



SCSI/SECI2143: PROBABILITY & STATISTICAL DATA ANALYSIS
2021/2022 – SEMESTER 2

ASSIGNMENT 3

(Total Marks: 80)

1. A study has been conducted to verify the mean of birth weight at Apple County. From the simple random sample collected, the mean was 3.433 kg and the standard deviation of all birth weights is 495 g.
 - a) Using a sample size of 75, construct a 95% confidence interval estimate of the mean birth weight in the Apple County.
 - b) Using a sample size of 75,000, construct a 95% confidence interval estimate of the mean birth weight in the Apple County.
 - c) Which of the preceding confidence intervals is wider? Why?

(8 marks)

2. Abbot Milk Company conduct a study about child growth and the relationship with milk consumption daily. Children that stay in an urban area might have better growth because their parent can afford to buy milk everyday compare to those stayed in rural area. To test the claim, they measured the heights of 36 children from urban area and obtained a mean of 92.8 cm. The population of children has heights with a mean of 91.4 cm and a standard deviation of 3.6 cm (based on annual report). Use a 0.05 significance level to test the claim that children from urban area have a mean height different from 91.4 cm.

(8 marks)

3. Zebra crossing is used by pedestrian to cross the road safely. However, there are few drivers that ignore it (they do not stop even when they saw the pedestrian would like to cross the road). A survey was conducted to identify which drivers (male or female) are more aggressive drivers and are more prone to go through zebra crossing. A CCTV camera was use to record the behavior at various times of the day and monitored a random sample of drivers. The results are as follows: -

	Drivers	No of drivers going through zebra crossing
Male	71	51
Female	53	31

Conduct a hypothesis test at the 5% level of significant to test the claim that male drivers are more prone to go through zebra crossing compare to female drivers.

(8 marks)

4. MamaSaby wanted to investigate the claim made by the consumer regarding the different in weight (gram) between the NutyFruty and Biscoff flavors cupcake. A sample of 9 NutyFruty flavor and 13 of Biscoff flavor cupcake are collected and their weight are recorded as shown in Table 1. Is there any difference between the mean of the weight? Use $\alpha = 0.05$ and assume the variances are equal.

Table 1

No. of observation	NutyFruty flavor (X)	Biscoff flavor (Y)
1	21.7	21.5
2	21.0	20.5
3	21.2	20.3
4	20.7	21.6
5	20.4	21.7
6	21.9	21.3
7	20.2	23.0
8	21.6	21.3
9	20.6	18.9
10		20.0
11		20.4
12		20.8
13		20.3

(10 marks)

5. There is a claim that playing e-sport increase food intake. The following table (Table 2) shows data on food intake (in Kcal) for 15 men on the 2 hours playing e-sport each night and for 15 men on the 4 hours playing e-sport each night. Conduct a hypothesis test with $\alpha = 0.05$ to determine there is a significant difference in mean food intake for two different group.

Table 2

2-hour playing e-sport	3586	4470	3068	5338	2221
	4791	4435	3099	3187	3901
	3868	3869	4878	3632	4518
4-hour playing e-sport	4965	3918	1987	4993	5220
	3653	3510	3338	4100	5792
	4547	3319	3336	4304	4057

(10 marks)

6. An EduConsult company claim that it new teaching method possible to increase memory recall for slow learners. Eight students from rehabilitation class were selected at random and given 10 minutes to memorize a list of 20 nonsense words. Each was asked to list as many of words as he or she could remember in the next 24 hour later. The similar group of students also trained with the new teaching method for a month and the test was repeated. Is there evidence to suggest that the mean number of words recalled after 1 hour exceeds the mean number of words recall after 24 hours by more than 3. Use 0.1 of significant level.

subject	1	2	3	4	5	6	7	8
Before treatment	14	12	18	7	11	9	16	15
After treatment	10	4	14	6	9	6	12	12

(10 marks)

7. A study has been conducted on tar level of cigarettes smoked for a sample of male smokers who subsequently died of lung cancer. The data are shown below.

Tar level	Frequency
0 – 7 mg	103
8 – 14 mg	378
15 – 21 mg	563
≥ 22 mg	150

Assume it is reasonable to regard the sample as representative of male smokers who die of lung cancer. Is there convincing evidence that the proportion of male smoker lung cancer deaths is not the same for the four given tar level categories? Let $\alpha = 0.05$.

(8 marks)

8. In a food laboratory, scientist use mango to test the theory that claims the frequencies of the four categories “yellow and sweet”, “green and sweet”, “yellow and juicy”, and “yellow and sour” will occur in the ratio 9:3:3:1. If a scientist counted 124, 30, 43, and 11, respectively, for these four categories, would these data support the theory? Let $\alpha = 0.01$.

(8 narks)

9. Two groups of workers have the same qualifications for a particular type of work. Their expertise in salaries is summarized by the following 2x5 contingency table.

Group	Salary (Thousands of Dollars)				
	27-29	29-31	31-33	33-35	35 and over
1	6	11	16	14	13
2	5	9	8	6	2

Test whether the group assignment and the salaries seem to be independent with these data at the $\alpha = 0.1$ significance level.

(10 marks)

~~~~~

**Deadline of submission: 21<sup>th</sup> June 2022**