

Soc 303 Lecture Notes 5 9/10/14

Sherri Verdugo
September 10, 2014

Contents

| | |
|-------------------------------------------------------------------|---|
| Administration Items | 1 |
| Key Concepts One | 1 |
| Key Concepts Two | 2 |
| Key Concepts Three | 2 |
| Prologue | 3 |
| Central Tendency | 3 |
| The Arithmetic Mean | 3 |
| Frequency Methods for Mean | 3 |
| Summation Symbol and Checkpoints using the Mean | 4 |
| The Median: Odd | 4 |
| The Median: Even | 5 |
| Cumulative Frequency | 5 |
| The Mode: Most Frequent | 5 |
| Graphing Terms | 6 |
| Skewness | 6 |
| Graphical Representations | 6 |
| Concepts of the Boxplots | 7 |
| Summary of Major Formulas and Concepts: Individual Data | 7 |
| Measuring Central Tendency | 7 |
| Summary of Major Formulas: Frequency Distributions | 7 |
| Measuring Central Tendency | 7 |
| Next Week | 8 |

Administration Items

1. Attendance (pass the roster around please)
2. Writing Assignment Draft (visit course website for information)
3. Chapter 4–Measuring Central Tendency
4. Next Week – Chapters 5 and 6
5. Organize your notes and items as the semester is moving forward to limit your stress levels!

Key Concepts One

| Term | Pg | Brief Definition |
|-------------------------------|-----|--------------------------------------------------------------|
| Measures of Central Tendency | 84 | averages or measures of location that reflects the middle |
| Measures of Dispersion | 84 | measures concerning the spread or dispersion around the mean |
| Mean/Arithmetic Mean | 85 | one form of the average accounting for distances of scores |
| Summation symbol (Σ) | 86 | Greek symbol: capital sigma Σ indicates summing |
| \bar{x} and \bar{y} | 86 | the mean value of a variable x or y (what is under the bar) |
| Median (Md.) | 90 | equal scores greater and less than the median (middle) |
| Median Position (Md. Pos.) | 90 | person, place, or thing possessing the median position |
| Array | 91 | a listing from highest to lowest or lowest to highest |
| Cumulative Frequency (cf) | 95 | determines which value of x is the median by frequencies |
| Mode | 98 | most frequent case |
| Modal class/ modal category | 99 | most frequent case for grouped data |
| x-axis | 100 | horizontal line @ the graph bottom perpendicular to y-axis |
| f or y -axis | 101 | vertical line @ the graph left side perpendicular to x-axis |

3 Measures of Central Tendency: Mean, Median, and Mode or M^3

Key Concepts Two

| Term | Pg | Definition |
|------------------------------|-----|-----------------------------------------------------------|
| Origin of a Graph | 100 | point where the x-axis and y-axis intersect |
| Frequency Polygon | 101 | connecting dots by drawing a straight line as x increases |
| Histogram | 101 | graph of bars representing frequency of scores |
| Smooth Curve | 101 | curved line fitting through the dots not a straight line |
| Continuous Variable | 103 | not limited to a finite number of scores |
| Unimodal Freq. Distribution | 103 | distribution with one mode |
| Bimodal Freq. Distribution | 103 | distribution with two modes |
| Trimodal Freq. Distribution | 103 | distribution with three modes |
| Modality | 103 | number of modes found in the frequency distribution |
| Skewness | 107 | extent to which the frequency deviates from symmetry |
| Symmetry | 107 | balance between the right and left halves of the curve |
| Symmetric Freq. Distribution | 107 | frequency distribution with no skewness |

3 Measures of Central Tendency: Mean, Median, and Mode or M^3

Key Concepts Three

| Term | Pg | Definition |
|---------------------------|-----|------------------------------------------------------------|
| Positively Skewed (right) | 109 | skewed in the direction of increasing positive values of x |
| Negatively Skewed (left) | 109 | skewed in the direction of decreasing negative values of x |
| Stem & Leaf Diagram | 112 | representation of histogram w/ scores |
| Box plot Graph | 113 | representation of small data sets |
| Box & Whisker Graph | 113 | representation of small data sets |
| Fractiles | 114 | measurement that divides scores into small equal groups |
| Quartiles | 114 | measurement that divides scores into sets of 4 |
| Deciles | 114 | measurement that divides scores into sets of 10 |
| Percentiles | 114 | measurement that divides scores into sets of 100 |

3 Measures of Central Tendency: Mean, Median, and Mode or \hat{M}^3

Prologue

This chapter is the first introduction into the mathematical concepts of statistics. If I want to measure something in a sample from the population, how would I go about it? Perhaps I would want to look at the measures of central tendency and how the numbers behave as a group. Be careful with the equations so that you know what they are used for.

Central Tendency

- Mean, Median and Mode

The Arithmetic Mean

Sophisticated and classy method of the three measures of central tendency

Formula: $\bar{x} = \frac{\Sigma}{n}$ where Σ = the sum of all the values and n = number of cases

Example Page 85:

Group A (\bar{A}): {8, 8, 8, 7, 7, 7, 6, 6, 6} where $n = 9$ (you count the items) $\bar{A} = \frac{63}{9} = 7$

Group B (\bar{B}): {9, 9, 8, 8, 8, 7, 7, 7, 6, 6} where $n = 10$ (you count the items) $\bar{B} = \frac{75}{10} = 7.5$

Based on the mean... which group did better?

Would changing n make a difference?

Frequency Methods for Mean

Using Group A Data again:

Group A (\bar{A}): {8, 8, 8, 7, 7, 7, 6, 6, 6} where $n = 9$ (you count the items) $\bar{A} = \frac{63}{9} = 7$

Frequencies: we can group similar data points... 8, 7, and 6... we add up the frequencies this time.

| x = | f = | fx = |
|-----|-----|------|
| 8 | 3 | 24 |
| 7 | 3 | 21 |
| 6 | 3 | 18 |

$n = 9$ and $\Sigma fx = 63$ or $24 + 21 + 18 = 63$

$$\bar{A} = \frac{\Sigma fx}{n} = \frac{63}{9} = 7.0$$

Summation Symbol and Checkpoints using the Mean

The summation symbol is a way of statisticians, mathematicians and programmers to write “please add up all the items from the first entry to the very last entry.” Using the professional way, we would write this as:

$$\bar{x} = \frac{\Sigma_{i=1}^n X_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

For this class we are using: $\bar{X} = \frac{\Sigma X}{n}$

Using Group A Data we can write this as:

$$\bar{A} = \frac{x_1 + x_2 + \dots + x_9}{9}$$

or simply: $\bar{A} = \frac{6+6+6+7+7+7+8+8+8}{9} = \frac{63}{9} = 7.00$

Check your work: Box 4.2 Page 89 For individual and ungrouped data points: $\Sigma(x - \bar{x}) = 0$

Check your work: Box 4.2 Page 89 For Grouped Frequencies: $\Sigma[(x - \bar{x})f] = 0$

The Median: Odd

When the number of items in an array are ODD we use the following approach to find the middle by locating the median position (middle position): Median does not always equal the Mean!

Equation: $Md.Pos. = \frac{n+1}{2}$ For Group A we have $Md.Pos. = \frac{9+1}{2} = 5$

| i = | x = |
|-------------------|--------------|
| 9 | 8 |
| 8 | 8 |
| 7 | 8 |
| 6 | 7 |
| 5 <i>Md. Pos.</i> | 7 <i>MD.</i> |
| 4 | 7 |
| 3 | 6 |
| 2 | 6 |
| 1 | 6 |

The Median: Even

When the number of items in an array are EVEN we use the following approach to find the middle by: 1) locating the median position (middle position) and 2) add the score above and below the Md. Pos. then divide by 2. **Equation:** $Md. = \frac{score_{above} + score_{below}}{2}$. For Group B Md. Pos. = $\frac{n+1}{2} = \frac{10+1}{2} = \frac{11}{2} = 5.5$. The Median = $\frac{7+8}{2} = \frac{15}{2} = 7.5$

| i = | x = |
|----------|----------|
| 10 | 9 |
| 9 | 9 |
| 8 | 8 |
| 7 | 8 |
| 6 | 8 |
| 5 | 7 |
| 4 | 7 |
| 3 | 7 |
| 2 | 6 |
| 1 | 6 |

Cumulative Frequency

A way to account for all the frequencies generated up to a specific value of x. It is used to determine which value of x is at the median position. This is great for large data sets.

For now, let's work through the example on page 97 in the textbook.

Next, take a moment at this point to partner up and work with SPSS using the ICU data file located on the course website in this week's folder and on the github repository. We will work on this together:

<http://libguides.library.kent.edu/SPSS/ImportData>

Find the Mean, Median and Cumulative Frequency for the Age Variable.

| Min. | 1st Quarter | Median | Mean | 3rd Quarter | Max |
|------|-------------|--------|------|-------------|------|
| 16.0 | 46.8 | 63.0 | 57.5 | 72.0 | 92.0 |

The Mode: Most Frequent

Mode: A category of a variable that contains more cases than can be found in either category adjacent to it. You can have more than one mode or no mode (nothing is repeated).

Modal Class or Modal Category: Where data have been grouped, a class interval or category that contains more cases than can be found in either category adjacent to it.

What is the mode? What is the modal category? Are these grouped or individual scores?

| Age Range | f = |
|-----------|-----|
| 50-59 | 3 |
| 40-49 | 9 |
| 30-39 | 4 |
| 20-29 | 2 |
| 10-19 | 7 |
| 0-9 | 1 |
| Total | 26 |

Graphing Terms

Frequency polygon: a straight line created by a connection of dots, as x increases.

Histogram: graph in which bars are created from one-half unit below each value of x to one-half unit above that value. The y-axis indicates the frequency of each score's occurrence.

Smooth curve: connection of dots similar to a frequency polygon but generated by a curved line fitting through the dots instead of a series of straight lines.

Continuous variable: not limited to a finite number of scores.

Unimodal, bimodal, trimodal: a distribution with one, two, and three modes.

Modality: the number of modes found in the frequency distribution.

The mode is the only measure of central tendency that is appropriate for ALL LEVELS of measurement.

https://github.com/verdu103/ExData_Plotting1/blob/master/README.md

Skewness

Skewness: the extent to which the frequency distribution deviates from symmetry

Symmetry: the balance between the right and left halves of the curve

Symmetric frequency distribution: a frequency distribution with no skewness (skewness value equals zero)

Skewed to the right or positively skewed: skewness is in the direction of increasing positive values on the x-axis

Skewed to the left or negatively skewed: skewness is in the direction of decreasing negative values on the x-axis

https://github.com/verdu103/ExData_Plotting1/blob/master/README.md

Graphical Representations

Stem and Leaf Display: A graphic representation that combines the visual effect of a histogram but preserves the actual scores in small to medium sized data sets.

Boxplots or Box and Whisker Plots: Useful representations of small data sets, where more traditional graphs cannot be used effectively.

Break Point: Use SPSS and the ICU data set to generate a stem and leaf display as well as a box and whiskers plot.

<http://www-01.ibm.com/support/knowledgecenter/>

Concepts of the Boxplots

Fractiles: measurement that divides scores into smaller groups of scores of approximately equal size

Quartiles: a number that divides scores into sets of 4 and indicates for each individual studied the number of people (or other units of analysis) his or her score exceeds

Deciles: a number that divides scores into sets of 10 and indicates for each individual studied the number of people (or other units of analysis) his or her score exceeds

Percentiles: a number that divides scores into sets of 100 and indicates for each individual studied the number of people (or other units of analysis) his or her score exceeds

Summary of Major Formulas and Concepts: Individual Data

Measuring Central Tendency

| Term | Equation | Definition |
|-----------------------|-------------------------------------|-----------------------------------------|
| Mean | $\bar{x} = \frac{\Sigma X}{n}$ | Average of scores |
| Median Position: Odd | $Md.Pos. = \frac{n+1}{2}$ | Score that equally divides the data set |
| Median: Odd | Find Md. Pos. | Score located directly at the Md. Pos. |
| Median Position: Even | $Md.Pos. = \frac{n+1}{2}$ | Will not be a whole number |
| Median | $\frac{scoreAbove + scoreBelow}{2}$ | Score combining the midpoints (pg. 93) |
| Mode | | Most frequent occurring score |

Summary of Major Formulas: Frequency Distributions

Measuring Central Tendency

| Term | Equation | Definition |
|-----------------------|-------------------------------------|--------------------------------------------|
| Mean | $\bar{x} = \frac{\Sigma fX}{f}$ | Average of scores |
| Median Position: Odd | $Md.Pos. = \frac{f+1}{2}$ | Score that equally divides the data set |
| Median: Odd | Find Md. Pos. | Score located directly at the Md. Pos. |
| Median Position: Even | $Md.Pos. = \frac{f+1}{2}$ | Will not be a whole number |
| Median | $\frac{scoreAbove + scoreBelow}{2}$ | Score combining the midpoints (pg. 93) |
| Mode | | Most frequent occurring score: frequencies |

Next Week

1. Recap of previous weeks
2. Chapter 5 – Measuring Dispersion
3. Chapter 6 – Constructing and Interpreting Contingency Tables
4. More SPSS and Excel interaction
5. Homework 1 is coming up
6. Writing Draft