**Confidence Intervals, Pooled and Separate Variances T**

The programs provided elsewhere ([SPSS](http://core.ecu.edu/psyc/wuenschk/SPSS/CI-d-SPSS.zip) and [SAS](http://core.ecu.edu/psyc/wuenschk/SAS/Conf-Interval-d2.sas)) for computing Cohen’s *d* (estimate of Cohen’s *δ* ) and putting a confidence interval about that estimate assume that one is willing to pool the error variances both for computing *d* and for obtaining the confidence interval. IMHO it makes sense, when computing *d*, to pool the within-group variances even when they greatly differ from one another (although in some cases it would be better to employ Glass’ Δ in the first place).

If you have unequal sample sizes and you supply (to the confidence interval program) a separate variances *t* value rather than a pooled variances *t* value, the value of *d* will be incorrectly computed (except in the special case where the sample sizes are equal, in which case *d* will be correctly computed regardless of which *t* you use, since in this special case the value of *t* (but not the value of *df*) for separate variances is identical to that for pooled variances). Obtaining the correct *d* does not, however, mean that the obtained confidence interval is also correct.

**My Current Advice**

Give the confidence interval program the **pooled** *t* and *df*, even when you will be reporting the separate variances *t* test. Then check to see if there is any inconsistency between the conclusion from the separate variances *t* test and the pooled variances confidence interval – for example, if the separate variances *t* test is not significant but the confidence interval excludes zero then there is an inconsistency. You are most likely to find such an inconsistency when the sample variances and sizes differ greatly between groups. If there is no inconsistency, report the pooled variances confidence interval. If there is an inconsistency, report an approximate confidence interval which is obtained by dividing each end of the separate variances confidence interval by the pooled standard deviation. Inform your audience that the confidence interval is approximate.

There are robust methods for estimating the standardized difference between two means, with homogeneous or heterogeneous variances. If you wish to learn about them, go to <http://dx.doi.org/10.1037/1082-989X.13.2.110.supp> .

**An SAS Example**

--------------------------------------------------------------------------------------------------

The TTEST Procedure

Variable: Pups

Group N Mean Std Dev Std Err Minimum Maximum

2F 8 40.8750 2.8504 1.0078 37.0000 44.0000

2M 15 56.8667 12.7552 3.2934 40.0000 84.0000

Diff (1-2) -15.9917 10.5438 4.6161

Group Method Mean 95% CL Mean Std Dev 95% CL Std Dev

2F 40.8750 38.4920 43.2580 2.8504 1.8846 5.8014

2M 56.8667 49.8031 63.9303 12.7552 9.3384 20.1162

Diff (1-2) Pooled -15.9917 -25.5913 -6.3921 10.5438 8.1119 15.0678

Diff (1-2) Satterthwaite -15.9917 -23.2765 -8.7069

Method Variances DF t Value Pr > |t|

Pooled Equal 21 -3.46 0.0023

Satterthwaite Unequal 16.456 -4.64 0.0003

Equality of Variances

Method Num DF Den DF F Value Pr > F

Folded F 14 7 20.02 0.0005

--------------------------------------------------------------------------------------------------

Notice that the groups differ greatly in variance (that in the 2M group being twenty times that in the 2F group) and sample size. The pooled standard deviation is 10.5438. The difference in means is 15.9917, and *d* is 15.9917/10.5438 = 1.52. If we use one of my programs for constructing a confidence interval for *δ* and use the value of pooled *t* and *df*, the output indicates that the point estimate is 1.52 and the confidence interval runs from .53 to 2.47. What happens if we provide the confidence interval program with the separate variances *t* and *df* ? We would get a point estimate of 2.013 and a confidence interval running from .913 to 3.113. This is more than a trivial error. Clearly we cannot use the separate variances *t* and *df* with the available programs for obtaining a standardized confidence interval.

**How to Compute an Approximate Standardized Confidence Interval, Pooled Variances**

Simply divide each end of the unstandardized confidence interval by the pooled standard deviation. The unstandardized confidence interval runs from 6.39 to 25.59. The pooled standard deviation is 10.54. Dividing each end of the unstandardized confidence interval by 10.54 yields an approximate standardized confidence interval that runs from .61 to 2.43. Notice that the exact confidence interval is a little bit wider than the approximate confidence interval, because the exact confidence interval takes into account error in the estimation of σpooled, but the approximate confidence interval does not.

**How to Compute an Approximate Standardized Confidence Interval, Separate Variances**

First, obtain the unstandardized separate variances confidence interval. Second, obtain the pooled standard deviation. SAS gives this with the TTEST output. With SPSS you get it by conducting an ANOVA comparing the two groups and then taking the square root of the Mean Square Error. Third, divide each end of the separate variances confidence interval by the pooled standard deviation.

From the output above, the unstandardized interval runs from 8.71 to 23.28. Divide each end of this confidence interval by the pooled standard deviation, obtaining an approximate standardized confidence interval running from .83 to 2.21. The midpoint of this confidence interval is still equal to the point estimate of *d*. As with the approximate pooled variances confidence interval, the width of the approximate confidence interval is less than that of the exact confidence interval.

**An SPSS Example**

Consider the data W\_LOSS2 at my [SPSS data page](http://core.ecu.edu/psyc/wuenschk/SPSS/SPSS-Data.htm). We are comparing the amount of weight loss by participants in two different weight loss programs.

Notice that both the mean and the standard deviation are higher in the first group than in the second group. The ratio of the larger variance to the smaller variance is great, 6.8.

****

****

If we inappropriately applied the **pooled test**, we would report a significant difference. The unstandardized confidence interval for the difference between means would run from 2.133 to 16.701. When we use the pooled *t* and *df* to construct the standardized confidence interval, we get *d* = 1.37 and a standardized confidence interval that runs from .266 to 2.44. The confidence interval probably should be wider, given the heterogeneity of variance.

If we appropriately applied the **separate variances test**, we would report that the difference fell short of statistical significance (by the usual .05 criterion) and the confidence interval runs from 1.766 pounds in one direction to 20.599 pounds in the other direction. If we use the pooled *t* and *df* to construct the exact standardized confidence interval, we get *d* = 1.37 and a standardized confidence interval that runs from .266 to 2.44. The *d* is correct, but the confidence interval is not (it should include zero). If we give the program the pooled value of *t* and the separate variances value for *df*, the confidence interval does get wider (.088, 2.584), but it still excludes zero.

If we were to use the separate variances *t* and *df*, the program would give us a value of 1.04 for *d* and -.133 to 2.152 for the standardized confidence interval. The value of *d* is incorrect. The *d* should be the same regardless of whether the variances are homogeneous or not, but it may well be appropriate for the width of the confidence interval to be affected by heterogeneity of variance.

The approximate separate variances confidence interval is easily calculated. A quick ANOVA on these data shows that the *MSE* is 47.224, so the pooled standard deviation is 6.872.

****

Dividing the confidence limits by 6.872 yields the approximate separate variance confidence interval: −0.257, 2.998. The midpoint of this confidence interval does equal the point estimate of *d*. The interval should be a bit wider than it is, because we have not taken into account error in our estimation of the pooled standard deviation, but I can live with that.

**Unequal Sample Sizes**

If we follow Donald Zimmerman’s advice (A note on preliminary tests of equality of variances, *British Journal of Mathematical and Statistical Psychology*, 2004, *57*, 173-181), we shall always use the separate variances *t* test when our sample sizes are not equal. This leads to the awkward possibility of obtaining a unstandardized confidence interval (using separate variances error and *df*) that includes the value 0 but a standardized confidence interval (using pooled error and *df*) that does not.

I am not uncomfortable with employing the pooled variances test when the sample sizes differ a little and the sample variances differ a little, but where should one draw the line between a little and not a little. When the sample sizes do not differ much and the sample data do not deviate from normal by much, many are comfortable with pooled *t* even when the one sample variance is up to four times the size of the other.

It would be nice to have a rule of thumb such as “If the ratio of the larger sample size to the smaller sample size does not exceed 2, then you may use the pooled variances test unless the ratio of the larger variance to the smaller variance exceeds 4,” but, to the best of my knowledge, nobody has done the Monte Carlo work that would be required to justify such a rule of thumb.

**What Does Smithson Say About This?**

I think we shall probably have to wait for future work to have a more comfortable solution here. I have asked Michael Smithson about this (personal communication, June 8, 2005) -- see his excellent page on [confidence intervals](http://psychology3.anu.edu.au/people/smithson/details/CIstuff/CI.html). Here is his response:

That's a good question and I'm not entirely confident about the answer I'll give, but here 'tis for what it's worth... My understanding is that the separate variances *t* would be the appropriate estimate of the noncentrality parameter simply because the separate variances formula is the appropriate estimate for the variance of the difference between the means. As for the *df*, intuitively I'd go for the downward adjustment if it's being used for hypothesis-testing because any confidence interval (CI) is an inverted significance-test. If you want a CI that agrees with your t-test then you'd need to use the same *df* as that test does. That said, I don't know how good the resulting approximation to a noncentral *t* would be (and therefore what the coverage-rate of the CI would be). In the worst case, I believe you'd get a conservatively wide CI.

Karl L. Wuensch, Dept. of Psychology, East Carolina University, Greenville, NC. USA

March, 2015.

[Return to My Statistics Lessons Page](http://core.ecu.edu/psyc/wuenschk/StatsLessons.htm)