AP Stats – Chap 18
Confidence Intervals for Proportions

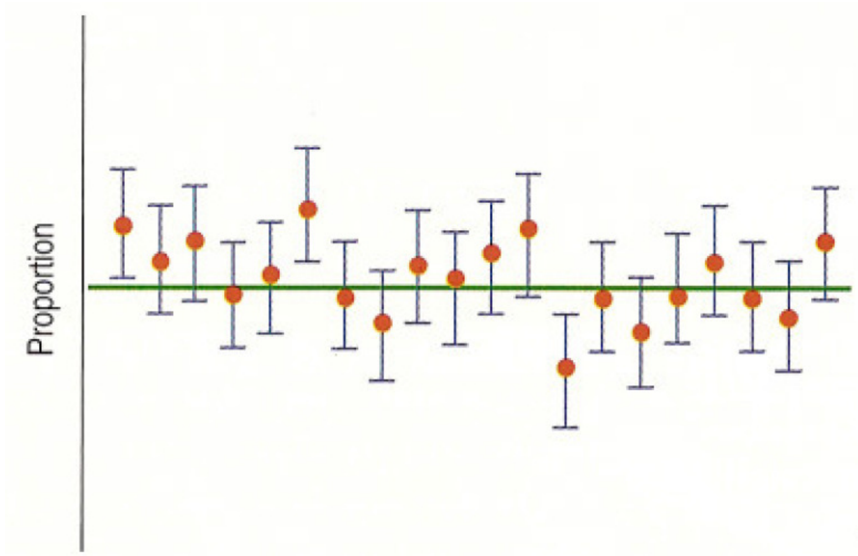
In many cases, we will not know p, the <u>true</u> proportion of success for the entire population, but we will know the proportion of success for our sample. This is denoted $\hat{p}$. Also…the proportion of failure of our sample will be $\hat{q}$.

Therefore, we can't find the SD of the sample, but we <u>can</u> find the **Standard Error of the Sample** using the formula…$SE = \sqrt{\frac{\hat{p}\hat{q}}{n}}$. (Does it look familiar?)

> ***"We are 95% confident that between 42.1% and 61.7% of Las Redes sea fans are infected."***

This type of statement is a **confidence interval**. They are also known as **one-proportion z-intervals**.

Confidence Intervals



The green line is…
The red dots are…
The blue segments are…
The CLT says…

Every confidence interval is a balance between _________ and __________.

Margin of Error (ME)
Critical Values

The extent of the blue segment on either side of $\hat{p}$ is called the **margin of error**.

$$ME = \left(z^*\right)\left(SE\right)$$

where z^* = the z-value (the **critical value**) associated with the specific confidence interval you want.

For 90% confidence interval, the critical value is 1.645.
For 95% confidence interval, the critical value is 1.96.
For 98% confidence interval, the critical value is 2.326.
For 99% confidence interval, the critical value is 2.58.

How 'bout an example? ☺

New Medicine

An experiment finds that 27% of 53 subjects report improvement after using a new medicine.

1. Check the conditions and assumptions!

2. State/calculate all needed values.

3. Create a 95% confidence interval for the actual cure rate.

4. Interpret the confidence interval in this context using all relevant terms in your explanation.

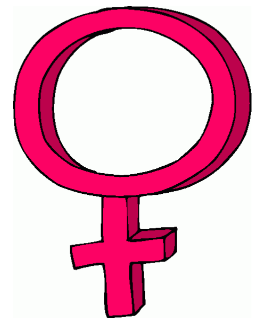
5. This interval is too wide. Make it narrower… 90% confidence.

6. What are the advantages and disadvantages of this narrower interval?

7. Explain what the phrase "90% confidence" means in this context.

8. What sample size would we need in a follow-up study if we want a margin of error of only 3% with 98% confidence?

How 'bout another one? ☺



Female Workers

The countries of Europe report that 46% of the labor force is female. The United Nations wonders is the percentage of females in the labor force is the same in the United States. Representatives from the United Stated Department of Labor plan to check a random sample of over 10,000 employment records on file to estimate a percentage of females in the US labor force.

9. The representatives from the Department of Labor want to estimate a percentage of females in the US labor force to within +/- 5%, with 90% confidence. How many employment records should they sample?

10. They actually select a random sample of 525 employment records, and find that 229 of the people are females. Create the confidence interval.

11. Interpret the confidence interval in this context.

12. Explain what 90% confidence means in this context.

13. Should the representatives from the Department of Labor conclude that the percentage of females in their labor force is lower than Europe's rate of 46%? Explain.

Final one! ☺

Private Schools

Virginia's Department of Education reports that 12% of the high school students in the Commonwealth attend private high schools. The local university wonders if the percentage is the same in their applicant pool for the upcoming fall semester. Admissions officers plan to check a random sample of the over 10,000 applicants on file to estimate the percentage of students applying for admission who attend private schools.

14. The admissions officers want to estimate the true percentage of private school applicants within +/- 4%, with 95% confidence. How many application should they sample?

15. They actually select a random sample of 450 applications, and find that 46 of those students attend private schools. Create the confidence interval.

16. Interpret the confidence interval in this context.

17. Explain what 95% confidence means in this context.

18. Should the admissions officers conclude that the percentage of private school students in their applicant pool is lower than the statewide enrollment rate of 12%? Defend your answer.

AP Stats – Chap 18
Highlights

This chapter provided your first formal exposure to **statistical inference** by introducing you to **confidence intervals**, a widely used technique. You learned how to construct a confidence interval for a population proportion (known as a "parameter"). You also examined how to interpret both the resulting interval and also what the **confidence level** means.

For example, using a 95% C-level gives you...

- 95% confidence that the interval contains the actual value of the unknown population parameter (the population proportion), and
- statistical support that 95& of all intervals generated by this procedure – in the long run – will succeed in "capturing" the unknown population value (this is the CLT).

You also investigated the effects of sample size and confidence level on the interval and its **margin of error**. The *ideal* confidence interval is very narrow with a very high C-level. But there's a trade-off: using a higher confidence level produces a wider interval if all else remains the same. One solution is to use a larger sample. The larger sample produces a narrower interval for the same confidence level.

You also learned how to plan ahead (working backwards) by determining, before collecting any data, the sample size you would need to achieve a certain margin of error for a given confidence level.

The general form of all confidence intervals is **Estimate +/- Margin of Error**.
And **Margin of Error = (Critical Value) x (Standard Error of the Estimate)**.

In symbols, this looks like $\hat{p} \pm \left(z^*\right)\left(\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$.

The margin of error is affected by several factors, primarily

- a higher confidence level produces a greater margin of error (a wider interval)
- a larger sample size produces a smaller margin of error (a narrower interval)

Common C-levels are 90%, 95%, 98%, and 99%. The phrase "95% confidence" **means** that if you were to take a large number of random samples and use the same confidence interval procedure on each sample, then – in the long run – 95% of those intervals would succeed in capturing the actual parameter value. Note that this is **NOT** the same as saying there is a 95% probability that the parameter is inside the calculated interval. It's technical in the wording, but the difference is IMPORTANT.

Always check the **conditions** before applying this 1-Propotion z-Interval.

- The sample must be large enough to guarantee **np** and **nq** are each greater than or equal to ten
- If the sample was selected randomly from the population, you're good to go. If not, can it be safely assumed?
- Is it safe to assume that the *measurements you're collecting* are independent from each other?
- The sample must be less than 10% of the population.

Coming Up...

You will continue to study inference procedures for a population proportion. You will also investigate the second major type of statistical inference procedure: **Tests of Significance**.