

# Tutorial 1

Statistics 2 for IBA

Tilburg University

# Introduction

- In this tutorial we will explore the `f150.sav` dataset, which featured in the Resit exam in January 2021.
- The dataset contains information from advertisements for second-hand Ford F-150 trucks on Craigslist, a US website similar to Marktplaats here in NL.
- The data contains the asking price, the year of manufacture, the mileage (odometer reading), the truck's color and an indicator for whether truck is in good condition or not.

# Dataset Description as seen in TestVision

## Introduction

The data set for this exam can be downloaded by clicking on the following link: [f150](#)

Craigslist is a popular website in the United States where people can place classified advertisements to sell second-hand items. It is similar to *marktplaats.nl* in the Netherlands. Many second-hand cars are advertised on the website. The most common vehicle advertised on the website is the Ford F-150 pickup truck. Here is an example of what a Ford F-150 pickup truck looks like:



You are interested in understanding how characteristics of a second-hand truck, such as its age and mileage, affect the price it can sell for. You have data on 500 different Ford F-150 advertisements on Craigslist, where for each advertisement you observe the following variables:

<i>price</i>	The asking price of the truck
<i>year</i>	The year the truck was purchased new
<i>odometer</i>	The total number of miles the truck has driven
<i>paint_color</i>	The color of the truck (character/string variable)
<i>good_condition</i>	Dummy variable for if the truck is in good condition or not (=1 for good condition and =0 for not in good condition).

The advertisements were collected at the end of the year in 2020. Therefore if *year* is equal to 2019, the *age* of the truck is 1 year. If *year* is equal to 2018, the *age* of the truck is 2 years, and so on.

## Exercises

1. Open the `f150.sav` data file and inspect the data. Look at *Data View* and *Variable View*.
2. Create a histogram of the variable *Price* using *Graphs*→*Legacy Dialogs*→*Histogram*.
3. Obtain descriptive statistics (mean/min/max/SD) of the variable *Price*.
4. Create a scatter plot of *Year* against *Price*, with *Year* on the horizontal axis and price on the vertical axis. Interpret it.
5. Compute the covariance and correlation between *Year* and *Price*. Interpret them.
6. Create the variable *Age* from *Year*. Keep in mind that the advertisements were shown in 2020.
7. Compute the covariance and correlation between *Age* and *Price*. Relate this to what you found in Q5.
8. Study the relationship between *Age* and *Year*. Create a scatter plot and compute the covariance and correlation. Interpret them.

## Bonus Questions: Q1 from the January 2021 Resit

What is the sample correlation coefficient between *price* and *odometer*?

123

You were asked to type a number into the box.

## Bonus Questions: Q2 from the January 2021 Resit

Choose the answer below which best describes the interpretation of the sample correlation coefficient between *price* and *year*.

**Note:** See the block intro for how the *year* of an F-150 relates to its *age*.

- ☐ There is a negative linear relationship between the *age* of an F-150 and its price.
- ☐ A one-year increase in the *age* of an F-150 on average increases its price.
- ☐ There is no relationship between the price of an F-150 and its *age*.
- ☐ There is a positive linear relationship between the *age* of an F-150 and its price.

Multiple choice question.