3. Means

3.1. Example.

Suppose that we gather the following data on the savings (in $1,000's of dollars) for the nine OU employees who retired last July.

2	2	2
10	17	24
26	34	45

Question. Is this census data or sample data?
Question. What are the "average" savings?
Note that there are several possibilities for the average:

- The most frequent observations is two (the *mode*).
- The 50th percentile is seventeen (the *median*).
- The *midrange* is 23.5 (average of highest and lowest).
- The arithmetic average or **mean** is eighteen.

$$\text{mean} = \frac{2+2+2+10+17+24+26+34+45}{9}$$

The mean is the kind of average that you will most frequently encounter; in this course we will use “mean” and “average” interchangeably.

If we have **census data** then the symbol for the mean is μ and we say we have found the **population mean**. If we have **sample data** then the symbol for the mean is $\bar{x}$ and we say we have found the **sample mean**. Whether we have census or sample data the mean is computed the same way:

$$\text{mean} = \frac{\text{sum of observations}}{\text{number of observations}}$$

- Given a choice, which you rather know: the sample mean or the population mean?
- Are you studying the population or the sample?

You are studying the sample in order to draw conclusions about the population.

Whenever you gather data you will be dealing with a population.

- For *census data* you have complete information on the entire population;
- For *sample data* you have information on only part of the population.

Thus

- The *population mean* μ involves **no uncertainty.**
- The *sample mean* $\bar{x}$ involves **uncertainty.**

Other things being equal census data is preferable to sample data. In the real world, however, often only sample data is available. Since sample data must always involve error – being incomplete – it is important to develop strategies to minimize error. Statistical tools permit the researcher to determine how large the error might be.

3.2. Example.

Suppose that nine OU employees retire in October with the following accumulated savings (again listed in $1,000's of dollars):

11	*15*	*16*
17	*18*	*20*
20	*20*	*25*

Find the mean.

Solution. To find the mean, first find the sum of all the numbers. In this case, that sum is 162. Now divide by the number of observations (in this case 9):

$$\text{mean} = \frac{162}{9} = 18.$$

This is easy to do with a hand calculator or with spreadsheets.

MEANS.XLSX pre-loads the Excel formulae for calculating means and related statistical measures.

We have computed the means of two data sets so far:

2	11
2	15
2	16
10	17
17	18
24	20
26	20
34	20
45	25

Each has nine observations and each has a mean of 18. Yet one data set seems to be somewhat more *variable* than the other. Measuring how variable a set of observations are from their mean is our next topic. Before doing that we will briefly introduce some mathematical notation.

If you take observations on n subjects ($n = 9$ in our previous two examples) then you have a list of n numbers:

$$x_1, x_2, \cdots, x_n$$

In the example we just completed, the x_i's have the particular values

x_1	2
x_2	2
x_3	2
x_4	10
x_5	17
x_6	24
x_7	26
x_8	34
x_9	45

Using this notation, then we can write the mean as

$$\text{mean} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

A shorthand notation for

$$x_1 + x_2 + \cdots x_n$$

is

$$x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i$$

and so we can also write

$$\text{mean} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

This shorthand notation frequently appears in statistics books. Sometimes calculators will use the capital sigma Σ to indicate a function summing a series of numbers. There is Σ button in Excel for summing a list of numbers.
To summarize, if you have sample data

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

and if you have census data

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$