

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.)
 - (a) What is the difference between a histogram and a bar plot? (Draw examples.) (2 p.)
 - (b) Give an example of a data set that you would visualize using a bar plot. Explain (using 1-2 sentences) why bar plot is a good choice for your example data. (2 p.)
 - (c) Give an example of a data set that you would visualize using a histogram. Explain (using 1-2 sentences) why histogram is a good choice for your example data. (2 p.)
2. (dependence, 6 p.) Consider independent and identically distributed (i.i.d.) bivariate observations $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$.
 - (a) What type of dependence can be measured using Pearson correlation coefficient? (2 p.)
 - (b) What type of dependence can be measured using Spearman correlation coefficient? (2 p.)
 - (c) Give an example of a dependence type that can neither be detected using Pearson correlation coefficient nor using Spearman correlation coefficient, but that can be detected from a scatter plot. (2 p.)
3. (one sample sign test, 6 p.)
 - (a) Give the general statistical assumptions needed for applying one sample sign test. (2 p.)
 - (b) Give the null hypothesis and the two sided alternative hypothesis of the one sample sign test. (4 p.)

4. (multivariate linear regression, 6 p.)

Consider multivariate linear regression model

$$y_i = b_0 + B^T x_i + \varepsilon_i, \quad i \in 1, \dots, n,$$

where the elements of 2×1 vector b_0 and 4×2 regression matrix B are unknown constants and the expected value of the residuals ε_i is $E[\varepsilon_i] = 0$.

- (a) What does variance inflation factor (VIF) measure? (1 p.)
- (b) Give the definition of VIF for the explanatory variable $(x_i)_3$ and explain how it is calculated. (3 p.)
- (c) Explain how VIF can be used in selecting explanatory variables. (2 p.)