

Sean McLean

ALY 6010

Module 3 – R Practice

The first question looks at the BMI variable in the data set and after looking at the range of the variable from the 1000 people surveyed, the null and alternate hypotheses were established.

Null Hypothesis: The population mean of the BMI from the people surveyed is equal to 24.

Alternate Hypothesis: The population mean of the BMI from the people surveyed is not equal to 24.

The proper coding was entered in R and the following results were calculated to show whether or not the BMI is equal to 24:

One sample t-test

```
data: health$BMI
t = 3.5913, df = 999, p-value = 0.0003451
alternative hypothesis: true mean is not equal to 24
95 percent confidence interval:
 24.17502 24.59672
sample estimates:
mean of x
 24.38587
```

The test statistic provides a positive value which indicates that the sample mean is higher than the hypothesized mean. The p-value of the one sample t-test is very small which would provide ample reason to reject the null hypothesis because it is well below 0.05. The actual mean of the sample is slightly above 24, so overall the results from the statistical test would lead to a decision to reject the null hypothesis in favor of the alternate hypothesis because the population mean is not equal to 24 (Bluman, 2018).

The second question focuses on the p-value of the exercise frequency variable in the data set. After looking at the range which is from only 0 to 2, a null and alternate hypothesis is created that would show what metrics like the p-value would provide enough evidence to reject or not reject the null hypothesis.

Null Hypothesis: The true mean of the exercise frequency variable is equal to the hypothesized value of 1.

Alternate Hypothesis: The true mean of the exercise frequency variable is not equal to the hypothesized value of 1.

The proper coding was entered in R and the following results were calculated to show whether or not the exercise frequency variable is equal to 1:

One sample t-test

```
data: health$ExerciseFrequency
t = -1.8247, df = 999, p-value = 0.06834
alternative hypothesis: true mean is not equal to 1
95 percent confidence interval:
 0.9024548 1.0035452
sample estimates:
mean of x
```

0.953

With the second question centered on how the p-value impacts the null hypothesis, the p-value is higher than 0.05 which indicates that it is too high to possibly reject the null hypothesis. The negative t-statistic means that the sample mean is lower than the hypothesized mean for this variable in the data set. With a sample mean of 0.953, the true mean is not equal to 1, but the high p-value suggests that the null hypothesis could potentially not be rejected because there is a lack of strong evidence from the results in R.

References:

Bluman, A. G. (2017). *Elementary Statistics: A Step by Step Approach*. 10th edition. McGraw-Hill Education.

Chat GPT. (2023, November 18th). *Default (GPT 3.5)*. <https://chat.openai.com/>