6. A survey given to a random sample of students at a university included a question about which of two well known comedy shows, S or F, students preferred. The students were asked the question, “Do you prefer S or F?” The responses are shown below.

Preference		
S	F	Total
185	139	324

(a) Based on the results of this survey, construct and interpret a 95% confidence interval for the proportion of students in the population who would respond S to the question, “Do you prefer S or F?”
(b) What is the meaning of “95% confidence” in part (a)?
(c) In a follow-up survey, a separate group of randomly selected students was asked “Do you prefer F or S?” The responses are shown below.

Preference		
S	F	Total
68	88	156

Based on these two surveys, is there evidence that the stated preference depends on the order in which the comedy shows were listed in the survey question? Justify you answer.

(d) Suppose the test in part (c) indicates that the order in which the shows were listed does make a difference. Is the pooled value (185 + 68) / (324 + 156) = 0.527 a reasonable estimate for the proportion of students at the university who would respond S? If so, justify you answer. If not, what would be a more reasonable estimate? Explain why.

AP-Stats-2002-Q6 (Answer)
Chapter 9 – Unit 1 and 2 – pgs. 475-488
Chapter 11 – Unit 3 – pg. 611

Part (a)

p = proportion of students at this university who would respond S to the question, “Do you prefer S or F?”

Large sample confidence interval for a population proportion.

$$\text{p-hat} \pm 1.96 \sqrt{\frac{\text{p-hat}(1\text{-p-hat})}{n}}$$

State and Check Assumptions and Conditions:

Simple random sample. Need large sample with np-hat ≥5 and n(1-p-hat) ≥ 5. Here,

np-hat = 185 n(1-p-hat) = 139 are both greater than 5 (or 10)

or $\text{p-hat} \pm 3 \sqrt{\frac{\text{p-hat}(1\text{- p-hat})}{n}}$ is entirely in the interval (0.1).

Calculations:

$$\text{p-hat} \pm 1.96 \sqrt{\frac{\text{p-hat}(1 - \text{p-hat})}{n}} = 0.571 \pm 1.96 \sqrt{\frac{(.571)(.429)}{324}} = 0.571 \pm 0.054 = (0.571, 0.625)$$

Calculator: (0.5171, 0.62488)

Interpretation:

Based on this sample, we can be 95% confident that the proportion of students at this university who would respond S to the question, “Do you prefer S or F?” is between 0.517 and 0.625.

Part (b)

In repeated sampling, 95% of the intervals produced using this method will contain the proportion of students at this university who would respond S to the question “Do you prefer S or F?”

Part (c)

Hypothesis Test – Two Proportion Z-test

Hypothesis:

Ho: $p_1 - p_2 = 0$
Ha: $p_1 - p_2 \neq 0$

Where
p_1 = proportion of students at this university who would respond S with the original question wording
p_2 = proportion of students at this university who would respond S with the revised question wording

Name Test and State and Check Assumptions and Conditions:

Two sample z-test for proportions

Large samples: n_1p-hat$_1$ = 185; n_1(1 - p-hat$_1$) = 139; n_2p-hat$_2$ = 68; n_2(1 - p-hat$_2$) = 88
All are greater than 10, so the sample sizes are large enough.

Calculations:

For two sample proportion z-test:

$$\text{p-hat}_1 = \frac{185}{324} = 0.571 \qquad \text{p-hat}_2 = \frac{68}{156} = 0.436$$

$$\text{p-hat} = \frac{185 + 68}{324 + 156} = \frac{253}{480} = 0.527$$

$$z = \frac{\text{p-hat}_1 - \text{p-hat}_2}{\sqrt{\frac{\text{p-hat(1-p-hat)}}{n_1} + \frac{\text{p-hat(1-p-hat)}}{n_2}}} = \frac{0.571 - 0.436}{\sqrt{\frac{(0.527)(0.473)}{324} + \frac{(0.527)(0.473)}{156}}} = \frac{0.135}{0.0478} = 2.77$$

P-value = 2(.0028) = .0056 from tables

Calculator: z = 2.776554085, P-value = .0054939656

Conclusion:
Since the P-value is so small, we reject the null hypothesis that the proportions of this university's students who would respond S to the two survey questions are equal. We believe the order in which the choices are given affects the students' response.

Part (d)
If the sample sizes had been equal, it would be reasonable to combine the data from the two samples by pooling, which would be equivalent to averaging the two proportions in this case. But since the wording of the question makes a difference, and more people were asked the original version than were asked the revised version, we cannot just pool.

One reasonable approach would be to scale sample 2 up to a sample size of 324 while maintaining the same sample proportion. To do this, the number of S's would be multiplied by a factor of 2.076923. This would result in two samples of sizes 324 with 185 S's in sample 1 and141 S's in sample 2. This would result in an estimate of those who prefer S of

$$\text{p-hat} = \frac{185 + 141}{648} = \frac{326}{648} = 0.503$$