**СТАНДОТКЛОН.В и СТАНДОТКЛОН.Г: в чем различие?**

Я уже несколько раз сталкивался с тем, что существуют две статистики, обе называемые стандартное отклонение, для описания меры рассеивания случайной величины. Недавно это произошло при чтении книги [Дональд Уилер, Дэвид Чамберс. Статистическое управление процессами](http://baguzin.ru/wp/?p=15577). Чтобы эти статистики различать, в книге они названы по-разному. Корень из среднего квадрата отклонения *sn*, определен для ряда, состоящего из *n* элементов формулой:

$$\left(1\right) s\_{n}= \sqrt{\frac{\sum\_{i=1}^{n}(X\_{i}–\overbar{X})^{2}}{n}}$$

Стандартное отклонение рассчитывается по формуле:

$$\left(2\right) s= \sqrt{\frac{\sum\_{i=1}^{n}(X\_{i}–\overbar{X})^{2}}{n –1}}$$

При этом авторы вскользь упоминают: «Несмотря на то что некоторые учебники говорят о применении *s* для генеральных совокупностей, а *sn* — для выборок, и то, и то — статистики, они обе — просто арифметические функции данных». На мой взгляд, то ли авторы, то ли переводчики напутали, потому что, как раз наоборот. «Некоторые учебники» трактуют *sn*, как меру рассеивания генеральной совокупности, а *s*, как меру рассеивания выборки.

Не добавляют понимания и [определения](https://ru.wikipedia.org/wiki/%D0%A1%D1%80%D0%B5%D0%B4%D0%BD%D0%B5%D0%BA%D0%B2%D0%B0%D0%B4%D1%80%D0%B0%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%BE%D0%B5_%D0%BE%D1%82%D0%BA%D0%BB%D0%BE%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5) в Википедии, где говорится, что *sn* – среднеквадратичное отклонение, а *s* – стандартное отклонение, или оценка среднеквадратического отклонения случайной величины *x* относительно её математического ожидания на основе **несмещённой оценки** её дисперсии. В статье [Несмещенная дисперсия](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D1%81%D0%BC%D0%B5%D1%89%D1%91%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B0) можно прочитать, что *sn* – выборочная дисперсия, а *s* – исправленная выборочная дисперсия. Причем *sn* является смещённой, а *s* – несмещённой оценками параметра σ2.

Excel не отстает от теории и содержит две функции для определения стандартного отклонения: СТАНДОТКЛОН.В – оценивает стандартное отклонение по выборке, и использует в формуле знаменатель (n – 1); СТАНДОТКЛОН.Г – вычисляет стандартное отклонение по генеральной совокупности, и в знаменателе – n.

Ситуация с двумя формулами прояснилась для меня при чтении книги Фишер. Статистический вывод. Фишер рассматривал генеральные совокупности, которые описываются *параметрами*, традиционно обозначаемыми греческими буквами. Параметры нам не известны. Мы пытаемся оценить их, извлекая отдельные выборки, и измеряя их *статистики* (статистика – число, характеризующее выборку; статистики традиционно обозначаются латинскими буквами). Фишер сформулировал несколько критериев хорошей оценки. И среди них – *смещение*. Оценка статистики Т считается правильной и несмещенной для параметра Θ, если среднее значение Т (по множеству выборок) стремится к истинному значению Θ. Иначе оценка считается смещенной.

Для иллюстрации я создал модель в Excel, и случайным образом задал 10 000 чисел в диапазоне от 0 до 100. А затем создал 100 выборок по 100 последовательных значений: от 1 до 100, от 101 до 200 и т.д. Далее построил три графика: для среднего значения, СТАНДОТКЛОН.Г и СТАНДОТКЛОН.В. На каждый график в виде пунктирной линии нанес среднее значение, СТАНДОТКЛОН.Г и СТАНДОТКЛОН.В для всех 10 000 случайных чисел, а также в виде точек – скользящее среднее этих статистик для последовательности выборок. Например, первая точка – значение статистики для первой выборки: 1…100, вторая точка – среднее статистик двух выборок: 1…100 и 101…200 и т.д. Видно, что среднее выборок стремится к своему истинному значению – среднему по всей совокупности, так же ведет себя и СТАНДОТКЛОН.В. А вот скользящее среднее СТАНДОТКЛОН.Г стремится к числу меньшему, чем значение СТАНДОТКЛОН.Г для всех 10 000 чисел. Это и означает, что статистика СТАНДОТКЛОН.В дает несмещенную оценку параметра дисперсии σ2, а СТАНДОТКЛОН.Г – смещенную.



Рис. 1. Три статистики – среднее значение, СТАНДОТКЛОН.Г и СТАНДОТКЛОН.В – дают оценку параметров генеральной совокупности; среднее значение и СТАНДОТКЛОН.В – несмещенную оценку, а СТАНДОТКЛОН.Г – смещенную; откройте Excel-файл и нажмите F9; случайные числа пересчитаются, и графики изменятся; неизменным будет только стремление среднего значения и СТАНДОТКЛОН.В по выборкам к своим истинным значениям (по всей совокупности), и постоянно заниженная оценка статистики СТАНДОТКЛОН.Г; Excel-файл тяжелый, поэтому пересчет происходит медленно

Если тема вас заинтересовала, но не полностью отложилась в голове, рекомендую также заметку [Выборочная несмещенная дисперсия](http://statanaliz.info/metody/opisanie-dannyx/15-vyborochnaya-dispersiya).