



# TA報告(4) 実証会計学 「統計学の基礎知識」

藤井ゼミ サブゼミ

5/17(Fri) @106演習室

京都大学大学院 経済学研究科

博士後期課程 1回生

渡邊 誠士

# 報告内容

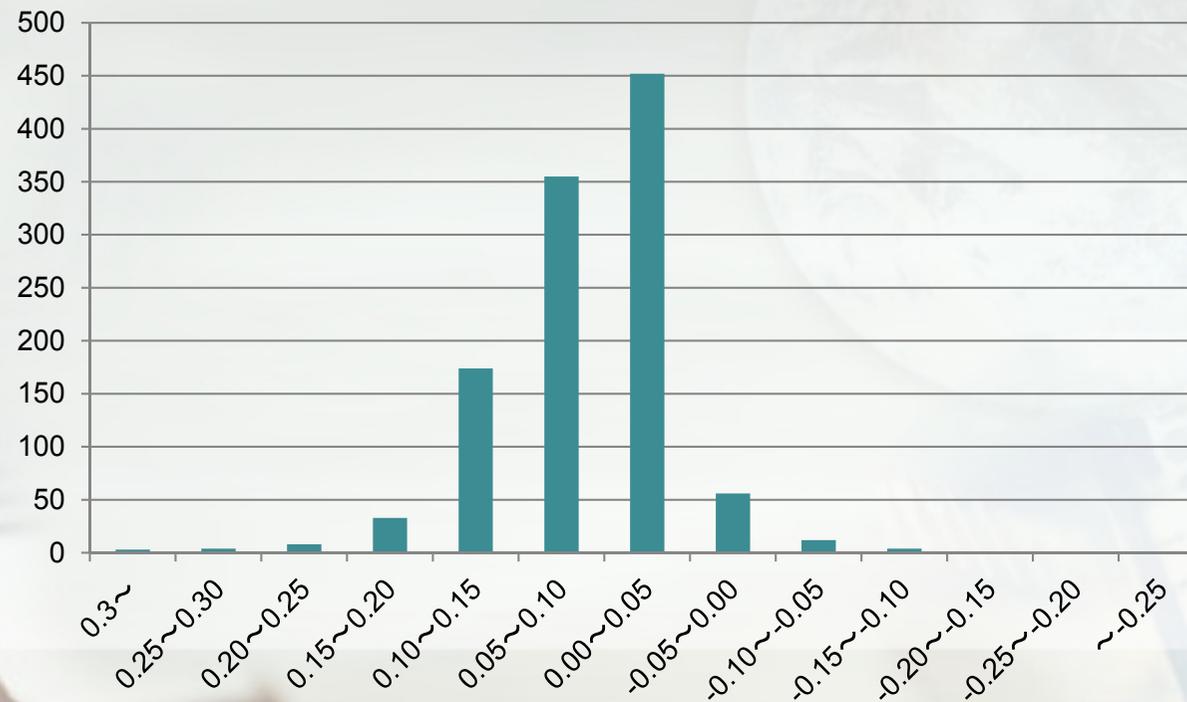
1. 母集団と標本
2. 分布
3. 記述統計量
4. 外れ値の除去
5. 仮説検定とは
6. 信頼区間
7. 有意水準
8. 推定量
9. 平均値の検定
10. Excelによる平均値の検定
11. Eviewsによる平均値の検定

# 母集団と標本

- 「製造業の収益性」について調査したいとする。
  - しかし、世界中すべての製造業(母集団)を実際に調査することは時間の問題、データ制約の問題上難しい(不可能)。
  - そこで、(本来であれば)世界中の製造業から無作為に一定数の企業をサンプル(標本)としてデータ収集する必要がある。
  - しかし、実際にはデータ制約の関係より、日本の東証一部上場企業などをサンプルとする場合が多い。
- 今回は、製造業の収益性を調べるため、2008年度東証1部上場企業をサンプルとして話を進める。

# 分布①

- 「製造業企業の収益性」を操作可能化し、「東証1部上場製造業の2008年度のROA」について調べることとする。
  - サンプルサイズ:1056社
- ヒストグラムを作成



# 記述統計量

- 変数の分布の特徴を表す, 代表的な(要約された)値。
  - 観測数(Observation)
  - 最大値(Max)
  - 最小値(Min)
  - 四分位数(Quantile)
    - 第一四分位数(Q1): 小さい方から25%目の数
    - 中央値(Median)
    - 第三四分位数(Q3): 小さい方から75%目の数
  - 平均値(Mean)
  - 標準偏差(Std. Dev.) ⇒ 分散(Variance)
  - 歪度(Skewness): 非対称性を表す値
  - 尖度(Kurtosis): とがり具合を表す値

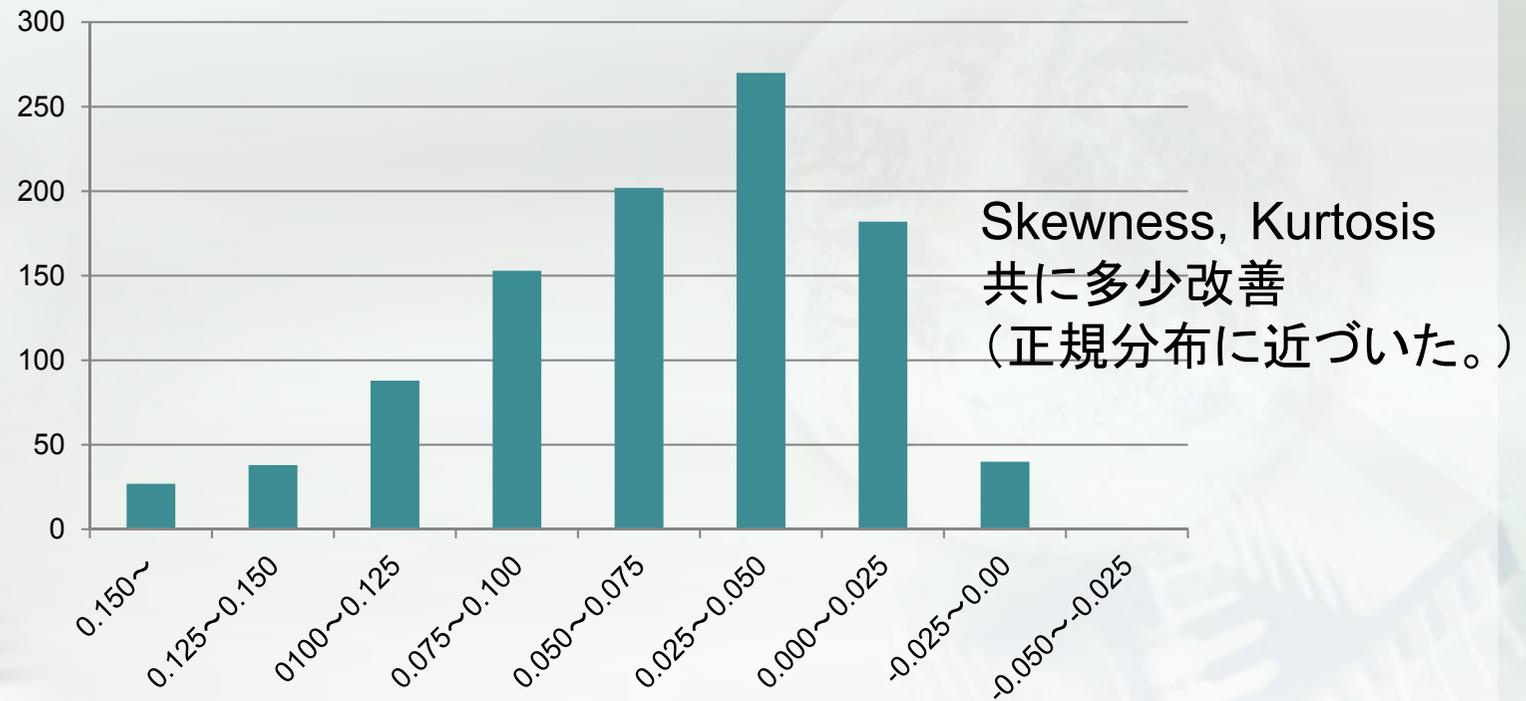
# 記述統計量

ROA

Percentiles	Smallest	外れ値?		
1%	-0.0855092	<b>-0.7224755 (Max)</b>		
5%	-0.0147496	-0.2226606		
10%	0.0076135	-0.1717053	Obs	<b>1056</b>
25%	<b>0.0255937</b>	-0.1120649	Sum of Wgt.	1056
50%	<b>0.0504667 (Median)</b>		Mean	<b>0.0568296</b>
		Largest	Std. Dev.	<b>0.0590756</b>
75%	<b>0.0852716</b>	0.2882749		
90%	0.1177539	0.3128935	Variance	<b>0.0034899</b>
95%	0.1452985	0.3793346	Skewness	<b>-1.411734</b>
99%	0.2178472	<b>0.4678797 (Min)</b>	Kurtosis	<b>35.42571</b>

## 分布②

- 外れ値として上下5%(各21社)をサンプルから外した場合の度数分布表

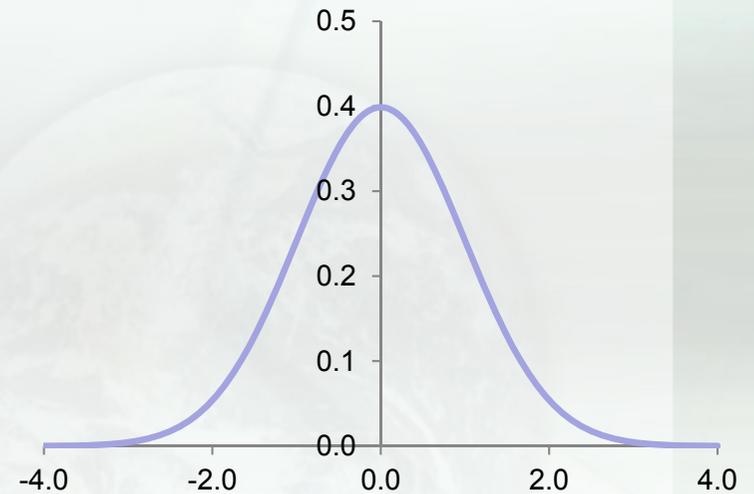


# 仮説検定とは

- ある仮説を立て、その仮説が間違っていることを統計学的に証明すること。
- 例えば、「製造業におけるROAの平均は5%より大きい」という仮説を置いたとする。
  - 標本平均(ある東証1部上場の製造業で実際に計算をしたROAの平均)の値から、母集団の平均(日本の製造業におけるROAの平均)が0.05だと言って差し支えないのか？
  - 帰無仮説 $H_0$ :「 $ROA=0.05$ 」
  - 対立仮説 $H_1$ :「 $ROA>0.05$ 」
  - $H_0$ を検定し、統計的に棄却(誤っていると)されれば $H_1$ が受容(正しいと)される。

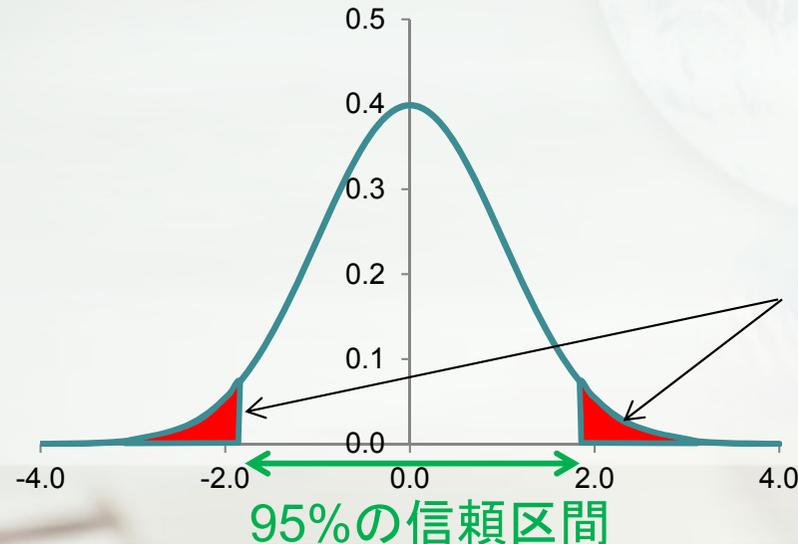
# 信頼区間

- 右図は(標準)正規分布
- この分布は, 平均値が0, 標準偏差1であるような確率分布を持つ時に, それぞれの値がどのような確率で生起するかを表したものと理解できる。
- 例えば, 1.64より大きな値を取る可能性は5%しかない稀なケースということになる。  
⇒ 標準正規分布の確率分布表参照



# 棄却域の設定

- 1.64より大きな値となった事象が観測された場合、**珍しい**ことが起こったと考えるよりも、そもそも設定していた「平均が0」という仮定(仮説)が誤っていたのではないかと考えることが合理的ではないか？
- どの程度で起こることなら珍しいと言えるのか？⇒有意水準(棄却域)の設定
  - 一般には10%, 5%, 1%など、いくつかの段階をもって有意水準を設定。
  - 有意水準の設定は慣習的なもの。cf) 自然科学の世界では0.001%なども



# 母集団の平均の推定

- 母集団の分布とデータの大小によって、それぞれ推定量を作り、その分布(の近似)を元に検定を行う。

データ数	正規分布	非正規分布
大	正規分布	正規分布
小	t分布	—

- 正規分布であれば,

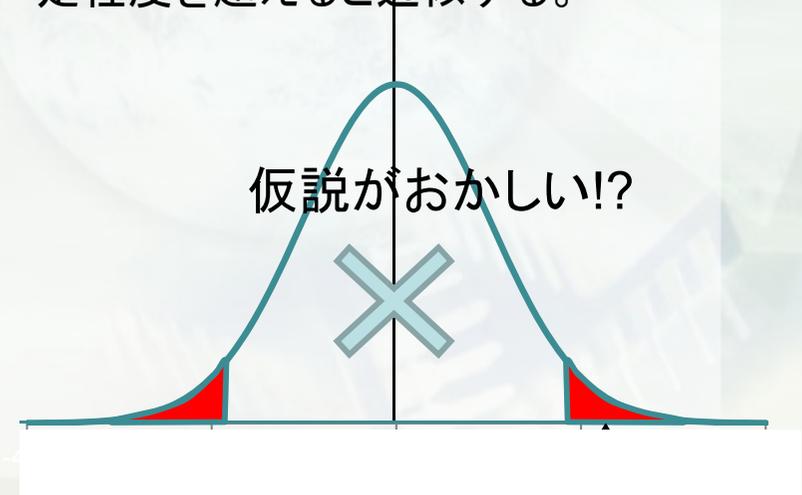
$$Z = \frac{\bar{X} - \mu}{s / \sqrt{n}} \approx N(0,1)$$

- t分布であれば

$$t = \frac{\bar{X} - \mu}{s / \sqrt{n}} \sim t(n - 1)$$

# 仮説検定のイメージ

1. 母集団の平均を仮定する。
  - 帰無仮説の設定
2. 標本の平均を計算する。
3. 母集団分布を想定し、統計量を作る。
  - 母集団の分布によらず、標本サイズが十分に大きければ正規分布を仮定。
  - 母集団の分布が正規分布で、標本サイズが小さい場合にはt分布を仮定。
  - ただし、標準正規分布とt分布は標本サイズが一定程度を超えると近似する。
4. 有意水準を設定する。
5. 帰無仮説の棄却・受容を決定する。



# 平均値の検定

- 母集団の平均を仮定する
  - 帰無仮説 $H_0$ :「ROA=0.05」
  - 対立仮説 $H_1$ :「ROA>0.05」
- 標本平均を計算する
  - サンプル(東証1部上場製造企業1013社, 2008年度3月決算)の平均値は0.0565
- 母集団の分布を想定し, 推定量を作る。
  - 母集団は正規分布である(統計処理ではこれが一般的)とし, サンプルサイズは十分ではないとする⇒t検定
  - $t = \frac{0.0565 - 0.05}{\frac{0.04163}{\sqrt{1014}}} = 5.0351$
- 有意水準を1%とする。
  - 標準正規分布と近似していると考えて,  $t=2.58$ が境界。
- よって帰無仮説は棄却され, 対立仮説を受容する。

# Excelによる平均値の検定

- 私の知る限りにおいて、Excelの「分析ツール」で今回の仮説検定はできない。
- ただし、以下の手順を経れば検定は可能。
  1. まず、"AVERAGE"関数で、標本の平均 $X$ を計算する。
  2. 続いて、"STDEV.S"関数で、標本の標準偏差 $s$ を計算する。
  3. " $=(X-0.05)/(s/(n)^{0.5})$ "で $t$ 値を計算する。
  4. "T.DIST.2T( $t$ 値, $n-1$ )"で両側検定の有意確率が出てくる。
    - ✓ つまり、この値が0.1より小さければ10%水準で、0.05より小さければ5%水準で、0.01より小さければ1%水準で帰無仮説を棄却できる。
- 少し面倒だが、これで $t$ 検定は可能。

# Eviewsによる仮説検定

- 藤井ゼミでは, Excelによる検定だけではなく(というか, Excelは使いません)Eviews等の統計専門ソフトを使っていくこととなります。
  - 理由① Excelの統計数値が信頼性に欠ける(もちろん, 一般的な意味での検定であれば問題ないが, 公表される論文としては信頼性に欠けるという意味)
  - 理由② Eviewsは数万円するソフトであり, なかなか自分で買って使用するという機会がない。大学ではそのソフトが自由に使えることから, 学習の良い機会だと思われる。
- Eviewsの使用方法については, 以下の資料参照。
  - 藤井先生のwebサイトの学部演習→2013年の中にある「Eviewsによる統計処理(1)」および「同(2)」
  - 産業経済研究所の研究員 小西葉子先生によるマニュアル  
<http://ykonishi.web.fc2.com/text.html>
    - ただし, 印刷ができないように設定されているので, どうしても印刷したい場合は自身の責任の下, 何らかの方法でロックを解除する等をしてください。
    - 「PDF, アンロック」などで検索すれば何らかの方法は出てくるのかも...