ps3

October 31, 2023

1 BA222 Problem Set 3

Ethan Chang, Fall 2023

```
[1]: import pandas as pd
     import pathlib
     import statsmodels.api as sm
```

```
[2]: path = pathlib.Path().cwd().parents[1] / 'CSVs' / 'USCarData.csv'
     df = pd.read_csv(path)
     df.head()
```

```
[2]:    price      brand    model  year  title_status  mileage   color
     0   6300     toyota  cruiser  2008  clean vehicle   274117   black  \
     1   2899       ford       se  2011  clean vehicle   190552  silver
     2   5350      dodge      mpv  2018  clean vehicle    39590  silver
     3  25000       ford     door  2014  clean vehicle    64146    blue
     4  27700  chevrolet     1500  2018  clean vehicle     6654     red

                     vin        lot        state country        condition
     0  jtezu11f88k007763  159348797  new jersey     usa    10 days left
     1  2fmdk3gc4bbb02217  166951262    tennessee     usa     6 days left
     2  3c4pdcgg5jt346413  167655728      georgia     usa     2 days left
     3  1ftfw1et4efc23745  167753855     virginia     usa  22 hours left
     4  3gcpcrec2jg473991  167763266      florida     usa  22 hours left
```

```
[3]: price = pd.DataFrame(df['price'])
     mileage = pd.DataFrame(df['mileage'])
```

```
[4]: X = sm.add_constant(mileage)
     Y = price

     model = sm.OLS(Y, X).fit()

     display(model.summary())

     print(f'The linear regression model is: \n Y = {model.params[0]:.3f} + {model.
       ↪params[1]:.3f} * X')
```

	coef	std err	t	P> |t|	[0.025	0.975]
Dep. Variable:	price		**R-squared:**		0.161	
Model:	OLS		**Adj. R-squared:**		0.160	
Method:	Least Squares		**F-statistic:**		478.0	
Date:	Tue, 31 Oct 2023		**Prob (F-statistic):**		4.25e-97	
Time:	10:31:30		**Log-Likelihood:**		-26823.	
No. Observations:	2499		**AIC:**		5.365e+04	
Df Residuals:	2497		**BIC:**		5.366e+04	
Df Model:	1					
Covariance Type:	nonrobust					

	coef	std err	t	P> |t|	[0.025	0.975]
const	2.302e+04	295.272	77.968	0.000	2.24e+04	2.36e+04
mileage	-0.0813	0.004	-21.863	0.000	-0.089	-0.074

Omnibus:	584.790	**Durbin-Watson:**	1.695
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1442.098
Skew:	1.269	**Prob(JB):**	0.00
Kurtosis:	5.721	**Cond. No.**	1.06e+05

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.06e+05. This might indicate that there are strong multicollinearity or other numerical problems.

```
The linear regression model is:
 Y = 23021.775 + -0.081 * X
```

1.0.1 Homework response

The coefficient of mileage is the change in price for each additional mile on a vehicle. For example, if a car had 100 more miles than a similar car, we could predict that the car with more miles would cost approximately $8 less.

In my opinion it is common knowledge that a more used item will fetch a lower price on the resale market. Therefore, I believe that the relationship between mileage and price is a causal one.