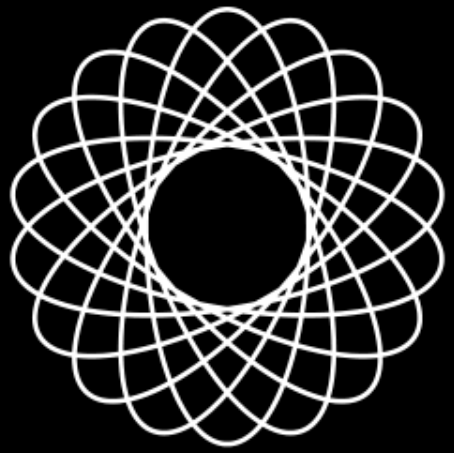
DATA SCIENCE



Agenda

Anova
- One Way
- Two Way
- Post Hoc Tests

Chi Square
- Association Tests
- Goodness-of-Fit Tests

Chi Square Parametric
- **Tests of Variance**

Chi-Square Test of Variance

- The Chi Square tests reviewed so far have been **non-parametric tests**

Chi-Square Test of Variance

- The Chi Square tests reviewed so far have been **non-parametric tests**

- To apply these tests, we do not need the underlying population to follow any specific distribution

Chi-Square Test of Variance

- The Chi Square tests reviewed so far have been **non-parametric tests**
- To apply these tests, we do not need the underlying population to follow any specific distribution
- There are many kinds of non-parametric tests, an equivalent one for every parametric test

Chi-Square Test of Variance

- The Chi Square tests reviewed so far have been **non-parametric tests**
- To apply these tests, we do not need the underlying population to follow any specific distribution
- There are many kinds of non-parametric tests, an equivalent one for every parametric test
- Which types of test are preferable? Non-parametric, or parametric?

Non-Parametric Tests

- There are tests in statistics that do not require a specific distribution for your data – **non-parametric tests**

- Non-parametric tests “better” than parametric tests because you are not bound to have a data distribution of a particular type

Non-Parametric Tests

- There are tests in statistics that do not require a specific distribution for your data – **non-parametric tests**
- Non-parametric tests “better” than parametric tests because you are not bound to have a data distribution of a particular type
- Why use parametric tests then?

Non-Parametric Tests

- There are tests in statistics that do not require a specific distribution for your data – **non-parametric tests**
- Non-parametric tests “better” than parametric tests because you are not bound to have a data distribution of a particular type
- Why use parametric tests then?
  - Non-parametric tests are less powerful than parametric tests in the sense that they use more information and are sometimes less flexible in terms of testing different kinds of hypothesis

Non-Parametric Tests

- There are tests in statistics that do not require a specific distribution for your data – **non-parametric tests**
- Non-parametric tests “better” than parametric tests because you are not bound to have a data distribution of a particular type
- Why use parametric tests then?
  - Non-parametric tests are less powerful than parametric tests in the sense that they use more information and are sometimes less flexible in terms of testing different kinds of hypothesis
  - Also, as sample size increases, it turns out that non-parametric test distributions approximate normal distributions

Chi-Square Parametric Test
(Exact Chi Square Test)

A Parametric Test

- This is a test of variance of sample tested against a population variance

Chi-Square Parametric Test
(Exact Chi Square Test)

A Parametric Test

- This is a test of variance of sample tested against a population variance

- The CLT posits that the distribution of sample means will follow a normal distribution

Chi-Square Parametric Test
(Exact Chi Square Test)

A Parametric Test

- This is a test of variance of sample tested against a population variance
- The CLT posits that the distribution of sample means will follow a normal distribution
- What about the variance of the samples?

Chi-Square Parametric Test
(Exact Chi Square Test)

A Parametric Test

- This is a test of variance of sample tested against a population variance
- The CLT posits that the distribution of sample means will follow a normal distribution
- What about the variance of the samples?
  - The variance of samples will follow a Chi Square distribution

Chi-Square Test of Variance

- The Chi Square distribution can also be used for testing variance and std deviation (so far all hypothesis testing we have reviewed was concerned with testing means)

Chi-Square Test of Variance

- The Chi Square distribution can also be used for testing variance and std deviation (so far all hypothesis testing we have reviewed was concerned with testing means)

- We cannot use the same tests to check for variance or std deviation (why not?)

- Test Statistic is computed as $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

Chi-Square Test of Variance

- The Chi Square distribution can also be used for testing variance and std deviation (so far all hypothesis testing we have reviewed was concerned with testing means)

- We cannot use the same tests to check for variance or std deviation (why not?)

- Test Statistic is computed as $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, where

  N = sample size

Chi-Square Test of Variance

- The Chi Square distribution can also be used for testing variance and std deviation (so far all hypothesis testing we have reviewed was concerned with testing means)

- We cannot use the same tests to check for variance or std deviation (why not?)

- Test Statistic is computed as $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, where

  N = sample size

  S = sample std. deviation

Chi-Square Test of Variance

- The Chi Square distribution can also be used for testing variance and std deviation (so far all hypothesis testing we have reviewed was concerned with testing means)

- We cannot use the same tests to check for variance or std deviation (why not?)

- Test Statistic is computed as $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, where

  N = sample size

  S = sample std. deviation

  σ = population std deviation

Chi-Square Test of Variance

Example:

A call center is experimenting with different approaches to improve customer experience, with the aim of consistent call resolution time.

Chi-Square Test of Variance

Example:

A call center is experimenting with different approaches to improve customer experience, with the aim of consistent call resolution time.

- Currently average resolution time is **6.5** minutes, with a variance of **4.5** minutes.

Chi-Square Test of Variance

Example:

A call center is experimenting with different approaches to improve customer experience, with the aim of consistent call resolution time.

- Currently average resolution time is **6.5** minutes, with a variance of **4.5** minutes.

- A new approach has been tested resulting in an average resolution time of **6** minutes, and a variance of **3** minutes across 30 calls.

Chi-Square Test of Variance

Example:

A call center is experimenting with different approaches to improve customer experience, with the aim of consistent call resolution time.

- Currently average resolution time is **6.5** minutes, with a variance of **4.5** minutes.

- A new approach has been tested resulting in an average resolution time of **6** minutes, and a variance of **3** minutes across 30 calls.

  – Is the new approach sufficiently different from the standard to justify investment in it?

Chi-Square Test of Variance

Example:

A call center is experimenting with different approaches to improve customer experience, with the aim of consistent call resolution time.

- If our aim is consistency, we check if there is significant reduction in variance of resolution time:

H0: Variance = 4.5 minutes

H1: Variance < 4.5 minutes

Chi-Square Test of Variance

Example:

A call center is experimenting with different approaches to improve customer experience, with the aim of consistent call resolution time.

- If our aim is consistency, we check if there is significant reduction in variance of resolution time:

H0: Variance = 4.5 minutes

H1: Variance < 4.5 minutes

Chi Square Statistic = 29 * 3^2 / (4.5^2) = 12.88

Chi-Square Test of Variance

Example:

A call center is experimenting with different approaches to improve customer experience, with the aim of consistent call resolution time.

- If our aim is consistency, we check if there is significant reduction in variance of resolution time:

H0: Variance = 4.5 minutes

H1: Variance < 4.5 minutes

Chi Square Statistic = 29 * 3^2 / (4.5^2) = 12.88

DF = 29

Chi-Square Tests - SAS

We could use a table to compare calculated Test Stat against a critical Value

OR

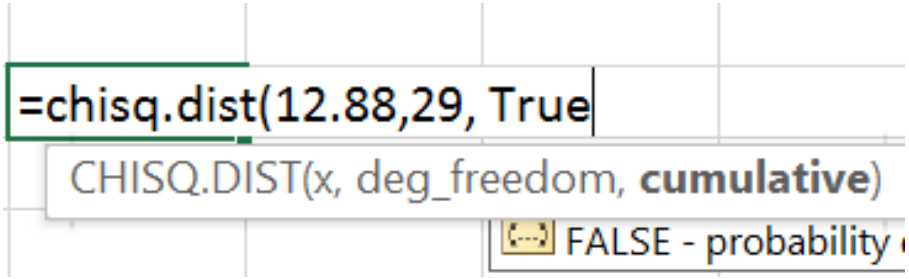
Directly calculate p-values in Excel

Chi-Square Tests - SAS

We could use a table to compare calculated Test Stat against a critical Value

OR

Directly calculate p-values in Excel

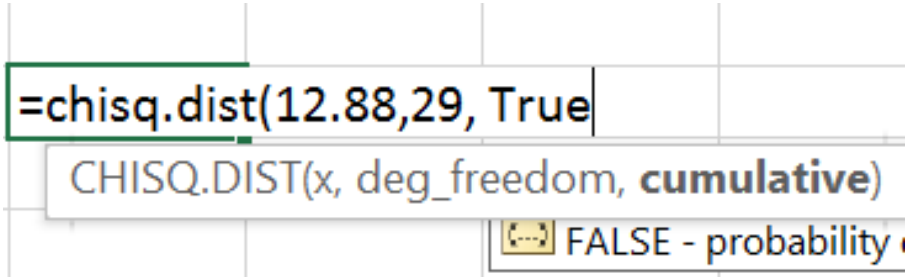


Chi-Square Tests - SAS

We could use a table to compare calculated Test Stat against a critical Value

OR

Directly calculate p-values in Excel



p-value = 0.002, therefore reject the null and conclude variance of calls has reduced

THANK YOU