22 Mar Assignment

March 29, 2023

1 Assignment 46

Q1. Pearson correlation coefficient is a measure of the linear relationship between two variables. Suppose you have collected data on the amount of time students spend studying for an exam and their final exam scores. Calculate the Pearson correlation coefficient between these two variables and interpret the result.

Ans. To calculate the Pearson correlation coefficient between the time students spend studying for an exam and their final exam scores, you first need to compute the covariance and standard deviations of the two variables.

Let X be the variable representing the time spent studying, and Y be the variable representing the exam scores. Let n be the number of observations in the dataset. Then, the Pearson correlation coefficient (r) can be calculated as:

r = (Σ((X - X_mean) * (Y - Y_mean))) / ((n - 1) * s_X * s_Y)

Where X_mean and Y_mean are the mean values of X and Y, respectively, and s_X and s_Y are the corresponding standard deviations.

Once you have calculated r, you can interpret the result as follows:

If r is close to 1, it indicates a strong positive linear relationship between the two variables, which means that as one variable increases, the other variable tends to increase as well.

If r is close to -1, it indicates a strong negative linear relationship between the two variables, which means that as one variable increases, the other variable tends to decrease.

If r is close to 0, it indicates a weak or no linear relationship between the two variables, which means that there is no significant association between them.

For example, if the Pearson correlation coefficient between the time spent studying and the exam scores is found to be 0.8, it suggests a strong positive linear relationship between the two variables. This means that students who spend more time studying tend to score higher on the exam. Conversely, if the correlation coefficient is found

to be -0.6, it suggests a strong negative linear relationship between the two variables. This means that students who spend less time studying tend to score higher on the exam.

[6]:
```
import pandas as pd
import numpy as np

# Create a sample dataset
time_spent = [5, 7, 6, 8, 4, 6, 7, 9, 8, 6]
exam_scores = [65, 82, 78, 92, 55, 74, 81, 95, 88, 73]

# Calculate the Pearson correlation coefficient
correlation_coefficient = np.corrcoef(time_spent, exam_scores)[0,1]

print("Pearson correlation coefficient:", correlation_coefficient)
```

```
Pearson correlation coefficient: 0.9841108417256759
```

Q2. Spearman's rank correlation is a measure of the monotonic relationship between two variables. Suppose you have collected data on the amount of sleep individuals get each night and their overall job satisfaction level on a scale of 1 to 10. Calculate the Spearman's rank correlation between these two variables and interpret the result.

[7]:
```
import pandas as pd

# Create a sample dataset
sleep = [8, 7, 6, 5, 7, 6, 9, 7, 8, 5]
job_satisfaction = [9, 8, 7, 5, 8, 6, 10, 7, 9, 5]
```

[8]:
```
# Rank the values of each variable
sleep_rank = pd.Series(sleep).rank()
job_satisfaction_rank = pd.Series(job_satisfaction).rank()
```

[9]:
```
sleep_rank
```

[9]:
```
0     8.5
1     6.0
2     3.5
3     1.5
4     6.0
5     3.5
6    10.0
7     6.0
8     8.5
9     1.5
dtype: float64
```

[10]:
```
job_satisfaction_rank
```

```
[10]: 0     8.5
      1     6.5
      2     4.5
      3     1.5
      4     6.5
      5     3.0
      6    10.0
      7     4.5
      8     8.5
      9     1.5
      dtype: float64
```

```
[11]: # Calculate the Spearman's rank correlation
      spearman_corr = sleep_rank.corr(job_satisfaction_rank, method='spearman')
```

```
[12]: print("Spearman's Rank Correlation Coefficient:", spearman_corr)
```

```
Spearman's Rank Correlation Coefficient: 0.974964745220317
```

Q3. Suppose you are conducting a study to examine the relationship between the number of hours of exercise per week and body mass index (BMI) in a sample of adults. You collected data on both variables for 50 participants. Calculate the Pearson correlation coefficient and the Spearman's rank correlation between these two variables and compare the results.

```
[20]: import pandas as pd
      import numpy as np

      # Create a sample dataset
      exercise_hours = [4, 6, 5, 2, 3, 1, 7, 8, 6, 5, 3, 4, 6, 2, 1, 3, 5, 4, 7, 8,␣
       ↪6, 3, 4, 5, 6, 7, 3, 2, 1, 4, 6, 7, 8, 9, 7, 5]
      bmi = [22, 26, 24, 28, 25, 29, 21, 20, 23, 27, 26, 24, 23, 29, 30, 27, 25, 26,␣
       ↪21, 20, 23, 27, 25, 24, 23, 22, 28, 29, 30, 26, 25, 24, 23, 22, 21, 27]
```

```
[23]: # Calculate Pearson correlation coefficient
      pearson_corr = np.corrcoef(exercise_hours, bmi)[0, 1]

      print("Pearson correlation coefficient:", pearson_corr)
```

```
Pearson correlation coefficient: -0.8747624691739074
```

```
[ ]: # Calculate Spearman's rank correlation coefficient
     exercise_hours_rank = pd.Series(exercise_hours).rank()
     bmi_rank = pd.Series(bmi).rank()
     spearman_corr = exercise_hours_rank.corr(bmi_rank, method='spearman')
```

```
[25]: print("Rank of Exercise Hours")
      print(exercise_hours_rank)
```

```
Rank of Exercise Hours
0     14.0
1     24.5
2     19.0
3      5.0
4      9.0
5      2.0
6     30.0
7     34.0
8     24.5
9     19.0
10     9.0
11    14.0
12    24.5
13     5.0
14     2.0
15     9.0
16    19.0
17    14.0
18    30.0
19    34.0
20    24.5
21     9.0
22    14.0
23    19.0
24    24.5
25    30.0
26     9.0
27     5.0
28     2.0
29    14.0
30    24.5
31    30.0
32    34.0
33    36.0
34    30.0
35    19.0
dtype: float64
```

[26]:

```
print("Rank of BMI")
print(bmi_rank)
```

```
Rank of BMI
0      7.0
1     23.5
2     15.5
3     30.5
4     19.5
```

```
5     33.0
6      4.0
7      1.5
8     11.0
9     27.5
10    23.5
11    15.5
12    11.0
13    33.0
14    35.5
15    27.5
16    19.5
17    23.5
18     4.0
19     1.5
20    11.0
21    27.5
22    19.5
23    15.5
24    11.0
25     7.0
26    30.5
27    33.0
28    35.5
29    23.5
30    19.5
31    15.5
32    11.0
33     7.0
34     4.0
35    27.5
dtype: float64
```

[27]:

```python
print("Spearman's rank correlation coefficient:")
print(spearman_corr)
```

```
Spearman's rank correlation coefficient:
-0.8681755461836058
```

Q4. A researcher is interested in examining the relationship between the number of hours individuals spend watching television per day and their level of physical activity. The researcher collected data on both variables from a sample of 50 participants. Calculate the Pearson correlation coefficient between these two variables.

[28]:

```python
import numpy as np


# Create a sample dataset
```

```
tv_hours = [2, 3, 4, 1, 5, 6, 3, 2, 4, 5, 1, 2, 3, 6, 5, 4, 2, 1, 5, 6, 3, 2,␣
 ↪1, 4, 5, 6, 3, 2, 1, 2, 3, 4, 5, 6, 3, 2, 4, 1, 5, 6, 3, 2, 4, 5, 1, 2, 3,␣
 ↪6, 5, 4]
physical_activity = [3, 5, 6, 2, 4, 5, 2, 3, 4, 5, 1, 2, 3, 5, 5, 4, 3, 1, 4,␣
 ↪5, 2, 2, 1, 4, 5, 5, 3, 3, 1, 2, 3, 4, 4, 5, 3, 3, 4, 2, 4, 5, 2, 3, 4, 5,␣
 ↪1, 2, 3, 5, 5, 4]

# Calculate Pearson correlation coefficient
corr_coeff = np.corrcoef(tv_hours, physical_activity)[0, 1]

print("Pearson correlation coefficient:", corr_coeff)
```

```
Pearson correlation coefficient: 0.883445792125462
```

Q5. A survey was conducted to examine the relationship between age and preference for a particular brand of soft drink. The survey results are shown below:

Age(Years)	Soft Drink Preference
25	Coke
42	Pepsi
37	Mountain Dew
19	Coke
31	Pepsi
28	Coke

[29]:
```
import pandas as pd

survey = pd.DataFrame({'Age(Years)' : [25, 42, 37, 19, 31, 28],
            'Soft Drink Preferences' : ['Coke', 'Pepsi', 'Mountain Dew', 'Coke',␣
 ↪'Pepsi', 'Coke']
})

print(survey)
```

```
   Age(Years) Soft Drink Preferences
0          25                   Coke
1          42                  Pepsi
2          37           Mountain Dew
3          19                   Coke
4          31                  Pepsi
5          28                   Coke
```

[30]:
```
# Correlation Analysis:

# Convert soft drink preferences to numeric codes
```

```
survey['Soft Drink Code'] = survey['Soft Drink Preferences'].astype('category').
 ↪cat.codes

# Calculate the Pearson correlation coefficient between age and soft drink␣
 ↪preference
corr = survey['Age(Years)'].corr(survey['Soft Drink Code'], method='pearson')

print('Pearson correlation coefficient:', corr)
```

```
Pearson correlation coefficient: 0.7691751415594736
```

Q6. A company is interested in examining the relationship between the number of sales calls made per day and the number of sales made per week. The company collected data on both variables from a sample of 30 sales representatives. Calculate the Pearson correlation coefficient between these two variables.

[31]:

```
import pandas as pd

# create a sample dataset with two variables
sales_data = pd.DataFrame({'sales_calls_per_day': [15, 20, 10, 12, 18, 22, 25,␣
 ↪30, 14, 19, 20, 24, 16, 8, 13, 17, 23, 21, 27, 29, 26, 28, 19, 18, 16, 12,␣
 ↪14, 10, 9, 11],
                           'sales_made_per_week': [2, 4, 1, 1, 3, 5, 6, 8, 2, 4,␣
 ↪4, 7, 3, 0, 1, 2, 6, 5, 8, 7, 6, 7, 4, 3, 3, 1, 2, 0, 0, 1]})

print(sales_data)


# calculate the Pearson correlation coefficient
corr_coef = sales_data['sales_calls_per_day'].
 ↪corr(sales_data['sales_made_per_week'], method='pearson')

print('Pearson correlation coefficient:', corr_coef)
```

```
    sales_calls_per_day  sales_made_per_week
0                    15                    2
1                    20                    4
2                    10                    1
3                    12                    1
4                    18                    3
5                    22                    5
6                    25                    6
7                    30                    8
8                    14                    2
9                    19                    4
10                   20                    4
11                   24                    7
12                   16                    3
```

```
13                    8                    0
14                   13                    1
15                   17                    2
16                   23                    6
17                   21                    5
18                   27                    8
19                   29                    7
20                   26                    6
21                   28                    7
22                   19                    4
23                   18                    3
24                   16                    3
25                   12                    1
26                   14                    2
27                   10                    0
28                    9                    0
29                   11                    1
Pearson correlation coefficient: 0.9797539496923989
```

```
[ ]:
```