Group 5, R Project 3

Kera Whitley, Laura Mathews, Nataliya Peshekhodko

2020-April-24

Purpose

Hypothesis testing is used to conduct statistical inference on hypotheses to determine if the data shows a consistency between different hypotheses regarding a parameter. Hypothesis testing can be conducted using a variety of test statistics and distributions, including the t distribution. These t-tests may be done in two different ways, depending on assumptions made about the variance of each sample. The pooled t-test assumes that the sample variances are equal, while the unpooled t-test assumes the sample variances are not equal.

Regardless of the type of hypothesis test, the significance level and power should be maximized to obtain the most effective statistical inference. In order to assess the effectiveness of a t test at controlling alpha and obtaining the greatest possible power, simulations may be used. Empirical values of alpha and power are calculated from simulated sample data under different sampling situations. The empirical values can be determined to evaluate the best test to be used on real sample data.

The purpose of this study is to compare the effectiveness of the pooled and unpooled t-tests at controlling alpha and maximizing power under a variety of situations.

Design

For each simulation, N = 10000 data sets were generated for the null and alternative hypotheses. A t-test was used to assess whether each trial would be accepted or rejected. The empirical alpha and power were calculated using the proportion of times the null was rejected. This simulation was done with both the pooled and non-pooled t-test. The result of each combination of variables was plotted to compare the use of the unequal and equal variance tests under different situations. Simulations were run for each combination of the following variances, sample sizes and mean differences; a total of 135 simulations for each test.

σ_1^2 = 1,3,9 n1 = 10, 25, 60 $\mu_1 - \mu_2$ = -5, -1, 0, 1, 5
σ_2^2 = 1 n_2=10,25,60
N=10000

A function for power calculation was written to most efficiently make calculations and manipulate parameter values.

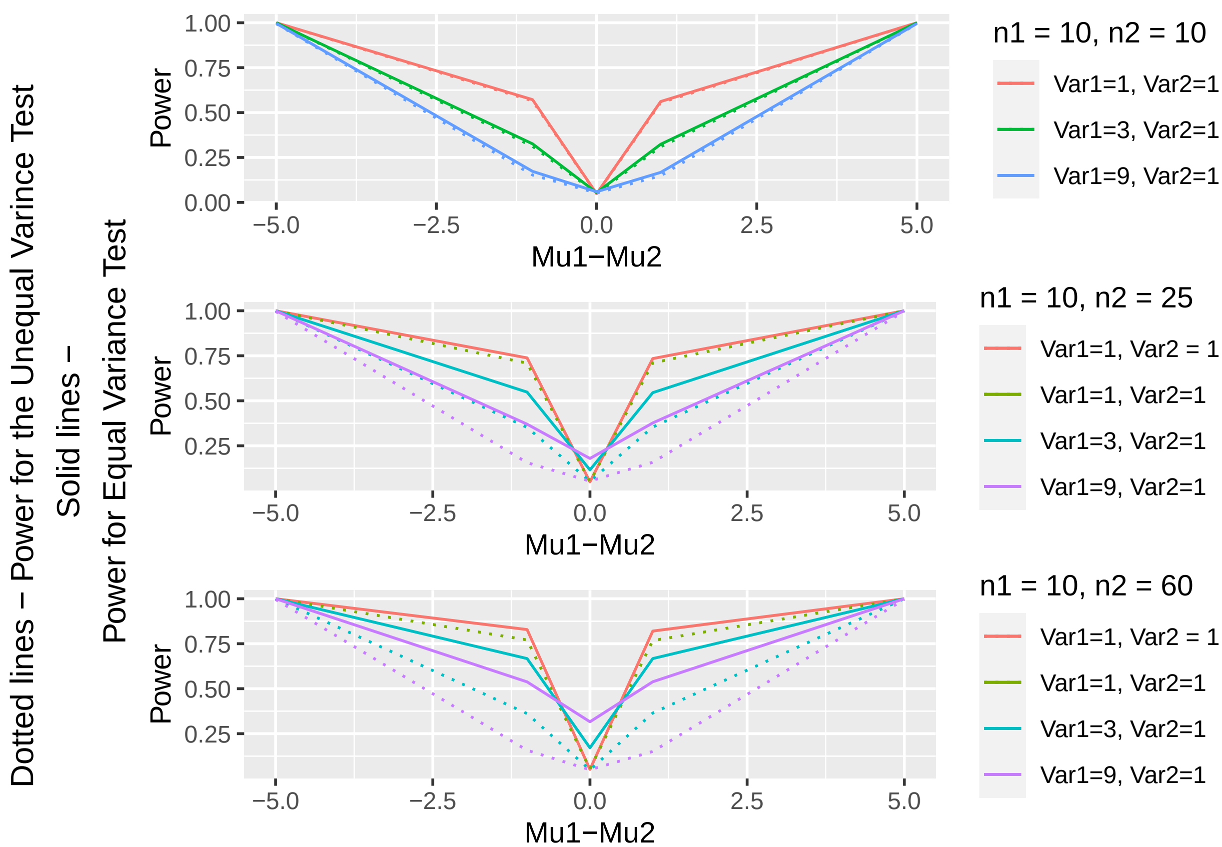
Results

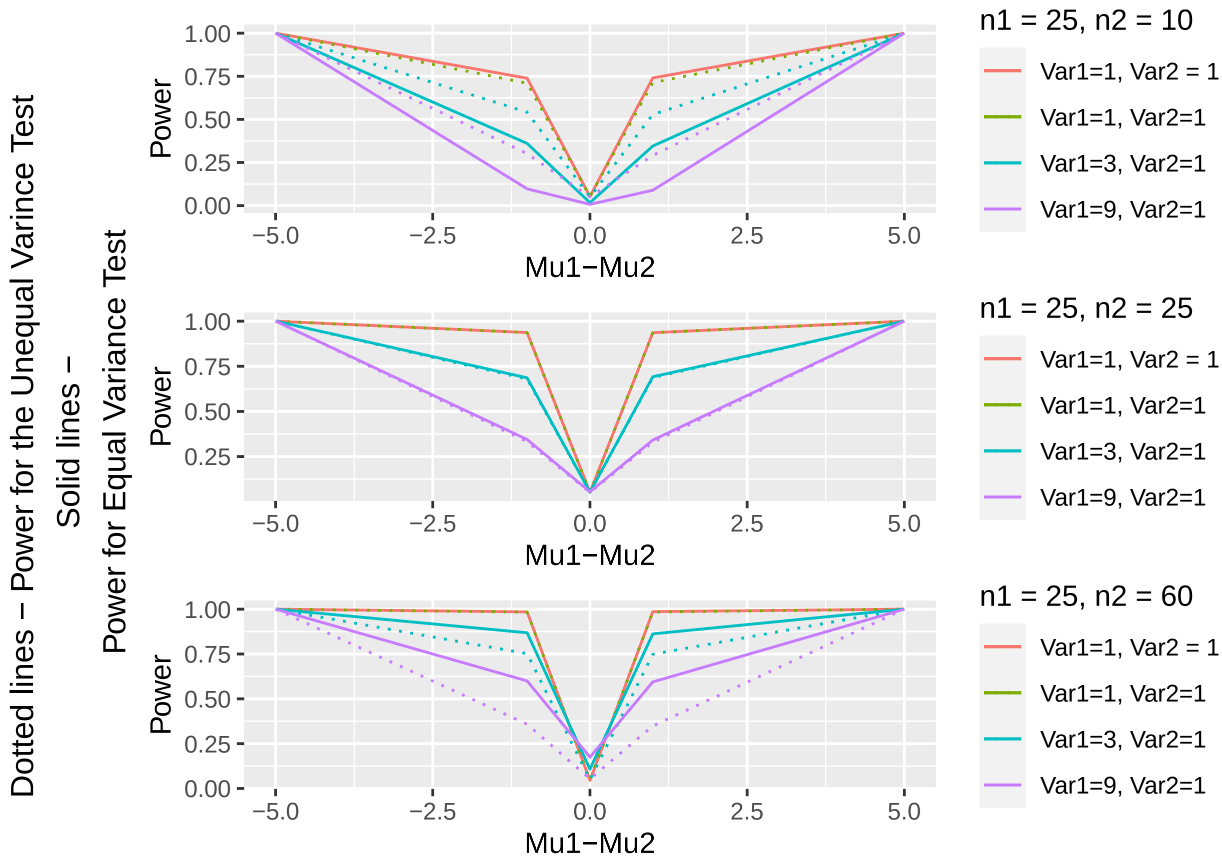
Using the plots generated in the simulation to compare sample sizes, the pooled and unpooled t-tests resulted in very similar power calculations when the sample sizes were equal. The tests also produced similar results when the sample variances were equal. This was to be expected as the only difference between each test would be the degrees of freedom used for the t distribution. It appears that after controlling for difference in µ and sample sizes, a greater difference in true sample variation produces a larger difference in power calculation between the pooled and unpooled t-tests. The largest difference in power calculation, when controlling for sample sizes and sample variances, appears to have been produced by the mu difference ±1. Because the t distribution is symmetric, the t-tests produced the same power values for the mu differences +1 and -1, and similarly for +5 and -5. This indicates that which sample is designated µ1 and which is µ2 is irrelevant.

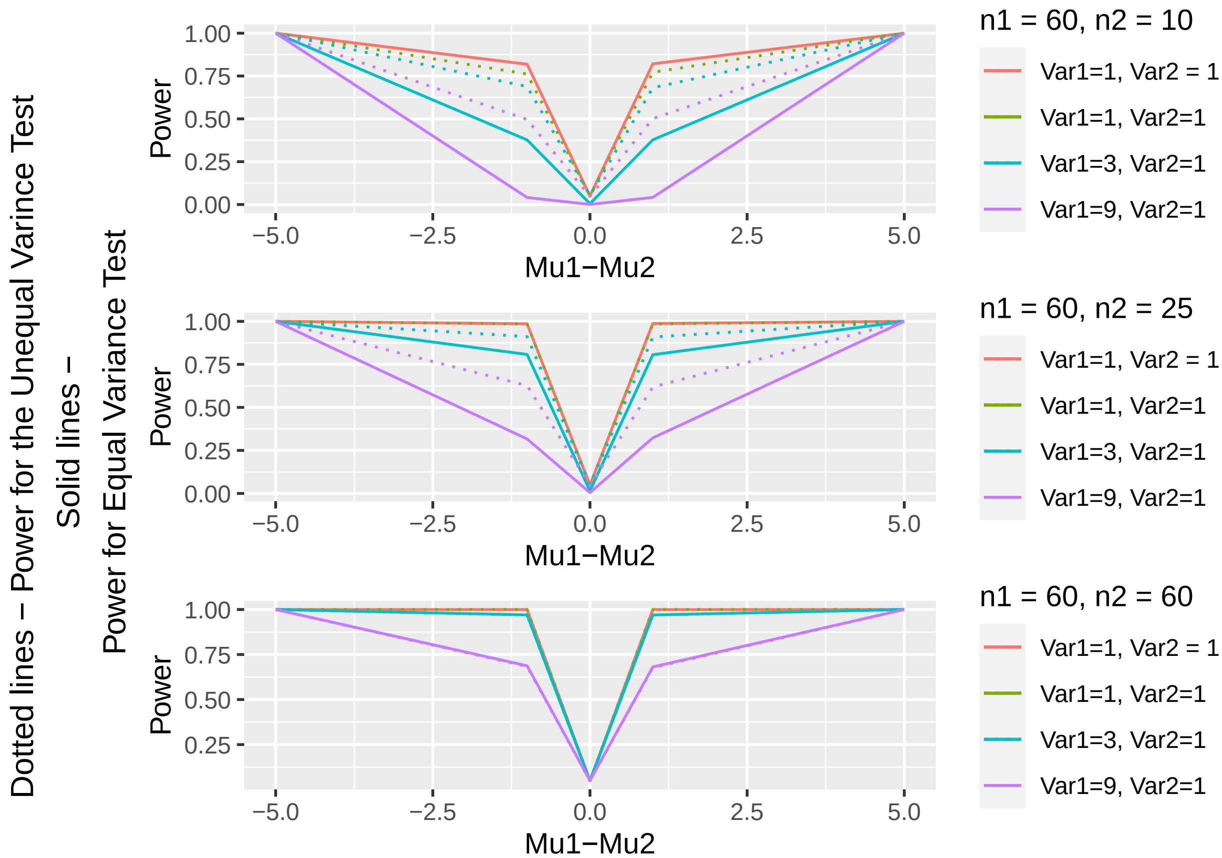
After controlling for sample sizes and sample variances, the power increases as the absolute value of the mean difference increases. A difference in sample means indicates a shift in the location of the distribution of each sample. A larger difference will have less overlap between the two sample distributions and thus will be easier to detect. This is consistent for both the pooled and unpooled t-tests.

The power is consistently higher for the unpooled t-test than for the pooled t-test across all simulations in which n1 is greater than n2, but lower for simulations in which n1 is less than n2. The sample variance was only increased for simulations with sample 1. The variance was constantly 1 for sample 2.

Finally, as the total number of samples (n1 + n2) increases, so does the power of both t-tests. It is particularly easy to see this result in the cases where mu difference is +-1. When using a given desired power and alpha level to calculate the sample size needed to produce such a test, the sample size and critical alpha and power values are directly proportional. This calculation explains the increase in power as total number of samples increases.




Dotted lines – Power for the Unequal Varince Test
Solid lines –
Power for Equal Variance Test
n1 = 25, n2 = 10
n1 = 25, n2 = 25
n1 = 25, n2 = 60
Power
Mu1−Mu2
Var1=1, Var2 = 1
Var1=1, Var2=1
Var1=3, Var2=1
Var1=9, Var2=1
1.00
0.75
0.50
0.25
0.00
−5.0
−2.5
0.0
2.5
5.0



Conclusion

TThrough these simulations, it was determined that the pooled t-test produces a higher power for simulations in which the sample variances were equal or n1 was less than n2. This may indicate that a larger sample size is better for the sample with the larger variance. As the difference in true sample variances increases, so too did the difference in the power value obtained by the pooled and unpooled t-test. Intuitively, this result makes sense as one would expect a smaller difference in true sample variance to be of less importance than a larger difference when determining which test to use. These simulations show power responds to the use of different t-tests under different sample situations. The results obtained in this study can be used to appropriately choose which test to use to maximize power when given a set of parameters. In general, it appears that the unpooled t-test is important to use when the projected difference in population variances is believed to be larger. Different sample sizes also may require different the use of a different t-test.