**Motivation**:

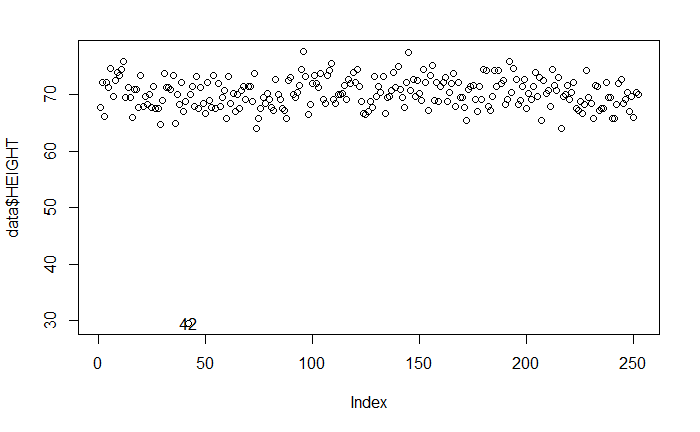
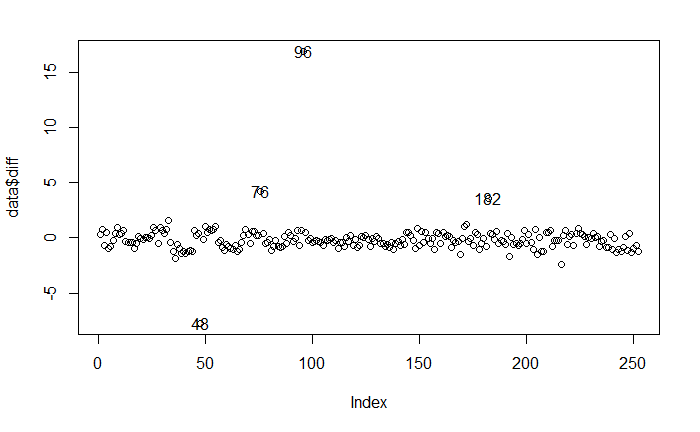
The percentage of body fat is an important indicator of personal health. However, it is hard to implement the accurate way of measurement of body fat. Thus, here we trying to discover a model to answer this question with other measurement which are much easier to get.

**Data Cleaning**:

The data set is about 252 men with measurements of their percentage of body fat and various body circumference measurements. (SOURCE). The Y, bodyfat percentage, range from 0 to 45.10, has mean 18.94 and median 19. It’s also worth noticing that the data is 0.1 level.

During checking the data set, there are several suspicious data points arise:

1. As we know from (), the bodyfat that calculated from the density shall be accurate. Yet there are several outliers whose bodyfat does not meet the calculated body fat from density. We can’t tell which one is wrong and the bodyfat is our Y. Thus we drop these data points. (48, 76, 182, 96)



1. Another suspicious point is point 42. It only has height of 30, which is below half of all other data points. So either there are some error in the data, or there is some very special disease (that person is over 40 years old, thus can’t be explained by age). Either way, we think it does not suit in our data set.
2. There are also some points (e.g. 39 and 41) that have measure values away from the majority. However after look deep into those data, we think the data is reasonable and can be explained with extremely obsess.

**Purposed Model:**

Our final model is:

And, in order to make it easy-to-use, we think the rule of thumb shall be: “(rounded)”.

So, for example, a man with xxx is expected to have body fat percentage of xxx, the 95PI is xxx.

1. Interpret your model (in *laymen’s terms*\*\*)
   1. Example 1: Our estimated coefficients are BLANK and BLANK, which are in the units of BLANK and BLANK. This means that for every BLANK increase in BLANK, the model predicts that body fat % will increase, on average, by BLANK.
2. Discuss, in some length, why you chose the final model. Include relevant tables/figures/etc (e.g. scatterplot of final model, table of key statistics, etc.)
   1. Example 1: We chose this model because of the following reasons. First, BLANK (e.g. background research). Second, BLANK (e.g. statistical analysis/figures/tables). Third, other models using BLANK had BLANK (e.g. some key statistics), which was BLANK in comparison to our final model.

**Statistical Analysis/Hypothesis Testing/Inference/**

1. Discuss any key statistical tests you have conducted to support your model. Make sure to interpret your results carefully.
   1. Example 1: We conducted the following test to see whether the predictor(s) we have chosen are significant in predicting the outcome. (formally state hypothesis; formally state which test statistic you used; state the p-value/test stat value; state the Type I error you’re willing to tolerate). From our statistical test/p-value, we can conclude that BLANK (interpret your statistical tests/p-values in a laymen’s term based on this data’s predictors; see lecture notes for details).
   2. Example 2: We found our R^2 to be BLANK, which implies (interpret it in layman’s terms)
   3. Example 3: The estimated slope and intercept are BLANK and BLANK, with 95% CIs BLANK and BLANK. This implies that BLANK (interpret in laymen’s terms). Also, based on the 95% CI, we can reject/retain the null hypothesis of BLANK and BLANK. In other words (interpret them in laymen’s terms)
2. This section may overlap with the previous section in some respects.

**Model Diagnostics**

1. Include any model diagnostics you did. Include relevant plots/tables/etc.
2. Explain your model diagnostics (e.g. why you did it, what assumption is this diagnostic checking, what is your conclusion from the diagnostic checks, how did you resolve any violations of model assumptions)
   1. Example 1: We checked the following four assumptions for SLR/MLR. First, we checked BLANK using BLANK (see Figure BLANK). Because BLANK, we believed BLANK is plausible, even though there is slight violations of BLANK. Second, we checked BLANK.
   2. Example 2: We also looked at three types of outliers in regression models: BLANK. For BLANK, we used BLANK and looked for BLANK; see Figure BLANK. We found BLANK and we removed/kept/fixed this outlier. After rerunning the diagnostics checks, we found BLANK.

**Model Strengths/Weaknesses**

1. Discuss strengths and weaknesses of your model
   1. Example 1: Some strengths of our model include BLANK, BLANK, and BLANK. In particular, our model satisfies the linear regression assumptions of BLANK and BLANK, brining credence to our results/interpretation in BLANK.
   2. Example 2: Some weaknesses of our model include BLANK, BLANK, and BLANK. In particular, for certain groups of males, the model may provide BLANK and BLANK because of BLANK.

**Conclusion/Discussion**: (summarize what you wrote above; final thoughts/discussions)

**Contributions**:

HT:

YZ:

ZJ:

Overall, we met online through video 4 times, spent about 6 hours discussing during the meeting. We also discuss in our IM group.

* 1. Example 2: JD: Figures BLANK, model strength/weakness, final editing, presentation slides BLANK. HK: figures BLANK, data cleaning

**References:** must be only on the 3rd page and may not exceed more than one page.