

Introductory Statistics

Midterm # 1

Learning Objectives covered in this midterm:

Question # 1 LOs:

- Name the population, sample, parameter and statistic in a given study
- Identify the observational units in a statistical study
- Determine and explain bias in a sampling method

Question # 2 LOs:

- Recognize observational studies and list the key components
- Identify the key components of experimental design

Question # 3 LOs:

- Read and interpret side-by-side and stacked bar graphs to compare different groups
- Create a graph that displays key information from quantitative data
- Explain the differences between different graphs that display the same quantitative data

Question # 4 LOs:

- Describe the graph of a data set using its shape, center, spread, and outliers
- Name and compare the measures of center shown in a graph
- Find and use the measures of center to explain the behavior of a data set
- Name the features of the distribution of a data set using statistical language
- Describe the connection between the distribution of a data set and its mean and median

Question # 5 LOs:

- Read information from a boxplot and make conclusions
- Compare boxplots
- Calculate the standard deviation and then describe what that value means

Question # 6 LOs:

- Describe the trend of bivariate data
- Calculate the correlation coefficient and explain what it means
- Find the estimated slope and y-intercept for a linear regression model
- Use the line of best fit to predict values

Question # 7 LOs:

- Find and describe the probability of one event AND/OR and NOT another event
- Find and describe conditional probability

- Describe the meaning of mutually exclusive and independence using probability

Question # 8 LOs:

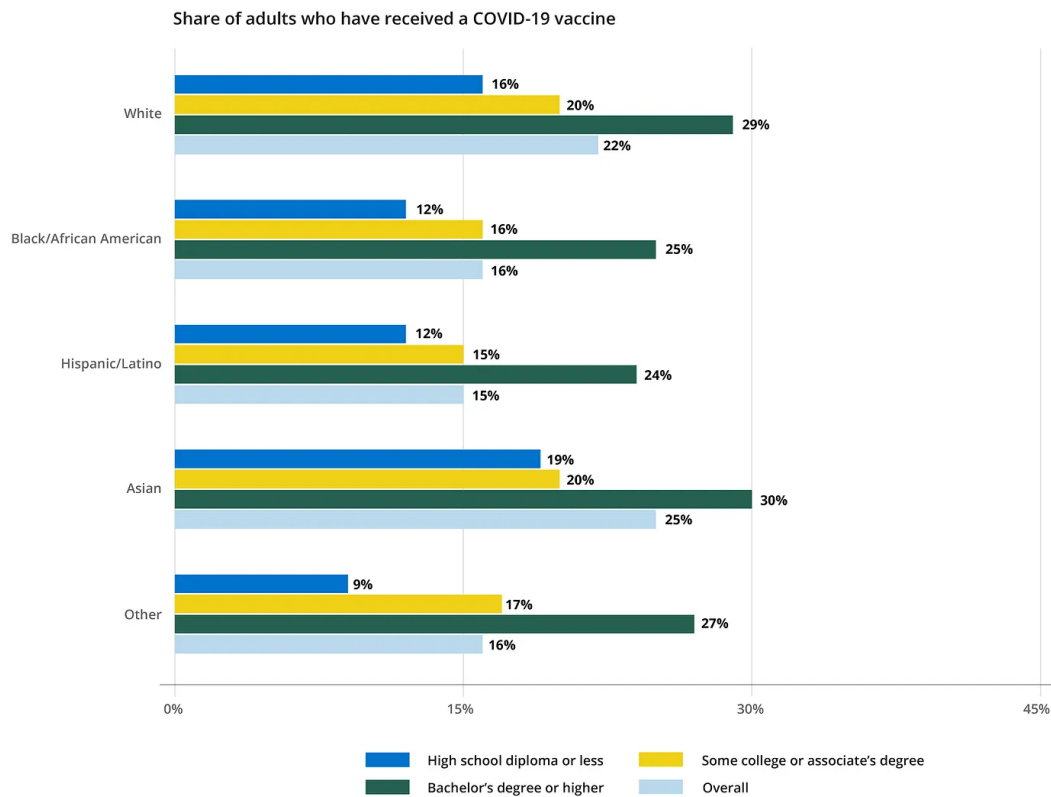
- Calculate z-scores to explain the location of data points
- Compare observations using z-scores and the Empirical Rule
- Calculate percentiles and probabilities for a normal distribution
- Describe the standard normal curve and calculate a z-score

Question # 9 LOs:

- Use a binomial distribution to calculate probability
- Determine if a probability model meets the conditions for a binomial distribution

3. Many Americans are eager to get back out into the world after a year of being cooped up at home because of the pandemic. The development, approval, and administration of vaccines will be major factors in the lifting of restrictions on people's activities. Progress is slowly but steadily occurring on this front: 32 percent of US adults (84 million adults) had received at least one COVID-19 vaccine dose by March 23, 2021. However, the vaccine rollout continues to encounter problems. One is ensuring that underserved communities have fair access to vaccinations, and another is overcoming some groups' entrenched distrust of the vaccines.³

Use the side-by-side bar graph to answer the questions below.



a. Based on the overall percentage, provide the list from the least likely to most likely groups of people to have received a COVID-19 vaccine.

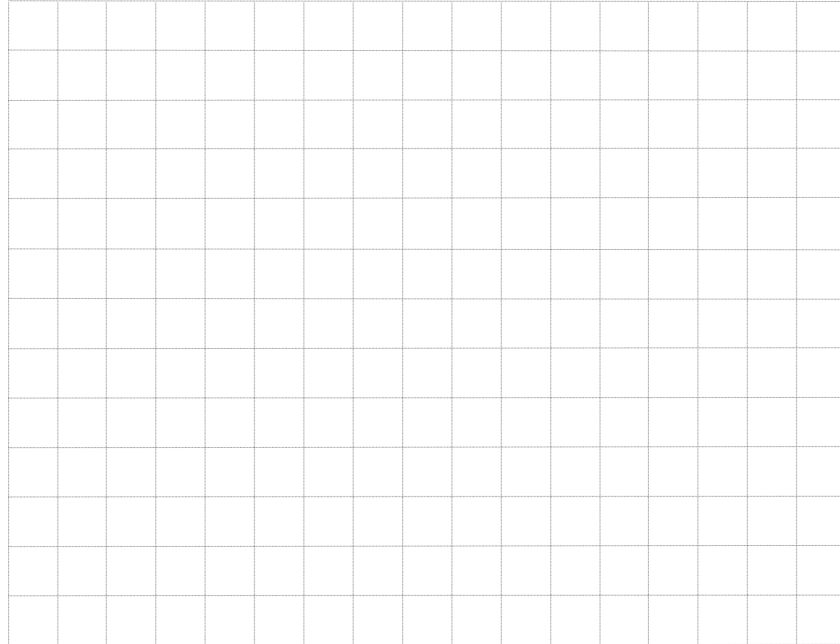
b. Based on the bar graph above, which level(s) of education were more likely to have received a vaccine during the early phases of the distribution?

³

<https://medium.com/georgetown-cew/vaccinations-have-ramped-up-but-has-distribution-been-equitable-ac76d91b6bfd>

c. Based on the bar graph above, for those who have a Bachelor's degree or higher level of education, which group were least likely to have received a vaccine during the early phases of the distribution?

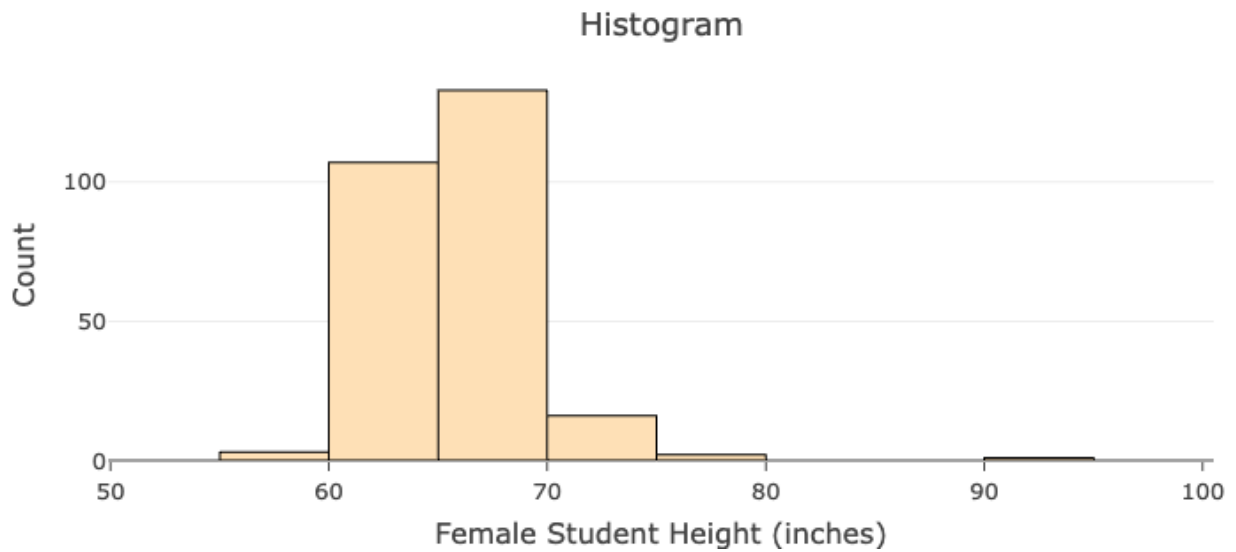
d. Create a stacked bar graph for the data set.



e. What would be the benefit of using the stacked bar graphs rather than the side-by-side bar graph presented above?

f. The title of the article where this graph came from is "Vaccinations Have Ramped Up, But Has Distribution Been Equitable?" What is the answer to the question based on the bar graph provided? Are there additional explanatory/extraneous variables we should consider to answer the question?

4. The height of 262 female students (in inches) are plotted in the histogram below.



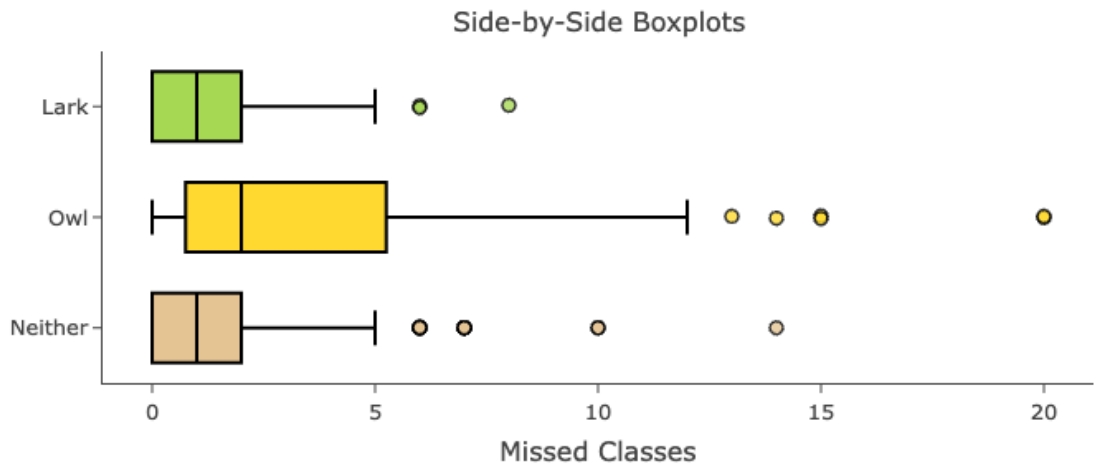
- a. Describe the shape of the distribution of the data set.

- b. What is the approximate mean and median of the data set? In a few sentences, interpret your findings.

- c. Which measure of center represents the data set best? Explain your choice.

- d. Instead of a histogram, you can represent the data set as a dotplot. What would be the advantages and disadvantages of having a dotplot rather than a histogram?

5. The sleep study data set investigated whether college students' chronotypes tend to be larks (morning people) or owls (night people) and measured the number of classes they missed in one semester. The boxplots are displayed below.



- Use the boxplots to describe the difference in variability between the groups.
- The mean and standard deviation of the "Lark" category is 1.56 and 2, respectively. Write a sentence describing the standard deviation in context of the problem.
- List the five-number summary for students who are neither lark nor owl.
- For the "Neither" category, what is the IQR? How many missed classes would be considered a lower outlier and upper outlier?
- For the "Lark" category, _____ % of students missed more than 2 classes.
- For the "Owl" category, _____ % of students missed less than 2 classes.

7. Gallup tracks Americans' views about global warming each March as part of its annual Environment poll. Below are the results of the latest survey conducted on March 1-23, 2023.⁴

Worry: Global Warming	18-34 years old	35-54 years old	55+ years old
A great deal	134	112	145
A fair amount	49	77	88
Only a little	38	60	83
Not at all	53	65	87

- a. If one adult American is chosen at random, what is the probability that he/she/they worry a great deal about global warming?

- b. If one adult American is chosen at random, what is the probability that he/she/they is 35-54 years old?

- c. If one adult American is chosen at random, what is the probability that he/she/they worry a great deal about global warming and are 35-54 years old?

- d. If one adult American is chosen at random, what is the probability that he/she/they worry a great deal about global warming or are 35-54 years old?

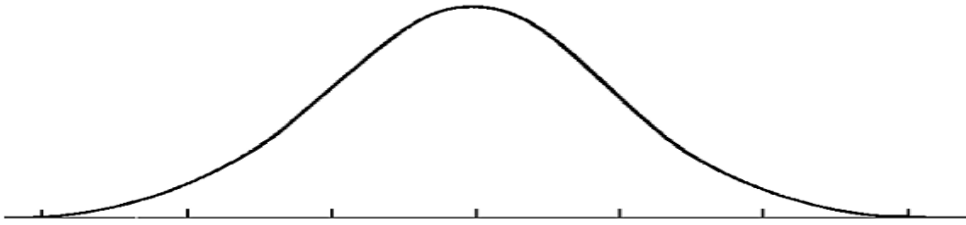
- e. Given that he/she/they is 35-54 years old, what is the probability that he/she/they worry a great deal about global warming?

- f. Is "35-54 years old" and "worry a great deal about global warming" mutually exclusive? Justify your answer.

- g. Is "35-54 years old" and "worry a great deal about global warming" independent? Justify your answer.

⁴ <https://news.gallup.com/poll/474542/steady-six-say-global-warming-effects-begun.aspx>

8. SAT scores are typically normally distributed. In 2021, the distribution of scores in the math section of the SAT had a mean of 528 and a standard deviation 120.⁵
- a. On the following graph, label each tick mark on the x-axis.



- b. Fill-in-the blanks below using the Empirical Rule.

About 68% of the scores are between _____ and _____.

About 95% of the scores are between _____ and _____.

About 99.7% of the scores are between _____ and _____.

- c. Using the empirical rule, answer the following.

$$P(x > 528) = \underline{\hspace{2cm}}$$

$$P(408 < x < 528) = \underline{\hspace{2cm}}$$

$$(408 < x < 768) = \underline{\hspace{2cm}}$$

$$P(x > 888) = \underline{\hspace{2cm}}$$

- d. Calculate and interpret the z-value for the SAT math section score of 600. Then, use the [statistical tool](#) to calculate the probability that an SAT test taker will score above 600.

- e. Use the [statistical tool](#) to answer: What SAT math section score corresponds to the 95th percentile?

⁵ https://nces.ed.gov/programs/digest/d21/tables/dt21_226.40.asp

9. In 2021, 62% of U.S. citizens who were 18 to 29 years old did not have retirement savings.⁶ Suppose that you want to test this theory and decide to pool your introductory statistics class. There are 30 students in the class who are 18 to 29 years old.
- Explain why this is a binomial distribution and not a normal distribution.
 - What are the probability of success (p) and the probability of failure ($1-p$) in this scenario?
 - Use the binomial distribution statistical tool to determine the probability that 15 out of 30 students in the class who are 18 to 29 years old did not have retirement savings.
 - Use the binomial distribution statistical tool to determine the probability more than 20 students in the class who are 18 to 29 years old did not have retirement savings.
 - Would you be surprised to find that more than half of your classmates did not have retirement savings? Justify your answer using probability.
 - Would you be surprised to find that less than a quarter of your classmates did not have retirement savings? Justify your answer using probability.

⁶ <https://www.statista.com/statistics/1273812/adults-with-no-retirement-savings-by-age-us/>