LAB: INTRO TO STAT ANALYSIS

Zeyi Qian

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

October 31, 2024

Question 1

Average happiness in the US is **7.6** based on a scale measuring happiness from 1 to 10 (where 10 is high levels of happiness). You think that people who have dogs are happier. You take a random sample of **50** dog owners and find that the mean happiness score in your sample is **8.2** with a standard deviation of **1**. Is this strong evidence that dog owners are happier than the general population? Use the 90% confidence interval to test your hypothesis and state your conclusion.

Question 1

Solution:

- **Hypotheses:**

$$H_0 : \mu = 7.6, \quad H_a : \mu > 7.6$$

- **Test Statistic:**

$$n = 50, \quad \bar{y} = 8.2, \quad stdev = 1, \quad SE = stdev/\sqrt{n}$$

- Use Excel "=T.INV.2T(0.10, 49)" or T table

$$df = n - 1 = 49$$
$$t^* = 1.676$$
$$\text{CI} = \bar{y} \pm t^* SE = 8.2 \pm 1.676 \times \frac{1}{\sqrt{50}} = [7.963, 8.437]$$

- **Conclusion:** Reject H_0; strong evidence that dog owners are happier.

Question 2

In the 1980s, it was generally believed that congenital abnormalities affected about **5%** of the nation's children. Some people believe that the increase in the number of chemicals in the environment has led to an increase in the incidence of abnormalities. A recent study examined **384** children and found that **46** of them showed signs of an abnormality. Is this strong evidence that the risk has increased? a) Write the appropriate hypotheses. b) Perform the test. What is the P-value? c) Explain what the p-value means in this context.

Question 2

Problem: Testing if the proportion of congenital abnormalities has increased from 5%.
Solution:

- **Hypotheses:**

$$H_0 : p = 0.05, \quad H_a : p > 0.05$$

- **Test Statistic:**

$$\hat{p} = \frac{46}{384} \approx 0.12, \quad n = 384$$
$$\text{SE} = \sqrt{\frac{0.05 \times 0.95}{384}} \approx 0.011$$

Question 2

$$z = \frac{0.12 - 0.05}{0.011} \approx 6.36$$

$$\text{P-value} \approx 0.00001$$

- Use Excel "=1 - NORM.S.DIST(6.29, TRUE)"or table to find P-Value
- **Conclusion:** Reject H_0; strong evidence of increased risk. **At any common level of significance (e.g., p-value less than 0.05 or 0.01), we reject the null hypothesis**.

Question 3

A survey of first-year and second-year students at a large university asked if students were satisfied with college life. We are interested in whether there is a difference between first and second-year students with respect to this question. You performed a hypothesis test and found there was no evidence of a difference between first and second-year students in response to this question. You based this conclusion on a test using alpha $= 0.01$. Would you have made the same decision at alpha $= 0.05$? Explain. (Hint: state the null and alternative hypotheses).

Question 3

Problem: Testing if there's a significant difference in satisfaction between first- and second-year students.

Solution:

- **Hypotheses:**
$$H_0 : p_1 = p_2, \quad H_a : p_1 \neq p_2$$
- **Conclusion:** No significant evidence to reject H_0; no strong evidence for satisfaction difference.

Question 4

Congress regulates corporate fuel economy and sets an annual gas mileage for cars. A company with a large fleet of cars hopes to meet the 2011 goal of 30.2mpg or better for their fleet of cars. To see if the goal is being met, they check the gasoline usage for 50 company trips chosen at random, finding a mean of 32.12mpg and a standard deviation of 4.83mpg. In this strong evidence that they have attained their fuel economy goal? a) Write the appropriate hypotheses. b) Find the P-value. (Hint: Use the excel function t.dist(.)) c) Explain what the P-value means in this context.

Hint: If we know the overall standard deviation , we can use the z-test because the z-test assumes that we have complete knowledge of the overall parameters (such as the overall mean and overall standard deviation). When we only have sample data, and especially when we only know the sample standard deviation but not the overall standard deviation, we need to use the t-test because it is better suited to deal with this kind of uncertainty.

Question 4

Problem: Has the company's fleet met the fuel efficiency goal of 30.2 mpg?
Solution:

- **Hypotheses:**

$$H_0 : \mu = 30.2, \quad H_a : \mu > 30.2$$

- **Test Statistic:**

$$\bar{y} = 32.12, \quad s = 4.83, \quad n = 50$$

$$t = \frac{32.12 - 30.2}{4.83/\sqrt{50}} \approx 2.81$$

$$\text{P-value} \approx 0.00355$$

- Use Excel "=T.DIST.RT(2.81, 49)" or t table to find p-value
- **Conclusion:** Since the p-value is less than 0.05, reject H_0; strong evidence that the fleet meets the goal.

Question 5

On many highways state police officers conduct inspections of driving logbooks from large trucks to see if the trucker has driven too many hours in a day. At one truck inspection station they issued citations to 49 of 348 truckers that they reviewed. Based on the results of this inspection station, construct the 95% confidence interval for the proportion of truck drivers that have driven too many hours in a day. Interpret the 95% confidence interval (i.e., put in a sentence).

Question 5

Problem: Construct a 95% confidence interval for the proportion of truck drivers who work overtime.

Solution:

- **Sample Proportion:**

$$\hat{p} = \frac{49}{348} \approx 0.128$$

- **Standard Error:**

$$\text{SE} = \sqrt{\frac{0.128 \times (1 - 0.128)}{348}} \approx 0.0185$$

- **Confidence Interval:**

$$\hat{p} \pm z^* \times \text{SE} = 0.128 \pm 1.96 \times 0.0185 = [0.091, 0.165]$$

- **Conclusion:** We are 95% confident that the true proportion of overworked truck drivers is between 9.1% and 16.5%.