Exercise Set 9

Goals

1) Practice test selection.

2) Perform a first ANOVA.

3) Compute tests on large real sets of data.

1 Different tests for different situations [basic]

State and explain which type of test you would use (z,t,Welch,$\chi 2$,ANOVA; one-/two-sided; one-sample,two-sample,differential;...):

a) You have produced two kinds of bioplastic, 5 samples in each batch. For each sample, you test their permeability for water. You want to know if their performance is the same or not.

b) You have measured the roughness of your sapphire substrate with a profilometer, which has a calibrated standard deviation of 1nm. To be more certain about the result, you have repeated the measurement 7 times. You only use substrates to grow superconductors which have a roughness below 3nm. Can you use your substrate?

c) You measure the structure of $C_6 0$ molecules using x-ray scattering, before and after illumination with a strong mid-infrared laser pulse. The measurement is repeated 8000 times and you have all the data. You want to know if laser illumination changes the interatomic distance in the molecule.

d) Your friend claims that all politicians use the word "illegal" the same amount in their interviews. You want to see if this statement can be statistically refuted. Using automated analysis of the interviews of 100 politicians, which accessed between 30 and 40 interviews for each, you get a "illegal-word-use per interview" data set.

e) You have developed a new method to produce glass. To test its quality, you analyse the number of microcracks in one square centimeter for 100 samples. Your first guess is that the number of cracks follows a normal distribution. How can you verify this guess.

2 Three pizzaiolos [basic]

Three brothers run a pizzeria in Lausanne, and they are in constant argument who makes the best pizza. At one point, they decide to do a pizza baking competition. They all make their best pizza Napoli and serve it to 15 random people who should rank the pizza taste on a scale of 0-100. The results are listed here. Is there any statistical evidence that their pizzas do not all taste the same on the 5% level of significance? What can you (not) conclude from the result?

Alberto	51	45	33	45	67
Francesco	23	43	23	43	45
Paolo	56	76	74	87	56

3 Radiation in Switzerland [Computational, normal]

All Swiss nuclear power plants publish the measured radiation levels around them.

"Radiation_KKW_Goesgen.csv" list the daily measurements of the radiation dose over three years at the Goesgen plant. The units are in μSh, (very low levels of radiation). To be safe, you set a limit of a the mean daily dose around a reactor to be less than 0.1μSh. Can you demonstrate at a level of significance of $\alpha = 0.001$ that the Goesgen plant fulfils that criterion?

4 Is the stockmarket a random process? [Computational, advanced]

In the file "Trading_Data_DAX.csv", you find the prices of the German stock index DAX. It contains the real daily high, low, opening and closing prices for three quarters of a year. In your financial advisory company, you want to test the hypothesis: Are the stock prices maybe just random, following a normal distribution?

- Find the most likely mean and variance of the normal distribution function $N(\mu, \sigma^2)$ that best describes the data. Estimate these parameters from the dataset. You can chose any of the columns for your analysis (high, low, start, close). In fact, the result does not depend on it.

- Form a strategy how you could investigate this question.

- Test the hypothesis at a level of significance $\alpha = 0.05$. Is the German stock markets index just normal-random?