Statistics MAT268

Contents

0.1 Chapter 10. Linear Correlation and Regression

0.1.1 Correlation

Linear Correlation Coefficient Formula:

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

Testing the coefficient with T test .

$$t_{test} = r\sqrt{\frac{n-2}{1-r^2}}$$

Obtain critical value as a two-tailed t-test.

$$df t_{crit} = n - 2$$

0.1.2 Regression

Linear Regression.

$$y' = a + bx$$

Calculating the y-intercept and slope.

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

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

0.1.3 Coefficient of Determination and Standard Error of the Estimate

Coefficient of Determination. Can also be calculated using correlation coefficient. Coefficient of determination is also a percentage.

$$r^2 = \frac{\text{explained variation}}{\text{total variation}}$$

Coefficient of Nondetermination

$$1 - r^2$$

Standard Error

$$s_{est} = \sqrt{\frac{\Sigma(y - y')^2}{n-2}}$$

Confidence Interval: Using the t-test values from before

$$y' - t_{\frac{\alpha}{2}} s_{est}\sqrt{1 + \frac{1}{n} + \frac{n(x - \bar{X})^2}{n\Sigma x^2 - (\Sigma x)^2}} < y < y' + t_{\frac{\alpha}{2}} s_{est}\sqrt{1 + \frac{1}{n} + \frac{n(x - \bar{X})^2}{n\Sigma x^2 - (\Sigma x)^2}}$$

0.2 Chapter 11. Other Chi-Square Tests

0.2.1 Test for Goodness of Fit

Determining if there is a statistical difference in your categorical data, you can perform a chi-square test. Given data set like this:

	Vanilla	Chocolate	Strawberry	Mango
Observed	38	50	28	44

Determine following:

- H0: There is no difference in ice cream flavor preferences
- H1: There is a difference in ice cream flavor preferences
- significance is given and a = 0.05 will be used

Using following degrees of freedom chi squared critical value of 7.815 is obtained

$$df\text{F table} = k - 1$$

Determine Expected by given k number of columns

$$E = \frac{\Sigma x}{k}$$

	Vanilla	Chocolate	Strawberry	Mango
Observed	38	50	28	44
Expected	40	40	40	40

Calculated Chi-Square

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

This example obtains chi squared test value of 6.6 and since 6.6 ¡ 7.815, we fail to reject null hypothesis.

0.2.2 Contingency Tables

If comparing for Independence.

	Vanilla	Chocolate	Strawberry
Men	100	80	20
Women	50	120	30

Calculate Expected of each cell by:

$$E = \frac{\text{row sum} * \text{column sum}}{\text{grand total}}$$

and perform hypothesis testing appropriately.

0.3 Chapter 12. Analysis of Variance

0.3.1 One-Way Analysis of Variance

Calculate the F-test value with following.

$$F = \frac{\text{variance between groups}}{\text{variance within groups}}$$

Calculate grand mean where X is a cell in the table and N is the total number of cells.

$$\bar{X}_{GM} = \frac{\Sigma X}{N}$$

Variance between groups:

$$s_B^2 = \frac{\Sigma n_i(\bar{X}_i - \bar{X}_{GM})^2}{k-1}$$

Variance within groups:

$$s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)}$$

Where Variance of the sample:

$$s_i^2 = \frac{n(\Sigma X^2) - (\Sigma X)^2}{n(n-1)}$$

To find F test Value:

$$d.f.N = k - 1$$

$$d.f.D = N - k$$

0.3.2 Scheffe and Tukey Test

Use Scheffe test to compare two different means

Obtain F prime as the new F test number

$$F' = (k-1)(C.V.)$$

Calculate F test between two different samples

$$F_{test} = \frac{(\bar{X}_i - \bar{X}_j)^2}{s_B^2(\frac{1}{n_i} + \frac{1}{n_j})}$$

Tukey test can be used to test if the samples have the same size.

$$q_{test} = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{s_W^2}{n}}}$$

Where the critical value of Tukey table is calculated using

$$v = N - k$$

where N is the total number of samples.

1 References

Elementary Statistics - Allan G Bluman