# MAT 2379 [A], Introduction to Biostatistics.

**Course Instructor: Oleksii Volkov**

## Assignment 3 (7 questions)

**Due date: Friday, November 15th, at 3:00pm.**

**Late assignments will not be accepted**

**Name**: ______________________________________________________

**Student ID**: __________________________

You can deposit your completed assignment into the corresponding box on the **second floor** of the **STEM** building. Please make sure you choose the one that says **MAT 2379 [A]** (your section) on it.

**Note:** You can also hand in your completed assignment in class or during the office hour **no later than Wednesday, November 13rd, at 2:30pm.**

- Solve the following exercises using a calculator.
- Late assignments will not be accepted.
- This assignment has 7 questions.
- You will have to answer all questions. It is possible that only part of the exercises will be corrected.
- All exercise numbers refer to the second edition of the book.

Q-1 (7.4).

The illegal traffic of rhinoceros horns is fueled by a huge demand in Asia where it is believed to have cancer-curing properties. The cost of rhino horn is estimated at $65,000 per kilogram on the black market. It is believed that there are no more than 11,000 white rhinos in the wild. In 2014, a record number of 1,020 rhinos have been poached in South Africa, most of them in the Kruger National Park, critically endangering this population. The following data gives the weight (in kg) for a sample of $n = 18$ white rhino horns:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1.1 | 2.9 | 1.3 | 2.1 | 2.3 | 1.3 | 2.2 | 1.6 | 1.5 |
| 2.0 | 3.1 | 1.6 | 2.4 | 1.8 | 2.7 | 3.0 | 0.9 | 2.1 |

(a) Find the mean and the standard deviation for the dataset.
(b) Find the median, the first quartile, the third quartile and the IQR for this dataset.
(c) Are there any outliers in this dataset?

Q-2 (7.6).

An irruption is an irregular migration of birds to a region where they are not usually found. In North America, approximately every four years, there is a winter eruption that causes snowy owls to fly south in larger numbers than usual. Many Ontarians were fortunate to sight a snowy owl in 2014. Suppose that the size of the body of a snowy owl is normally distributed with a mean 61.5 cm and a standard deviation of 4.75 cm.

We select n = 10 snowy owls from the population and compute the average body size (in cm). We consider this average as a point estimate of the population mean of the body size.

(a) Compute the standard error of the mean.
(b) What is the probability that the average body size of these 10 snowy owls will be less than 57cm?
(c) Suppose that we randomly select 5 samples (each of size $n = 10$) from this population. What is the probability that at most one of the 5 samples will have an average body size that is less than 57cm?

Q-3 (7.8).

With R, we produced descriptive statistics for two random samples.

> summary(sample1)

| Min. | 1st Qu. | Median. | Mean. | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| 42.97 | 52.92 | 55.34 | 55.38 | 58.79 | 64.09 |

> summary(sample2)

| Min. | 1st Qu. | Median. | Mean. | 3rd Qu. | Max. |
|---|---|---|---|---|---|
| 51.17 | 52.35 | 55.57 | 55.03 | 56.38 | 60.07 |

(a) Within each sample, are there any outliers? If so, are they below or above the median? Explain.
(b) For the first sample, do we have enough information to construct the corresponding boxplot? If so, construct the boxplot by hand. If not, explain why?
(c) For the second sample, do we have enough information to construct the corresponding boxplot? If so, construct the boxplot by hand. If not, explain why?

Q-4 (7.10).

The concentration of a reactant in a first-order chemical reaction that proceeds at a rate $k$ can be described as follows: $\ln(C) = \ln(C_0) - kt$, where $C$ is the concentration of the reactant at time $t$, $C_0$ is the initial concentration and $t$ is the elapsed time since the reaction started. Consider an initial concentration of $C_0 = 0.3$ mol/L. The experiment was repeated $n$ times to give a geometric mean of the concentration at time $t = 450$ seconds of 0.22 mol/L. The geometric standard deviation of the concentration at time $t = 450$ seconds is 1.17.

(a) Compute the mean of the rate constant $k$.
(b) Compute the standard deviation of the rate constant $k$.

***Hint***: Use the fact that $k$ is a linear function of $y = \ln(C)$.

Q-5 (7.12).

27 women with a diagnosis of inoperable or metastatic breast cancer have been followed-up for several years, while under continuous treatment with a medication called trastuzumab. The time (in months) each patient remained in remission was recorded. Below is the data

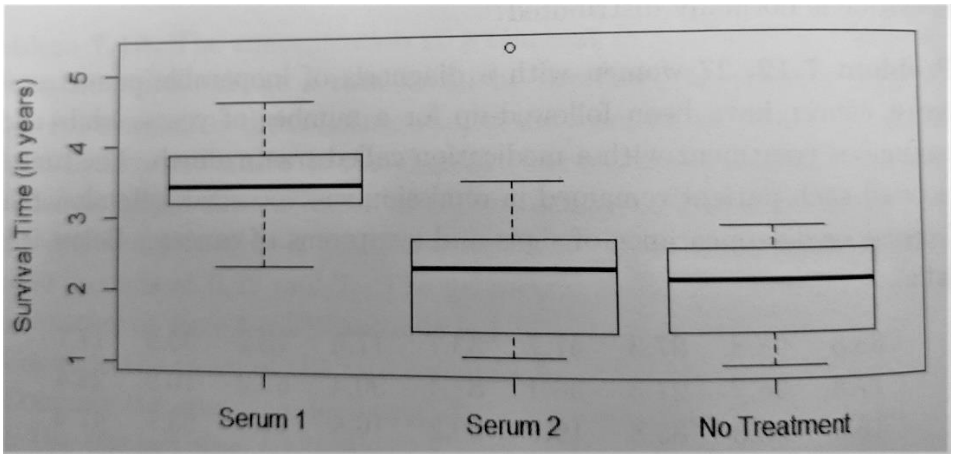
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 88.8 | 98.8 | 97.3 | 47.7 | 33.7 | 17.5 | 73.2 | 35.2 | 11.7 |
| 17.8 | 28.7 | 21.3 | 58.4 | 82.4 | 90.4 | 61.2 | 10.9 | 24.4 |
| 15.4 | 96.6 | 85.8 | 19.7 | 50.2 | 16.8 | 31.3 | 93.7 | 47.4 |

Answer the following questions using R.

(a) Compute the mean, median and the standard deviation of the remission time.
(b) Construct the boxplot of the remission time.
(c) Does the remission time appear to be normally distributed? Justify your answer using QQ-plot.
(d) Compute the mean, median and the standard deviation of the square root of remission time. Is there any relationship between these values and the values found in part (a)?
(e) Compute the geometric mean and the geometric standard deviation of the remission time.

Q-6 (7.14).

Consider 30 mice with an advanced stage of leukemia. Two new serums were developed in the lab to fight leukemia. We randomly divide the 30 mice into 3 groups of 10. The first group received serum 1, the second group received serum 2 and the third group received no treatment. This type of experiment is called a completely randomized design. Consider the following comparative boxplots.




(a) Which group has the largest survival time?
(b) Which group has the largest median survival time?
(c) Which group of survival times is the least dispersed?
(d) Which group has the largest range in survival times?
(e) Are there any groups with similar median survival times?
(f) Which groups of survival times are similarly dispersed?

Q-7 (7.16).

The body mass index (BMI) of a person is defined to be the person's body mass (in kg) divided by the person's height squared (in $\mathrm{m}^2$). Consider a population of males with a mean BMI of 3.2 $\mathrm{kg/m}^2$ and standard deviation of 0.17 $\mathrm{kg/m}^2$. We select $n = 33$ individuals from this population.

(a) Let $\bar{X}$ be the mean BMI for the 33 individuals. Give the expected value and the standard deviation of $\bar{X}$.

(b) Compute $P(\bar{X} > 3.3)$