STATISTICS TOPIC NOTES

Estimation of the Mean and Standard Deviation

and

Sampling Distributions

In this section of the Topic Notes, I discuss managing the uncertainty of sampling. We sample an uncertainty distribution in order to learn about it. The key information about a distribution is its mean and standard deviation. So in this Topic Note, I discuss estimation of the mean and standard deviation of an uncertainty distribution. I also discuss the related concept of sampling distribution, which lets us calculate the uncertainty of the estimates. The latter is a BIG CONCEPT in statistics. An estimate is uncertain and deviates from its target. But with the help of this Topic Note, we can manage the uncertainty of that deviation. To me it has always seemed magical that we can tell how close an estimate is likely to be to the truth without knowing much at all about what we are doing. Everywhere we turn, the Normal Distribution looms like a giant lighthouse to guide us safely into port.

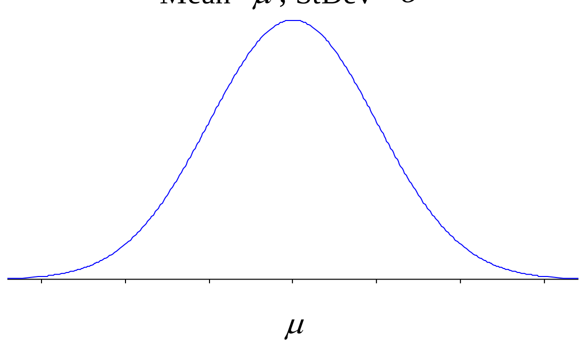
Estimation of the Mean (μ) and the Standard Deviation (σ)

Let us begin with an example to illustrate ideas. This example is more advanced than my usual baby-step starting examples. But I want to show you where we are heading before I retreat to more elementary pedagogical examples.

Consider the uncertain phenomenon that is the starting annual salary of a randomly selected graduate of U.S. MBA programs last year. Then the outcome set of the distribution is the set of all starting salaries of all graduates of U.S. MBA programs last year. The probability of picking any given graduate and his/her salary is the same as for any other graduate. However, most of the salaries clump together around a central value, with the frequency of salaries diminishing at both lower and higher levels. The graduates whose salaries are similar can be "stacked up" at their values. The "stacks" of individual graduates' salaries are highest for middle-level salaries, where most of the graduates are, and shorter away from the middle. The stacks make a distribution of salaries that *may* look similar to the following distribution:[1]

Possible distribution for Salary

Mean= μ , StDev= σ



[1] I have drawn the normal probability curve for convenience. The actual probability curve may not look like this. My discussion does not depend upon the normal distribution being the true distribution.

Suppose that I would like to learn the value of μ, which is the mean of all graduates' starting salaries. The only way to know this value for sure is to track down all graduates and convince them to state their salaries truthfully, then average all of their salaries. Such a brute force approach is likely to be an exercise in frustration, futility, and cost. Similarly, if I would like to learn the value of σ, once I have tracked down and persuaded or coerced all population members into revealing their salaries, I can compute the value of σ by brute force using the formula in the Topic Note on means and standard deviations.

An alternative approach to learning about the distribution of MBA starting salaries is to examine a small and hopefully representative sample of the population of all graduates. We can more easily and at less time and cost pursue a small sample of graduates to get their salaries and average them. This sampling approach yields a manageable project, but at the price of sacrificing certainty. However, if we follow simple statistical principles, we can manage the uncertainty of the sampling approach. We will do this by calculating how much the estimate can be expected to differ from the truth. This Topic Note will explain how to do this for the mean.[2]

Before I get into the meat of this Topic Note, I need to lay some groundwork by being sure that you understand the sample mean and sample standard deviation, for we will use them repeatedly as estimates of the population mean and population standard deviation.

The Sample Mean ($\bar{x}$) and the Sample Standard Deviation (s)

The sample mean and sample standard deviation are, with one minor difference, the same as the population mean and population standard deviation that were introduced in the Topic Note on means and standard deviations – except that the sample versions are applied to a sample of data instead of to the whole population. Here is an example.

Example 1. From a class of 65 MBA students, 10 students were randomly selected and anonymously provided the values of the annual salaries that they expect to make after graduation, to the nearest $1000. The 10 responses were 130, 125, 100, 118, 125, 105, 100, 125, 160, 72. The 10 responses are one sample. The 65 students' expected salaries are the population.

Since the 10 students were randomly selected, we may hope that the distribution of their responses is similar to the distribution of the 65 students. Justification for this hope is given below in Point 1 of this Topic Note. But the 10 may not be representative. They may be too high or too low, too spread out or too narrow. But it is harder for a randomly selected 10 to deviate materially from the 65 than it is for a randomly selected 5 to deviate materially, and easier than a randomly selected 20 to deviate. A sample of all 65 would not deviate at all.

If the distribution of the 10 is similar to the distribution of the 65, then the 10 could be used as a proxy for the 65. In particular, it would be reasonable to expect that the mean and standard deviation of the 10 would be reasonably close to the mean and standard deviation of the 65. This is justified below in Point 2 of this Topic Note. The mean and standard deviation of the 10 could be used as *estimates* for the mean and standard deviation of the 65.

[2] A somewhat different procedure can be used to manage the uncertainty of the standard deviation or variance estimates, but you will not need it for this course.

The figure below is an extract from an Excel worksheet that shows the calculation of the sample mean and sample standard deviation of the 10 anticipated salaries.

Annual Salary Expectations

Random sample of ten expected salaries in first year after MBA:

	Expected Salary	Mean	Deviation from mean	Absolute deviation	Squared deviation	
1	130	116	14	14	196	
2	125	116	9	9	81	
3	100	116	-16	16	256	
4	118	116	2	2	4	
5	125	116	9	9	81	
6	105	116	-11	11	121	
7	100	116	-16	16	256	
8	125	116	9	9	81	
9	160	116	44	44	1936	
10	72	116	-44	44	1936	
Sum	1160		0	174	4948	
Mean	116.0		0	17.4	494.8	549.778
					22.244	23.447

Variance = mean of squared deviations (except divide by 9 instead of 10)

Standard deviation = square root of variance

Mean deviation

Mean absolute deviation

Square root of mean sq deviations

Mean of squared deviations

To calculate the mean of the sample, sum the 10 values and divide the sum by 10. The answer is $\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{1160}{10} = 116$, where $\bar{x}$ (read, "x bar") is the symbol most commonly used for the sample mean of sample data represented by $x_1, x_2, ..., x_{10}$. This computes to the same numerical result as the formula for the mean of a distribution, $\mu = \sum_{all\ x} x \times pr(X = x)$, as was pointed out in the Topic Note on means and standard deviations – provided you treat the 10 sample data 130, 125, 100, 118, 125, 105, 100, 125, 160, 72 as a distribution with equal probabilities of 1/10 each. However – and this is important – the sample mean is NOT equal to the mean of the distribution, unless you are very, very lucky when you draw the sample. The mean of the distribution is the mean of the 65, and there is no reason to expect it will be identical to the mean of the 10. I sometimes see students write "$\bar{x} = \mu$". This is almost never true. A word to the wise!

The motivation for the sample standard deviation is the same as for the population standard deviation: To measure an average magnitude of the data deviations from the mean. To calculate the standard deviation of the sample, begin by calculating the sample mean (116). Next subtract the mean from each of the data to get the deviation from the mean for each datum. In the Excel extract above, these deviations are shown in the column labeled "Deviation from mean". Next, square these deviations. This converts negative deviations to positive because the standard deviation is concerned only with the *magnitudes* of the deviations. The squared deviations are shown in the column labeled "Squared deviation."

An alternative – and more intuitively natural approach to getting magnitudes – is absolute value, as shown in the column labeled "Absolute deviation". Absolute values are sometimes used in statistics, but squares are used far, far more often. (The reason is not readily apparent at the level of this course, so I ask you to trust me!)

Next, compute the mean of the squared deviations – except divide by 9 instead of 10.[3,4] This yields the sample variance = 549.778.

Finally, take the square root of the sample variance to yield the sample standard deviation = $\sqrt{549.778}$ = 23.447. This computes to the same numerical result as the formula for the standard deviation of a distribution, $\sigma = \sqrt{\sum_{all\ x}(x-\mu)^2 \times pr(X=x)}$, provided you treat the 10 sample data 130, 125, 100, 118, 125, 105, 100, 125, 160, 72 as a distribution with equal "probabilities" of 1/9 each (not 1/10). Again, a word to the wise! The sample standard deviation is NOT equal to the standard deviation of the distribution. "$s = \sigma$" is almost never true.

There is an Excel function, STDEV, that does all of this automatically. For example, STDEV(130, 125, 100, 118, 125, 105, 100, 125, 160, 72) = 23.447. You can also put a cell range as the argument. For example, STDEV (B5:B14) = 23.447.

The sample mean shares the same interpretation as the population mean: the total positive deviations from the sample mean exactly cancel the total negative deviations from the sample mean; and the sample mean is the point where a teeter totter would balance if equal weights were placed on the teeter totter at the sample data values.

Likewise, the sample standard deviation shares the same interpretation as the population standard deviation: it is a measure of the "average" magnitude of the deviations of the data from the sample mean. This statement is exactly true of the mean absolute deviation (17.4 in the Excel extract). It is approximately true of the sample standard deviation. Since the mean absolute deviation is easy to understand intuitively, I encourage you to think of the standard deviation in exactly the same intuitive manner as the mean absolute deviation, even though they are not exactly equal numerically.

It is very important to note that the sample standard deviation is a measure of the average deviation of the *individual* data in the sample from the sample mean. (And correspondingly, the population standard deviation is a measure of the average deviation of the *individual* data in the population from the population mean.) The sample standard deviation is NOT a measure of how far the sample mean deviates from the population mean. Explaining how to assess the latter is

---

[3] This question always comes up in class: Why divide by n-1 instead of n? Alas! I have tried but failed to give an enlightening answer at the level of this course. All of my past attempts to explain have merely muddied things and wasted time. Moreover, understanding why adds no appreciable value to the course. There are far more important points to master in statistics. I wish I could ignore the question. The difference between dividing by n and dividing by n-1 is small if n is of even moderate size. But all statistical software, including Excel's STDEV, divides by n-1 (with an option to divide by n for the whole population.) However, for the brave souls who are still reading and who have not been deterred so far, the next footnote starts to outline a reason for dividing by n-1.

[4] Why divide by one less than the sample size? If you have read the preceding footnote and still want to know, read on. The reason has to do with our desire to use the sample variance and sample standard deviation as estimates of the true but unknown population variance and standard deviation. It turns out that dividing by the number of data makes the sample variance a biased estimate of population variance – the estimate is a little low. Dividing by one less than the sample size makes the estimate larger – just enough to make it unbiased. (Trust me! There is a mathematical proof!)

what most of the rest of this Topic Note is about. But the sample standard deviation is an important starting point. (As a teaser preview: the sample mean can be expected to differ from the true distribution mean by approximately the sample standard deviation, divided by $\sqrt{n}$. For Example 1, that is 23.447 / $\sqrt{10}$ = 7.415.)

The formulas for the sample mean and sample standard deviation of *n* sample data $x_1, x_2, ..., x_n$:

$$\textbf{Sample mean } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}\text{ ; }\textbf{sample standard deviation } s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

The main takeaways:

- Sampling is a method for learning about a population.
- The mean and standard deviation are key information to learn about a population.
- The sample mean and sample standard deviation play the same role for samples as the population mean and population standard deviation play for populations: the sample mean assesses typical value of the sample; the sample standard deviation assesses typical deviation of the sample from the mean.

With the groundwork for your understanding of the sample mean and sample standard deviation now laid, I will resume the main story of managing the uncertainty of sampling.

In brief, our procedure will be:

- take a random sample from the population of interest,
- estimate μ by calculating the mean of the sample, and then
- quantify the uncertainty about μ by calculating the standard error of the sample mean.

The following two points set forth the intuition that justifies our procedure:

(Point 1) A randomly selected sample is (probably) like the population of interest.

Why do we select a sample *at random* from the population of interest? Instead, why not use our best judgment to deliberately pick representative values? Answer: because *random selection respects and preserves the proportions of values actually in the population of interest, without needing to know what those proportions are*! [5]

For example, say that (unknown to us) 20% of the population of MBA graduates have salaries between \$80,000 and \$100,000 and we draw a sample of 100 graduates at random. Then the sample will automatically have about 20% of its draws between \$80,000 and \$100,000. The reason is that there will be a 20% chance of any given draw being in the \$80,000 to \$100,000 range if that range has 20% of the graduates. So if 100 draws are made, about 20 of them should be in that range. If the true proportion is 25% instead of 20%, then the randomly drawn sample will automatically have about 25% of its draws between \$80,000 and \$100,000. Whatever the actual proportion is, we do not need to do anything differently – the randomly drawn sample will

[5] There is an additional answer: Because studies show that most people are very bad at deliberately picking representative samples.

automatically *tend* to have about the correct proportion of its draws in any specified range.

On the other hand, if we exercise our judgment in selecting a sample, we may not get the proportions right – we may believe the population is more (or less) concentrated in the \$80,000 - \$100,000 range than it really is, and then we will overdraw (or underdraw) from that range. For example, we may believe that 30% of the population lies between \$80,000 and \$100,000. If so, then we will intentionally draw 30 of our 100 sample items in that range. If the true proportion is 20%, then the \$80,000 to \$100,000 range will be over-represented in the sample and the sample will be biased.

You learned in an earlier Topic Note that the larger the randomly drawn sample, the closer the proportions in the data will come to the actual proportions, so the more representative the sample will be. *The larger the sample size, the more the randomly drawn sample looks like the population from which it was drawn.* This is an important point, and it is intuitive: If you keep on drawing more and more from the population, eventually you will draw out the entire population, and then the sample will be the population!

(Point 2) The mean of a randomly selected sample is (probably) like the population mean.

This follows from Point 1. Since a randomly drawn sample looks like the population from which it is drawn, then the sample distribution looks like the population distribution – with the resemblance increasing as the sample grows.

If the sample looks like its population, then the sample may be used as a proxy for the population. So the numerical characteristics of the sample can proxy the corresponding numerical characteristics of the population. Thus the mean of the sample will be close to the mean of the population; and the standard deviation of the sample will be close to the standard deviation of the population.

Mathematically, if we calculate the sample mean by using the relative-frequency weighted formula $\bar{x} = \sum_{all\ x} x \times \frac{count\ of\ x}{n}$,[6] then the larger the sample, the closer the relative frequency weights, $\frac{count\ of\ x}{n}$, will be to the probabilities $pr(X = x)$. Therefore, the closer the sample mean $\bar{x} = \sum_{all\ x} x \times \frac{count\ of\ x}{n}$ will be to $\sum_{all\ x} x \times pr(X = x)$, which is μ, the mean of the whole population. Similarly for the standard deviation.

The best estimate of the population mean μ is the sample mean $\bar{x}$.[7]

Similarly:

The best estimate of the population variance σ^2 is the sample variance S^2. And an excellent estimate of the population standard deviation σ is the sample standard deviation S.

[6] This "relative-frequency weighted" alternative to the usual "add 'em up and divide by n" formula was discussed in the Topic Note on Mean and Standard Deviations. If you do not recall, now would be a good time to review.

[7] This can actually be proven by advanced mathematical methods by showing that the sample mean is unbiased and that no other unbiased estimate has smaller standard deviation.

Keep in mind what I am trying to do. I want to justify sampling as a substitute for brute force examination of all members of the population of interest. So far, I have argued that random sampling is reasonable because it produces a sample that resembles the population. It follows that the sample can proxy the population. Therefore, the sample mean will be close to the population mean.[8]

But how close? How much can be sample mean be expected to vary from the true mean μ? That is the million-dollar question. I will answer it. But my answer will require the rest of this Topic Note to develop. To answer the question, I need to quantify the uncertainty of using the sample mean to estimate μ. To do this, I need to introduce you to the concept of sampling distribution. The sampling distribution is an uncertainty distribution. It is the uncertainty distribution of the sample mean (and more generally of any estimate). There are many possible values that the sample mean *could be*. And those values have probabilities. So the sample mean has an uncertainty distribution. Outcomes and probabilities. That's all the sampling distribution is!

But experience shows that sampling distribution is a difficult concept to master. In fact, *sampling distribution* is one of the two most difficult concepts in statistics for most students. (The other is the related concept of the difference between prediction error and estimation error.) Mastery of the sampling distribution concept is essential to fully understand statistics. So I will explore sampling distributions thoroughly, starting with very simple examples where you can see all of the moving parts, and proceeding to more complicated examples, where there are too many parts to write down everything in detail.

Sampling Distributions

Let me begin with a simple example.

Example 2. 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

The mean of X is $\mu = 0.5$ and the standard deviation of X is $\sigma = 0.5$.[9]

Now let me draw two times, at random, from the distribution of X. This is the same as tossing the coin twice. Let me calculate the mean of the two draws. Denote the outcome of the first draw by X_1 and the outcome of the second draw by X_2. Then X_1 and X_2 are random variables and each has the same distribution as X, shown above. Next, calculate the mean of the two draws: $\bar{X} = (X_1 + X_2)/2$. We call this statistic the **sample mean**.

[8] And the sample standard deviation will be close to the population standard deviation.

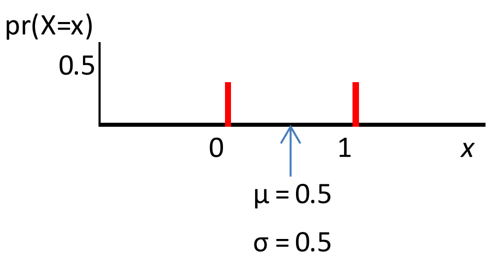
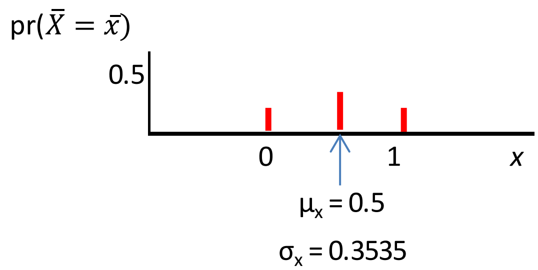
[9] These values were calculated in Examples 1 and 5 in Topic Note on Mean and Standard Deviation.

Now you need to be sure that you understand the following catechism: Is the sample mean $\bar{X}$ a random variable? (Yes.) Are we uncertain about its value? (Yes.) Therefore, does the sample mean have an uncertainty distribution? (Yes.) So does the sample mean therefore have a set of outcomes? (Yes.) And do the outcomes for the sample mean have corresponding probabilities? (Yes.) The distribution of the sample mean $\bar{X}$ – its outcomes and their probabilities – is called the **sampling distribution** of the sample mean. So conceptually, the sampling distribution is just another uncertainty distribution – a concept you are now familiar with – the distribution of an uncertain sample estimate.[10]

Where do the possible outcomes of the sample mean come from? From different samples. If I draw a second sample by tossing the coin two more times, I can compute another sample mean and the values of both sample means will be among all of the possible outcomes for the sample mean. To get ALL possible outcomes for the sample mean, I must examine ALL POSSIBLE samples and calculate the means of ALL of the samples. We are uncertain about the value of the sample mean because sample means vary. Sample means vary because samples vary.

So what is the actual sampling distribution of $\bar{X}$? We need the two things that constitute a distribution: all of the possible outcomes and their probabilities. That's it! A good way to think about this question is to list all of the possible samples, and for each such sample calculate what the sample mean outcome is, then stack up the resulting values. This is a three-step procedure. First, list the outcomes of the original population of interest. These are the 0 and 1 possible outcomes of the coin toss. Second, list all of the possible samples that you can make of these original outcomes: If you write down all of the possible combinations of X_1 and X_2 - in that order, X_1 first followed by X_2 – simple enumeration shows the following four possible samples: (0,0), (0,1), (1,0), (1,1). There are no other possible samples. Third, list the sample mean of each of these samples, and stack them up. The sample means for each are: (0+0)/2 = 0, (0+1)/2 = ½, (1+0)/2 = ½, (1+1)/2 = 1. There are no other possible outcomes for the sample mean. The stack at 0 has one outcome, the stack at ½ has two outcomes, and the stack at 1 has one outcome. Each of the four outcomes is equally likely if the original coin has equally likely head and tail probabilities, which we are given. So the sampling distribution of $\bar{X}$ has outcomes 0, ½, 1, with corresponding probabilities ¼ , ½, ¼. That is the sampling distribution of $\bar{X}$. The process is illustrated in the schematic below.

[10] If you are thinking ahead, you may wonder if other sample statistics can have sampling distributions. Yes. The sample standard deviation, sample proportion, sample median – every sample statistic – has a sampling distribution. We will look at other sampling distributions later in the course.

<table>
<tr><td>1</td><td>Original population</td><td>All possible samples</td><td>All possible means</td></tr>
<tr><td>2</td><td>0,1</td><td>(0,0)
(0,1)
(1,0)
(1,1)</td><td>0
½
½
1</td></tr>
<tr><td>3</td><td><table><tr><td>X</td><td>0</td><td>1</td></tr><tr><td>$pr(X = x)$</td><td>½</td><td>½</td></tr></table></td><td>(0,0)
(0,1)
(1,0)
(1,1)</td><td><table><tr><td>$\bar{x}$</td><td>0</td><td>½</td><td>1</td></tr><tr><td>$pr(\bar{X} = \bar{x})$</td><td>¼</td><td>½</td><td>¼</td></tr></table></td></tr>
<tr><td>4</td><td>
</td><td>(0,0)
(0,1)
(1,0)
(1,1)</td><td>
</td></tr>
<tr><td>5</td><td>"Population 1"</td><td>"Pop 2"</td><td>"Population 3"</td></tr>
</table>

Row 1 of the schematic labels the three steps from left to right: Original population, All possible samples, All possible means. Each of these steps is really a complete and exhaustive list. So each step is a complete "population" or "universe" of all possibilities: all original outcomes, all samples, all means. In recognition that each is an entirety, I call each step a "population" and number them Population 1, Population 2, and Population 3, respectively, in row 5 of the schematic. I call this way of thinking about sampling distributions the **"Three Populations" Concept**.[11] The original population (Population 1) gives rise to the population of all possible samples (Population 2), which gives rise to the population of all possible sample means (Population 3).

In row 2 of the schematic, you see arrows leading from the outcomes of the original population to each sample in Population 2. This reinforces the fact that each possible sample is obtained by drawing (twice) from the original population. You also see a second set of arrows leading from each sample to its mean at the right. This reinforces the fact that every possible outcome for the sample mean is obtained as the arithmetic mean of some sample. There are no other possibilities than these.

In row 3, I have displayed the distribution of the coin toss, with outcomes and

[11] Warning! The "Three Populations" Concept is a Sagerism. You will not find this term in any other book or reference, as far as I know, although it is implicit in the way that sampling distributions are obtained. I find the concept useful in forcing myself to distinguish among the three populations and therefore drive me away from confusing or entangling the *variability of individuals* (Pop 1) with the *variability of means* (Pop 3). You will see later in the course that distinguishing Pop 1 from Pop 3 is key to understanding the difference between prediction error and estimation error – along with sampling distributions the two most difficult concepts in statistics.

probabilities, as a table on the left. I have also collected the distribution of the sample mean in a table on the right. From row 2, you can see that there are three possible distinct outcomes for the sample mean: 0, ½, 1, which occur once, twice, and once, respectively. Since all original outcomes, 0 and 1, are equally likely, all samples are equally likely, and all four means are equally likely. So each has probability ¼. Since the value 0 occurs once, it has probability ¼ in the sampling distribution. Since the value ½ occurs twice, it has probability ¼ + ¼ = ½. Since the value 1 occurs once, it has probability ¼.

Row 4 shows the original and sampling distributions graphically in "stacked" form. In the sampling distribution, there is one mean stacked at 0, there are two means stacked at ½, and one mean stacked at 1. In row 4, I have also added the mean and standard deviation of the original distribution and the sampling distribution. You should be able to calculate these from row 3. If not, please review the Topic Note on Mean and Standard Deviation.

Now let me make an interpretation of the sampling distribution that will foreshadow our subsequent use of it in more realistic settings. Suppose I want to estimate the mean of Population 1. [I know … I can calculate it – the answer is ½ - so I do not need to estimate it. But bear with me, please. I am establishing a principle here, without the distracting and confusing clutter of the real world.] Suppose I will draw twice from the original distribution and calculate the arithmetic mean of my two draws. I will use the resulting sample mean as an estimate of the mean of Population 1. Is the sample mean of two draws uncertain? Yes. So is there an uncertainty distribution for the sample mean? Yes. What is it? Answer: The sampling distribution (Population 3), shown on the right in the schematic.

The sampling distribution embodies the uncertainty of estimation. The sample mean exactly equals the mean that I am estimating with probability ½. This occurs if I get either (0,1) or (1,0) as my sample, both of which have sample mean ½. So I have a 50% probability that my estimate will be exactly correct. My sample mean will be 0.5 too low ¼ of the time. This happens when I get the sample (0,0), which occurs with probability ¼. My sample mean will be 0.5 too high ¼ of the time. This happens when I get the sample (1,1), which occurs with probability ¼. Thus, *the sampling distribution tells me all about the accuracy of estimation*.

Let us look more closely. Notice that the mean of the sampling distribution is $\mu_{\bar{x}}$ = ½, which is exactly the same as the mean of the original population μ = ½, which I am trying to estimate. The interpretation is that the sample mean gets the correct answer, *on average*. But the sample mean gets the correct answer *exactly* right only 50% of the time (see preceding paragraph). The rest of the time, it *deviates* from the right answer. How much does the sample mean deviate from the correct answer? Answer: On average, the sample mean deviates from $\mu_{\bar{x}} = \mu$ = ½ by the standard deviation of the sampling distribution, namely $\sigma_{\bar{x}}$ = 0.3535. (Recall the interpretation of standard deviation as the average magnitude of deviation from the mean. If not, please see the Topic Note on Mean and Standard Deviation.) This tells us the average estimation error and is therefore a fundamental measure of the accuracy of estimation. It is called the **standard error**. *The standard error is just the standard deviation of the sampling distribution.* It tells us the average amount of error that the estimate makes. Sometimes the estimate will miss by more, sometimes by less. On average, it will miss by the standard error (approximately).

Furthermore, the standard deviation of the sampling distribution is related to the standard deviation of the original population: $\sigma_{\bar{x}}$ = 0.3535 = $0.5/\sqrt{2} = \sigma/\sqrt{2}$. That is, Stdev(Pop 3) = StDev(Pop 1)/$\sqrt{2}$. So there is *less* average error in the estimate than in the original population of interest because of dividing by $\sqrt{2}$. You can also see this graphically by looking at row 4 of the schematic. The sampling distribution "mounds up" a little in the center around $\mu_{\bar{x}}$ = ½.

In summary, two remarkable facts about the relationship between Population 3 and Population 1 emerge from this simple example: mean(Pop 3) = mean(Pop 1) and Stdev(Pop 3) = StDev(Pop 1)/ $\sqrt{2}$. In symbols, $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \sigma / \sqrt{2}$. These are very nice properties. The first says that the sample mean gets the right answer *on average*. The second says that the average error of the sample mean can be calculated from, and is smaller than, the average error of individual outcomes. Could these facts be true more generally? Or are they special properties of this simple Example 2? It turns out that they are true very generally. Here is another example.

Example 3.[12] Throw a fair die once and let *X* denote the number of spots on the side facing up. The distribution of *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

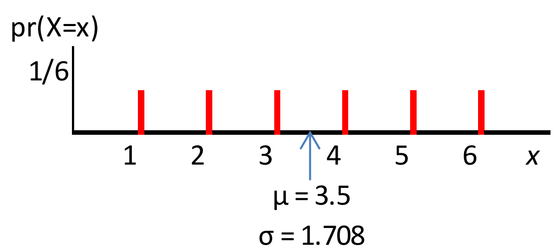
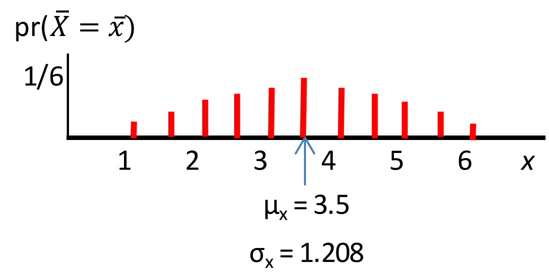
- In Example 2 of the Topic Note on Means and Standard Deviations, the mean of this distribution was calculated to be 3.5.
- In Example 6 of the Topic Note on Means and Standard Deviations, the standard deviation of this distribution was calculated to be 1.708.

Suppose that I want to estimate the mean of *X* by sampling. So throw the fair die twice. This draws a sample of size two from the distribution of *X*. Denote the two outcomes by X_1 and X_2, and let $\bar{X} = (X_1 + X_2)/2$ denote the sample mean of the two throws. The sample mean of the two throws is uncertain. So it has an uncertainty distribution, again called the sampling distribution of the sample mean.

The following schematic shows the gory details of finding the sampling distribution of the sample mean for Example 3. I have displayed the original population, the population of all possible samples, and the population of all possible sample means (the sampling distribution). I will omit detailed discussion in favor of a few brief comments. You will find more details on throwing two dice in Example 6 of the Topic Note on Uncertainty Distributions and Random Variables.

Just briefly, the six equally likely original outcomes can be paired with each other for two throws in 36 equally likely ways. This yields the 36 possible samples shown in the middle of the schematic. So there are 36 sample means (some duplicates) to be stacked up on the right. There are 11 distinct outcomes for the sample mean, with the outcome 7 occurring 6 times, 6 and 8 occurring 5 times, etc. This yields the triangularly-shaped distribution for $\bar{X}$ shown at the right in row 4.

[12] Based on Example 5 from Topic Note on Distributions and Random Variables.

<table>
<tr><td>1</td><td>Original population</td><td>All possible samples</td><td>All possible means</td></tr>
<tr><td>2</td><td>1,2,3,4,5,6</td><td>(1,1)(1,2)(1,3)(1,4)(1,5)(1,6)
(2,1)(2,2)(2,3)(2,4)(2,5)(2,6)
(3,1)(3,2)(3,3)(3,4)(3,5)(3,6)
(4,1)(4,2)(4,3)(4,4)(4,5)(4,6)
(5,1)(5,2)(5,3)(5,4)(5,5)(5,6)
(6,1)(6,2)(6,3)(6,4)(6,5)(6,6)</td><td>1, 1.5, 2, 2.5, 3, 3.5
1.5, 2, 2.5, 3, 2.5, 4
2, 2.5, 3, 3.5, 4, 4.5
2.5, 3, 3.5, 4, 4.5, 5
3, 3.5, 4, 4.5, 5, 5.5
3.5, 4, 4.5, 5, 5.5, 6</td></tr>
<tr><td>3</td><td><table><tr><td>X</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>pr(X=x)</td><td>1/6</td><td>1/6</td><td>1/6</td><td>1/6</td><td>1/6</td><td>1/6</td></tr></table></td><td>(1,1)(1,2)(1,3)(1,4)(1,5)(1,6)
(2,1)(2,2)(2,3)(2,4)(2,5)(2,6)
(3,1)(3,2)(3,3)(3,4)(3,5)(3,6)
(4,1)(4,2)(4,3)(4,4)(4,5)(4,6)
(5,1)(5,2)(5,3)(5,4)(5,5)(5,6)
(6,1)(6,2)(6,3)(6,4)(6,5)(6,6)</td><td><table><tr><td>$\bar{x}$</td><td>1.0</td><td>1.5</td><td>2.0</td><td>2.5</td><td>3</td></tr><tr><td>prob</td><td>1/36</td><td>2/36</td><td>3/36/</td><td>4/36</td><td>5/36</td></tr></table><table><tr><td>3.5</td><td>4.0</td><td>4.5</td><td>5.0</td><td>5.5</td><td>6.0</td></tr><tr><td>6/36</td><td>5/36</td><td>4/36</td><td>3/36</td><td>2/36</td><td>1/36</td></tr></table></td></tr>
<tr><td>4</td><td></td><td>(1,1)(1,2)(1,3)(1,4)(1,5)(1,6)
(2,1)(2,2)(2,3)(2,4)(2,5)(2,6)
(3,1)(3,2)(3,3)(3,4)(3,5)(3,6)
(4,1)(4,2)(4,3)(4,4)(4,5)(4,6)
(5,1)(5,2)(5,3)(5,4)(5,5)(5,6)
(6,1)(6,2)(6,3)(6,4)(6,5)(6,6)</td><td></td></tr>
<tr><td>5</td><td>"Population 1"</td><td>"Population 2"</td><td>"Population 3"</td></tr>
</table>

Now let us see if the same remarkable facts obtain in Example 3 that I noted at the end of the discussion of Example 2. You should be able to calculate the mean and standard deviation of the sampling distribution from the table at right in row 3. If you cannot, please review the Topic Note on Mean and Standard Deviation. I find that $\mu_{\bar{x}} = \mu$ = 3.5. Again, the mean of the sampling distribution exactly equals the mean of the original population of interest.

Furthermore, I calculate $\sigma_{\bar{x}}$ = 1.208 = $1.708/\sqrt{2} = \sigma/\sqrt{2}$. On average, the sample mean deviates from $\mu_{\bar{x}} = \mu$ = 3.5 by $\sigma_{\bar{x}}$ = 1.208. This tells us the average estimation error and is therefore a fundamental measure of the accuracy of estimation. Just as in Example 2, it is called the **standard error**. Just as in Example 2, there is *less* average error in the sample mean than in the individual outcomes because of dividing by $\sqrt{2}$. You can also see this graphically by looking at row 4 of the schematic. The sampling distribution "mounds up" a little in the center around $\mu_{\bar{x}}$ = 3.5.

The same two remarkable facts about the relationship between Population 3 and Population 1 emerge from this Example 3 as from Example 2: mean(Pop 3) = mean(Pop 1) and Stdev(Pop 3) = StDev(Pop 1)/$\sqrt{2}$. In symbols, $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \sigma/\sqrt{2}$.

These remarkable facts are true more generally. They apply to *any* original distribution. Here are two more examples, discussed in abbreviated manner, now that you are familiar with the principles. Example 4 is a lop-sided distribution, whereas Examples 2 and 3 are symmetrical.

Example 4.

An original population consists of three pairs of socks, priced at \$2, \$4, and \$8. They are equally likely. So the distribution of Population 1 is:

x	2	4	8
$pr(X=x)$	1/3	1/3	1/3

Two draws are made with replacement from the original population. There are 9 possible samples in Population 2, and they are equally likely. (See the "Three Populations" schematic below.) Each sample has a mean. So there are 9 possible sample means, and they are equally likely. The 9 possible sample means are collected in Population 3. So the distribution of Population 3 is:

$\bar{x}$	2	3	5	3	4	6	5	6	8
$pr(\bar{X}=\bar{x})$	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9

Since some means have the same values, we may count like-valued means and sort the values in increasing order to yield the conventional representation of the distribution of Population 3:

$\bar{x}$	2	3	4	5	6	8
$pr(\bar{X}=\bar{x})$	1/9	2/9	1/9	2/9	2/9	1/9

Let us check the relationship between the means and standard deviations of Populations 1 and 3:

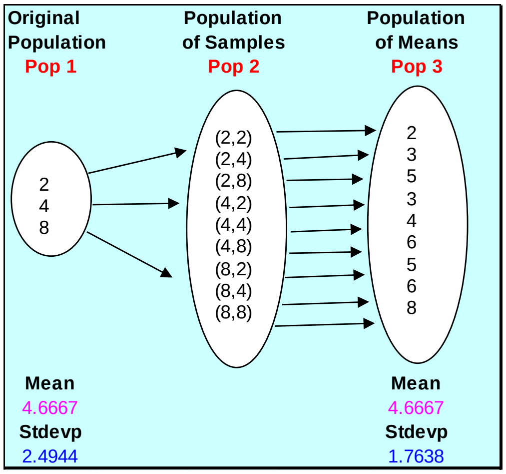
Mean of Pop 1 = $\mu = \sum_{all\,x} x \times pr(X=x) = 2\times\frac{1}{3}+4\times\frac{1}{3}+8\times\frac{1}{3}$ = 4.667.

Mean of Pop 3 = $\mu_{\bar{x}} = \sum_{all\,\bar{x}} \bar{x} \times pr(\bar{X}=\bar{x}) = 2\times\frac{1}{9}+3\times\frac{2}{9}+4\times\frac{1}{9}+5\times\frac{2}{9}+6\times\frac{2}{9}+8\times\frac{1}{9}$ = 4.667.

Or you can simply use the Excel functions **AVERAGE(2,4,8)** and **AVERAGE(2,3,5,3,4,6,5,6,8)** to obtain these figures since the respective values are equally likely. Again, the two means are identical.

Now for the standard deviations. This is more tedious to verify by hand. You may use the spreadsheet approach presented in the Topic Note on Mean and Standard Deviation. However, since the raw values are equally likely in each population, you may also compute the Excel functions **STDEVP(2,3,5,3,4,6,5,6,8)** = 1.7638 and **STDEVP(2,4,8)** ÷ $\sqrt{n}$ = 2.4944 ÷ $\sqrt{2}$ = 1.7638. Again, the standard error of the sample mean is reduced by the factor $1/\sqrt{2}$, compared with the standard deviation of individual outcomes.

Simplified Schematic for Example 4.



Example 5.

Suppose that we would like to know the mean starting salary of all graduates of U.S. MBA programs last year. To approach this problem by sampling, we set up the original population (Pop 1) to be the salaries of all graduates of U.S. MBA programs. Let us again draw two graduates' salaries X_1 and X_2 at random from Pop 1 and calculate their sample mean $\overline{X} = (X_1 + X_2)/2$. The uncertainty distribution for each sample draw consists of all salaries of U.S. MBA program graduates, and the probabilities are all equally likely since we will draw at random.

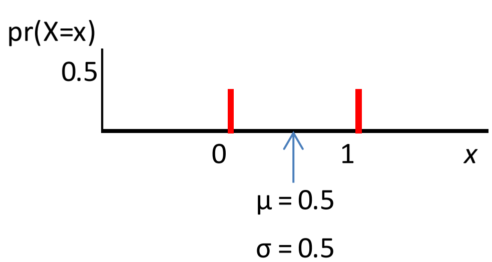
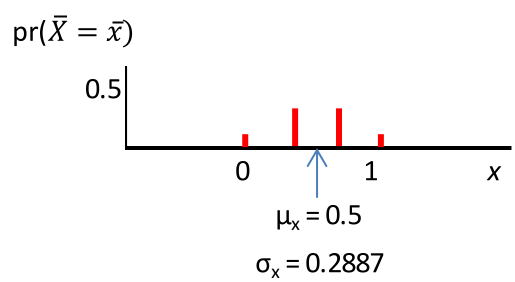
But there is an immediate problem. Unlike the preceding three examples, we do not know what the outcomes in Pop 1 are. We do not even know how many outcomes there are. So we cannot list all possible samples. Nor can we list all possible sample means. Nevertheless, Populations 1, 2, and 3 "exist" – in the sense that an all-knowing umpire could, in principle, track down all graduates of U.S. MBA programs last year and find out what their salaries are. There might be tens of thousands of them. Armed with the list of salaries, the umpire could list all possible samples of two salaries each. If there are 50,000 graduates, there would be 2,500,000,000 possible samples. There would also be 2,500,000,000 sample means. Although the counts are very large, the sampling distribution could, in principle, be created. Not a job I would want, though!

Given that no one is actually going to make a list of any of the Populations for Example 5, is there still anything we can say about the relationship between Pop 1 and Pop 3? Yup! The mean of Pop 3 is identical to the mean of Pop 1, and the standard deviation of Pop 3 equals the standard deviation of Pop 1 divided by $\sqrt{2}$.

How do we know this if we cannot, in practice, actually write down any of the distributions? The answer is that there is a mathematical theorem in statistics that says so. The theorem has a rigorous proof, so that we know it is true. But for our purposes, we can simply accept the result and use it. The theorem says that mean of Pop 3 = mean of Pop 1 and stdev of Pop 3 = stdev of Pop 1 ÷ $\sqrt{2}$, no matter whether we can write down any of the Populations or not, no matter whether the Populations are discrete or continuous, no matter whether the Populations are finite or discrete. All we need do is draw twice and at random.

But most samples in the real world have more than two draws. What can be said about more than two draws? Is the relationship between Pop 1 and Pop 3 for many draws similar to the case of two draws? The answer is yes, and here are a few examples.

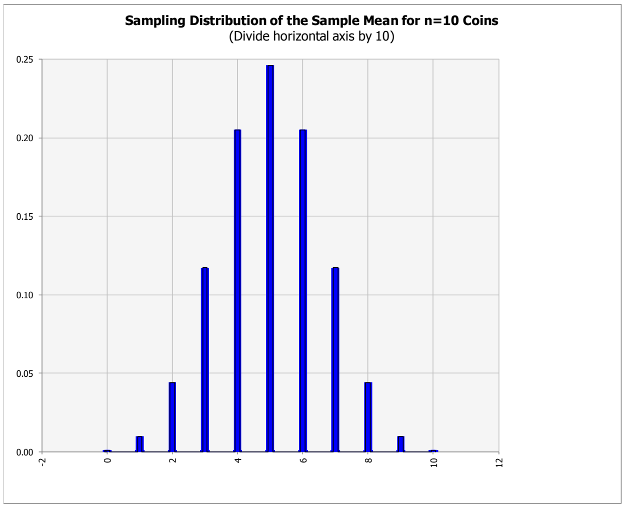
Example 6. [Extension of Example 2] Toss a fair coin three times, with head = 1 and tail = 0 on each toss. Denote the outcome of the first toss by X_1, second toss by X_2, and the third by X_3, and calculate the mean of the three draws: $\overline{X} = (X_1 + X_2 + X_3)/3$. What is the sampling distribution of $\overline{X}$? The details are given in the schematic below.

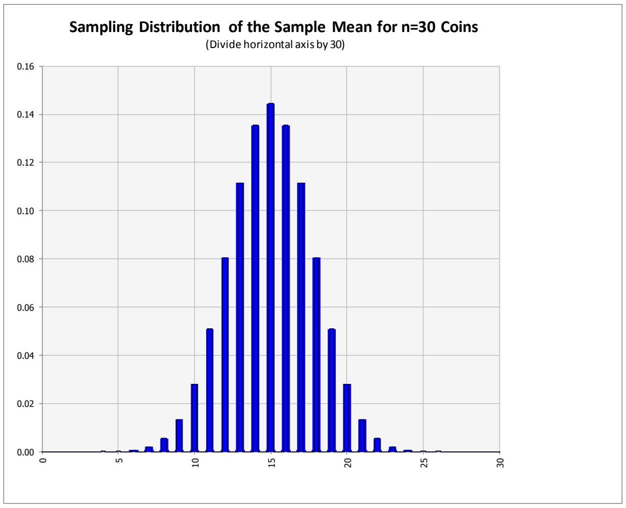
1	Original population	All possible samples	All possible means
2	0,1 →	(0,0,0) (0,0,1) (0,1,0) (0,1,1) (1,0,0) (1,0,1) (1,1,0) (1,1,1)	→ 0, 0.33 → 0.33, 0.67 → 0.33, 0.67 → 0.67, 1
3	X: 0, 1 $pr(X = x)$: ½, ½	(0,0,0) (0,0,1) (0,1,0) (0,1,1) (1,0,0) (1,0,1) (1,1,0) (1,1,1)	$\bar{x}$: 0, 0.33, 0.67, 1 $pr(\overline{X} = \bar{x})$: 1/8, 3/8, 3/8, 1/8
4		(0,0,0) (0,0,1) (0,1,0) (0,1,1) (1,0,0) (1,0,1) (1,1,0) (1,1,1)	
5	"Population 1"	"Pop 2"	"Population 3"

There are 8 possible samples, listed in "Population 2" in the schematic. Each is equally likely if the coin is fair. There are 4 unique possible values for the sample mean, with the values 1/3 and 2/3 each occurring on three outcomes, and the values 0 and 1 each occurring on only one outcome. This leads to the sampling distribution, as shown in rows 3 and 4 of the schematic.

Once again, we compute the mean of the sampling distribution and find that it is the same (1/2) as the mean of the original population. We also can compute (an exercise left for you) the standard deviation of the sampling distribution and find that it is equal to stdev(Pop 1) ÷ $\sqrt{3}$ = 0.5 ÷ $\sqrt{3}$ = 0.2887. Notice that the standard deviation of the sampling distribution is getting smaller in a predictable manner as the number of coin tosses (draws) increases. And the sampling distribution is mounding up even more (see row 4 of the schematic).

The pattern continues. Suppose that we denote by n the number of coin tosses or draws from the original population. Then here are the sampling distributions for the sample mean when n = 10 and n = 30 tosses or draws.



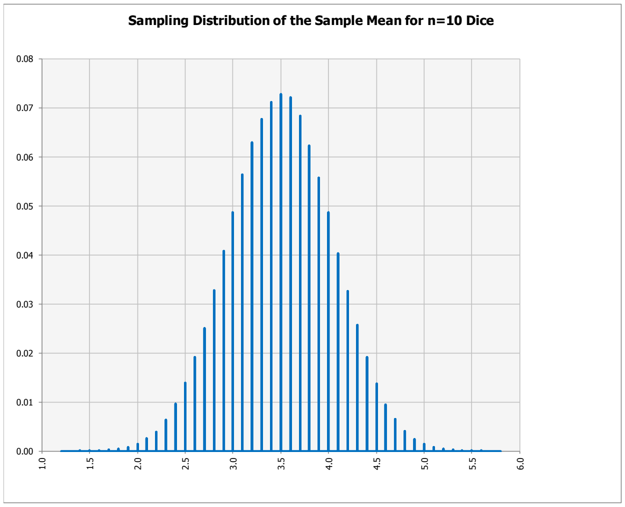


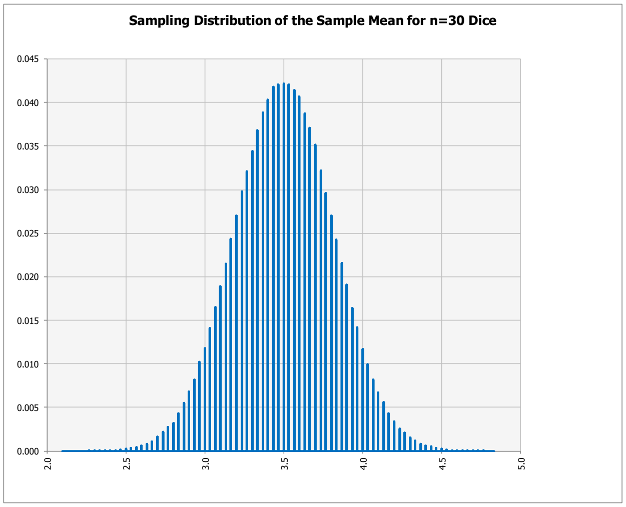
Mean = 0.5, Stdev = $0.5/\sqrt{10}$ = 0.1581

Mean = 0.5, Stdev = $0.5/\sqrt{30}$ = 0.0913.

The number of possible samples and sample mean outcomes for n = 10 is 2^{10} = 1,024.
The number of possible samples and sample mean outcomes for n = 30 is 2^{30} = 1,073,741,824.
Therefore, I do not list either the Population 2 samples or their Population 3 sample means.

For further evidence, here are the sampling distributions of the mean for n = 10 and n = 30 for Example 3 (dice) and Example 4 (socks):[13]

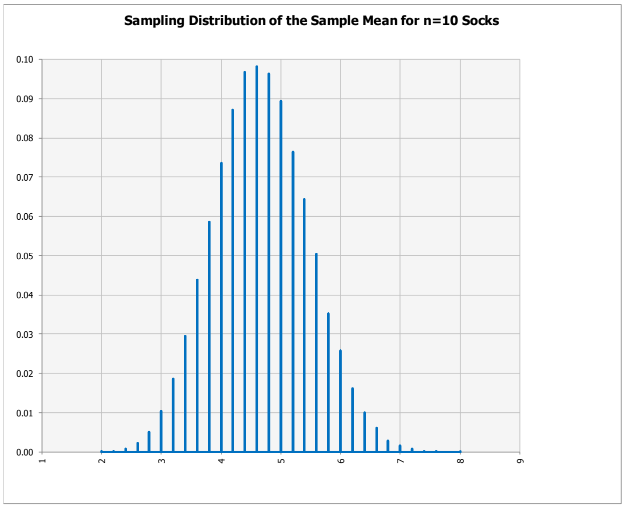


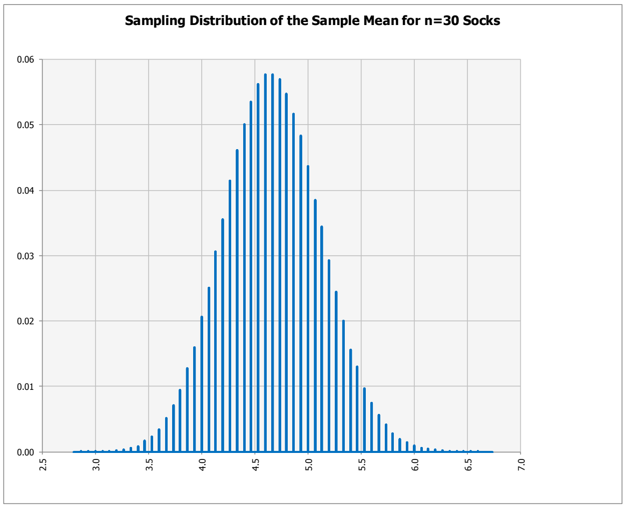


Mean = 3.5, Stdev = $1.7078/\sqrt{10}$ = 0.5400

Mean = 3.5, Stdev = $1.7078/\sqrt{30}$ = 0.3118.

[13] These 4 graphs show close approximations to the actual sampling distributions. In fact, they are based upon 500,000 random samples of size n=10 or n=30, drawn from the appropriate original dice or socks populations. It is too hard to compute the exact sampling distributions. For example, there are about 221 sextillion samples and sample means for 30 dice. But these simulations will be very, very close.





Mean = 4.6667, Stdev = 2.4944/ $\sqrt{10}$ = 0.7888 Mean = 4.6667, Stdev = 2.4944/ $\sqrt{30}$ = 0.4554.

In addition to the relationships between the means and standard deviations of Populations 1 and 3, the above charts for n = 10 and n = 30 now show clearly another feature of Population 3 that was merely hinted in the probability graphs for n = 2 and 3: The mounding up of the sampling distribution looks strikingly like a normal distribution as n increases. At n = 10, the bell shape has clearly emerged. At n = 30, the bell shape looks like a very good approximation to the actual probability curve.

The preceding preparation gives me enough credibility now to state a truly amazing fact: The sampling distribution of the sample mean gets closer and closer to a normal distribution as the number of sample items n (number of draws) gets larger – no matter whether we can write down any of the Three Populations or not, no matter what the original distribution is, no matter whether the original population is discrete or continuous, no matter whether the original population is finite, discrete, or infinite, no matter what the shape of the probability curve of the original population, no matter whether the original population is lop-sided or not – no matter what!

Moreover, in most cases, the sampling distribution is close enough to the normal distribution to use the normal distribution as a reasonable proxy for the sampling distribution if $n \geq 30$. ***This is the key point in all of the development of this Topic Note.*** If the normal distribution applies (approximately), we can calculate probabilities using the Normal Probability Table on Canvas or the Excel function NORMSDIST. Since the sampling distribution embodies the accuracy of estimation (see interpretation at end of discussion of Example 2), this means that we can calculate the approximate probability that the sample mean estimate will be accurate no matter what the original distribution! We can be completely ignorant of the original distribution!! In the real world, we usually are!!!

Before giving you a realistic example to apply these findings, let me collect the key three points in a convenient summary:

Suppose that we draw *n* items $X_1, X_2, ..., X_n$ *at random* from an original population. Denote the mean and standard deviation of that original population by μ and σ, respectively. μ and σ do not need to be known. Denote the mean of the sample by $\overline{X} = (X_1 + X_2 + ... + X_n)/n$. Then $\overline{X}$ is a random variable with possible outcomes and probabilities. The distribution of $\overline{X}$ has the following properties:

- **Property 1.** (**The Central Limit Theorem)** The uncertainty distribution of sample means is approximately normal provided that the number of items *n* in the sample is sufficiently large (30 or more usually works).
- **Property 2.** The population mean of the uncertainty distribution of sample means is μ (the same value as the population mean of the original distribution.)
- **Property 3.** The population standard deviation of the uncertainty distribution of sample means is $\sigma/\sqrt{n}$.

A few important notes:

Note 1. *n* is the number of items in the sample, not the number of outcomes in the original population, and not the number of possible sample means. In the sock example, *n* = 2, because I drew two socks from the three. Every sample list in Pop 2 has 2 sock prices in it. Every sample mean in Pop 3 is computed from 2 socks. In the Austin apartment example (see Example 7), *n* = 60. Every sample list in Population 2 has *n* = 60 apartments on it. Every mean in Population 3 is computed from 60 apartments. There is no "mix-and-match" of sample sizes. If I had drawn a sample of 30 apartments instead of 60, then *n* = 30 – and every sample list in Population 2 would have 30 apartments, and every mean in Population 3 would be computed from 30 apartments.

Note 2. Properties 2 and 3 do not require that *n* be "sufficiently large." They are exactly true for all sample sizes *n*.

Note 3. If Population 1 is normal, then Property 1 holds without any restriction on *n*. It holds when *n* = 1, 2, … and so on.

Note 4. Sampling with replacement versus sampling without replacement.

I need to issue a further explanation of the meaning of sampling "*at random*". Sampling at random means that on each draw the same outcomes and the same probabilities apply. In other words, each draw has the same uncertainty distribution. This will be true if after each draw, we replace the item that we drew on that draw. This is called **sampling with replacement**. The original distribution is restored prior to each subsequent draw. Sampling with replacement permits the same outcome to be drawn again on subsequent draws. ALL of the preceding examples implicitly or explicitly sample with replacement. For example, if you look at Example 4 (socks), you can see that three of the 9 samples include duplicates: (2,2) (4,4) and (8,8). This is NOT the way that we usually sample in the real world. In the real world, once we have seen a sample item, we do not need to see it again. To allow it to repeat would waste time and money. In the real world, we almost always use **sampling without replacement**. If we use sampling without replacement, then the number of outcomes in the original distribution is reduced by one on every draw, and the probabilities change correspondingly.

So what happens to my cherished Properties 1, 2, 3 above if we sample realistically without replacement? Answer: Properties 1 and 2 remain the same; Property 3 is revised to state that the population standard deviation of the distribution of sample means is $\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$, where N is the number of outcomes in the original population. The inserted multiplicative factor $\sqrt{1-\frac{n}{N}}$ is called the **finite population correction factor** (fpc). If I sample the entire original population without replacement, then $n = N$ and so fpc = 0. Then the standard deviation of the sample mean is zero! This makes sense, because sampling the entire original population tells me exactly what the whole original population is – including its mean – so there is no uncertainty.

Note also that if I draw a sample that is a relatively small fraction of the original population, then that fraction n/N will be small, and so the fpc will be close to 1. For example, if I draw a 5% sample, then fpc = $\sqrt{1-0.05}$ = 0.975. In this case, Property 3 without replacement is almost the same as Property 3 with replacement. Since Properties 1 and 2 remain the same regardless, we conclude that there is very little difference between sampling with and without replacement if the sample size is less than (say) 5% of the population size. In this course, I will ignore the difference between sampling with replacement and sampling without replacement if the sample size is less than 5% of the population size.

Note 5. Why does Population 3 mound up in the middle?

We get Pop 3 from Pop 2, which we get from Pop 1. Since we draw each sample at random from Pop 1, then each sample in Pop 2 has the same probability of being selected. So the mean of any sample has the same probability as the mean of any other sample. However, the *values* of these sample means are likely to clump in one place and diminish in frequency as you move away from the clump in either direction. The reason is that it is hard to get a low value for the sample mean, since most of the outcomes in that sample would have to be low. Similarly, it is hard to get a high value for a sample mean, since most of the sample outcomes would have to be high. When you take the average of a generic sample, the high outcomes in the sample tend to cancel the low outcomes in the sample, leaving an average value. So most of the sample means in Population 3 are likely to be average means, with the frequency tailing off away from the vicinity of the average. This is an intuitive argument that the sample means in Population 3 are likely to mound up in the middle and tail off away from the middle.

Note 6. The Central Limit Theorem.

In fact, the mound discussed in Note 5 is approximately normal. The only proviso is that the sample size n – the number of times we draw from the original population to get a sample – be sufficiently large. Remarkably, $n = 30$ is ordinarily sufficiently large. Even more remarkably, the normality of Population 3 applies REGARDLESS of what the distribution of the original population is. In particular, we do not need to know ANYTHING about the original distribution. We can be totally ignorant of both the possible values and the probability curve of the original population of interest. Whew! That is a relief! Because when we do statistics, we are usually trying to learn things about an original distribution. Statistics would be pointless if we had to know those things before we did the analysis. Finally – and most remarkably of all – using the normality of Population 3, we can calculate probabilities for the sample mean without knowing much of anything at all. This allows us to manage uncertainty through sampling. (See Example 7 later.) Getting normality for the probability curve of sample means is so important in statistics that it is called the **Central Limit Theorem** – "Central" because of its central importance.

Random Sampling

Usually, the data we have to analyze can be thought of as repeated draws from the same uncertainty distribution. It makes sense that most statistical data should be of this type. The reason is that when we collect data, we usually want to reduce our uncertainty about a specific phenomenon – Austin apartment rents, the salaries of graduating MBAs, how stock prices relate to earnings, etc. The more data we collect about that particular phenomenon the less uncertain we are. It makes sense that we should try to enforce uniformity in the data collection – drawing each time from the *same* uncertainty distribution, with the *same* outcomes and the *same* probabilities on every draw – in order to ensure that each draw gives us more information about the phenomenon that interests us and not about some other phenomenon.

This sort of data is so common and so important in statistics that there is a name for it: The **Random Sample** (**RS**). The Random Sample is a statistical model for repeated draws from the same uncertainty distribution. Examples 2-6 above and Example 7 to follow are all Random Samples. The importance of Random Sampling is so great that I almost made it into a separate Topic Note. In the end, I decided to integrate it throughout the current Topic Note, but to add some emphasis in this section. Here is the definition:

Definition. A **Random Sample** is a set of random variables $X_1, X_2, ..., X_n$, each of which has the same uncertainty distribution (same set of possible outcomes, same probabilities).

There are two types of Random Samples: **with replacement** and **without replacement**. In Random Sampling with replacement, the outcome that is selected on one draw is put back into the population of outcomes before the next draw. Putting the outcome back recreates exactly the same set of possible outcomes and probabilities. In Random Sampling without replacement, the outcome that is selected on one draw is kept out of the population of outcomes for the next draw.

Not replacing the draws changes the outcome set and their probabilities. Technically, this violates the defining condition of a Random Sample. But if the total number of sample draws is small relative to the population of outcomes, then neither the outcome set nor their probabilities change very much. (See Note 4 above.) In this case, we may treat the data as approximately a Random Sample with replacement.

Most real-world Random Sampling is without replacement. There are several good reasons for this:

- In order to avoid getting the same information again, it is efficient to require that each new draw be a new outcome.
- Sometimes replacing the outcome is not feasible. For example, in some types of product testing, the testing turns the product into a "used" item. For example, opening a can of Folger's coffee to weigh its contents and perform other quality checks. After the testing, the product will not be reassembled and put back into the output stream for consumer use.
- Usually, the sample is a small part of the population of interest. Then random sampling without replacement is approximately the same as random sampling with replacement and the differences may be ignored.
- If the sample is not a small part of the population of interest, the standard error may be adjusted by a simple formula (the fpc – see Note 4 above).

The key point that you need to know is that ***Random Sampling is a prerequisite for Properties 1, 2 , and 3.*** If your data satisfy the simple requirements for random sampling (repeated draws from the same uncertainty distribution), then you may employ Properties 1, 2 and 3. If your data do not satisfy the random sampling requirements, then Properties 1, 2, and/or 3 *may* not hold.

If you feel that you have a good grasp of Random Sampling, you may now skip ahead to Example 7. If you would like to read more about when data are or are not a Random Sample, I have included some examples next:

- RS ex 1: Toss a fair coin twice. This is a RS. Each toss has the same uncertainty distribution, with common outcome set {0,1} and equal probabilities (0.5).
- RS ex 2: Toss an unfair coin twice. This is a RS. Each toss has the same uncertainty distribution, with common outcome set {0,1} and the same probabilities on each toss. Although we do not know what the probabilities of 0 and 1 are, they remain the same on each toss because it is the same coin on each toss.
- RS ex 3: Toss two separate fair coins. This is a RS. Each toss has the same uncertainty distribution, with common outcome set {0,1} and equal probabilities (0.5). It makes no difference that two coins are used, as long as the premise that each is fair is correct.
- RS ex 4: Toss two separate but otherwise identical unfair coins. This is a RS. Each toss has the same uncertainty distribution, with common outcome set {0,1} and the same probabilities for 0 and 1 with each coin.
- RS ex 5: Toss one fair coin and one unfair coin. This is not a RS. Although the outcome set is the same {0,1}, the probabilities are different since one coin is fair and the other is not.
- RS ex 6: Toss a fair coin and roll a fair die. This is not a RS. Although the outcomes {0,1} for the coin are equally likely and the outcomes {1,2,3,4,5,6} for the die are equally likely, the outcome sets are not the same, and neither are the probabilities – e.g., 1 has probability 0.5 for the coin, but 1/6 for the die, and 0 has probability 0.5 for the coin but zero for the die.
- RS ex 7: Draw a sample of 60 Austin apartments at random from the population of all Austin apartments and record their rents. This is (approximately) a RS. If the draws are with replacement, this would be exactly a RS, because the outcome set (the rents) is restored for each draw by the replacement, and the probabilities of the apartments are equally likely and hence the same on every draw. If the draws are without replacement, then the outcome set and the probabilities are almost the same on every draw because 60 is very small in relation to the population of Austin apartments. Then we may treat this *as though* it is a RS.
- RS ex 8: Draw a sample of 10 different students at random from a class of 65 students. This is not a RS. "Different" clues you that this is sampling without replacement. The sample size is 10/65 = 15.4% of the population. This is too large a percentage of the population to ignore the differences between sampling with and without replacement. (Our somewhat arbitrary cut-off is 5%.) By the 10^{th} draw, the outcome set has changed materially, having only 56 of the original 65 students available for the 10^{th} draw. The probabilities have also changed materially, with each remaining student having probability 1/56 = 0.0179, compared with 1/65 = 0.0154 before the first draw. Nonetheless, Properties 1, 2, and 3 are still applicable if the **fpc** adjustment is made in Property 3.[14]

---

[14] However, Property 1 (the Central Limit Theorem) fails on other grounds, because a sample of size 10 is not "sufficiently large". So the sample mean will not be normally distributed unless the 65 students' expected salaries

- RS ex 9: Here is a more sophisticated example. Take a sample of 100 closed-out consumer loans from Big Bank. The sampling is done as follows: From the set of defaulted consumer loans of Big Bank, 50 loans are drawn at random without replacement. From the set of non-defaulted consumer loans of Big Bank, 50 loans are drawn at random without replacement. The goal is to get an accurate estimate of the bank's actual cost of servicing its consumer loans. It is expected that servicing costs are much higher for defaulted loans. However, a random sample of 100 consumer loans without regard for default status might include only a handful of defaulted loans – since most loans do not default – too few for an accurate estimate. Dividing the sample equally between defaulted and non-defaulted loans ensures that an adequate number of high-cost defaulted loans are included in the estimation.

  However, *the 100 loans are not a RS.* The outcome set for the first 50 loans (defaults) is different from the outcome set for the second 50 loans (non-defaults). The sub-population of non-defaulted loans is excluded from the outcome set of the first 50 draws, and the sub-population of defaulted loans is excluded from the outcome set of the second 50 draws. Moreover, the probabilities of selection also differ between the first and second 50 draws. The probability of selecting a non-default is zero on the first 50 draws; the probability of selecting a default is zero on the second 50 draws.

  Nonetheless, each set of 50 loans is a RS – but from two different distributions. Each of the 50 draws from the defaulted population has the same outcome set with the same probabilities. Likewise, each of the 50 draws from the non-defaulted population has the same outcome set with the same probabilities. If Big Bank has at least 1000 defaulted loans and at least 1000 non-defaulted loans, then the 50 loans in each set are a sufficiently small part of their populations that the difference between sampling with and without replacement can be ignored. So each separate sample is (approx.) a RS.

  The sample mean of the 100 loans would give a bad estimate of Big Bank's total cost of servicing its loans. The defaulted loans are (deliberately) over-represented in the sample of 100. So a straight sample mean of servicing costs will be too high. Properties 2 and 3 fail. (Property 1 still holds – but the mean and standard deviation of the normal distribution are wrong.)

  The correct way to get an accurate estimate in this example is to estimate the defaulted and non-defaulted populations separately and use a weighted combination of the separate estimates. You are not responsible for that. In this example, I am focusing only on the sampling features.

It is now time to put all of this together in a more realistic integrated example.

Example 7. Austin apartment rents.

Suppose that I would like to know the mean rent of all Austin apartments. My population of interest is the set of all of the rents of all Austin apartments. I know only vaguely how many Austin apartments there are. I also have only a vague idea what the probability curve of their rents looks like. I do not know what the mean of the population is, nor do I know the standard deviation of the population. So I do not know much about the distribution of Austin apartment

---

(say) are normal, which is unlikely.

rents. But the collection of all Austin apartment rents is my Population 1. Denote the mean of Population 1 by μ and its standard deviation by σ. I want to estimate μ.

To manage my uncertainty, I draw a sample of 60 Austin apartments at random from Population 1. I find the rent of each. Here are the 60 sample rents:

519	530	450	425	470	415	505	470	625	470	659	605
765	580	520	770	700	399	445	470	745	480	650	929
475	995	495	445	450	585	565	700	540	460	750	695
575	565	420	510	785	525	650	455	650	600	455	455
415	620	575	635	485	495	515	550	595	575	430	1050

I have previously established that the best estimate of the population mean is the sample mean. So I compute the mean of the 60 rents $\bar{x}$ = \$572.27 as my estimate of μ.

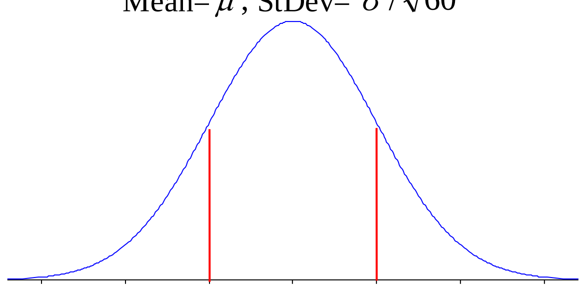
How good is my estimate? How much uncertainty do I now have about the value of μ? One way to address these questions is to reformulate them as follows:

- "How much can I expect my estimate to differ from the true value of μ?"
- "How much confidence can I have that my estimate will deviate by less than ____ from the true value of μ?" (Fill in the blank with a deviation of interest.)

These are questions about the variability of my estimate. Since my estimate is a sample mean, I should look to the variability of sample means to answer the questions. That is, I need to look at the distribution of sample means. By Properties 1, 2, and 3, the uncertainty distribution of sample means looks like the figure below.

By Property 1, the probability curve is (approx.) normal, since the sample size n = 60 is "sufficiently large".
By Property 2, the mean of Population 3 is the same as the mean of Population 1, namely μ.
By Property 3, the standard deviation of Population 3 is the standard deviation of Population 1, divided by $\sqrt{n}$, which is $\sigma/\sqrt{60}$.[15]

Population 3: The Distribution of Sample Mean Rents
Mean= μ, StDev= $\sigma/\sqrt{60}$



$$\mu-\frac{\sigma}{\sqrt{60}} \quad \mu \quad \mu+\frac{\sigma}{\sqrt{60}}$$

---

[15] I drew the sample of 60 without replacement. So I should use the finite population correction (fpc). However, 60 apartments is certainly fewer than 5% of Austin apartments, so it is OK in this case to ignore the fpc.

Therefore, when I calculate my sample mean and get $\bar{x}$ = \$572.27, I am in essence drawing this sample mean from the normal distribution depicted above.

- Therefore, I can expect that my sample mean differs from the true mean μ by about the "average" difference between sample means and the true mean – namely, by about $\sigma/\sqrt{60}$.
- Furthermore, the probability that I draw this sample mean from the region within one standard deviation of the mean of Population 3 is about 68%. (This is the central region between $\mu-\frac{\sigma}{\sqrt{60}}$ and $\mu+\frac{\sigma}{\sqrt{60}}$ in the normal curve above – between the two red lines.) If my sample mean is drawn from this central region, then my sample mean will differ from the true mean μ by less than $\sigma/\sqrt{60}$.[16] Therefore, I have about 68% confidence that my sample mean differs from the true mean μ by less than $\sigma/\sqrt{60}$.

Although these two bulleted statements are both correct assessments of the uncertainty of my sample mean, neither statement is yet useful in practice. This is because both statements reference an unknown quantity – σ. To remedy this problem, I will replace σ by a good, known estimate of σ. You may recall from earlier in this Topic Note that I justified the sample standard deviation as a good estimate of σ. The standard deviation of the 60 apartment rents is *s* = \$140.52. Let us accept for now it is OK to plug the sample standard deviation in for σ in the two bulleted statements. I will discuss the justification later. When I substitute, the two bullets become:

- Therefore, I can expect that my sample mean differs from the true mean μ by about the "average" difference between sample means and the true mean – namely, by about $140.52/\sqrt{60}$.
- Furthermore, the probability that I draw this sample mean from the region within one standard deviation of the mean of Population 3 is about 68%. (This is the central region between about $\mu-\frac{140.52}{\sqrt{60}}$ and $\mu+\frac{140.52}{\sqrt{60}}$ in the normal curve above.) If my sample mean is drawn from this central region, then my sample mean will differ from the true mean μ by less than $140.52/\sqrt{60}$. Therefore, I have about 68% confidence that my sample mean differs from the true mean μ by less than $140.52/\sqrt{60}$.

The standard deviation of the 60 apartment rents is *s* = \$140.52, which is my estimate of σ. Therefore, my estimate of $\sigma/\sqrt{60}$, the standard deviation of Population 3, is \$140.52 / $\sqrt{60}$ = \$18.14.

I now summarize the uncertainty of the sample mean as an estimate of μ in Example 7 as follows:

- I can reasonably expect that my sample mean differs from the true mean μ by about

[16] This is worth checking out. Look at the central region in the graph. Pick any point on the horizontal axis in that central region. Do you see that the distance between the point that you pick and the central value μ is less than the half-width ($\sigma/\sqrt{60}$) of the central region?

±\$18.14. This is because the "average" difference between all sample means and the true mean is about this much, and I expect my sample mean to be "average".

- I can be about 68% confident that my sample mean differs from the true mean μ by less than \$18.14. This is because about 68% of all sample means differ from μ by less than \$18.14.

Comment: It is truly remarkable that we do not need to know anything about the distribution of apartment rents to make these two statements. We do not need to know the mean μ nor the standard deviation σ of Population 1. We do not need to know the form of the probability curve nor a list of the possible outcomes of Population 1. All we need is a randomly selected sample of data from Population 1. Using the sample, we can estimate μ and summarize the uncertainty of the estimate.

Sidebar. A loose end: The uncertainty of σ – The T distribution. {You are not responsible for this sidebar.}

Why is it OK to plug in the sample standard deviation *s* in for σ when assessing the uncertainty of the sample mean? One answer is that *s* is a good estimate of σ, so the value of *s* is likely to be close to σ. But that answer is not entirely satisfactory because, although *s* may be close, it is still not σ - and *s* may not be close. After all, *s* has an uncertainty distribution of its own. So if I replace σ by *s*, am I not injecting additional uncertainty that I have not taken into account?

That is true. Statisticians have developed an additional distribution called (Student's) T distribution to deal with the additional uncertainty from using *s* instead of σ. The T distribution looks similar to the normal distribution, but has a somewhat wider probability curve to make the probability of the central region a little smaller. The probability within one standard deviation of the mean of a T distribution is a little less than 68%. However, for sample sizes of 30 or more, the difference between the normal distribution and the T distribution is mostly not worth bothering about. Since there is not much difference and statistics is already hard enough, I have chosen not to bother you with the additional complexity of the T distribution in this course. However, you will see some computer output labeled with "T-stat" or "pr(T > t)". These notations are referring to the T distribution. If you look in other statistics books, you will see statements about the T distribution. When you see these, just ignore the "T" and think "Z" or "normal" instead.

End sidebar.

More on Population 3 Variability

Recall that Property 3 states that the population standard deviation of the distribution of sample means is $\sigma/\sqrt{n}$. Property 3 deserves some more interpretation:

Property 3 says that the variability of sample means is much less than the variability of the individuals in Population 1. Why is this true? Population 1 is a population of individual values. Population 3 is a population of means. The standard deviation σ of Population 1 measures the variability of individual values. The standard deviation of Population 3 measures the variability of means of groups. The members of Population 1 (individual outcomes) are entirely different types of things from the members of Population 3 (group means). Intuitively, I expect means of groups of individuals to be much less variable than the individuals themselves. If two statistics classes take the same exam, you should expect much less difference between the class means than between the students in either class. The reason is that individual variability within a group is suppressed by averaging: the high individuals and the low individuals cancel each other out in an average. This is the reason (noted in Note 5 above) that Pop 3 mounds up in the middle. By Property 3, the variability of sample means is only a fraction of the variability of individuals.

Property 3 establishes the standard error, $\sigma/\sqrt{n}$, as a fundamental measure of how good the sample mean is as an estimator of the population mean. Property 3 says that we can expect the typical sample mean to differ from the true mean μ by about $\sigma/\sqrt{n}$. Thus, $\sigma/\sqrt{n}$ is a fundamental measure of the uncertainty of using the sample mean to estimate the population mean and is called the standard error of the sample mean.

Furthermore, Property 3 tells us that if we want to reduce this uncertainty by half, then we must quadruple the sample size: $\sigma/\sqrt{4n} = (\sigma/\sqrt{n}) \div 2$. This is another version of the law of diminishing returns: Additional data are valuable, but not as valuable as the initial data.

A key point is that Property 3 *links* the variability of sample means to the variability of individual outcomes. Property 3 says that I do not need a sample of means to estimate the variability of means. This is something that may not have occurred to you before: I can estimate the variability of individual outcomes because the sample (of 60 apartments) that I have is a sample of individual apartment rents. I can look at the 60 individual apartment rents and see how much they vary. The standard deviation of the 60 individual apartment rents is an excellent measure of the variability of individual apartment rents. But what about the variability of sample means? I have only one sample and one mean. I do not have a sample of sample means. I do not have 60 samples, nor 60 sample means. How can I estimate the variability of sample means if I have only one of them? Answer: By Property 3. I can get an excellent estimate of the standard deviation of sample means by dividing the standard deviation of *individual* apartment rents (which I do have) by the square root of 60. Property 3 links the variability of sample means to the variability of individual apartment rents. So I can bypass getting a sample of sample means and use my sample of 60 individual apartments.

To pursue this idea a little further in order to ensure that this horse gets beaten to death, let me propose a thought experiment: Imagine that I decide to collect a sample of sample means in order to estimate their variability directly. Suppose I take 30 samples of 60 apartments each and calculate the sample mean of each of the 30 samples. Then I would have 30 sample means. I could calculate the standard deviation of the 30 sample means and have an estimate of the variability of sample means. Nothing wrong with that – except that I must collect 30 x 60 = 1800 apartment rents rather than just 60. But if I did that, I would want to use all 1800 apartments to estimate μ. I would have a much better estimate of μ by averaging 1800 data than by averaging 60 data. But then I would want to know the variability of the 1800-data sample mean. I would be back where I started – but needing to know the variability of 1800-data means rather the variability of 60-data means.

Do not Construct Population 2 and/or Population 3

Examples 2, 3, 4, 6 are "toy" examples. They are very simple and devoid of most real world interest. In Examples 2, 3, 4, 6, you can actually work out the detailed relationships among Pop 1, 2, and 3, and you can actually verify the relationships among their means and standard deviations. It is useful to go through those simple exercises in order to learn the principles thoroughly. However, it is only in toy examples like Examples 2, 3, 4, 6 that you can actually construct either Population 2 or Population 3. In most real world applications, you cannot. There are several practical reasons for this.

- In any real problem there are just too many possible samples and too many possible sample means. For example, if 10 students are drawn from a class of 60 MBA students at random with replacement and noting the order of selection, then there are about 605 quadrillion possible samples and sample means. Even if you draw without replacement and do not distinguish the order of selection, there are still about 75 billion possible samples and means.
- In many real-world problems, you may not know what the possible values are in Population 1. So you cannot even count the number of samples and means, let alone write them down. For example, if you draw 10 students at random from 60, you can count the number of samples, but you do not know what the 60 possible values are – if you did, you would not need to draw a sample – you could calculate the answer to your question. If you want to sample the potential customers for your product, you do not even know how many there are.
- In the real world, you do not need to construct either Population 2 or Population 3. All that you really need to know about them is contained in Properties 1, 2, and 3. *Population 2 and Population 3 are conceptual devices to help you understand sampling distributions.* Populations 2 and 3 "exist" only in a virtual sense. It is not necessary to make them actual.[17]

So … about constructing Populations 2 and 3 for the real world: It's too hard. You do not want to do it. You do not need to do it. Just don't.

The Three Types of Mean and Variance

At this point in the course, the terms "mean" and "standard deviation" have become ambiguous. Mean of what? Standard deviation of what? We have the sample and two Populations (1 and 3) to which each term might apply. To avoid ambiguity it is a good idea to add further modifiers when you use these terms. The following are suggested phrasings. Can you explain the differences in meaning?

Population mean	Sample mean	Mean of the sampling distribution

Population standard deviation	Sample standard deviation	Standard deviation of the sampling distribution

[17] You can simulate Population 3. That is what I did in **Sampling distributions simulation (Normal).xls** and in **Sampling distributions simulation (U-shaped).xls**. But these Population 3's were incomplete – only some thousands of the means in Pop 3 were produced. I did these simulations for pedagogical purposes, for your benefit, to demonstrate that Properties 1,2,3 are true, rather because I need to do them. I do not need to.

- Population mean – μ – the mean of Population 1, the original population of interest
- Sample mean – $\bar{x}$ – the mean of a particular sample drawn from Population 1
- Mean of the sampling distribution – $\mu_{\bar{x}}$ – the mean of Population 3, which is all possible sample means (of a given size) that could be drawn from Population 1

- Population standard deviation – σ – the standard deviation of Population 1, the original population of interest
- Sample standard deviation – *s* – the standard deviation of the particular sample drawn from Population 1
- Standard deviation of the sampling distribution – $\sigma_{\bar{x}} = \sigma / \sqrt{n}$ – the standard deviation of Population 3, which is all possible sample means (of a given size) that could be drawn from Population 1

It is especially important to recognize the difference between the population standard deviation (of Pop 1) and the standard deviation of the sampling distribution (the standard deviation of Pop 3). In Example 7, the former is the variability of *individual* apartment rents. The latter is the variability of *groups* of apartments (in which each group is represented by the mean of a list of 60 apartments.) The latter also has another name that is commonly used: **standard error**. The standard error of an estimator (like the sample mean) is the standard deviation of its sampling distribution. In Example 7, the standard error of the sample mean is estimated to be $18.14.

SUMMARY

We sample an uncertainty distribution in order to learn about it.

The most important information to learn about an uncertainty distribution are its mean and standard deviation.

The best estimate of the population mean μ of a distribution is the mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ of a randomly drawn sample.

The best estimate of the population standard deviation σ of a distribution is the standard deviation $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$ of a randomly drawn sample.

The sample mean has an uncertainty distribution that consists of all possible values of the sample mean and their corresponding probabilities. This distribution is called the **sampling distribution of the sample mean**.

There are three important properties of the sampling distribution of the sample mean:

- **Property 1.** (**The Central Limit Theorem)** The distribution of sample means is approximately normal provided that the number of items *n* in the sample is sufficiently large (30 or more usually works).
- **Property 2.** The population mean of the distribution of sample means is μ (the same value as the population mean of the original distribution.)
- **Property 3.** The population standard deviation of the distribution of sample means is $\sigma / \sqrt{n}$ (in case of sampling without replacement, it is $\frac{\sigma}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$, where *N* is the population size.) If *n* is sufficiently large, σ may be replaced by its estimate, the standard deviation of the sample, when desired.

Sampling distributions are best understood through the **Three Populations Concept**: The original population is Population 1, which consists of all individual outcomes of interest. It has mean μ and standard deviation σ. Population 2 consists of all possible samples of size *n* from Population 1. Each sample in Pop 2 produces a sample mean – the totality of all such sample means constitutes Population 3. Pop 3 is the sampling distribution of the sample mean. Populations 2 and 3 are virtual populations – conceptual devices to understand sampling distributions. In the real world, no one actually lists them, except in toy problems for teaching and learning purposes.

The standard deviation of the sampling distribution is called the **standard error**. It is a fundamental measure of the accuracy of estimation. It is the amount, plus or minus, that we can expect the sample mean to deviate from the Population 1 mean that it is estimating.

The **Central Limit Theorem** goes further in assessing estimation accuracy by supplying probabilities. For example, it says that the probability that the sample mean will deviate by less than one standard error from the Population 1 mean that it is estimating is about 68%, by less than two standard errors is about 95%, by less than three standard errors is about 99.7%.

PROBLEMS

This Topic Note has introduced a number of new concepts that we will use repeatedly to support uncertainty calculations. If you would like to try your hand at some of the computational aspects, I have added some problems below. Solutions are provided at the end.

1. Suppose that the same two teams, A and B, play each other in the SuperBowl for three consecutive years. You are a fan of team A. Let X_1 be 1 if A wins the first year, 0 if B wins. Similarly for the outcomes X_2 and X_3 of the second and third years. The mean of X_1, X_2, X_3 is the mean number of times that your team wins. Suppose that the two teams are equal in skill, so that both teams are equally likely to win any given SuperBowl.

What is the outcome set for Population 1?

What is the distribution of Pop 1?

What is the mean of Pop 1?

What is the standard deviation of Pop 1?

List Population 2.

What is the outcome set for Population 3?

What is the sampling distribution of the sample mean?

What is the probability that your team will win all three years.

What is the mean of Pop 3?

What is the standard deviation of Pop 3?

2. Suppose that the population of salaries of UT MBA graduates has a mean of $100,000 and a standard deviation of $30,000. A random sample of 36 UT MBA graduates will be drawn and its sample mean will be calculated. (Assume that there are at least 720 UT MBA graduates in the population of interest.)

What is the mean of the population of sample means?

What is the standard deviation of the population of sample means?

What is the probability that the sample mean yet to be drawn will be between $95,000 and $105,000?

3. Suppose that a random sample of 36 UT MBA graduates has been drawn. Suppose that the sample mean salary is $105,000 and the sample standard deviation of salaries is $36,000. (Assume that there are at least 720 UT MBA graduates in the population of interest.)

Estimate the mean salary of all UT MBA graduates.

Estimate the magnitude of the difference between the truth and your estimate in the preceding question.

What is the probability that your estimate of mean salary will differ from the truth by less than $10,000?

4. Suppose that we give a 10%-off coupon to each of a random sample of 64 of our customers, which number at least 2,000. We observe the amount that each couponed customer spends in our store until coupon expiration. The sample mean expenditure is $50. The sample standard deviation of expenditure is $20.

Estimate the mean amount that ALL of our customers would spend under the same incentive if all had been given the 10%-off coupon.

Estimate the magnitude of the difference between the truth and your estimate in the preceding question.

How much confidence can you have that the truth differs from your estimate by no more than twice this magnitude?

SOLUTIONS to PROBLEMS

1. Suppose that the same two teams, A and B, play each other in the SuperBowl for three consecutive years. You are a fan of team A. Let X_1 be 1 if A wins the first year, 0 if B wins. Similarly for the outcomes X_2 and X_3 of the second and third years. The mean of X_1, X_2, X_3 is the mean number of times that your team wins. Suppose that the two teams are equal in skill, so that both teams are equally likely to win any given SuperBowl.

What is the outcome set for Population 1?
Ans. {0,1}

What is the distribution of Pop 1?

Value	0	1
Probability	½	½

What is the mean of Pop 1?
Ans. Mean = 0*1/2 + 1*1/2 = 1/2

What is the standard deviation of Pop 1?
Ans. Variance = (0-1/2)^2*1/2 + (1-1/2)^2*1/2 = 1/8 + 1/8 = ¼. Stan deviation = sqrt(1/4) = ½.

List Population 2.
Ans. (1,1,1,) (1,1,0) (1,0,1) (1,0,0) (0,1,1,) (0,1,0) (0,0,1) (0,0,0)

What is the outcome set for Population 3?
Ans. {1, 2/3, 2/3, 1/3, 2/3, 1/3, 1/3, 0} (in order corresponding to the order of Pop 2)

What is the sampling distribution of the sample mean?

sample mean	0	1/3	2/3	1
probability	1/8	3/8	3/8	1/8

What is the probability that your team will win all three years.
Ans. 1/8 (the probability that the sample mean = 1)

What is the mean of Pop 3?
Ans. 0*1/8 + 1/3*3/8 + 2/3*3/8 + 1*1/8 = ½

What is the standard deviation of Pop 3?

sample mean	0	1/3	2/3	1
Probability	1/8	3/8	3/8	1/8
sq mean deviation	(0-.5)^2 = 1/4	(1/3-.5)^2 = 1/36	(2/3-.5)^2 = 1/36	(1-.5)^2 = 1/4
pr*sq mean deviation	1/4*1/8 = 1/32	1/36*3/8 = 1/96	1/36*3/8 = 1/96	1/4*1/8 = 1/32

2. Suppose that the population of salaries of UT MBA graduates has a mean of $100,000 and a standard deviation of $30,000. A random sample of 36 UT MBA graduates will be drawn and its sample mean will be calculated. (Assume that there are at least 720 UT MBA graduates in the population of interest.)

What is the mean of the population of sample means?
Ans. $100,000 (The question asks for the mean of Pop 3, which is the same as the mean of Pop 1 by Property 2.)

What is the standard deviation of the population of sample means?
Ans. $30,000 ÷ sqrt(36) = $5,000 (By Property 3, the st dev of Pop 3 = st dev of Pop 1 ÷ sqrt(n). The assumption that there are at least 720 members of Pop 1 means that the fpc need not be employed.)

What is the probability that the sample mean yet to be drawn will be between $95,000 and $105,000?
Ans. Since the mean of Pop 3 = $100,000 and the st dev of Pop 3 = $5,000, then the range $95,000 to $105,000 covers one st dev below to one st dev above the mean of the (approximately) normal (by Property 1) distribution of Pop 3. Therefore, the answer is about 68%.

3. Suppose that a random sample of 36 UT MBA graduates has been drawn. Suppose that the sample mean salary is $105,000 and the sample standard deviation of salaries is $36,000. (Assume that there are at least 720 UT MBA graduates in the population of interest.)

Estimate the mean salary of all UT MBA graduates.
Ans. This is a RS (at least approximately: The sampling is evidently without replacement and the assumption that the sample size 36 is less than 5% of the Pop 1 size (which exceeds 720) means that the difference between with and without replacement may be ignored.) The best estimate of the mean of Pop 1 using a RS is the sample mean = $105,000.

Estimate the magnitude of the difference between the truth and your estimate in the preceding question.
Ans. The question calls for the standard error = st dev of Pop 3 = st dev of Pop 1 ÷ sqrt(n) = st dev of Pop 1 ÷ sqrt(36). Now st dev of Pop 1 is not known, but the best estimate of st dev of Pop 1 for a RS is the sample st dev = $36,000. So answer = (approx.) $36,000 ÷ 6 = $6,000.

What is the probability that your estimate of mean salary differs from the truth by less than $10,000?
Ans. This is the probability that the sample mean (your estimate) lies within $10,000 of the true but unknown mean of Pop 3. A ±$10,000 deviation is approx ±10,000 ÷ 6,000 = ±1.6667 standard deviations. Since Pop 3 is approx. normal by Property 1, then the answer is the probability between -1.6667 and +1.6667 in the Standard Normal distribution. Calculate this by use of the normal table on BlackBoard or by use of the Excel function NORMSDIST(1.6667) – NORMSDIST(-1.6667) = 0.9044.

4. Suppose that we give a 10%-off coupon to each of a random sample of 64 of our customers, which number at least 2,000. We observe the amount that each couponed customer spends in our store until coupon expiration. The sample mean expenditure is $50. The sample standard deviation of expenditure is $20.

Estimate the mean amount that ALL of our customers would spend under the same incentive if all had been given the 10%-off coupon.
Ans. Pop 1 is all of our customers. The best estimate of the mean of Pop 1 using a RS is the sample mean = $50.

Estimate the magnitude of the difference between the truth and your estimate in the preceding question.
Ans. This the standard error of your estimate = st dev of Pop 3 = st dev of Pop 1 ÷ sqrt(n) = st dev of Pop 1 ÷ sqrt(64). The st dev of Pop 1 is unknown but is best estimated by st dev of sample = $20. So answer = (approx.) $20 ÷ sqrt(64) = $2.50.

How much confidence can you have that the truth differs from your estimate by no more than twice this magnitude?
Ans. This is the probability that the sample mean (your estimate) lies within 2 standard errors = 2 st devs of Pop 3 of the true but unknown mean of Pop 3. Since Pop 3 is approx. normal by Property 1, then the answer is the probability between -2.0000 and +2.0000 in the Standard Normal distribution. Calculate this by use of the normal table on BlackBoard or by use of the Excel function NORMSDIST(2.0000) – NORMSDIST(-2.0000) = 0.9545.