AKHIL.S
2021MT12054

ADVANCED STATISTICAL TECHNIQUES FOR ANALYTICS ASSIGNMENT

Q1

Null Hypothesis → H_0: The course feedback averages given by the professors are equal $(\mu_A = \mu_B = \mu_C = \mu_D)$

Alternate Hypothesis → H_1: Atleast one mean is different from the rest

	A	B	C	D
X	4.5	4.2	3.3	3.6
Y	3.8	4.1	2.7	4
Z	4	3.4	3.4	3
$\sum$	12.3	11.7	9.4	10.6

$n = 12 \qquad n_i = r = 4 \qquad n_j = 3$

$= T = 44$

Correction Factor $CF = \frac{T^2}{n} = \frac{44^2}{12} = 161.333$

$$SST = \sum_{i=0}^{3}\sum_{j=0}^{2} x_{ij} - CF = \left(4.5^2 + 3.8^2 + \cdots + 4^2 + 3^2\right) - 161.333$$

$$= 3.0667$$

$$SSTR = \frac{\sum_{i=0}^{3}\left(\sum_{j=0}^{2} x_{ij}\right)^2}{n_j} - CF = \frac{12.3^2 + 11.7^2 + 9.4^2 + 10.6^2}{3} - 161.333$$

$$= 1.6334$$

$$SSE = SST - SSTR = 3.0667 - 1.6334$$

$$= 1.4333$$

$$MSTR = \frac{SSTR}{r-1} = \frac{1.6334}{3} = 0.5445$$

$$MSE = \frac{SSE}{n-r} = \frac{1.4333}{8} = 0.1792$$

$$F = \frac{MSTR}{MSE} = \frac{0.5445}{0.1792} = 3.0385$$

AKHIL.S
2021MT12054

Q1 CONT..

ANNOVA SUMMARY

SOURCE OF VARIATION	SUM OF SQUARES	d.o.f	MEAN SQUARES	F-VALUE
Between Sample (COURSE)	SSTR = 1.6334	$r-1 = 3$	MSTR = 0.5445	3.0385
Within Sample (ERROR)	SSE = 1.4333	$n-r = 8$	MSE = 0.1792	
TOTAL	SST = 3.0667	$n-1 = 11$		

Table Value $F_{(0.01\ \text{CONFIDENCE}, \text{d.o.f } 3,8)} = 7.591$ [From F Table for $\alpha = 0.01$]

$3.0385 < 7.591 \Rightarrow$ Null hypothesis is accepted at 1% level of Significance since calculated F value is less than Fvalue from table. We can conclude that course feedback of professors is statistically similar for all courses

Q2

PROCESS	SAMPLE SIZE	MEAN LIFE	STD DEVIATION
A	20	20400	100
B	25	21800	100

dof = 20+25-2
= 43

We are given two means and the samples dof > 30, hence To test the hypothesis we will employ the Z-test. We will use two tailed since we are checking if the samples are statistically different

Null hypothesis H_0: The two sample mean lives are similar

Alternate hypothesis H_1: The two sample means are significantly statistically different.

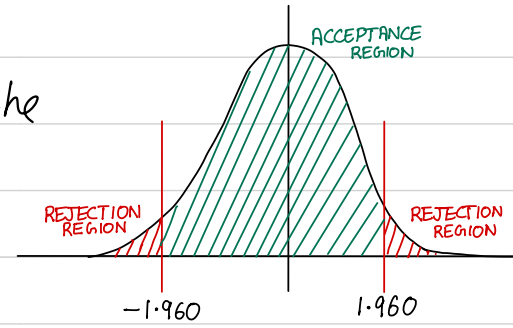
$\bar{x}_A = 20400 \quad \bar{x}_B = 21800 \quad n_A = 20 \quad n_B = 25 \quad S_A = S_B = 100$

$$Z = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} = \frac{20400 - 21800}{100\sqrt{\frac{9}{100}}} = \frac{-14}{3/10} = -46.667$$

AKHIL.S
2021MT12054

Q2 CONT..

For the two tail test at 5% significance the Z value is 1.960. The region of acceptance for the null hypothesis is between −1.96 to 1.96. The calculated Z (−46.667) is lesser than −1.96. hence we can reject the null hypothesis



Q3

x (QUIZ)	y (ASSIGNMENT)	$X = x - \bar{x}$	X^2	$Y = y - \bar{y}$	Y^2	XY
5	8.5	0.425	0.1806	−0.525	0.2756	−0.2231
4.5	9	−0.075	0.0056	−0.025	0.0006	0.0019
3	5.5	−1.575	2.4806	−3.525	12.4256	5.5519
5	9.5	0.425	0.1806	0.475	0.2256	0.2019
4.5	9	−0.075	0.0056	−0.025	0.0006	0.0019
4.5	9.5	−0.075	0.0056	0.475	0.2256	−0.0356
4.5	10	−0.075	0.0056	0.975	0.9506	−0.0731
5	9.5	0.425	0.1806	0.475	0.2256	0.2019
5	9	0.425	0.1806	−0.025	0.0006	−0.0106
4.5	9	−0.075	0.0056	−0.025	0.0006	0.0019
4.5	10	−0.075	0.0056	0.975	0.9506	−0.0731
4.5	10	−0.075	0.0056	0.975	0.9506	−0.0731
5	8.5	0.425	0.1806	−0.525	0.2756	−0.2231
3.5	9	−1.075	1.1556	−0.025	0.0006	0.0269
4.5	7.5	−0.075	0.0056	−1.525	2.3256	0.1144
5	9.5	0.425	0.1806	0.475	0.2256	0.2019
5	9.5	0.425	0.1806	0.475	0.2256	0.2019
4.5	9.5	−0.075	0.0056	0.475	0.2256	−0.0356
5	9	0.425	0.1806	−0.025	0.0006	−0.0106
4.5	9.5	−0.075	0.0056	0.475	0.2256	−0.0356
Σ 91.5	180.5		5.1375		19.7375	5.7125

$\bar{x} = 4.575$ $\bar{y} = 9.025$

AKHIL.S
2021MT12054

Q3 CONT...

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} = \frac{5.7125}{\sqrt{5.1375 \times 19.7375}} = 0.5673 \qquad (0 < r < 1)$$

The r value lies between 0 and 1, so this means that the two variables show a positive partial correlation

---

Q4

Null Hypothesis → H_0: The selling abilities of the three salesmen (X, Y & Z) are same

Alternate Hypothesis → H_1: Atleast one salesman's selling ability is different from the rest

	X	Y	Z	X^2	Y^2	Z^2
	5	3	5	25	9	25
	4	7	3	16	49	9
	7	4	5	49	16	25
	8	6	4	64	36	16
	6	-	3	36	-	9
Σ	30	20	20	190	110	84

$T = 30 + 20 + 20 = 70$

$T^2 = 4900$

$r = 3$

$n = 14$

$$CF = \frac{T^2}{n} = \frac{4900}{14} = 350$$

$$SST = (\sum X^2 + \sum Y^2 + \sum Z^2) - CF = 190 + 110 + 84 - 350 = 34$$

$$SSTR = \frac{(\sum X)^2}{n_x} + \frac{(\sum Y)^2}{n_y} + \frac{(\sum Z)^2}{n_z} - CF = \frac{900}{5} + \frac{400}{4} + \frac{400}{5} - 350 = 10$$

$$SSE = SST - SSTR = 34 - 10 = 24$$

$$MSTR = \frac{SSTR}{r - 1} = \frac{10}{2} = 5$$

$$MSE = \frac{SSE}{n - r} = \frac{24}{11} = 2.1818$$

$$F = \frac{MSTR}{MSE} = \frac{5}{2.1818} = 2.2917$$

AKHIL.S
2021MT12054

Q4 CONT..

ANNOVA SUMMARY

SOURCE OF VARIATION	SUM OF SQUARES	d.o.f	MEAN SQUARES	F-VALUE
Between Sample (PROFESSOR)	SSTR = 10	$r-1 = 2$	MSTR = 5	2.2917
Within Sample (ERROR)	SSE = 24	$n-r = 11$	MSE = 2.1818	
TOTAL	SST = 34	$n-1 = 13$		

Table Value $F(0.01_{\text{CONFIDENCE}}, d.o.f_{2,11}) = 7.206$

Table Value $F(0.05_{\text{CONFIDENCE}}, dof_{2,11}) = 3.9823$

Table Value $F(0.1_{\text{CONFIDENCE}}, dof_{2,11}) = 2.8595$

Since the F value (2.2917) is less than the table values for 1%, 5% & 10% level of significance we can accept Ho and say that all 3 salesmen have similar selling abilities at these levels of significance