COURSE:

SECI2143-02 PROBABILITY & STATISTICAL DATA ANALYSIS

FACULTY:

FACULTY OF ENGINEERING

SCHOOL:

SCHOOL OF COMPUTING

TITLE:

ASSIGNMENT 3

LECTURER'S NAME:

DR. CHAN WENG HOWE

NAME	MATRIC NO.
NURARISSA DAYANA BINTI MOHD SUKRI	A20EC0120
SAKINAH AL'IZZAH BINTI MOHD ASRI	A20EC0142
QAISARA BINTI ROHZAN	A20EC0133

Question 1

(a) $n = 25$ university students

$\bar{x} = 19.5$ seconds

$\sigma_{\text{(standard deviation)}} = 9.88$ seconds

$1 - \alpha = 0.90$

$\alpha = 0.1$, $\alpha/2 = 0.05$

so $Z_{\alpha/2} = Z_{0.05} = 1.645$

Therefore :

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 19.50 \pm 1.645 \frac{9.88}{\sqrt{25}}$$

$$= 19.50 \pm 3.25$$

$$= (16.25, 22.75)$$

∴ we are 90% confident that the mean number of time for women's 100-meter perfomance is between 16.25 and 22.75 seconds.

(b) $n = 360$ food stores, 160 food stores offers special promotions.

sample proportion, $\hat{p} = \frac{160}{360} = 0.44$

$$99\% \text{ of Confident Interval} = \hat{p} \pm (\text{Z critical value}) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= 0.44 \pm 2.58 \sqrt{\frac{0.44(1-0.44)}{360}}$$

$$= 0.44 \pm 0.07$$

$$= (0.37, 0.51)$$

∴ we are 99% confident that the populatio of food stores that offers promotion estimation is between 0.37 and 0.51.

QUESTION 2

a) $n = 350$

$H_0 : p = 0.1$

$H_1 : p < 0.1$

population proportion, $\hat{p} = \frac{33}{350}$ ← count of success / ← size of population

$\hat{p} = 0.09429$

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

$$z = \frac{\hat{p} - p}{\sqrt{\frac{(p)(1-p)}{n}}}$$

$$= \frac{0.09429 - 0.1}{\sqrt{\frac{(0.1)(0.9)}{350}}}$$

$z = -0.37$

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

∴ Test statistic does not fall under critical region, hence we fail to reject H_0. This shows that we do not have sufficient evidence to accept $p < 0.1$.

b) $\bar{x} = 58400$, $\sigma = 652$, $n = 16$, $\alpha = 0.01$ ← 99% confidence level

$H_0 : \mu = 58000$

$H_1 : \mu \neq 58000$

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

Test statistic

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

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

$z = 2.45$

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

QUESTION 3

a) sample size, $n = 10$, $\alpha = 0.05$

10.3
9.9
10.2
10.1
9.7
9.9
9.8
10.2
10.0
10.4

Hypothesis statement :

$H_0 : \mu = 10$

$H_1 : \mu \neq 10$

Mean, $\bar{x} = \frac{\Sigma x}{n}$

$$= \frac{10.3 + 9.9 + 10.2 + 10.1 + 9.7 + 9.9 + 9.8 + 10.3 + 10.0 + 10.4}{10}$$

$\bar{x} = 10.06$

Sample standard deviation, s

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

$$= \sqrt{\frac{(1012.54) - \frac{(100.6)^2}{10}}{9}}$$

$$= \sqrt{0.056}$$

$s = 0.236643$

$\therefore$ test statistic, $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

$$t = \frac{10.06 - 10}{0.23664/\sqrt{10}} = \cancel{0.8019}\ 0.8018$$

critical value :

$df = n-1 = 10-1$

$df = 9$

$\therefore t_{0.025, 9} = 2.262$

Result :

~~to~~ Fail to reject H_0. Since the critical region is $t > 2.262$ or $t < -2.262$, the ~~ta~~ t-score of 0.8018 falls outside the region. Therefore, ~~the~~ we can conclude that the average content volume of the Brand X car lubricants is 10 liters.

QUESTION 3

b) $n = 14$

$\sigma = 0.44$

$\alpha = 0.05$

Hypothesis statement.

$H_0 : \sigma = 0.44$

$H_1 : \sigma \neq 0.44$

x	x^2
3.70	13.69
4.31	18.576
3.73	13.9129
4.33	18.7489
3.33	11.0889
2.58	6.6564
4.47	19.9809
3.55	12.6025
4.66	21.7156
3.68	13.5424
3.02	9.1204
4.09	16.7281
2.36	5.5696
3.35	11.2225
$\Sigma x = 51.16$	$\Sigma x^2 = 193.1552$

$\chi^2_{0.05, 13}$

~~Va~~ sample variance, s^2

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

$$= \frac{(193.1552) - \frac{(51.16)^2}{14}}{14-1}$$

$$s^2 = 0.4771$$

Test statistic :

$$\therefore \chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(14-1)(0.4771)}{(0.44)^2}$$

$$\chi^2 = 32.035$$

Critical region :

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

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

Decision :

Reject Ho, since $\chi^2 = 32.035 > 24.736$ upper-tail critical value.

The test statistic χ^2 falls within the critical region of $\chi^2 = 24.736$.

Therefore, at 0.05 significance level, there is sufficient evidence to conclude that vitamin supplement has effects on birth weight.

Question 4

$H_0 : \sigma^2 = 0.18 \text{ (inch)}^2$
$H_1 : \sigma^2 < 0.18 \text{ (inch)}^2$

$n = 101$
$\alpha = 0.05$ (95%)

Test statistic :

$$\therefore \chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{100(0.165)}{0.18} = 91.667$$

$$\chi^2_{0.05, 100} = 77.929$$

conclusion : Since $77.929 < 91.667$, fail to reject H_0. There is sufficient evidence that the new joint inspector is not making satisfactory measurement. The satisfactory measurement is 0.18 (inch)^2

Question 5

$\bar{x}_1 = 76.40$
$\sigma_1^2 = 25.30$
$n_1 = 15$

$\bar{x}_2 = 71.20$
$\sigma_2^2 = 22.20$
$n_2 = 18$

$\alpha = 0.05$
C.V $= Z_{0.05/2} = Z_{0.025} = 1.96$

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

Test statistic :

$$Z_0 = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{76.40 - 71.20}{\sqrt{\frac{25.30}{15} + \frac{22.20}{18}}}$$

$$= \frac{5.20}{\sqrt{2.92}}$$

$$Z_0 = 3.04$$

conclusion : Since $Z_0 = 3.04 > 1.96$, we reject H_0 at the 0.05 level and conclude that there is the difference in the mean spending between these two populations, 18 - 34 age group and 35+ group.

QUESTION 6

Batch 1	Batch 2	
$n_1 = 16$	$n_2 = 21$	$\alpha = 0.05$
$\bar{x}_1 = 93.6$	$\bar{x}_2 = 98.5$	
$s_1 = 3.6$	$s_2 = 2.5$	

Hypothesis statement:

$H_0 : \sigma_1 = \sigma_2$

$H_1 : \sigma_1 > \sigma_2$

Test statistic:

$$F = \frac{(s_1)^2}{(s_2)^2} = \frac{(3.6)^2}{(2.5)^2}$$

$$F = 2.0736$$

Degree of Freedom:

- Numerator, $n_1 - 1 = 15$
- Denominator, $n_2 - 1 = 20$

Critical value:

$F_{0.05, 15, 20} = 2.2033$

Conclusion:

~~Since~~, Since $F = 2.0736 < F_{0.05,15,20} = 2.2033$, we fail to reject H_0. Hence, there is no significant evidence to support that the standard deviation of processors from batch 1 is greater than the standard deviation of batch 2.

QUESTION 7

X = Before Training

Y = After Training

n = 10

①

X	Y	XY	X^2	Y^2
203	225	45675	41209	50625
390	410	159900	152100	168100
389	402	156378	151321	161604
279	285	79515	77841	81225
333	355	118215	110889	126025
213	240	51120	45369	57600
410	444	182040	168100	197136
364	370	134680	132496	136900
470	501	235470	220900	251001
464	490	227360	215296	240100
3515 ← $\sum X$	3722 ← $\sum Y$	1390353 ← $\sum XY$	1315521 ← $\sum X^2$	1470316 ← $\sum Y^2$

② $$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$r = \frac{10(1390353) - (3515)(3722)}{\sqrt{[10(1315521) - (3515)^2][10(1470316) - (3722)^2]}}$$

$$r = \frac{820700}{\sqrt{(799985)(849876)}}$$

$r = 0.99533$

correlation coefficient

③ $H_0 : \rho = 0$ (no linear correlation)

$H_1 : \rho \neq 0$ (linear correlation exist)

$\alpha = 0.05$, $df = n-2 = 8$, $t_{\alpha/2} = 0.025, 8 = 2.306$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.99533}{\sqrt{\frac{1-(0.99533)^2}{8}}} = 29.1639$$

④ If test statistics > 2.306, we reject H_0. otherwise, we fail to reject H_0.

Since test statistics = 29.1639 > 2.306, we reject H_0. There is sufficient evidence to conclude that training can increase the number of words spelled correctly.

QUESTION 8

i) X = weight of plastic usage $n = 8$

Y = size of household

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
14.6	26	56.8	32.9632	104
↑	↑	↑	↑	↑
ΣX	ΣY	ΣXY	ΣX^2	ΣY^2

correlation coefficient, r:

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

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

$$r = \frac{14.8}{\sqrt{[50.5456][156]}}$$

$$r = 0.84236$$

ii) $H_0 : \rho = 0$ (No linear correlation)

$H_1 : \rho \neq 0$ (Linear correlation exists)

$\alpha = 0.05$ $\quad df = n - 2 = 8 - 2 = 6$ $\quad t_{\alpha/2 = 0.025, 6} = \pm 2.4469$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.84236}{\sqrt{\frac{1-(0.84236)^2}{6}}} = 3.8287$$

∴ If test statistics > 2.4469 or test statistics < -2.4469, reject H_0. otherwise fail to reject H_0. Since $t = 3.8287 > 2.4469$, H_0 is rejected There is sufficient evidence of a linear relationship between weight of plastic usage and size of household

iii) $\alpha = 0.01$

$t_{\alpha/2 = 0.005, 6} = \pm 3.707$

∴ Since test statistics $= 3.8287 > 3.707$, we reject H_0. Even if we increase the confidence level to 99%, it doesn't change the previous decision. There is sufficient evidence of a linear relationship between weight of plastic usage and size of household.

QUESTION 9

i)

	x	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
Total:	$\Sigma x = 1076$	$\Sigma y = 216$	133336	5228	$\Sigma xy = 25431$

$n = 9$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{1076}{9} = 119.56$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{216}{9} = 24$$

Least square Equation:

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

$$b_1 = \frac{-393}{4694.22}$$

$$b_1 = -0.08372$$

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

$$= (24) - (-0.08372)(119.56)$$

$$b_0 = 34.0096$$

QUESTION 9

$b_1 = -0.08372$

$b_0 = 34.01$

$\hat{y} = b_0 + b_1 x$

$\widehat{Mileage} = 34.01 + (-0.08372)(\text{engine volume})$

$\therefore$ The least square equation is

$\widehat{Mileage} = 34.01 - 0.08372 * volume$

ii) Estimate mileage for a car of 125 cubic km engine volume

$\widehat{Mileage} = 34.01 - 0.08372 * volume$

$= 34.01 - 0.08372(125)$

$= 23.545$

$\therefore$ The estimated mileage for 125 cubic km capacity is 23.545 km.

Question 10

$\alpha = 0.01$

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

H_1 : at least one mean is different

memory booster:

$n = 5$

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

$= 83.40$

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

$= \sqrt{\dfrac{179.56 + 40.96 + 0.16 + 43.56 + 184.96}{4}}$

$= 10.60$

Placebo:

$n = 5$

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

$= 50$

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

$= \sqrt{\dfrac{169 + 49 + 0 + 49 + 169}{4}}$

$= 10.44$

without treatment:

$n = 5$

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

$= 16.6$

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

$= \sqrt{\dfrac{184.96 + 43.56 + 0.16 + 40.96 + 179.56}{4}}$

$= 10.60$

mean between samples:

$$\bar{\bar{x}} = \frac{83.40 + 50.0 + 16.60}{k = 3} = 50.0$$

standard deviation between samples:

$$S_{\bar{x}} = \sqrt{\frac{(83.40 - 50)^2 + (50 - 50)^2 + (16.6 - 50)^2}{3 - 1}}$$

$$= \sqrt{\frac{1115.56 + 0 + 1115.56}{2}}$$

$$= 33.40$$

variance between samples:

$$nS_{\bar{x}}^2 = 5(33.40)^2 = 5577.80$$

variance within samples:

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

$$= 111.24$$

Test statistic, F

$$F = \frac{\text{variance between samples}}{\text{variance within samples}} = \frac{5577.80}{111.24} = 50.14$$

numerator $= k - 1 = 3 - 1 = 2$

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

F critical value $= 6.93$, $\alpha = 0.01$

conclusion: Since F_{test} statistic > F critical value ($50.14 > 6.93$), we reject the null hypothesis. There is enough evidence to state that at least one of the means is different.