Math 183 HW11, Name: Chandler Blaid Burgess, **PID:** A98029477

5.4 Find the p-value, Part II. An independent random sample is selected from an approximately normal population with an unknown standard deviation. Find the p-value for the given set of hypotheses and T test statistic. Also determine if the null hypothesis would be rejected at $\alpha = 0.01$.

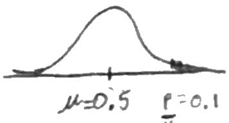
(a) $H_A : \mu > 0.5$, $n = 26$, $T = 2.485$

(b) $H_A : \mu < 3$, $n = 18$, $T = 0.5$

(Also, as you answer this question, draw a picture of the area you are finding, and write the command in R that you used to find the area. Confirm your area using the T-table in Appendix B.)

one-tail

a) From T-table p-value is 0.01 thus we can reject null Hypothesis. $p = \alpha$



b) One-tail

p-value is something > 0.10, thus we cannot reject H_0.

$p \simeq 0.31$

5.6 Working backwards, Part II. A 90% confidence interval for a population mean is (65, 77). The population distribution is approximately normal and the population standard deviation is unknown. This confidence interval is based on a simple random sample of 25 observations. Calculate the sample mean, the margin of error, and the sample standard deviation.

$\bar{x} = \frac{77-65}{2} + 65 = 71$

$z = 1.645$ $\quad SE = \frac{S_x}{\sqrt{n}}$

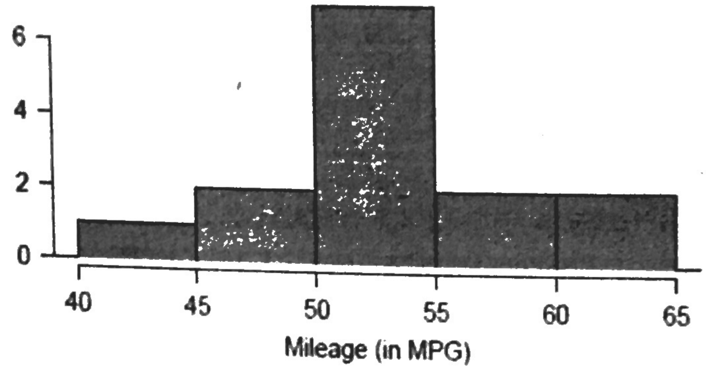
$ME = 0.10 = 1.645 \cdot SE$

$\Rightarrow 0.0608 = SE$

$S_x = 0.0608 \cdot \sqrt{25}$

$S_x = 0.304$

5.8 Fuel efficiency of Prius. Fueleconomy.gov, the official US government source for fuel economy information, allows users to share gas mileage information on their vehicles. The histogram below shows the distribution of gas mileage in miles per gallon (MPG) from 14 users who drive a 2012 Toyota Prius. The sample mean is 53.3 MPG and the standard deviation is 5.2 MPG. Note that these data are user estimates and since the source data cannot be verified, the accuracy of these estimates are not guaranteed.[9]



(a) We would like to use these data to evaluate the average gas mileage of all 2012 Prius drivers. Do you think this is reasonable? Why or why not?

(b) The EPA claims that a 2012 Prius gets 50 MPG (city and highway mileage combined). Do these data provide strong evidence against this estimate for drivers who participate on fueleconomy.gov? Note any assumptions you must make as you proceed with the test.

(c) Calculate a 95% confidence interval for the average gas mileage of a 2012 Prius by drivers who participate on fueleconomy.gov.

a) Yes, assuming that the 14 users were selected at random.

b) Assuming a near normal distribution. Assuming $\alpha = 0.01$

$H_0 \Rightarrow \mu = 50$ mph

$H_A \Rightarrow \mu \neq 50$ mph

$t = \frac{53.3 - 50}{SE}$ $SE = \frac{5.2}{\sqrt{14}}$

$t = \frac{3.3}{5.2/\sqrt{14}} = 2.3745 \Rightarrow p \approx 0.02$ $\therefore$ We reject H_0 as $p > \alpha$

c) $CI = \bar{x} \pm t \cdot SE$

$CI = 53.3 \pm 2.16 \cdot 5.2/\sqrt{14} = (56.3, 50.3)$

5.10 t^* vs. z^*. For a given confidence level, t^*_{df} is larger than z^*. Explain how t^*_{df} being slightly larger than z^* affects the width of the confidence interval.

If t^*_{df} is larger than z^*_i then the C.I will be larger, given that $\bar{x}+t^*_{df}\cdot SE > \bar{x}+z^*\cdot SE$ &

$\bar{x}-t^*_{df}\cdot SE < \bar{x}+z^*\cdot SE$

5.14 SAT scores. SAT scores of students at an Ivy League college are distributed with a standard deviation of 250 points. Two statistics students, Raina and Luke, want to estimate the average SAT score of students at this college as part of a class project. They want their margin of error to be no more than 25 points.

(a) Raina wants to use a 90% confidence interval. How large a sample should she collect?

(b) Luke wants to use a 99% confidence interval. Without calculating the actual sample size, determine whether his sample should be larger or smaller than Raina's, and explain your reasoning.

(c) Calculate the minimum required sample size for Luke.

a) $S_x=250$ $\quad ME = 25 = 1.29\cdot\frac{250}{\sqrt{n}}$

$n=(12.9)^2 \approx 165$

b) To be more confident, intuitively you would want to maximize your sample-size. Also, looking at $ME = z\cdot\frac{S_x}{\sqrt{n}} \Rightarrow n = \left(\frac{S_x\cdot z}{ME}\right)^2$ we can see that as z increases (w/ larger confidence) the sample size required also increases.

c) The min for Luke is $n=\left(\frac{S_x\cdot z}{ME}\right)^2=\left(\frac{250\cdot 2.33}{25}\right)^2=(10\cdot 2.33)^2=(23.3)^2\approx 539$

5.22 Auto exhaust and lead exposure. Researchers interested in lead exposure due to car exhaust sampled the blood of 52 police officers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. The blood samples of these officers had an average lead concentration of 124.32 μg/l and a SD of 37.74 μg/l; a previous study of individuals from a nearby suburb, with no history of exposure, found an average blood level concentration of 35 μg/l.

(a) Write down the hypotheses that would be appropriate for testing if the police officers appear to have been exposed to a higher concentration of lead.

(b) Explicitly state and check all conditions necessary for inference on these data.

(c) Test the hypothesis that the downtown police officers have a higher lead exposure than the group in the previous study. Interpret your results in context.

(d) Based on your preceding result, without performing a calculation, would a 99% confidence interval for the average blood concentration level of police officers contain 35 μg/l?

a) $H_0 \Rightarrow \mu \leq 35$

$H_A \Rightarrow \mu > 35$

b)
- The samples are independent? ✓
- The sample size ≥ 30? ✓
- population is not strongly skewed ✓

c) $t = \dfrac{\bar{x} - \mu}{s_x/\sqrt{n}} = \dfrac{124.32 - 35}{37.74/\sqrt{52}} = 17.067 \Rightarrow 17.07$

$p \approx 0$ ∴ reject H_0.

$p < \alpha$

We can conclude that downtown police officers have higher lead exposure than the group in the previous study w/ high favor of H_A over H_0.

d) Yes, because the p-value in (c) is nearly 0, thus the CI will be very similar to a dist. w/ 35 as the mean.

```
#~~~~~~~~~~~~~~ R1 ~~~~~~~~~~~~~~~~~

# Mean of baby weights
x_bar <- mean(smokingBabies$weight)

# SD of baby weights
sx <- sqrt(var(smokingBabies$weight))

# Histogram of the babies' weights
h <- hist(smokingBabies$weight)

#~~~~~~~~~~~~~~ R2 ~~~~~~~~~~~~~~~~~

#   The conditions of the sample were met to do justify
# inference on mu of the average weight of babies born to
# NC smokers. This is because the sample is less than 10%
# of the population of babies born to and participating
# smokers that gave birth were chosen at random, thus
# independent. Also, the number of people in the sample
# is large enough to get a graph that is near normal.

#~~~~~~~~~~~~~~ R3 ~~~~~~~~~~~~~~~~~

p_val <- pnorm(x_bar, mean = 7.18, sd = (sx/sqrt(50)), lower.tail = TRUE)

# [1] 0.0379445

#~~~~~~~~~~~~~~ R4 ~~~~~~~~~~~~~~~~~

ttest <- t.test(smokingBabies$weight, mu=7.18, alternative = "less")

#
# One Sample t-test
#
# data:  smokingBabies$weight
# t = -1.7751, df = 49, p-value = 0.04105
# alternative hypothesis: true mean is less than 7.18
# 95 percent confidence interval:
#   -Inf 7.157747
# sample estimates:
#   mean of x
# 6.779
```