cf. Ch19 in Rubin book: Causal Inference in Retrospective Studies

Math Notes Week 2

Stat 209

p.475 Neyman-Rubin-Holland Formulation From Holland (1988)

potential outcome $Y(u,s)$ response observed *if* unit u exposed to cause $s \in K$

U units in pop
K causes (treatments)
S assignment rule

p.476 actual exposure $S(u)$, observed response $Y(u, S(u)) = Y_S(u)$

Unit-level Causal effect (t, c in K)

$$Y(u,t) - Y(u,c) = T_{tc}(u)$$

unobserved counterfactual can't observe both

p.477 Unit homogeneity (lab setting)

$Y(u,s) = Y(v,s)$ all $s \in K$, $u, v \in U$

Causal Effect $T_{tc}(u) = Y_t(u) - Y_c(v)$ constant effect

Fisher's H_0 $T_{tc}(u) = 0$ $t,c \in K$, $u \in U$

---

p.478 ACE (p.478) unobservable average over u

$P(S=s)$ proportion units $S(u)=s$

$E(Y_S | S=t)$ average value of Y for units $S(u)=t$

$$ACE_{tc} = E(T_{tc}) \qquad ACE_{tc}(Y) = E(Y_t) - E(Y_c)$$

aka ATE (econ)

observable

$$FACE_{tc}(Y) = E(Y_t | S=t) - E(Y_c | S=c)$$

p.479 randomization makes independence assumption, S indep Y, plausible, so that $FACE_{tc}(Y) = ACE_{tc}(Y)$

p.480 more detail: for constant effect $T_{tc}(u) = \tau_{tc}$ same for all u

$$ACE_{tc}(Y) = \tau_{tc}$$

$$FACE_{tc}(Y) = \tau_{tc} + \{E(Y_c | S=t) - E(Y_c | S=c)\}$$

value of Y in control condition for treatment subjects

Covariates: X $X(u,s)$ does not depend on s "exogenous"

$$ATET = E(Y_t | S=t) - E(Y_c | S=t)$$ counterfactual MIT p.3

Math Notes p. 2 Week 2 Stat 209

Observational Studies
(week 5)

S_w not arranged
not indep of Y_s

pretest equivalence: $P\{X = x \mid S = s\}$ does not depend on s

conditional independence given X (strong ignorability)

$$\text{C-FACE}_{tc}(Y) = E\{E(Y_s \mid S = t, X) - E(Y_s \mid S = c, X)\}$$

what conditions equals $\text{ACE}_{tc}(Y)$
(regression, discontinuity, model selection function)

Freedman at Stanford

What would it take to make this stick?
Response schedules and invariance. Potential outcomes

There are two treatments (levels u and v), and a response variable Y. Both treatments may be applied to subject i. There are three parameters, a, b, and c. With no treatment at all, response level for subject i is a, up to random error. Each additional unit of treatment #1 adds b to the response. Likewise, each additional unit of treatment #2 adds c to the response. Constancy of parameters across subjects and levels of treatment is an assumption. If treatment #1 is applied at level u and #2 at level v, response is "potential"

$$Y_{i,u,v} = a + bu + cv + \epsilon_i.$$

Invariance of a, b, c, ϵ_i? My response is unaffected by your treatment?? Manipulation???

Statistical assumptions

In order to make the transition from a hypothetical experiment to the actual observational study, and to justify OLS, we assume:

(i) $E(\epsilon_i) = 0$.
(ii) ϵ_i are independent and identically distributed across subjects i.
(iii) Exogeneity. Nature chooses U_i, V_i independently of the random errors ϵ_i, and determines the response Y_i from the response schedule:

$$Y_i = Y_{i,U_i,V_i} = a + bU_i + cV_i + \epsilon_i.$$

Nature shows us U_i, V_i, Y_i. We're good to go. a) OLS works. b) Causal inferences justified—built into the response schedule. (With small samples, need to assume errors are normal.)