COSC478-24S2 ASSIGNMENT-2/TASK-1
Addendum to Proposal

David Ewing

2024.09.13

Addendum: Statistical Test Framework

In addition to the proposed chatbot, the following outlines a framework for developing a statistical test to validate hypotheses based on running a set of conditions with small changes. This framework includes bootstrapping techniques and adjustments for inflation, using multiple predictors.

1. Set of Conditions and Statistical Sampling

Define a set of initial conditions, each representing a variable or predictor. By systematically varying these conditions through small adjustments, we can simulate the behavior of the system under different scenarios. Running each variation multiple times allows us to collect data for statistical sampling. This robust dataset can then be used for subsequent analysis.

2. Hypothesis Testing

We propose the formulation of a hypothesis to be tested, such as:

> "The two predictors have a statistically significant effect on the outcome."

This hypothesis can be tested using the following statistical methods:

- **t-tests** or **ANOVA**: Useful for comparing means across different groups.
- **Regression Analysis**: To explore the relationship between predictors and outcomes.
- **Chi-Square Test**: Suitable for analyzing categorical data.

3. Bootstrapping

Bootstrapping will be used to generate confidence intervals and validate the test results by reducing assumptions about the data distribution. This technique involves resampling the collected data to create multiple simulated datasets, allowing for a robust estimate of the variability of statistics (such as means or variances).

4. Multiple Predictors and Inflation Adjustment

To adjust for inflation and include multiple predictors, a **multiple regression model** can be constructed. For example, the following equation can be used:

$$\text{outcome} = \beta_0 + \beta_1 \cdot \text{predictor1} + \beta_2 \cdot \text{predictor2} + \epsilon$$

In this case, one of the predictors could represent inflation, allowing for its effects to be considered in the model.

5. Example Workflow

1. **Generate Data**: Establish initial conditions and make small adjustments.
2. **Run Simulations**: Perform simulations or data collection for each set of conditions.
3. **Collect Data**: Gather the resulting data for statistical analysis.
4. **Hypothesis Testing**: Use appropriate statistical tests, such as regression analysis, to test hypotheses.
5. **Bootstrapping**: Apply bootstrapping to estimate variability and confidence intervals.
6. **Interpret Results**: Determine whether the hypothesis holds, and evaluate the effect of predictors (including inflation).

Tools and Techniques

- **R** or **Python**: These programming languages can be used to run statistical tests, bootstrapping, and regression analysis.
- **ggplot2** or **matplotlib**: Visualization libraries for generating plots to illustrate relationships and results.
- **StatKey**: An online tool for performing bootstrap analysis.

This statistical framework provides a robust approach to testing hypotheses in the chatbot context and could further enhance the chatbot's ability to detect and analyze biases using data science techniques.