

Probability Theory and Introductory Statistics

ALY 6010

Assignment 4

Title: Analysis of sales data in department stores using confidence levels

Rahul Iyer

Professor Name: Amin Karimpour

1. **CONFIDENCE LEVEL & INTERVAL**

**Introduction:**

The analysis is being performed on the sales data obtained from different department stores. The records in each of this data set are analyzed to draw a particular conclusion, either supporting or rejecting an underlying assumption (hypothesis) about the population. We are testing the data to validate different hypothesis based on confidence level, sample size and other parameters. In addition, we are determining the impact of increasing confidence level on the interval width and associated variables.

**Analysis:**

We begin the analysis by calculating sample size (n), standard deviation of the sample (s) and the value of z. The value of z can be calculated with the help of confidence level.

For instance, when confidence level= 90% (or 0.9)

α= (1-0.9)/2

α = 0.05

The value of z can be determined by using the table containing z values for different probabilities in a standard normal distribution curve. We estimate the value of z, corresponding to the probability of 1- α=0.95. Therefore, we obtain z (critical)=1.645.

Now, Margin of Error (M.E.) is given by the formula:

Substituting the values, we can obtain the margin of error. Consequently, this approach may be implemented in the reverse order to calculate sample size if the margin of error is known or assumed. This is particularly useful in determining the sample size for a very small margin of error. Rearranging the terms, we can calculate sample size (approximate to the next integer value).

However, this approach may have to be altered a bit if the parameter does not follow the standard normal distribution. In case of T-distribution, the process remains the same with the only difference that the z-table is no longer used and instead a t-table is utilized, following T-distribution.

In case of Chi-squared distribution, the parameters are calculated differently. Sample size (n) is used to determine degree of freedom. Degree of freedom = n – 1. Further, proportion of sample (p-hat) is evaluated by determining the probability of satisfying the hypothesis from the sample size. Consequently, margin of error (M.E.) is calculated by using the equation:

**Conclusion:**

We can use the margin of error as a foundation to estimate the confidence interval’s upper and lower limit by adding and subtracting it from the sample parameter being analyzed (for example- mean). The difference between the upper and lower limits represents the interval width. We observe that this interval width increases with an increase in confidence level. This holds true across all the sales data.

Let us take the first sales report, the interval width rose from 1666.02 for 90% confidence to 1985.19 for 95% and peaked at 2608.98 for 99% confidence level.

1. **HYPOTHESIS TESTING**

**Introduction**

Null hypothesis is always the opposite of the hypothesis under test. For instance, if the hypothesis states that the mean, µ>8000, then Null hypothesis would be µ<=8000 and correspondingly, the alternate hypothesis would be µ>8000. It starts with the assumption that the null hypothesis is true and different parameters are calculated to lend support or disapprove this hypothesis. If considerable evidence/data is collected to reject the assumption, the null hypothesis is rejected else we fail to reject the null hypothesis.

**Analysis**

The analysis begins by calculating the different parameters. In the mentioned case, sample size, mean and standard deviation would be calculated to obtain the value of z. Consequently, the margin of error can be found out by the difference between the sample parameter and the hypothesized parameter. In this example,

µsample=8894 & µhypothesis=8000

Therefore M.E. = 8894-8000

=894

This can be used to determine the value of z using the formula:

Where, n= size of the sample

& s= standard deviation of the sample

Once, we obtain the value of z, we can calculate the p-value using excel functions such as NORM.S.DIST(..) or CHISQ.DIST(..) depending on the distribution of that parameter as well as whether it is right-tailed, left-tailed or two-tailed test.

Comparing the values of p-value with α & the test-statistic value with critical value, we may infer whether we must reject the null hypothesis or fail to reject the null hypothesis.

**Conclusion:**

There are two grounds to reject the null hypothesis:  
If the p-value < α, we reject the null hypothesis

OR

If the test-statistic value is greater than the critical value, we reject the null hypothesis.

In our case, the z-value (critical)=1.64, α=0.05 and the calculated values such as test-statistic and p-values are 1.77 and 0.04 respectively.

In this case, the p-value is less than the value of α. At the same time, the test-statistic value is more than the critical value. Therefore, we must reject the null hypothesis. This means that the original assumption may or may not be correct: “The population mean may or may not be greater than 8000”. However, if we failed to reject this hypothesis, we would have committed a Type-II error (also known as False negative). It is a situation where we fail to reject the hypothesis even though we were supposed to reject it.

**REFERENCE:**

Lane, D.M. (n.d.). T-Distribution. Retrieved December 2, 2018 from <http://onlinestatbook.com/2/estimation/t_distribution.html>

Stephanie. (2012, September 3). Normal Distribution Probability in Excel: All Versions up to 2016. Retrieved December 1, 2018 from <https://www.statisticshowto.datasciencecentral.com/how-to-calculate-normal-distribution-probability-in-excel/>

Sullivan, L. (n.d.). Hypothesis Testing - Chi Squared Test. Retrieved December 1, 2018 from <http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_HypothesisTesting-ChiSquare/BS704_HypothesisTesting-ChiSquare_print.html>