Final Project - Checkpoint 2

STAT 327 Winter 2024: Statistics for the Physical and Environmental Sciences

Hala Abdelbaki, Zoya Malik

April 2, 2024

---

https://data.world/makeovermonday/2021w3

[1] Group Members

Hala Abdelbaki and Zoya Malik

[2] Our Plan

We plan to continue using the same dataset as initially described in Checkpoint 1.

[3] Research Question

In Checkpoint 1, we stated multiple different research questions that we would potentially focus on. We have decided to narrow it down, and focus on the research questions:

1.

Based on historical data, can we make predictions about future temperature anomalies for specific seasons or months?

Since we planned on using the Regression test type for this research question, after doing some research on our own about what regression is, we also thought of rephrasing the questions as:

How does the historical data of temperature anomalies predict future anomalies for specific seasons or months in the Northern and Southern Hemispheres?

This might be refined further after we cover Regression in class completely, but the main point and goal of the question will remain the same.

2.

Are there any significant differences in temperature anomalies between the Northern and Southern Hemisphere?

We plan on researching this question as well, as we think the question aligns well with the dataset's structure, specifically the Hemisphere variable.

Both these questions are significant to the main goal of our project, as our aim was to come up with questions with high relevance in climate research. Both questions will help us determine long-term historical temperature and predictions, as well as a current comparison of climate patterns in different hemispheres.

[4] Statistical Inference Tests

We still plan on using the same statistical inference tests that we indicated in Checkpoint 1. For the second research question, we have yet to decide whether to use the Mann-Whitney U test, or the two-sample t-test.

1. *How does the historical data of temperature anomalies predict future anomalies for specific seasons or months in the Northern and Southern Hemispheres?*

Statistical Inference Test: Regression

Reason: Regression is used to examine the relationship between variables and can help in making predictions about future temperature anomalies based on historical data.

2. *Are there any significant differences in temperature anomalies between the Northern and Southern Hemisphere?*

Statistical Inference Test: Two-Sample t-test or Mann-Whitney U test, depending on the distribution and variance or the temperature anomaly data.
Reason: These tests compare the means (or medians, in the case of non-parametric tests like the Mann-Whitney U test) of the temperature anomalies between the two hemispheres to determine if there are statistically significant differences.

We will use the two-sample t-test if the temperature anomaly data for both hemispheres are normally distributed. It will test whether the mean temperature anomaly in the Northern Hemisphere is significantly different from that in the Southern Hemisphere.

If the data are not normally distributed or have unequal variances, we will use the Mann-Whitney U test (a non-parametric test) as this test compares the medians between the two groups without assuming normal distribution.

[5] Issues or Concerns

The only issue or concern that may arise is if the test does not go as planned or does not work. Mainly for the hypothesis question involving a Regression test, the assumption of linearity is where we could come upon a problem. If the relationship between the independent and dependent variables is not linear, then the model might not accurately capture the trends which would lead to erroneous predictions. Also, any outliers in the dataset could skew our results.