

Как лгать при помощи статистики

Материал из Википедии — свободной энциклопедии

Как лгать при помощи статистики — книга, написанная Дареллом Хаффом в 1954 году. Она рассказывает о различных способах злоупотребления статистикой в целях обмана аудитории и манипулирования её мнением. Рассмотрено множество конкретных примеров, в основном из американской жизни (реклама, политика, пропаганда и агитация).

Первым эпиграфом к книге выбрана цитата из графа Бйконсфилда (Б. Дизраэли) о статистике: *«Существуют три вида лжи: ложь, наглая ложь и статистика»*.

Книга ориентирована на читателя-неспециалиста и снабжена яркими иллюстрациями. Материал излагается живо и в доступной форме, что обеспечило высокую популярность книги — она является одной из самых многотиражных публикаций, посвящённых статистике, за вторую половину XX века^[1].

Как лгать при помощи статистики	
Автор	Дарелл Хафф
Язык оригинала	английский
Оригинал издан	1954
Оформление	Ирвинг Гейс
Издатель	W.W. Norton & Company Inc
Страниц	142
ISBN	0-393-31072-8

Содержание

Содержание

- Выборка изначально необъективна
- Грамотно выбранное среднее
- Нюансы, о которых скромно умалчивают
- Много шума практически из ничего
- График — лучше не бывает
- Схематичная картинка
- Псевдообоснованная цифра
- Снова «после — значит вследствие»
- Как производить статистикуляции
- Как поставить статистика на место

Издания на других языках

- На русском языке

Примечания

Ссылки

Содержание

Выборка изначально необъективна

Объяснение того, что такое выборка, как интервьюеры неосознанно выбирают опрашиваемых и влияют на их ответы.

Грамотно выбранное среднее

Рассматриваются виды среднего:

- Среднее арифметическое
- Медиана
- Мода

На примерах показано, как выбор вида среднего значения влияет на его величину для одних и тех же выборок. Обращается внимание на возможность манипулирования неподготовленным читателем путём выбора «удобного» (для манипулятора) вида среднего.

Нюансы, о которых скромно умалчивают

В этой главе автор рассматривает важные нюансы статистического исследования, которые часто намеренно или неосознанно не сообщают в статьях, предназначенных для широкой публики.

Объясняется важность размера выборки и связь её с видом совокупности. Приводятся примеры манипуляций с размером выборки:

- **Испытания эффективности зубной пасты.** Проводятся лабораторные испытания эффекта использования зубной пасты на шести испытуемых. Иногда проводится ряд таких исследований, и исследование, показавшее выгодный заказчику (производителю пасты) результат, используется в рекламных кампаниях.
- **Тест противополиомиелитной вакцины.** 450 детей были привиты, 680 непривиты (контрольная группа). Вскоре после этого в местности произошла эпидемия, у привитых детей не было ни одного случая заражения полиомиелитом. Как и ни у одного из членов контрольной группы. Эксперимент был лишён смысла с самого начала из-за неверного выбора количества участников, так как в группе такого размера можно ожидать не более двух случаев заражения.

Вводятся понятия:

- Статистическая значимость
- Доверительный интервал
- Доверительная вероятность

На примере шкалы развития Гесселя рассказывается об опасности восприятия точечной (внеинтервальной) оценки среднего значения — родители начинают паниковать, если их ребёнок не соответствует норме (среднему значению).

Обращается внимание на важность формулировок в статьях, базируемых на статистике. В качестве примера рассматривается заявление электроснабжающих компаний (1948 г.): «Электроэнергия доступна более чем 3/4 ферм США». Слово «доступна», никак не определённое в заявлении, обесмысливает его — обычно это означает, что ЛЭП находится на расстоянии 10-

100 миль (16-160 км) от фермы, но под доступностью можно понимать и что-то иное. Автор также замечает, что при интерпретации тех же самых данных можно поставить противоположный акцент, написав: «Электроэнергия недоступна четверти ферм США».

Завершается глава напоминанием о важности наличия чисел на графиках — график роста прибыли компании за несколько лет (опубликованный в журнале Fortune) не говорит читателю ничего, поскольку ось ординат лишена числовых меток. По такому графику нельзя сказать, был ли рост прибыли значительным, средним или близким к нулю.

Много шума практически из ничего

Понятия доверительного интервала и доверительной вероятности иллюстрируются примерами из реальной жизни:

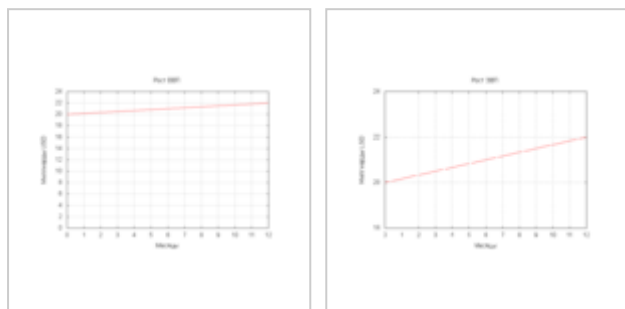
- Разница в результатах теста IQ между 98 и 101 не позволяет сказать, какой из испытуемых обладает более высоким показателем IQ, что видно из полной записи результатов теста: 98 ± 3 и 101 ± 3 соответственно.
- Измерение содержания вредных вещества в различных марках сигарет выявило отсутствие какой-либо существенной разницы между ними. Тем не менее, одна из марок оказалась на последнем месте по содержанию вредных веществ (пусть и с пренебрежимым отрывом от первого места!). Производитель этих сигарет (Old Gold) запустил рекламную кампанию, в которой утверждал, что сигареты Old Gold содержат меньше всего вредных веществ по данным независимой лаборатории.

График — лучше не бывает

Первая из глав, посвященных рассмотрению способов манипулирования при помощи графической информации.

Рассматривает способы искажения восприятия графиков:

- «Сжатие» части координатной сетки якобы в целях экономии места. В реальности приводит к затруднениям в восприятии масштаба



Полная версия "Сжатие" части
графика координатной сетки

- Изменение масштаба по осям абсцисс и ординат. Этот метод позволяет «превратить» (визуально) рост, близкий к нулю, в ярковыраженный устойчивый. В качестве примера приводится график роста правительственных субсидий, приведенный в одном из объявлений. Рост составил лишь 4 %, но визуально выглядел почти как 400 % за счет увеличения масштаба по оси ординат.

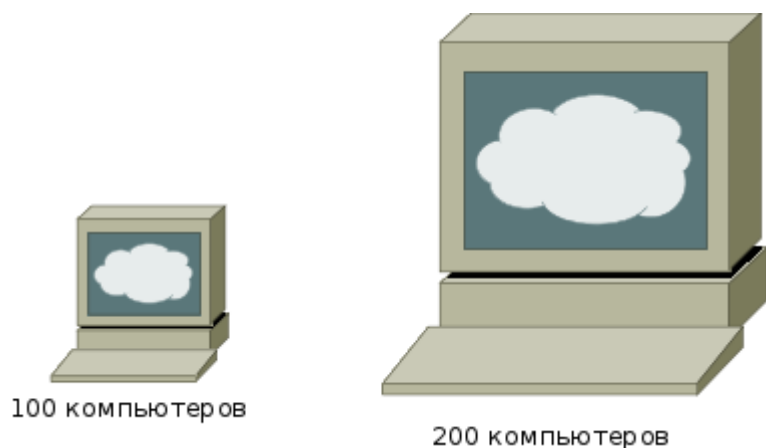
Схематичная картинка

Разбираются способы обманывания аудитории при помощи инфографики.

Использование графических объектов, ассоциативно связанных с представляемой информацией, открывает широкие возможности для злоупотреблений. Это утверждение поясняется рядом примеров:

- Для сравнения двух заработных плат можно использовать инфографику и нарисовать два мешка с деньгами. Если вторая заработная плата выше первой в два раза, то второй мешок будет не только выше, но и шире в два раза (что необходимо для сохранения пропорции). А поскольку мешок — это трёхмерный объект, то и контур второго мешка будет в два раза толще, чем первого. В результате наше зрение воспринимает второй мешок, как мешок в 8 (а не в 2!) раз больший, чем первый. Этот прием был использован журналом Newsweek.
- В рекламе американского института стали и сплавов была использована инфографика для показа увеличения выплавки стали между 1930 и 1940 гг на 4.25 млн тонн (с 10 млн до 14,25 млн). Техника инфографики (к рассмотренным ранее методам добавилось преднамеренное искажение пропорций) привело к тому, что указанное увеличение выплавки визуально воспринималось как 1500 %. Автор замечает, что это тот случай, когда «арифметика превращается в фантазию».
- Использование изображения коров разного размера для показа различных показателей удоев по годам. В добавление к уже рассмотренным эффектам, этот метод приводит к еще одному недоразумению — читатель может подумать, что не только удои стали выше, но и коровы больше.

На рисунке ниже показан пример злоупотребления инфографикой — второй объект визуально больше в 8 раз:



Псевдообоснованная цифра

Главу открывает язвительная рекомендация, — «Если вы не можете доказать то, чего хотите, то продемонстрируйте нечто другое и претендуйте на то, что эти вещи — одно и то же».

Приводится множество примеров такого жульничества. В частности:

- Выдача опроса *мнений людей* о равенстве возможностей белых и афроамериканцев устроиться на работу за реальную ситуацию на рынке труда. Этот опрос может показать тем лучшие результаты, чем больше в нём участвует людей с расовыми

предвзвещениями в отношении афроамериканцев, поскольку такие респонденты склонны считать, что никакой расовой дискриминации на рынке труда нет.

- Реклама сигарет использовала следующий аргумент: «более 27 % из большой выборки известных врачей курят Throaties, что является наивысшим показателем среди всех марок сигарет». Реклама неявно предполагала, что доктора знают нечто особенное, неизвестное остальным, о вреде, который причиняется сигаретами различных марок. Но это не так.
- Реклама соковыжималки утверждала, что лабораторные исследования показали, что она выжимает на 26 % больше сока. Когда был задан вопрос, — «чем что?», был получен ответ — «чем ручная конусная соковыжималка». Даже если исходное утверждение верно, оно не позволяет сравнить рекламируемое изделие с товарами конкурентов. Оно может являться и худшим на рынке, но тем не менее на 26 % превосходить ручную соковыжималку по эффективности отжима.
- Смертность на флоте во время Испанско-американской войны составляла 0,09 %. У гражданских же в Нью-Йорке в тот же период она была 0,16 %. Флот использовал эти цифры для агитации за службу в армии. Но эти показатели некорректно сравнивать — на флот набирают молодых и здоровых людей, а в гражданское население входят младенцы, старики, больные люди.

Снова «после — значит вследствие»

Эта глава рассказывает о понятии корреляции и часто возникающей путанице между причиной и следствием. Если явление А и В встречаются вместе, то это можно объяснить тремя способами:

1. Явление А является следствием явления В
2. Явление В является следствием явления А
3. Явления А и В являются следствием другого/других явления/явлений

Приводится ряд примеров ошибочных суждений о причинно-следственных связях. В частности:

- Исследования показали, что среди курящих студентов больше плохо успевающих, чем среди некурящих. Этот факт использовался в антитабачной кампании. Но из этого результата нельзя заключить, что курение отрицательно влияет на способности студентов. Возможно, что студенты начали курить из-за плохой успеваемости или же они плохо учатся и курят из-за какой-то третьей причины (например, тяжелых жизненных условий).
- Исследования показывают положительную корреляцию между уровнем образования и доходом. Из этого факта нельзя заключить, что если вы (ваш сын, дочь и т. д.) получаете высшее образование, то они непременно и обязательно будут иметь более высокий доход, чем если бы они его не получили. Более того, эта корреляция не позволяет вывести как всеобщее правило, что именно высшее образование ведет к более высокому доходу — возможно, лица, получившие его, происходят из богатых семей и именно поэтому получают более высокий доход в зрелом возрасте. Здесь рассмотрен пример ошибки post hoc. См. также Логические ошибки.
- Исследование 1500 типичных выпускников университетов среднего возраста показало, что 93 % мужчин были женаты (для всего населения этот показатель составил 83 %), а из женщин были замужем лишь 65 %. Из этого делается вывод, что образованная женщина имеет меньше шансов выйти замуж, чем необразованная. Но исследование не показывает причинно-следственную связь между этими явлениями. Возможно, эти незамужние женщины остались бы незамужними и в том случае, если бы не закончили университет.

Глава заканчивается почти анекдотическим (но реальным) примером перепутывания причины и следствия аборигенами Новых Гебрид. Они полагали, что наличие вшей ведёт к здоровью. Этот вывод делался на том основании, что больного человека вши покидали (так как вследствие повышенной температуры тела условия существования для них становились некомфортными), тогда как у всех здоровых людей они были (иными словами, наблюдалась положительная корреляция между здоровьем и наличием вшей).

Как производить статистикуляции

Статистикуляции — статистические манипуляции. В этой главе автор ещё раз на конкретных примерах показывает способы манипуляции статистическими данными. Однако он призывает не отвергать огульно статистические данные, но вдумчиво, основательно с осторожной недоверчивостью изучать их, прежде чем принять к сведению.

Как поставить статистика на место

Автор предлагает проверять статистические данные с помощью пяти простых вопросов:

- Кто это говорит?
- Откуда ему это известно?
- Чего не хватает?
- Не подменен ли объект исследования?
- Есть ли в этом смысл?

Издания на других языках

На русском языке

- *Дарелл Хафф*. Как лгать при помощи статистики = *How to Lie with Statistics*. — М.: Альпина Паблишер, 2015. — 163 с. — ISBN 978-5-9614-5212-9.

Примечания

1. «Over the last fifty years, *How to Lie with Statistics* has sold more copies than any other statistical text.» J. M. Steele. "Darrell Huff and Fifty Years of *How to Lie with Statistics* (<http://www-stat.wharton.upenn.edu/~steele/Publications/PDF/TN148.pdf>) . *Statistical Science*, 20 (3), 2005, 205—209.

Ссылки

Источник — https://ru.wikipedia.org/w/index.php?title=Как_лгать_при_помощи_статистики&oldid=119762594

Эта страница в последний раз была отредактирована 2 февраля 2022 в 07:59.

Текст доступен по лицензии Creative Commons Attribution-ShareAlike; в отдельных случаях могут действовать дополнительные условия.

Wikipedia® — зарегистрированный товарный знак некоммерческой организации Wikimedia Foundation, Inc.

