Measures of Variation

Suppose I propose that we play the following games. A fair six-sided dice is tossed and you will receive/pay according to the following charts.

Game 1

1	**2**	**3**	**4**	**5**	**6**
-2	-3	1	-3	2	5

Game 2

1	**2**	**3**	**4**	**5**	**6**
-200	-300	100	-100	200	300

Game 3

1	**2**	**3**	**4**	**5**	**6**
-1	1	-1	-1	1	1

Exercise: Calculate the expected value of the games above.

(1) Game 1

(2) Game 2

(3) Game 3

Question: What game would you rather play and why?

In order to quantify the riskiness of the game we have to be able to measure the *variations* of the possible outcomes.

There are a few ways to do this.

One possible way to measure the possible variations is by looking at the range of values that outcomes can be in.

Definition. The range of a set of n data point $S = \{x_1, x_2, \ldots, x_n\}$ is

MAXIMUM IN S - MINIMUM IN S.

Exercises: Find the range of the following.

(1) The set $\{100, 20, 4, 102, 58, 91, 3\}$.

(2) The outcomes from throwing a regular dice.

(3) The outcomes of the games from page 1.

(a) GAME 1

(b) GAME 2

(c) GAME 3

(4) The outcomes of the following games. A fair dice is tossed and you will receive/pay according to the following charts.

GAME 1

1	**2**	**3**	**4**	**5**	**6**
-10	-5	-2	-3	-4	-6

GAME 2

1	**2**	**3**	**4**	**5**	**6**
2	2	2	2	4	2

(a) GAME 1

(b) GAME 2

Question: What game would you rather play, GAME 1 from part (3) or GAME 1 from part (4) and why?

Note: The range does not tell us anything about the expected value of an experiment.

Consider the following exercise.

Exercise: Consider the following game. In an urn there are 9999 white balls and 1 red ball. A ball is drawn at random. If a white ball is chosen then I pay you $1, if the red ball is chosen the you pay me $9999. Let X be *your* payoff after playing this game.

(1) What is $E(X)$?

(2) What is the rage of the possible outcomes of X?

Question: Would you still play this game despite such a large range? Why?

Note: Even if you know the expected value of the range dose not tell you how ofter the deviations are expected to happen.

Consider the following data set $\{10, 10, 10, -10, -10, -10\}$.
Question: What is the mean?

Let's try to quantify how far away the data is from its mean.
TRY 1: We can try to take the sum of the difference between the data points and the mean.

What is the problem?

TRY 2: We can try to take the sum of squares of the difference between the data points and the mean.

In order to compare how far away data points are from the mean of a set, we use the two following concepts of *variance* and *standard deviation*.

Definition: Let $S = \{x_1, x_2, \ldots, x_n\}$ be a set of n data points. Assume that the mean of S is μ.

(1) The *variance* of S is

$$s^2 = \frac{\left(\sum_{i=1}^{n} x_i^2\right) - n\mu^2}{n-1}.$$

(2) The *standard deviation* is

$$s = \sqrt{\frac{\left(\sum_{i=1}^{n} x_i^2\right) - n\mu^2}{n-1}}.$$

WORDS OF CAUTION:

(1) Your book uses $n-1$ instead of n in the denominators above because is claims that it is more convenient.

(2) Generally, we will use the standard deviation to measure the variation of the outcomes from the mean, but we will make computations with the variance since it has no square root.

Exercises: Find the variance and standard deviations of the following.

(1) The set $\{1, -1, 1, -1, 1, -1, 1, 100, -1, 1, -1, -1, 1\}$.

(2) The set $\{2, 2, 2, 2\}$.

(3) The possible outcomes when tossing a dice.

Recall that in the last chapter we looked at how to take the average of data once it has been grouped. Similarly, we want to know how to take the variances and standard deviation of data one it has been grouped.

Suppose that a data set of 400 points has been grouped and plotted in the following histogram.

Question: What is the mean of the grouped distribution?

Definition: Assume that a set of data points has been grouped and plotted in a histogram with n bars. Let $\tilde{\mu}$ being the mean of the grouped distribution. Set x_i to be the midpoints of bar i, and f_i to be its frequency, and recall that $\sum_{i=1}^{n} f_i = n$. The *standard deviation for the grouped distribution* is

$$\tilde{s} = \sqrt{\frac{\left(\sum_{i=1}^{n} f_i x_i^2\right) - n\tilde{\mu}^2}{n-1}}.$$

Question: Calculate the standard deviation for the grouped distribution on the previous page.

Question: Calculate the standard deviation of the following grouped distribution

Interval	**Frequency**
0-10	4
10-20	3
20-30	6
30-40	2
40-50	3
50-60	4