Week 7 Math Notes

Stat 209

Formulation for Compliance Analyses

Bool Jo version — assume no controls get treatment

DAP: single cross-over Greenland Vitamin A

$Z = 1, 0$ T, C $\quad \pi_c = P(T|Z=1)$ compliance

T μ_1, (A) μ_{c1} (B) μ_{n1} $\quad \mu_1 = \pi_c \mu_{c1} + (1-\pi_c)\mu_{n1}$

C μ_0, (a) μ_{c0} (b) μ_{n0} $\quad \mu_0 = \pi_c \mu_{c0} + (1-\pi_c)\mu_{n0}$ (unobserved)

$ITT = \mu_1 - \mu_0 = \pi_c(\mu_{c1} - \mu_{c0}) + (1-\pi_c)(\mu_{n1} - \mu_{n0})$

$CACE = \mu_{c1} - \mu_{c0} = \frac{\mu_1 - \mu_0}{\pi_c}$ iff $\mu_{n1} = \mu_{n0}$ no effect of assignment

HIV ex: $\frac{\hat{ITT}}{\hat{\pi}_c} = \frac{.364}{.457} = .76$ IV est, AIR c.f AIR handout 2/19

David Freedman Analysis (modelobs talk, OXcauses paper)

Neyman model

Potential outcomes: overall in population

$\bar{T} = Ave(T_i)$ $\bar{C} = Ave(C_i)$

randomize to T, C

ITT estimates $= \bar{T} - \bar{C}$

	Group	Number	T	C
(Lisa)	Always Treat	αN	a	a
(Margo)	Compliers	βN	τ	c
(Homer)	Never-Treat	γN	η	η
(Bart)	Defiers	θN	τ_D	c_D

Ave response it in

IV estimate (pp. 704-5) "oxcauser"

assume: single crossover, $\alpha = \theta = 0 \Rightarrow \beta + \gamma = 1$, no Lisas, Barts

result $\hat{c} = (\bar{Y}^c - \gamma\hat{\eta})/\beta$, $\hat{\tau} = (\bar{Y}^T - \gamma\hat{\eta})/\beta \Rightarrow \hat{\tau} - \hat{c} = \frac{\bar{Y}^T - \bar{Y}^c}{\beta}$ (IV)

ave response control (sample)

CACE

per-protocol (A vs a+b)

$$\frac{\alpha a + \beta\tau}{\alpha+\beta} - \frac{\beta c + \gamma\eta}{\beta+\gamma}$$

if trial blind may work

As-treated (A vs a+b+B) $\quad \lambda = N_T/N_C$

$$\frac{\alpha\lambda a + \beta\lambda\tau + \alpha a + \theta c_D}{\alpha\lambda + \beta\lambda + \alpha + \theta} - \frac{\beta c + \gamma\eta + \gamma\lambda\eta + \theta\lambda\tau_D}{\beta + \gamma + \gamma\lambda + \theta\lambda}$$

Freedman see scelo "Statistical models for causation" Evaluation Review see 10

p. 706.

Two Achilles' Heels

Internal vs external validity: the study population may not be representative.

A threat to internal validity is crossover: some people assigned to treatment decline treatment, some controls insist on treatment.

The intention-to-treat principle is a response to the crossover problem: you measure the effect of assignment, not treatment

Other estimators

(i) per protocol, (ii) treatment received, (iii) IV to estimate effect of treatment

Summary on the other estimators

Per protocol & treatment received. Unless you have very good blinding, these are very bad options.

The IV estimator. Pretty good—if you have a 0–1 response, single crossover, no blocking. With multi-level response, double crossover, or blocking, it's a lot less clear what's being estimated.

"Blocking" means, randomize subjects within (small) strata. It's the least of the issues here.

Calibrate using the Neyman model

Some would say, the Rubin model, but this mistakes the history.

> D Dabrowska and TP Speed (1990). On the application of probability theory to agricultural experiments. Essay on principles. English translation of Neyman (1923). *Statistical Science*, 5: 463–80 (with discussion).

Index subjects by i running from 1 to N. If subject i is assigned to treatment, the response is T_i; if assigned to control, the response is C_i. If all subjects are assigned to treatment, the average response is

$$\overline{T} = \frac{1}{N}\sum_{i=1}^{N} T_i.$$

If all are assigned to control, the average response is

$$\overline{C} = \frac{1}{N}\sum_{i=1}^{N} C_i.$$

The intention-to-treat parameter is $\overline{T} - \overline{C}$. The mean in the treatment group minus the mean in the control group is an unbiased estimate: this is a theorem, not a tautology.

Let's say (i) open-label trial (everybody knows treatment status), (ii) response is 0–1 and so is compliance, (iii) response is to treatment not assignment, (iv) randomize some subjects to T = treatment, rest to C = control.

Group	No.	Ave. response if assigned to T	Ave. response if assigned to C
Always-treat	αN	A	A
Compliers	βN	T	C
Never-treat	γN	N	N
Defiers	θN	$\mathfrak{T}$	$\mathfrak{C}$

N is the number of subjects. The fractions $\alpha, \beta\ \gamma, \theta$ are parameters, constrained to be nonnegative, sum equals 1. The gothic (and very gothic) letters are parameters too. Not all identifiable.

Per-protocol estimand is

$$\frac{\alpha\mathrm{A} + \beta\mathrm{T}}{\alpha + \beta} - \frac{\beta\mathrm{C} + \gamma\mathrm{N}}{\beta + \gamma}$$

Treatment-received estimand is

$$\frac{\alpha\lambda\mathrm{A} + \beta\lambda\mathrm{T} + \alpha\mathrm{A} + \theta\mathfrak{C}}{\alpha\lambda + \beta\lambda + \alpha + \theta} - \frac{\beta\mathrm{C} + \gamma\mathrm{N} + \gamma\lambda\mathrm{N} + \theta\lambda\mathfrak{T}}{\beta + \gamma + \gamma\lambda + \theta\lambda}$$

Do these formulas look useless? Maybe that's because the estimators are useless. . . .

If there are no defiers, e.g., single crossover, IV estimand is T − C