Jacob Miller - Homework 6

April 7, 2020

1 Problem 8.5

Import required packages:

[1]:
```
import pandas as pd
import numpy as np
import patsy
import statsmodels.api as sm
```

Read in data, create X matrix and y vector:

[2]:
```
df = pd.read_excel('Data/data-table-B3.xlsx')
y, X = patsy.dmatrices('y ~ x10 + x11', df)
```

1.1 8.5.a

Create linear regression model, print out results table:

[3]:
```
model = sm.OLS(y, X)
results = model.fit()
results.model.data.design_info = X.design_info
print(results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.760
Model:                            OLS   Adj. R-squared:                  0.743
Method:                 Least Squares   F-statistic:                     45.91
Date:                Tue, 07 Apr 2020   Prob (F-statistic):           1.03e-09
Time:                        15:13:41   Log-Likelihood:                -81.056
No. Observations:                  32   AIC:                             168.1
Df Residuals:                      29   BIC:                             172.5
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     39.1919      2.557     15.327      0.000      33.962      44.422
x10           -0.0047      0.001     -4.975      0.000      -0.007      -0.003
```

```
x11           -2.6958      1.981     -1.361      0.184      -6.747       1.355
==============================================================================
Omnibus:                        0.125   Durbin-Watson:                   2.164
Prob(Omnibus):                  0.940   Jarque-Bera (JB):                0.292
Skew:                           0.120   Prob(JB):                        0.864
Kurtosis:                       2.598   Cond. No.                     1.85e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.85e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Linear regression model, transmission not significant:

[4]:
```
coef = np.round(results.params, 3)
print('\ny = {} + ({} * x10) + ({} * x11)'.format(coef[0], coef[1], coef[2]))
print('No. High p-value [{}], so transmission not significant\n'.\
      format(round(results.pvalues[2], 3)))
```

```
y = 39.192 + (-0.005 * x10) + (-2.696 * x11)
No. High p-value [0.184], so transmission not significant
```

1.2 8.5.b

Add interaction term, create linear regression model, print out results:

[5]:
```
y, X = patsy.dmatrices('y ~ x10 + x11 + x10 * x11', df)

# Create model with interaction
model = sm.OLS(y, X)
results = model.fit()
results.model.data.design_info = X.design_info
print(results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.849
Model:                            OLS   Adj. R-squared:                  0.833
Method:                 Least Squares   F-statistic:                     52.63
Date:                Tue, 07 Apr 2020   Prob (F-statistic):           1.24e-11
Time:                        15:13:41   Log-Likelihood:                -73.602
No. Observations:                  32   AIC:                             155.2
Df Residuals:                      28   BIC:                             161.1
Df Model:                           3
Covariance Type:            nonrobust
```

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     58.1084      5.078     11.443      0.000      47.707      68.510
x10           -0.0125      0.002     -6.090      0.000      -0.017      -0.008
x11          -26.7249      6.107     -4.376      0.000     -39.235     -14.215
x10:x11        0.0090      0.002      4.076      0.000       0.004       0.014
==============================================================================
Omnibus:                        0.087   Durbin-Watson:                   2.549
Prob(Omnibus):                  0.958   Jarque-Bera (JB):                0.291
Skew:                          -0.067   Prob(JB):                        0.865
Kurtosis:                       2.552   Cond. No.                     8.33e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 8.33e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Separate regression models for automatic and manual transmissions:

[6]:
```
# Linear regression model with interaction
coef = np.round(results.params, 3)
print('\ny = {} + ({} * x10) + ({} * x11) + ({} * x10 * x11)'.\
      format(coef[0], coef[1], coef[2], coef[3]))
print('\nFor automatic transmission, x11 = 1, therefore:')
print('y = {} + ({} * x10)'.format(round(coef[0] + coef[2], 3),
                                   round(coef[1] + coef[3], 3)))
print('\nFor manual transmission, x11 = 0, therefore:')
print('y = {} + ({} * x10)'.format(coef[0], coef[1]))
```

```
y = 58.108 + (-0.013 * x10) + (-26.725 * x11) + (0.009 * x10 * x11)

For automatic transmission, x11 = 1, therefore:
y = 31.383 + (-0.004 * x10)

For manual transmission, x11 = 0, therefore:
y = 58.108 + (-0.013 * x10)
```

For manual transmission engines, gasoline mileage decreases more quickly.