# Advanced Hypothesis Testing

Armand Junior DONGMO NOTUE

College of Engineering and Technology, Grand Canyon University

DSC-510-O500: Advanced Probability and Statistic

Edward Ofori

03/27/2024

## Task1: Hypothesis Testing for Paired Data

### Introduction

The Affordable Care Act (ACA) is a comprehensive healthcare reform law enacted in 2010. This report aims to analyze the impact of ACA on health insurance coverage across states in the USA. We will conduct a hypothesis test for paired data to determine if there has been a significant change in uninsured rates before and after Obamacare.

### Problem Statement

The research question we seek to address is: "Has the implementation of the Affordable Care Act led to a significant reduction in uninsured rates across states?"

### Data Description

The dataset used contains health insurance coverage data for each state and the nation as a whole. It includes variables such as uninsured rates before and after Obamacare, estimates of individuals covered by employer and marketplace healthcare plans, and enrollment in Medicare and Medicaid programs.

### Methodology

We will employ a paired t-test to compare uninsured rates before and after Obamacare implementation. The assumptions of the t-test include normality of differences and independence of observations.

### Hypothesis Testing

We will test the null hypothesis (H0): "There is no significant difference in uninsured rates before and after Obamacare." Against the alternative hypothesis (H1): "There is a significant difference in uninsured rates before and after Obamacare."

### Results and Discussion

The paired t-test yields a p-value of $< 0.05$, indicating a significant difference in uninsured rates before and after Obamacare. This suggests that the implementation of ACA has led to a reduction in uninsured rates across states.

### Visualization

Visual representation of uninsured rates before and after Obamacare using a before-and-after plot (Boxplot).

### Conclusion

The analysis indicates a statistically significant reduction in uninsured rates following the implementation of the Affordable Care Act. This suggests that ACA has been effective in expanding healthcare coverage and reducing the number of uninsured individuals across states.

## Task2: Life Expectancy in North America

### Introduction

This documentation provides an analysis of life expectancy data for countries in North America, namely the USA, Canada, and Bermuda. The report includes an overview of the dataset, statistical analysis, visualizations, and conclusions drawn from the analysis.

### Dataset Overview

The dataset contains information on life expectancy for each country in North America. The data includes the following variables:

- Country: Name of the country
- Life Expectancy: Average life expectancy in years

### Problem Statement

The objective of this analysis is to compare the life expectancy across different countries in North America and identify any significant differences.

### Statistical Analysis

We performed a descriptive analysis to understand the distribution of life expectancy values for each country. Additionally, we conducted hypothesis testing to compare the mean life expectancy across the three countries.

### Descriptive Analysis

- Mean Life Expectancy:
    - USA: 78.136 years
    - Canada: 80.99 years
    - Bermuda: 79.927years

### Hypothesis Testing

We used an analysis of variance (ANOVA) test to compare average life expectancy between these three countries. The null hypothesis (H0) assumes that there is no significant difference in mean life expectancy between countries.

.

### Results

The results of the ANOVA test indicate a statistically significant difference in the mean life expectancy across the three countries (F-statistic = 15.87, p-value < 0.05).

## Visualizations

We created a box plot to visualize the average life expectancy for each country in North America. The plot highlights the differences in life expectancy among the countries.

## Conclusion

In conclusion, our analysis reveals variations in life expectancy across North American countries. Canada has the highest average life expectancy, followed by Bermuda and the USA. These findings have implications for healthcare policies and interventions aimed at improving life expectancy in the region.

**Task3: Analysis of Poverty-Level Wages in the United States**

**Introduction:** This report aims to analyze poverty-level wages in the United States over the period from 1973 to 2022. The dataset provides information on poverty-level wages for different ethnic groups, including Hispanics, Blacks, and Whites. We will explore trends in poverty-level wages over time and compare wage disparities among these ethnic groups.

**Dataset Description:** The dataset contains wage data for Hispanics, Blacks, and Whites in the United States from 1973 to 2022. Each record includes the poverty-level wages for a specific year and ethnic group.

**Research Questions:**

1. How have poverty-level wages evolved over the years for Hispanics, Blacks, and Whites?
2. Are there significant differences in poverty-level wages among these ethnic groups?
3. What are the wage disparities between Hispanics, Blacks, and Whites over the 50-year period?

**Data Analysis:**

- We will begin by loading the dataset and performing exploratory data analysis (EDA) to understand the distribution of poverty-level wages over time for each ethnic group.
- Next, we will calculate summary statistics, including mean, median, and standard deviation, to describe the central tendency and variability of poverty-level wages.
- To compare wage disparities among ethnic groups, we will conduct hypothesis tests.
- Additionally, we will visualize trends in poverty-level wages using line plots and box plots to provide insights into wage disparities over time.

**Assumptions and Considerations:**

- We assume that the dataset is representative and accurately reflects poverty-level wages for Hispanics, Blacks, and Whites in the United States.
- We will consider potential biases or limitations in the data collection process that may affect the interpretation of results.
- Assumptions of statistical tests, including normality and homogeneity of variances, will be assessed to ensure the validity of our analysis.

**Results and Interpretation:**

This means that there is insufficient evidence to conclude that there are significant differences in variance between the revenues of different ethnic groups.

**Conclusion:**

This report will contribute to our understanding of poverty-level wages in the United States and shed light on disparities among ethnic groups. By examining historical trends and conducting

statistical analysis, we aim to identify areas for policy intervention and address socioeconomic inequalities in wages.

References:

Jim, F. Paired T Test: Definition & When to Use It. Statistics by Jim. https://statisticsbyjim.com/hypothesis-testing/paired-t-test/

statistics solutions. Directory of Statistical Analyses. https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/

**Videos :**

Video1 : https://www.loom.com/share/d4e19f76b9c14a59a935b451b4402b4d
Video2: https://www.loom.com/share/9c87f8586bf94e8a94a39cddfaf8f1a4
Video3: https://www.loom.com/share/6282e7fdc5e940d997e7910d5ac1e432

**Data Set:**

Dataset1: https://www.kaggle.com/datasets/hhs/health-insurance/data
Dataset2: https://www.kaggle.com/datasets/mjshri23/life-expectancy-and-socio-economic-world-bank
Dataset3: https://www.kaggle.com/datasets/asaniczka/poverty-level-wages-in-the-usa-dataset-1973-2022

**GitHub:**

GitHub: https://github.com/ARMAND-cod-eng/Advanced-Hypothesis-Testing/blob/main/Advanced%20Hypothesis%20Testing.ipynb