How would you explain the difference between correlation and covariance?

20

25

Following up on this question, How would you explain covariance to someone who understands only the mean?, which addresses the issue of explaining covariance to a lay person, brought up a similar question in my mind.

How would one explain to a statistics neophyte the difference between *covariance* and *correlation*? It seems that both refer to the change in one variable linked back to another variable.

Similar to the referred-to question, a lack of formulae would be preferable.

correlation covariance

share improve this question

edited May 17 '12 at 21:32
chl ♦
33.9k 6 100 218

asked Nov 8 '11 at 16:52
pmgjones
2,307 3 22 29

add a comment

5 Answers

active oldest votes

24

The problem with covariances is that they are hard to compare: when you calculate the covariance of a set of heights and weights, as expressed in (respectively) meters and kilograms, you will get a different covariance from when you do it in other units (which already gives a problem for people doing the same thing with or without the metric system!), but also, it will be hard to tell if (e.g.) height and weight 'covariate better' than, e.g. the length of your toes and fingers, simply because the 'scale' you calculate the covariance on is different.

The solution to this is to 'normalize' the covariance: you divide the covariance by something that represents the diversity and scale in both the covariates, and end up with a value that is assured to be between -1 and 1: the correlation. Whatever unit your original variables were in, you will always get the same result, and this will also ensure that you can, to a certain degree, compare whether two variables 'correlate' more than two others, simply by comparing their correlation.

Note: the above starts from the idea that the listener has already been explained what a covariance is.

share improve this answer

answered Nov 8 '11 at 19:20
Nick Sabbe
7,134 17 26

+1 Did you mean to write "correlation" instead of "covariance" in the last sentence? – whuber ♦ Nov 8 '11 at 19:36

Are you sure you can't compare covariances with different units? The units pass through covariance multiplied - if your X is in `cm`, and your Y is in `s`, then your $cov(X, Y) = z\ cm \cdot s$. And then you can just multiply by the result by the unit conversion factor. Try it in R: `cov(cars$speed,cars$dist) == cov(cars$speed/5,cars$dist/7)*(7*5)` – naught101 May 18 '12 at 1:03

@naught101 I suspect the point is that, if I told you that $\mathrm{Cov}(X, Y) = 10^10$ and nothing else, you would have no clue whether X is highly predictive of Y or not, whereas if I told you $\mathrm{Cor}(X, Y) = .9$ you would have something a little more interpretable. – guy Dec 20 '13 at 5:46

@guy: That would be covariances *without* units :P I think the important thing is that you can't easily compare covariances from two data sets that have different variances. For example, if you have the relation B=2*A, and two datasets, {A1, B1} and {A2, B2}, where A1 has a variance of 0.5 and A2 has a variance of 2, then the $cov(A2, B2)$ will be much larger than $cov(A1, B1)$, even though the relationship is exactly the same. – naught101 Dec 23 '13 at 23:31

So in simple terms corelation > covariance – Karl Morrison Jul 29 at 21:43

add a comment

Linked

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

37 What is covariance in plain language?

Related

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

5 Create positive-definite 3x3 covariance matrix given specified correlation values

1 How to prove the correlation between two variables using covariance

2 In what situations is covariance preferred to correlation?

0 correlation between adjacent pixels in an image

2 Intuition on the definition of the covariance

2 Difference between identity and diagonal covariance matrices

0 Covariance between normalised correlation functions

3 sparse covariance/correlation thresholding

1 How is it possible to have a significant correlation between two variables but a low covariance?

Hot Network Questions

20

The requirements of these types of questions strike me as a bit bizarre. Here is a *mathematical* concept/formula, yet I want to talk about it in some context completely devoid of mathematical symbols. I also think it should be stated that the actual algebra necessary to understand the formulas, I would think, should be taught to most individuals before higher education (no understanding of matrix algebra is needed, just simple algebra will suffice).

So, at first instead of completely ignoring the formula and speaking of it in some magical and heuristic types of analogies, lets just look at the formula and try to explain the individual components in small steps. The difference in terms of covariance and correlation, when looking at the formulas, should become clear. Whereas speaking in terms of analogies and heuristics I suspect would obsfucate two relatively simple concepts and their differences in many situations.

So lets starts out with a formula for the **sample covariance** (these I have just taken and adopted from wikipedia);

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

To get everyone up to speed, lets explicitly define all of the elements and operations in the formula.

- x_i and y_i are each measurements of two seperate attributes of the same observation
- $\bar{x}$ and $\bar{y}$ are the means (or average) of each attribute
- For $\frac{1}{N-1}$, lets just say this means we divide the final result by $N-1$.
- $\sum_{i=1}^{n}$ may be a foreign symbol to some, so it would likely be useful to explain this operation. It is simply the sum of all i seperate observations, and n represents the total number of observations.

At this point, I might introduce a simple example, to put a face on the elements and operations so to speak. So for example, lets just make up a table, where each row corresponds to an observation (and x and y are labeled appropriately). One would likely make these examples more specific (e.g. say x represents age and y represents weight), but for our discussion here it should not matter.

```
x y
---
2 5
4 8
9 3
5 6
0 8
```

At this point if you feel the sum operation in the formula may not have been fully comprehended, you can introduce it again in a much simpler context. Say just present that $\sum_{i=1}^{n}(x_i)$ is the same as saying in this example;

```
  x
 --
  2
  4
  9
  5
+ 0
 --
 20
```

Now that mess should be cleared up, and we can work our way into the second part of the formula, $(x_i-\bar{x})(y_i-\bar{y})$. Now, assuming people already know what the mean, $\bar{x}$ and $\bar{y}$ stand for, and I would say, being hypocritical of my own comments earlier in the post, one can just refer to the mean in terms of simple heuristics (e.g. the middle of the distribution). One can then just take this process one operation at a time. The statement $(x_i-\bar{x})$ is just examining the deviations/distance between each observation, and the mean of all observations for that particular attribute. Hence when an observation is further from the mean, this operation will be given a higher value. One can then refer back to the example table given, and simply demonstrate the operation on the x vector of observations.

```
x x_bar (x - x_bar)
2 4     -2
4 4      0
9 4      5
5 4      1
0 4     -4
```

The operation is the same for y vector, but just for reinforcement you can present that operation as well.

How to fly to multiple destinations and luggages?
Sign of researcher's professional maturity
How do muggles communicate with their magical children when they're at Hogwarts?
Do you have to beat or equal a DC?
Why is "a road" incorrect in this test?
HTML password fields
Why can I see a font even if it is not installed
Why don't two binaries of programs with only comments changed exactly match in gcc?
Invalid foreign Key error
Why do both sine and cosine exist?
Unable to install Octave
What does “I don't suffer from insanity” mean?
How can I make my languages structurally less like English?
Is this cracking foundation under a jack post a problem that requires repair?
Can pulling with Thorn Whip generate an Opportunity Attack?
What's good colors for illustrating Up and Down?
what does the 'g' flag do in the javascript RegExp
We are #1 in our industry. However because we aren't a hot startup anymore some workers leave. How to manage this?
How to handle an insubordinate employee faction when I can't replace them?
What is the earliest instance of Spock being called "Dr. Spock"?
Which polynomial's roots are its coefficients?
Help plot an illustration of "neighborhood" around a point on the real line
Fast strlen with bit operations
Elements and atoms

```
y y_bar (y - y_bar)
5  6     -1
8  6      2
3  6     -3
6  6      0
8  6      2
```

Now, the terms $(x_i - \bar{x})$ and $(y_i - \bar{y})$ should not be ambiguous, and we can go onto the next operation, multiplying these results together, $(x_i - \bar{x}) \cdot (y_i - \bar{y})$. As gung points out in the comments, this is frequently called the cross product (perhaps a useful example to bring back up if one were introducing basic matrix algebra for statistics).

Take note of what happens when multiplying, if two observations are both a large distance above the mean, the resulting observation will have an even larger positive value (the same is true if both observations are a large distance below the mean, as multiplying two negatives equals a positive). Also note that if one observation is high above the mean and the other is well below the mean, the resulting value will be large (in absolute terms) and negative (as a positive times a negative equals a negative number). Finally note that when a value is very near the mean for either observation, multiplying the two values will result in a small number. Again we can just present this operation in a table.

```
(x - x_bar) (y - y_bar)  (x - x_bar)*(y - y_bar)
-2            -1               2
 0             2               0
 5            -3             -15
 1             0               0
-4             2              -8
```

Now if there are any statisticians in the room they should be boiling with anticipation at this point. We can see all the seperate elements of what a covariance is, and how it is calculated come into play. Now all we have to do is sum up the final result in the preceding table, divide by $N - 1$ and *voila*, the covariance should no longer be mystical (all with only defining one greek symbol).

```
(x - x_bar)*(y - y_bar)
-----------------------
   2
   0
 -15
   0
+ -8
-----
 -21

-21/(5-1) = -5.25
```

At this point you may want to reinforce where the 5 is coming from, but that should be as simple as referring back to the table and counting the number of observations (lets again leave the difference between sample and population to another time).

Now, the covariance in and of itself does not tell us much (it can, but it is needless at this point to go into any interesting examples without resorting to magically, undefined references to the audience). In a good case scenario, you won't really need to sell why we should care what the covariance is, in other circumstances, you may just have to hope your audience is captive and will take your word for it. But, continuing on to develop the difference between what the covariance is and what the correlation is, we can just refer back to the formula for correlation. To prevent greek symbol phobia maybe just say ρ is the common symbol used to represent correlation.

$$\rho = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}}$$

Again, to reiterate, the numerator in the preceding formula is simply the covariance as we have just defined, and the denominator is the square root of the product of the *variance* of each individual series. If you need to define the variance itself, you could just say that the variance is the same thing as the covariance of a series with itself (i.e. $Cov(x,x) = Var(x)$). And all the same concepts that you introduced with the covariance apply (i.e. if a series has many values a far ways from its mean, it will have a high variance). Maybe note here that a series can not have a negative variance as well (which should logically follow from the math previously presented).

So the only new components we have introduced are in the denominator, $Var(x)Var(y)$. So we are dividing the covariance we just calculated by the product of the variances of each series. One could go into the treatment about why dividing by $\sqrt{Var(x)Var(y)}$ will always result in a value between -1 and 1, but I suspect the Cauchy–Schwarz inequality should be left off of the agenda for this discussion. So again, I'm a hypocrite and resort to some, *take my word for it*, but at this point we can introduce all the reasons why we use the correlation coefficient. One can then relate these math lessons back to the heuristics that have been given in the other statements, such as Peter Flom's response to one of the other questions. While this was critisized for introducing the concept in terms of causal statements, that lesson should be on the agenda at some point as well.

I understand in some circumstances this level of treatment would not be appropriate. *The senate needs the executive summary*. In that case, well you can refer back to the simple heuristics that people have been using in other examples, but Rome wasn't built in a day. And to the senate whom asks for the executive summary, if you have so little time perhaps you should just take my word for it, and dispense with the formalities of analogies and bullet-points.

share improve this answer

edited May 18 '12 at 0:03

answered Nov 8 '11 at 20:22

Andy W

10.3k 2 27 96

3 I completely concur with the notion that the question is somehow outside the purpose of this forum. The definition of covariance as

$$\operatorname{cov}(X,Y)=\mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])]$$

is the clearest explanation one can propose. It only uses the notion of expectation. Avoiding the formula is leading to necessarily incomplete and potentially misleading versions. And this cannot provide the reader with the man to compute the covariance/correlation in a new situation. Not the best way to fight innumeracy. – X'an Nov 10 '11 at 5:59

3 +1, this is quite good. I would not be so critical of conceptual introductions, however. I've worked w/ people w/ enough math anxiety that showing a formula is likely to lose them. I usually get them up to speed w/ the intuition 1st, and then walk through the math simply & thoroughly (much as you do here) *afterward*. That way, they're just learning how the math represents what they already know, & if they do drop out mentally, they still learned the big ideas. As a tangential point, I work though the math in Excel, which I find very good for this. – gung May 17 '12 at 17:18

1 A couple of nitpicks (sorry): in your top equation, you divide by N, but then (correctly) discuss dividing by $N-1$ in the associated bullet point; I might note that $(x_i-\bar{x})(y_i-\bar{y})$ is called the "cross product"; since you've been talking about the *sample* covariance, when you get to correlation, I might skip the stuff about ρ and just use r; lastly, the correlation is calculated from the covariance by scaling it relative to the *SDs*, not the variances, see here, eg. – gung May 17 '12 at 17:28

Thanks @gung, I changed the typo in the first formula and then for the correlation I took the square root of the multiplied variances (instead of defining the standard deviation). On using rho versus another symbol, I don't feel too strongly either way. If I was teaching and had a text book, I would likely just want to conform with the text. Hopefully one more greek symbol doesn't cause chaos! – Andy W May 17 '12 at 17:58

1 If I could upvote your answer 100 times I would. What a terrifically lucid explanation! – Julian Mar 2 at 18:11

show **2** more comments

1

If you are familiar with the idea of centering and standardizing, x-xbar is to center x at its mean. Same applies to y. So covariance simply centers the data. Correlation, however, not only centers the data but also scales using the standard deviation (standardize). The multiplication and summation is the dot-product of the two vectors and it tells how parallel these two vectors compare to each other (the projection of one vector onto the other). The division of (n-1) or taking the expected value is to scale for the number of observations. Thoughts?

share improve this answer

answered Dec 19 '13 at 23:26

user31180

21 1

add a comment

0

As far as I've understood it. Corelation is a "normalized" version of the covariance.

share improve this answer

answered Jul 29 at 21:49

Karl Morrison

133 9

We're looking for long answers that provide some explanation and context. Don't just give a one-line answer; explain why your answer is right, ideally with citations. Answers that don't include explanations may be removed.

1 As many posts attest, "normalize" has many different meanings. Which one are you using? – whuber ♦ Jul 29 at 22:20

add a comment

-2

Correlation is scaled to be between -1 and +1 depending on whether there is positive or negative correlation, and is dimensionless. The covariance however, ranges from zero, in the case of two independent variables, to Var(X), in the case where the two sets of data are equal. The units of COV(X,Y) are the units of X times the units of Y.

share improve this answer

answered Mar 26 '12 at 16:42

Nagaraj

5 1

6 The covariance can be negative, so it is not bounded at 0. It is also unclear to me what you mean by your last sentence, `The units of COV(X,Y) are the units of X times the units of Y.`, care to elaborate? – Andy W Mar 26 '12 at 17:09

@AndyW isn't the units part clear from the definition? $\mathrm{Cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$. The expectation operator is just a weighted average of the values of X/Y, and the units pass through. – naught101 May 18 '12 at 0:56

1 @naught101, The units pass through? My initial comment to Nagaraj was to prompt further clarity, as ambiguous statements such as the one quoted I would assert are not helpful to anyone. So, why can't we interpret the covariance as "the units of x multiplied by the units of y", because that isn't what it is. A potentially more correct statement (for the sample covariance) would be it is the "*average of the products of the* **mean deviations**". cont... – Andy W May 18 '12 at 12:26

Now, mean deviations are certainly not the same as the original units, and the resultant statistic for the covariance is not simply dependent on the mean and variance of the original attributes. The covariance, in and of itself, tells you nothing without knowing the variance of the original attributes. – Andy W May 18 '12 at 12:26

add a comment

Your Answer

Sign up or log in

Sign up using Google

Sign up using Facebook

Sign up using Stack Exchange

Post as a guest

Name

Email

required, but never shown

Post Your Answer

By posting your answer, you agree to the privacy policy and terms of service.

Not the answer you're looking for? Browse other questions tagged correlation covariance or ask your own question.

question feed