SCHOOL OF COMPUTING
Faculty of Engineering

SESSION 2020/2021 SEMESTER 2

SECI2143 – PROBABILITY & STATISTICAL DATA ANALYSIS

Dr. Nor Azizah Ali

Assignment 3

SECTION: 05

Group Name: Data Miners

List of Members:

Name	Matric No.
1. Omar Mokhtar Bin Yohan	**A20EC0128**
2. Muhammad Sulaiman Daud Syu'aib Bin Yaacob	**A20EC0100**
3. Ibtesham Ahmed Promit	**A20EC4027**
4. Md Mirza Shihab	**A16CS4016**

Question 1

(a)

$n = 25 \quad \bar{x} = 19.5 \quad s = 9.88 \quad \alpha = 0.10$ (90% confidence interval)

$df = 25 - 1 = 24$

$t_{0.05, 24} = 1.711$

$\bar{x} \pm (t\text{-critical value})\left(\frac{s}{\sqrt{n}}\right)$

$= 19.5 \pm 3.381$

$= (16.119, 22.881)$

$\therefore$ At 90% confidence level, the mean time for the population of 100-meter performance is between 16.119 and 22.8809

(b)

$n = 360 \quad \alpha = 0.01$ (99% confidence interval)

$\hat{p} = \frac{160}{360} = 0.444$

z-critical value $= 2.575$

$\hat{p} \pm (z\text{-critical value})\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$= 0.444 \pm 2.575\sqrt{\frac{0.444(1-0.444)}{360}}$

$= 0.444 \pm 0.026$

$= (0.418, 0.470)$

Question 2

(a)

$n = 350 \qquad \alpha = 0.05$

$H_0 : p = 0.1$
$H_1 : p < 0.1$

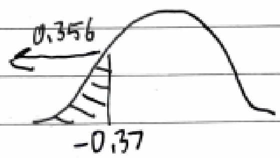
$$\hat{p} = \frac{33}{350} = 0.094$$

$p = 0.1 \quad q = 1 - p = 1 - 0.1 = 0.9$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.094 - 0.1}{\sqrt{\frac{(0.1)(0.9)}{350}}}$$

$$= -0.374$$

$P(z < -0.374) \approx P(z < -0.37) = 0.356$



$P\text{-value} = 0.356$

$\therefore$ Since p-value is greater than significance value $(0.356 > 0.05)$, we fail to reject H_0. There is insufficient evidence to support the claims that the incorrect test result will be less than 10%

Question 2

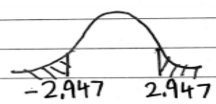
(b) $n = 16 \qquad \bar{x} = 58400 \qquad \sigma = 652 \qquad \alpha = 0.01\ (99\%)$

$H_0 : \mu = 58000$ psi
$H_1 : \mu \neq 58000$ psi

since sample size is 16 $(n \leq 30)$, test statistic is $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

$$t = \frac{58400 - 58000}{652/\sqrt{16}} = 2.454$$

$t_{0.005, 15} = 2.947$



$\therefore$ Since $(2.454 < 2.947)$, we fail to reject H_0. There is sufficient evidence that the true average compressive strength of steel is 58000 psi at 99% significance level.

Question 3

(a) $n = 10$ $\alpha = 0.01$

	x	$(x-\bar{x})^2$
1	10.300	0.058
2	9.900	0.026
3	10.200	0.020
4	10.100	0.002
5	9.700	0.130
6	9.900	0.026
7	9.800	0.068
8	10.300	0.058
9	10.000	0.004
10	10.400	0.116
Σ	100.600	0.508

$\bar{x} = \dfrac{100.6}{10} = 10.060$

$H_0: \mu = 10$ litres

$H_1: \mu \neq 10$ litres

$s = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n-1}}$

$= \sqrt{\dfrac{0.508}{9}}$

$= 0.283$

Test on mean, variance unknown

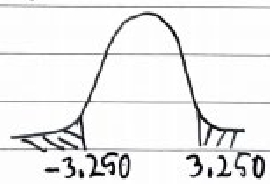
t-test is used since $n \leq 30$

$t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

$= \dfrac{10.06 - 10}{0.238/\sqrt{10}}$

$= 0.797$

$t_{0.005,9} = 3.250$



$\therefore$ Since $(0.797 < 3.250)$, we fail to reject H_0. There is sufficient evidence to prove that average content volume of the Brand x car lubricante is 10 litres at $\alpha = 0.01$

Question 3

(b)

$n = 14 \quad \sigma = 0.44 \quad \alpha = 0.05\,(95\%)$

x	3.7	4.31	3.73	4.33	3.33	2.58	4.47	3.55	4.66	3.68	3.02	4.09	2.36	3.35
$(x-\bar{x})^2$														

$$\bar{x} = \frac{3.7+4.31+3.73+4.33+3.33+2.58+4.47+3.55+4.66+3.68+3.02+4.09+2.36+3.35}{14}$$

$= 3.654$

$$\Sigma(x-\bar{x})^2 = (3.7-3.654)^2+(4.31-3.654)^2+(3.73-3.654)^2+(4.33-3.654)^2+(3.33-3.654)^2+(2.58-3.654)^2+(4.47-3.654)^2+(3.55-3.654)^2+(4.66-3.654)^2+(3.68-3.654)^2+(3.02-3.654)^2+(4.09-3.654)^2+(2.36-3.654)^2+(3.35-3.654)^2$$

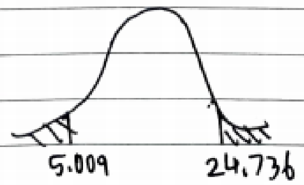
$= 6.202$

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{6.202}{14-1}} = 0.691$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(14-1)(0.691)^2}{(0.44)^2} = 32.062$$

(right) $\chi^2_{0.025,13} = 24.736$

(left) $\chi^2_{0.975,13} = 5.009$



$\therefore$ Since $32.062 > 24.736$, we reject H_0. There is strong evidence to prove that the vitamin supplement has effect on standard deviation of birth weight.

Question 4

$\sigma^2 = 0.18$ $\alpha = 0.05$

$s^2 = 0.165$

$n = 101$

$H_0 : \sigma = 0.18$

$H_1 : \sigma > 0.18$

$$\chi^2 = \frac{(101-1)(0.165)}{(0.18)} = 91.667$$

$$\chi^2_{0.05,100} = 124.342$$

Since $91.667 < 124.342$, we fail to reject H_0.
There is sufficient evidence that support the new joint inspector making more satisfactory measurement.

Question 5

Question 5

$H_0 : \mu_1 = \mu_2$ $\quad n_1 = 15$, $\bar{x}_1 = 76.40$, $s_1^2 = 25.30$

$H_1 : \mu_1 \neq \mu_2$ $\quad n_2 = 18$, $\bar{x}_2 = 71.20$, $s_2^2 = 22.20$

$\alpha = 0.05$

$$t_0 = \frac{76.40 - 71.20}{\sqrt{\frac{25.30}{15} + \frac{22.20}{18}}} = \cancel{8.82}\ 3.04$$

$$v = \frac{(2.92)^2}{\frac{\left(\frac{25.30}{15}\right)^2}{14} + \frac{\left(\frac{22.20}{18}\right)^2}{17}} = 29.13 \approx 29$$

$t_{0.025, 29} = 2.045$

or

$-t_{0.025, 29} = -2.045$

Since $t_0^* = 3.04 > t_{0.025,29} = 2.045$, we reject the null hypothesis. There is evidence to conclude that mean spending on 20 consumer items at the age 18-34 is different from the mean in the +35 group.

Question 6

Question 6

$H_0 : \mu_1 - \mu_2 = 0$ $\qquad \sigma_1 = 3.6\,^\circ C \rightarrow \sigma_1^2 = 12.96$

$H_1 : \mu_1 > \mu_2$ $\qquad \sigma_2 = 2.5\,^\circ C \rightarrow \sigma_2^2 = 6.25$

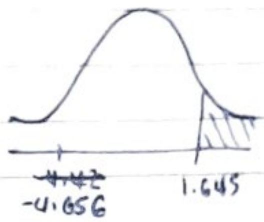
$\alpha = 0.05$ $\qquad \bar{x}_1 = 93.6\,^\circ C$

$n_1 = 16$ $\qquad \bar{x}_2 = 98.5\,^\circ C$

$n_2 = 21$

$Z_{0.05} = 1.645$

$$Z_0 = \frac{93.6 - 98.5}{\sqrt{\frac{12.96}{16} + \frac{6.25}{21}}} = \cancel{-4.42}\; -4.656$$



~~Therefore~~, Since $\cancel{-4.42}\; -4.656 < 1.645$, we fail to reject H_0 at 95% significance level. Therefore, the standard variation of process from batch is greater than the standard deviation of batch 2.

Question 7

Ans. to the Q-7

$H_0 : \mu_d = 0$

$H_1 : \mu_d < 0$

The degree of freedom = 9,

Critical value for left tailed test with 0.05 significance level is -1.833

difference, $d = x_1 - x_2$ [x_1 = before training, x_2 = after training]

$d = -22, -20, -13, -6, -22, -27, -34, -6, -31, -26$

$\Sigma d = -207$

$\Sigma d^2 = 5131$

Now, $\bar{d} = \frac{\Sigma d}{n} = \frac{-207}{10} = -20.7$

Standard deviation, $s_d = \sqrt{\frac{\Sigma d^2 - \frac{(\Sigma d)^2}{n}}{n-1}}$

$= \sqrt{\frac{5131 - \frac{(-207)^2}{10}}{10-1}}$

$= 9.69593$

variance unknown,

So t-test $= \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}} = -6.7512$

Since, $t = -6.7512 < t$ value $= -1.833$, we reject the Hypothesis. There are convincing evidence that training can increase the number of words spelled correctly.

Question 8

Ans to the Q-08

i)

Here, given $n = 8$, $\Sigma x = 14.6$, $\Sigma y = 26$

$\Sigma xy = 56.8$, $\Sigma x^2 = 32.9632$, $\Sigma y^2 = 104$

correlation coefficient, $r = \dfrac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$

$= \dfrac{74.8}{(7.1095)(12.49)}$

$= 0.8423645872$

Ans.

ii) $H_0 : \rho = 0$

$H_1 : \rho \neq 0$

$d.f = n - 2 = 8 - 2 = 6$, $\alpha = 0.05$

from r table critical value = 0.707

Here, $r = 0.8424 > 0.707$; so we reject the hypothesis. So we can say that there is significant correlates between x and y

iii) Here, $\alpha = 0.01$, critical value = 0.834

again $r = 0.8424 > 0.834$, so we reject hypothesis. So the decision won't change

Question 9

i)

Volume (x)	Mileage (y)	x^2	y^2	xy
97	24	9409	576	2328
85	29	7225	841	2465
98	26	9604	676	2548
105	24	11025	576	2520
120	24	14400	576	2880
151	22	22801	484	3322
140	23	19600	529	3220
134	23	17956	529	3082
146	21	21316	441	3066
$\sum x = 1076$	$\sum y = 216$	$\sum x^2 = 133336$	$\sum y^2 = 5228$	$\sum xy = 25431$

$$b_1 = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$= \frac{25431 - \frac{(1076)(216)}{9}}{133336 - \frac{(1076)^2}{9}}$$

$$= -0.0837$$

$$\bar{y} = \frac{216}{9} = 24 \qquad \bar{x} = \frac{1076}{9} = 119.5556$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$= 24 - (-0.0837)(119.5556)$$

$$= 34.0068$$

$\therefore$ The least squares regression equation is $y = 34.0068 - 0.0837x$

ii)

$$y = 34.0068 - 0.0837x$$

$$= 34.0068 - 0.0837(125)$$

$$= 23.5443$$

The estimated mileage for a car that has 125 cubic km engine volume is 23.5443 km

Question 10

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : at least one treatment will have different means

$\alpha = 0.01$

Σ Memory booster $= 70 + 77 + 83 + 90 + 97$
$= 417$

Σ Placebo $= 37 + 43 + 50 + 57 + 63$
$= 250$

Σ Without Treatment $= 3 + 10 + 17 + 23 + 30$
$= 83$

$n = 5$

Memory booster:

$\bar{x} = \frac{417}{5} \quad s = \sqrt{\frac{(70-83.4)^2 + (77-83.4)^2 + (83-83.4)^2 + (90-83.4)^2 + (97-83.4)^2}{5-1}}$

$= 83.4 \quad = 10.60$

Placebo:

$\bar{x} = \frac{250}{5} \quad s = \sqrt{\frac{(37-50)^2 + (43-50)^2 + (50-50)^2 + (57-50)^2 + (63-50)^2}{5-1}}$

$= 50 \quad = 10.44$

Without Treatment:

$\bar{x} = \frac{83}{5} \quad s = \sqrt{\frac{(3-16.6)^2 + (10-16.6)^2 + (17-16.6)^2 + (23-16.6)^2 + (30-16.6)^2}{5-1}}$

$= 16.6 \quad = 10.60$

$k = 3$

$\bar{\bar{x}} = \frac{83.4 + 50 + 16.6}{3} = 50$

$s_{\bar{x}} = \sqrt{\frac{(83.4-50)^2 + (50-50)^2 + (16.6-50)^2}{3-1}}$

$= 33.4$

$ns_{\bar{x}}^2 = 5(33.4)^2 = 5577.8$

$$s_p^2 = \frac{(10.60)^2 + (10.44)^2 + (10.60)^2}{3}$$

$= 111.24$

$$F = \frac{ns_{\bar{x}}^2}{s_p^2}$$

$$= \frac{5577.8}{111.24}$$

$= 50.142$

Numerator $= k-1 = 3-1 = 2$
Denominator $= k(n-1) = 3(5-1) = 12$

F-critical value $= 6.93$

$\therefore$ Since $F_{\text{test statistic}} > F_{\text{critical value}}$ $(50.142 > 6.93)$, we reject the null hypothesis

$\therefore$ There is sufficient evidence to claim that the treatments will have different effects.