STAT 535 Team Project

Analysis of Seasonal Temperature Variability Using Bootstrap and Monte Carlo Methods

Marco Lopez
Aiden Nguyen
Jerry Obour
Shumeng Zhang

University of Massachusetts Amherst

Contents

- Introduction
- Research Question and Hypothesis
- Objectives
- Methods
- Results & Discussion
- Conclusion

University of Massachusetts Amherst

Topic:

Comparative Analysis of Seasonal Temperature Variability in Boston and Amherst Using Bootstrap and Monte Carlo Methods

Data source:

- Accuweather.com
- https://www.accuweather.com/en/us/downtown-boston/02108/june-weather/2626565

Information of dataset:

- 4 variables: Date, Week, High Temperature, Low Temperature

Goal:

- Use daily high temperature data to determine whether one city experiences higher or lower average temperatures in winter and summer.

Introduction

Research Question and Hypothesis

Research Question:

Is there a significant difference in the average summer and winter temperatures between Amherst and Boston?

Hypothesis:

$\mu A = \mu B$

$\mu A \neq \mu B$

Research Question & Hypothesis

Objectives

- Use Bootstrapping to assess if there is a significant difference in the average summer and winter temperatures between Amherst and Boston.

- Perform a Monte Carlo simulation and conduct a Power Study to estimate the robustness of the observed mean temperature difference between Amherst and Boston.

Objectives

Methods

1. **Data Processing**:

- Load HTML data and scrape important details
  - dates and temperatures.
- Create a new data frame
  - Amherst Summer
  - Boston Summer

Description: df [6 × 2]

	Date <dbl>	High_Temperature <dbl>
1	1	83
2	2	87
3	3	87
4	4	85
5	5	85
6	6	82

6 rows

2. Bootstrap Sampling and Hypothesis Testing

- Setup Variables:
  - Ensure both data are of equal length
  - Set alpha = 0.05
- Bootstrap sampling:
  - Set up sampling indices ‘var_name’
  - Use ‘var_name’ to sample from both data
- Hypothesis Test
  - Perform a two sample t-test with alpha level of significance.
  - Construct a 95% confidence interval and record results

Methods

University of Massachusetts Amherst

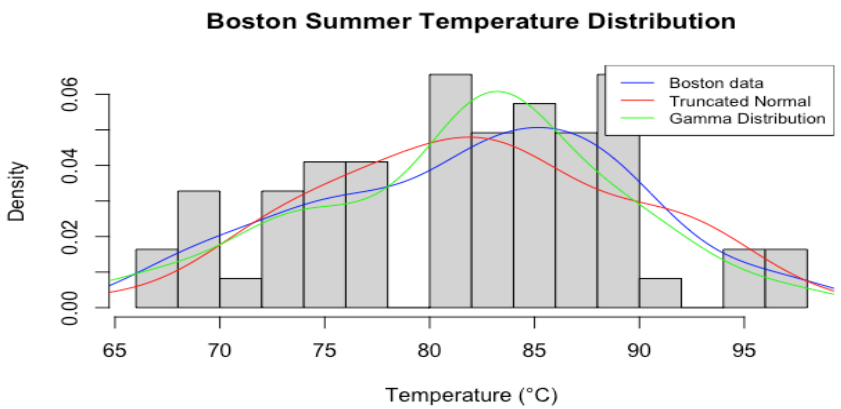
Bootstrap analysis of temperature data

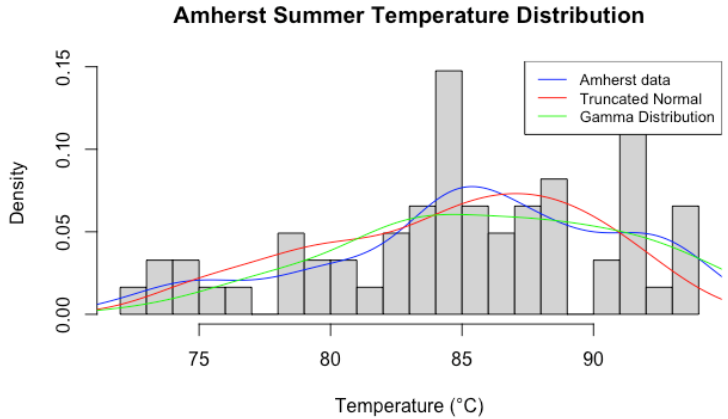
- **Bootstrap simulation**: Generated 10,000 bootstrap samples to estimate the distribution of difference in high temperature means between Amherst and Boston.
- **Confidence Interval**: A 95% confidence interval for difference in means was calculated (blue lines).
- **Mean difference**: The mean difference between the two cities was calculated (red line).

Methods

3. Monte-Carlo and Power Analysis

- Study the distribution of the data for simulation

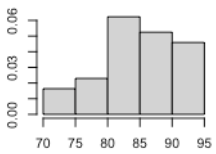


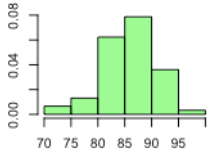


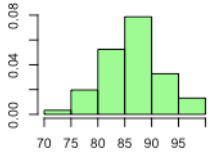
Methods

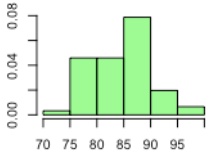
- Studying Distribution of data - Truncated Normal

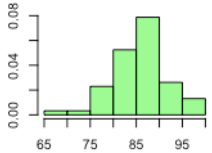
Amherst data

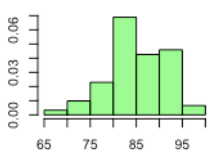




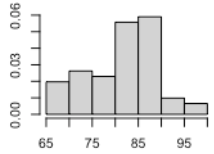


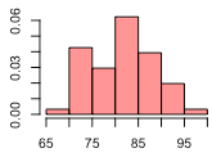


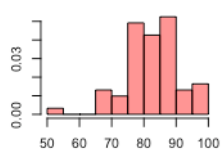


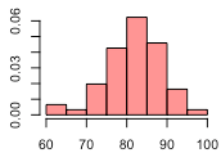


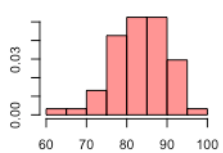
Boston data

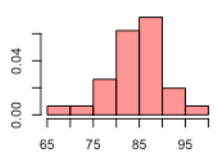








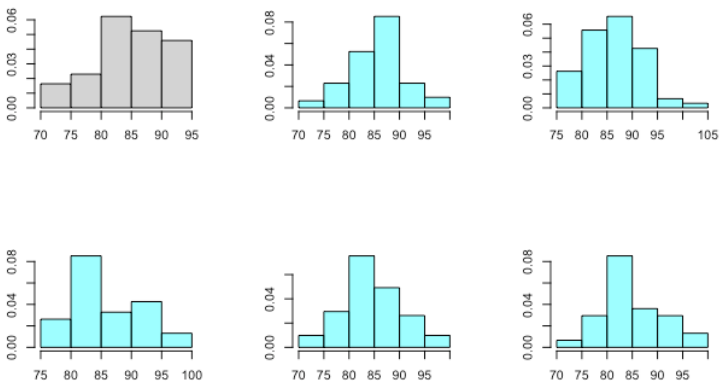




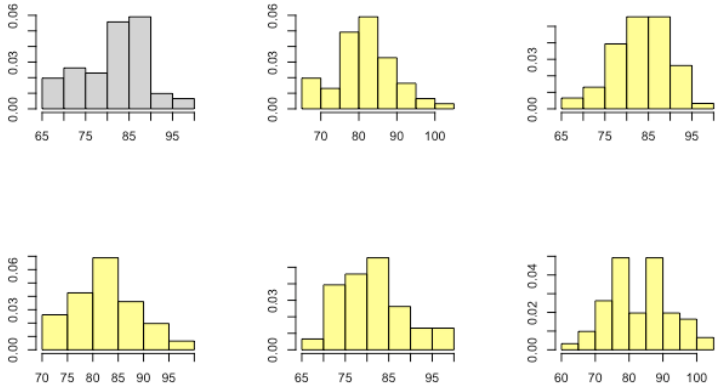
Methods

- Studying Distribution of data - Gamma Distribution

Amherst data



Boston data



Methods

- Choice of Distribution

  Truncated Normal / Gamma.

  Assumptions:

  - Skewness in the distribution of data is not extreme
  - Average temperature in the summer for both cities is in the range [40,100]

Methods

- Estimate parameters for both data

Description: df [2 × 4]

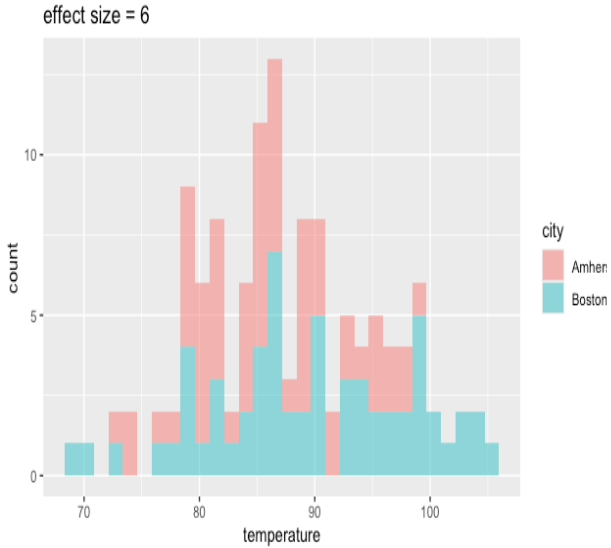
n <int>	mean <dbl>	sd <dbl>	city <chr>
61	85.42623	5.619784	Amherst
61	82.11475	7.583092	Boston

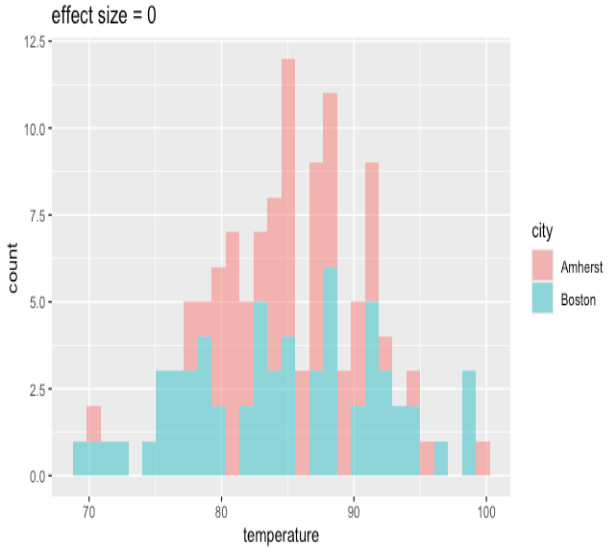
- Make a sampling function to generate data frame ensuring
  - Amherst sampling is done with sample from distribution
  - Boston sampling is done with sample from distribution + true difference in means + difference in means for true data
- Input arguments: data parameters & true difference

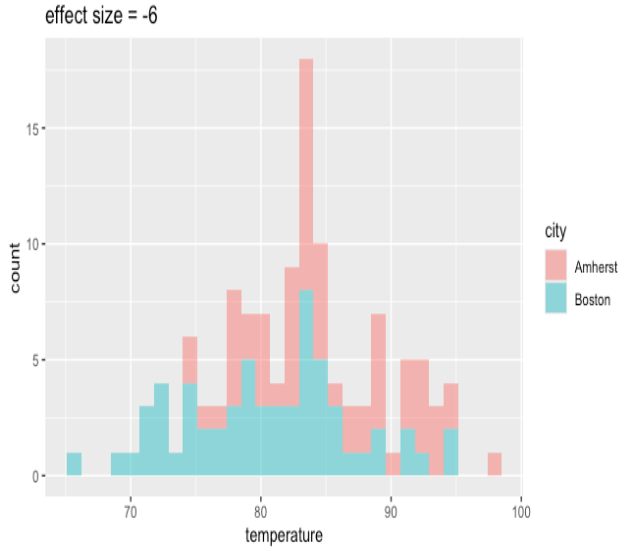
Methods

- Using estimated parameters & Effect size = {-6,0,6}

alpha = 0.05







Methods

- Find Power of the t-test

- This is accomplished with M = 2000 and alpha = 0.05
- Estimate the test at each iteration and return
  - 1 for p_value < alpha
  - 0 otherwise
- Estimate the probability of drawing 1's ; Which is the Power of the test:

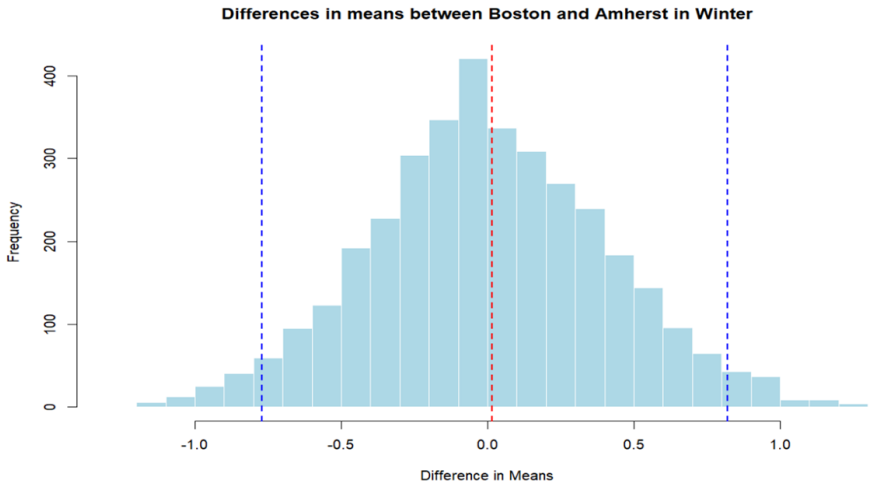
  P = sum(1's)/M

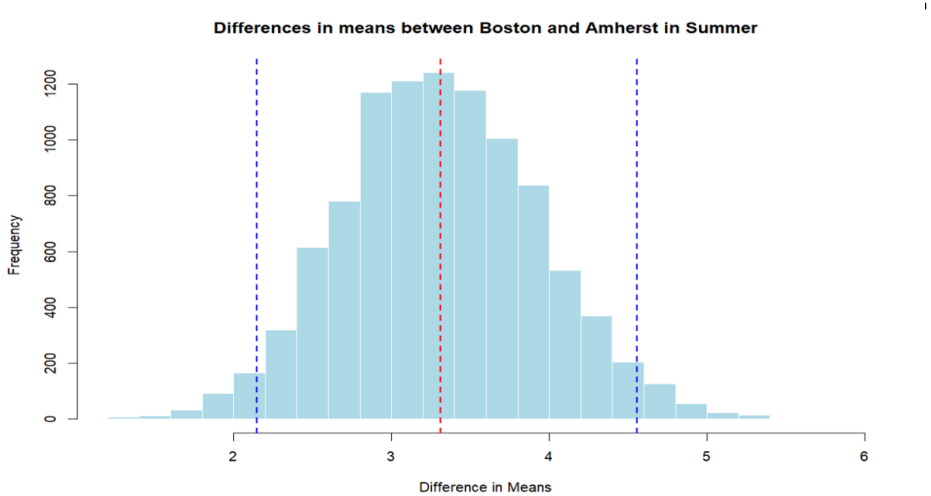
- Conduct a Power Study:

This is accomplished by varying the true difference parameter in the range [-6 , 6] with constant difference of 0.5

Methods

University of Massachusetts Amherst





Winter 95% CI: [-0.77, 0.82]

Summer 95% CI: [2.15, 4.56]

Based on this, the bootstrap says that there is no significant difference in winter and a significant difference in the summer.

Results & Discussion

University of Massachusetts Amherst

Results & Discussion

University of Massachusetts Amherst

- Study shows:
  - Test is most sensitive to large differences in mean temperatures.
  - Power is high as effect size approaches extremes
  - Power reduces as effect size approaches zero
- Implication:
  - Smaller temperature difference between the two locations are less likely to be detected as significant.
- It is a critical consideration for climatological studies that focus on subtle but potentially important differences.

Results & Discussion

University of Massachusetts Amherst

Conclusion:

- The findings of this study for the summer data are robust, indicating that the observed differences are unlikely to be due to chance. Such temperature differences could be attributed to various geographical and meteorological factors.
- For example, Boston's coastal proximity could cool its climate during the summer, whereas Amherst, being further inland, may experience warmer weather.
- These variations have implications for urban planning, agriculture, and energy management in both regions.
- The results of the power study suggest that while major temperature differences are readily detected, smaller variations might not always be identified, highlighting the need for enhanced sensitivity in measurement and analysis.

Conclusion

Thank you

University of Massachusetts Amherst