Week 3 Math Facts
Part 2 Thurs

Stat 209

Gold Standard Tests for Structural Equation Models
using individual unit models (response functions)

## A. Holland (1988) Encouragement Designs

individual unit model: (potential outcomes)

$R_G(u) = R_c(u) + \rho(u) G$ ; $Y_{Gr}(u) = Y_{co}(u) + \tau(u) G + \beta(u) r$

params $\rho$ $\beta$ $\tau$

Path Analysis

using obs on G R Y

$$\begin{tikzcd} & R \arrow[dr, "\gamma_2"] & \varepsilon_R \arrow[l] & \\ G \arrow[ur, "\gamma_1"] \arrow[rr, "\gamma_3"'] & & Y & \varepsilon_Y \arrow[l] \end{tikzcd}$$

$R = \alpha_R + \gamma_1 G + \varepsilon_R$ $\quad$ $Y = \alpha_Y + \gamma_2 R + \gamma_3 G + \varepsilon_Y$

Compare $\gamma_1$ $\gamma_2$ $\gamma_3$
with $\rho(u)$ $\beta(u)$ $\tau(u)$

Under ALICE $\rho(u) = \rho$ $\beta(u) = \beta$ $\tau(u) = \tau$
results $\gamma_1 = \rho$ $\quad \gamma_2 = \beta + \delta$ $\quad \gamma_3 = \tau - \rho\delta$

bias due to $\delta = \mathrm{cov}(Y_{co}(u), R_c(u)) / \mathrm{Var}(R_c(u))$
individual diff's matter

overall effect $\tau + \beta\rho = \gamma_3 + \gamma_1 \gamma_2$
G on Y

direct/indirect $= \beta\rho/\tau$ vs $\dfrac{\rho(\beta+\delta)}{\tau - \rho\delta}$ path analysis

Numerical Examples (slides)

set $\beta = 3$ $\tau = 1$ $\mathrm{Corr}(R_c, Y_{co}) = .75$, $\mathrm{Var}(Y_{co}) = 64$
for $\rho = 3$, Var $R_c = 4$)

$\gamma_1 = 3$ $\quad \gamma_2 = 6$ $\quad \gamma_3 = -8$

ind/direct $= -2.25$ (vs 9)

more exs on slides

Does it work?

Stat 209 [illegible]XX

# Path Analysis and Encouragement Design

drr summary

## Encouragement Designs: Effects of Interventions

***Exemplar Study:*** Random assignment of students to treatment-control conditions for intervention on improving study habits. *Measures:* Treatment/control assignment (G), amount of study (R), and outcome measure, achievement test score (Y).

*Questions:* **1.** Increase in study time from intervention? **2.** Increase in achievement from studying an hour longer (dose response)? **3.** Increase in achievement if no increase in study (placebo effect)? **4.** Total impact on achievement?

| ***Path Analysis Regressions*** |
|---|
| ***Path Coefficients:*** $\gamma_1, \gamma_2, \gamma_3$ |
| $R = \alpha_R + \gamma_1 G + \varepsilon_R$ |
| $Y = \alpha_Y + \gamma_2 R + \gamma_3 G + \varepsilon_Y$ |

*Counterfactual Data Formulation for Individual u.* **1.** $R_t(u) - R_c(u) = \rho(u)$, treatment/control difference in amount of study. **2.** $Y_{Gr}(u) - Y_{Gr'}(u) = \beta(u)*(r - r')$, increment to outcome from study amount r′ vs r. **3.** $Y_{tr}(u) - Y_{cr}(u) = \tau(u)$, treatment/control difference in outcome with same amount of study r. **4.** $Y_{tR}(u) - Y_{cR}(u) = \tau(u) + \rho(u)\beta(u)$, overall treatment/control difference.

*Individual Level Model.* $R_G(u) = R_c(u) + \rho(u)G$ ; $Y_{Gr}(u) = Y_{c0}(u) + \tau(u)G + \beta(u)r$.

| |
|---|
| ALICE specification: $\rho(u) = \rho$; $\beta(u) = \beta$; $\tau(u) = \tau$ . |
| **Path Analysis Results Under ALICE.** $\gamma_1 = \rho$ ; $\gamma_2 = \beta + \delta$ ; $\gamma_3 = \tau - \rho\delta$. |
| Indirect/Direct Effects: $\beta\rho/\tau$ under ALICE; $\rho(\beta + \delta)/(\tau - \rho\delta)$ from path analysis. |

Holland (1988) p.469

$$\mu_c(r) = \gamma + \delta r. \tag{51}$$

A positive $\delta$ means that the more a student would study when not encouraged, the *higher* he or she would score on the test without studying and without encouragement. A negative $\delta$ means that the more a student would study when not encouraged, the *lower* he or she would score without studying and without encouragement.

The quantities computed in path analysis are the conditional expectations

S = 0/1 (G)

$$E(R_S|S) = E(R_c) + \rho S \tag{52}$$

and

$$E(Y_{SR_S}|S, R_S) = \mu_c(R_S - \rho S) + \tau S + \beta R_S, \tag{53}$$

in which $S$ is a 1/0 indicator variable. If we make the untestable assumption that $\mu_c(r)$ is linear, e.g. (51), then (53) becomes

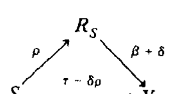
$$E(Y_{SR_S}|S, R_S) = \gamma + (\tau - \delta\rho)S + (\beta + \delta)R_S. \tag{54}$$

Equations (52) and (54) are both linear and may be combined into the empirical path diagram in Figure 5.

Comparing Figures 5 and 4, we see that even if the ALICE model holds and $\mu_c(r)$ is linear, the estimated path coefficients are

meaning of δ
individual differences
in outcome if
null R, G

FIGURE 5.



## B Longitudinal Data (panel data)

observe $X_1\ X_2\ X_3$ measured w/out error times $t_1\ t_2\ t_3$
following straight-line growth curve

$X_p(t) = X_p(0) + \theta_p t$ , individual unit model with individual differences

Goldstein-style path analysisis (canonical ex)

$$\begin{tikzcd} & \varepsilon_2 \arrow[d] & & \varepsilon_3 \arrow[d] \\ X_1 \arrow[r, "\beta_1"] \arrow[rrr, bend right, "\beta_3"'] & X_2 \arrow[rr, "\beta_2"] & & X_3 \end{tikzcd}$$

temporal = causal ordering

$$X_2 = \alpha_2 + \beta_1 X_1 + \varepsilon_2$$

$$X_3 = \alpha_3 + \beta_2 X_2 + \beta_3 X_1 + \varepsilon_3$$

Result (Rogosa 1989) $\beta_3 = \dfrac{t_2 - t_3}{t_2 - t_1}$ $\beta_2 = \dfrac{t_3 - t_1}{t_2 - t_1}$

also $R^2_{3\cdot 21} = 1$

modified results if X measured w/ error (Trento paper)
also individual unit model exponential growth (handout)

---

Simplex Model Example (Rogosa, Willett 1984)

auto regressive lag-1 $\eta_{i+1,p} = \beta_i \eta_p + \delta_{i+1,p} \Rightarrow \rho_{\eta_1 \eta_3 \cdot \eta_2} = 0$

fit via LISREL (see slides) to data from maximally "unsimplex" straight-line growth

$$\eta_{ip} = \eta_{1p} + \theta_p (t_i - t_1) \Rightarrow \rho_{\eta_1 \eta_3 \cdot \eta_2} = -1$$

wonderful fit from LISREL
(cf Breckler examples)

# Path Analysis II
# Longitudinal Data (Goldstein ea)

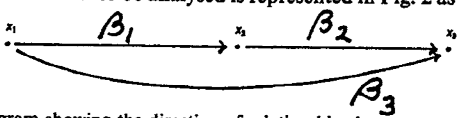
Rogosa 1988 Casual models.....

Does it work?

1979 GOLDSTEIN – *Analysing Longitudinal Data* 411

5. MODELS

The first and simplest model to be analysed is represented in Fig. 2 as a path diagram.




$n = 17000$
ages 7 11 16

FIG. 2. Path diagram showing the direction of relationships between measurements at three occasions.

Rogosa (1988)
individual-level model straight-line growth in reading

specifies a constant rate of change denoted by $\theta$. The straight-line growth curve for individual $p$ is written:

$$\xi_p(t) = \xi_p(0) + \theta_p t. \tag{1}$$

*Path Regressions*

Path analysis models for longitudinal data use the temporal ordering of the measurements to delimit the possible paths between the variables. Consider the example of a three-wave design with measures on $X$ at times $t_1$, $t_2$, $t_3$. The path regressions for the unstandardized variables are:

$$\begin{aligned} X_2 &= \alpha_2 + \beta_1 X_1 + e_2 \\ X_3 &= \alpha_3 + \beta_2 X_2 + \beta_3 X_1 + e_3. \end{aligned} \tag{2}$$

gression (path) coefficients from (2) are:

$$\beta_3 = \frac{t_2 - t_3}{t_2 - t_1} < 0$$

and $R^2 = 1$

$$\beta_2 = \frac{t_3 - t_1}{t_2 - t_1} > 0. \tag{3}$$

data-free causal analysis

[Note: the results in (3) are easily verified by using (1) to substitute $X_2 = X_1 + \theta(t_2 - t_1)$ into the $X_3$ equation in (2). Substituting the values for $\beta_2$ and $\beta_3$ from (3) and collecting terms yields $X_3 = X_1 + \theta(t_3 - t_1)$.]

additional results in Trento (dw'93)

### 3.2. Exponential growth

**For a population of individual growth curves given by (2) the population partial regression (structural) coefficients for the $\eta_3$ equation in (3) are:**

$$\beta_3 = \frac{\exp[-\gamma t_3] - \exp[-\gamma t_2]}{\exp[-\gamma t_1] - \exp[-\gamma t_2]} > 0$$

individ growth

$$\eta_p(t) = \lambda_p - (\lambda_p - \eta_p(0))e^{-\gamma t}$$

just depends on curvature

$$\beta_2 = \frac{\exp[-\gamma t_3] - \exp[-\gamma t_1]}{\exp[-\gamma t_2] - \exp[-\gamma t_1]} < 0.$$

asymptote $\lambda_p$ curvature $\gamma$

(7) $\frac{d\eta}{dt}$ linear state dependence