Basic Stochastics and the idea of testing

Dirk Metzler

http://evol.bio.lmu.de/_statgen

October 10, 2017

Assume a small population of 100 individuals, and a neutral allele A that has frequency 0.3 in this generation.

Assume a small population of 100 individuals, and a neutral allele A that has frequency 0.3 in this generation.

What will be the frequency X of A in the next generation?

Assume a small population of 100 individuals, and a neutral allele A that has frequency 0.3 in this generation.

What will be the frequency X of A in the next generation?

We don't know, as X is a random variable .

Assume a small population of 100 individuals, and a neutral allele A that has frequency 0.3 in this generation.

What will be the frequency X of A in the next generation?

We don't know, as X is a random variable .

However, we can ask, for example, for

$\mathbb{E}X$, the expectation value of X,

Assume a small population of 100 individuals, and a neutral allele A that has frequency 0.3 in this generation.

What will be the frequency X of A in the next generation?

We don't know, as X is a random variable .

However, we can ask, for example, for

$\mathbb{E}X$, the expectation value of X, or for

$\Pr(X = 0.32)$, the probability that X takes a value of 0.32.

Assume a small population of 100 individuals, and a neutral allele A that has frequency 0.3 in this generation.

What will be the frequency X of A in the next generation?

We don't know, as X is a random variable .

However, we can ask, for example, for

$\mathbb{E}X$, the expectation value of X, or for

$\Pr(X = 0.32)$, the probability that X takes a value of 0.32.

Even these values (especially the second on) depend on our model assumptions.

Contents

We start with a simpler Example: Rolling a dice, W is the result of the next trial.

$$\mathcal{S} = \{1, 2, \ldots, 6\}$$
$$\Pr(W = 1) = \cdots = \Pr(W = 6) = \tfrac{1}{6}$$
$$(\ \Pr(W = x) = \tfrac{1}{6} \text{ for all } x \in \{1, \ldots, 6\}\ )$$

A Random Variable is a
result of a random incident or experiment.

A Random Variable is a
result of a random incident or experiment.

The state space $\mathcal{S}$
of a random variable is the set of possible values.

A Random Variable is a
result of a random incident or experiment.

The state space $\mathcal{S}$
of a random variable is the set of possible values.

The distribution of a random variable X
assigns to each set $A \subseteq \mathcal{S}$
the probability $\Pr(X \in A)$
that X takes a value in A.

A Random Variable is a
result of a random incident or experiment.

The state space $\mathcal{S}$
of a random variable is the set of possible values.

The distribution of a random variable X
assigns to each set $A \subseteq \mathcal{S}$
the probability $\Pr(X \in A)$
that X takes a value in A.

In general, we use capitals for
random variables ($X, Y, Z, \ldots$),
and small letters ($x, y, z, \ldots$) for (possible) fixed values.

Notations for events

The event U that X takes a value in A can be written with curly brackets:

$$U = \{X \in A\}$$

We can interpret this as the set of results (elementary events) for which the event is fulfilled.

Notations for events

The event U that X takes a value in A can be written with curly brackets:

$$U = \{X \in A\}$$

We can interpret this as the set of results (elementary events) for which the event is fulfilled.
Thus, events have a lot in common with sets, and similar notations as for sets are used for events U and V:

$$U \cap V = U \text{ "and" } V$$

is the event that takes place if and only if both U and V take place.

$$U \cup V = U \text{ "or" } V$$

is the event that takes place if and only if U or V (or both) take place.

Example

Let X and Y be the results of two dice rolls, $A = \{1, 2, 3\}$, and $B = \{1, 3, 5\}$. Then:

$$\begin{aligned} \{X \in A\} \cap \{X \in B\} &= \{X \in A \cap B\} = \{X \in \{1, 3\}\} \\ &= \{X = 1\} \cup \{X = 3\} \end{aligned}$$

and

$$\{Y \in A\} \cup \{Y \in B\} = \{Y \in A \cup B\} = \{Y \in \{1, 2, 3, 5\}\}$$

Example

Let X and Y be the results of two dice rolls, $A = \{1, 2, 3\}$, and $B = \{1, 3, 5\}$. Then:

$$\begin{aligned} \{X \in A\} \cap \{X \in B\} &= \{X \in A \cap B\} = \{X \in \{1, 3\}\} \\ &= \{X = 1\} \cup \{X = 3\} \end{aligned}$$

and

$$\{Y \in A\} \cup \{Y \in B\} = \{Y \in A \cup B\} = \{Y \in \{1, 2, 3, 5\}\}$$

and

$$\{X \in A\} \cap \{Y \in B\} =$$

Example

Let X and Y be the results of two dice rolls, $A = \{1, 2, 3\}$, and $B = \{1, 3, 5\}$. Then:

$$\begin{aligned} \{X \in A\} \cap \{X \in B\} &= \{X \in A \cap B\} = \{X \in \{1, 3\}\} \\ &= \{X = 1\} \cup \{X = 3\} \end{aligned}$$

and

$$\{Y \in A\} \cup \{Y \in B\} = \{Y \in A \cup B\} = \{Y \in \{1, 2, 3, 5\}\}$$

and

$$\{X \in A\} \cap \{Y \in B\} = \{(X, Y) \in A \times B\},$$

Example

Let X and Y be the results of two dice rolls, $A = \{1, 2, 3\}$, and $B = \{1, 3, 5\}$. Then:

$$\begin{aligned} \{X \in A\} \cap \{X \in B\} &= \{X \in A \cap B\} = \{X \in \{1, 3\}\} \\ &= \{X = 1\} \cup \{X = 3\} \end{aligned}$$

and

$$\{Y \in A\} \cup \{Y \in B\} = \{Y \in A \cup B\} = \{Y \in \{1, 2, 3, 5\}\}$$

and

$$\{X \in A\} \cap \{Y \in B\} = \{(X, Y) \in A \times B\}, \text{where}$$

$$A \times B = \{(1, 1), (1, 3), (1, 5), (2, 1), (2, 3), (2, 5), (3, 1), (3, 3), (3, 5)\}.$$

The intersection

$$\{X \in A\} \cap \{X \in B\} = \{X \in A, X \in B\} = \{X \in A \cap B\}$$

is then the event that X takes a value that is in A and in B.

The intersection

$$\{X \in A\} \cap \{X \in B\} = \{X \in A, X \in B\} = \{X \in A \cap B\}$$

is then the event that X takes a value that is in A and in B.
The join

$$\{X \in A\} \cup \{X \in B\} = \{X \in A \cup B\}$$

is the event that the event that X takes a value in A or in B (or both).

The intersection

$$\{X \in A\} \cap \{X \in B\} = \{X \in A, X \in B\} = \{X \in A \cap B\}$$

is then the event that X takes a value that is in A and in B.
The join

$$\{X \in A\} \cup \{X \in B\} = \{X \in A \cup B\}$$

is the event that the event that X takes a value in A or in B (or both).
Sometimes the curly brackets are not written:

$$\Pr(X \in A, X \in B) = \Pr(\{X \in A, X \in B\})$$

Calculation rules:

Example Rolling a dice W:

$$\begin{aligned}\Pr(W \in \{2,3\}) &= \frac{2}{6} = \frac{1}{6} + \frac{1}{6} \\ &= \Pr(W = 2) + \Pr(W = 3) \\ \Pr(W \in \{1,2\} \cup \{3,4\}) &= \frac{4}{6} = \frac{2}{6} + \frac{2}{6} \\ &= \Pr(W \in \{1,2\}) + \Pr(W \in \{3,4\})\end{aligned}$$

Caution:

$$\begin{aligned}\Pr(W \in \{2,3\}) + \Pr(W \in \{3,4\}) &= \frac{2}{6} + \frac{2}{6} = \frac{4}{6} \\ &\neq \Pr(W \in \{2,3,4\}) = \frac{3}{6}\end{aligned}$$

Example: rolling two dice (W_1, W_2):
Let W_1 and W_2 the result of dice 1 and dice 2.

$$\begin{aligned}
&\Pr(W_1 \in \{4\}, W_2 \in \{2,3,4\}) \\
&= \Pr((W_1, W_2) \in \{(4,2),(4,3),(4,4)\}) \\
&= \frac{3}{36} = \frac{1}{6} \cdot \frac{3}{6} \\
&= \Pr(W_1 \in \{4\}) \cdot \Pr(W_2 \in \{2,3,4\})
\end{aligned}$$

In general:

$$\Pr(W_1 \in A, W_2 \in B) = \Pr(W_1 \in A) \cdot \Pr(W_2 \in B)$$

for all sets $A, B \subseteq \{1, 2, \dots, 6\}$

If S is the sum of the results $S = W_1 + W_2$,
what is the probability that $S = 5$,
if dice 1 shows $W_1 = 2$?

$$\Pr(S = 5 | W_1 = 2) \stackrel{!}{=} \Pr(W_2 = 3)$$
$$= \frac{1}{6} = \frac{1/36}{1/6} = \frac{\Pr(S{=}5, W_1{=}2)}{\Pr(W_1{=}2)}$$

If S is the sum of the results $S = W_1 + W_2$,
what is the probability that $S = 5$,
if dice 1 shows $W_1 = 2$?

$$\Pr(S = 5 | W_1 = 2) \stackrel{!}{=} \Pr(W_2 = 3)$$
$$= \frac{1}{6} = \frac{1/36}{1/6} = \frac{\Pr(S=5, W_1=2)}{\Pr(W_1=2)}$$

What is the probability $S \in \{4, 5\}$ under the condition $W_1 \in \{1, 6\}$?

$$\begin{aligned}
&\Pr(S \in \{4,5\} | W_1 \in \{1,6\}) \\
&= \frac{\Pr(S \in \{4,5\}, W_1 \in \{1,6\})}{\Pr(W_1 \in \{1,6\})} \\
&= \frac{\Pr(W_2 \in \{3,4\}, W_1 = 1)}{\Pr(W_1 \in \{1,6\})} \\
&= \frac{2/36}{2/6} = \frac{1}{6}
\end{aligned}$$

Calculation rules:

- $0 \leq \Pr(U) \leq 1$ for each event U (in the probability space).
- $\Pr(X \in \mathcal{S}) = 1$ if X is a random variable with state space $\mathcal{S}$.
- If the events U and V exclude each other, then

$$\Pr(U \cup V) = \Pr(U) + \Pr(V)$$

- The general rule is the inclusion-exclusion formula

$$\Pr(U \cup V) = \Pr(U) + \Pr(V) - \Pr(U \cap V)$$

- Definition of conditional probabilities:
  The probability of U under the condition V

$$\Pr(U|V) := \frac{\Pr(U, V)}{\Pr(V)}$$

  "Conditional probability of U given V"

Calculation rules:

- $0 \leq \Pr(U) \leq 1$ for each event U (in the probability space).
- $\Pr(X \in \mathcal{S}) = 1$ if X is a random variable with state space $\mathcal{S}$.
- If the events U and V exclude each other, then

$$\Pr(U \cup V) = \Pr(U) + \Pr(V)$$

- The general rule is the inclusion-exclusion formula

$$\Pr(U \cup V) = \Pr(U) + \Pr(V) - \Pr(U \cap V)$$

- Definition of conditional probabilities:
  The probability of U under the condition V

$$\Pr(U|V) := \frac{\Pr(U, V)}{\Pr(V)}$$

“Conditional probability of U given V”

Note: $\Pr(U, V) = \Pr(V) \cdot \Pr(U| V)$

How to say

$$\Pr(X \in A,\ Y \in B) = \Pr(X \in A) \cdot \Pr(Y \in B \mid X \in A)$$

in words:

How to say

$$\Pr(X \in A,\ Y \in B) = \Pr(X \in A) \cdot \Pr(Y \in B \mid X \in A)$$

in words:

The probability of $\{X \in A,\ Y \in B\}$ can be computed in two steps:

- First, the event $\{X \in A\}$ must take place.
- Multiply its probability with the conditional probability of $\{Y \in B\}$, given that $\{X \in A\}$ is already known to take place.

Stochastic Independence of events

Definition (stochastic independence)

Two events U and V are (stochastically) independent, if the identity

$$\Pr(U, V) = \Pr(U) \cdot \Pr(V)$$

holds for all events U, V.

Stochastic Independence of events

Definition (stochastic independence)

Two events U and V are (stochastically) independent, if the identity

$$\Pr(U, V) = \Pr(U) \cdot \Pr(V)$$

holds for all events U, V.

Note that $\Pr(U, V) = \Pr(U) \cdot \Pr(V)$ is equivalent to

$$\Pr(U|V) = \Pr(U) \text{ and also to } \Pr(V|U) = \Pr(V)$$

Stochastic Independence of random variables

Definition (stochastic independence)

Two random variables X and Y are (stochastically) independent, if the identity

$$\Pr(X \in A, Y \in B) = \Pr(X \in A) \cdot \Pr(Y \in B)$$

holds for all (measurable) subsets A and B of the state spaces of X and Y.

Stochastic Independence of random variables

Definition (stochastic independence)

Two random variables X and Y are (stochastically) independent, if the identity

$$\Pr(X \in A, Y \in B) = \Pr(X \in A) \cdot \Pr(Y \in B)$$

holds for all (measurable) subsets A and B of the state spaces of X and Y.

Example:

- Tossing two dice:
  $X =$ result dice 1, $Y =$ result dice 2.

$$\Pr(X = 2,\ Y = 5) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \Pr(X = 2) \cdot \Pr(Y = 5)$$

Contents

Bernoulli distribution

A Bernoulli experiment is an experiment with two possible oucomes "success" and "fail", or 1 or 0.

Bernoulli distribution

A Bernoulli experiment is an experiment with two possible oucomes "success" and "fail", or 1 or 0.

Bernoulli distribution

A Bernoulli experiment is an experiment with two possible oucomes "success" and "fail", or 1 or 0.

Bernoulli random variable X:

State space $\mathcal{S} = \{0, 1\}$.
Distribution:

$$\Pr(X = 1) = p$$
$$\Pr(X = 0) = 1 - p$$

The parameter $p \in [0, 1]$ is the success probability.

Bernoulli distribution

Examples:

- Tossing a coin: Possible outcomes are “head” and “tail”

Bernoulli distribution

Examples:

- Tossing a coin: Possible outcomes are “head” and “tail”
- Does the Drosophila have a mutation that causes white eyes? Possible outcomes are “yes” or “no”.

Bernoulli distribution

Examples:

- Tossing a coin: Possible outcomes are “head” and “tail”
- Does the Drosophila have a mutation that causes white eyes? Possible outcomes are “yes” or “no”.
- The sex of a newborn child has the values “male” or “female”.

Assume a Bernoulli experiment (for example tossing a coin) with success probability p is repeated n times *independently*.

Assume a Bernoulli experiment (for example tossing a coin) with success probability p is repeated n times *independently*.

What is the probability that it...

1. ...alway succeeds?

Assume a Bernoulli experiment (for example tossing a coin) with success probability p is repeated n times *independently*.

What is the probability that it...

1. ...alway succeeds?

$$p \cdot p \cdot p \cdots p = p^n$$

Assume a Bernoulli experiment (for example tossing a coin) with success probability p is repeated n times *independently*.

What is the probability that it...

1. ...alway succeeds?

$$p \cdot p \cdot p \cdots p = p^n$$

2. ...always fails?

Assume a Bernoulli experiment (for example tossing a coin) with success probability p is repeated n times *independently*.

What is the probability that it...

1. ...alway succeeds?

$$p \cdot p \cdot p \cdots p = p^n$$

2. ...always fails?

$$(1-p) \cdot (1-p) \cdots (1-p) = (1-p)^n$$

Assume a Bernoulli experiment (for example tossing a coin) with success probability p is repeated n times *independently*.

What is the probability that it...

1. ...alway succeeds?

$$p \cdot p \cdot p \cdots p = p^n$$

2. ...always fails?

$$(1-p) \cdot (1-p) \cdots (1-p) = (1-p)^n$$

3. ...first succeeds k times and then fails $n-k$ times?

Assume a Bernoulli experiment (for example tossing a coin) with success probability p is repeated n times *independently*.

What is the probability that it...

1. ...alway succeeds?

$$p \cdot p \cdot p \cdots p = p^n$$

2. ...always fails?

$$(1-p) \cdot (1-p) \cdots (1-p) = (1-p)^n$$

3. ...first succeeds k times and then fails $n-k$ times?

$$p^k \cdot (1-p)^{n-k}$$

4. ...succeeds in total k times and fails the other $n-k$ times?

Assume a Bernoulli experiment (for example tossing a coin) with success probability p is repeated n times *independently*.

What is the probability that it...

1. ...alway succeeds?

$$p \cdot p \cdot p \cdots p = p^n$$

2. ...always fails?

$$(1 - p) \cdot (1 - p) \cdots (1 - p) = (1 - p)^n$$

3. ...first succeeds k times and then fails $n - k$ times?

$$p^k \cdot (1 - p)^{n-k}$$

4. ...succeeds in total k times and fails the other $n - k$ times?

$$\binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$$

Note

$\binom{n}{k} = \frac{n!}{k!\cdot(n-k)!}$ ("*n* choose *k*") is the number of possibilities to choose *k* successes in *n* trials.

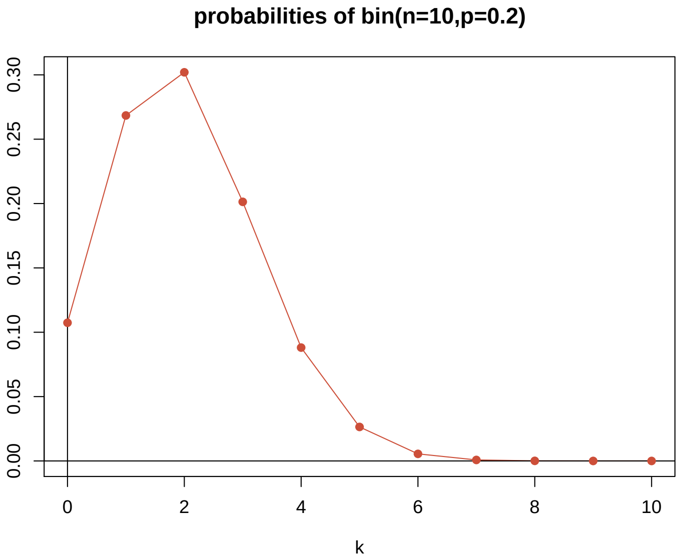
Binomial distribution

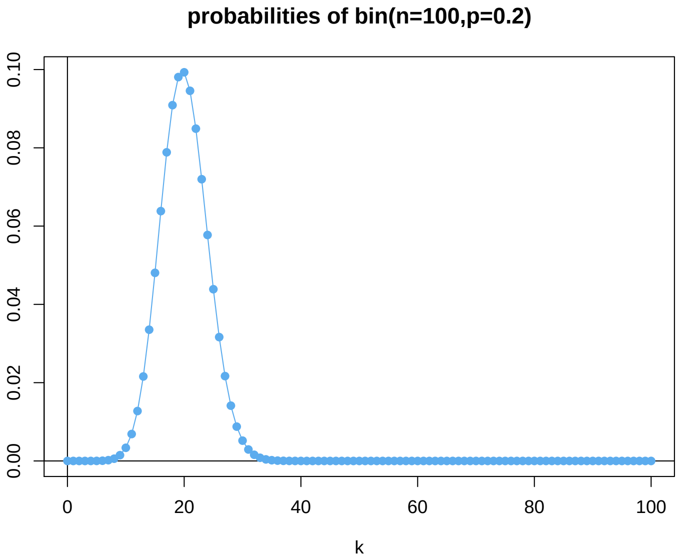
Let X be the number of successes in n independent trials with success probability of p each. Then,

$$\Pr(X = k) = \binom{n}{k} p^k \cdot (1 - p)^{n-k}$$

holds for all $k \in \{0, 1, \ldots, n\}$ and X is said to be *binomially distributed*, for short:

$$X \sim \text{bin}(n, p).$$


probabilities of bin(n=10,p=0.2)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
2
4
6
8
10
k


probabilities of bin(n=100,p=0.2)
0.00
0.02
0.04
0.06
0.08
0.10
0
20
40
60
80
100
k

With the binomial distribution we can treat our initial question

With the binomial distribution we can treat our initial question

Assume in a small populaiton of $n = 100$ individuals the neutral allele A has a frequency of 0.3.

With the binomial distribution we can treat our initial question

Assume in a small populaiton of $n = 100$ individuals the neutral allele A has a frequency of 0.3.

How probable is it that X, the frequency of A in the next generarion is 0.32?

With the binomial distribution we can treat our initial question

Assume in a small populaiton of $n = 100$ individuals the neutral allele A has a frequency of 0.3.

How probable is it that X, the frequency of A in the next generarion is 0.32?

$$\Pr(X = 0.32) = ?$$

With the binomial distribution we can treat our initial question

Assume in a small populaiton of $n = 100$ individuals the neutral allele A has a frequency of 0.3.

How probable is it that X, the frequency of A in the next generarion is 0.32?

$$\Pr(X = 0.32) = ?$$

We can only answer this on the basis of a probabilistic model, and the answer will depend on how we model the population.

Modeling approach

We make a few simplifying assumptions:

- Discrete generations
- The population is haploid, that is, each individual has exactly one parent in the generation before.
- constant population size $n = 100$

Modeling approach

We make a few simplifying assumptions:

- Discrete generations
- The population is haploid, that is, each individual has exactly one parent in the generation before.
- constant population size $n = 100$

$\Pr(X = 0.32)$ still depends on whether few individuals have many offspring or whether all individuals have similar offspring numbers.

Modeling approach

We make a few simplifying assumptions:

- Discrete generations
- The population is haploid, that is, each individual has exactly one parent in the generation before.
- constant population size $n = 100$

$\Pr(X = 0.32)$ still depends on whether few individuals have many offspring or whether all individuals have similar offspring numbers. $\Pr(X = 0.32)$ is only defined with additional assumptions, e.g.:

Modeling approach

We make a few simplifying assumptions:

- Discrete generations
- The population is haploid, that is, each individual has exactly one parent in the generation before.
- constant population size $n = 100$

$\Pr(X = 0.32)$ still depends on whether few individuals have many offspring or whether all individuals have similar offspring numbers. $\Pr(X = 0.32)$ is only defined with additional assumptions, e.g.:

- Each individual chooses its parent purely randomly in the generation before.

Modeling approach

We make a few simplifying assumptions:

- Discrete generations
- The population is haploid, that is, each individual has exactly one parent in the generation before.
- constant population size $n = 100$

$\Pr(X = 0.32)$ still depends on whether few individuals have many offspring or whether all individuals have similar offspring numbers. $\Pr(X = 0.32)$ is only defined with additional assumptions, e.g.:

- Each individual chooses its parent purely randomly in the generation before.

“purely randomly” means *independent of all others* and *all potential parents with the same probability*.

Our assumptions imply that each individuals of the next generations have a probability of 0.3 to obtain allele A, and they get their alleles independently of each other.

Our assumptions imply that each individuals of the next generations have a probability of 0.3 to obtain allele A, and they get their alleles independently of each other.

Therefore, the number K of individuels who get allele A is binomially distributed with $n = 100$ and $p = 0.3$:

$$K \sim \text{bin}(n = 100, p = 0.3)$$

Our assumptions imply that each individuals of the next generations have a probability of 0.3 to obtain allele A, and they get their alleles independently of each other.

Therefore, the number K of individuels who get allele A is binomially distributed with $n = 100$ and $p = 0.3$:

$$K \sim \text{bin}(n = 100, p = 0.3)$$

For $X = K/n$ follows:

$$\Pr(X = 0.32) =$$

Our assumptions imply that each individuals of the next generations have a probability of 0.3 to obtain allele A, and they get their alleles independently of each other.

Therefore, the number K of individuels who get allele A is binomially distributed with $n = 100$ and $p = 0.3$:

$$K \sim \text{bin}(n = 100, p = 0.3)$$

For $X = K/n$ follows:

$$\Pr(X = 0.32) = \Pr(K = 32) =$$

Our assumptions imply that each individuals of the next generations have a probability of 0.3 to obtain allele A, and they get their alleles independently of each other.

Therefore, the number K of individuels who get allele A is binomially distributed with $n = 100$ and $p = 0.3$:

$$K \sim \text{bin}(n = 100, p = 0.3)$$

For $X = K/n$ follows:

$$\Pr(X = 0.32) = \Pr(K = 32) = \binom{n}{32} \cdot p^{32} \cdot (1 - p)^{100-32}$$

Our assumptions imply that each individuals of the next generations have a probability of 0.3 to obtain allele A, and they get their alleles independently of each other.

Therefore, the number K of individuels who get allele A is binomially distributed with $n = 100$ and $p = 0.3$:

$$K \sim \text{bin}(n = 100, p = 0.3)$$

For $X = K/n$ follows:

$$\Pr(X = 0.32) = \Pr(K = 32) = \binom{n}{32} \cdot p^{32} \cdot (1 - p)^{100-32}$$

$$= \binom{100}{32} \cdot 0.3^{32} \cdot 0.7^{68}$$

Our assumptions imply that each individuals of the next generations have a probability of 0.3 to obtain allele A, and they get their alleles independently of each other.

Therefore, the number K of individuels who get allele A is binomially distributed with $n = 100$ and $p = 0.3$:

$$K \sim \text{bin}(n = 100, p = 0.3)$$

For $X = K/n$ follows:

$$\Pr(X = 0.32) = \Pr(K = 32) = \binom{n}{32} \cdot p^{32} \cdot (1 - p)^{100-32}$$

$$= \binom{100}{32} \cdot 0.3^{32} \cdot 0.7^{68} \approx 0.078$$

Contents

- We want to argue that some deviation in the data is not just random.

- We want to argue that some deviation in the data is not just random.
- To this end we first specify a null hypothesis H_0, i.e. we define, what “just random” means.

- We want to argue that some deviation in the data is not just random.
- To this end we first specify a null hypothesis H_0, i.e. we define, what “just random” means.
- Then we try to show: If H_0 is true, then a deviation that is at least at large as the observed one, is very improbable.

- We want to argue that some deviation in the data is not just random.
- To this end we first specify a null hypothesis H_0, i.e. we define, what “just random” means.
- Then we try to show: If H_0 is true, then a deviation that is at least at large as the observed one, is very improbable.
- If we can do this, we reject H_0.

- We want to argue that some deviation in the data is not just random.
- To this end we first specify a null hypothesis H_0, i.e. we define, what “just random” means.
- Then we try to show: If H_0 is true, then a deviation that is at least at large as the observed one, is very improbable.
- If we can do this, we reject H_0.
- How we measure deviation, must be clear *before* we see the data.

Statistical Testing: Imporatant terms

null hypothesis H_0 : says that what we want to substantiate is not true and anything that looks like evidence in the data is just random. We try to reject H_0.

significance level α : If H_0 is true, the probability to falsly reject it, must be $\leq \alpha$ (often $\alpha = 0.05$).

test statistic : measures how far the data deviates from what H_0 predicts into the direction of our alternative hypothesis.

p value : Probability that, if H_0 is true, a dataset leads to a test statistic value that is as least as extreme as the observed one.

- We reject the null hypothesis H_0 if the p value is smaller than α.

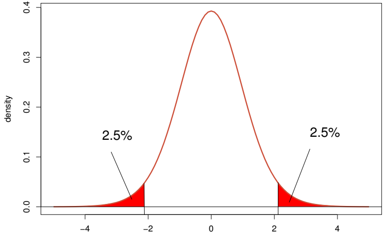
- We reject the null hypothesis H_0 if the p value is smaller than α.
- Thus, if H_0 is true, the probability to (falsely) reject it is α (not the p value).

- We reject the null hypothesis H_0 if the p value is smaller than α.
- Thus, if H_0 is true, the probability to (falsely) reject it is α (not the p value).
- This entails that a researcher who performs many tests with $\alpha = 0.05$ on complete random data (i.e. where H_0 is always true), will falsely reject H_0 in 5% of the tests.

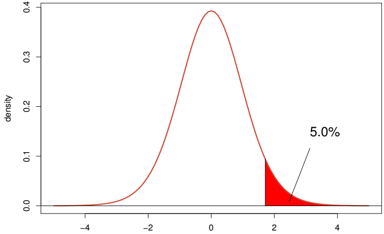
- We reject the null hypothesis H_0 if the p value is smaller than α.
- Thus, if H_0 is true, the probability to (falsely) reject it is α (not the p value).
- This entails that a researcher who performs many tests with $\alpha = 0.05$ on complete random data (i.e. where H_0 is always true), will falsely reject H_0 in 5% of the tests.
- Therefore it is a severe violation of academic soundness to perform tests until one shows significance, and to publish only the latter.

Testing two-sided or one-sided?

We observe a value of x that is much larger than the H_0 expectation value μ.



p-value=$\Pr_{H_0}(|X - \mu| \geq |x - \mu|)$



p-value=$\Pr_{H_0}(X \geq x)$

The pure teachings of statistical testing

- Specify a null hypothesis H_0, e.g. $\mu = 0$.

The pure teachings of statistical testing

- Specify a null hypothesis H_0, e.g. $\mu = 0$.
- Specify level of significance α, e.g. $\alpha = 0.05$.

The pure teachings of statistical testing

- Specify a null hypothesis H_0, e.g. $\mu = 0$.
- Specify level of significance α, e.g. $\alpha = 0.05$.
- Specify an event $\mathcal{A}$ such that

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

(or at least $\Pr_{H_0}(\mathcal{A}) \leq \alpha$).

The pure teachings of statistical testing

- Specify a null hypothesis H_0, e.g. $\mu = 0$.
- Specify level of significance α, e.g. $\alpha = 0.05$.
- Specify an event $\mathcal{A}$ such that

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

(or at least $\Pr_{H_0}(\mathcal{A}) \leq \alpha$).
e.g. $\mathcal{A} = \{\overline{X} > q\}$ or $\mathcal{A} = \{|\overline{X} - \mu| > r\}$

The pure teachings of statistical testing

- Specify a null hypothesis H_0, e.g. $\mu = 0$.
- Specify level of significance α, e.g. $\alpha = 0.05$.
- Specify an event $\mathcal{A}$ such that

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

(or at least $\Pr_{H_0}(\mathcal{A}) \leq \alpha$).
e.g. $\mathcal{A} = \{\overline{X} > q\}$ or $\mathcal{A} = \{|\overline{X} - \mu| > r\}$
in general: $\mathcal{A} = \{p\text{-value} \leq \alpha\}$

The pure teachings of statistical testing

- Specify a null hypothesis H_0, e.g. $\mu = 0$.
- Specify level of significance α, e.g. $\alpha = 0.05$.
- Specify an event $\mathcal{A}$ such that

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

  (or at least $\Pr_{H_0}(\mathcal{A}) \leq \alpha$).
  e.g. $\mathcal{A} = \{\overline{X} > q\}$ or $\mathcal{A} = \{|\overline{X} - \mu| > r\}$
  in general: $\mathcal{A} = \{p\text{-value} \leq \alpha\}$
- AND AFTER THAT: Look at the data and check if if $\mathcal{A}$ occurs.

The pure teachings of statistical testing

- Specify a null hypothesis H_0, e.g. $\mu = 0$.
- Specify level of significance α, e.g. $\alpha = 0.05$.
- Specify an event $\mathcal{A}$ such that

$$\Pr_{H_0}(\mathcal{A}) = \alpha$$

  (or at least $\Pr_{H_0}(\mathcal{A}) \leq \alpha$).
  e.g. $\mathcal{A} = \{\overline{X} > q\}$ or $\mathcal{A} = \{|\overline{X} - \mu| > r\}$
  in general: $\mathcal{A} = \{p\text{-value} \leq \alpha\}$
- AND AFTER THAT: Look at the data and check if if $\mathcal{A}$ occurs.
- Then, the probability that H_0 is rejected in the case that H_0 is actually true ("Type I error") is just α.

Violations against the pure teachings

"The two-sided test gave me a p-value of 0.06. Therefore, I tested one-sided and this worked out nicely."

Violations against the pure teachings

"The two-sided test gave me a p-value of 0.06. Therefore, I tested one-sided and this worked out nicely."

is as bad as:

Violations against the pure teachings

"The two-sided test gave me a *p*-value of 0.06. Therefore, I tested one-sided and this worked out nicely."

is as bad as:

"At first glance I saw that $\overline{x}$ is larger than μ_{H_0}. So, I immediately applied the one-sided test."

Important

The decision between one-sided and two-sided must not depend on the concrete data that are used in the test.

Important

The decision between one-sided and two-sided must not depend on the concrete data that are used in the test.
More generally: If $\mathcal{A}$ is the event that will lead to the rejection of H_0, (if it occurs) then $\mathcal{A}$ must be defined without being influenced by the data that is used for testing.

This means: Use separate data sets for exploratory data analysis and for testing.

This means: Use separate data sets for exploratory data analysis and for testing.
In some fields these rules are followed quite strictly, e.g. testing new pharmaceuticals for accreditation.

This means: Use separate data sets for exploratory data analysis and for testing.
In some fields these rules are followed quite strictly, e.g. testing new pharmaceuticals for accreditation.
In some other fields the practical approach is more common: Just inform the reader about the p-values of different null-hypotheses. Let the reader decide which null-hypothesis would have been the most natural one.

If H_0 is rejected on the 5%-level, which of the following statements is true?

- The null hypothesis is wrong.
-
-

If H_0 is rejected on the 5%-level, which of the following statements is true?

- ~~The null hypothesis is wrong.~~
-
-

If H_0 is rejected on the 5%-level, which of the following statements is true?

- ~~The null hypothesis is wrong.~~
- H_0 is wrong with a probability of 95%.
-

If H_0 is rejected on the 5%-level, which of the following statements is true?

- ~~The null hypothesis is wrong.~~
- ~~H_0 is wrong with a probability of 95%.~~
-

If H_0 is rejected on the 5%-level, which of the following statements is true?

- ~~The null hypothesis is wrong.~~
- ~~H_0 is wrong with a probability of 95%.~~
- If H_0 is true, you will see such an extreme event only in 5% of the data sets.

If H_0 is rejected on the 5%-level, which of the following statements is true?

- ~~The null hypothesis is wrong.~~
- ~~H_0 is wrong with a probability of 95%.~~
- If H_0 is true, you will see such an extreme event only in 5% of the data sets. ✓

If the test did not reject H_0, which of the following statements are true?

- We have to reject the alternative H_1.
-
-
-
-
-

If the test did not reject H_0, which of the following statements are true?

- ~~We have to reject the alternative H_1.~~
-
-
-
-
-

If the test did not reject H_0, which of the following statements are true?

- ~~We have to reject the alternative H_1.~~
- H_0 is true.
-
-
-
-

If the test did not reject H_0, which of the following statements are true?

- ~~We have to reject the alternative H_1.~~
- ~~H_0 is true~~
-
-
-
-

If the test did not reject H_0, which of the following statements are true?

- ~~We have to reject the alternative H_1.~~
- ~~H_0 is true~~
- H_0 is probably true.
-
-
-

If the test did not reject H_0, which of the following statements are true?

- ~~We have to reject the alternative H_1.~~
- ~~H_0 is true~~
- ~~H_0 is probably true.~~
-
-
-

If the test did not reject H_0, which of the following statements are true?

- ~~We have to reject the alternative H_1.~~
- ~~H_0 is true~~
- ~~H_0 is probably true.~~
- It is safe to assume that H_0 was true.
-
-

If the test did not reject H_0, which of the following statements are true?

- ~~We have to reject the alternative H_1.~~
- ~~H_0 is true~~
- ~~H_0 is probably true.~~
- ~~It is safe to assume that H_0 was true.~~
-
-

If the test did not reject H_0, which of the following statements are true?

- ~~We have to reject the alternative H_1.~~
- ~~H_0 is true~~
- ~~H_0 is probably true.~~
- ~~It is safe to assume that H_0 was true.~~
- Even if H_0 is true, it is not so unlikely that our test statistic takes a value that is as extreme as the one we observed.
-

If the test did not reject H_0, which of the following statements are true?

- ~~We have to reject the alternative H_1.~~
- ~~H_0 is true~~
- ~~H_0 is probably true.~~
- ~~It is safe to assume that H_0 was true.~~
- Even if H_0 is true, it is not so unlikely that our test statistic takes a value that is as extreme as the one we observed. ✓
-

If the test did not reject H_0, which of the following statements are true?

- ~~We have to reject the alternative H_1.~~
- ~~H_0 is true~~
- ~~H_0 is probably true.~~
- ~~It is safe to assume that H_0 was true.~~
- Even if H_0 is true, it is not so unlikely that our test statistic takes a value that is as extreme as the one we observed. ✓
- With this respect, H_0 is compatible with the data.

If the test did not reject H_0, which of the following statements are true?

- ~~We have to reject the alternative H_1.~~
- ~~H_0 is true~~
- ~~H_0 is probably true.~~
- ~~It is safe to assume that H_0 was true.~~
- Even if H_0 is true, it is not so unlikely that our test statistic takes a value that is as extreme as the one we observed. ✓
- With this respect, H_0 is compatible with the data. ✓

Contents

Definition (Expectation value)

Let X be a random variable with finite or countable state space $\mathcal{S} = \{x_1, x_2, x_3 \ldots\} \subseteq \mathbb{R}$.

Definition (Expectation value)

Let X be a random variable with finite or countable state space $\mathcal{S} = \{x_1, x_2, x_3 \dots\} \subseteq \mathbb{R}$. The *expectation value* of X is defined by

$$\mathbb{E}X = \sum_{x \in \mathcal{S}} x \cdot \Pr(X = x)$$

Definition (Expectation value)

Let X be a random variable with finite or countable state space $\mathcal{S} = \{x_1, x_2, x_3 \dots\} \subseteq \mathbb{R}$. The *expectation value* of X is defined by

$$\mathbb{E}X = \sum_{x \in \mathcal{S}} x \cdot \Pr(X = x)$$

It is also common to write μ_X instead of $\mathbb{E}X$.

If we replace probabilities by relative frequencies in this definition, we get the formula for the mean value (of a sample).

Definition (Expectation value)

If X is a random variable with finite or countable state space $\mathcal{S} = \{x_1, x_2, x_3 \ldots\} \subseteq \mathbb{R}$, the *expectation value* of X is defined by

$$\mathbb{E}X = \sum_{x \in \mathcal{S}} x \cdot \Pr(X = x)$$

Examples:

- Let X be Bernoulli distributed with success probability $p \in [0, 1]$. Then we get

$$\mathbb{E}X = 1 \cdot \Pr(X = 1) + 0 \cdot \Pr(X = 0) = \Pr(X = 1) = p$$

Definition (Expectation value)

If X is a random variable with finite or countable state space $\mathcal{S} = \{x_1, x_2, x_3 \dots\} \subseteq \mathbb{R}$, the *expectation value* of X is defined by

$$\mathbb{E}X = \sum_{x \in \mathcal{S}} x \cdot \Pr(X = x)$$

Examples:

- Let X be Bernoulli distributed with success probability $p \in [0, 1]$. Then we get

$$\mathbb{E}X = 1 \cdot \Pr(X = 1) + 0 \cdot \Pr(X = 0) = \Pr(X = 1) = p$$

- Let W be the result of rolling a dice. Then we get

$$\begin{aligned} \mathbb{E}W &= 1 \cdot \Pr(W = 1) + 2 \cdot \Pr(W = 2) + \dots + 6 \cdot \Pr(W = 6) \\ &= 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + \dots + 6 \cdot \tfrac{1}{6} = 21\tfrac{1}{6} = 3.5 \end{aligned}$$

Calculating with expectatins

Theorem (Linearity of Expectation)

If X and Y are random variables with values in $\mathbb{R}$ and if $a \in \mathbb{R}$, we get:

- $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}X$
- $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$

Calculating with expectatins

Theorem (Linearity of Expectation)

If X and Y are random variables with values in $\mathbb{R}$ and if $a \in \mathbb{R}$, we get:

- $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}X$
- $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$

Theorem (Only if independent!)

*If X and Y are **stochastically independent** random variables with values in $\mathbb{R}$, we get*

- $\mathbb{E}(X \cdot Y) = \mathbb{E}X \cdot \mathbb{E}Y.$

Calculating with expectatins

Theorem (Linearity of Expectation)

If X and Y are random variables with values in $\mathbb{R}$ and if $a \in \mathbb{R}$, we get:

- $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}X$
- $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$

Theorem (Only if independent!)

*If X and Y are **stochastically independent** random variables with values in $\mathbb{R}$, we get*

- $\mathbb{E}(X \cdot Y) = \mathbb{E}X \cdot \mathbb{E}Y.$

But in general $\mathbb{E}(X \cdot Y) \neq \mathbb{E}X \cdot \mathbb{E}Y$. Example:

$$\mathbb{E}(W \cdot W) = \frac{91}{6} = 15.167 > 12.25 = 3.5 \cdot 3.5 = \mathbb{E}W \cdot \mathbb{E}W$$

Proof of Linearity: If $\mathcal{S}$ is the state space of X and Y, and if $a, b \in \mathbb{R}$, we obtain:

$$\begin{aligned}
&\mathbb{E}(a \cdot X + b \cdot Y) \\
&= \sum_{x\in\mathcal{S}} \sum_{y\in\mathcal{S}} (a \cdot x + b \cdot y) \Pr(X = x, Y = y) \\
&= a \cdot \sum_{x\in\mathcal{S}} \sum_{y\in\mathcal{S}} x \Pr(X = x, Y = y) + b \cdot \sum_{x\in\mathcal{S}} \sum_{y\in\mathcal{S}} y \Pr(X = x, Y = y) \\
&= a \cdot \sum_{x\in\mathcal{S}} x \sum_{y\in\mathcal{S}} \Pr(X = x, Y = y) + b \cdot \sum_{y\in\mathcal{S}} y \sum_{x\in\mathcal{S}} \Pr(X = x, Y = y) \\
&= a \cdot \sum_{x\in\mathcal{S}} x \Pr(X = x) + b \cdot \sum_{y\in\mathcal{S}} y \Pr(Y = y) \\
&= a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y)
\end{aligned}$$

Proof of the product formula: Let $\mathcal{S}$ be the state space of X and Y, and let X and Y be (stochastically) independent.

$$\begin{aligned}
&\mathbb{E}(X \cdot Y) \\
&= \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} (x \cdot y) \Pr(X = x, Y = y) \\
&= \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} (x \cdot y) \Pr(X = x) \Pr(Y = y) \\
&= \sum_{x \in \mathcal{S}} x \Pr(X = x) \cdot \sum_{y \in \mathcal{S}} y \Pr(Y = y) \\
&= \mathbb{E}X \cdot \mathbb{E}Y.
\end{aligned}$$

Theorem

If X is random variable with finite state space $\mathcal{S} \subset \mathbb{R}$, and if $f : \mathbb{R} \to \mathbb{R}$ is a function, we obtain

$$\mathbb{E}(f(X)) = \sum_{x \in \mathcal{S}} f(x) \cdot \Pr(X = x)$$

Exercise: proof this.

Expectation of the binomial distribution

Let $Y_1, Y_2, \ldots, Y_n$ be the indicator variables of the n independent trials, that is

$$Y_i = \begin{cases} 1 & \text{if trial } i \text{ succeeds} \\ 0 & \text{if trial } i - \text{ fails} \end{cases}$$

Expectation of the binomial distribution

Let $Y_1, Y_2, \ldots, Y_n$ be the indicator variables of the n independent trials, that is

$$Y_i = \begin{cases} 1 & \text{if trial } i \text{ succeeds} \\ 0 & \text{if trial } i - \text{ fails} \end{cases}$$

Then $X = Y_1 + \cdots + Y_n$ is binomially distributed with parameters (n, p), where p is the success probability of the trials.

Expectation of the binomial distribution

Let $Y_1, Y_2, \ldots, Y_n$ be the indicator variables of the n independent trials, that is

$$Y_i = \begin{cases} 1 & \text{if trial } i \text{ succeeds} \\ 0 & \text{if trial } i - \text{ fails} \end{cases}$$

Then $X = Y_1 + \cdots + Y_n$ is binomially distributed with parameters (n, p), where p is the success probability of the trials.

Linearity of expectation implies

$$\begin{aligned} \mathbb{E}X &= \mathbb{E}(Y_1 + \cdots + Y_n) \\ &= \mathbb{E}Y_1 + \cdots + \mathbb{E}Y_n \end{aligned}$$

Expectation of the binomial distribution

Let $Y_1, Y_2, \ldots, Y_n$ be the indicator variables of the n independent trials, that is

$$Y_i = \begin{cases} 1 & \text{if trial } i \text{ succeeds} \\ 0 & \text{if trial } i - \text{ fails} \end{cases}$$

Then $X = Y_1 + \cdots + Y_n$ is binomially distributed with parameters (n, p), where p is the success probability of the trials.

Linearity of expectation implies

$$\begin{aligned} \mathbb{E}X &= \mathbb{E}(Y_1 + \cdots + Y_n) \\ &= \mathbb{E}Y_1 + \cdots + \mathbb{E}Y_n \\ &= p + \cdots + p \; = \; np \end{aligned}$$

Expectation of the binomial distribution

Let $Y_1, Y_2, \ldots, Y_n$ be the indicator variables of the n independent trials, that is

$$Y_i = \begin{cases} 1 & \text{if trial } i \text{ succeeds} \\ 0 & \text{if trial } i - \text{ fails} \end{cases}$$

Then $X = Y_1 + \cdots + Y_n$ is binomially distributed with parameters (n, p), where p is the success probability of the trials.

Linearity of expectation implies

$$\begin{aligned} \mathbb{E}X &= \mathbb{E}(Y_1 + \cdots + Y_n) \\ &= \mathbb{E}Y_1 + \cdots + \mathbb{E}Y_n \\ &= p + \cdots + p = np \end{aligned}$$

Note:

$$X \sim \text{bin}(n, p) \Rightarrow \mathbb{E}X = n \cdot p$$

Contents

Definition (Variance, Covariance and Correlation)

The *Variance* of a $\mathbb{R}$-valued random variable X is

$$\mathrm{Var}X = \sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}X)^2\right].$$

Definition (Variance, Covariance and Correlation)

The *Variance* of a $\mathbb{R}$-valued random variable X is

$$\text{Var}X = \sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}X)^2\right].$$

$\sigma_X = \sqrt{\text{Var } X}$ is the *Standard Deviation*.

Definition (Variance, Covariance and Correlation)

The *Variance* of a $\mathbb{R}$-valued random variable X is

$$\mathrm{Var} X = \sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}X)^2\right].$$

$\sigma_X = \sqrt{\mathrm{Var}\ X}$ is the *Standard Deviation*.
If Y is enother $\mathbb{R}$-valued random variable,

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}X) \cdot (Y - \mathbb{E}Y)\right]$$

is the *Covariance* of X and Y.

Definition (Variance, Covariance and Correlation)

The *Variance* of a $\mathbb{R}$-valued random variable X is

$$\mathrm{Var}X = \sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}X)^2\right].$$

$\sigma_X = \sqrt{\mathrm{Var}\ X}$ is the *Standard Deviation*.
If Y is enother $\mathbb{R}$-valued random variable,

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}X) \cdot (Y - \mathbb{E}Y)\right]$$

is the *Covariance* of X and Y.
The *Correlation* of X and Y is

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}.$$

The Variance

$$\text{Var}X = \mathbb{E}\left[(X - \mathbb{E}X)^2\right]$$

is the average squared deviation from the expectation.

The Variance

$$\mathrm{Var}X = \mathbb{E}\left[(X - \mathbb{E}X)^2\right]$$

is the average squared deviation from the expectation.

The Correlation

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

is always between in the range from -1 to 1. The random variables X and Y are

- positively correlated, if X and Y tend to be both above average or both below average.
- negatively correlated, if X and Y tend to deviate from their expectation values in opposite ways.

The Variance

$$\mathrm{Var}X = \mathbb{E}\left[(X - \mathbb{E}X)^2\right]$$

is the average squared deviation from the expectation.

The Correlation

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

is always between in the range from -1 to 1. The random variables X and Y are

- positively correlated, if X and Y tend to be both above average or both below average.
- negatively correlated, if X and Y tend to deviate from their expectation values in opposite ways.

If X and Y are independent, they are also uncorrelated, that is $\mathrm{Cor}(X, Y) = 0$.

Example: rolling dice

Variance of result from rolling a dice W:

$$\begin{aligned}
\text{Var}(W) &= \mathbb{E}\big[\big(W - \mathbb{E}W\big)^2\big] \\
&= \mathbb{E}\big[\big(W - 3.5\big)^2\big] \\
&= (1 - 3.5)^2 \cdot \frac{1}{6} + (2 - 3.5)^2 \cdot \frac{1}{6} + \ldots + (6 - 3.5)^2 \cdot \frac{1}{6} \\
&= \frac{17.5}{6} \\
&= 2.91667
\end{aligned}$$

Example: Empirical Distribution

If $x_1, \ldots, x_n \in \mathbb{R}$ are data and if X is the result of a random draw from the data (such that $\Pr(X = x_i) = \frac{1}{n}$), we get:

$$\mathbb{E}X = \sum_{i=1}^{n} x_i \Pr(X = x_i) = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$$

and

$$\operatorname{Var} X = \mathbb{E}\big[\big(X - \mathbb{E}X\big)^2\big] = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

Example: Empirical Distribution

If $x_1, \dots, x_n \in \mathbb{R}$ are data and if X is the result of a random draw from the data (such that $\Pr(X = x_i) = \frac{1}{n}$), we get:

$$\mathbb{E}X = \sum_{i=1}^{n} x_i \Pr(X = x_i) = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}$$

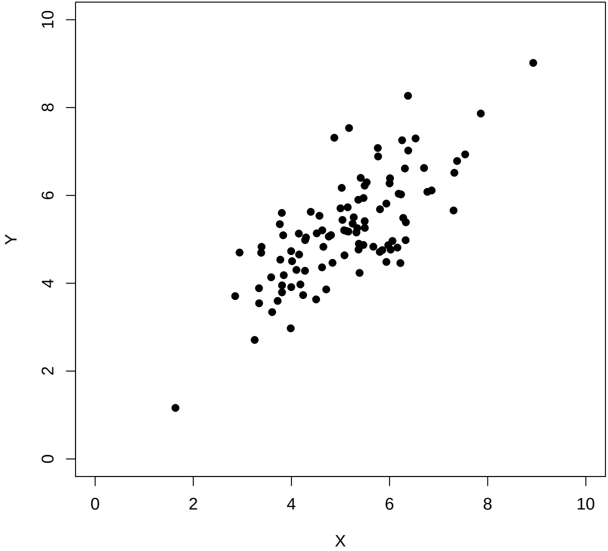
and

$$\text{Var } X = \mathbb{E}\left[\left(X - \mathbb{E}X\right)^2\right] = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

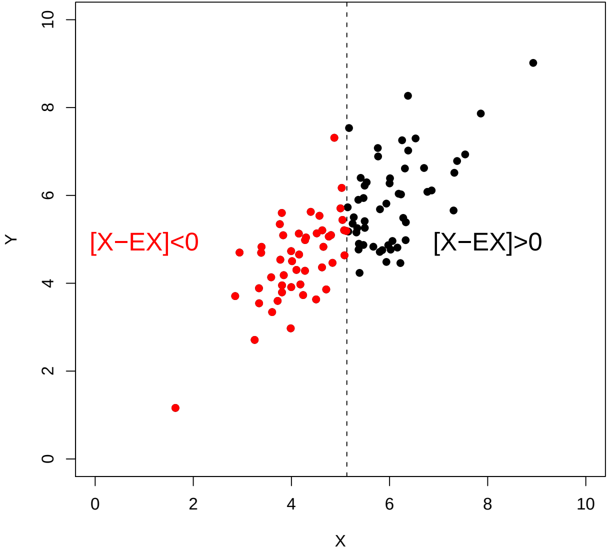
If $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R} \times \mathbb{R}$ are data if (X, Y) are drawn from the data such that $\Pr((X, Y) = (x_i, y_i)) = \frac{1}{n}$, we get

$$\text{Cov }(X, Y) = \mathbb{E}\left[\left(X - \mathbb{E}X\right)\left(Y - \mathbb{E}Y\right)\right] = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

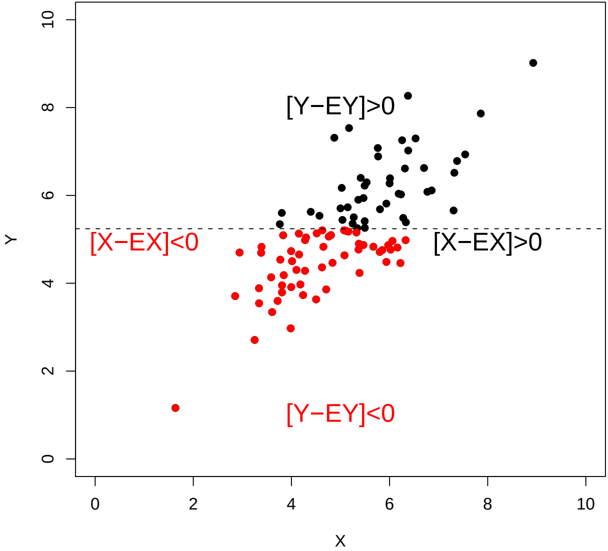
Why $\text{Cov}(X, Y) = \mathbb{E}([X - \mathbb{E}X][Y - \mathbb{E}Y])$?



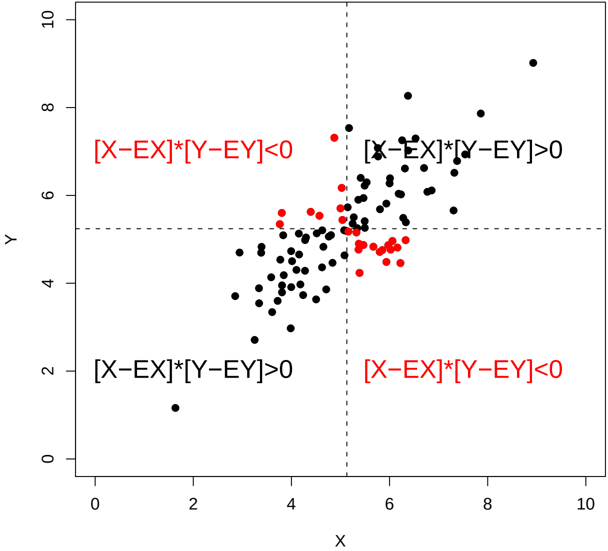
Why $\text{Cov}(X, Y) = \mathbb{E}([X - \mathbb{E}X][Y - \mathbb{E}Y])$?



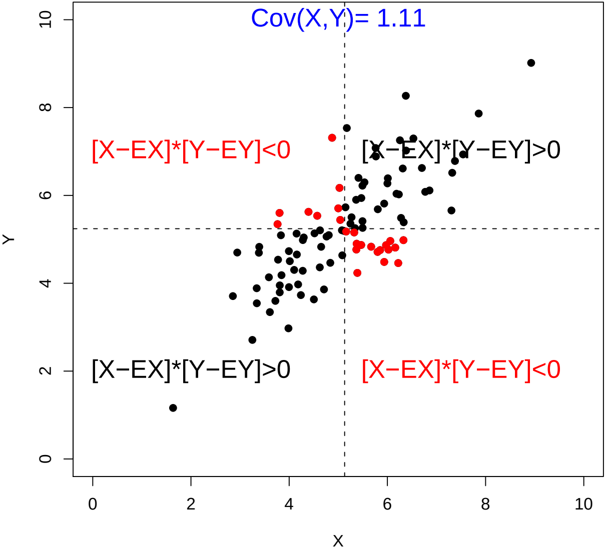
Why $\mathrm{Cov}(X, Y) = \mathbb{E}([X - \mathbb{E}X][Y - \mathbb{E}Y])$?



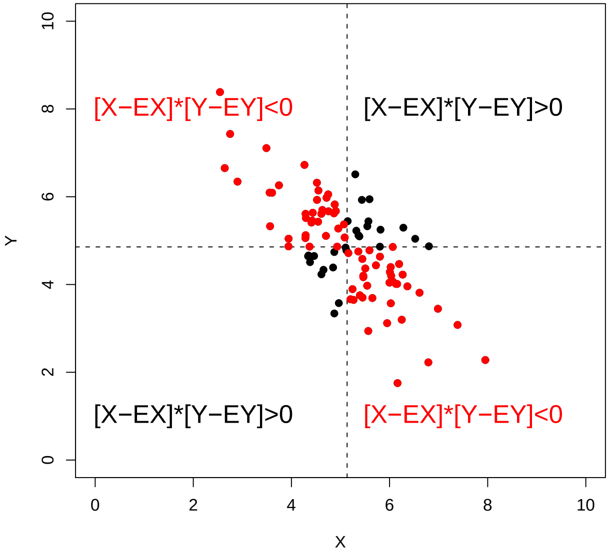
Why $\mathrm{Cov}(X, Y) = \mathbb{E}([X - \mathbb{E}X][Y - \mathbb{E}Y])$?



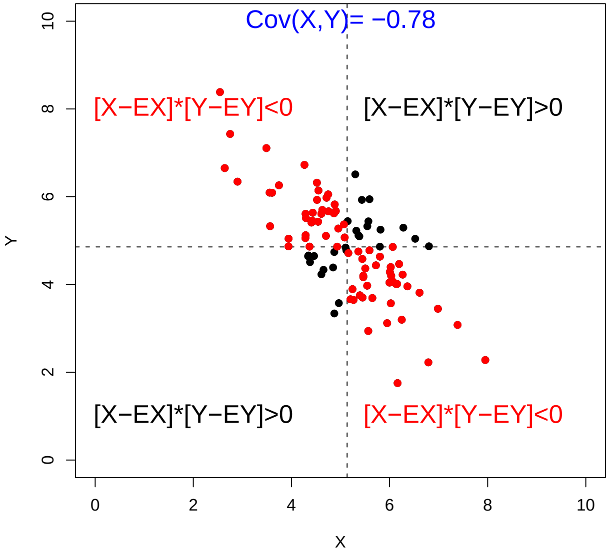
Why Cov$(X, Y) = \mathbb{E}([X - \mathbb{E}X][Y - \mathbb{E}Y])$?



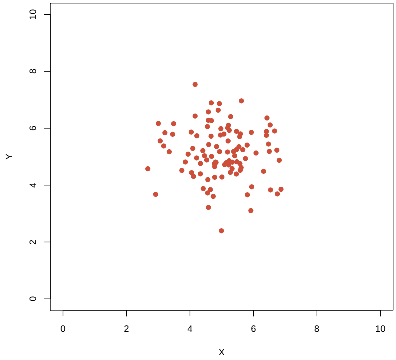
Why $\mathrm{Cov}(X, Y) = \mathbb{E}([X - \mathbb{E}X][Y - \mathbb{E}Y])$?



Why $\text{Cov}(X, Y) = \mathbb{E}([X - \mathbb{E}X][Y - \mathbb{E}Y])$?

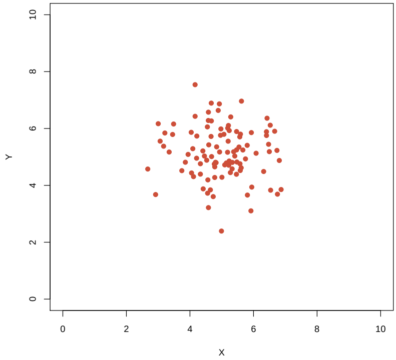


$\sigma_X = 0.95$, $\sigma_Y = 0.92$



$\sigma_X = 0.95$, $\sigma_Y = 0.92$

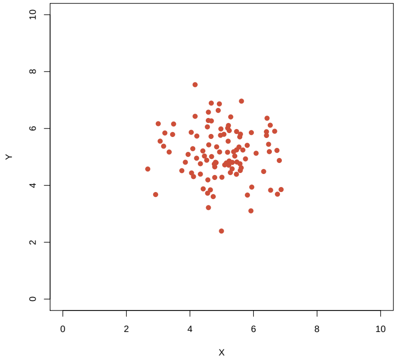
$\text{Cov}(X, Y) = -0.06$



$\sigma_X = 0.95$, $\sigma_Y = 0.92$

$\text{Cov}(X, Y) = -0.06$

$\text{Cor}(X, Y) = -0.069$

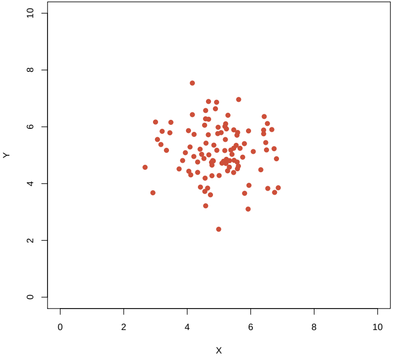


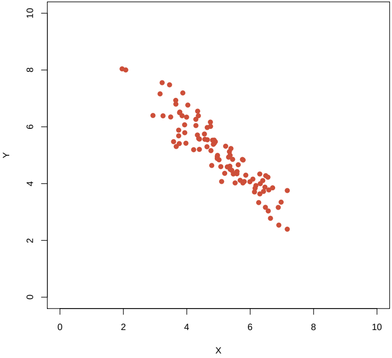
$\sigma_X = 0.95$, $\sigma_Y = 0.92$

$\text{Cov}(X, Y) = -0.06$

$\text{Cor}(X, Y) = -0.069$

$\sigma_X = 1.13$, $\sigma_Y = 1.2$





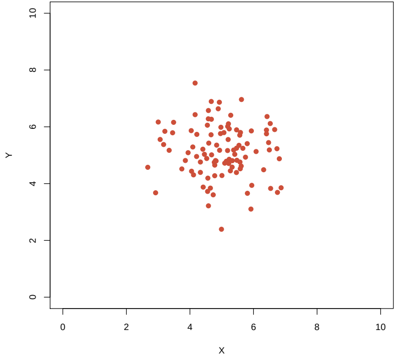
$\sigma_X = 0.95$, $\sigma_Y = 0.92$

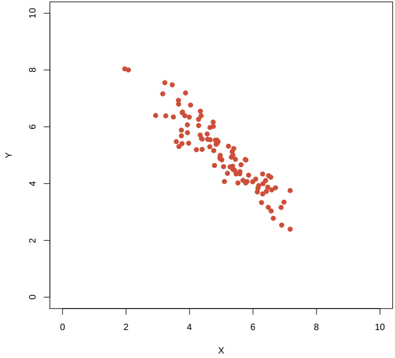
$\text{Cov}(X, Y) = -0.06$

$\text{Cor}(X, Y) = -0.069$

$\sigma_X = 1.13$, $\sigma_Y = 1.2$

$\text{Cov}(X, Y) = -1.26$





$\sigma_X = 0.95$, $\sigma_Y = 0.92$

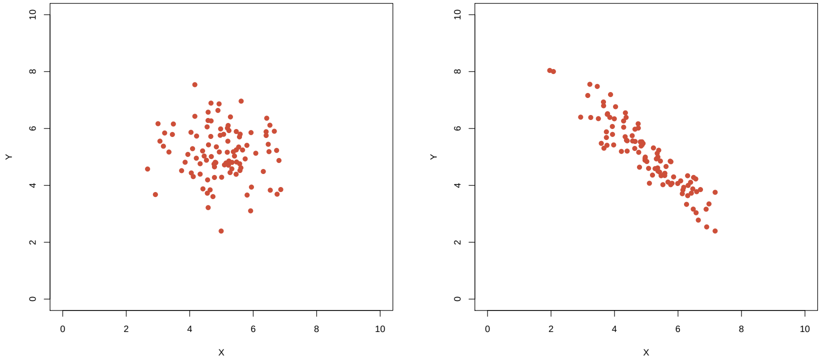
$\text{Cov}(X, Y) = -0.06$

$\text{Cor}(X, Y) = -0.069$

$\sigma_X = 1.13$, $\sigma_Y = 1.2$

$\text{Cov}(X, Y) = -1.26$

$\text{Cor}(X, Y) = -0.92$

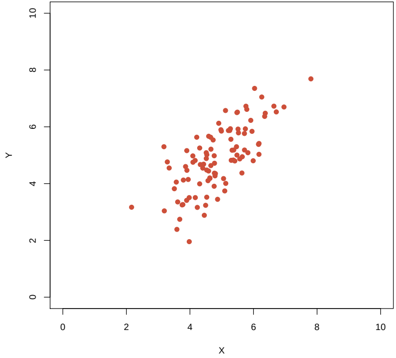


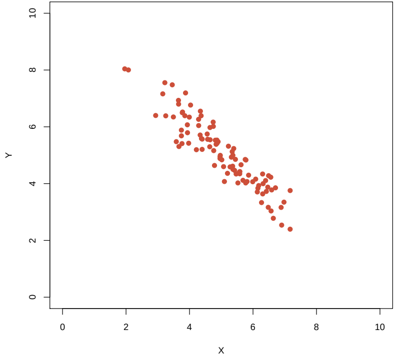
$\sigma_X = 1.14$, $\sigma_Y = 0.78$

$\sigma_X = 1.13$, $\sigma_Y = 1.2$

$\text{Cov}(X, Y) = -1.26$

$\text{Cor}(X, Y) = -0.92$





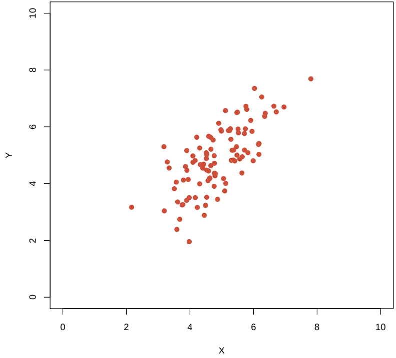
$\sigma_X = 1.14$, $\sigma_Y = 0.78$

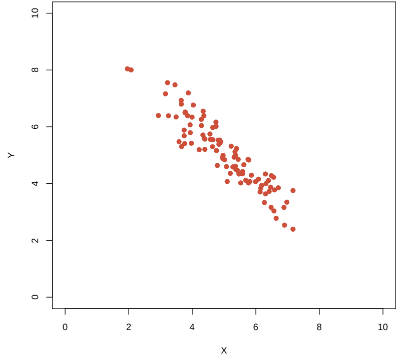
$\text{Cov}(X, Y) = 0.78$

$\sigma_X = 1.13$, $\sigma_Y = 1.2$

$\text{Cov}(X, Y) = -1.26$

$\text{Cor}(X, Y) = -0.92$





$\sigma_X = 1.14, \sigma_Y = 0.78$

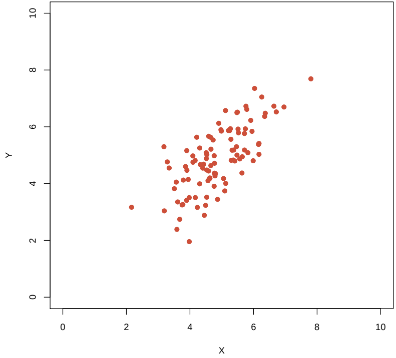
$\text{Cov}(X, Y) = 0.78$

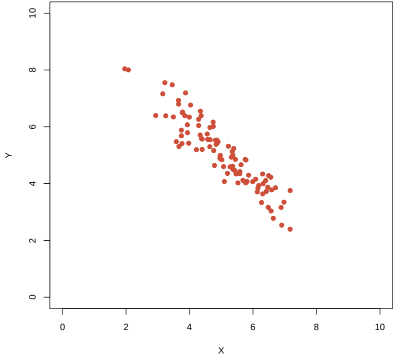
$\text{Cor}(X, Y) = 0.71$

$\sigma_X = 1.13, \sigma_Y = 1.2$

$\text{Cov}(X, Y) = -1.26$

$\text{Cor}(X, Y) = -0.92$



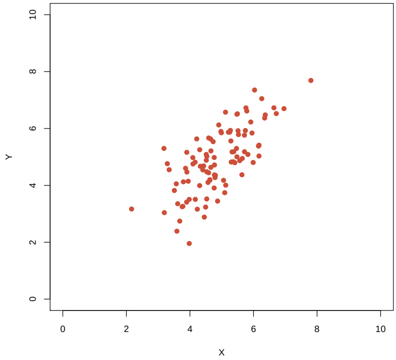


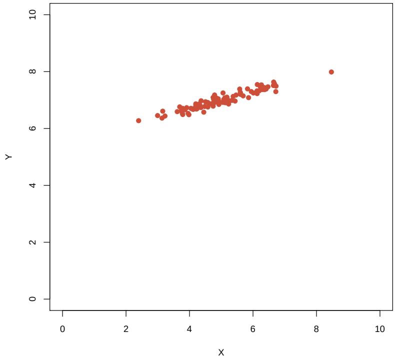
$\sigma_X = 1.14$, $\sigma_Y = 0.78$

$\text{Cov}(X, Y) = 0.78$

$\text{Cor}(X, Y) = 0.71$

$\sigma_X = 1.03$, $\sigma_Y = 0.32$

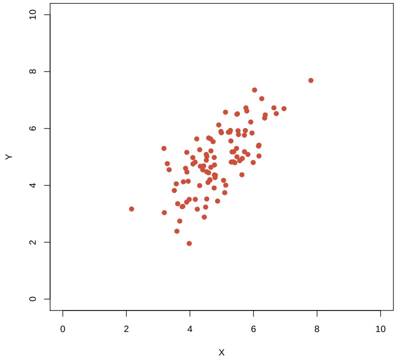




$\sigma_X = 1.14$, $\sigma_Y = 0.78$

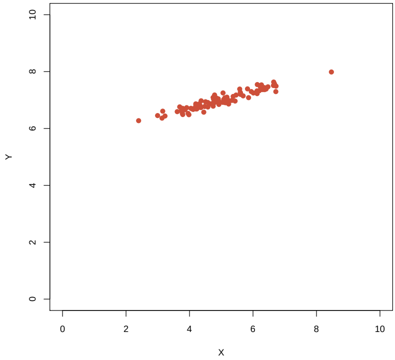
$\text{Cov}(X, Y) = 0.78$

$\text{Cor}(X, Y) = 0.71$



$\sigma_X = 1.03$, $\sigma_Y = 0.32$

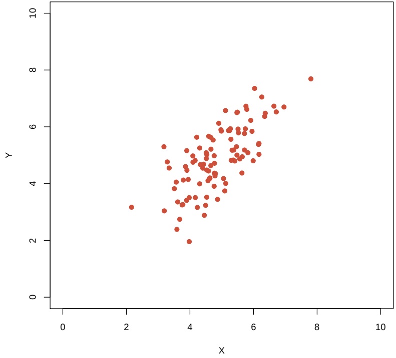
$\text{Cov}(X, Y) = 0.32$



$\sigma_X = 1.14$, $\sigma_Y = 0.78$

$\text{Cov}(X, Y) = 0.78$

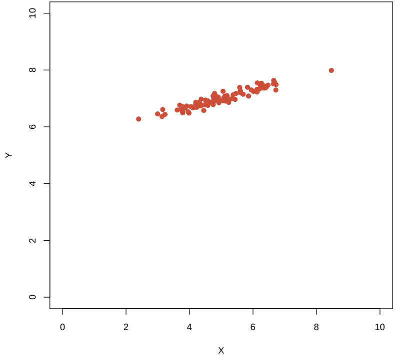
$\text{Cor}(X, Y) = 0.71$



$\sigma_X = 1.03$, $\sigma_Y = 0.32$

$\text{Cov}(X, Y) = 0.32$

$\text{Cor}(X, Y) = 0.95$

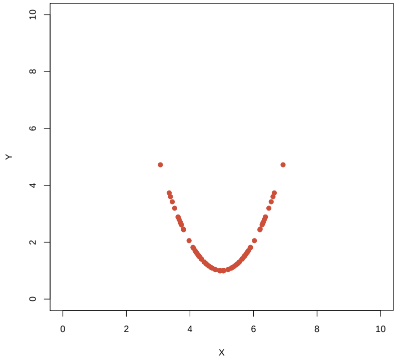


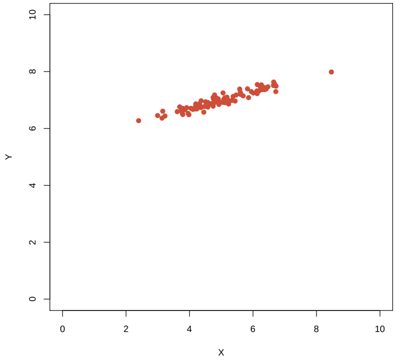
$\sigma_X = 0.91$, $\sigma_Y = 0.88$

$\sigma_X = 1.03$, $\sigma_Y = 0.32$

$\text{Cov}(X, Y) = 0.32$

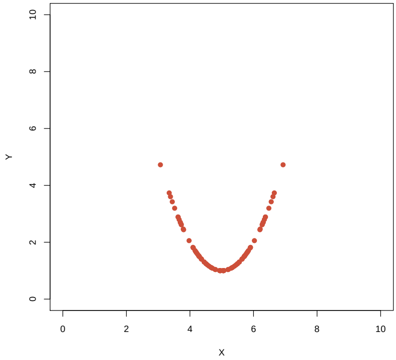
$\text{Cor}(X, Y) = 0.95$





$\sigma_X = 0.91$, $\sigma_Y = 0.88$

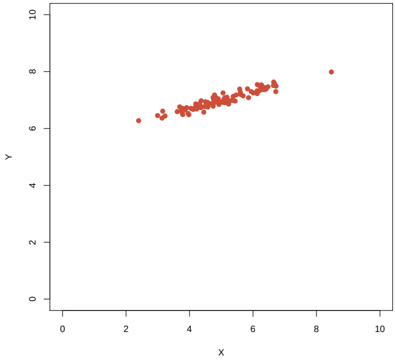
$\text{Cov}(X, Y) = 0$



$\sigma_X = 1.03$, $\sigma_Y = 0.32$

$\text{Cov}(X, Y) = 0.32$

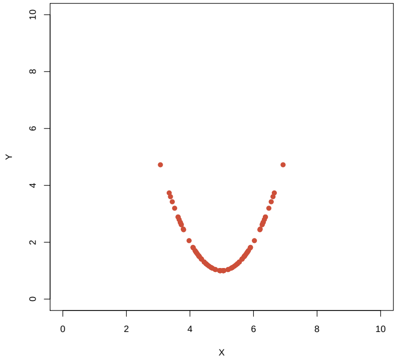
$\text{Cor}(X, Y) = 0.95$



$\sigma_X = 0.91$, $\sigma_Y = 0.88$

$\text{Cov}(X, Y) = 0$

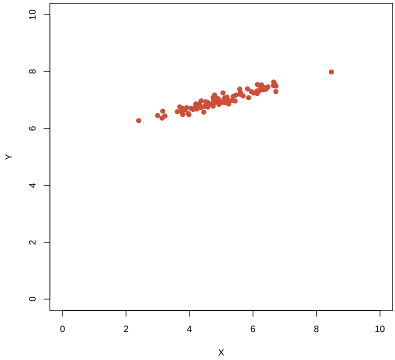
$\text{Cor}(X, Y) = 0$



$\sigma_X = 1.03$, $\sigma_Y = 0.32$

$\text{Cov}(X, Y) = 0.32$

$\text{Cor}(X, Y) = 0.95$



Calculation rules for variances

$$\text{Var}X = \mathbb{E}[(X - \mathbb{E}X)^2]$$

- $\text{Var}X = \text{Cov}(X, X)$

Calculation rules for variances

$\mathrm{Var}X = \mathbb{E}[(X - \mathbb{E}X)^2]$

- $\mathrm{Var}X = \mathrm{Cov}(X, X)$
- $\mathrm{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ (Exercise!)

Calculation rules for variances

$$\mathrm{Var}X = \mathbb{E}[(X - \mathbb{E}X)^2]$$

- $\mathrm{Var}X = \mathrm{Cov}(X, X)$
- $\mathrm{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ (Exercise!)
- $\mathrm{Var}(a \cdot X) = a^2 \cdot \mathrm{Var}X$

Calculation rules for variances

$$\mathrm{Var}X = \mathbb{E}[(X - \mathbb{E}X)^2]$$

- $\mathrm{Var}X = \mathrm{Cov}(X, X)$
- $\mathrm{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ (Exercise!)
- $\mathrm{Var}(a \cdot X) = a^2 \cdot \mathrm{Var}X$
- $\mathrm{Var}(X + Y) = \mathrm{Var}X + \mathrm{Var}Y + 2 \cdot \mathrm{Cov}(X, Y)$

Calculation rules for variances

$$\mathrm{Var}X = \mathbb{E}[(X - \mathbb{E}X)^2]$$

- $\mathrm{Var}X = \mathrm{Cov}(X, X)$
- $\mathrm{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ (Exercise!)
- $\mathrm{Var}(a \cdot X) = a^2 \cdot \mathrm{Var}X$
- $\mathrm{Var}(X + Y) = \mathrm{Var}X + \mathrm{Var}Y + 2 \cdot \mathrm{Cov}(X, Y)$
- $\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2 \cdot \sum_{j=1}^n \sum_{i=1}^{j-1} \mathrm{Cov}(X_i, X_j)$

Calculation rules for variances

$\text{Var}X = \mathbb{E}[(X - \mathbb{E}X)^2]$

- $\text{Var}X = \text{Cov}(X, X)$
- $\text{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ (Exercise!)
- $\text{Var}(a \cdot X) = a^2 \cdot \text{Var}X$
- $\text{Var}(X + Y) = \text{Var}X + \text{Var}Y + 2 \cdot \text{Cov}(X, Y)$
- $\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \cdot \sum_{j=1}^n \sum_{i=1}^{j-1} \text{Cov}(X_i, X_j)$
- If (X, Y) stochastically independent we get:

$$\text{Var}(X + Y) = \text{Var}X + \text{Var}Y$$

Calculation rules for Covariances

$$\text{Cov}(X, Y) = \mathbb{E}[(X - EX) \cdot (Y - \mathbb{E}Y)]$$

- If X and Y are independent, then $\text{Cov}(X, Y) = 0$ (but not the other way around!)

Calculation rules for Covariances

$\text{Cov}(X, Y) = \mathbb{E}[(X - EX) \cdot (Y - \mathbb{E}Y)]$

- If X and Y are independent, then $\text{Cov}(X, Y) = 0$ (but not the other way around!)
- Cov(X, Y)=Cov(Y, X)

Calculation rules for Covariances

$$\text{Cov}(X, Y) = \mathbb{E}[(X - EX) \cdot (Y - \mathbb{E}Y)]$$

- If X and Y are independent, then $\text{Cov}(X, Y) = 0$ (but not the other way around!)
- $\text{Cov}(X, Y)$=$\text{Cov}(Y, X)$
- $\text{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y$ (Exercise!)

Calculation rules for Covariances

$\text{Cov}(X, Y) = \mathbb{E}[(X - EX) \cdot (Y - \mathbb{E}Y)]$

- If X and Y are independent, then $\text{Cov}(X, Y) = 0$ (but not the other way around!)
- $\text{Cov}(X, Y)$=$\text{Cov}(Y, X)$
- $\text{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y$ (Exercise!)
- $\text{Cov}(a \cdot X, Y) = a \cdot \text{Cov}(X, Y) = \text{Cov}(X, a \cdot Y)$

Calculation rules for Covariances

$$\text{Cov}(X, Y) = \mathbb{E}[(X - EX) \cdot (Y - \mathbb{E}Y)]$$

- If X and Y are independent, then $\text{Cov}(X, Y) = 0$ (but not the other way around!)
- $\text{Cov}(X, Y)$=$\text{Cov}(Y, X)$
- $\text{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y$ (Exercise!)
- $\text{Cov}(a \cdot X, Y) = a \cdot \text{Cov}(X, Y) = \text{Cov}(X, a \cdot Y)$
- $\text{Cov}(X + Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y)$

Calculation rules for Covariances

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - EX) \cdot (Y - \mathbb{E}Y)]$$

- If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$ (but not the other way around!)
- $\mathrm{Cov}(X,Y)$=$\mathrm{Cov}(Y,X)$
- $\mathrm{Cov}(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y$ (Exercise!)
- $\mathrm{Cov}(a \cdot X, Y) = a \cdot \mathrm{Cov}(X,Y) = \mathrm{Cov}(X, a \cdot Y)$
- $\mathrm{Cov}(X + Z, Y) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(Z,Y)$
- $\mathrm{Cov}(X, Z + Y) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(X,Y)$

Calculation rules for Covariances

$$\text{Cov}(X, Y) = \mathbb{E}[(X - EX) \cdot (Y - \mathbb{E}Y)]$$

- If X and Y are independent, then $\text{Cov}(X, Y) = 0$ (but not the other way around!)
- $\text{Cov}(X, Y)$=$\text{Cov}(Y, X)$
- $\text{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y$ (Exercise!)
- $\text{Cov}(a \cdot X, Y) = a \cdot \text{Cov}(X, Y) = \text{Cov}(X, a \cdot Y)$
- $\text{Cov}(X + Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y)$
- $\text{Cov}(X, Z + Y) = \text{Cov}(X, Z) + \text{Cov}(X, Y)$

The last three rules describe the bilinearity of covariance.

Calculation rules for Correlations

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

- $-1 \leq \text{Cor}(X, Y) \leq 1$
- $\text{Cor}(X, Y) = \text{Cor}(Y, X)$
- $\text{Cor}(X, Y) = \text{Cov}(X/\sigma_X, Y/\sigma_Y)$

Calculation rules for Correlations

$\text{Cor}(X, Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$

- $-1 \leq \text{Cor}(X, Y) \leq 1$
- $\text{Cor}(X, Y) = \text{Cor}(Y, X)$
- $\text{Cor}(X, Y) = \text{Cov}(X/\sigma_X, Y/\sigma_Y)$
- $\text{Cor}(X, Y) = 1$ if and only if Y is an increasing, affine-linear function of X, that is, if $Y = a \cdot X + b$ for appropriate $a > 0$ and $b \in \mathbb{R}$.

Calculation rules for Correlations

$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$

- $-1 \leq \mathrm{Cor}(X, Y) \leq 1$
- $\mathrm{Cor}(X, Y) = \mathrm{Cor}(Y, X)$
- $\mathrm{Cor}(X, Y) = \mathrm{Cov}(X/\sigma_X, Y/\sigma_Y)$
- $\mathrm{Cor}(X, Y) = 1$ if and only if Y is an increasing, affine-linear function of X, that is, if $Y = a \cdot X + b$ for appropriate $a > 0$ and $b \in \mathbb{R}$.
- $\mathrm{Cor}(X, Y) = -1$ if and only if Y is an decreasing, affine-linear function of X, that is, if $Y = a \cdot X + b$ for appropriate $a < 0$ and $b \in \mathbb{R}$.

Bernoulli distribution

A Bernoulli distributed random variable Y with success probability $p \in [0, 1]$ has expectation value

$$\mathbb{E}Y = p$$

and variance

$$\text{Var } Y = p \cdot (1 - p)$$

Bernoulli distribution

A Bernoulli distributed random variable Y with success probability $p \in [0,1]$ has expectation value

$$\mathbb{E}Y = p$$

and variance

$$\text{Var } Y = p \cdot (1-p)$$

Proof: From $\Pr(Y = 1) = p$ and $\Pr(Y = 0) = (1-p)$ follows

$$\mathbb{E}Y = 1 \cdot p + 0 \cdot (1-p) = p.$$

Bernoulli distribution

A Bernoulli distributed random variable Y with success probability $p \in [0, 1]$ has expectation value

$$\mathbb{E}Y = p$$

and variance

$$\text{Var } Y = p \cdot (1 - p)$$

Proof: From $\Pr(Y = 1) = p$ and $\Pr(Y = 0) = (1 - p)$ follows

$$\mathbb{E}Y = 1 \cdot p + 0 \cdot (1 - p) = p.$$

variance:

$$\begin{aligned} \text{Var } Y &= \mathbb{E}\big(Y^2\big) - \big(\mathbb{E}Y\big)^2 \\ &= 1^2 \cdot p + 0^2 \cdot (1 - p) - p^2 = p \cdot (1 - p) \end{aligned}$$

Binomial distribution

Let $Y_1, \cdots, Y_n$ be independent Bernoulli distributed with success probability p. Then follows

$$\sum_{i=1}^{n} Y_i =: X \sim \text{bin}(n, p)$$

and we get:

$$\text{Var}\, X =$$

Binomial distribution

Let $Y_1, \cdots, Y_n$ be independent Bernoulli distributed with success probability p. Then follows

$$\sum_{i=1}^{n} Y_i =: X \sim \text{bin}(n, p)$$

and we get:

$$\text{Var}\, X = \text{Var}\left(\sum_{i=1}^{n} Y_i\right) =$$

Binomial distribution

Let $Y_1, \cdots, Y_n$ be independent Bernoulli distributed with success probability p. Then follows

$$\sum_{i=1}^{n} Y_i =: X \sim \mathrm{bin}(n, p)$$

and we get:

$$\mathrm{Var}\ X = \mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \mathrm{Var}\ Y_i =$$

Binomial distribution

Let $Y_1, \cdots, Y_n$ be independent Bernoulli distributed with success probability p. Then follows

$$\sum_{i=1}^{n} Y_i =: X \sim \text{bin}(n, p)$$

and we get:

$$\text{Var } X = \text{Var}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \text{Var } Y_i = n \cdot p \cdot (1 - p)$$

Binomial distribution

Theorem (Expectation value and variance of the binomial distribution)

If X is binomially distributed with parameters (n, p), we get:

$$\mathbb{E}X = n \cdot p$$

und

$$Var\,X = n \cdot p \cdot (1 - p)$$

Example: Genetic Drift

In a haploid population of n individuals, let p be the frequency of some allele A. We assume that (due to some simplifying assumptions?) the absolute frequency K of A in the next generation is (n, p)-binomially distributed.
For $X = K/n$, the relative frequency in the next generation follows:

$$\mathrm{Var}(X) = \mathrm{Var}(K/n) = \mathrm{Var}(K)/n^2$$

Example: Genetic Drift

In a haploid population of n individuals, let p be the frequency of some allele A. We assume that (due to some simplifying assumptions?) the absolute frequency K of A in the next generation is (n, p)-binomially distributed.
For $X = K/n$, the relative frequency in the next generation follows:

$$\mathrm{Var}(X) = \mathrm{Var}(K/n) = \mathrm{Var}(K)/n^2 = n \cdot p \cdot (1 - p)/n^2$$

Example: Genetic Drift

In a haploid population of n individuals, let p be the frequency of some allele A. We assume that (due to some simplifying assumptions?) the absolute frequency K of A in the next generation is (n, p)-binomially distributed.
For $X = K/n$, the relative frequency in the next generation follows:

$$\mathrm{Var}(X) = \mathrm{Var}(K/n) = \mathrm{Var}(K)/n^2 = n \cdot p \cdot (1-p)/n^2$$

$$= \frac{p \cdot (1-p)}{n}$$

Example: Genetic Drift

If we consider the change of allele frequencies over m generations, the variances add up. If m is a small number, such that p will not change much over m generations, the is variance of change of allele frequencies is approximately

$$m \cdot \text{Var}(X) = \frac{m \cdot p \cdot (1-p)}{n}$$

(because the changes per generation are independent of each other) and thus, the standard deviation is about

$$\sqrt{\frac{m}{n} \cdot p \cdot (1-p)}$$