

# КАК ЛГАТЬ ПРИ ПОМОЩИ СТАТИСТИКИ

*There are three kinds of lies:  
lies, damned lies, and statistics*

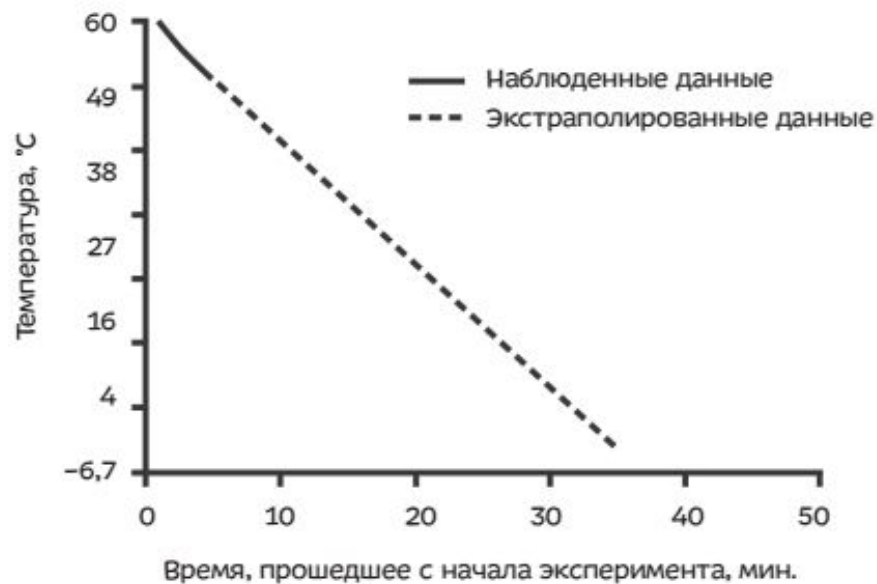
*Figures often beguile me particularly when I have the arranging of them myself*  
*Mark Twain*

*За последние 176 лет Нижняя Миссисипи стала короче на 242 мили. Это в среднем чуть больше одной мили и одной трети в год. Отсюда всякий здравомыслящий человек, если он не слепой и не идиот, может заключить, что в старом верхнем оолитическом силурийском периоде, чему в следующем ноябре исполняется миллион лет, Нижняя Миссисипи насчитывала свыше 1 300 000 миль в длину и торчала из Мексиканского залива, как удочка. Рассуждая аналогично, легко прийти к выводу, что через 742 года длина Нижней Миссисипи сократится всего лишь до одной мили и трёх четвертей, а Каир и Новый Орлеан объединятся под руководством одного мэра и общего городского совета. Есть всё-таки что-то притягательное в науке: можно получить внушительные дивиденды в предположениях от пустякового вложения фактов.*

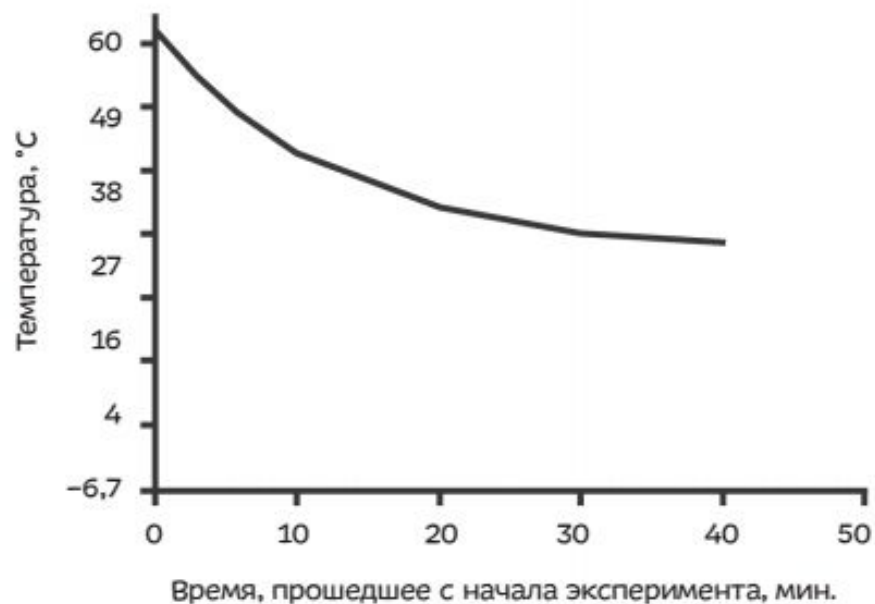
**Марк Твен Жизнь на Миссисипи**

# Необоснованная экстраполяция

Температура оставленного на столе кофе



Температура оставленного на столе кофе



Выборка изначально необъективна  
Грамотно выбранное среднее

## ГРАФИК И

указано количество проб  
подписаны оси  
указаны доверительные интервалы

## ТАБЛИЦ Ы

указано количество проб  
указаны коэффициент вариации или стандартн

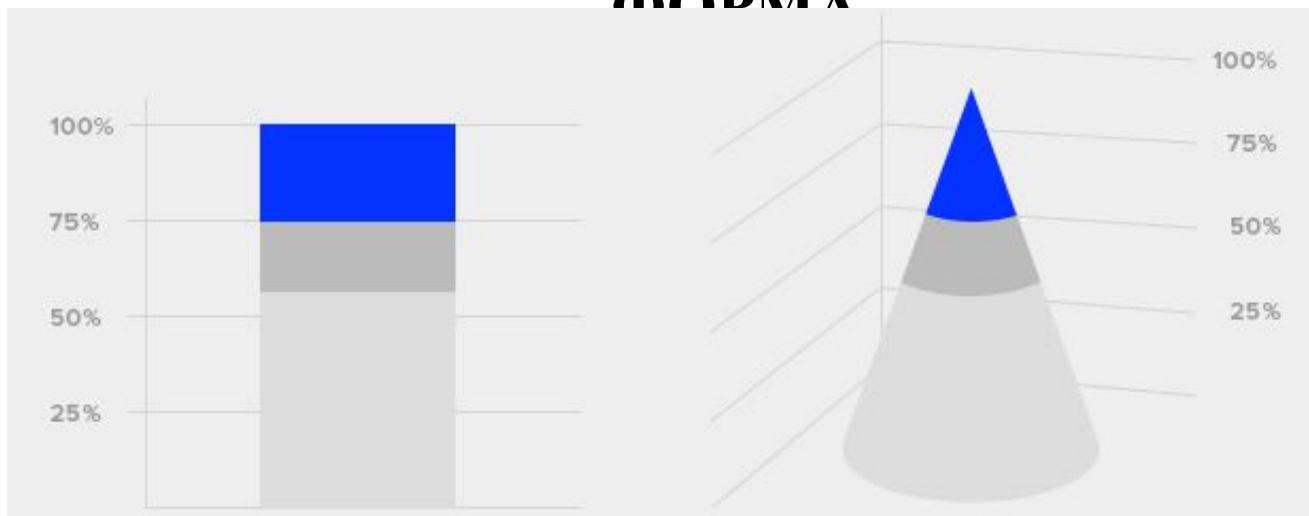
**чем больше времени требуется для  
расшифровки графика, тем больше  
вероятность, что вам пытаются пустить  
пыль в глаза.**

# СБИВАЮЩАЯ С ТОЛКУ ПЕРСПЕКТИВА



сдавшие статистику  
завалившие статистику

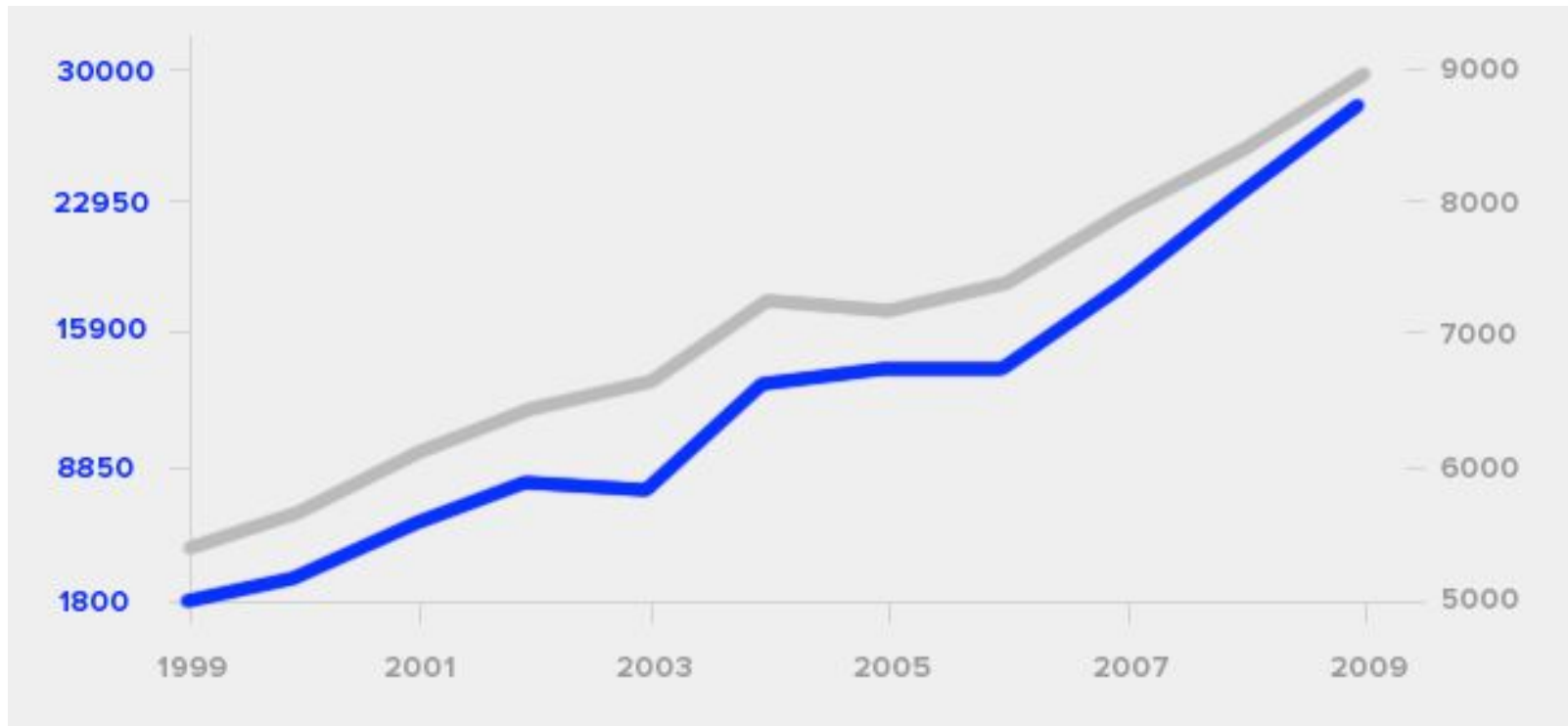
# ОБМАНЧИВАЯ ФОРМА



администрирование  
исследования  
гранты

# ПОИСК ТРЕНДОВ

(корреляция не подразумевает причинность)



Количество самоубийств через повешение и удушье  
Траты США на науку, космос и технологии

<http://www.tylervigen.com/spurious-correlations>

## **Cum hoc, ergo propter hoc**

*Вместе с этим, следовательно, по причине этого*

## **Post hoc, ergo propter hoc**

*После этого, следовательно, по причине этого*

**Почистил зубы – пошёл на учёбу!**



I USED TO THINK  
CORRELATION IMPLIED  
CAUSATION.



THEN I TOOK A  
STATISTICS CLASS.  
NOW I DON'T.



SOUNDS LIKE THE  
CLASS HELPED.  
WELL, MAYBE.



# КОВАРНЫЕ ГОРИЗОНТАЛИ



**маркетинг**  
**исследования**  
**продажи**

# ЛОВКОЕ МАСШТАБИРОВАНИЕ

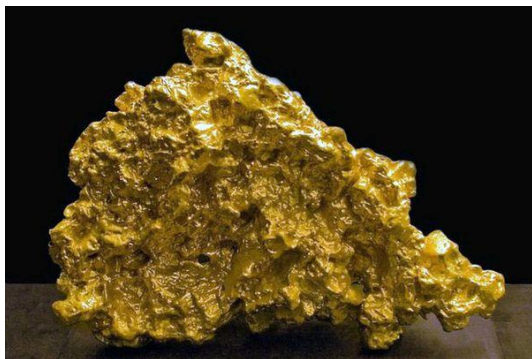


**КОЛИЧЕСТВО**  
**неаттестаций**

Царёво-Александровский прииск, Миасский район, Урал  
Никифор Сюткин, 1842 год

36 кг 15 г

27,5 см



39 см

55 см

*Открытие! В **два** раза больший самородок!!!*

288 кг 120 г



88 см

# Инфографика...

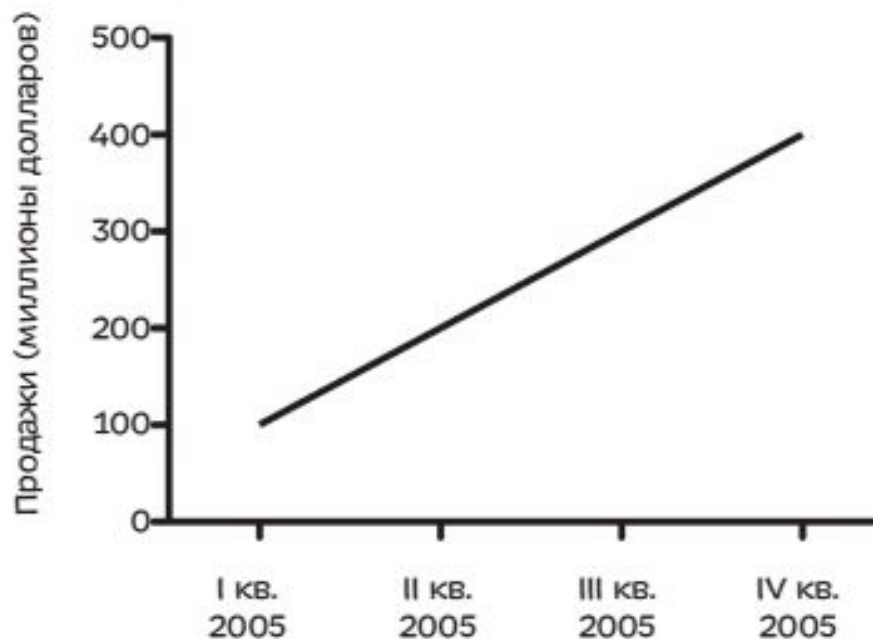


...такая инфографика

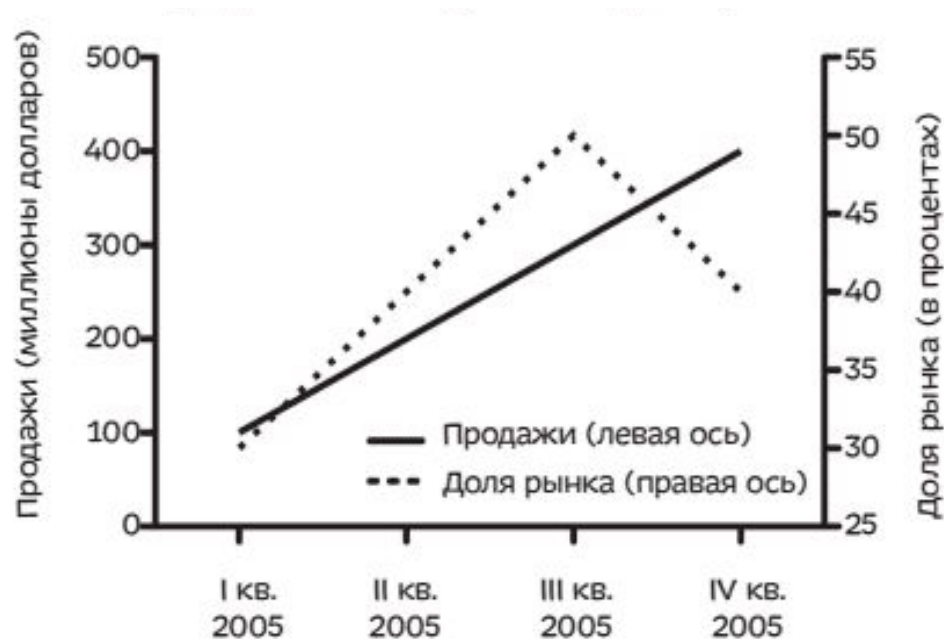


*Тот факт, что кто-то предъявляет вам статистику или показывает график, еще не означает, что все это относится к делу. ))*

### Продажи

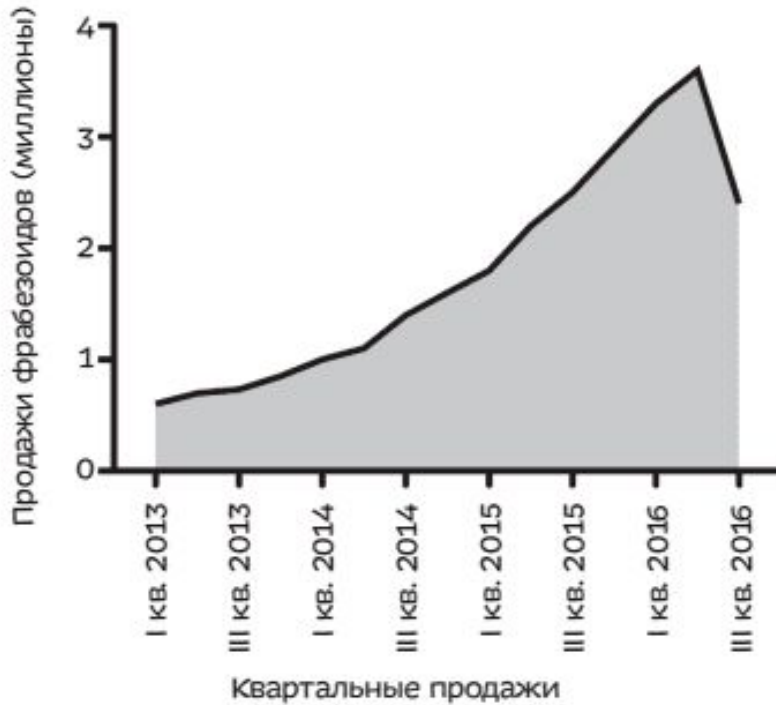


### Продажи и доля рынка

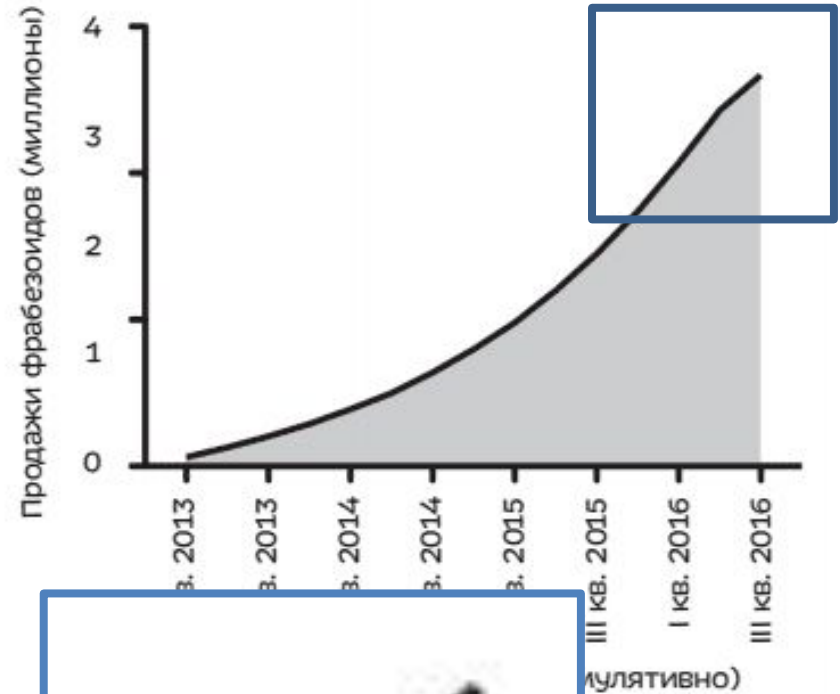


**А где прибыль?**

## Продажи яблок



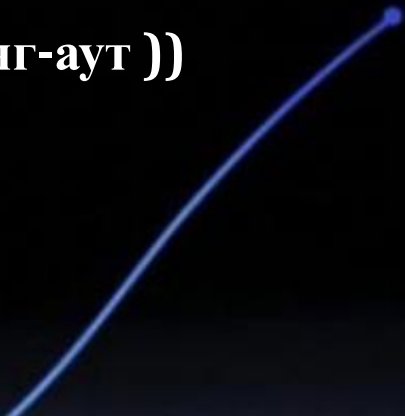
## Кумулятивные продажи яблок



# Cumulative iPad sales

В 2014 году совершён каминг-аут ))

180M  
150M  
120M  
90M



## Tim Cook

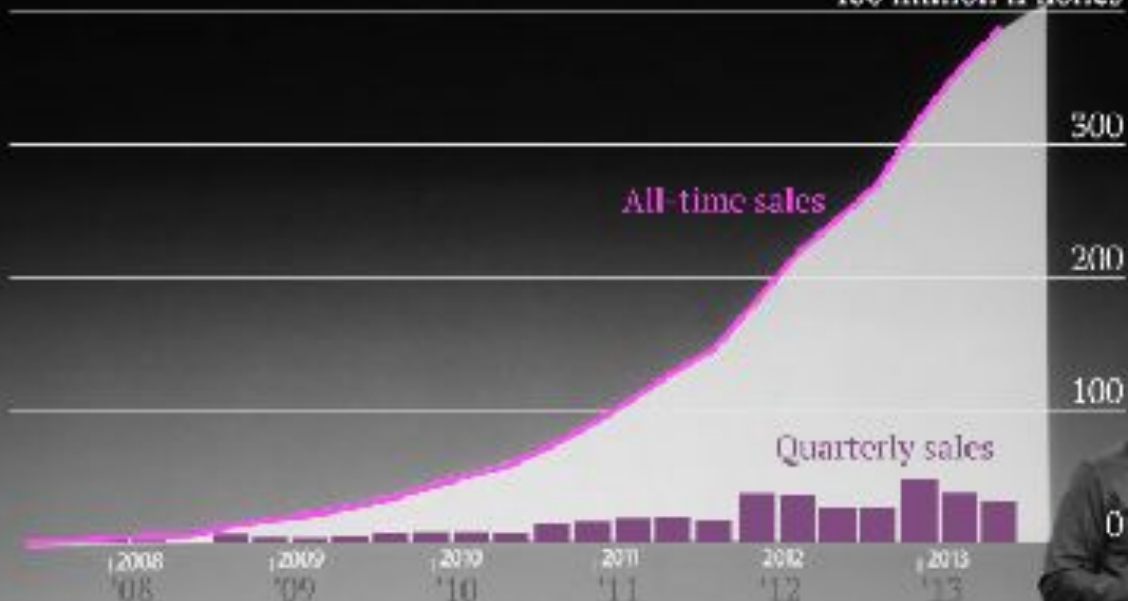
### CEO

(Chief Executive Officer)  
of Apple Inc.



# Cumulative iPhone sales

400 million iPhones



## СРЕДНЕ Е

Среднее бывает разное: медиана, мода, среднее арифметическое, среднее геометрическое, среднее урезанное, среднее пропорциональное

*Начальство ест мясо, я — капусту.  
В среднем мы едим голубцы.*



## СОГЛАСНО

## СТАТИСТИКЕ)))

Большая часть людей в течении 10 минут разговора врет 3 раза;

80% человечества врет каждый день;

65% людей считают, что нужно врать, чтобы не обидеть других людей;

45% населения планеты полагают, что просто необходимо преувеличить какие-то события, чтобы они казались интереснее;

95% тинейджеров врут своим родителям;

чем выше социальное положение, тем лучше человек лжет;

чем человек старше, тем лучше он врет;

50% людей считают, что воровство в интернете приемлемым, и всего лишь 5% людей считают воровство в магазине приемлемым;

70% девушек говорят неправду о своем весе;

Солгать по e-mail легче, чем в обычном письме;

15% сотрудников уличают во лжи прямо на рабочем месте;

20% парней лгут о количестве женщин, которые у них были;

30% пар врут друг другу о количестве денег, которые у них есть;

40% девушек считают нормальным, когда их молодой человек врет о том было ли ему хорошо с ней в постели;

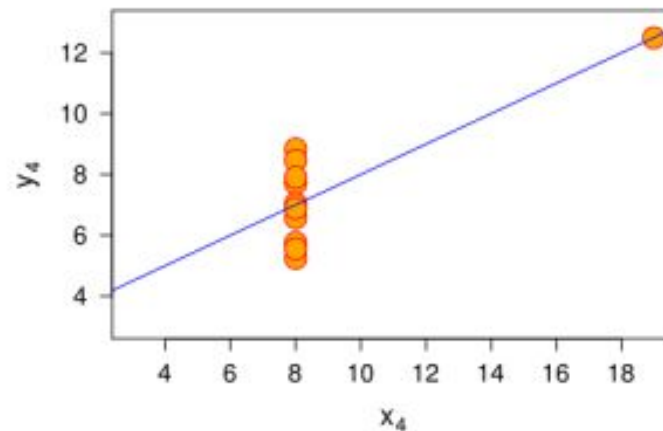
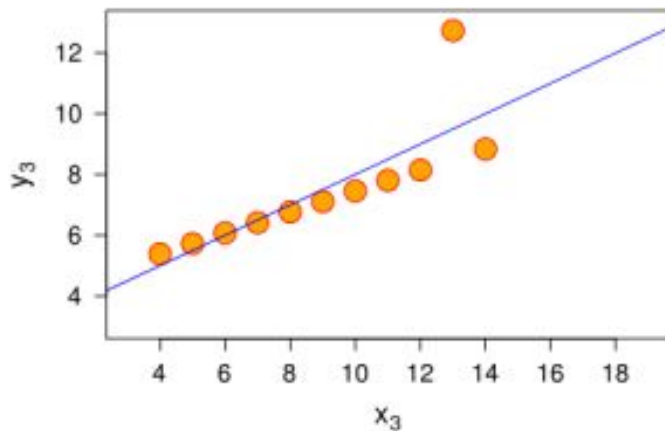
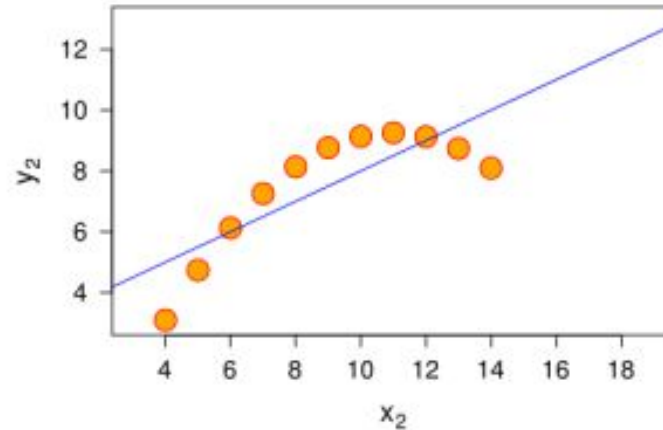
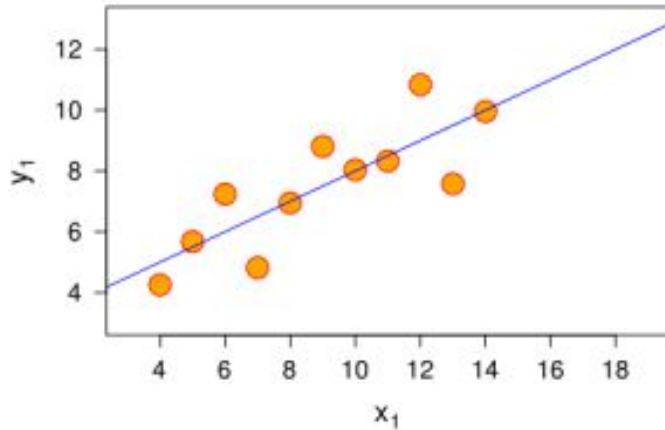
70% девушек полагают, что ничего страшного в том, что мужчина врет им по поводу их внешнего вида;

80% мужчин и женщин соврали бы о наличии у себя болезней, передающихся половым путем.

В геологии (до эксплуатации) очень много на доверии: никто за вас маршрут второй раз не пройдёт

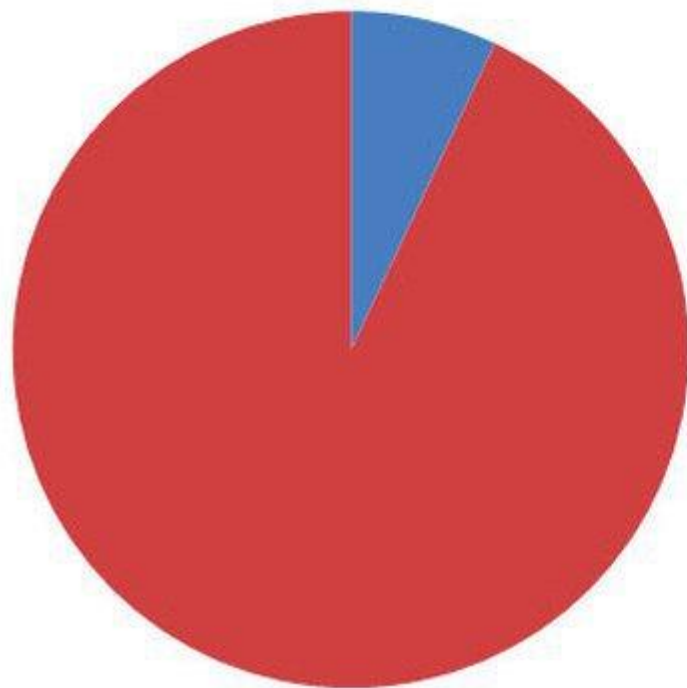
**Выбирайте результаты, которые подтверждают ваши идеи**

# Квартет Энскогомба



- Среднее значение переменной  $x$  для каждого набора данных равно 9.
- Среднее значение переменной  $y$  для каждого набора данных равно 7,5.
- Дисперсия (разброс) переменной  $x$  — 11, переменной  $y$  — 4,12.
- Корреляция между переменными  $x$  и  $y$  для каждого набора данных равна 0,816

# Статистика **не** врёт



■ НЕТ

■ ТОЖЕ НЕТ, НО СИНЕГО ЦВЕТА

«Четверо стоматологов из пяти рекомендуют  
зубную пасту Colgate»

...в ходе опроса стоматологи могли  
рекомендовать более одной зубной пасты. И,  
как оказалось, самого крупного конкурента  
Colgate рекомендовали почти так же часто,  
как и Colgate

результатов переработки мусора на самых разных улицах Лос-Анджелеса показывает, что одна конкретная улица перерабатывает в 2,2 раза больше, чем любая другая

самоубийства в Белорусии

слово «значимый», потому что для статистиков это не означает «заслуживающий внимания». В статистике это слово связано с тем, что данные были получены в результате статистических процедур, например проверки по критерию Стьюдента, Фишера и критерию хи-квадрат и т.д. (сотни их). Статистический уровень значимости представляет в количественной форме, насколько легко результаты объясняются чистой случайностью. При большом количестве наблюдений даже самые незначительные отклонения бывает сложно объяснить в рамках используемой статистической модели. Не критерии определяют, что заслуживает внимания, а что нет, — тут нужны человек и его оценка.

Не критерии определяют, что заслуживает внимания, а что нет, — тут нужны ~~человек~~ специалист и его оценка

значимый коэффициент корреляции при числе наблюдений 10 000

Чем больше у вас наблюдений в двух группах, тем вероятнее вы найдете между ними разницу



**Статистика : ни одному мужчине не удалось найти ядерную бомбу за девушкой**





# Time

На свете гораздо больше людей, у которых есть телефоны, нежели тех, у кого есть туалеты!!!

На деле в докладе ООН говорилось, что людей, у которых имелся ДОСТУП к мобильным телефонам, больше, нежели тех, у кого был доступ к туалетам, что существенно меняет картину

Доступ — одно из тех слов, которые вас должны насторожить, когда вы встречаете их в статистике. Люди, имеющие доступ к медицинскому обслуживанию, возможно, просто живут рядом с медцентром, но это не означает, что им дадут там консультацию или что они имеют достаточно денег, чтобы заплатить за лечение.

## **Мухи отдельно, котлеты отдельно**

в 2010 году погибло 3482 американских военнослужащих (при конфликте в Афганистане). Исходя из общего числа военнослужащих — а их 1 431 000 человек, — получим 2,4 смертельных случая на тысячу человек. На территории Соединенных Штатов количество смертей в 2010 году составило 8,2 на тысячу человек.

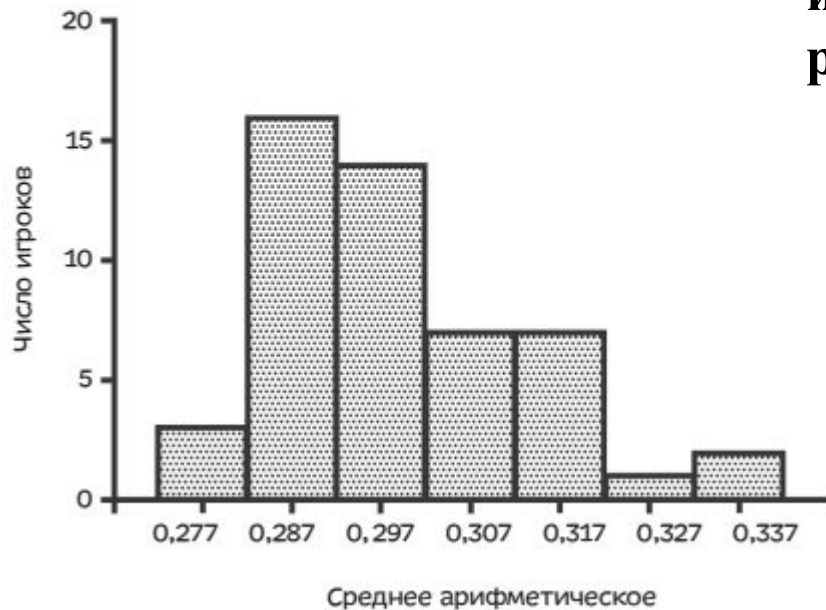
ТОЧНОЕ\_  
ВЕРНОЕ



«70% школьников в возрасте от 10 до 18 лет ведут половую жизнь». Нас не удивляет, что в этой категории 17- или 18-летние школьники, — но десятилетки! Такие заявления могут повергнуть в шок их бабушек и дедушек, которых придется отпаивать валерьянкой. Но ведь понятно же, что в общей категории, к которой отнесли и десятилетних детей, и 18-летних подростков, оказались и те, кто ведет половую жизнь, и те, кто нет. Гораздо правильнее было разбить всех участников исследования на группы, объединив по возрасту и имеющемуся у них жизненному опыту: например, 10-11 лет, 12-13, 14-15, 16-18.

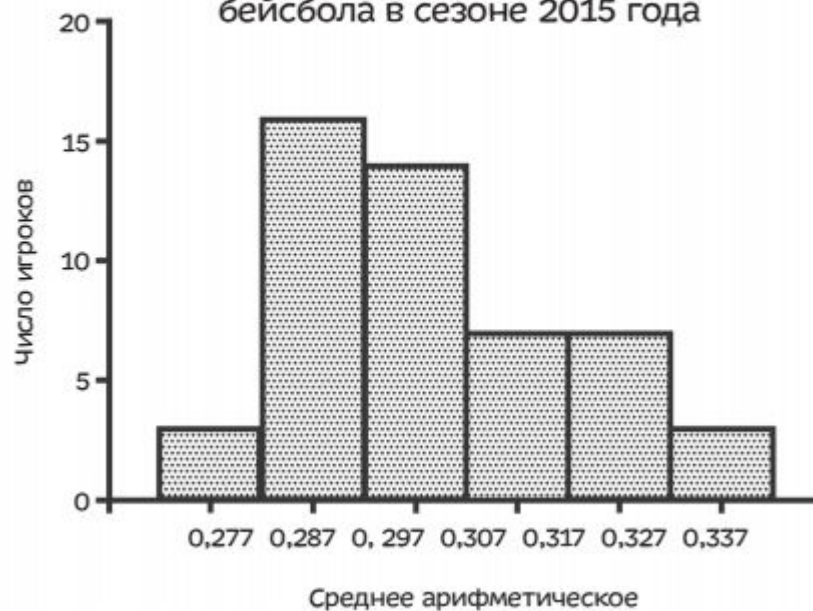
Что вообще имеется в виду под  
словосочетанием «вести половую жизнь»

Средний коэффициент результативности отбивания 50 лучших игроков Главной лиги бейсбола в сезоне 2015 года



## использование категорий разных размеров

Средний коэффициент результативности отбивания 50 лучших игроков Главной лиги бейсбола в сезоне 2015 года



## Обманчивое разделение на группы

болезни сердца: 611 105;  
рак: 584 881;  
хронические заболевания нижних  
дыхательных путей: 149 205.

острая ревматическая лихорадка и хроническое ревматическое заболевание  
сердца: 3260

гипертоническая болезнь сердца: 37 144

острый инфаркт миокарда: 116 793

сердечная недостаточность: 65 120

И так далее. Подобным же образом разбейте на подгруппы виды рака — и  
дело в шляпе! Заболевания нижних дыхательных путей становятся причиной  
смерти номер один.