

Universidad Simón Bolívar

Departamento de Cómputo científico y Estadística

Estadística para Ingenieros – CO3321

Enero-Marzo 2019

Resultados Laboratorio 1

Preparador: Andy Guevara

Arturo Yépez. Carnet: 15-11551

Antonella Requena. Carnet: 15-11196

Dada una muestra de datos de un mercado, que incluye en su recolección datos del genero del consumidor, su edad, qué fruta compró, el cárnico, los vegetales, y el consumo total que realizó en el mercado. Se cargan los datos proporcionados por el archivo `mercado.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 las variables vegetales y consumo, apoyándose en los gráficos y el resumen de los datos

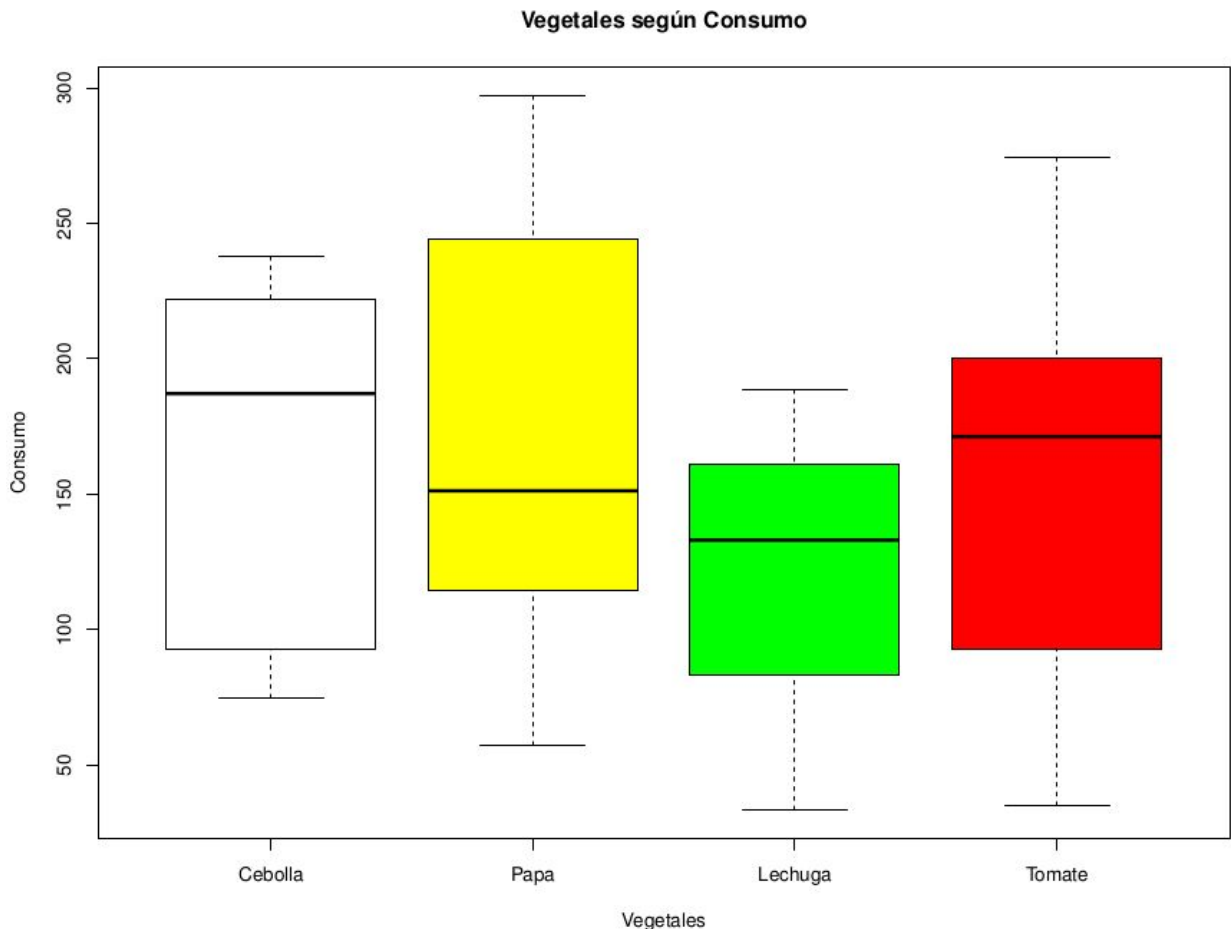


Figura 1. Gráficos de caja por tipo de verdura en relación a su consumo por ambos géneros

Según la gráfica, para *cebolla* podemos ver que la mediana es aproximadamente 180. El primer cuantil (25% de la muestra) consume entre 100 y 110 y el tercer cuantil (25% restante) consume entre 220 y 230.

Para *papa*, la mediana es aproximadamente 150. El primer cuantil (25% de la muestra) consume entre 50 y 120 y el tercer cuantil (25% restante) consume entre 230 y 300.

Para *lechuga*, la mediana de la muestra es aproximadamente 140. El primer cuantil (25% de la muestra) consume entre 30 y 80 y el tercer cuantil (25% restante) consume entre 160 y 190.

Para *tomate*, la mediana es aproximadamente 180. El primer cuantil (25% de la muestra) consume entre 30 y 90 y el tercer cuantil (25% restante) consume entre 200 y 290.

Ahora, con el comando `summary` de R podemos obtener los datos exactos:

```
> summary(consumo_cebolla)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 74.99 105.38 187.19 168.93 221.28 237.69

> summary(consumo_papa)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 57.41 114.58 151.32 168.99 244.13 297.25

> summary(consumo_lechuga)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 33.5   83.3   133.1   118.4   160.9   188.7

> summary(consumo_tomate)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 35.15  92.71 171.30 152.37 200.21 274.30
```

El rango de consumo por **cebolla** está entre 74.99 y 237.69

La mediana es 187.19

La media es 168.93

Para el primer cuartil se reporta un consumo de 105.38

Para el tercer cuartil se reporta un consumo de 221.28

El rango de consumo por **papa** está entre 57.41 y 297.25

La mediana es 151.32

La media es 168.99

Para el primer cuartil se reporta un consumo de 57.41

Para el tercer cuartil se reporta un consumo de 244.13

El rango de consumo por **lechuga** está entre 33.5 y 188.7

La mediana es 133.1

La media es 118.4

Para el primer cuartil se reporta un consumo de 83.3

Para el tercer cuartil se reporta un consumo de 160.9

El rango de consumo por **tomate** está entre 33.15 y 274.30

La mediana es 171.30

La media es 152.37

Para el primer cuartil se reporta un consumo de 92.71

Para el tercer cuartil se reporta un consumo de 200.21

Calcularemos ahora la moda implementando la función tomando el código proporcionado en las láminas del curso. Calculando la moda se obtiene que no se repite ningún valor.

Calculamos la varianza utilizando la función `var()` de R obteniendo los siguientes valores:

```
> var(consumo_cebolla)
[1] 4187.635

> var(consumo_papa)
[1] 5186.619

> var(consumo_lechuga)
[1] 6179.581

> var(consumo_tomate)
[1] 6899.713
```

La varianza de consumo de **cebolla** es 4187.635

La varianza de consumo de **papa** es 5186.619

La varianza de consumo de **lechuga** es 6179.581

La varianza de consumo de **tomate** es 6899.713

Por último, calculamos la desviación estándar utilizando la función `sd()` de R obteniendo los siguientes valores:

```
> sd(consumo_cebolla)
[1] 64.71194

> sd(consumo_papa)
[1] 72.01819

> sd(consumo_lechuga)
[1] 78.61031

> sd(consumo_tomate)
[1] 83.06451
```

La desviación estándar de consumo de **cebolla** es 64.711

La desviación estándar de consumo de **papa** es 72.018

La desviación estándar de consumo de **lechuga** es 78.610

La desviación estándar de consumo de **tomate** es 83.064

2. *Analice la relación en el dinero gastado en las mujeres que compraron cambur según el tipo de cárnico.*

Antes de realizar el análisis, se agruparon los datos en 3 vectores filtrados mediante el cárnico consumido, que fuese género femenino y que haya consumido de

fruta cambur. Para el análisis de estos cuatro vectores resultantes sacamos un summary de cada uno y luego un boxplot, correspondiente a la figura 2.

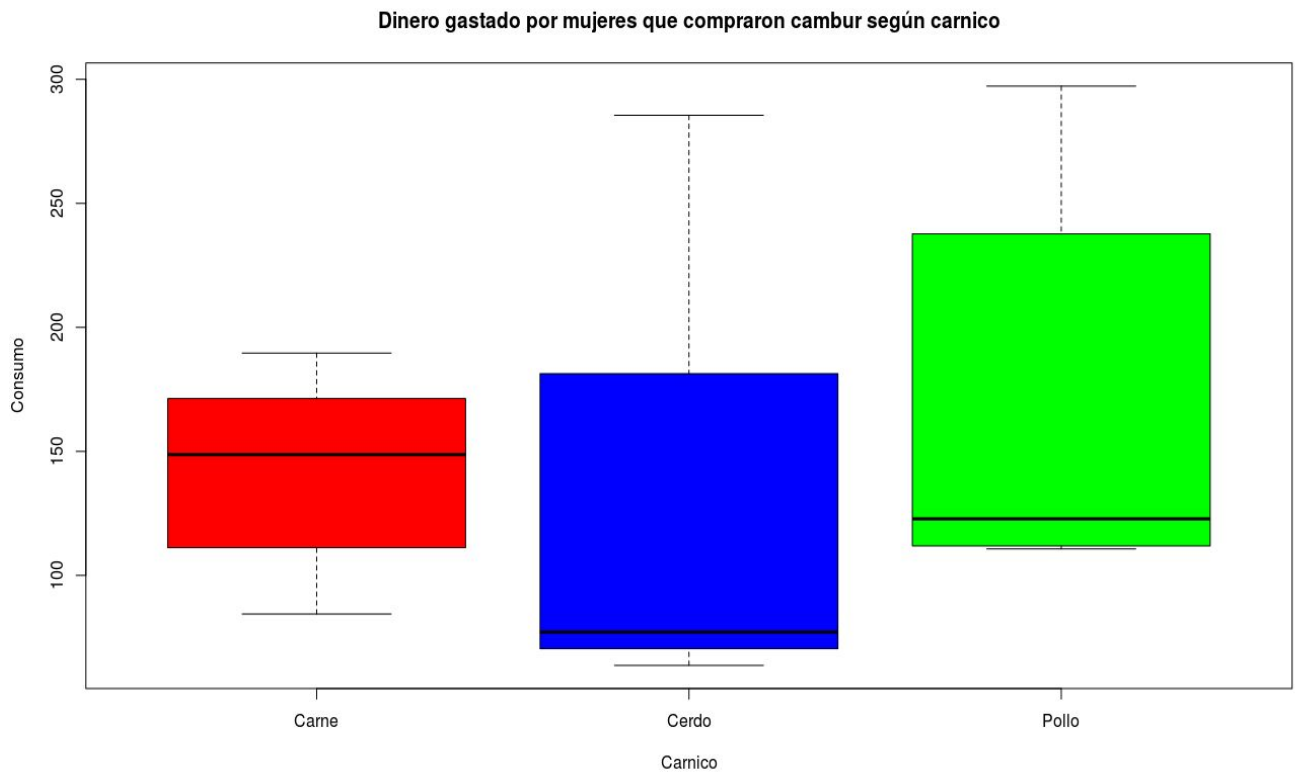


Figura 2. Gráficos de caja por tipo de cárnico en relación a su consumo (Mujeres que consumen cambur)

Carnes

De los datos con Carne como cárnico se obtuvo:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
84.43	111.19	148.68	141.05	171.30	189.63

Según los gráficos y los datos recogidos, podemos observar que la media aritmética del consumo de toda la compra en el mercado es de 141.05 en la unidad de consumo. Del boxplot podemos observar que comparado con el resto de cárnicos, los datos de la carne muestra que en su mayoría el consumo es promedio entre el de Cerdo y el de Pollo. A su vez podemos observar de aquí que los datos de consumo están más agrupados alrededor 150 y el resto de gastos están más dispersos.

Cerdo

De los datos con el Cerdo, se obtuvo

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
------	---------	--------	------	---------	------

63.72 70.45 77.18 142.13 181.34 285.50

Podemos observar directamente de lo obtenido, que en promedio se consume en el total de las compras 142.13. Aunque el promedio aritmético nos muestre que es ligeramente mayor que el consumo con Pollo, en el boxplot podemos observar directamente que el 50% de los datos están agrupados alrededor de 75 unidades de consumo y el resto de datos está disperso, lo que significa que aunque sea el 50% de dato restantes, estos presentan mayor diferencia entre valores.

Pescado

En el caso del pescado, no hubo ningún caso que cumpliera con el filtro. De esto podemos decir que no hubo ninguna compra de alguna mujer de pescado con cambur.

Pollo

De los datos obtenidos de forma analítica por R, obtuvimos que

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
110.7	113.4	122.8	167.2	210.2	297.2

Lo primero que observamos sobre los datos es que en general los consumidores de pollo tuvieron un consumo más alto en la compra final que los otros cárnicos, y se puede observar viendo que el mínimo en consumo de pollo es mayor que la mediana del consumo de carne y casi igual que la mediana del consumo de cerdo. En promedio, el pollo presentó una media aritmética de 210.2 de unidades de consumo, que apoya a lo dicho anteriormente con las medias aritméticas menores a 200 unidades. En cuestión de datos, por el boxplot los datos están concentrados alrededor de 150 con el resto de datos más dispersos y mucho más altos en general.

3. *¿Cuántos hombres mayores de 60 años compraron carne?*

Considerando los datos proporcionados, el resultado es 4.

Este cálculo se obtuvo utilizando el comando `length()` de R que indica la longitud de un vector, primero accediendo al vector columna de hombres mediante `mercado$sex`, y se filtra mediante la sintaxis de acceso de datos con índices lógicos.

```
> carnico [genero == "M" & carnico == "Carne" & edad > 60]
> cant_consumo_carne = length(consumo_carne)
[1] 4
```