EC3320

2016-2017

Michael Spagat

Lecture 2

An Economic Model of Counter Terrorism

The next few slides present the simple supply and demand model of Bruno Frey from his paper "Terrorism from the Rational Choice Point of View."

Frey uses the most basic supply and demand framework (next slide)

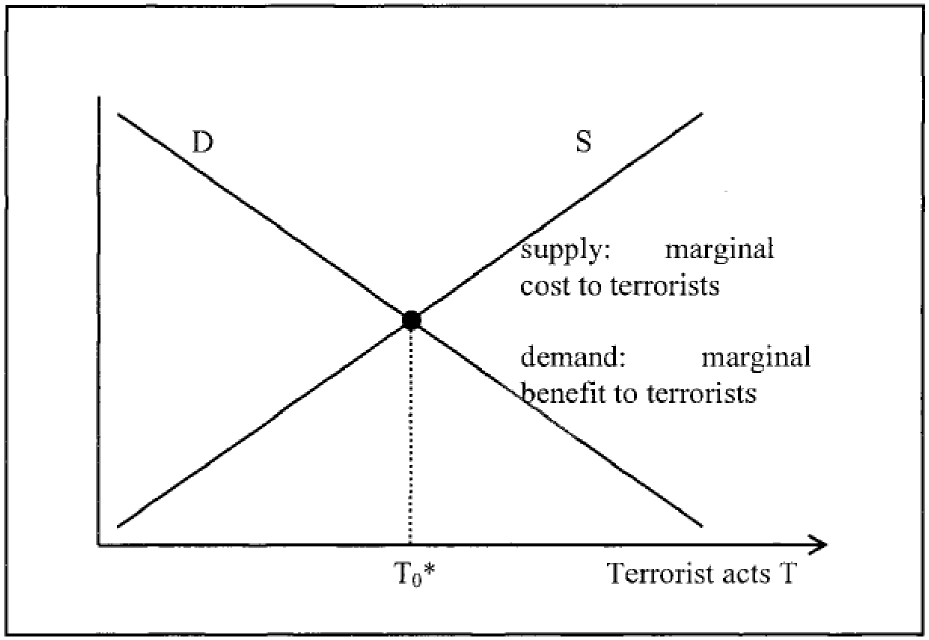


Figure 1: Supply of and Demand for Terrorism

The supply curve slopes upwards because the first targets chosen are the easiest ones to strike. Harder targets come later and are more costly to strike.

The demand curve slopes downwards on the theory that as a society is hit more and more by terrorism it becomes jaded and news of terrorist attacks has less and less impact. This assumption is not necessarily true - in some cases it could make sense for there to be an upward sloping demand curve.

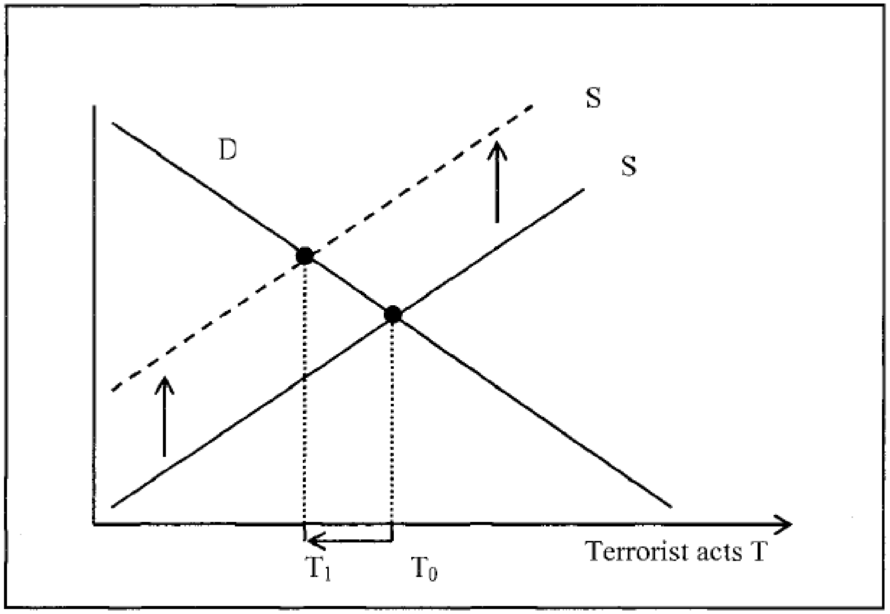


Figure 2: Increasing the Cost of Terrorism

We can think of policy measures aimed at deterring terrorists as shifting the supply curve up by raising the costs of terrorism.

One could imagine that terrorists are super motivated and, therefore, unaffected by costs but this would be a rather extreme and implausible assumption.

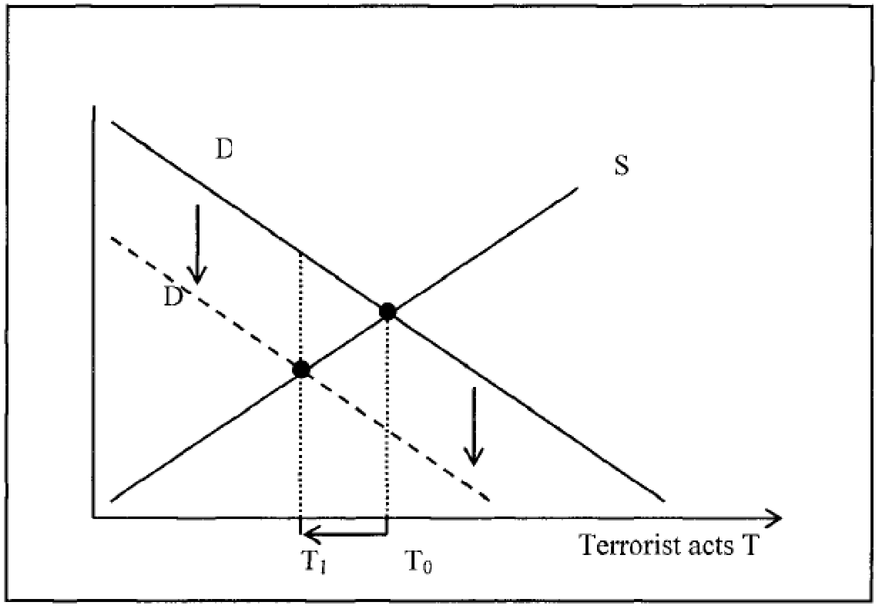


Figure 3: Reducing the Benefits of Terrorism

How can you reduce the benefits of terrorism, thereby shifting the demand curve down?

1. Deny media attention to terrorists. However, this is difficult to do in a country with an independent media.

2. Deny political advantages to terrorists. For example, do not release prisoners in response to terrorist threats.

3. Deny economic advantages to terrorists. For example, do not make unnecessary expenditures in response to terrorism such as cancelling plane flights.

Point 3 is the main insight of the paper in my view.

Spending money on deterring terrorism does shift the supply curve up, as depicted in figure 2 (slide 3). But, at the same time, expenditure on deterrence also increases the *benefits* the terrorists receive from attacking.

In other words, if terrorism leads to costly anti-terrorism expenditures that is a victory for the terrorists. So beefing up deterrence policy also shifts the demand curve up (opposite to the shift shown in figure 3 [slide 4]).

When both the demand and supply curve shift up the new intersection could be either to the left or to the right of the original intersections. If the demand curve shifts more than the supply curve then the intersection will be to the right. In this case counterterrorism policy is actually *counterproductive*, increasing the equilibrium level of terrorism.

Intuitively, this makes sense. *If terrorists know that their attacks will lead to massive expenditures to prevent further attacks then terrorists are incentivized to make attacks, leading countries to self-inflict wounds by overreacting to the threat of terrorism.*

We now turn to the Mueller and Stewart article – "Evaluating Counterterrorism Spending".

Mueller and Stewart argue that the costs of counterterrorism spending in the US far outweigh the benefits.

To make their case they need some way to place dollar values on the people who are killed in terrorist attacks.

Therefore, to understand the paper we need to start with a little excursion into the concepts that is used for this costing which is known as the *Value of a Statistical Life (VSL)*.

What does this strange phrase (value of a statistical life) mean?

This paper by Orley Ashenfelter is not assigned for the course but is recommended to anyone who gets fascinated by this VSL concept. I will lean on it heavily in my short discussion.

The VSL concept is based on how people answer a question like the following – *how much would you be willing to pay for a specific decrease in the probability of dying?*

For example, would you pay £30 for a bicycle helmet that decreases the probability that you will be killed in a bike crash by 0.000015?

If your answer is “yes” then we might say that your VSL is at least £2,000,000, i.e., £30/0.000015.

This calculation gives us a lower bound, not the exact VSL. If you had refused to purchase a helmet for this price then £2,000,000 would be an upper bound.

In principle, you can pin down the exact VSL for a person by asking enough questions. Would you pay £30? If "yes" would you pay £40? If "no" would you pay £35? Etc.

Such calculations tend to place VSL's around $6 million to $7 million in rich countries.

The VSL gives us a way of costing lives lost in terrorist attacks and progressing to a cost-benefit analysis. However, before doing this we should at least be aware of some of the *weaknesses of the concept*.

1. There is an implicit assumption that if I would be willing to pay exactly £30 to decrease my probability of getting killed by 0.000015 then I would pay another £30 for a further reduction of 0.000015, all the way down to when my chance of getting killed is 0 (A point which does not really exist but leave this detail aside.)

At the same time I would accept an increase by 0.000015 in the probability of dying if you pay me £30 and would keep doing this all the way up to 1.0. This brings us to an old joke by the comedian Jack Benny (quoted in the Ashenfelter paper):

Robber: “This is a stick-up. Your money or your life.”

Pause

Robber: “Look, bud. I said your money or your life.”

Jack Benny: “I’m thinking it over”

My point here is just that it is hazardous to extrapolate from trade-offs people make between small changes in risks and money to trade-offs over big changes in the same things.

2. Many economists would question such an approach because economists prefer data on what people really buy in reality rather than what they say they would buy under hypothetical circumstances. In other words, actions may speak louder than words.

Later in the course we will see some situations when we can calculate VSL's based on real market behaviour but we will not go off on this tangent now.

The take home point from slides 7-11 is that Mueller and Stewart use the VSL concept to cost the lives lost in terrorist attacks. This method is far from perfect but it at least can get us started on a cost-benefit calculation.

Here is their central table of Mueller and Stewart:

Table 1

How Many Terrorist Attacks Would Need to Occur Each Year in the Absence of All Counterterrorism Measures in Order to Begin to Justify a Counterterrorism Expenditure of $75 Billion

	Type of terrorist attack					
	Ft. Hood Shooting	*Boston, Times Square bombing*	*London bombing*	*9/11*	*Nuclear bomb, port*	*Nuclear bomb, Grand Central Station*
Losses per incident	*$100 million*	*$500 million*	*$5 billion*	*$200 billion*	*$1 trillion*	*$5 trillion*
Level of risk reduction assumed						
10 percent	7,500	1500	150	4	.75	.15
25 percent	3,000	600	60	2	.30	.06
50 percent	1,500	300	30	.75	.15	.03
75 percent	1,000	200	20	.50	.10	.02
90 percent	833	167	17	.42	.08	.02
100 percent	750	150	15	.38	.08	.02

Notes: If the $75 billion expenditure is expected to reduce the risk (the likelihood of, and/or the damage caused by, a successful terrorist attack) by 50 percent, those expenditures would need to deter, disrupt, or protect against at least half of the attacks in each entry in the 50 percent line. For the boxed entries, that would be 150 Boston-type attacks per year, 15 London-type attacks each year, or one 9/11-type attack about every three years.

There are two basic types of facts underlying this table.

1. Mueller and Stewart estimate that the US is spending about $75 billion per year *more* on terrorism *after* 9/11 happened than it was spending before 9/11. The real number is almost certainly bigger than this and, possibly, a lot bigger. But we go with $75 billion.

2. There are estimates of the costs of various terrorist attacks, both real and hypothetical built into the table. Substantial components of these estimates are VSL-based human costs.

How do we interpret the numbers in the table?

Let's just work through a simple example and then all should become clear. Take the "30" in the column on "London Bombing".

Notice first that Mueller and Stewart cost this bombing at $5 billion. (Obviously, this bombing was in London but it is used here as representing a type of attack with a certain cost that could happen in the US.)

$5 billion x 30 = $150 billion which is twice as large as the estimated $75 billion annual increase in counterterrorism expenditure in the US post 9/11. Thus, if this expenditures screens out 50% of attempted attacks the size of the London bombing then it would just pay for itself. In other words, if there are 60 attempts on the scale of the London bombing and only 30 succeed then the $75 billion spent is validated.

Of course, we are not living in such a world since the above calculation presumes that there were 30 successful London-scale attacks last year.

If the $75 million screens out all such attacks and there were 15 such attempts then, again, the investment would pay for itself.

The main point of Mueller and Stewart is that it is hard to believe that we are really screening out so many attacks.

The incomplete public record that exists suggests that we are really far from justifying such a large expenditure.

How do governments get their populations to agree to large expenditures on security such as the $75 billion per year of increased post 9/11 counterterrorism expenditures?

Flores-Macías and Kreps (short version here) give an interesting insight into this issue, specifically, that *taxpayers are more inclined to support a war if it is financed through borrowing rather than through taxes.*

Their approach is simple.

1. They describe a war scenario to one group and ask the members if they would support spending money on an important war to be financed through existing funds plus borrowing.

2. They ask the same question to another group except that now the finance will be through existing funds, borrowing *and a war tax*.

Flores-Macías and Kreps do these experiments in the US and the UK with similar results.

They find that support for a hypothetical war drops by about 10 percentage, i.e., from around 45% to around 35% when taxes are included as part of the funding package.

Note that this finding runs counter to the idea of Ricardian Equivalence, according to which taxpayers should not believe they have avoided paying for a war if the government finances it initially through borrowing money. This is because someday the government will have to pay down the debt which will require raising taxes above what they would have been without the war.

In practice, violations of Ricardian Equivalence are common. For example, some taxpayers may expect that taxes will only be raised after they die and are not especially concerned about the burden this will place on future generations.

Dispersion

Let’s bring back our old friends, the thirteen numbers:

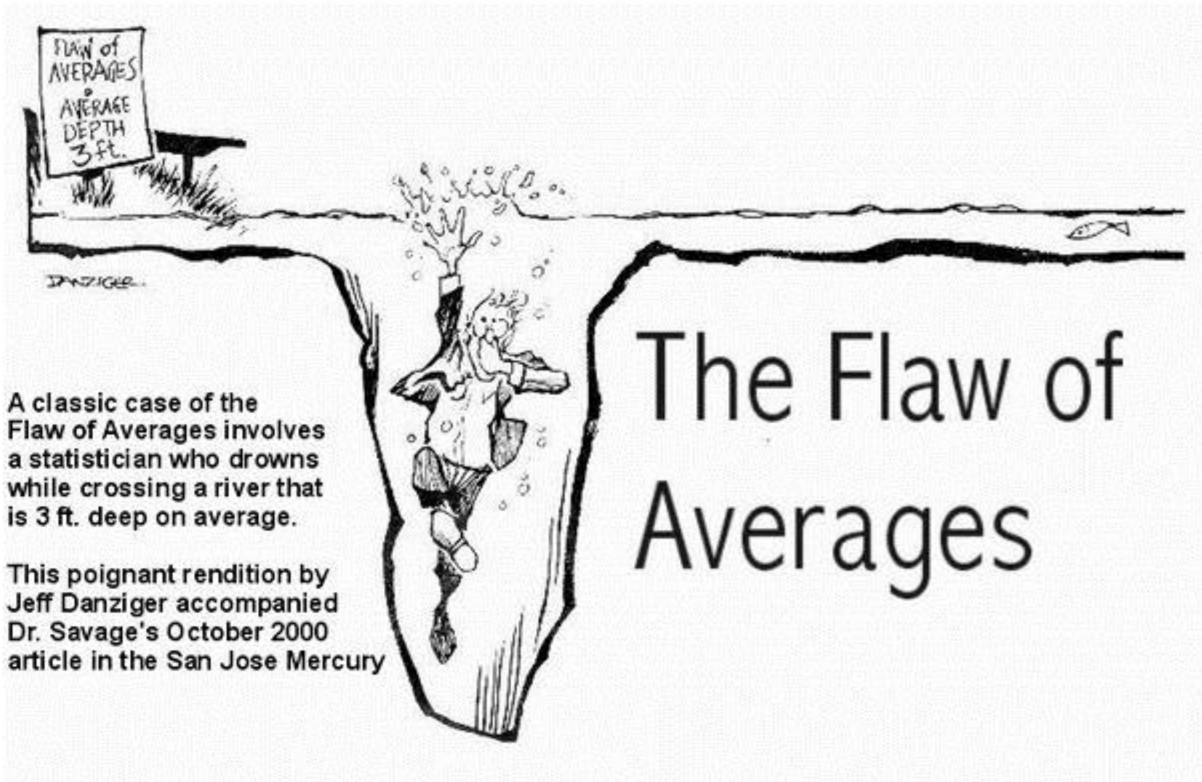
4.6 3.2 0.8 2.1 12 0.6 3.1 0.3 0.6 0.7 3.6 1.8 5.1

We would like to briefly summarize these numbers.

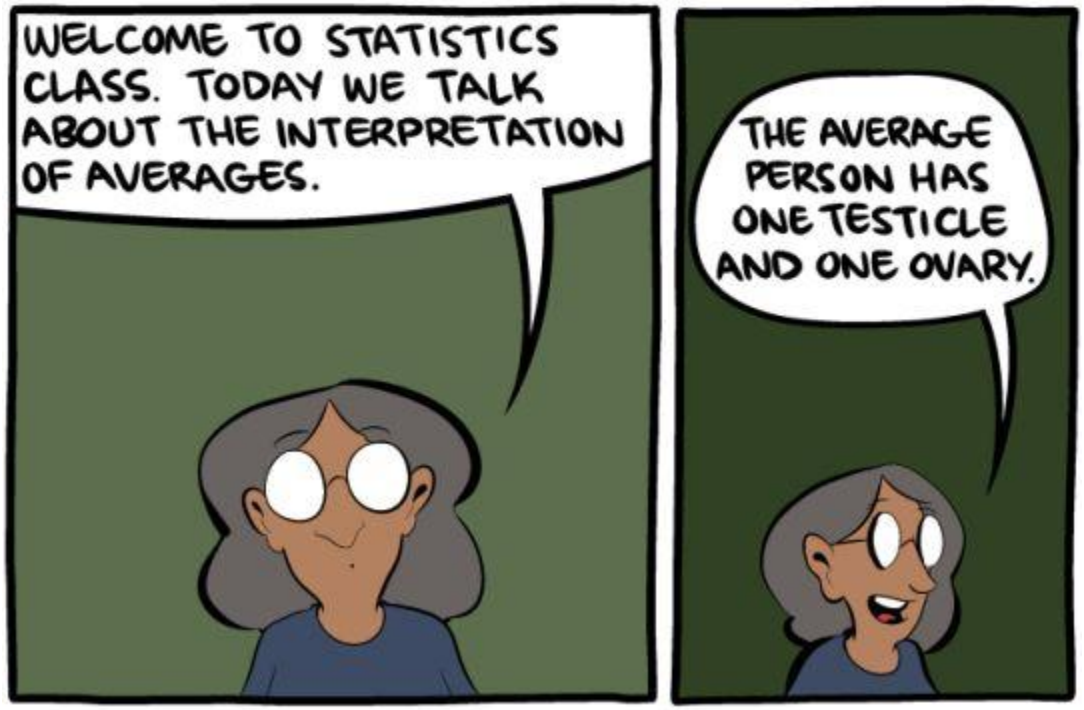
One way to do this is through one or several numbers capturing their “central tendency”.

That is, we can give their mean and/or their median and possibly also their mode.

However, if we stop there then we say nothing about how the numbers distribute themselves around the central tendency, that is, how they are “dispersed”.



Strangely, this cartoon fits perfectly with our 13 numbers since the mean is 3 but the biggest number is 12.



Again, dispersion about the mean is important.

The *variance* of this set of numbers is defined as the mean (average?) squared deviation from the mean. (Hmmmm…..what does that mean?)

$$\text{Variance} = \frac{(4.6-3.0)^2+(3.2-3.0)^2+(0.8-3.0)^2+\ldots+(5.1-3.0)^2}{13}=9.2$$

Note, again, the importance of the outlier, 12. If we throw out the 12 then the variance becomes *2.6, really different from 9.2*. Variances are more sensitive to outliers than means are because numbers squared increase much faster than the numbers themselves.

The *standard deviation* of this set of 13 numbers is just the square root of the variance. Rather coincidentally, this turns out to be right around 3.0 (so, by chance, the mean turns out to be equal to the standard deviation).

Without the outlier, 12, the standard deviation becomes 1.6. The difference between 3.0 and 1.6 is not nearly as large as the difference between 9.2 and 2.6 – taking square roots of the two numbers (with and without the outlier) squeezes them together.

We can use the mean and the standard deviation to make *predictions* about what will happen if we draw one of these 13 numbers at random. The outcome will probably be the mean plus or minus the standard deviation, i.e., 3.0 ± 3.0 or, in other words, a number between 0 and 6.0. In fact, that prediction works 12 out of 13 times – it fails only if we draw the outlier, 12.0.

We sure are great at predicting numbers!!!! Well, not really. That was quite a crude prediction. We allowed ourselves an error margin of plus or minus 100%.

This crudeness comes largely from the outlier.

Throwing out the outlier our prediction becomes 2.2 ± 1.6 or, in other words, 0.6 to 3.8. This is still a rather crude prediction, but it is much sharper than plus or minus 100%.

Our success rate with this prediction is OK, 9 out of 12. The ones we miss are 0.3, 4.6 and 5.1.

So far our discussion of statistics has only been about summarizing bunches of numbers. We now turn to *estimating* facts about bunches of numbers when we do not know in advance what all the numbers of interest are.

We can think of estimation techniques as machines that give us approximate answers to certain types of questions we put to them.

WARNING – We often analyse these estimation techniques by seeing how well they answer questions for which we already know the answers. This is a sensible thing to do but it can be confusing because you can easily get stuck wondering why we are estimating something we already know – we do this because we want to know how our technique will perform when we use it to answer questions for which we do not already know the answers. For example, suppose I design a new type of scale. It is sensible to start by using it to weigh some things with known weights to verify that it does well on these before moving on to weighing things with unknown weights.

Suppose I tell you that I have thirteen pieces of paper, each with a number written it. I want you to estimate the mean of these 13 numbers. You are allowed to pick 4 of the thirteen numbers at random to help you form your estimate.

What would you do?

To be clear, your answer is to be a single number. Later we will talk about how you might put error bands around your estimate. But, for now, you should give what we refer to is a *point estimate*.

Hopefully, your answer was that your estimate for the mean of all 13 numbers would be the mean of the 4 numbers that you are allowed to observe.

So your technique for estimating the mean of a bunch of numbers is to take a random sample of the numbers and calculate the mean of the sample of numbers. That *sample mean* is your estimate for the *population mean*.

To be concrete, let's apply this technique to our favourite list of 13 numbers.

Suppose our random sample is the first 4 numbers listed on slide 18 - 4.6 3.2 0.8 2.1. The mean of these four numbers is about 2.7. This is not too bad – the true mean is 3.0 and we estimate it to be 2.7.

Suppose we had a different random sample - 3.2 12 0.3 3.6. Now our estimate for the mean is 4.8. That darn outlier, 12, has messed us up again. We would have been off by more except that we also drew the smallest number, 0.3.

IMPORTANT POINT – Each random sample that we can draw gives us a different estimate of the mean of the 13 numbers. *Our estimate is random* – each random sample produces a different estimate.

In fact, we can write down every single sample with their corresponding estimates – there are 715 of them. Taking this one step further we can calculate the mean of these 715 estimates. What do you think this mean of means will be?

Answer – 3.0

This means that our estimation method is correct on average. We call such a method *unbiased*.

Notice that it is the *method* that is unbiased. In this case the method is picking a random sample and calculating the mean within the sample. This method produces correct answers *on average* but any particular sample mean is likely to be either higher or lower than the true mean.

A method is *biased* if it does not give a correct answer on average.

For example, suppose my method is to pick a random sample and make my estimate equal to the largest number in the sample. This method has a positive bias, i.e, the estimates tend to be too high. Making the estimates equal to the lowest number in the sample will have a negative bias.

Often the things we are measuring are qualitative but we can still convert them into quantities and use our technique for estimating means. For example, suppose we have a list of N people who have been killed in an armed conflict and we want to know what percentage of them were killed by small-arms gunfire rather than by some other method. The information here is qualitative, taking the form of "yes" (killed by small-arms gunfire) or "no" (killed by some other method). But it is easy to code the information quantitatively – "1" can mean "yes" and "0" can mean "no". The mean of these 1's and 0's multiplied by 100 will be, precisely, the percentage killed by small arms gunfire.

Similarly, we can estimate the percentage of females in this room by taking a sample of students, assign the number "1" to every female and "0" to every male in the sample, add up the numbers, divide by the sample size and multiply by 100 (to convert this into a percentage). This device is used a lot in survey research.

A more difficult task is to put error bands around an estimate. Recall our 13 numbers:

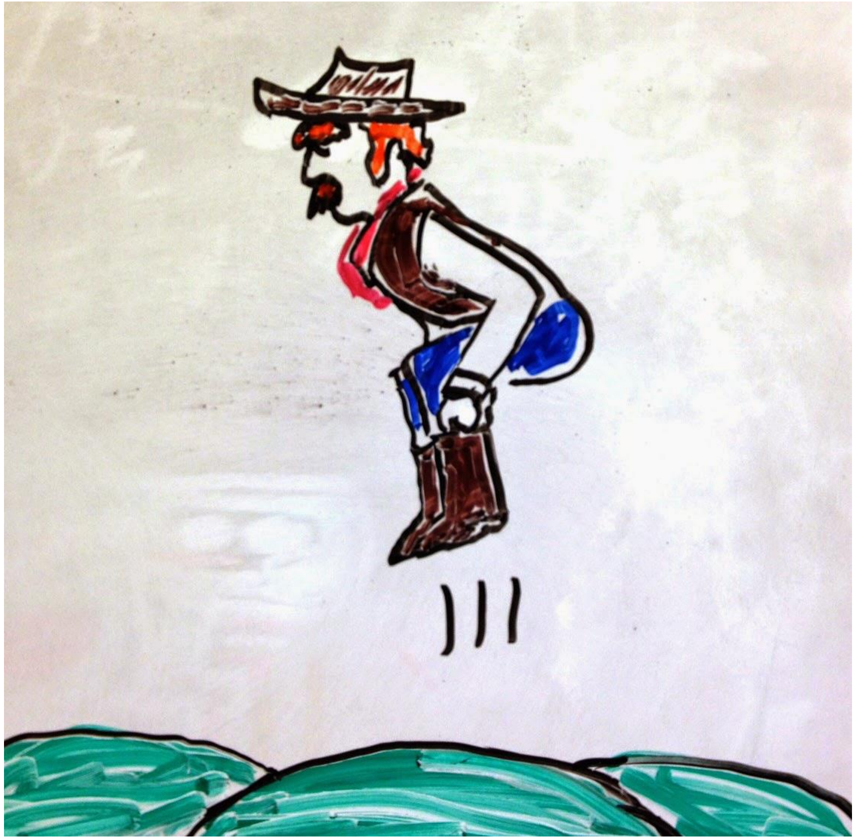
4.6 3.2 0.8 2.1 12 0.6 3.1 0.3 0.6 0.7 3.6 1.8 5.1

Once again, suppose we take a sample of 4 of these numbers and use the mean of these 4 numbers as our estimate of the mean of all 13 numbers. Suppose our sample is:

3.2 0.7 0.3 1.8

Our estimate for the mean is 1.5. This is not a spectacular estimate since the true mean is 3.0. But we do not know this. For all we know, the true mean might really be 1.5. On the other hand, we should be generally aware that our estimate could be fairly far from the truth.

The question is – how can we quantify how wrong we are likely to be? This situation is inherently frustrating – we can only use the data at hand to try to quantify our uncertainty about our estimate of 1.5. We only see the 4 numbers we drew and do not know that we might have picked a 4.6 or even the outlier, 12. Some of you will have heard the expression “pulling yourself up by your bootstraps.”



This is the type of situation we are in – we need to use what has actually happened, our sample of 4 numbers, to get a handle on what might have happened. There can be no way of doing this without sometimes making big mistakes.

There is no single correct way to proceed for calculating error bands around our estimates but there are a couple of standard procedures.

I will show you how these work in a general context, that is, not just for our 13 numbers.

Here is the standard procedure.

1. First get a fix on how dispersed the full set of numbers are. You do this by calculating how dispersed your sample of n is….but with just a tiny adjustment. Specifically, you calculate the variance of the numbers, except you divide the sum of squared deviations from the mean by n-1 rather than dividing by n. This is called a *sample variance*. (I actually hesitate to even mention this point because in this course we will generally work with samples of 1,000 or more so it does not matter whether you divide by 1,000 or by 999. So let's not focus on n versus n-1. But you can go here if you want to learn more about this.). Take the square root and we have what we call the *sample standard deviation*.

2. Divide the sample standard deviation by the square root of the sample size, n. Call what you get the *standard error*. (Note that I have just supplied the piece of information you need to fully understand the numbers in the table on slide 13 of Lecture 1.)

3. Our 95% confidence interval is then our sample mean plus or minus 2 standard errors.

Let’s master these formulas by looking at a couple of samples of 4 from our list of 13 numbers.

Before we do this I just want to make two caveats.

1. As noted above, in this course we are interested in cases where the sample size is very small compared to the total population, e.g., 2,000 compared to 4 million. However, for the next few slides we will look at cases where the sample size is not so small compared to the total population, 4 compared to 13. This case is useful for illustrating the formulas but to do this case right we have to add a slight tweak to the formulas known as the *finite sample correction*. I say this only because you might open up a textbook and find a formula that looks a little different from what we use below.

2. The final step where you form an interval extending plus or minus 2 standard errors around the sample mean is only valid when you have a sufficiently large sample size. A sample size of 4 is will not be large enough.

On the other hand, working with this tiny sample size is very convenient for mastering the mechanics of how the formulas work so we ignore these caveats in the next four slides.

Let's take the sample 3.2 0.7 0.3 1.8.

The sample mean is 1.5. This is our estimate of the mean of the 13 numbers. It is not a great estimate, since the real mean is 3.0, but we do not know this.

The sample standard deviation is:

$$\left(\frac{(3.2-1.5)^2+(0.7-1.5)^2+(0.3-1.5)^2+(1.8-1.5)^2}{3}\right)^{\frac{1}{2}}=1.3$$

We divide by 2 (the square root of 4) to get the standard error which turns out to be 0.65.

The 95% confidence interval is the sample mean plus or minus two standard errors. This is the interval [0.2, 2.8]. This interval does not quite make it up to 3.0 so even with a confidence interval attached to the estimate this is not a great sample.

Suppose we draw a different sample: 3.2   2.1   3.1   5.1.

Now the sample mean is 3.4, the sample standard deviation is 1.3, the standard error is 0.6. The 95% confidence interval is 2.1 to 4.7. Now the true mean is well within the confidence interval.

Notice that both the sample means and the confidence intervals are random. Every time we draw another sample we get a new estimated mean and a new confidence interval.

Now imagine we write a computer programme that draws 100 random samples, each of size 4. How many of them do you think would include 3.0?

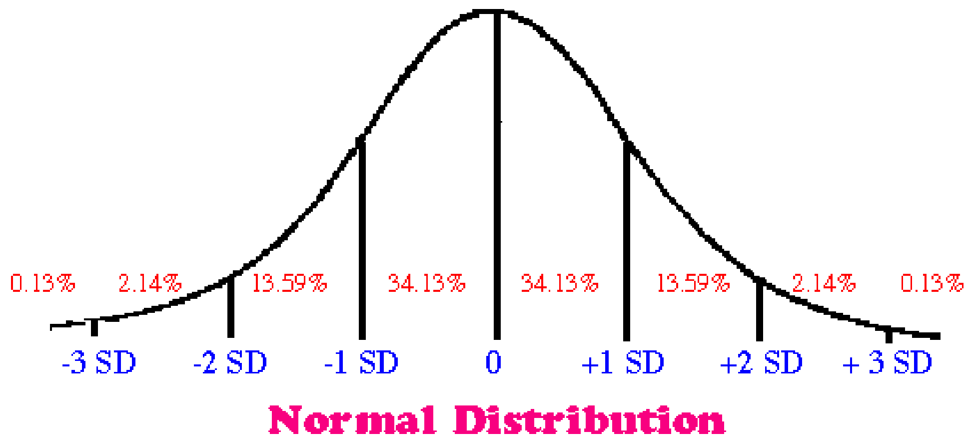
Answer – We'd expect that around 95 out of the 100 would include 3.0. That is what the term *95% confidence interval* means.

Let's now try to get some intuitive understanding of this method.

Step 1 (Sample Standard Deviation). More dispersion in your sample translates into a wider confidence interval.

Step 2 (Standard Error). This is the step where you divide the sample standard deviation by the square root of your sample size. This step captures the idea that the bigger your sample the closer your sample mean is likely to be to the true sample mean.

Step 3 (plus or minus 2 standard errors). This comes from the deep result that the distribution of means converges to a normal distribution (bell curve) as sample size grows.



Finally, I would like to introduce and important point to which we will return. The above method for calculating confidence intervals recognizes only one source of uncertainty – the fact that you base your estimate on a sample rather than the whole population. In most situations there are further sources of uncertainty so these conventional confidence intervals understate the true uncertainty surrounding an estimate.