Intro to Statistics Classwork/Homework Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Variance and Standard Deviation in Samples

Warm-up: If you purchased 6 cans of Brand A paint to study fading times of the paint brand, do you have a population or a sample of Brand A paint?



**Why do statisticians use a different formula for the variance of a sample than they do for a population?**

Samples tend to be very small compared to the population under study. The formula for population variance, when applied to a sample, usually underestimates the variance actually found for the full population.

Therefore, statisticians use a formula that provides a slightly larger estimation of variance than the one used for population. The formula for populations is:

An equivalent formula for samples would replace σ with s and N with n. What symbol would replace μ? Make these chances to the formula above.

What change to the **denominator** of a fraction increases the **value** of the fraction?

Instead of dividing by n, the size of the sample, statisticians divide by (n – 1) to obtain the variance. The standard deviation in a sample is obtained by taking the square root of the sample variance. Make that change above, as well.

Find the variance and standard deviation for Brand A if your six cans of paint are a sample, not a population. (Hint: you’ve already done most of this work on the last set of notes).

There is a shortcut formula for sample variance, shown below, which is mathematically equivalent to the formula we used to define sample variance. Why might this be faster?



**Try it:** use the shortcut formulas:

****

What are the steps for using the shortcut?

**Coefficient of Variation**

Sometimes you will want to compare data sets that used different units, such as the number of sales per salesperson and the commissions made by the same salespeople, to determine which data have more variation. However, standard deviation has the units of the data set. To compare data with different units, we will use the **coefficient of variation.**



Try it:



**Range Rule of Thumb**



If your data are unimodal and roughly symmetric, the range rule of thumb can be used to estimate the standard deviation quickly. This is especially useful in checking your work. Go back to Brand A. What does the range rule of thumb predict? Does this match the results of your other calculations?

Classwork/homework (you’re welcome to do this on a separate page so as to retain your notes): Solve problems 7-10:

