The “AB” of Random Effects Models

Rick Chappell, Ph.D.

Professor,
Department of Biostatistics and Medical Informatics
University of Wisconsin Madison

ADC Meetings, Baltimore, 10/2016

Wisconsin Alzheimer's Disease Research Center
UNIVERSITY OF WISCONSIN
SCHOOL OF MEDICINE AND PUBLIC HEALTH

Outline

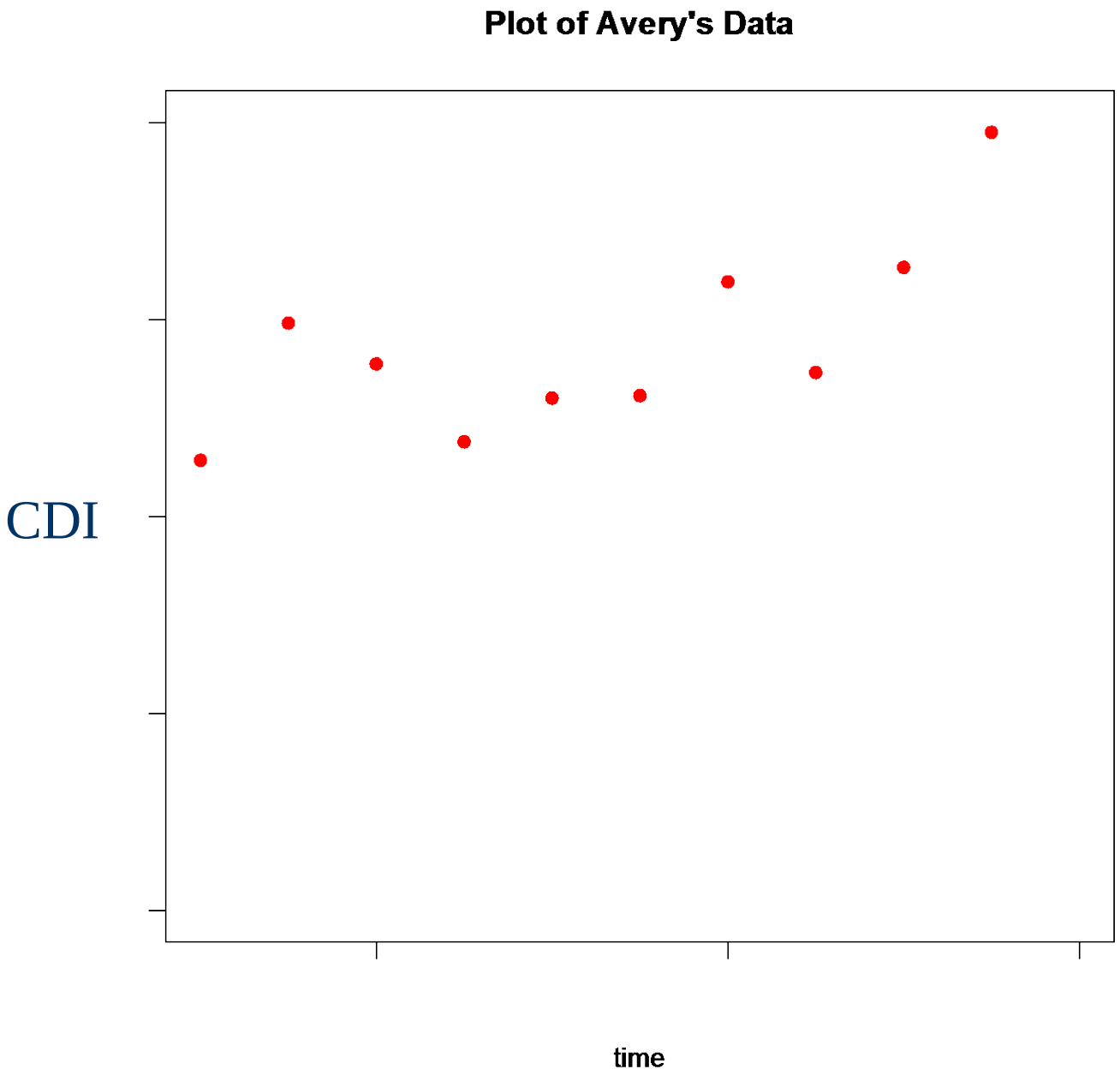
A. Linear Models for longitudinal models
B. Why can't we stick with "Fixed Effects" models?
C. Salvation via Random Effects models
D. Three useful answers

A. Linear Regression:

Consider a collection of "Avery's" *Cognitive Disability Index (CDI)* over time (red dots).

Time could be age, date, time since treatment, or other relevant scale.

For observational studies time often means age.

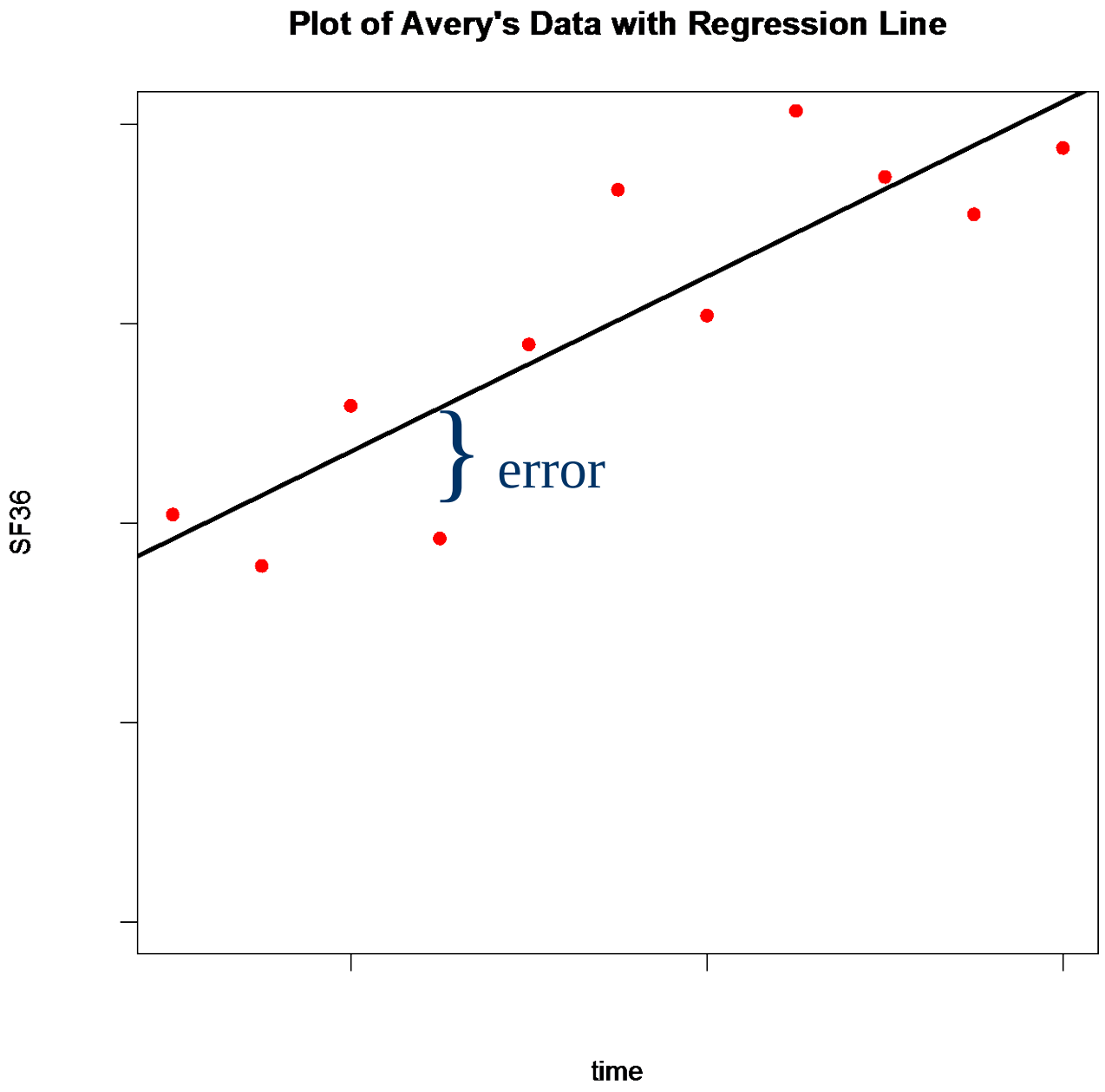

Plot of Avery's Data
CDI
time

Linear Regression:

Consider a collection of "Avery's" *CDI* over time (red dots).

The simplest model is a linear one, where at each time Avery's *CDI* is:

$$CDI = constant + \beta_{time} \times time + error$$



There are different errors (not shown) for the other 10 points.

Linear Regression:

But what if you now have Brenda's *CDI* measures (green dots)?

Even if they parallel Avery's changes over time, they still may be higher or lower than his on average:


Plot of Avery and Brenda's Data with Regression Line
CDI
error
error
time

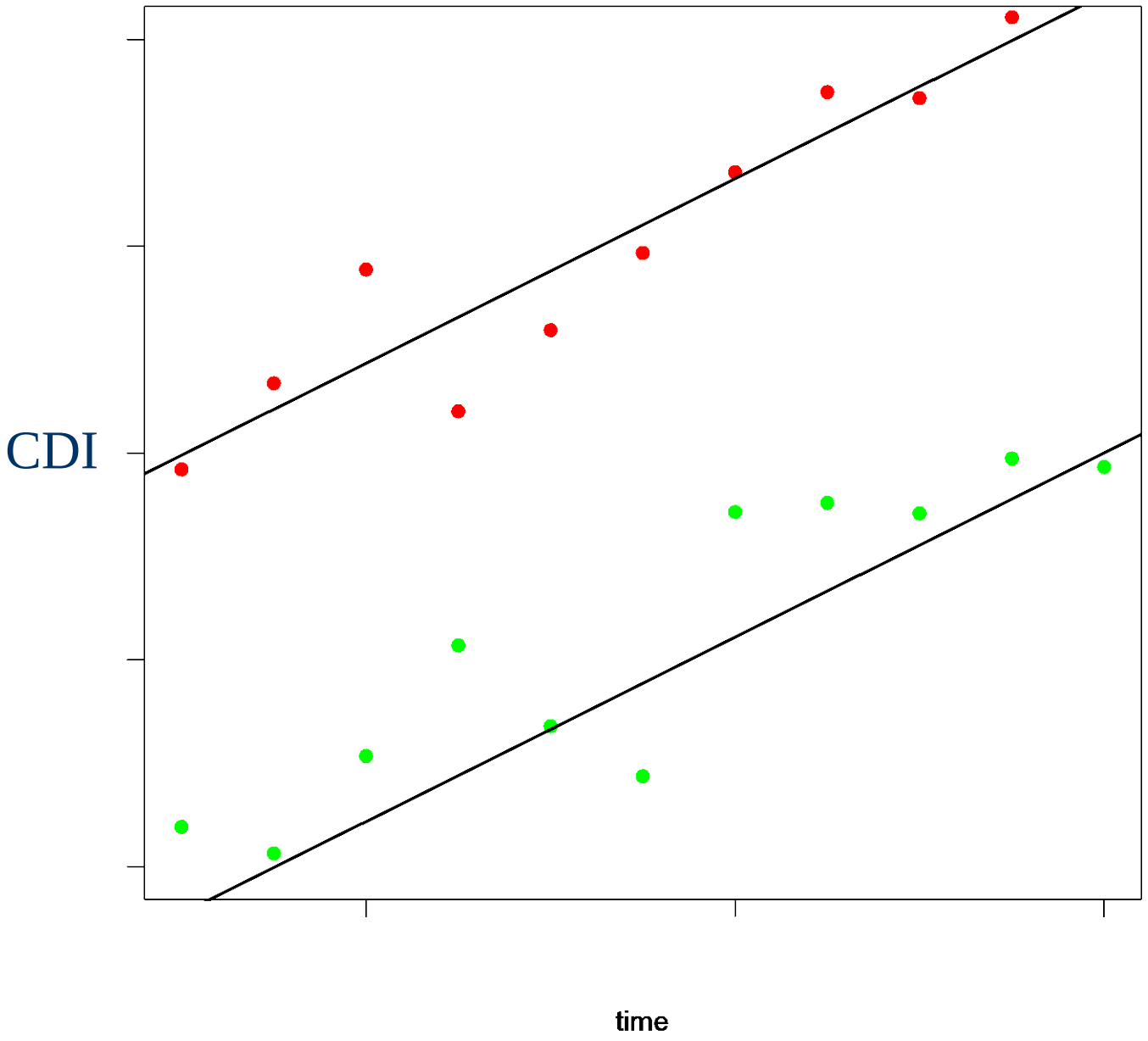
Linear Regression with a simple Fixed Effect:

Even if they parallel Avery's changes over time, they still may be higher or lower than his on average:

$$CDI = constant + \beta_{B\ vs.\ A} \times Ind(Brenda) + \beta_{time} \times time + error,$$

Where *Ind(Brenda)* = 1 for Brenda's data

0 for Avery's data


Plot of Avery and Brenda's Data with Fixed Effect Model
CDI
time

B. Why can't we stick with "Fixed Effects" models?

We need to include a lot more than two subjects.

We could have a fixed effects model with, say 500 parameters for 500 patients.

Alas:

1) This induces severe statistical problems.

2) It also doesn't answer the questions we ask.

1) Severe statistical problems:

- Tabulating and interpreting 500 estimates?

- The usual good properties of regression estimates require a *small* number of parameters.

1) Severe statistical problems:

- Tabulating and interpreting 500 estimates?

- The usual good properties of regression estimates require a *small* number of parameters.
- We can assess these properties by performing computer simulations of artificial datasets.

1) Severe statistical problems:

- Tabulating and interpreting 500 estimates?

- The usual good properties of regression estimates require a *small* number of parameters.
- We can assess these properties by performing computer simulations of artificial datasets.

Oh Joy.

2) [A fixed effects model] also doesn’t answer the questions we ask.

Do we want to be able to describe Avery, etc., or do we want to characterize humans in general?

C. Random Effects models:

We now have *two* error terms:

$$CDI = constant + \beta_{time} \times time$$

$$+ within\text{-}subject\ error$$

$$+ between\text{-}subject\ error$$

"within-subject error" is the same as plain "error" above

We now have *two* error terms:

$$CDI = constant + \beta_{time} \times time$$

$$+ within\text{-}subject\ error$$

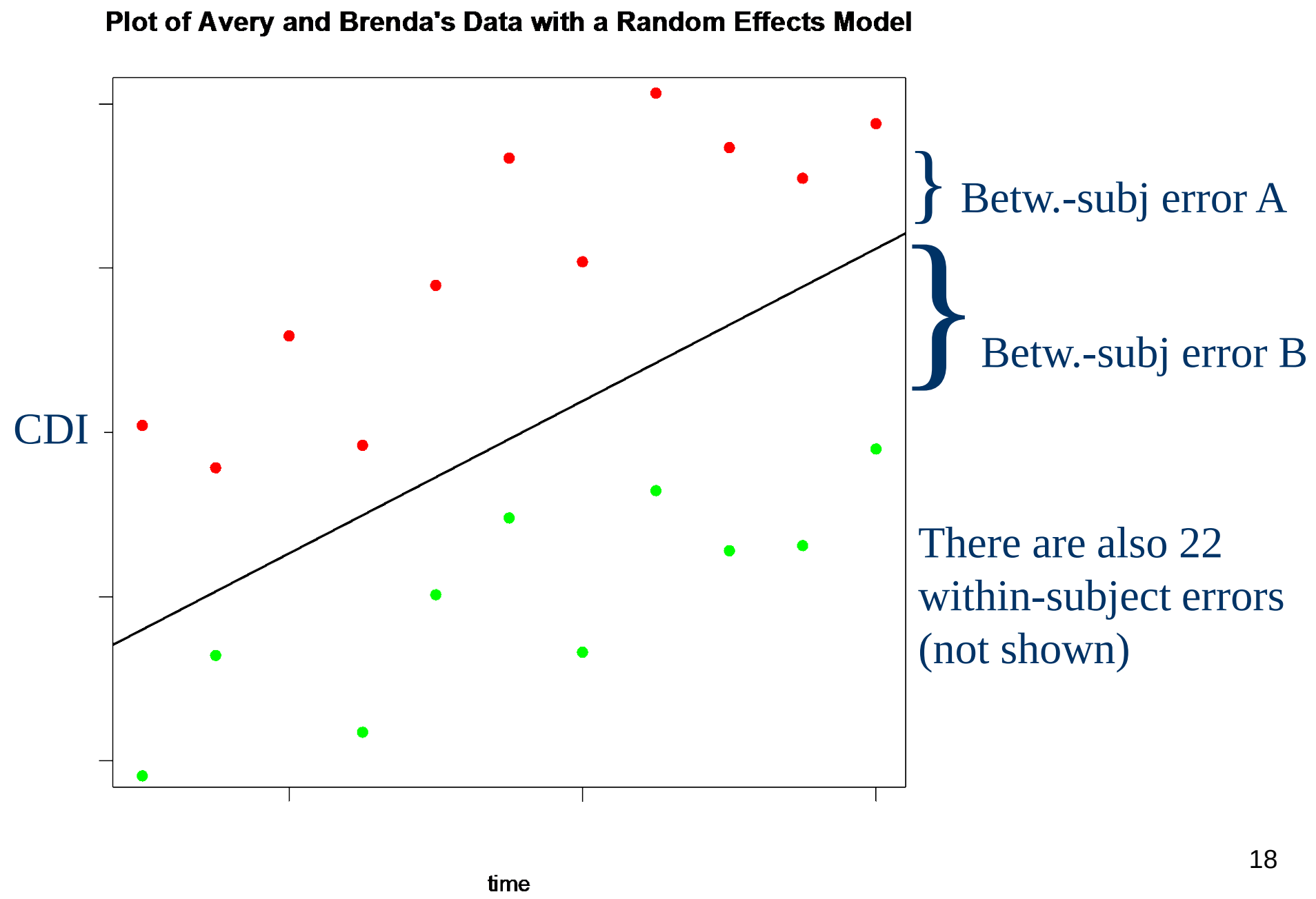
$$+ between\text{-}subject\ error$$

“Between-subject error” is new;

It allows for subjects to express individual variation as well as the usual ups and downs over time.

It is constant for each subject;

Here, it has one value for Avery and another for Brenda.


Plot of Avery and Brenda's Data with a Random Effects Model
CDI
time
Betw.-subj error A
Betw.-subj error B
There are also 22 within-subject errors (not shown)

D. Three useful answers

1. The example above showed only the case where the subject's regressions lines shifted up or down - the Random Intercept model.

Is there also a Random Slope model, analogous to fitting interaction terms for the fixed-effects model; where both the slope and intercept vary between subjects?

D. Three useful answers

1. The example above showed only the case where the subject's regressions lines shifted up or down - the Random Intercept model.

Is there also a Random Slope model, analogous to fitting interaction terms for the fixed-effects model; where both the slope and intercept vary between subjects?

YES.

2. What do you call a model with Random Effects for subjects and Fixed Effects for age, sex, markers etc.?

2. What do you call a model with Random Effects for subjects and Fixed Effects for age, sex, markers etc.?

A Mixed Effects model.

This is very common.

3. When can you use “derived variables”; when can you estimate the slope and intercept (or other summaries) for each subject and then do simple analyses of these?

E.g., can’t we just do a t-test of the difference between the slopes of marker + vs. marker - subjects?

3. When can you use “derived variables”; that is, suppose you estimate the slope and intercept for each subject and then do simple analyses of these.

E.g., can’t we just do a t-test of a difference between the slopes of marker + vs. marker - subjects?

This is appropriate when the study is balanced; that is, when every subject has measurements at a common set of ages with no missing values.

Much more often true in the lab than the clinic.