

Applied Biostatistics with R and rk.Teaching

Alfredo Sánchez Alberca (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

1	Frequency distributions and charts	1
1.1	Solved exercises	1
1.2	Ejercicios propuestos	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 frequencies bar chart.



1. Select the menu Teaching ▶ Gráficos ▶ Diagrama de barras.
2. In the dialog displayed, select the variable hijos in the field Variable and hacer clic en el botón Send.

- (d) Para la misma tabla de frecuencias anterior, dibujar también el diagrama de barras de las relative frequencies, el de absolutas acumuladas and el de relativas acumuladas, además de sus correspondientes polígonos.



Repetir los pasos del apartado anterior activando, en la solapa de Opciones de las barras, la opción Frecuencias relativas si se desea el diagrama de barras de relative frequencies, activando la opción Frecuencias acumuladas si se desea el diagrama de barras de frecuencias acumuladas and activando la opción Polígono para obtener el polígono asociado.

2. En un hospital se realizó un estudio sobre el número de personas que ingresaron en urgencias cada día del mes de noviembre. Los datos observados fueron:

15, 23, 12, 10, 28, 50, 12, 17, 20, 21, 18, 13, 11, 12, 26
30, 6, 16, 19, 22, 14, 17, 21, 28, 9, 16, 13, 11, 16, 20

Se pide:

- (a) Crear un conjunto de datos con la variable urgencias e introducir los datos.
- (b) Dibujar el diagrama de cajas. ¿Existe algún dato atípico? En el caso de que exista, eliminarlo and proceder con los siguientes apartados.



1. Select the menu Teaching►Gráficos►Diagrama de cajas.
2. In the dialog displayed, select the variable urgencias in the field Variables y click the button Send.
3. En la ventana que aparece con el diagrama de cajas identificar el dato atípico.
4. Ir a la ventana de edición de datos and eliminar la fila del dato atípico haciendo clic con el botón derecho del ratón en la cabecera de la fila and seleccionando Borrar esta fila.

(c) Construir la tabla de frecuencias agrupando en 5 clases.



1. Select the menu Teaching►Frequency distribution►Frequency tabulation.
2. In the dialog displayed select the variable urgencias.
3. En la solapa de Clases activar la casilla Agrupar en intervalos, marcar la opción Número de intervalos e introducir el número deseado de intervalos in the field Intervalos sugeridos and hacer clic sobre el botón Send.

(d) Dibujar el histograma de absolute frequencies correspondiente a la tabla anterior.



1. Select the menu Teaching►Gráficos►Histograma.
2. In the dialog displayed select the variable urgencias in the field Variable.
3. En la solapa de Clases activar la casilla Agrupar en intervalos, marcar la opción Número de intervalos e introducir el número deseado de intervalos in the field Intervalos sugeridos and hacer clic sobre el botón Send.

(e) Para la misma tabla de frecuencias anterior, dibujar también el histograma de las relative frequencies, el de absolutas acumuladas and el de relativas acumuladas, además de sus correspondientes polígonos.



Repetir los pasos del apartado anterior activando, en la solapa de Opciones del histograma, la opción Frecuencias relativas si se desea el histograma de relative frequencies, activando la opción Frecuencias acumuladas si se desea el histograma de frecuencias acumuladas and activando la opción Polígono para obtener el polígono asociado.

3. Los grupos sanguíneos de una muestra de 30 personas son:

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.

Se pide:

- (a) Crear un conjunto de datos con la variable grupo.sanguineo e introducir los datos.
- (b) Construir la tabla de frecuencias.



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

(c) Dibujar el diagrama de sectores.



1. Select the menu Teaching►Gráficos►Diagrama de sectores.
2. In the dialog displayed, select the variable grupo.sanguineo in the field Variables and hacer clic sobre el botón Send.

4. En un estudio de población se tomó una muestra de 27 personas, and se les preguntó por su edad and estado civil, obteniendo los siguientes resultados:

Estado civil	Edad									
Soltero	31	45	35	65	21	38	62	22	31	
Casado	62	39	62	59	21	62				
Viudo	80	68	65	40	78	69	75			
Divorciado	31	65	59	49	65					

Se pide:

- Crear un conjunto de datos con la variables estado.civil and edad e introducir los datos.
- Construir la tabla de frecuencias de la variable edad para cada categoría de la variable estado.civil.



- Select the menu Teaching►Frequency distribution►Frequency tabulation.
- In the dialog displayed, select the variable edad in the field Variable a tabular, activar la casilla Tabular por grupos, select the variable estado.civil in the field Variable de agrupación and click the button Send.

- Dibujar los diagramas de cajas de la edad según el estado civil. ¿Existen datos atípicos? ¿En qué grupo hay mayor dispersión?



- Select the menu Teaching►Gráficos►Diagrama de cajas.
- In the dialog displayed, select the variable edad in the field Variables, activar la casilla Dibujar por grupos, select the variable estado.civil in the field Variable de agrupación and click the button Send.

2 Ejercicios propuestos

- 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

Se pide:

- Construir la tabla de frecuencias.
 - Dibujar el diagrama de barras de las relative frequencies and de relative frequencies acumuladas.
 - Dibujar el diagrama de sectores.
- 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.

Se pide:

- Dibujar el histograma de las absolute frequencies agrupando desde 150 a 200 en clases de amplitud 10.

- (b) Dibujar el diagrama de cajas. ¿Existe algún dato atípico?.
3. 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. Se pide:
- Construir la tabla de frecuencias 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?
 - 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?
 - Construir la tabla de frecuencias 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?
 - 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?
 - 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.
 - Calcular el riesgo relativo de que un recién nacido tenga bajo peso cuando la madre fuma, frente a cuando la madre no fuma.
 - Construir el diagrama de barras de la puntuación Apgar al minuto. ¿Qué puntuación Apgar es la más frecuente?
 - Construir el diagrama de relative frequencies acumuladas de la puntuación Apgar al minuto. ¿Por debajo de que puntuación estarán la mitad de los niños?
 - 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?
 - 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?
 - 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?
 - 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?
 - 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?
 - Comparar el diagrama de cajas 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?
 - Comparar el diagrama de cajas de la puntuación Apgar al minuto and a los cinco minutos. ¿En qué variable hay más dispersión central?