

In the exam you may have your pens and pencils, a ruler and an eraser. On top of that you may have one A4 of notes. The rules for the note are: size A4, text on one side only, it must be hand-written, your name has to be on the top right corner of the note. Other materials, such as formulae books, calculators and tables, you may not have in the exam.

1. (visualization, 6 p.)

Explain how would you visualize the following data sets.

- (a) The eye colors of the students attending Introduction to Statistical Inference -course. (Draw an example.) (2 p.)
- (b) The heights of the male students attending Introduction to Statistical Inference -course. (Draw an example.) (2 p.)
- (c) The proportions of seats by political parties in the Finnish parliament. (Draw an example.) (2 p.)

Justify your answers.

2. (correlation and dependence, 6 p.)

Consider the following bivariate random sample:

$(3.1, 9.6), (-1.2, 1.5), (0.2, 0.1), (-1.8, 3.4), (0.8, 0.6), (-3.1, 9.5), (2.0, 4.1)$.

- (a) Draw a scatter plot of the data. (1.5 p.)
- (b) Estimate the Pearson sample correlation coefficient from the plot. (You do not have to calculate the value of the coefficient — rough numerical approximation is enough.) (1.5 p.)
- (c) Estimate the Spearman sample correlation coefficient from the plot. (You do not have to calculate the value of the coefficient — rough numerical approximation is enough.) (1.5 p.)
- (d) Describe the dependence structure between the variables. (1.5 p.)

3. (Wilcoxon one sample signed rank test, 6 p.)

(a) Give the general statistical assumptions needed for applying Wilcoxon one sample signed rank test. (2 p.)

(b) Give the null hypothesis and the two sided alternative hypothesis of the Wilcoxon one sample signed rank test. (4 p.)

4. (confidence intervals 6 p.)

(a) Let x_1, x_2, \dots, x_n be independent and identically distributed (i.i.d.) observations from a distribution F_x . Let θ be a parameter of the distribution F_x . Let $\hat{\theta}$ be an estimate of the parameter θ calculated using the sample x_1, x_2, \dots, x_n . Explain how to construct an estimate of the 95% confidence interval for $\hat{\theta}$ by using bootstrapping. (4 p.)

(b) When and why is it better to use bootstrap confidence intervals than exact confidence intervals? (1 p.)

(c) When and why is it better to use exact confidence intervals than bootstrap confidence intervals? (1 p.)