

# (一社)日本家政学会関東支部 統計講習会

～The ABCs of Statistics～

シム チュン キヤット

昭和女子大学

人間社会学部・現代教養学科

2015年4月25日(土)13:30-16:00

(於:昭和女子大学 学園本部館3階 大会議室)

# 今日のメニュー

1. 社会調査の略史
2. 社会調査法の分類
3. アンケート調査票を作る際の注意点
4. データの取り方(サンプリング)
5. データの種類と分析
6. 正規分布と標準化
7. 仮説検定の考え方
8. 統計分析法いろいろ

# 1. 社会調査の略史

- 古代エジプトでも古代中国でも、人口や領土に関する調査が行われた

**Nothing is certain but death & taxes !**  
**by Benjamin Franklin**

- 「社会統計」の台頭は17世紀のヨーロッパ; 社会分析者は、出生と死亡、結婚を数えようとした; 人口が増えるのは国家が健全な状態 (**state**) にある証拠だと信じられていたから
- このような数値的調査を行う人たちは、**statist** と呼ばれるようになり、やがて定量的な証拠を指す新しい言葉、**statistics** (統計学) が生まれた

## 2. 社会調査法の分類

- **調査票調査** Questionnaire Survey/Research
  - 質問紙調査、アンケート (Enquête) 調査ともいう
- **聞き取り調査** Interview Survey/Research
- **参与観察** Participatory Observation
  - 調査対象である集団に加わって観察し、資料を収集
- **ドキュメント分析** Document Analysis
  - 手紙、日記、新聞・雑誌記事などの文書や記録を収集して、それをデータとして分析
- **二次資料分析** Secondary Data Analysis
  - 既存統計資料分析ともいう

# Social Survey vs Social Research

- 単なるデータ収集や統計的計算の技術は狭義の社会調査  
→ **Social Survey**
- データを通じて、社会的世界を解釈的に探求し、  
新しい知見を提示することを目指す社会調査  
→ **Social Research**
- そのための方法が、「社会調査法」である

# 3. アンケート調査票を作る際の注意点

## 1. 曖昧な表現を避ける

- 政治家の**犯罪**についてどう思いますか
- **浮気**はやはり行けないことでしょうか

## 2. ステレオタイプ of 言葉を避ける

- 官僚の**天下り/再就職**、**ゆとり世代/近年の若手社員**

## 3. 難しい言葉を避ける

- あなたは、「**痘痕も靨**」の英語を知っていますか

## 4. 相手が普段考えていないことは聞かない

- 人工衛星がたくさん打ち上げられている現状について、あなたはどう考えますか

# 調査票を作る際の注意点

5. 似てはいるが微妙にニュアンスが違う質問に注意
  - あなたは、消費税を上げることに賛成ですか vs
  - 消費税を上げることはやむを得ないと思いますか
6. 誘導質問
  - 官僚の汚職が問題化していますが、あなたは公務員を信頼していますか
7. キャリー・オーバー効果 carry-over effect
  - 子どもの犯罪の増加についてどう思いますか
  - 子どもの躾は厳しくするべきだと思いますか

# 調査票を作る際の注意点

## 8. ダブル・バーレル double-barrel

- 親や先生の期待に応えるために、勉強しなければ、と思うことがありますか

## 9. 回答形式

- 単一回答？複数回答？順位回答？  
段階回答？（←奇数段階？偶数段階？）  
サブ・クエスチョン、自由回答を入れる？

10. 多くの質問を詰め込まず、Story のある調査票を

11. 仮説設定をしっかりと

12. Pretest もしっかりと

# いきなりですが...

- 昭和女子大学へのアクセスについて、どう感じましたか。
  1. 遠い
  2. どちらでも言えない
  3. 近い
- 統計学について、私は...
  1. かなり知っている
  2. ある程度知っている
  3. あまり知らない
  4. まったく知らない

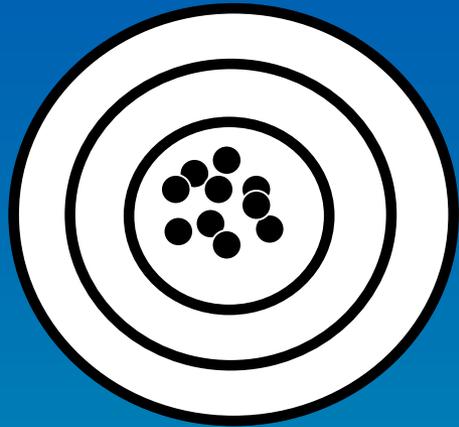
# 4. データの取り方(サンプリング)

- 単純無作為抽出 simple random sampling
  - 「人間の恣意を無くす」ことが醍醐味
- 系統抽出 systematic sampling
  - 母集団の名簿からあるスタート番号をランダムに決めたあと、一定の等しい間隔で対象者を抽出
- 多段抽出 multi-stage sampling
  - 例えば、三段抽出の場合は市区町村→地点→個人
- 層化抽出 stratified sampling
- Random Digit Dialing (RDD、乱数番号法)

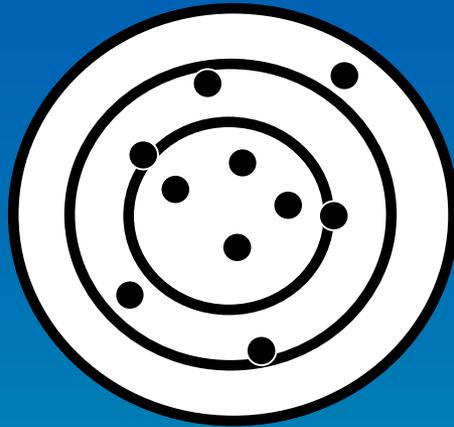
# アンケート調査の方法

- 個別面接法（自記式・他記式）
  - 留置調査法
  - 郵送調査法
  - 電話調査法
  - 街頭調査法（自記式・他記式）
  - 集合調査法
  - 電子メール・インターネット調査法
- 手間、費用、回収率、代表性・信頼性などを考慮

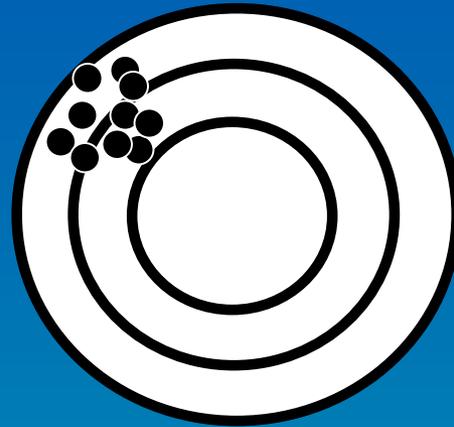
# 妥当性と信頼性の違い



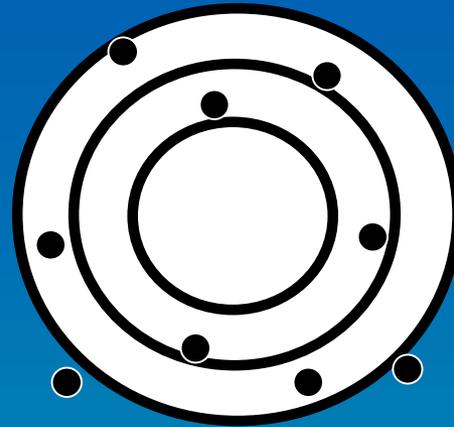
妥当性 ○  
信頼性 ○



妥当性 ○  
信頼性 ×



妥当性 ×  
信頼性 ○



妥当性 ×  
信頼性 ×

- 妥当性: 照準が合っているか
- 信頼性: 射手の腕前(安定性)
- 一般に(研究者の間で)妥当性の方がより重要

# 有効回収率は60%以上必要

- **回収率:**
  - 配布したアンケート調査に占める回収されたアンケート数の割合
- **有効回収率:** (電話調査などの場合は**有効回答率**)
  - 白紙など無効なアンケートを除いたものの割合
- 50%を切るような有効回収(回答)率では、なぜそれだけ多くの方が回答を拒否したのか、アンケートの手法に問題はなかったのか、あるいはアンケートの項目の中に拒否感を持たせるような内容はなかったのかなど再検討する必要がある

# 統計学 Statistics

- **統計：（統べて計る！）**  
ある集団の構成要素を個々に調査して得た数値（データ）を処理し、その集団の性質・傾向などを数量的に表すこと；  
また、その数量的に表されたもの
- **記述統計 Descriptive Statistics：**  
標本から得られたばらばらのデータを整理して、その特徴を説明する統計手法のこと
- **推定統計 Inferential Statistics：**  
標本から得られたデータをもとに、母集団を推測する統計手法のこと

# 5. データの種類と分析

## ■ 量的変数 Quantitative Variables

- 変数の値が、連続的な量としての性質を持っている
- 小数点以下の値・負の値を取ることがある
- 平均値を計算して意味がある

## ■ 質的変数 Qualitative Variables

- 変数の値が、カテゴリーを区別・分類するために使われている
- 正の整数のみで、小数点以下の値・負の値にはならない
- 平均値を計算してもあまり意味がない

# 量的変数・質的変数・その中間の変数

- 量的変数 → 例えば: 年齢、身長
- 質的変数 → 例えば: 性別、血液型
- 「幸福度」という変数は？
- 量的変数と質的変数の中間的な性質を持っている
- 「①＝幸せ、②...③...④＝不幸せ」なので、数値が大きくなるほど、不幸を感じる度合いが強くなることから、「幸福度」は量的な性質を持っている
- しかし④を選んだ人は①を選んだ人より4倍不幸である、とは言えない; 値の順序には明確な意味があるが、値それ自体は幸福度を直接的に反映していないことから、「幸福度」は質的変数であるとも言える

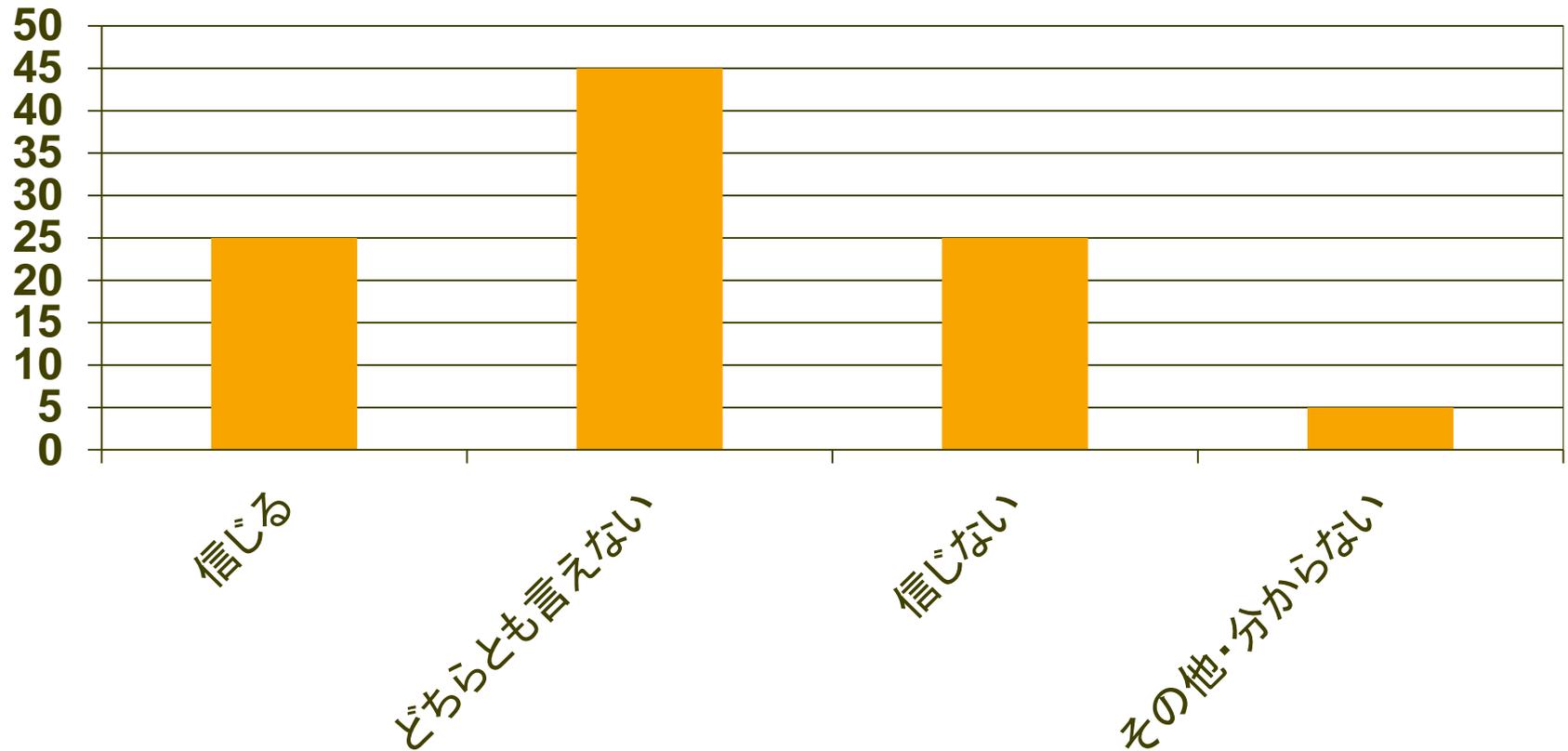
# データ分析とは

- データ:「**数値化された情報の集合**」のこと
- データ分析:「**変数**」を対象に行う
- 分析その1: 変数の性質・特徴を調べる
  - 例えば、平均「**身長**」を計算したり、最も多い「**血液型**」を調べたり
- 分析その2: 変数間の関係を調べる **←事前に仮説をしっかりと!**
  - 例えば、若い人ほど身長が高いのか (「**年齢**」と「**身長**」の関係)
  - 男と女とではどちらのほうが幸せを感じているか (「**性別**」と「**幸福度**」の関係)

# 度数分布表をグラフにする(棒グラフ)

- 「あなたは『神様』を信じますか」

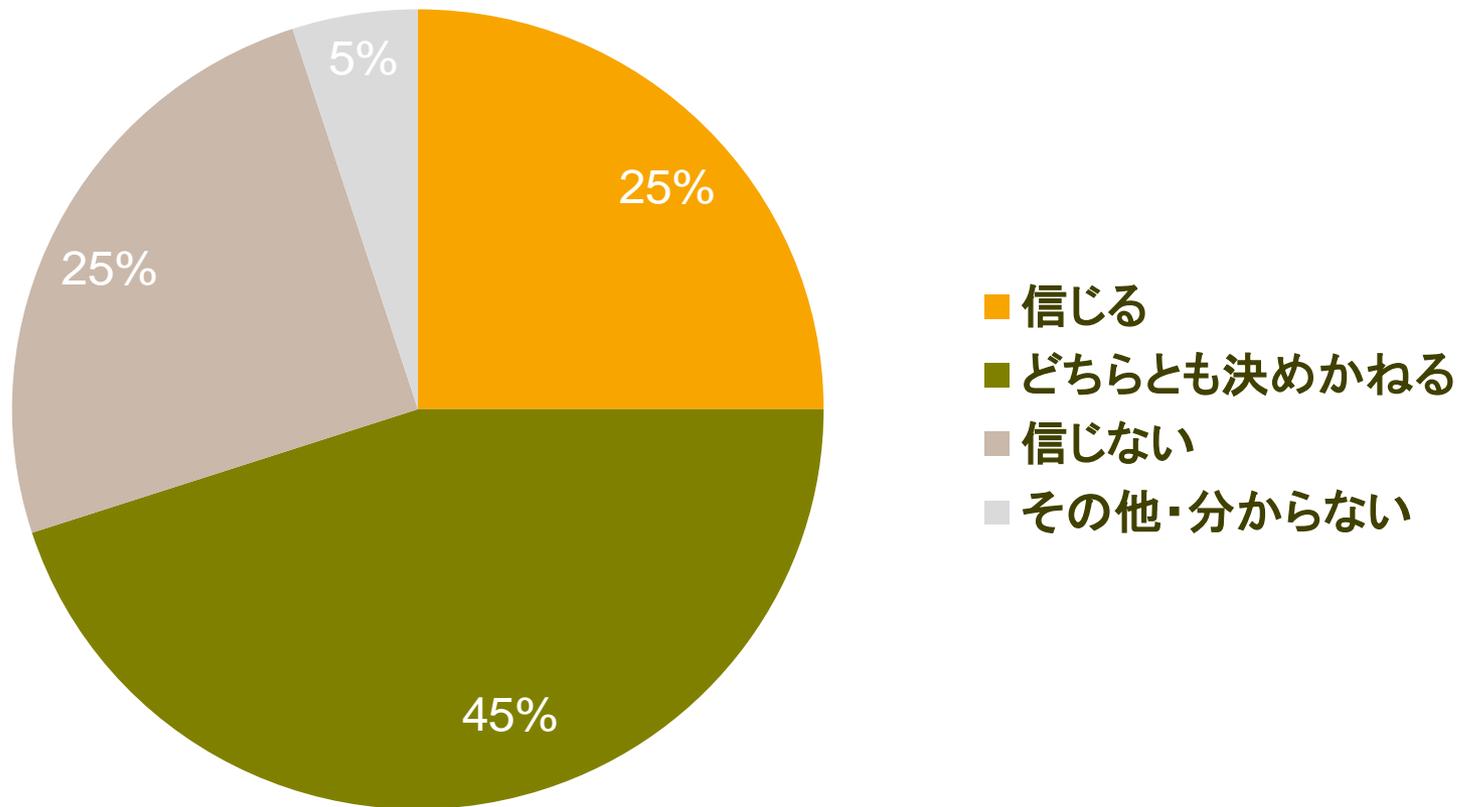
棒グラフ



# 度数分布表をグラフにする(円グラフ)

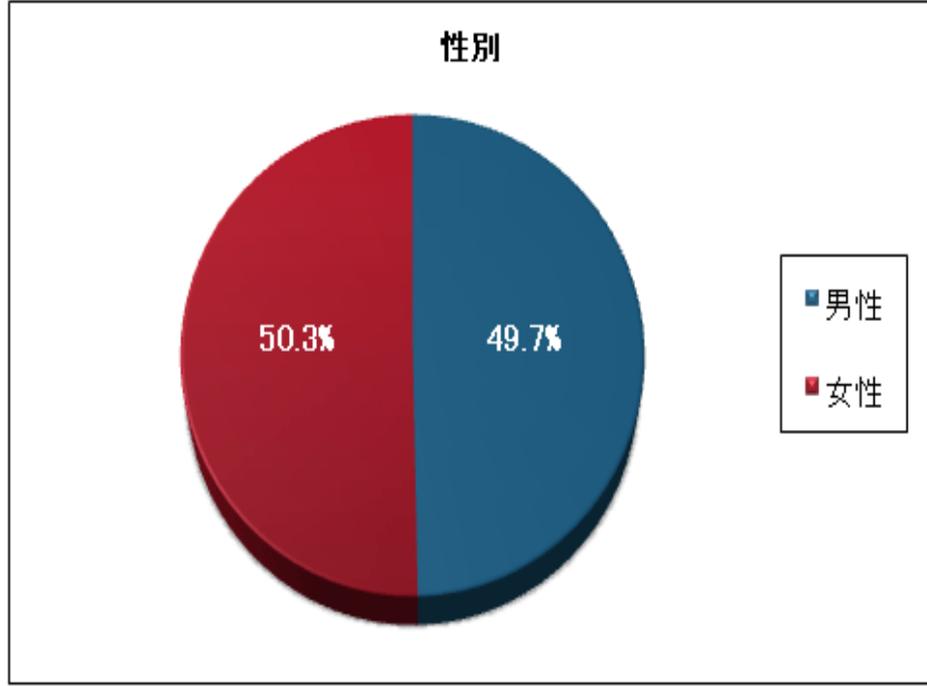
- 「あなたは『神様』を信じますか」

円グラフ



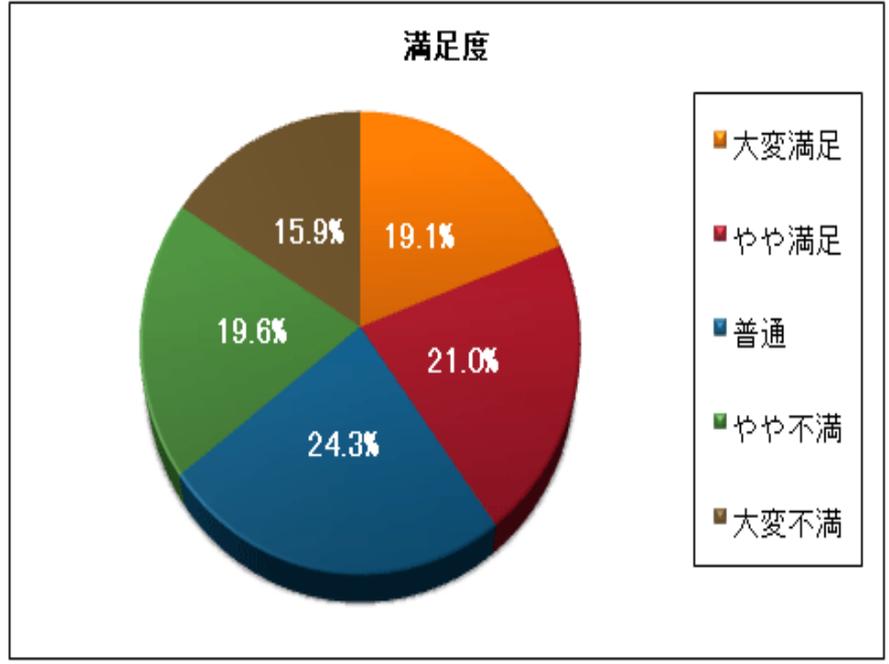
# 円グラフの他の例

	度数	%
男性	296	49.7%
女性	300	50.3%
合計	596	100.0%



満足度

	度数	%
大変満足	114	19.1%
やや満足	125	21.0%
普通	145	24.3%
やや不満	117	19.6%
大変不満	95	15.9%
合計	596	100.0%



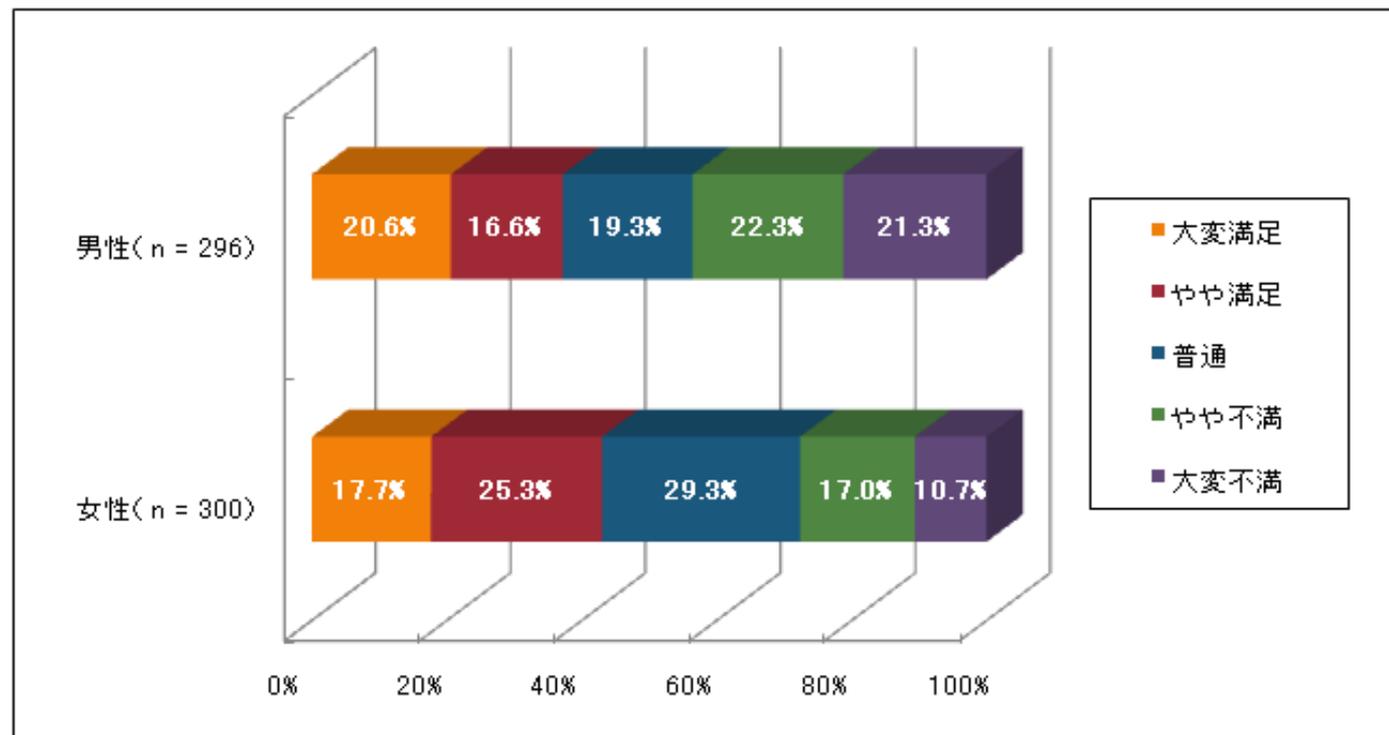
# 帯グラフ (性別 x 満足度)

性別 × 満足度

上段: 度数  
下段: %

	合計	大変満足	やや満足	普通	やや不満	大変不満
合計	596 100.0%	114 19.1%	125 21.0%	145 24.3%	117 19.6%	95 15.9%
男性	296 100.0%	61 20.6%	49 16.6%	57 19.3%	66 22.3%	53 21.3%
女性	300 100.0%	53 17.7%	76 25.3%	88 29.3%	51 17.0%	32 10.7%

これがクロス集計表！  
2重クロス表 (2変数のクロス表) もしくは単にクロス表ともいう

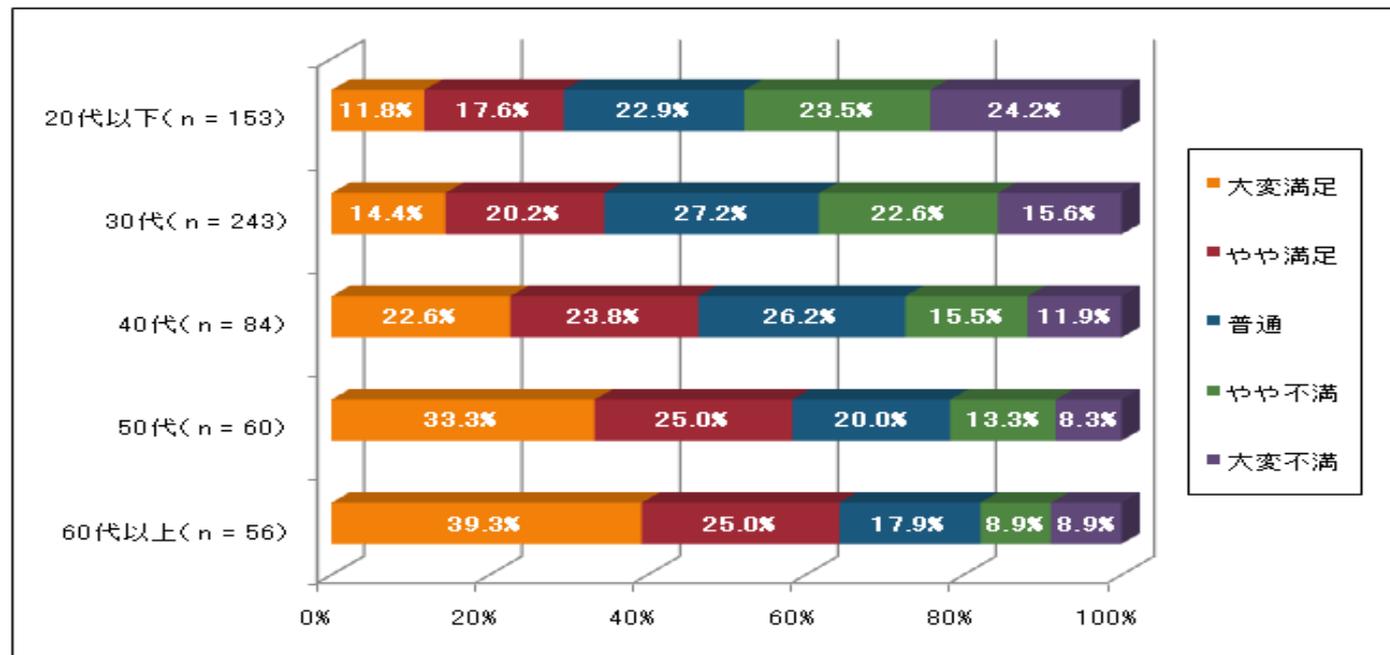


# 帯グラフ (年代 x 満足度)

これもクロス集計表！  
2重クロス表(2変数のクロス表)もしくは単にクロス表ともいう

年代 × 満足度

	合計	大変満足	やや満足	普通	やや不満	大変不満
合計	596	114	125	145	115	97
	100.0%	19.1%	21.0%	24.3%	19.6%	15.9%
20代以下	153	18	27	35	36	37
	100.0%	11.8%	17.6%	22.9%	23.5%	24.2%
30代	243	35	49	66	56	38
	100.0%	14.4%	20.2%	27.2%	22.6%	15.6%
40代	84	19	20	22	13	10
	100.0%	22.6%	23.8%	26.2%	15.5%	11.9%
50代	60	20	15	12	8	5
	100.0%	33.3%	25.0%	20.0%	13.3%	8.3%
60代以上	56	22	14	10	5	5
	100.0%	39.3%	25.0%	17.9%	8.9%	8.9%



# トリプルクロス表 (性別 x 年代 x 満足度)

3重クロス表  
ともいう

性別 × 年代 × 満足度

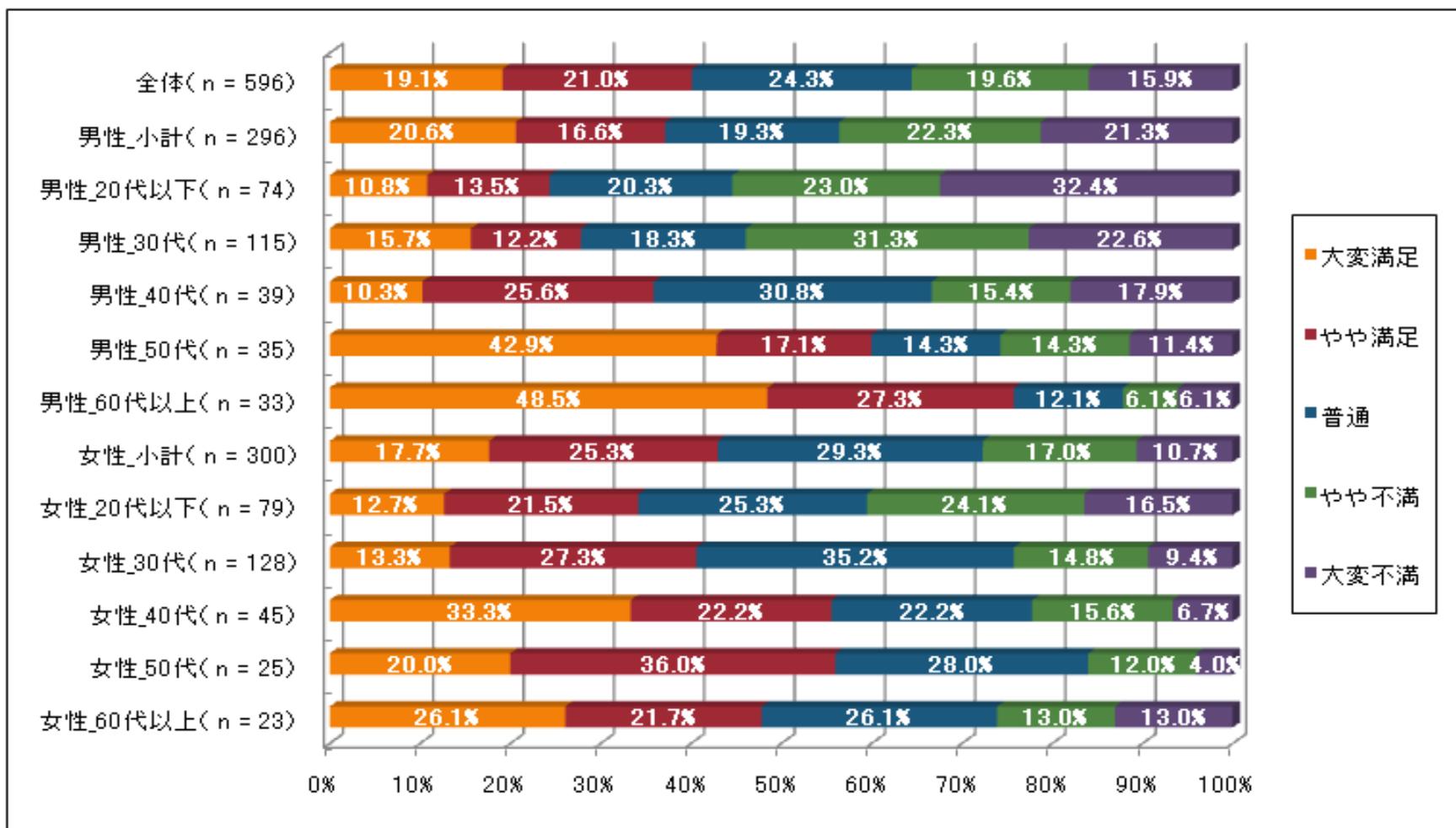
上段: 度数  
下段: %

		合計	大変満足	やや満足	普通	やや不満	大変不満
全体		596	114	125	145	117	95
		100.0%	19.1%	21.0%	24.3%	19.6%	15.9%
男性	小計	296	61	49	57	66	63
		100.0%	20.6%	16.6%	19.3%	22.3%	21.3%
	20代以下	74	8	10	15	17	24
		100.0%	10.8%	13.5%	20.3%	23.0%	32.4%
	30代	115	18	14	21	36	26
		100.0%	15.7%	12.2%	18.3%	31.3%	22.6%
	40代	39	4	10	12	6	7
	100.0%	10.3%	25.6%	30.8%	15.4%	17.9%	
50代	35	15	6	5	5	4	
	100.0%	42.9%	17.1%	14.3%	14.3%	11.4%	
60代以上	33	16	9	4	2	2	
	100.0%	48.5%	27.3%	12.1%	6.1%	6.1%	
女性	小計	300	53	76	88	51	32
		100.0%	17.7%	25.3%	29.3%	17.0%	10.7%
	20代以下	79	10	17	20	19	13
		100.0%	12.7%	21.5%	25.3%	24.1%	16.5%
	30代	128	17	35	45	19	12
		100.0%	13.3%	27.3%	35.2%	14.8%	9.4%
	40代	45	15	10	10	7	3
	100.0%	33.3%	22.2%	22.2%	15.6%	6.7%	
50代	25	5	9	7	3	1	
	100.0%	20.0%	36.0%	28.0%	12.0%	4.0%	
60代以上	23	6	5	6	3	3	
	100.0%	26.1%	21.7%	26.1%	13.0%	13.0%	

統制変数  
Control!

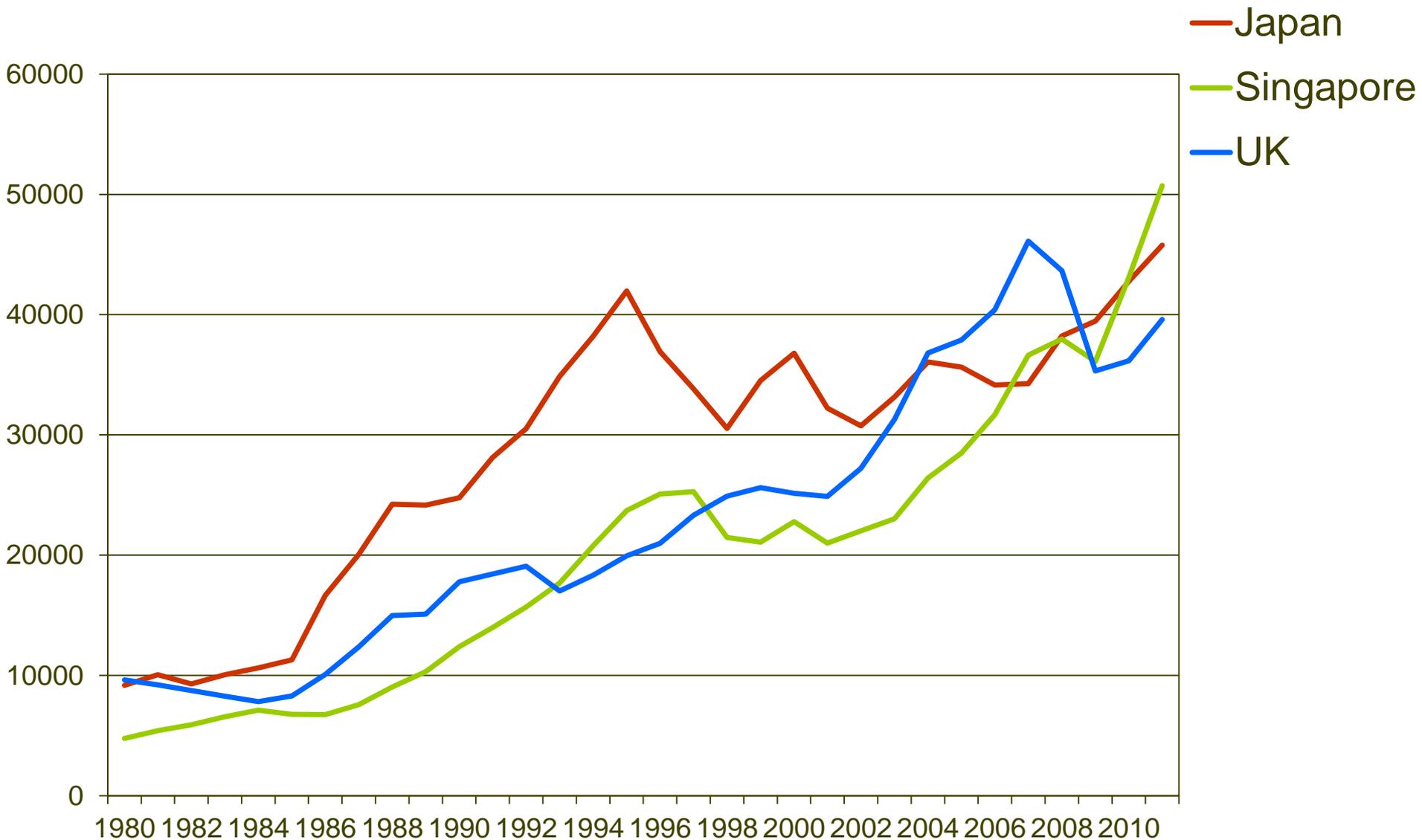
# トリプルクロス集計 (性別 x 年代 x 満足度)

## 3重クロス集計ともいう



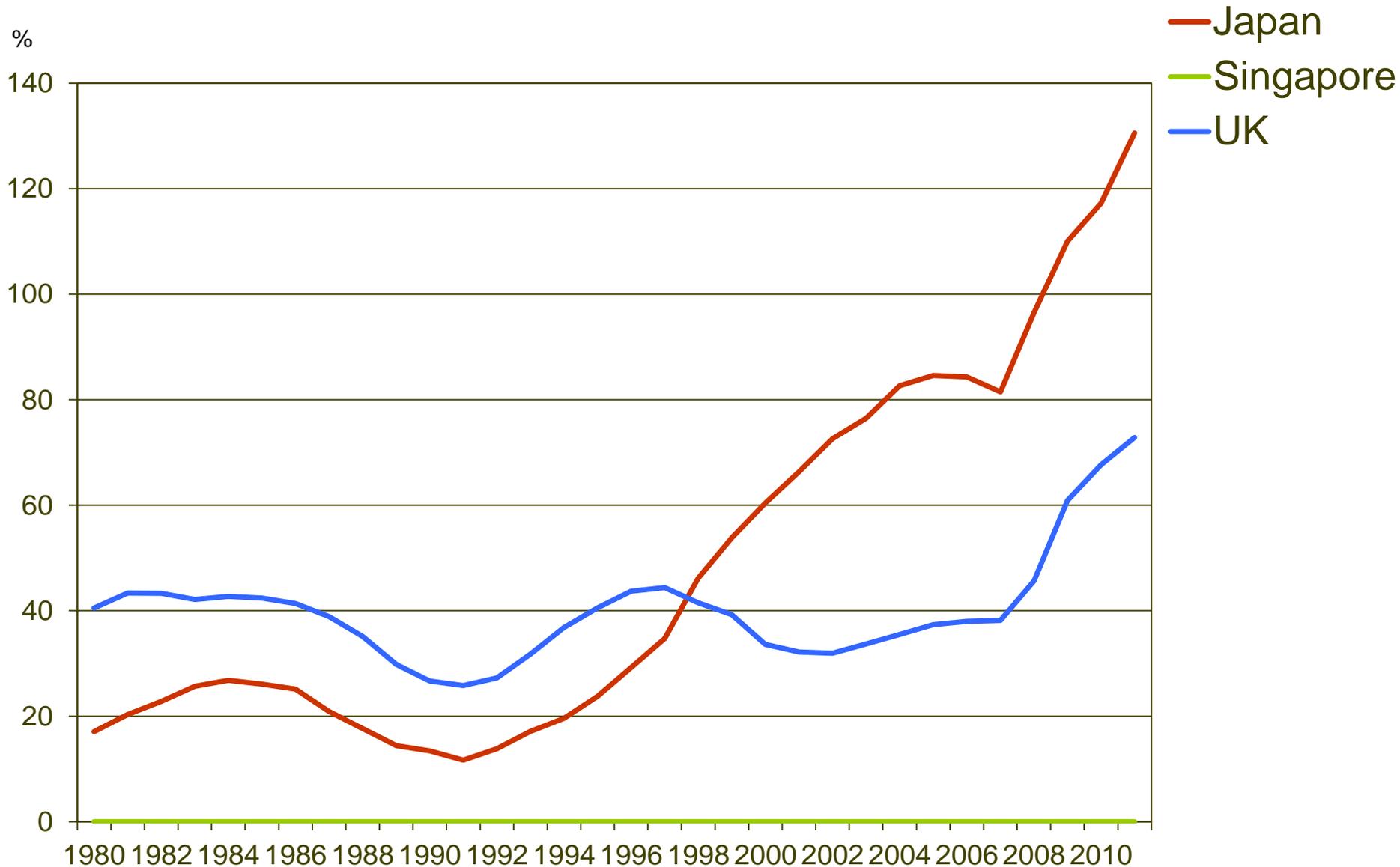
# GDP per capita (1980~2011)

Source: IMF World Economic Outlook 2011



# Government Net Debt by % of GDP

Source: IMF World Economic Outlook 2011



# どのグラフにすればいいのか

グラフの種類	表現したいこと	グラフのポイント
棒グラフ	ある数量(度数、平均、比率など)の大きさ	棒の長さが「数量の大きさ」を示す
円グラフ	各カテゴリーの全体に占める比率(構成比)	扇形の面積が比率の大きさを示す
帯グラフ	グループ別(性別、年代別など)の構成比	帯の面積が比率の大きさを示す
折れ線グラフ	ある数量の変化	折れ線の傾きが「変化の程度」を示す

# 6-1. 正規分布とは: まずは平均と分散

- (a) 5人で飲みで行って、それぞれ¥3000, ¥3500, ¥4000, ¥4500, ¥5000飲み食いした;さて、割り勘(平均)にすると、一人いくら?
  - $(¥3000 + ¥3500 + ¥4000 + ¥4500 + ¥5000) \div 5$   
 $= ¥20000 \div 5$   
 $= ¥4000$
- (b) 今度は5人で飲みで行って、それぞれ¥1000, ¥2000, ¥4000, ¥6000, ¥7000飲み食いした;割り勘にすると、一人いくら?
- 平均が同じでも、(a)より(b)のほうが分散が大きい
  - 分散: 散らばり、ばらつき
  - (a)より(b)のほうが「より散らばって分布している」

# 分散の求め方と標準偏差

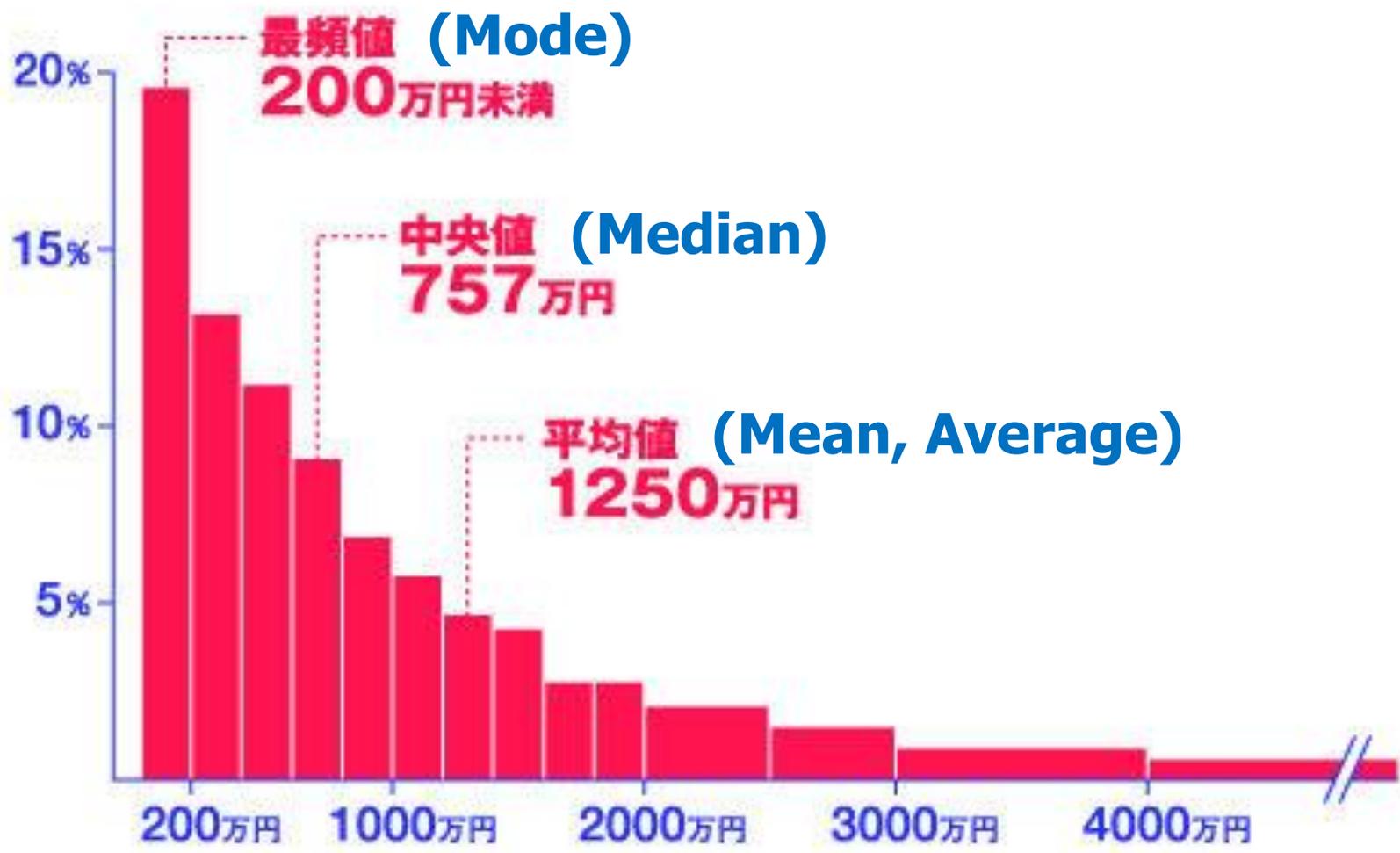
- 各ケース値と平均との差を二乗して、さらにその平均を
- (a) 5人で飲みで行って、それぞれ¥3000, ¥3500, ¥4000, ¥4500, ¥5000飲み食いした;さて、分散は?
  - $\{ (\text{¥}3000 - \text{¥}4000)^2 + (\text{¥}3500 - \text{¥}4000)^2 + (\text{¥}4000 - \text{¥}4000)^2 + (\text{¥}4500 - \text{¥}4000)^2 + (\text{¥}5000 - \text{¥}4000)^2 \} \div 5$
  - =  $\{ (-\text{¥}1000)^2 + (-\text{¥}500)^2 + (\text{¥}0)^2 + (\text{¥}500)^2 + (\text{¥}1000)^2 \} \div 5$
  - =  $\text{¥}^2 2,500,000 \div 5$
  - =  $\text{¥}^2 500,000$
- (b) の場合の分散は? → 答:  $\text{¥}^2 5,200,000$
- 分散の平方根を「標準偏差」という
  - (a) の標準偏差 =  $\sqrt{\text{¥}^2 500,000} = \text{¥} 707.11$
  - (b) の標準偏差 =  $\sqrt{\text{¥}^2 5,200,000} = \text{¥} 2280.35$

# 平均値・中央値・最頻値

- 平均値: 変数の「重心」と「中心」にあたる
  - 「真ん中」「普通」という意味ではない
  - 極端に大きい(小さい)「外れ値」(outlier)に要注意
- 中央値: ケースを値の小さい順に並べて真ん中にくるケースの値
- 最頻値: 文字通り、最も度数が大きい値

# 平均値と最頻値と中央値

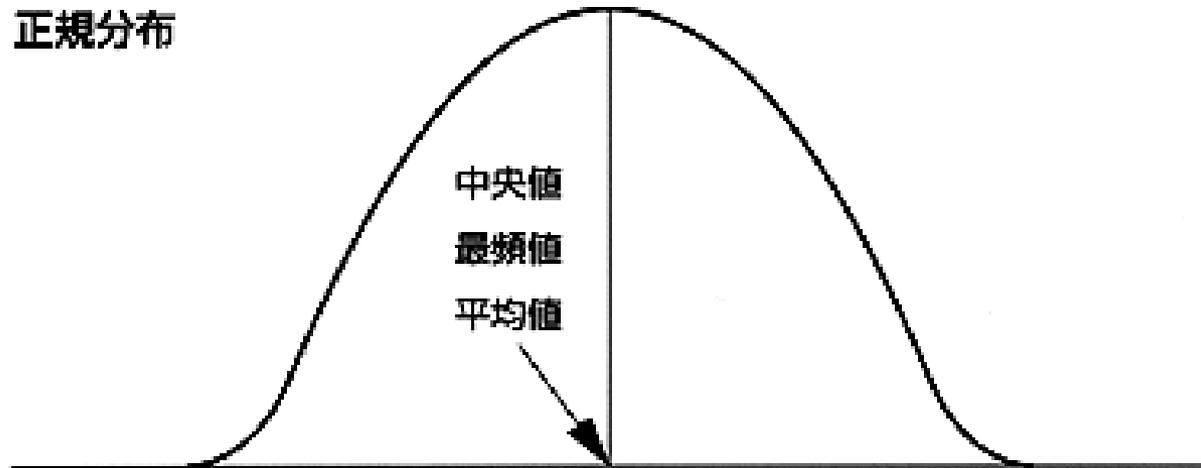
## 勤労者世帯の貯蓄額の分布



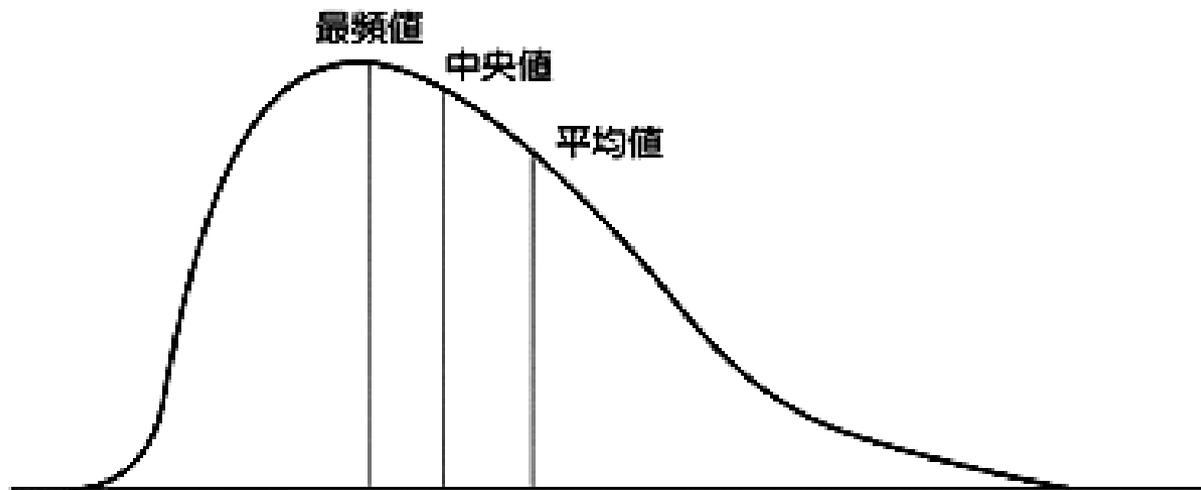
出典:総務省「家計調査」勤労者世帯の貯蓄現在高階級別世帯分布(2008年)

# 平均値=最頻値=中央値: 正規分布

正規分布



右に歪んだ分布



# 正規分布の重要性

1. 正規分布にしたがう自然・社会現象が多い
2. 正規分布を使うと、ある値が全体の中でどのあたりに位置するのか、確認することができる
3. 推定や検定の多くは、**母集団**や**標本平均の分布**が正規分布であることを前提にしている

## 6-2. 標準化

- 標準偏差は様々な応用が可能であるという点で重要；  
その一例が**標準化**である
- 例えば、センター試験で日本史を選択したA君と、  
世界史を選択したB君の点数が次のようだったとする：
  - A君の日本史の得点 = 73点  
(日本史の平均点 = 65点、標準偏差 = 17.0)
  - B君の世界史の得点 = 67点  
(世界史の平均点 = 59点、標準偏差 = 12.0)
- どっちがすごいですか

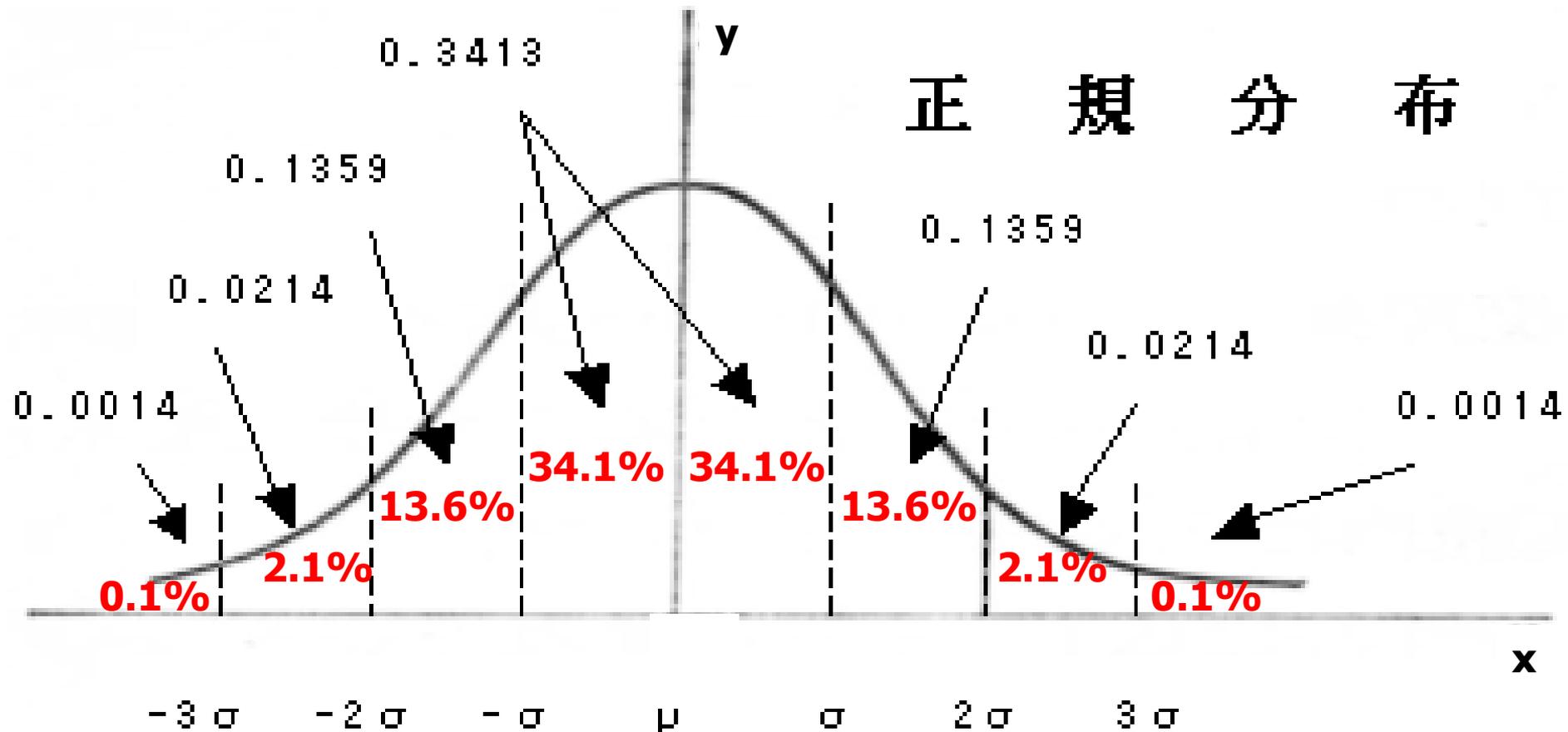
# 標準得点

- 変数を「平均=0、標準偏差=1」に変換
- 標準化された各ケースの値のことを**標準得点**と呼ぶ

$$\text{標準得点} = \frac{\text{得点} - \text{平均点}}{\text{標準偏差}}$$

- A君とB君の標準得点はそれぞれ何点になりますか  
(答え:A君=0.471, B君=0.667)
- B君のほうがすごい!

# 標準得点(正規分布)の性質



|| 平均点

|| 標準偏差

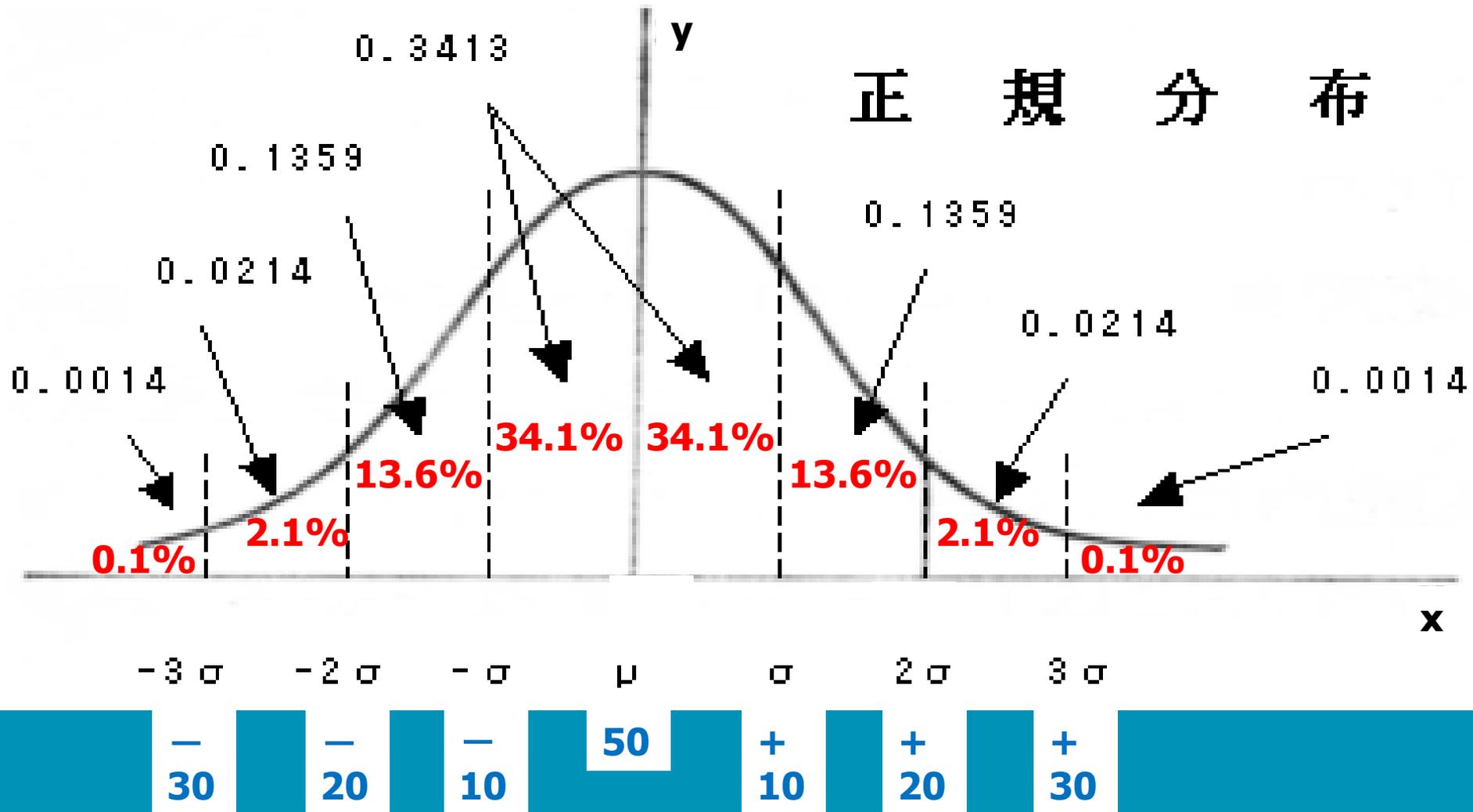
# 偏差値

- 日本で使われている偏差値は、以下のように計算される：

$$\text{偏差値} = 10 \times \text{標準得点} + 50$$

- 平均が50点、標準偏差が10点、つまり約2/3の生徒が40点から60点の間に入るように換算したもの
- >60点はトップ16% (1/6) ぐらい、<40点はボトム16% ぐらい

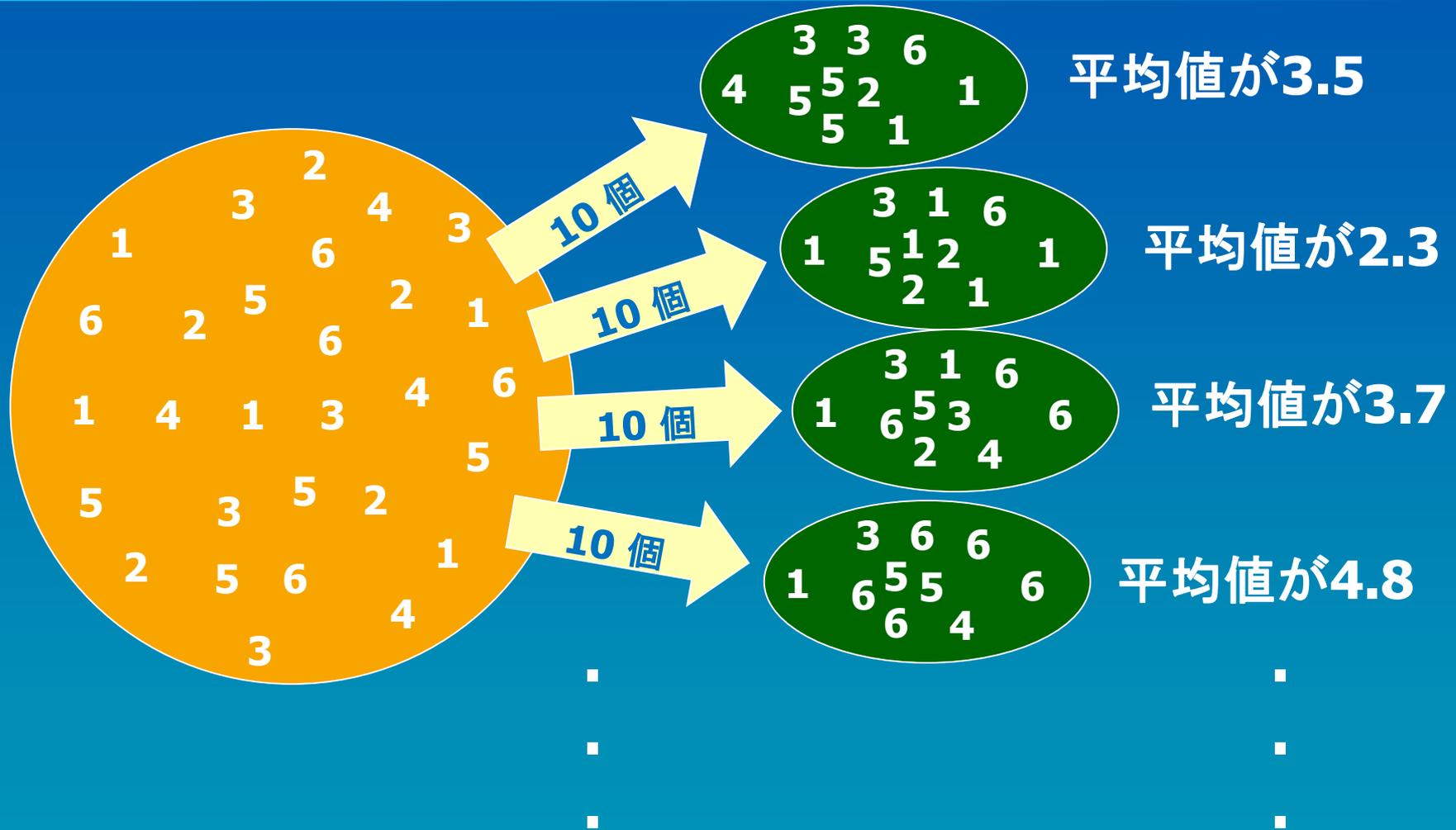
# 偏差值



# 推定はなぜ可能か：中心極限定理

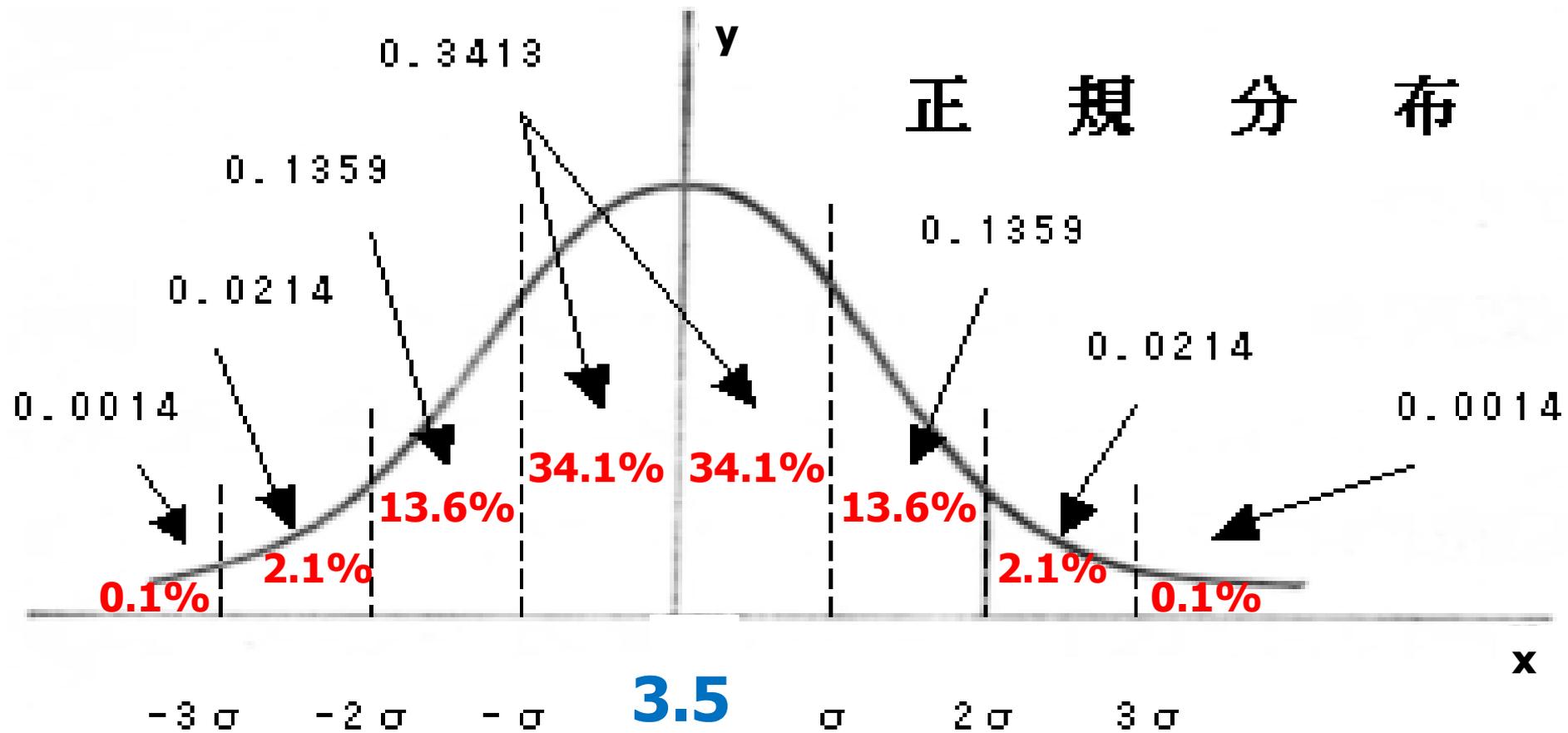
- **中心極限定理** Central Limit Theorem:  
標本平均の分布は、標本の大きさが大きくなると、それらが抽出された元の母集団が正規分布でなくとも、**正規分布に近づく**
- 中心極限定理は、「母集団から標本を抽出して標本平均を計算する」という操作を何度も行うことを前提にしている
- 標本平均の分布とは、1回目、2回目、3回目...  
...m回目に抽出した標本から得られた標本平均、  
によって作られた分布のこと

# 標本平均の分布



➤ 標本平均が母集団の平均に近づいていく

# 標本平均の分布(サイコロを使った例)



# 7. 仮説検定の考え方

- 母集団に関する仮説の真偽を標本調査データで検証することを、**仮説検定** (Hypothesis Test) という
- つまり、**調査仮説**が成り立つかどうかを検定する
- 調査仮説を否定する仮説を**帰無仮説** (Null Hypothesis,  $H_0$ ) という (Null = 0)

# なぜ帰「無」仮説なのか？

- 帰無仮説は調査仮説を否定する仮説のことで、「平均値に差が**無い**」「変数間の関連が**無い**」といったように、調査仮説が述べる傾向が実は存在しないという内容を述べた仮説である
- 社会学を含め、科学的研究では何かに違いがあることを主張したいことがほとんどである
  - 新しく開発した薬が従来の薬よりも効き目が大きい**(差がある)**と主張したい薬学の研究
  - ある政策を施行することが経済効果をもたらす**(差を作りだす)**と主張したい経済学の研究
  - 親の地位によって子どもの大学進学のお機に**差が見られる**と主張したい教育社会学の研究...などなど

# 帰無仮説

- しかし残念なことに、統計学では「**差がある**」ことを直接的に証明することはできない;なぜなら、差がある状態を無数に想定できるからである  
→ 差が 0.8 の場合、2 の場合、-4.7 の場合... (**無限!**)
- 一方、**差が無い**(**差がゼロである**)状態は1つしかない
- 「**差が無い**」とか「50%である」というような、状態が1つに定まった仮説ならば、しっかり検討することができる

# 「稀だよ」を評価する有意水準

- 「差がない」( $H_0$ )の下で、調査仮説(差がある)の確率が低ければ(稀)、帰無仮説は棄却され、調査仮説が採択される(差がある！)
- 反対に確率が高ければ、帰無仮説は棄却されず、調査仮説は採択されない(差があるとは言えない)
- 有意水準(p値、有意確率)と呼ばれる目安を決めて、帰無仮説が棄却されるかどうかの判断を行う
- 有意水準としては、通常 **10%、5%、1%、0.1%** などといった区切りのよいパーセンテージが選ばれる
- 因みに、コイン4枚が同時に表になる確率は6.25%で、5枚だとその確率は半分の3.125%になる

# 因みに...

- 統計的に有意であることは、統計的検定の結果として帰無仮説が棄却されるべきだという意味を有するというだけで、分析結果が有意義だとか、学問的に意味があると保証された、などの意味ではない
- 統計的に有意である = 正解でもない

# 8. 統計分析法いろいろ

- クロス表における2つのグループの比率差の検定  
カイ2乗  $\{\chi^2\}$  検定 (Chi-Square Test)
  - 調査仮説: 2つの母集団の比率に差がある
  - 帰無仮説: 2つの母集団の比率に差が**無い**
  - 内閣支持率は1ヶ月前と比べて本当に下落したのか
  - 原発再稼働について、女性のほうが反対する傾向が高いか

# 2つの平均値の差：t検定(t-test)

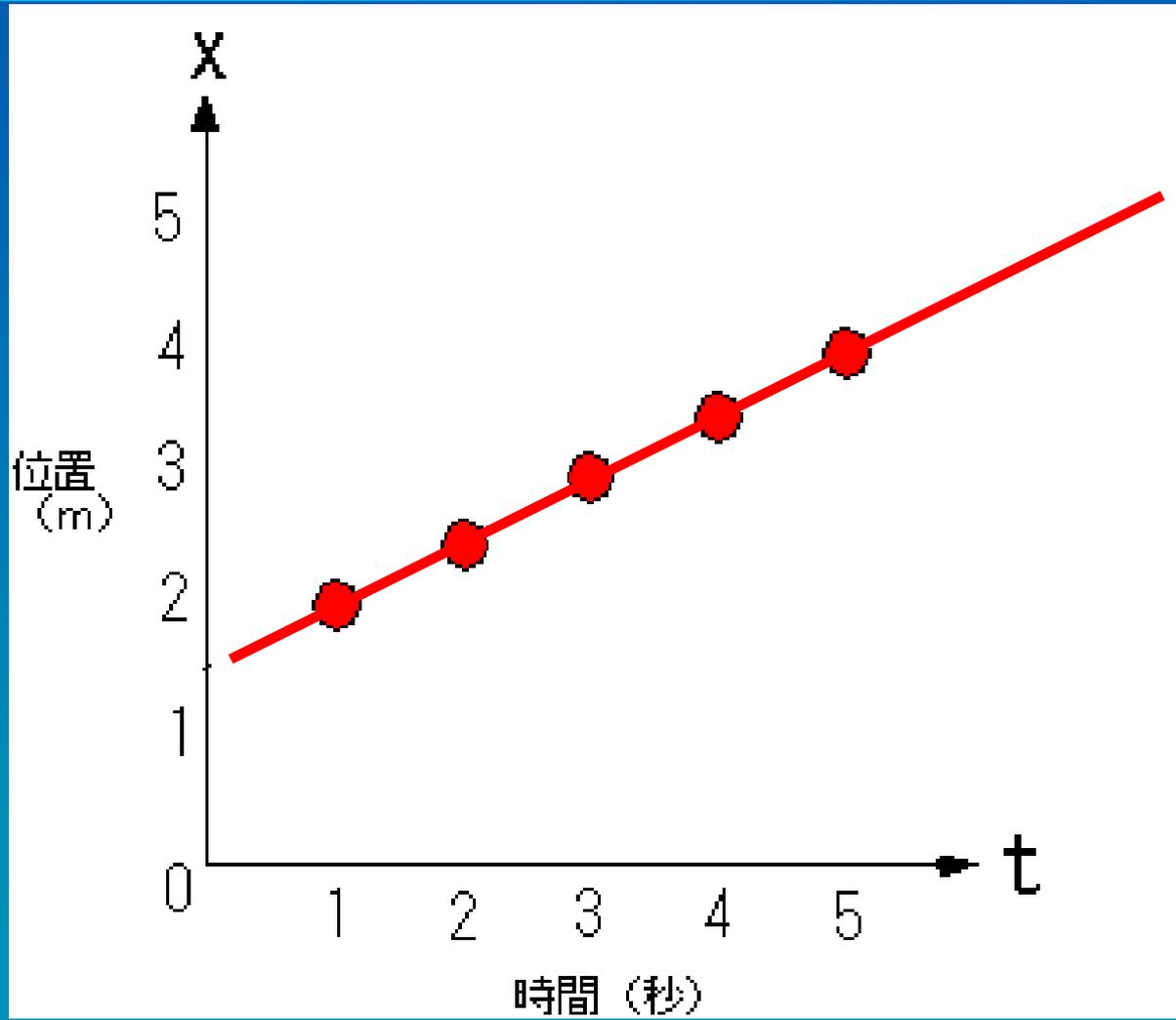
## ■ 平均の差の検定

- 調査仮説：A市とB市の平均世帯収入には差がある  
帰無仮説：A市とB市の平均世帯収入には差が**無い**
- その他に、男女間での所得格差、学力は低下したか、親の階層によって子供の勉強時間に差があるか、など

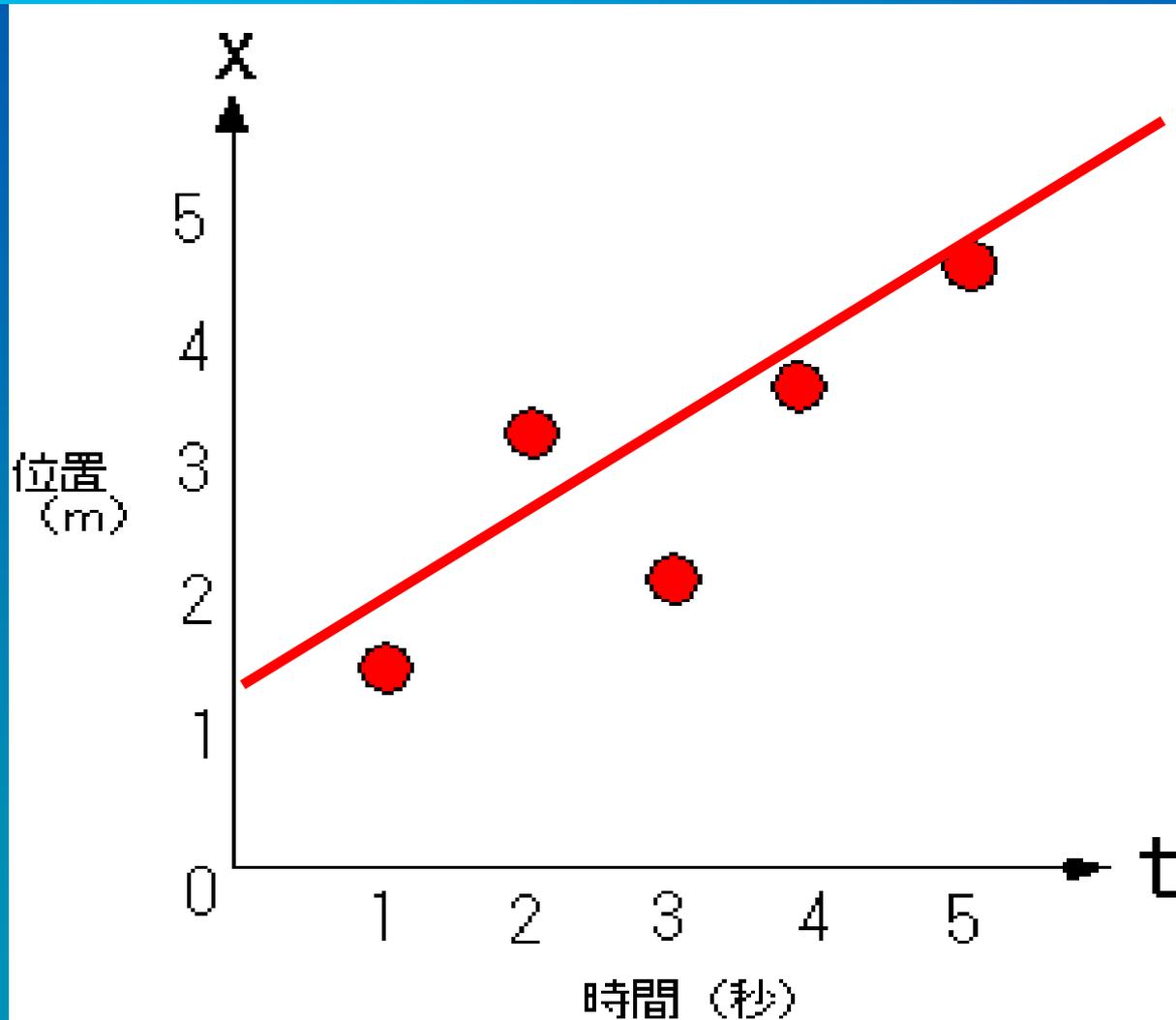
# 3つの平均値の差：分散分析(ANOVA)

- 調査仮説：すべてのグループの平均値のうち、少なくとも1つの平均値は等しくない
- 帰無仮説：すべてのグループの平均値に差が**無い**
- 途中計算で「分散」を使うため、分散分析という名前がついただけで、実質的には平均値を比較する手法
  - 異なる雇用形態(正規・非正規・自営)によって、平均睡眠時間が違うのかどうか
  - 学校タイプによって、平均勉強時間が違うのかどうか
- 全体として見た場合、各グループの平均値が違うかどうかを確かめているだけで、どのグループとどのグループの間で差があるのかは分からない

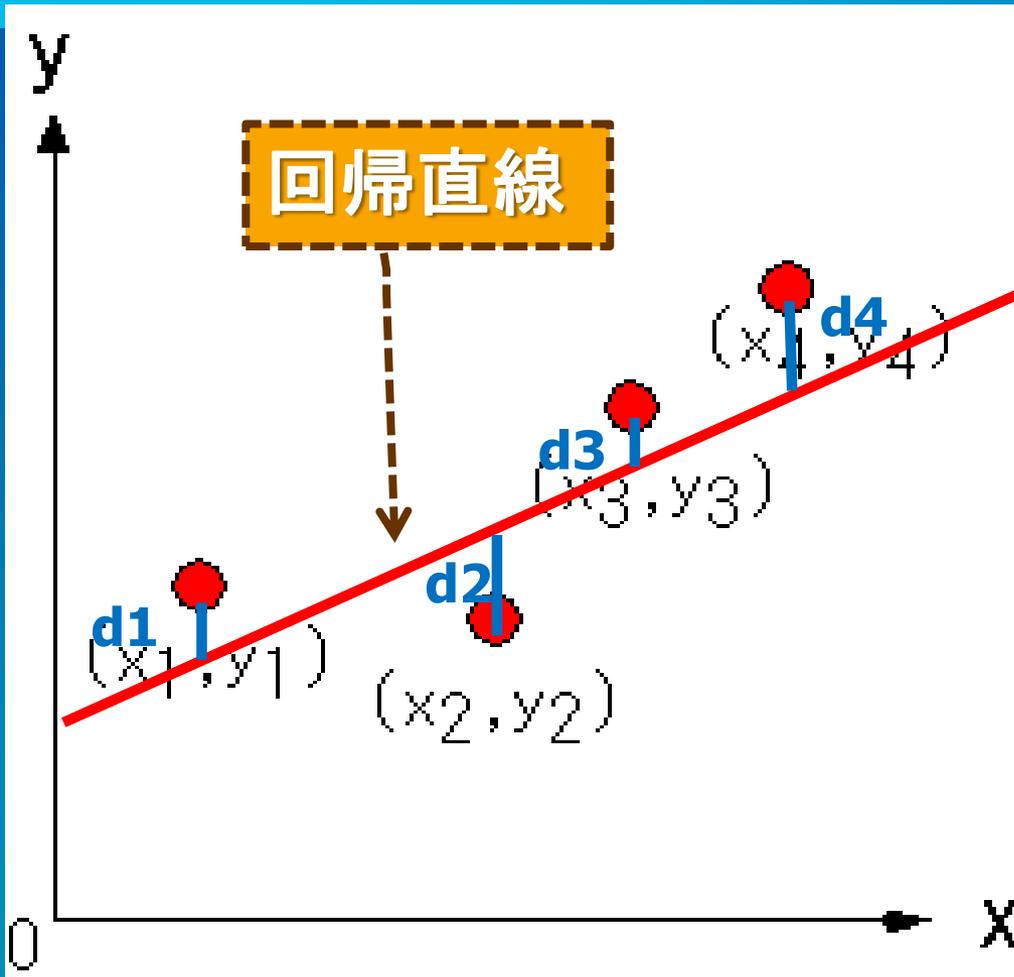
# 回歸分析 Regression Analysis



# 回歸分析 Regression Analysis



# 最小2乗法



( $d_1^2+d_2^2+d_3^2+d_4^2$ )の  
値を最小にするような直線  
 $Y=aX+b$  (回帰式)が  
「もっともらしい直線である」  
と考える

なお、 $d_1$ 、 $d_2$ 、 $d_3$ 、 $d_4$ はそれ  
ぞれの観測値と予測値との  
「誤差」を意味する

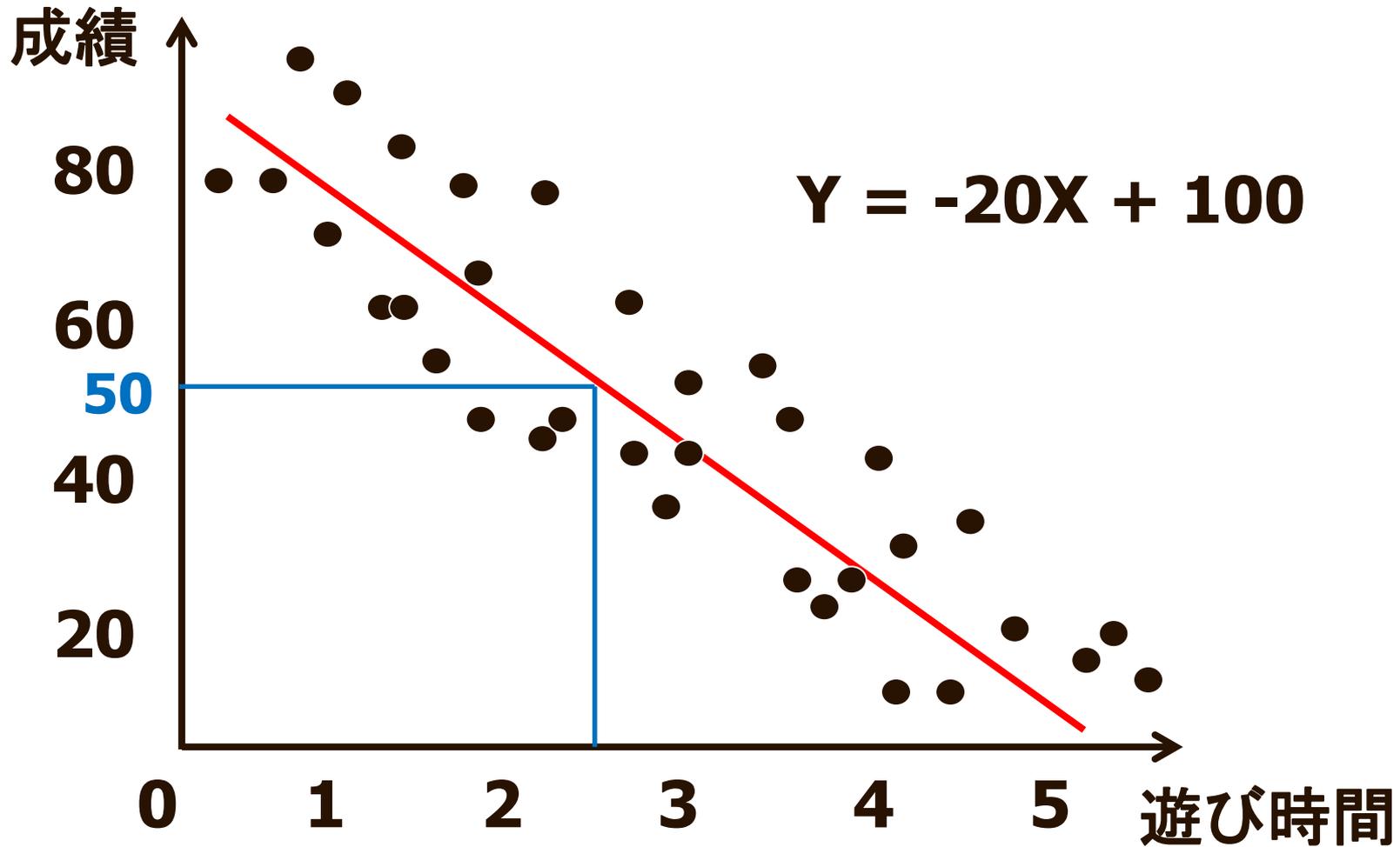
- 最小2乗法は誤差の2乗を合計した値が最小になる  
ように回帰式の  $a$ (回帰係数)と  $b$ (切片)を求めること

# 回帰分析 Regression Analysis

- 変数  $X$  と  $Y$  の散布図において点の分布を貫く直線を引き、この直線の式を求める分析のこと
- $X$  の値で  $Y$  の値を予測するときに使う  
(回帰式の  $X$  に具体的数値を代入して、 $Y$  がどんな値をとるかを予測)
- $X$  = 独立変数 (原因にあたる変数のこと)、  
 $Y$  = 従属変数 (結果にあたる変数のこと) という
- $X$  = 説明変数、 $Y$  = 被説明変数 と呼ばれることもある

# 回帰分析 Regression Analysis

- 例：学力と遊び時間の関係



# 回帰分析の精度を測る決定係数

- 前の例では、回帰直線と回帰式を使って「遊び時間が2.5時間の生徒の成績は50点になるだろう」というもっともらしい予測値をはじき出しても、回帰直線と回帰式のあてはまりが悪ければ、予測値は信用に値しない
- 回帰直線・回帰式のあてはまりのよさは、決定係数で評価できる

$$\text{決定係数 } R^2 = r_{XY}^2 \text{ (相関係数の2乗! )}$$

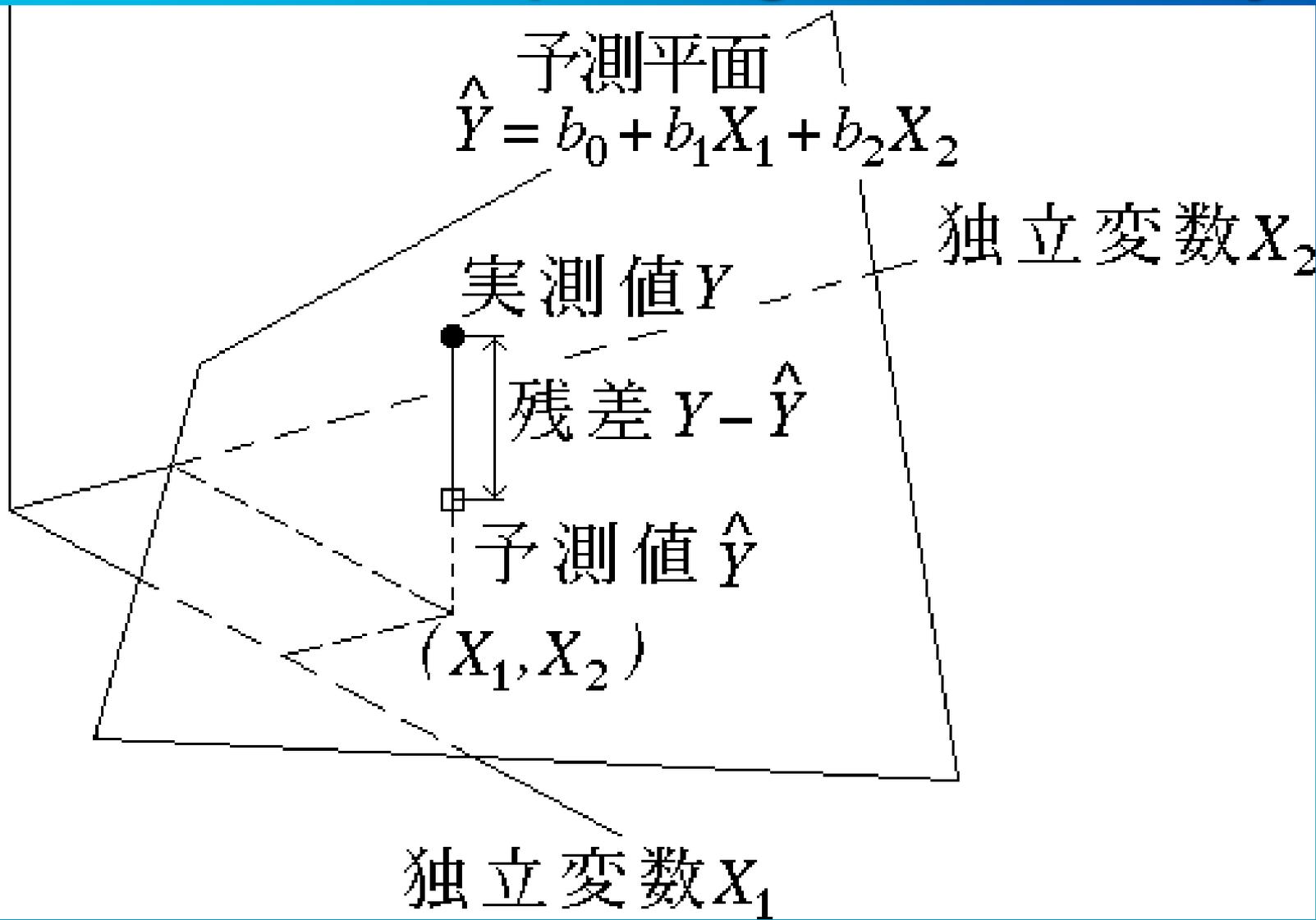
- 決定係数は0 ~ 1の値をとり、1のとき最もあてはまりが良く、0のとき最もあてはまりが悪い

# 重回帰分析 Multiple Regression Analysis

- 回帰分析: 独立変数  $X$  が1つ
- 重回帰分析: 独立変数  $X$  が2つ以上 ( $X_1$ 、 $X_2$ 、 $X_3$  ...)
- 独立変数が2つのとき、予測する重回帰式は以下のようになる
  - $Y = b_0 + b_1X_1 + b_2X_2$
- 独立変数が  $m$  個のとき、予測する重回帰式は以下のようになる
  - $Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \cdots + b_mX_m$

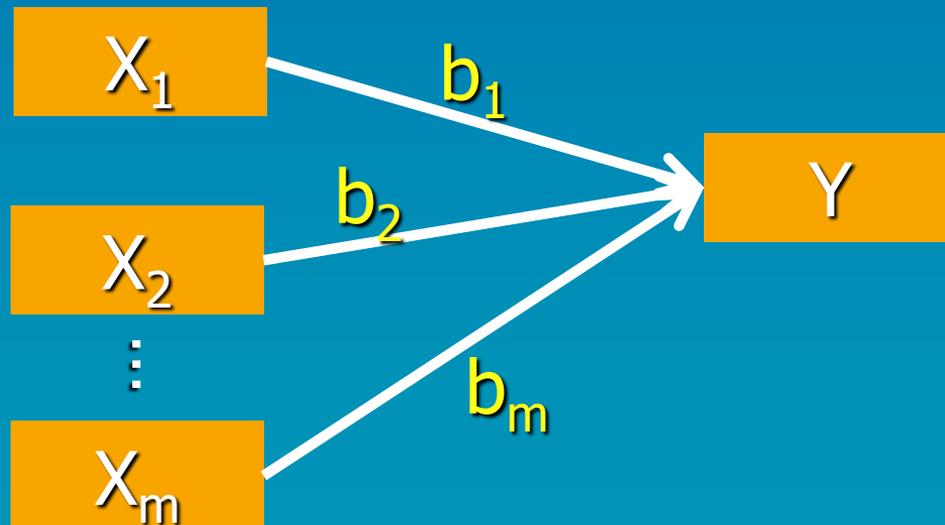
# 重回帰分析 Multiple Regression Analysis

従属変数 Y



# 偏回帰係数

- 重回帰のとき、 $b_1$ 、 $b_2$ ... $b_m$  は偏回帰係数とよばれる ( $b_0$ は切片)
- 偏回帰係数は、他の独立変数の影響を取り除いたときに(コントロール!)、ある独立変数の値が「1」だけ変わったとき、従属変数 $Y$ の予測値がどれだけ変化するかを表わしている



# 標準化偏回帰係数＝ベータ係数

- それぞれの独立変数  $X_1$ 、 $X_2$ 、 $X_3$  ... の単位や分散が違いため、全ての変数をあらかじめ標準化しておいて、その上で重回帰分析を行うことが多い；こうすると、変数の分散が全て1に統一されるので、ある意味で「標準化した観点」から偏回帰係数の大きさを考察 することができる
- このようにして求まる**標準化偏回帰係数**は、慣習的に**ベータ係数**ともよばれ、通常の偏回帰係数と一緒に出力されることが多い

# その他の統計分析法

- **ロジスティック回帰分析 Logistic Regression Analysis**
  - 重回帰分析の従属変数(Y)が量的変数であるのに対し、ロジスティック回帰分析の従属変数はダミー変数(1か0)
  - 回帰係数よりもオッズ比が重要
  - **オッズ比 Odds Ratio:**

	ビール好き	好きでない	オッズ	オッズ比
男	60%	40%	1.5	1.5 ÷ 0.25 = 6
女	20%	80%	0.25	

# その他の統計分析法

- 因子分析 Factor (要因、要素) Analysis
  - 多数の測定値を少数の未知な因子によって説明することを目的とする

第一因子	測定値	第二因子	測定値
外向 自信型	$X_1$		
	$X_2$		
	$X_3$		
		外見 自慢型	$X_4$
			$X_5$
			$X_6$

# その他の統計分析法

## ■ 主成分分析 Principal Component Analysis

- 多数の測定値を少数の総合的指標にする
- 国語の学力 =  $a_1$ 理解力 +  $a_2$ 分析力 +  $a_3$ 表現力
- 英語の学力 =  $b_1$ 理解力 +  $b_2$ 分析力 +  $b_3$ 表現力
- 数学の学力 =  $c_1$ 理解力 +  $c_2$ 分析力 +  $c_3$ 表現力
- 理科の学力 =  $d_1$ 理解力 +  $d_2$ 分析力 +  $d_3$ 表現力
- 地理の学力 =  $e_1$ 理解力 +  $e_2$ 分析力 +  $e_3$ 表現力
- 社会の学力 =  $f_1$ 理解力 +  $f_2$ 分析力 +  $f_3$ 表現力

.

.

.

このとおりやればすぐできる

# 社会調査

のための 神林博史・三輪哲◎著

# 統計学

生きた  
実例で  
理解する

基礎からやさしくわかる  
現場の統計学



## 「なんでもあり」の社会学を データ分析するには？

中学生でもわかる計算で  
社会調査がおもしろくなる

改訂版 朝日新聞社

## 朝日おとなの 学びなおし!

社会学

文系でもわかる

# 統計分析

すらすら  
読める!

東京大学大学院  
教育学研究科研究員

須藤康介

東京大学大学院  
総合文化研究科博士課程

古市憲寿

東京大学大学院  
教育学研究科教授

本田由紀

## 統計分析って、 使える おもしろい!

リアルなく教育調査データ>を  
使った実践本。統計ソフトSPSS対応



朝日新聞社

「習得保証コンテンツ」

# MULTIVARIATE ANALYSIS

まずはこの一冊から

## 意味がわかる 多変量解析

大人のための数学教室「和」講師

●●● 石井俊全 著  
Toshiaki Ishii

ストーンと胸に落ちる  
徹底解説で、  
これから  
学び始める人に  
最適の一冊!

- 手元に膨大なデータがあるけれど、  
どう分析してよいかわからない!
- 多変量解析のソフトの中では何が  
行なわれているのか、その原理が知りたい!

そんな、文系・理系出身者の方々に向け、必ずしも数式を覚まなくても  
多変量解析の原理を理解できるように解説していきます。

ビッグデータ時代の必読書!

「習得保証コンテンツ」



日本統計学会公式認定  
統計検定②級対応

# 統計学基礎

日本統計学会 編

東京図書

# 3 Kinds of Lies 三つの嘘？

- しかし19世紀のイギリス首相ベンジャミン・ディズレーリはこう言った：
- There are 3 kinds of lies 世の中には三つの嘘がある
  1. Lies 嘘
  2. Damn Lies 真っ赤な嘘
  3. and... Statistics! そして、統計！



Thank You!

