**Exploring the Correlation between GRE and First-Year GPA**

**with the Moderating Role of Demographic Factors**

Ziyu Ren

*University of Minnesota*

PSY 8712

**Exploring the Correlation between GRE and First-Year GPA**

**and the Influence of Demographic Variables**

In the current project, I explored the distribution of GRE scores and 1st-Year graduate school GPA, together with their correlations. Then, I showed the varying GRE-GPA correlations across different demographic groups with a shiny app. Finally, Machine learning models are built to predict GPA from GRE and demographic variables.

**Hypotheses and Research Questions (RQ)**

I explored the following three research questions.

H1 (RQ1): Is there a significant sex difference on GRE Verbal / Quantitative score? (This is technically a hypothesis, but I label it as RQ1 to align with other materials.)

RQ2: What are the correlations between GRE Verbal / Quantitative score with GPA across different demographic groups (separated by sex, age, citizenship, and major)?

RQ3: How well can I predict 1st-Year GPA with GRE and demographic information using machine learning models?

**Method**

**Open Science Materials**

A binder is an online code repository contains code and contents, together with configuration files to create a project from the scratch. When saving an R project in a binder, it saves the R version being used to create the project to avoid possible confusions in future replications. Therefore, I create a binder for the current project with mybinder.org, which is an online service to build and share binders. The binder for the current project can be found online via the link <https://mybinder.org/v2/gh/Lindsey-R/psy8712-final/HEAD?urlpath=rstudio> . Clicking this link will automatically open an R studio session built under R 4.2.3 (the version I use to run the project) with corresponding codes and data.

Similarly, material to create the current project can be accessed through GitHub following the link <https://github.com/Lindsey-R/psy8712-final> . Clicking on this link will lead to the GitHub page will all files, an README file with project descriptions can be accessed by screwing done to the bottom of the page.

**Participants**

The current project used pre-collect publicly available data from OpenICPSR (<https://www.openicpsr.org/openicpsr/project/155721/version/V1/view;jsessionid=CCA87775E2BAE63EE1B4FC92FF3AE409>). Participants are 3538 de-identified graduate students from IVY League universities.

**Measures**

Below are variables used in the current analysis and their corresponding measures.

**Sex** Male / Female as reported by students.

**Citizenship** US / International Citizens as reported by students.

**GRE Quantitative** GRE Quantitative scores.

**GRE Verbal** GRE Verbal Scores

**GRE Writing** GRE Writing Scores

**GraduateFieldProgram** The student’s major in their college.

**GPA**  GPA as reported.

**Stay** Whether the student stay in the program or not (i.e., dropout).

**Procedure**

Given that the data is obtained from online sources, the procedure to collect data is unknown to me. Likely self-report is used for demographic information, and GRE and GPA are obtained from school profile.

**Analysis**

**Descriptive Statistics and Static Visualizations**

Table 1 shows the descriptive data for overall scores of GRE (Verbal, Quantitative, Writing) and GPA. A detailed table of descriptive data for each demographic group can be found online through the GitHub links.

**Table 1.**

*Descriptive data of GRE and GPA.*

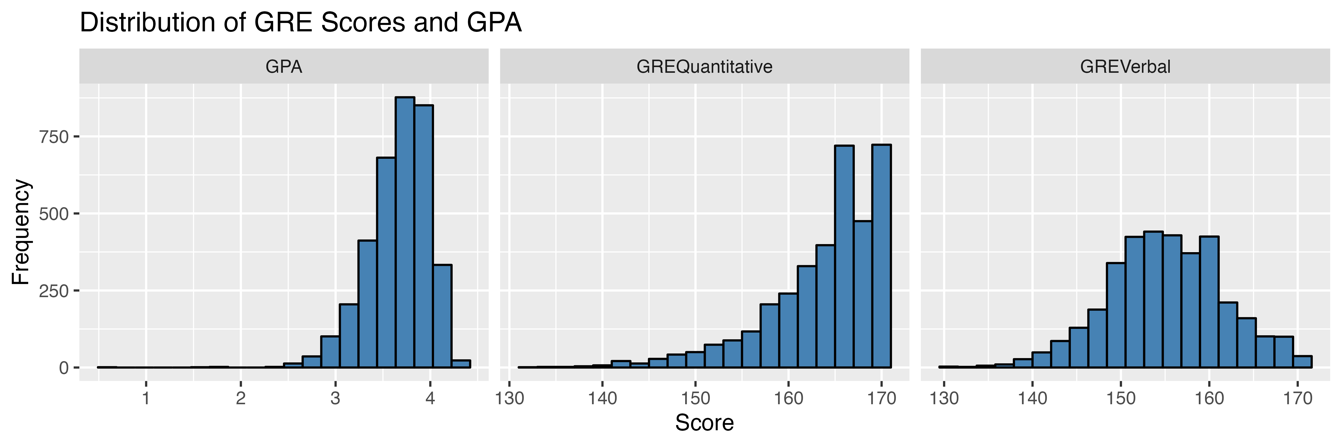
|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | ***N*** | **Mean** | **SD** |
| GRE Quantitative | 3538 | 164.11 | 5.98 |
| GRE Verbal | 3538 | 155.07 | 6.55 |
| GRE Writing | 3538 | 3.68 | 0.69 |
| GPA | 3538 | 3.67 | 0.32 |

*Note: N* = total sample size.

Figure 1 shows the distribution of GPA, GRE Quantitative, and GRE Verbal scores. While GPA and GRE Quantitative scores are skewed to the left, GRE Verbal is roughly normal skewed.

**Figure 1.**

*Descriptive plot of GRE and GPA.*



**Interactive Visualization**

An online shiny app (<https://purplefishlovespig.shinyapps.io/greshinyapp/>) was created for data visualization. The app was created as part of the process to answer RQ2, with interacting plot showing correlations based on different demographic groups. Specifically, allows user to choose from five options: Sex, Citizenship, Major (Graduate Field Program), whether student finished program (Stay), and GRE (GRE Verbal, Quantitative, or Sum score). Based on the user’s choice, the app returns 1) a scatter plot showing the correlations between the selected GRE and GPA, with a linear predicting line and 2) a line of text output the correlation.

**Data Cleaning**

In data cleaning, I kept all observations from participants who report a GRE and GPA scores. In addition, participants without information about whether they stay with the program or not were removed. Then, I selected all variables of interest: Sex, Citizenship, GRE scores, GPA, Graduate Field Program, and Stay. Finally, I kept majors with more than 300 students for future analysis.