HW1 Report

안예영 (2022311154)
2022-09-12

1. We first write a series of code in order to prompt the user for necessary information such as file name, separator type, response variable position, etc.

Prompt

```
In [2]: import_name=input("Please enter the name of the data file: ")
Please enter the name of the data file: harris.dat

In [3]: import_name
Out[3]: 'harris.dat'

In [4]: encode_sep=int(input("Please select the separator used in the data file (1 = whitespace or 2 = comma): "))
separator={encode_sep ==1 : " "}.get(True, ",")
Please select the separator used in the data file (1 = whitespace or 2 = comma): 1

In [5]: res_pos =int(input("Please enter the position of the response variable column (select from 1 to p): "))
Please enter the position of the response variable column (select from 1 to p): 1

In [6]: header=input("Does the data file include a column header? (y/n) : ")
Does the data file include a column header? (y/n) : n

In [7]: export_name=input("Please enter the name for the file to be exported (e.g. result.txt) : ")
Please enter the name for the file to be exported (e.g. result.txt) : HW1_output.txt
```

Next, we use the input data to formulate the response variable and predictor variables. We use the matrix approach such that $\hat{Y} = Zb$ where $b = (Z'Z)^{-1}Z'Y$. R^2 and MSE can each be calculated as $R^2 = 1 - \frac{SSE}{SST}$ and $MSE = \frac{SSE}{n-p}$, respectively, where p is the number of predictor variables.

```
In [14]: #Fitted values
Y_hat = Z.dot(B_hat)
Y_hat=round(Y_hat,3)
Y_hat
```

Out[14]:

	0
0	4630.068
1	4646.300
2	5315.187
3	4418.292
4	4396.536
...	...
88	6247.148
89	5815.688
90	5785.258
91	6328.435
92	6530.762

93 rows × 1 columns

```
In [15]: #Calculate R^2
SSE = sum((Y-Y_hat[0])**2)
SST = sum((Y-Y.mean())**2)
R_square = round(1-SSE/SST,4)
R_square
Out[15]: 0.5109

In [16]: #Calculate MSE
n = len(data)
p = len(Z.columns)
MSE = round(SSE/(n-p),4)
MSE
Out[16]: 257476.5621
```

We write our results on a separate file and export the file as shown below. The export file is also shown below.

Write File ¶

In [17]:
```
f = open(export_name,'w')
text = "Coefficients\n-------------\nConstant: "+str(B_hat.iloc[0,0])+"\n"
for i in range(1,len(Z.columns)):
    text += "Beta"+str(i)+": "+str(B_hat.iloc[i,0])+"\n"
text += "\nID, Actual values, Fitted values\n--------------------------------\n"
for j in range(5):
    text += str(j+1) + ", " + str(Y[j]) + ", " + str(Y_hat.iloc[j,0]) + "\n"
text += "(continue)\n\nModel Summary\n-------------\nR-square = " + str(R_square) + "\n"
text += "MSE = " + str(MSE)
f.write(text)
f.close()
```

Coefficients

Constant: 3526.422
Beta1: 90.02
Beta2: 1.269
Beta3: 23.406
Beta4: 722.461

ID, Actual values, Fitted values

1, 3900, 4630.068
2, 4020, 4646.3
3, 4290, 5315.187
4, 4380, 4418.292
5, 4380, 4396.536
(continue)

Model Summary

R-square = 0.5109
MSE = 257476.5621