ML 12 - Covariance and Correlation By Virat Tiwari

December 1, 2023

1 Covariance And Correlation By Virat Tiwari

[2]:
```
# For importing dataset we have to import seaborn becouse it provide lot of␣
↪built in dataset

import seaborn as sns
```

[4]:
```
# Here we import "healthexp" dataset
# df is a variable in which we store the dataset

df=sns.load_dataset("healthexp")
```

[5]:
```
# .head ( )  function is used for initialising first 5 datapoints

df.head()
```

[5]:

	Year	Country	Spending_USD	Life_Expectancy
0	1970	Germany	252.311	70.6
1	1970	France	192.143	72.2
2	1970	Great Britain	123.993	71.9
3	1970	Japan	150.437	72.0
4	1970	USA	326.961	70.9

[7]:
```
# . cov ( )  function is used for finding the " covariance "

df.cov()
```

```
/tmp/ipykernel_110/1545644723.py:1: FutureWarning: The default value of
numeric_only in DataFrame.cov is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  df.cov()
```

[7]:

	Year	Spending_USD	Life_Expectancy
Year	201.098848	2.571883e+04	41.915454
Spending_USD	25718.827373	4.817761e+06	4166.800912
Life_Expectancy	41.915454	4.166801e+03	10.733902

[21]:
```
# . corr ( )  function is used for finding the correlation

df.corr()
```

```
/tmp/ipykernel_110/1134722465.py:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  df.corr()
```

[21]:

	Year	Spending_USD	Life_Expectancy
Year	1.000000	0.826273	0.902175
Spending_USD	0.826273	1.000000	0.579430
Life_Expectancy	0.902175	0.579430	1.000000

[9]:
```
# Pearson correlation cofficient
# . corr ( method = " pearson " ) is used for finding the pearson correlation␣
 ↪coffiecient

df.corr(method="pearson")
```

```
/tmp/ipykernel_110/2979612414.py:3: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  df.corr(method="pearson")
```

[9]:

	Year	Spending_USD	Life_Expectancy
Year	1.000000	0.826273	0.902175
Spending_USD	0.826273	1.000000	0.579430
Life_Expectancy	0.902175	0.579430	1.000000

[10]:
```
# Spearman correlation cofficieant
# . corr  ( method ="spearman ") is sused for finding the spearman correlation␣
 ↪cofficient
df.corr(method="spearman")
```

```
/tmp/ipykernel_110/383025161.py:3: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  df.corr(method="spearman")
```

[10]:

	Year	Spending_USD	Life_Expectancy
Year	1.000000	0.931598	0.896117
Spending_USD	0.931598	1.000000	0.747407
Life_Expectancy	0.896117	0.747407	1.000000

[13]:
```
# For a PRACTICE  we have to take another DATASET " TIPS "
# We do all operation in this  " TIPS "  dataset as we done before in "␣
 ↪HEALTHEXP "  dataset

df1=sns.load_dataset("tips")
```

[14]:
```
df1.head()
```

[14]:

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

[16]:
```
df1.cov()
```

```
/tmp/ipykernel_110/3142585312.py:1: FutureWarning: The default value of
numeric_only in DataFrame.cov is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  df1.cov()
```

[16]:

	total_bill	tip	size
total_bill	79.252939	8.323502	5.065983
tip	8.323502	1.914455	0.643906
size	5.065983	0.643906	0.904591

[17]:
```
df1.corr(method="pearson")
```

```
/tmp/ipykernel_110/2620161816.py:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  df1.corr(method="pearson")
```

[17]:

	total_bill	tip	size
total_bill	1.000000	0.675734	0.598315
tip	0.675734	1.000000	0.489299
size	0.598315	0.489299	1.000000

[19]:
```
df1.corr(method="spearman")
```

```
/tmp/ipykernel_110/4248716929.py:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  df1.corr(method="spearman")
```

[19]:

	total_bill	tip	size
total_bill	1.000000	0.678968	0.604791
tip	0.678968	1.000000	0.468268
size	0.604791	0.468268	1.000000

[20]:
```
df1.corr()
```

```
/tmp/ipykernel_110/473017434.py:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  df1.corr()
```

[20]:

	total_bill	tip	size
total_bill	1.000000	0.675734	0.598315
tip	0.675734	1.000000	0.489299
size	0.598315	0.489299	1.000000

THANK YOU SO MUCH !!

VIRAT TIWARI :)