Стандартное отклонение и стандартная ошибка

Я читаю курс статистического мышления магистрам, и одна тема вызывает у них явные затруднения – чем стандартное отклонение отличается от стандартной ошибки, и в каких случаях, применять ту или иную статистику. А недавно в книге Искусство статистики Дэвида Шпигельхалтера я узнал про бутстрэппинг, и понял, как объяснить различия стандартного отклонения и стандартной ошибки.

Для начала зададим 100 значений стандартной нормально распределенной случайной величины. В этом контексте *стандартная* означает, что ее матожидание μ = 0, а среднеквадратичное отклонение σ = 1. Поскольку значения в Excel получены с помощью волатильной функции СЛМАССИВ(), после любого действия они пересчитываются. Поэтому диаграммы в заметке и в файле будут отличаться.

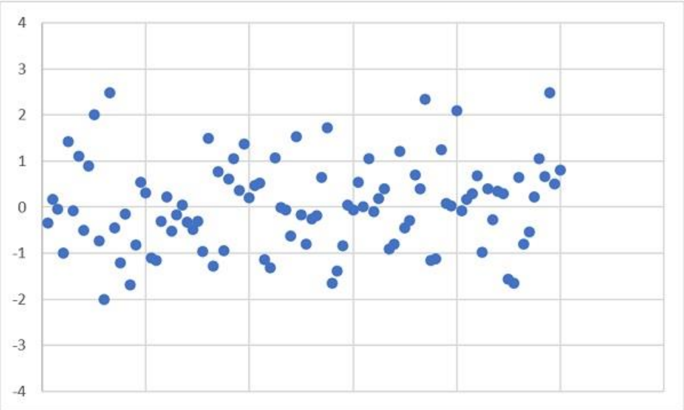


Рис. 1. Нормально распределенная случайная величина

Стандартное отклонение

Стандартное отклонение является наиболее распространенным показателем рассеивания значений случайной величины относительно её среднего арифметического.

Стандартное отклонение вычисляют по формуле:

$$(1)\quad s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$

где $\bar{X}$ – среднее арифметическое значений случайной величины (далее я буду называть его просто *средним*), X_i – отдельные значения случайной величины, n – число значений случайной величины.

Вообще термины разными авторами используются немного по-разному. Мне нравится следующий подход. Генеральную совокупность описывают параметрами, обозначаемыми греческими буквами: математическое ожидание μ и среднеквадратичное отклонение σ. Выборки описывают статистиками, обозначаемыми латинскими буквами: среднее арифметическое $\bar{X}$ и стандартное отклонение s. Стандартное отклонение иначе называют *оценкой среднеквадратичного отклонения*. Как правило, есть генеральная совокупность с неизвестным нам среднеквадратичным отклонением σ. Извлекая выборку, и вычисляя стандартное отклонение s, мы кое-что узнаем о среднеквадратичном отклонении генеральной совокупности σ. Поэтому и говорят, что s является оценкой сигмы.

На самом деле за термином *стандартное отклонение* стоят две немного отличающиеся статистики. Но эта заметка о другом)) Подробнее см. СТАНДОТКЛОН.В и СТАНДОТКЛОН.Г: в чем различие?

Нанесем на диаграмму линию среднего и границы, отстоящие от среднего на расстоянии ±2s.

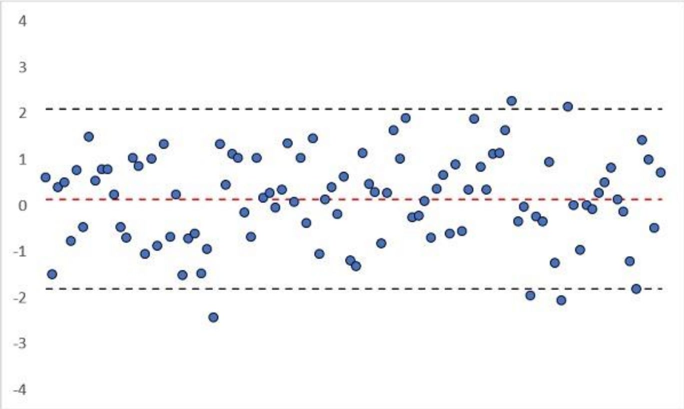


Рис. 2. Линия среднего и границы ±2s

Для стандартного нормального распределения за границы ±2s попадают 4,6% значений.

=(1-НОРМ.СТ.РАСП(2;ИСТИНА))*2 = 4,6%

И действительно 5 точек на рис. 2 лежат вне границ. Совпадение не обязано быть таким точным. Если вы откроете файл Excel на листе «Рис. 2» и понажимаете F9, принудительно изменяя случайные значения, то увидите, что вне границ может лежать от 2 до 8 точек. А если нажимать F9 достаточно долго, то вы получите более экстремальные числа точек вне границ. Для стандартного нормального распределения в пределах ±2s лежат приблизительно 95% значений. Поскольку s – оценка среднеквадратичного отклонения σ, которое в свою очередь равно 1, то 95% всех значений попадают в диапазон ≈ ±2.

Чем меньше s, тем кучнее значения случайной величины располагаются вокруг среднего. Итак

стандартное отклонение – мера разброса случайной величины

Среднее арифметическое выборки

Напомню, что мы задаем наши 100 значений с помощью генератора случайных чисел формулой в Excel

=НОРМ.СТ.ОБР(СЛМАССИВ(100;;0;1;ЛОЖЬ))

Хотя мы установили для генератора случайных чисел μ = 0 и σ = 1, значения $\bar{X}$ и s будут немного отличаться для каждой выборки.

$\bar{X}$	s
0,119	1,087
0,169	1,109
0,087	1,086
0,013	1,030
-0,053	0,921
0,164	1,108
0,067	1,035
-0,130	1,005
-0,016	0,893
0,030	1,010
-0,029	1,078
-0,043	1,020
-0,049	1,070
-0,118	0,916
0,028	0,978

Рис. 3. Среднее и стандартное отклонение для 15 выборок размером n = 100

Теперь мы хотим узнать, что можно сказать о неизвестном математическом ожидании генеральной совокупности μ, подсчитав среднее арифметическое конкретной выборки, например, первой $\bar{X}$ = 0,119?

Бутстрэп

Как пишет Евгения Поникарова, переводчик книги Дэвида Шпигельхалтера «Искусство статистики», слово bootstraps означает ремешки в виде ушка, которые прикрепляются к верхней части обуви, чтобы ее было проще натягивать. В английском языке есть выражение To pull oneself over a fence by one’s bootstraps (буквально — перетащить себя через ограду за ушки своей обуви), которое означает «выпутаться из своих проблем самому». Еще можно вспомнить барона Мюнхгаузена, который вытащил себя за волосы из болота.

Бутстрэп – компьютерный метод исследования распределения статистик, основанный на многократной генерации выборок методом Монте-Карло *на базе имеющейся одной выборки*. Термин ввел в 1977 году Брэдли Эфрон.

Итак, возьмем одну выборку из 100 случайных чисел и зафиксируем значения. Это наша исходная выборка (столбец A на рис. 4). Её среднее $\bar{X}(100)$ = 0,121, а стандартное отклонение $s(100)$ = 0,995. 95% значений попадают в диапазон ≈ 0,121 ± 1,990.

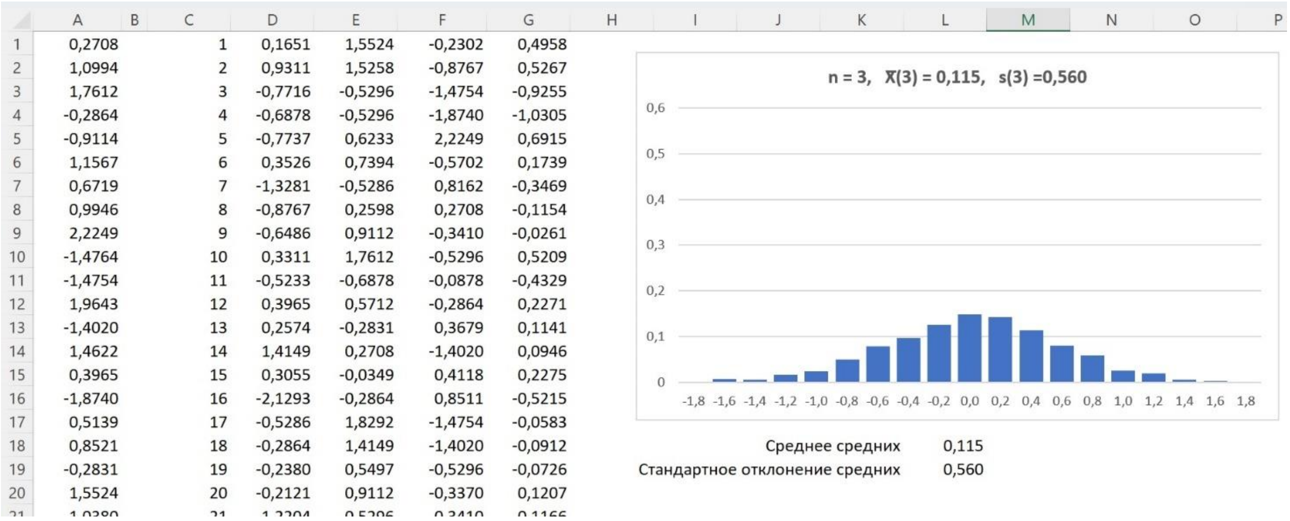
С помощью генератора случайных чисел будем формировать из исходной выборки бутстрэп-выборки разного размера. Хитрость заключается в том, что выбирать значения мы будем с возвращением. Т.е., все значения любой бутстрэп-выборки взяты из исходной, а вот уникальность значений будет потеряна. Например, выборка в столбце C содержит два значения 0,7394. Я подсветил их с помощью условного форматирования. Опять же, если вы откроете Excel-файл, то дублей может не быть, так как бутстрэп-выборка сформирована волатильной функцией СЛМАССИВ().

=ИНДЕКС(A1:A100;СЛМАССИВ(10;;1;100;ИСТИНА))

	A	B	C	D	E
1	0,2708		0,4118		
2	1,0994		0,7394		
3	1,7612		-0,0087		
4	-0,2864		-0,8716		
5	-0,9114		0,4798		
6	1,1567		-0,3094		
7	0,6719		1,2821		
8	0,9946		-0,1253		
9	2,2249		0,7394		
10	-1,4764		-0,5762		
11	-1,4754				
12	1,9643				
13	-1,4020				

Рис. 4. Бутстрэп-выборка может содержать повторения

Для удобства последующей обработки расположим значения бутстрэп-выборки по горизонтали. Начнем со значения n = 3. Извлечем 1000 бутстрэп-выборок (рис. 5). В столбце A исходная выборка, n = 100. В столбце C номер бутстрэп-выборки. В столбцах D, E и F извлеченные значения. В столбце G средние значения по выборкам. В ячейке G1 среднее D1:F1, в ячейке G2 – среднее D2:F2 и т.д. На диаграмме показано распределение средних значений бутстрэп-выборок для n = 3.



	A	B	C	D	E	F	G
1	0,2708		1	0,1651	1,5524	-0,2302	0,4958
2	1,0994		2	0,9311	1,5258	-0,8767	0,5267
3	1,7612		3	-0,7716	-0,5296	-1,4754	-0,9255
4	-0,2864		4	-0,6878	-0,5296	-1,8740	-1,0305
5	-0,9114		5	-0,7737	0,6233	2,2249	0,6915
6	1,1567		6	0,3526	0,7394	-0,5702	0,1739
7	0,6719		7	-1,3281	-0,5286	0,8162	-0,3469
8	0,9946		8	-0,8767	0,2598	0,2708	-0,1154
9	2,2249		9	-0,6486	0,9112	-0,3410	-0,0261
10	-1,4764		10	0,3311	1,7612	-0,5296	0,5209
11	-1,4754		11	-0,5233	-0,6878	-0,0878	-0,4329
12	1,9643		12	0,3965	0,5712	-0,2864	0,2271
13	-1,4020		13	0,2574	-0,2831	0,3679	0,1141
14	1,4622		14	1,4149	0,2708	-1,4020	0,0946
15	0,3965		15	0,3055	-0,0349	0,4118	0,2275
16	-1,8740		16	-2,1293	-0,2864	0,8511	-0,5215
17	0,5139		17	-0,5286	1,8292	-1,4754	-0,0583
18	0,8521		18	-0,2864	1,4149	-1,4020	-0,0912
19	-0,2831		19	-0,2380	0,5497	-0,5296	-0,0726
20	1,5524		20	-0,2121	0,9112	-0,3370	0,1207

Рис. 5. Распределений средних значений 1000 бутстрэп-выборок, n = 3

Среднее средних 1000 бутстрэп-выборок = 0,115, стандартное отклонение средних значений 1000 бутстрэп-выборок = 0,560. Напоминаю, что 95% исходных значений выборки попадают в диапазон 0,12 ± 1,99. Для бутстрэп-выборок n = 3 мы только что нашли, что 95% средних попадают в диапазон 0,115 ± 1,120 (0,560*2 = 1,120). Кажется естественным, что разброс средних меньше, чем разброс отдельных значений.

Повторим моделирование для n = 5, 20, 50.

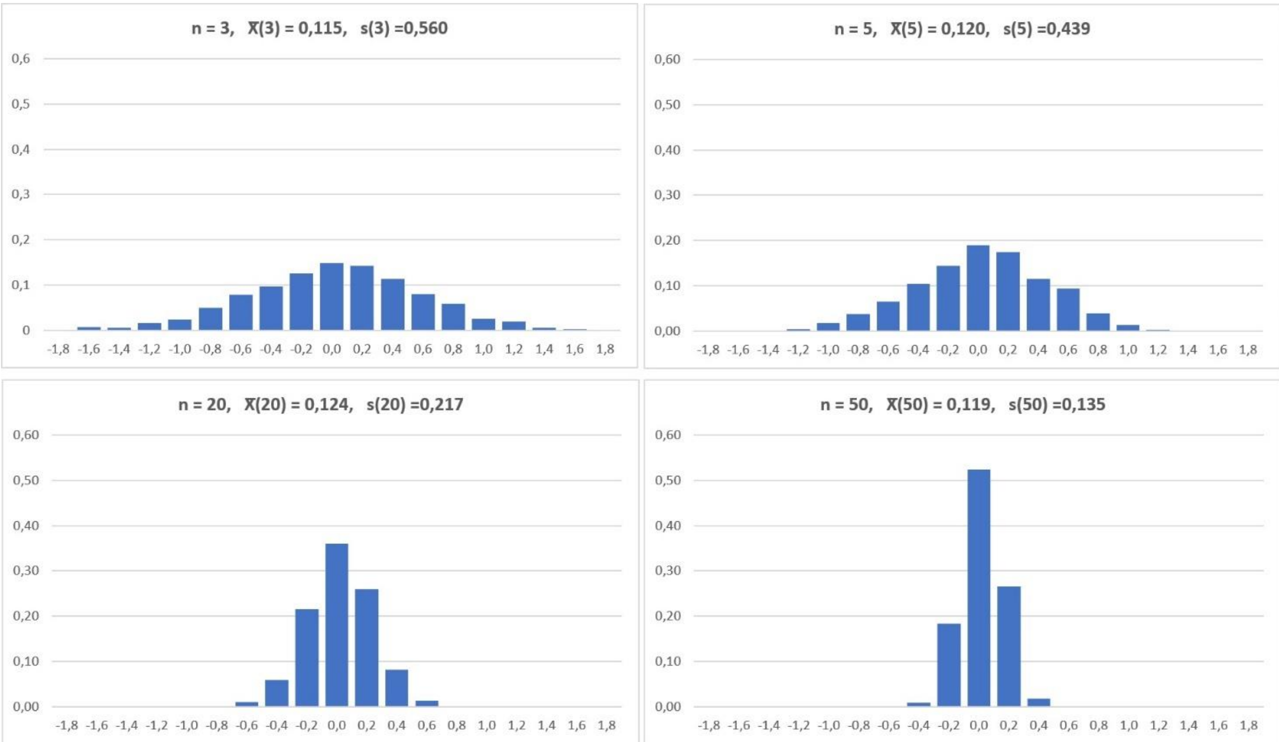


Рис. 6. С увеличением n стандартное отклонение средних значений бутстрэп-выборок уменьшается

Осмыслим, что мы получили. На рис. 6 представлены распределения средних значений бутстрэп-выборок разного размера из исходной выборки 100 случайных нормально распределенных чисел. Среднее каждого распределения близко к нулю (в нашей конкретной выборке из 100 чисел это среднее равно 0,121). А вот стандартное отклонение $s(n)$ уменьшается по мере роста размера бутстрэп-выборок: $s(3) = 0,560, s(5) = 0,439, s(20) = 0,217, s(50) = 0,135.$

Стандартна ошибка

…или стандартная ошибка среднего – статистика, характеризующая стандартное отклонение выборочного среднего, рассчитанное по выборке размера n из генеральной совокупности.

Ничего не напоминает!? А что за статистику $s(n)$ мы рассчитали выше в бутстрэп-анализе!? Да, это было стандартное отклонение выборочного среднего $\bar{X}(n)$.

Величина стандартной ошибки зависит от дисперсии генеральной совокупности σ^2 и объёма выборки n. Стандартная ошибка среднего вычисляется по формуле

$$(2)\ SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

где σ – величина среднеквадратического отклонения генеральной совокупности, и n – объём выборки. Поскольку дисперсия генеральной совокупности, как правило, неизвестна, то оценка стандартной ошибки вычисляется по формуле:

$$(3)\ SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

где s — стандартное отклонение случайной величины.

Сведем в одной таблице рассмотренные статистики:

I	J	K	L	M
Размер выборки, n	Среднее выборки, X̄	Стандартное отклонение выборки, s	Стандартная ошибка среднего, SE	Стандартное отклонение средних значений бутстрэп-выборок
3	-0,527	0,868	0,501	0,569
5	0,126	0,797	0,356	0,428
20	0,057	0,788	0,176	0,217
50	0,024	0,983	0,139	0,137
100	0,111	1,059	0,106	0,098

Рис. 7. Рассмотренные статистики

Здесь в столбцах J:L приведены статистики для одной выборки размера n, а в столбце M – статистики для бутстрэп-выборок соответствующего размера с рис. 6. Если в Excel-файле на листе «Рис. 7» понажимать F9, вы увидите, что не всегда совпадение между столбцами L и M будет таким хорошим, но тенденция будет прослеживаться.

Выше я писал, что мы исследуем неизвестное математическое ожидание генеральной совокупности μ на основе среднего арифметического выборки $\bar{X}(100)$ = 0,119.

Мы можем использовать статистику, именуемую *стандартной ошибкой*. Для нас она черный ящик – формула, выведенная на основе теории вероятностей. С другой стороны мы можем построить множество бутстрэп-выборок размера $n = 100$, и подсчитать стандартное отклонение средних этих бутстрэп-выборок. И мы показали, что стандартная ошибка для одной выборки и стандартное отклонение средних бутстрэп-выборок, это одно и то же! В нашем примере, получив $\bar{X}(100)$ = 0,119, мы можем сказать, что с вероятностью 95% математическое ожидание генеральной совокупности μ лежит в диапазоне 0,119 ± 0,212 (0,106*2=0,212). Итак

> *стандартная ошибка – мера оценки математического ожидания генеральной совокупности μ на основании статистик выборки*

Например, 95%-ный доверительный интервал для μ

$$(4)\ \bar{X} - 2 \cdot SE \leq \mu \leq \bar{X} + 2 \cdot SE,$$

где стандартная ошибка $SE = \frac{s}{\sqrt{n}}$

Понятно, что с увеличением размера выборки n доверительный интервал будет сужаться. В пределе при $n \to \infty$, $\bar{X} \to \mu$ и $SE \to 0$.