AP Statistics

Chapter 8:
Estimating with Confidence

Day 1

HW: Lesson 1 Practice Worksheet

Why Do We Use Confidence Intervals?

I am not sure what the population parameter is but I am somewhat confident that it falls between here and there.

We use confidence intervals when we do not have enough information to find the true population proportion (p) or what we believe to be the true population proportion ("p-hat").

Confidence intervals are based on taking many, many samples to come up with a range of values that we believe will contain the true population proportion.

Why Do We Take Samples?

Left-handed people

- To find or estimate the proportion of people that are left-handed.
- Or we are told that the true proportion of people that are left-handed was a certain proportion and we beleive it has changed over time (either a higher or lower proportion). So we take a sample to investigate that.
- Or if there is something new out there that we do not know anything about.

Why Do We Take Samples?

Coral reefs in Caribbean

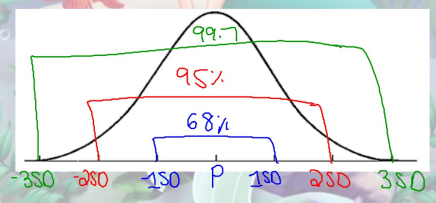
- The coral reefs in the Caribbean are slowly being infected by a new disease. So, I have no idea what the true population proportion is for the number of reefs that are infected. So I am going to try and estimate a range of values that I think will contain that true population proportion.

What We Know About Sampling Distributions

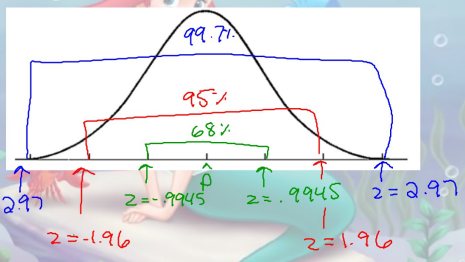
- A sampling distribution represents many, many samples with the true p in the middle, with a standard deviation of

$$\text{sample } \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$\sigma_{\hat{p}} = \sqrt{np(1-p)}$



The Truth About Standard Deviation

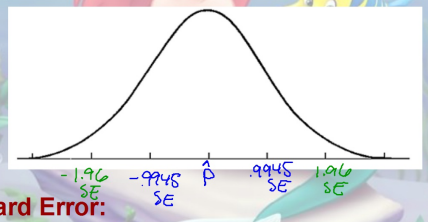


- We call these values **z*** or **critical z-scores**.

"2-stars"

How To Find Confidence Intervals

- Coral reefs in the Caribbean.

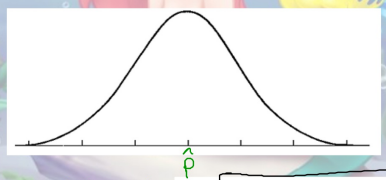


Standard Error:

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

95% Confidence Intervals

- The confidence interval is trying to find the true population proportion based on our samples.



Margin of error

Confidence Interval:

$$\hat{p} \pm 2^{*} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

point estimate

SE

95% Confidence Intervals

- Why do we use SE instead of SD?

We are using a range of values instead of 1 specific #. We do not know the pop. proportion or $\hat{p}$.

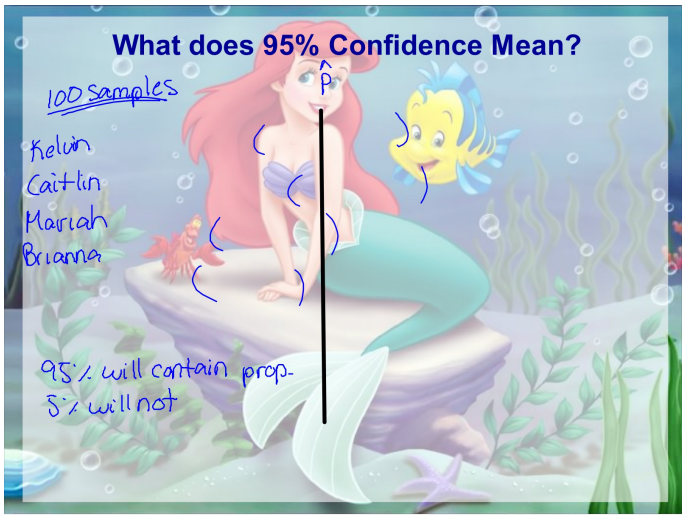
- Why do we use 1.96 instead of 2?

Because it is more exact.

What does 95% Confidence Mean?

- IT DOES NOT MEAN a 95% probability or chance that the mean lies within this interval!!!
- Confidence intervals should never be used to make predictions about a population because it is not a probability.
- It means that if you ran many, many samples, and found many, many "p-hats", 95% of those "p-hat" would create intervals that would contain the true population proportion.

What does 95% Confidence Mean?



The Conditions for Confidence Intervals

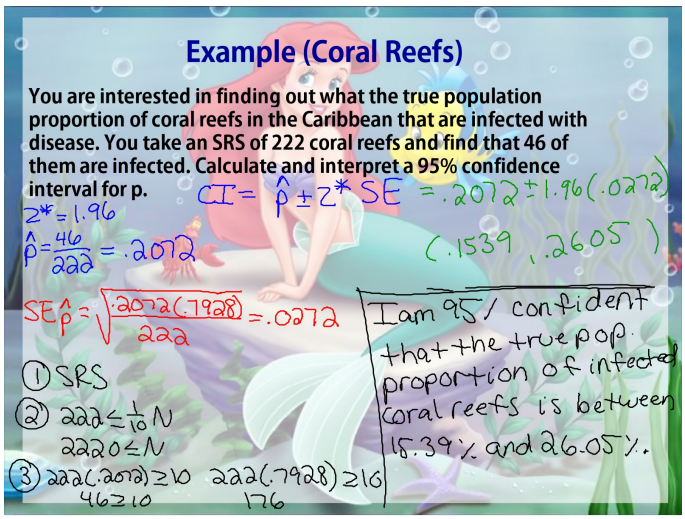
- Is it a random sample? Stated
- 10% condition to use the standard error formula.

$n \leq \frac{1}{10} N$

- Normal condition to use the z* scores.

$np \geq 10$ Successes

$n(1-p) \geq 10$ failure


Example (Coral Reefs)
You are interested in finding out what the true population proportion of coral reefs in the Caribbean that are infected with disease. You take an SRS of 222 coral reefs and find that 46 of them are infected. Calculate and interpret a 95% confidence interval for p.
$z^* = 1.96$
$\hat{p} = \frac{46}{222} = .2072$
$CI = \hat{p} \pm z^* SE = .2072 \pm 1.96(.0272)$
$(.1539, .2605)$
$SE_{\hat{p}} = \sqrt{\frac{.2072(.7928)}{222}} = .0272$
I am 95% confident that the true pop. proportion of infected coral reefs is between 15.39% and 26.05%.
1. SRS
2. $222 \leq \frac{1}{10}N$
$2220 \leq N$
3. $222(.2072) \geq 10$ $222(.7928) \geq 10$
$46 \geq 10$ 176