Strategy and Performance Management

Prediction champion - Timp Health II - Group Pre-Assignment

Group 4 : Kirtesh Patel, Nils Marthiensen, Neelesh Bhalla, Chia-Jung Chang

GitHub - https://github.com/nilsmart96/fs-spm-assignments/tree/main

```
# Import of necessary libraries

import pandas as pd
import numpy as np
import statsmodels.api as sm

# The equation obtained from our model is
print('log(Predicted price) = 0.0451*Month + 0.2229*NLISDummy +
0.5723*LISCHOSERDummy + 0.0009*RiskScore + 1.2389*SpecialtyDummy +
0.1841*AdjudicationDays + -0.0052*Age + -0.0672*Gender')
print('OR')
print('Predicted price = exp(0.0451*Month + 0.2229*NLISDummy +
0.5723*LISCHOSERDummy + 0.0009*RiskScore + 1.2389*SpecialtyDummy +
0.1841*AdjudicationDays + -0.0052*Age + -0.0672*Gender)')
```

```
log(Predicted price) = 0.0451*Month + 0.2229*NLISDummy +
0.5723*LISCHOSERDummy + 0.0009*RiskScore + 1.2389*SpecialtyDummy +
0.1841*AdjudicationDays + -0.0052*Age + -0.0672*Gender
OR
Predicted price = exp(0.0451*Month + 0.2229*NLISDummy +
0.5723*LISCHOSERDummy + 0.0009*RiskScore + 1.2389*SpecialtyDummy +
0.1841*AdjudicationDays + -0.0052*Age + -0.0672*Gender)
```

```
# Import of the data and removal of non-relevent columns
data_url =
'https://github.com/nilsmart96/fs-spm-assignments/blob/main/Prediction
_champion-Timp_Health_II-%20Group_Pre-Assignment/Timp_Health_data%20-
%20including%20month%207-1.txt?raw=true'
df = pd.read_csv(data_url, delimiter='\t')
df.drop(['RecordID','MemberID'],axis=1,inplace=True)
df
```

```
       Month  GrossDrugCost  NLISDummy  LISCHOSERDummy  RiskScore  \
0          7         184.26          0               0      336.0
1          6        1242.17          0               0      668.4
2          1         625.86          0               0      290.0
3          6          27.91          0               0      477.2
4          6       46451.23          0               0     2135.6
...      ...            ...        ...             ...        ...
52069      3         337.39          1               0      245.6
```

```
52070      4          358.93           1                0      245.6
52071      5           73.92           1                0      245.6
52072      6          771.96           1                0      245.6
52073      7          861.55           1                0      245.6
```

```
       SpecialtyDummy  AdjudicationDays  Age  Gender  FrailityDummy  \
0                   0                21   52       1              0
1                   1                21   96       1              0
2                   1                21   59       1              0
3                   1                21   43       1              0
4                   0                21   67       0              0
...               ...               ...  ...     ...            ...
52069               0                22   66       1              0
52070               1                22   66       1              0
52071               0                22   66       1              0
52072               0                22   66       1              0
52073               1                22   66       1              0
```

```
       HospiceDummy  InstitutionDummy  ESRDDummy
0                 0                 0          0
1                 0                 0          0
2                 0                 0          0
3                 0                 0          0
4                 0                 0          0
...             ...               ...        ...
52069             0                 0          0
52070             0                 0          0
52071             0                 0          0
52072             0                 0          0
52073             0                 0          0
```

```
[52074 rows x 13 columns]
```

```
# retaining data for 7th month exclusively
df_7th_month = df[df['Month']==7]
df_7th_month
```

```
       Month  GrossDrugCost  NLISDummy  LISCHOSERDummy  RiskScore  \
0          7         184.26          0               0      336.0
5          7           6.66          0               0      228.8
8          7         442.12          0               0      539.6
9          7          53.66          0               0      637.6
12         7           0.00          0               0      434.4
...      ...            ...        ...             ...        ...
52054      7          85.05          0               1      330.0
52058      7         306.12          0               1      396.4
52059      7         874.11          0               1      569.6
52068      7          28.31          1               0      248.8
52073      7         861.55          1               0      245.6
```

```
       SpecialtyDummy  AdjudicationDays  Age  Gender  FrailityDummy  \
0                   0                21   52       1              0
5                   0                21   77       1              0
8                   0                21   24       0              0
9                   0                21   86       0              0
12                  0                21   45       1              0
...               ...               ...  ...     ...            ...
52054               0                22   66       0              0
52058               0                22   66       1              0
52059               0                22   60       1              0
52068               0                22   66       0              0
52073               1                22   66       1              0

       HospiceDummy  InstitutionDummy  ESRDDummy
0                 0                 0          0
5                 0                 0          0
8                 0                 0          0
9                 0                 0          0
12                0                 0          0
...             ...               ...        ...
52054             0                 0          0
52058             0                 0          0
52059             0                 0          0
52068             0                 0          0
52073             0                 0          0

[21914 rows x 13 columns]
```

```
# Evaluating predicted cost using the model equation

df_7th_month['Predicted Cost']= np.exp(0.0451*df_7th_month['Month']+
0.2229*df_7th_month['NLISDummy'] +
0.5723*df_7th_month['LISCHOSERDummy'] +
0.0009*df_7th_month['RiskScore'] +
1.2389*df_7th_month['SpecialtyDummy'] +
0.1841*df_7th_month['AdjudicationDays'] - 0.0052*df_7th_month['Age'] -
0.0672*df_7th_month['Gender'])
df_7th_month
```

```
/var/folders/wx/1_76tj0s15gc4yxmndvw4l_00000gn/T/
ipykernel_16749/4184236895.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  df_7th_month['Predicted Cost']= np.exp(0.0451*df_7th_month['Month']+
0.2229*df_7th_month['NLISDummy'] +
0.5723*df_7th_month['LISCHOSERDummy'] +
```

```
0.0009*df_7th_month['RiskScore'] +
1.2389*df_7th_month['SpecialtyDummy'] +
0.1841*df_7th_month['AdjudicationDays'] - 0.0052*df_7th_month['Age'] -
0.0672*df_7th_month['Gender'])
```

```
       Month  GrossDrugCost  NLISDummy  LISCHOSERDummy  RiskScore  \
0          7         184.26          0               0      336.0
5          7           6.66          0               0      228.8
8          7         442.12          0               0      539.6
9          7          53.66          0               0      637.6
12         7           0.00          0               0      434.4
...      ...            ...        ...             ...        ...
52054      7          85.05          0               1      330.0
52058      7         306.12          0               1      396.4
52059      7         874.11          0               1      569.6
52068      7          28.31          1               0      248.8
52073      7         861.55          1               0      245.6

       SpecialtyDummy  AdjudicationDays  Age  Gender  FrailityDummy  \
0                   0                21   52       1              0
5                   0                21   77       1              0
8                   0                21   24       0              0
9                   0                21   86       0              0
12                  0                21   45       1              0
...               ...               ...  ...     ...            ...
52054               0                22   66       0              0
52058               0                22   66       1              0
52059               0                22   60       1              0
52068               0                22   66       0              0
52073               1                22   66       1              0

       HospiceDummy  InstitutionDummy  ESRDDummy  Predicted Cost
0                 0                 0          0       63.218691
5                 0                 0          0       50.406493
8                 0                 0          0       93.938471
9                 0                 0          0       74.324452
12                0                 0          0       71.633296
...             ...               ...        ...             ...
52054             0                 0          0      133.219747
52058             0                 0          0      132.232270
52059             0                 0          0      159.435587
52068             0                 0          0       87.314802
52073             0                 0          0      280.995878

[21914 rows x 14 columns]
```

```
# calculating square of differences between predicted and actual drug
cost in a column called 'square_of_differences'
df_7th_month['square_of_difference'] = (df_7th_month['Predicted Cost']
```

```
- df_7th_month['GrossDrugCost'] ) ** 2
df_7th_month
```

```
/var/folders/wx/1_76tj0s15gc4yxmndvw4l_00000gn/T/
ipykernel_16749/131239660.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  df_7th_month['square_of_difference'] = (df_7th_month['Predicted
Cost'] - df_7th_month['GrossDrugCost'] ) ** 2
```

	Month	GrossDrugCost	NLISDummy	LISCHOSERDummy	RiskScore	\
0	7	184.26	0	0	336.0	
5	7	6.66	0	0	228.8	
8	7	442.12	0	0	539.6	
9	7	53.66	0	0	637.6	
12	7	0.00	0	0	434.4	
...	...	...	...	...	...	
52054	7	85.05	0	1	330.0	
52058	7	306.12	0	1	396.4	
52059	7	874.11	0	1	569.6	
52068	7	28.31	1	0	248.8	
52073	7	861.55	1	0	245.6	

	SpecialtyDummy	AdjudicationDays	Age	Gender	FrailityDummy	\
0	0	21	52	1	0	
5	0	21	77	1	0	
8	0	21	24	0	0	
9	0	21	86	0	0	
12	0	21	45	1	0	
...	...	...	...	...	...	
52054	0	22	66	0	0	
52058	0	22	66	1	0	
52059	0	22	60	1	0	
52068	0	22	66	0	0	
52073	1	22	66	1	0	

	HospiceDummy	InstitutionDummy	ESRDDummy	Predicted Cost	\
0	0	0	0	63.218691	
5	0	0	0	50.406493	
8	0	0	0	93.938471	
9	0	0	0	74.324452	
12	0	0	0	71.633296	
...	...	...	...	...	
52054	0	0	0	133.219747	
52058	0	0	0	132.232270	
52059	0	0	0	159.435587	

```
52068             0                  0          0       87.314802
52073             0                  0          0      280.995878
```

```
       square_of_difference
0              14650.998501
5               1913.755667
8             121230.377406
9                427.019589
12              5131.329159
...                     ...
52054           2320.324553
52058          30236.942529
52059         510759.516004
52068           3481.566641
52073         337043.088632

[21914 rows x 15 columns]
```

```
# calculating RMSE
RMSE = np.sqrt(df_7th_month['square_of_difference'].sum() /
len(df_7th_month))
RMSE
```

```
1584.5939663860904
```