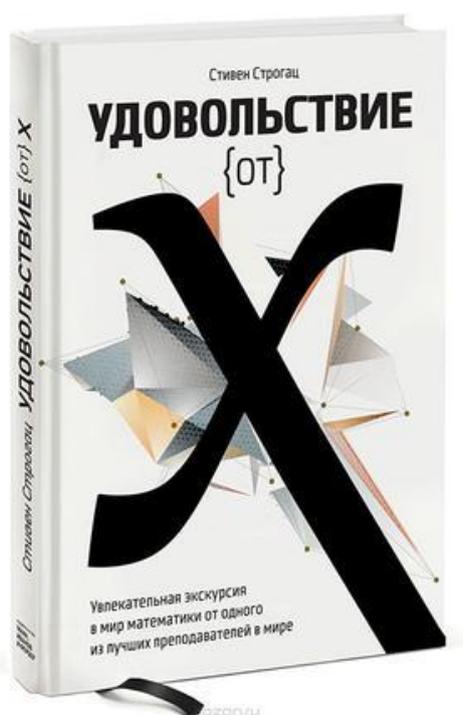


## Удовольствие от X: Новая нормальность. Глава из книги Стивена Строгаца

Отрывок из книги одного из лучших преподавателей математики



[Манн, Иванов и Фербер](#)

***Насколько полезны числа для изучения окружающего мира, в чем прелесть геометрии, насколько изящны интегральные счисления и важна статистика? Обо всем этом в своей книге «Удовольствие от X» рассказывает Стивен Строгац. Автор объясняет фундаментальные математические идеи просто и элегантно, приводя примеры, понятные каждому. Indicator.Ru публикует одну из глав книги, опубликованной в издательстве «Манн, Иванов и Фербер».***

Статистика внезапно стала сверхмодным направлением. С появлением Интернета, электронной торговли, социальных сетей, проекта по расшифровке генома человека, а также в связи с развитием цифровой культуры в целом мир стал захлебываться в данных. Маркетологи изучают наши вкусы и привычки. Разведывательные службы собирают информацию о нашем местонахождении, электронной переписке и телефонных звонках. Специалисты по спортивной статистике жонглируют цифрами, решая, каких игроков покупать, кого набирать в команду, а кого посадить на скамью запасных. Каждый стремится

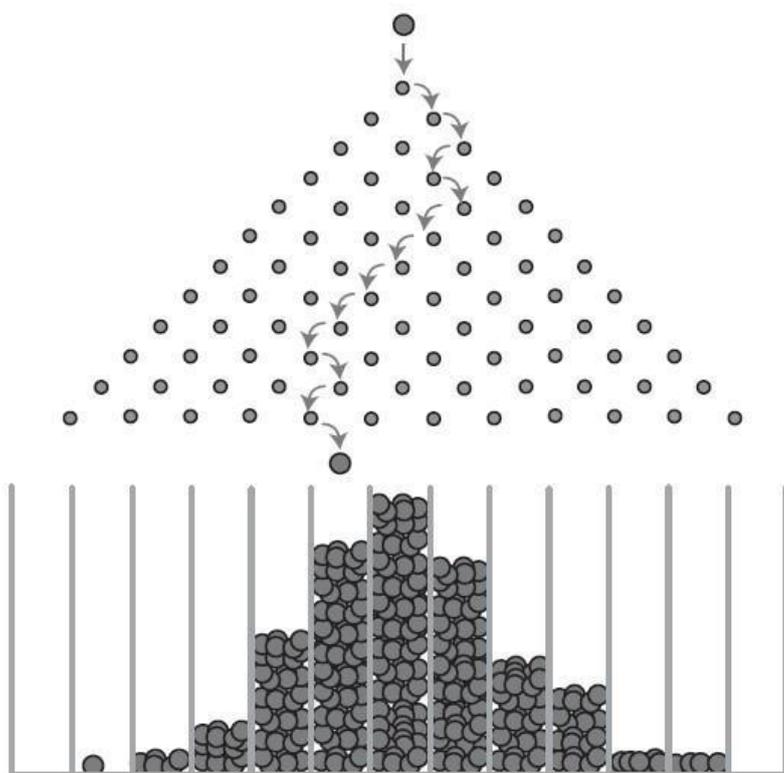
объединить точки в график и обнаружить закономерность в беспорядочном скоплении данных.

Неудивительно, что эти тенденции отражаются и в обучении. «Давайте обратимся к статистике», — увещевает в своей колонке газеты New York Times Грег Мэнкью, экономист из Гарвардского университета.

«В учебной программе по математике в средней школе слишком много времени уделяется традиционным темам, таким как евклидова геометрия и тригонометрия. Эти полезные для обычного человека умственные упражнения, однако, малоприменимы в повседневной жизни. Учащимся было бы гораздо полезнее больше узнать о теории вероятности и статистике». Дэвид Брукс идет еще дальше. В своей статье, посвященной дисциплинам, заслуживающим внимания для получения достойного образования, он пишет: «Возьмите статистику. Вот увидите, окажется, что знание того, что такое стандартное отклонение, вам очень пригодится в жизни».

Вполне вероятно, а еще неплохо разбираться в том, что такое распределение. Это первое, о чем я намерен поговорить. И хотел бы заострить на нем внимание, поскольку в этом заключается один из главных уроков статистики: вещи кажутся безнадежно случайными и непредсказуемыми при рассмотрении их по отдельности, однако в совокупности в них обнаруживается закономерность и предсказуемость.

Возможно, вы видели демонстрацию этого принципа в каком-нибудь научном музее (если нет, видеоролики можно найти в Интернете). Типичный экспонат представляет собой приспособление под названием доска Гальтона, которая чем-то напоминает автомат для игры в пинбол, только без флипперов. Внутри его с равными интервалами располагаются ровные ряды штырьков.



Доска Гальтона

Строгац С. Удовольствие от X. — М.: Манн, Иванов и Фербер, 2014.

Опыт начинается с того, что в верхнюю часть доски Гальтона запускаются сотни шариков. При падении они сталкиваются со штырьками и с равной вероятностью отскакивают то вправо, то влево, а затем распределяются внизу доски, попадая в отсеки одинаковой ширины. Высота столбика из шариков показывает, с какой вероятностью шарик может оказаться в данном месте. Большинство шариков размещаются примерно в середине, по бокам их уже меньше, и еще меньше — по краям.

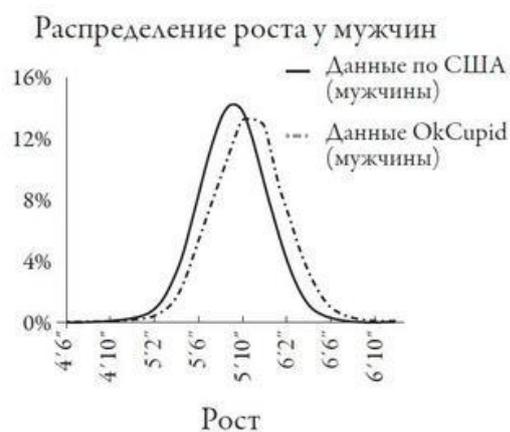
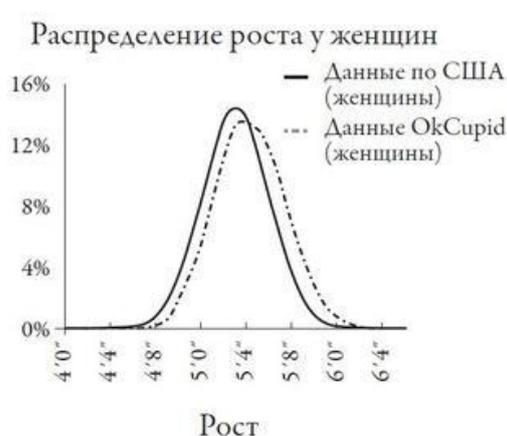
В общем, картина чрезвычайно предсказуема: шарики всегда образуют распределение в форме колокола, хотя предугадать, где окажется каждый отдельно взятый шарик, невозможно.

Каким образом отдельные случайности превращаются в общие закономерности? Но именно так действует случайность. В среднем столбике скопилось больше всего шариков потому, что, прежде чем скатиться вниз, многие из них совершат примерно одинаковое количество прыжков вправо и влево и в результате окажутся где-то посередине. Несколько одиноких шариков, расположившихся по краям, образуют хвосты распределения — это те шарики, которые при столкновении со штырьками отскакивали всегда в одном направлении. Такие отскоки маловероятны, поэтому по краям так мало шариков.

Подобно тому как местоположение каждого шарика определяется суммой множества случайных событий, многие явления в этом мире являются следствием множества мелких обстоятельств и тоже подчиняются колоколообразной кривой. По этому принципу работают страховые компании. Они с высокой точностью могут назвать количество своих клиентов, которые умирают каждый год. Однако не знают, кому именно не повезет на этот раз.

Или возьмем, к примеру, рост человека. Он зависит от бесчисленного количества случайностей, связанных с генетикой, биохимией, питанием и окружающей средой. Следовательно, велика вероятность, что при рассмотрении в совокупности рост взрослых мужчин и женщин будет представлять собой колоколообразную кривую.

В одном блоге под названием «Ложные данные, которые люди сообщают о себе в Интернете» статистическая служба сайта знакомств OkCupid недавно опубликовала график роста своих клиентов или, скорее, указанных ими значений. Обнаружилось, что показатели роста представителей обоих полов, как и ожидалось, образуют колоколообразную кривую. Однако удивительно то, что оба распределения были примерно на два дюйма смещены вправо относительно ожидаемых значений.



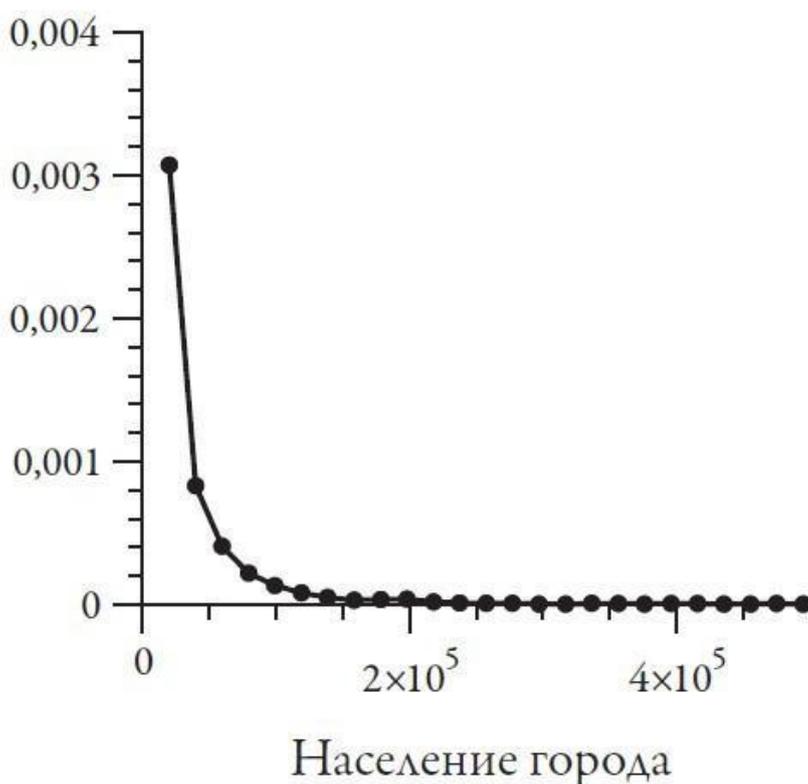
Таким образом, либо рост клиентов, опрошенных компанией OkCupid, превышает средний, либо при описании себя в Интернете они прибавляют к своему росту еще пару дюймов.

Идеализированной версией подобных колоколообразных кривых является то, что математики называют нормальным распределением. Это одно из важнейших понятий в статистике, имеющее теоретическое обоснование. Можно доказать, что нормальное распределение возникает при сложении большого количества мелких случайных факторов, причем каждый из них действует независимо от других. И многие события происходят именно таким образом.

Но не все. И это второй пункт, на который я хотел бы обратить внимание. Нормальное распределение не такое уж вездесущее, как кажется. На протяжении сотни лет, и особенно в последние несколько десятилетий, ученые и специалисты в области статистики отмечают существование множества явлений, отклоняющихся от этой кривой и следующих собственному графику. Любопытно, что подобные типы распределений практически не упоминаются в учебниках по элементарной статистике, а если и встречаются, то обычно рассматриваются как некие патологии.

Это странно. Я попытаюсь объяснить, что многие явления современной жизни приобретают больший смысл при условии понимания этих «патологических» распределений. Это новая нормальность. Возьмем, к примеру, распределение размеров городов в США. Вместо того чтобы скапливаться вокруг некоей средней величины колоколообразной кривой, подавляющее большинство городов имеют небольшой размер и, следовательно, скапливаются в левой части графика.

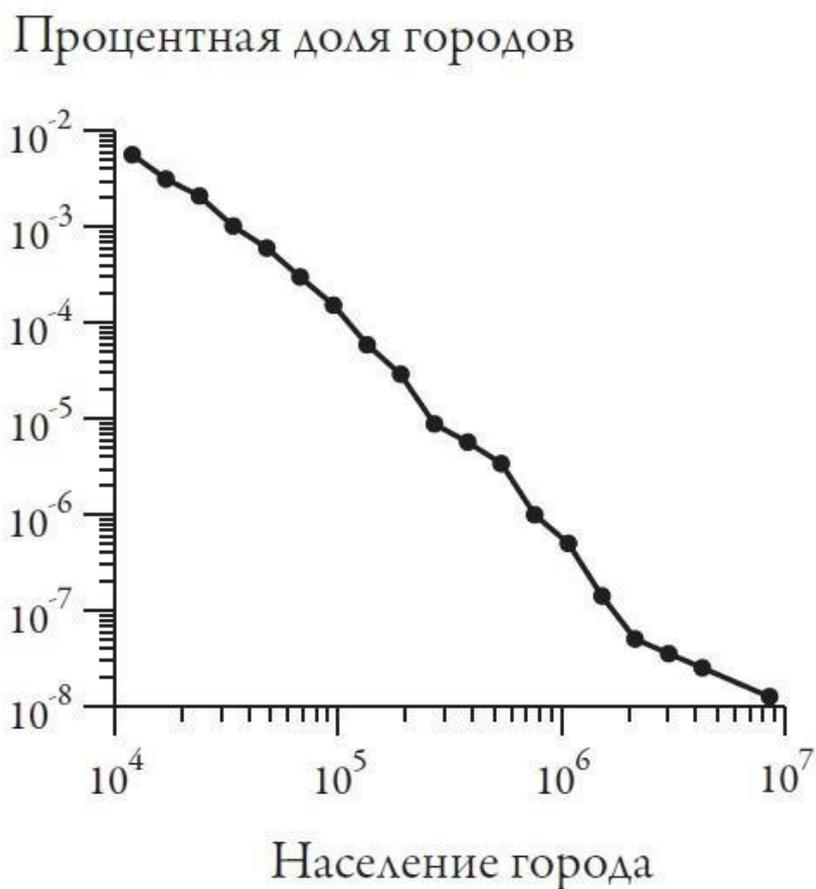
### Процентная доля городов



И чем больше население города, тем реже такие города встречаются. Иначе говоря, в совокупности распределение будет представлять собой скорее кривую в форме буквы L, чем колоколообразную кривую.

И в этом нет ничего удивительного. Все знают, что мегаполисов гораздо меньше, чем маленьких городов. Хотя это не так очевидно, размеры городов подчиняются простому красивому распределению — если посмотреть на них в логарифмическом масштабе.

Будем считать, что различие между двумя городами одно и то же, если их население отличается в одно и то же число раз (подобно тому как две любые клавиши рояля, отстоящие на октаву, всегда разнятся вдвое по частоте). И сделаем то же самое на вертикальной оси.



Строгац С. Удовольствие от X. — М.: Манн, Иванов и Фербер, 2014.

Теперь данные располагаются на кривой, представляющей собой почти идеальную прямую линию. Исходя из свойств логарифмов, нетрудно вывести, что исходная L-образная кривая представляет собой степенную зависимость, которая описывается функцией вида

$$y = \frac{c}{x^a}$$

где  $x$  — население города,  $y$  — количество городов, имеющих такой размер,  $c$  — константа, а показатель степени  $a$  (показатель степенной зависимости) определяет отрицательный наклон прямой линии.

Степенные распределения имеют некоторые нелогичные, с точки зрения традиционной статистики, свойства. Например, в отличие от нормального распределения, их моды, медианы и средние значения не совпадают из-за скошенной асимметричной формы L-образных кривых.

Президент Буш извлек из этого немалую пользу, заявив в 2003 году, что сокращение налогов позволило каждой семье сэкономить в среднем 1586 долларов. Хотя математически это верно, здесь он к своей выгоде взял за основу среднее значение вычета, под которым скрывались огромные вычеты в сотни тысяч долларов, полученные 0,1% богатейшего населения страны. Известно, что «хвост» в правой части распределения дохода следует степенной зависимости, и в подобной ситуации использование средней величины вводит в заблуждение, поскольку она далека от своего реального значения. В действительности большинству семей вернули менее 650 долларов. В данном распределении медиана значительно меньше, чем среднее значение.

Этот пример демонстрирует важнейшее свойство распределений степенной зависимости: они имеют «тяжелые хвосты» по сравнению по крайней мере с маленькими «жидкими хвостиками» нормального распределения. Подобные большие хвосты хотя и редкость, но встречаются чаще в распределениях данных, чем обычные колоколообразные кривые.

В «черный понедельник», 19 октября 1987 года, промышленный индекс Доу-Джонса упал на 22%. По сравнению с обычным уровнем нестабильности на фондовом рынке это падение составило более двадцати стандартных отклонений. Согласно традиционной статистике (в которой используется нормальное распределение), подобное событие практически невозможно: его вероятность составляет менее чем один случай на 100 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 000 (10 в 50 степени). Однако это произошло — поскольку колебания цен на фондовом рынке не соответствовали нормальному распределению.

Для их описания лучше подходят распределения с «тяжелым хвостом». Подобное происходит с землетрясениями, пожарами и наводнениями, что усложняет страховым компаниям задачу управления рисками.

Такая же математическая модель описывает число погибших в результате войн и террористических атак, а также другие, гораздо более мирные вещи, такие как количество слов в романе или число сексуальных партнеров у человека.

Хотя прилагательные, используемые для описания длинных хвостов, выставляют их в не слишком выгодном свете, «хвостатые» распределения гордо несут свои хвосты. Жирный, тяжелый и длинный? Да, это так. Но в таком случае покажите, какой нормальный?

.....

Теги [#Книжный клуб](#)