## HW Question Week 5

## STA2101F 2021

## Due October 21 2021 11.59 pm

## Homework to be submitted through Quercus

Part 1: Data set for project (brought forward from HW4)

Please submit details about the data you will use for your project. Ideally the data will have a single response or outcome variable of interest, and several potential explanatory variables. You should submit with this homework:

- (1) the data source: both bibliographic and a web link
- (2) the number of observations and the number of potential explanatory variables
- (3) a description of the response variable
- (4) a description of the potential explanatory variables
- (5) the scientific question(s) of interest

When you submit the final project, it will consist of the parts listed in Slide 3 on October 6.

Part 2: Question for marking:

LM-2, Ex.16.6. The "High School and Beyond" data is found in hsb.

- (a) Model the math score in terms of the five factors: gender, race, ses, schtyp and prog. Include all second-order interactions but no hgiher-order interactions. How many regression parameters does your model use? Explain how this can be calculated in terms of the number of levels for each factor.
- (b) Determine whether some two-way interactions can be eliminated using the **anova** function.
- (c) Determine whether some two-way interactions can be eliminated but now using the drop1 function. Why do the results differ from the previous question? Which method of testing do you think is better, and why?
- (d) Fit a model with only main effects and compare it to the model with all two-way interactions. Which model do you prefer and why?
- (e) Check the regression diagnostics for the main effects only model and report on any anomalies.

- (f) Summarize your conclusions from the analysis of the data in plain language, at most one paragraph.
- (g) Bonus/PhD: week off yay!