**Assignment NO. 4 Week12 -Week14**

**Student Full Name**:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ .

**Student ID**:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ .

**CRN No**:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ .

|  |  |
| --- | --- |
| **STATISTICS**  **(STAT-101)** |  |
|  | |

**Branch:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

|  |  |
| --- | --- |
| **Total Points** |  |
| **True/False**  **MCQ**  **Short Answer** | \_\_\_\_/6  \_\_\_\_/6  \_\_\_\_/18 |
|  |  |
| **Total** | \_\_\_\_/30 |

**Good Luck**

**STATISTICS (STAT-101)**

Marks- 30

Answer all the Questions on the same question paper*.*

# Section-I

***State whether the following statements are True or False***. ***(6 marks, 1 Mark each)***

1. The term correlation refers to how two variables vary independently of one another. False
2. Outliers can increase or decrease the value of r. True
3. The linear correlation coefficient may assume any value -1 ≤ r ≤ 1. 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. ANOVA is the preferred method for finding differences among several population proportions. False
6. If P-value ≤ *α* ,reject *H*1. False

# Section-II

***Multiple choice questions. (6 marks, 1 Mark Each)***

1. In goodness of fit test. If k= 25 ( number of different categories or cells) and n =500 ( number of trials) , and then expected frequency for each cell is:
2. 1
3. 1.5
4. 20
5. 0.5
6. In contingency table. The corresponded degree of freedom for Chi-square in test of independence for 3x4 table is:
7. 12
8. 2
9. 3
10. 6
11. Analysis of variance (ANOVA) is a method for testing the hypothesis:
12. That three or more population means are equal.
13. Those at most three population means are equal.
14. That three or more population means are NOT equal.
15. That two population means are equal.
16. If the equation of regression is given by 9X -10Y +77 = 0, then the values of *b*0 ( intercept) and *b*1 (slope) are respectively given by
    1. 7.7, 0.9
    2. 7.7, 0.7
    3. 0.7, 7.
    4. 0.9, 7.7
17. If SS (Total) =100 and SS (error) =15 then the value of SS (Treatment) in 1 way ANOVA is?
18. 75
19. 115.
20. 85
21. None of the above.
22. A fitted linear regression model is. If *x =* 1 and the corresponding observed value of *y* = 11, the residual at this observation is:
23. +1
24. -1
25. 0
26. -2

# Section-III

**Answer the following Essay Type Question (18 marks, 3 Mark Each)**

1. Five pair of shoes print length and height were used to conduct a formal hypothesis test of the claim that there is linear correlation between the two variables. Use the value of r =0.591and find the appropriate test statistics method at 0.05 significance level. Also based on the result compare and conclude whether it is reject hypothesis or fail to reject the null hypothesis. ( P-value is 0.2937)

**Solution:** Since the value of r is given so we assume that the requirement for the linear correlation are satisfied.

To claim that there is a linear correlation is to claim that population linear correlation coefficient is different from 0.

Step 1: (There is no linear correlation)

Step 2: (There is linear correlation)

Step 3: Significance level is 0.05

Step 4: Given r= 0.591 and n =5 we use the test statistic formula

Degree of freedom df = n-2 = 5-2 =3

Step 5: From the given observation P-value = 0.2937 is greater than the significance level 0.05, we fail to reject null hypothesis .

1. Given the sample data below, find the F test statistic value

|  |  |  |  |
| --- | --- | --- | --- |
|  | Brand 1 | Brand 2 | Brand 3 |
| n | 16 | 16 | 16 |
| Mean | 2.09 | 3.48 | 1.86 |
| S2 | 0.37 | 0.61 | 0.45 |

**Solution:**

1. According to Berford’s law, a variety of different data sets includes numbers with leading (first) digits that follow the distribution shown in the nine row of Table 1. The bottom row lists the frequencies of leading digits of the populations of all 120 countries from New York and California combined. Test the claim that those 120 countries have populations with leading digits that follow Benford’s law.

|  |  |  |
| --- | --- | --- |
| Leading digit | Benford’s law:Distibution of leading digit | CA and NY country population |
| 1 | 30.1% | 33 |
| 2 | 17.6% | 22 |
| 3 | 12.5% | 10 |
| 4 | 9.7% | 15 |
| 5 | 7.9% | 10 |
| 6 | 6.7% | 9 |
| 7 | 5.8% | 5 |
| 8 | 5.1% | 7 |
| 9 | 4.6% | 9 |

Table 1

(From Table A-4 the Critical Value for at degree of freedom =8 is 15.507 and p-value=0.652)

**Solution:**

Requirements are satisfied:

The sample data are assumed to be a random sample. (2) The sample data consist of frequency counts. (3) Each expected frequency is at least 5.

Step 1: The original claim is that leading digits fit distribution given as Benford’s law.

We can express this claim as p1 =0.301 and p2 =0.176 and … and p9 =0.046.

Step 2: If the original claim is false, then at least one of the proportions does not have the value as claimed.

Step 3: The null hypothesis must contain the condition of equality, so we have

H0: p1 =0.301 and p2 =0.176 and … and p9 =0.046.

H1: At least one of the proportions is not equal to the given claimed value.

Step 4: The significance level is not specified, so we use common choice of

Step 5: Because we are testing a claim that the distribution of leading digits fits the distribution given by Benford’s law, we use goodness-of-fit test. The *χ*2Distribution is used for the test statistics.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Leading digit | Observed Frequency (O) | Expected Frequency (E)=np |  |  |  |
| 1 | 33 | 120\*0.301=36.12 | -3.12 | 9.7344 | 0.2695 |
| 2 | 22 | 120\*0.176=21.12 | 0.88 | 0.7744 | 0.0367 |
| 3 | 10 | 120\*0.125=15 | -5 | 25.00 | 1.6666 |
| 4 | 15 | 120\*0.097=11.64 | 3.36 | 11.2896 | 0.9698 |
| 5 | 10 | 120\*0.079=9.48 | 0.52 | 0.2704 | 0.0285 |
| 6 | 9 | 120\*0.067=8.04 | 0.96 | 0.9216 | 0.1146 |
| 7 | 5 | 120\*0.058=6.96 | -1.96 | 3.8416 | 0.5519 |
| 8 | 7 | 120\*0.051=6.12 | 0.88 | 0.7744 | 0.1265 |
| 9 | 9 | 120\*0.046=5.52 | 3.48 | 12.1104 | 2.1939 |
| Total | 120 |  |  |  | 5.958 |
|  | | | | | |

Table 2

Step 6: Table 2 shows *χ* 2 test statistics for the leading digits 1 to 9. The test statistic *χ*2 = 5.958 and the degree of freedom (k-1) = 9-1= 8. The critical value is *χ*2 = 15.507 (value given in the question which is obtained from the A-4 table at degree of freedom= 8 and significance level =0.05). The p-value is 0.652 (for *χ*2 = 15.507 and degree of freedom = 8)

Step 7: The p-value of 0.652 is greater than the significance level =0.05. There is not the sufficient evidence to reject the null hypothesis (Alsodoes not fall in the critical region bounded by the critical value of , so there is no evidence to reject the null hypothesis).

Step 8: There is no sufficient evidence to warrant rejection of the claim that 120 countries have populations with leading digits that fit the distribution given by Benford’s law.

1. Find Linear correlation coefficient between X and Y and also obtain the Regression equation for the following data:

|  |  |
| --- | --- |
| Hours(X) | Score(Y) |
| 1 | 1 |
| 1 | 3 |
| 3 | 2 |
| 4 | 5 |
| 6 | 4 |
| 7 | 5 |
| 8 | 7 |
| 8 | 8 |

**Solution:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Hours  (X) | Score  (Y) | X2 | Y2 | XY |
| 1 | 1 | 1 | 1 | 1 |
| 1 | 3 | 1 | 9 | 3 |
| 3 | 2 | 9 | 4 | 6 |
| 4 | 5 | 16 | 25 | 20 |
| 6 | 4 | 36 | 16 | 24 |
| 7 | 5 | 49 | 25 | 35 |
| 8 | 7 | 64 | 49 | 56 |
| 8 | 8 | 64 | 64 | 64 |
| **ΣX = 38** | **ΣY = 35** | **ΣX2=240** | **ΣY2=193** | **ΣXY= 209** |

The Linear correlation Coefficient

The Linear correlation Coefficient () = 0.878

The Regression equation is given by

Where the value of the slope and y intercept

Now

Therefore Regression equation is given by

1. While conducting a one-way ANOVA comparing five treatments with 10 observations per treatment, computed value for SST = 250 and MSE = 3 given. Find the value of F and also construct the ANOVA table?

**Solution:**

Number of treatment (k) =5, Number of total treatment (n) =10,

SS (Total) =250, and MS (Error) = 3

The missing data are: N= n = 5x10=50

Numerator: df1= =5-1 =4

Denominator: df2= =50-5=45

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | Df | SS | MS | F |
| Treatment | 4 | 115 | 28.75 | 9.583 |
| Error | 45 | 135 | 3 |  |
| Total | 49 | 250 |  |  |

1. Find the expected frequency for the given Contingency Table, where the Observed frequencies are 88 and 10. Use a 0.05 significance level to test the claim that getting an infection (cold) is independent of the treatment group. What does the result indicates about the effectiveness of Echinacea as a treatment for colds?

**Solution:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Treatment Group | | | |  |
|  | Placebo | Echinacea:20%extract | Echinacea:60%extract | Total |
| Infected | 88 | 48 | 42 | 178 |
| Not Infected | 15 | 4 | 10 | 29 |
| total | 103 | 52 | 52 | 207 |

The “grand total” is the sum of all frequencies in the table, which is 207.

At the observed frequency 88, Row total =178 and Column total =103.

The expected frequency

At the observed frequency 10, Row total =29 and Column total =52.

The expected frequency

There is a discrepancy between *O* = 88 and *E* = 88.570, and *O* = 10 and *E* = 7.285 such discrepancies are key components of the test statistic.

Requirements are satisfied: randomly assigned to treatment groups, frequency counts, expected frequencies are all at least 5.

Step 1:

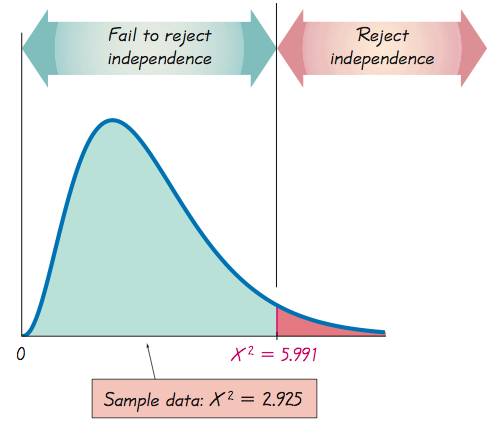
*H*0: Getting an infection is independent of the treatment

*H*1: Getting an infection and the treatment are dependent

Step 2: Significance level is *α* **=** 0.05.

Step 3: Contingency table: *χ*2 distribution is used

Step 4: The critical value of *χ*2 = 5.991 is found from Table A-4 with *α* = 0.05 in the right tail and the number of degrees of freedom given by  
(*r* – 1)(*c* – 1) = (2 – 1)(3 – 1) = 2.



Step 5: Because the test statistic does not fall within the critical region, we fail to reject the null hypothesis of independence between getting an infection and treatment. It appears that getting an infection is independent of the treatment group. This suggests that Echinacea is not an effective treatment for colds.