**EDF 5401 Practice Final 2023 Questions**

*For all questions, include an explanation.*

1. **Multiple Regression**

A.1. Linearity

Look at the initial scatterplot. Is there any reason to think that the relationship between the pretest and posttest is non-linear?

A.2. Distribution shape by condition

Look at the boxplots by condition. Is there any reason to think that the gain scores have a different distribution in different conditions?

A.3 Demographic Variables

Is there any reason to believe that the relationship between pretest and posttest is different for different demographic groups? If so, which ones?

A.4 Base Model `ECTpost ~ ECT + Supports`

Write the equation of the fitted base model. (Hint replace the `Supports` with two dummy variable, `COG` and `COG+AFF`.

A.5 Aptitude Treatment Interaction Model.

Compare the no interaction model `ECTpost ~ ECT + Supports` with the model with the interaction, `ECTpost~ECT\*Supports`. Is there an interaction between the support condition and the pretest score?

A.6 Model Selection

Which demographic variable improve the fit of the model? How much additional variance is explained by the new variables?

A.7 Demographic Effects

For each additional variable added to the model in the pevious step, describe the strength and direction of the effects.

A.8 Influential points

Compare the models fit with and without the influential points. Describe the effect of the influential cases on the model.

A.9 High and Low ECT pretest scores

One way to fit a nonlinear model is to add a dummy variable for different regions. Here `LowECT` represents scores below the median of 9.

Compare the model with and without the `LowECT` model, which one is better and why?

A.10 LowECT:Support interaction

Are the treatment effects different for high and low ECT students? Explain your answer.

A.11 Best Model

Of all the models we’ve considered, which one do you think is best? Explain your choice.

A.12 Treatment Effects

What are the effects of the two treatments in your best model? Are they significantly different from zero?

1. **Logistic Regression**

B.1 Scatterplot

Look at the scatterplot. Describe the relationship between Physics Score and probability of enjoyment.

B.2 Difference by gender

Look at the scatterplot with gender and the table. Is the relationship between physics score and enjoyment the same for boys and girls?

B.3 Model Equation

Write the equation of the model without sex.

B.4 Relationship between physics and enjoyment

Describe the relationship between the physics score and enjoyment according to the model.

B.5 Sex intercept

Are girls or boys more likely to enjoy the game? How big is the difference? Is it plausible that there is no difference?

B.6 Sex slope

Is the relationship between physics and enjoyment for both boys and girls?

B.7 Enjoyment Rates

Describe the predicted enjoyment of the game for a student with average physics ability, for a boy with average physics ability, and for a girl with average ability.