

Cardinal Rule of Statistical Inference

- **NEVER- NEVER- NEVER- NEVER- NEVER report a point estimate without a corresponding report of the uncertainty (variability) of the estimate.**
- One method of reporting uncertainty is by stating the margin of error
- Another method of reporting uncertainty is by constructing a confidence interval

Proportions # 1

Inference Types

- **Statistical inferences** are made when the target and sampled populations are the same
- **Judgement inferences** are the result of studies done where the target and sampled populations are not the same

Proportions # 2

Confidence Interval

- A **confidence interval** for a population parameter is a numeric interval that is computed from the sample data.
- The associated confidence is really a probability that the computed interval actually contains the true population parameter of interest.

Proportions # 3

General CI - Formula

- Any confidence interval has the general formula:

Parameter = Point Estimate +/- Margin of Error

Proportions # 4

Margin of Error (ME)

- The **margin of error** is the measure of the variability associated with the point estimate at the desired level of confidence
- Small margins of error imply a higher precision than large margins of error
- The higher the desired confidence level, the larger the margin of error

Proportions # 5

Endpoints

- A confidence interval is made up of two endpoints which enclose a range of values.
- The lowest value in the computed confidence interval is called the **Lower Endpoint**.
- The largest value in the computed confidence interval is called the **Upper Endpoint**.

Proportions # 6

Confidence Level

- This is a proportion associated with ANY confidence interval.
- Most often confidence levels of 0.90, 0.95, 98, and 0.99 are used.
- You might also hear the intervals associated with these values referred to as 90, 95, 98, and 99 percent confidence intervals

Proportions # 7

Interpretation of Confidence Intervals

- One popular interpretation is that, “After the experiment is done the investigator is 95% (or 90 or 99%) certain (confident) that the computed interval contains the **TRUE** population parameter of interest.”
- We will use this definition for class purposes.

Proportions # 8

The Point Estimate

- Just as we used the sample mean as the point estimate for the true population mean – so we use the sample proportion as the point estimate for the true population proportion.
- The sample proportion is denoted as: \hat{p}

Proportions # 9

Computing the point estimate

- To compute p-hat, use the following formula:

$$\hat{p} = \frac{\text{number of outcomes of interest}}{n}$$

Proportions # 10

Margin of Error (ME)

To Compute ME for the sample proportion you use:

$$ME = z \cdot \sqrt{\frac{(\hat{p})(1 - \hat{p})}{n}}$$

Proportions # 11

CI's for Proportions

Parameter = Point Estimate +/- Margin of Error

$$p = \hat{p} \pm z \cdot \sqrt{\frac{(\hat{p})(1 - \hat{p})}{n}}$$

Proportions # 12

Example

A political pollster would like to estimate the true proportion of the population of county residents that favor a controlled growth ballot issue. A SRS of 350 county residents found that 230 of them favored the issue. Use this information to construct a 95% CI for the true proportion of residents who favor the issue.

Proportions # 13

Compute the Point Estimate

- Another way to think of successes is to view a “success” as an outcome of interest. In this example the outcome of interest is “favoring” the ballot issue
- So the point estimate is:

$$\hat{p} = \frac{230}{350} = 0.657$$

Proportions # 14

Find the z-value

- There are 4 popular levels of confidence. Each one has its own corresponding z-value
- In this example CL = 95% so the z-value is 1.96

CL	z-value
90%	1.645
95%	1.96
98%	2.33
99%	2.58

Proportions # 15

Compute the ME - 1

$$ME = z \cdot \sqrt{\frac{(\hat{p})(1 - \hat{p})}{n}}$$

$$ME = 1.96 \cdot \sqrt{\frac{(0.657)(1 - 0.657)}{350}}$$

Proportions # 16

Compute the ME - 2

$$\begin{aligned}ME &= 1.96 \cdot \sqrt{\frac{(0.657)(0.343)}{350}} \\ &= 1.96 \cdot \sqrt{.000644} \\ &= 1.96 (0.0254) \\ &= 0.0497\end{aligned}$$

Proportions # 17

Confidence Interval

- Now put everything together

Parameter = Point Estimate +/- Margin of Error

$$p = 0.657 \pm 0.0497$$

$$p_l = 0.657 - 0.0497 = 0.607$$

$$\begin{aligned}p_u &= 0.657 + 0.0497 = 0.707 \\ &= (0.607, 0.707)\end{aligned}$$

Proportions # 18

Interpretation of Confidence Intervals

- The English interpretation follows the format described for interpreting confidence intervals described in a previous slide
- In this case, we are 95% confident that the true proportion of county voters in favor of the ballot issue is between 0.607 and 0.707
- Note that you might also interpret in terms of percentages

Proportions # 19

Sample Size Determination Population Proportions

- Researchers often want to plan a study to insure that the uncertainty in the point estimate will not exceed some specified value.
- Another way to state this is that the investigators would like to be sure that, at the end of the study, a CI for the parameter of interest does not exceed some prescribed width or that the ME doesn't exceed some specified value.

Proportions # 20

Sample Size Determination

- The superintendent of a large school district wants to estimate p , the proportion of first-graders who have not had their immunization shots. She plans to use a SRS of first-graders to obtain the estimate and she wants to be 95% confident that the point estimate \hat{p} will be no more than 0.05 units from the true value of p .

Proportions # 21

Sample Size Determination – The conservative approach

- In order to obtain a CI at a specified level of confidence and ME you need to take “n” observations where:

$$n = \frac{z^2}{4E^2}$$

Proportions # 22

Sample Size Determination if you have an estimate of p

- What do we know?
 - $E = 0.05$
 - $\hat{p} = 0.10$
 - $(1 - \hat{p}) = 1 - 0.10 = 0.90$
 - z comes from the t -table. For a 95% CI the value of $z = 1.96$

Proportions # 23

Sample Size Determination if you have an estimate of p

- Now it's just plug-n-chug

$$\begin{aligned} n &= \frac{z^2}{4E^2} \\ &= \frac{1.96^2}{4(0.05)^2} \\ &= 384.2 \rightarrow 385 \end{aligned}$$

- Note that we ALWAYS round up!!

Proportions # 24

The Estimation of Population Size

- In order to wisely manage the annual elk hunting season the division of wildlife would like to estimate the total number of elk in the state.
- In order to determine whether or not to remove a species from the endangered species list the EPA wants to estimate the population size of that species.

Proportions # 25

Capture/Mark/Recapture

- Estimation of the sizes of “hard to census” populations is often carried out via the method of “capture/recapture”.
- This is a two-stage method.
- During the first stage a group from the population of interest is captured and marked in some way.
- After a prescribed time has passed a second sample is “re”captured and the proportion of marked units in this group is used to compute a sample proportion.

Proportions # 26

Capture/Mark/Recapture

- The value of the sample proportion is used further to estimate the true population size.
- The underlying premise is, as usual, that the sample is a good representation of the population.
- In other words, if my resample contains 10% marked units then we say that the entire population must contain 10% marked units.
- This means that the number of units caught and marked at the first stage must have been about 10% of the entire population of interest.

Proportions # 27

Setting

- A wildlife biologist would like to estimate the number of black bear living in the western part of the state. She spends the summer capturing and marking 82 black bear. The following year she collects another sample of 95 bear. 6 of these had been marked. Develop a point estimate and corresponding 95% CI for the population size of black bear in the western part of the state.

Proportions # 28

Variables and Values

- Let m be the number of bear in the first sample that were caught and marked:
→ $m = 82$
- Let n be the number of bear in the caught in the recapture phase:
→ $n = 95$
- Let k be the number of marked bear present in the recaptured sample:
→ $k = 6$

Proportions # 29

Step 1A: Compute \hat{p}

$$\begin{aligned}\hat{p} &= \frac{k}{n} \\ &= \frac{6}{95} \\ &= 0.0632\end{aligned}$$

Proportions # 30

Step 1B: Interpret \hat{p}

- We can interpret this statistic as:
The estimated proportion of marked bears in the entire population of bears
- Since we know the total number of marked bears - because we put the marks on'em - we can estimate the total number of bears in the population

Proportions # 31

Step 2: Solve For \hat{N}

$$\begin{aligned}\hat{p}\hat{N} &= m \\ \hat{N} &= \frac{m}{\hat{p}} \\ &= \frac{82}{0.0632} \\ &= 1298.3 \rightarrow 1298\end{aligned}$$

Proportions # 32

CI for N

- We'll get a CI for N by using the CI for p
- p is the true proportion of marked elk in the population.
- We don't have this value but we have an estimate of it. Recall that p-hat is 0.0632.

Proportions # 33

CI for p

Recall that we can construct the CI of a population proportion by using the relation:

Parameter = Point Estimate +/- Margin of Error

$$p = \hat{p} \pm z \cdot \sqrt{\frac{(\hat{p})(1 - \hat{p})}{n}}$$

Proportions # 34

CI for p

- Here \hat{p} is the estimated proportion of marked bear.
- n is the number of marked critters in the re-captured sample.
- Since we want to develop a 95% CI, the appropriate z-value is 1.96

Proportions # 35

CI for p

$$\begin{aligned} p &= 0.0632 \pm 1.96 \sqrt{\frac{(0.0632)(1 - 0.0632)}{95}} \\ &= 0.0632 \pm 0.0489 \\ &= (0.0143, 0.112) \end{aligned}$$

Proportions # 36

CI for N

- Just as we used \hat{p} to obtain an estimate for N so we'll use the interval endpoints to obtain a CI for N
- Call the upper endpoint: \hat{p}_u
- And call the lower endpoint: \hat{p}_l

Proportions # 37

CI for N

- Now use the endpoints of the CI for p to get the endpoints for the CI for N
- The upper endpoint is: $\frac{m}{\hat{p}_l}$
- And the lower endpoint will be: $\frac{m}{\hat{p}_u}$

Proportions # 38

CI for N

- Now use the endpoints of the CI for p to get the endpoints for the CI for N
- The upper endpoint is: $\frac{82}{0.0143} = 5734 .3 \rightarrow 5734$
- And the lower endpoint will be: $\frac{82}{0.112} = 732 .1 \rightarrow 732$

Proportions # 39

Interpretation of the CI

- This CI is interpreted like all of the other confidence intervals we've looked at thus far.
- That is: We are 95% confident that the true number of bears in the population is between 732 and 5734.
- Note that this is a very wide interval and would probably be of little use!

Proportions # 40

The Estimation of Population Totals

- In 2002, the estimated total number of highway deaths where alcohol was a contributing factor is: _____
- The estimated number of practicing Catholics in Larimer County is: _____
- The estimated number of college undergraduates who use recreational drugs is: _____

Proportions # 41

The Estimation of Population Totals

- These statistics are computed in three steps
 - 1) Estimate the proportion of the population with the characteristic of interest
 - 2) Construct a CI for this estimate
 - 3) Use the values computed in steps 1 and 2 to obtain estimates and limits for the total number of units in a population that exhibit the characteristic of interest

Proportions # 42

The Estimation of Population Totals

In 2004 a survey was commissioned by the Wyoming state legislature to estimate the number of its citizens that were living at or below the poverty line. In a random sample of 504 of its residents 63 reported a annual income that placed them at or below the poverty line.

Proportions # 43

Step 1A: Compute \hat{p}

$$\hat{p} = \frac{63}{504} = 0.125$$

Proportions # 44

Use census estimates for the population

To estimate the total number of persons living at or below the poverty line we need to know the total number of people living in the state during 2004.

This information is usually obtained through the national census database

In 2004 the estimated population was 506,529

Proportions # 45

The Estimation of Population Totals

So the estimated total number of persons living at or below poverty is:

$$0.125(506,529) = 63,316$$

Recall that all reported estimates need to include a report of the uncertainty

Proportions # 46

Uncertainty in estimates of Population Totals

To report uncertainty we'll use the limits of the CI for the proportion of Wyoming residents that are living at or below the poverty line.

Proportions # 47

CI for p

For purposes of illustration we'll use a 90% level of confidence

$$\begin{aligned} p &= \hat{p} \pm z \cdot \sqrt{\frac{(\hat{p})(1 - \hat{p})}{n}} \\ &= 0.125 \pm 1.645 \cdot \sqrt{\frac{(0.125) \cdot (1 - 0.125)}{504}} \\ &= (0.101, 0.149) \end{aligned}$$

Proportions # 48

CI for N

- The lower limit for N:
 - $N_l = 0.101(506,529) = 51,159$
- The upper limit for N:
 - $N_u = 0.149(506,529) = 75,473$

Proportions # 49

Interpret the CI for N

- The estimated number of Wyoming residents living at or below the poverty line is 63,316
- We are 90% confident that the true number is between 51,159 and 75,473

Proportions # 50