Многомерные случайные величины, независимость случайных величин

Сурова София, БПИ191

https://github.com/avorus

README.md

Теория вероятностей и математическая статистика 2021-2022

materials

задачник «Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами», 2007

homeworks

hw01 стр. 42-43 №4, 5, 6, 7, 11, 12, 13, 14, 15

`комбинаторика` `классическое определение вероятности`

hw02 стр. 44-52 №19, 22, 27, 28, 37, 41, 46, 82, 83

`формула умножения вероятностей` `условная вероятность` `независимость событий`

hw03 стр. 43-48 №18, 25, 26, 47, 50, 51, 52, 53, 54, 55

`схема Бернулли`

hw04 стр. 49-51 №62, 64, 65, 66, 71, 72, 76, 77, 78, 79

`формула полной вероятности` `формула Байеса`

hw05 стр. 90-94 №4, 9, 19, 28, 32, 53; стр. 90 №11, 12, 14

`дискретные СВ` `ряд распределения` `функция распределения` `ряд распределения` `математическое ожидание и дисперсия дискретной СВ`

Случайное событие vs Случайная величина

Случайной величиной называют числовую величину, значение которой зависит от того, какой именно *элементарный исход* произошел в результате эксперимента со случайным исходом. Множество всех значений, которые случайная величина может принимать, называют ***множеством возможных значений*** этой случайной величины.

Следовательно, для задания случайной величины необходимо каждому элементарному исходу поставить в соответствие число — значение, которое примет случайная величина, если в результате испытания произойдет именно этот исход. Обозначать случайные величины (СВ) принято греческими буквами ξ, η, ζ т.д.

Рассмотрим примеры.

Пример 4.1. В опыте с однократным бросанием игральной кости случайной величиной является число ξ выпавших очков. Множество возможных значений случайной величины ξ имеет вид

$$\{1;2;\ldots;6\}.$$

Если вспомнить, как выглядит *пространство элементарных исходов* в этом опыте, то будет очевидно следующее соответствие между элементарными исходами ω и значениями случайной величины ξ:

$$\begin{array}{ccccccc} \Omega & = & \{\omega_1 & \omega_2 & \ldots & \omega_6\} \\ & & \downarrow & \downarrow & \ldots & \downarrow \\ \xi & = & \{1 & 2 & \ldots & 6\}. \end{array}$$

Многомерные случайные величины

Определение 7.1. ***Многомерной (n-мерной) случайной величиной*** (или ***n-мерным случайным вектором***) называется вектор $\xi = (\xi_1, \quad \ldots, \quad \xi_n)$, компонентами которого являются случайные величины $\xi_1 = \xi_1(\omega), \quad \ldots, \quad \xi_n = \xi_n(\omega)$, определённые на одном и том же *вероятностном пространстве* $(\Omega, \mathscr{A}, \mathsf{P})$.

Пример 7.1. Отклонение точки разрыва снаряда от точки прицеливания при стрельбе по плоской цели можно задать двумерной случайной величиной (ξ, η), где ξ — отклонение по дальности, а η — отклонение в боковом направлении.

При стрельбе по воздушной цели необходимо рассматривать трехмерный случайный вектор (ξ, η, ζ), где ξ, η, ζ — координаты отклонения точки разрыва зенитного снаряда от точки прицеливания в некоторой пространственной системе координат.

Пример 7.2. При испытании прибора на надежность совокупность внешних воздействий в некоторый момент времени можно описать случайным вектором $(\xi, \eta, \zeta, \ldots)$. Здесь, например, ξ — температура окружающей среды, η — атмосферное давление, ζ — амплитуда вибрации платформы, на которой установлен прибор и т.д. Размерность этого вектора зависит от количества учитываемых факторов и может быть достаточно большой. #

Функция распределения

$F(x) = P\{\xi \le x\}$

$0 \le F(x) \le 1$

$F(-\infty) = 0$

$F(+\infty) = 1$

$P\{a < \xi \le b\} = F(b) - F(a)$

$F(x,y) = P\{\xi \le x, \eta \le y\}$

$0 \le F(x,y) \le 1$

$F(-\infty, y) = F(x, -\infty) = 0$

$F(+\infty, +\infty) = 1$

$P\{a_1 < \xi \le b_1, a_2 < \eta \le b_2\} = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2)$

$F_\xi(x) = F(x, +\infty)$

$F_\eta(y) = F(+\infty, y)$

Плотность распределения

$F(x) = \int\limits_{-\infty}^{x} f(t)dt$

$P\{a < \xi \le b\} = \int\limits_{a}^{b} f(x)dx$

$F(x,y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f(s,t)dsdt = \int\limits_{-\infty}^{y} dt \int\limits_{-\infty}^{x} f(s,t)ds$

$P\{a_1 < \xi \le b_1, a_2 < \eta \le b_2\} = \int\limits_{a_1}^{b_1} dx \int\limits_{a_2}^{b_2} f(x,y)dy$

$f_\xi(x) = \int\limits_{-\infty}^{+\infty} f(x,y)dy$

$f_\eta(y) = \int\limits_{-\infty}^{+\infty} f(x,y)dx$

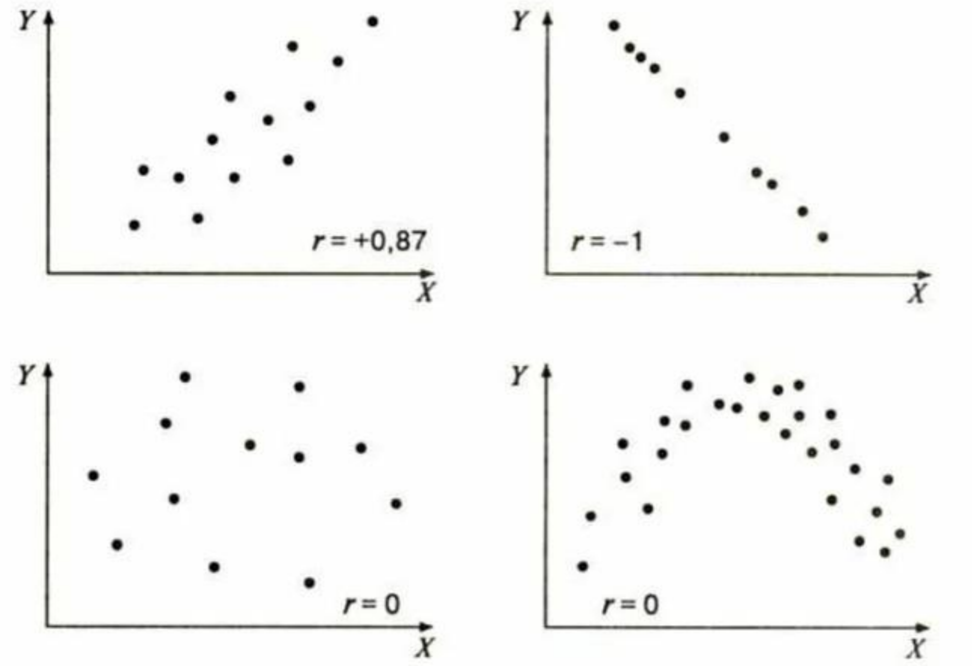
Ковариацией $cov(\xi, \eta)$ случайных величин ξ и η называется математическое ожидание произведения центрированных случайных величин.

$$cov(\xi, \eta) = E[(\xi - E\xi)(\eta - E\eta)] = E(\xi\eta) - E(\xi)E(\eta)$$

Ковариация нормированных случайных величин ξ и η называется **коэффициентом корреляции**

$$\rho(\xi, \eta) = \frac{cov(\xi, \eta)}{\sqrt{D_\xi}\sqrt{D_\eta}}$$

Случайные величины ξ и η называют коррелированными (корреляционно зависимыми), если $cov(\xi, \eta) \neq 0$, и некоррелированными, если $cov(\xi, \eta) = 0$.
Случайные величины ξ и η называют положительно коррелированными, если если $cov(\xi, \eta) > 0$, и отрицательно коррелированными, если $cov(\xi, \eta) < 0$.



1. $cov(\xi, \xi) = D\xi$
2. $D(\xi + \eta) = D\xi + D\eta + 2cov(\xi, \eta)$
3. $cov(a_1\xi + b_1, a_2\eta + b_2) = a_1 a_2 cov(\xi, \eta)$
4. $\rho(\xi, \xi) = 1$
5. $\rho(a_1\xi + b_1, a_2\eta + b_2) = sgn(a_1 a_2)\rho(\xi, \eta)$

ξ, η независимые $\Leftrightarrow F(x,y) = F_\xi(x)F_\eta(y) \quad \forall x, y \in \mathbb{R}$

ξ, η независимые $\Leftrightarrow P\{\xi = x_i, \eta = y_j\} = P\{\xi = x_i\}P\{\eta = y_j\} \quad \forall x_i, y_j$ - для дискретных СВ

ξ, η независимые $\Leftrightarrow f(x,y) = f_\xi(x)f_\eta(y) \quad \forall x, y \in \mathbb{R}$ - для непрерывных СВ

ξ, η независимые $\Rightarrow cov(\xi, \eta) = 0$

ξ, η независимые $\Rightarrow \rho(\xi, \eta) = 0$

$cov(\xi, \eta) = 0 \cancel{\Rightarrow} \xi, \eta$ независимые

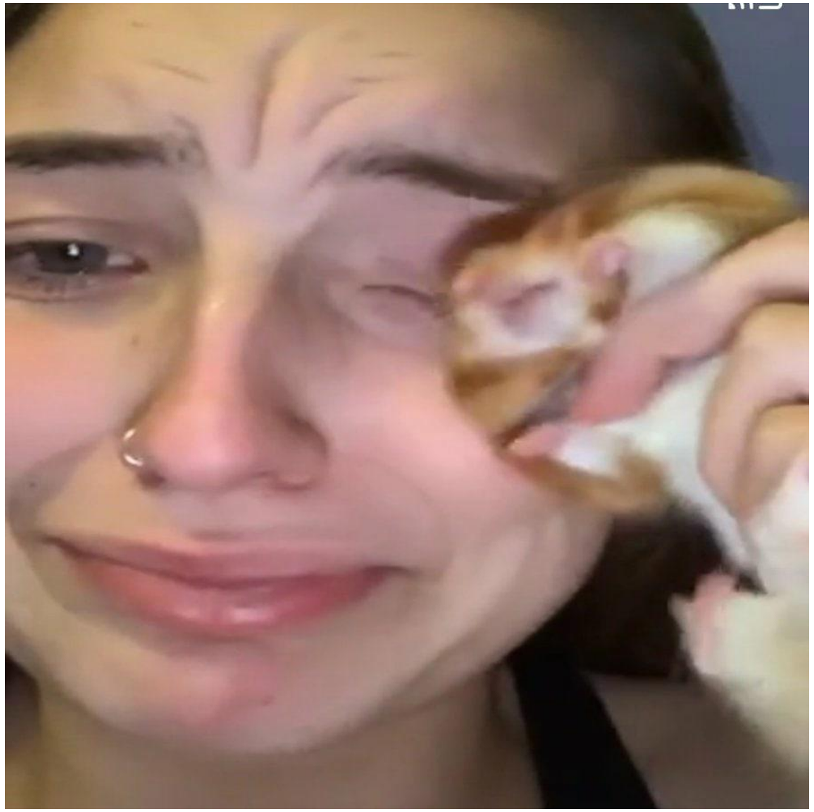
$\rho(\xi, \eta) = 0 \cancel{\Rightarrow} \xi, \eta$ независимые

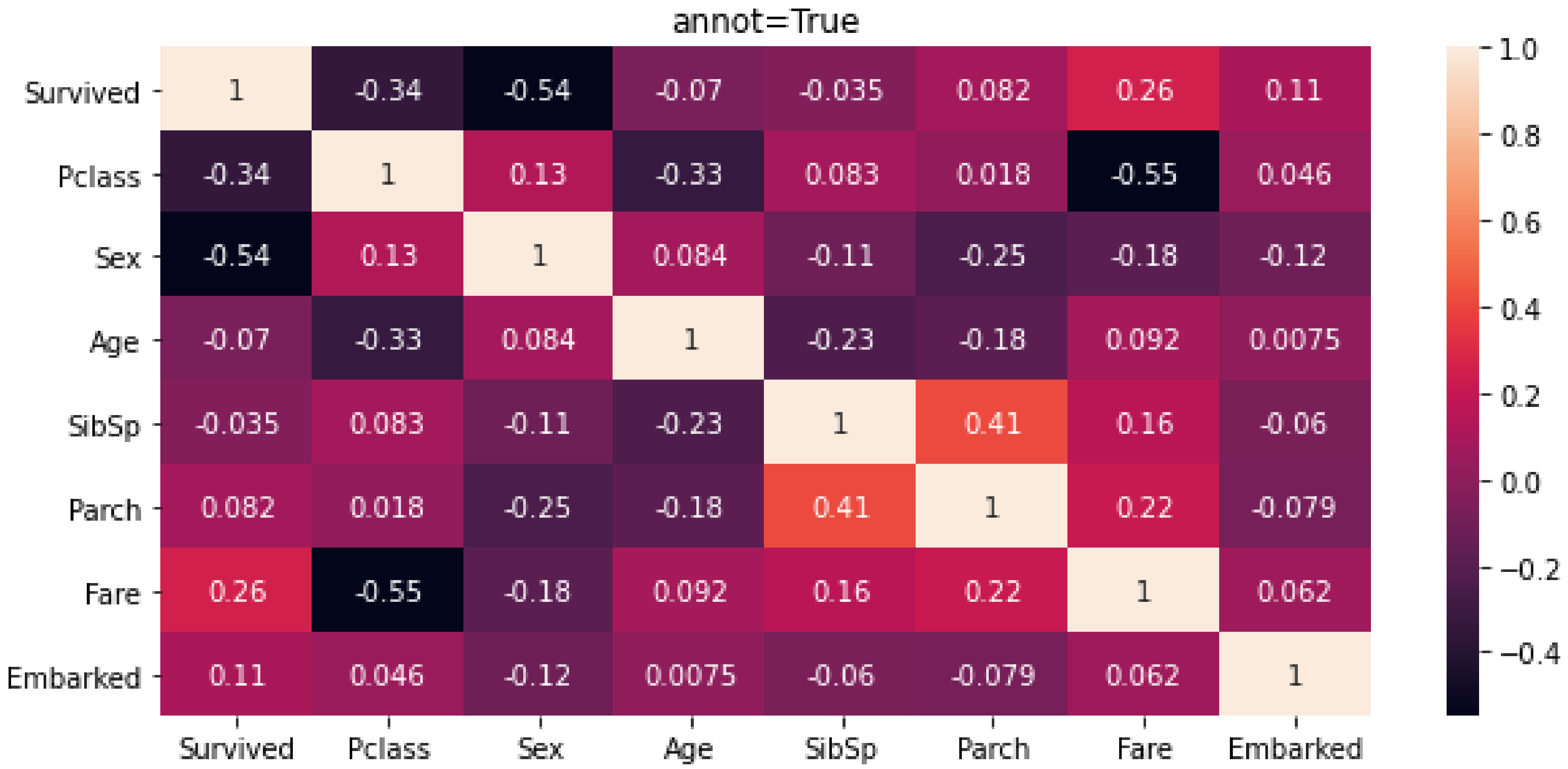
$cov(\xi, \eta) \neq 0 \Rightarrow \xi, \eta$ зависимые

$\rho(\xi, \eta) \neq 0 \Rightarrow \xi, \eta$ зависимые

$\rho(\xi, \eta) = \pm 1 \Rightarrow \xi, \eta$ линейно зависимые

кто-то: коэффициент корреляции 0, значит СВ независимы
я:




annot=True
Survived
Pclass
Sex
Age
SibSp
Parch
Fare
Embarked
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

стр.132, №5

Задан закон распределения случайного вектора $Z = (X, Y)$

Y/X	-1	1
1	1/6	1/3
2	0	1/6
3	1/3	0

Требуется:

а) найти закон распределения случайной величины $X + Y$

б) проверить справедливость равенств $E(X+Y) = EX + EY$, $D(X+Y) = DX + DY + 2cov(X, Y)$

Решение

$P\{X = x_i\} = \sum\limits_j P\{X = x_i, Y = y_j\} = \sum\limits_j P\{Z = z_{ij}\}$ (в данном случае i - столбец, j - строка)

$P\{Y = y_j\} = \sum\limits_i P\{X = x_i, Y = y_j\} = \sum\limits_i P\{Z = z_{ij}\}$ (в данном случае i - столбец, j - строка)

Найдём сумму по строке и по столбцу

Y/X	-1	1	
1	1/6	1/3	1/2
2	0	1/6	1/6
3	1/3	0	1/3
	1/2	1/2	

Тогда

X	-1	1
p	1/2	1/2

Y	1	2	3
p	1/2	1/6	1/3

Y/X	-1	1	
1	1/6	1/3	1/2
2	0	1/6	1/6
3	1/3	0	1/3
	1/2	1/2	

Тогда

X	-1	1
p	1/2	1/2

Y	1	2	3
p	1/2	1/6	1/3

$P(AB) \neq P(A)P(B)$

$P\{X+Y=0\} = P\{X=-1, Y=1\} = 1/6$
$P\{X+Y=1\} = P\{X=-1, Y=2\} = 0$
$P\{X+Y=2\} = P\{X=-1, Y=3\} + P\{X=1, Y=1\} = 1/3 + 1/3 = 2/3$
$P\{X+Y=3\} = P\{X=1, Y=2\} = 1/6$
$P\{X+Y=4\} = P\{X=1, Y=3\} = 0$

X+Y	0	1	2	3	4
p	1/6	0	2/3	1/6	0

Математическое ожидание дискретной случайной величины X

$EX = \sum x_i p_i = -1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0$

Математическое ожидание дискретной случайной величины Y

$EY = \sum y_i p_i = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{3} = \frac{11}{6}$

Математическое ожидание дискретной случайной величины $X+Y$

$E(X+Y) = 0 \cdot \frac{1}{6} + 1 \cdot 0 + 2 \cdot \frac{2}{3} + 3 \cdot \frac{1}{6} + 4 \cdot 0 = \frac{11}{6}$

$E(X+Y) = EX + EY$ - верно

Y/X	-1	1	
1	1/6	1/3	1/2
2	0	1/6	1/6
3	1/3	0	1/3
	1/2	1/2	

X	-1	1
p	1/2	1/2

Y	1	2	3
p	1/2	1/6	1/3

X+Y	0	1	2	3	4
p	1/6	0	2/3	1/6	0

Дисперсия случайной величины X

$$E(X^2) = \sum x_i^2 p_i = (-1)^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = 1$$

$$DX = E(X^2) - (EX)^2 = 1 - 0 = 1$$

Дисперсия случайной величины Y

$$E(Y^2) = \sum y_i^2 p_i = 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{3} = \frac{25}{6}$$

$$DY = E(Y^2) - (EY)^2 = \frac{25}{6} - \frac{121}{36} = \frac{29}{36}$$

Дисперсия случайной величины $X + Y$

$$E((X+Y)^2) = 0 \cdot \frac{1}{6} + 1 \cdot 0 + 2^2 \cdot \frac{2}{3} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot 0 = \frac{25}{6}$$

$$D(X+Y) = \frac{25}{6} - \frac{121}{36} = \frac{29}{36}$$

Ковариация случайных величин X и Y

$$E(XY) = \sum x_i y_j p_i j = -1 \cdot 1 \cdot \frac{1}{6} - 1 \cdot 3 \cdot \frac{1}{3} + 1 \cdot 1 \cdot \frac{1}{3} + 1 \cdot 2 \cdot \frac{1}{6} = -\frac{1}{2}$$

$$cov(X,Y) = E(XY) - E(X)E(Y) = -\frac{1}{2} - 0 \cdot \frac{11}{6} = -\frac{1}{2}$$

$$D(X+Y) = D(X) + D(Y) + 2cov(X,Y) = 1 + \frac{29}{36} + 2 \cdot (-1)\frac{1}{2} = \frac{29}{36}$$ - верно

стр.132, №7

В продукции завода брак вследствие дефекта A составляет 3%, а вследствие дефекта B - 4.5%. Годная продукция составляет 95%. Найти коэффициент корреляции дефектов A и B.

Решение

Пусть случайная величина A - наличие дефекта A, а случайная величина B - наличие дефекта B.

Тогда

A	0	1
p	0.97	0.03

B	0	1
p	0.955	0.045

A/B	0	1
0	0.95	?
1	?	?

Но

A/B	0	1	
0	0.95	?	0.97
1	?	?	0.03
	0.955	0.045	

Тогда

A/B	0	1
0	0.95	0.02
1	0.005	0.025

A	0	1
p	0.97	0.03

B	0	1
p	0.955	0.045

A/B	0	1
0	0.95	0.02
1	0.005	0.025

Математическое ожидание дискретной случайной величины A
$EA = \sum a_i p_i = 0 * 0.97 + 1 * 0.03 = 0.03$
Математическое ожидание дискретной случайной величины B
$EB = \sum b_i p_i = 0 * 0.955 + 1 * 0.045 = 0.045$

Дисперсия случайной величины A
$E(A^2) = 0 * 0.97 + 1 * 0.03 = 0.03$
$D(A) = E(A^2) - (EA)^2 = 0.03 - 0.03^2 = 0.0291$
Дисперсия случайной величины B
$E(B^2) = 0 * 0.955 + 1 * 0.045 = 0.045$
$D(B) = E(B^2) - (EB)^2 = 0.045 - 0.045^2 = 0.042975$

Ковариация A и B
$E(AB) = \sum a_i b_j p_{ij} = 0 * 0 * 0.95 + 0 * 1 * 0.02 + 1 * 0 * 0.05 + 1 * 1 * 0.025 = 0.025$
$cov(A, B) = E(AB) - E(A)E(B) = 0.025 - 0.03 * 0.045 = 0.02365$

Коэффициент корреляции A и B

$$\rho(A, B) = \frac{cov(A, B)}{\sqrt{DA}\sqrt{DB}} = \frac{0.02365}{\sqrt{0.0291}\sqrt{0.042975}} \approx 0.669$$

Ответ: 0.669

Центр исследований гражданского общества и некоммерческого сектора НИУ ВШЭ провёл социологическое исследование, в котором изучалась проблема участия населения в благотворительной деятельности. Среди вопросов, задаваемых респондентам, были, в частности, вопросы о материальном положении и образовании респондента.

Пусть СВ ξ (материальное положение) принимает три значения - 1 (бедный), 2 (средний уровень благосостояния), 3 (высокий уровень благосостояния), а СВ η (уровень образования) принимает значения - 1 (образование ниже среднего), 2 (среднее и среднее специальное образование), 3 (высшее образование).

Распределение случайного вектора (η, ξ), соответствующее репрезентативной выборке 2009 года, представлено следующей таблицей.

η/ξ	1	2	3
1	0.083	0.035	0.001
2	0.31	0.375	0.026
3	0.04	0.116	0.014

Найдите коэффициент корреляции СВ ξ и η. Являются ли СВ ξ и η некоррелированными? Являются ли СВ ξ и η зависимыми? Прокомментируйте полученный результат.

Решение

ξ	1	2	3
p	0.433	0.526	0.041

η	1	2	3
p	0.119	0.711	0.17

η/ξ	1	2	3
1	0.083	0.035	0.001
2	0.31	0.375	0.026
3	0.04	0.116	0.014

ξ	1	2	3
p	0.433	0.526	0.041

η	1	2	3
p	0.119	0.711	0.17

$$E\xi = 1 * 0.433 + 2 * 0.526 + 3 * 0.041 = 1.608$$
$$E\eta = 1 * 0.119 + 2 * 0.711 + 3 * 0.17 = 2.051$$

$$E(\xi^2) = 1 * 0.433 + 4 * 0.526 + 9 * 0.041 = 2.906$$
$$D\xi = E(\xi^2) - (E\xi)^2 = 2.906 - 1.608^2 = 0.320336$$
$$E(\eta^2) = 1 * 0.119 + 4 * 0.711 + 9 * 0.17 = 4.493$$
$$D\eta = E(\eta^2) - (E\eta)^2 = 4.493 - 2.051^2 = 0.286399$$

$$E(\xi\eta) = 1*1*0.083+1*2*0.035+1*3*0.001+2*1*0.31+2*2*0.375+2*3*0.026+3*1*0.04+3*2*0.116+3*3*0.014 = 3.374$$
$$cov(\xi,\eta) = E(\xi\eta) - E\xi E\eta = 0.075992$$

$$\rho(\xi,\eta) = \frac{cov(\xi,\eta)}{\sqrt{D\xi}\sqrt{D\eta}} = \frac{0.075992}{\sqrt{0.320336 * 0.286399}} \approx 0.250887631$$

Центр исследований гражданского общества и некоммерческого сектора НИУ ВШЭ провёл социологическое исследование, в котором изучалась проблема участия населения в благотворительной деятельности. Среди вопросов, задаваемых респондентам, были, в частности, вопросы о материальном положении и образовании респондента.
Пусть СВ ξ (материальное положение) принимает три значения - 1 (бедный), 2 (средний уровень благосостояния), 3 (высокий уровень благосостояния), а СВ η (уровень образования) принимает значения - 1 (образование ниже среднего), 2 (среднее и среднее специальное образование), 3 (высшее образование).
Распределение случайного вектора (η, ξ), соответствующее репрезентативной выборке 2009 года, представлено следующей таблицей.

Найдите коэффициент корреляции СВ ξ и η. Являются ли СВ ξ и η некоррелированными? Являются ли СВ ξ и η зависимыми? Прокомментируйте полученный результат.

$$\rho(\xi, \eta) = \frac{cov(\xi, \eta)}{\sqrt{D\xi}\sqrt{D\eta}} = \frac{0.075992}{\sqrt{0.320336 * 0.286399}} \approx 0.250887631$$

ξ, η независимые $\Leftrightarrow F(x, y) = F_\xi(x) F_\eta(y) \quad \forall x, y \in \mathbb{R}$

ξ, η независимые $\Leftrightarrow P\{\xi = x_i, \eta = y_j\} = P\{\xi = x_i\} P\{\eta = y_j\} \quad \forall x_i, y_j$ - для дискретных СВ

ξ, η независимые $\Leftrightarrow f(x, y) = f_\xi(x) f_\eta(y) \quad \forall x, y \in \mathbb{R}$ - для непрерывных СВ

ξ, η независимые $\Rightarrow cov(\xi, \eta) = 0$

ξ, η независимые $\Rightarrow \rho(\xi, \eta) = 0$

$cov(\xi, \eta) = 0 \cancel{\Rightarrow} \xi, \eta$ независимые

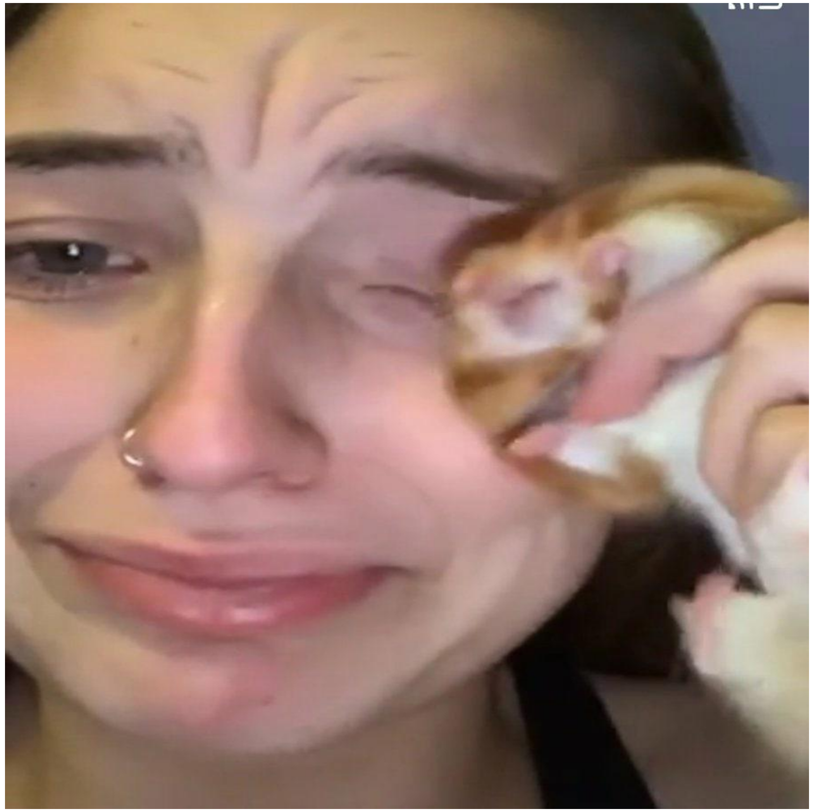
$\rho(\xi, \eta) = 0 \cancel{\Rightarrow} \xi, \eta$ независимые

$cov(\xi, \eta) \neq 0 \Rightarrow \xi, \eta$ зависимые

$\rho(\xi, \eta) \neq 0 \Rightarrow \xi, \eta$ зависимые

$\rho(\xi, \eta) = \pm 1 \Rightarrow \xi, \eta$ линейно зависимые

кто-то: коэффициент корреляции 0, значит СВ независимы
я:



Центр исследований гражданского общества и некоммерческого сектора НИУ ВШЭ провёл социологическое исследование, в котором изучалась проблема участия населения в благотворительной деятельности. Среди вопросов, задаваемых респондентам, были, в частности, вопросы о материальном положении и образовании респондента.
Пусть СВ ξ (материальное положение) принимает три значения - 1 (бедный), 2 (средний уровень благосостояния), 3 (высокий уровень благосостояния), а СВ η (уровень образования) принимает значения - 1 (образование ниже среднего), 2 (среднее и среднее специальное образование), 3 (высшее образование).
Распределение случайного вектора (η, ξ), соответствующее репрезентативной выборке 2009 года, представлено следующей таблицей.

Найдите коэффициент корреляции СВ ξ и η. Являются ли СВ ξ и η некоррелированными? Являются ли СВ ξ и η зависимыми? Прокомментируйте полученный результат.

$$\rho(\xi,\eta) = \frac{cov(\xi,\eta)}{\sqrt{D\xi}\sqrt{D\eta}} = \frac{0.075992}{\sqrt{0.320336 * 0.286399}} \approx 0.250887631$$

$\rho(\xi,\eta) \neq 0 \Rightarrow \xi$ и η коррелированы
$\rho(\xi,\eta) > 0 \Rightarrow \xi$ и η коррелированы положительно

$\rho(\xi,\eta) \neq 0 \Rightarrow \xi$ и η зависимы

Вывод:
Чем продвинутее уровень образования, тем успешнее материальное положение!

Вывод:
Чем лучше материальное положение, тем более продвинутый уровень образования можно себе позволить!

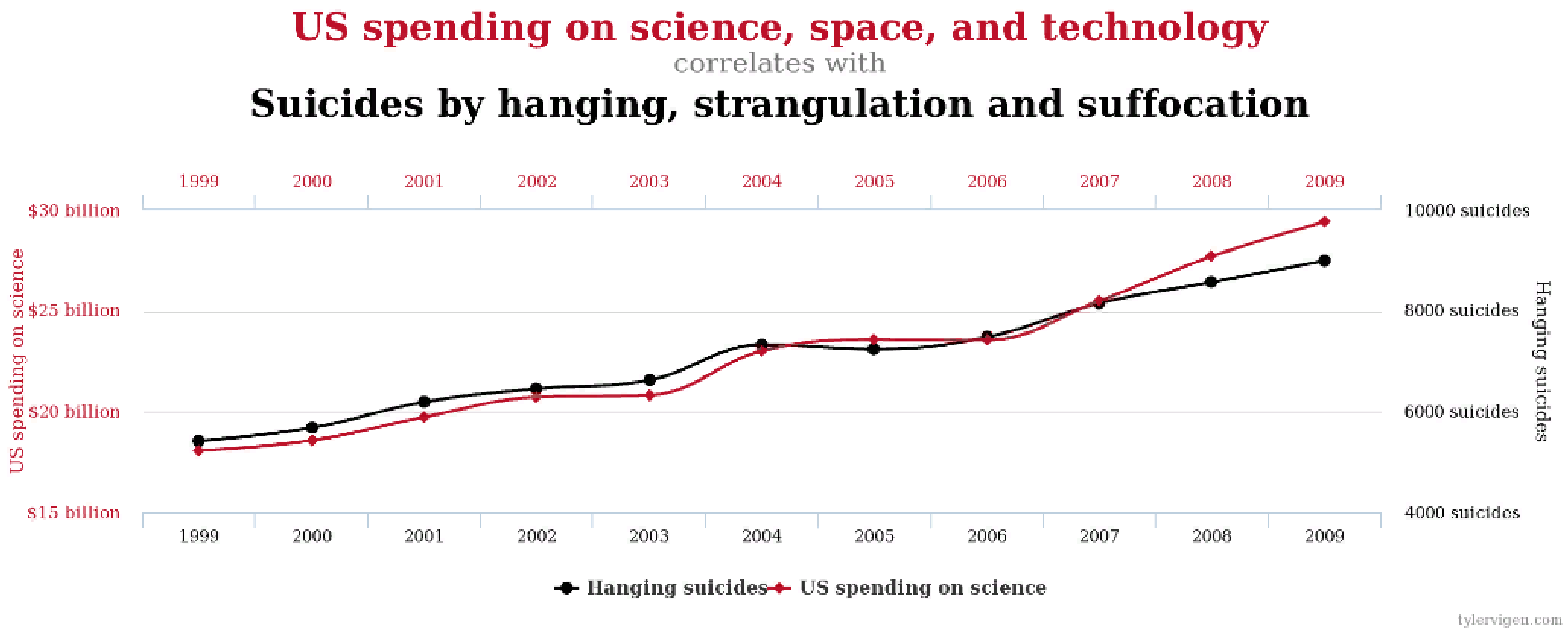
Вывод:
Существует взаимосвязь между материальным положением и уровнем образования среди выбранных респондентов, данные величины положительно кореллированы, коэффициент корреляции $\rho = 0.25$

Площадь пожара

Количество ущерба и людских потерь

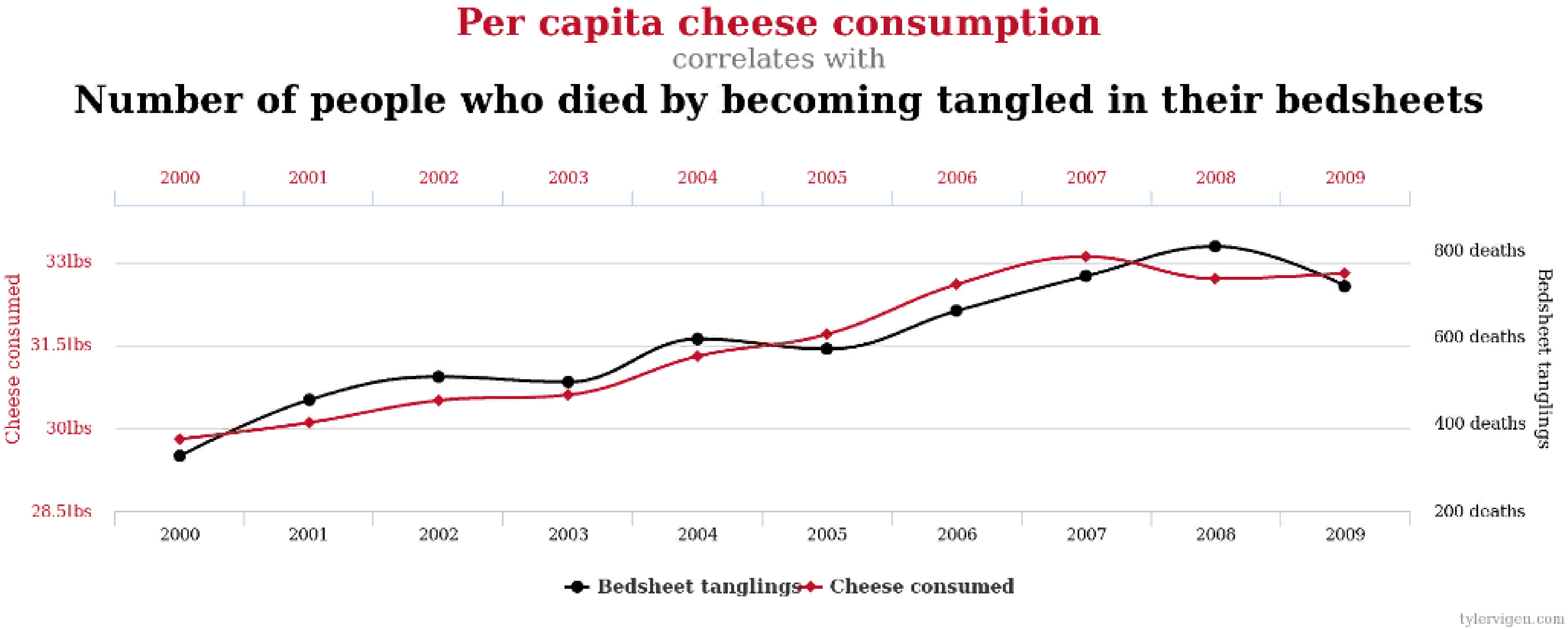
?

Количество задействованных пожарных бригад



Затраты США на науку, космос и технологии / Суициды путем повешения и удушения. Корреляция 99,79%.

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



Потребление сыра / Число до смерти запутавшихся в простынях. Корреляция 94,71%

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

Иллюзорная корреляция

Материал из Википедии — свободной энциклопедии

[править | править код]

Иллюзорная корреляция (англ. *illusory correlation*) — когнитивное искажение преувеличенно тесной связи между переменными, которая в реальности или не существует, или значительно меньше, чем предполагается. Типичным примером могут служить приписывание группе этнического меньшинства отрицательных качеств. Иллюзорная корреляция считается одним из способов формирования стереотипов.

Феномен иллюзорной корреляции чаще всего наблюдается, когда события необычные, заметные и бросаются в глаза.

Содержание [скрыть]

Классическое исследование [править | править код]

Иллюзорная корреляция впервые[*источник не указан 1988 дней*] была обнаружена в ходе экспериментов Чэпмен и Чэпмен (Chapman and Chapman) (1967). Понятием «иллюзорная корреляция» они обозначили тенденцию людей переоценивать связь между двумя группами, когда представлена отличительная и необычная информация.

В ходе их эксперимента испытуемым, не имеющим медицинской подготовки, предлагалась информация о гипотетических душевнобольных. Затем им предлагалось оценить частоту, с который каждый диагноз (например, паранойя) сопровождался особенностями рисунка (например, большие глаза). В ходе эксперимента выяснилось, что испытуемые преувеличивали частоту совпадения естественных ассоциаций (большие глаза — паранойя). Данные эксперимента подвергли сомнению состоятельность проективных методик как психодиагностического инструмента[1][2].

https://ru.wikipedia.org/wiki/Иллюзорная_корреляция