Data interpretation practicum

Science, Statistics

Correlation Introduction Correlation is used to investigate the extent of linear relationship between any two variables (Cohen, Cohen, West, & Leona 48). In this analysis, the association between injury rate and number of hours worked will be investigated. Under this test, it is assumed that the observations are independent. A correlation test is the most preferred as it will reveal whether there is a corresponding increase (or decrease) in the number of injuries as compared to the number of hour worked. Regression could also be used in place of correlation as it would show how the number on injuries (dependent variable) changes as a result of increased number of working hours (independent variable). A regression procedure would further help in predicting the injury rate based on working hours. However, discriminant analysis cannot be used.

Hypothesis

The correlation test will test the level of association between injury rate and number of hours worked will be investigated consequently, our hypotheses are as follows:

Null Hypothesis, H0: Injury rate and hours worked are correlated

Alternative Hypothesis, H1: Injury rate and hours worked are not correlated

A scatterplot of the data is shown below:

From this plot, it is seen that the injury rate is inversely proportional to hours worked, i. e. the two variables exhibit a negative correlation.

Descriptive statistics of the data is shown below:

Descriptives

Statistic

Std. Error

Hours Worked

Mean

49960. 78

2183. 070

95% Confidence Interval for Mean

Lower Bound

45575. 96

Upper Bound

54345. 61

Std. Deviation

15590. 236

Minimum

10400

Maximum

93600

InjuryRate

Mean

15. 175696

2. 4469443

95% Confidence Interval for Mean

Lower Bound

10. 260864

Upper Bound

20. 090528

Std. Deviation

17. 4746773

Minimum

. 0000

Maximum

76. 9231

The average working hours in the three states is 2183. 07 hour while the average injury rate in the three states is 2. 4446. The true population mean for average working hours in the three states is bound between 45575. 96 and 54345. 61 while true injury rate mean for average working hours in the three states is bound between 10. 26 and 20. 09.

Correlation Analysis

The output for the correlation test is shown below:

Correlations

Hours Worked

InjuryRate

Hours Worked

Pearson Correlation

1

-. 636**

Sig. (2-tailed)

. 000

N

51

51

InjuryRate

Pearson Correlation

-. 636**

1

Sig. (2-tailed)

. 000

N

51

51

**. Correlation is significant at the 0. 01 level (2-tailed).

From this output, the correlation coefficient between hours worked and injury rate is -0. 636. This implies that as work hours increases, injury rate reduces (p-value ~ 0. 000). The test is significant, hence we reject the null hypothesis and conclude that the two variables are correlated. This value is consistent with the observation from a scatterplot of the two variables shown above.

A possible explanation for the observation made is that only a few injuries are normally witnessed, hence, increasing the hours worked does not necessarily lead to an increase in the number of injuries. Since injury rate is obtained by dividing the number of hours worked by the number of injuries, the values reduces as hours worked increases. The value of the correlation coefficient does not imply that increasing the number of working hours results into less injuries.

Reference

Cohen, J., Cohen, P., West, S., & Leona, S. A. (2002). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). NY: Psychology Press.