***University of Toronto***

***Department of Computer & Mathematical Sciences***

**STAB57: an Introduction to Statistics**

Assignment Nr 2

*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: sections 5.1,5.2,5.3 and 5.5
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:**

- random sample, iid random variables
- the normal distribution, the central limit theorem, approximating distributions (see problem 3)
- the idea of inference, statistical model
- descriptive statistics, sample mean, sample quantile
- injective, surjective and bijective functions

## Problem 1

Practice your skills on descriptive statistics by doing problems 5.5.16, 5.5.17, 5.5.22 (note that empirical distribution is the same as the sample distribution)

## Problem 2

Let $X_1, \dots X_n$ be a random sample with mean $\mu$ and variance $\sigma^2$ and define $M_n = \frac{1}{n}\sum X_i$. Show that

$$\sqrt{n}\left(\frac{M_n - \mu}{\sigma}\right)$$

has mean 0 and variance 1.

## Problem 3

For this problem, you'll need a little more background: for any $n$, let $(X_1, \dots, X_n)$ be a sequence of random samples with mean $\mu$ and variance $\sigma^2$. We wish to approximate the distribution of $\sum_i X$ using the central limit theorem.

- rewrite the central limit theorem to conclude that $\lim_i P_{\frac{S_n - m\mu}{\sqrt{n}\sigma}} \sim N(0,1)$

In other words, if $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}}$, we conclude that

$$\lim_i P\left(\frac{S_n - \mu}{\sqrt{n}\sigma} \leq x\right) = \Phi(x)$$

- use the above observation to argue why it makes sense to approximate

$$P(S_n \leq x) \text{ by } \Psi\left(\frac{x-\mu}{\sqrt{n}\sigma}\right) \text{ and } P(x \leq S_n \leq y) \text{ by } \Psi\left(\frac{y-\mu}{\sqrt{n}\sigma}\right) - \Psi\left(\frac{x-\mu}{\sqrt{n}\sigma}\right)$$

- Work through example 4.4.10 to see how this works
- Finally, do exercises 4.4.5,6 and 7

## Problem 4

Let $X$ be the set of the 143 students in STAB57 . This is a probability space when given the uniform probability.
Let $X \longrightarrow \mathbb{R}$ be the random variable that associates to each student their quiz 1 mark. Assume that the three (sample) quantiles for $P_X$ are respectively $3, 6$ and $8$.

- What are the outliers?

Assume now that we split up $X$ into two groups $X_1$ and $X_2$, giving a probability of $\frac{2}{143}$ to students in $X_1$ and $\frac{1}{286}$ in $X_2$

- How do the quartiles and outliers change?
- Are there any conditions under which the outliers do not change?

## Problem 5

Let $X$ be the set of students in STAB57. Find a set $Y$ and a function $f : X \longrightarrow Y$ which is

- surjective and not injective
- injective and not surjective
- bijective

## Problem 6

Let $f : X \longrightarrow Y$ and $g : Y \longrightarrow Z$ be two functions, recall that $g \circ f : X \longrightarrow Z$ is the function defined by $g \circ f(x) = g(f(x))$. Then show the following statements:

- if both $f$ and $g$ are injective (resp. surjective) then so is $g \circ f$
- if $g \circ f$ is injecitve, then so is $f$, but $g$ doesn't need to be (give a counterexample)
- if $g \circ f$ is surjective then so is $g$, but $f$ doesn't need to be (give a counterexample)