Homework Assignment 02

Numerical Statistics Fall, 2022

Jake Underland - 1A193008

2022-01-04

1.

Consider a stock that grows by 12% in year one, declines by 8% in year two, declines by 3% in year three and grows by 42% in year four. In this case calculate the average growth rate of the stock, and then express it in the form of a decimal to the second decimal places. (When calculating the fourth root of a number using a calculator, you need only apply the square root twice.) (2 points)

Solution:

$$\sqrt[4]{(1+0.12)\times(1-0.08)\times(1-0.03)\times(1+0.42)} \approx 1.09\ldots\square$$

```
sqrt(sqrt(1.12 * (1 - 0.08) * (1 - 0.03) * (1 + 0.42)))
```

```
## [1] 1.091482
```

2.

Suppose we want to make a histogram based on a data $x_1, x_2, \ldots, x_{128}$ using the Sturges rule, where we expect the histogram to be symmetrical. In this case find the number of classes, k, in the corresponding frequency distribution. (2 points)

Solution:

$$k \approx 1 + \log_2 128 = 8\ldots\square$$

```
1 + log(128, 2)
```

```
## [1] 8
```

3.

Consider the linear regression of y on x based on a data:

x	0	1	2	3	4
y	2	1	4	3	5

Using the Pearson's correlation coefficient r_{xy}, calculate the coefficient of determination R^2 of the regression, and then express it as an irreducible fraction. (7 points)

Solution:

$$\bar{x} = \frac{(0+1+2+3+4)}{5} = 2$$
$$\bar{y} = \frac{(2+1+4+3+5)}{5} = 3$$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
0	2	-2	-1	2	4	1
1	1	-1	-2	2	1	4
2	4	0	1	0	0	1
3	3	1	0	0	1	0
4	5	2	2	4	4	4
Total		0	0	8	10	10

$$\implies r_{xy} = \frac{8}{\sqrt{10\times 10}} = \frac{4}{5} \ldots \square$$

4.

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

	a	b	c	d	e	f
X	50	45	25	30	35	10
Y	96	75	88	54	29	42

In this case, calculate the Kendall rank correlation coefficient r_k, and then express it as an irreducible fraction. (9 points)

Solution:

The above data points ranked are as below:

	a	b	c	d	e	f
X	6	5	2	3	4	1
Y	6	4	5	3	1	2

Since there are 6 data points, the sum of the number of concordant pairs (C) and discordant pairs (D) is $C + D = \binom{6}{2} = \frac{6\cdot 5}{2} = 15$. I count the number of concordant and discordant pairs below, using the given definition:

```
def countCD(data):
  C = D = 0
  for i, tup1 in enumerate(data[:-1]):
    for tup2 in data[i+1:]:
      if (tup1[0] > tup2[0] and tup1[1] > tup2[1]) \
      or (tup1[0] < tup2[0] and tup1[1] < tup2[1]):
        C += 1
      else:
        D += 1
  return C, D
```

```
data = [(6, 6), (5, 4), (2, 5), (3, 3), (4, 1), (1, 2)]

print "Number of concordant pairs is:", countCD(data)[0], \
"\nNumber of discordant pairs is:", countCD(data)[1]
```

```
## Number of concordant pairs is: 10
## Number of discordant pairs is: 5
```

Thus, using the formula for the Kendall rank correlation coefficient,

$$r_k = \frac{C - D}{C + D} = \frac{10 - 5}{15} = \frac{1}{3} \ldots \square$$