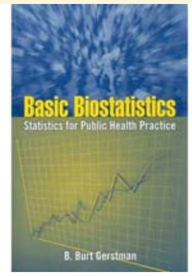
Basic Biostatistics
Statistics for Public Health Practice
B. Burt Gerstman

Chapter 16: Inference About a Proportion

April 10

Structure of the book

- **Section 1**
  (Ch 1 – 10)
  Basic concepts and techniques
- **Section 2**
  (Ch 11 – 15):
  Inference for quantitative outcomes
- **Section 3**
  (Ch 16 – 19):
  Inference for categorical outcomes



2

In Chapter 16:

+ ≡ covered in introductory recorded lecture

3

Binary Variables

Binary variable ≡ two possible outcomes: “success” or “failure”

Examples

- Current smoker (Y/N)
- Gender (M/F)
- Survival 5+ years (Y/N)
- Case or non-case (Y/N)

4

Binomial Proportions

Start by calculating sample proportion “p-hat”:

$$\hat{p} = \frac{x}{n} \quad \text{where } x \equiv \text{no. of successes}$$

Illustration: An SRS of 57 adults identifies 17 smokers. Therefore, the sample proportion is:

$$\hat{p} = \frac{x}{n} = \frac{17}{57} = .2982$$

5

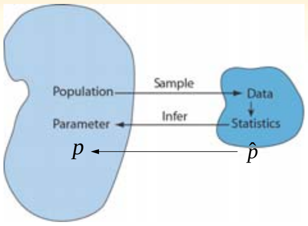
Two Special Proportions

Incidence proportion ≡ proportion at risk that develop a condition over a specified period of time ≡ “average risk”

Prevalence proportion ≡ proportion with the characteristic or condition at a particular time

6

Inference about p

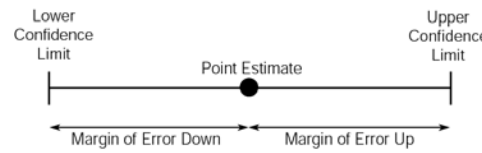


$\hat{p}$ is the point estimator of parameter p

7

Confidence Interval for p

The true value of parameter p will never be known with absolute certainty, but can be estimated with $(1 - \alpha)100\%$ confidence



9

Plus-four CI method for p

$$(1-\alpha)100\% \text{ CI for } p = \tilde{p} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\tilde{p}\tilde{q}}{\tilde{n}}}$$

where

$\tilde{n} = n + 4$ $\quad$ $\tilde{x} = x + 2$

$\tilde{p} = \frac{\tilde{x}}{\tilde{n}}$ $\quad$ $\tilde{q} = 1 - \tilde{p}$

10

Illustration

Data: 17 smokers in an SRS on n = 57

$\tilde{n} = 57 + 4 = 61$ $\quad$ $\tilde{x} = 17 + 2 = 19$

$\tilde{p} = \frac{19}{61} = .3115$ $\quad$ $\tilde{q} = 1 - .3115 = .6885$

For 95% confidence, use $z_{1-.05/2}$ = 1.96

$$95\% \text{ CI for } p = \tilde{p} \pm z \cdot \sqrt{\frac{\tilde{p}\tilde{q}}{\tilde{n}}} = .3115 \pm (1.96)\sqrt{\frac{(.3115)(.6885)}{61}}$$

$$= .3115 \pm .1162$$

$$= (.1953, .4277)$$

11

Testing a Proportion

A. Hypothesis statements

B. z statistic or binomial procedure*

* In small samples (less than 5 successes expected) the exact binomial procedure is required

C. *P*-value with interpretation

13

Hypothesis Testing Example

A survey in a particular community shows a sample prevalence of 17 of 57 (.2982)

The National Center for Health Statistics reports a national prevalence of 21%.

National Center for Health Statistics

Do data support the *claim* that the particular community has a prevalence that exceeds 21%?

Under the null hypothesis: H_0: p = .21
Note: The value of p under the null hypothesis (call this p_0) comes from the research question, NOT the data.

H_a: p > .21 (one-sided) or H_a: $p \neq .21$ (two-sided)

14

Z statistic

$$z_{\text{stat}} = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

where

$p_0 \equiv$ expected proportion under H_0

$q_0 = 1 - p_0$

$$\text{Data}: \hat{p} = \frac{17}{57} = .2982$$

$$z_{\text{stat}} = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} = \frac{.2982 - .21}{\sqrt{(.21)(1 - .21)/57}} = 1.63$$

15

P-value (example)

Use Table B or a computer program to determine area under curve beyond the z_{stat} of 1.63

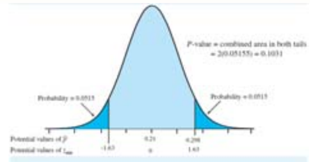


FIGURE 16.5 *P*-value, smoking prevalence illustrative example.

From Table B:

One-tailed *P-value* = .0515

Two-tailed *P-value* = 2 × .0515 = .1030

Marginal evidence against H_0 (marginal significance)

16

Summary of Illustrative Test

- H_0: p = .21
- 17 of 57 (.2982) were smokers
- Two-tailed *P*-value = .10
- The evidence against the claim made by H_0 is marginally significant

My *P* value is marginally small!?

17

Hypothesis Test with the CI

- H_0: $p = p_0$ can be tested at the α-level of significance with the (1–α)100% CI
- Example:
  Test H_0: p = .21 at α = .05
- 95% CI for p = (.195 to .428) [prior slide]
- CI does NOT exclude a proportion of .21
- ∴ The difference is NOT significant at α = .05

18