PROBABILITY & STATISTICAL DATA ANALYSIS

SECI2143

Assignment 3

Lecturer:

Rozilawati Dollah @ Md Zain

Group Members:

No	Name	Matric No
1	AIMAN HAIKAL BIN ZAINUDDIN	A21EC0154
2	ARSYAD WIRATAMA	A20EC0290
3	MUHAMAD AMSYAR BIN IBRAHIM	A21EC0058
4	MUHAMMAD DANIAL WAJDI BIN SAFIAY	A21EC0071

Question 1

Date :

Assignment 3

Question 1

$\mu = 3.433$ kg

$\sigma = 495$ g

a) $n = 75$ $ci = 95\%$

z ~~score~~ score at 95% confidence interval = 1.96

$\tilde{X} \pm z \frac{s}{\sqrt{N}} = 3.433 \pm 1.96 \left(\frac{495}{\sqrt{75}}\right)$

$= 3.433 \pm 112.029$

$= (-108.596, 115.462)$ #

b) $N = 75000$ $Ci = 95\%$

z score = 1.96

$\tilde{X} \pm z \frac{s}{\sqrt{N}} = 3.433 \pm 1.96 \left(\frac{495}{\sqrt{75000}}\right)$

$= 3.433 \pm 3.543$

$= (-0.110, 6.976)$ #

c) Which of the preceding confidence interval is wider?

- Confidence interval with size 75 is wider. A larger sample size and lower variability will result in a tighter confidence interval with a smaller margin of error. A smaller sample size or a higher variability will result in a wider confidence interval with a larger margin of error. The level of confidence also affects the interval width. Increasing the sample size makes the confidence interval narrower. ~~Increasing~~ Decreasing the sample size makes the confidence interval wider.

*enlivo

Question 2

Question 2

$n = 36$, $\bar{x} = 92.8$, $\mu = 91.4$, $\sigma = 3.6$, $\alpha = 0.05$

Test hypothesis : $H_0 : \mu = 91.4$

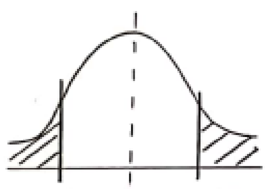
$H_1 : \mu \neq 91.4$

Test statistic :

Since σ is known ~~and n > 30~~, $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

$\therefore z = \dfrac{92.8 - 91.4}{\frac{3.6}{\sqrt{36}}} = 2.33$

Critical value :



-0.025   0.025

$z = -1.96$

$= 0.5 - 0.025$

$= 0.475$

$z = 1.96$

Since it is a two tailed test, $\alpha = \frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

From z-table, $z = 1.96$

Since test statistic value, $z = 2.33$ > critical value, $z = 1.96$, we reject H_0. Thus, there is sufficient evidence that children from urban area have a mean height different from 91.4 cm.

Question 3

Assignment 3

Question 3

$X_m = 51 \qquad X_2 = 31 \qquad \alpha = 0.05$

$N_m = 71 \qquad N_2 = 53$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{51}{71} = 0.7183 \qquad \bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{31}{53} = 0.5849 \qquad = \left(\frac{51+31}{71+53}\right)$$

$$= 0.6613$$

$H_0 :- p_1 = p_2$

$H_1 :- p_1 > p_2$ (right tailed test)

$$z \quad z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.7183 - 0.5849)}{\sqrt{(0.6613)(1-0.6613)\left(\frac{1}{71} + \frac{1}{53}\right)}}$$

$z = 1.55$ (test statistics)

p-value with $z = 1.55$ is p-value $= p(z > 2065)$

$= p(z > 1.55)$

$= 0.0606$ ∴ p value is 0.0606

∴ Since p-value is greater than $\alpha = 0.05$, we failed to reject the null hypothesis.

There is not sufficient evidence to conclude that male drivers are more prono to go through zebra crossing compare to female drivers.

Question 4

4. Nutty Fruty : x_i

$n_1 = 9$

$$\bar{x}_1 = \frac{21.7 + 21.0 + 21.2 + 20.7 + 20.4 + 21.9 + 20.2 + 21.6 + 20.6}{9}$$

$= 21.033$

$$S_x^2 = \frac{\Sigma (x_i - \bar{x})^2}{n_1 - 1}$$

$= \frac{2.93984}{8}$

$= 0.36748$

Biscoff : y_{2i}

$n_2 = 13$

$$\bar{x}_2 = \frac{21.5 + 20.5 + 20.3 + 21.6 + 21.7 + 21.3 + 23.0 + 21.3 + 18.9 + 20.0 + 20.4 + 20.8 + 20.3}{13}$$

$= 20.8923$

$$S_y^2 = \frac{\Sigma (y_i - \bar{y})^2}{n_2 - 1}$$

$= \frac{12.0028826}{12}$

$= 1.0002$

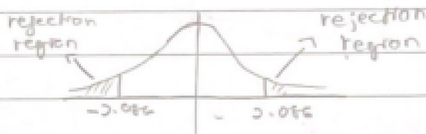
$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{n_1 s_x^2 + n_2 s_y^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{21.033 - 20.8923}{\sqrt{\frac{9(0.36748) + 13(1.0002)}{20}\left(\frac{1}{13} + \frac{1}{9}\right)}}$$

$$= 0.3592$$



critical value = 2.086

$\therefore$ fail to reject $H_0 : \mu_1 = \mu_2$

Question 5

5. 2-hour playing e-sport :

$\Sigma x = 3586 + 4470 + 3068 + 5338 + 2221 + 4791 + 4435 + 3099 + 3187 + 3901 + 3868 + 3869 + 4878 + 3632 + 4518$
$= 58861$

$\Sigma x^2 = (3586)^2 + (4470)^2 + (3068)^2 + (5338)^2 + (2221)^2 + (4791)^2 + (4435)^2 + (3099)^2 + (3187)^2 + (3901)^2 + (3868)^2 + (3869)^2 + (4878)^2 + (3632)^2 + (4518)^2$
$= 240\ 610\ 699$

$n_1 = 15$

$\bar{x}_1 = \frac{\Sigma x}{n_1} = \frac{58861}{15}$
$= 3924.0667$

$\sigma_1 = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}}$

$= \sqrt{\frac{240610699 - \frac{(58861)^2}{15}}{14}}$

$= 829.6389$

4-hour playing e.sport :

$\Sigma x = 4965 + 3918 + 1987 + 4993 + 5220 + 3653 + 3510 + 3388 + 4100 + 5792 + 4547 + 3319 + 3336 + 4304 + 4057$
$= 61039$

$\Sigma x^2 = (4965)^2 + (3918)^2 + (1987)^2 + (4993)^2 + (5220)^2 + (3653)^2 + (3510)^2 + (3388)^2 + (4100)^2 + (5792)^2 + (4547)^2 + (3319)^2 + (3336)^2 + (4304)^2 + (4057)^2$
$= 261\ 096\ 115$

$n_2 = 15$

$\bar{x}_2 = \frac{\Sigma x}{n_2} = \frac{61039}{15}$

$= 4069.2667$

$\sigma_2 = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}}$

$= \sqrt{\frac{261\,096\,115 - \frac{(61039)^2}{15}}{14}}$

$= 952.8958$

Null and alternative hypothesis :

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

Pooled variance : $S^2_p = \frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{(n_1-1)+(n_2-1)}$

$= \frac{(829.6389)^2(14) + (952.8958)^2(14)}{28}$

$= 798155.6381$

Test statistic :

$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{3924.0667 - 4069.2667}{\sqrt{\frac{(829.6389)^2}{15} + \frac{(952.8958)^2}{15}}}$

$= -0.4451$

$df = n_1 + n_2 - 2 = 28$

p-value = 0.6597 , ∴ do not reject null hypothesis as p-value > a, not enough evidence to conclude at 0.05 significance level

Question 6

6.

Before treatment	After treatment	(d_i)	$(d_i)^2$
14	10	4	16
12	4	8	64
18	14	4	16
7	6	1	1
11	9	2	4
9	6	3	9
16	12	4	16
15	12	3	9

$H_0 : \mu d \leq 3$, $n = 8$, $df = 7$, $a = 0.1$

$H_a : \mu d > 3$

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

$= 3.625$

critical value, $t_c = 1.4150$

standard deviation, $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

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

$= 2.066$

standard error, $S.E = s/\sqrt{n}$

$= \frac{2.066}{\sqrt{8}} = 0.7304$

~~tel~~ test statistic, $t = \frac{(\bar{d}-\mu)}{S.E} = \frac{3.625 - 3}{0.7304}$

$= 0.856$

$\therefore t < t_c$, do not reject H_0

p-value $= 0.2102$

$\therefore$ p-value $> a = 0.1$, do not reject H_0 as not enough evidence to support

Question 7

Question 7

Hypotheses:

$H_0: p1 = p2 = p3 = p4 = 0.25$

Ha: Proportion of male smoker lung cancer ~~det~~ deaths is not the same for the four given tar level categories.

Expected frequencies table

O	p	E = p * 1194
103	0.25	298.5
378	0.25	298.5
563	0.25	298.5
150	0.25	298.5
1194	1	1194

Calculation for test statistics

O	E	$\frac{(O-E)^2}{E}$
103	298.5	128.041
378	298.5	21.173
563	298.5	234.373
150	298.5	73.877
1194	1194	457.464

$\alpha = 0.05$

Test statistics:

$$\sum \frac{(O-E)^2}{E} = 457.464$$

Degree of freedom:
$= 4-1 = 3$

$\chi^2_{crit} = 7.815$

Rejection region
If $\chi^2 >= 7.815$, reject H0

↳ Because test statisics in rejection region, reject null hypothesis

P value: P-value = 0
↳ p-value is less than level of significance, reject null hypothesis

Thus there are no evidence to conclude proportion of male smoker lung cancer dead is not the same for the four giver tar level categories.

Masterpints

Question 8

Question 8

Hypothesis

H_0: There is a fit between observes and expected frequencies

H_1: - There is no fit between observed and expected frequencies.

We have given observes frequencies

-> 124. 30, 43. 11

ratio of -> 9 : 3 : 3 : 1

$(124 + 30 + 43 + 11) = 208$

test statistic

Expected values

$124 \rightarrow \frac{9}{16} \times 208 = 117$, $30 \rightarrow \frac{3}{16} \times 208 = 39$

$43 \rightarrow \frac{3}{16} \times 208 = 39$, 11 $\quad 11 \rightarrow \frac{1 \times 208}{16} = 13$

O	E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
124	117	49	0.419
30	39	81	2.077
43	39	16	0.410
11	13	4	0.308
208	208		3.214

- Test statistics: $\chi^2 = 3.214$
- level of significance: $\alpha = 0.01$
- Degrees of freedom $4 - 1 = 3$
- Critical value 11.345

-> Since calculated value of χ^2 is less than critical value of χ^2, accept the null hypothesiss H_0.

Conclusion

There is a fit between expected frequencies and observed frequencies.

~~So these~~

Question 9

Question 9

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

$\alpha = 0.1$

Test hypothesis: H_0: The variables are independent

H_1: The variables are related (dependent)

	Salary (Thousands of Dollars)					
Group	27-29	29-31	31-33	33-35	35 and over	Total
1	Obs = 6 Exp = 7.33	Obs = 11 Exp = 13.33	Obs = 16 Exp = 16.00	Obs = 14 Exp = 13.33	Obs = 13 Exp = 10.00	60
2	Obs = 5 Exp = 3.67	Obs = 9 Exp = 6.67	Obs = 8 Exp = 8.00	Obs = 6 Exp = 6.67	Obs = 2 Exp = 5.00	30
Total	11	20	24	20	15	90

Obs = Observed, Exp = Expected

Test statistic value:

Cell. i,j	Observed Count, o_{ij}	Expected count, e_{ij}	$(o_{ij} - e_{ij})^2 / e_{ij}$
1,1	6	$= \frac{60 \cdot 11}{90} = 7.33$	0.2413
1,2	11	$= \frac{60 \cdot 20}{90} = 13.33$	0.4073
1,3	16	$= \frac{60 \cdot 24}{90} = 16.00$	0.0000
1,4	14	$= \frac{60 \cdot 20}{90} = 13.33$	0.0337
1,5	13	$= \frac{60 \cdot 15}{90} = 10.00$	0.9000
2,1	5	$= \frac{30 \cdot 11}{90} = 3.67$	0.4820
2,2	9	$= \frac{30 \cdot 20}{90} = 6.67$	0.8139
2,3	8	$= \frac{30 \cdot 24}{90} = 8.00$	0.0000
2,4	6	$= \frac{30 \cdot 20}{90} = 6.67$	0.0673
2,5	2	$= \frac{30 \cdot 15}{90} = 5.00$	1.8000

$$\chi^2 = \sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

$= 4.7458$

Critical value : $\chi^2_{k,\alpha}$

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

$= 4$

$\therefore$ From chi square table, $\chi^2_{k=4,\alpha=0.1} = 7.779$

$\therefore$ Since test statistic value < critical value $(4.7455 < 7.779)$, we do not reject H_0 at $\alpha = 0.1$. Thus, there is evidence that group assignment and the salaries are independence at $\alpha = 0.1$ significance level.