HW2 (Due: April 18 - Monday)

1. Using tlc data, reproduce results in S95-108.
2. Using tlc data, analyze the same data with the different methods:
3. Mixed model (e.g., for longitudinal data)
4. ANOVA (=one outcome at a time)
5. repeated measures ANOVA
6. MANOVA (=ANOVA for multiple outcomes)

And comment on the differences, advantages (if any) and findings.

1. With the first 2 visits in tlc data, analyze the data with different methods for pre vs. post data:
2. ANOVA
3. Change analysis (i.e., change as outcome)
4. ANCOVA
5. Paired t-test

And comment on the differences, advantages (if any) and findings. [There is a file   
“ANOVAvsChangevsANCOVA.pdf” in the Optional folder. You can check if you want.]

1. How to estimate and interpret “Adjusted means” (=LS means) from #3?
2. How and why to use “Contrast” statement in SAS/R?
3. After reading Tomasetti and Vogestein (Science 2015), aka, “bad luck cancer paper, a renowned statistician stated “*The R2 value doesn't matter, because it is not cor(Y, X). Instead, it is cor(Ybar, X). (Ybar is substituted for Y in their analysis at each of 31 values of X, that is, cancer incidences are averages).Their R2 could be 0.9999 while cor(Y,X) could be 0.0001.*

*Also, you can have R2=1 while the slope is barely off zero.*”

Please justify if this statement is correct. [For example, you can simulate a toy data. Other

options are permitted.]

Disclaimers: Some of traditional methods were not discussed in class but they are old and tons of resources in textbooks and internet. There are no strictly right or wrong answers, so feel free to analyze and summarize as you wish (but succinctly!).