ASSIGNMENT- 4 STAT101

Student Full Name:…………………………………………………..

Student ID:………………………………………………………….

CRN No:…………………………………………………………….

Section-I

State whether the following statements are True or False.

1- There may be some relationship between X and Y even though there is no linear correlation. (True)

2- The graph of regression line is not effected by the outlier. (False)

3- McNemar's test is used for frequency counts from match pairs of nominal data. (True)

4- When conducting a hypothesis test with chi-square analysis, the rejection region in a chi-square distribution is always in the upper or right tail. (True)

5- In ANOVA, if the sum of squares for error is 400, the sum of squares for treatment is 180, then total sum of squares is 480. (False)

Section-II

Circle the right answer from the answers given below.

1. If r = 0.591 and n =5, then test statistics t is equal to

   a. 1.69

   b. 1. 45

   c. 1. 269

   d. 0.91

2. To analyze data cross-classified in a contingency table, how are the degrees of freedom found?
   a. N–1
   b. Number of Rows – Number of Columns
   c. (Number of Rows) x (Number of Columns)
   d. (Number of Rows – 1) x (Number of Columns – 1)

3. The value of McNemar test statistic for the discordant pairs b=10 and c= 15 is
   a. 0.46
   b. 0.64
   c. 0.69
   d. 0.76

4- If the line of regression is given by 8X -10Y +66 = 0, then the values of b_0 and b_1 are respectively given by
   a. 6.6, 0.8
   b. 6.6, 0.6
   c. 0.6, 6.6
   d. 0.8, 6.6

5- The distribution used to find test statistic for ANOVA is …
   a. t-distribution.
   b. z-distribution.
   c. F-distribution.
   d. Chi-square distribution.

Section-III

Answer the following Essay Type Questions

1- For the following data, calculate linear correlation coefficient r and also obtain the equation of the line of regressions:

X:	1	2	3	4	5
Y:	1	1	2	2	4

Solution:

LINEAR CORRELATION COEFFICIENT

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

So we construct the following table:

x	y	x^2	y^2	xy
1	1	1	1	1
2	1	4	1	2
3	2	9	4	6
4	2	16	4	8
5	4	25	16	20
$\sum x$=15	$\sum y$=10	$\sum x^2$=55	$\sum y^2$=26	$\sum xy$=37

$$r = \frac{5\times37-15\times10}{\sqrt{[5\times55-(15)^2][5\times26-(10)^2]}} = \frac{35}{38.73} = 0.904$$

Regression equation between x and y is given by

$$\widehat{Y} = b_0 + b_1\overline{X}$$

Where

$$b_1 = \frac{\sum xy - n\bar{x}\,\bar{y}}{\sum x^2 - n\bar{x}^2} = \frac{37-30}{55-45} = 0.7$$

and

$$\hat{b}_0 = \bar{y} - \hat{b}_1\bar{x} = 2 - 0.7\times3 = -0.1$$

Thus, simple regression line is

$\hat{y}$=-0.1+0.7x

2. The following figures shows the distribution of digits in number chosen at random from a mobile directory

Digits:	0	1	2	3	4	5	6	7	8	9
Frequency:	1026	1107	997	966	1075	933	1107	972	964	853

Test whether the digits may be taken to occur equally frequently in the directory at 5% level of significance

Solution: Ho: The digits are uniformly distributed in the directory.

H1: The digits are not uniformly distributed in the directory.

O	E	$(O-E)^2$	$(O-E)^2/E$
1026	1000	676	0.676
1107	1000	11449	11.449
997	1000	9	0.009
966	1000	1156	1.156
1075	1000	5625	5.625
933	1000	4489	4.489
1107	1000	11449	11.449
972	1000	784	0.784
964	1000	1296	1.296
853	1000	21609	21.609
Total			58.542

The number of degree of freedom = 9

The tabulated value of Chi square at 5% level of significance and 9 degree of freedom = 16.919

Since, the calculated chi square is much greater than tabulated value, it is highly significant and we reject the null hypothesis.

3. Set up an analysis of variance table for the following data for three varieties of wheat each grown on 4 plots and state if the variety differences are significant.

Plot of land	Per acre production data		
	Variety of Wheat		
	X_1	X_2	X_3
1	6	5	5
2	7	5	4
3	3	3	3
4	8	7	4

Solution: Ho: There is no difference in sample means

H1: there is a difference between the sample means

$\overline{X}_1 = 6 \quad \overline{X}_2 = 5 \quad \overline{X}_3 = 4$

Mean of the sample mean= 5

Sum of squares between the samples = 8

Sum of Squares within samples = 24

Total Sum of squares = 32

ANOVA Table

Source of Variation	degree of freedom	Sum of Squares	Mean sum of Squares	F Ratio
Between samples	2	8	4	1.5
Within samples	9	24	2.67	
Total	11	32		

Tabulated F at 5% level of significance with degree of freedom being 2 and 9 is 4.26. So, this analysis support the null hypothesis of no difference in sample means.