Unit 28: Inference for Proportions

PREREQUISITES

Students need the background on *z*-procedures covered in Unit 24, Confidence Intervals, and Unit 25, Tests of Significance.

ADDITIONAL TOPIC COVERAGE

Additional coverage of inference for proportions can be found in *The Basic Practice of Statistics*, Chapter 20, Inference about a Population Proportion, and Chapter 21, Comparing Two Proportions.

ACTIVITY DESCRIPTION

In this activity, students revisit the data they collected for Unit 21's activity. However, here the focus is on using sample proportions to estimate population proportions. The population is children born to brown-eyed parents, each of whom has a recessive gene for blue eyes. The characteristic of interest is blue eyes. In this case, we know the population proportion of blue-eyed children is $p = 0.25$. In this activity, students use the simulated data collected in Unit 21's activity to make a number of estimates of p using samples of different sizes.

MATERIALS

Completed Table 21.1, simulated data for Unit 21's activity.

In Unit 21, each sample represented a family. Here each size-4 sample can be thought of as a simple random sample from the population of all children born to brown-eyed parents with the recessive gene for blue eyes.

In question 2, students calculate the sample proportions and use them as estimates of the population proportion of blue-eyed children born to brown-eyed parents with recessive genes for blue eyes. Since the data contain 30 size-4 samples, students will have 30 values of $\hat{p}$. Next, they make a histogram of their sample proportions to get a sense of the shape of the sampling distribution of the sample proportion. (The concept of a sampling distribution of a sample statistic is difficult for students to grasp.) Students' histograms won't closely resemble a normal density curve because the sample size is too small. However, this question can serve as a good springboard to a discussion of the sampling distribution of the sample proportion when the sample size is large.

In questions 4, 5, and 6, students combine data from the first 10 samples, first 20 samples, and then all 30 samples. They use these data to construct three 95% confidence intervals for *p*. This gives students an opportunity to see that the size of the margin of error shrinks as the sample size increases.

Question 8 involves a little algebra. (Tell students to skip this problem if you don't want to deal with the algebra. In that case, don't assign Review Question 4(d).) In question 8, students find the sample size that guarantees that the margin of error will be less than 0.5. If *E* represents the margin of error, then the sample size needed to guarantee a margin of error less than *E* in a 95% confidence interval is:

$$n = \left(\frac{1.96}{E}\right)^2 \hat{p}(1-\hat{p})$$

The problem is that $\hat{p}$ varies and is not known until after the sample is collected. Since $\hat{p}(1-\hat{p})$ has a maximum value of 0.25 when $\hat{p}$ = 0.5, we can substitute this value into the formula for *n*, which gives a conservative value for *n*. The result is the following formula for determining the sample size needed to guarantee that the margin of error is less than *E* regardless of the value of $\hat{p}$:

$$n = 0.25\left(\frac{1.96}{E}\right)^2$$

Finally, in the world of simulation – unlike in the real world – we know that p = 0.25 for the sample data collected for this activity. Question 9 gives students an opportunity to see if their confidence intervals got it right – in other words, contained the true population proportion of 0.25.

THE VIDEO SOLUTIONS

1. Using information from samples to make inferences about population proportions.

2. People were able to solve problems creatively and come up with new ideas when they felt most motivated and excited about their work.

3. Progress is paramount to people feeling positive and highly motivated about their work.

4. We would expect 20% of the managers to select progress.

5. A z-test statistic was used.

6. The values in the confidence interval all fell below 20%.

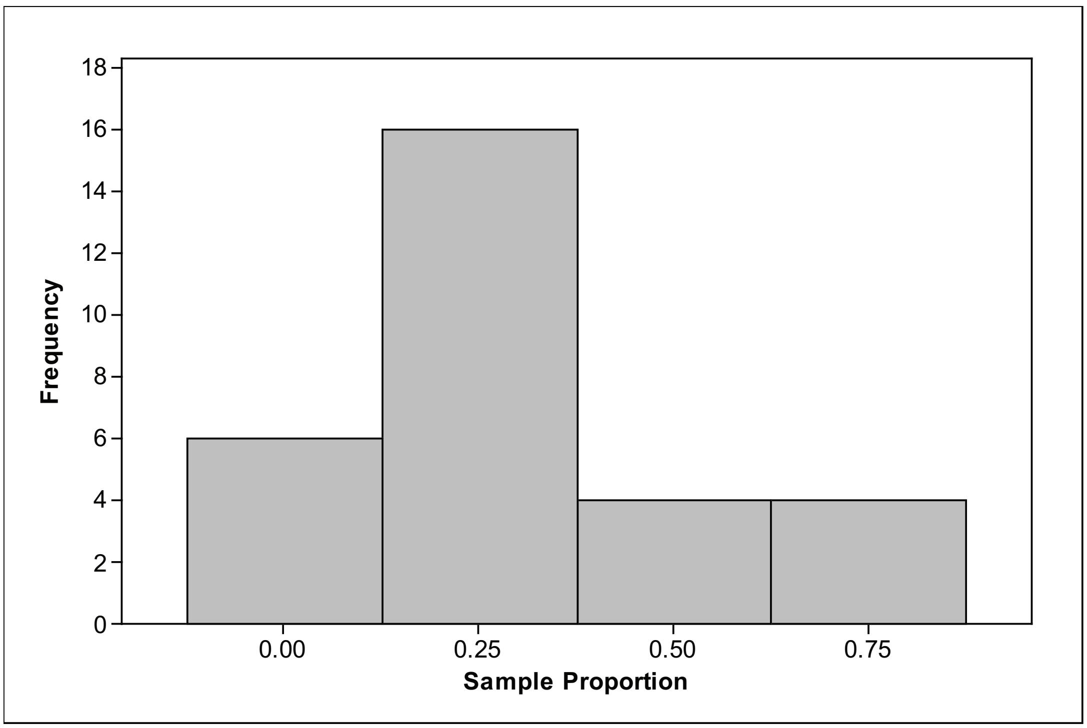
UNIT ACTIVITY:
PROPORTIONS OF BLUE EYES SOLUTIONS

1. See sample answer to 3(b). The sample data in the solution to 3(b) will be used for sample answers to this activity.

2. a. See sample answer to 3(b).

b. Sample answer: The smallest sample proportion was 0 and the largest was 0.75.

c. The histogram does not appear to have a normal distribution. There were only four observed values for the sample proportion. However, there is one peak at value 0.25.



3. a. See sample answer to 3(b).
(On next page...)

3. b.

Sample Number	Number of Blue-Eyed Children *n = 4*	Sample Proportion Blue-Eyed Children *n* = 4	Running Total Number of Children	Running Total Number of Blue-Eyed Children
1	1	0.25	4	1
2	0	0	8	1
3	2	0.5	12	3
4	1	0.25	16	4
5	0	0	20	4
6	1	0.25	24	5
7	1	0.25	28	6
8	3	0.75	32	9
9	1	0.25	36	10
10	1	0.25	40	11
11	1	0.25	44	12
12	3	0.75	48	15
13	0	0	52	15
14	3	0.75	56	18
15	1	0.25	60	19
16	1	0.25	64	20
17	2	0.5	68	22
18	1	0.25	72	23
19	2	0.5	76	25
20	1	0.25	80	26
21	1	0.25	84	27
22	1	0.25	88	28
23	1	0.25	92	29
24	2	0.5	96	31
25	1	0.25	100	32
26	0	0	104	32
27	3	0.75	108	35
28	0	0	112	35
29	0	0	116	35
30	1	0.25	120	36

4. a. Sample answer based on data shown in 3(b): $\hat{p} = 11/40 = 0.275$

b. $0.275 \pm (1.96)\sqrt{\dfrac{(0.275)(1-0.275)}{40}} = 0.275 \pm 0.138$; from 0.137 to 0.413

c. The margin of error is 0.138.

5. a. Sample answer based on data shown in 3(b): $\hat{p} = 26/80 = 0.325$

b. $0.325 \pm (1.96)\sqrt{\dfrac{(0.325)(1-0.325)}{80}} = 0.325 \pm 0.103$; from 0.222 to 0.428

c. The margin of error is 0.103.

6. a. Sample answer based on data shown in 3(b): $\hat{p} = 36/120 = 0.3$

b. $0.300 \pm (1.96)\sqrt{\dfrac{(0.300)(1-0.300)}{120}} = 0.325 \pm 0.082$; from 0.218 to 0.382

c. The margin of error is 0.082.

7. The margin of error decreased as the sample size increased.

8. a. $n = (1536.64)(\hat{p})(1-\hat{p})$

b.

$\hat{p}$	$\hat{p}(1-\hat{p})$
0.1	0.09
0.2	0.16
0.3	0.21
0.4	0.24
0.5	0.25
0.6	0.24
0.7	0.21
0.8	0.16
0.9	0.09

c. n = (1536.64)(0.25) = 384.16. So, a sample size of at least n = 385 will guarantee that the margin of error is less than 0.05.

9. Sample answer: All of our confidence intervals contained the value 0.25 and gave correct results.

EXERCISE SOLUTIONS

1. a. $\hat{p} = \dfrac{2098}{2454} \approx 0.855$, or around 85.5%

b. $0.855 \pm 1.96\sqrt{\dfrac{(0.855)(1-0.855)}{2454}}$

0.855 ± (1.96)(0.007) = 0.855 ± 0.014; between 0.841 and 0.869 or between 84.1% and 86.9%

c. A 90% confidence interval would be narrower. Using a less reliable procedure gives more precision. Instead of using $z^* = 1.96$, we would use $z^* = 1.654$, which would make for a smaller margin of error.

2. a. $z = \dfrac{0.466 - 0.50}{\sqrt{\dfrac{(0.50)(0.50)}{440}}} \approx -1.426$; p = 2(0.07693) ≈ 0.154.

There is insufficient evidence to reject the null hypothesis.

b. $0.466 \pm 1.96\sqrt{\dfrac{(0.466)(1-0.466)}{440}} \approx 0.466 \pm 0.047$; from 0.419 to 0.513, or 41.9% to 51.3%

c. Since 0.5 is in the confidence interval, we cannot reject the null hypothesis. This decision is the same as the decision made using the *z*-test statistic in (a). (It should be noted that in the case of proportions, results from significance tests and confidence intervals do not always lead to the same results.)

3. a. The population is all American households in which a member of the household owns or uses a computer at home.

b. $H_0 : p = 0.90$
$H_a : p > 0.90$

c. $\hat{p} = 1816/1910 \approx 0.951$

$$z = \frac{0.951 - 0.90}{\sqrt{\dfrac{(0.90)(0.10)}{1910}}} \approx 7.43$$

The *p*-value is essentially 0.

Conclusion: Reject the null hypothesis and conclude the population percentage is greater than 90%.

4. $0.951 \pm 1.96\sqrt{\frac{(.951)(1-0.951)}{1910}} \approx 0.951 \pm 0.010$; between 0.941 and 0.961

Expressed as a percentage: between 94.1% and 96.1%

REVIEW QUESTIONS SOLUTIONS

1. a. $\hat{p} = \dfrac{2998}{5462} \approx 0.549$, or around 54.9%

b. $H_0: p = 0.5$
$H_a: p > 0.5$

c. $z = \dfrac{0.549 - 0.50}{\sqrt{\dfrac{(0.50)(1-0.50)}{5462}}} \approx 7.24$; $p \approx 0.000$.

Hence, we can conclude that the physical education teacher is correct.

2. a. The result of the poll was that between 45.4% and 50.6% (that's 48% ± 2.6%) of the voters intended to vote for Obama. This result was obtained by a method that gives correct results 95% of the time when used repeatedly.

b. The two confidence intervals, Obama's 45.4% to 50.6% and Romney's 44.4% to 49.6%, overlap. If Obama's actual percentage is at the low end of his confidence interval and Romney's actual percentage is at the high end of his confidence interval, then Romney would win. On the other hand, if Obama's actual percentage is at the high end of his confidence interval and Romney's is at the low end of his interval, then Obama would win.

3. a. $\hat{p} = \dfrac{255}{296} \approx 0.861$

$$0.861 \pm 1.96\sqrt{\frac{(0.861)(1-0.861)}{296}} \approx 0.861 \pm 0.039;$$

between 0.822 and 0.900, or between 82.2% and 90.0%.

b. $\hat{p} = \dfrac{234}{298} \approx 0.785$

$$0.785 \pm 1.96\sqrt{\frac{(0.785)(1-0.785)}{298}} \approx 0.785 \pm 0.047;$$

between 0.738 and 0.832, or between 73.8% and 83.2%.

c. $\hat{p} = \frac{174}{297} \approx 0.586$

$0.586 \pm 1.96\sqrt{\frac{(0.586)(1-0.586)}{297}} \approx 0.586 \pm 0.056$; between 0.530 and 0.642.

4. a. $0.63 \pm 1.95\sqrt{\frac{0.63(1-0.63)}{1000}} \approx 0.63 \pm 0.03$; from 0.60 to 0.66, or from 60% to 66%.

The margin of error, to two decimals, is 0.03 or $\pm 3\%$.

(It was 2.99%, which we rounded to 3%.)

b. They match.

c. Corresponding to 50%: margin of error is $1.96\sqrt{\frac{(.5)(.5)}{1000}} \approx 0.031$; 3.1%.

So, the margin of error would be 3% only if we round to the nearest whole percent.

Corresponding to 80%: margin of error is $1.96\sqrt{\frac{(.8)(.2)}{1000}} \approx 0.025$; 2.5%

d. $n = \left(\frac{1.96}{.03}\right)^2 (0.25) \approx 1067.11.$

In order to guarantee that the margin of error was less than 3%, a sample size of at least 1,068 should have been used.