DERIVATIONS: SECTIONS 3 AND 4 OF STAT 2.1X

Ani Adhikari
University of California, Berkeley
Spring 2013

This handout contains algebraic derivations of some results about averages and SDs that are covered in Sections 3 and 4 of Stat 2.1X.

Quick review of sigma notation
Let $x_1, x_2, \ldots, x_n$ be a list of numbers. Here n is a fixed positive integer that denotes the number of entries in the list. The notation

$$\sum_{i=1}^{n} x_i$$

is used to denote the sum of the entries in the list.

Basic properties of sums

1. Let $y_1, y_2, \ldots, y_n$ be another list, also consisting of n entries. Then

$$\sum_{i=1}^{n}(x_i + y_i) \ = \ \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$$

2. Let a and b be constants. Then

$$\sum_{i=1}^{n} ax_i \ = \ a\sum_{i=1}^{n} x_i \qquad\qquad \sum_{i=1}^{n}(x_i + b) \ = \ \sum_{i=1}^{n} x_i \ + \ nb$$

THE AVERAGE

The *average* or *mean* of the list will be denoted by μ, which is the Greek letter *mu*. The average is defined to be

$$\mu \ = \ \frac{1}{n}\sum_{i=1}^{n} x_i$$

Two ways of "smoothing": By Property 2 of sums,

$$\mu \ = \ \frac{1}{n}\sum_{i=1}^{n} x_i \ = \ \sum_{i=1}^{n} \frac{1}{n} \cdot x_i \ = \ \sum_{i=1}^{n} \frac{x_i}{n}$$

This formalizes the idea that the we can think of the average as a smoothing operation conducted in one of two ways:
• The definition of the average says, "Put everyone's contribution into one big pot and then divide the pot equally among the people."
• That is equivalent to saying, "Divide each person's contribution equally among all people."

Markov's inequality: Statement 1. Suppose $x_i \geq 0$ for $1 \leq i \leq n$. Let $k > 0$ be a constant.The proportion of entries that are greater than or equal to $k\mu$ is at most $1/k$.

Proof. First note that an equivalent statement is:

Markov's inequality: Statement 2. Suppose $x_i \geq 0$ for $1 \leq i \leq n$. Let $c > 0$ be a constant. The proportion of entries that are greater than or equal to c is at most μ/c.

That the two statements are equivalent can be seen by setting $c = k\mu$. We will now prove the second form.

$$
\begin{aligned}
n\mu &= \sum_{\text{all entries}} x_i \\
&= \sum_{\text{all entries}<c} x_i + \sum_{\text{all entries}\geq c} x_i \\
&\geq 0 + \sum_{\text{all entries}\geq c} c \\
&\geq c(\text{number of entries} \geq c)
\end{aligned}
$$

The first term in the third line uses the fact that the entries in the list are non-negative. Divide by n on both sides to get

$$\mu \geq c(\text{proportion of entries} \geq c)$$

which is equivalent to

$$(\text{proportion of entries} \geq c) \leq \mu/c$$

as was to be proved.

Deviations from average. Let $d_i = x_i - \mu$ be the *ith deviation from average.* Then $\sum_{i=1}^{n} d_i = 0$, and hence the average of the list of deviations is 0.

Proof.

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (x_i - \mu) = \sum_{i=1}^{n} x_i - n\mu = n\mu - n\mu = 0$$

The second equality follows from Property 2 of sums.

THE STANDARD DEVIATION

Definition. The *standard deviation* of a list of numbers is defined by

$$\text{SD} = \text{root mean square of the deviations from average}$$

The SD will be denoted by σ, which is the Greek letter *sigma*. By definition,

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$

The quantity inside the square root is called the *variance* of the list:

$$\text{variance} = \text{mean square of the deviations from average}$$

That is, the variance is

$$\sigma^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

The unit of measurement of variance is the square of that of the list, so it is often hard to interpret physically. However, variance has mathematical properties that make it easy to compute. Therefore many results about SDs are derived by first finding the variance using its computational properties, and as a final step taking the square root to get the SD which has units that make sense.

Computational formula. Here is a useful property of the variance, which speeds up computation.

$$\text{variance} = (\text{mean of squares}) - (\text{square of mean})$$

That is,

$$\sigma^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}x_i^2 \;-\; \mu^2$$

Proof.

$$\begin{aligned}
\sigma^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \;&=\; \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\mu + \mu^2)\\
&=\; \frac{1}{n}\sum_{i=1}^{n}x_i^2 \;-\; \frac{1}{n}\sum_{i=1}^{n}2x_i\mu \;+\; \frac{1}{n}\sum_{i=1}^{n}\mu^2\\
&=\; \frac{1}{n}\sum_{i=1}^{n}x_i^2 \;-\; 2\mu\frac{1}{n}\sum_{i=1}^{n}x_i \;+\; \frac{1}{n}n\mu^2\\
&=\; \frac{1}{n}\sum_{i=1}^{n}x_i^2 \;-\; 2\mu^2 \;+\; \mu^2\\
&=\; \frac{1}{n}\sum_{i=1}^{n}x_i^2 \;-\; \mu^2
\end{aligned}$$

Note. Since $\sigma^2 \geq 0$, the computational formula implies that

$$\text{mean of squares} \geq \text{square of mean}$$

for all lists; and that the mean of the squares is equal to the square of the mean if and only if the variance is 0, that is, if and only the standard deviation is 0, that is, if and only if all the numbers in the list are the same.

Chebychev's inequality: In any list of numbers, the proportion of entries that are k or more SDs away from the average is at most $1/k^2$.

Proof. The idea of the proof is to notice that the variance is the mean of the list of squared deviations, and to apply Markov's inequality to that list.

Step 1. Let $d_i = x_i - \mu$ be the ith deviation from average, as before. Let $w_i = d_i^2 = (x_i - \mu)^2$ be the ith squared deviation. The variance is the mean squared deviation, and so

$$\sigma^2 \ = \ \frac{1}{n}\sum_{i=1}^{n} w_i$$

In other words, $w_1, w_2, \ldots, w_n$ is a list of non-negative numbers whose average is σ^2. Markov's inequality can be applied to this list. Statement 2 of Markov's inequality says that for any $c > 0$,

$$(\text{proportion of entries such that } w_i \geq c) \ \leq \ \sigma^2/c$$

Step 2. For x_i to be k or more SDs away from the average, its distance from μ must be at least $k\sigma$. So it must satisfy $|x_i - \mu| \geq k\sigma$.

So our job is to show that

$$(\text{proportion of entries such that } |x_i - \mu| \geq k\sigma) \quad \leq \quad 1/k^2$$

Equivalently, we have to show that

$$(\text{proportion of entries such that } (x_i - \mu)^2 \geq k^2\sigma^2) \quad \leq \quad 1/k^2$$

Equivalently we have to show that, in the notation of Step 1,

$$(\text{proportion of entries such that } w_i \geq k^2\sigma^2) \quad \leq \quad 1/k^2$$

Step 3. So take $c = k^2\sigma^2$ in the result of Step 1. This leads to

$$(\text{proportion of entries such that } w_i \geq k^2\sigma^2) \ \leq \ \sigma^2/k^2\sigma^2 \ = \ 1/k^2$$

Linear transformations. Let a and b be constants. Construct a new list whose ith element is $y_i = ax_i + b$. Because we will now have a couple of means and SDs in our calculations, let us give them names that distinguish them from each other:

μ_x = average of the list of x's $\quad\quad$ σ_x = SD of the list of x's
μ_y = average of the list of y's $\quad\quad$ σ_y = SD of the list of y's

Then

$$\mu_y \ = \ a\mu_x \ + \ b \qquad\qquad \sigma_y = |a|\sigma_x$$

Proof.

$$\mu_y \ = \ \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) \ = \ \frac{1}{n}\sum_{i=1}^{n} ax_i \ + \ \frac{1}{n}\cdot nb \ = \ \frac{1}{n}\cdot a\sum_{i=1}^{n} x_i \ + \ b \ = \ \frac{1}{n}\cdot a \cdot n\mu_x \ + \ b \ = \ a\mu_x \ + \ b$$

To prove the formula for the SDs, notice that the ith deviation of the y's is

$$y_i - \mu_y \ = \ (ax_i + b) \ - \ \mu_y \ = \ (ax_i + b) - (a\mu_x + b) \ = \ a(x_i - \mu_x)$$

In other words, each deviation of the y's is equal to a times the corresponding deviation of the x's. This implies that each squared deviation of the y's is equal to a^2 times the corresponding squared deviation of the x's. Thus

$$\sigma_y^2 \;=\; a^2\sigma_x^2 \qquad \text{and so} \qquad \sigma_y \;=\; |a|\sigma_x$$

because the SD is the positive square root of the variance.

Standard units. For each i, let $z_i =$ "x_i in standard units" be defined by

$$z_i \;=\; \frac{x_i - \mu_x}{\sigma_x}$$

Then $\mu_z = 0$ and $\sigma_z = 1$.

Proof. For each i,

$$z_i \;=\; \frac{1}{\sigma_x} \cdot x_i \;-\; \frac{\mu_x}{\sigma_x} \;=\; ax_i + b$$

where $a = 1/\sigma_x$ and $b = -\mu_x/\sigma_x$. Apply the results about the mean and SD of a linear transformation to get

$$\mu_z \;=\; a\mu_x + b \;=\; \frac{1}{\sigma_x} \cdot \mu_x - \frac{\mu_x}{\sigma_x} = 0$$

and

$$\sigma_z \;=\; \frac{1}{\sigma_x} \cdot \sigma_x \;=\; 1$$

Note: We are assuming $\sigma_x > 0$, because if σ_x were equal to 0 then we could not divide by it. But if $\sigma_x = 0$ then all the x's would be equal, which is not a case that requires analysis by conversion into other units.