Chapter 2 Modeling Distributions of Data
Lecture Notes & Examples Chapter 2.1 Part A

Learning Targets

- Find and interpret the percentile of an individual value in a distribution of data.
- Estimate percentiles and individual values using a cumulative relative frequency graph.
- Find and interpret the standardized score (z-score) of an individual value in a distribution of data.

Section 2.1 – Describing Location in a Distribution

1. Measuring Position: Percentiles

Definition: ***Percentile***: The percent of values that are less than or equal to a given value

Example: The stemplot below shows the number of wins for each of the 30 Major League Baseball teams in 2009.

5 | 9 → 1
6 | 2455 → 4
7 | 00455589 → 8
8 | 0345667778 → 10
9 | 123557 → 6
10 | 3 → 1

Key: 5 | 9 represents a team with 59 wins.

Find the percentiles for the following teams: (a) The Colorado Rockies, who won 92 games; (b) The New York Yankees, who won 103 games; (c) the Kansas City Royals and the Cleveland Indians, who both won 65 games.

A) Rockies have 25 teams at or below it. $25/30 \approx .83$ 83rd percentile

B) NY Yankees has 30 teams at or below it $30/30 = 1$ 100th percentile.

C) Royals and Indians have 5 teams at or below them. $5/30 \approx .17 \approx$ 17th percentile

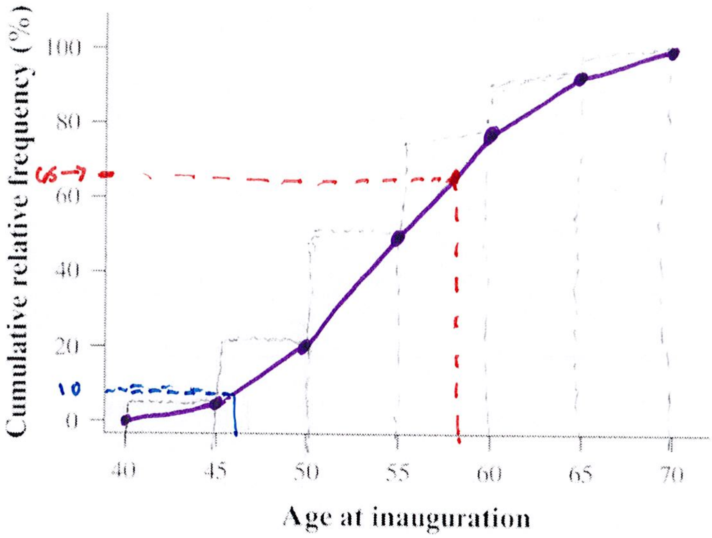
2. Cumulative Relative Frequency Graphs

When you are given a *frequency table* for a quantitative variable, it is possible to graphs that depict the *percentiles*. The table gives the inauguration ages of the first 44 US Presidents.

Age	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
40-44	2	$2/44 \approx 4.5\%$	2	$2/44 \approx 4.5\%$
45-49	7	$7/44 \approx 15.9\%$	9	$9/44 \approx 20.5\%$
50-54	13	$13/44 \approx 29.5\%$	22	$22/44 = 50\%$
55-59	12	$12/44 \approx 27.3\%$	34	$34/44 \approx 77.3\%$
60-64	7	$7/44 \approx 15.9\%$	41	$41/44 \approx 93.2\%$
65-69	3	$3/44 \approx 6.8\%$	44	$44/44 = 100\%$

use cumulative relative
frequency to graph percentiles

To make a cumulative relative frequency graph, we plot a point corresponding to the cumulative relative frequency in each class at the smallest value of the next class. For example, for the 40 to 44 class, we plot a point at a height of 4.5% above the age value of 45. This means that 4.5% of presidents were inaugurated before they were 45 years old. (In other words, age 45 is the 4.5th percentile of the inauguration age distribution.) It is customary to start a cumulative relative frequency graph with a point at a height of 0% at the smallest value of the first class (in this case, 40). The last point we plot should be at a height of 100%. We connect consecutive points with a line segment to form the graph. The figure below shows the completed cumulative relative frequency graph.



Cumulative Relative Frequency Graph - Used to examine location within a distribution. Cumulative relative frequency graphs begin by grouping the observations into equal-width classes. The completed graph shows the accumulating percent of observations as you move through the classes in increasing order.

Interpreting Cumulative Relative Frequency graphs

Using the graph above, answer the following:

A) Was Barack Obama, who was inaugurated at age 47, unusually young?

Yes, around 10th percentile. About 10 % of US Presidents were Obama's age or younger when inaugurated.

B) Estimate and interpret the 65th percentile of the distribution.

The 65th percentile is around the age of 57. 65% of presidents were 57 years old or younger at time of inaugeration.

CHECK YOUR UNDERSTANDING

1. Multiple choice: Select the best answer. Mark receives a score report detailing his performance on a statewide test. On the math section, Mark earned a raw score of 39, which placed him at the 68th percentile. This means that

(a) Mark did better than about 39% of the students who took the test.
(b) Mark did worse than about 39% of the students who took the test.
(c) Mark did better than about 68% of the students who took the test.
(d) Mark did worse than about 68% of the students who took the test.
(e) Mark got fewer than half of the questions correct on this test.

2. Mrs. Munson is concerned about how her daughter's height and weight compare with those of other girls of the same age. She uses an online calculator to determine that her daughter is at the 87th percentile for weight and the 67th percentile for height. Explain to Mrs. Munson what this means.

87% of girls of the same age have a weight that's less than or equal to her daughter's weight. 67% of the girls of the same age have a height that is less than or equal to her daughter's height.

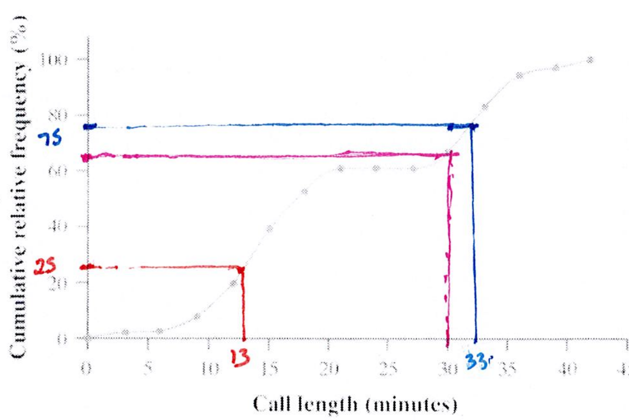
Questions 3 and 4 relate to the following setting. The graph displays the cumulative relative frequency of the lengths of phone calls made from the mathematics department office at Gabalot High last month.

3. About what percent of calls lasted less than 30 minutes? 30 minutes or more?

About 65% of calls are less than or equal to 30 minutes.

About 35% of calls are more than 30 minutes

Percentile



4. Estimate Q1, Q3, and the IQR of the distribution.

Q_1 = 25th percentile $\quad Q_1 \approx 13$ min

Q_3 = 75th percentile $\quad Q_3 \approx 33$ min

IQR = $Q_3 - Q_1 = 33 - 13 = 19$ minutes

3. Measuring Position: Z-Scores

Another way of *measuring position* is to determine how many *standard deviations* above or below the mean an individual data point is. This is called computing a ***z-score***. This process is known as ***standardizing.***

Definition - **Standardized value (z-score)**:

aka z-score

If x is an observation from a distribution that has a known mean and standard deviation, the **standardized value** of x is

value

$$Z = \frac{\text{value} - \text{mean}}{\text{standard deviation}}$$

value

This measure tells how many standard deviations the given data point is from the mean.

Example: 2009 MLB Wins (revisited)

5 | 9 6 | 2455 7 | 00455589 8 | 0345667778 9 | 123557 10| 3 Key: 5 | 9 represents a team with 59 wins.	Mean: 81 Median: 83.5 StDev: 11.43 Minimum: 59 Maximum: 103 Q1: 74 Q3: 88

Use the information provided to find the standardized scores for the (a) Boston Red Sox with 95 wins; (b) Atlanta Braves with 86 wins; and (c) Washington Nationals with 59 wins.

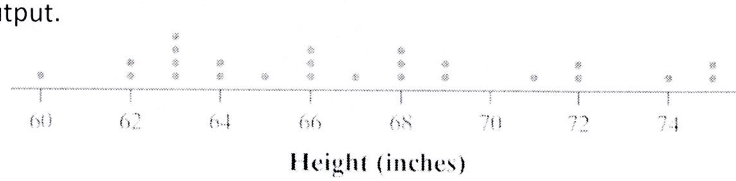
A) Red Sox 95 wins $Z_{95} = \frac{95-81}{11.43} = \frac{14}{11.43} \approx 1.22$

B) Braves $Z_{86} = \frac{86-81}{11.43} = \frac{5}{11.43} = .44$ ← closest to the mean

C) Nationals $Z_{59} = \frac{59-81}{11.43} = \frac{-22}{11.43} = -1.92$ ← below the mean

CHECK YOUR UNDERSTANDING

Mrs. Navard's statistics class has recorded their heights. The figure below shows a dotplot of the class's height distribution, along with summary statistics from computer output.



Variable	n	$\bar{x}$	s_x	Min	Q_1	M	Q_3	Max
Height	25	67	4.29	60	63	66	69	75

1. Lynette, a student in the class, is 65 inches tall. Find and interpret her z-score.

$Z_{65} = \frac{65-67}{4.29} = \frac{-2}{4.29} = -0.47$

Lynette's height is .47 standard deviations below the mean.

2. Another student in the class, Brent, is 74 inches tall. How tall is Brent compared with the rest of the class? Give appropriate numerical evidence to support your answer.

$Z_{74} = \frac{74-67}{4.29} = \frac{7}{4.29} = 1.63$

Brent's height is 1.63 standard deviations above the mean.

3. Brent is a member of the school's basketball team. The mean height of the players on the team is 76 inches. Brent's height translates to a z-score of –0.85 in the team's height distribution. What is the standard deviation of the team members' heights?

$Z = \frac{\text{value} - \text{mean}}{SD}$ $-0.85 = \frac{74-76}{SD}$ $-0.85 = \frac{-2}{SD}$ $SD(-0.85) = -2$

$SD = \frac{-2}{-0.85} \approx 2.35$

Summarize Big Ideas:

Learning Target #1:

Learning Target #2:

Learning Target #3: