

Foundations of Statistical Modeling

Prof. Dr. Stefan Kettemann

Spring term 2024

Exercise sheet 5, submit by Wednesday April 17th th, 2024 on

TEAMS

Your name: _____



1. Find and Download a Data Set [5 Points]

See the project guidelines on TEAMS for hints how to find and choose interesting data sets. The data set should have at least 2 features. It does not have to be huge, but should contain at least several hundred data points. Describe the raw dataset you have chosen for the project in mathematical correct formalism, define the universe in which the data has been taken, the RV functions, and the data value spaces S . Describe whether you had to clean the data, whether there are any missing values. How did you deal with missing values? Show excerpts of the raw data.

2. Distribution, Moments [5 Points]

Choose from this data set 1 numerical features where you have at least 100 measurements with results in a discrete DVS S (if it is a continuous S , choose intervals with some bin width d , to map it on a discrete S).

- Plot the corresponding histograms and the pmf of this discrete distribution.
- Calculate the mean value, the standard deviation, the skewness and the kurtosis of this distribution.
- Check all kinds of distributions which we reviewed in the lecture. To which kind of distribution function does the pmf of the feature you choose resemble most closely? Find the hyperparameters of that distribution function which minimise the MSE.