CS 5A

Nov 21

Final Project Logistics

- Final project notebook submission is due on **Wednesday, December 4 (11:59 p.m.)**
- Only one member will turn in the final project notebook as an .ipynb file
  - and **they must add their partner to their submission** via Gradescope (instructions can be found at the bottom of the notebook)
- Of the available notebooks, **you only should complete one**, the notebook assigned to your presentation slot on the sign up sheet

Late submissions will not be accepted for the Final

Presentation Logistics

- Each team will prepare a **6-8 min presentation**, on results of notebook
- Presentations during section on **Thursday, December 5**
  - list your partners' names in your presentation time slot on the sign up sheet, by **Friday Nov 22 (11:59pm)**
- All team members need to attend their own presentation
  - You can sign up for any slot, and it does not have to be in your regular section
- Oral presentation slides are due to the Google Drive by **Thursday, December 5 at 9 a.m**
  - Ensure the presentation title includes both partners' full names and the topic of your final (ex: Bob_Smith+John_Doe_Titanic)

FInal Grading

- Final project is worth 25% of overall grade and is scored out of 100 pts
  - 70 points for notebook submission and 30 points for presentation
- Please **compile all ChatGPT prompts** used in their corresponding questions on submitted notebook
  - You are not required to use ChatGPT, but if you do, you must include the prompt
- New Collaboration Guidelines
  - Split up responsibilities before starting in writing and include this at the top of the notebook where you write your names
  - After presentations, we will send out a survey where you note what you completed and whether your teammate(s) held up their end of the bargain
    - Each team member completes this independently and anonymously
  - Your self-evaluation and your teammates evaluation of you is not directly part of the rubric.

Presentation Grading

- Oral presentation is 6-8 mins, and will be timed
  - Presentations that exceed 8 mins **will be cutoff**, to ensure time for everyone
- Focus of slides should be displaying your **figures, analysis, and findings** from your final
  - You should not include your code or large paragraphs of text
  - We want to see graphs and most important takeaways
- **All members of your group should be there for the presentation**
- See EdStem post for oral presentation rubric

Lab and Worksheets Logistics

- Lab 6 is an extra credit lab (optional)
  - Due **Wednesday, December 4 (11:59 p.m.)**
- Worksheets 7 and 8 released (optional)
  - Due **Wednesday, November 27 (11:59 p.m.)**

Lab 06 - Assessing a Model

A model is a set of assumptions about data i.e assumptions about chance processes that affect the distribution of data

-> A model helps us understand data and allows us to make predictions

We can evaluate the quality of a model, with previous techniques

Simulation -> Chosen Statistic -> Expectation -> Evaluation

If the behavior die deviates significantly from a fair die, we can assume (or model) the die is unfair

Lab 06 - Null and Alternative Hypothesis

Null Hypothesis states that there is no effect, no difference, or no relationship in the population being studied. It represents the default or "status quo" assumption.

Example: "There is no difference in average test scores between two groups."

Alternative Hypothesis states that there is an effect, a difference, or a relationship in the population. It is what the researcher aims to support or prove.

Example: "There is a difference in average test scores between two groups."

Lab 06 - p-value

p-value is the probability of observing the data (or something more extreme), assuming the null hypothesis is True.

- A small p-value (typically less than 0.05) indicates that the observed data is unlikely under the null hypothesis, leading to its rejection in favor of the alternative hypothesis
- A large p-value suggests that the data is consistent with the null hypothesis, so there is no strong evidence to reject null hypothesis

p-value is a means to evaluate the quality of a model

Lab 06 - A/B Testing

A/B testing is a statistical method used to determine if two numerical samples come from the same underlying distribution, by checking if there is a significant difference between them.

1) **Hypothesis**: null hypothesis assumes there is no difference between the two groups; alternative hypothesis assumes that there is a significant difference.
2) **Randomization**: randomly assign individuals to either Group A or Group B, to ensure that any differences observed between the groups are due to the treatment
3) **Data Collection**: We collect data from both groups, measuring a specific variable of interest
4) **Statistical Analysis**: Using statistical techniques, we analyze the collected data to determine if there is a significant difference between the two groups (t-tests, chi-square tests, or ANOVA)
5) **Interpretation**: Based on the statistical analysis, we either reject or fail to reject the null hypothesis

Section 11/21

- Review Siddharth's EdStem Post so you are aware of all the final deadlines and grading criteria
  - reach out to teaching staff with questions

Work Period for rest of Section