TO: Professor Johnson-Hall

FROM: Yaseen Lotfi

SUBJECT: PLE Assignment 3

DATE: 4/17/17

Problem Statement

Chapter 7: Statistical Inference

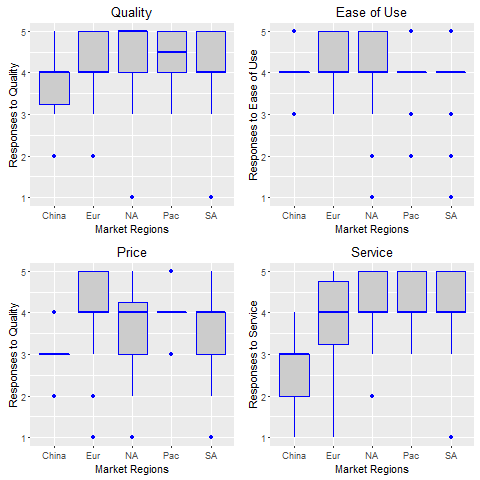
Chapter 8: Regression Analysis

**Analysis & Methodology: Statistical Inference**

The following points of interest are outlined in detail as follow:

1. Are there any significant differences between the elements of the 2014 Customer Survey
2. Assess the significance of on-time delivery improvements since 2010
3. Have the number of defects changed significantly over the past five years
4. Why new processes are not better than the current process for transmission costs
5. Are gender, educational level, and residence significant factors for employee retention
6. *Differences of Variance Across Survey Categories*

PLE wants to know how survey responses differ across categories, alluding to analysis of the variance (ANOVA). It is a technique for testing multiple mean hypotheses at once, but it makes critical assumptions about the data. Assumptions include that the data has a Gaussian (normal) distribution and the variance across variables must be the same. Before jumping into ANOVA, the data was visualized and then prepared for two statistical tests. These tests inspect the data’s homogeneity of variance (homoscedasticity) to ensure accurate analysis. The statistical software environment, R, supports native, non-parametric tests for homoscedasticity; the Bartlett test was used in this case, but others output similar results.

First, the boxplots show a large degree of variation based on a cursory look at the survey data. Considering that the data is not continuous, the boxplots look stratified along the result values. That said, there are many outliers in the data in which a point is beyond the outer quartiles. Furthermore, there is no consistent pattern that one could immediately see. In other words, the survey seems to be all over the place region to region and, more importantly, category to category.

Next, the data was prepared for the statistical tests. First, region labels were dropped from the data as PLE is interested in the differences across survey topics and not market regions. The survey results were then vectorized wherein a 200x4 matrix is flattened to an 800x1 column vector. A second array is constructed which defines the classes for the test. In this case, the survey categories result in another 800x1 column vector containing 200 elements for each class. At this point, the Bartlett test can output a chi-squared value and p-value for its hypothesis test on variance homogeneity. These metrics are compared against a normally distributed chi-square value at three degrees of freedom for a 95 percent confidence interval.

*(b - e) Basic Hypothesis Testing Methodology*

Other than the first problem, basic hypothesis tests were constructed. The hypothesis tests all followed a similar analytical structure. First, simple data import and cleaning was conducted. For the most part, this meant vectorizing columns, creating binary labels where necessary, and identifying test samples and hypothesis samples. Then, sample and hypothesized means were computed. Depending on whether population variation was known, either the sample or population standard deviations were also calculated. If and only if the population variation was known **and** the test sample size was greater than or equal to 30, the z-score was used to compute the p-value. Otherwise, the t-score was used.

Several equations were used to determine either the z or t statistic for the sample and hypothesized means. The first test is concerned with proportions, so it uses a different equation which is presented in its overview. To find the test statistic that is used for computing the p-value, the following equations were used:

Unknown Variance and

Known Variance and

The only difference is if population variance is known or not, in which case either a sample standard deviation, , or population standard deviation, is used. The other important factor is sample size, n, which determines the test-statistic as well.

1. *Hypothesis Test: On-Time Deliveries*

The hypothesize claims that the mean fill rate in 2014 is greater than the historical mean fill rate. This statement will test if the proportion of on-time deliveries is higher in 2014 than in previous years. Due to the inherent directionality of the hypothesis, it is structured as a lower-tail test formally written as:

In this case, p̂ is the sample proportion and p0 is the hypothesized proportion. Each proportion is computed as the mean fill rate for each column-wise vector from the original data.[[1]](#footnote-1) The z-statistic can be calculated with the following equation:

The z-statistic is used to calculate the p-value of the lower-tail. If the probability area under the curve is greater than or equal to a confidence interval of 0.05, then the initial hypothesis is statistically significant. Otherwise, mean fill-rate improvement is not clearly founded.

*(c) Hypothesis Test: Defects Over Time*

PLE is interested in any significant changes in terms of defects over the last five years. For this test, the data is manipulated like the last case. Two subsets are created based on year. The hypothesized mean therefore covers 2010 to 2013 and the sample mean reflects 2014. It makes sense to frame this test with some directionality, as well. Thus, we hypothesize that if a positive change exists, the mean defect count in 2014 is less than the historical mean defect count. This upper-tail test is formalized as:

Although the population variance is known, the sample size, , is well under 30 observations. As outlined in the previous methodology section, it is best practice to use the t-statistic. As such, the sample standard deviation, , is also used. The p-value can now be computed based on the t-statistic and a 95 percent confidence interval as well. Again, if the p-value is greater than 0.05, the initial hypothesis is grounded, indicating that defects over time decreased significantly over the last five years.

*(d) Hypothesis Test: Transmission Costs*

The engineering team was unable to reach a conclusion if alternative process costs were more effective than the current one. To confirm their findings, two separate hypothesis tests can be constructed. If the sample means of Process A and B, and , are equal to the hypothesized mean of the current process, , then they are not statistically better. There is no inherent directionality in this test, so this is a two-tailed test formally written as:

This is a two-step process where the following parameters are calculated twice, once for each hypothesis. The z-statistic was used because the sample size, , was equal to thirty and population variance,, could be established. Two standard deviations were calculated on separate vectors; one of the current data with Process A and another with Process B instead. The p-values were then computed and compared against a confidence interval of 95 percent. If either p-value was greater than 0.05, then there is reason to believe that the corresponding processes are not statistically different from the current process.

(e) Hypothesis Test: Employee Retention

The final hypothesis test is constructed in a similar fashion as the previous one. Instead of two tests, three were conducted based on each factor.

Two-Tailed Tests where xbar = mu0

The sample mean of the target variables are equal to the mean number of years for the PLE employees.

**Assumptions and Limitations**

The most obvious assumption made is that the confidence interval was held constant across all analysis. This is mainly done for simplicity and lack of domain knowledge. The tests that were conducted are highly reproducible and so they can be repeated by someone with more expertise. One might prefer to use a higher or lower confidence interval depending on the application.

Drilling into the previous assumption, an obvious limitation exists for the transmission cost analysis. It is not clear if the engineering team also used the same confidence interval. These results can potentially be misleading as the analysis attempts to reproduce their findings with minimal input into the methods that they implemented.

One-way ANOVA makes significant assumptions about the data and underlying variances that it is assessing. First, it assumes that the data was randomly and independently measured. This is a fair assumption to make about the provided data given that it is out of the analysis’ control. Second is that the data normally distributed, which can be easily tested by created basic histogram charts. The most important assumption that it makes is that the variables have equal variances, meaning that the data is homoscedastic. The primary limitation of ANOVA is its sensitivity to the last assumption of equal variances. When violated, it can affect the confidence interval and beta of the test, therefore influencing the conclusion.

**Conclusions and Recommendations**

1. *Differences of Variance Across Survey Categories*

The Bartlett test output a p-value of close to zero suggesting the variance of each factor is not homogenous. It also provides a chi-squared value of 29.3, which is well above the chi-squared value of 7.81 at three degrees of freedom on a normal distribution with a 95 percent confidence interval. As such, the variance of each category differs significantly and further analysis with ANOVA was not pursued due to its underlying assumptions being violated by the data.

*(b) Hypothesis Test: On-Time Deliveries*

In 2014, the average sample proportion, p̂, was 99.07 percent. The hypothesized average proportion, p0, was 98.69 percent. The z-statistic is 0.1138 which is used to calculate the p-value. The p-value turns out to be 0.5453 which is well over the 0.05 confidence interval threshold. Therefore, there is significant statistical cause to rely on the initial hypothesis claiming that the proportion of on-time deliveries improved since 2010.

*(c) Hypothesis Test: Defects Over Time*

The average sample defect count in 2014, , is 496.25 and the average hypothesized mean of the historical sample, , is 779.69. The sample standard deviation, , is 54.22 and sample size, , is only 12. The t-statistic is therefore -18.11 which determines the p-value of 0.9999. In this case, the test suggests that defects have decreased over time because the p-value is above 0.05. The outcome is confirmed by checking the critical values, as well. The t-score for an alpha value of 0.05 is 1.796 which is greater than the calculated t-statistic of -18.11.

The mean processing costs for both options, and , are $285.50 and $298.43, respectively. The mean cost for the current process, , is $289.60. The population standard deviations, and , are $55.59 and $35.31. With a sample size of exactly 30, the z-statistics for each test are -0.404 and 1.370, respectively. Finally, the p-values for both were 0.6892 and 1.819, respectively. Both p-values are well beyond a significance value of 0.05, suggesting neither alternative process is better than the current one. This test confirms the findings of the engineering team.

1. Note that the fill rate is calculated by dividing the total orders by the difference of total orders and on-time orders. [↑](#footnote-ref-1)