**ASSIGNMENT 4**

**Submitted By**

**Zajiba Sadia Islam**

**Student ID: 501279357**

**Group-09**

**Submitted for**

**Program: MEng in Biomedical Engineering**

**Course: Stats for the Health Sciences**

**BP8101, Winter 2024**

**Class Timing: Wednesday 3:00-6:00 PM**

**Analysis of Variances**

1.

a. H0: The means for the final exams are the same for all statistics class delivery types.

Or, H0: μ1 = μ2= μ3

b. Ha: At least one mean is different from the others.

c. df(n) = Number of groups,k - 1 = 3 - 1 = 2

df(d) = Total number of observations(n) - Number of groups(k) = 16 - 3 = 13

d. F-distribution will be used for the test.

e. Calculation of test statistic:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 20.17321  (SST) | 2 | 10.08661  (MST) | 0.638815 | 0.54371 | 3.805565 |
| Within Groups | 205.2643  (SSE) | 13 | 15.78956  (MSE) |  |  |  |
| Total | 225.4375 | 15 |  |  |  |  |

f. The p-value has been calculated using excel and found to be 0.54371.

g.

A graph of a curve

Description automatically generated

h. Since the p-value associated with an F-statistic of 0.638815 is greater than 0.05. Hence, the null hypothesis is failed to be rejected.

2.

Degrees of Freedom (df):

Mixture: Number of groups - 1 = 6 - 1 = 5

Error: Total number of observations - Number of groups = 26 \* 6 - 6 = 150

Total: Total number of observations - 1 = 26 \* 6 - 1 = 155

Sum of Squares (SS):

Total SS = Sum of squares for Mixture + Sum of squares for Error

Mean Square (MS):

Mixture: MS = Sum of squares for Mixture / df(Mixture)

Error: MS = Sum of squares for Error / df(Error)

F-ratio:

F = MS(Mixture) / MS(Error)

Table:



Based on these calculations, it is seen that F-ratio is very large, suggesting that there is a significant difference in electrical resistivity among the different low-permeability concrete bridge deck mixtures. The critical F value at 0.05 significance for (5,150) df is 2.46. Since the obtained F-ratio exceeds the critical value, we will reject the null hypothesis. At least one of the mixtures has a different electrical resistivity.

3.

a. H0: There is no significant difference in effectiveness among the doses.

b. Ha: There is a significant difference in effectiveness among the doses.

c. df(n) = Number of groups - 1 = 6 - 1 = 5

df(d) = Total number of observations - Number of groups = 30 - 6 = 24

d. F-distribution will be used for the test.



e. The test statistic is F=5.6

f. The p-value is found to be 0.0015 for f value of 5.6 and degree of freedom 5, and 24.

g.

A graph on a piece of paper

Description automatically generated

h. Since the p value is very low compared to significance level of 0.05 and f value is higher than critical f value, the null hypothesis can be rejected with enough evidence. Hence, it can be said that there is significant difference in effectiveness among the doses.Top of Form

Top of Form

**Regression and Correlation**

4.

The given dataset:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| x | 6.7 | 5.1 | 4.2 | 3.3 | 2.1 | units: mm Hg/10 |
| y | 43.6 | 32.9 | 26.2 | 16.2 | 13.9 | units: mm Hg/10 |

The linear regression model is:

y=mx+b

Where, m is the slope and b is the y-intercept.

Formula for slope is:

A mathematical equation with numbers and symbols

Description automatically generated

So, m=47.56

Formula for y-intercept



b = −176.95

So, the linear regression model for the given dataset is:

Y=47.56x−176.95

For x=4, y^=13.29

The standard error is

A square root of a mathematical equation

Description automatically generated

SE=13.83

From the t-distribution table, t=3.182 for a 90% confidence interval with 3 degrees of freedom. Hence, margin of error, Me=t\*SE=44

Confidence interval:



=(-30.71,57.29)

So, the 90% confidence interval for y when x=4.0 is approximately (-30.71, 57.29).

5.

a) The independent variable is the year the state entered the Union, and the dependent variable is the number of letters in the state name.

b)

c) From inspection of the scatter plot, there is no clear relationship between the number of letters in a state name and the year the state entered the Union. The points are scattered and do not seem to follow a linear trend.

d) Let, the number of letters be (y) and the year entered the Union be (x).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year entered the Union(x) | #letters in name(y) | -)^2 | -)^2 | - | - | -)\*( - |
| 1819 | 7 | 407.3058 | 0.07438 | -20.18181818 | -0.27272727 | 5.504132 |
| 1876 | 8 | 1355.579 | 0.528926 | 36.81818182 | 0.727272727 | 26.77686 |
| 1959 | 6 | 14356.4 | 1.619835 | 119.8181818 | -1.27272727 | -152.496 |
| 1846 | 4 | 46.4876 | 10.71074 | 6.818181818 | -3.27272727 | -22.314 |
| 1788 | 8 | 2619.579 | 0.528926 | -51.18181818 | 0.727272727 | -37.2231 |
| 1821 | 8 | 330.5785 | 0.528926 | -18.18181818 | 0.727272727 | -13.2231 |
| 1787 | 9 | 2722.942 | 2.983471 | -52.18181818 | 1.727272727 | -90.1322 |
| 1803 | 4 | 1309.124 | 10.71074 | -36.18181818 | -3.27272727 | 118.4132 |
| 1788 | 13 | 2619.579 | 32.80165 | -51.18181818 | 5.727272727 | -293.132 |
| 1896 | 4 | 3228.306 | 10.71074 | 56.81818182 | -3.27272727 | -185.95 |
| 1848 | 9 | 77.76033 | 2.983471 | 8.818181818 | 1.727272727 | 15.2314 |
|  |  |  |  |  |  |  |
| 1839.182 | 7.272727 | 29073.64 | 74.18182 | -6.82121E-13 | 2.66454E-15 | -628.545 |
| Mean | Mean | Sum | Sum | Sum | Sum | Sum |

∑(-)^2=29073.64

∑-)^2=74.18182

∑-)\*( --628.545

Slope:

b=

b=-0.02162

y-intercept:

a=yˉ​−bxˉ

a= 47.03415778

The equation of the least squares line is:

y^​=a+bx

=47.034-0.022x

e) Correlation coefficient, r=/√

r=-0.428

This indicates moderate to low inverse relationship between year and number of letters in name.

f) Using the equation of least square line, if it entered the Union in 1900 there would be

47.034-0.022\*1900=5.234 or 5 letters would be present.

In 1940, 4.354 or 4 letters would have been present.

g) Since the points in the scattered plot are scattered and do not follow a linear trend, a line may not be the best way to fit the data.

h) If a new state entered union this year, it would have 47.034-0.022\*2024=2.5 or 2 letters. This prediction may not be accurate because the relationship between the number of letters in a state name and the year entered the Union is not linear.

6.

a)

Given,



Sample correlation formula:

A black and white math equation

Description automatically generated

Plugging in the values, we obtain r=0.966

The sample correlation coefficient r is approximately 0.771. This value indicates a strong positive linear relationship between shear force and percent fiber dry weight.

b) Given that there is strong positive correlation between shear force and dry fiber weight, a larger value of shear force will create larger percent dry fiber weight.

c) If shear force is expressed in pounds instead of kilograms, the value of r would not change. This is because the correlation coefficient is a measure of the strength and direction of a linear relationship between two variables, and it is not affected by the units in which the variables are measured.

d) Squaring the correlation coefficient:

r^2=0.933.

So, approximately 93.3% of the observed variation in percent fiber dry weight could be explained by the linear relationship with shear force.

e) Null hypothesis: There is no linear association (r=0),

Alternative hypothesis: There is a positive linear association (r>0).

Test Statistics, t=r\*√(n-2)/(1-r^2)=14.94

n−2=18−2=16 degrees of freedom.

The critical value for a one-tailed test at α=0.01 is approximately 2.583. Since t=14.94 is greater than 2.583, the null hypothesis is rejected. There is sufficient evidence to conclude that there is a positive linear association between shear force and percent fiber dry weight at the 0.01 significance level.

7.

Sample correlation formula:

A black and white math equation

Description automatically generated

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | x | y | x^2 | y^2 | xy |
|  | 7.98 | 0.28 | 63.6804 | 0.0784 | 2.2344 |
|  | 24.52 | 0.65 | 601.2304 | 0.4225 | 15.938 |
|  | 12.47 | 0.32 | 155.5009 | 0.1024 | 3.9904 |
|  | 6.92 | 0.27 | 47.8864 | 0.0729 | 1.8684 |
|  | 24.11 | 0.81 | 581.2921 | 0.6561 | 19.5291 |
|  | 35.71 | 0.57 | 1275.204 | 0.3249 | 20.3547 |
|  |  |  |  |  |  |
| Sum | 111.71 | 2.9 | 2724.794 | 1.6572 | 63.915 |

r=0.77

Test Statistics, t=r\*√(n-2)/(1-r^2)=2.42

The critical t-value for a significance level of 0.05 with df=6−2=4 is approximately 2.776.

Since the calculated t-value (2.42) is less than the critical t-value (2.776), the null hypothesis is not rejected. There is no significant linear relationship between stiffness and thickness of the flame-retardant fabric samples at the 0.05 significance level.

Top of Form