JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

JOHNS HOPKINS BLOOMBERG SCHOOL *of* PUBLIC HEALTH

Section H

Small Sample Considerations for Confidence Intervals for Population Proportions

The Central Limit Theorem (CLT)

- The Central Limit Theorem (CLT) is a powerful mathematical tool that gives several useful statistics including:
  - The sampling distribution of sample proportions based on all samples of same size *n* is *approximately* normal
  - Mother/infant transmission example, placebo group:
  - CLT 95% CI:
    (can be done by hand)

$$(.158,.282) \approx (.16,.28)$$

$$\rightarrow \textit{16\% to 28\%}$$

  - Exact 95% CI:
    (requires computer, always correct)

```
From Òcii 183 40Ó command

 [95% Conf. Interval]

---------------------

.160984    .2855248
```

Notes on 95% Confidence Interval for Proportion

- The CLT based formula for a 95% CI is only *approximate*; it works very well if you have enough data in your sample

- The approximation works better the bigger $n \times \hat{p} \times (1 - \hat{p})$

- "Large sample" for binary outcomes is not only a function of total sample size *n*, but the split between "yes" and "no" outcomes

Mother/Infant Transmission: AZT Group

- Mother/infant transmission example, AZT group:
  - (n = 180, $\hat{p} = \frac{13}{180} = .07$)

- CLT 95% CI:
  (can be done by hand)

$$(.032,.108) \approx (.03,.11)$$

$\rightarrow$ *3% to 11%*

- Exact 95% CI:
  (requires computer,
  always correct)

```
From Òcii 180 13Ó command

 [95% Conf. Interval]

---------------------

 .0390137    .1203358
```

Mother/Infant Transmission CIs

- In the placebo sample

$$n \times \hat{p}_{plac} \times (1 - \hat{p}_{plac}) =$$
$$183 * .22 * .78 \approx 31$$

- In the AZT sample

$$n \times \hat{p}_{AZT}(1 - \hat{p}_{AZT}) =$$
$$180 * .07 * .93 \approx 12$$

Notes on 95% Confidence Interval for Proportion

- You do not use the t-correction for small sample sizes like we did for sample means
  - We use exact binomial calculations

- Interpretation of 95% CIs exactly the same with either method
  - In real life, using computer will always give valid result
  - CLT only breaks down with “small” sample sizes
  - In testing situations you will not be required to do exact CIs!

Really Small Sample Example for Illustration

- Random sample of 16 patients on drug A: two of sixteen patients experience drug failure in first month
  - CLT 95% CI: $\hat{p} \pm 2 \times S\hat{E}(\hat{p}) \rightarrow$

    $$\frac{2}{16} \pm 2 \times \sqrt{\frac{(2/16) \times (1 - 2/16)}{16}} \rightarrow$$

    $$(-0.05, 0.28)$$

  - Exact 95% CI: (0.02, 0.38)

```
. cii 16 2

                                                         -- Binomial Exact --
    Variable |        Obs        Mean    Std. Err.       [95% Conf. Interval]
-------------+---------------------------------------------------------------
             |         16        .125    .0826797        .0155136    .3834762
```