Week 9 Longitudinal Data

Stat 209
[redacted]

Time 1 - Time 2 Data, measurement of change

Observe Y_1, Y_2 (meas w/ error)

diff score

$Y_1 = \eta_1 + \varepsilon_1 \quad Y_2 = \eta_2 + \varepsilon_2 \quad \varepsilon \sim (0, \sigma^2)$

$D = Y_2 - Y_1$ unbiased for θ, when is unbiased not good enough?

Reliability of D $\quad rel(D) = (1 + 2\sigma_\varepsilon^2/\sigma_\theta^2)^{-1}$

$t_2 - t_1 = 1$

$\varepsilon_1, \varepsilon_2$ uncorr

big for σ_θ^2 big, can't detect individ diffs that don't exist

"Fixing" D, $R = Y_2 \cdot Y_1 = D \cdot Y_1$ useful identity

Residual change: How much change if all started out equal?

counterfactual

Exogenous var W, correlates of change

Pow: is W associated w/ growth? what kinds of people grow fastest?

Psychometrics: $Corr(\eta_2 \cdot \eta_1, W) = Corr(\Delta \cdot \eta_1, W)$

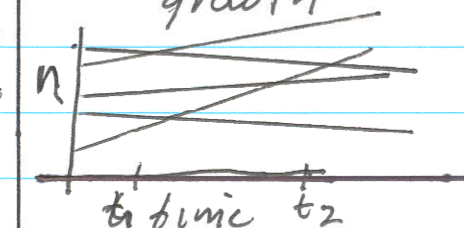
$Corr(\Delta W \cdot \eta_1) = Corr(\eta_2 \cdot \eta_1, W \cdot \eta_1) = Corr(\eta_2 W \cdot \eta_1)$

Underlying Growth Curve

Model: $\eta(t) = \eta(0) + \theta t$

$\theta_j, \eta_j(0)$ vary over indiv.

$\sigma_\theta^2 \to$ individ diffs in growth



$\Delta = \eta_2 - \eta_1 = \theta(t_2 - t_1)$

Thurs, HW9 fit collection of growth curves ($T > 2$) via lme, HLM, SAS etc.

Rogosa '85 Psych Mtka

Lord's Paradox (1967 - forever?)

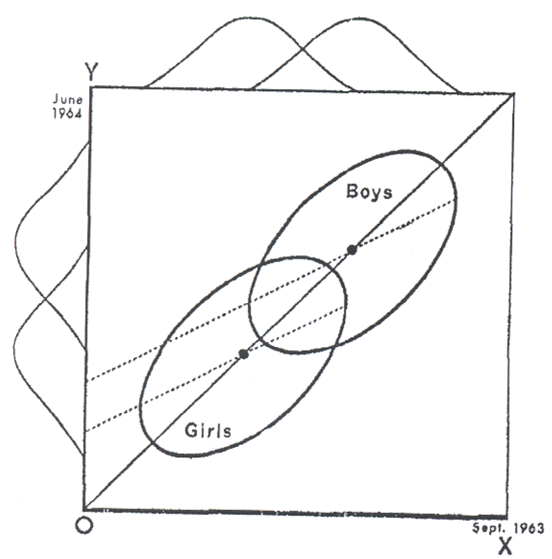


FIG. 1. Hypothetical scatterplots showing initial and final weight for boys and for girls.

A large university is interested in investigating the effects on the students of the diet provided in the university dining halls and any sex difference in these effects. Various types of data are gathered. In particular, the weight of each student at the time of his arrival in September and his weight the following June are recorded.

At the end of the school year, the data are independently examined by two statisticians. Both statisticians divide the students according to sex. The first statistician examines the mean weight of the girls at the beginning of the year and at the end of the year and finds these to be identical. On further investigation, he finds that the frequency distribution of weight for the girls at the end of the year is actually the same as it was at the beginning.

He finds the same to be true for the boys. Although the weight of individual boys and girls has usually changed during the course of the year, perhaps by a considerable amount, the group of girls considered as a whole has not changed in weight, nor has the group of boys. A sort of dynamic equilibrium has been maintained during the year.

The whole situation is shown by the solid lines in the diagram. Here the two ellipses

Distillation of Lord's Paradox
(and Neyman-Rubin-Holland) via Rubin, Wainer, Maris

Stat 209
[redacted]
p. 2

Wainer '91

		Treatment S	Sub-Population G	Outcome Y_t	Y_c	Concomitant Variable X_t	X_c
Population U of units	1 2 3 . . . N	t or c All diet t	1 or 2 male 1 fem 2 gender	June wt		Sept wt	

Figure 1. A framework for causal inference (From "On Lord's Paradox" [p. 5] by P. W Holland & D. B. Rubin, 1983, in H. Wainer & S. ... Principals of modern psychological measurement.

Average Causal Effect ACE

$E(Y_t - Y_c) = E(Y_t) - E(Y_c)$

observe in comparative studies

$E(Y_t | S=t) \quad E(Y_c | S=c)$

observables $\qquad$ potential outcomes

In whole pop $\quad E(Y_t) = E(Y_t | S=t)P(S=t) + E(Y_t | S=c)P(S=c)$

$E(Y_c) = E(Y_c | S=c)P(S=c) + E(Y_c | S=t)P\{S=t\}$

in Lord's paradox (also Wainer MCAT, rat heart rate) c doesn't occur. [no "control" diet]

Gender Effect $\qquad$ male $\qquad$ female

$\Delta = E(Y_t - Y_c | G=1) - E(Y_t - Y_c | G=2)$

$= \underbrace{E(Y_t | G=1)}_{\text{obs}} - \underbrace{E(Y_t | G=2)}_{\text{obs}} - E(Y_c | G=1) + E(Y_c | G=2)$

response function $\quad Y_c = X \quad$ under "control" diet same wt in June

$\Delta = E(Y_t | G=1) - E(Y_t | G=2)$
$- E(X | G=1) + E(X | G=2)$

diff of gains gives Δ

if instead response function $\quad Y_c = a + bX \quad$ wt in June linear funct of Sept.

$\Delta = E(Y_t | G=1) - E(Y_t | G=2)$
$- [E(a+bX | G=1) - E(a+bX | G=2)]$

or $\quad - b(E(X | G=1) - E(X | G=2))$

ancova works for Δ! (MCAT)

Rubin's Model for Causal Inference

The structure used to unravel this mystery involves Rubin's model (Rubin, 1974, 1977, 1978, 1980; Holland, 1986a, 1986b) for the analysis of causal effects. This model allows absolute explicitness about certain distinctions and elements that are often left implicit in other accounts. This model is not meant to find the cause of an effect; rather it tells how to measure the effect of a cause. This purpose is made explicit in Equation 1.

The basic elements of the model are as follows:

1. A population of units, U
2. An "experimental manipulation," with levels t or c, and its associated indicator variable, S
3. A subpopulation indicator variable, G
4. An outcome variable, Y
5. A concomitant variable, X.

This ...