Standard Deviation

Motivation

The mean (expected value) is a measure of central tendancy, it measures the position of the center of data. The mean summarizes data by telling us its center. The mean gives us one bit of information about the data, but is it enough information?

Consider the following two frequency distributions. In each distribution, x denotes an outcome of an experiment.

Distribution A

x	Frequency
1	10
9	10

$$\text{Mean} = \frac{1 \cdot 10 + 9 \cdot 10}{10 + 10} = 5$$

Distribution B

x	Frequency
4	10
6	10

$$\text{Mean} = \frac{4 \cdot 10 + 6 \cdot 10}{10 + 10} = 5$$

Both distributions have the same mean (5), and both distributions have the same number of outcomes ($10 + 10 = 20$). However, in distribution **A**, the distance from an outcome to the mean is 4 ($5 - 1 = 4$ and $9 - 5 = 4$), while in distribution **B**, the distance from an outcome to the mean is 1 ($5 - 4 = 1$ and $6 - 5 = 1$). You can see how the mean is not an adequate summary of data.

Distributions **A** and **B** provide a hint of another quantity that we can use to summarize data. That quantity is the average distance between an outcome and the mean. This average distance is 4 for distribution **A**, and it is 1 for distribution **B**. You can see that the outcomes of distribution **A** are spread out further from the mean than the outcomes of distribution **B**. In other words, the outcomes of distribution **A** are more *dispersed* around the mean than the outcomes of distribution **B**. The average distance between an outcome and the mean is called a *measure of dispersion*.

There are different ways of computing the average distance between an outcome and the mean. We computed the mean distance between an outcome and the mean for distributions **A** and **B**. However, the mean distance between an outcome and the mean is not the usual average that is used as a measure of dispersion for technical reasons. The usual average that is computed is the *standard deviation*.

Standard Deviation

The standard deviation measures the (weighted) average distance between a value of a random variable and the mean. It tells us how values of a random variable *deviate* from the mean on average. It shows us how much the values of a random variable are spread out (dispersed) around the mean. The standard deviation is the *standard* measure of dispersion.

Three definitions of standard deviation are given below. These three definitions are equivalent to each other. The definition that is used to compute standard deviation depends on the situation.

The symbol μ (pronounced "mew") denotes the mean and the expected value. The symbol σ (pronounced "sigma") denotes the standard deviation.

Definition 1: The *standard deviation* of numbers $x_1, x_2 \ldots, x_n$ which have mean μ is

$$\sigma = \sqrt{\frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2}{n}}$$

Definition 2: Suppose X is a discrete random variable with possible values $x_1, x_2, \ldots, x_n$ and suppose X has the following frequency distribution (x denotes a value of X):

x	Frequency
x_1	f_1
x_2	f_2
$\vdots$	$\vdots$
x_n	f_n

Let $m = f_1 + f_2 + \cdots + f_n$ and $P(X = x_i) = f_i \,/ m$ (the relative frequency of x_i) for $1 \le i \le n$. If $E(X) = \mu$, then the *standard deviation* of X is

$$\sigma = \sqrt{\frac{f_1\cdot(x_1-\mu)^2+f_2\cdot(x_2-\mu)^2+\cdots+f_n\cdot(x_n-\mu)^2}{f_1+f_2+\cdots+f_n}}$$

Definition 3: Suppose X is a discrete random variable with possible values $x_1, x_2, \ldots, x_n$ and suppose X has the following probability distribution (x denotes a value of X):

x	$P(X=x)$
x_1	p_1
x_2	p_2
$\vdots$	$\vdots$
x_n	p_n

If $E(X) = \mu$, then the *standard deviation* of X is

$$\sigma = \sqrt{p_1 \cdot (x_1 - \mu)^2 + p_2 \cdot (x_2 - \mu)^2 + \cdots + p_n \cdot (x_n - \mu)^2}$$

Remember that if σ is small, then the values of a random variable are close to the mean on average. If σ is large, then the values of a random variable are far away from the mean on average.

Computing Standard Deviation

You should have noticed the complexity of the standard deviation formulas in the three definitions. This complexity makes it difficult to compute the standard deviation by hand. To avoid this complexity, we will compute the standard deviation using the statistical functions built into our calculators.

Example: Suppose X is a discrete random variable which has the following frequency distribution (x denotes a value of X):

x	Frequency
0	4
1	7
2	19
3	16
4	8
5	2

Compute μ and σ.

Notes: When you are given the frequency distribution of a random variable, you may assume that $P(X = x)$ is the relative frequency of x. In addition, the instructions given below are summarized on the supplement available here: Standard Deviation

Calculator Instructions

The first step is to put the calculator into the proper mode. Since the data in this problem uses only a single variable x, we need to put the calculator into 1-variable statistics mode. This is achieved by pressing the following keys on the calculator:

2nd **DATA** **ENTER**

The next step is to put the data into the calculator. Press **DATA** to access the data entry function of the calculator. Since the first x in the frequency distribution table is 0, press **0** to get $X_1 = 0$ on the screen of the calculator. Now press **▼** and then **4** to get $FRQ = 4$ on the screen of the calculator. You have now successfully told the calculator that data value 0 has frequency 4.

The following keystrokes will tell the calculator that data value 1 has frequency 7. Watch the screen carefully as you press the keys.

▼ **1** **▼** **7**

The following keystrokes will tell the calculator that data value 2 has frequency 19. Watch the screen carefully as you press the keys.

▼ **2** **▼** **1** **9**

The following keystrokes will tell the calculator that data value 3 has frequency 16. Watch the screen carefully as you press the keys.

▼ **3** **▼** **1** **6**

The following keystrokes will tell the calculator that data value 4 has frequency 8. Watch the screen carefully as you press the keys.

▼ **4** **▼** **8**

The following keystrokes will tell the calculator that data value 5 has frequency 2. Watch the screen carefully as you press the keys.

▼ **5** **▼** **2**

The data values and their frequencies have now been successfully entered into the calculator.

The final step is to get the calculator to do the tedious work of computing the expected value (μ) and the standard deviation (σ). Press **STATVAR** to tell the calculator to do the computations.

All we need to do now is to retrieve the expected value and the standard deviation from the calculator.

Press **▶** to see the expected value. The calculator uses the symbol $\bar{x}$ for the expected value, but we use the symbol μ. Your screen should show that $\mu = 2.410714286$.

Press ▶ ▶ to see the standard deviation. The calculator uses the symbol σx for the standard deviation, but we use the symbol σ. Your screen should show that $\sigma = 1.191889043$.

If your calculator failed to show the correct μ and σ, then it is likely that you did not enter the data values and frequencies correctly. Press **DATA** and then use ▼ and ▲ to check each data value and its frequency. Change any numbers that are incorrect and press **STATVAR** to recompute μ and σ.

If you cannot obtain the correct μ and σ even after attempting to correct the data values and their frequencies, then you can clear out all of the data by pressing the following sequence of keys. Watch the screen carefully as you do this.

2nd **DATA** ▶ ▶ **ENTER**

Now press **DATA** and enter the data values and their frequencies by following the instructions given above. Finally, press **STATVAR** to recompute μ and σ.

Example: Compute the mean and the standard deviation of the following scores from a math exam: 72, 54, 95, 74, 50, 74, 64, 66, 87, 88. Round each statistic to the nearest $1/10$th.

The first step is to clear out the data values and frequencies from the previous example. Pressing the following sequence of keys to clear out the old data values and their frequencies. Watch the screen carefully as you do this.

2nd **DATA** ▶ ▶ **ENTER**

Now press **DATA** and enter the data values and their frequencies. In this case, each frequency will be 1. Once the data values and frequencies are entered correctly, press **STATVAR** to compute μ and σ. Retrieve μ and σ by following the instructions given in the previous example.

You should get $\mu = 72.4$ and $\sigma = 13.87227451$ if you followed the steps correctly.

Since each statistic must be rounded to the nearest $1/10$th, the final results are $\mu = 72.4$ and $\sigma = 13.9$.

Example: A group of people were surveyed and asked: "Which of the numbers 1, 2, 3, and 4 is your favorite?" Let X be a random variable that denotes the favorite number of a person responding to the survey. The probability distribution of X is given below. Compute μ and σ.

x	$P(X = x)$
1	0.24
2	0.15
3	0.41
4	0.20

Since we cannot put probabilities into the calculator, we need to convert the probability distribution to a frequency distribution. To do this, we will assume that 100 people responded to the survey. Since $P(X = 1) = 24\%$, 24% of the 100 people (24 people) said that 1 was their favorite. The other frequencies are determined in a similar way, and we have the following frequency distribution:

x	Frequency
1	24
2	15
3	41
4	20

Now follow the procedure given in the first example. You should get $\mu = 2.57$ and $\sigma = 1.060707311$.

You should exit 1-variable statistics mode once you are finished computing statistics such as the mean and the standard deviation. This will enable your calculator to operate normally for non-statistical computations. Press the following key sequence to exit 1-variable statistics mode:

2nd **STATVAR** **ENTER**

Author: Scott P. Randby
Email: srandby@uakron.edu

Made with GNU Emacs 24.5.1 and Org 8.3.1

Updated: 2015-10-22 Thu 14:01

Mathematics for Modern Technology
TI-30X IIS Handout

Enter 1-Variable Statistics Mode

2nd DATA ENTER

Data Input

x	Frequency
0	4
1	7
2	19
3	16
4	8
5	2

To put the above data into the calculator, do the following steps:

DATA $X_1 = 0$ ▼ $FRQ = 4$ ▼ $X_2 = 1$ ▼ $FRQ = 7$ ▼ $X_3 = 2$ ▼ $FRQ = 19$ ▼ $X_4 = 3$ ▼ $FRQ = 16$ ▼ $X_5 = 4$ ▼ $FRQ = 8$ ▼ $X_6 = 5$ ▼ $FRQ = 2$

Compute Statistics

STATVAR ▶ $\mu = 2.410714286$ ▶ ▶ $\sigma = 1.191889043$

Clear Out Data

2nd DATA ▶ ▶ ENTER

Exit 1-Variable Statistics Mode

2nd STATVAR ENTER