I do not really understand is the Box-Cox transformation.  In addition to the derivation being difficult, I do not fully understand when we would want to apply this in a real world setting.  In other words, what situation would we need Box-Cox instead of just doing a log or square root transform?

ANOVA table and the exact meaning of each terms.

I am still a little bit confused regarding the H1 hat matrix. I am also still a little bit confused regarding the process for model selection when looking purely at diagnostic plots. I feel like I know what they mean much better now, as I previously mentioned, but I think it is extremely difficult sometimes to be able to tell which model actually looks better, especially when they are relatively similar in strength of fit.

When completing the second homework I became pretty confused on how to write out a hat matrix. I understand that the hat matrix multiplied with y creates yhat, but when I was attempting to fully write out a hat matrix and do the multiplication for myself it became complicated and confused me more. I'm not sure if it's just not something we typically write out since there are so many parts involved in it or if it's not an important thing to communicate. Also I could use more practice with knowing how and when to apply transformations because I feel like my previous stats courses glossed over the criteria for determining when a transformation is needed. I do understand the concept of applying transformations to achieve a more normal distribution of points, but I'm just not sure how to tell which transformations should be applied in situations like the PGA golf data where it's unclear sometimes. I think if a data set has a clear trend towards a universally identifiable pattern like a negative quadratic, it's easier to see that it should be transformed as such. In future problems if I encountered an amorphous blob/circle of data points in a scatterplot however, I wouldn't know how to proceed.

Early on the matrices were a bit overwhelming because I haven't done linear algebra since Freshmen year. However, HW2 helped me brush up on some matrix properties and ideas (idempotent, symmetry, identity matrix, etc.) so now I feel like I'm on more solid ground. But I still have a bit more ways to go before I feel 100% comfortable with matrices. Generally, I feel that the implementation of the ideas we learn in class into a data analysis helps make the theory more digestible.

I'm having some difficulty understanding Variance-Stabilizing Transformations (covered in Lecture 4).

 The concept of Box-Cox is not so intuitive and I am trying to understand how it really works. It was the first I learned the concept, so I am still trying to understand the fundamentals behind the concept and the graphical interpretation behind it.

I'm not very sure when we should include interaction terms in a model, and what the criteria of a good interaction terms would be.

It seems like the transformations we've seen so far address skewness in the data. What transformations should we use to address kurtosis? (Are non-monotonic transformations a bad idea in this process?)

Something that confuses me is whether the Scale-Location plot and Residuals vs Fitted plot shows whether the variance of the $\hat{e}$, $\epsilon$, $y$ or $\hat{y}$ is non-constant or constant. Also, I'm not sure how to derive variances and covariances of $\hat{e}$, $\epsilon$, $y$ or $\hat{y}$.

I think I am bit confused about Box-Cox transformation and standardized residual.

I do not really understand how could I judge which model is better or the best. There are many information we can look to a model, such as "R-squared", "Residual plots", "AIC" and so on, I am not sure how to balance between these staffs.

Something that we covered but still confuses me is how we calculate leverage for the Residuals vs Leverage plots. I understand from a high level what leverage is: data points that have a large affect on the model. However, I still get confused on how you would go about calculating leverage by hand.

I still feel shaky on the theory behind VST, specifically determining when and where we should be using certain transformations. This feeling does extend towards the application of all these methods we are learning in the wild on data not from homeworks. I don't feel confident with providing justifications and understanding what assumptions we can make (and should make) when making linear models.

I actually have a confusion of the GIR variable in question 1. Based on the normal QQ plot and histogram of the GIR variable, it has a little long right tail and maybe a good or a little long left tail. Do I need to make a transformation for this variable and what kind of simple transformation can I use.

In fact, except for the Normal Q-Q plot, I still don't know exactly how to judge the fit of the other diagnostic plots, although I have learned about it on the Internet, I don't know if my knowledge is correct.

How do high-leverage points affect a linear model?

Not really, but I feel like it will gets harder in the later courses though... I'm a statistic major so I've seen most of the stuff so far but I'm still worrying about what comes after.

Something that still mystifies me in this course so far is the concept of box-cox method. I believe as it is covered more and I see it come up somewhere it will make more sense, but I'm a bit confused on it.

I am really confused on how the box cox likelihood is generated in such a complicated formula.

\*\*\*\* The matrix algebra is something that still mystifies me. I am working on getting a better understanding of it but I find it extremely challenging.

Something that I am still a little confused about is the application of interaction terms, especially with multiple quantitative variables. I was confused about identifying good uses for them without negatively impacting the model. I remember covering in class that omitted variable bias has a far greater negative effect than the penalty for additional variables, but other than testing all interactions (while still following the hierarchy principle), if there is a quicker/better way other than substantive reasoning. I am used to conducting regression with interactions that involve a categorical or dummy variable, so the use, selection, and interpretation with quantitative interaction variables is something I hope to clarify over the semester.

We've learnt hat matrix in this course, but I don't know the application of hat matrix, what's the meaning of hat matrix?

To be honest, at first I did not fully understand the implications made in Sheather's book, especially because it has been so long since I looked into these statistical equations. My major issue was that I was not able to make connections between these mathematical terms and real-world business problems. However, as the semester progresses, this point became much more clear, and I am actually re-reading these textbooks thoroughly to solidify my understanding right now. This will take some time since it is quite a bit of reading, but I am making a good progress. I still have a little bit of issue with above mentioned connection (eg. Why would we need to use Hat Matrix in the real-world business context?), but I believe such shortfalls will be resolved soon as I complete my re-reading.

I still have a hard time interpreting the diagnostic plots, specifically the Scale-Location plot. In the case where there's functional dependence of variance on y, what additional information would Scale-Location plot give compared to Residual vs Fitted plot?

The simple transformation of variables can only deal with the situation that both left and right tails lay on the same side of the line? For instance, the Q-Q plot for variable GIR in question 1 does not look perfect. However, as I tried several simple transformation, such as sqrt, log, square, none of these transformation makes the plot looks better. So I think the transformation we would like to choose should based on the distribution of the data points. For instance, the Q-Q plot for PrizeMoney looks like a logarithm, so we choose log(PrizeMoney) instead to make it more normally distributed. Thus, if we need to improve the Q-Q plot for GIR, a more complicated transformation might works. Is it correct?

How does the Box-Cox function choose the best lambda?

The use of transformations which can make the variance of residuals constant.

I was struggled with Problem 2(b)(iii) and 2(c) in homework2. Other than that, the lecture materials and other homework questions are fine.

What is the maximum likelihood estimation used for? And is adjusted R square a mainstream judging method on how well the model fits?

Since it has been a long time from my last linear model courses in undergraduate, the matrix and transformation in linear model confuse me a lot.

Sometimes it is ambiguous to judge whether the results of diagnosis plots satisfy the assumptions.

I often use a trial-and-error approach when fitting y and x1, x2... to determine the form of x (whether it is the combination of x1 and x2,  or log(x1) or other forms), is there a more logical way to find its form?

How to analyze the diagnostic plots, especially the leverage examples (don't know how to tell high/low leverage and residual).

I don't understand how Hat matrix works in real-world applications since we already have the package. I also don't quite understand how some equations we have in HW#2 but never defined.

I don't really understand about the reducing leverage: powers of X, especially the Box-cox.

How should we generally check which transformation we need to make?  Seems like we need to check each column of data to see if they are normally distributed. What if the dataset is too large?

I understand all the class materials for far, the only question I have is, why should we prove the matrix in our HW2 questions 1 & 2 because we don't need to use that info in R to run the data.