

正規分布、正規乱数とプログラミングによるグラフの描画。

実験や臨床情報から得られるデータには常に「ばらつき」が含まれ、臨床研究においては、例えば動物実験のように体重や週齢を揃えたりできないため、実験条件を制御することに限界がある。データのばらつきは「分布」として考えヒストグラム (histogram、頻度分布を表すグラフ) で視覚的に理解する。例数が少ない場合には全体の分布がわからない場合もあるが、臨床研究で数十例やそれ以上のデータが得られた場合にはヒストグラムからおおよその分布の形状がわかることが多い。

統計学ではいろいろな分布の形状が数式として定義されているが、薬学領域では正規分布、対数正規分布、あるいは二項分布などをよく目にする。詳細な数式は統計学の書籍に掲載されている。

正規分布：平均値と分散から、左右対称の分布が得られる。特に平均値を 0、分散を 1 としたとき「標準正規分布」と呼び、データ解析をする上で基本となる分布である。

対数正規分布：もとの数値の対数値が正規分布と仮定できる場合の分布。薬物血中濃度などの自然現象は対数正規分布で表されることがある。

二項分布：「コインを N 回投げたときにオモテ (あるいはウラ) が k 回上となる確率」と例えられる。

一様分布： x がどんな値でもその発生確率が同じ場合の分布。

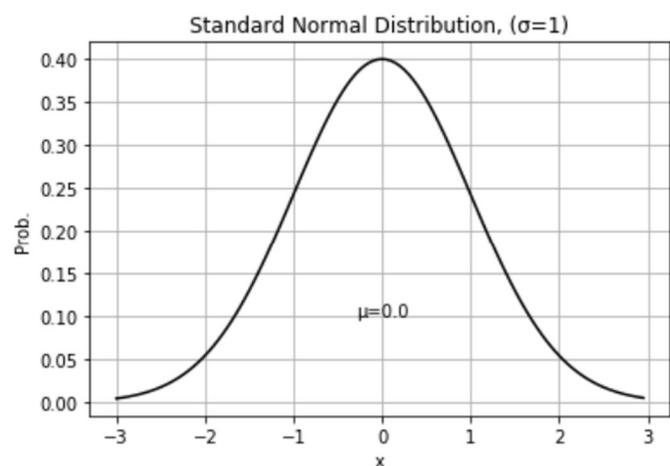
```
import matplotlib.pyplot as plt
import numpy as np

mu = 0.0
sigma = 1.0

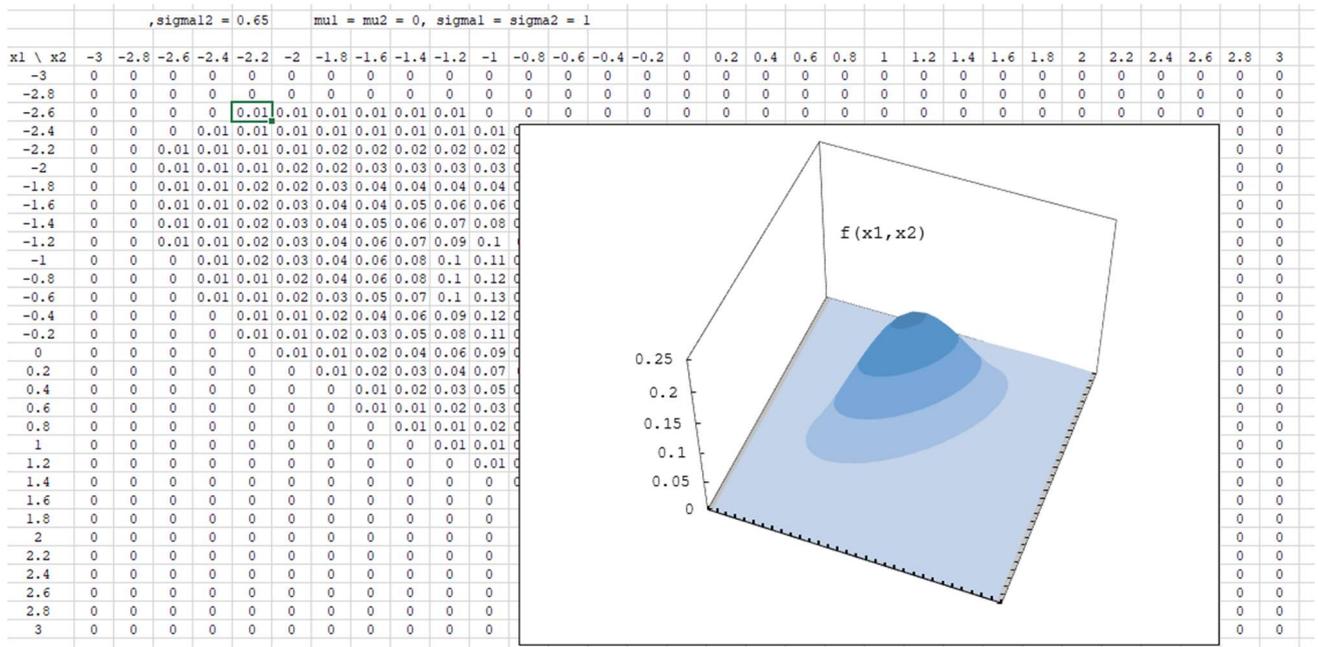
x = np.arange(-3.0, 3.0, 0.05)
y = (1 / np.sqrt(2 * np.pi * sigma)) * np.exp(-(x - mu) ** 2 / (2 * sigma))
print(x, y)

plt.plot(x, y, color = 'black')
plt.grid()
plt.xlabel('x')
plt.ylabel('Prob.')
plt.text(mu, 0.1, 'μ='+str(mu), ha='center')
plt.title('Standard Normal Distribution, (σ=1)')
plt.show()
```

これらの分布を表す数式が定義されているので、シミュレーションでグラフを作成できる。例えばプログラミング言語 Python では上のようなコードで標準正規分布を描画できる。実際には Python や R 言語ではより簡便な組み込みコマンドでグラフ化できるが、ここでは標準正規分布の式を定義し、値を小刻みに算出してグラフ化してみた (右図)。EXCEL でも `norm.dist()` といった関数があり描画できる。



正規分布を 2 つ以上同時に考えるとき「多変量正規分布」となり、2 変量までなら 3 次元的な図に描画できる。図に EXCEL で作図した例を示す。詳細な計算手順の説明は省略するが、EXCEL シートの各セルに 2 変量正規分布の式を定義し x_1 、 x_2 の値を変えて計算し、三次元グラフで示している。



データ解析において、これらの分布に従う数値をシミュレートしたい場合がある。特に一様分布に従う乱数（一様乱数）は、EXCELであれば rand() 関数で、R 言語なら unif 関数が使え。下図は EXCEL を用いて一様乱数を 1,000 個発生させてそのヒストグラムを描いたものである。ほぼ一定の高さにはなっているが、その高さにはまだばらつきがあり 1,000 個のシミュレーションでは十分に精度のよいシミュレーションができていないと思われる。正規乱数も EXCEL 関数や R 言語であれば rnorm() という関数がある。この場合もほぼ標準正規分布の形状にはなっているが、まだ十分にはなめらかではない。

