Biostatistics Midterm 2 Prep

Problem 1

In a report produced by the US Department of Health and Human Services it is recommended that adults should do at least 150 minutes a week of moderate-intensity, or 75 minutes a week of vigorous-intensity aerobic physical activity, or an equivalent combination of moderate- and vigorous-intensity aerobic activity. A poll of 85 randomly selected undergrads at the University of Iowa was taken to estimate the proportion of undergrads at the University of Iowa who meet these guidelines. Of the 85 polled, 36 reported that they met these physical activity guidelines.

What is your best estimate for the proportion of undergrads at the University of Iowa who meet these physical activity guidelines?

$\hat{p} = \frac{36}{85} = 0.4235294$

What is the standard error of this estimate?

$se(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.424)(0.576)}{85}} = 0.0536$

Create a 95% one-proportion z-interval confidence interval for this estimate:

- **P**
  - **parameter of interest:** Let p = proportion of undergrads at the University of Iowa who meet US HHS guidelines
- **A**
  - **check assumptions:**
    * data is randomly sampled ✓
    * we assume each sampled value (person polled) is independent ✓
    * we assume that $85 < 10\%$ of all UI undergrads ✓
    * we have 36 "successes" and 49 "failures" ✓
- **N**
  - **name the interval:** One Proportion z-interval
- **I**
  - **95% confidence interval is:**
    * $(\hat{p} - z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}) = (0.424 - (1.96)(0.0536), 0.424 + (1.96)(0.0536)) = (0.318944, 0.529056)$
- **C**
  - **state your conclusion:**
    * I am 95% confident that the interval from 0.318944 to 0.529056 contains the true proportion of University of Iowa undergrads who meet US Department of Health and Human Services physical activity guidelines.

What does the 95% confidence interval mean?

If many random samples were taken under the same conditions, 95% of the confidence intervals produced would contain the true proportion of undergrads at the University of Iowa who meet the US Department of Health and Human Service's physical activity guidelines.

Problem 2

We are given a confidence interval of (0.52, 0.65) for estimating the proportion of residents in Iowa City who eat fast-food at least once per week:

What is the point estimate for the proportion?

$$\hat{p} + moe = 0.65$$

$$\hat{p} - moe = 0.52$$

$$2\hat{p} = 0.65 + 0.52 = 1.17 \rightarrow \hat{p} = 0.585$$

What is the margin or error?

$$\hat{p} + moe = 0.65 \rightarrow moe = 0.65 - \hat{p} = 0.65 - 0.585 = 0.065$$

If this were given as a 95% confidence interval, what sample size n would have been used to produce it?

$$moe = 0.065 = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{1.96\sqrt{(0.585)(0.415)}}{\sqrt{n}} \rightarrow \sqrt{n} = \frac{1.96\sqrt{(0.585)(0.415)}}{0.065} \rightarrow$$

$$\rightarrow n = \frac{1.96^2(0.585)(0.415)}{0.065^2} = 220.7442$$

Se we need a sample of $n = 221$ to produce this confidence interval

Let's say that we are given that the sample size $n = 155$, approximately what percent confidence interval would this be?

First we need to determine z^*

$$moe = 0.065 = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \rightarrow z^* = \frac{0.065\sqrt{n}}{\sqrt{\hat{p}(1-\hat{p})}} = \frac{0.065\sqrt{155}}{\sqrt{(0.585)(0.415)}} = 1.642394$$

For a 90% confidence interval we use $z^* = 1.645$, so for the given confidence interval values and a sample size of $n = 155$, we have approximately a 90% confidence interval.

Problem 3

Assume we wish to estimate the proportion of Iowa City and Coralville residents whose favorite season is Fall. For the most conservative estimate of p, what size sample would we need in order to estimate the proportion to within one percent with 95% confidence?

The standard error for $\hat{p}$ is greatest when $\hat{p} = 0.5$. If we assume $\hat{p} = 0.5$, then

$$se(\hat{p}) = \frac{0.5}{\sqrt{n}} \to moe = 0.01 = 1.96 * \frac{0.5}{\sqrt{n}} \to \sqrt{n} = \frac{(1.96)(0.5)}{0.01} \to n = (\frac{(1.96)(0.5)}{0.01})^2 = 9604$$

So if we wanted to produce a 95% confidence interval to within one percent of our estimate we would want a sample size of 9604 people.

Problem 4

Several years worth of research suggests that Excedrin is effective at relieving migraines in about 36% of subjects. A company claims to make an herbal remedy for migraines that is just as effective as Excedrin but using only natural ingredients and no acetaminophen, aspirin or caffeine. 600 migraine sufferers were randomly prescribed to take the herbal remedy for their migraines, and of those 600 subjects, 192 reported that their migraines abated. Perform a hypothesis test to assess the company's claim.

- **P**
  - **parameter of interest:** Let p = proportion of migraine sufferers whose migraines get better after using the herbal remedy
- **H**
  - **hypothesis about parameter:**
    - Null hypothesis $H_0 : p = p_0 = 0.36$ meaning herbal remedy is just as effective as Excedrin
    - Alternative hypothesis $H_A : p < p_0 = 0.36$
- **A**
  - **check assumptions:**
    - subjects randomly assigned to study ✓
    - we assume each sampled value is independent ✓
    - we assume that $600 < 10\%$ of all migraine sufferers ✓
    - we have 192 "successes" and 408 "failures" ✓
- **N**
  - **name the test:** One Proportion z-test
- **T**
  - **find the test statistic:**

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \text{ where } \hat{p} = \frac{192}{600} = 0.32,\ p_0 = 0.36, \text{ and } n = 600 \to z = \frac{0.32 - 0.36}{\sqrt{\frac{(0.36)(0.64)}{600}}}$$

$$z = \frac{-0.04}{0.01959592} = -2.041241$$

- **O**
  - **obtain a p-value:**

$$P(Z < -2.04) \approx 0.021 \text{ so our p-value is 0.021 which we'll test against } \alpha = 0.05$$

- **M**
  - **make a decision:**
    - $0.021 < 0.05$ so we reject H_0
- **S**
  - **state conclusion**
    - Since the p-value is 0.021, which is less than the significance level of 0.05, I reject H_0 and conclude that there is sufficient evidence that the herbal remedy is less effective than Excedrin in relieving migraines.

What does the p-value mean here?

The p-value of 0.021 says that if the true population proportion of migraine sufferers whose migraines get better after using the herbal remedy is 0.36, then samples of size $n = 600$ can be expected to have an observed proportion of 0.32 or less 2.1% of the time