Chapter 6 Group Activity

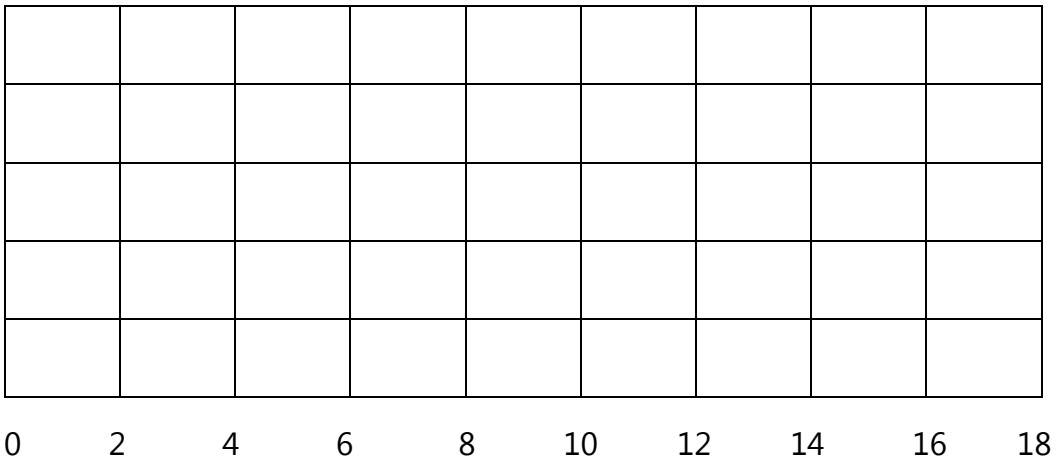
Group Activity

Summarizing a Distribution

1. The following data are the number of credit hours taken by Math 105 students during a summer term. You will be analyzing the shape, center and spread of this data, and whether there are any outliers.

a. On the grid, make a histogram of the data using a bin-width of 2 credits. Give your histogram a descriptive title and label the units on the bottom. Label the vertical axis.

Number of Credit Hours Taken in Summer Term
8
4
16
12
4
7
4
4
6
8
8
4
6
12

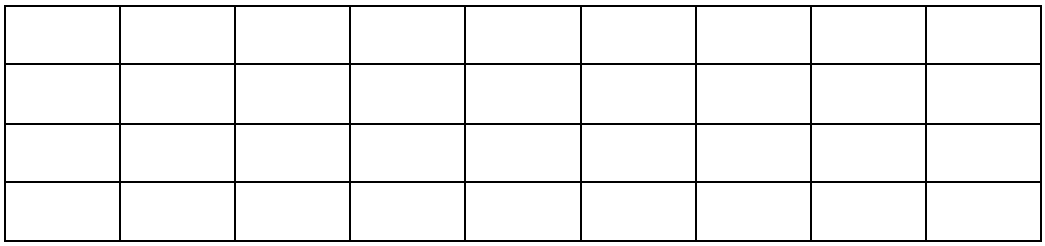


b. What is the shape of the histogram? If you are not sure yet, compute and compare the mean and the median in the next question.

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

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

e. Use the 5-number summary to draw a boxplot on the grid below. Make your horizontal scale match your histogram for comparison. Give your boxplot a descriptive title and label the horizontal axis.



f. How does the shape of the boxplot compare with the shape of the histogram? Can you see the shape of the distribution on the boxplot?

g. Do you think any of the students are outliers? Why or why not?

h. Write a few complete sentences summarizing the four characteristics of this distribution.

Comparing Distributions with Boxplots

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)		

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

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

Class 2:

Mean = _____, n = _____

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

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

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

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

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

More on Shapes (Distributions) of Data

Measurements like heights and weights of people tend to follow a Normal distribution. Measurements of machined parts or test scores on very complex tests, like the SAT's, also follow a Normal distribution.

Variables that have a limit on one side tend to be skewed away from that side. For example, the number of tattoos people have tends to be skewed to the right, because there is a lower limit of 0.

3. Match the description of the data with its most likely shape:

a. The amount of liquid in soda cans (ml)	i. Normally distributed
b. Lengths of newborn babies (in)	ii. Skewed to the right
c. Household incomes in the U.S. (dollars)	iii. Skewed to the left

4. For each quantitative variable in the class survey, write the shape of the distribution. Write n/a for qualitative variables.

a. How do you identify your gender?

b. What is your age?

c. How many credits are you taking this term?

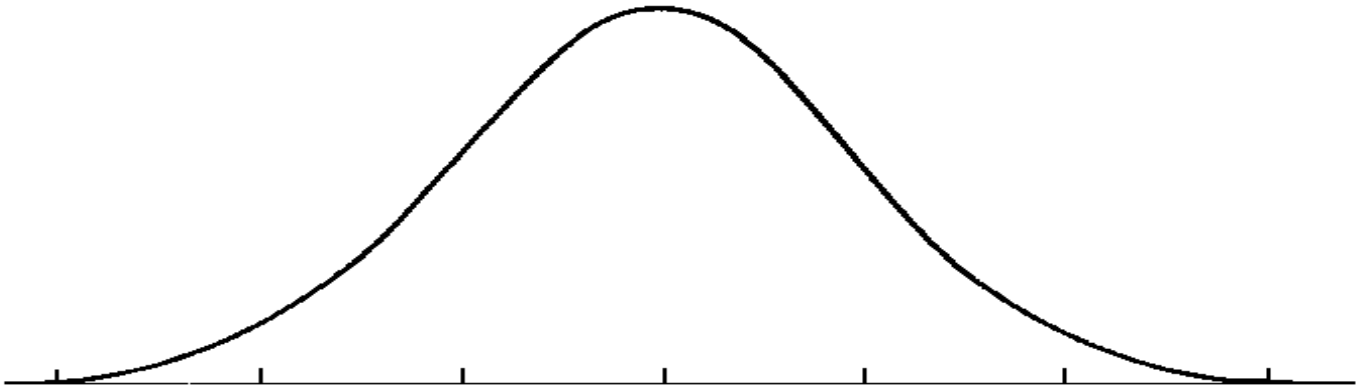
d. What is your main mode of transportation to campus?

e. How long is your pinky finger to the nearest half-centimeter?

The Normal Distribution

Theoretically, pinky measurements should follow a Normal distribution. From several terms of Math 105 students, the mean pinky length is 6 cm with a standard deviation of 0.86 cm.

5. Label the horizontal scale of the Normal model to represent the pinky length of PCC students. Then divide the curve into sections and write the probabilities from the empirical rule in each section.



6. Using the **Empirical rule**, find the percentage of PCC students who would have pinky lengths

a. Less than 5.14 cm

b. Between 6 and 7.72 cm

c. Greater than 6.86 cm

Z-Scores

7a. Calculate the Z-score for a person with a pinky length of 7.72 cm. What does this mean?

b. Calculate the Z-score for a person with a pinky length of 5.57 cm. What does this mean?

Percentiles

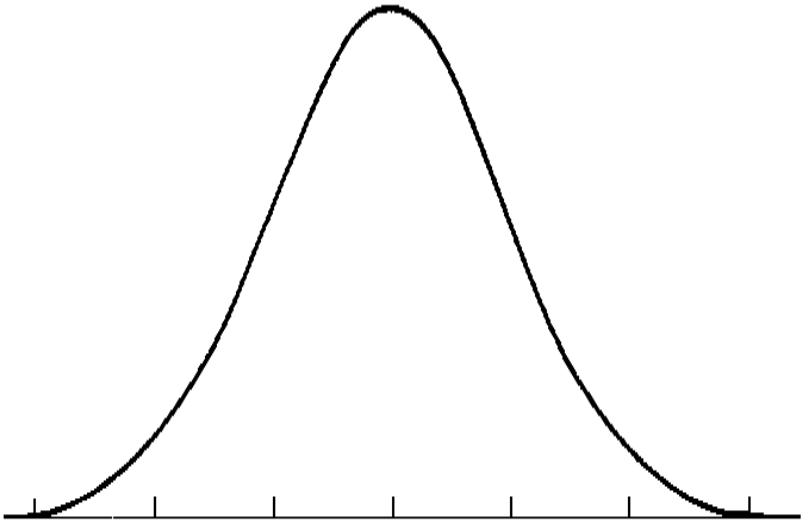
8a. In what percentile is a student with a pinky length of 5.14 cm? What does this mean?

b. What pinky length is the 84th percentile? What does this mean?

More Practice

The thickness of the plastic on an interior car door part follows a Normal distribution with a mean of 3 mm and a standard deviation of 0.5 mm.

9. Label the horizontal scale of the Normal model to represent the pinky length of PCC students. Then divide the curve into sections and write the probabilities from the empirical rule in each section.



10. Using the **Empirical rule**, find the percentage of the parts that have a thickness

a. Between 2.5 and 4 mm

b. Greater than 2mm

c. Less than 3.5 mm

Z-Scores

11a. Calculate the Z-score for a part with a thickness of 3.75 mm. What does that mean?

b. Calculate the Z-score for a part with a thickness of 1.5 mm. What does that mean?

Percentiles

12a. In what percentile is a part with a thickness of 4.5 mm?

b. What thickness is the 34^{th} percentile?