

Probability Theory and Introductory Statistics

ALY 6010

Assignment 5

Title: Confidence Interval & Hypothesis Testing of Differences: Two-sample Tests

Rahul Iyer

Professor Name: Amin Karimpour

**INTRODUCTION:**

Often, we may have to make conclusions based on two independent samples. In such cases, the two sample tests are implemented to increase the reliability and accuracy of the estimate. Depending on the sample size and whether or not we know the standard deviation associated with that sample, the Z-distribution or T-distributed may be used. For large samples, typically greater than 30 with known standard of deviation: we use Z-distribution. If any of the samples are small (less than 30) with unequal and unknown standard deviation, the T-test hypothesis is performed to test different conditions for the population mean. It may be used to quantitatively compare the quality of product, manufactured from two different factories.

On the other hand, certain cases have to test the difference between two dependent paired samples. In those conditions, we use T-test by calculating a variable D, which is the mathematical difference between two samples (X1-X2). Here, the sample size is equal n1=n2=n and the degrees of freedom is given by df=n-1. It may be applied to test the effectiveness of a medication before and after administering a pill.

Z-test hypothesis can also be used to test the hypothesis for difference of two population proportion (p1-p2); p1 is the proportion of success in sample 1 while p2 is the proportion of success in sample p2. Mathematically, p1=x1/n1 & p2=x2/n2. X1, X2 is the size of sample that meets the success criterion for samples 1 and 2 of sample size n1 and n2. The application could be to conclude on the performance of male and female students with GPA greater than 3.4.

F-distribution may be performed to test the difference/ratio of two population variances/standard deviation, However, the two samples must be independent, selected from separate normally distributed population. An example of F-distribution is the variation in climatic conditions such as humidity in two different cities.

The procedure for testing all the hypothesis remains the same with few variations in the formula which shall be discussed below.

**ANALYSIS**

PART I:

Let us explore the first data sheet and perform hypothesis testing to test if the average company’s sales is greater than the competitor’s average sales at 0.05 level of significance. We begin the null hypothesis, assuming that the company’s average sales is less than or equal to the competitor’s average sales, opposite of the claim.

The Null Hypothesis: µ1-µ2<=0

Alternate Hypothesis (Claim): µ1-µ2>0

It is a right-tailed problem.

Now, ***α=0.05*** (given)

Solution:

First, we calculate the size(n), mean(X) and variance of samples(S2),

n1=46

n2=34

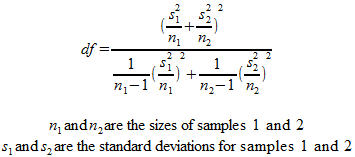
X1= 7692.282

X2= 7012.994

S12= 3289016.2

S22= 7283256.5

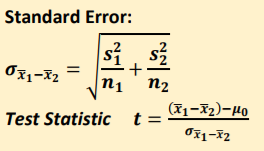
Now, we calculate the degrees of freedom (df):

**

After substituting values, we obtain:

df=54.27~55

Test statistic (t) and standard error (also known as sampling error) can be determined using:

**

Therefore, standard error = **534.52**

***t (test-statistic) = 1.271***

t (critical) can be found out using the Excel formula

*=T.INV(1-α, df)*

Substituting α=0.05 & df = 55

***t (critical) = 1.673***

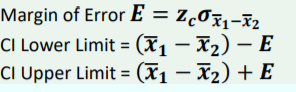
Further, p-value is obtained using the Excel function:

*=1 – T.DIST(t, df, 1)*

Substituting t (test-statistic) = 1.271 and df =55:

***p-value = 0.1046***

In other instances, Margin of error can be calculated by the product of z (critical) and Standard error. The procedure to calculate z (critical) remains similar to the excel formula for t-critical, except that the degree of freedom is eliminated.



With the help of margin of error, we can calculate confidence interval’s upper and lower limit by adding and subtracting this margin of error to the difference of the two sample means, respectively.

PART II:

The procedure for F-distribution is a bit different as we are testing the difference/ratio of two population standard deviation in the sales data of the fourth Excel sheet.

Ho: σ1 = σ2 or σ1/σ2 = 1

HA: σ1≠σ2 or σ1/σ2 ≠ 1

This is a two-tailed test to determine whether there is a difference in the standard deviation of sales of the company and the competitor.

We begin by calculating size, degrees of freedom and variance for each sample separately.

n1 = 15

n2 = 21

s12= 2830002.183

s22= 6314466.103

df1 = 14

df2 = 20

Since, the value of s22 > s12

*Test-statistic (f) =*

***f (test-statistic) = 2.23***

For 0.05 level of significance ***(α=0.05),***

*f (critical) = F.INV(1 – α, df1, df2)*

Therefore, ***f (critical) = 2.60***

*p-value = 1 – F.DIST(ftest-statistic, df1, df2,1)*

***p-value = 0.0988***

**CONCLUSION:**

The conclusion for any hypothesis can be decided by comparing the values obtained (highlighted in bold and italic). The 4 parameters that influence the decision are p-value, alpha value, test statistic and critical value. Depending on whether it is a right-tailed/left-tailed or two-tailed test, we can draw conclusion to support or reject the null hypothesis and validate out claim (alternate hypothesis)

PART I:

In case of the two-sample t-test (Right-tailed):

|  |  |
| --- | --- |
| α | 0.05 |
| t-critical | 1.6730 |
| t-test statistic | 1.271 |
| p-value | 0.1046 |

Since, the p-value > α & (We fail to reject the null hypothesis)

the t (test-statistic) < t (critical) (We fail to reject the null hypothesis)

This means that we are 95% confident that the company’s average sales is not greater than the competitor’s average sales.

PART II:

In case of the two-sample f-test (Two-tailed):

|  |  |
| --- | --- |
| test statistic | 2.23 |
| P-value | 0.0988 |
| Alpha-value | 0.05 |
| Critical F value | 2.60 |

Since, the p-value > α & (We fail to reject the null hypothesis)

the f (test-statistic) < f (critical) (We fail to reject the null hypothesis)

This means that we are 95% confident that there is no difference in the standard deviation of sales of the company and the competitor.

**REFERENCE:**

Albright, E.A. (2018). Two Independent Samples Unequal Variance. Retrieved December 7, 2018 from <https://sites.nicholas.duke.edu/statsreview/means/welch/>

Chieh, C.J. (n.d.). Making sense of the two-sample t-test. Retrieved December 8, 2018 from <https://www.isixsigma.com/tools-templates/hypothesis-testing/making-sense-two-sample-t-test/>