Sampling Distribution and Interval Estimation Of Population Parameters

Sampling Distribution of $\bar{X}$

$\bar{X}$

A natural estimator for the population mean μ is the sample mean

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}.$$

Consider $\overline{x}$ to be a single realization of a random variable $\overline{X}$ over all possible samples of size n.

The sampling distribution of $\overline{X}$ is the distribution of values of $\overline{x}$ over all possible samples of size n that could be selected from the population.

Sampling Distribution of $\overline{X}$

The average of the sample means ($\overline{x}$'s) when taken over a large number of random samples of size n will approximate μ.

Let $X_1, \cdots, X_n$ be a random sample from some population with mean μ. Then for the sample mean $\overline{X}$, $E(\overline{X}) = \mu$.

$\overline{X}$ is an unbiased estimator of μ.

Variance

Let $X_1, \cdots, X_n$ be a random sample from some population with mean μ. and variance σ^2.

The variance of the sample mean $\overline{X}$ is given by

$$Var(\overline{X}) = \sigma^2/n.$$

The standard deviation of the sample mean is given by $\sigma/\sqrt{n}$. This quantity is called the standard error (of the mean).

The standard error $\sigma/\sqrt{n}$ is estimated by $s/\sqrt{n}$.

The standard error measures the variability of sample means from repeated samples of size n drawn from the same population.

A larger sample provides a more precise estimate $\overline{X}$ of μ

Let $X_1, \cdots, X_n$ be a random sample from a population that is normally distributed with mean μ and variance σ^2.

Then the sample mean $\overline{X}$ is normally distributed with mean μ and variance σ^2/n.

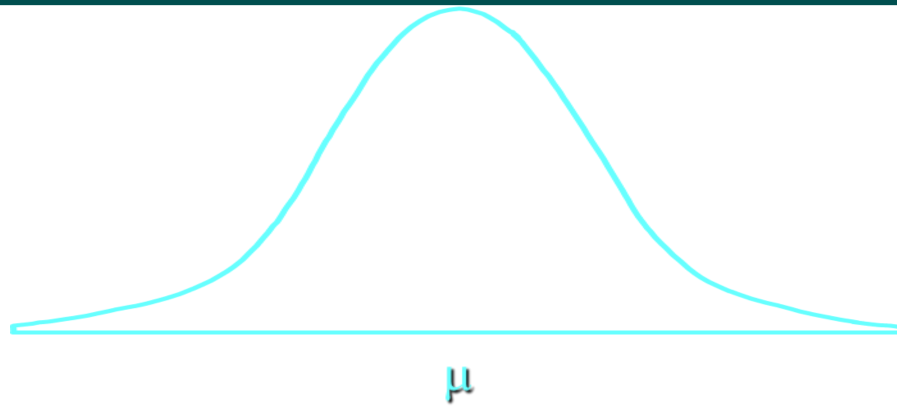
That is

$$\overline{X} \sim N(\mu, \sigma^2/n).$$

The Sampling Distribution of the Sample Mean

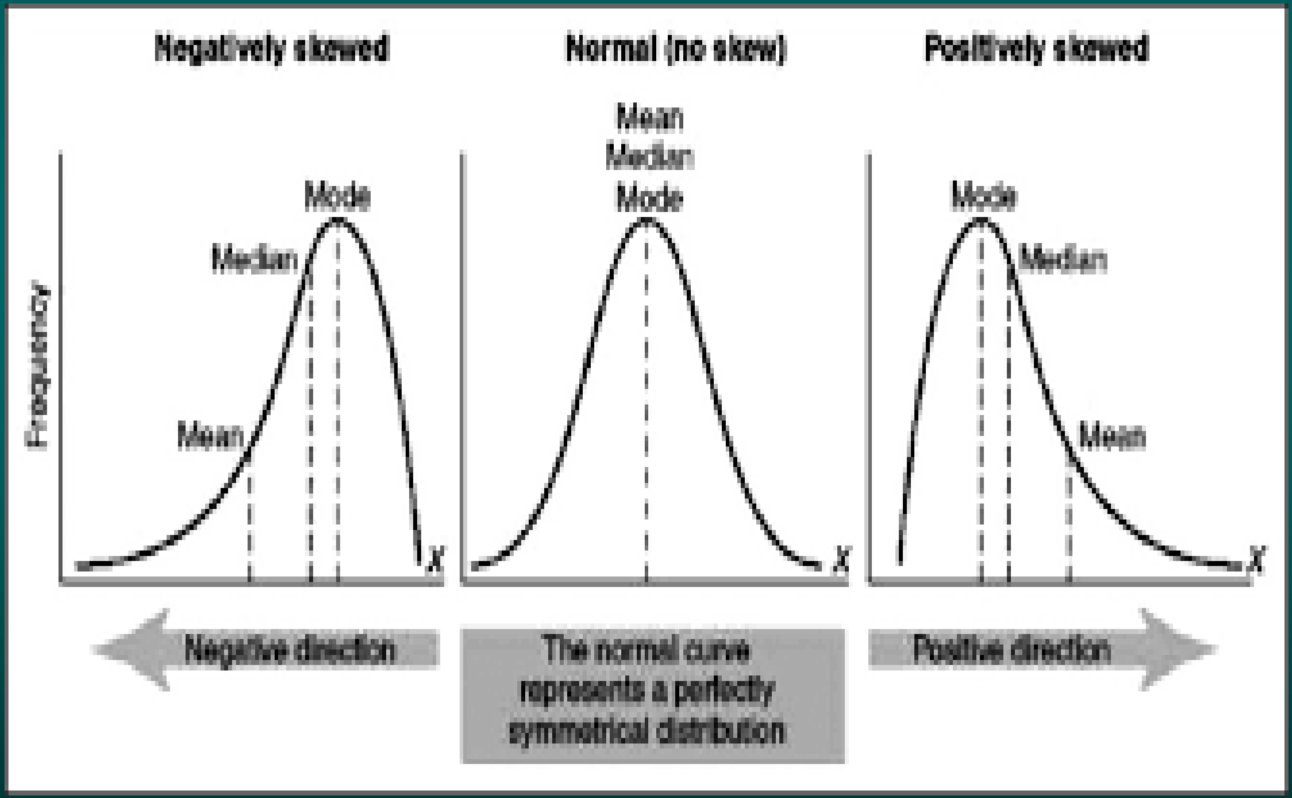
1. $\mu_{\bar{x}} = \mu_x$

2. $\sigma^2_{\bar{x}} = \dfrac{\sigma^2_x}{n}$



3. *If x is* normal, $\bar{x}$ is normal. If x is nonnormal $\bar{x}$ is approximately normally distributed for sufficiently large sample size.

- The normal distribution is a continuous, symmetric, bell-shaped distribution of a variable.
- Any particular normal distribution is determined by two parameters 1) Mean, μ 2) Standard Deviation, σ



Central Limit Theorem

CENTRAL LIMIT THEOREM If all samples of a particular size are selected from any population, the sampling distribution of the sample mean is approximately a normal distribution. This approximation improves with larger samples.

Central Limit Theorem

Let $X_1, \cdots, X_n$ be a random sample from any population with mean μ and variance σ^2.

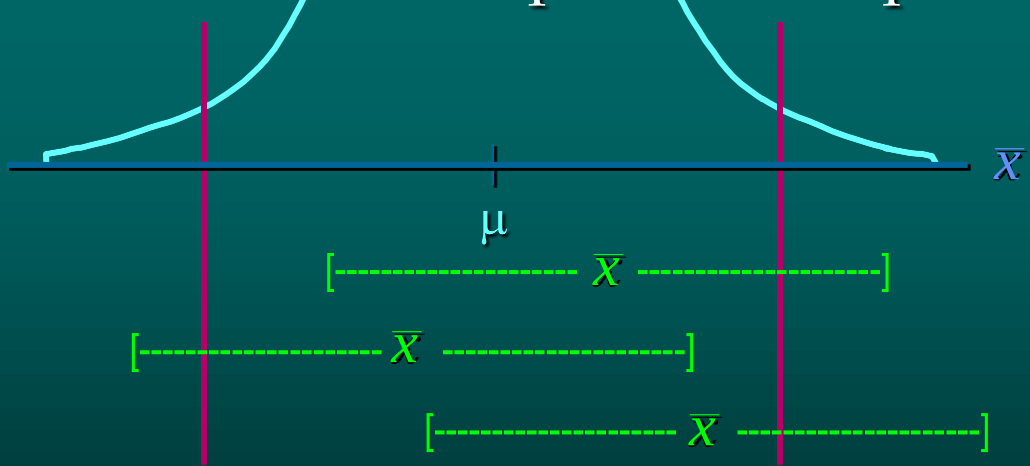
Then the sample mean $\overline{X}$ is approximately normally distributed with mean μ and variance σ^2/n.

Concepts of Estimation

- In general we are interested in determining the value of a population parameter (e.g. μ) on the basis of a sample statistic. We use:
  - Point estimator (estimates the parameter by a single value, a statistic).
  - Interval estimator (estimates the parameter by a range of values).

Interval Estimation and Determination of the Sample Size

- Interval Estimation of a Population Mean:
  Large-Sample Case
- Interval Estimation of a Population Mean:
  Small-Sample Case
- Determining the Sample Size
- Interval Estimation of a Population Proportion



Interval Estimation of a Population Mean: Large-Sample Case

- Sampling Error
- Probability Statements about the Sampling Error
- Constructing an Interval Estimate:
  Large-Sample Case with σ Known
- Calculating an Interval Estimate:
  Large-Sample Case with σ Unknown

Sampling Error

- The absolute value of the difference between an unbiased point estimate and the population parameter is called the sampling error.
- For the case of a sample mean estimating a population mean, the sampling error is

$$\text{Sampling Error} = |\bar{x} - \mu|$$

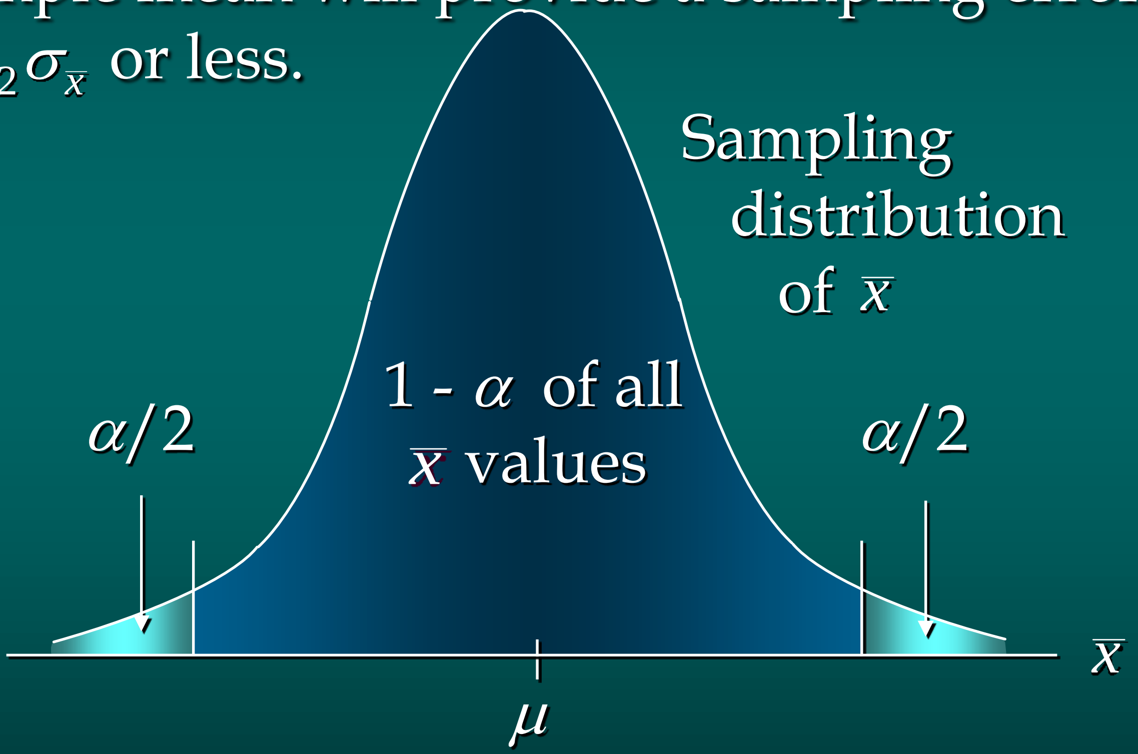
Probability Statements About the Sampling Error

- Knowledge of the sampling distribution of $\bar{x}$ enables us to make probability statements about the sampling error even though the population mean μ is not known.
- A probability statement about the sampling error is a precision statement.

Probability Statements About the Sampling Error

- Precision Statement

  There is a $1 - \alpha$ probability that the value of a sample mean will provide a sampling error of $z_{\alpha/2}\sigma_{\bar{x}}$ or less.



Interval Estimate of a Population Mean: Large-Sample Case ($n \geq 30$)

- With σ Known

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

where: $\bar{x}$ is the sample mean

$1 - \alpha$ is the confidence coefficient

$z_{\alpha/2}$ is the z value providing an area of $\alpha/2$ in the upper tail of the standard normal probability distribution

σ is the population standard deviation

n is the sample size

Interval Estimate of a Population Mean: Large-Sample Case ($n \geq 30$)

- With σ Unknown

  In most applications the value of the population standard deviation is unknown. We simply use the value of the sample standard deviation, *s*, as the point estimate of the population standard deviation.

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

Example: National Discount, Inc.

National Discount has 260 retail outlets throughout the United States. National evaluates each potential location for a new retail outlet in part on the mean annual income of the individuals in the marketing area of the new location.

Sampling can be used to develop an interval estimate of the mean annual income for individuals in a potential marketing area for National Discount.

A sample of size n = 36 was taken. The sample mean, $\bar{x}$, is $21,100 and the sample standard deviation, s, is $4,500. We will use .95 as the confidence coefficient in our interval estimate.

Example: National Discount, Inc.

- Precision Statement

  There is a .95 probability that the value of a sample mean for National Discount will provide a sampling error of $1,470 or less……. determined as follows:

  95% of the sample means that can be observed are within $\pm$ 1.96 $\sigma_{\bar{x}}$ of the population mean μ.

  If $\sigma_{\bar{x}} = s/\sqrt{n} = 4{,}500/\sqrt{36} = 750$, then 1.96 $\sigma_{\bar{x}}$ = 1,470.

Example: National Discount, Inc.

- Interval Estimate of the Population Mean: σ Unknown

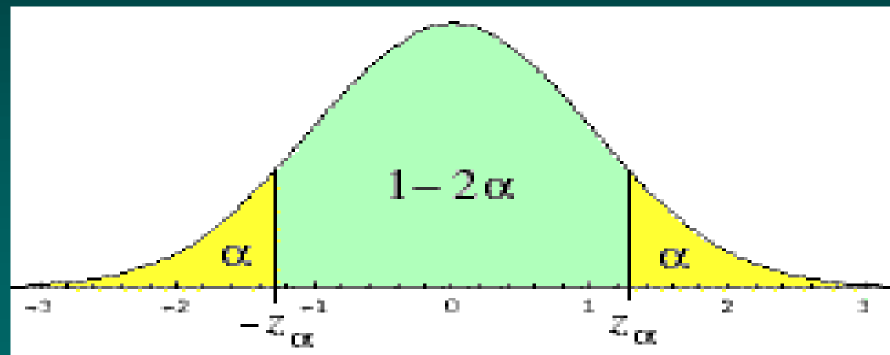
Interval Estimate of μ is:

$$\$21,100 \pm \$1,470$$

or $\$19,630$ to $\$22,570$

We are 95% confident that the interval contains the population mean.

The most common five critical values of z_α are summarized in the following table.



α = tail area	central area = 1 – 2α	z_α
0.05	0.90	$z_{.05} = 1.645$
0.025	0.95	$z_{.025} = 1.96$
0.01	0.98	$z_{.01} = 2.33$
0.005	0.99	$z_{.005} = 2.58$

Interval Estimation of a Population Mean: Small-Sample Case (n < 30)

- Population is Not Normally Distributed
  The only option is to increase the sample size to $n \geq 30$ and use the large-sample interval-estimation procedures.
- Population is Normally Distributed and σ is Known
  The large-sample interval-estimation procedure can be used.
- Population is Normally Distributed and σ is Unknown
  The appropriate interval estimate is based on a probability distribution known as the *t* distribution.

t Distribution

- The *t* distribution is a family of similar probability distributions.
- A specific *t* distribution depends on a parameter known as the degrees of freedom.
- As the number of degrees of freedom increases, the difference between the *t* distribution and the standard normal probability distribution becomes smaller and smaller.
- A *t* distribution with more degrees of freedom has less dispersion.
- The mean of the *t* distribution is zero.

Interval Estimation of a Population Mean: Small-Sample Case (n < 30) with σ Unknown

- Interval Estimate

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

where $1 - \alpha$ = the confidence coefficient

$t_{\alpha/2}$ = the t value providing an area of $\alpha/2$ in the upper tail of a t distribution with $n - 1$ degrees of freedom

s = the sample standard deviation

Example: Apartment Rents

- Interval Estimation of a Population Mean:
  Small-Sample Case ($n < 30$) with σ Unknown

  A reporter for a student newspaper is writing an article on the cost of off-campus housing. A sample of 10 one-bedroom units within a half-mile of campus resulted in a sample mean of $550 per month and a sample standard deviation of $60.

  Let us provide a 95% confidence interval estimate of the mean rent per month for the population of one-bedroom units within a half-mile of campus. We'll assume this population to be normally distributed.

Example: Apartment Rents

- *t* Value

  At 95% confidence, $1 - \alpha = .95$, $\alpha = .05$, and $\alpha/2 = .025$.

  $t_{.025}$ is based on $n - 1 = 10 - 1 = 9$ degrees of freedom.

  In the *t* distribution table we see that $t_{.025} = 2.262$.

Degrees	**Area in Upper Tail**				
of Freedom	**.10**	**.05**	**.025**	**.01**	**.005**
.	.	.	.	.	.
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
.	.	.	.	.	.

Student's T Critical Values, t_α and $t_{\alpha/2}$

Conf. Level	50%	80%	90%	95%	98%	99%
One Tail(α)	0.250	0.100	0.050	0.025	0.010	0.005
Two Tail($\alpha/2$)	0.500	0.200	0.100	0.050	0.020	0.010
df = 1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.695	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861

Conf. Level	50%	80%	90%	95%	98%	99%
One Tail(α)	0.250	0.100	0.050	0.025	0.010	0.005
Two Tail($\alpha/2$)	0.500	0.200	0.100	0.050	0.020	0.010
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.319	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
40	0.681	1.303	1.684	2.021	2.423	2.704
50	0.679	1.299	1.676	2.009	2.403	2.678
60	0.679	1.296	1.671	2.000	2.390	2.660
70	0.678	1.294	1.667	1.994	2.381	2.648
80	0.678	1.292	1.664	1.990	2.374	2.639
90	0.677	1.291	1.662	1.987	2.368	2.632
100	0.677	1.290	1.660	1.984	2.364	2.626
z	0.674	1.282	1.645	1.960	2.326	2.576

Example: Apartment Rents

- Interval Estimation of a Population Mean: Small-Sample Case (n < 30) with σ Unknown

$$\bar{x} \pm t_{.025} \frac{s}{\sqrt{n}}$$

$$550 \pm 2.262 \frac{60}{\sqrt{10}}$$

$$550 \pm 42.92$$

or $507.08 to $592.92

We are 95% confident that the mean rent per month for the population of one-bedroom units within a half-mile of campus is between $507.08 and $592.92.

Sample Size for an Interval Estimate of a Population Mean

- Let E = the maximum sampling error mentioned in the precision statement.
- E is the amount added to and subtracted from the point estimate to obtain an interval estimate.
- E is often referred to as the margin of error.
- We have

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Solving for n we have

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

Example: National Discount, Inc.

- Sample Size for an Interval Estimate of a Population Mean

  Suppose that National's management team wants an estimate of the population mean such that there is a .95 probability that the sampling error is $500 or less.

  How large a sample size is needed to meet the required precision?

Example: National Discount, Inc.

- Sample Size for Interval Estimate of a Population Mean

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 500$$

At 95% confidence, $z_{.025} = 1.96$.

Recall that $\sigma = 4{,}500$.

Solving for n we have

$$n = \frac{(1.96)^2(4,500)^2}{(500)^2} = 311.17$$

We need to sample 312 to reach a desired precision of $\pm$ \$500 at 95% confidence.

Example: Estimation Using SPSS

A survey by Accountemps asked a sample of 200 executives to provide data on the number of minutes per day office workers waste trying to locate mislabeled, misfiled, or misplaced items. Data consistent with this survey are contained in the data file ActTemps.

a. Use ActTemps to develop a point estimate of the number of minutes per day office workers waste trying to locate mislabeled, misfiled, or misplaced items.
b. What is the sample standard deviation?
c. What is the 95% confidence interval for the mean number of minutes wasted per day?

Interval Estimation of a Population Proportion

- Interval Estimate

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

where: $1 - \alpha$ is the confidence coefficient

$z_{\alpha/2}$ is the z value providing an area of $\alpha/2$ in the upper tail of the standard normal probability distribution

$\bar{p}$ is the sample proportion

Example: Political Science, Inc.

- Interval Estimation of a Population Proportion

  Political Science, Inc. (PSI) specializes in voter polls and surveys designed to keep political office seekers informed of their position in a race. Using telephone surveys, interviewers ask registered voters who they would vote for if the election were held that day.

  In a recent election campaign, PSI found that 220 registered voters, out of 500 contacted, favored a particular candidate. PSI wants to develop a 95% confidence interval estimate for the proportion of the population of registered voters that favors the candidate.

Example: Political Science, Inc.

- Interval Estimate of a Population Proportion

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

where: $n = 500$, $\hat{p} = 220/500 = .44$, $z_{\alpha/2} = 1.96$

$$.44 \pm 1.96\sqrt{\frac{.44(1-.44)}{500}}$$

$$.44 + .0435$$

PSI is 95% confident that the proportion of all voters that favors the candidate is between .3965 and .4835.

Sample Size for an Interval Estimate of a Population Proportion

- Let E = the maximum sampling error mentioned in the precision statement.
- We have

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

- Solving for n we have

$$n = \frac{(z_{\alpha/2})^2 p(1-p)}{E^2}$$

Example: Political Science, Inc.

- Sample Size for an Interval Estimate of a Population Proportion

  Suppose that PSI would like a .99 probability that the sample proportion is within + .03 of the population proportion.

  How large a sample size is needed to meet the required precision?

Example: Political Science, Inc.

- Sample Size for Interval Estimate of a Population Proportion

At 99% confidence, $z_{.005} = 2.576$.

$$n = \frac{(z_{\alpha/2})^2 p(1-p)}{E^2} = \frac{(2.576)^2(.44)(.56)}{(.03)^2} \cong 1817$$

Note: We used .44 as the best estimate of *p* in the above expression. If no information is available about *p*, then .5 is often assumed because it provides the highest possible sample size. If we had used *p* = .5, the recommended *n* would have been 1843.

For Practice see book (Chapter 8)

- Chapter 8