

# Midterm Exam

1. **Format:** Submit the exam to gradescope. For this exam, please make sure that your document has correctly compiled (check all of your figures). Use the `echo = T` option to make sure your code is visible (so that you can be awarded partial credit).
2. **Advice:** Be sure to adequately justify your answers and appropriately reference any sources used. Even if you are not able to answer a question completely, do your best to provide an answer and discuss solutions that you tried. Include your code and graphics for each question and create neat output by using options like `kable()` for tables and writing results in line with R commands.
3. **Resources and Citations:** While the exam is open book and you can use any resources from class or freely available on the internet, this is strictly an individual endeavor and **you should not discuss the problems with anyone outside the course instructor including class members.** All resources, including websites, should be acknowledged.
4. **Exam Questions:** If clarification on questions is required, please email the course instructor: [andrew.hoegh@montana.edu](mailto:andrew.hoegh@montana.edu). Clarifying questions will be addressed, but troubleshooting of R code will not be provided.
5. **A note on sharing / reusing code:** This is a huge volume of code is available on the web to solve any number of problems. For this exam you are allowed to make use of any online resources (e.g., StackOverflow, ChatGPT) but you must explicitly cite where you obtained any code you directly use (or use as inspiration). Any code that is discovered and is not explicitly cited will be treated as plagiarism.

## **Academic Honesty Statement**

Include the following statement at the beginning of your submission.

I, \_\_\_\_ (your full name here) \_\_\_\_, hereby state that I have not communicated with or gained information in any way from my classmates or anyone other than the course instructor during this exam, and that all work is my own or appropriately cited.

In the event that you have inadvertently violated the above statement, you should not sign above and instead discuss the situation with the course instructor.

### Question 1

Recall the questions you answered on the first day of class about your self-reported enjoyment and/or ability in R. The reported values for the class can be seen in Table 1 below. Notice that either you or some of your classmates didn't report actual numbers.

Table 1: Fall 2024 STAT 408 pre-class self-reported R ability and enjoyment scores (on a scale of 1 - 5).

r_ability	r_enjoyment
3.5	3.5
3	4
2	2
4	5
3	4
3	4
3	3
2-3	3-4
3-4	2
2	2
2	2.5
3	5
3	4
3.3	4
3	4-5!
4	4
2-3ish	4
2	2
3-4	4
4	4
3	5
3	3
3	2

Calculate the average r ability of the course. For any value reported as 2-3, 3-4 or similar use the midpoint between the two numbers.

### Question 1a (5 points)

Add a column to the dataset to show the r ability as a numeric value. For full credit, you'll need to manipulate the dataset directly and create a new column that contains the numeric

values. Print your table so it can be seen in the compiled document.

### Question 1b (5 points)

Calculate the average reported reliability of this STAT 408 class. For full credit, report this value in a sentence using the inline r.

### Question 2

For this question, we'll use a dataset on charging for electric vehicles, [https://www.kaggle.com/datasets/valakhorasani/electric-vehicle-charging-patterns?select=ev\\_charging\\_patterns.csv](https://www.kaggle.com/datasets/valakhorasani/electric-vehicle-charging-patterns?select=ev_charging_patterns.csv)

Note: don't be alarmed if the dataset seems a little off. I noticed a few problems, but thought it would still make for an interesting exam question.

```
ev <- read_csv('https://raw.githubusercontent.com/stat408/Data/refs/heads/main/ev_charging_patterns.csv')
```

### Question 2a (5 points)

Note that the dataset contains a `Time of Day` variable with four outcomes: "Evening", "Morning", "Afternoon", and "Night." However, this variable seems incorrect as it doesn't correspond to the `Charging Start Time` in a coherent manner.

Create a new variable where the four times of the day correspond to `Charging Start Time` such that:

- Morning: 4AM - 11:59 AM or in military time (4 - 11:59)
- Afternoon: 12PM - 3:59PM or in military time (12 - 15:59)
- Evening: 4PM - 7:59PM or in military time (16 - 19:59)
- Night: 8PM - 12PM & 12PM - 3:59AM or in military time (20 - 24 and 0 - 3:59)

Print out the first 25 rows of the dataset that contains `Charging Start Time` and this new variable

### Question 2b (5 points)

Create a figure to investigate whether charging times are longer at different times of day. Use your new variable from Question 2A for this question. Make sure your caption addresses whether you see any differences.

### Question 3

For this question, we will use a cheese dataset <https://www.kaggle.com/datasets/noahjanes/canadian-cheese-directory>

```
cheese <- read_csv('https://raw.githubusercontent.com/stat408/Data/refs/heads/main/cheese_data.csv')
```

#### Question 3a (5 points)

Print out a table the average moisture content by MilkTypeEn. Write a written summary describing the results in this table. Hint: You can, and should, filter out NAs.

#### Question 3b (10 points)

Create a single figure that explores MoisturePercent by RindType, MilkTreatmentTypeEN, and FatLevel1. Hint: Again, you can, and should, filter out NAs. Make sure your caption describes the results.

#### Question 3c (5 points)

How many of the 1042 cheeses are classified as a type of cheddar in the CheeseName?

### Question 4

This question uses a bread basket dataset with transactions from a bakery in Edinburgh <https://www.kaggle.com/datasets/mittalvasu95/the-bread-basket>.

- date: date of purchase
- time: time of purchase
- transaction: transaction number (there can be multiple items with each transaction)
- item: item type

```
bread <- read_csv('http://math.montana.edu/ahoegh/teaching/timeseries/data/BreadBasket.csv')
```

#### Question 4a (2 points)

Pose a question.

**Question 4b (10 points)**

Answer your question with a data visualization. For full credit, include informative labels, a caption, and at least one annotation.