

Complete Data Science and Machine Learning Using Python

By
Jitesh Khurkhuriya

Hypothesis Testing

What is Hypothesis?

An idea or explanation of something that is based on a few known facts but that has not yet been proved to be true or correct.

-- Oxford Dictionary

What is Hypothesis?

Testing a known assumption that is generally accepted as the truth.

Testing a claim that is supposed to change the current facts.

Null Vs Alternate Hypothesis

H_0

- There is no change from the stated facts.
- Chocolate bars weigh 100 gms or more

$=$
 \leq
 \geq

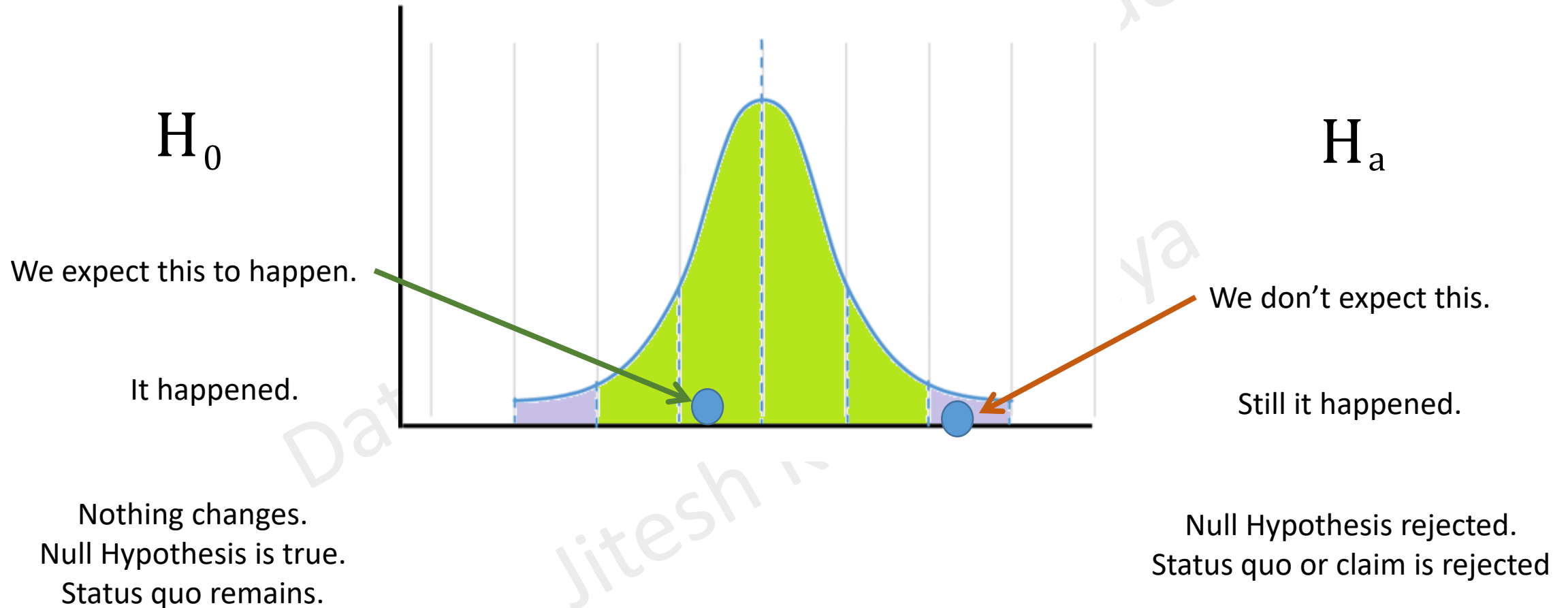
H_a

- The stated facts are incorrect
- Chocolate bars weigh less than 100 gms

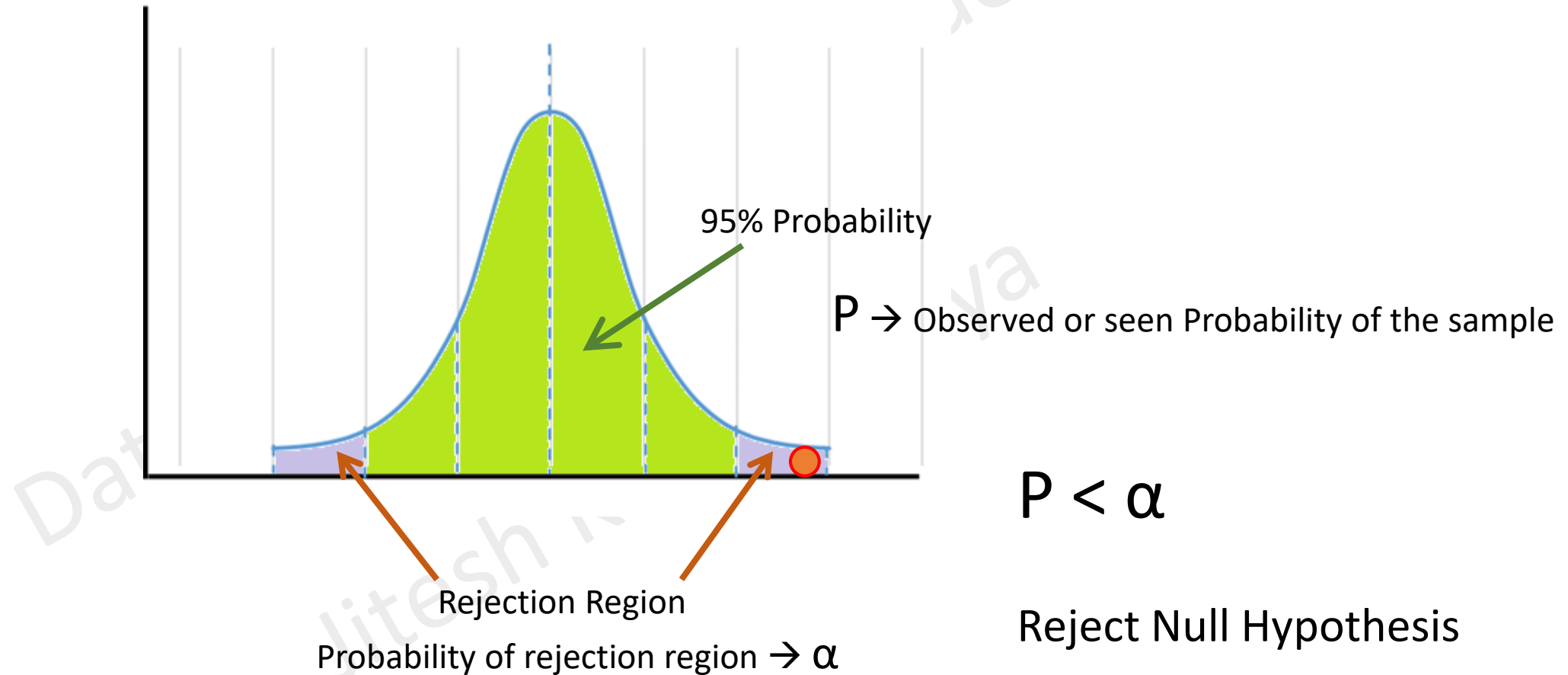
\neq
 $<$
 $>$

Statistical Significance

Statistical Significance



Important terms – Statistical Significance



Hypothesis Testing

Current average waiting period for the customers who call the customer service helpline is 100 seconds with a standard deviation of 20 seconds. Certain changes were recently done to the IVR menu options as well as the overall customer service processes. After a week, the management picked-up a sample of 100 calls and found that the average waiting period was 95 seconds. Have the process implementations resulted in the waiting period reduction?

H_0 : null hypothesis : There is no change in the waiting period.

H_a : alternate hypothesis : The waiting period has reduced.

Significance Level; $\alpha = 0.05$ or 5%

1. State Population parameters and Sample statistics

$\mu = 100$; $\sigma = 20$; $N = 100$; $\bar{X} = 95$;

2. Compute Sample Standard Deviation and Z-Value

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

3. Compute p using Z-Score for the Z-value

$p = 0.62\%$

Hypothesis Testing

Current average waiting period for the customers who call the customer service helpline is 100 seconds with a standard deviation of 20 seconds. Certain changes were recently done to the IVR menu options as well as the overall customer service processes. After a week, the management picked-up a sample of 100 calls and found that the average waiting period was 95 seconds. Have the process implementations resulted in the waiting period reduction?

H_0 : null hypothesis : There is no change in t

H_a : alternate hypothesis : The waiting perio

Significance Level; $\alpha = 0.05$ or 5%

$$p < \alpha$$

Reject Null

Population parameters and Sample statistics

$$\mu = 100; \sigma = 20; N = 100; \bar{x} = 95;$$

Compute Sample Standard Deviation and Z-Value

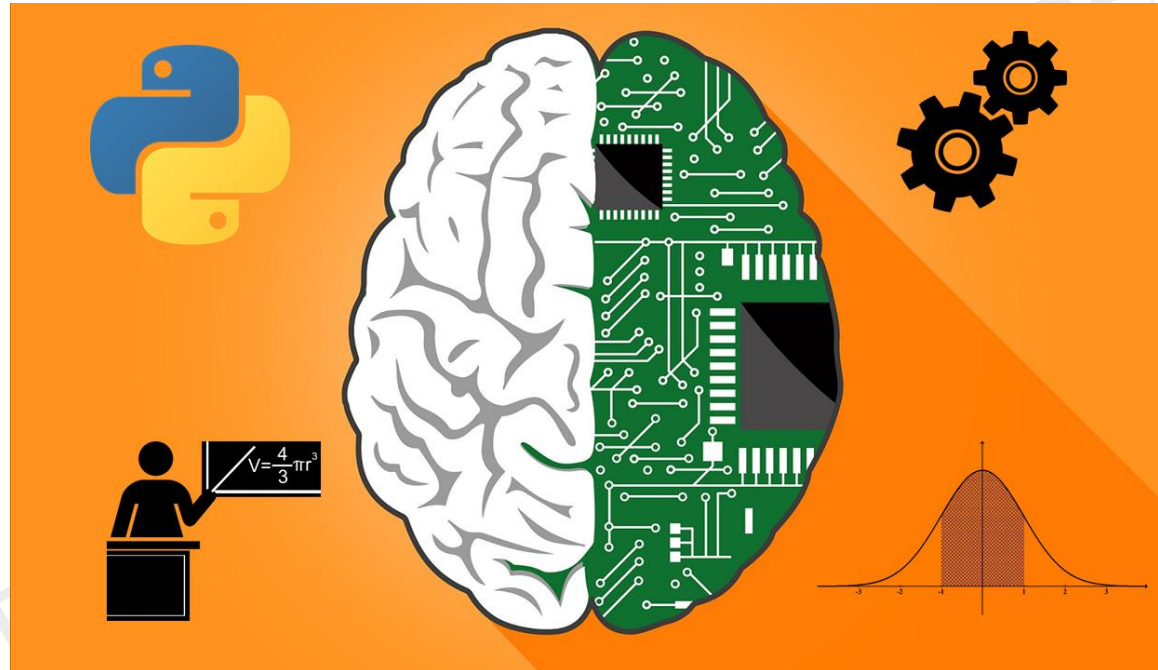
$$s = \frac{\sigma}{\sqrt{n}}$$

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

3. Compute p using Z-Score for the Z-value

$$p = 0.62\%$$

Complete Data Science and Machine Learning Using Python



Thank You!