Modeling Data Distributions: Part 1

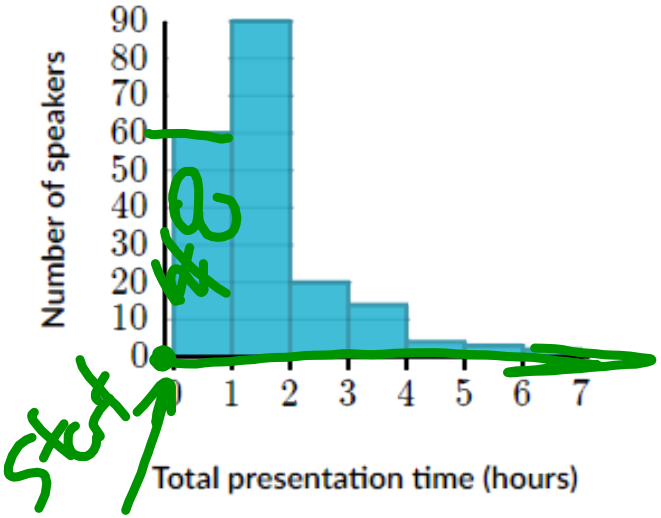
- Topics: Percentiles, Z-scores, and Area

- Objective: Students will be able to interpret data percentiles, calculate z-scores, and find understand density curves with relation to area under a graph.

- Standards: AP Stats: UNC-1 (EU), UNC-1.I (LO), UNC-1.I.5 (EK)

Calculating Percentiles

- You may need to determine what percentile a particular datum is in a complete set of data.
- Example 1: At a convention, 192 speakers gave one or more presentations of varying lengths. The histogram summarizes the total presentation time, in hours, of each speaker.



192 * .25

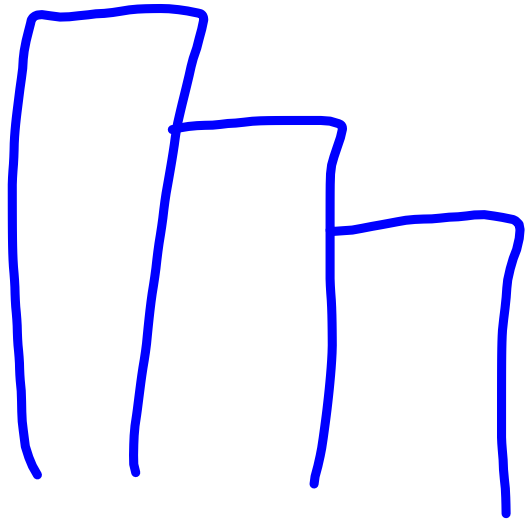
convert to decimal

0-1

- What interval contains the 25th percentile for this data?

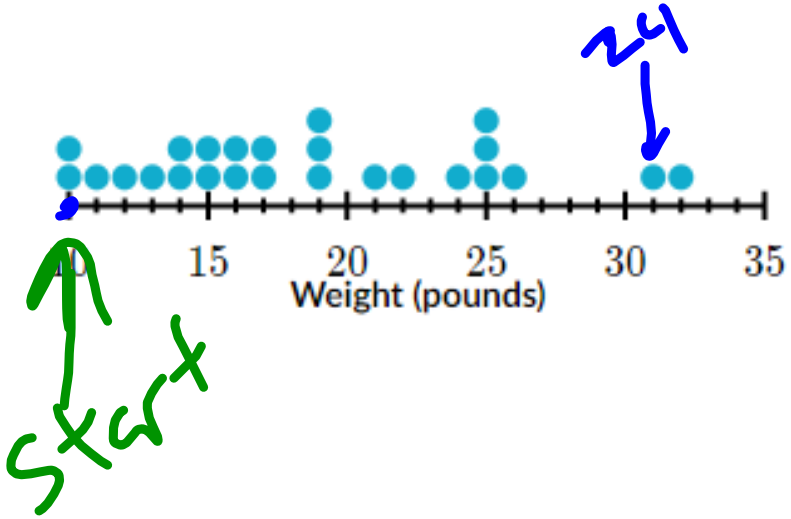
433 members

433 x .30 = 130



Calculating Percentiles

- You may need to determine what percentile a particular datum is in a complete set of data.
- Example 2: The dot plot shows the weight, in pounds, of 25 students' filled backpacks. Each dot represents one backpack.



$\frac{24}{25} = .96$

- Which of the following is the closest estimate to the percentile for the backpack that weighs 31 pounds?

96%

Calculating Z-scores

- What are z-scores?
  - > A z-score measures exactly how many standard deviations above or below the mean a data point is.
  - > Formula for calculating z-scores:

Your data point

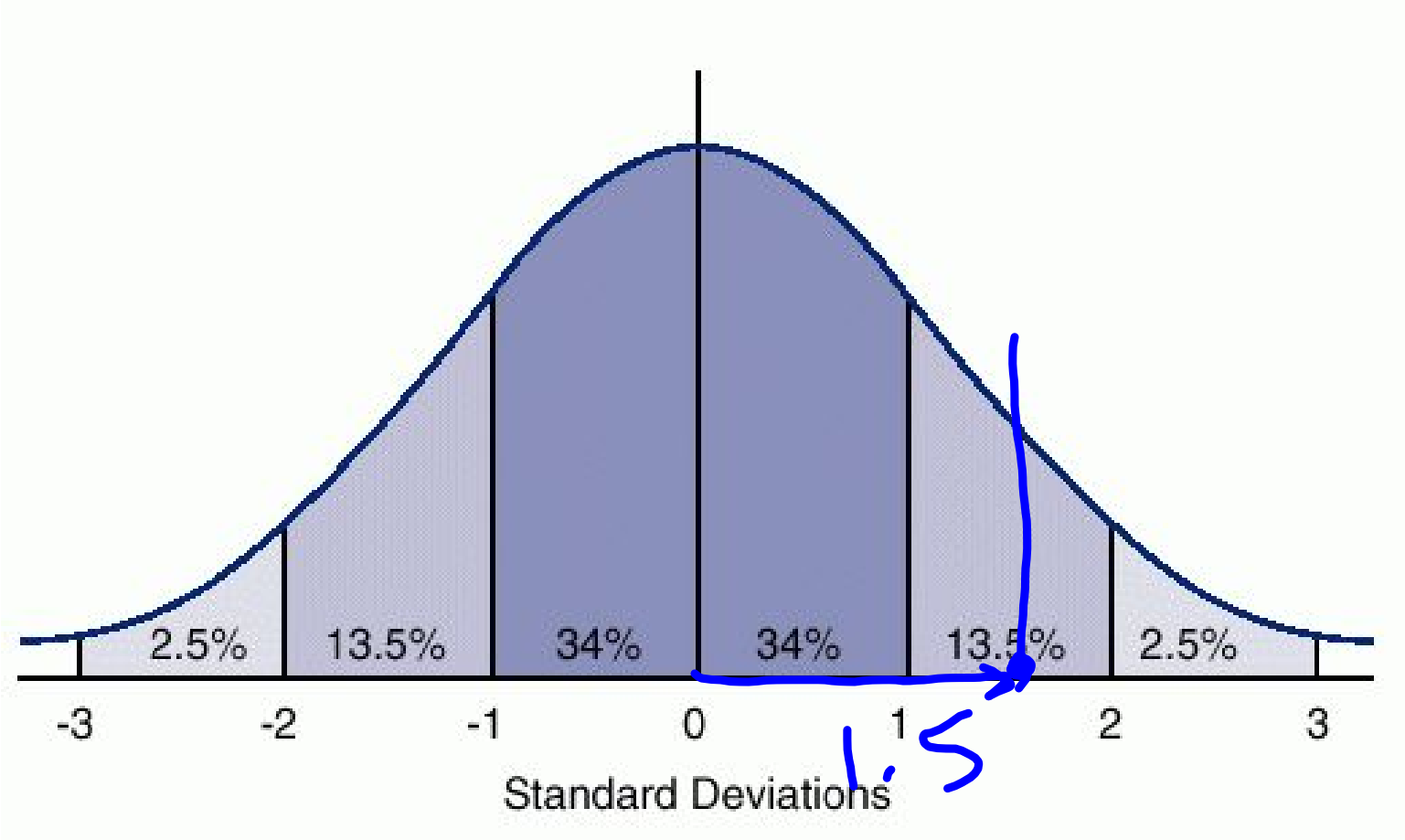
$$z = \frac{\text{data point} - \text{mean}}{\text{standard deviation}}$$

$$z = \frac{x - \mu}{\sigma}$$

mu: mean median

z score

sigma (lowercase) Standard Deviation


2.5%
13.5%
34%
34%
13.5%
2.5%
-3
-2
-1
0
1
2
3
Standard Deviations
1.5

Calculating Z-scores

- Example 1: The grades on a history midterm at Almond have a mean of μ=85 and a standard deviation of σ=2.
- Michael scored 86 on the exam.
- Find the z-score for Michael's exam grade.

$$z = \frac{\text{data point} - \text{mean}}{\text{standard deviation}}$$

$$z = \frac{x - \mu}{\sigma}$$

$$\frac{86-85}{2}$$

$$\frac{1}{2} = .5$$

Z-Score = .5000

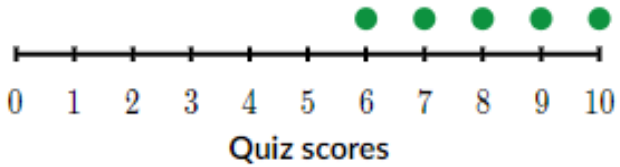
Transforming Data

- It is very common to take data and apply the same transformation to every data point in the set.

- For example, we may take a set of temperatures taken in degrees Fahrenheit and convert them all to degrees Celsius.
- How would this conversion impact the measures of center of spread in the data set? Let's look at a simpler example to think about this situation.

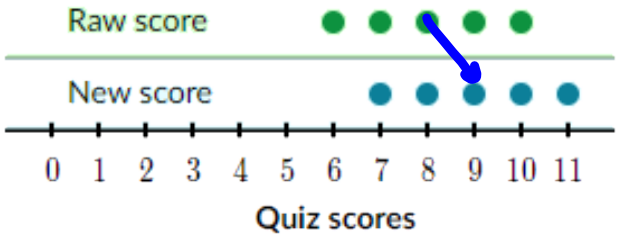
Transforming Data

- Example 1: Adding a constant
  - > Five friends took a 10 question multiple choice quiz in class. Their raw scores on the quiz are shown in the dotplot below along with summary statistics.



	$\bar{x}$	s_x	M	IQR	range
Scores	8	1.41	8	3	4

  - > The teacher told everyone that she would add 1 to every student's score as extra credit. Their new scores are shown below.

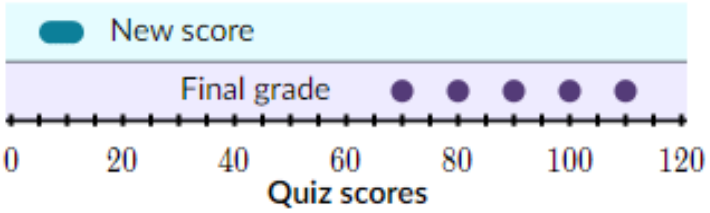


	$\bar{x}$	s_x	M	IQR	range
Raw scores	8	1.41	8	3	4
New scores	9	1.41	9	3	4

  - > Find the summary statistics for the new scores.

Transforming Data

- Example 1: Multiplying a constant
  - > Five friends took a 10 question multiple choice quiz in class. Their raw scores on the quiz are shown in the dotplot below along with summary statistics.
  - > The teacher always scores her quizzes out of 100 points. For this 10-question quiz, she multiplies the new scores by 10 to get the students' final grades which are shown in the dotplot below.



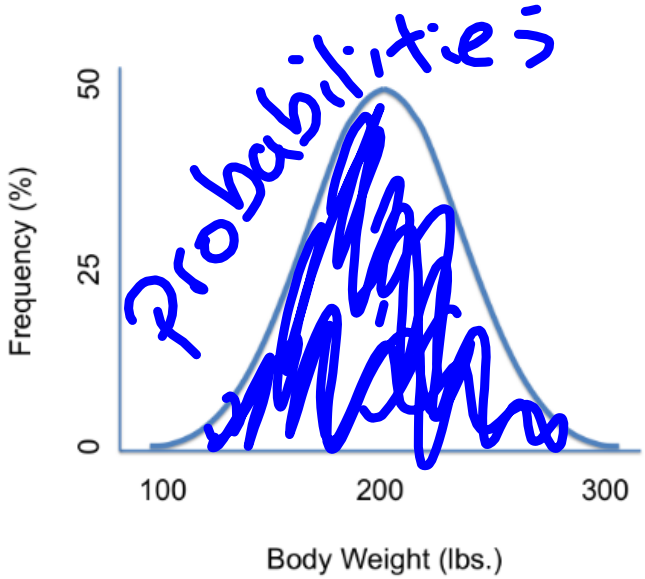
	$\bar{x}$	s_x	M	IQR	range
Scores	8	1.41	8	3	4
New scores	9	1.41	9	3	4
Final grades	?	?	?	?	?

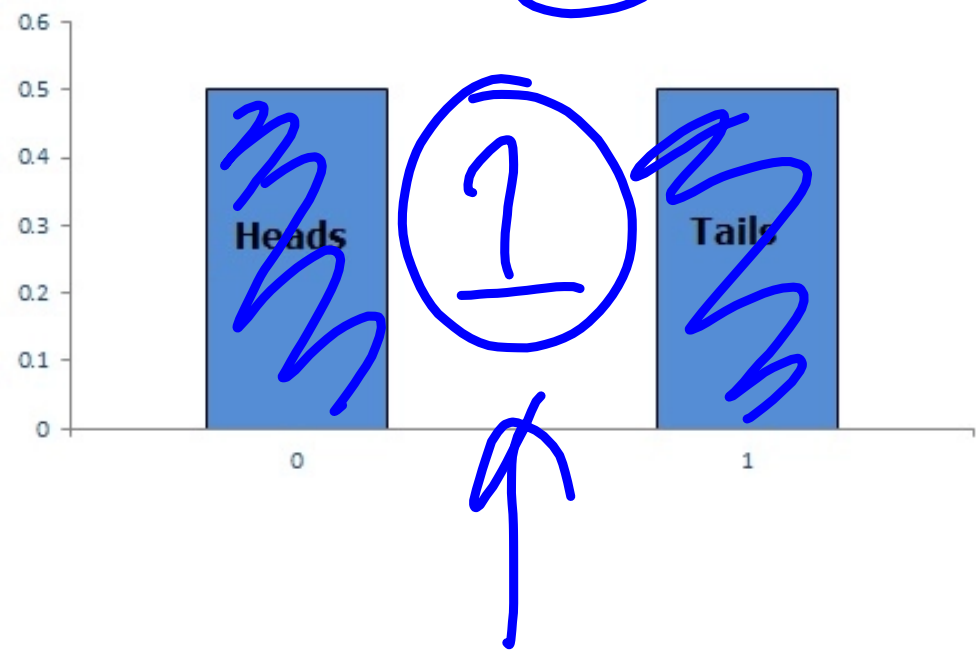
90 14.1 90 30 40

  - > Find the mean and median of the final grades.

Density Curves

- What is a density curve?
  - > A density curve is a graph that shows probability. The area under the density curve is equal to 100 percent of all probabilities. As we usually use decimals in probabilities you can also say that the area is equal to 1 (because 100% as a decimal is 1).



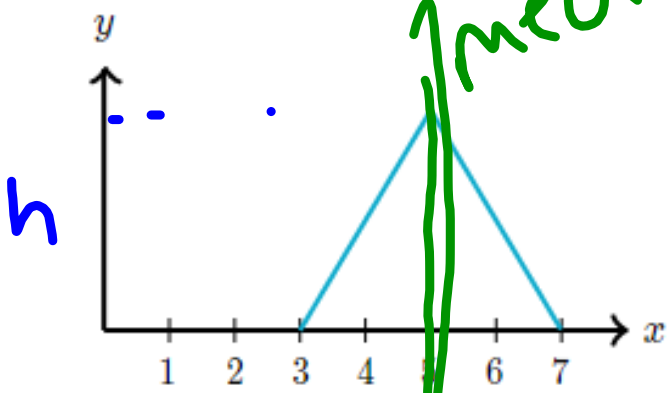


Density Curves

- Formulas you may need:
  - > Area of Rectangle: $a = l*w$
  - > Area of a Triangle: $a= 1/2\ b*h$
  - > Area of a Trapezoid: $a = (b_1 + b_2)/2 * h$

Density Curves

- Example 1: The density curve below represents the number of minutes Joseph spent getting dressed for school each morning this past school year.



median

h

$A = \frac{1}{2}bh$

$1 = \frac{1}{2}(4)x$

$1 = 2x$

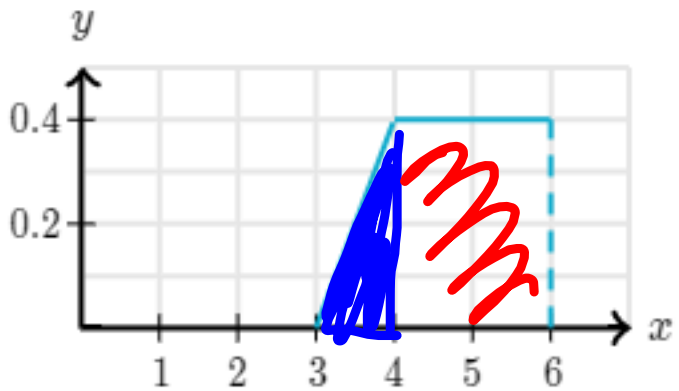
$\frac{1}{2} = \frac{2x}{2}$

$h = \frac{1}{2}$

- Which of the following statements are true?
- Choose all answers that apply:
  1. The distribution of time spent getting dressed is symmetric. ✓
  2. The median time spent getting dressed is 5 minutes. ✓
  3. The height of the density curve is 1/5. ✗

Density Curves

- Area Under Density Curves
- Example 1: Consider the density curve below.
- Find the percent of the area under the density curve where x is more than 3.



$x < 4$

$a = \frac{1}{2}bh$

$= \frac{1}{2}(1)(.4)$

$a = .2$

2

$a > 4$

$a = bh$

$= 2(.4)$

$a = .8$

Displaying and Comparing Quantitative Data

You should be working on the following skills:

1. Calculating percentiles
2. Calculating z-scores
3. Transforming data
4. Properties of density curves
5. Area under density curves