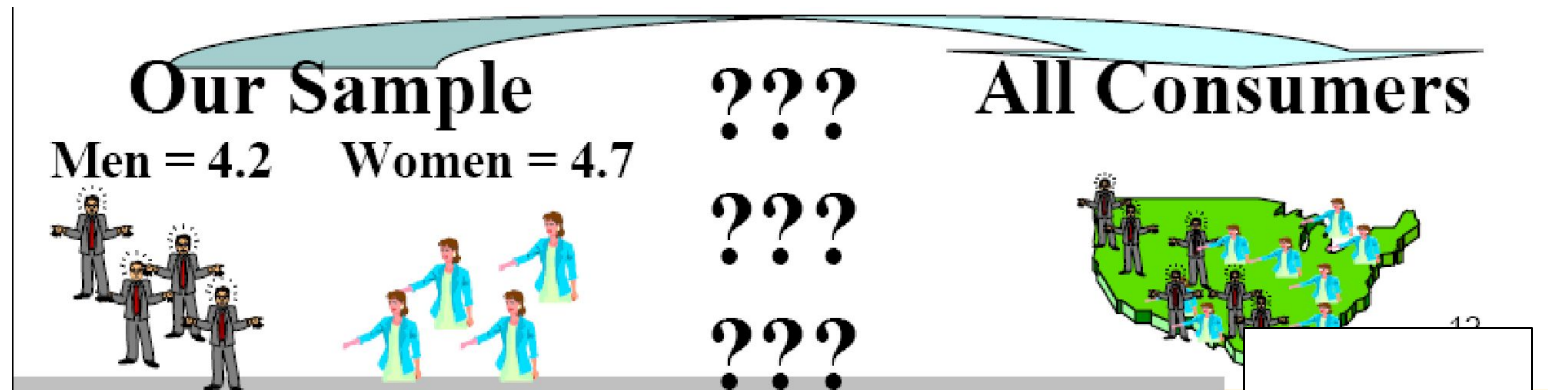

Hypothesis Testing

MD Arshad Ahmad
15 Years+ Experience in Data Science
Mentored 100+ people



Why Significance Testing ?

- Significance testing allows us to examine results from a sample of people, and make inferences about the total population.
 - Significance tests ask: Given the differences observed in our sample, what are the odds that there are no true differences in the population?



Why Hypothesis Testing ?



Why Hypothesis Testing ?

MD Arshad Ahmad
15 Years+ Experience in Data Science
Mentored 100+ people

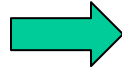


Video Explaining the scenarios where we need to use it

<https://youtu.be/hq40GN4HmP4>

Hypothesis Testing

Medicine Experiment

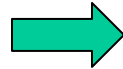


Average duration of cold = 8.5 days

$n = 250$

Recovery time = 7.3 days

Questions



Is the result significant?

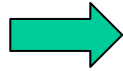
Could this sample just be the result of chance or did this drug have an impact?

Should the drug be tested further?

Does this mean this new drug should be approved for use?

How to test a hypothesis in four steps

Medicine Experiment



Town population = 35,000
Percentage men = 50%
Percentage women = 50%



Next Sample
14 Men
36 Women

Step 1a) Design Hypothesis

Step 1a) Develop Hypotheses

H_0 = Null hypothesis

Jury numbers happened by chance

Shows odds of woman being picked as 50%

H_0 is $p \leq 0.50$

H_a = Alternative hypothesis

Jury numbers not by chance

Shows odds of women being picked as higher than 50%

H_a is $p > 0.50$



Decoding Data
Science

Step 1b) State Significance Level

Set threshold for test

Significance level = 5%. If 36 or more women ending up on a jury have less than a 5% chance of occurring at random, then we will reject our null hypothesis.

Step 2) Identify Test Statistic

Binomial probability

$$p = 0.50$$

Step 3) Determine P-Value

Looking for the probability that the number of women chosen for the jury-panel would be 36 or more

Probability = 0.13%

Step 4) Compare P-Value to Significance Level

There was only a 0.13% chance that at random 36 or more women would be chosen for a panel of 50 potential jurors.

Alpha = 0.05 or 5%

$p\text{-value} < \text{significance level}$

It much more likely for a woman to be chosen versus a man.

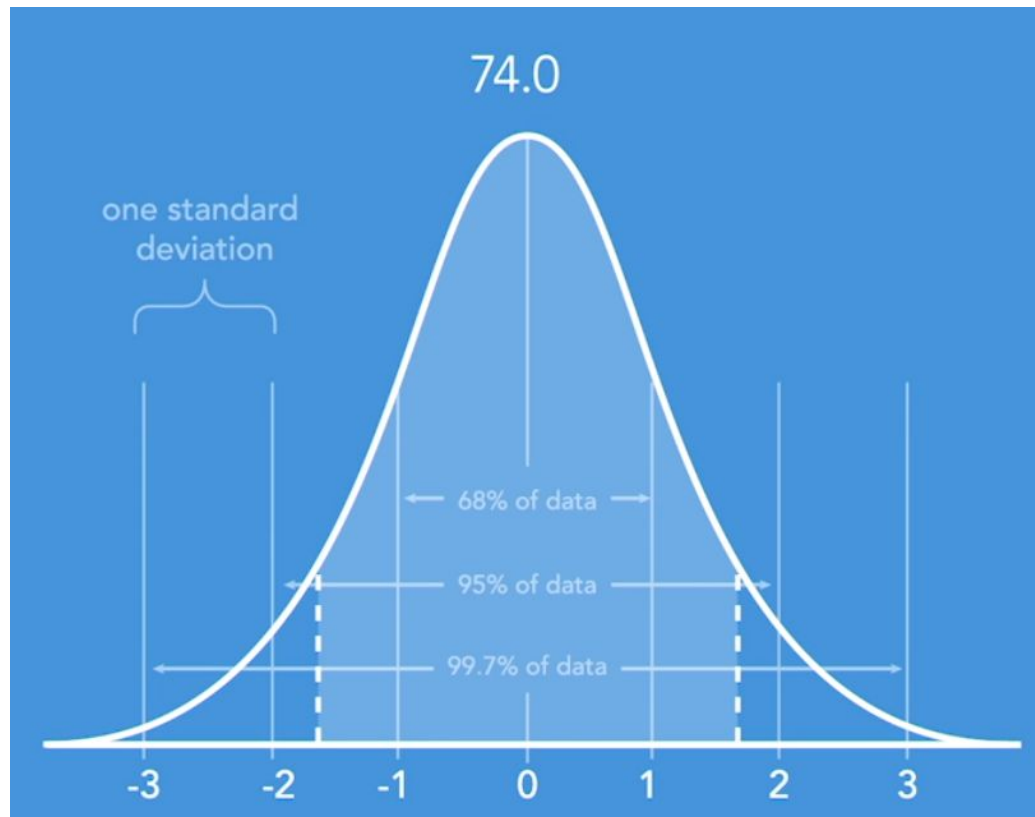


Scenarios for hypothesis testing

A recent national study found that the average American between the ages of 18 and 24 checks their phone 75 times per day. A mobile service provider questions these results

$$H_0: \mu = 74.0$$

$$H_a: \mu \neq 74.0$$



As you can see, we have two rejection areas here, one rejection area in the positive direction, greater than the mean. The other in the negative direction, less than the mean. This is considered a two tailed test because the null hypothesis is tested in both directions

Thank You!
To know more Get In Touch!

Kick start your Data Science Career



[Book Mentoring Session](#)

decodingdatascience.com

[Analytics Blog](#)