Quiz 01

Instructions

Please read the quiz instructions and each question carefully. If you have questions about the quiz, please direct message the teaching team in Canvas. Only clarifications questions will be answered. The quiz must be completed in markdown or quarto and be reproducible (you only need to upload to Gradescope for this quiz and may use any template of your choosing).

The quiz is due on Monday February 10 at 5 PM.

There is no time limit.

The quiz may only be submitted one time. Answers may not be changed once the quiz is submitted.

No late work is accepted on the quiz and no makeups are given for missed quizzes. (Note: The lowest quiz grade will be dropped at the end of the semester.)

The quiz is open book, open note, open internet; however, you maynot discuss the quiz with anyone else. See the course syllabus for more information on the academic honesty policy, which you agree to by submitting your work to both Canvas and Gradescope.

This quiz covers content from lectures that correspond with Beyond Multiple Linear Regression chapters 1 - 3, including all reading from these chapters.

By taking this quiz, you agree to abide by the Duke Community Standard:

To uphold the Duke Community Standard:

I will not lie, cheat, or steal in my academic endeavors;

I will conduct myself honorably in all my endeavors; and

I will act if the Standard is compromised.

Question 1

> **Note**
>
> Which member of the teaching team has office hours Tuesdays 10:30 - 11:30am?

Question 2

> **Note**
>
> If you have a question about the course content, how should you ask it? Select all that apply.

a. Ask in office hours

b. Post in Slack

c. Ask in Slack

Question 3

> **Note**
>
> If a homework assignment is due on Wednesday at 11:59pm, when is the latest it can be turned in for credit?

Question 4

> **Note**
>
> The course policy on the use of artificial intelligence is based on two guiding principles: (1) *Cognitive dimension* and (2) *Ethical dimension*. Briefly describe what each dimension means.

Question 5

> **Note**
>
> Suppose data from 100 randomly selected elementary school students are used to fit a regression model to predict height (in inches) based on age (in months). Which of the following best describes the normality assumption for this model?

Question 6

> **Note**
>
> Let $Y \sim \text{Binomial}(n, \theta)$. Write the likelihood of the Binomial distribution and define it to be $L(\theta)$, where n is known and θ is unknown.

Question 6

> **Note**
>
> Assume the information from question 6. Derive using calculus the maximum likelihood estimator of θ, called $\hat{\theta}_{\text{MLE}}$. Show all steps/derivations for full credit.

Question 7

> **Note**
>
> Provide the **three** assumptions of a generalized linear model that may deviate from a linear model. State the **one** assumption they share in common.

Question 8

> **Note**
>
> Describe, in your own words, the importance of exploratory data analysis (in less than two sentences).

Question 9

> **Note**
>
> Define the maximum likelihood estimator. Is it always optimal or the best estimator that we should consider; explain.

Question 10

> **Note**
>
> Define a likelihood using words (and without notation). Can we make probability statements regarding the likelihood? Explain.