# EXERCISES OF STATISTICS

Subject: Statistics Applied tho the Health Sciences
Course: $2^{nd}$
Degree: Physiotherapy
Year: 2015-2016

Authors: Santiago Angulo Díaz-Parreño (sangulo@ceu.es)
José Miguel Cárdenas Rebollo (cardenas@ceu.es)
Anselmo Romero Limón (arlimon@ceu.es)
Alfredo Sánchez Alberca (asalber@ceu.es)

CEU
*Universidad San Pablo*

# Contents

# 1 Descriptive Statistics

1. Classify the following variables

   (a) Daily hours of exercise.

   (b) Nationality.

   (c) Blood pressure.

   (d) Severity of illness.

   (e) Number of sport injuries in a year.

   (f) Daily calorie intake.

   (g) Size of clothing.

   (h) Subjects passed in a course.

2. The number of injuries suffered by the members of a soccer team in a league were

   | | | | | | | | | | | | |
   |---|---|---|---|---|---|---|---|---|---|---|---|
   | 0 | 1 | 2 | 1 | 3 | 0 | 1 | 0 | 1 | 2 | 0 | 1 |
   | 1 | 1 | 2 | 0 | 1 | 3 | 2 | 1 | 2 | 1 | 0 | 1 |

   (a) Construct the frequency distribution table of the sample.

   (b) Draw the bar chart of the sample and the polygon.

   (c) Draw the cumulative frequency bar chart and the polygon.

3. A survey about the daily number of medicines consumed by people over 70 years, shows the following results:

   | | | | | | | | | | | | | | | | | | | |
   |---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
   | 3 | 1 | 2 | 2 | 0 | 1 | 4 | 2 | 3 | 5 | 1 | 3 | 2 | 3 | 1 | 4 | 2 | 4 | 3 | 2 |
   | 3 | 5 | 0 | 1 | 2 | 0 | 2 | 3 | 0 | 1 | 1 | 5 | 3 | 4 | 2 | 3 | 0 | 1 | 2 | 3 |

   (a) Construct the frequency distribution table of the sample.

   (b) Draw the bar chart of the sample and the polygon.

   (c) Draw the cumulative relative frequency bar chart and the polygon.

4. In survey about the dependency of older people, 23 persons over 75 years were asked about the help they need in daily life. The answers were

   | | | | | | | | | | | | | | | | | | | | | | |
   |---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
   | B | D | A | B | C | C | B | C | D | E | A | B | C | E | A | B | C | D | B | B | A | A | B |

   where the meanings of letters are:

   A No help.

   B Help climbing stairs.

   C Help climbing stairs and getting up from a chair or bed.

   D Help climbing stairs, getting up and dressing.

   E Help for almost everything.

   Construct the frequency distribution table and the suitable chart.

5. 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 |

   (a) Construct the frequency distribution table of the sample.

   (b) Draw a suitable chart for the frequency distribution.

(c) Draw a suitable chart for the cumulative frequency distribution.

6. The following frequency distribution table represents the distribution of time (in min) required by people attended in a medical dispensary.

| Time | $n_i$ | $f_i$ | $N_i$ | $F_i$ |
|---|---|---|---|---|
| $[0, 5)$ | 2 | | | |
| $[5, 10)$ | | | 8 | |
| $[10, 15)$ | | | | 0.7 |
| $[15, 20)$ | 6 | | | |

(a) Complete the table.

(b) Draw the ogive.

7. Use the data of exercise 2 to calculate the following statistics and interpret them.

(a) Mean.

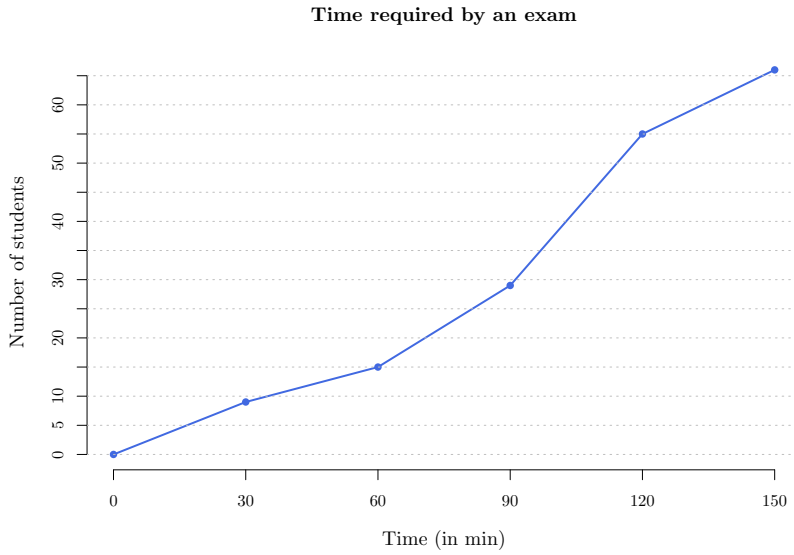
(b) Median.

(c) Mode.

(d) Quartiles.

(e) Percentile 32.

8. The chart below shows the cumulative distribution of the time (in min) required by 20 students to do an exam.




(a) A which time have finished half of the students? And 90% of students?

(b) Which percentage of students have finished after 100 minutes?

(c) Which is the time that best represent the time required by students in the sample to finish the exam? Is this value representative or not?

9. In a study about the children growth two samples where drawn, one for newborns and the other for one year old. The height in cm of children in both samples were

Newborn children: 51, 50, 51, 53, 49, 50, 53, 50, 47, 50
One year old children: 62, 65, 69, 71, 65, 66, 68, 69

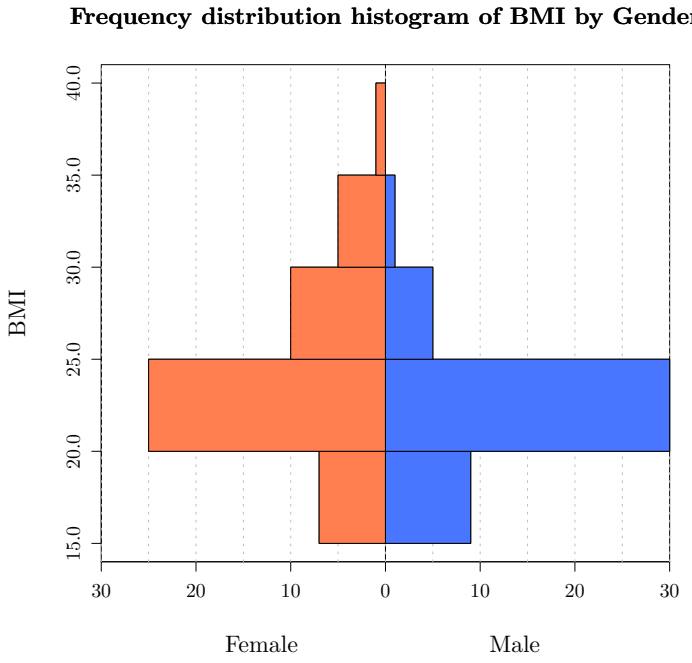
In which group is more representative the mean? Justify the answer.

10. To determine the accuracy of a method for measuring hematocrit in blood, the measurement was repeated 8 times on the same blood sample. The results in percentage of hematocrit in plasma were

    42.2 42.1 41.9 41.8 42 42.1 41.9 42

    What do you think about the accuracy of the method?

11. The histogram below shows the frequency distribution of the body mass index (BMI) of a group of people by gender.




    (a) Draw the pie chart for the gender.

    (b) In which group is more representative the mean of the BMI?

    (c) Calculate the mean for the whole sample.

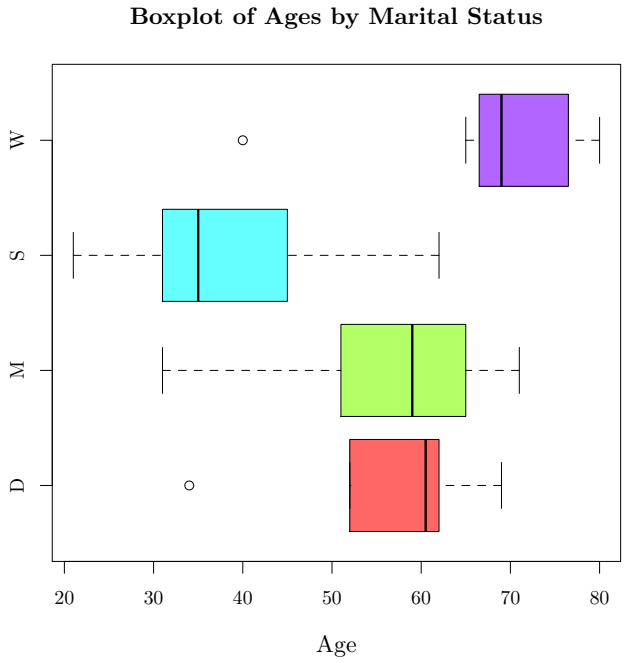
    Use the following sums
    Males: $\sum x_i = 1002$ kg/m$^2$ $\sum x_i^2 = 22781$ kg$^2$/m$^4$
    Females: $\sum x_i = 1160$ kg/m$^2$ $\sum x_i^2 = 29050$ kg$^2$/m$^4$

12. The following table represents the frequency distribution of the yearly uses of a health insurance in a sample of clients of a insurance company.

| Uses: | 0 | 1 | 2 | 3 | 4 | 5 | 7 |
|---|---|---|---|---|---|---|---|
| Clients: | 4 | 8 | 6 | 3 | 2 | 1 | 1 |

    Draw the box plot. How is the symmetry of the distribution?

13. The box plots below correspond to the age of a sample of people by marital status.




(a) Which group has higher ages?

(b) Which group has lower central dispersion?

(c) Which groups have outliers?

(d) Which group has a distribution of ages more asymmetric?

14. The following table represents the frequency distribution of ages at which a group of people suffered a heart attack.

| Age | [40-50) | [50-60) | [60-70) | [70-80) | [80-90) |
|---|---|---|---|---|---|
| Persons | 6 | 12 | 23 | 19 | 5 |

Could we assume that the sample comes from a normal population?

Use the following sums: $\sum x_i = 4275$ years, $\sum (x_i - \bar{x})^2 = 7462$ years$^2$, $\sum (x_i - \bar{x})^3 = -18249$ years$^3$, $\sum (x_i - \bar{x})^4 = 2099636$ years$^4$.

15. To compare two rehabilitation treatments $A$ and $B$ for an injury, every treatment was applied to a different group of people. The number of days required to cure the injury in every group is shown in the following table:

| Days | $A$ | $B$ |
|---|---|---|
| 20-40 | 5 | 8 |
| 40-60 | 20 | 15 |
| 60-80 | 18 | 20 |
| 80-100 | 7 | 7 |

(a) In which treatment is more representative the mean?

(b) In which treatment the distribution of days is more skew?

(c) In which treatment the distribution is more peaked?

Use the following sums:
$A$: $\sum x_i = 3040$ days, $\sum (x_i - \bar{x})^2 = 14568$ days$^2$, $\sum (x_i - \bar{x})^3 = 17011.2$ days$^3$, $\sum (x_i - \bar{x})^4 = 9989603$

days$^4$
*B*: $\sum x_i = 3020$ days, $\sum(x_i - \bar{x})^2 = 16992$ days$^2$, $\sum(x_i - \bar{x})^3 = -42393.6$ days$^3$, $\sum(x_i - \bar{x})^4 = 12551516$ days$^4$

16. The systolic blood pressure (in mmHg) of a sample of persons is

    135 128 137 110 154 142 121 127 114 103

    (a) Calculate the central tendency statistics.

    (b) How is the relative dispersion with respect to the mean?

    (c) How is the skewness of the sample distribution?

    (d) How is the kurtosis of the sample distribution?

    (e) If we know that the method used for measuring the blood pressure is biased, and, in order to get the right values, we have to apply the linear transformation $y = 1.2x - 5$, which are values of the statistics required to answer the previous questions for the corrected values of the blood pressure?

    Use the following sums: $\sum x_i = 1271$ mmHg, $\sum(x_i - \bar{x})^2 = 2188.9$ mmHg$^2$, $\sum(x_i - \bar{x})^3 = 2764.32$ mmHg$^3$, $\sum(x_i - \bar{x})^4 = 1040080$ mmHg$^4$.

17. The table below contains the frequency of pregnancies, abortions and births of a sample of 999 women in a city.

| Num | Pregnancies | Abortions | Births |
|---|---|---|---|
| 0 | 61 | 751 | 67 |
| 1 | 64 | 183 | 80 |
| 2 | 328 | 51 | 400 |
| 3 | 301 | 10 | 300 |
| 4 | 122 | 2 | 90 |
| 5 | 81 | 2 | 62 |
| 6 | 29 | 0 | 0 |
| 7 | 11 | 0 | 0 |
| 8 | 2 | 0 | 0 |

    (a) How many birth outliers are in the sample?

    (b) Which variable has lower spread with respect to the mean?

    (c) Which value is relatively higher, 7 pregnancies or 4 abortions? Justify the answer.

    Use the following sums:
    Pregnancies: $\sum x_i = 2783$, $\sum x_i^2 = 9773$.
    Abortions: $\sum x_i = 333$, $\sum x_i^2 = 559$.
    Births: $\sum x_i = 2450$, $\sum x_i^2 = 7370$.