Homework Assignment 02

Numerical Statistics Fall, 2022

Jake Underland - 1A193008

2022-01-24

1.

Suppose that five people, a, b, c, d and e, entered their photos in an exhibition, and judges X and Y scored the photos within the range of 0 to 50 points and 0 to 100 points, respectively:

	a	b	c	d	e
X	32	18	39	47	26
Y	90	65	45	80	70

In this case calculate the Spearman rank correlation coefficient r_s, and then express it as an irreducible fraction.

Solution:
The table converted to express only the ordinal evaluations of X and Y is as below:

	a	b	c	d	e
X	3	1	4	5	2
Y	5	2	1	4	3

The formula for Spearman rank correlation coefficient is

$$r_s = 1 - \frac{6}{n^3 - n} \sum_{i=1}^{n} (R_i - R_i')^2$$

Thus, plugging the ranked data in we get

$$\begin{aligned} r_s &= 1 - \frac{6}{5^3 - 5} \left\{ (3-5)^2 + (1-2)^2 + (4-1)^2 + (5-4)^2 + (2-3)^2 \right\} \\ &= \frac{24}{120} = \frac{1}{5} \end{aligned}$$

Code for calculating the Spearman rank correlation coefficient is below.

```
import numpy as np
X = [32, 18, 39, 47, 26]
Y = [90, 65, 45, 80, 70]

def rank_vector(vec):
    return [(sorted(vec).index(x) + 1) for x in vec]

def compute_spearman(X, Y):
    '''
    inputs: two vectors
    outputs: spearman coefficient of two vectors
```

```
    '''
    assert len(X) == len(Y)
    n = len(X)
    data = np.array([rank_vector(X), rank_vector(Y)])
    rs = 1 - (6 / (n**3 - n)) * sum((data[0,] - data[1,])**2)
    return rs
```

2.

Consider the following data:

x	-1	0	1	3	4	5
y	8	6	5	2	0	-3

(2-1) Find the linear regression line, $y = \alpha + \beta x$, of y on x based on the data, and then express α and β as irreducible fractions.

Solution.

In the following simple regression model $y = \beta_0 + \beta_1 x + e_i$, the best linear predictors for the two coefficients under squared loss are$(\hat{\beta}_0, \hat{\beta}_1) \in \text{argmin}_{\beta_0,\beta_1} \sum_i (y_i - \beta_0 - \beta_1 x_i)^2$. We can compute these as,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\beta}_1 = \frac{\sum_{i=1}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}(x_i - \bar{x})^2}$$

Then, following the above formula,

$$\begin{aligned}
\bar{x} &= 2 \\
\bar{y} &= 3 \\
\beta &= \frac{(-1-2)(8-3) + (0-2)(6-3) + (1-2)(5-3) + (3-2)(2-3) + (4-2)(0-3) + (5-2)(-3-3)}{(-1-2)^2 + (0-2)^2 + (1-2)^2 + (3-2)^2 + (4-2)^2 + (5-2)^2} \\
&= \frac{-12}{7} \\
\alpha &= 3 - \frac{-12}{7} \times 2 = \frac{45}{7} \ldots \square
\end{aligned}$$

Code can be found below:

```
def estimate_beta1(x, y):
    x = np.array(x)
    y = np.array(y)
    numerator = np.sum((x - np.mean(x)) * (y - np.mean(y)))
    denominator = np.sum((x - np.mean(x)) ** 2)
    return reduce_frac(numerator, denominator)

def reduce_frac(num, denom):
    x = gcd(num, denom)
    return "{}/{}".format(num/x, denom/x)

def gcd(m, n):
    r = m % n
    return n if not r else gcd(n, r)
```

```
x = [-1, 0, 1, 3, 4, 5]
y = [8, 6, 5, 2, 0, -3]
print(estimate_beta1(x, y))
```

```
## -12.0/7.0
```

(2-2) Find the linear regression line, $x = \gamma + \delta y$ of x on y based on the data, and then express γ and δ as irreducible fractions.

Solution.
Similarly,

```
print(estimate_beta1(y, x))
```

```
## -4.0/7.0
```

Thus,

$$\delta = -\frac{4}{7}$$
$$\gamma = 2 - \frac{-4}{7} \times 3 = \frac{26}{7}$$