SCHOOL OF COMPUTING
Faculty of Engineering

# Assignment 3 (Group)
# SECI2143-07
# Probability & Statistical Data Analysis

SEMESTER II, SESSION 2021/2022

Lecturer: Dr. Nor Azizah Ali

Group Name: Little Group

| Name | Student ID |
|---|---|
| TIA SIAW XUEN | A21EC0233 |
| CHUA XIN LIN | A21EC0020 |
| MA ZE JUN | A21EC4009 |
| KEE LE WEI | A21EC0189 |

Subject: SECI2143 _ PROBABILITY & STATISTICAL DATA ANALYSIS

No:

Date: 21/6/2022

ASSIGNMENT 3

1. $\bar{x} = 3.433\text{kg}$

$\sigma = 495\text{g} = 0.495\text{ kg}$

a) $n = 75$

$1-\alpha = 0.95$, $\therefore \alpha/2 = 0.025$

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 3.433 \pm Z_{0.025} \frac{0.495}{\sqrt{75}}$$

$$= 3.433 \pm 1.96 \frac{0.495}{\sqrt{75}}$$

$$= 3.433 \pm 0.1120$$

$$= (3.321, 3.545)$$

b) $n = 75000$

$1-\alpha = 0.95$, $\therefore \alpha/2 = 0.025$

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 3.433 \pm Z_{0.025} \frac{0.495}{\sqrt{75000}}$$

$$= 3.433 \pm 1.96 \frac{0.495}{\sqrt{75000}}$$

$$= 3.433 \pm 0.0035$$

$$= (3.4295, 3.4365)$$

c) The confidence intervals of a) is more wider. This is because the sample size will affect the width of the confidence intervals. A larger sample size will result in a narrow confidence interval with a smaller margin of error. A smaller sample size will result in a wider confidence interval with larger margin of error.

Subject:

No:

Date: 21/6/2022

2. $H_0 : \mu = 91.4$ cm

$H_1 : \mu \neq 91.4$ cm

$n = 36$

$\bar{x} = 92.8$ cm

$\sigma = 3.6$ cm

$\alpha = 0.05$

$\alpha/2 = 0.025$

critical value $= Z_{0.025} = 1.96$

Test statistic: $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

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

$= \dfrac{1.4}{0.6}$

$= 2.33$

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

$P(z < -2.33) = 0.0099$

P-value $= 2(0.0099) = 0.0198$

Since $|z| = 2.33 > Z_{0.025} = 1.96$, reject $H_0$. There is sufficient evidence to conclude that children from urban area have a mean height different from 91.4 cm.

Subject:

No:
Date: 21/6/2022

3. $H_0: p_1 = p_2$
$H_1: p_1 > p_2$

$\alpha = 0.05$

$\hat{p}_1 = \frac{51}{71}$
$= 0.718$

$\hat{p}_2 = \frac{31}{53}$
$= 0.585$

$\bar{p} = \frac{51+31}{71+53} = \frac{82}{124}$
$= 0.661$

$\bar{q} = 1 - \bar{p} = 1 - 0.661$
$= 0.339$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(0.718 - 0.585) - (0)}{\sqrt{0.661(0.339)\left(\frac{1}{71} + \frac{1}{53}\right)}}$$

$$= \frac{0.133}{0.0859}$$

$$= 1.55$$

P-value $= P(z > 1.55) = 1 - (0.9394) = 0.0606$

Critical value $= Z_\alpha = Z_{0.05} = 1.645$

Since $z = 1.55 < Z_{0.05} = 1.645$, thus we fail to reject the null hypothesis. There is no sufficient evidence to support the claim that male drivers are more prone to go through zebra crossing compare to female drivers.

Q4:

X: Y:

mean: $\bar{x} = 21.0333$ $\bar{Y} = 20.8923$

$s_1 = 0.06218$ $s_2 = 1.00703$

$n_1 = 9$ $n_2 = 13$.

$\alpha = 0.05$, $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$

$df = 9 - 1 = 8$.

standard error: $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{0.36775/9 + 1.00141/13} \approx 0.3447$

Test statistic $= (21.0338 - 20.8923) / \sqrt{0.36755/9 + 1.014107/13} \approx 0.408$

p-value $= 2 \times p(t > |0.408|) = 0.69321$

~~tail~~, $t_c = t_{0.025}, df = \pm 2.306$

rejection two tail to reject the null hypothesis. not evidence to conclude $\mu_1$ is different from $\mu_2$.

5.

for 2-hour

$\Sigma x = 58861$ , $n_1 = 15$

mean = $58861 / 15 = 3924.067$

standard deviation = $\sqrt{240610699 - 58861^2 / 15 - 1} = 829.639$

for: 4-hour

$n_2 = 15$

$\Sigma x = 61039$

mean: $61039 / 15 = 4069.267$

standard deviation = $\sqrt{\dfrac{261099615 - (61039)^2}{(15-1)}} = 952.896$

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

pooled variance:

$s_p^2 = (15-1) \times 829.639 + (15-1) \times 952.896^2) / (15+15-2) = 798155.6$

Test statistic: $t = (\bar{x}_1 - \bar{x}_2) / \sqrt{s_p^2 (\frac{1}{n_1} + \frac{1}{n_2})} = -0.4451$

df: 28.

p-value = 0.6597. > $\alpha$ so not reject the null hypothesis

so, not enough evidence.

年 月 日

No: Date:

6. $n = 8$

$H_0 : \mu_D = 3$

$H_1 : \mu_D > 3$

| subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
|---|---|---|---|---|---|---|---|---|---|
| Before treatment ($x_1$) | 14 | 12 | 18 | 7 | 11 | 9 | 16 | 15 | |
| After treatment ($x_2$) | 10 | 4 | 14 | 6 | 9 | 6 | 12 | 12 | |
| $D = x_1 - x_2$ | 4 | 8 | 4 | 1 | 2 | 3 | 4 | 3 | $\Sigma D = 29$ |
| $D^2 = (x_1 - x_2)^2$ | 16 | 64 | 16 | 1 | 4 | 9 | 16 | 9 | $\Sigma D^2 = 135$ |

$$\bar{D} = \frac{\Sigma D}{n} = \frac{29}{8} = 3.625$$

$$S_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{n}}{n-1}} = \sqrt{\frac{135 - \frac{(29)^2}{8}}{8-1}} = 2.066$$

$$t = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}} = \frac{3.625 - 3}{\frac{2.066}{\sqrt{8}}} = 0.856$$

t-value from table = 1.415

∴ Since $0.856 < 1.415$, hence fail to reject null hypothesis.
There is not enough evidence to support the claim mean number of words after 1 hour exceeds the mean number of words is more than 3.

Step by Step®

No: Date:

7. $H_0$ : Tar level categories are the same
$H_1$ : At least 1 is different

$$E = \frac{n}{k} = \frac{103+378+563+150}{4} = 298.5$$

| Tar level | Frequency | Expected | $(O-E)^2/E$ |
|---|---|---|---|
| 0 – 7 mg | 103 | 298.5 | 128.04 |
| 8 – 14 mg | 378 | 298.5 | 21.17 |
| 15 – 21 mg | 563 | 298.5 | 234.37 |
| ≥ 22 mg | 150 | 298.5 | 73.88 |

$$\chi^2 = \Sigma \frac{(O-E)^2}{E} = 128.04 + 21.17 + 234.37 + 73.88 = 457.46$$

$$\chi^2_{3, 0.05} = 7.815$$

Since 457.46 > 7.815, we reject hypothesis null.
Hence, we claimed that proportion of male smoker lung cancer deaths are not the same for the four given tar level categories.

No: Date:

Question 8

$p_1$ = yellow and sweet
$p_2$ = green and sweet
$p_3$ = yellow and juicy
$p_4$ = yellow and sour

$H_0: p_1 = 9/16, p_2 = 3/16, p_3 = 3/16, p_4 = 1/16$
$H_1$: At least one of the proportions is different from claimed value.

| Frequency | Ratio | p | Expected Frequency | $x^2$ |
|---|---|---|---|---|
| 124 | 9 | 9/16 | $208(9/16) = 117$ | $7^2/117 = 0.4188$ |
| 30 | 3 | 3/16 | $208(3/16) = 39$ | $-9^2/39 = 2.0769$ |
| 43 | 3 | 3/16 | $208(3/16) = 39$ | $4^2/39 = 0.4103$ |
| 11 | 1 | 1/16 | $208(1/16) = 13$ | $-2^2/13 = 0.3077$ |

$x^2 = 3.2137$
$= 3.214$
$x^2_{0.01,3} = 11.345$

$\therefore$ Test statistic value ($x^2 = 3.214$) < critical value ($x^2_{0.01,3} = 11.345$), we do not reject $H_0$. These data support the theory.

No: Date:

## Question 9

$H_0$ = No relationship between variables.
$H_1$ = Variables has relationship.

$\alpha = 0.1$
$df = (2-1)(5-1) = 4$

| | 27-29 | 29-31 | 31-33 | 33-35 | >35 | |
|---|---|---|---|---|---|---|
| G1 | 6 | 11 | 16 | 14 | 13 | 60 |
| G2 | 5 | 9 | 8 | 6 | 2 | 30 |
| | 11 | 20 | 24 | 20 | 15 | 180 |

| Cell$_{ij}$ | $O_{ij}$ | $E_{ij}$ | $x^2$ |
|---|---|---|---|
| 11 | 6 | 3.6667 | 1.4848 |
| 12 | 11 | 6.6667 | 2.8166 |
| 13 | 16 | 8.0000 | 8.0000 |
| 14 | 14 | 6.6667 | 8.0666 |
| 15 | 13 | 5.0000 | 12.8000 |
| 21 | 5 | 1.8333 | 5.4699 |
| 22 | 9 | 3.3333 | 9.6354 |
| 23 | 8 | 4.0000 | 4.0000 |
| 24 | 6 | 3.3333 | 2.1334 |
| 25 | 2 | 2.5000 | 0.1000 |

$x^2 = 54.5067$
$= 54.507$
$x^2_{0.1,4} = 7.779$

$\therefore$ Test statistic value ($x^2 = 54.507$) > critical value ($x^2_{0.1,4} = 7.779$), we reject $H_0$.
There is evidence of a relationship between the variables.