

This homework assignment is to be presented on exercise session on **January 21, 2020**. Students should tick in TUWEL problems they have solved and are prepared to present their detailed solutions on blackboard. The solved problems should be ticked by **23:30h on January 20, 2020**.

(1) χ^2 -Test for independence (without R)

100 students from major computer science of three Viennese Universities were randomly chosen and asked which lecture of mathematics (a: calculus, b: algebra, c: probability) they enjoyed most. The frequencies are given in the following table:

	Uni A	Uni B	Uni C
calculus	10	5	5
algebra	10	20	10
probability	20	5	

Perform a χ^2 -test to test whether the preference for a lecture is independent from the university, on a 5% significance level. Only use the following table which gives the 95%-quantile q of the χ^2 -distribution with df degrees of freedom.

df	1	2	3	4	5	6	7	8	9
q	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92

(2) χ^2 -Test for independence (with R)

- (a) Solve the previous exercise using R.
- (b) Can you also reject for a significance level of $\alpha = 0.1\%$?
- (c) Double the observed frequencies in each cell and perform the test on the 0.1% level.

(3) χ^2 -statistic

In the situation of the χ^2 -tests (both goodness of fit and test for independence) what happens to the test statistic if in each cell the observed frequencies are doubled?

(4) Bachelor's theses

For each of the three programs (a: computer science, b: mathematics and c: physics) at a certain university five students that recently handed in their Bachelor's thesis were asked for the number of pages (unified layout) of the thesis. The results were

computer sc.	mathematics	physics
51	62	76
49	59	51
57	48	70
49	74	80
66	60	88

Visualize the data and test the null hypothesis (on a 5%-level) that there is no difference in the mean lengths of the Bachelor's theses between the three programs. Formulate a result.

(5) **Cigarettes - yummy (part 1)**

100 Viennese students were asked of whether they smoke. The possible answers were (1) 'yes, I do' (2) 'no, I stopped smoking', or (3) 'no, I never smoked'. The survey resulted in 22 smokers and 55 that never smoked. Everybody answered.

Test the null hypothesis that the proportion of Viennese students that never smoked is double of those that do smoke and also double of those that stopped smoking. Choose a 1%-significance level. Formulate a result.

(6) **Cigarettes - yummy (part 2)**

In the survey of the previous exercise additionally the smokers were asked how many cigarettes they approximately smoke per day and also those that stopped smoking were asked for the number of cigarettes they smoked before they quit.

The data is given in the file `cigarettes.Rdata`. Visualize the data. Then test the null hypothesis that there is no difference in the mean cigarette consumption between the two groups. Use a 1%-significance level. Formulate a result.

(7) **'Probability and Statistics' exercises**

A tutor of 'probability and statistics' exercises claims that every student that is asked to present the solution for a problem on blackboard is randomly sampled (uniformly) with replacement among all students, for each problem. Every sheet contains of seven problems. After the sixth exercise session a motivated student mentions the following:

- (a) 'Every student in class was asked to present exactly once.' What is the probability for that? Do the data support the tutor's statement?
- (b) 'Two students were asked to present half of the exercises each.' What is the probability for that, if the number of students is the same as in (a)?
- (c) What is the probability for the observations in (a) and (b) if the students were sampled without replacement?