# Weekly Exercise - Week 1

Julian D. Karch

February 12, 2023

Write a function that implements the following pseudo code:

- Input sample size $n$.
- Generate $n$ observations with a single predictor variable $x \sim \mathcal{U}(-3, 3)$
- Generate the response $y = 8 \sin(x) + \epsilon$, with $\epsilon \sim \mathcal{N}(0, 1)$
- return x, y as a `data.frame` or `pandas.DataFrame`.

Use this function to generate a training set of size 50 and a test set of size 10000. Train two polynomial regression models, one with a degree of 3 and one with a degree of 15, and estimate their mean squared errors (MSE) on the test set. Repeat the process with a training set of size 10000. Answer the following questions:

- What is the best possible prediction rule $f$ in this case? Obtain the test MSE of the best prediction rule as well. (BONUS: Explain, considering how we generated the data, why this value is not a surprise.)
- Report the 4 obtained test mean squared error values (for degree 3 and 15 and for both training set sizes). Explain the obtained numbers using bias and variance. Hints: Start with the results for the large training set, taking the optimal MSE into account as well can help.

**Generate and upload one pdf using either `RMarkdown` or `Python Notebook` including both your code as well as the textual answers.**