§7.1 Estimating a Proportion

Goal: Create a confidence Interval for a population proportion

1. What is a CI?
2. Steps to Make a CI.
3. Interpret CI.
4. Theory
5. How Large a Sample do we Need? Find n.

Requirements before Making a Confidence Interval for a proportion

(1) Good Sample - SRS
- Simple Random Sample
- Statistically Reliable Sample (No Bias)

(2) Binomial with $np > 5$ and $nq > 5$
- n is large enough so Normal Approximates the Binomial.
- i) Two Outcomes
- ii) Independent
- iii) P = Same for every trial
- iv) n = fixed Number of trial

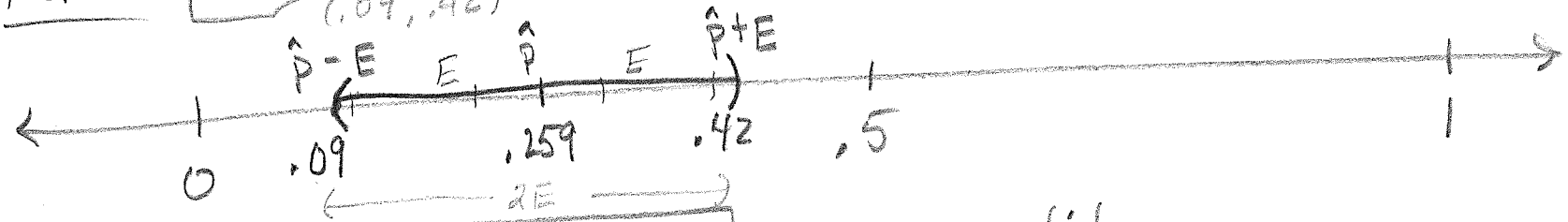
(3) $np > 5$ & $nq > 5$

$$\hat{P} = \frac{x}{n} = \frac{7}{27} = .259$$

(1) What is a Confidence Interval?

Method 1 $(.09396, .42456)$ This is an interval on x-axis

$(.09, .42)$



Method 2 $.09 < p < .42$ Inequality

Method 3 the population proportion is between .09 and .42

Method 4 $\hat{p} \pm E = .259 \pm .165$

$E = \frac{UB - LB}{2} = .165$

$SE = \frac{E}{2} = .0825$

$\hat{p} = \frac{x}{n} = \frac{7}{27}$ = Sample Proportion = .259

$\hat{p} = \frac{UB + LB}{2} = \frac{.42456 + .09396}{2} = .259$

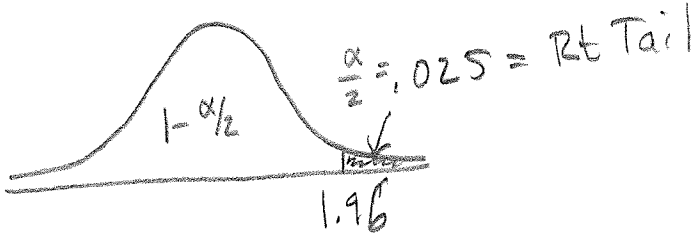
E = Margin of Error $= UB - \hat{p} = .42456 - .259 = .16556$

Class Try

Ex Find a 95% Confidence Interval for the proportion of SRHS Students who will Report that they have been drunk in the last Month. Your Sample indicates that 21 out of 110 Surveyed have been drunk.

① Requirements ✓

② PE: $\hat{p} = \frac{x}{n} = \frac{21}{110} = .191$



③ CV: $z_{\alpha/2} = \text{invnorm}(1-\alpha/2, 0, 1) = 1.96$

④ $E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96 \cdot \sqrt{\frac{.191 \cdot .809}{110}} = .0734$

⑤ CI: $\hat{p} \pm E = .191 \pm .0734$

$(\hat{p} - E, \hat{p} + E) = (.1176, .2644)$

$\hat{p} - E < p < \hat{p} + E \qquad .1176 < p < .2644$

⑥ Interpret We are 95% confident that the Proportion of SRHS Students who have been drunk is between 12% and 26%.

Quiz 7 1) 30 of 50 Stats Students are registered to vote.
Make a 98% CI for the proportion of stat students who are registered to vote.

4) If $E = .03$, find n.

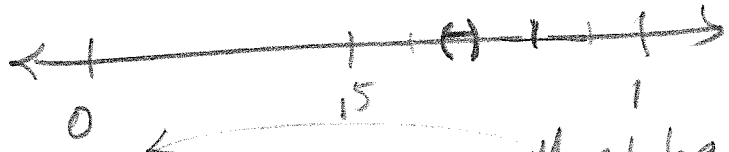
Ex The PD reports that 67% of voters plan to vote in the presidential Election, with a margin of Error of ±3%.

a) Write this CI in 3 ways

$\hat{p} = .67 \quad E = .03$

(1) $\hat{p} \pm E = .67 \pm .03$

(2) $(.67-.03, .67+.03) = (.64, .70)$

Graph 

(3) $.64 < p < .70$ Must be p Not

b) The population prop will Not differ from .67 by more than .03.

§7.1b Confidence Intervals Proportions

1-PropZTest x = # of Successes n = Sample Size CL = .95
C-Level = .95

STAT ▸ TESTS ▽

(4) Finding Sample Size n.

How large does the Sample Need to be So that E, The Margin of Error is less than .07?

x = 7 had power shut off out of 27 Students.

Gives $\hat{p} = \frac{7}{27} = .259$.

(2) Steps for Making a Confidence Interval

(1) Check Requirements $np > 5$, $n\hat{q} > 5$ $np = 7 = x$ $nq = 20$ ✓

(2) PE: $\hat{p} = \frac{x}{n} = \frac{\#\text{Successes}}{\text{Sample Size}} = \frac{7}{27} = .259 = \hat{p}$ $n = 27$

↑ Point Estimate

PE $\hat{p} = .259$

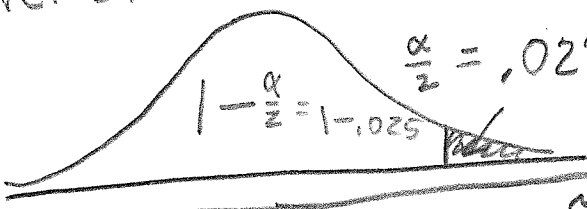
(3) CV: $Z_{\alpha/2} = \text{invnorm}(1 - \frac{\alpha}{2}, 0, 1) = 1.96 \approx 2$

Critical Value

$\alpha = 1 - CL$, CL = Confidence level of Interval

$\alpha = 1 - .95 = .05$, $CL = 95\% = .95$

$\frac{\alpha}{2} = .025$ = Area in Rt Tail



CV $Z_{\alpha/2} = 1.96$

$\hat{q} = 1 - \hat{p} = 1 - .259 = .741$

(4) $E = Z_{\alpha/2} \cdot SE = Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} = 1.96 \cdot \sqrt{\frac{.259 \cdot .741}{27}}$

$E = .165$ = CV · SE $SE = \sqrt{\frac{\hat{p}\hat{q}}{n}}$

(5) CI: $\hat{p} \pm E = .259 \pm .165$

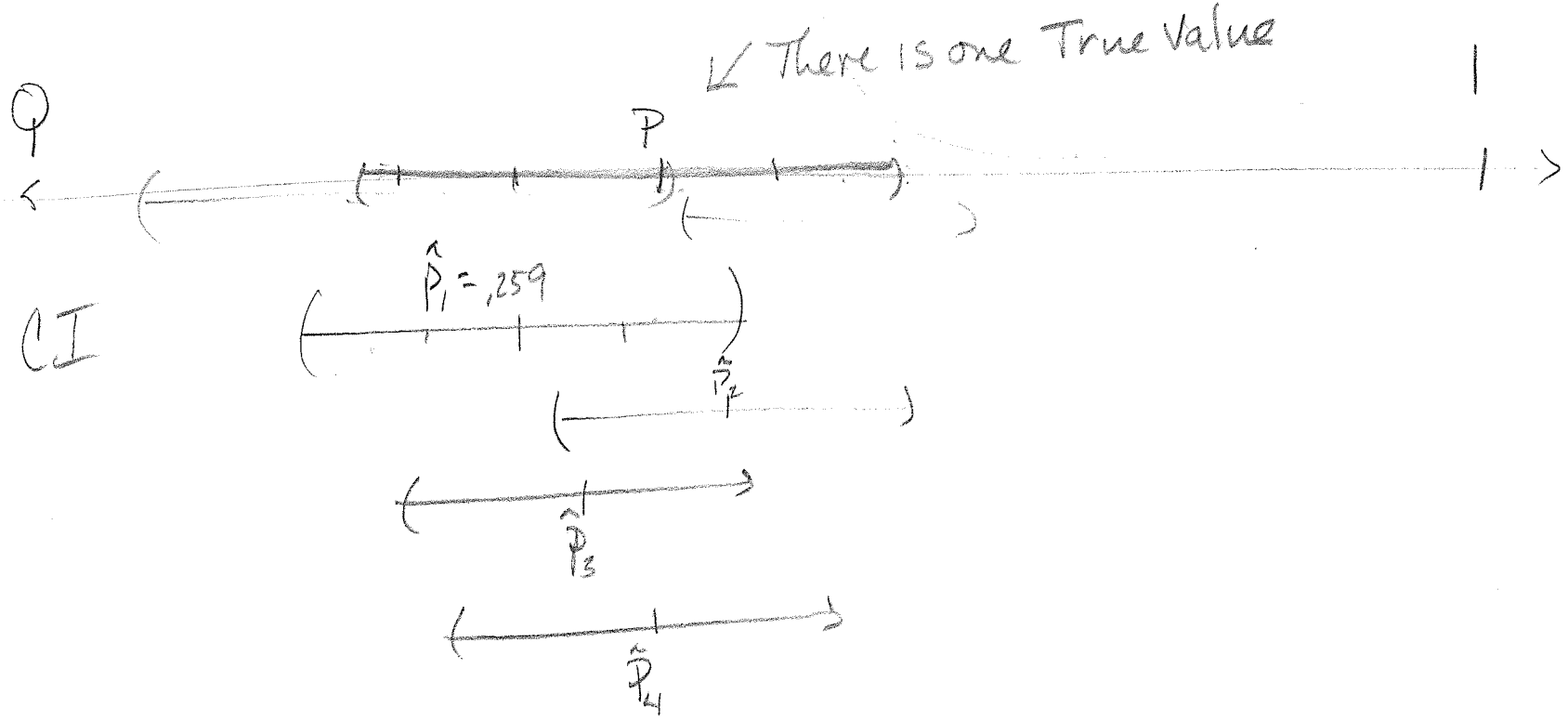
$(LB, UB) = (.259 - .165, .259 + .165) = (.094, .424)$

$.094 < p < .424$

$\hat{p} - E < p < \hat{p} + E$

(6) (3) Interpret Confidence Interval

We are 95% confident that the True proportion of all SRJC Students (p) is between .094 and .425

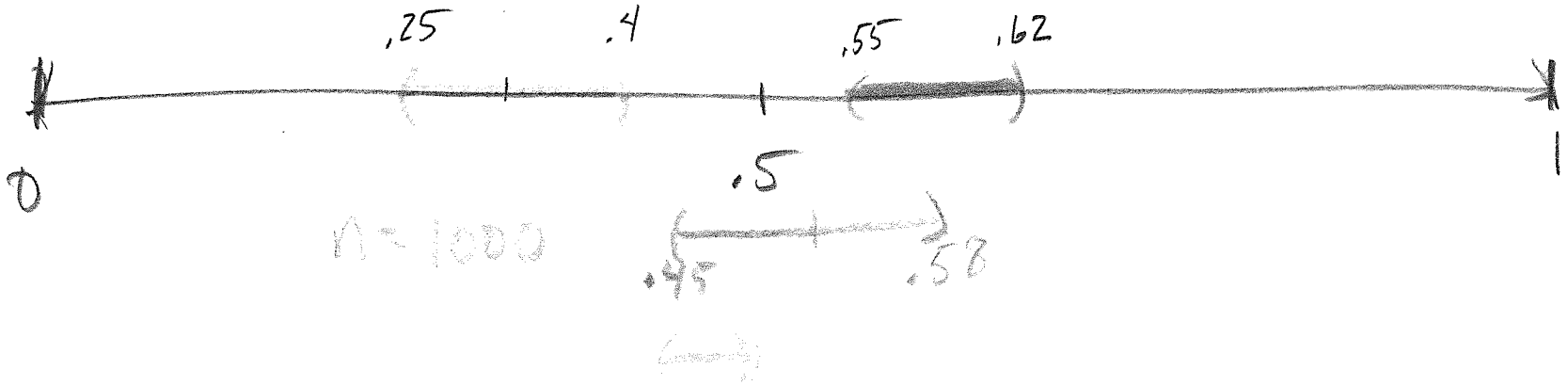


Election

Hillary

$\hat{p}$ = proportion who support Trump

Trump



n = 1000

.45 .58

Calculator

Stat ▶▶ TESTS

A: 1-PropZINT

X: 21

n: 110

C-Level: .95

Calculate

(5) Find Sample Size = n

$$E = Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\frac{E}{Z_{\alpha/2}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\frac{E^2}{Z_{\alpha/2}^2} = \frac{\hat{p}\hat{q}}{n}$$

$$E^2 n = Z_{\alpha/2}^2 \hat{p}\hat{q}$$

If $\hat{p}$ is known

$$\boxed{n = \frac{Z_{\alpha/2}^2 \cdot \hat{p}\hat{q}}{E^2}} \quad \text{or} \quad n = \frac{Z_{\alpha/2}^2 \cdot .25}{E^2}$$

Do the Algebra Solve for n.

$$E = Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$$

Square both sides

$$n \cdot E^2 = Z_{\alpha/2}^2 \cdot \frac{\hat{p}\,\hat{q}}{n} \cdot n$$

$$n E^2 = Z_{\alpha/2}^2 \cdot \hat{p}\,\hat{q}$$

$$\boxed{n = \frac{Z_{\alpha/2}^2 \cdot \hat{p}\cdot\hat{q}}{E^2}}$$

If $\hat{p}$ an estimate of p is given

Find n, $E = .07$

$CL = .95$ $\alpha = 1 - CL = 1 - .95 = .05$

If $\hat{p} = .259$ is an estimate of p

$Z_{\alpha/2} = \text{invnorm}(1 - \alpha/2, 0, 1)$

$= \text{invnorm}(.975, 0, 1) = 1.96$

$\hat{q} = 1 - \hat{p} = 1 - .259 = .741$

$n = (1.96^2 \cdot .259 \cdot .741) / (.07)^2 = 150.46$

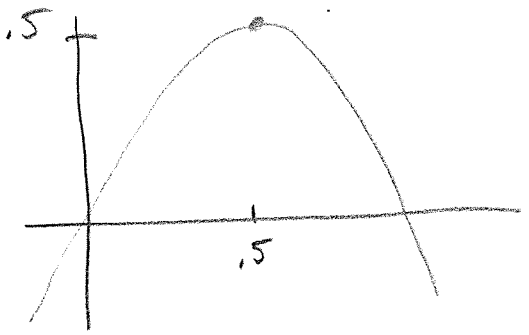
150 is Not enough

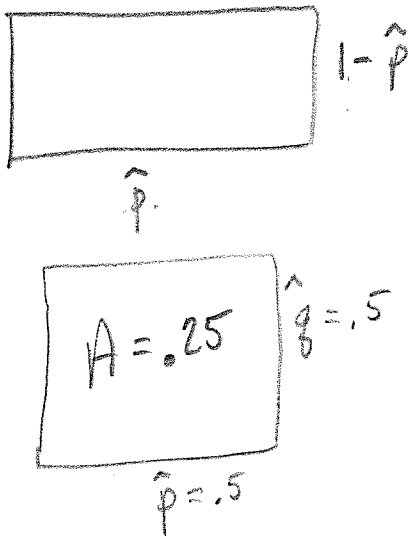
Round Up on n $\boxed{n = 151 \text{ people}}$

$$\hat{p} \cdot \hat{q} = \hat{p} \cdot (1 - \hat{p}) \leq .25$$

$$= \hat{p} - \hat{p}^2 = A$$

$$A \leq .25$$





If $\hat{p}$ is unknown then

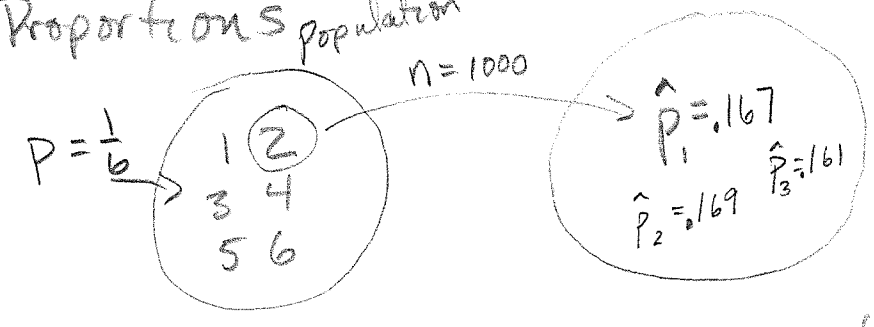
$$n > \frac{Z_{\alpha/2}^2 \cdot .25}{E^2} = \frac{1.96^2 \cdot .25}{(.07)^2} = 196 \text{ people}$$

Theory Consider Central Limit Theorem for Proportions

(1) Normal Dist of Sample Proportions

(2) $\mu_{\hat{p}} = p$

(3) $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$



Ex Toss dice 1000 times, Count the Number of 2's, Calculate the proportion of 2's

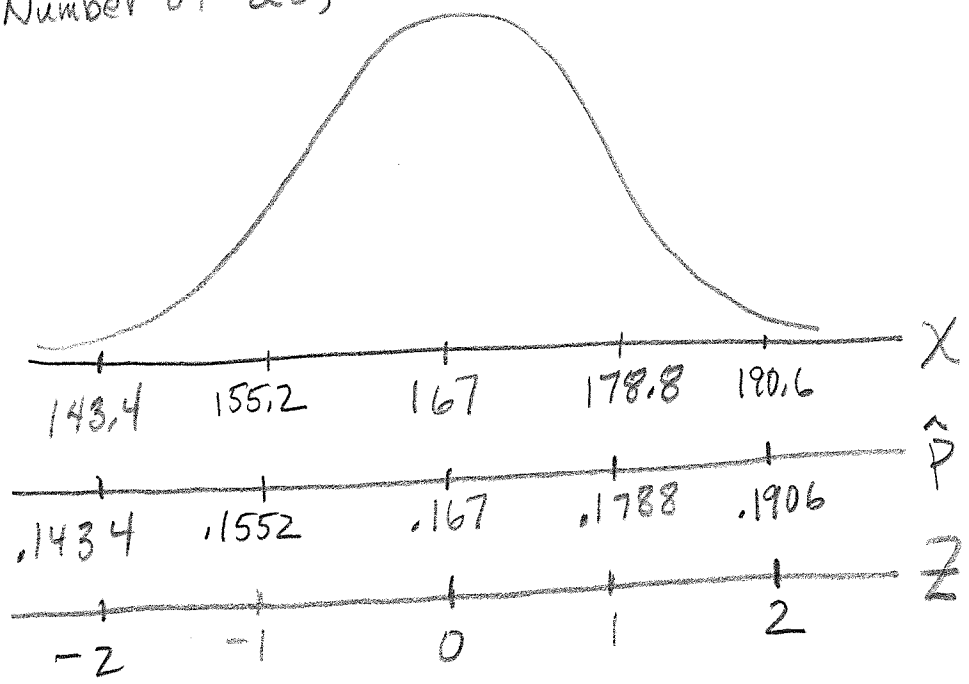
$\mu = 167 = np$

$\sigma = \sqrt{npq} = \sqrt{1000 \cdot 1/6 \cdot 5/6} = 11.8$

$\mu_{\hat{p}} = .167 = \frac{1}{6}$

$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{.167 \cdot 5/6}{1000}} = .0118 = SE$

$Z = 0$



Quiz 8 7.2 Math 15 Jones Name_____

A statistics professor asked her students whether or not they were registered to vote. In a sample of 50 of her students (randomly sampled from her 700 students), 30 said they were registered to vote.

1) (20 Points) Find a 98% confidence interval for the true proportion of the professor's students who were registered to vote. (Make sure to check any necessary conditions and to state a conclusion in the context of the problem.)
   a) What is the sample and the population in this problem.

   b) What point estimate of the population proportion p does this survey give?_____

   c) What is the level of confidence for this interval? _____ Find α_____
   d) What is the critical value use it's symbol
   e) Find the margin of error?
   Show formula and all values you use to find this by hand. CV:_____ E=_____

   f) Find the confidence interval. $.438 < p < .761$

2) Assuming that this class is representative of the population of all statistics studetents. What is the probability that the true proportion of the statistics students who were registered to vote is in your confidence interval?

3) According to a Gallup poll, about 73% of 18- to 29-year-olds said that they were registered to vote. Does the 73% figure from Gallup seem reasonable for the professor's students? Explain.

Yes .73 is in (.438, .761) so it is possible that her students are regestered at about the same rate as the National average. Her $\hat{p} = .6$ is less But Not Statistically low. A larger Sample might change this conclusion

4) If the professor only knew the information from the Gallup poll and wanted to estimate the percentage of her students who were registered to vote to within ±3% with 98% confidence, how many students should she sample?

CL=.98 ↑E=.03 ↑find n use $\hat{p}$=.73

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{E^2}$$

$$n = \frac{(2.33)^2 \cdot .73 \cdot .27}{(.03)^2} = 1189$$

n=1448 $\hat{p}$=.6
1448 $z_{\alpha/2}$=2.33

n=1185 z=2.326
1189 z=2.33

The Importance of Sample Size

we learn that the population proportion of college graduates in the US is $p = 0.275$, and Figure 3.2 shows the sampling distribution for the sample proportion of college graduates when repeatedly taking samples of size $n = 200$ from the population. How does this distribution change for other sample sizes? Figure 3.3 shows the distributions of sample proportions for many (simulated) random samples of size $n = 50$, $n = 200$, and $n = 1000$. Discuss the effect of sample size on the center and variability of the distributions.