

Рис. 4.33. Ветер, генерируемый вентилятором, охлаждает электронную печатную плату. Воздух в вентиляционном отверстии (справа) немного теплее, чем у вентиляторов (слева)


4.5. ТЕСТИРОВАНИЕ

Исследуя данные и выявляя закономерности, мы часто разрабатываем теории, объясняющие, что именно наблюдаем, и многие из них оказываются ошибочными или уводящими в сторону. Как сказал космолог Карл Саган в своем последнем интервью, *«наука — это нечто большее, чем совокупность знаний. Это способ мышления, способ скептически исследовать вселенную с ясным пониманием того, что человек склонен ошибаться»*.

Может показаться неестественным подвергать сомнению собственные предположения и, как следствие, трудно убедительно объяснять закономерности. К счастью, ученые ценят скептицизм и разработали ряд полезных инструментов.

ГИПОТЕЗЫ

Чтобы ответить на вопрос или оценить интуитивное знание, не делая при этом неточных выводов, начните с правильной формулировки. Выразите это как **гипотезу**: утверждение, которое может быть поддержано или опровергнуто данными, которые вы можете собрать. Придется проверять гипотезу, поэтому пока *не* предполагайте, что она верна или ложна. Сделайте свою гипотезу простой и объективной, чтобы ее было легче проверить.

ДЕЛО ВКУСА  Ваша кофейня в Миннеаполисе теряет своих клиентов-веганов. Десять дней назад вы решили обновить меню и предложили зеленый чай и мороженое с кокосовым молоком. Вы сделали это без каких-либо исследований, основываясь на следующих интуитивных предположениях.





Клиенты-веганы жаждали зеленого чая в дни занятия йогой.



Клиентам не нравился старый выбор мороженого.

Теперь, когда новое меню готово, вы хотите провести исследование, чтобы определить, было ли такое изменение хорошим решением. Можете ли вы выразить каждое из своих интуитивных представлений в виде гипотезы? Попробуйте сформулировать такую гипотезу, которая относится к вашим целям и которую вы можете легко проверить.

Если говорить об интуиции , то трудно найти объективную меру для понятия «жаждать». Однако одной из целей вашего бизнеса должна быть продажа вашей продукции, поэтому вы можете выразить гипотезу следующим образом: *клиенты-веганы покупают зеленый чай в дни йогой*. Проверка этой гипотезы может помочь вам понять, сможет ли ваше решение вернуть лояльность клиентов-веганов. Однако это потребует определенного труда: вы должны провести опрос, чтобы определить, которые из ваших клиентов являются йогой-веганами.

«Интуицию » сложно проверять, так как опрос давних клиентов может привести к ошибке выборки: вы можете опросить только тех, кто вернулся и кого, скорее всего, вполне устраивало старое меню. Но здесь имеется тесная связь с вашей бизнес-целью в виде продажи продуктов, поэтому гипотеза может быть такой.



Продажи мороженого с новым меню стали выше.

Эта гипотеза не дает особого представления о том, *почему* вы достигаете или не достигаете своих целей в вопросе мороженого, но, по крайней мере, ее легко проверить: ваша система учета, вероятно, собирала данные о продажах как до, так и после изменения меню. У вас уже есть все необходимые данные! Используя среднее значение, подведем итоги и сравним последние 100 дней старого меню с первыми десятью днями нового (рис. 4.34).

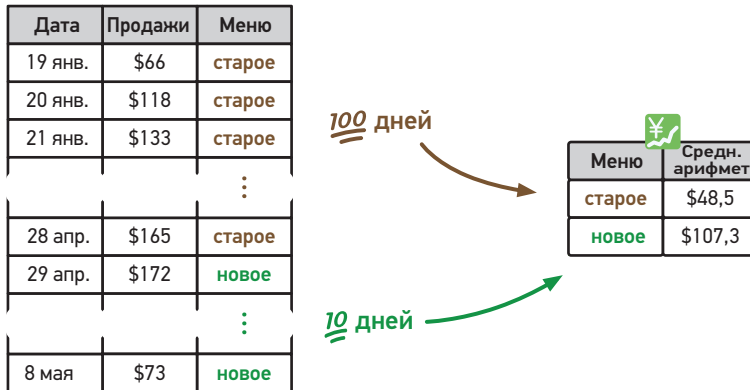



Рис. 4.34. Обобщение ежедневных продаж для проверки гипотезы 

Эти данные согласуются с гипотезой: продажи стали выше при использовании нового меню. Достаточно ли этих доказательств, чтобы подтвердить гипотезу о том, что новое меню увеличит продажи мороженого?

Конечно же нет. *Невозможно* прийти к заключению, сравнивая средние значения по этому небольшому набору данных. На данные могли повлиять многие другие факторы. Изменения в поведении клиентов могут быть вызваны другими изменениями, привнесенными в вашу кофейню, такими как, например, расположение столов. Они могут быть вызваны факторами, находящимися вне вашего контроля, такими как погода (рис. 4.35).

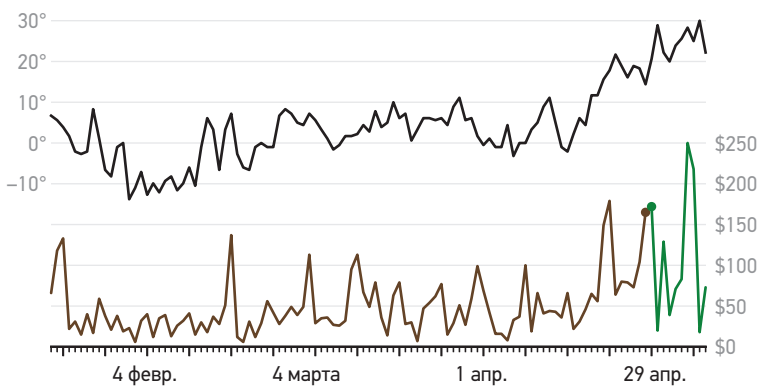



Рис. 4.35. Ежедневные пиковые температуры и продажи мороженого в вашей кофейне в Миннеаполисе. Температура показана сверху черным цветом (левая шкала), продажи мороженого показаны снизу цветом (правая шкала)

Независимо от вашего меню, клиенты наверняка съедят больше мороженого в жаркий день, чем во время снежного бурана. Возможно, именно более высокие температуры вызвали рост продаж, а вовсе не новое меню. Может быть, вам просто повезло. И если вы не можете быть уверены в том, что именно вызвало изменение продаж, то как можете узнать, отражает ли результат проверки гипотезы истину? Далее подробно рассмотрим методы, разработанные учеными для тщательного решения этой проблемы.

ЭКСПЕРИМЕНТЫ

Хорошие исследователи всегда разрабатывают процедуры, позволяющие уменьшить неопределенность вокруг их результатов. Научная процедура проверки гипотезы называется **экспериментом**. Гипотезы могут быть должным образом проверены только с помощью хорошо продуманных экспериментов.

Первым шагом в разработке эксперимента является определение переменных, которые необходимо измерить, а также данных, которые требуется собрать. Для гипотезы  этими переменными являются *продажи* и *меню*. Они уже представлены в виде столбцов на рис. 4.34.

Всегда есть и другие переменные, которые влияют на тех, кого вы хотите изучить. Мы называем их **внешними переменными**, и будьте внимательны: они вполне могут привести к неправильным выводам. Попробуйте собрать данные касательно внешних переменных, если они еще не включены в ваши записи. Если вы внимательно изучите рис. 4.35, то заметите, что пики продаж кажутся выше, когда погода теплее. Продажи довольно нерегулярны, но мы наблюдаем пик продаж каждые семь дней. Может быть, люди едят больше мороженого по выходным. Мы уже определили две внешние переменные: *температуру* и *день недели*.

Когда вы изучаете данные, продолжайте искать подсказки, чтобы идентифицировать различные внешние переменные. Когда вы их найдете, убедитесь, что записи соответствующим образом обновлены (рис. 4.36).

В хорошо продуманном эксперименте гипотеза проверяется с использованием записей, в которых дисперсия внешних переменных минимизируется. Но это может серьезно ограничить количество используемых

записей. Например, если мы рассмотрим только один день недели с аналогичной пиковой температурой, останутся только две записи данных (рис. 4.37).




| Дата | Продажи | Меню | День | Темп. |
|---------|---------|--------|------|-------|
| 19 янв. | \$66 | старое | Пт | 6,7° |
| 20 янв. | \$118 | старое | Сб | 5,6° |
| 21 янв. | \$133 | старое | Вс | 3,9° |
| | | ⋮ | | |
| 28 апр. | \$165 | старое | Сб | 14,4° |
| 29 апр. | \$172 | новое | Вс | 20,6° |
| | | ⋮ | | |
| 8 мая | \$73 | новое | Вт | 22,2° |

Рис. 4.36. Сбор данных о факторах, которые могут повлиять на продажи



| Дата | Продажи | Меню | День | Темп. |
|---------|---------|--------|------|-------|
| 22 апр. | \$179 | старое | Вс | 17,8° |
| 29 апр. | \$172 | новое | Вс | 20,6° |

Рис. 4.37. Сравнение продаж в аналогичных условиях

В этом эксперименте продажи были ниже с новым меню. Данные не подтверждают гипотезу , поэтому мы на шаг приблизились к ее отклонению. Но достаточно ли этого, чтобы ее отвергнуть?

Конечно же нет. Даже при разработке эксперимента в контролируемой среде мы никогда не в силах контролировать *всё*. Всегда есть факторы, влияющие на данные, некоторые из которых трудно идентифицировать. Могли ли вы знать, что ваши продажи увеличила группа евших мороженое на спор, а также непредвиденное число поломок морозильников? И наоборот, могли ли вы знать, что неожиданный всплеск аллергии на пыльцу удерживал людей дома, снизив потенциальные продажи? Мир

слишком сложен, чтобы даже при всем старании можно было вычлнить *все* внешние переменные.

ШУМ В некотором масштабе сложность нашей Вселенной всегда будет давать случайные флуктуации, независимые от любых внешних переменных, которые мы способны обнаружить. Такие случайные необъяснимые колебания в просторечии называются **шумом**. Когда доступно слишком мало точек данных, трудно понять, какие колебания являются шумом, а какие — нет, и истина становится размытой. В конечном счете чем больше у нас данных, подтверждающих гипотезу, тем ближе мы подходим к ее подтверждению. И наоборот, чем больше данных, не согласующихся с гипотезой, тем ближе мы подходим к ее отклонению. К сожалению, на рис. 4.37 есть только две точки данных, и их недостаточно, чтобы подтвердить или отвергнуть гипотезу. Нужно больше данных.

Поскольку это уменьшает неопределенность, может возникнуть соблазн продолжать бесконечный сбор данных. Давайте попробуем определить, *сколько* данных будет достаточно, чтобы удостовериться в том, что результат проверки гипотезы отражает истину.

Р-ЗНАЧЕНИЯ

Поскольку всегда имеется риск невезения и получения данных, подтверждающих ложную гипотезу, специалисты по статистике разработали методы оценки того, *насколько маловероятно*, что это произошло. Такой метод называется **статистической проверкой гипотезы** (рис. 4.38).



Рис. 4.38. Статистическая проверка гипотезы в действии

Вероятность того, что нам не повезло, называется ***p*-значением**. Чем меньше *p*-значение, тем ближе мы к подтверждению гипотезы. Если гипотеза верна, то сбор большего количества данных будет иметь тенденцию к снижению *p*-значения.

Существует множество различных типов проверки гипотез, каждый из которых лучше всего подходит для конкретного сценария. Чтобы *p*-значение было значимым, проведите исследование и убедитесь, что вы выбрали именно тот тип, который соответствует вашим данным и гипотезе¹. Все они работают одинаково: вы вводите экспериментальные данные, а они выводят *p*-значение для проверяемой гипотезы.

Вероятность — это всегда число от 0 до 1, и *p*-значение не является исключением. Получение $p = 0,2$ означает, что вероятность того, что ваши данные подтверждают неверную гипотезу из-за вашего невезения, составляет 20 %. Если вас не устраивает такой уровень риска, соберите больше данных. Запустите проверку еще раз, и если ваша гипотеза верна, *p*-значение должно уменьшиться. Будет ли этого наконец достаточно, чтобы подтвердить гипотезу с абсолютной уверенностью?

Неа. Значение *p* никогда не бывает равным 0 или 1. Другими словами, вы никогда не получите 0 или 100 % вероятности того, что гипотеза ложная. А еще будьте очень внимательны при интерпретации этих вероятностей, потому что результат проверки гипотезы по надежности не может превышать сами данные. Если данные страдают от ошибки выборки, то ваше *p*-значение не имеет смысла. Проверка гипотезы не заменяет контролируемые эксперименты, а лишь дополняет их.

СТАТИСТИЧЕСКАЯ ЗНАЧИМОСТЬ Даже притом, что *p*-значение никогда не может быть равным нулю, наступит момент, когда мы будем уже в состоянии сделать выводы. Обычно мы определяем максимально допустимое *p*-значение, которое должна дать наша проверка гипотезы, чтобы ее можно было бы считать *скорее всего* верной. Этот порог называется **уровнем значимости**. Когда *p*-значение ниже этого порога, можно утверждать, что имеются **статистически значимые** доказательства, подтверждающие нашу гипотезу. Ученые обычно работают с уровнями значимости от 0,05 до 5 %.

¹ Обзор распространенных проверок гипотез и способов использования каждой проверки см. в разделе <http://code.energy/hypothesis-test>.