15: Covariance

David Varodayan

February 10, 2020

Adapted from slides by Lisa Yan

CS109 roadmap

Multiple events:

intersection	conditional probability	independence
$P(E \cap F) = P(EF)$	$P(E\|F) = \frac{P(EF)}{P(F)}$	$P(EF) = P(E)P(F)$

Joint (Multivariate) distributions

joint PMF/PDF	conditional distributions	independent RVs
$p_{X,Y}(x,y)$ $f_{X,Y}(x,y)$	$p_{X\|Y}(x\|y)$ $f_{X\|Y}(x\|y)$	sum of independent RVs

Tracking in 2-D space

- You have a prior belief about the 2-D location of an object, (X, Y).
- You observe a noisy distance measurement, $D = 4$.
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

Recall Bayes terminology:

posterior belief

likelihood (of evidence)

prior belief

$$f_{X,Y|D}(x, y|d) = \frac{f_{D|X,Y}(d|x, y) f_{X,Y}(x, y)}{f_D(d)}$$

normalization constant

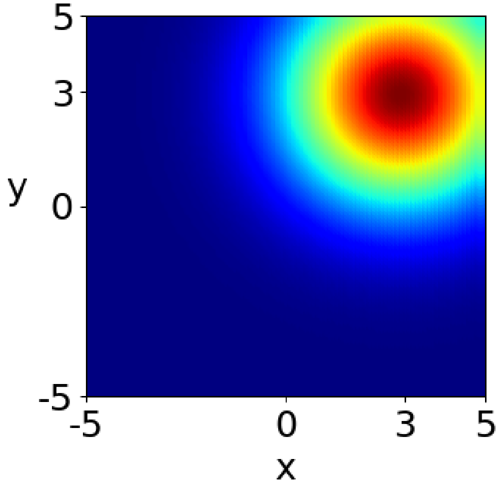
Tracking in 2-D space

- You have a prior belief about the 2-D location of an object, (X, Y).
- You observe a noisy distance measurement, $D = 4$.
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

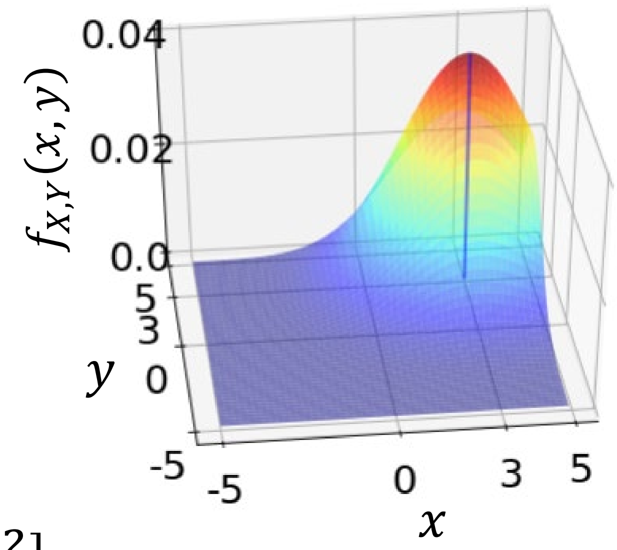
Let (X, Y) = object's 2-D location.
(your satellite is at (0,0)

Suppose the prior distribution is a symmetric bivariate normal distribution:

Top-down view



3-D view

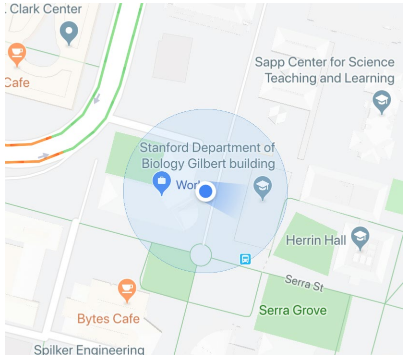


$$f_{X,Y}(x,y) = \frac{1}{2\pi 2^2} e^{-\frac{[(x-3)^2+(y-3)^2]}{2(2^2)}} = K_1 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}$$

normalizing constant

Tracking in 2-D space

- You have a prior belief about the 2-D location of an object, (X, Y).
- You observe a noisy distance measurement, $D = 4$.
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?



Let D = distance from the satellite (radially).
Suppose you knew your actual position: (x, y).

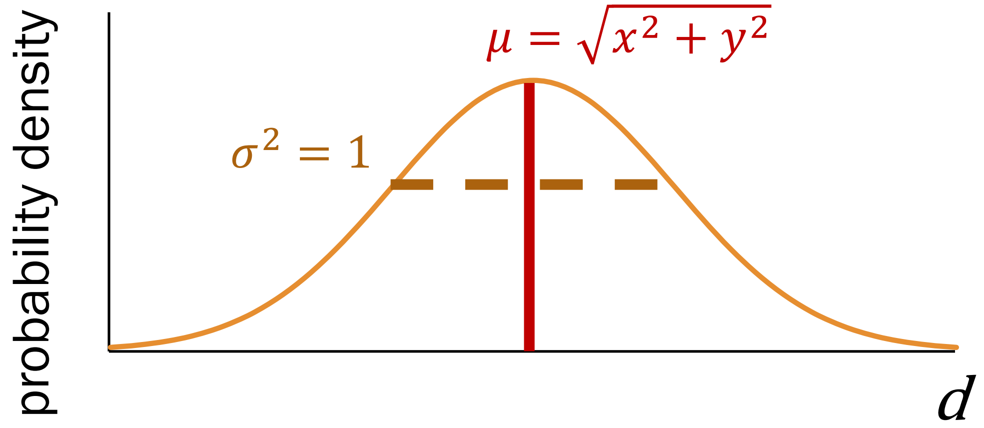
- D is still noisy! Suppose noise is unit variance: $\sigma^2 = 1$
- On average, D is your actual position: $\mu = \sqrt{x^2 + y^2}$

If you knew your actual location (x, y), you could say how likely a measurement $D = 4$ is!!

Tracking in 2-D space

- You have a prior belief about the 2-D location of an object, (X, Y).
- You observe a noisy distance measurement, $D = 4$.
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

If you knew your actual location (x, y), you could say how likely a measurement $D = 4$ is!!



If noise is normal: $$D|X,Y \sim N\left(\mu = \sqrt{x^2 + y^2}, \sigma^2 = 1\right)$$

Distance measurement of a ping is normal with respect to the true location.

Tracking in 2-D space

- You have a prior belief about the 2-D location of an object, (X, Y).
- You observe a noisy distance measurement, $D = 4$.
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

If you knew your actual location (x, y), you could say how likely a measurement $D = 4$ is!!

$$D|X,Y \sim \mathcal{N}\left(\mu = \sqrt{x^2 + y^2}, \sigma^2 = 1\right)$$

$$f_{D|X,Y}(D = d|X = x, Y = y) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(d-\mu)^2}{2\sigma^2}}$$

substitute μ and σ^2

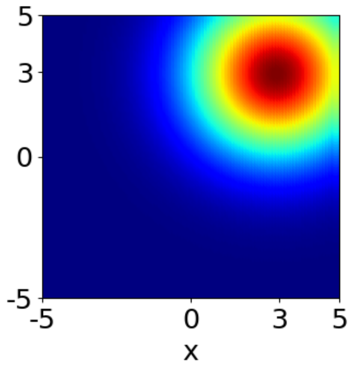
$$= \frac{1}{\sqrt{2\pi}} e^{\frac{-\left(d-\sqrt{x^2+y^2}\right)^2}{2}} = K_2 \cdot e^{\frac{-\left(d-\sqrt{x^2+y^2}\right)^2}{2}}$$

normalizing constant

Tracking in 2-D space

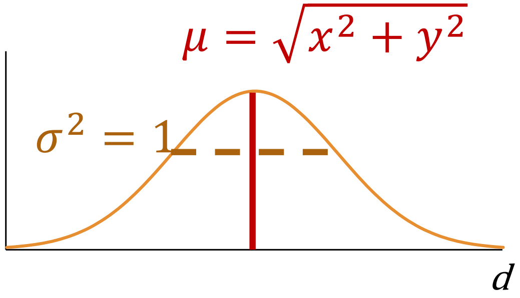
- You have a prior belief about the 2-D location of an object, (X, Y).
- You observe a noisy distance measurement, $D = 4$.
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

Prior belief



Top-down view

Observation likelihood



$$f_{X,Y}(x, y) = K_1 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}$$

$$f_{D|X,Y}(d|x, y) = K_2 \cdot e^{\frac{-\left(d-\sqrt{x^2+y^2}\right)^2}{2}}$$

Posterior belief

$$f_{X,Y|D}(x, y|4) = f_{X,Y|D}(X = x, Y = y|D = 4)$$

Tracking in 2-D space

What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

likelihood of $D = 4$ — prior belief

$$f_{X,Y|D}(X = x, Y = y|D = 4) = \frac{f_{D|X,Y}(D = 4|X = x, Y = y)f_{X,Y}(x, y)}{f(D = 4)}$$

Bayes' Theorem

$$= \frac{K_2 \cdot e^{-\frac{\left(4-\sqrt{x^2+y^2}\right)^2}{2}} \cdot K_1 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}}{f(D = 4)}$$

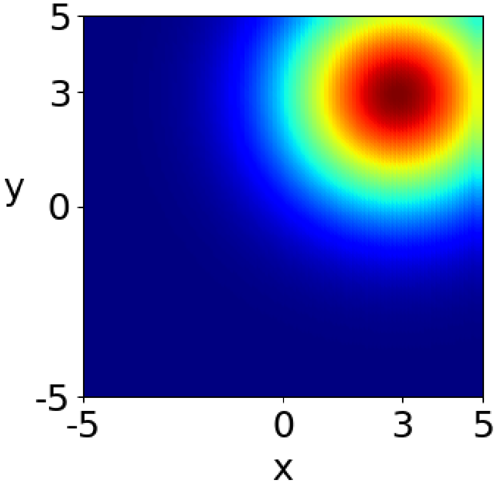
$$= \frac{K_3 \cdot e^{-\left[\frac{\left(4-\sqrt{x^2+y^2}\right)^2}{2} + \frac{[(x-3)^2+(y-3)^2]}{8}\right]}}{f(D = 4)}$$

$$= K_4 \cdot e^{-\left[\frac{\left(4-\sqrt{x^2+y^2}\right)^2}{2} + \frac{[(x-3)^2+(y-3)^2]}{8}\right]}$$

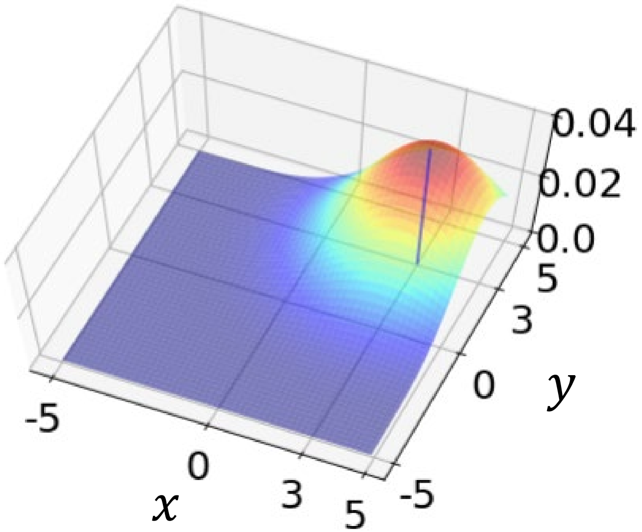
Tracking in 2-D space: Posterior belief

Prior belief

Top-down view



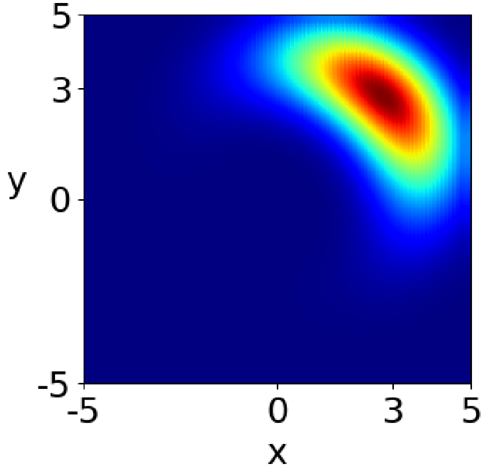
3-D view



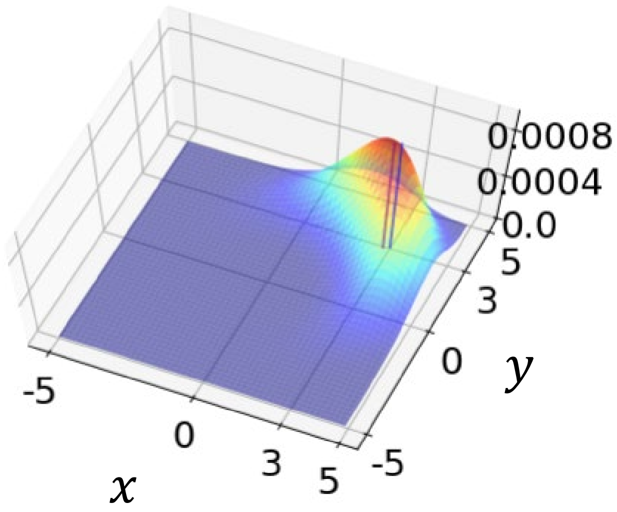
$$f_{X,Y}(x,y) = K_1 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}$$

Posterior belief

Top-down view



3-D view



$$f_{X,Y|D}(x,y|4) = K_4 \cdot e^{-\left[\frac{\left(4-\sqrt{x^2+y^2}\right)^2}{2} + \frac{[(x-3)^2+(y-3)^2]}{8}\right]}$$

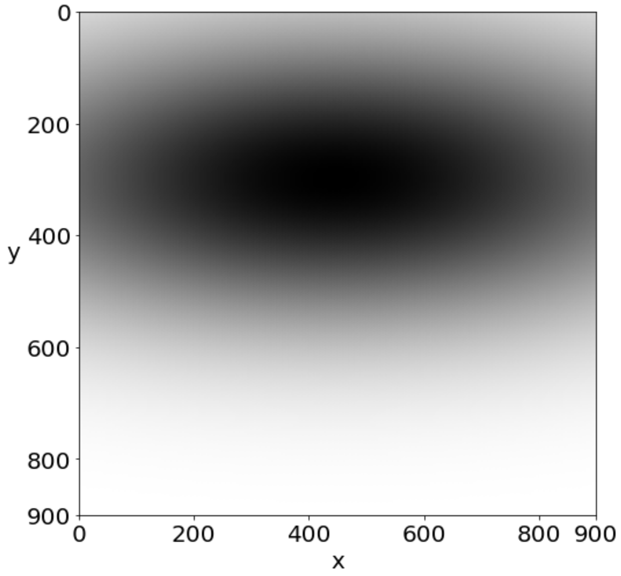
Today's plan

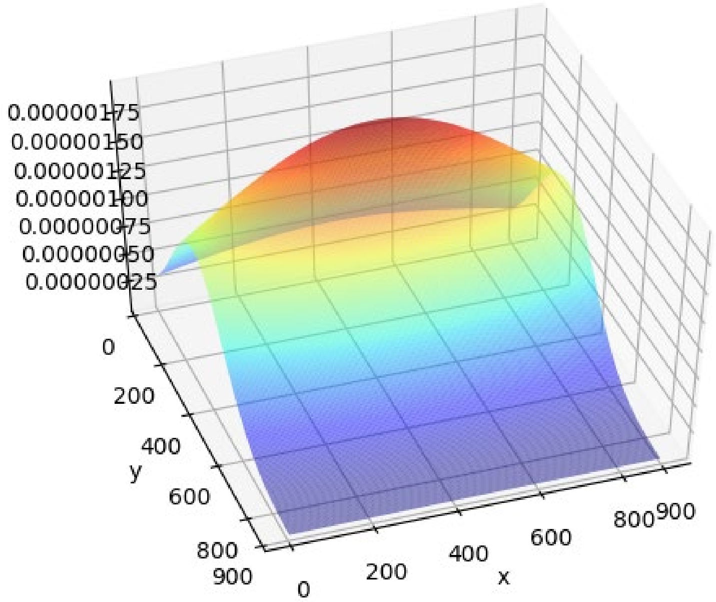
Covariance

Variance/covariance of independent RVs

Correlation

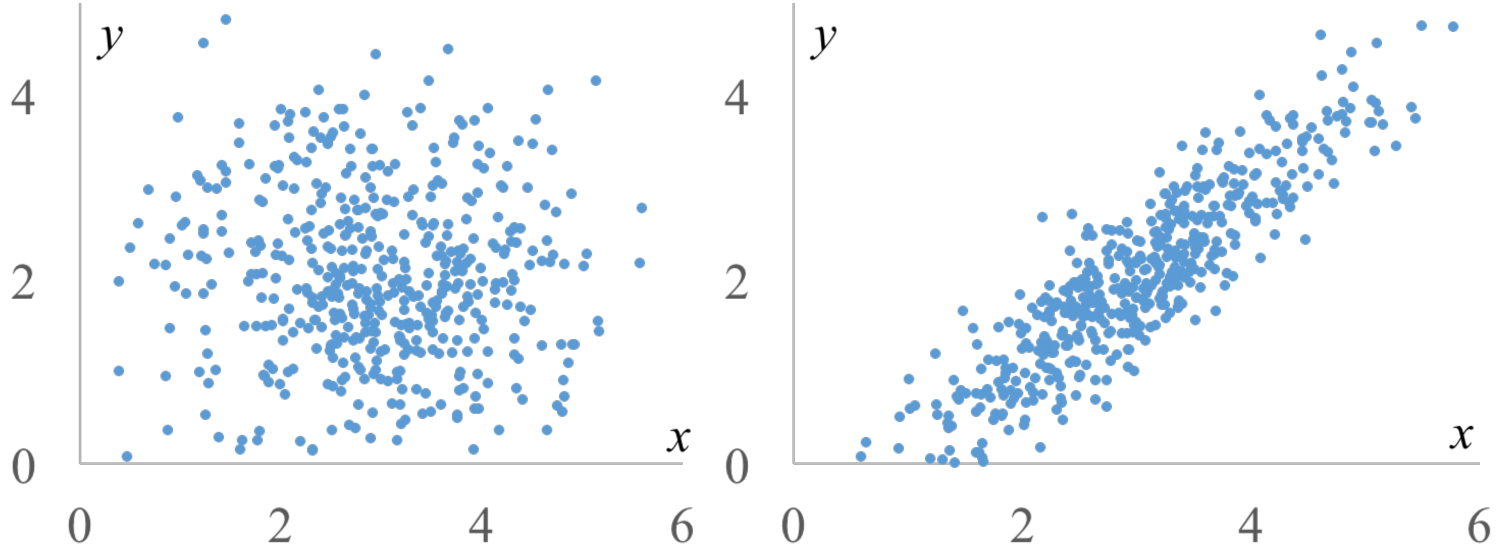
A word about today's diagrams:





Spot the difference

How do the following distributions of two variables differ?



In both distributions: $E[X] = E[Y], \quad \text{Var}(X) = \text{Var}(Y)$

Covariance

The **covariance** of two variables X and Y is:

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$
$$= E[XY] - E[X]E[Y]$$

Proof of second part:

$$\begin{aligned}\text{Cov}(X,Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E\big[XY - XE[Y] - E[X]Y + E[X]E[Y]\big] \\ &= E[XY] - E\big[XE[Y]\big] - E\big[E[X]Y\big] + E\big[E[X]E[Y]\big] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

(linearity of expectation)

($E[X]$, $E[Y]$ are scalars)

Covarying humans

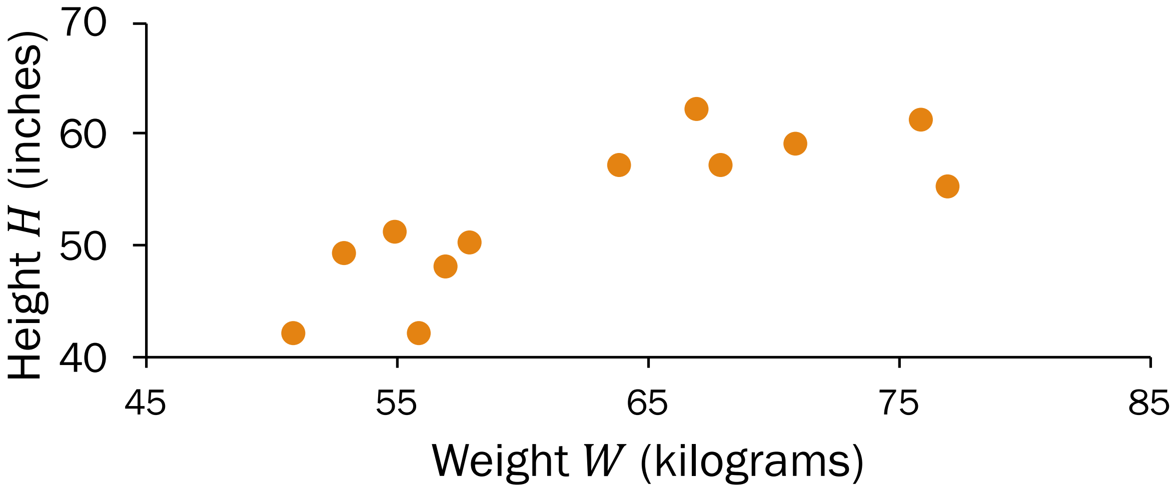
$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$
$= E[XY] - E[X]E[Y]$

Weight (kg)	Height (in)	W · H
64	57	3648
71	59	4189
53	49	2597
67	62	4154
55	51	2805
58	50	2900
77	55	4235
57	48	2736
56	42	2352
51	42	2142
76	61	4636
68	57	3876

$E[W] = 62.75$ $E[H] = 52.75$ $E[WH] = 3355.83$

What is the covariance of weight W and height H?

$$\begin{aligned} \text{Cov}(W, H) &= E[WH] - E[W]E[H] \\ &= 3355.83 - (62.75)(52.75) \\ &= 45.77 \quad \text{(positive)} \end{aligned}$$



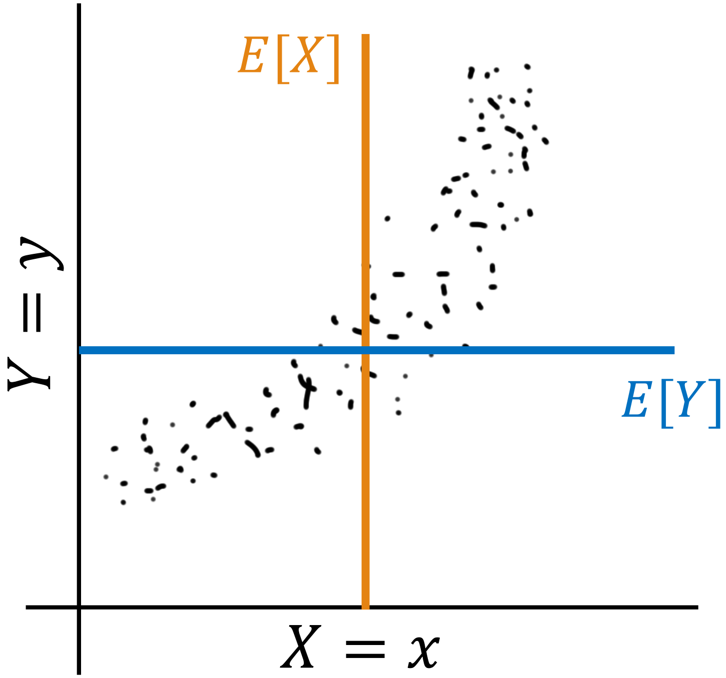
Positive covariance = as one variable increases, so does the second variable.

Covariance reps

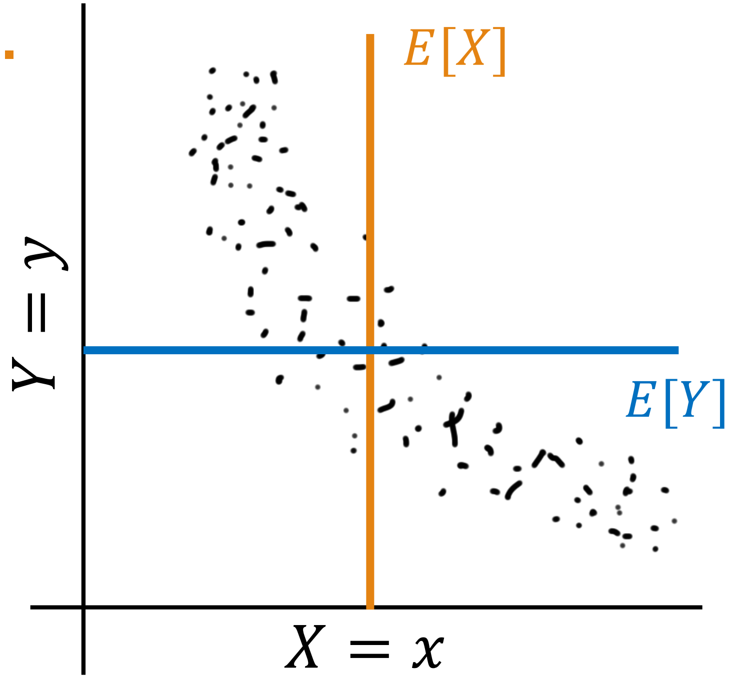
$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$
$$= E[XY] - E[X]E[Y]$$

Is the covariance positive, negative, or zero?

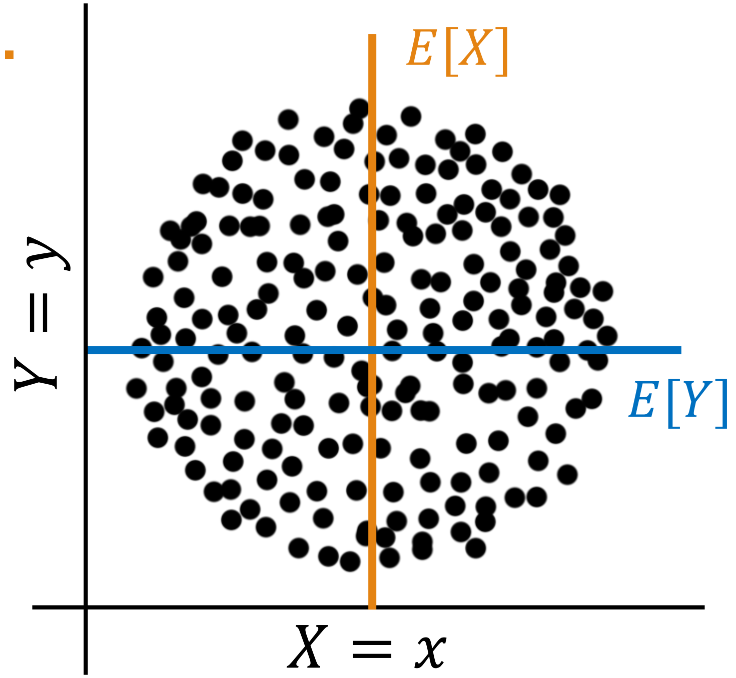
1.



2.



3.



Properties of Covariance

The covariance of two variables X and Y is:

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$
$$= E[XY] - E[X]E[Y]$$

True/False:

1. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
2. $\text{Cov}(X, X) = E[X \cdot X] - E[X]E[X] = \text{Var}(X)$
3. $\text{Cov}(aX + b, Y) = a\text{Cov}(X, Y)$
4. $\text{Cov}\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j \text{Cov}\left(X_i, Y_j\right)$

Announcements

Midquarter feedback (optional but appreciated)

Link posted in announcement on CS109 webpage

https://forms.gle/6JC6a4oyrH5hEGTy7

Closes: Wednesday February 12, 11:59pm

Today's plan

Covariance

Variance/covariance of sum of RVs

Correlation

Variance of sum of RVs

If X and Y are random variables, then

$$\text{Var}(X+Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X,Y) + \text{Var}(Y)$$

Proof:

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

$\text{Var}(X) = \text{Cov}(X,X)$

$$= \text{Cov}(X,X) + \text{Cov}(X,Y) + \text{Cov}(Y,X) + \text{Cov}(Y,Y)$$

covariance of all pairs

$$= \text{Var}(X) + 2 \cdot \text{Cov}(X,Y) + \text{Var}(Y)$$

Symmetry of covariance + $\text{Cov}(X,X) = \text{Var}(X)$

More generally:

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \text{Cov}\left(X_i, X_j\right)$$

(proof in extra slides)

Variance of sum of independent random variables

If X and Y are independent, then:

$$E[XY] = E[X]E[Y]$$

(proof in extra slides)

Therefore for independent X and Y :

$$\text{Cov}(X, Y) = 0$$
$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Proof of covariance:

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$ def. of covariance

$$= E[X]E[Y] - E[X]E[Y]$$ X and Y are independent

$$= 0$$

NOT bidirectional: $\text{Cov}(X, Y) = 0$ does NOT imply independence of X and Y!

Zero covariance does not imply independence

Let X take on values $\{-1,0,1\}$ with equal probability $1/3$.

Define $Y = \begin{cases} 1 & \text{if } X = 0 \\ 0 & \text{otherwise} \end{cases}$

What is the joint PMF of X and Y?

A.

Y \ X	-1	0	1
0	1/6	1/6	1/6
1	1/6	1/6	1/6

B.

Y \ X	-1	0	1
0	1/3	0	1/3
1	0	1/3	0

C.

Y \ X	-1	0	1
0	0	1/3	0
1	1/3	0	1/3

Zero covariance does not imply independence

Let X take on values $\{-1,0,1\}$ with equal probability $1/3$.

Define $Y = \begin{cases} 1 & \text{if } X = 0 \\ 0 & \text{otherwise} \end{cases}$

	X: -1	0	1	
Y: 0	1/3	0	1/3	2/3
1	0	1/3	0	1/3
	1/3	1/3	1/3	

Marginal PMF of Y, $p_Y(y)$

Marginal PMF of X, $p_X(x)$

1. $E[X] =$ $\qquad E[Y] =$
2. $E[XY] =$
3. $\text{Cov}(X, Y) =$
4. Are X and Y independent?

Variance of sum of independent random variables

If X and Y are independent, then:

$$E[XY] = E[X]E[Y]$$

(proof in extra slides)

Therefore for independent X and Y:

$$\text{Cov}(X, Y) = 0$$

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

Proof of variance:

$$\text{Var}(X + Y) = \text{Var}(X) + 2 \cdot \text{Cov}(X, Y) + \text{Var}(Y)$$ (proved earlier)

$$= \text{Cov}(X, X) + \text{Cov}(Y, Y)$$ X and Y are independent

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

1. Also not bidirectional
2. Does not apply to dependent X and Y

Variance of the Binomial

$X \sim \text{Bin}(n, p) \quad \text{Var}(X) = np(1-p)$

Let $\quad X = \sum_{i=1}^{n} X_i$

Let X_i = ith trial is heads

$X_i \sim \text{Ber}(p)$

$\text{Var}(X_i) = p(1-p)$

X_i are independent (by definition)

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{n} X_i\right)$$

$$= \sum_{i=1}^{n} \text{Var}(X_i)$$

X_i are independent, therefore variance of sum = sum of variance

$$= \sum_{i=1}^{n} p(1-p)$$

Variance of Bernoulli

$$= np(1-p)$$

Today's plan

Covariance

Variance/covariance of sum of independent RVs

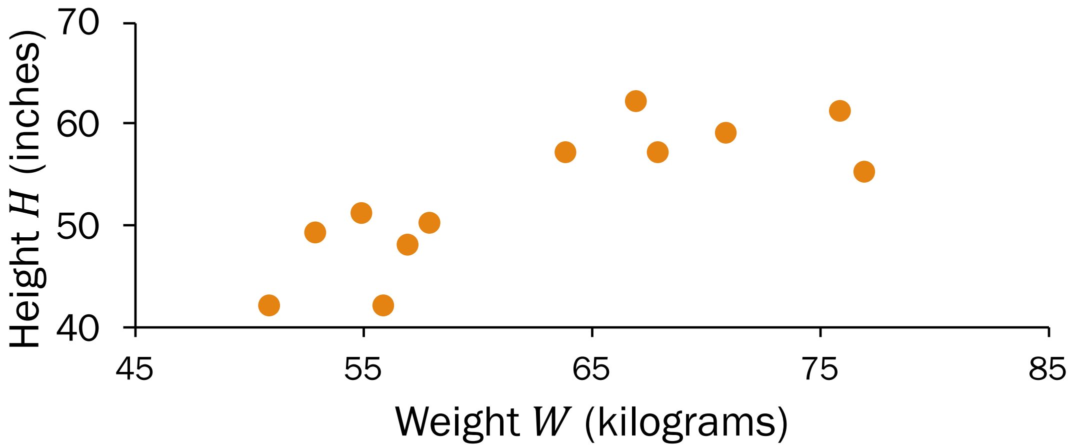
Correlation

Covarying humans

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$
$$= E[XY] - E[X]E[Y]$$

What is the covariance of weight W and height H?

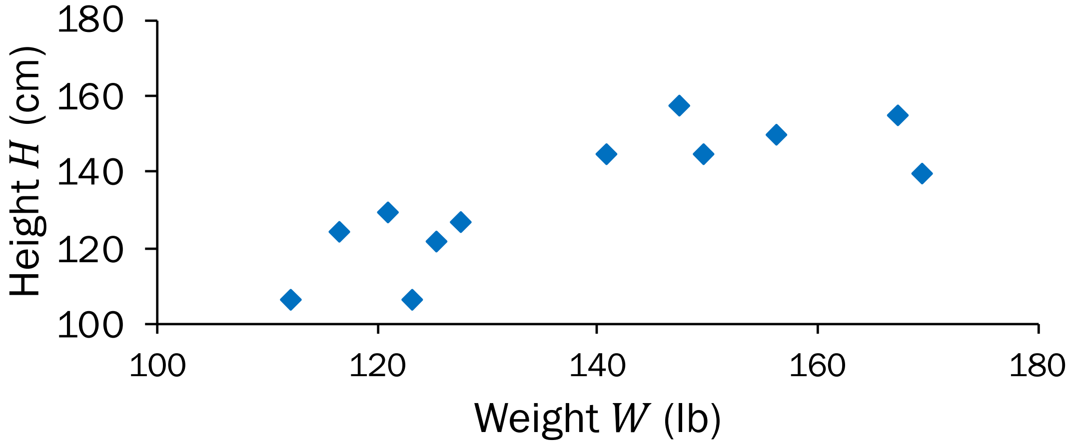
$$\text{Cov}(W,H) = E[WH] - E[W]E[H]$$
$$= 3355.83 - (62.75)(52.75)$$
$$= 45.77 \quad \text{(positive)}$$



What about weight (lb) and height (cm)?

$$\text{Cov}(2.20W, 2.54H)$$
$$= E[2.20W \cdot 2.54H] - E[2.20W]E[2.54H]$$
$$= 18752.38 - (138.05)(133.99)$$
$$= 255.06 \quad \text{(positive)}$$

Covariance depends on units!



For covariance, the sign (+/–) is more meaningful than the value.

Correlation

The correlation of two variables X and Y is:

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \, \sigma_Y}$$

$\sigma_X^2 = \text{Var}(X)$,
$\sigma_Y^2 = \text{Var}(Y)$

- Note: $-1 \leq \rho(X,Y) \leq 1$
- Correlation measures the linear relationship between X and Y:

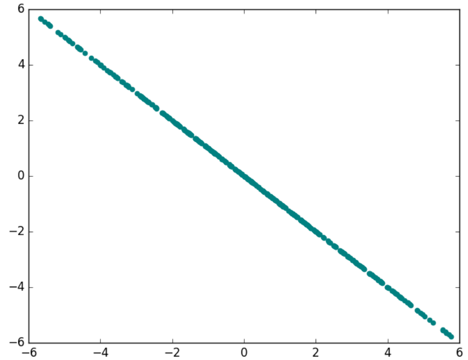
$\rho(X,Y) = 1 \quad \Rightarrow Y = aX + b$, where $a = \sigma_Y / \sigma_X$
$\rho(X,Y) = -1 \quad \Rightarrow Y = aX + b$, where $a = -\sigma_Y / \sigma_X$
$\rho(X,Y) = 0 \quad \Rightarrow$ "uncorrelated" (absence of linear relationship)

Correlation reps

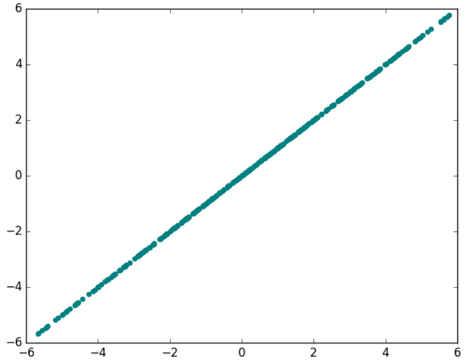
A. $\rho(X,Y) = 1$
B. $\rho(X,Y) = -1$
C. $\rho(X,Y) = 0$
D. Other

What is the correlation coefficient $\rho(X,Y)$?

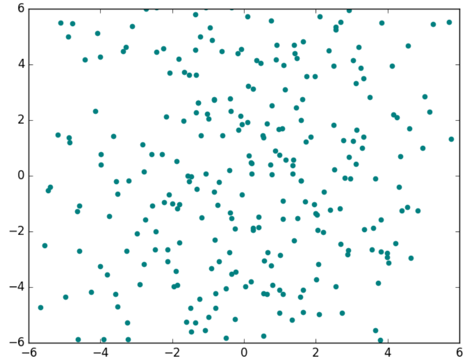
1.



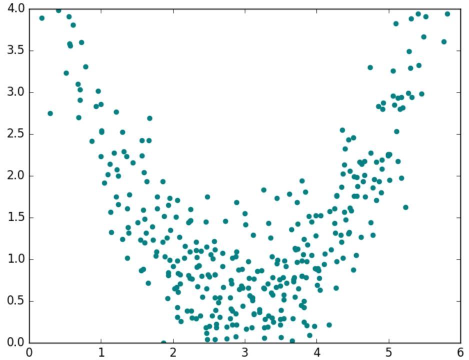
2.



3.



4.

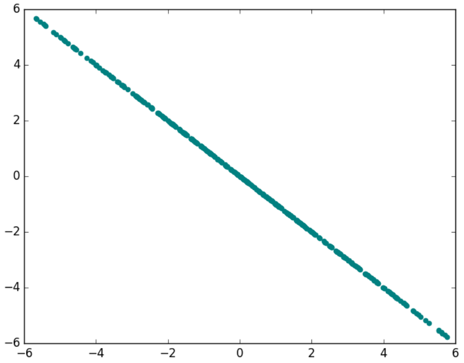


Correlation reps

A. $\rho(X, Y) = 1$
B. $\rho(X, Y) = -1$
C. $\rho(X, Y) = 0$
D. Other

What is the correlation coefficient $\rho(X, Y)$?

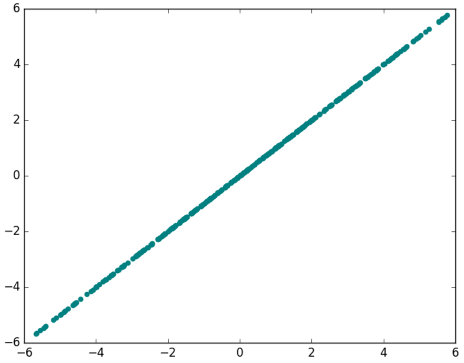
1.



B. $\rho(X, Y) = -1$

$$Y = -\frac{\sigma_Y}{\sigma_X} X + b$$

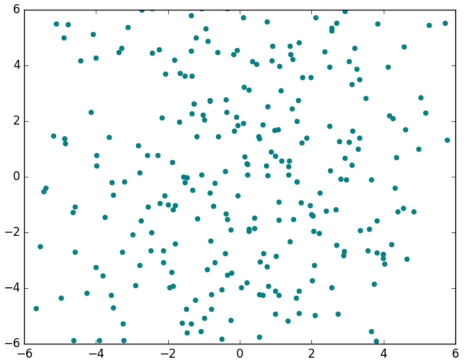
2.



A. $\rho(X, Y) = 1$

$$Y = \frac{\sigma_Y}{\sigma_X} X + b$$

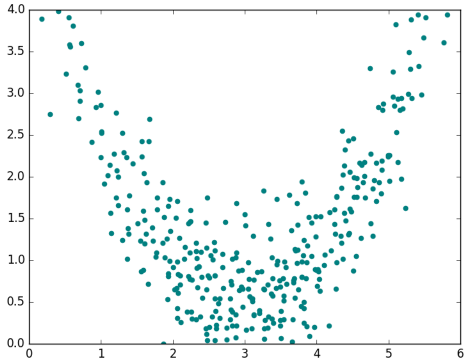
3.



C. $\rho(X, Y) = 0$

"uncorrelated"

4.



C. $\rho(X, Y) = 0$

$$Y = X^2$$

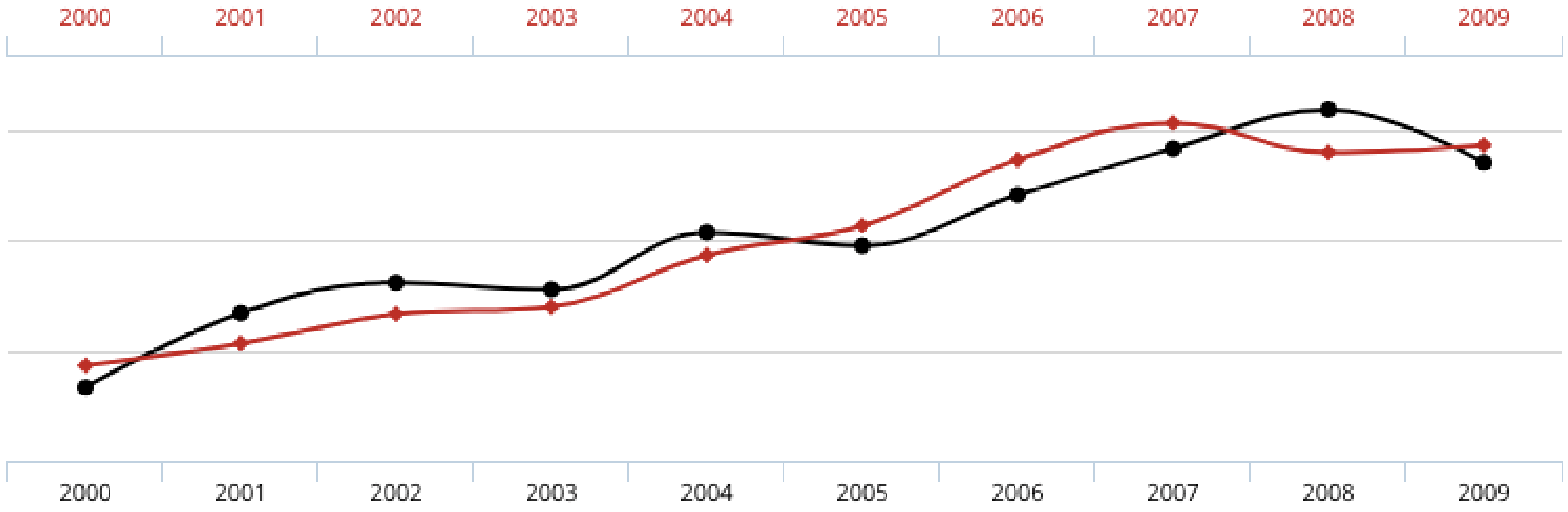
Correlation measures linearity.
X and Y can be nonlinearly related even if $\text{Cov}(X, Y) = 0$

Spurious Correlations

“Correlation does not imply causation”

$\rho(X, Y)$ is used a lot to statistically quantify the relationship b/t X and Y.

Correlation:
0.947091



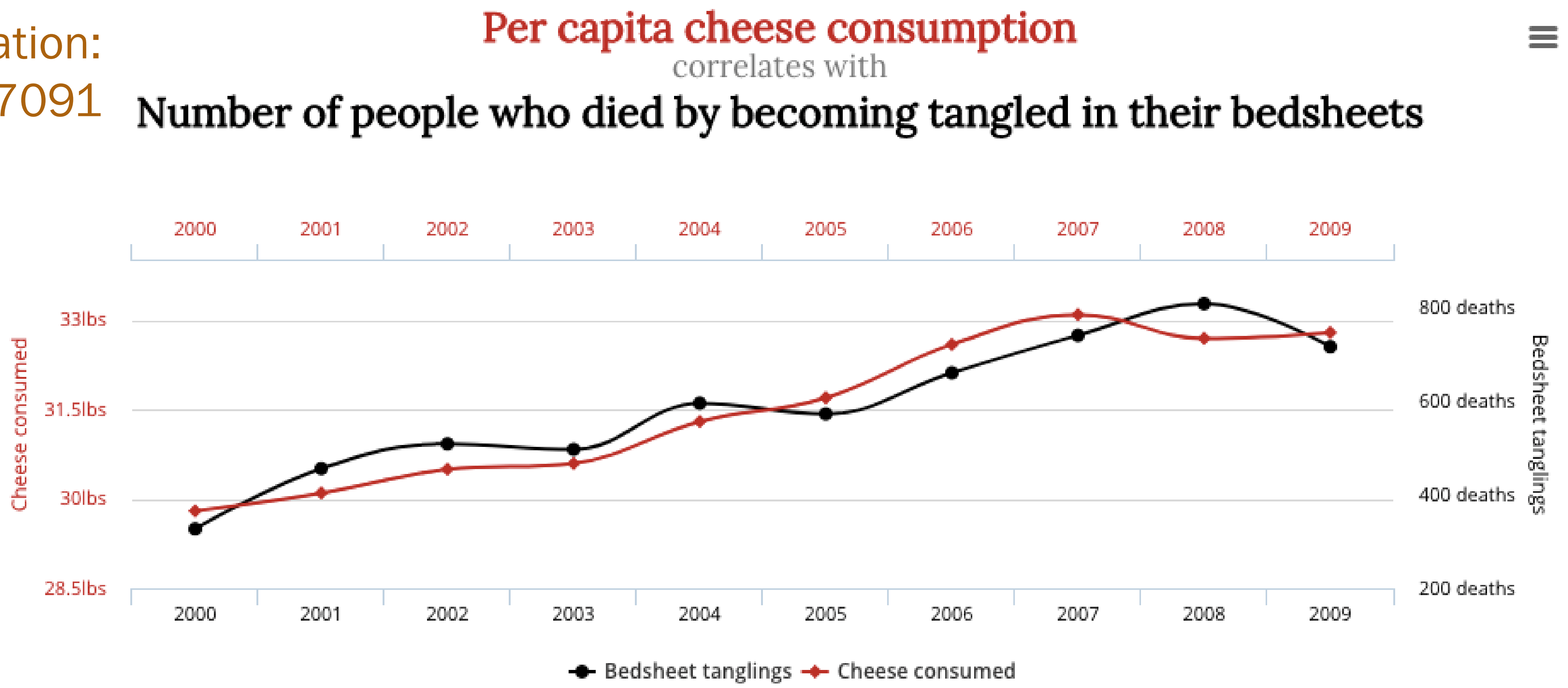
https://www.tylervigen.com/spurious-correlations

Spurious Correlations

"Correlation does not imply causation"

$\rho(X, Y)$ is used a lot to statistically quantify the relationship b/t X and Y.

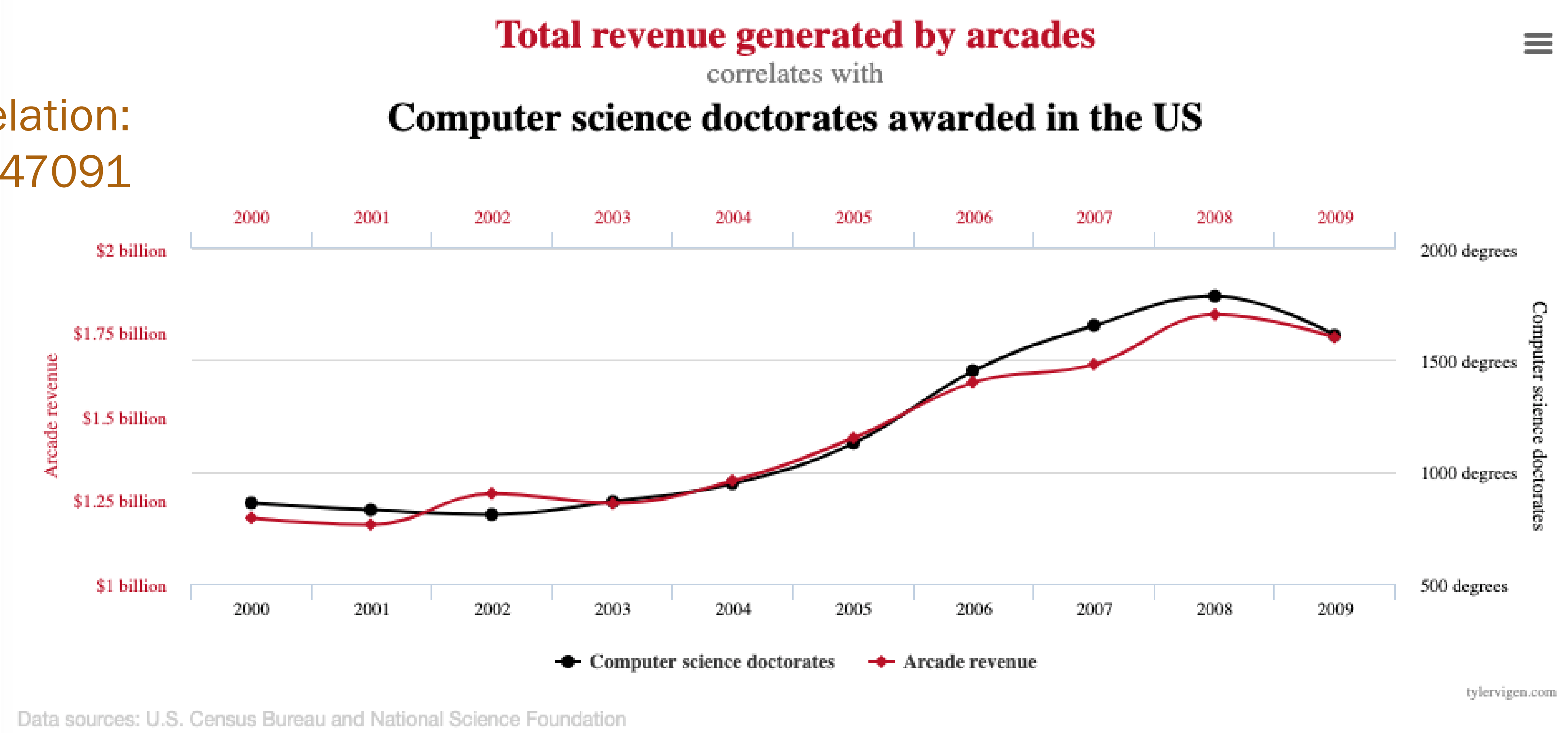
Correlation: 0.947091



Arcade revenue vs. CS PhDs

"Correlation does not imply causation"

Correlation: 0.947091



https://www.tylervigen.com/spurious-correlations

Extra slides

Expectation of a product of independent RVs

Variance of sums of variables

Expectation of product of independent RVs

If X and Y are independent, then:

$$E[XY] = E[X]E[Y]$$

More generally,

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Proof: $E[g(X)h(Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{X,Y}(x,y)dx\,dy$ (for discrete proof, replace integrals with summations)

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dx\,dy$ — X and Y are independent

$= \int_{-\infty}^{\infty} h(y)f_Y(y)dy \int_{-\infty}^{\infty} g(x)f_X(x)dx$ — Terms dependent on y are constant in integral of x

$= \left(\int_{-\infty}^{\infty} g(x)f_X(x)dx\right)\left(\int_{-\infty}^{\infty} h(y)f_Y(y)dy\right)$ — Integrals separate

$= E[g(X)]E[h(Y)]$

Variance of Sums of Variables

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=\text{i}+1}^{n} \text{Cov}\left(X_i, X_j\right)$$

For 2 variables: $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$

Proof:

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}\left(X_i, X_j\right)$$

$\text{Var}(X)$ $= \text{Cov}(X,X)$

covariance of all pairs

$$= \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \text{Cov}\left(X_i, X_j\right)$$

Symmetry of covariance
$\text{Cov}(X,X) = \text{Var}(X)$

$$= \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i=1}^{n}\sum_{j=\text{i}+1}^{n} \text{Cov}\left(X_i, X_j\right)$$

Adjust summation bounds