第2章　正規乱数と正規分布

第 2 章では，正規乱数を用いたシミュレーションを行い，正規分布の性質を調べます。それは平均，標準偏差，確率，分布です。正規乱数は科学実験のデータと見なすことができます。特に，これらのデータが時間の経過とともに得られたとすると，身近で観測できるホワイトノイズ[1]と見なすこともできます（図 2.1)。

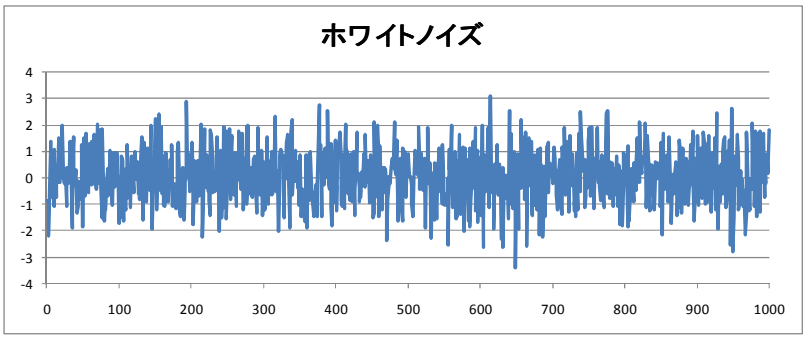


図 2.1　ホワイトノイズの例

2.1　正規乱数の発生と表示

本章では，正規乱数を 5000 個発生させ，ヒストグラムを作ります。図 2.1 は，最初の 1000 個だけを表示しています。行 1 のセル A1~K1 まで，「実験番号」，「実験結果 1」，「平均」，「標準偏差」，「区間配列」，「度数」「実験結果 2」，「平均」，「標準偏差」，「区間配列」，「度数」と入力します（図 2.2)。実験結果 1 と実験結果 2 では，標準偏差を変えてシミュレーションを行います。

最初に，列 A に 1～5000 までの番号を入力します。第 1 章のようにフィルハンドル＋をドラッグして入力することも可能ですが，対象とするセル全体が画面に収まらないため手間取りますので，1.1 で用いた手法（Ctrl キー＋Enter キー）[2]を応用します。最初に，セル A2 に「1」と入力します。次に数式を入力する範囲 A3～A5001 を選択するために，セル A3 をクリックし，ファンクションキーF5[3]を押して[ジャンプ]ダイアログボックスを表示します（図 2.3)。[参照先（R）:]に「A3: A5001」と入力し Enter キーを押せば，A3 から A5001 まで選択されますので，そのまま「=A2+1」とタイプし[4]，Ctrl キーを押しながら

---

[1] ホワイトノイズには，すべての波長の波が同じ強さで含まれています。ホワイトノイズの名称の由来は，すべての波長の光が同じ強さで含まれている白色光です。

[2] これは，範囲を指定し，数式を入力し，その範囲に一括して数式を設定する方法です。

[3] ファンクションキーF5 は[ジャンプ]機能のショートカットキーであり，[ホーム]タブ→[編集]→[検索と選択]→[ジャンプ]と同じです。

[4] この時，数式バーの数式ボックス（図 1.7 参照）に「=A2+1」と表示されます。

Enter キー（または[OK]ボタン）を押せば完了です。

	A	B	C	D	E	F	G	H	I	J	K
1	実験番号	実験結果1	平均	標準偏差	区間配列	度数	実験結果2	平均	標準偏差	区間配列	度数
2	1	0.876742	0.004444	0.997552	-4	0	3.478708	-0.02451	1.996902	-8	0
3	2	-0.25104			-3.9	0	0.244099			-7.9	0
4	3	0.600866			-3.8	0	1.692283			-7.8	0
5	4	0.203545			-3.7	0	-0.58576			-7.7	0
6	5	0.378445			-3.6	0	2.473098			-7.6	0
7	6	0.505108			-3.5	0	-1.35316			-7.5	0
8	7	1.295889			-3.4	0	-2.73783			-7.4	0
9	8	-0.05377			-3.3	0	1.152248			-7.3	0
10	9	1.328829			-3.2	2	-1.05532			-7.2	0
11	10	-0.41386			-3.1	2	2.844367			-7.1	0
12	11	0.866564			-3	0	-3.29763			-7	1
13	12	1.059088			-2.9	3	2.320344			-6.9	1
14	13	0.388573			-2.8	2	0.829057			-6.8	0
15	14	-1.5299			-2.7	2	-0.85078			-6.7	0
16	15	-1.5512			-2.6	7	-1.26726			-6.6	0
17	16	1.916515			-2.5	8	-3.53055			-6.5	1
18	17	-0.78876			-2.4	11	-1.08728			-6.4	1

図 2.2　データ入力

図 2.3　[ジャンプ]ダイアログボックス

以下では，正規乱数の発生は，[データ]タブの[分析]グループにある[データ分析]を使います。[データ分析]が表示されていない場合は，付録の操作に従って分析ツールを読み込みます。正規乱数は[データ]タブから行いますが，一様乱数は[数式]タブから行うという違いがあります。この違いについては，付録を参照してください。

実験結果 1 の列 B（図 2.2）に正規乱数を発生させます。 [データ分析]をクリックし，現れた[データ分析]ダイアログボックス（図 2.4）の[乱数発生]を選択し，[OK]ボタンを押せば，[乱数発生]ダイアログボックスが表示されます（図 2.5）。

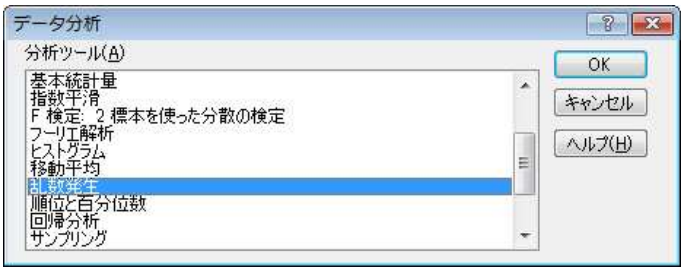


図 2.4　[データ分析]ダイアログボックス

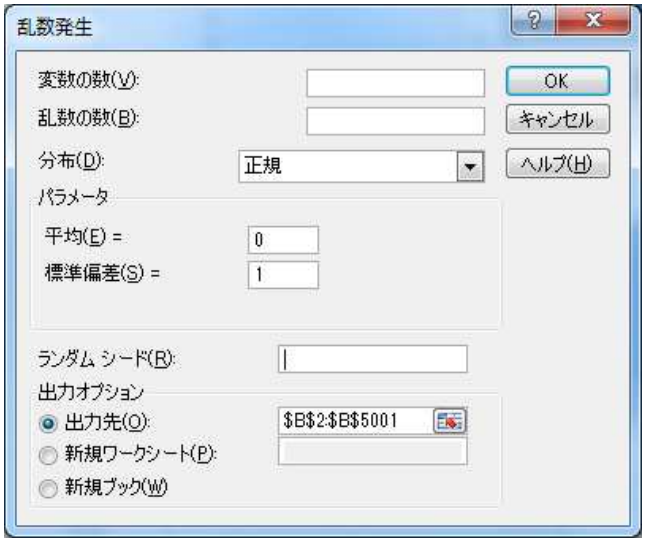


図 2.5　[乱数発生]ダイアログボックス

[乱数発生]ダイアログボックス（図 2.5）は，次のように設定します：

変数の数（V）：　空白

乱数の数（B）：　空白

分布（D）：　正規

平均（E）＝　0

標準偏差（S）＝　1

ランダムシード（R）：　空白

出力オプション　出力先（O）：$B$2:$B$5001

です。出力オプションの入力は，[出力先（O）：]ラジオボタンをクリックし，右側のテキストボックスをクリックした後，セル B2 をクリックすれば「$B$2」と入力されますので，「$B$2:$B$5001」と入力後，[OK]ボタンを押します。「$」は配列を固定する意味があります。詳細は，付録を参照してください。

グラフを作るために，B2 をクリックし，Ctrl キーと Shift キーを押しながら↓方向キーを押して，X-軸のすべてのデータを選択します。次に，[挿入]→グラフ[散布図]→[散布図（直線）]と進めば，デフォールトのグラフが得られます。図 2.1 のように体裁を整える方法は付録のグラフ体裁の変更（1）にあります。

2.2 データの独立性

第 1 章と第 2 章で発生させた乱数は，互いに独立です。つまり，1 回目のさいころの目と 2 回目のさいころの目とは無関係であり，互いに影響することはありません。同様に，発生させた正規乱数は互いに独立であり，その発生過程において，お互いに影響を与えることはありません。

図 2.1 の X-軸を実験番号，Y-軸を実験結果の数値と考えると，これは実験値を並べたことになります。一方，X-軸を時間，Y-軸を電圧などの観測値と考えると，時系列データの典型的なモデルであるホワイトノイズと見なせます。ホワイトノイズの例としては，ある装置の電流の時間変動などが思い当たります。これらは，バックグラウンドノイズ，ベースラインノイズなどと呼ばれている，いわゆるノイズです。

図 2.1 の X 軸を実験番号から時間に変えると，まったく異なった解釈ができることは興味深いことです。実験番号の場合，各実験結果は独立と考えれば，結果の順番を入れ替えても統計解析上の問題はありません。ホワイトノイズの場合も，各時刻のノイズは互いに独立ですが，実験結果が時間的に並んでいるため，ノイズの順番を入れ替えることはできません。

ホワイトノイズは自然現象だけでなく，社会現象でも観測されることがあります。たとえば，薬局で毎日調剤されている薬の量を図 2.1 のようにプロットするとホワイトノイズと見なせる場合があります。胃の薬などがこれにあたります。しかし，感染症の薬（インフルエンザの薬）の毎日の調剤量はホワイトノイズではなく，もう少し複雑な時系列になります。

地域社会においては，胃の疾患はランダムに発生し，患者は互いに無関係であると考えられます。つまり，胃の疾患は非感染症ですから，ある患者が他の患者の病状に影響を与えることはありません。しかし，インフルエンザの場合，患者の発生はランダムですが，ある患者は他の人に病気を感染させることがありますので，患者同士は無関係ではありません。X-軸に日付，Y-軸に薬の販売量をプロットすると，地域社会における罹患状況の時間変化を調べられます[2]。

2.3 正規乱数のヒストグラム

正規乱数のヒストグラムを作るには，図 2.2 の列 E の区間配列を作る必要があります。ここでは，各区間の幅を 0.1 に設定します。そのために，セル E2 に「－4」，E3 に「－3.9」を入力し，セル E2 と E3 を同時に選択した後，フィルハンドル「＋」を E82 までドラッ

グします。

FREQUENCY 関数の結果（度数）を入力する範囲 F2~F82 を選択状態にします[5]。 [数式]タブ→[その他の関数]→[統計]→[FREQUENCY]の順で選択し，[関数の引数]ダイアログボックスの [データ配列]に「B2:B5001」，[区間配列]に「E2:E82」と入力します（図 1.8 参照)。Ctrl キーと Shift キーを同時に押しながら[OK]ボタンを押せば度数分布の計算は完了です。

度数と範囲の関係を述べます。セル F2 の数値は，－4 以下の範囲にあるデータ B2~B5001 の度数です。F3 の数値は，－4 より大きく，－3.9 以下の範囲の度数です。以下同様です。

最後にヒストグラムを作ります。棒グラフのデータを選択するために，F1~F82 を選択し，[挿入]タブ→[縦棒]→[集合縦棒]の順でクリックすれば，デフォールトのヒストグラムが得られます。グラフの体裁を整えれば，図 2.6 のヒストグラムが得られます（付録のグラフ体裁の変更（2）を参照)。

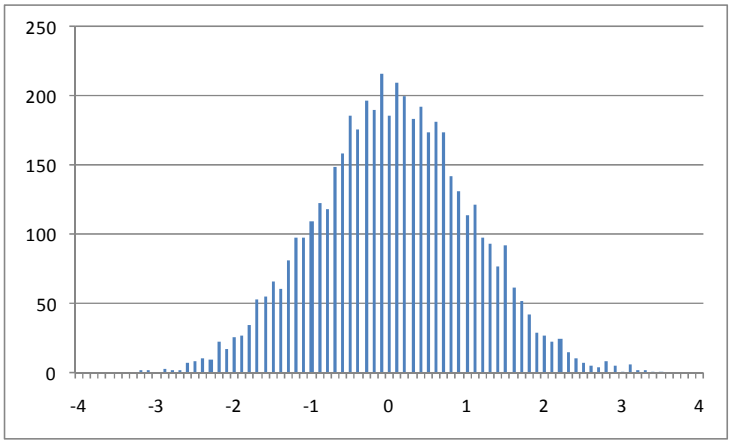


図 2.6　平均 0，標準偏差 1 の正規乱数のヒストグラム（n=5000）

5000 回のシミュレーションから得られたヒストグラム（図 2.6）は，すこしデコボコしていますが，正規分布らしく見えます。第 1 章の 30 回のくり返し実験では，一様分布は，一様からは程遠いものでしたが，くり返し数を 5000 に増やすと，それらしい分布が得られることが分かります。

図 2.6 のシミュレーションの平均と標準偏差を求めます。方法は第 1 章と同じですから，ここでは割愛します。真の値（平均は 0，標準偏差は 1）にかなり近い値がシミュレーションから得られたと思います。第 1 章の 30 回のくり返し実験と第 2 章の 5000 回のくり変えし実験では，平均と標準偏差を推定する精度が大きく異なることが分かります。くり返し数と推定の確からしさの関係は，第 3 章と第 4 章で学習します。

[5] マウスでドラッグして選択するのが簡単です。

2.4　正規分布の標準偏差と確率

次は数学の本によくある記述です：

平均をμ，標準偏差をσとします。正規分布の確率密度関数は

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} \tag{2.1}$$

であり，分布関数は，

$$\int_{-\infty}^{x} f(z;\mu,\sigma)dz \tag{2.2}$$

です。

ランダム変数 x が範囲 (x, x + dx) に存在する確率は，近似的に f(x; μ, σ)dx と表せます。x のディメンジョンを長さとすれば，密度関数 f(x; μ, σ)のディメンジョンは「確率密度＝確率÷長さ」であり，f(x; μ, σ)と長さ dx の積は「確率密度×長さ＝確率」となります。

近似的にではなく，正確に確率を求める場合は分布関数を使います。式（2.2）はランダム変数が値 x より小さい場合の確率を表します。式（2.2）は確率を表すので，－∞から＋∞まで積分すれば積分値は 1 になります。

正規分布は特筆すべき性質は，平均 μ と標準偏差 σ がどんな値であっても，ランダム変数 x が範囲

$$-\lambda\sigma \le x-\mu \le \lambda\sigma \tag{2.3}$$

に存在する確率は，λ だけに依存して決まることです。たとえば，λ=1.96 の場合，x が（μ －1.96 σ，μ ＋1.96 σ）に存在する確率は 95%です（図 2.7 参照)。x が μ ＋1.96 σ より大きい確率は 2.5%で，μ －1.96 σ より小さい確率が 2.5%です。この場合，片側で 2.5%，両側で 5%ということがあります。片側と両側が紛らわしいことが往々にしてありますので，これらの区別には注意する必要があります。

図 2.7 の関係を，もう少し一般的に表 2.1 に示します。λ を任意の実数としたとき，ランダム変数 X が（μ －λ σ，μ ＋λ σ）の範囲の値を取る確率を $P\{|X-\mu|\le\lambda\sigma\}$と記述しています。x，$\mu$，σ は同じディメンジョンであることは知っている必要があります。

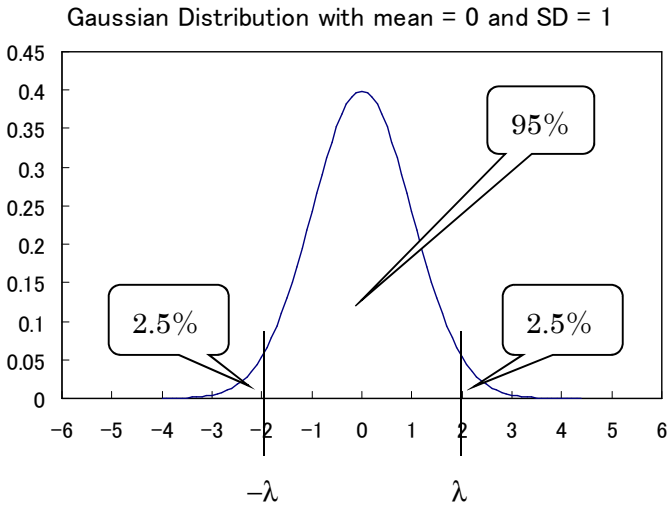


図 2.7　正規分布の確率密度関数（λ＝1.96 の範囲と確率）[6]

表 2.1　正規分布の確率と標準偏差

λ	1	1.65	1.96	2	2.58	3
P{|X−μ|≤λσ}	0.683	0.900	0.950	0.954	0.990	0.997

λ=3 の場合，興味深い事実があります。この場合の確率は，「±3σ 以内には 99.7%が含まれる。」または「±3σ 以内に含まれないのは 0.3%である。」と言われます。統計学的な会話においては，3σ 以上の確率とはめったに起きない事象を意味します。一方，千三（せんみつ）という日本語があり，これは，真実を言うのは千のうちわずか三つだけという，うそつきを意味します。1000 回のうち 3 回しか起きない事態とは，洋の東西を問わず，異常な事態なのでしょう。

話は前後しますが，確率 f(x; μ, σ)dx を，図 2.6 のデータから検証してみます。使っていないセルをクリックし，[数式]タブ→関数ライブラリ[オート SUM]→[合計]と進みます。列 E の値−1 から＋1 に対応する列 F の区間を選択し，Enter キーを押します。最初に選択したセルに関数が「=SUM(F33:F52)」と入力されているはずです。このセルの値は区間 E33~E52 の度数の合計ですから，確率に変換するためにくり返し数 5000 で割ります。関数を「=SUM(F33:F52)/5000」と書きかえれば，確率が得られます。この値は，0.683 に近い値のはずです。同様に，区間−2~＋2 などに対応する確率を計算すると，表 2.1 に近い値が得られるはずです。

[6] 正規分布の密度関数を表示するには，NORMDIST 関数を使います。

2.5　正規分布の標準偏差σの性質

表 2.1 は，正規分布の確率はλによって決まり，実際のσの値には無関係であることを示しています。つまり，σ=1 のときの範囲 2 > x >－2 の確率とσ=2 のときの範囲 4 > x >－4 の確率は同じです（=0.954)。この項では，σ=2 場合のシミュレーションを行い，既に調べたσ=1 の場合と比べます。

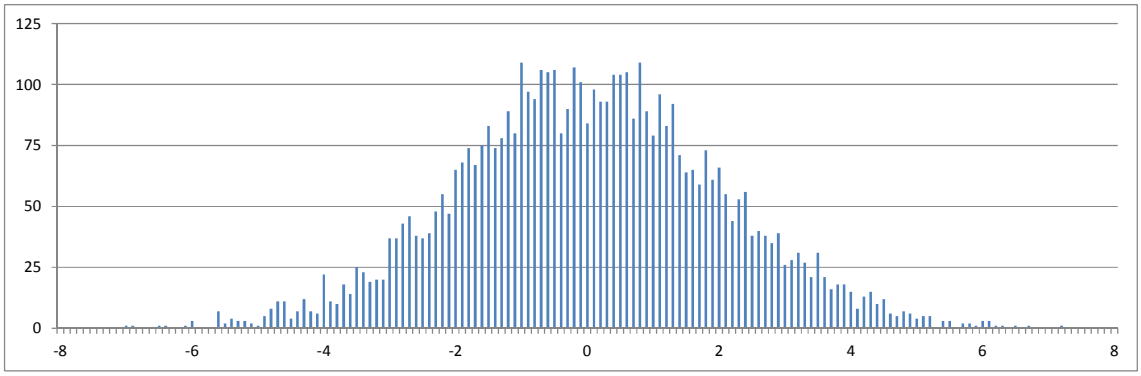
図 2.2 にあるように列 G から列 K まで数値を入力します。列 G に正規乱数を入力するため, [データ分析]にある正規乱数の[乱数発生]ダイアログボックスを次のように設定します:
変数の数（V）：　空白
乱数の数（B）：　空白
分布（D）：　正規
平均（E）＝　0
標準偏差（S）＝　2
ランダムシード（R）：　空白
出力オプション　出力先（O）: $G$2:$G$5001
です。標準偏差が 2 であることに注意してください。

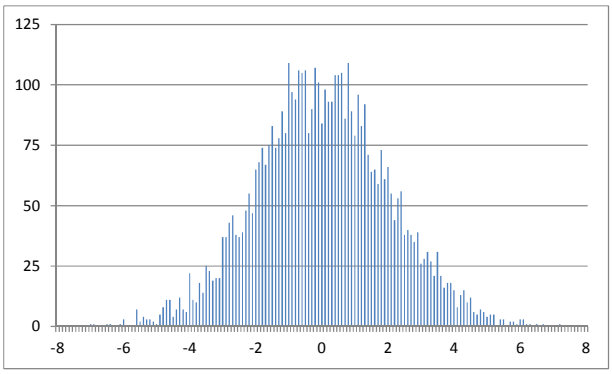
列 J の区間配列は，－8~＋8 の間を 0.1 間隔で設定します。つまり，σ=1 の場合（図 2.6）と同じ間隔です。J2~J162 に区間配列が設定されるはずです。列 K に度数を入力し，ヒストグラムを作ると図 2.8（上）のようになります。ここでは，X 軸の目盛を，σ=1 の場合（図 2.8（下））と同じになるように表示しています。

σ=2 の場合は分布がσ=1 の場合（図 2.8(下)）に比べて広がっていることが分かります。σ=2 の場合は－8~＋8 に亘って度数が分布していますが，σ=1 の場合は－4~＋4 に亘っています。しかし，どちらの場合も，±4σの範囲であることは共通しています。度数の合計が 5000 であることも共通しています。

図 2.8（中）は，σ=2 の場合の図（図 2.8（上））の X 軸のスケールを 1/2 にしたものです。有限のくり返し数のシミュレーションですから，分布がギザギザしていることを除けば，σ=2 の分布とσ=1 の分布の形は同じになります。つまり，X 軸をσを単位として表示すれば，σの値には無関係に，正規分布の形は同じになります。これが，正規分布の確率はλによって決まることの理由です。

図 2.8（中）と（下）では，Y 軸のスケールが違います。σ=2 の場合は，分布の最大値が 100 近辺にあるのに対して，σ=1 の場合は，200 近辺にあります。この理由を考えてみます。どちらの場合でもくり返しは同じですから（=5000)，図 2.8（上）と（下）では，度数の合計を表すブロックの全面積は同じです。しかし，図 2.8（中）は X 軸に沿って半分に縮めてありますので，面積を保つためには，Y 軸に沿って倍に伸ばす必要があります。すると，図 2.8（中）と（下）のブロックの全面積は同じになります。





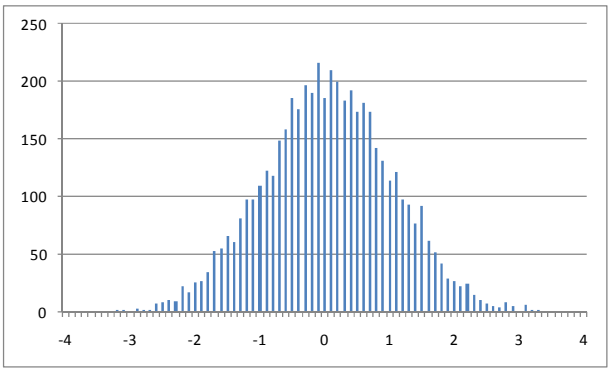


図 2.8　σ=2 の場合のシミュレーション（上と中）と σ=1 の場合のシミュレーション（下）

2.6　付録

分析ツールの読み込み

正規乱数を発生させるツールを提供する[データ分析]を読み込むためには，次の操作を行います：

1. [ファイル]タブをクリックし，[オプション] をクリックします。
2. [Excel のオプション]ダイアログボックスの左側にある[アドイン] をクリックし，[管理：] コンボボックスのから[Excel アドイン] をクリックします。

3. [設定] ボタンをクリックします。
4. [アドイン]ダイアログボックスの[有効なアドイン] 一覧の [分析ツール] チェック ボックスをオンにし，[OK] をクリックします。
5. 「分析ツールが現在コンピュータにインストールされていない」というメッセージボックスが現れた場合は，[はい] ボタンをクリックします。

一様乱数と正規乱数のツールの違い

第 1 章では[数式]タブにある乱数発生のツール（RANDBETWEEN）を使いましたが，第 2 章では[データ]タブにある[分析]の乱数発生機能を使いました。一様乱数は，セルに関数を記述し発生させましたが，正規乱数は，予め発生させた乱数を数値データとしてセルに書き込みます。前者の場合，ファンクションキーF9 を押すと再計算が実行されましたが，後者では再計算は行われません。乱数を発生する関数がセルに入力されるのと，すでに発生された乱数が数値データとしてセルに入力される違いがあるからです。[数式]タブと[データ]タブに別々に分類されているのも，この違いのためです。なお，数式バーの数式ボックス（図 1.11 参照）にも違いが現れます。第 1 章で設定したセル B2 をクリックすると関数が表示されますが，第 2 章のセル B2 をクリックすると数値が表示されます。

グラフ体裁の変更（1）

項 2.1 の操作では，X 軸が Y 軸のゼロを通っています。そこで，図 2.1 のように，X 軸の位置を移動します。そのために，Y 軸の目盛の上を右クリックし，ポップアップメニューから[軸の書式設定（F)...]を選択します。[軸の書式設定]ダイアログボックスの[横軸との交点： 軸の値（E）:]をオンにして，テキストボックスに「－4」と入力します。 [閉じる]ボタンをクリックします。なお，凡例を除去し，横軸の[軸の書式設定（F)...]において最大値を 1000 に固定しています。

グラフ体裁の変更（2）

項 2.3 の操作では，図 2.9 の図が得られます。今の段階では，X 軸が正しく設定されていませんので，これを修正します。目的のグラフをクリックして選択した後，[デザイン]タブ→[データ]→[データの選択]の順で進むと，[データソースの選択]ダイアログボックスが現れます[7]。 [横（項目）軸ラベル（C）]の下にある[編集（T）]ボタンを押します。現れた[軸ラベル]ダイアログボックスの[軸ラベルの範囲 (A) :]の下のテキストボックスに「E2:E82」を入力します。[OK]ボタンを押した後，[データソースの選択]ダイアログボックスの[OK]ボタンを押します。

[7] この操作はグラフのプロットエリア内を右クリックして現れるポップアップメニューから[データの選択]をクリックしても行えます。

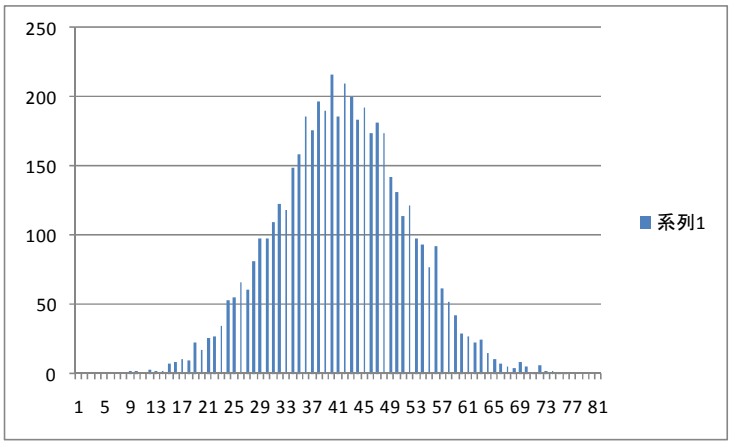


図 2.9　デフォールトの正規乱数のヒストグラム

X 軸の目盛を標準偏差（=1）と同じ間隔にします。まず，目盛の上をクリックし選択状態にした後，その上を右クリックし，ポップアップメニューから[軸の書式設定（F）...]を選択し，[間隔の単位（S）]をクリックし選択状態にしてから，右にあるテキストボックスに「10」と入力します。以上の操作で，図 2.8 と同じ X 軸の表示が得られます。

参照する配列を固定する方法（F4 の利用）

関数の引数として設定する配列を固定する方法はファンクションキーF4 を使います。キーボードの数字 4 を Shift を押しながら押すと「$」が表示されることに対応していると覚えればよいでしょう。

セルをクリックし，数式バーの数式ボックス内にある固定したい配列を選択し，F4 を押します。F4 を押すたびに，$の表示位置が変わりますので，試してください。

関数の引数を設定するときに，F4 を使って引数の配列を固定することもできます。たとえば，関数 FREQUENCY の[関数の引数]ダイアログボックスにある[データ配列]を「B2:B1001」とタイプした後に，F4 を押せば「$B$2:$B$1001」となります。もう一度 F4 を押すと$の表示が変わります。何度か押すと，元の表示に戻ります。