SECTION 9.1 – Estimating a Population Proportion

Point Estimate

A **point estimate** is the value of a statistic that estimates the value of a parameter.

Point estimate of a population proportion, $p \rightarrow \hat{p}$.

Confidence-Interval

A **confidence interval** for an unknown parameter consists of an interval of numbers based on a point estimate.

Level of Confidence (or Confidence Level)

The level of confidence represents the expected proportion of intervals that will contain the parameter if a large number of different samples is obtained. The level of confidence is denoted $(1-\alpha)\cdot 100\%$.

Confidence-Interval Estimate

The confidence interval estimates for the population proportion are of the form

Point estimate $\pm$ margin of error.

Constructing a $(1-\alpha)\cdot 100\%$ Confidence Interval for a Population Proportion, *p*

Suppose that a simple random sample of size *n* is taken from a population or the data are the result of a randomized experiment. A $(1-\alpha)\cdot 100\%$ confidence interval for *p* is given by the following quantities:

Lower bound: $\hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ Upper bound: $\hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

or simply,

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

Note: It must be the case that $n\hat{p}(1-\hat{p}) \geq 10$ and $n \leq 0.05N$ to construct this interval.

Table of Areas and Critical Values for the 90%, 95%, & 99% Confidence Intervals

Level of Confidence, $(1-\alpha)\cdot 100\%$	Area in each tail, $\frac{\alpha}{2}$	Critical Value, $z_{\frac{\alpha}{2}}$
90%	0.05	1.645
95%	0.025	1.96
99%	0.005	2.575

Interpretation of a Confidence Interval

A $(1-\alpha)\cdot 100\%$ confidence interval indicates that $(1-\alpha)\cdot 100\%$ of all simple random samples of size *n* from the population whose parameter is unknown will result in an interval that contains the parameter.

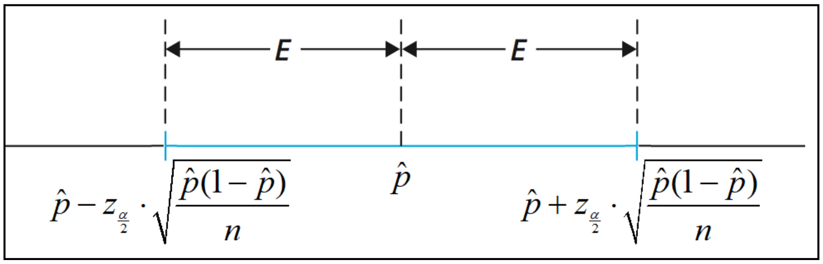
SECTION 9.1 – Estimating a Population Proportion

Margin of Error for the Estimate of *p*

The **margin of error, *E***, in a $(1-\alpha)\cdot 100\%$ confidence interval for a population proportion is given by

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

Note: This is simply half the length of the confidence interval.



Sample Size Needed for Estimating the Population Proportion *p*

The sample size required to obtain a $(1-\alpha)\cdot 100\%$ confidence interval for p with a margin of error of at most E is given by

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

rounded up to the nearest integer, where $\hat{p}$ is a prior estimate (or guess) of p.

If a prior estimate of p is unavailable, the sample size required is

$$n = 0.25\left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2$$

rounded up to the next integer.

☺ **Exercises:**

1) Using Table V, find the critical value $z_{\alpha/2}$ that corresponds to a 95% level of confidence.

2) Using Table V, find the critical value $z_{\alpha/2}$ that corresponds to a 90% level of confidence.

3) Determine the point estimate of the population proportion, the margin of error for a confidence interval that has a lower bound of 0.054, an upper bound of 0.074, and the number of individuals in the sample with a sample size of 1120.

SECTION 9.1 – Estimating a Population Proportion

☺ **Exercises:**

4) **Drinking Habits.** A *Reader's Digest/Gallup Survey* on the drinking habits of Americans estimated the percentage of adults across the country who drink beer, wine, or hard liquor, at least occasionally. Of the 1516 adults interviewed, 985 said that they drank.

a) Construct a point estimate for the population of American adults across the country who drink beer, wine, or liquor occasionally.

b) Verify that the requirements for constructing a confidence interval about p are satisfied.

c) Construct a 95% confidence interval for the population proportion of American adults across the country who drink beer, wine, or liquor occasionally. Also, interpret the interval.

d) Find the margin of error for the estimate of p.

5) When 460 junior college students were surveyed, 100 said they have a passport. Construct a 90% confidence interval for the proportion of all junior college students that have a passport and interpret your results.

SECTION 9.1 – Estimating a Population Proportion

☺ **Exercises:**

6) **Hypertension.** In a random sample of 678 adult males 20 to 34 years of age, it was determined that 58 of them have hypertension (high blood pressure).

Source: The Centers for Disease Control.

a) Obtain a point estimate for the proportion of adult males 20 to 34 years of age who have hypertension.

b) Construct a 95% confidence interval for the proportion of adult males 20 to 34 years of age who have hypertension. Interpret the confidence interval.

c) Find the margin of error for the estimate of p.

d) You wish to conduct your own study to determine the proportion of adult males 20 to 34 years old who have hypertension. What sample size would be needed for the estimate to be within 3 percentage points with 95% confidence if you use the point estimate obtained in part **a)**?

e) You wish to conduct your own study to determine the proportion of adult males 20 to 34 years old who have hypertension. What sample size would be needed for the estimate to be within 3 percentage points with 95% confidence if you don't have a prior estimate?

SECTION 9.1 – Estimating a Population Proportion

☺ **Exercises:**

7) A researcher wants to estimate the proportion of X-ray machines that malfunction and produce excess radiation. A random sample of 60 machines is taken and 15 of the machines were labeled as malfunctioned. Find a 99% confidence interval on p the proportion of machines that malfunction in the population.

8) In a college student poll, it is of interest to estimate the proportion p of students in favor of changing from a quarter-system to a semester-system. How many students should be polled so that we can estimate p to within 0.09 using a 90% confidence interval?

9) A pollster wishes to estimate the number of left-handed scientists. How large a sample is needed in order to be 95% confident that the sample proportion will not differ from the true proportion by more than 3%? A previous study indicates that the proportion of left-handed scientists is 9%.

SECTION 9.1 – Estimating a Population Proportion

<u>Interpretation and the Meaning of Confidence Intervals for *p*</u>

Consider **Exercise #4c**:

> *Construct a 95% confidence interval for the population proportion of American adults across the country who drink beer, wine, or liquor occasionally.*

So, we found that the 95% confidence interval for the population proportion, p, was (0.626, 0.674). Students and teachers sometimes refer to this by saying that "We are 95% confident that the population proportion, p, of American adults across the country who drink beer, wine, or liquor occasionally is between 0.626 and 0.674".

However, while this wording is okay for student use in the classroom, it is technically incorrect, and would not be used by actual statisticians when interpreting the meaning of a confidence interval. Because you would never describe yourself as being 75% happy or 50% in love with someone, you also can't be 95% confident about something. Do not confuse what we perceive as 'confident' with what it really means to be confident when it relates to statistics.

We know the word 'confident' means full of conviction or having assurance and self-reliance.

However, in statistics, **a 95% level of confidence means that 95% of all possible samples result in confidence intervals that include the paramete**r (and 5% of all possible samples result in confidence intervals that do not include the parameter).

Another wording is as follows: A 95% confidence interval means that the method used to obtain the confidence interval produces an interval that captures the unknown, p, 95% of the time. To be more precise a statistician would say that if we were to take 100 samples of the same from the population, and from those samples construct 100 sample proportions, and then construct 100 confidence intervals around the sample proportions at the 95% level, then 95 of the 100 confidence intervals constructed from the samples would include the true population proportion, while the other 5 of the 100 would not. So, for an individual sample of size 1516 from the American adult population there is a 95% chance that its corresponding confidence interval encapsulates the population proportion.

Let's illustrate what "95% confidence" with regards to a 95% confidence interval using a simulation and a visual. We will simulate obtaining 100 different random samples of size $n = 1516$ from a population with $p = 0.60$. (Here we will assume that this is the true population proportion of Americans adults that drink occasionally. The figure to the right shows the confidence intervals in a group of 100. Note that each confident interval is exactly the same length of 0.048, with all different centers of $\hat{p}$. A black interval is a 95% confidence interval that includes the population proportion, 0.60. A red interval is a confidence interval that does not include the population proportion. Notice that the red intervals that do not capture the population proportion 0.6 have centers that are far away (more than 1.96 standard errors or more than 0.024 standard deviations) from 0.6. Of the 100 confidence intervals obtained, 5 (the red intervals) do not include the population proportion. For example, the first interval to miss has a sample proportion that is too small to result in an interval that captures 0.6. So, 95/100 = 0.95 (or 95%) of the samples have intervals that capture the population proportion. Again, a 95% level of confidence means that 95% of all possible samples result in confidence intervals that include the parameter (and 5% of all possible samples result in confidence intervals that do not include the parameter).