

# Applied Biostatistics with R and rk.Teaching

Alfredo Sánchez Alberca ([asalber@ceu.es](mailto:asalber@ceu.es))

Department of Applied Math and Statistics  
CEU San Pablo

March 1, 2016



CEU

*Universidad  
San Pablo*

---

## Applied Biostatistics with R

Alfredo Sánchez Alberca (asalber@ceu.es)

### License terms

This work is licensed under an Attribution-NonCommercial-ShareAlike 4.0 International Creative Commons License. <http://creativecommons.org/licenses/by-nc-sa/4.0/>

You are free to:

- Share – copy and redistribute the material in any medium or format
- Adapt – remix, transform, and build upon the material

Under the following terms:



**Attribution.** You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



**NonCommercial.** You may not use the material for commercial purposes.



**ShareAlike.** If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

# Contents

<b>1</b>	<b>Frequency distributions and charts</b>	<b>1</b>
1.1	Solved exercises . . . . .	1
1.2	Proposed exercises . . . . .	3



# Frequency distributions and charts

## 1 Solved exercises

1. The number of children in a sample of 25 is

1, 2, 4, 2, 2, 2, 3, 2, 1, 1, 0, 2, 2, 0, 2, 2, 1, 2, 2, 3, 1, 2, 2, 1, 2.

Do the following operations:

- (a) Create a data frame with the variable children and enter the data.
- (b) Create the frequency table.



1. Select the menu Teaching > Frequency distribution > Frequency tabulation .
2. In the dialog displayed, select the variable children in the field Variable to tabulate and click the button Send.

- (c) Create the absolute frequency bar chart.



1. Select the menu Teaching > Charts > Bar chart .
2. In the dialog displayed, select the variable children in the field Variable and click the button Send.

- (d) Create also the relative frequency, cumulative absolute frequency and cumulative relative frequency bar charts, with their respective polygons.



Follow the steps above checking, in the Bar options tab, the box Relative frequencies for the relative frequencies bar chart, the box Cumulative frequencies for the cumulative absolute bar chart, and both of them for the cumulative relative frequency bar chart. Check the box Polygon, to plot the corresponding polygon.

2. The number of people treated in the emergency service of a hospital every day of November was

15	23	12	10	28	7	12	17	20	21	18	13	11	12	26
30	6	16	19	22	14	17	21	28	9	16	13	11	16	20

Do the following operations:

- (a) Create a data frame with the variable emergencies and enter the data.
- (b) Create the box plot. Are there some outlier? In that case, remove the outliers and proceed with the next part.



1. Select the menu **Teaching** » **charts** » **Box plot**.
2. In the dialog displayed select the variable emergencies in the field Variables and click the button Send.
3. In the output windows with the box plot identify the outliers.
4. In the data frame edition tab, remove the rows with the outliers right-clicking the row header and selecting **Delete this row**.

(c) Create the frequency table grouping data into 5 classes.



1. Select the menu **Teaching** » **Frequency distribution** » **Frequency tabulation**.
2. In the dialog displayed select the variable emergencies.
3. In the Classes tab check the box Grouping intervals, check the option Number of intervals, enter the desired number of intervals in the field Suggested intervals and click the button Send.

(d) Create the absolute frequencies histogram.



1. Select the menu **Teaching** » **charts** » **Histogram**.
2. In the dialog displayed select the variable emergencies in the field Variable.
3. In the Classes tab, check the box Grouping intervals, check the box Number of intervals, enter the desired number of intervals in the field Suggested intervals and click the button Send.

(e) Create also the relative frequency, cumulative absolute frequency and cumulative relative frequency histograms, with their respective polygons.



Follow the steps above checking, in the Histogram options, the box Relative frequencies for the relative frequency histogram, the box Cumulative frequencies for the cumulative absolute frequency histogram, and both of them for the cumulative relative frequency histogram. Check the box Polygon to plot the corresponding polygons.

3. The blood type of a sample of 30 persons are:

A, B, B, A, AB, 0, 0, A, B, B, A, A, A, AB,  
A, A, A, B, 0, B, B, B, A, A, A, 0, A, AB, 0.

Do the following operations:

- (a) Create a data frame with the variable blood.type and enter the data.
- (b) Create the frequency table.



1. Select the menu **Teaching** » **Frequency distribution** » **Frequency tabulation**.
2. In the dialog displayed, select the variable blood.type in the field Variable to tabulate and click the button Send.

(c) Create the pie chart.



1. Select the menu **Teaching** » **charts** » **Pie chart**.
2. In the dialog displayed, select the variable blood.type in the field Variables and click the button Send.

4. The age and the marital status of a sample of 27 persons are:

Marital status	Age									
Single	31	45	35	65	21	38	62	22	31	
Married	62	39	62	59	21	62				
Widow(er)	80	68	65	40	78	69	75			
Divorced	31	65	59	49	65					

Do the following operations:

- Create a data frame with the variables marital.status and age and enter the data.
- Create the frequency table of the variable age for every marital status.



- Select the menu **Teaching > Frequency distribution > Frequency tabulation**.
- In the dialog displayed, select the variable age in the field Variable to tabulate, check the box Tabulate by groups, select the variable marital.status in the field Grouping variable(s) and click the button Send.

- Create the box plots of age for every marital status. Are there outliers? Which group have more spread in ages?



- Select the menu **Teaching > charts > Box plot**.
- In the dialog displayed, select the variable age in the field Variables, check the box Plot by groups, select the variable marital.status in the field Grouping variable and click the button Send.

## 2 Proposed exercises

- El número de lesiones padecidas durante una temporada por cada jugador de un equipo de fútbol fue el siguiente:

0, 1, 2, 1, 3, 0, 1, 0, 1, 2, 0, 1, 1, 1, 2, 0, 1, 3, 2, 1, 2, 1, 0, 1

Do the following operations:

- Construir la frequency table.
  - Dibujar el diagrama de barras de las relative frequencies and de relative frequencies acumuladas.
  - Dibujar el pie chart.
- Para realizar un estudio sobre la estatura de los estudiantes universitarios, seleccionamos, mediante un proceso de muestreo aleatorio, una muestra de 30 estudiantes, obteniendo los siguientes resultados (medidos en centímetros):

179, 173, 181, 170, 158, 174, 172, 166, 194, 185,  
162, 187, 198, 177, 178, 165, 154, 188, 166, 171,  
175, 182, 167, 169, 172, 186, 172, 176, 168, 187.

Do the following operations:

- Dibujar el histograma de las absolute frequencies agrupando desde 150 a 200 en clases de amplitud 10.
  - Dibujar el box plot. ¿Existe algún outlier?.
- El conjunto de datos neonatos del paquete rk.Teaching, contiene información sobre una muestra de 320 recién nacidos en un hospital durante un año que cumplieron el tiempo normal de gestación. Do the following operations:

- (a) Construir la frequency table de la puntuación Apgar al minuto de nacer. Si se considera que una puntuación Apgar de 3 o menos indica que el neonato está deprimido, ¿qué porcentaje de niños está deprimido en la muestra?
- (b) Comparar las distribuciones de frecuencias de las puntuaciones Apgar al minuto de nacer según si la madre es mayor o menor de 20 años. ¿En qué grupo hay más neonatos deprimidos?
- (c) Construir la frequency table para el peso de los neonatos, agrupando en clases de amplitud 0.5 desde el 2 hasta el 4.5. ¿En qué intervalo de peso hay más niños?
- (d) Comparar la distribución de relative frequencies del peso de los neonatos según si la madre fuma o no. Si se considera como peso bajo un peso menor de 2.5 kg, ¿En qué grupo hay un mayor porcentaje de niños con peso bajo?
- (e) Si en los recién nacidos se considera como peso bajo un peso menor de 2.5 kg, calcular la prevalencia del bajo peso de recién nacidos en el grupo de madres fumadoras and en el de no fumadoras.
- (f) Calcular el riesgo relativo de que un recién nacido tenga bajo peso cuando la madre fuma, frente a cuando la madre no fuma.
- (g) Construir el diagrama de barras de la puntuación Apgar al minuto. ¿Qué puntuación Apgar es la más frecuente?
- (h) Construir el diagrama de relative frequencies acumuladas de la puntuación Apgar al minuto. ¿Por debajo de qué puntuación estarán la mitad de los niños?
- (i) Comparar mediante diagramas de barras de relative frequencies las distribuciones de las puntuaciones Apgar al minuto según si la madre ha fumado o no durante el embarazo. ¿Qué se puede concluir?
- (j) Construir el histograma de pesos, agrupando en clases de amplitud 0.5 desde el 2 hasta el 4.5. ¿En qué intervalo de peso hay más niños?
- (k) Comparar la distribución de relative frequencies del peso de los neonatos según si la madre fuma o no. ¿En qué grupo se aprecia menor peso de los niños de la muestra?
- (l) Comparar la distribución de relative frequencies del peso de los neonatos según si la madre fumaba o no antes del embarazo. ¿Qué se puede concluir?
- (m) Construir el diagrama de caja and bigotes del peso. ¿Entre qué valores se considera que el peso de un neonato es normal? ¿Existen datos atípicos?
- (n) Comparar el box plot and bigotes del peso, según si la madre fumó o no durante el embarazo and si era mayor o no de 20 años. ¿En qué grupo el peso tiene más dispersión central? ¿En qué grupo pesan menos los niños de la muestra?
- (o) Comparar el box plot de la puntuación Apgar al minuto and a los cinco minutos. ¿En qué variable hay más dispersión central?