13: Statistics of Multiple RVs

Lisa Yan

May 4, 2020

Quick slide reference

Expectation of Common RVs

Linearity of Expectation is useful

Expectation is a linear mathematical operation. If $X = \sum_{i=1}^{n} X_i$:

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

- Even if you *don't know* the distribution of X (e.g., because the joint distribution of $(X_1, \dots, X_n)$ is unknown), you can still compute *expectation* of the sum!!

- Problem-solving key: Define X_i such that

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

Most common use cases:

- $E[X_i]$ easy to calculate
- Sum of dependent RVs

Expectations of common RVs: Binomial

Review

$X \sim \text{Bin}(n, p) \quad E[X] = np$

\# of successes in n independent trials with probability of success p

Recall: $\text{Bin}(1, p) = \text{Ber}(p)$

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

Let $X_i = i$th trial is heads
$X_i \sim \text{Ber}(p), E[X_i] = p$

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np$$

Expectations of common RVs: Negative Binomial

$Y \sim \text{NegBin}(r, p) \quad E[Y] = \dfrac{r}{p}$

# of independent trials with probability of success p until r successes

Recall: $\text{NegBin}(1, p) = \text{Geo}(p)$

$$Y = \sum_{i=1}^{?} Y_i$$

1. How should we define Y_i?
2. How many terms are in our summation?

Expectations of common RVs: Negative Binomial

$Y \sim \text{NegBin}(r, p) \quad E[Y] = \frac{r}{p}$

# of independent trials with probability of success p until r successes

Recall: $\text{NegBin}(1, p) = \text{Geo}(p)$

$$Y = \sum_{i=1}^{?} Y_i$$

Let Y_i = # trials to get ith success (after $(i-1)$th success)

$Y_i \sim \text{Geo}(p), E[Y_i] = \frac{1}{p}$

$$E[Y] = E\left[\sum_{i=1}^{r} Y_i\right] = \sum_{i=1}^{r} E[Y_i] = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$$

Coupon Collecting Problems

Linearity of Expectation is useful

Expectation is a linear mathematical operation. If $X = \sum_{i=1}^{n} X_i$:

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

- Even if you *don't know* the distribution of X (e.g., because the joint distribution of $(X_1, \ldots, X_n)$ is unknown), you can still compute *expectation* of the sum!!

- Problem-solving key:
  Define X_i such that

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

Most common use cases:

- $E[X_i]$ easy to calculate
- Sum of dependent RVs

Coupon collecting problems: Server requests

The coupon collector's problem in probability theory:

- You buy boxes of cereal.
- There are k different types of coupons
- For each box you buy, you "collect" a coupon of type i.

<u>Servers</u>

requests

k servers

request to server i

1. How many coupons do you expect after buying n boxes of cereal?

What is the expected number of utilized servers after n requests?

* 52% of Amazon profits

** more profitable than Amazon's North America commerce operations

[source]

Computer cluster utilization

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Consider a computer cluster with k servers. We send n requests.

- Requests independently go to server i with probability p_i
- Let X = # servers that receive ≥ 1 request.

What is $E[X]$?

Computer cluster utilization

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Consider a computer cluster with k servers. We send n requests.

- Requests independently go to server i with probability p_i
- Let X = # servers that receive ≥ 1 request.

What is $E[X]$?

1. Define additional random variables.

Let: A_i = event that server i receives ≥ 1 request

X_i = indicator for A_i

$$P(A_i) = 1 - P(\text{no requests to } i)$$
$$= 1 - (1 - p_i)^n$$

Note: A_i are dependent!

2. Solve.

$$E[X_i] = P(A_i) = 1 - (1 - p_i)^n$$

$$E[X] = E\left[\sum_{i=1}^{k} X_i\right] = \sum_{i=1}^{k} E[X_i] = \sum_{i=1}^{k} \left(1 - (1 - p_i)^n\right)$$

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

Coupon collecting problems: Hash tables

The **coupon collector's problem** in probability theory:

- You buy boxes of cereal.
- There are k different types of coupons
- For each box you buy, you "collect" a coupon of type i.

Servers	Hash Tables
requests	strings
k servers	k buckets
request to server i	hashed to bucket i

1. How many coupons do you expect after buying n boxes of cereal? → What is the expected number of utilized servers after n requests?

2. How many boxes do you expect to buy until you have one of each coupon? → What is the expected number of strings to hash until each bucket has ≥ 1 string?

Stay tuned for live lecture!

Covariance

Statistics of sums of RVs

For any random variables X and Y,

$$E[X+Y] = E[X] + E[Y]$$

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

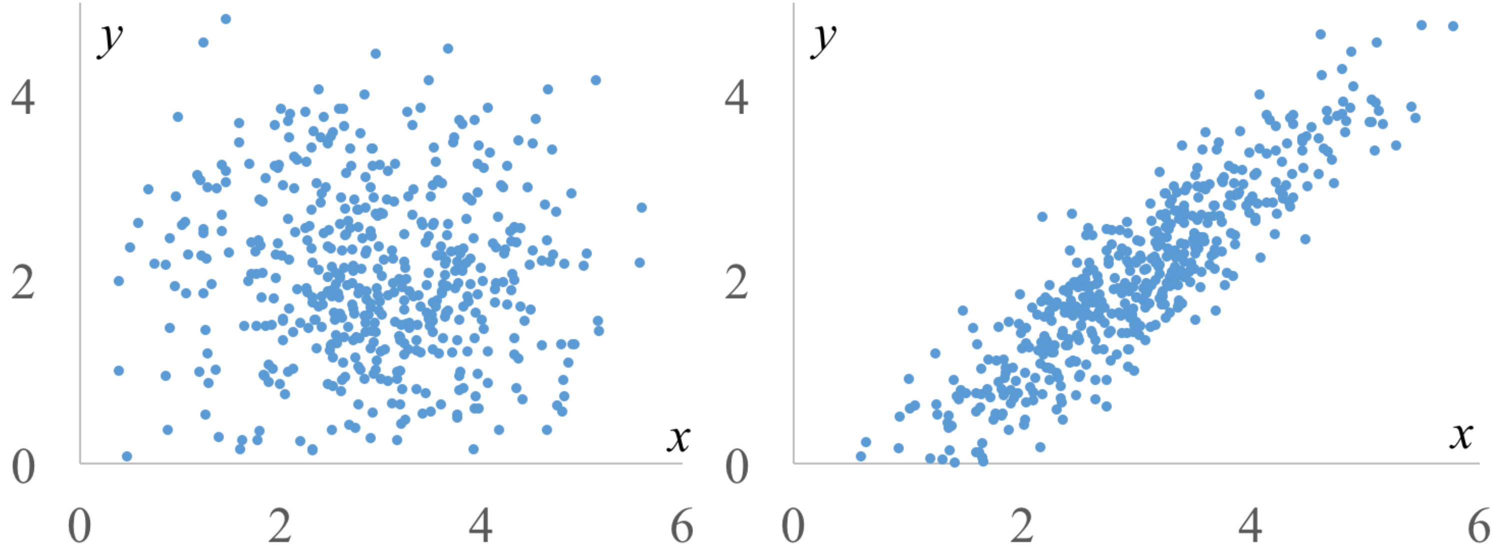
But first...
a new statistic!

Spot the difference

Compare/contrast the following two distributions:

Assume all points are equally likely.

$$P(X = x, Y = y) = \frac{1}{N}$$



Both distributions have the same $E[X]$, $E[Y]$, Var(X), and Var(Y)

Difference: how the two variables vary with ***each other***.

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

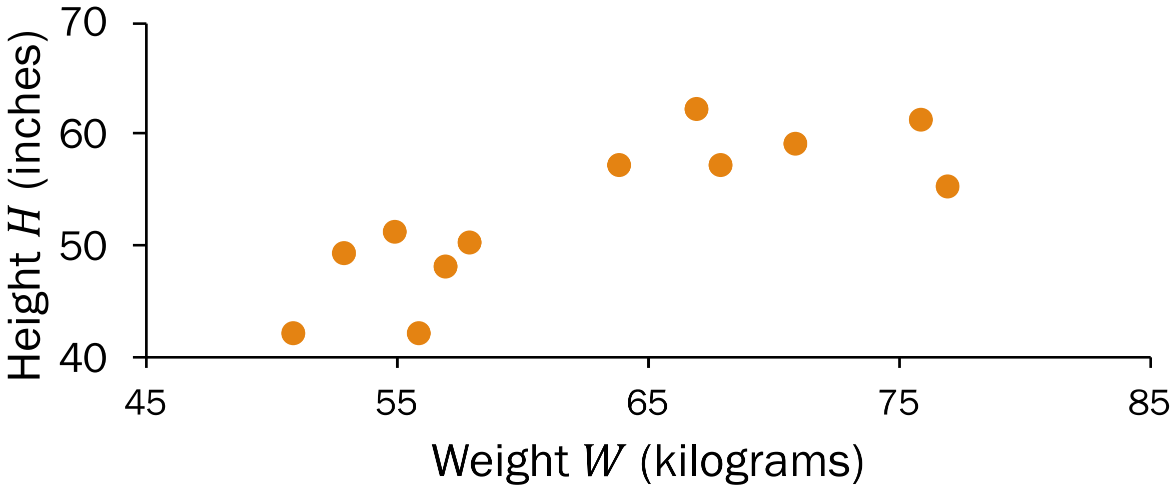
$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])] \\ = E[XY] - E[X]E[Y]$$

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

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

What is the covariance of weight W and height H?

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



Covariance > 0: one variable ↑, other variable ↑

Properties of Covariance

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

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

Properties:

1. $\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \mathrm{Cov}(X,X)$
2. Symmetry
3. Non-linearity
4. Covariance of sums

(to be discussed in live lecture)

Variance of sums of RVs

Statistics of sums of RVs

For any random variables X and Y,

$$E[X+Y] = E[X] + E[Y]$$

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

Variance of general sum of RVs

For any random variables X and Y,

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

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)

Statistics of sums of RVs

For any random variables X and Y,

$$E[X+Y] = E[X] + E[Y]$$

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

For **independent** X and Y,

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

(Lemma: proof in extra slides)

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

Variance of sum of independent RVs

For **independent** X and Y,

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

Proof:

1. $\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$ def. of covariance
   $= E[X]E[Y] - E[X]E[Y]$ X and Y are independent
   $= 0$

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

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

Proving Variance of the Binomial

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

To simplify the algebra a bit, let $q = 1 - p$, so $p + q = 1$.

So:

$$
\begin{aligned}
\mathrm{E}\left(X^2\right) &= \sum_{k \ge 0}^{n} k^2 \binom{n}{k} p^k q^{n-k} && \text{Definition of Binomial Distribution: } p + q = 1 \\
&= \sum_{k=0}^{n} kn \binom{n-1}{k-1} p^k q^{n-k} && \text{Factors of Binomial Coefficient: } k\binom{n}{k} = n\binom{n-1}{k-1} \\
&= np \sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} && \text{Change of limit: term is zero when } k - 1 = 0 \\
&= np \sum_{j=0}^{m} (j+1) \binom{m}{j} p^j q^{m-j} && \text{putting } j = k - 1, m = n - 1 \\
&= np\left(\sum_{j=0}^{m} j \binom{m}{j} p^j q^{m-j} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j}\right) && \text{splitting sum up into two} \\
&= np\left(\sum_{j=0}^{m} m \binom{m-1}{j-1} p^j q^{m-j} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j}\right) && \text{Factors of Binomial Coefficient: } j\binom{m}{j} = m\binom{m-1}{j-1} \\
&= np\left((n-1)p \sum_{j=1}^{m} \binom{m-1}{j-1} p^{j-1} q^{(m-1)-(j-1)} + \sum_{j=0}^{m} \binom{m}{j} p^j q^{m-j}\right) && \text{Change of limit: term is zero when } j - 1 = 0 \\
&= np\left((n-1)p(p+q)^{m-1} + (p+q)^m\right) && \text{Binomial Theorem} \\
&= np((n-1)p + 1) && \text{as } p + q = 1 \\
&= n^2p^2 + np(1-p) && \text{by algebra}
\end{aligned}
$$

Then:

$$
\begin{aligned}
\mathrm{var}(X) &= \mathrm{E}\left(X^2\right) - (\mathrm{E}(X))^2 \\
&= np(1-p) + n^2p^2 - (np)^2 && \text{Expectation of Binomial Distribution: } \mathrm{E}(X) = np \\
&= np(1-p)
\end{aligned}
$$

as required.

proofwiki.org

Let's instead prove this using independence and variance!

Proving Variance of the Binomial

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

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

$$= np(1-p)$$

(live)

13: Statistics of Multiple RVs

Slides by Lisa Yan
July 20, 2020

Where are we now? A roadmap of CS109

Last week: Joint distributions

$p_{X,Y}(x, y)$

Today: Statistics of multiple RVs!

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

$E[X + Y]$

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

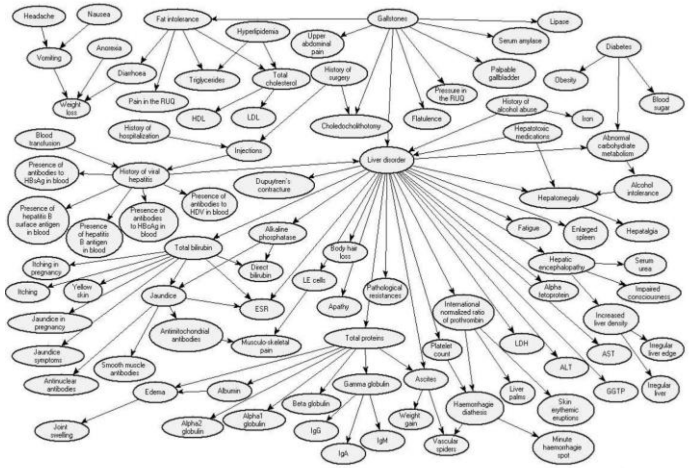
$\rho(X, Y)$

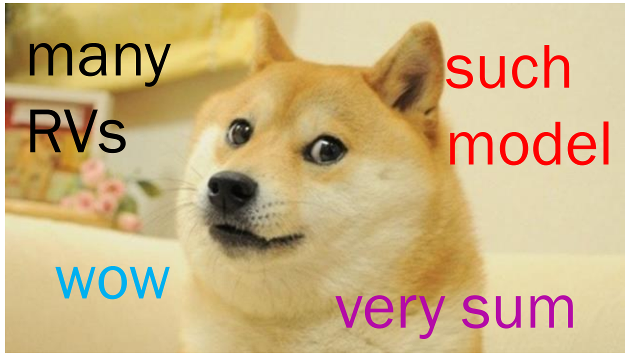
Wednesday: Conditional distributions

$p_{X|Y}(x|y)$

$E[X|Y]$

Also Wednesday: Modeling with Bayesian Networks





Don't we already know linearity of expectation?

Review

Expectation is a linear mathematical operation. If $X = \sum_{i=1}^{n} X_i$:

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

We covered this back in Lecture 6 (when we first learned expectation)!

- Proved binomial: sum of 1s or 0s
- Hat check (section): sum of 1s or 0s
- We ignored (in)dependence of events.

Why are we learning this again???

- Now we can prove it!
- We can now ignore (in)dependence of random variables.
- Our approach is still the same!

Coupon collecting problems: Hash tables

The **coupon collector's problem** in probability theory:

- You buy boxes of cereal.
- There are k different types of coupons
- For each box you buy, you "collect" a coupon of type i.

Servers	Hash Tables
requests	strings
k servers	k buckets
request to server i	hashed to bucket i

1. How many coupons do you expect after buying n boxes of cereal? → What is the expected number of utilized servers after n requests?

2. How many boxes do you expect to buy until you have one of each coupon? → What is the expected number of strings to hash until each bucket has ≥ 1 string?

Breakout Rooms

Check out the properties on the next slide.
Post any clarifications here!

https://us.edstem.org/courses/667/discussion/93095

Breakout rooms: 4 min. Introduce yourself!

Hash Tables

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Consider a hash table with k buckets.

- Strings are equally likely to get hashed into any bucket (independently).
- Let Y = # strings to hash until each bucket ≥ 1 string.

What is $E[Y]$?

1. Define additional random variables.

   How should we define Y_i such that $Y = \sum_i Y_i$?

2. Solve.

Hash Tables

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Consider a hash table with k buckets.

- Strings are equally likely to get hashed into any bucket (independently).
- Let Y = # strings to hash until each bucket ≥ 1 string.

What is $E[Y]$?

1. Define additional random variables.

Let: Y_i = # of trials to get success after i-th success

- Success: hash string to previously empty bucket
- If i non-empty buckets: $P(\text{success}) = \frac{k-i}{k}$

2. Solve.

$$P(Y_i = n) = \left(\frac{i}{k}\right)^{n-1}\left(\frac{k-i}{k}\right)$$

Equivalently, $Y_i \sim \text{Geo}\left(p = \frac{k-i}{k}\right)$ $\quad E[Y_i] = \frac{1}{p} = \frac{k}{k-i}$

Hash Tables

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

Consider a hash table with k buckets.

- Strings are equally likely to get hashed into any bucket (independently).
- Let Y = # strings to hash until each bucket $\geq$ 1 string.

What is $E[Y]$?

1. Define additional random variables.

Let: Y_i = # of trials to get success after i-th success

$$Y_i \sim \text{Geo}\left(p = \frac{k-i}{k}\right), \qquad E[Y_i] = \frac{1}{p} = \frac{k}{k-i}$$

2. Solve.

$$Y = Y_0 + Y_1 + \cdots + Y_{k-1}$$

$$E[Y] = E[Y_0] + E[Y_k] + \cdots + E[Y_{k-1}]$$

Even if Y_i dependent, it wouldn't affect expectation!

$$= \frac{k}{k} + \frac{k}{k-1} + \frac{k}{k-2} + \cdots + \frac{k}{1} = k\left[\frac{1}{k} + \frac{1}{k-1} + \cdots + 1\right] = O(k \log k)$$

Covariance

The covariance of two variables X and Y is:

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

Think

The next slide has a question to go over by yourself.

Post any clarifications here!

https://us.edstem.org/courses/667/discussion/93095

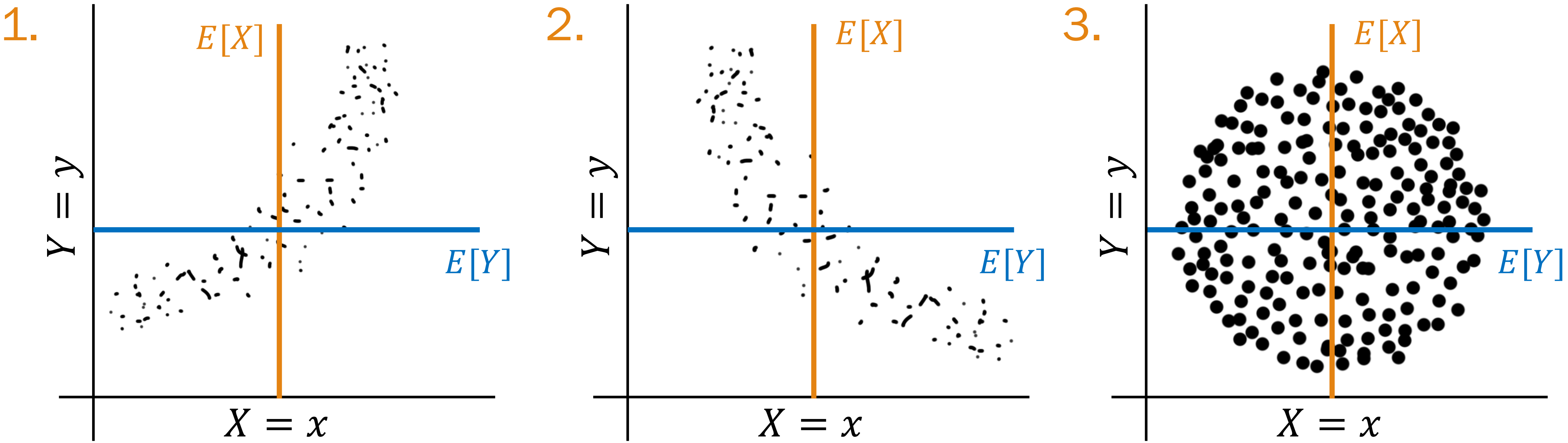
Think by yourself: 1 min

(by yourself)

Feel the covariance

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

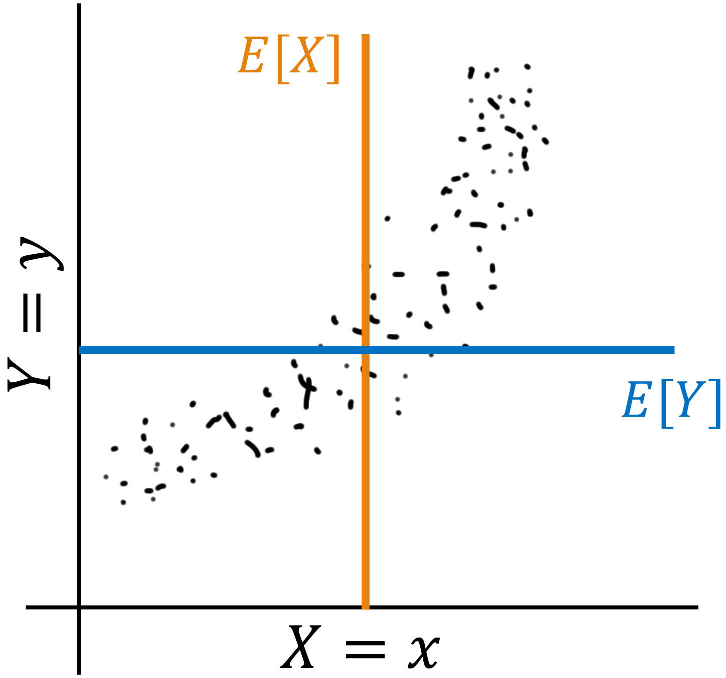
Is the covariance positive, negative, or zero?



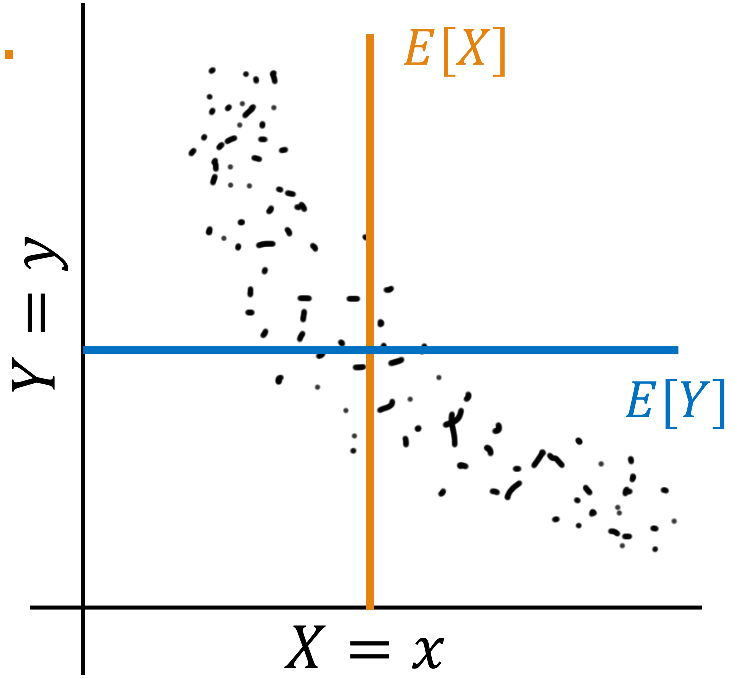
Feel the covariance

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

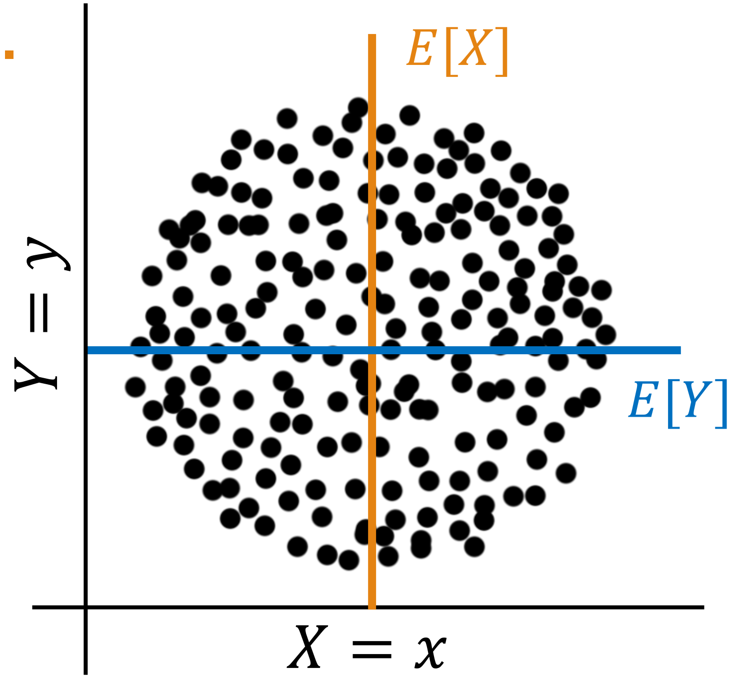
Is the covariance positive, negative, or zero?

1. 

positive

2. 

negative

3. 

zero

Properties of Covariance

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

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

Properties:

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

   Covariance is non-linear: $\text{Cov}(aX + b, Y) = a\text{Cov}(X,Y)$

Statistics of sums of RVs

Review

For any random variables X and Y,

$$E[X+Y] = E[X] + E[Y]$$

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

For **independent** X and Y,

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

(Lemma: proof in extra slides)

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

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

Breakout Rooms

Check out the properties on the next slide.
Post any clarifications here!

https://us.edstem.org/courses/667/discussion/93095

Breakout rooms: 4 min. Introduce yourself!

Zero covariance does not imply independence

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

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

Y \ X	-1	0	1	Marginal PMF of Y, $p_Y(y)$
0	1/3	0	1/3	2/3
1	0	1/3	0	1/3
Marginal PMF of X, $p_X(x)$	1/3	1/3	1/3	

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

Zero covariance does not imply independence

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

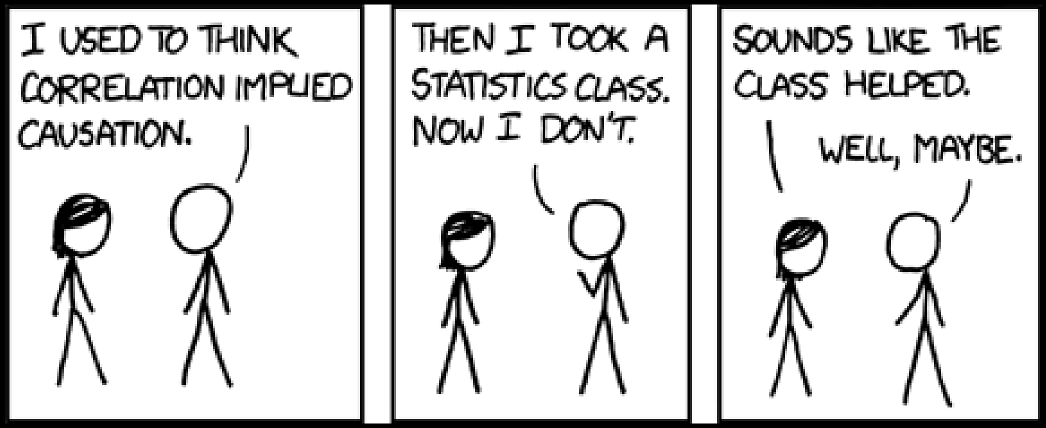
Define $Y = \begin{cases} 1 & \text{if } X = 0 \\ 0 & \text{otherwise} \end{cases}$

		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. $E[X] = -1\left(\frac{1}{3}\right) + 0\left(\frac{1}{3}\right) + 1\left(\frac{1}{3}\right) = 0$ $\qquad E[Y] = 0\left(\frac{2}{3}\right) + 1\left(\frac{1}{3}\right) = 1/3$
2. $E[XY] = (-1 \cdot 0)\left(\frac{1}{3}\right) + (0 \cdot 1)\left(\frac{1}{3}\right) + (1 \cdot 0)\left(\frac{1}{3}\right)$
   $= 0$
3. $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$
   $= 0 - 0(1/3) = 0$ ⚠ does not imply independence!
4. Are X and Y independent? ✕
   $P(Y = 0 | X = 1) = 1$
   $\neq \quad P(Y = 0) = 2/3$



Interlude for fun/announcements

Announcements

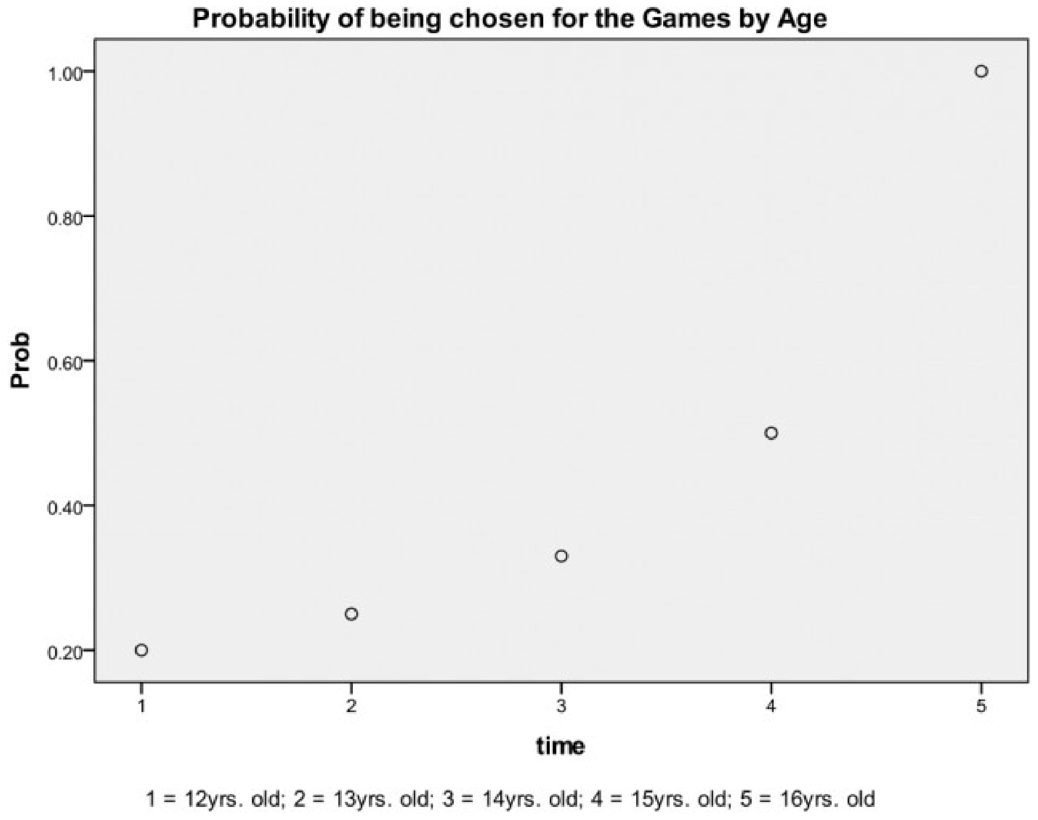
Midterm Quiz

Start: Today (Mon) 5PM PDT – find on Website
Due: Tomorrow (Tue) 5PM PDT – submit to Gradescope

More notes: (no office hours tomorrow, Ed will be set to private-questions-only mode, we'll make clarifications via Ed)

Interesting probability news

Probability and Game Theory in *The Hunger Games*



https://www.wired.com/2012/04/probability-and-game-theory-in-the-hunger-games/

"Suppose the parents in a given district gave birth to only...five girls, and that all of these kids were born at the same time."

- Not a probability mass function
- Also duh? (P(you get chosen if you're the only person) = 1)
- You now know enough Python/ probability to write a better simulation to model the Reaping!!!!
- (game theory part of the article is good)

Ethics in Probability: Smoking and Cancer

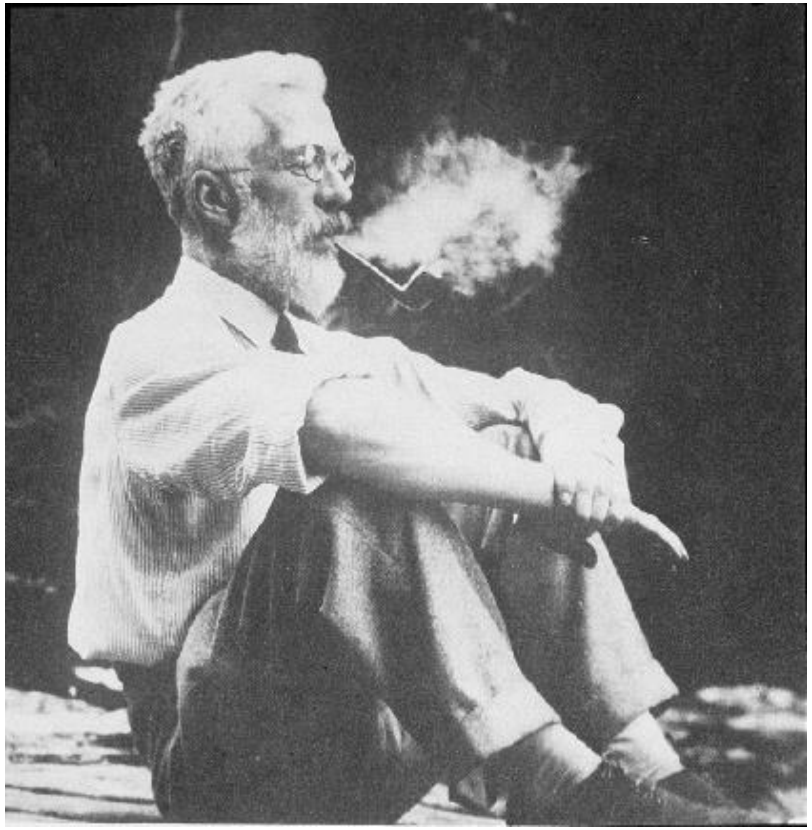
Correlation does not imply causation

Does lung cancer cause smoking?

https://towardsdatascience.com/correlation-does-not-imply-causation-92e4832a6713

> *"Is it possible then, that lung cancer — that is to say, the pre-cancerous condition which must exist and is known to exist for years in those who are going to show over lung cancer — is one of the causes of smoking cigarettes? I don't think it can be excluded."*

- Statistician R.A. Fisher



http://www.economics.soton.ac.uk/staff/aldrich/fisherguide/Doc1.htm

How, then, do we think about correlation and causation?

Fisher's Paper (1958): https://www.nature.com/articles/182596a0

A reference paper from Judea Pearl on causal inference: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2836213/

LIVE

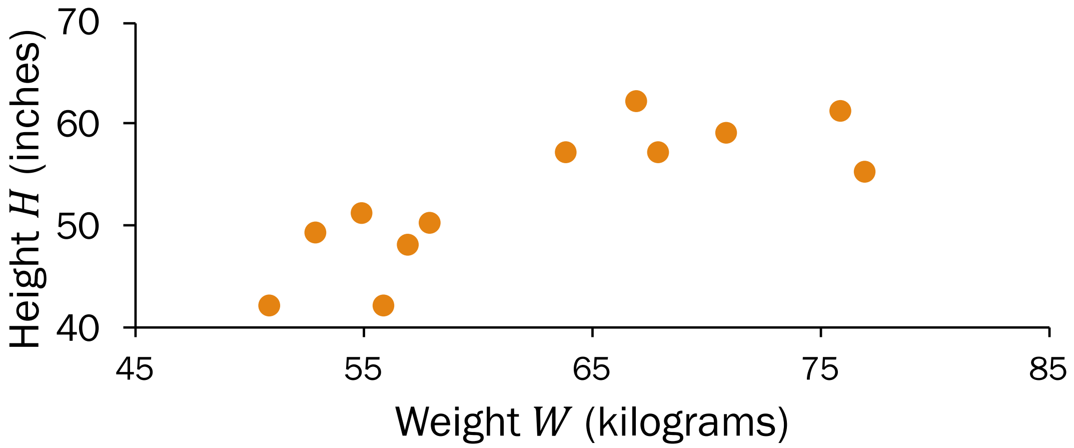
Correlation

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?

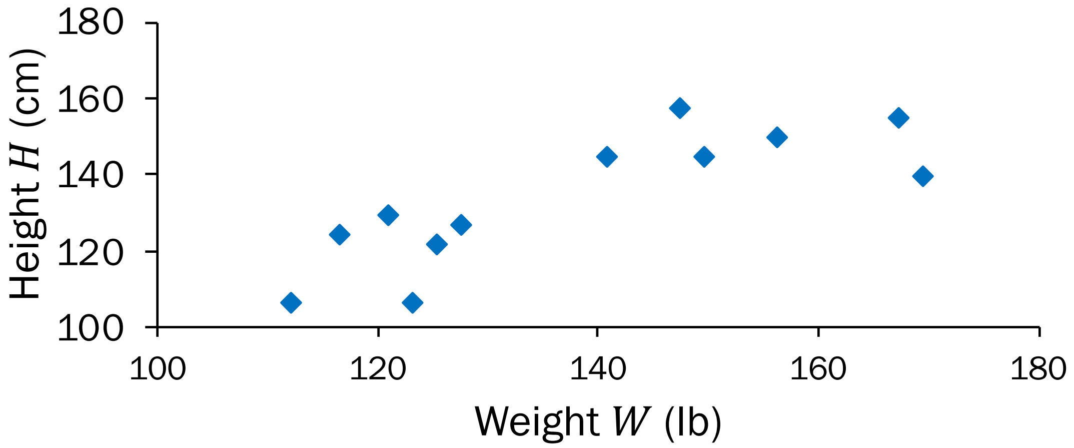
$$\text{Cov}(W,H) = E[WH] - E[W]E[H]$$
$$= 3355.83 - (62.75)(52.75)$$
$$= 45.77 \quad \text{(positive)}$$



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!



Sign of covariance (+/–) more meaningful than magnitude

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:

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

Think

The next slide has a question to go over by yourself.

Post any clarifications here!

https://us.edstem.org/courses/667/discussion/93095

Think by yourself: 1 min

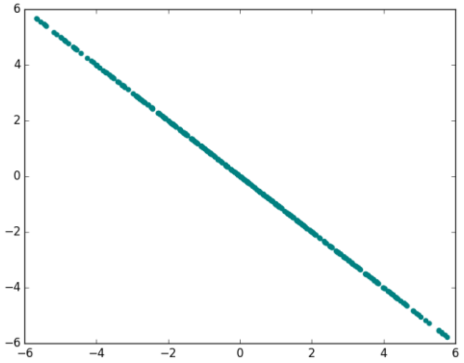
(by yourself)

Correlation reps

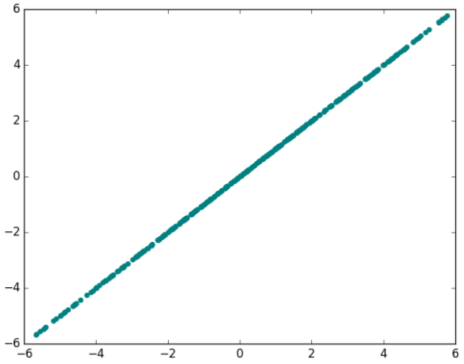
A. $\rho(X, Y) = 1$
B. $\rho(X, Y) = -1$
C. $\rho(X, Y) = 0$
D. Other

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

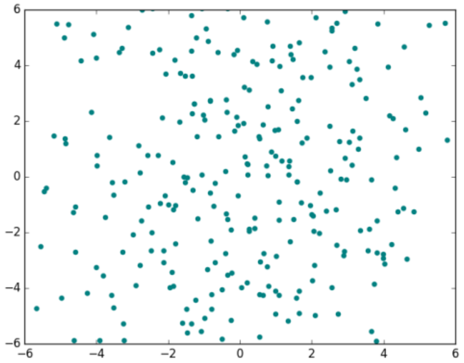
1.



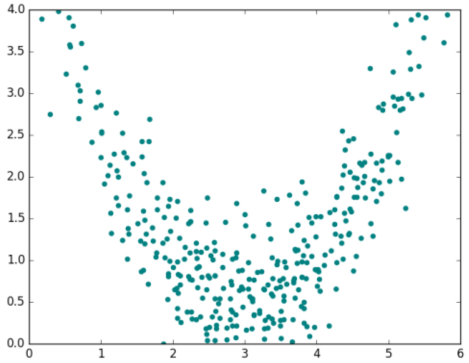
2.



3.



4.

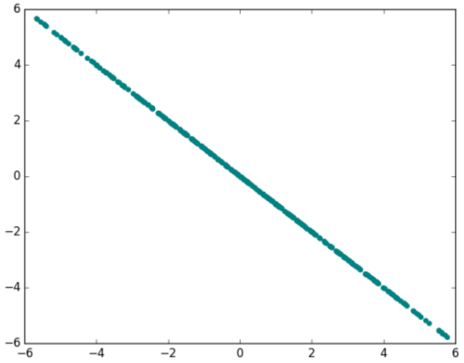


Correlation reps

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

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

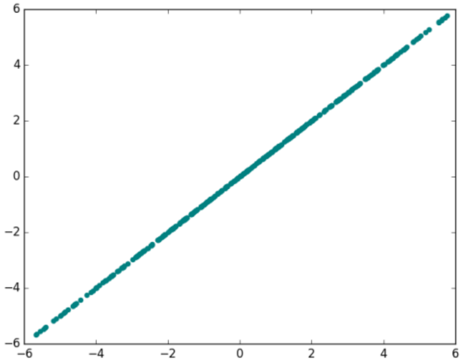
1.



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

$$Y = -aX + b$$
$$a > 0$$

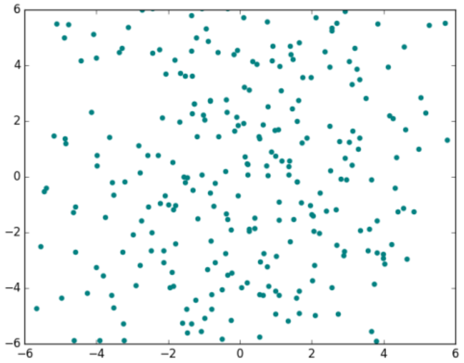
2.



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

$$Y = aX + b$$
$$a > 0$$

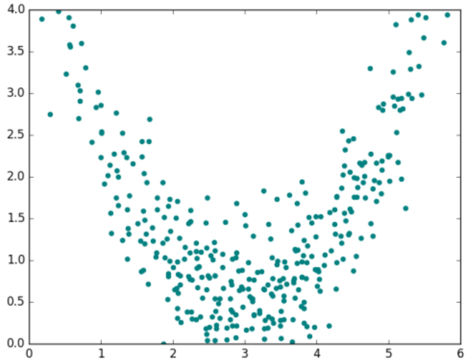
3.



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

"uncorrelated"

4.



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

$$Y = X^2$$

X and Y can be nonlinearly related even if $\rho(X,Y) = 0$.

CS103: Conditional statements

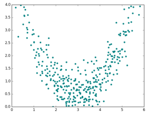
Statement $P \to Q$: Independence → No correlation ☑

Contrapositive $\neg Q \to \neg P$: Correlation → Dependence ☑ (logically equivalent)

Inverse $\neg P \to \neg Q$: Dependence → Correlation? ✗ (not always)

$Y = X^2$

$\rho(X, Y) = 0$



Converse $Q \to P$: No correlation → Independence? ✗ (not always)

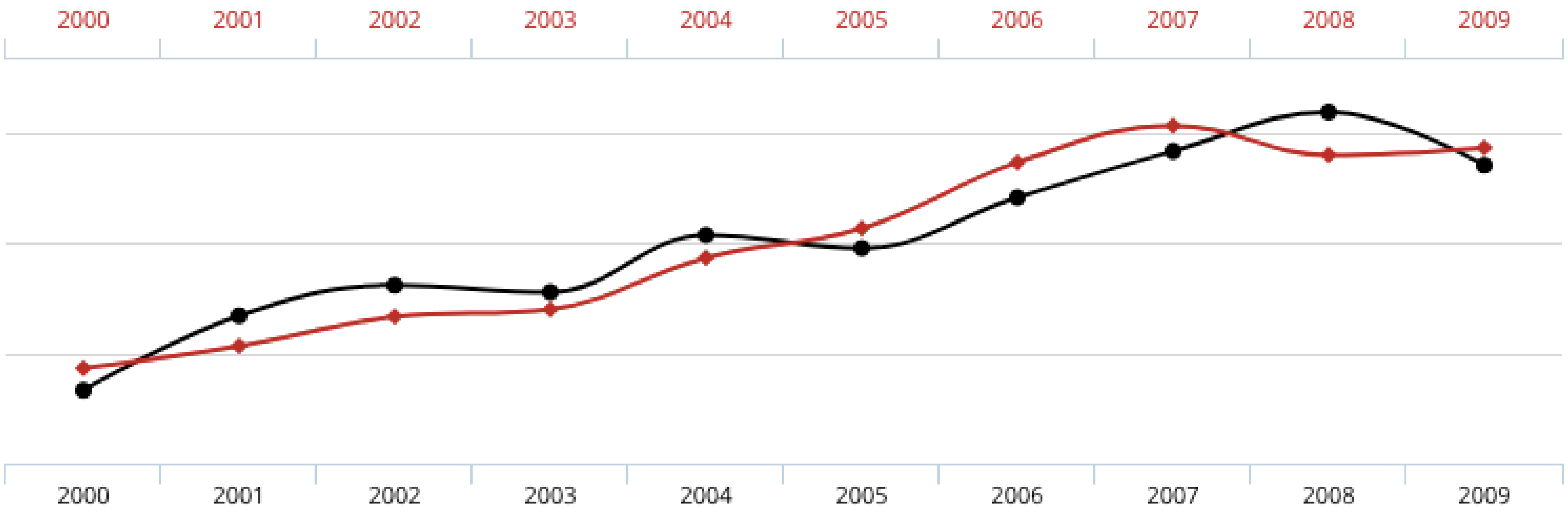
Slide 46

"Correlation does not imply causation"

Spurious Correlations

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

Correlation:
0.947091

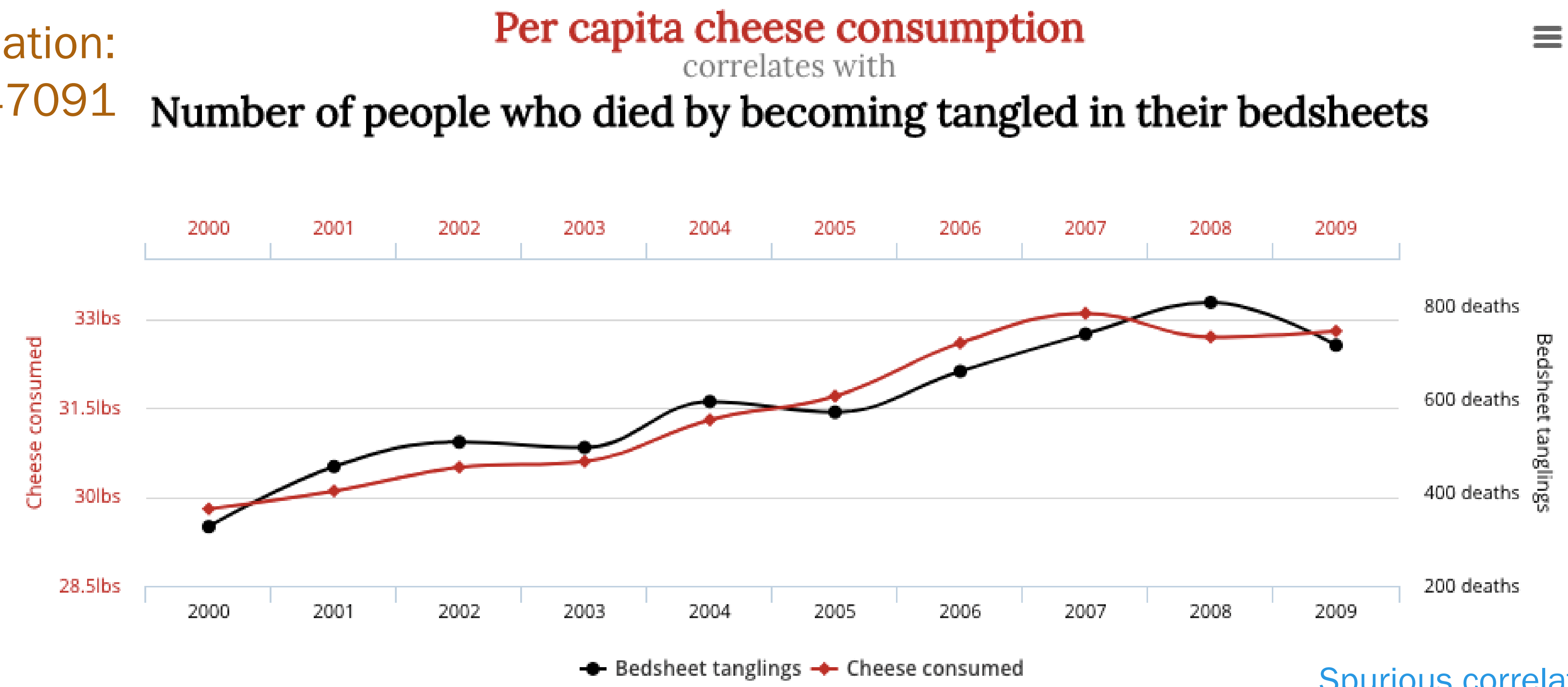


[Spurious correlations]

Spurious Correlations

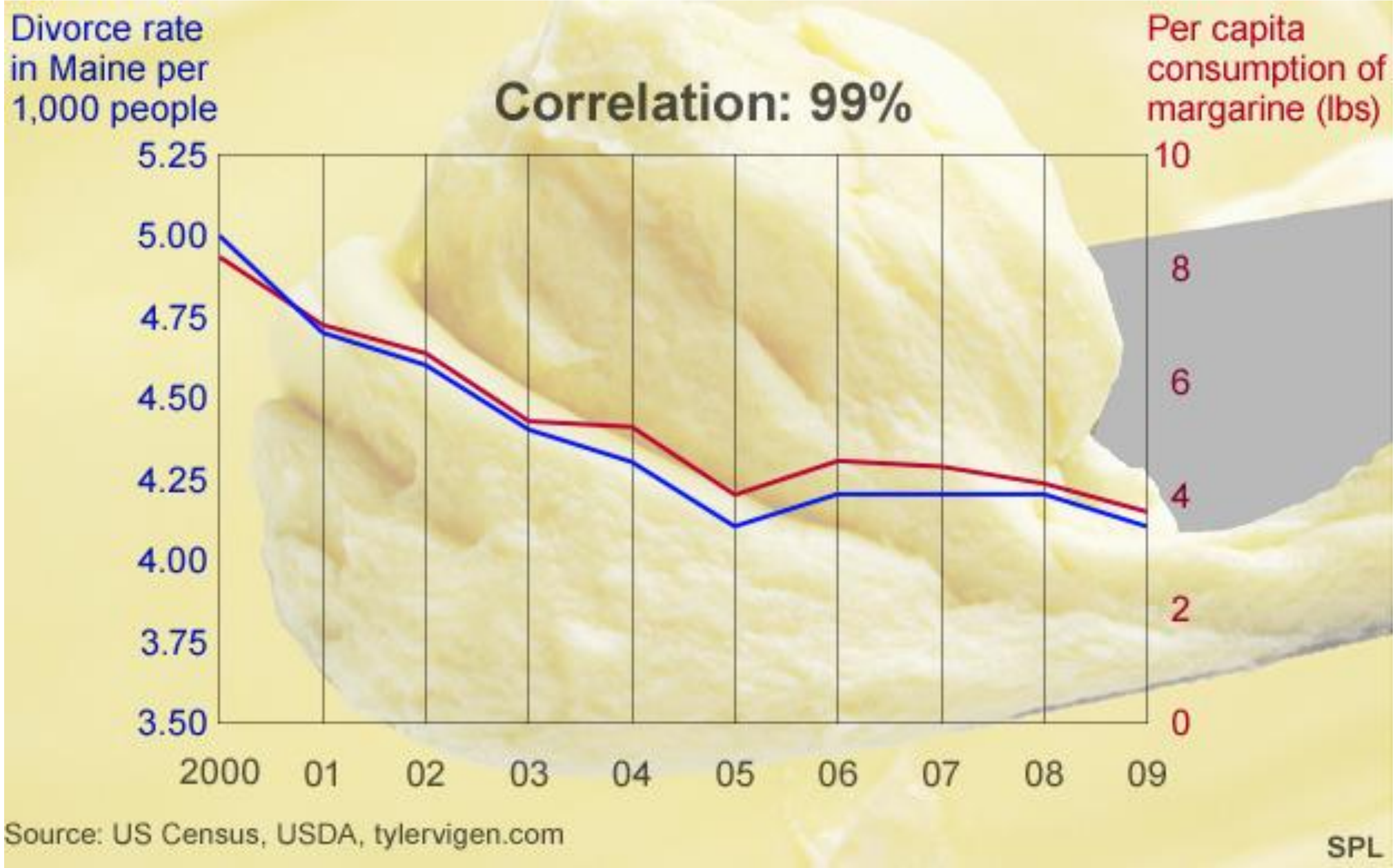
$\rho(X, Y)$ is used a lot to statistically quantify the relationship b/t X and Y.

Correlation: 0.947091



[Spurious correlations]

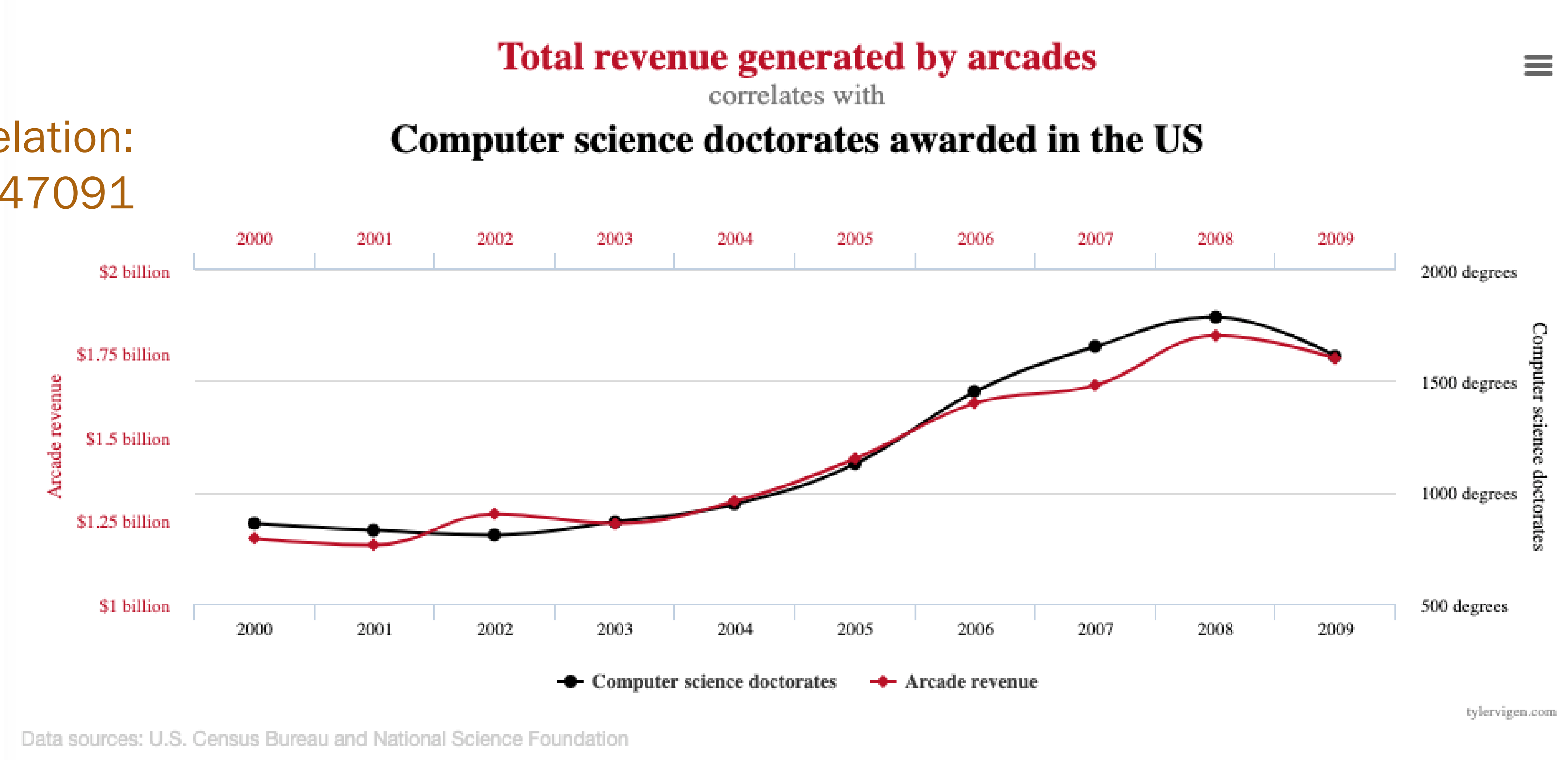
Divorce vs. Butter



http://www.bbc.com/news/magazine-27537142

Arcade revenue vs. CS PhDs

Correlation: 0.947091



[Spurious correlations]

Extra

Expectation of product of independent RVs

If X and Y are independent, then

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

Proof:

$$E[g(X)h(Y)] = \sum_y \sum_x g(x)h(y)p_{X,Y}(x,y)$$

(for continuous proof, replace summations with integrals)

$$= \sum_y \sum_x g(x)h(y)p_X(x)p_Y(y)$$

X and Y are independent

$$= \sum_y \left(h(y)p_Y(y) \sum_x g(x)p_X(x) \right)$$

Terms dependent on y are constant in integral of x

$$= \left(\sum_x g(x)p_X(x) \right) \left(\sum_y h(y)p_Y(y) \right)$$

Summations separate

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

Variance of Sums of Variables

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

Proof:

Var(X) = Cov(X,X)

covariance of all pairs

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

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

Symmetry of covariance
Cov(X,X) = 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