Analysis of Student Exam Preferences Under Various Conditions

STA4211 Written Report

Zachary Allen

December 6, 2023

Objective

The objective of this project is to observe how student's prefer a course based on its exam format, and whether this preference changes once the exam averages for each format are revealed to the student. Observations will be recorded through a survey which also contains demographic questions, and analysis will be conducted using methods learned throughout STA4211. Primarily, this involves determining whether certain main and interaction demographic effects exist and if group means/medians are statistically different from each other.

Data Collection

Data was collected using a Qualtrics survey that was sent to students on UF-centric online communities such as Discord and Reddit. After 2 weeks the survey accumulated 51 respondents from UF students. In short, each respondent answered the following 6 questions.

1. Educational Status at UF (1st year Undergrad through Masters/PhD)

2. Primary Field of Study (Computer Science, Stats/Data Science, Math, Other)

3. Would you prefer a course with a 1 week long take-home (open-book) midterm worth 40% of your grade [Format A] or 3 timed in-person and closed-book exams worth 75% of your grade [Format B].

4. Rate your preference over the alternative from 1 to 5 (or 0 if no preference)

5. Format A has a midterm average of 61% and Format B has averages 73%, 76%, and 75%. Which format do you prefer with this new information?

6. Re-rate using same scale as Q4.

For the full details of each question see **Appendix**. For the resulting data set see `https://github.com/ZackAllen1/exam-preference`

Data Processing

To create slightly more balanced groups, Responses from Question 1 regarding educational status were split into two groups: low (1st to 3rd Year Undergrad, **24**) and high (4th and 5th+ Undergrad, Masters/PhD, **27**). Additionally, responses from Question 2 regarding primary field of study were regrouped into 4 categories: Computer Science (CS, **35**), Quantitative Sciences (SCI, **7**), Life Sciences (LI, **5**), and Other (O, **4**).

Notation and New Variables

Let the initial rating given in Question 4 be denoted by IR and the new rating given in Question 6 be NR. Additionally, enforce NR and IR to be positive whenever Format A is selected and negative whenever Format B is selected. We can now define our response variable `cip` (change in preference) as $NR - IR$ which will be positive when Format A becomes the preferred format or Format B becomes less preferred once exam averages are given. Likewise, `cip` will be negative when Format B becomes the preferred format or Format A becomes less preferred when exam averages are given.

Data Analysis

The following section will describe how two primary forms of data analysis were conducted on the resulting dataset. The first of these forms is to determine whether the factors within the Education Status and Field of Study variables have statistically different means/medians.

Residual Analysis for ANOVA Models

Before using an ANOVA model to determine if group means are statistically different, we must first check if the model residuals are normally distributed and the within group variances are equal. As shown in the Figures below, it is clear that the residuals for each model are left-skewed.

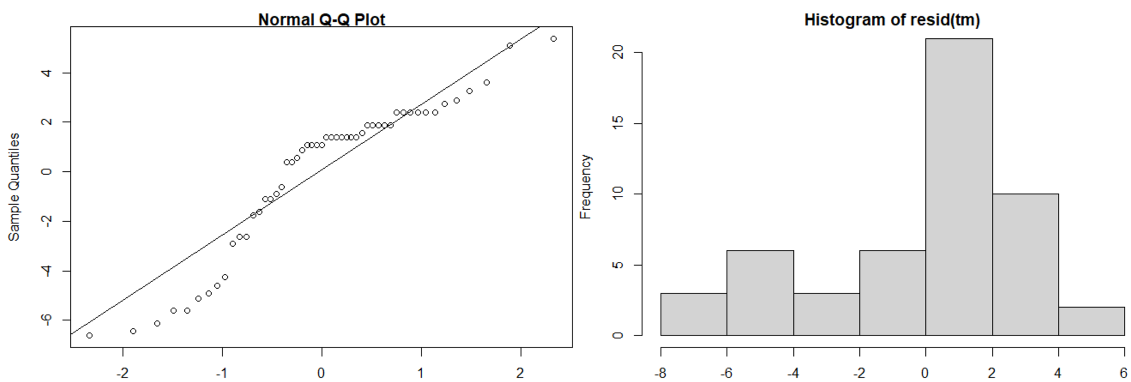

Figure 1: Q-Q Plot and Histogram for FieldOfStudy Model Residuals

To test if group variances are equal, Bartlett's Test was used for each model. For the model containing the high/low educational status, Bartlett's Test ($H_0 : \sigma_1^2 = \sigma_2^2$), resulted in a test-statistic of 0.48342 with 1 degree of freedom and a p-value of 0.4869. For the model containing the field of study, Bartlett's Test ($H_0 : \sigma_1^2 = \ldots = \sigma_4^2$), resulted in a test-statistic of 1.1032 with 3 degrees of freedom and a p-value of 0.7763. Based on these tests we can

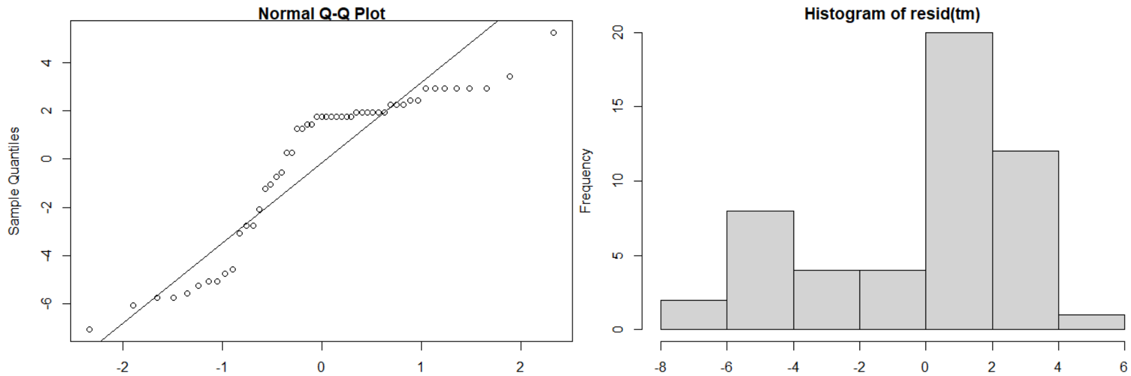

Figure 2: Q-Q Plot and Histogram for high/low Model Residuals

conclude that the group variances are not statistically different for each factor.

Since the residuals are non-normal and the group variances are not statistically different per factor, we can deploy a Kruskal-Wallis Test as a non-parametric equivalent to ANOVA to determine if high/low and FieldOfStudy group medians are equivalent.

Kruskal-Wallis Test

To conduct this test we rank all observations across treatments from 1 to n_T, assigning average ranks if ties occur. We define $\overline{R}_{i\bullet}$ as the average rank of the observations in each group. The test statistic is:

$$X^2_{KW} = \left[\frac{12}{n_T(n_T+1)\sum_{i=1}^{R}\frac{R^2_{i\bullet}}{n_i}}\right] - 3(n_T+1) = \frac{SSTR_R}{\left[\frac{SSTO_R}{n_T-1}\right]}$$

$$RR : X^2_{KW} \geq \chi^2(1-\alpha, r-1)$$

For `cip` by high/low, we get a test statistic of 0.28418 with 1 degree of freedom and a p-value of 0.594. For `cip` by FieldOfStudy, we get a test statistic of 2.8703 with 3 degrees of freedom and a p-value of 0.4121. Thus, we can conclude that the `cip` medians for high/low are approximately equal and the `cip` medians for FieldOfStudy groups are approximately equal.

Unbalanced 2-Factor Studies

We now transition to the second objective of this analysis, which is to determine whether main/interaction effects exist for high/low and FieldOfStudy when predicting `cip`. Since the factors for FieldOfStudy are significantly unbalanced, we can utilize the following model

$$\begin{aligned} Y_{ijk} =& \mu_{\bullet\bullet} + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + \beta_3 X_{ijk,4} \\ &+ (\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + (\alpha\beta)_{13} X_{ijk1} X_{ijk4} + \epsilon_{ijk} \end{aligned}$$

where

$$X_1 = \begin{cases} 1 & \text{low in high/low} \\ -1 & \text{high in high/low} \end{cases} \quad X_2 = \begin{cases} 1 & \text{LI in FoS} \\ -1 & \text{CS in FoS} \\ 0 & \text{otherwise} \end{cases} \quad X_3 = \begin{cases} 1 & \text{O in FoS} \\ -1 & \text{CS in FoS} \\ 0 & \text{otherwise} \end{cases} \quad X_4 = \begin{cases} 1 & \text{SCI in FoS} \\ -1 & \text{CS in FoS} \\ 0 & \text{otherwise} \end{cases}$$

and $\sum \alpha_i = \sum \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$.

We can create three additional reduced models from this full model, where each one removes either a main effect or the interaction term. Comparing each of these new models to the full one using an F-test allows us to determine if main or interaction effects exist. The general test statistic and rejection region is

$$F^* = \frac{\left[\frac{SSE(R) - SSE(F)}{df_E(R) - df_E(F)}\right]}{SSE(F)/df_E(F)} \qquad RR: F^* \geq F(1-\alpha, df_E(R) - df_E(F), df_E(F))$$

Testing H_0 : All Interaction Effects $= 0$ versus H_a : Not all $(\alpha\beta)_{ij} = 0$ results in a test statistic $F^* = 3.7013$ and a critical value of $F(0.95, 3, 46) = 2.807$. Testing $H_0 : \alpha_1 = \alpha_2 = 0$ versus H_a : Not all $\alpha_i = 0$ results in a test statistic $F^* = 0.0735$ and a critical value of $F(0.95, 1, 46) = 4.051749$. Testing $H_0 : \beta_1 = \ldots = \beta_4 = 0$ versus H_a : Not all $\beta_j = 0$ results in a test statistic $F^* = 1.9020$ and a critical value of $F(0.95, 3, 46) = 2.807$.

Conclusions and Limitations

From the result above, we can conclude that interaction effects between education status and primary field of study exist, whereas both main effects do not. We also concluded previously that the group medians within each factor are not statistically different from each other.

There do exist some limitations of this analysis. For example, the midterm averages shown in the survey were fixed. It could be possible that if different reported averages were shown to the respondents that different responses would have occurred. However, there were not enough subjects to test the plausibility of this claim. Additionally, not collecting whether the student has experienced either exam format in the past may also impact results. Lastly, since respondents chose if they wanted to be as part of the sample this lead to volunteer bias, and as a result the conclusions drawn from this sample may not be representative of the entire population.

Appendix

Survey Format

Question 1

What is your current educational status at UF? **(Num. Repondents)**

- 1st Year Undergraduate **(5)**
- 2nd Year Undergraduate **(4)**
- 3rd Year Undergraduate **(15)**
- 4th Year Undergraduate **(16)**
- 5th Year Undergraduate **(8)**
- Graduate Student **(3)**

Question 2

Please select your primary field of study for your degree: **(Num. Repondents)**

- Computer Science, Computer Engineering, or DAS **(35)**
- Statistics or Data Science **(2)**
- Mathematics **(0)**
- Other **(14)**

Question 3

For a typical course in your primary field of study, which of the following exam formats would you most prefer?

- **Format A:** A 1 week long take-home (open-book) midterm worth 40% of your grade. No final exam.
- **Format B:** Three timed in-person and closed-book exams worth a combined 75% of your grade. No final exam.
- No Preference

Question 4

How much do you prefer the option selected in Q3 over the alternative?

- 1 (Slight Preference) to 5 (Extreme Preference) in 0.5 intervals.
- "No Preference" in Q3 = 0

Question 5

In a previous semester, both formats were tested and the following exam averages were reported

- **Format A:** 61% Midterm Average
- **Format B:** 73%, 76%, and 75% Average for Each Exam

Assuming you would receive an exam(s) of similar difficulty, does this new information change which exam format you would most prefer?

Question 6

Preference rating of selection with new information over alternative. No preference in Question 5 = 0.