Body Fat Project: Grading Guidelines

Total Points: 60 points

Deliverables and Deadlines:

|  |  |
| --- | --- |
| Deliverables (for **Tuesday Group**) | Due Date (All times are local to Madison, WI) |
| Presentation slides (.pptx, .ppt, .pdf)  Note: You’ll give an in-class presentation on Tuesday, Oct. 15. | Monday, Oct. 14, 2024, 11:59pm |
| Two-page executive summary (.pdf) | Monday, Oct. 14, 2024, 11:59pm |
| Github repo containing code (a web link) | Monday, Oct. 14, 2024, 11:59pm |
| Shiny (or Web-based) app (a web link) | Monday, Oct. 14, 2024, 11:59pm |

|  |  |
| --- | --- |
| Deliverables (for **Thursday Group**) | Due Date (All times are local to Madison, WI) |
| Presentation slides (.pptx, .ppt, .pdf)  Note: You’ll give an in-class presentation on Thursday, Oct. 17. | Wednesday, Oct. 16, 2024, 11:59pm |
| Two-page executive summary (.pdf) | Wednesday, Oct. 16, 2024, 11:59pm |
| Github repo containing code (a web link) | Wednesday, Oct. 16, 2024, 11:59pm |
| Shiny (or Web-based) app (a web link) | Wednesday, Oct. 16, 2024, 11:59pm |

**All deliverables** **must be submitted to Canvas**. For the presentation and the executive summary, you must submit (i) one file for the presentation and (ii) one file for the two-page executive summary. For the Github and Shiny app, you must provide (i) a link to the Github page that is hosting all the code and (ii) a link hosting your functioning Shiny App in a Shiny App server of your choosing (or an equivalent web-based app). We'll only grade the latest submission that were submitted before the due date. Only one member of the group should submit all three deliverables for the group. No late submissions will be accepted.

If possible, we encourage you to submit all the deliverables only once.

Groups:

You will work in groups of three or four. Each group will be responsible for all the deliverables.

You will choose your groups on Canvas (under People’s tab). For the Tuesday group, you must choose your groups by **the end of class on** **Sept. 24**. For the Thursday group, you must choose your groups by **the end of the class on Sept. 26**. If you do not choose your groups by then, I’ll randomly assign you into a group.

The teaching staff has the final say on group assignments.

Grading and Distribution of Points:

Unless otherwise noted, all group members will receive the same grade for the project. However, if the teaching staff finds that there were **serious, unequitable contributions** or there was a **lack of regular communication** between team members, we’ll assign points individually within a group. Please remember that this may also mean that **some group members who may have contributed the most to the project** may still lose points if they didn’t make an earnest attempt to keep others informed about their progress or they weren’t being a

“team player” with respect to sharing the workload.

See the Grading Rubric below for the exact breakdown of points by each deliverable. **All members of the group must contribute to each deliverable.**

Presentations:

The goal of the presentation is to practice presenting your statistical findings in an organized, concise and clear manner. The presentation should include key evidence (e.g. plots, tables, inferential methods, etc.) that support your findings. Your presentation must be clear enough that **any employee with a quantitative background** (not necessarily in statistics) should be able to understand what statistical analysis you used and how you have reached your conclusion. The exact grading rubric for the presentation is outlined below. If you would like, you can also demo your Shiny app (see below) during the presentation. But, this is not a requirement.

Your group will prepare a **8 minute** in-class presentation of your data analysis; your presentation can be shorter than 8 minutes, but it cannot be longer than 8 minutes. All members of the group must work on the presentationand **each person must speak** **for at least one minute** during the in-class presentation. The exact time of your group’s presentation will be determined randomly on the first day of the presentation.

Due to time constraints, the time limit will be**strictly enforced**. To encourage this behavior, every additional 30 seconds after the time limit will incur **a penalty of 2 points**. It is **your responsibility** to rehearse your presentation so that it stays under the time limit.

Each group will submit **a single** presentation slide (in .pptx, .ppt, .pdf) to Canvas.It is your responsibility to check with the teaching staff that your slides can be displayed properly on the projector in the lecture hall before the presentation day.

Two-Page Executive Summary

Your group must submit a **two-page** executive summary of the data analysis, with at most **one additional page** for references & contributions. All relevant statistical analysis, plots, tables, figures must fit in the two- page limit and the one-page reference section can only be usedfor references to other works & contributions (see below).

Your summary must include (i) your overall findings and the final model, (ii) relevant and important evidence to support your findings (e.g., plots, tables), and (iii) important details of your statistical analysis (e.g., type of model used, inferential quantities, outliers, leverage points, modeling assumptions, etc.). Your summary should be detailed enough that **any data scientist** can understand your summary and replicate your analysis. The exact grading rubric is outlined below. All members of the group must contribute to the summary.

On the reference page, you may follow any reasonable style for references (e.g. MLA, APA, Chicago Manual of Style, etc.). Also on the reference page, the group must **clearly indicate** **each member’s contribution** to the presentation, the summary, the code, and the Shiny app. For example, you can add a table such as:

|  |  |  |  |
| --- | --- | --- | --- |
| Contributions | Hyunseung Kang | John Doe | Jane Doe |
| Presentation | Responsible for slides 1-4 (introduction and data cleaning).  Reviewed/edited slide 5-10 (results). | Responsible for slides 5-10 (results).  Reviewed/edited slides 1-4. | Reviewed/edited and provided feedback on all slides. |
| Summary | Responsible for introduction, data cleaning, conclusion, and references.  Reviewed/edited data analysis section. | Responsible for Figures 1 and 2.  Reviewed/edited and provided feedback on whole document. | Responsible for data results.  Reviewed/edited the introduction, data cleaning, and conclusion |
| Code | Responsible for data cleaning code.  Reviewed code for analysis section | Responsible for methods/results for final models and code to replicate Figures 1 and 2.  Reviewed data cleaning code. | Responsible for methods/results under different models. Reviewed data cleaning code. |
| Shiny App | Responsible for Shiny app | Reviewed/edited and provided feedback on Shiny app | Reviewed/edited and provided feedback on Shiny app |

The summary must be typed in 12-inch Calibri, Aptos, Times New Roman, Sans Serif, or Computer Modern font. The summary must be single-spaced with 1-inch margins and include all relevant figures/tables and equations. All figures and tables must be legible when printed on an 8in. x 11in. paper.

Each group will submit **a single electronic copy (in .pdf)** to Canvas. It is your responsibility to submit the file on time and that the file can be opened in a standard PDF or Word viewer.

Github Repository

Your group must publish a Github repository (repo) and provide **a web link** to the repository. The Github repo must contain the following:

1. a data folder containing the raw and (if relevant) cleaned data
2. a code folder containing all the code for your analysis (e.g. cleaning the data, running the analysis, producing figures/tables, Shiny app base code, etc.)
3. an image folder containing any figures/images/tables produced in your analysis.
4. The two-page, pdf summary file above.
5. The final presentation slides above
6. a README Markdown filesummarizing the contents of the repository and directions on how to use the code.

The code must **replicate every part of your analysis** from the start (i.e. reading in the data) to the end (the figures/tables/results in your presentation and two-page summary). Replication code includes, but is not limited to: data cleaning, outlier removal, model building, evaluation of different models, statistical testing, prediction, and any and all intermediary plots, tables, and analysis. Your R code must reproduce the **exact tables, plots, and other analysis** in your summary and the presentation (i.e. exact labels for axis, color shading plots, etc.). Also, your R code also must be **well-documented** so that **any data scientist** can read and understand your code. This is important for reproducibility and to track down potential bugs/errors in your analysis. All members of the group must contribute to the R code. All members of the group must contribute to the Github repository.

It is **strongly encouraged that you use Github pull/push/commit** functions to manage your project and to record who contributed/worked on different parts of the code.

Shiny App:

In some settings, data scientists are expected to make “actionable” prototypes/products based on your data analysis. To replicate this practice in industry, you will create a Shiny (or a web-based) application that will demonstrate your analysis in real-time and **submit a link to your** **live/running Shiny app** for grading.

Shiny is an easy-to-use, R-based platform to turn your R code into a web application. While you do not have to use Shiny (if you have app development experience, feel free to use other languages/platforms!), all applications must run on the latest Chrome browser and be accessible to the teaching staff for grading. For more information about Shiny, visit: <https://shiny.posit.co/>.

We’ll leave the user-interface and other graphical specifications of the Shiny app up to you. We’ll grade the Shiny application based on the Grading Rubric below. All members of the group must contribute to the development of the Shiny App.

Grading Rubric:

We will use the following grading rubric to grade your deliverables.

|  |  |
| --- | --- |
|  | Possible Points |
| Presentation | 25 |
| 1. Clear, takeaway message with a “rule-of-thumb”/final model 2. A simple, illustrative example to demonstrate the final model 3. Relevant, concise, and clear summary of statistical analysis 4. Relevant (**no extraneous and/or “R-dump” plots**!) and visually accurate plots 5. Strengths and weaknesses of the analysis 6. Overall, did the group present convincing evidence for their finding? 7. Overall, was the delivery clear and easy to understand? 8. Overall, was the presentation under the time limit? |  |
| Executive Summary | 25 |
| 1. Introduction and motivation for model 2. Statement of the final model and rationale for why the final model was chosen 3. Concise and relevant summary about the model, which may include R^2, standard errors, confidence intervals, p-values, hypothesis testing statements, etc. **No “data/R printout dump”** (properly format your tables/plots so they look presentable!) 4. Clear, laymen’s interpretation of the statistical estimates and inferential quantities 5. Model diagnostics with meaningful plots and discussion of model assumptions 6. Strengths and weakness of the model, conclusion or discussion. 7. Does it follow instructions concerning the two-page limit? |  |
| Shiny App | 10 |
| 1. Does it run in real time? 2. Is the application robust to user inputs (e.g., gives warning if you put in too high/low weights)? 3. Does it provide useful and insightful information to the user? 4. Is there contact information in case if users have any questions about the app? |  |

The Github repository is **graded for completion only**. If you do not submit the link to the Github repo containing the elements described above, **your project will not be graded** and you will get a zero for the entire project.

A Reminder: Academic Integrity

Each year in Stat 628, we have a few students who violate the academic integrity standards laid out in class. We take this opportunity to remind students of the policies regarding academic integrity.

By virtue of enrollment, each student agrees to uphold the high academic standards of the University of Wisconsin–Madison; academic misconduct is behavior that negatively impacts the integrity of the institution. Cheating, fabrication, plagiarism, unauthorized collaboration and helping others commit these previously listed acts are examples of misconduct which may result in disciplinary action. Examples of [disciplinary sanctions](https://conduct.students.wisc.edu/academic-misconduct/) include, but are not limited to, failure on the assignment/course, written reprimand, disciplinary probation, suspension or expulsion.

The members of the faculty of the Department of Statistics at UW-Madison uphold the highest ethical standards of teaching, data, and research. They expect their students to uphold the same standards of ethical conduct. Standards of ethical conduct in data analysis and data privacy are detailed on the ASA website (<https://www.amstat.org/your-career/ethical-guidelines-for-statistical-practice>), and include:

* Use methodology and data that are relevant and appropriate; without favoritism or prejudice; and in a manner intended to produce valid, interpretable, and reproducible results.
* Be candid about any known or suspected limitations, defects, or biases in the data that may affect the integrity or reliability of the analysis. Obviously, never modify or falsify data.
* Protect the privacy and confidentiality of research subjects and data concerning them, whether obtained from the subjects directly, other persons, or existing records.

Specific examples include, but are not limited to,

1. Copying, plagiarizing, or stealing any of the deliverables from other groups, students outside of the class, or from previous iteration of the course. You cannot ask anyone besides your own group members to clean the data, create tables, figures or plots, or provide parts of their summary. If you are unsure, you are always welcome to ask the teaching staff.
2. Using someone else’s analysis of the dataset, either in its entirety or in parts. You are also not allowed to copy, steal, plagiarize, paraphrase, or use any analysis that was already conducted on this data (or a derivative thereof) from others (e.g., data science courses online, someone’s blog post or R markdown, Google Cloud’s API platform, AWS Machine Learning API, Azure ML, etc.).

However, you are **strongly encouraged** to browse through resources on body fat, health, and other relevant information to gather background information. You are strongly encouraged to use the information from your background research **to complement** your own analysis and **provide proper attributions/references**.

If you have any questions about this, please come talk to the teaching staff.

1. Sabotaging others’ work by deleting, copying, damaging, misrepresenting, or falsifying information about the data, any of the deliverables, or the project.
2. Using ChatGPT or other generative AI to write any portion of your two-page summary or to create any part of your presentation.

However, **for this course only**, you may use ChatGPT or generative AI to help you with your R code, especially to debug R code or to improve your existing R code that you originally wrote. **If you do this, you must provide proper attributions of your code** (e.g. this portion of the code was written by ChatGPT, version BLANK, access date BLANK, prompt BLANK). Also, if you do this, please note that **you ultimately bear the responsibility of the code generated from generative AI**, specifically its correctness and accuracy. We generally discourage you from using ChatGPT to generate all your R code for the project.

By registering for this course, you are implicitly agreeing to conduct yourself with the utmost integrity throughout the semester.