How would you explain covariance to someone who understands only the mean?

...assuming that I'm able to augment their knowledge about variance in an intuitive fashion (Understanding "variance" intuitively) or by saying: It's the average distance of the data values from the 'mean' - and since variance is in square units, we take the square root to keep the units same and that is called standard deviation.

Let's assume this much is articulated and (hopefully) understood by the 'receiver'. Now what is covariance and how would one explain it in simple English without the use of any mathematical terms/formulae? (I.e., intuitive explanation. ;)

Please note: I do know the formulae and the math behind the concept. I want to be able to 'explain' the same in an easy to understand fashion, without including the math; i.e., what does 'covariance' even mean?

variance covariance intuition

edited Nov 8 '11 at 4:17
Mike Wierzbicki
1,951 2 11 26

asked Nov 7 '11 at 19:41
PhD
1,588 2 17 27

@Xi'an - 'how' exactly would you define it *via simple linear regression*? I'd really like to know... – PhD Nov 8 '11 at 2:08

1 Assuming you already have a scatterplot of your two variables, x *vs.* y, with origin at (0,0), simply draw two lines at x=mean(x) (vertical) and y=mean(x) (horizontal): using this new system of coordinates (origin is at (mean(x),mean(y)), put a "+" sign in the top-right and bottom-left quadrants, a "-" sign in the two other quadrants; you got the sign of the covariance, which is basically what @Peter said. Scaling the x- and y-units (by SD) lead to a more interpretable summary, as discussed in the ensuing thread. – chl ♦ Nov 9 '11 at 22:54

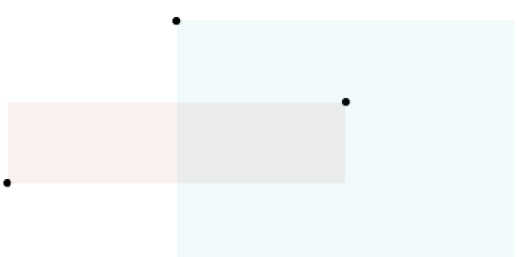
@chl - could you please post that as an answer and maybe use graphics to depict it! – PhD Nov 10 '11 at 4:03

6 Answers

Sometimes we can "augment knowledge" with an unusual or different approach. I would like this reply to be accessible to kindergartners and also have some fun, so **everybody get out your crayons!**

Given paired (x,y) data, draw their scatterplot. (The younger students may need a teacher to produce this for them. :-) Each pair of points (x_i,y_i), (x_j,y_j) in that plot determines a rectangle: it's the smallest rectangle, whose sides are parallel to the axes, containing those points. Thus the points are either at the upper right and lower left corners (a "positive" relationship) or they are at the upper left and lower right corners (a "negative" relationship).

Draw all possible such rectangles. Color them transparently, making the positive rectangles red (say) and the negative rectangles "anti-red" (blue). In this fashion, wherever rectangles overlap, their colors are either enhanced when they are the same (blue and blue or red and red) or cancel out when they are different.



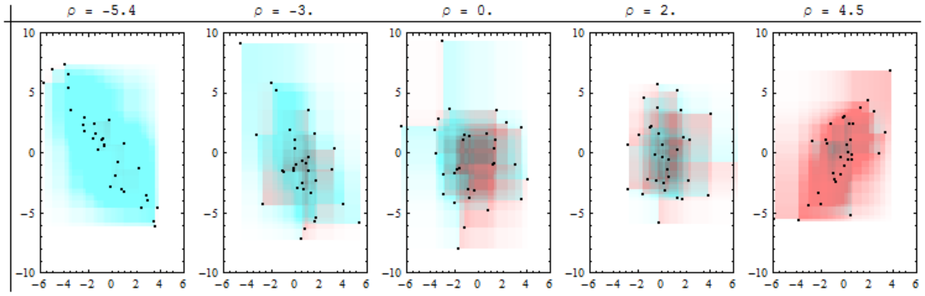
(*In this illustration of a positive (red) and negative (blue) rectangle, the overlap ought to be white; unfortunately, this software does not have a true "anti-red" color. The overlap is gray, so it will darken the plot, but on the whole the* net *amount of red is correct.*)

Now we're ready for the explanation of covariance.

Now we're ready for the explanation of covariance.

The covariance is the net amount of red in the plot (treating blue as negative values).

Here are some examples with 32 binormal points drawn from distributions with the given covariances, ordered from most negative (bluest) to most positive (reddest).



Let's deduce some properties of covariance. Understanding of these properties will be accessible to anyone who has actually drawn a few of the rectangles. :-)

- **Bilinearity.** Because the amount of red depends on the size of the plot, covariance is directly proportional to the scale on the x-axis and to the scale on the y-axis.
- **Correlation.** Covariance increases as the points approximate an upward sloping line and decreases as the points approximate a downward sloping line. This is because in the former case most of the rectangles are positive and in the latter case, most are negative.
- **Relationship to linear associations.** Because non-linear associations can create mixtures of positive and negative rectangles, they lead to unpredictable (and not very useful) covariances. Linear associations can be fully interpreted by means of the preceding two characterizations.
- **Sensitivity to outliers.** A geometric outlier (one point standing away from the mass) will create many large rectangles in association with all the other points. It alone can create a net positive or negative amount of red in the overall picture.

Incidentally, this definition of covariance differs from the usual one only by a universal constant of proportionality (independent of the data set size). The mathematically inclined will have no trouble performing the algebraic demonstration of the equivalence.

edited Nov 10 '11 at 18:45

answered Nov 10 '11 at 17:14
whuber ♦
74.2k 7 119 260

19 +1 Wow. This even works for explaining covariance to those who already thought they knew what it was. – Aaron Nov 10 '11 at 17:24

3 +1 I really enjoy reading your response. I will draw some rectangles, and let my son paint them :) – chl ♦ Nov 10 '11 at 18:01

10 Now if only all introductory statistical concepts could be presented to students in this lucid manner … – MannyG Nov 10 '11 at 18:26

2 +1 Wow! Very nice indeed. – Dilip Sarwate Nov 10 '11 at 19:12

2 This is beautiful. And very very clear. – Benjamin Mako Hill Jun 2 '12 at 15:37

To elaborate on my comment, I used to teach the covariance as a measure of the (average) co-variation between two variables, say x and y.

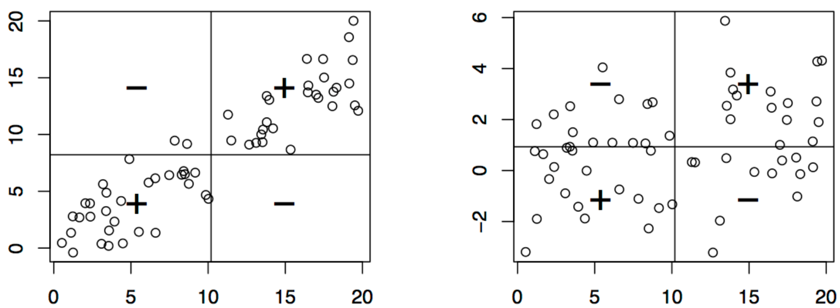
It is useful to recall the basic formula (simple to explain, no need to talk about mathematical expectancies for an introductory course):

$$ \text{cov}(x,y)=\frac{1}{n}\sum_{i=1}^n(x_i-\bar x)(y_i-\bar y) $$

so that we clearly see that each observation, (x_i,y_i), might contribute positively or negatively to the covariance, depending on the product of their deviation from the mean of the two variables, $\bar x$ and $\bar y$. Note that I do not speak of magnitude here, but simply of the sign of the contribution of the ith observation.

This is what I've depicted in the following diagrams. Artificial data were generated using a linear model (left, $y = 1.2x + \varepsilon$; right, $y = 0.1x + \varepsilon$, where ε

linear model (left, $y = 1.2x + \varepsilon$; right, $y = 0.1x + \varepsilon$, where ε were drawn from a gaussian distribution with zero mean and $\text{SD}=2$, and x from an uniform distribution on the interval $[0,20]$).



The vertical and horizontal bars represent the mean of x and y, respectively. That mean that instead of "looking at individual observations" from the origin $(0,0)$, we can do it from $(\bar x, \bar y)$. This just amounts to a translation on the x- and y-axis. In this new coordinate system, every observation that is located in the upper-right or lower-left quadrant contributes positively to the covariance, whereas observations located in the two other quadrants contribute negatively to it. In the first case (left), the covariance equals 30.11 and the distribution in the four quadrants is given below:

	+	-
+	30	2
-	0	28

Clearly, when the x_i's are above their mean, so do the corresponding y_i's (wrt. $\bar y$). Eye-balling the shape of the 2D cloud of points, when x values increase y values tend to increase too. (But remember we could also use the fact that there is a clear relationship between the covariance and the slope of the regression line, i.e. $b=\text{Cov}(x,y)/\text{Var}(x)$.)

In the second case (right, same x_i), the covariance equals 3.54 and the distribution across quadrants is more "homogeneous" as shown below:

	+	-
+	18	14
-	12	16

In other words, there is an increased number of case where the x_i's and y_i's do not covary in the same direction wrt. their means.

Note that we could reduce the covariance by scaling either x or y. In the left panel, the covariance of $(x/10,y)$ (or $(x,y/10)$) is reduced by a ten fold amount (3.01). Since the units of measurement and the spread of x and y (relative to their means) make it difficult to interpret the value of the covariance in absolute terms, we generally scale both variables by their standard deviations and get the correlation coefficient. This means that in addition to re-centering our (x,y) scatterplot to $(\bar x, \bar y)$ we also scale the x- and y-unit in terms of standard deviation, which leads to a more interpretable measure of the linear covariation between x and y.

answered Nov 10 '11 at 10:50
chl ♦
30.1k 5 80 186

Covariance is a measure of how much one variable goes up when the other goes up.

answered Nov 7 '11 at 23:33
Peter Flom ♦
36.5k 5 31 83

Is it always in the 'same' direction? Also, does it apply for inverse relations too (i.e., as one goes up the other goes down)? – PhD Nov 8 '11 at 2:07

I think that that's what determines the sign of covariance...as per my posted 'answer' – PhD Nov 8 '11 at 2:29

3 @nupul Well, the opposite of "up" is "down" and the opposite of "positive" is "negative". I tried to give a one sentence answer. Yours is much more complete. Even your "how two variables change together" is more complete, but, I think, a little harder to understand. – Peter Flom ♦ Nov 8 '11 at 11:37

1 +1 for fitting it in a single, simple sentence, but isn't that correlation? I mean, I know greater cov=> greater corr, but with that sentence, I'd expect something like "80%" as an answer, which corresponds to corr=0.8. Doesn't cov also describe the variance within the data? ie. "Covariance is proportional to how much one variable goes up when the other goes up, and also proportional to the spread of the data in both variables", or something? – naught101 Feb 28 '12 at 5:44

1 That's right, Peter, which is why @naught101 made that comment: you description sounds like a rate of change, whose units will therefore be [units of one variable] / [units of the other variable] (if we interpret it like a derivative) or will just be [units of one variable] (if we interpret as a pure difference). Those are neither covariance (whose unit of measure is the product of the units for the two variables) nor correlation (which is unitless). – whuber ♦ Aug 13 '13 at 20:15

I *am* answering my own question, but I thought It'd be great for the people coming across this post to check out some of the explanations on this page.

I'm paraphrasing one of the very well articulated answers (by a user'Zhop'). I'm doing so in case if that site shuts down or the page gets taken down when someone eons from now accesses this post ;)

> Covariance is a measure of how much two variables change together. Compare this to Variance, which is just the range over which one measure (or variable) varies.
>
> In studying social patterns, you might hypothesize that wealthier people are likely to be more educated, so you'd try to see how closely measures of wealth and education stay together. You would use a measure of covariance to determine this.
>
> ...
>
> I'm not sure what you mean when you ask how does it apply to statistics. It is one measure taught in many stats classes. Did you mean, when should you use it?
>
> You use it when you want to see how much two or more variables change in relation to each other.
>
> Think of people on a team. Look at how they vary in geographic location compared to each other. When the team is playing or practicing, the distance between individual members is very small and we would say they are in the same location. And when their location changes, it changes for all individuals together (say, travelling on a bus to a game). In this situation, we would say they have a high level of covariance. But when they aren't playing, then the covariance rate is likely to be pretty low, because they are all going to different places at different rates of speed.
>
> So you can predict one team member's location, based on another team member's location when they are practicing or playing a game with a high degree of accuracy. The covariance measurement would be close to 1, I believe. But when they are not practicing or playing, you would have a much smaller chance of predicting one person's location, based on a team member's location. It would be close to zero, probably, although not zero, since sometimes team members will be friends, and might go places together on their own time.
>
> However, if you randomly selected individuals in the United States, and tried to use one of them to predict the other's locations, you'd probably find the covariance was zero. In other words, there is absolutely no relation between one randomly selected person's location in the US, and another's.

Adding another one (by 'CatofGrey') that helps augment the intuition:

> In probability theory and statistics, covariance is the measure of how much two random

> variables vary together (as distinct from variance, which measures how much a single variable varies).
>
> If two variables tend to vary together (that is, when one of them is above its expected value, then the other variable tends to be above its expected value too), then the covariance between the two variables will be positive. On the other hand, if one of them is above its expected value and the other variable tends to be below its expected value, then the covariance between the two variables will be negative.

These two together have made me understand covariance as I've never understood it before! Simply amazing!!

answered Nov 8 '11 at 2:23
PhD
1,588 2 17 27

8 Although these descriptions are qualitatively suggestive, sadly they are incomplete: they neither distinguish covariance from correlation (the first description appears to confuse the two, in fact), nor do they bring out the fundamental assumption of *linear* co-variation. Also, neither addresses the important aspect that covariance depends (linearly) on the scale of each variable. – whuber ♦ Nov 8 '11 at 14:35

@whuber - agreed! And hence haven't marked mine as the answer :) (not as yet ;) – PhD Nov 9 '11 at 0:45

Two variables that would have a high positive covariance (correlation) would be the number of people in a room, and the number of fingers that are in the room. (As the number of people increases, we expect the number of fingers to increase as well.)

Something that might have a negative covariance (correlation) would be a person's age, and the number of hair follicles on their head. Or, the number of zits on a person's face (in a certain age group), and how many dates they have in a week. We expect people with more years to have less hair, and people with more acne to have less dates.. These are negatively correlated.

answered Nov 9 '11 at 9:28
Adam
495 3 13

2 Covariance is not necessarily interchangeable with correlation - the former is very unit dependent. Correlation is a number between -1 and 1 a unit-less scalar representing the 'strength' of the covariance IMO and that's not clear from your answer – PhD Nov 9 '11 at 18:24

I would simply explain correlation which is pretty intuitive. I would say "Correlation measures the strength of relationship between two variables X and Y. Correlation is between -1 and 1 and will be close to 1 in absolute value when the relationship is strong. Covariance is just the correlation multiplied by the standard deviations of the two variables. So while correlation is dimensionless, covariance is in the product of the units for variable X and variable Y.

answered May 6 '12 at 23:56
Michael Chernick
21.5k 2 20 46

6 This seems inadequate because there is no mention of linearity. X and Y could have a strong quadratic relationship but have a correlation of zero. – mark999 May 7 '12 at 0:46

Answer This Question