CHAPTER SEVEN: CONFIDENCE INTERVALS

Review February 15 ↝ Test February 24

A c% confidence interval for a mean is a range centered around the sample mean that is c% likely to include the population mean. If σ is known, the confidence interval can be found by adding and subtracting $z(\frac{\sigma}{\sqrt{n}})$ from the sample mean, where z is the number of standard errors needed to achieve the desired percentage as in chapter 6. If σ is not known, Student's t distribution is used instead of the normal distribution, which is similar to the normal distribution but more spread out and dependent on sample size. Confidence intervals for other parameters such as proportion can be calculated as well.

7-A Confidence Intervals for a Mean — Monday • 2/6

confidence interval • confidence level • critical value • *t* distribution • *t* chart • degrees of freedom • margin of error

❶ Find a critical value z_0 or t_0.

❷ Find a confidence interval for a mean and interpret it in words.

7-B Confidence Intervals for a Proportion — Friday • 2/10

❶ Find a confidence interval for a proportion and interpret it in words.

❷ Interpret a poll result based on its margin of error.

7-C Sample Size Needed for a Specified Margin of Error — Wednesday • 2/15

❶ Estimate the sample size needed to achieve a specified margin of error for a confidence interval for a mean.

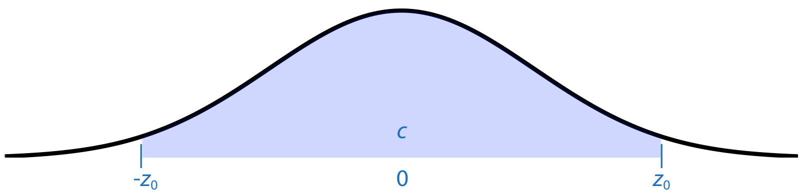
❷ Estimate the sample size needed to achieve a specified margin of error for a confidence interval for a proportion.

7-A Confidence Intervals for a Mean

A CONFIDENCE INTERVAL is a range around a sample statistic (such as $\bar{x}$) that is likely to include the true population parameter (in this case μ).

The CONFIDENCE LEVEL c, typically 90%, 95%, or 99%, is the probability, based on the sample data, that the parameter is in fact in the confidence interval.

The CRITICAL VALUE z_0 (or z_c) is the z score distance from the middle to the edge of the confidence interval.



Since z is based on σ, z cannot be used if σ is unknown. Instead, Student's T DISTRIBUTION can be used.

Like z, t is the number of standard deviations above the mean, and $t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$. The t distribution appears exactly the same as the normal (z) distribution except that it is a little flatter. The larger the sample size is, the closer the t distribution is to a perfect normal distribution.

The T table (Table 6) in the book states the t value needed to include c of the t distribution. This value is based not only on c but also on the degrees of freedom.

DEGREES OF FREEDOM (df) is the number of data values that are free to vary and still satisfy the given statistics. In the case of a single statistic such as a mean, once all but one of the data values are known, the last data value can be calculated. Therefore, for a mean, $df = n - 1$.

❶ Find a critical value z_0 or t_0.

1. If σ is known, find z_0 by doing any of the following:
   - Cut c in half, add it to 50%, and find this amount inside the z chart.
   - Find c at the top of the t chart and go down to the bottom row.
   - Memorize the common critical values: $z_0 = 1.645$ for $c = .90$, $z_0 = 1.960$ for $c = .95$, and $z_0 = 2.575$ for $c = .99$.
2. If σ is unknown, find t_0 by finding the degrees of freedom df and cross referencing this number on the left side of the t table with the confidence level c at the top.

❶ Find the critical value for the following confidence intervals.

a) $c = 95\%$, $\sigma = 8.1$, $n = 45$
$z_0 = 1.96$

b) $c = 80\%$, $\sigma = 0.72$, $n = 11$
$z_0 = 1.28$ yields .90 inside the z-chart

c) $c = 90\%$, $s = 21$, $n = 8$
$df = 8 - 1 = 7$, so $t_0 = 1.895$

The MARGIN OF ERROR E is the raw score distance from the middle to the edge of the confidence interval. This distance is z_0 (or t_0) times the standard error $\sigma_{\bar{x}}$ or $s_{\bar{x}}$: $E = z_0 \frac{\sigma}{\sqrt{n}}$ (or $t_0 \frac{s}{\sqrt{n}}$).

❷ Find a confidence interval for a mean and interpret it in words.

1. Find the critical value z_0 if σ is known or t_0 is σ is unknown.
2. Calculate the error E.
3. The confidence interval is $(\bar{x} - E) < \mu < (\bar{x} + E)$.
4. We are c confident that μ is between $\bar{x} - E$ and $\bar{x} + E$.

❷ The heights of 6 random turner trees after one year average 84 cm with standard deviation 10 cm. Use $c = .95$.

1. $df = 6 - 1 = 5$, so $t_0 = 2.571$
2. $E = 2.571\left(\frac{10}{\sqrt{6}}\right) \approx 10.5$
3. $(84 - 10.5) < \mu < (84 + 10.5)$ which is $73.5 < \mu < 94.5$
4. We are 95% confident that the population mean height of one-year-old turner trees is between 73.5 cm and 94.5 cm.

7-B Confidence Intervals for a Proportion

Standard error for a proportion is $\sqrt{\frac{pq}{n}}$ instead of $\frac{\sigma}{\sqrt{n}}$. Therefore, for proportions $E = z_0\sqrt{\frac{pq}{n}}$. Always use z for proportions, never t.
To use this formula, np and nq should both be greater than 5. $np \approx n\hat{p}$ = number of successes and $nq \approx n\hat{q}$ = number of failures.

❶ Find a confidence interval for a proportion and interpret it in words.

1. Identify the critical value z_0 (see 8-A ❶).
2. Identify $\hat{p}$ and $\hat{q}$ and use them as estimates for p and q. If they are fractions, do not round them (that is, do not change them to decimals).
3. Calculate the error E.
4. The confidence interval is $(\hat{p} - E) < p < (\hat{p} + E)$.
5. We are c confident that p is between $\hat{p} - E$ and $\hat{p} + E$.

❶ In a Gallup survey in January 2016, 1017 American adults were asked, "Is there any candidate running who you think would make a good president?" 671 said yes. Create a 95% confidence interval.

1. $z_0 = 1.96$
2. $p \approx \frac{671}{1017}$, $q \approx \frac{346}{1017}$
3. $E \approx 1.96\sqrt{\left(\frac{671}{1017}\right)\left(\frac{346}{1017}\right) / 1017} = .029$
4. $\left(\frac{671}{1017} - .029\right) < p < \left(\frac{671}{1017} + .029\right)$ which is $63.1\% < p < 68.9\%$
5. We are 95% confident that the proportion of Americans who, 10 months before the election, felt that there is a good presidential candidate was between 63.1% and 68.9%.

When pollsters state a margin of error they are referring to E for a 95% confidence interval. In this case, $E \approx \frac{1}{\sqrt{n}}$ is a reasonable estimate.

❷ Interpret a poll result based on its margin of error.

1. Add and subtract the margin of error E provided.
2. Write it as a 95% confidence interval between the two values calculated in step 1.
3. They are c confident that p is between $\hat{p} - E$ and $\hat{p} + E$ (or that μ is between $\bar{x} - E$ and $\bar{x} + E$).

❷ In a Gallup survey last year, 13% of Americans said they would like to see gays and lesbians less widely accepted in America. The margin of error was ±4 percentage points.

1. $\hat{p} + E = 13\% - 4\% = 9\%$
   $\hat{p} - E = 13\% + 4\% = 17\%$
2. $9\% < p < 17\%$
3. They are 95% confident that the proportion of Americans (as of 2016) who would like to see gays and less widely accepted in America is between 9% and 17%.

7-C Sample Size Needed for a Specified Margin of Error

The E formulas involving z can be solved for n so that pollsters can determine how large a sample they should collect in order to achieve the desired margin of error

For a mean, $n = \left(\frac{z_0 \sigma}{E}\right)^2$. A preexisting estimate of σ is needed.

❶ Estimate the sample size needed to achieve a specified margin of error for a confidence interval for a mean.

1. Identify the critical value z_0 (see ❶ in 8-A).
2. If σ is unknown, use $\sigma \approx s$.
3. Calculate n.

❶ Tanner wants to find the average speed of drivers of drivers on Navarra Drive. Given $s = 4.5$ mph from a previous survey, how many cars should he record to have a margin of error of 1.5 mph for a 98% confidence interval?

1. $z_0 = 2.33$ for $c = .98$
2. $\sigma \approx s = 4.5$
3. $n \approx \left(2.33 \bullet \frac{4.5}{1.5}\right)^2 = 49$

For a proportion, $n = pq\left(\frac{z_0}{E}\right)^2$. If no there is no preexisting estimate of p, $p = \frac{1}{2}$ can be used because that will result in the highest (safest) estimate.

❷ Estimate the sample size needed to achieve a specified margin of error for a confidence interval for a proportion.

1. Identify the critical value z_0 (see ❶ in 8-A).
2. Use $pq \approx \hat{p}\hat{q}$. If there is no preexisting estimate $\hat{p}$, use $pq \le \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$.
3. Calculate n.

❷ Sheena wants to find out what percentage of Santa Cruz county adults were born here. How many people should she survey to have an 8% margin of error for a 95% confidence interval?

1. $z_0 = 1.96$ for $c = .95$
2. p is unknown and there is no existing estimate $\hat{p}$, so use $pq \le \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$.
3. $n \le \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1.96}{.08}\right)^2 \approx 150$