Chapter 16 - Confidence Intervals for Proportions

16.1, 16.2, & 16.3 A Confidence Interval

Confidence Interval is an interval computed from the sample data (statistic) that has a probability, C, of producing an interval that contains the true value of the population parameter.

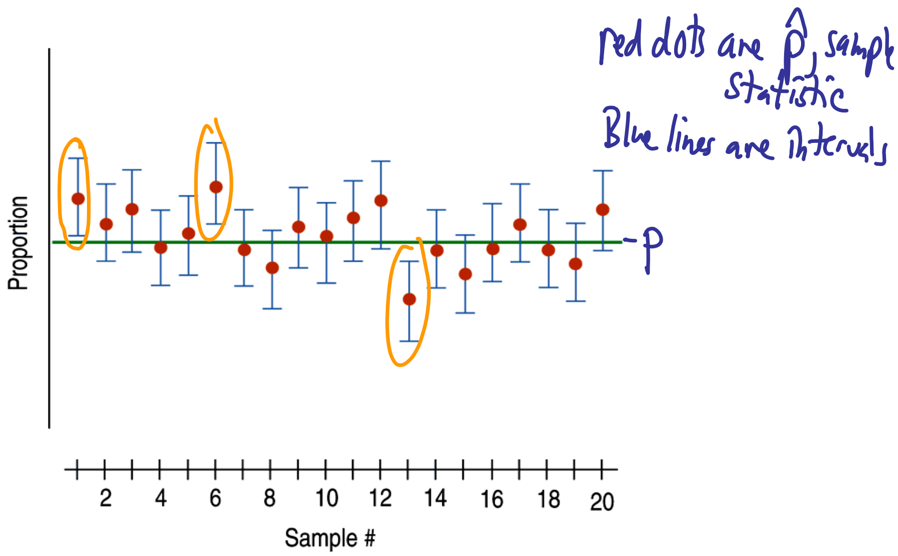
(We have been finding intervals when applying the 68-95-99.7% Rule; however, those were approximations.)

C = Confidence Level

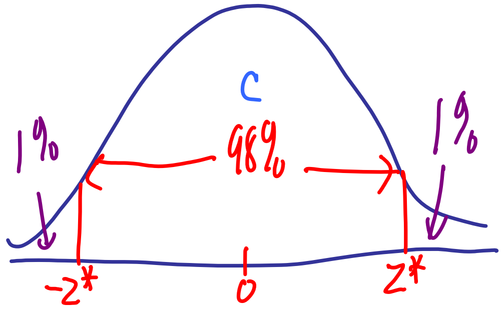
Interval = Estimate ± Margin of Error
(Estimate is the sample statistic)

95% confidence interval has 95% chance of including the parameter or in the long run 95% of the intervals found will contain the population parameter.

p.432



In Chapter 15 we learned that if the sample size, n, is large along with other assumptions that the sampling distribution for $\hat{p}$ is approximately normal with mean, $\mu_{\hat{p}} = p$, and a standard deviation, $\sigma_{\hat{p}} = \sqrt{\dfrac{pq}{n}}$.



InvNorm(.99) ⇒ $z^* = 2.326$

InvNorm(.975) ⇒ $z^* = 1.95996$

Critical Values

C	90%	95%	98%	99%
z^*	1.645	1.960	2.326	2.576

Confidence Interval for Population Proportion

$$\underbrace{\hat{p}}_{\text{Estimate}} \pm \underbrace{z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}}_{\text{Margin of Error}}$$

Remember that $SE_{\hat{p}} = \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$ (The Standard Error) which we use since p is not known. Margin of Error is $ME = z^* SE_{\hat{p}} = z^* \sqrt{\dfrac{\hat{p}\hat{q}}{n}}$.

16.4 Assumptions and Conditions

As in Chapter 15 Assumptions must be met, but once again we will check conditions instead.

- Independence Assumption: The data values are assumed to be independent from each other. We check three conditions to decide whether independence is reasonable.
  - Plausible Independence Condition: Is there any reason to believe that the data values somehow affect each other? This condition depends on your knowledge of the situation—you can't check it with data.
  - Randomization Condition: Were the data sampled at random or generated from a properly randomized experiment? Proper randomization can help ensure independence.
  - 10% Condition: Is the sample size no more than 10% of the population?

- Sample Size Assumption: The sample needs to be large enough for us to be able to use the CLT.
  - Success/Failure Condition: We must expect at least 10 "successes" and at least 10 "failures."

$n\hat{p} \geq 10$ & $n\hat{q} \geq 10$

Examples

p. 447: 26

population is 200000

$n = 1000$ $\hat{p} = \frac{123}{1000} = .123$ $\hat{q} = 1 - .123 = .877$

a) Random, Should be independent.

10% cond: $n = 1000$ is less than 10% of mailing list (200000)

S/F cond: $n\hat{p} = 1000 \times .123 = 123 \geq 10$

$n\hat{q} = 1000 \times .877 = 877 \geq 10$

Conditions are met

Normal Distribution can be used.

$C = 90\%$ or $.9 \Rightarrow z^* = 1.645$

$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = .123 \pm 1.645 \sqrt{\frac{.123 \times .877}{1000}} = .1059$ to $.1401$

or 10.59% to 14.01%

b) We are 90% confident that the population parameter, p, is between 10.59% to 14.01%.

c) 90% of similarly constructed intervals will contain p.

d) Company should do mailing since interval is well above the 5%.

p. 448: 30

Newspaper: 53% of 1200 ($\hat{p}$, n)

StatsClass: 54% of 450

$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

a) Stats Class will have larger Margin of Error since sample is smaller.

b) $C = 95\% \Rightarrow z^* = 1.96$

$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Newspaper: $.53 \pm 1.96 \sqrt{\frac{.53 \times .47}{1200}} = .502$ to $.558$

or 50.2% to 55.8%

Stats Class: $.54 \pm 1.96 \sqrt{\frac{.54 \times .46}{450}} = .494$ to $.586$

49.4% to 58.6%

c) Stats Class will say race is too close to call since part of interval is below 50%

Choosing a Sample Size

If the Margin of Error is too big then the Confidence Interval will not be of much use. One way to decrease the Margin of Error is to decrease the Confidence Level. Rarely have Confidence Intervals with Confidence Levels below 80%. Confidence Levels of 95% and 99% will be more common.

How else can the Margin of Error, $ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$, be decreased? How can the variability of the sample proportion be lessened?

To find how large a sample size is needed to obtain a certain Margin of Error one of the following formulas is used.

- If a previous proportion is known then that can be used for p^* along with the formula: $n = \frac{(z^*)^2 p^* q^*}{ME^2}$ p^* is a previous proportion
- If a previous proportion is not known then the worst p^* could be is 0.5 and the above formula becomes $n = \left(\frac{z^* \cdot 0.5}{ME}\right)^2$

(To be safe with both formulas ROUND UP to the next integer!)

Examples (see p. 438-9)

p. 449, #42 $C = .98 \Rightarrow z^* = 2.326$

$p^* = .22$ $q^* = 1 - .22 = .78$ $ME = 4\% = .04$

$$n = \frac{z^{*2} p^* q^*}{ME^2} = \frac{2.326^2 \times .22 \times .78}{.04^2} = 580.25$$

Round up and use $n = 581$

#40 $C = .99 \Rightarrow z^* \approx 2.576$ $ME = 4\% = .04$

$p^* = ?$

$$n = \left(\frac{z^* \times .5}{ME}\right)^2 = \left(\frac{2.576 \times .5}{.04}\right)^2 = 1036.84$$

use $n = 1037$