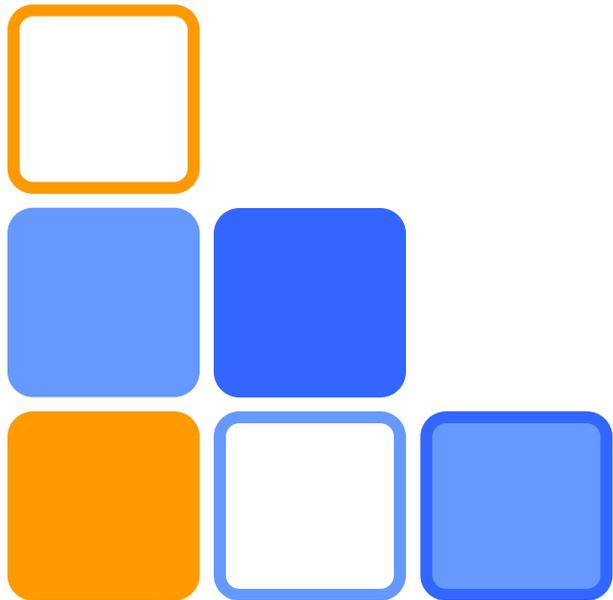


薬学情報処理演習 第2回

表計算ソフトによる統計 処理

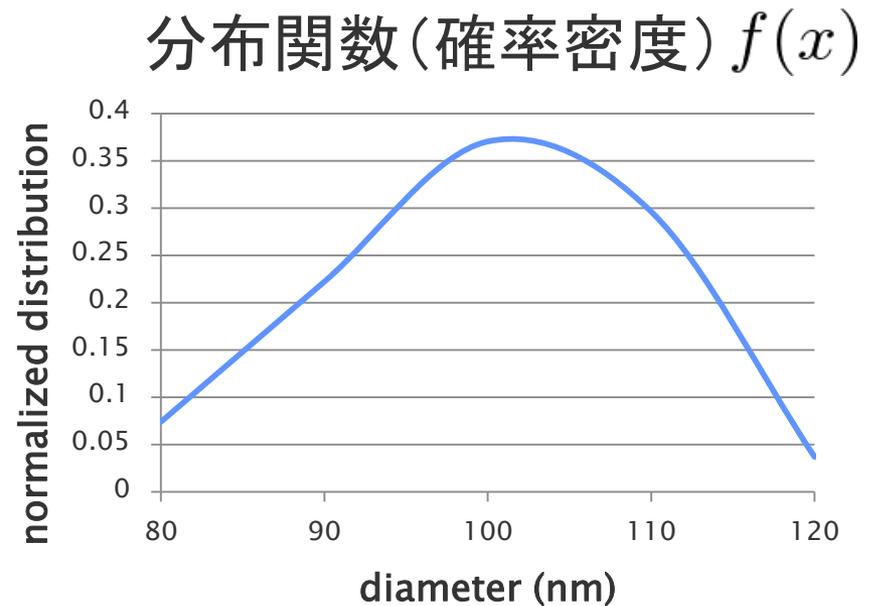
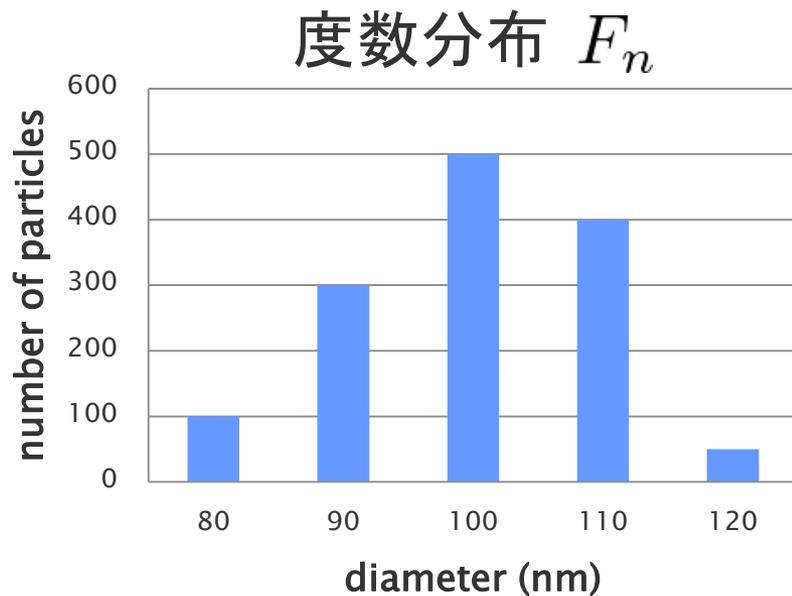


奥菌 透

コロイド・高分子物性学

度数分布

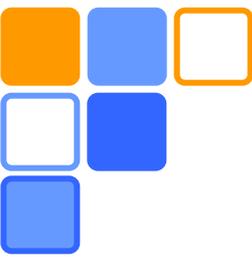
- 例：多数のコロイド粒子の粒径 x_i ($i = 1, 2, \dots, N$) の測定値から、度数分布を作成する。



F_n : 粒径が $x_n \leq x_i < x_n + \Delta x$
を満たす粒子の個数
 $\Delta x = x_{n+1} - x_n$

\longrightarrow
 $N \rightarrow \infty$
 $\Delta x \rightarrow 0$

$f(x)dx$: 粒径が x と $x + dx$
の間にある確率



度数分布(ヒストグラム)を作る

- 与えられたデータ(実験データなど)

$$x_1, x_2, \dots, x_N$$

に対して、データ区間

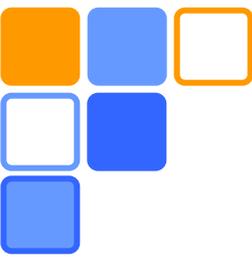
$$X_1, X_2, \dots, X_M \quad (X_{i+1} = X_i + \Delta X)$$

を定め、 $X_i \leq x_k < X_{i+1}$ を満たすデータ(x_k)の個数を度数(頻度) F_i とする。

- エクセルで度数分布を作る方法はいろいろあるが、ここでは「分析ツール」を使う。これを使用可能とするには、

Officeボタン/Excelのオプション/アドイン/設定

で「分析ツール」を選択して「OK」をクリックする。



ヒストグラムを作る練習

□ データを作る

- 0から99までの数をでたらめに30個程度一列に入力する

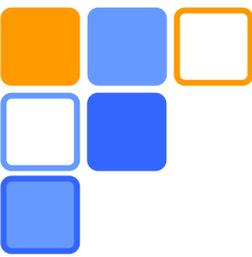
□ データ区間を入力する

- 間隔を20くらいにとる

□ 分析ツールを使う

- データ/分析/データ分析/ヒストグラム
- 入力範囲、データ区間を指定
- 出力先を選択・指定
- グラフ作成をチェック(オプション)

データ	データ区間
24	0
56	20
19	40
48	60
97	80
2	100
76	



演習課題

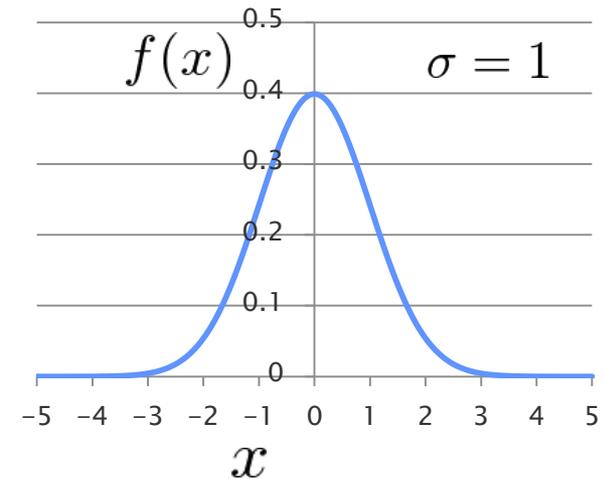
- 関数RAND()を用いてデータを作成する。
 - RAND()は呼び出されるたびに、0と1の間のでたらめな数(疑似乱数)を返す関数である。
 - この関数で生成された乱数は0と1の間で一様な分布をもつ(一様乱数)。
- 一様乱数から正規分布をもつ乱数を生成する。(後述)
- 上記のデータに対するヒストグラムを作る。
 - 分析ツールを用いて、度数分布を作る。
 - データを規格化して理論曲線と比較する。

規格化:
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

正規分布(ガウス分布)

- 実験的に測定される量には“ばらつき”がある。
ばらつき=平均値からのずれは以下のガウス分布に従うことが多い。なぜか？

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (\sigma^2 : \text{分散})$$



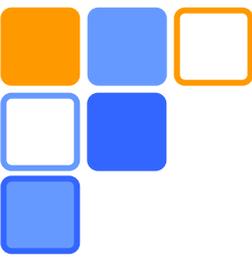
- 中心極限定理

n 個の独立な確率変数 u_i (分散 s_i^2 平均値0) からなる確率変数

$$x_n = (u_1 + u_2 + \cdots + u_n) / \sqrt{\sigma_n^2} \quad \sigma_n^2 = s_1^2 + s_2^2 + \cdots + s_n^2$$

は、 $n \rightarrow \infty$ で分散1, 平均値0の正規分布に従う。

- ばらつき=多数の確率的事象の和



データの作成(1)

- RAND()を使って乱数を生成する(数千個)。

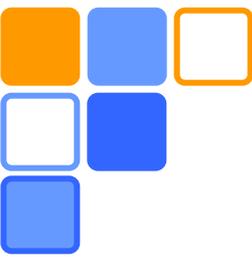
uniform	n=2	n=4	n=12
-0.04527	0.073737	0.366643	2.003184
-0.01196	-0.07254	-1.29495	0.527577
0.496424	0.442248	0.015446	1.058621

0.08482	-0.80747	-0.36108	2.226332
-0.05235	1.227829	-0.67171	0.602208

各データに対する統計量の出力:

- データのチェック
- データ範囲の目安

平均値			
-0.001	0.014706	0.023185	0.018525
分散			
0.084827	1.001033	0.978528	1.019962
最大値			
0.498811	2.39257	2.875021	3.597864
最小値			
-0.49964	-2.2876	-2.84531	-3.78472



データの作成(2)

- RAND()で生成される乱数は一様分布関数

$$p(x) = 1 \quad (0 \leq x < 1)$$

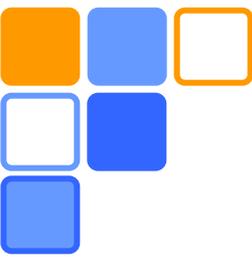
に従い、平均と分散は、

$$\langle x \rangle = \int_0^1 xp(x)dx = \frac{1}{2} \quad \langle (x - \langle x \rangle)^2 \rangle = \int_0^1 (x - \langle x \rangle)^2 p(x)dx = \frac{1}{12}$$

となるので、平均0の一様乱数と、それを n 個足し合わせた乱数を以下のように生成する。

$$u = x - 0.5 \quad x_n = (u_1 + u_2 + \cdots + u_n) \sqrt{12/n}$$

uniform	n=2
=RAND()-0.5	=(RAND()+RAND()-1)*SQRT(6)



度数分布の作成(1)

- データ区間を入力する
 - -4.0から4.0まで0.2刻みで入力
 - 代表値も作っておくとよい
- 理論曲線を描くためのデータを作成する

データ区間	代表値	正規分布
-4.0		
-3.8	$= (F2 + F3) / 2$	$= \text{NORM.DIST}(G3, 0, 1, \text{FALSE})$
-3.6		
-3.4		

度数分布の作成(2)

- 分析ツールを使って度数分布を出力する。
 - グラフは後で作成する(チェックしない)
- 度数分布を規格化し確率密度分布を計算する。

$$\sum_i F_i \Delta X = \mathcal{N} \quad \rightarrow \quad \sum_i f_i \Delta X = 1, \quad f_i = F_i / \mathcal{N}$$

データ区間	頻度	確率分布
-4.5	0	
-4.3	0	0
-4.1	0	0
-3.9	0	0
-3.7	0	0
-3.5	2	0.005
4.5	0	0
次の級	0	
積分値	400	

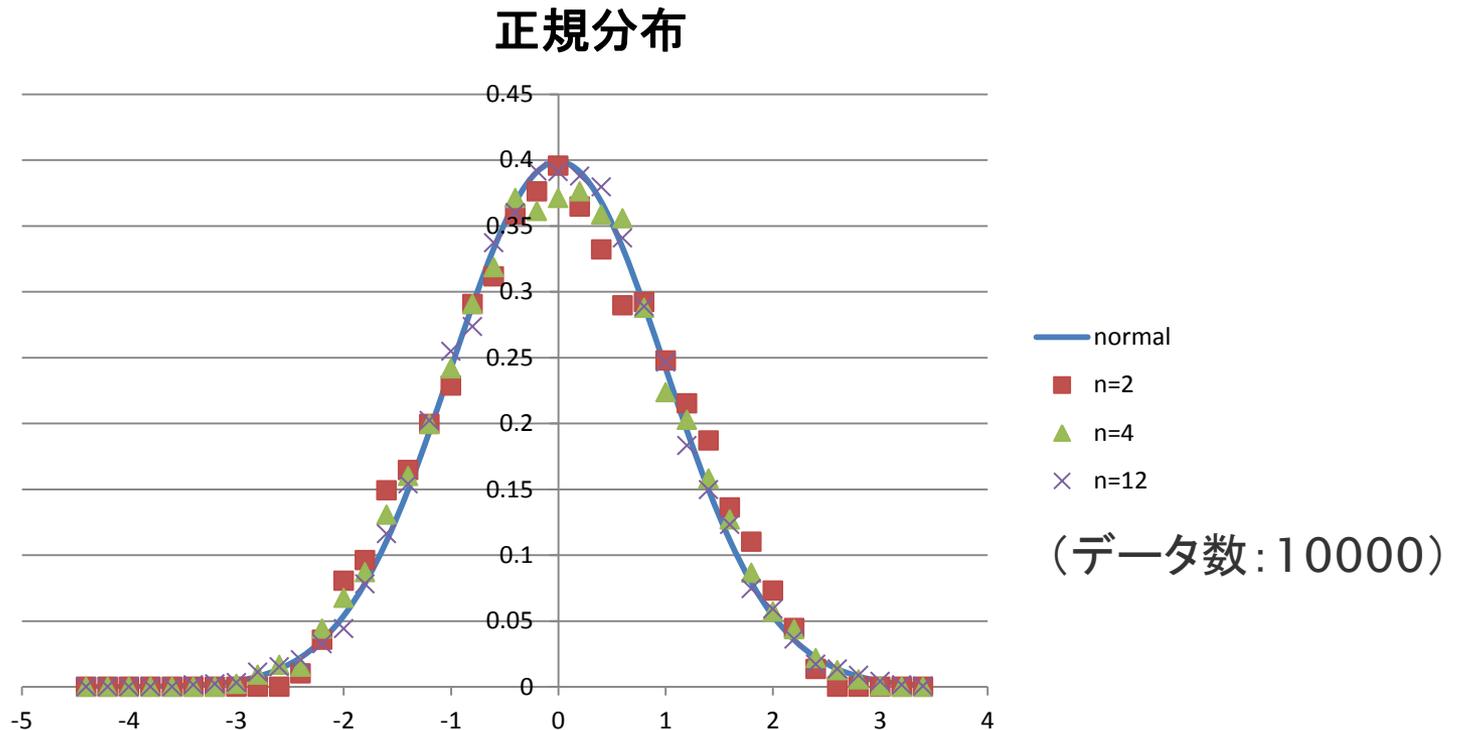
$$\leftarrow f_i = F_i / \mathcal{N}$$

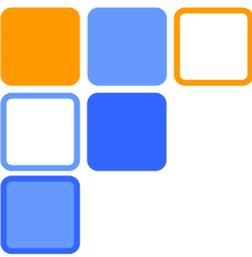
$$\leftarrow \mathcal{N}$$

グラフの作成

□ 正規分布のグラフ

- 横軸に代表値、縦軸に確率密度分布をとる。
- 理論値と度数分布から得られたデータを比較する。





レポート

- 演習課題をレポートとしてA4用紙1枚にまとめ、学籍番号、氏名(自筆)を明記してこの時間内に提出。