イノベーションマネジメント研究科
技術経営専攻

コーポレートファイナンス

第６講：証券投資に関する理論（１）

東京工業大学イノベーションマネジメント研究科
中川　秀敏
E-mail: nakagawa@mot.titech.ac.jp
Office Hour : 13:00-14:30, every Tuesday, at W9-105

※配布資料がある場合、講義前日の午後5時までにはアップするので事前に
http://www.craft.titech.ac.jp/~nakagawa/dir2/lecture.html#TIT2006_2
から各自でダウンロードして用意すること。

今回の目標

- リスクとは？
- 平均、標準偏差
- 共分散、相関係数

Uncertainty と Risk

- Uncertainty: ある対象の不確定要素について「法則」すら分かっていない不確実性
  - 確率分布を特定することが不可能な状況
- Risk: 不確定要素について「法則」は分かる、または仮定できる不確実性（≠危険）
  - 確率分布を特定したうえで議論可能
- 金融における不確定要因は一般に「リスク」と称するので、以後は上の区別と関係なく、定量化する（あるいはされた）不確実性を「リスク」と呼ぶことにする。

（参考）Uncertainty と Risk

- 危険(risk)
  - 経済主体が起こす行動（意思決定）のおのおのに対応して特定の既知の結果が生じるような確実性下の世界に反して、どのような結果が生じるかが既知でないとき、その世界には危険（リスク）があるという。狭義には、生じうる可能な結果のおのおのにその確率が付与され、意思決定主体がその確率を既知としている場合を危険と呼び、確率を既知としない不確実性とは区別される
- 不確実性(uncertainty)
  - 広義には、危険と同じ意味で用いられるが、狭義には、意思決定主体の行動のそれぞれに対応して、特定の結果が生じる確率が主体には不明であるか、あるいは、そのような確率を付与すること自体が無意味であるような場合をいう。

（金森久雄他編「経済辞典」第4版, 有斐閣より）

「リスク」の表現

① 「将来どのくらい損を被るか」という「想定すべき損害の程度」という観点
  - 例えば、1年後に資産価値がどのくらい目減りする可能性があるか？

② 「将来資産価値がおよそどのくらいの値の範囲に収まるか」という「不確実性の度合い」という観点
  - 例えば、1年後の資産価値が、およそ何円～何円の範囲に収まるか？

③ 対象に影響を与える要素が現段階で1単位変動したときに、その対象の価値がどれだけ変化するか
  - 例えば、あるマクロ変数が1単位減少すると、その資産価値はどのくらい増加 or 減少するか？

リターンの計算

- 通常のリターンの定義

$$\text{日次収益率} = \frac{\text{終値}}{\text{前日の終値}} - 1$$

- リターンを計算する際に価格比の対数を取る場合もある

$$\text{日次対数収益率} = \log \frac{\text{終値}}{\text{前日の終値}}$$

※N日分の日次対数リターンの和は、N日対数リターンに一致するが、ポートフォリオの対数リターンは各資産の対数リターンの加重和に一致しない。

- 個別の株式の場合は配当権利落ちや株式分割等の影響を考慮して、修正した株価を利用するのが一般的

復習：統計学の基礎

- 平均・分散（標準偏差）・中央値
- 度数分布・相対度数分布・ヒストグラム

統計学の基礎

- 統計学は観測データ（確率変数の標本）があることを前提とした話
- 以下、N日分のTOPIX日次リターンの観測データ

$$x_1, x_2, \cdots, x_N$$

が与えられているとして話を進める。

平均・分散（標準偏差）・中央値

- 平均とは次の計算式で与えられるもの

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

※EXCEL関数
AVERAGE

- 平均はデータ集合の特性を表す有効な量であるが、分布の形状が対称型でないときは注意が必要。

平均・分散（標準偏差）・中央値

- 詳しいことをいうと先の平均は「標本平均」と呼ばれるもの
- データの元になる確率変数の真の平均は「母平均」と呼ばれる
- 本当に知りたいのは「母平均」であり、「標本平均」はあくまでも「母平均」の推定値である。

平均・分散（標準偏差）・中央値

- 分散とは次の計算式で与えられるもの・・・

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

※EXCEL関数
VAR

- 各データ値と平均とのずれを2乗したものをデータ集合全体で平均化したもの
- 標準偏差は分散の正の平方 $\hat{\sigma}$

※EXCEL関数
STDEV

平均・分散（標準偏差）・中央値

- 詳しいことをいうと先の分散は「標本分散」と呼ばれるもの
  （さらに詳しく言うと「不偏標本分散」というもの）
- データの元になる確率変数の真の分散は「母分散」と呼ばれる
- 本当に知りたいのは「母分散」であり、「標本分散」はあくまでも「母分散」の推定値である。

平均・分散（標準偏差）・中央値

- 中央値とは文字通り、データを昇順に並べたときの中央に位置する値　※EXCEL関数　MEDIAN
- 中央値は分布の形状によらずに、データ集合の特性を表す量と考えられる。
- 他にも、第1四分意点（25%点）、第3四分意点（75%点）や、あるいは一般的に「〇〇%点」のようなものを考えることもある。

※EXCEL関数
QUARTILE，PERCENTILE

平均・分散（標準偏差）・中央値

- （平均）絶対偏差

※EXCEL関数 AVEDEV

$$\frac{1}{N}\sum_{i=1}^{N}|x_i-\bar{x}|$$

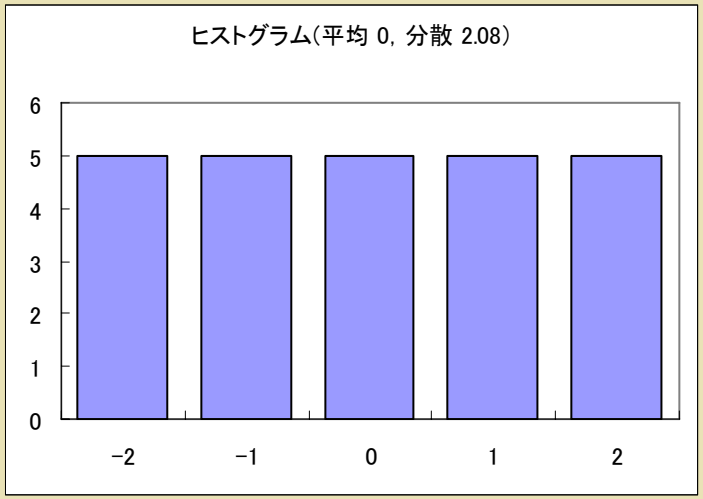
- 元々、各データ値と平均の差の絶対値を「偏差」と呼ぶ。従って、上の絶対偏差がデータ集合のばらつき度合いを表すものとして直感にあう。
- 理論的な計算の都合や正規分布との相性の良さもあり、標準偏差の方が研究されてきたという経緯もある。

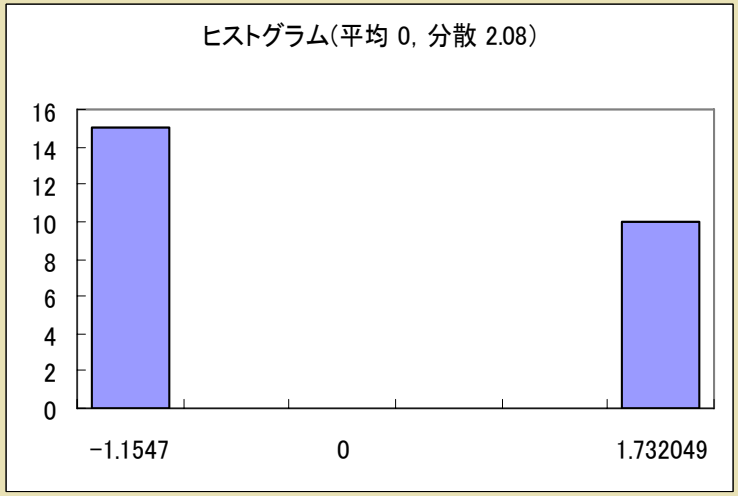
度数分布・ヒストグラム

- 平均や分散だけでは、データ集合の様子を正しくとらえられない場合もある
- データ区間をいくつかの階級に分けて、各階級に何個のデータが入っているかをまとめて表すとデータの散らばり具合がわかりやすい
  - 表で表せば「度数分布表」
  - グラフで表せば「ヒストグラム」
- 各階級のデータ数のデータ全体に対する割合を「相対度数」という。データ個数の異なる集合どうしで散らばり具合を比較するときは、「相対度数」を使う方が適している

ヒストグラム

☆ 同じ平均・分散をもつが、散らばり具合の全く異なる例





☆スタージェスの公式：データ個数とそれに対応する適切な階級数の関係を表したとされる式（実際には良い式とは思えない）

$$\text{階級数} = 1 + \log_2(\text{データ数}) \cong 1 + 3.3\log(\text{データ数})$$

※実際には個別の検討が必要に思う

統計学の基礎

- 推定と検定のお話

（参考：東北大学統計グループ「これだけは知っておこう！統計学」
有斐閣ブックス）

推定のお話

- あるデータ集合を何かしらの確率変数Xの標本値の集合と見なし、その確率分布を考えることは自然な考え方
- かしこまった言い方をすると、上のような考え方は観測している対象の「モデル」を与えることを意味する

推定のお話

- ただし、確率分布を最初から完全に特定することはできず、一般にはある未知数を含んだ形で「〇〇分布」ということだけ仮定する。
  - 例えば、あるデータ集合が「正規分布」に従うと仮定できても、「平均」と「標準偏差」を与えないと完全に特定されたことにならず、一般にこの2つを最初は未知数と考える
  - このような未知数のことを「パラメータ」や「母数」と呼ぶ
- 最終的にはパラメータを与える必要がある

推定のお話

- 推定とは、データ集合から適当な方法を用いて、モデルとなる確率分布のパラメータの値を決めること
- 推定には「誤差」がつきものであることに注意
  - パラメータの推定値≠パラメータの真の値
  - 区間推定という考え方
- 推定する方法はいくつかあり、問題に応じて使い分けられたりする。
  - 最尤法、最小2乗法、モーメント法など

推定のお話

- 例：視聴率調査
  - ある番組の視聴率を調査したい。視聴率 p がここでのパラメータとなる。
  - 400世帯をサンプルとして、28世帯がその番組を視聴したことがわかった。
  - p の最も自然な推定値は

$$\hat{p} = \frac{28}{400} = 0.07 = 7\%$$

推定のお話

- 例：視聴率調査（つづき）
  - 一方で日本全体には数千万（？）の世帯があり、7%という視聴率がどの程度確からしいか問題になる。
  - 実は真の視聴率は95%の確率で、4.5%～9.5% の区間に含まれていることが計算で求められる。
    （区間推定の結果は、このように確率（信頼水準とよばれる）の設定の仕方で変わる）

推定のお話

- 信頼水準95%における平均の信頼区間

$$\bar{x} - 1.96\frac{s}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{s}{\sqrt{n}}$$

標準誤差

μ: 母平均, n : データ数（十分多い）

$\bar{x}$: 標本平均, s : 標本標準偏差

※ $1.96 = \mathrm{NORMSINV}(0.975)$

※厳密には標準正規分布のところを、自由度 n-1 のt分布に基づいて考えるべきだが、ここでは詳細は割愛

推定のお話

- 先の視聴率調査の場合

$$0.07 - 1.96\frac{\sqrt{0.07(1-0.07)}}{\sqrt{400}} < p < 0.07 + 1.96\frac{\sqrt{0.07(1-0.07)}}{\sqrt{400}}$$

$$\Downarrow$$

$$0.045 < p < 0.095$$

検定のお話

- 一方、観測対象について、何らかの仮説を立てそれをデータを用いて肯定的に主張したいと考えることも一般的である
  - 株価リターンは正規分布に従う
  - 自己資本比率が低い企業は倒産しやすい、などなど
- しかし、データを用いて、このような仮説が100%成り立つと断定することは難しい。
- そこで何らかのルールに従って、仮説が妥当かどうかを調べる方法論が必要になる

検定のお話

- 検定とは、仮説が正しいかどうかをデータを用いて判定すること
  - 自分が主張する仮説を否定する「帰無仮説」を考える。自分が主張したい仮説は「対立仮説」とする。
  - 帰無仮説が正しいという仮定のもとで計算された統計量の値について、そのような値が観測される確率が非常に小さいとき（実際は有意水準と呼ばれる確率の基準を与えて、それと比較する）、帰無仮説は「棄却」されると言う。

検定のお話

- 「帰無仮説」はそもそも棄却したい仮説であり、めでたく「帰無仮説」が検定によって棄却されると、その「対立仮説」すなわち自分が主張したかった仮説が採択されるという理屈になる。
  - その意味で検定の問題の設定は難しい
  - ２種類の誤りに注意
    - 第１種の過誤：「帰無仮説が正しい」にもかかわらず、「正しくない」としてしまう誤り
    - 第２種の過誤：「帰無仮説が正しくない」にもかかわらず、「正しい」としてしまう誤り

検定のお話

- 例：再び視聴率調査
  - ある番組の視聴者年齢層を調査した。
  - 20代と50代それぞれ400人について調査したところ、20代でその番組を見たのが40人で、50代では28人であった。
  - それぞれの年齢層での視聴率を単純に推定すると、20代が10%で、50代が7%となる。
  - この結果だけで、この番組は50代よりも20代に人気がある（より多く視聴されている）と言えるだろうか？

検定のお話

- 例：再び視聴率調査（つづき）
  - そこで次のような統計的検定を行う
    - 帰無仮説：20代と50代での当該番組の視聴率は同じ
    - 対立仮説：20代の方が50代よりも当該番組の視聴率は高い
  - 検定のための統計量を計算してみると、帰無仮説のもとでそのような統計量が観測される確率（p値と呼ばれる）は6.4%となり、5%の有意水準では帰無仮説は棄却されない→この番組が20代の方に人気が高いことは言えない
  - もし、50代の視聴者が24人（視聴率の推定値6%）だと、p値は1.9%となり、帰無仮説は棄却される。

検定のお話

- ２群の比率の片側検定（例：視聴率調査）
  - 仮説の設定

p_{20}：20代の視聴率，p_{50}：50代の視聴率

$H_0 : p_{20} = p_{50}$（帰無仮説）

$H_1 : p_{20} > p_{50}$（対立仮説）

  - 検定統計量

$$t = \frac{p_{20} - p_{50}}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_{20}} + \frac{1}{n_{50}}\right)}}$$

$\bar{p}$：20代と50代合わせた視聴率

n_{20}：20代サンプル数

n_{50}：50代サンプル数

検定のお話

- ２群の比率の片側検定（例：視聴率調査）
  - p値の算出

$$p\text{値} = 1 - NORMSDIST(\text{検定統計量の絶対値})$$

  - 例の場合

$$t = \frac{0.10 - 0.07}{\sqrt{0.085(1-0.085)\left(\frac{1}{400}+\frac{1}{400}\right)}} = 1.521$$

$$p\text{値} = 1 - NORMSDIST(1.521) = 0.064 > 0.05$$

帰無仮説を棄却できない

復習:統計学の基礎

- 共分散・相関係数
- 散布図

共分散・相関係数

- 対応関係のあるデータ集合どうしの関係を分析したくなることがある
  - 複数の株価の連動性
  - 為替レートと物価の連動性
  - 売上高の変化率と利益の変化率の関係
- 以下のような対応する2つのデータ集合の関係性を1つの数値で表すことを考える

$$(x_1, x_2, \cdots, x_N),$$
$$(y_1, y_2, \cdots, y_N)$$

共分散・相関係数

- 共分散とは次の計算式で与えられるもの

$$Cov(X,Y)=\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})$$

※EXCEL関数：COVARは *N-1* ではなく、*N* で割ったものを計算

- X,Y それぞれのデータ値が平均からどれくらいずれているかをお互いのずれの方向性の一致具合も符号でとらえて、全体で平均化したもの

共分散・相関係数

- 共分散は2つのグループのデータ値の水準（単位）に影響される
- データの単位に影響されないように、共分散をそれぞれの標準偏差で割った料を相関係数と呼ぶ。

$$\rho(X,Y) = \frac{Cov(X,Y)}{\hat{\sigma}_X \hat{\sigma}_Y}$$

※EXCEL関数
CORREL

$\hat{\sigma}_X, \hat{\sigma}_Y : X, Y$ それぞれの標準偏差

共分散・相関係数

- 相関係数は次の関係を満たす

$$-1 \le \rho(X,Y) \le 1$$

$$Y = aX + b$$

$$\Rightarrow \rho(X,Y) = 1(a>0),\ -1(a<0), 0(a=0)$$

- 「相関係数が0に近い」ということは2つのグループが線形関係にはないということを示している
- 「全く関係がない」ということを意味しているのではないことに注意

共分散行列

- 3つ以上の変数の共分散は次のように行列を用いて表す

$X_1, \cdots, X_N$ というN個の確率変数があるとする。

$\sigma_i^2 := Var(X_i),$

$\sigma_{i,j} := Cov(X_i, X_j)$

$$\Sigma := \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \cdots & \sigma_{1,N} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \cdots & \sigma_{2,N} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \cdots & \sigma_{3,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,N} & \sigma_{2,N} & \sigma_{3,N} & \cdots & \sigma_N^2 \end{pmatrix}$$

相関行列

- 同様に3つ以上の変数の相関は次のように行列を用いて表す

$X_1, \cdots, X_N$ というN個の確率変数があるとする。

$$\rho_{i,j} := \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)}\sqrt{Var(X_j)}}$$

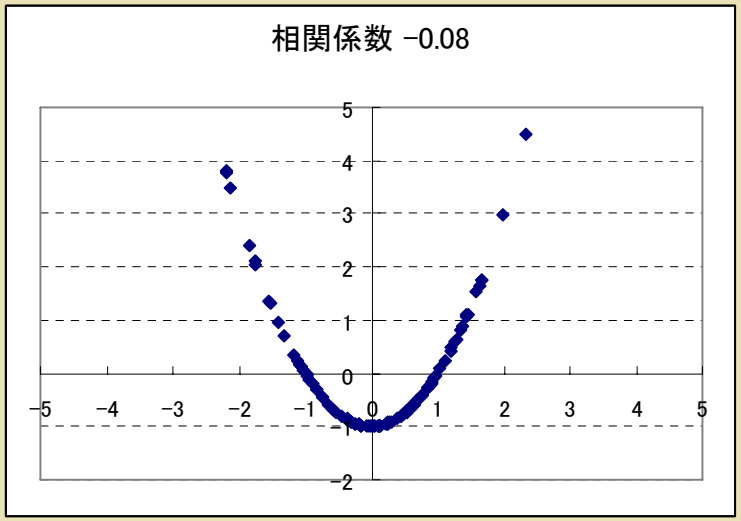
$$\Phi := \begin{pmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \cdots & \rho_{1,N} \\ \rho_{1,2} & 1 & \rho_{2,3} & \cdots & \rho_{2,N} \\ \rho_{1,3} & \rho_{2,3} & 1 & \cdots & \rho_{3,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{1,N} & \rho_{2,N} & \rho_{3,N} & \cdots & 1 \end{pmatrix}$$

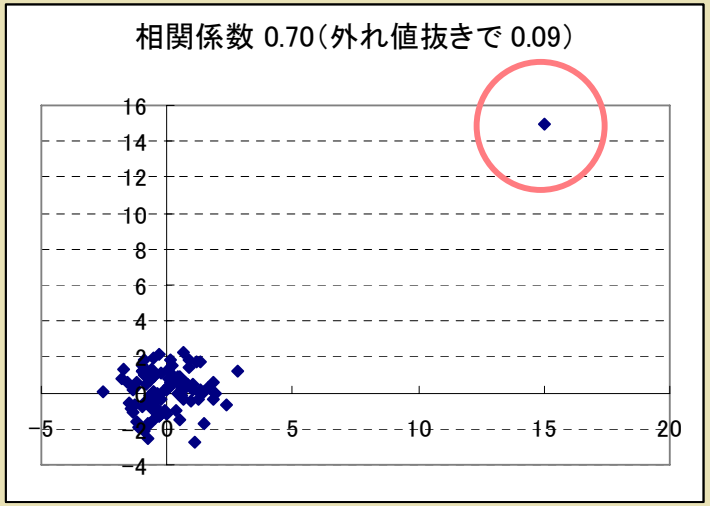
散布図

- 共分散や相関係数だけでは2つのデータ集合間の関係を正しくとらえられない場合もある
- 散布図を描いて、両者の関係を視覚的にとらえることは有効

散布図

☆ 相関係数と「依存関係」についての注意





相関は0に近いが、両者の間には明確な関係が認められる

ほとんどのデータは固まっていて、相関もほとんどないが、大きく外れた1点の影響が全体に及んでいる

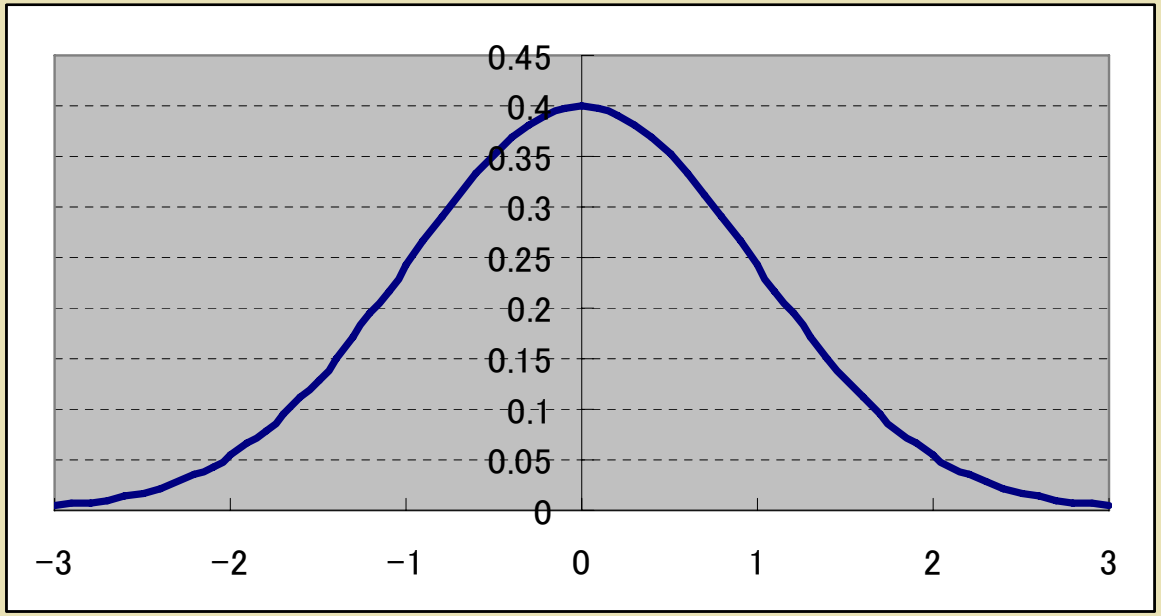
知っておきたい確率分布～正規分布

- 正規分布(Normal Distribution)
  - 連続値をとる確率変数の確率分布の中で特に重要な働きをする
    - 収益率（リターン）の分布として良く用いられる
    - 中心極限定理（同じ分布に従う確率変数の独立な観測値の和の分布は観測数を増加させると正規分布に近づいていく）
  - 平均と分散（＝標準偏差の２乗）の２つのパラメータで決定される

    （平均、分散、標準偏差の統計的意味については次回）

知っておきたい確率分布～正規分布

- 標準正規分布（平均0,分散1）の密度関数グラフ



- 平均（正規分布では中央値に一致する）のところが最も密度が高い
- 平均に関して左右対称

知っておきたい確率分布～正規分布

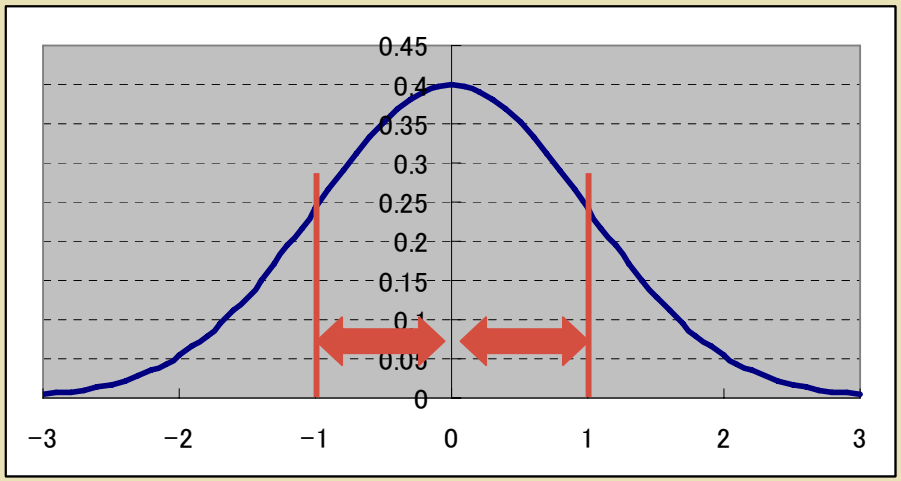
- Xが標準正規分布（平均0,分散1）に従うとき

$$P(X \geq 0) = P(X \leq 0) = 0.5$$

$$P(-1 \leq X \leq 1) \cong 0.6827$$

$$P(-1.96 \leq X \leq 1.96) \cong 0.95$$

$$P(X \leq 2.33) \cong 0.99$$



知っておきたい確率分布～正規分布

標準正規分布関数に対しては
NORMSDIST(c) がある

- 正規分布の計算（Excel）
  - NORMDIST(c, μ , σ ,TRUE or FALSE)
  - ４番目の引数を TRUE にすると「分布関数」

$$F(c) = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

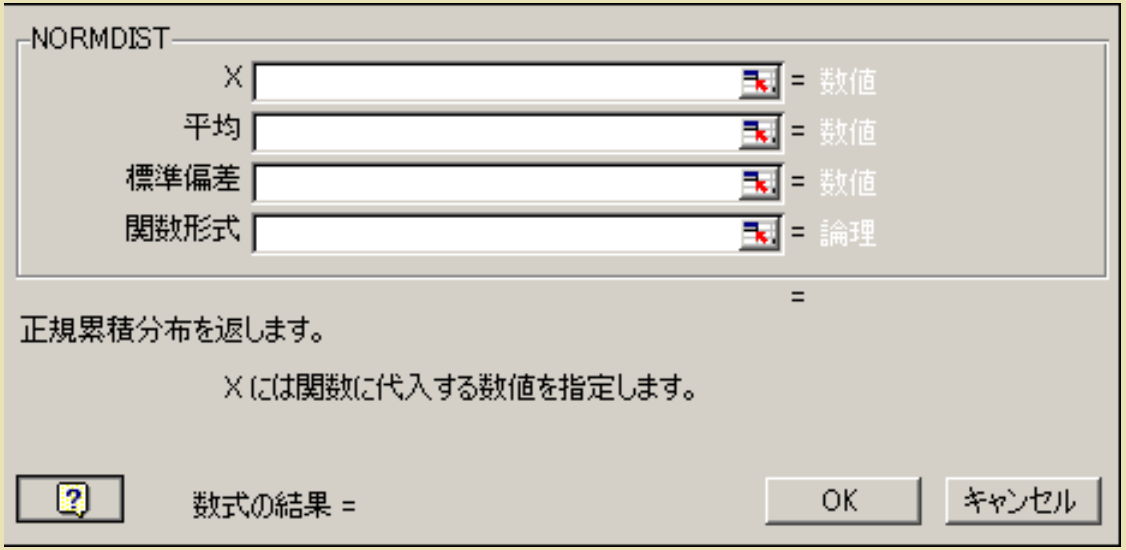
  - ４番目の引数を FALSE にすると「密度関数」

$$f(c) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(c-\mu)^2}{2\sigma^2}}$$

NORMDIST(0, 0, 1,TRUE) ＝？, NORMDIST(0, 0, 1, FALSE) ＝？

知っておきたい確率分布～正規分布

– 関数ウィザードで NORMDIST 関数を利用



知っておきたい確率分布～正規分布

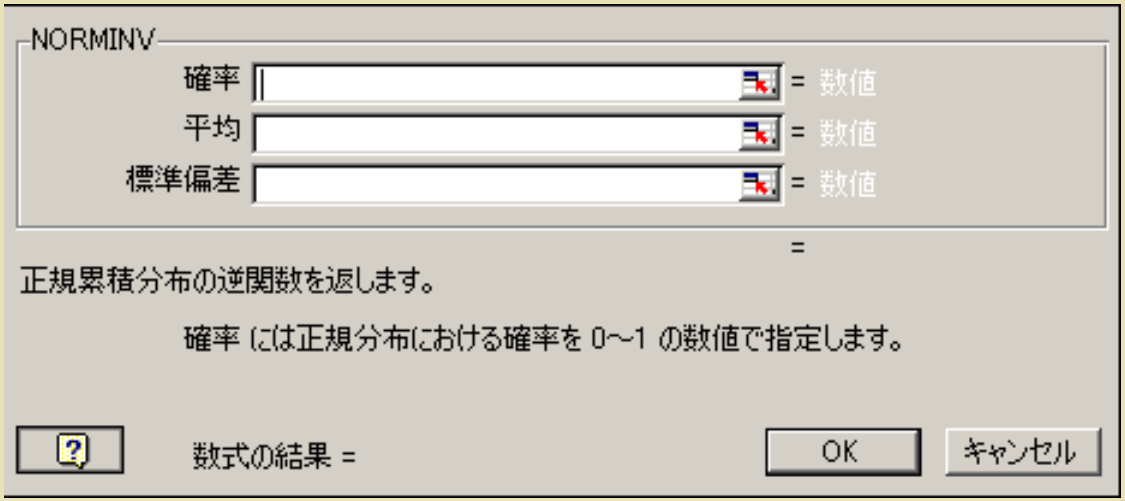
標準正規分布関数の逆関数
に対しては
NORMSINV(p) がある

- 正規分布の逆関数
  - NORMINV(p, μ , σ)
  - 確率 p が与えられたときに下の関係式を満たす c の値を求める

？

$$p = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

NORMINV(0.95, 0,1) = ？

知っておきたい確率分布～正規分布

– 関数ウィザードで NORMINV関数を利用