

Нормальное распределение обычно ассоциируется с правилом 68-95-99,7, которое вы можете видеть на изображении выше. 68% данных находятся в пределах 1 стандартного отклонения (σ) от среднего (μ), 95% данных находятся в пределах 2 стандартных отклонений (σ) от среднего (μ) и 99,7% данных находятся в пределах 3 стандартных отклонений отклонения (σ) от среднего (μ).

Функция плотности вероятности

Чтобы иметь возможность понять, откуда берутся проценты, важно знать о функции плотности вероятности (PDF). PDF используется для указания вероятности того, что случайная величина попадает в определенный диапазон значений, а не принимает какое-либо одно значение. Эта вероятность определяется интегралом PDF этой переменной в этом диапазоне, то есть она определяется площадью под функцией плотности, но выше горизонтальной оси и между самым низким и самым большим значениями диапазона. Это определение может не иметь особого смысла, поэтому давайте проясним его, построив график функции плотности вероятности для нормального распределения. Приведенное ниже уравнение представляет собой функцию плотности вероятности для нормального распределения.

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Давайте упростим это, предположив, что у нас есть среднее значение (μ) равное 0 и стандартное отклонение (σ) равное 1.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Теперь, когда функция стала проще, давайте нарисуем график этой функции в диапазоне от -3 до 3.

```
# Import all libraries for the rest of the blog post
from scipy.integrate import quad
import numpy as np
import matplotlib.pyplot as plt

x = np.linspace(-3, 3, num = 100)

constant = 1.0 / np.sqrt(2*np.pi)
pdf_normal_distribution = constant * np.exp((-x**2) / 2.0)

fig, ax = plt.subplots(figsize=(10, 5));
ax.plot(x, pdf_normal_distribution);
ax.set_ylim(0);
ax.set_title('Normal Distribution', size = 20);
ax.set_ylabel('Probability Density', size = 20);
```

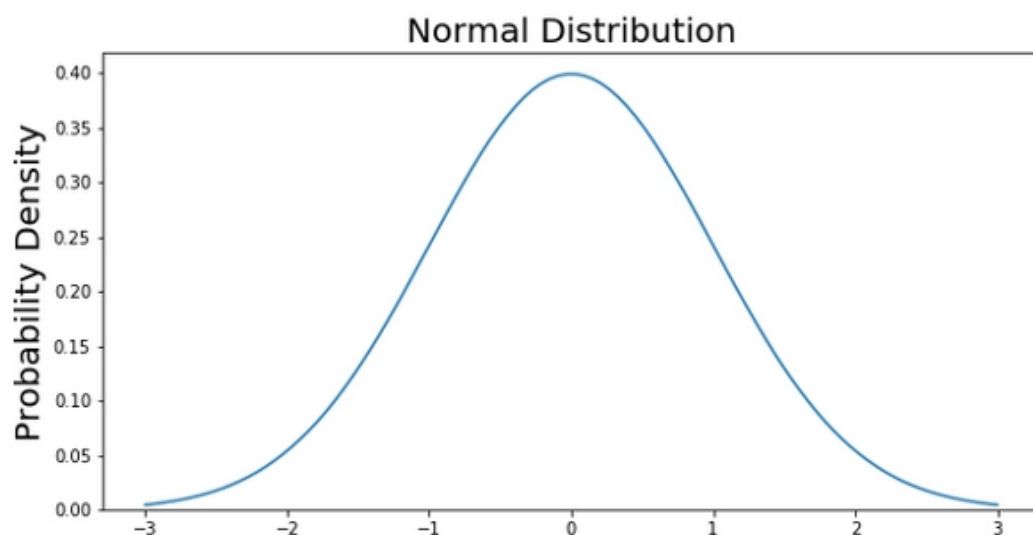


График выше показывает не вероятность событий, а плотность их вероятности. Чтобы получить вероятность события в заданном диапазоне, нам нужно будет интегрировать. Предположим, нас интересует вероятность того, что случайная точка данных приземлится в пределах 1 стандартного отклонения от среднего, нам нужно интегрировать от -1 до 1. Это можно сделать с помощью SciPy.

```
# Make a PDF for the normal distribution a function
def normalProbabilityDensity(x):
    constant = 1.0 / np.sqrt(2*np.pi)
    return(constant * np.exp((-x**2) / 2.0) )

# Integrate PDF from -1 to 1
result, _ = quad(normalProbabilityDensity, -1, 1, limit = 1000)
print(result)
```

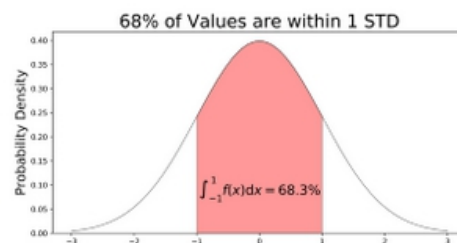
Math Expression

$$\int_{-1}^1 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

```
# Make the PDF for the normal distribution a function
def normalProbabilityDensity(x):
    constant = 1.0 / np.sqrt(2*np.pi)
    return(constant * np.exp((-x**2) / 2.0) )

# Integrate normal distribution from -1 to 1
result, _ = quad(normalProbabilityDensity, -1, 1, limit = 1000)
print(result)

0.682689492137086
```



68% данных находятся в пределах 1 стандартного отклонения (σ) от среднего значения (μ).

Если вас интересует вероятность того, что случайная точка данных окажется в пределах 2 стандартных отклонений от среднего значения, вам необходимо интегрировать от -2 до 2.

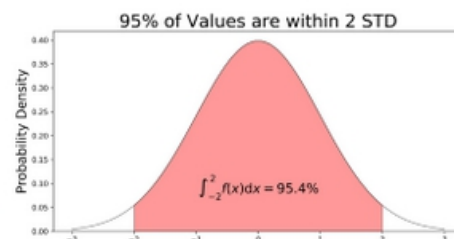
Math Expression

$$\int_{-2}^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

```
# Make the PDF for the normal distribution a function
def normalProbabilityDensity(x):
    constant = 1.0 / np.sqrt(2*np.pi)
    return(constant * np.exp((-x**2) / 2.0) )

# Integrate PDF from -2 to 2
result, _ = quad(normalProbabilityDensity, -2, 2, limit = 1000)
print(result)

0.9544997361036417
```



95% данных находятся в пределах 2 стандартных отклонений (σ) от среднего значения (μ).

Если вас интересует вероятность того, что случайная точка данных окажется в пределах 3 стандартных отклонений от среднего значения, вам необходимо интегрировать от -3 до 3.

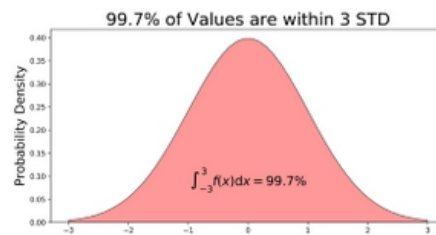
Math Expression

$$\int_{-3}^3 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

```
# Make the PDF for the normal distribution a function
def normalProbabilityDensity(x):
    constant = 1.0 / np.sqrt(2*np.pi)
    return(constant * np.exp((-x**2) / 2.0) )

# Integrate PDF from -3 to 3
result, _ = quad(normalProbabilityDensity, -3, 3, limit = 1000)
print(result)

0.9973002039367399
```



99,7% данных находятся в пределах 3 стандартных отклонений (σ) от среднего значения (μ).

Важно отметить, что для любой PDF площадь под кривой должна быть равна 1 (вероятность выпадения любого числа из диапазона функции всегда равна 1).

Вы также обнаружите, что наблюдения также могут падать на 4, 5 или даже больше стандартных отклонений от среднего, но это очень редко, если у вас нормальное или почти нормальное распределение.

