STA101 Problem Set 5 - KEY

Summer I, 2021, Duke University

Exercises from the OpenIntro book - 100pts total

Problems include: Chapter 5 exercises 5.4, 5.16, 5.26, 5.30 and 1 additional problem

5.4 (35pts total)

(a) **(5pts for "all US adults" or equivalent wording)** The sample is from all adults in the United States, so US adults is the population under consideration.

(b) **(5pts for "fraction/proportion" in answer)** The fraction of US adults who could not cover a $400 expense without borrowing money or selling something.

(c) **(5pts)** We estimate the parameter by computing the observed value in the data:

$$\hat{p} = \frac{322}{765} = 0.421.$$

(d) **(5pts for "standard error" or "SE")** We quantify this uncertainty using the *standard error*, which may be abbreviated as SE.

(e) **(5pts)** We can compute the standard error using the SE formula and plugging in the point estimate $\hat{p} = 0.421$ for p:

$$SE = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.421(1-0.421)}{765}} = 0.0179.$$

(f) **(5pts for "surprised")** The standard error can be thought of as the standard deviation of $\hat{p}$. A value of 0.50 would be over 4 standard errors from the observed value, which would represent a very uncommon observation. The news pundit should be surprised by the data.

(g) **(5pts for minimal change or equivalent wording)** The recomputed standard error is

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.4(1-0.4)}{765}} = 0.0177$$

This value is hardly different at all, since 0.4 isn't too different from 0.421.

5.16 (10pts total)

(a) **(5pts)** $H_0 : \mu = 1100$ (The current average calorie intake is 1100 calories)
$H_A : \mu \neq 1100$ (The current average calorie intake is different than 1100 calories.)

(b) **(5pts)** $H_0 : p = 0.7$ (The fraction of Wisconsin adults who consume alcohol is 0.7)
$H_A : p \neq 0.7$ (The fraction of Wisconsin adults who consume alcohol is different from 0.7)

5.26 (20pts total)

(a) **(5pts for "Scenario I")** Scenario I is higher. Recall that a sample mean based on less data tends to be less accurate and have larger standard errors.

(b) **(5pts for "Scenario I")** Scenario I is higher. The higher the confidence level, the higher the corresponding margin of error.

(c) **(5pts for "equal")** They are equal. The sample size does not affect the calculation of the p-value for a given Z-score.

(d) **(5pts for "Scenario I")** Scenario I is higher. If the null hypothesis is harder to reject (lower α), then we are more likely to make a Type 2 Error when the alternative hypothesis is true.

5.30 (20pts)

(a) **(5pts for "True")** True.

(b) **(3pts for "False", 2pts for correction)** False. The significance level is the probability of the Type 1 Error.

(c) **(3pts for "False", 2pts for correction)** False. Failure to reject H_0 only means there wasn't sufficient evidence to reject it, not that it has been confirmed.

(d) **(5pts for "True")** True.

Problem 5 (15pts total)

(5pts for confidence interval, 5pts for interpretation, 5pts for comparison)

The z-score associated to a 90% confidence interval is 1.65, so the confidence interval is given by

$$0.52 \pm (1.65)(0.024) = 0.52 \pm 0.0396 \quad \text{or equivalently } (0.4804, 0.5596).$$

This tells us that there is a 90% chance that this confidence intervals contains the true proportion of U.S. adult Twitter users who get at least some news on Twitter.

The confidence interval would be *wider*. The reason is that the width is determined by the margin of error $z \cdot SE$. Increasing the coverage (90% to 99%) results in a higher z-score, which increases the margin of error, and therefore, the width of the confidence interval.