Sampling Distribution--

Distribution of possible values of a sample statistic assuming it is from a random sample.

Sample Proportion sampling distribution–

$E(\hat{P}) = P$ $\qquad SE \equiv SD(\hat{P}) = \sqrt{\frac{p(1-p)}{n}}$

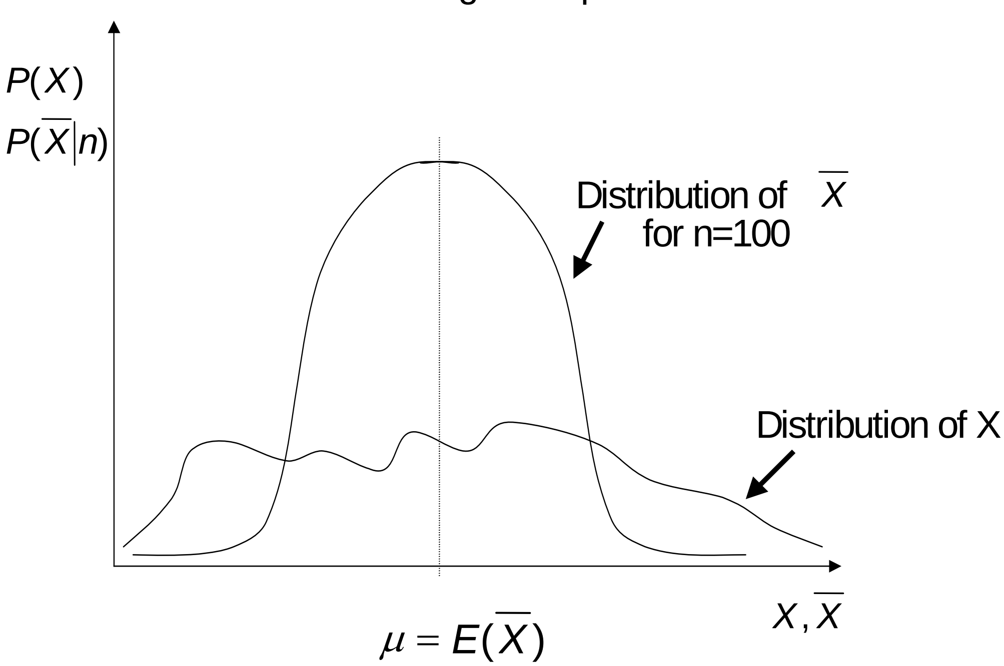
Sample Mean sampling distribution –

$E(\overline{X}) = \mu$ $\qquad SE \equiv SD(\overline{X}) = \frac{\sigma}{\sqrt{n}}$

Central Limit Theorem

- Sampling distribution of the sample mean is approximately normal *regardless* of the original variable distribution for large samples.
- Sampling distribution gets "more normal" as sample size increases. T

The $\overline{X}$ distribution with a large sample size.



Making Inference and Confidence Intervals

How can we make assertions about the unknown population distribution of X?

Scientific Method

Theory ⇒	**Hypothesis** ⇒	**Empirical Evidence**
(model of how the world works)	(specific statement about the world if theory is correct)	(test of the hypothesis using sample)

Statistical Inference

Uses information from samples to test hypotheses about populations. Empirical evidence often comes from samples.

Tools for Samples

Point estimate

Best guess of a population parameter based upon a sample.

Confidence Interval

Range estimate around point estimate

Hypothesis test

Decision rule for rejecting hypothesized population values (null hypotheses)

p value

Continuous measure of support for a hypothesis about the population (a probability)

Tools work for :

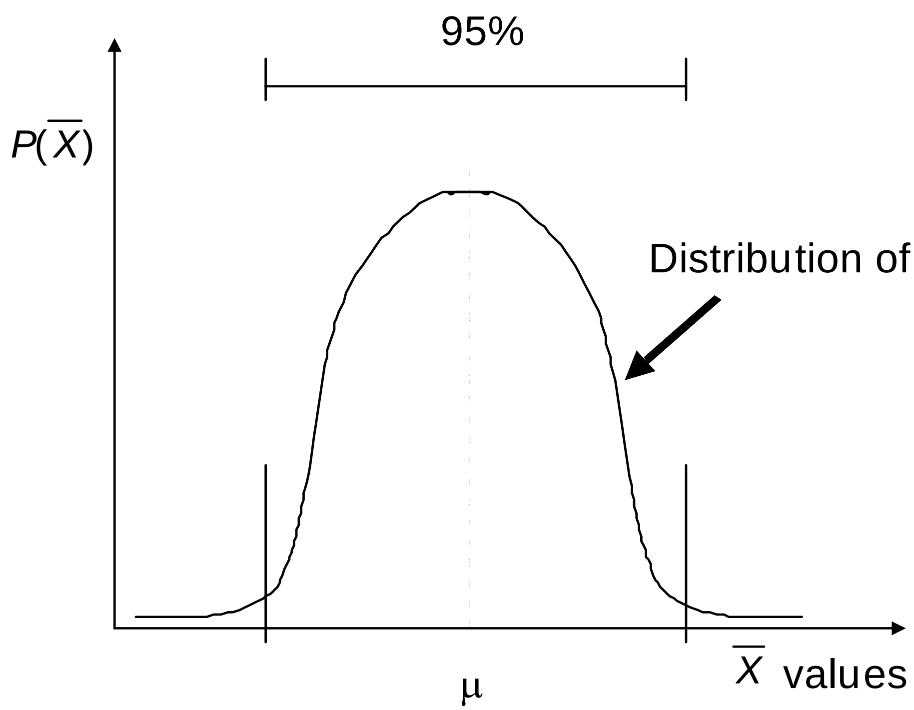
- **One mean or proportion OR**
- **Difference between 2 means or proportions OR**
- **Estimates of other parameters (next quarter)**

Confidence Intervals

--Range estimate for a random sample that reflects a fixed probability (e.g.: 99%, 95%, or 90%) that the range will "capture" the population mean.

$$\bar{X} \pm Z^*(SE)$$

- CIs are built around the sample mean (the point estimate).
- CIs vary with sample size, variable variance, and significance level.



What is the chance that $\bar{x}$ will be within 2 SE of μ?

What is the chance that our random sample will give us a CI that will include μ?

Confidence Interval for mean from Large Sample

A recent study of smoking for a sample of **11,237** US teens showed that they faced an average price of cigarettes of **$1.88** per pack with a standard deviation of **$.22.** What range estimate could you construct that will have a 95% chance of including the (unknown) population mean?

Want 95% confidence interval (for a large sample mean):

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

1. Get standard error. $SE = \frac{S}{\sqrt{n}} = \frac{\$.22}{\sqrt{11{,}237}} = .0021$

2. Plug in Z, SE, and sample mean:

$\bar{X} \pm 1.96 \frac{S}{\sqrt{n}}$ =1.88 - (1.96)(.0021) to 1.88 + (1.96)(.0021)

= $1.876 to $1.884

So, on average, in 1996 US teens faced a cigarette price of $1.876 to $1.884.

Confidence Interval for Proportion from Large Sample

In a recent study of smoking for a sample of **11,237** US teens, 27.7 percent reported that they had smoked within the last 30 days.

What range estimate could you construct that will have a 99% chance of including the (unknown) population proportion?

Want 99% confidence interval (for a large sample proportion):

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

1. Get standard error. $SE = \sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = \sqrt{\frac{.277(1-.277)}{11{,}237}} = .0042$

2. Plug in Z, SE, and sample proportion:

$$\hat{P} \pm 2.58\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} = .277 \pm 2.58(.0042) = .266 \text{ to } .288$$

So, on average, between 27 and 29 percent of US teens smoked within the last 30 days.

Confidence Interval for <u>Small Sample</u> Sample Mean

T distribution—distribution of sample means for small samples from approximately normal population.

- Depends on degrees of freedom (n-1 for sample mean)
- For large samples is same as Z distribution

Suppose I have a sample of 100 households with a mean income of $35,000 and a standard deviation (S) of $33,000. What is the 90 percent confidence interval?

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

1) Find SE estimate:

$$SE = \frac{S}{\sqrt{n}} = \frac{33{,}000}{\sqrt{100}} = 3300$$

2) Find t score for 90 percent CI (5 percent in each tail)

   Degrees of freedom = n-1 = 100-1= 99
   ➔tc=1.67

3) Plug in t score, sample mean, and SE:

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 35{,}000 \pm 1.67(3300) = \$29{,}489 \, to \, \$40{,}511$$

So, our sample evidence suggests that the average household income is somewhere between $29,489 and $40,511.

Confidence Interval for Small Sample Proportion

Suppose that you only had a sample of 25 teens and wanted a range estimate of their life time smoking rate. In your sample 15 of the teens had ever smoked. (90 percent confidence)

$$p^* \pm z_{\alpha/2}\sqrt{\frac{p^*(1-p^*)}{n+4}} \qquad p^* = \frac{x+2}{n+4}$$

1) Get p*

$$p^* = \frac{x+2}{n+4} = \frac{15+2}{25+4} = .586$$

2) Get SE

$$SE = \sqrt{\frac{p^*(1-p^*)}{n+4}} = \sqrt{\frac{.586(1-.586)}{25+4}} = 0.091$$

3) Calculate confidence interval:

$$p^* \pm z_{\alpha/2}\sqrt{\frac{p^*(1-p^*)}{n+4}} = .586 \pm 1.65 * (.091)$$

So between 42 and 72 percent of the population of teens are estimated to have ever smoked.

Calculating the needed sample size for a margin of error

Suppose you were designing a new survey of teen smokers to better understand when and why they start smoking. You want to assess their average age at which they first smoked and be able to identify a range estimate that is no more than 2 years wide.

What sample size do you need to be sure you will get a CI 2 years wide or less?

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{B^2}$$

1) **Choose B (the margin of error), Z, and estimate** σ^2

Let's choose the margin of error: B= 1 (to get a CI two years wide),

A 95 percent significant level Z=1. 96,

and let's estimate $\sigma^2 = 9$

2) **Calculate needed sample size**

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{B^2} = \frac{(1.96)^2 9}{1^2} = 34.6$$

So we need a sample of at least 35 smokers to get a CI less than 2 years wide.

What if S=5 instead of 3?