More on two sampled t-tests

I. Quick review:

A. The basic procedure for hypotheses tests:

- set up your hypotheses
- decide on $\alpha$
- verify assumptions (which is what's covered in this set of notes)
- calculate $t^*$ (or whatever your test statistic is)
- compare to the tabulated value of $t$, using the correct d.f.
  - (or compare $p$ to $\alpha$)
- make your decision to reject or “fail to reject”

B. The only difference between this and other statistical procedures will be how to calculate your test statistic ($t^*$ in our case), and which tables (and how) to use.

- we’ve already discussed how to calculate $t^*$.

- also discussed how to look up the tabulated t-value (aren’t we happy we have computers!!)

II. Now, what about our assumption of $\sigma_1 \neq \sigma_2$??

A. Let’s make sure we understand this:

- we’re assuming that the POPULATION standard deviations (or, equivalently, the variances) are not the same. It’s important to realize that the SAMPLE standard deviations may or may not be the same, regardless of this assumption.

B. But suppose we can assume that they are?? Does that change anything?

- YES! Because now we can use a “pooled” estimate for our standard error; in other words, our standard error is now based on both samples.

- What does $t^*$ look like now?

$$
\begin{align*}
t^* &= \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s^2_{pooled} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \\
\text{where} & \\
s^2_{pooled} &= \frac{(n_1 - 1)s^2_1 + (n_2 - 1)s^2_2}{n_1 + n_2 - 2}
\end{align*}
$$
- $s^2_{pooled}$ is simply a weighted average of $s_1$ and $s_2$. In fact, if $n_1 = n_2$:

$$s^2_{pooled} = \frac{s_1^2 + s_2^2}{2}$$

- Now note that the formula is almost identical to the previous one, except that now we have $s^2_{pooled}$ in the denominator, instead of two separate estimates.

One can prove that if $n_1 = n_2$ OR $s_1 = s_2$ then the two formulas are identical (that's really easy to see in the case of $s_1 = s_2$)

This formula can give you quite a different result on other cases.

Regardless, it turns out that the formula for the d.f. also changes:

$$d.f. = v = n_1 + n_2 - 2$$

This is MUCH easier than the other formula!

C. Some comments:

1) As mentioned, if either the $n$’s or the $s$’s are the same, the two formulas give the same result. But if the $s$’s and $n$’s are not the same, you get two different results.

2) So why all the fuss if the two formulas are the same for equal $s$’s, and we’re assuming $\sigma_1 = \sigma_2$?

REMEMBER: $s$ is an estimate of $\sigma$. Thus, unless we get exactly the same $s$ out of our samples (VERY unlikely, even if $\sigma_1 = \sigma_2$), the pooled formula will do a better job if $\sigma_1 = \sigma_2$ is true.

Why? because it combines the estimates (the two values for $s$) in a better way (i.e., mathematically) if this assumption is true. This is a fine point!

3) If $\sigma_1 = \sigma_2$, the results from our two formulas are generally pretty close. Again, so why all the fuss?

BECAUSE, if $\sigma_1 = \sigma_2$, then the t-test using the pooled $s^2$ is actually more powerful! In other words,

$$\Pr\{\text{rejecting } H_0 \} \text{ if } H_0 \text{ is really false}$$

is better with this test if $\sigma_1 = \sigma_2$ is really true.
D. But note very carefully!!

- if \( \sigma_1 \neq \sigma_2 \) is really true, and if \( n_1 \neq n_2 \) (remember, the two formulas are the same if \( n_1 = n_2 \)), then the \( t \)-test based on a pooled \( s^2 \) can give you VERY bad results, and can be very misleading!

E. So what do you do?

- Unless you’re pretty certain about \( \sigma_1 = \sigma_2 \), use the \( t \)-test based on the unpooled standard deviations. If \( \sigma_1 = \sigma_2 \), this test is not quite as powerful, but it's only a small difference in power.

- ONLY if you’re pretty sure about \( \sigma_1 = \sigma_2 \), then use the pooled standard deviation.

F. It is possible to test and find out if \( \sigma_1 = \sigma_2 \). But there are two problems.

1) Remember that we never “prove” our \( H_0 \), so we can't prove \( H_0: \sigma_1 = \sigma_2 \).

2) The test to do this is not very powerful (in fact, it's awful!).

Therefore, we won’t bother learning this test here. Stick with the suggestions in item (E) instead.

G. Incidentally, this assumption is usually called the “equal variances” assumption. This is the original \( t \)-test. The previous \( t \)-test we learned (for unequal variances) is often called Welch's test.

III. Other assumptions.

A. We just got done dealing with our first assumption. As you can tell, dealing with assumptions can be complicated.

B. The \( t \)-test has several other assumptions we need to know about. Probably the most important is the “normal distribution assumption”. This, incidentally, is true for all \( t \)-tests, and a lot of the tests we’ll be looking at in the future.

C. Normally distributed:

1. As you can guess, this implies that the data (sample) come from a population that is normally distributed.

2. If you take a “sample” from a normally distributed population, it makes sense that your sample will be normally distributed.
- you’re more likely to pick items near the average, because more items will be NEAR the average!

3. So how do you figure out if your population is actually normally distributed?

- graphical methods:
  - histograms
  - box plots
  - QQ plots

- statistical tests:
  - goodness of fit (very bad!, but some people will use this)
  - Shapiro-Wilks test (not great, but okay, particularly for smaller sample sizes).
  - Kolmogorov test (also not great, but works better for larger samples sizes).

- a lot of people don’t much care for “tests” of normality. We won’t be covering them in class, but you’ll know what they’re called and can look them up if you need them.


- main problem - a little subjective. Sometimes it’s hard to decide.

- histograms - this is fairly easy:

  - make a histogram of your data (but remember - the shape of your histogram can change depending on how many “categories” you use)

  - before you do anything else, does it look normal? If it looks grossly abnormal, obviously it’s not going to be normal.

  - overlay a normal curve on your histogram. How??

    - the easy way - let the computer do it.

    - the hard way - don’t bother. You can do this, but it involves plugging in the equation for the normal curve, etc.
- box plots - not very good for this, but if these don’t look somewhat symmetrical, there’s a problem

- but neither histograms or box-plots are ideal (and can be misleading). The best way is a “normal probability plot”, also known as q-q plot (actually, a normal probability plot is a particular type of q-q plot).

- in general, we’ll let the computer do this. What the computer does is roughly as follows:

  - It takes your data, figures out your sample size, and then calculates what you should “expect” from a sample of that size.

  - It then plots these “expected” values against what you actually got. If this turns out to be a straight line, then things are probably normal.

- Your text has a pretty good description on p. 138 [136].

  - For example, if you took many, many samples of size 11, where would you expect the smallest value to be? The second smallest? etc. (this is easy to see for the median, but you can do all of your values in a similar way).

- We’ll look at the math a little more in a minute, but let’s do an example first (see p. 137 [134])

The vertical axis is basically just a dot plot. BUT, of course, there’s now a horizontal axis. Where do you put your dot on the horizontal axis? You calculate a “normal score”

- there are several ways to do this:

  The second and third edition do things one way (see the footnote on p.138 [136], though this description is slightly wrong if you use the tables in your text).

  The fourth edition does things differently and calculates “expected” values.

  - this is easier to understand (perhaps), but no one really does it this way, so we’ll stick with the 2nd and 3rd edition approach.
where \( \alpha = \frac{i - \frac{1}{3}}{n + \frac{1}{3}} \)

- (Your text (2\textsuperscript{nd} and 3\textsuperscript{rd} edition) has “1-\(\alpha\)” as a subscript for \(Z\), which is exactly backwards of the way your tables are set up).

- The \(Z\) value is what you look up in the \(z\)-tables. You just need to calculate \(\alpha\). “\(i\)” is the \(i\)\textsuperscript{th} normal score, 1\textsuperscript{st}, 2\textsuperscript{nd}, 3\textsuperscript{rd}, etc., and \(n\) is our usual sample size.

- For instance, let’s look at the example of heights on p. 137. There are 11 women, so the first normal score would be:

\[
\alpha = \frac{1 - \frac{1}{3}}{11 + \frac{1}{3}} = 0.0588
\]

so we look up

\[Z_\alpha = Z_{0.0588} = -1.56\]

- The second normal score would be:

\[
\alpha = \frac{2 - \frac{1}{3}}{11 + \frac{1}{3}} = 0.1471
\]

so we look up

\[Z_\alpha = Z_{0.1471} = -1.05\]

- the 4th edition calculates normal scores a bit differently, and gets different numbers for the normal scores (-1.69 and -1.10 for the two calculations above).

- (remember, you’re finding \(\alpha\), then doing a reverse lookup for your \(Z\)-value.). The other 9 numbers would be calculated the same way - aren’t you glad the computer can do all this?
- If we calculated all 11 values and plotted them, we’d get the graph on p. 137. You plot your smallest observed value against the 1st Z-score you calculate ($i=1$), the second smallest against the 2nd Z-score, and so on.

- If things are perfectly normal, then all the dots should line up on a straight line (think of this a little like plotting $x=y$. If things are exactly normal, all your Z-scores should have exactly the same ratio to your actual values, so everything should be on a straight line).

- Anyway, you hardly ever do this by hand. You have the computer graph your samples, and then look at them.

- Question - what is sufficiently straight? That takes some practice. Some basics:

  - don’t worry about every little bump. There’ll be lots of bumps.

  - worry if:

    - you see a strong overall curve (e.g. fig. 4.26 on p. 140)

    - if you see a backwards “S” (if the ends curve towards the vertical) - this indicates long tails which is BAD. If you see a regular “S” that indicates short tails, that’s usually okay, even if it isn’t quite normal.

- Some folks (& software (e.g., Minitab) like to reverse the axes, so always make sure you look at the labels of the axes before you interpret a probability plot.

  - if the axes are reversed, then the above comments need to be “reversed” as well (a regular “S” becomes long tailed).

- Exercise 4.15, p. 143 is pretty good for a start.

5. Some comments on the normal assumption.

  - Depending on how the data are “not normal”, the t-test will probably still do pretty good, though there are often “more powerful” tests.
D. What are some other assumptions? Here are all of them:

1. The data must be collected “randomly” (a “random sample”)

2. In the case of a two sample test, each sample needs to be independent - for instance, if I choose someone for group A, this can’t influence how I choose someone else for group B.

3. The data should be normal, BUT, if your sample size is nice and large, you can get away with non-normal data (remember the CLT).
   - what is large? For a two-sample test, if each sample is 20 or over that is often good enough. But it depends a little on how non-normal the data is.
   - If the data are strongly non-normal, you might need a larger sample size; but if it’s almost, but not quite normal, you can get away with a smaller sample size. Worry more if you have long tails in a q-q plot.

4. For a two-sample t-test you should figure out whether or not you can use the assumptions of equal variances. Unless you're pretty sure, you should not assume equal variances.