Homework #10

Problem #1

Write your own one-sample t-test function for the " $\neq$ " alternative ($H_0 : \mu = \mu_0 \ vs \ H_a : \mu \neq \mu_0$), name it *my.t.test*(), for example. It should execute steps of the testing procedure I outlined in the class. Run your function and *R*'s *t.test*() function on the same input, compare the outputs. **You can not utilize function *t.test*() inside your own *my.t.test*() function definition**.

Make sure your function takes in just those three parameters as input:

- data vector of quantitative values,
- hypothesized mean value,
- confidence level,

and returns a list with following elements:

- t (**TS value**, has to equal the "$t =$ " from the *t.test*() call output for same data)
- df (**degrees of freedom**, has to equal the "$df =$ " from the *t.test*() call output)
- p.value (has to equal the "$p\text{-}value =$ " from the *t.test*() call output)
- CI (has to be a 2-element vector: first element = first value under "95 percent CI" of *t.test*() output, second element = second value)
- sample.estimate (just the sample mean, has to equal to the value under 'mean of x' in *t.test*() output)

Make sure to provide one example for classical confidence level (0.95), and one - for non-typical level (e.g. 0.99, or 0.90). Some code to guide you:

```
my.t.test <- function(vec, mu, ci.lvl=0.95){
  # Here the floor is yours.
}
my.t.test(c(1:10),mu=5) # Exemplary call and output for your function.
$t
[1] 0.522233
$df
[1] 9
$p.value
[1] 0.6141173
$CI
[1] 3.334149 7.665851
$sample.estimates
[1] 5.5

t.test(c(1:10), mu=5) # Comparative t.test() call.

my.t.test(c(1:10), mu=5, ci.lvl=0.99) # Example for non-typical CI level.

t.test(c(1:10), mu=5, conf.level=0.99) # Compare output with similar t.test() call.
```

Problem #2

This question deals with data on working hours per week.

a. Conduct a t-test on finding out whether the US population average of weekly working hours differs from the standard of $40hrs$. Make sure to formulate the hypotheses (in parameter notation), report both the p-value and (Cohen's) effect size. Interpret the result. Does the confidence interval agree with hypothesis test? Explain.
b. Conduct a t-test on finding out whether the **male** population average of weekly working hours differs from the standard of $40hrs$. Make sure to formulate the hypotheses (in parameter notation), report both the p-value and (Cohen's) effect size. Interpret the result. Does the confidence interval agree with hypothesis test? Does confidence interval also allow to judge the practical significance? Explain.
c. For parts (a) and (b) check if:
    - There are any strong outliers.
    - Given that, should we be concerned about the legitimacy of t-test results? Why/why not?

Problem #3

This question deals with our beloved Airbnb data.

1. Conduct a t-test on finding out whether the average price of NYC listings significantly differs from 150$ per night. What population we are trying to infer about when conducting this test? Make sure to formulate the hypotheses (in parameter notation), report both the p-value and (Cohen's) effect size. Interpret the result. Does the confidence interval agree with hypothesis test?

2. Is the normality assumption satisfied? Are there are any extreme outliers? Given that, should we be concerned about the legitimacy of the conducted t-test?

3. (**BONUS 1.5PTS; Similar to what we did with $TVhours$ data set in lecture code**). Building on part 2, proceed to pretend as if your data set is the "entire population" of NYC Airbnb listings (of size $48,864$). Repeat the following process $10,000$ times:

    i. Randomly sample 30 listings from that "population".
    ii. Conduct the t-test based on this sample of whether population mean price differs from $mean(listings\$price)$ (which is the true mean of our "population").
    iii. Record the p-value.

    In the end, calculate the % of times your p-value was less than 0.05. Is it what you expected with significance level $\alpha = 0.05$? Why/Why not?

Problem #4

This question deals with data on working hours per week.

1. Proceed to check the normality assumption for weekly working hours of 1) males and 2) females. Does it look satisfied? If not - do you believe that t-test results on comparing males and females with respect to weekly working hours might be suspect? Why/why not?

2. Conduct a two-sample t-test comparing males and females with respect to weekly working hours. Make sure to formulate the hypotheses (in parameter notation), report both the p-value and (Cohen's) effect size. Interpret the result (along with confidence interval). Does the confidence interval agree with hypothesis test? Explain.

Problem #5 (Make sure to use R when appropriate for carrying out the calculations, showing your work)

9.49

9.52

9.54 (In addition, provide the Cohen's d value)

9.56

10.14

10.24

10.49, 10.58 (please use R here, along with $t.test()$ function to find confidence intervals)