Unit 3 – Introduction to Nonparametrics
Practice Problems
Due: Thursday September 26, 2024
Last Date to Submit for Credit (-20 points): Thursday October 3, 2024

1. (*Source: Moore, D and McCabe, GP. Introduction to the Practice of Statistics, Third Edition)*

Food products are often enriched with vitamins and other supplements. Does the level of a supplement decline over time so that the user receives less than the manufacturer intended? The following are n=9 observations of vitamin C levels (milligrams per 100 grams) in a wheat soy blend, a flour-like product supplied by international aid programs mainly for feeding children. Each bag was measured twice, first at the factory and a second time five months later in Haiti. Researchers suspect that vitamin C levels are generally higher at the factory than they were five months later. We would like to test the hypotheses:

Null, H_O: vitamin C has the same distribution at both times
Alternative, H_A: vitamin C is systematically higher at the factory (one sided)

Bag	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Factory	45	32	47	40	38	41	37	52	37
Haiti	38	40	35	38	34	35	38	38	40

(a) What is the correct nonparametric test here?

(b) Produce a copy of the table above that shows the differences and the ranks of the *absolute* differences.

(c) Corresponding to your answer to "a", what is the corresponding normal theory test that would have been performed if the assumptions were met?

(d) By any means you like, perform the nonparametric test you gave in "a". In 1-2 sentences at most, report. What do you conclude?

(e) By any means you like (and just for comparison), perform the normal theory test you gave in "c", even knowing that it is not appropriate here.

2. (*Source: Moore, D and McCabe, GP. Introduction to the Practice of Statistics, Third Edition).*

The most used measure of economic growth is the rate of growth in a country's total output of goods and services gauged by the gross domestic product (GDP) adjusted for inflation. The level of a country's GDP growth reflects the growth of businesses, jobs, and personal income. The following are World Bank data on the average growth of GDP (percent per year) for the period 2010 to 2013 in developing countries of Europe:

Developing Countries: Europe

Country	Growth	Country	Growth
Albania	2.3	Macedonia, FYR	2.1
Armenia	4.4	Moldova	5.5
Azerbaijan	3.2	Montenegro	1.7
Belarus	4.0	Romania	1.3
Bosnia and Herzegovina	0.4	Serbia	0.9
Bulgaria	0.9	Turkey	6.0
Georgia	5.6	Ukraine	2.9
Kosovo	3.4		

Developing Countries: Central Asia

Country	Growth	Country	Growth
Uzbekistan	8.2	Kyrgyz Republic	4.0
Turkmenistan	11.3	Kazakhstan	6.5
Tajikistan	7.2		

(a). Suppose we are interested in the similarity of average growth of GDP in the two groups of developing countries: Europe versus Central Asia. State the null and alternative hypotheses.

(b). Produce a copy of the table above that shows the ranks of the 20 observations. Take care in your ranking to handle ties!

(c). What is the correct nonparametric test here?

(d) Consider your answer to "c". What is the corresponding normal theory test that would have been performed if the assumptions were met?

(e) By any means you like, perform the nonparametric test you gave in "c". In 1-2 sentences at most, report. What do you conclude?

(f) By any means you like (and just for comparison), perform the normal theory test you gave in "d", even knowing that it is not appropriate here.

Supplement (NOT part of your homework assignment; NOTHING to turn in)

Learn R

Practice, practice, practice (and maybe enjoy it along the way!

Practice 1 - Vectors.

__1a) Create a vector of 100 elements such that the first 20 elements are 1, 2, …, 20, the next 10 elements are 10, 20, 30, …. 100, and the last 70 elements are 31, 32, …., 100.

TIP #1: You will need to use the function seq( ). To learn more about this, at the console window, type ?seq and then ready the information provided in the HELP tab, which is located together with the tabs "ENVIRONMENT", "FILES", "PLOTS", etc.

TIP #2 If you type seq really slowly in the console, notice that R Studio provides autofill options as you go along. This is a nice way to learn how to work with a new command.

__1b) Display

Practice 2 - Matrices.

__2a) Create the following two matrices

m1

1	2	3	4
5	6	7	8

m2

1	3	5	7
2	4	6	8

__2b) Display

Practice 3 - Probability Distribution Calculations.

__3a) Write the code to calculate Pr [Normal (mean=75, sd=4) $\leq$ 76.1].
Show your result.

__3b) Write the code to calculate Pr [Normal (mean=75, sd=4) $\geq$ 76.1].
Show your result.

__3c) Write the code to calculate Pr [Student t-distribution (degrees of freedom=8) $\geq$ 1.645].
Show your result.

__3d) Write the code to obtain
the value of the 90^{th} percentile of a Chi Square distribution with degrees of freedom = 3.
Show your result