Chapter 6 Group Activity - SOLUTIONS

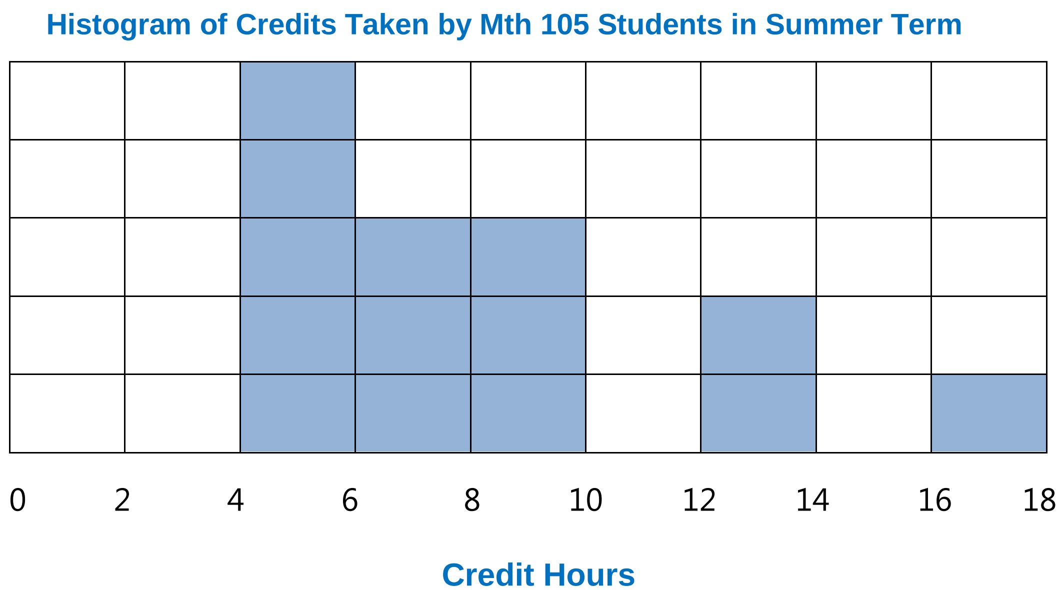
Group Activity

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.

The data are skewed to the right. The mean should be higher than the median.

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

Mean = (8+4+16+12+4+7+4+4+6+8+8+4+6+12)/14 = 103/14 = 7.4 credits.

Median: Put the numbers in order: 4, 4, 4, 4, 4, 6, 6, 7, 8, 8, 8, 12, 12, 16. The average of the middle two, 6 and 7 is 6.5 credits.

Mode: The center of the highest bar is 5 credits; however, no one took 5 credits. I would say the mode is 4 credits because all 5 people in that bar are taking 4 credits.

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

[4, 4, 4, 4, 4, 6, 6] [7, 8, 8, 8, 12, 12, 16]

5-Number summary (Min, Q1, Med, Q3, Max)

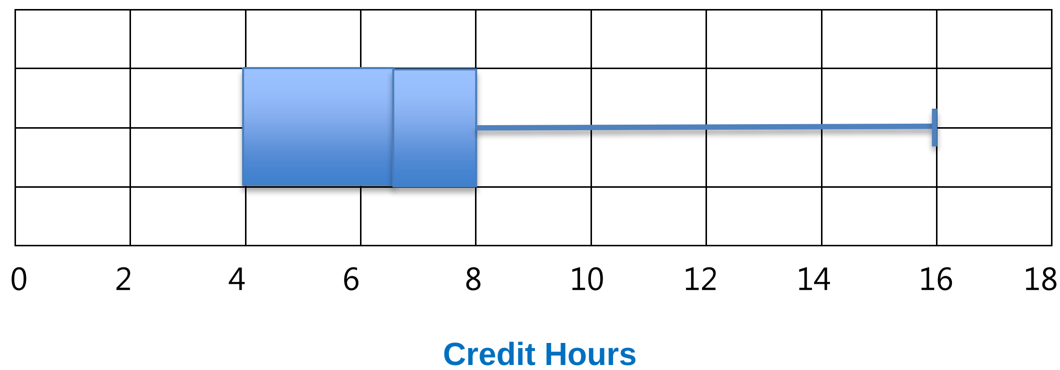
5-Number summary: (4, 4, 6.5, 8, 16) credits

IQR = Q3-Q1 = 8 – 4 = 4 credits

Range = Max – Min = 16 – 4 = 12 credits

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.

Boxplot of Credits Taken by Mth 105 Students in Summer Term



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?

From the boxplot we can see that the scores are skewed to the right because the box and tail are longer on the right side.

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

The student taking 16 credits may be an outlier. We will learn a test for outliers in Math 243. (Fences)

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

The distribution of the number of credits taken by Math 105 students in the summer is unimodal and skewed to the right. The median number of credits taken is 6.5 credits with the middle 50% taking 4-8 credits for an IQR of 4 credits. The student taking 16 credits may be an outlier.

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

[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:

Mean: 79.3 points

5-Number summary:
(55, 71, 79.5, 89, 99) points

IQR = Q3-Q1 = 89 – 71 = 18 points

Range = Max – Min = 99 – 55 = 44 points

Class 2:

Mean: 78.4 points

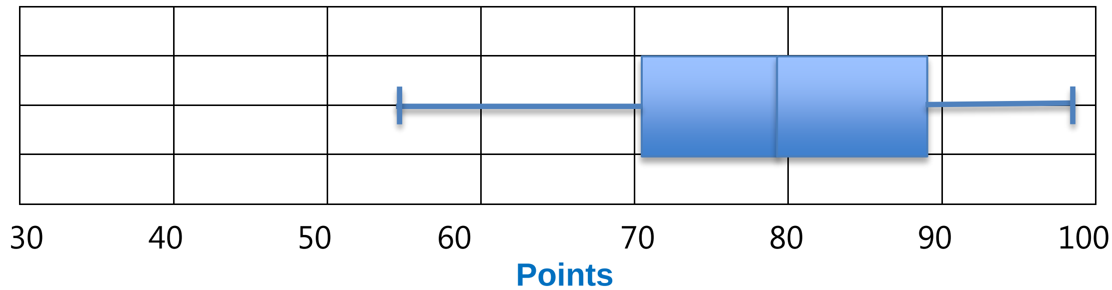
(65, 70, 78, 86.5, 94) points

IQR = Q3-Q1 = 86.5 – 70 = 16.5 points

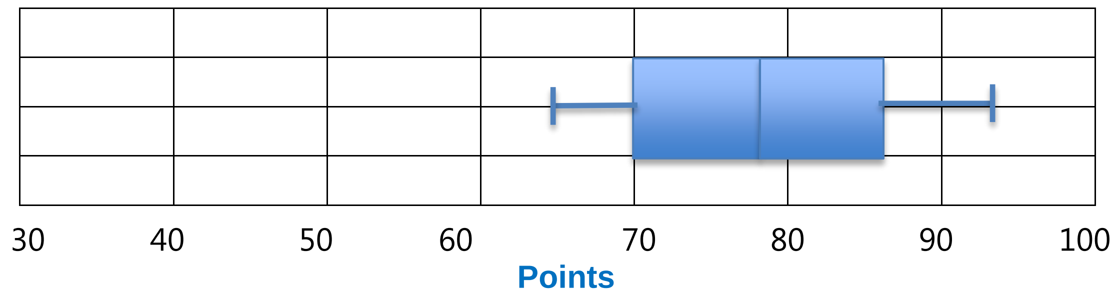
Max – Min = 94 – 65 = 29 points

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.

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	→ iii
b. Lengths of newborn babies (in)	ii. Skewed to the right	→ i
c. Household incomes in the U.S. (dollars)	iii. Skewed to the left	→ ii

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? **n/a, qualitative**

b. What is your age? **Skewed to the right**

c. How many credits are you taking this term? **Skewed to the right**

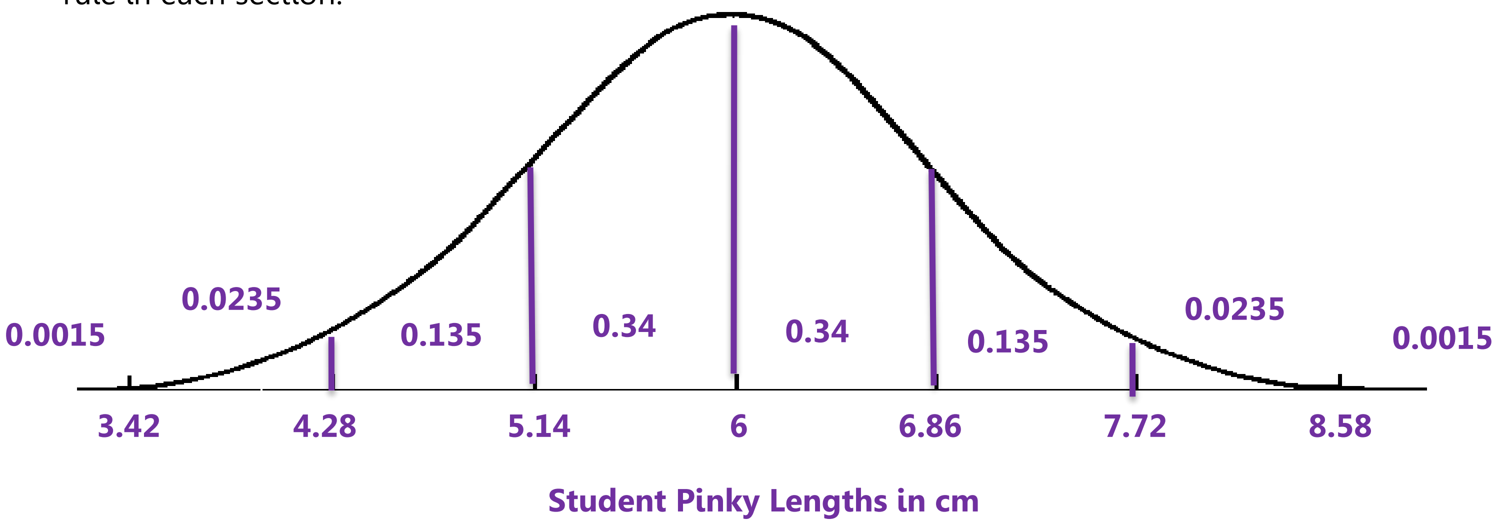
d. What is your main mode of transportation to campus? **n/a qualitative**

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

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 **0.0015 + 0.0235 + 0.135 = 0.16 or 16%**

b. Between 6 and 7.72 cm **0.34 + 0.135 = 0.475 or 47.5%**

c. Greater than 6.86 cm **0.135 + 0.0235 + 0.0015 = 0.16 or 16%**

Z-Scores

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

$Z = \frac{7.72 - 6}{0.86} \approx$ **2 standard deviations. This person is 2 standard deviations above the mean, which is rare.**

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

$Z = \frac{5.57 - 6}{0.86} \approx$ **-0.5 standard deviations. This person is half a standard deviation below the mean.**

Percentiles

8a. In what percentile is a student with a pinky length of 5.14 cm? What does this mean?
0.135 + 0.0235 + 0.0015 = 0.16. This person is in the 16th percentile, which means their pinky length is longer than 16% of the population.

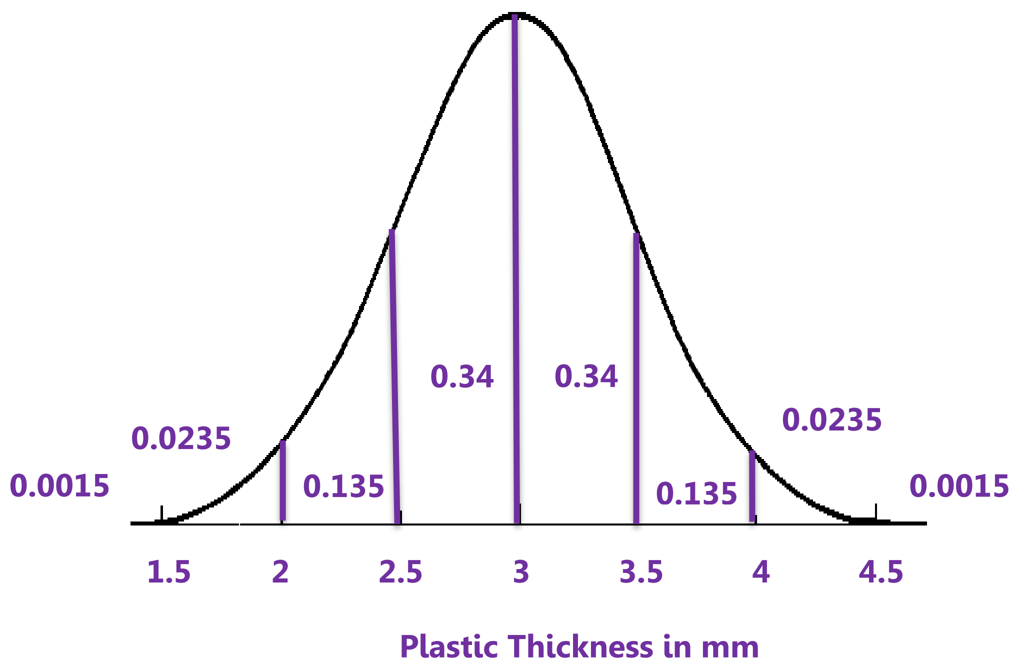
b. What pinky length is the 84th percentile? What does this mean?
0.0015 + 0.0235 + 0.135 + 0.34 + 0.34 = 0.84 or 84%

The 84th percentile corresponds to one standard deviation above the mean, so the 84th percentile is 6.86 cm.

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 **0.34 + 0.34 + 0.135 = 0.815 or 81.5%**

b. Greater than 2mm **0.135 + 0.34 + 0.34 + 0.135 + 0.0235 + 0.0015 = 97.5%**

c. Less than 3.5 mm **0.34 + 0.34 + 0.135 + 0.0235 + 0.0015 = 0.84 or 84%**

Z-Scores

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

$Z = \frac{3.75-3}{0.5} \approx$ **1.5 standard deviations. This part is 1.5 standard deviations above the mean.**

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

$Z = \frac{1.5-3}{0.5} \approx$ **-3 standard deviations. This part is 3 standard deviations below the mean, which is extremely rare.**

Percentiles

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

1 – 0.0015 = 0.9985. (Or add up all the decimals to the left of 4.5.) This part is in the 99.85th percentile, which means the part's is thicker than 99.85% of the parts.

b. What thickness is the 16th percentile?

0.0015 + 0.0235 + 0.135 = 0.16 or 16%

The 16th percentile corresponds to one standard deviation below the mean, so the 16th percentile is a part with thickness of 2.5 mm. That means a part with a thickness of 2.5 mm is thicker than 16% of the parts.