

Soluciones laboratorio semana 2 Matemáticas II

Índice de contenido

Soluciones laboratorio semana 2 Matemáticas II.....	1
Introducción a las soluciones.....	1
Problema 1.5.....	2
Solución 1.5:.....	2
Problema 1.6.....	6
Solución 1.6:.....	6
Problema 1.7.....	8
Solución 1.7:.....	8

Introducción a las soluciones

En este laboratorio consiste resolver , en clase de laboratorio y de forma manual, los problemas 1.5, 1.6 y 1.7.

Se dieron unas indicaciones adicionales para fijar la resolución de los problemas. De forma que quedaron enunciados como se expone en estas soluciones.

Este documento está realizado utilizando la tecnología OpenOffice+R+OdfWeave que se expuso en el laboratorio de la semana 1.

Podéis ver los ficheros entrada y salida de este documento [aquí](#).

En las soluciones dejaremos a la vista el código R necesario. De todas formas vuestra solución debe ser el papel.

Antes de empezar haremos la carga de funciones y librerías, si cabe, usadas. En este caso hemos definido función que dibuja el histograma de frecuencias absolutas con su correspondiente polígono de frecuencias.

```
<<echo=TRUE>>
```

```
histograma.absolutas <- function(Lj,Xj,nj){
#numero de datos
n <- sum(nj)
#numero de intervalos
k <- length(Lj)-1
#Amplitudes
Aj <- Lj[2:(k+1)]-Lj[1:k]
#Alturas
hj <- nj/Aj
#calculo de coordenadas de los puntos y marcas
aux <- rep(3,times=(k-1))
x <- rep(Lj, times=c(2,aux,2))
```

```
#x <- c(L[1],L[1],rep(Li[2:k],times=c(rep(3,times=k-
2)),L[k+1],L[k+1]))
h<-c(0,hj[1],hj[1],0)
for(i in c(2:k)){h<- c(h,hj[i],hj[i],0)}
#Dibujo de tipo escalera
plot(x,h,type="S",
main="Histograma y polígono de \n frecuencias absolutas",
ylim=c(0,max(hj)+0.1*Aj[k]),xlim=c(min(Lj)-0.1*Aj[1],max(Lj)
+0.1*min(Lj)),xlab="Intervalos y macas de clase",
sub="El área de cada rectángulo es la frecuencia absoluta",
ylab="Altura",xaxs="i",yaxs="i",bty="l",xaxt="n",yaxt="n")
for(i in c(1:k)){text(Xj[i],hj[i]/2,nj[i])}
#poligono de frecuencias absolutas
x <- c(Lj[1],Xj,Lj[k+1])
y <- c(0,hj,0)
lines(x,y,col="red")
axis(1,at=Lj, labels=Lj)
axis(1,at=Xj, labels=Xj)
axis(2,at=hj, labels=round(hj,2))
}
```

@

Problema 1.5

1.5.- A 50 aspirantes a un determinado puesto de trabajo se les sometió a una prueba. Las puntuaciones obtenidas fueron:

4	4	2	10	1	9	5	3	4	5
6	6	7	6	8	7	6	8	7	6
5	4	4	4	5	6	6	7	5	6
6	7	5	6	6	7	5	6	4	3
2	6	6	7	7	8	8	9	8	7

- Agrupando con la metodología expuesta en clase y con la regla de *Sturges*, construir la tabla de frecuencias absolutas, absolutas acumuladas, relativas y relativas acumuladas. Hacer la representación gráfica para el histograma de frecuencias absolutas.
- Encontrar la puntuación que seleccione al 20% de los mejores candidatos.

Solución 1.5:

- La precisión de los datos es 1.

Calculemos con R los elemento básicos, maximo, mínimo..... , el valor del número de clases k.

```

> precision <- 1
> x <- c(4, 4, 2, 10, 1, 9, 5, 3, 4, 5, 6, 6, 7, 6, 8, 7, 6, 8,
+       7, 6, 5, 4, 4, 4, 5, 6, 6, 7, 5, 6, 6, 7, 5, 6, 6, 7, 5,
+       6, 4, 3, 2, 6, 6, 7, 7, 8, 8, 9, 8, 7)
> n <- length(x)
> m <- min(x)
> M <- max(x)
> k1 <- 1 + log(n, 2)
> k <- ceiling(1 + log(n, 2))
> rango <- M - m
> candidatoA <- rango/k
> A <- 2
> L1 <- m - precision/2
> aux <- A * c(0:k)
> L <- L1 + aux

```

Ahora vamos a seguir los pasos de la página 25 de las notas de clase de este tema.

- La observación mayor es $M=10$ la más pequeña es $m=1$.
- El número de datos es $n=50$.
- El rango es $\text{rango}=M-m = 9$.
- $1 + \log_2(n) = 1 + \log_2(50) = \log_2(2) = 6.64385618977472$.
- Ahora lo redondeamos al entero superior $k = 7$.
- Elegiremos amplitud constante para todos los intervalos.
- Así el candidato a amplitud es $\text{rango}/k = 1.28571428571429$ y lo redondeamos por exceso la unidad de la precisión por lo tanto $A=2$.
- Ahora tenemos que calcular los extremos de los intervalos de clase.
- El extremo inferior es $L_1 = m - \frac{\text{precision}}{2} = 0.5$.

Por lo tanto los extremos de los intervalos son:

```

> odftable(L, colnames = paste("L", seq(1, k + 1, 1), sep = ""),
+         horizontal = T)

```

L1	L2	L3	L4	L5	L6	L7	L8
0.5	2.5	4.5	6.5	8.5	10.5	12.5	14.5

Como los dos últimos intervalos superan el máximo, tendrán frecuencia 0. Así que los eliminamos y $k=5$ intervalos.

Calculemos las frecuencias:

```

> k <- 5
> L1 <- m - precision/2
> aux <- A * c(0:k)
> Lj <- L1 + aux
> conteo <- table(cut(x, breaks = Lj))
> nj <- as.numeric(conteo)
> Nj <- cumsum(nj)
> fj <- nj/n

```

```

> Fj <- cumsum(fj)
> Xj <- (Lj[2:(k + 1)] + Lj[1:k])/2
> tablafrecuencias <- as.data.frame(cbind(Lj[1:k], Lj[2:(k + 1)],
+   Xj, nj, Nj, fj, Fj))
> names(tablafrecuencias) <- c("Lj", "Lj+1", "Xj", "nj", "Nj",
+   "fj", "Fj")
> odfTable(tablafrecuencias, colnames = names(tablafrecuencias),
+   useRowNames = F, digits = 2)

```

Lj	Lj+1	Xj	nj	Nj	fj	Fj
0.5	2.5	1.5	3	3	0.06	0.06
2.5	4.5	3.5	9	12	0.18	0.24
4.5	6.5	5.5	21	33	0.42	0.66
6.5	8.5	7.5	14	47	0.28	0.94
8.5	10.5	9.5	3	50	0.06	1.00

Ahora nos piden el histograma de frecuencias absolutas acumuladas. Para ello tenemos que calcular las altura de cada barra que son $h_j = \frac{n_j}{A}$.

Las alturas son

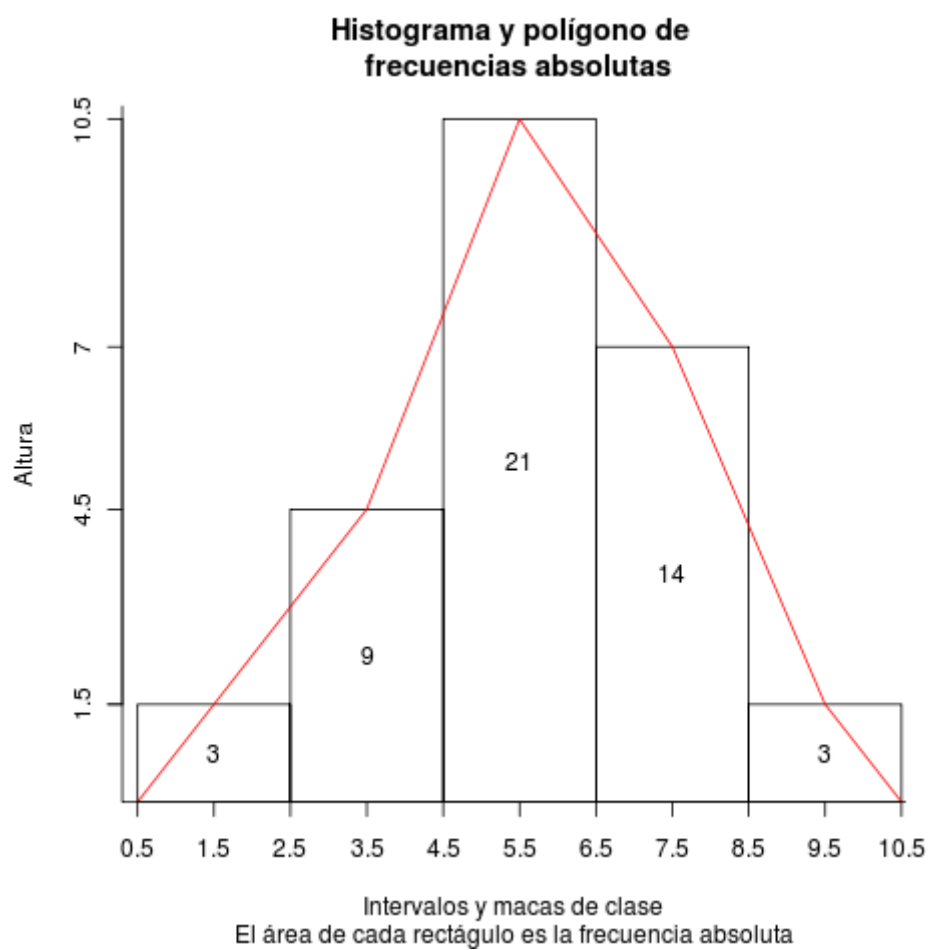
```

> hj <- nj/A
> odfTable(hj, colnames = "hj", digits = 2)

```

hj
1.5
4.5
10.5
7.0
1.5

El histograma es, (en este caso no os pongo el código, vosotros lo dibujáis manualmente):



b) La puntuación que selecciona al 20% de los mejores candidatos es el cuantil 0.8, l cuantil $Q_{0.8}$. El intervalo crítico es el primero en el que la frecuencia absoluta acumulada supera o iguala $0.8 \cdot n = 0.8 \cdot 50 = 40$. El intervalo crítico es $c=4$. Necesitamos:

1. El extremo inferior del intervalo crítico $L_c = L_4 = 6.5$
2. La frecuencia absoluta acumulada del intervalo anterior al crítico $N_{c-1} = N_3 = 33$
3. La frecuencia absoluta del intervalo crítico $n_c = n_4 = 14$
4. La amplitud del intervalo crítico es $A_c = A_4 = 2$
5. El valor $0.8 \cdot n = 0.8 \cdot 50 = 40$.

Entonces el cuantil pedido es $Q_{0.8} = 6.5 + 2 \frac{40-33}{14} = 7.5$.

Problema 1.6

1.6.- En la población de estudiantes de la facultad, se seleccionó una muestra de 20 alumnos y se obtuvieron las siguientes tallas en cm.:

162, 168, 174, 168, 166, 170, 168, 166, 170, 172, 188, 182, 178, 180, 176, 168, 164, 166, 164, 172.

Se pide:

- Descripción numérica (tabla de frecuencias absolutas y relativas y representación gráfica del diagrama de barras de las frecuencias absolutas)
- Media aritmética, mediana y moda.

Solución 1.6:

a) Calculemos las frecuencias sin agrupar:

```
> x <- c(162, 168, 174, 168, 166, 170, 168, 166, 170, 172, 188,
+       182, 178, 180, 176, 168, 164, 166, 164, 172)
> tabla.frecuencias <- table(x)
> Xj <- as.numeric(names(tabla.frecuencias))
> nj <- as.numeric(tabla.frecuencias)
> n <- sum(nj)
> Nj <- cumsum(nj)
> fj <- nj/n
> Fj <- cumsum(fj)
> tabla.frecuencias <- as.data.frame(cbind(Xj, nj, Nj, fj, Fj))
> odftable(tabla.frecuencias, colnames = names(tabla.frecuencias),
+          useRowNames = F, digits = 2)
```

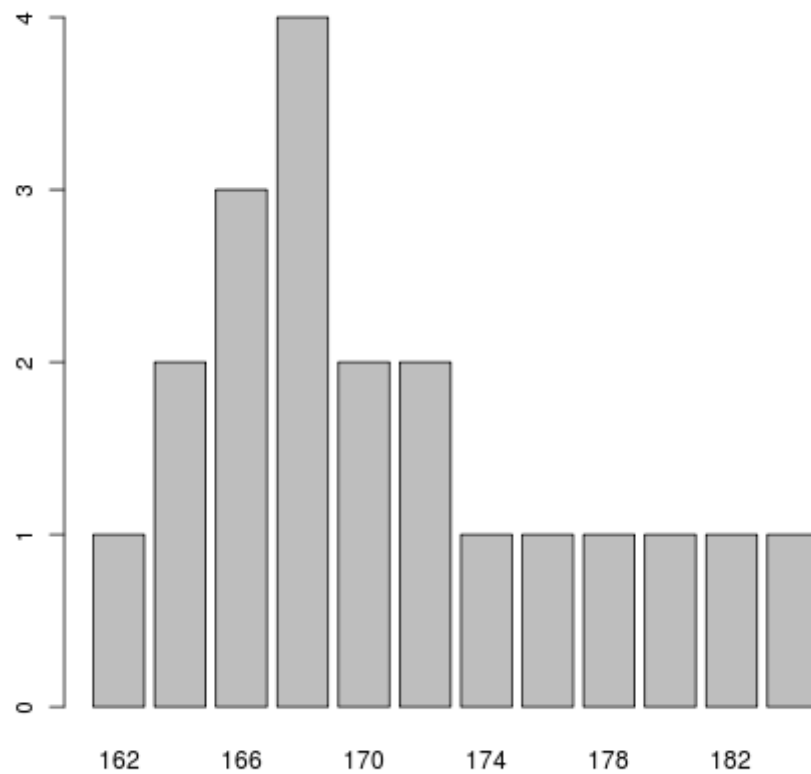
Xj	nj	Nj	fj	Fj
162	1	1	0.05	0.05
164	2	3	0.10	0.15
166	3	6	0.15	0.30
168	4	10	0.20	0.50
170	2	12	0.10	0.60
172	2	14	0.10	0.70
174	1	15	0.05	0.75
176	1	16	0.05	0.80
178	1	17	0.05	0.85
180	1	18	0.05	0.90
182	1	19	0.05	0.95

Xj	nj	Nj	fj	Fj
188	1	20	0.05	1.00

b) La media es 171.1, no se pide pero la varianza es 43.4987, y la mediana es 169, la moda es 168 pues su frecuencia es la máxima 4.

¡Ah! faltaba el diagrama de barras de las frecuencias absolutas. Las alturas son las frecuencias absolutas, fijaros en los "agujeros"

```
> barplot(table(x))
```



Problema 1.7

Agrupando los datos del ejercicio anterior en intervalos de amplitud 10 cm., se pide:

- a) Descripción numérica y representación gráfica del histograma de las frecuencias absolutas
- b) Media aritmética, mediana y moda.
- c) Analizar los cálculos hechos y los errores de agregación comparándolos con el ejercicio anterior.

Solución 1.7:

```
> precision = 1
> m <- min(x)
> M <- max(x)
> L1 <- m - precision/2
> k <- 3
> A <- 10
> aux <- A * c(0:k)
> Lj <- L1 + aux
> conteo <- table(cut(x, breaks = Lj))
> nj <- as.numeric(conteo)
> Nj <- cumsum(nj)
> fj <- nj/n
> Fj <- cumsum(fj)
> Xj <- (Lj[2:(k + 1)] + Lj[1:k])/2
> tablafrecuencias <- as.data.frame(cbind(Lj[1:k], Lj[2:(k + 1)],
+   Xj, nj, Nj, fj, Fj))
> names(tablafrecuencias) <- c("Lj", "Lj+1", "Xj", "nj", "Nj",
+   "fj", "Fj")
> odfTable(tablafrecuencias, colnames = names(tablafrecuencias),
+   useRowNames = F, digits = 2)
```

Lj	Lj+1	Xj	nj	Nj	fj	Fj
162	172	166	12	12	0.6	0.6
172	182	176	6	18	0.3	0.9
182	192	186	2	20	0.1	1.0

Ahora calculamos la media, la varianza (que no se pedía):

```
> media <- sum(nj * Xj)/n
> varianza <- sum(nj^2)/n - media^2
```

Por lo tanto la media es 171.5 y la varianza es -29403.05.

La mediana es el cuantil $Q_{0.5}$. El intervalo crítico es el primero en el que la frecuencia absoluta acumulada supera o iguala $0.5 \cdot n = 0.5 \cdot 20 = 10$. El intervalo crítico es $c=1$. Necesitamos:

1. El extremo inferior del intervalo crítico $L_c = L_1 = 161.5$
2. La frecuencia absoluta acumulada del intervalo anterior al crítico $N_{c-1} = 0$
3. La frecuencia absoluta del intervalo crítico $n_c = n_1 = 12$
4. La amplitud del intervalo crítico es $A_c = A_1 = 10$
5. El valor $0.5 \cdot n = 0.5 \cdot 20 = 10$.

Entonces la mediana es $Q_{0.5} = 161.5 + 10 \frac{10-0}{12} = 169.8333$.

Por último el intervalo modal es el de mayor frecuencia que es también el primero con frecuencia 12.

c) Evidentemente el resultado de la media, varianza, mediana y moda varía de tratar los datos de forma agrupada a sin agrupar. Cada uno habrá interpretado a su manera estos cambios. Lo que sí es evidente es que el histograma de frecuencia agrupadas es más correcto y aporta más significado a la distribución de los datos que el diagrama de barras.

El histograma de frecuencias absolutas es :

```
> histograma.absolutas(Lj, Xj, nj)
```

