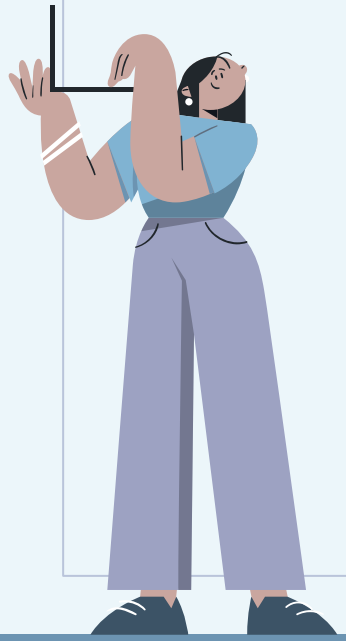
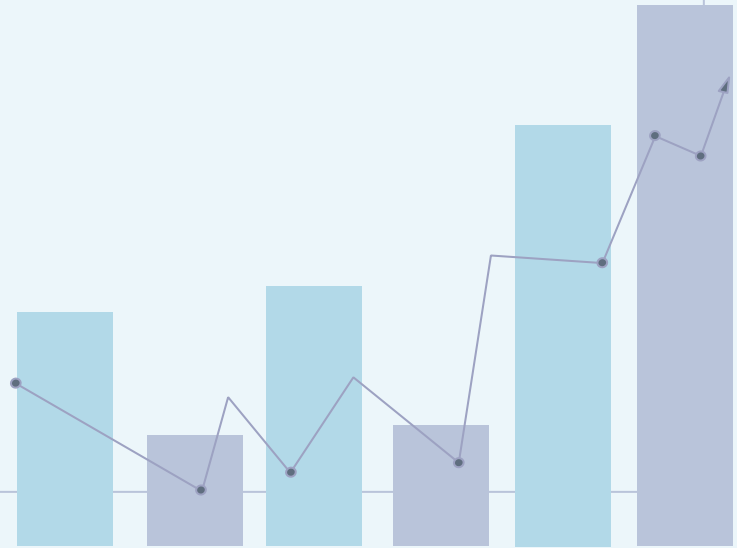
Authentic Data Explorations: Investigating the Normal Distribution through Comparative Rent Data



Maria Cruciani, Michigan State University
Justin Post, North Carolina State University
Jennifer Green, Michigan State University
Sunghwan Byun, North Carolina State University

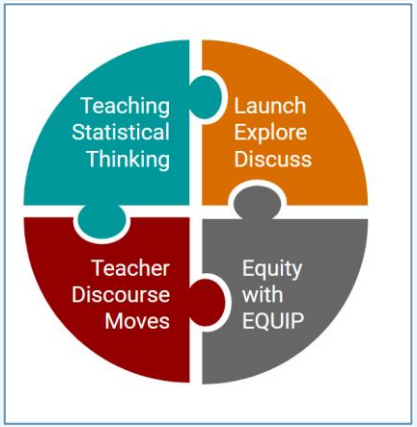
Research Overview and Purpose



GTAs-LEAD

- The authentic data exploration activities are a part of the NSF funded project: **G**raduate **T**eaching **A**ssistants **L**earning to Teach **E**quitably with **A**uthentic **D**ata (GTAs-LEAD)

- This project aims to create modules to help new teaching assistants learn to teach equitably, foster group work, and engage students with authentic data in statistics lessons



Implement

Three Authentic Data Explorations:

- **The Normal Distribution**
- **Confidence Intervals for P**
- **Linear Regression**

Why Develop Additional Activities?

- GAISE College Report (2016) has recommended integrating real multivariate data with a context and a purpose, teaching statistics as an investigative process, and fostering active learning.
- However, it is widely reported that statistics courses are not adhering to these recommendations, and are focusing on procedural aspects of statistical skills (Allen et al., 2010; Zieffler et al., 2008) or are using pseudo contexts (Gould, 2010; Horton 2015)
- Lessons and activities focused on the normal distribution didn't include engaging contexts and were focused on learning about the normal distribution through calculations

Development of The First Authentic Data Exploration:

Investigating the Normal Distribution through Comparative Rent Data

Overview

Lesson Title: Investigating the Normal Distribution through Comparative Rent Data
Level: Introductory Statistics at the High School or Postsecondary Levels
Duration: 60 Minutes
Objectives:

Task 1:

- Students will calculate, interpret, and draw conclusions from z-scores, percentiles, and proportions in addition to standardizing and comparing normally distributed data.
- Students will consider how data sources are gathered and fit to models and engage with an online applet

Task 2:

- Students will engage in a real-world scenario and make data-based decisions

Context: Students will be exploring the distributions of rent in each state. Rent is standardized as a percentage of income.

Data Exploration Goals

1. Cultivate Statistical Literacy

2. Engage with Authentic Data

3. Develop Technological Familiarity

4. Center Group Work

1. Cultivating Statistical Literacy

Task 1: Forming Groups and Choosing a State

Let's explore! Write down the state you chose and its distribution below:

_____~Normal(_____,_____)
Mean Standard Deviation

1. Using the mean and standard deviation for your chosen state, **calculate by hand** the percentage of rent that is one standard deviation greater than the mean.

2. **Use the applet** to find the probability that someone in the state is paying rent that is one standard deviation above the mean.

3. **Use the applet** to find the probability that someone is paying rent between the 30th and 70th percentile.

- Students will form groups, each choose a different state, and write down their state's distribution
- Students will first find probabilities and percentiles to describe the spread of their chosen state's distribution

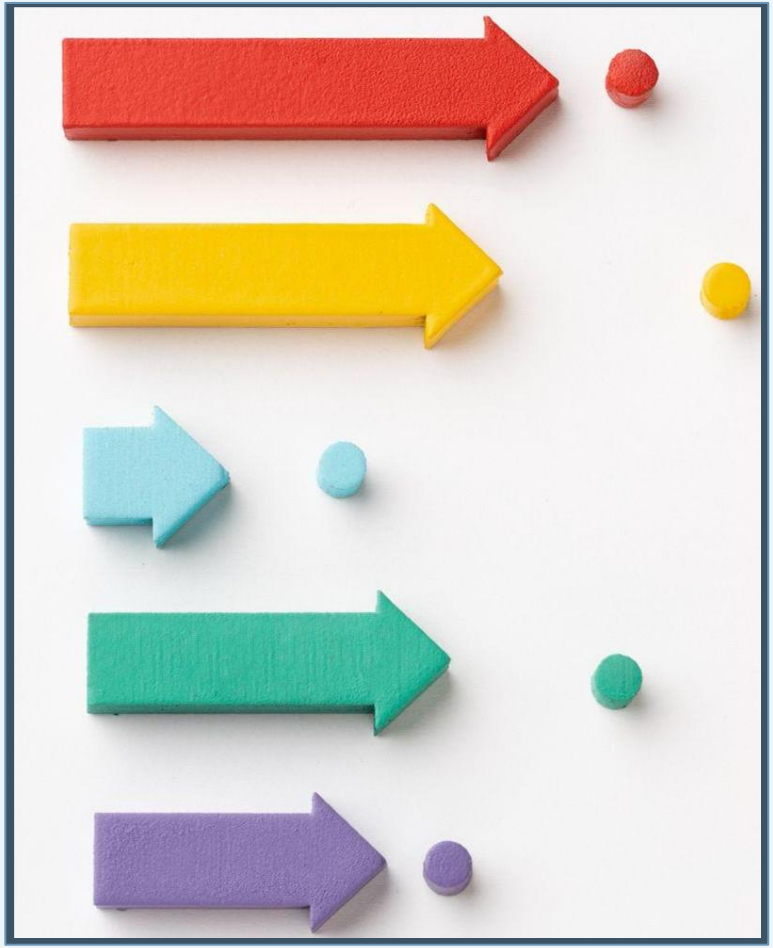
Task 1: Forming Groups and Choosing a State

4. Compare your answer to questions #1-3 **with your group**. Rank the chosen states by most to least variable. How did you decide this order?

6. Compare your answer to question #4 **with your group**.
   a. Which state had the largest percentile?

   b. Do you think we compare different populations based off percentiles? Why or why not?

- Students will begin to compare distributions, consider how to compare different distribution, and consider if their comparison methods are sound



Importance of Context:

This activity is set in a context that is meaningful to students and that is or will be impacting their lives. Students are seeing the fundamental purpose of fostering their statistical literacy while becoming of aware of phenomena that can be modeled by statistics and is of social and personal importance (Gal, 2004).

A Discussion on Rent and Income

- Students will discuss what they pay for rent and what income they have
- Students will look up the average local rent and average local salary

Consider: Income and Rent

Consider or look up:
How much is the average rent in your area? What is the average salary in your area? How much of your salary goes to rent? Do you think this is a fair amount?

We are going to consider the percentage of income that goes to paying monthly rent in each state using the normal distribution. Rent is measured as a percentage of household. Why do you think they chose to measure rent in this way? What problems might there have been with measuring rent just as a dollar amount?

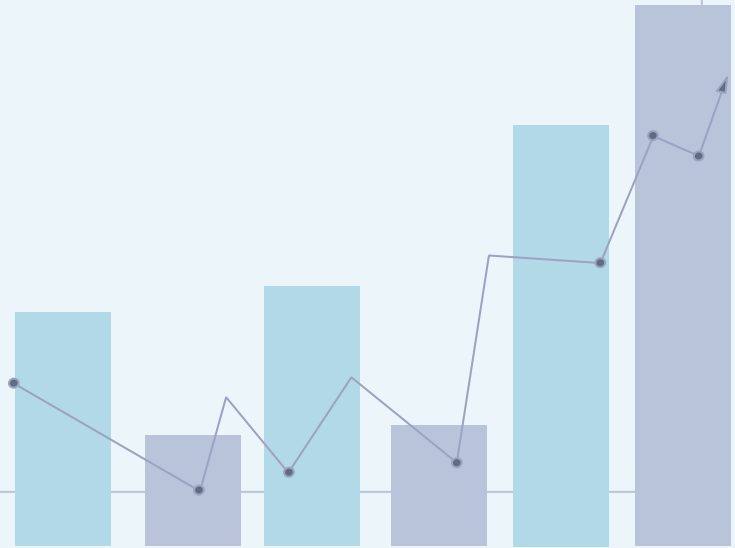
Task 2: Comparing and Standardizing Two Distributions

7. Use your answers to questions #3-6 to compare the spread of average rent paid for the two chosen states.

8. What questions would this exploration be helpful to answer? If you were writing an article about your exploration what else may readers want to know?

- Students will find percentiles and probabilities and make claims on their own after working in a group and exploring spread in Task 1.
- Students will extend their findings and consider real-world applications of the normal distribution
- Discussion will follow the activity

2.
Engaging with Authentic Data



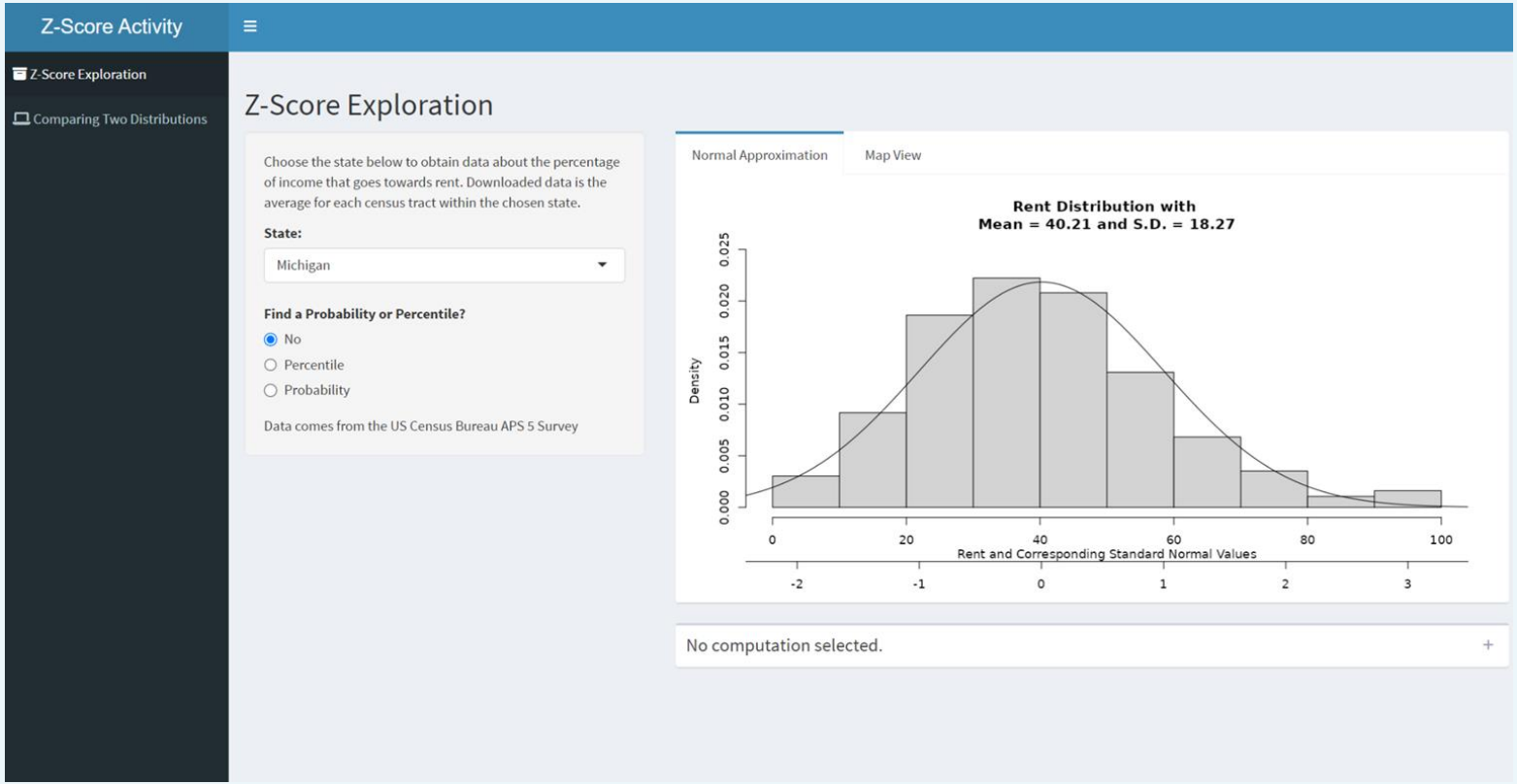
Developing Investigative Skills with Data

Purpose:
"Nearly 70% of the workforce will be expected to use and analytically interpret data by 2025" (Pothier & Condon, 2023)

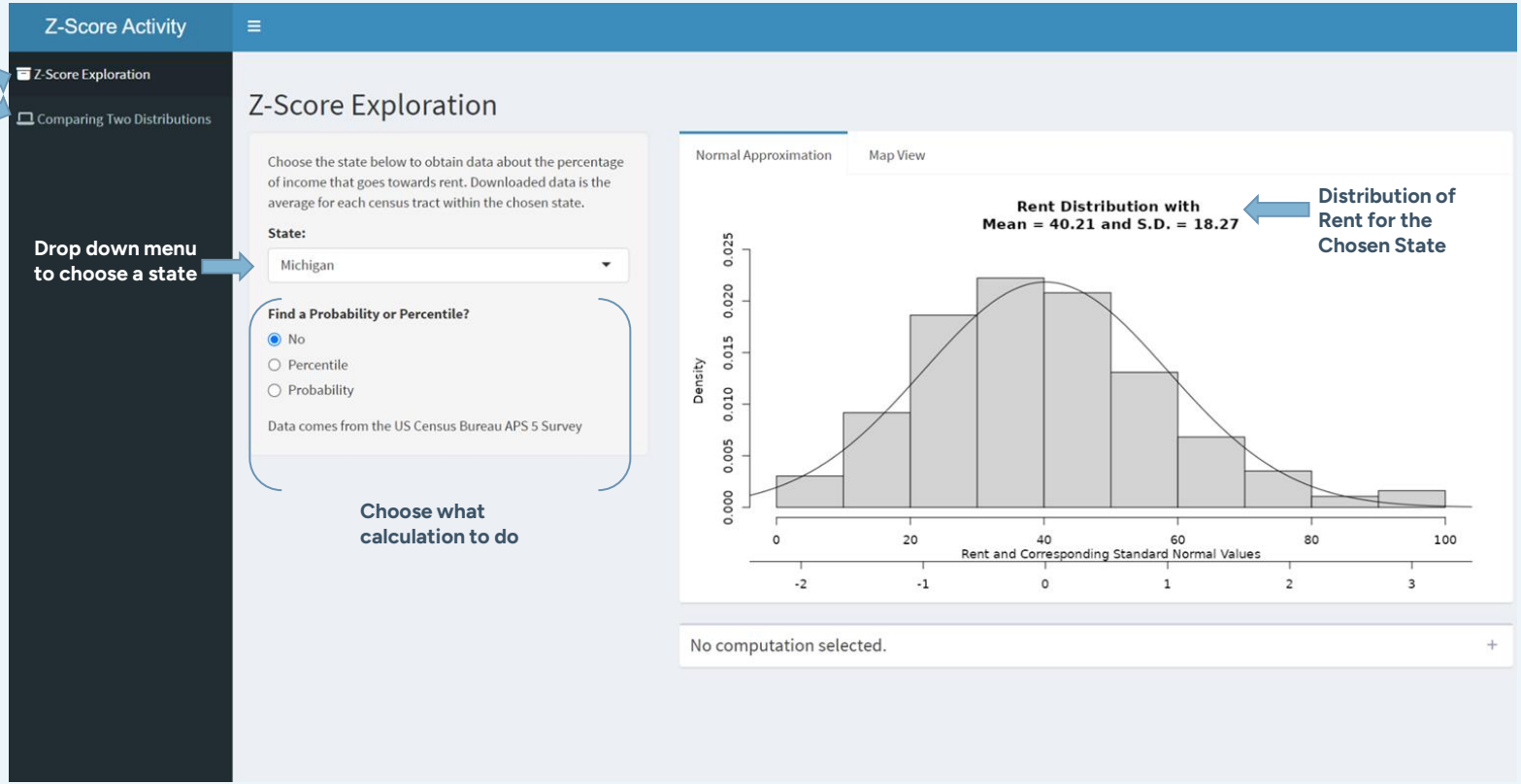
Elements in the Activity:
Students will develop data literacy through considering the source of this data, exploring how the US census bureau collects data on the general public, and what assumptions we can make about the data to create statistical models (Gould, 2017).

3. Developing Technological Familiarity

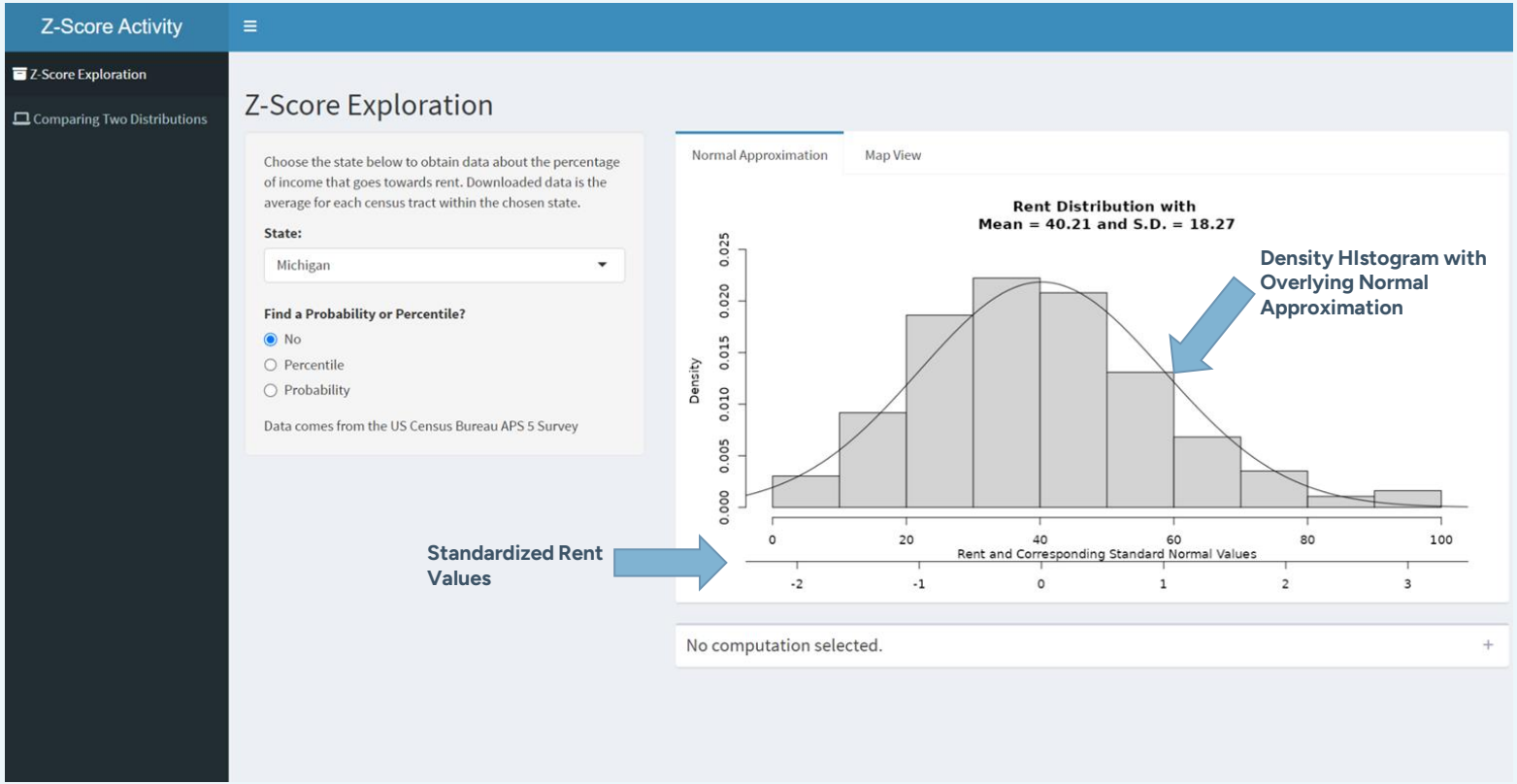
Intro: Getting Acquainted with the Normal Distribution and the Applet



Task 1: Describing a State's Distribution of Rent



Intro: Getting Acquainted with the Normal Distribution and the Applet



Task 1: Describing a State's Distribution of Rent

Z-Score Activity

Z-Score Exploration

Comparing Two Distributions

Z-Score Exploration

Choose the state below to obtain data about the percentage of income that goes towards rent. Downloaded data is the average for each census tract within the chosen state.

State:

Michigan

Find a Probability or Percentile?

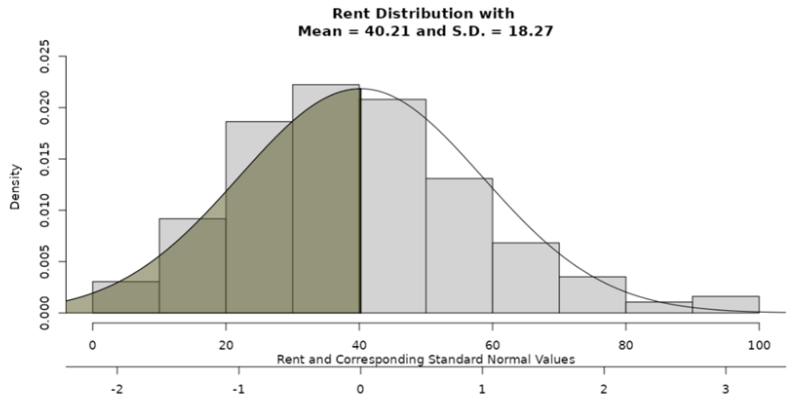
- No
- Percentile
- Probability

Percentile to find for Michigan:

0.5

Data comes from the US Census Bureau APS 5 Survey

Normal Approximation | Map View



Percentile value = 40.21

Task 1: Describing a State's Distribution of Rent

Z-Score Activity

Z-Score Exploration

Comparing Two Distributions

Z-Score Exploration

Choose the state below to obtain data about the percentage of income that goes towards rent. Downloaded data is the average for each census tract within the chosen state.

State:

Michigan

Find a Probability or Percentile?

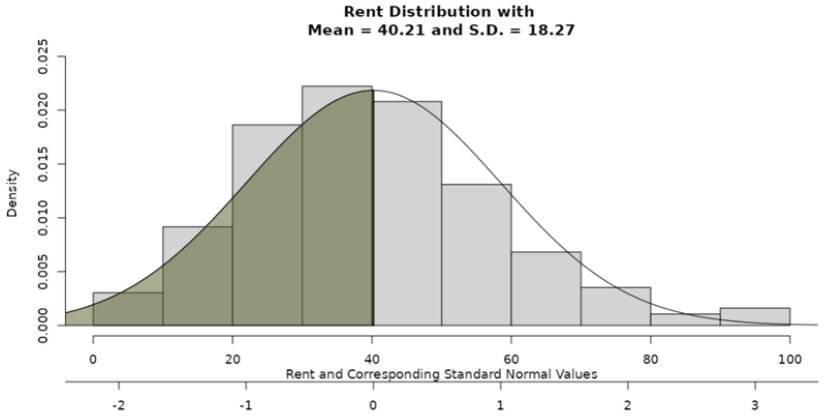
- No
- Percentile
- Probability

Percentile to find for Michigan:

0.5

Data comes from the US Census Bureau APS 5 Survey

Normal Approximation | Map View



Percentile value = 40.21

Calculation:

- 0.5 value from the standard normal is 0
- Converted to the distribution of rent we use:

```
rent value = mean + standard deviation * quantile
40.21 = 40.21 + 18.27 * 0
```

Task 2: Comparing and Standardizing Two Distributions

Z-Score Activity

Z-Score Exploration

Comparing Two Distributions

Comparing Two Distributions

Choose states below to obtain data about the percentage of income that goes towards rent. Downloaded data is the average for each census tract within the chosen states.

First State:

Louisiana

Second State:

Wisconsin

Two State Drop Down Menus

Find a Probability or Percentile for Louisiana?

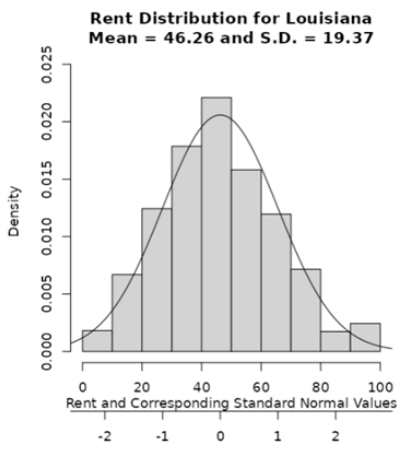
- (•) No
- ( ) Percentile
- ( ) Probability

Find a Probability or Percentile for Wisconsin?

- (•) No
- ( ) Percentile
- ( ) Probability

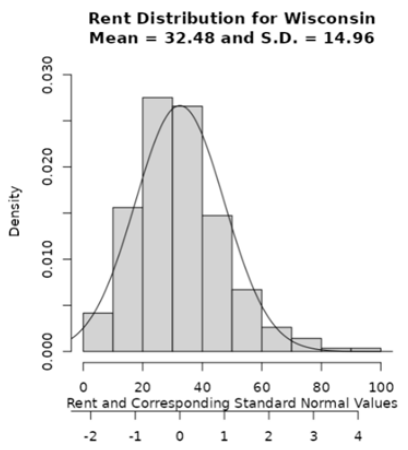
Data comes from the US Census Bureau APS 5 Survey

Normal Approximation | Map View



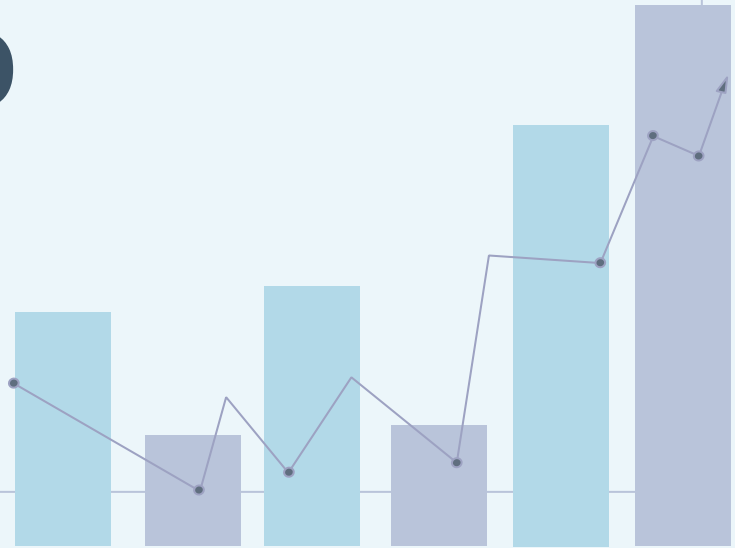
No computation selected. +

Normal Approximation | Map View



No computation selected. +

4.
A Pedagogical Focus on Group Work



Promoting Group Discussion:

4. Compare your answer to questions #1-3 **with your group**. Rank the chosen states by most to least variable. How did you decide this order?

6. Compare your answer to question #4 **with your group**.
   a. Which state had the largest percentile?

   b. Do you think we compare different populations based off percentiles? Why or why not?

- Questions throughout prompt students to work and discuss with their groups
- Specific questions within tasks 1 and 2 ask students to gather information from their groups, and compare what they have found

Next Steps:

- This activity along with the second and third are undergoing review
- Pilot runs of this activity will be implemented at two universities
- We look forward to sharing these three authentic data explorations with the broader community!

Acknowledgements

- The National Science Foundation
- US Census Bureau APS 5 Survey

This material is based upon work supported by the National Science Foundation under Grant #DUE: 2315434, 2315435. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Allen, R., Folkard, A., Lancaster, G., & Abram, B. (2010). Statistics for the biological and environmental sciences: Improving service teaching for postgraduates. In C. Reading (Ed.), Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics, Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.

GAISE College Report ASA Revision Committee (GAISE Committee). (2016). Curriculum guidelines for undergraduate programs in statistical science. American Statistical Association.

Gal, I. (2003). Expanding conceptions of statistical literacy: An analysis of products from statistics agencies. Statistics education research journal, 2(1), 3-21.

Gal, I. (2004). Statistical Literacy: Meanings, Components, Responsibilities. In The Challenge of Developing Statistical Literacy, Reasoning and Thinking. (pp. 47–78).

Gould, R. (2010). Statistics and the modern student. International Statistical Review, 78(2), 297-315.

Gould, R. (2017). Data literacy is statistical literacy. Statistics Education Research Journal, 16(1), 22-25.

Gelman, A., & Glickman, M. E. (2000). Some class-participation demonstrations for introductory probability and statistics. Journal of Educational and Behavioral Statistics,

Horton, N. J. (2015). Challenges and opportunities for statistics and statistical education: looking back, looking forward. The American Statistician, 69(2), 138-145.

Kaplan, J. J., & Thorpe, J. (2010). Post secondary and adult statistical literacy: Assessing beyond the classroom. In Data and context in statistics education: towards an evidence-based society. Proceedings of the eighth international conference on teaching statistics. Voorburg, The Netherlands: International Statistical Institute.

Pothier, W., & Condon, P. (2023). Cultivating a Data Literate Workforce: Considerations for Librarians. portal: Libraries and the Academy, 23(4), 629-636.25(1), 84-100.

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. Journal of Statistics Education, 16(2).