Why Not Use Simple Analyses

EPSY
Longitudinal Data Analysis

School of Statistics
University of Minnesota

Feb 3, 2011

Simple Analyses

On Tuesday we discussed simple analyses, including...

- Independent t-tests on individual times
- Paired t-tests on only two times
- Difference of difference tests
- Testing the mean for each subject or the slope from separate regressions
- Removing subjects with missing data to make these possible.

These are sometimes reasonable ... but BEWARE! Problems can arise.

Problems?

What kind of problems?

- **Efficiency:** We lose information, so our inferences are less powerful.
- **Bias:** Summary data from different individuals can be measuring slightly different things.
- **Misses Richness in Data:** There's more going on than these methods can handle.

Omitting Subjects

In order to do some analyses on the pediatric pain data

- we omitted subjects with missing data
- which made our sample size smaller
- and potentially increased the standard error of our estimates

We'd prefer methods that can handle missing data.

Omitting Observations

For the pediatric pain data, we discussed a t-test of Time 4 against Time 2

- Why use just Time 2? Time 1 and 3 were also before treatment. We might believe all of them have information about the baseline.
- Not only that, we expect the baseline to be the same regardless of which treatment they get. We might want to use information across groups to estimate the baseline. Or pool the standard error across groups.

Prefer methods that allow us to specify how to combine data from different times.

Removing Missing Data

In the pediatric pain data, some were missing because of broken arms... what if they're different?

Summary Statistics

Suppose you took a mean across time for each subject.
You might not be measuring the same thing if...

- There is missing data
- The subjects were measured at different times

Prefer methods that somehow adjust estimates based on missingness and

Summary Statistics

- Not only that, the means might have different variances.
- Other summary statistics (like the slope) have that same problem.

Prefer methods that handle variance appropriately.

Difference of Differences

See handout

Single Measures

Analyzing single measures separately can miss relatedness

- For example, slope / intercept for each subject;
- What if slope and intercept are related??

Prefer methods that handle several measures together

Complex Patterns

Prefer methods that build a model to capture lots of things about the data, not tools that test just one thing.

In short, a **multivariate normal model.**