Biost 578A: Special Topics Statistical Design of Clinical Trials

Scott S. Emerson, M.D., Ph.D.
Professor of Biostatistics
University of Washington

Lecture 2:

Ensuring Precision of Inference:

Sample Size Calculation

April 3, 2013

1

Lecture Outline

- Probability Models for Inference
  - Frequentist vs Bayesian Inference
  - (Semi)Parametric vs Distribution-Free Probability Models
- Ensuring Adequate Precision of Inference
  - Sample Size Calculation

2

Basic Premise

- There are two types of people in the world:
  - those who dichotomize everything, and
  - those who don't.

3

Classes of Statistical Models

- Breiman (2000): The two approaches to data analysis
  - Model based (e.g., regression)

    vs

    Algorithmic (e.g., trees, neural nets)
- This lecture:
  - (Semi)Parametric vs distribution-free (nonparametric)
  - Frequentist vs Bayesian

4

Clinical Trial Setting

- Clinical Trials: Experimentation in human volunteers
  - Designed experiments
    - Scientific (epi, basic, clinical) optimality criteria
    - Efficiency
  - Human volunteers
    - Individual and group ethics
  - Industrial sponsors
    - Economic optimality criteria
  - Regulatory agencies
    - Require credible data, data analysis
    - Entire analysis plan specified prior to gathering any data

5

Ultimate Goal

- At the end of the study we want to provide estimates of treatment effect and quantification of the strength of statistical precision
  - Estimate of the treatment effect
    - Single best estimate
    - Range of reasonable estimates
  - Decision for or against hypotheses
    - Binary decision
    - Quantification of strength of evidence

6

Inferential Paradigms

- Two complementary ways to quantify the evidence about hypotheses
  - (Different people have different standards of evidence)
- Are the observed data what we would reasonably expect to see under a specific hypothesis?
  - (The frequentist approach using Pr $(X \mid \theta)$)
- Based on the observed data, what is our strength of belief about a specific hypothesis?
  - (The Bayesian approach using Pr $(\theta \mid X)$)

7

(Semi)Parametric vs Distribution-Free

- Analysis of the data is generally in the context of a parametric, semiparametric, or distribution-free (nonparametric) model
  - Parametric models assume a known shape for the distribution of the data
  - Semiparametric models assume that the shape is similar in some way across groups, but do not otherwise make any assumptions about the exact shape of the distribution
  - Distribution-free models make no assumption about how the shape of the distribution might be similar (or different) across groups

8

Probability Models for Inference

Frequentist vs Bayesian

Where am I going?

Scientific proof is adversarial. We have to address the concerns of skeptical observers.

The frequentist and Bayesian paradigms for inference are complementary, and we will want to accommodate both approaches

9

Notation

- A general framework

Target of inference: $\vec{\theta}$ (finite dimensional)

Observations: $Y_1, \ldots, Y_n$

Joint distribution: $p\left(\vec{y}, \vec{\theta}\right)$

10

Frequentist Probability Model

- Frequentist inference considers the sampling distribution of the statistic across conceptual replications of the experiment

$$p\left(\vec{Y} \mid \vec{\theta}\right) = \frac{p\left(\vec{Y}, \vec{\theta}\right)}{\int p\left(\vec{Y}, \vec{\theta}\right) d\vec{Y}}$$

- A particular value of $\vec{\theta}$ used in the conditional sampling density is termed a "hypothesis"
- Frequentists usually do not explicitly consider the joint distribution of the data and the parameter, instead they just consider every conditional distribution separately

11

Frequentist Point Estimation

- Typical methods for finding point estimates
  - The hypothesis for which the observed statistic is the mean, median, or mode of the sampling distribution, or
  - The hypothesis for which the sampling density at the observed data is highest
- Optimality of point estimates typically judged by
  - Consistency: With an infinite sample size we know the truth

$$\forall \vec{\theta} \in \Theta: \quad \left(\hat{\vec{\theta}} \mid \vec{\theta}\right) \rightarrow_p \vec{\theta}$$

  - Minimal bias (or MSE) across repeated experiments

$$bias\left(\vec{\theta}\right) = E\left(\hat{\vec{\theta}} \mid \vec{\theta}\right) - \vec{\theta}$$

12

Frequentist Interval Estimation

- Confidence intervals (confidence sets)
  - The set of hypotheses for which we might reasonably expect to obtain the observed data
- Typical methods for two sided CI
  - For some level of confidence defined by α, and
  - For some definition *O* of the ordering of $\Re^n$

$$CI(\vec{y};\alpha)=\left\{\vec{\theta}:\frac{\alpha}{2}<\Pr\left(\vec{Y}\le_O \vec{y}\mid\vec{\theta}\right)<1-\frac{\alpha}{2}\right\}$$

- These CI will have the desired "coverage probability"

$$\Pr\left(\vec{\theta}\in CI\left(\vec{Y};\alpha\right)\mid\vec{\theta}\right)=1-\alpha$$

13

Frequentist Interval Estimation

- Usually the ordering is based on some statistic

$$\vec{y}_1\le_T \vec{y}_2 \quad\Leftrightarrow\quad T(\vec{y}_1)\le T(\vec{y}_2)$$

$$CI(\vec{y};\alpha)=\left\{\vec{\theta}:\frac{\alpha}{2}<\Pr\left(T\left(\vec{Y}\right)\le T(\vec{y})\mid\vec{\theta}\right)<1-\frac{\alpha}{2}\right\}$$

14

Frequentist Interval Estimation

- Sometimes we choose orderings that also depend upon $\vec{\theta}$
  - E.g., likelihood ratio ordering from "inverting LR test"

$$\vec{y}_1\le_{T,\vec{\theta}_0}\vec{y}_2 \quad\Leftrightarrow\quad \frac{p\left(\vec{y}_1\mid\vec{\theta}=\hat{\hat{\theta}}\right)}{p\left(\vec{y}_1\mid\vec{\theta}=\vec{\theta}_0\right)}\le\frac{p\left(\vec{y}_2\mid\vec{\theta}=\hat{\hat{\theta}}\right)}{p\left(\vec{y}_2\mid\vec{\theta}=\vec{\theta}_0\right)}$$

15

Frequentist Hypothesis Testing

- Hypotheses

$$\text{For }\Theta_0\cap\Theta_0=\phi:\quad H_0:\vec{\theta}\in\Theta_0\quad vs\quad H_1:\vec{\theta}\in\Theta_1$$

- Reject a hypothesis for which the observed data is too rare
  - For some critical value, the type 1 error is

$$\alpha_{\vec{c}}=\max_{\vec{\theta}_0\in\Theta_0}\Pr\left(\vec{Y}\le_O\vec{c}\mid\vec{\theta}_0\right)$$

  - For some specified alternative hypothesis the power is

$$Pwr_{\vec{c}}\left(\vec{\theta}\right)=\Pr\left(\vec{Y}\le_O\vec{c}\mid\vec{\theta}\right)$$

16

Frequentist Hypothesis Testing

- Typically, for some specified ordering
  - We fix the type 1 error to some suitably low level α
  - We then find a critical value c that achieves that type 1 error
  - We then compute the power curve

17

Frequentist Hypothesis Testing

- In some settings we can define "optimal" tests
  - Simple hypotheses: Neyman-Pearson lemma tells us that the likelihood ratio ordering provides the MP-α test
  - Composite hypotheses with monotone likelihood ratio: Karlin-Rubin theorem provides one-sided UMP-α test
  - In two-sided tests, we sometimes appeal to uniformly most powerful unbiased (UMPU) tests
    - Unbiased tests:

$$\forall \vec{\theta}_0 \in \Theta_0, \vec{\theta}_1 \in \Theta_1 : \quad Pwr_{\vec{c}}\left(\vec{\theta}_0\right) \leq Pwr_{\vec{c}}\left(\vec{\theta}_1\right)$$

18

Bayesian Probability Model

- Bayesian inference considers the probability distribution for the true summary measure conditioned on the observed data and an assumed prior distribution

$$p\left(\vec{\theta} \mid \vec{Y}\right) = \frac{p\left(\vec{Y} \mid \vec{\theta}\right)\lambda\left(\vec{\theta}\right)}{\int p\left(\vec{Y} \mid \vec{\theta}\right)\lambda\left(\vec{\theta}\right)d\vec{\theta}}$$

where

$$\lambda\left(\vec{\theta}\right) = \int p\left(\vec{Y}, \vec{\theta}\right)d\vec{Y} \text{ is a prior distribution for } \vec{\theta}$$

- Bayesians most often specify a frequentist probability model and a prior distribution to induce the joint distribution

19

Bayesian Probability Space

- Bayesian inference considers the probability distribution for the parameter measuring treatment effect

$$\lambda\left(\vec{\theta}\right) = \int p\left(\vec{Y}, \vec{\theta}\right)d\vec{Y} \text{ is a prior distribution for } \vec{\theta}$$

- Possible scientific relevance of the prior distribution
  - Describes behavior of scientists: A frequentist probability
    - "Of all experimental hypotheses that might have been selected for further investigation, how likely is the selected treatment to be truly beneficial?"
  - Quantifies subjective uncertainty at an individual level
    - "What is a particular scientists' belief about the treatment effect?"
  - Consensus subjective prior
    - "What is the population average of individual priors?
      - But posterior probabilities using an average prior may not correspond to average of individual posterior probabilities

20

Bayesian Estimation

- Point estimates:
  - A summary measure of the posterior probability distribution (mean, median, mode)

$$\text{Posterior mean} \quad \hat{\vec{\theta}} = E\left(\vec{\theta} \mid \vec{Y}\right)$$

- Interval estimates: Credible intervals
  - A set of hypotheses having the highest posterior density

$$1-\alpha_c \text{ HPD credible interval:}$$

$$CI(c) = \left\{\vec{\theta} : p\left(\vec{\theta} \mid \vec{Y}\right) > c\right\}$$

$$\alpha_c = 1 - \Pr\left(\vec{\theta} \in CI(c) \mid \vec{Y}\right)$$

21

Bayesian Decisions

- Tests:
  - Reject a hypothesis for which the posterior probability is too low
  - Quantify the posterior probability of the hypothesis

$$\text{Posterior probability of null} \quad \Pr\left(\vec{\theta} \in \Theta_0 \mid \vec{Y}\right)$$

22

Statistical Inference

- Information required for inference
  - Frequentist
    - Tests: need the sampling distribution under the null
    - Estimates: need the sampling distribution under all hypotheses
  - Bayesian
    - Tests and estimates: need the sampling distribution under all hypotheses and a prior distribution

23

Statistical Inference

- Both approaches have their adherents
- Frequentist
  - A precise (objective) answer to not quite the right question
  - Well developed nonparametric and moment based analyses (e.g., GEE)
  - Conciseness of presentation
- Bayesian
  - A vague (subjective) answer to the right question
  - Adherence to likelihood principle in parametric settings (and coarsened approach)

24

Ideal Results

- Goals of "drug discovery" are similar to those of diagnostic testing in clinical medicine
- We want a "drug discovery" process in which there is
  - A low probability of adopting ineffective drugs
    - High specificity (low type I error)
  - A high probability of adopting truly effective drugs
    - High sensitivity (low type II error; high power)
  - A high probability that adopted drugs are truly effective
    - High positive predictive value
    - Will depend on prevalence of "good ideas" among our ideas

25

Distinctions without Differences

- There is no such thing as a "Bayesian design"
- Every RCT design has a Bayesian interpretation
  - (And each person may have a different such interpretation)
- Every RCT design has a frequentist interpretation
  - (In poorly designed trials, this may not be known exactly)

26

Diagnostic Medicine: Evaluating a Test

- **We condition on diagnoses** (from gold standard)
  - Frequentist criteria: We condition on what is unknown in practice
- **Sensitivity: Do diseased people have positive test?**
  - Denominator: Diseased individuals
  - Numerator: Individuals with a positive test among denominator
- **Specificity: Do healthy people have negative test?**
  - Denominator: Healthy individuals
  - Numerator: Individuals with a negative test among denominator

27

Diagnostic Medicine: Using a Test

- **We condition on test results**
  - Bayesian criteria: We condition on what is known in practice
- **Pred Val Pos: Are positive people diseased?**
  - Denominator: Individuals with positive test result
  - Numerator: Individuals with disease among denominator
- **Pred Val Neg: Are negative people healthy?**
  - Denominator: Individuals with negative test result
  - Numerator: Individuals who are healthy among denominator

28

Points Meriting Special Emphasis

- Discover / evaluate tests using frequentist methods
  - Sensitivity, specificity
- Consider Bayesian methods when interpreting results for a given patient
  - Predictive value of positive, predictive value of negative
- Possible rationale for our practices
  - Ease of study: Efficiency of case-control sampling
  - Generalizability across patient populations
    - Belief that sensitivity and specificity might be
    - Knowledge that PPV and NPV are not
  - Ability to use sensitivity and specificity to get PPV and NPV
    - But not necessarily vice versa

29

Bayes' Rule

- Allows computation of "reversed" conditional probability
- Can compute PPV and NPV from sensitivity, specificity
  - **BUT: Must know prevalence of disease**

$$PPV = \frac{sensitivity \times prevalence}{sens \times prevalence \quad + \quad (1-spec)\times(1-prevalence)}$$

$$NPV = \frac{specificity \times (1-prevalence)}{spec \times (1-prevalence) \quad + \quad (1-sens)\times prevalence}$$

30

Application to Drug Discovery

- We consider a population of candidate drugs
- We use RCT to "diagnose" truly beneficial drugs
- Use both frequentist and Bayesian optimality criteria
  - Sponsor:
    - High probability of adopting a beneficial drug (frequentist power)
  - Regulatory:
    - Low probability of adopting ineffective drug (freq type 1 error)
    - High probability that adopted drugs work (posterior probability)
  - Public Health (frequentist sample space, Bayes criteria)
    - Maximize the number of good drugs adopted
    - Minimize the number of ineffective drugs adopted

31

Slightly Different Setting

- Usually we are interested in some continuous parameter
  - E.g., proportion of infections cured is $0 < p < 1$
- "Prevalence" is replaced by a probability distribution
  - Prior (subjective) probability of selecting a drug to test that cures proportion p of the population
- Sum over two hypotheses replaced by weighted average (by some subjective prior) over all possibilities

$$\Pr(p \mid \hat{p}) = \frac{\Pr(\hat{p} \mid p)\times \Pr(p)}{\int \Pr(\hat{p} \mid p)\times \Pr(p)\, dp}$$

$$= \frac{freq\ samp\ distn \ \times \ prior\ prob}{weighted\ average\ freq\ samp\ distn}$$

32

Frequentist Inference

- Control type 1 error: False positive rate
  - Based on specificity of our methods
- Maximize statistical power: True positive rate
  - Sensitivity to detect specified effect
- Provide unbiased (or consistent) estimates of effect
- Standard errors: Estimate reproducibility of experiments
- Confidence intervals
- Criticism: Compute probability of data already observed
  - "A precise answer to the wrong question"

33

Bayesian Inference

- Hypothesize prior prevalence of "good" ideas
  - Subjective probability
- Using prior prevalence and frequentist sampling distribution
  - Condition on observed data
  - Compute probability that some hypothesis is true
    - "Posterior probability"
  - Estimates based on summaries of posterior distribution
- Criticism: Which presumed prior distribution is relevant?
  - "A vague answer to the right question"

34

Frequentist vs Bayesian

- Frequentist and Bayesian inference truly complementary
- I contend that both frequentist and Bayesian inference can provide evidence for treatment effects
- Frequentist: Design so the same data not likely from null / alt
  - (This is in some sense placing equal emphasis on both hypotheses.)
- Bayesian: Explore updated beliefs based on a range of priors

35

Frequentist vs Bayesian

- Bayes rule tells us that we can parameterize the positive predictive value by the type I error and prevalence
  - Maximize new information by maximizing Bayes factor
  - With simple hypotheses, the steeper the power curve, the greater discrimination between hypotheses

$$PPV = \frac{power \times prevalence}{power \times prevalence \;+\; type\ I\ err \times (1 - prevalence)}$$

$$\frac{PPV}{1-PPV} = \frac{power}{type\ I\ err} \times \frac{prevalence}{1-prevalence}$$

$$posterior\ \ odds = Bayes\ Factor \times prior\ \ odds$$

36

Bayes Factor

- In the more general case, Bayes Factor can be computed to compare two sets of hypotheses
  - But it will then depend on the prior

$$\frac{\Pr(\Theta_1 \mid \vec{Y}=\vec{y})}{\Pr(\Theta_0 \mid \vec{Y}=\vec{y})} = \frac{\Pr(\vec{Y}=\vec{y} \mid \Theta_1)}{\Pr(\vec{Y}=\vec{y} \mid \Theta_0)} \times \frac{\Pr(\Theta_1)}{\Pr(\Theta_0)}$$

$$= \frac{\Pr(\Theta_1, \vec{Y}=\vec{y})}{\Pr(\Theta_0, \vec{Y}=\vec{y})} = \frac{\int_{\Theta_1} p(\vec{y} \mid \theta)\lambda(\theta)\, d\theta}{\int_{\Theta_0} p(\vec{y} \mid \theta)\lambda(\theta)\, d\theta}$$

37

Statistical Inference

- I take the view that both approaches need to be accomodated in every analysis
  - Goal of the experiment is to convince the scientific community, which likely includes believers in both standards for evidence
  - Bayesian priors should be chosen to reflect the population of priors in the scientific community

38

The Problem As I See It

- Scientific criteria dictate using distribution-free probability models insofar as possible and presenting both frequentist and Bayesian inference
- Commonly used frequentist statistics are often easily interpreted as distribution-free
  - (We often calculate the variance wrong for my preferred scientific hypotheses)
- Bayesian methods are most often couched in parametric models
  - Distribution-free methods only poorly developed

39

The Ultimate Goal

- Development of a framework for the analysis of clinical trial data in which
  - the hypotheses being tested are defined for distribution-free probability models
    - (distribution-free model is superset of parametric alternatives)
  - tests and estimates are consistent
  - both Bayesian and frequentist inference are possible
  - the methods can easily be used by nonstatisticians

40

The Ultimate Goal

- Major Issues:
  - Distribution-free interpretation of summary measures
  - Specification of hypotheses
  - Estimation of sampling distributions
    - Variance and mean-variance relationships
  - Dual Bayesian approaches for each test
    - Suitable families of priors

41

Probability Models for Inference

(Semi)Parametric vs Distribution-Free Models

Where am I going?

Statistical methods are most often derived in the context of parametric or semi-parametric probability models.

The "unconscious" frequentist gravitates towards models that are robust across distributions.

We will want to use the same models for Bayesian inference.

42

Setting: Two Arm Clinical Trials

Theorem (Fulghum):

All you really need to know,
can be learned in kindergarten.

Corollary (Emerson):

Most the statistics you really
need to know can be learned
in Biost 514 and Stat 512.

43

Setting: Two Arm Clinical Trials

"Because the simplest thing statisticians
need to do is compare two groups.
And we don't know how to do it."

- Attributed to Fred Mosteller when asked by Dr. Elliot Antman (a well known cardiologist) to explain why we need so many types of two sample comparison procedures.

44

Probability Models

- I define parametric, semiparametric, and nonparametric models in the two independent sample setting
  - My definition of semiparametric models is a little stronger than some statisticians
    - There are probability models intermediate to my semiparametric and distribution-free categories (e.g., "stochastically ordered")
  - The distinction is to isolate models with assumptions that I think too strong
- Notation for two sample probability model

$$\text{Treatment}: \quad Y_1, \ldots, Y_n \overset{iid}{\sim} F$$

$$\text{Control}: \quad X_1, \ldots, X_m \overset{iid}{\sim} G$$

45

Parametric Probability Models

- *F, G* are known up to some finite dimensional parameter vectors

$$F(t) = \Psi\left(t, \vec{\Phi}_Y\right)$$
$$G(t) = \Psi\left(t, \vec{\Phi}_X\right)$$

where:

$\Psi(\cdot,\cdot)$ has known form (in both t and $\vec{\Phi}$)

$\vec{\Phi}$ is finite dimensional and unknown

46

Parametric Probability Models

- Examples

$$\text{Normal}: \quad Y_i \sim N\left(\mu, \sigma^2\right) \quad X_j \sim N\left(\nu, \tau^2\right)$$

$$\text{Bernoulli}: \quad Y_i \sim B(1, \mu) \quad X_j \sim B(1, \nu)$$

$$\text{Poisson}: \quad Y_i \sim P(\mu) \quad X_j \sim P(\nu)$$

$$\text{Exponential}: \quad Y_i \sim E(\mu) \quad X_j \sim E(\nu)$$

47

Parametric Probability Models

- Target of inference
  - Because the shape of the distribution is entirely known, the target of inference can be expressed as a function of the unknown parameters

$$\vec{\theta} = \vec{h}\left(\vec{\Phi}_X, \vec{\Phi}_Y\right)$$

  - If the goal is to estimate some particular functional of the distribution, it must be recognized that the target of inference in the most general case will involve the shape of the distribution
    - "functional" includes operations on its function arguments such as integration, inversion, etc

$$\vec{\theta} = \vec{\eta}\left(\vec{\Phi}_X, \vec{\Phi}_Y, \Psi(\cdot)\right)$$

48

Semiparametric Probability Models

- Forms of F, G are unknown, but related to each other by some finite dimensional parameter vector
  - F can be determined from G and a finite dimensional parameter
  - (Most often: under the null hypothesis, F = G)

49

Semiparametric Probability Models

- Forms of *F, G* are unknown, but related to each other by some finite dimensional parameter vector
  - *F* can be determined from *G* and a finite dimensional parameter

$$F(t) = \Psi\left(t, \vec{\Phi}_Y\right)$$
$$G(t) = \Psi\left(t, \vec{\Phi}_X = \vec{0}\right)$$

where:

$\Psi(\cdot,\cdot)$ has unknown form (in t)

$\vec{\Phi}_X = \vec{0}$ is finite dimensional and known (identifiability)

$\vec{\Phi}_Y$ is finite dimensional and unknown

50

Semiparametric Probability Models

- Examples

$$\text{Shift}: \quad F(t) = G(t-\mu)$$
$$\text{Shift-scale}: \quad F(t) = G\left(\frac{t-\mu}{\sigma}\right)$$
$$\text{Accel failure}: \quad F(t) = G(t\gamma)$$
$$\text{Prop hzd}: \quad 1-F(t) = [1-G(t)]^{\gamma}$$

51

Semiparametric Probability Models

- Target of inference
  - Most often, the target of inference is some function of the unknown finite dimensional parameter
    - Recall that for identifiability, that parameter measures the relationship between *F* and *G*

$$\vec{\theta} = \vec{h}\left(\vec{\Phi}_Y\right)$$

  - If the goal is to estimate some particular functional of the distribution, it must be recognized that the target of inference in the most general case will also involve the infinite dimensional parameter measuring the shape of the distribution

$$\vec{\theta} = \vec{\eta}\left(\vec{\Phi}_Y, \Psi(\cdot)\right)$$

52

Distribution-Free Probability Models

- Forms of *F, G* are completely arbitrary and unknown
  - Each distribution is an infinite dimensional parameter
- An infinite dimensional parameter is needed to derive the form of *F* from *G*

53

Distribution-Free Probability Models

- Target of inference
  - The target of inference in the most general case is regarded as some functional contrast across the distribution functions *F* and *G*

$$\vec{\theta} = \vec{\eta}\left(F(\cdot), G(\cdot) \right)$$

  - Most often, however, this is a contrast of functionals across the distributions

$$\vec{\theta} = \vec{h}\left(\vec{\eta}(F(\cdot)), \vec{\eta}(G(\cdot)) \right)$$

54

The Problem

- In the development of statistical models, and even moreso in the teaching of statistics, parametric probability models have received undue emphasis
- Examples:
  - t test is typically presented in the context of the normal probability model
  - theory of linear models stresses small sample properties
  - random effects specified parametrically
  - Bayesian (and especially hierarchical Bayes) models are replete with parametric distributions

55

The Problem

- ASSERTION: Such emphasis is not typically in keeping with the state of knowledge as an experiment is being conducted
- The parametric assumptions are more detailed than the hypothesis being tested, e.g.,:
  - Question: How does the intervention affect the first moment of the probability distribution?
  - Assumption: We know how the intervention affects all of the 2nd, 3rd, …, ∞ central moments of the probability distribution.

56

Arguments Against (Semi)Parametrics

- Conditions under which an intervention might be expected to affect many aspects of a probability distribution
- Example 1: Cell proliferation in cancer prevention
  - Within subject distribution of outcome is skewed (cancer is a focal disease)
  - Such skewed measurements are only observed in a subset of the subjects
  - The intervention affects only hyperproliferation (our ideal)

57

Arguments Against (Semi)Parametrics

- Conditions under which an intervention might be expected to affect many aspects of a probability distribution (cont.)
- Example 2: Treatment of hypertension
  - Hypertension has multiple causes
  - Any given intervention might treat only subgroups of subjects (and subgroup membership is a latent variable)
  - The treated population has a mixture distribution
    - (and note that we might expect greater variance in the group with the lower mean)

58

Arguments Against (Semi)Parametrics

- Conditions under which an intervention might be expected to affect many aspects of a probability distribution (cont.)
- Example 3: Effects on rates
  - The intervention affects rates
  - The outcome measures a cumulative state
  - Arbitrarily complex mean-variance relationships can result

59

The Problem

- These and other mechanisms would seem to make it likely that the problems in which a fully parametric model or even a semiparametric model is correct constitute a set of measure zero
  - Exception: independent binary data must be binomially distributed in the population from which they were sampled randomly (exchangeably?)

60

Optimality of Inference

- Impact on what we teach about optimality of statistical models
  - Clearly, parametric theory may be irrelevant in an exact sense (though as guidelines it is still useful)
  - Much of what we teach about the optimality of nonparametric tests is based on semiparametric models
    - e.g., Lehmann, 1975: location-shift models

61

Example: Wilcoxon Rank Sum Test

- Common teaching:
  - Not too bad against normal data
  - Better than t test when data have heavy tails
- More accurate guidelines:
  - Above holds when a shift model holds for some monotonic transformation of the data
  - If propensity to outliers (mixture distributions) is different between groups, the t test may be better even in presence of heavy tails
  - In the general case, the t test and the Wilcoxon are not testing the same summary measure

62

Distribution-Free Probability Models

- Target of inference
  - The target of inference in the most general case is regarded as some functional contrast across the distribution functions *F* and *G*

$$\vec{\theta} = \vec{\eta}\left(F(\cdot), G(\cdot) \right)$$

  - Most often, however, this is a contrast of functionals across the distributions

$$\vec{\theta} = \vec{h}\left(\vec{\eta}(F(\cdot)), \vec{\eta}(G(\cdot)) \right)$$

63

Targets of Inference

$$\vec{\theta} = \vec{h}\left(\vec{\eta}(F(\cdot)), \vec{\eta}(G(\cdot)) \right)$$

- Difference (or ratio) of mean blood pressures
- Ratio of geometric mean blood pressures
- Ratio (or difference) of median blood pressures
- Difference (or ratio) of proportion with SBP < 120
- Ratio (or difference) of odds of having SBP < 120
- Ratio (or difference) of average hazard for time to SBP < 120

$$\vec{\theta} = \vec{\eta}\left(F(\cdot), G(\cdot) \right)$$

- Cox PH ratio of hazard for time to SBP < 120
- Median of difference in blood pressures
- Probability that $X < Y$
- Supremum of $| F(t) - G(t) |$

64

Statistical Models

- How are (semi)parametric assumptions really used in statistical models?
  - **Choice of functional for comparisons**
  - Formula for computing the estimate of the functional
  - Distributional family for the estimate
  - Mean-variance relationship across alternatives
  - Shape of distribution for data

65

Choice of Functional for Comparisons

- Parametric: Driven by efficiency of functional for the particular parametric family
  - Normal: use mean
  - Lognormal: use (log) geometric mean
  - Double exponential: use median
  - Uniform: use maximum
- Semiparametric: Choose functional for scientific relevance, etc., then adopt a semiparametric model in which desired functional is basic to model
  - Survival data: consider hazard ratio and use proportional hazards

66

Hierarchy for Choice of Functional

- Better bases for choosing summary measure for decisions in order of importance (nonparametric)
  - Current state of scientific knowledge
  - Scientific (clinical) relevance
  - Potential for intervention to affect the measure
  - Statistical accuracy and precision of analysis

67

Criteria: Current State of Knowledge

- Scientific investigation proceeds through a series of studies/experiments
- Initial studies refine hypotheses to be examined in later studies
  - Relevance of a binary search: Divide the set of all possible hypotheses into two sets to be discriminated between in the initial experiment
- E.g., First try to characterize first order trends, later try to determine exact shape of dose-response

68

Criteria: Scientific Relevance

- E.g., Goal is predicting totals in a larger population
  - Health services research: mean cost of different health care strategies
- E.g., Sensitivity to detecting differences in tendency to outlying values
  - Cancer prevention: mean cell proliferation rates within a person (cancer is an outlier)
  - Economic policy: Sociology of wealth distribution (median) vs economic force (mean)
- E.g., Important clinical thresholds
  - Sepsis trials: 28 day mortality rather than time to death

69

Criteria: Potential for Effect

- E.g., Treatment is designed to affect outliers
  - Aspirin only lowers temperature in fever
  - Ideal cancer therapy only decreases proliferation of cancer cells

70

Criteria: Statistical Precision

- E.g., Outliers decrease the precision of estimating means relative to precision of estimating geometric means or medians
- E.g., Dichotomization of data may result in a loss of efficiency

71

Statistical Models

- How are (semi)parametric assumptions really used in statistical models?
  - Choice of functional for comparisons
  - **Formula for computing the estimate of the functional**
  - Distributional family for the estimate
  - Mean-variance relationship across alternatives
  - Shape of distribution for data

72

Parametric Estimate of Functional

- Estimate parameters and then derive summary measures from parametric model
- E.g., estimating the median with aymptotically efficient MLE
  - Normal: sample mean
  - Exponential: sample mean / log(2)
  - Lognormal: sample geometric mean
  - Double exponential: sample median

73

Semiparametric Estimate of Functional

- Parameter is typically fundamental to probability model
  - Transform one group by the parameter and obtain the same distribution as the other group
- E.g., proportional hazards model
  - Hazard ratio estimate averages hazard ratios at each failure time
- E.g., survival cure model (Ibrahim, 1999, 2000)
  - Proportion p_i is cured (survival probability 1 at ∞) in the *i*-th group
  - Noncured group has survival distribution modeled parametrically (e.g., Weibull) or semiparametrically (e.g., proportional hazards)
  - The problem as I see it: Incorrect assumptions about the nuisance parameter can bias the estimation of the treatment effect $\theta = p_1 - p_0$

74

Distribution-Free Estimate of Functional

- Estimate summary measures from nonparametric empirical distribution functions
- E.g., use sample median for inference about population medians
- Note:
  - Often the nonparametric estimate agrees with a commonly used parametric or semiparametric estimate
    - Interpretation may depend on sampling scheme, however
  - The difference will come in the computation of the standard errors

75

Statistical Models

- How are (semi)parametric assumptions really used in statistical models?
  - Choice of functional for comparisons
  - Formula for computing the estimate of the functional
  - **Distributional family for the estimate**
  - Mean-variance relationship across alternatives
  - Shape of distribution for data

76

(Semi)Parametric Distribution of Estimate

- Parametric: Use probability theory to derive distribution of estimate
  - E.g., estimating the mean
    - Normal: sample mean is normal
    - Exponential: sum is gamma
    - Lognormal: log geometric mean is normal
- Semiparametric:
  - Small sample properties: Conditional distributions based on permutation
  - Large sample properties: Asymptotics

77

Distribution Free Estimates

- Nonparametric: Asymptotic normal theory (almost always)
  - Most nonparametric estimators involve a sum somewhere: Central limit theorem holds (like it or not)
    - Thus gamma distributions converge to a normal…
  - Estimates derived from empirical CDF, which converges asymptotically to Brownian bridge process

78

Later: Reliance on Asymptotics

- We use asymptotic theory as justification for approximations based on the normal distribution

$$\sqrt{n}\left(\hat{\vec{\theta}} - \vec{\theta}\right) \rightarrow_d N\left(0, V\left(\vec{\theta}\right)\right)$$

- In RCT, the sample size at which the approximation holds depends on four aspects (in approx order from lowest to highest)
  - Approximate normality
    - Depends on how far out in the tails of the distribution we need
  - Nearly constant mean-variance relationship
  - Discreteness of data
  - Local alternatives used in experimental sequential design
    - We tend to design studies to have power less than 1

79

Statistical Models

- How are (semi)parametric assumptions really used in statistical models?
  - Choice of functional for comparisons
  - Formula for computing the estimate of the functional
  - Distributional family for the estimate
  - **Mean-variance relationship across alternatives**
  - Shape of distribution for data

80

Sampling Distributions

- Asymptotically, most the summary measures considered can be shown to have a limiting normal distribution
  - (exception is the supremum of the difference between the cdf's)
- In this setting, we need only estimate the variance of the sampling distribution under specific hypotheses
  - Formulas
  - Bootstrapping within groups (Population model)
  - Permutation distributions (Randomization model)

81

Sampling Distributions

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$X_{0.5} \sim N\left(mdn(X), \frac{1}{4f^2\left(mdn(X)\right)}\right)$$

$$U = \sum_{i,j} 1_{[X_i \le Y_j]} \overset{H_0}{\sim} N\left(\frac{mn}{2}, \frac{mn(m+n+1)}{12}\right)$$

82

Sampling Distributions

- In most cases, however, it must be recognized that we can only estimate the variance under the truth, which may not correspond to a hypothesis of interest
- If the intervention can affect the variance of the summary measures, then we must account for a mean-variance relationship when considering different hypotheses

83

Sampling Distributions

- Example: Two sample test of binomial proportion

$$\hat{p}_X \sim \left(p_X, \frac{p_X(1-p_X)}{n}\right) \quad \hat{p}_Y \sim \left(p_Y, \frac{p_Y(1-p_Y)}{m}\right)$$

$$Var\left(\hat{\theta} = \hat{p}_X - \hat{p}_Y\right) = \frac{p_X(1-p_X)}{n} + \frac{p_Y(1-p_Y)}{m}$$

84

Sampling Distributions

- Example: Two sample test of binomial proportion
  - Estimated variance is subject to
    - Sampling variability
    - Difference between the truth and the hypothesis

$$\hat{Var}\left(\hat{\theta}\right) = \frac{\hat{p}_X\left(1-\hat{p}_X\right)}{n} + \frac{\hat{p}_Y\left(1-\hat{p}_Y\right)}{m}$$

$$\hat{Var}\left(\theta_0\right) = \frac{\bar{p}\left(1-\bar{p}\right)}{n} + \frac{\bar{p}\left(1-\bar{p}\right)}{m}$$

85

Sampling Distributions

- Estimating mean variance relationships
  - May not be too important for frequentist tests of the null hypothesis, because convention often dictates the null variance we should use
    - Use randomization and/or population variances in adversarial argument
  - However confidence intervals and all Bayesian inference are statements about what data would arise under a variety of hypotheses
    - We must have some idea about how the variance might change with the mean

86

Statistical Hypotheses

- In comparing the distributions across groups using some summary measure, there are two general formulations of the hypotheses
  - Randomization model - H_0: F = G
    - Allows us to ignore possible treatment effects on aspects of the distribution beyond that measured by θ
  - Population model - H_0: $\theta = \theta_0$
    - Sensitive only to the value of the summary measure, thereby allowing for the possibility that the intervention has other effects

87

Statistical Hypotheses

- Example: t test assuming equal variances

$$X_1, \ldots, X_n \overset{iid}{\sim} \left(\mu, \sigma^2\right) \quad Y_1, \ldots, Y_m \overset{iid}{\sim} \left(\nu, \tau^2\right) \quad \pi = \frac{m}{m+n}$$

$$T = \frac{\left(\bar{X} - \bar{Y}\right) - \theta_0}{s_p\sqrt{1/n + 1/m}}$$

$$\xrightarrow{H_0}_d N\left(0, \frac{\pi\sigma^2 + (1-\pi)\tau^2}{(1-\pi)\sigma^2 + \pi\tau^2}\right)$$

88

Statistical Hypotheses

- Example: t test assuming equal variances (cont.)
  - Type I error of test
    - Randomization model: Correct, because unequal variances is an alternative hypothesis
    - Population model: Incorrect if variances are unequal and sample sizes are unequal

89

Statistical Hypotheses

- Example: t test assuming equal variances (cont.)
  - Consistency of test: With an infinite sample size, will every alternative hypothesis be rejected with probability 1?
    - Randomization model: Inconsistent test, because will not reject with probability 1 unless means are different
    - Population model: Consistent test

90

Statistical Hypotheses

- Example: t test assuming unequal variances
  - Population or randomization models
    - Correct type I error
    - Consistent test
    - (slightly less efficient under randomization model)

$$T = \frac{\left(\overline{X} - \overline{Y}\right) - \theta_0}{\sqrt{s_X^2/n + s_Y^2/m}} \xrightarrow{H_0} d \quad N(0,1)$$

91

Sampling Distributions

- Example: Wilcoxon rank sum statistic
  - It basically estimates Pr(X<Y) and the null variance is based on a permutation distribution
    - Inconsistent to test the randomization hypothesis
    - Wrong size to test the population hypothesis
      - (consider a bimodal distribution in one group)
  - Bootstrapping could be used to find a consistent test of the population hypothesis (under the truth)
    - (Note, however, that the Wilcoxon is based on a bivariate functional that is intransitive for location)

92

Statistical Hypotheses

- It seems most in keeping with the scientific setting to consider the population model as the primary hypothesis
  - However: The experiment must convince the scientific community, and some skeptics might want proof under both models

93

Sampling Distributions

- Possible approaches to the mean-variance relationship estimation
  - Explore various mean-variance relationships
    - Bootstrap tilting could be used here
  - Assume no mean-variance relationship
  - Sensitivity analyses intermediate to the two, e.g.

$$Var\left(\hat{\theta}\right) = \omega\theta^{r}$$

94

Sampling Distributions

- Possible approaches to the mean-variance relationship estimation (cont.)
  - A key issue is deciding how many observations are present for estimating the mean-variance relationship
    - If the control group can be used to estimate behavior under the null and the treatment group under the alternative, then possibly have two
    - If an active intervention modifies the response in both groups or in population model, then may only have one

95

Statistical Models

- How are (semi)parametric assumptions really used in statistical models?
  - Choice of functional for comparisons
  - Formula for computing the estimate of the functional
  - Distributional family for the estimate
  - Mean-variance relationship across alternatives
  - **Shape of distribution for data**

96

Statistical Models

- Shape of distribution for data
  - Only really an issue for prediction, which is not considered here

97

Probability Models for Inference

Distribution-Free Bayesian Inference

Where am I going?

A simple approach to providing Bayesian inference in a distribution-free probability model.

98

Bayesian Posterior Distribution

- Derivation based on
  - Density for data
  - Prior for parameter

$$p\left(\vec{\theta} \mid \vec{X}, \vec{Y}\right) = \frac{p\left(\vec{X}, \vec{Y} \mid \vec{\theta}\right)\lambda\left(\vec{\theta}\right)}{\int p\left(\vec{X}, \vec{Y} \mid \vec{\theta}\right)\lambda\left(\vec{\theta}\right)d\vec{\theta}}$$

where

$\lambda\left(\vec{\theta}\right)$ is a prior distribution for $\vec{\theta}$

99

Nonparametric Bayes

- Dirichlet process priors have been proposed for Bayesian inference in the nonparametric setting
- Motivation from categorical distributions
  - Support of distribution is finite set of discrete points
  - Multinomial distribution parameterizes probability of each value
  - Dirichlet distribution is conjugate prior for multinomial
- Application to nonparametric probability model
  - The data is presumed to arise from a mixture distribution
  - A Dirichlet distribution is presumed for the mixing parameters
  - Possibly infinite number of component distributions allows modeling of continuous distributions over a common support
    - ➔ Dirichlet process

100

Dirichlet Process Priors

- The basic idea is most easily seen as the process of deriving the Bayesian estimates
  - Derivation considers a predictive data generation process
- Basic idea
  - *H* is some base distribution (possibly continuous)
    - Defines the support of the family of distributions
    - The presumed distribution in the absence of any data
  - α is a concentration (or precision) parameter measuring the relative belief in *H* or the empirical distribution of the data in Bayesian predictive probability for next observation
    - *n*th observation comes from *H* with probability *α / (α + n – 1)*
    - *n*th observation comes from empirical cdf $\hat{\mathbf{F}}_{n-1}$ with probability *(n-1) / (α + n – 1)*

101

Dirichlet Process Priors: Properties

- It can be shown that with enough data, the Dirichlet Process prior will be consistent for a true distribution having the same support as *H*
  - The closer the true distribution is to *H*, the better the small sample behavior (especially with a high value of the concentration parameter *α*
- However, it is difficult to quantify how the prior places mass on particular distributions in a distribution-free sense
  - E.g., how much mass is placed on bimodal distributions?

102

"Coarsened" Bayesian Models

- Modification regards estimate of summary measure as the data
  - Use asymptotic distributions under population model

$$p\left(\vec{\theta} \mid \hat{\vec{\theta}}\right) = \frac{p\left(\hat{\vec{\theta}} \mid \vec{\theta}\right)\lambda\left(\vec{\theta}\right)}{\int p\left(\hat{\vec{\theta}} \mid \vec{\theta}\right)\lambda\left(\vec{\theta}\right)d\vec{\theta}}$$

where

$\lambda\left(\vec{\theta}\right)$ is a prior distribution for $\vec{\theta}$

103

Impact of "Coarsening"

- Relative to full parametric approach
  - We treat the estimate as if it is sufficient
  - We ignore nuisance parameters, invoking consistency of estimates
  - We model a mean-variance relationship
  - We use the approximate normal distribution based on asymptotics instead of the exact distribution
- In many commonly used parametric or semi-parametric models, the only loss is the use of the asymptotic approximation
  - Sample mean is MLE in the regular normal (known variance), binomial, Poisson, exponential probability models
  - Hazard ratio is semi-parametric sufficient in proportional hazards

104

Advantages / Disadvantages

- Distribution-free consistent estimates of parameter measuring treatment effect
- Specification of prior distributions on the parameter measuring treatment effect
- Furthermore, choice of normal priors allows a standardized exploration of Bayesian inference across a space of priors

105

Standardized Presentation of Inference

- The chief advantage of frequentist inference (to my mind) is that it presents a standard for concise presentation of results
  - Estimates, standard errors, P values, CI
- Bayesian analysis, on the other hand, requires such a presentation for every prior
  - Your prior does not matter to me
  - A consensus prior will not capture the diversity of prior opinion

106

Sensitivity of Inference to Priors

- Papers can present frequentist sampling distribution as the "sufficient statistic" for Bayesian inference
- In the context of the coarsened Bayes approach, we can adopt a standard based on normal priors
  - Conjugate distribution in the absence of a mean-variance relationship
- Two dimensional space of prior distributions
  - Prior mean (pessimism)
  - Prior standard deviation (dogmatism)
    - Also can be measured as information in prior relative to that in planned sample
- Bayesian inference as a contour plot for each inferential quantity
  - Posterior mean, limits of credible intervals, posterior probabilities

107

"Coarsened" Bayesian Posterior Distn

$$\hat{\theta}_n \dot{\sim} N\left(\theta, \frac{V}{n}\right)$$

Prior distribution

$$\theta \sim N\left(\zeta, \tau^2\right)$$

Posterior distribution

$$\theta \mid \hat{\theta}_n \sim N\left(\frac{\frac{n}{\sigma^2}\hat{\theta}_n + \frac{1}{\tau^2}\zeta}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right)$$

108

"Coarsened" Bayesian Posterior Distn

$$\hat{\theta}_n \sim N\left(\theta, \frac{V}{n}\right)$$

Prior distribution

$$\theta \sim N\left(\zeta, \tau^2 = \frac{\sigma^2}{n_0}\right)$$

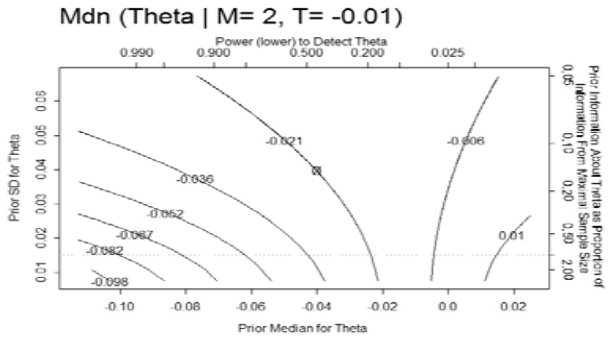
Posterior distribution

$$\theta \mid \hat{\theta}_n \sim N\left(\frac{\frac{n}{\sigma^2}\hat{\theta}_n + \frac{n_0}{\sigma^2}\zeta}{\frac{n}{\sigma^2} + \frac{n_0}{\sigma^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{n_0}{\sigma^2}}\right) = N\left(\frac{n\hat{\theta}_n + n_0\zeta}{n + n_0}, \frac{\sigma^2}{n + n_0}\right)$$

109

Example

- Contour plots for Bayesian inference: Posterior mean



0

Sensitivity Analyses

- In the presence of a mean-variance relationship, a more complicated posterior results
  - Can be numerically integrated or MCMC
- To the extent that people can only describe the first two moments of their prior:
  - A convenient standard for presentation
- But, normal prior is less informative than other priors having the same mean and variance
- In any case, so long as the estimate and standard error is presented in a paper, a reader can apply a more complicated prior

111

Mean-Variance Relationship

- Provide a prior distribution for summary measure that incorporates a prior on the mean-variance relationship
- Note that the concept of updating the prior is probably not valid here, because there is really no added information about mean-variance relationship
  - The mean variance relationship is observed at two points (at most)

112

Ramifications

- The approach to using estimates as the data does mean that in some cases we cannot regard that we are continually updating our posterior
- E.g.: The sample median of the combined sample is not necessarily a weighted mean of the sample median from two separate samples

113

Ensuring Precision

Sample Size Estimation

Where am I going?

A common formula can be used for sample size estimation in the frequentist distribution-free setting

We consider how study design affects sample size

114

Statistical Planning

- Satisfy collaborators as much as possible
  - Discriminate between relevant scientific hypotheses
    - Scientific and statistical credibility
  - Protect economic interests of sponsor
    - Efficient designs
    - Economically important estimates
  - Protect interests of patients on trial
    - Stop if unsafe or unethical
    - Stop when credible decision can be made
  - Promote rapid discovery of new beneficial treatments

115

Sample Size Calculation

- Traditional approach
  - Sample size to provide high power to "detect" a particular alternative
- Decision theoretic approach
  - Sample size to discriminate between hypotheses
    - "Discriminate" based on interval estimate
    - Standard for interval estimate: 95%
      - Equivalent to traditional approach with 97.5% power

116

Reporting Inference

- At the end of the study analyze the data
- Report three measures (four numbers)
  - Point estimate
  - Interval estimate
  - Quantification of confidence / belief in hypotheses

117

Reporting Frequentist Inference

- Three measures (four numbers)
- Consider whether the observed data might reasonably be expected to be obtained under particular hypotheses
  - Point estimate: minimal bias? MSE?
  - Confidence interval: all hypotheses for which the data might reasonably be observed
  - P value: probability such extreme data would have been obtained under the null hypothesis
    - Binary decision: Reject or do not reject the null according to whether the P value is low

118

Reporting Bayesian Inference

- Three measures (four numbers)
- Consider the probability distribution of the parameter conditional on the observed data
  - Point estimate: Posterior mean, median, mode
  - Credible interval: The “central” 95% of the posterior distribution
  - Posterior probability: probability of a particular hypothesis conditional on the data
    - Binary decision: Reject or do not reject the null according to whether the posterior probability is low

119

Parallels Between Tests, CIs

- If the null hypothesis not in CI, reject null
  - (Using same level of confidence)
- Relative advantages
  - Test only requires sampling distn under null
  - CI requires sampling distn under alternatives
  - CI provides interpretation when null is not rejected

120

Scientific Information

- "Rejection" uses a single level of significance
  - Different settings might demand different criteria
- P value communicates statistical evidence, not scientific importance
- Only confidence interval allows you to interpret failure to reject the null:
  - Distinguish between
    - Inadequate precision (sample size)
    - Strong evidence for null

121

Hypothetical Example

- Clinical trials of treatments for hypertension
- Screening trials for four candidate drugs
  - Measure of treatment effect is the difference in average SBP at the end of six months treatment
  - Drugs may differ in
    - Treatment effect (goal is to find best)
    - Variability of blood pressure
  - Clinical trials may differ in conditions
    - Sample size, etc.

122

Reporting P values

Study	P value
A	0.1974
B	0.1974
C	0.0099
D	0.0099

123

Point Estimates

Study	SBP Diff
A	27.16
B	0.27
C	27.16
D	0.27

124

Point Estimates

Study	SBP Diff	P value
A	27.16	0.1974
B	0.27	0.1974
C	27.16	0.0099
D	0.27	0.0099

125

Confidence Intervals

Study	SBP Diff	95% CI	P value
A	27.16	-14.14, 68.46	0.1974
B	0.27	-0.14, 0.68	0.1974
C	27.16	6.51, 47.81	0.0099
D	0.27	0.06, 0.47	0.0099

126

Interpreting Nonsignificance

- Studies A and B are both “nonsignificant”
  - Only study B ruled out clinically important differences
  - The results of study A might reasonably have been obtained if the treatment truly lowered SBP by as much as 68 mm Hg

127

Interpreting Significance

- Studies C and D are both statistically significant results
  - Only study C demonstrated clinically important differences
  - The results of study D are only frequently obtained if the treatment truly lowered SBP by 0.47 mm Hg or less

128

Bottom Line

- If ink is not in short supply, there is no reason not to give point estimates, CI, and P value
- If ink is in short supply, the confidence interval provides most information
  - (but sometimes a confidence interval cannot be easily obtained, because the sampling distribution is unknown under the null)

129

Full Report of Analysis

Study	n	SBP Diff	95% CI	P value
A	20	27.16	-14.14, 68.46	0.1974
B	20	0.27	-0.14, 0.68	0.1974
C	80	27.16	6.51, 47.81	0.0099
D	80	0.27	0.06, 0.47	0.0099

130

Sample Size Calculation

- Traditional approach
  - Sample size to provide high power to “detect” a particular alternative
- Decision theoretic approach
  - Sample size to discriminate between hypotheses
    - “Discriminate” based on interval estimate
    - Standard for interval estimate: 95%
      - Equivalent to traditional approach with 97.5% power

131

Issues

- Summary measure
  - Mean, geometric mean, median, proportion, hazard…
- Structure of trial
  - One arm, two arms, k arms
  - Independent groups vs cross over
  - Cluster vs individual randomization
  - Randomization ratio
- Statistic
  - Parametric, semi-parametric, nonparametric
  - Adjustment for covariates

132

Refining Scientific Hypotheses

- Scientific hypotheses are typically refined into statistical hypotheses by identifying some parameter θ measuring difference in distribution of response
  - Difference/ratio of means
  - Ratio of geometric means
  - Difference/ratio of medians
  - Difference/ratio of proportions
  - Odds ratio
  - Hazard ratio

133

Inference

- Generalizations from sample to population
- Estimation
  - Point estimates
  - Interval estimates
- Decision analysis (testing)
  - Quantifying strength of evidence

134

Measures of Precision

- Estimators are less variable across studies
  - Standard errors are smaller
- Estimators typical of fewer hypotheses
  - Confidence intervals are narrower
- Able to statistically reject false hypotheses
  - Z statistic is higher under alternatives

135

Std Errors: Key to Precision

- Greater precision is achieved with smaller standard errors

Width of CI: $2 \times (crit\ val) \times se(\hat{\theta})$

Test statistic: $Z = \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})} \sim N\left(\frac{\theta - \theta_0}{se(\hat{\theta})}, 1\right)$

Power: $Pwr(\theta) = \Pr(Z \geq z_{1-\alpha} \mid \theta) = 1 - \Phi\left(z_{1-\alpha} - \frac{\theta - \theta_0}{se(\hat{\theta})}\right)$

Typically: $se(\hat{\theta}) = \sqrt{\frac{V}{n}}$

(V related to average "statistical information")

136

Increasing Precision

- Options
  - Increase sample size
  - Decrease V
    - Alter statistical summary measure
    - Improve reliability of measurements
    - Alter study design (e.g., cross-over)
    - Alter eligibility (decrease heterogeneity)
    - Alter clinical endpoint
  - (Decrease confidence level)

137

Without Loss of Generality

- It is sufficient to consider a one sample test of a one-sided hypothesis
  - Generalization to other probability models is immediate
    - We will interpret our variability relative to average statistical info
  - Generalization to two sided hypothesis tests is straightforward
- Fixed sample one-sided tests
  - Test of a one-sided alternative ($\theta_+ > \theta_0$)
    - Upper Alternative: H_+: $\theta \geq \theta_+$ (superiority)
    - Null: H_0: $\theta \leq \theta_0$ (equivalence, inferiority)
  - Decisions based on some test statistic T:
    - Reject H_0 (for H_+) $\Longleftrightarrow$ $T \geq c$
    - Reject H_+ (for H_0) $\Longleftrightarrow$ $T \leq c$

138

Notation

Potential data : $Y_1, Y_2, Y_3, \ldots, Y_{N_j}$

Probability model : $Y_i \overset{iid}{\sim} (\theta, V)$

Interim estimates : $\hat{\theta}_{N_j} = \hat{\theta}(Y_1, \ldots, Y_{N_j})$

Without sequential sampling :

Approximate distn : $\hat{\theta}_j = \hat{\theta}_{N_j} \sim N(\theta, V / N_j)$

Indep increments : $Cov(\hat{\theta}_{N_j}, \hat{\theta}_{N_{j+1}}) = V / N_{j+1}$

Interim test statistics : $Z_j = Z_{N_j} = \dfrac{\hat{\theta}_j - \theta_0}{\sqrt{V / N_j}}$

139

Ex: One Sample Mean

$$iid\ Y_i \sim (\mu, \sigma^2),\ i = 1, \ldots, n$$

$$\theta = \mu \qquad \hat{\theta} = \overline{Y}$$

$$V = \sigma^2 \qquad se(\hat{\theta}) = \sqrt{\frac{\sigma^2}{n}}$$

140

Ex: One Sample Geometric Mean

$$iid\ W_i = \log(Y_i) \sim (\mu, \sigma^2),\ i = 1, \ldots, n$$

$$\theta = \mu = \log(GM) \qquad \hat{\theta} = \overline{W}$$

$$V = \sigma^2 \qquad se(\hat{\theta}) = \sqrt{\frac{\sigma^2}{n}}$$

141

Ex: Difference of Indep Means

$$ind\ Y_{ij} \sim (\mu_i, \sigma_i^2),\ i = 1,2;\ j = 1, \ldots, n_i$$
$$n = n_1 + n_2; \quad r = n_1 / n_2$$

$$\theta = \mu_1 - \mu_2 \qquad \hat{\theta} = \overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}$$

$$V = (r+1)\left[\sigma_1^2 / r + \sigma_2^2\right] \qquad se(\hat{\theta}) = \sqrt{\frac{V}{n}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

142

Ex: Diff of Indep Proportions

$$ind\ Y_{ij} \sim B(1, p_i),\ i = 1,2;\ j = 1, \ldots, n_i$$
$$n = n_1 + n_2; \quad r = n_1 / n_2$$

$$\theta = p_1 - p_2 \qquad \hat{\theta} = \hat{p}_1 - \hat{p}_2 = \overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}$$
$$\sigma_i^2 = p_i(1 - p_i)$$

$$V = (r+1)\left[\sigma_1^2 / r + \sigma_2^2\right] \qquad se(\hat{\theta}) = \sqrt{\frac{V}{n}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

143

Ex: Difference of Paired Means

$$Y_{ij} \sim (\mu_i, \sigma_i^2),\ i = 1, 2; \quad j = 1, \ldots, n$$
$$corr(Y_{1j}, Y_{2j}) = \rho; \quad corr(Y_{ij}, Y_{mk}) = 0 \text{ if } j \neq k$$

$$\theta = \mu_1 - \mu_2 \qquad \hat{\theta} = \overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}$$

$$V = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 \qquad se(\hat{\theta}) = \sqrt{\frac{V}{n}}$$

144

Ex: Mean of Clustered Data

$Y_{ij} \sim (\mu, \sigma^2), i = 1, \ldots, n;\ j = 1, \ldots, m$

$corr(Y_{ij}, Y_{ik}) = \rho$ if $j \neq k$; $corr(Y_{ij}, Y_{mk}) = 0$ if $i \neq m$

$\theta = \mu_1 - \mu_2 \qquad \hat{\theta} = \bar{Y}_{1\bullet} - \bar{Y}_{2\bullet}$

$V = \sigma^2\left(\frac{1 + (m-1)\rho}{m}\right) \qquad se(\hat{\theta}) = \sqrt{\frac{V}{n}}$

145

Ex: Independent Odds Ratios

ind $Y_{ij} \sim B(1, p_i), i = 1,2;\ j = 1, \ldots, n_i$

$n = n_1 + n_2; \quad r = n_1 / n_2$

$\theta = \log\left(\frac{p_1/(1-p_1)}{p_2/(1-p_2)}\right) \qquad \hat{\theta} = \log\left(\frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)}\right)$

$\sigma_i^2 = \frac{1}{p_1(1-p_1)} = \frac{1}{p_1 q_1}$

$V = (r+1)\left[\sigma_1^2 / r + \sigma_2^2\right] \qquad se(\hat{\theta}) = \sqrt{\frac{V}{n}} = \sqrt{\frac{1}{n_1 p_1 q_1} + \frac{1}{n_2 p_2 q_2}}$

146

Ex: Hazard Ratios

ind censored time to event (T_{ij}, δ_{ij}),

$i = 1, 2;\ j = 1, \ldots, n_i; n = n_1 + n_2; \quad r = n_1 / n_2$

$\theta = \log(HR) \qquad \hat{\theta} = \hat{\beta}$ from PH regression

$V = \frac{(1+r)(1/r+1)}{\Pr[\delta_{ij} = 1]} \qquad se(\hat{\theta}) = \sqrt{\frac{V}{n}} = \sqrt{\frac{(1+r)(1/r+1)}{d}}$

147

Ex: Linear Regression

ind $Y_i \mid X_i \sim (\beta_0 + \beta_1 \times X_i, \sigma^2_{Y|X}),\ i = 1, \ldots, n$

$\theta = \beta_1 \qquad \hat{\theta} = \hat{\beta}_1$ from LS regression

$V = \frac{\sigma^2_{Y|X}}{Var(X)} \qquad se(\hat{\theta}) = \sqrt{\frac{\sigma^2_{Y|X}}{nVar(X)}}$

148

Statistics to Address Variability

- At the end of the study we perform frequentist and/or Bayesian data analysis to assess the credibility of clinical trial results
  - Estimate of the treatment effect
    - Single best estimate
    - Precision of estimates
  - Decision for or against hypotheses
    - Binary decision
    - Quantification of strength of evidence

149

Criteria for Precision

- Standard error
- Width of confidence interval
- Statistical power: Probability of rejecting the null hypothesis
  - Select level of significance
    - Standard: One-sided 0.025; two-sided 0.05
    - Pivotal: One-sided 0.005; two-sided 0.01
  - Select "design alternative"
    - Minimal clinically important difference
      - To detect versus declaring significant
      - May consider what is feasible
    - Minimal plausible difference
  - Select desired power
    - High power for a decision theoretic approach

150

Sample Size Determination

- Based on sampling plan, statistical analysis plan, and estimates of variability, compute
  - Sample size that discriminates hypotheses with desired power,

    OR
  - Hypothesis that is discriminated from null with desired power when sample size is as specified, or

    OR
  - Power to detect the specific alternative when sample size is as specified

151

Standardized Setting

Typically: $\hat{\theta} \sim N\left(\theta, \frac{V}{n}\right)$

Width of CI: $2 \times z_{1-\alpha/2} \times \sqrt{\frac{V}{n}}$

Test statistic: $Z = \sqrt{n}\frac{\hat{\theta}-\theta_0}{\sqrt{V}} \sim N\left(\delta = \sqrt{n}\frac{\theta-\theta_0}{\sqrt{V}}, 1\right)$

Power: $Pwr(\theta) = 1 - \Phi\left(z_{1-\alpha/2} - \sqrt{n}\frac{\theta-\theta_0}{\sqrt{V}}\right)$

152

Sample Size Computation: CI

- Confidence interval discriminates between null, design alternative

Typically: $\hat{\theta} \sim N\left(\theta, \frac{V}{n}\right)$

Width of CI: $2 \times z_{1-\alpha/2} \times \sqrt{\frac{V}{n}} = \theta_1 - \theta_0$

Sample size: $n = \frac{\left(z_{1-\alpha/2} + z_{1-\alpha/2}\right)^2 V}{\left(\theta_1 - \theta_0\right)^2}$

153

Sample Size Computation: Power

Typically: $\hat{\theta} \sim N\left(\theta, \frac{V}{n}\right)$

Test statistic: $Z = \sqrt{n}\frac{\hat{\theta} - \theta_0}{\sqrt{V}} \sim N\left(\delta = \sqrt{n}\frac{\theta - \theta_0}{\sqrt{V}}, 1\right)$

Power: $Pwr(\theta) = 1 - \Phi\left(z_{1-\alpha/2} - \sqrt{n}\frac{\theta - \theta_0}{\sqrt{V}}\right) = \beta$

Sample size: $z_\beta = \sqrt{n}\frac{\theta - \theta_0}{\sqrt{V}} - z_{1-\alpha/2} \Rightarrow n = \frac{\left(z_{1-\alpha/2} + z_\beta\right)^2 V}{\left(\theta_1 - \theta_0\right)^2}$

154

Extension to Sequential Sampling

Standardized level α test (n = 1): $\delta_{\alpha\beta}$ detected with power β

Level of significance α when $\theta = \theta_0$

Design alternative $\theta = \theta_1$

Variability V within 1 sampling unit

Required sampling units: $n = \frac{\left(\delta_{\alpha\beta}\right)^2 V}{\left(\theta_1 - \theta_0\right)^2}$

(Fixed sample test: $\delta_{\alpha\beta} = z_{1-\alpha/2} + z_\beta$)

155

When Sample Size Constrained

- Often (usually?) logistical constraints impose a maximal sample size
  - Compute power to detect specified alternative

Find β such that $\delta_{\alpha\beta} = \sqrt{\frac{n}{V}}\left(\theta_1 - \theta_0\right)$

  - Compute alternative detected with high power

$$\theta_1 = \theta_0 + \delta_{\alpha\beta}\sqrt{\frac{V}{n}}$$

156

Special Case: Binomial Model

- Issues in applying normal approximation in small samples
  - Approximate normal distribution
  - Mean-variance relationship
  - Discreteness
- In one sample case we can just use the exact binomial distribution
  - But we can use this case to explore the relative contributions of the above three issues
- In two sample settings it is generally important to use other methods when sample sizes will be small
  - Fishers exact test is too conservative ➔ loss of power
  - Recommend "unconditional exact tests" instead

157

Illustration: One Sample Binomial

- Methods for construction of confidence intervals
  - Exact distribution
  - Exact distribution with half-p
  - Wald intervals
  - Continuity corrected Wald intervals
  - Score intervals
  - Continuity corrected Score intervals
- Recall we construct CI by inverting tests

$$CI(\vec{y};\alpha)=\left\{\vec{\theta}:\frac{\alpha}{2}<\Pr\left(\vec{Y}\le_O \vec{y}\mid\vec{\theta}\right)<1-\frac{\alpha}{2}\right\}$$

158

Exact P Values

- Uses binomial distribution

$$\text{Lower P}: \quad \Pr\left(\hat{p}\le\frac{k}{n}\mid p_0\right)=\sum_{i=0}^{k}\binom{n}{i}p_0^i(1-p_0)^{n-i}$$

$$\text{Upper P}: \quad \Pr\left(\hat{p}\ge\frac{k}{n}\mid p_0\right)=\sum_{i=k}^{n}\binom{n}{i}p_0^i(1-p_0)^{n-i}$$

159

Exact P Values: Half - P

- Uses binomial distribution, but half the probability of being equal to observed value

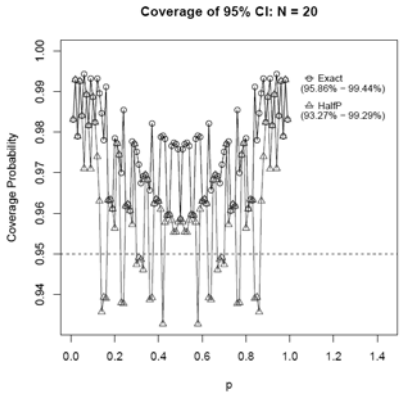
$$\text{Lower P}: \quad \sum_{i=0}^{k-1}\binom{n}{i}p_0^i(1-p_0)^{n-i}+\frac{1}{2}\binom{n}{k}p_0^k(1-p_0)^{n-k}$$

$$\text{Upper P}: \quad \sum_{i=k+1}^{n}\binom{n}{i}p_0^i(1-p_0)^{n-i}+\frac{1}{2}\binom{n}{k}p_0^k(1-p_0)^{n-k}$$

160

Exact (Half P) CI Coverage

- Coverage as function of *p* for *n*=20



161

Wald P Values

- Uses normal approximation with MLE in variance

$$\text{Lower P}: \quad \Pr\left(\hat{p} \le \frac{k}{n} \mid p_0\right) = \Phi\left(\sqrt{n}\,\frac{\frac{k}{n} - p_0}{\sqrt{\hat{p}(1-\hat{p})}}\right)$$

$$\text{Upper P}: \quad \Pr\left(\hat{p} \ge \frac{k}{n} \mid p_0\right) = 1 - \Phi\left(\sqrt{n}\,\frac{\frac{k}{n} - p_0}{\sqrt{\hat{p}(1-\hat{p})}}\right)$$

162

Asymptotic CI: Elevator Stats

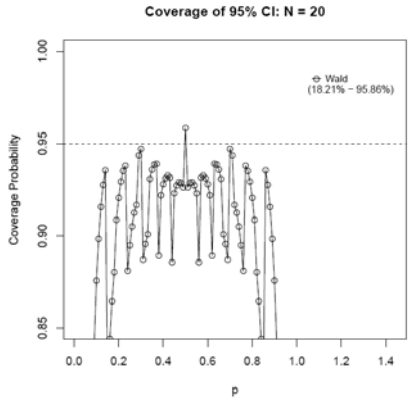
- Often we can just use best estimate of *p* in standard error for confidence intervals and ignore the continuity correction
  - *np* and *n(1-p)* must be large

$$100(1-\alpha)\%\ \text{CI for}\ p: \qquad \hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

163

Wald CI Coverage

- Coverage as function of *p* for *n*=20



164

Continuity Corrected Wald P Values

- Uses normal approximation with MLE in variance, but adjusts for discreteness

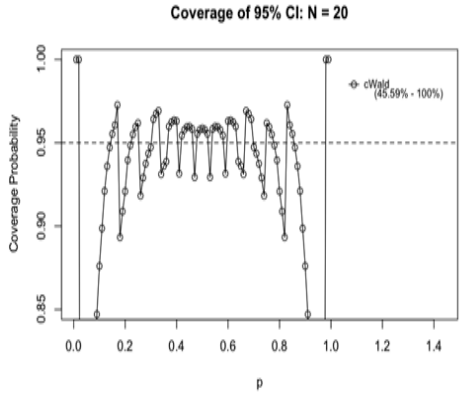
$$\text{Lower P}: \quad \Pr\left(\hat{p} \le \frac{k+0.5}{n} \mid p_0\right) = \Phi\left(\sqrt{n}\,\frac{\frac{k}{n}+\frac{1}{2n}-p_0}{\sqrt{\hat{p}(1-\hat{p})}}\right)$$

$$\text{Upper P}: \quad \Pr\left(\hat{p} \ge \frac{k-0.5}{n} \mid p_0\right) = 1-\Phi\left(\sqrt{n}\,\frac{\frac{k}{n}-\frac{1}{2n}-p_0}{\sqrt{\hat{p}(1-\hat{p})}}\right)$$

165

Continuity Corrected Wald CI Coverage

- Coverage as function of *p* for *n*=20



166

Score P Values

- Uses normal approximation with hypothesized variance

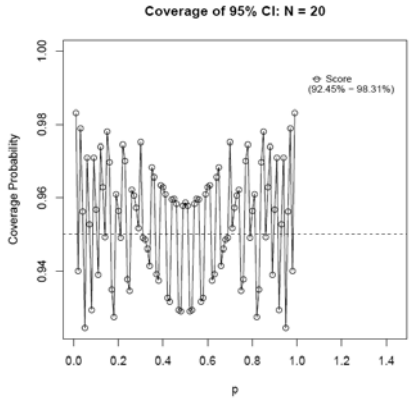
$$\text{Lower P}: \quad \Pr\left(\hat{p} \le \frac{k}{n} \mid p_0\right) = \Phi\left(\sqrt{n}\,\frac{\frac{k}{n}-p_0}{\sqrt{p_0(1-p_0)}}\right)$$

$$\text{Upper P}: \quad \Pr\left(\hat{p} \ge \frac{k}{n} \mid p_0\right) = 1-\Phi\left(\sqrt{n}\,\frac{\frac{k}{n}-p_0}{\sqrt{p_0(1-p_0)}}\right)$$

167

Score CI Coverage

- Coverage as function of *p* for *n*=20



168

Continuity Corrected Score P Values

- Uses normal approximation with hypothesized variance, but adjusts for discreteness

$$\text{Lower P}: \quad \Pr\left(\hat{p} \le \frac{k+0.5}{n} \mid p_0\right) = \Phi\left(\sqrt{n}\,\frac{\frac{k}{n}+\frac{1}{2n}-p_0}{\sqrt{p_0(1-p_0)}}\right)$$

$$\text{Upper P}: \quad \Pr\left(\hat{p} \ge \frac{k-0.5}{n} \mid p_0\right) = 1-\Phi\left(\sqrt{n}\,\frac{\frac{k}{n}-\frac{1}{2n}-p_0}{\sqrt{p_0(1-p_0)}}\right)$$

169

Asymptotic CI: Best Approach

- We do best by considering mean-variance relationship and continuity correction
  - Requires quadratic formula or iterative search
  - (Quadratic formula can be easily implemented in Excel, etc.)

$100(1-\alpha)\%$ CI for p: $\quad (\hat{p}_L, \hat{p}_U)$

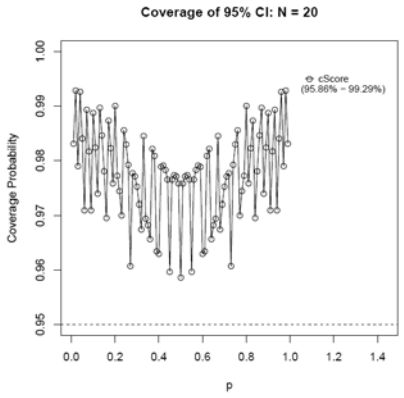
$$\hat{p}_L = \hat{p} - \frac{1}{2n} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}_L(1-\hat{p}_L)}{n}}$$

$$\hat{p}_U = \hat{p} + \frac{1}{2n} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}_U(1-\hat{p}_U)}{n}}$$

170

Continuity Corrected CI Coverage

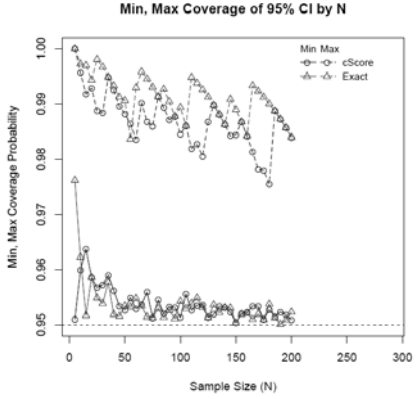
- Coverage as function of *p* for *n*=20



171

Continuity Corrected 95% CI Coverage

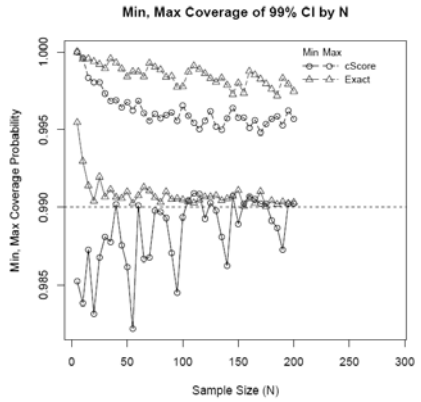
- Coverage as function of *n*



172

Continuity Corrected 99% CI Coverage

- Coverage as function of *n*



173

Two Sample Binomial

- With two sample binomial, we are interested in $\theta = p_1 - p_0$
- There is a nuisance parameter p_0 that affects the distribution of the statistic
- Classical choices include
  - Fishers exact test
  - Chi square test (score test)
  - Wald test
  - Likelihood ratio test
- Better choice in small samples: unconditional exact tests

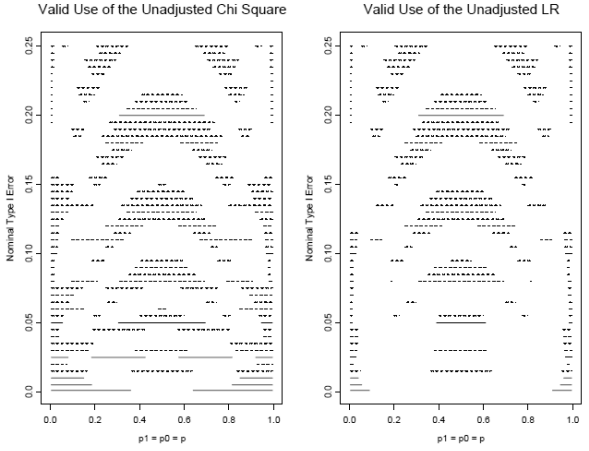
174

Modifications: Unconditional Exact Test

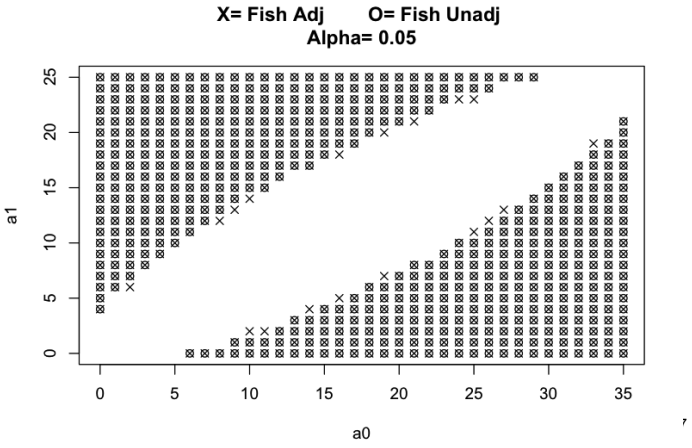
- Use the classical statistic
- Don't presume the classical distribution
  - Don't assume chi squared statistic has chi square distribution
  - Don't assume Fisher's Exact P value has uniform distribution
- Consider all possible values of p common to both groups, and use exact distribution
  - Then take worst case

175

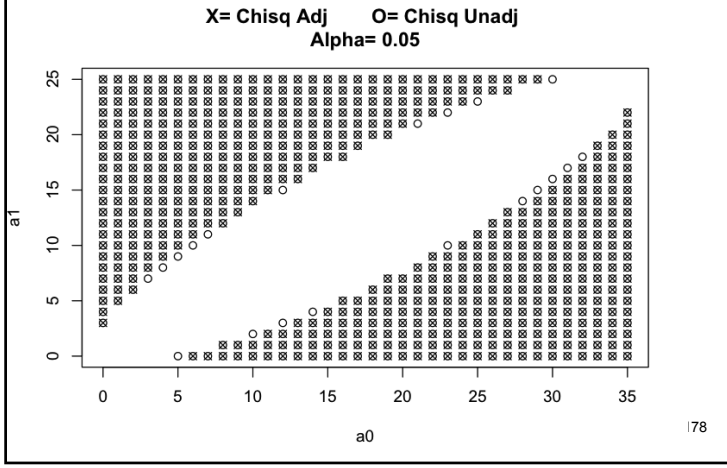
Regions of Correct Size for Asymptotics



Critical Regions for Fishers Exact Test



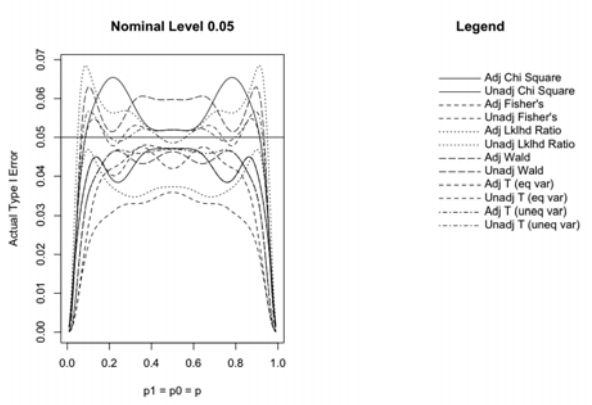
Critical Regions for Chi Square Test



78

Actual Type 1 Error

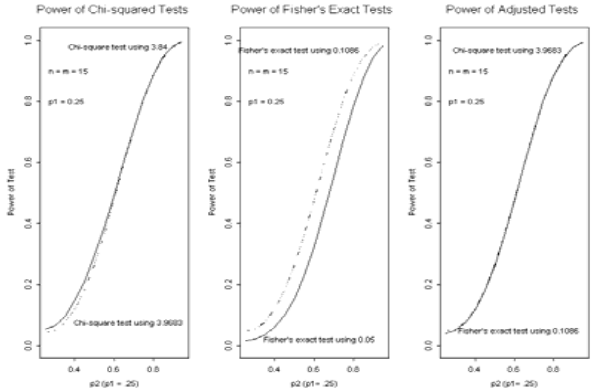
- Various unadjusted and adjusted tests as function of *p*



179

Gains in Power

- Power of unadjusted, adjusted level .05 tests



180

General Comments

- It is generally immaterial whether the Fisher's exact test P value or the chi square statistic or likelihood ratio statistic is used as the basis for the exact test
- In any case, the critical value is dependent upon the sample sizes
- Using this approach, substantial improvement in power is obtained in low sample sizes
- I strongly recommend its use when confronted with small samples in real life

181

Increasing Precision: Options

- Increase sample size
- Decrease V
  - Statistically
    - Alter statistical summary measure
    - Adjust for prognostic covariates
    - Alter study design (e.g., cross-over)
  - Scientifically
    - Improve reliability of measurements
    - Alter eligibility (decrease heterogeneity)
    - Alter clinical endpoint
- (Decrease confidence level)

182

Summary Measures: Relative Efficiency

- Example: Dichotomization of exponential time to event

$$iid\ Y_i \sim E(\mu), i = 1, \ldots, n$$
$$\hat{\theta} = \overline{Y} \quad V = \mu^2$$

$$W_i = 1_{[Y_i > c]} \sim B\left(1,\ p = e^{-c/\mu}\right)$$
$$\hat{p} = \overline{W} \quad \hat{\theta}^* = -\frac{c}{\log \hat{p}} \quad V = ??$$

183

Summary Measures: Relative Efficiency

- Example: Dichotomization of exponential time to event

By CLT : $\hat{p} \dot{\sim} N\left(p = e^{-c/\mu}, \frac{p(1-p)}{n}\right)$

Using δ method : $\theta = g(p) = -\frac{c}{\log p}$

$$g'(p) = -\frac{c}{p \log^2 p}$$

$$\hat{\theta}^* = -\frac{c}{\log \hat{p}} \dot{\sim} N\left(\theta, \frac{c^2(1-p)}{np \log^4 p} = \frac{\mu^4\left(1 - e^{-c/\mu}\right)}{nc^2 e^{-c/\mu}}\right)$$

184

Summary Measures: Relative Efficiency

- Example: Dichotomization of exponential time to event

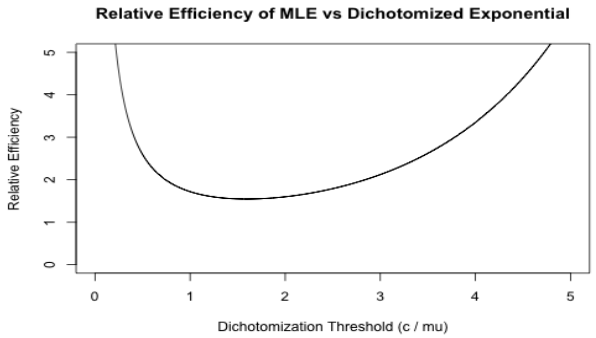
Relative Efficiency :

$$RE\left(\hat{\theta},\hat{\theta}^*\right)=\frac{\dfrac{\mu^4\left(1-e^{-c/\mu}\right)}{nc^2e^{-c/\mu}}}{\dfrac{\mu^2}{n}}=\frac{\left(1-e^{-c/\mu}\right)}{\left(c/\mu\right)^2e^{-c/\mu}}$$

185

Summary Measures: Relative Efficiency

- Minimum relative efficiency: 1.544 at c=1.594 (p= 0.203)



186

Dichotomization of Measurements

- In general, dichotomization (or categorization) of measurements will result in a loss of precision when assessed in a parametric model
  - The relative efficiency loss will depend upon the exact parametric model
- When viewed in a distribution-free manner, however, failure to dichotomize the data may mean that inference relates to the wrong measure of treatment effect
  - Quite often, we dichotomize data to reflect scientific importance
    - E.g., attaining normal SBP or blood glucose
    - E.g., surviving at least 28 days in critical care

187

Controlling Variation

- In a two sample comparison of means, we might control some variable in order to decrease the within group variability
  - Restrict population sampled
  - Standardize ancillary treatments
  - Standardize measurement procedure

188

Adjusting for Covariates

- When comparing means using stratified analyses or linear regression, adjustment for precision variables decreases the within group standard deviation
  - Var (Y | X) vs Var (Y | X, W)

189

Ex: Linear Regression

Unadj: $Y_i \mid X_i \overset{ind}{\sim} \left(\beta_0 + \beta_1 \times X_i, \sigma^2_{Y|X}\right),\ i = 1, \ldots, n$

Adj: $Y_i \mid X_i, W_i \overset{ind}{\sim} \left(\gamma_0 + \gamma_1 \times X_i + \gamma_2 \times W_i, \sigma^2_{Y|X,W}\right),\ i = 1, \ldots, n$

$\theta = \beta_1 = \gamma_1$ if balanced, stratified

$\theta = E\left(\hat{\beta}_1 \mid X\right) = E_W\left[E\left(\hat{\gamma}_1 \mid X, W\right)\right]$ if complete randomization

$se\left(\hat{\beta}_1\right) = \sqrt{\dfrac{\sigma^2_{Y|X}}{nVar(X)}}$ $\quad se\left(\hat{\gamma}_1\right) = \sqrt{\dfrac{\sigma^2_{Y|X,W}}{nVar(X)\left(1 - r^2_{XW}\right)}}$

$\sigma^2_{Y|X,W} = \sigma^2_{Y|X} - \gamma_2^2 Var(W \mid X)$

$r^2_{XW} = 0$ if balanced randomization, else $r^2_{XW} \approx 0$

190

Precision with Proportions

- When analyzing proportions (means), the mean variance relationship is important
  - Precision is greatest when proportion is close to 0 or 1
  - Greater homogeneity of groups makes results more deterministic
    - (At least, I always hope for this)
  - Hence, we should get lower within-group variance upon adjusting for prognostic variables

191

Ex: Diff of Indep Proportions

$ind\ Y_{ij} \sim B(1, p_i),\ i = 1,2;\ j = 1, \ldots, n_i$

$n = n_1 + n_2; \quad r = n_1 / n_2$

$\theta = p_1 - p_2 \qquad \hat{\theta} = \hat{p}_1 - \hat{p}_2 = \bar{Y}_{1\bullet} - \bar{Y}_{2\bullet}$

$\sigma_i^2 = p_i(1 - p_i)$

$V = (r+1)\left[\sigma_1^2 / r + \sigma_2^2\right] \qquad se\left(\hat{\theta}\right) = \sqrt{\dfrac{V}{n}} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

192

Precision with Odds

- When analyzing odds (a nonlinear function of the mean), adjusting for a precision variable results in more extreme estimates
  - odds = *p / (1-p)*
  - odds using average of stratum specific *p* is not the average of stratum specific odds
- Generally, little "precision" is gained due to the mean-variance relationship
  - Unless the precision variable is highly prognostic

193

Precision with Hazards

- When analyzing hazards, adjusting for a precision variable results in more extreme estimates
- The standard error tends to still be related to the number of observed events
  - Higher hazard ratio with same standard error ➔ greater precision

194

Adjustment for Covariates

- We "adjust" for other covariates
  - Define groups according to
    - Predictor of interest, and
    - Other covariates
  - Compare the distribution of response across groups which
    - differ with respect to the Predictor of Interest, but
    - are the same with respect to the other covariates
      - "holding other variables constant"

195

Unadjusted vs Adjusted Models

- Adjustment for covariates changes the scientific question
  - Unadjusted models
    - Slope compares parameters across groups differing by 1 unit in the modeled predictor
      - Groups may also differ with respect to other variables
  - Adjusted models
    - Slope compares parameters across groups differing by 1 unit in the modeled predictor but similar with respect to other modeled covariates

196

Interpretation of Slopes

- Difference in interpretation of slopes

$$\text{Unadjusted Model}: \quad g\left[\theta \,|\, X_i\right] = \beta_0 + \beta_1 \times X_i$$

  - β_1 = Compares θ for groups differing by 1 unit in X
    - (The distribution of W might differ across groups being compared)

$$\text{Adjusted Model}: \quad g\left[\theta \,|\, X_i, W_i\right] = \gamma_0 + \gamma_1 \times X_i + \gamma_2 \times W_i$$

  - γ_1 = Compares θ for groups differing by 1 unit in X, but agreeing in their values of W

197

Comparing models

Unadjusted $g\left[\theta \,|\, X_i\right] = \beta_0 + \beta_1 \times X_i$

Adjusted $g\left[\theta \,|\, X_i, W_i\right] = \gamma_0 + \gamma_1 \times X_i + \gamma_2 \times W_i$

Science: When is $\gamma_1 = \beta_1$?

When is $\hat{\gamma}_1 = \hat{\beta}_1$?

Statistics: When is $se(\hat{\gamma}_1) = se(\hat{\beta}_1)$?

When is $\hat{se}(\hat{\gamma}_1) = \hat{se}(\hat{\beta}_1)$?

198

General Results

- These questions can not be answered precisely in the general case
  - However, in linear regression we can derive exact results
  - These will serve as a basis for later examination of
    - Logistic regression
    - Poisson regression
    - Proportional hazards regression

199

Linear Regression

- Difference in interpretation of slopes

$$\text{Unadjusted Model}: \quad E\left[Y_i \,|\, X_i\right] = \beta_0 + \beta_1 \times X_i$$

  - β_1 = Diff in mean Y for groups differing by 1 unit in X
    - (The distribution of W might differ across groups being compared)

$$\text{Adjusted Model}: \quad E\left[Y_i \,|\, X_i, W_i\right] = \gamma_0 + \gamma_1 \times X_i + \gamma_2 \times W_i$$

  - γ_1 = Diff in mean Y for groups differing by 1 unit in X, but agreeing in their values of W

200

Relationships: True Slopes

- The slope of the unadjusted model will tend to be

$$\beta_1 = \gamma_1 + \rho_{XW}\frac{\sigma_W}{\sigma_X}\gamma_2$$

- Hence, true adjusted and unadjusted slopes for X are estimating the same quantity only if
  - $\rho_{XW} = 0$ (X and W are truly uncorrelated), OR
  - $\gamma_2 = 0$ (no association between W and Y after adjusting for X)

201

Relationships: Estimated Slopes

- The estimated slope of the unadjusted model will be

$$\hat{\beta}_1 = \hat{\gamma}_1\left(1 + \hat{\gamma}_2 r_{XW}\left[\frac{s_W}{s_X\left(r_{YX} - r_{YW}r_{XW}\right)}\right]\right)$$

- Hence, estimated adjusted and unadjusted slopes for X are equal only if
  - $r_{XW} = 0$ (X and W are uncorrelated in the sample, which can be arranged by experimental design), OR
  - $\hat{\gamma}_2 = 0$ (which cannot be predetermined, because Y is random)
  - $s_W = 0$ (*W* is controlled at a single value in which case $r_{XW} = 0$)

202

Relationships: True SE

Unadjusted Model $\left[se\left(\hat{\beta}_1\right)\right]^2 = \frac{Var\left(Y \mid X\right)}{nVar\left(X\right)}$

Adjusted Model $\left[se\left(\hat{\gamma}_1\right)\right]^2 = \frac{Var\left(Y \mid X, W\right)}{nVar\left(X\right)\left(1 - r_{XW}^2\right)}$

$$Var\left(Y \mid X\right) = \gamma_2^2 Var\left(W \mid X\right) + Var\left(Y \mid X, W\right)$$

$$\sigma_{Y|X}^2 = \gamma_2^2\sigma_{W|X}^2 + \sigma_{Y|X,W}^2$$

203

Relationships: True SE

Unadjusted Model $\left[se\left(\hat{\beta}_1\right)\right]^2 = \frac{Var\left(Y \mid X\right)}{nVar\left(X\right)}$

Adjusted Model $\left[se\left(\hat{\gamma}_1\right)\right]^2 = \frac{Var\left(Y \mid X, W\right)}{nVar\left(X\right)\left(1 - r_{XW}^2\right)}$

$$Var\left(Y \mid X\right) = \gamma_2^2 Var\left(W \mid X\right) + Var\left(Y \mid X, W\right)$$

Thus, $se\left(\hat{\beta}_1\right) = se\left(\hat{\gamma}_1\right)$ if

$$r_{XW} = 0$$

AND

$$\gamma_2 = 0 \quad \text{OR} \quad Var\left(W \mid X\right) = 0$$

204

Relationships: Estimated SE

Unadjusted Model $\left[s\hat{e}\left(\hat{\beta}_1\right)\right]^2 = \dfrac{SSE\left(Y \mid X\right)/(n-2)}{(n-1)s_X^2}$

Adjusted Model $\left[s\hat{e}\left(\hat{\gamma}_1\right)\right]^2 = \dfrac{SSE\left(Y \mid X, W\right)/(n-3)}{(n-1)s_X^2\left(1-r_{XW}^2\right)}$

$$SSE\left(Y \mid X\right) = \sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 \times X_i\right)^2$$
$$SSE\left(Y \mid X, W\right) = \sum\left(Y_i - \hat{\gamma}_0 - \hat{\gamma}_1 \times X_i - \hat{\gamma}_2 \times W_i\right)^2$$

205

Relationships: Estimated SE

Unadjusted Model $\left[s\hat{e}\left(\hat{\beta}_1\right)\right]^2 = \dfrac{SSE\left(Y \mid X\right)/(n-2)}{(n-1)s_X^2}$

Adjusted Model $\left[s\hat{e}\left(\hat{\gamma}_1\right)\right]^2 = \dfrac{SSE\left(Y \mid X, W\right)/(n-3)}{(n-1)s_X^2\left(1-r_{XW}^2\right)}$

Thus, $s\hat{e}\left(\hat{\beta}_1\right) = s\hat{e}\left(\hat{\gamma}_1\right)$ if

$r_{XW} = 0$

AND

$SSE\left(Y \mid X\right)/(n-2) = SSE\left(Y \mid X, W\right)/(n-3)$

206

Residual Squared Error

$$SSE\left(Y \mid X\right) = \sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 \times X_i\right)^2$$
$$SSE\left(Y \mid X, W\right) = \sum\left(Y_i - \hat{\gamma}_0 - \hat{\gamma}_1 \times X_i - \hat{\gamma}_2 \times W_i\right)^2$$

When calculated on the same data :

$SSE\left(Y \mid X\right) \geq SSE\left(Y \mid X, W\right)$

207

Relationships: Estimated SE

$$SSE\left(Y \mid X\right) = \sum\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 \times X_i\right)^2$$
$$SSE\left(Y \mid X, W\right) = \sum\left(Y_i - \hat{\gamma}_0 - \hat{\gamma}_1 \times X_i - \hat{\gamma}_2 \times W_i\right)^2$$

Now $\hat{\beta}_1 = \hat{\gamma}_1$ if

$\hat{\gamma}_2 = 0,$ in which case $SSE\left(Y \mid X\right) = SSE\left(Y \mid X, W\right)$

OR

$r_{XW} = 0,$ and $SSE\left(Y \mid X\right) > SSE\left(Y \mid X, W\right)$ if $\hat{\gamma}_2 \neq 0$

208

Special Cases

- Behavior of unadjusted and adjusted models according to whether
  - X and W are uncorrelated (no association in means)
  - W is associated with Y after adjustment for X

	$r_{XW} = 0$	$r_{XW} \neq 0$
$\gamma_2 \neq 0$	Precision	Confounding
$\gamma_2 = 0$	Irrelevant	Var Inflation

209

Simulations

- Unadjusted and adjusted estimates of treatment effect as a function of
  - Effect of a third covariate on mean outcome
  - Association between third covariate and treatment
    - Difference in mean covariate
    - Difference in median covariate

$$\text{Sampling} : E\left[W_i \middle| X_i\right] = \alpha_0 + \alpha_1 \times X_i$$

$$\text{Unadjusted} : g\left[\theta \middle| X_i\right] = \beta_0 + \beta_1 \times X_i$$

$$\text{Adjusted} : g\left[\theta \middle| X_i, W_i\right] = \gamma_0 + \gamma_1 \times X_i + \gamma_2 \times W_i$$

210

Linear Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.19)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	1.0	1.0 (0.28)	1.0 (0.20)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.28)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.3 (0.29)	0.0 (0.21)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.22)

211

Linear Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.19)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	1.0	1.0 (0.28)	1.0 (0.20)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.28)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.3 (0.29)	0.0 (0.21)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.22)

212

Linear Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.19)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	1.0	1.0 (0.28)	1.0 (0.20)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.28)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.3 (0.29)	0.0 (0.21)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.22)

213

Linear Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.19)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	1.0	1.0 (0.28)	1.0 (0.20)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.28)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.3 (0.29)	0.0 (0.21)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.22)

214

Linear Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.19)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	1.0	1.0 (0.28)	1.0 (0.20)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.28)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.3 (0.29)	0.0 (0.21)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.22)

215

Linear Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.19)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.28)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	1.0	1.0 (0.28)	1.0 (0.20)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.28)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.3 (0.29)	0.0 (0.21)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.22)

216

Logistic Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.42)	0.0 (0.42)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.40)	0.0 (0.42)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.42)	0.0 (0.43)
Precision	0.0	0.0	0.00	1.0	1.0	0.8 (0.43)	1.0 (0.49)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.43)	0.0 (0.48)
Confound	0.0	0.3	0.15	1.0	0.0	0.2 (0.41)	0.0 (0.47)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.41)	0.0 (0.47)

217

Logistic Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.42)	0.0 (0.42)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.40)	0.0 (0.42)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.42)	0.0 (0.43)
Precision	0.0	0.0	0.00	1.0	1.0	0.8 (0.43)	1.0 (0.49)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.43)	0.0 (0.48)
Confound	0.0	0.3	0.15	1.0	0.0	0.2 (0.41)	0.0 (0.47)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.41)	0.0 (0.47)

218

Logistic Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.42)	0.0 (0.42)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.40)	0.0 (0.42)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.42)	0.0 (0.43)
Precision	0.0	0.0	0.00	1.0	1.0	0.8 (0.43)	1.0 (0.49)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.43)	0.0 (0.48)
Confound	0.0	0.3	0.15	1.0	0.0	0.2 (0.41)	0.0 (0.47)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.41)	0.0 (0.47)

219

Logistic Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.42)	0.0 (0.42)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.40)	0.0 (0.42)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.42)	0.0 (0.43)
Precision	0.0	0.0	0.00	1.0	1.0	0.8 (0.43)	1.0 (0.49)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.43)	0.0 (0.48)
Confound	0.0	0.3	0.15	1.0	0.0	0.2 (0.41)	0.0 (0.47)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.41)	0.0 (0.47)

220

Logistic Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.42)	0.0 (0.42)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.40)	0.0 (0.42)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.42)	0.0 (0.43)
Precision	0.0	0.0	0.00	1.0	1.0	0.8 (0.43)	1.0 (0.49)
Confound	0.3	0.3	0.15	1.0	0.0	0.3 (0.43)	0.0 (0.48)
Confound	0.0	0.3	0.15	1.0	0.0	0.2 (0.41)	0.0 (0.47)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.41)	0.0 (0.47)

221

Proportional Hazards Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.21)	0.0 (0.22)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.21)	0.0 (0.21)
Precision	0.0	0.0	0.00	1.0	1.0	0.7 (0.21)	1.0 (0.22)
Confound	0.3	0.3	0.15	1.0	0.0	0.2 (0.21)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.1 (0.20)	0.0 (0.22)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.23)

222

Next

Proportional Hazards Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.21)	0.0 (0.22)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.21)	0.0 (0.21)
Precision	0.0	0.0	0.00	1.0	1.0	0.7 (0.21)	1.0 (0.22)
Confound	0.3	0.3	0.15	1.0	0.0	0.2 (0.21)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.1 (0.20)	0.0 (0.22)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.23)

223

Next

Proportional Hazards Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.21)	0.0 (0.22)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.21)	0.0 (0.21)
Precision	0.0	0.0	0.00	1.0	1.0	0.7 (0.21)	1.0 (0.22)
Confound	0.3	0.3	0.15	1.0	0.0	0.2 (0.21)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.1 (0.20)	0.0 (0.22)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.23)

224

Next

Proportional Hazards Regression

- Simulation results

	Truth					Estimates	
	ΔMdn	α_1	r_{XW}	γ_2	γ_1	β_1	γ_1
Irrelevant	0.0	0.0	0.00	0.0	0.0	0.0 (0.20)	0.0 (0.20)
Precision	0.0	0.0	0.00	1.0	0.0	0.0 (0.21)	0.0 (0.22)
Precision	- 0.3	0.0	0.00	1.0	0.0	0.0 (0.21)	0.0 (0.21)
Precision	0.0	0.0	0.00	1.0	1.0	0.7 (0.21)	1.0 (0.22)
Confound	0.3	0.3	0.15	1.0	0.0	0.2 (0.21)	0.0 (0.21)
Confound	0.0	0.3	0.15	1.0	0.0	0.1 (0.20)	0.0 (0.22)
Var Inflatn	0.0	1.0	0.45	0.0	0.0	0.0 (0.20)	0.0 (0.23)

225

Next

Precision: Linear Regression

- E.g., X, W independent in population (or completely randomized experiment) AND W associated with Y independent of X

$$\rho_{XW} = 0 \qquad \gamma_2 \neq 0$$

	True Value	Estimates
Slopes	$\beta_1 = \gamma_1$	$\hat{\beta}_1 \approx \hat{\gamma}_1$
Std Errs	$se(\hat{\beta}_1) > se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) > s\hat{e}(\hat{\gamma}_1)$

226

Precision: Logistic Regression

- Adjusting for a precision variable
  - Deattenuates slope away from the null
  - Standard errors reflect mean-variance relationship
    - Substantially increased power only in extreme cases
      - » (OR > 5 for equal samples sizes of binary W)

		True Value	Estimates
Slopes	$\beta_1 > 0$:	$\beta_1 < \gamma_1$	$\hat{\beta}_1 < \hat{\gamma}_1$
	$\beta_1 < 0$:	$\beta_1 > \gamma_1$	$\hat{\beta}_1 < \hat{\gamma}_1$
Std Errs		$se(\hat{\beta}_1) < se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) < s\hat{e}(\hat{\gamma}_1)$

227

Precision: Poisson Regression

- Adjusting for a precision variable
  - No effect on the slope (similar to linear regression)
    - log ratios are linear in log means
  - Standard errors reflect mean-variance relationship
    - Virtually no effect on power

	True Value	Estimates
Slopes	$\beta_1 = \gamma_1$	$\hat{\beta}_1 \approx \hat{\gamma}_1$
Std Errs	$se(\hat{\beta}_1) \approx se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) \approx s\hat{e}(\hat{\gamma}_1)$

228

Precision: PH Regression

- Adjusting for a precision variable
  - Deattenuates slope away from the null
  - Standard errors stay fairly constant
    - (Complicated result of binomial mean-variance)

		True Value	Estimates
Slopes	$\beta_1 > 0$:	$\beta_1 < \gamma_1$	$\hat{\beta}_1 < \hat{\gamma}_1$
	$\beta_1 < 0$:	$\beta_1 > \gamma_1$	$\hat{\beta}_1 > \hat{\gamma}_1$
Std Errs		$se(\hat{\beta}_1) \approx se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) \approx s\hat{e}(\hat{\gamma}_1)$

229

Lin Reg: Stratified Randomization

- Stratified (orthogonal) randomization in a designed experiment

$r_{XW} = 0 \qquad \gamma_2 \neq 0$

	True Value	Estimates
Slopes	$\beta_1 = \gamma_1$	$\hat{\beta}_1 = \hat{\gamma}_1$
Std Errs	$se(\hat{\beta}_1) = se(\hat{\gamma}_1)$	$s\hat{e}(\hat{\beta}_1) > s\hat{e}(\hat{\gamma}_1)$

230

Special Case: Baseline Adjustment

- Options
  - Final only (throw away baseline)
    - $V = 2\sigma^2$
  - Change (final – baseline)
    - $V = 4\sigma^2 (1 - \rho)$
  - ANCOVA (change or final adj for baseline
    - $V = 2\sigma^2 (1 - \rho^2)$

231

Ex: ANCOVA (Baseline Adjustment)

$$ind\ Y_{fi} \mid X_i \sim \left(\beta_0 + \beta_1 \times X_i + \beta_1 \times Y_{0i}, \sigma^2_{Y|X,Y_0}\right),$$
$$i = 1, \ldots, n \qquad \rho = corr\left(Y_{0i}, Y_{fi}\right)$$

$\theta = \beta_1 \qquad \hat{\theta} = \hat{\beta}_1$ from LS regression

$$V = \frac{\sigma^2_{Y|X}\left(1-\rho^2\right)}{Var(X)} \qquad se(\hat{\theta}) = \sqrt{\frac{\sigma^2_{Y|X}\left(1-\rho^2\right)}{nVar(X)}}$$

232

Comparison of Study Designs

- Single Arm: Mean; absolute reference N= 25
- Single Arm: Mean; historical data 50

- Two Arms : Diff in Means 100
- Two Arms : Diff in Mean Change (r = 0.3) 140
- Two Arms : Diff in Mean Change (r = 0.8) 40

- Two Arms : ANCOVA (r = 0.3) 81
- Two Arms : ANCOVA (r = 0.8) 36

- Cross-over: Diff in Means (r = 0.3) 70
- Cross-over: Diff in Means (r = 0.8) 20

233

General Comments: Alternative

- What alternative to use?
  - Minimal clinically important difference (MCID)
    - To detect? (use in sample size formula)
    - To declare significant? (look at critical value)
  - Subterfuge: 80% or 90%

234

General Comments: Level

- What level of significance?
  - "Standard": one-sided 0.025, two-sided 0.05
  - "Pivotal": one-sided 0.005?
    - Do we want to be extremely confident of an effect, or confident of an extreme effect

235

General Comments: Power

- What power?
  - Science: 97.5%
    - Unless MCID for significance→ ~50%
  - Subterfuge: 80% or 90%

236

Role of Secondary Analyses

- We choose a primary outcome to avoid multiple comparison problems
  - That primary outcome may be a composite of several clinical outcomes, but there will only be one CI, test
- We select a few secondary outcomes to provide supporting evidence or confirmation of mechanisms
  - Those secondary outcomes may be
    - alternative clinical measures and/or
    - different summary measures of the primary clinical endpoint

237

Secondary Analysis Models

- Selection of statistical models for secondary analyses should generally adhere to same principles as for primary outcome, including intent to treat
- Some exceptions:
  - Exploratory analyses based on dose actually taken may be undertaken to generate hypotheses about dose response
  - Exploratory cause specific time to event analyses may be used to investigate hypothesized mechanisms

238

Subgroups

- Testing for effects in K subgroups
  - Does the treatment work in each subgroup?
  - Bonferroni correction: Test at α / K
    - No subgroups: N = 100
    - Two subgroups: N = 230
- Testing for interactions across subgroups
  - Does the treatment work differently in subgroups?
    - Two subgroups: N = 400

239

Additional Constraints

- Safety analyses
  - Often there is a minimal number needed to treat in order to have enough data to rule out unacceptably high rates of extremely serious adverse events
    - "3 over n rule" as confidence bound when no such events observed
- Subgroup analyses
  - May need sufficient data to examine effects in important subgroups

240