【補足資料】
確率・統計の基礎知識

２０１１年４月
日本銀行金融機構局
金融高度化センター

目　次

(注) 本資料はセミナー内容の理解を助けるために作成した補足資料です。
確率・統計理論を体系的に説明するものではありません。数学的な厳密さよりも直感的に理解することに重点を置いた記載も含まれています。
確率・統計理論をしっかりと習得したい方は、別途、初等統計学のテキストをご利用ください。

1. 基本統計量(1変量)

（1） 平　　均

（2） 分　　散

（3） 標準偏差

（4） パーセント点

（1）平　均

- 平均は、観測データセットの「中心の位置」を示す指標の1つ。

$$\overline{X} = \frac{\text{データの和}}{\text{データの数}}$$

$$= \frac{X_1 + X_2 + \cdots + X_N}{N}$$

- Excelでは、関数AVERAGE（データ範囲）を使って求める。

(2)分　散

- 分散は、観測データセットの「バラツキ」を示す指標の1つ。

―― データの「偏差平方和」(平均との差を2乗して合計)を求めて「データ数－1」で割る(ここでは 分散を推測統計＜後述＞の立場で定義)。

―― 分散の「単位」は、データの持つ「単位」の2乗。

$$V = \sigma^2 = \frac{\text{データの偏差平方和}}{\text{データ数}-1}$$

$$= \frac{(X_1-\overline{X})^2+(X_2-\overline{X})^2+\cdots+(X_N-\overline{X})^2}{N-1}$$

- Excelでは、関数VARA(データ範囲)を使って求める。

（3） 標準偏差

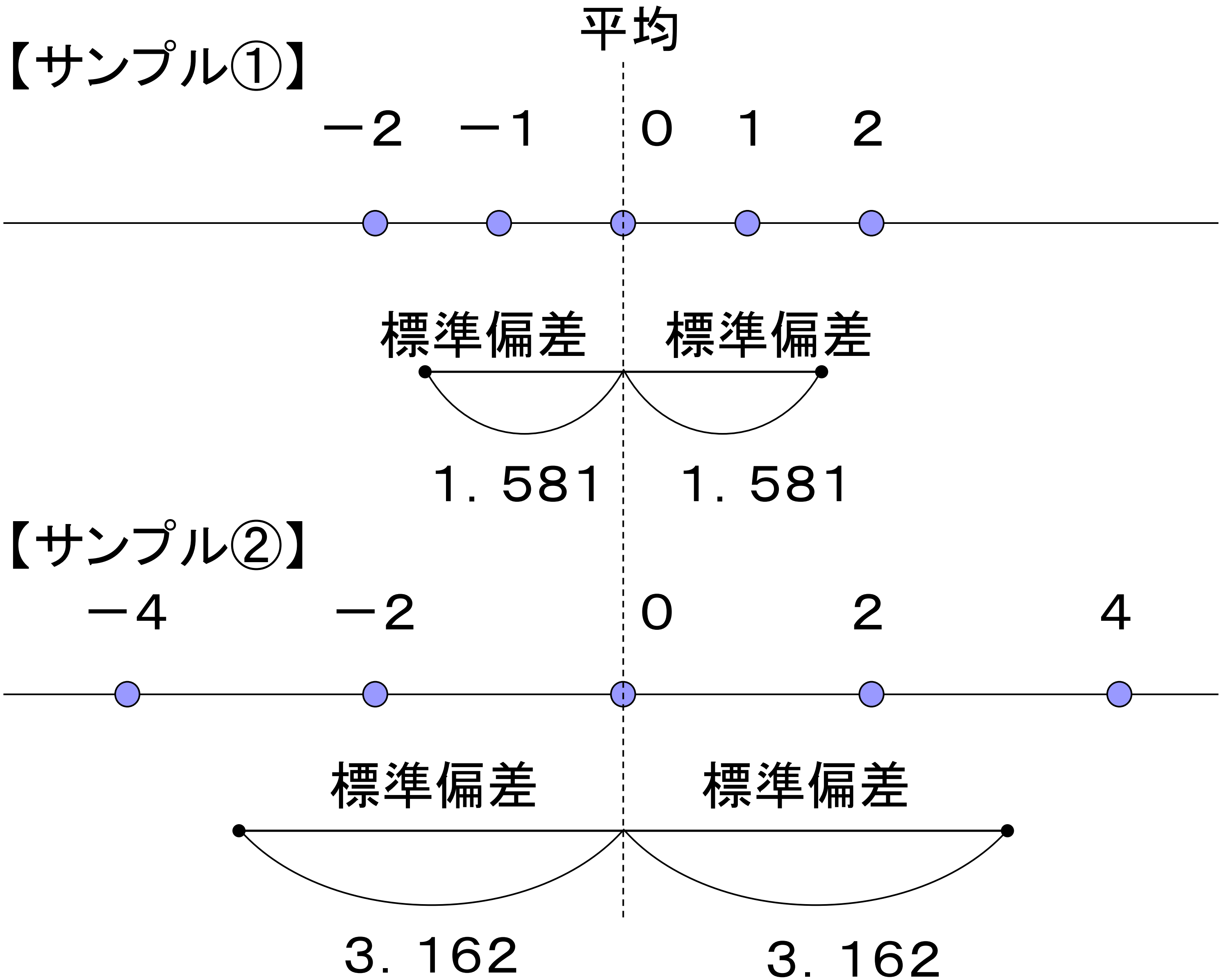
- 標準偏差は、観測データセットの「バラツキ」を示す指標の1つ。分散の平方根（ルート）をとって定義する。

　－－　標準偏差の「単位」は、データの持つ「単位」と同じ。

$$\sigma = \sqrt{\frac{\text{データの偏差平方和}}{\text{データ数}-1}}$$

$$= \sqrt{\frac{(X_1-\overline{X})^2+(X_2-\overline{X})^2+\cdots+(X_N-\overline{X})^2}{N-1}}$$

- Excelでは、関数STDEVA（データ範囲）を使って求める。


平均
【サンプル①】
－2
－1
0
1
2
標準偏差
標準偏差
1．581
1．581
【サンプル②】
－4
－2
0
2
4
標準偏差
標準偏差
3．162
3．162

（4）パーセント点

- パーセント点とは、観測データを小さい順に並べたときに、その値よりも小さな値の割合が指定された割合（百分率）になるデータの値として定義される。

- 例えば、99パーセント点というのは、その値より小さなデータの割合が99%となるデータの値のことを指す。
  - 50パーセント点のことを中央値（メジアン）と呼ぶ。
  - 25パーセント点を第1四分位点、75パーセント点を第3四分位点と呼ぶ。

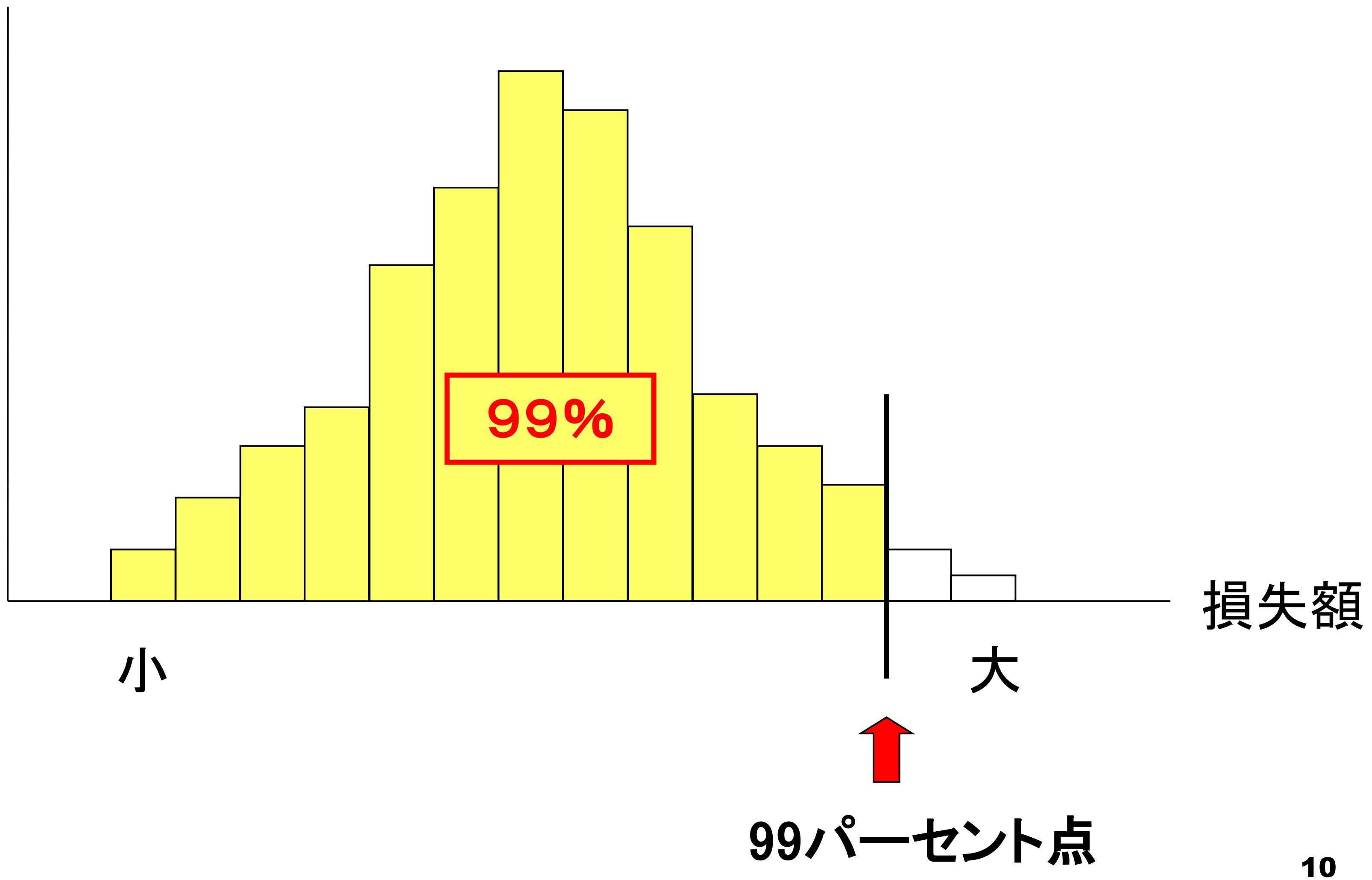
- Excelでは、関数PERCENTILE（データ範囲,率）を使って求める。

（例） 1000個の損失データが観測されている場合、
99%点というのは、損失額を小さい順に並べて
990番目になるデータ値のこと。

順位	百分位	損失額
985 番目	98.5%	529
986 番目	98.6%	558
987 番目	98.7%	589
988 番目	98.8%	618
989 番目	98.9%	621
990 番目	99.0%	632
991 番目	99.1%	654
992 番目	99.2%	671
993 番目	99.3%	698
994 番目	99.4%	703
995 番目	99.5%	712
996 番目	99.6%	776
997 番目	99.7%	794
998 番目	99.8%	810
999 番目	99.9%	831
1000 番目	100.0%	869

99%点

ヒストグラムで表したときの９９パーセント点



（参考1）対数変化率

- VaRの計測にあたり、観測データ・セットとして、リスクファクターの変化率をみることがある。

- このとき、統計的に扱い易い「**対数変化率**」を採用することが多い。

⇒　「対数変化率」の定義は？

どんな特徴があるのか？

対数変化率の定義

日次対数変化率

$$\log\frac{X_t}{X_{t-1}} \fallingdotseq \frac{X_t - X_{t-1}}{X_{t-1}} = \frac{X_t}{X_{t-1}} - 1$$

10日間対数変化率

$$\log\frac{X_t}{X_{t-10}} \fallingdotseq \frac{X_t - X_{t-10}}{X_{t-10}} = \frac{X_t}{X_{t-10}} - 1$$

- 対数変化率は、通常の変化率と近似的に等しいことが知られている。
- log(自然対数)は、Excelでは関数LN(・)で与えられる。

対数変化率の特徴

- 対数変化率は、同率の低下、上昇により、元の値に戻る。
- 10日間対数変化率は、日次対数変化率(10日分)の和となる。

	変化率(日次)	対数変化率(日次)
100	0.0101	0.0101
99	−0.0100	−0.0101
100	0.0526	0.0513
95	−0.0500	−0.0513
100	0.1111	0.1054
90	−0.1000	−0.1054
100	0.2500	0.2231
80	−0.2000	−0.2231
100	0.4286	0.3567
70	−0.3000	−0.3567
100	0.6667	0.5108
60	−0.4000	−0.5108
100	1.0000	0.6931
50	−0.5000	−0.6931
100	—	—

		対数変化率(日次)
X10	100	0.2877
X9	75	−0.4700
X8	120	1.3863
X7	30	−0.6931
X6	60	−0.9163
X5	150	0.5108
X4	90	1.0986
X3	30	−0.6931
X2	60	−0.2877
X1	80	−0.1178
X0	90	—
$\Sigma \log(X_t/X_{t-1})$		0.1054

	対数変化率(10日間)
log(X10/X0)	0.1054

（参考2）対数変化率と√T倍法の適用

- 10日間対数変化率は、日次対数変化率（10日間）の「和」となる。

  0日目 X_0　1日目 X_1　2日目 X_2　・・・　10日目 X_{10}

  数式で表すと　$\log(X_{10}/X_0)$

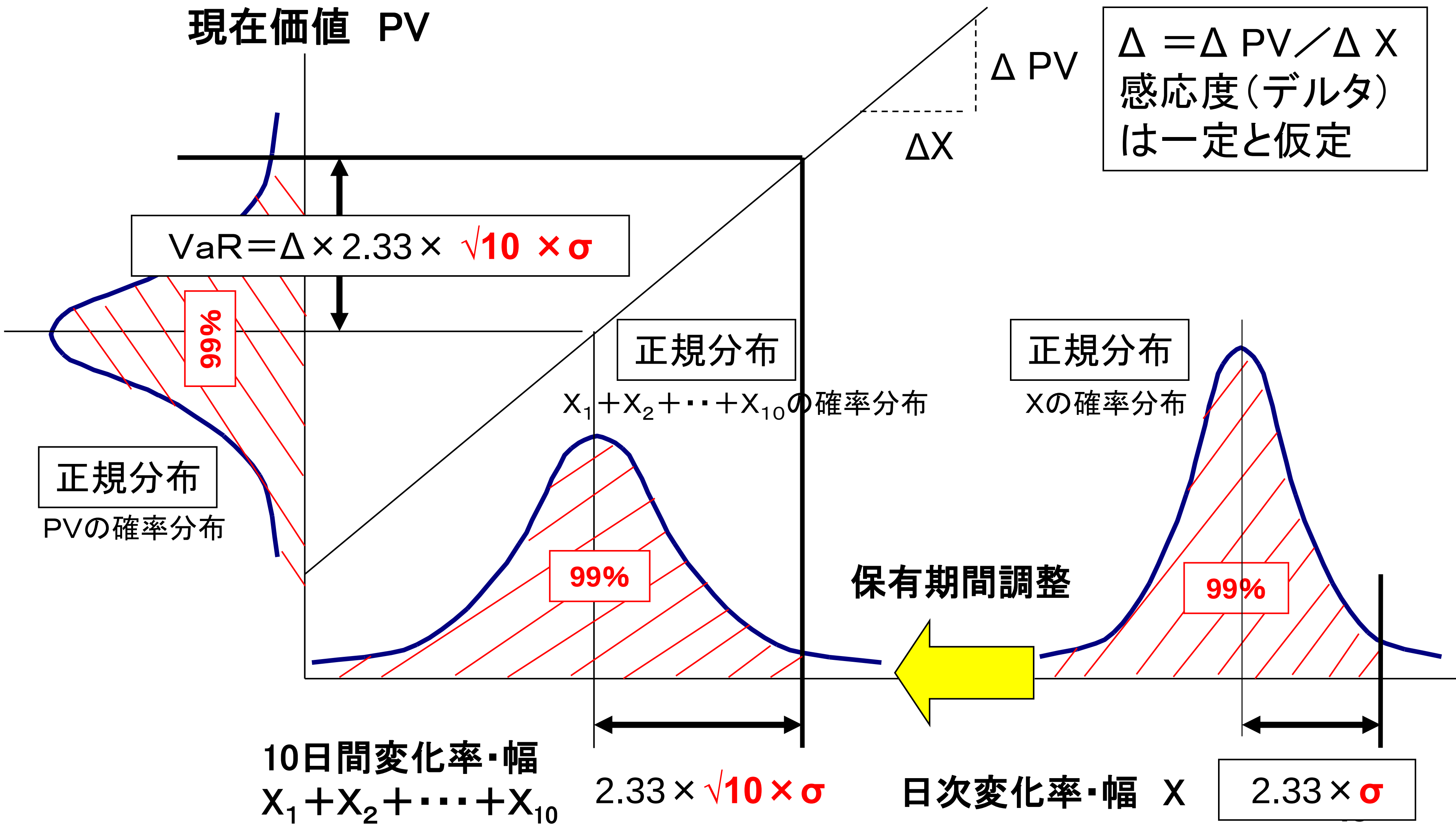
  $= \log\{(X_{10}/X_9)(X_9/X_8)\cdots(X_1/X_0)\}$

  $= \log(X_{10}/X_9) + \log(X_9/X_8) + \cdots + \log(X_1/X_0)$

- 『日次変化率が、互いに独立な確率変数であり、
  その分散がσ^2（標準偏差がσ）のとき、
  10日間対数変化率の分散は $10\sigma^2$（標準偏差は $\sqrt{10}\sigma$）
  となる』 ことが知られている。

- リスクファクターの日次対数変化率が、互いに独立で分散（標準偏差）の等しい確率変数であるとすれば、√T倍法を適用可能となる。

√T倍法による保有期間調整（イメージ図）



２. 基本統計量(２変量)

（１）散布図

（２）共分散

（３）相関係数

（４）相関行列

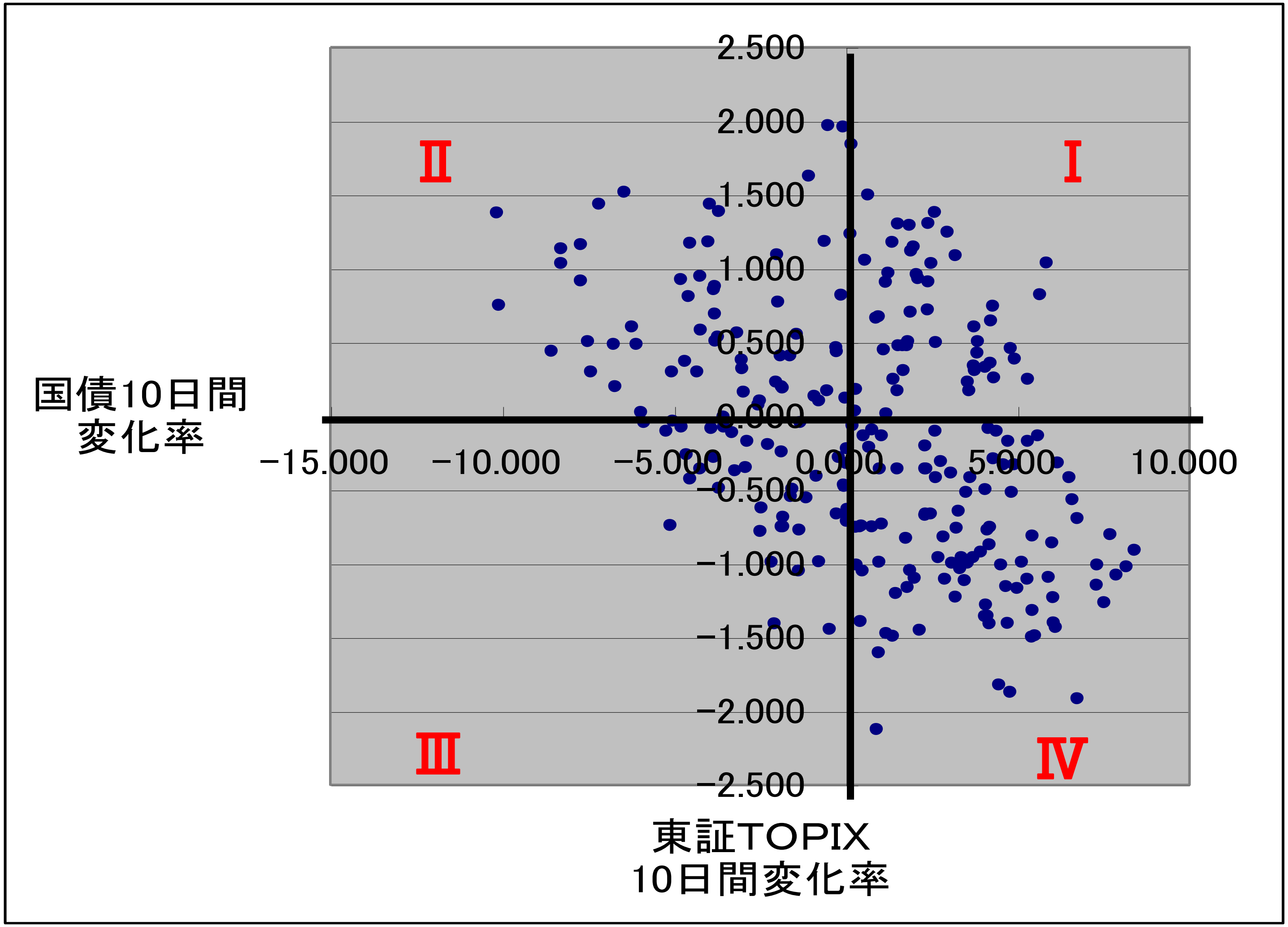
（1） 散布図

- 以下のような2変量の関係を調べるためには、散布図を書くのが直感的に理解しやすい。

	東証TOPIX 10日間変化率 (X)	10年割引国債 10日間変化率 (Y)
2006/9/29	0.785	−0.098
2006/9/28	1.194	0.010
2006/9/27	0.319	0.177
2006/9/26	−2.994	0.315
2006/9/25	−3.783	0.688
2006/9/22	−3.139	0.560
2006/9/21	−3.894	−0.088
2006/9/20	−5.040	0.295
2006/9/19	−3.538	−0.010
2006/9/15	−2.474	0.098
⋮	⋮	⋮

国債と株価の相関関係

◆ Ⅱ、Ⅳのエリアに分布が多く、「負の相関」が観察される。



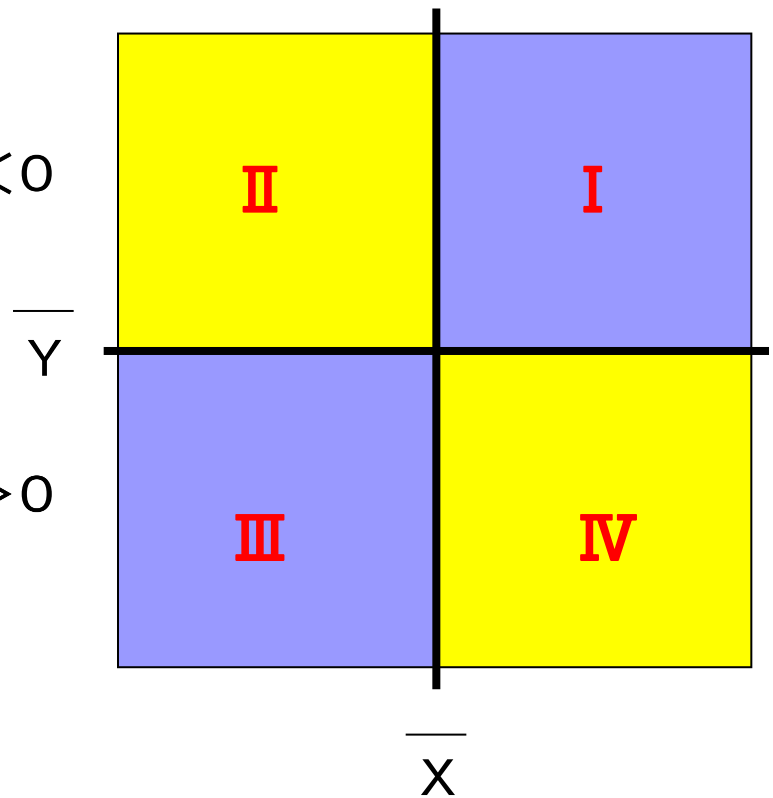
偏差積和

$$= (X_1-\overline{X})(Y_1-\overline{Y})+(X_2-\overline{X})(Y_2-\overline{Y})+\cdots+(X_N-\overline{X})(Y_N-\overline{Y})$$

Ⅰ、ⅢのエリアIに多く分布 ⇒ 偏差積和 ＞ 0 ： 正の相関
Ⅱ、Ⅳのエリアに多く分布 ⇒ 偏差積和 ＜ 0 ： 負の相関

$(X_i-\overline{X})(Y_i-\overline{Y})<0$

$(X_i-\overline{X})(Y_i-\overline{Y})>0$



$(X_i-\overline{X})(Y_i-\overline{Y})>0$

$(X_i-\overline{X})(Y_i-\overline{Y})<0$

（2）共分散

- 共分散は、2つの変量（X、Y）の間の「直線的な比例関係の強さ」を示す指標。

－－　データの「偏差積和」を求めて、「データ数－１」で割る。

－－　共分散の「単位」は、Xの持つ「単位」 掛ける Yの持つ「単位」。

$$\mathrm{COV}(X,Y)=\frac{\text{データの偏差積和}}{\text{データ数}-1}$$

$$=\frac{(X_1-\overline{X})(Y_1-\overline{Y})+(X_2-\overline{X})(Y_2-\overline{Y})+\cdot\cdot+(X_N-\overline{X})(Y_N-\overline{Y})}{N-1}$$

- Excelでは、関数COVAR（データ範囲（X）、データ範囲（Y））を使って求める。

（注）Excelでは、データの偏差積和をN－１ではなく、Nで割って共分散を定義しているため、調整を行う必要がある。

（3）相関係数

- 相関係数は、2つの変量（X、Y）間の「直線的な比例関係の強さ」を示す指標。共分散を、それぞれの標準偏差の積で割って定義する。

  －－　相関係数は　－1～　＋1　までの値をとる。

  －－　相関係数は「単位」を持たない無名数。

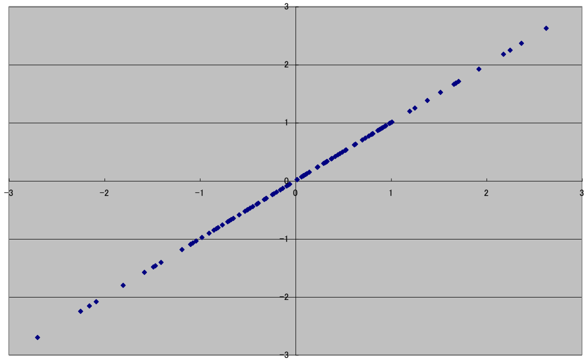
$$\rho(X, Y) = \frac{COV(X, Y)}{\sigma(X)\,\sigma(Y)}$$

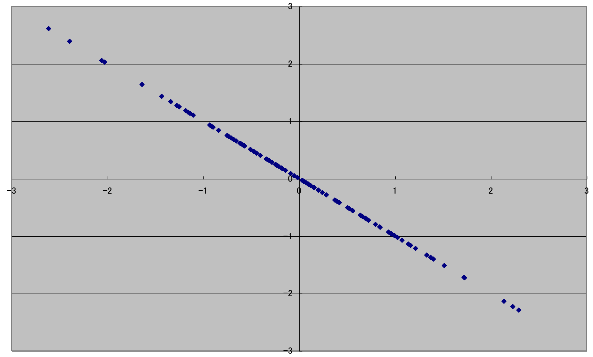
$$= \frac{(X_1-\overline{X})(Y_1-\overline{Y})+\cdots+(X_N-\overline{X})(Y_N-\overline{Y})}{\sqrt{(X_1-\overline{X})^2+\cdots+(X_N-\overline{X})^2}\sqrt{(Y_1-\overline{Y})^2+\cdots+(Y_N-\overline{Y})^2}}$$

- Excelでは、関数CORELL（データ範囲（X）、データ範囲（Y））を使って求める。

相関係数と散布図

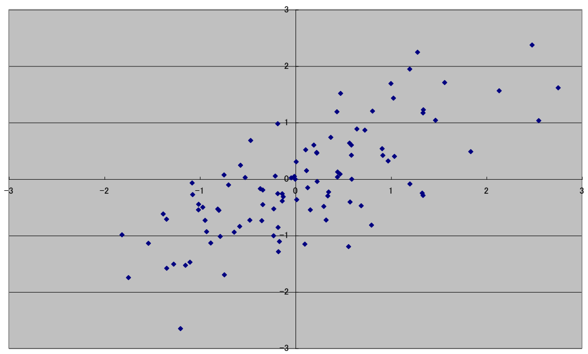
ρ＝1.0
（正の完全相関）

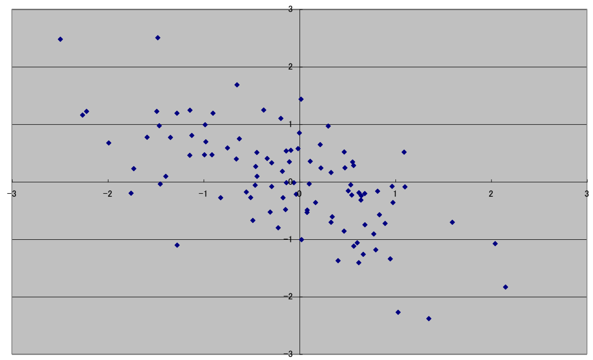




ρ＝－1.0
（負の完全相関）

ρ＝0.7





ρ＝－0.7

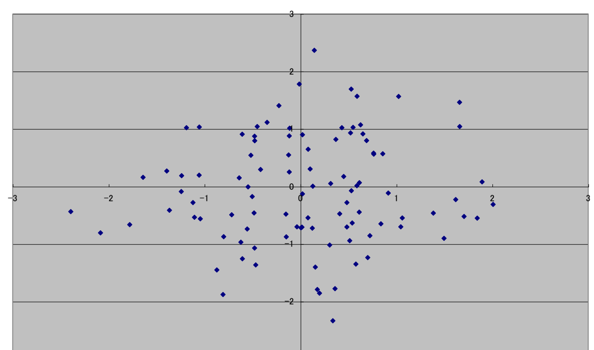
相関係数の定義

ρ xy＝ COV(X,Y)／σ xσ y

COV(X,Y) ： X,Yの共分散 ＝(1/N-1)*Σ (Xt－EX)(Yt－EY)

σ x ： Xの標準偏差 EX ： Xの平均値

σ y ： Yの標準偏差 EY ： Yの平均値



ρ＝0
（無相関）

（4）相関行列と分散共分散行列

相関行列

	X_1	X_2	X_3	⋯	X_N
X_1	1	ρ（X_1、X_2）	ρ（X_1、X_3）	⋯	ρ（X_1、X_N）
X_2	**ρ（X_2、X_1）**	1	ρ（X_2、X_3）	⋯	ρ（X_2、X_N）
X_3	**ρ（X_3、X_1）**	**ρ（X_3、X_2）**	1	⋯	ρ（X_1、X_2）
⋮	⋮	⋮	⋮	⋱	⋮
X_N	**ρ（X_N、X_1）**	**ρ（X_N、X_2）**	**ρ（X_N、X_3）**	⋯	1

ρ（X_i、X_i）＝1　：　同じ変量（X_{ii}）同士の相関は1

ρ（X_i、X_j）＝ρ（X_j、X_i）：　2つの変量（X_i、X_j）の順序を変えて計算しても相関係数の値は同じ。

分散共分散行列

	X_1	X_2	X_3	・・・	X_N
X_1	V_{X1}	COV（X_1、X_2）	COV（X_1、X_3）	・・・	COV（X_1、X_N）
X_2	COV（X_2、X_1）	V_{X2}	COV（X_2、X_3）	・・・	COV（X_2、X_N）
X_3	COV（X_3、X_1）	COV（X_3、X_2）	V_{X3}	・・・	COV（X_1、X_2）
⋮	⋮	⋮	⋮	⋱	⋮
X_N	COV（X_N、X_1）	COV（X_N、X_2）	COV（X_N、X_3）	・・・	V_{XN}

相関考慮後のVaR計算式①（分散共分散法）

相関考慮後のポートフォリオVaR　＝

$$
\sqrt{
\overset{\text{（単独VaR）}}{\begin{pmatrix} VaR(X_1) & VaR(X_2) & \cdots & VaR(X_N) \end{pmatrix}}
\overset{\text{（相関行列）}}{\begin{pmatrix} 1 & \rho(X_1, X_2) & \cdots & \rho(X_1, X_N) \\ \rho(X_1, X_2) & 1 & \cdots & \rho(X_2, X_N) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(X_1, X_N) & \rho(X_N, X_2) & \cdots & 1 \end{pmatrix}}
\overset{\text{（単独VaR）}}{\begin{pmatrix} VaR(X_1) \\ VaR(X_2) \\ \vdots \\ VaR(X_N) \end{pmatrix}}
}
$$

相関考慮後のVaR計算式②（分散共分散法）

ポートフォリオ現在価値の標準偏差（σ_p） ＝

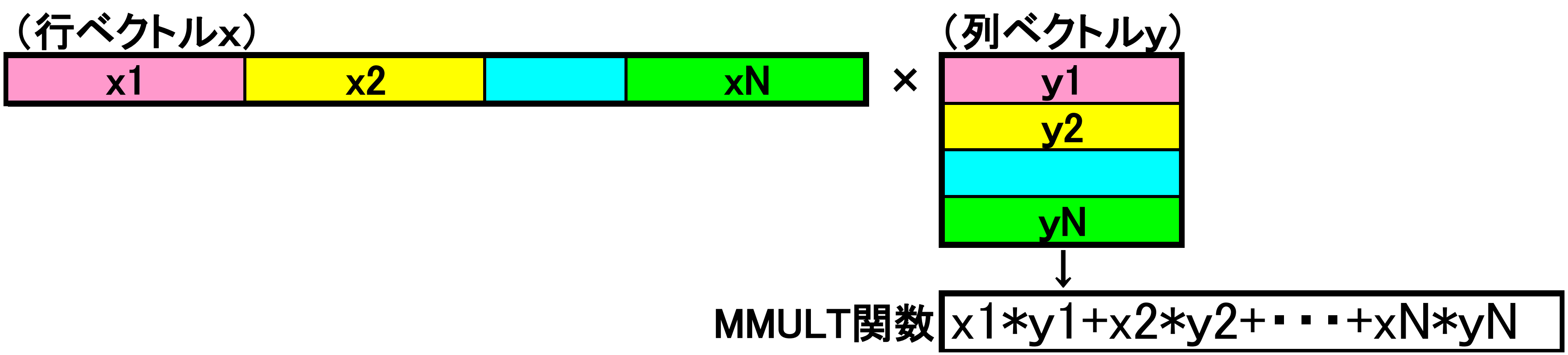
$$
\sqrt{
\overset{\text{（デルタ）}}{\begin{pmatrix} \Delta_{X1} & \Delta_{X2} & \cdots & \Delta_{XN} \end{pmatrix}}
\overset{\text{（分散共分散行列）}}{\begin{pmatrix}
V_{X1} & \mathrm{COV}(X_1、X_2) & \cdots & \mathrm{COV}(X_1、X_N) \\
\mathrm{COV}(X_1、X_2) & V_{X2} & \cdots & \mathrm{COV}(X_2、X_N) \\
\vdots & \vdots & \ddots & \vdots \\
\mathrm{COV}(X_1、X_N) & \mathrm{COV}(X_N、X_2) & \cdots & V_{XN}
\end{pmatrix}}
\overset{\text{（デルタ）}}{\begin{pmatrix} \Delta_{X1} \\ \Delta_{X2} \\ \vdots \\ \Delta_{XN} \end{pmatrix}}
}
$$

相関考慮後のポートフォリオVaR ＝ 信頼係数× σ_p

（参考）行列計算式（基本型）

- 行ベクトル（1行×N列）と列ベクトル（N行×1列）の掛け算はExcelでは、MMULT関数を利用して行う。

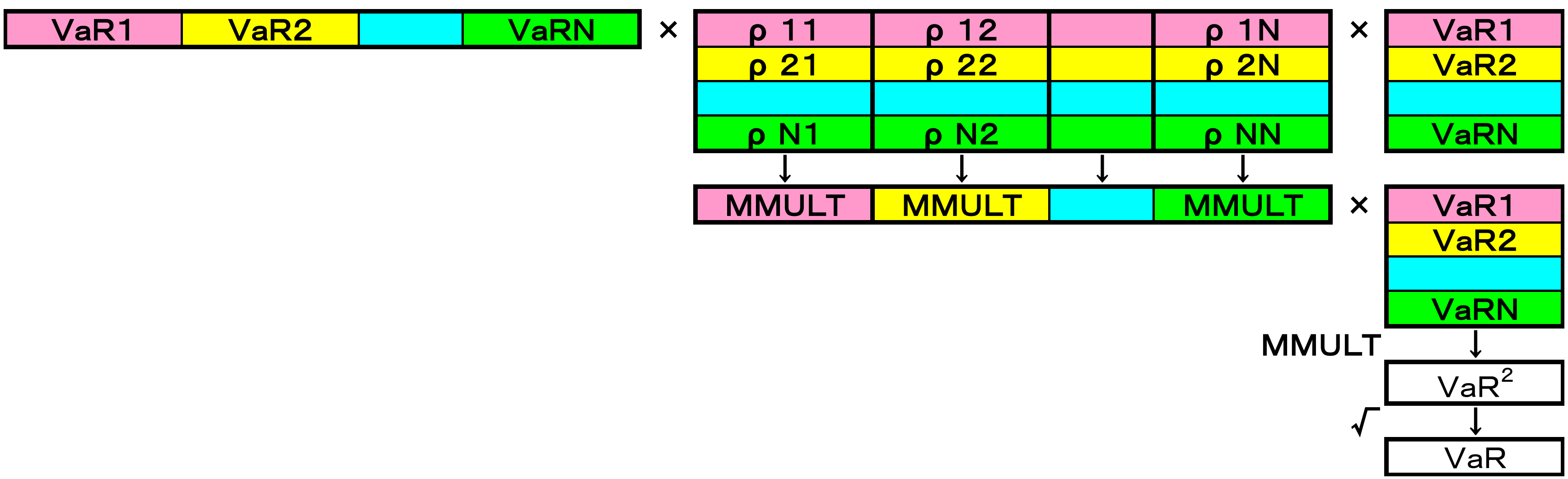
行列計算式の基本型



（参考）行列計算式（相関考慮後のVaR）

- 行列の掛け算は、MMULT関数を利用した基本型の繰り返しで計算できる。

相関考慮後VaRの行列計算式



３．確率変数と確率分布

（１）確率変数

（２）確率分布

－　確率密度関数、分布関数

（３）様々な確率分布

－　一様分布、正規分布、2項分布

（４）確率変数の期待値

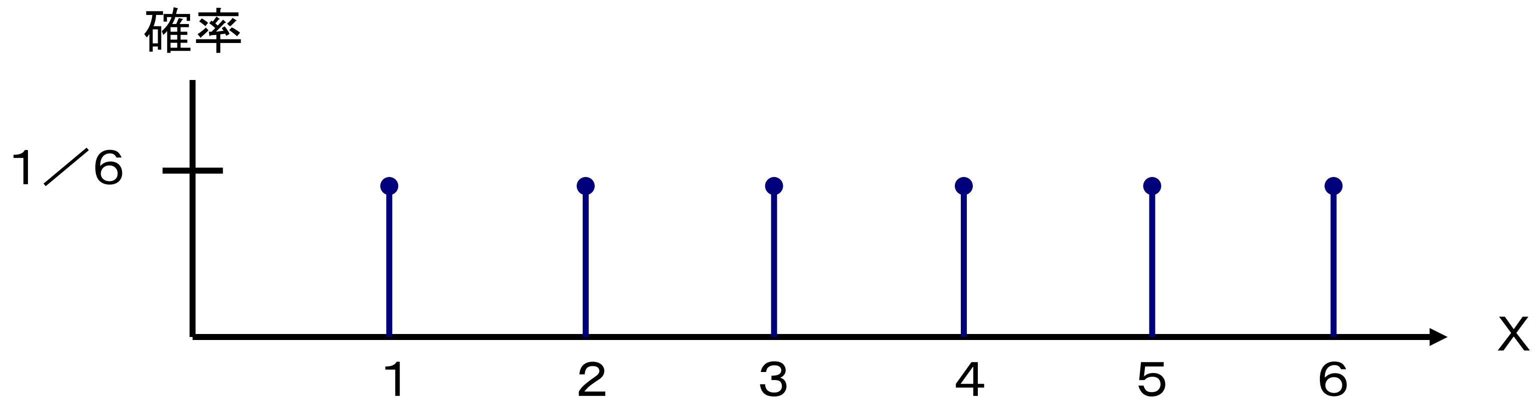
（５）確率変数の独立

（1）確率変数

- 予め定まった確率にしたがって値が変動する数のことを「確率変数」という

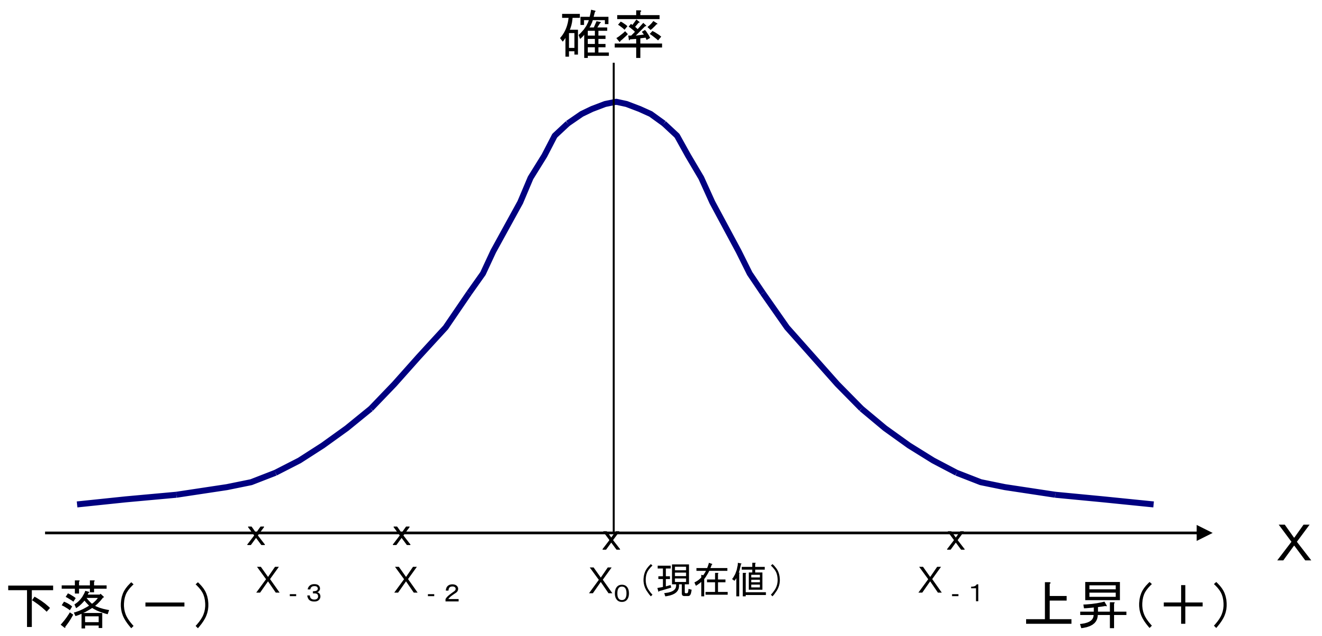
（例）サイコロを振ったときに出る目の数

サイコロの目（X）	1	2	3	4	5	6
確　　率	1／6	1／6	1／6	1／6	1／6	1／6



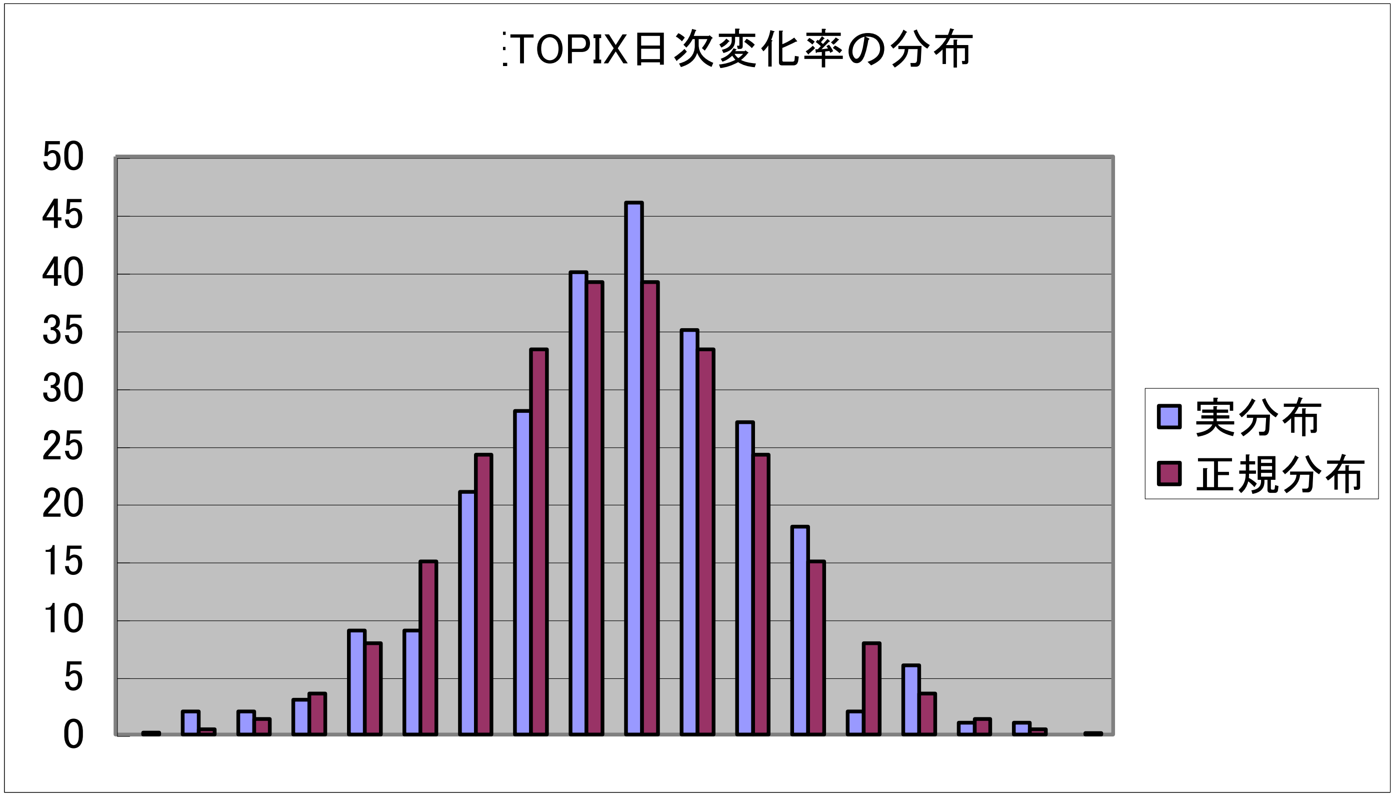
- 株価、金利、為替等のリスクファクターの変化率について「確率変数」として捉えることもできる。

（例）TOPIXの変化率（X）



- リスクファクターの変化率の分布は、正規分布（後述）にしたがうと想定されることが多い。
- しかし、実際の分布をみると、歪み、偏りやファット・テール（注）が観察されることも少なくない。

（注）両端部分の裾野の分布が厚くなることをいう。



(2)確率分布

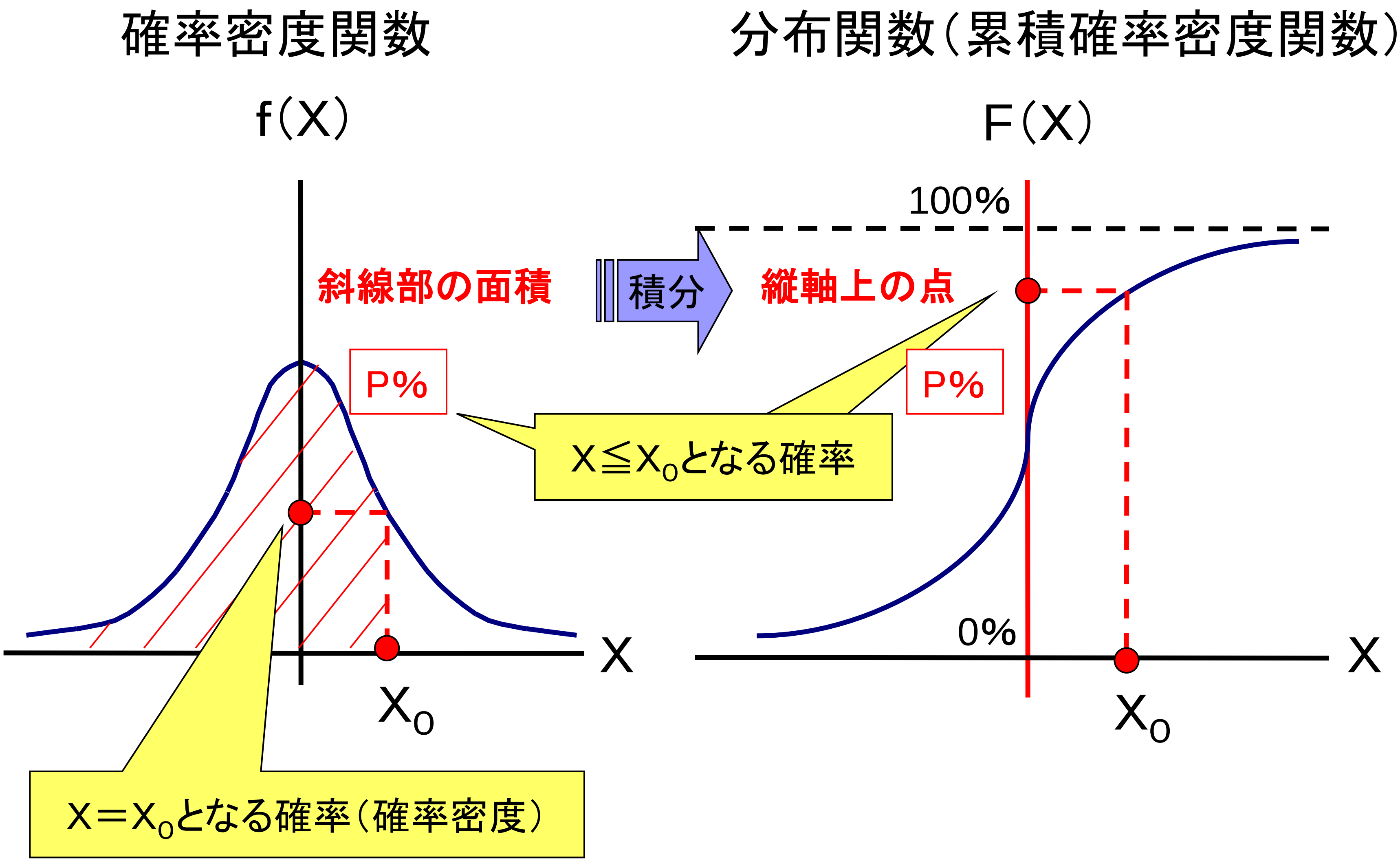
- 確率分布を表わすとき、2種類の関数がある。

① 確率密度関数

確率変数(X)が **「ある値」をとる確率(確率密度)** を表わす関数

② 分布関数(累積確率密度関数)

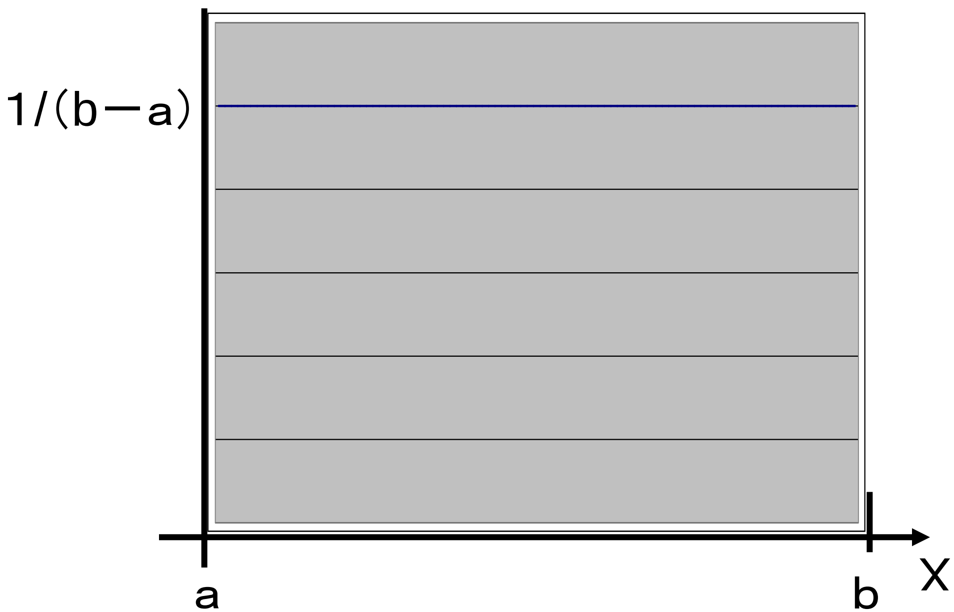
確率変数(X)が **「ある値以下」になる確率** を表わす関数


確率密度関数
分布関数（累積確率密度関数）
f（X）
F（X）
斜線部の面積
積分
縦軸上の点
100%
0%
P%
P%
X≦X₀となる確率
X
X
X₀
X₀
X＝X₀となる確率（確率密度）

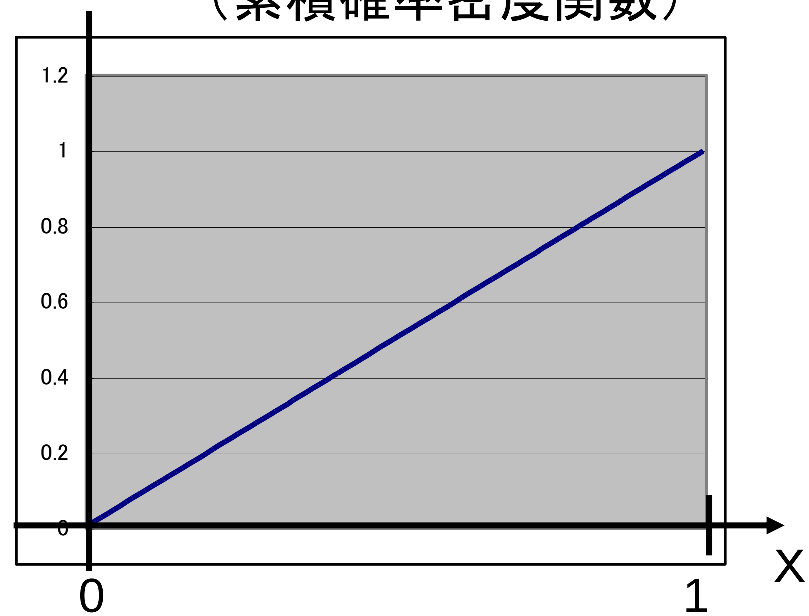
（2）様々な確率分布

- **一様分布**：　ある区間の中の値が同じ確率で生起する分布。

f(X)　確率密度関数



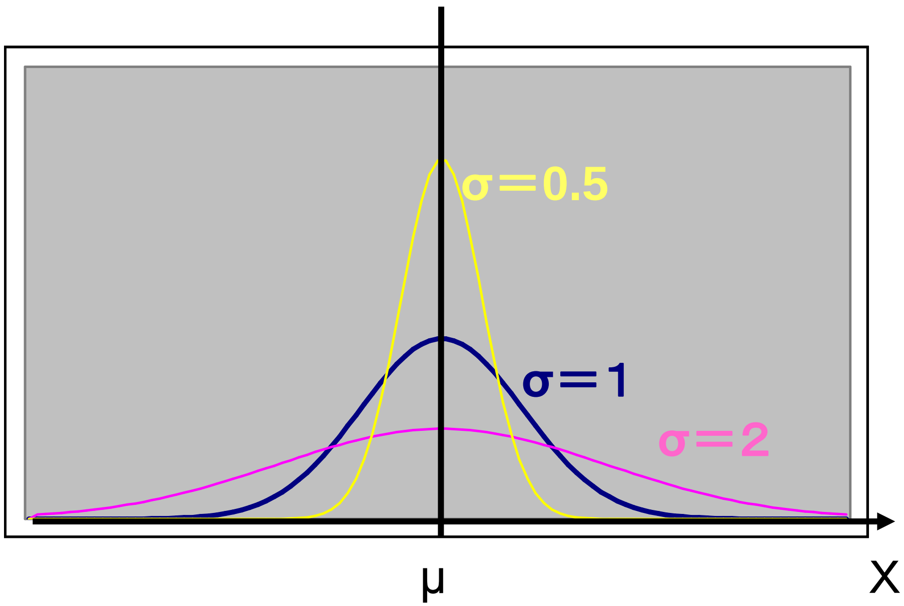
F(X)　分布関数
（累積確率密度関数）



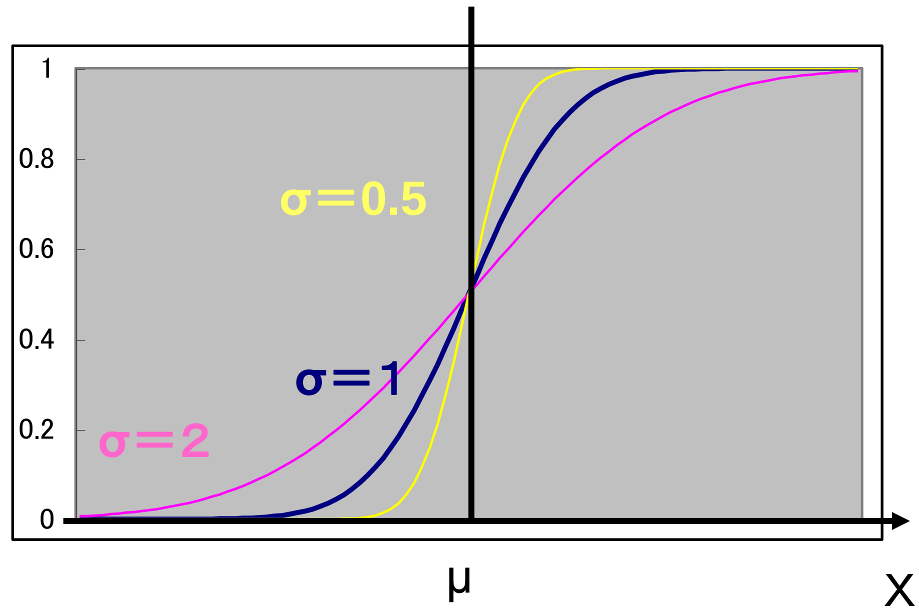
- 一様分布にしたがう乱数（一様乱数）は、Excel関数RAND()を使って生成することができる。

- **正規分布**：　左右対称の釣鐘型をした確率分布。
  平均（μ ）、標準偏差（σ ）を与えると分布の形状が決まるため、N（μ ,σ ^{2}）と表す。

f（X）　確率密度関数



F（X）　分布関数
（累積確率密度関数）



- 平均（μ ）＝0、標準偏差（σ ）＝1の正規分布を**標準正規分布**と言い、N（0,1）と表す。

正規分布の特徴

- 平均からどれだけ離れているか（標準偏差の何倍か）という情報から、X以下の値をとる確率が分かる。
- 例えば、XがN(0, σ^2)の正規分布にしたがって生起するとき

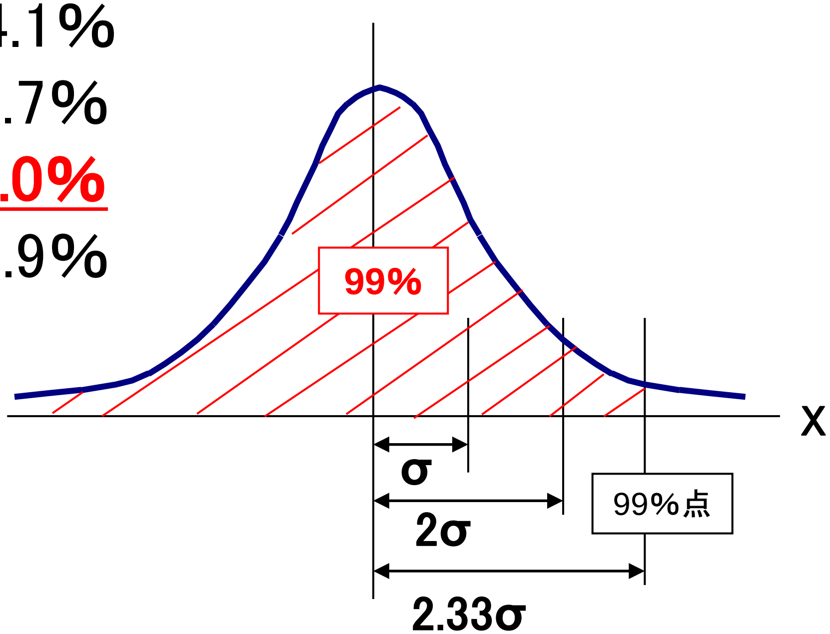
$X \leqq \sigma$ となる確率は　84.1%

$X \leqq 2\sigma$ となる確率は　97.7%

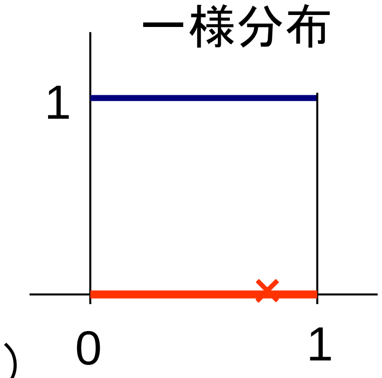
$X \leqq 2.33\sigma$ となる確率は　99.0%

$X \leqq 3\sigma$ となる確率は　99.9%

となることが知られている。



正規乱数の生成方法(一様乱数から作る方法)



（ｉ） 一様乱数を作る(右図 ×)。

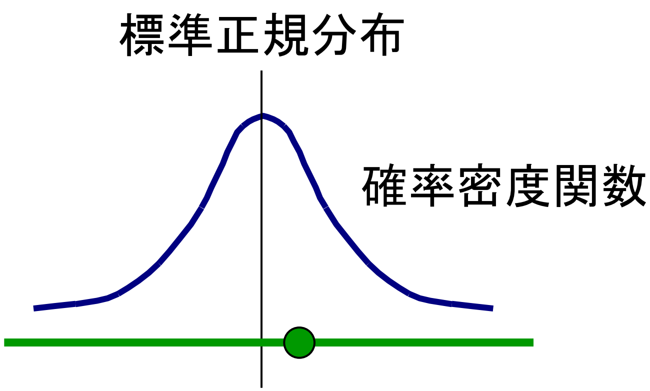
Rand()

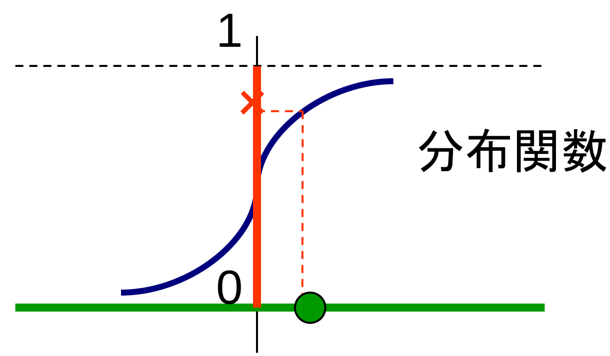
： 0以上で1より小さい乱数を発生させる。

（ⅱ） 一様乱数を標準正規乱数 に変換する(下図 × ●)

Normsinv(Rand())

： 一様乱数の値を、**標準正規分布の「分布関数の逆関数」**に代入すると、標準正規乱数に変換される。





（ⅲ） 標準正規乱数を(ⅱ)×σ ＋μ により、正規乱数～N(μ 、σ ^{2}) に変換する。

（ⅳ） 正規乱数の生成方法には、様々なものがあり、どの方法が優れているか研究の対象となっている。上記方法は一例に過ぎない

- **２項分布**： 結果が2通りある試行（実験）をN回繰り返したとき、２通りの結果のうち一方が起こる回数の確率分布

（例）サイコロを10回振って １の目が出る回数（K）

0回　$f(0) = {}_{10}C_0 (1/6)^0 (5/6)^{10}$

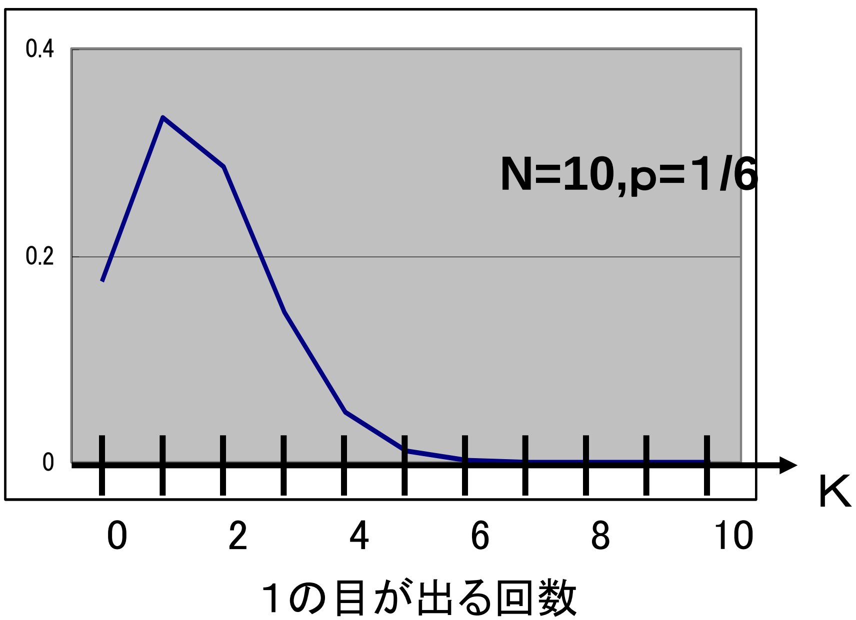
1回　$f(1) = {}_{10}C_1 (1/6)^1 (5/6)^9$

2回　$f(2) = {}_{10}C_2 (1/6)^2 (5/6)^8$

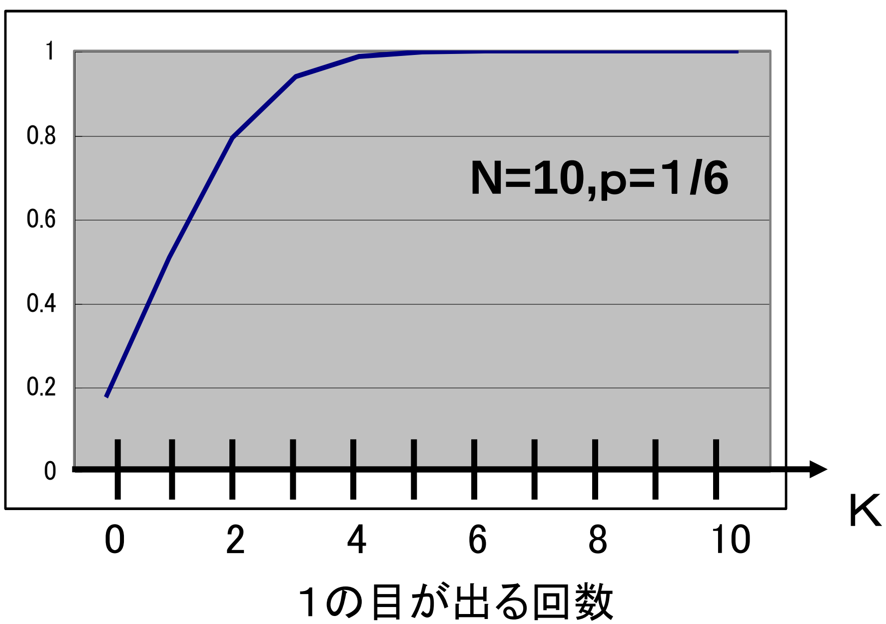
⋮　　⋮　　⋮

10回　$f(10) = {}_{10}C_{10} (1/6)^{10} (5/6)^0$

f(K)　確率



F(K)　分布関数（累積確率）



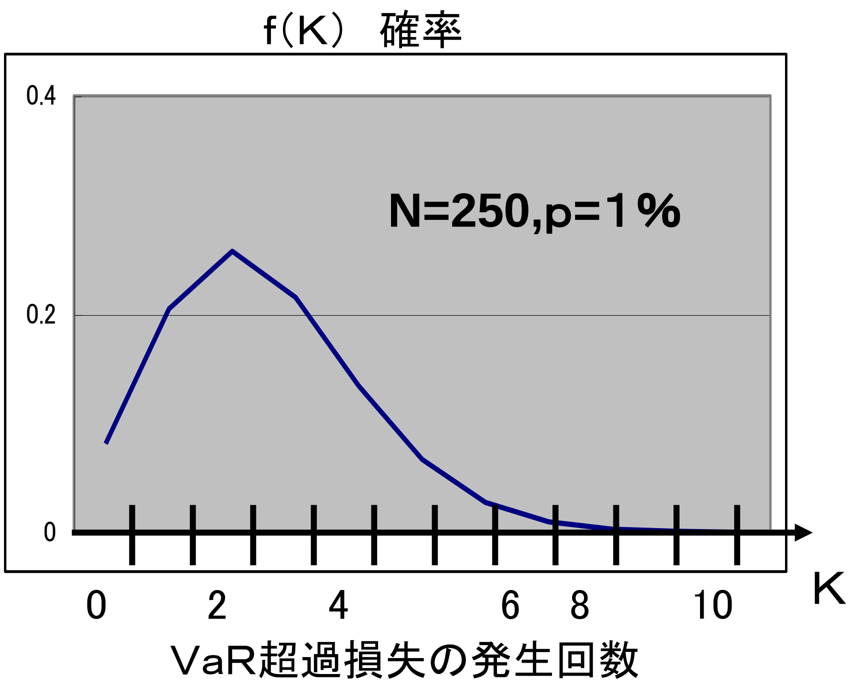
（例）　VaRを超過する損失が発生する回数（K）

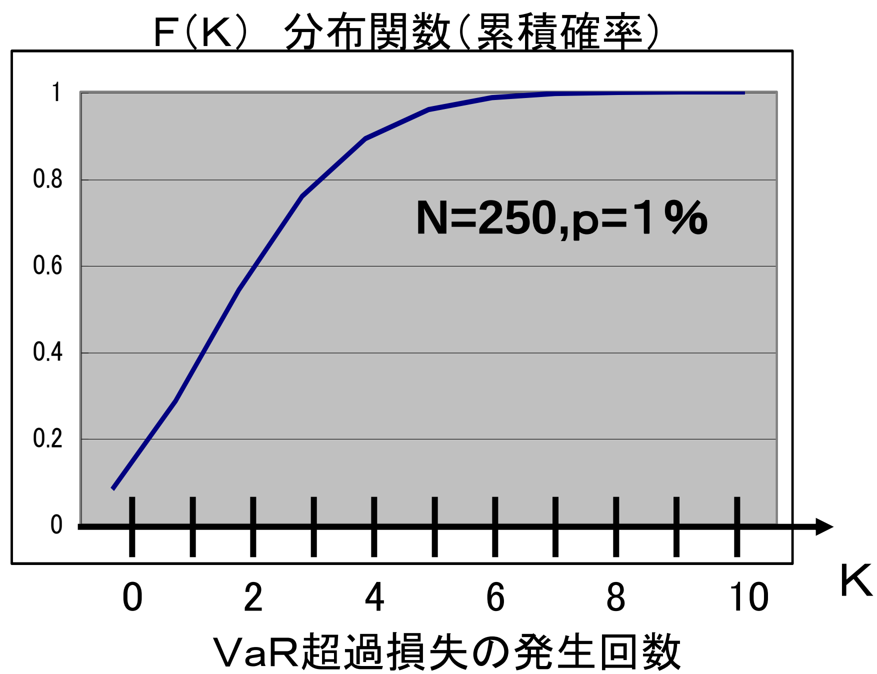
VaRを超過する確率　　$p = 1\%$

VaRを超過しない確率　$1-p = 99\%$（信頼水準）

VaRの計測個数　　$N=250$

発生確率 $f(K) = {}_{250}C_K\,(0.01)^K\,(0.99)^{250-K}$

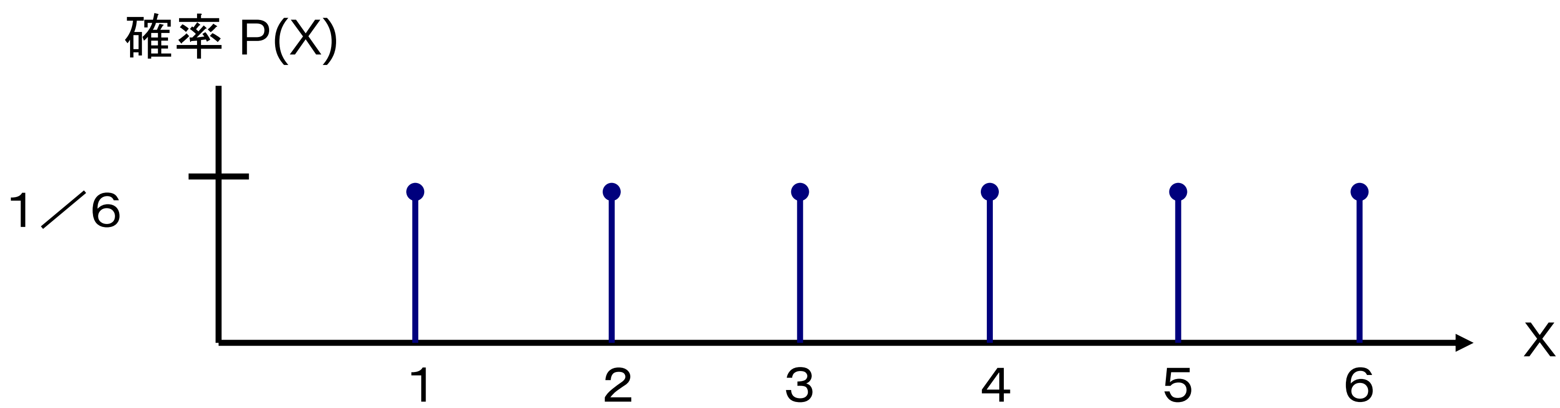




（4）確率変数の期待値

- 確率変数（X）は、平均的にみてどんな値をとるのか？

（例）サイコロを振ったときに出る目の数



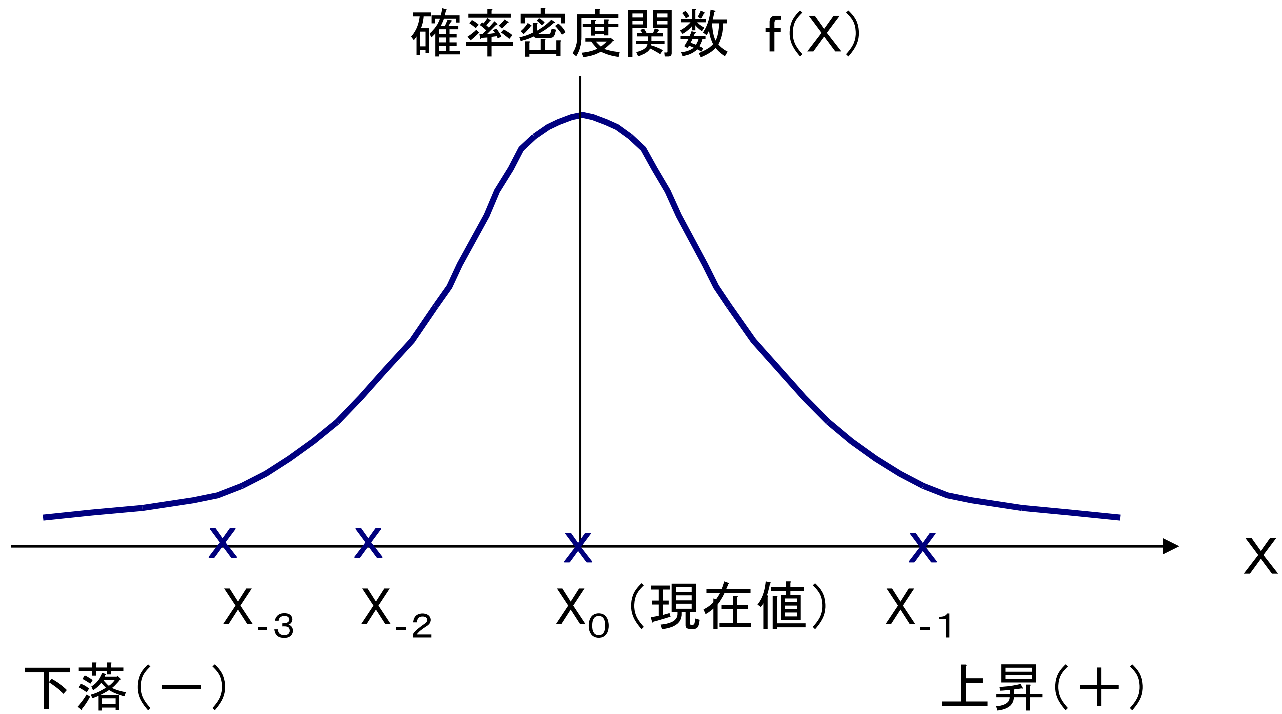
サイコロを振ったときに出る目の数の「期待値」

$$\sum_{X=1}^{6} XP(X)$$

$$= 1\times(1/6) + 2\times(1/6) + 3\times(1/6) + 4\times(1/6) + 5\times(1/6) + 6\times(1/6)$$

$$= 3.5$$

（例） TOPIXの変化率(X)



TOPIXの変化率(X)の期待値

$$\int_{-\infty}^{+\infty} X\, f(X)\, dX$$

（5）確率変数の独立

【定義】

- 確率変数　X、Y が互いに影響されず、それぞれの確率分布にしたがって値をとるとき、確率変数　X、Yは、互いに「独立」であるという。

数式で表すと　　$P(X=a, Y=b)=P(X=a)P(Y=b)$

【定理】

- 確率変数　X、Y が互いに「独立」のとき、以下のことが 成り立つ。

① 確立変数 XY の期待値は、それぞれの確率変数の期待値の積になる。

$$E(XY)=E(X)E(Y)$$

② 確率変数 X＋Y の分散は、それぞれの確率変数の分散の和に等しい。

$$V(X+Y)=V(X)+V(Y)$$

③ 確率変数 X と Y は無相関である。

$$\rho(X, Y)=0$$

（証明省略）

（例）サイコロを振ったときに出る目の数

１回目：　X_1　＝　１、　２回目：　X_2　＝　１

３回目：　X_3　＝　？

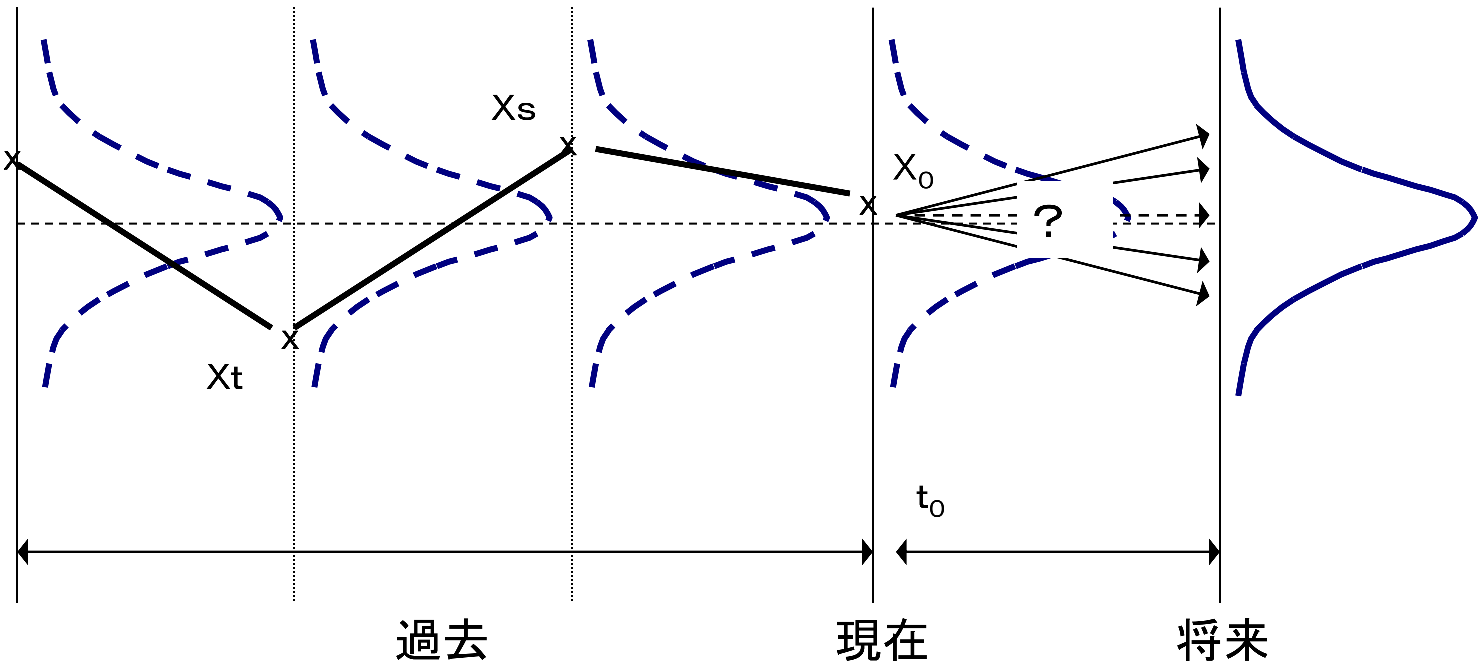
サイコロの目（X_3）	１	２	３	４	５	６
確　　率	１／６	１／６	１／６	１／６	１／６	１／６

- ２回続けて1の目が出ても、３回目の結果には影響を及ぼさない。
- ３回目は、いずれの目が出る確率も１／６。

（例）株価、金利、為替等リスクファクターの変化率

- 過去の変化率（実績）が、将来の変化率（予想）に影響を及ぼすことはないと考えて、互いに独立な確率変数として捉えることが多い。

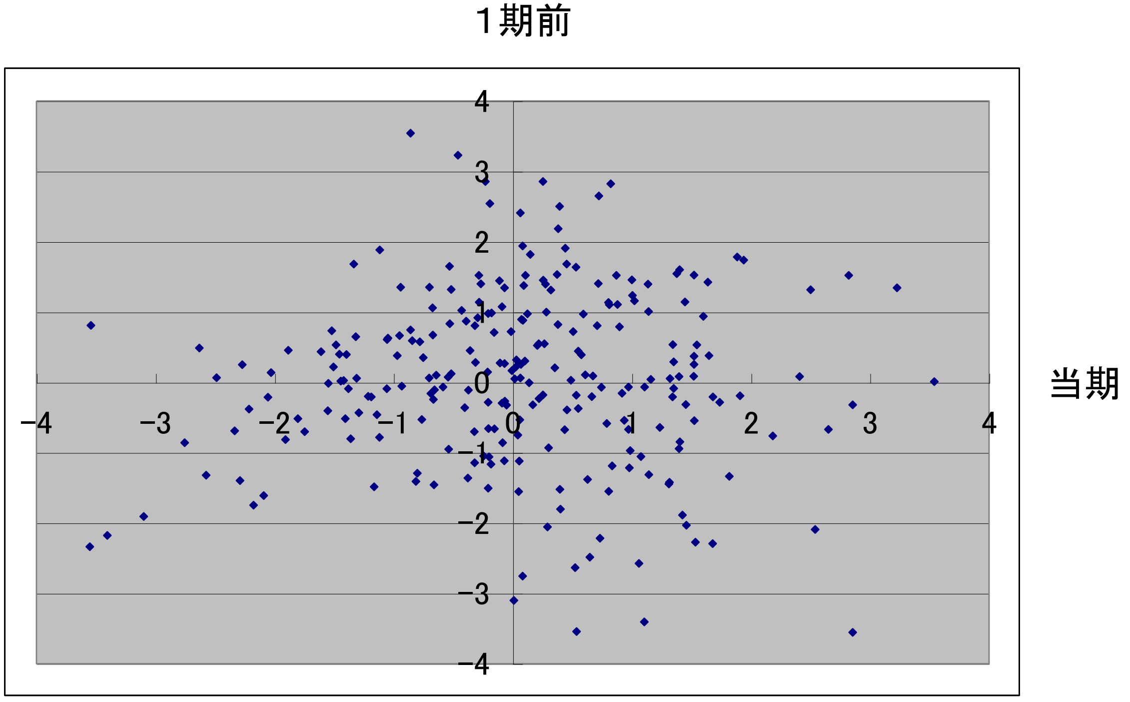
リスクファクター（X）の推移と、その確率分布



（注）山下智志（「市場リスクの計量化とVaR」2000）を参考に日本銀行が作成

- しかし、リスクファクターの変化率が時点間で独立とは限らず、相関関係が認められることも少なくないので注意を要する。

  - 下図は、TOPIX・日次対数変化率1期前の変化率との相関をみたもの。独立の判定には、様々なタイムラグを置いて相関の有無をみる必要。



相関係数ρ＝0.037

４．推定と検定

（１）記述統計と推測統計

（２）推定

（３）検定

(1)記述統計と推測統計

- 記述統計　：　基本統計量の算定や図表、グラフを利用して観測データが持つ特性を分析・記述する。

(例)特定の集団(N人)の身長の平均と分散を計算する。

平均

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

分散

$$Vp = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_N - \overline{X})^2}{N}$$

- 推測統計　：　標本として集めた一部の観測データに基づき、母集団の特性について推測し、検証する。

（例）任意に抽出したN人（標本）の身長を計測して、日本人全体（母集団）の身長の平均と分散を推定する。

平均

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_N}{N}$$

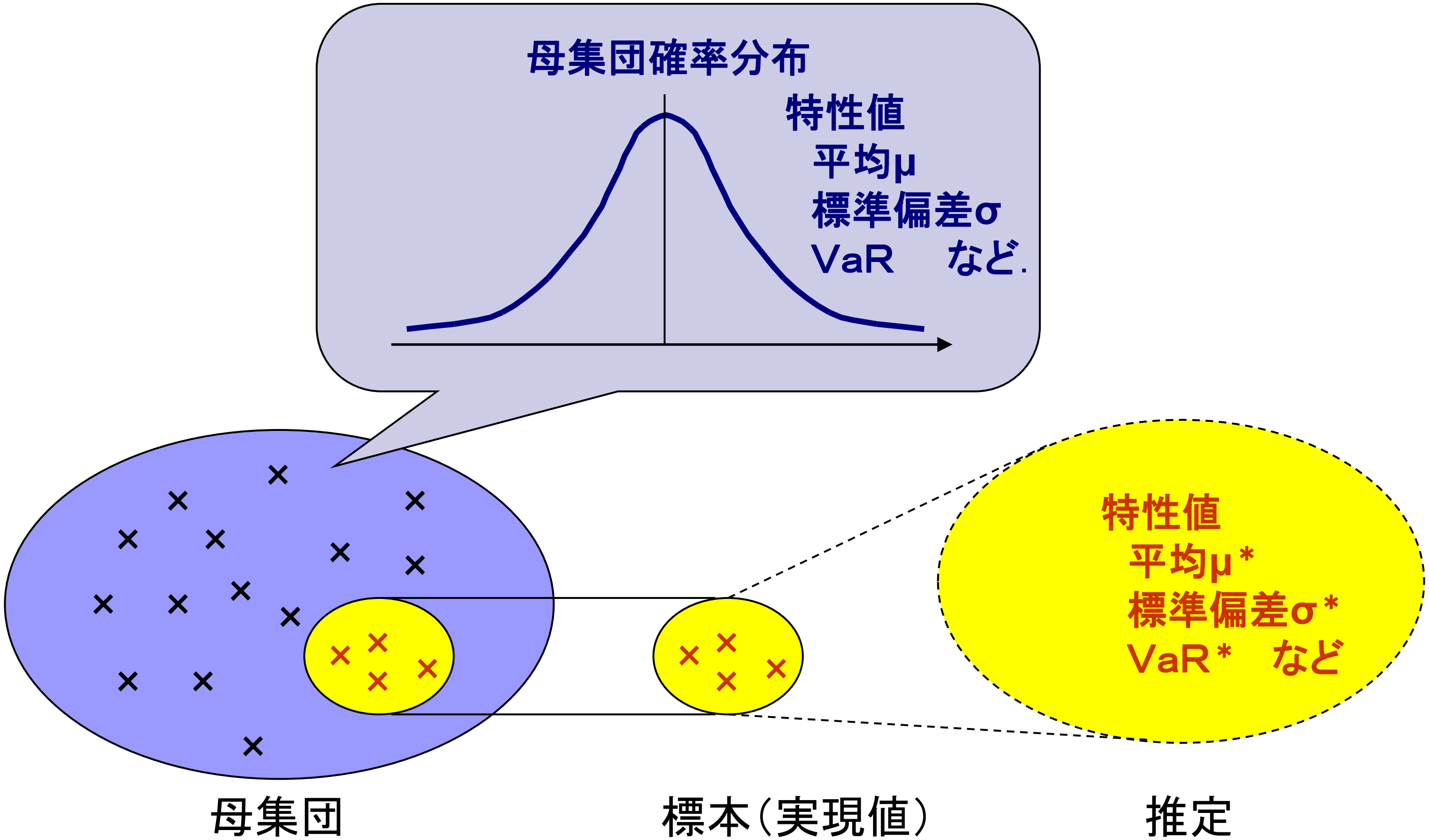
分散（不偏標本分散）

$$Va = \frac{(X_1 - \overline{X})^2 + (X_2 - \overline{X})^2 + \cdots + (X_N - \overline{X})^2}{N-1}$$

（注）上記定義（偏差平方和をN－１で割る）による標本分散Vaについては、理論上、「その期待値が母集団の分散となる」ことが知られている。Vaは母集団の分散を偏りなく推定する統計量となるため、「不偏標本分散」と言う。

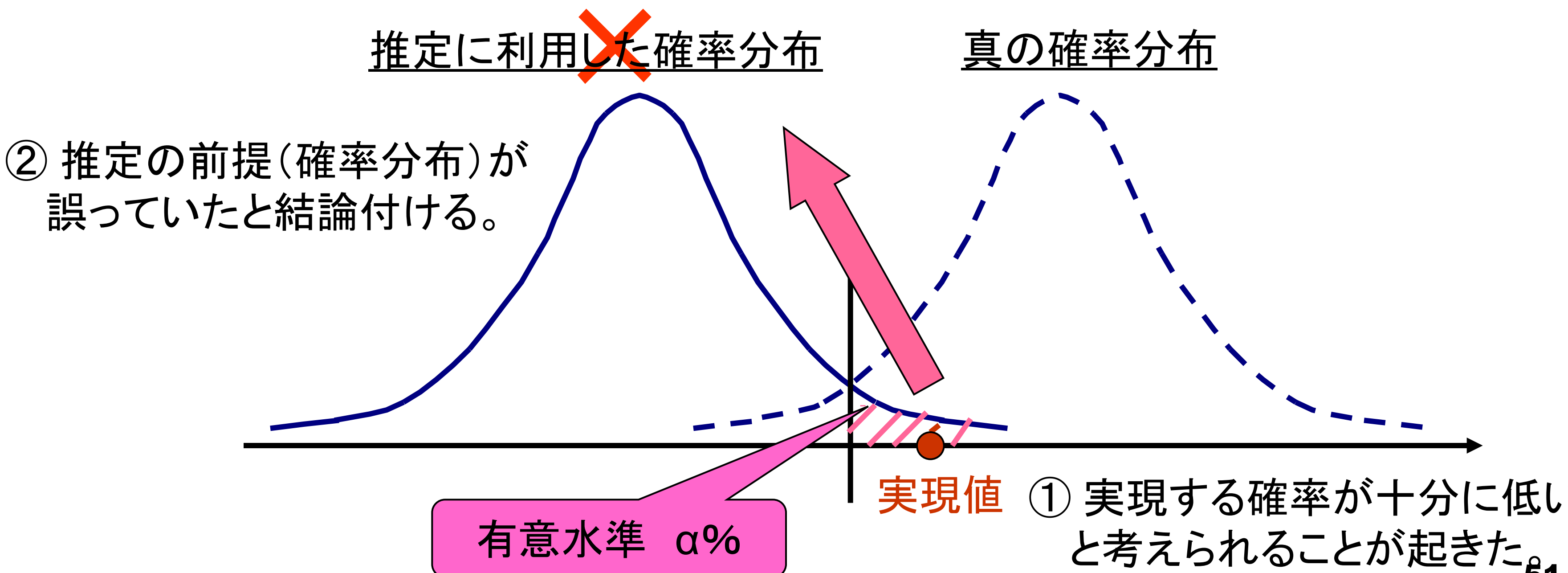
（2）推　定

- 母集団の確率分布、特性値は、誰にも分からない。
- 標本の特性値から母集団の特性値を統計的に推測する。



（3）検　定

- 一定の確率分布を前提にして推定した値について、
その値をとる確率（有意水準α　%）が十分に低いとき、
「偶然、珍しいことが起きた」と考えるのではなく、
「推定の際に置いた前提（帰無仮説） が誤っていた」
と結論付ける。



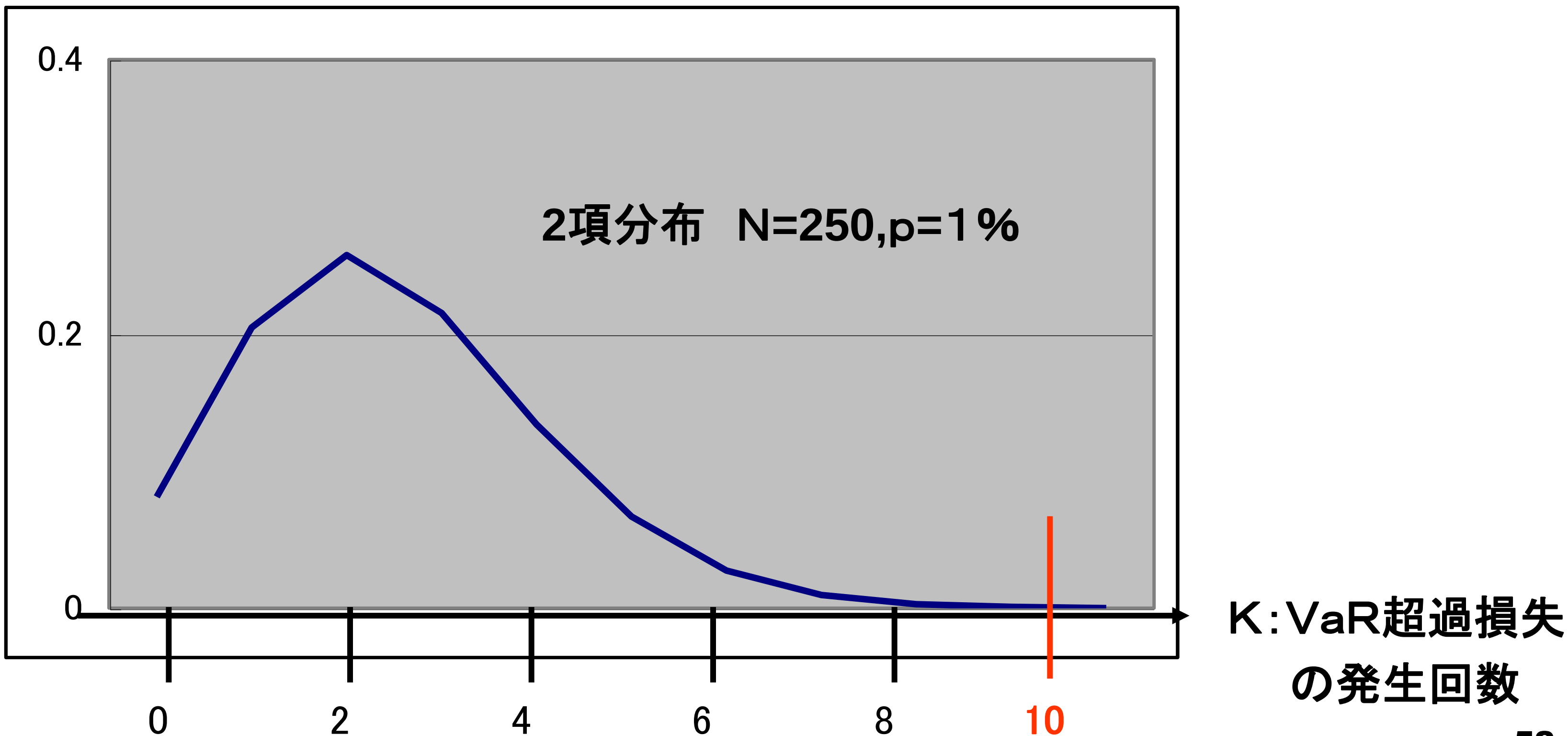
VaRを超過する損失が発生する回数(K)とその確率

VaRを超過する確率　　p　＝　1 %

VaRを超過しない確率　1－p ＝ 99%(信頼水準)

VaRの計測個数　　　N＝250

発生確率　$f(K) = {}_{250}C_K\,(0.01)^K\,(0.99)^{250-K}$



バックテスト（２項検定）

観測データ数	250	N回	N回の観測で、K回、VaRを超過する確率
信頼水準	99%		２項分布 ${}_{N}C_{K}\, p^{K}(1-p)^{N-K}$
1ー信頼水準	1%	p%	

VaR超過回数（K回）	確率	確率	VaR超過回数（K回以上）
0	8.11%	100.00%	0回以上
1	20.47%	91.89%	1回以上
2	25.74%	71.42%	2回以上
3	21.49%	45.68%	3回以上
4	13.41%	24.19%	4回以上
5	6.66%	10.78%	5回以上
6	2.75%	4.12%	6回以上
7	0.97%	1.37%	7回以上
8	0.30%	0.40%	8回以上
9	0.08%	0.11%	9回以上
10	0.02%	0.03%	10回以上
11	0.00%	0.01%	11回以上
12	0.00%	0.00%	12回以上
13	0.00%	0.00%	13回以上
14	0.00%	0.00%	14回以上
15	0.00%	0.00%	15回以上

バックテストは「検定」の考え方にしたがって行う

- VaR計測モデルは正しい(帰無仮説)。

↓

- VaR超過損失の発生が、250回中、10回以上発生した。

↓

- VaR超過損失の発生が、250回中、10回以上発生する確率は0.03%と極めて低い。

↓

- VaR計測モデルは誤っている(結論)

２種類の過誤

- 「検定」では、次の2通りの「過誤」（エラー）が起きる可能性がある。したがって、バックテストの結果も「過誤」（エラー）を伴っている可能性がある点、注意を要する。

第１種の過誤（エラー）

本当は帰無仮説（VaR計測モデル）が正しいのに、

検定の結果、

帰無仮説（VaR計測モデル）が誤っていると結論付けてしまう。

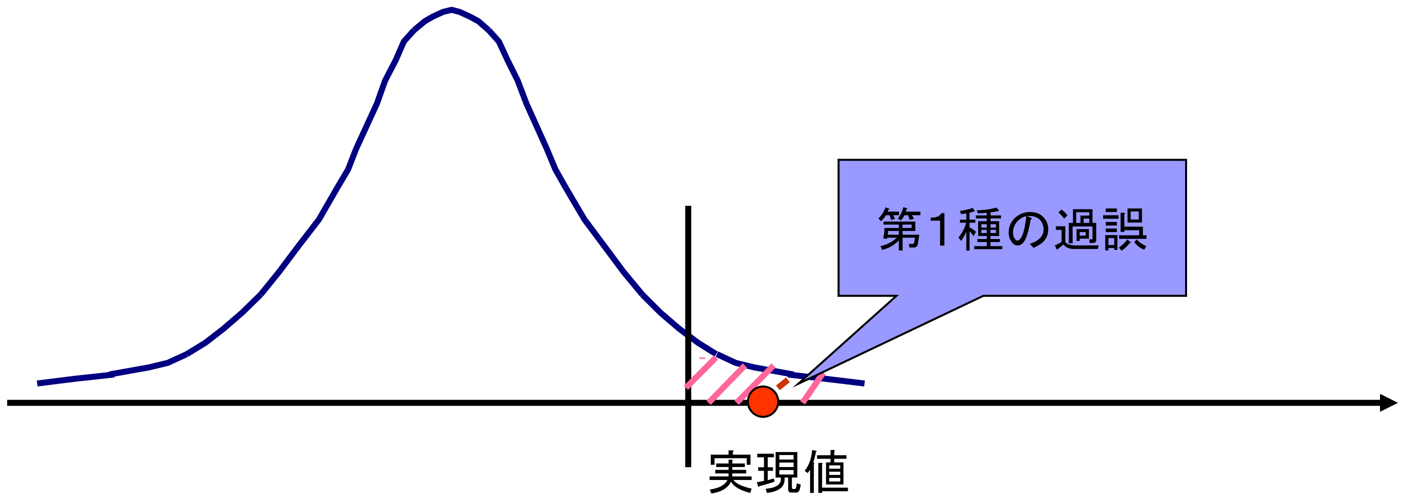
第２種の過誤（エラー）

本当は帰無仮説（VaR計測モデル）が正しくないのに、

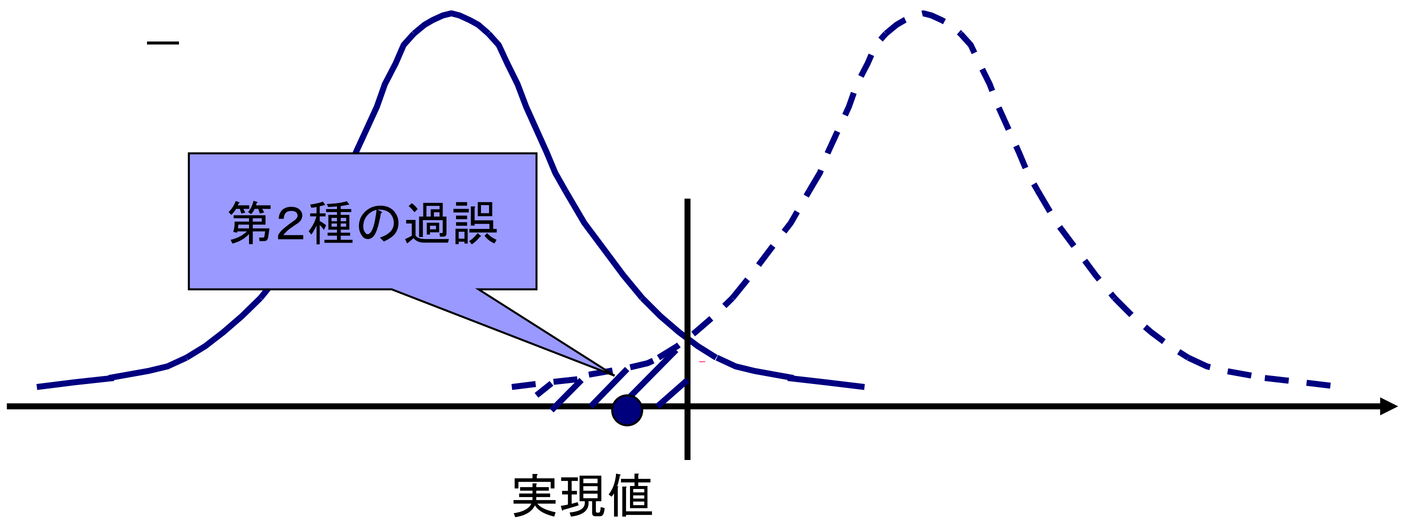
検定の結果、

帰無仮説（VaR計測モデル）が正しいと結論付けてしまう。

推定に利用した確率分布 ＝ 真の確率分布



推定に利用した確率分布 ≠ 真の確率分布



５．線形回帰分析

（１）線形回帰分析とは

（２）Excel分析ツールの利用

（３）チェック項目（決定係数、Ｐ値）

（1）線形回帰分析とは

- X_i と Y_i の間に「直線的な比例関係」があることを前提にして、X_i と Y_i の散布図の中の各点のなるべく近くに直線を描く。

$$Y_i = aX_i + b + e_i$$

変数 Y を変数 X で説明する。

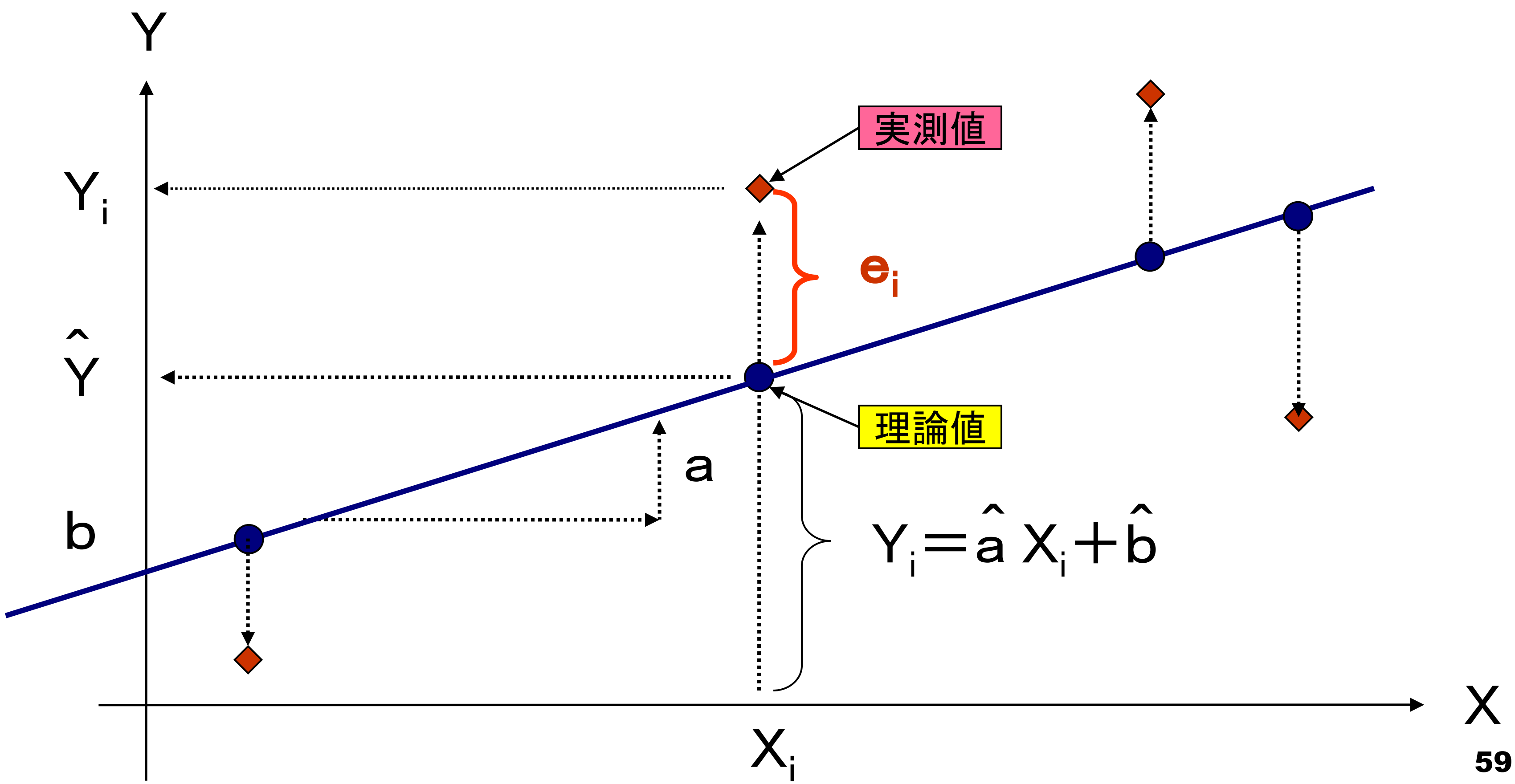
Y_i ： 被説明変数（目的変数）
X_i ： 説明変数
a ： 回帰係数
b ： 定数項（切片）
e_i ： 残差

（注）本例のように、説明変数が1つの場合、単回帰分析という。説明変数が2つ以上の場合、重回帰分析という。

最小2乗法

- 残差　e_i ＝　$Y_i - aX_i - b$　の2乗和を最小にするように

　a 、bを推定する。それぞれの推定値を $\hat{a}$ 、$\hat{b}$ と表記する。



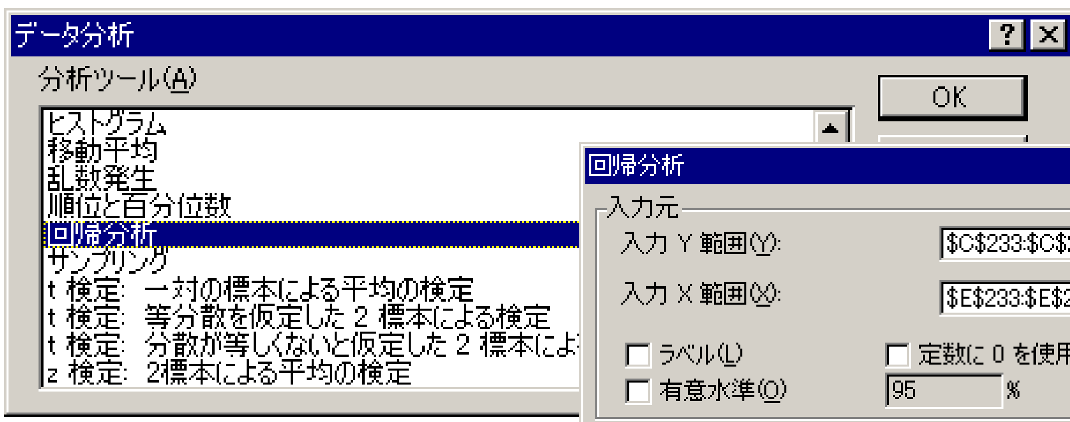
（2）Excel分析ツールを利用した回帰分析

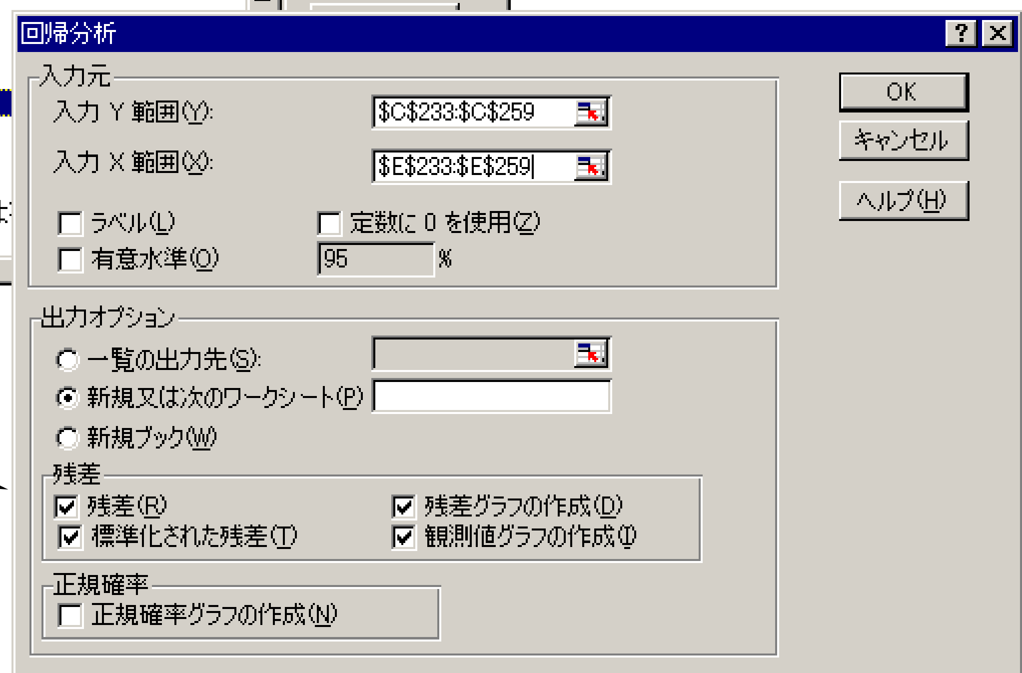
【手順】

①「ツール」メニューから「分析ツール」を起動。

②ボックスの中の「回帰分析」を選択してOKをクリック。

③「入力Y範囲」、「入力X範囲」に、それぞれデータ範囲を入力。





チェックを入れると観測値、残差のグラフ等をを表示

（注）PCによっては、分析ツールのアドインが必要です。

(例)Excel分析ツール・回帰分析の出力結果

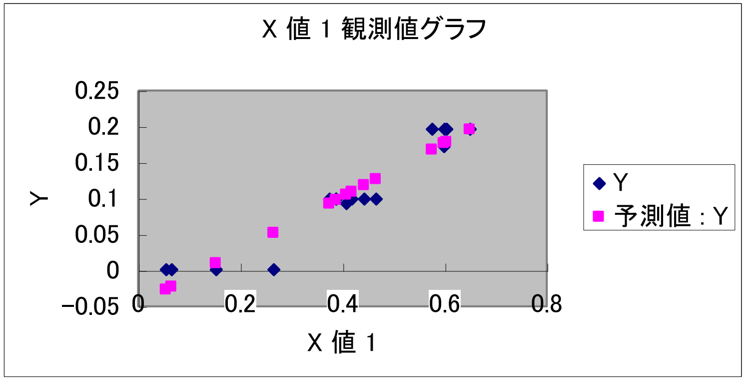
概要

回帰統計	
重相関 R	0.956320779
重決定 R2	0.914549432
補正 R2	0.90844582
標準誤差	0.022258115
観測数	16

分散分析表

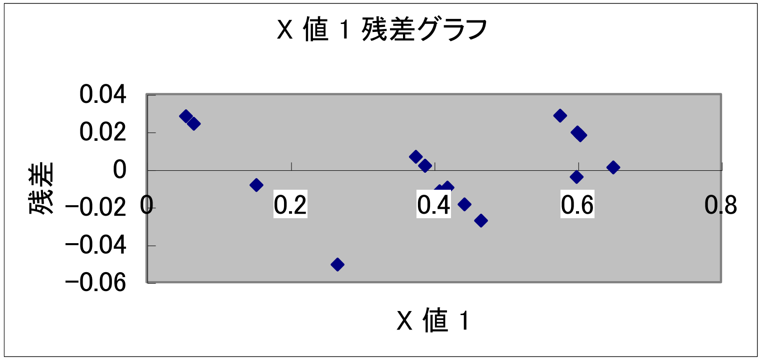
	自由度	変動	分散	観測された分散比	有意 F
回帰	1	0.074233006	0.074233006	149.8374126	7.24E-09
残差	14	0.006935932	0.000495424		
合計	15	0.081168938			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	-0.047846512	0.013516678	-3.539813066	0.003266347	-0.07684	-0.018856096	-0.076836928	-0.018856096
X 値 1	0.37369024	0.03052823	12.24080931	7.24475E-09	0.308214	0.439166839	0.308213641	0.439166839



残差出力

観測値	予測値 : Y	残差	標準残差
1	-0.027293549	0.028293549	1.315772009
2	-0.023182956	0.024182956	1.124611728
3	0.009328095	-0.008328095	-0.387292319
4	0.051555092	-0.050555092	-2.351029759
5	0.104619106	-0.011619106	-0.540338532
6	0.092287328	0.006712672	0.312168184
7	0.097145301	0.001854699	0.086251488
8	0.097145301	0.001854699	0.086251488
9	0.108729699	-0.009729699	-0.452472943
10	0.117698264	-0.018698264	-0.869549921
11	0.12629314	-0.02729314	-1.269248692
12	0.175993942	-0.003993942	-0.185735522
13	0.177862393	0.018137607	0.843476924
14	0.167399066	0.028600934	1.330066732
15	0.176367632	0.019632368	0.912989753
16	0.195052144	0.000947856	0.044079382



回帰分析を行うときのチェック項目（必要最低限度）

概要

回帰統計	
重相関 R	0.956320779
重決定 R2	0.914549432
補正 R2	0.90844582
標準誤差	0.022258115
観測数	16

分散分析表

	自由度	変動	分散	観測された分散比	有意 F
回帰	1	0.074233006	0.074233006	149.8374126	7.24E-09
残差	14	0.006935932	0.000495424		
合計	15	0.081168938			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	-0.047846512	0.013516678	-3.539813066	0.003266347	-0.07684	-0.018856096	-0.076836928	-0.018856096
X 値 1	0.37369024	0.03052823	12.24080931	7.24475E-09	0.308214	0.439166839	0.308213641	0.439166839

定数項（切片）（bの推定値）

回帰係数（aの推定値）

決定係数（R^2）：モデルの当てはまりの良さを示す指標（1に近いほど良い）

– Yの偏差平方和（全変動）に占める、$\hat{a}X+\hat{b}$の偏差平方和（モデルで説明できる変動）の割合として定義される（重回帰分析の場合は、自由度補正後の補正R^2をみる）

P－値　　：回帰係数、定数項の有意性を示す指標（ゼロに近いほど良い）

– 回帰係数、定数項がゼロであると仮定した（帰無仮説）ときに、それぞれの推定値が実現する確率。ゼロに近ければ、検定の考え方にしたがって、帰無仮説を棄却できる。回帰係数、定数項はゼロではない　→　回帰係数、定数項は Yを説明するのに有効。

- 本資料に関する照会先

  日本銀行金融機構局金融高度化センター

  企画役　碓井 茂樹

  Tel　03(3277)1886　E-mail shigeki.usui@boj.or.jp

- 本資料の内容について、商用目的での転載・複製を行う場合は予め日本銀行金融機構局金融高度化センターまでご相談ください。転載・複製を行う場合は、出所を明記してください。
- 本資料に掲載されている情報の正確性については万全を期しておりますが、日本銀行は、利用者が本資料の情報を用いて行う一切の行為について、何ら責任を負うものではありません。