Day 4 (27.03.2024)

1. You have collected exam scores from two different groups of students. Write python code for the following:
i. using point estimates, calculate the mean score for each group. Then, construct 95% confidence intervals for the mean scores.
ii. Perform a hypothesis test to determine if there is a significant difference between the two groups.
2. In a manufacturing setting, you are tasked with estimating the average lifespan of two different types of products. Develop python code for the following.
i. point estimates for the mean lifespan of each product type and construct 90% confidence intervals.
ii. Perform a hypothesis test to assess whether there is a statistically significant difference in the lifespans.

3. You are a data scientist working for an e-commerce company. The marketing team has conducted an A/B test to evaluate the effectiveness of two different website designs (A and B) in terms of conversion rate. They randomly divided the website visitors into two groups, with one group experiencing design A and the other experiencing design B. After a week of data collection, you now have the conversion rate data for both groups. You want to determine whether there is a statistically significant difference in the mean conversion rates between the two website designs. Write python code for above said scenario.

4. You are a researcher working in a medical lab, investigating the effectiveness of a new treatment for a specific disease. You have collected data from a clinical trial with two groups: a control group receiving a placebo, and a treatment group receiving the new drug. Your goal is to analyze the data using hypothesis testing and calculate the p-value to determine if the new treatment has a statistically significant effect compared to the placebo. Write a Python program for the scenario and also use matplotlib library to visualize the data and the p-value.

5. Imagine you are an analyst for a popular online shopping website. Your task is to analyze customer reviews and provide insights on the average rating and customer satisfaction level for a specific product category. You have been provided with a CSV file named "customer_reviews.csv," which contains customer ratings for products in the chosen category. Your task is to write a python program using pandas library to calculate confidence intervals to estimate the true population mean rating. The following are columns of CSV file.

id, name, brand categories, manufacturer, manufacturer Number, reviews date,
reviews id, reviews rating, reviews text