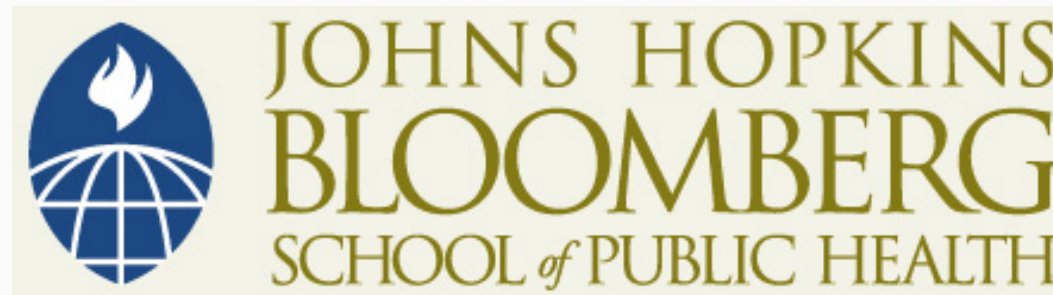


This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



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 E

---

The Sample Proportion as a Summary Measure for Binary Outcomes and the CLT

## Proportions (p)

- Proportion of individuals with health insurance
- Proportion of patients who became infected
- Proportion of patients who are cured
- Proportion of individuals who are hypertensive
- Proportion of individuals positive on a blood test
- Proportion of adverse drug reactions
- Proportion of premature infants who survive

## Proportions (p)

- For each individual in the study, we record a binary outcome (Yes/No; Success/Failure) rather than a continuous measurement

## Proportions (p)

- Compute a sample proportion,  $\hat{p}$  (pronounced “p-hat”), by taking observed number of “yes” responses divided by total sample size
  - This is the key summary measure for binary data, analogous to a mean for continuous data
  - There is a formula for the standard deviation of a proportion, but the quantity lacks the “physical interpretability” that it has for continuous data

# Example 1

- Proportion of dialysis patients with national insurance in 12 countries (only six shown..)<sup>1</sup>

**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 <sup>a</sup>	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 <sup>b</sup>	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: Canada:  $\hat{p} = \frac{400}{503} = 0.796$

Notes: <sup>1</sup> Hirth, R., et al. (2008). Out-of-pocket spending and medication adherence among dialysis patients in twelve countries, *Health Affairs*, 27 (1).

## Example 2

- Maternal/infant transmission of HIV<sup>1</sup>
- 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

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

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

Notes: <sup>1</sup>Spector, S., et al. (1994). A controlled trial of intravenous immune globulin for the prevention of serious bacterial infections in children receiving zidovudine for advanced human immunodeficiency virus infection, *New England Journal of Medicine* 331 (18).

## Proportions (p)

- What is the sampling behavior of a sample proportion?
- In other words, how do sample proportions, estimated from random samples of the same size from the same population, behave?

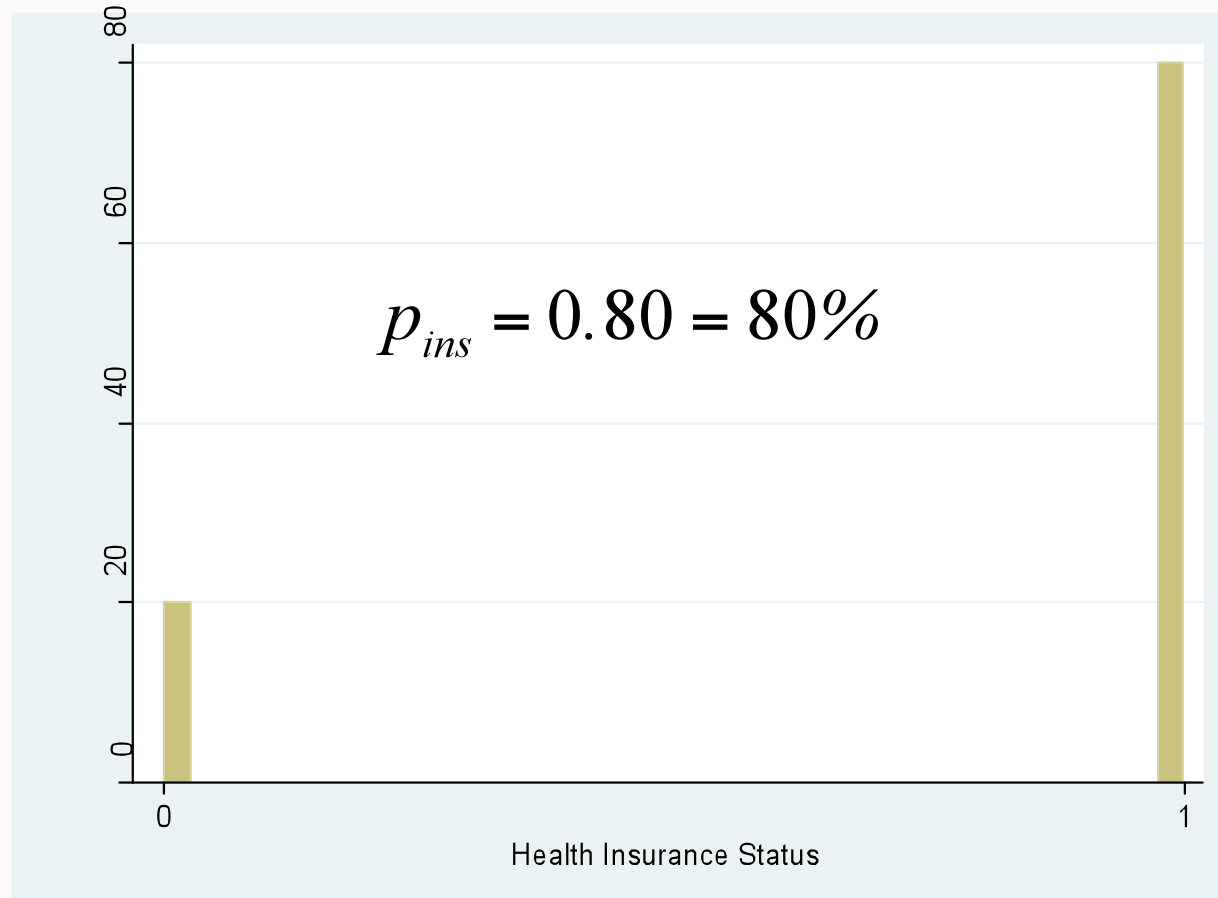


## Proportions (p)

- Suppose we have a population in which 80% of persons have some form of health insurance and 20% have no health insurance

## Example: Health Insurance Coverage

- Assume the population distribution is given by the following:

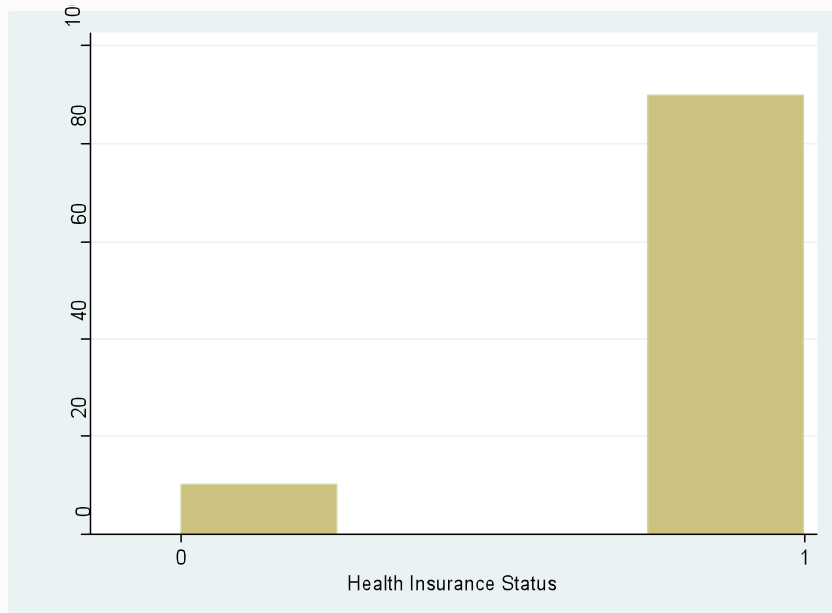


## Example: Health Insurance Coverage

- Suppose we had all the time in the world (leftover from last time)
- We decide to do another set of experiments
- We are going to take 500 separate random samples from this population, each with 20 subjects
- For each of the 500 samples, we will plot a histogram of the sample proportion of insured individuals and record the sample proportion
- Ready, set, go . . .

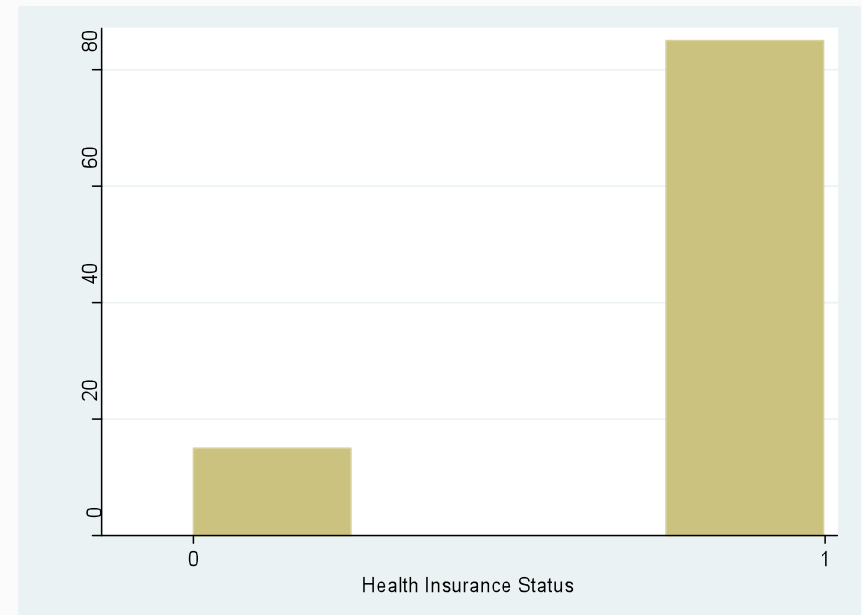
# Random Samples

■ Sample 1:  $n = 20$



$$\hat{p}_{ins} = 0.90 = 90\%$$

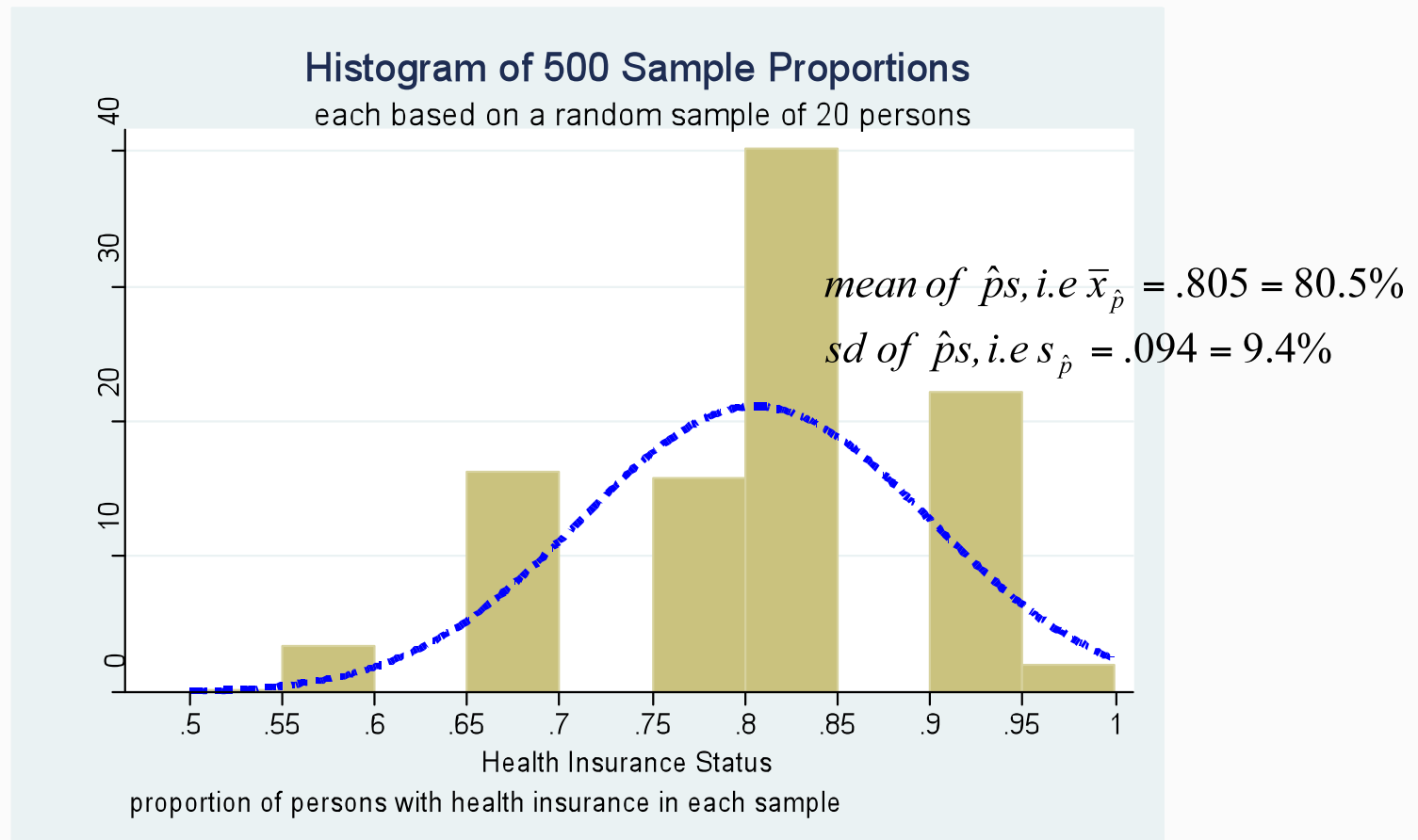
■ Sample 2:  $n = 20$



$$\hat{p}_{ins} = 0.85 = 85\%$$

# Estimated Sampling Distribution

- So we did this 500 times: now let's look at a histogram of the 500 proportions

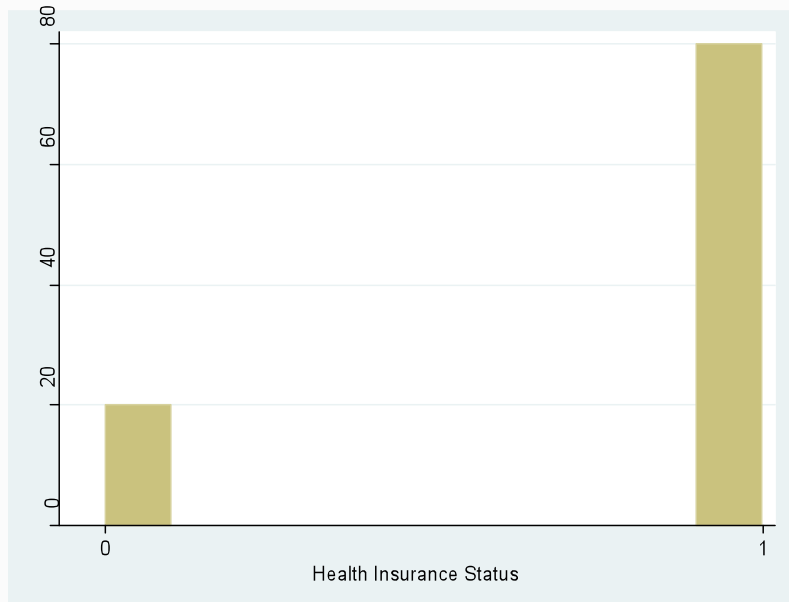


## Example: Health Insurance Coverage

- We decide to do one more experiment
- We are going to take 500 separate random samples from this population, each with 100 subjects
- For each of the 500 samples, we will plot a histogram of the sample proportion of insured individuals and record the sample proportion
- Ready, set, go . . .

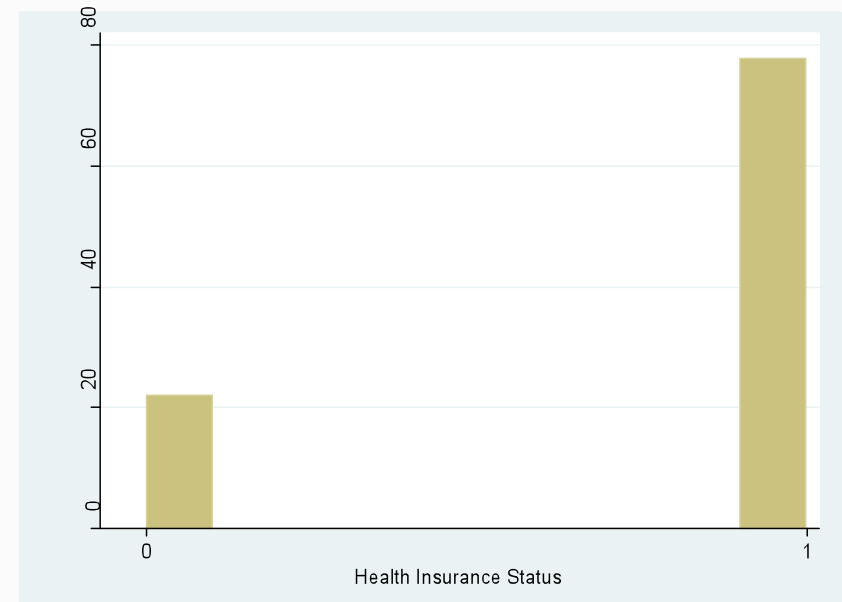
# Random Samples

■ Sample 1:  $n = 100$



$$\hat{p}_{ins} = 0.80 = 80\%$$

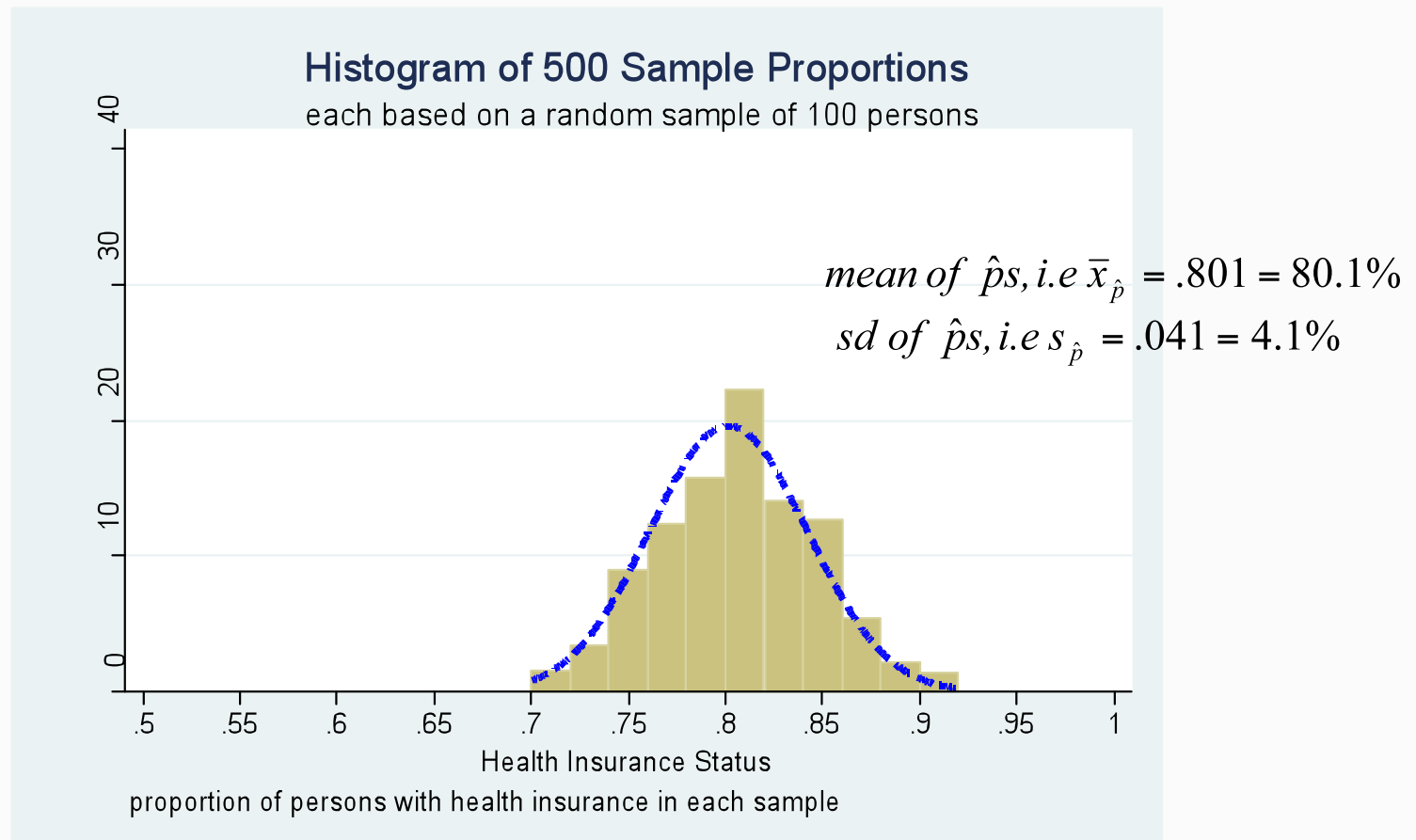
■ Sample 2:  $n = 100$



$$\hat{p}_{ins} = 0.78 = 78\%$$

# Example: Blood Pressure of Males

- So we did this 500 times: now let's look at a histogram of the 500 proportions



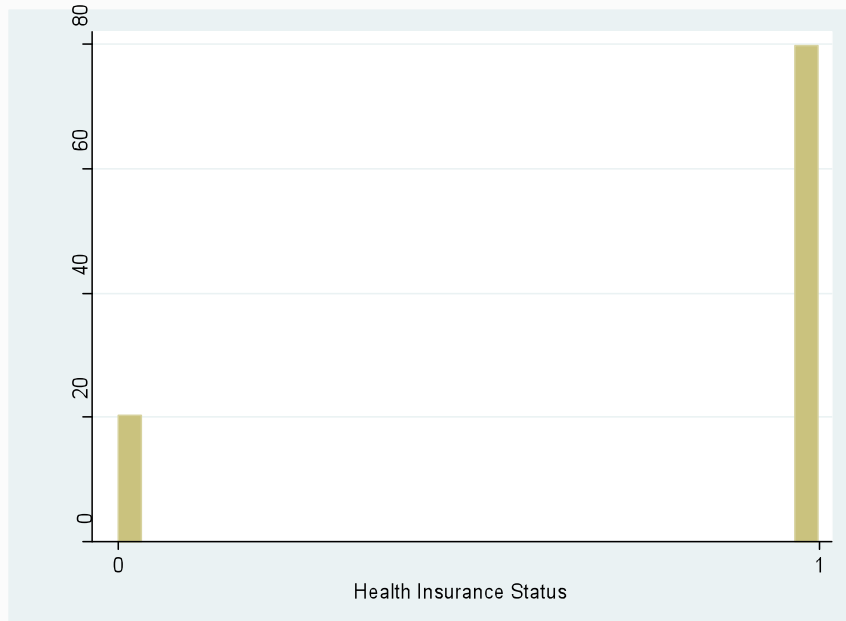


## Example: Health Insurance Coverage

- We decide to do one more experiment
- We are going to take 500 separate random samples from this population, each with 1,000 subjects
- For each of the 500 samples, we will plot a histogram of the sample proportioned of insured individuals, and record the sample proportion
- Ready, set, go . . .

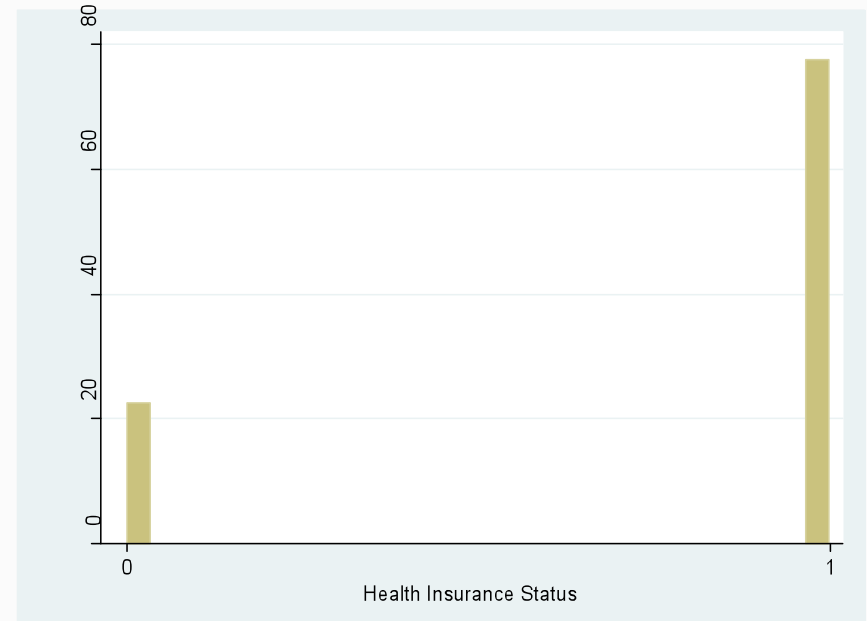
# Random Samples

- Sample 1:  $n = 1,000$



$$\hat{p}_{ins} = 0.798 = 79.8\%$$

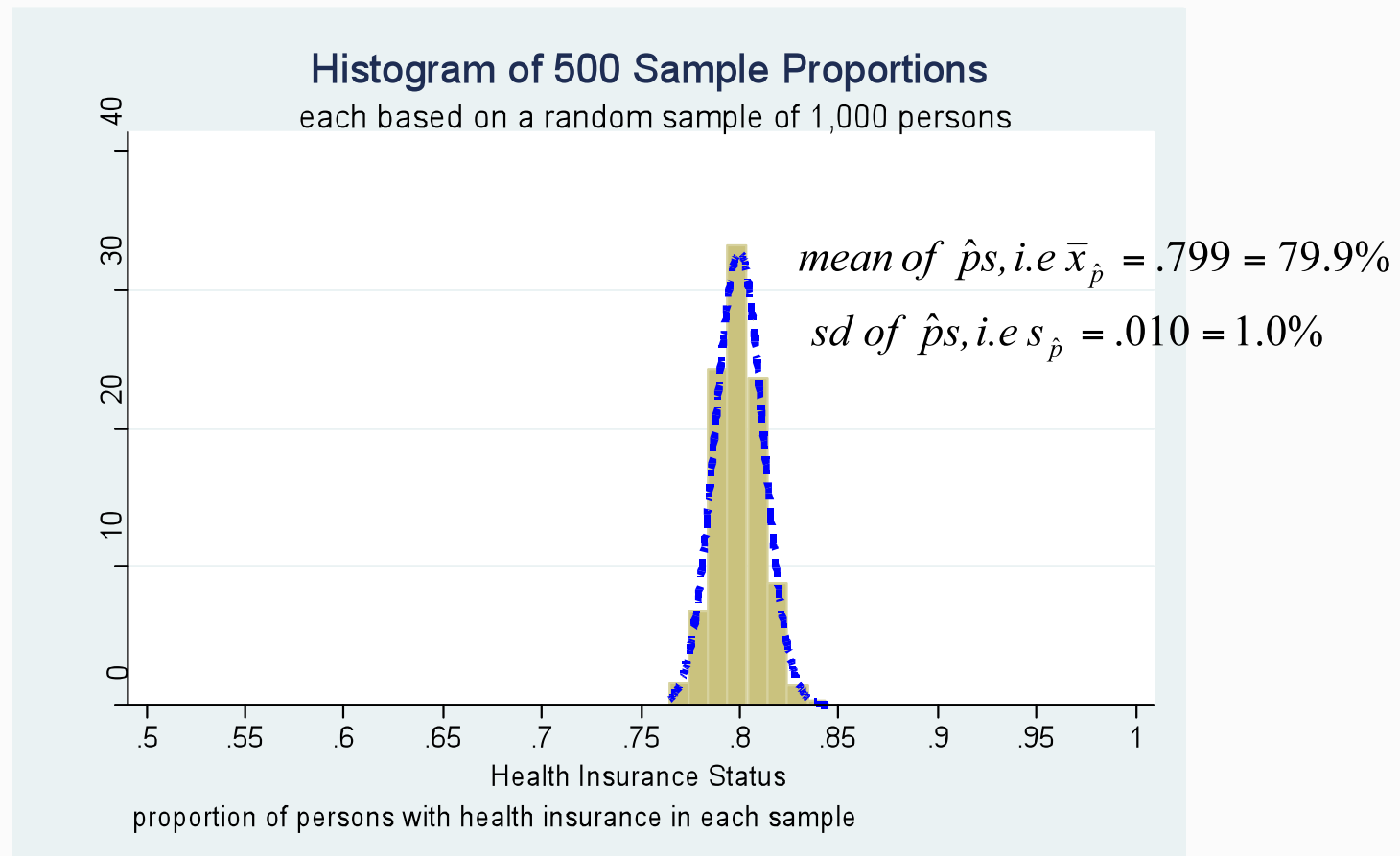
- Sample 2:  $n = 100$



$$\hat{p}_{ins} = 0.777 = 77.7\%$$

# Example: Blood Pressure of Males

- So we did this 500 times: now let's look at a histogram of the 500 proportions



## Example 2: Hospital Length of Stay

- Let's review the results
- True proportion of insured:  $p = 0.80$
- Results from 500 random samples:

Sample Sizes	Means of 500 Sample Proportions	SD of 500 Sample Proportions	Shape of Distribution of 500 Sample Proportions
$n = 20$	0.805	0.094	Approaching normal?
$n = 100$	0.801	0.041	Approximately normal
$n = 1,000$	0.799	0.012	Approximately normal

## Example 2: Hospital Length of Stay

- Let's review the results

