Assignment

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:The Pearson correlation coefficient, denoted by

ρ

r, is a measure of the linear relationship between two variables. It ranges from -1 to 1, where:

- $\rho = 1$
- $r = 1$ indicates a perfect positive linear relationship,
- $\rho = -1$
- $r = -1$ indicates a perfect negative linear relationship,
- $\rho = 0$
- $r = 0$ indicates no linear relationship.

The formula for calculating the Pearson correlation coefficient (

ρ

r) between two variables, X and Y, is given by:

$$\rho = \sum(\rho\rho - \bar{\rho})(\rho\rho - \bar{\rho}) \sum(\rho\rho - \bar{\rho})2 \sum(\rho\rho - \bar{\rho})2$$

$r =$

$\sum(X$

i

$\bar{}$

X

-

$$)$$

$$2$$

$$\sum(Y$$

$$i$$

$$-$$

$$Y$$

$$-$$

$$)$$

$$2$$

$$\sum(X$$

$$i$$

$$-$$

$$X$$

$$-$$

$$)(Y$$

$$i$$

$$-$$

$$Y$$

$$-$$

$$)$$

Here's how you could interpret the result in the context of your example:

- If
- ρ
- r is close to 1, it suggests a strong positive linear relationship. As students spend more time studying, their final exam scores tend to increase.
- If
- ρ
- r is close to -1, it suggests a strong negative linear relationship. As students spend more time studying, their final exam scores tend to decrease.
- If
- ρ
- r is close to 0, it suggests no clear linear relationship between the time spent studying and the final exam scores.

To calculate

ρ

r, you would need the data on the amount of time each student spent studying (X) and their corresponding final exam scores (Y). Once you have these data points, you can plug them into the formula to obtain the Pearson correlation coefficient.

In Python, you could use libraries like NumPy or SciPy to calculate the correlation coefficient. For example:

python

Copy code

```
import       as

                    10  20  30  40  50
                  60  70  80  90  100

                                                                    0
1
```

```
print "Pearson correlation coefficient:"
```

Interpretation:

- If the correlation coefficient is close to 1, it indicates a positive linear relationship.
- If the correlation coefficient is close to -1, it indicates a negative linear relationship.
- If the correlation coefficient is close to 0, it suggests no clear linear relationship.

Remember that correlation does not imply causation, and other factors may influence the relationship between the variables.

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.
Ans:Spearman's rank correlation coefficient, denoted by

ρ (rho), is a non-parametric measure of the monotonic relationship between two variables. It assesses how well the relationship between two variables can be described using a monotonic function. Spearman's correlation is suitable for variables measured on ordinal or interval scales.

The formula for calculating Spearman's rank correlation coefficient is as follows:

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

i

2

Where:

- ††
- d
- i
-
- is the difference between the ranks of corresponding pairs of observations.
- †
- n is the number of observations.

Here's how you could interpret the result:

- If
- †=1
- $\rho=1$, it indicates a perfect positive monotonic relationship. As one variable increases, the other variable consistently increases.
- If
- †=−1
- $\rho=-1$, it indicates a perfect negative monotonic relationship. As one variable increases, the other variable consistently decreases.
- If
- †=0
- $\rho=0$, it suggests no monotonic relationship between the variables.

In Python, you can use libraries like SciPy to calculate Spearman's rank correlation coefficient.

Here's an example:

python

Copy code

```
from              import
```

```
                    7  6  8  5  7
                     9  5  8  3  7



print "Spearman's rank correlation coefficient:"
print "p-value:"
```

Interpretation:

- If
-
- ρ is close to 1, it suggests a strong positive monotonic relationship. More sleep is associated with higher job satisfaction.
- If
-
- ρ is close to -1, it suggests a strong negative monotonic relationship. More sleep is associated with lower job satisfaction.
- If
-
- ρ is close to 0, it suggests no clear monotonic relationship between the amount of sleep and job satisfaction.

The p-value is also important. If the p-value is below a chosen significance level (e.g., 0.05), you may reject the null hypothesis that there is no monotonic relationship. Always consider both the correlation coefficient and the p-value in the interpretation.

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.
Ans:To calculate the Pearson correlation coefficient and Spearman's rank correlation between the number of hours of exercise per week and body mass index (BMI), you can use Python with libraries like NumPy and SciPy. Here's an example:
python

Copy code

```
import       as
from            import


                    2  3  4  5  2  1  3  4  5  1
 3  2  4  5  1  3  2  4  5  1
 3  2  4  5  1  3  2  4  5  1
 3  2  4  5  1  3  2  4  5  1
 3  2  4  5  1  3  2  4  5  1

       25  26  24  28  23  27  25  26  24  28
 23  27  25  26  24  28  23  27  25  26
 24  28  23  27  25  26  24  28  23  27
 25  26  24  28  23  27  25  26  24  28
 23  27  25  26  24  28  23  27  25  26
```

```
print "Pearson correlation coefficient:"
print "Spearman's rank correlation coefficient:"
```

Interpretation:

- If the Pearson correlation coefficient is close to 1, it indicates a strong positive linear relationship between hours of exercise and BMI.
- If the Spearman's rank correlation coefficient is close to 1, it indicates a strong monotonic relationship between hours of exercise and BMI.

Note: The interpretation of the correlation coefficients depends on the context of the study and the characteristics of the data. Additionally, correlation does not imply causation, and other factors may influence the relationship between variables.

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.
Ans:To calculate the Pearson correlation coefficient between the number of hours individuals spend watching television per day and their level of physical activity, you can use Python with libraries like NumPy. Here's an example:
python

Copy code

```
import       as


               2  3  4  5  2  1  3  4  5  1
 3  2  4  5  1  3  2  4  5  1
 3  2  4  5  1  3  2  4  5  1
 3  2  4  5  1  3  2  4  5  1
 3  2  4  5  1  3  2  4  5  1

                    25  26  24  28  23  27  25  26  24  28
 23  27  25  26  24  28  23  27  25  26
 24  28  23  27  25  26  24  28  23  27
 25  26  24  28  23  27  25  26  24  28
 23  27  25  26  24  28  23  27  25  26


                                                        0  1


print "Pearson correlation coefficient:"
```

Interpretation:

- If the Pearson correlation coefficient is close to 1, it indicates a strong positive linear relationship between the number of hours individuals spend watching television per day and their level of physical activity.
- If the Pearson correlation coefficient is close to -1, it indicates a strong negative linear relationship.
- If the Pearson correlation coefficient is close to 0, it suggests no clear linear relationship between the two variables.

Note: The interpretation of the correlation coefficient depends on the context of the study and the characteristics of the data. Additionally, correlation does not imply causation, and other factors may influence the relationship between variables.

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)
25 Coke
42 Pepsi
37
19
31
28

Mountain dew
Coke
Pepsi
Coke
Soft drink Preference
Ans:It seems like there might be a typo or formatting issue in the data you provided. It appears to be a table with two columns (Age in Years and Soft Drink Preference), but the association between each age value and the corresponding soft drink preference is not clear. Could you please clarify or provide a more structured format for the data?

For example:

markdown

Copy code

Once you provide a clearer representation of the data, I can assist you in calculating relevant statistics, such as correlation or association measures, to examine the relationship between age and soft drink preference.

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.

Ans:To calculate the Pearson correlation coefficient between the number of sales calls made per day and the number of sales made per week for a sample of 30 sales representatives, you can use Python with libraries like NumPy. Here's an example:

python

Copy code

```
import       as


                      15  20  18  25  22  17  19  23  21  16
 20  24  19  26  28  15  22  18  25  20
 17  19  23  21  16  20  24  19  26  28

                   50  60  55  70  65  50  58  68  62  48
 55  72  56  78  80  45  68  52  75  58
 48  58  70  64  47  55  75  60  80  85


                                                          0  1


print "Pearson correlation coefficient:"
```

Interpretation:

- If the Pearson correlation coefficient is close to 1, it indicates a strong positive linear relationship between the number of sales calls made per day and the number of sales made per week.
- If the Pearson correlation coefficient is close to -1, it indicates a strong negative linear relationship.
- If the Pearson correlation coefficient is close to 0, it suggests no clear linear relationship between the two variables.

Note: The interpretation of the correlation coefficient depends on the context of the study and the characteristics of the data. Additionally, correlation does not imply causation, and other factors may influence the relationship between variables.