

# Teaching a Prestatistics Course: Propelling Non-STEM Students Forward

Jay Lehmann

College of San Mateo

MathNerdJay@aol.com

[www.pearsonhighered.com/lehmannseries](http://www.pearsonhighered.com/lehmannseries)

## Learning Is in the Details

Detailing concepts at a microscopic level affords quantum leaps in learning.



# Outline

- 1 Motivation for Prestatistics Course
- 2 Structure of Course
- 3 Prestatistics Content
- 4 Present Concepts in Detail
- 5 In-Class Activities
- 6 Projects
- 7 Cool Data Sets and Drawing Comparisons
- 8 Promoting the Course

## Non-Accelerated Students Through Dev Math

Ten-year study:

- Elementary algebra: 53.1% success rate
- Persistence into intermediate algebra: 65.8%
- Intermediate algebra: 56.4% success rate



**Complete path: 19.7%**

## Non-Accelerated Students Through Dev Math

Ten-year study for **split courses**:

- First half of elementary: 58%
- Persistence into second half of elementary: 90%
- Second half of elementary: 53%
- Persistence into first half of intermediate: 65.8%
- First half of Intermediate: 61%
- Persistence into second half of intermediate: 90%
- Second half of intermediate: 61%



**Complete path: 6.1%**

# Course Structure

- 6-unit course
- some lecture, some group work
- 2 hours on Tuesdays, 1 hour other weekdays
- Ti-84
- online homework
- 5 tests, 10 quizzes, 1 project, 1 final



Chris Joseph Taylor

## Who Can Take the Course?

- Non-STEM, non-business, non-architecture, non-etc majors
- Prerequisite: Arithmetic

## Prestatistics Course Content

In California, we can offer courses equivalent to intermediate algebra.

Course contains:

- Quick arithmetic review that foreshadows statistics
- Descriptive statistics
- Algebra required for elementary statistics
- Additional algebra not absolutely necessary but closely tied to elementary statistics



# Prestatistics More Detailed Course Content

- Arithmetic review
- Statistical diagrams
- Measures of center and spread
- Probability
  - ▶ Probability laws
  - ▶ Normal distribution
- Linear regression
  - ▶ Algebra
  - ▶ Sum of squared residuals
  - ▶ Linear regression equation
  - ▶ Residual plots
  - ▶ Interpret slope and vertical intercept
- Exponential regression
  - ▶ Algebra
  - ▶ Similar to linear regression concepts

## Learn From Our Mistakes

What did we learn from math reform in the 90s?



## What the Course Should Not Be

Acceleration should not mean . . .

- Deleting topics because they are challenging.
- Dumbing-down the curriculum.
- Duplicating the first half of a statistics course.



Avoid the 3 Ds!

## Algebra tied to Linear Regression

- Graph linear equations
- Interpret slope and vertical intercept
- Simplify linear expressions
- Solve linear equations
- Solve linear inequalities
- Formulas
- Functions
- Use two points to find an equation of a line

## Goal of Course

Introduce crucial concepts at an arithmetic level and propel students' understanding beyond what's attained in statistics.



# Implementation

But How?



## Learning Is in the Details

By detailing concepts at a much higher level than in statistics.



## Goal of Course

For example, today we will detail the following concepts and draw comparisons between them:

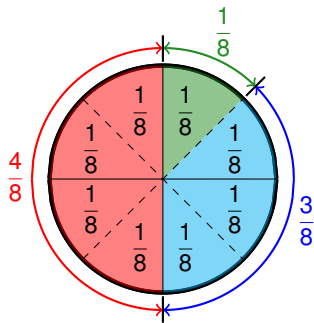
- Proportions
- Density histograms
- Probability laws
- Normal curve





## Groundwork for Probability: Proportion

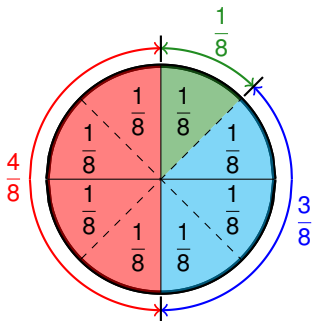
A **proportion** is a fraction of the whole.



## Properties of Proportions

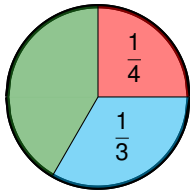
- A proportion is always between 0 and 1, inclusive.
- If an object is made up two or more parts, the sum of the proportions equals 1.
- Let  $\frac{a}{b}$  be the proportion of the whole that has a certain characteristic. Then the proportion of the whole that does NOT have that characteristic is

$$1 - \frac{a}{b}$$



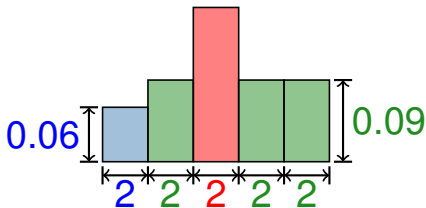
## Problem Solving with Proportions

Find the proportion of the disk that is green.



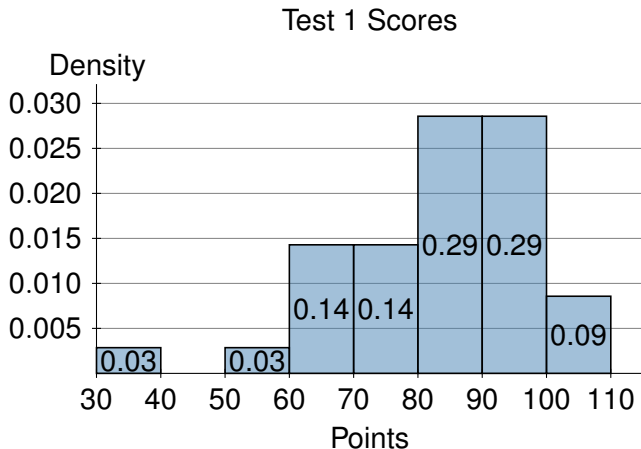
## Problem Solving with Proportions

Use the fact that the total area of the object is 1 to find the area of the red bar. The object has not been drawn to scale.

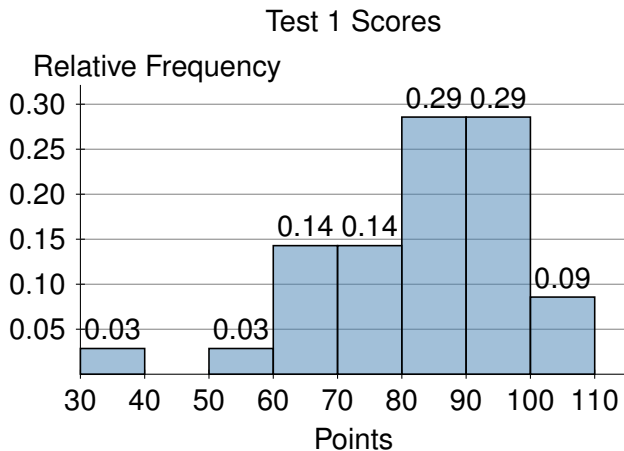


## Definition of Density Histogram

For a *density histogram*, the vertical axis has units called *density* so that the area of each bar is the relative frequency of the bar's class.



# Compare with Relative Frequency Histogram

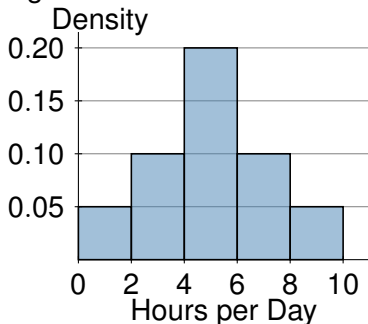


# Television Viewing Durations



## Total Area of Bars of Density Histogram

Television Viewing Durations in the Summer by College Students



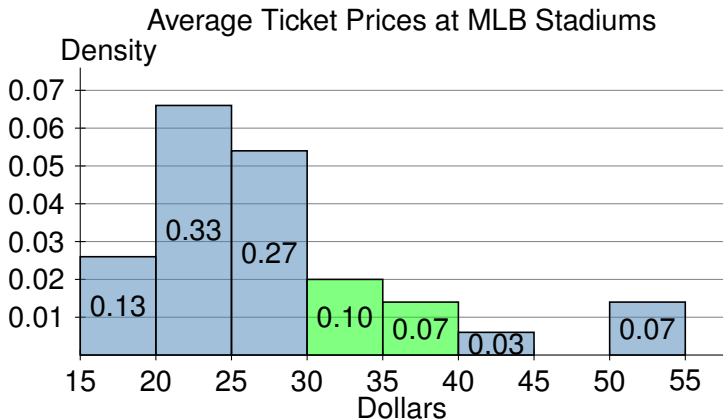
- 1 Compute the area of each of the five bars in the density histogram.
- 2 Find the total area of the five bars.
- 3 Explain why the total area of the bars of *any* density histogram is always equal to your result in part (b).



## Average Ticket Prices at MLB Stadiums

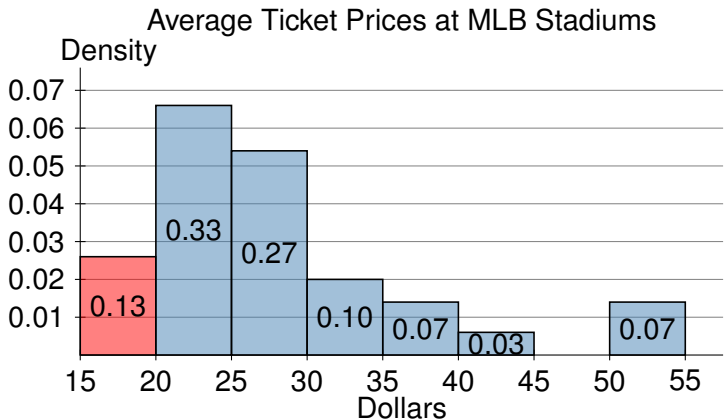


## Density Histogram and Proportion



Find the proportion of stadiums whose average price of 2014 MLB tickets are between \$30 and \$40.

# Density Histogram and Law of the Complement



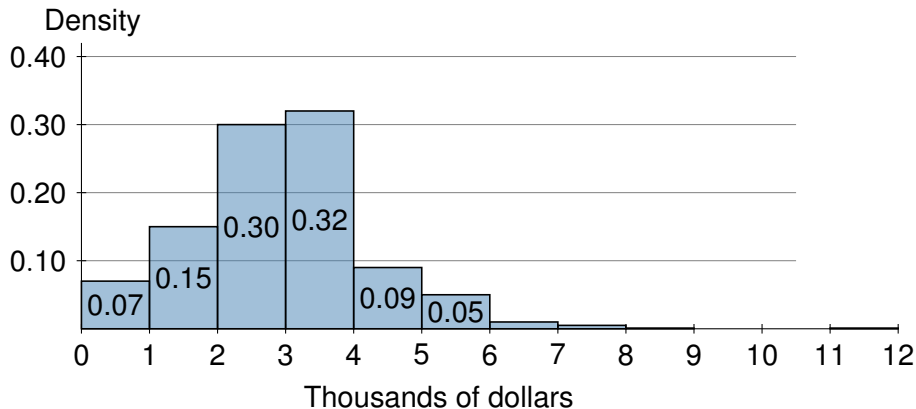
Find the proportion of stadiums whose average price of 2014 MLB tickets are at least \$20.

# Tuitions of 2-Year, Public Colleges



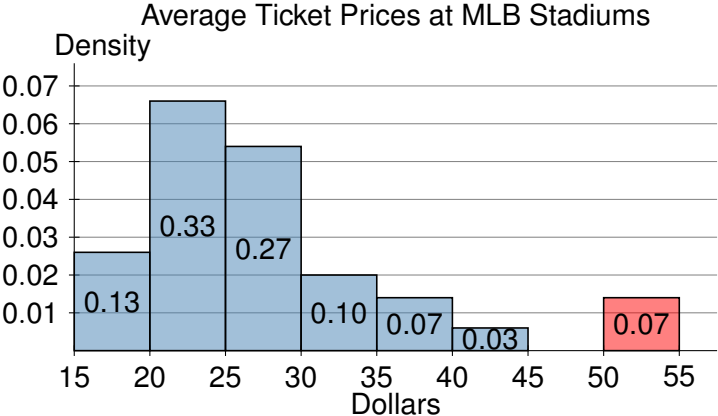
# Density Histogram and Law of the Complement

Tuitions of 2-Year, Public Colleges



What proportion of the colleges had tuitions greater than \$6 thousand?

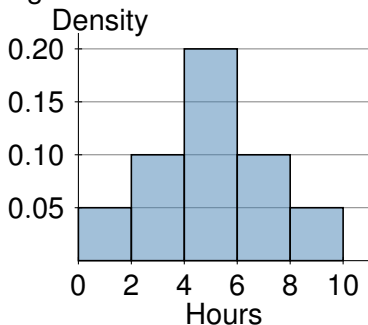
# Density Histogram and Percentile



Estimate the average ticket price at the 93rd percentile.

## Total Area of Bars of Density Histogram

Television Viewing Durations in the Summer by College Students



Once the fall semester begins, the mean viewing time decreases by 2 hours and the standard deviation decreases. Sketch a density histogram that might describe the TV viewing distribution during the fall semester.

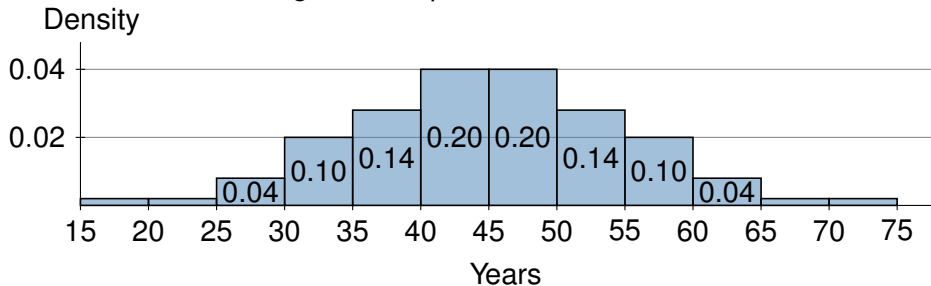
## Ages of People Who Attend a Movie





# Density Histogram and the Empirical Rule

Ages of People Who Attend a Movie



1 Estimate  $\bar{x}$ .

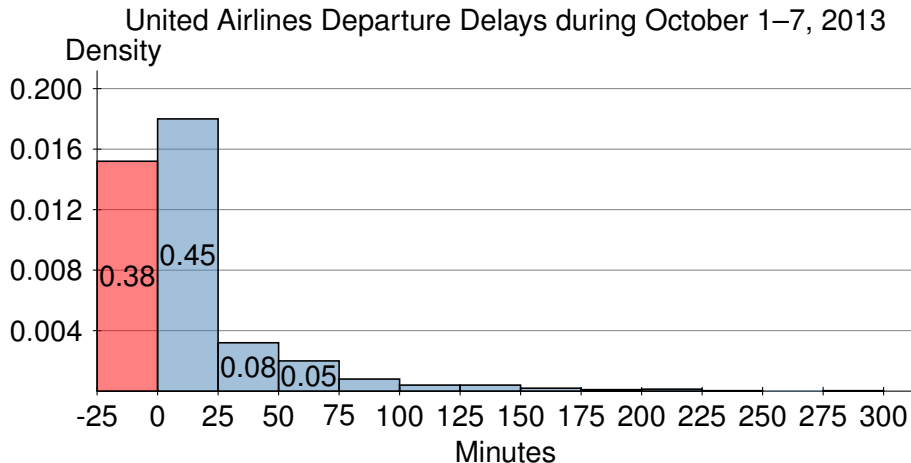
2 Estimate  $s$ .

## United Airlines Departure Delays



	GATE #	STATUS
ON	A23	DELAYED
ES	C72	DELAYED
	B34	DELAYED
	A14	DELAYED
	C89	DELAYED
	G12	DELAYED
	C5	DELAYED
	D13	DELAYED
0	A4	DELAYED
	B22	DELAYED
	A22	DELAYED

## Finding Probabilities with Density Histograms



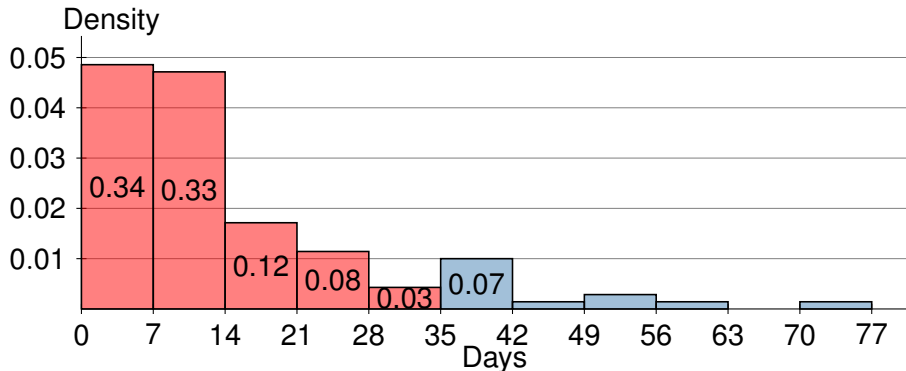
Find the probability that a flight randomly selected from the 1219 flights departed on time OR late.

# Mean Response Time to Fix Potholes in Chicago



# Probability-Proportion-Percentile Connection

Mean Response Time to Fix Potholes in Chicago



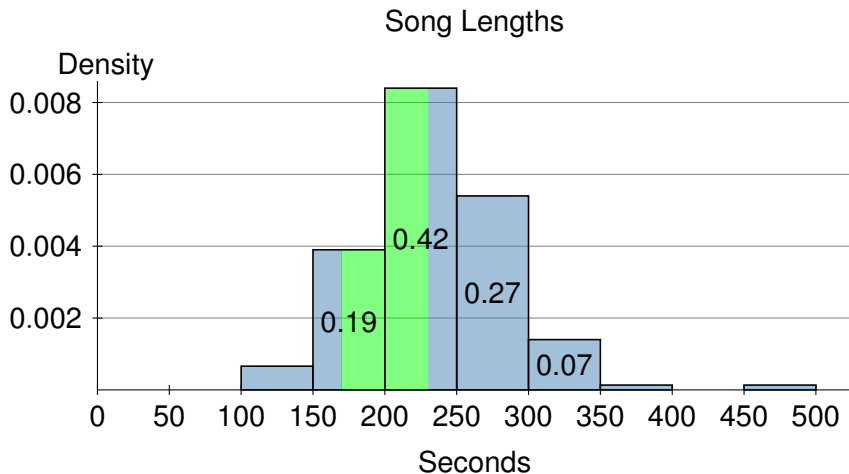
Interpret the area of the red region in the density histogram as ...

- 1 a probability.
- 2 a proportion.
- 3 a percentile.

## Song Lengths Played by Live 105

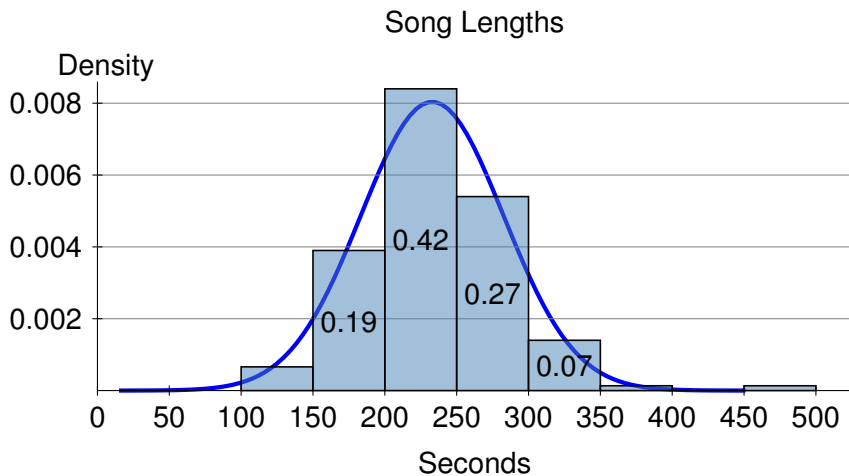


# Motivating the Normal Curve



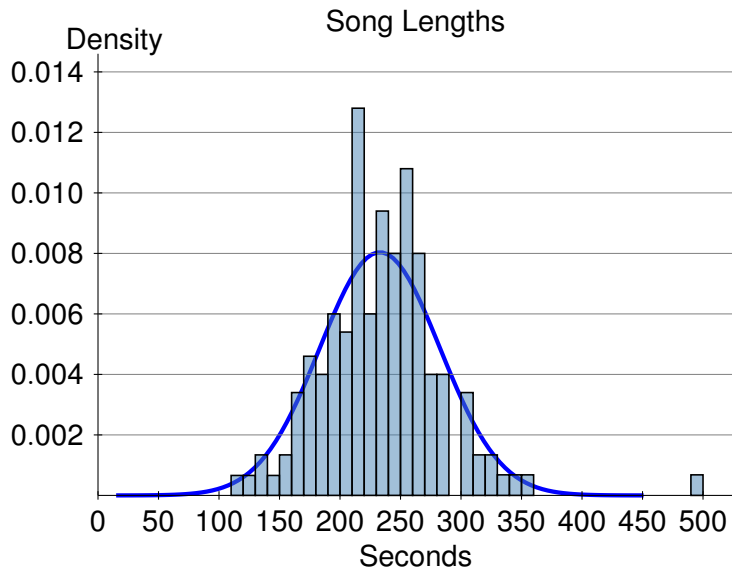
What is the probability of randomly selecting a song length between 170 and 230 seconds?

# Introducing the Normal Curve

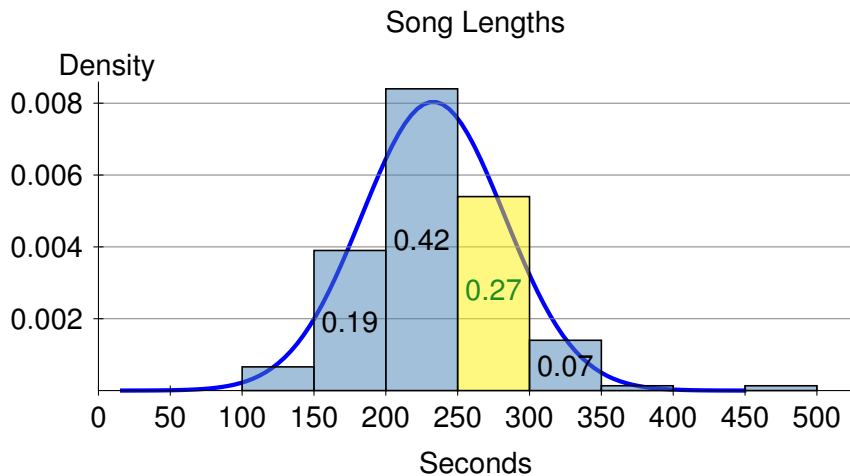




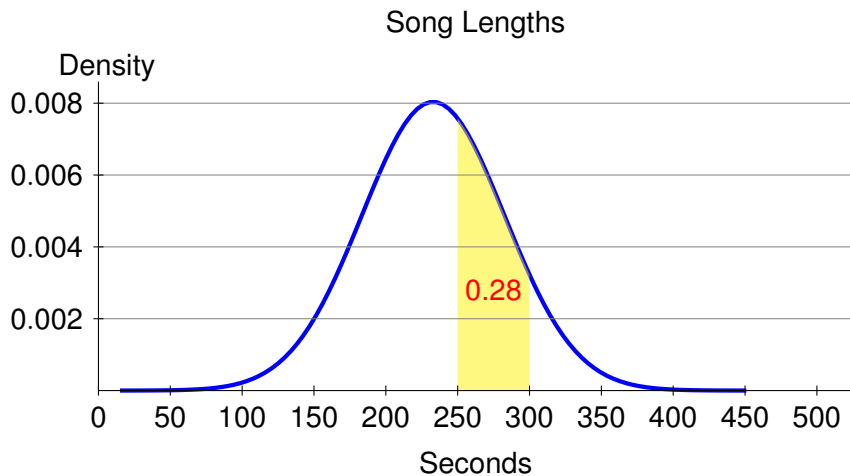
# Using Smaller Class Sizes



# Area of a Bar versus Area Under Normal Curve



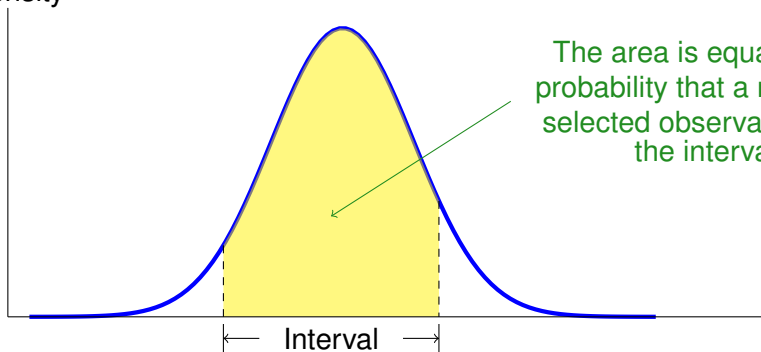
# Area of a Bar versus Area Under Normal Curve



# Area is Equal to Probability

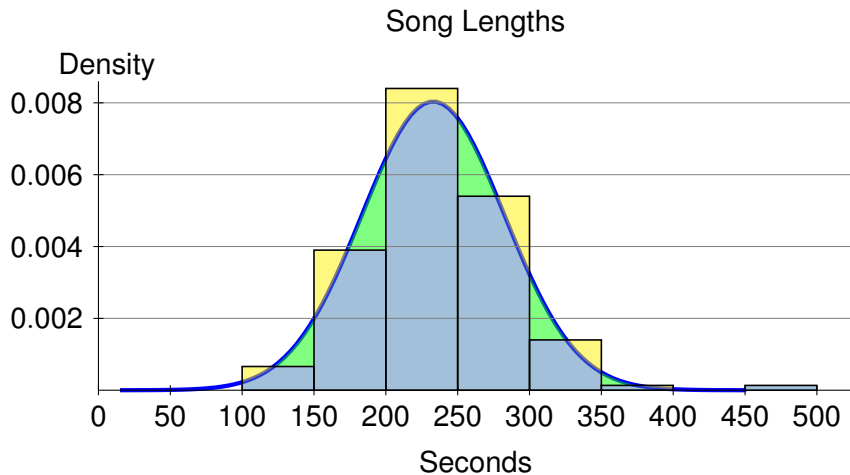
A Normal Curve

Density



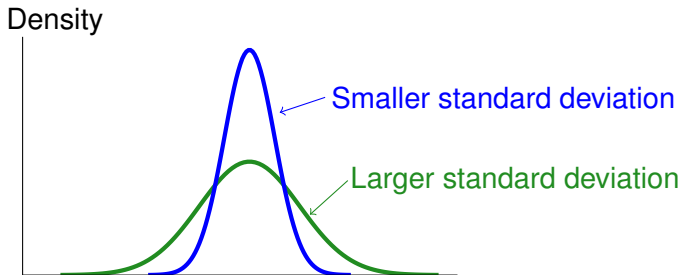
The area is equal to the probability that a randomly selected observation is in the interval

# Total Area Under Normal Curve

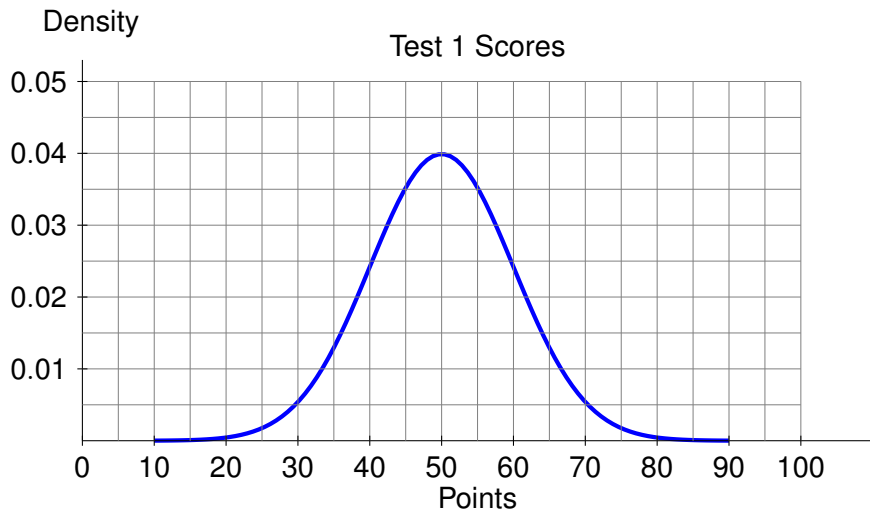


The total area under a normal curve is equal to 1.

# Normal Distributions with Different Standard Deviations

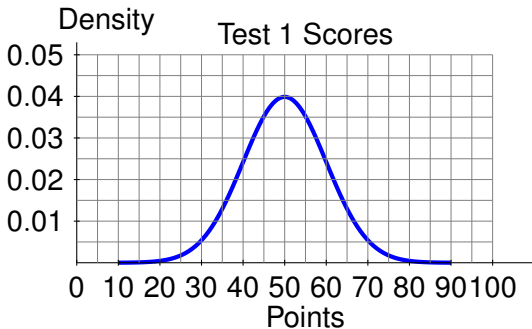


# Estimating Quartiles



Estimate  $Q_1$ ,  $Q_2$ , and  $Q_3$ .

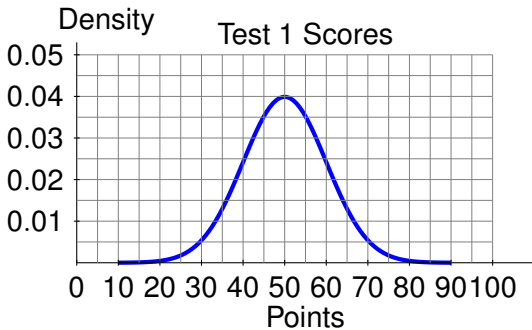
## In-Class Activity



- 1 Find the probability that a randomly selected test score is between 50 and 70 points.
- 2 Find the probability that a randomly selected test score is between 50 and 60 points.
- 3 Explain why it makes sense that your result in Problem 2 is *not* half your result in Problem 1, even though the interval in Problem 2 is half as long as the interval in Problem 1.

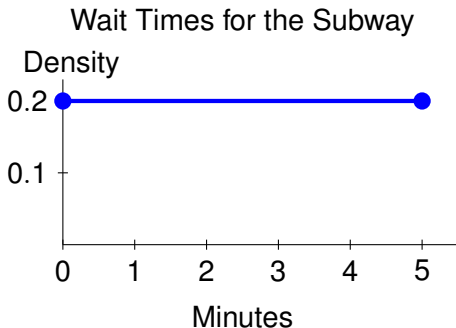


## In-Class Activity



Find the probability that a randomly selected test score is between 40 and 60 points. Is the probability that a randomly selected test score is between 50 and 60 half as much? Explain.

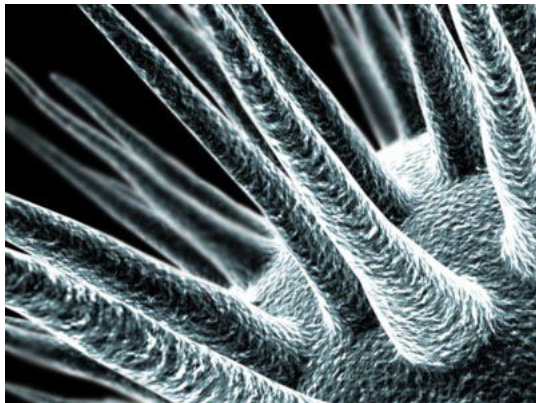
## In-Class Activity



Find the probability that a person will wait between 0 and 4 minutes on a certain day. Is the probability that a person will wait between 0 and 2 minutes half as much? Explain.

## Learning Is in the Details

Detailing concepts at a microscopic level affords quantum leaps in learning.



# Procrastinistas



Information



## Other Examples

- Residuals, Regression, residual plots, and Coefficient of Determination
- Two-way tables, proportion, and probability
- Shape, center, spread, and outliers
- Four characteristics of an association between two numerical variables



## Projects

- Hypothesis: Select two numerical variables you suspect are linearly associated.
- Collect data.
- Construct scatterplot.
- Describe the four characteristics of the association.
- Find regression equation.
- Construct residual plot.
- Interpret parameters.
- Make predictions, if appropriate.

## UFO Sightings

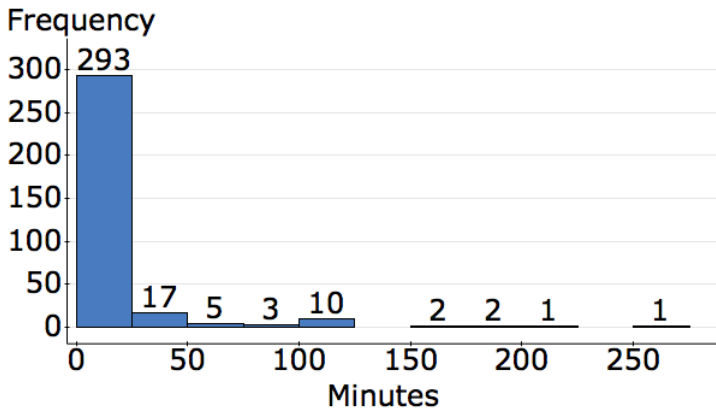
Reported to UFO National Reporting Center during the period May 14–31, 2014





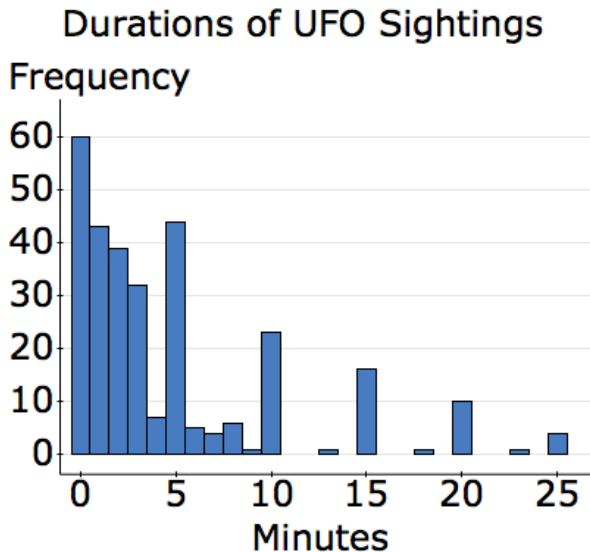
# UFO Sightings

Durations of UFO Sightings



Describe the shape of the distribution. Which is larger, the mean or the median? Which is the better measure of the center? Explain.

## Zooming In



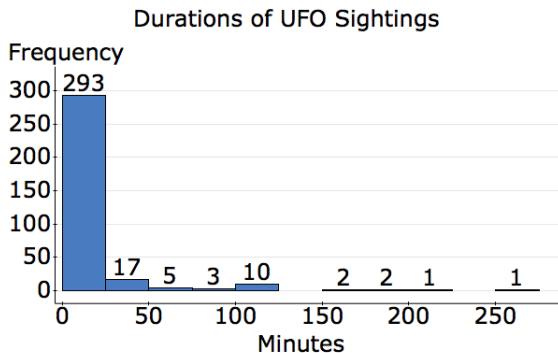
Guess why there are spikes at the following times (in minutes): 5, 10, 15, 20, and 25.

## Comparing Histograms

By inspecting both histograms, is it reasonable to assume that the mean is even greater than the median than if one were to inspect only the first histogram? Explain.

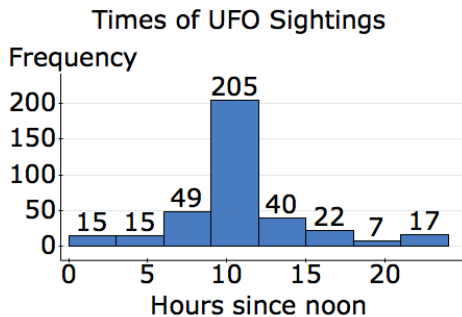
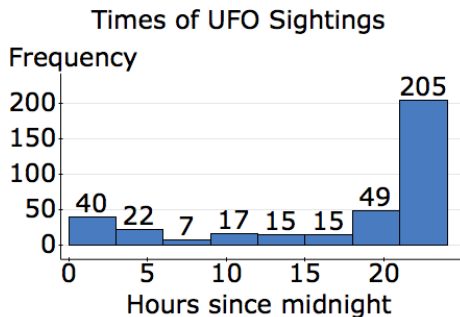


# Outliers



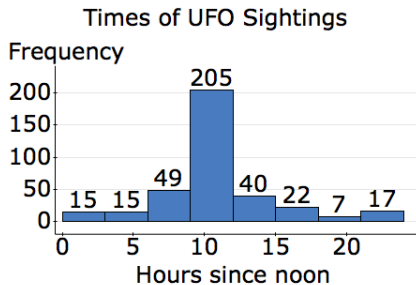
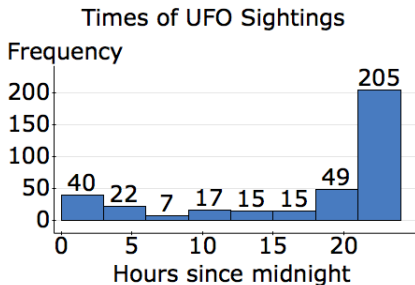
Each observation more than 25 minutes is an outlier. How could the entire right tail consist of outliers? Three durations not described by the histogram were reported as “ongoing.” Taking into account these three reports, what is the total number of outliers?

## Misleading Histogram



Explain why one distribution is bimodal and the other is unimodal. Which of the two distributions is somewhat misleading? What variable about sightings would have 365 modes for one year of data?

# Contextual Thinking

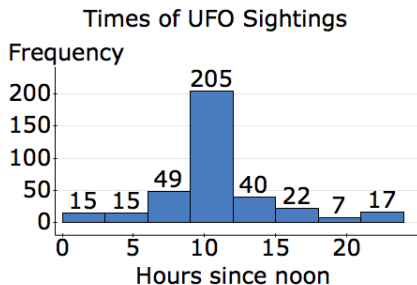
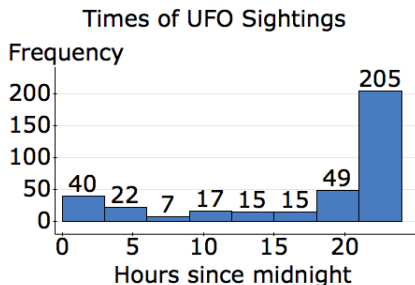


A student thinks the medians shown below must be wrong, because the values should be 12 hours apart, not 11 hours. What would you tell the student?

Summary statistics:

Column	n	Mean	Median
Hours since midnight	370	16.910811	21
Hours since noon	370	10.489189	10

# Contextual Thinking



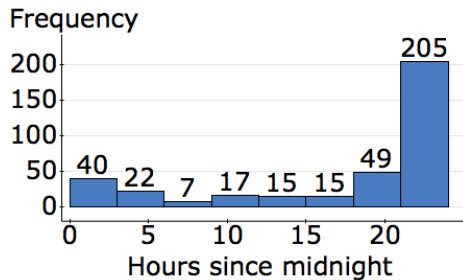
How is it possible that the means of the two distributions differ by only approximately 6.4 hours when the medians differ by so much more (11 hours)?

Summary statistics:

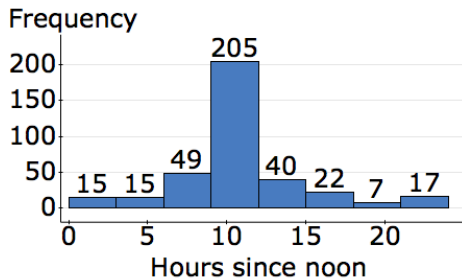
Column	n	Mean	Median
Hours since midnight	370	16.910811	21
Hours since noon	370	10.489189	10

# Comparing Results

Times of UFO Sightings



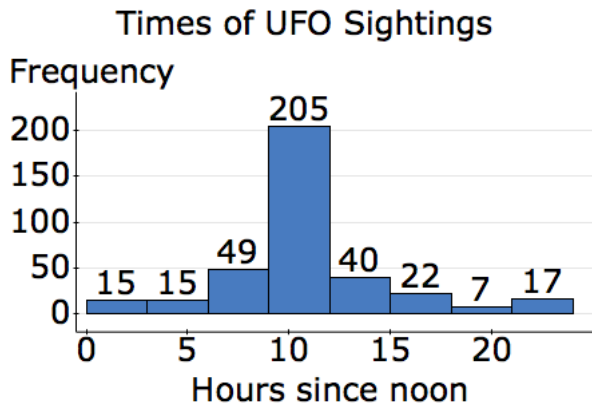
Times of UFO Sightings



During what three-hour time slot were there the most sightings? Do you get the same result by inspecting the two distributions?



## Contextual Thinking



Give three possible explanations why there are so many sightings between 9 pm and midnight.

## Promoting the Course

- Meet with counselors
- Visit arithmetic classes
- Visit high schools
- New student day

## Emphasize Non-STEM



- Course catalog
- Counselors
- First day of class

## Learning Is in the Details

Detailing concepts at a microscopic level affords quantum leaps in learning.

