LAB: Estimation of the Amount of Water Stored on Mount David[1]

Introduction

Many years ago, I went for a hike in the Sierra Nevada of California in early summer. I came over a ridge into a deep valley filled with ponderosa pine and other California mountain conifers (including incense cedar, the tree from which commercial pencils have traditionally been made). The forests of the Sierras are relatively sparse, and this was no exception. As I came over the ridge, I spotted a tall red and white tower, 30 feet or more in height, with, of all things, a shovel tied to its top. It was one of the towers used by the state of California to estimate the amount of water stored in the snow pack of the Sierras. The snow pack is the primary direct or indirect source of drinking and irrigation water for the state.

These towers are scattered at strategic locations throughout the Sierras. The painted red and white bands on the towers allow observers from the air or from nearby ridges rapidly to estimate the depth of snow on the mountains. As the shovel tied to the top of my tower attested, just a few months before my hike, the snow in this particular valley was so deep, the entire tower was buried, and someone had found a convenient way to make it a couple of meters taller.

So, how do we determine exactly how much water is stored in the snow pack?

In the Sierras, the professionals collect information on the depth of snow at certain locations. From calculations and prior experience the hydrologists can estimate with fair precision the amount of water that will turn up at downstream reservoirs. The problem of estimating the amount of water in the snow pack, therefore has impeccable, practical roots.

Of course, the amount of water stored on the snow pack on Mount David has no such practical importance. While melt water from the snow on Mount David may briefly overwhelm the combined sewers of Lewiston, thus flushing raw sewage into the Androscoggin River, for the most part, we are interested in estimating the total water stored in the snow pack there for purely pedagogical reasons.

For our purposes, this question is useful largely because it provides us with a setting in which students need to confront the difficulties of measurement and estimation in the face of the nearly omnipresent reality of spatial and temporal variability in the real world.

How Much Water is Stored on Mount David?

This is a superficially simple question, one that has, at least in principle, a simple answer. The answer could come in the form of a single number. The snow pack on Mount David contains just so many cubic meters of water, no more and no less. While the number would be different if you measured it some other time, at any given time there is a certain amount of water stored in the snow. We could n in principal vacuum all the snow on Mount David,

[1] Mount David is a small hill on the Bates College campus. This lab could be done on any small area near your campus or modified slightly could be done with leaf litter rather than snow.

weigh it and determine the amount of water stored there. Of course in practice, we do no such thing. It's much too impractical.

The Assignment

Your job is to estimate the amount of water stored in the snow pack of Mount David. There are many ways you could do that. Your job during the first week of this lab is to decide on the method you will use to come up with an estimate.

To make this problem both more interesting and more representative of real environmental science practice, we want you to confront the real trade-offs between cost of developing such an estimate and the accuracy that you can achieve. In general, one could get a more accurate estimate of the amount of snow on Mount David either by taking more, or more accurate, measurements. However each measurement costs money, and typically, more accurate measurements are more expensive than less expensive measurements.

To simulate that situation, we want you prepare a "bid" to submit to us summarizing the method of measurement you will use, what it will cost, and what you think the accuracy of the resulting estimates will be.

The "costs" you will use to establish your bid will be as follows:

Item	***Cost***
Field Work	$50 per hour per person
Meter Stick Measurements of DEPTH	$1.00 each
Pesola Scale Measurements	$10.00 each
Mettler Balance Measurements	$50.00each

Other measurements (such as measurements of volume, area, etc.) are free, except for the time involved in making them. You need not charge for your time taking measurements, estimating areas, making calculations, and so on, if that work occurs in the laboratory. You only need to charge for time spent in the field.

Your "bid"should consist of

a) A short (less than 1 page) narrative description of the measurements you plan to take
b) An estimate of the "cost" of taking those measurements, and
c) An estimate of the precision with which you will be able to estimate the amount of water stored in the snow pack on Mount David.

Your "bid" is due at the start of lab next week (January 23 or 24).

During the second week of the lab, you will have the opportunity to head out to Mount David (wear suitable warm clothes!) and collect your data. Be certain to record how many measurements you actually take and how long you actually are in the field, so that you can "charge" for time and measurements actually taken. In practice, it is unlikely that the estimated "cost" from your bid will exactly match the time and measurements you actually use.

From your data, you should then calculate your estimate of the amount of water stored on Mount David, as well as the precision with which you can estimate it, expressed in terms of the "standard error" of the estimate (see below). Finally, you should write up a brief (2 page) report summarizing the methods you used, presenting your results, and submitting a final "bill". This report will be due at the beginning of lab in two weeks (January 30 or 31).

Suggested Approach

Perhaps the easiest way to estimate the amount of water stored in the snow pack is to break the problem down into pieces. First, estimate the density of the snow pack, in grams (or kg, or cubic meters, or whatever units you find convenient) of water per square meter. Second, estimate the area over which you want to estimate the amount of stored water, and multiply.

This general approach can be applied either to the entire area of Mount David, or to several sub-areas. Conditions in different parts of Mount David are likely to be quite different. There is likely to be rather less snow on the exposed top of the rock than on the protected flanks of the hill. There may be less snow under the white pines (which trap significant snow before it reaches the ground) than under the hardwoods (or maybe there is more snow there because the trees shade the snow and minimize melting). More snow will accumulate in hollows where it is blown by the wind than on hillocks from which it is scoured. Therefore you may get a more accurate estimate of the total amount of water stored on Mount David if you subdivide Mount David into several sub-areas, and estimate snow pack characteristics for each area separately. The total amount of water stored on Mount David will then just be the sum of the amount of water stored in each sub-area. If you are clever about dividing Mount David into sub-areas, you can get a more accurate estimate of the total snow cover. If you do it badly, however, your estimates will be less accurate.

How do you allocate measurement effort to get the best possible estimate of the snow pack?

To answer this question, we need to look at some underlying statistical thinking. We start by breaking down each measurement of snow density into several pieces that are of interest to us in thinking through the problem.

A model of measurement of the density of the snow pack per unit area

Model 1:

Measurement of "density" =
A. Average density of snow on Mount David +
B. Variation that's not of interest to us +
C. Measurement error

Estimating the average snow pack density

What this model suggests, is that any single measurement we make is a rough approximation of the value of the "true" average density of snow on Mount David. Unfortunately, each measurement includes both measurement error (which could be either positive or negative), and some variation that just happens to be of no interest to us (which could also be either positive or negative). Therefore, if all we have is one measurement, we are unlikely to have gotten a measurement that (by chance) just happens to be close to the "true" average we are looking for. The measurement may be a bit larger then the "true" average, or a bit smaller. We don't know, and without taking additional measurements, there is no way for us to know.

Luckily, if we take SEVERAL measurements of snow density (and those measurements are not biased in any way) we will usually get a few measurements that are a little too large and a few measurements that are a little too small. If we average those measurements, the small measurements should cancel out the large measurements to some extent. The average of our measurements, therefore, has a very good chance of being a better estimate of the "true" average than the single measurements. In fact, the more measurements we take, the more likely it is that large measurements will be cancelled out by small measurements, and the closer the average of those measurements is likely to come to the "true" average.

Based on this simple logic, we can conclude:

- An average of several measurements of snow density provides us with an ESTIMATE of the "true" average. There is no way, short of vacuuming up all the snow on Mount David and weighing it, of finding out exactly what the "true" average is.
- An average of several measurements gives us a better estimate of the "true" average than a single measurement would.
- The more measurements we take, the better our estimate of the "true" average becomes.

How close is our estimate of the "TRUE" average?

Under remarkably robust assumptions, we can calculate how accurate an average of several measurements is likely to be. I say how accurate our average is "likely" to be, because no matter how many measurements we take, there is always a chance that we were unlucky, and measured lots of small measurements (or lots of large measurements), and our average is wildly inaccurate.

The calculations are based on the idea that if individual measurements vary a lot, then the average of a few of those measurements will also vary a lot. We can get an accurate estimate of the "true" average of our measurements either if our original measurements are all similar, or if we take many measurements.

Statisticians generally lump two of the terms in our model of measurement together. They take "variation that is not of interest to us" and "measurement error", and combine them into a single term, called "experimental error". Experimental error is all the variability among measurements that we are not paying direct attention to. It is the "noise" in our measurements, the portion of the measurement that we either chose not to explain, or that we are unable to explain. "Experimental error" is what makes some observations larger than average, and others smaller. Other things being equal, a smaller "experimental error" would allow us a more accurate estimation of the "true" average snow pack on Mount David, and thus to get a better estimate of the total water stored on Mount David.

The Standard Deviation

The variability of a group of measurements is frequently measured in terms of a statistic calculated from the values of the measurements called the standard deviation. The standard deviation can be thought of as a measure of how far away from the "true" average a typical measurement will be. Some measurements will be closer to the "true" average and some will be farther away.

For our purposes, we will use the sample standard deviation (Sometimes written as SD^+. On calculators it is often shown as S_{n-1}, or _$_{n-1}$).

Given a list of n measuments , $X_1, X_2, X_3 \cdots X_{n-1}, X_n$,

the sample standard deviation is defined as :

$$SD^+ = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}}$$

X_i is the ith measurement you took

$\bar{X}$ is the average of all the measurements

This is an imposing-looking formula, but the ideas underlying it are not hard to understand. The core of the formula (inside the parentheses) is the difference between a measurement and the average of all the measurements. It is a pretty good estimate of how far away from the "true" average any particular measurement is. It is called, in statistical parlance, the "deviation" of the measurement from the average. We want to describe how far away from the "true" average a "typical" measurement is, so what we want is some sort of an "average" of these deviations.

Unfortunately, taking a simple average does not work. The deviations of the measurements that are a bit smaller than average are negative, while the deviations of the numbers that are a bit larger than average are positive. You can prove pretty easily that the negative deviations will exactly cancel out the positive deviations, and their average will always be zero. Not much help there.

One way to ensure that the negative values don't just cancel out the positive values is to square the deviations before you take their average. That turns out to work especially nicely. The resulting number (the "mean square deviation", or variance) plays a central part in modern statistics. It corresponds to the value in the formula inside the square root sign.

The problem is, the variance is measured in weird, generally incomprehensible squared units. If our original measurements of the snow pack density were in kg(water)/m^3, the variance is in units of kg^2/m^6. Ugh! The solution? Take the square root. That returns us to units that make some sense. The result is sometimes called the "root mean squared deviation", but more generally, it is called the standard deviation.

If you have been really paying attention, you may be wondering why the sum of the squared deviations was divided by n-1 and not by n. After all, if we take the average of a list of n numbers (in this case the deviations), we add them up and divide by n, not by n-1. The answer to that question gets us deeper into statistical theory than we need to go. Basically, dividing by n-1 corrects for a bias that creeps into our value of the standard deviation because we don't know the "true" average.

The standard deviation is probably best calculated using EXCEL's STDEV() function, or a scientific calculator. For all but the shortest lists of measurements, it is a bit of a hassle to calculate by hand, although you can do it if you are careful.

The Standard Error

O.K. Now we have a measure of how variable our measurements were. Using that information it is easy to calculate how close the average of our measurements will typically be to the "true" (but still unknown, indeed unknowable) average.

The statistic we use to describe the accuracy of our average is called the "standard error" of the average. As its name suggests, it gives an indication of how far away from the "true" average the average of our measurements is likely to be. The logic here can be a bit tricky to wrap your mind around. Remember that the average of our measurements is still nothing but an estimate of the "true" average. I hope that by now you are beginning to realize that no matter what we do, we will never know what the "true" average is. By taking a gazillion measurements, we can get arbitrarily close to it, but there will always be a certain amount of uncertainty in the result. If we repeated our entire study and collected all the measurements again, the average of all of our measurements would be slightly different.

As a matter of fact, one interpretation of the standard error is that it is a prediction of how much variability we would observe if we did go to all the work of repeating our study dozens of times.

Remember, the standard deviation describes variation among our original measurements, While the standard error describes the expected variation among averages, if we repeated our study over and over.

The standard error of our average is just the standard deviation of our measurements, divided by the square root of the number of measurements combined to make up the average.

$$SE = \frac{SD^{+}}{\sqrt{n}}$$

The most obvious implication of this formula is that the more measurements you take, the closer the average of your measurements is likely to be to the "true" average. However, it takes a lot of measurements to dramatically improve the accuracy of your study. If you quadruple the number of measurements you take, the accuracy of your study only doubles. And if you want to improve your estimate by a factor of four, you'll need 16 times as many measurements.

Remember the original problem?

We are almost at the point where we can think about how to allocate our resources to get the best possible estimate of the snow pack on Mount David. We have made significant progress. We have laid the groundwork to be able to think about trade-offs between the number of measurements we take and their relative accuracy. What remains, is to develop a way to think about the relative advantages and disadvantages of treating Mount David as a single area, or of dividing it into several smaller areas with more (we hope) more uniform snow cover.

Subdividing Mount David

To a first approximation (I'm skipping over several statistical niceties here), the smaller the experimental error, the more accurately we can estimate the average snow pack in any given area. If we can subdivide Mount David into several smaller areas that have more uniform snow packs than Mount David as a whole, we might be able to estimate the snow pack in each of those areas considerably more accurately than we can estimate the snow pack on the entire hill. The result, once we add together the (more accurate) estimated amount of snow in each area, would be a more accurate estimate of the snow on the hill as a whole.

To justify that idea, we need to look at a slight modification of our original model of a measurement of the snow pack. If we believe that certain areas on the hill have thicker snow packs, while other areas have thinner ones, than we might want to think about each measurement of the snow pack as follows:

Model 2:

Measurement of "density" =

- A. Average density of snow on Mount David +
- B1. Difference between overall average and the average in specific area +
- B2. Variation that is not of interest to us +
- C. Measurement error

This second model s nearly identical to the model we used at the beginning of this discussion. The only difference is that the new model takes each measurement and breaks it down into four pieces instead of three. Two of the pieces in the second model (Parts B1 and B2) are produced by breaking up one part of the first model (Part B). The other two parts of the model (the average density of snow on Mount David and the measurement error, remain the same. What we have done is look more closely at the "variation that is not of interest to us", and decided that some of it may be of interest after all.

What effect does that have? Well, first we now are no longer estimating the average density of snow on Mount David as a whole. We are instead estimating the average density of snow in each of several sub-areas. But more importantly, if we have selected sub-areas of Mount David that really are different from one another, we have removed some variation from "variation that is not of interest to us". That will reduce the experimental error associated with each measurement (recall that experimental error is the sum of “variation that is not of interest to us” and measurement error). Lower experimental error means smaller standard deviations among measurements taken within each area. And, all things being equal, that should produce a lower standard error of the estimated mean snow density within each area.

Unfortunately, all things generally are not equal. Time constraints or cost usually limit us to a small number of measurements. That means the more areas we subdivide Mount David into, the fewer measurements we can use to estimate average conditions within each sub-area. There is a trade-off here. A lower standard deviation among measurements within an area makes for a lower standard error of our estimated mean. The smaller number of measurements that contribute to each average, however, tends to increase the standard error of the estimated mean. If we do this right, we replace one uncertain estimate of an average with several more accurate estimates of averages. If we do it poorly, we end up with several poorly characterized averages, made with too few measurements, and our overall estimate of the snow pack is worse that if we had just calculated a single average.

How do we determine the appropriate balance? Often we do so based primarily on prior experience. Where experience is not available, there are detailed statistical methods available for answering the question, however, they are beyond the scope of this course. You can get a pretty good idea of the effect of different study designs if you can guess (or measure) the amount of variability among measurements within each area, and calculate the standard error of the mean that you would EXPECT to see, if your guesses about variability among measurements are correct.

A few rules of thumb may help. Generally, if we divide an area into sub-areas that have very different snow characteristics, we should end up with improved estimates, provided the number of samples within each area does not get too small. How small a sample is too small? Generally, you will need at least 5 measurements within each sub-area. Increasing the sample size in each area to 20 would double the precision of your estimates, but an additional doubling of precision would require you to take 80 measurements in each area – which is likely to be a prohibitively large number of measurements to make in each of several areas.

Standard Error of the TOTAL Amount of Water in the snow pack

So far, we have focussed on how to estimate the AVERAGE density of water in the snow pack. However, we are not really interested in the average density of water in the snow pack, we want to know the TOTAL amount of water stored on Mount David.

To convert from average snow pack densities to the total water stored on Mount David, you calculate a weighted sum. (This formula assumes you chose to sample several sub areas of Mount David. If you did not, the calculations are a bit simpler).

$$\text{Estimated Total Water Stored} = \sum_{i=1}^{n} A_i D_i$$

A_i = Area of the ith subarea of Mount David (in square meters)

D_i = Average stored density of water (per meter squared) in the ith sub - area

The problem here is that we don't actually know the true average density of water in the snow pack. We have an estimate of that number. That estimate is associated with a certain amount of uncertainty. When we multiply an uncertain number by a large value (in this case, the area), we get a much larger number, and we also get a much more uncertain number.

The standard error of a weighted sum is easy to calculate. It is found as follows:

Standard error of estimated total water stored $= \sqrt{\sum_{i=1}^{n} A_i^2 (SE_{E_i})^2}$

A_i = Area of the ith sub-area of Mount David (in square meters)

E_i = Average stored density of water (per meter squared) in the ith sub-area

SE_{E_i} = Standard error of E_i. That is, the standard error associated with the estimated average density of water stored in the ith sub-area.

Available Equipment

- Meter Sticks
- Pesola Scales
- Ziplock bags
- Plastic Beakers
- Mettler Balance
- Shovels
- Maps of Mount David
- 50-meter fiberglass measuring tapes (only a few)

Suggested approach for thinking through measurement choices

1. Decide what measurements you will take to come up with different measurements of snow density per meter squared. Presumably you will need to measure the amount of water in a certain area or volume of snow.

2. Consider the level of accuracy you need for each measurement. Remember, if measurement error is much smaller than the level of real variation among measurements, than it will have little effect on the standard deviation of your measurements, and thus will have only a tiny effect on the accuracy of your estimated snow density.

3. Chose your measurement technologies.

4. Try to estimate the likely variability among measurements of snow density using your technology. You may want to walk around on Mount David to help you come up with these guesses. A meter stick or other measurement equipment might help you with this step.

5. Run through the calculations laid out on this handout with various sample sizes, plugging in your guesses for variability among measurements. This will give you a pretty good idea of how accurately you can estimate the snow pack, and provide you with a sense of the advantage of larger sample sizes. This is your estimate of how well you can estimate the total snow pack if you do not subdivide Mount David into smaller areas.

6. Try to choose a few "intelligent" subdivisions of Mount David, and estimate the difference in the snow pack between different areas (that meter stick may come in handy again). Sketch in the different areas on a copy of the map of Mount David.

7. Again, you can run through the calculations laid out on this handout and get a pretty good estimate of the accuracy you might expect if you subdivide Mount David into smaller areas. Be realistic about the number of measurements you plan to take. Is it worth subdividing?

A Final Note on Random Sampling

All of the statistical methods presented in this handout are based on the idea that the measurements you take are not biased in any way. In environmental work, this is often a difficult standard to meet. If you just walk up to Mount David, and start taking measurements, chances are you will take more measurements in easily accessible areas, and take fewer measurements in areas that are difficult to get to. Your measurements may also be concentrated in areas of more open ground, or in areas protected from the wind. A variety of human behavioral traits can lead to conscious, or more commonly, unconscious bias in where samples are collected. Any such bias will result in your measurements being unrepresentative of the snow pack on Mount David.

The gold standard for environmental sampling is to select your sampling locations at random. A fully random sample would mean that every point on mount David had an equal probability of being sampled. In practice, we have to be a bit pragmatic about collecting random samples. But the closer you can get to a random sample, the less likely you are to collect data that is not representative of conditions on Mount David.