Comparing Two Means (Part 2)

Ray Block, Jr.
PS #585 (Research Methods)
Fall 2003

Today's Blueprint

Last Class

- Correlation (A Recap)
- Comparing Two Means
  - The Logic Behind Means Difference Tests

Today's Class

- Comparing the Means of 2 Dependent Populations
- Comparing the Means of 2 Independent Populations
  - Difference Tests
  - Confidence Intervals (T-Interval)

T-Tests: A Brief Review

T-Tests: A Brief Review

- They are tests of significance
- They determine whether there is a significant difference between the mean values of two groups

T-Tests: A Brief Review

- Two Types:
  - *T-Tests for Independent Means*
    - AKA: The Pooled T-Tests
  - *T-Tests for Dependent Means*
    - AKA: The Paired T-Test

T-Tests: A Brief Review

- *T-Tests for Independent Means* = Tests the difference between the means of 2 unrelated groups
  - Who is in the second sample doesn't depend on who is in the first sample

T-Tests: A Brief Review

- *T-Tests for Dependent Means* = Tests the difference between the means of 2 related groups
  - Who is in the second sample depends on who is in the first sample

T-Tests: A Brief Review

- Allow researchers to compare groups:
  - Determine whether one group is meaningfully different from another group
  - In an experiment, researchers can determine whether differences between groups are due to the manipulation

T-Tests: A Brief Review

- Data in each group follow a *normal distribution*
- For pooled test, the *variances* for each group *are equal* (no *between group variation*)
- The samples are independent

T-Tests: A Brief Review

- But what if samples are not independent?!
  - If samples are not independent, then we say that they are *dependent*, *correlated*, or *paired*

T-Tests: A Brief Review

- Ways That Pairing Can Occur:
  - When subjects in one group are "matched" with a similar subject in the second group.

- When subjects serve as their own control by receiving both of two different treatments.
- When, in "before and after" studies, the same subjects are measured twice.

T-Tests: A Brief Review

- When doing an *experiment*
- When doing a comparative observational study

T-Tests: A Brief Review

- *Experiments* = Manipulating one variable to observe its effect on another variable
  - One dependent variable (Y)
  - One independent variable (X) manipulated 2 ways
    - Treatment group
    - Control group

T-Tests: A Brief Review

- *Comparative Observation* = A research study in which two or more groups are compared with respect to some measurement or *response*
  - The groups, determined by their natural characteristics, are merely "*observed*"

Comparing the Means of 2 *Independent* Populations

The Pooled T-Test…

The Pooled T-Test

- Let μ_M = the average fastest speed of *all male students*
- And let μ_F = the average fastest speed of *all female students*
- Then we want to know whether $\mu_M \neq \mu_F$
  - This is equivalent to knowing whether $\mu_M - \mu_F \neq 0$

The Pooled T-Test

- In general, we can always compare two means by seeing how their difference compares to 0:

The Pooled T-Test

- Null hypothesis: $\mathbf{H_0: \mu_M = \mu_F}$
  - This is equivalent to saying that $\mu_M - \mu_F = 0$
- Alternative hypothesis: $\mathbf{H_a: \mu_M \neq \mu_F}$
  - This is equivalent to saying that $\mu_M - \mu_F \neq 0$

The Pooled T-Test

- Assume null hypothesis is true
  - That is, assume $\mu_M = \mu_F$
  - Or, equivalently, assume $\mu_M - \mu_F = 0$

The Pooled T-Test

- P-value = "How likely is it that our *sample means* would differ by as much as 16.35 m.p.h. if the difference in *population means* really is 0?"
- We choose a significance level of 0.05

The Pooled T-Test

- The P-value is determined by standardizing, that is, by calculating the *two-sample test statistic:*

The Pooled T-Test

- If variances of the measurements of the two groups are *not equal*...
- Estimate the standard error of the difference as:

The Pooled T-Test

- If variances of the measurements of the two groups are *equal*...

The Pooled T-Test

- Analyze ➔ Compare Means
- Specify whether you want an independent sample's T-Test or a Paired Samples T-Test
- Place the dependent variable (driving speed) in the "Test variable" box
- Place the independent variable (gender) in the "grouping variable" box
- Select your confidence interval (default is 95%) from the "options" box

The Pooled T-Test

```
Two sample T for Fastest

Gender       N      Mean     StDev    SE Mean
female      21     85.71      9.39        2.0
male        17     102.1      17.1        4.1

95% CI for mu (female) - mu (male  ): ( -25.9,  -6.8)
T-Test mu (female) = mu (male  ) (vs not =): T = -3.54
                                             P = 0.0017
                                             DF = 23
```

The Pooled T-Test

- Our obtained value (0.0017) is smaller than this (0.001)
- This tells us that our sample result is not likely if the null hypothesis is true
- Therefore, we can reject the null hypothesis

The Pooled T-Test

- There is sufficient evidence, at the 0.05 level of significance, to conclude that the average reported fastest driving speed of *all male college students* differs from the average reported fastest driving speed of *all female students*

The Pooled T-Test

- We can be "such-and-such" confident that the difference in the population means falls in the interval:

The Pooled T-Test

- Interpreting a confidence interval for the difference in 2 means:

Example 2: Comparing the Means of 2 *Dependent* Populations

The Paired T-Test

Do males earn higher average starting salaries than females?

Hypotheses for Paired T-test

- Null hypothesis: $\mathbf{H_0: \mu_M = \mu_F}$
  - This is equivalent to saying that $\mu_M - \mu_F = 0$
- Alternative hypothesis: $\mathbf{H_a: \mu_M \neq \mu_F}$
  - This is equivalent to saying that $\mu_M - \mu_F \neq 0$

Data analyzed as Paired T (See Slides)

Interpret Results

- P = 0.016. This is smaller than 0.05.
- Therefore, reject null.
- Sufficient evidence to conclude that average starting salaries differ between males and females.