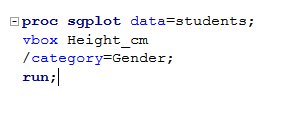
Lab 3

Adam Chlebek

1a. The question is stated to round to the nearest full cm with your shoes on. This could give a different answer based on the thickness of the shoe.

1b. Height\_cm

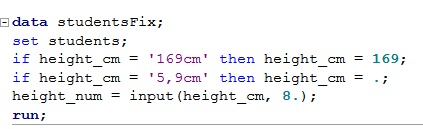
2a.



2b. ERROR: The VBOX variable must be numeric.

3a. Row 64 and Row 70 have ended with cm. Also, row 70 used a comma as well.

3b.

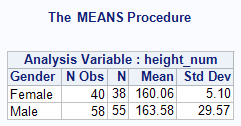


3c. 14, 65, 72, and 203

3d. The shortest human to ever live was 54cm tall. This sets 14 out of the picture. I don’t believe that 65 and 72 are reliable either, unless there is a student that is that short. 203cm is possible, however, since that is a possible height of someone in 8th grade.

3e. The student who wrote 65 could have been thinking in terms of inches. This is a reasonable size for a student in 8th grade.

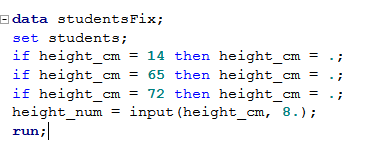
4a.



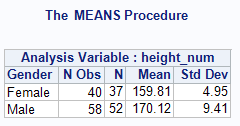
4bi. There is a possibility that 2 people completely skipped that question since it is left blank in the dataset.

4bii. N Obs is the observable sample size (the count of the rows based off of the gender) and N is the actual data that is collected. If you look at the dataset, there are a few rows missing in the height column where there is a gender. This is why the two values differ.

5a.



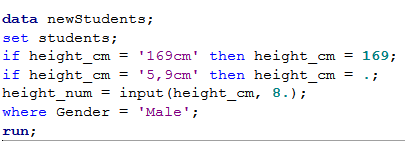
5b.



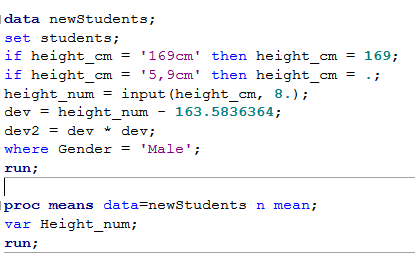
5ci. N decreases by 3 because there were 3 values removed.

5cii. Since 3 outstanding outliers were removed, the mean went up (since the values were below the mean).

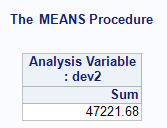
6a.



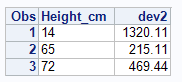
6b.



6c.



6d.



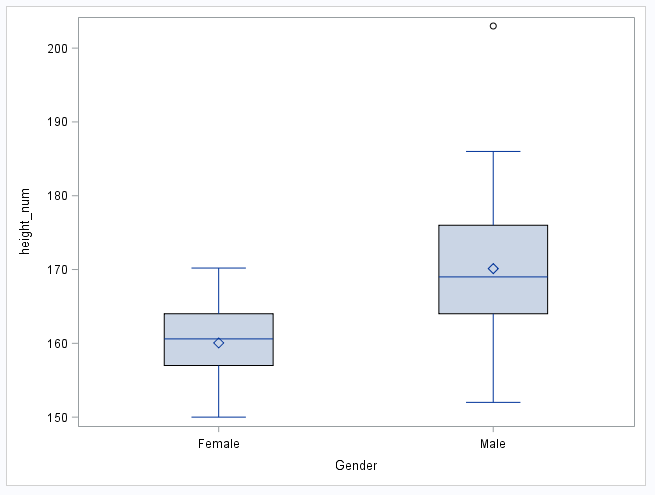
6e.

14cm = 2.8% of the sum

65cm = 0.45% of the sum

72cm = 0.99% of the sum

7.



8. The TTest shows a clear belle curve for the height and gender correlation. These follow the normal distribution and that is shown because the red and blue lines are very close together throughout the graph. The points on the scatterplot closely follow the regression line as well, which proves again that this correlation is normal.

9a. The population of the 8th grade students providing answers to the survey.

9b. Since all of the students in the United States didn’t respond to this survey, it would be impossible to relate this dataset to every student in the US. There are many variables that could change this outcome.

10a.

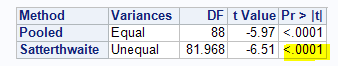
μ1 = The mean height for all females in 8th grade that take the survey.

μ2 = The mean height for all males in 8th grade that take the survey.

Ho : μ1 >= μ2 VS. Ha: μ1 < μ2, where μ1 = females mean height for 8th grade students that took this survey and μ2 = males mean height for 8th grade students that took this survey

10b.

Test statistic = -6.51 and p-value = <0.0001

10c. With alpha level of .01 and the p-value of <0.0001, we will reject Ho.

10d. At the alpha = 0.01 significance level, there is statistically significant evidence that the males mean height for 8th graders that took this survey is greater than the females in 8th grade that took this survey.

11.