# Appendix

**Robustness of Student’s t-test when sample sizes are equal across groups**

Simulations show that Student’s *t*-test is strongly robust to violations of equal variances assumption, even for big SDR, as long as sample sizes are the same between groups. Whatever the SDR, Student’s *t*-test and Welch’s *t*-test remain close to the nominal Type 1 error rate (see Table A1).

**Table A1**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Student's t-test | | | Welch's t-test | | |
| Nominal size of alpha risk | 1% | 5% | 10% | 1% | 5% | 10% |
| SDR |  |  |  |  |  |  |
| 1,2 | 0,01 | 0,0502 | 0,1004 | 0,0099 | 0,05 | 0,1003 |
| 1,4 | 0,0102 | 0,0504 | 0,1002 | 0,01 | 0,0501 | 0,0998 |
| 1,6 | 0,0101 | 0,0501 | 0,1 | 0,0099 | 0,0496 | 0,0994 |
| 1,8 | 0,0103 | 0,0504 | 0,0999 | 0,0099 | 0,0496 | 0,0991 |
| 2 | 0,0106 | 0,0512 | 0,1014 | 0,0102 | 0,0502 | 0,1003 |
| 2,2 | 0,0107 | 0,051 | 0,1014 | 0,0102 | 0,0499 | 0,1002 |
| 2,4 | 0,0106 | 0,0512 | 0,1012 | 0,01 | 0,0499 | 0,0997 |
| 2,6 | 0,0108 | 0,0516 | 0,1016 | 0,0101 | 0,0501 | 0,0999 |
| 2,8 | 0,0108 | 0,0514 | 0,1016 | 0,0101 | 0,0499 | 0,0998 |
| 3 | 0,0109 | 0,0519 | 0,1023 | 0,01 | 0,0501 | 0,1004 |
| 3,2 | 0,011 | 0,0519 | 0,102 | 0,01 | 0,0501 | 0,1001 |
| 3,4 | 0,0111 | 0,0518 | 0,1019 | 0,0101 | 0,05 | 0,0998 |
| 3,6 | 0,011 | 0,0523 | 0,1025 | 0,01 | 0,0503 | 0,1003 |
| 3,8 | 0,011 | 0,052 | 0,1023 | 0,01 | 0,05 | 0,1001 |
| 4 | 0,0109 | 0,0517 | 0,1019 | 0,0098 | 0,0497 | 0,0996 |
| 4,2 | 0,0111 | 0,0523 | 0,1029 | 0,01 | 0,0502 | 0,1004 |
| 4,4 | 0,0111 | 0,0524 | 0,1024 | 0,01 | 0,0502 | 0,1 |
| 4,6 | 0,011 | 0,0519 | 0,1024 | 0,0099 | 0,0497 | 0,0998 |
| 4,8 | 0,0111 | 0,0522 | 0,1028 | 0,01 | 0,05 | 0,1002 |
| 5 | 0,0114 | 0,0523 | 0,1026 | 0,0102 | 0,05 | 0,1001 |
| 5,2 | 0,0111 | 0,0524 | 0,1025 | 0,0099 | 0,0501 | 0,0999 |
| 5,4 | 0,0109 | 0,0519 | 0,1022 | 0,0098 | 0,0495 | 0,0995 |
| 5,6 | 0,0112 | 0,0525 | 0,1029 | 0,01 | 0,0502 | 0,1002 |
| 5,8 | 0,0111 | 0,0517 | 0,1022 | 0,0099 | 0,0495 | 0,0995 |
| 6 | 0,0112 | 0,0523 | 0,1029 | 0,01 | 0,0499 | 0,1002 |
| 6,2 | 0,0111 | 0,0521 | 0,103 | 0,0099 | 0,0498 | 0,1002 |
| 6,4 | 0,0113 | 0,0526 | 0,1027 | 0,01 | 0,0502 | 0,1 |
| 6,6 | 0,0113 | 0,0524 | 0,1029 | 0,0101 | 0,05 | 0,1001 |
| 6,8 | 0,0112 | 0,0521 | 0,1028 | 0,0099 | 0,0496 | 0,1 |
| 7 | 0,0114 | 0,0526 | 0,1026 | 0,0101 | 0,0502 | 0,0999 |
| 7,2 | 0,0112 | 0,0521 | 0,1026 | 0,01 | 0,0497 | 0,0999 |
| 7,4 | 0,0112 | 0,0524 | 0,1027 | 0,01 | 0,05 | 0,0998 |
| 7,6 | 0,0113 | 0,0528 | 0,103 | 0,01 | 0,0503 | 0,1002 |
| 7,8 | 0,0113 | 0,0526 | 0,1029 | 0,01 | 0,0501 | 0,1 |
| 8 | 0,0111 | 0,0526 | 0,1029 | 0,0098 | 0,0501 | 0,1001 |
| 8,2 | 0,0115 | 0,0528 | 0,1032 | 0,0102 | 0,0503 | 0,1003 |
| 8,4 | 0,0112 | 0,0525 | 0,1029 | 0,0099 | 0,05 | 0,1001 |
| 8,6 | 0,0112 | 0,0522 | 0,1026 | 0,0098 | 0,0496 | 0,0996 |
| 8,8 | 0,0115 | 0,0528 | 0,1028 | 0,0101 | 0,0503 | 0,0999 |
| 9 | 0,0112 | 0,0523 | 0,1027 | 0,0099 | 0,0498 | 0,0998 |
| 9,2 | 0,0112 | 0,0526 | 0,1025 | 0,01 | 0,0501 | 0,0997 |
| 9,4 | 0,0113 | 0,0522 | 0,1026 | 0,01 | 0,0497 | 0,0997 |
| 9,6 | 0,0112 | 0,0525 | 0,103 | 0,0099 | 0,05 | 0,1001 |
| 9,8 | 0,0113 | 0,0528 | 0,1032 | 0,01 | 0,0502 | 0,1002 |
| 10 | 0,0112 | 0,0524 | 0,1025 | 0,0099 | 0,0498 | 0,0996 |

**Standard deviations in samples and in populations**

To check if the sample SD show the same pattern as the population SD, we simulated 1.000.000 sets of samples for each sample size from 10 to 85 under 5 population SD: 1.84, 0.92, 1.11, 1.32 and 1.63. Because the SD sampling distribution was quite close to a normal distribution, the mean and standard deviation are good estimators of the quality of the estimation by the sample SD (see Table A2).

**Table A2**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| POPULATIONS  SD | | 0.92 | | 1.11 | | 1.32 | | 1.63 | | 1.84 | | |
| **N** | Mean | | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| **10** | 0,895 | | 0,214 | 1,080 | 0,258 | 1,284 | 0,307 | 1,586 | 0,378 | 1,79 | 0,428 |
| **15** | 0,904 | | 0,172 | 1,090 | 0,208 | 1,297 | 0,247 | 1,601 | 0,305 | 1,807 | 0,345 |
| **20** | 0,908 | | 0,148 | 1,095 | 0,179 | 1,303 | 0,212 | 1,609 | 0,263 | 1,816 | 0,296 |
| **25** | 0,91 | | 0,132 | 1,098 | 0,159 | 1,306 | 0,19 | 1,613 | 0,234 | 1,821 | 0,264 |
| **30** | 0,912 | | 0,12 | 1,101 | 0,145 | 1,309 | 0,173 | 1,616 | 0,213 | 1,824 | 0,241 |
| **35** | 0,913 | | 0,111 | 1,102 | 0,134 | 1,31 | 0,159 | 1,618 | 0,197 | 1,826 | 0,222 |
| **40** | 0,914 | | 0,104 | 1,103 | 0,125 | 1,312 | 0,149 | 1,619 | 0,184 | 1,829 | 0,208 |
| **45** | 0,915 | | 0,098 | 1,104 | 0,118 | 1,312 | 0,14 | 1,621 | 0,173 | 1,83 | 0,195 |
| **50** | 0,915 | | 0,093 | 1,104 | 0,112 | 1,313 | 0,133 | 1,622 | 0,164 | 1,831 | 0,185 |
| **55** | 0,916 | | 0,088 | 1,105 | 0,107 | 1,314 | 0,127 | 1,622 | 0,157 | 1,831 | 0,177 |
| **60** | 0,916 | | 0,084 | 1,105 | 0,102 | 1,314 | 0,121 | 1,623 | 0,15 | 1,832 | 0,169 |
| **65** | 0,916 | | 0,081 | 1,106 | 0,098 | 1,315 | 0,116 | 1,623 | 0,144 | 1,833 | 0,162 |
| **70** | 0,917 | | 0,078 | 1,106 | 0,094 | 1,315 | 0,112 | 1,624 | 0,138 | 1,833 | 0,156 |
| **75** | 0,917 | | 0,076 | 1,106 | 0,091 | 1,316 | 0,108 | 1,624 | 0,134 | 1,834 | 0,151 |
| **80** | 0,917 | | 0,073 | 1,106 | 0,088 | 1,316 | 0,105 | 1,625 | 0,13 | 1,834 | 0,146 |
| **85** | 0,917 | | 0,071 | 1,107 | 0,086 | 1,316 | 0,102 | 1,625 | 0,126 | 1,834 | 0,142 |

The smaller is the sample size, the further the average standard deviation is from the population standard deviation, and the bigger the dispersion around this average.

# Power of Student’s *t*-test and Welch’s *t*-test when variances are unequal

Figure 2b shows that when variances are equal, but sample sizes are unequal, Welch’s *t*-test incurs a very small loss in power compared to Student’s *t*-test. Table A3 is a complete overview of the exact numbers of power plotted in Figure 2b.

**Table A3**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SDR** | n1 | n2 | SD1= SD2 | power\_Welch (%) | power\_Student (%) |
| **1** | 5 | 95 | 2 | 14.03 | 19.04 |
| **1** | 10 | 90 | 2 | 27.86 | 31.78 |
| **1** | 15 | 85 | 2 | 39.62 | 42.39 |
| **1** | 20 | 80 | 2 | 48.98 | 50.82 |
| **1** | 25 | 75 | 2 | 56.14 | 57.30 |
| **1** | 30 | 70 | 2 | 61.45 | 62.13 |
| **1** | 35 | 65 | 2 | 65.25 | 65.60 |
| **1** | 40 | 60 | 2 | 67.77 | 67.92 |
| **1** | 45 | 55 | 2 | 69.22 | 69.25 |
| **1** | 50 | 50 | 2 | 69.69 | 69.69 |
| **1** | 55 | 45 | 2 | 69.22 | 69.25 |
| **1** | 60 | 40 | 2 | 67.77 | 67.92 |
| **1** | 65 | 35 | 2 | 65.25 | 65.60 |
| **1** | 70 | 30 | 2 | 61.45 | 62.13 |
| **1** | 75 | 25 | 2 | 56.14 | 57.30 |
| **1** | 80 | 20 | 2 | 48.98 | 50.82 |
| **1** | 85 | 15 | 2 | 39.62 | 42.39 |
| **1** | 90 | 10 | 2 | 27.86 | 31.78 |
| **1** | 95 | 5 | 2 | 14.03 | 19.04 |

# Type 1 error rate of Student’s *t*-test and Welch’s *t*-test when variances are unequal

Assuming an Type 1 error rate of 5% under then null, a test can yield either significant result *p*-value < 5%; or a “false positive” -FP) or non-significant result *p*-value > 5%; or a “true negative”-TN).

The specificity is the relative frequency of effects detected as non-significant, under the null:

=

The alpha risk is the complement of the specificity:

=

In order to estimate the Type 1 error rate for Student’s *t*-test and Welch’s *t*-test, we simulated 1.000.000 simulations of two samples generated under the assumption of normality), under 64 different conditions yielding 64.000.000 simulations in total). In each condition, the first sample is generated from a population where σ1=2, and its sample size vary from 10 to 40 in a step of 10. The standard deviation and the sample size of the second sample is a function of the sample size ratio SSR = ; ranging from 0.5 to 2 in step of 0.5) and SDR ranging from 0.5 to 2 in step of .0.5). Consistently with research conducted by Minitab statisticians available at <http://support.minitab.com/en-us/minitab/17/Assistant_Two_Sample_t.pdf>, simulations show that the Type 1 error rate for Student’s *t*-test can differ noticeably from the nominal Type 1 error rate i.e. 5%) when the groups have different variances, while the Type 1 error rate of Welch’s *t*-test remains close of the nominal size i.e. 5%). On the other hand, when the groups have equal variances, Type 1 error rate of both Welch’s *t*-test and Student’s *t*-test remains close of the nominal size (see the grey lines in Table A4).

**Table A4**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **SSR** | **SDR** | n1 | n2 | SD1 | SD2 | alpha\_Welch | alpha\_Student |
| **0,5** | **0,5** | 10 | 5 | 2 | 1 | 0,047 | 0,021 |
| **0,5** | **1** | 10 | 5 | 2 | 2 | 0,050 | 0,050 |
| **0,5** | **1,5** | 10 | 5 | 2 | 3 | 0,054 | 0,086 |
| **0,5** | **2** | 10 | 5 | 2 | 4 | 0,056 | 0,117 |
| **1** | **0,5** | 10 | 10 | 2 | 1 | 0,050 | 0,055 |
| **1** | **1** | 10 | 10 | 2 | 2 | 0,049 | 0,050 |
| **1** | **1,5** | 10 | 10 | 2 | 3 | 0,049 | 0,052 |
| **1** | **2** | 10 | 10 | 2 | 4 | 0,050 | 0,054 |
| **1,5** | **0,5** | 10 | 15 | 2 | 1 | 0,051 | 0,087 |
| **1,5** | **1** | 10 | 15 | 2 | 2 | 0,050 | 0,050 |
| **1,5** | **1,5** | 10 | 15 | 2 | 3 | 0,049 | 0,036 |
| **1,5** | **2** | 10 | 15 | 2 | 4 | 0,049 | 0,030 |
| **2** | **0,5** | 10 | 20 | 2 | 1 | 0,051 | 0,114 |
| **2** | **1** | 10 | 20 | 2 | 2 | 0,050 | 0,050 |
| **2** | **1,5** | 10 | 20 | 2 | 3 | 0,050 | 0,027 |
| **2** | **2** | 10 | 20 | 2 | 4 | 0,050 | 0,019 |
| **0,5** | **0,5** | 20 | 10 | 2 | 1 | 0,049 | 0,019 |
| **0,5** | **1** | 20 | 10 | 2 | 2 | 0,050 | 0,050 |
| **0,5** | **1,5** | 20 | 10 | 2 | 3 | 0,051 | 0,086 |
| **0,5** | **2** | 20 | 10 | 2 | 4 | 0,051 | 0,114 |
| **1** | **0,5** | 20 | 20 | 2 | 1 | 0,050 | 0,053 |
| **1** | **1** | 20 | 20 | 2 | 2 | 0,049 | 0,050 |
| **1** | **1,5** | 20 | 20 | 2 | 3 | 0,050 | 0,051 |
| **1** | **2** | 20 | 20 | 2 | 4 | 0,050 | 0,052 |
| **1,5** | **0,5** | 20 | 30 | 2 | 1 | 0,050 | 0,084 |
| **1,5** | **1** | 20 | 30 | 2 | 2 | 0,050 | 0,050 |
| **1,5** | **1,5** | 20 | 30 | 2 | 3 | 0,050 | 0,035 |
| **1,5** | **2** | 20 | 30 | 2 | 4 | 0,050 | 0,029 |
| **2** | **0,5** | 20 | 40 | 2 | 1 | 0,050 | 0,112 |
| **2** | **1** | 20 | 40 | 2 | 2 | 0,050 | 0,050 |
| **2** | **1,5** | 20 | 40 | 2 | 3 | 0,050 | 0,027 |
| **2** | **2** | 20 | 40 | 2 | 4 | 0,050 | 0,018 |
| **0,5** | **0,5** | 30 | 15 | 2 | 1 | 0,049 | 0,018 |
| **0,5** | **1** | 30 | 15 | 2 | 2 | 0,050 | 0,050 |
| **0,5** | **1,5** | 30 | 15 | 2 | 3 | 0,050 | 0,085 |
| **0,5** | **2** | 30 | 15 | 2 | 4 | 0,050 | 0,113 |
| **1** | **0,5** | 30 | 30 | 2 | 1 | 0,050 | 0,052 |
| **1** | **1** | 30 | 30 | 2 | 2 | 0,050 | 0,050 |
| **1** | **1,5** | 30 | 30 | 2 | 3 | 0,050 | 0,050 |
| **1** | **2** | 30 | 30 | 2 | 4 | 0,050 | 0,051 |
| **1,5** | **0,5** | 30 | 45 | 2 | 1 | 0,050 | 0,083 |
| **1,5** | **1** | 30 | 45 | 2 | 2 | 0,050 | 0,050 |
| **1,5** | **1,5** | 30 | 45 | 2 | 3 | 0,050 | 0,035 |
| **1,5** | **2** | 30 | 45 | 2 | 4 | 0,050 | 0,028 |
| **2** | **0,5** | 30 | 60 | 2 | 1 | 0,050 | 0,111 |
| **2** | **1** | 30 | 60 | 2 | 2 | 0,050 | 0,050 |
| **2** | **1,5** | 30 | 60 | 2 | 3 | 0,050 | 0,026 |
| **2** | **2** | 30 | 60 | 2 | 4 | 0,050 | 0,017 |
| **0,5** | **0,5** | 40 | 20 | 2 | 1 | 0,050 | 0,018 |
| **0,5** | **1** | 40 | 20 | 2 | 2 | 0,050 | 0,050 |
| **0,5** | **1,5** | 40 | 20 | 2 | 3 | 0,051 | 0,086 |
| **0,5** | **2** | 40 | 20 | 2 | 4 | 0,051 | 0,111 |
| **1** | **0,5** | 40 | 40 | 2 | 1 | 0,050 | 0,052 |
| **1** | **1** | 40 | 40 | 2 | 2 | 0,050 | 0,050 |
| **1** | **1,5** | 40 | 40 | 2 | 3 | 0,050 | 0,050 |
| **1** | **2** | 40 | 40 | 2 | 4 | 0,050 | 0,051 |
| **1,5** | **0,5** | 40 | 60 | 2 | 1 | 0,050 | 0,083 |
| **1,5** | **1** | 40 | 60 | 2 | 2 | 0,050 | 0,050 |
| **1,5** | **1,5** | 40 | 60 | 2 | 3 | 0,050 | 0,034 |
| **1,5** | **2** | 40 | 60 | 2 | 4 | 0,050 | 0,028 |
| **2** | **0,5** | 40 | 80 | 2 | 1 | 0,050 | 0,110 |
| **2** | **1** | 40 | 80 | 2 | 2 | 0,050 | 0,050 |
| **2** | **1,5** | 40 | 80 | 2 | 3 | 0,050 | 0,026 |
| **2** | **2** | 40 | 80 | 2 | 4 | 0,050 | 0,017 |