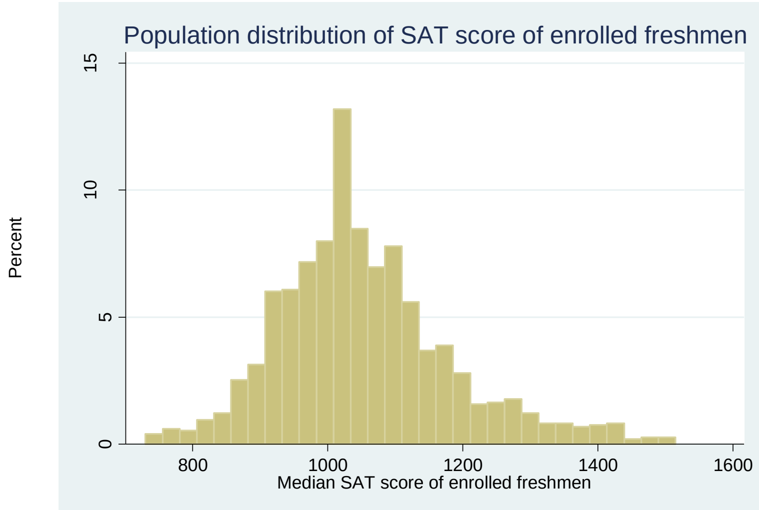
Lecture 6 notes - sampling distribution, estimation

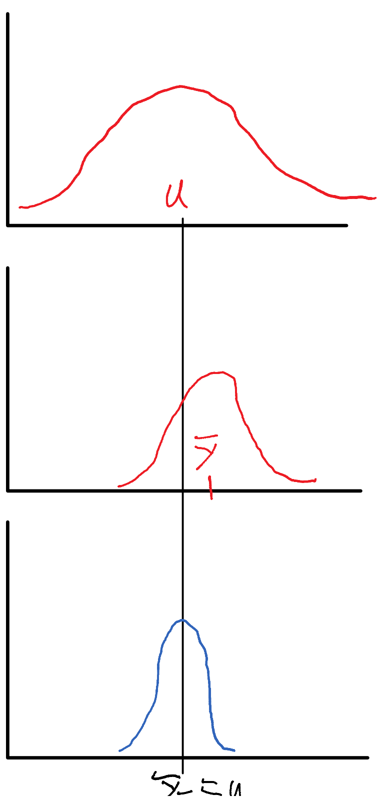
Saturday, September 24, 2011
9:54 AM




```
. sum satp50

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      satp50 |      1463     1053.66    129.1057        730       1515
```

Population → Sample → Sampling Distribution



Population distribution (we don't see this)

Population mean u

Sample data (we see this).

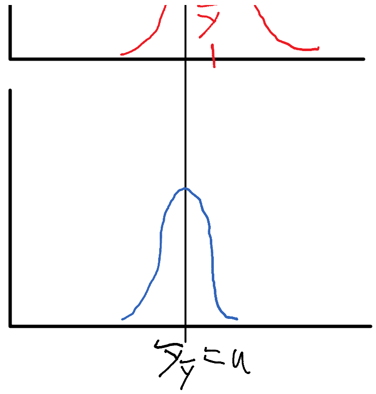
We usually only have one sample

Sample mean $\bar{y}$

Distribution of sample means (we don't see this)

- Each observation in histogram is a sample mean, y
- Mean of sample means, y_y = population mean, u

Sample mean $\bar{y}$

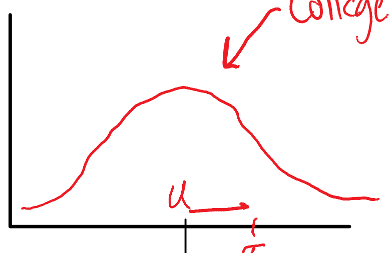


Distribution of sample means (we don't see this)

- Each observation in histogram is a sample mean, y
- Mean of sample means, y_y = population mean, u

Standard deviation → Standard error

College tuition



Population distribution

$Population\ Standard\ Deviation, \sigma$

Average distance of an obs from the mean

$u = 20{,}000$

Sample of colleges

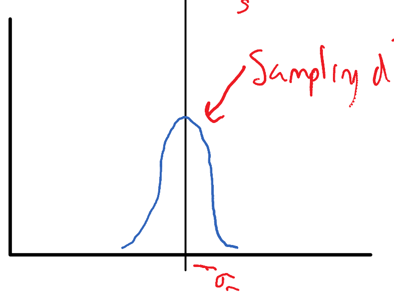
Sample data

$Sample\ standard\ deviation, s$

Average distance of an obs from the sample mean

$\bar{y} = 21{,}000$

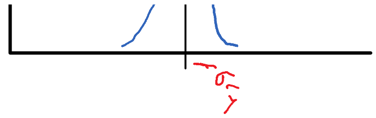
Sampling dist



Distribution of sample means

- Standard error, σ_y, is the standard deviation of the sampling distribution.
  - Average distance of a single sample mean from the mean of the sample means

$\bar{y}_{\bar{y}} = u = 20{,}000$



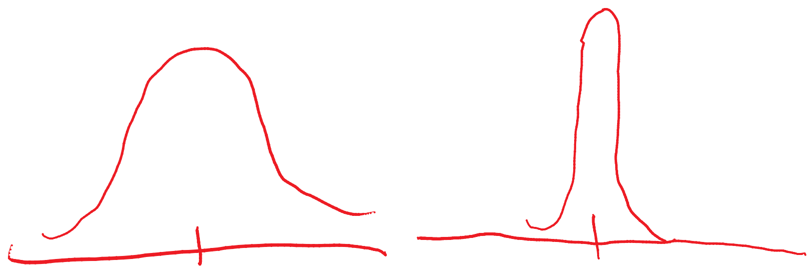
sample means

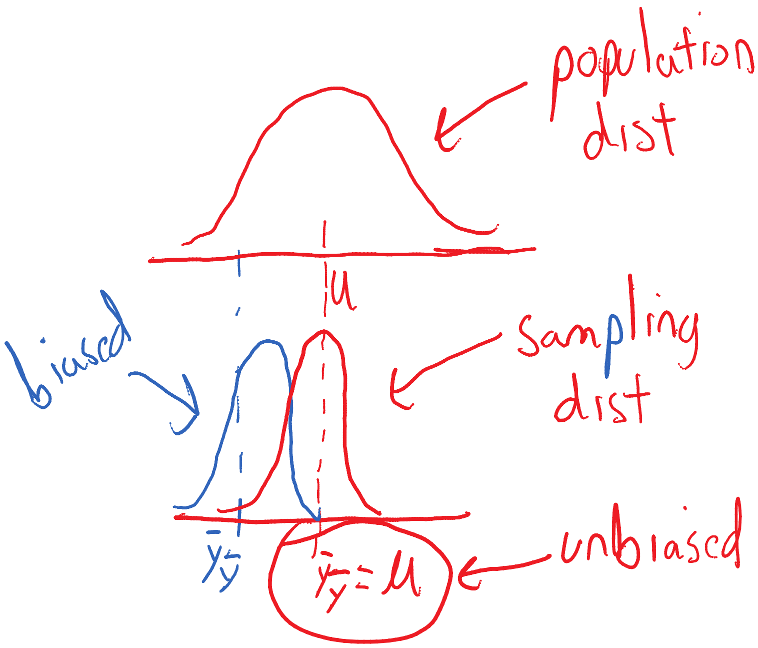
$\bar{y}_{\bar{y}} = \mu = 20{,}000$

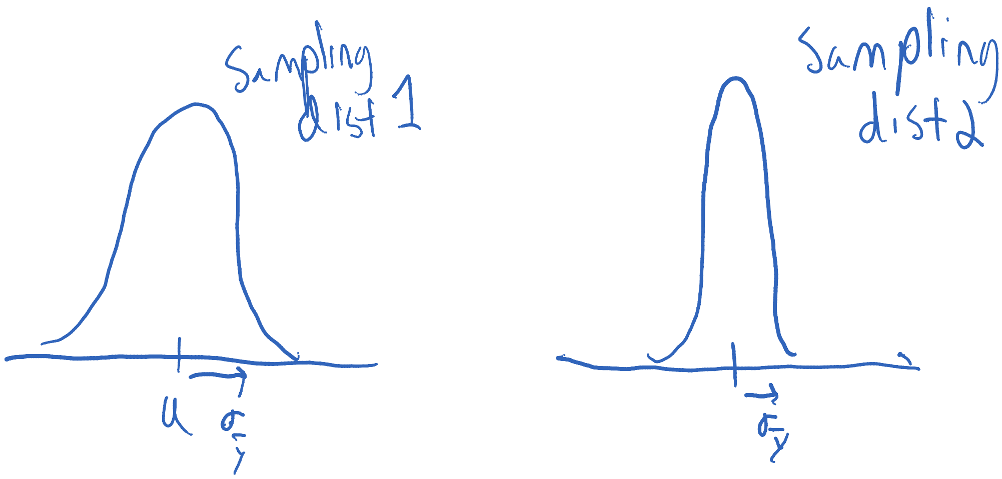
$\sigma_{\bar{y}}$ = average distance between the population mean and the sample mean trition

For one random sample

Chapter 5



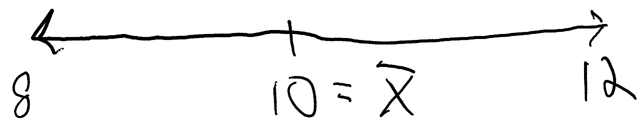




CI: $\bar{X} \pm 2 \cdot \frac{\text{std dev}}{\sqrt{n}}$

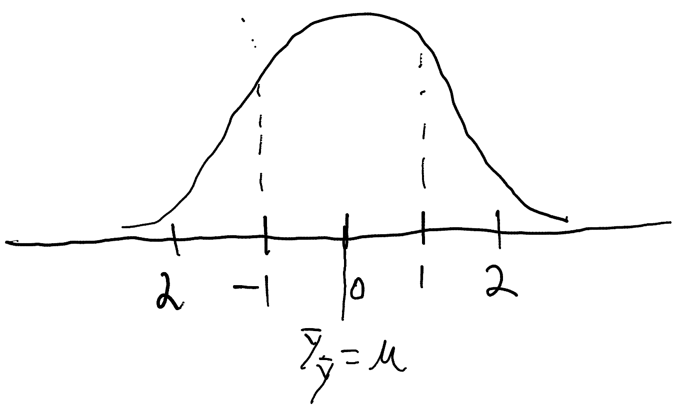
Confidence interval for population mean

How many hours per week do you spend on internet?



We are 95% sure that the population mean number of hours spent on the internet is between 8 and 12

Sampling distribution



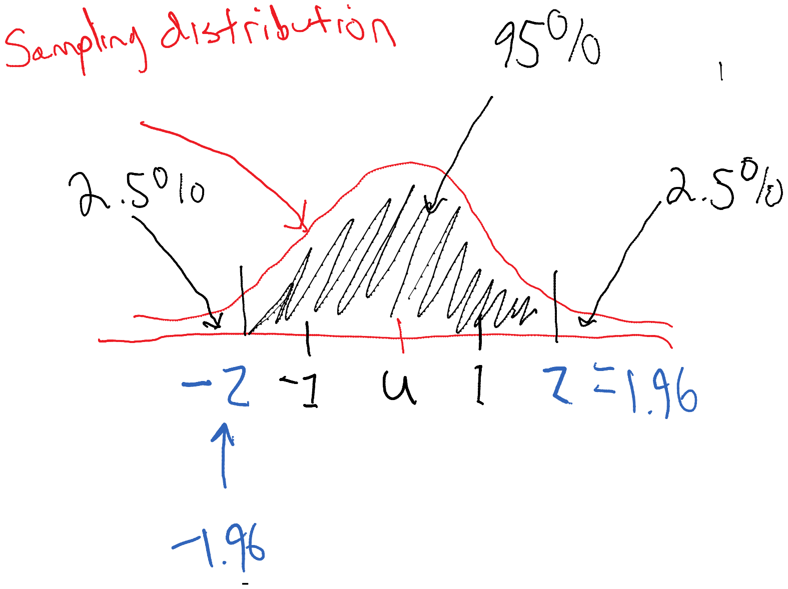
Z = number of standard deviations from population mean

What percentage of observations fall within one standard deviation of the population mean?
What percentage of sample means fall within one standard deviation of the population mean?
What percentage of sample means fall within one standard error of the population mean?
What percentage of sample means fall within two standard errors of the population mean

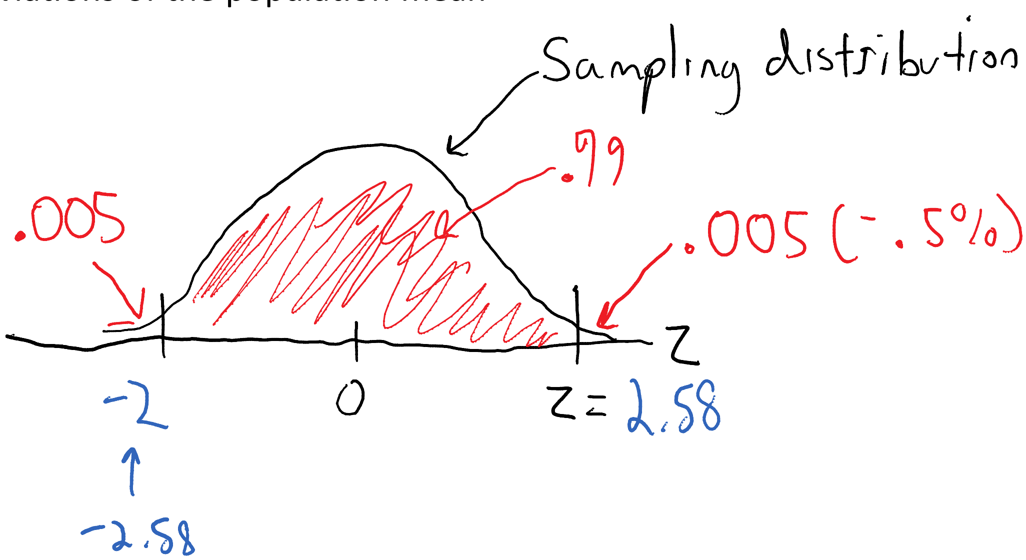
Z-scores and confidence intervals

Ask this question: What is the z-score such that 95% of observations will be within z-standard deviations of the mean?

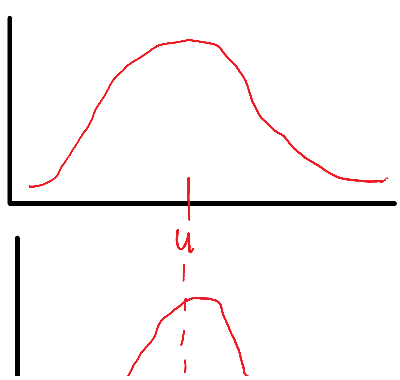
Look at Z score table



What is the Z-score such that 99% of observations would be within Z standard deviations of the population mean

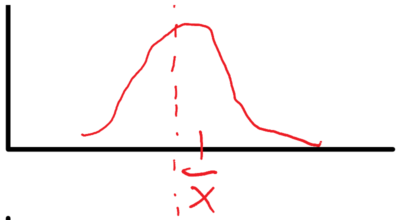


What we see and what we don't see

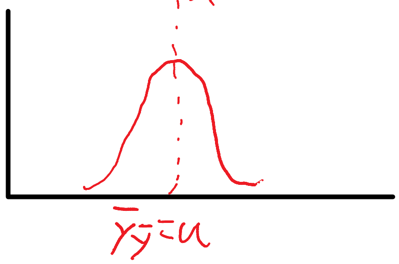


Population Distribution
(We don't see this)

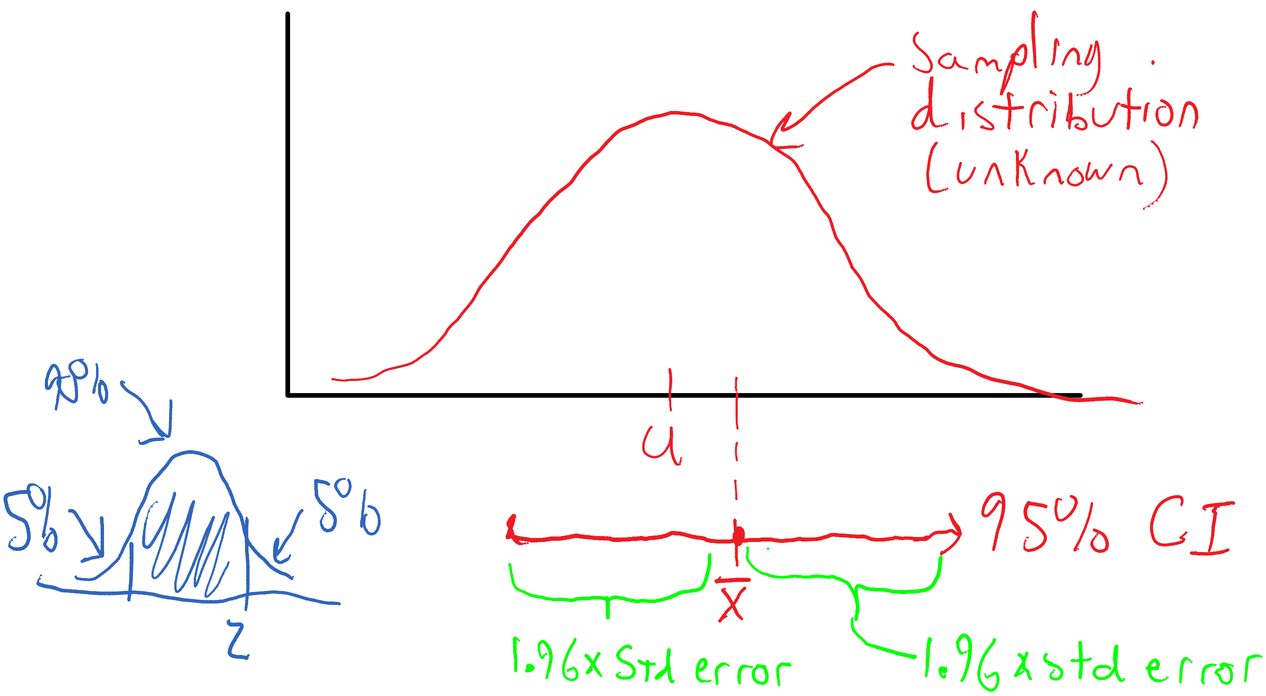
Sample distribution
(We see this for one sample)



(We see this for one sample)



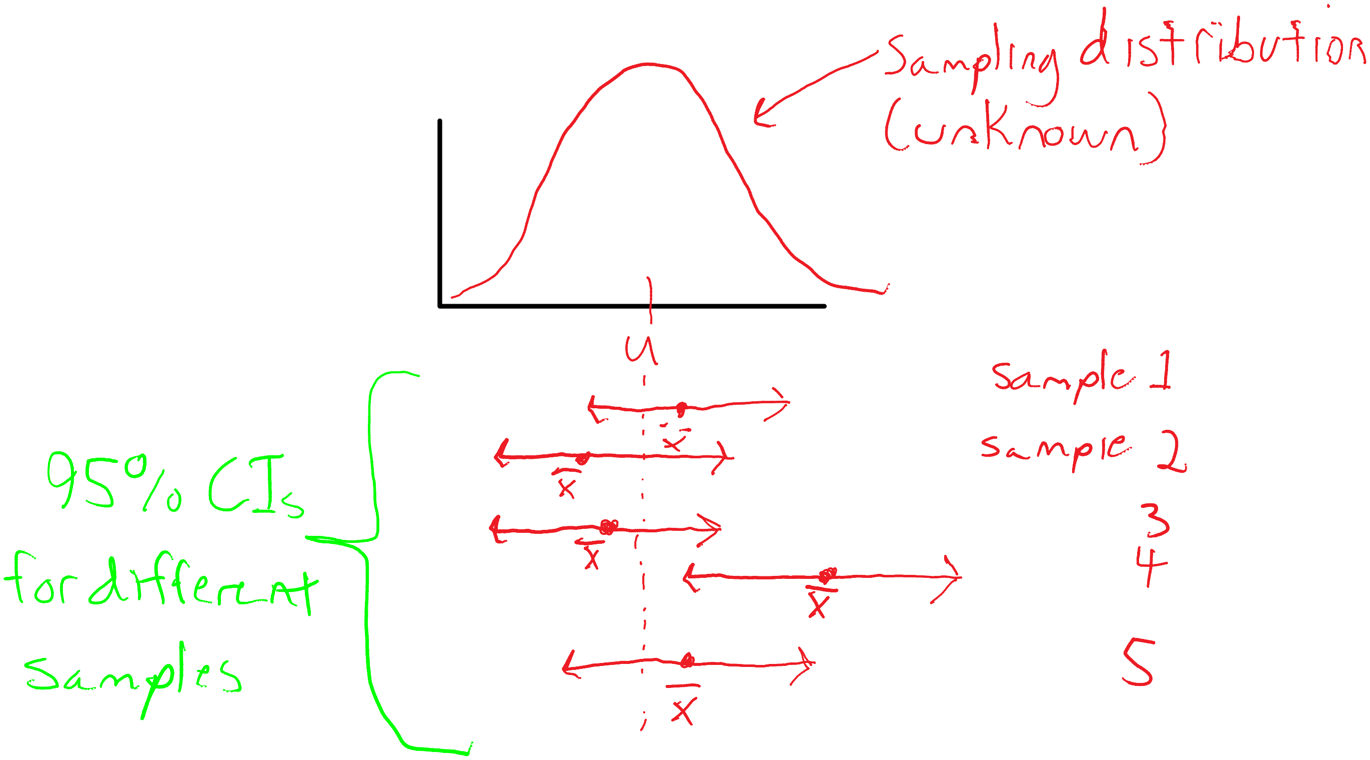
Sampling distribution
(We don't see this)



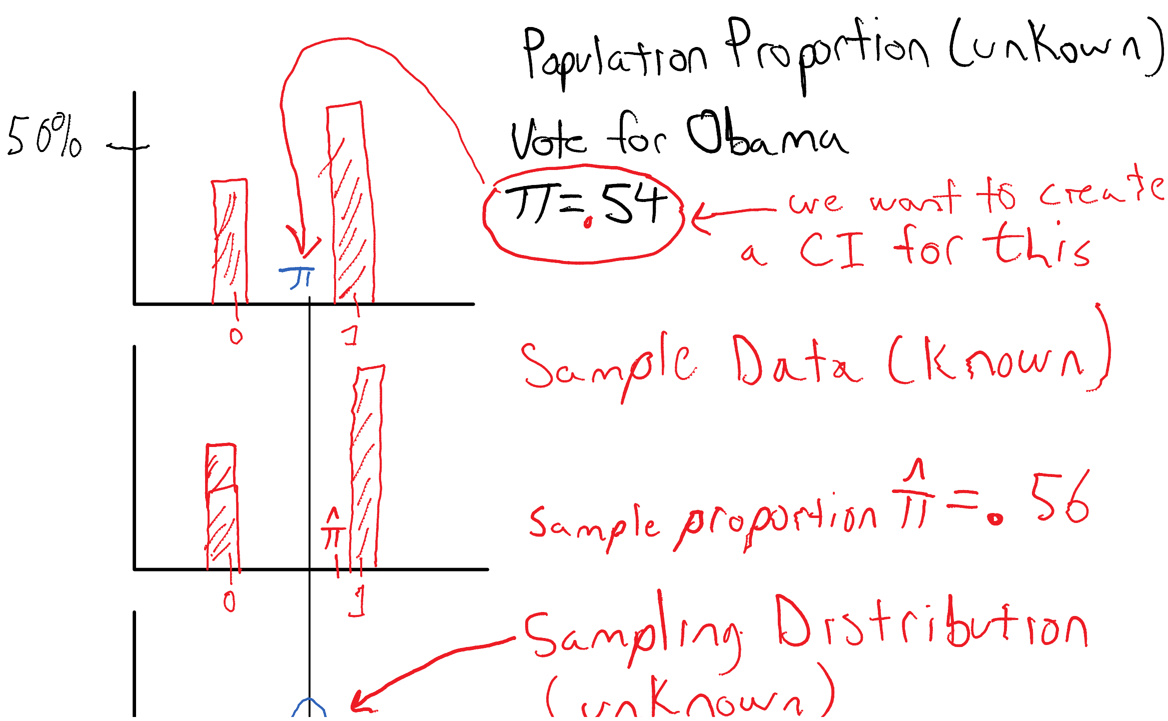
Confidence Interval = sample mean ± margin of error

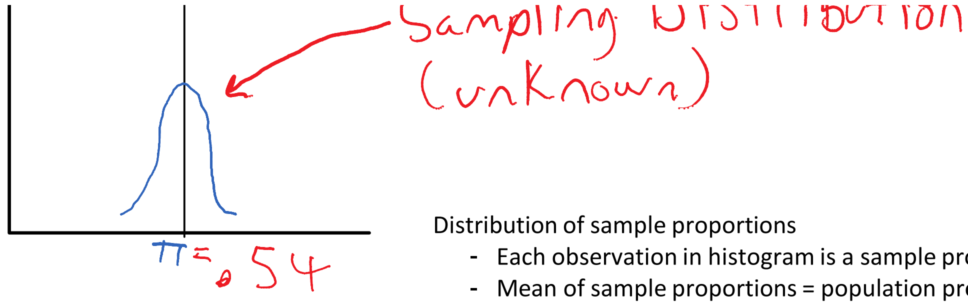
margin of error = (standard error) x (z-score associated with desired confidence level)

Confidence interval will usually contain the population mean, but sometimes it won't. (And we do not do whether it contains population mean or not because we do not know the sampling distribution)



Confidence Intervals for Proportions



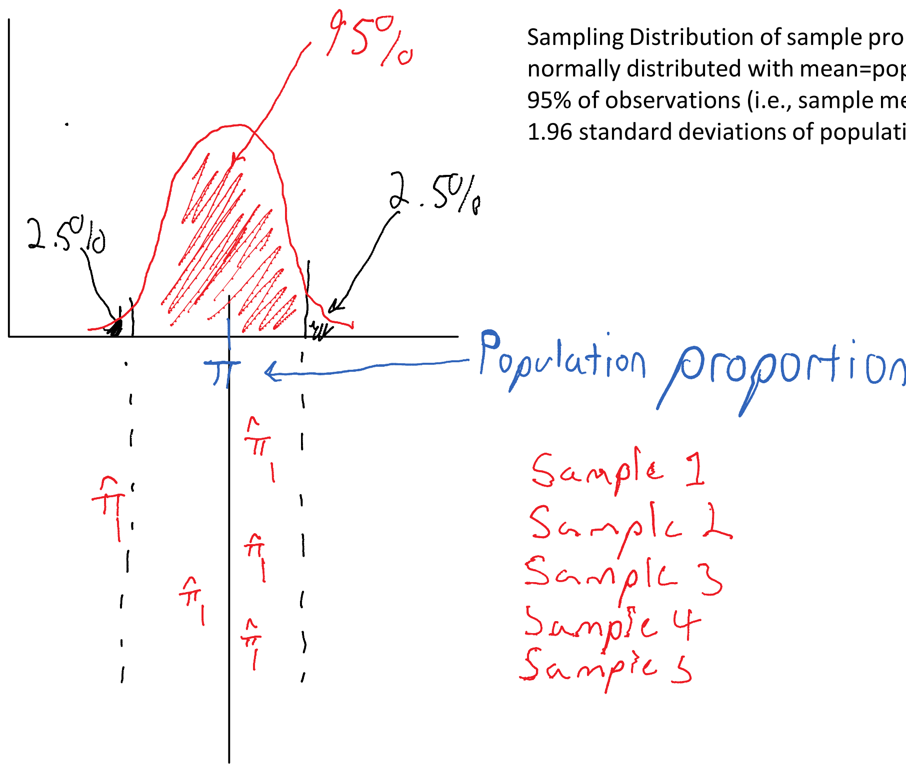


Distribution of sample proportions
- Each observation in histogram is a sample proportion, $\hat{\pi}$
- Mean of sample proportions = population proportion
- Goal: we want an estimate of the population proportion π

- 95% of sample proportions (from the sampling distribution) will be within 1.96 standard deviations of the population proportion

Sampling Distribution of sample proportion: If we knew the sampling distribution

Sampling Distribution of sample proportion: normally distributed with mean=population mean. 95% of observations (i.e., sample means) fall within 1.96 standard deviations of population mean



- Equivalently, if we select a random sample and calculate the sample proportion, there is a 95% chance that the population proportion will be within 1.96 standard deviations of the sample proportion

Confidence intervals in green

We are 95% confident that
The true population proportion
lies somewhere in this
confidence interval.