

3.3-3.4: Summary Statistics: Measures of Center and Spread

Group Activity

1. Halloween Data. A data scientist has been counting the number of trick-or-treaters that come to his house every year. We will analyze the shape, center and spread of this data.

Source: <https://www.dataplusscience.com/HalloweenData.html>

a. On the top grid, make a histogram of the number of trick-or-treaters using a bin-width of 100 trick-or-treaters. On the bottom grid you will make a boxplot, but not yet 😊. Follow the steps in order. Label your axes.

Histogram of Trick-or-Treaters

Year	Number of Trick-or-Treaters
2008	492
2009	542
2010	726
2011	869
2012	673
2013	391
2014	454
2015	747
2016	822

Boxplot of Trick-or-Treaters

b. What is the shape of the histogram? If you are not sure yet, compare the mean and the median first.

c. Find the mean, median and mode, include units.

d. Find the 5-number summary, IQR and range, including units.

e. Use the 5-number summary to draw a boxplot on the second grid above. Make your horizontal scale match your histogram scale.

f. Do you think there are any outliers? Why or why not?

Comparing Distributions

2. Below are fictitious student test scores from a Math 105 midterm in two different classes. You will be making a boxplot for each to compare their distributions.

Class 1: 72, 86, 65, 99, 86, 71, 55, 86, 92, 73, 95, 71 points

Class 2: 75, 94, 82, 81, 69, 71, 85, 92, 88, 78, 73, 65, 66 points

a. Find the mean, 5-number summary, IQR and range for each class, including units.

Class 1:

Class 2:

b. Draw the boxplot for each class using the same scale.

Boxplot for Class 1

Boxplot for Class 2

c. What is the shape of the data for each class? How can you tell?

Calculating Standard Deviation, s

d. Using your means rounded to one decimal place, find the standard deviation for each class, including units. The variable n refers to the number of data values.

Class 1:

Mean = _____, n = _____

Test Score (points)	Deviation from the mean	Squared deviation
72		
86		
65		
99		
86		
71		
55		
86		
92		
73		
95		
71		
Sum of the squared deviations (numerator)		

$$s = \sqrt{\frac{\sum (x - \text{mean})^2}{n - 1}}$$

$$= \sqrt{\frac{\text{_____}}{-1}}$$

Class 2:

Mean = _____, n = _____

Test Score (points)	Deviation from the mean	Squared deviation
Sum of the squared deviations (numerator)		

$$s = \sqrt{\frac{\sum (x - \text{mean})^2}{n - 1}}$$

$$= \sqrt{\frac{\quad}{-1}}$$

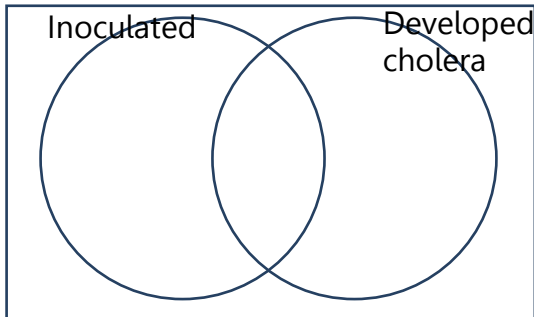
e. Write a few complete sentences summarizing the four characteristics of the distribution of class 1 (shape, center, spread and unusual features).

f. Which class did better on the test? Use the vocabulary and values for center and spread in your answer.

4.1: Contingency Tables and Probability

Cholera Inoculation Study, 1894-96:

A group of 818 people who were exposed to cholera in Calcutta, India in 1894-6 were studied. Of this group, 279 were inoculated with Haffkine's anti-cholera vaccine, while the remaining 539 had not been inoculated. Overall, 69 people developed cholera. Three of those who were vaccinated developed cholera. Use this information to complete the Venn diagram and the contingency table. (Source: <https://mysite.du.edu/~jcalvert/econ/two-by-two.htm>)



	Developed cholera	Did not develop cholera	Total
Inoculated			
Not inoculated			
Total			

1. Find the marginal, "and," and "or" probabilities. If a randomly selected person from the study was chosen, what is the probability they

MARGINAL

a. were inoculated?

b. developed cholera?

AND

c. were inoculated and developed cholera?

d. were not inoculated and did not develop cholera?

OR

e. were inoculated or did not develop cholera?

f. were not inoculated or developed cholera?

CONDITIONAL

g. Given that a person was inoculated, what is the probability they developed cholera?

h. If a person was not inoculated, what is the probability they developed cholera?

i. Do you think the vaccine was effective? Why or why not?

Our Class Data:

The survey data below is from four classes of Math 105 students. Their gender identities and modes of transportation to PCC are summarized in a contingency table.

	Bike	Bus	Drive Self	Ride with Another	Walk	Total
Female	0	12	25	5	3	45
Non-binary or Genderqueer	0	3	2	0	0	5
Male	1	4	16	3	1	25
Total	1	19	43	8	4	75

2. Find the following marginal, "and", and "or" probabilities.

If we were to randomly select a student who took the survey, what is the probability they:

- identify as female?
- identify as non-binary or genderqueer?
- walk to PCC?
- bus to PCC?
- walk and identify as male?
- identify as non-binary or genderqueer and drive them self to PCC?
- identify as female or ride with another?
- identify as male or walk to PCC?

3. Calculate these conditional probabilities:

a. Given that a student from the survey identifies as female, what is the probability they take the bus to PCC?

b. If a student drives themselves to campus, what is the probability they identify as non-binary or genderqueer?

c. What is the probability that a student walks, given they identify as male?

d. Of those students who identify as female, what is the probability they ride to campus with another?