# Linear Regression Assumptions Lesson

|  |  |  |  |
| --- | --- | --- | --- |
| **Learning Objectives** | **Discussion(s)** | **Assignment(s)** | **Learning Resources + Media** |
| After this lesson, you will be able to …   * Explain what linear regression is and how it can be used * Evaluate where strong correlations exist * Know what requirements a dataset must meet in order to use linear regression * Understand the importance of practicing good data science | What are some potential uses of linear regression in our field? Why is using linear regression beneficial in these examples?  Why is it important to practice good data science? Discuss the importance of addressing missing data, outliers and high leverage points, and the linear regression assumptions | **Assignment 1**  Have students select whichever level of lesson they are comfortable with. Have students answer the questions in each lesson by walking through the R notebook.  **Bonus Assignment**  Students comfortable with programming can change the code in the R notebook to read in a new dataset. Find a dataset relevant to your field, especially one where linear regression is already applied. See if missing data and outliers were addressed, and if the linear regression assumptions were met. If not, what does this say about the validity of the linear regression model? | **Articles:**   * Linear Regression: <https://www.geeksforgeeks.org/ml-linear-regression/#:~:text=Linear%20Regression%20is%20a%20machine,value%20based%20on%20independent%20variables.&text=Linear%20regression%20performs%20the%20task,given%20independent%20variable%20(x)>. * Linear Regression for Machine Learning: <https://machinelearningmastery.com/linear-regression-for-machine-learning/> * MCAR: <http://www.hubresearch.ca/bridging-the-data-gap-how-to-deal-with-missing-data-in-observational-studies/#:~:text=As%20a%20rule%20of%20thumb,any%20significant%20ramifications%20(3).&text=In%20this%20case%2C%20it%20is,fill%20in%20the%20missing%20data>. * Types of Missing Data: <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-017-0442-1> * Imputation Methods: <http://www.stat.columbia.edu/~gelman/arm/missing.pdf> * Cook’s Distance: <https://www.statisticshowto.com/cooks-distance/> * Cook’s Distance: <https://online.stat.psu.edu/stat462/node/173/> * Linear Regression Assumptions: <http://www.sthda.com/english/articles/39-regression-model-diagnostics/161-linear-regression-assumptions-and-diagnostics-in-r-essentials/> * Linear Regression Assumptions: <https://blog.uwgb.edu/bansalg/statistics-data-analytics/linear-regression/what-are-the-four-assumptions-of-linear-regression/> * Residuals: <https://www.statisticshowto.com/residual/#:~:text=A%20residual%20is%20the%20vertical,at%20that%20point%20is%20zero>. * Linear Regression Results: <http://www.learnbymarketing.com/tutorials/linear-regression-in-r/#:~:text=lm()%20Function-,Linear%20Regression%20Example%20in%20R%20using%20lm()%20Function,variable%20from%20your%20new%20model>. |