Testing of Hypothesis - II (Z,T and Chi Square Tests)

AJITH M.S

Z test

- A Z-test is a statistical test used to determine whether two population means are different when the variances are known and the sample size is large.
- The Z-test calculates a Z-score which represents the number of standard deviations a value is from the mean.
- It then compares this Z-score to the standard normal distribution to determine the significance of the difference between the means.
- This test is used in hypothesis testing to determine if an observed difference between two means is significant or if it could have occurred by chance.

The equation for a Z-test is as follows:

$Z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$

- Where:
- $\bar{x}$ is the sample mean
- μ is the population mean
- σ is the population standard deviation
- n is the sample size

The Z-score calculated using this equation represents the number of standard deviations that the sample mean deviates from the population mean.

Procedure

1. Ho : Mu = Mu o

H1 : Mu != Mu o (Two tailed)

H1 : Mu < Mu o (One tailed)

H1 : Mu >Mu o (one tailed)

2. Decide the test

If the sample size is large use Z test

IF you know the sigma value and the sample size is large use Z test

If sigma is not known and sample is small use T test

Step 3

Find Alpha(Level of Significance) And DOF

DOF for Z test = infinity

DOF for T test = n-1

- Where n = Sample size

Step 4

Find the value of Test static

For Z test, Z is the test static

For T test , T is the test static

Formula to find Test static

Z= (Sample mean – Population Mean)/SE

SE- Standard error

How to find Standard Error(SE)

If you know the sigma(Standard deviation) and the sample size is large or small

Use

SE= Sigma/ Sqrt(n)

You don’t know sigma and the sample is large

SE = S/Sqrt(n)

where S is the standard deviation of the sample

You know the Standard Deviation of the Sample(s), and the sample is small

SE = S/Sqrt(n-1)

STEP 5

Find the Table Value. Table value is denoted by TV.

Table value depends on level of Significance (alpha) and degree of freedom(D.O.F)

STEP 6

CV<TV (Numerically less Do not consider the sign)➔ Can accept the null hypothesis

CV>TV ➔ can reject the null values

A real-life example of a Z-test could be in the field of medical research, where the efficacy of a new drug is being tested against a placebo. The researchers take a random sample of patients and divide them into two groups: one group receives the new drug and the other receives the placebo. After a certain period of time, the researchers measure the reduction in symptoms (such as pain relief) for each patient and calculate the sample mean for each group.

To determine if the new drug is more effective than the placebo, they would perform a Z-test using the following steps:

1. State the null and alternative hypothesis H0: μd = μp (the population mean reduction in symptoms for the drug is equal to the population mean reduction in symptoms for the placebo) Ha: μd > μp (the population mean reduction in symptoms for the drug is greater than the population mean reduction in symptoms for the placebo)
2. Calculate the sample mean and standard deviation for each group
3. Calculate the Z-score using the equation provided earlier
4. Compare the Z-score to the standard normal distribution to determine the significance of the difference between the means.

T Test

- A t-test is a statistical test used to determine whether two population means are different when the variances are unknown and the sample size is small.
- It is similar to a Z-test, but instead of using the standard normal distribution to compare the calculated test statistic, it uses the t-distribution.
- The t-distribution is a family of distributions that are used to make inferences about a population mean when the population standard deviation is unknown and estimated from the sample data.
- There are two types of t-tests: the one-sample t-test and the two-sample t-test. The one-sample t-test is used to compare the mean of a single sample to a known population mean, while the two-sample t-test is used to compare the means of two independent samples. The t-test is commonly used in hypothesis testing to determine if an observed difference between two means is significant or if it could have occurred by chance.

Procedure

1. Ho : Mu = Mu o

H1 : Mu != Mu o (Two tailed)

H1 : Mu < Mu o (One tailed)

H1 : Mu >Mu o (one tailed)

2. Decide the test

If the sample size is large use Z test

IF you know the sigma value and the sample size is large use Z test

If sigma is not known and sample is small use T test

Step 3

Find Alpah(Level of Significance) And DOF

DOF for Z test = infinity

DOF for T test = n-1

- Where n = Sample size

Step 4

Find the value of Test static

For Z test, Z is the test static

For T test , T is the test static

Formula to find Test static

Z= (Sample mean – Population Mean)/SE

SE- Standard error

How to find Standard Error(SE)

If you know the sigma(Standard deviation) and the sample size is large or small

Use

SE= Sigma/ Sqrt(n)

You don’t know sigma and the sample is large

SE = S/Sqrt(n)

where S is the standard deviation of the sample

You know the Standard Deviation of the Sample(s), and the sample is small

SE = S/Sqrt(n-1)

STEP 5

Find the Table Value. Table value is denoted by TV.

Table value depends on level of Significance (alpha) and degree of freedom(D.O.F)

STEP 6

CV<TV (Numerically less Do not consider the sign)➔ Can accept the null hypothesis

CV>TV ➔ can reject the null values

Example

A real-life example of a t-test could be in the field of education, where the effectiveness of a new teaching method is being evaluated. The researchers take a random sample of students and divide them into two groups: one group receives the new teaching method and the other receives the traditional teaching method. After a certain period of time, the researchers measure the test scores for each student and calculate the sample mean for each group.

To determine if the new teaching method is more effective, they would perform a two-sample t-test using the following steps:

1. State the null and alternative hypothesis H0: μ1 = μ2 (the population mean test score for the new teaching method is equal to the population mean test score for the traditional teaching method) Ha: μ1 > μ2 (the population mean test score for the new teaching method is greater than the population mean test score for the traditional teaching method)
2. Calculate the sample mean and standard deviation for each group
3. Calculate the t-statistic using the equation provided earlier
4. Compare the t-statistic to the t-distribution with (n1-1) and (n2-1) degrees of freedom to determine the significance of the difference between the means.

If the calculated t-statistic is higher than the critical value from the t-distribution, the researchers can reject the null hypothesis and conclude that the new teaching method is more effective than the traditional teaching method.

CHI Square Test

- A statistical test in which a test statistics follows , chi square distribution
- The Chi-Square test is a statistical method used to determine if there is a significant difference between the observed frequency distribution and the expected frequency distribution in one or more categories.
- It can be used for both goodness-of-fit tests and test of independence.
- The test statistic calculated is the Chi-Square ($\chi 2$) value, and the decision to reject the null hypothesis is based on the p-value obtained from comparing the calculated $\chi 2$ value to the critical value from the Chi-Square distribution table with a given degree of freedom.

The equation for the Chi-Square test statistic is:

$\chi 2 = \Sigma ((O - E)^2 / E)$

Where:

- $\chi 2$: the calculated Chi-Square test statistic
- O: the observed frequency
- E: the expected frequency, calculated as (row total * column total) / grand total
- Σ: the sum of all the categories being compared

The critical value for the test is obtained from the Chi-Square distribution table with a given degree of freedom, which is calculated as (number of categories - 1). If the calculated test statistic is greater than the critical value, the null hypothesis is rejected, indicating a significant difference between the observed and expected frequency distributions.

A real-life example of using a Chi-Square test could be in market research to determine if there is a significant association between the type of coffee drink a customer prefers and their age group. The null hypothesis would be that there is no relationship between the two variables, and the alternate hypothesis would be that there is a relationship.

The researcher would collect data by surveying a sample of customers and categorizing their coffee drink preference (such as espresso, latte, cappuccino, etc.) and their age group (such as 18-25, 26-35, 36-45, etc.). The observed frequency would be the number of customers in each category, and the expected frequency would be calculated as the product of the row and column totals divided by the grand total.

The researcher would then perform a Chi-Square test to determine if the difference between the observed and expected frequency is significant. If the calculated test statistic is greater than the critical value, the null hypothesis would be rejected, and the researcher would conclude that there is a significant association between the type of coffee drink a customer prefers and their age group.

Characteristic of Chi Square Test

1. it’s a non parametric test

2.Its a distribution free test

3.Easy to calculate chi square value

4.It describes the magnitude difference of Observed frequency and Expected frequency

Uses

1. To test the goodness of fit between observed and expected frequency.

2.To test the independence of attributes

3.To test the homogeneity of the data.

4.To test population varience.

Applications

The chi-square test has many real-world applications in various fields, including:

1. Medical Research: The chi-square test is used in medical research to determine if there is a relationship between a disease and a risk factor, such as age, gender, or lifestyle habits.
2. Marketing: Market researchers use the chi-square test to determine if there is a significant relationship between consumer behavior and product preferences.
3. Education: The chi-square test is used in education to determine if there is a significant difference in the performance of students based on factors such as gender, ethnicity, or socioeconomic status.
4. Political Science: Political scientists use the chi-square test to determine if there is a significant relationship between political attitudes and demographic characteristics, such as age, education, and income.
5. Quality Control: The chi-square test is used in quality control to determine if there is a significant difference between the observed and expected number of defects in a manufacturing process.

Procedure

1.Ho: there is goodness of fit between OF&EF
H1: there is no goodness of fit
2. We apply chi square test
3.Find Level of significance and degree of freedom
In chi square DOF= n-r-1
n = sample size, r= no of parameters computed from the data(eg- mean, median)

4. $\chi 2 = \Sigma ((O - E)^2 / E)$

Where:

- $\chi 2$: the calculated Chi-Square test statistic
- O: the observed frequency
- E: the expected frequency, calculated as (row total * column total) / grand total
- Σ: the sum of all the categories being compared

5. We will find the table value

If calculated value < Table value ➔ Accept Ho

If calculated value> Table Value ➔Reject Ho

Inclusys Org Foundation
Empowering the Neurodivergent

Thanks