Midterm 2

2023-05-10

Part 1: T/F Problems

1. A p-value of 0.05 means there is a 5% chance that the null hypothesis is true.

(False)

2. For a dataset listing the students' ID and students' letter grade, we can use `geom_histogram` in `ggplot2` package to visualize the grade distribution of the students.

(False)

3. In the dataset `nycflights` in package `nycflights13`, command `dim(flights)` give you the result: `[1] 336776 19`, based on the result we should use `geom_hex` to visualize the relationship between departure delay and arrival delay.

(True)

4. A z-score is a measure of how many standard deviations a data point is from the mean.

(True)

5. A Type I error occurs when we reject an alternative hypothesis that is actually true.

(False)

6. The null hypothesis is always the hypothesis we want to prove.

(False)

7. You calculate the sample correlation on the variable of study hours per week and the variable of final exam score. The result correlation is 0.7653. So you proved study hard will cause students to get higher final exam score.

(False)

8. In homework 5 we used the Monte Carlo simulation to calculate complex function integration. Due to the central limit theorem (CLT), the more points we simulate, the more precise the integration result will be.

(False)

9. `for` loop, `sapply`, and `replicate` can be used to repetitively execute some functions.

(True)

10. We use pipes to add additional layers (such as labels, texts, lines) when we using `ggplot2` to generate plots.

(False)

11. The Poisson distribution can be used to model the number of defects in a manufacturing process. For this Poisson distribution, its mean is equal to its variance.

(True)

12. For Binomial distribution $Y \sim Binomial(n, p)$, there is only 1 parameter.

(True)

Part 2: Multiple Choice Problems

Part 3: Code sorting problem

Copy your solution here for grading

1-12											
13-24											

Solution for code sorting