

ST502: Final Project - Part 2

Apostolos Stamenos & Tyler Pollard

4/19/2022

In the previous section, we concluded that the t-test with the Satterthwaite approximation is more appropriate when there is no evidence that the population variances are equal. In order to understand the properties of both tests and to verify that our conclusion makes sense, we conducted a simulation study. A numerical simulation study can be used to evaluate the two testing procedures because unlike in a real data analysis, we know exactly what the true population parameters are. By generating enough datasets through a data-generating mechanism that we control, we can evaluate how the two testing procedures perform when drawing inferences about the underlying populations.

For this simulation study, two random samples of data were generated from normal distributions with varying population characteristics outlined in Table 1. All combinations of these population characteristics were used to generate data for a total of 135 unique combinations. Any data that was generated for a true mean difference of 0 correspond to data from a null situation and the rest of the generated data correspond to data from an alternative situation. For a given combination of population characteristics, 100 unique data sets of two samples were randomly generated from their corresponding normal distribution. Both two-sample t-tests when equal variance is assumed (pooled) and when unequal variances are assumed (with Satterthwaite approximation for the degrees of freedom) were conducted, using the p-value method, on each of the 100 randomly generated data sets for all 135 combinations of population characteristics. The p-value for each of the randomly generated data sets was compared to the set significance level of $\alpha = 0.05$. If the p-value was less than alpha then the test would reject the null hypothesis of equal population means. The proportion of times we rejected the null hypothesis was used to estimate alpha when the data was generated from a null situation and power when the data was generated from an alternative situation.

Table 1: **Varying Population Characteristics**

Population Characteristics	Values
True Variance	$\sigma_1^2 = 1, 4, 9$ and $\sigma_2^2 = 1$
Sample Size	$n_1 = 10, 30, 70$ and $n_2 = 10, 30, 70$
True Mean Difference	$\Delta = \mu_1 - \mu_2 = -5, -1, 0, 1, 5$

The simulated alpha value curves corresponding to the Type I error rate for the assumed equal variance and assumed unequal variance tests, in Figure 1, were approximately equal when the variance of both populations were equal (i.e., $\sigma_1^2 = \sigma_2^2$), regardless of the sample sizes. This makes sense because the equal variance assumption holds and the tests should be expected to produce similar results. When the sample sizes are the same (i.e., $n_1 = n_2$), the simulated alpha values of the two tests are also approximately equal, regardless of the varying σ_1^2 . For $n_1 > n_2$ and $\sigma_1^2 \neq \sigma_2^2$, the simulated alpha values for the pooled variance t-test are less than that of the Satterthwaite t-test and approach 0 as σ_1^2 increases, whereas for $n_1 < n_2$, the alpha values for the pooled variance t-test are much larger than that of the Satterthwaite t-test as σ_1^2 increases. Since we used the significance level of $\alpha = 0.05$ in our rejection criteria, we should expect the simulated Type I error rate to be close to 0.05. Regardless of the population characteristics used to generate the data, the assumed unequal variance t-test performed well, estimating the Type I error rate to be within ± 0.05 of the expected value of 0.05. The assumed unequal variance test performed well when $\sigma_1^2 = \sigma_2^2$ or $n_1 = n_2$, however when $\sigma_1^2 \neq \sigma_2^2$ and $n_1 \neq n_2$, the test performed poorly, estimating the Type I error rate to be far from the expected value.

Regardless of n_1 and n_2 , the simulated power of both the pooled and Satterthwaite t-tests is greater for smaller values of σ_1^2 . This result makes sense intuitively. When the variance of one population is small, it is easier for the hypothesis testing procedure to detect the true mean difference. When the sample sizes are the same (i.e., $n_1 = n_2$), the simulated powers of the pooled and Satterthwaite tests are approximately equal, regardless of σ_1^2 . For $n_1 > n_2$, the Satterthwaite t-test tends to perform better, whereas for $n_1 < n_2$, the pooled variance t-test tends to perform better. As n_1 and n_2 increase, so does the power of both the pooled and Satterthwaite t-tests. This makes sense because for both tests, the standard error (i.e., the denominator of the test statistic) gets smaller and the degrees of freedom get larger for large sample sizes.

Figure 1: **Approximate alpha curves of the Pooled and Satterthwaite t-tests for different sample sizes and population variances**

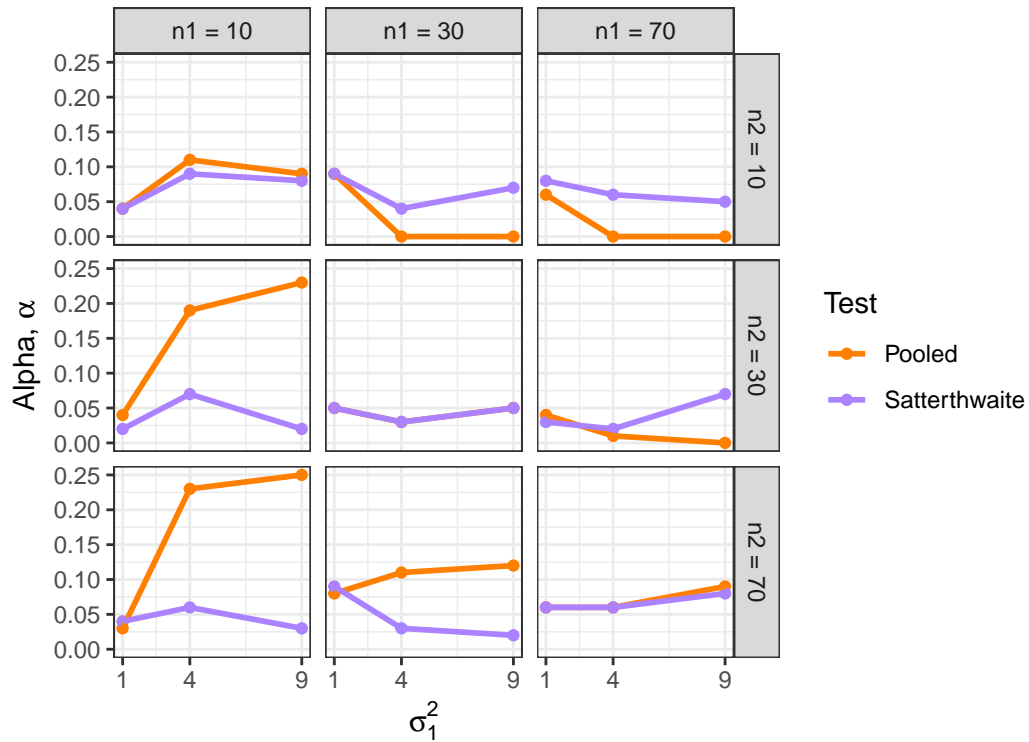
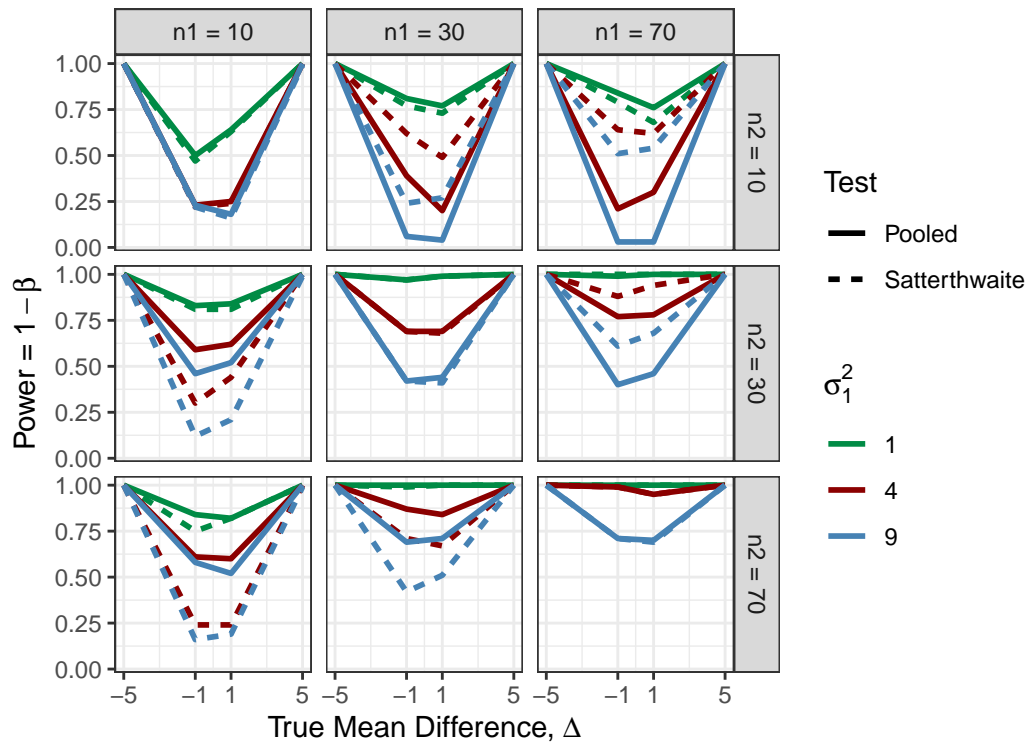


Figure 2: **Approximate power curves of the pooled and Satterthwaite t-tests for different samples sizes and population variances**



Discuss conclusions based on plots and other theoretical arguments (Apostolos)

Need a conclusion paragraph

Appendix A: R Code

Team Member Contributions

These reports are the result of extensive collaboration facilitated by Zoom meetings and GitHub. Both group members took turns coding and writing the report. We had long discussions to figure out the best way to conduct the hypothesis tests and simulation study, and how to effectively present our findings.

Apostolos: performed the pooled variance t-test, wrote the code for many of the plots comparing smokers and nonsmokers, wrote the section assessing the normality assumption, wrote the section on testing the equality of variances, wrote a function to generate datasets for the simulation, wrote the section justifying the use of a simulation study, and wrote the section discussing the power of the tests.

Tyler: managed pull requests to the GitHub repository, performed the Satterthwaite t-test, wrote a function to calculate the probability of rejecting H_0 for the simulation, wrote code to calculate and process the simulated values of α and power, wrote the section discussing the design of the simulation study, and wrote the section discussing the Type I error rate of the tests.