Non parametric and chi-square distribution

ASSIGN BUSTER

Brief Summary

I worked for a logistic company. My majorresponsibilitywas in charge of the storage and transportation of parts of cars between two areas, which are about 1400 miles apart. One of my jobs is collecting the goods from suppliers and arranging the trucks to deliver them. There are five truck drivers, and each of them is assigned to deliver on each weekday throughout a whole year. Before the delivery, we will check the quality of the goods and Make sure that there are no damaged goods.

When arriving at the destination, the staff will check the goods again and record the damaged goods that occurred in transit. At the end of every month, we will pay for the compensation according to the number of the defective goods. In order to reduce the number of the damaged goods during the delivery, I want to identify the reasons why they are damaged. In this study, I want to find out that whether some drivers are more prone to make the goods damaged during their delivery.

Variable to be measured: Two variables are to be measured. The first variable is just the five truck drivers, and the second one is the quality of the goods after the delivery.

Determination of Population

Population in this case is defined as the all goods delivered from Tianjin area to Guangzhou area. Statistical method: To analyze relationship between the two variables above which are both nominal in terms of data type, I decide to use Chi-squared test of a contingency table. Sample Selection:

The information about delivery is recorded in our computer system, including the delivery date, name of the driver, the number of damaged goods and so on. I take out the data about 52 weeks during the previous year and record them into the following table: | Quality| | Truck Driver| Passed| Damaged| Total| Driver A| | | | Driver B| | | | Driver C| | | | Driver D| | | | Driver E| | | | Total| | | | Hypothesis: The objective is to describe whether there is a relationship between the five drivers and the number of damaged goods.

The null hypothesis will specify that there is no relationship between the two variables: H0 : The two variables are independent The alternative hypothesis specifies one variable affects the other, expressed as: H1 : The two variables are dependent Use the formula: ? 2= i= 1k(fi-ei)2ei ? =(r-1) (c-1) which calculates the test statistic. Or use the Excel by importing the data into the function of Data Analysis Plus, Contingency Table. Then I could acquire chi-squared Stat, p-value. The number of degrees of freedom v=(r-1)(c-1)=(5-1) (2-1)= 4.

If I employ a 5% significance level, the rejection region is X2> X2 a, v = X2 . 05, 4 = 9. 49 Comparing the results, if the p-value is greater than 9. 49, there is not enough evidence to infer that there is a relationship between the five drivers and the number of damaged goods; if the p-value is not greater than 9. 49, I can reject the null hypothesis in favor of the alternative, which means there is a relationship between the five drivers and the number of damaged goods. So I can reduce the number of damaged goods through improving the drivers’ conditions.