Statistical Inference: Hypothesis Testing

Prof. Abhinava Tripathi

Introduction: Hypothesis Testing

Introduction

- Inferring insights from sample data is called inferential statistics
- Consider an example, where you have joined a mobile phone manufacturer as business analyst
- Company has come up with a new brand that is expected to charge in 30-mins for a full day operation : A day's power within 30-mins
- You want to check whether this claim is valid or not: some of the units do not conform to this number
- Marketing team wants you to conduct statistical test to be 95% sure of this claim

Introduction

- One approach to solve this problem is to collect a sample of 100 phones
- Chances are that some of these phones may take less than 30-mins and some may take more than 30-mins
- With this information you can come-up with a confidence interval that the charging time is in the range of 24-29 minutes with 95% confidence: I can say that charging time is less than 30 mins
- But what if the confidence interval is from 26-32 mins: I can not say that charging time is less than 30-mins

Introduction

- This confidence interval approach is perfectly valid
- However, as I increase my confidence interval from 95% to 99%, it will become increasing difficult to make any statistical claim that charging time is less than 30-mins
- This is so because chances are that 30-min value may fall in this interval
- A more robust and efficient method to test this claim is provided by hypothesis testing

Applications of Hypothesis Testing

Applications of Hypothesis Testing

- Hypothesis testing is an efficient way to test the statistical credibility of a claim
- You gather the evidence from the sample and check if the claim can be rejected or not
- One of the most common uses of hypothesis testing is campaign effectiveness
- A pizza delivery firm plans to test the effectiveness of their campaign in the test population and control population

Applications of Hypothesis Testing

- These kind of problems are referred to as AB testing: for example, you want to compare the response rate of two different webpages
- You divide the population in two groups (Group 1 and 2) that are exposed to different versions of the product: version A and version B
- Another example is the application of quality claim checks: a lightbulb manufacturer claims that his product will last more than 5000 hours
- Here, hypothesis testing can add a lot of value to analysis

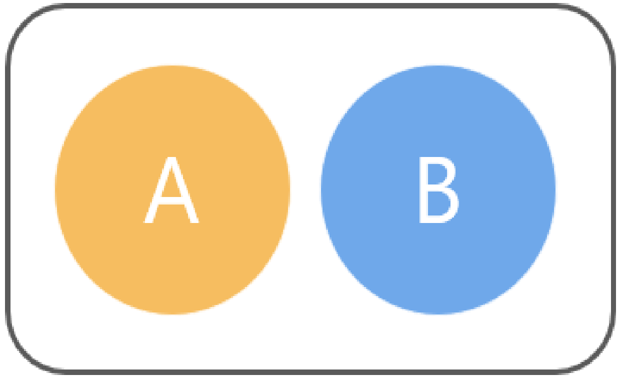
Hypothesis Testing: Part I

Hypothesis Testing

- We will continue with our charging time problem
- Let us start by assuming that the claim is true: It takes 30-minutes to charge the phone
- As a first step, we define the null and alternate hypothesis
- Null H0: The phone fully charges in exactly 30-mins
- Alternate H1: The phone does not charge in 30-mins
- If the null is rejected then either it takes more or less than 30-minutes to charge the phone

Hypothesis Testing

- This was a simple case of framing null and alternate hypothesis
- Null (A) and alternate (B) hypothesis are always mutually exclusive events



- This was a simple case of framing null and alternate hypothesis
- Null (A) and alternate (B) hypothesis are always mutually exclusive events
- Any other non-mutually exclusive configuration will be rejected

Hypothesis Testing

- Also the null and alternate must be collectively exhaustive, i.e., at least one of them must be true
- So the Null (H0) and alternate (Ha) are always mutually exclusive and exhaustive
- Null H0: Charging time is less than equal to (≤) 30-mins; Alternate H1: Charging time is more than (>) 30-mins
- As per the convention, Null hypothesis must contain equality sign

Hypothesis Testing

- In a converse manner, if someone claimed that the phone takes at least 30-mins then the following null and alternate would be framed
- Null H0: Charging time is more than (≥) 30-mins
- Alternate H1: Charging time is less than (<) 30-mins
- To summarize, (a) Null hypothesis captures the status quo, (b) alternate, which we are trying to prove, is the complement to the null, (c) as a convention, null contains equality (=, ≤, ≥,)

Hypothesis Testing

- In our earlier example, H0 is that 'Charging time is = 30 min'
- Alternative hypothesis, is that 'Charging time is NOT equal to 30 min'
- If the computations favor alternate hypothesis, then you reject the null or accept the alternate hypothesis
- That is, the claim that charging time is not equal to 30-mins is true
- And the required action is based on whether the charging time is more or less than 30-mins

Hypothesis Testing

- If the calculations favor the null hypothesis that charging time is equal to 30-mins that means you fail to reject the null (not that null is correct or accepted)
- Not that null hypothesis has been provided to be true
- Sample properties may different from population, and as more and more sample arrive, null may be disapproved
- For example, all swans are white, till the one single black swan was found to reject the null

Hypothesis Testing Part II: Critical Value Method

Hypothesis Testing: Critical Value Method (CVM)

- We will discuss the CVM method for testing the claim that charging time is equal to 30-mins
- We start with the assumption that null is true
- You sample 100 phones and find that their mean charging time is 30.37-mins and population standard deviation (σ) is equal to 2.477
- Please note that these 100 data points make-up for one sample
- We will employ CLT here to implement the hypothesis testing

Hypothesis Testing: Critical Value Method (CVM)

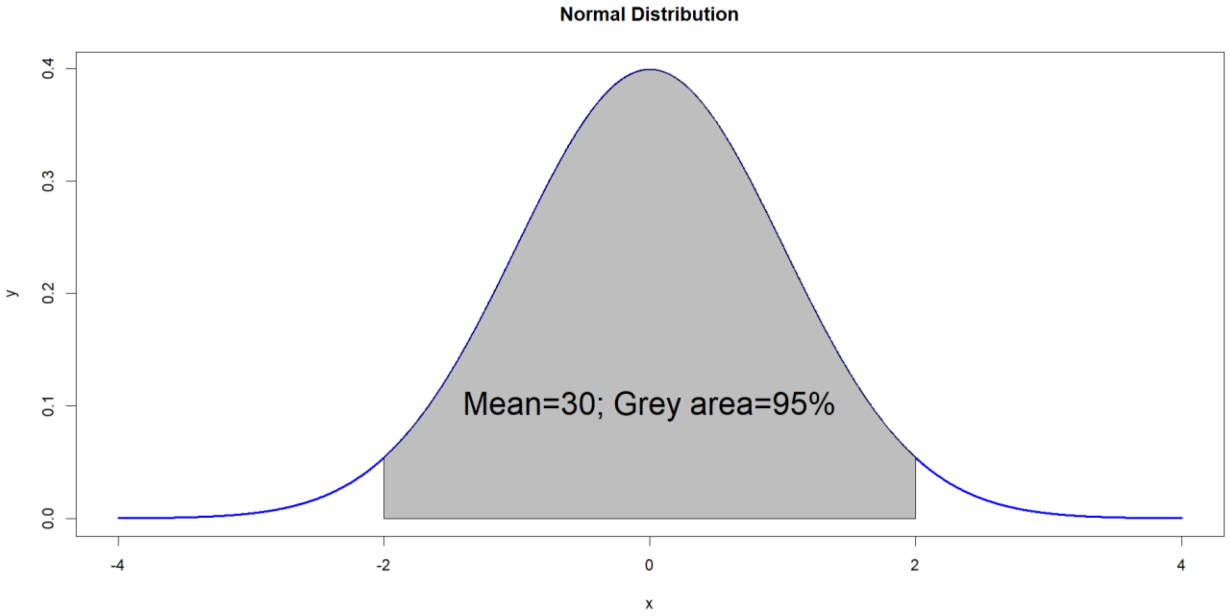
- Choose a sufficiently large sample (>30) so that sampling distribution is same as normal distribution
- CLT suggests that the mean of sampling distribution is the same as population mean: 30
- Also, if the population SD is 2.477, the SD of sampling distribution is 0.2477
- You are happy as long as the sample mean lies in the 95% confidence level interval of 30-mins

Hypothesis Testing: Critical Value Method (CVM)

- You are happy as long as the sample mean lies in the 95% confidence level interval of 30-mins

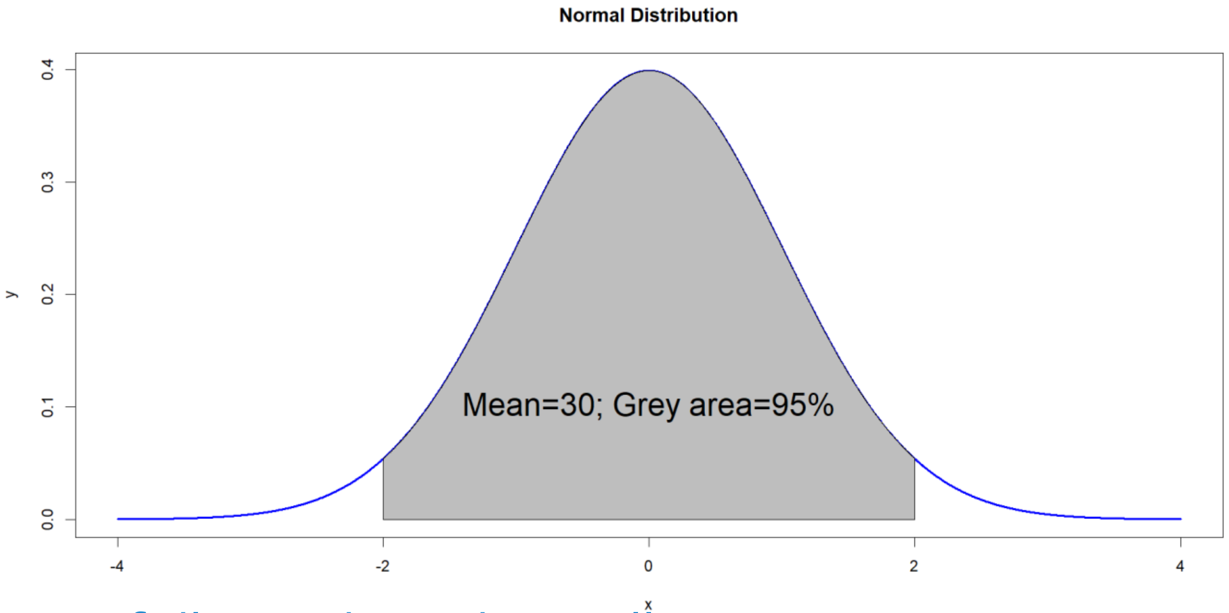
- Recall the properties of the normal curve

- Around 68% of the area lies in ± 1 SD, 95% in ± 2 SD, and 99.7% in ± 3 SD



Hypothesis Testing: Critical Value Method (CVM)

- The z-score corresponding to 30.37 is calculated as shown here: $\frac{30.37-30}{0.2477} = 1.4937$
  - Sample mean lies 1.4937 SD away from the center
  - This is less than the critical score of z=2



- Since this is less than the critical score, you fail to reject the null

Hypothesis Testing: Critical Value Method (CVM)

- For example, if sample mean was 30.62, then $z = \frac{30.62-30}{0.2477} = 2.5$
- This would fall outside the 95% region and you would be able to reject the null
- Let us recap the problem: (1) Frame the null and alternate hypothesis; (2) Decide the appropriate confidence interval; (3) Calculate the critical z value; (4) Compute the sample z-score; (5) Compare the sample z-score with the critical z value

Hypothesis Testing: Critical Value Method (CVM)

- For example, if sample mean was 30.62, then $z = \frac{30.62-30}{0.2477} = 2.5$
- This would fall outside the 95% region and you would be able to reject the null
- Let us recap the problem: (1) Frame the null and alternate hypothesis; (2) Decide the appropriate confidence interval; (3) Calculate the critical z value; (4) Compute the sample z-score; (5) Compare the sample z-score with the critical z value

Hypothesis Testing Part III: One Tailed Test (CVM)

One Tailed Test

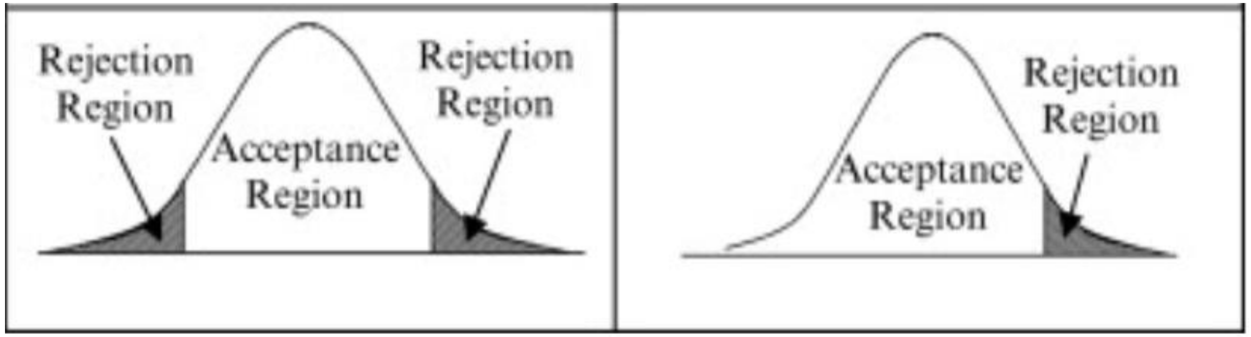
- It is sometimes sufficient to test only one side of the sample mean distribution; this is called a one-tailed test
- In the previous example, we performed the test on both sides of the normal distribution, that is, we would have rejected the null hypothesis if the sample was significantly different in either direction
- As a customer, you want to test whether a day's charge is less than 30-mins

One Tailed Test

- Step 1 is the same, i.e., frame the null and alternate hypothesis
- H0: Charging time is less than or equal (≤) 30-mins
- H1: Charging time is less than or equal (≥) 30-mins
- Step 2: Decide the confidence level (95%);
- Step 3: Find the corresponding critical z-value; here, we can reject the null if we can prove that the sample mean is significantly greater than 30-minutes
- Hence the rejection region is on the right side of the curve: right tailed test

One Tailed Test

- Let us compare the rejection regions in one-tailed vs two-tailed test



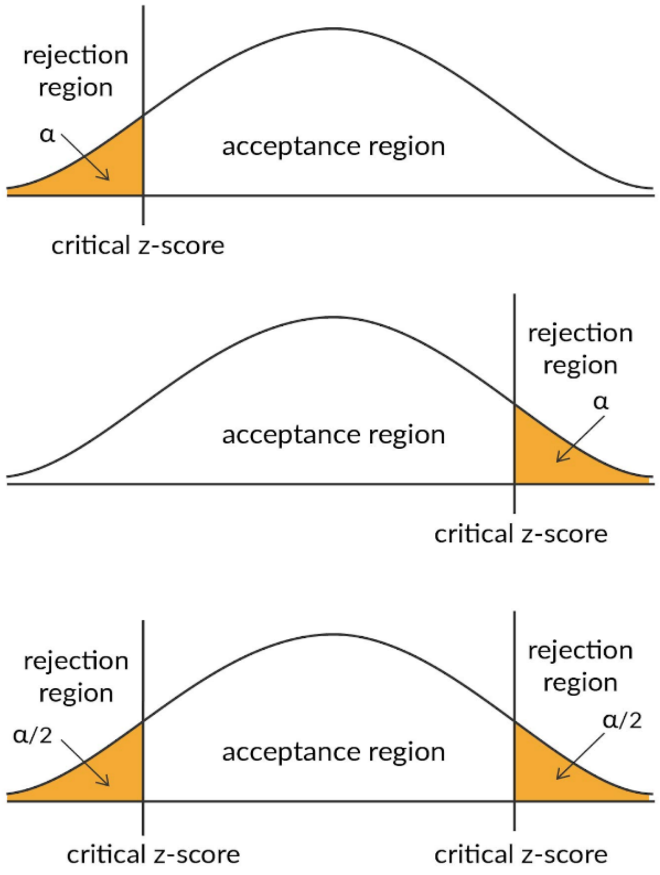
- In the two tailed test, the unshaded region is the 2.5% on the right and left
- In the single tailed test, the entire 5% is on the right side, thus the critical value will not be the same
- In the previous case, critical z value was 1.96. Now the complete area on the left of the curve is 95% and the corresponding z-value is 1.645
- The rejection region is on the right of this z-value

One Tailed Test

- Now that we have our z-value, let us compute the sample z-score
- For our sample of 100 phones, the sample mean was 30.79, and we found the population SD to be 2.477
- Sample z-score= $\frac{30.79-30}{\frac{2.477}{sqrt(100)}} = 3.19$; which is much larger than 1.645
- So we reject the null for a right tailed test

One Tailed Test

- So there are three types of tests: (1) two tailed tests; (2) Right tailed tests; (3) Left tailed tests
- If the alternate hypothesis has < sign, then it is a left tailed test
- If the alternate hypothesis has > sign, then it is a right tailed test
- If the alternate hypothesis includes a '≠' sign, then it is a two tailed test



One Tailed Test

- Let us consider one hypothetical case of left-tailed test
- Cadbury states that the average weight of a particular brand of its chocolate is 60g; as an analyst, you want to test if the weight is less than 60g or not at 2% significance
- Here H0: Weight is less than (≥) 60g; H1: Weight is less than (<) 60g
- Again the hypothesis can be solved in 5 standard steps
- We need to compare the critical z value (corresponding to 2% level) on the left tail with the corresponding z-statistic from the sample

Hypothesis Testing Part IV: P-value method

P-value method

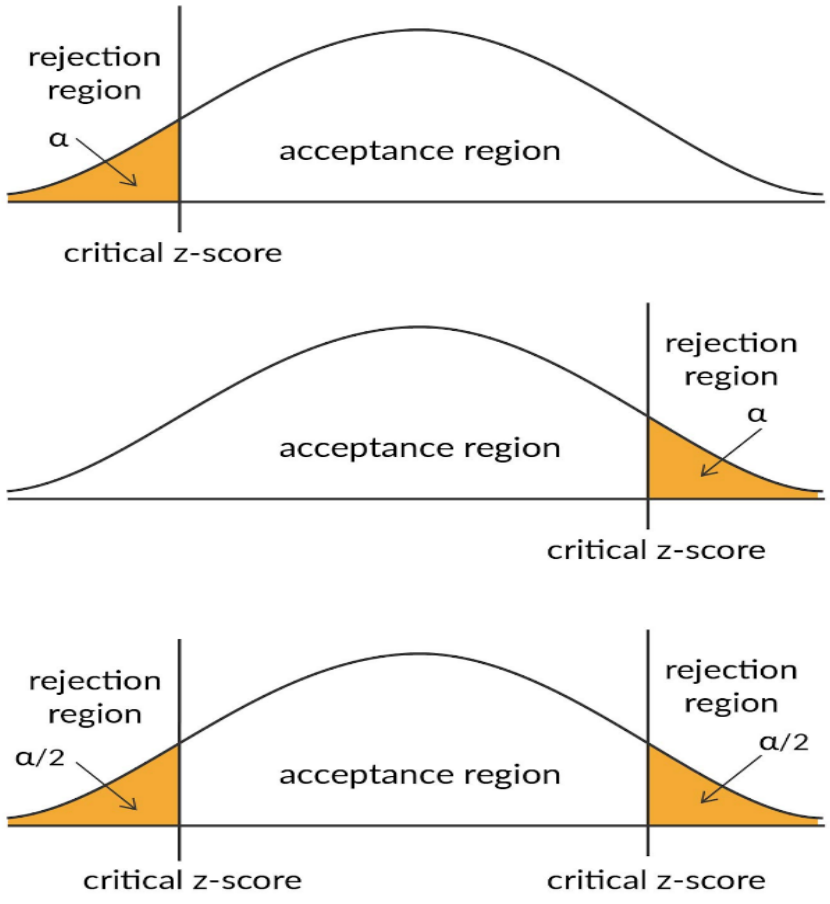
- Let us get introduced to the P-value approach to hypothesis testing
- In the same mobile phone example, first we formulate null and alternate
- H0: Mean charging time=30-mins; H1: Mean charging time ≠ 30-mins
- Decide the level of significance (alpha=5%) or level of confidence 1-0.05=0.95
- To compute the p-value, we need the corresponding z-score: we already calculated this value as 1.4937
- P-value or significance level or the area in the tail of the normal probability distribution

P-value method

- In our mobile phone example, p-value represents area from $-\infty$ to -1.4937 and 1.4937 to $+\infty$
- Since the curve is symmetric, we can calculate one value and multiply it with 2
- The value can be computed with R software. The value corresponding to one tail-works out to 0.068 and thus, the total p-value becomes 2*0.068=0.136
- If p-value is less alpha, then we reject the null

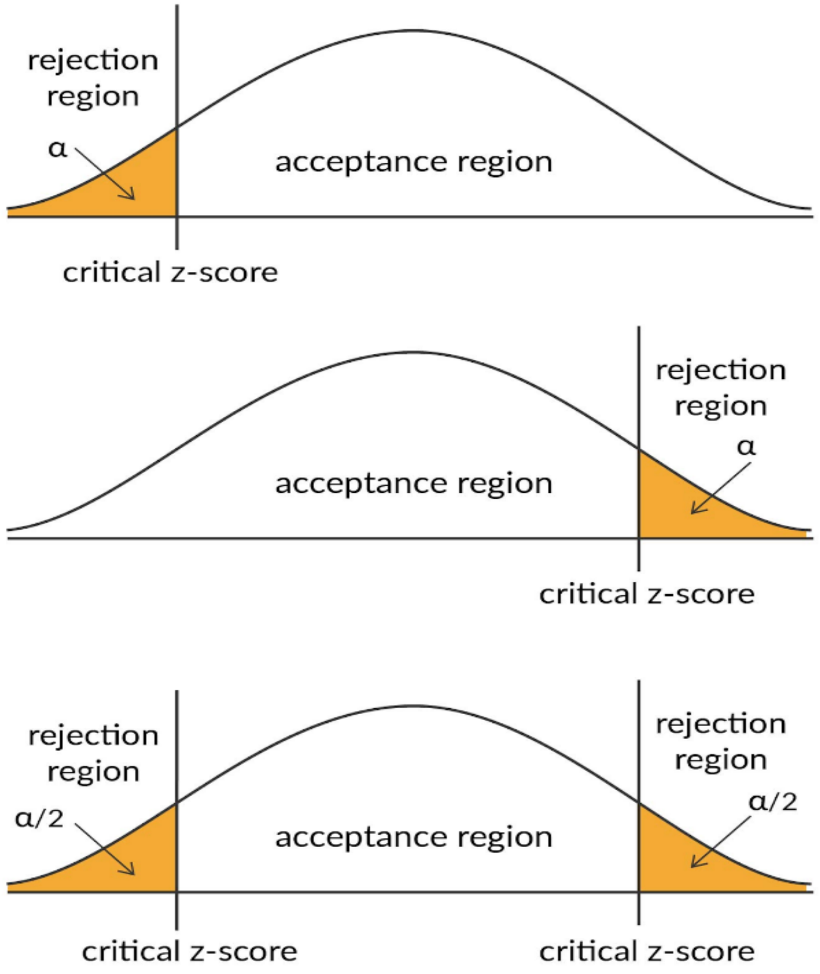
P-value method

- Here our p-value is 0.136, which is more than 0.05, we fail to reject our null at 95% confidence (or 5% significance level)
- For a two tailed test, p-value falls on both sides
- For the left tailed test, p-value will be on the left side of the curve



P-value method

- For the right tailed test, p-value will be on the right side of the curve
- Once the p-value is available, the condition to reject the null remain the same



Summary and Concluding Remarks

Summary

- We discussed hypothesis testing to check the validity of claims with statistical rigor
- You defined the null and alternate hypothesis
- You computed the z-score from sample data points and compared it to the critical z-value
- You also saw the p-value method, which was the probability at tails (significance level)
- A lower p-value meant the higher chance (confidence) of rejecting the null

Thanks!