

STA310 Quiz 1

Olivia Fu

2025-02-10

Question 1

No member of the teaching team holds office hours on Tuesdays from 10:30 to 11:30 AM (as indicated in the syllabus).

Question 2

A. Ask in office hours

We can ask and post questions on Ed discussion, instead of Slack.

Question 3

If a homework assignment is due on Wednesday at 11:59 PM, the latest it can be turned in for credit is **Wednesday at 11:59 PM** (the exact time of the deadline), because there are no late extensions on any assignments.

Question 4

The **cognitive dimension** refers to the principle that working with AI should not hinder our thinking and learning. Therefore, if we use AI tools, we should still think independently and critically about the questions. We should not fully rely on AI to handle the thinking process for us.

The **ethical dimension** refers to the principle of being transparent about the use of AI and strictly adhering to academic integrity.

- If we use technology to generate code for assignments, we must explicitly cite the source, such as AI-generated content.
- We are not allowed use AI for narratives, which means we cannot let AI to write or answer questions for us. The written work we submit should reflect our own understanding of the course material.

Question 5

The normality assumption means that the height in inches (response Y) of elementary school students follows a normal distribution at each level/value of age in months (predictor X).

Question 6

The probability mass function of a Binomial random variable is given by:

$$P(Y = y) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}, \quad \text{for } y = 0, 1, \dots, n$$

Since y represents a single observation (and not a set of observations like y_1, y_2, \dots, y_m), the likelihood function is:

$$L(\theta) = P(Y = y) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

The log-likelihood function is:

$$\begin{aligned} \ell(\theta) &= \log(L(\theta)) \\ &= \log \left(\binom{n}{y} \theta^y (1 - \theta)^{n-y} \right) \\ &= \log \left(\binom{n}{y} \right) + \log(\theta^y) + \log((1 - \theta)^{n-y}) \\ &= \log \left(\binom{n}{y} \right) + y \log(\theta) + (n - y) \log(1 - \theta) \end{aligned}$$

The first derivative of the log-likelihood function is:

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{y}{\theta} - \frac{n - y}{1 - \theta}$$

We then set the first derivative equal to zero to find the MLE:

$$\begin{aligned} \frac{y}{\theta} - \frac{n - y}{1 - \theta} &= 0 \\ \Rightarrow \frac{y}{\theta} &= \frac{n - y}{1 - \theta} \\ \Rightarrow y(1 - \theta) &= (n - y)\theta \\ \Rightarrow y - y\theta &= n\theta - y\theta \\ \Rightarrow y &= n\theta \\ \Rightarrow \hat{\theta}_{\text{MLE}} &= \frac{y}{n} \end{aligned}$$

The second derivative of the log-likelihood function is:

$$\frac{\partial^2 \ell(\theta)}{\partial \theta^2} = -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}$$

Since θ is between 0 and 1, and both y and $n-y$ are non-negative (≥ 0), the second derivative is negative, confirming that the critical point corresponds to a maximum.

Question 7

Three assumptions of a generalized linear model that may deviate from a linear model

- In generalized linear models, the relationship between the response and predictor(s) can be nonlinear. On the other hand, a linear model assumes a linear relationship between the mean of the response Y and the predictor X .
- Generalized linear models allow the response variable to be non-normal, such as binary outcomes. In contrast, linear models are constrained by normality assumption, meaning the response variable Y follows a normal distribution at each level of the predictor X .
- In generalized linear models, the variance of Y at each level of X can differ. Linear models, however, assume equal variance, which means the variance of the response Y remains constant across all levels of the predictor X .

One assumption they share in common

- The independence assumption must hold for both models. That means each observation of the predictor X and response Y pair is independent of the others. There is no connection between how far any two data points lie from the regression line.

Question 8

Firstly, univariate exploratory data analysis helps us understand the **distribution** of specific variables of interest, along with key characteristics such as mean, median, and variance, which are important for us to choose appropriate analysis techniques.

Additionally, bivariate exploratory data analysis allows us to identify **relationships** between variables (response Y and predictors X), which can provide basic information about the correlation—whether increasing or decreasing, linear or quadratic—and pave the foundation for further regression analysis.

Question 9

A likelihood is function that tells us how likely we are to observe our data for a given parameter value. Maximum likelihood estimator, as implied by its name, is the value of the parameter that **maximizes** the likelihood function. Therefore, the MLE of the unknown parameter is the value that makes the observed data most probable.

No, MLE is not always the best estimator to consider.

- Firstly, MLE assumes that we know the correct model or distribution and can explicitly write out the likelihood function. However, in some cases, the distribution of the data is unclear, or it may be difficult or even impossible to specify an accurate likelihood function, which can make the MLE biased or even non-existent.
- Additionally, when the likelihood function is extremely complex (e.g., high-dimensional parameter space), MLE can be computationally expensive to calculate.
- Also, when the number of observations (sample size) is very small, MLE can be biased, as it relies solely on the given (empirical) data to find the parameter value that maximizes the likelihood.

Question 10

A likelihood is function that tells us how likely we are to observe our data for a given parameter value.

No, the likelihood function is not a probability. A probability function has fixed parameter values and takes different observed values as inputs, returning the probability of those observations given a specific parameter value. In contrast, a likelihood function is constructed by fixing the observed data and taking different parameter values, returning the likelihood of the fixed data for each possible parameter value.