



5. Los salarios semanales de 65 empleados de la empresa X vienen expresados en la siguiente distribución de frecuencias. Se pide:

- El límite inferior de la 6ª clase.
- El límite superior de la 4ª clase.
- La marca de la 3ª clase.
- Las fronteras de la 5ª clase.
- Las anchuras del 5º intervalo.
- La frecuencia absoluta de la 3ª clase.
- La frecuencia relativa de la 3ª clase
- El intervalo de clase con máxima frecuencia (intervalo de clase modal).
- El porcentaje de empleados que cobran menos de 280,00 € a la semana.
- El porcentaje de empleados que cobran menos de 300,00 € pero al menos 260,00 € a la semana.

Salario semanal (€)	Nº de empleados
250,00 – 259,99	8
260,00 – 269,99	10
270,00 – 279,99	16
280,00 – 289,99	14
290,00 – 299,99	10
300,00 – 309,99	5
310,00 – 319,99	2

Solución:

Aquí los datos están expresados con una precisión de 1 céntimo (0,01 €). Por lo tanto los límites de los intervalos se expresan en céntimos de euro, mientras que las fronteras de los intervalos se expresan en milésimas de euro:

```
> frecAbs <- c(8,10,16,14,10,5,2)
> frecAbsAcum <- cumsum(frecAbs)
> frecAbsAcum
[1] 8 18 34 48 58 63 65
> frecRel <- frecAbs/sum(frecAbs)
> frecRel
[1] 0.12307692 0.15384615 0.24615385 0.21538462
[5] 0.15384615 0.07692308 0.03076923
> frecRelAcum <- cumsum(frecRel)
> frecRelAcum
```

```

[1] 0.1230769 0.2769231 0.5230769 0.7384615 0.8923077
[6] 0.9692308 1.0000000
> frecRelAcum[3]
[1] 0.5230769
> frecRelAcum[5]-frecRelAcum[1]
[1] 0.7692308

```

Clase	Intervalo	Front. Inf.	Front. Sup.	Límite Inf.	Límite Sup.	Marca	Frec. Abs.	Frec. Abs.Ac.
1	250,00 – 259,99	249,995	259,995	250,00	259,99	254,995	8	8
2	260,00 – 269,99	259,995	269,995	260,00	269,99	264,995	10	18
3	270,00 – 279,99	269,995	279,995	270,00	279,99	274,995	16	34
4	280,00 – 289,99	279,995	289,995	280,00	289,99	284,995	14	48
5	290,00 – 299,99	289,995	299,995	290,00	299,99	294,995	10	58
6	300,00 – 309,99	299,995	309,995	300,00	309,99	304,995	5	63
7	310,00 – 319,99	309,995	319,995	310,00	319,99	314,995	2	65

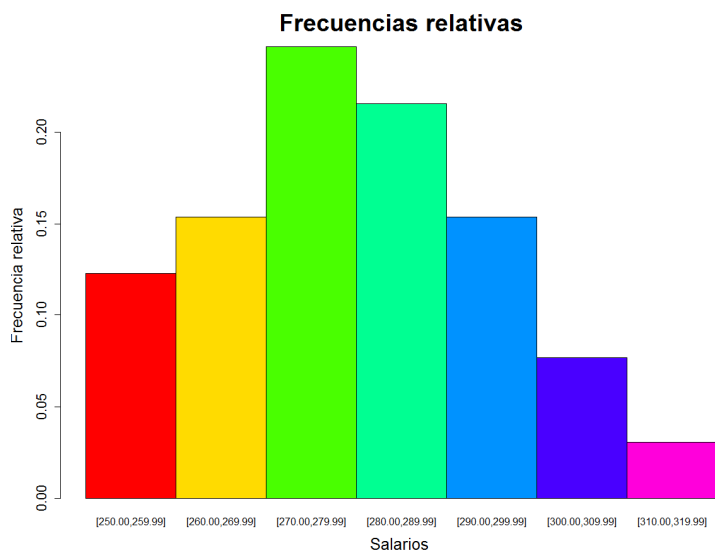
- a) El límite inferior de la 6ª clase → 300,00
- b) El límite superior de la 4ª clase → 289,99
- c) La marca de la 3ª clase → 274,995
- d) Las fronteras de la 5ª clase → Inferior: 289,995, superior: 299,995
- e) Las anchuras del 5º intervalo → $299,995 - 289,995 = 10$
- f) La frecuencia absoluta de la 3ª clase → 16
- g) La frecuencia relativa de la 3ª clase → $16/65 = 0,246$, es decir, 24.6%
- h) El intervalo de clase con máxima frecuencia (intervalo de clase modal) → [270,00, 279,99]
- i) El porcentaje de empleados que cobran menos de 280,00 € a la semana → Frecuencia absoluta acumulada de la 3ª clase = $8 + 10 + 16 = 34$. Frecuencia relativa = $34/65 = 0,523$, es decir, 52,3%
- j) El porcentaje de empleados que cobran menos de 300,00 € pero al menos 260,00 € a la semana → Frecuencia absoluta acumulada de la 5ª clase – Frecuencia absoluta acumulada de la 1ª clase = $10 + 16 + 14 + 10 = 50$. Frecuencia relativa = $0,769$, es decir, 76,9%.

6. Con los datos del ejercicio anterior, se pide:
- Calcular la tabla de frecuencias relativas.
 - Dibujar el histograma de frecuencias relativas.
 - Calcular las frecuencias relativas acumuladas.
 - Dibujar el polígono de frecuencias relativas acumuladas.

Solución:

Versión con los datos agrupados: En este caso el histograma debemos dibujarlo con la función `barplot`, indicándole expresamente que las barras deben ser contiguas sin espacios intermedios porque la variable es continua.

```
> intervalos <- c("[250.00,259.99]", "[260.00,269.99]",
"[270.00,279.99]", "[280.00,289.99]", "[290.00,299.99]",
"[300.00,309.99]", "[310.00,319.99]")
> names(frecAbs) <- intervalos
> frecAbs
[250.00,259.99] [260.00,269.99] [270.00,279.99]
           8           10           16
[280.00,289.99] [290.00,299.99] [300.00,309.99]
           14           10           5
[310.00,319.99]
           2
> frecRel <- frecAbs/sum(frecAbs)
> frecRel
[250.00,259.99] [260.00,269.99] [270.00,279.99]
           0.12307692           0.15384615           0.24615385
[280.00,289.99] [290.00,299.99] [300.00,309.99]
           0.21538462           0.15384615           0.07692308
[310.00,319.99]
           0.03076923
> barplot(frecRel,main="Frecuencias relativas", xlab="Salarios",
ylab="Frecuencia relativa",col=rainbow(7),cex.lab=1.6,cex.axis=1.4,
cex.main=2.4,space=FALSE)
```



```
> frecRelAcum <- cumsum(frecRel)
> frecRelAcum
[250.00,259.99] [260.00,269.99] [270.00,279.99]
```

```

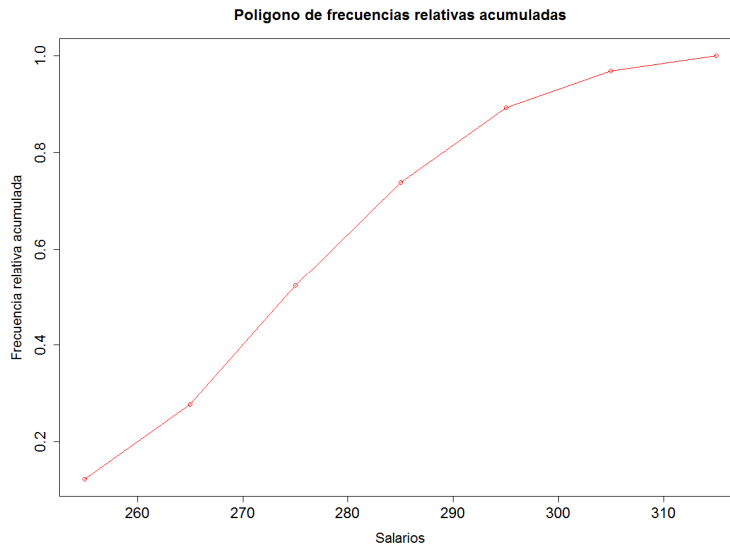
0.1230769      0.2769231      0.5230769
[280.00,289.99] [290.00,299.99] [300.00,309.99]
0.7384615      0.8923077      0.9692308
[310.00,319.99]
1.0000000

```

```

> marcas<-seq(254.995,314.995,10)
> plot(marcas,frecRelAcum,type="o",xlab="Salarios",
ylab="Frecuencia relativa acumulada",
main="Poligono de frecuencias relativas acumuladas",col="red",
cex.axis=1.5,cex.main=1.5,cex.lab=1.4)

```

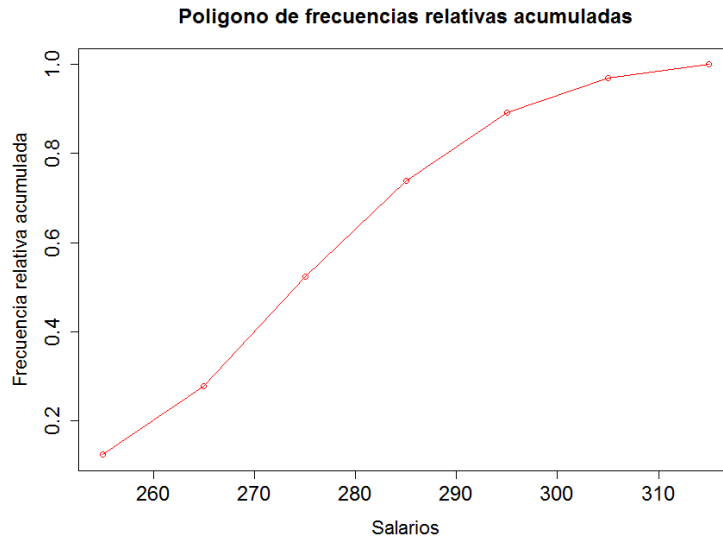


Versión con los datos sin agrupar: Podemos reconstruir los salarios originales de los trabajadores, suponiendo que todos los salarios correspondientes a un determinado intervalo coinciden con su marca de clase.

```

> salarios <- rep(marcas,frecAbs) # rep = repetir
> salarios
 [1] 254.995 254.995 254.995 254.995 254.995 254.995
 [7] 254.995 254.995 264.995 264.995 264.995 264.995
[13] 264.995 264.995 264.995 264.995 264.995 264.995
[19] 274.995 274.995 274.995 274.995 274.995 274.995
[25] 274.995 274.995 274.995 274.995 274.995 274.995
[31] 274.995 274.995 274.995 274.995 284.995 284.995
[37] 284.995 284.995 284.995 284.995 284.995 284.995
[43] 284.995 284.995 284.995 284.995 284.995 284.995
[49] 294.995 294.995 294.995 294.995 294.995 294.995
[55] 294.995 294.995 294.995 294.995 304.995 304.995
[61] 304.995 304.995 304.995 314.995 314.995
> h <- hist(salarios,col=rainbow(7),main="Frecuencias absolutas",
xlab="Salarios",ylab="Frecuencia absoluta",cex.lab=1.6,cex.axis=1.4,
cex.main=2.4)
> plot(h$mids,cumsum(h$counts)/65,type="o",xlab="Salarios",
ylab="Frecuencia relativa acumulada",
main="Poligono de frecuencias relativas acumuladas",col="red",
cex.axis=1.5,cex.main=1.5,cex.lab=1.4)

```



OJO: Las marcas o puntos medios calculados por la función `hist` presuponen que tenemos una precisión infinita (en vez de 0,01 €). Es por ello que difieren ligeramente de las marcas de clase que hemos utilizado previamente. A efectos prácticos, y salvo que necesitemos una perfección total, ambas gráficas son iguales.

```
> h$mids
[1] 255 265 275 285 295 305 315
```

7. Tras la contratación de 5 nuevos empleados, repetir el histograma de salarios, donde ahora la tabla queda modificada con intervalos de amplitud no constante.

Salario semanal (€)	Nº de empleados
250,00 – 259,99	8
260,00 – 269,99	10
270,00 – 279,99	16
280,00 – 289,99	15
290,00 – 299,99	10
300,00 – 319,99	8
320,00 – 379,99	3

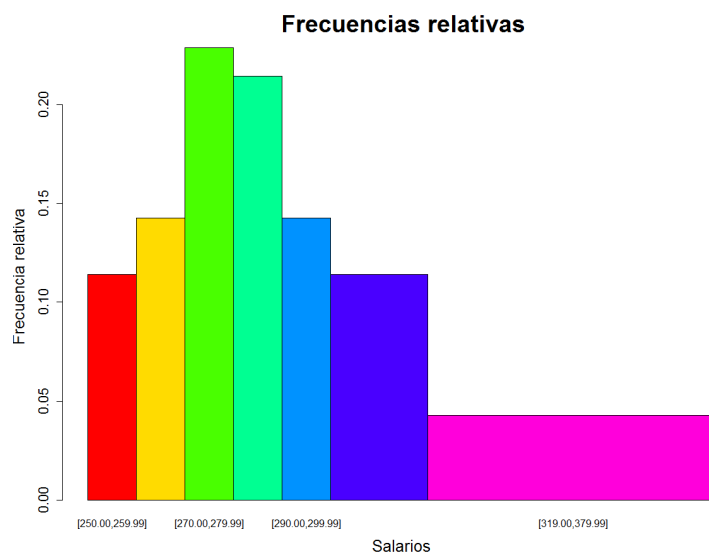
Solución:

```
> marcas<-c(seq(254.995,294.995,10),309.995,349.995)
> marcas
[1] 254.995 264.995 274.995 284.995 294.995 309.995
[7] 349.995
> intervalos <-
c("[250.00,259.99]", "[260.00,269.99]", "[270.00,279.99]",
"[280.00,289.99]", "[290.00,299.99]", "[300.00,319.99]",
"[319.00,379.99]")
> names(frecAbs) <- intervalos
> frecAbs
```

```

[250.00,259.99] [260.00,269.99] [270.00,279.99]
      8              10             16
[280.00,289.99] [290.00,299.99] [300.00,319.99]
      15              10             8
[319.00,379.99]
      3
> frecRel <- frecAbs/sum(frecAbs)
> frecRel
[250.00,259.99] [260.00,269.99] [270.00,279.99]
      0.11428571      0.14285714      0.22857143
[280.00,289.99] [290.00,299.99] [300.00,319.99]
      0.21428571      0.14285714      0.11428571
[319.00,379.99]
      0.04285714
> anchuras <- c(rep(10,5),20,60)
> barplot(frecRel,main="Frecuencias relativas",xlab="Salarios",
ylab="Frecuencia relativa",col=rainbow(7),cex.lab=1.6,cex.axis=1.4,
cex.main=2.4,space=FALSE,width=anchuras)

```



A veces interesa que el área total de histograma sea 1, con lo que en vez de frecuencias relativas se representan densidades. Esto se logra con la función hist con el parámetro freq=FALSE.

```

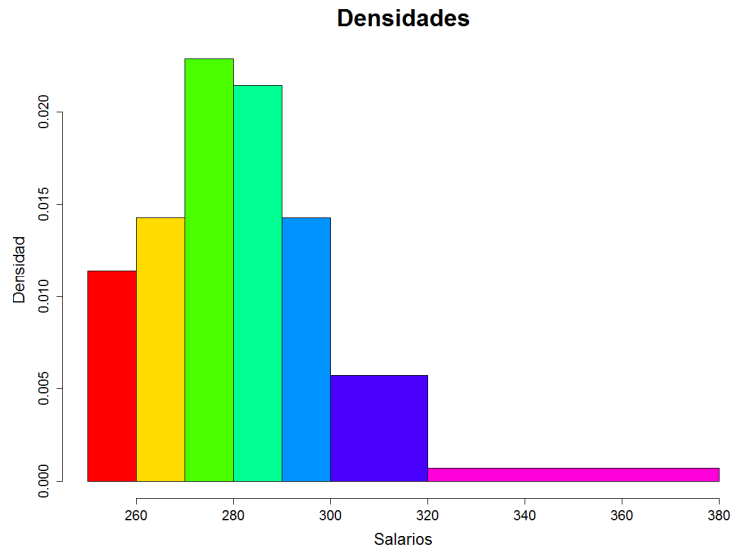
> marcas
[1] 254.995 264.995 274.995 284.995 294.995 309.995
[7] 349.995
> frecAbs
[250.00,259.99] [260.00,269.99] [270.00,279.99]
      8              10             16
[280.00,289.99] [290.00,299.99] [300.00,319.99]
      15              10             8
[319.00,379.99]
      3
> salarios <- rep(marcas,frecAbs)
> salarios
[1] 254.995 254.995 254.995 254.995 254.995 254.995
[7] 254.995 254.995 264.995 264.995 264.995 264.995
[13] 264.995 264.995 264.995 264.995 264.995 264.995
[19] 274.995 274.995 274.995 274.995 274.995 274.995
[25] 274.995 274.995 274.995 274.995 274.995 274.995
[31] 274.995 274.995 274.995 274.995 284.995 284.995
[37] 284.995 284.995 284.995 284.995 284.995 284.995

```

```

[43] 284.995 284.995 284.995 284.995 284.995 284.995
[49] 284.995 294.995 294.995 294.995 294.995 294.995
[55] 294.995 294.995 294.995 294.995 294.995 294.995
[61] 309.995 309.995 309.995 309.995 309.995 309.995
[67] 309.995 349.995 349.995 349.995
> h <- hist(salarios,breaks=c(seq(250,300,10),320,380),freq=FALSE,
col=rainbow(7),main="Densidades",xlab="Salarios",ylab="Densidad",
cex.lab=1.6,cex.axis=1.4,cex.main=2.4)

```



Nótese cómo la altura de la barra del intervalo [300.00, 319.99] se ha dividido por 2, y la del intervalo [320.00, 379.99] se ha reducido en un factor 6, respecto de la gráfica anterior.

14. En una fábrica de tuercas se toma una muestra de 40 unidades para un estudio de calidad. Los pesos de las tuercas (con precisión de 0,1 g) vienen indicados en la siguiente tabla:

Peso (gramos)	Frecuencia absoluta
11,8 – 12,6	3
12,7 – 13,5	5
13,6 – 14,4	9
14,5 – 15,3	12
15,4 – 16,2	5
16,3 – 17,1	4
17,2 – 18,0	2

Se pide calcular:

- a) Identificar las marca de clase, los límites de los intervalos y las fronteras.
- b) La media aritmética.
- c) La mediana.
- d) La moda.

Solución:

a)

Intervalo	Front. Inf.	Front. Sup.	Límite Inf.	Límite Sup.	Marca	Frec. Abs.
11,8 – 12,6	11,75	12,65	11,8	12,6	12,2	3
12,7 – 13,5	12,65	13,55	12,7	13,5	13,1	5
13,6 – 14,4	13,55	14,45	13,6	14,4	14,0	9
14,5 – 15,3	14,45	15,35	14,5	15,3	14,9	12
15,4 – 16,2	15,35	16,25	15,4	16,2	15,8	5
16,3 – 17,1	16,25	17,15	16,3	17,1	16,7	4
17,2 – 18,0	17,15	18,05	17,2	18,0	17,6	2

b) Como son datos agrupados, la media aritmética se calcula multiplicando la marca de cada clase por el valor de la frecuencia absoluta, sumando todos estos productos y dividiendo por el total de datos.

```
> frecAbs <- c(3,5,9,12,5,4,2)
> marcas <- seq(12.2,17.6,0.9)
> marcas
[1] 12.2 13.1 14.0 14.9 15.8 16.7 17.6
> N <- sum(frecAbs)
> N
[1] 40
> media <- sum(frecAbs * marcas)/N
> media
[1] 14.6975
```

c) La mediana para datos agrupados se calcula como:

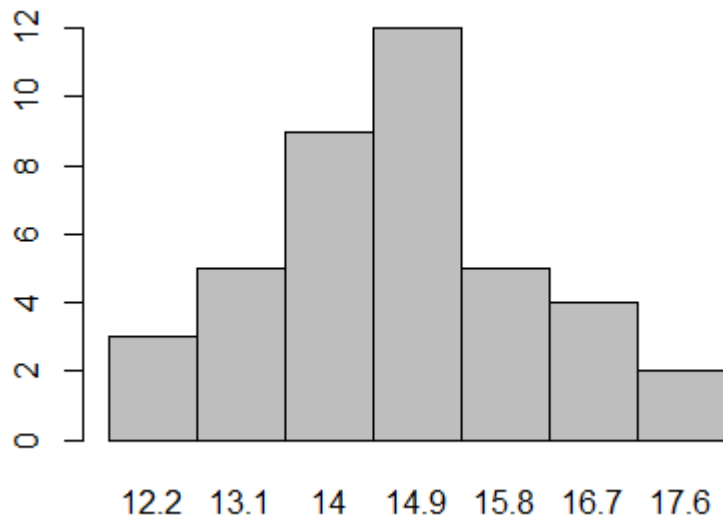
```
> frecAbsAcum <- cumsum(frecAbs)
> frecAbsAcum
[1] 3 8 17 29 34 38 40
```

Como no hay ningún valor cuya frecuencia absoluta acumulada corresponda a 20 (la mitad de los datos), identificamos el intervalo que supera por primera vez este valor. Se trata del cuarto intervalo, 14,5 – 15,3, cuyas fronteras son:

```
> ai <- 14.45 # Frontera inferior
> aimas1 <- 15.35 # Frontera superior
> Nimenos1 <- 17 # Frecuencia absoluta acumulada del intervalo anterior
> ni <- 12 # Frecuencia absoluta del intervalo que contiene la mediana
> mediana <- ai + (N/2 - Nimenos1)*(aimas1-ai)/ni
> mediana
[1] 14.675
```

Esta mediana interpolada corresponde al punto que divide el histograma en dos partes iguales, cada una correspondiente al 50% del área total.

```
> barplot(frecAbs, names=marcas, space=0)
```

Cada intervalo se representa mediante un rectángulo de base igual a la anchura de los intervalos (0,9 unidades) y una altura igual a la frecuencia absoluta correspondiente.

```
> frecAbs
[1] 3 5 9 12 5 4 2
> sum(frecAbs*0.9) # Área total del histograma
[1] 36
```

Luego necesitamos calcular la abcisa para la cual el área es 18.

```
> cumsum(frecAbs*0.9)
[1] 2.7 7.2 15.3 26.1 30.6 34.2 36.0
```

Efectivamente en el cuarto intervalo es cuando se supera ese valor, pues se pasa de 15,3 como área acumulada de las 3 primeras barras a 26,1 para las 4 primeras barras. Nos hemos pasado la siguiente área:

```
> 18-15.3
[1] 2.7
```

Para un rectángulo de altura 12, corresponde a una anchura de la barra de:

```
> 2.7/12
[1] 0.225
```

Este valor es la distancia entre la mediana y el comienzo del cuarto intervalo, medido desde la frontera inferior del mismo. Luego:

```
> 14.45 + 0.225
[1] 14.675
```

Que es el valor que habíamos obtenido.

d) La moda para datos agrupados se calcula como:

```
> frecAbs  
[1] 3 5 9 12 5 4 2
```

El intervalo modal es aquel con la mayor frecuencia absoluta. Es el cuarto intervalo.

La diferencia de frecuencia con respecto al intervalo premodal es:

```
> delta1 <- 12-9
```

La diferencia de frecuencia con respecto al intervalo postmodal es:

```
> delta2 <- 12-5  
> moda <- ai + delta1 * (aimas1-ai)/(delta1+delta2)  
> moda  
[1] 14.72
```

30. Las notas de 14 alumnos de Algoritmos y de Interacción Persona-Ordenador son las siguientes:

Alg.	3,3	1	6	5	8	2	5,4	7	7,5	6,5	9	10	5,1	5,8
IPO	2,4	3,5	5,5	5	9	1	4	7,5	6	7	10	8	4,3	3

Se pide representar los datos en un diagrama de dispersión y calcular el mejor ajuste lineal a los datos.

Solución:

```
> algoritmos <- c(3.3,1,6,5,8,2,5.4,7,7.5,6.5,9,10,5.1,5.8)  
> ipo <- c(2.4,3.5,5.5,5,9,1,4,7.5,6,7,10,8,4.3,3)  
> plot(algoritmos,ipo,main="Notas",xlab="Algoritmos",ylab="IPO",  
col="red",pch=16,cex=1.5,cex.main=2,cex.lab=1.5,cex.axis=1.5)  
> ajuste <- lm(ipo ~ algoritmos)  
> ajuste  
Call:  
lm(formula = ipo ~ algoritmos)  
  
Coefficients:  
(Intercept) algoritmos  
0.2760 0.8865
```

Es decir, que se predice que la nota de IPO y la de Algoritmos están relacionadas mediante la siguiente expresión lineal:

$$\text{Nota de IPO} = 0,2760 + 0,8865 * \text{Nota de Algoritmos}$$

```
> abline(ajuste,col="blue")
```

