EXERCISE SET 1,
MS-A0503, FIRST COURSE IN PROBABILITY AND STATISTICS

About the exercises

The weekly exercise sheets will have two parts: explorative problems and homework exercises. This first week, however, there are only explorative problems.

I will expect that you study the explorative problems BEFORE the first lecture of the week. In a manner of speaking, the lectures will contain the "solutions" to the explorative problems, meaning that by solving the problems you have taken important steps towards developing the theory on your own, before I present the theory on the lectures. For the explorative problems, it is very strongly recommended that you work on them in groups. It is often not difficult to come up with a solution to these problems (especially not if one looks in the textbook), but the difficulty lies in comparing your solutions to those of others, and that is also how you will learn *new* ways to think about statistics.

The homework problems are reported during the second exercise session of the week. Also for those exercises, it is a good idea to work in groups, but every student should be prepared to present the solutions individually. During the last exercise session of the week, the teacher will ask you to mark what problems you have solved, and you get points according to how many problems you marked as solved. If you mark a problem as solved, however, you should also be prepared to present your solution in front of the class.

Explorative exercises

Problem 1. In the public debate, it is often said that "one can prove anything with statistics". Discuss this statement; is it true in your opinion? In such case, why do we even bother learning statistics?

Problem 2. By polling a sample of the voting population, we are trying to predict the outcome of the next general election. Which of the following methods of selection is likely to yield a useful sample?

(1) Poll all people of voting age currently sitting in the university library
(2) Poll the first 1000 names from the voter registation list.
(3) Poll 1000 names selected randomly from the voter registation list (with any voter having the same probability of being chosen).
(4) Have a major radio station ask its listeners to call in and name the party they plan to vote for.

Problem 3. There are at least three different notions that describe the "typical" value of a numerical sample (fancy way of saying "list of numbers"). Recall the definitons of the *mode*, *mean*, and *median*. Consult section 2.3.1. in the book if you are unsure.

(1) Consider the sample $S = \{-8, 0, 1, 1, 2, 2, 2\}$. Compute the mean, mode, and median of S.
(2) Construct an example of a sample for which mean $<$ mode $<$ median.
(3) Construct an example of a sample for which mode $<$ mean $<$ median.

Problem 4. The following is an extract of data concerning what percentage of the population in northeastern US take public transportation to work.

State	Percent using public transport
Maine	0.9
New Hampshire	0.7
Vermont	0.7
Massachusetts	8.3
Rhode Island	2.5
Connecticut	3.9
New York	24.8
New Jersey	8.8
Pennsylvania	6.4
Entire North-east	12.8

(1) Convince yourself that both the mean, median, and mode of the upper part of the table are (much) smaller than 12.8.
(2) Explain how the total average can still be 12.8! Has there been a mistake, or is there some other explanation? (Hint: you are allowed but not obliged to use what you know about US demographics.)

Problem 5. The following was overheard on the bus by yours truly (true story): *"Probability theory is just bogus, because when I come home, either my house has burned down or it has not. So if probability theory were right, the probability that my house has burned down would be* 50%*. But that would mean my house would burn down every second day, which it doesn't."* What has the person on the bus misunderstood? Discuss.

Problem 6.

(1) A fair six-sided die is rolled. What is the probability that it comes up 6?
(2) Two fair six-sided dice are rolled. What is the probability that at least one of them will come up 6?
(3) Three fair six-sided dice are rolled. What is the probability that at least one of them will come up 6?

$\vdots$

(n) n fair six-sided dice are rolled. What is the probability that at least one of them will come up 6?