University of Toronto

Department of Computer & Mathematical Sciences

STAB57: an Introduction to Statistics

Assignment Nr 3

taught by Louis de Thanhoffer de Volcsey

-email me

-website

-textbook

This week's list of problems is based on the material from:
Chapter 5 and Chapter 6, §1
You are expected to work on this list of problems prior to the upcoming tutorial.
Problems have the following tags:
: difficult, : Book exercise, : extra exercise

Terminology and Concepts to learn:

- descriptive statistics
- sample mean, quantile, variance, correlation (see problem 2)
- the boxplot
- likelihood function

Problem 1

Practice your skills on descriptive statitistics by doing problems 5.5.1-2-3 (as a note, the word estimates is a synonym for the word sample)

Problem 2

Practice your skills on likelihood functions by doing problems 6.1.1-2-3-4

Problem 3

The correlation of two random variables X and Y with means μ_X, μ_Y and variance σ_X^2, σ_Y^2 is given by

$$corr(X,Y) = \frac{\mathbb{E}(X-\mu_X)\cdot\mathbb{E}(Y-\mu_Y)}{\sigma_1\sigma_2}$$

Let $\Delta \subset S$ be a dataset, $\overline{x}$ and $\overline{y}$ be the sample means for the dataset and s_x^2, s_y^2 the sample variances. Show that the sample correlation (also called Pearson correlation) is given by:

$$\frac{\sum(x_1 - \overline{x})\sum(y_j - \overline{y})}{(n-1)s_x s_y}$$

Problem 4

In class, I showed that the sample p-th quantile for a dataset $x_1 \leq x_2 \ldots \leq x_n$ is given by x_i, where $i = \lceil np \rceil$. The book defines it as x_i where i is the index such that

$$\frac{i-1}{n} \leq p \leq \frac{i}{n}$$

Argue why these two definitions coincide.