AUA Fall 2023
DS 110 Statistics 2
Homework 2

- For the problems 1-2 show all calculations on the paper.
- Please use R or Python only in the case the statement of the problem contains (R, Python) at the beginning.
- You can complete all tasks, including non-coding tasks, using Jupyter Notebook or R Markdown.
- If you choose to submit online, the final solution should be converted to PDF format.
- For offline submission, you must submit a hard copy of the homework paper by the deadline

Problem 1 (5 points)

What is the number of dummy variables of the regression model for a categorical independent variable with n categories? Explain your answer.

Problem 2 (5 points)

Consider the following model $Ln(y) = \beta_0 + \beta_1 x$. If x is a dummy variable, how to interpret β_1?

Problem 3 (R, Python) (30 points)

The file houseprice.csv contains data on 1,080 houses sold in Baton Rouge, Louisiana, during mid-2005.We will be concerned with the selling price (PRICE), the size of the house in square feet (SQFT), and the age of the house in years (AGE). Use all observations to estimate the following regression model and report the results

$$Price = \beta_1 + \beta_2 Sqft + \beta_3 Age + \gamma(Sqft \cdot Age)$$

1. Explain the meaning of γ and its estimate. (5 points)

2. Find estimates of the marginal effect of Age on $Price$ for the smallest and largest houses in the sample. Comment on these values? (10 points)

3. Find estimate of the marginal effect $Sqft$ on $Price$ for a house that is 20 years old and interpret it. (5 points)

4. Test the joint null hypothesis about non significance of β_2 and β_3 at the level of significance 0.05. (10 points)

Problem 4 (R, Python) (60 points)

In September 1998, a local TV station contacted an econometrician to analyze some data for them. They were going to do a Halloween story on the legend of full moons' affecting behavior in strange ways. They collected data from a local hospital on emergency room cases for the period from January 1, 1998, until mid-August. There were 229 observations (fullmoon.csv). During this time, there were eight full moons and seven new moons (a related myth concerns new moons) and three holidays (New Year's Day, Memorial Day, and Easter). If there is a full-moon effect, then hospital administrators will adjust numbers of emergency room doctors and nurses, and local police may change the number of officers on duty.

1. Examine the histogram of Cases. What do you observe? Create the variable $ln(Cases)$ and examine its histogram. Comment on the difference. (10 points)

2. Estimate a regression of $ln(Cases)$ on the remaining variables. Discuss the estimation results. Comment on the signs and significance of all variables. (10 points)

3. Discuss the effect of full moon on the emergency room cases. (10 points)

4. Test a joint hypothesis of non-significance of FRIDAY and SATURDAY variables. (10 points)

5. Drop FRIDAY from the model and re-estimate the equation. Now, what is the estimated effect of a full moon? Discuss why the estimate is different from that obtained in part (2). (10 points)

6. Transform the numeric $time$ variable (time trend, $time = 1, 2, \dots, 229$) to the interval variable with 0-100 days, 101-200 days, 200+ days. In regression of point 2) replace initial $time$ variable with this new $time$ variable. Interpret the coefficients of the $new\ time$ variable. (10 points)