Chapter 6

Estimates and Sample Sizes

Chapter 6
Estimates and Sample Sizes

Overview

This chapter presents the beginning of inferential statistics using sample data to :

1. estimate a population parameter
2. test a claim (hypothesis) about a population

Overview

This chapter presents methods for:

- ❖estimating population proportions, means, and variances
- ❖determining sample sizes

6-2

Estimating a Population Proportion

Assumptions

1. The sample is a simple random sample.
2. The conditions for the binomial distribution are satisfied (See Section 4-3.)
3. The normal distribution can be used to approximate the distribution of sample proportions because $np \geq 5$ and $nq \geq 5$ are both satisfied.

We will do stuff like:

Example: 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

Find the 95% confidence interval for the population proportion *p*.

$$0.476 < p < 0.544$$

Notation for Proportions

$p =$ **population** proportion

$\hat{p} = \dfrac{x}{n}$ **sample** proportion of x successes in a sample of size n

(pronounced 'p-hat')

$\hat{q} = 1 - \hat{p} =$ **sample** proportion of x failures in a sample size of n

Definition

Point Estimate

A point estimate is a single value (or point) used to approximate a population parameter.

Definition

Point Estimate

The sample proportion $\hat{p}$ is the best point estimate of the population proportion p.

Definition

Confidence Interval

(or Interval Estimate)

a range (or an interval) of values <u>used to estimate</u> the true value of the population parameter

Lower # < population parameter < Upper #

As an example

Lower # < p < Upper #

Definition

Confidence Interval

(or Interval Estimate)

a range (or an interval) of values <u>used to estimate</u> the true value of the population parameter

Lower # < population parameter < Upper #

As an example

$$0.476 < p < 0.544$$

Definition

Confidence Level

(degree of confidence or confidence coefficient)

the probability 1 - α (often expressed as the equivalent percentage value) that is the <u>proportion of times the confidence interval actually does contain the population parameter</u>, assuming that the estimation process is repeated a large number of times

usually 90%, 95%, or 99%

(α = 10%), (α = 5%), (α = 1%)

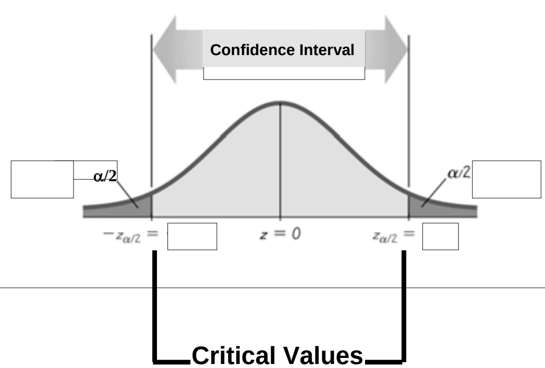
Interpreting a Confidence Interval

$$0.476 < p < 0.544$$

Correct: We are 95% confident that the interval from 0.476 to 0.544 actually does contain the true value of *p*. This means that if we were to select many different samples of size 829 and construct the confidence intervals, 95% of them would actually contain the value of the population proportion *p*.

Wrong: There is a 95% chance that the true value of *p* will fall between 0.476 and 0.544.

Finding Critical Values



Critical Value Observations

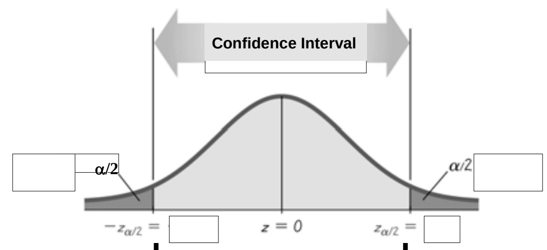
1. **The sampling distribution of sample proportions can be approximated by a normal distribution.**
2. **Sample proportions have a relatively small chance (denoted by α) of falling in one of the red tails.**
3. **Denoting the area of each shaded region by α/2, there is a total probability of α that a sample proportion will fall in either of the two red tails.**

Critical Value Observations

4. By the rule of complements, there is a probability of 1- α that the sample proportion will fall within the green-shaded region.
5. The *z* score separating the right-tail region is denoted by $z_{\alpha/2}$, and is referred to as a <u>critical value</u> because it is on the borderline separating sample proportions that are likely to occur from those that are unlikely to occur.

   (The value of $-z_{\alpha/2}$ is at the vertical boundary for area $\alpha/2$ in the left tail.)

Finding Critical Values



Critical Values

Definition

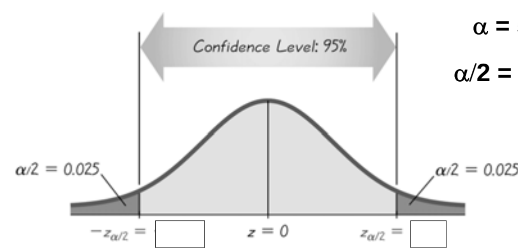
Critical Value

the number on the borderline separating sample statistics that are likely to occur from those that are unlikely to occur. The number $z_{\alpha/2}$ is a critical value that is a z score with the property that it separates an area $\alpha/2$ in the right tail of the standard normal distribution.

Finding $\pm z_{\alpha/2}$ for 95% Degree of Confidence

α = 5%

$\alpha/2$ = 2.5% = .025



Critical Values

NEGATIVE Z Scores   Table A-2

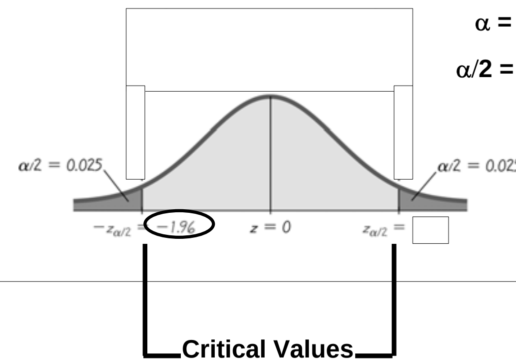
TABLE A-2 Standard Normal (z) Distribution: Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 and lower	.0001									
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	* .0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	* .0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559

Finding $\pm z_{\alpha/2}$ for 95% Degree of Confidence

α = 5%

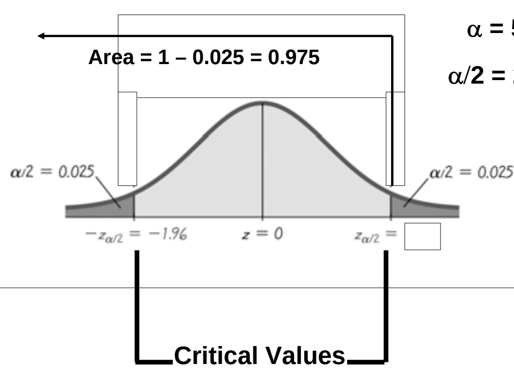
$\alpha/2$ = 2.5% = .025



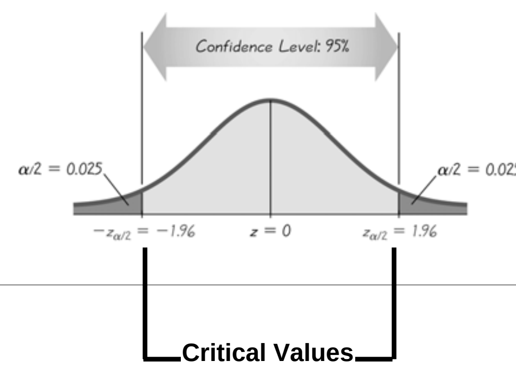
Finding $\pm z_{\alpha/2}$ for 95% Degree of Confidence

α = 5%

$\alpha/2$ = 2.5% = .025



Finding $\pm z_{\alpha/2}$ for 95% Degree of Confidence



Definition

Margin of Error

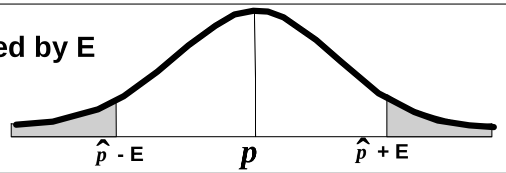
When data from a simple random sample are used to estimate a population proportion *p*, the margin of error, denoted by *E*, is the maximum likely (with probability $1 - \alpha$) difference between the observed proportion $\hat{p}$ and the true value of the population proportion *p*.

Definition

Margin of Error

is the maximum likely difference between observed sample proportion $\hat{p}$ and true population proportion *p*.

denoted by E



$$\hat{p} - E < p < \hat{p} + E$$

lower limit — upper limit

Margin of Error of the Estimate of *p*

Formula 6-1

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

Confidence Interval for Population Proportion

$$\hat{p} - E < p < \hat{p} + E$$

where

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$

Confidence Interval for Population Proportion

$$\hat{p} - E < p < \hat{p} + E$$

$$p = \hat{p} \pm E$$

$$(\hat{p} - E, \hat{p} + E)$$

Round-Off Rule for Confidence Interval Estimates of *p*

Round the confidence interval limits to

three significant digits

Procedure for Constructing a Confidence Interval for *p*

1. Verify that the required assumptions are satisfied. (The sample is a simple random sample, the conditions for the binomial distribution are satisfied, and the normal distribution can be used to approximate the distribution of sample proportions because $np \geq 5$, and $nq \geq 5$ are both satisfied).
2. Refer to Table A-2 and find the critical value $z_{\alpha/2}$ that corresponds to the desired confidence level.
3. Evaluate the margin of error $E = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Procedure for Constructing a Confidence Interval for *p*

4. Using the calculated margin of error, *E* and the value of the sample proportion, $\hat{p}$, find the values of $\hat{p} - E$ and $\hat{p} + E$. Substitute those values in the general format for the confidence interval:

$$\hat{p} - E < p < \hat{p} + E$$

5. Round the resulting confidence interval limits to three significant digits.

Example: 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

a) Find the margin of error E that corresponds to a 95% confidence level.

b) Find the 95% confidence interval estimate of the population proportion p.

c) Based on the results, can we safely conclude that the majority of adult Minnesotans oppose use the the photo-cop?

Example: In the Chapter Problem, we noted that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

a) Find the margin of error E that corresponds to a 95% confidence level

First, we check for assumptions. We note that $n\hat{p} = 422.79 \geq 5$, and $n\hat{q} = 406.21 \geq 5$.

Next, we calculate the margin of error. We have found that $\hat{p} = 0.51$, $\hat{q} = 1 - 0.51 = 0.49$, $z_{\alpha/2} = 1.96$, and $n = 829$.

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Example: In the Chapter Problem, we noted that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

a) Find the margin of error E that corresponds to a 95% confidence level

First, we check for assumptions. We note that $n\hat{p} = 422.79 \geq 5$, and $n\hat{q} = 406.21 \geq 5$.

Next, we calculate the margin of error. We have found that $\hat{p} = 0.51$, $\hat{q} = 1 - 0.51 = 0.49$, $z_{\alpha/2} = 1.96$, and $n = 829$.

$$E = 1.96\sqrt{\frac{(0.51)(0.49)}{829}}$$

$$E = 0.03403$$

Example: In the Chapter Problem, we noted that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

b) Find the 95% confidence interval for the population proportion p.

We substitute our values from Part a into:

$$\hat{p} - E < p < \hat{p} + E$$

Example: In the Chapter Problem, we noted that 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

b) Find the 95% confidence interval for the population proportion *p*.

We substitute our values from Part a to obtain:

$$0.51 - 0.03403 < p < 0.51 + 0.03403,$$
$$0.476 < p < 0.544$$

Example: 829 adult Minnesotans were surveyed, and 51% of them are opposed to the use of the photo-cop for issuing traffic tickets. Use these survey results.

c) Based on the results, can we safely conclude that the majority of adult Minnesotans oppose use of the photo-cop?

Based on the survey results, we are 95% confident that the limits of 47.6% and 54.4% contain the true percentage of adult Minnesotans opposed to the photo-cop. The percentage of opposed adult Minnesotans is likely to be any value between 47.6% and 54.4%. However, a majority requires a percentage greater than 50%, so we *cannot* safely conclude that the majority is opposed (because the *entire* confidence interval is not greater than 50%).

Example: In a survey of 1002 people, 701 said that they voted in a recent presidential election (based on data from ICR Research Group). Voting records show that 61% of eligible voters actually did vote.

a) Find a 99% confidence interval estimate of the proportion of people who say they voted.

b) Are the survey results consistent with the actual voter turnout or 61%? Why or why not?