

INFERENCEAL STATISTICS BY EXAMPLE: MICROSOFT EXCEL & EPI INFO

Learning Objectives:

- Statistical tests in Microsoft Excel. (Exercise 3, 4)
- Statistical tests in Epi Info. (Exercise 4,)
- Student t-test for two independent samples. (Exercise 1–3, & 5)
- Chi-square test. (Exercise 4 & 6)
- Fisher exact test. (Exercise 7)

Create in the **Lab10** folder a new Microsoft Excel file named **StatisticalTests.xlsx** and a new Microsoft Word file named **StatisticalTests.docx**.

EXERCISE 1

A study has been conducted to compare the efficacy of a new drug in treatment of arterial hypertension. Systolic blood pressure decreased on average with 13.7 mmHg on subjects who received the new drug, and with 5.5 mmHg on subjects who received placebo. There has been tested if a significant difference between the two means exists and a p-value smaller than 0.001 was obtained [Giles TD, Khan BV, Lato J, Brener L, Ma Y, Lukic T. Nebivolol monotherapy in younger adults (younger than 55 years) with hypertension: a randomized, placebo-controlled trial. *J Clin Hypertens* (Greenwich). 2013 Sep;15(9):687-93. doi: 10.1111/jch.12169. Epub 2013 Jul 16.]

Provide an answer for the following requests in the **StatisticalTests.docx** file. In green you will find the solution for the first Exercise!

1. Formulate the null hypothesis for this test: There is no statistically significant difference between the decrease in systolic blood pressure on subjects who received the new drug and the subjects who received placebo.
2. Formulate the null hypothesis for this test: There is a statistically significant difference between the decrease in systolic blood pressure on subjects who received the new drug and the subjects who received placebo.
3. Based on the obtained p-value, which hypothesis will be accepted? With a 5% risk of error, we reject the H_0 .
4. Based on the obtained p-value, specify the decision for this test (the difference between these two decreases is statistically significant?): The decrease in systolic blood pressure on subjects who received the new drug proved to be statistically significant higher compared with the decrease in systolic blood pressure on subjects who received placebo.
5. Explain the reason for chosen decision in Request 4:

EXERCISE 2

For the same new drug as presented in Exercise 1 the decreasing of the heart rate has also been investigated. The heart rate decreased on average with 10.7 heartbeats on subjects who received the new drug and with 2.4 heartbeats on subjects who received placebo. The heartbeats followed a normal distribution and the calculated statistical parameter was equal with 9.54 (critical value equal with 1.964). Therefore, it has been tested if a significant difference exists between the new drug and placebo in regards of heartbeats.

1. Formulate the null hypothesis for this test:
2. Formulate the null hypothesis for this test:
3. Determine the critical region:

4. Determine the acceptance region:
5. Based on the obtained results, which hypothesis will be accepted?
6. Specify if the heartbeats difference between the two investigated groups is statistically significant:
7. Explain the reason for chosen decision in Request 6:

EXERCISE 3

For the same new drug as presented in Exercise 1 the decreasing in diastolic blood pressure has been investigated. The diastolic blood pressure decreased on average with 11.8 mmHg (and a standard deviation of 8.8 mmHg, n = 423) for subjects who received the new drug and with 5.5 mmHg (and a standard deviation of 9.5 mmHg, n = 214) for subjects who received placebo. Diastolic blood pressure proved to be normal distributed and the variances proved not to be significantly different. The critical value for a significance level of 5% is equal with 1.964.

1. Formulate the null hypothesis for this test:
2. Formulate the null hypothesis for this test:
3. Determine the critical region:
4. Determine the acceptance region:
5. Calculate in Microsoft Excel the value of statistical test. For calculation, insert a new sheet in the **StatisticalTests.xlsx** file and name it Exercise 3. Create in this sheet the following structure and used the provided functions:

	A	B
1	m(new drug)	11.8
2	s(new drug)	8.8
3	n(new drug)	423
4	m(placebo)	5.5
5	s(placebo)	9.5
6	n(placebo)	214
7	α(significance level)	0.05
8	t(critical value)	1.964
9	S ²	
10	t(calculated value)	
11	p-value	

- For t(critical value) – cell B8 use the following formula: `=ROUND(TINV(B7,B3+B6-2),3)`
- For S² – cell B9 use the following formula: $S^2 = \frac{(n_1 - 1) \times S_1^2 + (n_2 - 1) \times S_2^2}{n_1 + n_2 - 2}$, where $S_1^2 = s(\text{new drug})^2$, $S_2^2 = s(\text{placebo})^2$, $n_1 = n(\text{new drug})$, $n_2 = n(\text{placebo})$. The implementation of this formula in Excel is: `=(B3-1)*B2^2+(B6-1)*B5^2)/(B3+B6-2)`
- For t(calculated value) use the following formula: $t = \frac{m_1 - m_2}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}}$ where $m_1 = m(\text{new drug})$, $n_1 = n(\text{new drug})$, $m_2 = m(\text{placebo})$, $n_2 = n(\text{placebo})$. The implementation of this formula in Excel is: `=(B1-B4)/(SQRT(B9/B3+B9/B6))`
- For p-value use the following formula: `=T.DIST.2T(B10,B3+B6-2)`

6. Which hypothesis will be accepted?

7. Specify if the difference between the two decreases in diastolic blood pressure is statistically significant:
8. Explain the reason for chosen decision in Request 7:

EXERCISE 4

For the same study as described in Exercise 1, in the group of subjects who received the new drug, 120 out of 427 were women while in the placebo group 95 out of 214 were women. A test has been applied to identify if there is any significant differences between the investigated groups (subjects who received the new drug and subjects who received placebo) in regards of gender. The test has been applied at a significance level of 5% and the critical value was equal with 3.84.

1. Formulate the null hypothesis for this test:
2. Formulate the null hypothesis for this test:
3. Determine the critical region:
4. Create in Microsoft Excel and fill with values the observed contingency table using the model presented below. For calculation, insert a new sheet in the **StatisticalTests.xlsx** file and name it Exercise 4. Create in this sheet the following structure and fill it with data presented above:

	A	B	C	D
1	Observed contingency table			
2		Nebivolol	Placebo	Total
3	Women			=SUM(B3:C3)
4	Men			0
5	Total			GrandTotal

Use the SUM predefined function to calculate the total for columns (Nebivolol and Placebo), rows (Women and Men) and respectively GrandTotal.

5. Create in the same sheet theoretical/expected contingency table using the formulas provided below:

	F	G	H	I
1	Expected contingency table			
2		Nebivolol	Placebo	Total
3	Women			215
4	Men			426
5	Total	427	214	641

Formulas shown in the image:

- Cell G3: $=I3*G5/I5$
- Cell H3: $=I3*H5/I5$
- Cell F4: $=I4*G5/I5$
- Cell I4: $=I4*H5/I5$

6. Compute the value of statistical test (Chi-square test) using the formula:

$$\chi^2 = \sum_{i=1}^{rc} \frac{(O - E)^2}{E} = \sum_{i=1}^{rc} \frac{(f_i^O - f_i^E)^2}{f_i^E}$$

where χ^2 is Chi-square statistics, O = observed value, E = expected /theoretical value, r = number of rows in contingency table, c = number of columns in contingency table, f = frequency.

To calculate the Chi-square statistics use the formulas as in the image below:

G11				fx				=(B3-G3)^2/G3+(C3-H3)^2/H3+(B4-G4)^2/G4+(C4-H4)^2/H4			
	A	B	C	D	E	F	G	H	I		
1	Observed contingency table					Expected contingency table					
2		Nebivolol	Placebo	Total			Nebivolol	Placebo	Total		
3	Women	120	95	215		Women	143	72	215		
4	Men	307	119	426		Men	284	142	426		
5	Total	427	214	641		Total	427	214	641		
6											
7						Critical value	3.84				
8						Significance level	0.05				
9						df(degree of freedom)	1				
10						p-value					
11						Chi-square statistics	16.97				

7.

Predefined Excel function that could be used to calculate the p-value associated to Chi-square statistics is CHISQ.TEST where Actual_range represents the observed values and Expected_range represents the expected values:

	A	B	C	D	E	F	G	H	I
1	Observed contingency table					Expected contingency table			
2		Nebivolol	Placebo	Total			Nebivolol	Placebo	Total
3	Women	120	95	215		Women	143	72	215
4	Men	307	119	426		Men	284	142	426
5	Total	427	214	641		Total	427	214	641
6									
7						Critical value	3.84		
8						Significance level	0.05		
9						df(degree of freedom)	1		
10						p-value	C4,G3:H4		
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									

Function Arguments

CHISQ.TEST

Actual_range B3:C4 = {120,95;307,119}

Expected_range G3:H4 = {143,221528861154,71.7784711388...}

= 3.79875E-05

Returns the test for independence: the value from the chi-squared distribution for the statistic and the appropriate degrees of freedom.

Actual_range is the range of data that contains observations to test against expected values.

Formula result = 3.80E-05

[Help on this function](#)

OK Cancel

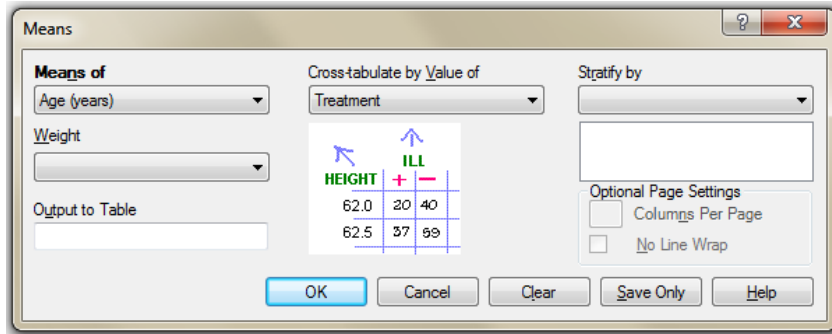
8. Which hypothesis will be accepted?
9. Specify if there is a statistically significant difference between the two groups in regards of gender:
10. Explain the reason for chosen decision (Request 9):

For the next 3 exercises apply the requests in Epi Info provide the answer to the questions in **StatisticalTests.docx** file. Save the Nebivolol.xlsx file in the Lab10 and import all data from in Epi Info.

EXERCISE 5

Test whether there is a statistically significant difference between the average age by treatment (nebivolol or placebo).

1. Apply the proper test using the Means option under Statistics in Epi Info:



- Copy the obtained value of the statistical test and associated p-value (method Pooled) in the StatisticalTests.docx under the Exercise 5 heading.

T-Test

	Method	Mean	95% CL Mean	Std Dev
Diff (Group 1 - Group 2)	Pooled	-1.8630	-2.9925 -0.7334	6.8679
Diff (Group 1 - Group 2)	Satterthwaite	-1.8630	-2.9814 -0.7445	

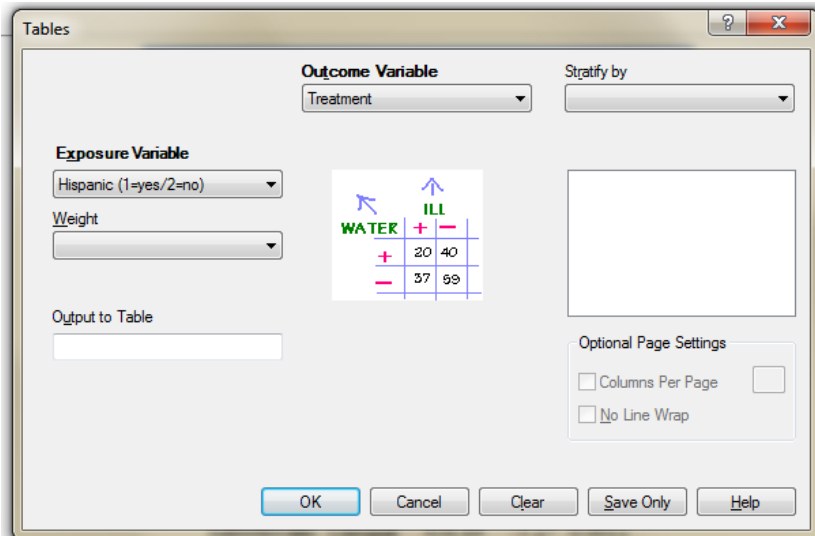
Method	Variances	DF	t Value	Pr > t
Pooled	Equal	639	-3.24	0.0013
Satterthwaite	Unequal	438.89	-3.27	0.0011

- Is there any statistical difference between the two investigated groups? Motivate your answer.

EXERCISE 6

Test whether there is a statistically significant difference in terms of ethnicity (Hispanic vs. Non-hispanic) by treatment (nebivolol or placebo).

- Apply the proper test using the Tables option under Statistics in Epi Info:



- Write in the **StatisticalTests.docx** file under the Exercise 6 heading if there is any value in the contingency table smaller than 5 (yes/no).
- Specify the name of the test used to answer this question (Chi-square test if all values in the contingency table are higher than 5, otherwise it will be use the Fisher exact test).
- Write in the **StatisticalTests.docx** file the value of Chi-square statistics (Chi-square - uncorrected):

Single Table Analysis

	Point	95% Confidence Interval	
	Estimate	Lower	Upper
PARAMETERS: Odds-based			
Odds Ratio (cross product)	0.8633	0.6080	1.2257 (T)
Odds Ratio (MLE)	0.8635	0.6085	1.2289 (M)
		0.5998	1.2473 (F)
PARAMETERS: Risk-based			
Risk Ratio (RR)	0.9513	0.8426	1.0740 (T)
Risk Difference (RD%)	-3.2973	-11.2204	4.6257 (T)
(T=Taylor series; C=Cornfield; M=Mid-P; F=Fisher Exact)			
STATISTICAL TESTS	Chi-square	1-tailed p	2-tailed p
Chi-square - uncorrected	0.6763		0.4108558805
Chi-square - Mantel-Haenszel	0.6753		0.4112210977
Chi-square - corrected (Yates)	0.5362		0.4640182156
Mid-p exact		0.2060493231	
Fisher exact		0.2315111221	0.4183802971

5. Knowing that the critical value is equal with 3.84, indicate whether there is a statistically significant difference using the obtained Chi-square value. Please explain why.

EXERCISE 7

Test whether there is a statistically significant difference in the occurrence of peripheral edema (swollen feet) by treatment (nebivolol or placebo). Provide the answers for the following questions in the **StatisticalTests.docx** file under the Exercise 7 heading.

1. Indicate whether there is in the contingency table at least one cell with value smaller than 5 (yes/no):
2. Provide the name of the statistical test proper to be used in this case.
3. Write the p-value:
4. Please indicate whether there is a statistically significant difference between the two groups according to peripheral edema.