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

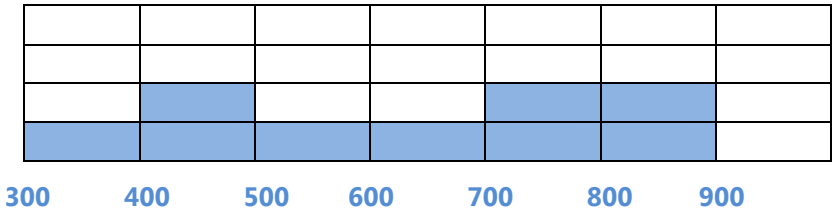
Group Activity

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

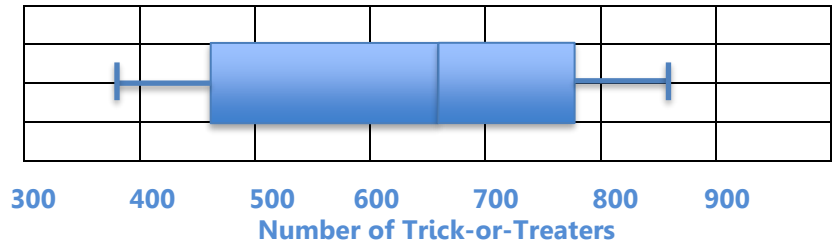
a. On the top grid, make a histogram of the number of trick-or-treaters using a bin-width of 100 trick-or-treaters. On the bottom grid you will make a boxplot, but not yet☺. Follow the steps in order. Label your axes.

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

Histogram of Trick-or-Treaters



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.
The Shape is bimodal and skewed to the left because the mean is less than the median.

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

Put in order: 391, 454, 492, 542, 673, 726, 747, 822, 869
The mean is 635.1 trick-or treaters. The median is 673 trick-or-treaters.
The modes are the centers of the highest peaks at 450 and 800 trick-or-treaters.

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

5-number summary: 391, 473, 673, 784.5, 869 trick-or-treaters
IQR = Q3 – Q1 = 784.5 – 473 = 311.5 trick-or-treaters
Range = 869 – 391 = 478 trick-or-treaters

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? **There do not seem to be any years that are outliers.**

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

[55, 65, 71, 71, 72, 73], [86, 86, 86, 92, 95, 99]

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

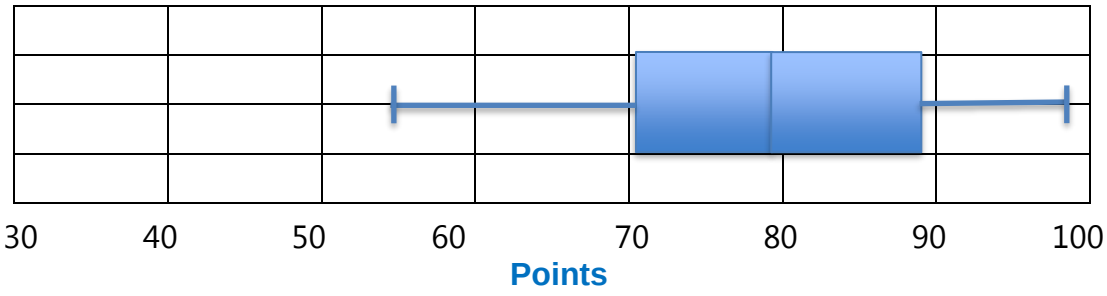
[65, 66, 69, 71, 73, 75], 78, [81, 82, 85, 88, 92, 94]

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

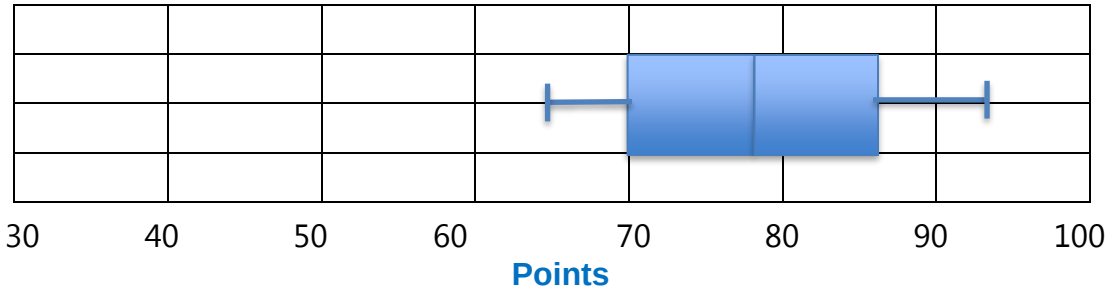
Class 1:	Class 2:
Mean: 79.3 points	**Mean: 78.4 points**
5-Number summary: **(55, 71, 79.5, 89, 99) points**	**(65, 70, 78, 86.5, 94) points**
IQR = Q3-Q1 = 89 – 71 = 18 points	**IQR = Q3-Q1 = 86.5 – 70 = 16.5 points**
Range = Max – Min = 99 – 55 = 44 points	**Max – Min = 94 – 65 = 29 points**

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?

The shapes of both are approximately symmetric. The boxplots are fairly symmetric, and the medians are very close to the means for both classes.

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 = ___**79.3 points**_____, n = ___**12 students**_____

Test Score (points)	Deviation from the mean	Squared deviation
72 **– 79.3**	**– 7.3**	**53.29**
86 **– 79.3**	**6.7**	**44.89**
65 **– 79.3**	**– 14.3**	**204.49**
99 **– 79.3**	**19.7**	**388.09**
86 **– 79.3**	**6.7**	**44.89**
71 **– 79.3**	**– 8.3**	**68.89**
55 **– 79.3**	**– 24.3**	**590.49**
86 **– 79.3**	**6.7**	**44.89**
92 **– 79.3**	**12.7**	**161.29**
73 **– 79.3**	**– 6.3**	**39.69**
95 **– 79.3**	**15.7**	**246.49**
71 **– 79.3**	**– 8.3**	**68.89**
Sum of the squared deviations (numerator)		**1956.28**

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

$$= \sqrt{\frac{1956.28}{12-1}}$$

$$\approx 13.34 \text{ points}$$

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

Mean = ___78.4 points_____, n = ____13 students____

Test Score (points)	Deviation from the mean	Squared deviation
75 – 78.4	– 3.4	11.56
94 – 78.4	15.6	243.36
82 – 78.4	3.6	12.96
81 – 78.4	2.6	6.76
69 – 78.4	– 9.4	88.36
71 – 78.4	– 7.4	54.76
85 – 78.4	6.6	43.56
92 – 78.4	13.6	184.96
88 – 78.4	9.6	92.16
78 – 78.4	– 0.4	0.16
73 – 78.4	– 5.4	29.16
65 – 78.4	– 13.4	179.56
66 – 78.4	– 12.4	153.76
Sum of the squared deviations (numerator)		1101.08

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

$$= \sqrt{\frac{1101.08}{13-1}}$$

$$\approx 9.58 \text{ points}$$

e. Write a few complete sentences summarizing the four characteristics of the distribution of class 1.

The test scores for class 1 are approximately symmetric with a mean of 79.3 points and a standard deviation of 13.35 points. The student who scored 55 points may be an outlier.

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

Class 1 had a slightly higher mean (79.3 vs 78.4 points) and median (79.5 vs 78 points) and students scored higher, but some students also scored lower. It could be said that Class 1 did better.

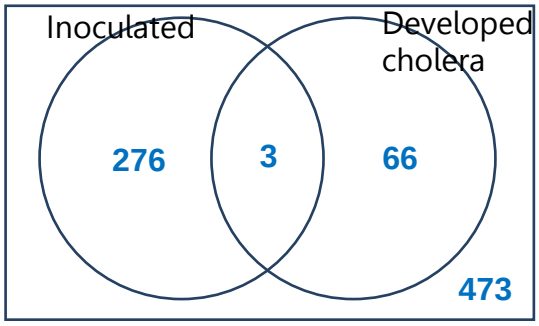
It could also be said that Class 2 did better because they have a smaller range (29 vs 44 points) and standard deviation (9.58 vs 13.34 points), so their scores were less spread out and more consistent. The lowest score in class 2 was 65.

The lowest score in Class 1 was 55 points which may be an outlier due to a student having a bad day, illness or emergency, etc.

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/twobytwo.htm)



	Developed cholera	Did not develop cholera	Total
Inoculated	**3**	**276**	**279**
Not inoculated	**66**	**473**	**539**
Total	**69**	**749**	**818**

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?
P(inoculated) = 279/818 = .3411

b. developed cholera?
P(cholera) = 69/818 = .0844

AND

c. were inoculated and developed cholera?
P(inoculated and cholera) = 3/818 = .0037

d. were not inoculated and did not develop cholera?
P(not inoculated and no cholera) = 473/818 = .5782

OR

e. were inoculated or did not develop cholera?
P(inoculated or no cholera) = (3+276+473)/818 = .9193

f. were not inoculated or developed cholera?
P(not inoculated or cholera) = (3+66+473)/818 = .6626

CONDITIONAL

g. Given that a person was inoculated, what is the probability they developed cholera?
P(cholera given inoculated) = 3/279 = .0108

h. If a person was not inoculated, what is the probability they developed cholera?
P(cholera given not inoculated) = 66/539 = .1224

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

The chance of getting cholera if vaccinated is about 1% versus 12% if not vaccinated. It seems like the vaccine was effective.

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:

a. identify as female?
P(female) = 45/75 = .6

b. identify as non-binary or genderqueer?
P(non-binary or genderqueer) = 5/75 = .0667

c. walk to PCC?
P(walk) = 4/75 = .0533

d. bus to PCC?
P(bus) = 19/75 = .2533

e. walk and identify as male?
P(walk and male) = 1/75 = .0133

f. identify as non-binary or genderqueer and drives them self to PCC?
P(non-binary or genderqueer and drives self) = 2/75 = .0267

g. identify as female or ride with another?
P(female or ride with another) = (12+25+5+3+3)/75 = 48/75 = .64

h. identify as male or walk to PCC?
P(male or walk) = (1+4+16+3+1+3)/75 = 28/75 = .3733

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?

P(bus given female) = 12/45 = .2667

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

P(non-binary or genderqueer given drives self) = 2/43 = .0465

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

P(walk given male) = 1/25 = .04

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

P(ride with another given female) = 5/45 = .1111