

STATISTICS I

SECTION R3

SPRING 2010

FEBRUARY 12, 2010

TODAY'S PLANS

- Homework and classwork review
- Quick review of measures of dispersion
 - Standard deviation = “average” distance from the middle
- Measures of position
 - Percentiles and quartiles
 - Five-number summary
 - z-score, interquartile range, and outliers
- Brief open-book quiz

Classwork Review

- What happens to the mean and standard deviation when you:
 - Add the same number to all values?
 - Mean: changes by the same number
 - Standard deviation: nothing
 - Subtract the same number from all values?
 - Mean: changes by the same number
 - Standard deviation: nothing
 - Multiply all values by the same number?
 - Mean: changes by the same number
 - Standard deviation: changes by the same number
 - Divide all values by the same number?
 - Mean: changes by the same number
 - Standard deviation: changes by the same number

Homework Review

1. The average salary of the graduates of the class of 1980 is \$32,500.
2. It is estimated that in Podunk there are 27,256 cats.
3. Only 3% of the men surveyed read Cosmopolitan magazine.
4. Based on a recent mail survey, 85% of the respondents favored gun control.
5. A recent study showed that high school dropouts drink more coffee than students who graduated; therefore, coffee dulls the brain.
6. Since most automobile accidents occur within 15 miles of a person's residence, it is safer to make long trips.

Bad Reasons

- I agree with the results of the study because I like that candidate too.
- I disagree with the results of the study because I think they're biased.
- I disagree with the results of the study because they didn't ask anyone I know.

Good Reasons

- I disagree with the results of the study because not everyone in the population has the same chance of being selected.
- I disagree with the results of the study because they implied causation on the basis of correlation.

**CORRELATION IS NOT
NECESSARILY CAUSATION!**

MEASURES OF POSITION: PERCENTILE

- Percentile – not percentage!
 - Divide the data into 100 equal groups
- Two calculations
 - Find the data value corresponding to a given percentile
 - Find the percentile that goes with a given data value

PERCENTILE

continued

- 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

example 1

- Numbers are 6, 7, 12, 15, 21, 33, 42, 43, 45
 - Find the number at the 30th percentile:
 - Calculate the position: $c = \frac{9 \cdot 30}{100} = \frac{270}{100} = 2.7$
 - c is not a whole number, so round it up to 3
 - The value in position 3 is at the 30th percentile
 - That value is 12

PERCENTILE

example 2

- Numbers are 6, 7, 12, 15, 21, 33, 42, 43, 45, 46
 - Find the number at the 30th percentile:
 - Calculate the position:
$$c = \frac{10 \cdot 30}{100} = \frac{300}{100} = 3$$
 - c is a whole number, so take the average of the values at positions c and $c + 1$
 - The value at position 3 is 12
 - The value at position 4 is 15
 - The average of 12 and 15 is 13.5

PERCENTILE

continued

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

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

PERCENTILE

example 3

- Numbers are 6, 7, 12, 15, 21, 33, 42, 43, 45, 46
 - Find the percentile that goes with 42:
 - Number of values below 42 = 6

$$\textit{percentile} = \frac{6 + 0.5}{10} \cdot 100\% = 65\%$$

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