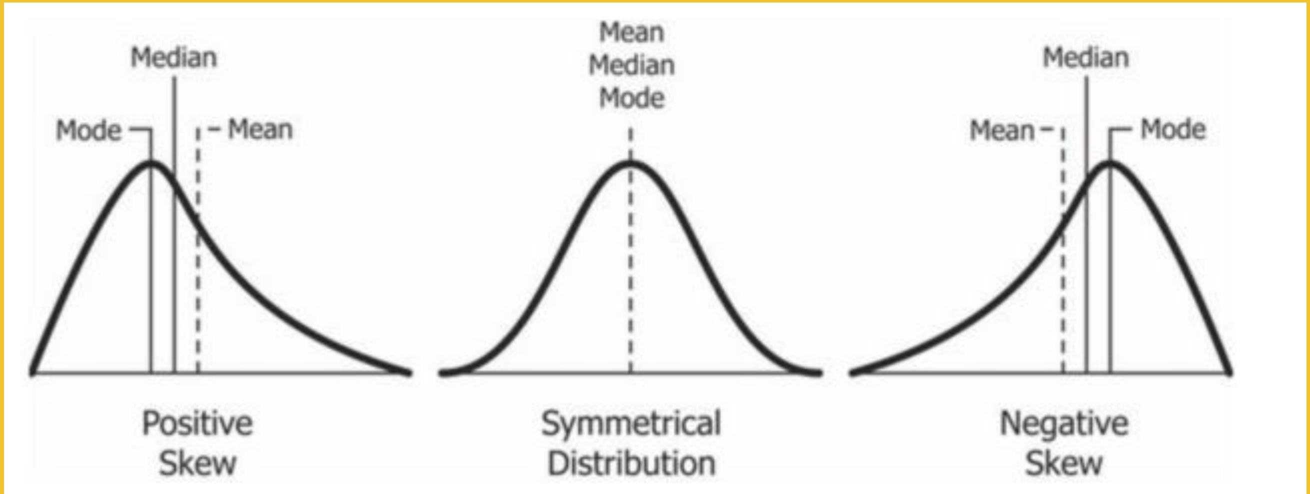


Econ 2250: Stats for Econ

Fall 2022

[Source for pic stats above.]

Announcements

- Homework 6 is due on Tuesday

These are still really useful:

- https://www.probabilitycourse.com/chapter3/3_2_2_expectation.php
- https://mixtape.scunning.com/02-probability_and_regression#variance

What we will do today?

- Quick revisit summary operator, E(X), Variance
- Revisit Covariance
- Introduce Correlation

Summary Operator

$$\Sigma X = x_1 + x_2 + \ldots + x_n$$

Summary Operator Property 3:

For any constant a and b: $\sum_{i=1}^{n}(ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{j=1}^{n} y_i$

$$\Sigma_i^n (a * x_i + b * y_i) =$$
$$(a * x_1 + b * y_1) + (a * x_2 + b * y_2) + (a * x_3 + b * y_3) =$$
$$a * x_1 + b * y_1 + a * x_2 + b * y_2 + a * x_3 + b * y_3 =$$
$$a(x_1 + x_2 + x_3) + b(y_1 + y_2 + y_3) =$$
$$a\Sigma_i^n x_i + b\Sigma_i^n y_i$$

Show that…using x = (1,2,3)

$$\sum_{i}^{n} \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}$$

And

$$\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i\right)^2$$

Expected Value Operator

$$E(x) = \Sigma\, x_i * Pr(x_i)$$

Expected Value Operator Property 2:

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

x = [3,6,2]

p(x) = ⅓

a = 5

b = 4

E(aX + b) = E(aX) + E(b)

$= ax_1*p(x_1) + ax_2*p(x_2) + ax_3*p(x_3) + b = a(x_1*p(x_1) + x_2*p(x_2) + x_3*p(x_3)) + b$

= a(E(x)) + b = 5(3*⅓ + 6*⅓ + 2*⅓) + 4 = 22⅓

Variance

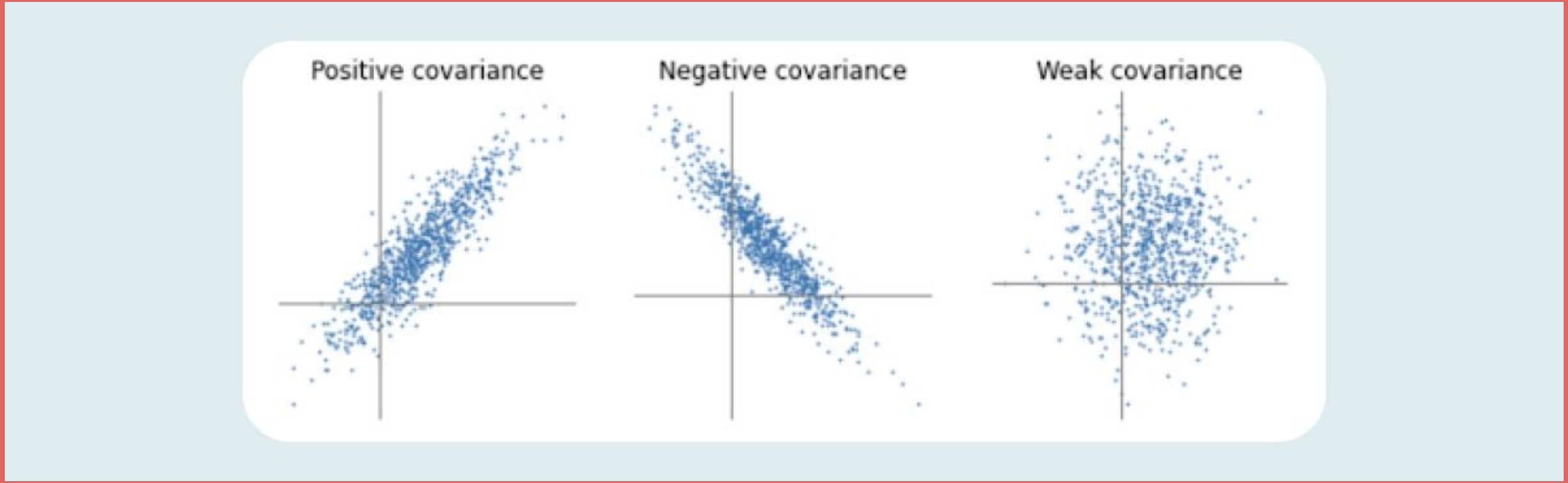
$$V(X) = E((X - E(X))^2)$$

Expectation is our best guess of what something will equal, so take the expectation of the squared deviation

$$E[(X-\mu_x)^2] = \sum (x_i - \mu_x)^2 * P(x_i)$$

if $P(x_i)$ is $\frac{1}{n}$ for all $i = 1, 2, \ldots, n$

$$V(X) = \sum (x_i - \mu_x)^2 * \frac{1}{n} = \frac{1}{n} \sum (x_i - \mu_x)^2$$



Covariance

$Cov(X, Y)$ = E[(X-E(X)(Y-E(Y)]

image

Nice correlation app

https://shiny.rit.albany.edu/stat/rectangles/

Covariance

$$
\begin{aligned}
Cov(x, y) &= E[(X - E(X))(Y - E(Y))] \\
&= \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n - 1}
\end{aligned}
$$

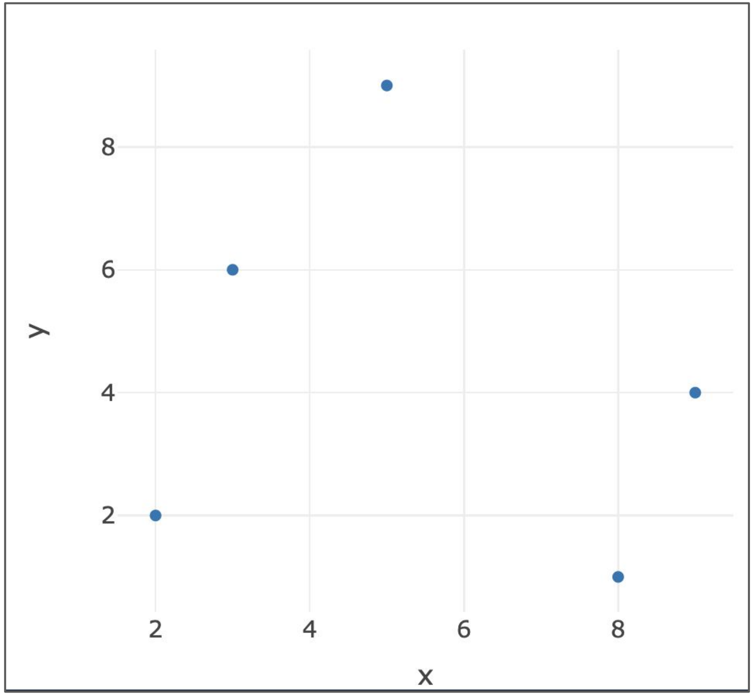
Example covariance

$$\frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n - 1}$$

x	y	demean_x	demean_y	demean_x*demean_y
3	6	-2.4	1.6	-3.84
5	9	-0.4	4.6	-1.84
2	2	-3.4	-2.4	8.16
8	1	2.6	-3.4	-8.84
9	4	3.6	-0.4	-1.44

-7.8	sum
-1.95	sum/(n-1)

mean_y	4.4
mean_x	5.4



Correlation

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \mu_x}{\sigma_x} \right) \left(\frac{y_i - \mu_y}{\sigma_y} \right)$$

Correlation

$$\rho_{X,Y} = \operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\operatorname{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}, \quad \text{if } \sigma_X \sigma_Y > 0$$

$$r_{xy} \stackrel{\text{def}}{=} \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \mu_x}{\sigma_x} \right) \left(\frac{y_i - \mu_y}{\sigma_y} \right)$$

Example

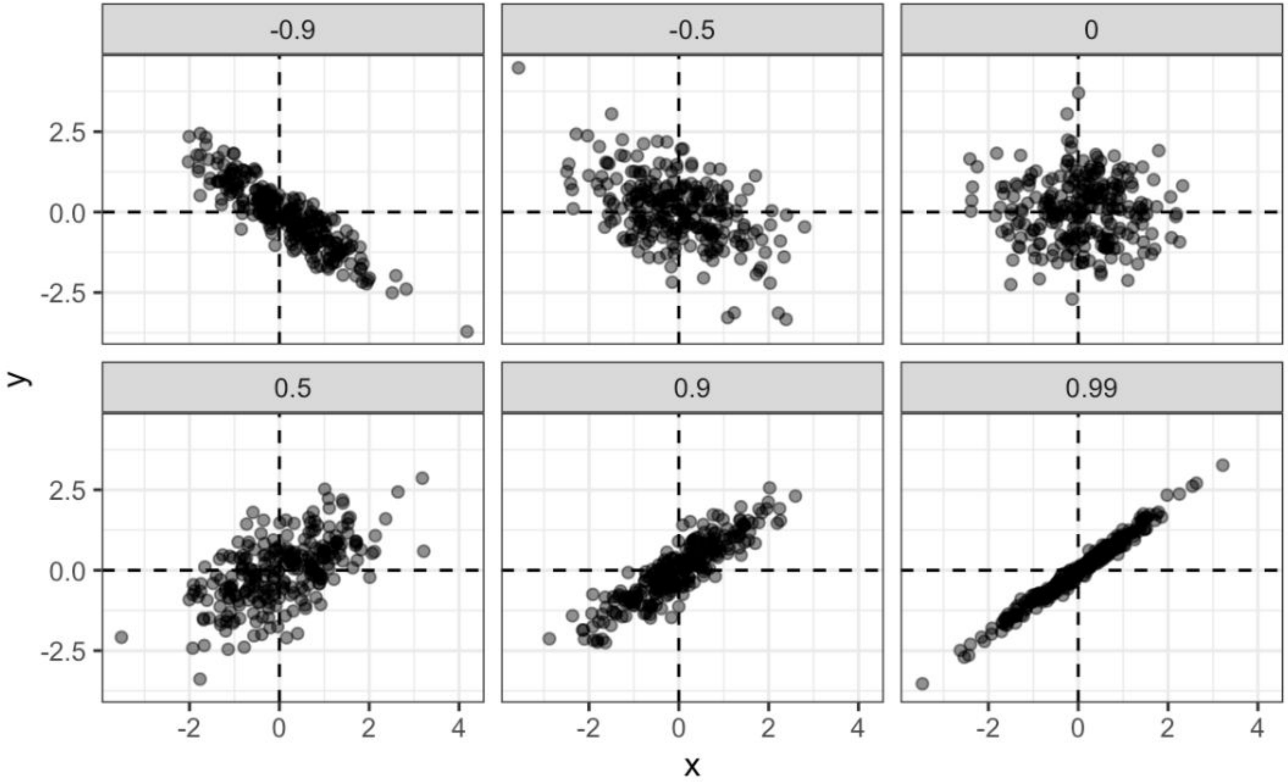
$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

x	y	demean_x	demean_x_sq	demean_y	demean_y_sq	demean_x*demean_y	
3	6	-2.4	5.76	1.6	2.56	-3.84	
5	9	-0.4	0.16	4.6	21.16	-1.84	
2	2	-3.4	11.56	-2.4	5.76	8.16	
8	1	2.6	6.76	-3.4	11.56	-8.84	
9	4	3.6	12.96	-0.4	0.16	-1.44	
			37.2		**41.2**	-7.8	sum
						-1.95	sum/(n-1)

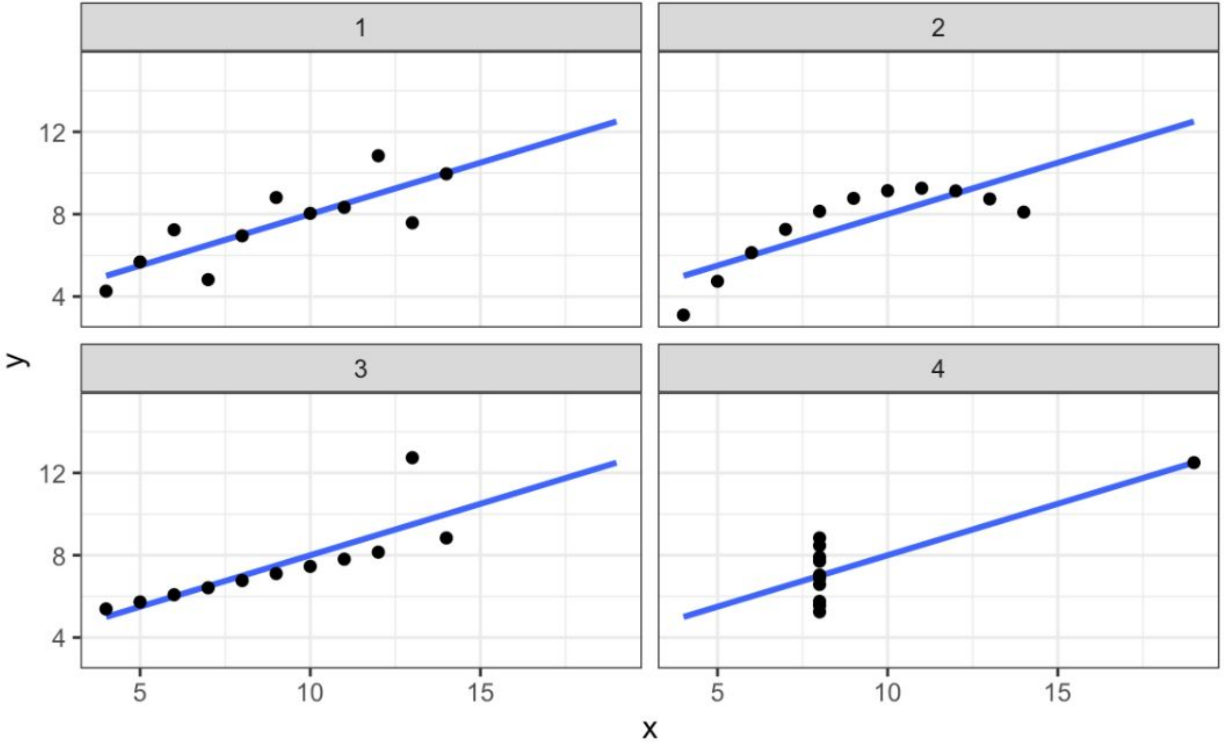
mean_y	4.4
mean_x	5.4

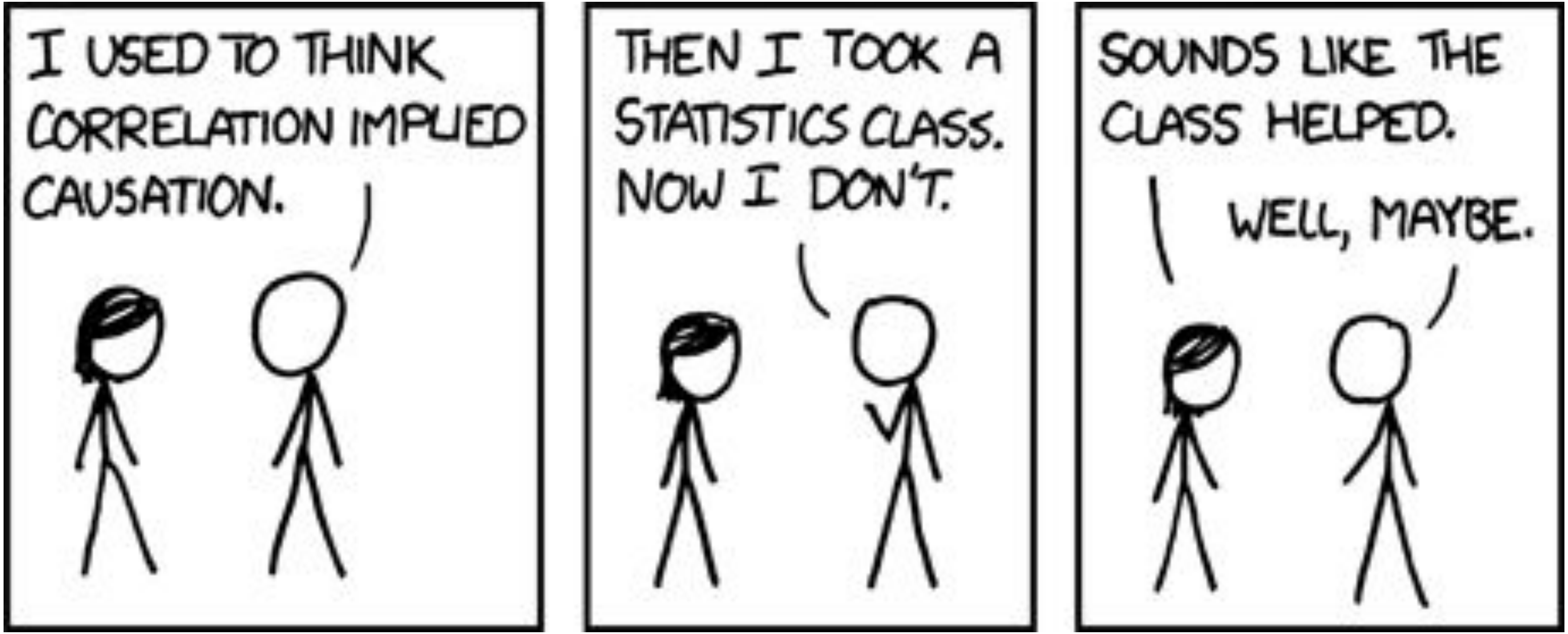
numerator	**-1.95**	-1.95	**-0.22**	**correlation**
denom	sqrt(**37.2** + **41.2**)	8.85		

Bounded by [-1,1]



Not always useful




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

And certainly does **not** imply causation

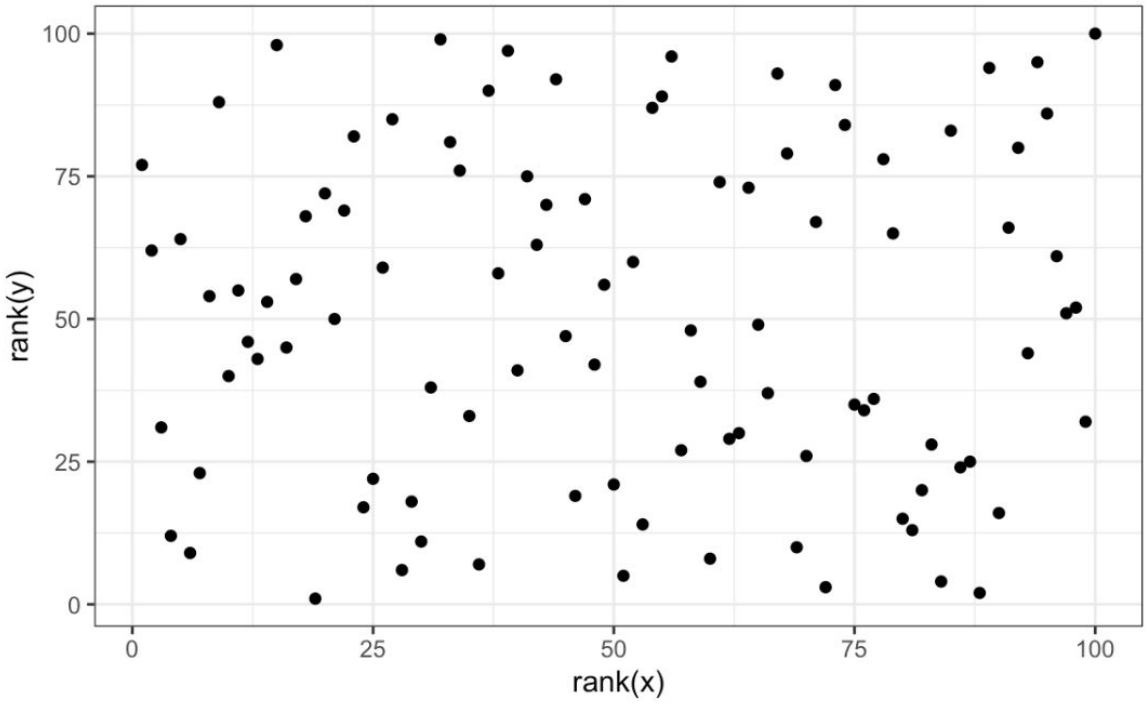
Possible outcomes

- A causes B (direct causation);
- B causes A (reverse causation);
- A and B are both caused by C (common causation);
- There is no connection between A and B; the correlation is a coincidence.

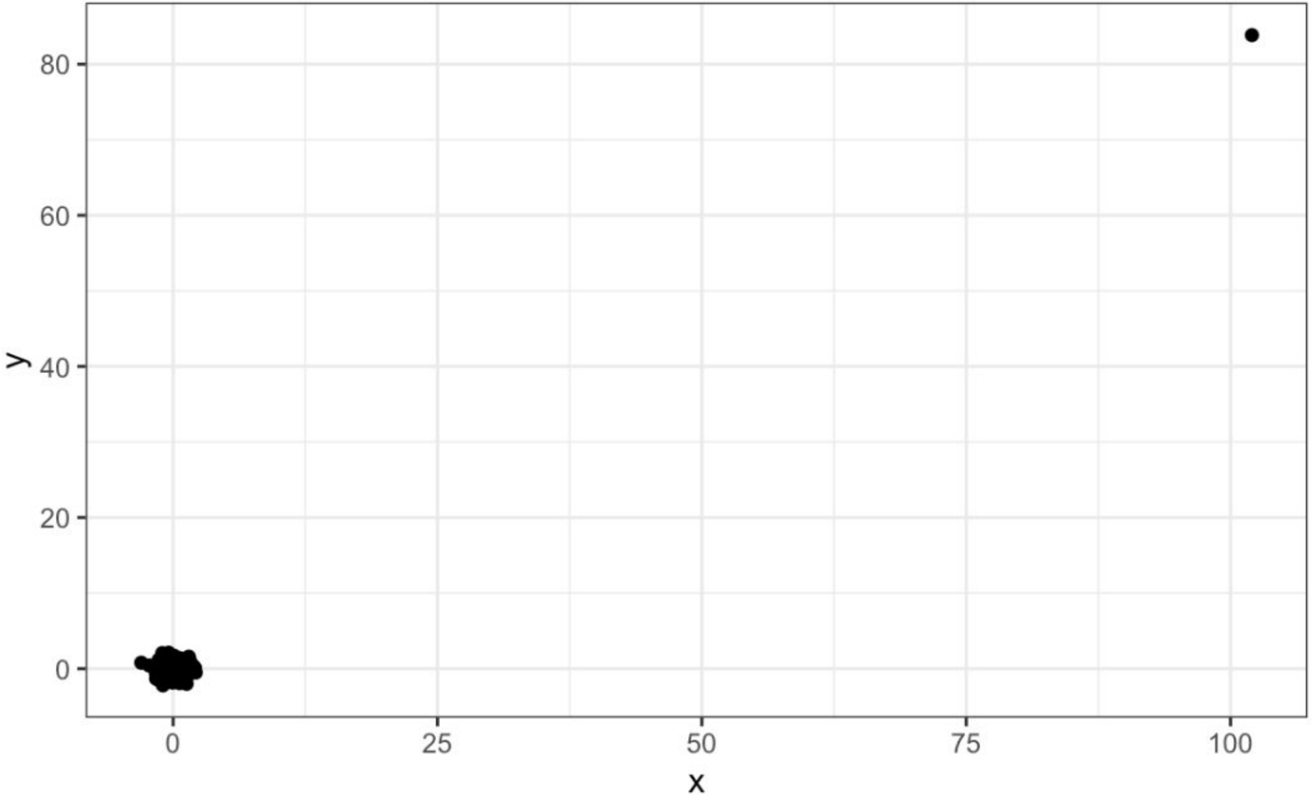
Spurious Correlation

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

Outliers



Outliers con’t



Confounders

- Tutors make students worse
- Students who wear uniforms perform better academically
- People who eat vegetables have longer life spans
- Increased ice cream sales cause more drownings

Simpson’s paradox: batting averages

Year / Batter	1995		1996		Combined	
Derek Jeter	12/48	.250	183/582	.314	195/630	**.310**
David Justice	104/411	**.253**	45/140	**.321**	149/551	.270

Simpson's paradox

Famous Berkeley Graduate Admissions Example

	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
Total	12,763	41%	8,442	44%	4,321	35%

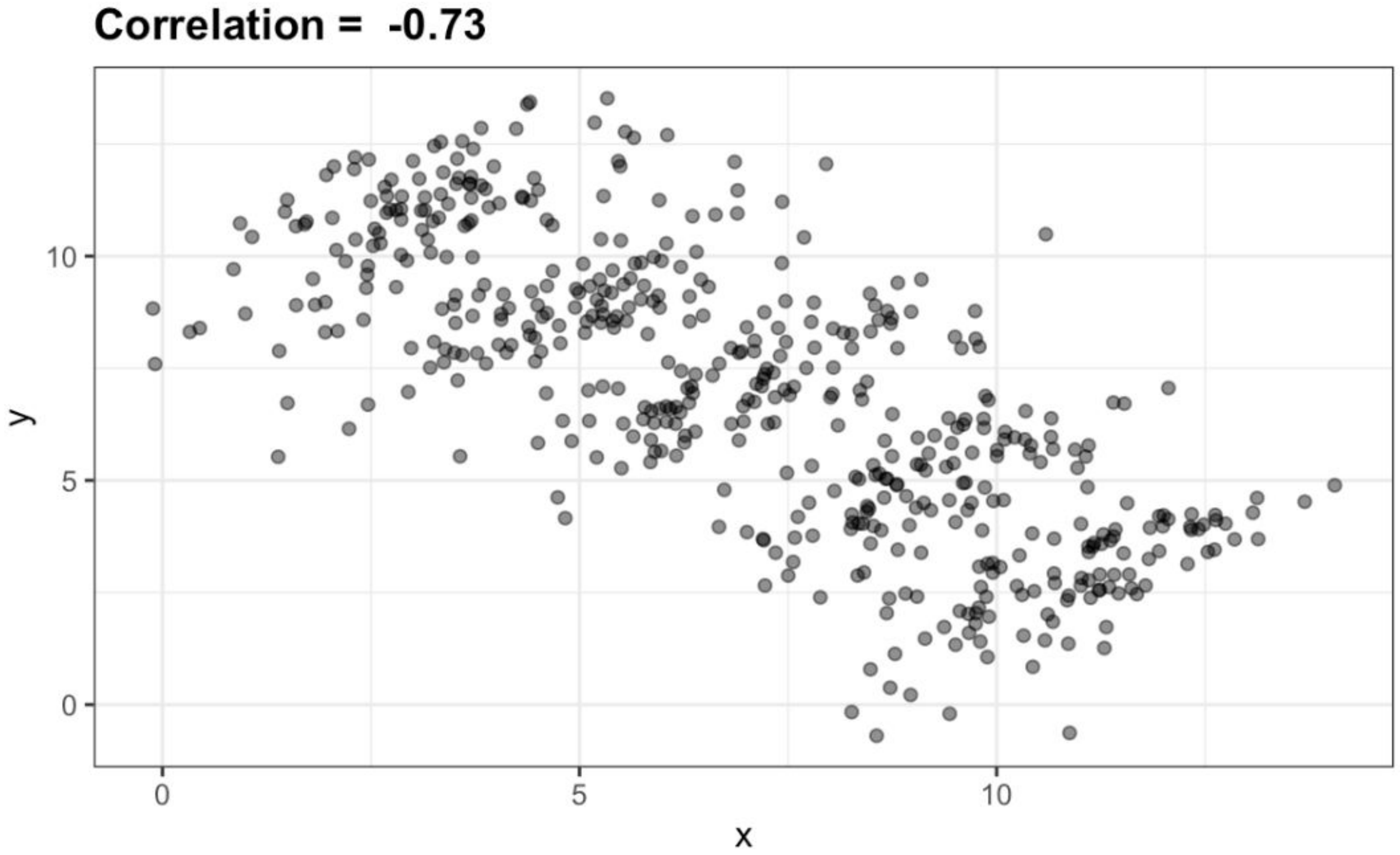
Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
A	933	64%	**825**	62%	108	82%
B	585	63%	**560**	63%	25	68%
C	918	35%	325	37%	**593**	34%
D	792	34%	417	33%	375	35%
E	584	25%	191	28%	**393**	24%
F	714	6%	373	6%	341	7%
Total	**4526**	**39%**	**2691**	**45%**	**1835**	**30%**

Legend:

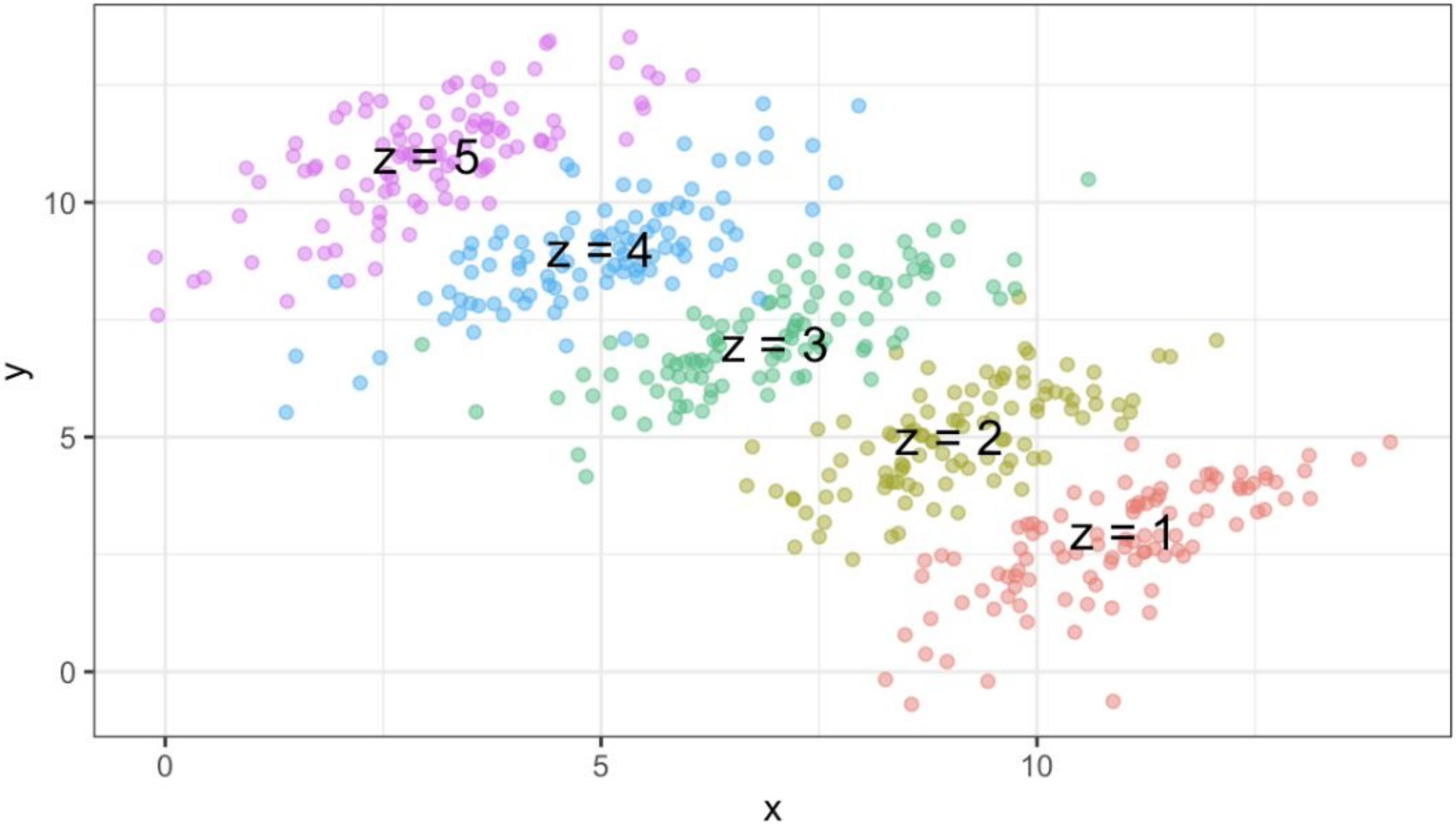
greater percentage of successful applicants than the other gender

greater number of applicants than the other gender

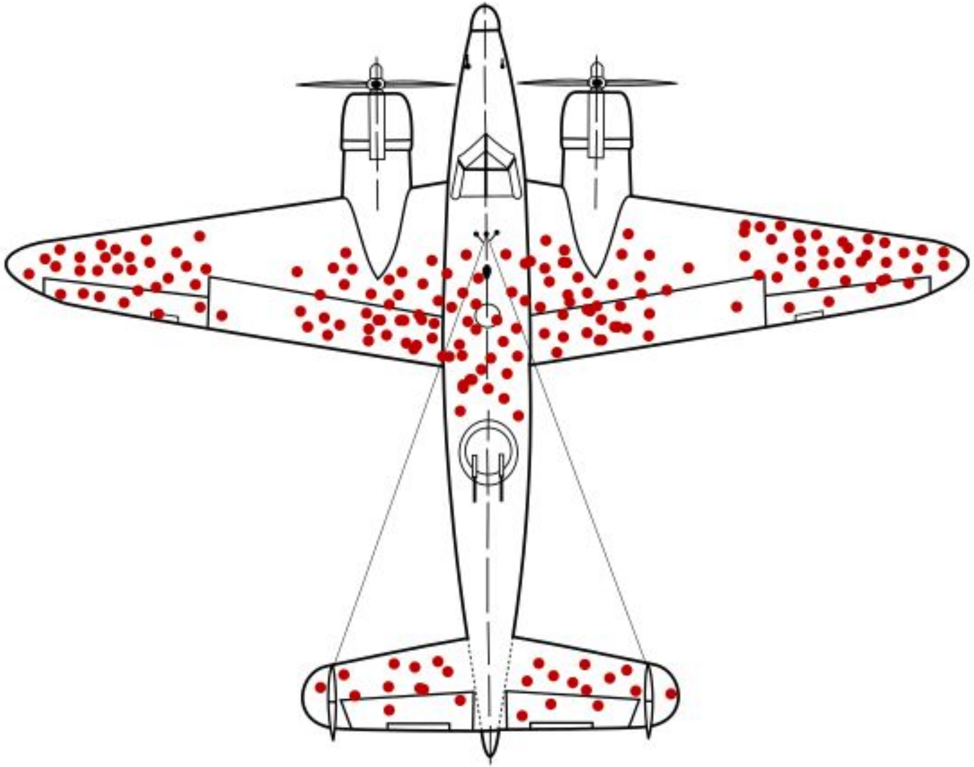
bold - the two 'most applied for' departments for each gender


Correlation = -0.73
y
x
0
5
10
0
5
10

Correlations = 0.74 0.65 0.76 0.72 0.71



Survivorship bias



End of class form

https://forms.gle/kgT2w9wPZo3vJcjA8