**Lab on Standard Deviation by Paul Kinion**

Introduction: We have defined three different concepts for standard deviation. The population standard deviation σ, the biased estimator $s\_{n}$, and the sample standard deviation $ s\_{n-1}$ or just $s$, also known as the unbiased estimator. Clearly, when the sample size n is the same as the population size N, the biased estimator $s\_{n}$ is the same as the parameter σ. So, where is the bias, why use $s\_{n-1}$? This lab will give a statistical justification for the practice.

Purpose: To show that often, $s\_{n-1}$ is a better predictor of σ, than $s\_{n}$.

Method: Our population consists of a class of 32 students. On a scale of 1 to 5 they responded (7, 12, 4, 0, 9) to a survey question. This means 7 answered “1”, 12 answered “2” and so on. Ten times we will randomly sample 16 or half of the population and record the sample standard deviation $s\_{n-1}$. A simple calculation gives $s\_{n}$ . Of the ten trials, we will record the proportion of times $s\_{n}$ was at least as good of a predictor as $s\_{n-1}$. The sample proportion will be used to estimate the population’s proportion. Measurements will be taken with *Sampling Distributions for Small Samples* by Kinion and Haxton.

Hypothesis: Let π be the proportion of trials where $s\_{n}$ is at least as good as a predictor of σ as $s\_{n-1}$. The null hypothesis is π ≥ 0.5, leaving π < 0.5 as the alternative. Use a 0.10 level of significance.

Question: Out of ten trials, how few successes would you tolerate before rejecting the null hypothesis and adopting$ s\_{n-1}$ as the predictor that is most often the best predictor of the population standard deviation? Explain

Step 1: Open *Sampling Distributions for Small Samples*. Enter the population as described and 16 for the sample size. Generate data.

Step 2: The top graph is the population. Hover the mouse over it and right click. Record population parameters μ and σ to the nearest ten-thousandth on page two of the data sheet.

Step 3: The middle graph is a random sample. Hover the mouse over it and right click. Record the first sample standard deviation$s\_{n-1}$. Close the sample window and push the “Generate Data” button. Record the second standard deviation. Repeat until you have ten sample standard deviations recorded.

Step 4: Calculate the conversion factor F = $ \sqrt{\frac{n-1}{n}}$ Use the formula $s\_{n}$ = F$ s\_{n-1}$ to complete the second column in the data sheet.

Step 5: Both predictors perform the same when σ - $s\_{n}$ = $s\_{n-1}$ - σ. Make the substitution $s\_{n}$ = F$ s\_{n-1}$ and solve the resulting equation for $s\_{n-1}$. Did you obtain 2σ/( 1 + F)? Calculate and record on the data sheet. This value serves as a boundary. If $s\_{n-1}$ is greater than or on the boundary, then $s\_{n}$ is the better predictor, or at least as good. Check each qualifying trial on page one of the data sheet. Divide the number of qualifiers by 10 to obtain the sample proportion $\hat{p}$.

Step 6: Find the critical value for a hypothesis test done with a 0.10 level of significance.

Step 7: Find the test statistic $z= \frac{\hat{p}-0.5}{\sqrt{\frac{0.25}{n}}}$.

Step 8: Use a complete sentence to state the conclusion of your hypothesis test.

Step 9: Repeat Steps 1 through 8 with a Population of your own choosing

Data Sheet Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Trial Number $s\_{n-1}$ $s\_{n}$ Check if $s\_{n-1}\geq B$

 1 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 2 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 3 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 4 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 5 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 6 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 7 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 8 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 9 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 10 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

Population \_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_,

Population Size (N) \_\_\_\_\_\_\_\_\_\_\_

Sample Size (n) \_\_\_\_\_\_\_\_\_\_\_

Population Mean (μ) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Population Standard Deviation (σ) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Conversion Factor (F) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Boundary Value (B) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Sample proportion ($\hat{p}$) \_\_\_\_\_\_\_\_\_\_\_\_\_

Test Statistic (z) \_\_\_\_\_\_\_\_\_\_\_\_\_

Conclusion

Data Sheet Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Trial Number $s\_{n-1}$ $s\_{n}$ Check if $s\_{n-1}\geq B$

 1 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 2 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 3 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 4 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 5 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 6 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 7 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 8 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 9 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

 10 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

Population \_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_,

Population Size (N) \_\_\_\_\_\_\_\_\_\_\_

Sample Size (n) \_\_\_\_\_\_\_\_\_\_\_

Population Mean (μ) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Population Standard Deviation (σ) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Conversion Factor (F) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Boundary Value (B) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Sample proportion ($\hat{p}$) \_\_\_\_\_\_\_\_\_\_\_\_\_

Test Statistic (z) \_\_\_\_\_\_\_\_\_\_\_\_\_

Conclusion