

ОГЛЯНИСЬ ВОКРУГ

Глеб Погудин,

Екатерина Антоненко

*Существуют три вида
лжи: ложь, наглая ложь
и статистика.*

Старая шутка

КАК ВЫГЛЯДИТ
«СРЕДНЕСТАТИСТИЧЕСКИЙ СЛОН»?

Какой длины в среднем хобот у слона? Сколько среднестатистический человек тратит времени в день на завязывание шнурков? Каков процент девочек с голубыми волосами среди женского населения России?

Получить точный ответ на любой из этих вопросов крайне сложно: вам пришлось бы познакомиться со всеми слонами мира, провести много часов с секундомером в руке в самых разных точках земного шара и, что теперь кажется не таким уж сложным, пересчитать всех девочек с голубыми волосами в России.

Не странно ли, что на многие вопросы такого типа ответ можно найти в книгах или в интернете? Эти ответы получаются обычно в ходе «статистического исследования». Например, можно пойти гулять по Африке, измерить длину хобота только у первой тысячи встреченных слонов и в качестве ответа взять среднее из этих чисел. Скорее всего, полученное число будет близко к правильному ответу. На такой же логике основаны и всевозможные опросы общественного мнения. Однако у такого способа выяснения истины есть весьма неожиданные подводные камни.

Например, те объекты, которые вы выбрали для измерений, в нашем случае это первая тысяча встреченных слонов, могут оказаться «нетипичными». На такой ошибке основана шутка: «Интернет-опрос показал, что 100 процентов россиян пользуются интернетом». Разумеется, если человек принял участие в интернет-опросе, он воспользовался интернетом.

ПЕРВЫМ ДЕЛОМ – САМОЛЁТЫ

Это всего лишь шутка, но история знает достаточно примеров, когда в эту ловушку попадались неглупые и даже порою учёные люди.

Например, во время Второй мировой войны довольно остро стоял вопрос дополнительной защиты бомбардировщиков: многие из них не возвращались с задания. Первое, что приходит в голову, – обшить самолёт бронёй целиком, как танк. Однако тогда он просто не сможет взлететь из-за собственной тяжести. Военные осмотрели бомбардировщики, вернувшиеся



на базу, и отметили места, где повреждений было больше всего, предложив защищать бронёй эти места. Однако одобрено было неожиданное предложение математика Абрахама Вальда: защищать те участки, где повреждений почти не было. Ведь все части самолёта подвергаются удару одинаково часто (например, ракета не целится в определённое место самолёта, а взрывается рядом с ним, поражая осколками). А значит, самолётов, у которых повреждены «чистые» места, примерно столько же. Мы не видим их только потому, что они попросту не долетали до базы из-за этих повреждений. Следовательно, «чистые» места и есть наиболее уязвимые.

НЕУДАЧЛИВЫЕ ПРЕДСКАЗАТЕЛИ

Одной из самых известных ошибок такого рода является неудачное предсказание результатов президентских выборов в США в 1936 году, опубликованное журналом «Literary Digest». Тут стоит немного рассказать о том, как в Соединённых Штатах выбирают президента. Дело в том, что на выборах обычно соревнуются ровно два кандидата – по одному от каждой из основных партий (республиканской и демократической). Исторически так сложилось, что более богатые слои населения голосуют, как правило, за кандидата-республиканца, а более бедные – за демократа.

В 1936 году журнал «Literary Digest» провёл массовый опрос, разослав читателям анкеты, в которых требовалось написать, какому из кандидатов они отдают предпочтение. Было известно, что среди подписчиков журнала преобладают республиканцы, поэтому в число опрошенных были включены люди из телефонных книг и регистрационных списков автомобилей, так как адреса и тех, и других были доступны. Из ответивших 57 процентов собиралось голосовать за республиканца Альфа Лэндона. Весьма неожиданно для авторов опроса, выиграл выборы демократ Франклин Рузвельт с 62 процентами. Можно было бы думать, что причина в том, что в опросе приняло участие слишком мало людей. Однако одновременно с «Literary Digest»



ОГЛЯНИСЬ ВОКРУГ

другой исследователь Джон Гэллап получил результат, очень близкий к правильному, опросив в 50 раз меньше человек.

Этот пример стал классическим и вошёл во многие учебники. Стандартное объяснение такое: дело в том, что телефоны и автомобили в то время были менее распространены, чем сейчас, а значит, опрос среди их владельцев автоматически затрагивал более обеспеченную часть населения, которая обычно голосует за республиканцев. Однако более поздние исследования показали, что эта ошибка в проведении опроса была не единственной. Дело в том, что из разосланных анкет вернулась только четверть. Таким образом, «Literary Digest» собрали мнение не просто более обеспеченной части населения, а ещё и тех из них, кто был готов участвовать в опросе журнала, симпатизирующего республиканцам! Именно из-за сочетания этих двух ошибок разница между предсказанием и исходом выборов оказалась такой существенной.

ВОПРОСЫ ДЛЯ РАЗМЫШЛЕНИЯ

Напоследок предлагаем вам подумать самим над правильностью выводов, сделанных из статистических исследований.

1. Опросы, проведённые в штатах Флорида, Калифорния и Мэн, показали, что 55% опрошенных за последний год провели хотя бы две недели на океанском побережье. Можно ли отсюда сделать вывод, что 55% всех американцев проводят на берегу океана не менее двух недель в году? (*Подсказка: посмотрите на карту США.*)

2. Известно много случаев, когда дельфины спасали утопающих, поднимая их на поверхность воды и толкая в сторону суши. Однако некоторые исследования показали, что дельфины воспринимают человека как мячик и толкают в произвольном направлении. Может ли одно согласовываться с другим?

3. Многие университеты периодически рассылают своим выпускникам анкеты, чтобы выяснить, сколько те зарабатывают. Так вычисляется средняя зарплата выпускника данного учебного заведения. Оказалось, что результаты зачастую были завышенными. Как бы вы это объяснили?