PROBABILITY & STATISTICAL DATA ANALYSIS

SEMESTER II 2020/2021

SECI2143-05

ASSIGNMENT 3

NAME OF STUDENTS	MATRIC NO.
BRENDAN DYLAN GAMPA ANAK JOSEPH DUSIT @ DUSIT	A20EC0021
HAFIY HAKIMI BIN SAIFUL REDZUAN	A20EC0040
MOHAMAD HAZIQ ZIKRY BIN MOHAMMAD RAZAK	A20EC0079
AUM JEEVAN A/L AUM NIRANGKAR	A20EC0017

LECTURER: DR. NOR AZIZAH BINTI ALI

SUBMISSION DATE: 3RD JULY 2021

Question 1

a) $n = 25$ $\sigma = 9.88$ confidence interval $= 90\%$ $\bar{x} = 19.5$

$\downarrow$

$Z_{0.05} = 1.645$

confidence interval $= \bar{x} \pm (z_{\alpha/2})\left(\frac{\sigma}{\sqrt{n}}\right)$

$= 19.5 \pm (1.645)\left(\frac{9.88}{\sqrt{25}}\right)$

$= 19.5 \pm 3.25052$

$=$ lower confidence limit : 16.25 , upper confidence limit : 22.75

b) $\hat{p} = \frac{160}{360} = \frac{4}{9} = 0.4444$

$n = 360$, $a = 99\%$

$z_{a/2} = 2.575$

confidence interval $= \hat{p} \pm (z_{a/2})\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$= 0.4444 \pm (2.575)\sqrt{\frac{0.4444(0.5556)}{360}}$

$= 0.4444 \pm 0.0674$

$= (0.3770, 0.5118)$

QUESTION 2 **(15 MARKS)**

(a) The company ABC introduce a new drug test method, 350 subjects were tested and results from 33 subjects were incorrect (either false positive or false negative). The company claims that the incorrect test results will be less than 10%. Test this claim using α=0.05.

$H_0 : p = 0.1$

$H_1 : p < 0.1$

$$\hat{p} = \frac{33}{350} = 0.09429$$

$p = 0.1$, $\hat{p} = 0.09429$, $n = 350$

$$z = \frac{0.094 - 0.1}{\sqrt{\frac{(0.1)(0.9)}{350}}} = \frac{-0.006}{0.016} = -0.37$$

$z = -0.37$

critical value $= Z_{0.05} = -1.645$

Test statistic does not fall into critical region, hence fail to reject H_0. Does not have sufficient evidence to accept $p < 0.1$.

(b) A random sample of 16 steel beams has a mean compressive strength of 58400 psi (pound per square inch) and population standard deviation of 652 psi. Using 99% significance level, test if the true average compressive strength of steel is 58000 psi.

$\bar{x} = 58400$ $\sigma = 652$ $n = 16$

$H_0 : \bar{x} = 58000$

$H_1 : \bar{x} \neq 58000$

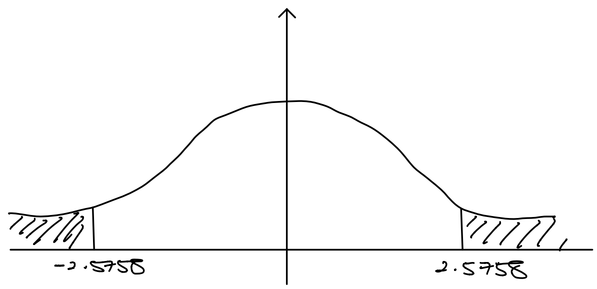
Test statistic

$$z = \frac{58400 - 58000}{\frac{652}{\sqrt{16}}} = 2.45$$

$\alpha = 0.01$

<u>Since it is two tailed</u>

critical value $= Z_{\alpha/2} = 2.5758$



<u>Conclusion</u>

Test statistics does not fall into critical region, hence fail to reject H_0. $\mu = 58000$.

QUESTION 3

a) $\alpha = 0.01$, $\alpha/2 = 0.005$

$H_0 : \mu = 10$ litres

$H_1 : \mu \neq 10$ litres

* test on mean, variance UNKNOWN

$$\cancel{s =} \quad \bar{x} = \frac{10.3 + 9.9 + 10.2 + 10.1 + 9.7 + 9.9 + 9.8 + 10.3 + 10.0 + 10.4}{10}$$

$$= 10.06$$

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

$$= \sqrt{\frac{0.0576 + 0.0256 + 0.0196 + 0.0016 + 0.1296 + 0.0256 + 0.0676 + 0.0576 + 0.0036 + 0.1156}{10 - 1}}$$

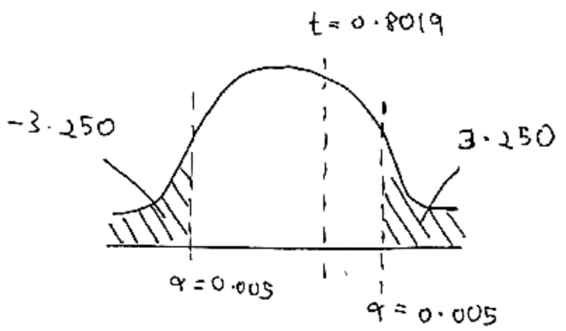
$$S = 0.2366$$

since $n < 30$, then use formula

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{10.06 - 10.00}{0.2366/\sqrt{10}}$$

$$t = 0.8019$$

critical value $t_{0.005,9} = \pm 3.250$



* since $0.8019 < 3.250$, fail to reject H_0. There is sufficient evidence that the average content volume of the Brand X car lubricants is 10 litres.

b.) $\sigma = 0.44$ kg

$a = \cancel{0.95}\ 0.05$ (right-tailed), $n = 14$ $\quad a/2 = 0.0025$

* Test on variance / standard deviation

$H_0 : \sigma = 0.44$

$H_1 : \sigma \neq 0.44$

$$\bar{x} = \frac{3.70 + 4.31 + 3.73 + 4.33 + 3.33 + 2.58 + 4.47 + 3.55 + 4.66 + 3.68 + 3.02 + 4.09 + 2.36 + 3.35}{14}$$

$\bar{x} = 3.654$

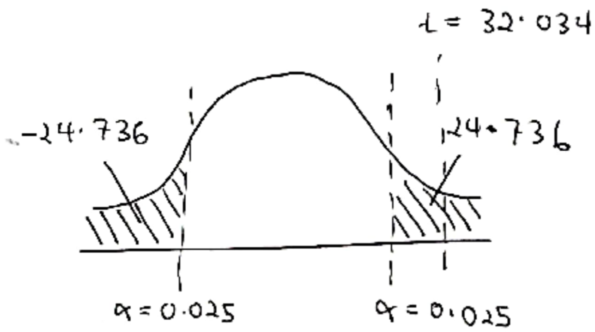
$$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

$$= \sqrt{\frac{0.002116 + 0.4303 + 0.005776 + 0.4570 + 0.1050 + 1.1535 + 0.6659 + 0.0108 + 1.012036 + 0.000676 + 0.402 + 0.1901 + 1.6744 + 0.0924}{14-1}}$$

$= 0.6907$

$$\chi^2 = \frac{(n-1)(s)^2}{\sigma^2} = \frac{(14-1)(0.6907)^2}{(0.44)^2} = 32.034$$

critical value $\chi^2_{0.025,13} = \pm 24.736$



* since $32.034 > 24.736$, reject H_0. There is sufficient evidence that the vitamin supplements has affected the standard deviation of the birth weight.

Question 4

$H_0: \sigma^2 = 0.18$ $\quad$ $95\% = \alpha = 0.05$ $\quad$ $\alpha/2 = 0.025$

$H_1: \sigma^2 \neq 0.18$ $\quad$ $n = 101$

$s^2 = 0.165$

Test statistic:

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

$$= \frac{(101-1)(0.165)}{0.18}$$

$$= 91.67$$

Critical values: $\chi^2_{0.0025,100} = 129.591$

$\chi^2_{0.975,100} = 74.22$

Since $\chi^2 = 91.67$ does not lie in critical region, Ho fails to be rejected. There is no sufficient evidence to conclude that the new inspector is making satisfactory measurements.

Question 5

Given in question,

$$n_1 = 15$$
$$\bar{x}_1 = 76.40$$
$$s_1^2 = 25.30$$
$$s_1 = 5.03$$

Hypothesis, $H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \neq 0$

rejection region $\alpha = 0.05$

Test ~~statistics~~ $\cancel{t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}}$

$t = \cancel{76.40} - 71.20$

df ~~total~~ $= \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\left(\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}\right)}$

$= \dfrac{\left(\dfrac{25.30}{15} + \dfrac{22.20}{18}\right)^2}{\left(\dfrac{(25.30/15)^2}{15-1} + \left(\dfrac{(22.20/18)^2}{18-1}\right)\right)}$

$= 29.1299$

$\therefore t_\alpha = 2.045$

test statistics $t = \dfrac{\bar{x}_1 - x_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

$= \dfrac{76.40 - 71.20}{\sqrt{\dfrac{(5.03)^2}{15} + \dfrac{(4.71)^2}{18}}}$

$= 3.043$

$|t| = 3.043$

$t_\alpha = 2.045$

$3.043 > 2.045$

$\therefore$ H_0 is rejected. There is not enough proof to claim that population mean ~~the~~ between two groups at 0.05

QUESTION 6

Note:

σ_1 : batch 1 — numerator

σ_2 : batch 2 — denominator

~~$\alpha = 0.05$~~

$\alpha = 0.05$

$H_0 : \sigma_1 = \sigma_2$

$H_1 : \sigma_1 > \sigma_2$ (claim)

Find variances, σ_1^2 and σ_2^2

$\sigma_1^2 = (3.6)^2 = 12.96$

$\sigma_2^2 = (2.5)^2 = 6.25$

d.f Numerator ; $n_1 - 1 = 16 - 1 = 15$

Denominator, $n_2 - 1 = 21 - 1 = 20$

$$F = \frac{S_1^2}{S_2^2} = \frac{12.96}{6.25} = 2.0736$$

critical value $F_{(0.05, 15, 20)} = 2.20$

since $2.0736 < 2.20$, ~~failed~~ fail to reject H_0. There is sufficient evidence that the standard deviation of processor ~~is less~~ from Batch 1 is less than the standard deviation of processor from Batch 2 at 95% significance level.

QUESTION 7 **(15 MARKS)**

Table 4 shows the number of words spelled correctly by 10 participants of an international spelling bee competition before and after training. Using 0.05 significance level, is there sufficient evidence to conclude that training can increase the number of words spelled correctly?

Table 4: Number of words spelled correctly before and after training.

Before Training	203	390	389	279	333	213	410	364	470	464
After Training	225	410	402	285	355	240	444	370	501	490

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

$\bar{d} = \frac{-207}{10} = -20.7$

$\sigma = \sqrt{\frac{10(5131) - 42849}{10(9)}} = 9.6959$

$H_0 : \mu_d = 0$

$H_1 : \mu_d < 0$

since $\alpha = 0.05$

test statistic

$t = \frac{(-20.7 - 0)}{9.6959 / \sqrt{10}} = -6.7512$

$d.f = n-1 = 9 \quad , \quad \alpha = 0.05$

critical value $= -1.833$

since $-6.7512 < -1.833$, reject H_0. There is sufficient evidence to conclude that training can increase the number of words spelled correctly.

Question 8

i)

x	y	xy	x^2	y^2
0.27	2	0.54	0.0729	4
1.41	3	4.23	1.9881	9
2.19	3	6.57	4.7961	9
2.83	6	16.98	8.0089	36
2.19	4	8.76	4.7961	16
1.81	2	3.62	3.2761	4
0.85	1	0.85	0.7225	1
3.05	5	15.25	9.3025	25
$\Sigma = 14.6$	$\Sigma = 26$	$\Sigma = 56.8$	$\Sigma = 32.9632$	$\Sigma = 104$

$n = 8$

$$r = \frac{\Sigma xy - (\Sigma x \cdot \Sigma y)/n}{\sqrt{\left[(\Sigma x^2) - (\Sigma x)^2/n\right]\left[(\Sigma y^2) - (\Sigma y)^2/n\right]}}$$

$$r = \frac{(56.8) - (14.6)(26)/8}{\sqrt{\left[(32.9632) - (14.6)^2/8\right]\left[(104) - (26)^2/8\right]}}$$

$$r = 0.8424$$

ii) $H_0 : \rho = 0$ $\quad a = 0.05$, $d.f = n-2$

$H_1 : \rho \neq 0$ $\quad a/2 = 0.025$ $\quad = 8-2$

$= 6$

critical value ; $t_{0.025, 6} = 2.447$

Test statistic: $t = \dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} = \dfrac{0.8424}{\sqrt{\dfrac{1-(0.8424)^2}{8-2}}} = 3.8293$

Since $t = 3.8293 > 2.447$, Ho is rejected. There is sufficient evidence that X and Y really correlates using 95% confidence level.

iii) 99% = 0.01 confidence level, $a/2 = 0.005$

critical value ; $t_{0.005, 6} = 3.707$

Since $t = 3.8293 > 3.707$, Ho is still rejected. There is sufficient evidence that X and Y really correlates using 95% confidence level. Thus, the decision won't change.

QUESTION 9

i.) Find least square equation. ($\hat{y} = b_0 + b_1 x$)

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

$$b_1 = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}} = \frac{25431 - \left(\frac{(1076)(216)}{9}\right)}{133336 - \left(\frac{(1076)^2}{9}\right)}$$

-393

4692.222

$$b_1 = -0.08376$$

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

$$= \left(\frac{216}{9}\right) - (-0.08376)\left(\frac{1076}{9}\right)$$

$$= 34.014$$

$$\therefore \hat{y} = 34.014 - 0.08376x$$

ii.) given that $x = 125$,

$$\hat{y} = 34.014 - 0.08376(125)$$

$$= 23.544$$

$\therefore$ 23.544 km #

Question 10

Step 1: $H_0 : \mu_1 = \mu_2 = \mu_3$

$H_1 : \mu_1 \neq \mu_2 \neq \mu_3$

Step 2:

Memory booster

$n = 5$

$$\bar{x} = \frac{70 + 77 + 83 + 90 + 97}{5}$$

$$= 83.4$$

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

$$= 10.60$$

Placebo

$n = 5$

$$\bar{x} = \frac{37 + 43 + 50 + 57 + 63}{5}$$

$$= 50.0$$

$$S = \sqrt{\frac{(37-50)^2 + (43-50)^2 + (50-50)^2 + (57-50)^2 + (63-50)^2}{5-1}}$$

$$= 10.44$$

Without Treatment

$n = 5$

$$\bar{x} = \frac{3 + 10 + 17 + 23 + 30}{5} = 16.6$$

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

$$= 10.60$$

Step 3a: $\bar{\bar{x}} = \frac{83.4 + \cancel{50}50.0 + 16.6}{k = 3}$

$= 50.0$

Step 3b: $S_{\bar{x}} = \sqrt{\frac{(83.4-50.0)^2 + (\cancel{50.0}\,50.0-50.0)^2 + (16.6-50.0)^2}{\cancel{34}3-1}}$

$= 33.4$

Step 3c: $n s_{\bar{x}}^{2} = 5(33.4)^2$

$= 5577.8$

Step 4: $S_p^{2} = \frac{(10.60)^2 + (10.44)^2 + (10.60)^2}{k = 3}$

$= 111.24$

Step 5: $F = \frac{n s_{\bar{x}}^{2}}{S_p^{2}}$

$= \frac{5577.8}{111.24}$

$= 50.14$

Step 6: Numerator $= k-1 = 3-1 = 2$

Denominator $= k(n-1)$

$= 3(5-1)$

$= 12$

Step 7: Find critical value with $\alpha = 0.01$ from F-distribution table

F-critical value $= 6.927$

Step 8: Calculated value of F is greater than critical value of F. So, reject H_0 at $\alpha = 0.05$ level of significance.

Thus, there is a difference in the mean scores of 3 different group.