

STATISTICS I

SECTION R3

SPRING 2010

FEBRUARY 17, 2010

TODAY'S PLANS

- Homework and quiz
 - Page 162:16,17
 - Other Questions?
- Measures of position
 - Quick review of percentiles, quartiles, and 5-number summary
 - z-score and outliers
 - Interquartile range (IQR) and outliers
- Preliminary data analysis
- Reminder: Test 1 is on Friday!

PERCENTILE

- To find the value corresponding to the p^{th} percentile in a list of n numbers
 - Arrange the numbers in order from smallest to largest
 - Calculate the position:
$$c = \frac{n \cdot p}{100}$$
 - If c is not a whole number, round up to the next highest whole number; the number at that position is the p^{th} percentile
 - If c is a whole number, take the average of the numbers in positions c and $c + 1$; that's the p^{th} percentile

PERCENTILE

- The percentile corresponding to a value X is given by the formula:

$$\textit{percentile} = \frac{\textit{number of values below } X + 0.5}{\textit{total number of values}} \cdot 100\%$$

Quartiles

- Quartiles divide the data into 4 equal groups
 - 25th percentile = 1st quartile
 - 50th percentile (median) = 2nd quartile
 - 75th percentile = 3rd quartile
- Five number summary
 - Low, first quartile, median, third quartile, high

z-score

- z-score – tells how far from the middle a data value is
 - Calculate z using data value, mean, and standard deviation

- Formula:

$$z = \frac{x - \bar{x}}{s}$$

- An outlier has a z-score > 3 or < -3

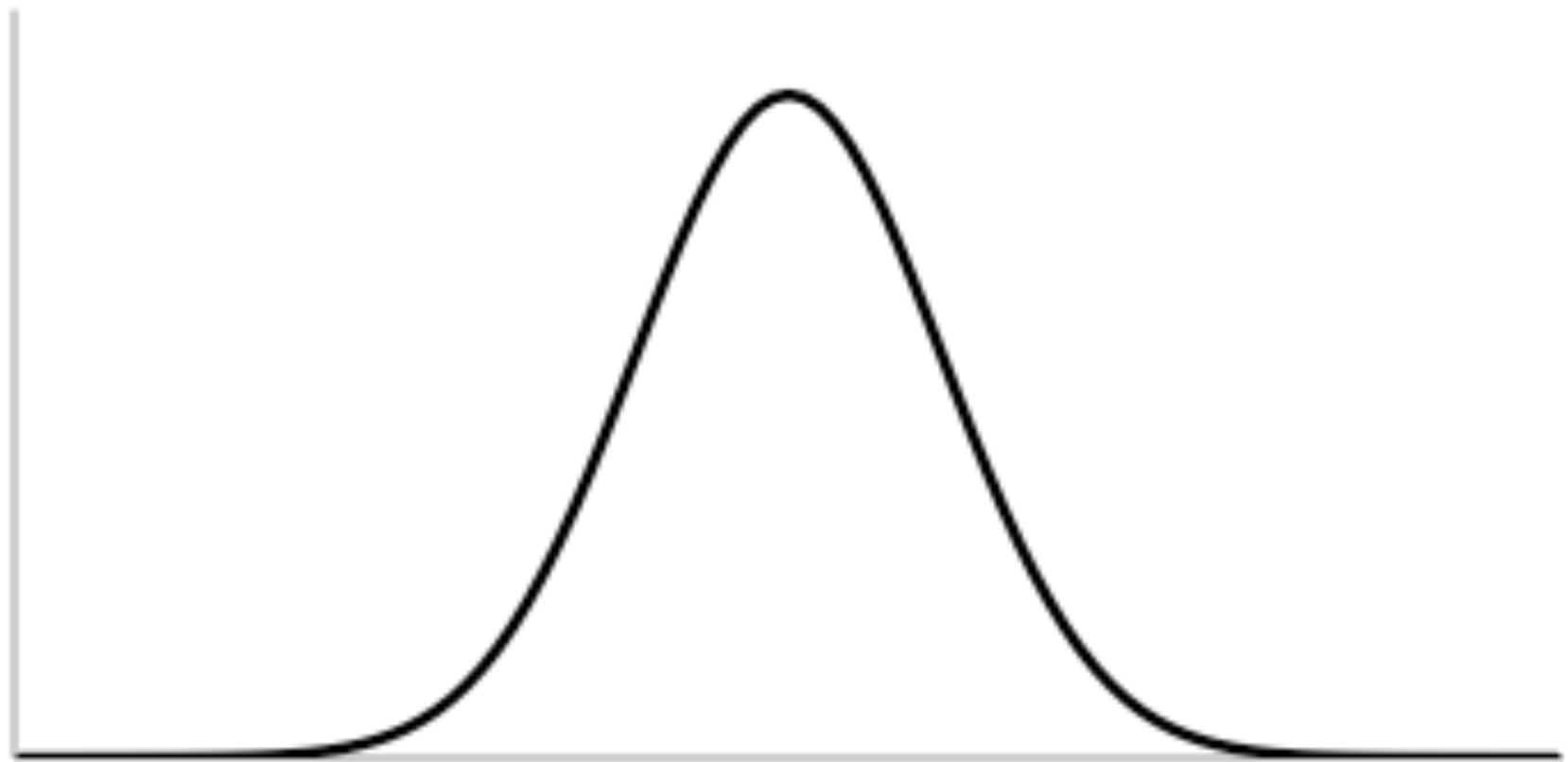
Interquartile Range

- IQR = 75th percentile – 25th percentile
 - Another formulation: $\text{IQR} = Q_3 - Q_1$
- To find outliers:
 - Calculate $J = 1.5 \cdot \text{IQR}$
 - Calculate $Q_1 - J$ and $Q_3 + J$
 - Any value that is less than $Q_1 - J$ is an outlier
 - Any value that is more than $Q_3 + J$ is an outlier

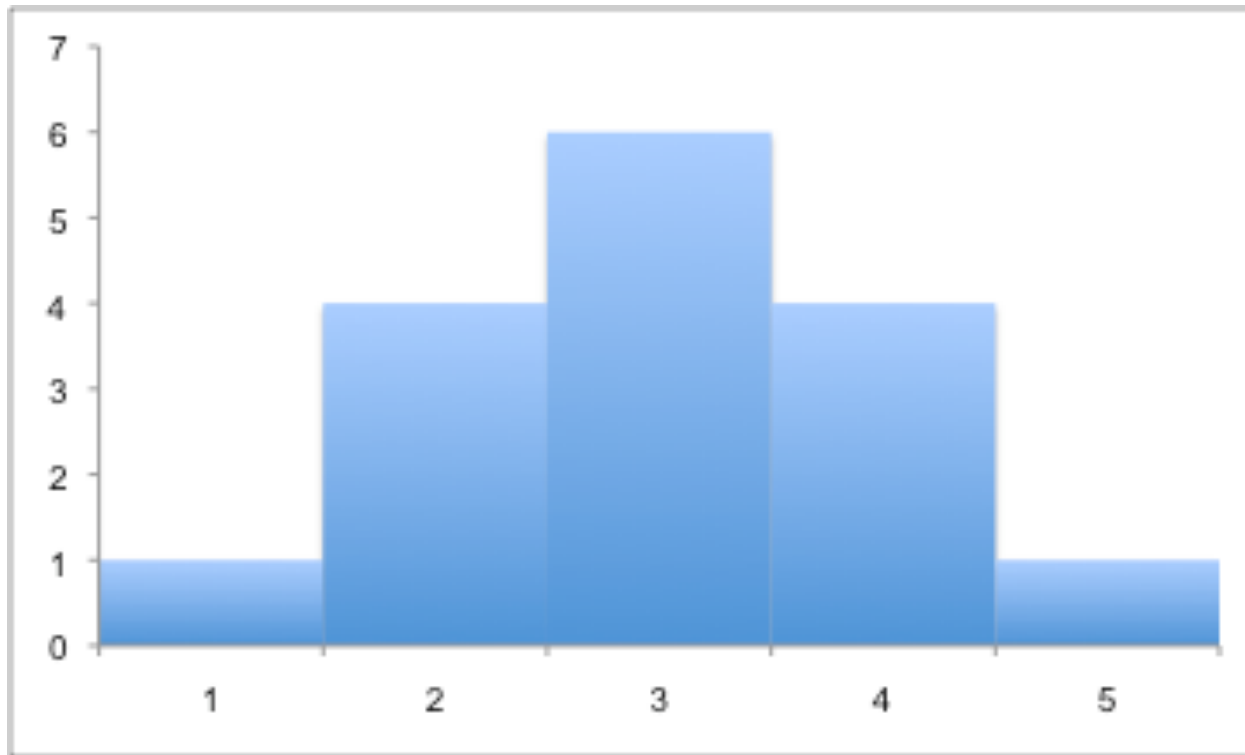
Preliminary Data Analysis

- When data is normally distributed:
 - The mean, median, and mode are the same
 - Empirical rule:
 - About 68% of the data is within 1 standard deviation of the mean
 - About 95% of the data is within 2 standard deviations of the mean
 - About 99.7% of the data is within 3 standard deviations of the mean

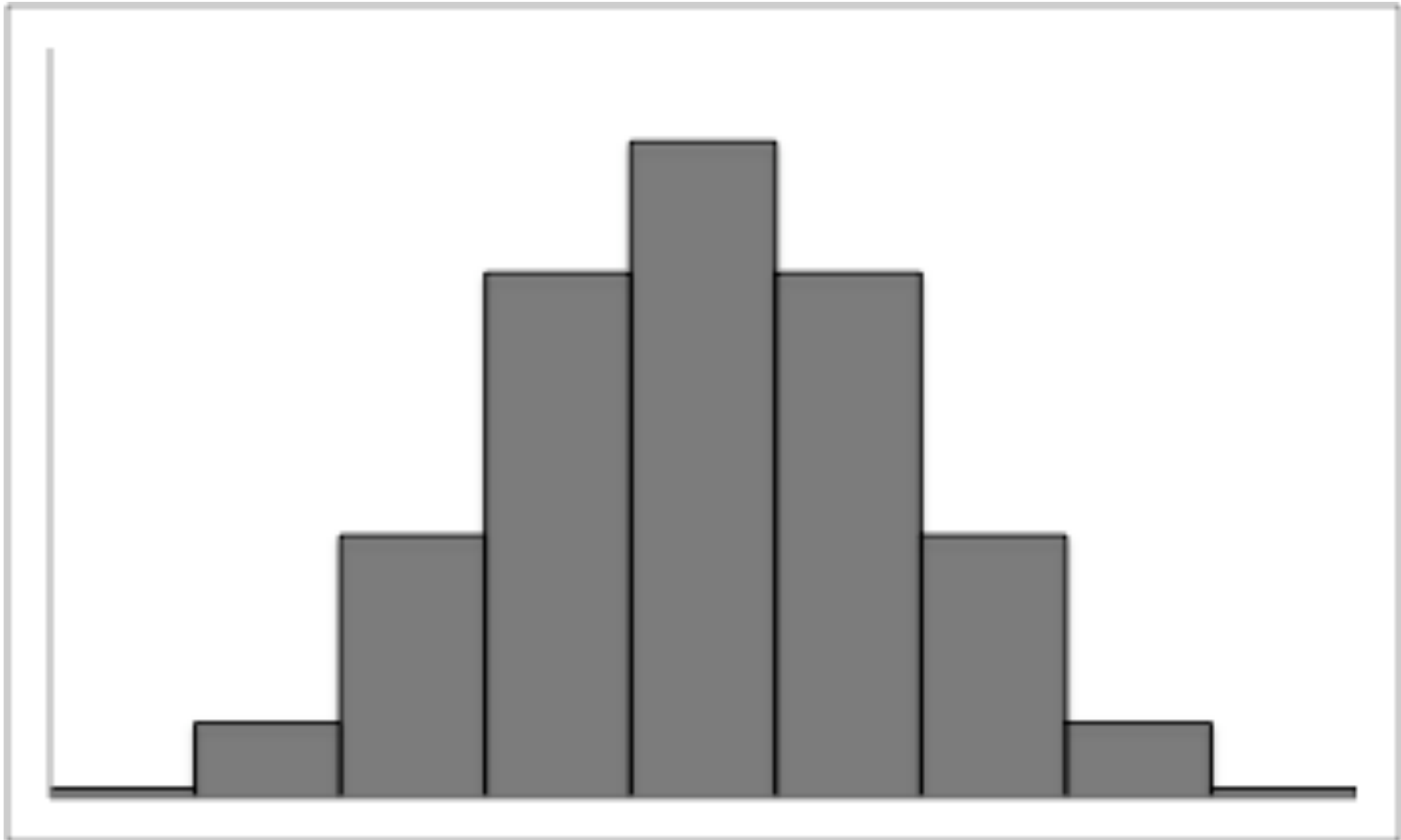
The Normal Curve



Normally Distributed Data Example 1

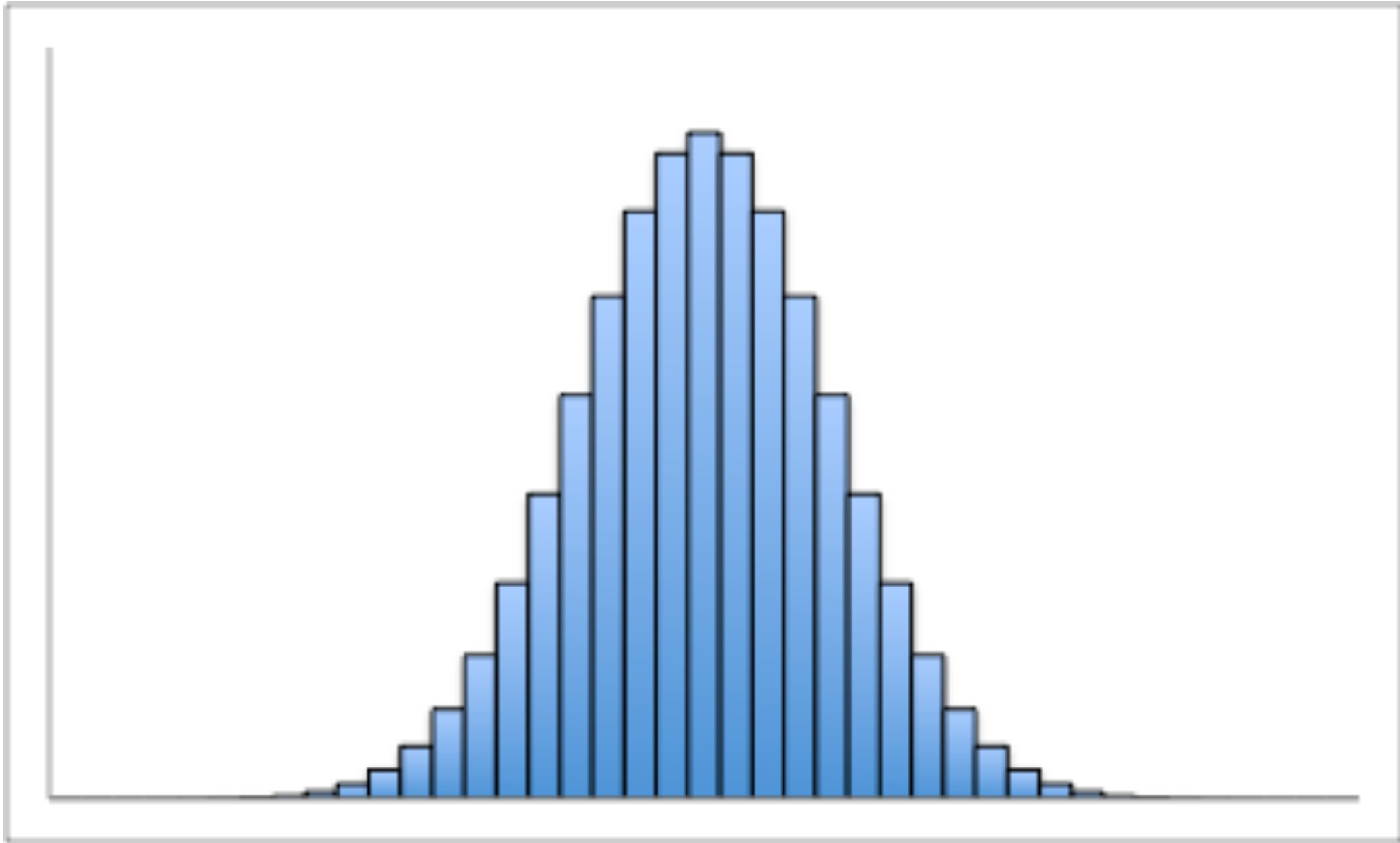


Normally Distributed Data Example 2



Normally Distributed Data

Example 3



Study Guide for Test 1

- Descriptive and inferential statistics
 - When you can make a valid inference
 - Correlation is not necessarily causation!
- Measures of central tendency
 - Mean, median, mode, and midrange – calculating and interpreting
- Measures of variation
 - Range and standard deviation
- Measures of position
 - Percentiles, quartiles, and 5-number summaries
- Preliminary data analysis
 - Outliers and other uses of the z-score