# Inferential statistics:2

# Case study 1:

Niki.ai is a startup. They develop different AI solutions and provide these solutions as a service. When they started the company, they deployed all there AI models in AWS and provided APIs to customers to communicate with those models. They observed that cloud service was expensive and the research that they were doing was very good, hence they decided to move all the models to in house servers.

After moving all the APIs and models to in house server, technical team made a observation on response time of individual models. Sample of those observations are given below.

| Services | AWS response time | In House response time |
|---|---|---|
| Sentiment Analysis | 32 | 96 |
| Named entity detection | 16 | 78 |
| Category classification | 12 | 64 |
| Snippeting tool | 34 | 84 |

How would you find is there any significant difference after moving models to in house servers?

# Plotting distribution

# Plotting distribution for IR service
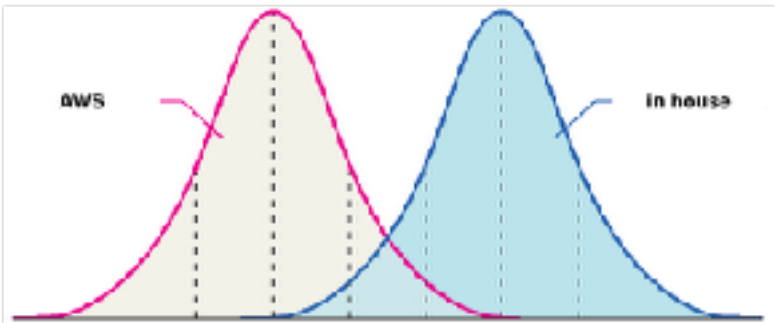


Mean of IR service response time in AWS = <mark>5 sec</mark>
Mean of IR service response time in house = <mark>20 sec</mark>

mean can not show change in distribution because distribution has other parameters too. so, it can not infer that the change has significant.
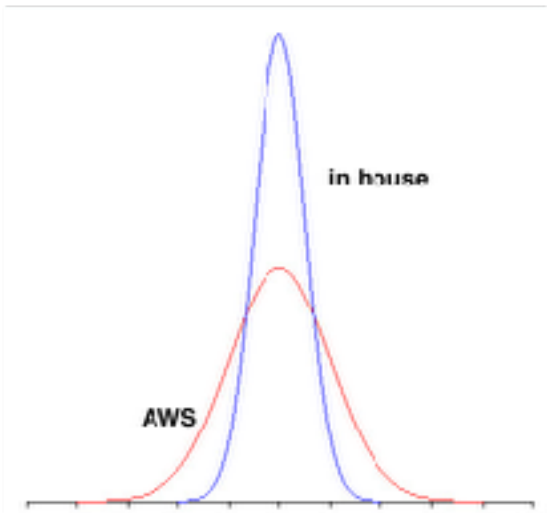


Mean of IR service response time in AWS = 5 sec
Mean of IR service response time in house = 5.3 sec

Is there any significant change in the new distribution?

| Person A | Yes |
|----------|-----|
| Person B | Yes |
| Person C | No  |
| Person D | No  |

Can not say.

# T Test

1. What is t-test?
2. What is the purpose of t-test?
3. What is critical value?
4. Are you conducting a one-tailed or two tailed test?
5. Is the data is paired or not?

## What is t-test?

A t-test is a statistic that checks if two means are reliably different from each other .

## What is the purpose of t-test?

Statistic way of finding is there any significant change in the new distribution compared to the old one.

# What is the critical value?

1. Let's say we wan't to represent whether there is a significant change or not how would you represent it?

$$P(\text{ significant change } | (OD, ND)) + p(\text{ no significant change } | (OD, ND)) = 1$$

OD = old distribution
ND = New distribution

2. Critical value probability = $p(\text{ no significant change } | (OD, ND))$

$$P_{critical} = 0.05$$

$$p(\text{ no significant change } | (OD, ND)) = 0.05$$

$$P(\text{ significant change } | (OD, ND)) + 0.05 = 1$$

$$P(\text{ significant change } | (OD, ND)) = 0.95$$

$$P_{critical} >= 0.05$$

There is only less than 5% chance that there is significant change in the New data distribution compared to old

# Are you conducting a one-tailed or two tailed test?

Two Tailed test: Is there any reliable significant change?
One Tailed test: Is new distribution performed well over old distribution?

https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/hypothesis-testing/one-tailed-test-or-two/

# Is the data is paired or not?

| | Height | | | | Boy Height | |
|---|---|---|---|---|---|---|
| | Boys | girls | | | Age 5 | Age 15 |
| | 72 | 66 | | | 54 | 69 |
| | 68 | 67 | | | 45 | 78 |
| | 69 | 68 | | | 56 | 64 |
| | 78 | 70 | | | 47 | 63 |
| | 64 | 61 | | | 45 | 61 |
| | 63 | 62 | | | 48 | 62 |
| | 77 | 71 | | | 39 | 71 |
| | 74 | 70 | | | 44 | 70 |
| | 73 | 68 | | | 45 | 68 |
| Mean | 70.88888889 | 67 | | Mean | 47 | 67.33333333 |
| std deviation | 5.30199124 | 3.5 | | std deviation | 5.196152423 | 5.431390246 |

# How we do it?

1. For given dataset define a hypothesis:

   $H_o$ = mean of the new distribution is same as the mean of the old distribution
   $H_a$ = mean of the new distribution is not same as the mean of the old distribution

2. Calculate the t-value:

$$= \frac{\left| \overline{X_1} - \overline{X_2} \right|}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

   $X_1$ = mean of old distribution
   $X_2$ = mean of new distribution
   $S_1$ = St.Deviation of old distribution
   $S_2$ = St.Deviation of new distribution
   n1 = Number of samples in old distribution
   n2 = Number of samples in new distribution

3. Define probability for critical value:

   Let's say I want my inference to be 95% sure about mean of new distribution is significantly vary with respect to old mean. That is,

   P( significant change | (OD,ND ) ) = 95%
   $P_{critical}$ = 0.05

4. Using probability for critical value find critical value using t table

5. If t-value > critical value then reject the null hypothesis or else accept the null hypothesis

# Case study 1:

Niki.ai is a startup. They develop different AI solutions and provide these solutions as a service. When they started the company they deployed all there AI models in AWS and provided APIs to customers to communicate with those models. They observed that cloud service was expensive and the research that they were doing was very good, hence they decided to move all the models to in house servers.

After moving all the APIs and models to in house server, technical team made a observation on response time of individual models. Sample of those observations are given below.

| Services | AWS response time | In House response time |
|---|---|---|
| Sentiment Analysis | 32 | 96 |
| Named entity detection | 16 | 78 |
| Category classification | 12 | 64 |
| Snippeting tool | 34 | 84 |

How would you find is there any significant difference after moving models to in house servers?

# Case study 1 solution

# Example:

The technical team of amazon has observed that the particular API is slow in server A compared to Server B. Given response time of that API in both the server as below,

| Server1 | Server2 |
|---------|---------|
| 11 | 12 |
| 12 | 14 |
| 15 | 12 |
| 11 | 12 |
| 12 | 14 |
| 15 | 15 |
| 12 | 12 |
| 15 | 14 |
| 11 | 12 |
| 12 | 14 |

How will you validate(or infer) the proposed argument is right or wrong?

$H_o$ : mean of the response time of API in server A = mean of the response time of API in server B
$H_a$ : mean of the response time of API in server A != mean of the response time of API in server B

# Case study 2:

The technical team of amazon has observed that the particular API is slow in server A compared to Server B & C. Given response time of that API in all the server as below,

| Server1 | Server2 | Server3 |
|---|---|---|
| 11 | 12 | 12 |
| 12 | 14 | 13 |
| 15 | 12 | 14 |
| 11 | 12 | 12 |
| 12 | 14 | 13 |
| 15 | 15 | 14 |
| 12 | 12 | 12 |
| 15 | 14 | 12 |
| 11 | 12 | 14 |
| 12 | 14 | 13 |
| 12 | 12 | 14 |
| 15 | 14 | 13 |
| 11 | 14 | 13 |

How will you validate(or infer) the proposed argument is right or wrong?

$H_o$ : mean of the RT of API in server A = mean of the RT of API in server B = mean of the RT of API in server C
$H_a$ : At least one difference among the means

# ANOVA

Analysis of variance

$$F = \frac{between\ groups}{within\ groups}$$

# Case study 2 solution

# Chi-squared Test

Let' say John is flipping a **fair** coin. He flipped the coin for 50 times and he observed that 28 times he has got head and 22 times he has got tail. Now he wants to understand whether the coin is actually behaving as a fair coin or not. How you can help John to infer this?

## Solution

https://www.youtube.com/watch?v=CS_BKChyPuc

$H_o$ : There is no significant difference between observed and expected frequencies
$H_a$ : There is significant difference between observed and expected frequencies

Let's plot a chart which represents what he observed during the experiment and what John was accepting from the Experiment

|  | Head | Tail |
|---|---|---|
| Expected | 25 | 25 |
| Observed | 28 | 22 |

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(28 - 25)^2}{25} + \frac{(22 - 25)^2}{25} = 0.72$$

Degree of freedom = Number of possible outcomes - 1
= 2 - 1
= 1

Prob(Critical value) = 0.05

Critical value = 3.84

Thank you