

Simple Randomized Trials (full compliance)

Consequences of randomization (using potential outcomes framework)

Let R denote a subject's randomization indicator (in simple trials, $R=A$ for all subjects)

Write $\underline{Y} \equiv \{Y^1, Y^0\}$ for vector of all potential outcomes

What can one say about association of R and potential outcomes \underline{Y} ?

Treatment assignment under randomization is independent of the potential outcomes:

Potential outcomes considered fixed before randomization; thus, randomization cannot affect potential outcomes, just what is observed

Different but equivalent ways to express independence:

1. $\underline{Y} \perp\!\!\!\perp R$ (independence:
 $A \perp\!\!\!\perp B | C$: read A is independent of B given C)
2. $pr(R=1|\underline{Y})=pr(R=1)$ (ignorability)
Note: for strong ignorability, require $0 < pr(R=1) < 1$; will discuss later
3. $f(\underline{Y}|R=1)=f(\underline{Y}|R=0)=f(\underline{Y})$ (comparability)

Subjects randomized to level $R=1$ (or $R=0$) are representative of all subjects with respect to potential outcomes \underline{Y}

Consequences of comparability/representativeness $f(\underline{Y}|R=1)=f(\underline{Y}|R=0)=f(\underline{Y})$

$$f(Y^1)=f(Y^1|R=1)=f(Y|R=1)$$

The observed outcomes among subjects randomized to receive level $R=1$ of treatment are representative of the potential outcomes Y^1 in the population

Similarly for $R=0$, $f(Y^0)=f(Y|R=0)$

Consequences: under randomization we can use observable outcomes Y , together with randomization indicator R , to learn about (identify) distribution of potential outcomes, and thereby about causal effects

Which type of causal effects?

1. Comparison of distributions of potential outcomes in population (population averaged)

or

2. Distributions of contrasts in potential outcomes in population (subject-specific; causal types)

Comparison of distributions of potential outcomes in population (population averaged)

Cannot (usually) learn precisely about distribution of causal types (joint distribution of potential outcomes)

Can get bounds/limits: e.g., the proportion of subjects for whom treatment causative (i.e., $Y^0=0$, $Y^1=1$) is no less than the causal risk difference $pr(Y^1=1) - pr(Y^1=0)$ and no greater than the proportion of subjects for whom $Y^1=1$ (i.e., $pr(Y^1=1)$); can obtain tighter bounds

Each of these quantities equals an observable quantity in a randomized trial

Consequences for attributable fractions: cannot (under these assumptions) learn about the proportion of subjects for whom treatment causes outcome (but can get limits on these quantities)

In contrast, can learn about aggregate effects: excess in number/proportion of cases because of treatment/exposure

Names “attributable fraction” and “attributable fraction in the exposed,” when applied to usual formula (e.g., $1 - 1/RR$) is somewhat misleading; “excess fraction” (in exposed) would be more accurate (but has not been widely adopted in practice)

The formulas for “attributable fraction” in the exposed actually calculate the excess fraction in the exposed, which is no greater than the proportion of the exposed with the outcome whose outcome is due to exposure

$$1 - \frac{1}{RR} = 1 - \frac{pr(Y=1|R=0)}{pr(Y=1|R=1)} = 1 - \frac{pr(Y^0=1)}{pr(Y^1=1)} = \frac{pr(Y^1=1) - pr(Y^0=1)}{pr(Y^1=1)} \leq \frac{pr(Y^1=1, Y^0=0)}{pr(Y^1=1)}$$

Statistical inference:

In randomized trials with very large populations, can estimate population causal parameters (e.g., causal risks and risk differences) simply by observing the corresponding empirical quantities in the different randomized groups

In practical-sized studies, require methods for statistical inference which take into account uncertainty due to random variation/fluctuations

How would one learn about causal effects in finite samples, make probability or confidence statements about causal parameters in randomized trial?

Methods of analysis:

1. Standard statistical modeling/estimation
2. Bayesian analysis
3. Tests of ignorable treatment assignment

Will try to unify 1 and 3 later

Write likelihood/probability model for joint distribution of potential outcomes \underline{Y} , treatment R in randomized trial: $f(\underline{Y}, R)$

2 general approaches: please suggest

how will this relate to likelihood for observed outcomes?

$$f(\underline{Y}, R) = f(R)f(\underline{Y}|R)$$

$$= f(\underline{Y})f(R|\underline{Y})$$

$$= f(R)f(\underline{Y})$$

(Under randomization)

likelihood for observed data obtained by integrating out missing/counterfactual outcome Y^{1-R}

$$f(Y, R) = \int_{Y^{1-R}} f(\underline{Y}, R) dY^{1-R} = \int_{Y^{1-R}} f(R) f(Y^0, Y^1 | R) dY^{1-R}$$

$$= \int_{Y^{1-R}} f(R) f(Y^R, Y^{1-R} | R) dY^{1-R} = f(R) f(Y^R | R) = f(R) f(Y | R)$$

can use all these formulations

1st: standard inference (e.g., likelihood based)

use “forward” factorization to learn about aggregate effects:

$$f(Y,R) = f(R)f(Y|R)$$

should one use information about distribution of R (i.e., first part of likelihood) in inference about causal effects?

In likelihood-based inference, one would generally not

information about causal effects is contained in $f(Y|R)$

if have distinct parameters $\alpha, \beta : f(Y, R; \alpha, \beta) = f(R; \alpha) f(Y|R; \beta)$

$f(R; \alpha)$ is ancillary; plays no role in likelihood-based inference (demonstrate)

can use usual likelihood-based methods for estimating $f(Y|R; \beta)$

Tests of ignorable treatment assignment:

strongly ignorable treatment assignment: $pr(R=1|\underline{Y})=pr(R=1)$

not strictly testable: \underline{Y} not observable

under various theories about causal effects, can test

strongly ignorable treatment assignment implies $pr(R=1|Y^0)=pr(R=1)$ (Will explore such assumptions in more detail later)

what if treatment has no effect on outcome for any subject?

Rewrite above assumption:

$$Y^0 = Y$$

then

$$pr(R=1|Y) = pr(R=1)$$

then, tests of above are tests of null hypothesis

how can one test above?

Permutation tests

Large-sample tests

Permutation/exact tests:

for each subject, have pair $\{Y_i, R_i\}$

compute test statistic that compares outcomes among subjects with $R=1$ and

$$R=0; \text{ e.g., } T = \frac{\sum_i Y_i R_i}{\sum_i R_i} - \frac{\sum_i Y_i (1-R_i)}{\sum_i (1-R_i)}$$

Permutation distribution of test statistic: keep subjects' values of Y the same, permute assignment of R

Let $N = \sum_i 1$ be # of subjects in study

Let $Z = \sum_i R$ be # of treated subjects

Let $J \equiv \binom{N}{Z}$ be # of ways of choosing Z treated subjects from N subjects
(sampling without replacement)

Let R_i^j denote the value of R subject i has in j th permutation

$$T^j = \frac{\sum_i Y_i R_i^j}{\sum_i R_i^j} - \frac{\sum_i Y_i (1 - R_i^j)}{\sum_i (1 - R_i^j)} : \text{value of test statistic in } j\text{th permutation}$$

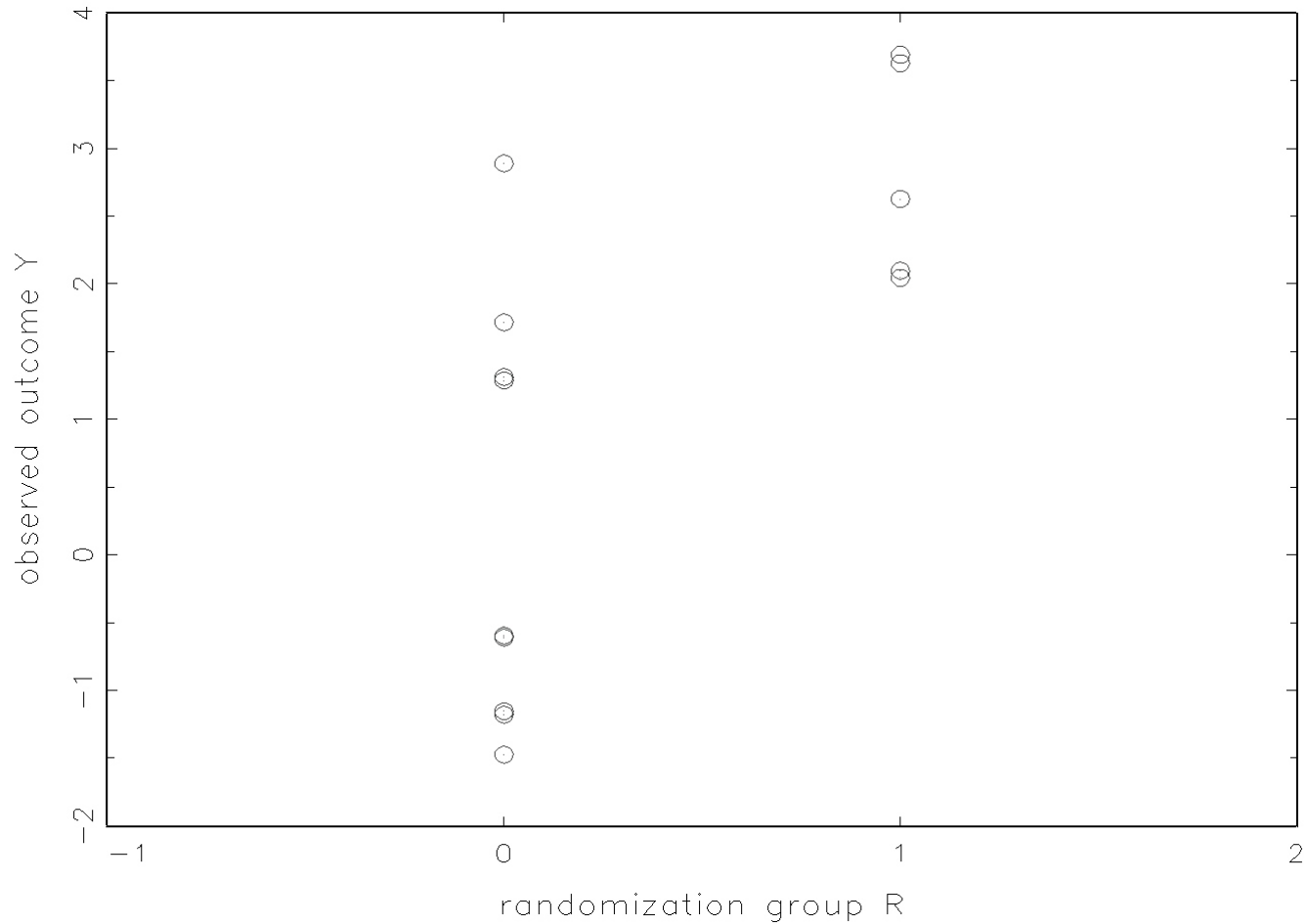
permutation distribution: $pr(T^j \leq t) = J^{-1} \sum_j I(T^j \leq t)$

get p-value by comparing observed value of T to permutation distribution

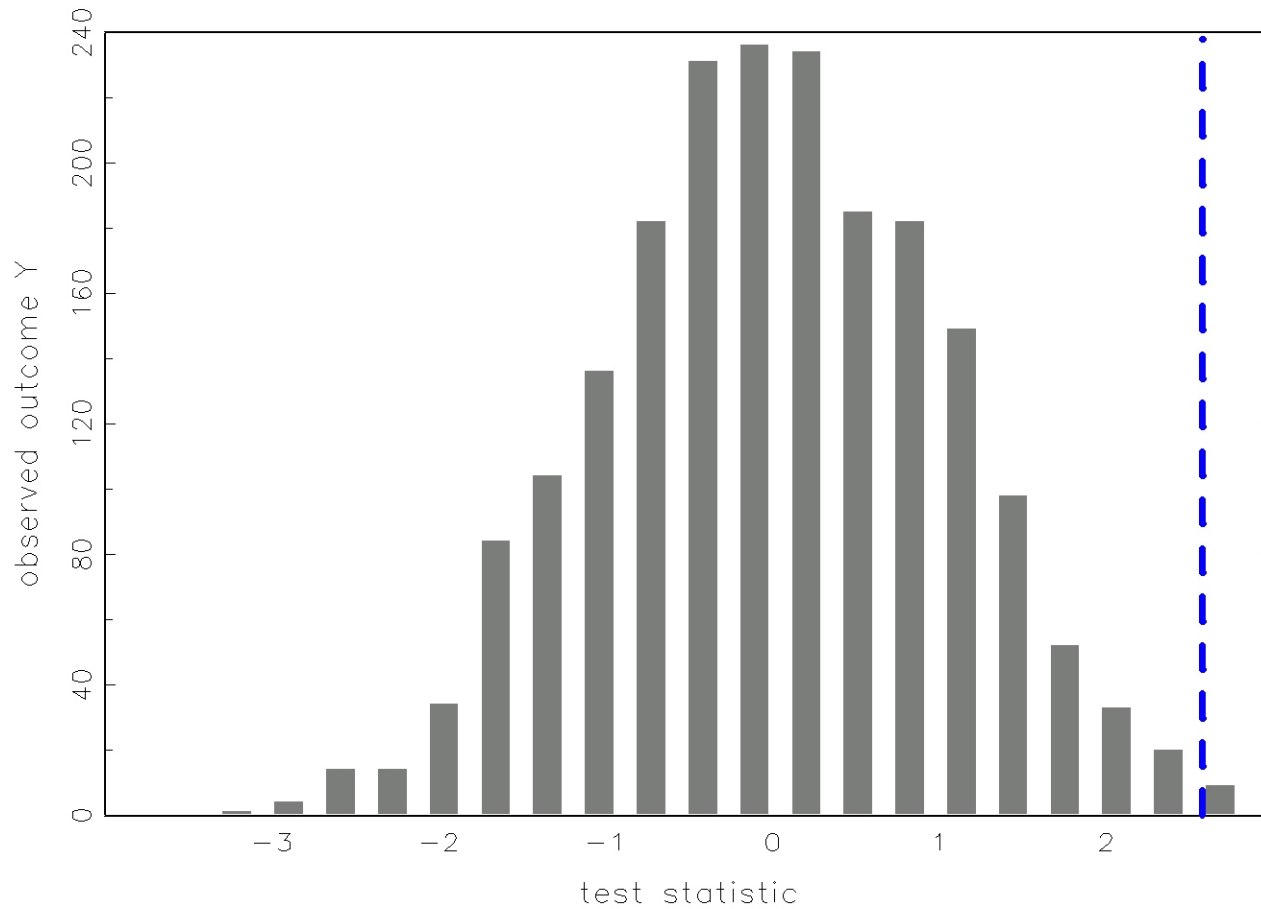
Illustrate: 14 subjects, 5 treated

Raw data:

R	
0	1
2.89	2.63
1.72	3.63
-0.60	2.04
-1.18	2.09
-1.16	3.69
-1.48	
1.31	
-0.61	
1.29	
Mean	
0.24	2.82



$\binom{14}{5} = 2002$ permutations



observed value (dashed blue line) 7th largest; p (2-sided) = 0.007

What is source of probability in this model/approach?

To whom does inference apply?

Randomization/permutation is source of probability here

Inference formally applies to subjects studied

Suppose that randomization done individually (e.g., by flip of a coin)

Why should consider only possibilities in which # of treated subjects Z equals observed value?

Z is ancillary statistic

in general, preferable to have inference conditional on observed value of ancillary statistic

classic case: inference for mean μ of distribution:

$Y \sim N(\mu, 1)$ if $X=0$, $Y \sim N(\mu, 100)$ if $X=1$; $pr(X=1)=0.5$

observe X (common for all subjects), Y

in inference for μ , would not want to consider hypothetical repetitions in which X took some value other than that observed

consider hypothetical repetitions that are most similar to those observed

Consider other assumed causal models:

simple shift model:

$$Y^1 = Y^0 + \Psi$$

derive formula for Y^0 using observable quantities $Y, R; \Psi$

$$Y^0 = Y - R\Psi$$

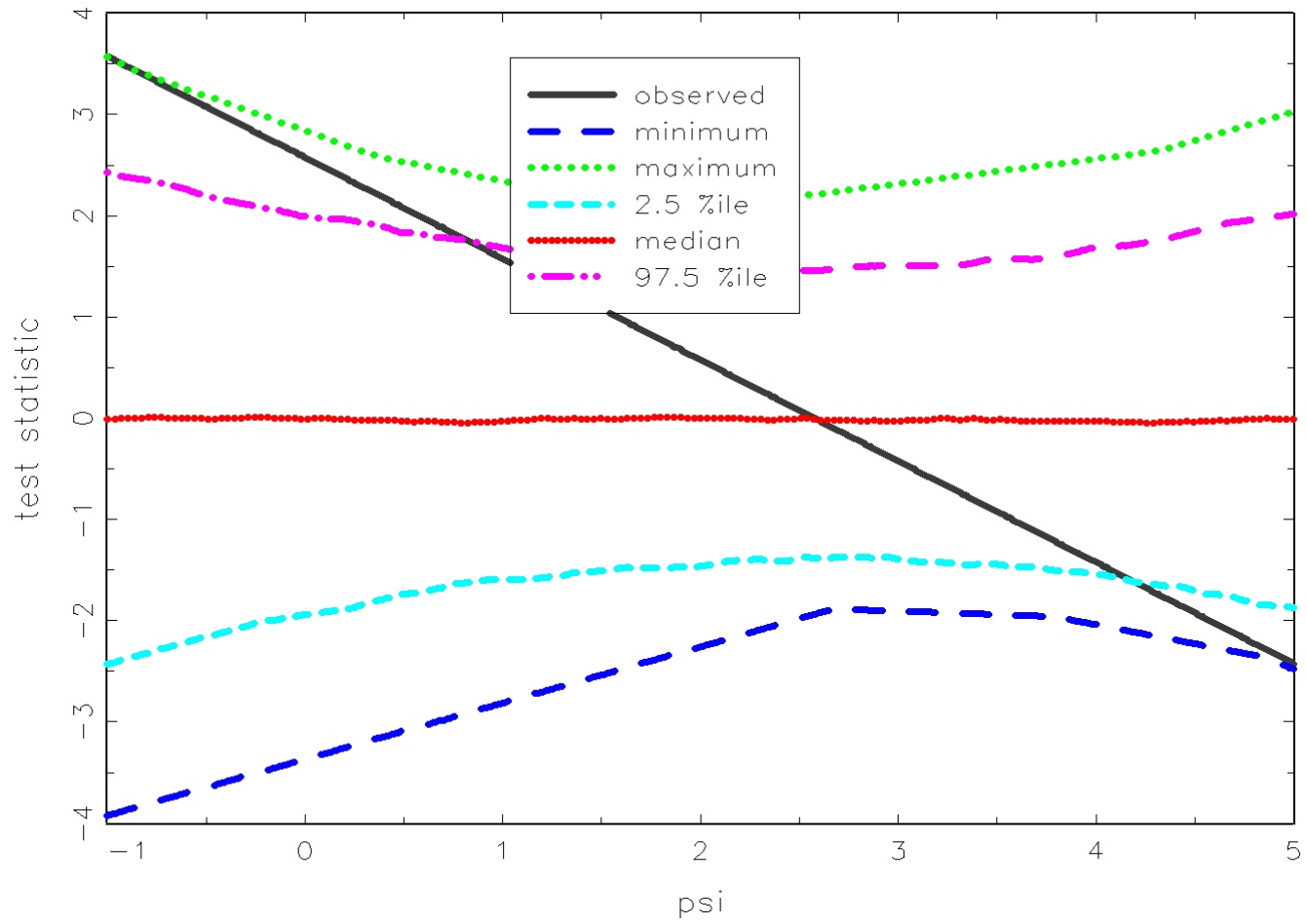
how could one test whether some putative value for Ψ is correct?

Derive putative value for baseline potential outcome $Y^0(\Psi) = Y - R\Psi$

If causal theory indexed by Ψ is true, $Y^0(\Psi)$ will be independent of R

Test independence of $Y^0(\Psi)$ and R using permutation test

Illustrate:

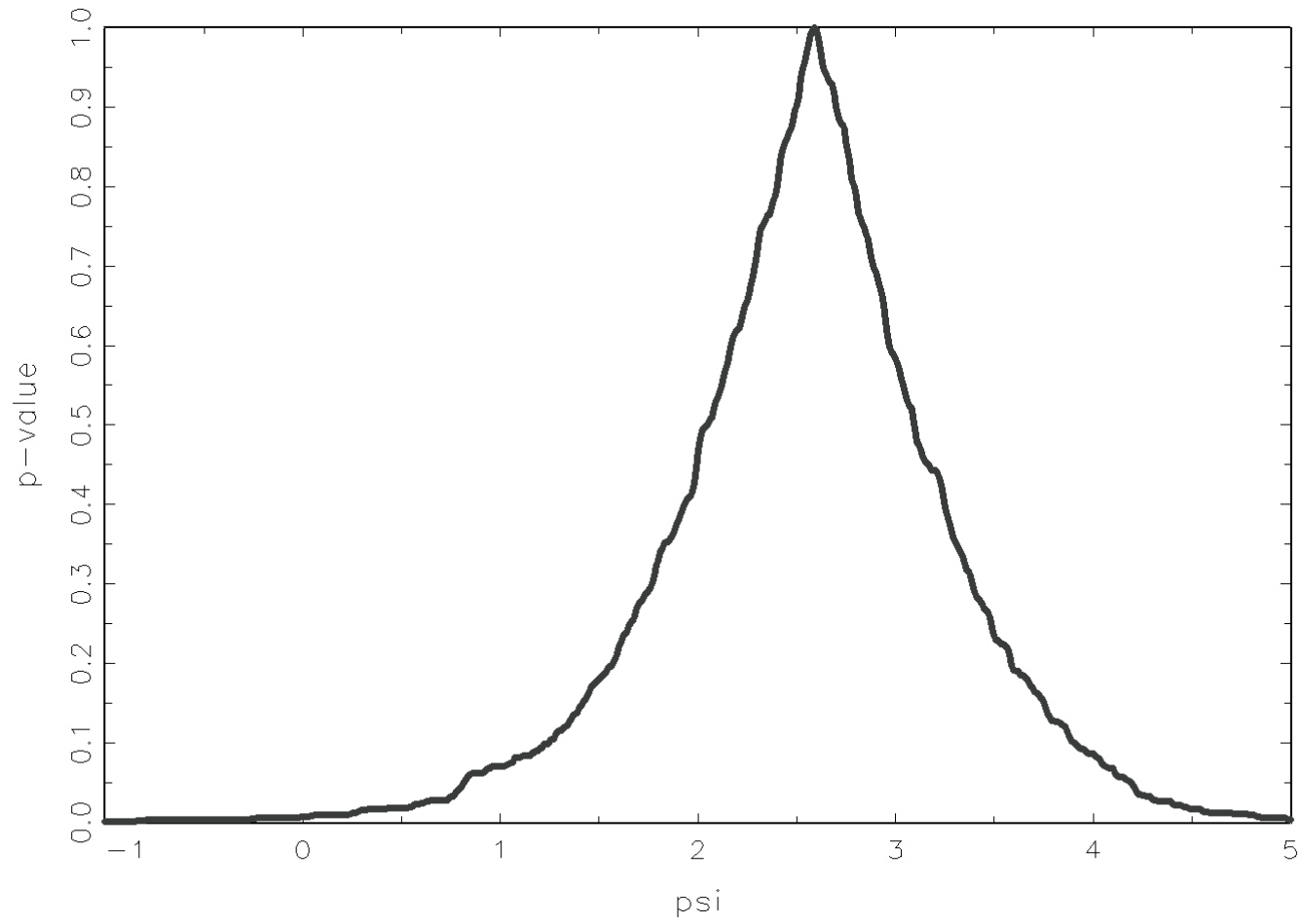


How would one obtain point estimate, confidence interval?

Point estimate: value of Ψ for which p-value closest to 0

$(1-\alpha)*100\%$ confidence interval: values of Ψ for which α level test fails to reject

convert to p-value curve:



Can also use large-sample test of independence

will sketch one approach

recall: $pr(R=1|Y^0)=pr(R=1)$;

implies $pr\{R=1|Y^0(\Psi)\}=pr(R=1)$ for true Ψ

consider logistic regression model: $logit[pr\{R=1|Y^0(\Psi)\}]=\alpha+Y^0(\Psi)\beta$

what should β be

- If Ψ is true value? Why?
- If Ψ is some other value? Why?

If Ψ is true value, $Y^0(\Psi)=Y^0$, which is not associated with R

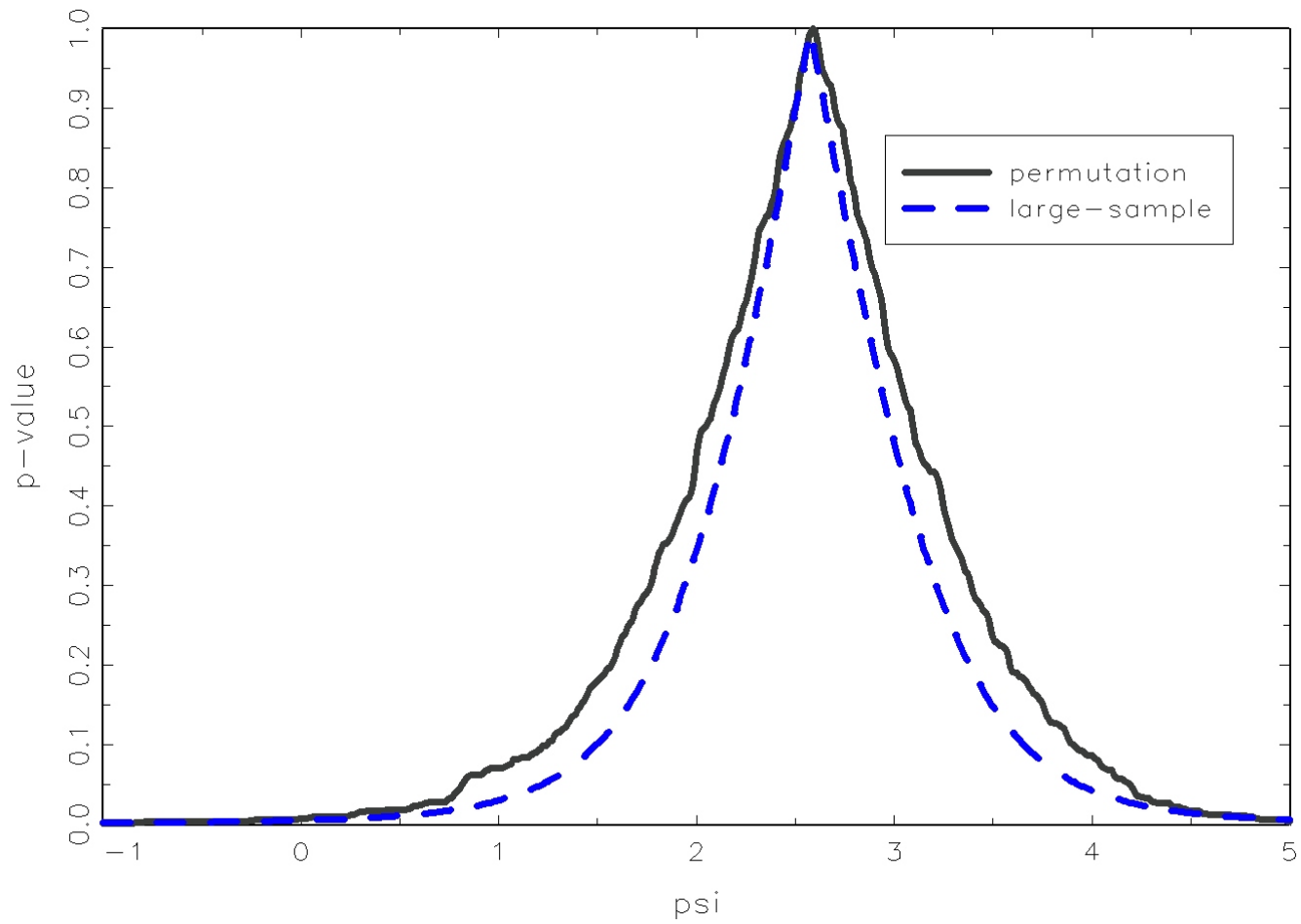
Thus, $\beta=0$

For other values of $\Psi \neq \Psi_0$ (true value), $Y^0(\Psi)=Y^0-R(\Psi-\Psi_0)$

Then, $Y^0(\Psi)$ is associated with R , and $\beta \neq 0$

Do large-sample test of $\beta=0$; invert test

Score test



compare with permutation test

Can also show score statistic:

$$E\{\sum (R-p)Y^0(\Psi)\} = 0$$

If randomization probability known, how would one get variance of score statistic?

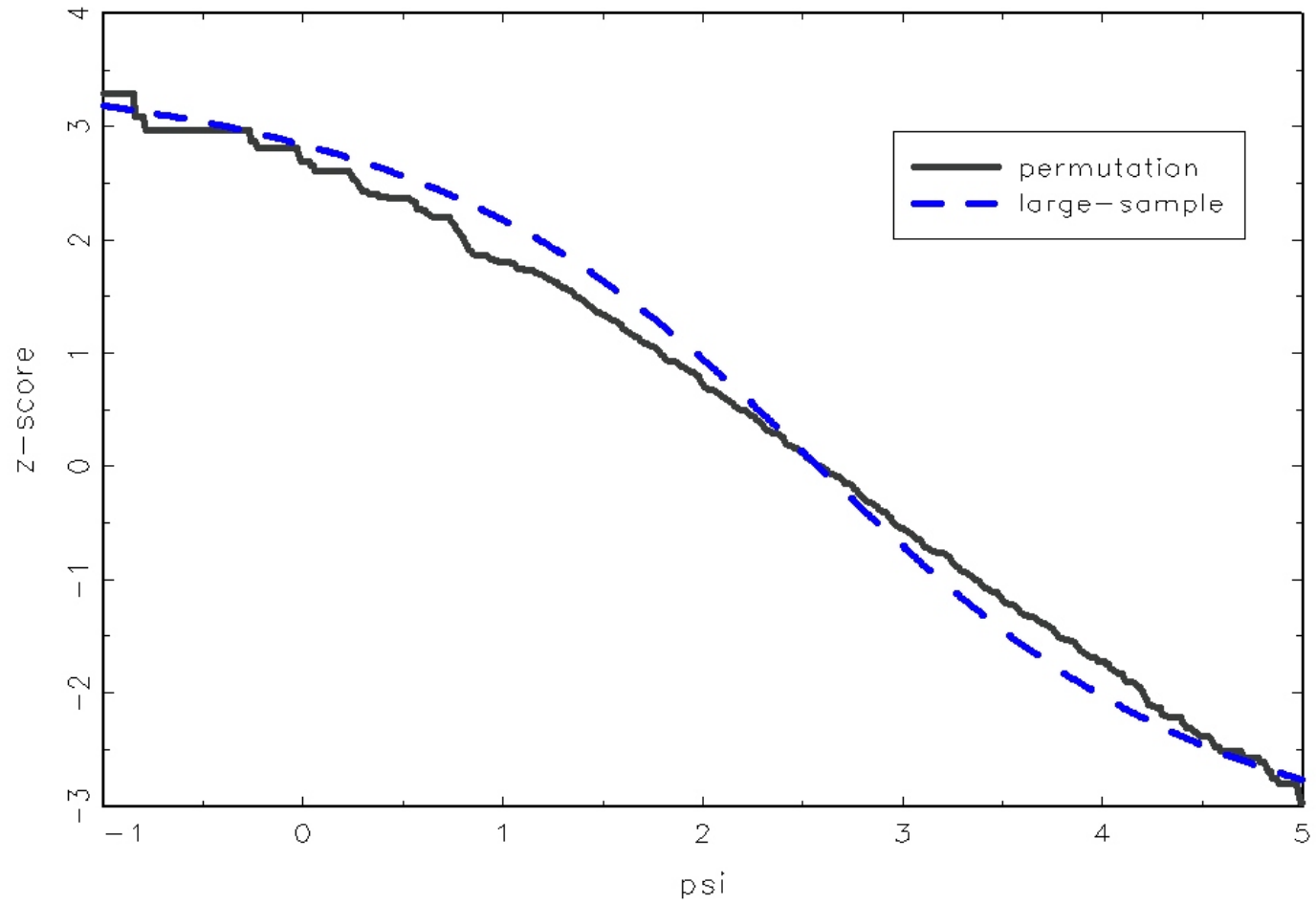
Treat $Y^0(\Psi)$ as fixed

from logistic regression, take derivative of equation with respect to β

or use sum:

$$E\left\{\sum S_i(\Psi)^2\right\}, \text{ where } S_i(\Psi) = (R-p)Y^0(\Psi)$$

show score statistic curve:



Does inference depend on knowing anything about distribution of potential outcomes Y^0 ?

Validity of argument above does not depend on knowing distribution

Thus, can get consistent estimates of Ψ without knowing distribution

However, efficiency of estimate may depend on distribution

Will consider later how to obtain efficient estimator

How might one obtain alternate estimators (large-sample or permutation-based)?

Large-sample:

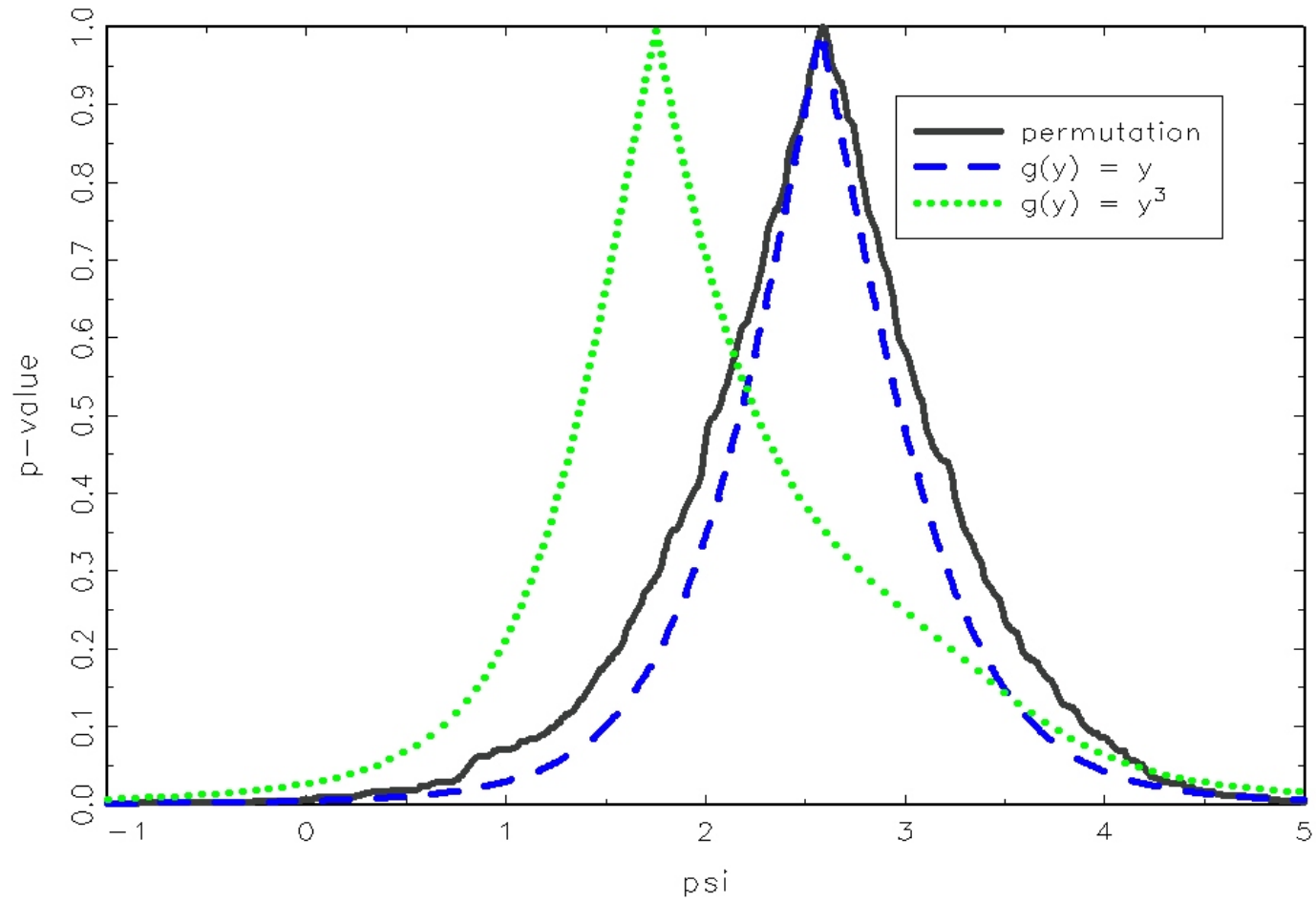
consider other functions to add to logistic regression:

$$\text{logit}[\text{pr}\{R=1|Y^0(\Psi)\}] = \alpha + g\{Y^0(\Psi)\}\beta$$

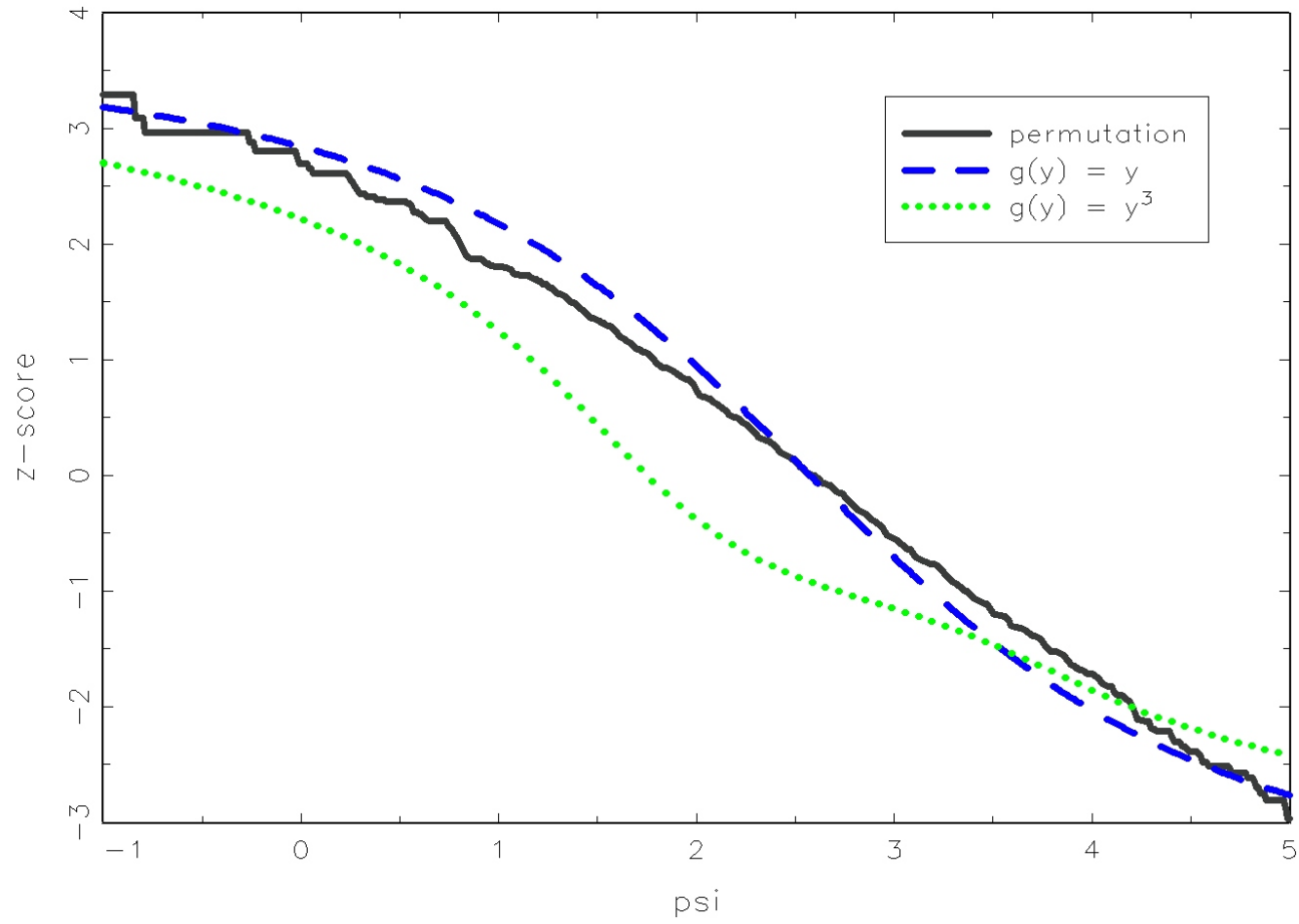
for any arbitrary function $g(\cdot)$, will have independence of $g(Y^0)$ and R ; thus,
can use any arbitrary function $g(\cdot)$ in regression

will consider later how choice might matter

show for example:



score test statistic:



Extend to permutation tests: how?

Permutation test:

can consider other test statistics

Is assumption of constant treatment effect reasonable?

Is it useful?

In general, not reasonable

Nonetheless, useful

Indistinguishable in data from following model

$$Y^0 = \mu + \epsilon^0$$

$$Y^1 = \mu + \Psi + \epsilon^1$$

$$f(\epsilon^0) = f(\epsilon^1)$$

Why is deterministic model indistinguishable from above?

Because not possible to observe both potential outcomes simultaneously

Is above model parametric or semiparametric?

Depends on whether distribution given for ϵ^a

If no distribution specified for ϵ^a , then semiparametric

Semiparametric: part of model for joint distribution of observables unspecified or infinite dimensional

Can imagine density for $f(\epsilon; \gamma)$, where γ is unknown parameter indexing density

If γ is infinite dimensional, then model is semiparametric

Otherwise, model is parametric (finite # of unknowns in joint density of observables)

What if distribution of error terms ϵ^a different for different values of a ?

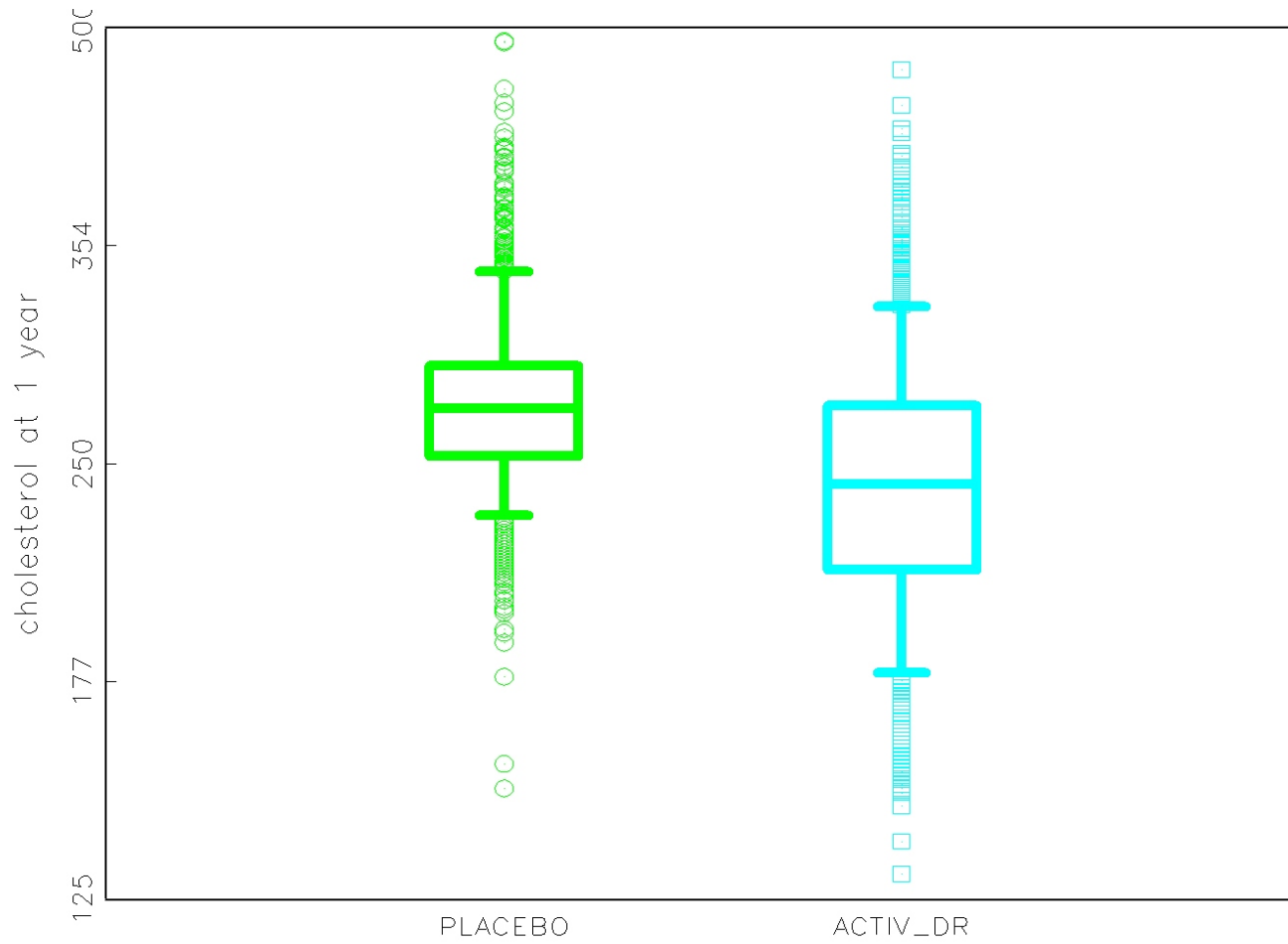
- Describe how this might happen
- Is model sketched above correct?
- Can deterministic model considered above be correct?
- Is statistical inference procedure sketched above correct?

Might happen if treatment has substantially different effects on different subjects; magnitude of individual level effects larger for some subjects than others

Consider data from Lipid Research Clinics/Coronary Primary Prevention Trial (LRC-CPPT)

Randomized trial of cholestyramine, a drug used to lower blood cholesterol

Data:



Describe data/findings

Model not correct

greater spread of cholesterol (shown here on log scale) for subjects on active treatment than placebo

consider model:

$$Y^r = \mu + \Psi + \epsilon^0 + \delta$$

$$\text{corr}(\delta, \epsilon^0) = 0$$

Y^1 has larger variance than Y^0 unless variance of δ is 0

speculate further on why randomization might have different effects on different subjects

different biological effects of drug

effects of randomization on compliance

data can be used to choose explanation; will consider later

Useful to specify correct model to know if inference formally justified

Additional models/generalizations:

Mean models:

$$E(Y^r) = \mu + r\Psi$$

or (more properly) $E(Y|R=r) - E(Y^0|R=r) = r\Psi$

Models for percentiles of distribution

define $y_q^a \equiv y: F_{Y^a}(y) = q$ (or $y_q^a \equiv y: F_{Y^a|A=a}(y|a) = q$, $y_q^{a,0} \equiv y: F_{Y^0|A=a}(y|a) = q$)

1 type of model: $y_q^a = y_q^0 + r\Psi_q$ ($y_q^a = y_q^{a,0} + r\Psi_q$)

can have model

- for single value of q (e.g., 0.5 - median)
- for multiple values of q simultaneously

if Ψ_q constant for all q , then have distribution model specified above

can have different models for different q ; can parametrize Ψ_q in terms of finite
of parameters

Are estimation procedure sketched above valid for mean model?

Not in general

For g-estimation, valid if $g(Y^0) = Y^0$ (No transformation)

Why should mean model have fewer valid estimating equations than distribution model?

Weaker model, puts fewer restrictions on data

Consider following scenario

desire to estimate mean

no assumptions: only consistent estimator is sample mean

assume distribution is symmetric: can use trimmed mean, median, trimean, etc.

stronger assumptions allow larger number of consistent estimators

More general formulation of mean and distribution models

Mean model:

$$h\{E(Y^a)\} - h\{E(Y^0)\} = \Psi_a$$

Distribution model

$$h(y_q^a) - h(y_q^0) = \Psi_a$$

or make conditional on observed treatment

$$h\{E(Y^a|A=a)\} - h\{E(Y^0|A=a)\} = \Psi_a$$

will consider further generalizations later

Nature of probability in models discussed

exact/permutation-based methods

- probability comes from randomization
- other elements deterministic

3 other possible probability models

- stochastic process(es)
- sampling (typically from infinite population)
- subjective (typically Bayesian)

Will discuss inference for marginal parameters (e.g., $pr(Y^a=1)$) first, then inference for contrasts.

Thus, base inference for Y^0 on distribution of Y in control arm

Can imagine:

1. Probabilistic (stochastic) process:

Subjects in finite population assigned to treatment

Probability that each of these subjects dies is π ; imagine flipping a coin with probability π of heads; if coin lands heads, subject dies, otherwise lives

Repeat process many times

Interest then is in estimating characteristic π of the process

Suggested probability model: binomial distribution

Suppose N subjects; probability that one will see x subjects die is

$$p(X=x) = \binom{N}{x} \pi^x (1-\pi)^{N-x}$$

In approach, π is fixed parameter (will use Greek letters to denote *parameters*, which denote characteristics of distribution or underlying process giving rise to data)

Typically, π unknown (although restricted: $0 \leq \pi \leq 1$) but considered fixed

Goal: learn about $\pi = \text{pr}(Y=1|R=0)$ from data

Conceptual difficulties with model?

Model treats all subjects as comparable; each has same probability π of dying

Suppose we know something about subjects' prognoses: can divide subjects into 2 classes-good and poor prognosis

Simple binomial model no longer appropriate: conceptually or mathematically

Conceptually: subjects are not comparable/exchangeable

Mathematical consequence: given knowledge of number of subjects in each group, variance of mean probability less than that predicted by binomial distribution

$\pi_1 \equiv$ probability of failure in stratum 1 (good prognosis)

$\pi_0 \equiv$ probability of failure in stratum 0 (poor prognosis)

$p_1 \equiv$ probability of being in stratum 1

$n_1 \equiv$ proportion of subjects in sample in stratum 1

Average probability of failure in *this* study:

$$\pi^* \equiv \frac{n_1 \pi_1 + (N - n_1) \pi_0}{N}$$

Given that n_0, n_1 known, derive variance of estimator

$$\hat{\pi}^* \equiv X/N = \frac{n_1 \frac{X_1}{n_1} + n_0 \frac{X_0}{n_0}}{N} = \frac{n_1 \hat{\pi}_1 + n_0 \hat{\pi}_0}{N} = \frac{n_1}{N} \hat{\pi}_1 + \frac{n_0}{N} \hat{\pi}_0$$

For independent variables A, B:

$$\text{Var}(A+B) = \text{Var}(A) + \text{Var}(B)$$

More generally,

$$\text{Var}(A+B) = \text{Var}(A) + \text{Var}(B) + 2\text{Cov}(A, B)$$

Also, for constant c,

$$\text{Var}(cA)=c^2\text{Var}(A)$$

$$\text{Var}(\hat{\pi}^*)=\text{Var}\left(\frac{n_1}{N}\hat{\pi}_1+\frac{n_0}{N}\hat{\pi}_0\right)=$$

$$\text{Var}\left(\frac{n_1}{N}\hat{\pi}_1\right)+\text{Var}\left(\frac{n_0}{N}\hat{\pi}_0\right)=\left(\frac{n_1}{N}\right)^2\text{Var}(\hat{\pi}_1)+\left(\frac{n_0}{N}\right)^2\text{Var}(\hat{\pi}_0)=$$

$$\frac{1}{N^2}\left\{n_1^2\frac{\pi_1(1-\pi_1)}{n_1}+n_0^2\frac{\pi_0(1-\pi_0)}{n_0}\right\}=\frac{1}{N^2}\{n_1\pi_1(1-\pi_1)+n_0\pi_0(1-\pi_0)\}$$

$$\leq \pi^*(1-\pi^*)/N$$

Now, if n_1, n_0 random, can (sometimes) get back familiar variance formula:

If prognosis fixed for each subject in study, where does randomness in n_1 come from?

Sampling or randomization (can be combined with deterministic or stochastic model for outcomes given treatment):

Randomization: under potential outcomes model (with SUTVA), Y^0 in subjects randomized to receive placebo are simple random sample without replacement of all values of all potential outcomes Y^0 in study

However, if inference is about $pr(Y^0=1)$ in study population, again usual standard error too high

Problem: we have observed $\frac{1}{2}$ of the true values of $pr(Y^0=1)$ in the study

For finite population then, variance $\frac{1}{2}$ of that computed under formula

Finite population corrections; not much used in epidemiology

For some further discussion, see Robins (1988)

Other possible justification:

2. Sampling from much larger population: imagine that subjects in trial are random sample from much larger population

Then, Y 's among subjects randomized to receive placebo are random sample of Y^0 's from much larger population

Sampling unconditionally results in same variance, distributions of failures as stochastic model

Again, conditioning on measured covariates reduces variance

In typical epidemiologic studies, both randomized and not, random sampling not employed; superpopulation from which subjects sampled is convenient fiction

Possible to have hybrid conceptual model:

Study subjects randomly sampled from large population

Within strata defined by measured covariates and treatment, outcome Y is stochastic

Results in same distributions as “pure” models based only on sampling or stochastic process

Subjective probability:

$pr(Y=1)$ describes one's uncertainty