

# Inference for Paired Observations

Mini-Assignment - MTH 361 A/B - Spring 2023

## Instructions:

- Please provide complete solutions for each problem. If it involves mathematical computations, explanations, or analysis, please provide your reasoning or detailed solutions.
- Note that some problems have multiple solutions or ways to solve it. Make sure that your solutions are clear enough to showcase your work and understanding of the material.
- Creativity and collaborations are encouraged. Use all of the resources you have and what you need to complete the mini-assignment. Each student must take personal responsibility and submit their work individually. Please abide by the University of Portland Academic Honor Principle.
- **Please save your work as one pdf file, don't put your name in any part of the document, and submit it to the Teams Assignments for this course. Your document upload will correspond to your name automatically in Teams.**
- If you have questions or concerns, please feel free to ask the instructor.

## I. Paired Data Inference

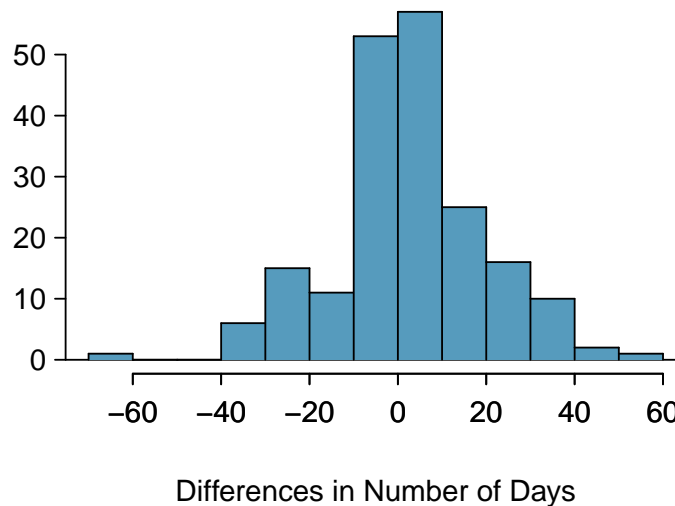
### Materials

The exercises below are derived from the textbook [OpenIntro Statistics \(4th edition\)](#) by David Diez, Mine Cetinkaya-Rundel, and Christopher Barr.

### Exercises

1. **Paired or not?.** In each of the following scenarios, determine if the data are paired.
  - a. Compare pre- (beginning of semester) and post-test (end of semester) scores of students.
  - b. Assess gender-related salary gap by comparing salaries of randomly sampled men and women.
  - c. Compare artery thicknesses at the beginning of a study and after 2 years of taking Vitamin E for the same group of patients.
  - d. Assess effectiveness of a diet regimen by comparing the before and after weights of subjects.
  - e. We would like to know if Intel's stock and Southwest Airlines' stock have similar rates of return. To find out, we take a random sample of 50 days, and record Intel's and Southwest's stock on those same days.
  - f. We randomly sample 50 items from Target stores and note the price for each. Then we visit Walmart and collect the price for each of those same 50 items.
  - g. A school board would like to determine whether there is a difference in average SAT scores for students at one high school versus another high school in the district. To check, they take a simple random sample of 100 students from each high school.

2. **Global warming, Part I.** Let's consider a limited set of climate data, examining temperature differences in 1948 vs 2018. We sampled 197 locations from the National Oceanic and Atmospheric Administration's (NOAA) historical data, where the data was available for both years of interest. We want to know: were there more days with temperatures exceeding 90°F in 2018 or in 1948? ([Environmental Information \(NCEI\), n.d.](#)) The difference in number of days exceeding 90°F (number of days in 2018 - number of days in 1948) was calculated for each of the 197 locations. The average of these differences was 2.9 days with a standard deviation of 17.2 days. We are interested in determining whether these data provide strong evidence that there were more days in 2018 that exceeded 90°F from NOAA's weather stations.



- Is there a relationship between the observations collected in 1948 and 2018? Or are the observations in the two groups independent? Explain.
  - Write hypotheses for this research in symbols and in words.
  - Check the conditions required to complete this test. A histogram of the differences is given to the right.
  - Calculate the test statistic and find the p-value.
  - Use  $\alpha = 0.05$  to evaluate the test, and interpret your conclusion in context.
  - What type of error might we have made? Explain in context what the error means.
  - Based on the results of this hypothesis test, would you expect a confidence interval for the average difference between the number of days exceeding 90°F from 1948 and 2018 to include 0? Explain your reasoning.
3. (Outstanding Question) **Global warming, Part II.** We considered the change in the number of days exceeding 90°F from 1948 and 2018 at 197 randomly sampled locations from the NOAA database in Exercise (2). The mean and standard deviation of the reported differences are 2.9 days and 17.2 days.
- Calculate a 90% confidence interval for the average difference between number of days exceeding 90°F between 1948 and 2018. We've already checked the conditions for you.
  - Interpret the interval in context.
  - Does the confidence interval provide convincing evidence that there were more days exceeding 90°F in 2018 than in 1948 at NOAA stations? Explain.

**References**

Environmental Information (NCEI), N. C. for. (n.d.). Climate data online: Dataset discovery. In *Datasets / Climate Data Online (CDO) / National Climatic Data Center (NCDC)*, April 24, 2019. <https://www.ncdc.noaa.gov/cdo-web/datasets>