LAB: INTRO TO STAT ANALYSIS

Zeyi Qian

zeqian@clarku.edu
Office Hours: JC 201, Tuesday 4-5 PM & Thursday 3-4 PM

October 24, 2024

Question 1

It is generally believed that electrical problems affect about 14% of new cars. An automobile mechanic conducts diagnostic tests on 128 new cars on the lot

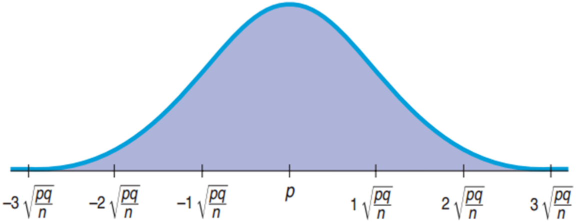
- a. Describe the sampling distribution for the sample proportion by naming the model and telling its mean and standard deviation. Justify your answer
- b. Sketch and clearly label the model using the "68%-95%-99.7%" rule
- c. What is the probability that in this parking lot less than 9% of cars have electrical problems?

Question 1: a

- Check the sampling distribution of the sample proportion is approximately normal
- Mean of the sampling distribution: $p = 0.14$
- The standard deviation $\sigma = \sqrt{\frac{p(1-P)}{n}} = 0.031$
- Thus, the sampling distribution is approximately normal with: $\hat{p} \sim N(0.14, 0.031)$

Question 1: b

- About 68% of the data falls within 1 standard deviation from the mean
- About 95% falls within 2 standard deviations
- About 99.7% falls within 3 standard deviations



Question 1: c

- $z = \frac{0.09 - 0.14}{0.031} = -1.61$
- Using a z-table
- Or using Excel "$= NORM.S.DIST(-1.61, TRUE)$"
- $P = 0.053$

Question 2

According to Gallup, about 33% of Americans polled said they frequently experience stress in their daily lives. Suppose you are in a class of 45 students

- a. Describe the sampling distribution model for the proportion of students who experience stress in the class
- b. What is the probability that no more than 12 students in the class will say that they frequently experience stress in their daily lives?
- c. If 20 students in the class said they frequently experience stress in their daily lives, would you be surprised? Explain, and use statistics to support your answer

Question 2: a

- $p = 0.33$
- $\sigma = \sqrt{\frac{p(1-P)}{n}} = -0.9$
- Thus, the sampling distribution is approximately normal with: $\hat{p}\ N(0.33, 0.07)$

Question 2: b

- $\hat{p} = 12/45 = 0.267$
- $z = \frac{0.267 - 0.33}{0.07} = -0.9$
- Using a z-table
- Or using Excel "$= NORM.S.DIST(-0.9, TRUE)$"
- $P = 0.18$

Question 2: c

- $\hat{p} = 20/45 = 0.44$
- $z = \frac{0.44-0.33}{0.07} = 1.629$
- Using a z-table
- Or using Excel "$= NORM.S.DIST(1.629, TRUE)$"
- $P = 0.94$
- $1 - 0.94 = 0.5$
- It would be somewhat surprising

Question 3

A 2011 Gallup poll found that 76% of Americans believe that high achieving high school students should be recruited to become teachers. This poll was based on a random sample of 1002 Americans

- a. Find a 90% confidence interval for the proportion of Americans who would agree with this.
- b. How many people should we interview if we want to have 95% confidence level with 3% margin of error?

Question 3: a

- Sample Proportion ($\hat{p}$):

$$\hat{p} = 0.76$$

- Sample Size (n):

$$n = 1002$$

- Standard Error (SE):

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx 0.0135$$

- Z-Score for 90% Confidence Level ($= NORM.S.INV(0.95)$):

$$z \approx 1.645$$

Question 3: a

- Margin of Error (ME):

$$ME = z \cdot SE \approx 0.0222$$

- Confidence Interval:

$$(0.76 - 0.0222, 0.76 + 0.0222) \approx (0.7378, 0.7822)$$

Question 3: b

- Z-Score for 95% Confidence Level ($= NORM.S.INV(0.975)$):

$$z \approx 1.96$$

- Desired Margin of Error (E):

$$ME = 0.03$$

- Sample Size Formula:

$$n = \left(\frac{z^2 \cdot \hat{p}(1 - \hat{p})}{ME^2} \right)$$

- Plugging Values:

$$n \approx \left(\frac{(1.96)^2 \cdot 0.76 \cdot 0.24}{(0.03)^2} \right) \approx 779$$

Question 4

Statistics from Cornell's Northeast Regional Climate Center indicate that Ithaca, New York, gets an average of 35.4" of rain each year with a standard deviation of 4.2". Assume that a Normal model applies. A Cornell University student is in Ithaca for 4 years. What is the probability that those 4 years average less than 30" of rain?

Question 4

- Average rainfall (μ): 35.4 inches
- Standard deviation (σ): 4.2 inches
- Number of years (n): 4
- Mean of the sample mean:

$$\mu_{\bar{x}} = 35.4 \text{ inches}$$

- Standard deviation of the sample mean (Standard Error):

$$\sigma_{\bar{x}} = \frac{4.2}{\sqrt{4}} = 2.1 \text{ inches}$$

- Standardizing the value of 30 inches to find the z-score:

$$z = \frac{30 - 35.4}{2.1} \approx -2.57$$

Question 4

- Finding the probability:

$$P(Z < -2.57)$$

- From the standard normal distribution table:

$$P(Z < -2.57) \approx 0.0051$$

Question 5

A father is concerned that his teenage son is watching too much television each day, since his son watches an average of 2 hours per day. His son says that his TV habits are no different than those of his friends. Since this father has taken a stats class, he knows that he can actually test to see whether or not his son is watching more TV than his peers. The father collects a random sample of television watching times from boys at his son's high school and gets the following data:

1.9 2.3 2.2 1.3 1.6 2.3 1.8 2.0 2.0 1.8

Is the father right? That is, is there evidence that other boys average about 2 hours of television per day like his son? Use the 90% confidence interval to test his hypothesis, making sure to state your conclusions in the context of the problem.

Question 5: Hypothesis Test

Problem Statement:

- Father is concerned about his son watching too much television (average 2 hours/day).
- Claims that his son's habits are similar to those of his peers.
- Random sample of television watching times (hours):

$$1.9,\ 2.3,\ 2.2,\ 1.3,\ 1.6,\ 2.3,\ 1.8,\ 2.0,\ 2.0,\ 1.8$$

Hypotheses:

- Null Hypothesis (H_0): $\mu = 2$ hours
- Alternative Hypothesis (H_a): $\mu \neq 2$ hours

Question 5: Calculations

- Sample Mean ($\bar{x}$):

$$\bar{x} = \frac{1.9 + 2.3 + 2.2 + 1.3 + 1.6 + 2.3 + 1.8 + 2.0 + 2.0 + 1.8}{10} = 1.81$$

- Sample Standard Deviation (s):

$$s \approx 0.316$$

- Standard Error (SE):

$$SE = \frac{s}{\sqrt{n}} = \frac{0.316}{\sqrt{10}} \approx 0.1$$

- 90% Confidence Interval:

$$CI : \bar{x} \pm z^* \cdot SE \quad (z^* \approx 1.645)$$

$$CI : 1.81 \pm 1.645 \cdot 0.1 \Rightarrow (1.645, 1.975)$$

Question 5: Conclusion

- Since 2 hours is not within the 90% confidence interval (1.645, 1.975),
- We reject H_0. There is evidence that the average TV watching time of boys at this high school differs from 2 hours/day.