

Homework 1

EXERCISE 1

Without performing all the calculations you should be able to indicate whether list P or list Q has the smaller sample standard deviation.

- a. List P contains the 2000 integers
1, 1, 2, 2, 3, 3, 4, 4, \dots , 1000, 1000.

List Q contains the 3000 integers
1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4, \dots , 999, 999, 999, 1000, 1000, 1000.

- b. List P has 50 values.
The number 180 appears 10 times.
The number 200 appears 30 times.
The number 220 appears 10 times.

List Q has 50 values.
The number 180 appears 5 times.
The number 200 appears 40 times.
The number 220 appears 5 times.

- c. List P contains 500 values.
The value 18 appears 100 times.
The value 19 appears 100 times.
The value 20 appears 100 times.
The value 21 appears 100 times.
The value 22 appears 100 times.

List Q contains 500 values.
The value 18 appears 150 times.
The value 20 appears 200 times.
The value 22 appears 150 times.

EXERCISE 2

Let x_1, x_2, \dots, x_n be the observations of a sample of size n , and let \bar{x} and s^2 be the sample mean and sample variance respectively.

- a. Suppose that on each observation a constant a is added. Find the new sample mean and sample variance in terms of \bar{x} and s^2 .
- b. Suppose that each observation is multiplied by a constant a . Find the new sample mean and sample variance in terms of \bar{x} and s^2 .

EXERCISE 3

In the the U.S. temperature is recorded in Farenheit degrees, while in most of the other countries it is recorded in Celcius degrees. Suppose a tourist from a country where temperature is recorded in Celcius degrees will visit Los Angeles this summer. He was told that the July average in Los Angeles is 85 Farenheit degrees, with a standard deviation of 10 Farenheit degrees. Help this tourist understand the weather conditions in Los Angeles.

EXERCISE 4

You have just calculated that the sample mean and sample standard deviation of a sample of size 101 are 240.0 and 25.88 respectively. Unfortunately, a check of the list uncovers two errors:

A number listed as 230 should be 200.

A number listed as 250 should be 280.

- a. After you make the corrections, what will be the sample mean?
- b. What will be the standard deviation, after you make the correction?

EXERCISE 5

The sample variance is given by the following formula:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

where n is the sample size and \bar{x} is the sample mean. By expanding the above formula show that:

$$s^2 = \frac{1}{n - 1} \left[\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right] = \frac{1}{n - 1} \left[\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right]$$

EXERCISE 6

The sample covariance is given by $\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$. Expand this formula to show that

$$\text{cov}(x, y) = \frac{1}{n - 1} \left[\sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right) \right].$$

EXERCISE 7

Show mathematically what will happen to the correlation coefficient between x and y if we multiply x by a constant $c_1 > 0$ and y by a constant $c_2 > 0$.

EXERCISE 8

Construct the empirical cumulative distribution function (**ecdf**) of the following data:

1.5, 1.5, 1.5, 1.5, 1.5, 3.6, 5.8, 10.5, 10.5, 13.0, 16.0, 16.0, 16.0, 20.0, 25.0, 30.0, 31.0, 50.0, 60.0, 61.0.