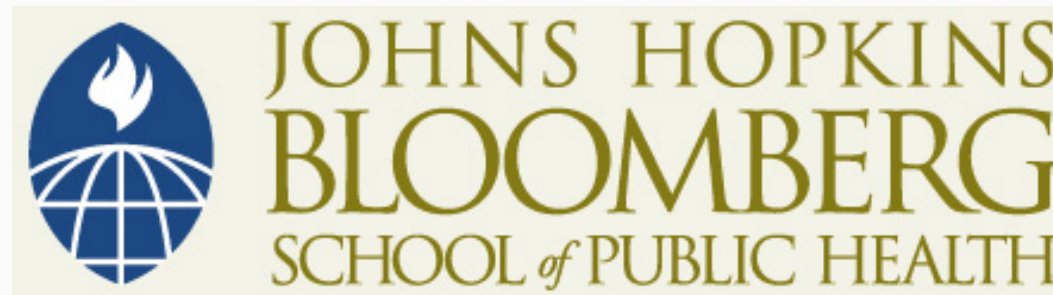


This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2009, The Johns Hopkins University and John McGready. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.



JOHNS HOPKINS
BLOOMBERG
SCHOOL *of* PUBLIC HEALTH

Section G

Estimating Confidence Intervals for the Proportion of a Population Based on a Single Sample of Size n : Some Examples

Estimating a 95% Confidence Interval

- In last section we defined a 95% confidence interval for the population proportion p

- Interval given by $\hat{p} \pm 2SE(\hat{p}) : \hat{p} \pm 2 * \sqrt{\frac{p \times (1 - p)}{n}}$

- Problem: we don't know p

- Can estimate with \hat{p} , such that our estimated SE is

- $SE(\hat{p}) = \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{n}}$

- Estimated 95% CI for based on a single sample of size n

- $\hat{p} \pm 2 * \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{n}}$

Example 1

- Proportion of dialysis patients with national insurance in 12 countries (only six shown . . .)

EXHIBIT 1
Descriptive Measures Of The Prevalent Cross-Sectional Patient Sample, Dialysis Patients In Twelve Countries, 2002-2004

	A/NZ (n = 561)	BEL (n = 468)	CAN (n = 503)	FRA (n = 481)	GER (n = 524)	ITA (n = 540)
Mean age (years)	59.9 (14.7)	66.2 (13.4)	62.1 (14.7)	64.1 (14.5)	61.7 (14.1)	64 (13.7)
Minority ^a	21.5%	5.3%	18.7%	7.1%	0.4%	0.4%
Income (\$US)						
<\$20,000	85.0%	73.4%	71.8%	67.0%	59.7%	78.3%
\$20,000-\$39,000	9.1	17.5	20.8	21.8	27.1	17.4
≥\$40,000	5.9	9.1	7.4	11.2	13.1	4.2
Insurance type						
National only	69.8%	74.1%	79.6%	45.5%	95.4%	99.6%
Private only	5.4	0.4	0.2	0.2	2.9	0.0
Mean number of comorbid conditions ^b	3.7 (2)	3.9 (2.1)	4.1 (2.1)	3.1 (1.9)	3.4 (2.1)	2.7 (1.9)
Mean number of prescribed medications	8.7 (3.6)	9.9 (4.1)	12.6 (4.8)	7.7 (3.5)	9.7 (3.5)	6.4 (3.6)

- Example, France: $\hat{p} = \frac{219}{481} = .46$

Example 1

- Estimated confidence interval

$$\hat{p} \pm 2 \times \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{n}}$$

$$.46 \pm 2 \times \sqrt{\frac{.46 \times (1 - .46)}{481}}$$

$$.46 \pm 2 \times .023$$

$$.46 \pm .046$$

$$(.414, .505) \approx (.41, .51)$$

→ 41% to 51%

Example 1 in Stata

- Can use *cii* command for binary outcomes to get CIs for p
- Syntax: *cii n y*
 - Where n is the total sample size, y is number of “yes” outcomes
- National health insurance in France

```
. cii 481 219, bin
```

Variable	Obs	Mean	Std. Err.	-- Binomial Exact -- [95% Conf. Interval]	
	481	.4553015	.0227068	.4101514	.5010042

Example 2

- Maternal/infant transmission of HIV
 - HIV-infection status was known for 363 births (180 in the zidovudine (AZT) group and 183 in the placebo group)
 - Thirteen infants in the zidovudine group and 40 in the placebo group were HIV-infected

$$\hat{p}_{AZT} = \frac{13}{180} = 0.07 = 7\%$$

$$\hat{p}_{PLAC} = \frac{40}{183} = 0.22 = 22\%$$

Example 2

- Estimated confidence interval for transmission percentage in the placebo group

$$\hat{p} \pm 2 \times \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{n}}$$

$$.22 \pm 2 \times \sqrt{\frac{.22 \times (1 - .22)}{183}}$$

$$.22 \pm 2 \times .031$$

$$.46 \pm .062$$

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

→ 16% to 28%

Example 2 in Stata

- Results from *cii* command

```
. cii 183 40
```

```
Variable | Obs      Mean      Std. Err.      -- Binomial Exact --  
          |          [95% Conf. Interval]  
-----+-----  
          | 183      .2185792    .0305507    .160984    .2855248
```

Notes on 95% Confidence Interval for Proportion

- Sometimes $\pm 2 \text{SE}(\hat{p})$ is called
 - 95% error bound
 - Margin of error