Tutku Öztel, 62234. INTL550 Homework 2 Report

Hypothesis. As the amount of km gone with a second hand car increases, the price of the second hand car will decrease, controlled for the car brand.

Data retrieval procedure. Price, km and brand information of various second hand cars were retrieved from https://www.araba.com/otomobil?sayfa=2.

Brand decoding. Brand information of the car was retrieved as character variable (i.e., string). To convert brand variable to an appropriate data type for the multiple regression modeling approach, each unique brand datum was given a unique numerical tag. As a result, the data was converted in a numerical type to allow matrix multiplication, but was in nominal type in nature, where, in essence, numerical tags did not have any real "numerical" value.

Hypothesis testing. Multiple regression analysis was performed using the km of the second hand car as a predictor of the price, controlling for car brand. The multiple regression model performed is as depicted below:

Multiple Regression Model:

$$price \sim km + brand + constant + e$$

where price is the outcome variable, e is the error term and ~ stands for "predicted from". The depicted model was fit to raw data (as the brand variable was in nominal type, it could not be scaled).

Results. There was an expected negative relationship between km and price when assuming constant brand (ß = -0.437, SE = 0.0147, 95% CI = [-0.770, -0.105]). However, the brand of the car was not a significant predictor of the price when assuming constant km (ß =7272.108, SE = 3483.374, 95% CI = [-0.607.832, 15152.048]).

The linear relationship between km and price is as illustrated below:

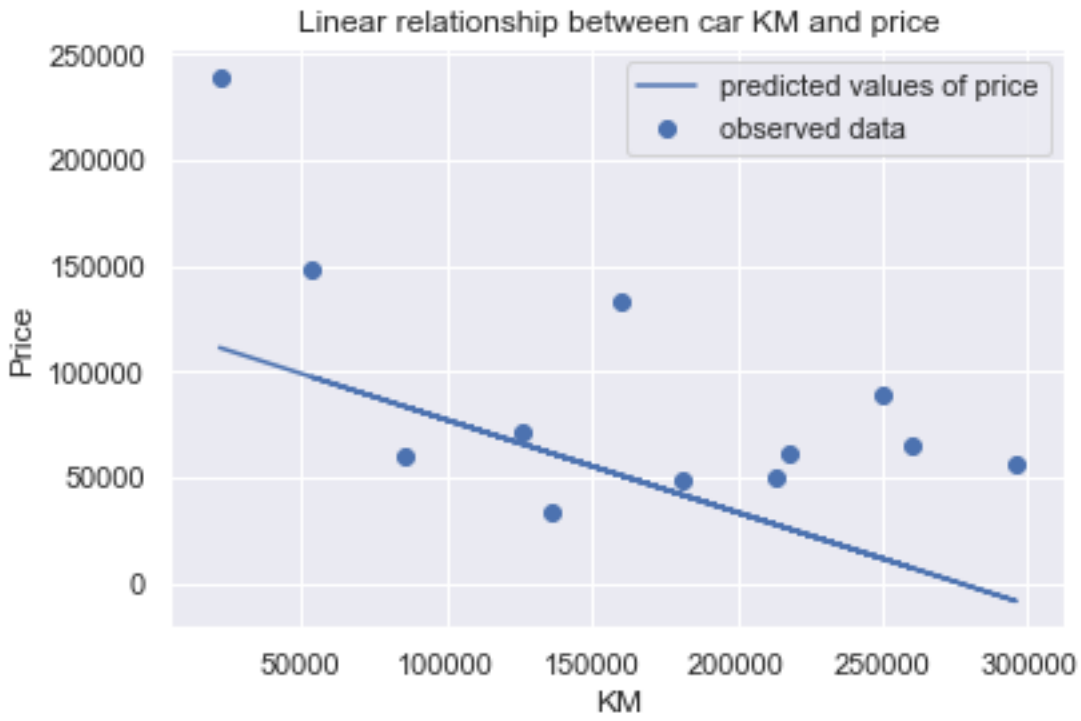

Figure. The linear relationship between car km and price.