

Assignment4 — Linear Regression and Hypothesis Test

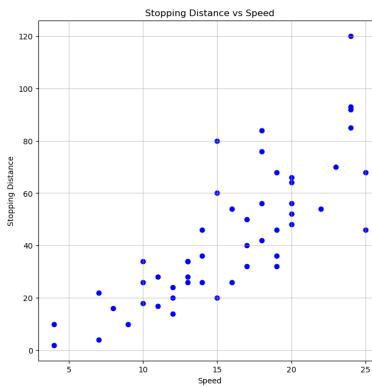
1. Import the data using the following code

```

import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
#1. Load Data
df = pd.read_csv('cars.csv')
# Display the first few rows to check
df.head()

```

2. Visual inspection



The graph clearly shows a trend that as speed increases, stopping distance increases too;
A straight line trend is reasonable.

3. Model statement

Simple Linear equation is: Distance = B₀ + B₁ * Speed + e

Distance is Dependent and Speed is Independent

4. Estimate coefficients

Intercept (B₀): -16.516

Slope (B₁): 3.828

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

H₀: Speed has no effect on stopping distance. B₁ = 0

H₁: Speed DOES have an effect on stopping distance. B₁ ≠ 0

We use the t-stat because it is a fairly small set of data and we don't know the variance

t-statistic: 9.361337676953

p-value: 2.099716476318138e-12 Very Small!

Reject the null hypothesis.

Since p < 0.05, there is a strong relationship between speed and stopping distance.

6. R² (coefficient of determination):

R-squared = 0.646

This means that ~65% of the variation in stopping distance is determined by the speed of the vehicle.