

統計学 講義

第1回 イントロダクション&ガイダンス

2016年4月12日(火) 3限

担当教員: 唐渡 広志

研究室: 経済学研究棟4階432号室

email: kkarato@eco.u-toyama.ac.jp

website: <http://www3.u-toyama.ac.jp/kkarato/>

受講上の注意

- 授業時間帯：火曜 3 限 + 金曜 1 限
- スライドの印刷資料を配布するが、すべての情報が書かれているわけではないので、必ずノート、メモをとる。
- 板書が多くなることもある。
- 平方根 $\sqrt{\quad}$ が計算できる電卓を毎回持参する。

参考書

- 白砂堤津耶 [2015] 『例題で学ぶ初歩からの統計学 第2版』 日本評論社, ISBN 453555790X, ¥2700.
- 鳥居泰彦 [1994] 『はじめての統計学』 日本経済新聞社, ISBN 4532130743, ¥2233.
- 大屋幸輔 [2011] 『コアテキスト 統計学 第2版』 新世社. ISBN 4883840506, ¥2100.
- 神林博史・三輪哲[2011] 『社会調査のための統計学 – 生きた実例で理解する –』 技術評論社, ISBN-10: 4774146803, ¥2138.
- 加納・浅子・竹内[2011] 『入門 経済のための統計学 第3版』 日本評論社, ISBN 4535556571, ¥3672
- 日本統計学会 (編集)[2012] 『日本統計学会公式認定 統計検定 2 級対応 統計学基礎』 東京図書, ISBN 4489021224, ¥2376
- 日本統計学会 (編集)[2012] 『日本統計学会公式認定 統計検定 3 級対応 データの分析』 東京図書, ISBN 4489021321, ¥2376

講義予定

#	月日	曜日	講義内容	#	月日	曜日	講義内容
1	4月12日	火曜日	イントロダクション	16	6月10日	金曜日	母平均の区間推定
2	4月15日	金曜日	データの種類と整理	17	6月14日	火曜日	母比率の区間推定
3	4月19日	火曜日	母集団と標本	18	6月17日	金曜日	母分散の区間推定
4	4月22日	金曜日	データの要約	19	6月21日	火曜日	仮説検定 (1)
5	4月26日	火曜日	和の記号	20	6月24日	金曜日	仮説検定 (2)
6	5月6日	金曜日	データの変換	21	6月28日	火曜日	仮説検定 (3)
7	5月10日	火曜日	階級別データの整理	22	7月1日	金曜日	仮説検定 (4)
8	5月13日	金曜日	確率分布の基礎 (1)	23	7月5日	火曜日	2変量データの分析(1)
9	5月17日	火曜日	確率分布の基礎 (2)	24	7月8日	金曜日	2変量データの分析(2)
10	5月20日	金曜日	確率分布の基礎 (3)	25	7月12日	火曜日	2変量データの分析(3)
11	5月24日	火曜日	標準正規分布 (1)	26	7月15日	金曜日	回帰分析(1)
12	5月27日	金曜日	標準正規分布 (2)	27	7月19日	火曜日	回帰分析(2)
13	5月31日	火曜日	中間試験	28	7月22日	金曜日	回帰分析(3)
14	6月3日	金曜日	中心極限定理	29	7月26日	火曜日	回帰分析(4)
15	6月7日	火曜日	標本平均の分布	30	7月29日	金曜日	まとめ

成績評価方法

■ 評価対象と評価ウェイト

- 授業中に行う演習問題の提出 20%
- 宿題の提出 10%
- 中間試験 20%
- 期末試験 50%

■ 評価基準

- 総合得点：評価対象にウェイトをつけて100点に換算
- 評価
 - A（優）： 総合得点 ≥ 80
 - B（良）： 総合得点 ≥ 70
 - C（可）： 総合得点 ≥ 60
 - F（不合格）： 総合得点 < 60
 - H（未評価）： 期末試験を受験しない場合

統計的思考と推測

度数分布表とヒストグラム

20XX年度統計学受講者の身長 [男性] 104名

観測データ表

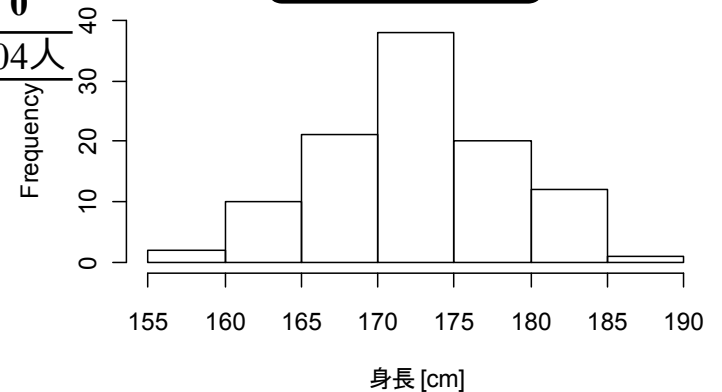
番号	身長
1	177
2	174
3	176
4	173
5	164
6	175
⋮	
⋮	
⋮	
102	168
103	168
104	171

度数分布表

身長	度数
150-155cm	0
155-160	2
160-165	10
165-170	23
170-175	36
175-180	20
180-185	12
185-190	1
190-195	0
合計	104人

170~175cmの人は36人いる

ヒストグラム



統計的（常識的）な判断 (1)

【例】 ある部屋に2人の大人がいる

身長150cmぐらいの人 (A)

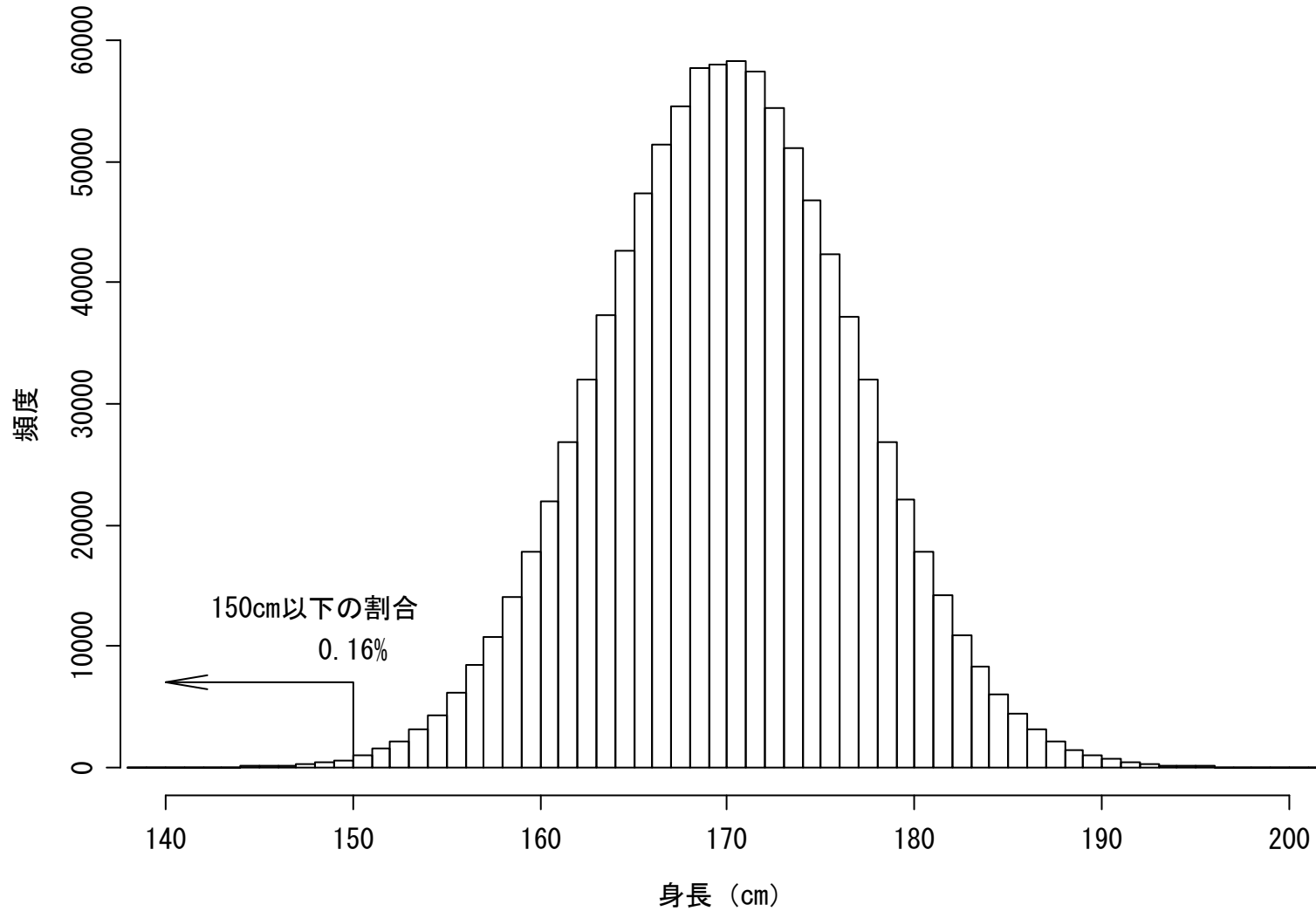
身長180cmぐらいの人 (B)

(A) さんは男性だろうか？女性だろうか？

(B) さんは男性だろうか？女性だろうか？

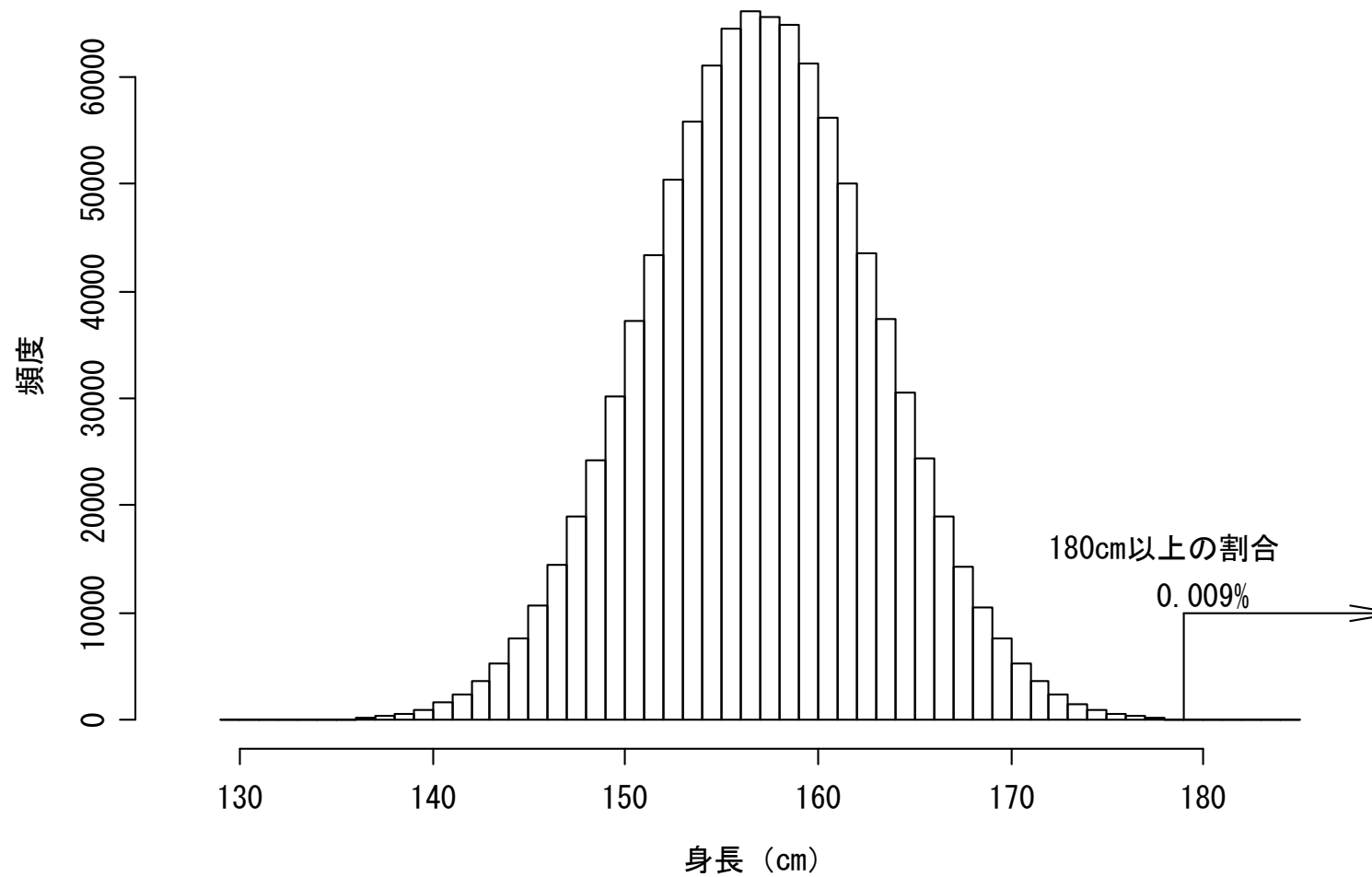
男性の身長分布

身長分布：100万人の成人男性



女性の身長分布

身長分布：100万人の成人女性



統計的（常識的）な判断 (2)

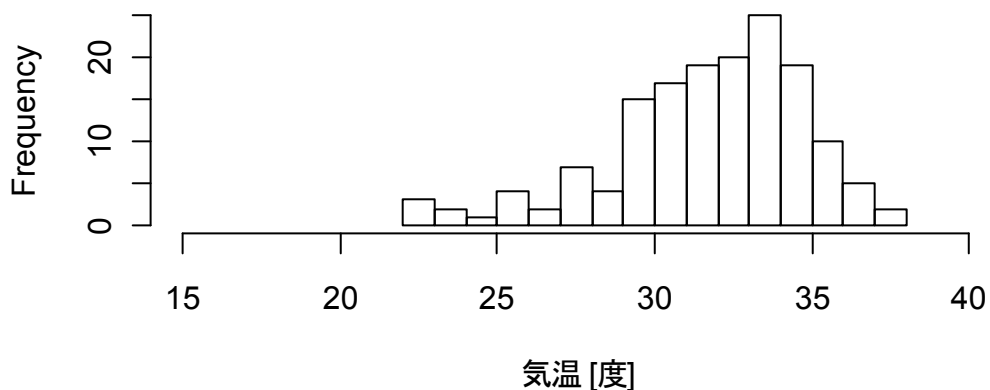
- 東京は札幌よりも暑い。
- このやかんでお湯を沸かすと2分後に沸騰する。
- 自宅から学校まで歩いて10分かかる。

蓄積された経験に基づく常識的な判断 = 統計的思考
100%確実とは言い切れないが、可能性の問題としてありうる事柄の判断

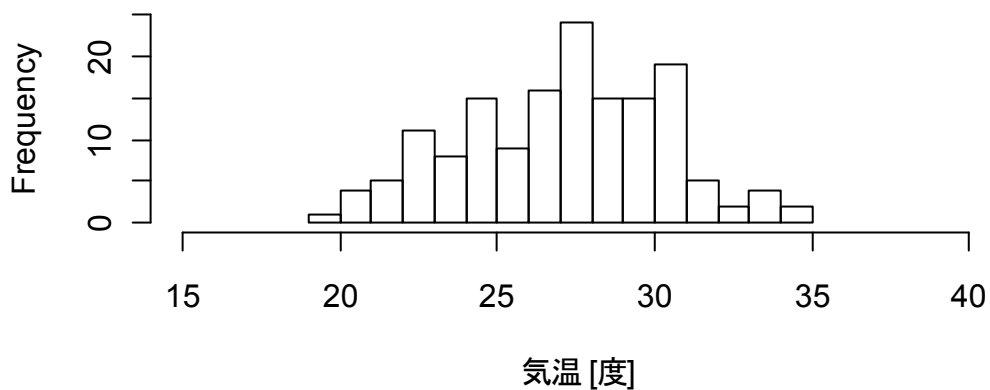
8月の最高気温の比較

31日×5年

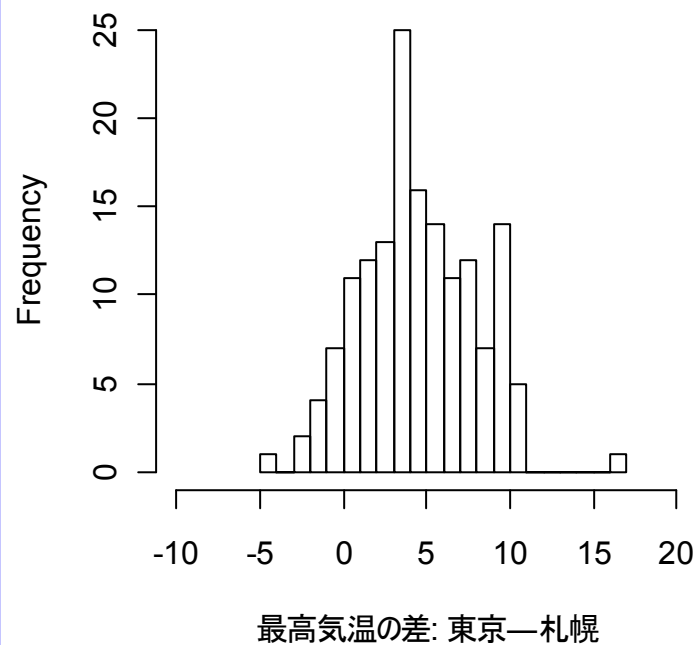
東京の最高気温(2007-2011年)



札幌の最高気温(2007-2011年)



東京と札幌の気温差



8月中, 札幌の方が暑い日が5年間で13日(8.4%)あった。

統計的推測

- 経験だけでは判断できないケース

- 例. ある市長選挙における当選者の予測
 - 投票者数は10万人
 - 候補者は現役市長を含めて4人
 - 選挙当日に投票所で出口調査（100人に聞き取り）
 - 100人中40人が現役市長に投票したと回答
 - 可能性の問題として現役市長は当選確実と考えてよいだろうか？

- 類似の例
 - 開票率わずか 1% 未満で「当確」情報が出るのはなぜ？

開票率 0% (1) : 実験

開票終了後の得票状態(開票率 100%)

候補者	得票	得票率
A (現職市長)	39,000	39%
B	30,000	30%
C	17,000	17%
D	14,000	14%
合計	100,000	100%

20人分だけ開票

有権者番号	候補者
1	B
2	A
3	B
4	B
5	B
6	C
7	C
8	C
9	D
10	A
11	A
12	B
13	B
14	C
15	B
16	B
17	B
18	B
19	A
20	C

すべて開票するまでこの結果はわからない

20人分の集計結果

	A	B	C	D
得票数	4	10	5	1

B候補者がトップ得票

開票率 0% (2) : 実験

20人分だけ開票する場合, 20人の中でA氏がトップ(最多得票)になる可能性はどの程度か?
 100カ所の投票所で20票だけ調べることができるものとしよう。

→10万票の中から「無作為」に 20 票だけ抜き出す作業を100回繰り返す。

	A	B	C	D	Aが得票最多のとき*
投票所-1	4	10	5	1	
投票所-2	13	2	5	0	*
投票所-3	2	11	4	3	
投票所-4	9	6	2	3	*
投票所-5	13	5	2	0	*
・	・	・	・	・	・
・	・	・	・	・	・
・	・	・	・	・	・
投票所-99	14	5	1	0	*
投票所-100	7	6	5	2	*
最多得票の 投票所数	66	29	2	3	



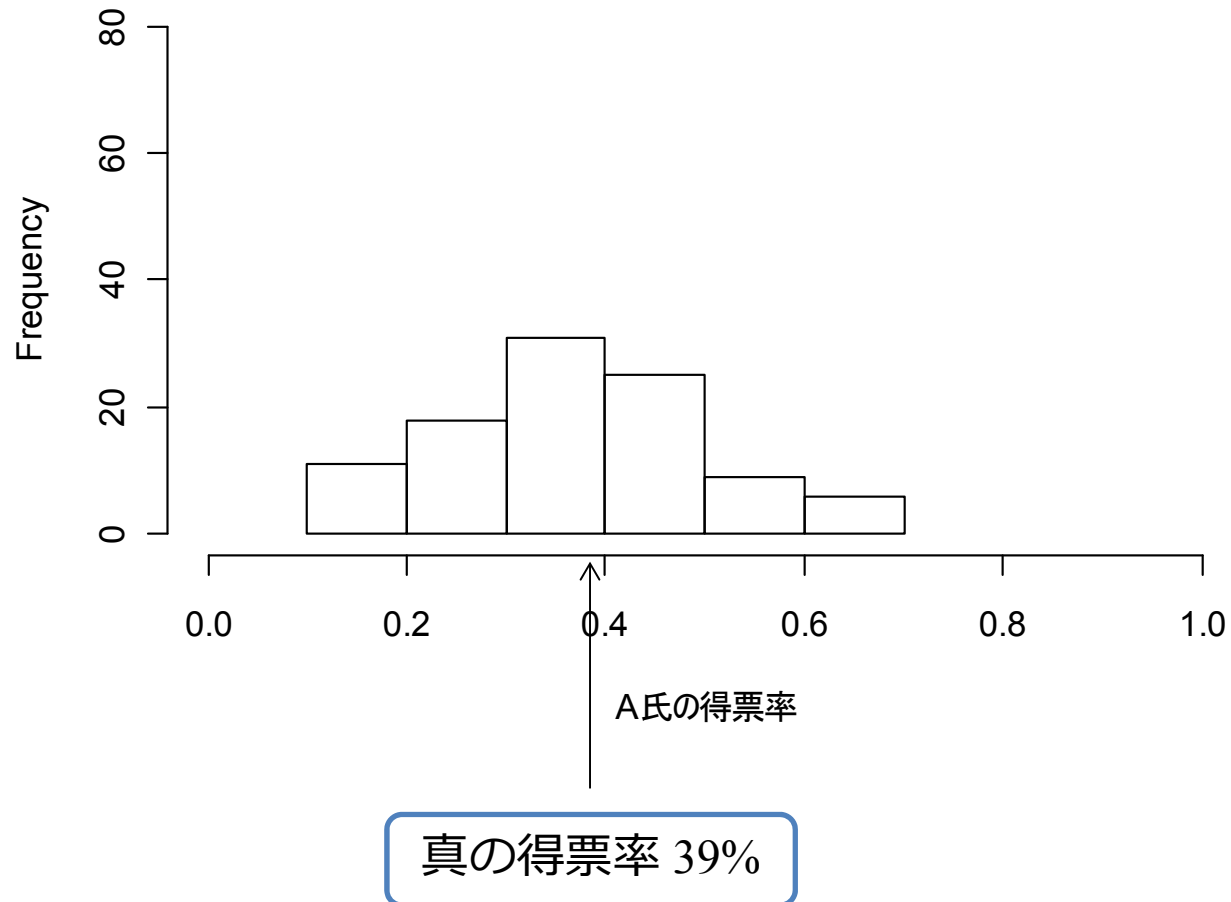
www.nikkei.com

この実験では100カ所中 66 の投票所でAがトップ得票
 20 人分調べる場合, A氏がトップである可能性は 66%
 (情報の精度はやや低い, 10万分の20票なので開票率は0.02%)
 30, 50, 100票, それ以上の場合はどうか?
 →10万票中何票開票すれば「当確」が出せるか?

開票率 0% (3) : 実験

20人開票でのA氏の得票率分布

開票数 = 20

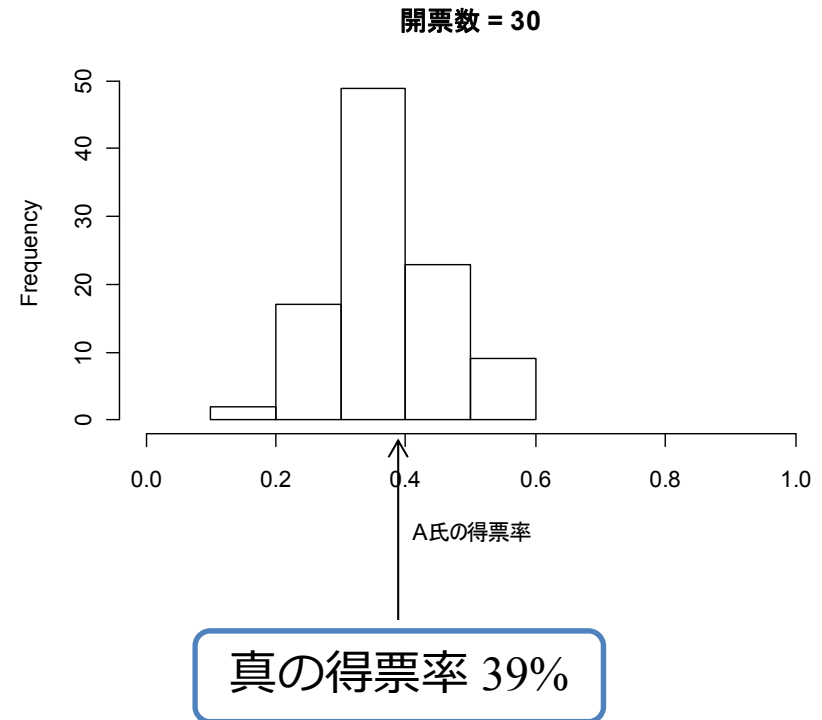


開票率 〇% (4) : 実験

30人分だけ開票する場合, 30人の中でA氏がトップになる可能性はどの程度か?
100カ所の投票所で調べることができるものとしてよう。

→10万票の中から「無作為」に 30 票だけ抜き出す作業を100回繰り返す。

	A	B	C	D	
投票所-1	10	7	4	9	*
投票所-2	9	7	9	5	*
投票所-3	8	13	4	5	
投票所-4	10	14	4	2	
投票所-5	12	8	6	4	*
⋮					
⋮					
⋮					
投票所-99	11	10	5	4	*
投票所-100	14	10	3	3	*
最多得票の 投票所数	77	22	1	0	



100カ所中 77 の投票所でAがトップ得票
30 人分調べる場合, A氏がトップである可能性は 77%
(開票率は10万分の30票 = 0.03%)

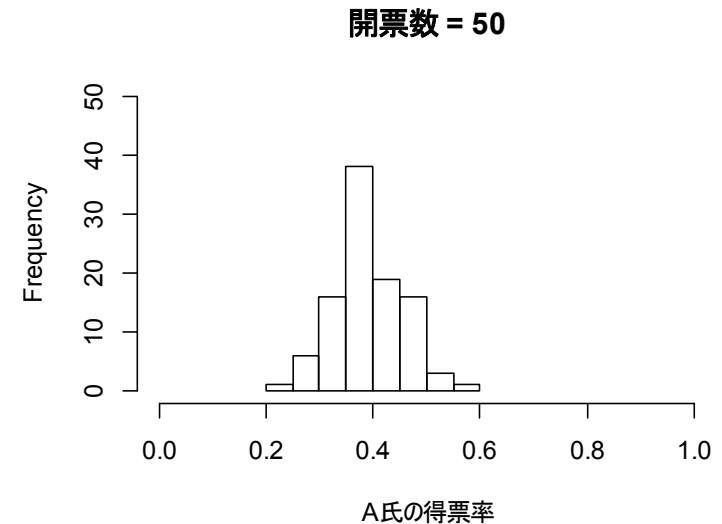
* 印は A がトップの投票所を示す。

開票率 0% (5) : 実験

50人分だけ開票する場合, 50人の中でA氏がトップになる可能性はどの程度か?
100カ所の投票所で調べることができるものとして。

→10万票の中から「無作為」に 50 票だけ抜き出す作業を100回繰り返す。

	A	B	C	D	
投票所-1	20	17	5	8	*
投票所-2	14	13	12	13	*
投票所-3	29	7	7	7	*
投票所-4	17	14	6	13	*
投票所-5	13	20	12	5	
・	・	・	・	・	
・	・	・	・	・	
・	・	・	・	・	
投票所-99	20	14	7	9	*
投票所-100	18	15	7	10	*
最多得票の 投票所数	83	17	0	0	



100カ所中 83 の投票所でAがトップ得票
50 人分調べる場合, A氏がトップである可能性は 83%
(開票率は10万分の50票 = 0.05%)

* 印は A がトップの投票所を示す。

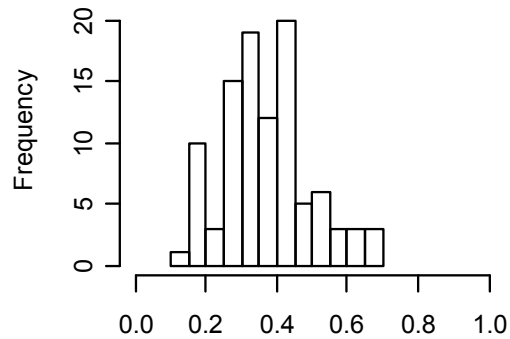
開票率 〇% (6) : 実験結果の集計

開票数 (10万票中)	開票率	A氏がトップ の投票所数 (100箇所中)	当選確実 の可能性
20	0.02%	66	0.66
30	0.03%	77	0.77
50	0.05%	83	0.83
100	0.1%	81	0.81
150	0.15%	89	0.89
200	0.2%	93	0.93
250	0.25%	96	0.96
300	0.3%	99	0.99
500	0.5%	100	1
1000	1%	100	1

開票数を増やしていけば(当然ながら)当確予想の精度は上昇するが、10万票すべてを開票する必要はない。誰がトップになるかどうかは、たかだか 0.5%の開票で答えははっきりする。

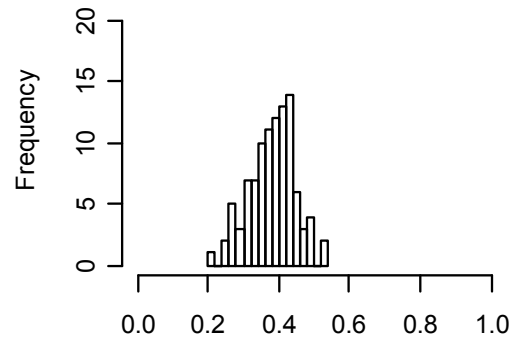
開票率 〇% (7)

開票数 = 20



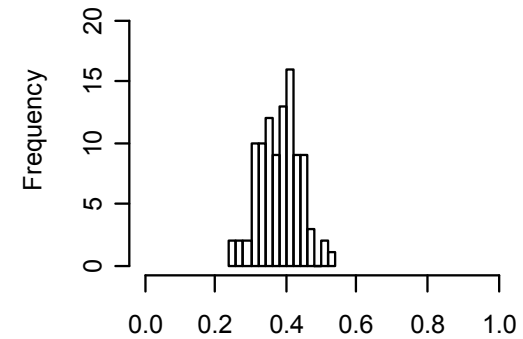
A氏の得票率

開票数 = 50



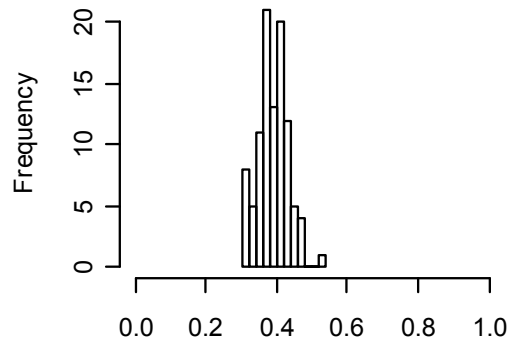
A氏の得票率

開票数 = 100



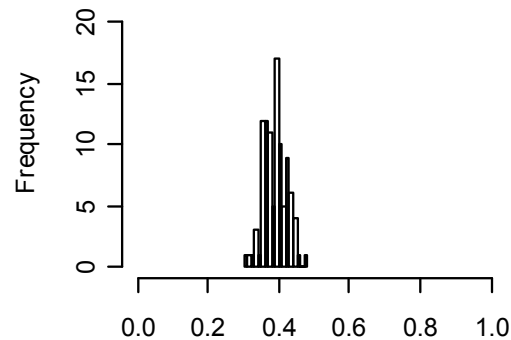
A氏の得票率

開票数 = 150



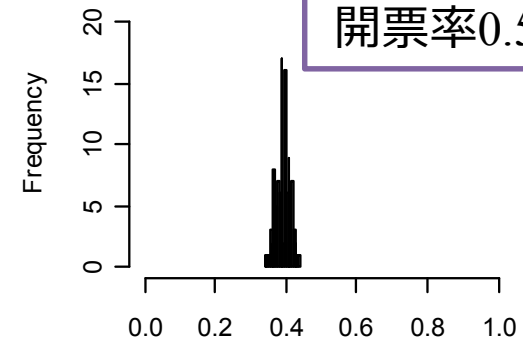
A氏の得票率

開票数 = 250



A氏の得票率

開票数 = 500

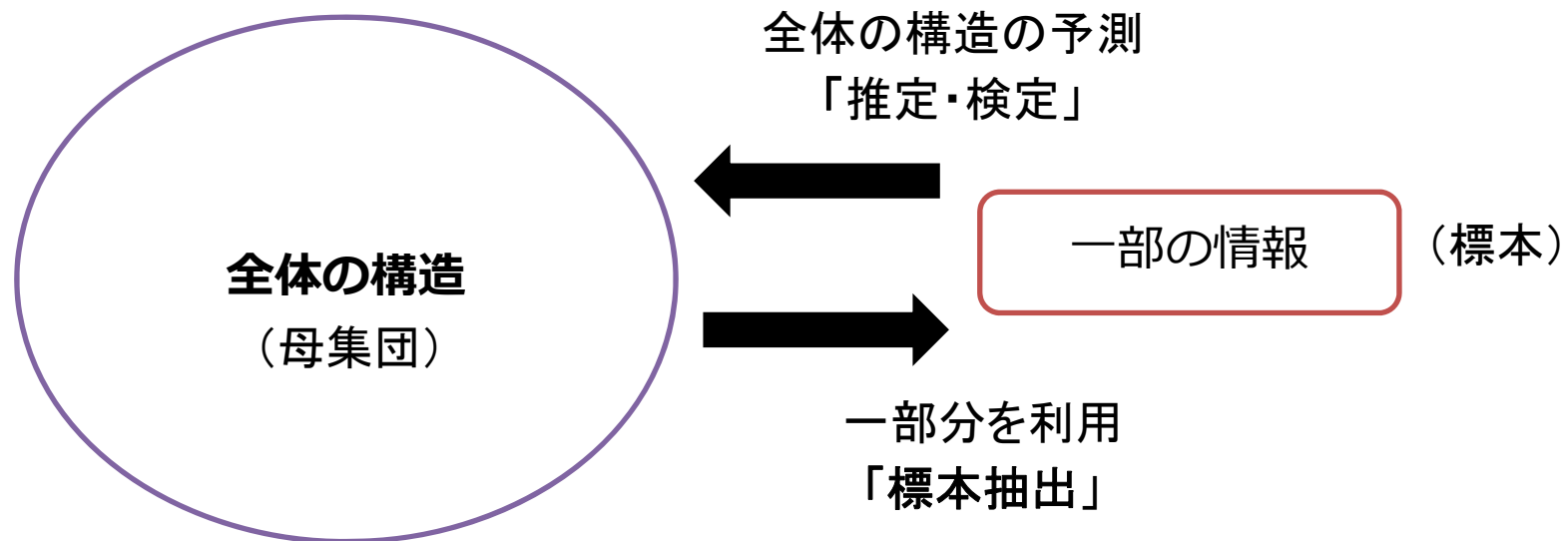


A氏の得票率

開票数を増やしていくと、A氏の得票率の分布の中心は「真の得票率」0.39に近づいていく。この性質を利用すると、開票率がわずかでも十分な予測になる。

統計学の目的

- 一部の情報だけを利用して全体の構造を予測
 - すべての情報を観測・利用することができないケースがある。
 - 調査における時間的, 金銭的制約がある。



統計学ではデータがどのような仕組みで生じたのか、その背後にあるものに思いをめぐらす。

統計学はどこで使われているか

- 実験・調査・観察によって得られた「データ」の分析に利用
- 各学問分野における理論の検証
 - 医学, 疫学, 経済学, 経営学, 心理学, 生物学, 教育学, 社会学, 工学など（経済学→とくに「計量経済学」）
 - 文系・理系の区別なく必須。あらゆる学問の基礎になっている。
 - 理論仮説の検証
 - ・ 最低賃金の水準を上げると失業が増える。
 - ・ マネタリーベースを増やすとインフレになる。
- 社会, 企業経営, スポーツにおける意思決定（昨年
の新聞記事[1]-[9]を参照）。

[1] すかいらーく、「ガスト」アプリ分析——仮説と検証作業 1 週間で。
2015/03/23 日経産業新聞

- 同じ店舗でも時間帯ごとに、同じ時間帯でも店の立地ごとに、それぞれ客層は異なる。モーニングメニューのクーポンでも、高齢者層の多い店か女性層が中心かによってクーポンを出し分ける。さらに高齢者だからあっさりの和食、と単純に考えず、「たまの外食だから洋食に関心があるのでは」と仮説を立てて検証する。

中略

- 神谷ディレクターは「クラウドやスマホの普及で、データを集めて分析するハードルは下がった。重要になるのはデータを解釈する能力だ」と話す。仮説を立てるには事業の課題を理解する能力、結果を読み解くには統計学の知識など、IT（情報技術）にとどまらない多様なスキルが必要になる。こうした多様な能力が、これからのマーケティング担当者には求められると強調する。

- いわゆる「ビッグデータ分析」ブームをけん引した著名なデータサイエンティスト（分析専門技術者）が次々と独立起業している。企業のマーケティングや経営判断に統計分析を活用する機運の高まりに呼応し、使いやすい分析サービスの提供を目指している。

中略

- 「統計分析で成果が出るのは明らかなのに手を付けていない企業が実に多い」と酒巻氏は起業の理由を話す。同社は顧客企業へのコンサルティングと受託分析が当面は業務の中心。特に分析基盤の構築支援を重視する。「やる気はあるのに分析対象のデータを適切に集められていない企業を助きたい」（酒巻氏）。

以下略

- テレビCMやインターネット、雑誌向けなどの広告出稿は、消費者向け製品・サービスを手がける企業にとって顧客に特徴やブランドイメージを直接アピールできる重要な接点だ。ただ広告出稿が売り上げにどう貢献し、媒体別にどう配分することが最も効果的なのかを知るのは難しい。ソフトバンク・テクノロジーは実績豊富な分析官が統計学を活用し、科学的に割り出すサービスを始めた。

中略

- まず顧客企業から対象製品の売り上げと、媒体別の広告出稿額などのデータを提示してもらおう。その上で売り上げをベースとなる額と、例えば「夏になるとビールが売れる」といった季節要因、広告による要因などに大別。専門の分析官が貢献度を割り出すための複雑な数式から成る統計モデルを考える。

以下略

- 日産自動車で初となる女性の専務執行役員だ。市場調査を担当する常務執行役員から昇格し、4月から国内営業とマーケティングの両本部を指揮する。男性中心といわれがちな自動車販売の業界だが、データ分析の知見を織り交ぜ、国内市場に新たな風を起こそうと奔走している。
- 日本債券信用銀行（現あおぞら銀行）を経て、米ノースウエスタン大学ケロッグ経営大学院に自費留学した。マーケティングを学んだ後、日本の調査会社に入った。統計学を取り入れた客観性の高い市場分析は食品や金融など各業界で評判を呼び、2001年に役員へ抜てきされた。

以下略

- 最近、地震や火山活動が活発になっている気がするわ。大地震の前触れだったら心配だわ。地震予測の精度って高まっているのかしら。
- 地震予測をテーマに、飯塚三枝子さん（46）と広瀬真理名さん（26）が久保田啓介編集委員に話を聞いた。

中略

- 地震予測はどのようにしているのですか。
- 「阪神大震災後、政府は一部地域を除いて予知を諦め、長期的な予測に切り替えました。政府の地震調査委員会が活断層やプレート境界ごとに『今後10～50年間に地震が起きる確率は〇%』などと公表しています。地震には周期性があり、発生間隔と直近の地震の年代がわかれば、統計学の手法で確率を計算できます。地震が起きた直後の確率はゼロに近く、周期上の“満期”になれば約50%、それを超えると100%に近づくという考え方です」

以下略

- コーヒーを1日にたくさん飲む人は、飲まない人に比べて心臓病や脳卒中などで死亡するリスクが大きく下がることがわかった。研究したのは日本を代表する国立がん研究センターや東京大学などだ。コーヒーの愛飲者にとっては朗報だが、実はこの結果は疫学研究という手法で出てきた。コーヒーは本当に体にいいのか。疫学研究のカラクリを探ってみた。

中略

- コレラもかっけも原因菌や物質が突き止められたのは疫学研究よりも後になる。原因は分からないものの予防法を見いだすというのは疫学の強みともいえそうだ。
- ただ、日本の疫学研究の体制は欧米に比べると脆弱であると専門家は指摘する。予算の規模も小さいほか、医師以外にも統計学などの専門知識を持った人が必要だが不足している。20～30年も追跡調査をして初めて結果が出てくるものも多いため、5年程度で結果を求められる現状の日本の研究支援の枠組みにそぐわず、研究者が育ちにくいという指摘もある。

- デロイトトーマツファイナンシャルアドバイザー
合同会社（東京・千代田）は、日本企業の訴訟などを
支援するコンサルティングを強化する。このほど
米国公認会計士や統計学の専門家で構成するチーム
を結成。訴訟で経済学や統計学に基づいた損害賠償
額の算定や意見書の作成をする。
- 特許侵害などの知的財産訴訟では、本来得られたはずの利益に基づく損害額の算定や、専門家の証言をする。企業再編関連では、合併が市場価格に与える影響などを分析して意見書を作成。公正取引委員会の審査が円滑に進むように支援する。

- 米プロフットボール、NFLで低迷が続くブラウンズに、米大リーグで旋風を巻き起こした「マネーボール」の手法が導入される。戦略責任者に大リーグのメッツで育成部門の幹部だったポール・デポデスタ氏（43）の就任が決まった。名門ハーバード大卒で、野球に統計学を持ち込んだことで有名になったデポデスタ氏は「私の経験と考えを注ぎたい」と語り、3勝13敗で5季連続の地区最下位に沈んだチームの再建に意欲を示した。
- 2000年代前半に、詳細なデータを用いて選手の能力や成績を分析する「セイバーメトリクス」で好成績を残したアスレチックスで、ビリー・ビーン・ゼネラルマネジャー（GM）の右腕として働いた。その躍進ぶりを描いたベストセラー書籍「マネーボール」は、ブラッド・ピット主演で映画化もされた。

以下略

- 関西経済の成長のけん引役として期待が高まっているのが「健康・医療」産業だ。産官学の連携が進み、iPS細胞など画期的な研究開発が進む。ロボット技術などのイノベーションも相次ぎ生まれており、高齢化に向かう中でも経済に新たな活力を与えようとしている。

中略

- 人材育成について、産官学それぞれの役割を果たすことも確認した。三菱総合研究所の稲垣公雄関西センター長は「人材不足も問題だ」と指摘。医学知識に加えてデータ分析に必要な統計学に精通した人材の育成を産官学でやるべきだとした。医療用ロボットを開発する川崎重工業の河野行伸執行役員は「工学的知識も必要」と述べ、大阪大学の澤芳樹教授は「人材育成は大学の使命。企業などと協力体制をつくり貢献したい」と話した。
- I o Tや人工知能など先端技術を経営に生かす手法を議論する分科会では、情報が素早く処理され、生産や消費に大きな影響を与える現状が報告。ネットワークの標準化や情報保護、技術の普及には産官学の連携不足といった課題も挙げられた。

以下略

最も期待される学問

- あらゆる情報が観察されて、計測することができる世界になりつつある (Microsoft JobsBlog)。
- 「いつも言っていることだけれど、これからの10年で魅力的な職業は統計分析になるでしょう」 (Google チーフエコノミスト Hal Varian)

統計学の需要

- 社会人のための統計学関連の書籍はおどろくほど多い。
- 必要とする業種や職種は多い。
- 資格試験の科目になっている。
 - 公認会計士, 証券アナリスト, 財務専門官, 中小企業診断士, 国家総合職, CFP, アクチュアリー, クォンツ など

分布：データの特徴を知る

データの特徴を知る (1) : 株価の変化率

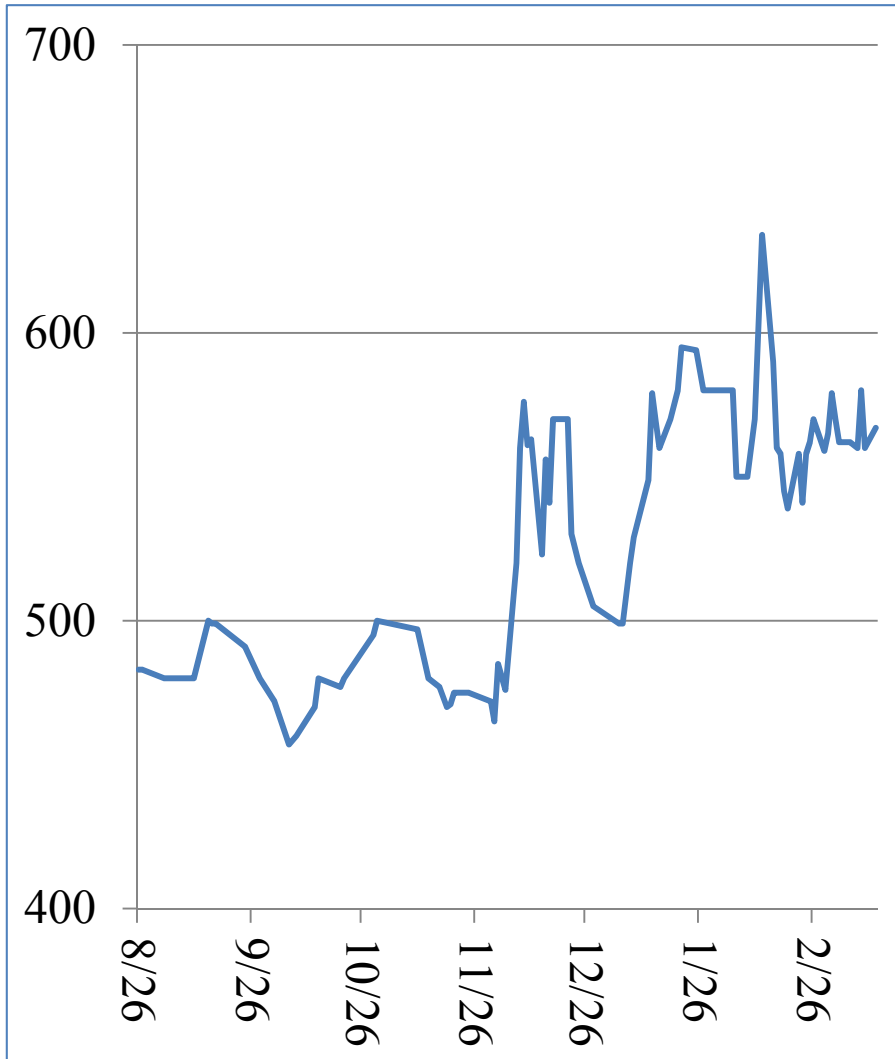
ある会社の株価(円/一株)

	(朝)	(夕方)		
	始値[円]	終値[円]	終値-始値 [円]	変化率(%)
08/23	538	525	-13	-2.4 ← $100\% \times \frac{525 - 538}{538}$
08/24	525	543	18	3.4
08/25	551	558	7	1.3
08/26	569	541	-28	-4.9
⋮	⋮	⋮	⋮	⋮

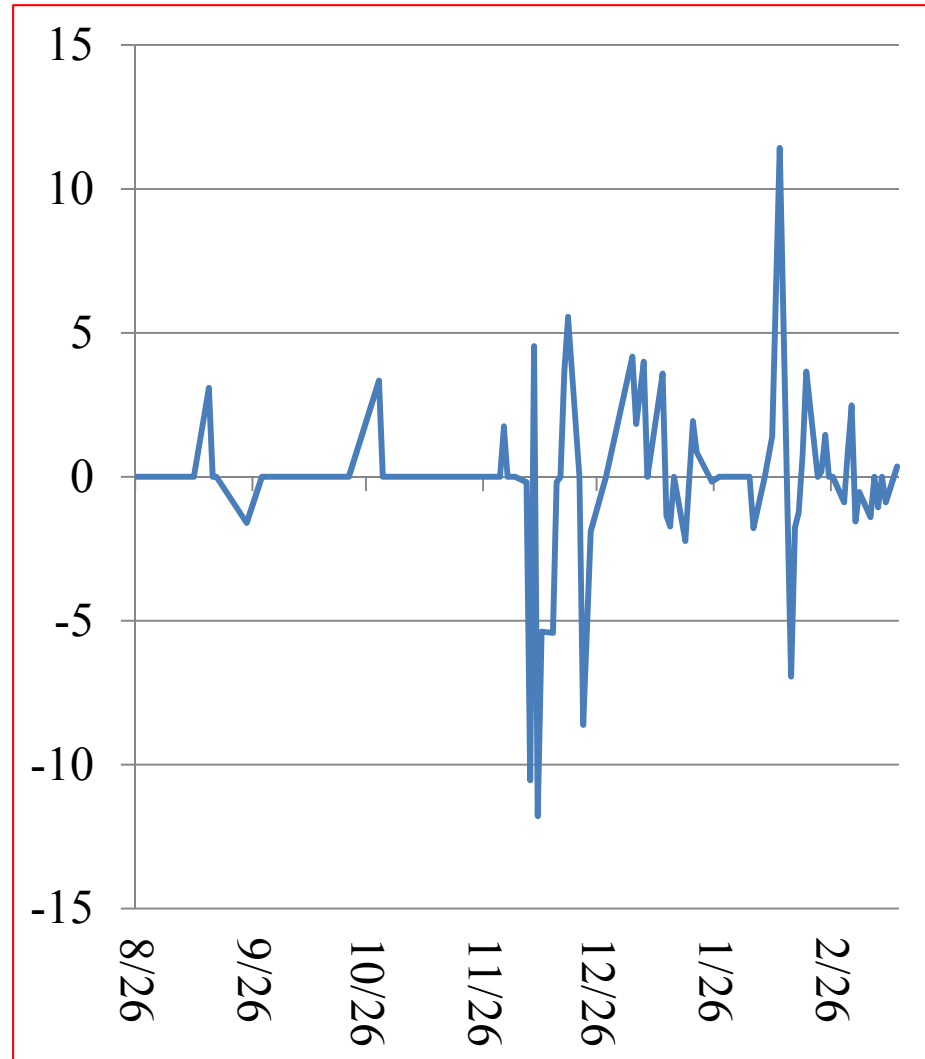
$$\text{変化率}[\%] = 100\% \times \frac{\text{終値} - \text{始値}}{\text{始値}}$$

時系列グラフ

株価[円]



株価の変化率[%]



データの特徴を知る (2) : 変化率データの整理

A社6か月分の変化率データ

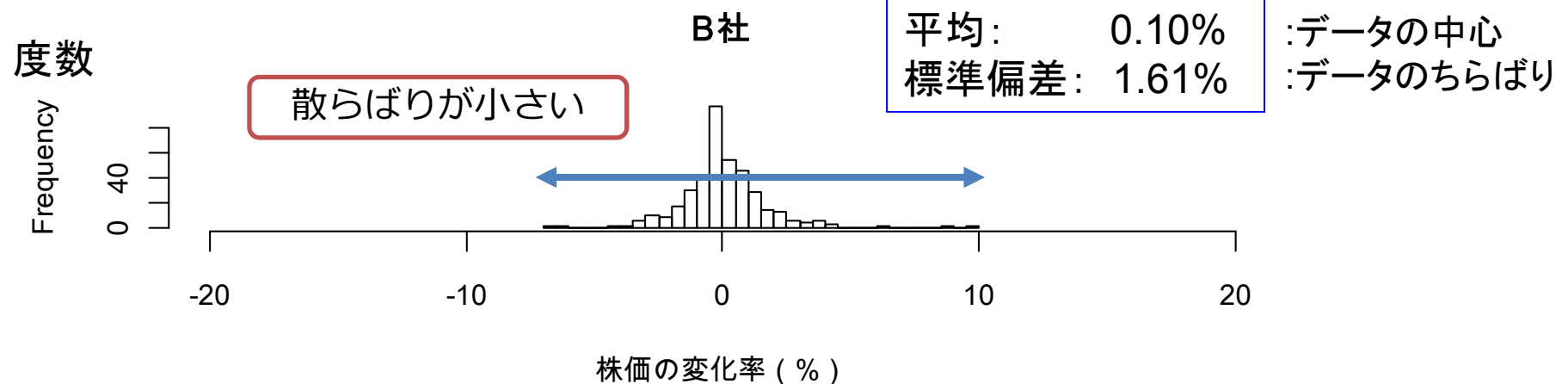
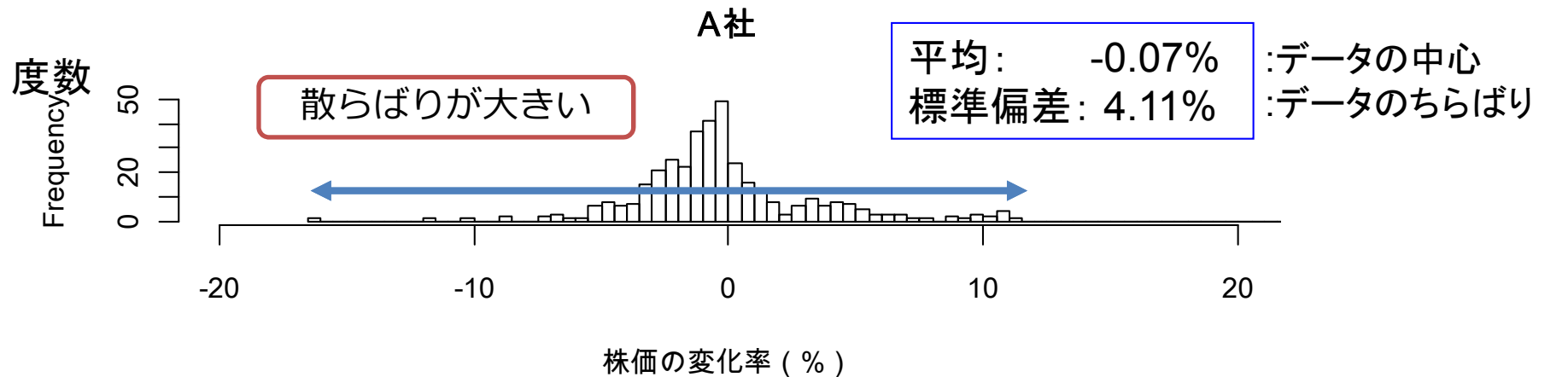
-2.4 3.4 1.3 -4.9 -1.5 0.2 0.0 -1.7 -1.0 -2.9
 -3.1 0.6 -1.8 -3.2 -2.3 -5.0 4.5 4.4 -4.7 -3.9
 -6.6 -1.6 11.2 -3.3 -3.1 -3.2 4.9 -3.9 1.3 0.0
 0.3 -4.3 -0.6 -4.5 -7.5 10.4 10.8 -4.6 -6.4 -0.9
 6.5 5.4 1.7 -10.1 -0.9 -1.7 -4.0 5.3 -0.5 -1.6
 -2.1 -0.7 -1.8 -0.5 -2.1 -2.0 -3.0 4.3 -2.3 -1.0
 9.8 -5.0 -0.5 -1.5 3.6 -3.0 -0.3 0.3 -1.0 -0.8
 -3.5 5.5 1.8 -0.8 1.8 3.5 -1.9 -1.9 3.7 -1.2
 -0.7 -1.5 -1.2 -0.8 -0.3 6.2 -1.6 1.2 0.5 4.8
 -1.8 -0.2 -1.8 4.2 11.0 -1.3 -2.4 -4.4 -4.9 5.1
 10.2 -6.8 3.8 0.0 0.3 -2.1 -0.8 -0.8 -1.9 -1.1
 4.1 -1.5 -0.3 -8.6 -4.8 0.0 -2.6 6.7 3.3 -4.2
 -0.6 -1.5 -4.5 -1.9 -0.6 1.9 0.0 -2.2 0.3 -0.6
 0.9 -2.7 0.6 0.3 -2.3 -2.9 -0.9 0.9 -1.2 -0.6
 -0.3 0.6 -0.3 -0.6 0.0 -2.5 -2.9 -0.3 -0.3 -0.3
 -3.1 -1.0 -0.3 -0.7 0.7 0.0 -0.3 -0.3 1.6 -0.3
 -2.0 -1.2 0.3 -0.6 4.4 -1.1 1.1 0.3 -3.0 2.5
 4.0 0.3 0.3 -1.0 1.0 0.3 0.3 -1.8 0.3 0.0
 3.6 -3.7 3.1 -2.0 -2.8 2.6 -0.7 1.2 0.7 4.7
 1.9 2.5 -0.8 -2.3 -1.1 -1.3 7.0 3.3 -2.3 -0.6

度数分布表

-15%未満		1
-15%以上 -10%未満		2
-10%以上 -8%未満		2
-8%以上 -6%未満		6
-6%以上 -4%未満		21
-4%以上 -2%未満		68
-2%以上 0%未満		124
0%以上 2%未満		85
2%以上 4%未満		24
4%以上 6%未満		23
6%以上 8%未満		8
8%以上 10%未満		5
10%以上 15%未満		8
15%以上		2

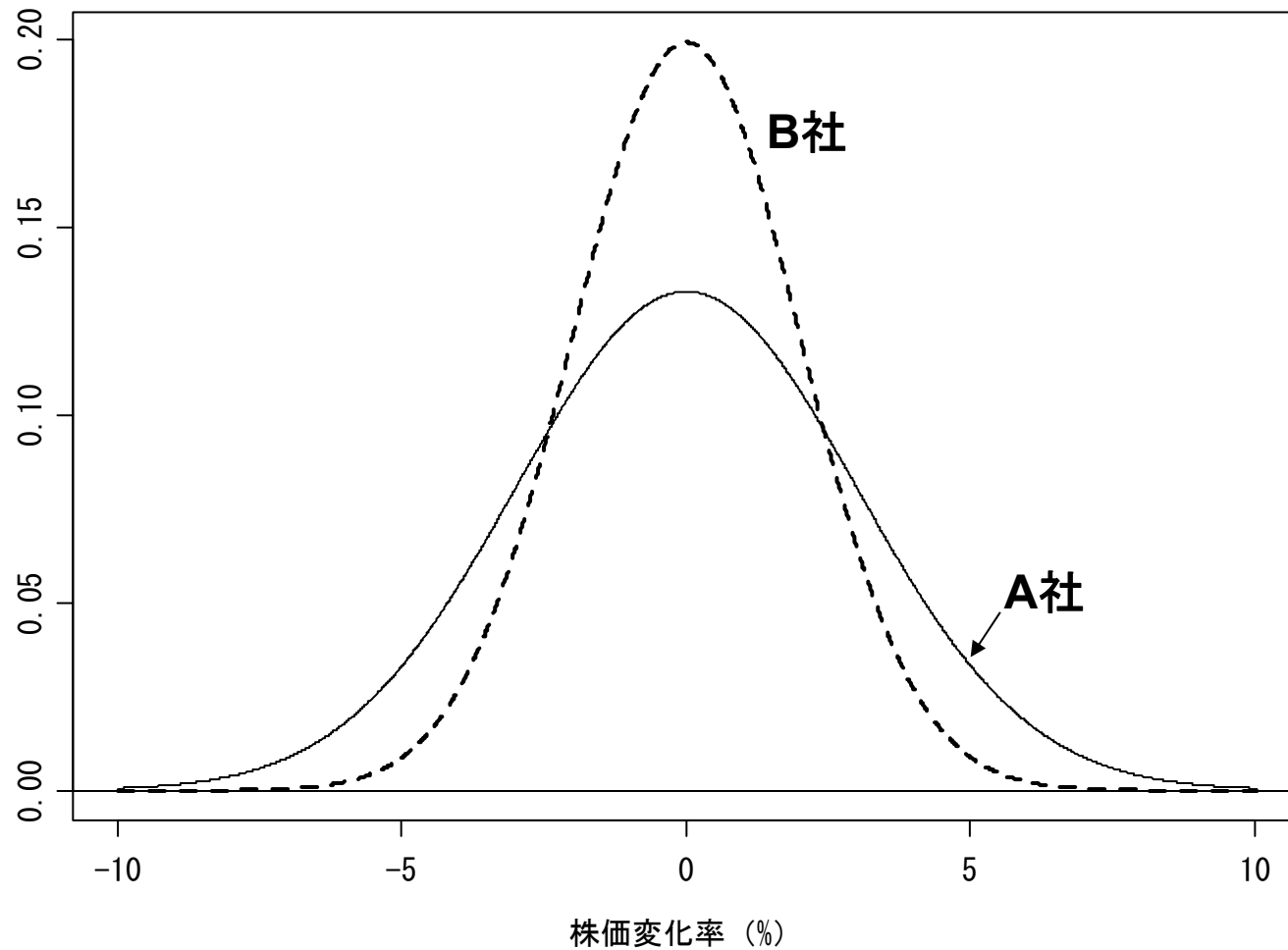
株価の変化率 (1)

始値で買って、終値で売ったときの株価の変化率[%] : 過去6ヶ月

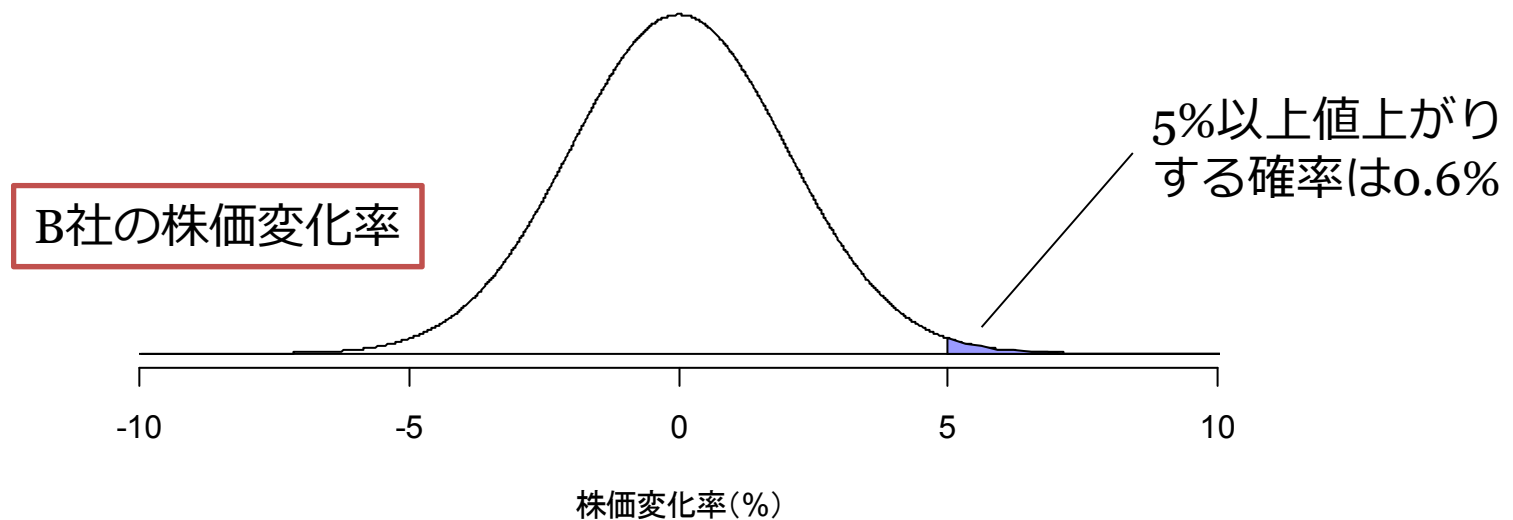
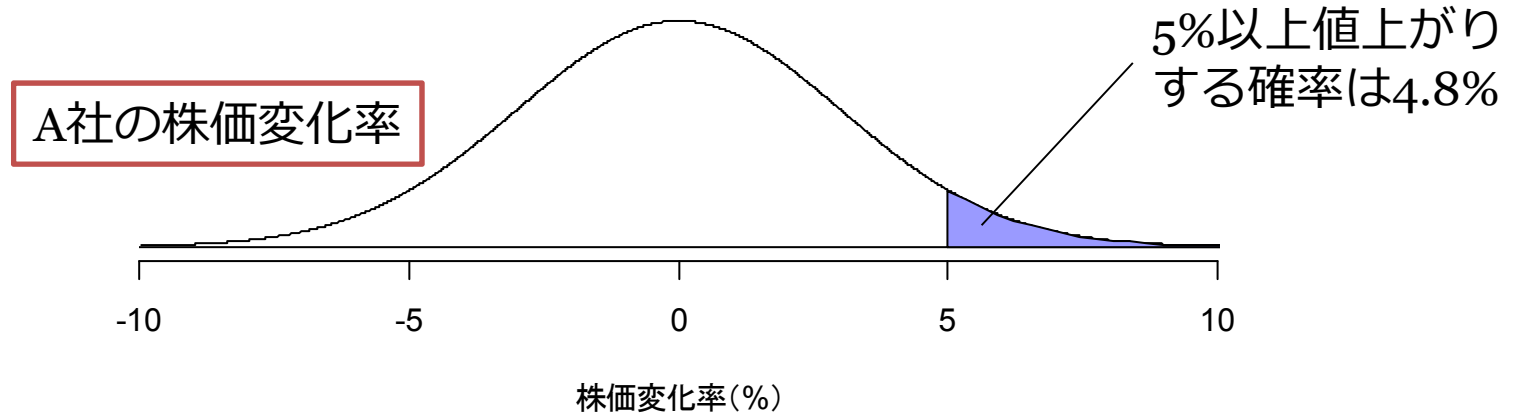


株価の変化率 (2)

5%以上値上がりする可能性をA社とB社で比較

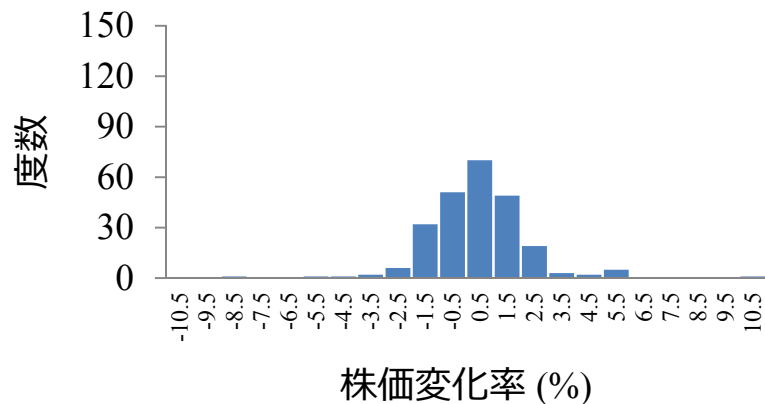


値上がり（値下がり）する確率

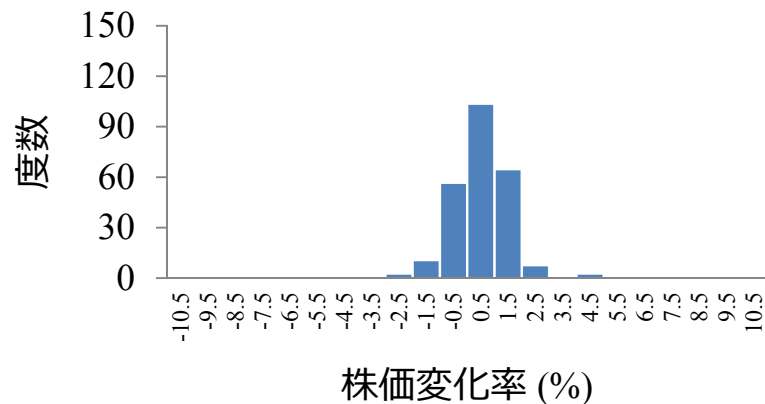


リスクとリターン (2014年1年間の株価変化の分布)

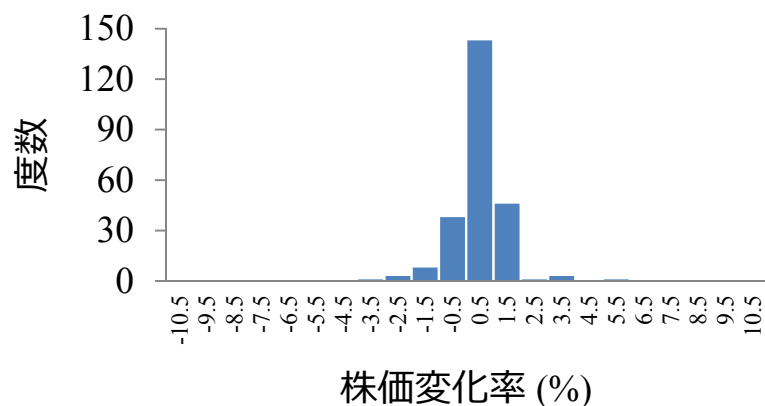
平均 **-0.02%**
標準偏差**2.01%** 任天堂



平均 **0.01%**
標準偏差**0.94%** JR西日本



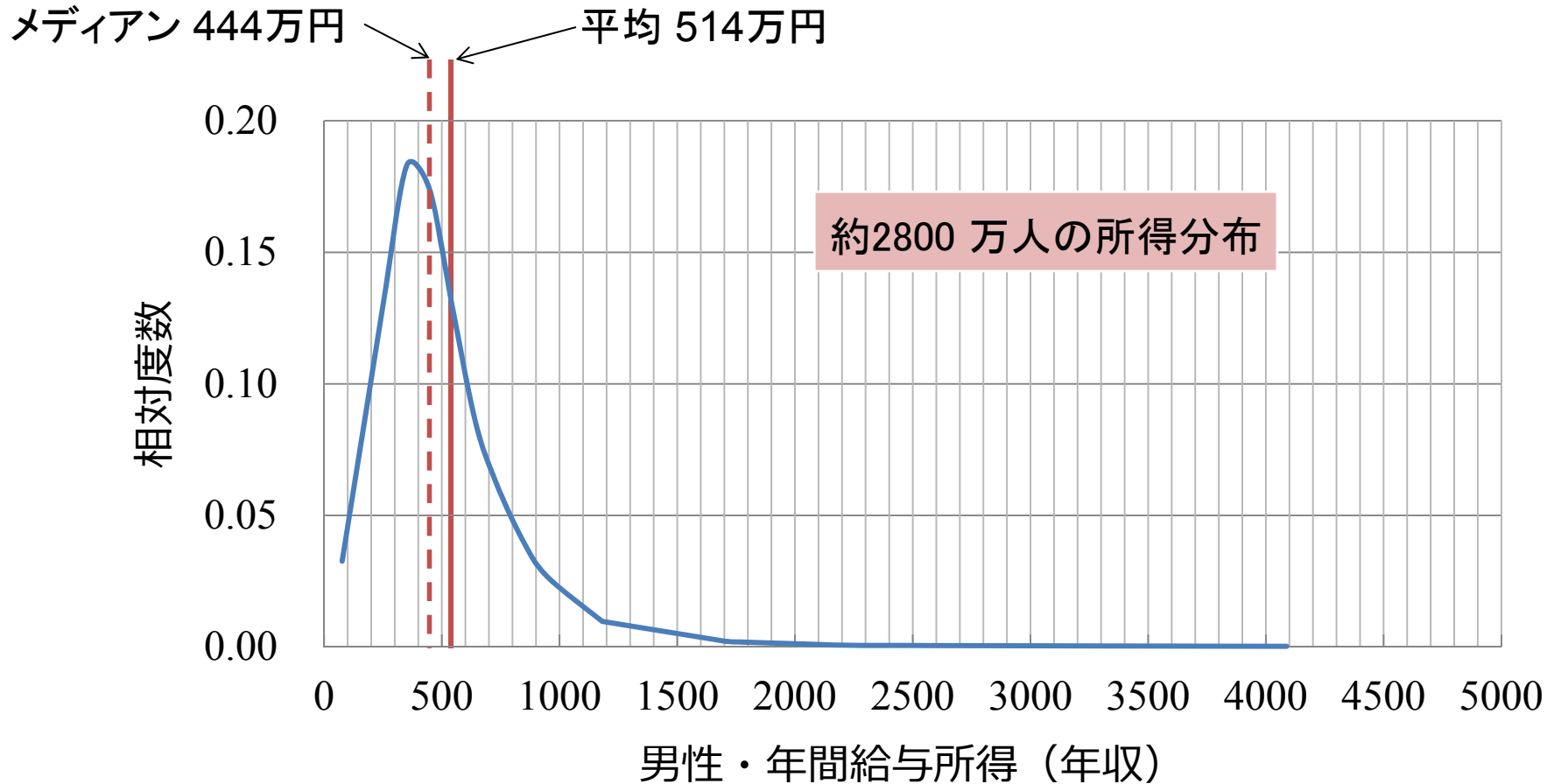
平均 **-0.01%**
標準偏差**0.89%** マクドナルド



データの平均だけでなく、
散らばりも重要

データ出所 : <http://finance.yahoo.co.jp/>

所得分布（男性）



- 右裾が重くなるのが所得分布の特徴（平均値が大きい値に引っ張られる）。
- 2800万人の平均所得は514万円。
- 下から数えてちょうど真ん中の人の年収は444万円。

出所：国税庁『民間給与実態統計調査』平成26年度

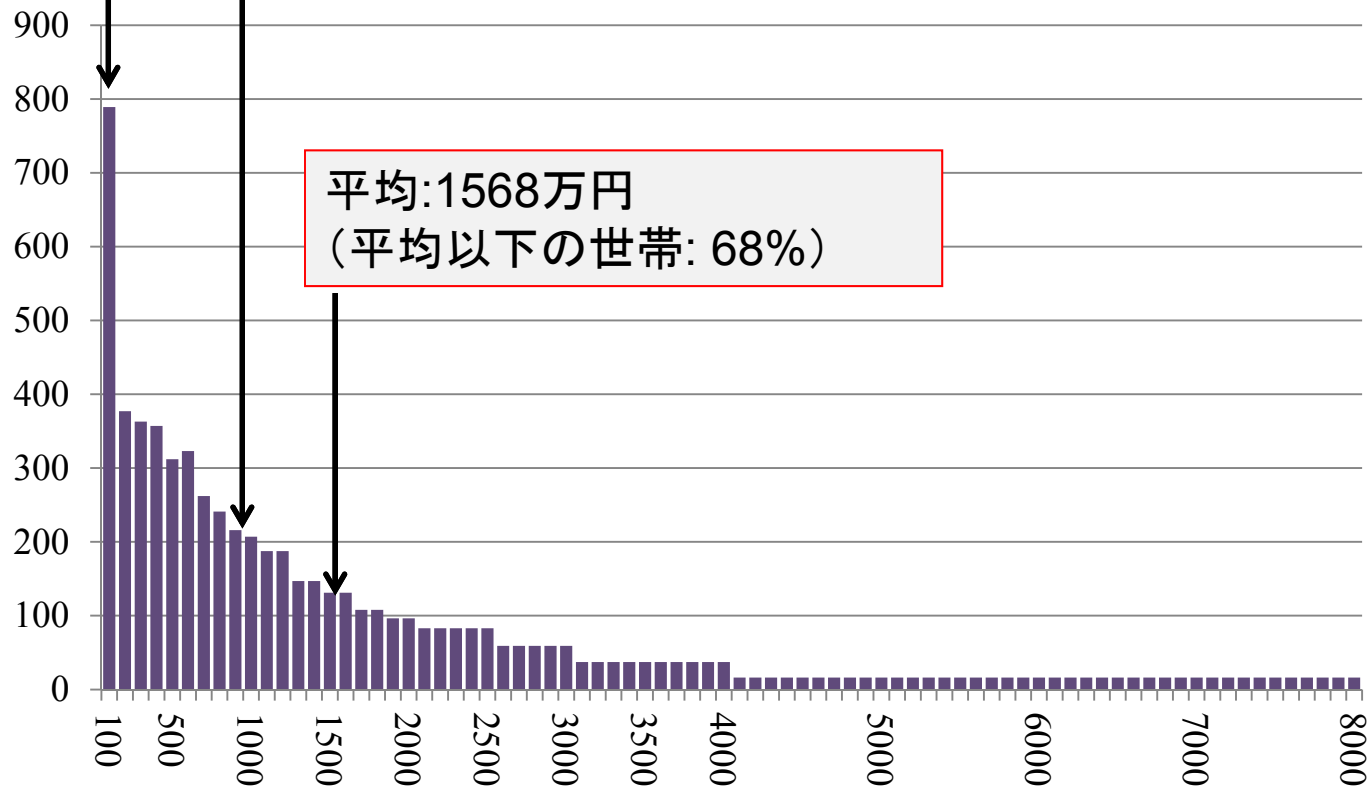
世帯の貯蓄残高の分布

モード：100万円未満
(最頻値)

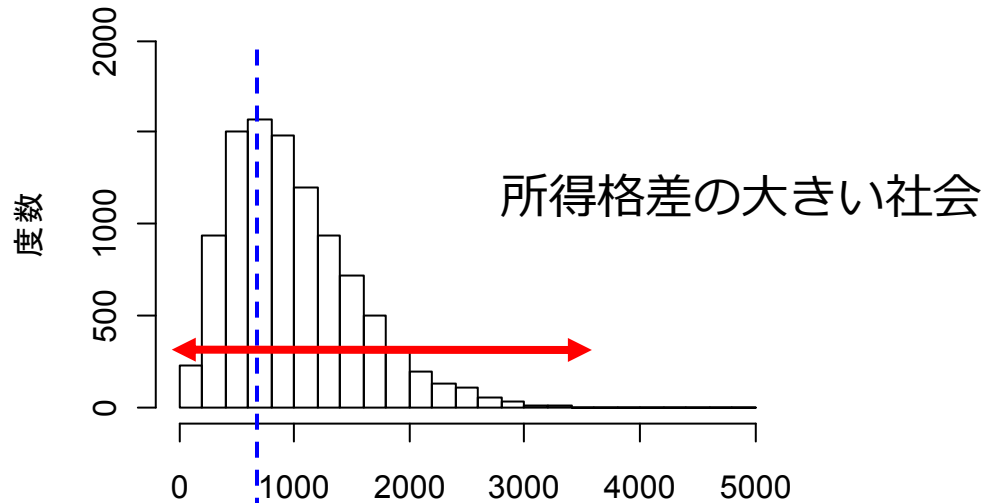
メディアン：909万円
(909万以下の世帯:50%)

平均：1568万円
(平均以下の世帯: 68%)

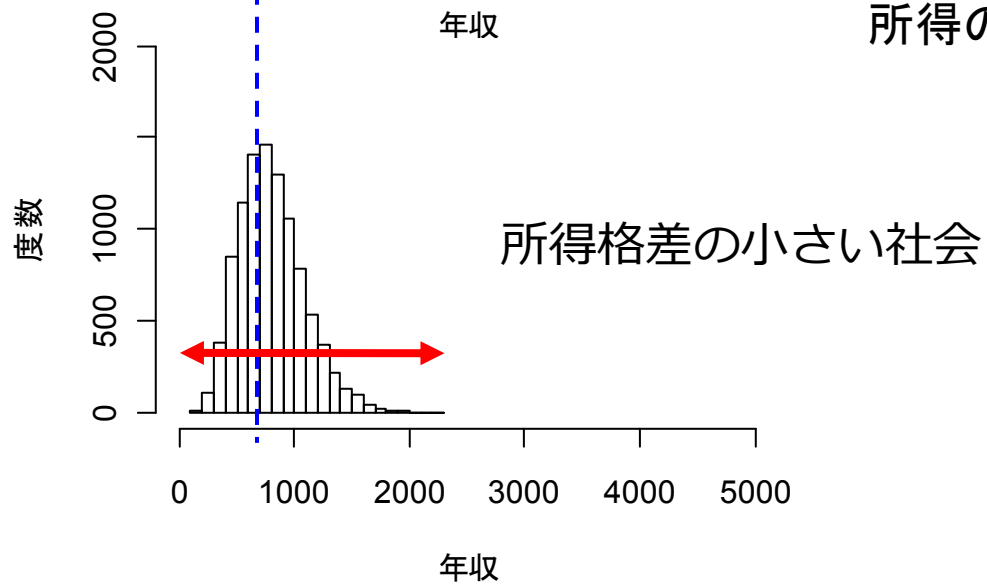
8000世帯の貯蓄残高分布
(出所：総務省「家計調査」)



所得の格差



所得の平均値は同じ。散らばりは異なる。



上位10%の人は下位10%の人の何倍の所得？

Country	P90/P10	Country	P90/P10	Country	P90/P10
Chile	8.7	United Kingdom	4.2	Austria	3.3
Turkey	6.3	Estonia	4.1	Netherlands	3.3
Israel	6.2	Italy	4.1	Sweden	3.3
United States	5.9	Poland	4.0	Slovak Republic	3.2
Japan	5.2	Ireland	3.8	Slovenia	3.2
Spain	4.9	Germany	3.6	Finland	3.1
Korea	4.8	Switzerland	3.6	Iceland	3.1
Portugal	4.8	France	3.5	Czech Republic	3.0
Greece	4.3	Luxembourg	3.5	Norway	2.9
Canada	4.2	Belgium	3.4	Denmark	2.8
New Zealand	4.2	Hungary	3.4		

出所: OCED stat

日本の男性の場合：下位10% = 200万円以下, 上位10% = 1000万円以上

八百長？

出所：S. レヴィット, S. ダブナー (2006)「ヤバい経済学」

理論上の勝率

7勝7敗の力士の8勝6敗の力士に対する**期待勝率**

48.7%

実際の勝率

7勝7敗の力士の8勝6敗の力士に対する**実際の勝率**

79.6%

7勝7敗の力士の9勝5敗の力士に対する**期待勝率**

47.2%

7勝7敗の力士の9勝5敗の力士に対する**実際の勝率**

73.4%

- 7勝7敗で千秋楽を迎えた力士の15日目の勝率は恐ろしく高い。
- 統計学では、それが本当に「ありえないほど珍しいこと」なのかどうかを「確率」の考え方を利用して調べることができる。

統計学のはじめの一步

- データにはどのような種類があり、どのような特徴を持っているのか？
- データを集約・整理するにはどのようなやり方があるか？
- 分布を特徴づけているものは何か？「確率」の考え方はどのように活かされているか？