The important assumptions for using the *t*-test

S-012

Helpful advice for assignment 3

Assumptions

- These are called assumptions, but these are not things that we should assume are always true.
- These are really the important conditions that should hold in order for the t-test to work accurately
- "Assumption" is more in the sense of: "Imagine the following conditions are true, then what might we expect."
- Example: "Imagine that there are two populations (two schools) that have identical means on the achievement test."
  - "What might happen if we take a sample from each school and compare the sample means?"
- Or a related example: "Imagine that there is a single population of scores that are normally distributed (μ,σ)"
  - "What might happen if we were to take two samples from this population?"

Think of these as "things we should always check"
Do not just assume!

Things we should check
Conditions for an effective t-test

- 1. Independence
- 2. Normality
- 2. Equal variances

Independence

- Are the observations independent?
  - Do we have separate samples of people in the groups?
  - Or do we have the same people at two different times?

Different (independent)

Obs	Group	Score
101	0	75
102	0	80
103	0	70
104	0	76
201	1	81
202	1	85
203	1	75
204	1	77

Repeated measures (paired groups)

Obs	Pre	Post
501	75	81
502	80	81
503	70	85

We use different t-tests for these two situations.
For now we are learning about the t-test for independent groups.

Other combinations

Between-group and within-subjects
(Mixed models)

Normality

- Is the distribution normal?
  - The test works best when the distribution of the scores is approximately normal
  - This is how Gossett (and Fisher) first developed the ideas for the t-distribution and the t-test of significance
  - What will things look like if the null hypothesis is true? (I know! They will look like this . . . and the t-test will be a good test!)
- If not normal
  - The test might not work so well
  - Can we try a transformation?
  - (Later: should we try a test that does not rely on having a normal distribution?)

When there is a normal (bell shaped, symmetric, well-behaved) distribution:

- the test works well
- rejects H0 when it should
- does not reject when it should not
- It is a powerful statistical test

When the distribution is "strange" (skewed, "wacky" etc.):

- High (or low) values have a big influence on mean and SD
- Sample results will vary a lot
- The test will be less reliable
  - Sometime we reject H0 when we should not
  - Sometimes we do not reject when we should

Normal vs. non-normal

Equal variances

- This is a bit more subtle, but it is also important
- We are trying to compare the means of two groups
  - If the SDs are similar, then things are okay and the t-test will work well
  - But if the SDs are not similar, then be careful
    - The t-test was developed by assuming that the SDs are equal

So we need to check the SDs. And this is always a good idea.

To check the SDs, we use Fisher's test – using the F-distribution.
The logic is exactly the same as the t-test.

Null hypothesis: $\sigma^2 = \sigma^2$ (variances are equal) (SDs are equal)
Alternative: Not equal

Steps:

1. State null and alternative
2. Decide on alpha level (criterion)
3. Compute F-statistic and the probability value
4. Reject or do not reject

The F-test for the Standard Deviations

Fortunately we have alternatives: The classic method and the modern (well a bit more modern) method

- Classic t-test (the "equal variance approach")
  - Use the pooled SD approach
  - If the SDs are similar, then pool then (get the average – the weighted average – and use that for the t-test
- The alternative (the "unequal variance method")
  - Use Satterthwaite's adjustment (or maybe Welch's adjustment)
  - When the SDs are not similar (when the F-test tells us they are "significantly different") then use this idea:
    - Do not pool the SDs. Keep them separate. (The separate-variance method)
    - But now adjust the DF. (Make this df smaller – a kind of penalty for not having equal variances.)

This approach is brilliant!

- When the variances are different, this method adjusts things
  - The t-statistic, the df, and the probability value are adjusted
- When the variance are not different, there is no adjustment

I like this approach! It is the modern approach.
Why do we not use this all the time? Tradition? Let's be familiar with both.

Assumptions are important to check

Let's look at some more of the Stata examples

Remember:

- We test the means
- We also check the SDs
- Why?
  - This is important to check
  - This also helps us choose the correct t-test method