Tutorial 2

1. Relevant material: Unit 2.

   A grain exporting company is trying to understand how much grain they might export next month. Let's denote X the amount of exported grain next month (in tonnes). Based on previous exporting records, the company estimates that the exporting probabilities and amounts are the following (in tonnes):

x	P(x)
100	0.05
150	0.20
200	0.5
250	0.20
300	0.05

   (a) What is the probability that the company does not export more than 250 tonnes?

   (b) What is the expected amount of tonnes that the company will export?

   (c) What is the standard deviation?

   (d) Now assume that the company is also exporting salt. Let's denote Y the amount of exported salt (in tonnes) next month. In contrast to grain, the contracts for salt allow for salt being sold in much smaller quantities, making it a continuous random variable. The variable has a mean (i.e., expected value) of 120 and a standard deviation of 30, and it follows the normal distribution. What is the probability that the amount of exported salt next month will not be bigger than 160 tonnes?

   (e) What is the probability that the amount of exported salt next month will be between 80 and 160 tonnes?

2. Relevant material: Unit 3.

   We are studying a population which consists of all first-year primary school students in a specific neighborhood ($N = 500$). We are interested in studying if students who attended nursery for a longer time have different outcomes when starting school than students who attended nursery for a shorter time.

   We have administrative data form the council on the full population of students. Using this data, we calculate that the population mean of the number of years spent in nursery is $\mu = 3$ and the population standard deviation is $\sigma = 1.6$.

We want to study children's outcomes in specific tasks, but to measure outcomes in these tasks, we need to run additional tests. These tests are expensive and therefore we select only a small random sample from the population to be tested.

We pick a random sample of $n = 47$ students who will go to tests. In the sample that we had chosen, it turns out that the sample mean is relatively high compared to the population mean. Because the sample mean is different from the population mean, we are worried that we might have made a mistake somewhere. We wish to know if this sample mean is something that we can expect with a high or a small probability, given the population values. To do so, answer the following questions.

We denote by $\bar{X}$ the random variable "sample mean" and by $\bar{x}$ its realization.

(a) Explain what do we mean by "the sampling distribution of the sample mean ($\overline{X}$)"?

(b) Using your knowledge about sampling (without making any calculations), characterize the sampling distribution of $\overline{X}$. In other words, what does the sampling distribution look like? Why do we know this?

(c) Calculate the mean of the sampling distribution. Calculate the standard deviation of the sampling distribution.

(d) Based on your answers to parts a-c, describe in words: What values of $\overline{X}$ are likely? Would, for example, a sample mean $\overline{x} = 3.5$ be likely or unlikely?

3. Relevant material: Unit 3. (*Exercise 6.3 Textbook*)

A population contains two million zeros an nine million ones.

(a) What is the sampling distribution of the sample mean, when the sample size is $n = 4$?

Hint: Make a table which contains all the possible values of the sample mean when $n = 4$ and their probabilities.

Hint: Remember that when the data is binary (0/1), there is a link between the sample mean and the sample proportion.

(b) What is the approximate sampling distribution of the sample mean, when the sample size is $n = 50$?

Hint: Do not make a table.

4. **Advanced.** Relevant material: Unit 3. (Not included in exam topics.)

   A small company has six employees, whose years of experience are

$$\{2 \quad 4 \quad 6 \quad 6 \quad 7 \quad 8\}.$$

   Two of these employees are to be chosen randomly for a particular work group.

   Let us consider the average number of years of experience of the two employees $\{X_1, X_2\}$ chosen randomly from the population of six. We denote by $\bar{X}_2$ the random variable sample means and by $\bar{x}_2$ its realization (the subscript 2 refers to sample size).

   Hint: We are in a case where the sample size is large relative to the population and we are think about sampling *without* replacement.

   (a) Compute the population mean μ_X and the population variance σ_X^2.

   (b) Compute the sampling distribution of $\bar{X}_2$.

   Hint: It could be easier to start by constructing a matrix which represents all different possible work groups (different possible samples).

   Hint: the sampling is done without replacement.

   (c) Compute the expectation of $\bar{X}_2$ using the above sampling distribution and compare it with μ_X.

   (d) Compute the variance of $\bar{X}_2$ using the sampling distribution computed above. Nota bene: this means the variance of the sampling distribution of $\bar{X}_2$, not the sample variance!

   (e) Verify that when sampling is done without replacement and the sample size is large relative to the population, then

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

The population variance σ_X^2 is usually unknown. Let us consider the sample variance of the sample $\{X_1, X_2\}$, denoted by s_2^2.

(f) Compute the sampling distribution of s_2^2.

(g) Compute the expectation of s_2^2 using the above sampling distribution and compare it with σ_X^2.

(h) Verify that $\mathbb{E}(s_2^2) = \frac{N}{N-1}\sigma^2$, where $N = 6$.

(*See also Tables 6.1 and 6.2 in the Textbook, and Exercise 6.56*)