**Нормальное распределение. Построение графика в Excel**

Наверное, не все знают, что в Excel есть встроенная функция для построения *нормального распределения*. Графики *нормального распределения* часто используются для демонстрации идей статистической обработки данных.

В качестве примера обратимся к Excel-файлу.

Функция НОРМРАСП имеет следующий синтаксис:

*НОРМРАСП (Х; среднее; стандартное\_откл; интегральная)*

*Х* — аргумент функции; фактически НОРМРАСП можно трактовать как y=f(x); при этом функция возвращает вероятность реализации события *Х*

*Среднее (µ)* — среднее арифметическое распределения; чем дальше Х от среднего, тем ниже вероятность реализации такого события

*Стандартное\_откл (*σ) — стандартное отклонение распределения; мера кучности; чем меньше σ, тем выше вероятность у тех Х, которые расположены ближе к *среднему*

*Интегральная* — логическое значение, определяющее форму функции. Если «интегральная» имеет значение ИСТИНА, функция НОРМРАСП возвращает интегральную функцию распределения, тот есть суммарную вероятность всех событий для аргументов от -∞ до *Х*; если «интегральная» имеет значение ЛОЖЬ, возвращается вероятность реализации события *Х,* точнее говоря, вероятность событий находящихся в некотором диапазоне вокруг *Х*

Например, для µ=0 имеем:



2,14 13,59 34,13 34,13 13,59 2,14

Здесь по оси абсцисс единица измерения – σ, или (что то же самое), можно сказать, что график построен для σ = 1. То есть, «-2» на графике означает -2σ. По оси ординат шкала убрана умышленно, так как она лишена смысла. Точнее говоря, высота кривой зависит от плотности точек на оси абсцисс, по которым мы строим график. Например, если на интервал от 0 до 1σ приходится 10 точек, то высота в максимуме составит 4%, а если 20 точек – 2%. Здесь проценты означают вероятность попадания случайной величины в узкий диапазон окрестности точки на оси абсцисс. Зато имеет смысл площадь под кривой на определенном интервале. И эта площадь не зависит от плотности точек. Так, например, площадь под кривой на интервале от 0 до 1σ составляет 34,13%. Это значение можно интерпретировать следующим образом: с вероятностью 68,26% случайная величина Х попадет в диапазон µ ± σ.

Теперь, наверное, вам будет лучше понятен смысл выражения «качество шести сигм». Оно означает, что производство налажено таким образом, что случайная величина Х (например, диаметр вала) находясь в диапазон µ ± 6σ, всё еще удовлетворяет техническим условиям (допускам). Это достигается за счет значительного уменьшения сигмы, то есть случайная величина Х очень близка к нормативному значению µ. На графике ниже представлено три ситуации, когда границы допуска остаются неизменными, а благодаря повышению качества ([уменьшению вариабельности](http://baguzin.ru/wp/?p=236), сужению сигма) доля брака сокращается:



На первом рисунке только 1,5σ попадают в границы допуска, то есть только 86,6% деталей являются годными. На втором рисунке уже 3σ попадают в границы допуска, то есть 99,75% являются годными. Но всё еще 25 деталей из каждых 10 000 произведенных являются браком. На третьем рисунке целых 6σ попадают в границы допуска, то есть в брак попадут только две детали на миллиард изготовленных!

 Вообще-то говоря, измерение качества в терминах сигм используют не совсем нормальное распределение. ☺ Вот что пишет на эту тему [Википедия](http://ru.wikipedia.org/wiki/%D0%A8%D0%B5%D1%81%D1%82%D1%8C_%D1%81%D0%B8%D0%B3%D0%BC%D0%B0):

*Опыт показывает, что показатели процессов имеют тенденцию изменяться с течением времени. В результате со временем в промежуток между границами поля допуска будет входить меньше, чем было установлено первоначально. Опытным путём было установлено, что изменение параметров во времени можно учесть с помощью смещения в* ***1,5 сигма****. Другими словами, с течением времени длина промежутка между границами поля допуска под кривой нормального распределения уменьшается до 4,5 сигма вследствие того, что среднее процесса с течением времени смещается и/или среднеквадратическое отклонение увеличивается.*

*Широко распространённое представление о «процессе шесть сигма» заключается в том, что такой процесс позволяет получить уровень качества 3,4 дефектных единиц на миллион готовых изделий при условии, что длина под кривой слева или справа от среднего будет соответствовать 4,5 сигма (без учёта левого или правого конца кривой за границей поля допуска). Таким образом, уровень качества 3,4 дефектных единиц на миллион готовых изделий соответствует длине промежутка 4,5 сигма, получаемых разницей между 6 сигма и сдвигом в 1,5 сигма, которое было введено, чтобы учесть изменение показателей с течением времени. Такая поправка создана для того, чтобы предупредить неправильною оценку уровня дефектности, встречающееся в реальных условиях.*

С моей точки зрения, не вполне внятное объяснение. Тем не менее, во всем мире принята следующая таблица соответствия числа дефектов и уровня качества в сигмах:

|  |  |  |  |
| --- | --- | --- | --- |
| **Число сигм** | **Число дефектов на миллион измерений** | **Процент дефектов (несоответствий)** | **Уровень качества** |
| 6σ | 3,4 | 0,00034% | идеал |
| 5σ | 233 | 0,023% | мировой уровень |
| 4σ | 6210 | 0,62% | приемлемый уровень |
| 3σ | 66 807 | 6,68% | недостаточный уровень |
| 2σ | 308 537 | 30,9% | неприемлемый уровень |
| 1σ | 691 462 | 69,1% |

Для сравнения приведу таблицу для нормального распределения:

|  |  |  |
| --- | --- | --- |
| **Сигма окрестность среднего значения** | **Число случаем на миллион за пределами сигма окрестности** | **Процент случаев за пределами сигма окрестности** |
| 6σ | 0,002 | 0,0000002% |
| 5σ | 0,6 | 0,00006% |
| 4σ | 63 | 0,01% |
| 3σ | 2 700 | 0,27% |
| 2σ | 45 500 | 4,6% |
| 1σ | 317 311 | 31,7% |