Lecture 6: Joint distributions

MVE055 / MSG810
Mathematical statistics and discrete mathematics

Moritz Schauer

Last updated September 13, 2021, 2021

GU & Chalmers University of Technology

Bivariate distributions

Definition

Informal: A two-dimensional or bivariate random variable (X, Y) produces a pair of random numbers.

Bivariate distributions

Definition

Informal: A two-dimensional or bivariate random variable (X, Y) produces a pair of random numbers.

For discrete random variables we have the joint density (probability mass function)

$$f_{X,Y}(i,j) = \mathsf{P}(X = i, Y = j) = \mathsf{P}(X = i \text{ and } Y = j).$$

Here $f_{X,Y}(i,j) \geqslant 0$ and $\sum_{\text{all } i,j} f_{X,Y}(i,j) = 1$.

Example

Let X and Y be the number of girls, respectively boys in a randomly chosen Swedish family. The joint density function $f_{X,Y}(x, y)$ is given in the table below.

Y / X	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

Example

Let X and Y be the number of girls, respectively boys in a randomly chosen Swedish family. The joint density function $f_{X,Y}(x,y)$ is given in the table below.

X \ Y	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$$\sum_{\text{all } x,y} f_{X,Y}(x,y) = 1$$

Example

Let X and Y be the number of girls, respectively boys in a randomly chosen Swedish family. The joint density function $f_{X,Y}(x,y)$ is given in the table below.

X \ Y	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$$\sum_{\text{all } x,y} f_{X,Y}(x,y) = 1$$

$$\mathrm{P}(X = 0 \text{ and } Y = 1) = f_{X,Y}(0,1) = 0.16$$

Example

Let X and Y be the number of girls, respectively boys in a randomly chosen Swedish family. The joint density function $f_{X,Y}(x,y)$ is given in the table below.

X \ Y	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$$\sum_{\text{all } x,y} f_{X,Y}(x,y) = 1$$

$\mathrm{P}(X = 0 \text{ and } Y = 1) = f_{X,Y}(0,1) = 0.16$

$\mathrm{P}(X = 2) = f_{XY}(2,0) + f_{X,Y}(2,1) + f_{XY}(2,2) = 0.08$

Expected values ♡

$$\mathsf{E}(h(X,Y)) = \sum_{\text{all } i,j} h(i,j) f_{X,Y}(i,j).$$

For example:

$$\mathsf{E}(X+Y]) = \sum_{\text{all } i,j} (i+j) f_{X,Y}(i,j)$$

with $h(i,j) = i + j$.

Expected number of children

X and Y be the number of girls, respectively boys in a randomly chosen Swedish family.

$\boxed{\qquad\qquad}$ is the expected number of girls + boys = children.
So $h(i, j) =$.

X \ Y	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$E(X + Y) =$

Expected number of children

X and Y be the number of girls, respectively boys in a randomly chosen Swedish family.

$E(X+Y)$ is the expected number of girls + boys = children. So $h(i,j) =$.

Y / X	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$E(X+Y) =$

Expected number of children

X and Y be the number of girls, respectively boys in a randomly chosen Swedish family.

$E(X+Y)$ is the expected number of girls + boys = children. So $h(i,j) = \boxed{}$.

Y / X	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$E(X+Y) =$

Expected number of children

X and Y be the number of girls, respectively boys in a randomly chosen Swedish family.

$E(X+Y)$ is the expected number of girls + boys = children. So $h(i,j) = i + j$.

X \ Y	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$E(X+Y) =$

Expected number of children

X and Y be the number of girls, respectively boys in a randomly chosen Swedish family.

$E(X+Y)$ is the expected number of girls + boys = children. So $h(i,j) = i + j$.

Y X	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$E(X+Y) =$

Expected number of children

X and Y be the number of girls, respectively boys in a randomly chosen Swedish family.

$E(X+Y)$ is the expected number of girls + boys = children. So $h(i,j) = i + j$.

Y / X	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$E(X+Y) =$

Expected number of children

X and Y be the number of girls, respectively boys in a randomly chosen Swedish family.

$E(X+Y)$ is the expected number of girls + boys = children. So $h(i,j) = i + j$.

X \ Y	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$$E(X+Y) = \boxed{(0+0)\cdot 0.38 + (1+0)\cdot 0.17 + ... = 1.08}$$

Marginal distributions

Given a pair of discrete random variables (X, Y) with joint density $f_{X,Y}$ density for X and Y are given by

$$f_X(i) = \sum_{\text{all } j} f_{X,Y}(i, j)$$
$$f_Y(j) = \sum_{\text{all } i} f_{X,Y}(i, j).$$

and called **marginal densities** (marginal p.m.f.'s.)

Y / X	0	1	2	3	4	f_X
0	0.38	0.16	0.04	0.01	0.01	0.60
1	0.17	0.08	0.02			0.27
2	0.05	0.02	0.01			0.08
3	0.02	0.01				0.03
4	0.02					0.02
f_Y	0.64	0.27	0.07	0.01	0.01	1

Continuous bivariate random variables

For a pair of continuous random variables: a function $f_{X,Y}(x,y)$ with properties

1. $f_{X,Y}(x,y) \geqslant 0$,
2. $\int\int f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y = 1$, and
3. $\mathsf{P}(a \leqslant X \leqslant b \text{ and } c \leqslant Y \leqslant d) = \int_a^b \int_c^d f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$.

Marginal distributions

For a bivariate continuous random variable (X, Y), the probability density functions for X and Y are given by

$$f_X(x) = \int f_{X,Y}(x, y)\mathrm{d}y$$

$$f_Y(y) = \int f_{X,Y}(x, y)\mathrm{d}x$$

Expected value

For a two-dimensional random variable (X, Y), the expected values of X and Y are given by

$$\mathsf{E}(X) = \begin{cases} \sum\limits_{\text{all } i,\, j} i f_{X,Y}(i,j), & \text{for } X \text{ discrete,} \\ \int \int x f_{X,Y}(x,y) \mathrm{d}x \mathrm{d}y, & \text{for } X \text{ continuos,} \end{cases}$$

Expected value

For a two-dimensional random variable (X, Y), the expected values of X and Y are given by

$$\mathsf{E}(X) = \begin{cases} \sum_{\text{all } i,\ j} i f_{X,Y}(i,j), & \text{for } X \text{ discrete,} \\ \int\int x f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y, & \text{for } X \text{ continuos,} \end{cases}$$

and

$$\mathsf{E}(Y) = \begin{cases} \sum_{\text{all } i,\ j} j f_{X,Y}(i,j), & \text{for } Y \text{ discrete,} \\ \int\int y f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y, & \text{for } Y \text{ continuous.} \end{cases}$$

Expected value

For a two-dimensional random variable (X, Y), the expected values of X and Y are given by

$$\mathsf{E}(X) = \begin{cases} \sum\limits_{\text{all } i,\ j} i f_{X,Y}(i,j), & \text{for } X \text{ discrete,} \\ \int \int x f_{X,Y}(x,y) \mathrm{d}x \mathrm{d}y, & \text{for } X \text{ continuos,} \end{cases}$$

Conditional distribution

The *conditional distribution of* X *given* $Y = y$ is defined by its density

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)},$$

provided that $f_Y(y) > 0$.

Conditional distribution

The *conditional distribution of X given $Y = y$* is defined by its density

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)},$$

provided that $f_Y(y) > 0$.

Independent random variables

Two random variables X and Y are called independent if their bivariate density can be written as product of the marginal densities:

$$f_{X,Y}(u,v) = f_X(u) f_Y(v).$$

There is no "samvariation", knowing X does not explain Y, etc.

Covariance

Covariance

Covariance between random variables X and Y is defined as $\mathrm{Cov}(X, Y) = \mathsf{E}[(X - \mu_X)(Y - \mu_Y)]$, where $\mu_X = \mathsf{E}(X)$ and $\mu_Y = \mathsf{E}(Y)$.

Covariance

Covariance

Covariance between random variables X and Y is defined as $\mathrm{Cov}(X, Y) = \mathsf{E}[(X - \mu_X)(Y - \mu_Y)]$, where $\mu_X = \mathsf{E}(X)$ and $\mu_Y = \mathsf{E}(Y)$.

- According to the definition,

$$\mathrm{Cov}(X, Y) = \begin{cases} \sum_{\text{all } i,\ j} (i - \mu_X)(j - \mu_Y) f_{X,Y}(i, j), & \text{discrete} \\ \int\int (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y) \mathrm{d}x \mathrm{d}y, & \text{cont.} \end{cases}$$

- Note that $\mathrm{Cov}(X, X) = \mathsf{V}(X)$.
- **If** X and Y are independent, **then** $\mathrm{Cov}(X, Y) = 0$ and $\mathsf{E}(XY) = \mathsf{E}(X)\mathsf{E}(Y)$.
- Unit??

Rules for covariance

$\mathrm{Cov}(X, Y)$ can be calculated as
$\mathrm{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$.

Rules for covariance

$\mathrm{Cov}(X, Y)$ can be calculated as
$\mathrm{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$.

For two random variables X and Y, and two numbers a and b we have

$$\mathsf{V}(aX + bY) = a^2\mathsf{V}(X) + b^2\mathsf{V}(Y) + 2ab\ \mathrm{Cov}(X, Y).$$

Rules for covariance

$\mathrm{Cov}(X, Y)$ can be calculated as
$\mathrm{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$.

For two random variables X and Y, and two numbers a and b we have

$$\mathsf{V}(aX + bY) = a^2\mathsf{V}(X) + b^2\mathsf{V}(Y) + 2ab\ \mathrm{Cov}(X, Y).$$

Examples:

$$\mathsf{V}(2X) = \mathsf{V}(X + X) = \mathsf{V}(X) + \mathsf{V}(X) + 2\ \mathrm{Cov}(X, X) = 4\mathsf{V}(X)$$

$$\mathsf{V}(X + Y) = \mathsf{V}(X) + \mathsf{V}(Y) \text{ when } X \text{ and } Y \text{ are independent}$$

Rules for covariance

$\mathrm{Cov}(X, Y)$ can be calculated as
$\mathrm{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$.

For two random variables X and Y, and two numbers a and b we have

$$\mathsf{V}(aX + bY) = a^2\mathsf{V}(X) + b^2\mathsf{V}(Y) + 2ab\ \mathrm{Cov}(X, Y).$$

Examples:

$$\mathsf{V}(2X) = \mathsf{V}(X + X) = \mathsf{V}(X) + \mathsf{V}(X) + 2\ \mathrm{Cov}(X, X) = 4\mathsf{V}(X)$$

$$\mathsf{V}(X + Y) = \mathsf{V}(X) + \mathsf{V}(Y) \text{ when } X \text{ and } Y \text{ are independent}$$

(“Fun” thing to do: look up the law of cosines.)

Correlation and independence

Correlation

The correlation coefficient is defines as

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathsf{V}(X)\mathsf{V}(Y)}}.$$

Correlation and independence

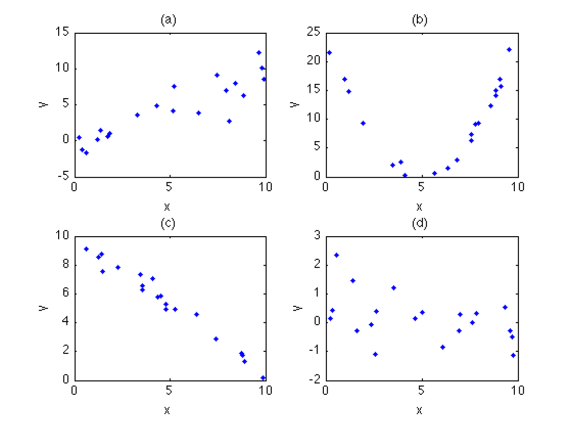
Correlation

The correlation coefficient is defines as

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathsf{V}(X)\mathsf{V}(Y)}}.$$

- A measure of linear relationship (linjär samvariation) of X and Y.
- It holds $-1 \leqslant \rho \leqslant 1$.
- X and Y are called uncorrelated if $\rho(X,Y) = 0$ (there is no "linjär samvariation") .
- Unit??

Visualisation



Assume 2d measurements (x_i, y_i). A scatter plot is a two-dimensional plot in which each (x_i, y_i) measurement is represented as a point in the x-y-plane.

Descriptive statistic for bivariate data

The sample covariance is defined as,

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

Descriptive statistic for bivariate data

The sample covariance is defined as,

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

and sample correlation coefficient is defined as

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

Descriptive statistic for bivariate data

The sample covariance is defined as,

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

and sample correlation coefficient is defined as

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

Descriptive statistic for bivariate data

The sample covariance is defined as,

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

and sample correlation coefficient is defined as

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

Descriptive statistic for bivariate data

The sample covariance is defined as,

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

and sample correlation coefficient is defined as

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

The sample correlation is a measure of linear dependence.

Descriptive statistic for bivariate data

The sample covariance is defined as,

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

and sample correlation coefficient is defined as

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

The sample correlation is a measure of linear dependence.

In the picture $r_{xy} = 0.8067$ i (a), $r_{xy} = 0.2912$ i (b), $r_{xy} = -0.9884$ i (c), och $r_{xy} = 0.3640$ i (d).

We have the following relationship between dependence and correlation:

- If X and Y are independent, then they are also uncorrelated.
- (Thus if X and Y are uncorrelated, they do not need to be independent.)

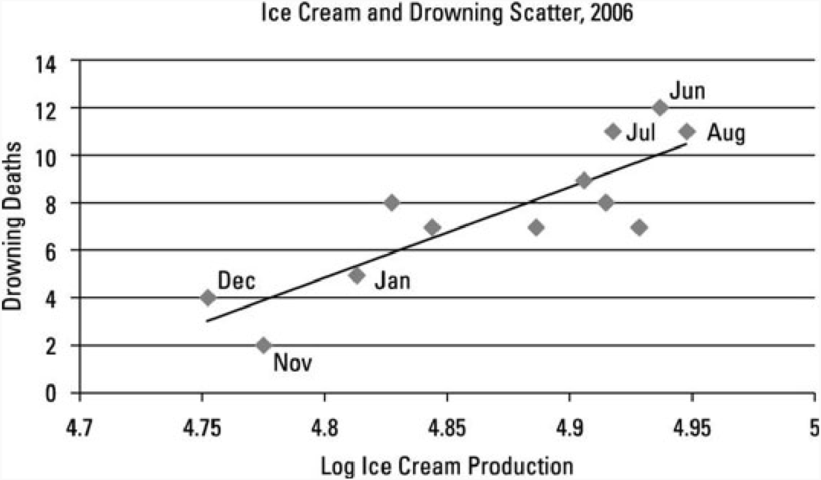
This is natural because two random variables are independent if there is no "samvariation" at all, while they are not correlated if there is no "*linjär* samvariation".

Correlation, dependence and causality

- Correlation does not say anything about causality!*
- Sometimes correlation is present but can be explained by a third variable which was not measured.
- Month with high ice cream sales tend to have more drowning accidents. Time to ban ice cream?

Correlation, dependence and causality

- Correlation does not say anything about causality!*
- Sometimes correlation is present but can be explained by a third variable which was not measured.
- Month with high ice cream sales tend to have more drowning accidents. Time to ban ice cream? In this example, an important variable which perhaps was not measured is the sunshine. Such variables are sometimes called confounding variables.



`https://twitter.com/dannagal/status/1244082688899919872`, September 13, 2021

Causality

- Correlation can also be introduced by selection effects.

Causality

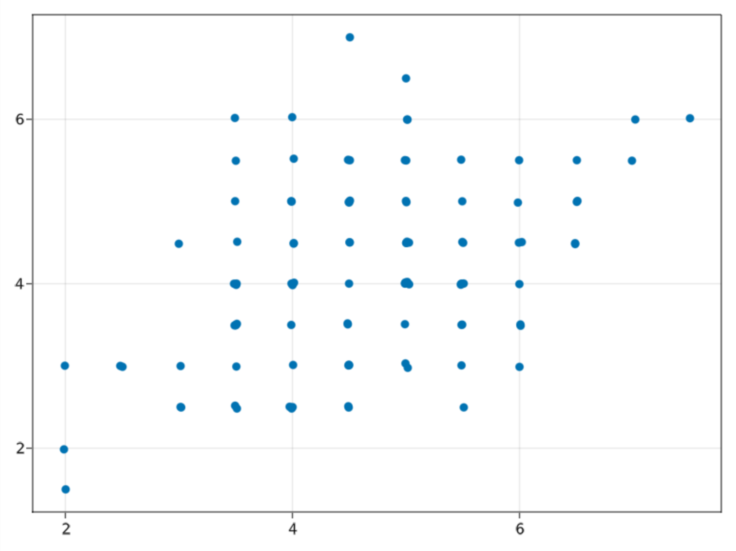
- Correlation can also be introduced by selection effects.
- Exam with two questions, one difficult, one easy. A student achieves X out of 10 points on the easy question, Y out of 10 points on the difficult question (random).

Causality

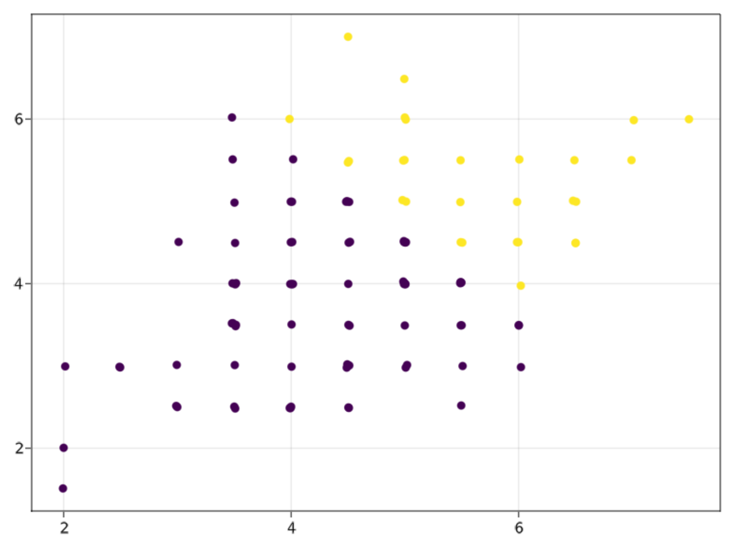
- Correlation can also be introduced by selection effects.
- Exam with two questions, one difficult, one easy. A student achieves X out of 10 points on the easy question, Y out of 10 points on the difficult question (random).
- Say X and Y slightly positively correlated. *But* only students with $X + Y \geqslant 10$ pass. Say I tell you the student has passed.

- Correlation can also be introduced by selection effects.
- Exam with two questions, one difficult, one easy. A student achieves X out of 10 points on the easy question, Y out of 10 points on the difficult question (random).
- Say X and Y slightly positively correlated. *But* only students with $X + Y \geqslant 10$ pass. Say I tell you the student has passed.
- Passing students performance on easy questions may now be negatively correlated with performance on the difficult question.

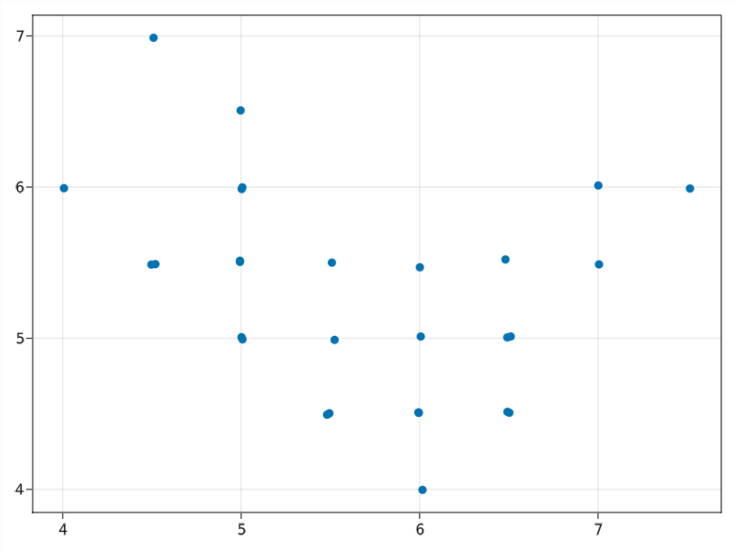
Exam points



Exam points



Exam points



Causality

- If we want to know/predict what will change if we perform an action we need insight into causality.

Causality

- If we want to know/predict what will change if we perform an action we need insight into causality.
- Will the number of drowning accidents change if we ban ice?

Causality

- If we want to know/predict what will change if we perform an action we need insight into causality.
- Will the number of drowning accidents change if we ban ice?
- There are many causal statements in the news!

Causality

- If we want to know/predict what will change if we perform an action we need insight into causality.
- Will the number of drowning accidents change if we ban ice?
- There are many causal statements in the news!
  - "Do not skip breakfast if you want to reduce the risk of coronary heart disease"

Causality

- If we want to know/predict what will change if we perform an action we need insight into causality.
- Will the number of drowning accidents change if we ban ice?
- There are many causal statements in the news!
  - "Do not skip breakfast if you want to reduce the risk of coronary heart disease"
- Be careful...

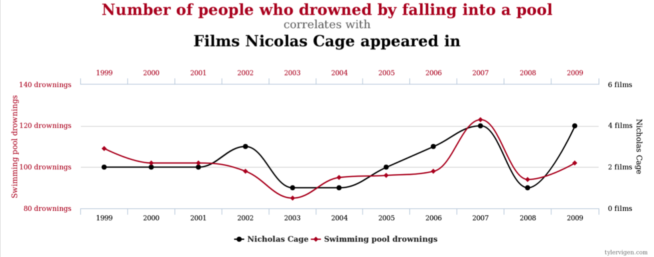
Causality

- If we want to know/predict what will change if we perform an action we need insight into causality.
- Will the number of drowning accidents change if we ban ice?
- There are many causal statements in the news!
    - "Do not skip breakfast if you want to reduce the risk of coronary heart disease"
- Be careful...
    - Candidate for a confounding variable: stress.

Causality

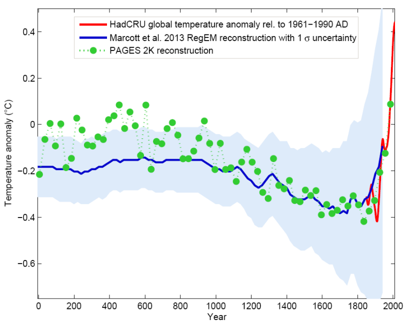
- If we want to know/predict what will change if we perform an action we need insight into causality.
- Will the number of drowning accidents change if we ban ice?
- There are many causal statements in the news!
  - "Do not skip breakfast if you want to reduce the risk of coronary heart disease"
- Be careful...
  - Candidate for a confounding variable: stress.
- We need to understand the science to answer causal questions! We will come back to this later.

Cherry picking



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

Thinking statistics: Global warming



Two millennia of mean surface temperatures according to different reconstructions from climate proxies with the instrumental temperature record overlaid in red.

Stefan Rahmstorf: Paleoclimate: The End of the Holocene.

`http://www.realclimate.org/index.php/archives/2013/09/paleoclimate-the-end-of-the-holocene/`.

Web. 3 Feb. 2019.