Biost 518
Applied Biostatistics II

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

Lecture 2:
Precision of Inference

January 9, 2008

1

Lecture Outline

- Statistical Inference
- Measures of Precision
  - Standard errors
  - Width of confidence intervals
  - Statistical power

2

General Methods for Statistical Inference

3

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

4

Criteria for Summary Measure

- In order of importance
  - Scientifically (clinically) relevant
    - Also reflects current state of knowledge
  - Is likely to vary across levels of the factor of interest
    - Ability to detect variety of changes
  - Statistical precision
    - Only relevant if all other things are equal

5

Inference

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

6

Approximate Sampling Distn

- Most often we choose estimators that are asymptotically normally distributed

For large n: $\hat{\theta} \dot{\sim} N\left(\text{mean } \theta, \text{var} \frac{V}{n}\right)$

V is related to average "statistical information" from each observation

Often: V depends on the value of θ

7

Typical Method for 100(1-α)% CI

- When estimate is approximately normal

$$100(1-\alpha)\% \text{ confidence interval is } (\theta_L, \theta_U)$$

$$\theta_L = \hat{\theta} - z_{1-\alpha/2}\, s\hat{e}(\hat{\theta})$$

$$\theta_U = \hat{\theta} + z_{1-\alpha/2}\, s\hat{e}(\hat{\theta})$$

$$(estimate) \pm (crit\ val) \times (std\ error)$$

8

Computing P values using Z

Standardized statistic $Z = \frac{est - hyp}{std\ err} = \frac{\hat{\theta} - \theta_0}{s\hat{e}(\hat{\theta})} \sim N(0,1)$

Stata commands

Lower one-sided P value $\text{norm}\left(\frac{\hat{\theta} - \theta_0}{s\hat{e}(\hat{\theta})}\right)$

Upper one-sided P value $1 - \text{norm}\left(\frac{\hat{\theta} - \theta_0}{s\hat{e}(\hat{\theta})}\right)$

Two-sided P value $2 \times \text{norm}\left(-\text{abs}\left(\frac{\hat{\theta} - \theta_0}{s\hat{e}(\hat{\theta})}\right)\right)$

9

Aside: Comparing Estimates

- Comparisons across strata or studies
  - This is easy, if estimates are independent and approximately normally distributed

For independent $\hat{\theta}_1 \dot{\sim} N(\theta_1, se_1^2)$; $\hat{\theta}_2 \dot{\sim} N(\theta_2, se_2^2)$

$$\hat{\theta}_1 + \hat{\theta}_2 \dot{\sim} N(\theta_1 + \theta_2,\ se_1^2 + se_2^2)$$

$$\hat{\theta}_1 - \hat{\theta}_2 \dot{\sim} N(\theta_1 - \theta_2,\ se_1^2 + se_2^2)$$

$$\hat{\theta}_1 / \hat{\theta}_2 \dot{\sim} N\left(\frac{\theta_1}{\theta_2},\ \frac{1}{\theta_2^2}\left(se_1^2 + \frac{\theta_1^2}{\theta_2^2} se_2^2\right)\right)$$

10

Aside: Correlated Estimates

- If estimates are correlated and approximately normally distributed

For correlated $\hat{\theta}_1 \dot{\sim} N(\theta_1, se_1^2)$; $\hat{\theta}_2 \dot{\sim} N(\theta_2, se_2^2)$

$$\omega = corr(\hat{\theta}_1, \hat{\theta}_2)$$

$$\hat{\theta}_1 + \hat{\theta}_2 \dot{\sim} N(\theta_1 + \theta_2,\ se_1^2 + se_2^2 + 2\omega\, se_1\, se_2)$$

$$\hat{\theta}_1 - \hat{\theta}_2 \dot{\sim} N(\theta_1 - \theta_2,\ se_1^2 + se_2^2 - 2\omega\, se_1\, se_2)$$

11

Measures of Precision

12

Classical Hypothesis Tests

- Only stated in terms of null hypothesis

One-sided test of greater alternative:

$$H_0 : \theta \le \theta_0 \quad vs \quad H_1 : \theta > \theta_0$$

One-sided test of lesser alternative:

$$H_0 : \theta \ge \theta_0 \quad vs \quad H_1 : \theta < \theta_0$$

Two-sided test:

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \ne \theta_0$$

13

Classical Conclusions

- Either
  - Reject null hypothesis
    - Because data is atypical of what would be expected when null is true
  - Do not reject null hypothesis
    - Because we cannot say that the data is atypical of what would be expected when null is true
      - Either null is true, or
      - Null is false and we "lack power"

14

Decision Theoretic Approach

- For suitably chosen "design alternative"

One-sided test of greater alternative:

$$H_0 : \theta \le \theta_0 \quad vs \quad H_1 : \theta \ge \theta_1$$

One-sided test of lesser alternative:

$$H_0 : \theta \ge \theta_0 \quad vs \quad H_1 : \theta \le \theta_1$$

Two-sided test:

$$H_0 : \theta = \theta_0 \quad vs \quad H_1 : \theta \le -\theta_1 \; or \; \theta \ge \theta_1$$

15

Decision Theoretic Conclusions

- Either
  - Reject null hypothesis
    - Because data is atypical of what would be expected when null is true
  - Reject null hypothesis
    - Because data is atypical of what would be expected when design alternative is true
      - Ideally, we use the same definition of "atypical" when considering the alternative as when we consider the null

16

Choice of Design Alternative

- The design alternative is ideally the "minimal important difference to detect"
  - Considerations include
    - Differences having clinical impact
      - Is it important to lower mean cholesterol by 1 mg / dl
    - Differences having economic value
      - New drug is not marketable unless there is large effect
    - Feasibility of study
      - Availability of subjects might limit studies to search for interventions having large impact

17

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

18

Sample Size Determination

- Based on CI: Choose sample size such that
  - 95% CI will not include both the null and the design alternative
- Based on statistical power: Choose sample size such that when alternative it true, there is a high probability of rejecting the null hypothesis
  - Minimize the type II error

19

Statistical Power

- The probability of rejecting the null hypothesis

$$Pwr(\theta) = \Pr(reject\ H_0 \mid \theta = \theta_0)$$

$$\text{Most often}: \quad \hat{\theta} \sim N\left(\theta, \frac{V}{n}\right) \qquad \text{Test statistic } Z = \frac{\hat{\theta} - \theta_0}{\sqrt{V/n}}$$

$$\text{Under } H_0 : Z \sim N(0,1) \quad \rightarrow \quad \text{Reject } H_0 \text{ if } |Z| > z_{1-\alpha/2}$$

$$\text{Under } H_1 : Z \sim N\left(\frac{\theta_1 - \theta_0}{\sqrt{V/n}},\ 1\right)$$

20

Power Function / Curve

- Power is a function of the true value of θ
  - *Pwr* (θ_0) = α , the type I error
- Hence:
  - For every θ we can compute the power,
  - For any choice of desired power, there is some θ such that the study has that power
  - In particular, *Pwr* (θ_0) = α , the type I error

21

Std Errors: Key to Precision

- Greater precision is achieved with smaller standard errors

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

$$(V \text{ related to average "statistical information"})$$

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

$$\text{Test statistic}: \quad Z = \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})}$$

22

Ex: One Sample Mean

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

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

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

23

Ex: Difference of Indep Means

$$ind\ Y_{ij} \sim (\mu_i, \sigma_i^2),\ i = 1,2;\ j = 1, \ldots, n_i$$

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

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

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

24

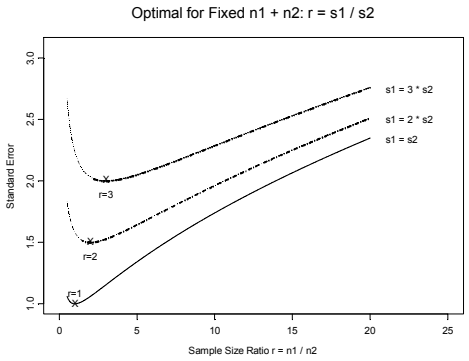
Comment: Optimal *r* (Fixed *n*)

- Suppose we are constrained by maximal sample size $n = n_1 + n_2$
  - Smallest *V* when

$$r = \frac{n_1}{n_2} = \frac{\sigma_1}{\sigma_2}$$

25

Comment: Optimal *r* (Fixed *n*)



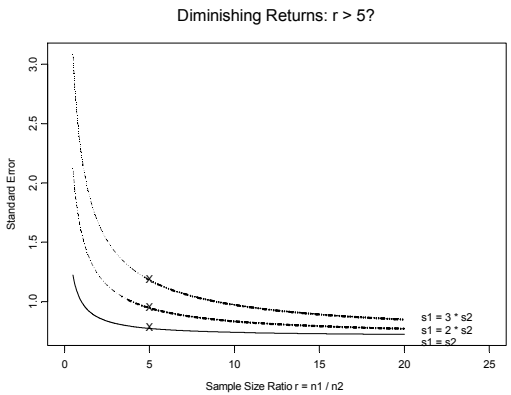
26

Comment: Diminishing Returns

- When we are unconstrained by maximal sample size we still hit a point of diminishing returns
  - Often quoted: *r = 5*

27

Comment: Diminishing Returns



28

Ex: Difference of Paired Means

$$Y_{ij} \sim \left(\mu_i, \sigma_i^2\right), i = 1,2;\ j = 1,\ldots,n$$
$$corr\left(Y_{1j}, Y_{2j}\right) = \rho; \quad corr\left(Y_{ij}, Y_{mk}\right) = 0 \text{ if } j \neq k$$
$$\theta = \mu_1 - \mu_2 \qquad \hat{\theta} = \overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}$$

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

29

Comment

- Note that gains in precision are only obtained if the data on matched observations are positively correlated
  - But this is usually the case
  - Possible exceptions
    - Sleep on successive nights?
    - Intrauterine growth of littermates?
    - Mileage on successive tanks of gas?

30

Clustered Data

- Experiments with clustered data occur when "treatments" are assigned on the basis of
  - Households
  - Schools
  - Clinics
  - Cities

31

Ex: Mean of Clustered Data

$$Y_{ij} \sim \left(\mu, \sigma^2\right), i = 1,\ldots,n;\ j = 1,\ldots,m$$
$$corr\left(Y_{ij}, Y_{ik}\right) = \rho \text{ if } j \neq k; \quad corr\left(Y_{ij}, Y_{mk}\right) = 0 \text{ if } i \neq m$$
$$\theta = \mu_1 - \mu_2 \qquad \hat{\theta} = \overline{Y}_{1\bullet} - \overline{Y}_{2\bullet}$$

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

32

Comment

- Even small correlations are important to consider if cluster sizes are large
  - Equal precision achieved with
    - 1000 clusters with m= 1, ρ = 0.01 (Tot N = 1000)
    - 650 clusters with m= 2, ρ = 0.30 (Tot N = 1300)
    - 550 clusters with m= 2, ρ = 0.10 (Tot N = 1100)
    - 190 clusters with m= 10, ρ = 0.10 (Tot N = 1900)
    - 109 clusters with m= 10, ρ = 0.01 (Tot N = 1090)
    - 20 clusters with m= 100, ρ = 0.01 (Tot N = 2000)

33

Ex: Independent Odds Ratios

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

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

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

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

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

34

Comment

- Note that maximal precision is achieved when the underlying odds are near 1
  - Proportions are near 0.5
- If we were considering the difference in proportions, maximal precision is achieved when the underlying proportions are near 0 or 1

35

Ex: Hazard Ratios

$$ind \text{ censored time to event } (T_{ij}, \delta_{ij}),$$

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

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

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

36

Comment

- In the proportional hazards model, statistical information is roughly proportional to the number of observed events *d*
  - Hence the importance of reporting number of events in a paper

37

Ex: Linear Regression

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

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

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

38

Comment

- Precision tends to increase as the predictor is measured over a wider range
  - Somewhat related to “leverage”
- Precision also related to the within group variance

39

Increasing Precision

- Options
  - Increase sample size
  - Decrease *V*
  - (Decrease confidence level)

40

Criteria for Precision

- Standard error
- Width of confidence interval
- Statistical power
  - Probability of rejecting the null hypothesis
    - Select "design alternative"
    - Select desired power

41

Sample Size Computation

Number of "sampling units" to obtain desired precision

$$\text{Level of significance } \alpha \text{ when } \theta = \theta_0$$

$$\text{Power } \beta \text{ when } \theta = \theta_1$$

$$\text{Variability } V \text{ within 1 sampling unit}$$

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

42

When Sample Size Constrained

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

$$\beta = \Phi\left(\frac{\left(\theta_1 - \theta_0\right)}{\sqrt{V/n}} - z_{1-\alpha/2}\right)$$

  - Compute alternative detected with high power

$$\theta_1 = \theta_0 + \left(z_{1-\alpha/2} + z_\beta\right)\sqrt{\frac{V}{n}}$$

43

General Comments

- Sample size required behaves like the square of width of CI
- Positively correlated observations within the same group provide less precision than same number of independent observations
- Positively correlated observations across groups provide more precision

44

What Power to Use

- Science versus subterfuge
  - Most popular: 80% or 90%
  - Most rational (I think): 97.5%
    - Recall The Scientist Game

45