Math 114 ACTIVITY 2: Numerically summarizing data – Measures of Central Tendency: Use of calculation tools

Why

In order to use sample information to draw conclusions about populations, we need to summarize data (the same vocabulary is used in describing populations as in describing sample data). To describe a distribution of values, we think mainly in terms of three ideas – 1.) The *Center*: Where is the "middle" of the values (what is a "typical" value)? 2.) The *spread* How much variation is there in the values? (How much do they spread out?). 3.) The *Shape* What does the distribution (the histogram) of the values look like?

Today we look at numerical descriptions of Center, and at the most basic ways of describing Shape. We also take the first step in learning the tools (calculator, software) necessary for doing the calculations necessary for working with sets of data.

LEARNING OBJECTIVES

1. Learn the most common measures of center for data sets
2. Learn the strengths and weaknesses of the different measures of center
3. Develop some sense of the *shape* descriptions used for distributions of quanatitative variables.
4. Continue developing skills in working with teams.

CITERIA

1. Success in working as a team and in fulfilling the team roles.
2. Success in involving all members of the team in the conversation.
3. Success in completing the exercises.

RESOURCES

1. The course syllabus
2. The team role desk markers (handed out in class for use during the semester)
3. Your text - especially section 3.1
4. Your calculator and the "Using your calculator for statistics" page from the Web for *your* calculator
5. 40 minutes

PLAN

1. Select roles, if you have not already done so, and decide how you will carry out steps 2 and 3 (5 minutes)
2. Work through the exercises given here - be sure everyone understands all results & procedures(30 minutes)
3. Assess the team's work and roles performances and prepare the Reflector's and Recorder's reports including team grade (5 minutes).

TERMINOLOGY

The *arithmetic mean* of a collection of numbers $\{x_1, x_2, \ldots, x_n\}$ is the number $\dfrac{x_1 + x_2 + \cdots + x_n}{n} = \dfrac{\sum x_i}{n}$. It gives the location at which the distribution (think of a histogram cut out of cardboard) would balance. Most useful with symmetric distributions – does not represent "typical" values (but still marks balance point) for skewed distributions.

The *median* of a collection of numbers is the number in the middle when the numbers are sorted in order. If there is an even number of items, the median is taken to be halfway between the two middle numbers. It gives the point at which the *area* of the histogram divides in half.

The *mode* of a collection of numbers is the value with the highest frequency.

A measure is *resistant* if it is not greatly affected by a few extreme values – the mean is not resistant, the median is resistant. (Resistant measures are more useful with non-symmetric distributions – and especially with skewed distributions)

A distribution of values is *uniform* if all values have (approximately) equal frequencies.

A distribution is *symmetric* if there shape of the distribution to the left of the its center mirrors the shape to the right.

A distribution is *mound-shaped* if it is roughly symmetric and the frequencies become larger closer to the center.

A distribution is *skewed to the right* if there is a long "tail" of values with decreasing frequencies extending out to the right (not matched on the left).

A distribution is *skewed to the left* if there is a long "tail" of values with decreasing frequencies extending out to the left (not matched on the right).

In discussing shapes of histograms, we always visually "smooth out" the corners and roughness (we may disregard a single bar that is lower/higher than its neighbors, if it does not match the overall pattern).

EXERCISE

1. Use the formulas/definitions to obtain the mean, median, and mode of this collection of values:
   2, 3, 8, 6, 5, 6, 2, 4, 9, 7, 4, 2

2. Use the built-in functions on your calculator (on TI-83/84 this will be the lists and the "1-var Stat" command) to calculate the mean and median of this set of numbers (there are 13 numbers here) 5.6, 8.2, 7.1, 4.6, 4.5, 3.9, 1.7, 2.2, 9.4, 12.4, 8.9, 7.6, 5.4
   Round the mean to *one decimal place* more than the given values.

3. Use Minitab (statistics package on the campus network) to calculate the mean and median of the "self-concept" values for the data from Activity 1. You need to log in to a coomputer on the network and go to the Blackboard site for the course.

   (a) The data file is available (you do not have to enter any figures) in the Activities area of the Blackboard site. Double-click on the file, it will download and start Minitab [You may have to click "Open with" and select Minitab 16].

   (b) The top of the screen is the "Session" window, where results will appear. The bottom is the "data" window, containing the data file. Each column is a variable (corresponding to the original typed data list).

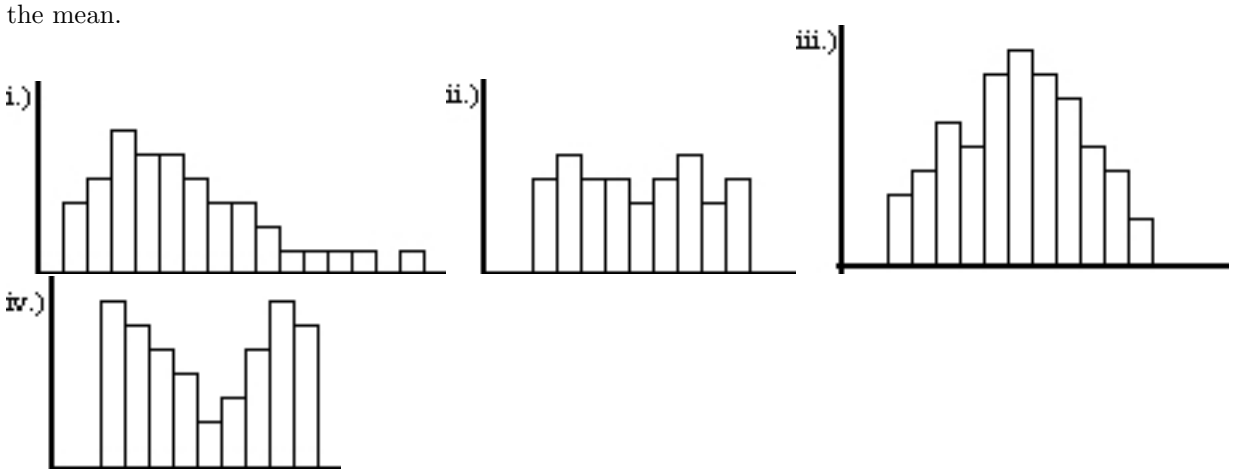
   (c) On the "Stat" menu, point to "Basic Statistics" and slide over to "Display Descriptive Statistics"

   (d) In the window that opens: With the cursor in the "Variables" pane, highlight "C5 self-concept" in the left-hand pane and click "Select" at the bottom.

   (e) Click OK

   (f) A row of values (with labels above) appears in the "Session" window. Read off the mean and the median (we'll discuss most of the others during the semester).

4. For each of these distributions: a.) Describe the shape ["smooth out" the stairsteps and corners as you look at the graph]. b.) Say whether you expect the median to be noticeably larger, noticeably smaller, or about the same as the mean.



READING ASSIGNMENT (in preparation for next class)
In Sullivan, read section 3.2 on measures of dispersion ("spread") for Monday

SKILL EXERCISES:(hand in - individually - at next class meeting) Sullivan p.125 #14, 16, 18, 23, 36, 38, 40, 41