	DEPARTMENT OF APPLIED COMPUTING		
UTM UNIVERSITI TEKNOLOGI MALAYSIA	**SUBJECT:** PROBABILITY & STATISTICAL DATA ANALYSIS		
	ASSESSMENT: EXERCISE 1	**CODE:** SECI 2143	**WEEK:** 2

Name : SOH ZEN REN
WONG HUI SHI
TEOH WEI JIAN

Matric ID / IC No : A20EC0152
A20EC0169
A20EC0229

Question 1,2,3,4	Wong Hui Shi
Question 5,6,7	Soh Zen Ren
Question 8,9,10	Teoh Wei Jian

ASSIGNMENT 3: Hypothesis 1 sample, 2 sample, Chi-Square test, Correlation, Regression, Anova

ANSWERS SHEET

Assignment 3

Group 12

Question 1

Given

$n = 25$ $\bar{x} = 19.5$ $\sigma = 9.88$ $\alpha = 0.1$

$\alpha/2 = 0.05$

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

Therefore:

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 19.5 \pm Z_{0.1/2} \frac{(9.88)}{\sqrt{25}}$$

$$= 19.5 \pm Z_{0.05} \frac{9.88}{\sqrt{25}}$$

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

$$= 19.5 \pm 3.25$$

The lower and upper confidence limits are 16.25 and 22.75.

confidence interval = (16.25 , 22.75)

∴ We are 90 % confident that the mean time of 100-meter performance for the university students is between 16.25 and 22.75.

Question 1

(b) Given

total food stores = 360   food stores offers special promotion = 160

$$\text{point estimate} = \hat{p} = \frac{\text{food stores offer special promotion}}{\text{total food stores}}$$

$$= \frac{160}{360}$$

$$= 0.44$$

for $\alpha = 0.01$, $Z_{0.01} = 2.58$ (99% confidence interval)

$$99\%\ CI = \hat{p} \pm Z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.44 \pm 2.58\left(\sqrt{\frac{(0.44)(1-0.44)}{360}}\right)$$

$$= 0.44 \pm 0.0675$$

$$= (0.3725, 0.5075)$$

∴ We are 99% confident that the true proportion population of food stores that offers promotion is between 0.3725 and 0.5075.

Question 2

(a) Given

total tested : 350 incorrect = 33

statement of Hypothesis :

$H_0 = \mu = 0.1$

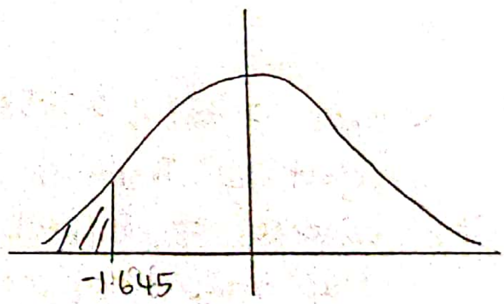
$H_1 : \mu < 0.1$

For $\alpha = 0.05$, $Z_{0.05} = -1.645$

$$\text{point estimate} = \hat{p} = \frac{\text{incorrect}}{\text{total tested}} = \frac{33}{350} = 0.094$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.094 - 0.1}{\sqrt{\frac{(0.1)(0.9)}{350}}} = -0.3742$$

∴ since $-0.3742 > -1.645$, fail to reject H_0 , there are insufficient evidence to support that the incorrect test result will be less than 10%.



Question 2

(b) Given

$n=16$ $\bar{x} = 58400$ $\sigma = 652$ $\alpha = 100\% - 99\%$
$= 1\%$
$= 0.01$

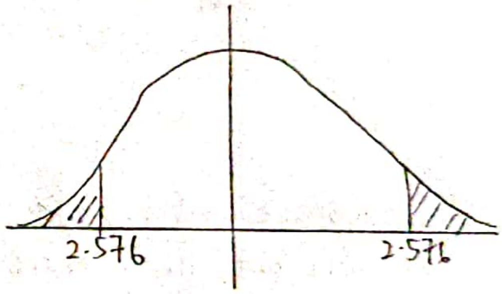
For two tail, $\alpha/2 = 0.01/2$
$= 0.005$

Statement of hypothesis:
$H_0 : \mu = 58000$
$H_1 : \mu \neq 58000$

For $\alpha = 0.005$, $Z_{0.005} = \pm 2.5758$

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{58400 - 58000}{652/\sqrt{16}} = 2.454$$

p-value $= P(Z = 2.454) = 0.00734$



$\therefore$ Since $2.454 < 2.5758$, fail to reject H_0. There is insufficient evidence to support that the true value compressive strength of steel is 58000 psi.

Question 3

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

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

$\bar{x} = 10.06$

For this data given:

The standard deviation is

$s = 0.2366$

Statement of Hypothesis:

$H_0 : \mu = 10$

$H_1 : \mu \neq 10$

$df = 10 - 1$

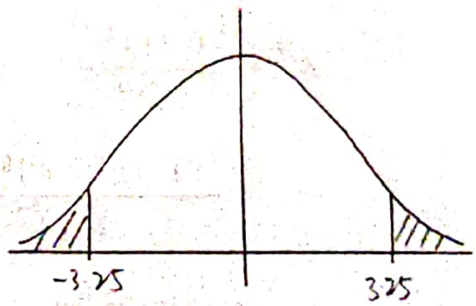
$= 9$

$t_{0.01, 9} = 3.25$

Test statistic:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{10.06 - 10}{0.2366/\sqrt{10}} = 0.8019$$

$\therefore$ Since $0.8019 < 3.25$, fail to reject H_0. There are sufficient evidence to support the average content volume of the Brand X car lubricant is 10 litres.



Question 3

(b) Given

$n = 14$ $\alpha = 0.95$

For the data given:

The standard deviation is $s = 0.6907$

Statement of hypothesis:

$H_0 : \sigma = 0.44$

$H_1 : \sigma \neq 0.44$

$df = 14 - 1$
$= 13$

$\alpha/2 = 1 - 0.95/2$
$= 0.05/2$
$= 0.025$

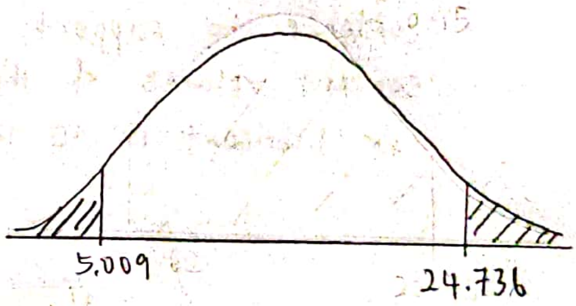
For $\chi^2_{0.025, 13} = 24.736$ (right tail)
$= 5.009$ (left tail)

Test statistic:

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

$$= \frac{(14-1)(0.6907)^2}{0.44^2}$$

$$= 32.0344$$



$\therefore$ Since $32.0344 > 24.736$, reject H_0. There are insufficient evidence to prove that standard deviation of weight of baby is equal to 0.44 kg.

Question 4

Given

$\sigma^2 = 0.18$ $n = 101$ σ^2 0.165

$\alpha/2 = 1 - 0.95/2$
$= 0.05/2$
$= 0.025$

Statement of hypothesis :

$H_0 : \sigma^2 = 0.165$

$H_1 : \sigma^2 \neq 0.165$

$df = 101 - 1$
$= 100$

For $\chi^2_{0.025, 100} = 129.561$ (right tail)
$= 74.222$ (left tail)

Test statistics :

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

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

$$= 109.091$$

$\therefore$ Since $109.091 < 124.342$, fail to reject H_0. There are sufficient evidence to conclude that the new joint is making satisfactory measurements.

Question 5

$n_1 = 15,\ \bar{x}_1 = 76.4,\ s_1^2 = 25.3$

$n_2 = 18,\ \bar{x}_2 = 71.2,\ s_2^2 = 22.2$

① Hypothesis

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

② (i) To get test statistic

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{76.4 - 71.2}{\sqrt{\frac{25.3}{15} + \frac{22.2}{18}}}$$

$$= 3.0431$$

(ii) To get degree of freedom

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2}{n_2}\right)^2}{n_2 - 1}\right)} = \frac{\left(\frac{25.3}{15} + \frac{22.2}{18}\right)^2}{\left(\frac{\left(\frac{25.3}{15}\right)^2}{14} + \frac{\left(\frac{22.2}{18}\right)^2}{17}\right)} = 29.1321$$

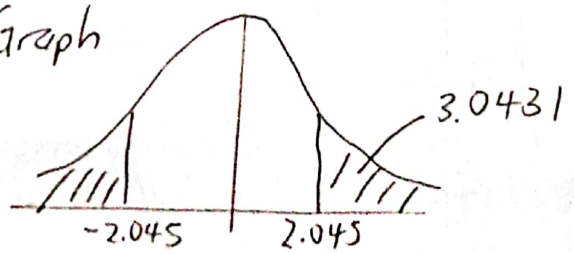
$$\approx 29$$

(iii) $1 - 0.95 = 0.05 \qquad \alpha = \frac{0.05}{2} = 0.025$

$t_0^* > t_{0.025, 29} = 2.045$

$t_0^* < -t_{0.025, 29} = -2.045$

④ Graph



⑤ Conclusion :

Since test statistic value is greater than critical value which is $3.041 > 2.045$, we reject the null hypothesis. There is sufficient evidence to conclude that there is difference in the mean spending between 2 populations in 95% significance level.

Question 6

Given

$n_1 = 16$ & $n_2 = 21$

$s_1 = 3.6$ & $s_2 = 2.5$

① Hypothesis

$H_0 : \sigma_1 = \sigma_2$

$H_1 : \sigma_1 > \sigma_2$

② (i) Test statistic

$$F = \frac{s_1^2}{s_2^2} = \frac{(3.6)^2}{(2.5)^2} = 2.0736$$

(ii) Degree of freedom

Numerator, $n_1 - 1 = 16 - 1 = 15$

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

③ $\alpha = 0.95$

$f_{0.95, 15, 20} = 2.203$

$\therefore$ Since $F_{\text{test statistic}} < F_{\text{critical value}}$, we fail to reject H_0.
Hence, there is insufficient evidence to conclude that standard variance of processors from batch 1 is greater than the standard deviation of batch 2.

estion 7

$H_0 : \mu_D = 0$

$H_1 : \mu_D < 0$

$df = 9$

$\alpha = 0.05$

(In order for the training to be effective, the before training must be less than after training, hence mean must less than zero.)

2)

Before	After	$D = (x_1 - x_2)$	$D^2 = (x_1 - x_2)^2$
203	225	-22	484
390	410	-20	400
389	402	-13	169
279	285	-6	36
333	355	-22	484
213	240	-27	729
410	444	-34	1156
364	370	-6	36
470	501	-31	961
464	490	-26	676
		$\Sigma D = -207$	$\Sigma D^2 = 5131$

3) (i) $S_D = \sqrt{\dfrac{\Sigma D^2 - \frac{(\Sigma D)^2}{n}}{n-1}}$

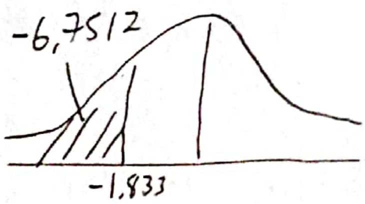
$= \sqrt{\dfrac{5131 - \frac{(-207)^2}{10}}{9}}$

$= 9.6959$

(ii) $t = \dfrac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} = \dfrac{\frac{-207}{10} - 0}{\frac{9.6959}{\sqrt{10}}}$

$= -6.7512$

4) $cv = t_{0.05, 9} = -1.833$



Since the test statistic value is smaller than critical value, which is $-6.7512 < -1.833$, we reject the null hypothesis. Hence, there is sufficient evidence to conclude that the training can increase the number of words spell correctly.

Question 8

i)

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

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

$$r=\frac{56.8-(26\times14.6)/8}{\sqrt{[(32.9632)-(14.6)^2/8][(104)-(26)^2/8]}}$$

$r=0.8424$

ii) $t=\dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$

$=\dfrac{0.8424}{\sqrt{\dfrac{1-(0.8424)^2}{8-2}}}$

$=3.8293$

$H0: p=0$

$H1: p\neq0$

d.f. $=8-2=6$

$t_{0.05,6}=2.447$

$\therefore$ Since $t=3.8293>t_{0.05,7}=2.447$, the null hypothesis is rejected. There is linear correlation exists between weight of plastic usage and size of household.

iii) $H_0: p=0$

$H_1: p\neq0$

$t=3.8293$

$t_{0.01,6}=3.707$

$\therefore$ Since $t=3.8293>t_{0.05,6}=3.707$, the decision does not change. The null hypothesis is rejected. There is linear correlation exists between weight of plastic usage and size of household.

Question 9

i)

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

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

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

$$= -0.0837 1993941$$

$$b_0 = \frac{216}{9} - (-0.08371993941)\left(\frac{1076}{9}\right)$$

$$= 34.00918387$$

$$\hat{y} = 34.0092 - 0.0837x$$

ii) $\hat{y} = 34.0092 - 0.0837(125)$

$= 23.5467 \text{ km}$

Scanned with CamScanner

Question 10

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

H_1: at least one mean is different

Category 1: Memory booster

$n = 5$

$$\bar{x} = \frac{70+77+83+90+97}{5} = 83.4$$

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

$= 10.5972$

Category 2: Placebo

$n = 5$

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

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

$= 10.4403$

Category 3: Without Treatment

$n = 5$

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

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

$= 10.5972$

mean between samples

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

standard deviation between samples

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

$= 33.4$

Variance between samples

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

Variance within samples

$$s_p^2 = \frac{(10.5972)^2 + (10.4403)^2 + (10.5972)^2}{3}$$
$$= 111.2004$$

Test statistic, F

$$F = \frac{ns_{\bar{x}}^2}{s_p^2}$$
$$= \frac{5577.8}{111.2004}$$
$$= 50.1599$$

d.f.

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

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

$a = 0.01$

$F_{c.v.} = 6.93$

$\therefore$ Since $F = 33.4399 > F_{c.v.} = 6.93$, the null hypothesis is fail to reject. There is sufficient evidence to support the claim that the treatments will have different effects.

Scanned with CamScanner