.54.

2. *Determine un intervalo de confianza para la media de los precios de la carne de guisar para la Zona 1 y así mismo, para los precios de la Zona 3. (confianza de 99 %).*

Lo primero que hacemos es crear las listas que contengan los precios de la carne de guisar en la Zona 1 y el de la Zona 3, extrayendo los datos de precios de la Zona correspondiente y filtramos con el tipo de carne de Guisar.

Luego, con la función manual para intervalos de confianza calculamos el intervalo de confianza para la media del precio de la carne de guisar de la Zona 1 es:

[1] 145.4041 160.0773 174.7506

Esto lo que nos dice que el intervalo está comprendido entre (145.4041, 174.7506), así aseguramos que el precio medio de la carne para guisar en la Zona 1, un 99% está comprendido entre esos valores.

Por último, los resultados del intervalo de confianza de la Zona 3 son:

[1] 366.4699 388.8188 411.1678

De este intervalo de confianza comprendido entre (366.4699, 388.8188) podemos asegurar que el 99% de la veces que se compra carne para guisar en la Zona 3, su precio medio está comprendido entre eso.

De ambos resultados podemos observar que en general, sale más económico comprar carne para guisar en la Zona 1 que en la Zona 3.

3. ¿Cuál de los dos alimentos logra mayor precio en la carne de guisar, la Zona 2 o la Zona 4? Suponga $1 - \alpha = 0,99$

Como queremos obtener los vectores de precios de carne de tipo guisada de la zona 2 y 4, utilizamos la sintaxis de R para hacerlo

```
### PREGUNTA 3

# Creamos una lista con los precios solos de la carne de guisar de la zona 2 y zona 4
carne2 = as.vector(mercado$Zona2)
carne4 = as.vector(mercado$Zona4)

# Filtramos la columna con los tipos
tipo_carne2 = as.vector(mercado$Tipo2)
tipo_carne4 = as.vector(mercado$Tipo4)

# Filtramos los vectores de carne de zona 2 y 4 por Tipo Guisar
carne_guisar2 = carne2 [tipo_carne2 == "Guisar"]
carne_guisar4 = carne4 [tipo_carne4 == "Guisar"]
```

Una vez extraídos esos dos vectores, hacemos una prueba de hipótesis con el comando `var.test` que hace una prueba por hipótesis con hipótesis nula $\text{varianza1} = \text{varianza2}$, ya que si 1 está en el intervalo de confianza del 99%, podemos asumir que son iguales y calcular el intervalo de confianza del 99% usando las medias como estimadores con distribución t.

```
# Verificamos que podamos asumir que las varianzas sean iguales (i.e. que 1 este en el intervalo)
var.test(carne_guisar2, carne_guisar4, conf.level = 0.99)

# Como el resultado es, que 1 esta en el intervalo, podemos decir que las varianzas son iguales con un
# nivel de confianza de 99%

# Ahora calculamos el intervalo de confianza de las medias
t.test (carne_guisar2, carne_guisar4, var.equal = T, conf.level = 0.99 )
```

La salida por consola fue:

F test to compare two variances

data: carne_guisar2 and carne_guisar4

F = 1.7809, num df = 23, denom df = 21, p-value = 0.1883

alternative hypothesis: true ratio of variances is not equal to 1

99 percent confidence interval:

0.5620605 5.5018641

sample estimates:

ratio of variances

1.780853

Como 1 pertenece al intervalo (0.5620605,5.5018641), podemos decir que las varianzas son iguales con un 99% de confianza, es decir, rechazamos la hipótesis nula.

Así, al usar el comando t.test la salida fue

Suponiendo que el precio de la carne guisada en la zona 2 es en promedio, mayor que la de la zona 4

```
> t.test (carne_guisar2, carne_guisar4, var.equal = T, conf.level = 0.99, alternative = "greater" )
```

Two Sample t-test

data: carne_guisar2 and carne_guisar4

t = -5.3304, df = 44, p-value = 1

alternative hypothesis: true difference in means is greater than 0

99 percent confidence interval:

-20.46125 Inf

sample estimates:

mean of x mean of y

40.48833 54.57136

Suponiendo que el precio de la carne guisada en la zona 4 es en promedio, mayor que la de la zona 2

```
> t.test (carne_guisar4, carne_guisar2, var.equal = T, conf.level = 0.99, alternative = "greater" )
```

Two Sample t-test

data: carne_guisar4 and carne_guisar2

t = 5.3304, df = 44, p-value = 1.608e-06

alternative hypothesis: true difference in means is greater than 0

99 percent confidence interval:

7.704808 Inf

sample estimates:

mean of x mean of y

54.57136 40.48833

Como en el primer caso, con hipótesis alternativa $\text{zona2} > \text{zona4}$ nos queda que el p-valor es 1 y nuestro alfa es 0.01, entonces no podemos rechazar la hipótesis nula, es decir $\text{zona2} = \text{zona4}$

Luego, en el segundo caso, que $\text{zona4} > \text{zona2}$, el p-valor nos da un numero mas pequeño que 0.001 así que si podemos rechazar la hipótesis nula, es decir, el precio de la carne de tipo guisada en la zona 4 es mayor que el precio de la carne guisada en la zona 2 con un 99% de confianza.

4. Los investigadores presumen que la diferencia del gasto promedio en carne de la zona 2 y 4 es menor a 1 \$ es esta afirmación cierta?

Para averiguar si la afirmación es cierta, sacamos un intervalo de confianza de la diferencia de medias con el comando `t.test()` entre la lista de precios de carnes de la Zona 2 y la Zona 4. Como no especifican, el intervalo de confianza elegido fue de 99%.

Luego de aplicarlo, podemos ver que el intervalo de confianza fue:

99 percent confidence interval:

-23.27155 5.48240

De allí sacamos que el intervalo está comprendido entre (-23.27155, 5.4820), como el 0 está incluido en el intervalo de confianza no existe ninguna tendencia y no se puede llegar a ninguna conclusión respecto a los resultados, por eso no podemos afirmar que la diferencia sea menor a \$1.

5. Calcule un intervalo de confianza al 90 % para estimar la proporción de precios mayores a 300\$ de la Zona 3.

Como nos piden un intervalo de confianza de la proporción de precios mayores a \$300 en la Zona 3, utilizamos el comando `binom.test()` utilizando dos listas de precios, la lista de precios de carnes en la zona 3 y la lista de precios de carnes mayores a \$300 en la Zona 3.

Al aplicar, tenemos de resultados:

90 percent confidence interval:

0.8737934 0.9859991

De aquí, podemos concluir que un 90% de las veces la proporción de precios mayores a \$300 se encuentra entre (0.8737934, 0.9859991) de forma que, podemos asegurar que la mayoría de las veces encontraremos precios de carne mayores a \$300.

6. Una semana después se volvió a realizar la encuesta en la Zona 1, en los mismos abastecimientos. La información se encuentra en la columna Zona1B. El investigador presume que los precios han aumentado en dicha zona con relación a la semana anterior. ¿Es esta información correcta?

Despues de correr el comando t.test de R se obtuvo por consola:

```
# Utilizamos prueba por hipotesis
# HN: media de carne 1B = media de carne 1
# HA: media de carne 1B > media de carne 1
t.test(carne1B,carne1,alternative = "greater",conf.level = 0.95)
```

Welch Two Sample t-test

data: carne1B and carne1

t = 24.716, df = 59.281, p-value < 2.2e-16

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

1990.198 Inf

sample estimates:

mean of x mean of y

2290.4203 155.9127

Como el p-valor obtenido es mucho mas pequeño que 0.01, es poco probable no rechazar la hipótesis nula, además que nuestro alfa, que es 0.05, es mucho mayor, así que podmeos rechazar la hipótesis nula, es decir, que es probable que los precios hayan aumentado en una semana en la Zona 1.

CÓDIGO DEL LABORATORIO

Laboratorio 2

Alumnos: Arturo Yepez 15-11551, Antonella Requena 15-11196

LECTURA DE DATOS

Lectura de Tabla

```
mercado = read.table("zonas_carne.txt", header = T)
```

```
# Se crea la función de intervalos de confianza
```

```
intervalo <- function(x, alfa) {  
  n <- length(x)  
  nu <- n - 1 #grados de libertad  
  cuantil <- qt( 1 - alfa/2, df = nu )  
  LS <- mean( x ) + cuantil*sqrt( var( x ) / n )  
  LI <- mean( x ) - cuantil*sqrt( var( x ) / n )  
  return( c( LI, mean( x ), LS ) )  
}
```

```
#### PREGUNTA 1
```

```
# Datos a extraer
```

```
# Extraemos vector con los precios solos de la carne de guisar de  
las zonas
```

```
# Zona 1
```

```
carne1 = as.vector(mercado$Zona1)  
tipo_carne1 = as.vector(mercado$Tipo1)
```

```
carne_guisar1 = carne1 [tipo_carne1 == "Guisar"]
```

```
# Zona 2
```

```
carne2 = as.vector(mercado$Zona2)  
tipo_carne2 = as.vector(mercado$Tipo2)
```

```
carne_guisar2 = carne2 [tipo_carne2 == "Guisar"]
```

```
# Zona 3
```

```
carne3 = as.vector(mercado$Zona3)
```

```
tipo_carne3 = as.vector(mercado$Tipo3)
```

```
carne_guisar3 = carne3 [tipo_carne3 == "Guisar"]
```

```
# Zona 4
```

```
carne4 = as.vector(mercado$Zona4)
```

```
tipo_carne4 = as.vector(mercado$Tipo4)
```

```
carne_guisar4 = carne4 [tipo_carne4 == "Guisar"]
```

```
# Histogramas
```

```
hist(carne_guisar1, main = "Precios de carne de guisar en Zona 1",  
ylab = "Abastecimientos", xlab = "Precios")
```

```
hist(carne_guisar2, main = "Precios de carne de guisar en Zona 2",  
ylab = "Abastecimientos", xlab = "Precios")
```

```
hist(carne_guisar3, main = "Precios de carne de guisar en Zona 3",  
ylab = "Abastecimientos", xlab = "Precios")
```

```
hist(carne_guisar4, main = "Precios de carne de guisar en Zona 4",  
ylab = "Abastecimientos", xlab = "Precios")
```

```
# Boxplots
```

```
boxplot(carne_guisar1, ylab="Precios", xlab="Zona 1", col=c("blue"))
```

```
boxplot(carne_guisar2, ylab="Precios", xlab="Zona 2", col=c("red"))
```

```
boxplot(carne_guisar3, ylab="Precios", xlab="Zona 3",  
col=c("yellow"))
```

```
boxplot(carne_guisar4, ylab="Precios", xlab="Zona 4",  
col=c("green"))
```

```
# Tabla
```

```

M = data.frame(stringsAsFactors = FALSE,
Zonas=c('Zona1', 'Zona2', 'Zona3', 'Zona4'),
Minimos=c(min(carne_guisar1),min(carne_guisar2),min(carne_guisar3),min(carne_guisar4)),
Q1=c(quantile(carne_guisar1,0.25),quantile(carne_guisar2,0.25),quantile(carne_guisar3,0.25),quantile(carne_guisar4,0.25)),
Q2=c(quantile(carne_guisar1,0.50),quantile(carne_guisar2,0.50),quantile(carne_guisar3,0.50),quantile(carne_guisar4,0.50)),Q3=c(quantile(carne_guisar1,0.75),quantile(carne_guisar2,0.75),quantile(carne_guisar3,0.75),quantile(carne_guisar4,0.75)),Media=c(mean(carne_guisar1),mean(carne_guisar2),mean(carne_guisar3),mean(carne_guisar4)),
DesviacionEstandar=c(sd(carne_guisar1),sd(carne_guisar2),sd(carne_guisar3),sd(carne_guisar4)))

```

PREGUNTA 2

```

# Creamos una lista con los precios solos de la carne de guisar de la zona 1

```

```

carne1 = as.vector(mercado$Zona1)
tipo_carne1 = as.vector(mercado$Tipo1)

```

```

carne_guisar1 = carne1 [tipo_carne1 == "Guisar"]

```

```

# Creamos una lista con los precios solos de la carne de guisar de la zona 3

```

```

carne3 = as.vector(mercado$Zona3)
tipo_carne3 = as.vector(mercado$Tipo3)

```

```

carne_guisar3 = carne3 [tipo_carne3 == "Guisar"]

```

```

# Determinamos el intervalo de Confianza de la Zona 1

```

```

int_conf1 = intervalo(carne_guisar1, 0.01)

```

```
# Determinamos el intervalo de Confianza de la Zona 3
```

```
int_conf3 = intervalo(carne_guisar3, 0.01)
```

```
#### PREGUNTA 3
```

```
# Creamos una lista con los precios solos de la carne de guisar de la zona 2 y zona 4
```

```
carne2 = as.vector(mercado$Zona2)
```

```
carne4 = as.vector(mercado$Zona4)
```

```
# Filtramos la columna con los tipos
```

```
tipo_carne2 = as.vector(mercado$Tipo2)
```

```
tipo_carne4 = as.vector(mercado$Tipo4)
```

```
# Filtramos los vectores de carne de zona 2 y 4 por Tipo Guisar
```

```
carne_guisar2 = carne2 [tipo_carne2 == "Guisar"]
```

```
carne_guisar4 = carne4 [tipo_carne4 == "Guisar"]
```

```
# Verificamos que podamos asumir que las varianzas sean iguales (i.e. que 1 este en el intervalo)
```

```
var.test(carne_guisar2, carne_guisar4, conf.level = 0.99)
```

```
# Como el resultado es, que 1 esta en el intervalo, podemos decir que las varianzas son iguales con un
```

```
# nivel de confianza de 99%
```

```
# Ahora calculamos el intervalo de confianza de las medias
```

```
t.test (carne_guisar2, carne_guisar4, var.equal = T, conf.level = 0.99 )
```

PREGUNTA 4

```
# Creamos una lista con los precios de carne de la Zona 2
carne2 = as.vector(mercado$Zona2)
```

```
# Creamos una lista con los precios de carne de la Zona 4
carne4 = as.vector(mercado$Zona4)
```

```
# Ahora sacamos el intervalo de confianza
int_conf_dif = t.test(carne2, carne4, conf.level = 0.9)
```

PREGUNTA 5

```
# Tomamos carne3 como la lista de precios de la Zona 3 y sacamos su longitud
```

```
len_carne3 = length(carne3)
```

```
# Calculamos una lista con los precios de la Zona 3 mayores a 300 y sacamos su longitud
```

```
max_carne3 = carne3 [carne3 >= 300]
```

```
len_max_carne3 = length(max_carne3)
```

```
# Ahora, calculamos el intervalo de confianza de para proporciones
```

```
int_conf3_propor = binom.test(len_max_carne3, len_carne3, conf.level = 0.99)
```

PREGUNTA 6


```
# Asumimos que el estadístico de prueba es la media

# Creamos una lista con los precios de la zona 1B (zona 1 ya lo
tenemos)
carne1B = as.vector(mercado$Zona3B)

# Utilizamos prueba por hipótesis
# H0: media de carne 1B = media de carne 1
# Ha: media de carne 1B > media de carne 1
t.test(carne1B,carne1,alternative = "greater",conf.level = 0.95)
```