SECI 2143-04 PROBABILITY & STATISTICAL DATA ANALYSIS – SEMESTER 2

Assignment 3

GROUP MEMBERS	MATRIC NUMBER
AFIF BIN GENARI@AZHARI	A21EC0003
MUHAMMAD ADAM HAIKAL BIN MOHD SHUHADZLEE	A21EC0062
MUHAMMAD FITRI BIN ISMAIL	A21EC0076
MUHAMMAD SYAIF ALFARIZ BIN ILYAS SUSANTO	A21EC0094

Deadline of submission: 23 June 2022

QUESTION 1

No.: Date:

ASSIGNMENT 3

Question 1

$\bar{x} = 3.433$ $\sigma = 0.495$

a) $N = 75$ $CI = 95\%$

$z_{a/2} = 1.96$

Therefore $= 3.433 \pm (1.96)\left(\frac{0.495}{\sqrt{75}}\right)$

$= 3.433 \pm 0.112$

$= (3.321, 3.545)$

We are 95% confident that the mean number of the birth weight in the Apple Country is between 3.321 and 3.545

b) $N = 75\ 000$ $CI = 95\%$

$Z_{a/2} = 1.96$

Therefore $= 3.433 \pm (1.96)\left(\frac{0.495}{\sqrt{75\ 000}}\right)$

$= 3.433 \pm 0.004$

$= (3.429, 3.437)$

We are 95% confident that the mean number of the birth weight in the Apple Country is between 3.429 and 3.437

c) The size of confidence interval depend on three things, one of them is sample size. The larger the sample size, the smaller the confidence interval. This is because the larger the sample, the more sure you can be that the answers truly reflect the population. Thus, with the same confidence level of 95%, the confidence interval in 1 (a) is wider than 1 (b) due to the smaller sample size which is $75 < 75\ 000$

QUESTION 2

No. Date

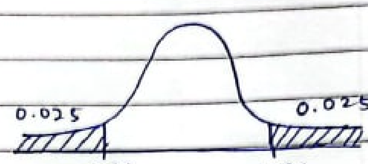
2. $H_0 : \mu = 91.4$ $\alpha = 0.05$ $\bar{x} = 92.8$ $\sigma = 3.6$

$H_1 : \mu \neq 91.4$ $\alpha/2 = 0.025$ $n = 36$

$Z_{\alpha/2} = 1.96$

$$Z = \frac{92.8 - 91.4}{3.6/\sqrt{36}}$$

$Z = 2.33$



$P(Z > 2.33) = 1 - 0.9901 = 0.0099$

$P(Z < 2.33) = 0.0099$

P-value $= 0.0099 + 0.0099 = 0.0198$

Decision = $P_{test\ statistic} = 0.0198 < P_{critical\ value} = 0.05$, thus reject H_0.

Conclusion = There is sufficient evidence that children from urban area have a mean height different from 91.4 cm.

QUESTION 3

Assignment 3 PSDA

Question 3

The claim that male drivers are more prone to go through zebra crossing compare to female drivers.

$H_0 : P_1 = P_2$ Let P_1 = male

$H_1 : P_1 > P_2$ P_2 = female

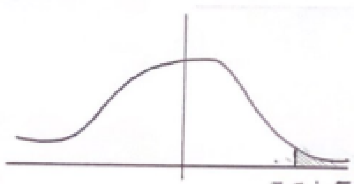
$\alpha = 0.05$

$\hat{P}_1 = \frac{51}{71}$ $\hat{P}_2 = \frac{31}{53}$ $\bar{P} = \frac{51+31}{71+53} = 0.6613$

$\bar{q} = 1 - \bar{p} = 1 - 0.6613 = 0.3387$

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}} = \frac{\left(\frac{51}{71} - \frac{31}{53}\right) - 0}{\sqrt{\frac{(0.6613)(0.3387)}{71} + \frac{(0.6613)(0.3387)}{53}}}$$

$= 1.55$



$z = 1.55$

$P(Z > 1.55) = 0.0606$

Since $0.0606 > 0.05$, fail to reject H_0

There is no sufficient evidence to support the claim that male drivers are more prone to go through zebra crossing compare to female drivers.

QUESTION 4

Question 4

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 \neq \mu_2$

$$\bar{x}_x = \frac{21.7 + 21.0 + 21.2 + 20.7 + 20.4 + 21.9 + 20.2 + 21.6 + 20.6}{9}$$

$= 21.033$

$$S_x = \sqrt{\frac{(21.7-21.033)^2 + (21.0-21.033)^2 + (21.2-21.033)^2 + (20.7-21.033)^2 + (20.4-21.033)^2 + (21.9-21.033)^2 + (20.2-21.033)^2 + (21.6-21.033)^2 + (20.6-21.033)^2}{9}}$$

$= 0.57$

$$\bar{x}_y = \frac{21.5 + 20.5 + 20.3 + 21.6 + 21.7 + 21.3 + 23.0 + 21.3 + 18.9 + 20.0 + 20.4 + 20.8 + 20.3}{13}$$

$= 20.892$

$$S_y = \sqrt{\frac{(21.5-20.892)^2 + (20.5-20.892)^2 + (20.3-20.892)^2 + (21.6-20.892)^2 + (21.7-20.892)^2 + (21.3-20.892)^2 + (23.0-20.892)^2 + (21.3-20.892)^2 + (18.9-20.892)^2 + (20.0-20.892)^2 + (20.4-20.892)^2 + (20.8-20.892)^2 + (20.3-20.892)^2}{13}}$$

$= 1.01$

$$S_p^2 = \frac{(n-1)S^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

$$= \frac{(9-1)(0.57)^2 + (13-1)(1.01)^2}{9+13-2}$$

$= 0.74202$

Question 4

$$v = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{\left[s_1^2/n_1\right]^2}{n_1 - 1} + \frac{\left[s_2^2/n_2\right]^2}{n_2 - 1}}$$

$$= \frac{\left[\frac{(0.57)^2}{9} + \frac{(1.01)^2}{13}\right]^2}{\frac{\left[(0.57)^2/9\right]^2}{8} + \frac{\left[(1.01)^2/13\right]^2}{12}} = 19.42 \approx 19$$

$$t_{0.025,19} = 2.093$$

$$-t_{0.025,19} = -2.093$$

$$t_0 = \frac{\bar{x}_x - \bar{x}_y - 0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{21.033 - 20.892}{0.74202\sqrt{\frac{1}{9} + \frac{1}{13}}} = 0.438$$

Since $0.438 < 2.093$, fail to reject H_0. That is, at the 0.05 level significance, we do not have strong evidence to conclude that different in weight between Nutty Fruty and Biscoff flavors cupcake.

QUESTION 5

No.: Date:

Question 5

Sample 1

$\Sigma x_1 = 58861$

$\bar{x}_1 = 3924.0667$

$s_1 = 829.6389$

$n = 15$

Sample 2

$\Sigma x_2 = 61039$

$\bar{x}_2 = 4069.2667$

$s_2 = 952.8958$

$n = 15$

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$$S_p^2 = \frac{(15-1)(829.6389)^2 + (15-1)(952.8958)^2}{(15+15-2)}$$

$$= 798155.555$$

$$S_p = 893.3955$$

Reject Ho if $t_0 > t_{0.025,14} = 2.145$ or if $t_0 < -t_{0.025,14} = -2.145$

$$t_0 = \frac{3924.0667 - 4069.2667 - 0}{893.3955\sqrt{\frac{1}{15}+\frac{1}{15}}}$$

$$= -0.4451$$

Since $-2.145 < t_0 = -0.4451 < 2.145$, Fail to reject Ho.

That is, at the level 0.05 level of significance we do not have strong evidence to conclude that there is significance difference in mean food intake for two different group.

QUESTION 6

No. Date

6)

Before treatment	After treatment	differences (d)
14	10	4
12	4	8
18	14	4
7	6	1
11	9	2
9	6	3
16	12	4
15	12	3

$$\bar{x}d = \frac{4+8+4+1+2+3+4+3}{8}$$

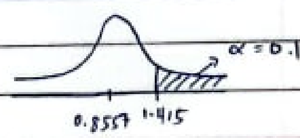
$$\bar{x} = \frac{29}{8} \quad \bar{x} = 3.625$$

$H_0 : \mu d \leq 3$ $\alpha : 0.1$ $df = n-1 = 8-1 = 7$

$H_1 : \mu d > 3$ $n = 8$

$$s^2 = \frac{(4-3.625)^2+(8-3.625)^2+(4-3.625)^2+(1-3.625)^2+(2-3.625)^2+(3-3.625)^2+(4-3.625)^2+(3-3.625)^2}{8-1}$$

$$= \frac{29.875}{7} = 4.26786$$



$s = 2.0659$

$n \leq 30$ t critical value $= 1.415$

$$t_{test\ statistic} = \frac{3.625-3}{2.0659/\sqrt{8}} = 0.8557$$

Decision = Since $0.8557 < 1.415$, thus we fail to reject H_0.

Conclusion = There is not enough evidence to suggest that the mean number of words recalled after 1 hour exceeds after 24 hours by more than 3.

QUESTION 7

Question 7.

Test hypothesis

H_0 : Proportion of male smoker lung cancer death is the same for the four level categories

H_1 : Proportion of male smoker lung cancer death is not same for the four level categories.

- Relative freq : 0.250

Tar level	0-7mg	8-14 mg	15-21 mg	≥ 22 mg	Total
obs. freq	103	378	563	150	1194
exp. freq	298.50	298.50	298.50	298.50	
$(O-E)^2/E$	128.04	21.17	234.37	73.88	

Exp value:

0-7mg, $E = 1194 \times 0.25$
$= 298.50$

8-14mg, $E = 1194 \times 0.25$
$= 298.50$

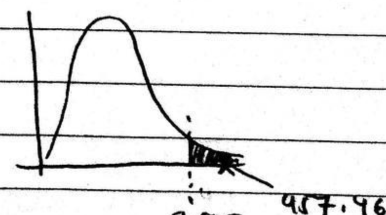
15-21mg, $E = 1194 \times 0.25$
$= 298.50$

≥22mg, $E = 1194 \times 0.25$
$= 298.50$

Test statistic, $\chi^2 = 128.04 + 21.17 + 234.37 + 73.88$
$= 457.46$

Critical value, $\chi^2_{k=3,\ \alpha=0.05} = 7.815$

$df = 4 - 1 = 3$



- Since test statistic > critical value, thus we reject H_0 at $\alpha = 0.05$

- There is evidence that the proportion of male smoker lung cancer deaths is not the same for the four given tar level categories.

QUESTION 8

Question 8

	Yellow & sweet	green & sweet	yellow & juicy	yellow & sour	total	
	124	30	43	11	208	(n sample)
Ratio	9	3	3	1	16	

H_0 : There is goodness of fit between observed and expected frequencies.

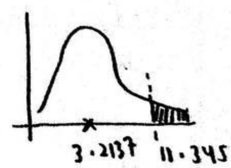
H_1 : There is no goodness of fit between obserd and expected frequencies.

Cell ij	Obs. count, O_{ij}	exp. count, e_{ij}	$[O_{ij} - e_{ij}]^2 / e_{ij}$
1,1	124	$(9)(208)/16 = 117$	0.4188
1,2	30	$(3)(208)/16 = 39$	2.0769
1,3	43	$(3)(208)/16 = 39$	0.4103
1,4	11	$(1)(208)/16 = 13$	0.3077
			$\chi^2 = 3.2137$

Test statistic : $\chi^2 = 3.2137$

Critical value : $\chi^2_{k=3,\ \alpha=0.01} = 11.345$

$df = (2-1)(4-1) = 3$



- Since test statistic < Critical value, thus we fail to reject H_0 at $\alpha = 0.01$.
- There is goodness of fit between observed and expected frequencies, so these data support the theory.

QUESTION 9

No.: Date:

Question 9.

	Salary (Thousands of dollar)					
Group	27-29	29-31	31-33	33-35	35 and over	
1	6	11	16	14	13	60
2	5	9	8	6	2	30
	11	20	24	20	15	90 (n sample).

H_0: Group assignment is independent of the salary

H_A: Group assignment is not independent of the salary.

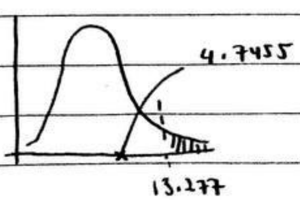
Cell, ij	Obs. count, O_{ij}	Exp. count, e_{ij}	$[o_{ij}-e_{ij}]^2/e_{ij}$
1,1	6	(60)(11)/90 = 7.33	0.2413
1,2	11	(60)(20)/90 = 13.33	0.4073
1,3	16	(60)(24)/90 = 16.00	0.0000
1,4	14	(60)(20)/90 = 13.33	0.0337
1,5	13	(60)(15)/90 = 10.00	0.9000
2,1	5	(30)(11)/90 = 3.67	0.4820
2,2	9	(30)(20)/90 = 6.67	0.8139
2,3	8	(30)(24)/90 = 8.00	0.0000
2,4	6	(30)(20)/90 = 6.67	0.0673
2,5	2	(30)(15)/90 = 5.00	1.8000
			$\chi^2 = 4.7455$

Test statistic: $\chi^2 = 4.7455$

Critical value: $\chi^2_{k=4, \alpha=0.1} = 13.277$

$df = (2-1)(5-1)$

$= 4$



- Since test statistic < critical value, thus we fail to reject H_0 at $\alpha = 0.01$
- There is evidence that group assignment and the salaries are independent.