Group 5, R Project 3

Kera Whitley, Laura Mathews, Nataliya Peshekhodko

2020-April-23

Purpose

The goal of this simulation is to analyze performance of the two tests in terms of controlling α and having larger power while varying true variances, sample sizes and true means differences. We got empirical power by simulation as a function of different combinations of μ differences, different variances and sample sizes.

Design

During this simulation next set of combinations were used:

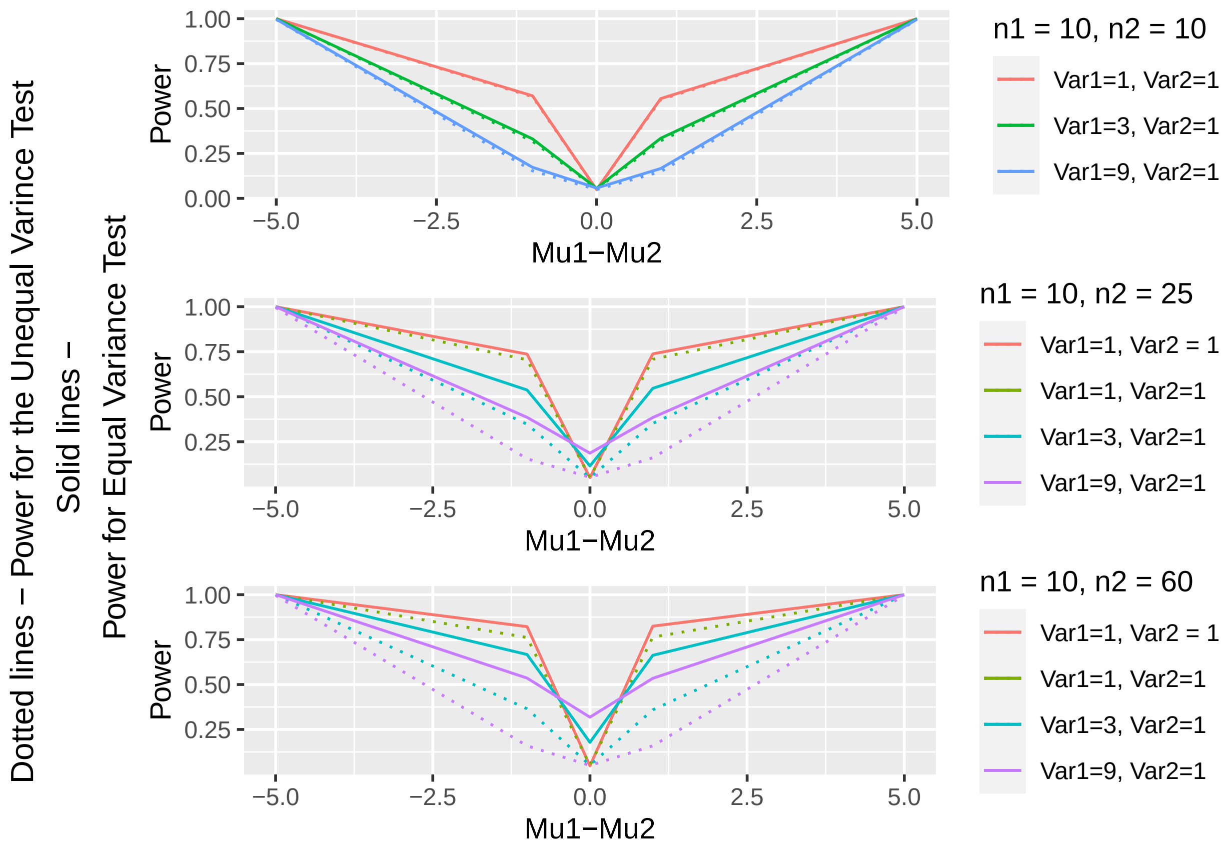
$\sigma_1^2 = 1{,}3{,}9$ $\sigma_2^2 = 1$
$n_1 = 10$, 25, 60 n_2=10,25,60
$\mu_1 - \mu_2 =$ -5, -1, 0, 1, 5
N=10000

Under each combination we generated a sample under H_A N times. Then we performed t.tests for equal and unequal variances and calculated number of times H_0 is rejected. Empirical power for both tests is calculated and analyzed.

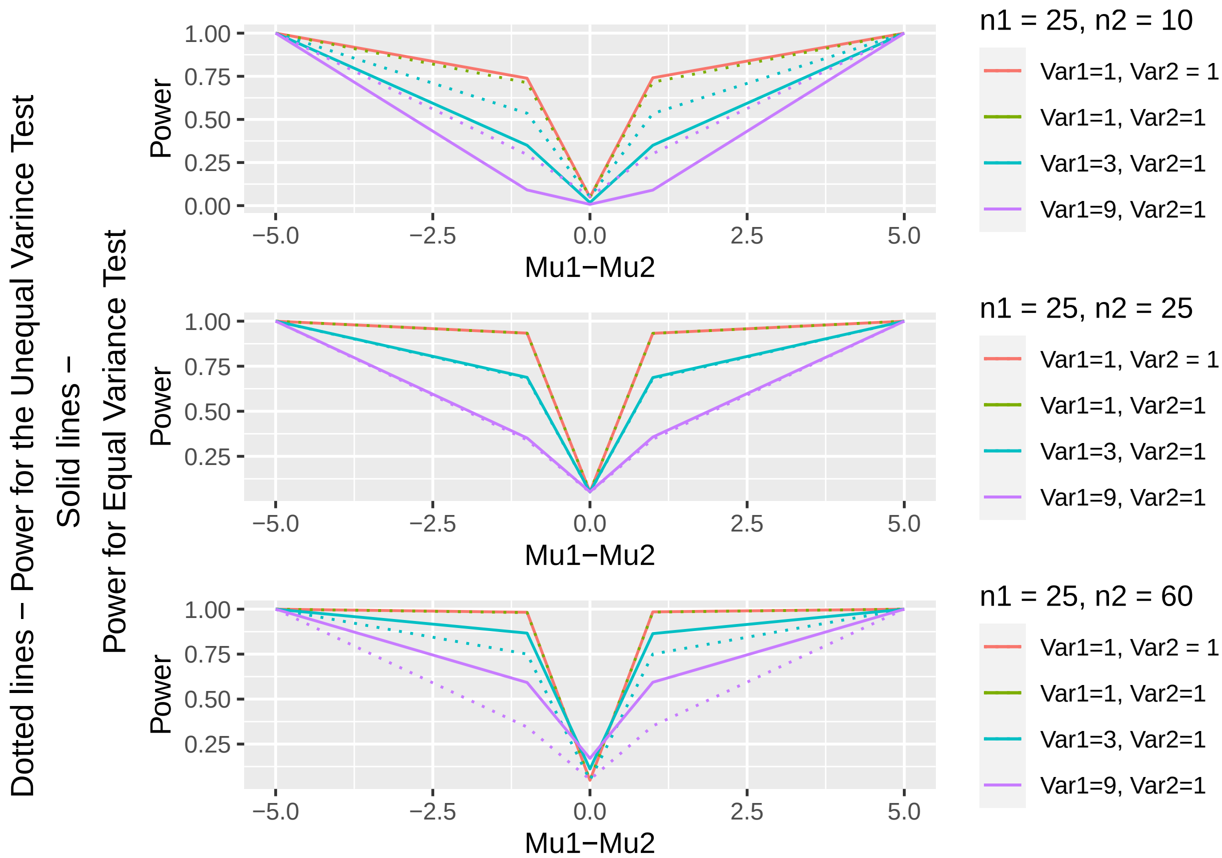
Since we need to simulate data as a function of different combinations of several parameters and make some manipulations with this data, a function for power calculation is created. This function takes a parameter sample sizes n_1 and n_2. We used while creating data for different combinations.

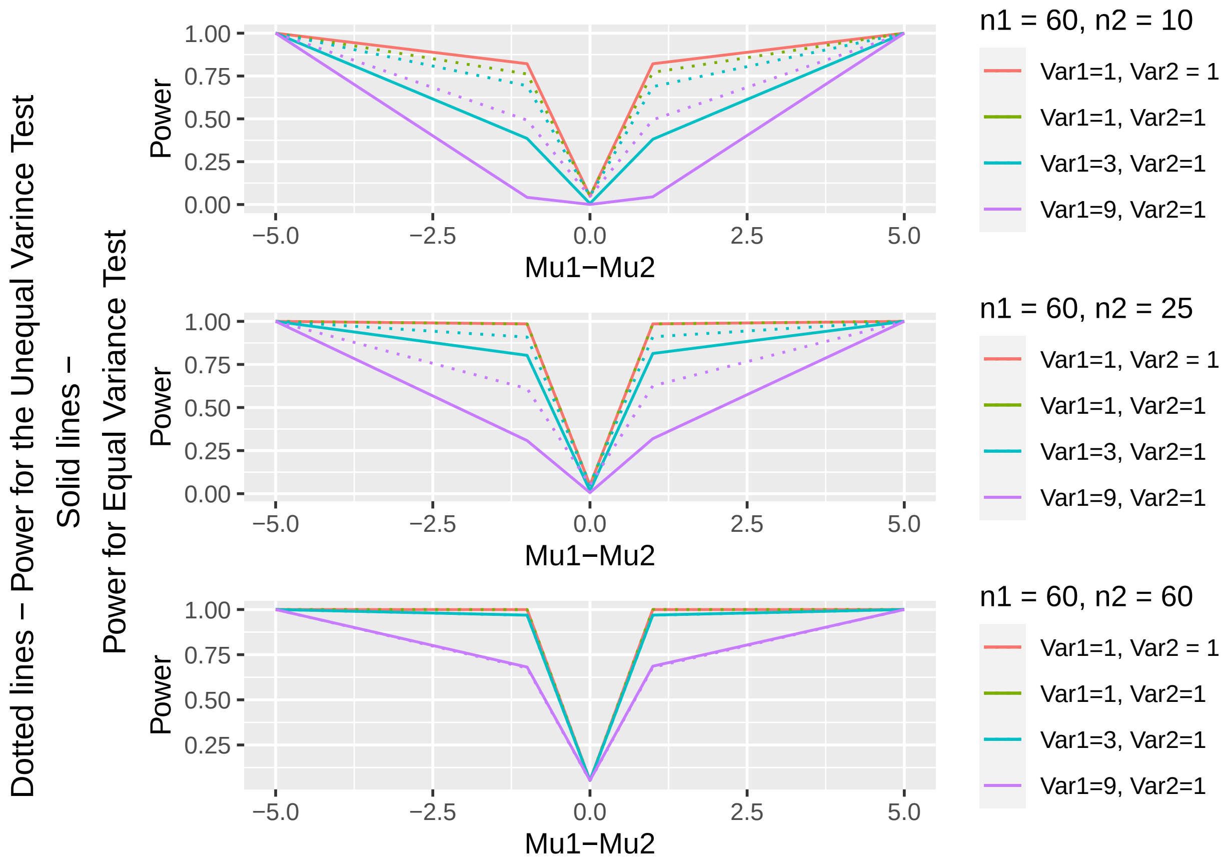
Results

Visualize different combinations of power as function of sample sizes, sample veriances and mean differences.



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


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



Conclusion

The results from the power analysis indicate that the t.tests take into account the absolute value of the difference in means between the two sample populations both when the variances are equal and not equal. This is shown by the fact that each of the graphs is symmetrical around mu1-mu2=0. The relationship between the difference in means and the difference in variance is that as the difference in variances between the sample populations increases it takes and larger absolute value of the difference in means to raise the power. This can be seen for both the equal variance t.test and the unequal t.test, though the uneual test does increase the power faster than the equal one. The impact that sample size has on the t.tests are that as the sample size increases so does the rate in which power is increased. The tests for both the equal and unequal variances give very similar results when the sample size, n, is equal. The reason for this is that when the n's are equal the fraction for the denominator for the t.tests are much closer to each other as in equal variance: $\sqrt{(sp^2/n_1) + (sp^2/n_2)} => \sqrt{\frac{(s_1^2+s_2^2)}{2n}}$ and unequal variance: $\sqrt{(s_1^2/n_1) + (s_2^2/n_2)} => \sqrt{(s_1^2 + s_2^2)/2n)}$ which means that the denominator is equal for each of these. So now the variation we see in the plots for these comparisons are from the variability in the data sets generated with these conditions.