

Matemáticas Computacionales

Practica 2: Estudio de una Base de Datos

1860560 Rangel Delgado Jesus Angel

07 de Marzo del 2021

1. Introducción

En esta practica estudiaremos una base de datos con estadística descriptiva en R, analizaremos todos los datos que nos proporciona la base de datos como atributos, tipos de dato, etc. Se detalla todo el procedimiento realizado para la vista de tablas, gráficas, etc.

2. Base de Datos ChickWeight

La base de datos que seleccione consiste en el estudio de los pesos corporales de pollitos, partiendo del día 0 y cada dos días, hasta el día 20 además también se tomo el peso del día 21 **Observación:** Los pollitos estaban separados en 4 grupos diferentes y cada grupo estaba sometido a una dieta diferente

2.1. Observaciones de la base de datos

En primer lugar observemos en la siguiente la tabla que representa la cantidad de cada grupo de pollitos en base a su dieta

Grupo	Cantidad
1	20
2	10
3	10
4	10

Tabla 1: Cantidad de pollitos por grupo

Entonces observe que de la tabla 1 tenemos que en total se trabajaron con 50 pollitos

Ahora el código en r nos genera una tabla con estadística mayormente acertada en los primeros datos de las tablas generadas por la base de datos en R (Los pesos de los pollos)

```
> summary(Chickweight)
  weight      Time      chick      Diet
Min.   : 35.0   Min.   : 0.00  13    : 12  1:220
1st Qu.: 63.0   1st Qu.: 4.00   9     : 12  2:120
Median :103.0   Median :10.00  20    : 12  3:120
Mean    :121.8   Mean    :10.72  10    : 12  4:118
3rd Qu.:163.8   3rd Qu.:16.00  17    : 12
Max.    :373.0   Max.    :21.00  19    : 12
              (other):506
```

Figura 1: Tablas generadas al ejecutar la instrucción summary

Observe que en el primer apartado que corresponde a la tabla de pesos el menor peso registrado es de 35, que al buscarlo en la base de datos propuesta corresponde al pollo numero 18 correspondiente al grupo 1 quien para el día 2 que se hizo la prueba perdió peso En la siguiente ilustración se muestra lo ya explicado en el párrafo anterior

	weight	Time	Chick	Diet
196	35	2	18	1
26	39	2	3	1
195	39	0	18	1
293	39	0	27	2
305	39	0	28	2
317	39	0	29	2
365	39	0	33	3
401	39	0	36	3
543	39	0	48	4
13	40	0	2	1
221	40	0	21	2
269	40	0	25	2
...

Figura 2: Base de Datos ChickWeight en R

Así como el ejemplo anterior, se pueden realizar mas comparativas con los diferentes datos.

Ahora observamos que la tabla generada con la función **summary** también nos proporciona otros datos como la Mediana de los datos capturados para los pesos la cual es: **103**

Observación ¿Que es la mediana?

La mediana : es el valor "más céntrico", una vez que los datos se ordenan según su tamaño. Si el número de observaciones es impar, la mediana es el valor de la observación en la posición

$$\frac{n+1}{2} \quad (1)$$

y si el número de observaciones es par, la mediana se define como la media (el promedio) de las observaciones en las posiciones

$$\frac{n}{2} \quad y \quad \frac{n+2}{2} \quad (2)$$

Otro dato que también podemos recabar de la tabla ya mostrada anteriormente es la media la cual es igual a **121.8**

Observación

¿Que es la media?

La media : es la suma de las observaciones dividida entre el tamaño de la muestra o de la población.

Otro dato que también nos ofrece la tabla son los cuartiles en su primer y tercer cuartil esto es igual a **63.0** para el primer cuartil y para el tercero este es igual a **163.8**

Observación

¿Que es el cuartil?

Los cuartiles: son valores que dividen una muestra de datos en cuatro partes iguales. Utilizando cuartiles puede evaluar rápidamente la dispersión y la tendencia central de un conjunto de datos, que son los pasos iniciales importantes para comprender sus datos.

- Primer cuartil: 25 % de los datos es menor que o igual a este valor
- Tercer cuartil: 75 % de los datos es menor que o igual a ese valor

Por ultimo la tabla del inicio nos genera el valor máximo de los pesos obtenidos por los pollitos el cual es: **373** el cual se encuentra ubicado con el pollito numero 35 que esta en el grupo de la dieta 3 y este peso lo presento en el ultimo dia de la prueba

	weight	Time	Chick	Diet
400	373	21	35	3
399	361	20	35	3
388	341	21	34	3
398	332	18	35	3
232	331	21	21	2
387	327	20	34	3
554	322	21	48	4
460	321	21	40	3
231	318	20	21	2
328	309	21	29	2
230	307	18	21	2
84	305	21	7	1
---	---	--	--	-

Figura 3: Base de Datos ChickWeight en R

2.2. Tablas de Frecuencia

Antes de comenzar con las gráficas observe que el programa de R también ofrece mas funciones para estadística como la tabla de frecuencias Para ejemplificar este apartado decidí obtener la tabla de frecuencias para ver cuanto es el por ciento de pollitos en cada grupo con su dieta.

Grupo	Porcentaje
1	38.06 %
2	20.76 %
3	20.76 %
4	20.41 %

Tabla 2: Cantidad de pollitos por grupo en porcentaje

```
10 y <-Chickweight$Diet
11 cbind(freq=table(y), percentage=prop.table(table(y))*100)|
```

Figura 4: Código en R para generar la tabla de frecuencias

2.3. Calculo de Correlaciones

Otra de las funciones que nos ofrece el programa R es poder realizar el calculo de correlaciones Para ejemplificar este punto realizare la correlación entre el peso de los pollitos y los días que duro el experimento

Peso	Tiempo
1.0000000	0.8371017
0.8371017	1.0000000

Tabla 3: Correlación entre el peso y tiempo

```

14 #correlacion
15 correlations <- cor(Chickweight[,1:2])
16 print(correlations)
17

```

Figura 5: Código en R para generar la Correlación

Observación

¿Que es la correlación?

La Correlación: indica la fuerza y la dirección de una relación lineal y proporcionalidad entre dos variables estadísticas. Se considera que dos variables cuantitativas están correlacionadas cuando los valores de una de ellas varían sistemáticamente con respecto a los valores homónimos de la otra: si tenemos dos variables (A y B) existe correlación entre ellas si al disminuir los valores de A lo hacen también los de B y viceversa. La correlación entre dos variables no implica, por sí misma, ninguna relación de causalidad

3. Gráficas de la base de datos

Otra de las ventajas de trabajar bases de datos en R es que podemos realizar graficas con todos los datos que estemos trabajando, a continuación unos ejemplos:

3.1. Histogramas

Observación

¿Qué es un histograma?

El histograma: es un gráfico de rectángulos que tiene su base en el eje x, de igual ancho cuando representa el comportamiento de una variable discreta, o de anchura proporcional a la longitud del intervalo cuando se desea representar una variable continua. En este último caso el punto central de la base de los rectángulos equivale al punto medio de cada clase

Ejemplo de un Histograma:

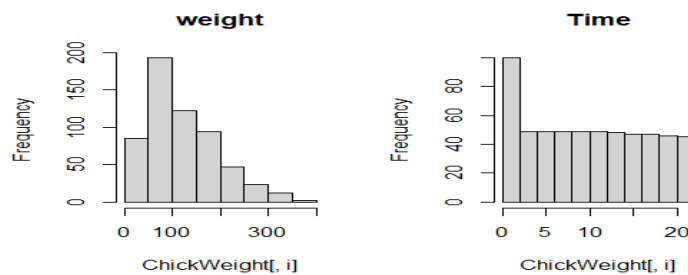


Figura 6: Histograma en base a los pesos y el tiempo

```

17
18 #Histograma
19 data("chickweight")
20 par(mfrow=c(1,2))
21 for(i in 1:2) {
22   hist(chickweight[,i], main=names(chickweight)[i])
23 }

```

Figura 7: Código para generar el histograma

3.2. Gráfica de correlaciones

Ejemplo de una Gráfica de Correlación:

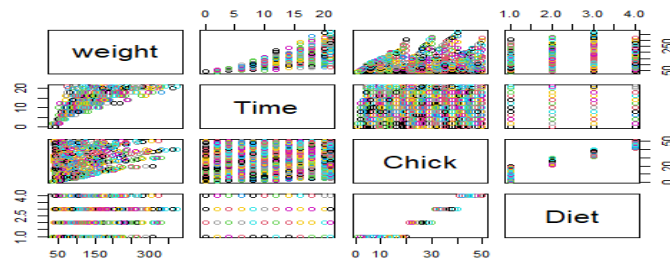


Figura 8: Gráfica de Correlación en base a los pesos y el tiempo

```

25 #grafica de correlaciones
26 library(corrplot)
27 correlations <- cor(chickweight[,1:2])
28 corrplot(correlations, method="circle")
29 pairs(chickweight)
30 pairs(weight~., data=chickweight, col=chickweight$weight)

```

Figura 9: Código para generarlas gráficas de correlaciones

3.3. Gráficos de densidad por clase

Otro de los gráficos que nos proporciona el programa de R es el de crear gráficos de densidad por clase para ejemplificar esto decidí hacer un gráfico del tiempo y peso con respecto a la dieta

Observación

¿Qué es un Gráfico de densidad?

El gráfico de densidad: visualiza la distribución de datos en un intervalo o período de tiempo continuo. Este gráfico es una variación de un Histograma que usa el suavizado de kernel para trazar valores, permitiendo distribuciones más suaves al suavizar el ruido. Los picos de un gráfico de densidad ayudan a mostrar dónde los valores se concentran en el intervalo.

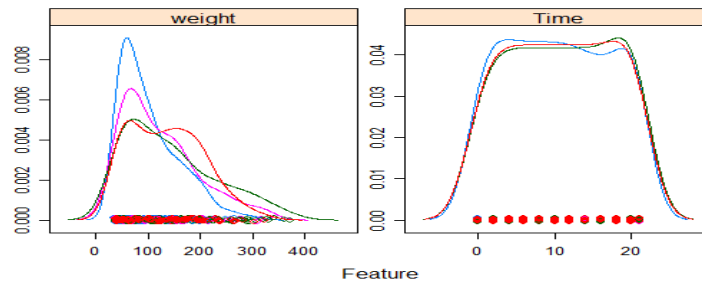


Figura 10: Gráfica de densidad por clase

```
33 #Gráficos de densidad por clase
34 library(caret)
35 x <- chickweight[,1:2]
36 y <- chickweight[,4]
37 scales <- list(x=list(relation="free"), y=list(relation="free"))
38 featurePlot(x=x, y=y, plot="density", scales=scales)
```

Figura 11: Código para generarlas gráficas de densidad por clase

3.4. Diagramas de caja

Por ultimo veremos un ejemplo del diagrama de caja otro gráfico que nos facilita el estudio de la base de datos

Observación

¿Qué es un diagrama de caja?

El diagrama de caja: es un gráfico utilizado para representar una variable cuantitativa (variable numérica). El gráfico es una herramienta que permite visualizar, a través de los cuartiles, cómo es la distribución, su grado de asimetría, los valores extremos, la posición de la mediana, etc.

Para esté apartado realice dos ejemplos donde observamos el diagrama de caja en base al peso y la dieta asignada y en el segundo caso observemos que el gráfico esta diseñado con respecto al peso y el pollo.

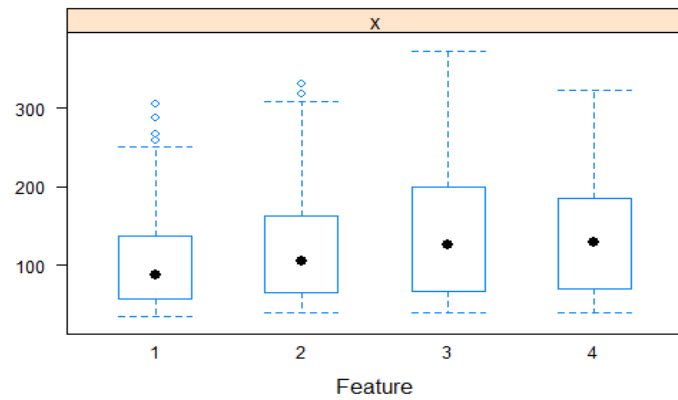


Figura 12: Diagrama de clase 1

```

41 #Diagramas de caja #1
42 x <- chickweight[,1]
43 y <- chickweight[,4]
44 featurePlot(x=x, y=y, plot="box")
45

```

Figura 13: Código para el diagrama de clase 1

```

46 #Diagrama de caja #2
47 x <- chickweight[,1]
48 y <- chickweight[,3]
49 featurePlot(x=x, y=y, plot="box")
50

```

Figura 14: Código para el diagrama de clase 2

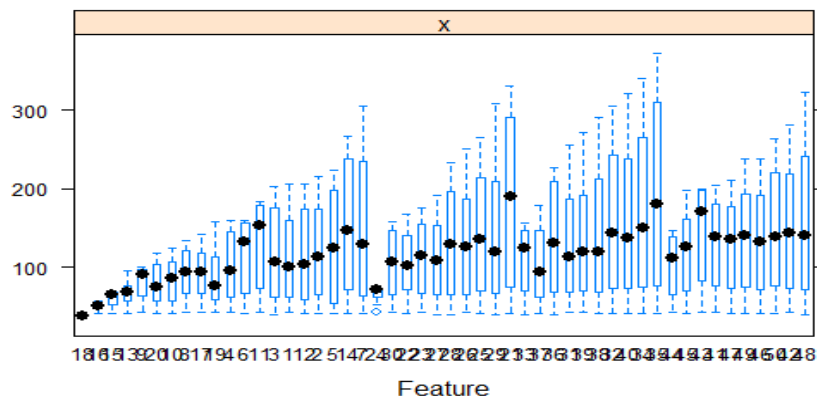


Figura 15: Diagrama de clase 2

4. Conclusión

Como conclusiones de la base de datos observe que los pollitos que registraron el mayor peso se encuentran en el grupo que consumió la dieta 3 y los pollitos con el menor peso la mayoría pertenecen a los que consumieron la dieta 1, además también podemos observar como algunos pollitos desde el inicio del consumo de la dieta 1 perdieron peso, lo cual se ejemplifico anteriormente en el documento Otra cosa que también podemos concluir es que el pollo con el mayor peso es el registrado con el número 35 (pertenece a la dieta 3) y el de menor peso al finalizar el experimento es el registrado con el número 24 (pertenece a la dieta 2). Otra cosa que también puede ser notoria es que un pollito del grupo de la dieta 4 no logro terminar el experimento ya que después del día 18 no se toma registro alguno de esto este pollito estaba nombrado como 44. Esto es notorio en base la cantidad de datos al finalizar la prueba. Como conclusión de la practica fue muy interesante observar las ventajas que te ofrece el programa de R para trabajar la estadística, es una muy buena base, para trabajar numerosas cantidades de datos ya que existen muchas funciones que te ayudan a resumir datos que trabajando a mano seria una tarea muy tediosa.

Referencias

- [1] Jhon E. Freund, Estadística matemática con aplicaciones, 2000
- [2] Datasets <https://www.rdocumentation.org/packages/datasets/versions/3.6.2>Datasets
- [3] Jesus Rangel, repositorio de GitHub<https://github.com/JesusRangel07/MatematicasComputacionales>Repositorio de Github