Testing of Statistical Hypothesis

Adejumo Ridwan Suleiman

2022-06-30

What is a Statistical Hpothesis

- This is a formal statement about the nature of a population.
- Example;
  - Are all Africans rich?
  - Do all footballers make above $100k/week.
  - Do Ronaldo and Messi have the same goal ratio.

Hypothesis Testing

- This is the use of statistical tests to make an inference about a population.
- The method of inference used depends on the nature of the data and reason for analysis.

Steps in conducting a Statistical Hypothesis Test

- Stating the Null and Alternative Hypothesis
- Stating the Level of Significance
- Test Statistics to be used
- Critical Region
- Decision Rule

Null Hypothesis

- Know as the Hypothesis of no difference.
- This Hypothesis states that the relationship the researcher is trying to investigate does not exist.
- OUr hope is always to prove the **Null hypothesis** wrong.
- The **Null Hypothesis** is represented mathematically as H_0

Alternative Hypothesis

- Opposite of the **Null Hypothesis**
- This is usually the hypothesis of the researcher.
- Represented as H_A

Case Study

- Let's say the mean price of diamonds is less than $10000
- In this case the **Null Hypothesis** can be stated as;
  "The mean price of diamonds is $10000"
- The **Alternative hypothesis** can be stated as ;
  "The mean price of diamonds is less than $1000"
- The parameter of interest in this example is the **Mean**.

Type 1 and Type 2 error

- Situation when you reject the **Null Hypothesis** when it is actually true is know as **Type 1** error.
- Situation when you fail to reject the **Alternative Hypothesis** when it is actually false is know as **Type 2** error.

Level of Significance (α)

- Situation where an experiment occurred beyond chance, then it is said to **Statistically Significant**.
- If we are to perform an experiment a number of times and the result of that experiment does not occurs by chance the experiment is declared **Statistically Significant**
- The level of significance denoted by α is the probability of rejecting the null hypothesis when it is actually true.
- The level of significance is mostly used as **0.05**.
- We can say that this is the probability of committing a **Type 1** error.

p-Value

- This is the probability of having the result of an experiment by chance when reproduced many times.
- For example a p-value of 0.32 means that when the experiment is repeated a number of times, there is a 32% chance that the result of your experiment actually occurred by chance.
- The lower the p-value the higher the statistical significance of the test.

Correlation Test

- This statistics test if two continuous variables are related to each other.
- The value ranges from -1 to +1 and the higher the value the stronger the relationship.
- Let's say from the diamonds data example we want to see if there is a relationship between diamonds carat and price.

```
##
##  Pearson's product-moment correlation
##
## data:  diamonds$price and diamonds$carat
## t = 551.41, df = 53938, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to
## 95 percent confidence interval:
##  0.9203098 0.9228530
## sample estimates:
##       cor
## 0.9215913
```

t-test