Applied Biostatistics with R and rk.Teaching

Alfredo Sánchez Alberca (asalber@ceu.es)

Department of Applied Math and Statistics
CEU San Pablo

April 3, 2016

CEU
*Universidad
San Pablo*

Applied Biostatistics with R
Alfredo Sánchez Alberca (asalber@ceu.es)

License terms ©

Contents

Hypothesis tests

1 Solved exercises

1. The active ingredient concentration of a random sample of 10 drug containers drawn from a batch are (in mg/mm^3)

$$17.6 \quad 19.2 \quad 21.3 \quad 15.1 \quad 17.6 \quad 18.9 \quad 16.2 \quad 18.3 \quad 19.0 \quad 16.4$$

Do the following operations:

(a) Create a data set with the variable concentration and enter the data of the sample.

(b) Test the two-sided hypothesis: H_0: $\mu = 18$ vs H_1: $\mu \neq 18$ with a significance level 0.05.

> 1. Select the menu [Teaching] ≫ [Parametric tests] ≫ [Means] ≫ [t-test for the mean of one population].
> 2. In the dialog displayed insert the variable concentration in the field Mean of.
> 3. In the Test options tab enter 18 in the field Null hypothesis, check the box Two-sided for the alternative hypothesis and click the button Submit.

(c) Test the two-sided hypothesis: H_0: $\mu = 19.5$ vs H_1: $\mu \neq 19.5$ with significance levels 0.05 and 0.01. How affects the significance level to the test?

> 1. Select the menu [Teaching] ≫ [Parametric tests] ≫ [Means] ≫ [t-test for the mean of one population].
> 2. In the dialog displayed insert the variable concentration in the field Mean of.
> 3. In the Test options tab enter 19.5 in the field Null hypothesis, check the box Two-sided for the alternative hypothesis and click the button Submit.

(d) Test the two-sided hypothesis: H_0: $\mu = 17$ vs H_1: $\mu \neq 17$ with a significance level 0.05. Test also the one-sided hypothesis: H_0: $\mu = 17$ vs H_1: $\mu > 17$ with a significance level 0.05. What is the relation between the p-value of the two-sided and the one-sided tests?

> 1. Select the menu [Teaching] ≫ [Parametric tests] ≫ [Means] ≫ [t-test for the mean of one population].
> 2. In the dialog displayed insert the variable concentration in the field Mean of.
> 3. In the Test options tab enter 17 in the field Null hypothesis, check the box Two-sided for the two-sided alternative hypothesis or Greater for the one-sided alternative hypothesis and click the button Submit.

(e) If the manufacturer of the drug affirms that has increased the active ingredient concentration with respect to previous batches where the concentration was 17 mg/mm^3, can we believe it?

(f) What is the sample size required to detect an increase in the concentration mean of 0.5 mg/mm^3 with a significance level $\alpha = 0.05$ and a power $1 - \beta = 0.8$?

To compute the sample size is required to know the population standard deviation or an estimate of it. To compute an estimate of the population standard deviation:

1. Select the menu Teaching 〉 Descriptive statistics 〉 Statistics.
2. In the dialog displayed enter the variable concentration in the field Variable.
3. In the Basic statistics tab check the box Corrected standard deviation and click the button Submit.

To compute the sample size:

1. Select the menu Teaching 〉 Parametric tests 〉 Means 〉 Sample size for the t-test.
2. In the dialog displayed check the box One population for the type of test, check the box one-sided for the alternative hypothesis, enter 0.5 in the field Difference between the means, enter the corrected standard deviation in the field Standard deviation, enter the value 0.05 in the field Significance level, enter the value 0.8 in the field Power and click the button Submit.

2. In a survey performed by a university about the use of the library, a random sample of 34 students has been asked whether they go to the library at least once a week. The answers are shown below.

no   yes   no   no   no   yes   no   yes   yes   yes   yes   no   yes   no   yes   no   no
no   yes   yes   yes   no   no   yes   no   no   yes   yes   no   no   yes   no   yes   no

(a) Create a data set with the variable answer and enter the data of the sample.

(b) Test if the percentage of students that uses the library at least once a week is greater than 40%.

1. Select the menu Teaching 〉 Parametric test 〉 Proportions 〉 Test for one proportion.
2. In the dialog displayed insert the variable answer in the field Variable and enter yes in the field Proportion of.
3. In the Test options tab enter 0.4 in the field Null hypothesis, check the box Greater for the alternative hypothesis and click the button Submit.

3. A study tries to determine if there is difference between the ages at which two populations of babies A and B begin to walk by themselves. For testing it a random sample of every population was draw and the ages in moths at which the babies began to walk was recorded. The data is shown in the following table.

| A | 9.5 | 10.5 | 9.0 | 9.8 | 10.0 | 13.0 | 10.0 | 13.5 | 10.0 | 9.8 | | |
| B | 12.5 | 9.5 | 13.5 | 13.8 | 12.0 | 13.8 | 12.5 | 9.5 | 12.0 | 13.5 | 12.0 | 12.0 |

(a) Create a data set with the variables age and population and enter the data of the sample.

(b) Test if there is difference between the age means at which babies begin to walk in both populations with a significance level 0.05.

1. Select the menu Teaching 〉 Parametric tests 〉 Means 〉 t-test for comparing the means of two independent populations.
2. In the dialog displayed insert the age in the field Compare mean of and the variable population in the field According to.
3. In the frame Populations to compare insert the value A in the field Compare and the value B in the field With.

There are differences between the means if the p-value is less than 0.05.

4. Some researchers have observed a greater airways resistance in smokers than in non-smokers. To test the hypothesis the percentage of tracheobronchial retention was measured in a sample of persons when they were smokers an after one year of quitting. The data is shown in the following table.

Percentage of tracheobronchial retention

Smoking	One year after quitting
60.6	47.5
12.0	13.3
56.0	33.0
75.2	55.2
12.5	21.9
29.7	27.9
57.2	54.3
62.7	13.9
28.7	8.90
66.0	46.1
25.2	29.8
40.1	36.2

(a) Create a data set with the variables before and after and enter the data of the sample.

(b) Perform the test to confirm or reject the hypothesis of researchers.

> 1. Select the menu `Teaching` ≫ `Parametric tests` ≫ `Means` ≫ `t-test for comparing the means of two paired populations`.
> 2. In the dialog displayed insert the variable before in the field Compare mean of population and the variable after in the field With mean of population.
> 3. In the Test options check the box Greater for the alternative hypothesis and click the button Submit.

5. In a course there are two groups of students, one in the morning and the other in the afternoon. In the morning group 55 students of 80 passed, while in the afternoon group 32 students of 90 passed. Are there significant differences between the percentages of students that passed in the morning and in the afternoon? Can we conclude that the timetable is the cause of the differences? Justify the answer.

> (a) Select the menu `Teaching` ≫ `Parametric tests` ≫ `Proportions` ≫ `Test for comparing two proportions`.
> (b) In the dialog displayed check the box Manual entry of frequencies, enter 55 in the field Sample frequency 1, enter 80 in the field Sample size 1, enter 32 in the field Sample frequency 2 and enter 90 in the field Sample size 2.
> (c) In the Test options tab check the box Two-sided and click the button Submit.

2 Proposed exercises

1. The data set neonates of the package rk.Teaching, contains information about the pulse of a sample of people after doing different exercises: resting pulse in beats per minute (pulse1), pulse after doing exercise in beats per minute (pulse2), type of exercise (ran, 1=running, 2=walking), gender (gender, 1=male, 2=female) and the weight (weight). Do the following operations:

   (a) Test if the resting pulse is less than 75 beats per minute.

   (b) What sample size is required to detect an increment of 2 beats per minute in the resting pulse with significance level 0.05 and power 0.9?

   (c) Test if the pulse after running is greater than 85 beats per minute.

   (d) A person has slight tachycardia if the resting pulse is greater than 90 beats per minute. Test if the percentage of people with slight tachycardia is greate than 5%.

   (e) Can we conclude that exercise increases the pulse with significance level 0.05? And with significance level 0.01? Justify the answer.

   (f) Is there difference in the pulse means after walking and running? Justify the answer.

   (g) Is there difference between the resting pulse means of males and females? And between the pulse means after running? Justify the answer.