

[Close](#)

for Educators

[CLICK TO PRINT PAGE](#)

Here is the article in a print-friendly format. [Click the button above to print this page.](#)

Why Variances Add -- And Why It Matters

by Dave Bock
Cornell University
Ithaca, New York

The Pythagorean Theorem of Statistics

Quick: What's the most important theorem in statistics? That's easy; it's the Central Limit Theorem, hands down. Okay, how about the second most important theorem? I'd argue that it's the fact that for the sum or difference of independent random variables, variances add:

For independent random variables X and Y ,

$$\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$$

I like to refer to this statement as the Pythagorean Theorem of Statistics, for several reasons:

1. When written in terms of standard deviations, it looks like the Pythagorean theorem:

$$SD^2(X \pm Y) = SD^2(X) + SD^2(Y)$$

2. Just as the Pythagorean theorem applies only to **right** triangles, this relationship applies only to **independent** random variables.
3. The name helps kids remember both the relationship and the restriction.

As you may suspect, this analogy is more than a mere coincidence. There's a nice geometric model that represents random variables as vectors whose lengths correspond to their standard deviations. When the variables are independent, the vectors are orthogonal, and then the standard deviation of the sum or difference of the variables is just the hypotenuse of a right triangle. You probably won't discuss orthogonal vectors with your AP Statistics students, but that's no excuse for not giving the Pythagorean theorem the emphasis it deserves. When your students understand and learn to use this theorem, many doors will open. They'll gain important insights in dealing with binomial probabilities, inference, and even the Central Limit Theorem itself -- and they'll gain an important problem-solving skill sure to pay off on the AP Exam.

Some Questions

Let's start by taking a look at the theorem itself. Three questions come immediately to mind:

1. Why do we add the **variances**?
2. Why do we add even when working with the **difference** of the random variables?
3. Why do the variables have to be **independent**?

I'll try to answer these questions on two levels. On one level is the proof, just to make you feel better. While some teachers may decide to show this proof to their classes, most won't inflict it on AP Stats students. Instead, a plausibility argument should

suffice; by that I mean a series of justifications that stop short of being formal proofs yet provide kids with clear examples that make the theorem believable, rather than just a meaningless rule to memorize.

Proving the Theorem: The Mathematics

First, then, let's have a look at a formal proof. I think it's a rule of mathematics that no proof should be taken seriously unless there's a lemma, so here's ours. The proof uses the fact that, because expected values are basically summations, they are additive:

$$E(X \pm Y) = E(X) \pm E(Y).$$

Lemma: $Var(X) = E(x^2) - \mu^2.$

Proof: $Var(X) = E[(x - \mu)^2]$

$$= E(x^2 - 2x\mu + \mu^2)$$

$$= E(x^2) - 2\mu E(x) + E(\mu^2)$$

$$= E(x^2) - 2\mu \cdot \mu + \mu^2$$

$$= E(x^2) - \mu^2.$$

And now, the Pythagorean Theorem of Statistics:

Theorem: If X and Y are independent, $Var(X \pm Y) = Var(X) + Var(Y).$

Proof: $Var(X \pm Y) = E[(x \pm y)^2] - (\mu_{x \pm y})^2$

$$= E[(x \pm y)^2] - (\mu_x \pm \mu_y)^2$$

$$= E(x^2 \pm 2xy + y^2) - (\mu_x^2 \pm 2\mu_x\mu_y + \mu_y^2)$$

$$= E(x^2) \pm 2E(xy) + E(y^2) - \mu_x^2 \mp 2\mu_x\mu_y - \mu_y^2$$

$$= E(x^2) - \mu_x^2 \pm 2[E(xy) - \mu_x\mu_y] + E(y^2) - \mu_y^2.$$

Consider that middle term: $E(xy) - \mu_x\mu_y.$

$E(xy)$ is the sum of all terms of the form

$x_i y_j \cdot P(x_i \cap y_j).$ The

product $\mu_x \mu_y$ is the

sum of all terms of the

form $x_i P(x_i) \cdot$

$y_j P(y_j).$ If X and Y

are independent, each

term in the first sum is

equal to the corresponding term in the second sum; hence that middle term is 0. Thus:

$$Var(X \pm Y) = E(x^2) - \mu_x^2 + E(y^2) - \mu_y^2$$

$$= Var(X) + Var(Y).$$

Note that this proof answers all three questions we posed. It's the variances that add. Variances add for the sum and for the difference of the random variables because the plus-or-minus terms all dropped out along the way. And independence was why part of the expression vanished, leaving us with the sum of the variances.

Teaching the Theorem: Building Understanding

While that proof may make you feel better about the theorem (or not), it's not likely to warm the hearts of most of your students. Let's have a look at some arguments you can make in class that should convince your students that the theorem makes sense.

Question 1: Why add variances instead of standard deviations?

We always calculate variability by summing squared deviations from the mean. That gives us a variance -- measured in square units (square dollars, say, whatever those are . . .). We realign the units with the variable by taking the square root of that variance, giving us the standard deviation (now in dollars again). To get the standard deviation of the sum of the variables, we need to find the square root of the sum of the squared deviations from the mean. Kids have learned in algebra that they shouldn't add the square roots, because $a + b \neq \sqrt{a^2 + b^2}$. Although $3 + 4 = 7$, we need $\sqrt{3^2 + 4^2} = 5$, the Pythagorean approach. We add the variances, not the standard deviations.

Question 2: Why add even for the difference of the variables?

We buy some cereal. The box says "16 ounces." We know that's not precisely the weight of the cereal in the box, just close; after all, one corn flake more or less would change the weight ever so slightly. Weights of such boxes of cereal vary somewhat, and our uncertainty about the exact weight is expressed by the variance (or standard deviation) of those weights.

Next we get out a bowl that holds 3 ounces of cereal and pour it full. Our pouring skill certainly is not very precise, so the bowl now contains about 3 ounces with some variability (uncertainty).

How much cereal is left in the box? Well, we'd assume about 13 ounces. But notice that we're less certain about this remaining weight than we were about the weight before we poured out the bowlful. The variability of the weight in the box has **increased** even though we subtracted cereal.

Moral: Every time something happens at random, whether it adds to the pile or subtracts from it, uncertainty (read "variance") increases.

Question 2 (follow-up): Okay, but is the effect exactly the same when we subtract as when we add?

Suppose we have some grapefruit weighing between 16 and 24 ounces and some oranges weighing between 9 and 13 ounces. We pick one of each at random.

- Consider the total possible weight of the two fruits. The maximum total is $24 + 13 = 37$ ounces, and the minimum is $16 + 9 = 25$ ounces -- a range of 12 ounces.
- Now consider the possible weight difference. The maximum difference is $24 - 9 = 15$ ounces, and the minimum is $16 - 13 = 3$ ounces -- again a range of 12 ounces. So whether we're adding or subtracting the random variables, the resulting range (one measure of variability) is exactly the same. That's a plausibility argument that the standard deviations of the sum and the difference should be the same, too.

Question 3: Why do the variables have to be independent?

Consider a survey in which we ask people two questions: During the last 24 hours, how many hours were you asleep? And how many hours were you awake?

There will be some mean number of sleeping hours for the group, with some standard deviation. There will also be a mean and standard deviation of waking hours. But now let's sum the two answers for each person. What's the standard deviation of this sum? It's 0, because that sum is 24 hours for everyone -- a constant. Clearly variances did not add here.

Why not? These data are paired, not independent, as required by the theorem. Just as we can't apply the Pythagorean theorem without first being sure we are dealing with a right triangle, we can't add variances until we're sure the random variables are independent. (This is yet another place where students must remember to check a condition before proceeding.)

Why Does It Matter?

Many teachers wonder if teaching this theorem is worth the struggle. I argue here that getting kids to understand this key concept (1) is not that difficult and (2) pays off repeatedly throughout the rest of the course, on the AP Exam, and in future work our students may do in statistics. Indeed, it arises so frequently that the statement "For sums or differences of independent random variables, variances add" becomes something of a mantra in my classroom. Let's take a tour of some of the many places the Pythagorean Theorem of Statistics holds the key to understanding.

Working with Sums

Remember Matt and Dave's Video Venture, a multiple-choice question from the 1997 AP Exam? At Matt and Dave's, every Thursday was Roll-the-Dice Day, allowing patrons to rent a second video at a discount determined by the digits rolled on two dice. Students were told that these second movies would cost an average of \$0.47 with a standard deviation is \$0.15. Then they were asked:

If a customer rolls the dice and rents a second movie every Thursday for 30 consecutive weeks, what is the approximate probability that the total amount paid for these second movies will exceed \$15.00?

One route to the solution adds variances.

First we note that the total amount paid is the sum of 30 daily values of a random variable.

$$T = X_1 + X_2 + X_3 + \cdots + X_{30}$$

We find the expected total.

$$\begin{aligned} E(T) &= 0.47 + 0.47 + 0.47 + \cdots + 0.47 \\ &= \$14.10. \end{aligned}$$

Because rolls of the dice are independent, we can apply the Pythagorean theorem to find the variance of the total, and that gives us the standard deviation.

$$\begin{aligned} Var(T) &= Var(X_1) + Var(X_2) + \cdots + Var(X_{30}) \\ &= 0.15^2 + 0.15^2 + 0.15^2 + \cdots + 0.15^2 \\ &= 0.675. \\ SD(T) &\approx 0.822. \end{aligned}$$

The Central Limit Theorem tells us that sums (essentially the same thing as means) of independent random variables approach a normal model as n increases. With $n = 30$ here, we can safely estimate the probability that $T > 15.00$ by working with the model $N(14.10, 0.822)$.

Working with Differences

On the 2000 AP Exam, the investigative task asked students to consider heights of men and women. They were given that the heights of each sex are described by a normal model; means were given as 70 inches for men and 65 inches for women, with standard deviations of 3 inches and 2.5 inches respectively. Among the questions asked was:

Suppose a married man and a married woman are each selected at random. What is the probability the woman will be taller than the man?

Again, we can solve the problem by adding variances:

First, define the random variables.

M = Height of the chosen man, W = Height of the woman.

We're interested in the difference of their heights.

Let D = Difference in their heights: $D = M - W$.

$$\begin{aligned} E(D) &= E(M) - E(W) \\ &= 70 - 65 \\ &= 5. \end{aligned}$$

Since the people were selected at random, the heights are independent, so we can find the standard deviation of the difference using the Pythagorean theorem.

$$\begin{aligned} SD(D) &= \sqrt{3^2 + 2.5^2} \\ &\approx 3.905. \end{aligned}$$

The difference of two normal random variables is also normal, so we can now find the probability that the woman is taller using the z-score for a difference of 0.

$$\begin{aligned} P(W > M) &= P(M - W < 0) \\ &= P(D < 0) \\ &= P\left(z < \frac{0 - 5}{3.905}\right). \end{aligned}$$

Standard Deviation for the Binomial

How many 4s do we expect when we roll 600 dice? 100 seems pretty obvious, and students rarely question the fact that for a binomial model $\mu = np$. However, the standard deviation is not so obvious. Let's derive that formula.

We start by looking at a probability model for a single Bernoulli trial.

Let X = The number of successes.

$$P(x = 1) = p, \quad P(x = 0) = q.$$

We find the mean of this random variable.

$$\begin{aligned} E(X) &= 1 \cdot p + 0 \cdot q \\ &= p. \end{aligned}$$

And then the variance.

$$\begin{aligned} Var(X) &= (1-p)^2 p + (0-p)^2 q \\ &= q^2 p + p^2 q \\ &= pq(q+p) \\ &= pq. \end{aligned}$$

Now we count the number of successes in n independent trials.

$$Y = X_1 + X_2 + X_3 + \dots + X_n.$$

The mean is no surprise.

$$\begin{aligned} E(Y) &= E(x_1) + E(x_2) + \dots + E(x_n) \\ &= p + p + p + \dots + p \\ &= np. \end{aligned}$$

And the standard deviation? Just add variances.

$$\begin{aligned} Var(Y) &= Var(x_1) + Var(x_2) + \dots + Var(x_n) \\ &= pq + pq + pq + \dots + pq \\ &= npq. \\ SD(Y) &= \sqrt{npq}. \end{aligned}$$

The Central Limit Theorem

By using the second most important theorem in statistics, we can derive part of the most important theorem. The Central Limit Theorem tells us something quite surprising and beautiful: when we sample from any population, regardless of shape, the behavior of sample means (or sums) can be described by a normal model with increasing accuracy as the sample size increases. That result is not just stunning, it's also quite fortunate, as most of the rest of what we teach in AP Statistics would not exist were it not true!

The full proof of the Central Limit Theorem is well beyond the scope of this article. What's within our grasp here is the theorem's quantification of the variability in these sample means, and the key is (drum roll!) adding variances.

The mean is basically the sum of n independent random variables, so:

$$\begin{aligned}
 \text{Var}(\bar{x}) &= \text{Var}\left(\frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}\right) \\
 &= \frac{1}{n^2} \text{Var}(x_1 + x_2 + x_3 + \cdots + x_n) \\
 &= \frac{1}{n^2} [\text{Var}(x_1) + \text{Var}(x_2) + \cdots + \text{Var}(x_n)] \\
 &= \frac{1}{n^2} [\sigma^2 + \sigma^2 + \cdots + \sigma^2] \\
 &= \frac{1}{n^2} \cdot n\sigma^2 \\
 &= \frac{\sigma^2}{n}.
 \end{aligned}$$

Hence,

$$SD(\bar{x}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}.$$

Inference for the Difference of Proportions

The Pythagorean theorem also lets kids make sense of those otherwise scary-looking formulas for inferences involving two samples. Indeed, I've found that students can come up with the formulas for themselves. Here's how that plays out in my classroom.

To set the stage for this discussion, we have just started inference. We first developed the concept of confidence intervals by looking at a confidence interval for a proportion. We then discussed hypothesis tests for a proportion, and we have spent a few days practicing the procedures. By now kids understand the ideas and can write up both a confidence interval and a hypothesis test, but only for one proportion. When class starts, I propose the following scenario:

Will a group counseling program help people who are using "the patch" actually manage to quit smoking? The idea is to have people attend a weekly group discussion session with a counselor to create a support group. If such a plan actually proved to be more effective than just wearing the patch, we'd seek funding from local health agencies. Describe an appropriate experiment.

We begin by quickly drawing a flowchart for the experiment -- a good review lesson. Start with a cohort of volunteer smokers who are trying to quit. Randomly divide them into two groups. Both groups get the patch; one group also attends these support/counseling sessions. Wait six months, then compare the success rates in the two groups.

Earlier in the course, this is where the discussion ended, but now we are ready to finish the job and actually "compare the success rates":

After six months, 46 of the 143 people who had worn the patch and participated in the counseling groups successfully quit smoking. Among those who received the patch but no counseling, 30 of 151 quit smoking. Do these results provide evidence that the counseling program is effective?

Students recognize that we need to test a hypothesis, and they point out that this is a different situation because there are two groups. I bet them they can figure out how to do it and start writing their suggestions on the board. They propose a hypothesis that the success rates are the same: $H_0: p_C = p_{NC}$. I agree, and then I add that we may also write this hypothesis as a statement of no difference: $H_0: p_C - p_{NC} = 0$.

They dictate the randomization, success/failure, and 10 percent conditions that allow the use of a normal model. They compute the sample proportions and find the observed difference, $\hat{p}_C - \hat{p}_{NC} = 0.322 - 0.199 = 0.123$. Here's where it gets interesting. They need to find the probability of observing a difference in proportions at least as large as

0.123, when they were expecting a difference of zero. They start to find the z -score, $z = \frac{0.123 - 0}{???}$, but get stumped by the denominator. I wait. If necessary, I point out that we need to find the standard deviation of the difference of the sample proportions. Soon someone gets it: "Add the variances!"

Good idea, but can we add the variances? Only if the groups are independent. Why are they independent? Randomization! We return to the list of conditions and add one more: the independent groups condition. At this point, rather than memorizing a list of conditions, everyone clearly realizes why this condition must be met.

Next we look at what happens when we add the variances:

$$\begin{aligned} SD(\hat{p}_C - \hat{p}_{NC}) &= \sqrt{SD^2(\hat{p}_C) + SD^2(\hat{p}_{NC})} \\ &= \sqrt{\left(\sqrt{\frac{p_C q_C}{n_C}}\right)^2 + \left(\sqrt{\frac{p_{NC} q_{NC}}{n_{NC}}}\right)^2} \\ &= \sqrt{\frac{p_C q_C}{n_C} + \frac{p_{NC} q_{NC}}{n_{NC}}}. \end{aligned}$$

Voila! The students have derived the formula for the standard deviation of the difference of sample proportions; thus it makes sense to them. We still need to talk about issues like using the sample proportions as estimates and pooling, but the basic formula is at hand and understood.

Inference for the Difference of Means

There's no need to include here a long-winded example that's analogous to the situation for proportions. It's enough to see that the standard deviation for the difference of sample means is also based on adding variances and that students can derive it on their own the first time you test a hypothesis about the difference of means of independent groups:

$$\begin{aligned} SD(\bar{x}_1 - \bar{x}_2) &= \sqrt{SD^2(\bar{x}_1) + SD^2(\bar{x}_2)} \\ &= \sqrt{\left(\frac{\sigma_1}{\sqrt{n_1}}\right)^2 + \left(\frac{\sigma_2}{\sqrt{n_2}}\right)^2} \\ &= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}. \end{aligned}$$

Two-Sample t-Procedures, or Matched Pairs -- It Matters!

Without advance fanfare, I propose to the class that we construct a confidence interval to see how many extra miles per gallon we might get if we use premium rather than regular gasoline. I give the students data from an experiment that tried both types of fuel in several cars (a situation involving matched pairs, but I don't point that out). When we start constructing the confidence interval, invariably someone will question the assumption that two measurements made for the same car are independent. I force them to clearly explain why that matters. The insight that lack of independence prevents adding variances, which in turn renders the formula for a two-sample t-interval incorrect, makes it forever clear to students that they must think carefully about the design under which their data were collected before plunging into the analysis. There's never a "choice" whether to use a paired differences procedure or a two-sample t-method.

Create One Confidence Interval, or Two?

| | Diet S | Diet N |
|-------------|---------|---------|
| n | 36 | 36 |
| Mean | 55 lbs. | 53 lbs. |
| SD | 3 lbs. | 4 lbs. |

Suppose we wonder if a food supplement can increase weight gain in feeder pigs. An experiment randomly assigns some pigs to one of two diets, identical except for the inclusion of the supplement in the feed for Group S but not for Group N. After a few weeks, we weigh the pigs; summaries of the weight gains appear in the table. Is there evidence that the supplement was effective? A reasonable way to try to answer this question is by using confidence intervals, but which approach is correct?

Plan A: Compare confidence intervals for each group. The separate intervals indicate we can be 95 percent confident that pigs fed this dietary supplement for this period of time would gain an average of between 53.90 and 56.01 pounds, while the average gain for those not fed the supplement would be between 51.65 and 54.35 pounds.

Note that these intervals overlap. It appears to be possible that the two diets might result in the same mean weight gain, hence we lack evidence that the supplement is effective.

Plan B: Construct one confidence interval for the difference in mean weight gain. We can be 95 percent confident that pigs fed the food supplement would gain an average of between 0.34 and 3.66 pounds more than those who were not fed this supplement. Because 0 is not in this confidence interval, we have strong evidence that this food supplement can make feeder pigs gain more weight.

Clearly the conclusions are contradictory, yet it may not be immediately clear which is correct. Students often see nothing wrong with Plan A, yet that analysis is inappropriate. Whether the two intervals overlap depends on whether the two means are farther apart than the sum of the margins of error. The mistake rests in the fact that we shouldn't add the margins of error. Why not? A confidence interval's margin of error is based on a standard deviation (well, standard error to be more exact), but standard deviations don't add; variances do. Plan B's confidence interval for the difference bases its margin of error on the standard error for the difference of two sample means, calculated by adding the two variances. That's the correct approach -- one confidence interval, not two.

Future Topics in Statistics

If you've stuck with the discussion this far, you should be convinced that adding variances plays a key role in much of the

statistics we teach in the AP course. As our students expand their knowledge of statistics by taking more courses beyond AP Stats, they will encounter the Pythagorean theorem again and again. To cite a few places on the horizon:

- **Prediction intervals:** When we use a regression line to make a prediction, we should include a margin of error around our estimate. That uncertainty involves three independent sources of error: (1) the line may be misplaced vertically because our sample mean only approximates the true mean of the response variable, (2) our sample data only gives us an estimate of the true slope, and (3) individuals vary above and below the line.
- **Multiple regression:** When we use several independent factors to arrive at an estimate for the response variable, we assess the strength of the model by looking at the total amount of variability it explains, and we are further able to attribute some of that variability to each of the explanatory variables.
- **ANOVA:** As the name "analysis of variance" suggests, we compare the effects of treatments on multiple groups or assess the effects of several treatments in a multifactor design by comparing the variability seen within groups to the total variability across groups. Here again, the idea of adding variances lies at the heart of the statistics.

Let's summarize: Variances add, and, yes, it matters!

Dave Bock has been a high school math teacher since 1969. He holds a BA in mathematics and an MS in statistics from the University at Albany. He has taught statistics at Ithaca High School, Cornell University, Ithaca College, and Tompkins-Cortland Community College, and has been teaching AP Statistics since its inception. He has served as an AP Statistics Reader and consultant since 1999, and leads the St. Johnsbury Summer Institute for teachers of AP Statistics. Dave is co-author of Barron's AP Calculus review book and of Stats: Modeling the World, an AP Statistics text published by Addison-Wesley