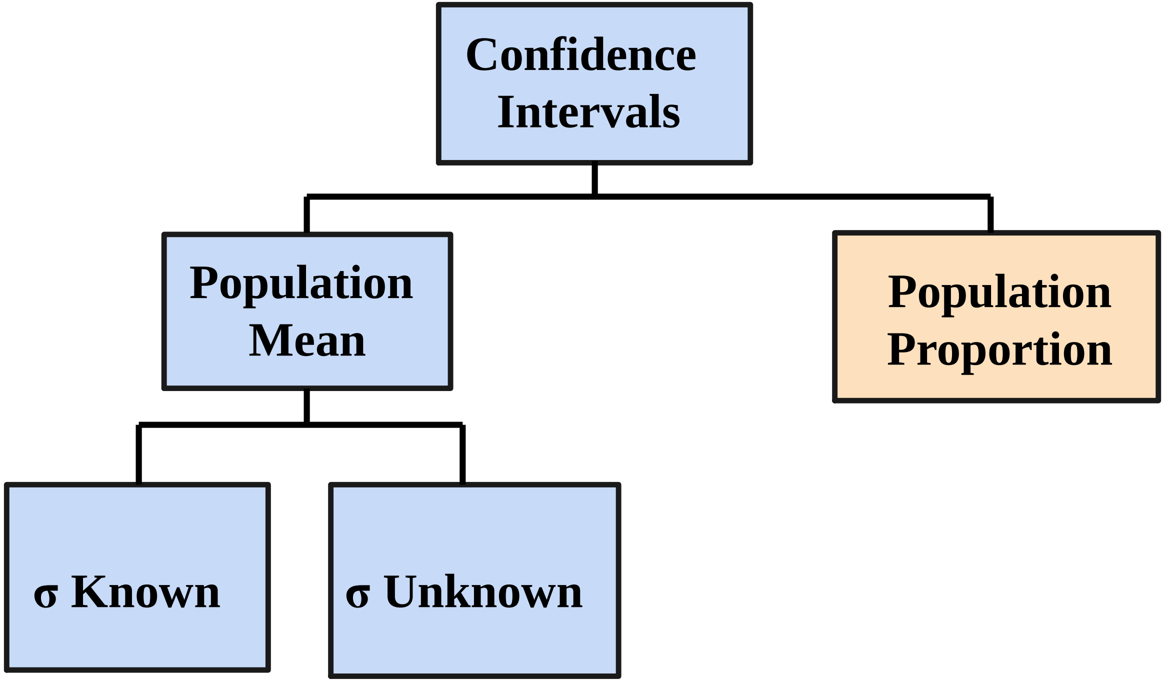
Confidence Intervals



Confidence Intervals for the **Population Proportion, π**

- An interval estimate for the population proportion (π) can be calculated by adding an allowance for uncertainty to the sample proportion (p)

Confidence Intervals for the Population Proportion, π

(continued)

- Recall that the **distribution of the sample proportion is approximately normal if the sample size is large**, with standard deviation

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

- We will **estimate** this with **sample d**ata:

$$\sqrt{\frac{p(1-p)}{n}}$$

Confidence Interval Endpoints

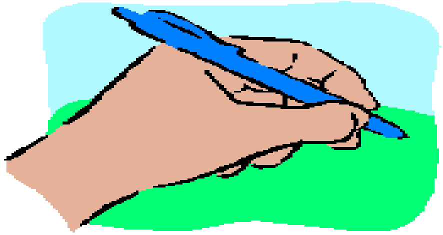
- Upper and lower confidence limits for the population proportion are calculated with the formula

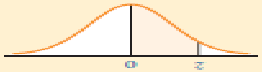
$$\boxed{p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}}$$

- where
  - $Z_{\alpha/2}$ is the standard normal value for the level of confidence desired
  - p is the sample proportion
  - n is the sample size
- Note: must have $np > 5$ and $n(1-p) > 5$

Example

- A random sample of 100 people shows that 25 are left-handed.
- Form a 95% confidence interval for the true proportion of left-handers





SECOND DECIMAL PLACE IN z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998									
4.0	.49997									
4.5	.499997									
5.0	.4999997									
6.0	.499999999									

Example

(continued)

- A random sample of 100 people shows that 25 are left-handed. Form a 95% confidence interval for the true proportion of left-handers.

$$p \pm Z_{\alpha/2}\sqrt{p(1-p)/n}$$

$$= 25/100 \pm 1.96\sqrt{0.25(0.75)/100}$$

$$= 0.25 \pm 1.96(0.0433)$$

$$\boxed{= 0.1651 \le \pi \le 0.3349}$$

Interpretation

- We are 95% confident that the true percentage of left-handers in the population is between 16.51% and 33.49%.

- Although the interval from 0.1651 to 0.3349 may or may not contain the true proportion, 95% of intervals formed from samples of size 100 in this manner will contain the true proportion.

Dr. Benjamin Shockley, a noted social psychologist, surveyed 150 top executives and found that 42 percent of them were unable to add fractions correctly.

a) Estimate the standard error of the proportion.
b) Construct a 99 percent confidence interval for the true proportion of top executives who cannot correctly add fractions.

Solution:

n = 150 $\bar{p} = 0.42$

a) $\hat{\sigma}_{\bar{p}} = \sqrt{\dfrac{\bar{p}\ \bar{q}}{n}} = \sqrt{\dfrac{0.42\ (0.58)}{150}} = 0.0403$

b) $\bar{p}\ \pm\ 2.05\sigma_{\bar{p}} = 0.42 \pm 2.58(0.0403) = 0.42 \pm 0.104$

$= (0.316,\ 0.524)$

Interval Estimation of a Population Proportion

- Example: Political Science, Inc.

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, PSI interviewers ask registered voters who they would vote for if the election were held that day.

Interval Estimation of a Population Proportion

■ Example: Political Science, Inc.

In a current election campaign, PSI has just found that 220 registered voters, out of 500 contacted, favor a particular candidate.

PSI wants to develop a 95% confidence interval estimate for the proportion of the population of registered voters that favor the candidate.

Interval Estimation of a Population Proportion

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

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

$$.44 \pm 1.96\sqrt{\frac{.44(1-.44)}{500}} = .44 +/- .0435$$

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

Sample Size for an Interval Estimate of a **Population** Proportion

- Margin of Error

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

- Solving for the necessary sample size, we get

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

However, $\bar{p}$ will not be known until after we have selected the sample. We will use the planning value p^* for $\bar{p}$.

For a test market, find the sample size needed to estimate the true proportion of consumers satisfied with a certain new product within $\pm$ 0.04 at the 90 percent confidence level. Assume you have no strong feeling about what the proportion is.

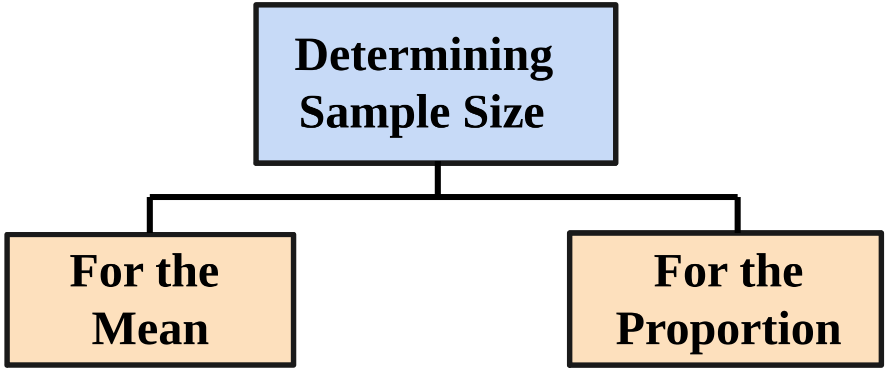
Solution:

Assume p = q = 0.5

$$0.04 = 1.64\sqrt{\frac{pq}{n}} = 1.64\sqrt{\frac{0.5\,(0.5)}{n}}$$

$$\text{So } n = \left(\frac{1.64(0.5)}{0.04}\right)^2 = 420.25 \text{ i.e., } n \geq 421$$

Determining Sample Size



Sampling Error

- The required sample size can be found to reach a desired margin of error (e) with a specified level of confidence (1 - α)

- The margin of error is also called sampling error
  - the **amount of imprecision in the estimate** of the population parameter
  - the amount added and subtracted to the point estimate to **form the confidence interval**

Determining Sample Size

Determining Sample Size

For the Mean

$$\overline{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Sampling error (margin of error)

$$e = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Determining Sample Size

(continued)

Determining Sample Size

For the Mean

$$e = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Now solve for n to get

$$\boxed{n = \frac{Z_{\alpha/2}^{2}\sigma^{2}}{e^{2}}}$$

Determining Sample Size

(continued)

- To determine the required sample size for the mean, you must know:
  - The desired level of confidence ($1 - \alpha$), which determines the critical value, $Z_{\alpha/2}$
  - The acceptable sampling error, e
  - The standard deviation, σ

Required Sample Size Example

If $\sigma = 45$, what sample size is needed to estimate the mean within $\pm$ 5 with 90% confidence????

Required Sample Size Example

If $\sigma = 45$, what sample size is needed to estimate the mean within $\pm$ 5 with 90% confidence?

$$n = \frac{Z^2 \sigma^2}{e^2} = \frac{(1.645)^2 (45)^2}{5^2} = 219.19$$

So the required sample size is **n = 220**

(Always round up)

If σ is unknown

- If unknown, σ can be estimated when using the required sample size formula
  - Use a value for σ that is expected to be at least as large as the true σ
  - Select a **pilot sample and estimate $\boldsymbol{\sigma}$ with** the sample standard deviation, S

1. A statistic is said to be an efficient estimator of a population parameter if, with increasing sample size, it becomes almost certain that the value of the statistic comes very close to that of the population parameter.
2. An interval estimate is a range of values used to estimate the shape of a population's distribution.
3. If a statistic tends to assume values higher than the population parameter as frequently as it tends to assume values that are lower, we say that the statistic is an unbiased estimate of the parameter.
4. The probability that a population parameter will lie within a given interval estimate is known as the confidence level.
5. With increasing sample size, the t distribution tends to become flatter in shape.
6. We must always use the t distribution, rather than the normal, whenever the standard deviation of the population is not known.

7. We may obtain a crude estimate of the standard deviation of some population if we have some information about its range.
8. When using the t distribution in estimation, we must assume that the population is approximately normal.
9. Using high confidence levels is not always desirable because high confidence levels produce large confidence intervals.
10. There is a different t distribution for each possible sample size.
11. A point estimate is often insufficient because it is either right or wrong.
12. A sample mean is said to be an unbiased estimator of a population mean because no other estimator could extract from the sample additional information about the population mean.