

# 第5章 確率分布の考え方

## —金融指標は正規分布に従うのか

データ・マイニングI

# 第4章で学んだこと

---

- 日本の金融市場における代表的な指標（日経平均など）を対象にこれらの指標の変動を分析した
  - 暦年や月次のデータに関しては、時間の経過に伴う各データ系列の変動を観察した
    - 暦年データや月次データについては、過去の変動を明確にし、今後の変動に見通しを付けた
  - 日次データにおいては、時間の経過に伴う変動を検討し、時間の経過を意識しないヒストグラムをベースとした原データ、日次変化、日次変化率に関する分析も行った
    - 日次データを観察する目的は、上記に加え、短期的な指標の変動に関するより詳細な情報を得ること

# 第4章で学んだこと

---

- さらに金融指標の統計学的性質について
  - 代表的な記述統計量の意味を説明しつつ,
  - 具体的な計算方法と併せて学び,
  - これを利用して各記述統計量がデータ分析で果たす役割を検討した

# 第5章で学ぶこと 確率分布

---

- 金融工学における各種の分析は金融指標の変動に多様な確率分布を想定して分析される
- そのための準備として、現実を観察された金融指標に基本的な加工を施し、これと理論的に想定される確率分布を結びつけて考えていく
  - 本書で利用する確率分布としては、一様分布、正規分布とその派生分布に限定する
  - データの観察を中心に据えることにより、理論的な説明のみでは感じをつかむことが難しい確率分布に関する考え方についてイメージを形成することができる

# 第5章で学ぶこと 株価と確率分布

---

- 金融データの統計的性質を分析するために確率分布の考え方を導入し、ヒストグラムの持つ確率的性質を考えていく
- 確率分布による分析は、データの発生状況を正確に説明することとはできないが、変化の可能性の程度を説明することが目的
  - 明日の株式市場で、日経平均株価がいくらになるかを正確に予測することは不可能だが、
  - 今日に比べて明日の株価が何円上昇あるいは下落する可能性がどれ位あるのかを知ることは株式市場で取引をしている人にとっては重要な情報である

# 第5章で学ぶこと 母集団とサンプル

---

- 各種金融指標の記録に残されている全ての対象期間を記録したデータが母集団になり,
- 特定の期間などを対象として記録したデータをサンプルと考える
  - 日経平均株価指数の動向を詳細に分析し全体像を正確に把握することは簡単ではない
  - そこで, いくつかのサンプルを分析することによって全体の動向を推察することが必要となる
    - 2003年から2011年のデータをベースに, 2003, 2005, 2008, 2011年の金融指標の各1年分をそれぞれの指標の母集団から得られた4種類のサンプルであると考え,
    - 各サンプルに関する分析から全体のデータに対する情報をどのように得ればよいのかという問題を考えていく

# 第5章で学ぶこと 一様分布と正規分布

---

上記の課題を考える為には確率分布に関するイメージを持つことが基本

- 確率分布の考え方を乱数によって求められる一様分布から出発する
- 統計学における各種の議論として利用されることの多い正規分布を説明し、正規分布の性質を理解する
- 前章で各種の金融指標やその変化、変化率に関して描いたヒストグラムの形状を正規分布として考えることが可能であるのかどうかを判断していく

# 離散一様分布

---

- $n$ は自然数とする
- 離散一様分布 $DU(n)$ とは、確率関数が

$$f_X(x) = \frac{1}{n}, x = 1, \dots, n$$

となるような確率変数 $X$ に対する分布

- 確率変数 $X$ が数値 $(1, \dots, n)$ を等確率 $1/n$ で取る



# 連続一様分布

---


- $a$ と $b$ は実数とする
- 連続一様分布 $U(a, b)$ とは、

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

となるような確率変数 $X$ に対する分布

- 確率変数 $X$ が区間 $(a, b)$ 上の数値が等確率で取る

# 【Excel】 確率分布の基本ー一様分布

- 246個（246営業日に対応）の一樣乱数を10組作ることにする
- 新しいワーク・シートを用意し，セルA2からセルA247に1から246までの通し番号を入力した上で，セルB1からセルK1にケース1からケース10と入力する
- この準備の下で， データ ，  データ分析 をクリックし，先に利用したヒストグラムより下方の乱数発生を選択してクリックすると，図5-1が表示される

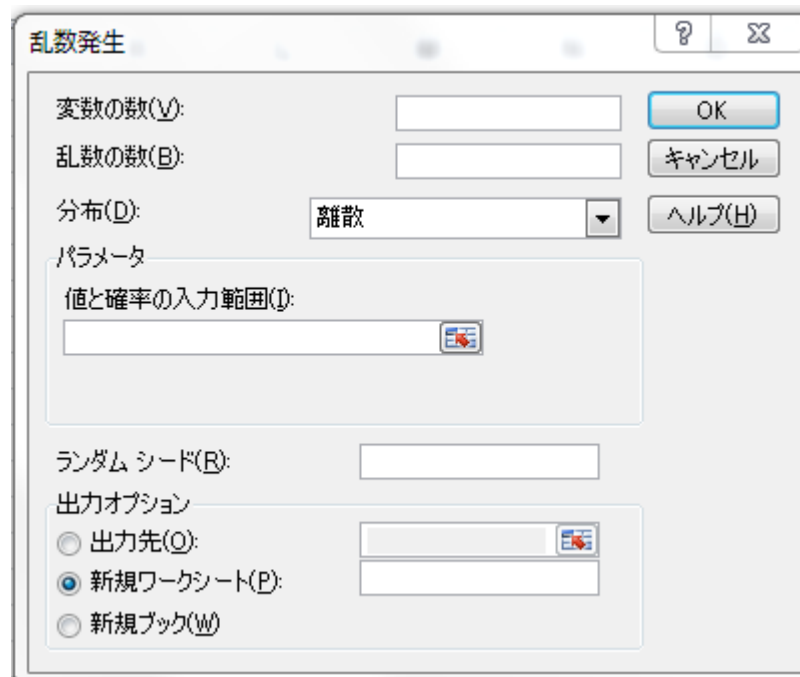


図5-1 乱数発生作業(1)

# 【Excel】 確率分布の基本——様分布

- 「分布(D)」の右側の下向きの矢印をクリックして表示される図5-2で、「均一」を選択する

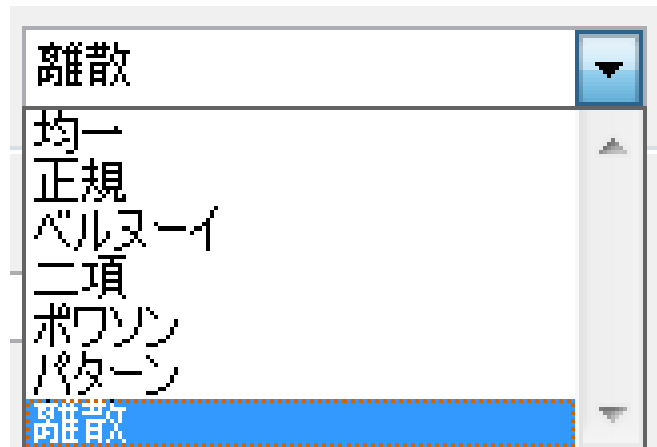




図5-2 乱数発生作業(2)

# 【Excel】 確率分布の基本——様分布

---

- 「出力先(O)」の右横の  マークをクリックし、セルB2からセルK247の範囲をクリック・アンド・ドラッグした上で、 をクリックしてOKをクリックする
  - 指定された範囲が0から1までの数値で埋められる
  - ここに示されたのが、0から1までの一様乱数である
- コンピュータによって作成された乱数は、乱数を発生させるための公式に従って計算された数値なので、擬似乱数と言う
  - 正10面体のサイコロを振って出る数字を記録し、この桁を区切って作成された真の乱数とは区別される
  - 擬似乱数には何らかの規則性が残存しており、規則性のまったくない真の乱数とは異なる

# 【Excel】 確率分布の基本——様分布

- 作成された10組の一樣乱数の一部を示したのが図5-3である
  - ここに示した数値と他の人が同じ手続きに従って作成した乱数の数値とは必ずしも一致しない
  - よって、数値が異なっても間違いではないのであわてる必要はない

	A	B	C	D	E	F	G	H	I	J	K
1		ケース1	ケース2	ケース3	ケース4	ケース5	ケース6	ケース7	ケース8	ケース9	ケース10
2	1	0.960234	0.435896	0.861324	0.353252	0.095828	0.144017	0.878048	0.420728	0.917417	0.476516
3	2	0.881863	0.22779	0.573962	0.67394	0.586993	0.823542	0.842555	0.13242	0.123997	0.359935
4	3	0.136845	0.501877	0.369793	0.226203	0.729209	0.145573	0.193945	0.435133	0.216376	0.867428
5	4	0.902188	0.417737	0.76104	0.312784	0.658376	0.882168	0.139531	0.949736	0.166234	0.893429
6	5	0.128086	0.036348	0.392224	0.830103	0.257393	0.480819	0.610462	0.374157	0.468368	3.05E-05
7	6	0.551347	0.190344	0.698965	0.48381	0.934935	0.451125	0.934141	0.969787	0.17011	0.863735
8	7	0.750053	0.136326	0.501419	0.355266	0.883328	0.719321	0.384045	0.072451	0.857326	0.758629
9	8	0.368694	0.851863	0.195959	0.08121	0.277596	0.817225	0.257057	0.926939	0.952544	0.259926
10	9	0.215613	0.122532	0.35667	0.45909	0.229163	0.711753	0.310892	0.596973	0.315744	0.445601
11	10	0.646382	0.426771	0.976043	0.041169	0.05414	0.848323	0.091037	0.613758	0.413587	0.048402
12	11	0.126347	0.189093	0.880215	0.361705	0.174047	0.7022	0.514756	0.508744	0.864223	0.38847
13	12	0.734581	0.366894	0.676351	0.973815	0.368328	0.93112	0.682882	0.569689	0.59566	0.620075
14	13	0.255074	0.6798	0.55266	0.58449	0.763695	0.223823	0.884884	0.173467	0.804346	0.985595
15	14	0.028504	0.215491	0.980102	0.39082	0.501755	0.269906	0.771538	0.021577	0.986755	0.505509
16	15	0.526322	0.772973	0.484115	0.362224	0.098819	0.058901	0.320933	0.652516	0.171575	0.972961
17	16	0.293313	0.257942	0.789575	0.810266	0.725181	0.07831	0.138707	0.857356	0.724265	0.388623
18	17	0.099643	0.855586	0.416547	0.260537	0.608234	0.04471	0.440626	0.204627	0.353893	0.192938
19	18	0.482864	0.689596	0.847224	0.519089	0.135044	0.694479	0.387616	0.914304	0.855403	0.950713
20	19	0.963164	0.368053	0.31428	0.029359	0.612781	0.006195	0.906339	0.930998	0.282937	0.646107

図5-3 乱数発生作業(3)

# 【Excel】 確率分布の基本――一様分布

- RAND関数を使って、一様乱数を生成することも可能である
  - セルB2に「=RAND()」と入力する
  - 他のセルにコピーする

	A	B	C	D	E	F	G	H	I	J	K
1		ケース1	ケース2	ケース3	ケース4	ケース5	ケース6	ケース7	ケース8	ケース9	ケース10
2	1	0.960234	0.435896	0.861324	0.353252	0.095828	0.144017	0.878048	0.420728	0.917417	0.476516
3	2	0.881863	0.22779	0.573962	0.67394	0.586993	0.823542	0.842555	0.13242	0.123997	0.359935
4	3	0.136845	0.501877	0.369793	0.226203	0.729209	0.145573	0.193945	0.435133	0.216376	0.867428
5	4	0.902188	0.417737	0.76104	0.312784	0.658376	0.882168	0.139531	0.949736	0.166234	0.893429
6	5	0.128086	0.036348	0.392224	0.830103	0.257393	0.480819	0.610462	0.374157	0.468368	3.05E-05
7	6	0.551347	0.190344	0.698965	0.48381	0.934935	0.451125	0.934141	0.969787	0.17011	0.863735
8	7	0.750053	0.136326	0.501419	0.355266	0.883328	0.719321	0.384045	0.072451	0.857326	0.758629
9	8	0.368694	0.851863	0.195959	0.08121	0.277596	0.817225	0.257057	0.926939	0.952544	0.259926
10	9	0.215613	0.122532	0.35667	0.45909	0.229163	0.711753	0.310892	0.596973	0.315744	0.445601
11	10	0.646382	0.426771	0.976043	0.041169	0.05414	0.848323	0.091037	0.613758	0.413587	0.048402
12	11	0.126347	0.189093	0.880215	0.361705	0.174047	0.7022	0.514756	0.508744	0.864223	0.38847
13	12	0.734581	0.366894	0.676351	0.973815	0.368328	0.93112	0.682882	0.569689	0.59566	0.620075
14	13	0.255074	0.6798	0.55266	0.58449	0.763695	0.223823	0.884884	0.173467	0.804346	0.985595
15	14	0.028504	0.215491	0.980102	0.39082	0.501755	0.269906	0.771538	0.021577	0.986755	0.505509
16	15	0.526322	0.772973	0.484115	0.362224	0.098819	0.058901	0.320933	0.652516	0.171575	0.972961
17	16	0.293313	0.257942	0.789575	0.810266	0.725181	0.07831	0.138707	0.857356	0.724265	0.388623
18	17	0.099643	0.855586	0.416547	0.260537	0.608234	0.04471	0.440626	0.204627	0.353893	0.192938
19	18	0.482864	0.689596	0.847224	0.519089	0.135044	0.694479	0.387616	0.914304	0.855403	0.950713
20	19	0.963164	0.368053	0.31428	0.029359	0.612781	0.006195	0.906339	0.930998	0.282937	0.646107

図5-3 乱数発生作業(3)

# 【Excel】 確率分布の基本——様分布

- 各ケースごとに平均と分散がどのようなになっているかを確認するため、各ケースの下にケースごとのデータの平均と分散とを計算して表示する
  - 図5-4に示すように、ケースごとに値がかなり異なっていることがわかる

	A	B	C	D	E	F	G	H	I	J	K
1		ケース1	ケース2	ケース3	ケース4	ケース5	ケース6	ケース7	ケース8	ケース9	ケース10
247	246	0.260475	0.453139	0.036897	0.200323	0.26606	0.999969	0.707724	0.248299	0.958495	0.360973
248	平均	0.491842	0.472109	0.538055	0.475704	0.492367	0.509576	0.474421	0.514513	0.520341	0.505357
249	分散	0.082832	0.087205	0.079534	0.086579	0.084633	0.090316	0.084228	0.08473	0.080872	0.087379

図5-4 乱数発生作業(4)

# 【Excel】 一様分布のヒストグラム


---

- ケース1からケース10の全データを対象とした平均と分散も併せて表示する
  - 個別ケースのデータの平均にはばらつきが認められるが、データ全体を用いた平均は、理論的に予想される0.5に非常に近い
- ヒストグラムを描いてこれらの数値の特性を観察する
  - 0から1の間の一様乱数を作成したので、0から1の間を0.1刻みで区切り、各範囲にいくつの乱数が存在するのかを度数分布表とヒストグラムによって確認すればよい



# 【Excel】 一様分布のヒストグラム

---

- 図5-5のように、セルM2からセルM11の間に0.1刻みの数値を入力した上で、 データ ,  データ分析 を選択し、

M
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

図5-5 乱数発生作業(5)

# 【Excel】 一様分布のヒストグラム

- ヒストグラムの画面で、ケース1のデータを対象にして、図5-6のように入力して、OKをクリックする

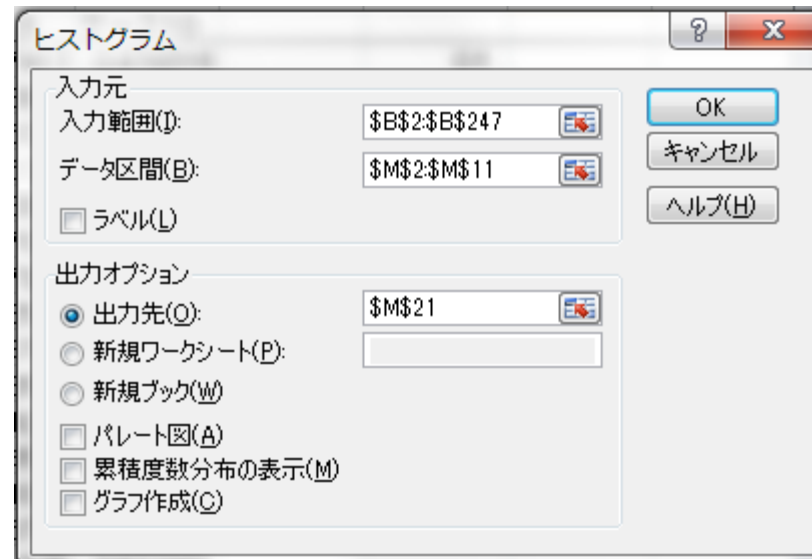


図5-6 乱数発生作業(6)

# 【Excel】 一様分布のヒストグラム

- 図5-7のように度数分布表とヒストグラムが表示される
  - 一様乱数といってもその分布が完全に均等になっていないことを認識すること

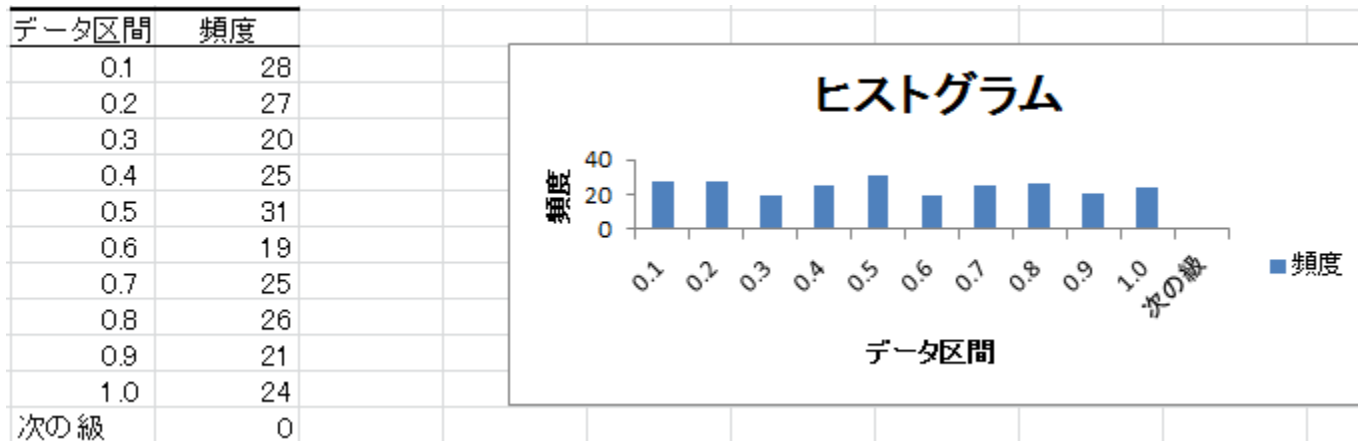


図5-7 乱数発生作業(7)

## 【練習問題5-1】 提出課題15:2点

---

- ケース2からケース10までのヒストグラムを,「入力範囲(I)」と「出力先(O)」を変更して9組作成し,これらの結果を見比べてみよ  
(結果は本章の付録参照)
- Excelとして提出せよ

## 【練習問題5-1】 提出課題15:2点

- 乱数というのは与えられた条件を満たしつつ、不規則に発生させたデータである
  - どのケースをとっても各区間に均等に分布が散らばっていることはない
  - どのケースにも共通でデータ数が多くなっている区間も存在しない
- データ数が多くなるにつれ乱数発生が規則が明確に示されるので、ここで作成した全ての乱数をひとつのヒストグラムに描いた図5-8では、各区間における散らばり方の均質性が高くなっている

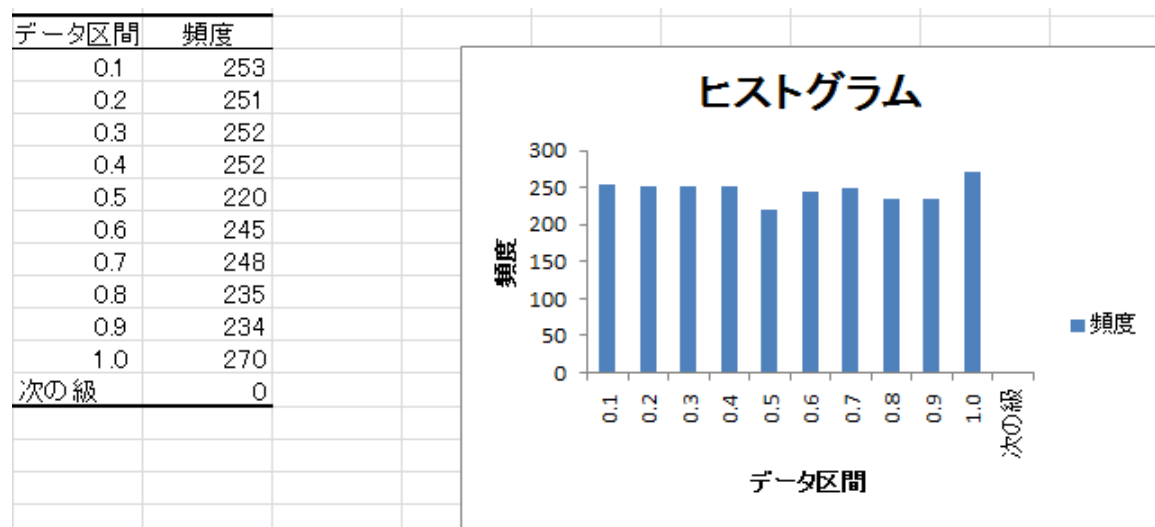


図5-8 乱数発生作業(8)

## 【練習問題5-1】 提出課題15:2点

---

- 上で説明したことが、母集団とサンプルとの間の関係である
  - 最後に表示したデータがいわゆる母集団の分布に対応し、ケースごとのデータがサンプルの分布に対応している
  - サンプルとして取られたデータにおける各種の性質から母集団の性質を知ることがいわゆる統計的推定や検定の問題となる
- サンプルのデータを見ただけで母集団における性質を知ること  
は簡単ではないので、各種の統計学の手法が用意されている

# 正規分布

---

- 統計分析において最も利用される確率分布は正規分布である
- 正規分布が幅広く利用される理由は,
  1. 数学の確率論で議論されるように, 各確率変数が独立で同じ分布に従う確率変数の平均が正規分布に帰着できるという中心極限定理があるため
  2. たった2種類のパラメータ(平均と分散)の組み合わせによって分布を近似できるという計算の簡便性もある
- ここでは, 正規分布に従う確率変数の組を発生させ, これらのデータを用いて正規分布の性質を体得する

# 正規分布と標準正規分布

---

- 平均が0で分散が1の正規分布を標準正規分布と呼び, 正規分布の基本型として利用される
  - 任意の平均と分散を持つ正規分布は変数変換により容易に標準正規分布に変換することが可能
  - 標準正規分布の性質がわかれば任意の正規分布の性質を理解することができる



# 正規分布の確率密度関数

---

- 正規分布の確率密度関数(各確率変数が特定の値をとる確率を示す関数である)は, 以下の式で示される

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (x - \mu)^2 \right] \quad (5-1)$$

- $\pi$ は円周率
- $\exp$ は自然対数の底 $e$ に関する指数関数
- $\mu$ は正規分布の平均
- $\sigma$ は正規分布の標準偏差( $\sigma^2$ は正規分布の分散)
- 正規分布を規定するのは平均 $\mu$ と分散 $\sigma^2$ という2つのパラメータだけ
  - 任意の確率変数 $X$ が平均 $\mu$ , 分散 $\sigma^2$ の正規分布に従う(こうした分布から得られた値である)ことを,  $N(\mu, \sigma^2)$ と表現する

# 標準正規分布の確率密度関数

---

- 標準正規分布に従う確率変数の確率密度関数は、平均が0、分散が1であるので以下の式で示すことができる

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} (z)^2 \right] \quad (5-2)$$

- 確率変数 $z$ が標準正規分布に従うことを、 $N(0, 1)$ と表現することも多い
- なお、正規分布は釣鐘状の形をする

# 【Excel】正規分布に従う確率変数の生成

---

- Excelによって、正規分布に従う確率変数を生み出す方法は2つある
  1. 第1の方法は、「NORM.DIST」というExcelの関数を利用する方法
    - 上の正規分布の定義式に従って計算される各 $x$ に対応する密度関数 $f(x)$ の値が求められる
  2. 第2の方法は、正規分布に従う乱数(正規乱数)を作成する分析ツールを利用する方法

# 【Excel】 正規分布の確率密度と累積分布

- ワーク・シートのセルA2からセルA14に以下の図5-9のように、セルA1に、セルB1に $f(x)$ と書き込んだ上で、-3から+3間の数値を0.5刻みで入力する

	A	B
1	x	f(x)
2	-3	
3	-2.5	
4	-2	
5	-1.5	
6	-1	
7	-0.5	
8	0	
9	0.5	
10	1	
11	1.5	
12	2	
13	2.5	
14	3	

図5-9 正規分布作業(1)

# 【Excel】 正規分布の確率密度と累積分布

- セルB2に「=NORM.DIST(A2,0,1,0)」と入力し、セルB14までコピーすると図5-10のように結果が表示される

	A	B
1	x	f(x)
2	-3	0.004432
3	-2.5	0.017528
4	-2	0.053991
5	-1.5	0.129518
6	-1	0.241971
7	-0.5	0.352065
8	0	0.398942
9	0.5	0.352065
10	1	0.241971
11	1.5	0.129518
12	2	0.053991
13	2.5	0.017528
14	3	0.004432

図5-9 正規分布作業(1)

# 【Excel】 正規分布の確率密度と累積分布

---

- NORM.DISTの括弧内には4種類のパラメータを入力することになるが,
  1. 1番目のパラメータは, 密度関数の値を計算する正規分布に従う変数の値あるいは値が入力されているセル番地
  2. 2番目は正規分布の平均
  3. 3番目は正規分布の標準偏差
  4. 4番目は0か1
    - 0の場合には通常密度関数の値を表示することを,
    - 1の場合には累積分布関数の値を表示することを指示している

# 【Excel】 正規分布の確率密度と累積分布

- セルC2に「=NORM.DIST(A2,0,1,1)」を入力して計算すると、結果は図5-11のようになる

	A	B	C
1	x	f(x)	af(x)
2	-3	0.004432	0.00135
3	-2.5	0.017528	0.00621
4	-2	0.053991	0.02275
5	-1.5	0.129518	0.066807
6	-1	0.241971	0.158655
7	-0.5	0.352065	0.308538
8	0	0.398942	0.5
9	0.5	0.352065	0.691462
10	1	0.241971	0.841345
11	1.5	0.129518	0.933193
12	2	0.053991	0.97725
13	2.5	0.017528	0.99379
14	3	0.004432	0.99865

図5-11 正規分布作業(3)

# 【Excel】 正規分布の確率密度と累積分布

- 列Bのデータによって折れ線グラフを作成したのが図5-12である

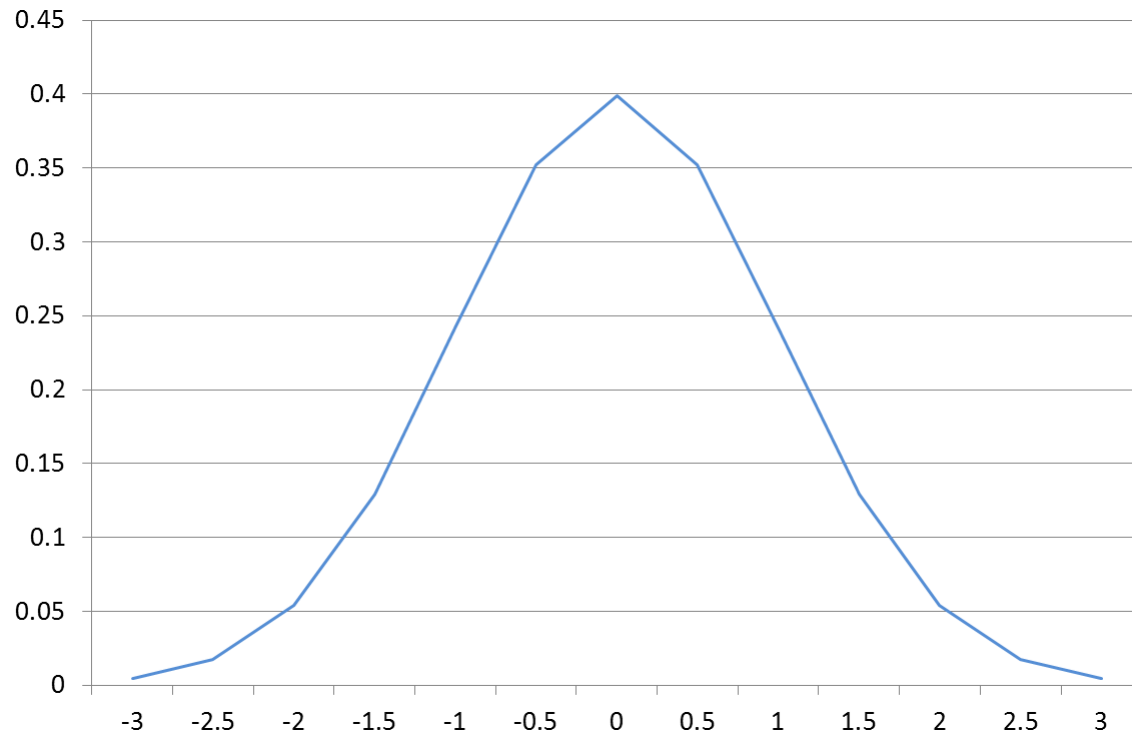


図5-12 正規分布作業(4)



# 【Excel】 正規分布の確率密度と累積分布

- 列Cのデータによって折れ線グラフを作成したのが図5-13である
  - 累積分布関数は各 $x$ 以下の確率密度関数の値の合計値である
  - 数学的には, 密度関数を $-\infty$ から $x$ まで積分した値になっている
  - 密度関数を $-\infty$ から $\infty$ まで積分すればその値は1になる

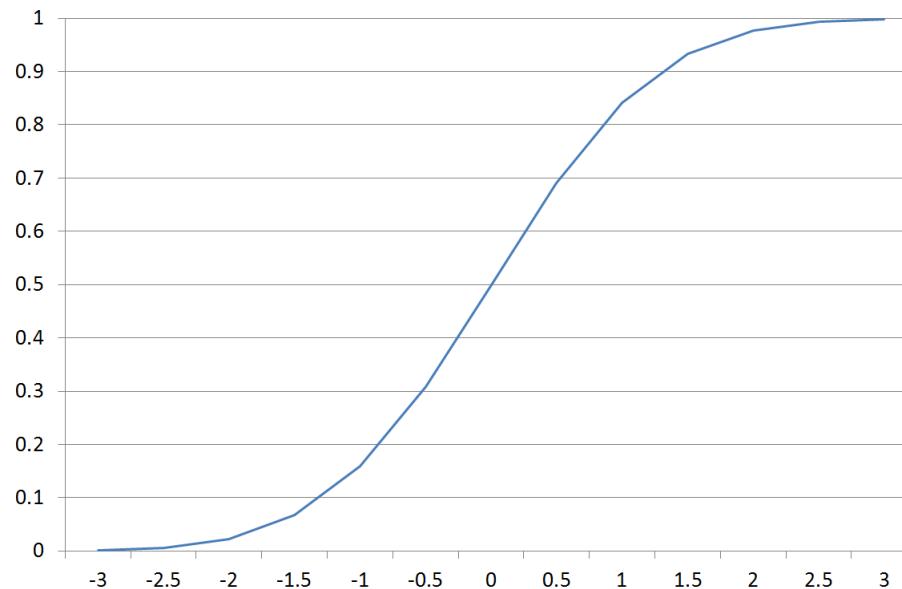


図5-13 正規分布作業(5)

# 【Excel】 正規分布の確率密度と累積分布

- グラフは全体としてギザギザした印象を受けるので、-3.5から+3.5の範囲を0.25刻みで取ってグラフを描く
  - このように調整して示したのが図5-14及び5-15である
  - 点の取り方を多くすることによってグラフがなめらかになる

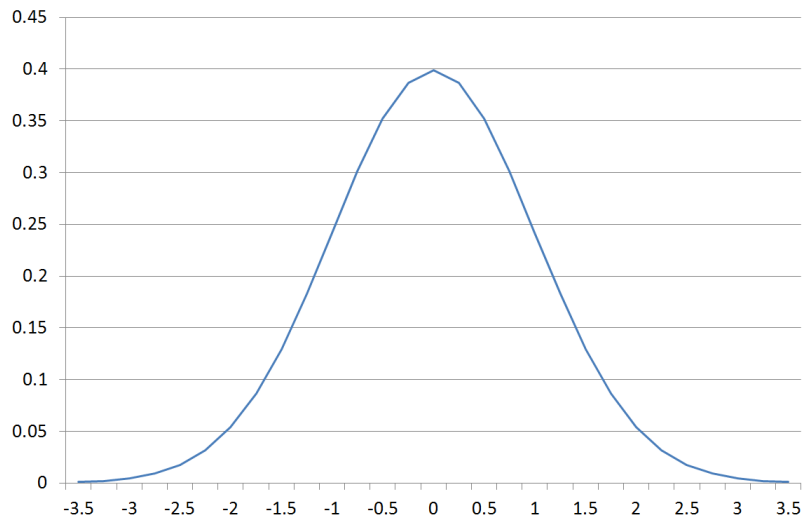


図5-14 正規分布作業(6)

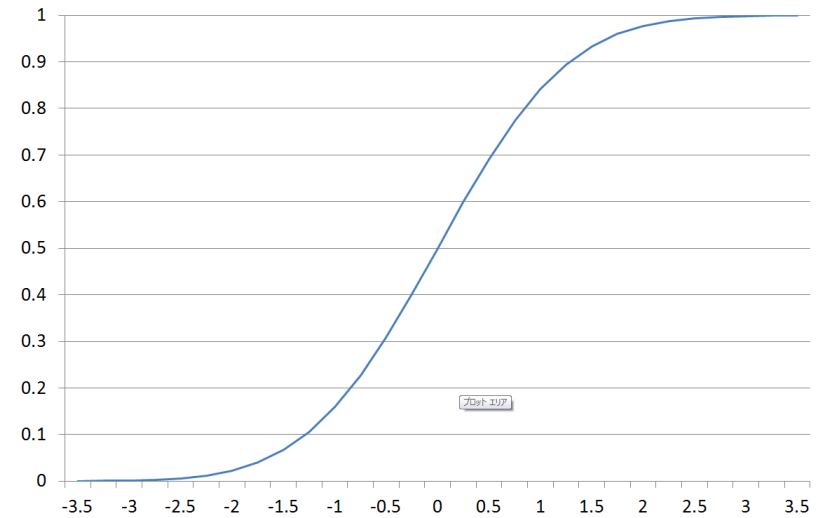



図5-15 正規分布作業(7)

# 【Excel】 正規乱数の生成

次に正規分布に従う乱数を作成する分析ツールを利用する方法を説明

- ワーク・シートに246個のデータによる10組の正規乱数を出力する場所を用意した上で、データ ,  データ分析 を選択し、乱数発生をクリックして表示された画面で、図5-19に示す形で入力すればよい
  - セルH2からセルH247に通し番号を入力し、セルI1からセルR1にケース1からケース10を入力した上で、セルI2からセルR247に正規乱数を作成することを前提に説明をする

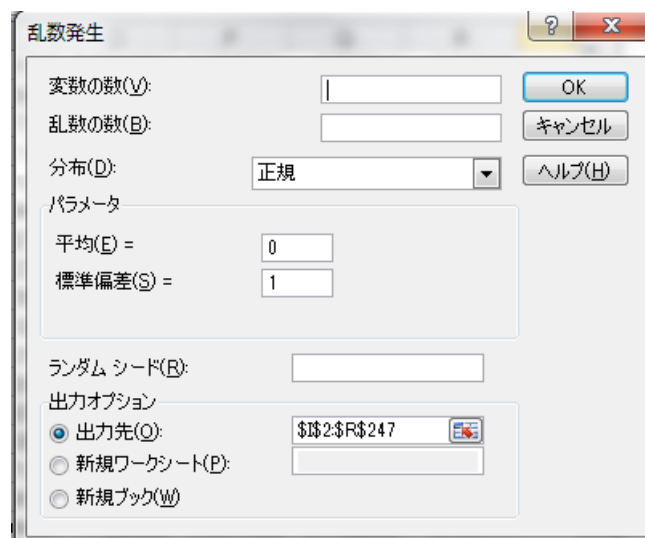
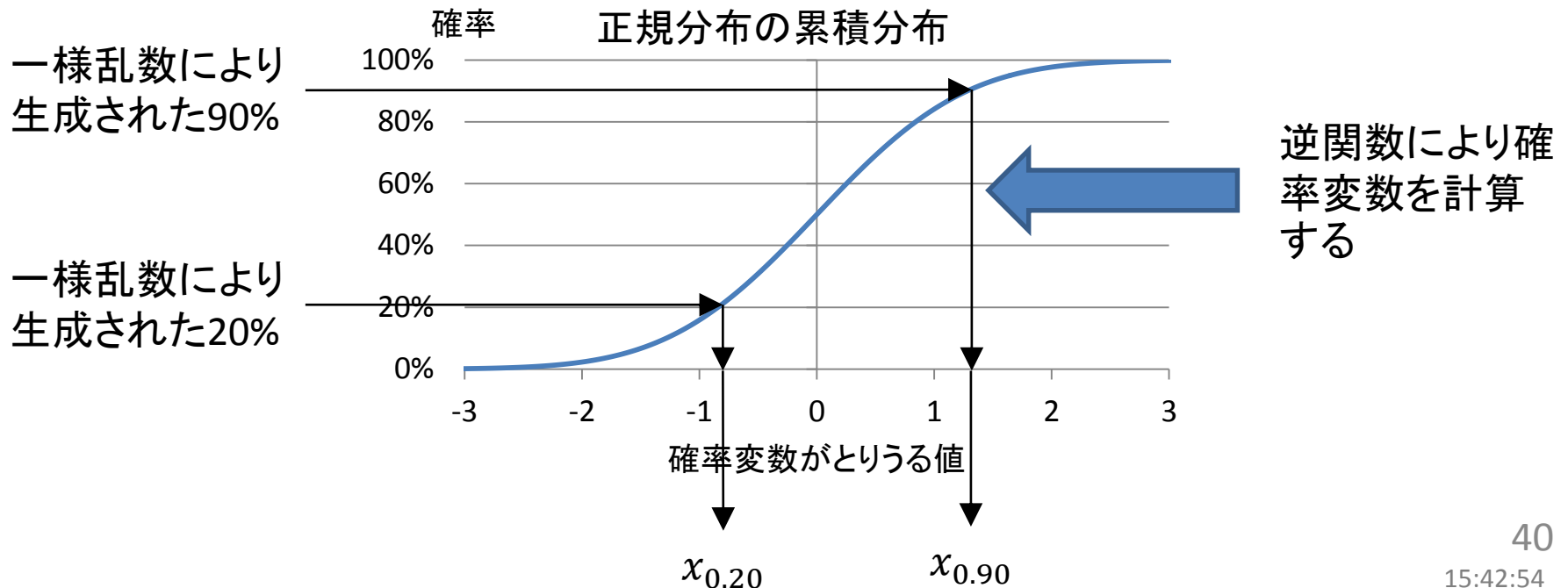


図5-19 正規分布作業(11)

# 【Excel】 正規乱数の生成

RAND関数を使って、正規乱数を生成させることができる

- セルI2「=NORM.INV(RAND(),0,1)」を入力し、その他のセルにコピーする
- NORM.INV関数は正規分布の累積分布関数の逆関数である
  - 第1引数は確率
  - 第2引数は平均で、ここでは0とした
  - 第3引数は標準偏差で、ここでは1とした(従って、標準正規乱数を生成している)



# 【Excel】 正規乱数の生成

---

- 標準正規分布を前提なので、「平均(E)」は0,「標準偏差(S)」は1である
  - 分散が1なので, この平方根である標準偏差も1である
- 得られたサンプルが標準正規分布の条件を満たしているかどうかをケースごとに平均と分散(標準偏差)を計算して確認する
  - 10ケースの全データの平均と分散, ケースごとのヒストグラムと全てのデータを対象としたヒストグラムによって確認
  - 平均と分散の計算結果とケース1及び全データを利用したヒストグラムの作成例を示し, 残りは練習問題とする

# 【Excel】正規乱数の生成

- ケースごと及び全データに関して求めた平均、分散及び標準偏差は図5-20に示す通りである
  - 乱数の性質上、皆さんが計算して求めた値と数値的に一致していないことも十分ありうる
  - ケースごとの平均や分散と比べて、全データを用いて計算した平均と分散の方が標準正規分布の条件に近い

	H	I	J	K	L	M	N	O	P	Q	R	S
1		ケース1	ケース2	ケース3	ケース4	ケース5	ケース6	ケース7	ケース8	ケース9	ケース10	全データ
247	246	-0.01304	1.252481	-1.09586	-1.62401	2.125926	1.050814	0.421373	-0.6938	-1.004	1.215626	
248	平均	0.036345	0.057567	0.033266	0.096105	-0.07863	0.167814	0.077235	0.054088	-0.03926	-0.04007	0.036446
249	分散	1.021613	1.012742	1.113868	0.901646	1.140329	1.011526	1.060902	1.117725	1.150076	1.031571	1.057144
250	標準偏差	1.010749	1.006351	1.0554	0.949551	1.067862	1.005747	1.030001	1.057225	1.072416	1.015663	1.028175

図5-20 正規分布作業(12)

# 【Excel】 正規乱数の生成

- ケースごとのヒストグラムを描いてみる
  - セルA2からセルA13に-3から+3の0.5刻みの区分が入力されている前提でヒストグラムをケース1に関して作成する
  - ヒストグラム作成画面での指定は図5-21に示してある

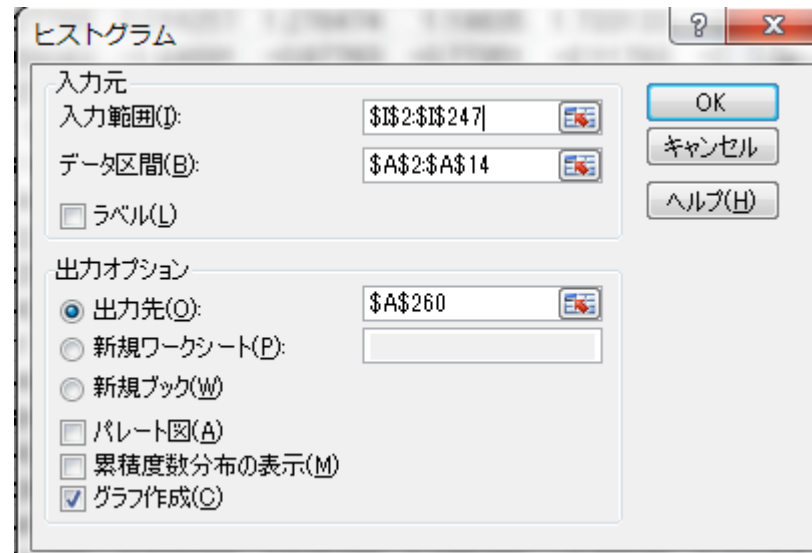


図5-21 正規分布作業(13)

# 【Excel】 正規乱数の生成

---

- 結果は図5-23にケース1に関するヒストグラムを-3から+3の0.5刻みの区分で示す

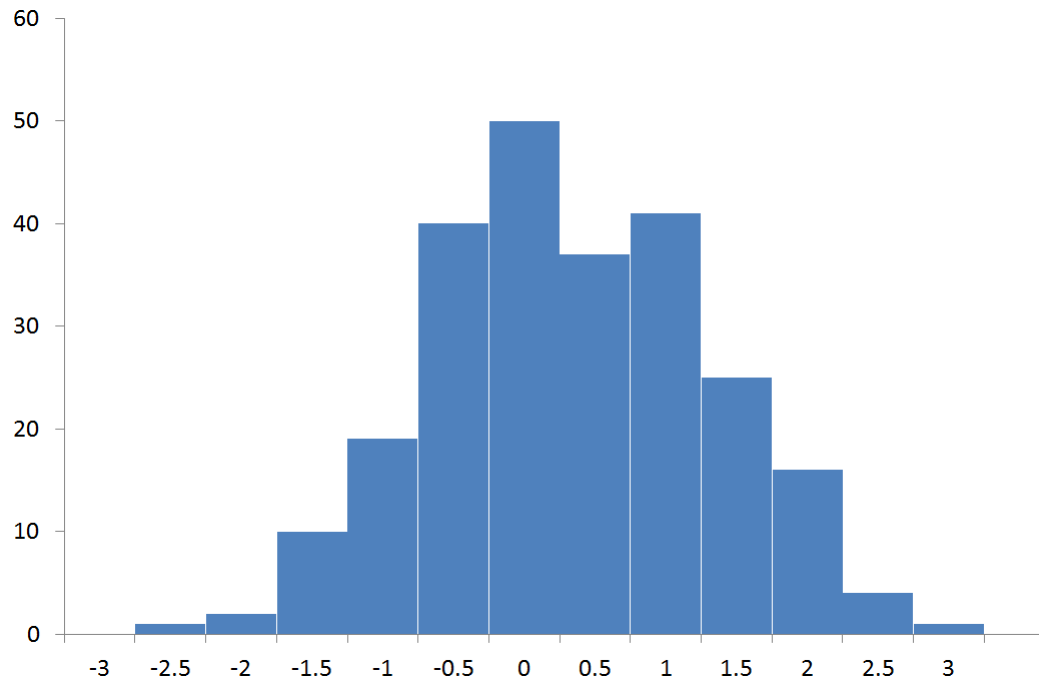


図5-23 正規分布作業(15)



# 【Excel】 正規乱数の生成

- ケースごとのヒストグラムを描いてみる
  - セルE2からセルE30に-3.5から+3.5に0.25刻みの区分が入力されている前提でヒストグラムをケース1に関して作成する
  - ヒストグラム作成画面での指定は、図5-22に示してある

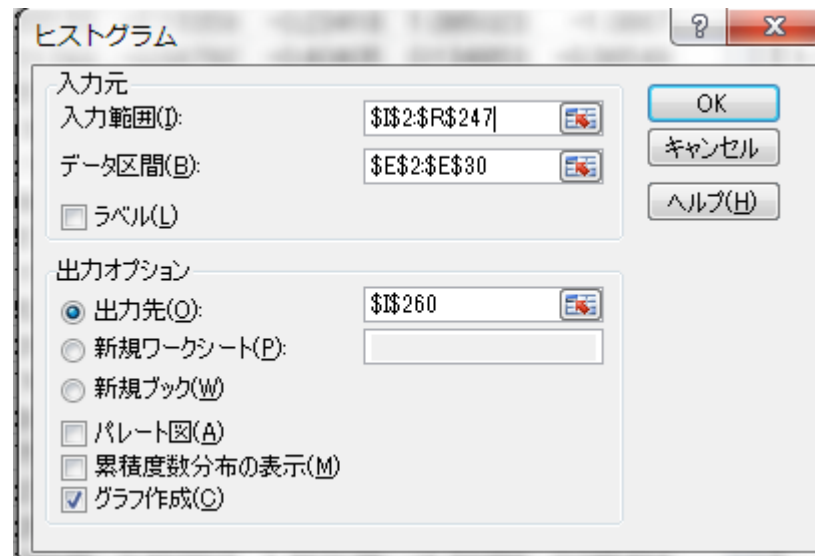


図5-22 正規分布作業(14)

# 【Excel】 正規乱数の生成

- 結果は図5-24にケース1に関するヒストグラムを-3.5から+3.5に0.25刻みの区分で示す

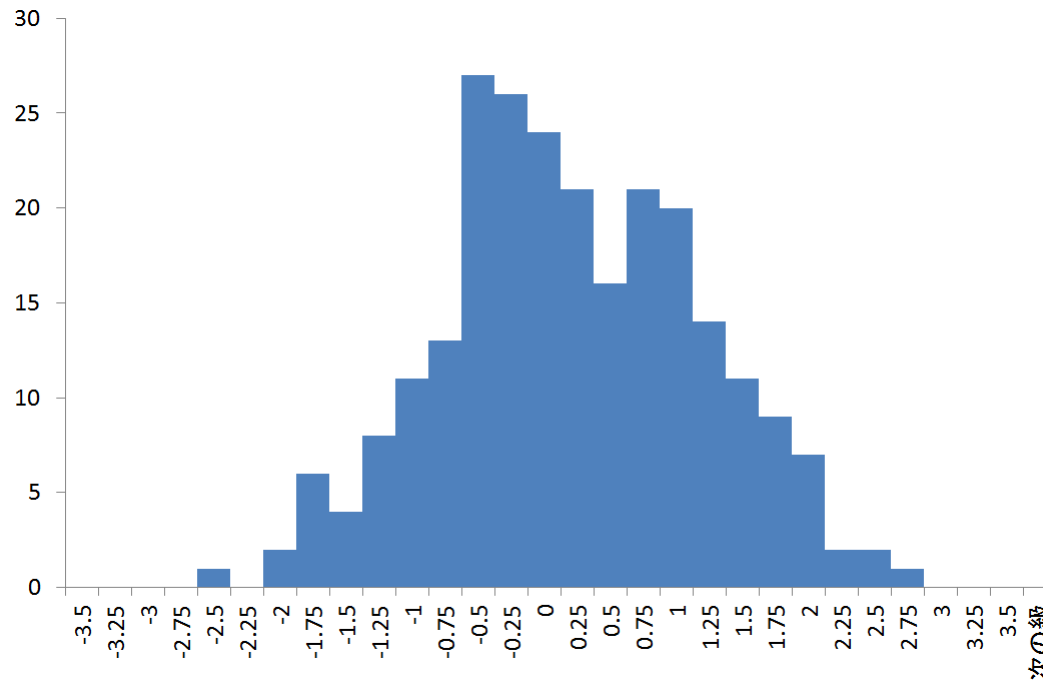


図5-24 正規分布作業(16)

# 【Excel】 正規乱数の生成

- 全データを対象とした同様のヒストグラムが図5-25(-3から+3の0.5刻みの区分)と5-26(-3.5から+3.5に0.25刻みの区分)に示す

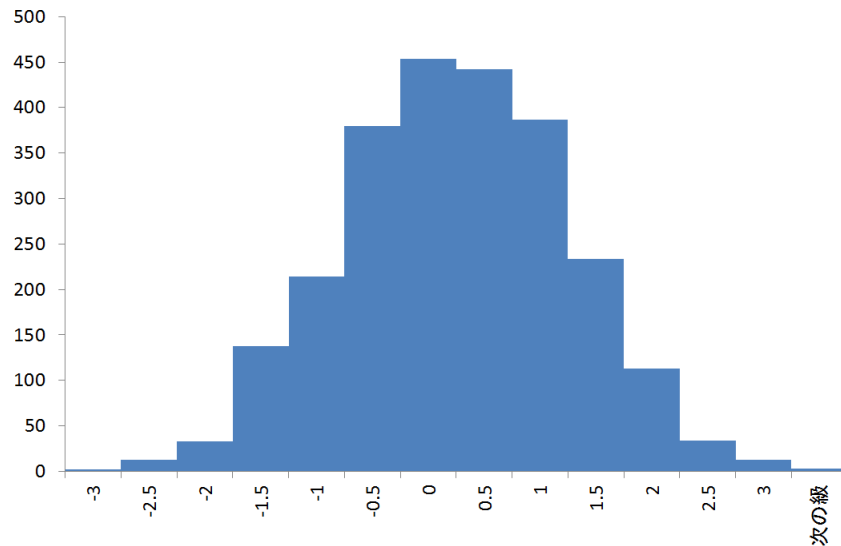


図5-25 正規分布作業(17)

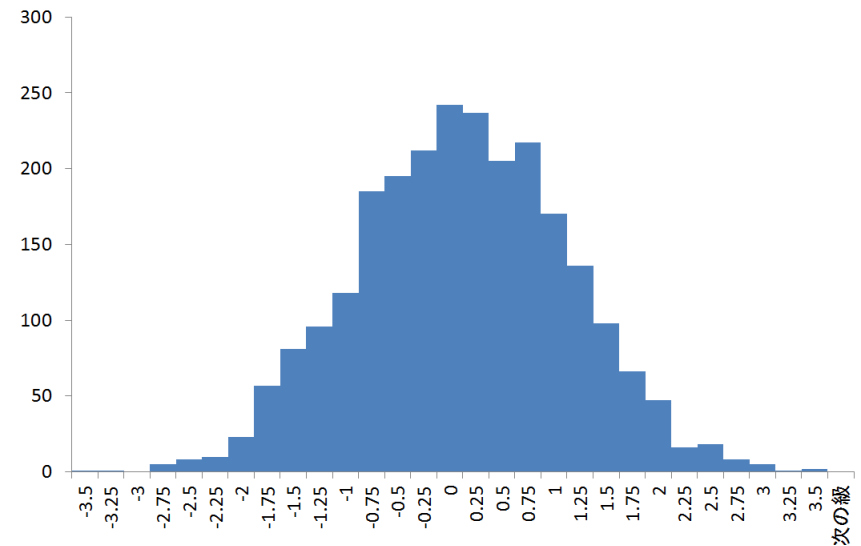


図5-26 正規分布作業(18)

# 【Excel】 正規乱数の生成

---

- 刻みを細かくとった方が滑らかなグラフになっているわけではなく、刻みが荒い方が滑らかな印象を受ける
  - ケース1のデータのみを用いた場合には極端に出る
  - 全データを用いた場合にも同様の印象を受ける
- ⇒ 刻みの取り方は難しい
- ⇒ 理論的な正規分布の確率密度関数とここに示したヒストグラムの形状とは相当異なるので、現実のデータの分布を見た場合に正規分布と判断できるかどうかの評価も難しい

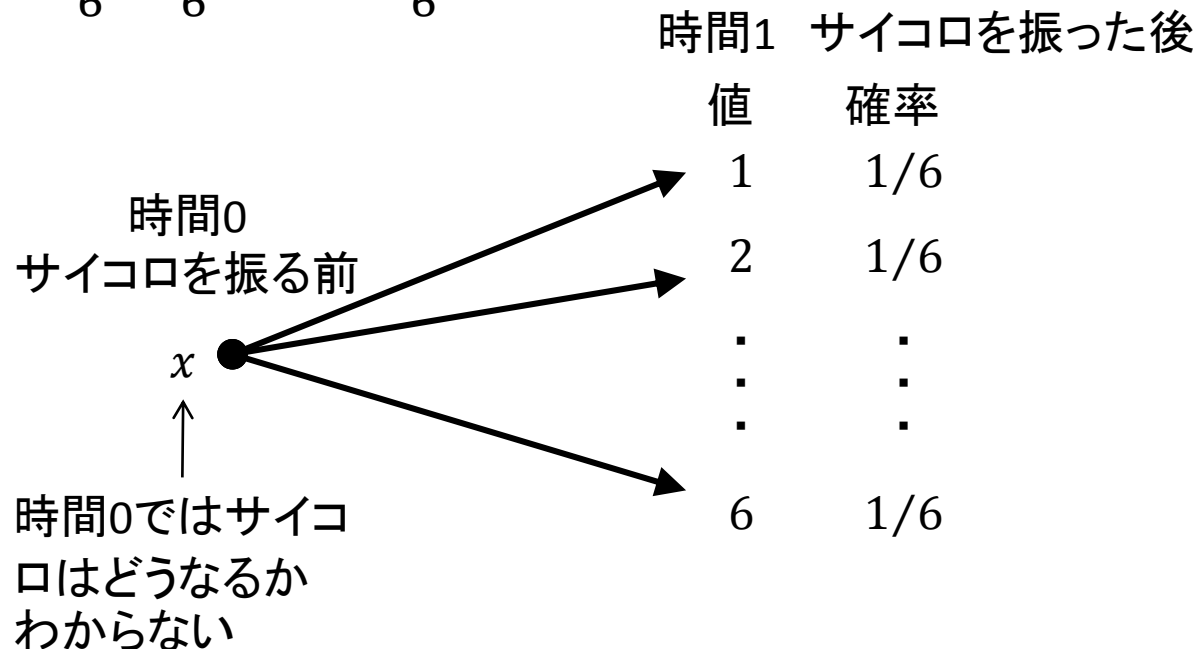
## 【練習問題5-2】 提出課題16:3点

---

- 10ケースの正規乱数のそれぞれの組に関して、荒い刻みのヒストグラムを作成し、平均と標準偏差の計算結果を表として、それぞれ、MS-Wordに貼り付けて提出せよ
- また、Excelファイルも提出せよ

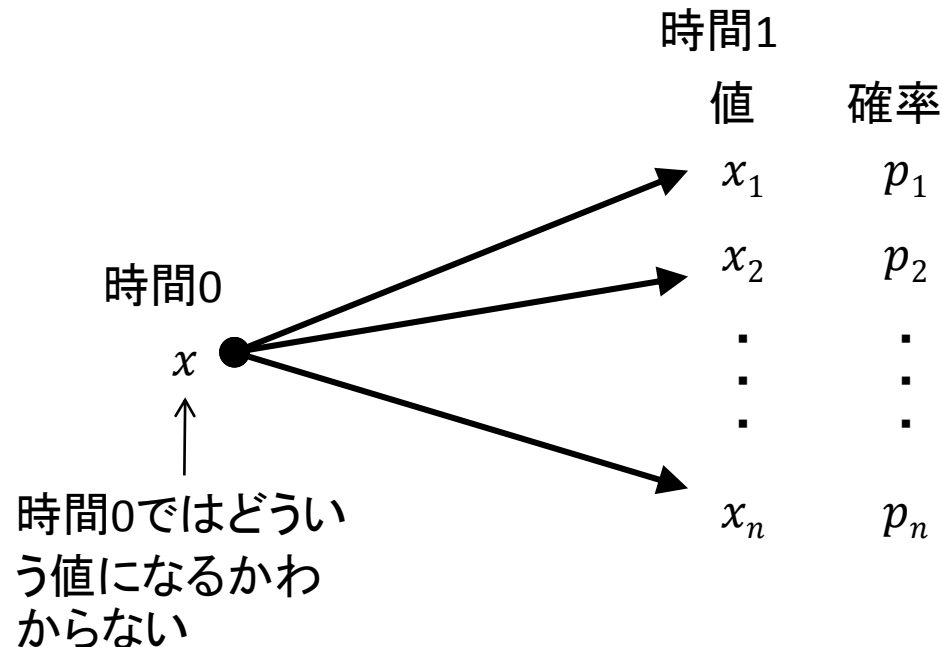
# 離散的な確率変数の例

- サイコロの目は
  - 1, 2, ..., 6の中から取り得る変数であり,
  - 1, 2, ..., 6はそれぞれの値が発生する確率 $\frac{1}{6}, \frac{1}{6}, \dots, \frac{1}{6}$ が対応している
  - 確率は $\frac{1}{6} \geq 0, \frac{1}{6} + \frac{1}{6} + \dots + \frac{1}{6} = 1$ を満たす



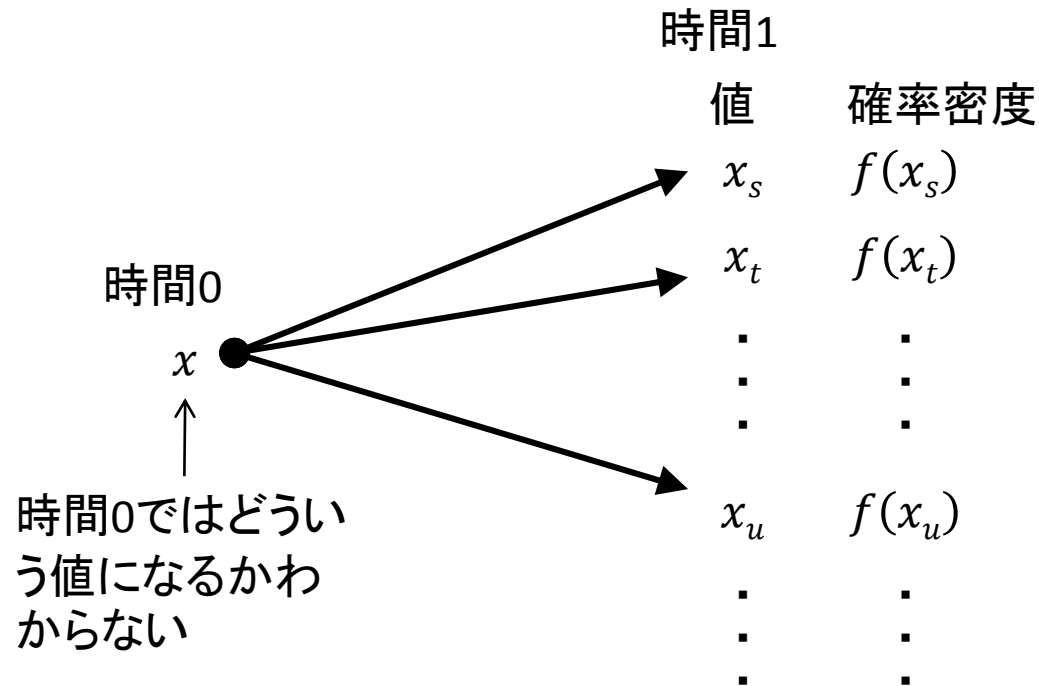
# 離散的な確率変数

- 離散的な確率変数 $x$ は,
  - $x_1, x_2, \dots, x_n$ の中から取り得る変数であり,
  - $x_1, x_2, \dots, x_n$ にはそれぞれの値が発生する確率 $p_1, p_2, \dots, p_n$ が対応している
  - 確率は $p_i \geq 0, p_1 + p_2 + \dots + p_n = 1$ を満たす



# 連続的な確率変数

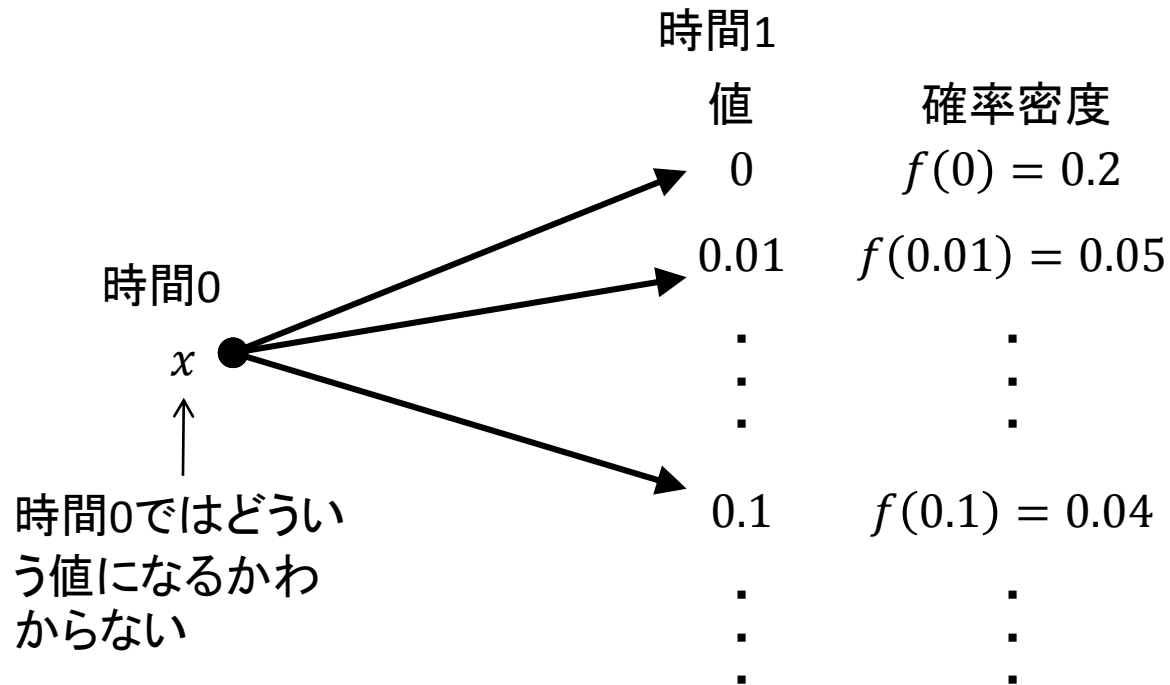
- 連続的な(実数値)確率変数 $x$ を離散的な確率変数の場合のアナロジーで説明すると,
  - 実数 $(-\infty, \infty)$ の中から取り得る変数であり,
  - 各実数値 $x$ にはそれぞれの値が発生する確率密度 $f(x)$ が対応している
  - 確率密度は $f(x) \geq 0, \int_{-\infty}^{\infty} f(x)dx = 1$ を満たす
  - 離散の場合の確率と連続の場合の確率密度は若干概念が異なる





# 連続的な確率変数のイメージ

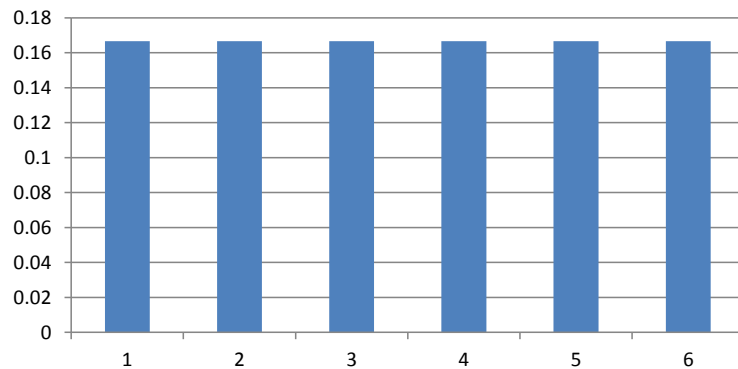
- 連続的な(実数値)確率変数 $x$ の例は以降に挙げる



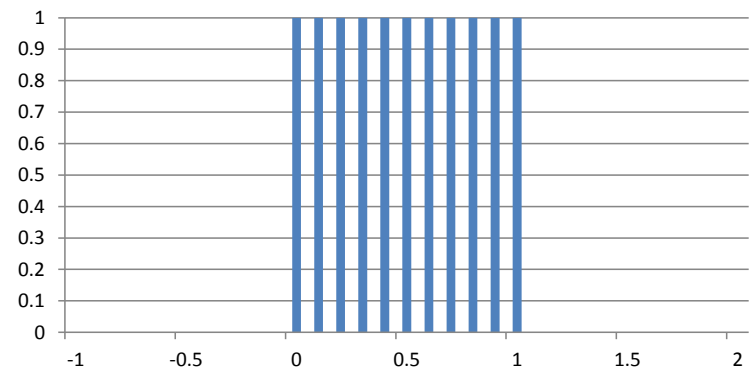
# 確率密度と累積分布

- 確率はいわゆる確率
- 確率密度は確率ではないが,  
確率  $\div$  確率密度  $\times$  確率変数を取る区間の幅
- 確率あるいは確率密度のグラフがヒストグラムに似ているのは偶然ではない

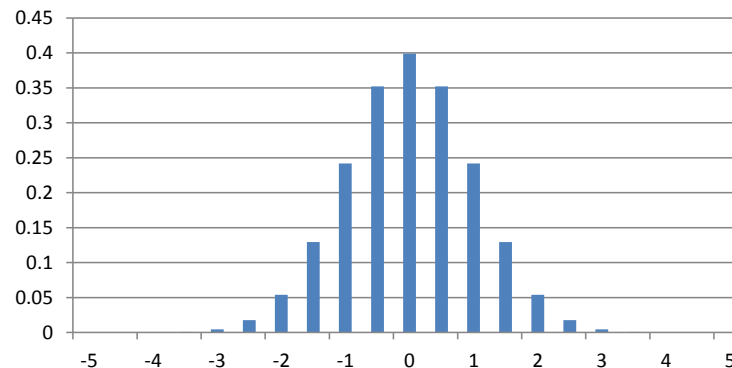
サイコロの目の確率



一様分布の確率密度



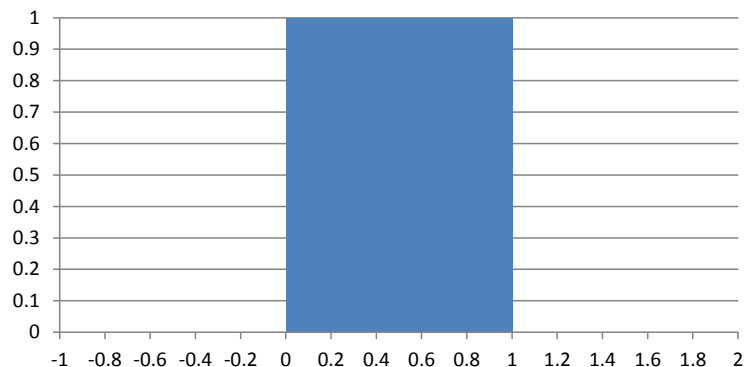
正規分布の確率密度



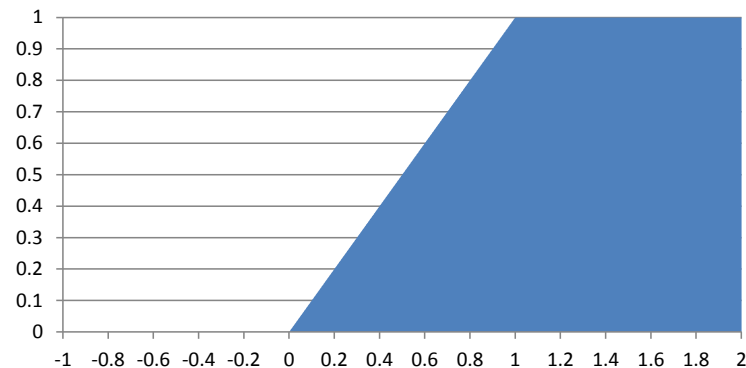
# 確率密度と累積分布

- もっと正確に言うと累積分布が確率
- 累積分布とは確率密度の面積
  - $\int_{-\infty}^y f(x)dx$ : ある点  $y$  においてそれまでの確率密度を足し合わせたもの

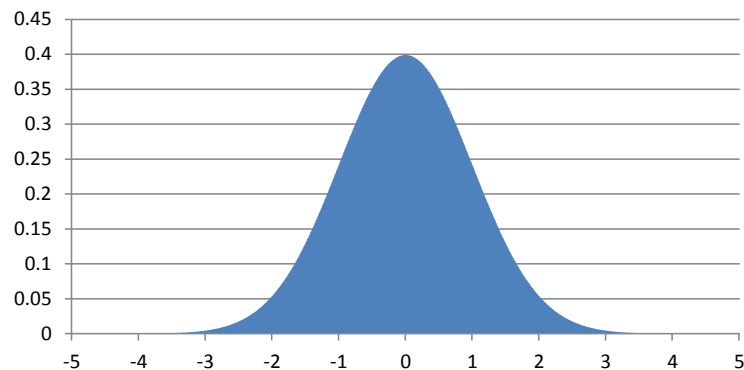
一様分布の確率密度



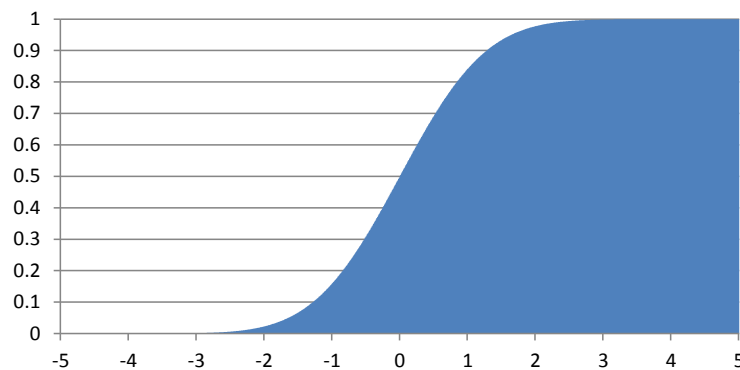
一様分布の累積分布



正規分布の確率密度



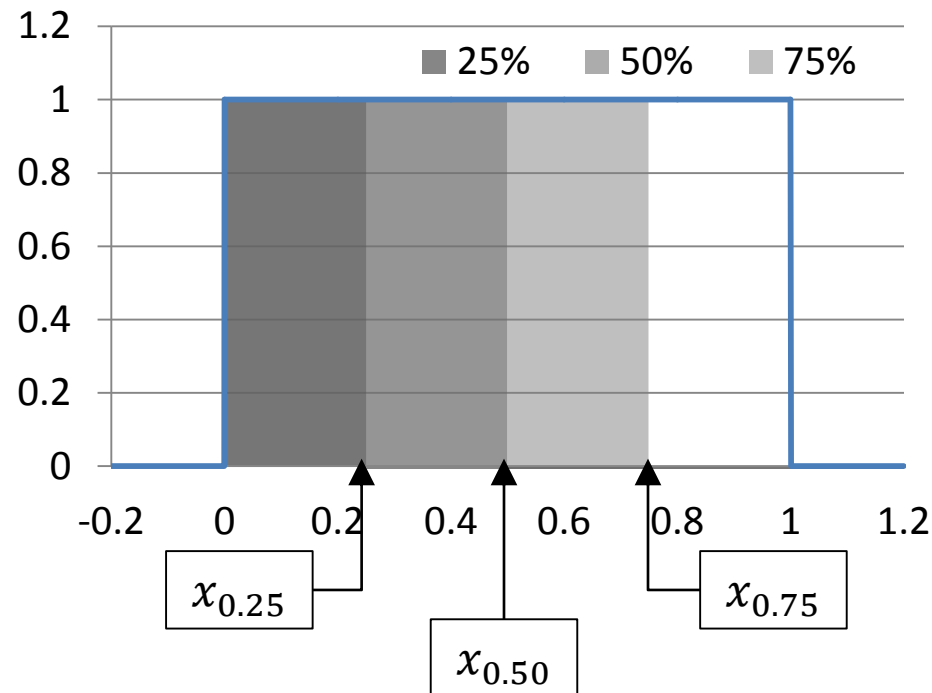
正規分布の累積分布



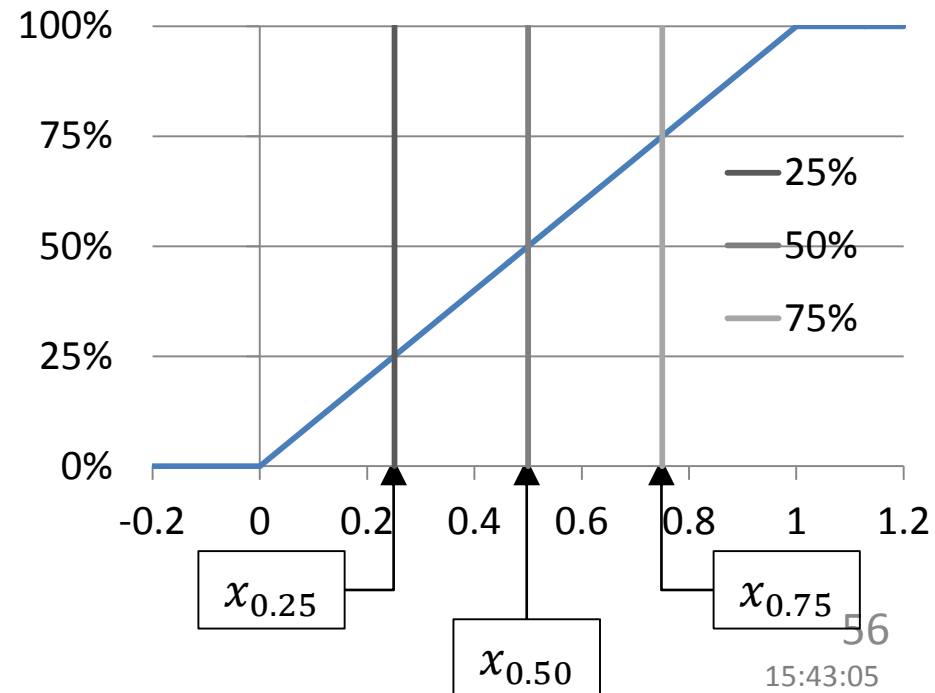
# 一様分布の確率密度と累積分布

- 確率密度関数の面積が確率
- $x$ と $x$ までの面積を計算した確率を、改めてプロットしたのが累積分布関数
  - $x_{0.25}$  (確率25%に対する $x$ )と25%
  - $x_{0.50}$  (確率50%に対する $x$ )と50%
  - $x_{0.75}$  (確率75%に対する $x$ )と75%

確率密度関数



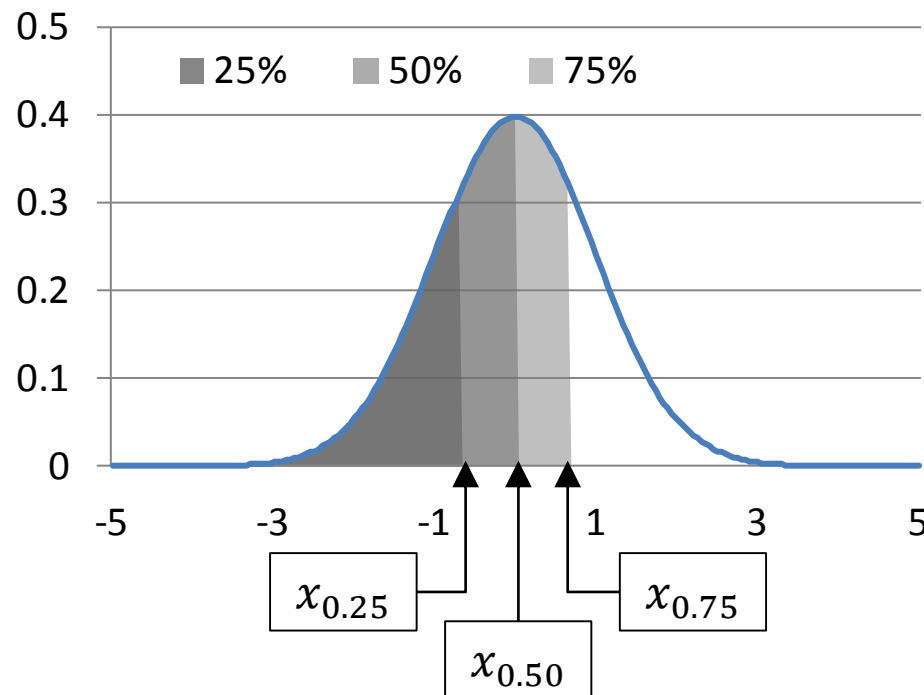
累積分布関数



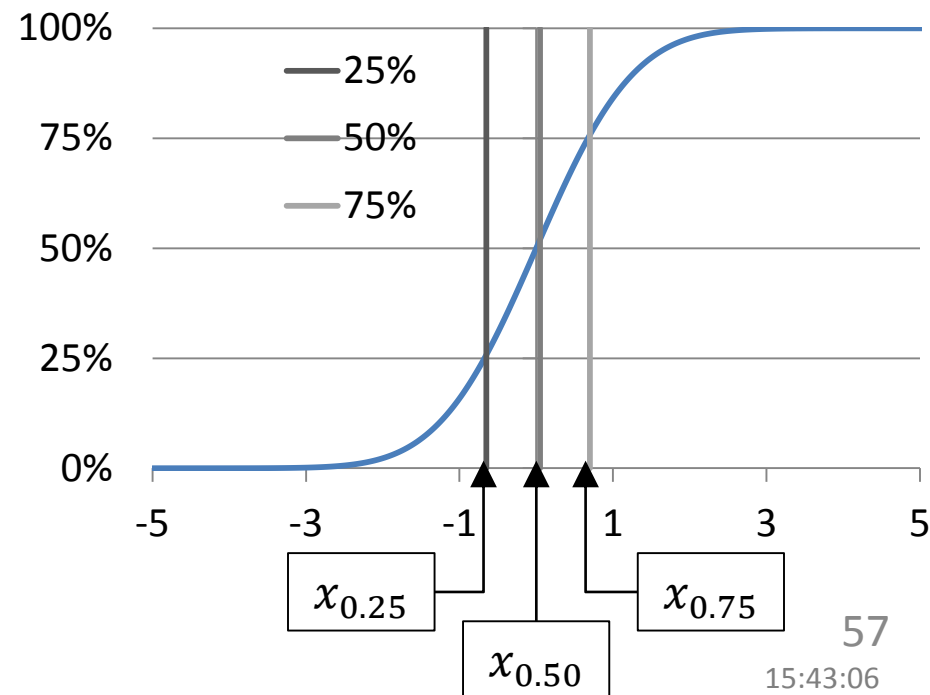
# 正規分布の確率密度と累積分布

- 確率密度関数の面積が確率
- $x$ と $x$ までの面積を計算した確率を、改めてプロットしたのが累積分布関数
  - $x_{0.25}$  (確率25%に対する $x$ )と25%
  - $x_{0.50}$  (確率50%に対する $x$ )と50%
  - $x_{0.75}$  (確率75%に対する $x$ )と75%

確率密度関数



累積分布関数



# 一般の正規分布と標準正規分布の関係

---

- 前節では, Excelの正規乱数発生プロセスによって発生させたデータの分布と正規分布の理論分布の形状について説明した
  - 確率分布のイメージが形成できた

# 一般の正規分布と標準正規分布の関係

---

- 標準正規分布以外の正規分布の形状を確認し、一般の正規分布を標準正規分布に変換するための方法を学ぶ
  - 正規分布の平均が変化する時には、分布の山の位置(分布の中心)を左右に移動させるだけ
  - 分散が変化した場合に正規分布の形状がどのように変化するのかを理論的に計算される正規分布によって確認する

# 【Excel】 分散が異なる正規分布の 確率密度と累積分布の確認

---

- 標準偏差が, 0.5, 1, 2の場合に付いて正規分布の確率密度の分布を求め, これをグラフ化して形状の相違を確認する
- -3.5から+3.5の範囲で, 刻みを0.25として, 各標準偏差に対応する確率密度の分布を求める



# 【Excel】 分散が異なる正規分布の 確率密度と累積分布の確認

---

1. 新しいワーク・シートのセルA1に「x」を入力し, セルB1に「STD=0.5」, セルC1に「STD=1.0」, セルD1に「STD=2.0」を入力
  - A2からA30まで, -3.5から+3.5の範囲で, 刻みを0.25として入力する
2. セルB2に「=NORM.DIST(A2,0,0.5,0)」を入力して計算し, セルB30までコピー
  - 以下同様に列Cに, 標準偏差が1の場合, 列Dに標準偏差が2の場合の正規確率密度を計算
3. セルB2からセルD30をクリック・アンド・ドラッグして白黒反転した上でマウスを右クリックする
4. セルの「書式設定(F)」をクリックして表示される画面で, 表示形式で, 「数値」を選択し, 小数点以下の桁数に10を入力して桁をそろえる

# 【Excel】 分散が異なる正規分布の 確率密度と累積分布の確認

- 折れ線グラフを描いたのが図5-27である

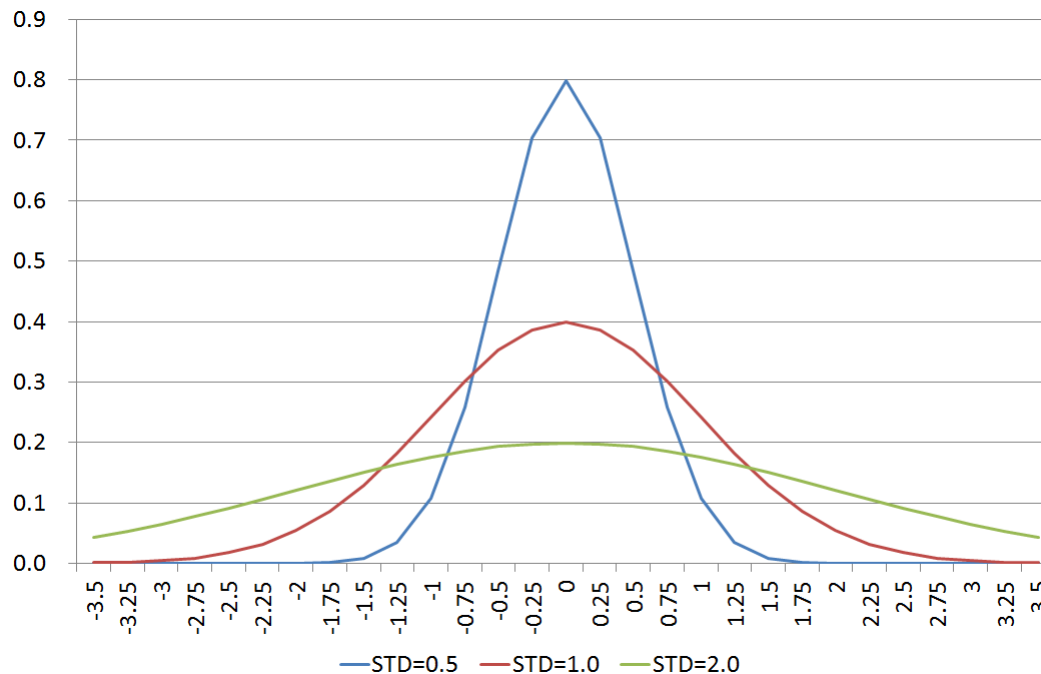


図5-27 正規分布作業(1)

# 【Excel】 分散が異なる正規分布の 確率密度と累積分布の確認

- 列E, 列F, 列Gに累積分布関数を作成し, これらを折れ線グラフで示したのが図5-28である
  - 「=NORM.DIST(A2,0,0.5,1)」のように最後のパラメータを1にすると示される

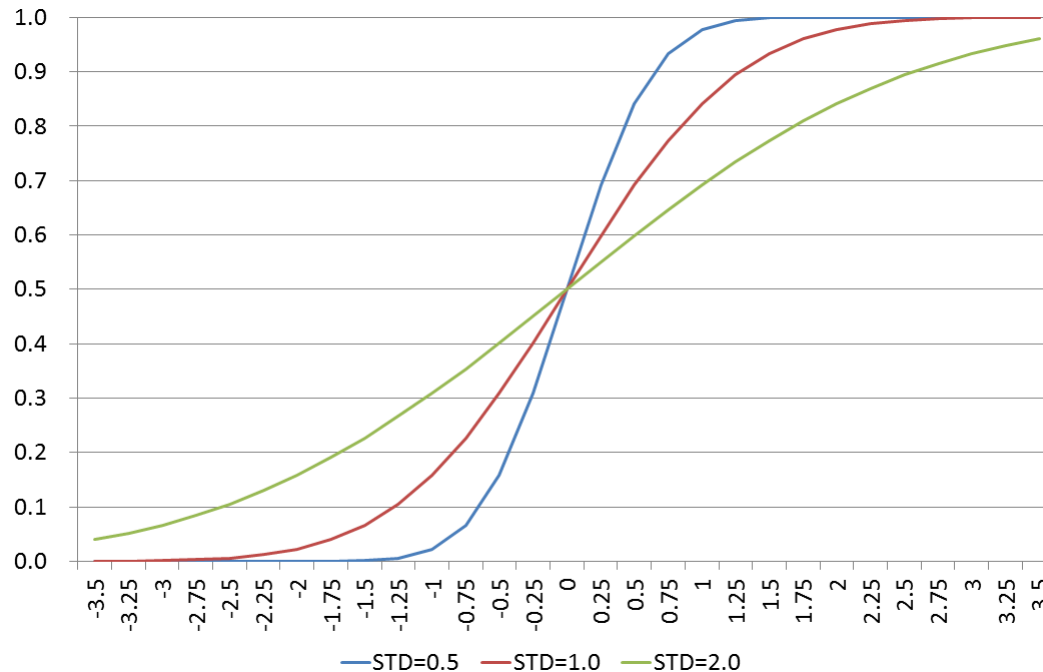


図5-28 正規分布作業(2)

# 【Excel】 分散が異なる正規分布の 確率密度と累積分布の確認

- 平均が共通の0であっても、分散が異なることによって、正規分布に従う確率変数の散らばり方の程度は非常に大きく異なる
- 標準偏差が2の正規分布の累積分布関数はほぼ直線的に上昇しており、平均の周辺に山が存在することが不明確である

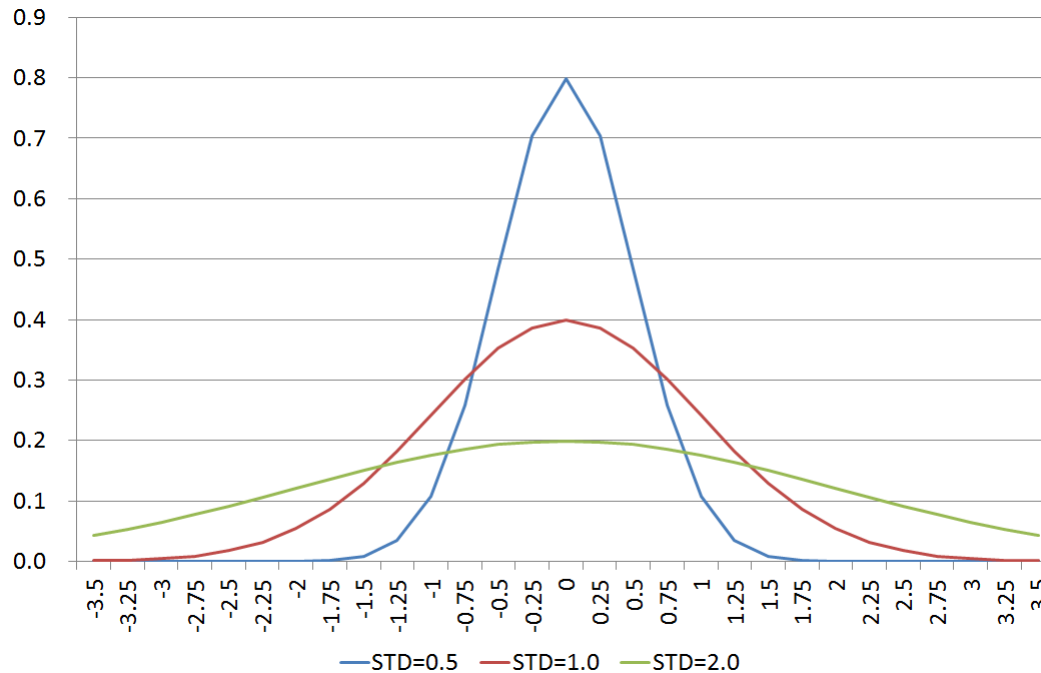


図5-27 正規分布作業(1)

# 【Excel】 正規乱数の標準化

---

- 一口に正規分布と言ってもパラメータの値が異なることにより、分布の形状は多様に変化する
  - 現実のデータを眺めた場合、その形状が正規分布であるのかどうかを判断することは簡単でない
- しかし、標準正規分布に関してはその形状の幅はある程度限定することが可能
  - 上と同様に平均が0で標準偏差が0.5, 1.0, 2.0であることを条件に,
    1. 正規乱数を発生させ、これらの分布の形状と,
    2. 公式に従って標準化(正規分布から標準正規分布への変換)した分布の形状とを見比べ、一般の正規分布を標準化することの意味を確認

# 【Excel】 正規乱数の標準化

---

1. 列Iにヒストグラムを描く際の区間を入力する
  - 分散の取り方によって、データの分布が広がるので、-3.5から+3.5の範囲
  - 正規乱数をベースにヒストグラムを描く際、区間の幅を狭く取り過ぎると分布の形状がなめらかにならないので、ここで区間の幅を0.5刻みとする

⇒ セルI1に「x」を入力した上で、セルI2からセルI16の範囲に-3.5から0.5刻みで+3.5までの数値を入力する
2. セルJ2からセルJ247に1から246までの通し番号を入力
3. セルK1に「STD=0.5」、セルL1に「STD=1.0」、セルM1に「STD=2.0」と書き込み、先に標準正規分布に従う乱数を発生させた場合と同様の方法で,
  - a. セルK2からセルK247に平均0, 標準偏差0.5の正規乱数,
  - b. セルL2からセルL247に平均0, 標準偏差1.0の正規乱数,
  - c. セルM2からセルM247に平均0, 標準偏差2の正規乱数を発生させる

# 【Excel】 正規乱数の標準化

- 上で求めた正規乱数の平均, 分散, 標準偏差をデータごとに計算する
  - 求めた平均や標準偏差は, 正規乱数の性質を確認するだけでなく, これらの乱数を標準正規分布に変換するためにも利用する
- 標準偏差が0.5の場合のヒストグラムが図5-29

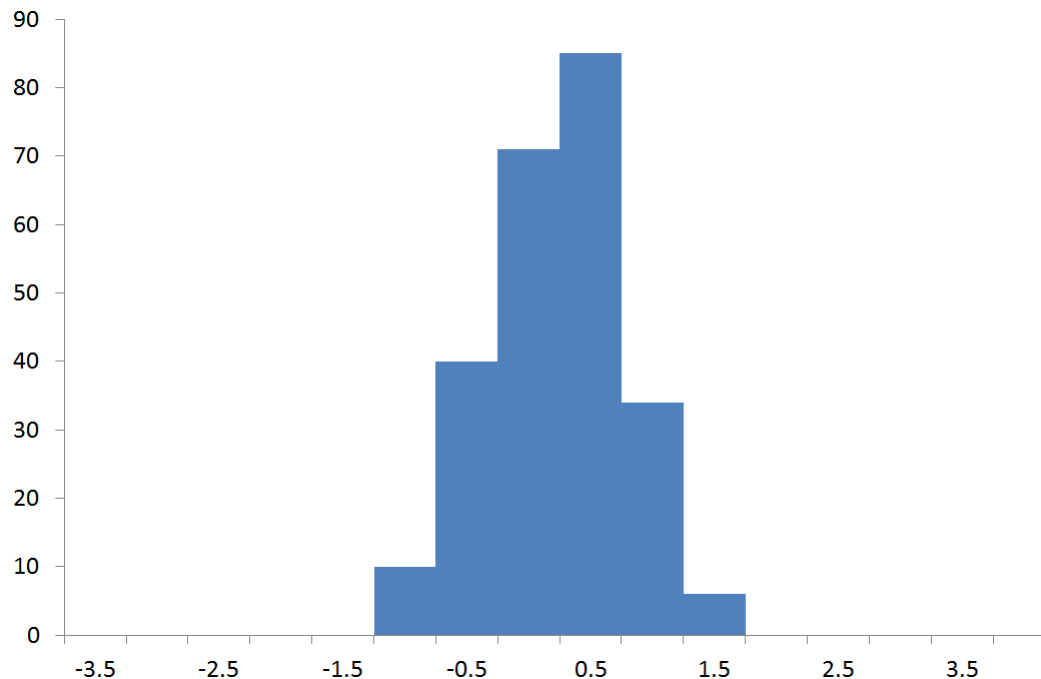


図5-29 正規分布作業(3)

# 【Excel】 正規乱数の標準化

---

- 標準偏差が1.0の場合のヒストグラムが図5-30

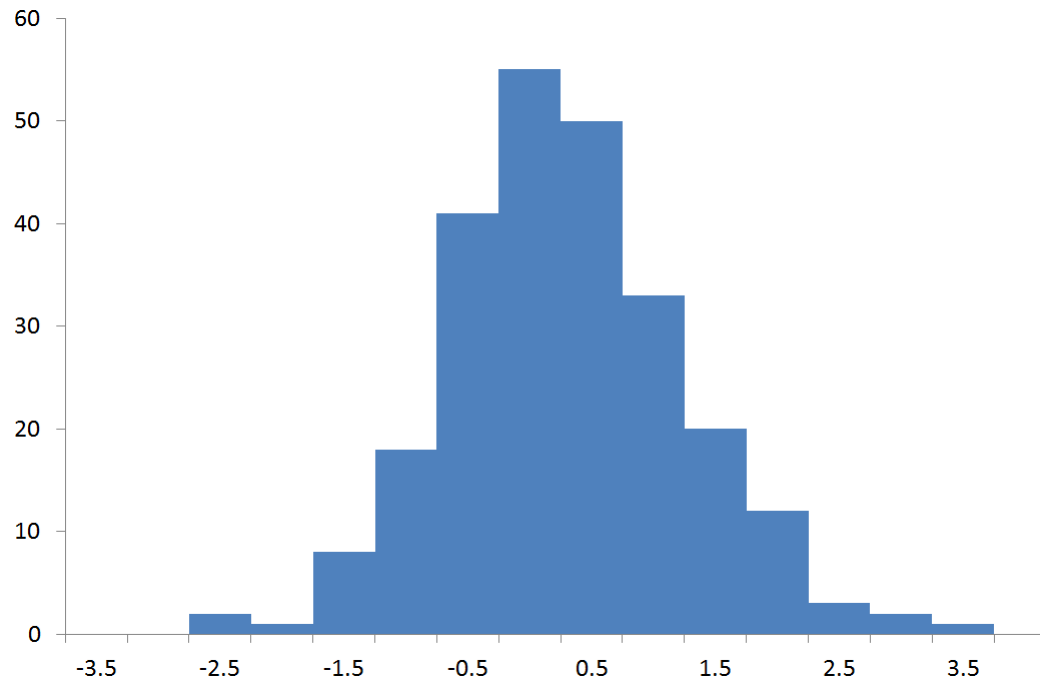


図5-30 正規分布作業(4)



# 【Excel】 正規乱数の標準化

- 標準偏差が2の場合のヒストグラムが図5-31である

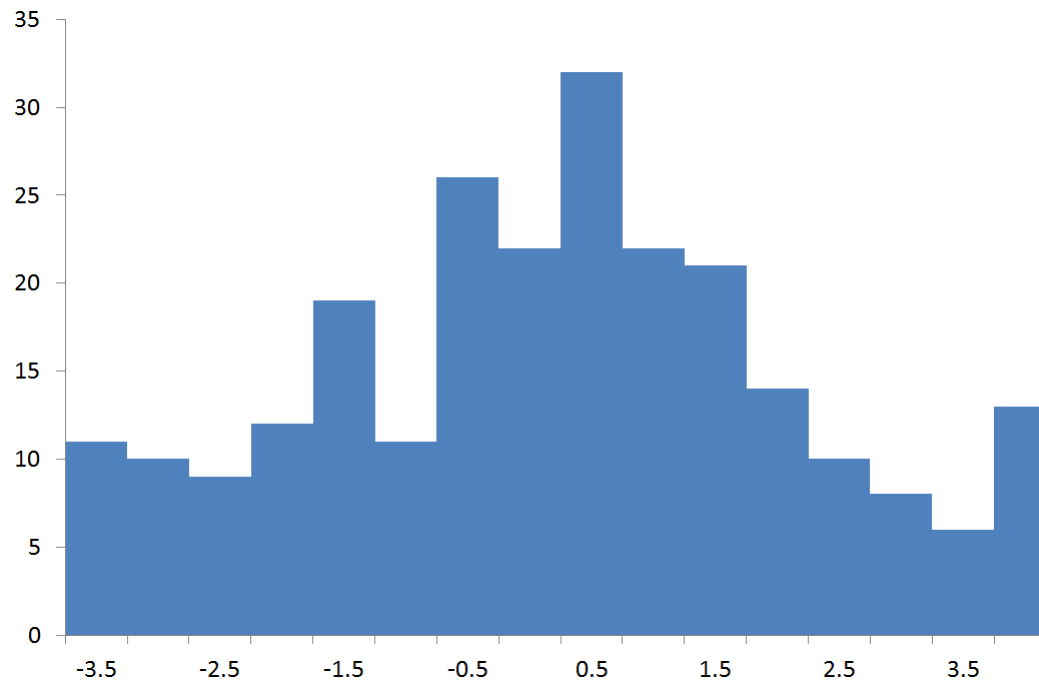


図5-31 正規分布作業(5)

# 【Excel】 正規乱数の標準化

---

- 発生させた正規乱数の分布は,
  - 標準偏差が0.5の場合には, -1.5から+1.5の間に入っており,
  - 標準偏差が1.0の場合には, ここで想定した-3.5から+3.5の間にほぼ入っているのに対して,
  - 標準偏差が2.0の場合には, ここで想定した範囲よりも幅を広く取った方が滑らかなグラフを描くことが可能である

# 【Excel】 正規乱数の標準化

---

- ここで用いた3種類の標準偏差に従う正規乱数を標準正規分布に変換して再度ヒストグラムを描き、標準偏差の違いがヒストグラムの形状に影響を与えなくなるかどうかを確認する
- 図5-29から図5-31に示されたヒストグラムの平均や分散を求め、それらの分布が正規分布に従っていることを前提として、標準正規分布に変換する
  - ただし、各種のデータを観察する場合に、先験的に分布の形状がわかっていることはほとんどない
  - 変換されたデータが標準正規分布に従っていることを前提として各種の統計処理を行う

# データの標準化

---

- 列Kから列Mのデータを標準正規分布に変換するため、以下の関係式を用いる

$$z = \frac{x - \hat{\mu}}{\hat{\sigma}} \quad (5-3)$$

- $x$ は正規分布に従うと想定されている確率変数の値
- $z$ は正規確率変数 $x$ を標準正規分布に従う確率変数に変換した値
- $\hat{\mu}$ は観測されたデータの平均値(母集団平均の推定値)
- $\hat{\sigma}$ は観測されたデータの標準偏差(母集団標準偏差の推定値)

# 【Excel】 正規乱数の標準化

---

上に記した変換公式を用いて、列Kから列Mのデータを標準正規分布に変換する

- セルP1に「SND(0.5)」, セルQ1に「SND(1.0)」, セルR1に「SND(2.0)」と書き込む

- セルP2に,

「 $= (K2 - \text{AVERAGE}(K\$2:K\$247)) / \text{STDEV.P}(K\$2:K\$247)$ 」

を入力して計算し、他のセルにコピーする

- 列Kから列Mに示された各正規確率変数が、列Pから列Rに標準正規確率変数に変換されて表示される

## 【Excel】 正規乱数の標準化

- 列Pから列Rのデータが標準正規分布に従っていることを確認するために、これらの系列の平均、分散、標準偏差を計算すると、図5-32に示したように標準正規分布の条件を満たしている

	O	P	Q	R
247	246	-1.03587	0.76028	0.848318
248	平均	5.05E-17	7.67E-18	1.26E-17
249	分散	1	1	1
250	標準偏差	1	1	1

図5-32 正規分布作業(6)

# 【Excel】 正規乱数の標準化

- これらのデータを対象にしてヒストグラムを描いたものが、以下の図5-33, 図5-34, 図5-35のグラフである

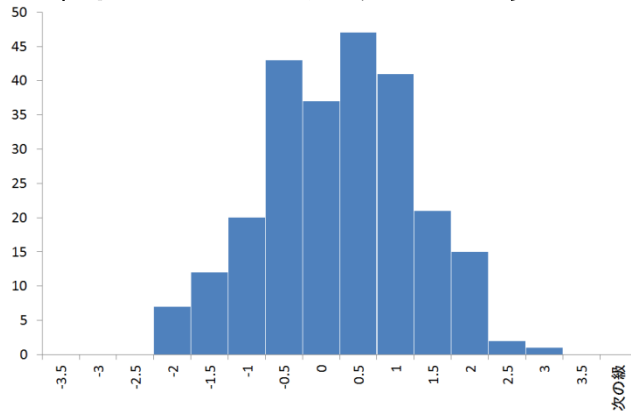


図5-33 正規分布作業(7)

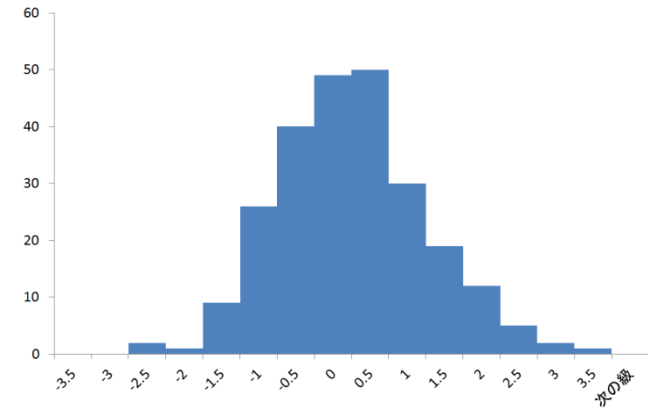


図5-34 正規分布作業(8)

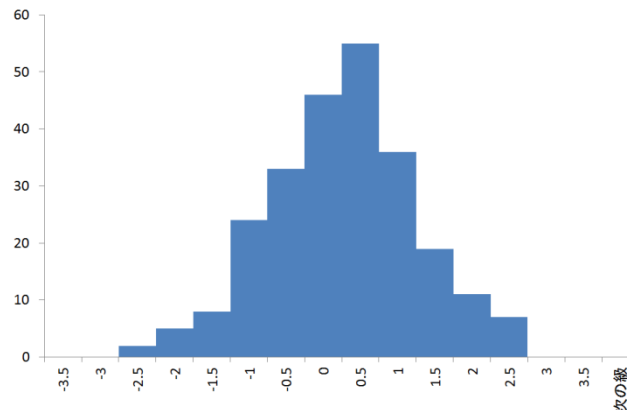


図5-35 正規分布作業(9)

# 【Excel】 正規乱数の標準化

---

- これらのグラフを先の図5-29, 5-30, 5-31のグラフと見比べた場合, 標準正規分布に変換して示したグラフの方がはるかに類似した分布の形状を示している
  - しかし, 同じ標準正規分布と言ってもここに示した各ヒストグラムの形状にはそれぞれの特徴がある
  - また, グラフを眺めただけで標準正規分布の条件を満たしているかどうかを判断することは容易ではない



# 【Excel】 正規乱数の標準化

ここに示した3ケースのヒストグラムの元になる度数分布表の値と標準正規分布の理論分布を1枚のグラフ上に折れ線グラフで表示し、形状の比較を行う

- ワーク・シートに図5-36のような表を作成する: 1行目とU列を入力

	U	V	W	X	Y
1	区間	SND	SND(0.5)	SND(1.0)	SND(2.0)
2	-3.5	0.000873	0	0	0
3	-3	0.004432	0	0	0
4	-2.5	0.017528	2	1	1
5	-2	0.053991	6	8	3
6	-1.5	0.129518	10	12	14
7	-1	0.241971	19	22	25
8	-0.5	0.352065	38	29	37
9	0	0.398942	54	39	47
10	0.5	0.352065	42	58	41
11	1	0.241971	31	37	39
12	1.5	0.129518	32	26	21
13	2	0.053991	9	11	15
14	2.5	0.017528	2	2	3
15	3	0.004432	1	1	0
16	3.5	0.000873	0	0	0
17	合計	1.999697	246	246	246

図5-36 正規分布作業(10)

# 【Excel】 正規乱数の標準化

---

SNDの列に示されているのは、データ区間として取られている値に対応する標準正規分布の理論値である

1.  $\Sigma$  をクリックして、「その他の関数(E)」をクリックする
2. 統計関数のグループの中から、NORM.DISTを選択して、
  - 各区間の端点の入力されているセル番地
  - 正規分布の平均
  - 標準偏差
  - 区間の対応する点の確率密度関数かその点までの累積分布関数を0か1で選択して入力する
3. 区間がセルU2からU16に入力されている前提で、セルV2に、「=NORM.DIST(U2,0, 1, 0)」を入力し、セルV16までコピーすると、図5-36に示した値が表示される

# 【Excel】 正規乱数の標準化

- SND(0.5)は標準偏差を0.5として発生させた正規乱数を標準正規分布に変換したデータの度数分布(以下同様)
- SND(0.5)等の下に示されている合計は度数分布の合計
  - グラフを作成するには度数ではなく、各度数を合計で割った値である相対度数を用いる
  - SNDの合計も2に近い値になるので、合計で割って相対度数にする

	U	V	W	X	Y	Z	AA	AB	AC
1	区間	SND	SND(0.5)	SND(1.0)	SND(2.0)	RSND	RSND(0.5)	RSND(1.0)	RSND(2.0)
2	-3.5	0.000873	0	0	0	0.000436	0	0	0
3	-3	0.004432	0	0	0	0.002216	0	0	0
4	-2.5	0.017528	2	1	1	0.008765	0.00813	0.004065	0.004065
5	-2	0.053991	6	8	3	0.027	0.02439	0.03252	0.012195
6	-1.5	0.129518	10	12	14	0.064769	0.04065	0.04878	0.056911
7	-1	0.241971	19	22	25	0.121004	0.077236	0.089431	0.101626
8	-0.5	0.352065	38	29	37	0.176059	0.154472	0.117886	0.150407
9	0	0.398942	54	39	47	0.199501	0.219512	0.158537	0.191057
10	0.5	0.352065	42	58	41	0.176059	0.170732	0.235772	0.166667
11	1	0.241971	31	37	39	0.121004	0.126016	0.150407	0.158537
12	1.5	0.129518	32	26	21	0.064769	0.130081	0.105691	0.085366
13	2	0.053991	9	11	15	0.027	0.036585	0.044715	0.060976
14	2.5	0.017528	2	2	3	0.008765	0.00813	0.00813	0.012195
15	3	0.004432	1	1	0	0.002216	0.004065	0.004065	0
16	3.5	0.000873	0	0	0	0.000436	0	0	0
17			0	0	0	0	0	0	0
18	合計	1.999697	246	246	246	1	1	1	1

図5-37 正規分布作業(11)

# 【Excel】 正規乱数の標準化

- SND(0.5)の列について
  - SND(0.5)の列については、セルW2 からW16を指定して、
  - セルW2に「=FREQUENCY(P\$2:P\$247, \$U2:\$U16)」を入力し、「Ctrl+Shift+Enter」を押す
- その他の列はSND(0.5)の列をコピーすればよい
- RSND, RSND(0.5), RSND(1.0), RSND(2.0)については、例えばセルZ2に「=V2/V\$17」を入力して、これを他のセルにコピーすればよい

	U	V	W	X	Y	Z	AA	AB	AC
1	区間	SND	SND(0.5)	SND(1.0)	SND(2.0)	RSND	RSND(0.5)	RSND(1.0)	RSND(2.0)
2	-3.5	0.000873	0	0	0	0.000436	0	0	0
3	-3	0.004432	0	0	0	0.002216	0	0	0
4	-2.5	0.017528	2	1	1	0.008765	0.00813	0.004065	0.004065
5	-2	0.053991	6	8	3	0.027	0.02439	0.03252	0.012195
6	-1.5	0.129518	10	12	14	0.064769	0.04065	0.04878	0.056911
7	-1	0.241971	19	22	25	0.121004	0.077236	0.089431	0.101626
8	-0.5	0.352065	38	29	37	0.176059	0.154472	0.117886	0.150407
9	0	0.398942	54	39	47	0.199501	0.219512	0.158537	0.191057
10	0.5	0.352065	42	58	41	0.176059	0.170732	0.235772	0.166667
11	1	0.241971	31	37	39	0.121004	0.126016	0.150407	0.158537
12	1.5	0.129518	32	26	21	0.064769	0.130081	0.105691	0.085366
13	2	0.053991	9	11	15	0.027	0.036585	0.044715	0.060976
14	2.5	0.017528	2	2	3	0.008765	0.00813	0.00813	0.012195
15	3	0.004432	1	1	0	0.002216	0.004065	0.004065	0
16	3.5	0.000873	0	0	0	0.000436	0	0	0
17			0	0	0	0	0	0	0
18	合計	1.999697	246	246	246	1	1	1	1

# 【Excel】 正規乱数の標準化

- 正規乱数から変換されたデータでは, 細部で理論値と隔たりがあるが, 全体的な動向はある程度類似している
  - 統計学的にデータが正規分布に従っていると判断するための検定方法も存在するが, ここではグラフによる確認にとどめる

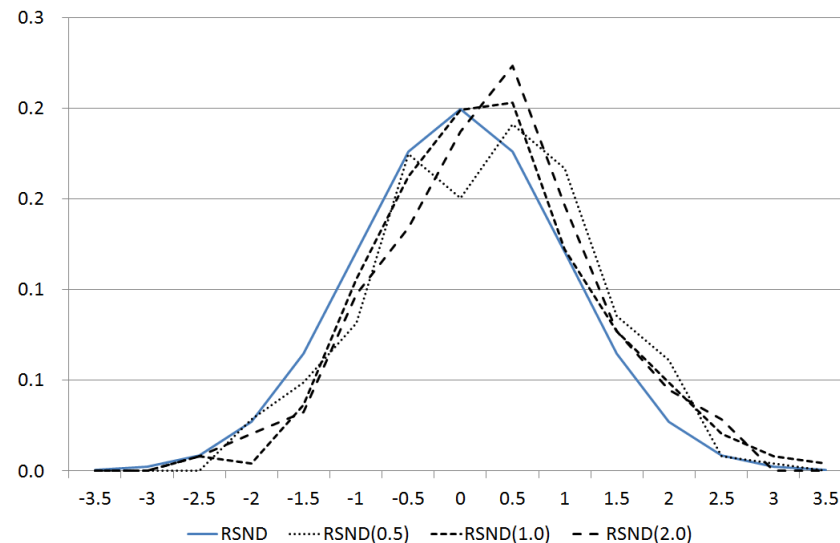


図5-38 正規分布作業(12)

## 【練習問題5-3】 提出課題17:3点

---

- Excelによって、平均10、標準偏差10の正規乱数を1組につき250個のデータを4組発生させよ
- 各組ごとに、平均と標準偏差を計算し、計算された各組の平均と標準偏差を用いて、各組の乱数を標準正規分布に変換し、これら4組の標準正規分布に変換されたデータを1枚のグラフに示せ
  - 結果は、本章の付録参照

# 金融データの標準化

---

- 前章では、日次の金融指標データ、その変化および変化率に関する分布を検討した
- 本章では、こうしたデータの分布を正確に評価するための準備として一様乱数と正規乱数を用いて、これらに関する確率分布の考え方を説明してきた

# 金融データの標準化

---

- 以下では、前章で求めた各種のヒストグラムと各種の統計量を利用して、
  - 標準化し、
  - これらの形状を標準正規分布の理論値と比較し、金融指標の分布を正規分布と判断することの妥当性を考える
- 本節の説明は2003年の日経平均株価を対象に行い、他の金融指標に関しては練習問題とするので、各自で試みること



# 【Excel】 金融データの標準化

- 新しいワーク・シートを用意し, 先に利用した「日次原データ (2003-2012).xls」のデータ・ファイルから以下の図5-40のように当面必要な日経平均に関するデータをコピーする

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1			2003年	日経平均(2003年)	変化(2003)	変化率(2003)	2005年	日経平均(2005年)	変化(2005)	変化率(2005)	2008年	日経平均(2008年)	変化(2008)	変化率(2008)	2011年	日経平均(2011年)	変化(2011)	変化率(2011)
2	1		2003/1/6	8713.33			2005/1/4	11517.75			2008/1/4	14691.41			2011/1/4	10398.1		
3	2		2003/1/7	8656.5	-56.83	-0.652219	2005/1/5	11437.52	-80.23	-0.696577	2008/1/7	14500.55	-190.86	-1.299126	2011/1/5	10380.77	-17.33	-0.166665
4	3		2003/1/8	8517.8	-138.7	-1.602264	2005/1/6	11492.26	54.74	0.4786003	2008/1/8	14528.67	28.12	0.1939237	2011/1/6	10529.76	148.99	1.43525
5	4		2003/1/9	8497.93	-19.87	-0.233276	2005/1/7	11433.24	-59.02	-0.513563	2008/1/9	14599.16	70.49	0.4851786	2011/1/7	10541.04	11.28	0.1071249
6	5		2003/1/10	8470.45	-27.48	-0.323373	2005/1/11	11539.99	106.75	0.9336811	2008/1/10	14388.11	-211.05	-1.445631	2011/1/11	10510.68	-30.36	-0.288017
7	6		2003/1/14	8553.06	82.61	0.9752729	2005/1/12	11453.39	-86.6	-0.750434	2008/1/11	14110.79	-277.32	-1.927425	2011/1/12	10512.8	2.12	0.02017
8	7		2003/1/15	8611.75	58.69	0.6861872	2005/1/13	11358.22	-95.17	-0.830933	2008/1/15	13972.63	-138.16	-0.979109	2011/1/13	10589.76	76.96	0.73206
9	8		2003/1/16	8609.17	-2.58	-0.029959	2005/1/14	11438.39	80.17	0.7058324	2008/1/16	13504.51	-468.12	-3.350264	2011/1/14	10499.04	-90.72	-0.856677
10	9		2003/1/17	8690.25	81.08	0.9417865	2005/1/17	11487.1	48.71	0.4258466	2008/1/17	13783.45	278.94	2.0655322	2011/1/17	10502.86	3.82	0.0363843
11	10	01	2003/1/20	8558.82	-131.43	-1.512385	2005/1/18	11423.26	-63.84	-0.555754	2008/1/18	13861.29	77.84	0.5647352	2011/1/18	10518.91	16.05	0.1528155
12	11		2003/1/21	8708.58	149.76	1.7497739	2005/1/19	11405.34	-17.92	-0.156873	2008/1/21	13325.94	-535.35	-3.862195	2011/1/19	10557.1	38.19	0.3630604
13	12		2003/1/22	8611.04	-97.54	-1.120045	2005/1/20	11284.77	-120.57	-1.057136	2008/1/22	12573.05	-752.89	-5.649808	2011/1/20	10437.31	-119.79	-1.134687
14	13		2003/1/23	8790.92	179.88	2.0889463	2005/1/21	11238.37	-46.4	-0.411174	2008/1/23	12829.06	256.01	2.0361806	2011/1/21	10274.52	-162.79	-1.559693
15	14		2003/1/24	8731.65	-59.27	-0.674218	2005/1/24	11289.49	51.12	0.4548702	2008/1/24	13092.78	263.72	2.0556455	2011/1/24	10345.11	70.59	0.6870394
16	15		2003/1/27	8609.47	-122.18	-1.399277	2005/1/25	11276.91	-12.58	-0.111431	2008/1/25	13629.16	536.38	4.0967617	2011/1/25	10464.42	119.31	1.1532985
17	16		2003/1/28	8525.39	-84.08	-0.976599	2005/1/26	11376.57	99.66	0.8837527	2008/1/28	13087.91	-541.25	-3.971265	2011/1/26	10401.9	-62.52	-0.597453
18	17		2003/1/29	8331.08	-194.31	-2.279192	2005/1/27	11341.31	-35.26	-0.309935	2008/1/29	13478.86	390.95	2.9871079	2011/1/27	10478.66	76.76	0.7379421
19	18		2003/1/30	8316.81	-14.27	-0.171286	2005/1/28	11320.58	-20.73	-0.182783	2008/1/30	13345.03	-133.83	-0.992888	2011/1/28	10360.34	-118.32	-1.129152
20	19		2003/1/31	8339.94	23.13	0.2781114	2005/1/31	11387.59	67.01	0.5919308	2008/1/31	13592.47	247.44	1.8541734	2011/1/31	10237.92	-122.42	-1.181621
21	20		2003/2/3	8500.79	160.85	1.9286709	2005/2/1	11384.4	-3.19	-0.028013	2008/2/1	13497.16	-95.31	-0.701197	2011/2/1	10274.5	36.58	0.3572991

図5-40 日経平均作業(1)

# 【Excel】 金融データの標準化

---

- 列S以下に各データ系列を標準正規分布に変換した数値を計算して入力する
  - 系列によって長さに相違があるので注意する
  - 列により開始行が異なることにも注意する
- 各列の最初のセルに入力する式を以下に示す
$$T2 := (D2 - \text{AVERAGE}(D\$2:D\$247)) / \text{STDEV.P}(D\$2:D\$247)$$
$$U3 := (E3 - \text{AVERAGE}(E\$3:E\$247)) / \text{STDEV.P}(E\$3:E\$247)$$
  - 
  - 
  -
$$AE3 := (R3 - \text{AVERAGE}(R\$2:R\$246)) / \text{STDEV.P}(R\$2:R\$246)$$

# 【Excel】 金融データの標準化

- これらの計算結果は、図5-41に示す通りである
- 元々のデータの値とは異なり、どの系列の値を見てもほぼ-2から+2の間にあることがわかる

	A	B	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
1			HSD(2003)	HSD(2003D)	HSD(2003R)	HSD(2005)	HSD(2005D)	HSD(2005R)	HSD(2008)	HSD(2008D)	HSD(2008R)	HSD(2011)	HSD(2011D)	HSD(2011R)
2	1		-0.57735724			-0.65603453			1.20844189			1.318954675		
3	2		-0.63266512	-0.470526	-0.517387	-0.71439019	-0.790574	-0.879077	1.117659438	-0.559131	-0.390279	1.295464993	-0.067323	-0.062683
4	3		-0.76765021	-1.064603	-1.176627	-0.6745748	0.2866266	0.3525544	1.1310347	0.1742139	0.1233526	1.497411198	1.1279483	1.0152472
5	4		-0.78698801	-0.202332	-0.226681	-0.71750326	-0.621296	-0.687271	1.164563229	0.3161074	0.2235487	1.5127005	0.1382848	0.1215508
6	5		-0.81373199	-0.257553	-0.2892	-0.63985816	0.7017204	0.8294967	1.064177415	-0.626746	-0.440679	1.471549505	-0.160964	-0.144341
7	6		-0.7333346	0.5412973	0.6119345	-0.70284707	-0.841413	-0.935521	0.932270313	-0.848678	-0.606424	1.474423026	0.0724557	0.0630387
8	7		-0.67621654	0.3677257	0.411337	-0.7720694	-0.909811	-1.019887	0.866554586	-0.382643	-0.280188	1.578737275	0.6102991	0.5420701
9	8		-0.67872744	-0.07687	-0.085599	-0.71375739	0.4895844	0.5907025	0.643893571	-1.487651	-1.095902	1.455772246	-0.594746	-0.526992
10	9		-0.59981907	0.5301951	0.5886982	-0.67832795	0.2385009	0.2972666	0.776571225	1.0141884	0.7672144	1.460950006	0.0846729	0.0739494
11	10	01	-0.72772888	-1.011849	-1.114259	-0.72476226	-0.659765	-0.731489	0.813595778	0.3407219	0.2509174	1.482704732	0.1725647	0.152296
12	11		-0.58198002	1.0285603	1.1493632	-0.73779646	-0.293276	-0.313446	0.55895685	-1.712798	-1.272014	1.53446878	0.3316756	0.29377
13	12		-0.67690753	-0.765932	-0.842013	-0.8254936	-1.112529	-1.256957	0.200845144	-2.441321	-1.88698	1.372101267	-0.80366	-0.714065
14	13		-0.50184539	1.2471212	1.384716	-0.85924285	-0.520576	-0.579963	0.322616157	0.9373978	0.757117	1.151450065	-1.112683	-1.000052
15	14		-0.55952792	-0.488232	-0.532653	-0.82206049	0.2577352	0.3276844	0.448054427	0.9632179	0.7638132	1.247130195	0.5645206	0.5117757
16	15		-0.67843548	-0.944728	-1.035773	-0.83121061	-0.250657	-0.265821	0.703183273	1.8763326	1.4659883	1.408847099	0.9146507	0.8255218
17	16		-0.7602635	-0.668262	-0.742475	-0.75872245	0.6451349	0.7771698	0.445738014	-1.732557	-1.309536	1.324105327	-0.392085	-0.35256
18	17		-0.94936914	-1.468128	-1.646349	-0.78436897	-0.431667	-0.473861	0.631693163	1.3893001	1.0842505	1.428148488	0.6088618	0.5460282
19	18		-0.96325693	-0.161697	-0.183666	-0.79944703	-0.315702	-0.340601	0.568036998	-0.368143	-0.284929	1.267773464	-0.793096	-0.710341
20	19		-0.94074644	0.1096903	0.128172	-0.750707	0.3845539	0.4713292	0.685731695	0.9086976	0.6945038	1.101841158	-0.822561	-0.745648
21	20		-0.78420461	1.1090331	1.2735004	-0.75302726	-0.175715	-0.178396	0.640397547	-0.239142	-0.184583	1.151422956	0.3201052	0.2898932

図5-41 日経平均作業(2)

# 【Excel】 金融データの標準化

- これらのデータを利用してヒストグラムと度数分布を求める
- この結果を示しているのが以下の図5-42である

区分	HSD	HSD(2003)	HSD(2003D)	HSD(2003R)	HSD(2005)	HSD(2005D)	HSD(2005R)	HSD(2008)	HSD(2008D)	HSD(2008R)	HSD(2011)	HSD(2011D)	HSD(2011R)
-3.75	8.8417E-05	0	1	0	0	1	2	0	0	1	0	2	2
-3.25	0.00048861	0	1	1	0	1	0	0	1	0	0	0	0
-2.75	0.00240274	0	1	1	0	1	0	0	2	3	0	0	0
-2.25	0.00924471	0	4	4	0	3	1	1	2	2	0	1	1
-1.75	0.02783468	0	6	7	0	3	6	23	6	5	1	2	1
-1.25	0.06559062	26	11	12	0	3	8	30	15	9	24	9	9
-0.75	0.12097758	54	20	21	49	27	23	4	21	13	58	32	34
-0.25	0.17466632	36	41	39	100	61	61	6	44	55	20	45	45
0.25	0.19741265	18	59	59	22	61	57	29	55	62	40	61	61
0.75	0.17466632	31	48	50	25	39	36	97	43	55	44	40	40
1.25	0.12097758	51	35	29	17	20	25	55	32	24	16	37	34
1.75	0.06559062	29	11	15	8	12	16	0	16	6	35	11	11
2.25	0.02783468	1	5	6	13	7	6	0	5	5	7	2	4
2.75	0.00924471	0	2	1	10	3	1	0	0	2	0	0	0
3.25	0.00240274	0	0	0	1	0	1	0	1	0	0	1	1
3.75	0.00048861	0	0	0	0	2	0	0	0	1	0	1	0
	8.8417E-05	0	0	0	0	0	1	0	1	1	0	0	1
SUM	1	246	245	245	245	244	244	245	244	244	245	244	244

図5-42 日経平均作業(3)

# 【Excel】 金融データの標準化

---

標準正規分布の理論値であるHSD列は以下のように計算

1. 「=NORM.DIST(-3.75,0,1,1)」を入力: ある確率変数が-3.75以下の値を取る確率を求める
2. 区分-3.75に対応するHSDの値は, 「=NORM.DIST(-3.75,0,1, 1)」
3. 区分-3.25に対応するHSDの値は, 確率変数が-3.75から-3.25の間の値を取る確率密度を求めるので,  
「=NORM.DIST(-3.25,0,1,1)-NORM.DIST(-3.75,0,1,1)」
4. 以下同様に, 区分3.75まで対応するHSDの値を求める
5. 最後に区分3.75の下は空白になっているが, ここには確率変数が3.75以上の値を取る場合を計算することに対応
  - 各区分の確率密度は+0.25の区分の確率密度をはさんで対称となっているので, 区分-3.75に対応する確率密度をコピー

# 【Excel】 金融データの標準化

- FREQUENCY関数でも作成することができる
  - 例えば, HSD(2003)列のセルAI2からAI18を指定し,
  - AI2に「=FREQUENCY(T:T, \$AG2:\$AG17)」と入力し, 「Ctrl+Shift+Enter」を押す
  - セルAI19は「=SUM(AI2:AI18)」と入力する

	A	B	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT
1			区分	HSD	HSD(2003)	HSD(2003D)	HSD(2003R)	HSD(2005)	HSD(2005D)	HSD(2005R)	HSD(2008)	HSD(2008D)	HSD(2008R)	HSD(2011)	HSD(2011D)	HSD(2011R)
2	1		-3.75	8.84173E-05	0	1	0	0	1	2	0	0	1	0	2	2
3	2		-3.25	0.000488608	0	1	1	0	1	0	0	1	1	0	0	0
4	3		-2.75	0.002402738	0	1	1	0	1	0	0	2	2	0	0	0
5	4		-2.25	0.009244709	0	4	4	0	3	1	1	2	2	0	1	1
6	5		-1.75	0.027834684	0	6	7	0	3	6	23	6	6	1	2	1
7	6		-1.25	0.065590617	26	11	12	0	3	8	31	15	8	24	9	9
8	7		-0.75	0.120977579	54	20	21	50	27	23	3	21	13	58	32	34
9	8		-0.25	0.174666322	36	41	39	99	61	61	6	44	55	20	45	45
10	9		0.25	0.197412651	18	59	59	22	61	57	29	55	62	40	61	61
11	10	01	0.75	0.174666322	31	47	50	25	39	36	97	43	55	44	40	40
12	11		1.25	0.120977579	50	34	29	17	20	25	55	32	24	16	37	34
13	12		1.75	0.065590617	30	13	15	8	12	16	0	16	6	35	10	11
14	13		2.25	0.027834684	1	5	6	13	7	6	0	5	5	7	3	4
15	14		2.75	0.009244709	0	2	1	10	3	1	0	0	2	0	0	0
16	15		3.25	0.002402738	0	0	0	1	0	1	0	1	0	0	1	1
17	16		3.75	0.000488608	0	0	0	0	2	0	0	0	1	0	1	0
18	17			8.84173E-05	0	0	0	0	0	1	0	1	1	0	0	1
19	18		SUM	1	246	245	245	245	244	244	245	244	244	245	244	244

図5-42 日経平均作業(3)

# 【Excel】 金融データの標準化

- 標準正規分布の理論値と各経済指標を標準正規分布に変換した数値を比較するためには、図5-43の度数分布を相対度数分布に変換する必要がある
- これらの結果は、図5-43に示す通りである

区分	HSD	HSD(2003)	HSD(2003D)	HSD(2003R)	HSD(2005)	HSD(2005D)	HSD(2005R)	HSD(2008)	HSD(2008D)	HSD(2008R)	HSD(2011)	HSD(2011D)	HSD(2011R)
-3.75	8.84E-05	0.00000	0.00408	0.00000	0.00000	0.00410	0.00820	0.00000	0.00000	0.00410	0.00000	0.00820	0.00820
-3.25	0.000489	0.00000	0.00408	0.00408	0.00000	0.00410	0.00000	0.00000	0.00410	0.00000	0.00000	0.00000	0.00000
-2.75	0.002403	0.00000	0.00408	0.00408	0.00000	0.00410	0.00000	0.00000	0.00820	0.01230	0.00000	0.00000	0.00000
-2.25	0.009245	0.00000	0.01633	0.01633	0.00000	0.01230	0.00410	0.00408	0.00820	0.00820	0.00000	0.00410	0.00410
-1.75	0.027835	0.00000	0.02449	0.02857	0.00000	0.01230	0.02459	0.09388	0.02459	0.02049	0.00408	0.00820	0.00410
-1.25	0.065591	0.10569	0.04490	0.04898	0.00000	0.01230	0.03279	0.12245	0.06148	0.03689	0.09796	0.03689	0.03689
-0.75	0.120978	0.21951	0.08163	0.08571	0.20000	0.11066	0.09426	0.01633	0.08607	0.05328	0.23673	0.13115	0.13934
-0.25	0.174666	0.14634	0.16735	0.15918	0.40816	0.25000	0.25000	0.02449	0.18033	0.22541	0.08163	0.18443	0.18443
0.25	0.197413	0.07317	0.24082	0.24082	0.08980	0.25000	0.23361	0.11837	0.22541	0.25410	0.16327	0.25000	0.25000
0.75	0.174666	0.12602	0.19592	0.20408	0.10204	0.15984	0.14754	0.39592	0.17623	0.22541	0.17959	0.16393	0.16393
1.25	0.120978	0.20732	0.14286	0.11837	0.06939	0.08197	0.10246	0.22449	0.13115	0.09836	0.06531	0.15164	0.13934
1.75	0.065591	0.11789	0.04490	0.06122	0.03265	0.04918	0.06557	0.00000	0.06557	0.02459	0.14286	0.04508	0.04508
2.25	0.027835	0.00407	0.02041	0.02449	0.05306	0.02869	0.02459	0.00000	0.02049	0.02049	0.02857	0.00820	0.01639
2.75	0.009245	0.00000	0.00816	0.00408	0.04082	0.01230	0.00410	0.00000	0.00000	0.00820	0.00000	0.00000	0.00000
3.25	0.002403	0.00000	0.00000	0.00000	0.00408	0.00000	0.00410	0.00000	0.00410	0.00000	0.00000	0.00410	0.00410
3.75	0.000489	0.00000	0.00000	0.00000	0.00000	0.00820	0.00000	0.00000	0.00000	0.00410	0.00000	0.00410	0.00000
	8.84E-05	0.00000	0.00000	0.00000	0.00000	0.00000	0.00410	0.00000	0.00410	0.00410	0.00000	0.00000	0.00410
SUM	1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

図5-43 日経平均作業(4)

# 【Excel】 金融データの標準化

- これらのデータによって、項目ごとに2003年の理論値と現実値とを1枚の折れ線グラフに示したのが、以下の図5-44から図5-46のグラフである

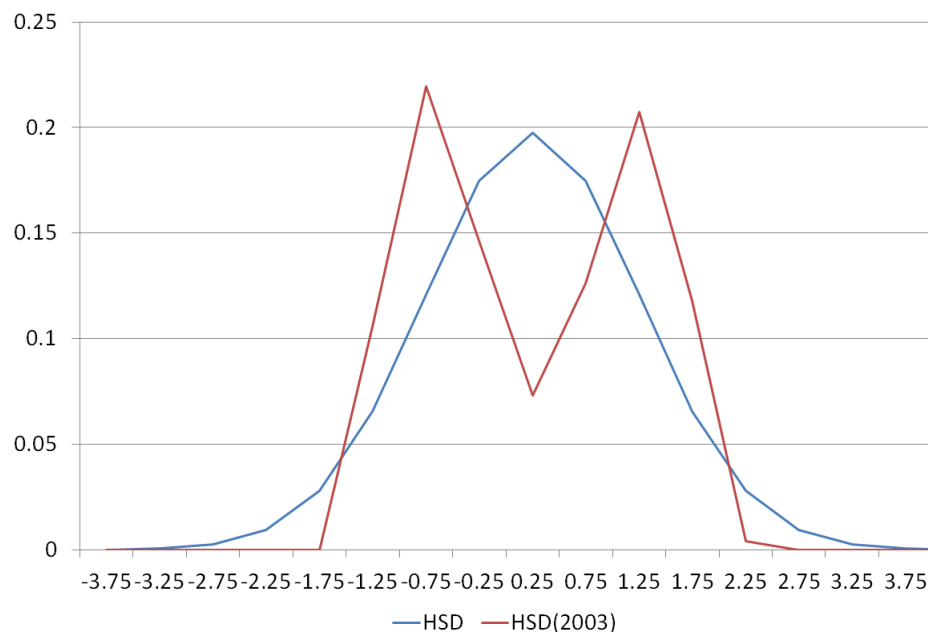


図5-44 2003年日経平均

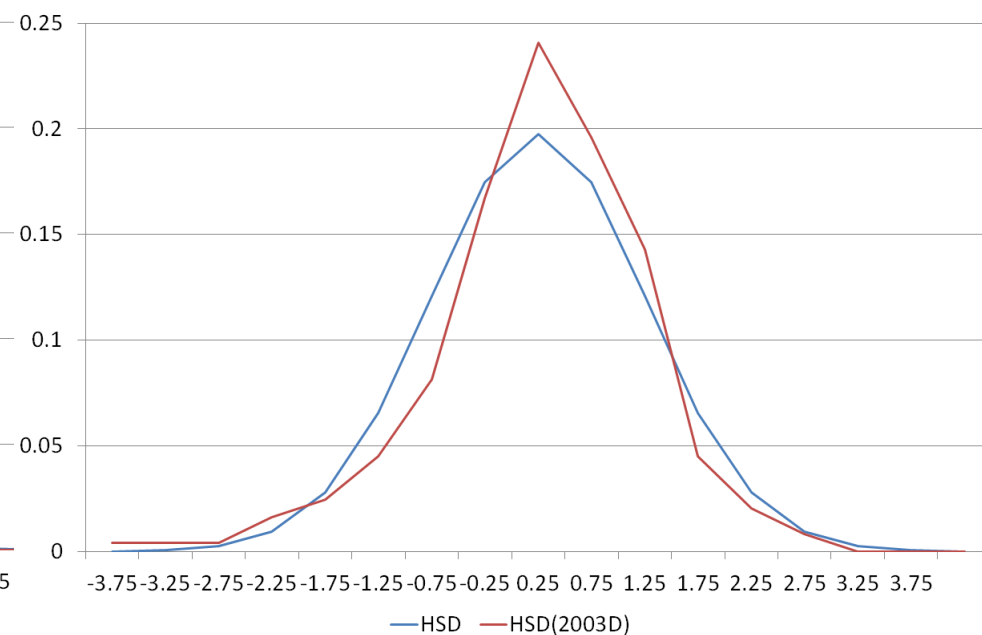


図5-45 2003年日経平均変化



# 【Excel】 金融データの標準化

- これらのデータによって、項目ごとに2003年の理論値と現実値とを1枚の折れ線グラフに示したのが、以下の図5-44から図5-46のグラフである

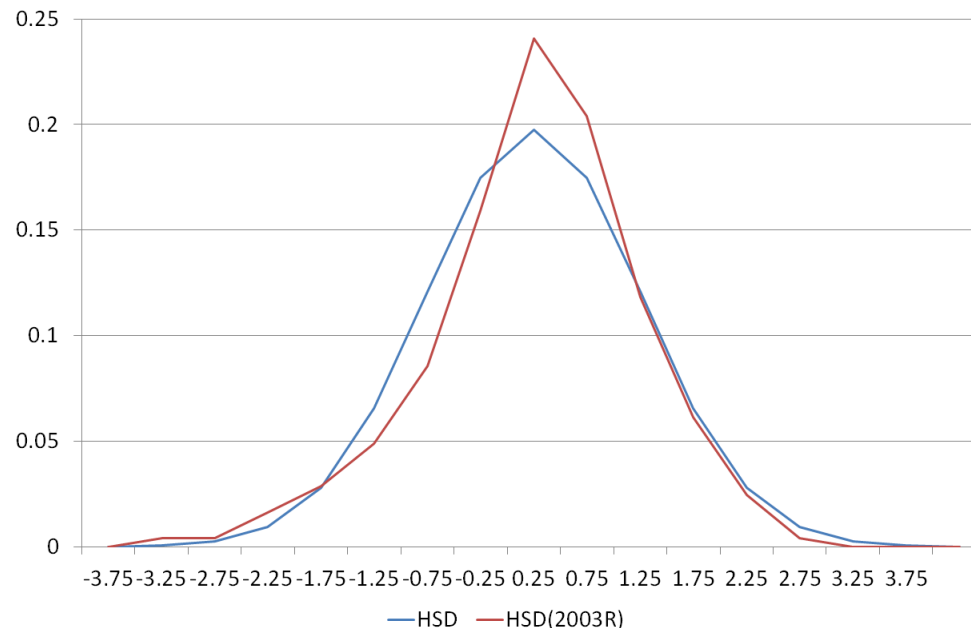


図5-46 2003年日経平均変化率

# 【Excel】 金融データの標準化

---

- 図5-44の日経平均の原データの変動を標準正規分布に変換したグラフにおいて理論値と現実値との間の開きが大きい
  - 2003年の日経平均の分布を正規分布と考えることは無理
  - 2003年は4月までの日経平均と9月以降の日経平均とでは平均的な水準が大きく異なっているため
    - この間の5月から8月の間は株価の上昇スピードが速かった

# 【Excel】 金融データの標準化

---

- 変化や変化率における理論値と現実値との間の相違は大きくなくそれぞれの分布が標準正規分布に従っていると考えて大きな問題はない
  - 変化や変化率のように日々の変化の記録をベースに分布の形状を分析すれば、原データにおいて上昇トレンドが存在してもその影響を強く受けることはなく、標準正規分布に近い
  - しかし、ここでも詳しく図を眺めると、図5-46のグラフでは分布の形状が右に偏っていると判断される
  - 年によって左に偏ったものも存在するだろう

## 【練習問題5-4】 提出課題18:3点

---

- 2003年における日経平均のデータを用いて、原データ、変化、変化率に関して、これらが正規分布に従うことを前提として、標準正規分布に変換した上で、理論分布と現実の分布とをグラフによって比較した
- ここでは同じ作業を2005年、2008年、2011年のデータを利用して行い、年ごとの比較と、データの種類による比較に関してコメントを記し、A4用紙に貼り付けてWordとして提出すること
- 作成するために用いたExcelも提出すること