Biometry practical 7

Illustrated (imperfect) practical guide

Preparatory work

1. Open in MS Excel the questionary data (file analysed already in previous practicals),
2. insert new worksheet, rename it to ’Praks7’ (or ’Practical 7’) and
3. make a copy of the data table (from worksheet ’Andmed’/’Data’) and paste it into the upper left corner of the new worksheet ’Praks7’.

Exercise.

Does the ownership of a car depend on gender (is the proportion of students with car different among male and female students)?

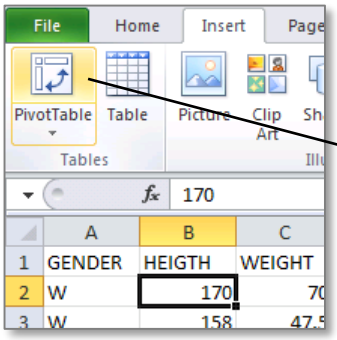
1. Create a two-way frequency table of variables 'GENDER' and 'CAR '.
2. Add into the table also relative frequencies (both, column and row percentages).
3. Comment the table using both row and column percentages.
4. To test the statistical significance of relationship between car ownership and gender, write down the hypothesis pair.
5. Create a new two-way frequency table of variables 'GENDER' and 'CAR', containing only absolute frequencies;
6. calculate theoretical (expected under null hypothesis) frequencies and
7. perform chi-square test (calculate a p-value).
8. Make a final conclusion (Is the relationship statistically significant? Why do you think so? What is the nature of the relationship?).

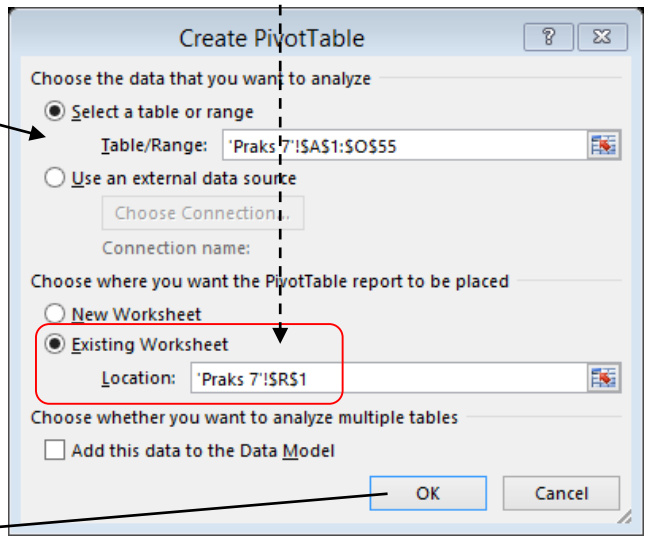
Guide

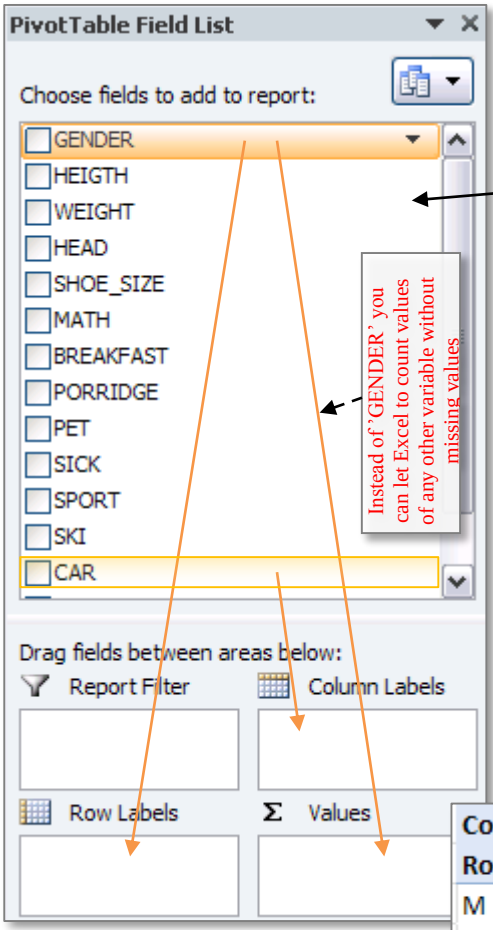
1. Create a two-way frequency table of variables 'GENDER' and 'CAR'.

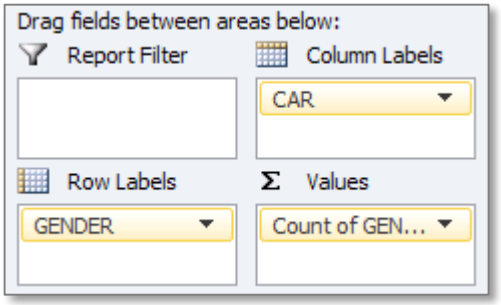
- Put the cursor into arbitrary cell in data table (in worksheet ’Practical 7’).
- *Insert*-tab → *PivotTable*

Place the *PivotTable* into the same ’Practical 7’ worksheet:









Result:

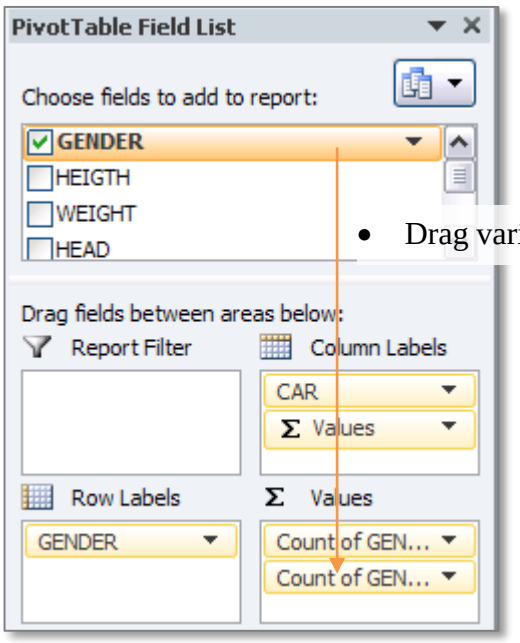
Count of GENDER	Column Lab			
Row Labels	no	yes	(blank)	Grand Total
M	2	8		10
W	20	23	1	44
Grand Total	22	31	1	54

- Omit the student,
  who does not know has she/he a car or not, from the future analyses.

Result:

Count of GENDER	Column La		
Row Labels	no	yes	Grand Total
M	2	8	10
W	20	23	43
Grand Total	22	31	53

2. Add into the table also relative frequencies (both, column and row percentages).

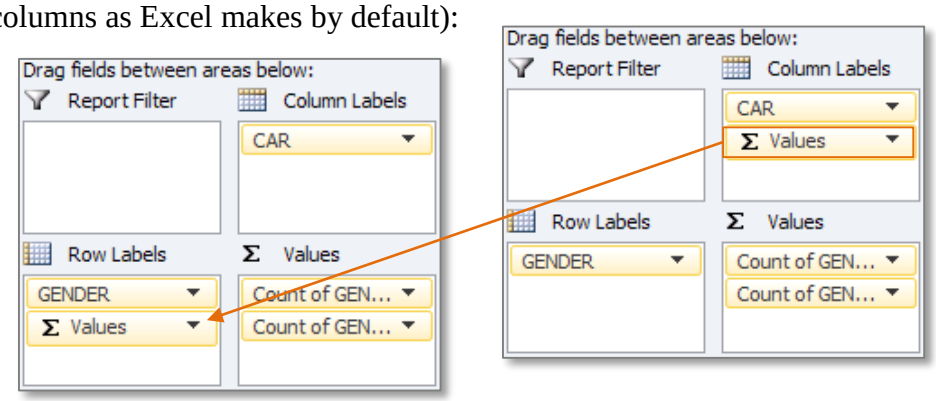


- Drag variable 'GENDER' once again into *Values*-cell.

Result:

	Column Labels					
	no		yes		Total Count of GENDER	Total Count of GENDER2
Row Labels	Count of GENDER	Count of GENDER2	Count of GENDER	Count of GENDER2		
M	2	2	8	8	10	10
W	20	20	23	23	43	43
Grand Total	**22**	**22**	**31**	**31**	**53**	**53**

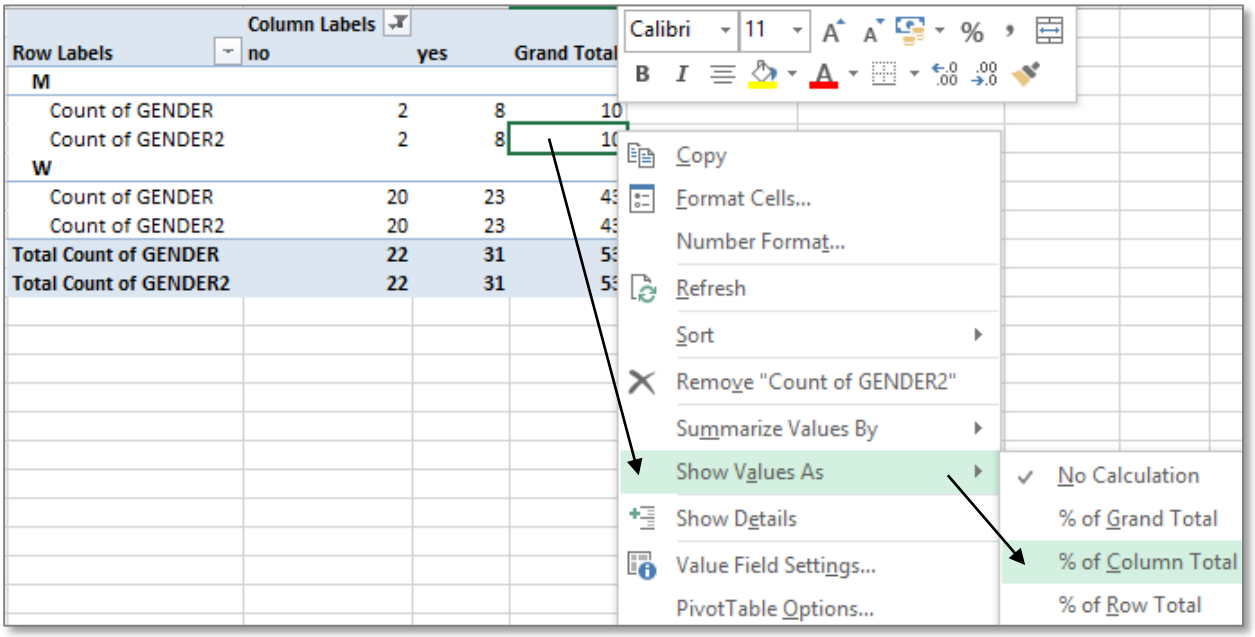
- For more clear presentation put different calculated values into different rows (not into different columns as Excel makes by default):



Result:

	Column		
Row Labels	no	yes	Grand Total
M			
Count of GENDER	2	8	10
Count of GENDER2	2	8	10
W			
Count of GENDER	20	23	43
Count of GENDER2	20	23	43
Total Count of GENDER	**22**	**31**	**53**
Total Count of GENDER2	**22**	**31**	**53**

- To present the second count of students as column percentage:



Result:

Row Labels	no	yes	Grand Total
M			
Count of GENDER	2	8	10
Count of GENDER2	9.09%	25.81%	18.87%
W			
Count of GENDER	20	23	43
Count of GENDER2	90.91%	74.19%	81.13%
Total Count of GENDER	**22**	**31**	**53**
Total Count of GENDER2	**100.00%**	**100.00%**	**100.00%**

- Calculate analogically the row percentages … (% of Row Total).

(start as described in previous page under point 2)

Row Labels	no	yes	Grand Total
M			
Count of GENDER	2	8	10
Count of GENDER2	9.09%	25.81%	18.87%
Count of GENDER3	20.00%	80.00%	100.00%
W			
Count of GENDER	20	23	43
Count of GENDER2	90.91%	74.19%	81.13%
Count of GENDER3	46.51%	53.49%	100.00%
Total Count of GENDER	**22**	**31**	**53**
Total Count of GENDER2	**100.00%**	**100.00%**	**100.00%**
Total Count of GENDER3	**41.51%**	**58.49%**	**100.00%**

3. Comment the table

(write down conclusions using at least one row and one column frequency)**!**

4. Write down the **hypothesis pair** (to be clear, what do you want to test).

5. Create a new two-way frequency table for variables 'GENDER' and 'CAR', containing **only absolute frequencies**.

	Column		
Row Labels	**no**	**yes**	**Grand Total**
M			
Count of GE	2	8	10
Count of GE	9.09%	25.81%	18.87%
Count of GE	20.00%	80.00%	100.00%
W			
Count of GE	20	23	43
Count of GE	90.91%	74.19%	81.13%
Count of GE	46.51%	53.49%	100.00%
Total Count of G	**22**	**31**	**53**
Total Count of G	**100.00%**	**100.00%**	**100.00%**
Total Count of G	**41.51%**	**58.49%**	**100.00%**

Comments .....

H_0: Ownership of a car does not depend on gender.

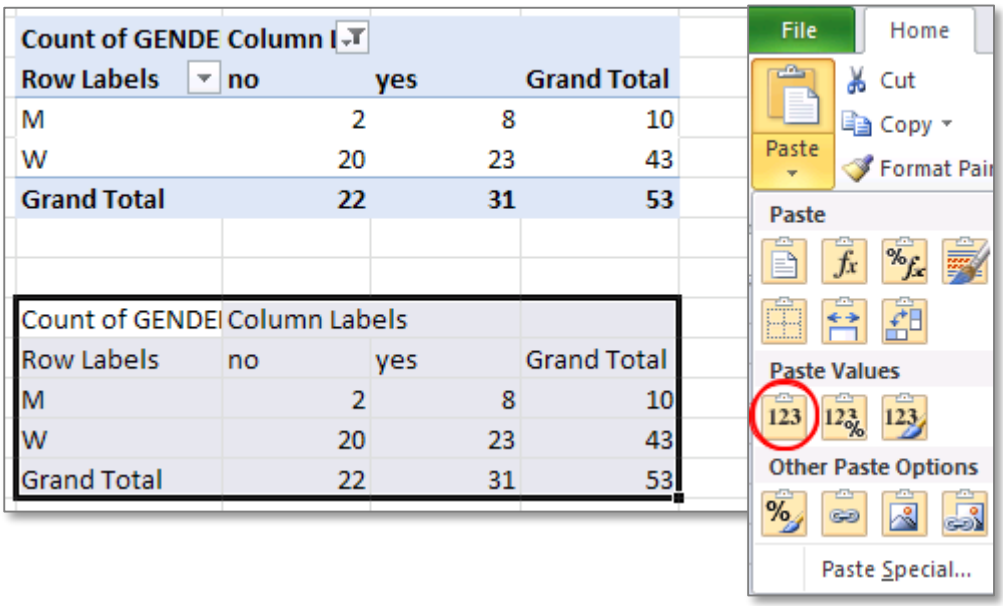
H_1: Ownership of a car depends on gender.

Count of GENDE	Column		
Row Labels	**no**	**yes**	**Grand Total**
M	2	8	10
W	20	23	43
Grand Total	**22**	**31**	**53**

6. Create a new frequency table analogous to the previous table but containing instead of observed frequencies theoretical frequencies.

How? Follow the next guide!

- Make a copy of the structure and **values** of *PivotTable* result:

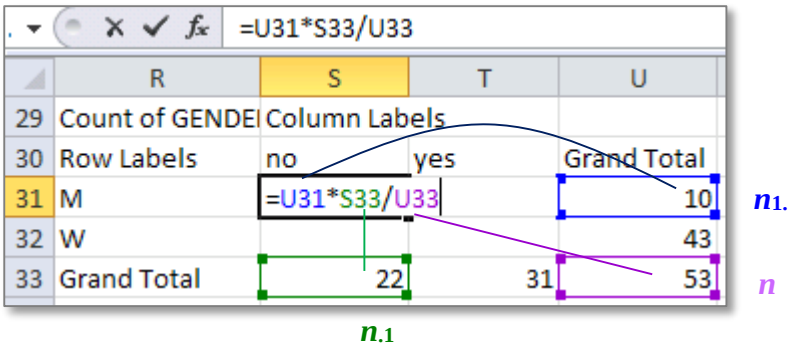


Count of GENDE	Column Labels		
Row Labels	no	yes	Grand Total
M	2	8	10
W	20	23	43
Grand Total	22	31	53

- delete the **content** of copied table (**except row and column totals**)

Count of GENDE	Column Labels		
Row Labels	no	yes	Grand Total
M			10
W			43
Grand Total	22	31	53

- calculate **theoretical**
(expected under null hypothesis) **frequencies** following the formula $n_{ij} = n_{i.} \times n_{.j} / n$.



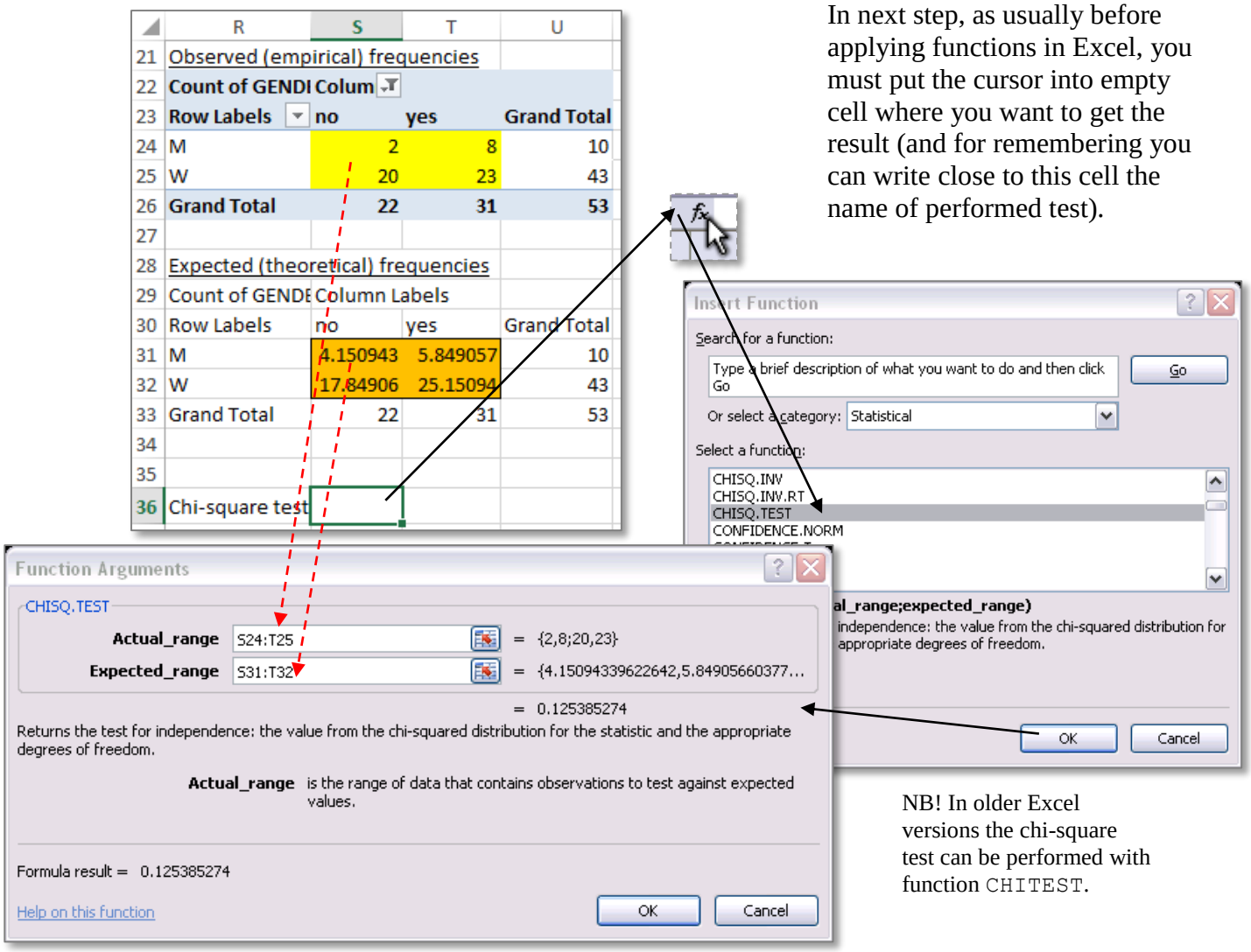
……

Count of GENDE	Column Labels		
Row Labels	no	yes	Grand Total
M	4.1509434	25.1509434	10
W	17.849057	=U32*T33/U33	43
Grand Total	22	31	53

$n_{2.}$ n $n_{.2}$

7. Perform chi-square test (calculate *p*-value) – function CHISQ.TEST.

For better understanding, what the chi-square test is comparing and which data ranges to specify for function CHISQ.TEST, you can colour the compared frequencies and write over the tables, which frequencies they contain.



In next step, as usually before applying functions in Excel, you must put the cursor into empty cell where you want to get the result (and for remembering you can write close to this cell the name of performed test).

NB! In older Excel versions the chi-square test can be performed with function `CHITEST`.

Result:

H_0: Ownership of a car does not depend on gender.

H_1: Ownership of a car depends on gender.

Observed (empirical) frequencies

Count of GENDI Colum			
Row Labels	**no**	**yes**	**Grand Total**
M	2	8	10
W	20	23	43
Grand Total	**22**	**31**	**53**

Expected (theoretical) frequencies

Count of GENDE	Column Labels		
Row Labels	no	yes	Grand Total
M	4.150943	5.849057	10
W	17.84906	25.15094	43
Grand Total	22	31	53

Chi-square test	0.125385	= p > 0.05 -> H_0: Ownership of a car does not depend on gender.

8. Final conclusion.

(Is the relationship statistically significant? Why do you think so? What is the nature of the relationship?).

If among male students 80.0% have a car, then among female students only 53.5% have a car. From all students with car 25.8% are men and 74.2% women (NB! this reflects that there are just a lot of more women than men and not this, that women have more cars). However, according to chi-square test the relationship between car ownership and gender is not statistically significant ($p = 0.125 > 0.05$). This means, that considering this dataset as a sample from all first year students of Estonian University of Life Sciences, it can't be concluded that the proportion of car owners is different between male and female first year students (the probability that this general conclusion is wrong is 12.5%).