Вероятности и распределения

Основы биостатистики, осень 2022

Марина Варфоломеева

- Вероятность
- Распределения вероятностей
- Действия с вероятностями
- Деревья вероятностей
- Зависимые события
- Условная вероятность
- Нормальное распределение

Вероятность

STATISTICS

means never having to say you're certain

Вероятность

Представьте, что ваш плейлист состоит из 1000 песен и вы нажимаете кнопку `shuffle`.

- Какова вероятность, что единственная самая любимая вами песня будет первой?
- Какова вероятность, что первой будет не самая любимая песня?

Вероятность

Представьте, что ваш плейлист состоит из 1000 песен и вы нажимаете кнопку `shuffle`.

- Какова вероятность, что единственная самая любимая вами песня будет первой?
- Какова вероятность, что первой будет не самая любимая песня?

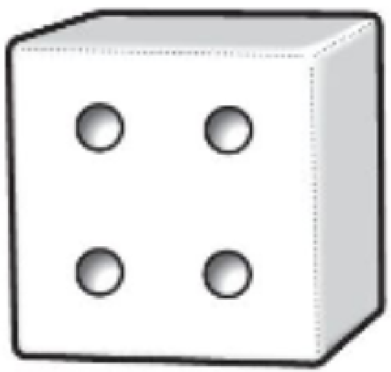
Представьте, что вы в лифте в 15 этажном доме, но нажали на случайную кнопку.

- Какова вероятность, что вы приедете на нужный вам этаж с первой попытки?
- Какова вероятность того, что вы попадете на неправильный этаж?

Случайное испытание

В результате случайного **испытания** происходит или не происходит случайное **событие**.

Случайному событию могут благоприятствовать один или несколько **элементарных исходов**.



Испытание:
бросок шестигранного кубика

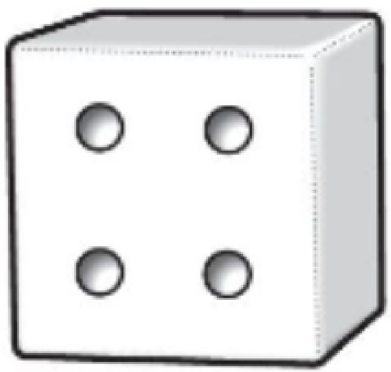
Элементарные исходы:
может выпасть 1, 2, 3, 4, 5, 6

Событие:
на кубике выпало 4

Случайное испытание

В результате случайного **испытания** происходит или не происходит случайное **событие**.

Случайному событию могут благоприятствовать один или несколько **элементарных исходов**.



Испытание:
бросок шестигранного кубика

Элементарные исходы:
может выпасть 1, 2, 3, 4, 5, 6

Событие:
на кубике выпало 4

Испытание:
нажимаем shuffle

Элементарные исходы:
первым может оказаться любой из треков плейлиста

Событие:
первый трек — любимая песня

Испытание:
нажимаем кнопку лифта 15-этажного дома

Элементарные исходы:
лифт может приехать на любой из 15 этажей

Событие:
попали на нужный этаж

Вероятность

Вероятность события — это доля случаев, когда происходит это событие в ряду испытаний.

$$0 \leq P \leq 1$$

Вероятность

Вероятность события — это доля случаев, когда происходит это событие в ряду испытаний.

$$0 \leq P \leq 1$$

$P(A)$ — вероятность того, что произошло некоторое событие A:

Вероятность

Вероятность события — это доля случаев, когда происходит это событие в ряду испытаний.

$$0 \leq P \leq 1$$

$P(A)$ — вероятность того, что произошло некоторое событие A:

- $P(\text{на кубике выпало 4})$

Вероятность

Вероятность события — это доля случаев, когда происходит это событие в ряду испытаний.

$$0 \leq P \leq 1$$

$P(A)$ — вероятность того, что произошло некоторое событие A:

- $P(\text{на кубике выпало 4})$
- $P(\text{на кубике выпало четное число})$

Вероятность

Вероятность события — это доля случаев, когда происходит это событие в ряду испытаний.

$$0 \leq P \leq 1$$

$P(A)$ — вероятность того, что произошло некоторое событие A:

- $P(\text{на кубике выпало 4})$
- $P(\text{на кубике выпало четное число})$
- $P(\text{в семье из 3 детей все девочки})$

Вероятность

Вероятность события — это доля случаев, когда происходит это событие в ряду испытаний.

$$0 \leq P \leq 1$$

$P(A)$ — вероятность того, что произошло некоторое событие A:

- $P(\text{на кубике выпало 4})$
- $P(\text{на кубике выпало четное число})$
- $P(\text{в семье из 3 детей все девочки})$
- $P(\text{в последовательности из 10 нуклеотидов только G})$

Вероятность

Вероятность события — это доля случаев, когда происходит это событие в ряду испытаний.

$$0 \leq P \leq 1$$

$P(A)$ — вероятность того, что произошло некоторое событие A:

- $P(\text{на кубике выпало 4})$
- $P(\text{на кубике выпало четное число})$
- $P(\text{в семье из 3 детей все девочки})$
- $P(\text{в последовательности из 10 нуклеотидов только G})$
- в случайной выборке людей оцениваем долю рыжеволосых

Вероятность

Вероятность события — это доля случаев, когда происходит это событие в ряду испытаний.

$$0 \leq P \leq 1$$

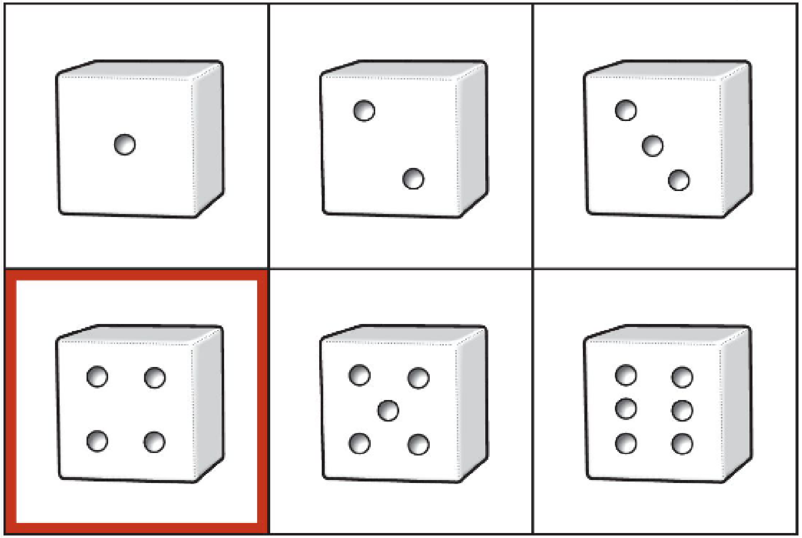
$P(A)$ — вероятность того, что произошло некоторое событие A:

- $P(\text{на кубике выпало 4})$
- $P(\text{на кубике выпало четное число})$
- $P(\text{в семье из 3 детей все девочки})$
- $P(\text{в последовательности из 10 нуклеотидов только G})$
- в случайной выборке людей оцениваем долю рыжеволосых
- в случайной выборке новорожденных считаем долю детей с синдромом Дауна

Диаграмма Венна

Диаграмма Венна (= диаграмма Эйлера-Венна) — схематическое изображение множества (= пространства) всех элементарных исходов и его подмножеств, соответствующих определенным событиям.

- На кубике выпало 4

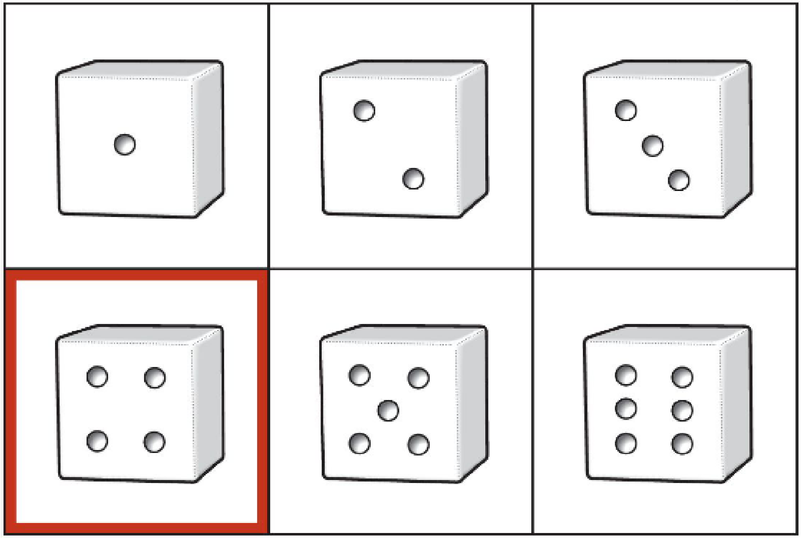


Whitlock, Schluter, 2015, fig.5.2-1-2

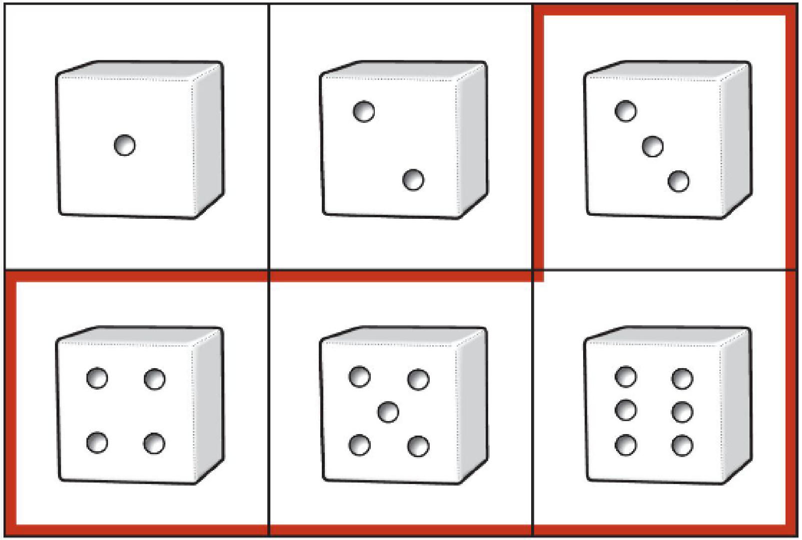
Диаграмма Венна

Диаграмма Венна (= диаграмма Эйлера-Венна) — схематическое изображение множества (= пространства) всех элементарных исходов и его подмножеств, соответствующих определенным событиям.

- На кубике выпало 4



- На кубике выпало 3 или больше

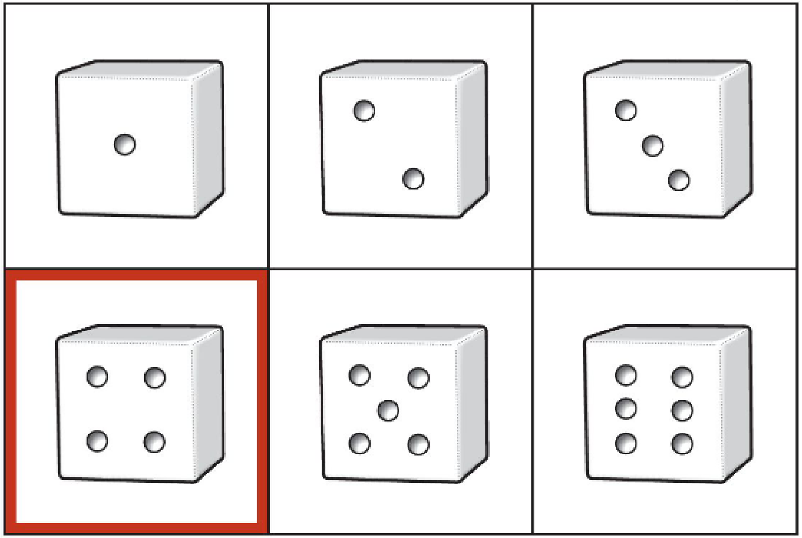


Whitlock, Schluter, 2015, fig.5.2-1-2

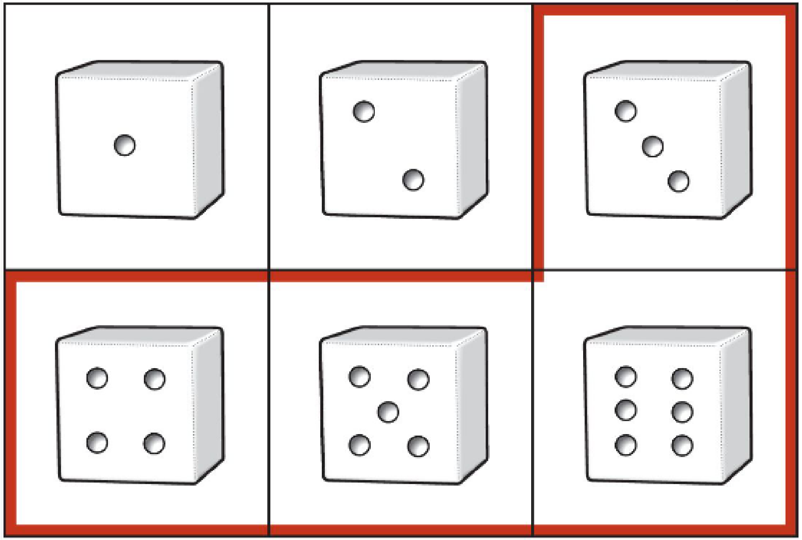
Диаграмма Венна

Диаграмма Венна (= диаграмма Эйлера-Венна) — схематическое изображение множества (= пространства) всех элементарных исходов и его подмножеств, соответствующих определенным событиям.

- На кубике выпало 4



- На кубике выпало 3 или больше



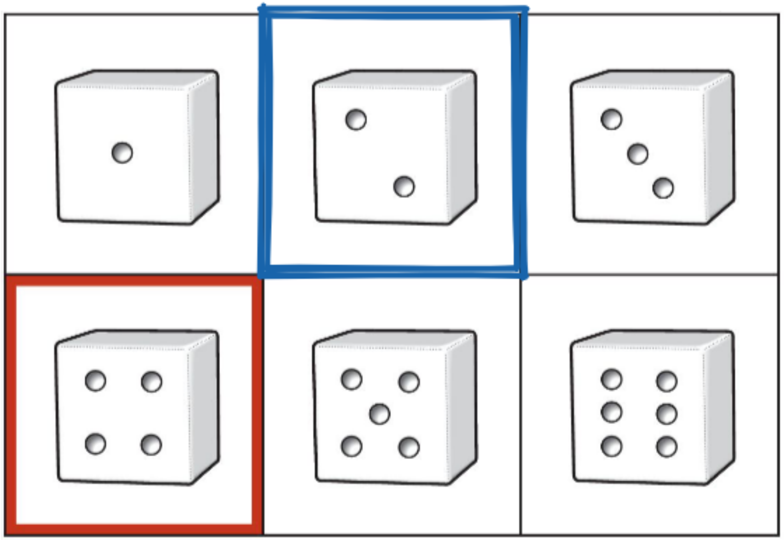
Whitlock, Schluter, 2015, fig.5.2-1-2

Вероятность события легко оценить, посчитав долю элементарных исходов, благоприятствующих этому событию.

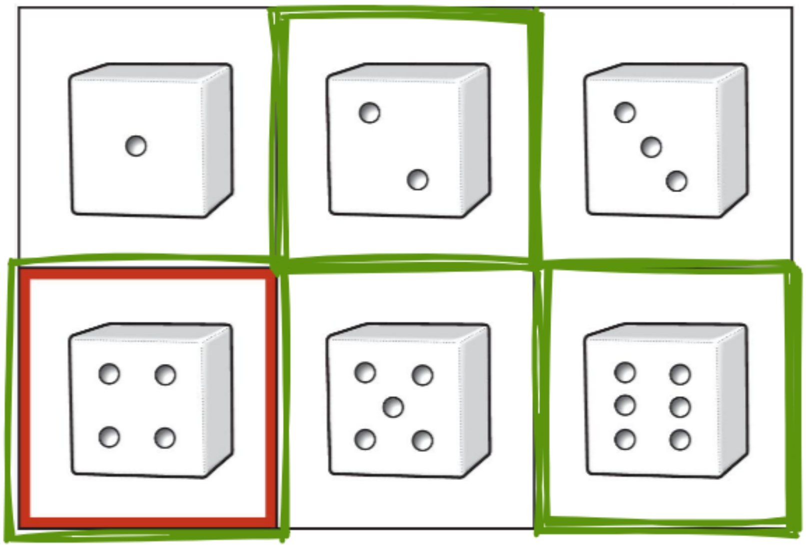
Совместные и несовместные события

События A и B называются **несовместными**, если $P(\text{A и B}) = 0$

- Выпало 4 и выпало 2 — **несовместные** события (mutually exclusive events).



- Выпало 4 и выпало четное число — **совместные** события.



Распределения вероятностей

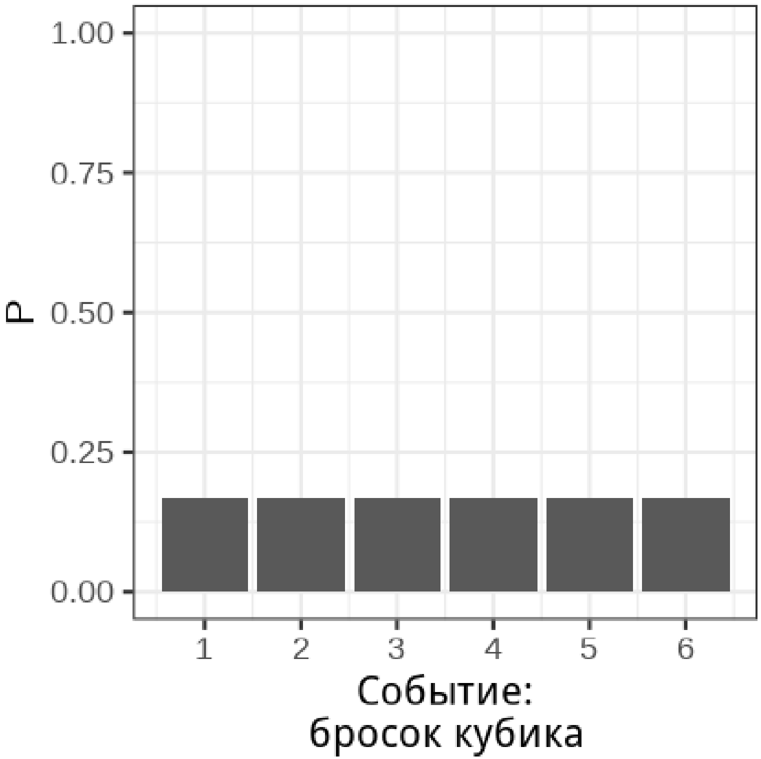
Распределение вероятностей

Все возможные исходы и их вероятности.

Дискретные и непрерывные распределения вероятностей

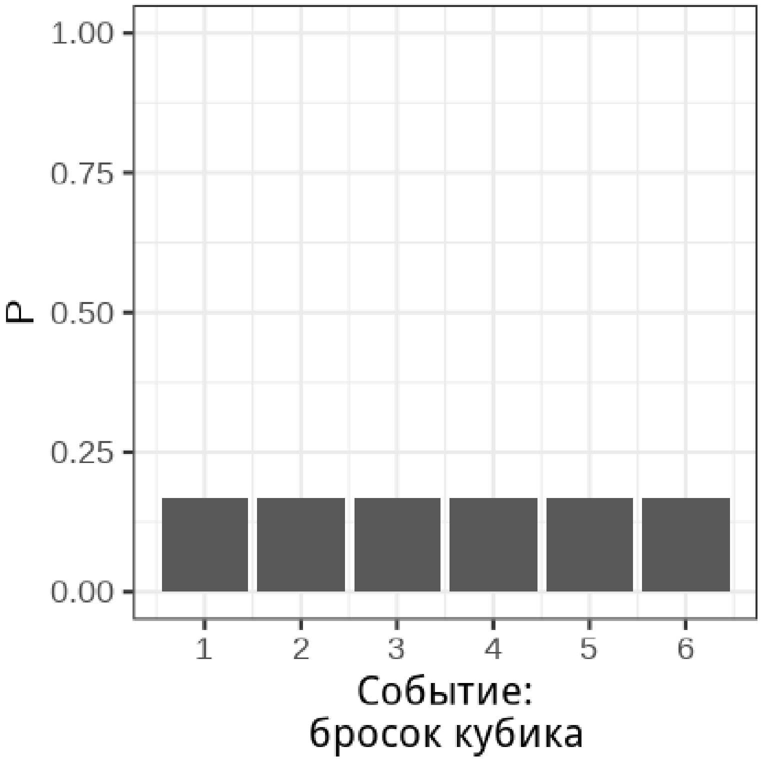
Дикретные распределения вероятностей

Для кубика вероятность каждого исхода — 1/6.

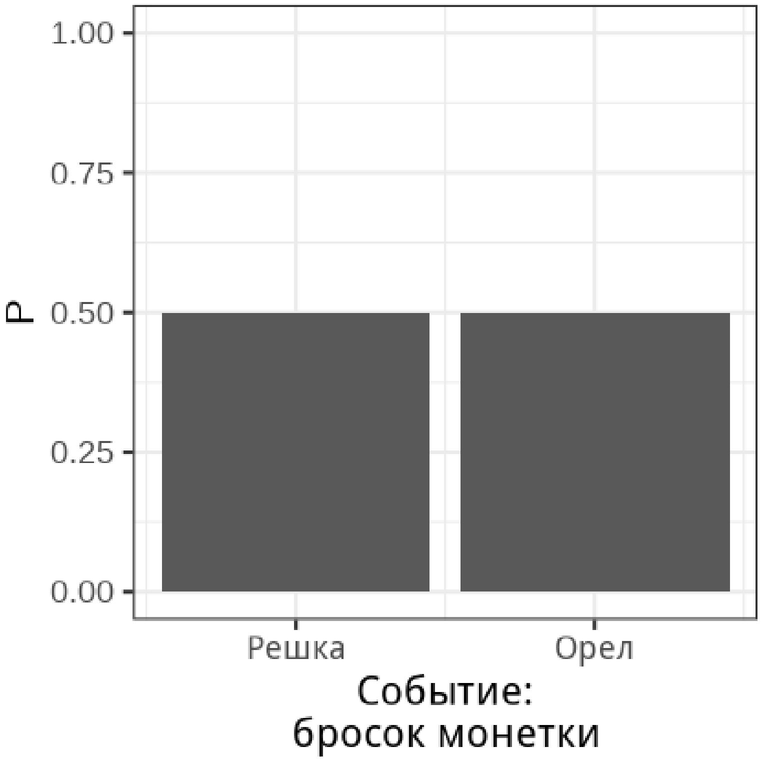


Дикретные распределения вероятностей

Для кубика вероятность каждого исхода — 1/6.

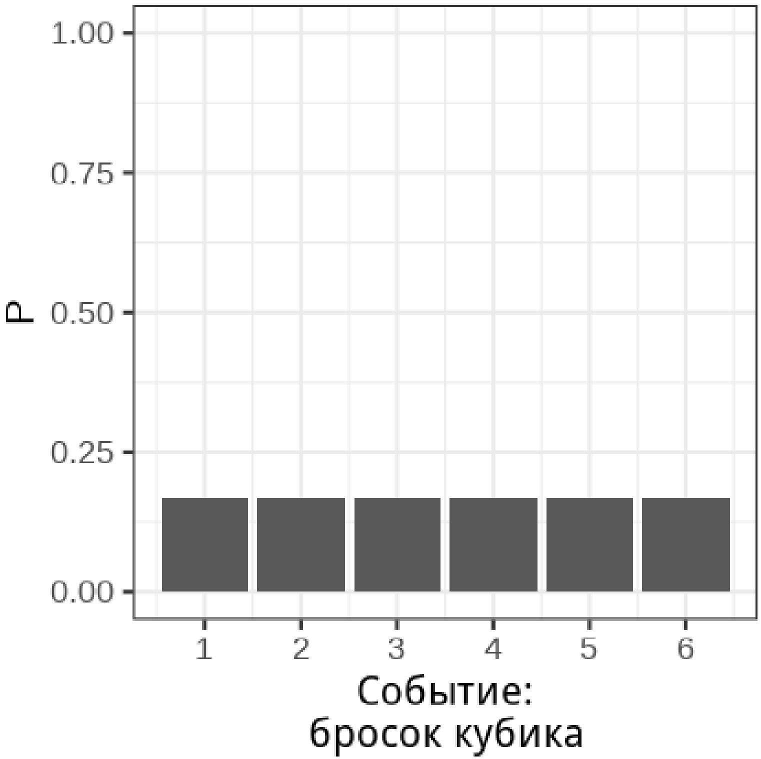


Для честной монетки вероятность каждого исхода — 1/2

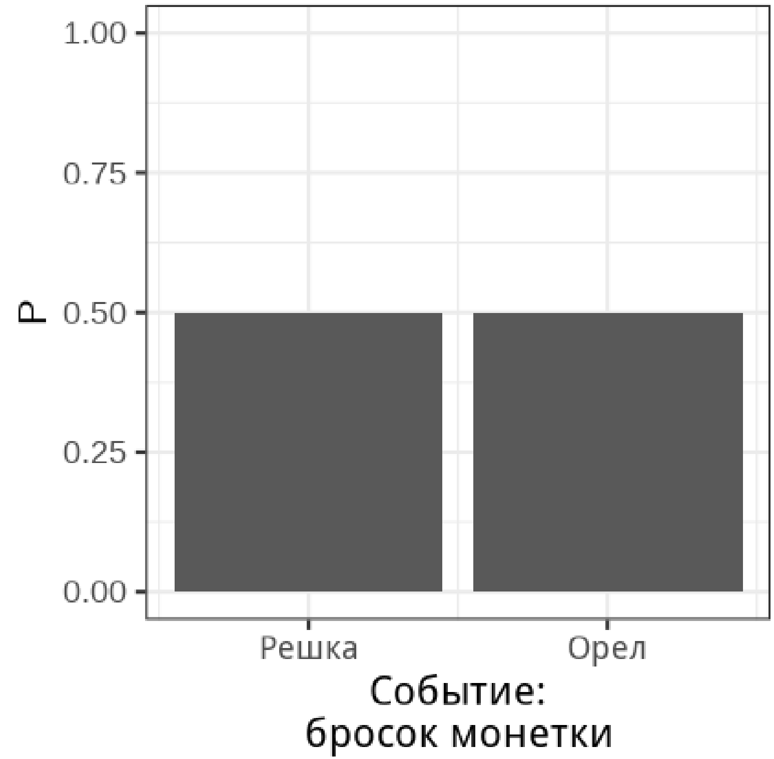


Дикретные распределения вероятностей

Для кубика вероятность каждого исхода — 1/6.



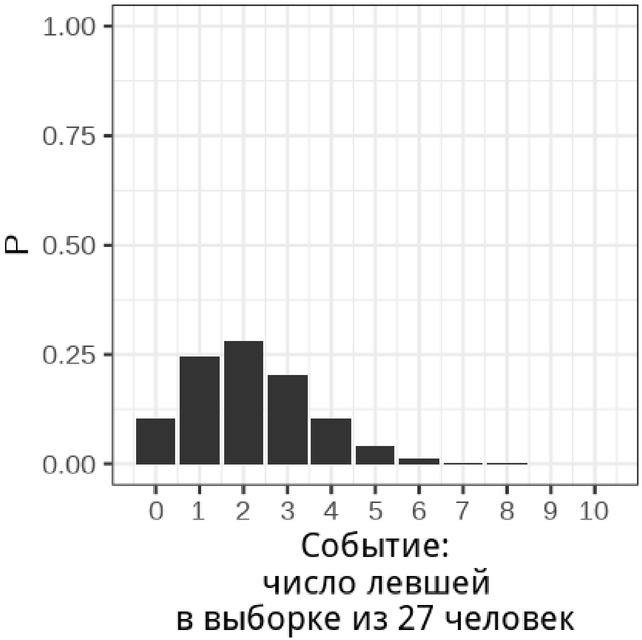
Для честной монетки вероятность каждого исхода — 1/2



Теоретическое распределение дискретной случайной величины x описывает вероятность получения определенного значения x.

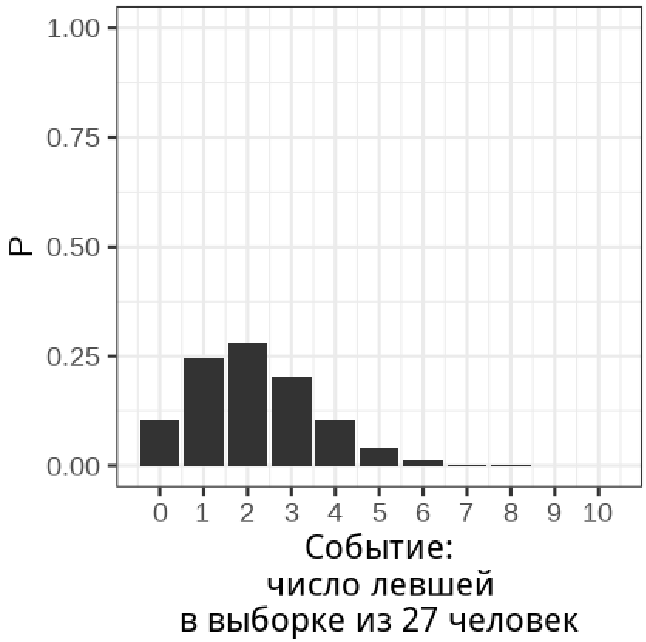
Другие примеры дискретных распределений

Распределение числа левшей в случайной выборке из 27 человек. Ожидаемая вероятность леворукости — 0.08.

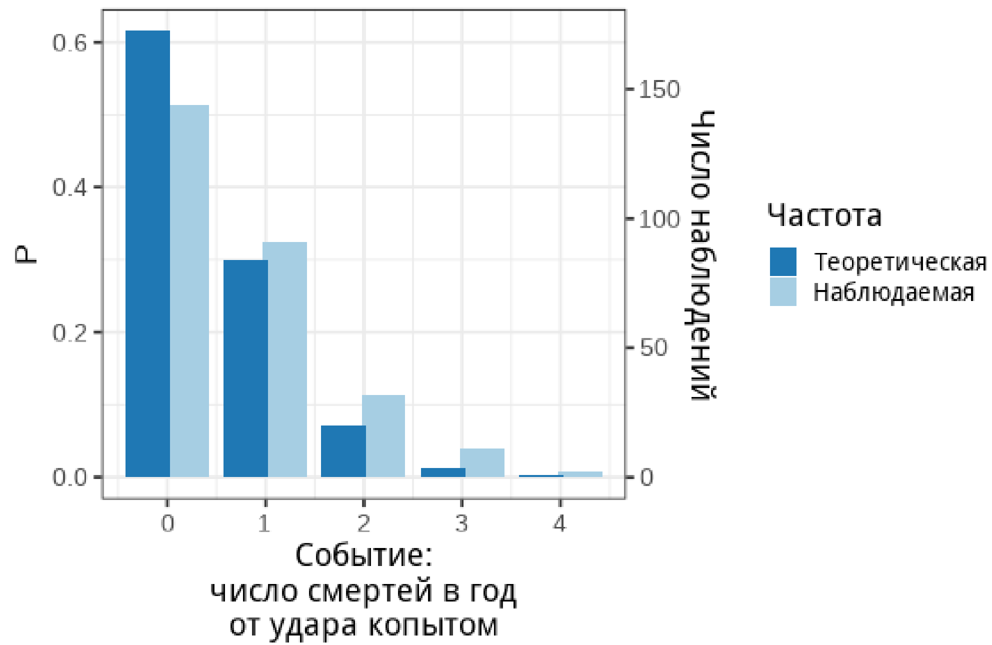


Другие примеры дискретных распределений

Распределение числа левшей в случайной выборке из 27 человек. Ожидаемая вероятность леворукости — 0.08.

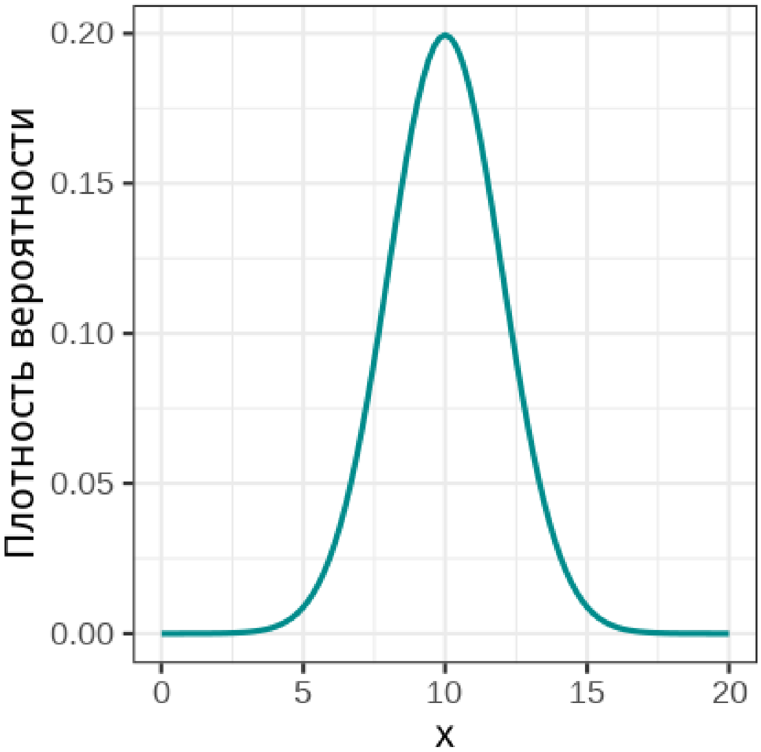


Распределение числа смертей в год от удара копытом лошади или мула в корпусах Прусской армии. Ожидаемая вероятность смерти от удара копытом за год в корпусе армии — 0.486.



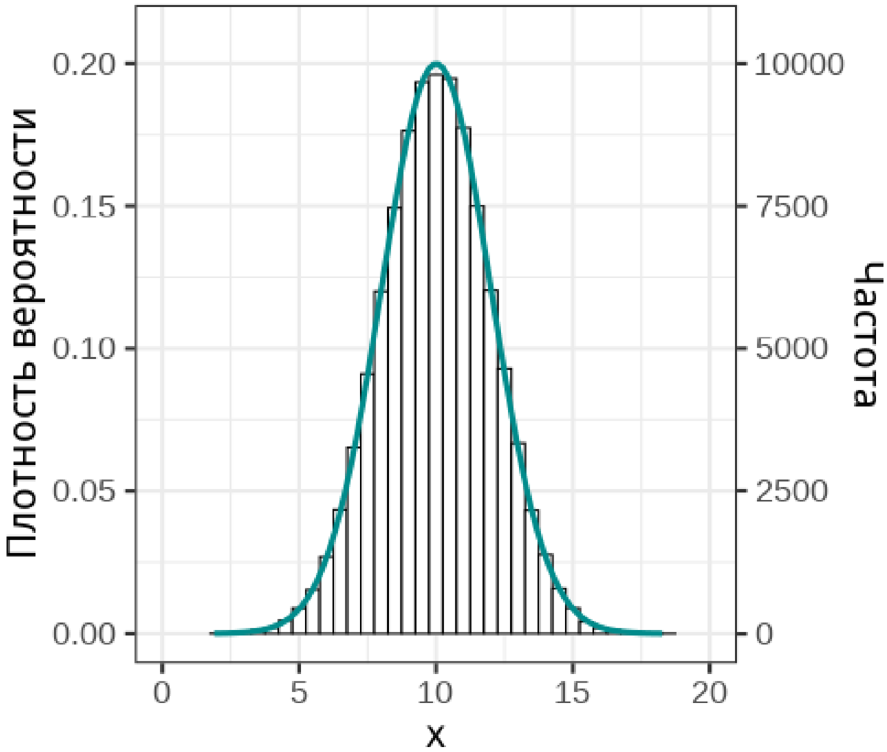
Данные: von Bortkiewicz, 1898; источник: пакет `pscl`

Непрерывные распределения вероятностей



Описывают, какие значения может принимать непрерывная случайная величина.

Относительная частота и плотность вероятности

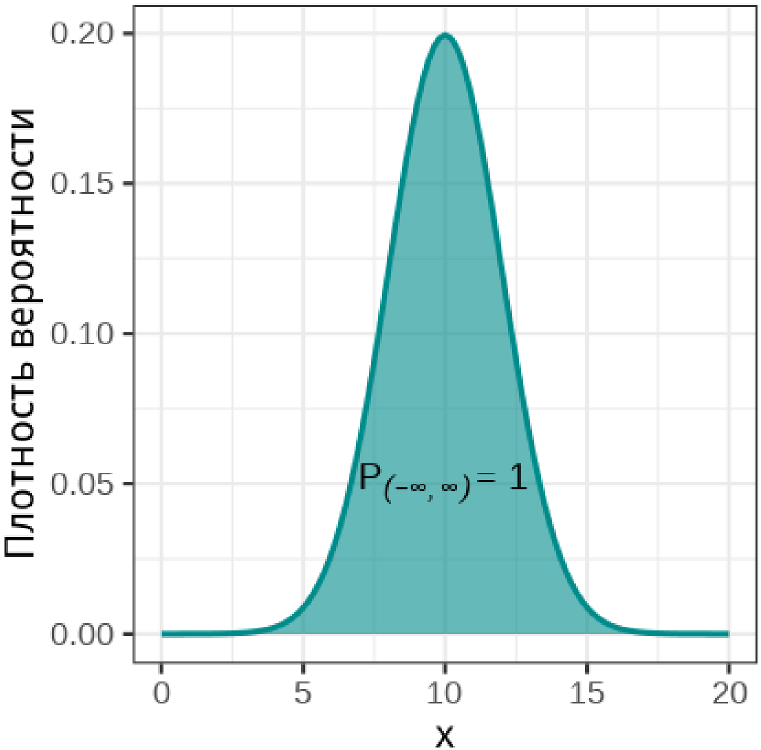


На “сырых” данных мы можем посчитать число наблюдений с разными значениями x. **Частота** — это то, что мы бы нарисовали на гистограмме.

Теоретическое распределение непрерывной случайной величины x описывает вероятность получить значение x в определенном диапазоне т.е. **плотность вероятности**.

Плотность вероятности $f(x)$ — это способ задания вероятности непрерывной случайной величины x на любом диапазоне значений.

Вероятности — площади под кривой распределения



Площадь под всей кривой $= 1$.

Вероятность встречи значений x из определенного промежутка можно узнать, проинтегрировав функцию распределения $f(x)$.

Вероятность конкретного значения нельзя определить, т.к. это точка, а под точкой нет площади.

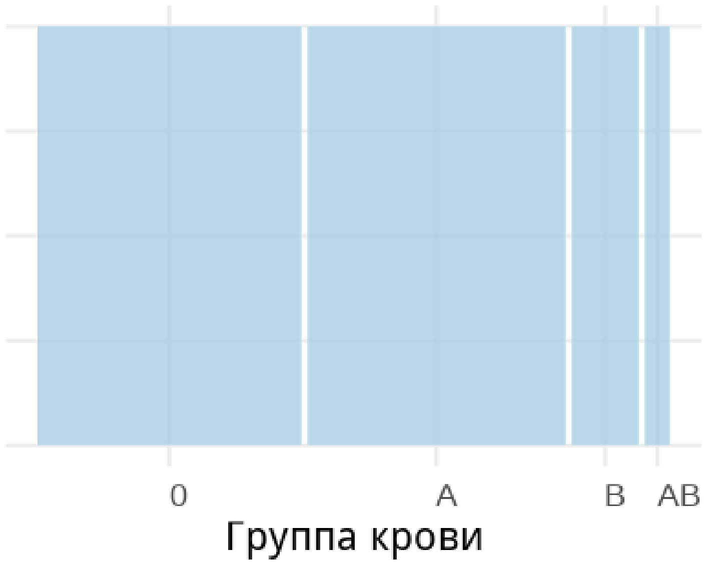
Действия с вероятностями

Группы крови

В среднеевропейской популяции частота встречаемости групп крови по системе AB0 представлена в таблице.

Группа крови	Вероятность
0	0.43
A	0.42
B	0.11
AB	0.04

http://www.almazovcentre.ru/, 18.08.2022



Нельзя иметь одновременно две группы крови по системе AB0, поэтому это несовместные события.

Рассмотрим на их примере действия с вероятностями.

Сложение вероятностей

Если события A и B несовместны, то вероятность того, что произойдет одно или другое — это сумма их вероятностей.

$P(A \; or \; B) = P(A) + P(B)$

Сложение вероятностей

Если события A и B несовместны, то вероятность того, что произойдет одно или другое — это сумма их вероятностей.

$$P(A \; or \; B) = P(A) + P(B)$$

Какова вероятность, что у человека кровь одной из трех групп: (A, B или AB)?

Группа крови	Вероятность
0	0.43
A	0.42
B	0.11
AB	0.04

http://www.almazovcentre.ru/, 18.08.2022

Сложение вероятностей

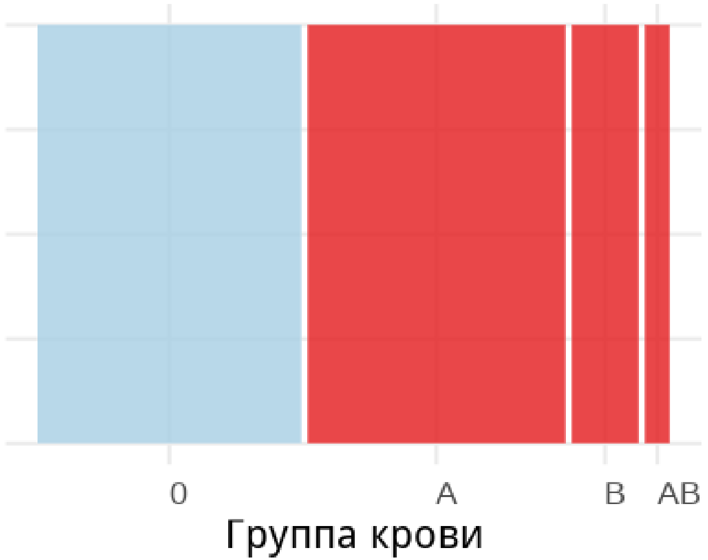
Если события A и B несовместны, то вероятность того, что произойдет одно или другое — это сумма их вероятностей.

$$P(A \; or \; B) = P(A) + P(B)$$

Какова вероятность, что у человека кровь одной из трех групп: (A, B или AB)?

Группа крови	Вероятность
0	0.43
A	0.42
B	0.11
AB	0.04

http://www.almazovcentre.ru/, 18.08.2022



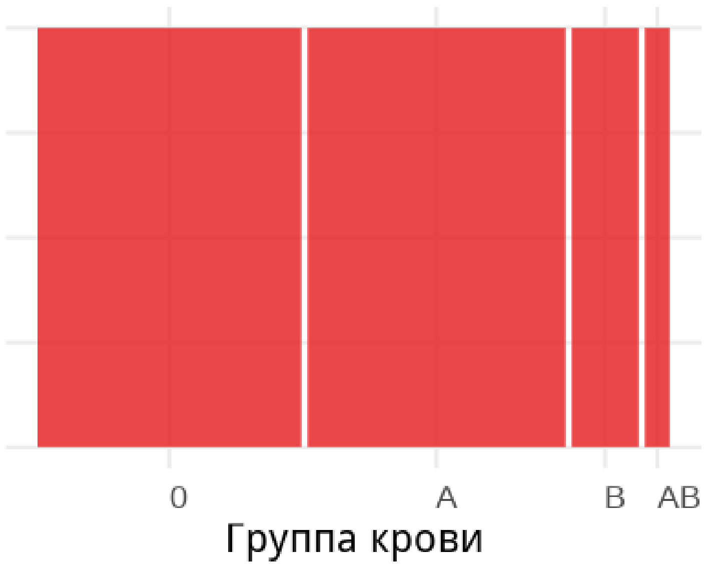
$$P(A \; or \; B \; or \; AB) = P(A) + P(B) + P(AB) = 0.57$$

Пространство всех событий

Сумма вероятностей всех возможных несовместных событий равна 1

Группа крови	Вероятность
0	0.43
A	0.42
B	0.11
AB	0.04

http://www.almazovcentre.ru/, 18.08.2022



$$P(0) + P(A) + P(B) + P(AB) = 1$$

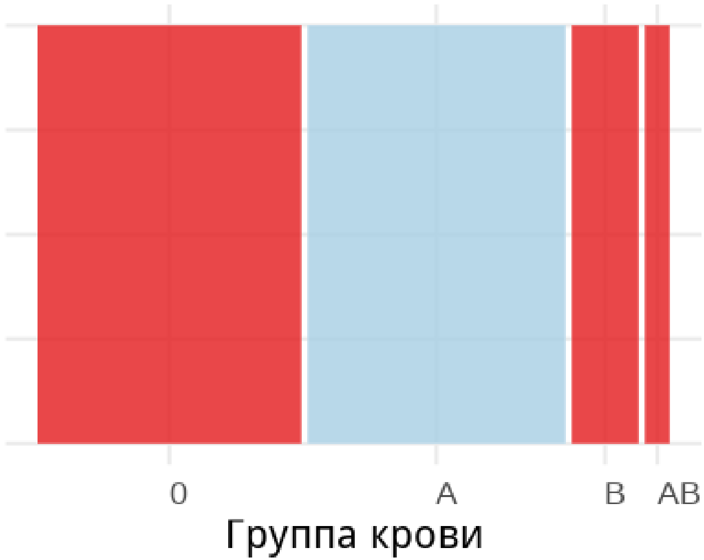
Отрицание

Вероятность того, что событие не произойдет равна 1 минус вероятность того, что оно произойдет.

$P(not A) = 1 - P(A)$

Группа крови	Вероятность
0	0.43
A	0.42
B	0.11
AB	0.04

http://www.almazovcentre.ru/, 18.08.2022



$$P(not A) = 1 - P(A) = 0.58$$

Независимые события

События **независимы**, если то, что произошло одно из них, никак не влияет на то, что произойдет второе.

Помимо групп крови по системе AB0 есть еще резус фактор Rh + или -.
Эти признаки взаимно независимы.

В европейской популяции частота Rh+ 0.85.

Группа крови	Вероятность
0	0.43
A	0.42
B	0.11
AB	0.04

Произведение независимых событий

Если события A и B независимы, то вероятность того, что произошли оба события одновременно равна произведению их вероятностей. (Это справедливо для любого числа независимых событий).

$$P(A \; and \; B) = P(A) \times P(B)$$

В европейской популяции частота Rh+ 0.85.

Группа крови	Вероятность
0	0.43
A	0.42
B	0.11
AB	0.04

Каковы вероятности групп крови AB0 с учетом Rh (если предположить, что они независимы)?

Произведение независимых событий

Если события A и B независимы, то вероятность того, что произошли оба события одновременно равна произведению их вероятностей. (Это справедливо для любого числа независимых событий).

$$P(A\ and\ B) = P(A) \times P(B)$$

В европейской популяции частота Rh+ 0.85.

Группа крови	Вероятность
0	0.43
A	0.42
B	0.11
AB	0.04

Каковы вероятности групп крови AB0 с учетом Rh (если предположить, что они независимы)?

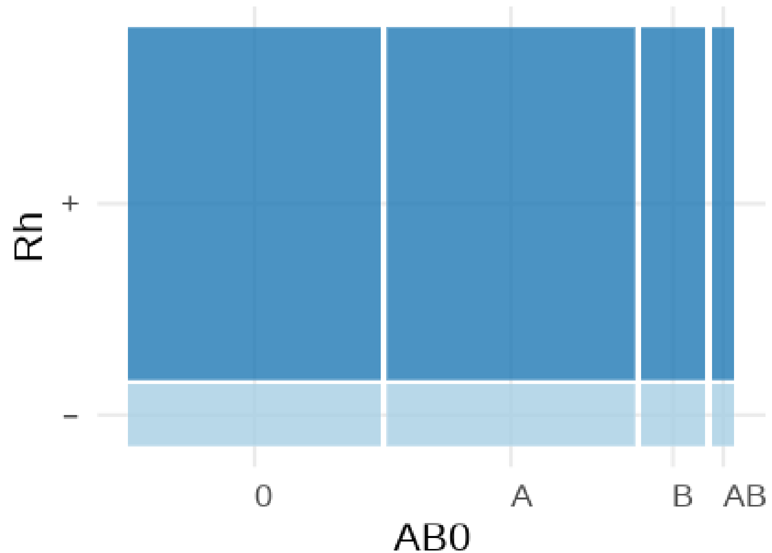
AB0	Rh	Вероятность
0	+	0.366
0	–	0.064
A	+	0.357
A	–	0.063
B	+	0.094
B	–	0.016
AB	+	0.034
AB	–	0.006

Независимые события и их произведение

В европейской популяции частота Rh+ 0.85. Каковы вероятности групп крови AB0 с учетом Rh (если предположить, что они независимы)?

AB0	Rh	Вероятность
0	+	0.366
0	–	0.064
A	+	0.357
A	–	0.063
B	+	0.094
B	–	0.016
AB	+	0.034
AB	–	0.006

Поскольку Rh и AB0 независимы, то соотношение Rh+ и Rh- будет одинаково в каждой группе AB0.



Деревья вероятностей

Деревья вероятностей (probability trees)

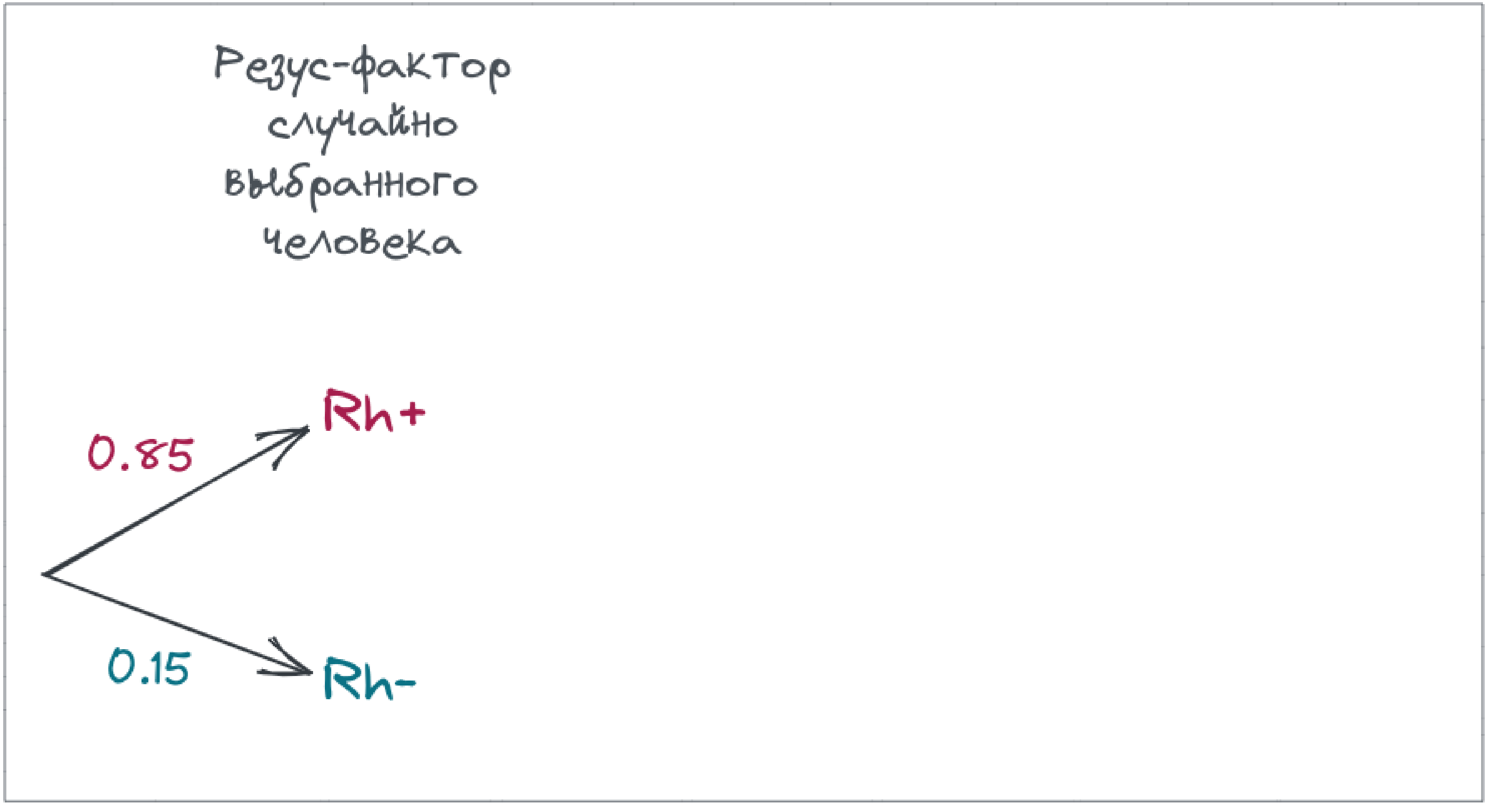
Дерево вероятностей — это способ изобразить вероятности сочетаний нескольких случайных событий.

В европейской популяции частота Rh+ 0.85.

Какова вероятность,

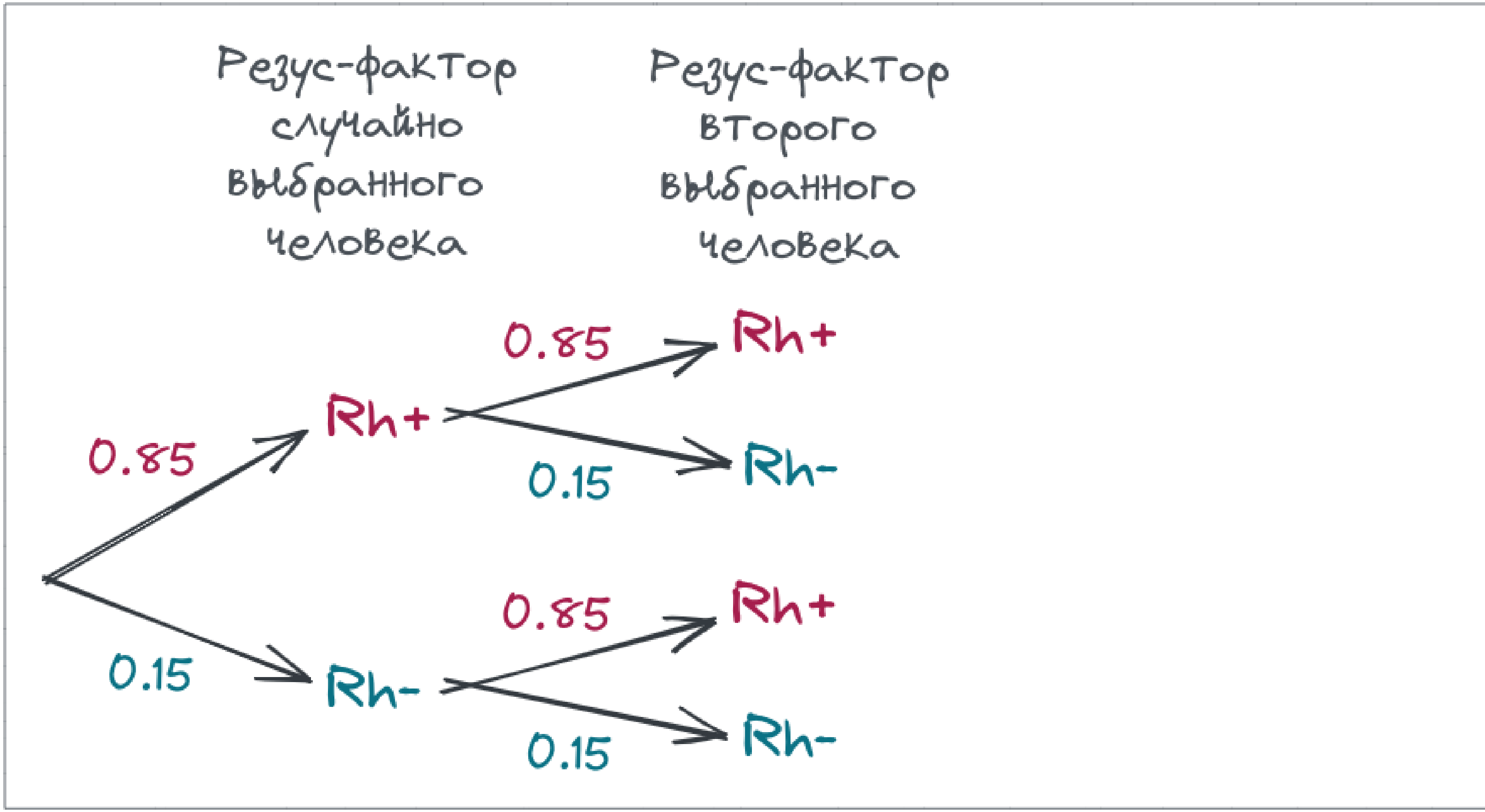
- что у случайно выбранного человека будет Rh-?
- что у двух случайных людей будет Rh-?
- что у двух случайных людей будет одинаковый резус-фактор?

Нарисуем дерево вероятностей



Если речь идет об одном случайно выбранном человеке — все просто.

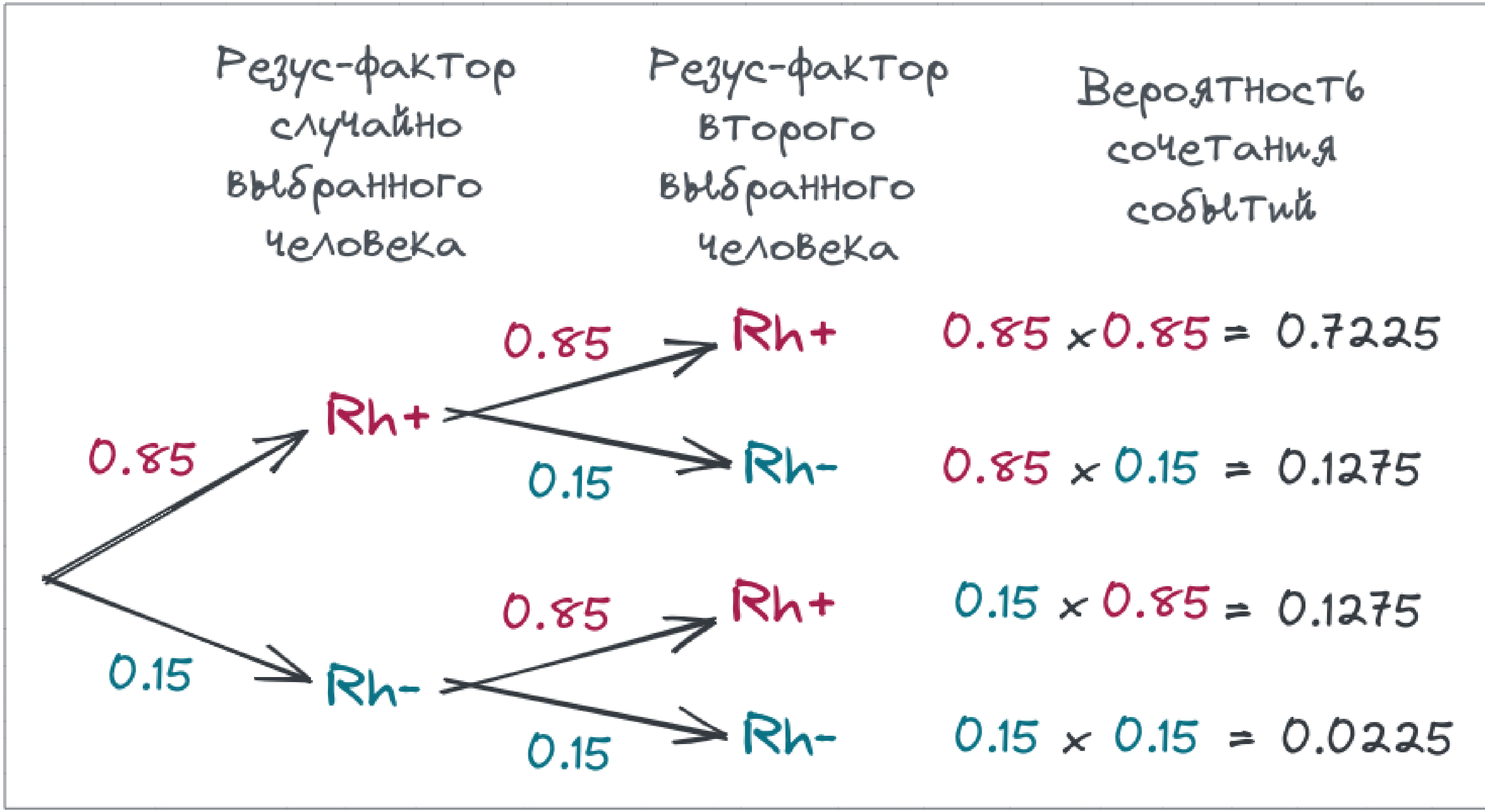
Нарисуем дерево вероятностей



Добавляем второго случайно выбранного человека.

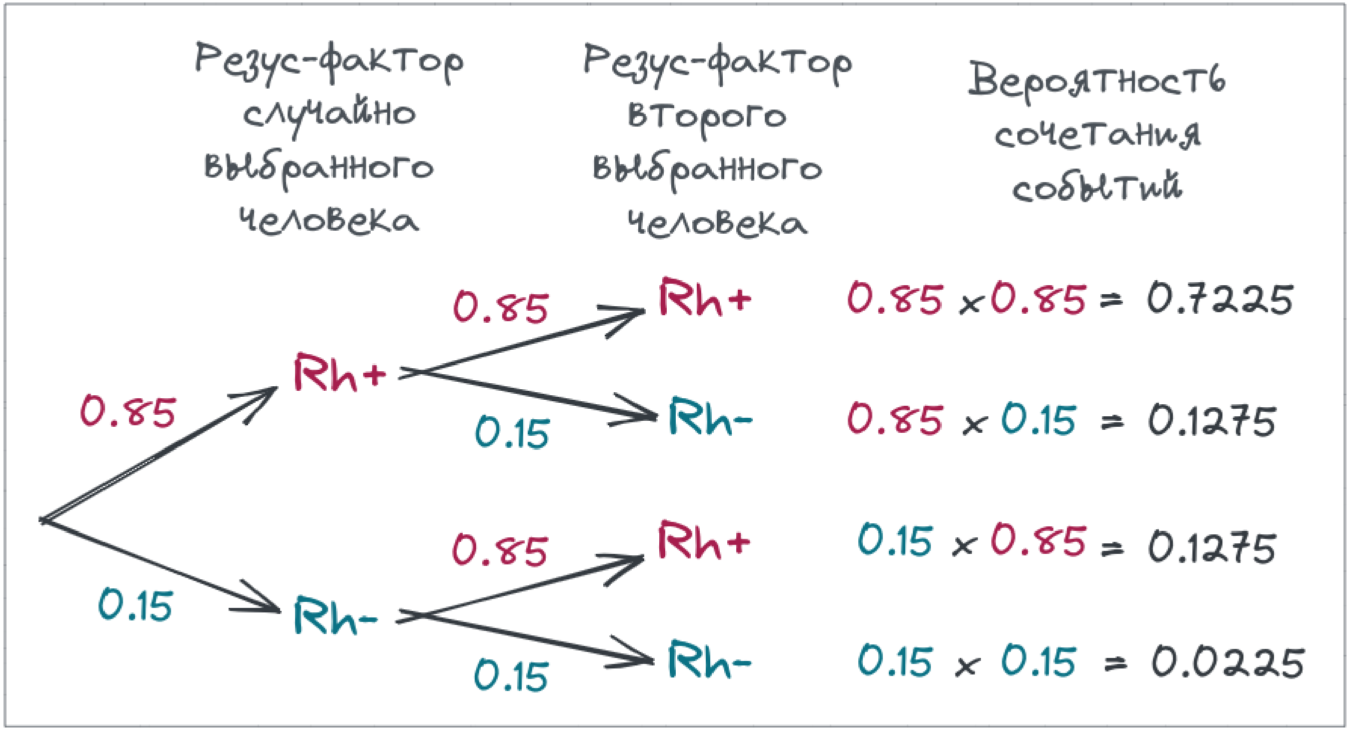
Резус фактор второго случайно выбранного человека не зависит от первого.

Нарисуем дерево вероятностей



Поскольку резус фактор второго случайно выбранного человека не зависит от первого, вероятность сочетаний этих двух независимых событий можно посчитать, перемножив вероятности.

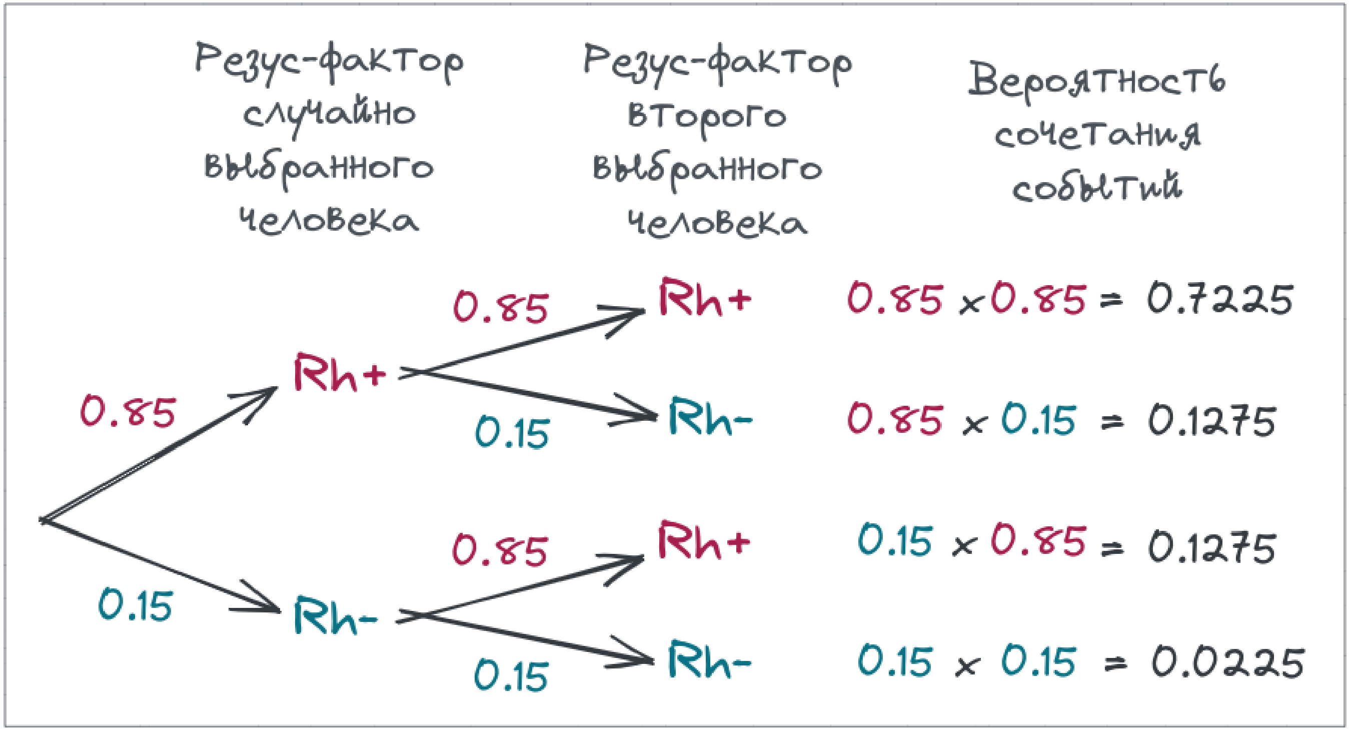
Теперь можно ответить на вопросы



Какова вероятность, что

- у случайно выбранного человека Rh-?
- у двух случайных людей Rh-?
- у двух случайных людей одинаковый резус-фактор?

Теперь можно ответить на вопросы



Какова вероятность, что

- у случайно выбранного человека Rh-?
- у двух случайных людей Rh-?
- у двух случайных людей одинаковый резус-фактор?

- 0.15
- 0.0225
- 0.7225 + 0.0225 = 0.745

Зависимые события

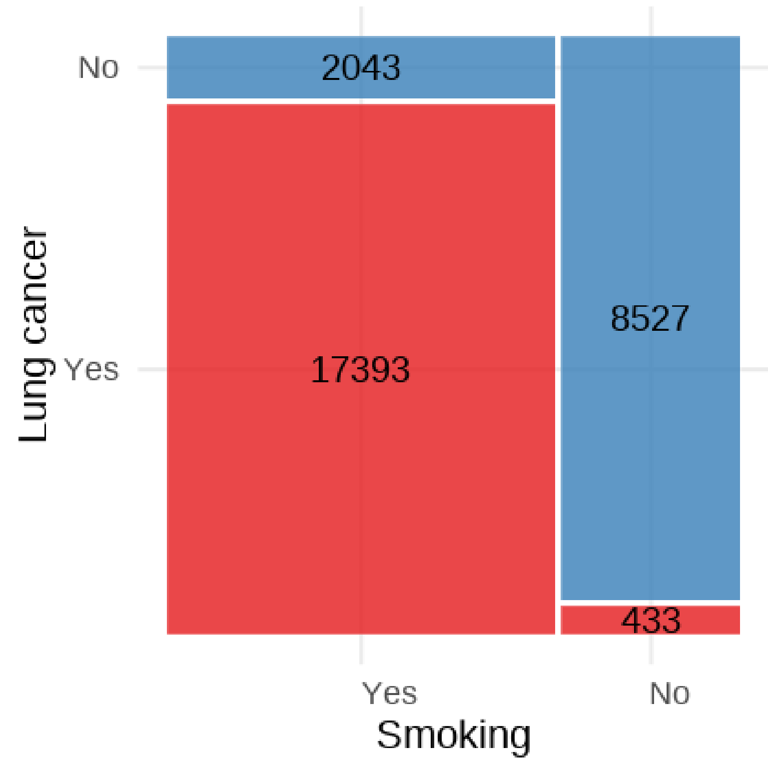
Зависимые события (dependent events)

События **зависимы**, если от появления одного из них зависит вероятность появления другого.

По результатам мета-анализа исследований влияния курения на возникновение рака легких видно, что для курящих вероятность появления рака легких выше, чем для некурящих. Это зависимые события.

	Smoking		
Lung_cancer	**Yes**	**No**	**Total**
Yes	17393	433	**17826**
No	2043	8527	**10570**
Total	**19436**	**8960**	**28396**

Данные Barukčić, 2019, table 7; DOI: 10.22270/jddt.v9i1-s.2273



Задание

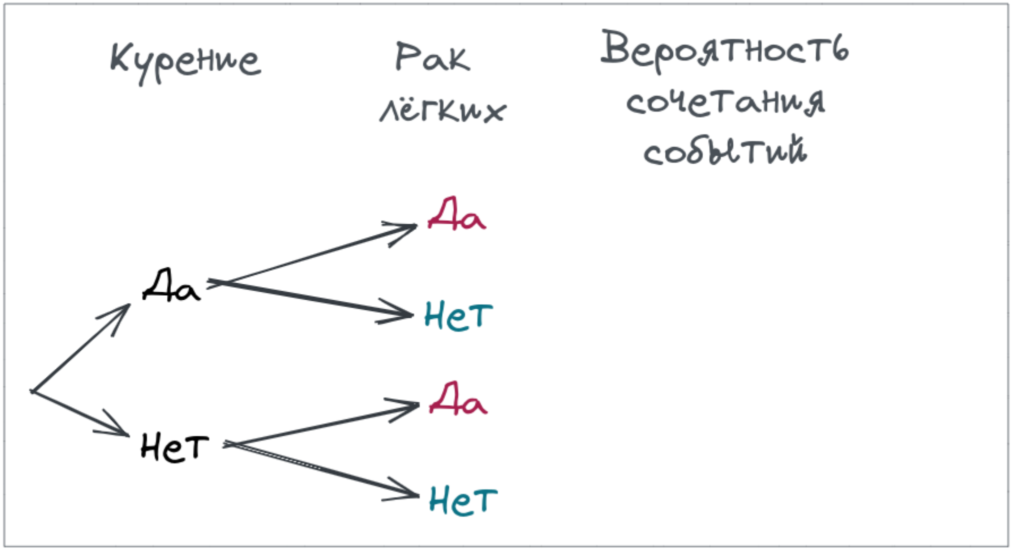
Нарисуйте дерево вероятностей

	Smoking		
Lung_cancer	**Yes**	**No**	**Total**
Yes	17393	433	**17826**
No	2043	8527	**10570**
Total	**19436**	**8960**	**28396**

Задание

Нарисуйте дерево вероятностей

	Smoking		
Lung_cancer	Yes	No	Total
Yes	17393	433	**17826**
No	2043	8527	**10570**
Total	**19436**	**8960**	**28396**



Вероятность курения/некурения среди всех:

$P_{\text{курит=да}} =$
$P_{\text{курит=нет}} =$

Вероятность рака среди курящих:

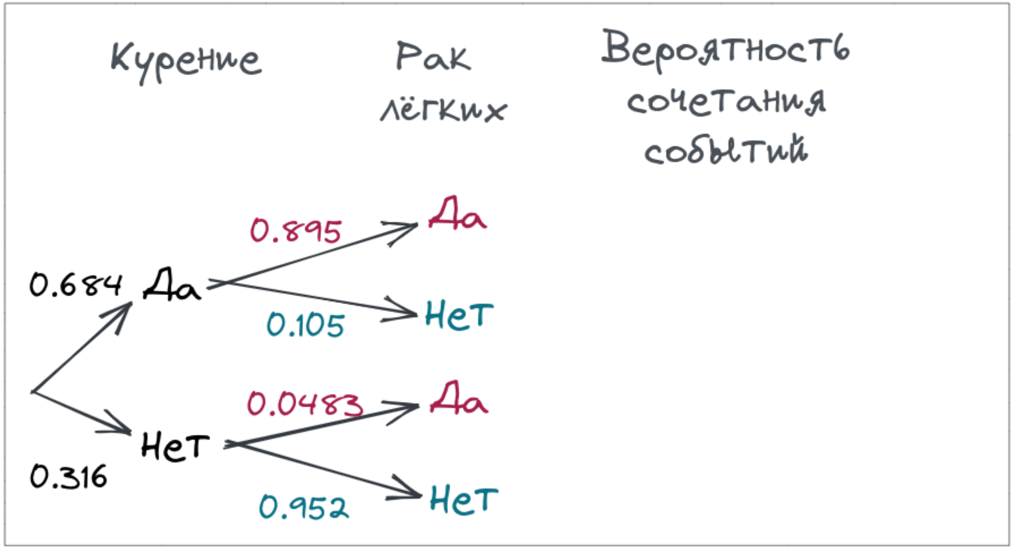
$P_{\text{рак=да}|\text{курит=да}} =$
$P_{\text{рак=нет}|\text{курит=да}} =$

Вероятность рака среди некурящих:

$P_{\text{рак=да}|\text{курит=нет}} =$
$P_{\text{рак=нет}|\text{курит=нет}} =$

Решение

	Smoking		
Lung_cancer	Yes	No	Total
Yes	17393	433	**17826**
No	2043	8527	**10570**
Total	**19436**	**8960**	**28396**



Вероятность курения/некурения среди всех:

$$P_{\text{курит=да}} = 19436/28396 = 0.684$$
$$P_{\text{курит=нет}} = 8960/28396 = 0.316$$

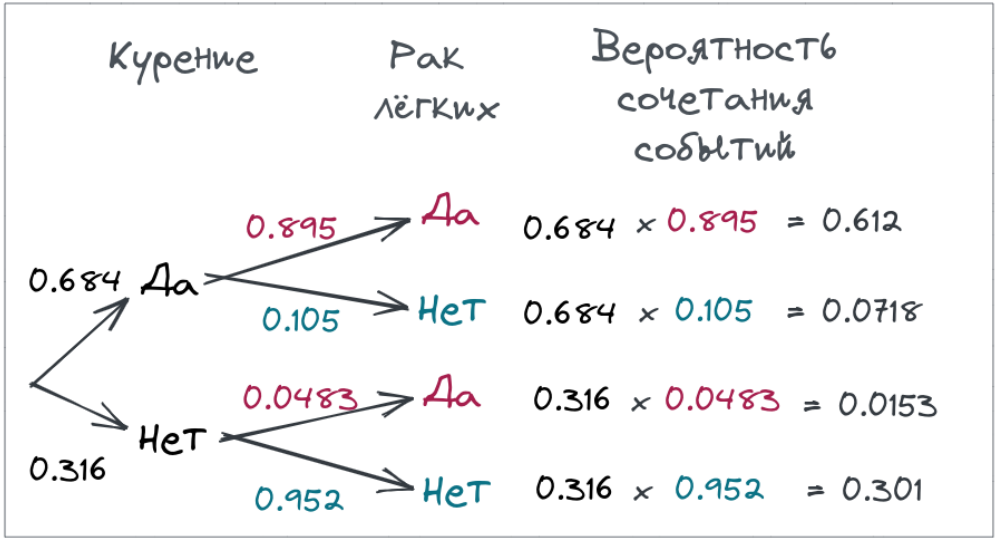
Вероятность рака среди курящих:

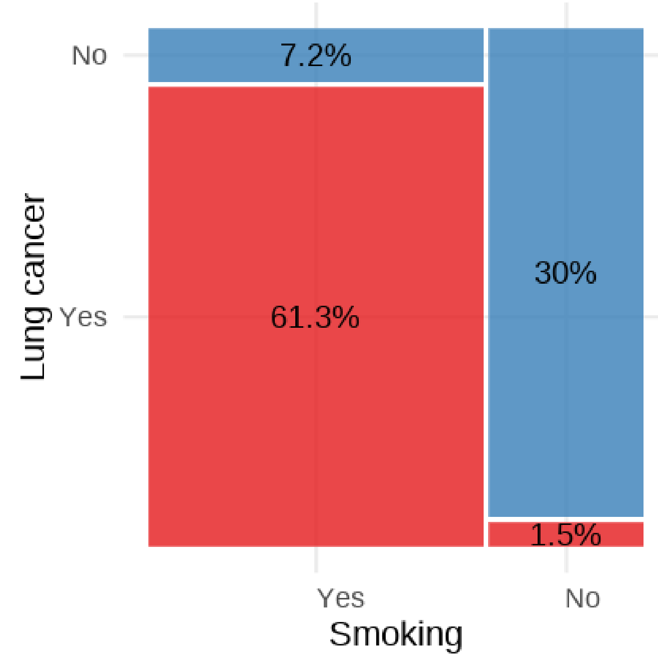
$$P_{\text{рак=да|курит=да}} = 17393/19436 = 0.895$$
$$P_{\text{рак=нет|курит=да}} = 2043/19436 = 0.105$$

Вероятность рака среди некурящих:

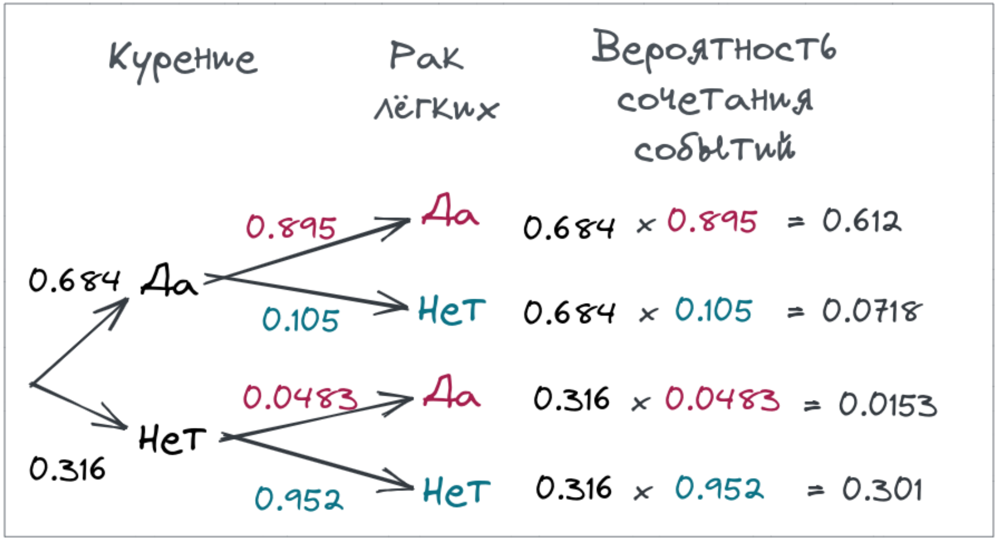
$$P_{\text{рак=да|курит=нет}} = 433/8960 = 0.048$$
$$P_{\text{рак=нет|курит=нет}} = 8527/8960 = 0.952$$

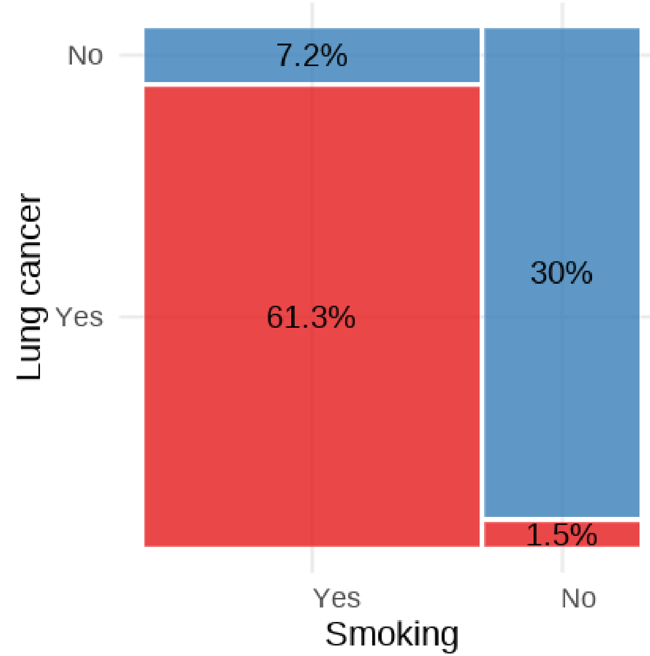
Дерево вероятностей для этих данных





Дерево вероятностей для этих данных

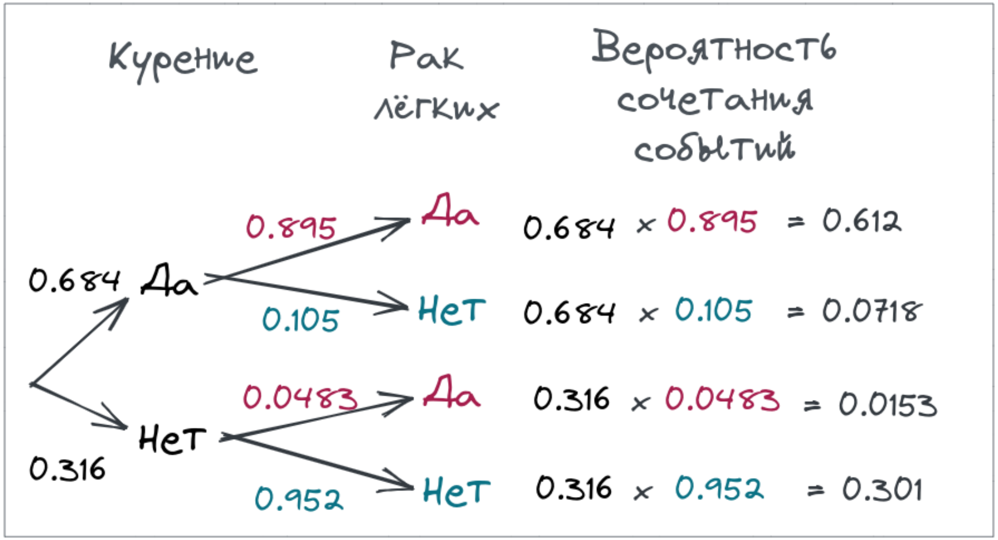


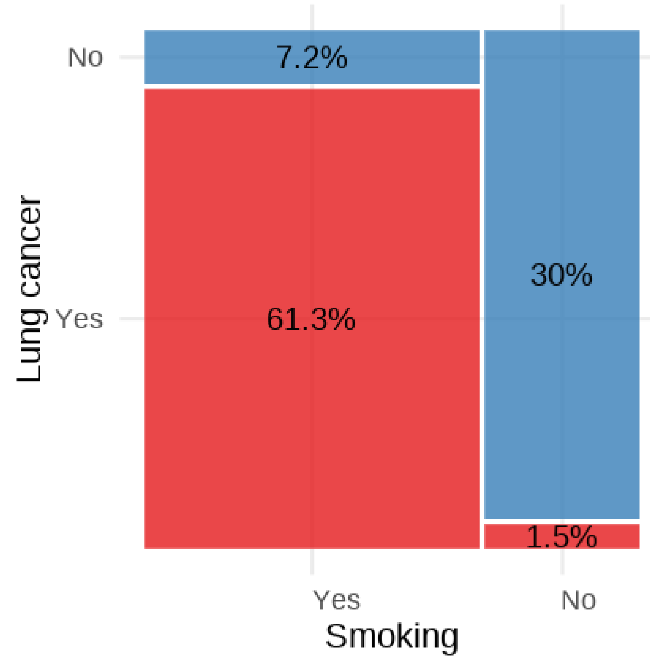


Осторожно! Поскольку события “курение” и “рак легких” — зависимые, правило умножения не выполняется:

$$P(\text{курение и рак}) \neq P(\text{курение}) \times P(\text{рак})$$

Дерево вероятностей для этих данных





Осторожно! Поскольку события “курение” и “рак легких” — зависимые, правило умножения не выполняется:

$$P(\text{курение и рак}) \neq P(\text{курение}) \times P(\text{рак})$$

На самом деле $P(\text{курение и рак}) = 0.612$.

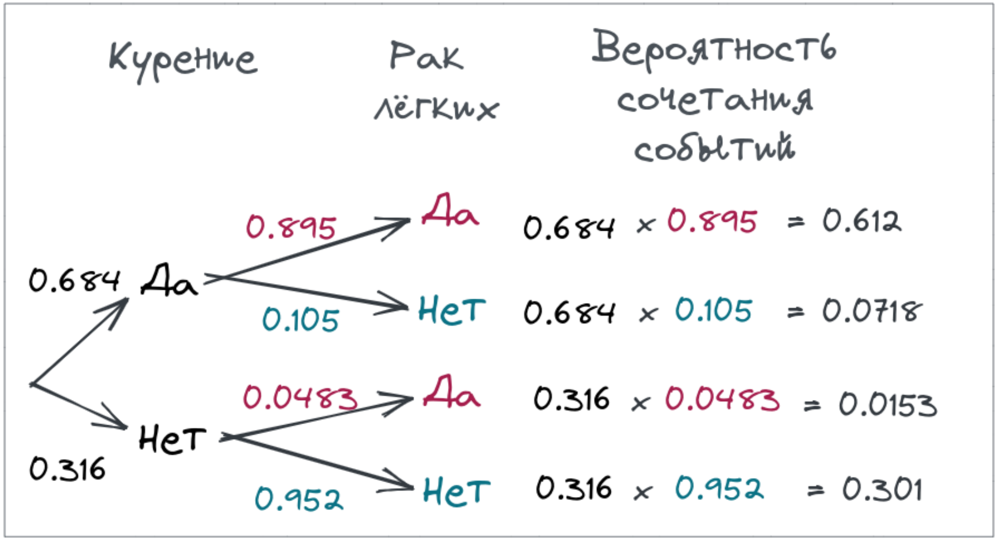
А если ошибочно считать по формуле для независимых событий, то

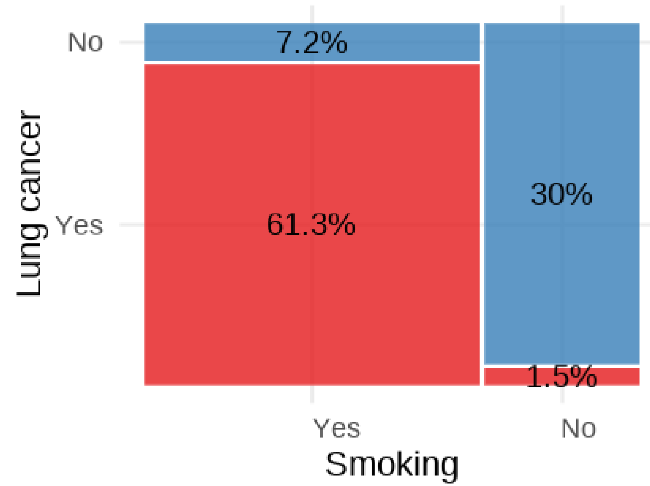
$$P(\text{курение}) \times P(\text{рак}) = 0.684 \times (0.895 + 0.0483) = 0.645.$$

Условная вероятность

Условная вероятность (conditional probability)

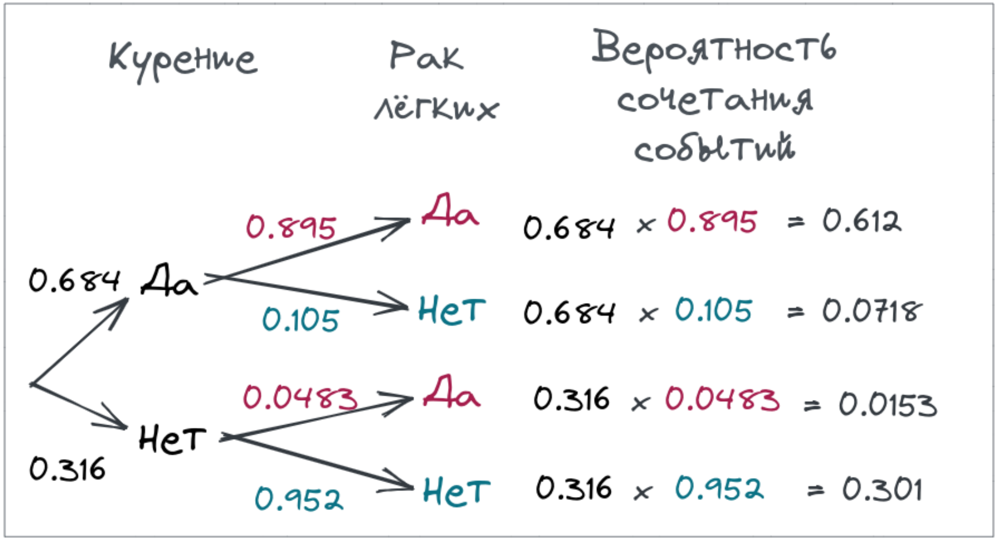
Условная вероятность — вероятность события, при каком-то условии (например, при условии, что произошло какое-то другое событие или события).

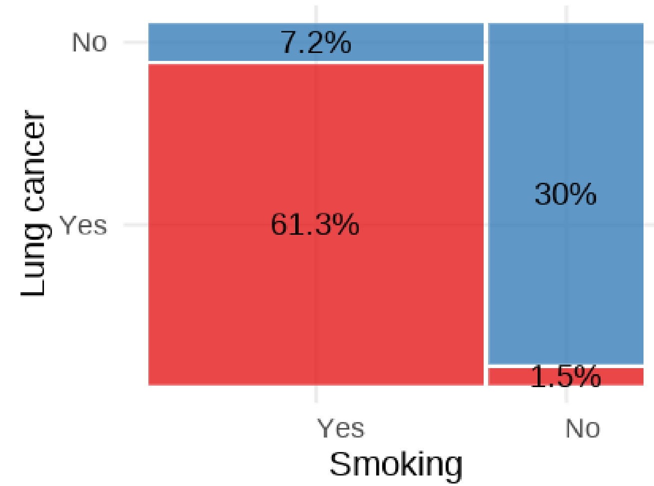




Условная вероятность (conditional probability)

Условная вероятность — вероятность события, при каком-то условии (например, при условии, что произошло какое-то другое событие или события).





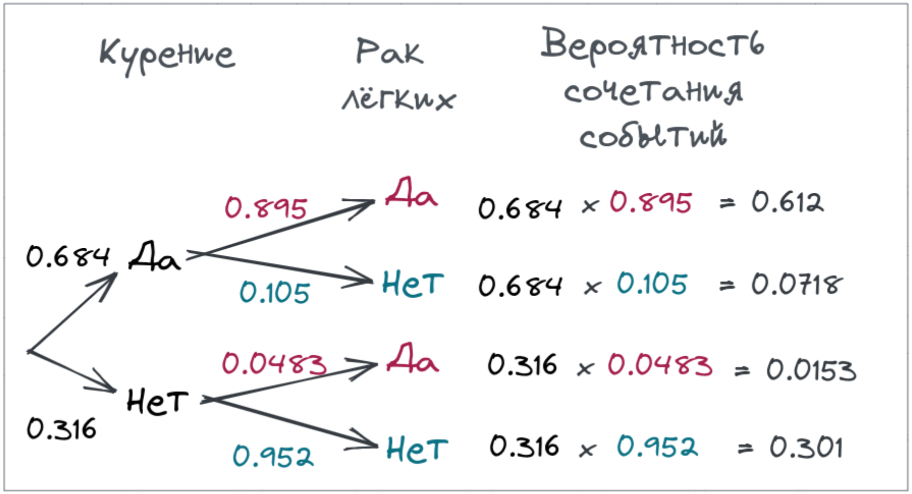
$P(\text{рак} = \text{да}|\text{курит} = \text{да}) = 0.895$ —
вероятность рака легких при условии, что человек курит

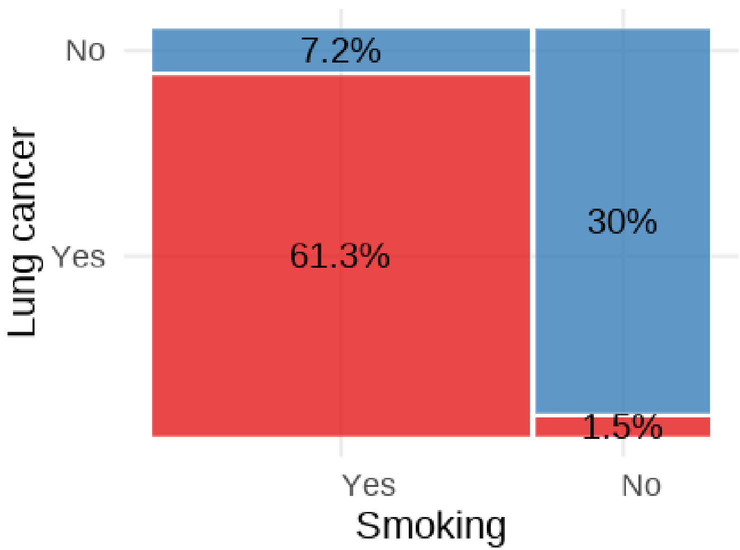
$P(\text{рак} = \text{да}|\text{курит} = \text{нет}) = 0.0483$ —
вероятность рака легких при условии, что человек НЕ курит

Формула полной вероятности (law of the total probability)

Вероятность события A можно вычислить исходя из его вероятностей при условии каждого из несовместных событий B_i.

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

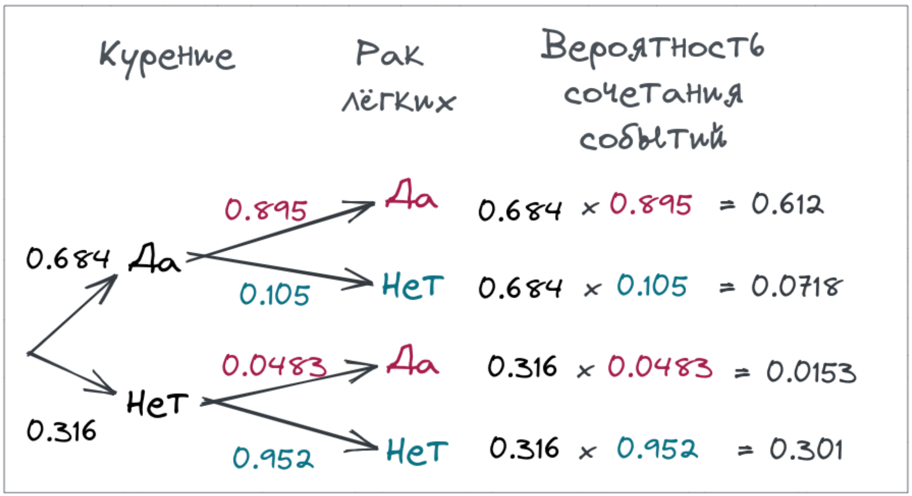


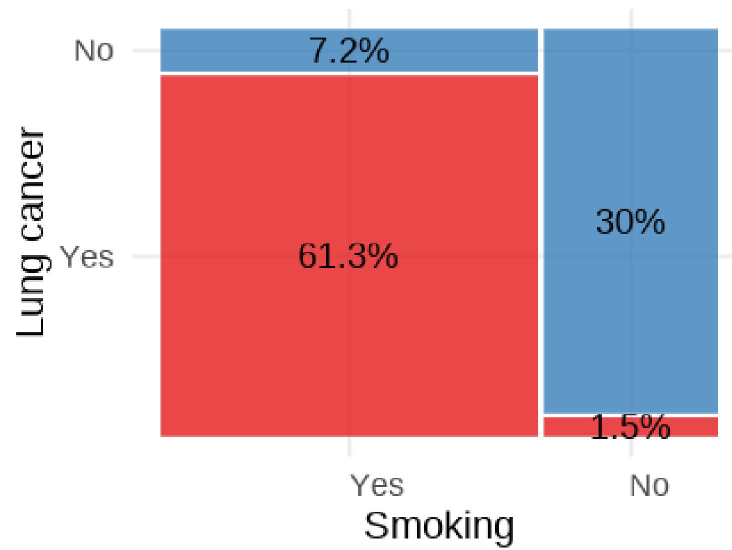


Формула полной вероятности (law of the total probability)

Вероятность события A можно вычислить исходя из его вероятностей при условии каждого из несовместных событий B_i.

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$



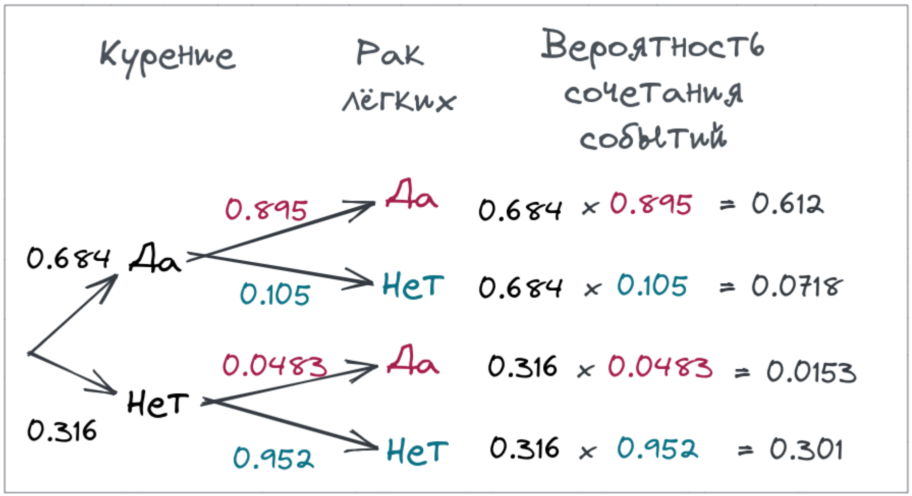


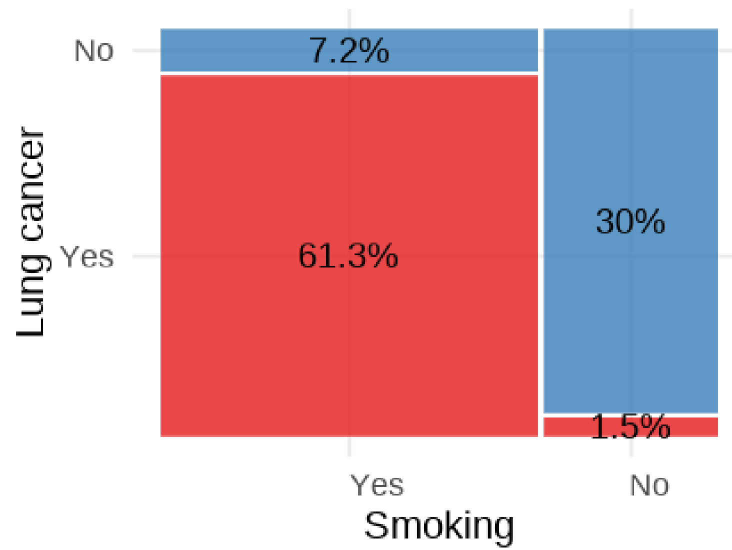
И действительно, вероятность того, что у случайно выбранного человека рак, будет складываться из площадей красных прямоугольников.

Обобщенное правило умножения вероятностей (generel multiplication rule)

Вероятность того, что произошли оба события, даже если они зависимы

$$P(A \; and \; B) = P(A) \cdot P(B|A) =$$
$$= P(B) \cdot P(A|B)$$





Задание

В 80-е годы в канаде 52% взрослых мужчин курили. По оценкам исследователей вероятность развития рака лежких в течение жизни у мужчин курильщиков была 17.2 %, а у некурящих 1.2 % (Villneuve and Mao, 1994).

- Какова условная вероятность нажить себе рак для канадца, если он курил в 80е годы?
- Какова вероятность того, что канадец курил в 80-е годы и у него развился рак?
- Какова вероятность того, что канадец не курил в 80-е годы и не получил рак?

Задание

В 80-е годы в канаде 52% взрослых мужчин курили. По оценкам исследователей вероятность развития рака лежких в течение жизни у мужчин курильщиков была 17.2 %, а у некурящих 1.2 % (Villneuve and Mao, 1994).

- Какова условная вероятность нажить себе рак для канадца, если он курил в 80е годы?
- Какова вероятность того, что канадец курил в 80-е годы и у него развился рак?
- Какова вероятность того, что канадец не курил в 80-е годы и не получил рак?

Используйте обобщенное правило умножения вероятностей

$$P(A\ and\ B) = P(A) \cdot P(B|A) =$$
$$= P(B) \cdot P(A|B)$$

Задание

В 80-е годы в канаде 52% взрослых мужчин курили. По оценкам исследователей вероятность развития рака лежких в течение жизни у мужчин курильщиков была 17.2 %, а у некурящих 1.2 % (Villneuve and Mao, 1994).

- Какова условная вероятность нажить себе рак для канадца, если он курил в 80е годы?
- Какова вероятность того, что канадец курил в 80-е годы и у него развился рак?
- Какова вероятность того, что канадец не курил в 80-е годы и не получил рак?

Используйте обобщенное правило умножения вероятностей

$$P(A\ and\ B) = P(A) \cdot P(B|A) =$$
$$= P(B) \cdot P(A|B)$$

Ради интереса и самопроверки, нарисуйте дерево вероятностей и сравните результаты.

Решение

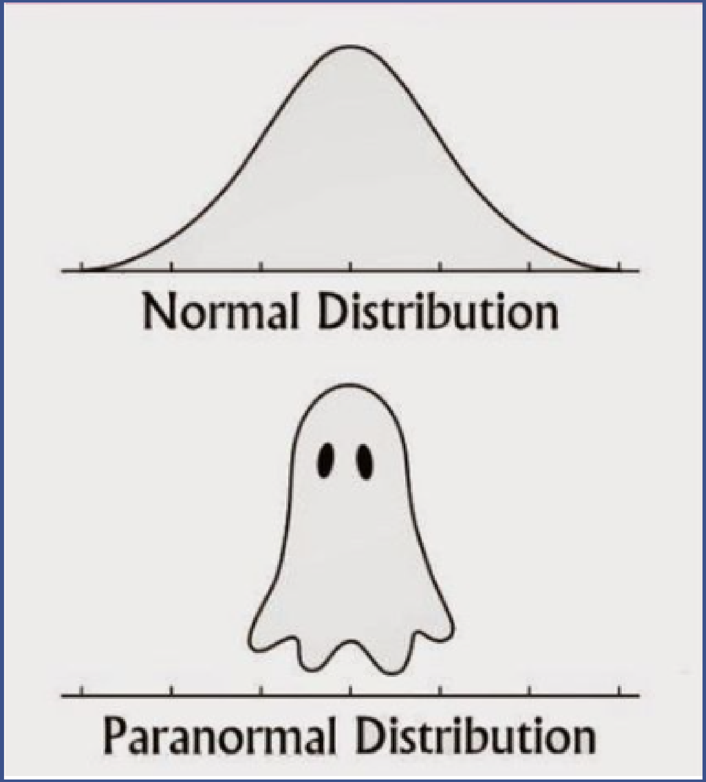
В 80-е годы в канаде 52% взрослых мужчин курили. По оценкам исследователей вероятность развития рака лежких в течение жизни у мужчин курильщиков была 17.2 %, а у некурящих 1.2 % (Villneuve and Mao, 1994).

- Какова условная вероятность нажить себе рак для канадца, если он курил в 80е годы?
  $P_{\text{рак}|\text{курильщик}} = 0.172$
- Какова вероятность того, что канадец курил в 80-е годы и у него развился рак?
  $P_{\text{курильщик и рак}} = P_{\text{курильщик}} \times P_{\text{рак}|\text{курильщик}} = 0.52 \times 0.172 = 0.089$
- Какова вероятность того, что канадец не курил в 80-е годы и не получил рак?
  $P_{\text{курильщик и нет рака}} = P_{\text{некурящий}} \times P_{\text{нет рака}|\text{некурящий}} = 0.48 \times 0.987 = 0.474$

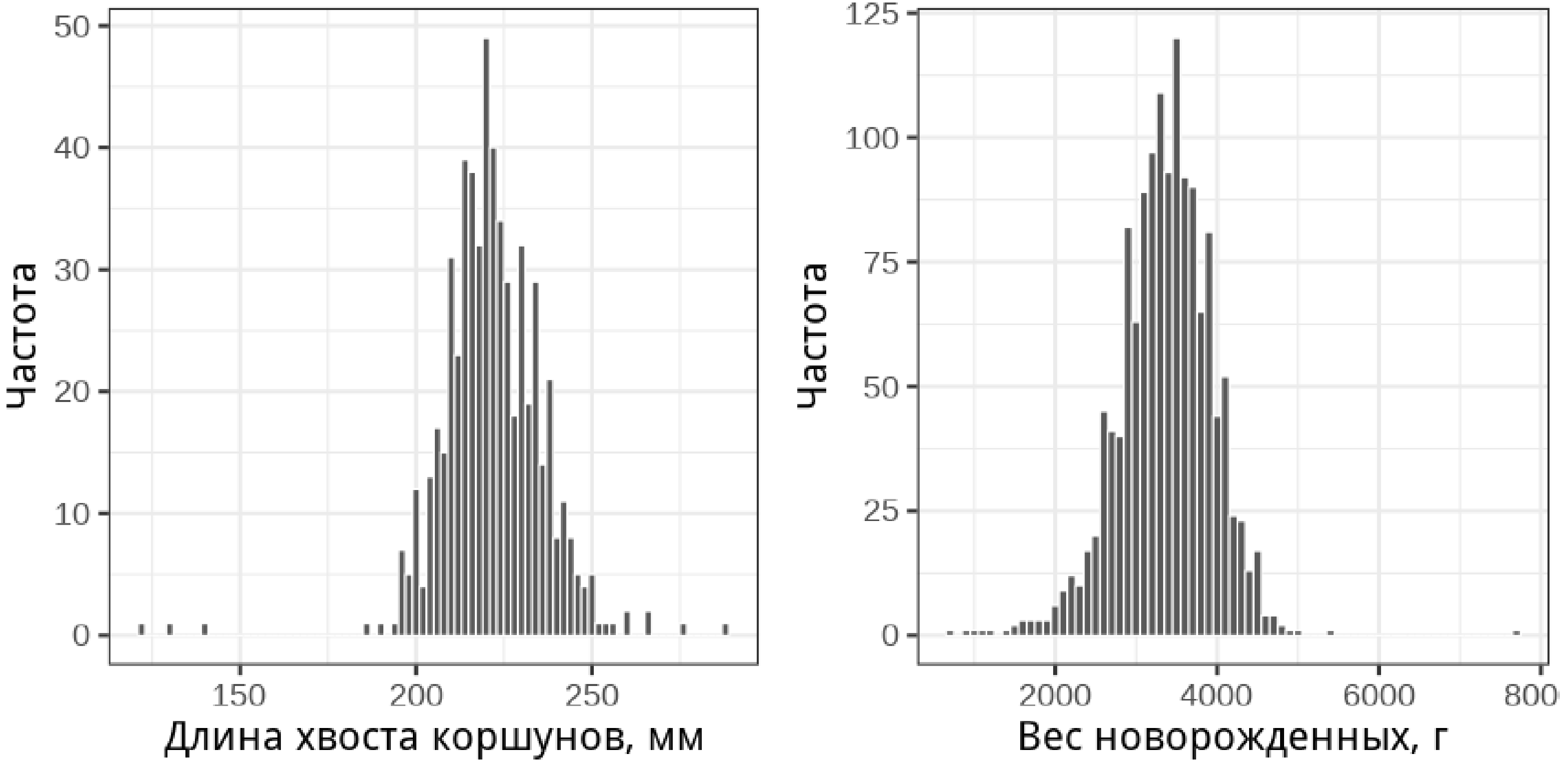
Обобщенное правило умножения вероятностей

$$P(A \; and \; B) = P(A) \cdot P(B|A) =$$
$$= P(B) \cdot P(A|B)$$

Нормальное распределение

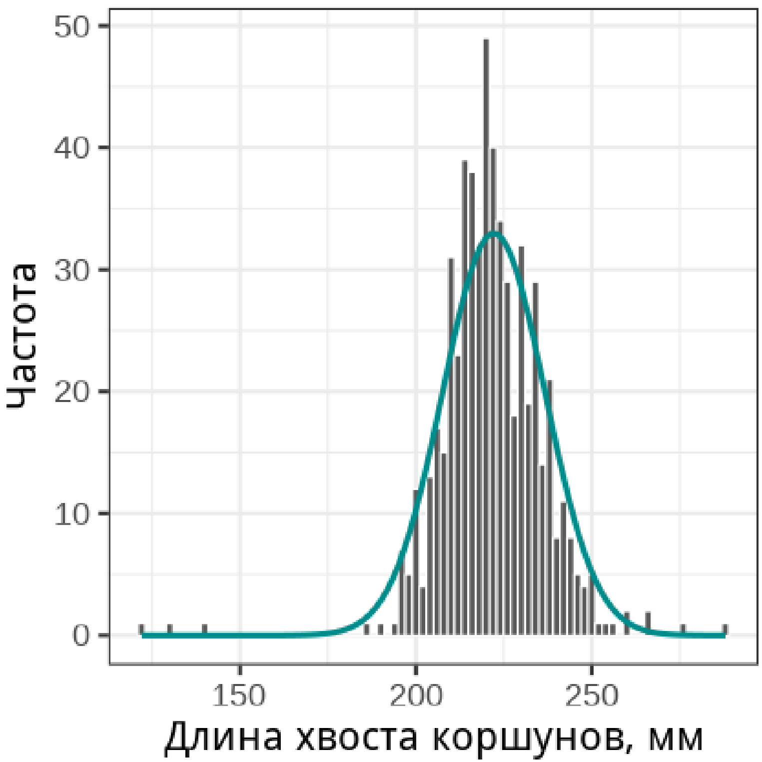


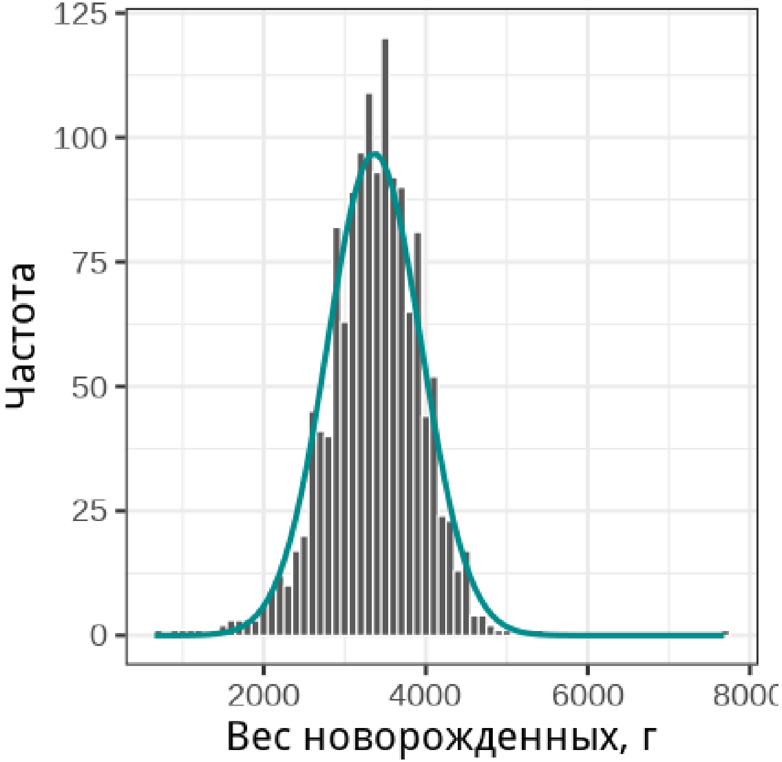
Частотные распределения мерных признаков



Частотные распределения многих мерных признаков имеют колоколообразную форму.

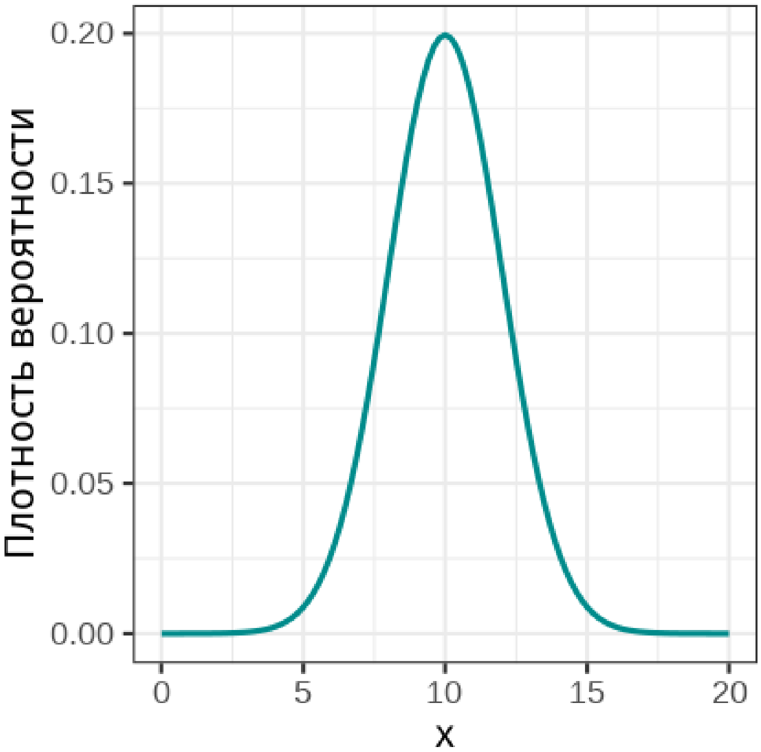
Нормальное распределение





Теоретическое распределение, которое описывает многие из таких колоколообразных кривых, называется **нормальное распределение**.

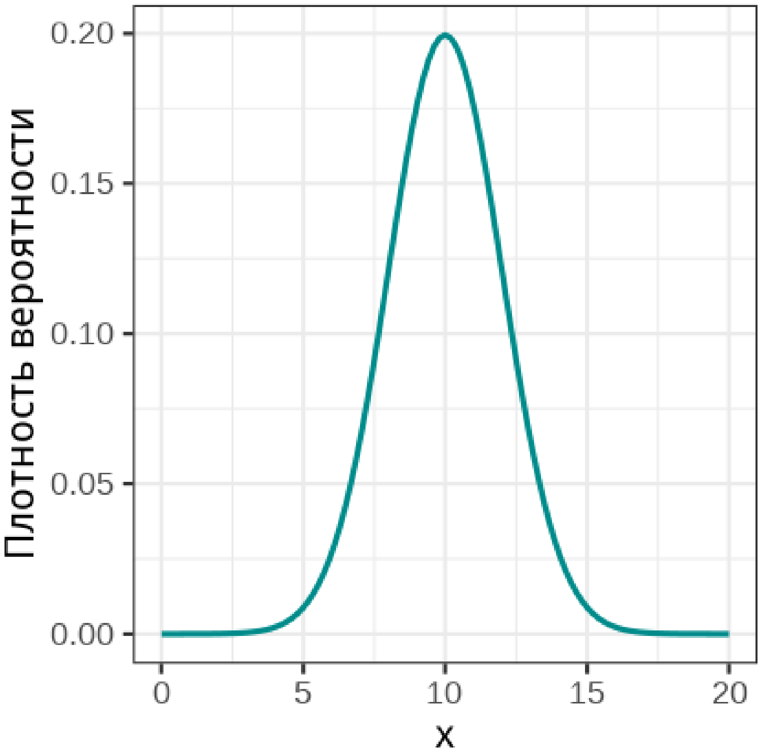
Нормальное распределение



Нормальное распределение

- Симметричное
- Унимодальное
- Непрерывное
- Наибольшая плотность вероятности — там, где среднее значение
- $-\infty \leq x \leq \infty$

Формула нормального распределения



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

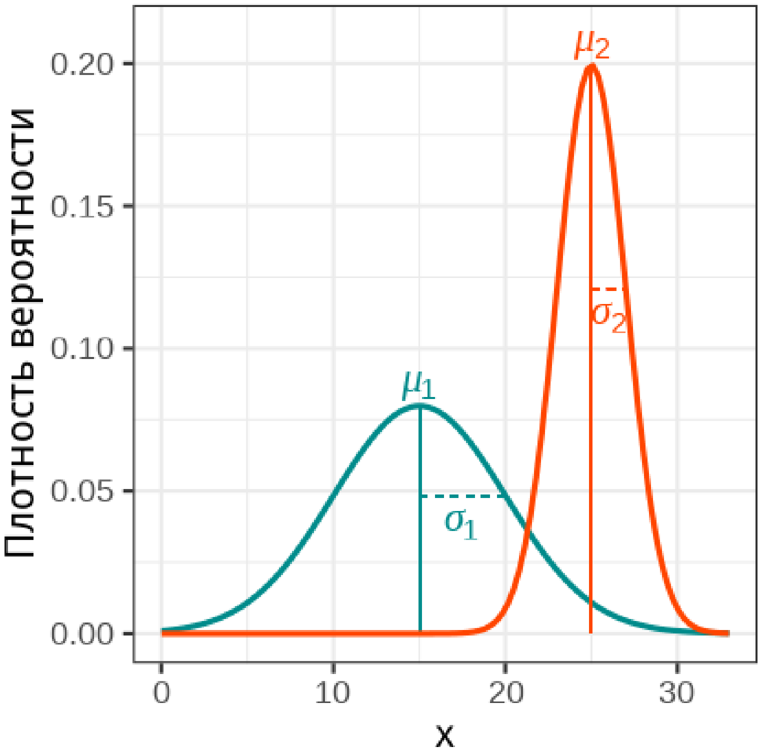
Параметры:

- μ — среднее значение
- σ — стандартное отклонение

Случайная величина x подчиняется нормальному распределению со средним μ и стандартным отклонением σ.

Это кратко записывается как $x \sim N(\mu, \sigma)$.

Параметры нормального распределения

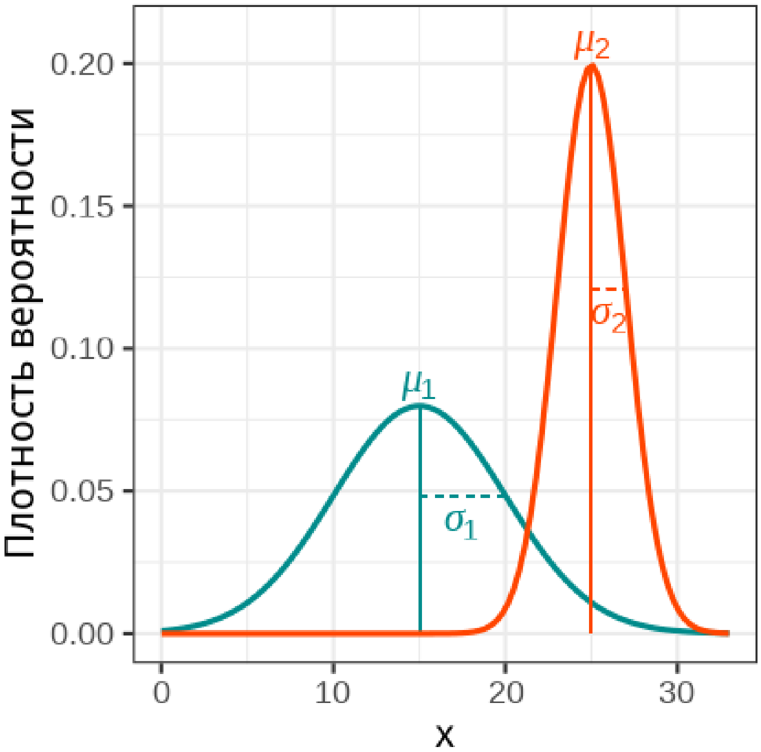


$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$

Параметры:

- μ — среднее значение — задает положение вершины по оси X
- σ — стандартное отклонение — задает размах кривой распределения

Параметры нормального распределения



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Параметры:

- μ — среднее значение — задает положение вершины по оси X
- σ — стандартное отклонение — задает размах кривой распределения

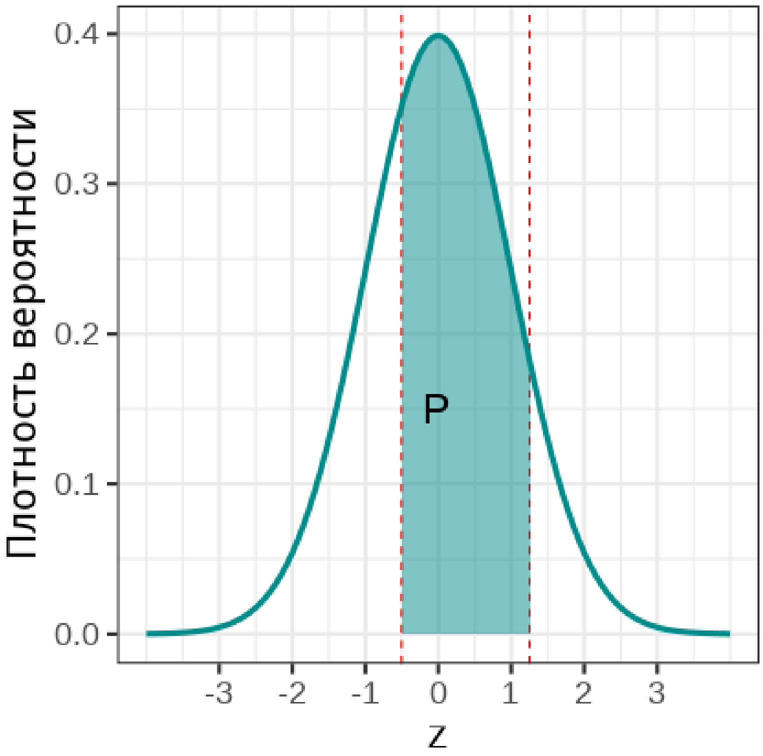
На рисунке распределения:

- $x_1 \sim N(\mu_1, \sigma_1)$
- $x_2 \sim N(\mu_2, \sigma_2)$

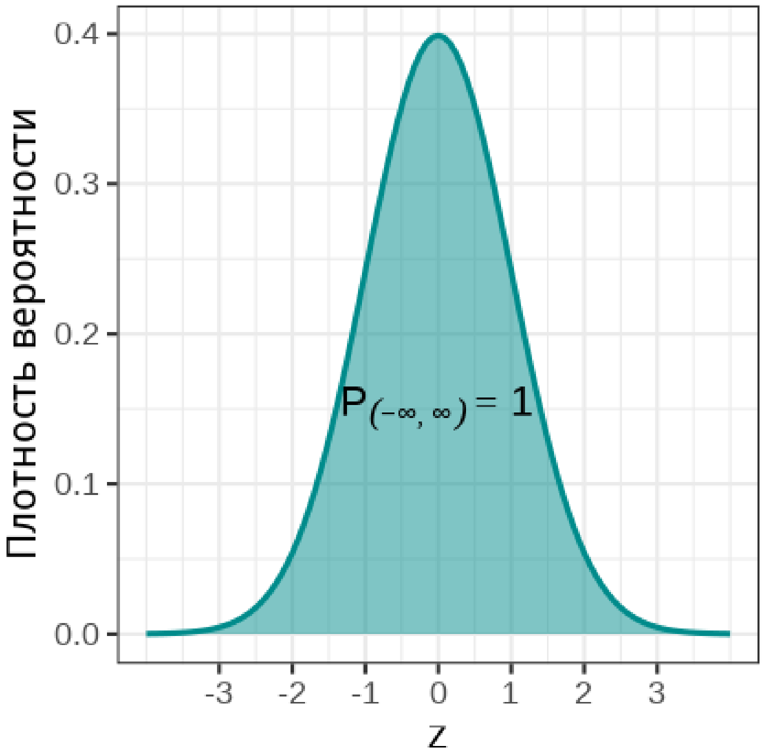
Параметры этих распределений:

- $\mu_1 < \mu_2$
- $\sigma_1 > \sigma_2$

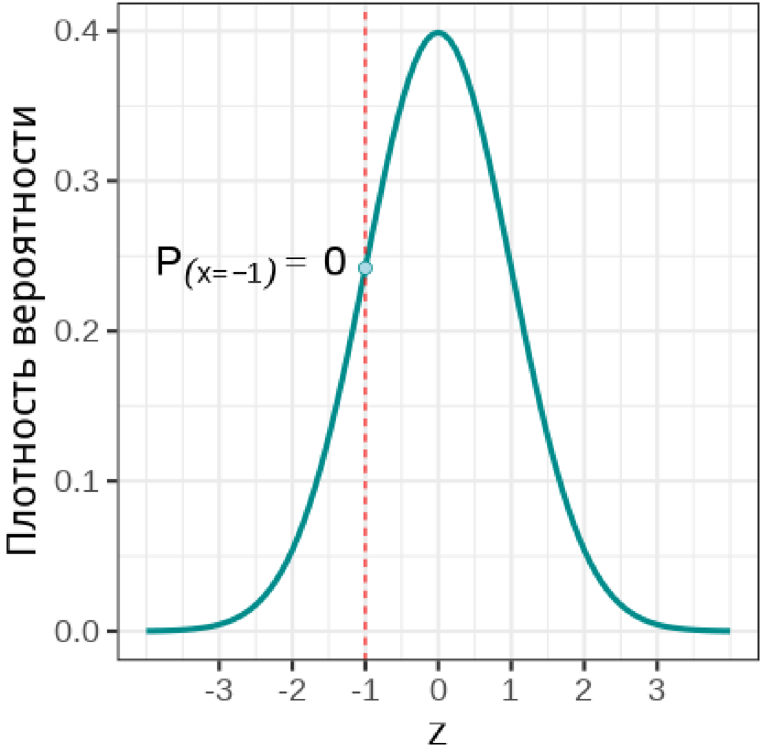
Кривые распределений можно использовать для оценки вероятностей



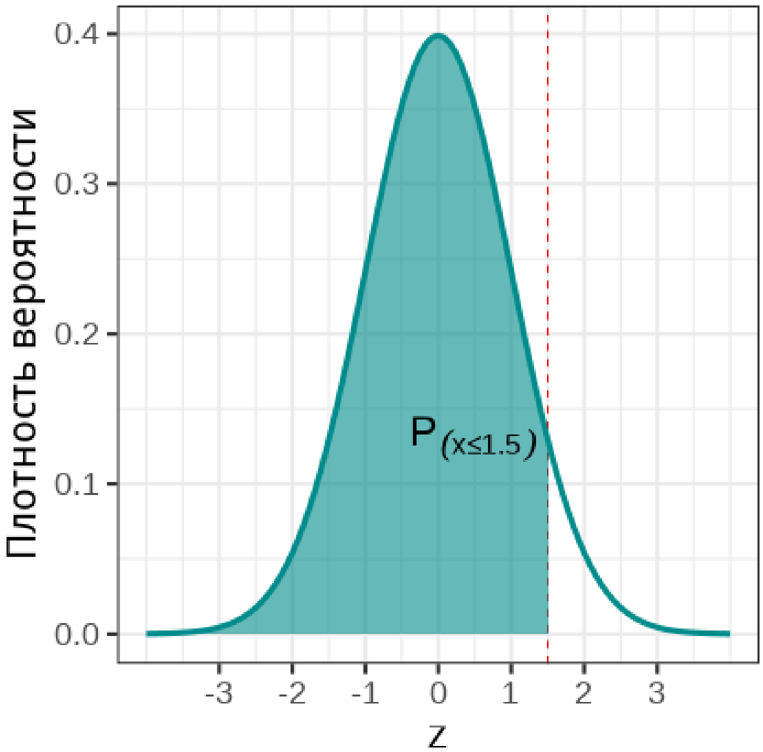
Площадь под всей кривой распределения равна 1



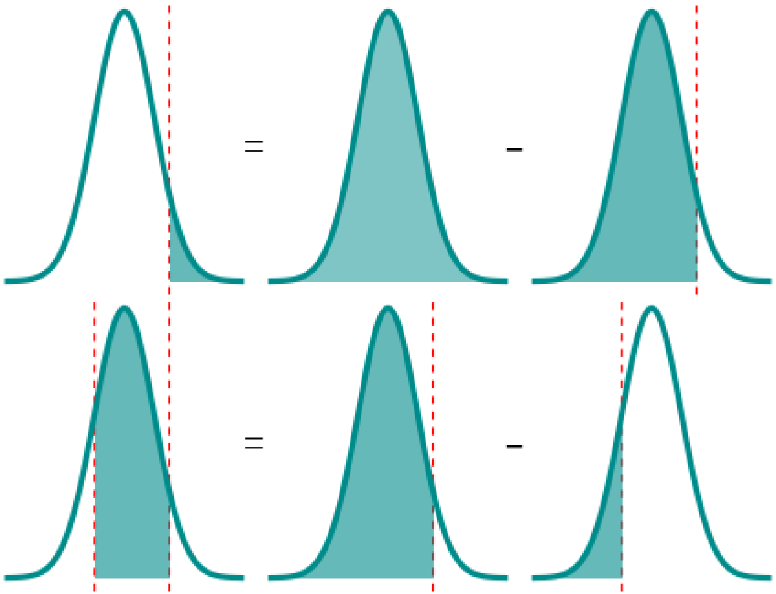
Вероятность конкретного значения нельзя определить



Можно определить вероятность того, что значение будет меньше заданного

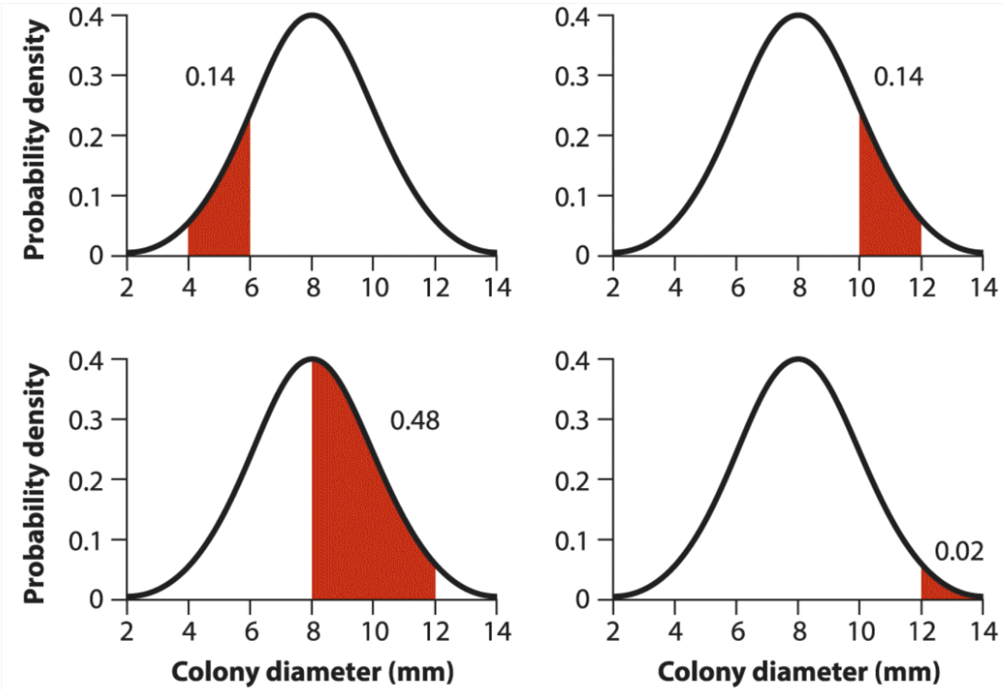


Остальные площади можно найти при помощи арифметических действий



Задание

Распределение диаметра колоний бактерий



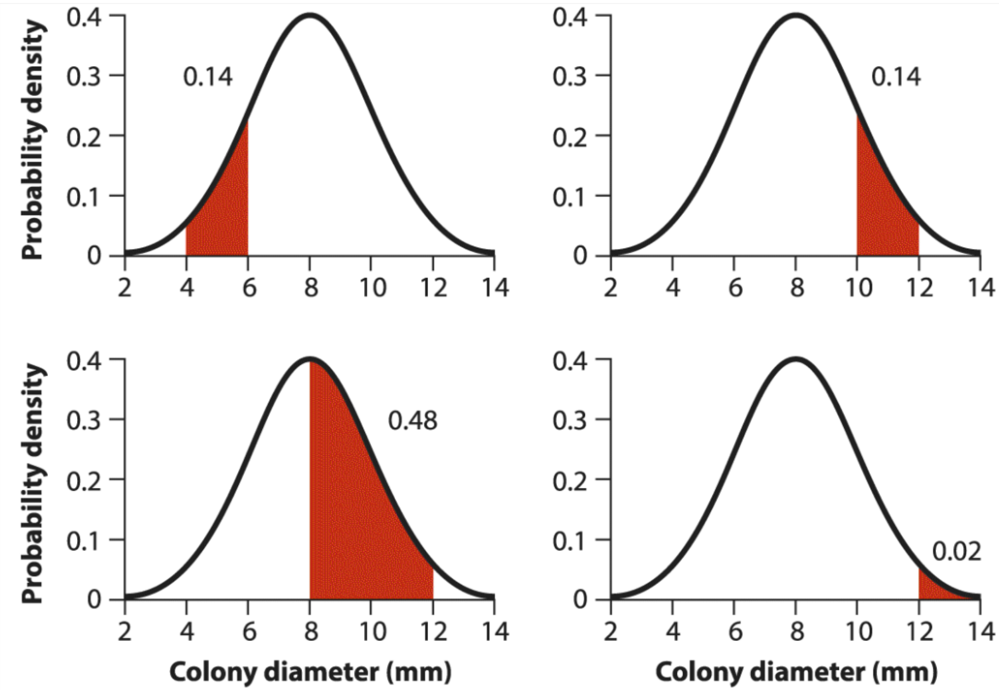
Какова вероятность, что диаметр случайно выбранной колонии будет лежать в заданных пределах?

- между 4 и 6 мм
- между 8 и 12 мм
- больше 10 мм
- между 8 и 10 мм
- меньше 6 мм
- меньше 4 или больше 12 мм

Whitlock and Shluter 2015, p.146

Решение

Распределение диаметра колоний бактерий



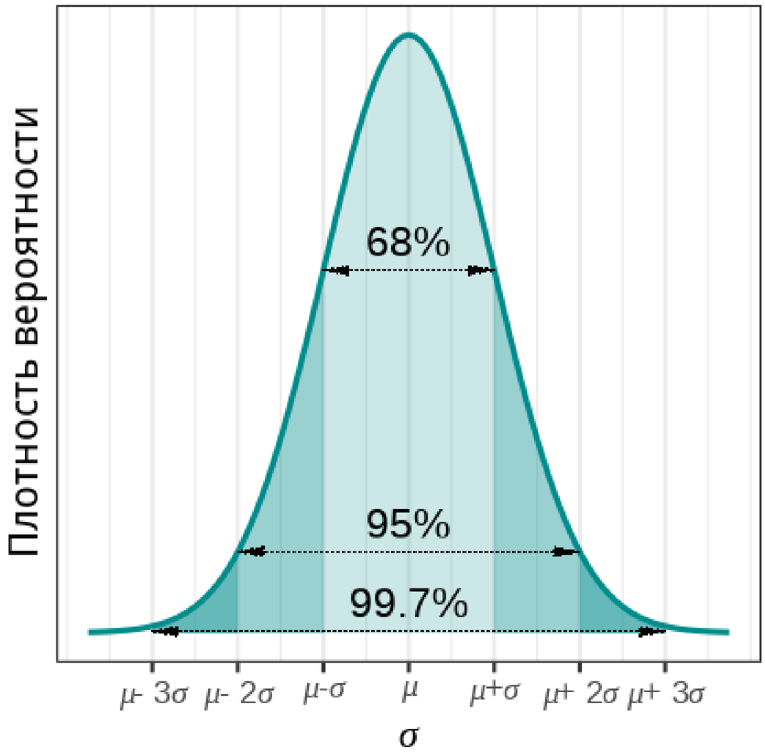
Какова вероятность, что диаметр случайно выбранной колонии будет лежать в заданных пределах?

- между 4 и 6 мм $P = 0.14$
- между 8 и 12 мм $P = 0.48$

- больше 10 мм
  $P = 0.14 + 0.02 = 0.16$
- между 8 и 10 мм
  $P = 0.48 - 0.14 = 0.34$
- меньше 6 мм
  $P = 0.14 + 0.02 = 0.16$
- меньше 4 или больше 12 мм
  $P = 0.02 + 0.02$

Whitlock and Shluter 2015, p.146

Площади под кривой нормального распределения



Правило 68–95–99.7 %

Если по оси x отложить стандартные отклонения от среднего значения, то окажется, что приблизительно

- $\sim$ 68% значений в пределах $1\ \sigma$
- $\sim$ 95% — в пределах $2\ \sigma$
- $\sim$ 99.7% — в пределах $3\ \sigma$

Summary

Summary: основные действия с вероятностями

- Вероятность события это доля случаев, когда это событие происходит в ряду повторных испытаний.
- Распредедение вероятностей описывает вероятности различных исходов случайного испытания.
- События называются несовместными, если не могут произойти одновременно.
- Суммарная вероятность всех несовместных событий равна 1.
- Вероятность того, что произойдет хотя бы одно из двух несовместных событий, равна сумме их вероятностей.
- Вероятность того, что произойдет одно из двух необязательно несовместных событий равна сумме их вероятностей за вычетом вероятности того, что они произошли оба.

Summary: зависимые и независимые события

- Деревья вероятностей помогают записать и вычислить вероятности цепочек событий.
- Условная вероятность события - это вероятность события, при условии, что уже произошло какое-то другое событие.
- Вероятность того, что произошли одновременно два независимых события, равна произведению их вероятностей.
- Вероятность того, что произошли два зависимых события, равна произведению полной вероятности одного из них на вероятность другого при условии, что первое произошло.
- Формула полной вероятности гласит, что вероятность события можно вычислить как сумму его вероятностей при условии каждого из несовместных событий.

Summary: нормальное распределение

- Нормальное распределение имеют многие мерные величины.
- Параметры нормального распределения задают положение его центра (среднее) и разброс (стандартное отклонение).
- Площади под кривой нормального распределения можно использовать для вычисления вероятности того, что значение нормально распределенной величины попадает в определенный диапазон, если известны параметры распределения этой величины в генеральной совокупности.

Что почитать

- Bluman, A. G. (2005). Probability Demystified (Vol. 1). McGraw-Hill Professional.
- Whitlock, M., & Schluter, D. (2015). The analysis of biological data (Second edition)**. Roberts and Company Publishers.