

ADDITIONAL NOTES RE: TWO SAMPLE t TESTS

The classic t test for the difference between two population means ... makes an assumption that the two population variances are equal (or equal SDs). Not close ... but equal. But, we know that when we have two independent samples ... each of the sample variances is ... under the homogeneous variance assumption ... simply an estimate of the SAME thing. Thus ... since we have TWO samples, we could “pool” the data between the two samples ... sort of average the two sample variances (so to speak) ... and get a better estimate of the variance in either or both of the target populations. Such a “pooled” formula for the SD is shown below:

Est. of SD in each pop. (pooled version) or est. of COMMON SD:

$$\text{Formula 1: } s_p = \sqrt{[(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2] / (n_1 + n_2 - 2)}$$

Now, after you get an estimate of the pooled standard deviation, we need to use it in the following formula to get an estimate of the standard error of the difference in means. It is ...

$$\text{Formula 2: } s_{\bar{x}_1 - \bar{x}_2} = s_p \sqrt{1/n_1 + 1/n_2}$$

Thus, the combination of 1 and 2 ... produce the standard error we need for the denominator for the two sample t test.

The formula in the book on page 212 is given as the formula for the standard error of the differences in the means. Well ... you might now be wondering ... is the formula on page 212 in the book ... which is:

$$\text{Formula 3: } s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

really the same as the combination of 1 and 2 above?

Well, the answer to this is YES ... when ns are equal. When ns are equal both formulas really boil down to (and prove this for a real FUN time!) the following:

$$\text{Formula 4: } \sqrt{(s_1^2 + s_2^2) / n} \text{ for either sample}$$

Now, when ns are NOT the same for both samples ... the combination of formulas 1 and 2, and 3 ... **DON'T NECESSARILY** produce the same standard error of the difference in means. Whether they do or not depends on whether the sample variances are the same ... if they are ... then the combo of formulas 1 and 2, produce the same value as does formula 3. But, if the variances in the samples are not the same ... then the combination of formulas 1 and 2, will produce a different value than formula 3. Is this clear as mud? Maybe a table will help.

FORMULAS FOR ERROR	ns for samples	variances for samples
1+2, 3 = SAME	same	same
1+2, 3 = SAME	same	different
1+2, 3 = SAME	different	same
1+2, 3 = DIFFERENT	different	different

The bottom line is that ... for cases where ns are different since the variances in the two samples are NOT likely to be the same ... then the standard errors for the difference in means will be different when you make an assumption of equal population variances then when you do not. The first case is referred to as the “CLASSIC t TEST” and the second is normally referred to as the “BEHRENS-FISHER” t.

Thus, the question is ... what is the impact of making the assumption of equal variances or not? Clearly, if you use the appropriate procedures to estimate the standard error of the difference in means, the resulting standard errors will NOT necessarily be the same ... and if they are not the same, then the t test value will not necessarily be the same. And, if the t test values are not the same ... then ones chances for retaining or rejecting the null are not the same ... if we assume using the SAME critical t value ... based on degrees of freedom. Thus, if your chances of retaining or rejecting the null are somewhat different ... then the chance of making a type I error (alpha) will be different too.

The bottom line problem is usually this. If you make an assumption that the two populations have equal variances, AND you are not correct, then it is likely that you will be using a type I error rate that is not exactly correct. That is ... if you find what you think are the correct degrees of freedom (N - 2) and use that to find the critical value ... then the p value that you are setting the critical value for ... is not likely to be correct. The rate could be greater than what you are thinking ... or it could be smaller than what you are thinking. The exact extent of this mistake is dependent upon both the unequalness of the ns ... AND the unequalness of the sample variances ... AND ... whether the larger n is paired with the larger/smaller sample variance.

Here, all I will say is that ... if your equal population variance assumption is not correct ... then the actual p value that you are working with could be much smaller OR larger than what you think.

It is very likely that even if you make the equal population variance assumption, that in reality ... the two population variances are NOT EXACTLY the same. Thus, we would be better off generally speaking using FORMULA 3 ... which does not make that assumption in the process of finding a value for the standard error of the difference in means ... and using it as the denominator for the two sample t test. This will tend to provide a more

realistic estimate of the standard error. However, this still does not completely get us out of the woods ...

In the book on page 212, what you have is the formula for finding the standard error of the difference in means where NO assumption about the equality of the population variance is made. So, on this score, the estimation of the standard error seems to be ok. But, remember that the procedure for finding the degrees of freedom value ... based on the classic t test with the homogeneous variance assumption is: $n_1 + n_2 - 2$. Thus, in two samples where both ns are 30, the df value would be 58. But, if the variance assumption is made ... but is not true ... then using df as 58 in this case ... will lead us to a possible type I error rate that is not correct. So, how do we adjust for this problem? Here is a formula that makes an adjustment in the df value ... based on the combination of the ns for both samples and the variances from both samples.

$$\text{Formula 5: Adjust df} = (s_1^2 + s_2^2)^2 / [((s_1^2)^2 / (n_1 - 1)) + ((s_2^2)^2 / (n_2 - 1))]$$

In Minitab, the default way to produce the standard error for the difference in means is to NOT make the equal population variance assumption. Now, even in the case where the ns are equal ... it is likely that the sample variances will not be equal ... and thus Minitab will see this as a red flag (ie, the population variances are also not likely to be the same) and makes the correction to the degrees of freedom value using FORMULA 5 above. If you look at the printout for the TWOSAMPLE t test ... you will notice that the df value given is not $n_1 + n_2 - 1$... as we normally figure df ... but rather a smaller one ... based on the adjustment produced in formula 5. IF THE ns are the same but ... the sample variance differences are small ... only a small reduction if df will be noticed. However, if ns are not the same AND there is a large difference in sample variances, there could be a rather drastic reduction in the df value (depending upon whether the larger n is accompanied by the larger sample variance, or other combinations).

The bottom line is ... if the df you need to use is less ... then the critical value you will find from the table will be larger ... and if it is larger, then it will be more difficult to reject the null hypothesis. In this sense ... using the corrected df value makes the test MORE CONSERVATIVE ... IE, a more conservative test is one where your chances of making a type I error (ie, rejecting the null when you should not) are smaller ... since it will not allow you to reject the null as often.

In addition, if you are a confidence interval "person", what happens above could have an impact on the width of the confidence intervals. Note that in 3 of the 4 conditions shown in the table above ... there is NO difference between the standard error formula results in both cases ... only when both ns are different and the sample variances are different will the two standard error formulas produce a different value. Of course, in that last case ... t test values and confidence intervals will be different since, the standard error is used in both ... as a divisor in the t test and as a multiplier when building a confidence interval. But, generally, especially if ns are the same, the standard error values would be identical using

both formulas. So ... how do the confidence interval change widths? Well, if there is an adjustment to the df value that you would use ... from formula 5 ... then the critical value that you look up will be somewhat different ... usually somewhat LARGER. Thus, when using it as a multiplier in a confidence interval building process it will produce a wider interval. In that sense, with a wider interval ... one is not able to be quite as "pin pointy" as to where the true difference in means might be. You would have just as much confidence ... in the normal sense of that word but ... the PRECISION of the estimate would be less.

In general, while some don't agree with this summary, my position is that it would be better to NOT make the equal variance assumption UNLESS YOU HAVE SOME OTHER INFORMATION THAT IS STRONG THAT SAYS THAT THEY ARE.

POINT 1: So, when calculating the standard error for the difference in means, formula 3 would be the way to go.

POINT 2: Then, if you use that ... then one should logically use formula 5 to make an adjustment to the df value to use when finding a critical value ... for use in either the t test or when building a confidence interval.

This is an overall conservative approach ... but, on the whole, appears to me to be the right course of action.