Unit #3: Expectation, Variance, and Covariance

3.4.1, 3.4.2

Photo by Kimberly Farmer on Unsplash

Learning Objectives

At the end of this unit, students should be able to:

$\bar{X}$

1. Define the mean/expected value, variance, and standard deviation of X, where X is either a discrete or continuous random variable.

2. Compute the mean, variance, and standard deviation of a given random variable X (either where the pmf/pdf is given explicitly, or where X is a particular type of rv, e.g., binomial).

3. Compute the mean, variance, and standard deviation of a function of a random variable (i.e., $g(X)$).

4. Simplify the calculations in 3 when g is a linear function.

5. Define, compute, and interpret the covariance between two random variables X and Y.

6. Define,compute,and interpret the correlation between two random variables X and Y.

7. State and prove important properties about the mean (i.e., linearity), the variance, and covariance.

8. Compute expectations, variances, covariances directly and through simulation in R.

Consider a university having $15{,}000$ students and let X equal the number of courses for which a randomly selected student is registered.

The pdf of X is given to you as follows:

x	1	2	3	4	5	6	7
$p(x) = P(X = x)$	.01	.03	.13	.25	.39	.17	.02

Students pay more money when enrolled in more courses, and so the university wants to know what the average number of courses students take per semester.

$1, 2, \ldots, n$

$P(1) = P(2) = \ldots = \frac{1}{n}$

For a discrete random variable X with pmf $f(x) = P(X = x)$, the **expected value** or **mean value** of X is denoted as $E(X)$ and is defined as:

$$E(X) = \sum_{x} x P(X = x)$$

Consider a university having $15{,}000$ students and let X equal the number of courses for which a randomly selected student is registered.

The pdf of X is given to you as follows:

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02

What is $E(X)$?

$$E(X) = (1)(0.01) + (2)(0.03) + \dots + (7)(0.02)$$

$$= 4.57$$

Photo by Jan Huber on Unsplash

$f(x) = \frac{1}{b-a}$ on $[a, b]$

For a continuous random variable X with pdf $f(x)$, the **expected value** or **mean value** of X is defined as:

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

Photo by Mika Baumeister on Unsplash

Example: The lifetime (in years) of a certain brand of battery is exponentially distributed with $\lambda = 1/4$.

$u = x \qquad v = -e^{-\lambda x}$

$du = dx \qquad dv = \lambda e^{-\lambda x} dx$

$$\int u\,dv = uv - \int v\,du$$

$$= -xe^{-\lambda x} + \int e^{-\lambda x} dx$$

$$= -xe^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} + C$$

How long, on average, will the battery last?

$$E(X) = \int_0^\infty x \lambda e^{-\lambda x} dx = \lim_{b \to \infty} \left[-xe^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} \right]_0^b$$

$$= \lim_{b \to \infty} \left[-be^{-\lambda b} \underbrace{- \frac{1}{\lambda} e^{-\lambda b}}_{\to 0} + 0 + \frac{1}{\lambda} e^{-\lambda (0)} \right]$$

$$= \lim_{b \to \infty} -\frac{b}{e^{\lambda b}} + \frac{1}{\lambda} \overset{LHR}{=} \underbrace{\lim_{b \to \infty} \frac{-1}{\lambda e^{\lambda b}}}_{=0} + \frac{1}{\lambda}$$

$$= \frac{1}{\lambda}$$

If a *discrete* random variable X has a a pmf $P(X = x)$, then the expected value of any function $g(X)$, computed as:

$$E(g(X)) = \sum_x g(x) P(X=x)$$

Note that $E[g(X)]$ is computed in the same way that $E(X)$ itself is, except that $g(x)$ is substituted in place of x.

Photo by Jan Huber on Unsplash

If a continuous random variable X has a pdf $f(x)$, then the expected value of any function $g(X)$, computed as:

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x)\,dx$$

Note that $E[g(X)]$ is computed in the same way that $E(X)$ itself is, except that $g(X)$ is substituted in place of x.

Example: A random variable X has pdf:

$$f(x) = \begin{cases} \frac{3}{4}(1-x^2) & -1 \le x \le 1 \\ 0 & o/w \end{cases}$$

What is $E(X^3)$? $g(X) = X^3$

$$E(X^3) = \int_{-1}^{1} x^3 \left(\tfrac{3}{4}\right)(1-x^2)\,dx$$

$$= \int_{-1}^{1} \left(\frac{3}{4}x^3 - \frac{3}{4}x^5\right) dx = \dots = 0$$

If $g(X)$ is a *linear function* of X (*i.e.*, $g(X) = aX + b$) then $E[g(X)]$ can be easily computed from $E(X)$.

Theorem: Let a and b be real numbers and X a random variable with distribution function $f(x)$. Then:

$$E(aX+b) = aE(X) + b$$

Proof:

$$E(aX+b) = \int_{-\infty}^{\infty} (ax+b) f(x)\,dx = \int_{-\infty}^{\infty} [ax f(x) + b f(x)]\,dx$$

$$= \underbrace{a\int_{-\infty}^{\infty} x f(x)\,dx}_{E(X)} + \underbrace{b\int_{-\infty}^{\infty} f(x)\,dx}_{=1} = aE(X) + b$$

Note: This works for continuous and discrete random variables.

Photo by Nathan Dumlao on Unsplash

Consider a university having $15{,}000$ students and let X equal the number of courses for which a randomly selected student is registered.

The pdf of X is given to you as follows:

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02

$g(x) = 500x + 100$

Earlier, we calculated that $E(X)$ was 4.57. If students pay $\$500$ per course plus a $\$100$ per-semester registration fee, what is the average amount of money the university can expect a student to pay each a semester?

$E(g(x)) = E(500X + 100) = 500\,E(X) + 100$

$= \$2{,}385$

Photo by Jan Huber on Unsplash

$\mu = E(X)$

For a discrete random variable X with pmf $f(x)$, the **variance** of X is denoted as $Var(X)$ and is calculated as:

$$\sigma^2 = Var(X) = \sum_x (x-\mu)^2 P(X=x)$$

$$= E\left[(x-\mu)^2\right]$$

The standard deviation (SD) of X is

$$\sigma = sd(X) = \sqrt{Var(X)}$$

in R: var() $= s_x^2 \neq \sigma^2$

Photo by Nathan Dumlao on Unsplash

Consider a university having $15{,}000$ students and let X equal the number of courses for which a randomly selected student is registered.

The pdf of X is given to you as follows:

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02

Earlier, we calculated $\overset{\mu}{E(X)} = 4.57$. What is *Var(X)*? What about $sd(X)$?

$$Var(X) = (1-4.57)^2(0.01) + (2-4.57)^2(0.03) + \ldots + (7-4.57)^2(0.02)$$

$$\approx 1.27$$

$$sd(X) = \sqrt{Var(X)} \approx \sqrt{1.27}$$

Photo by Jan Huber on Unsplash

For a continuous random variable X with pdf $f(x)$, the **variance** of X is denoted as $Var(X)$ and is defined as:

The standard deviation (SD) of X is

Photo by Jan Huber on Unsplash

The variance can also be calculated using an alternative formula:

$$Var(X) = E(X^2) - [E(X)]^2$$

Proof?

$$Var(X) = E[(X - \overset{E(X)}{\mu})^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

$$= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x)\,dx$$

$$= \underbrace{\int_{-\infty}^{\infty} x^2 f(x)\,dx}_{E(X^2)} - 2\mu \underbrace{\int_{-\infty}^{\infty} x f(x)\,dx}_{\mu = E(X)} + \mu^2 \underbrace{\int_{-\infty}^{\infty} f(x)\,dx}_{=1}$$

Why would we use this equation instead?

$$= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \underbrace{\mu^2}_{[E(X)]^2} \quad \checkmark$$

The variance of $g(X)$ is calculated as follows:

Discrete: $Var(g(X)) = \sum_x (g(x) - \mu_g)^2 P(X = x)$

Continuous:

$$Var(g(X)) = \int_{-\infty}^{\infty} (g(x) - \mu_g)^2 f(x)\,dx$$


Photo by Jan Huber on Unsplash

As with the expected value, there is also a shortcut formula if $\underbrace{g(X)}_{=aX+b}$ is a linear function of X:

$$Var(g(X)) = Var(aX+b) = a^2 Var(X)$$

$$\boxed{\mu = E(X),\ \mu_g = E(aX+b) = a\mu + b}$$

Can we do a simple proof to show this is true?

$$Var(aX+b) = \int_{-\infty}^{\infty} (ax+b\ \underbrace{-a\mu-b}_{\mu_g})^2 f(x)\,dx$$

$$= \int_{-\infty}^{\infty} (ax-a\mu)^2 f(x)\,dx = a^2 \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

$$= a^2 Var(X) \checkmark$$


Photo by Nathan Dumlao on Unsplash

Consider a university having $15{,}000$ students and let X equal the number of courses for which a randomly selected student is registered. The pdf of X is given to you as follows:

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02

$g(X) = 500X + 100$

Earlier, we calculated $E(X) = 4.57$. If students pay $\$500$ per course plus a $\$100$ per-semester registration fee, what is the expected **standard deviation** of the amount of money students to pay each a semester?

$\cancel{E(g(X)) = 500(4.57) + 100 =}$

$Var(500X + 100) = 500^2 Var(X) \approx 315{,}000$

$sd(g(X)) \approx 561.25$

Photo by Eduardo Soares on Unsplash

Let $X \sim Bin(n, p)$. Then:

1. $E(X) = np$.

2. $Var(X) = np(1-p)$.

Proof?

1.) Let $X \sim Bin(1, p)$ $(n=1)$

$$E(X) = (0)P(X=0) + (1)P(X=1)$$
$$= 0 + (1)p^1(1-p)^0\binom{1}{1} = p$$

proof uses moment generating functions

Let $Y = \sum_{i=1}^{n} X_i$, $X_i \sim Bin(1,p)$. $Y \sim Bin(n,p)$

$$E(Y) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np$$

Photo by Eduardo Soares on Unsplash

Example: A biased coin is tossed 10 times, so that the $\underbrace{\text{odds of heads}}_{O_H}$ are $3:1$.

$$P_H = \frac{O_H}{1+O_H} = \frac{3}{1+3} = \frac{3}{4}$$

What notation do we use to describe X?

$$X \sim Bin\left(10, \frac{3}{4}\right)$$

What is the mean of X? The variance?

$$E(X) = np = 10\left(\frac{3}{4}\right) = \frac{15}{2}$$

$$Var(X) = np(1-p) = 10\left(\frac{3}{4}\right)\left(\frac{1}{4}\right) = \frac{15}{8}$$

Mean and Variance for Other Distributions

Distribution	E(X)	Var(X)
Geom(π)	$1/\pi$ $\pi = p = \text{prob. success}$	$(1-\pi)/\pi$
NB(r, π)	$r*\pi/(1-\pi)$	$r*\pi/(1-\pi)^2$
Poisson(λ)	λ	λ
Uniform(a,b)	$½(a+b)$	$1/12*(b-a)^2$
Exp(λ)	$1/\lambda$	$1/\lambda^2$
Weib(α,β)	$\beta\Gamma(1+1/\alpha)$	$\beta^2\{\Gamma(1+2/\alpha)-[\Gamma(1+1/\alpha)]^2\}$
Beta(α,β)	$\alpha/(\alpha+\beta)$	$\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$

Photo by Jan Huber on Unsplash

When two random variables X and Y are not independent, it is frequently of interest to assess how strongly they are related to one another.

$$\mu_X = E(X)$$

$$\mu_Y = E(Y)$$

The **covariance** between two random variables X and Y is defined as:

and ∩

$$\mathrm{Cov}(X,Y) = \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y)P(X=x, Y=y) \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f(x,y)\,dx\,dy \end{cases}$$

If both variables tend to deviate in the same direction (both go above their means or below their means at the same time), then the covariance will be positive.

If the opposite is true, the covariance will be negative.

If X and Y are not strongly related, the covariance will be near 0.

$$Var(X) = E(X^2) - [E(X)]^2$$

The following shortcut formula for $Cov(X, Y)$ simplifies the computations.

<u>Theorem</u>:

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

According to this formula, no intermediate subtractions are necessary. This is analogous to the shortcut for the variance computation we saw earlier.

Covariance

$Cov(X,Y) > 0$ $\quad$ $Cov(X,Y) < 0$ $\quad$ $Cov(X,Y) \approx 0$

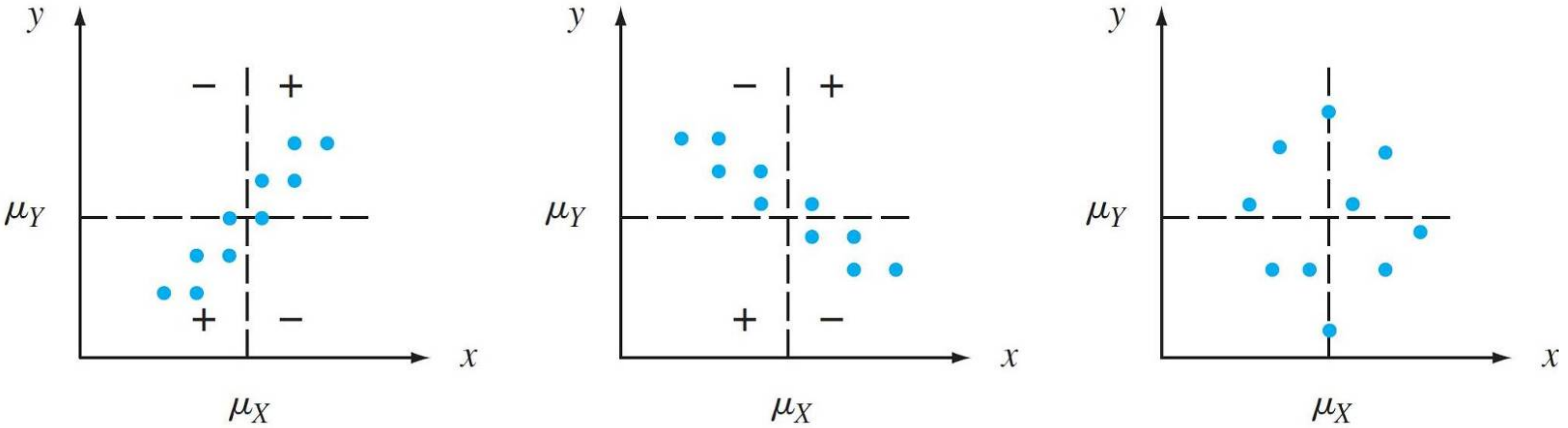


Photo by Jan Huber on Unsplash

Example: An insurance agency services customers who have both a homeowner's policy and an automobile policy. For each type of policy, a deductible amount must be specified. For an automobile policy, the choices are $\$100$ and $\$250$, whereas for a homeowner's policy, the choices are $\$0$, $\$100$, and $\$200$.

Suppose an individual—Bob—is selected at random from the agency's files. Let $X =$ his deductible amount on the auto policy and $Y =$ his deductible amount on the homeowner's policy.


Photo by Jan Huber on Unsplash

Suppose the joint pmf is given by the insurance company in the accompanying **joint probability table:**

$Cov(X,Y) = E(XY) - \mu_X \mu_Y = 1875$

$p(x, y) = P(X=x, Y=y)$		y (home): 0	100	200	$P(X=x)$
x (auto)	100	.20 +	.10 +	.20	= 0.5
	250	.05 +	.15 +	.30	= 0.5
$P(Y=y)$		0.25	0.25	0.5	

What is the covariance between X and Y?

$\mu_X = E(X) = (100)P(X=100) + (250)P(X=250)$

$= 175$

$\mu_Y = \ldots = 125$

$E(XY) = \sum_x \sum_y xy P(X=x, Y=y)$

$= (100)(100)(0.1) + (100)(200)(0.2) + (250)(100)(0.15)$

$+ (250)(200)(0.3) = ?$

The **correlation coefficient** of X and Y, denoted by $\rho_{X,Y}$ (or just ρ when the context is clear), is defined by:

$$\rho = \frac{Cov(X,Y)}{sd(X)\,sd(Y)}$$

It represents a “scaled” covariance – correlation ranges between -1 and 1.

Photo by Jan Huber on Unsplash

In the insurance example, what is the correlation between X and Y?

$p(x, y)$		y: 0	100	200
x	100	.20	.10	.20
	250	.05	.15	.30

- $\mathrm{Cov}(X,Y) = 1875$
- $sd(X) = 75$
- $sd(Y) = 83$

$$\rho = \frac{\mathrm{Cov}(X,Y)}{sd(X)\,sd(Y)} \approx 0.3$$

Photo by Jan Huber on Unsplash

Let $X_1, \ldots, X_n$ be random variables and let $a_1, \ldots, a_n$ be constants. Then:

1. $E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$

$\left(sgn(a_1 a_2) = \begin{cases} 1 & \text{if } a_1 a_2 > 0 \\ -1 & \text{if } a_1 a_2 < 0 \end{cases} \right)$

2. $Var\left(a_1 X_1 \pm a_2 X_2\right) =$

$a_1^2 Var(X_1) + a_2^2 Var(X_2) \pm \boxed{2} a_1 a_2 Cov(X_1, X_2)$

3. $Cov(a_1 X_1 + a_3, a_2 X_2 + a_4) =$

$a_1 a_2 Cov(X_1, X_2)$

4. $Corr(a_1 X_1 + a_3, a_2 X_2 + a_4) =$

$sgn(a_1 a_2) Corr(X_1, X_2)$

5. For any two random variables X and Y, $-1 \leq Corr(X, Y) \leq 1$

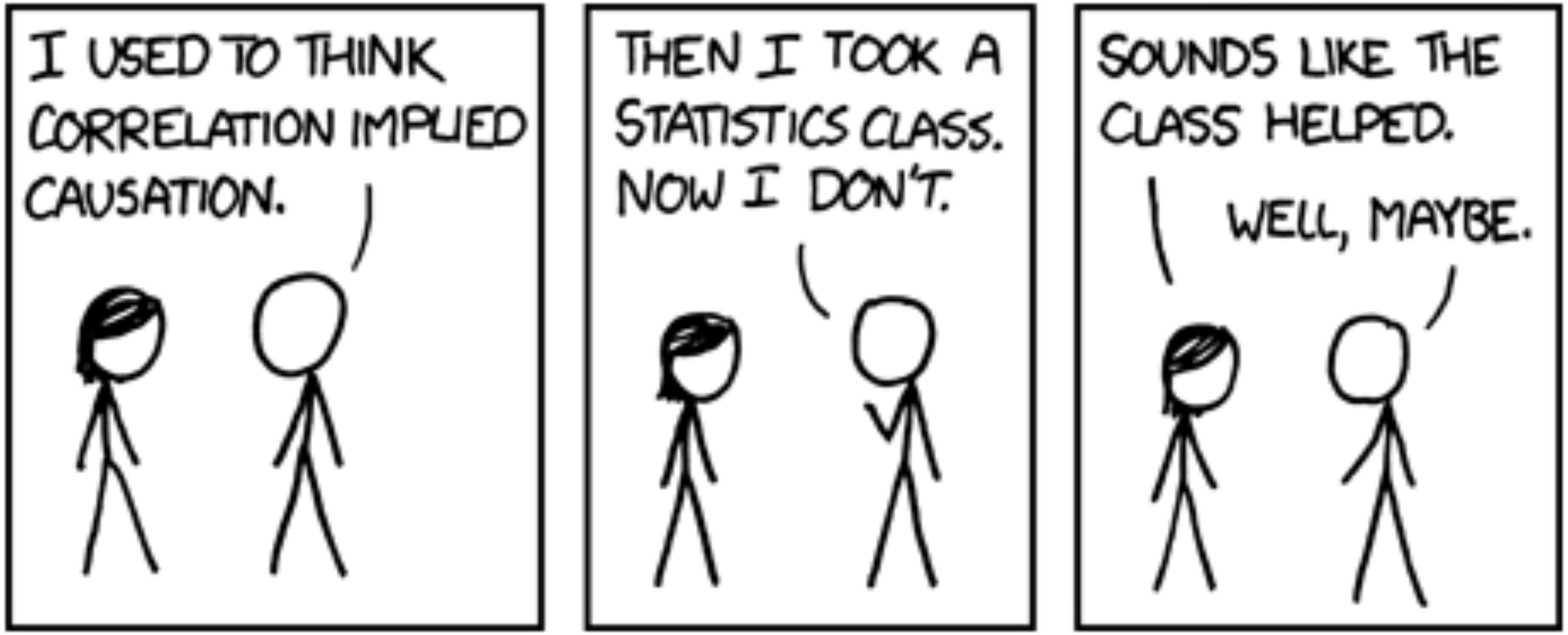
$Cov(X,Y) = 0 = \rho$

If X and Y are independent, then they are uncorrelated, but uncorrelated does not imply independence.

$\rho_{X,Y}$ is a measure of the **linear relationship** between X and Y.

Two variables can be uncorrelated yet highly dependent because there is a strong nonlinear relationship.

So be careful not to conclude too much from knowing that $\rho_{X,Y} \approx 0$.


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

xkcd.com/552/