some thing wrong

Md. Chiral
https://jhossain.me/
September 10, 2021

Lead Organizer, Introduction to Scientific Computing for Biologists
Founder, Health Data Research Organization

Variable Distribution Type Tests (Gaussian)

Variable Distribution Type Tests

- Shapiro-Wilk Test
- D'Agostino's k^2 Test
- Anderson-Darling Test

Shapiro-Wilk Test

Tests whether a data sample has a Gaussian distribution/normal distribution.

Assumptions

Observations in each sample are independent and identically distributed (iid).

Interpretation

- H0: The sample has a Gaussian/normal distribution.
- Ha: The sample does not have a Gaussian/normal distribution.

D'Agostino's k^2 Test

Tests whether a data sample has a Gaussian distribution/normal distribution.

Assumptions

Observations in each sample are independent and identically distributed (iid).

Interpretation

- H0: The sample has a Gaussian/normal distribution.
- Ha: The sample does not have a Gaussian/normal distribution.

Remember

- If Data Is Gaussian:
  - Use Parametric Statistical Methods
- Else:
  - Use Nonparametric Statistical Methods

Variable Relationship Tests (correlation)

Variable Relationship Tests

- Pearson's Correlation
- Coefficient
- Spearman's Rank Correlation
- Kendall's Rank Correlation
- Chi-Squared Test

Correlation Test

Correlation Measures whether greater values of one variable correspond to greater values in the other. Scaled to always lie between $\pm$ 1 .

- Correlation is Positive when the values increase together.
- Correlation is Negative when one value decreases as the other increases.
- A correlation is assumed to be linear.
- 1 is a perfect positive correlation
- 0 is no correlation (the values don't seem linked at all)
- –1 is a perfect negative correlation

Correlation Methods

- **Pearson's Correlation Test:**
  assumes the data is normally distributed and measures linear correlation.
- **Spearman's Correlation Test:**
  does not assume normality and measures non-linear correlation.
- **Kendall's Correlation Test:**
  similarly does not assume normality and measures non-linear correlation, but it less commonly used.

Difference Between Pearson's and Spearman's

Pearson's Test	Spearman's Test
Paramentric Correlation	Non-parametric
Linear relationship	Non-linear relationship
Continuous variables	continuous or ordinal variables
Propotional change	Change not at constant rate

Pearson's Correlation Coefficient

Tests whether two samples have a linear relationship.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample are normally distributed.
- Observations in each sample have the same variance.

Interpretation

- H0: There is a relationship between two variables
- Ha: There is no relationship between two variables

Spearman's Rank Correlation Test

Tests whether two samples have a monotonic relationship.

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.

Interpretation

- **H0 hypothesis:**
  There is is relationship between variable 1 and variable 2
- **H1 hypothesis:**
  There is no relationship between variable 1 and variable 2

Kendall's Rank Correlation Test

Assumptions

- Observations in each sample are independent and identically distributed (iid).
- Observations in each sample can be ranked.

Interpretation

- **H0 hypothesis:**
  There is is relationship between variable 1 and variable 2
- **H1 hypothesis:**
  There is no relationship between variable 1 and variable 2

Chi-Squared Test (Con..)

- The Chi-square test of independence tests if there is a significant relationship between two categorical variables The test is comparing the observed observations to the expected observations.
- The data is usually displayed in a cross-tabulation format with each row representing a category for one variable and each column representing a category for another variable.
- Chi-square test of independence is an omnibus test. Meaning it tests the data as a whole. This means that one will not be able to tell which levels (categories) of the variables are responsible for the relationship if the Chi-square table is larger than 2×2
- If the test is larger than 2×2, it requires post hoc testing. If this doesn't make much sense right now, don't worry. Further explanation will be provided when we start working with the data.

Chi-Squared Test (Con..)

Assumptions

- It should be two categorical variables(e.g; Gender)
- Each variables should have at leats two groups(e.g; Gender = Female or Male)
- There should be independence of observations(between and within subjects)
- Large sample size
  - The expected frequencies should be at least 1 for each cell.
  - The expected frequencies for the majority(80%) of the cells should be at least 5.

If the sample size is small, we have to use **Fisher's Exact Test**

Fisher's Exact Test is similar to Chi-squared test, but it is used for small-sized samples.

Chi-Squared Test

Interpretation

- The H0 (Null Hypothesis): There is a relationship between variable one and variable two.
- The Ha (Alternative Hypothesis): There is no relationship between variable 1 and variable 2.

Contingency Table

Contingency table is a table with at least two rows and two columns(2x2) and its use to present categorical data in terms of frequency counts.

Thank You