

Name \_\_\_\_\_

## Lab 1: Measurement, Sampling, & Probability

### 1. Summation Notation

The Greek capital letter sigma ( $\Sigma$ ) is used to indicate summation. The summation notation is actually a bit more complex, because it formally specifies the range of the variable over which you will sum:

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

That says, add together all the values for a variable ( $x$ ) that takes on values  $x_1, x_2, x_3, \dots, x_n$ . For our purposes, and as a means of simplification, we won't bother with the subscripts, but simply say  $\Sigma x$ .

In using summation notation, we also need to be careful about the order of operations, so here's an exercise to clarify the impact of the order of operations on summation notation. In general, you know that certain mathematical operations take precedence over other operations. As seen in your notes for Chapter 1:

#### Order of Mathematical Operations

1. Any calculation within parentheses is done first.
2. Squaring (or raising to other exponents) is done second.
3. Multiplying and/or dividing is done third (from left to right).
4. Summation ( $\Sigma$ ) is done next.
5. Finally, addition and/or subtraction is done.

Use the data below for an exercise like that found in your textbook (p. 27).

x	y
4	3
2	1
1	5
5	2
3	4

Formula	Result	Formula	Result
$\Sigma x$		$\Sigma y$	
$\Sigma x^2$		$\Sigma xy$	
$(\Sigma x)^2$		$(\Sigma x)(\Sigma y)$	
$\Sigma x-2$		$\Sigma(x-2)^2$	
$\Sigma(x-2)$		$(\Sigma(x-2))^2$	

## 2. Measurement Issues and the “Rule of Thumb”

Science relies heavily on measurement. You may have heard of the distinction between the hard sciences and the soft sciences. (And, I’m sure that you know where psychology fits in that dichotomy!) But another dichotomy that has been suggested is between the hard sciences and the easy sciences. That distinction is largely based on the ease with which scientists are able to measure their variables of interest. The variables that interest psychologists are damned difficult to measure precisely! First of all, think of some variables that you know (or imagine) would be of interest to psychologists, then try to figure out how you would measure them (essentially an operational definition).

Variable	How to Measure

We’ll talk about how psychologists might approach measuring their variables in this class, as well as other classes that you’ll take in psychology (and especially in Experimental Psychology). However, for now, let’s take on an “easier” measurement task.

You’ve heard of a “Rule of Thumb.” The origins of this “rule” are not clear. Some people claim that it originated as a measurement of the diameter of the rod with which a husband was allowed to beat his wife. The evidence for that origin is not extremely strong, though there is some support in some court rulings (ouch!). The OED defines the Rule of Thumb as “A method or procedure derived entirely from practice or experience, without any basis in scientific knowledge; a roughly practical method. Also, a particular stated rule that is based on practice or experience.”

Hmmm. Well, regardless of the etymological origins of the term, do you know how wide your thumb is? As I think you’ll soon learn, the little rascal is a bit difficult to measure. What I want you to do is to figure out a means of measuring your thumb, then take your own measurement. (Use the metric system.) Then I want you to have three other people in the class take a measurement of your thumb. Their methods may differ from yours, and their actual measurements may differ as well.

Person Measuring	Yourself			
<b>Measurement</b>				

OK, so you probably see some variability in the measurement of your thumb. How do you know which measurement to use?

### 3. Sampling from a known population

Suppose that you wanted to describe a population. You could do so by using the appropriate descriptive parameters and by describing the shape of the population. For example, suppose that the data below are the population of interest. Describe the population as fully as you can.

Population data:

1, 2, 3, 3, 4, 4, 4, 4, 5, 5, 5, 7, 9, 12, 15, 20

$\Sigma X =$	$\Sigma X^2 =$
--------------	----------------

Shape	
Central Tendency	<p>Mode =</p> <p>Median =</p> <p>Mean (<math>\mu</math>) = <math>\frac{\sum X}{N}</math></p>
Variability	<p>Range =</p> <p>Sum of Squares = <math>\sum X^2 - \frac{(\sum X)^2}{N} =</math></p> <p>Variance (<math>\sigma^2</math>) = <math>\frac{SS}{N} =</math></p> <p>Standard Deviation (<math>\sigma</math>) = <math>\sqrt{\sigma^2} =</math></p>

Of course, you would not ordinarily be able to measure all the members of a population. However, in a statistics class we can define some arbitrary and small populations in an effort to learn more about the relationship between populations and samples (as we are doing here). Now you can take some samples from this population and look at the relationship between the sample statistics and the population parameters for this distribution.

The plastic bag contains all the data from the population, but it's not necessarily in any random state. First, figure out how you would ensure that your samples are fairly random samples from the population. Then conduct the sampling as indicated below. Be sure to sample with replacement (record the score you draw from the population, then return it so that it might get sampled again).

Take a sample of  $n = 5$  and record the mean, variance, and standard deviation below:

	X	X <sup>2</sup>
Sum		
Mean ( $\bar{X}$ )		
Sum of Squares		
Variance ( $s^2$ )		
Standard Deviation ( $s$ )		

$$SS = \sum X^2 - \frac{(\sum X)^2}{n} =$$

$$s^2 = \frac{SS}{n-1} =$$

$$s = \sqrt{s^2} =$$

Take a sample of  $n = 16$  and record the mean, variance, and standard deviation below:

	X	X <sup>2</sup>
Sum		
Mean ( $\bar{X}$ )		
Sum of Squares		
Variance ( $s^2$ )		
Standard Deviation ( $s$ )		

$$SS = \sum X^2 - \frac{(\sum X)^2}{n} =$$

$$s^2 = \frac{SS}{n-1} =$$

$$s = \sqrt{s^2} =$$

Which sample seems to provide a more accurate estimation of population parameters? Why might that be the case?

4. Sampling from an “unknown” population (playing with your food)

As you know, it’s much more common for the population parameters to be entirely unknown. Have you ever wondered about the distribution of candy colors? It turns out that the companies actually have production data on their colors, so that it is possible to “know” the population parameters.

Population information about M&Ms (and maybe Skittles) can be found online. Just go to their web sites.

But presume that you couldn’t just use the web to determine the distribution of colors in the population. How would you go about estimating the population parameters? Estimation from samples, of course!

First, let’s sample from your sample. Close your eyes and pick a piece of candy from your bag. Record the color of the piece and place it back in the bag. Then repeat the procedure a total of 10 times. Record the colors below. How often did you get a red piece of candy?

--	--	--	--	--	--	--	--	--	--

Now you need to count the candy pieces in your bag, including a count of the number of red pieces that are present.

Red pieces	
Total number of pieces	
Proportion of red pieces	

What is the probability that you would randomly sample a piece of red candy?

Is that probability consistent with the blind picks you made earlier? Why or why not?
