Корреляция

Пусть у нас есть выборка X, посчитаем ее выборочное среднее

$$\overline{X} = \frac{\sum X_i}{n}$$

Так же мы можем посчитать для каждого наблюдения его отклонение от среднего выборочного

$$\Delta X_i = X_i - \overline{X}$$

Пусть у нас есть выборка Y, в которой наблюдения спарены с наблюдениями из выборки X (то есть i-му наблюдению из X соответствует i-е наблюдение из Y)

Аналогично:

$$\overline{Y} = \frac{\sum Y_i}{n}$$

$$\Delta Y_i = Y_i - \overline{Y}$$

Давайте теперь посмотрим, на произведение соответствующих отклонений

$$\Delta X_i \Delta Y_i = (Y_i - \overline{Y})(X_i - \overline{X})$$

Очевидно, что это произведение больше 0, когда оба отклонения имеют одинаковый знак и меньше 0, когда разный

Введем ковариацию - выборочное среднее произведений соответствующих отклонений. Делим на n -1 по причине схожей с той, по которой выборочная дисперсия считается с коэффициентом 1/(n-1), а не 1/n

$$cov(X, Y) = \frac{1}{n-1} \sum_i (Y_i - \overline{Y})(X_i - \overline{X})$$

Ковариация

$$cov(X, Y) = \frac{1}{n-1} \sum_i (Y_i - \overline{Y})(X_i - \overline{X})$$

Эта величина больше 0, если положительным отклонениям от среднего в X соответствуют положительные отклонения от среднего в Y и наоборот

Если положительным отклонениям от среднего в X соответствуют отрицательные отклонения от среднего в Y и наоборот

Ковариация

Эта величина больше 0, если положительным отклонениям от среднего в X соответствуют положительные отклонения от среднего в Y и наоборот

Если положительным отклонениям от среднего в X соответствуют отрицательные отклонения от среднего в Y и наоборот

Если X и Y из независимых распределений, то ковариация в среднем будет равна 0.

Ковариация может изменяться в пределах от -бесконечности до + бесконечности (зависит от величин и их масштаба)

Разделим ее на произведение стандартных отклонений X и Y

Корреляция

$$cor(X, Y) = \frac{\frac{1}{n-1}\sum_i (Y_i - \overline{Y})(X_i - \overline{X})}{s_x \cdot s_y}$$

$$s_x^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2$$

$$s_y^2 = \frac{1}{n-1}\sum_i (Y_i - \overline{Y})^2$$

Корреляция

$$cor(X,Y) = \frac{\frac{1}{n-1}\sum_i (Y_i - \overline{Y})(X_i - \overline{X})}{s_x \cdot s_y}$$

$$s_x^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2 \qquad s_y^2 = \frac{1}{n-1}\sum_i (Y_i - \overline{Y})^2$$

$$s_x \cdot s_y = \frac{1}{n-1}\sqrt{\sum_i (Y_i - \overline{Y})^2 \sum_i (X_i - \overline{X})^2}$$

$$cor(X,Y) = \frac{\sum_i (Y_i - \overline{Y})(X_i - \overline{X})}{\sqrt{\sum_i (Y_i - \overline{Y})^2 \sum_i (X_i - \overline{X})^2}}$$

Корреляция

$$s_x^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2 \qquad s_y^2 = \frac{1}{n-1}\sum_i (Y_i - \overline{Y})^2$$

$$s_x \cdot s_y = \frac{1}{n-1}\sqrt{\sum_i (Y_i - \overline{Y})^2 \sum_i (X_i - \overline{X})^2}$$

$$s_{xy} = cov(X, Y) = \frac{1}{n-1}\sum_i (Y_i - \overline{Y})(X_i - \overline{X}) = \frac{n}{n-1}(\overline{XY} - \overline{X}\cdot\overline{Y}) = s_{xy}$$

$$r = cor(X, Y) = \frac{\sum_i (Y_i - \overline{Y})(X_i - \overline{X})}{\sqrt{\sum_i (Y_i - \overline{Y})^2 \sum_i (X_i - \overline{X})^2}} = \frac{s_{xy}}{s_x s_y}$$

Корреляция = Корреляция Пирсона

$$s_x^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2 \qquad s_y^2 = \frac{1}{n-1}\sum_i (Y_i - \overline{Y})^2$$

$$s_x \cdot s_y = \frac{1}{n-1}\sqrt{\sum_i (Y_i - \overline{Y})^2 \sum_i (X_i - \overline{X})^2}$$

$$cov(X, Y) = \frac{1}{n-1}\sum_i (Y_i - \overline{Y})(X_i - \overline{X}) = \overline{XY} - \overline{X}\cdot\overline{Y} = s_{xy}$$

$$r = cor(X, Y) = \frac{\sum_i (Y_i - \overline{Y})(X_i - \overline{X})}{\sqrt{\sum_i (Y_i - \overline{Y})^2 \sum_i (X_i - \overline{X})^2}} = \frac{s_{xy}}{s_x s_y}$$

Корреляция Пирсона

Эта величина больше 0, если положительным отклонениям от среднего в X соответствуют положительные отклонения от среднего в Y и наоборот

Если положительным отклонениям от среднего в X соответствуют отрицательные отклонения от среднего в Y и наоборот

Если X и Y из независимых распределений, то корреляция в среднем будет равна 0.

Корреляция изменяется в пределах от [-1, 1].

Если мы говорим о генеральных совокупностях, то корреляция равна 1 и -1 тогда и только тогда, когда X и Y линейно зависимы.

Корреляция Пирсона

Корреляция также не меняется при умножении одной из выборок/случайных величин на число или прибавления числа.

Верно ли следующее рассуждение:

В общем, если за меру взаимосвязи взять такую «нормированную величину», то она получится очень удобной и очень универсальной: будет показывать, связаны ли величины между собой, давая для прямой связи единицу, для обратной — минус единицу, а для несвязанных величин — ноль…

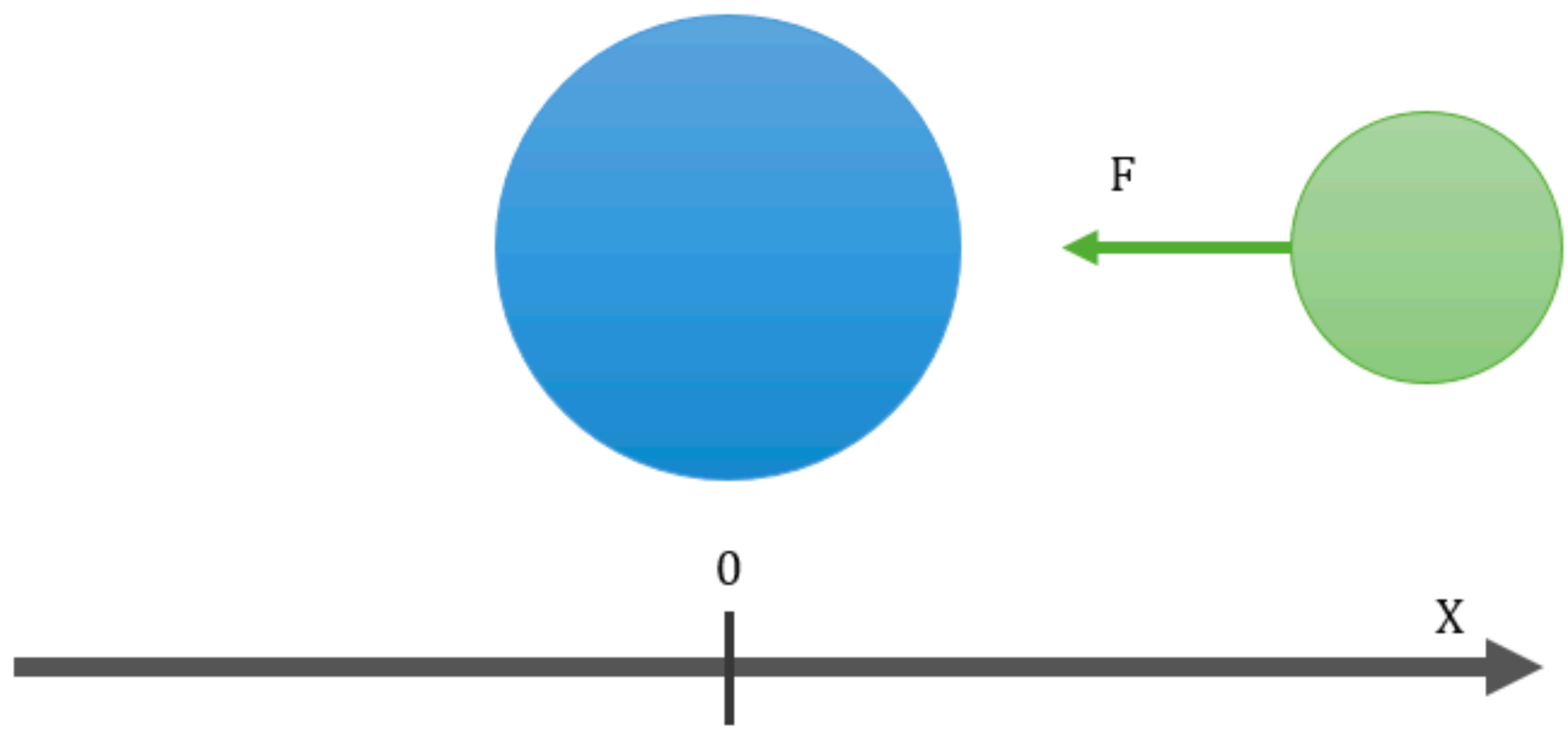
Верно ли следующее рассуждение:

В общем, если за меру взаимосвязи взять такую «нормированную величину», то она получится очень удобной и очень универсальной: будет показывать, связаны ли величины между собой, давая для прямой связи единицу, для обратной — минус единицу, а для несвязанных величин — ноль…

Корреляция Пирсона показывает только линейный вклад. Если это игнорировать, то получается “универсальный способ для поточной генерации совершенно неверных, но зато наукообразных выводов”.

Выстрелы в ногу

Выстрел 1: Ньютон всё наврал

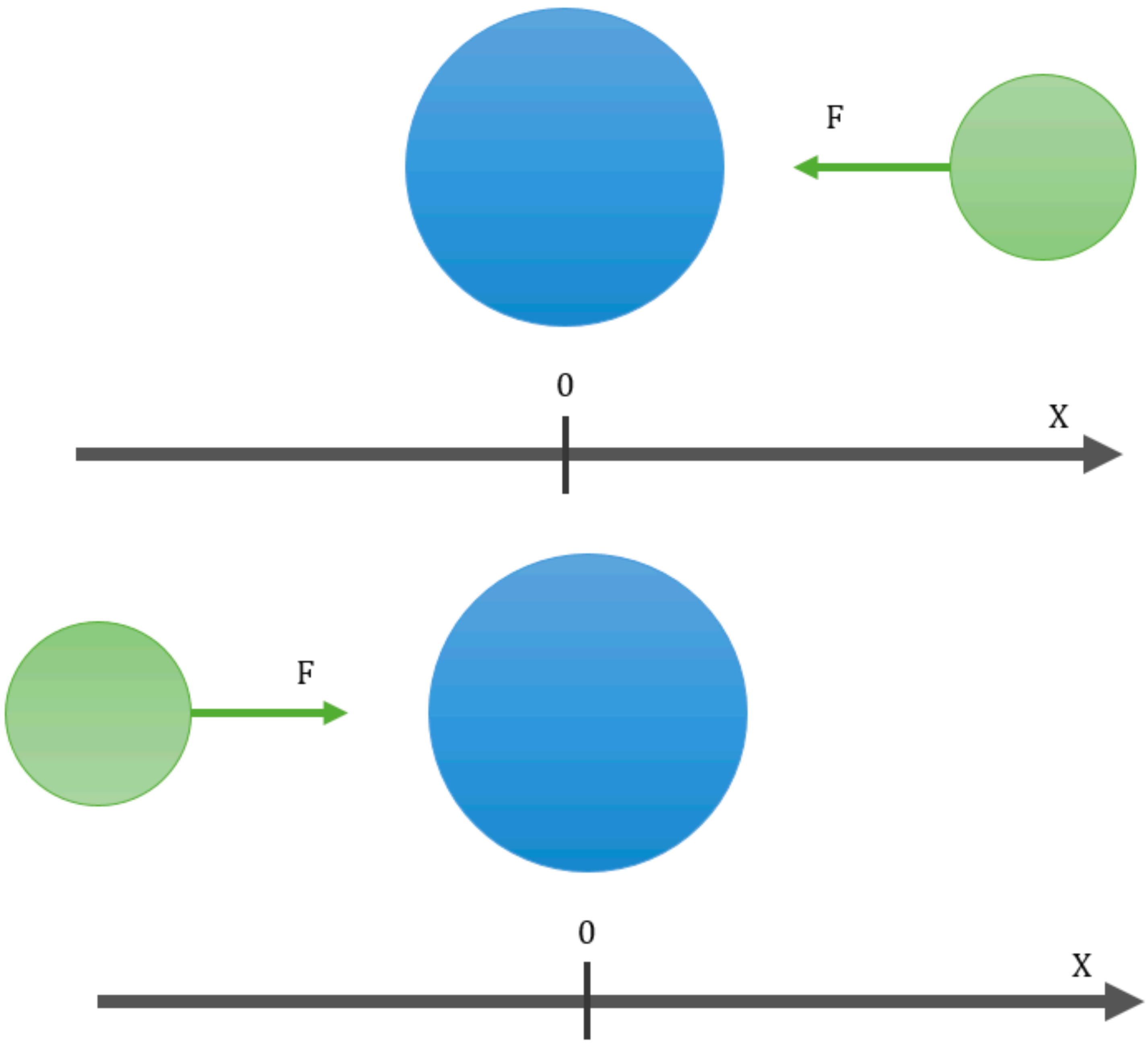


Со всей тщательностью исследователь измерил силу, действующее на тело, отнесённое от центра на самые разные расстояния. Запротоколировал данные.
А потом посчитал корреляцию между координатой изображённого тут зелёным тела и действующей на него силой.

$$r \approx -0{,}68$$

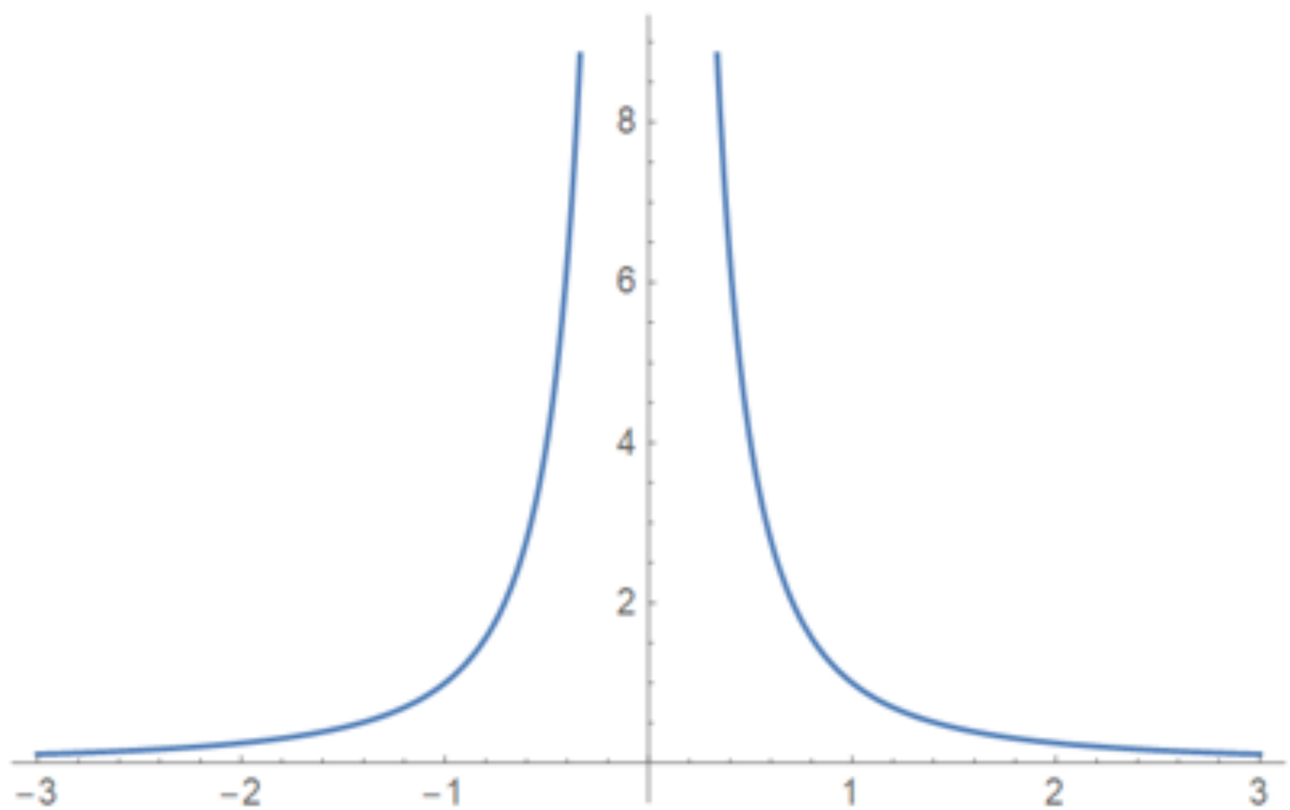
То есть связь между расстоянием между телами и притяжением между ними неполная… ???

Выстрел 1: Ньютон всё наврал



Будем подносить с двух сторон.. Корреляция равна 0

Зависимость силы от координаты второго тела



Отклонения компенсируют друг друга - получается 0

Координата	dx	dF	dx*dF
1	1	dF_1	dF_1
−1	−1	dF_1	$-dF_1$

Почему когда измеряли с одной стороны была корреляция не 1 - позже

Выстрел 2: электричество

Однажды физик обнаружил у себя дома удивительный артефакт: электрическую розетку.

Про розетки он слышал, что в них есть электрический ток, которым как раз и питаются электроприборы. А у тока есть напряжение. Которое, вроде бы, в розетках переменное. Так вот, интересно, связано ли это переменное напряжение со временем? Или же, напротив, оно там совершенно хаотичное и меняется как попало?

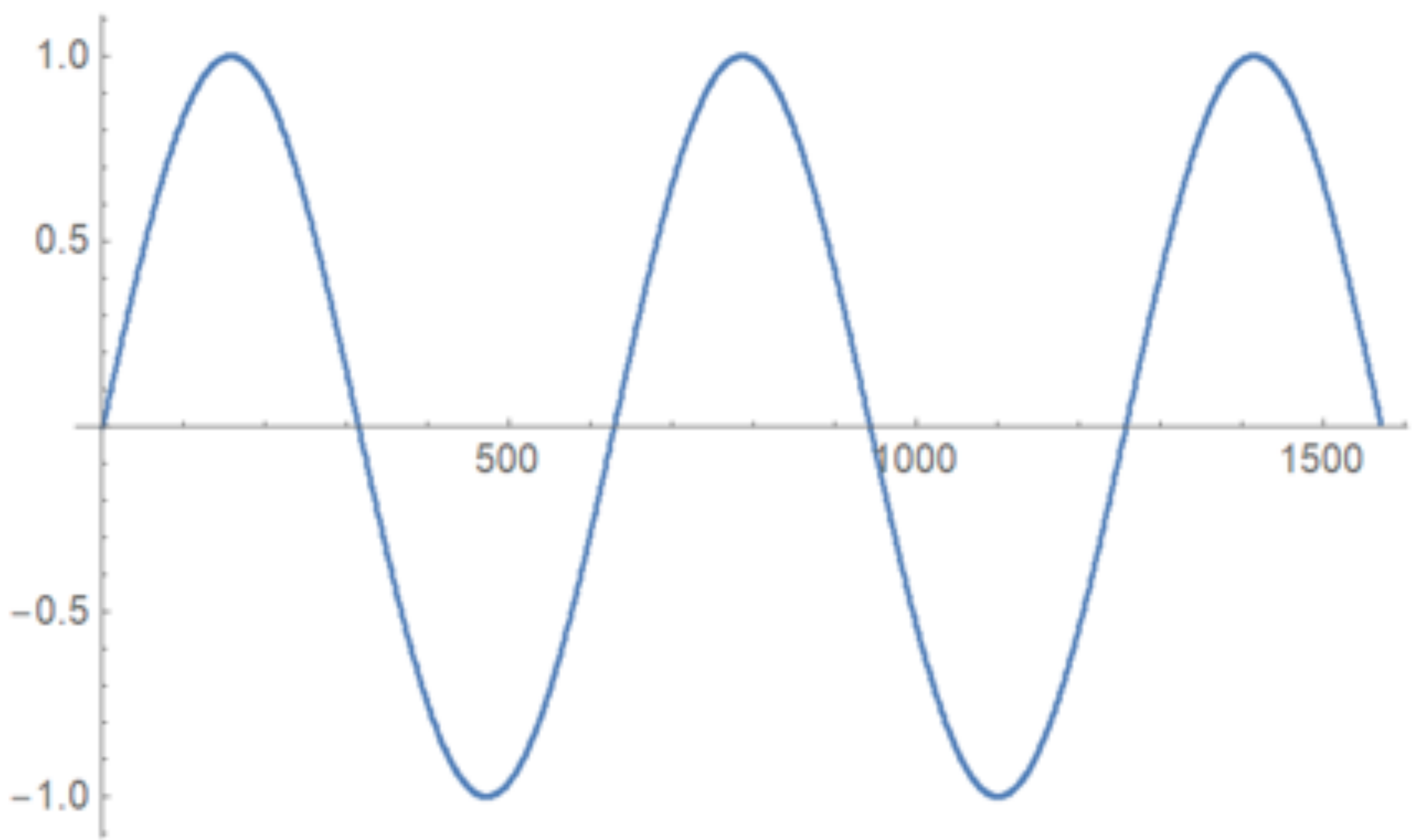
Для ответа на этот вопрос физик собрал хитрую схему из компьютера и вольтметра, которая через равные промежутки времени измеряет напряжение в сети.

Выстрел 2:

Какую корреляцию получит физик?

Выстрел 2:

Зависит от периода времени, через который он будет измерять



Выстрел 2:

Зависит от периода времени, через который он будет измерять

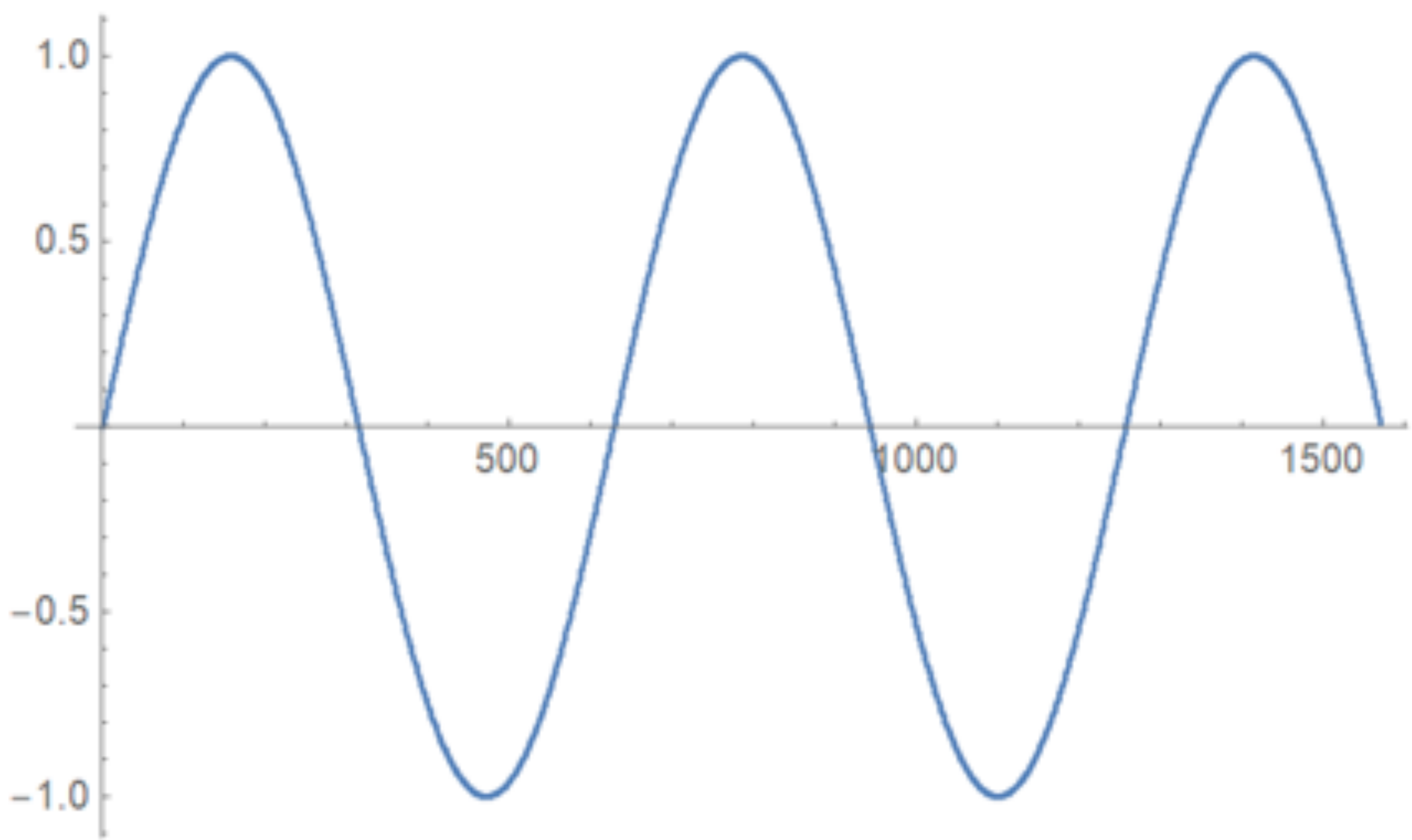


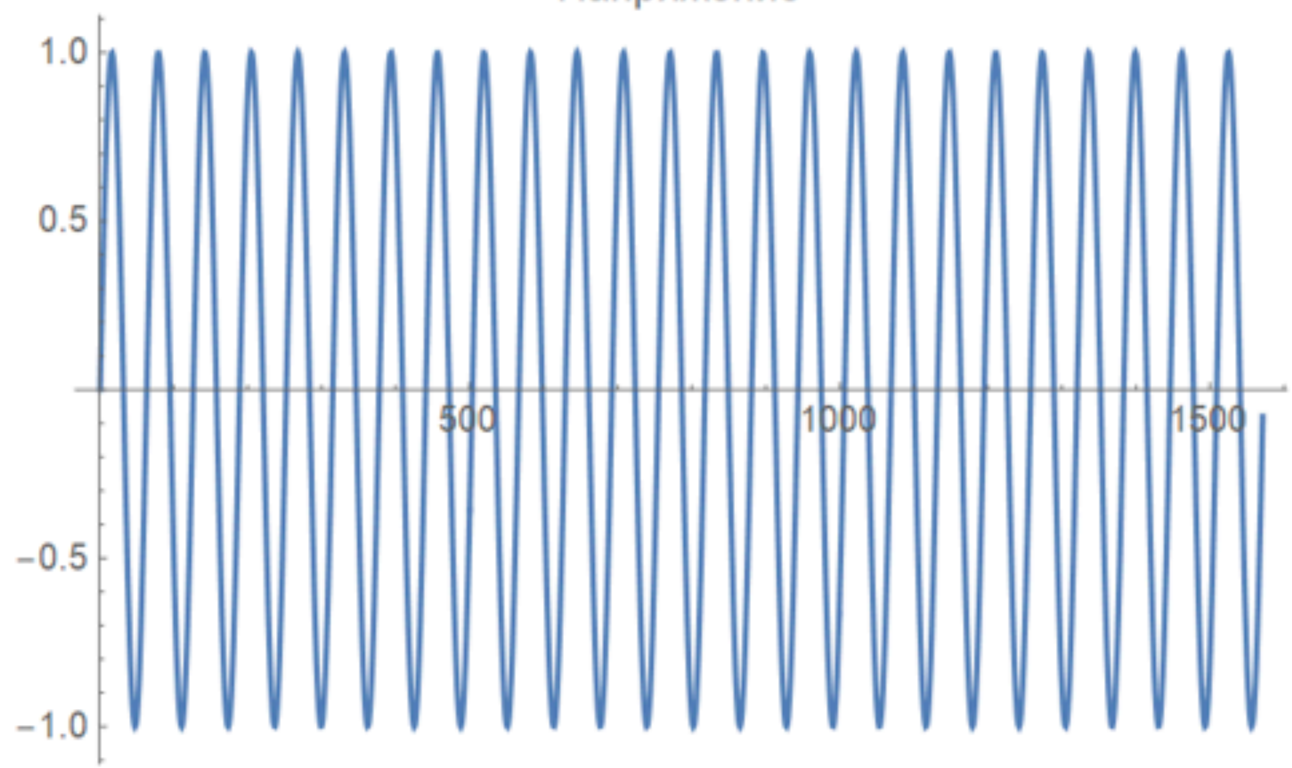
График напряжения переменного тока от времени

Выстрел 2:

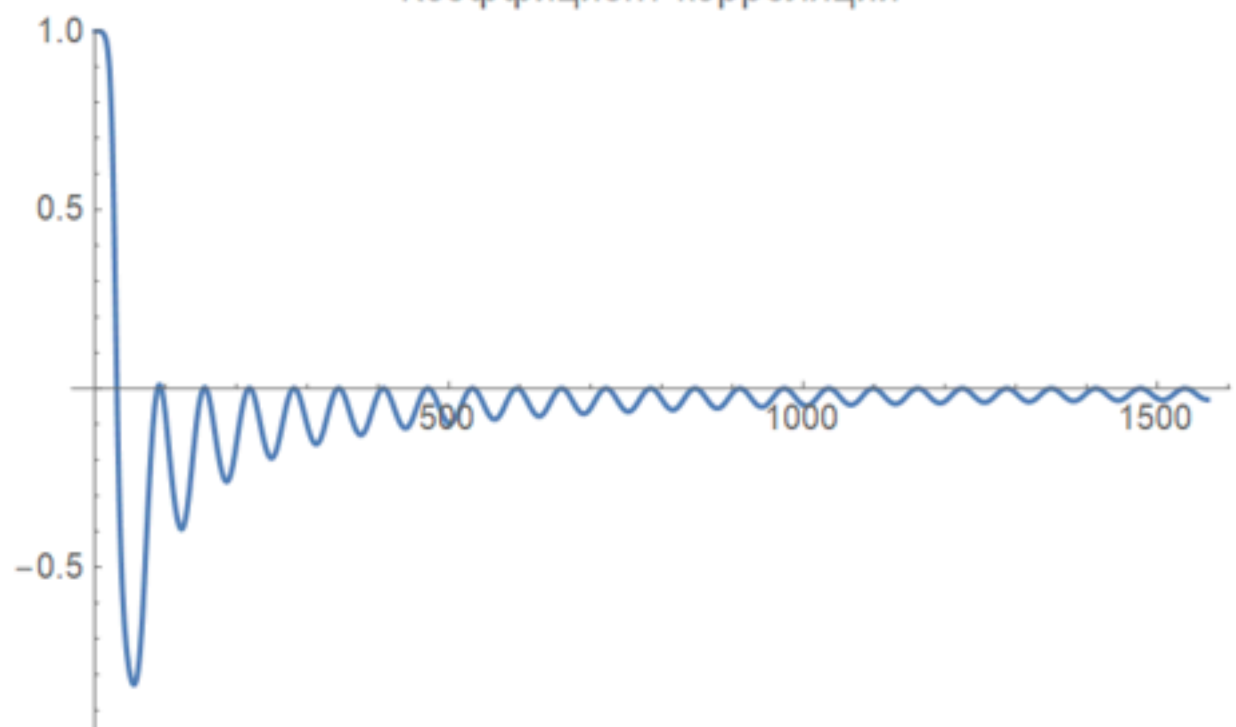
Если попасть в период, то

$r \to 0$

Напряжение

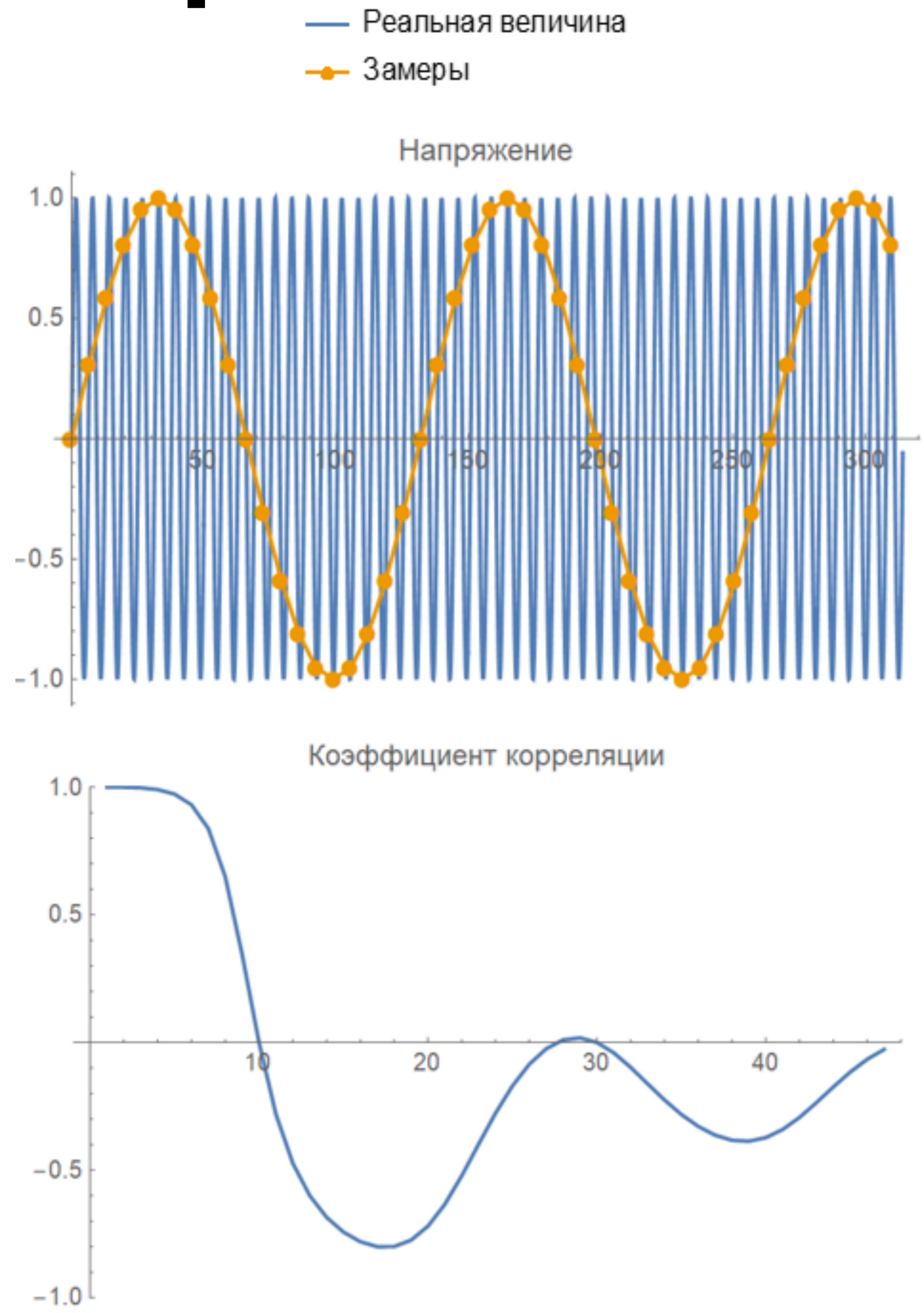


Коэффициент корреляции



Выстрел 2:

Если не попасть в период, то получим случайное число от 0 до 1



Выстрел 3: неубедительный экстрасенс

Однажды к группе заинтересованных в изучении паранормальных явлений пришёл человек, обладающий удивительной способностью — он абсолютно безошибочно умел угадывать выпавшее на кубиках.

Правда, по воле паранормальных сил эта способность у него была выражена в весьма странной форме: если кто-то собирался бросить четыре кубика, то Космос тут же шептал экстрасенсу сумму выпавшего на них в тридцатой степени. Увы, экстрасенс был безграмотен (он просто слушал голос Космоса), а потому извлекать корень тридцатой степени не умел.

Впрочем, паранормологи этого делать тоже не умели: ни извлекать корень тридцатой степени, ни возводить в тридцатую степень. Однако они предположили, что его предсказания всё равно должны коррелировать с суммой выпавшего на кубиках (которую они считать всё-таки умели) — ведь между числом и его тридцатой степенью имеется однозначная математическая связь.

Не поленившись, паранормологи провели 10 000 испытаний.

Выстрел 3: неубедительный экстрасенс

В результате корреляция была равна 0.2. Значимо ли это?

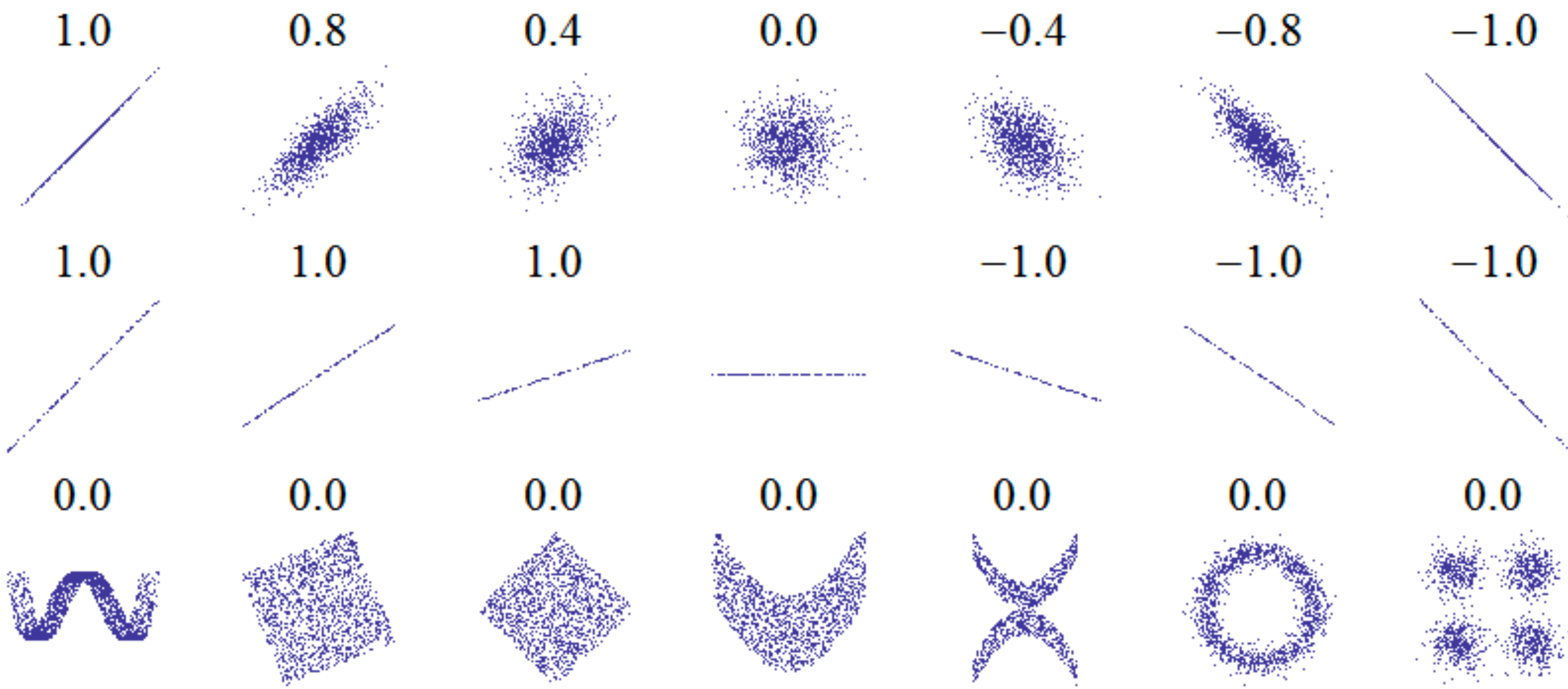
Выстрел 3: неубедительный экстрасенс

В результате корреляция была равна 0.2. Значимо ли это?

Экстрасенс угадал все разы. 100% точность. Но между суммой и ее 30й степенью зависимость не линейная

Аналогично объясняется первая часть выстрела 1.

Корреляционные поля



Значимость коэффициента корреляции

Коэффициент корреляции двух выборок - случайная величина. Как она распределена? Сложно.

Если мы предполагаем, что настоящий коэффициент корреляции (посчитанный для генеральных совокупностей) равен 0 (наша H0), то вот такая величина распределена по Стьюденту

$$t = \frac{r}{\sqrt{1 - r^2}} \sqrt{n - 2} \sim t_{n-2}$$

Можем посчитать насколько значимо отличается наш коэффициент корреляции от 0. Это поможет в выстреле 3 не прогнать правильного экстрасенса

Коэффициент Спирмена

Заменим Xi на его ранг - место в отсортированном X (например, если Xi меньше всех элементов из X, кроме двух, то его ранг - 3).
Аналогично для Yi.
Посчитаем корреляцию рангов - получится коэффициент Спирмена, который учитывает любую **монотонную зависимость**.
От прочих недостатков корреляции это его не избавляет

Выстрел 4: убедительный экстрасенс

К тем же самым паранормологам через некоторое время пришёл ещё один посетитель. Он тоже имел паранормальную связь с кубиками, но его метод был проще: он бросал один кубик, а потом предлагал экспериментатору бросить второй. Благодаря тому, что в его руках была заключена особая магия, первый бросок кубика мистическим образом влиял на всю ситуацию в целом.

О нет, выпавшее на втором кубике он не брался предсказать, однако утверждал, что первый бросок окажет очень сильное влияние на второй бросок, суть коего влияния он, впрочем, не брался объяснить. Но предлагал просто взять и посчитать корреляцию выпавшего на первом кубике с суммой выпавшего на двух кубиках.

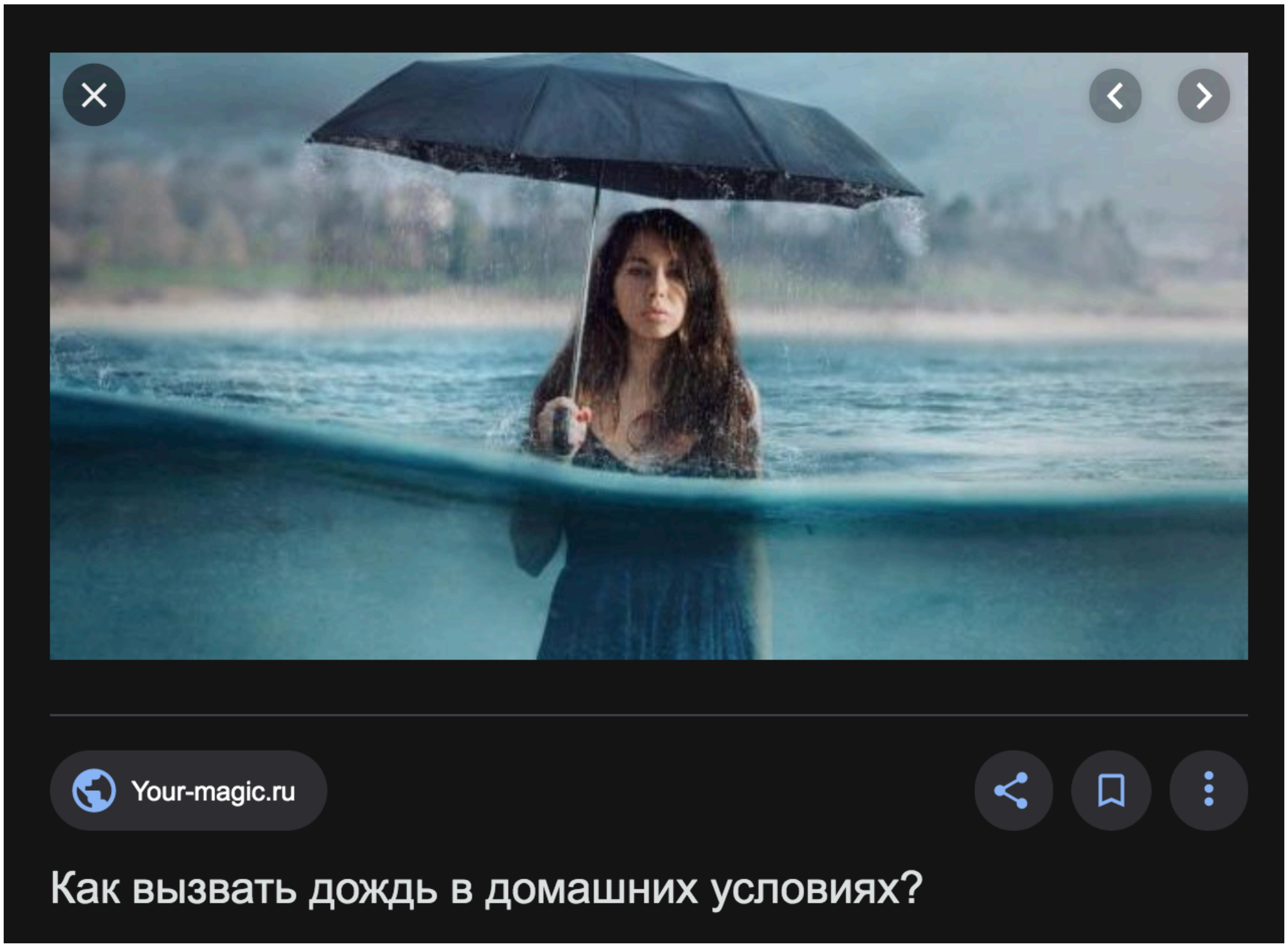
Чему оказался равен коэффициент корреляции?

0.75

Выстрел 4

Разумеется, выпавшее на первом кубике влияет на сумму, поскольку оно входит туда как слагаемое. Однако корреляция показала всего лишь линейную связь первого со вторым. Выпавшего на первом кубике и случайной величины, полученной прибавлением к уже выпавшему второй случайной величины.

Выстрел 5: хороший способ вызывать дождь



Выстрел 5: хороший способ вызывать дождь

Как-то раз антрополог наблюдал за обычным городским жителем (конечно же, с его, этого жителя, согласия). И через некоторое время он заметил удивительную закономерность: если этот житель брал с собой зонт, то в этот день шёл дождь.

Конечно, так случалось не каждый раз: иногда объект наблюдения брал зонт, но дождь не шёл, иногда не брал, но дождь всё равно шёл, — однако слишком часто наличие зонта с наличием дождя и отсутствие зонта с отсутствием дождя случались одновременно.

Корреляция была очень высокой: 0,95. Эти два события однозначно были связаны.

Выстрел 5: хороший способ вызывать дождь

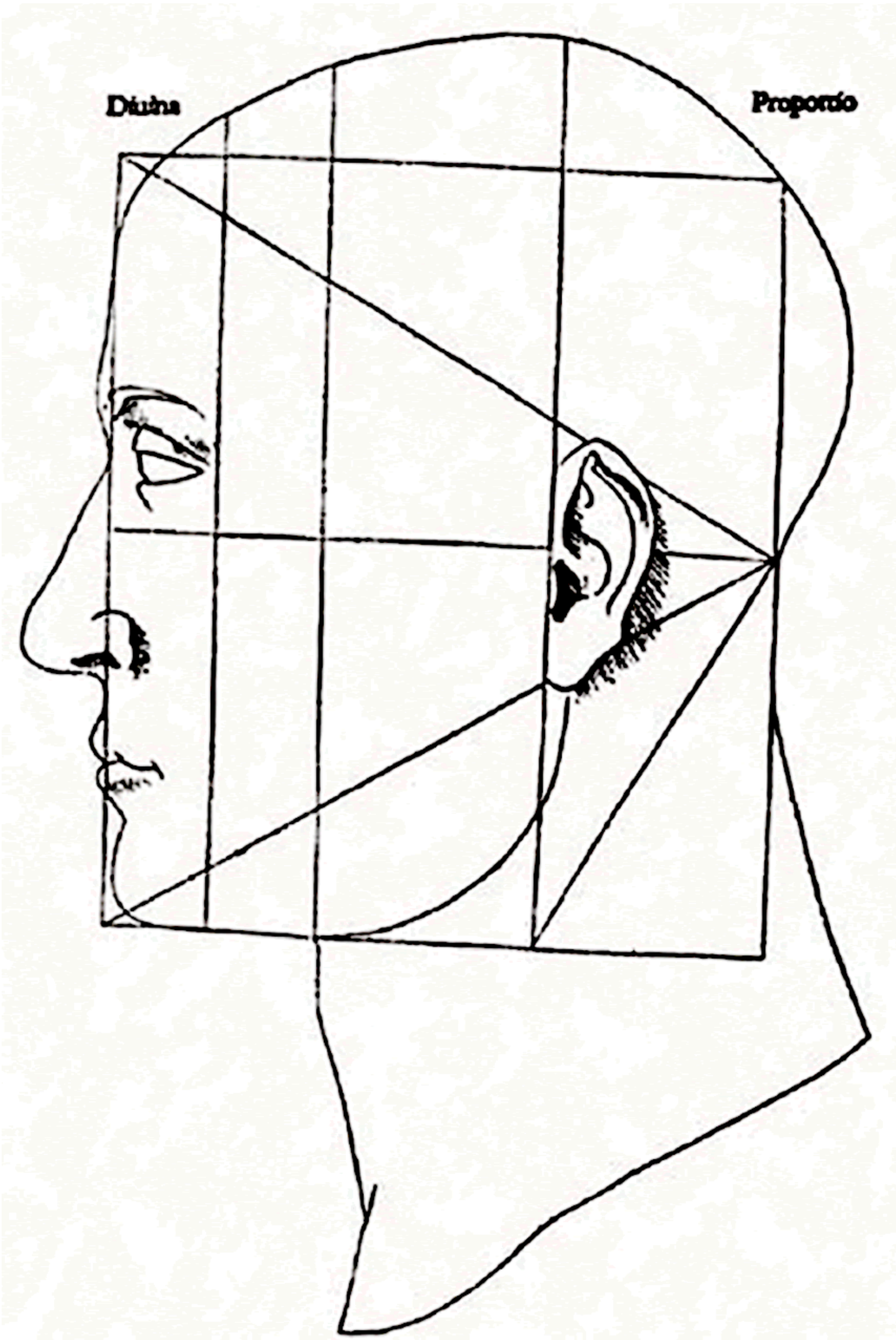
Как-то раз антрополог наблюдал за обычным городским жителем (конечно же, с его, этого жителя, согласия). И через некоторое время он заметил удивительную закономерность: если этот житель брал с собой зонт, то в этот день шёл дождь.

Конечно, так случалось не каждый раз: иногда объект наблюдения брал зонт, но дождь не шёл, иногда не брал, но дождь всё равно шёл, — однако слишком часто наличие зонта с наличием дождя и отсутствие зонта с отсутствием дождя случались одновременно.

Корреляция была очень высокой: 0,95. Эти два события однозначно были связаны.

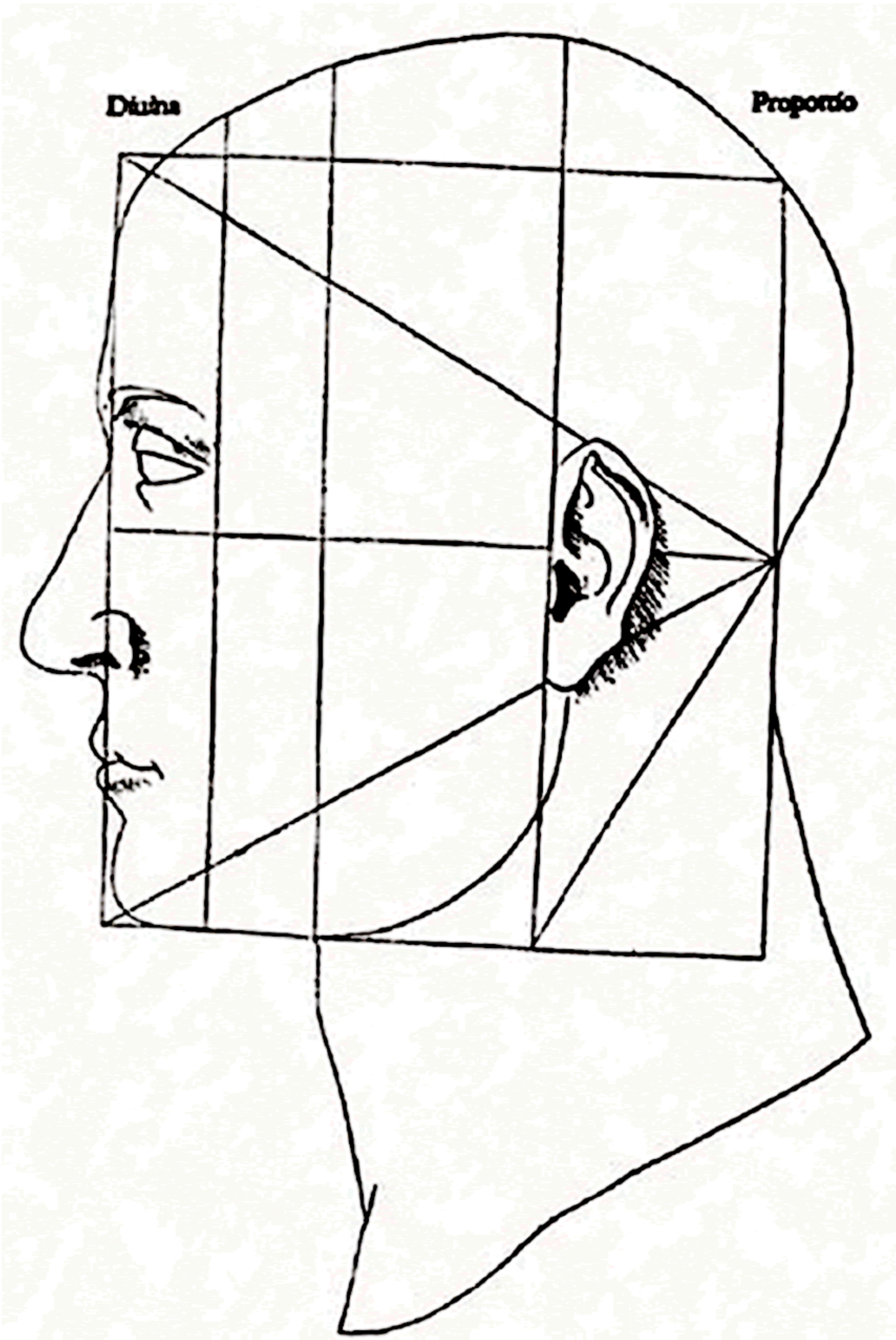
correlation doesn't imply causation

Физиогномика



Люди с “детскими” чертами лица склонны вести себя более по-детски

Физиогномика



Люди с “детскими” чертами лица склонны вести себя более по-детски

Очевидно, что черты лица на личность не завязаны - закладываются в разное время

Людей с “детскими” чертами лица воспринимают более по-детски и соответственно относятся - со временем они адаптируются к этому, как вариант - ведут себя более по-детски

“Важные” корреляции

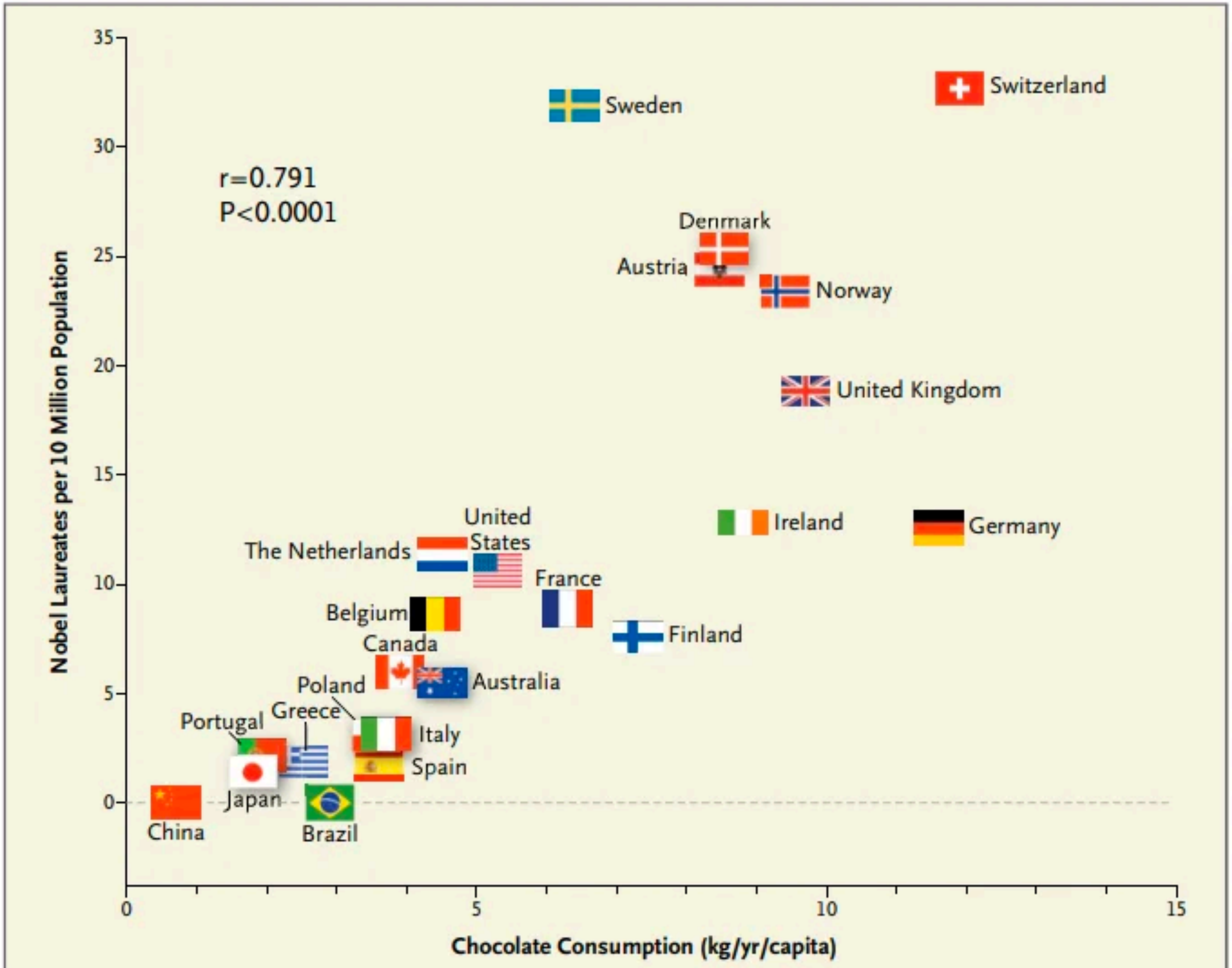


Figure 1. **Correlation between Countries’ Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

“Важные” корреляции

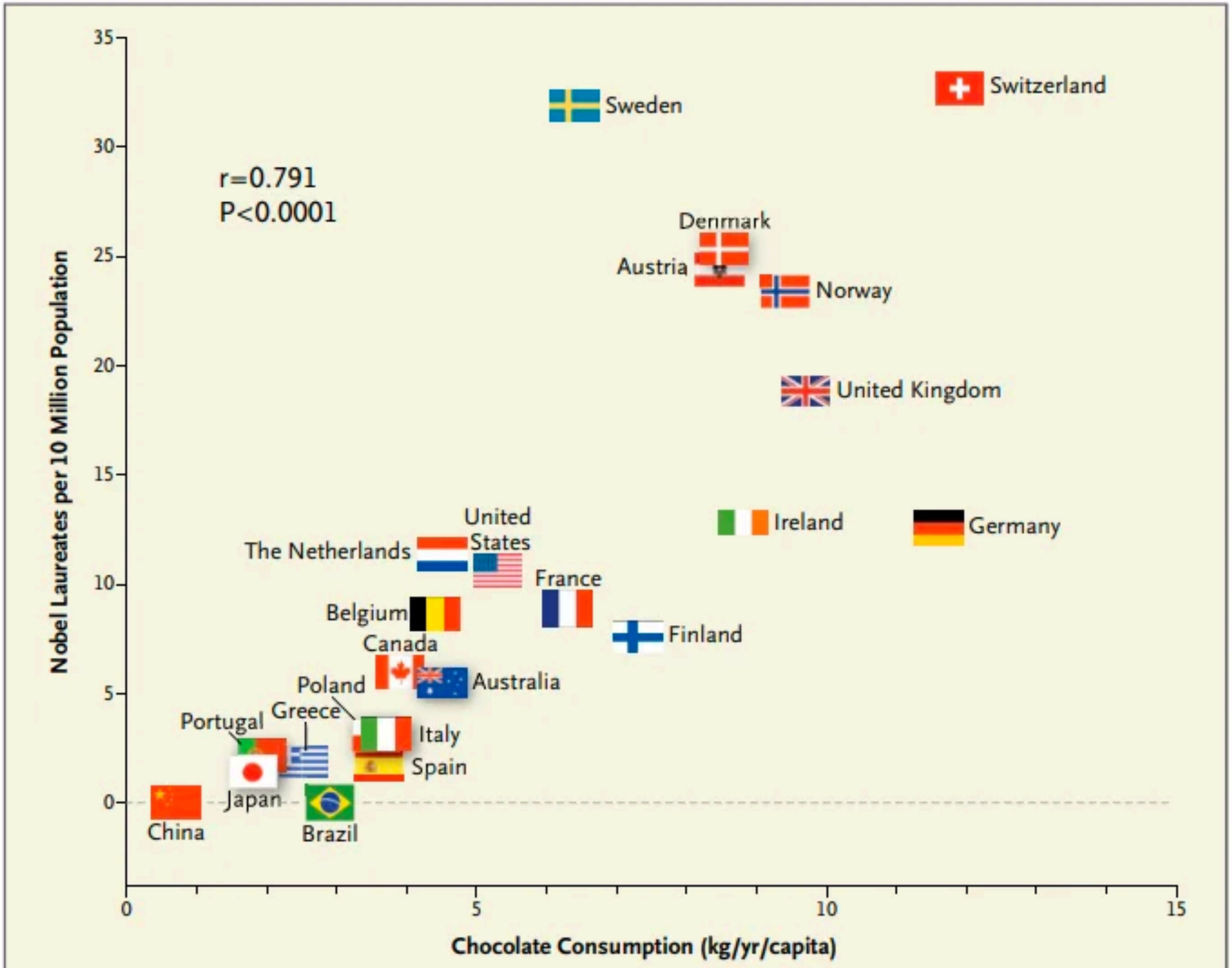
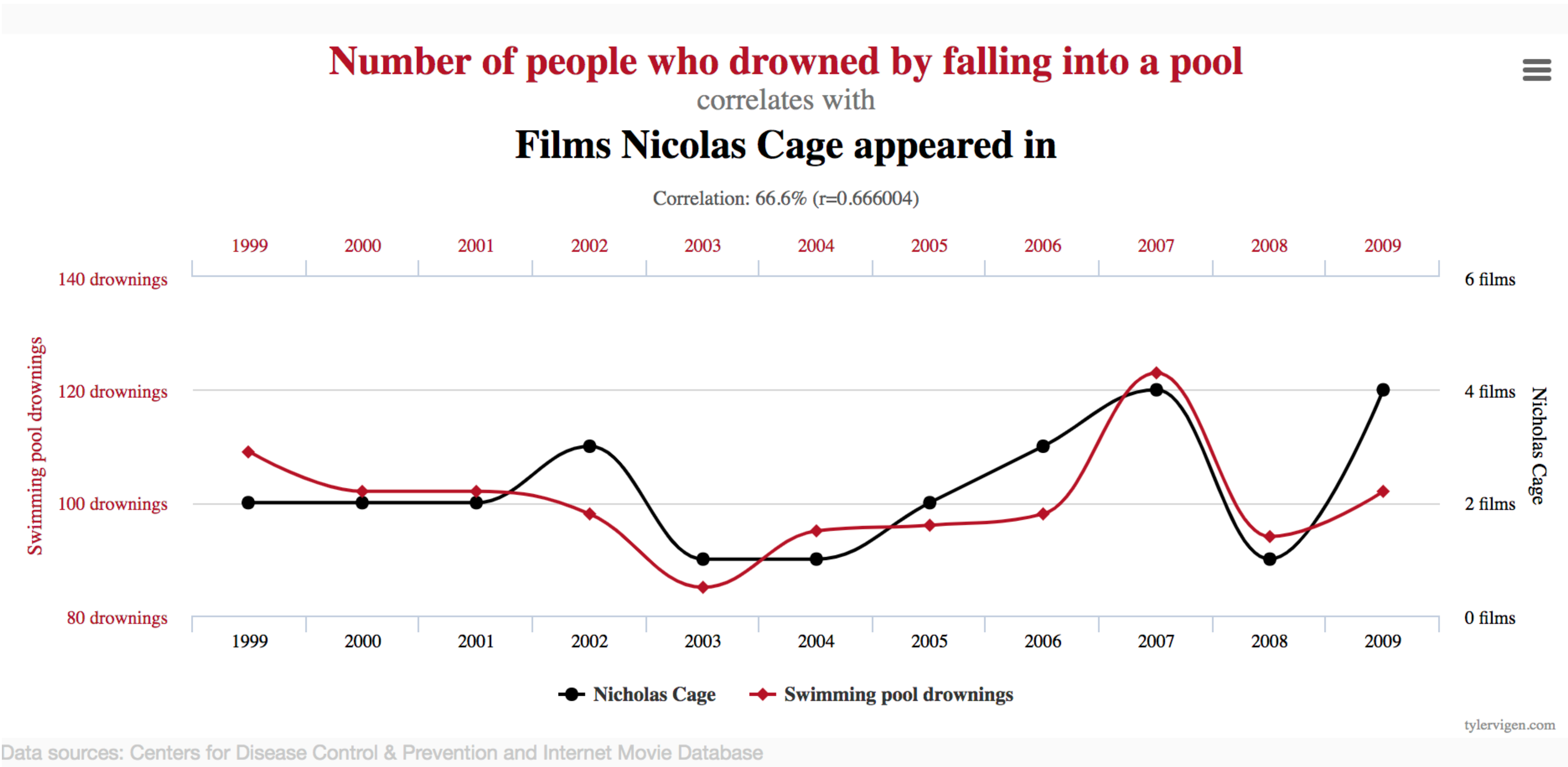


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

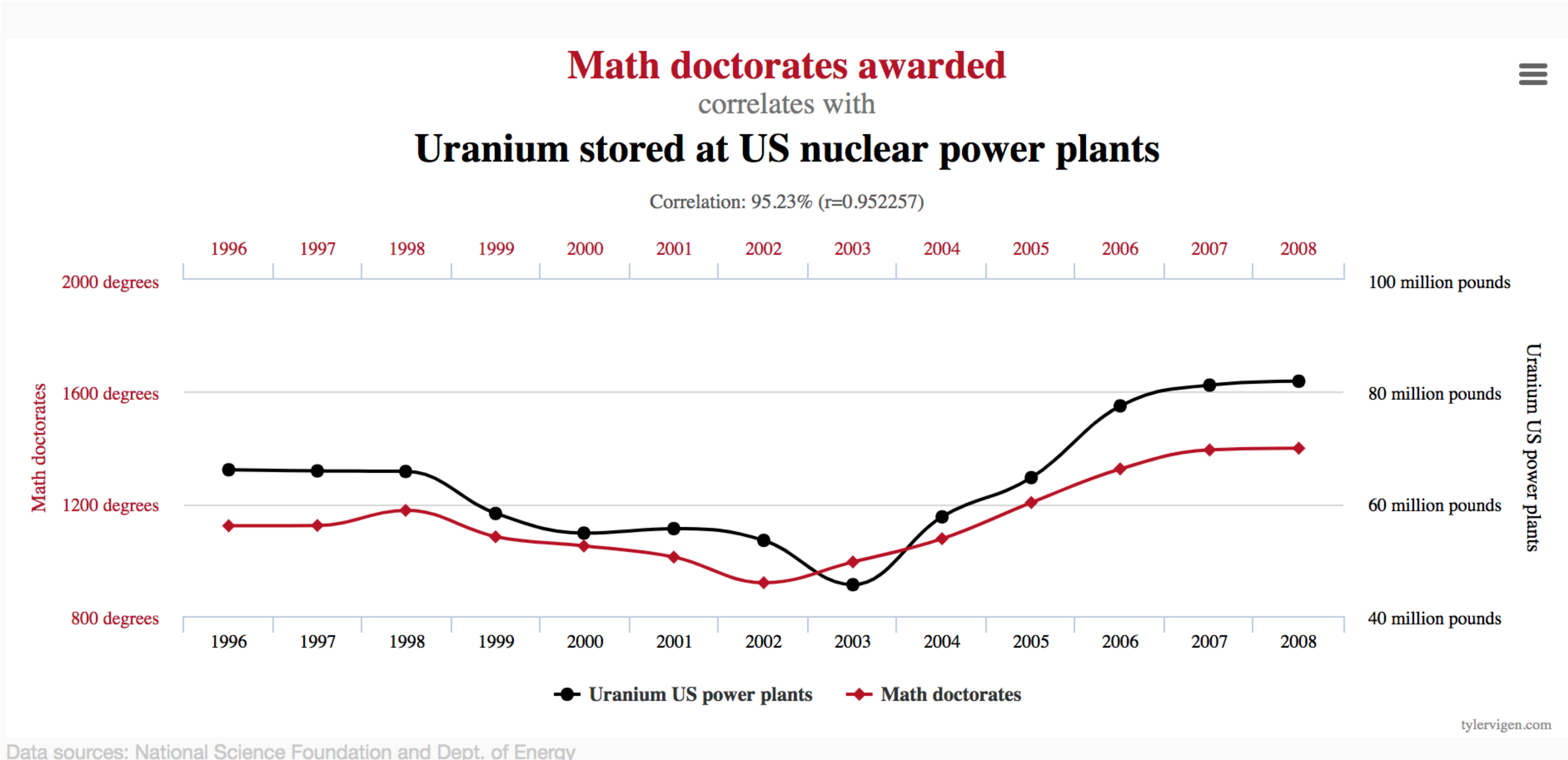
Здесь “хотя бы” есть третья переменная, которая все объясняет - богатство стран

“Важные” корреляции



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

“Важные” корреляции

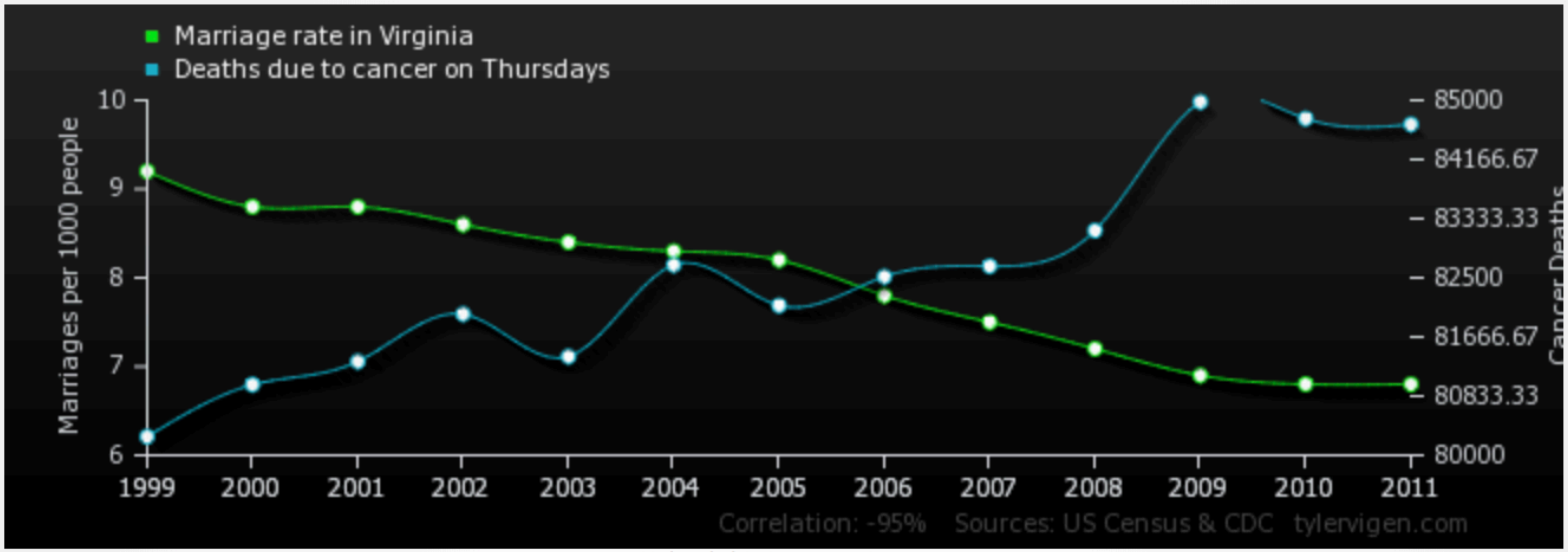


“Важные” корреляции

Marriage rate in Virginia

inversely correlates with

Deaths due to cancer on Thursdays



Upload this image to imgur

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Marriage rate in Virginia *Marriages per 1000 people (US Census)*	9.2	8.8	8.8	8.6	8.4	8.3	8.2	7.8	7.5	7.2	6.9	6.8	6.8
Deaths due to cancer on Thursdays *Cancer Deaths (CDC)*	80,262	80,994	81,321	81,988	81,390	82,682	82,109	82,516	82,663	83,166	84,978	84,742	84,660

Correlation: -0.951857

“Важные” корреляции

Как объяснить предыдущие три корреляции?

“Важные” корреляции

Как объяснить предыдущие три корреляции?

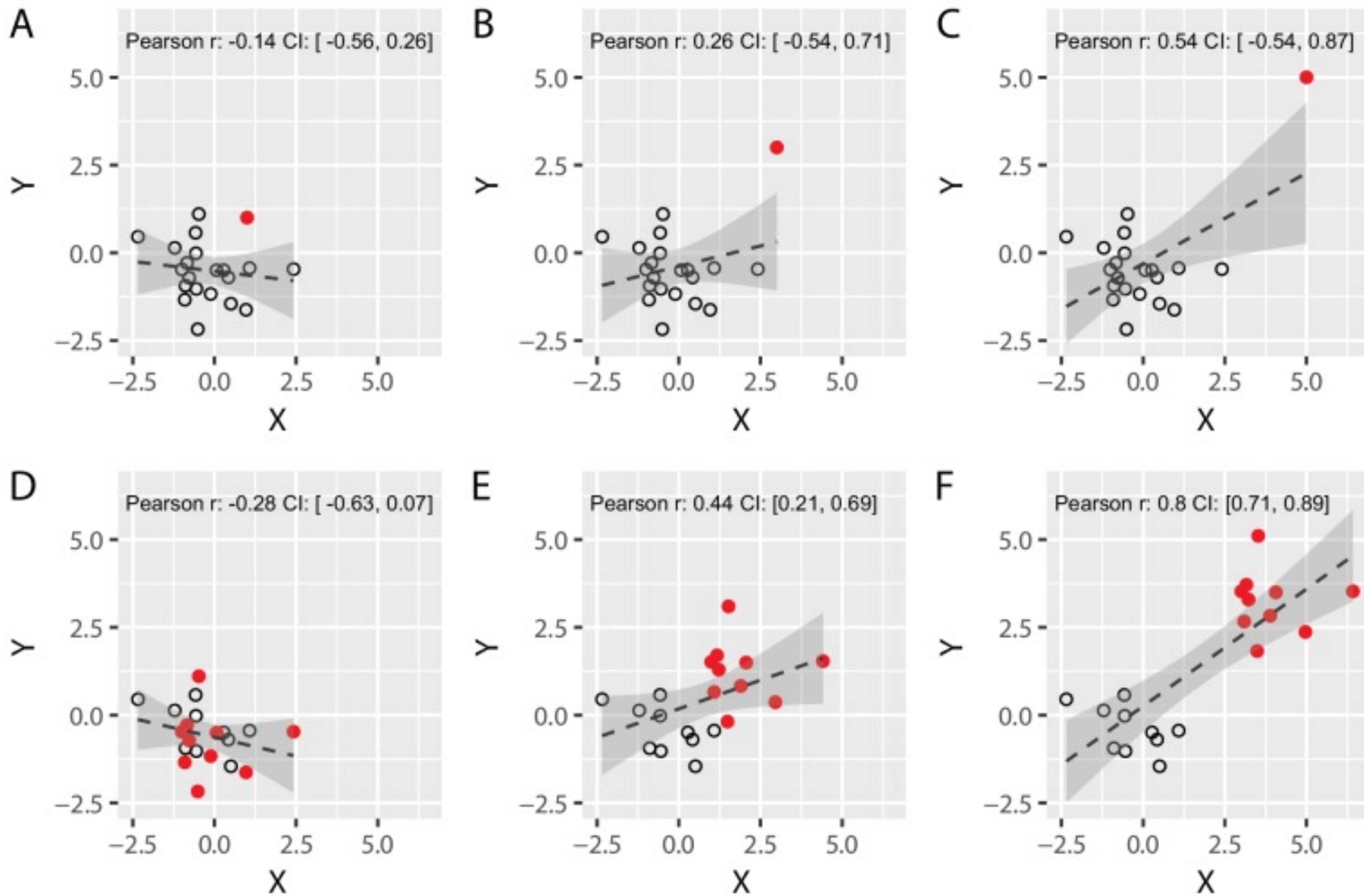
Указание:
Выборочный коэффициент корреляции на выборках - тоже случайная величина

“Важные” корреляции

Как объяснить предыдущие три корреляции?

Выборочный коэффициент корреляции - тоже случайная величина. Есть вероятность, особенно на **маленьких выборках**, получить большой и даже “значимый” коэффициент корреляции.
Если посчитать корреляцию многих-многих событий, то из-за множественного тестирования что-то, да найдется

Выстрел 6. Влияние выбросов и кластеров в данных



Выстрел 7: дискриминация

	Количество кандидатов	**Доля поступивших**
Мужчины	8442	**44%**
Женщины	4321	35%

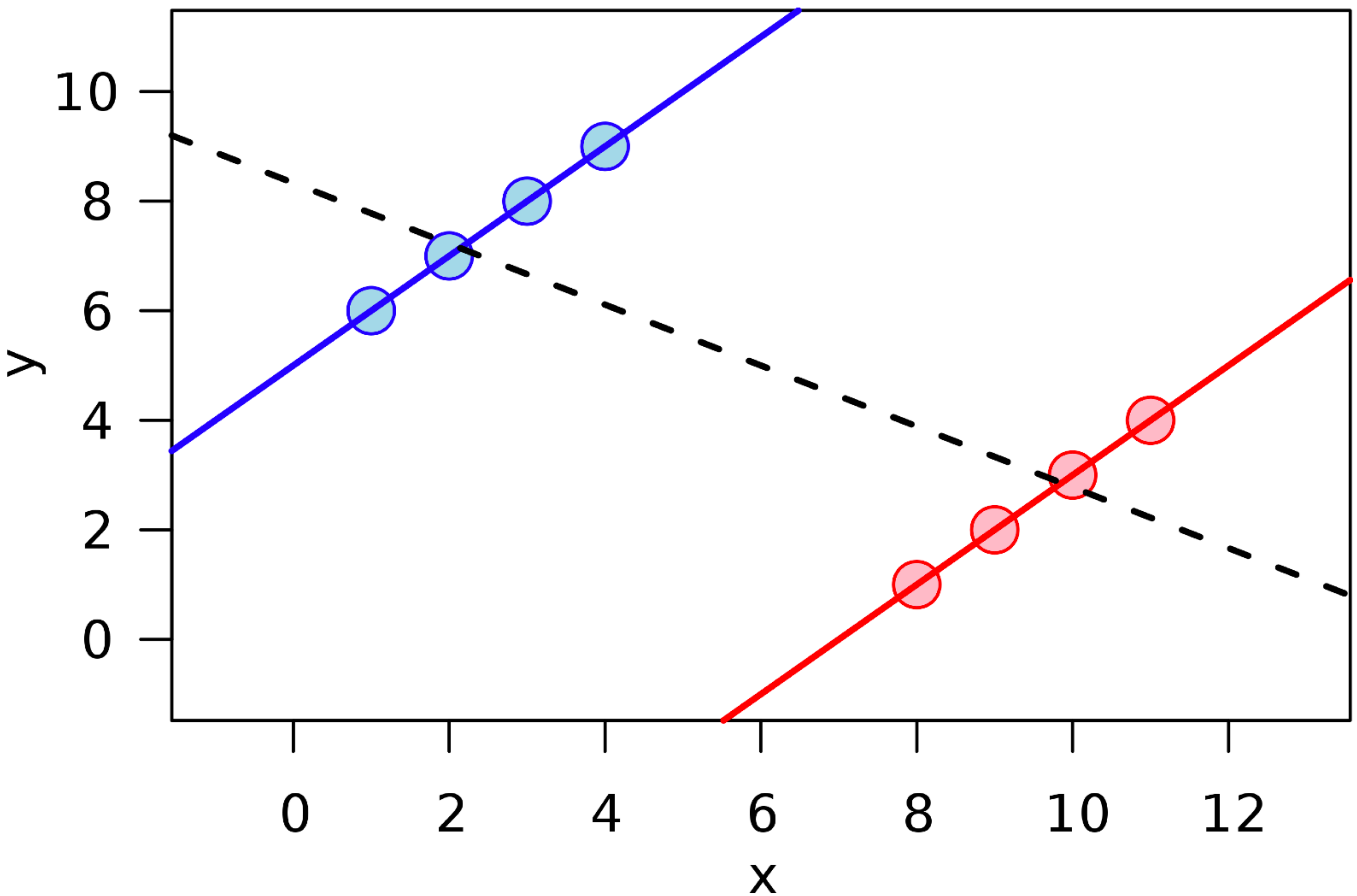
Дискриминируют женщин

Выстрел 7: дискриминация

	Мужчины		Женщины	
	Количество кандидатов	**Доля поступивших**	Количество кандидатов	**Доля поступивших**
Кафедра 1	825	62%	108	**82%**
Кафедра 2	560	63%	25	**68%**
Кафедра 3	325	**37%**	593	34%
Кафедра 4	417	33%	375	**35%**
Кафедра 5	191	**28%**	393	24%

Дискриминируют мужчин)

Выстрел 6: парадокс Симпсона



Для разделенных синей и красной выборок корреляция имеет один знак, для объединенных - разный

Какой коэффициент корреляции могут иметь зависимые величины?

Какой коэффициент корреляции могут иметь полностью зависимые величины?

Любой

	z	X = sin(z)	Y = cos(z/2)
1	0	0	1
2	pi /2	1	0.7
3	Pi	0	0

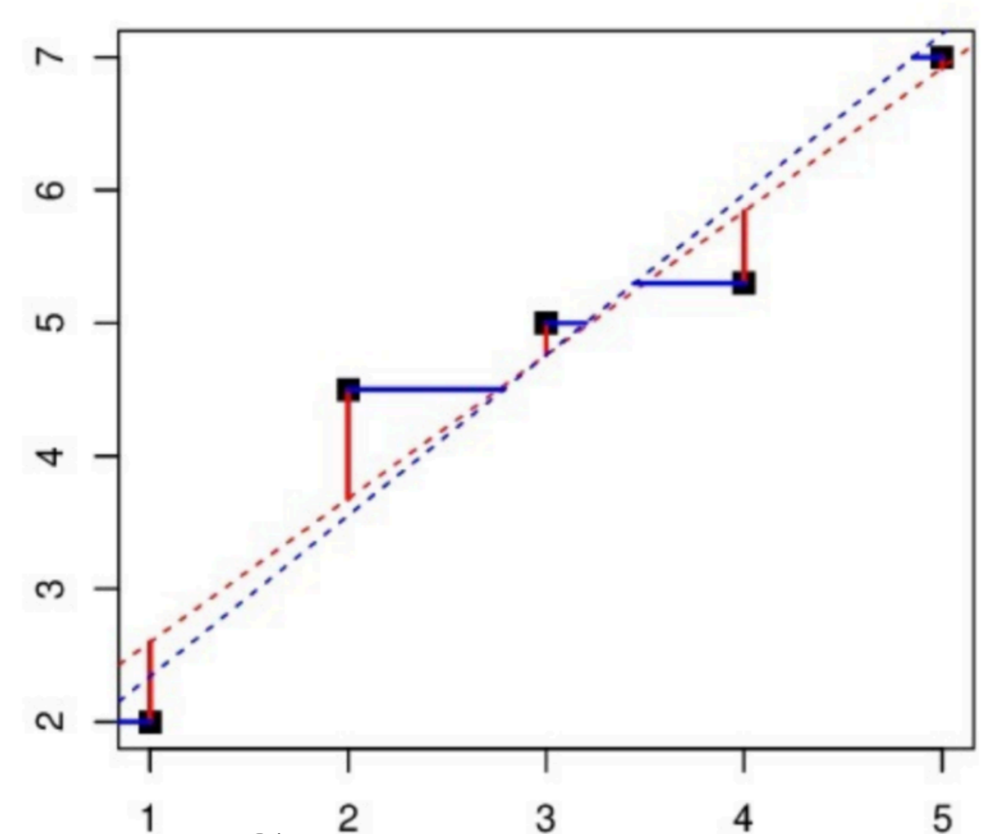
X и Y полностью зависимы, но корреляция “лишь” 0.23

Связь коэффициента корреляции Пирсона и коэффициента линейной регрессии

$$r = \frac{s_{xy}}{s_x s_y}$$

$$b = \frac{s_{xy}}{s_x^2} = r \cdot \frac{s_y}{s_x}$$

Связь коэффициента корреляции Пирсона и коэффициента линейной регрессии



$$y = a + bx$$

$$x = a' + b'y$$

$$b = \frac{s_{xy}}{s_x^2} \qquad b' = \frac{s_{xy}}{s_y^2} \qquad r = \frac{s_{xy}}{s_x s_y}$$

$$R^2 = r^2 = b \cdot b'$$

Коэффициент детерминации, доля дисперсии y объясняемая линейным вкладом переменной x

Коэффициент корреляции является мерой

A. того, насколько изменение одной переменной вызывает изменение другой
B. линейной взаимосвязи между двумя категориальными переменными
C. взаимосвязи между числовыми переменными, не обязательно линейной
D. линейной взаимосвязи между двумя числовыми переменными
E. измеряющей в процентах линейную взаимосвязь между числовыми переменными

Коэффициент корреляции между энергией, измеряемой в Джоулях, и температурой, измеряемой в градусах Цельсия, равен 0.23. Чему будет равен коэффициент корреляции между энергией, измеряемой в калориях, и температурой, измеряемой в градусах Фаренгейта? Связь между шкалами описывается уравнениями 1 cal = 4.184 J и $T_F = 1.8 * T_C + 32$.

A. 0.06

B. 0.10

C. 0.23

D. 0.53

E. 0.96

Коэффициент корреляции между энергией, измеряемой в Джоулях, и температурой, измеряемой в градусах Цельсия, равен 0.23. Чему будет равен коэффициент корреляции между энергией, измеряемой в калориях, и температурой, измеряемой в градусах Фаренгейта? Связь между шкалами описывается уравнениями 1 cal = 4.184 J и $T_F = 1.8 * T_C + 32$.

A. 0.06

B. 0.10

C. 0.23

D. 0.53

E. 0.96

Коэффициент корреляции инвариантен к изменению масштаба переменных и к прибавлению к ним константы

4. Коэффициент корреляции Пирсона, r, показывает силу и направление линейной зависимости между переменными. Однако надежность его предсказаний зависит от числа наблюдений. Например, маленькие выборки несвязанных между собой количеств – например, размер обуви человека (x) и средний счет за электричество, который он оплачивает в этом месяце – могут давать ненулевые коэффициенты корреляции из-за ошибки семплирования.

   (a) **(1 pt)** Рассмотрим величины x и y, которые рассмотрены выше. Выборка объема $n = 12$ дает коэффициент детерминации $R^2 = 0.5$. Чему равно пи-велью при тестировании гипотезы о том, что между x и y нет связи? Интерпретируйте свой результат в контексте задачи.

   (b) **(1 pt)** Для выборки объема $n = 12$, чему равно минимальное значение коэффициента корреляции, которое будет значимо отличаться от нуля на 5% уровне значимости? Используйте двусторонний тест.

   (c) **(1 pt)** При каком наименьшем объеме выборки значение коэффициента корреляции $r = 0.1$ становится статистически отличимо от нуля на 5% уровне значимости? Используйте двусторонний тест.

4. Коэффициент корреляции Пирсона, r, показывает силу и направление линейной зависимости между переменными. Однако надежность его предсказаний зависит от числа наблюдений. Например, маленькие выборки несвязанных между собой количеств – например, размер обуви человека (x) и средний счет за электричество, который он оплачивает в этом месяце – могут давать ненулевые коэффициенты корреляции из-за ошибки семплирования.

   (a) **(1 pt)** Рассмотрим величины x и y, которые рассмотрены выше. Выборка объема $n = 12$ дает коэффициент детерминации $R^2 = 0.5$. Чему равно пи-велью при тестировании гипотезы о том, что между x и y нет связи? Интерпретируйте свой результат в контексте задачи.

   (b) **(1 pt)** Для выборки объема $n = 12$, чему равно минимальное значение коэффициента корреляции, которое будет значимо отличаться от нуля на 5% уровне значимости? Используйте двусторонний тест.

   (c) **(1 pt)** При каком наименьшем объеме выборки значение коэффициента корреляции $r = 0.1$ становится статистически отличимо от нуля на 5% уровне значимости? Используйте двусторонний тест.

$$t = \frac{r}{\sqrt{1 - r^2}}\sqrt{n - 2} \sim t_{n-2}$$

$$\Pr(R^2 \geqslant 0.5) = 2\Pr(r > \sqrt{0.5}) = 2 * \Pr(t(10) > \frac{\sqrt{0.5}\sqrt{10}}{\sqrt{0.5}}) = 2\Pr(t(10) > 3.16)$$

$t_{0.005}(10) = 3.17 > 3.16$, but $t_{0.01}(10) = 2.76 < 3.16$. Hence

$\Pr(R^2 \geqslant 0.5) \in [0.01, 0.02]$.

Если регрессия y против x имеет наклон -2.54, а регрессия x против y имеет наклон -0.31, то коэффициент корреляции равен

A. -0.887

B. 0.887

C. -0.78

D. 0.78

E. None of the above

Если регрессия y против x имеет наклон -2.54, а регрессия x против y имеет наклон -0.31, то коэффициент корреляции равен

A. -0.887

B. 0.887

C. -0.78

D. 0.78

E. None of the above

$$R^2 = r^2 = b \cdot b'$$

Две выборки, X и Y, таковы, что $\overline{X} = 4$, $\overline{Y} = 3$, $s_x = 2.16$, $s_y = 2$. Что из следующего могло бы быть уравнением регрессии Y от X?

A. $Y = 0.11 + 0.32 * X$

B. $Y = 0.84 + 0.54 * X$

C. $Y = -0.80 + 0.95 * X$

D. $Y = -1.48 + 1.12 * X$

E. $Y = -1.92 + 1.23 * X$

Две выборки, X и Y, таковы, что $\overline{X} = 4$, $\overline{Y} = 3$, $s_x = 2.16$, $s_y = 2$. Что из следующего могло бы быть уравнением регрессии Y от X?

A. $Y = 0.11 + 0.32 * X$

B. $Y = 0.84 + 0.54 * X$

C. $Y = -0.80 + 0.95 * X$

D. $Y = -1.48 + 1.12 * X$

E. $Y = -1.92 + 1.23 * X$

Через какую точку обязательно проходит регрессионная прямая?

Две выборки, X и Y, таковы, что $\overline{X} = 4$, $\overline{Y} = 3$, $s_x = 2.16$, $s_y = 2$. Что из следующего могло бы быть уравнением регрессии Y от X?

A. $Y = 0.11 + 0.32 * X$

B. $Y = 0.84 + 0.54 * X$

C. $Y = -0.80 + 0.95 * X$

D. $Y = -1.48 + 1.12 * X$

E. $Y = -1.92 + 1.23 * X$

Через какую точку обязательно проходит регрессионная прямая? - через $(\overline{X}, \overline{Y})$

Две выборки, X и Y, таковы, что $\overline{X} = 4$, $\overline{Y} = 3$, $s_x = 2.16$, $s_y = 2$. Что из следующего могло бы быть уравнением регрессии Y от X?

A. $Y = 0.11 + 0.32 * X$

B. $Y = 0.84 + 0.54 * X$

C. $Y = -0.80 + 0.95 * X$

D. $Y = -1.48 + 1.12 * X$

E. $Y = -1.92 + 1.23 * X$

Через какую точку обязательно проходит регрессионная прямая? - через $(\overline{X}, \overline{Y})$

Какие есть ограничения на коэффициент b?

$$b = r \cdot \frac{s_y}{s_x}$$

Две выборки, X и Y, таковы, что $\overline{X} = 4$, $\overline{Y} = 3$, $s_x = 2.16$, $s_y = 2$. Что из следующего могло бы быть уравнением регрессии Y от X?

A. $Y = 0.11 + 0.32 * X$

B. $Y = 0.84 + 0.54 * X$

C. $Y = -0.80 + 0.95 * X$

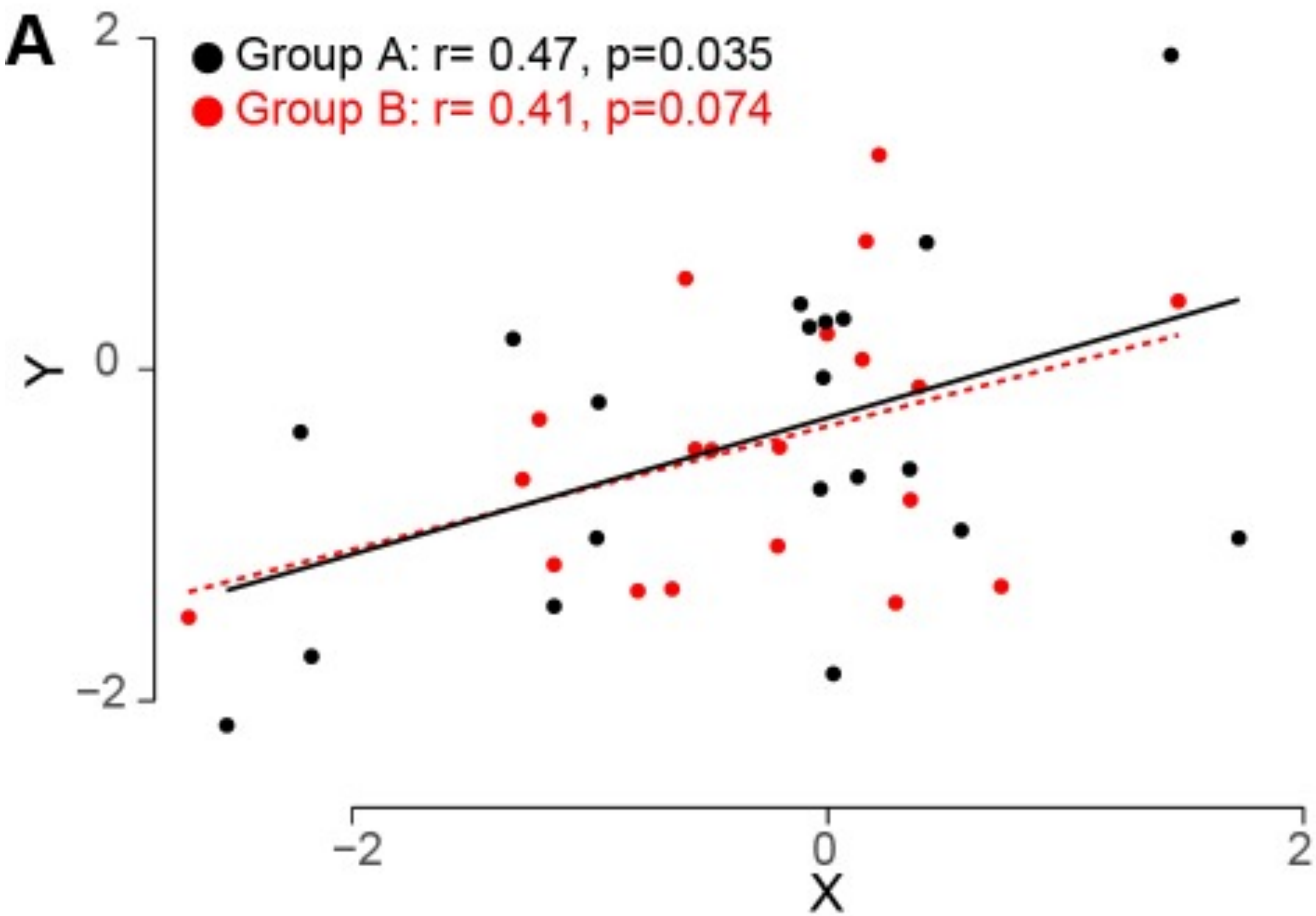
D. $Y = -1.48 + 1.12 * X$

E. $Y = -1.92 + 1.23 * X$

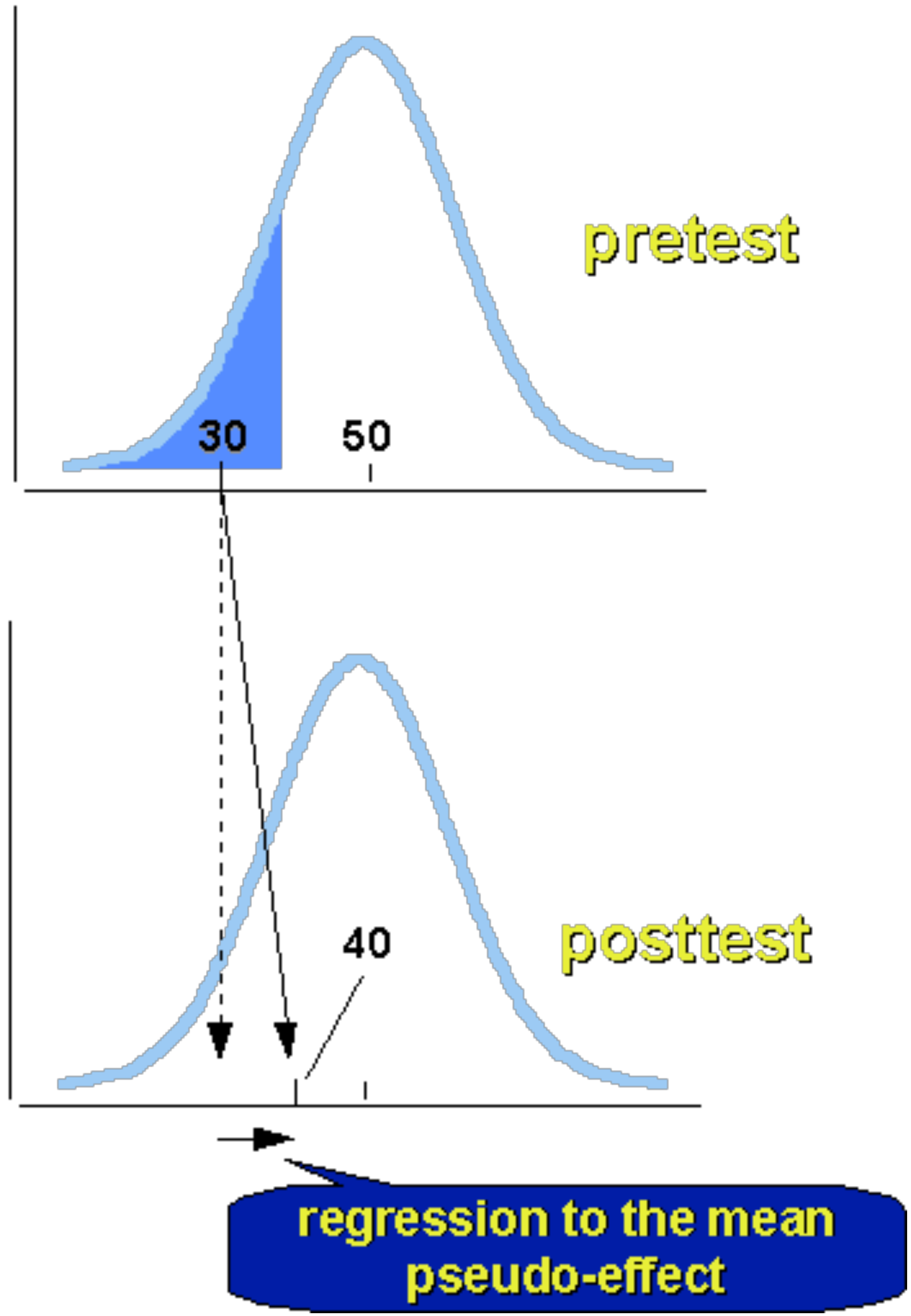
Через какую точку обязательно проходит регрессионная прямая? - через $(\overline{X}, \overline{Y})$

Какие есть ограничения на коэффициент b? $b = r \cdot \dfrac{s_y}{s_x}$ $\quad |r| < 1$

Выстрел, не только в корреляции



Можно ли сделать вывод, что сила связи в группах А и В отличаются?


pretest
30
50
posttest
40
regression to the mean
pseudo-effect