３．株式投資の
リスクとリターン

経済統計分析
（2015年度春学期）

株式投資のリスクとリターン

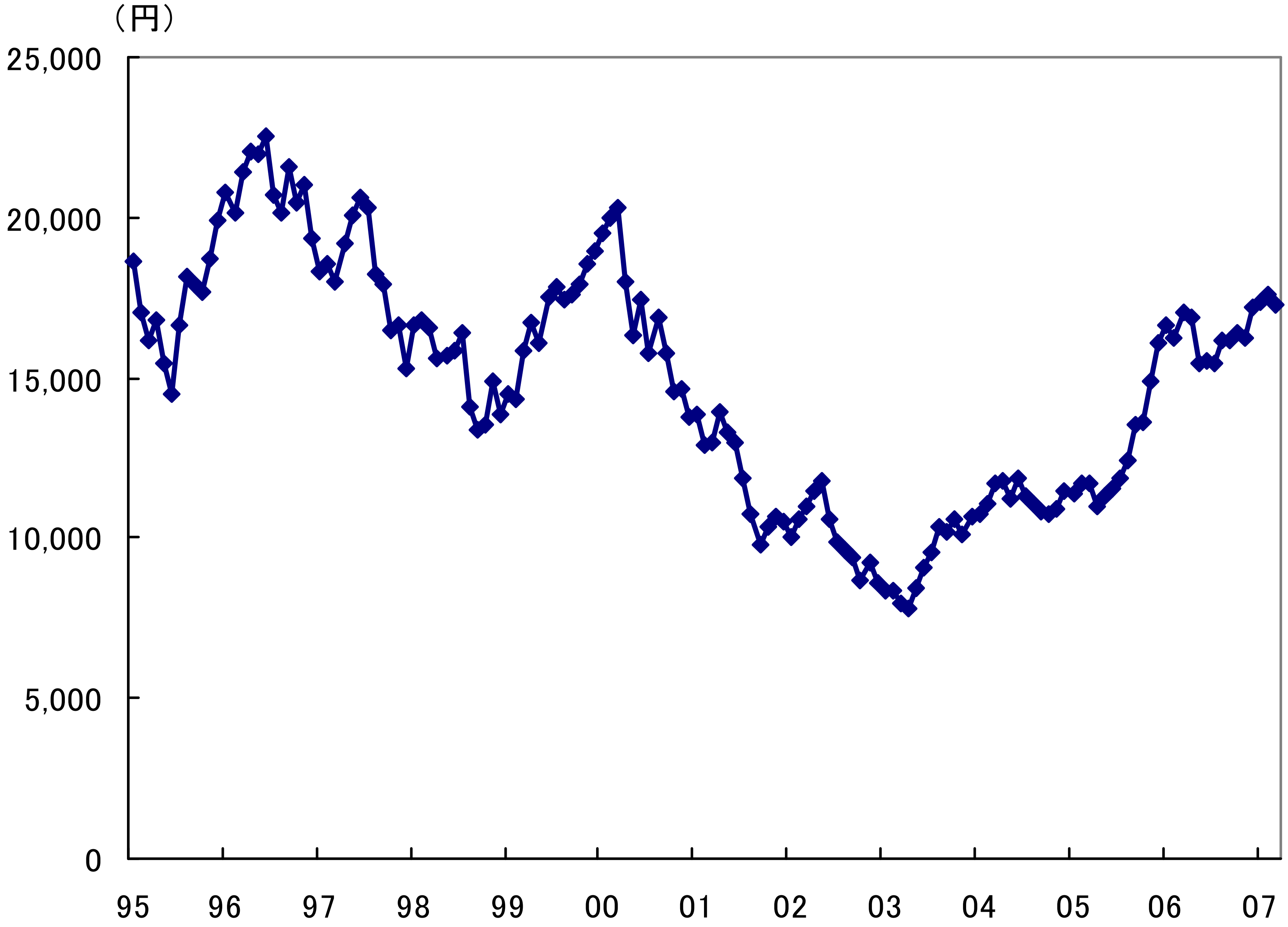
（統計分析手法）

- 成長率（株価上昇率）、指数
- 平均、分散、標準偏差
- 相関係数
- 分布と確率の計算
- 信頼区間の推定（点推定と区間推定）
- 仮説検定（平均値の検定、平均差の検定）

（経済理論等との関連）

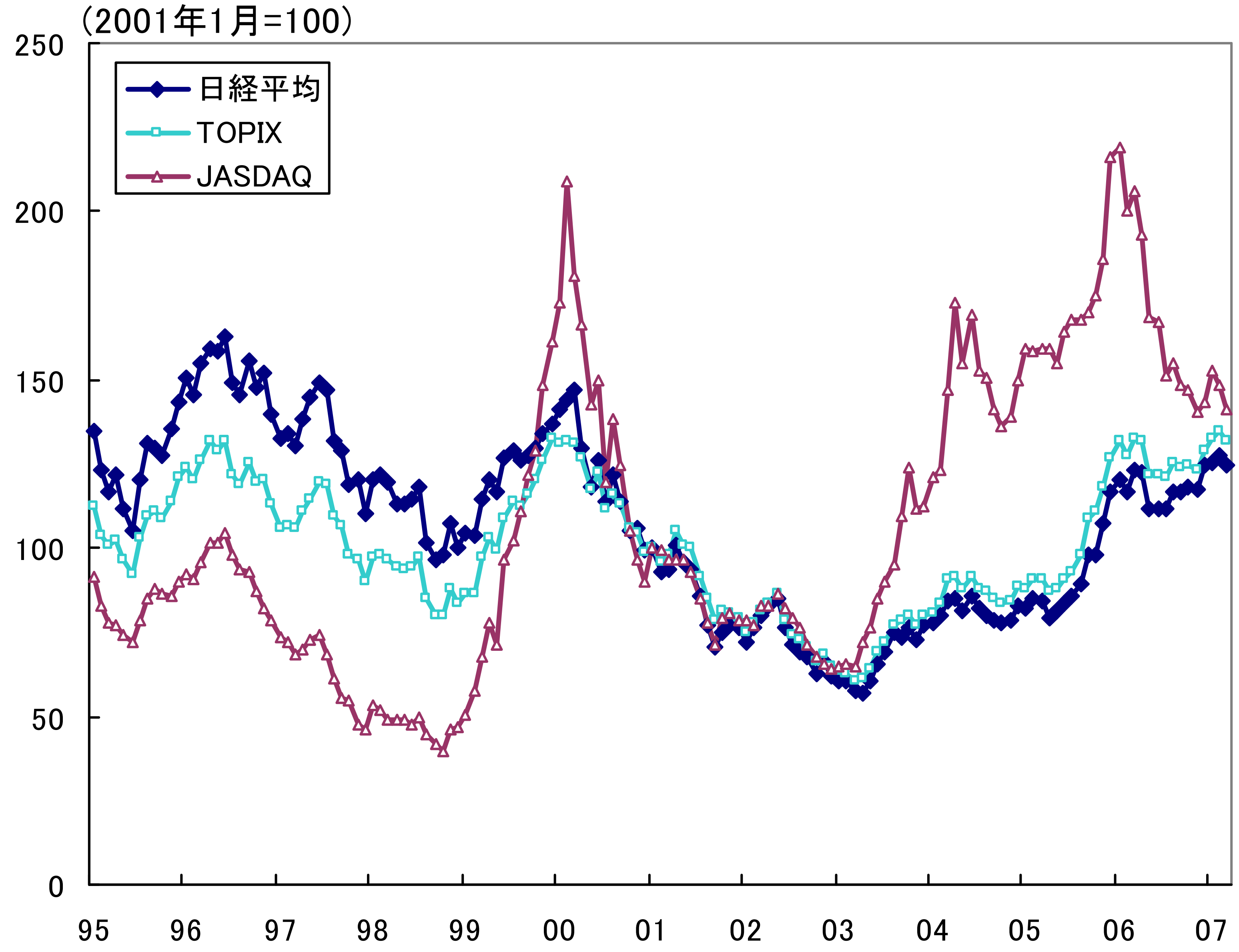
- 金融資産価格（株価、債券価格、為替レート）の決定要因、相互関連
- リスクとリターンの関係
- 投資理論（分散投資の効果等）

日経平均株価の推移



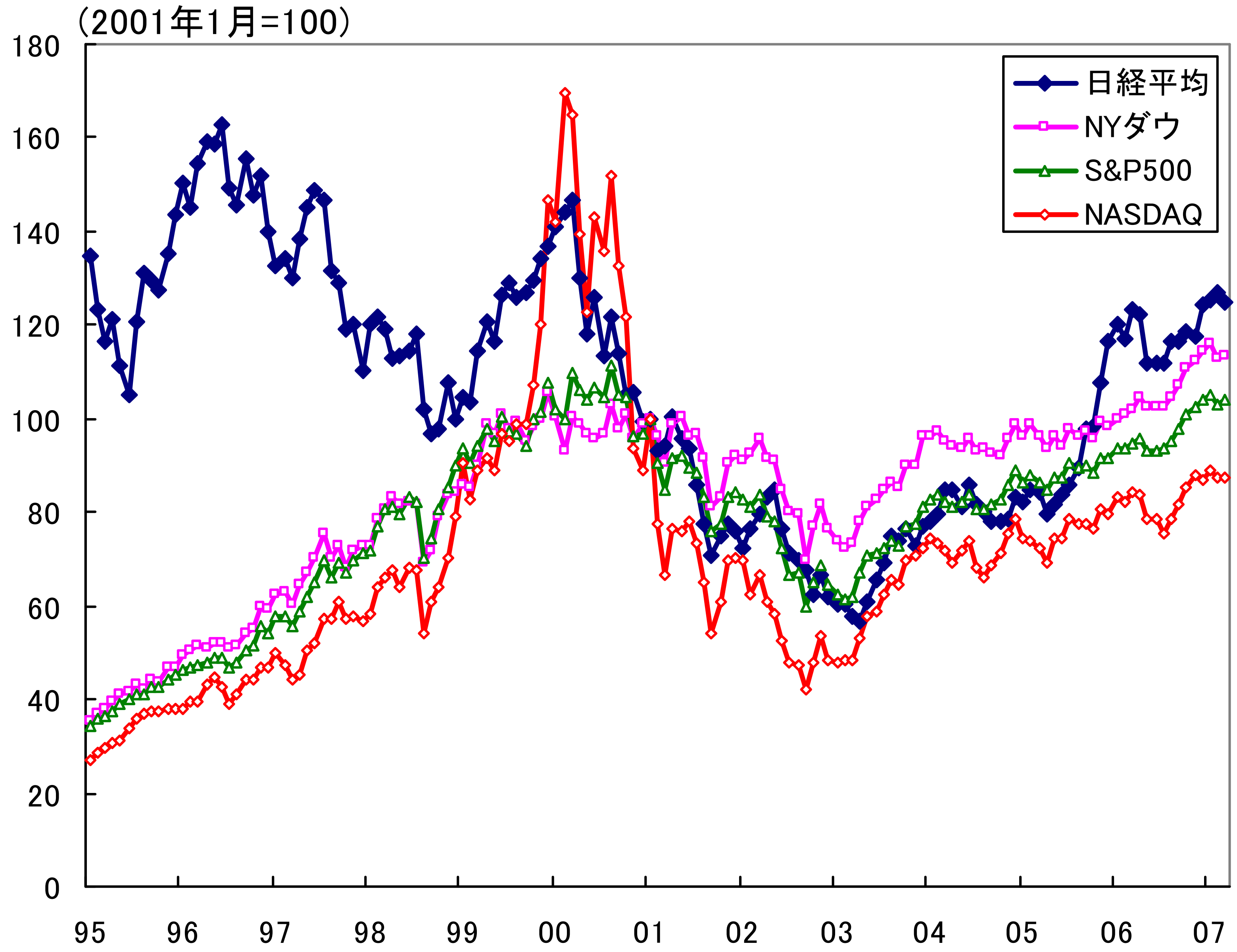
（データ）日経NEEDS

日本株価の推移



(データ)日経NEEDS

日米株価の推移

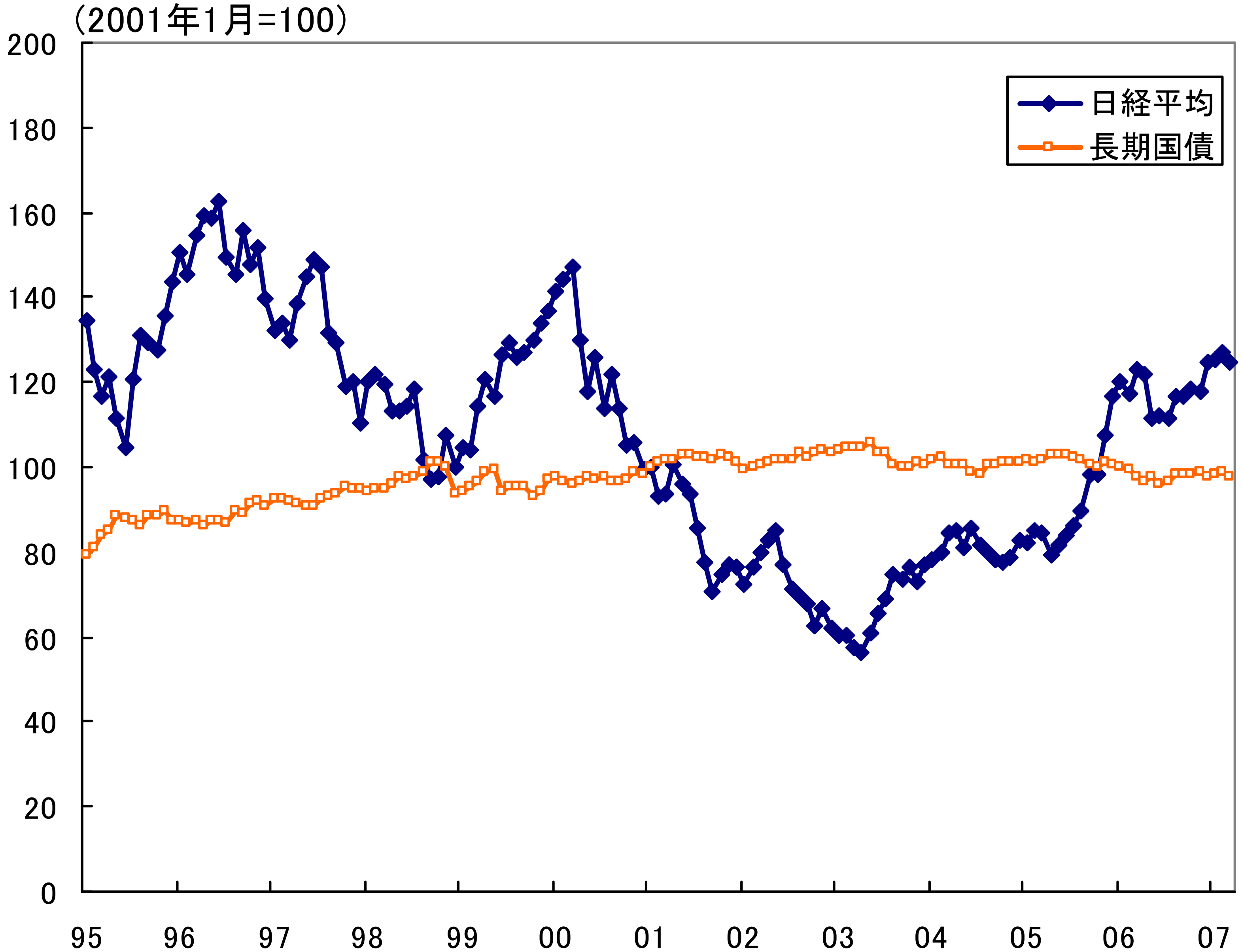


（データ）日経NEEDS, Yahoo Finance

外国為替(ドル・ユーロ)の推移

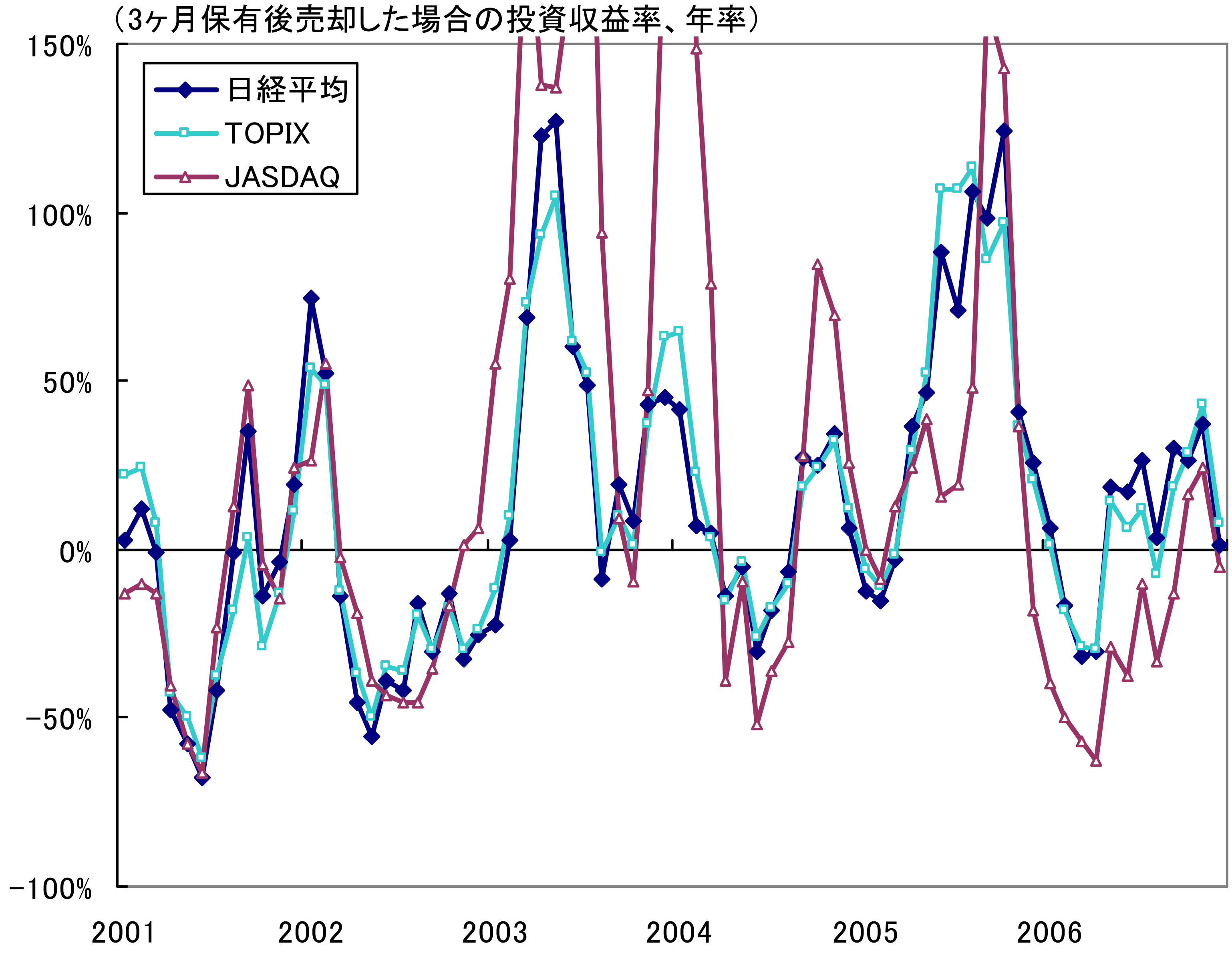
(データ)日経NEEDS, 日本銀行『金融経済統計』

国債価格と株価の推移



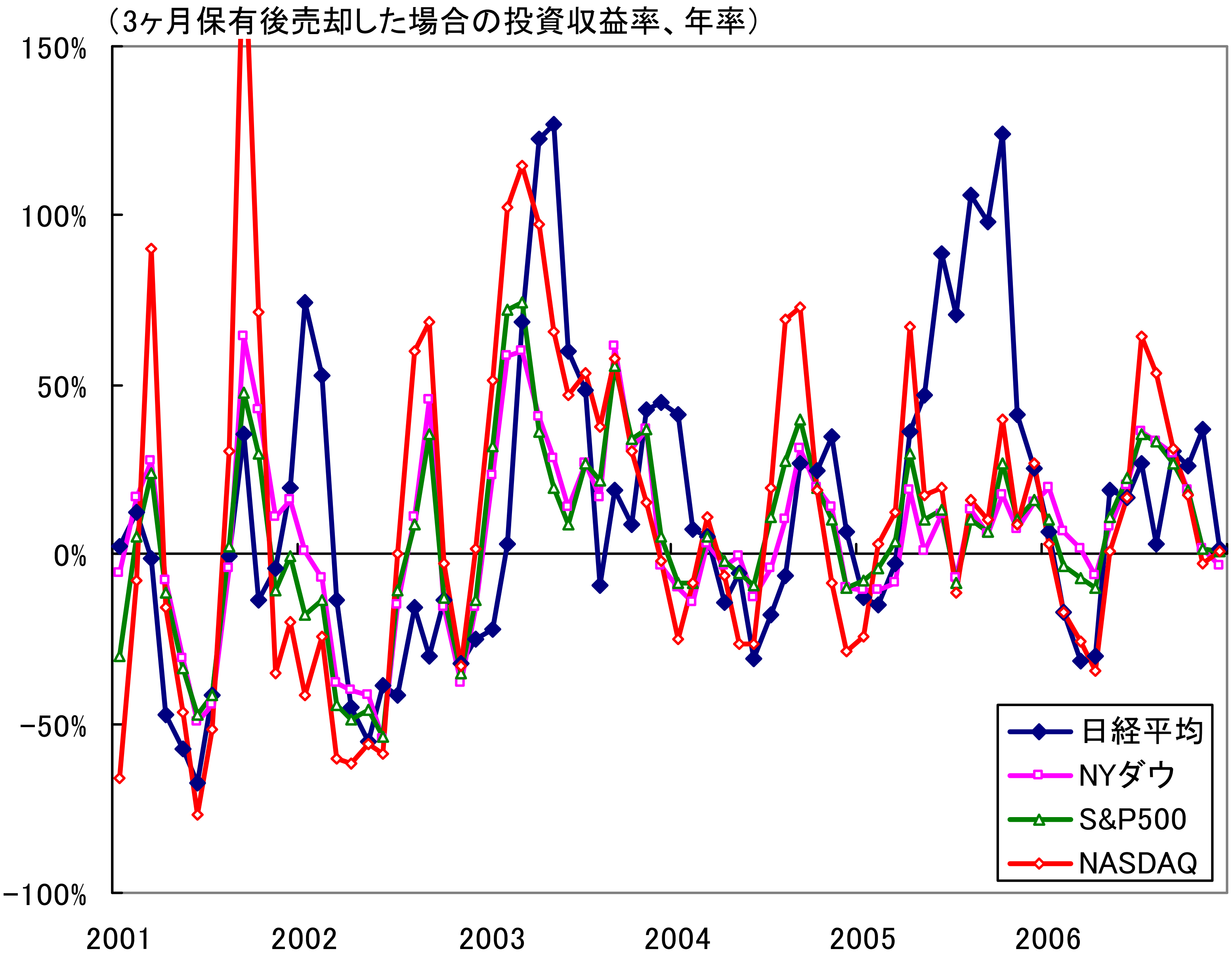
（データ）日経NEEDS 、日本経済新聞

日本株投資の収益率



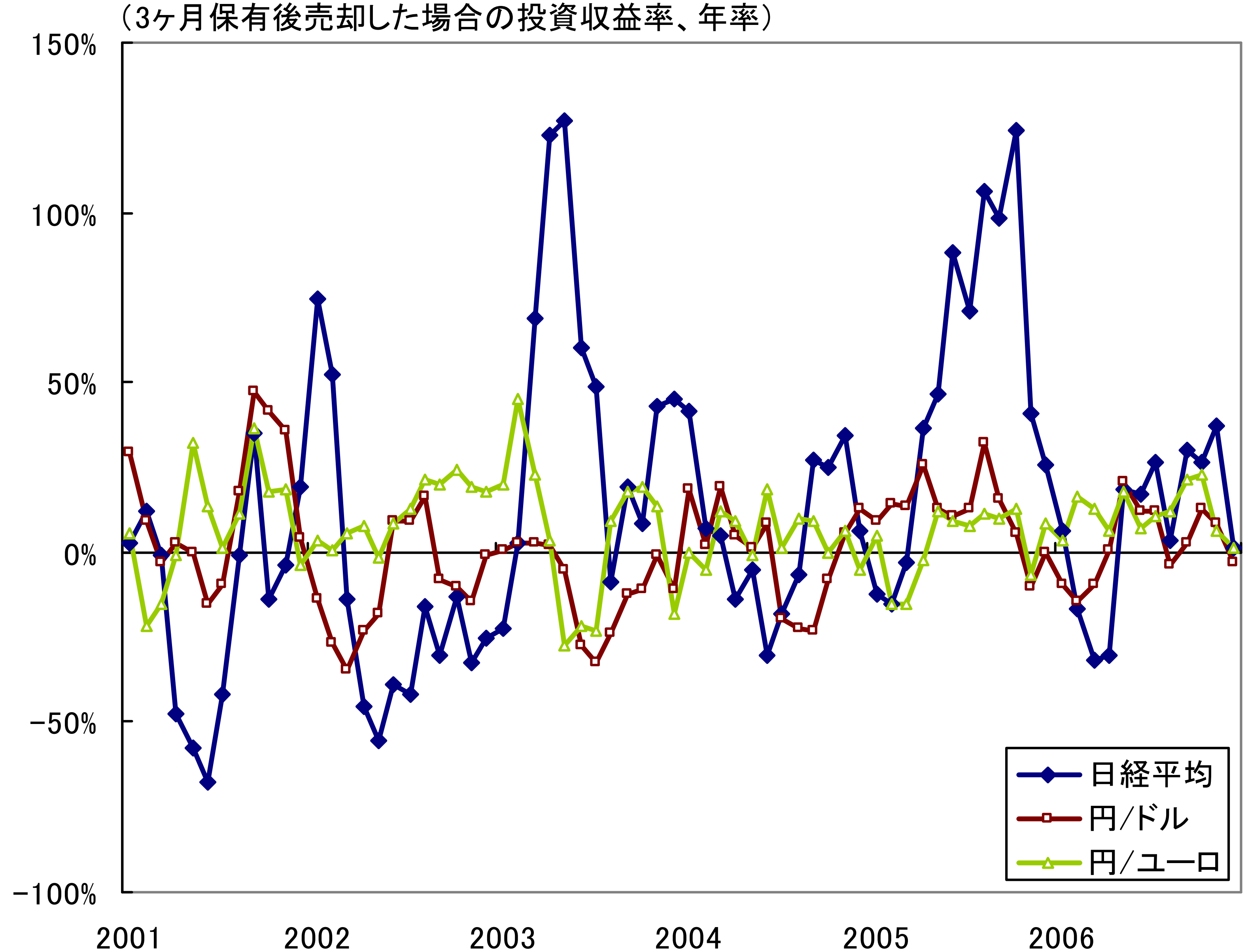
（データ）日経NEEDS

日米株式の収益率の推移



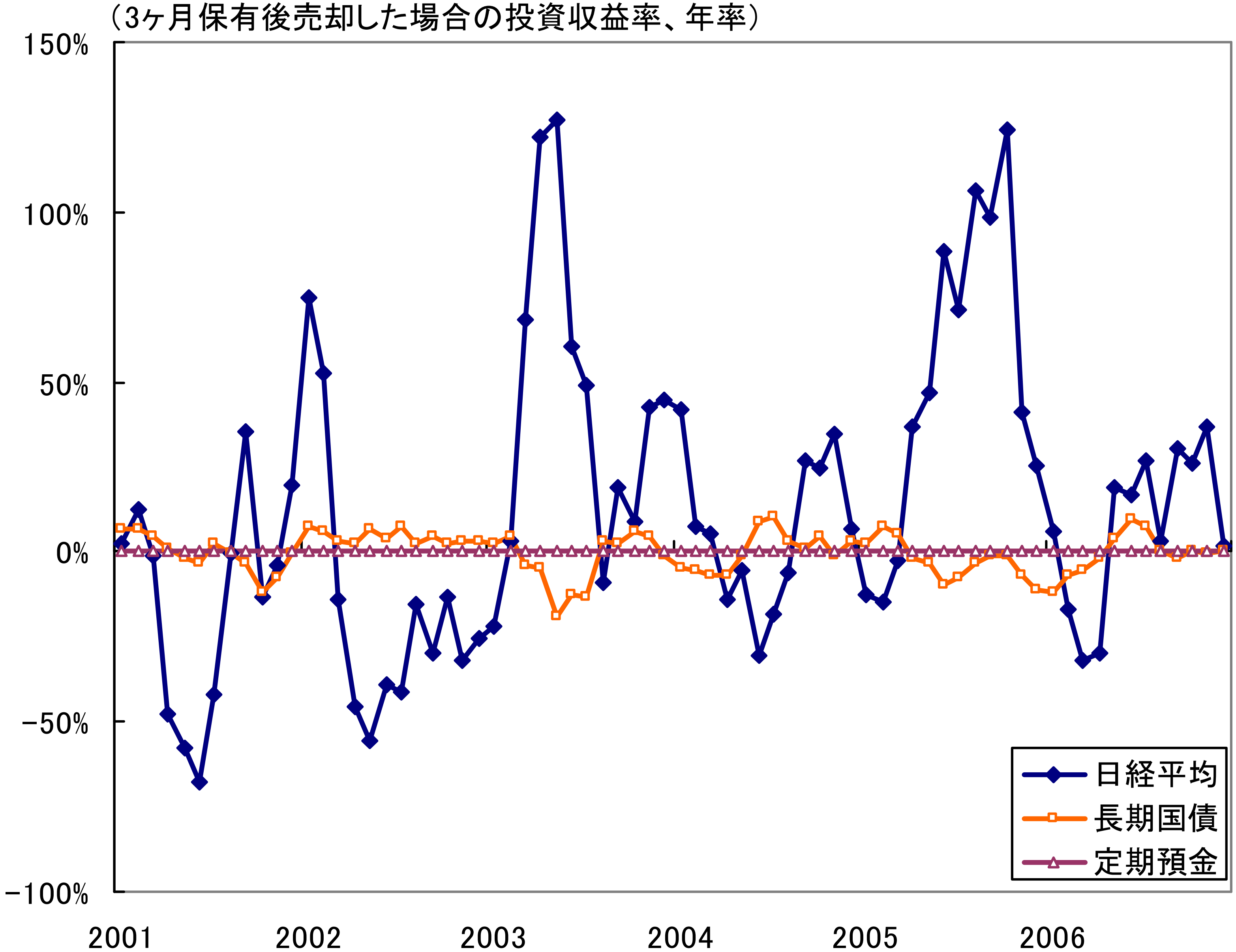
（データ）日経NEEDS, Yahoo Finance

外国為替投資の収益率推移



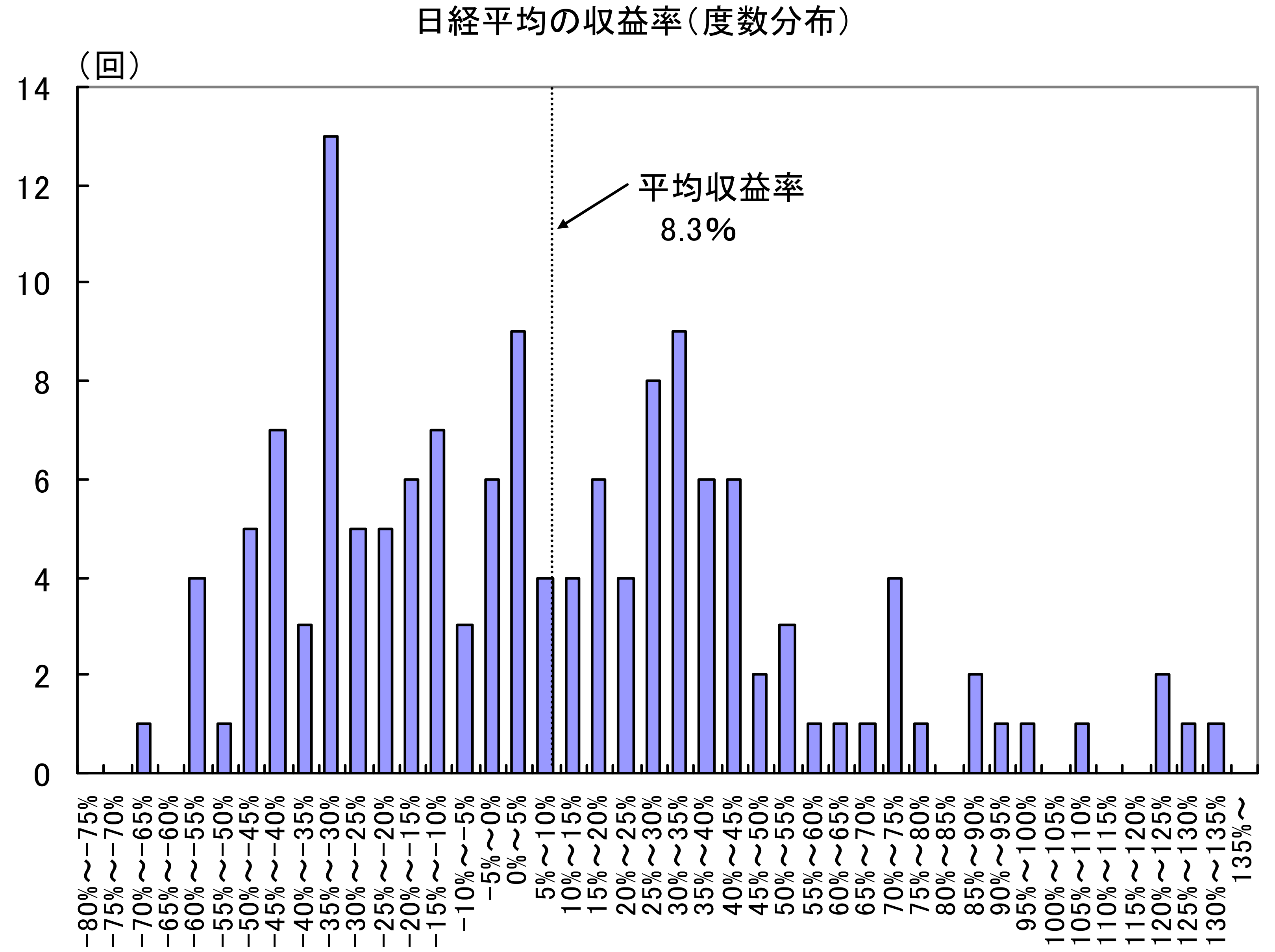
（データ）日経NEEDS, 日本銀行『金融経済統計』

国債投資の収益率推移



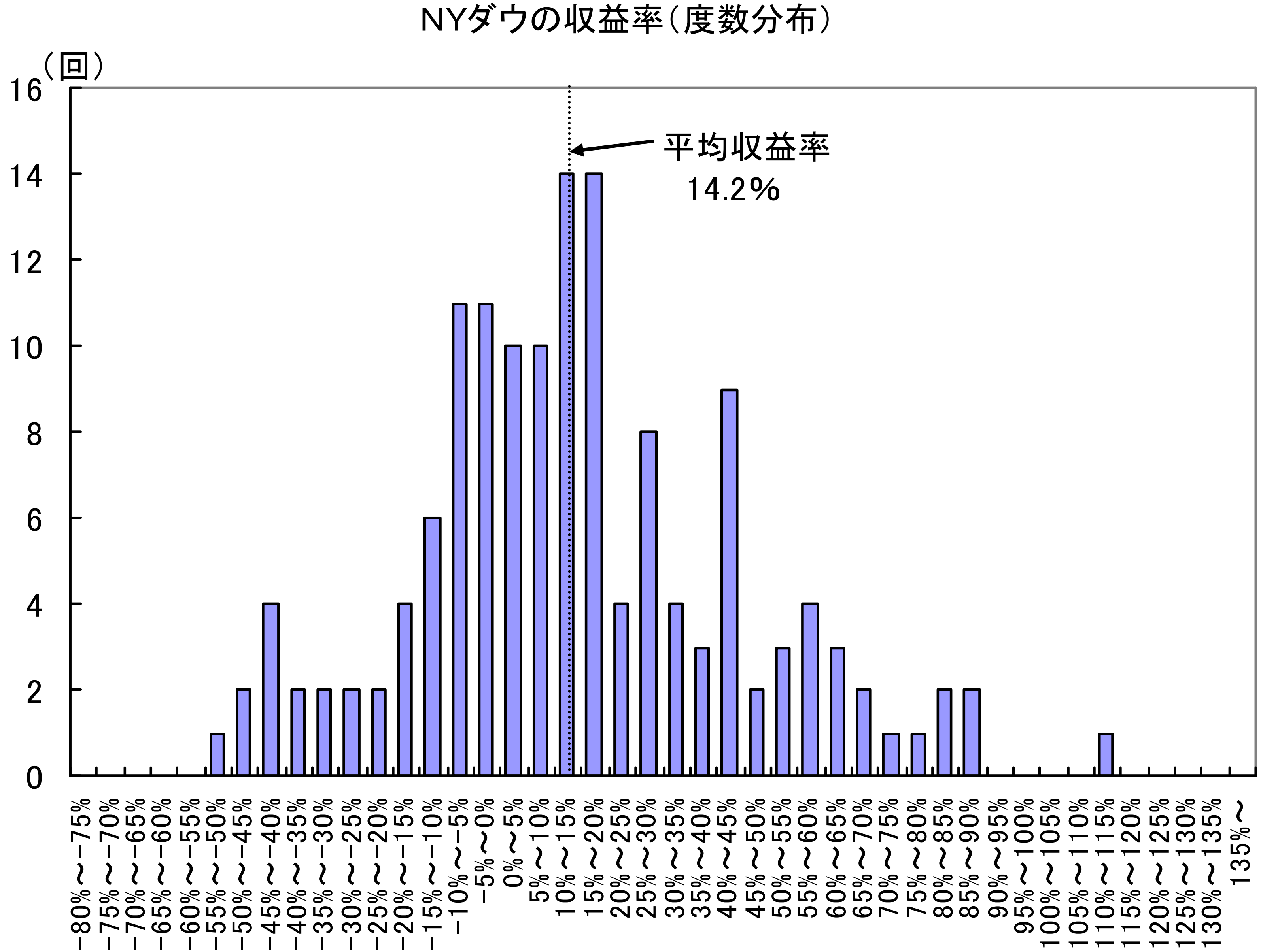
（データ）日経NEEDS、日本銀行『金融経済統計』、日本経済新聞

日本株投資の収益率の分布



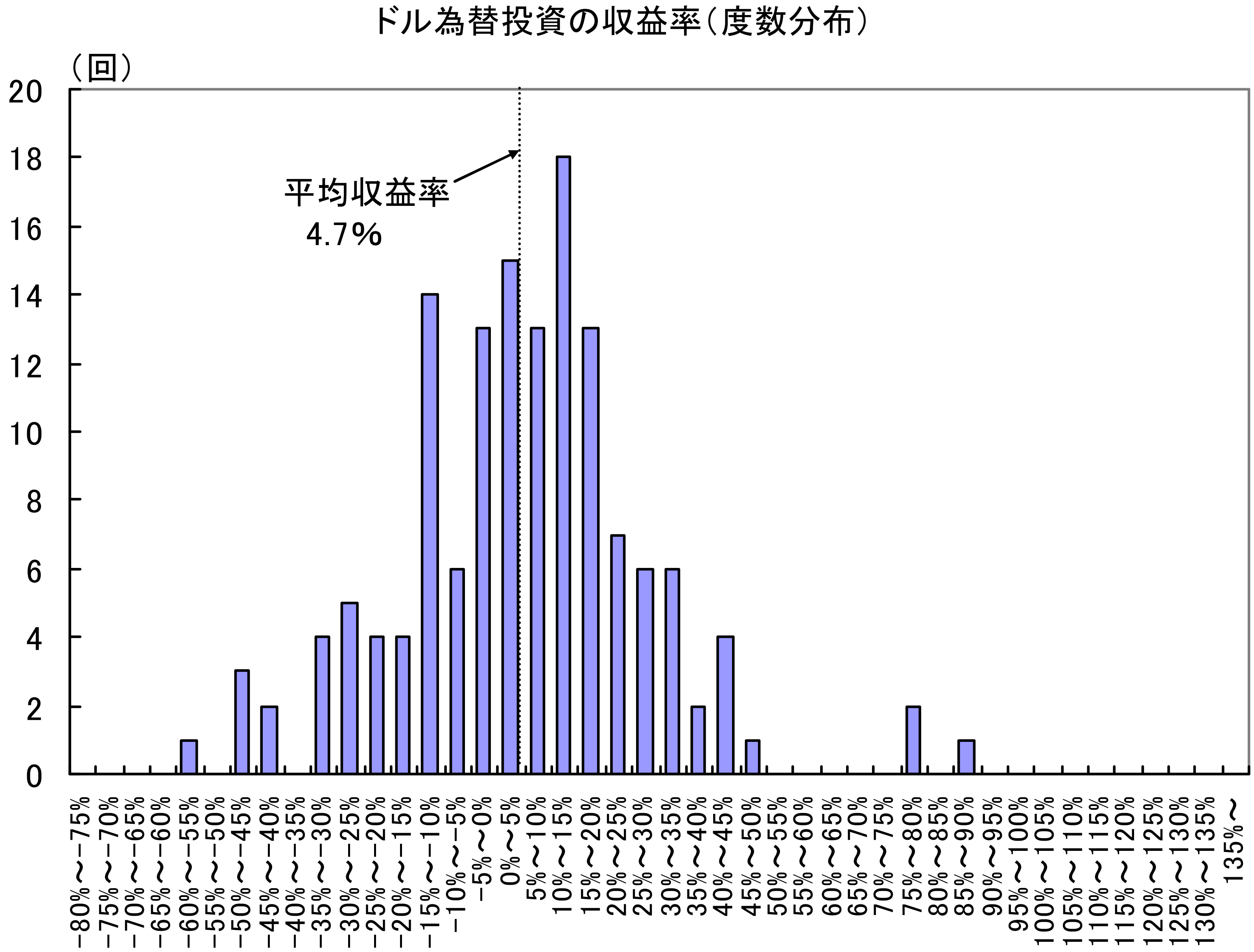
（データ）日経NEEDS

米国株投資の収益率の分布



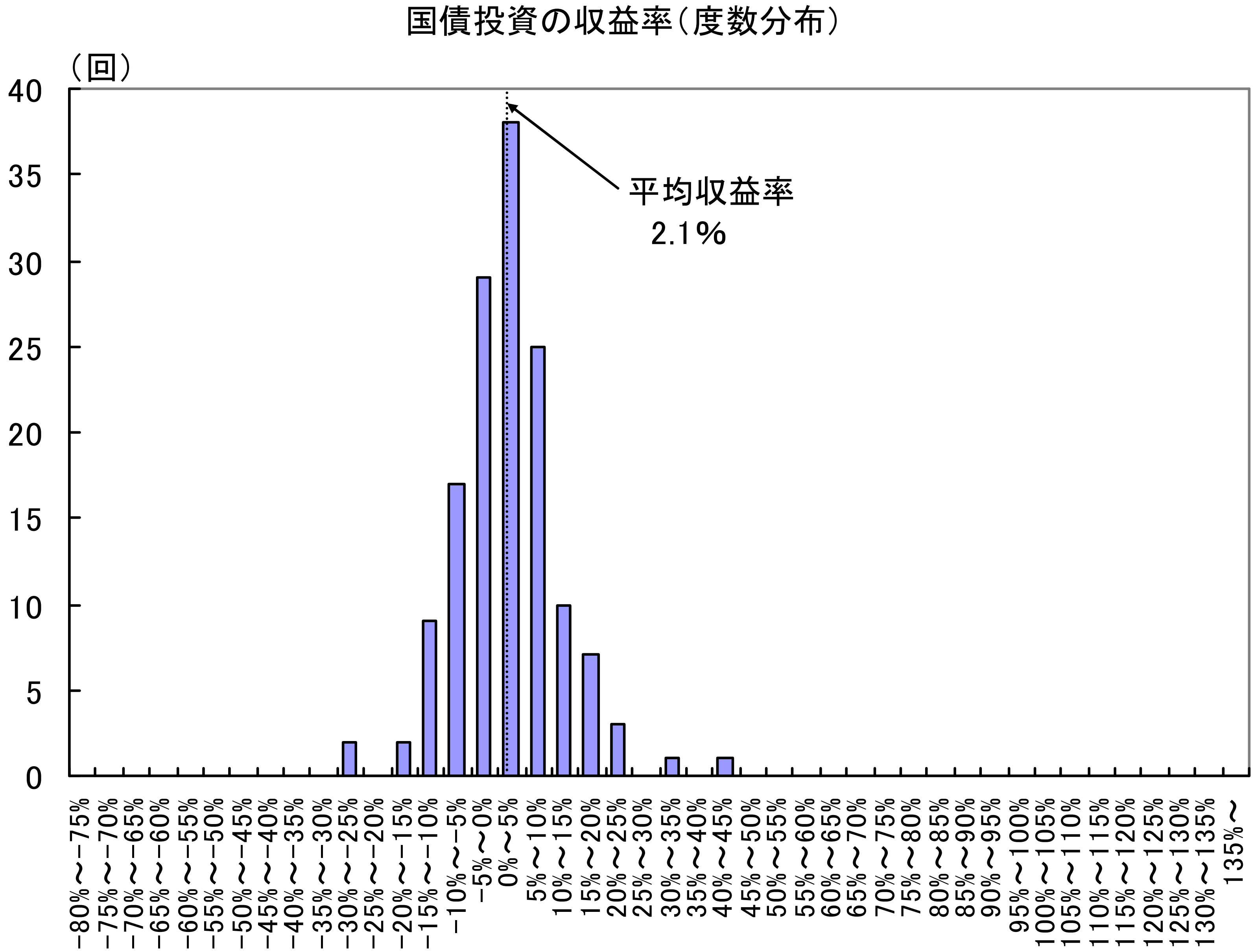
（データ）Yahoo Finance

外国為替投資の収益率の分布



（データ）日本銀行『金融経済統計』

国債投資の収益率の分布



（データ）日本経済新聞

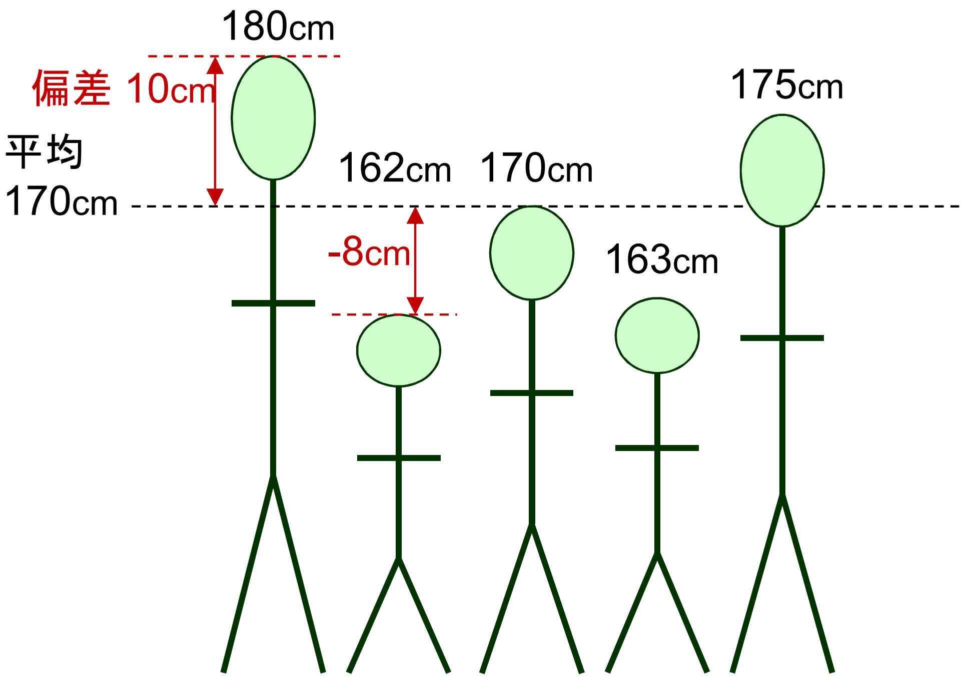
投資のリターンとリスク：収益率の期待値（平均）と分散

	日本株式	米国株式	為替		国債	預金
	日経平均	NYダウ	ドル	ユーロ	10年先物	3ヶ月定期
リターン 平均$\bar{x}$						
リスク 分散 s_x^2 標準偏差s_x						

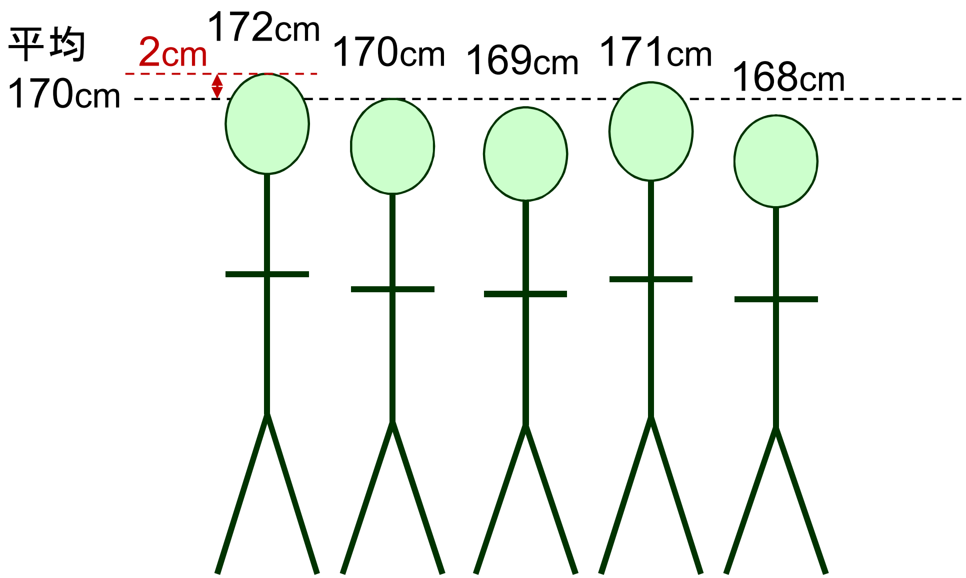
- リターンの指標・・・収益率の期待値（平均）
- リスクの指標・・・収益率の分散・標準偏差
  - 1回毎の投資の収益率は＋100%超～－50%超までバラツキ（リスク）・・・分散・標準偏差
  - 100回、200回と投資を繰り返せば平均的な収益率（リターン）は収斂＝期待収益率（収益率の期待値）

分散の考え方

【グループA】



【グループB】



身長	180	162	170	163	175	計
偏差						
(偏差)^2						

身長	172	170	169	171	168	計
偏差						
(偏差)^2						

$$\text{(母)分散} = \frac{(\text{偏差})^2\text{の合計}}{\text{人数}} = \frac{238}{5} = 47.6$$

$$\text{(母)標準偏差} = \sqrt{\text{(母)分散}} = \sqrt{47.6} = 6.9(cm)$$

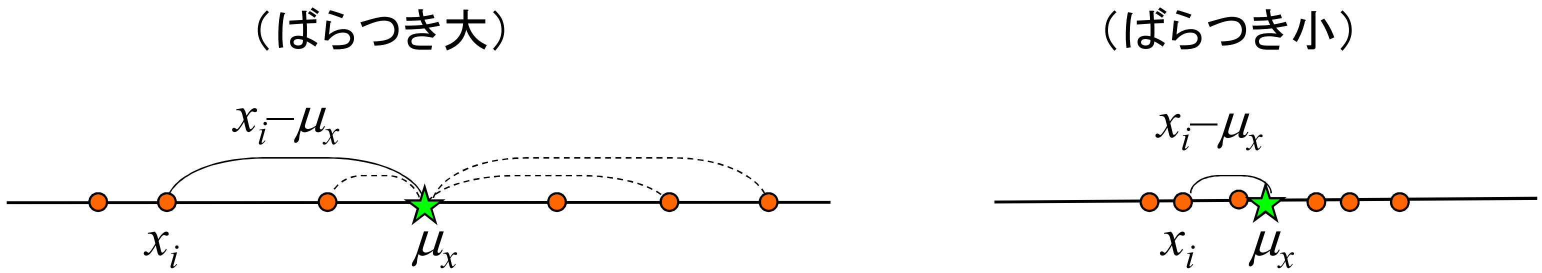
$$\text{(母)分散} = \frac{10}{5} = 2$$

$$\text{(母)標準偏差} = \sqrt{2} = 1.414(cm)$$

分散と標準偏差

- 分散（Variance）の考え方 ･･･ 変数の（平均回りの）ばらつきを表す

  $x_i - \mu_x$ … x_i の偏差（Deviation）: 要素 x_i の平均からの乖離



全ての要素 $x_1, x_2, \ldots, x_N$ を総合的に見た偏差の程度を表す指標が欲しい

- 偏差を単純に平均化 [$\Sigma(x_i - \mu_x)/N$] すると正負が相殺してしまう
- 偏差を二乗して平均化 $\Sigma(x_i - \mu_x)^2/N$ ⇒ 分散の概念

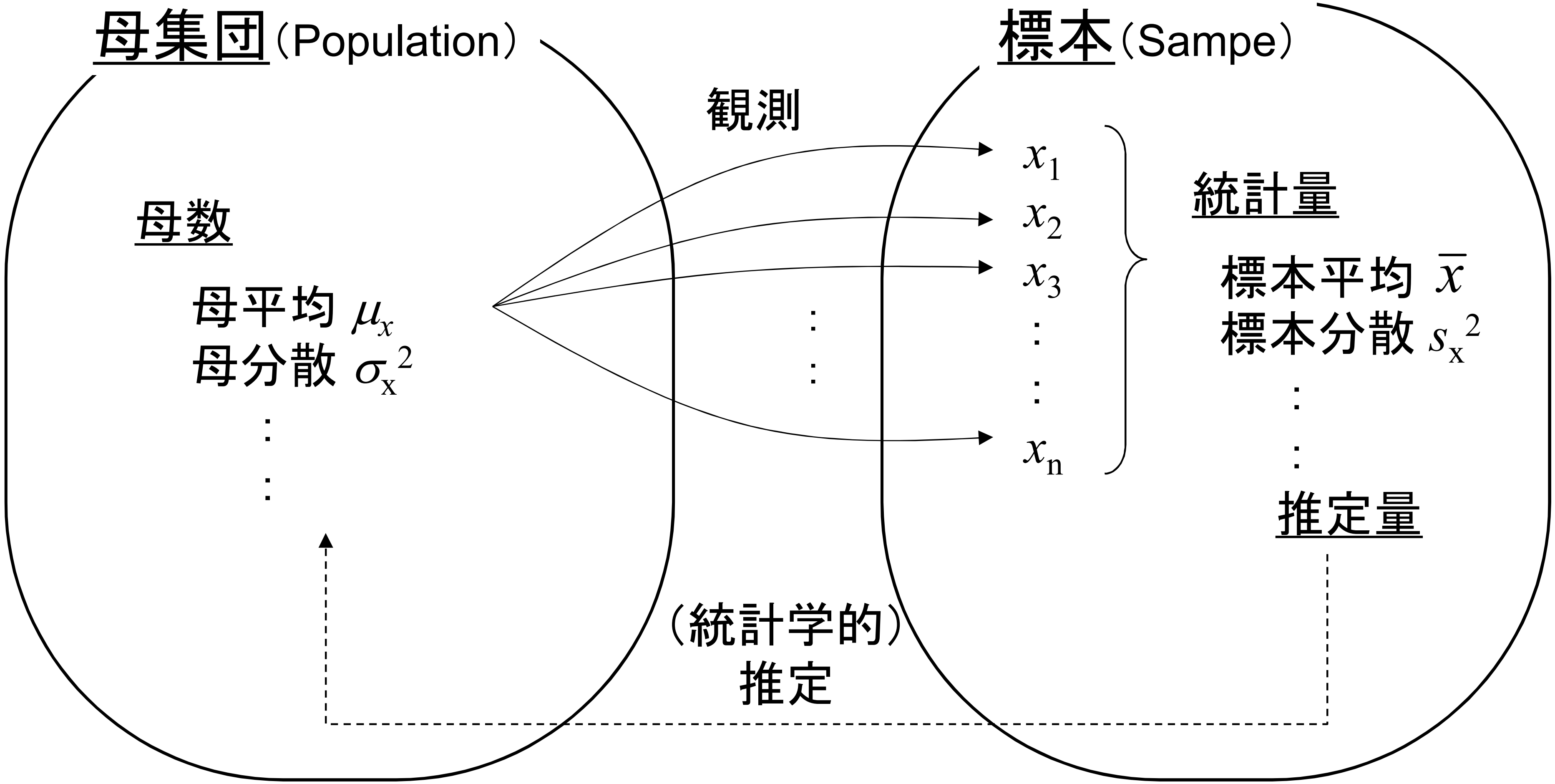
- 標準偏差（Standard Deviation） ― 分散の平方根

$$\sigma_x = \sqrt{\mathrm{var}(x)}$$

統計学的推定：
母集団と標本／母数と推定量

母集団/母数（全数調査の場合）		標本/推定量（標本調査の場合）	
期待値（母平均）	$\mu_x = \dfrac{\sum_{i=1}^{N} x_i}{N}$	標本平均	$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$
母分散	$\sigma_x^2 = \dfrac{\sum_{i=1}^{N} (x_i - \mu_x)^2}{N}$	標本分散	$s_x^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$
母標準偏差	$\sigma_x = \sqrt{\sigma_x^2}$	標本標準偏差	$s_x = \sqrt{s_x^2}$
母共分散	$\sigma_{xy} = \dfrac{\sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)}{N}$	標本共分散	$s_{xy} = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$
母相関係数	$\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$	標本相関係数	$r_{xy} = \dfrac{s_{xy}}{s_x s_y}$

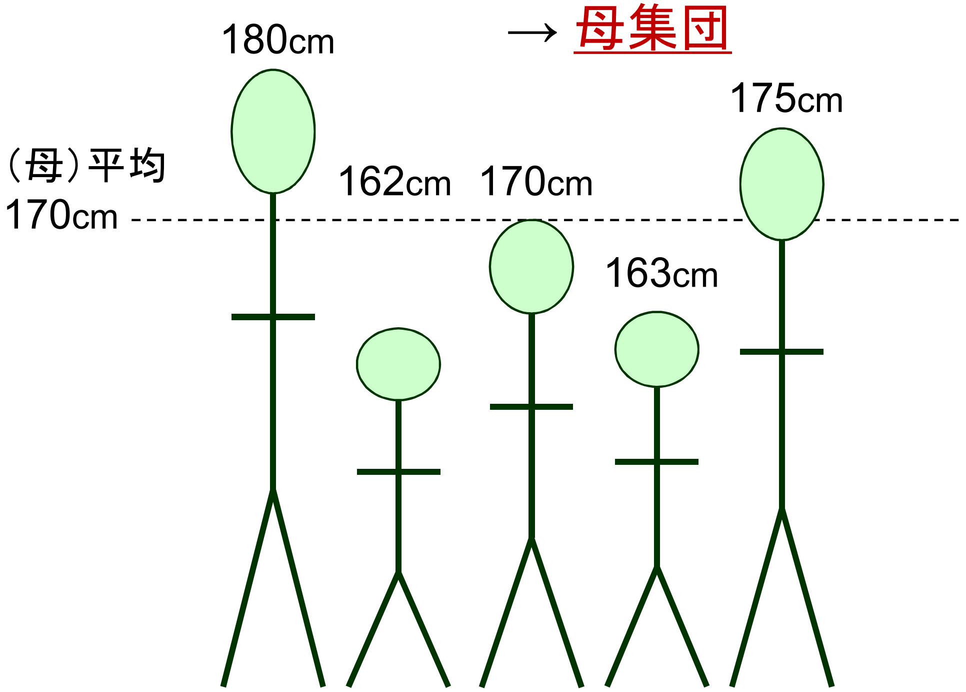
統計学的推定: 母集団と標本／母数と推定量



- 母数(Parameter): 母集団の特性を表わす値
- 統計量(Statistics): 標本の関数
- 推定量(Estimator): 母数の推定のために用いられる統計量
- 推定値(Estimates): 実際の標本を用いて計算した推定量の値

母分散と標本分散

【グループA（全員）】
→ 母集団



身長	180	162	170	163	175	計
偏差	10	-8	0	-7	5	0
(偏差)^2	100	64	0	49	25	238

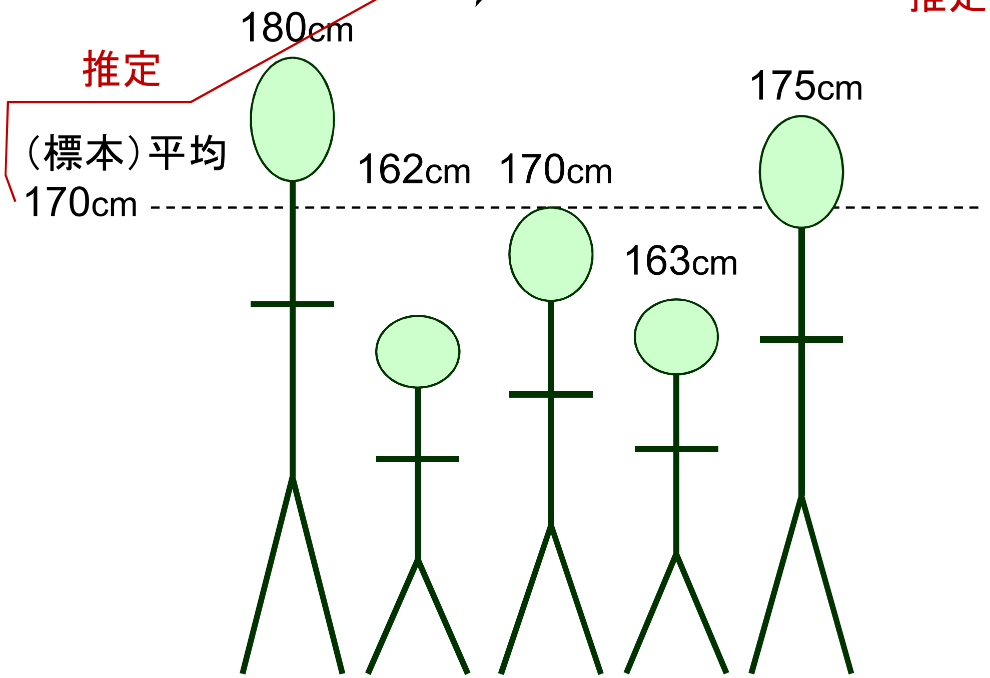
$$(\text{母})\text{分散} = \frac{(\text{偏差})^2\text{の合計}}{\text{人数}} = \frac{238}{5} = 47.6$$

母集団
【グループX（全員）】
母平均：？
母分散：？

うち５人
【標本】

推定

推定



身長	180	162	170	163	175	計
偏差	10	-8	0	-7	5	0
(偏差)^2	100	64	0	49	25	238

$$\text{標本分散} = \frac{(\text{偏差})^2\text{の合計}}{\text{人数}-1} = \frac{238}{4} = 59.5$$

標本平均、標本分散の性質

- 標本平均、標本分散はそれぞれ母平均、母分散の
  - □不偏推定量

    ・・・過大・過小といった偏りのない推定量

    $$\boxed{\mathrm{E}(\bar{x}) = \mu_x, \mathrm{E}(s_x^2) = \sigma_x^2}$$

  - □一致推定量

    ・・・標本数が増えると真の値に限りなく近づく推定量

    $$\boxed{n \to \infty \quad \text{のとき} \quad \bar{x} \to \mu_x, s_x^2 \to \sigma_x^2}$$

  である

※「大数の法則」

標本数が十分に大きければ標本平均は母平均に近づく

（参考）日本代表の監督別戦績

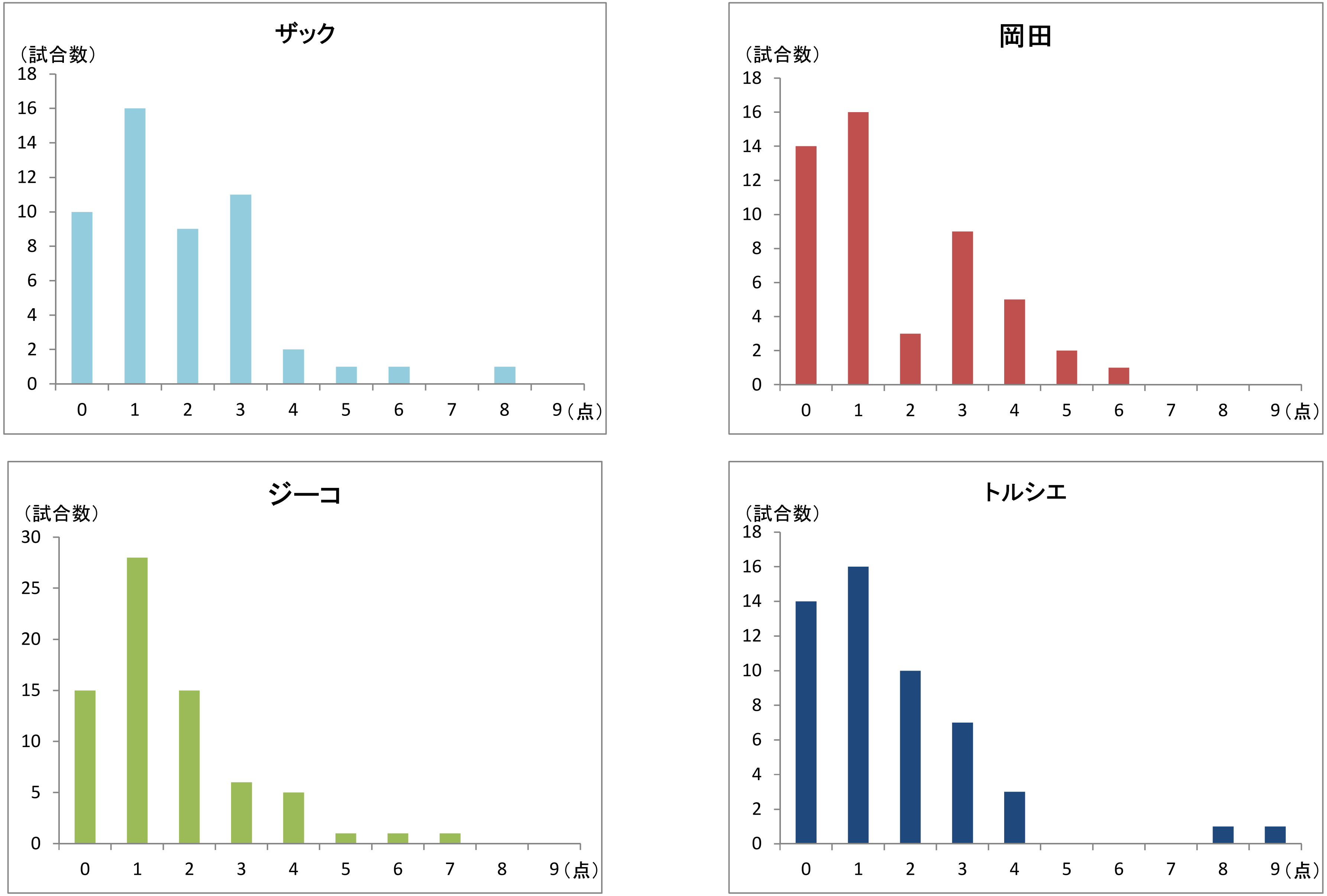
	ザック	岡田	オシム	ジーコ	トルシエ
試合数	51	50	20	72	52
勝ち数（勝率）	29（57%）	26（52%）	12（60%）	38（53%）	24（46%）
負け数（敗率）	11（22%）	11（22%）	3（15%）	19（26%）	11（21%）
引き分け（引分率）	11（22%）	13（26%）	5（25%）	15（21%）	17（33%）

	得点					失点				
	ザック	岡田	オシム	ジーコ	トルシエ	ザック	岡田	オシム	ジーコ	トルシエ
平均 $\bar{x}$	1.84	1.70	1.80	1.58	1.65	1.04	0.82	0.70	0.97	0.92
分散 s_x^2	2.69	2.66	1.96	2.13	3.33	1.32	1.01	0.85	1.32	1.29
標準偏差 s_x	1.64	1.63	1.40	1.46	1.82	1.15	1.00	0.92	1.15	1.13

	得失点差				
	ザック	岡田	オシム	ジーコ	トルシエ
平均 $\bar{x}$	0.80	0.88	1.10	0.61	0.73
分散 s_x^2	4.40	4.19	1.88	3.34	5.34
標準偏差 s_x	2.10	2.05	1.37	1.83	2.31

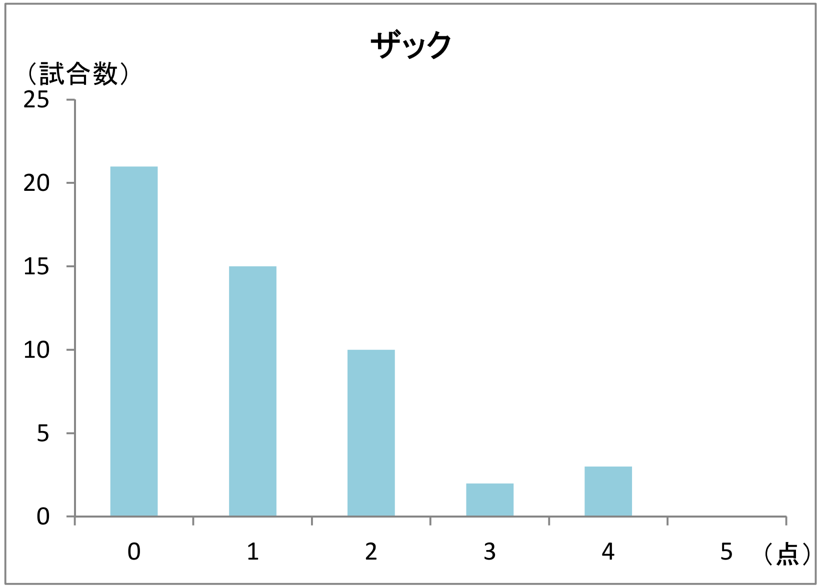
（データ）日本代表サッカー倶楽部（http://kick.or.tv/）

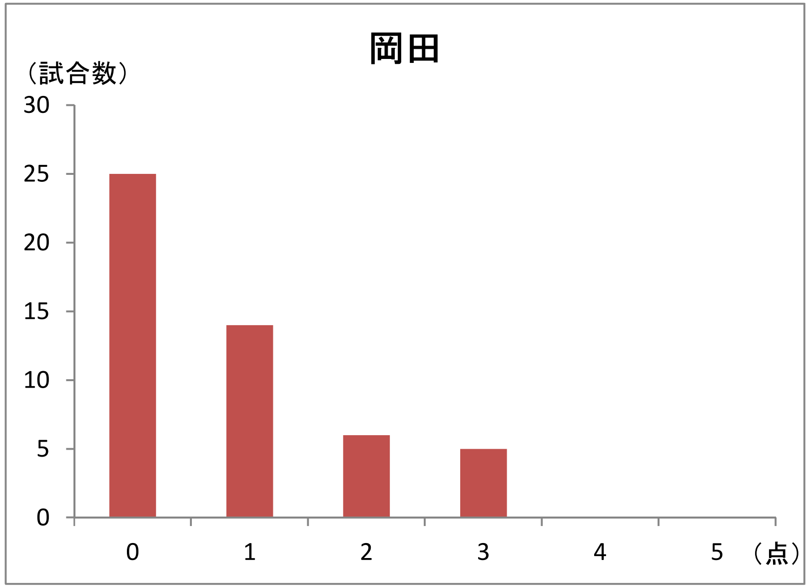
（参考）監督別日本代表の得点分布

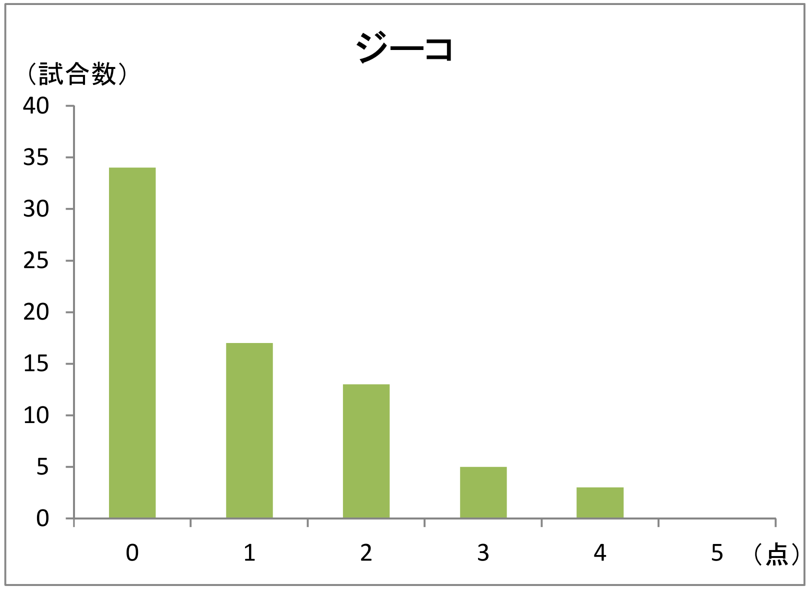


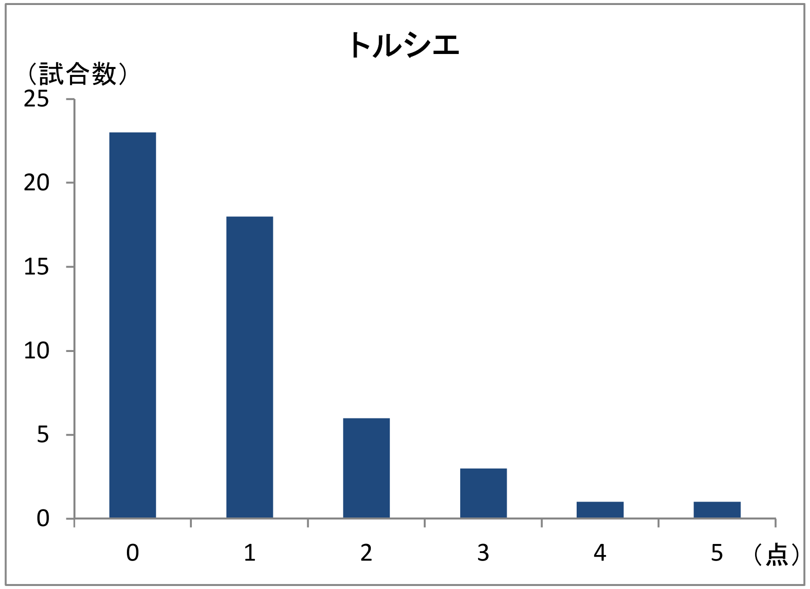
（データ）日本代表サッカー倶楽部（http://kick.or.tv/）

（参考）監督別日本代表の失点分布



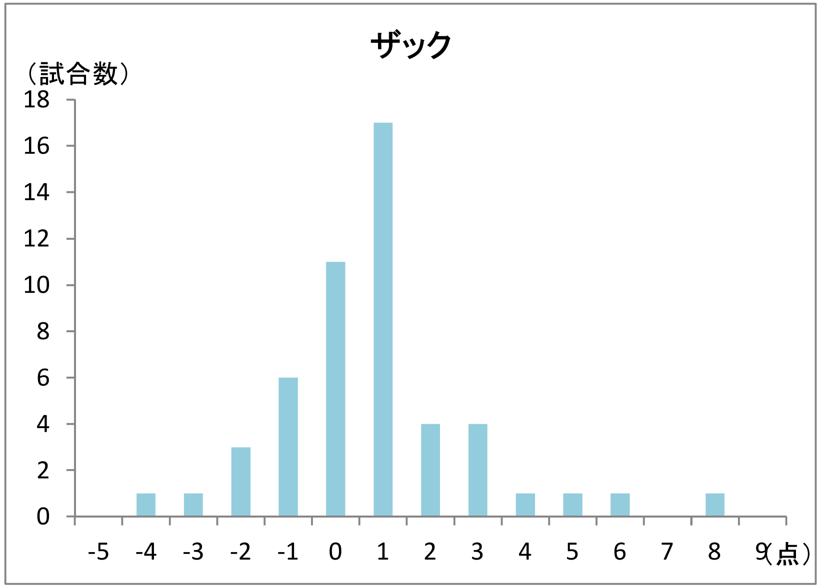


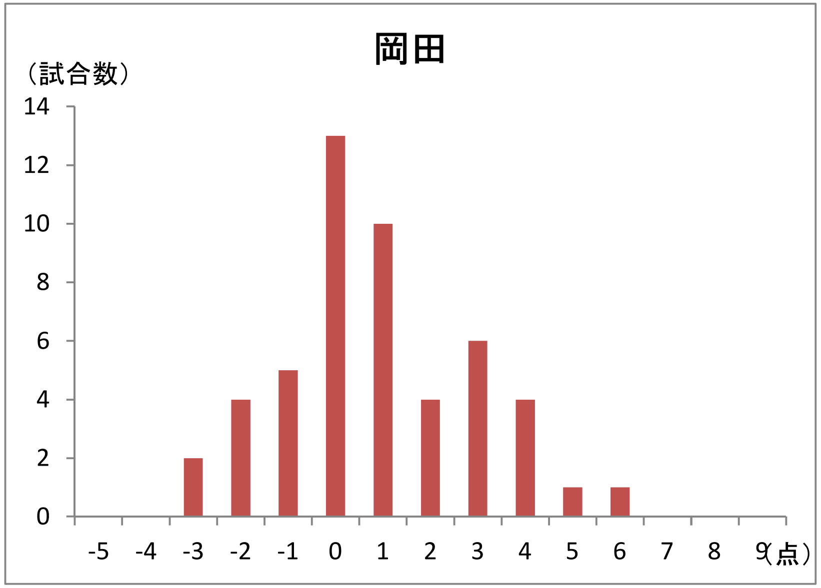


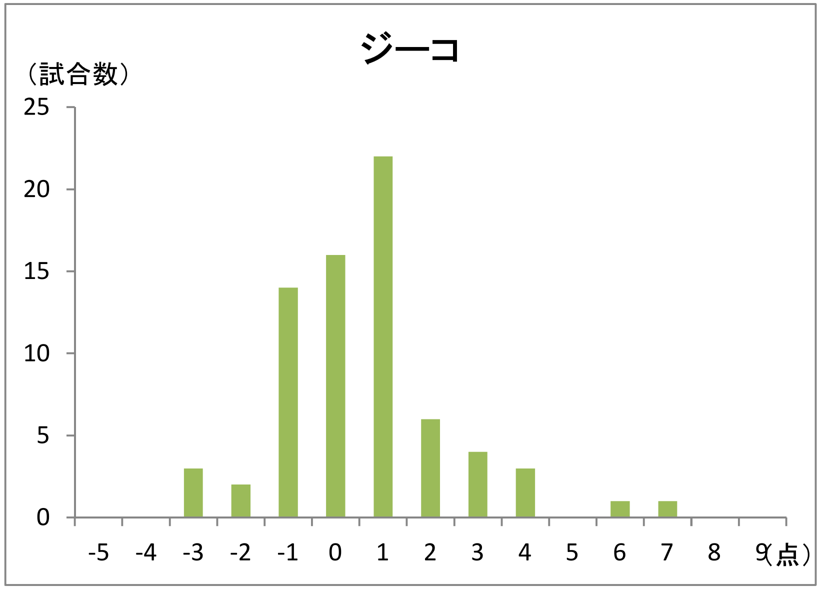


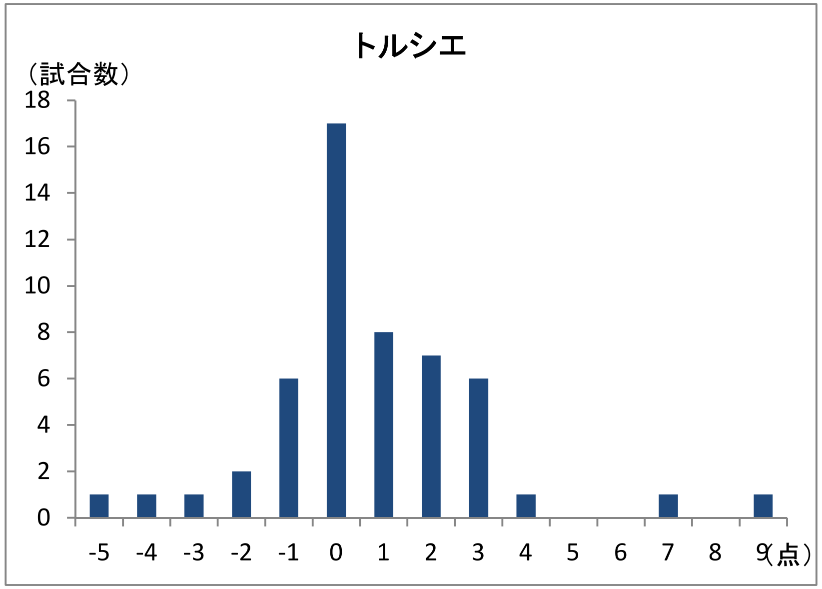
（データ）日本代表サッカー倶楽部（http://kick.or.tv/）

（参考）監督別日本代表の得失点差分布



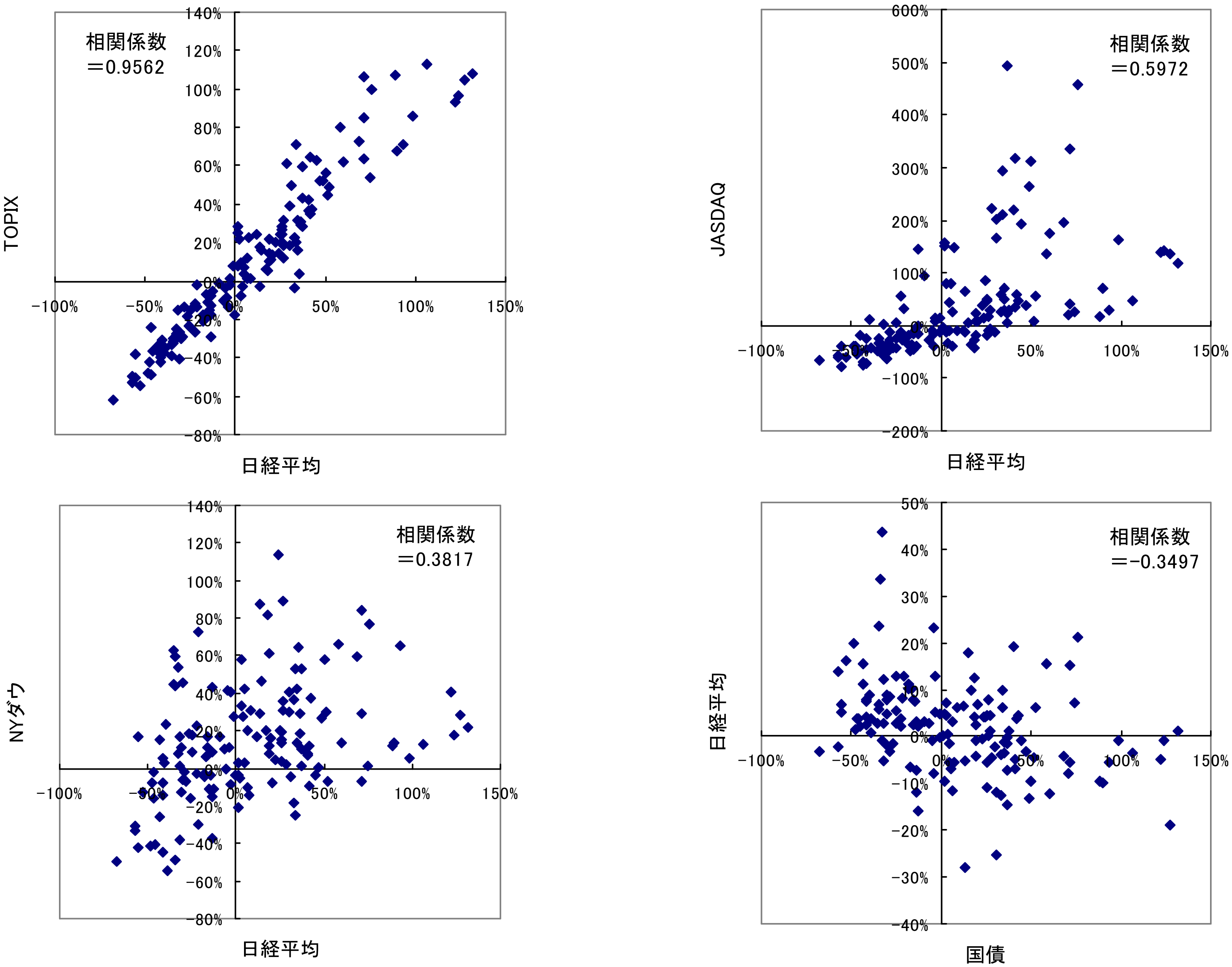






（データ）日本代表サッカー倶楽部（http://kick.or.tv/）

資産別収益率の相関（散布図）



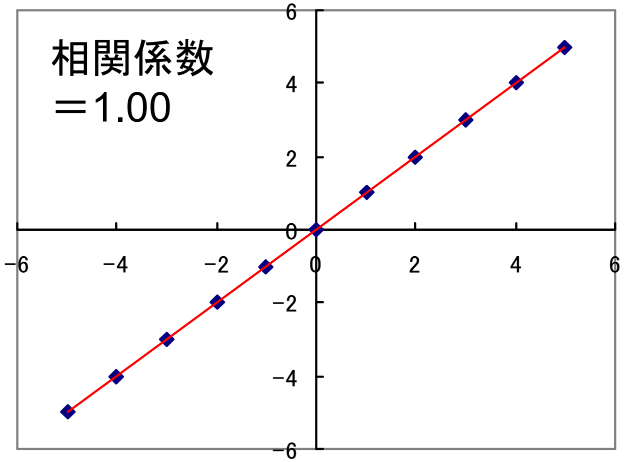
資産別収益率の相関係数

	日本株式			米国株式	為替		国債	預金
	日経平均	TOPIX	JASDAQ	NYダウ	ドル	ユーロ	10年先物	3ヶ月定期
日経平均								
TOPIX								
JASDAQ								
NYダウ								
ドル								
ユーロ								
国債								
3ヶ月定期								

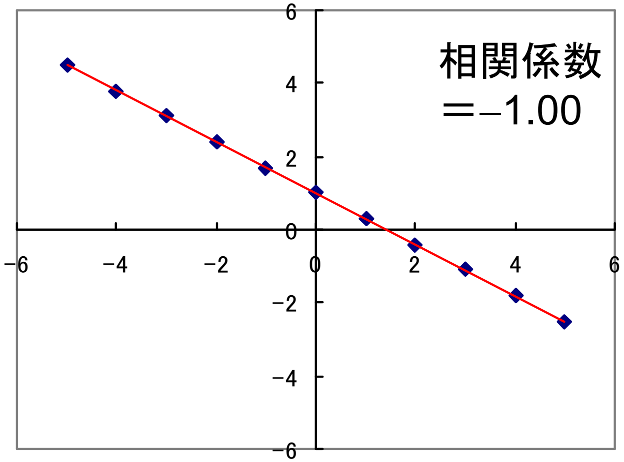
相関係数の性質

- <u>直線的関係</u>の強さを表す。正の直線的関係が強いほど1に、負の直線的関係が強いほど－1に近づく。非線形の関係を検出する力はない。

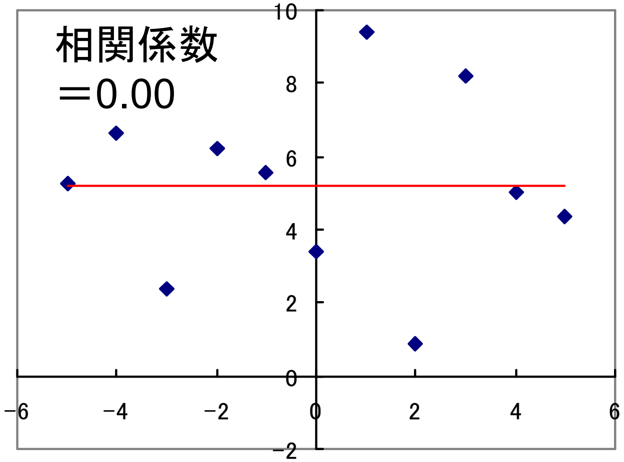
完全な正の相関①



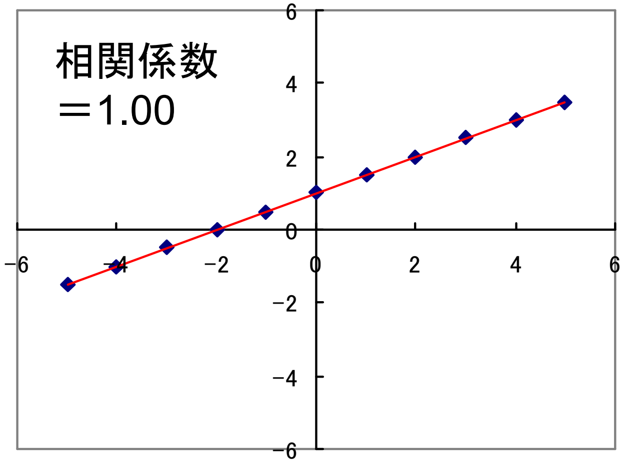
完全な負の相関



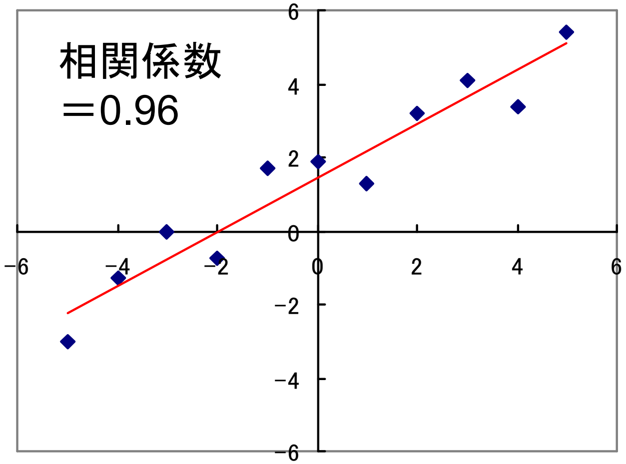
完全な無相関



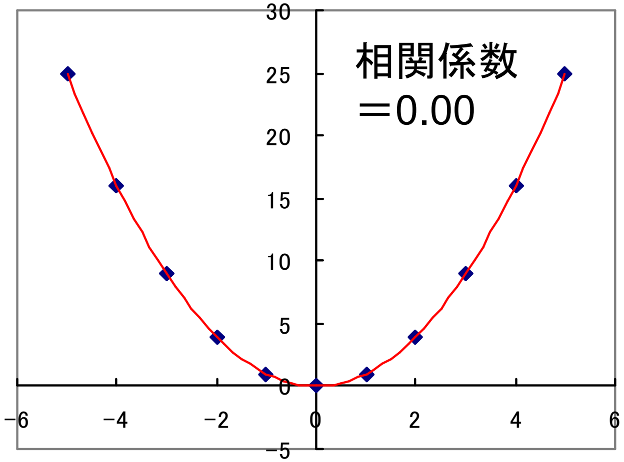
完全な正の相関②



強い正の相関



厳密な2次の関係



分散投資の効果

- ２種類の資産(x, y)に、$\alpha:\beta$ の割合で投資($\alpha+\beta=1$)した場合のリターン(期待収益率)とリスク(分散)は、
- リターン(収益率の期待値)

$$E(\alpha x+\beta y) = \alpha E(x) + \beta E(y)$$

・・・各資産のリターンの加重平均

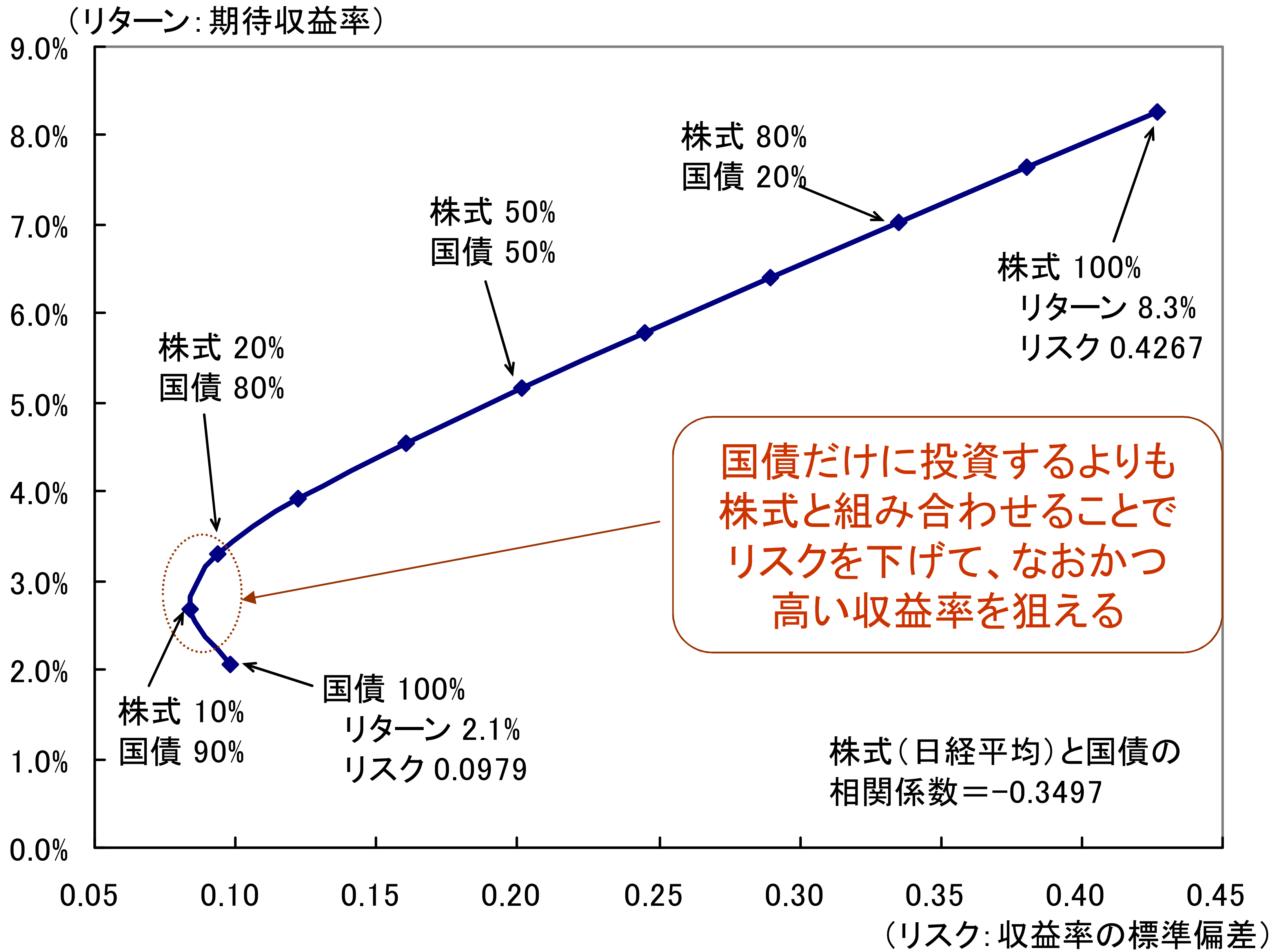
- リスク(分散)

$$Var(\alpha x+\beta y) = \alpha^2 Var(x) + \beta^2 Var(y) + 2\alpha\beta Cov(x,y)$$

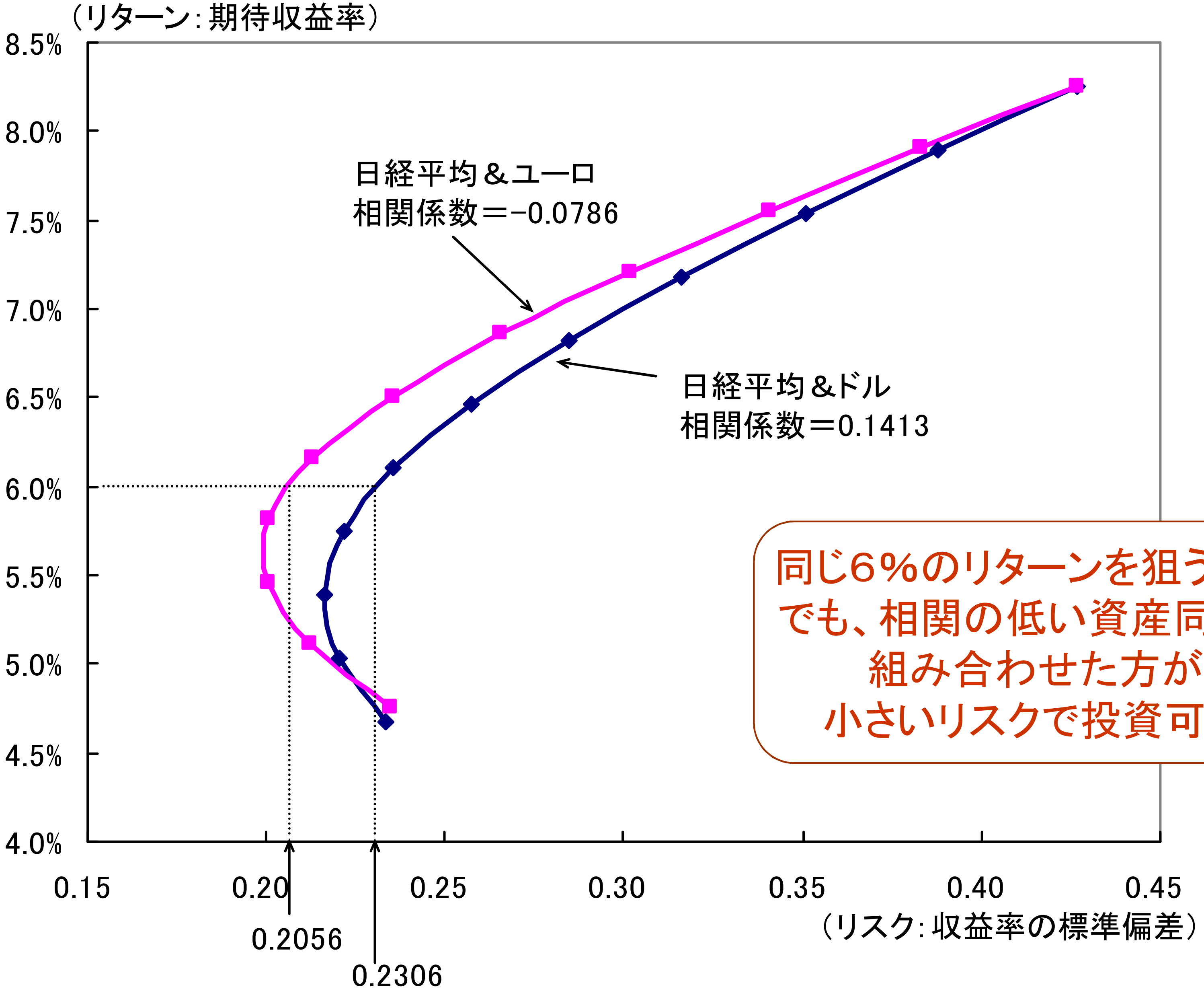
・・・負に相関する資産(Covが負)を組み合わせるとリスク縮小の効果が大きい

※ E( ), Var( ), Cov( )はそれぞれ期待値、分散、共分散を表す
上式の導出は、統計学の教科書等の和の平均と分散の公式を参照

分散投資の効果
例1： 株式(日経平均)と国債の組合せ



分散投資の効果
例2：株式(日経平均)と外国為替の組合せ



同じ6%のリターンを狙う場合でも、相関の低い資産同士を組み合わせた方が小さいリスクで投資可能

正規分布の仮定と確率計算

- 変数xが平均μ_x, 分散σ_x^2の正規分布に従うとき、以下のように標準化した変数zは平均0, 分散1の正規分布（＝標準正規分布）に従う

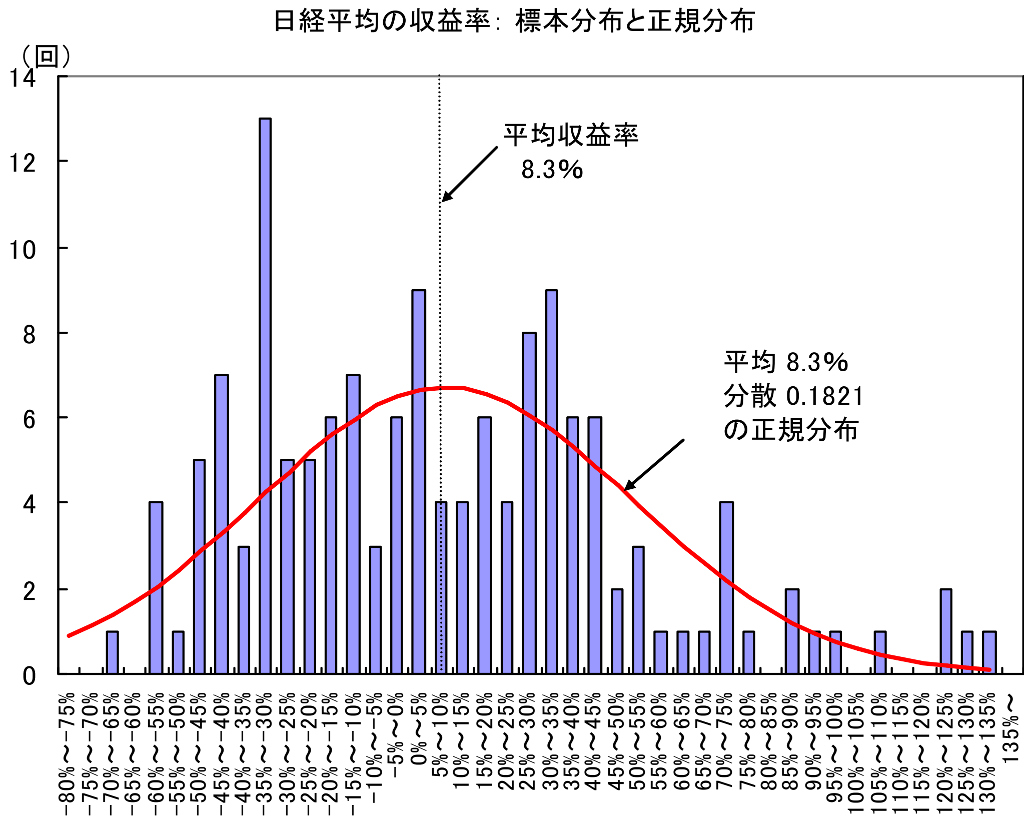
$$\boxed{x \sim N(\mu_x, \sigma_x^2) \quad \Rightarrow \quad z = \frac{x - \mu_x}{\sigma_x} \sim N(0,1)}$$

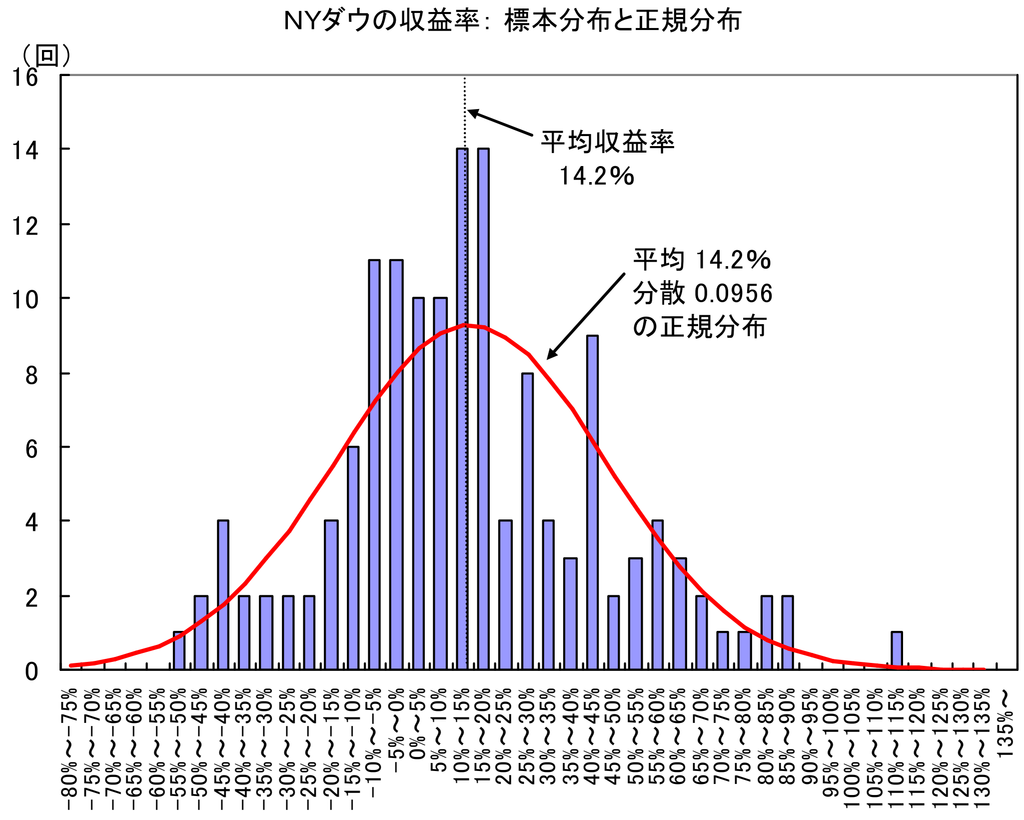
- したがって、xがx^*以上となる確率は、標準正規分布表を利用してzがz^*以上となる確率として求めることができる

$$\mathrm{Prob}(x \geq x^*) = \mathrm{Prob}(z \geq z^*)$$

$$z^* = \frac{x^* - \mu_x}{\sigma_x} \left(\cong \frac{x^* - \bar{x}}{s_x} \right)$$

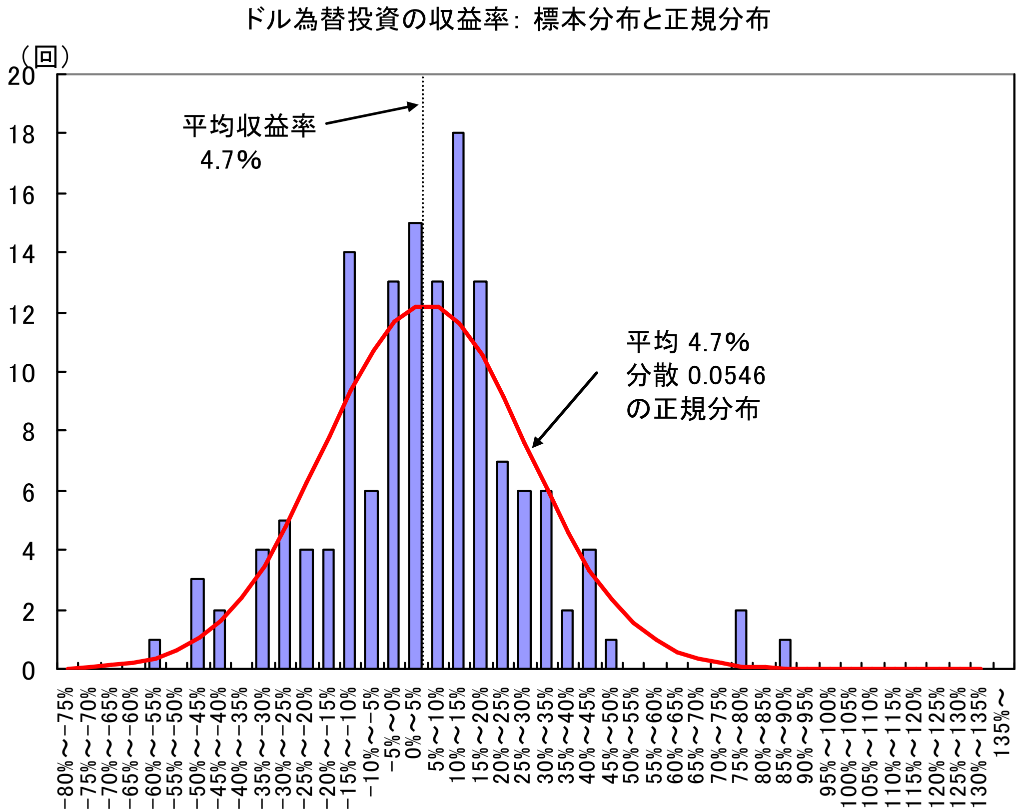
日米株投資の収益率分布と正規分布

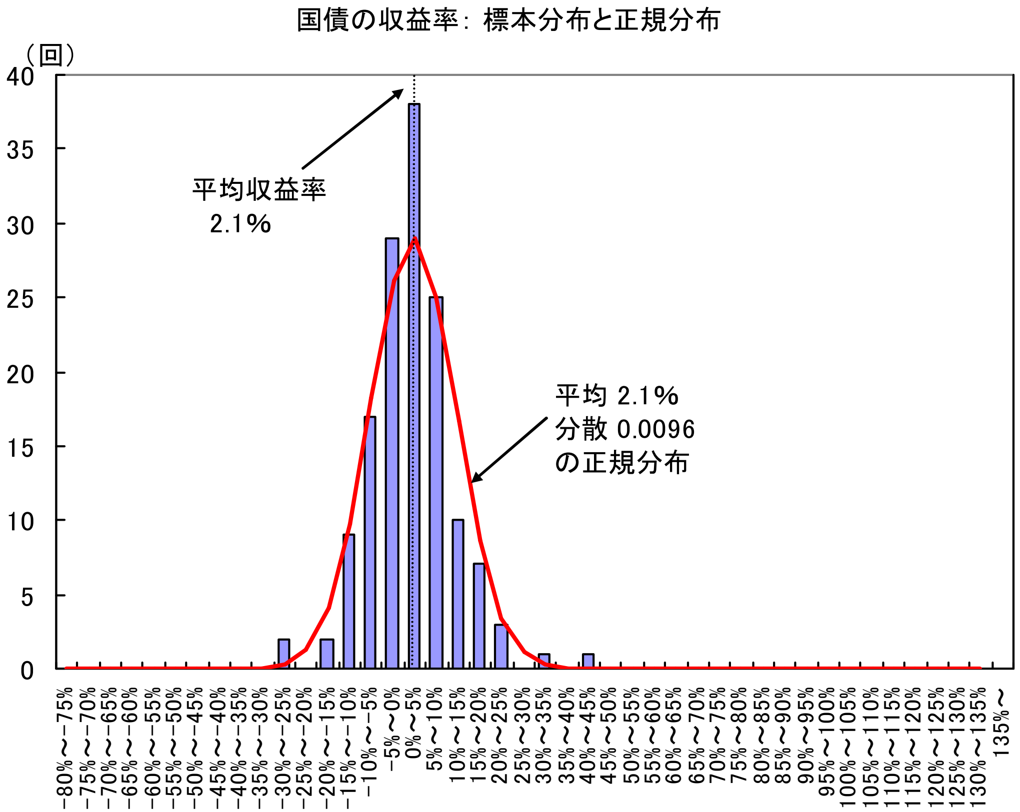




（データ）日経NEEDS、Yahoo Finance

外国為替投資、国債投資の収益率分布と正規分布





(データ)日本銀行『金融経済統計』、日本経済新聞

正規分布の仮定と確率計算(具体例)

	z 値	確率	実際に起こった回数（確率）
日本株が３カ月で年率10%以上上昇			65/144(45.1%)
日本株が３カ月で年率20%以上上昇			55/144(38.2%)
日本株が３カ月で年率20%以上下落			44/144(30.6%)
国債が３カ月で年率15%以上上昇			12/144(8.3%)
国債が３カ月で年率15%以上下落			4/144(2.8%)

（参考）

	z値	確率	実際に起こった回数・確率	
1. ザック・ジャパンが2点以上得点	0.067	0.473	15/33	45.5%
2. ザック・ジャパンが2点以上失点	1.524	0.064	4/33	12.1%
3. ザック・ジャパンが2点差以上で勝利	0.343	0.366	9/33	27.3%
4. 岡田ジャパンが2点以上得点	0.205	0.419	20/50	40.0%
5. 岡田ジャパンが2点差以上で勝利	0.613	0.270	16/50	32.0%

日本株投資の収益率の「標準化」後の分布と標準正規分布

x（日本株の収益率）の平均 $= 8.3\%$

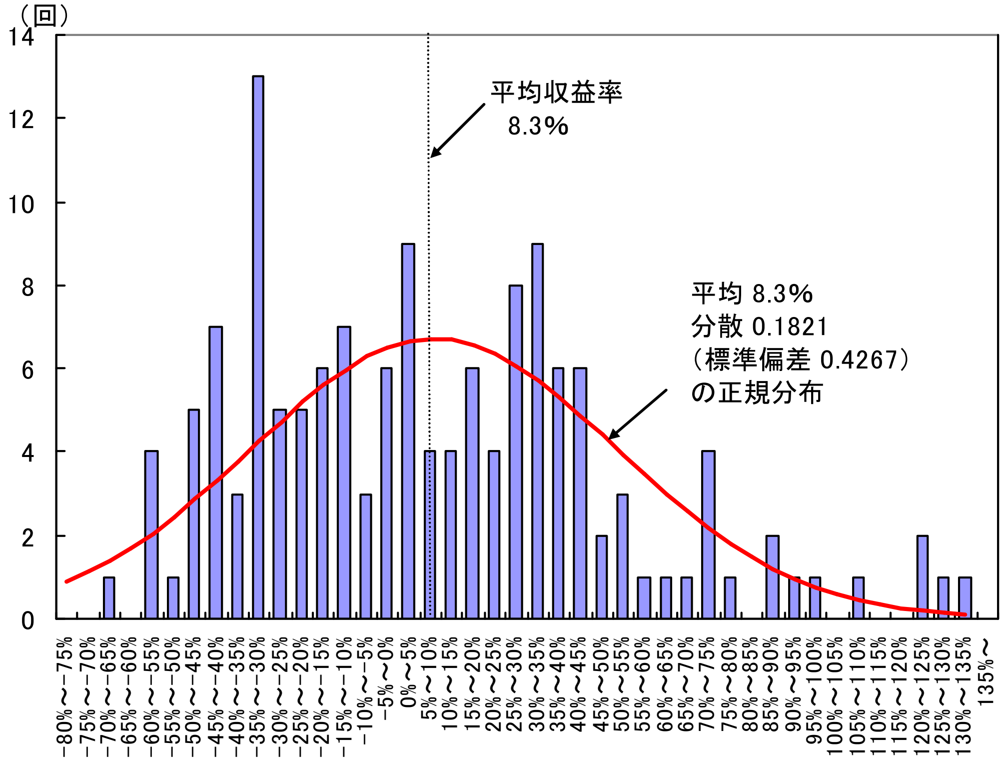
標準化

$$z = \frac{x - 8.3\%}{0.4267}$$

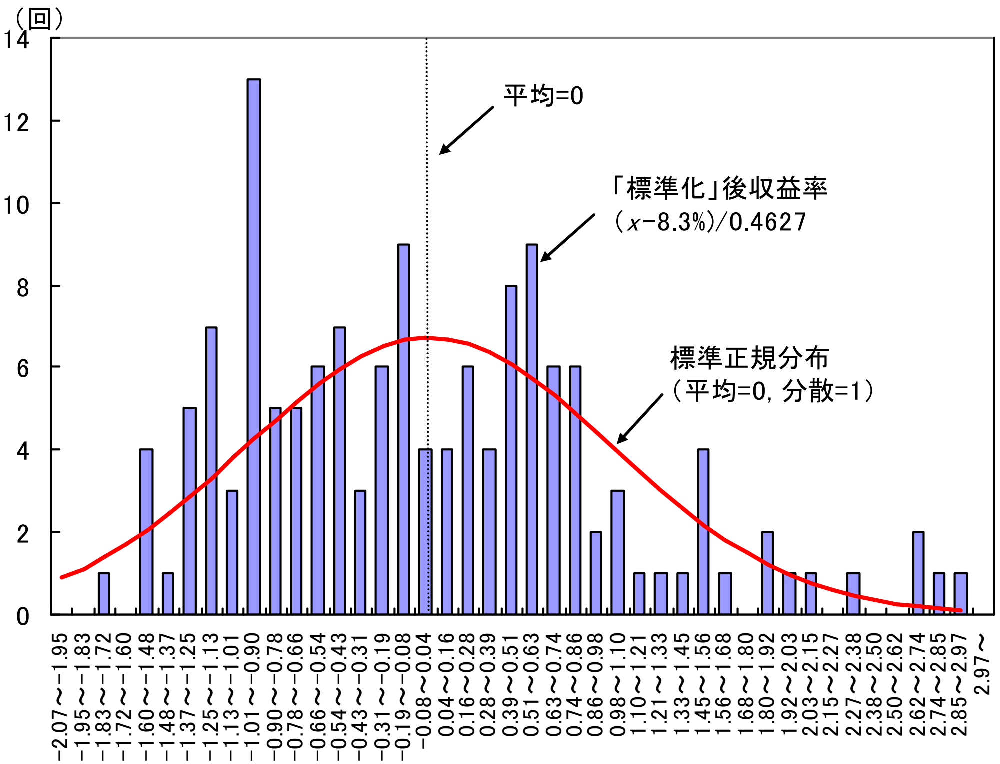
x の標準偏差 $= 0.4267$

$x \rightarrow z$

日経平均の収益率：標本分布と正規分布

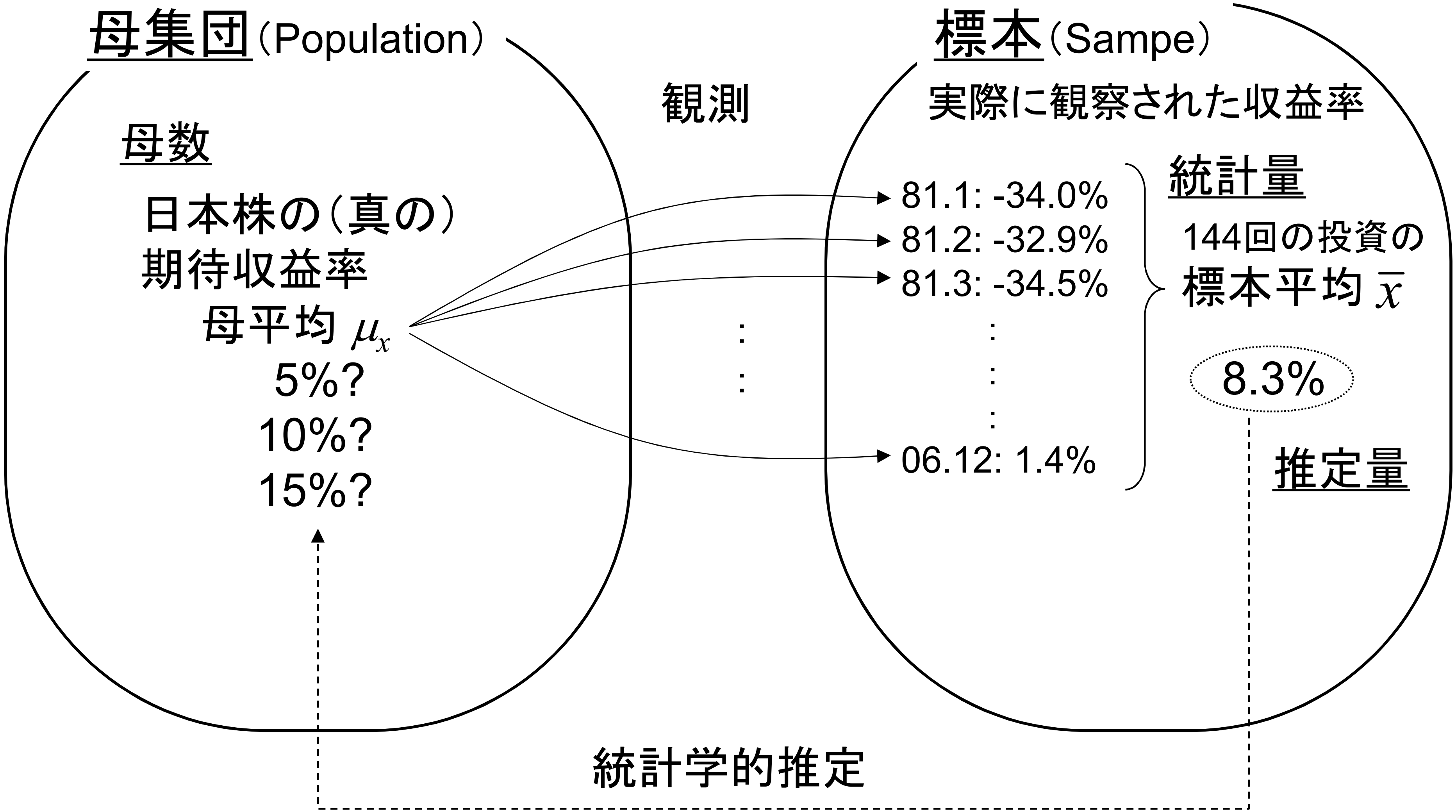


日経平均の「標準化」後収益率と標準正規分布



(データ) 日本銀行『金融経済統計』, IMF International Financial Statistics

統計学的推定：母集団と標本／母数と推定量



①区間推定，②仮説検定

点推定と区間推定

（例1）

- 日本の株式の収益率の標本平均は8.3%･･･点推定
- しかしこれが日本株の真の収益率（母平均）と完全に一致しているとは限らない　⇒ 幅をとって見る必要がある
- 幅をとって「日本株の真の期待収益率は95%の確率で1.2%～15.3%の範囲内にある」と言うことは出来る･･･区間推定

（例2：参考）

- 岡田ジャパンの得点の標本平均は1.70･･･点推定
- しかし岡田ジャパンの真の得点力（母平均）が1.70と完全に等しいとは限らない
- 幅をとって「岡田ジャパンの真の得点力は95%の確率で1.24～2.16の範囲内にあった」と言うことは出来る･･･区間推定

母平均の区間推定(概要)

	日本株式 日経平均	米国株式 NYダウ	国債 10年先物
標本数 n	144	144	144
標本平均 $\bar{x}$	8.3%	14.2%	2.1%
標本平均の分散 s_x^2/n	0.0013	0.0007	0.0001
標本平均の標準誤差 $\sqrt{s_x^2/n}$	0.0356	0.0258	0.0082
標本分散 s_x^2	0.1821	0.0956	0.0096
標本標準偏差 s_x	0.4267	0.3092	0.0979

標本平均
・・・母平均の点推定

標本平均で母平均を推定する場合の標準誤差
(=標本分散/標本数の平方根)
・・・どのくらい信頼区間の幅をとれば良いかの基準となる

- 一定数の標本数がある場合(概ね $n = 20$ 以上)、おおよそ
  - 95%信頼区間 ≒ 標本平均 ± 2×標準誤差

    (例)日経平均の期待収益率の95%信頼区間

    $\fallingdotseq [8.3\% - 2 \times 3.56\%,\ 8.3\% + 2 \times 3.56\%] = [1.1\%,\ 15.4\%]$
  - 90%信頼区間 ≒ 標本平均 ±1.7×標準誤差

    (例)日経平均の期待収益率の90%信頼区間

    $\fallingdotseq [8.3\% - 1.7 \times 3.56\%,\ 8.3\% + 1.7 \times 3.56\%] = [2.2\%,\ 14.3\%]$

95%信頼区間:
標本平均の上下に標準誤差のおよそ2倍の幅をとる

90%信頼区間:
標準誤差のおよそ1.7倍の幅をとる

※ 前後に標準誤差の何倍の幅をとるかは、より正確には(あるいは標本数が20より少ない場合には)t分布表を使用

母平均の区間推定

- 母平均の点推定(標本平均) $=\bar{x}$
- 標本平均の標準誤差 $=\sqrt{s_x^2/n}$ (s_x^2は x の標本分散)
- 標本数$=n$ ・・・ 自由度 $=n-1$

のとき、母平均の信頼区間は

自由度$n-1$のt分布の
2.5%(両側5%)分位点

95%信頼区間 $\boxed{\bar{x}\pm t(n-1,2.5\%)\times\sqrt{s_x^2/n}}$

90%信頼区間 $\boxed{\bar{x}\pm t(n-1,5\%)\times\sqrt{s_x^2/n}}$

自由度$n-1$のt分布の
5%(両側10%)分位点

☆ したがって、信頼区間は、

- 標本 x の分散 s_x^2 が小さいほど
- 標本数 n が大きいほど

狭い範囲に求めることができる(推定精度が高くなる)

資産別期待収益率の区間推定(結果)

【収益率の区間推定】

日経平均	
NYダウ	
ドル	
国債	
定期預金	

【参考：得点力の区間推定】

	点推定	95% 信頼区間
ザック	1.84	1.38~2.30
岡田	1.70	1.24~2.16
オシム	1.80	1.15~2.45
ジーコ	1.58	1.24~1.93
トルシエ	1.65	1.15~2.16

【参考：失点数の区間推定】

	点推定	95% 信頼区間
ザック	1.04	0.72~1.58
岡田	0.82	0.53~1.29
オシム	0.70	0.27~1.13
ジーコ	0.97	0.70~1.51
トルシエ	0.92	0.61~1.45

平均値の検定（例）

- 仮説： 日本株式の真の収益率は10％
  - 収益率の平均（実績）は8.3％・・・本来10％の実力があるのにたまたま実力より低い収益率となることが多かったのか、もともとの実力が10％もないからなのか？
- 仮説： 為替変動の方向に偏った傾向はない（＝為替変動の期待値は0％）
  - 実績は平均で年率4.7％のドル高・・・本来はドル高にもドル安にもなり得たのに、たまたまドル高となることが多かったのか、それともこの期間はドル高傾向となる要因が存在したのか？
- （参考）仮説： 岡田ジャパンの得点力は2点以上
  - 実績は1試合平均1.70・・・本来は2点とる実力があったのに、たまたま調子の悪い試合が続いたのか、もともと2点とるまでの実力はなかったのか？

平均値の検定（t検定）

- 標本 $(x_1, x_2, \ldots x_n)$ から母平均がμ_0に等しいかどうかを検定

(1) 仮説を立てる

帰無仮説 H_0: $\mu_x = \mu_0$　　［ v.s.　対立仮説 H_1: $\mu_x \neq \mu_0$ ］

(2) 検定統計量（t値）と自由度を計算する

$$t_0 = \frac{\bar{x} - \mu_0}{\sqrt{s_x^2 / n}} \quad \left(= \frac{\text{標本平均} - \text{仮説値}}{\text{標準誤差}} \right) \quad : \quad \text{自由度} = n - 1$$

(3) 検定統計量（t値）をt分布表と比較する

- $t_0 > t(n-1, 2.5\%)$ → H_0を5%有意水準で棄却（＝H_1を支持）
  ・・・H_0は誤り（H_1が正しい）と考える
- $t_0 \leq t(n-1, 2.5\%)$ → H_0は5%有意水準で棄却されない（ H_0を受容）
  ・・・H_0は誤りとは言えない（正しいとも言えない）

※良く用いられる有意水準は、5%のほか、10%, 1%など（それぞれt分布表の対応する値（ $t(n-1, 5\%)$, $t(n-1, 0.5\%)$ ）を使用

平均値の検定（t検定）：実例①

- 日本株の期待収益率が10%か否かを検定

(1) 仮説を立てる

帰無仮説 H_0: 日本株の収益率の母平均は10%

［対立仮説 H_1: 日本株の収益率の母平均は10% ではない］

(2) 検定統計量（t値）と自由度を計算する

$$t_0 = \frac{\overbrace{8.3\%}^{\text{標本平均}} - \overbrace{10\%}^{\text{仮説値}}}{\underbrace{0.0356}_{\text{標準誤差}}} = 0.491 \quad : \quad \text{自由度} = \overbrace{144}^{\text{標本数}} - 1 = 143$$

(3) 検定統計量（t値）をt分布表と比較する

- 自由度143のt分布（※表にないので自由度100で代用）の片側2.5%（両側5%）分位点は1.984 > 0.491
  ⇒ 帰無仮説H_0は5%の有意水準で棄却されない
  ⇒ 日本株の期待収益率は10%と考えても間違いとは言えない（正しいとも言えない

平均値の検定

母集団(Population)

母数

日本株の(真の)
期待収益率
母平均 μ_x

(1) 仮説

H_0: μ_x=10%?
or H_1: $\mu_x \neq$10%?

(3)判定

確率低い
→ H_0×, H_1○
確率高い
→ H_0?, H_1?

観測

標本(Sampe)

実際に観察された収益率

81.1: -34.0%
81.2: -32.9%
81.3: -34.5%
:
:
:
06.12: 1.4%

統計量

144回の投資の
標本平均 $\bar{x}$

8.3%

推定量

(2) 検定統計量

$$t_0 = \frac{8.3\% - 10\%}{0.0356}$$
$$= 0.491$$

統計学的推定〔仮説検定〕

仮説H_0が正しい(真の期待収益率が10%の)ときに
144回投資してこのような結果(標本平均が8.3%)が出る確率は?
(=t値が0.491となる確率は?)

（参考）平均値の検定（t検定）：実例②

- 岡田ジャパンの得点力が1点か否かを検定

(1) 仮説を立てる

帰無仮説 H_0: 岡田ジャパンの得点力は1試合1点程度

［対立仮説 H_1: 岡田ジャパンの得点力は1試合1点程度ではない］

(2) 検定統計量（t値）と自由度を計算する

$$\boxed{t_0 = \frac{\overset{\text{標本平均}}{1.76} - \overset{\text{仮説値}}{1}}{\underset{\text{標準誤差}}{0.2431}} = 3.129 \ : \quad \text{自由度} = \overset{\text{標本数}}{46} - 1 = 45}$$

(3) 検定統計量（t値）をt分布表と比較する

- 自由度44のt分布の片側2.5%（両側5%）分位点は2.014 < 3.129
  ⇒ 帰無仮説H_0は5%の有意水準で棄却される（H_1が支持される）
  ⇒ 岡田ジャパンの得点力は平均して1試合1点程度しか期待できないと考えるのは誤り（1試合1点よりも期待できる）

✍仮説検定の一般的な手順

（1） 仮説を立てる

帰無仮説 H_0（Null Hypothesis）:

通常は、棄却されることが想定されている（例：日本株の収益率は10%）

対立仮説 H_1（Alternative　Hypothesis）:

帰無仮説と対立する仮説（例：日本株の収益率は10% ではない）

帰無仮説を棄却することによって、これを間接的に証明する

（2） 帰無仮説の下での検定統計量（t値など）を計算する

（3） 検定統計量を用いて、得られた標本が帰無仮説の下で生じる確率を求める

- 確率が低い（1%以下, 5%以下,10%以下など）

  ⇒ 帰無仮説 H_0を棄却 ＝ 対立仮説 H_1を支持

  （例：日本株の収益率が10%である確率は5%以下＝95%の確率で「10%ではない」と言える）

- 確率が高い（5%以上,10%以上など）

  ⇒ 帰無仮説 H_0を受容（＝H_0、H_1 とも正しいとも誤りとも言えない）

  （例：日本株の収益率が10%であってもおかしくはないが、確かなことは言えない）

平均値の検定(結果)

仮説	標本平均 $\bar{x}$	仮説値 μ_0	標準誤差 $\sqrt{s_x^2/n}$	t値	自由度 $n-1$	p値	有意水準5%	有意水準10%
1. 日本株の収益率＝10%								
2. 日本株の収益率＝15%								
3. 日本株の収益率＝0%								
4. ドルの収益率＝0%								
5. 日本国債の収益率＝5%								

【参考】

仮説	標本平均 $\bar{x}$	仮説値 μ_0	標準誤差 $\sqrt{s_x^2/n}$	t値	自由度 $n-1$	p値	有意水準 5%	有意水準 10%
A. 岡田Jの得点力＝2得点	1.70	2	0.2308	−1.300	49	0.200	受容	受容
B. 岡田Jの得点力＝1得点	1.70	1	0.2308	3.033	49	0.004	棄却	棄却
C. ジーコJの得点力＝2得点	1.58	2	0.1722	−2.420	71	0.018	棄却	棄却
D. ジーコJの得点力＝1得点	1.58	1	0.1722	3.388	71	0.001	棄却	棄却
E. ジーコJの実力＝1点差勝利	0.61	1	0.2154	−1.806	71	0.075	受容	棄却

P値の意味

- P値＝帰無仮設H_0が正しいとしたときに、n個の標本を得たときの平均が $\bar{x}$ になる確率
  - P値（確率）が低い
    ＝H_0が正しいとしたら、そんなこと（n個の標本の平均が$\bar{x}$になる）は、「ほとんど起こり得ない」→ H_0は正しくない・・・H_0を棄却、H_1を支持
  - P値（確率）が高い
    ＝ H_0が正しいとしても、そういうことは、「十分起こり得る」→ H_0は正しくないとは言い切れない（絶対に正しいとも言えない）・・・H_0を受容

	確率（P値）が低い	確率（P値）が高い
帰無仮説H_0	×（棄却）	？（受容）
対立仮説H_1	○（支持）	？

P値の意味（例）

- 日本株の真の収益率が10%（H_0）のときに、144回投資して、その平均が8.3%である確率（P値）は、0.624（62.4%）

  ＝ 十分起こりえる

  → 日本株の収益率は10%と考えてもおかしいとは言えない（H_0受容）

- 日本株の真の収益率が15%（H_0）のときに、144回投資して、その平均が8.3%である確率（P値）は、0.060（6.0%）

  ＝微妙？（十分起こりえる？ほとんど起こりえない？）・・・判断基準をどこにおくか

  - 確率が5%以下のとき「起こり得ない」と判断するなら（有意水準5%）

    ・・・「起こり得る」→真の収益率15%と考えてもおかしくない（H_0受容）

  - 確率が10%以下のとき「起こり得ない」と判断するなら（有意水準10%）

    ・・・「起こり得ない」→真の収益率15%は正しくない（H_0棄却、H_1支持）

- 日本国債の真の収益率が5%（H_0）のときに、144回投資して、その平均が2.1%である確率（P値）は、0.000（0.0%）

  ＝ ほとんど起こりえない

  → 日本国債の収益率が5%と考えるのは正しくない（ H_0棄却、H_1支持）

✍標本平均の分布

- 正規分布
  - □ xが正規分布する変数である場合、その標本平均 $\bar{x}=\sum_{i=1}^{n}x_i\Big/n$ も正規分布する

標本平均の期待値は
母平均に等しい（不偏性）

$$x \sim iid\ N(\mu_x,\sigma_x^2) \quad \Rightarrow \quad \bar{x} \sim N(\mu_x,\frac{\sigma_x^2}{n})$$

平均をとると分散は
（標本数が多いほど）縮小

- 中心極限定理
  - □ xが正規分布する変数でない場合でも、標本数が十分大きければ、標本平均 $\bar{x}$ は正規分布する

$$x \sim iid\ (\mu_x,\sigma_x^2),\ n \to \infty \quad \Rightarrow \quad \bar{x} \to N(\mu_x,\frac{\sigma_x^2}{n})$$

✍検定統計量の分布

- 標本平均 $\bar{x}$ が正規分布するとき

  ①x が正規分布するとき
  ②(x が正規分布しなくとも)標本数が十分大きいとき

$$\bar{x} \sim N(\mu_x, \frac{\sigma_x^2}{n})$$

- $\bar{x}$ を標準化した z は標準正規分布する

$$\Rightarrow z = \frac{\bar{x} - \mu_x}{\sqrt{\sigma_x^2 / n}} \sim N(0,1)$$

- xの母標準偏差σ_x(＝未知)を標本標準誤差s_x(＝既知)で置き換えた t は自由度$n-1$の t 分布に従う

$$\boxed{\Rightarrow t = \frac{\bar{x} - \mu_x}{\sqrt{s_x^2 / n}} \sim t(n-1)}$$

平均値のt検定が有効な場合

	標本が少ないとき	標本が多いとき
正規分布すると考えられる変数の場合	○	○（※）
正規分布しないと考えられる変数の場合	×	○（※）

（※）z検定（標準正規分布表を用いた検定）も可
（標本が多い時はz検定とt検定は実質的に同じ）

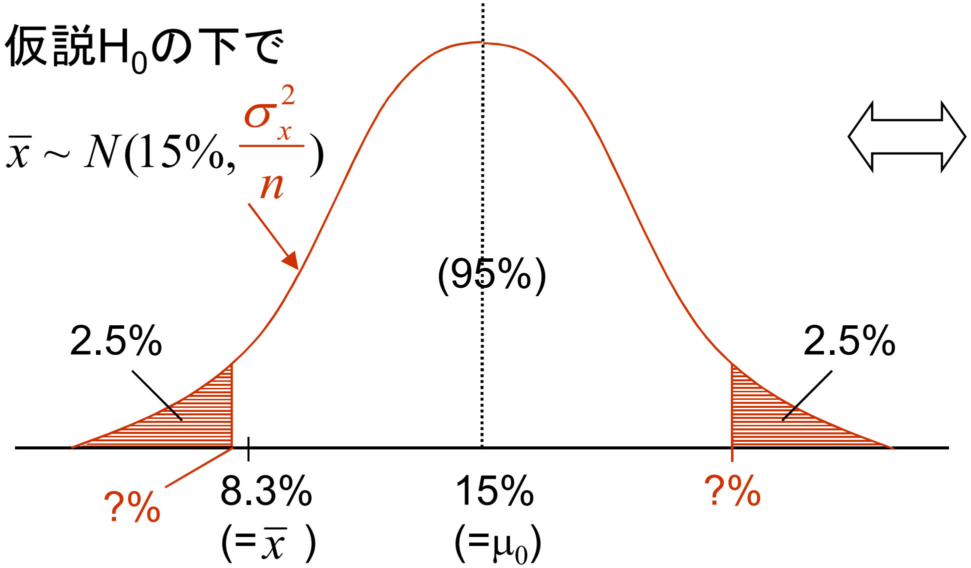
✍平均値の検定の考え方①

仮説H_0: μ_0=15%
(日本株の収益率の母平均=15%)
有意水準: 両側5%

標本平均 $\bar{x}$ が<u>正規分布</u>するとき
(x 正規分布 or 大標本のとき)

仮説H_0の下で

$\bar{x} \sim N(15\%, \frac{\sigma_x^2}{n})$

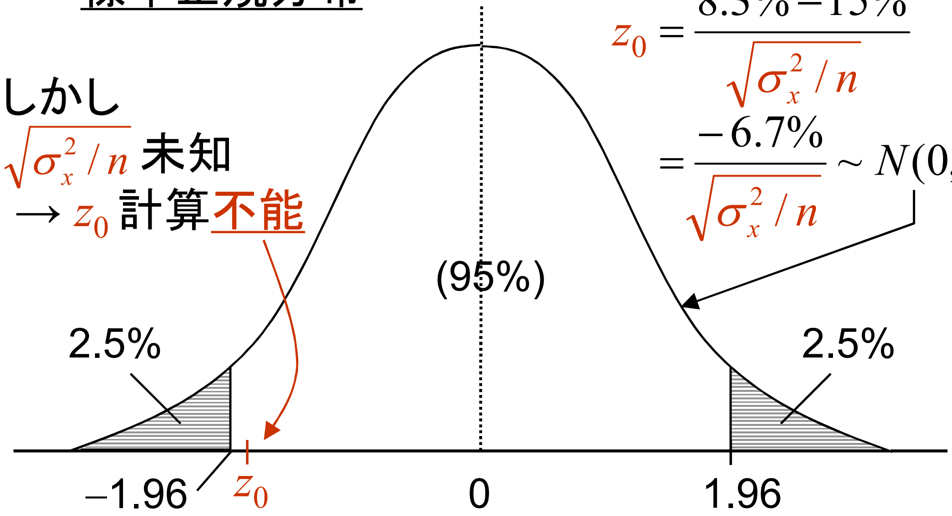


「標準化」すれば
<u>標準正規分布</u>

しかし
$\sqrt{\sigma_x^2/n}$ 未知
→ z_0 計算<u>不能</u>

仮説H_0の下で

$$z_0 = \frac{8.3\% - 15\%}{\sqrt{\sigma_x^2/n}} = \frac{-6.7\%}{\sqrt{\sigma_x^2/n}} \sim N(0,1)$$



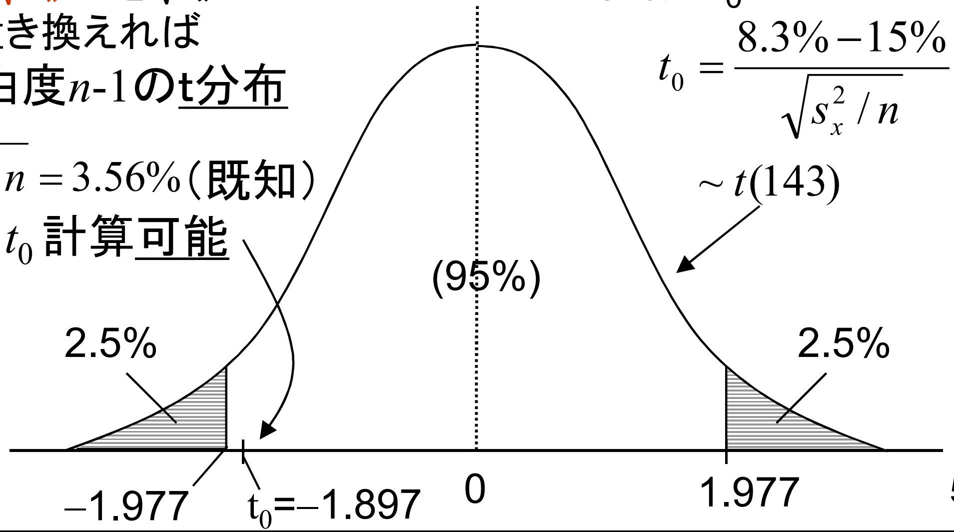
Z_0の$\sqrt{\sigma_x^2/n}$を$\sqrt{s_x^2/n}$
で置き換えれば
自由度n-1の<u>t分布</u>

$\sqrt{s_x^2/n} = 3.56\%$(既知)
→ t_0 計算<u>可能</u>

仮説H_0の下で

$$t_0 = \frac{8.3\% - 15\%}{\sqrt{s_x^2/n}}$$

$\sim t(143)$



✕ $|t_0| \geq 1.977$(自由度143のt分布の両側5%(片側2.5%)の境界値)のとき
=t_0が棄却域(横線部)にある
=8.3%が棄却域(横線部)にある
=真の収益率が15%なのに144個の観測値の平均が8.3%になることはほとんどない(確率5%以下)
⇒H_0を棄却(真の収益率は15%ではない)

◎ $|t_0| < 1.977$のとき
=t_0が棄却域(横線部)にない
=8.3%が棄却域(横線部)にない
=真の収益率が15%でも144個の観測値の平均が8.3%になることは十分あり得る(確率5%以上)
⇒H_0を受容(真の収益率は15%かもしれない)

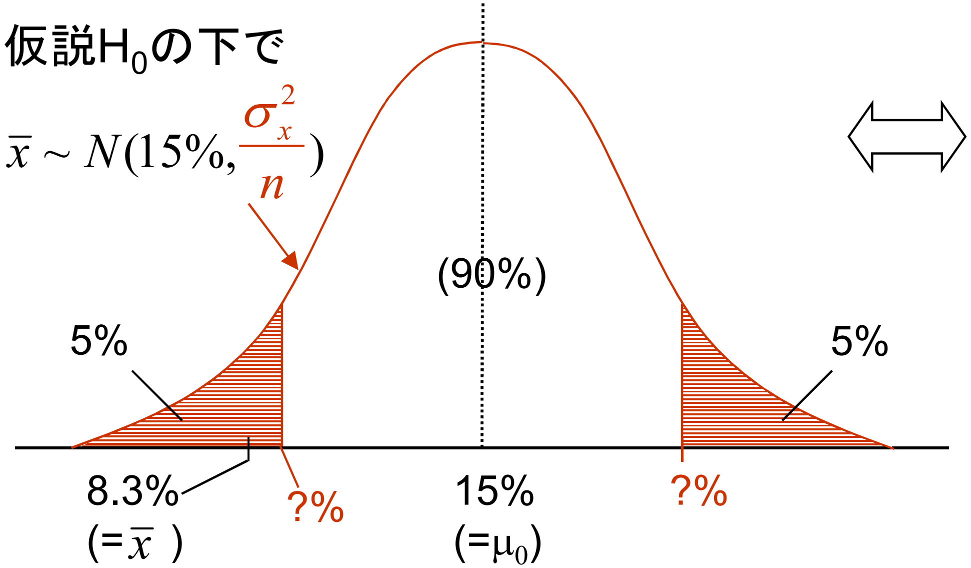
✍平均値の検定の考え方②

仮説H_0: μ_0=15%
（日本株の収益率の母平均=15%）
有意水準：両側10%

標本平均 $\bar{x}$ が正規分布するとき
（x 正規分布 or 大標本のとき）

仮説H_0の下で

$$\bar{x} \sim N(15\%, \frac{\sigma_x^2}{n})$$



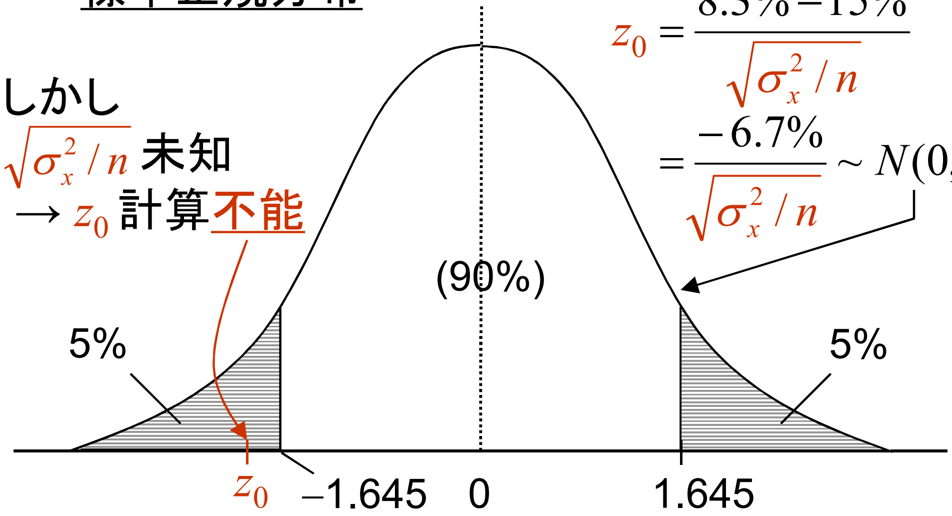
⇔

「標準化」すれば
標準正規分布

仮説H_0の下で

$$z_0 = \frac{8.3\% - 15\%}{\sqrt{\sigma_x^2 / n}} = \frac{-6.7\%}{\sqrt{\sigma_x^2 / n}} \sim N(0,1)$$

しかし
$\sqrt{\sigma_x^2 / n}$ 未知
→ z_0 計算不能



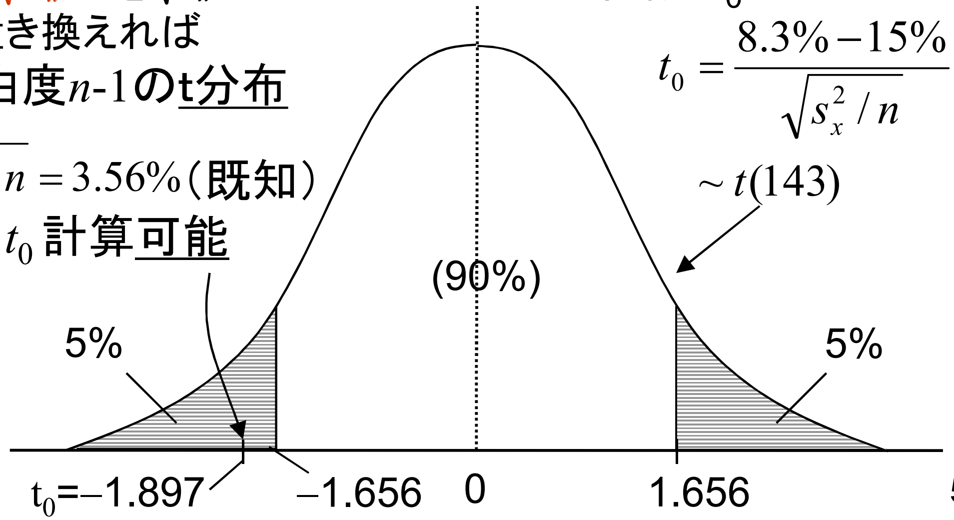
⇕

Z_0の$\sqrt{\sigma_x^2 / n}$を$\sqrt{s_x^2 / n}$
で置き換えれば
自由度n-1のt分布

仮説H_0の下で

$$t_0 = \frac{8.3\% - 15\%}{\sqrt{s_x^2 / n}} \sim t(143)$$

$\sqrt{s_x^2 / n} = 3.56\%$（既知）
→ t_0 計算可能



⦿ $|t_0| \geq 1.656$（自由度143のt分布の両側10%（片側5%）の境界値）のとき
=t_0が棄却域（横線部）にある
=8.3%が棄却域（横線部）にある
=真の収益率が15%なのに144個の観測値の平均が8.3%になることはほとんどない（確率10%以下）
⇒H_0を棄却（真の収益率は15%ではない）

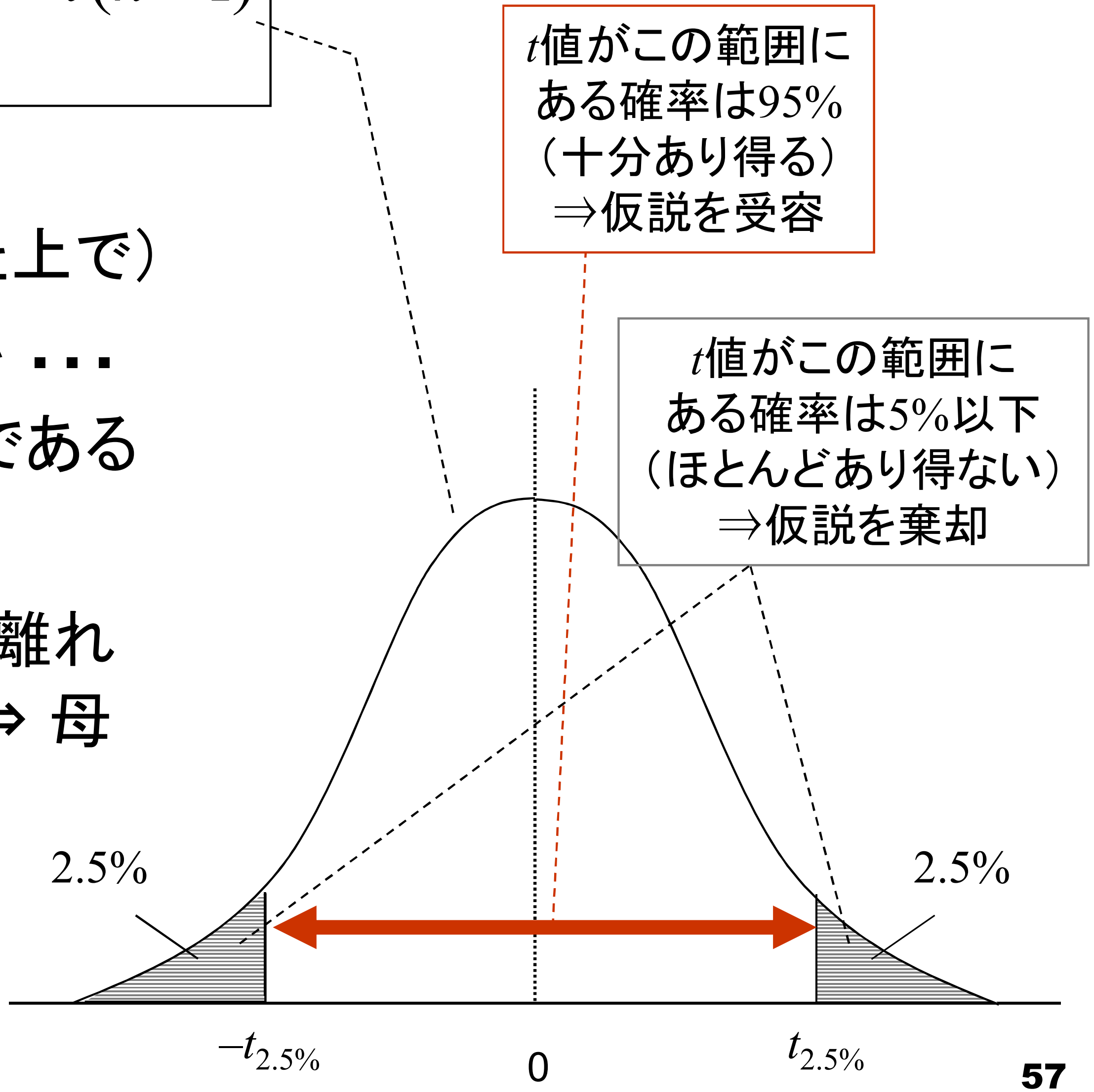
✕ $|t_0| < 1.656$のとき
=t_0が棄却域（横線部）にない
=8.3%が棄却域（横線部）にない
=真の収益率が15%でも144個の観測値の平均が8.3%になることは十分あり得る（確率10%以上）
⇒H_0を受容（真の収益率は15%かもしれない）

✍ t検定の考え方（直観）

検定統計量 $$t_0 = \frac{\bar{x} - \mu_0}{\sqrt{s_x^2 / n}} \sim t(n-1)$$

（$\bar{x}$の標準誤差$\sqrt{s_x^2 / n}$を考慮した上で）

- 標本の平均$\bar{x}$がμ_0に近いとき・・・ t値は小さい ⇔ 母平均がμ_0である確率は高い⇒仮説を受容
- 標本の平均 $\bar{x}$がμ_0から大きく離れているとき・・・ t値は大きい ⇔ 母平均がμ_0である確率は小さい ⇒ 仮説を棄却



区間推定と仮説検定の関係

☆ 区間推定と仮説検定は表裏の関係

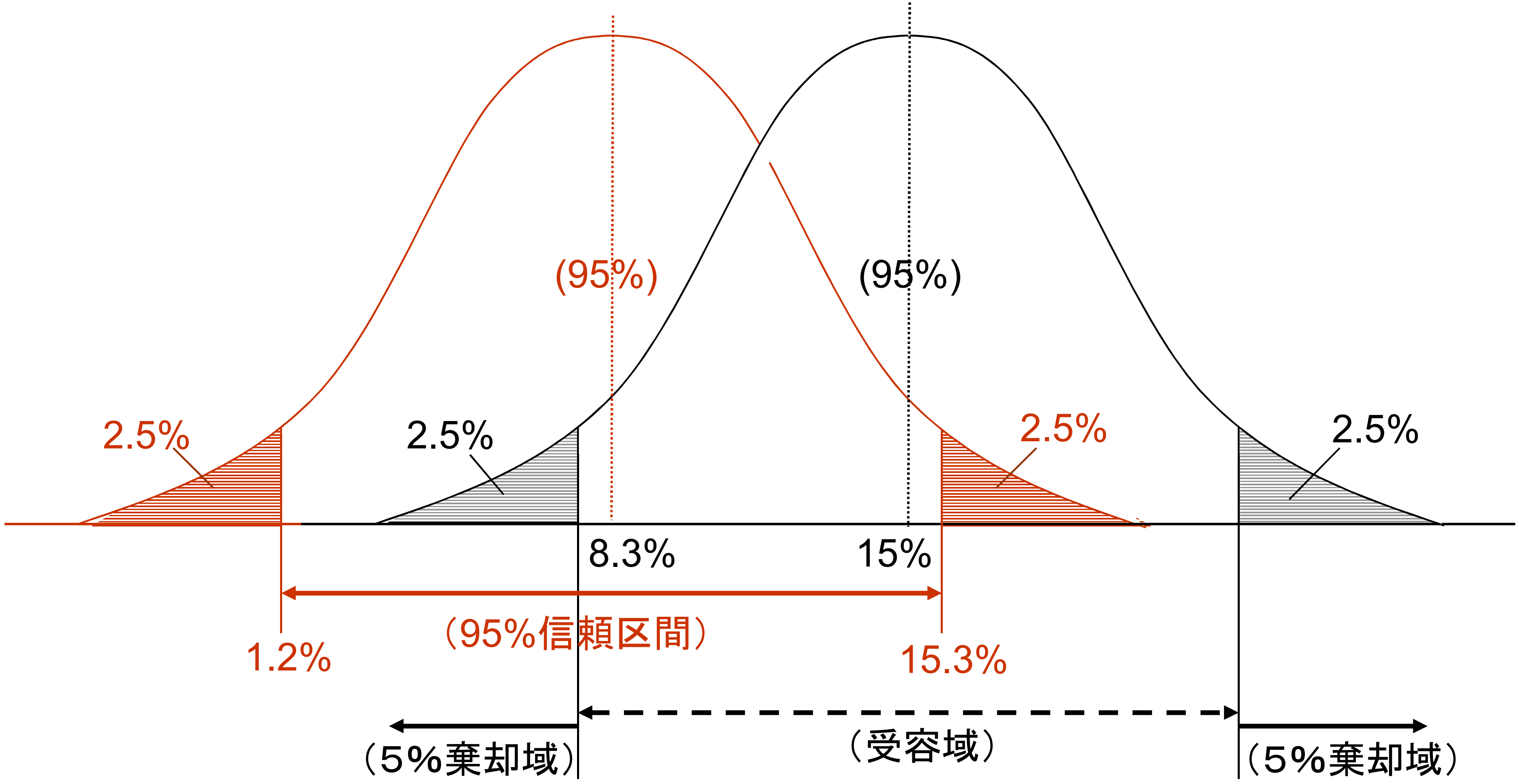
- 株式の収益率の95%信頼区間(1.2~15.3%)に15%が含まれる
  ⇔「株式の収益率は15%」という仮説は5%有意水準で棄却されない
  (株式の収益率が15%である確率は5%以上はある)

- 株式の収益率の90%信頼区間(2.4~14.1%)に15%が含まれない
  ⇔「株式の収益率は15%」という仮説は10%有意水準で棄却される
  (株式の収益率が15%である確率は10%以下)

- 国債の収益率の95%信頼区間(0.5~3.7%)に5%が含まれない
  ⇔「国債の収益率は5%」という仮説は5%有意水準で棄却される
  (国債の収益率が5%である確率は5%以下)

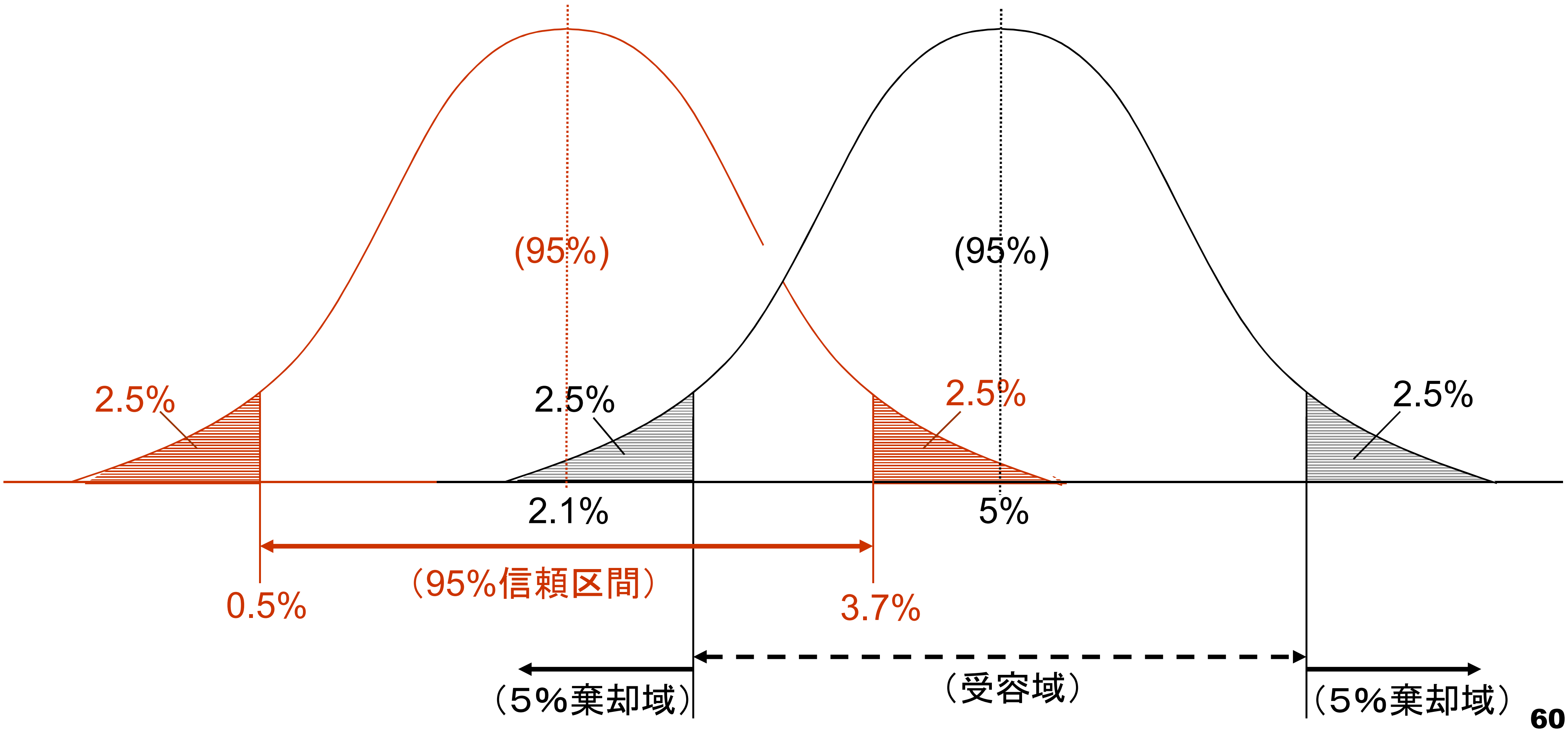
✍区間推定と仮説検定の関係（図示①）

- 「株式の真の収益率」の95%信頼区間（1.2~15.3%）に15%が含まれる
⇔「真の収益率は15%」という仮説（H_0）の下で、実際に観測された標本平均値（8.3%）は、5%有意水準の棄却域に含まれない（受容域にある）



✍区間推定と仮説検定の関係(図示②)

- 「国債の真の収益率」の95%信頼区間(0.5~3.7%)に5%が含まれない
⇔「国債の収益率は5%」という仮説(H_0)の下で、実際に観測された標本平均値(2.1%)は、5%有意水準の棄却域にある



平均の差の検定（例）

- 仮説： 日本株と米国株の収益力に差はない
  - 95年以降のデータで日本株の方が平均収益率が低いのは、たまたまなのか、もともと日本株（日本企業）の方が実力が劣っていたからなのか？
- 仮説： 株式と国債のリターンに差はない
  - 株式の方が平均収益率が高いのはたまたまなのか、それとの本来的に株式の方がハイ・リターンの金融商品であるのか

（参考）

- 仮説： 岡田ジャパンとジーコ・ジャパンの得点力は同等
  - 実績は岡田J＝1.70、ジーコJ＝1.58。ジーコ時代よりも代表の得点力は高まったのか？チームの調子や相手の出来によってたまたま良い成績を得ただけなのか？

平均差の検定

- ２つの変数の系列$x_i=(x_1, x_2, x_3, \ldots x_n)$ と $y_i=(y_1, y_2, y_3, \ldots y_m)$ の平均が等しいかどうかを検定

  ⇒　$\bar{x}-\bar{y}$　がゼロかどうかを検定

$$\boxed{t=\frac{\bar{x}-\bar{y}}{\sqrt{s^2}} \sim t(df)}$$

ただし、s^2（$\bar{x}-\bar{y}$ の分散の推定値）、df（自由度）は、以下により計算

- xとyの分散が等しいと考えられる場合

$$s^2=\frac{(n-1)s_x^2+(m-1)s_y^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right),\quad df=n+m-2$$

- xとyの分散が等しくないと考えられる場合

$$s^2=\frac{s_x^2}{n}+\frac{s_y^2}{m},\quad df=\frac{(s_x^2/n+s_y^2/m)^2}{(s_x^2/n)^2/(n-1)+(s_y^2/m)^2/(m-1)}$$

平均差の検定

母集団(Population)

日本株の期待収益率
〔母平均〕 μ_{Japan}

米国株の期待収益率
〔母平均〕 $\mu_{\text{U.S.}}$

(1) 仮説

H_0: $\mu_{\text{Japan}} = \mu_{\text{U.S.}}$?

or H_1: $\mu_{\text{Japan}} \neq \mu_{\text{U.S.}}$?

標本(Sampe)

実際に観察された収益率

95.1: -34.0%
95.2: -32.9%
:
:

144回平均
8.3%

95.1: 59.7%
95.2: 53.6%
:
:

144回平均
14.2%

(2) 検定統計量

$$t_0 = \frac{8.3\% - 13.6\%}{\sqrt{s^2}} = -1.346$$

統計学的検定〔平均差の検定〕

(3)判定

確率低い
→ H_0×, H_1○
確率高い
→ H_0?, H_1?

仮説H_0が正しい(日米株の期待収益率が等しい)ときに
144回投資してこのような結果(日米の収益率の標本平均が8.3%と14.2%)
が出る確率は？(＝t値が-1.346となる確率は？)

平均差の検定（結果）

（仮説） 期待収益率について:	平均1 $\bar{x}$	平均2 $\bar{y}$	分母 $\sqrt{s^2}$	t値	自由度 df	p値	有意水準 5%	有意水準 10%
6. 日本株＝米株								
7. ドル＝ユーロ								
8. 株式＝国債								
9. 株式＝定期預金								
10. 国債＝定期預金								

※ 非等分散の2系列の平均差の検定結果

【参考】

仮説	平均1 $\bar{x}$	平均2 $\bar{y}$	分母 $\sqrt{s^2}$	t値	自由度 df	p値	有意水準 5%	有意水準 10%
F. 得点力に関して、岡田J＝ジーコJ	1.70	1.58	0.2879	0.405	98	0.686	受容	受容
G. 守備力に関して、岡田J＝ジーコJ	0.82	0.97	0.1963	-0.775	114	0.440	受容	受容
H. 得失点差に関して、岡田J＝ジーコJ	0.88	0.61	0.3608	0.745	98	0.458	受容	受容
I. 岡田Jは、平均得点数＝平均失点数	1.70	0.82	0.2710	3.248	81	0.002	棄却	棄却
J. 得失点差に関して、岡田J＝オシムJ	0.88	1.10	0.4219	-0.521	52	0.604	受容	受容

※ 非等分散の2系列の平均差の検定結果

平均差の検定（結果②）Excel 分析ツールによる結果

- 収益率の差の検定
  日経平均 vs NYダウ

t-検定：分散が等しくないと仮定した2標本による検定

	日経平均	NYダウ
平均	8.3%	14.2%
分散	0.1821	0.0956
観測数	144	144
仮説平均との差異	0	
自由度	261	
t	-1.346	
P(T<=t) 片側	0.090	
t 境界値 片側	1.651	
P(T<=t) 両側	0.179	
t 境界値 両側	1.969	

有意水準10%でも棄却されない
＝収益率に差があったとは言えない

- 収益率の差の検定
  日経平均 vs 国債

t-検定：分散が等しくないと仮定した2標本による検定

	日経平均	国債
平均	8.3%	2.1%
分散	0.1821	0.0096
観測数	144	144
仮説平均との差異	0	
自由度	158	
t	1.696	
P(T<=t) 片側	0.046	
t 境界値 片側	1.655	
P(T<=t) 両側	0.092	
t 境界値 両側	1.975	

有意水準10%で棄却（5%で受容）
＝10%有意水準で差があったと言える
（有意水準5%では言えない）

主な分析結果（まとめ）

【基本統計量】

- 投資の「リターン」を収益率の<u>平均</u>、「リスク」を収益率の<u>分散</u>で見ると、
  - 株式投資はハイリスク（収益率の分散（ばらつき）が大）・ハイリターン（収益率の平均が大）
  - 国債投資はローリスク（収益率の分散（ばらつき）が小）・ローリターン（平均収益率が小）
  - 為替投資は、ミドルリスク・ミドルリターン

【相関】

- 株式と他の資産の収益率との相関をみると、
  - 株式と国債は負に相関
  - 日本株と米国株は正に相関
  - 株式と為替の相関は薄い（ドルはやや正、ユーロはやや負に相関）

主な分析結果（まとめ②）

【区間推定】

- 日本の株式の（真の）収益率の95%信頼区間は、1.2%~15.3%
- 国債投資の（真の）収益率の95%信頼区間は、0.5~3.7%

【平均値の検定】

- 「株式に投資すれば年率15%の収益が期待できる（期待収益率=15%）」という仮説は、5%有意水準では受容、10%有意水準では棄却される
- 「株式に投資しても儲けは期待できない（期待収益率=0%）」という仮説は、5%有意水準で棄却される

【平均差の検定】

- 「日本株と米国株の収益率に差はない」という仮説は、10%有意水準でも棄却できない（はっきりと差があるとは言えない）
- 「株式と国債の収益率に差はない」という仮説は、10%有意水準で棄却される（株式の方がハイリターンだと言える）