DONE BY LUKE YEO, NOTE THIS IS ALL DONE FOR LEARNING , NOT FOR ANYTHING ELSE

CREDITS TO:

FREECODECAMP: https://youtu.be/FTpmwX94_Yo?si=j64-jlPtn7MiqfzQ

RESOURCES: https://github.com/TatevKaren/CaseStudies/tree/main/AB%20Testing

Definition:

A/B testing is a method used in marketing and product development to compare two versions (A and B) of a webpage, app screen, or marketing material to determine which one performs better. Here's what it involves:

1. **Purpose**: A/B testing helps businesses understand which version of a webpage, email, or app design leads to better conversion rates or user engagement.
2. **Process**: It involves dividing your audience into two groups and showing each group a different version (A and B). Statistical analysis is then used to determine which version performs better based on predefined metrics (like click-through rates, conversions, or time spent on page).
3. **Benefits**: A/B testing allows companies to make data-driven decisions, optimize their marketing efforts, and improve user experience by identifying which design or content resonates better with their audience.

Overall, A/B testing is a valuable tool for optimizing digital products and marketing campaigns by leveraging real user data to improve performance.

It is basically use for testing new stuff (feature, UI, ect) using a control group (current version) and an experimental group (new version). This allows direct feedback for the business.

A/B Testing process:

1) Hypothesis of A/B Test
   – how to measure success using the primary matric (what we want)
2) A/B Test Design / Power Analysis
   – design the test, make assumptions (what we expect), minimum datasize, ect
3) Run the A/B Test
   – Ensuring that the test runs smoothly and successfully
4) Result Analysis / Statistical Significance
   – Using statistic like Z-test, T-test, Chi-square Test, ect on the data gathered.
5) Result Analysis / Practical Significance

Hypothesis of A/B testing:

- Business Hypothesis: Compare A and B, what is desire outcome, will affect the Key Performance Indicator (KPIs). Normally by brainstorming from different perspective.
  E.g. Changing color will change the amount of engagement

- Primary Metric (only 1): is a way to measure the performance of the product. it will be use to identify if there is a statistically significant difference between the 2 groups.
  To test accuracy, and choosing the Primary Metric, use the Metric Validity Question (if this chosen metric were to increase a lot while everything else stay put, would we achieve our goal and address the problem)
  E.g. Revenue: Conversion rate = Number of conversion/total number of visitors X 100%

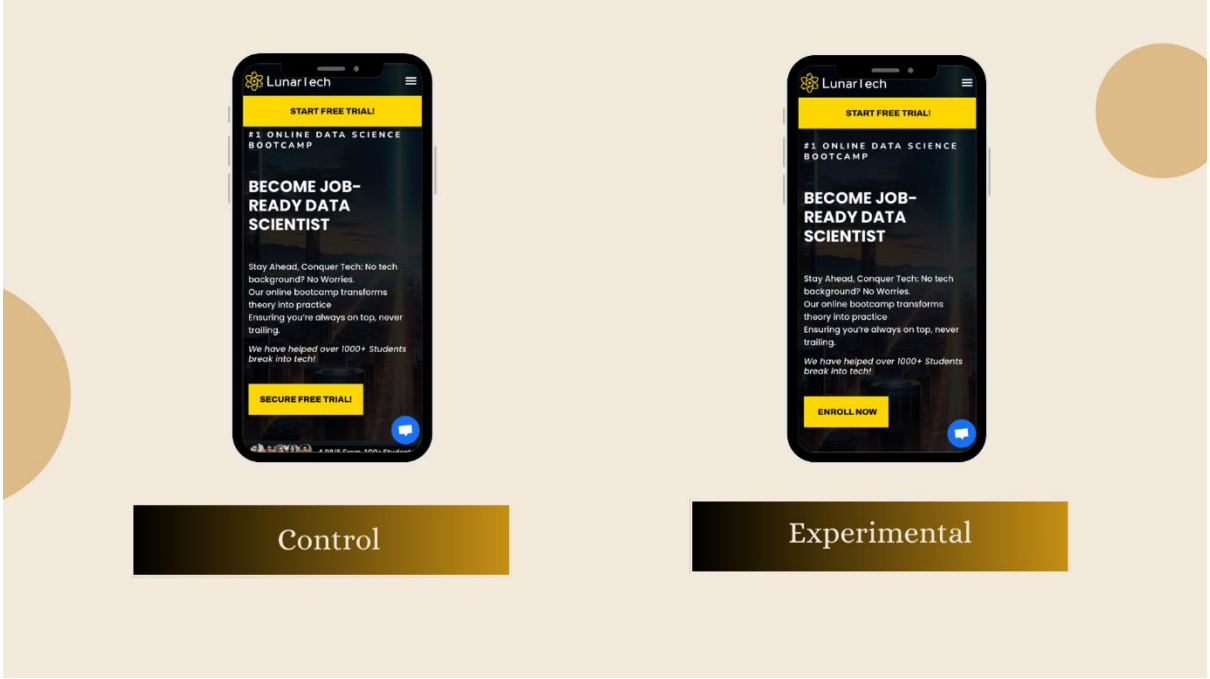
- Statistical Hypothesis: a statistical procedure that is use to determine whether there is a diff between observed data and expected data. This will help to find and fix the solution (Null hypothesis and Alternative hypothesis)

- A/B Test design: steps are (Power Analysis , Minimum Sample Size, Test duration)

Follow Base guide on what to do

End goal: To understand if we should release the new button or keep the old one

Want to find out which does customers click more, which is more engaging, remember, the result must be data driven:



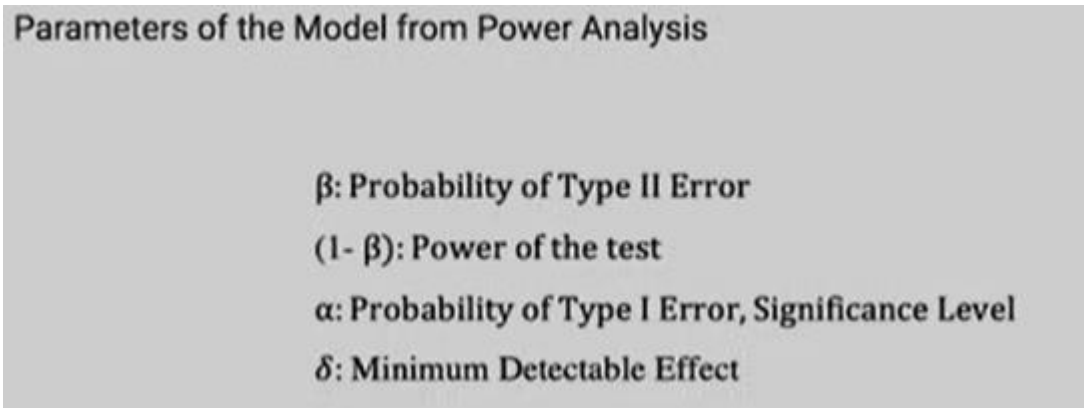
In stats meaning:

H0: Pcon = Pexp -> The old secure free trail is better than the new version

H1: Pcon != Pexp -> The new enroll now is better than the old version

Coding Steps:

1) Read the file, group it by total number of clicks
2) Show a graph of the click distribution between the 2 groups
3) Power analysis



4) Get more details in regards to the 2 groups
5) Calculate the Pooled Estimate, Pooled Variance, Pooled Standard Error and Test statistic -> Z critical value
   - **Pooled Estimate:**A pooled estimate is a combined estimate of a parameter (like a mean or proportion) derived from multiple groups or datasets. This method is particularly useful when you want to create a single summary statistic that represents all the groups collectively.
   - **Pooled Variance:** Pooled variance is a method for estimating the variance (or spread) of a population when data is collected from different groups that are assumed to have the same true variance but different sample means. This method pools the variances of the different groups into a single estimate that represents the common variance across the groups. It's commonly used in t-tests, particularly when comparing the means of two independent samples (assuming equal variances).
   - A 2-sample t-test (also known as an independent t-test) is a statistical test used to determine whether there is a significant difference between the means of two independent groups. This test is commonly used in experiments or studies where two groups are compared, such as a treatment group versus a control group.

     When to Use:

     - When comparing the means of two independent groups.
     - When the data is continuous (like height, weight, or test scores).
     - When the data from both groups are approximately normally distributed.
     - When the variances of the two groups are similar (homogeneity of variance).

   - **Z Critical Value:**
     The Z critical value is a point on the standard normal distribution (a normal distribution with a mean of 0 and a standard deviation of 1) that corresponds to a specified confidence level. It is commonly used in Z-tests, which are used when the population standard deviation is known or the sample size is large.
     When to Use:

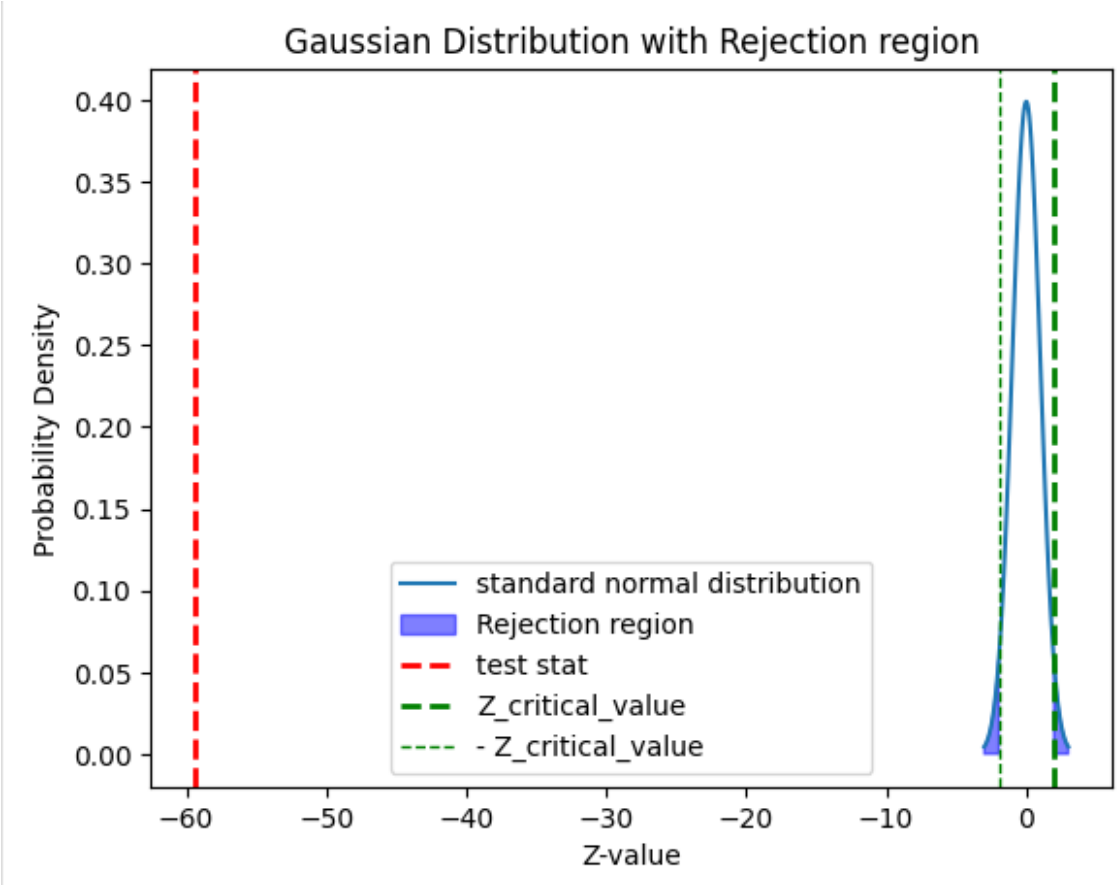
- In hypothesis testing when the population standard deviation is known.
- In large sample sizes (typically n>30n > 30n>30).
- When constructing confidence intervals for a population mean or proportion.

Confidence Levels and Z Critical Values:

For a **95% confidence level**, the Z critical value is approximately **1.96**.

For a **99% confidence level**, the Z critical value is approximately **2.58**.

6) Calculate the P-test of the Z-test.
   - If P-value is <= 0.05, it indicates strong evidence against the null hypothesis, so we reject the null hypothesis
   - If P-value is >= 0.05, it indicates weak evidence against the null hypothesis, so we fail to reject the null hypothesis
7) Calculate show the confidence interval of the test:



The graph you've generated represents a standard normal distribution along with marked rejection regions, the test statistic, and critical Z-values. Here's what the graph shows:

1. Standard Normal Distribution:

- The bell-shaped curve on the right side of the plot is the standard normal distribution, centered around 0 with a mean (`mu`) of 0 and a standard deviation (`sigma`) of 1.

- The curve represents the probability density function (PDF) of the normal distribution, which is highest at the mean (0) and decreases symmetrically as you move away from the mean.

2. Rejection Region (Shaded Blue Area):

- The blue shaded areas on either side of the standard normal curve represent the rejection regions.

- These regions are beyond the critical Z-values, where the null hypothesis would be rejected. The graph shows that these regions are close to the tails of the distribution.

3. Test Statistic (Red Dashed Line):

- The red dashed vertical line on the far left at approximately -59.44 represents the test statistic.

- This value is extremely far from the mean of the distribution, which indicates that the test statistic is highly unusual if the null hypothesis were true.

4. Z Critical Values (Green Dashed Lines):

- The green dashed lines on either side of the mean at \(Z_{critical}\) and \(-Z_{critical}\) represent the critical Z-values.

- These lines mark the boundaries of the rejection regions.

Interpretation of the Graph:

-Extreme Test Statistic: The test statistic is far outside the range of the standard normal distribution, beyond the Z-critical values. This indicates that if the null hypothesis were true, the probability of obtaining such a test statistic would be extremely low.

- Rejecting the Null Hypothesis: Since the test statistic falls well within the rejection region (actually, it is far beyond it), you would reject the null hypothesis with strong evidence against it.

The extreme position of the test statistic suggests that the observed data is very unlikely under the null hypothesis, leading to a rejection of the null hypothesis in favor of the alternative hypothesis.

8) Testing for Practical Significance