

Assignment4 — Linear Regression and Hypothesis Test

We will analyze how a car's **speed (mph)** affects its **stopping distance (ft)** using the file **cars.csv**.

Academic honesty: You must complete the work on your own. You may discuss questions with classmates or visit office hours. However, if you and a classmate submit identical solutions, it will be treated as cheating and will receive zero credit—unless you can prove the similarity is an extraordinary coincidence. For more complex assignments or projects, collaboration may be allowed and even encouraged.

Folder submission: Place all files in one folder. Name the folder exactly as your name appears in NEIUPORT. **For example, if your name is Jessica Fatima, submit the folder as Jessica_Fatima.**

Due date: Friday (11/28).

Grace period: You may submit up to three days late without a penalty of 20%.

Late policy: Submissions more than one week past the due date will not be accepted or graded.

Deliverables

- **Python notebook** performing Steps 1 – 4 and printing the required numbers. (60 pts)
 - **One-page write-up** answering Steps 2, 3, 5, 6 clearly and concisely. (40 pts)
 - **Do not report any additional statistics.**
-

Instructions

To get full credit for this assignment, you are required to use the designated statistical library called **statsmodels** we have demonstrated using the file named “linear_regression.py”. This file can be found inside the folder named “Week 13: Regression” on D2L.

1. Import the data using the following code

```
import pandas as pd
```

```
df = pd.read_csv('cars.csv')
```

2. Visual inspection

Create a scatter-plot of **Stopping Distance** (y-axis) versus **Speed** (x-axis). In one sentence, comment on whether a straight-line trend seems reasonable.

3. Model statement

State the simple linear equation you will estimate.

4. Estimate coefficients

Using the code provided on D2L, compute and report the intercept and slope to three decimal places.

5. Hypothesis test for the slope ($\alpha = 0.05$)

- Clearly state the null and alternative hypothesis.
- Should you use Z-statistic or t-statistic? Report the appropriate statistic and p-value, then state “**Reject the null**” or “**Fail to reject the null**” with a one-sentence justification.
- HINT: After `model = sm.OLS(...).fit()`, inspect `model.tvalues` and `model.pvalues` to find the t-statistic and p-value for each coefficient.

6. R^2 (coefficient of determination):

Report the R^2 value and interpret the percentage of variation in stopping distance it explains.