STATISTICS TOPIC NOTES

Mean and Standard Deviation

The concept *distribution* describes uncertainty by listing the possible outcomes and their probabilities. Once you have understood *distribution* at this general level, you are prepared to learn the major features of distributions. The two most important features of any distribution are the *location* and the *variability* of its outcomes.

- The **location** means how big or small are the outcomes – are the values typically more like 100 or 2,000?
- The **variability** means how spread out are the outcomes – are they typically close together or far apart?

The most important quantitative measures of location and variability are the mean and the standard deviation.

- The **mean** of a distribution is a numerical measure of the typical size (location) of its values.
- The **standard deviation** of a distribution is a numerical measure of the spread (variability) of its values.

In this Topic Note, I will explain the mean and the standard deviation. So far in this Topic Note I have just given you general descriptions of these concepts to jump start your thinking. The above comments are not definitions. There is much more to say. But keep these descriptions in mind as you read further.

Mean

In the forward to the first Topic Note of this course, I extolled the pedagogical principle of starting with simple examples and working up. In keeping with that principle, let us begin with the very simple Example 1 from the Topic Note on Uncertainty Distributions and Random Variables:

Example 1. Toss a fair coin. Let $X = 1$ if a head occurs. Let $X = 0$ if a tail occurs.

Then X is a random variable with the following distribution:

x	$Pr(X = x)$
0	½
1	½

(If any of this is not crystal clear to you, please review the concepts of distribution and random variable.)

Let us ask the question, what do you expect the average value of X to be if you draw repeatedly at random a large number of times from this distribution – that is, if you toss the coin repeatedly?

To answer this question, it is helpful to recall what I wrote about interpreting probabilities for this example in the Topic Note on Distribution and Random Variable. The long-run relative frequency interpretation is:

> "If we repeatedly toss the coin a large number of times, then the proportion of heads will get closer and closer to the probability of heads, namely ½. Likewise the proportion of tails will get closer and closer to the probability of tails, namely ½."

So we expect that about half of the outcomes will be 1 and about half will be 0. So if I calculate the average value over n tosses, I will calculate

$$\frac{1+1+\cdots+1+0+0+\cdots+0}{n} \cong \frac{1\times\frac{n}{2}+0\times\frac{n}{2}}{n} = 1\times\frac{1}{2}+0\times\frac{1}{2} = \frac{1}{2}$$

The ones do not all come first, followed by all of the zeroes – I just wrote it that way for convenience in adding them up, since I am permitted to add them up in any order and still get the same total. The "$\cong$" means "approximately equal to". About half the tosses are ones, and about half are zeroes. So there are approximately *n*/2 ones and approximately *n*/2 zeroes. The total of all of the ones and zeroes is about *n*/2. So the average value of *X* is about *n*/2 ÷ *n* = ½. According to the long-run frequency interpretation cited above, the larger the value of *n* (the more times I toss), the closer to ½ the proportion of heads will be, and the closer to ½ the proportion of tails will be. Thus, we will expect the average value of *X* will get closer to ½ the closer the proportion of 1's and 0's gets to the probabilities of 1 and 0.

So it is reasonable to say that the expected value of *X* is ½.

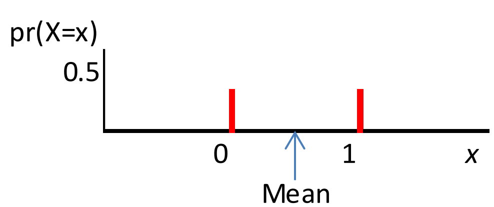
Notice that the next to the last step in the calculation above is the expression $1\times\frac{1}{2}+0\times\frac{1}{2}$. This says, weight the outcome 1 by its probability and weight the outcome 0 by its probability. The result is a weighted average of the outcomes, where the weights are given by the probabilities of the outcomes. In fact, this is the definition of the mean of a distribution (or of a random variable):

Definition. The **mean** of a distribution (or of a random variable) is obtained as a weighted average of all possible outcomes, weighted by the corresponding probabilities. We use the Greek letter μ (mu, pronounced "myoo") for the mean of a distribution (or random variable). The mathematical formula is

$$\textbf{Mean} = \mu = \sum\nolimits_{all\ x} x \times pr(X = x)$$

In this formula, *x* is the value of an outcome; $pr(X = x)$ is the probability of *x*; $x \times pr(X = x)$ is the weighted value of the outcome; and $\sum\nolimits_{all\ x} x \times pr(X = x)$ is the sum of all of the weighted outcomes. Applying this formula to Example 1, we have $\sum\nolimits_{all\ x} x \times pr(X = x) =$ $1\times pr(X=1)+0\times pr(X=0)$ = $1\times 1/2+0\times 1/2$ = ½.

Another interpretation. Suppose we make a teeter-totter and put a ½ pound weight at 0 and a ½ pound weight at 1. At what point should we put the fulcrum to make the teeter-totter balance? The picture below shows the solution:



The teeter-totter balances if the fulcrum is placed at ½. This is the *center of gravity* of the system. The center of gravity of the distribution is the mean of the distribution.

Example 2. (Same as Example 5 in the Topic Note on Distribution and Random Variable.) Throw a fair die once and let X denote the number of spots on the side facing up. So the distribution of the random variable X may be represented in table form as:

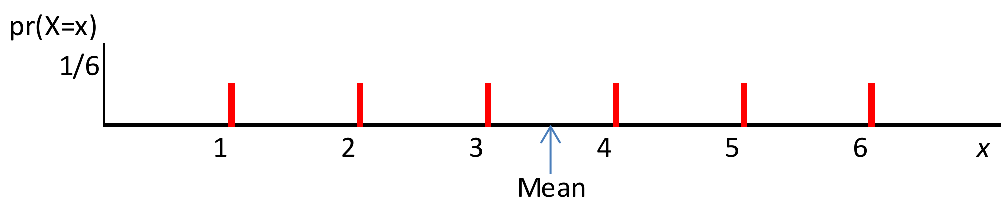
X	1	2	3	4	5	6
$pr(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

What is the mean of X? Here are three different solutions:

1. In a large number n of throws, about n/6 throws will be 1, about n/6 will be 2's, etc. So the average value of X will be about

$$\frac{1+\cdots+1+2+\cdots+2+\cdots+6+\cdots+6}{n} \cong \frac{1\times\frac{n}{6}+\cdots+6\times\frac{n}{6}}{n} = 1\times\frac{1}{6}+2\times\frac{1}{6}+\cdots+6\times\frac{1}{6} = 3.5$$

2. The center of gravity of a teeter-totter with 1/6 pound weights at each of the possible outcomes 1, 2, …, 6 is 3.5, as shown below:



3. From the mathematical formula,

Mean = $\mu = \sum_{all\ x} x \times pr(X=x) = 1\times\frac{1}{6}+2\times\frac{1}{6}+\cdots+6\times\frac{1}{6} = 3.5$

Example 3. Ten students were randomly selected from a fulltime UT MBA class. Each student was asked to provide anonymously, to the nearest \$1000, the annual salary that he/she anticipates to make in his/her first job after graduation (including bonus). The ten students reported the following anticipated salaries (in \$1,000s): 110, 160, 120, 95, 175, 80, 120, 130, 120, 110. Suppose one student is picked at random from the ten. Let X denote his/her anticipated salary. We are uncertain what the value of X will be. The distribution of X is:

X	80	95	110	120	130	160	175
$pr(X=x)$	0.1	0.1	0.2	0.3	0.1	0.1	0.1

What is the mean of X? The response will answer the question, how much can we expect a typical one of these students to make?

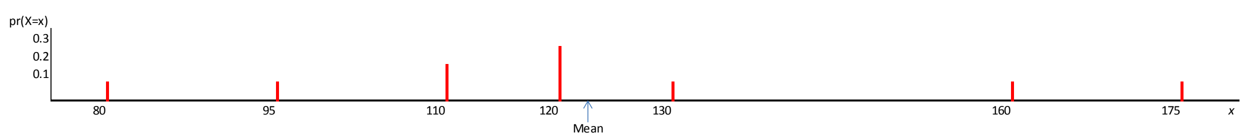
Here are three different solutions:

1. In a large number n of repeated pickings, about $0.1n$ picks will be 80, about $0.1n$ will be 95, about $0.2n$ will be 110, etc. So the average value of X will be about

$$\frac{80+\cdots+80+95+\cdots+95+\cdots+175+\cdots+175}{n} \cong \frac{80\times0.1n+\cdots+175\times0.1n}{n}$$

$$=80\times0.1+95\times0.1+110\times0.2+\cdots+175\times0.1=122.$$

2. The center of gravity of a teeter-totter with fractional pound weights at each of the possible outcomes 80, 95, …, 175 is 122, as shown below:



3. From the mathematical formula, Mean = $\mu = \sum_{all\ x} x \times pr(X = x) = 80 \times 0.1 + 95 \times 0.1 + 110 \times 0.2 + \cdots + 175 \times 0.1 = 122$

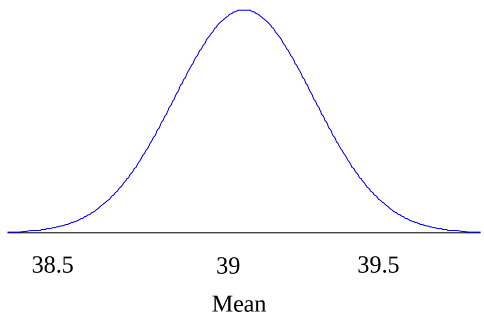
If we do a little further work with Example 3, we will gain further insight into the mean. Suppose we calculate the ordinary arithmetic average of the ten students' anticipated salaries: $\frac{110+160+120+95+175+80+120+130+120+110}{10} = 122$, as well! Are you surprised? There is a good reason why the results are the same. Suppose we rearrange the calculation of the arithmetic average by grouping like outcomes together in the numerator and then tally each group, as follows: $122 = \frac{80+95+(110+110)+(120+120+120)+130+160+175}{10} =$ $\frac{80\times 1+95\times 1+110\times 2+120\times 3+130\times 1+160\times 1+175\times 1}{10} =$ $80\times\frac{1}{10}+95\times\frac{1}{10}+110\times\frac{2}{10}+120\times\frac{3}{10}+130\times\frac{1}{10}+160\times\frac{1}{10}+175\times\frac{1}{10}$. The latter is exactly the formula for the mean.

In general, the arithmetic average of any set of numbers can be rearranged as in the preceding example to coincide with the formula for the mean of a distribution. This is because you can split up the set of outcomes into subgroups of identical values (like the one 80 and the two 110's and the three 120's, etc., in the preceding example). Then the sum of all of the outcomes can be found by multiplying the common value in each subgroup by its count (like 80 × 1 and 110 × 2 and 120 × 3) and adding the results. So the sum of the outcomes = $\sum_{all\ x} x \times count\ of\ x$. Then the arithmetic average is (supposing *n* numbers) = $\frac{sum\ of\ the\ outcomes}{n}$ = $\frac{\sum_{all\ x} x \times count\ of\ x}{n} = \sum_{all\ x} x \times \frac{count\ of\ x}{n} = \sum_{all\ x} x \times "probability"\ of\ x$. This is the same formula as the mean of a distribution, in which the "outcomes" are the *n* numbers and their "probabilities" are their proportions in the data. The mean of a distribution (or random variable) uses the same formula, but with the distribution probabilities instead of the data proportions. Since the data proportions get closer and closer to the distribution probabilities as the number of draws from the distribution increases, then the arithmetic average gets closer and closer to the mean of the distribution.

The takeaway is that the ordinary arithmetic mean of a bunch of numbers and the mean of an uncertainty distribution are calculated in exactly the same way. Both are just the ordinary arithmetic average of all of the *outcomes* – or equivalently the relative frequency-weighted average of all of the *values*. When you draw from an uncertainty distribution, all *outcomes* have the same probability of selection, but the *values* can have different probabilities of selection, since multiple outcomes can have the same value.

Example 4. The weight of a can of Folger's coffee. (Same as Example 9 in Topic Note on Uncertainty Distributions and Random Variables.)

Suppose we pick a can of Folger's coffee at random from the section of the grocery store shelf where 39-ounce cans of Folger's coffee are displayed. Let X denote the actual net weight of the coffee in the can. Since we are uncertain about the actual value, then X is a random variable with a distribution. Suppose that the distribution of X is shown below. The possible outcomes are represented by the horizontal axis and the probabilities by the curve in the following figure.



As pointed out in the Topic Note on Uncertainty Distributions and Random Variables, X is a continuous random variable. Intuitively, it is clear that the mean of this distribution is 39 ounces. This can be seen from the center-of-gravity interpretation of the mean: If a shape like the distribution shown above is placed on a teeter-totter, the teeter-totter will balance if the fulcrum is placed at 39, as shown above.

Here is yet another interpretation of the mean: *The negative deviations from the mean exactly cancel the positive deviations from the mean.*

This is a very important point for understanding what the mean really does. It requires some explanation. Suppose we pick a value c to represent the "typical" value of the outcomes of a distribution. (Mnemonically, "c" means "center".) Then c will represent some values in the distribution better than other values. To determine how well c represents each value, we can calculate the deviation of each value x from c. This deviation is $x - c$. Some of these deviations will be positive (if $x > c$). Others will be negative (if $x < c$). The smaller the magnitude of the deviation, the better c represents x; the larger the magnitude of the deviation, the worse c represents x.

We want to choose the best c that we can to represent the typical value in the distribution. Depending upon which value we choose for c, the positive deviations may predominate or the negative deviations may predominate. Choosing c less than all of the x-values is not a good idea, for then all of the deviations will be positive and they could all be reduced by moving c up toward the low end of the distribution. Similarly, choosing c greater than all of the x-values is not a good idea, for then all of the deviations will be negative and they could all be reduced by moving c down toward the high end of the distribution. In fact, c should be chosen so that it lies somewhere within the range of x-values, since then there will be at least some positive and at least some negative deviations to cancel each other.

Is there a c for which all negative deviations will exactly cancel all positive deviations? If so, then neither the positive nor the negative deviations will predominate. Yes! That c is the mean of the distribution. This can be stated more simply: Means work as a measure of the center because the deviations in the distribution above the mean exactly balance the deviations below the mean.

I will illustrate this fact for the data values from Example 3: 110, 160, 120, 95, 175, 80, 120, 130, 120, 110. The mean was computed to be 122. So the deviations from the mean are 110-122, 160-122, 120-122, 95-122, 175-122, 80-122, 120-122, 130-122, 120-122, 110-122, that is, -12, 38, -2, -27, 53, -42, -2, 8, -2, -12. The total of the positive deviations is 38 + 53 + 8 = 99. The total of the negative deviations is (-12) + (-2) + (-27) + (-42) + (-2) + (-2) + (-12) = -99. The sum of the positive deviations and the negative deviations is +99 - 99 = 0, as claimed.

The fact that the positive and negative deviations from the mean always exactly cancel and that the mean is the only number enjoying this property can be proven mathematically. If you are interested (it is not required), the following sidebar provides the argument.

Extreme sidebar! {Warning: Only for the mathematically inclined! You are not responsible for this.}

To see that the positive deviations exactly cancel the negative deviations when c = mean, note that the deviation $x - c$ occurs with relative frequency $pr(X=x)$ in the distribution since the value x occurs with relative frequency $pr(X=x)$ in the distribution. So the mean value of the *deviation* is $\sum_{all\,x}(x-c)\times pr(X=x)$ = $\sum_{all\,x}[x\times pr(X=x)-c\times pr(X=x)]$ = $\sum_{all\,x}[x\times pr(X=x)]$ – $\sum_{all\,x}[c\times pr(X=x)]$ = $\mu - c\sum_{all\,x}[pr(X=x)]$ = $\mu - c\times 1$ = $\mu - c$. So if the total deviation is to be zero, then the mean deviation must be zero. But the mean deviation is $\mu - c$, as just shown. So $\mu - c$ = 0, which implies that $\mu = c$. That is, the mean of a distribution is the value from which the positive deviations exactly cancel the negative deviations.
[End sidebar.]

Standard deviation

The standard deviation is intended to measure how much the outcomes in a distribution, collectively, are spread out. Intuitively, if the distribution values are all tightly clustered together, then there is not much uncertainty about what the outcome will be, because there is not much difference among the possible outcomes. On the other hand, if the distribution values are widely spread out, then there is a lot of uncertainty about where the outcome will be, since the possible values are widely dispersed. Thus, the standard deviation, in addition to measuring the variability of outcomes, can also measure uncertainty. In fact, there is an intimate connection between variability and uncertainty: *Uncertainty arises from variability*. If everything were always the same, there would be no variability and no uncertainty. The less variability, the less uncertainty. The more variability, the more uncertainty. By measuring variability, the standard deviation becomes a fundamental measure of uncertainty. Throughout this course, we will use standard deviations of various types to assess uncertainty of estimates that we will make. (More on the "types" of standard deviations later.) So it is critical that you understand the concept of standard deviation. I will lead up to it by first explaining two plausible alternatives that we will not use. This will allow me not only to lay some ground work, but also to deal pre-emptively with the natural student question, "Why do we have to use the complicated standard deviation when the range and MAD are easier and more intuitive?"

So how should the spread of a distribution be measured? One idea is to calculate the difference between the largest possible value and the smallest possible value in the distribution. This measure is called the **range**. However, a little reflection suggests that the range may not be a very good measure. One reason is that some distributions have no limit on how large and/or how small their possible values can be, *in principle*.[1] Another reason is that this measure of spread depends entirely on only two of the possible outcomes – the largest and the smallest – and so may not reflect the distribution values generally: For example, if most of the distribution values are tightly clustered, but one is very large, then this measure of spread will misleadingly signal wide dispersion.

A better idea is to measure spread so that ALL of the distribution outcomes may contribute.

For example, we can measure the deviation of each possible outcome *x* from some common benchmark value *c*. Since the mean μ is a satisfactory measure of the center of a distribution, we can use μ as the common benchmark *c*. Then the deviation of *x* from μ is $x - \mu$. Then we might think of calculating the average deviation as a measure of spread. However, if you remember from the preceding discussion of the mean in this Topic Note, the average deviation from the mean μ is zero (you do remember, don't you?) That is, the negative deviations from μ exactly cancel the positive deviations from μ. So this try merits only Honorable Mention.

There is nothing inherently wrong with the idea of calculating average deviation as a measure of spread as long as we calculate the average *magnitude* of the deviation, treating negative deviations the same as positive deviations. So we could calculate $\sum_{all\,x} |x - \mu| \times pr(X = x)$. This is the mean of the *magnitude* of the deviations, called the **mean absolute deviation (MAD)**.

The mean absolute deviation is actually used for some purposes in statistics. But only for some. The standard deviation receives far more use for very good reasons that would not be very illuminating for you at this time. (Trust me!) So I ask you to lay aside the quite reasonable and perfectly intuitive mean absolute deviation in favor of the still reasonable but somewhat less intuitive standard deviation as a measure of dispersion. Please do not be too upset with this. The motivation for the standard deviation is the same as for MAD – namely, to measure an "average" magnitude of outcome deviations from the mean – but the standard deviation just goes about it a little differently.

Let me explain how the standard deviation does this by comparing it to MAD. Note that the mean absolute deviation converts the negative deviations to positive deviations (magnitudes) by using absolute values. The standard deviation also converts the negative deviations to magnitudes, but by squaring the deviations: If you replace the $|x - \mu|$ in the formula for MAD by $(x - \mu)^2$, you get $\sum_{all\,x} (x - \mu)^2 \times pr(X = x)$. This is called the **variance** of a distribution:

$$\textbf{variance} = \sigma^2 = \sum_{all\,x} (x - \mu)^2 \times pr(X = x).$$

The variance is usually represented symbolically by "σ^2" [pronounced "sigma-squared"], using the Greek letter σ ["sigma"]. The variance is the mean of the *squared* deviations. Variance is important in its own right and is widely used in statistics, but variance is not quite what we want right now. The only thing really wrong with variance is that it is in the wrong units – whatever units the outcomes *x* are in (ounces, dollars, etc.), the units for variance are squared (*squared* ounces, *squared* dollars, etc.) The simplest way to restore the correct units is to take the square root of variance. This yields:

[1] For instance, see Examples 9 and 10 in the Topic Note on Distribution and Random Variable.

the **standard deviation** of a distribution = $\sigma = \sqrt{\sum_{all\,x}(x-\mu)^2 \times pr(X=x)}$

The standard deviation is usually represented symbolically by "σ" [the Greek letter pronounced "sigma"].

I hasten to assure you that the purpose of the standard deviation is the same as the purpose of the mean absolute deviation – namely, to measure an average spread of the distribution values from the center (mean) of the distribution. As its name implies, the standard deviation attempts to obtain a standard, typical, average, customary, usual deviation – a *standard* deviation – of the possible outcomes from the mean. The standard deviation and the mean absolute deviation do not calculate out to be the same value.[2] However, the values of the two are usually in the same ballpark.[3] Therefore, if you find it difficult to wrap your mental fingers around the concept of standard deviation, I encourage you to think of it in the same intuitive way that you can think of the mean absolute deviation. Even if you are one of those strange people who can wrap your mental fingers around the concept of standard deviation, I still encourage you to think about standard deviation as the average amount of deviation between the outcomes and their mean. That is the way that I will talk about it in class.

To gain a basic insight about standard deviation, look at the formula: (I know, it's not very pretty)

$$\sigma = \sqrt{\sum_{all\,x}(x-\mu)^2 \times pr(X=x)}$$

What if all of the outcomes are the same? Then, intuitively, there is no variability. If all of the outcomes are the same, then they are all equal to the mean. So, in the formula, all of the $(x-\mu)^2$ terms are equal to 0. So their weighted sum is also 0, and the standard deviation is zero – agreeing with our intuition. On the other hand, suppose that there is a lot of spread in the outcomes. Then many of the outcomes will be a long way from μ. These outcomes will have squared deviations $(x-\mu)^2$ that are very large. The weighted squared deviations will also be large. So the standard deviation will be large – agreeing with our intuition.

Before discussing more about the standard deviation, I want to give you two computing examples to insure that you can calculate the value of the standard deviation.

Example 5. (Same as Example 1.) Toss a fair coin. Let *X* = 1 if a head occurs. Let *X* = 0 if a tail occurs.

Then X is a random variable with the following distribution:

X	$pr(X = x)$
0	½
1	½

What are the mean absolute deviation, the variance, and the standard deviation of *X*?

Solution: First we need the mean of *X*. Recall that in the section on the mean, we calculated the mean to be $\mu = \sum_{all\,x} x \times pr(X=x) = 1\times\frac{1}{2}+0\times\frac{1}{2}$ = ½. To calculate the mean absolute deviation,

[2] The standard deviation is always greater than (sometimes equal to) the mean absolute deviation.

[3] For the normal distribution that you will read about later, the standard deviation is exactly 25% larger than the mean absolute deviation.

the variance, and the standard deviation of *X*, it is convenient to insert additional columns into the table of the distribution:[4]

X	$\lvert x-\mu \rvert$	$(x-\mu)^2$	$pr(X = x)$
0	$\lvert 0-1/2 \rvert=1/2$	$(0-1/2)^2=1/4$	½
1	$\lvert 1-1/2 \rvert=1/2$	$(1-1/2)^2=1/4$	½

x	$\lvert x-\mu \rvert \times pr(X = x)$	$(x-\mu)^2 \times pr(X = x)$	$pr(X = x)$
0	$\lvert 0-1/2 \rvert \times 1/2=1/4$	$(0-1/2)^2 \times 1/2=1/8$	½
1	$\lvert 1-1/2 \rvert \times 1/2=1/4$	$(1-1/2)^2 \times 1/2=1/8$	½

Then the sum of column 2 immediately above is the mean absolute deviation = ¼ + ¼ = ½.
The sum of column 3 immediately above is the variance = 1/8 + 1/8 = ¼.
The square root of the variance is the standard deviation = $\sqrt{1/4}=1/2$. We interpret this to say that the average distance of the possible outcomes from the mean is ½ (which is exactly true in this example because both 0 and 1 are ½ unit from the mean of 1/2.) In this example, the standard deviation and the mean absolute deviation are the same number.

Example 6. (Same as Example 2.) Throw a fair die once and let *X* denote the number of spots on the side facing up. So the distribution of the random variable *X* may be represented in table form as:

X	1	2	3	4	5	6
pr(*X*=*x*)	1/6	1/6	1/6	1/6	1/6	1/6

What are the mean absolute deviation, the variance, and the standard deviation of *X*?

<u>Solution:</u> First we need the mean of *X*. Recall that in the section of this Topic Note on the mean, we calculated the mean to be $\mu=\sum_{all\ x} x\times pr(X=x)=1\times\frac{1}{6}+2\times\frac{1}{6}+\cdots+6\times\frac{1}{6}=3.5$. To calculate the mean absolute deviation, the variance, and the standard deviation of *X*, it is convenient to insert additional rows into the table of the distribution:[5]

x	1	2	3	4	5	6
$\lvert x-\mu \rvert$	|1-3.5|=2.5	|2-3.5|=1.5	|3-3.5|=0.5	|4-3.5|=0.5	|5-3.5|=1.5	|6-3.5|=2.5
$(x-\mu)^2$	$(1-3.5)^2$ = 6.25	$(2-3.5)^2$ = 2.25	$(3-3.5)^2$ = 0.25	$(4-3.5)^2$ = 0.25	$(5-3.5)^2$ = 2.25	$(6-3.5)^2$ = 6.25
pr(*X*=*x*)	1/6	1/6	1/6	1/6	1/6	1/6
$\lvert x-\mu \rvert \times$ *pr*(*X*=*x*)	2.5 × 1/6 = 0.417	1.5 × 1/6 = 0.25	0.5 × 1/6 = 0.083	0.5 × 1/6 = 0.083	1.5 × 1/6 = 0.25	2.5 × 1/6 = 0.417
$(x-\mu)^2$ × *pr*(*X*=*x*)	6.25 × 1/6 = 1.042	2.25 × 1/6 = 0.375	0.25 × 1/6 = 0.042	0.25 × 1/6 = 0.042	2.25 × 1/6 = 0.375	6.25 × 1/6 = 1.042

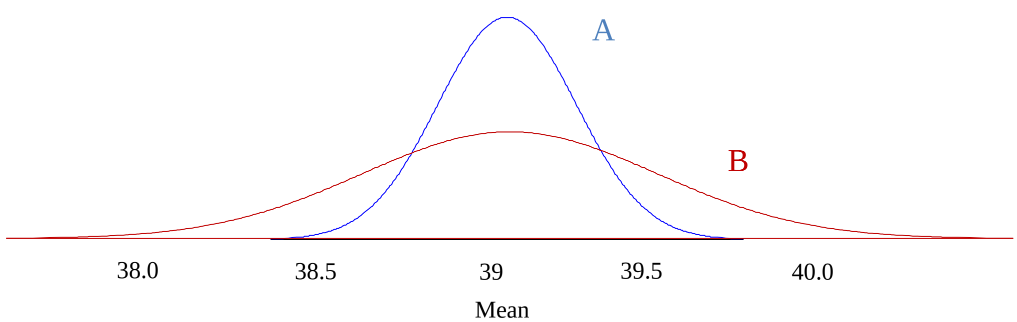
[4] This is really easy to do in a spreadsheet like Excel.
[5] This is really easy to do in a spreadsheet like Excel.

By summing the 5th row, we obtain the mean absolute deviation = 0.417 + 0.25 + 0.083 + 0.083 + 0.25 + 0.417 = 1.5.
By summing the 6th row, we obtain the variance = 1.042 + 0.375 + 0.042 + 0.042 + 0.375 + 1.042 = 2.918.
The square root of the variance is the standard deviation = $\sqrt{2.918} = 1.708$. We interpret this to say that the "average" distance of the possible outcomes from the mean is 1.708.[6]

Now with the computing examples done, let us continue to explore the meaning of standard deviation. It is easiest to grasp the intuitive meaning of standard deviation graphically, as in the following example.

Example 7. The net weight of a can of Folger's coffee.[7]

Suppose we pick a can of Folger's coffee at random from the section of the grocery store shelf where 39-ounce cans of Folger's coffee are displayed. Let X denote the net weight of the coffee in the can. Since we are uncertain about the actual value, then X is a random variable with a distribution. The following graph shows two possible distributions for X. The possible outcomes are represented by the horizontal axis and the probabilities by the curve.



Both distributions have mean = 39 ounces. Curve A in blue shows distribution A with standard deviation = 0.15 ounces. Curve B in red shows distribution B with standard deviation = 0.30 ounces. It is reasonable to say that distribution B is twice as spread out as distribution A. Intuitively, it is clear that we are considerably less certain about the net weight of the coffee if distribution B is the truth than if distribution A is the truth. It would be reasonable to say that the uncertainty is twice as much for distribution B as for distribution A.

<u>The standard deviation as a measure of risk.</u> In Example 7, a consumer's risk of buying an underweight can is the same for both distributions – namely, there is a 50% chance that a can will weigh less than 39 ounces whichever distribution is the truth (assuming unrealistically that these are the only two possible distributions.) However, there is a greater risk that B will be seriously underweight than that A will be. For example, one can calculate that about 16% of the cans from A weigh less than 38.84 ounces. However, about 31% of the cans from B weigh less than 38.84 ounces. You will learn how to calculate these percentages in the section on normal probabilities.

[6] The exact average distance is the mean absolute deviation, 1.5, which is a little smaller than the standard deviation.
[7] Expanded from Example 9 in the Topic Note on Uncertainty Distributions and Random Variables, and from Example 4 in the Mean section of this Topic Note.

Sidebar. The standard deviation as a measure of estimation uncertainty. {You are not responsible for this paragraph – yet!} I may be getting a little ahead of myself with this point, but you might like a sneak preview of the major use that we will make use of standard deviations in this course. If not, don't worry about it. We will study data and make estimates. For example, we may analyze past sales data and forecast that sales next quarter will be $15 billion. Are we sure? Of course not! How could we assess our uncertainty? If we *were* certain, we would expect no deviation between our forecast and actual sales. The more uncertain we are, the more deviation we expect between the forecast and actual sales. Is there any way we can estimate how much deviation there will be between the sales forecast and actual sales? Since we are uncertain about future sales, there is an uncertainty distribution. Our forecast is one of the possible outcomes of that uncertainty distribution. The forecasting methodology allows us to estimate the standard deviation of that uncertainty distribution. (The details on how and why will come later.) This tells us how much we can expect our forecast to differ from actual sales. (But we have no idea whether the deviation will be positive or negative – i.e., whether our forecast will be too big or too small – we can only estimate its *magnitude*.) The actual deviation may be bigger or smaller, but the average (standard) deviation gives us an idea. So if our forecast is $15 billion and the standard deviation is estimated to be $2 billion, then we can expect actual sales to differ by about ±$2 billion from our forecast.
[End Sidebar.]

Choosing the right distribution. You have probably noticed that there are often a number of possible distributions that *could* apply to an uncertain phenomenon. For example, for Folger's coffee, we could have a distribution with standard deviation = 0.15 or 0.30. Is there any reason why the standard deviation could not be some other value? And for that matter, does the mean have to equal 39? What if the cans are being underfilled on average? Then couldn't the mean be less than 39? And why does the distribution have to be bell-shaped? These are very good questions. The answer to all of them is that the true distribution *could* be otherwise.

How do we determine what the true distribution is? From one point of view, that is what this course is all about! When we propose a distribution for an uncertain phenomenon, we are *modeling* that phenomenon. We do not know what the true distribution is. But we may have some data values (outcomes) from that distribution or from related distributions. Since the data values come from the true distribution, they may offer clues to infer what the true distribution is. We will use available data to select a best model from the available candidate models. We will be able to assess how good the best model is and to use that model to make predictions – for example, to forecast sales for next quarter. The details of how we do this will be revealed as the course progresses.

Your immediate take away from this discussion is that multiple distributions *could* apply to an uncertain phenomenon and we often do not know which distribution applies – but, with the help of relevant data, we can often make a good choice. To manage uncertainty, we build statistical models, which involves proposing a model, verifying the model with available data, and then using the model to make inferences and forecasts.

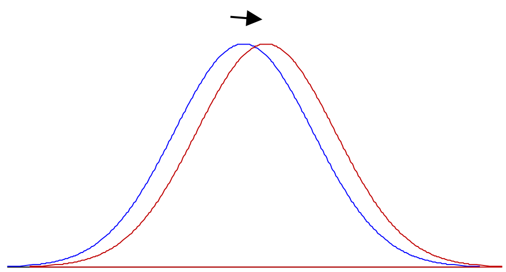
Two technical notes. I will use these properties on occasion later on.

Tech Note 1. The effect of adding a constant to a random variable.

What happens to the distribution of a random variable if we add a constant to the random variable? Here is an example of why this question might be of interest:

Example 8. Let X denote Dell Computer Company sales for next quarter. X is a random variable, as we already know.[8] Suppose that Dell learns that a major customer will add \$1 billion to the order that the customer was planning to place next quarter. Then X will increase by \$1 billion. Let Y be the new value of sales.

What happens to the distribution of uncertainty for sales? How does Y relate to X? Clearly, every possible value in the old X distribution will increase by \$1 billion since total sales will increase by \$1 billion, regardless of the value that old sales would have been. Thus $Y = X + 1$. Clearly, the probabilities will be carried along as well: The \$1 billion increase is certain. So whatever uncertainty there was about sales being \$14 billion before the \$1 billion increase, there is the same uncertainty about sales being \$15 billion after the \$1 billion increase. That is, if the probability of x was 0.05 in the original distribution before the order increase, then the probability of x + \$1 billion will also be 0.05 in the new distribution after the order increase. So $pr(Y = x + 1) = pr(X + 1 = x + 1) = pr(X = x)$. The graph below tells the story. The red curve shows the distribution after the order increases all x outcomes by \$1 billion. The shape does not change; only the location changes.



The take away: When the entire distribution shifts by a constant amount, the mean also shifts by the same amount, but the spread (the standard deviation) is unaffected.

Here is a calculating example:

Example 9.[9] Ten students were randomly selected from a fulltime UT MBA class. Each student was asked to provide anonymously, to the nearest \$1000, the annual salary that he/she anticipates to make in his/her first job after graduation (including bonus). The ten students reported the following anticipated salaries (in \$1,000s): 110, 160, 120, 95, 175, 80, 120, 130, 120, 110. Suppose one student is picked at random from the ten. Let X denote his/her anticipated salary. The distribution of X is:

X	80	95	110	120	130	160	175
$pr(X=x)$	0.1	0.1	0.2	0.3	0.1	0.1	0.1

In the Mean section, we learned that the mean of X is

$$\sum_{all\ x} x \times pr(X = x) = 80 \times 0.1 + 95 \times 0.1 + 110 \times 0.2 + \cdots + 175 \times 0.1 = 122$$

Suppose that each student learns that he or she will be awarded an additional \$10,000 bonus. Then the appropriate random variable changes from X to $Y = X + 10$. The new distribution is shown below:

y	90	105	120	130	140	170	185
$pr(Y=y)$	0.1	0.1	0.2	0.3	0.1	0.1	0.1

[8] Cf. Example 10 in the Topic Note on Distribution and Random Variable.

[9] Based on Example 3 from the discussion of the Mean.

Using the distribution of Y, we can now calculate

$$\sum_{all\ y} y \times pr(Y = y) = 90 \times 0.1 + 105 \times 0.1 + 120 \times 0.2 + \cdots + 185 \times 0.1 = 132$$

The mean of Y increases by 10.
Since every value increases by 10 and the mean also increases by 10, then the deviations from the mean all remain the same, and the standard deviation remains the same.

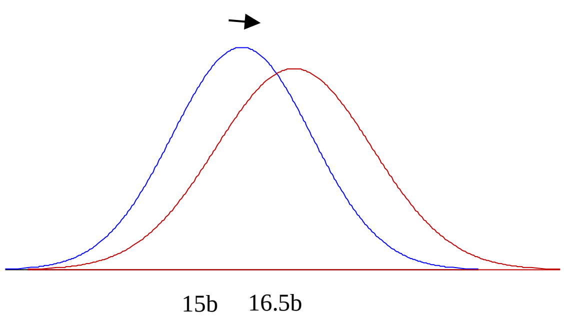
Tech Note 2. The effect of multiplying a random variable by a constant.

What happens to the distribution of a random variable if we multiply the random variable by a constant? Here is an example of why this question might be of interest:

Example 10. Let X denote Dell Computer Company sales for next quarter. X is a random variable, as we already know.[10] Suppose that an improving economy leads Dell to expect that each customer will increase purchases by 10%. Then X will increase by $0.1X$. Let Y be the new value of sales.

What happens to the distribution of uncertainty for sales? How does Y relate to X? Clearly, every possible value in the old X distribution will increase from x to $x + 0.1x = 1.1x$, regardless of the value that old sales would have been. Thus $Y = 1.1X$. Clearly, the probabilities will be carried along as well: That is, if the probability of x was 0.05 in the original distribution, then the probability of $1.1x$ will also be 0.05 in the new distribution. So $pr(Y = 1.1x) = pr(1.1X = 1.1x) = pr(X = x)$.

The distributions of X and Y are shown below. Both the mean and the standard deviation are multiplied by 1.1. If you look closely, you can see that the red distribution is a little wider (10% wider) than the blue distribution. Since the total probability under the red curve must remain at 1, the red curve probabilities also decline (by 10%), which is reflected in the lower height of the red curve. So the red curve is both wider and shorter than the blue curve in order to keep the total area at 1.



Here is a calculating example:

Example 11.[11] Throw a fair die once and let X denote the number of spots on the side facing up. So the distribution of the random variable X may be represented in table form as:

x	1	2	3	4	5	6
$pr(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

[10] Cf. Example 10 in the Topic Note on Uncertainty Distributions and Random Variables.
[11] From Example 6 in discussion of Standard Deviation.

Now suppose that we double the number of spots on each side of the die. Let *Y* be the number of spots that will be on the up side when the new die is thrown. What are the mean and standard deviation of *Y*?

Note that *Y* = 2*X*. So all of the outcomes of *X* are multiplied by 2 and the corresponding probabilities are carried along. The distribution of Y is as shown in the table below:

Y	2	4	6	8	10	12
pr(Y=y)	1/6	1/6	1/6	1/6	1/6	1/6

Recall that in the section on the mean, we calculated the mean of *X* to be

$\sum_{all\ x} x \times pr(X = x) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5$. The mean of *Y* is calculated as

$\sum_{all\ y} y \times pr(Y = y) = 2 \times \frac{1}{6} + 4 \times \frac{1}{6} + \cdots + 12 \times \frac{1}{6} = 7.0$. The mean of *Y* is twice as big.

The mean absolute deviation, the variance, and the standard deviation of *Y* may be calculated as follows:

y	2	4	6	8	10	12
$\mid y - \mu \mid$	|2-7|=5	|4-7|=3	|6-7|=1	|8-7|=1	|10-7|=3	|12-7|=5
$(y - \mu)^2$	$(2-7)^2$ = 25	$(4-7)^2$ = 9	$(6-7)^2$ = 1	$(8-7)^2$ = 1	$(10-7)^2$ = 9	$(12-7)^2$ = 25
pr(Y=y)	1/6	1/6	1/6	1/6	1/6	1/6
$\mid y - \mu \mid \times$ *pr(Y=y)*	5 × 1/6 = 0.833	3 × 1/6 = 0.5	1 × 1/6 = 0.167	1 × 1/6 = 0.167	3 × 1/6 = 0.5	5 × 1/6 = 0.833
$(y - \mu)^2 \times$ *pr(Y=y)*	25 × 1/6 = 4.167	9 × 1/6 = 1.5	1 × 1/6 = 0.167	1 × 1/6 = 0.167	9 × 1/6 = 1.5	25 × 1/6 = 4.167

By summing the 5th row, we obtain the mean absolute deviation = 0.833 + 0.5 + 0.167 + 0.167 + 0.5 + 0.833 = 3.0.

By summing the 6th row, we obtain the variance = 4.167 + 1.5 + 0.167 + 0.167 + 1.5 + 4.167 = 11.667.

Recall from the Standard Deviation section that the mean absolute deviation and the variance for *X* were 1.5 and 2.918, respectively. So the mean absolute deviation of *Y* is two times that of *X*, and the variance of *Y* is four times that of *X*.

The square root of the variance is the standard deviation of *Y* = $\sqrt{11.667} = 3.416$. We see that the mean and standard deviation of *Y* are exactly 2 times the mean and standard deviation of *X*.

The take away: When the entire distribution is multiplied by a constant amount, both the mean and also the spread (the standard deviation) are multiplied by the constant amount.

Where will these two technical notes be used?

- One place will be in calculating probabilities for normally distributed random variables. It turns out that probability calculations for normally distributed random variables can be simplified by converting them to a standard form. This involves subtracting the mean (a constant) and then dividing by the standard deviation (another constant). Then only one table or only one computer function need be used for normal distributions.

- Another place will be in regression. Regression proposes an uncertainty model for the relationship between a random variable *Y* and another variable *X* as *Y* = $\alpha + \beta x + error$. By changing the constant α , the distribution of *Y* is moved up or down to find the right intercept. By adjusting the constant β , the distribution of *X* is multiplied to find the right slope.

SUMMARY

The mean and standard deviation are the two most important numerical summary measures of a distribution.

The mean measures the location of the outcomes of a distribution – whether they tend to be big or small.

- The mean is a weighted average of the outcome values in a distribution: $\mu = \sum_{all\ x} x \times pr(X = x)$
- The mean is the value we expect the average of repeated draws from the distribution to converge to.
- The mean is the center of gravity of the distribution.
- The mean is the value from which the negative deviations exactly cancel the positive deviations.

The standard deviation measures the typical variability of the outcomes of a distribution – whether they tend to be close together or far apart.

- The standard deviation is the square root of a weighted average of the squared deviations of the outcome values from the mean: $\sigma = \sqrt{\sum_{all\ x} (x-\mu)^2 \times pr(X = x)}$.
- The standard deviation is best thought of as the average magnitude of the deviations of the outcome values from the mean.
- Uncertainty originates from variability.
- So the standard deviation is a fundamental measure of uncertainty. The larger the standard deviation, the more uncertainty.

Adding a constant to a random variable increases the mean by the value of the constant, but does not affect the standard deviation.

Multiplying a random variable by a constant multiplies the mean by the constant and multiplies the standard deviation by the absolute value of the constant.