Jared Yu
Module 8 Discussion

In the first question I agree with what you had said regarding what type of approach we should be using. It seems to me that in the case of the variance being proportional to x, then that reminds me of the table in the textbook that indicates that we should be using some sort of transformation in such a case. I am guessing that the square-root is appropriate, since the textbook seems to say that it's a good choice for $E(y)$, where this includes directly the x term. Afterwards, I think it also makes sense to analyze their effects to see which is best, since there's no guarantee that one transformation would work better than the other. I'm curious also if domain knowledge of the dataset would prove useful in such a situation. For example if you know that the full range of values instead included negative values and so the square root would not prove useful in such a situation.

You mention also an interesting point regarding the use of transformations. I think you are discussing the limit of using them in certain situations. For example, if the range of values is limited, then the impact of the transformation is also limited. In my response, I was thinking instead about how transformations will alter the unit measurement of the data, so the results will be in some alternate unit of measurement. Likewise, doing inverse transformation of the results doesn't provide the same meaning since the $E(y)$ is now about the median rather than the mean.

I think you make a good point also about continuous variables having a lower probability of there being replicate values. An addition that I can guess is that if the points are rounded to perhaps the first or second decimal place then it can become more likely. However, I often see in the textbook for example datasets where this is not the case and so replicate values are not commonplace. I also agree that we can do lack-of-fit if we at least have some replicate values. I wonder though how useful it could be if there are only one or two such values.

I agree with your explanation of what happens in the case of nonconstant variance. I also mentioned that the coefficients will be unbiased but lack the minimum-variance property. Therefore, they are no longer BLUE as in the case of the OLS estimators. I think there are some alternatives that the textbook mentions, such as using generalized or weighted least squares.