CS 466
Introduction to Bioinformatics

Instructor: Jian Peng

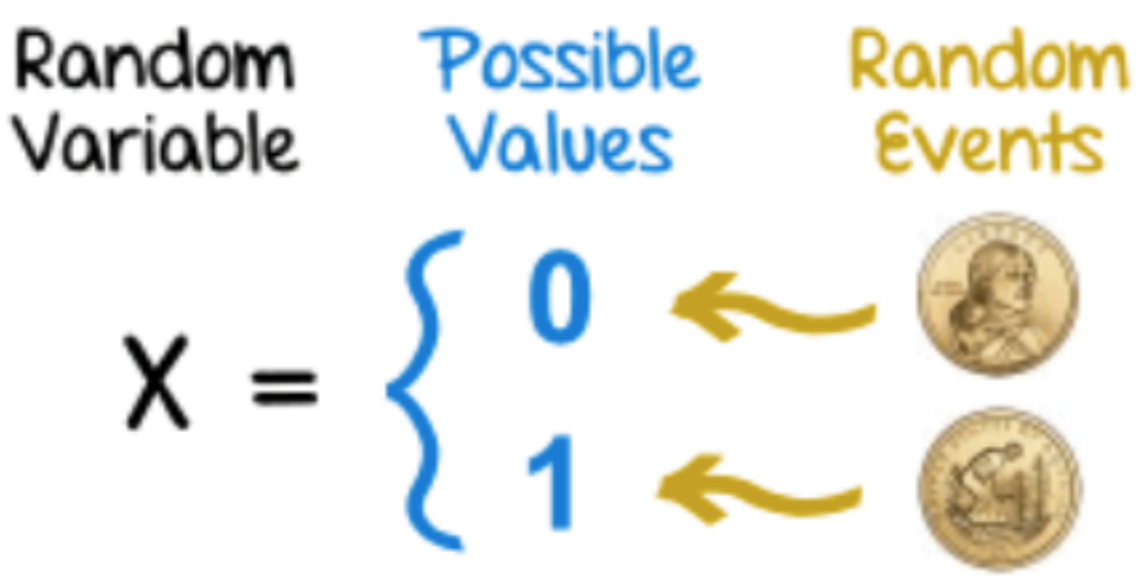
Probability and Statistics

Random Variables and Expectations

Random Variable

Quite commonly, we would like to deal with numbers that are random. We can do so by linking numbers to the outcome of an experiment. We define a **random variable**:

> **Definition: 4.1** *Discrete random variable*
>
> Given a sample space Ω, a set of events $\mathcal{F}$, a probability function P, and a countable set of of real numbers D, a discrete random variable is a function with domain Ω and range D.



Random Variable

Quite commonly, we would like to deal with numbers that are random. We can do so by linking numbers to the outcome of an experiment. We define a **random variable**:

Definition: 4.1 *Discrete random variable*

Given a sample space Ω, a set of events $\mathcal{F}$, a probability function P, and a countable set of of real numbers D, a discrete random variable is a function with domain Ω and range D.

Example: 4.1 *Numbers from coins*

We flip a coin. Whenever the coin comes up heads, we report 1; when it comes up tails, we report 0. This is a random variable.

Example: 4.2 *Numbers from coins II*

We flip a coin 32 times. We record a 1 when it comes up heads, and when it comes up tails, we record a 0. This produces a 32 bit random number, which is a random variable.

Probability distribution

Definition: 4.2 *Probability distribution of a discrete random variable*

The probability distribution of a discrete random variable is the set of numbers $P(\{X = x\})$ for each value x that X can take. The distribution takes the value 0 at all other numbers. Notice that the distribution is non-negative. **Notation warning:** probability notation can be quirky. You may encounter $p(x)$ with the meaning "some probability distribution" or $p(x)$ meaning "the value of the probability distribution $P(\{X = x\})$ at the point x" or $p(x)$ with the meaning "the probability distribution $P(\{X = x\})$". Context may help disambiguate these uses.

Worked example 4.1 *Numbers from coins III*

We flip a biased coin 2 times. The flips are independent. The coin has $P(H) = p$, $P(T) = 1 - p$. We record a 1 when it comes up heads, and when it comes up tails, we record a 0. This produces a 2 bit random number, which is a random variable taking the values 0, 1, 2, 3. What is the probability distribution and cumulative distribution of this random variable?

Solution: Probability distribution: $P(0) = (1 - p)^2$; $P(1) = (1 - p)p$; $P(2) = p(1 - p)$; $P(3) = p^2$. Cumulative distribution: $f(0) = (1 - p)^2$; $f(1) = (1 - p)$; $f(2) = p(1 - p) + (1 - p) = (1 - p^2)$; $f(3) = 1$.

Joint distribution

Definition: 4.4 *Joint probability distribution of two discrete random variables*

Assume we have two random variables X and Y. The probability that X takes the value x and Y takes the value y could be written as $P(\{X = x\} \cap \{Y = y\})$. It is more usual to write it as

$$P(x, y).$$

This is referred to as the **joint probability distribution** of the two random variables (or, quite commonly, the **joint**). You can think of this as a table of probabilities, one for each possible pair of x and y values.

Marginal distribution

Definition: 4.6 *The marginal probability of a random variable*

Write $P(x, y)$ for the joint probability distribution of two random variables X and Y. Then

$$P(x) = \sum_y P(x, y) = \sum_y P(\{X = x\} \cap \{Y = y\}) = P(\{X = x\})$$

is referred to as the **marginal probability distribution** of X.

Independent variables

Definition: 4.7 *Independent random variables*

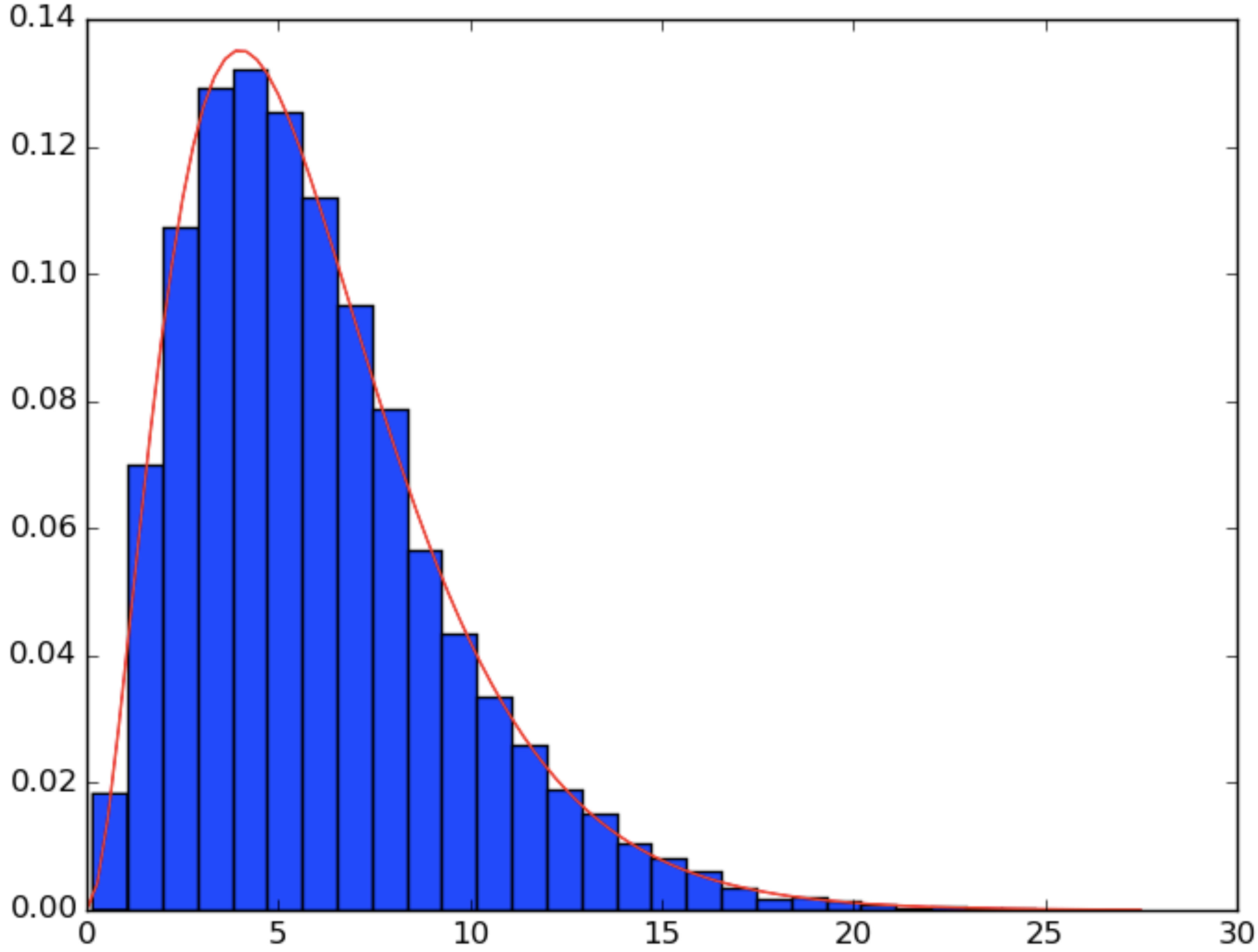
The random variables X and Y are **independent** if the events $\{X = x\}$ and $\{Y = y\}$ are independent. This means that

$$P(\{X = x\} \cap \{Y = y\}) = P(\{X = x\})P(\{Y = y\}),$$

which we can rewrite as

$$P(x, y) = P(x)P(y)$$

Continuous probability distribution



Continuous distribution: density function

$$p(x)dx = P(\{\text{event that } X \text{ takes a value in the range } [x, x + dx]\}).$$

Useful Facts: 4.1 *Properties of probability density functions*

- Probability density functions are non-negative. This follows from the definition; a negative value at some u would imply that $P(\{x \in [u, u + du]\})$ was negative, and this cannot occur.
- For $a < b$

  $$P(\{X \text{ takes a value in the range } [a, b]\}) = \int_a^b p(x)dx.$$

  which we obtain by summing $p(x)dx$ over all the infinitesimal intervals between a and b.
- We must have that

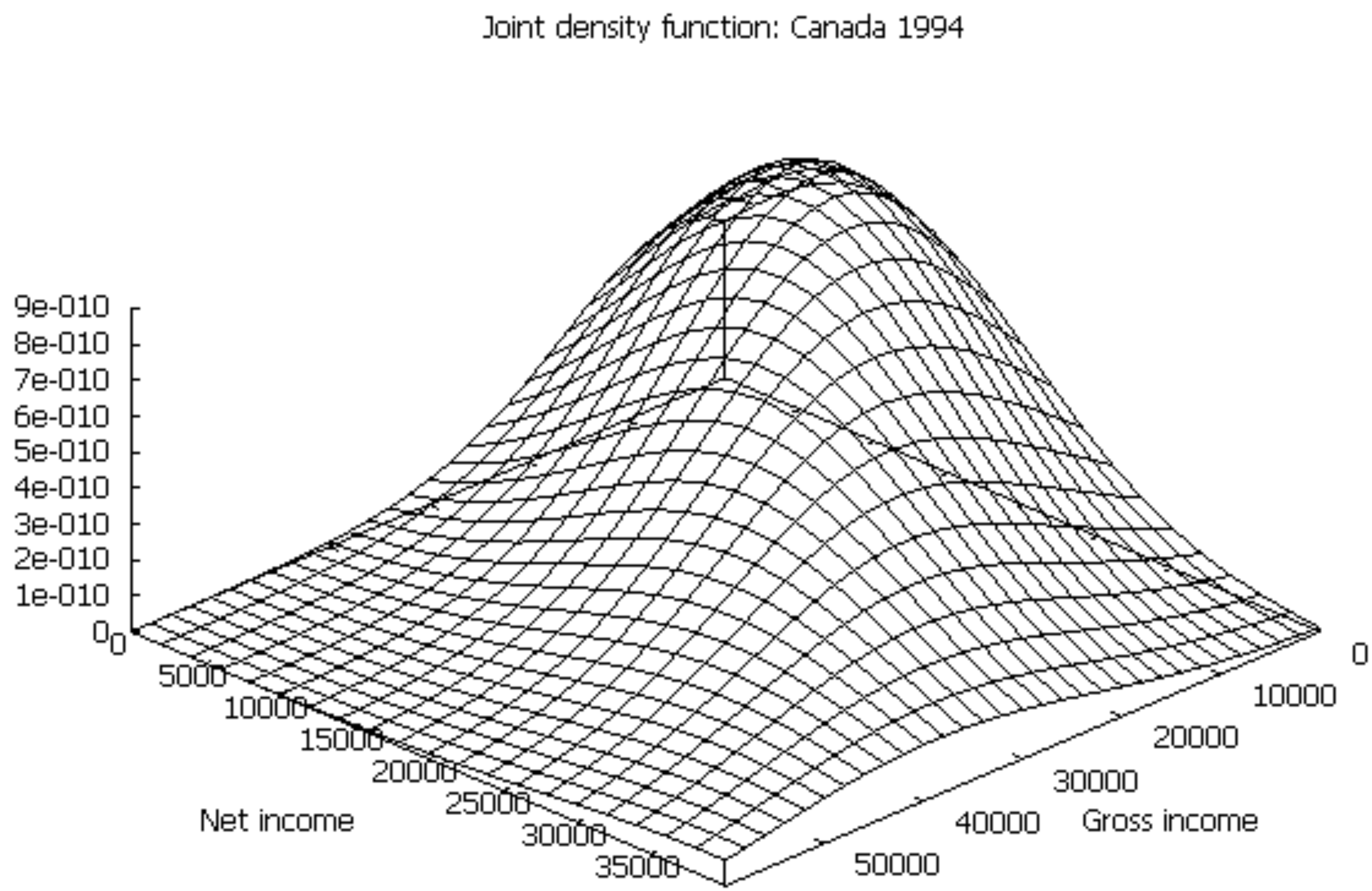
  $$\int_{-\infty}^{\infty} p(x)dx = 1.$$

  This is because

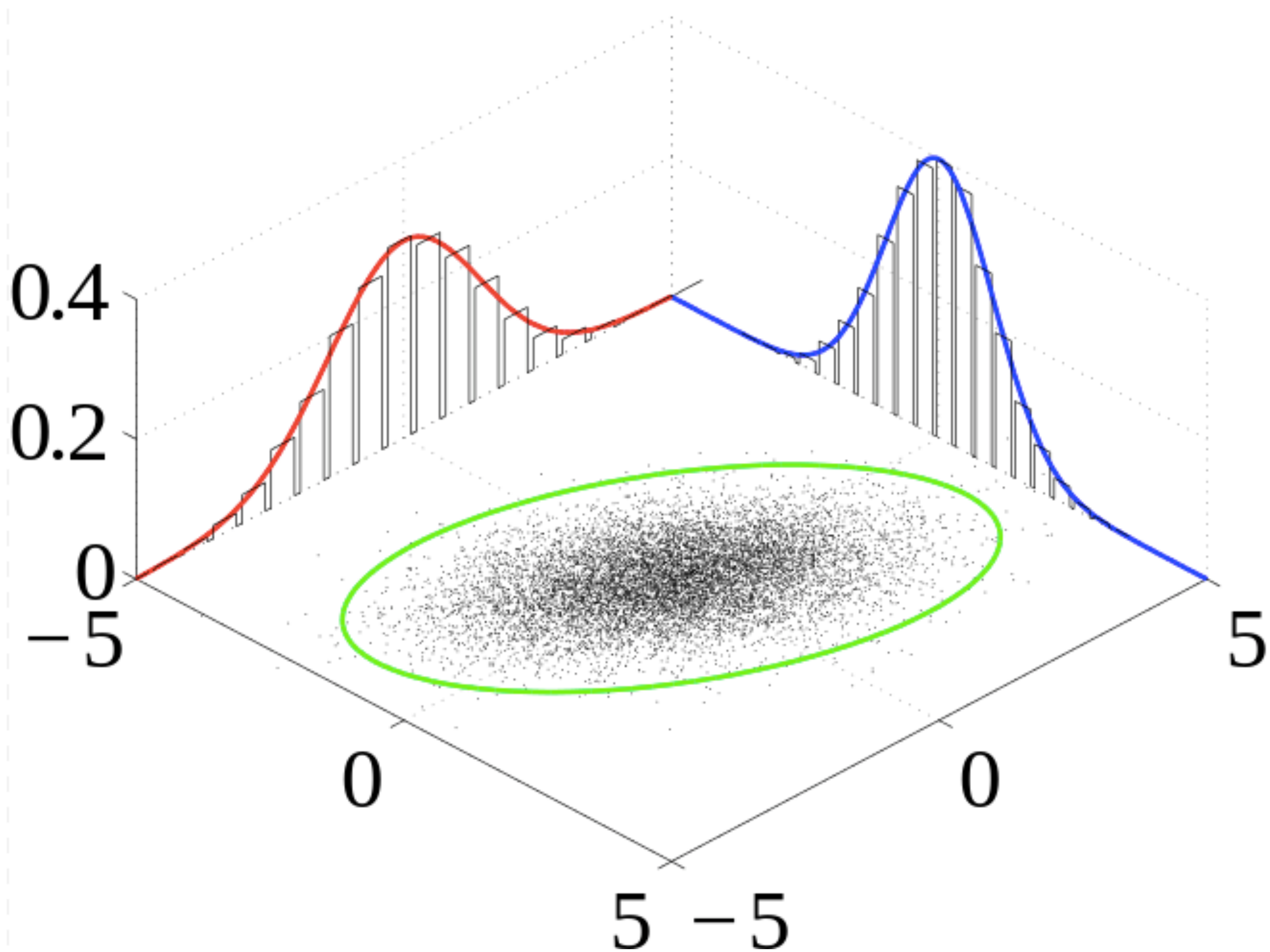
  $$P(\{X \text{ takes a value in the range } [-\infty, \infty]\}) = 1 = \int_{-\infty}^{\infty} p(x)dx$$

- Probability density functions are usually called pdf's.
- It is quite usual to write all pdf's as lower-case p's. If one specifically wishes to refer to probability (as opposed to probability density), one writes an upper case P, as in the previous points.

Joint distribution



Marginal distribution



Expected value

Definition: 4.8 *Expected value*

Given a discrete random variable X which takes values in the set $\mathcal{D}$ and which has probability distribution P, we define the expected value

$$\mathbb{E}[X] = \sum_{x \in \mathcal{D}} x P(X = x).$$

This is sometimes written $\mathbb{E}_P[X]$, to clarify which distribution one has in mind

Example: 4.5 *Betting on coins*

We agree to play the following game. I flip a fair coin (i.e. $P(H) = P(T) = 1/2$). If the coin comes up heads, you pay me 1; if the coin comes up tails, I pay you 1. The expected value of my income is 0, even though the random variable never takes that value.

Expectation

Definition: 4.9 *Expectation*

Assume we have a function f that maps a discrete random variable X into a set of numbers $\mathcal{D}_f$. Then $f(X)$ is a discrete random variable, too, which we write F. The expected value of this random variable is written

$$\mathbb{E}[f] = \sum_{u \in \mathcal{D}_f} uP(F = u) = \sum_{x \in \mathcal{D}} f(x)P(X = x)$$

which is sometimes referred to as "the expectation of f". The process of computing an expected value is sometimes referred to as "taking expectations".

Definition: 4.10 *Expected value of a continuous random variable*

Given a continuous random variable X which takes values in the set $\mathcal{D}$ and which has probability distribution P, we define the expected value

$$\mathbb{E}[X] = \int_{x \in \mathcal{D}} xp(x)dx.$$

This is sometimes written $\mathbb{E}_p[X]$, to clarify which distribution one has in mind.

Some properties of expectation

Useful Facts: 4.2 *Expectations are linear*

Write f, g for functions of random variables.

- $\mathbb{E}[0] = 0$
- for any constant k, $\mathbb{E}[kf] = k\mathbb{E}[f]$
- $\mathbb{E}[f + g] = \mathbb{E}[f] + \mathbb{E}[g]$.

Mean, Variance and Covariance

Definition: 4.12 *Mean or expected value*

The mean or expected value of a random variable X is

$$\mathbb{E}[X]$$

Definition: 4.13 *Variance*

The variance of a random variable X is

$$\mathsf{var}[X] = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big]$$

Definition: 4.14 *Covariance*

The covariance of two random variables X and Y is

$$\mathsf{cov}\,(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Examples

Worked example 4.9 *Mean of a coin flip*

We flip a biased coin, with $P(H) = p$. The random variable X has value 1 if the coin comes up heads, 0 otherwise. What is the mean of X? (i.e. $\mathbb{E}[X]$).

Solution: $\mathbb{E}[X] = \sum_{x \in D} xP(X = x) = 1p + 0(1 - p) = p$

Worked example 4.10 *Variance of a coin flip*

We flip a biased coin, with $P(H) = p$. The random variable X has value 1 if the coin comes up heads, 0 otherwise. What is the variance of X? (i.e. $\mathsf{var}[X]$).

Solution: $\mathsf{var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = (1p - 0(1-p)) - p^2 = p(1 - p)$

Worked example 4.11 *Variance*

Can a random variable have $\mathbb{E}[X] > \sqrt{\mathbb{E}[X^2]}$?

Solution: No, because that would mean that $\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] < 0$. But this is the expected value of a non-negative quantity; it must be non-negative.

Properties of variance and covariance

Useful Facts: 4.3 *Properties of variance*

1. For any constant k, $\mathsf{var}[k] = 0$
2. $\mathsf{var}[X] \geq 0$
3. $\mathsf{var}[kX] = k^2\mathsf{var}[X]$
4. if X and Y are independent, then $\mathsf{var}[X + Y] = \mathsf{var}[X] + \mathsf{var}[Y]$
5. $\mathsf{var}[X] = \mathsf{cov}(X, X)$.

1, 2, and 5 are obvious. You will prove 3 and 4 in the exercises.

Useful Facts: 4.6 *Independent random variables have zero covariance*

1. if X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
2. if X and Y are independent, then $\mathsf{cov}(X, Y) = 0$.

If 1 is true, then 2 is obviously true (apply the expression of useful facts 4.5). I prove 5 below.

Properties

Useful Facts: 4.4 *A useful expression for variance*

$$\begin{aligned}
\mathsf{var}[X] &= \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] \\
&= \mathbb{E}\left[(X^2-2X\mathbb{E}[X]+\mathbb{E}[X]^2)\right] \\
&= \mathbb{E}[X^2]-2\mathbb{E}[X]\mathbb{E}[X]+\mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2]-(\mathbb{E}[X])^2
\end{aligned}$$

Useful Facts: 4.5 *A useful expression for covariance*

$$\begin{aligned}
\mathsf{cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\
&= \mathbb{E}[(XY-Y\mathbb{E}[X]-X\mathbb{E}[Y]+\mathbb{E}[X]\mathbb{E}[Y])] \\
&= \mathbb{E}[XY]-2\mathbb{E}[Y]\mathbb{E}[X]+\mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y].
\end{aligned}$$

Proposition: *If X and Y are independent random variables, then* $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof: Recall that $\mathbb{E}[X] = \sum_{x \in D} xP(X = x)$, so that

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{(x,y) \in D_x \times D_y} xyP(X = x, Y = y) \\
&= \sum_{x \in D_x} \sum_{y \in D_y} (xyP(X = x, Y = y)) \\
&= \sum_{x \in D_x} \sum_{y \in D_y} (xyP(X = x)P(Y = y)) \\
& \quad \text{because } X \text{ and } Y \text{ are independent} \\
&= \sum_{x \in D_x} \sum_{y \in D_y} (xP(X = x))\,(yP(Y = y)) \\
&= \left(\sum_{x \in D_x} xP(X = x) \right) \left(\sum_{y \in D_y} yP(Y = y) \right) \\
&= (\mathbb{E}[X])(\mathbb{E}[Y]).
\end{aligned}$$

This is certainly not true when X and Y are not independent (try $Y = -X$).

Statistics

Mean

One simple and effective summary of a set of data is its **mean**. This is sometimes known as the **average** of the data.

Definition: 1.1 *Mean*

Assume we have a dataset $\{x\}$ of N data items, $x_1, \ldots, x_N$. Their mean is

$$\mathsf{mean}\left(\{x\}\right) = \frac{1}{N} \sum_{i=1}^{i=N} x_i.$$

Standard deviation and Variance

Definition: 1.2 *Standard deviation*

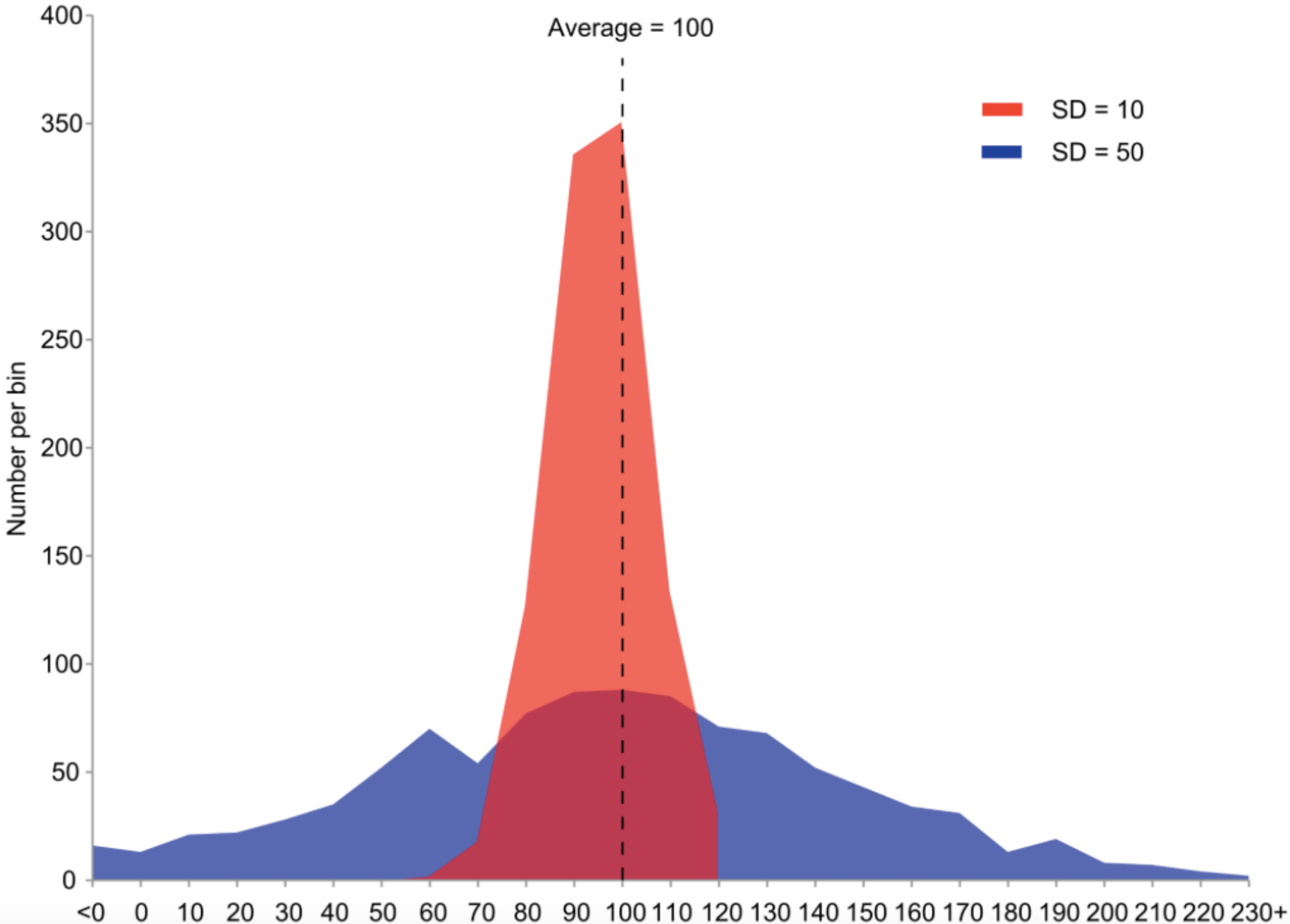
Assume we have a dataset $\{x\}$ of N data items, $x_1, \ldots, x_N$. The standard deviation of this dataset is is:

$$\mathsf{std}\left(\{x_i\}\right) = \sqrt{\frac{1}{N}\sum_{i=1}^{i=N}(x_i - \mathsf{mean}\left(\{x\}\right))^2} = \sqrt{\mathsf{mean}\left(\{(x_i - \mathsf{mean}\left(\{x\}\right))^2\}\right)}.$$

Definition: 1.3 *Variance*

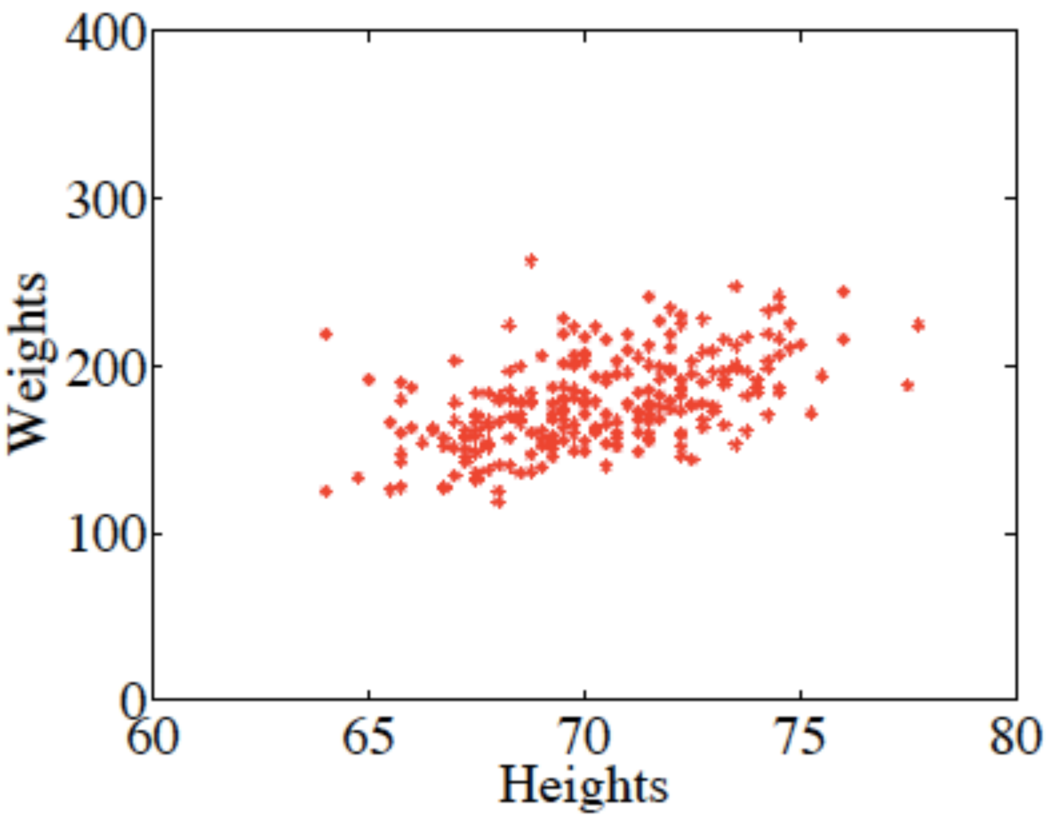
Assume we have a dataset $\{x\}$ of N data items, $x_1, \ldots, x_N$. where $N > 1$. Their variance is:

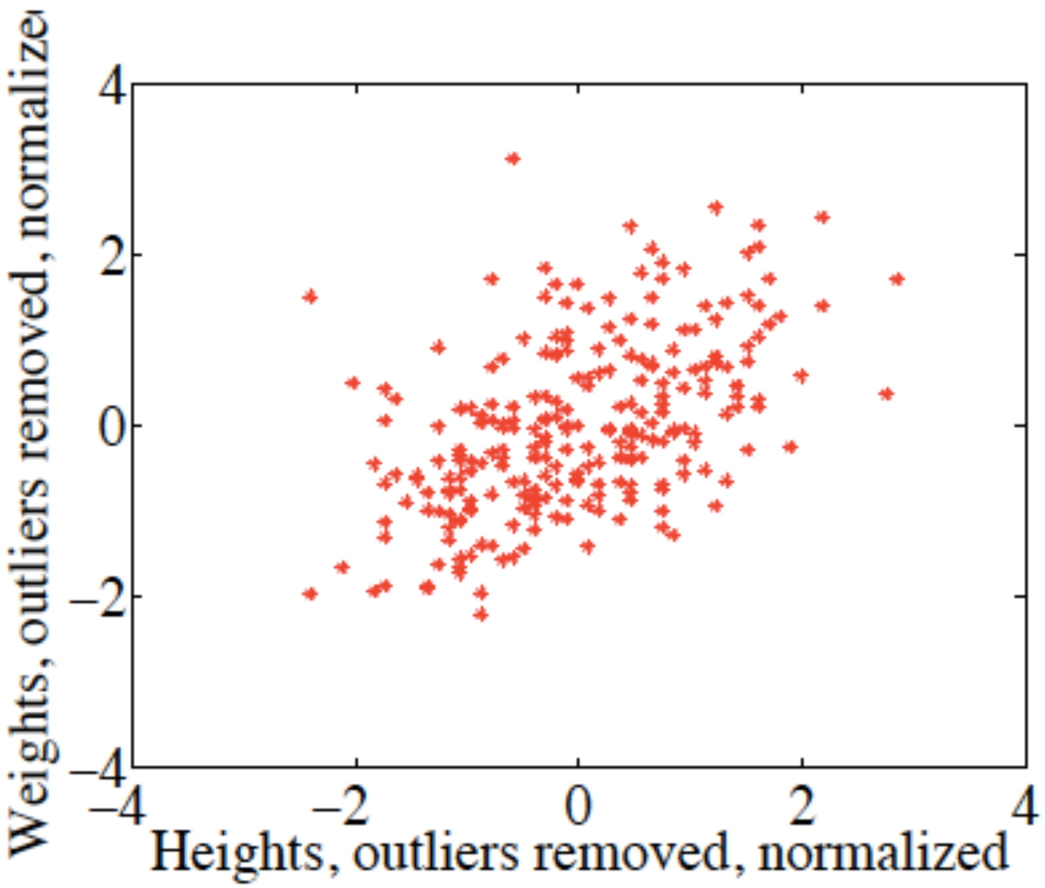
$$\mathsf{var}\left(\{x\}\right) = \frac{1}{N}\left(\sum_{i=1}^{i=N}(x_i - \mathsf{mean}\left(\{x\}\right))^2\right) = \mathsf{mean}\left(\left\{(x_i - \mathsf{mean}\left(\{x\}\right))^2\right\}\right).$$



Average = 100
SD = 10
SD = 50
Number per bin
400
350
300
250
200
150
100
50
0
<0
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230+

Normalization





$$\hat{x}_i = \frac{(x_i - \text{mean}(\{x\}))}{\text{std}(\{x\})}.$$

Correlation

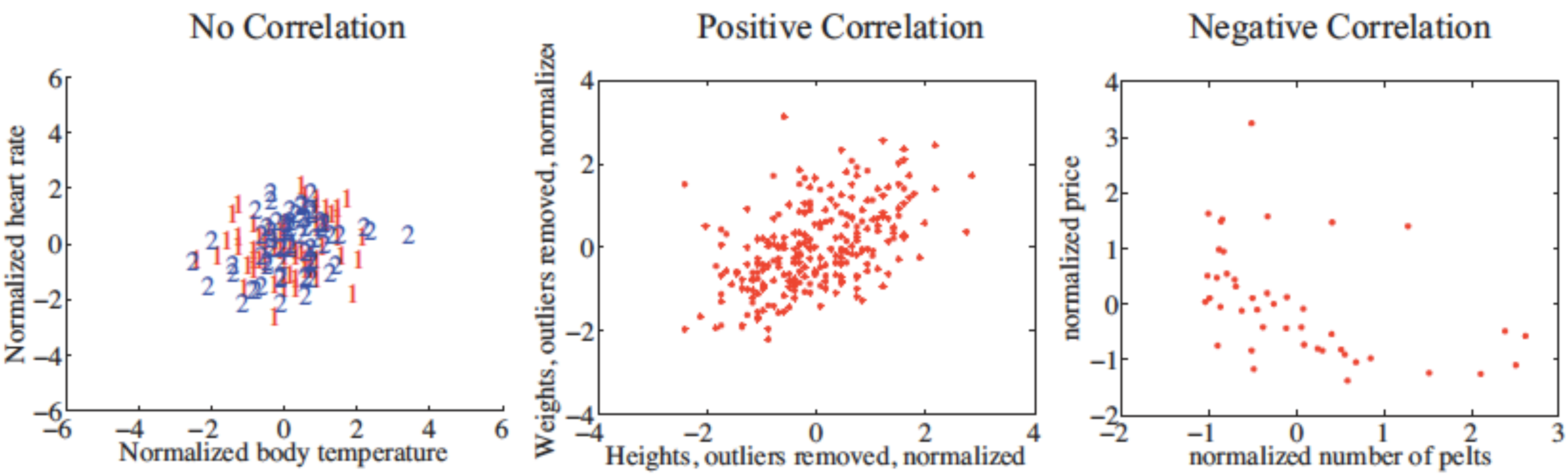


FIGURE 2.16: *The three kinds of scatter plot are less clean for real data than for our idealized examples. Here I used the body temperature vs heart rate data for the zero correlation; the height-weight data for positive correlation; and the lynx data for negative correlation. The pictures aren't idealized — real data tends to be messy — but you can still see the basic structures.*

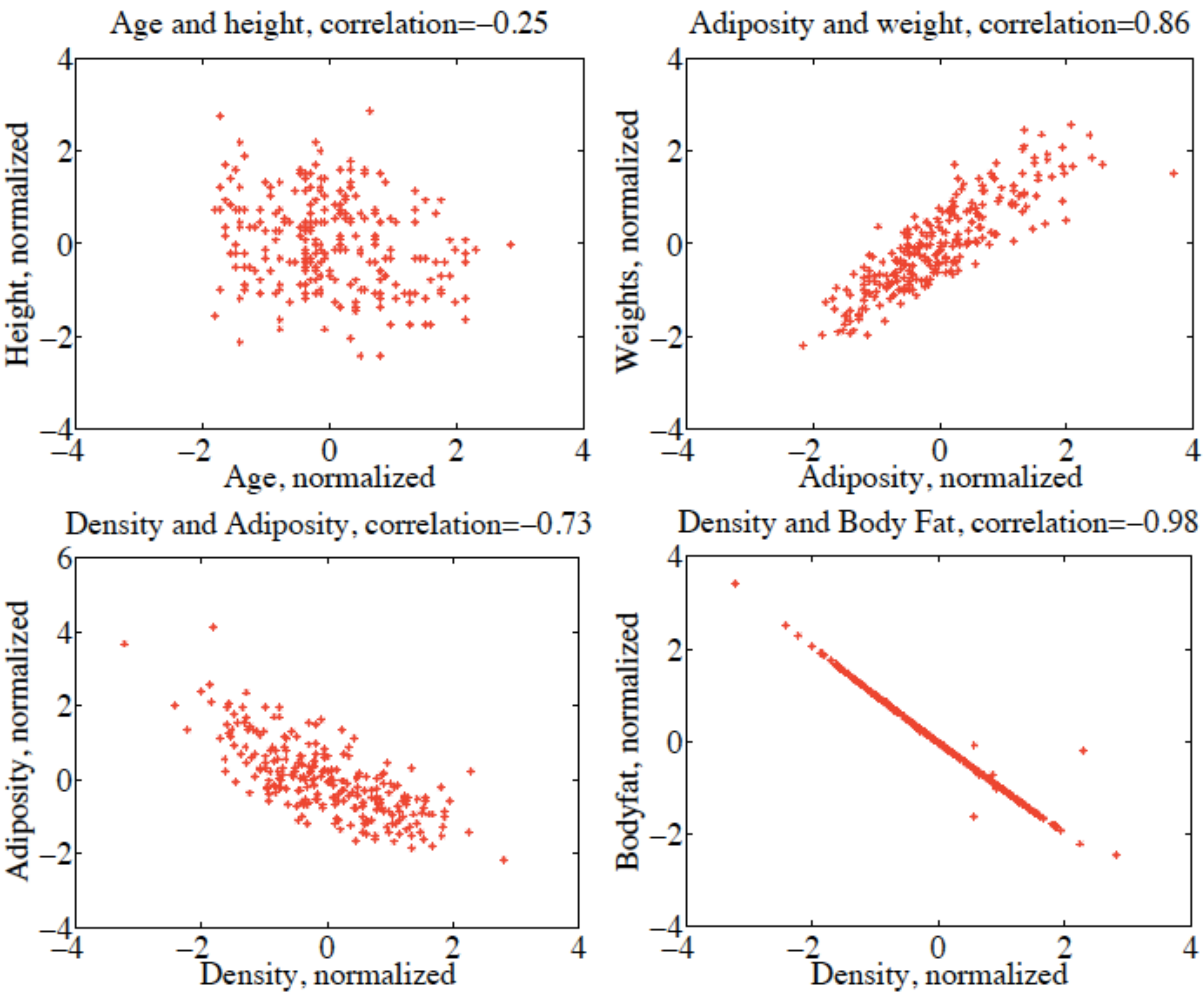
Correlation coefficient

Definition: 2.1 *Correlation coefficient*

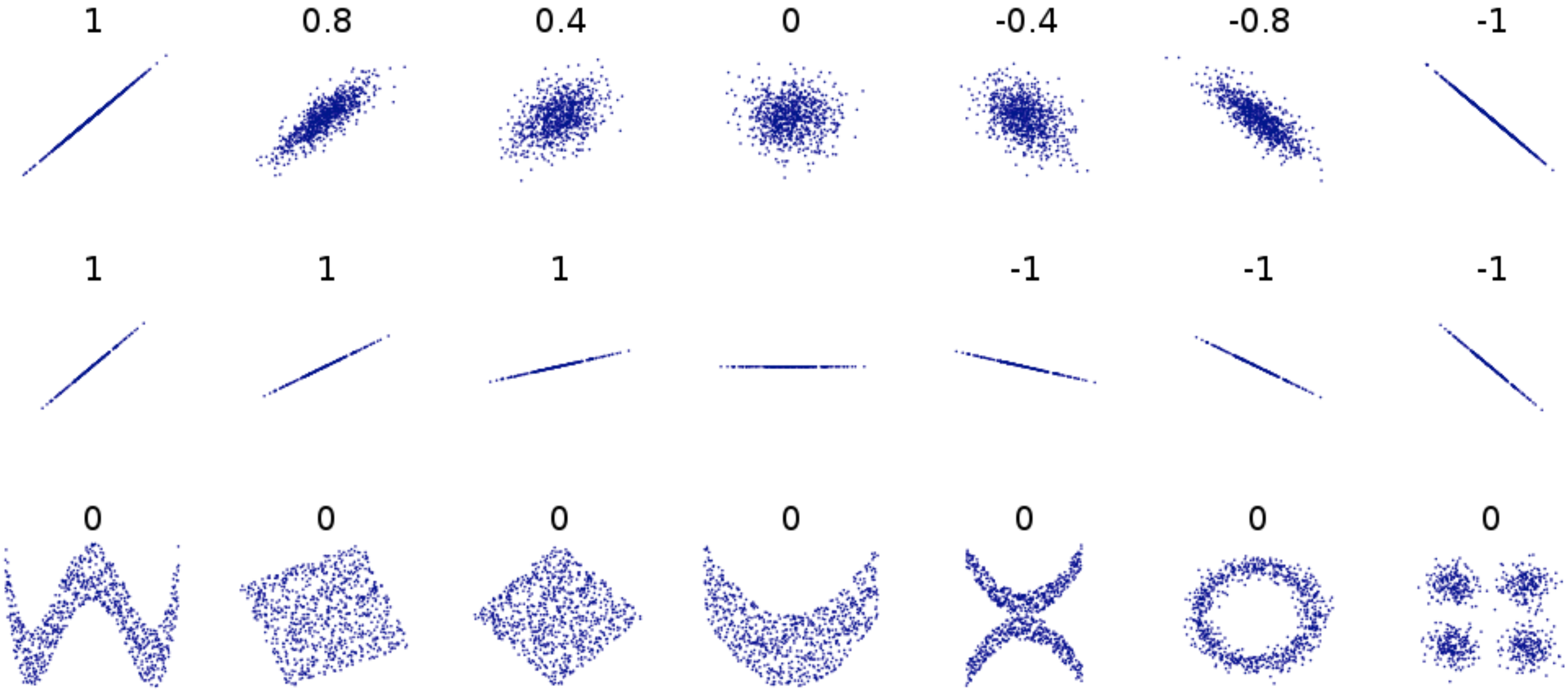
Assume we have N data items which are 2-vectors $(x_1, y_1), \ldots, (x_N, y_N)$, where $N > 1$. These could be obtained, for example, by extracting components from larger vectors. We compute the correlation coefficient by first normalizing the x and y coordinates to obtain $\hat{x}_i = \frac{(x_i - \mathsf{mean}(\{x\}))}{\mathsf{std}(x)}$, $\hat{y}_i = \frac{(y_i - \mathsf{mean}(\{y\}))}{\mathsf{std}(y)}$. The correlation coefficient is the mean value of $\hat{x}\hat{y}$, and can be computed as:

$$\mathsf{corr}\left(\{(x, y)\}\right) = \frac{\sum_i \hat{x}_i \hat{y}_i}{N}$$

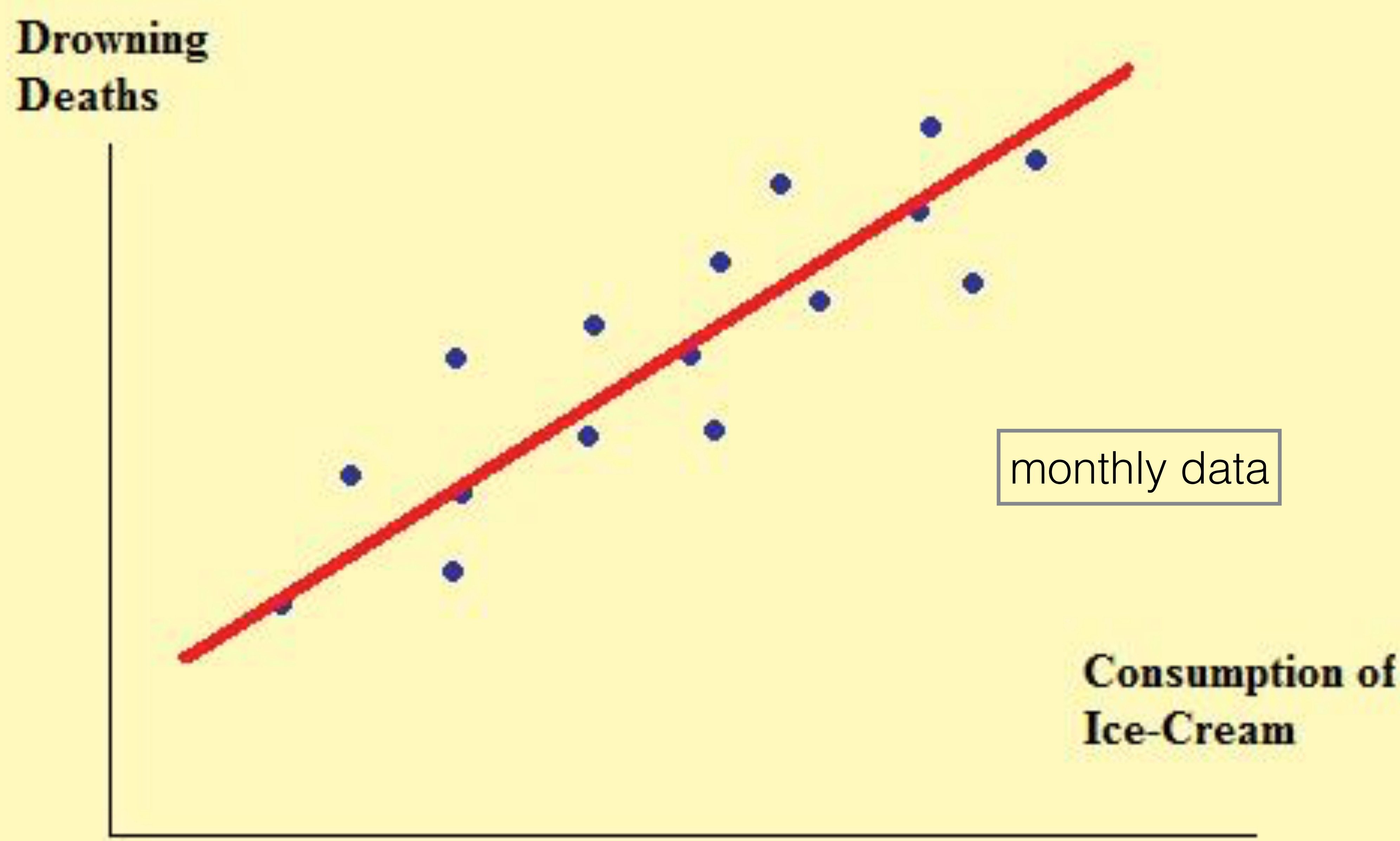
Also called **Pearson Correlation Coefficient**


Age and height, correlation=–0.25
Height, normalized
Age, normalized
Adiposity and weight, correlation=0.86
Weights, normalized
Adiposity, normalized
Density and Adiposity, correlation=–0.73
Adiposity, normalized
Density, normalized
Density and Body Fat, correlation=–0.98
Bodyfat, normalized
Density, normalized

Correlation coefficient vs Relationship

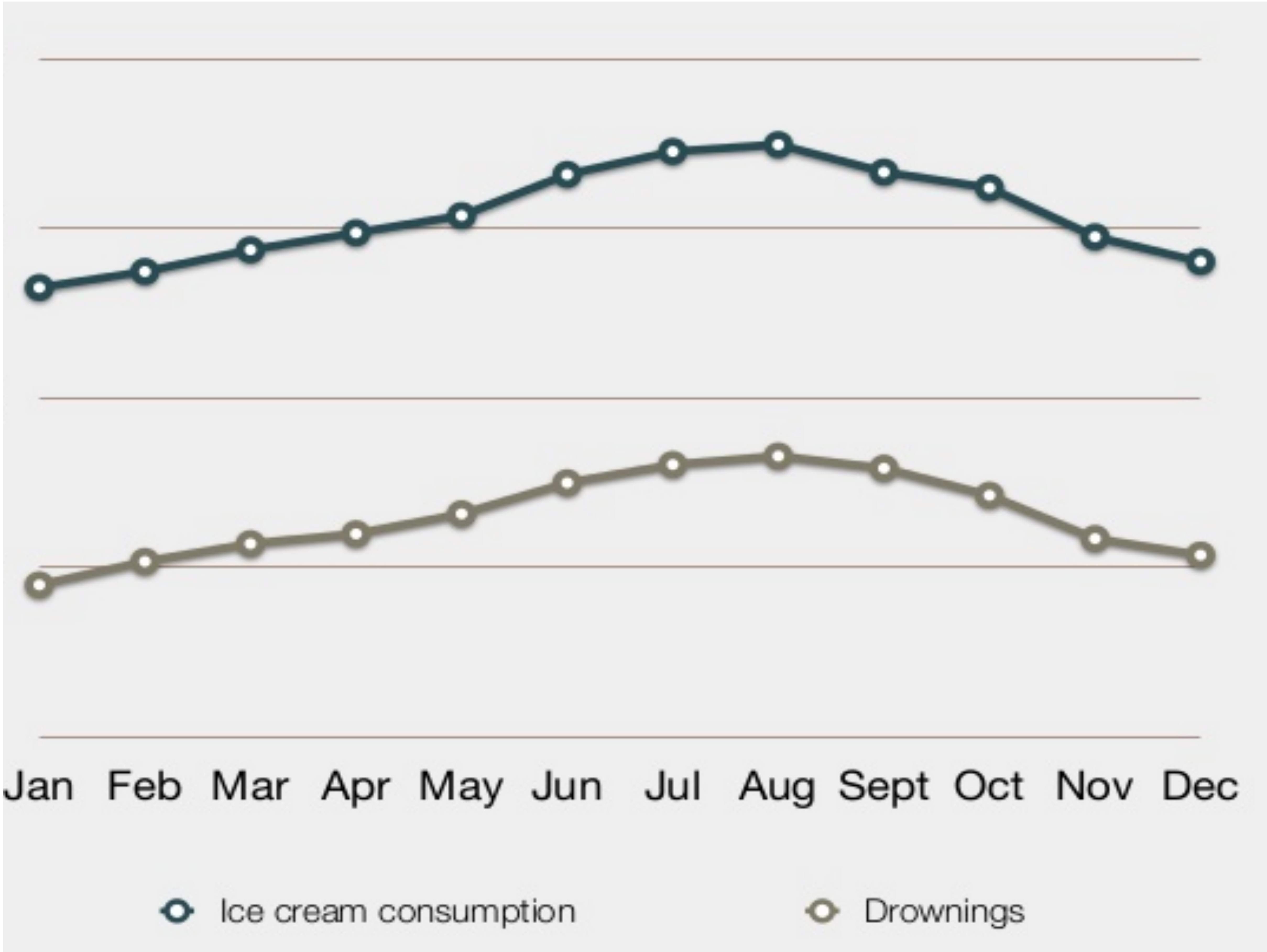


Correlation and Causality

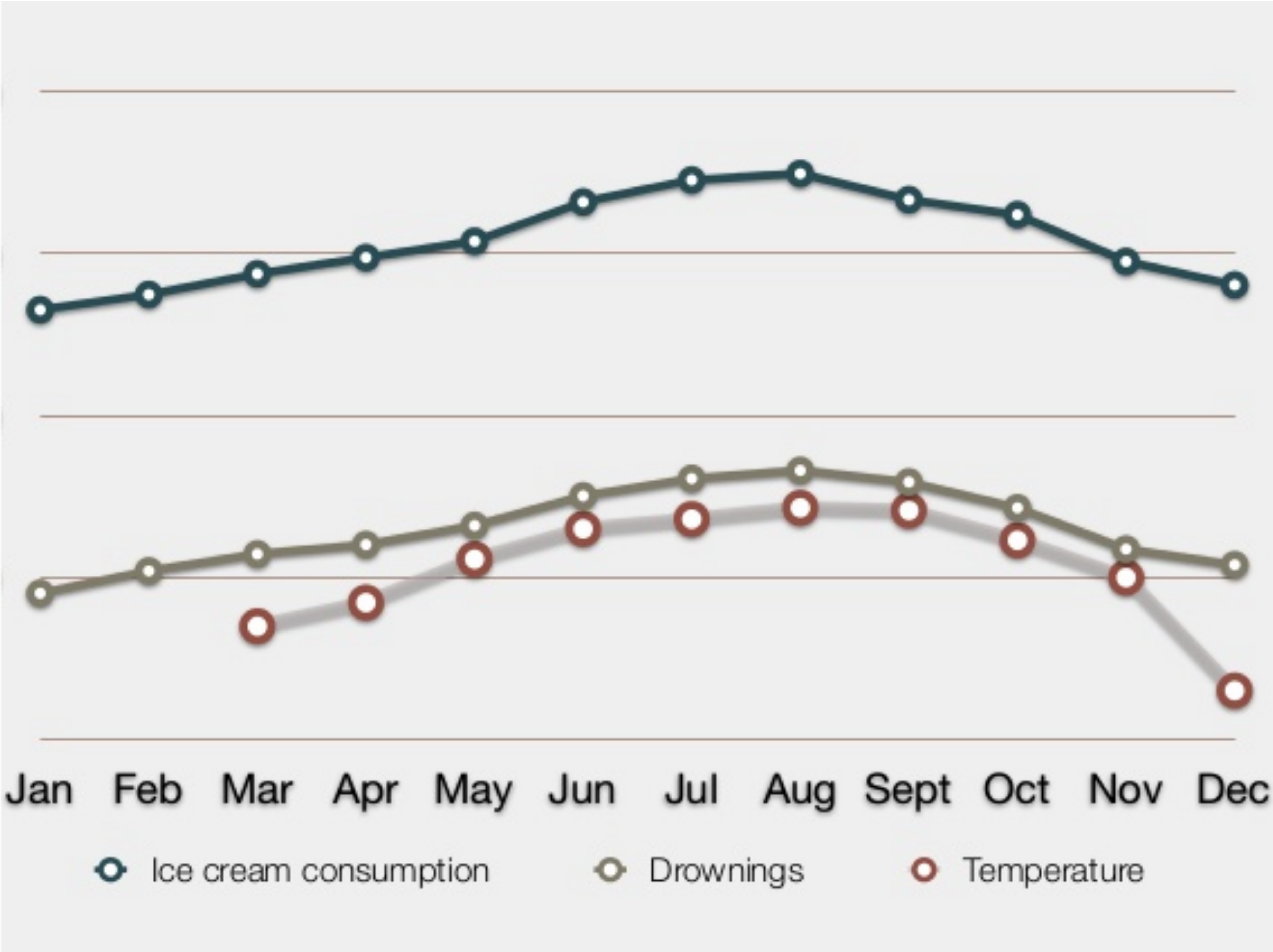


Ice Cream vs Drowning

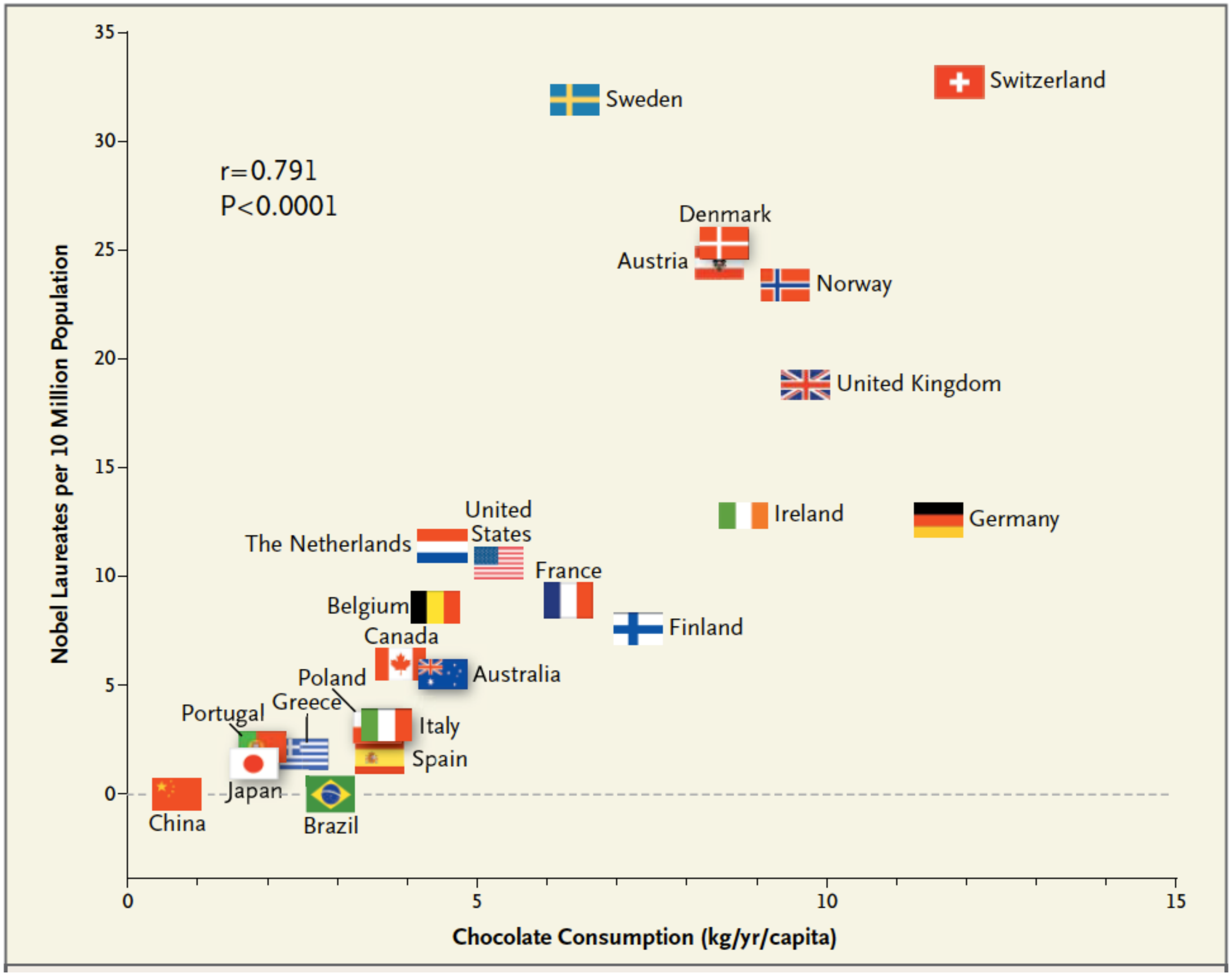
Ice Cream vs Drowning



Ice Cream vs Drowning



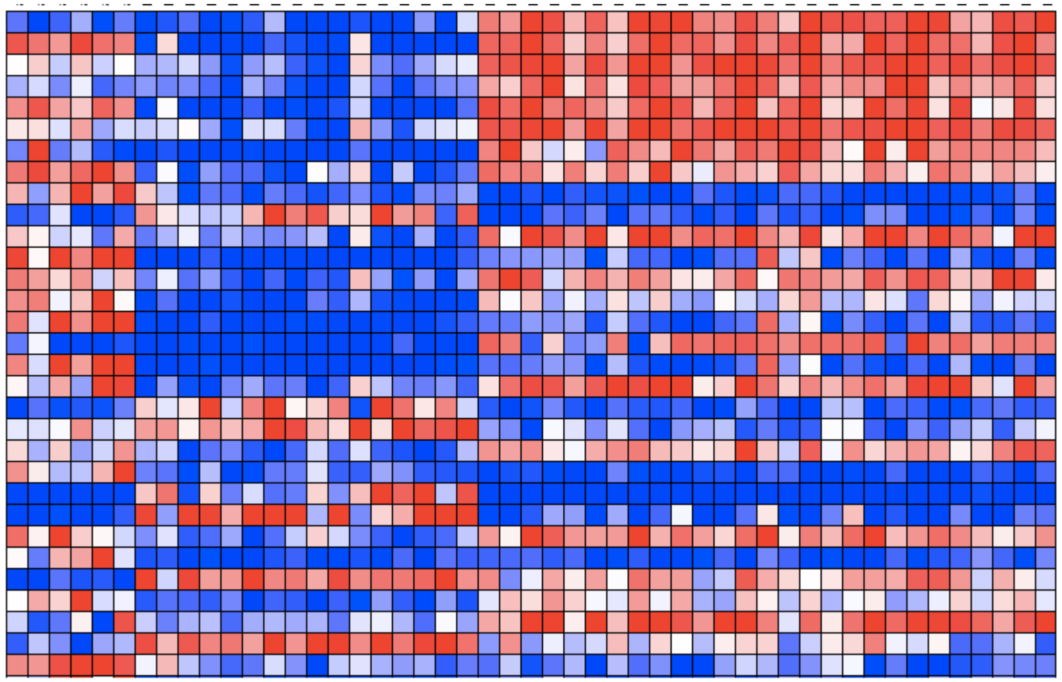
Chocolate vs Nobel Prizes



credit: NEJM, 2012

Gene expression analysis

sample



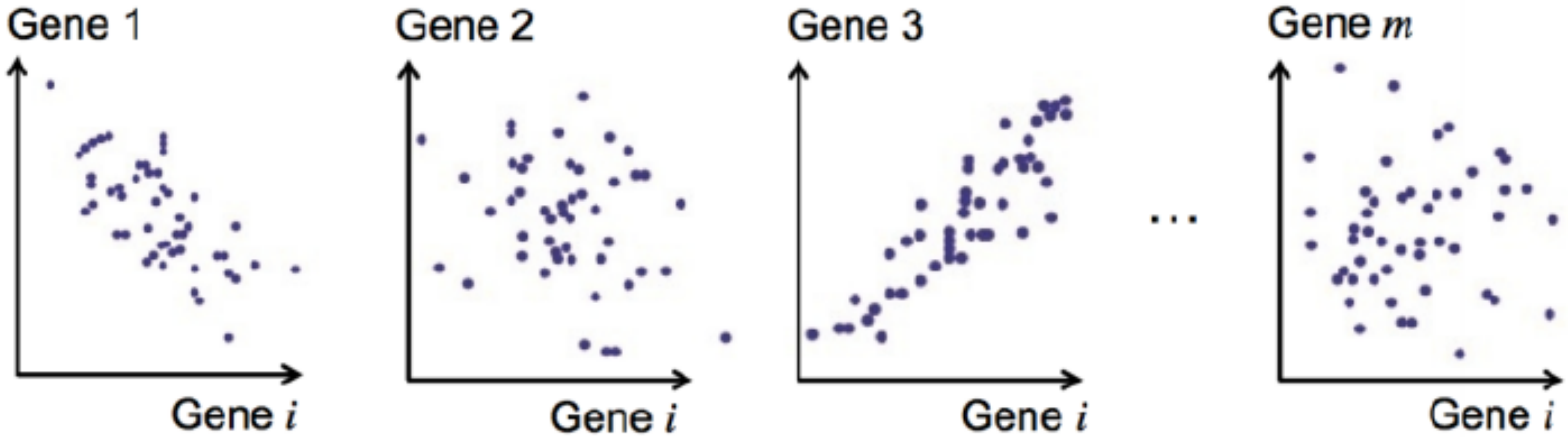
gene

Correlation of genes across experimental conditions ⇨ coregulation of genes

Correlation analysis

	Sample 1	Sample 2	···	Sample n
Gene 1	X_{11}	X_{12}	···	X_{1n}
Gene 2	X_{21}	X_{22}	···	X_{2n}
⋮	⋮	⋮	⋮	⋮
Gene m	X_{m1}	X_{m2}	···	X_{mn}

$$r = \frac{\sum(X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum(X - \overline{X})^2}\sqrt{\sum(Y - \overline{Y})^2}}$$



r=-0.8 r=-0.2 r=0.85 r=-0.15