Homework 5

Minoo Ahmadi

**Part 1: Mixed Effects ANOVA**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | SS | df | MS | F |
| A | 100 | 1 | 100 | 100/4.54 = 22.02 |
| B | 210 | 4 | 52.5 | 52.5/20 = 2.625 |
| AB | 80 | 4 | 20 | 20/4.54 = 4.4 |
| Within | 300 | 66 | 4.54 | N/A |
| Total | 690 | 69 | 10 | N/A |

1. Main effect of A:

Main effect of B:

Interaction of A and B (AB):

Critical F(1,66) = 3.98, and Critical F(4,66) = 2.51. All the F-values exceed the Fcritical , so the null hypotheses are rejected for both of the main effects, as well as the interaction.

**Part 2: Journal Article Review**

* Briefly describe the purpose and procedures of the study and the variables analyzed. What are the research questions as they pertain to the regression analyses? Identify the dependent and independent variables.

I chose the Smith, 1996 for this assignment.

This study was performed to answer the question whether taking algebra courses prior to high school would affect students’ later attainment in mathematics (the DV) while in high school. Early access to algebra is the IV, and high school math attainment (the DV) has been operationalized in terms of two outcome measures: “the number of years of advanced mathematics courses taken” and “senior-year mathematics achievement”.

* Briefly describe the sample data used (who was included in the sample?), the sample size, how the sample was selected, and the intended population of generalization. Use that information to comment on the adequacy of the sample size for the analyses performed, and the representativeness of the sample for the intended population of generalization.

Sample comes from two groups of students in their sophomore year, selected from the High School and Beyond (HS&B): students who have taken an algebra course in high school (5,818 students) and students who had taken an algebra course before high school (1,076 students).The population of generalization is American high schoolers. The sample size seems adequate, but it would be better if the two groups had roughly equal sizes. Overall, sample size seems large enough to be representative of the intended population.

* Describe the regression analyses performed in the paper.

The researcher does a forward selection stepwise linear regression, with “early access to algebra” as the first step and entering 2 sets of covariates as next 2 steps (social background [including African American, Hispanic, female, and social class] and 10th grade characteristics [including educational aspiration, in the academic track, and 10th grade achievement]) into the model. R2 changes are reported in the tables.

She also reports mean differences for each of the characteristics between the two samples.

There’s also a path analysis which is looking at the effect of taking algebra in 10th grade on the achievement in 12th grade and the possibility of this variable as being the mediator between pre-high-school algebra and 12th grade math attainment.

* Interpret the results of the regression analyses performed in the paper.

The stepwise regression models (both for early algebra on years of math and early algebra on math achievement in high school) show that including both sets of covariates (social background and 10th grade characteristic) add to the value of the model in predicting the outcome. Hence, ANCOVA is useful here. Taking social background and 10th grade characteristic into account, early algebra taking predicts math attainment (years of math and early algebra on math achievement) in high school.

* Discuss whether the (regression-related) conclusions drawn in the paper follow from the regression analyses reported.

They talk about the “True” impact of early algebra being somewhere between stage 2 and 3 of the model, which doesn’t make much sense to me. The model seems to be significantly predictive of the outcome at all stages and becomes better after entering the 10th grade characteristics.

I don’t quite follow her report on mean differences (from multiple (15) t-tests corrected with Bonferroni adjustments) and how it adds anything to the analysis and I think is quite unnecessary if not incorrect at all.

I’m not sure if I understand the path analysis either.

* Discuss the extent to which the authors tested and/or addressed the assumptions underlying multiple regression.

I don’t see any assumption checking being reported in the paper. This is especially important since sample sizes are not equal. In case there’s a violation of homoscedasticity, this will cause the larger n to have larger variance and lead to higher error rates.

* Discuss weaknesses or problems with the analyses reported in the paper and describe alternative analyses that could have, or should have, been performed. If the authors did not test or address the assumptions underlying multiple regression, what are some potential weakness or problems with their quantitative analyses? What analyses should be performed to address these assumptions?

Checking normality, homoscedasticity, and other assumptions would be very helpful. If there’s any violation they should be fixed by appropriate transform methods.