



# ВВЕДЕНИЕ

Всем привет!

Меня зовут Вова Савельев. В 2016 году я написал книгу «Статистика и котики». К тому моменту я уже четыре года преподавал «Математические методы в психологии» в Уральском Федеральном Университете. Объяснять психологам, что такое дисперсия или t-критерий Стьюдента — нетривиальная задача. Поэтому я решил делать это на котиках. Потому что все любят котиков и с ними можно придумать много интересных примеров.

В апреле 2016 года я нарисовал первого котика. Выглядел он так.



Мне говорили, что он напоминает известного тогда Grumpy Cat. Но на самом деле я скопировал его со своей кошки Музы.

Как раз тогда у меня закончился первый поток студентов-психологов. И не слишком довольный результатами, я переделал лекции для второго потока, добавив туда котиков. Справедливости ради — средняя успеваемость не увеличилась. Но зато ходить на лекции стали чаще.

К осени я дописал черновик книги. А потом — открыл краудфандинг на Boomstarter, собрав чуть больше 300 тыс. руб. Этого хватило на печать первого тиража. С тех пор книга регулярно переиздается и, как говорят, пользуется популярностью.

Но время идет. Пока я пишу это введение, я уже четыре года работаю аналитиком в компании «Контур». За это время я начал программировать, научился работать с базами данных и проводить исследования для бизнеса. Я увидел, как те вещи, которые я преподавал студентам, можно применить к реальным задачам. И в этой книге я поделюсь этим с вами.

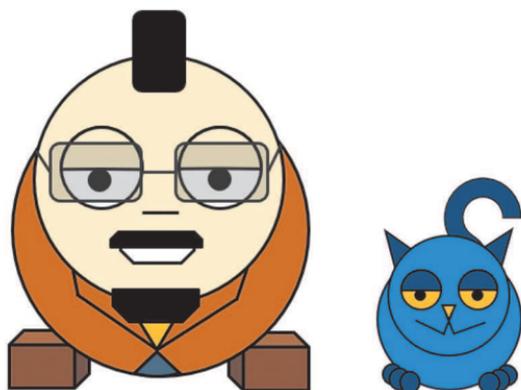
Первая книга была написана для студентов, желающих разобраться в статистике. Она — вариация университетского учебника, но облегченная и с забавными картинками.

Эта книга глубже. Я попытался не просто рассказать о методах, но и обозначить сферы их применения и пользу, которую вы можете получить. Кроме того, практически в каждой главе есть примеры вычислений с использованием языка R. Вы

можете просто скопировать их к себе на компьютер и посмотреть, как они работают.

Аудитория этой книги — маркетологи, специалисты по управлению персоналом, менеджеры. Люди, которым нужно принимать решения. Статистика помогает делать это лучше. И надеюсь, что после прочтения этой книги, вы научитесь использовать ее для решения своих бизнес-задач.

Желаю успехов!



# ГЛАВА 1

## ЗАЧЕМ СТАТИСТИКА БИЗНЕСУ?

**П**редставьте, что вы директор завода по производству кормов для котиков.

У вас на столе две новейшие разработки. Первая — корм со вкусом соленых огурчиков. Вторая — корм со вкусом болгарского перца. Денег немного — производить можно что-то одно.

Что из этого выбрать?



Можно выбрать корм, который вам больше нравится. Но не факт, что котикам он понравится тоже.

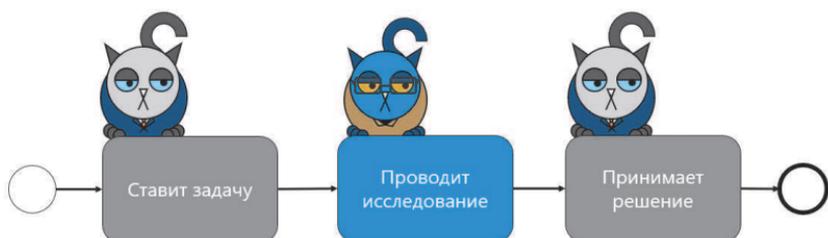
Можно кинуть монетку или положиться на интуицию. Но в этих случаях высок риск ошибиться. А в бизнесе это значит потерять деньги. Не угадали вкусы котиков, закупили ароматизаторов и вкусовых добавок на миллионы, а корм не продается. И вот вы в убытке.

Снизить риск ошибки можно если узнать, что нравится котикам. Чтобы это сделать, нужно провести исследование.

**ИССЛЕДОВАНИЕ (В ОБЩЕМ СЛУЧАЕ)** — комплекс мероприятий для получения достоверных знаний.

Как правило, исследованиями занимаются специальные сотрудники. Их называют исследователями или аналитиками.

Вы можете поставить задачу такому исследователю. Он узнает какие корма нравятся котикам и расскажет вам. И тогда вы сможете принять хорошее решение.



Работа исследователя состоит из четырех этапов:

1. Исследователь ставит цели и решает, как именно он будет проводить исследование. Также он находит нужные ресурсы — деньги, людей, оборудование. Это этап подготовки исследования.

2. Затем он собирает данные. Например, проводит опрос среди котиков. Или выпускает пробные партии кормов и смотрит как они продаются. Или читает результаты схожих исследований.

3. Как только он соберет достаточно данных, он садится за их обработку.

4. После обработки данных он оформляет результаты исследования в виде отчетов и презентаций.



Статистика в основном нужна на этапе обработки данных и чуть-чуть при подготовке исследования. В большинстве случаев без нее нельзя быть уверенным в достоверности полученных результатов. Почему так — поговорим в следующей главе.

Итак, подведем итоги:

1. Статистика — инструмент для проведения исследований.

2. Исследования нужны для получения достоверных знаний.

3. Достоверные знания нужны чтобы принимать правильные решения.

4. Правильные решения ведут к процветанию бизнеса.

## ГЛАВА 2

# ВЫБОРКА И ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ

**П**редположим, мы проводим маркетинговое исследование, чтобы определить какой корм нравится котикам (с перцем или с огурчиками).

На первый взгляд все просто:

1. Возьмем группу котиков.
2. Дадим им попробовать оба варианта.
3. Попросим оценить каждый по 10-балльной шкале (0 — отвратительно, 10 — восхитительно).
4. Посчитаем средний балл по каждому варианту.



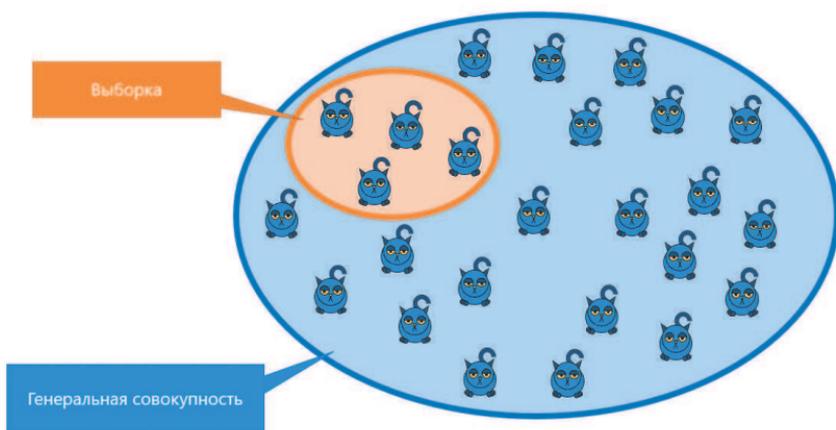
Кажется очевидным, что корм с наивысшим средним баллом больше нравится котикам и именно его надо производить. Но действительно ли это так? Давайте разбираться.

Вспомним, что наша цель — выбрать вариант, который нравится котикам. И мы хотим, чтобы результат был верен для всех существующих котиков. Все котики представляют собой генеральную совокупность.

Однако опросить всех котиков нереально — их слишком много. Поэтому для исследования мы берем их небольшую часть — выборку.

**ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ** — все объекты, о которых мы хотим получить некоторое знание.

**ВЫБОРКА** — часть генеральной совокупности, на которой мы проводим исследование с целью получить это знание



Итак, мы проводим исследования на выборке, а затем переносим его результаты на генеральную совокупность. Собственно статистические методы помогают сделать это правильно. Но чтобы они работали, выборка должна соответствовать двум требованиям.

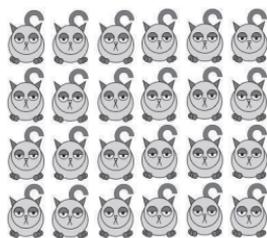
Начнем с того, что выборка должна быть похожа на генеральную совокупность. Степень похожести называется репрезентативностью.

**РЕПРЕЗЕНТАТИВНОСТЬ** — соответствие характеристик выборки и генеральной совокупности.

Если генеральная совокупность — котики, живущие в России, то некорректно набирать выборку из котиков, живущих во Франции. У французских котиков могут быть особые вкусовые предпочтения, разительно отличающиеся от российских.



Генеральная совокупность  
российских котиков



Французские  
котики

Также плохо набирать выборку только из котиков, живущих в Нижнем Тагиле. Тагильская специфика может быть нехарактерной для всей России.

К слову, подобная ситуация возникает, когда исследователи не хотят напрягаться и отбирают тех, кто находится под рукой. Такие выборки называют «удобными»\*.



«Удобная» выборка

Таким образом, если ваша генеральная совокупность — котики со всей России, то выборка должна делаться по всей России. А если вы хотите распространить выводы на весь мир, то и выборку набирайте со всего мира.

---

\* Хрестоматийный пример удобной выборки — психологическое исследование, проведенное на студентах-психологах. Невероятное число знаний было получено из этого неисчерпаемого источника. К сожалению, авторы не имеют привычки оговаривать, что полученные ими результаты нельзя распространять ни на кого, кроме этих самых студентов.