

W271 Group Lab

Bike share demand

Please fill in with your names.

Contents

1	Introduction (5 points)	1
2	Data (20 points)	2
2.1	Description (5 points)	2
2.2	EDA (15 points)	2
3	Model Development (40 points)	2
3.1	Poisson regression (10 points)	2
3.2	Model Comparison (10 points)	2
3.3	Model Assessment (10 points)	3
3.4	Alternative Specification (10 points)	3
4	Conclusion (5 points)	3

1 Introduction (5 points)

Complete the following task. In your final submission, please remove this question prompt so that your report reads as a report.

The `SeoulBikeData` data set measures the number of bike rentals (bikers) per hour in Seoul, South Korea, from December 1, 2017, to November 30, 2018. There are 8760 records/observations and 14 variables in this data set. The ‘Seoul Bike Sharing Demand Data Set’ can be downloaded from UC Irvine Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets/Seoul+Bike+Sharing+Demand>

Initiate your analysis by formulating a hypothesis related to bike share demand that can be tested using a regression model. This hypothesis aims to provide decision-makers with insights into the factors influencing bike usage and demand, facilitating the optimal availability of rental bikes to the public when needed.

2 Data (20 points)

2.1 Description (5 points)

Complete the following task. In your final submission, please remove this question prompt so that your report reads as a report.

Describe the data that you are using. How is this data generated, what is the sampling process that brought it to your availability, etc. If it is helpful, you might describe the population (i.e. the Random Variables) that exist and how samples are produced from these random variables.

2.2 EDA (15 points)

Complete the following task. In your final submission, please remove this question prompt so that your report reads as a report.

Examine the dataset and investigate anomalies, missing values, and convert variables when necessary for future analysis.

Conduct a thorough EDA of the data set, giving special attention to the relationships relevant to your hypothesis.

- This should include both graphical and tabular analysis as taught in this course.
- Since the report has a page-limit, you will have to be selective when choosing visuals to illustrate your key points, associated with a concise explanation of the visuals.

3 Model Development (40 points)

3.1 Poisson regression (10 points)

Complete the following task. In your final submission, please remove this question prompt so that your report reads as a report.

Estimate a simple poisson regression model (`model_poisson_1`) that includes variables as linear terms in the model based on what is essential for your hypothesis. Evaluate, using likelihood ratio tests and the statistical significance of each explanatory variable in the model. Also evaluate, using the context and data understanding that you have created in the Data section of this report, the practical significance of each explanatory variable in the model.

3.2 Model Comparison (10 points)

Expand your initial poisson model through a two-step process:

- Incorporate additional variables from the dataset to avoid potential omitted variable bias. (`model_poisson_2`)

- Introduce quadratic terms or interactions when they can potentially explain variations in bike rental demand. For instance, you might consider assessing whether the effect of Hour on bike demand differs between weekdays and weekends. (`model_poisson_3`)

Finally, use three information criteria for the three models (`model_poisson_1`, `model_poisson_2`, `model_poisson_3`) and rank the models based on each criterion. Are the results consistent? Why or why not?

3.3 Model Assessment (10 points)

Complete the following task. In your final submission, please remove this question prompt so that your report reads as a report.

Use suitable tools like residual or standardized residual plots to assess if your final model is a good approximation of the data. Additionally, check the results for indications of potential overdispersion.

- How would you address any potential issues, including any overdispersion?
- Produce an adjusted version of your best model that accounts for issues noted above.

3.4 Alternative Specification (10 points)

Complete the following task. In your final submission, please remove this question prompt so that your report reads as a report.

Using the same set of explanatory variables from your final model, estimate a linear regression model. Examine the summary of the model and interpret the coefficients. Obtain fitted values from both OLS and Poisson regression models and plot them. Discuss which regression model, linear or Poisson, is more suitable for this case and explain why.

4 Conclusion (5 points)

Complete the following task. In your final submission, please remove this question prompt so that your report reads as a report.

In conclusion, summarize your entire work to leave a lasting impression on the reader. Ensure you:

- Restate the main hypothesis or research question.
- Report OLS, Poisson, or any other regressions you estimated.
- Summarize key findings.
- Discuss implications.
- Address limitations.