Test of Hypothesis

Mohan M J

Predictive Modeling

- Set of methods to arrive at quantitative solutions to problems of business interest
- Predictive modeling is a process that uses data mining and probability to forecast outcomes
- Part of Data Science or Statistical Learning
- High importance in recent past because data availability rising exponentially
- Examples
  - Preventive Maintenance: Automotive manufacturer want to predict the occurrence of fault or failure (or classify the condition of vehicles) through the sensor data captured
  - Insurance company want to classify drivers as very risky, risky, safe, very safe etc. on the basis of captured driving habits so that insurance premium can be intelligently fixed

Hypothesis Testing

- Hypothesis are statements about a given problem
- Hypothesis testing consists of determining the plausibility of the statements on the basis of data
- Examples
  1. Increasing number of years of education increases earning potential
  2. Design A produces a lower defect rate compared to design B
  3. A particular design of a web page leads to more conversion compared to another

Introduction

- In many situations it is required to accept or reject a statement or claim about some parameter
- Example
  - Average cycle time is less than 24hours
  - Percentage rejection is only 1%
- The statement is called hypothesis
- Procedure for decision making about the hypothesis is called hypothesis testing
- Advantages
  - Handles uncertainty in decision making
  - Minimizes subjectivity in decision making
  - Helps to validate assumptions or verity conclusions

Commonly used hypothesis tests

- Checking mean equal to a specified value ($mu = mu_0$)
- Two means are equal or not ($mu_1 = mu_2$)
- Two variances are equal or not ($sigma_1^2 = sigma_2^2$)

Test of hypothesis

- Null Hypothesis
  - A statement about the status quo
  - One of no difference or no effect
  - Denoted by H0
- Alternative Hypothesis
  - One in which some difference or effect is expected
  - Denoted by H1

General procedure

- Formulate the null hypothesis H0 and alternate hypothesis H1
- Gather evidence (data collection)
- Based on evidence take a decision to accept or reject H0

Methodology demo

- To test mean = specific value (mu = mu_0)
- Suppose we want to test whether mean of a process characteristic is 5 based on the following sample data from the process

4	4	5	5	6
5	4.5	6.5	6	5.5

- Calculate the mean of the sample, xbar = 5.5
- Or xbar – specific value = xbar -5 with 0
- If xbar-5 is close to 0
- Then conclude mean =5
- Else mean ≠ 5

Methodology demo

- To test mean = specific value (mu = mu_0)
- Suppose we want to test whether mean of a process characteristic is 5 based on the following sample data from the process

400	400	500	500	600
500	450	650	600	550

- Calculate the mean of the sample, xbar = 550
- Here xbar – 500 = 550 – 500 = 50
- Can we conclude mean ≠ 500?
- **Conclusion**: Difficult to say mean = specified value by looking at xbar – specific value alone

Methodology demo

- In Test of Hypothesis the test statistic is calculated by dividing (xbar – specific value) by a function of SD (Standard Deviation)
- To test mean = specific value

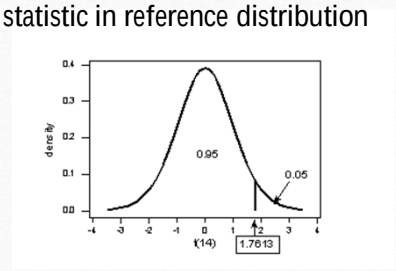
  Test Statistc , t_0 = (xbar – specific value)/(SD/√n)

  If the test statistic is close to 0, conclude that Mean = Specific value

  To check whether the test statistic is close to 0 find out P value from the sampling distribution of test statistic

P Value

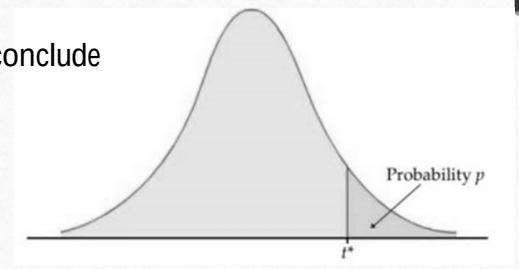
- The probability that such evidence or result will occur when H0 is true
- Based on the reference distribution of test statistic
- The tail area beyond the value of test statistic in reference distribution



P value

- If the test statistic t_0 is close to 0 then p will be high
- If the test statistic t_0 is not close to 0 when p will be small
- If p is small p<0.05 (with alpha =0.05), conclude that t $\neq$ 0 then

  Mean $\neq$ Specified value, H0 is rejected



Hypothesis testing: steps

1. Formulate the null hypothesis H0 and the alternate hypothesis H1
2. Select the appropriate statistical test and the corresponding test statistic
3. Choose level of significance – alpha (generally taken as 0.05)
4. Collect the data and calculate the value of test statistic
5. Determine the probability associated with test statistic under the null hypothesis using sampling distribution of the test statistic
6. Compare the probability associated with the test statistic with the level of significance specified

Exercise 1

- One sample t test

- **A company claims that on an average it takes only 40 hours to process any purchase order. Based on the data given, validate the claim? Data given in PO_Processing.csv**

One sample t test: Exercise 1

- **Hypothesis**
  - **Null Hypothesis H0: Mean processing time =40**
  - **Alternate Hypothesis H1: Mean processing time != 40**

Python Code

```
import pandas as mypanda
from scipy import stats as mystats
myData=mypanda.read_csv(“.\datasets\PO_Processing.csv")
print(myData) #print data
PT=myData.Processing_Time
print(PT) #print processing time
mystats.ttest_1samp(PT,40)
   Out[]: Ttest_1sampResult(statistic=3.7031497788267194,

pvalue=0.00035052328791173307)
#Conclusion is H0 rejected since p<0.05 ==> average processing
time is more than 40hours
```

Exercise 2:

- **A computer manufacturing company claims that on an average it will respond to any complaint logged by the customer from anywhere in the world in 24hours. Based on the data, validate the claim? Data is given in Complaint_Response_Time.csv**

Python Code

```
import pandas as mypandas
from scipy import stats as mystats
myData=mypandas.read_csv('.\datasets\Complaint_Response_Time.csv')
RT=myData.Response_Time
mystats.ttest_1samp(RT,24)
       Out[]: Ttest_1sampResult(statistic=6.9166494239747873,
                                pvalue=1.3350896696104558e-07)
#p value <0.05 ==> claim is not true - Null Hypothesis H0 rejected
```

Types of errors

- The decision procedure may lead to either of the two wrong conclusions
- Type I error
  - Rejecting the null hypothesis H0 when it is true
- Type II error
  - Failing to reject the null hypothesis H0 when it is false
- Alpha(Significance Level) = Probability of making type I error
- Beta = Probability of making the type II error
- Power = 1- Beta
  - Probability of correctly rejecting a false null hypothesis

Thank you