Assignment 14: Create a Function to Calculate the T-Value

74 Points scaled to 20 Points

Introduction

In this assignment, you will create a function that performs a T-Test and returns the resulting T-Value, p-value, degrees of freedom, Group 1 mean, and Group 2 mean. You will then use your function to assess for differences in mean annual temperature averaged over a county ("tempmn") within two different subregions of the United States (SUB_REGIONS"): Mid-Atlantic and New England. You will make use of the "us_county_ttest.csv" dataset. This is a subset of the "us_counties_data.csv" data in which 40 random counties were sampled within the two subregions of interest. Make sure to read the associated description of the data ("us_counties_data_ DESCRIPTION.pdf") prior to undertaking the assignment. Note that not all of the assumptions of a T-Test may be met. For example, there is likely some spatial autocorrelation in the data, which may invalidate the assumption of independent samples. However, we will not worry about testing assumptions and assume that the results are valid.

Objectives

- *Conduct a two-tailed, 2-Sample T-Test and interpret the results*
- *Create a function that performs a set of calculations and returns results*

Deliverables

- *Jupyter Notebook (Python) or R Markdown file (R) with all code and answers to the questions embedded. Files can be rendered to HTML webpages if your instructor requires this. Questions should be stated and answered within Markdown cells.*

Background Questions

Question 1. State the T-Test null hypothesis for this specific research question. (4 Points)

Question 2. State the T-Test alternative hypothesis for this specific research question. (4 Points)

Question 3. Explain how the degrees of freedom are determined for a 2-Sample T-Test? (4 Points)

Question 4. What is the difference between a 2-Sample T-Test and a Paired T-Test? (8 Points)

Question 5. Explain the difference in calculating the T-Value when the two groups are assumed to have the same variance in comparison to when they are assumed to have a different variance. (8 Points)

T-Test Function and Questions

This assignment can be conducted using either Python or R, whichever you prefer or whichever you instructor requires. Generate code to perform the following tasks and answer the associated questions.

Task 1. Read in the data table. (4 Points)

Task 2. Create a grouped boxplot that shows the distribution of mean annual temperature within the two groups. (6 Points)

Task 3. Generate a function that calculates and returns the T-Value, p-value, degrees of freedom, Group 1 mean, and Group 2 mean. The calculations should assume that the groups do not have equal variance and that the test is two-sided. You cannot use a function that performs a T-Test directly (e.g., t.test() in R). Instead, you must perform the required calculations in code. You can use a function that calculates a p-value if provided the T-Value and the degrees of freedom. This is too difficult to code on your own. (20 Points)

Question 6. Based on the grouped boxplot, what do you expect to be the results of the T-Test? Explain your reasoning. (8 Points)

Question 7. State the findings of the T-Test using standard statistical language. (8 Points)