Shiny App User Guide (DATA2902)

Toan Anh Dung Pham - SID: 520417215

Github:
https://github.sydney.edu.au/tpha0236/shiny-app-1.git

Shiny App:
https://oscarpham2090.shinyapps.io/shiny-app/

1 Introduction

This Shiny application was developed using RStudio, employing several key R packages for its construction. The application architecture is divided into two primary components: the User Interface (UI) and the Server. The UI file is responsible for rendering the layout and elements visible to the user, while the Server file handles all backend computations and data processing.

The dataset utilized in this application, titled `df_cleaned.csv`, is a curated and cleaned dataset containing selected columns from the original data collected through an online survey. This survey was conducted as part of Assignment 1 in the DATA2X02 course[1], ensuring the dataset's relevance and alignment with the course objectives.

2 Data Cleaning

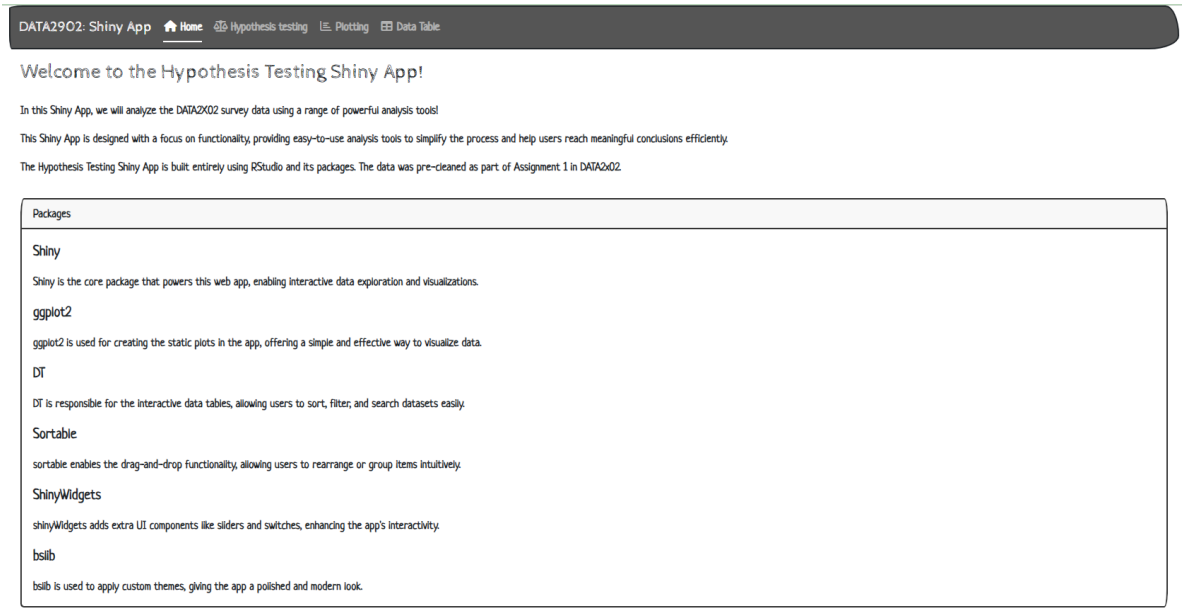
In the data cleaning process, columns with more than 50% missing data were removed to ensure completeness, while missing values in numeric and categorical data were imputed using the median and mode, respectively. Outliers, such as study hours exceeding 120 hours per week, were filtered out to improve data accuracy. Sleep hours and driver's license information were standardized for consistency, and unnecessary columns, such as timestamps and irrelevant preferences, were removed. Categorical text data was converted into factors for easier analysis. This cleaning ensured a more reliable and structured dataset for analysis.

3 Shiny App Implementation

The Shiny application is organized into four distinct tabs: "Home," "Hypothesis Testing," "Plotting," and "Data Table." This section provides a brief overview of the functionality of each tab.

3.1 Home

The "Home" tab serves as the introductory page of the application. It provides users with a summary of the app's purpose and instructions on how to navigate and use its features. This tab is designed to offer a clear entry point for users unfamiliar with the app.

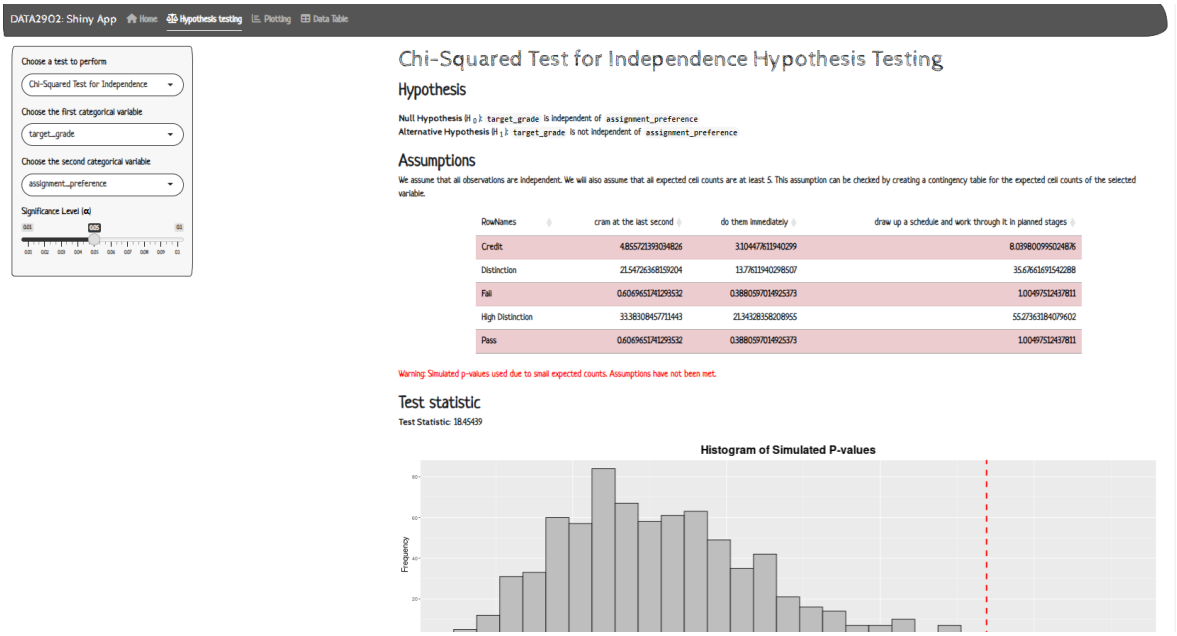


3.2 Hypothesis Testing

The "Hypothesis Testing" tab allows users to perform a range of statistical tests on pre-cleaned quantitative and categorical variables. Available tests include the Chi-square test for independence, the Chi-square goodness-of-fit test, and the two-sample t-test.

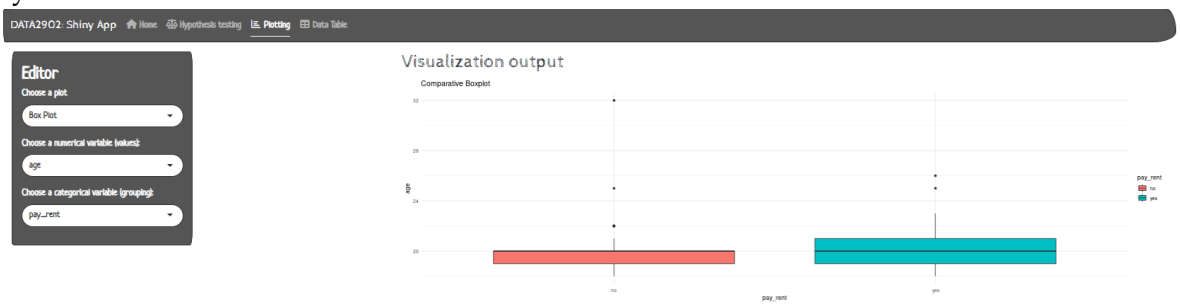
For Chi-square tests, users can select one or two categorical variables, depending on the test type. If the cell count assumptions for the test are not met, the app will automatically conduct a permutation test. Results are presented in a standard format (test statistic, p-value, and conclusion), with the significance level set by the user.

For the two-sample t-test, the app prompts users to select a quantitative variable and a categorical variable. Users can then choose two specific levels from the categorical variable for group comparison. The app checks assumptions by generating Q-Q plots and boxplots, allowing users to visually assess normality and variance. If assumptions are violated, the app provides options to perform a permutation test or a Welch's t-test as alternatives.



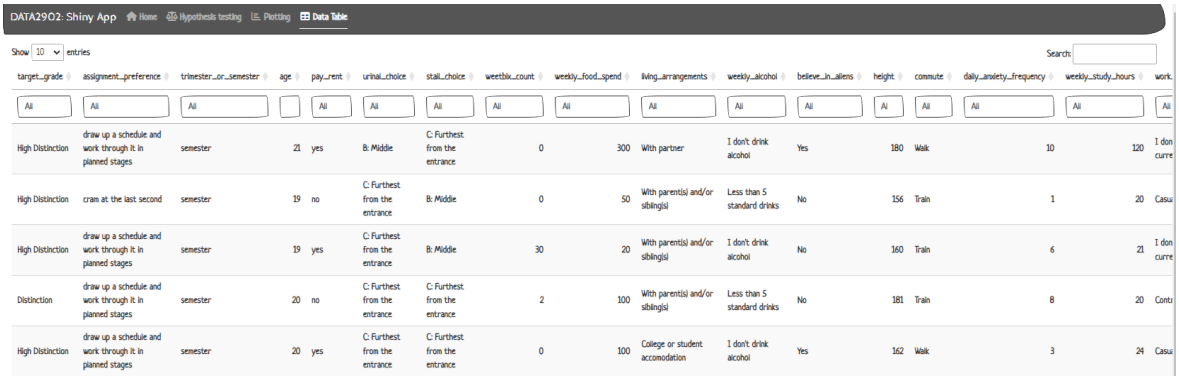
3.3 Plotting

In the "Plotting" tab, users can visualize the data based on the selected variables. Various plots, such as boxplots, are available to illustrate the distribution of the quantitative variable across the levels of the categorical variable. This provides a visual interpretation of the data, aiding in the understanding of the statistical analysis.



3.4 Data Table

The "Data Table" tab allows users to view the dataset in a tabular format. This tab is useful for exploring the raw data or reviewing specific observations. Users can interact with the table, applying filters or sorting data as needed.



4 App Design and Limitations

4.1 ui.R and server.R

ui.R:

- Title Panel: Displays the hypothesis testing title.
- Sidebar Layout:
  - Variable Selection: Dropdown to select variables.
  - Instructions: Guides how to select appropriate variables.
- Main panel: Displays the Hypothesis Test in HATPC format.

server.R:

- Data Handling: Loads a CSV dataset, converting character columns to factors.
- Dynamic Variable Updates: Observes dataset and updates dropdowns for quantitative and categorical variables.
- Test Execution: Performs the corresponding hypothesis test in R and outputs the results.
- Error Handling: Displays appropriate messages for invalid selections or insufficient data for the tests.

4.2 Limitations

- As for the t-test, the app assumes that the data is appropriate for a t-test (e.g., normality of the data distribution), but it does not automatically check assumptions like variance equality (homoscedasticity) or data normality, and instead relies on the user's decision, which could lead to misleading results in some cases.
- Independence assumption is not guaranteed since the data originates from an online survey of DATA2X02 students, which is not a random sample with a lot of biases.

- The app is designed to work with a specific dataset (`df_cleaned.csv`), which must have clean, structured data with numeric and categorical variables. The user cannot load their own data on.
- The app only offers 3 hypothesis tests, and doesn't allow for customization or flexibility in terms of options for testing.
- The cleaned data is based on a specific sample and may not be representative of the entire population.
- The use of simulated p-values (Monte Carlo) is based on the assumption of fixed rows and column totals, which, in the context of this online survey dataset, is not not satisfied.

5 Acknowledgements

This App was created with the help of ChatGPT[2], specifically in parts in debugging, syntax templates (Template for ui.R and server.R, as well as sidebar panel layouts) and finding appropriate functions(dropdowns, sliders, and tabs) for the app.

This App's theme is created by Bootswatch[3]. All the functions and syntax layout were utilized with the help of `bslib` documentation.

References

[1] G. Tarr, "Build software better, together," 2024.

[2] O. ChatGPT, "Chatgpt," 2024.

[3] B. Bootswatch, "Bootswatch: Free themes for bootstrap," 2024.