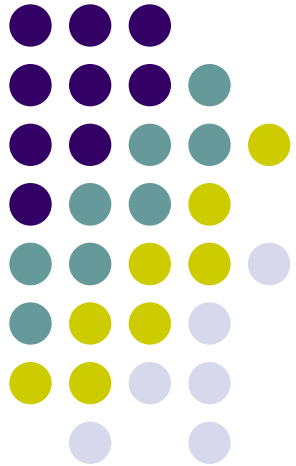
長崎大学医学部
臨床特論（臨床疫学、医療情報学）
平成24（2012）年4月　対象5年

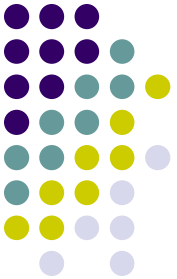
# 医療統計学のエッセンス
## ー統計的仮説検定・信頼区間の推定を学ぶー

長崎大学大学院・医歯薬学総合研究科
社会医療科学講座・医療情報学分野
（病院・医療情報部）

本多正幸

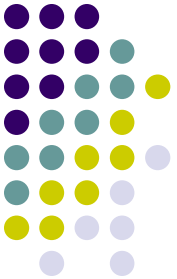


# 話の進め方(1)



- 導入(参考としたテキスト紹介)
- データの種類と尺度
- パラメトリック統計とノンパラメトリック統計
- 平均値と中央値
- 変動の測度(標準偏差と標準誤差)
- 正規分布と偏差値
- 標本平均の分布
- 区間推定ー信頼区間の構築

# 話の進め方(2)



- 統計的仮説検定
  - ー検定の例、概説
  - ー１標本における平均値に関する推測
  - ー第１種の過誤と第２種の過誤
  - ー片側検定と両側検定
  - ー標本の大きさの決定
  - ー統計的有意性と医学的有意性
  - ー対応がない場合のt検定
  - ー対応がある場合のt検定
  - ー分散分析
- CRAP　Detectors
- 医学関係論文における統計解析の記述のガイドライン

# 統計学とは

- 集団の特性や状況を数字で表し、そこに潜む規則性や傾向性を明らかにする学問

- 例：喫煙と肺ガン

| | | 肺ガン死亡者 | 死亡率（人口10万対） | リスク |
|---|---|---|---|---|
| 喫煙群 | 10万人 → | 227人 | 227／10万 × 10万 | 32.4倍 |
| 非喫煙群 | 10万人 → | 7人 | 7／10万 × 10万 | |

# 参考書


中野正孝，他：JJNスペシャル看護研究のための統計学入門，医学書院

中野正孝，他：系統看護学講座基礎8情報科学，医学書院

丹後俊郎：新版医学への統計学，朝倉書店

宮原・丹後編：医学統計ハンドブック，朝倉書店

中野正孝，他：論文が読める！早わかり統計学，メディカルサイエンスインターナショナル

野尻雅美，他：論文が読める！早わかり疫学，メディカルサイエンスインターナショナル

丹後俊郎：統計学のセンスデザインする視点・データを見る目，朝倉書店

# 論文が読める！早わかり統計学

メディカルサイエンスインターナショナル

副題：臨床研究データを理解するためのエッセンス

英文：PDQ　Statistics

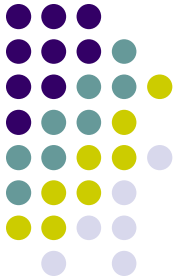
Geoffery R. Norman, David L. Streiner 著

中野正孝、本多正幸、宮崎有紀子、野尻雅美　訳

メディカル・サイエンス・インターナショナル社発行

スライドの上に
PDQと表示

# ラーメンの嗜好アンケートについて



アンケート結果1

| ラーメンの種類 | |
|---|---|
| しょうゆラーメン | 15 |
| とんこつラーメン | 20 |
| 味噌ラーメン | 10 |
| 塩ラーメン | 5 |

アンケート結果2

| ラーメン好き嫌い | |
|---|---|
| 大好き | 15 |
| 好き | 20 |
| 普通 | 10 |
| 嫌い | 5 |

平均値？意味？

# データの種類と尺度

| データの種類 | 尺度 | 四則演算 |
|---|---|---|
| 質的データ | 名義尺度 | 不可 |
| | 順序尺度 | 不可 |
| 量的データ | 間隔尺度 | 加算,減算のみ |
| | 比尺度 | 可 |

- データと集計方法（1変数の場合）
- 概念枠組みと仮説検証の方法
- データと集計方法（2変数の場合）

# データと集計方法（1変数の場合）

- 喫煙量についての質問例
  質問1．あなたのタバコを吸う量は他の人に比べてどうですか？
  {1．多い　2．どちらでもない　3．少ない}
  質問2．あなたの1日の喫煙量は？
  {1日に、＿＿＿本}

質的データの場合

| 喫煙量 | 人数 | % |
|---|---|---|
| 多い | 20 | 40 |
| どちらでもない | 25 | 50 |
| 少ない | 5 | 10 |
| 計 | 50 | 100 |

量的データの場合

| 喫煙量 | 人数 | % |
|---|---|---|
| 9本以下 | 8 | 16 |
| 10～19本 | 12 | 24 |
| 20～29本 | 18 | 36 |
| 30～39本 | 7 | 14 |
| 40本以上 | 5 | 10 |
| 計 | 50 | 100 |

※平均,標準偏差の計算可

# 概念枠組と仮説検証の方法

検証したいこと：産褥体操は身体的回復を促進する

| | （概念） | （概念） | 一般的仮説 |
|---|---|---|---|
| | ↓ | ↓ | 作業仮説 |
| 統計的検証 | （指標） | → | （指標） |
| A案：産褥体操をどの程度実施？ {1.よく実施　2.あまり実施しない} | | | 子宮硬度 {1.良好　2.不良} |
| B案：産褥体操をどの程度実施？ {1.よく実施　2.あまり実施しない} | | | 子宮硬度 （＿＿横指） |
| C案：産褥体操を何項目実施？ （＿＿＿項目） | | | 子宮硬度 （＿＿横指） |

**A案：　質的データ×質的データ**

**B案：　質的データ×量的データ**

**C案：　量的データ×量的データ**

# データと集計方法（2変数の場合）

**A案：質的データ×質的データ**

**クロス集計表（%）**

| | | | |
|---|---|---|---|
| | | | |
| | | | |
| | 14.9 3 | 1.7 | 15.100 |
| | 17.8 9 | 2.1 | 19.100 |

**クラマ-の関連係数**

**φ係数**

**$\chi^2$検定**

**比率の差の検定**

**B案：質的データ×量的データ**

**平均値±標準偏差**

**実施状況　　子宮底長**

**よく実施　　5.2 ±3.8**

**少し実施　　6.1 ±4.5**

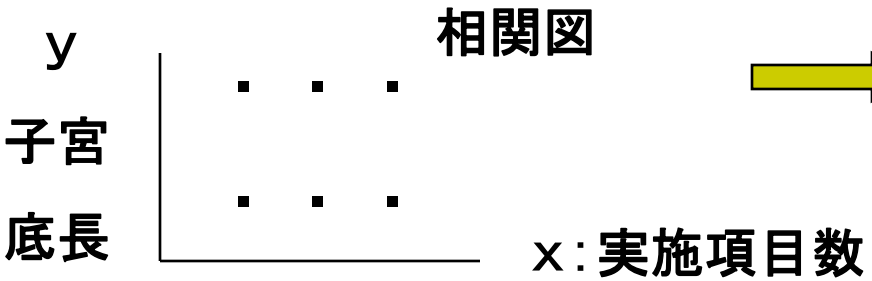
**相関比**

**相関比の検定**

**平均値の差の検定**

**一元配置分散分析**

**C案：量的データ×量的データ**




**相関図**

**相関係数**

**相関係数の検定**

# 変数の型

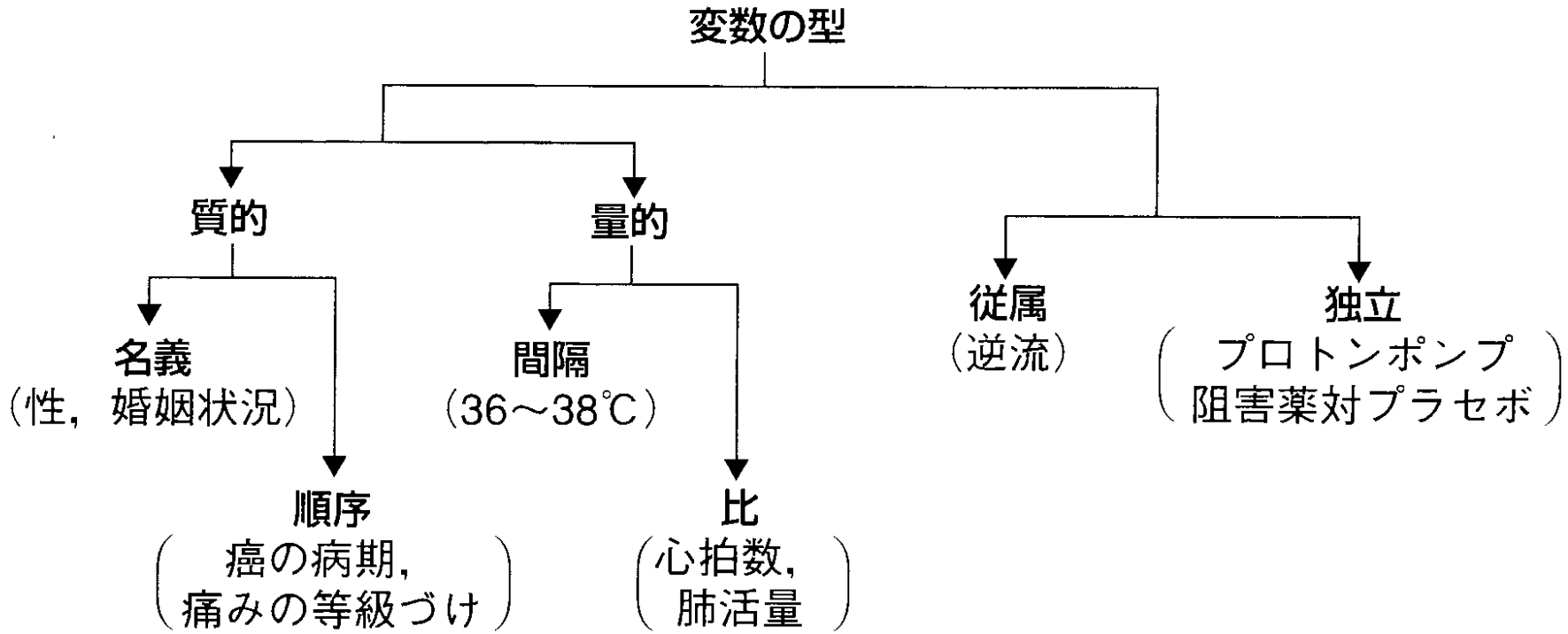



図 1-1　変数の型

# パラメトリック検定と ノンパラメトリック検定

PDQ

- 質的データ ⟶ ノンパラメトリック検定
  - $\chi^2$検定
  - $U$ 検定
  - クラスカル・ウォリス検定
  - 順位相関係数と検定
- 量的データ ⟶ パラメトリック検定
  - $t$ 検定
  - 一元配置分散分析
  - 相関係数と検定

PDQ

Part2　パラメトリック統計学
3. 統計的推測
　標本と母集団、標準誤差、
　1標本における平均値に関する推測、
　第1種の過誤と第2種の過誤、
　αレベルとβレベル
　片側検定と両側検定
　信頼区間
　パラメトリック統計とノンパラメトリック統計
　標本の大きさの計算
　統計的有意性と医学的有意性
4. 2つの標本平均の比較：t検定
5. 多群の平均間の比較：分散分析
6. 間隔変数と比例変数の間の関係
　　：回帰分析と関連手法
7. 共分散分析
8. 時系列分析

Part3　ノンパラメトリック統計学
9. ノンパラメトリック検定
　名義データの場合の検定、
　2項検定、
　フィッシャーの正確な検定、
　マクネマー検定
　順序変数の場合のノンパラメトリック検定、
　マンーホイットニーのU検定
　中央値検定
　クラスカルーウォリス検定
　コロモゴロフースミルノフ検定
　符号検定
　ウイルコクソン検定、・・・

# 度数分布(質的変数)

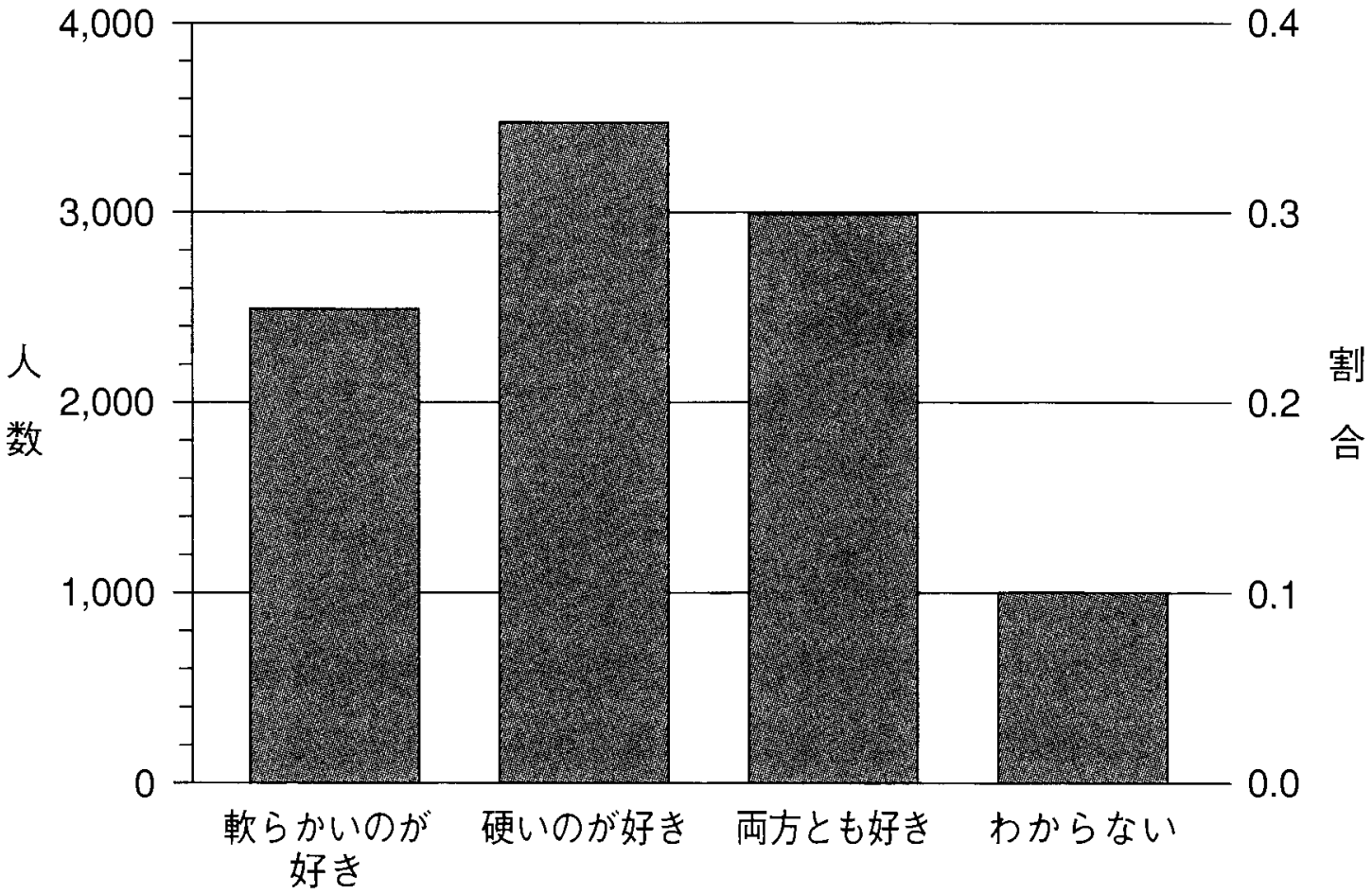



図 2-1　10,000 人の子どものアイスクリームの好みの分布

# 度数分布(量的変数)

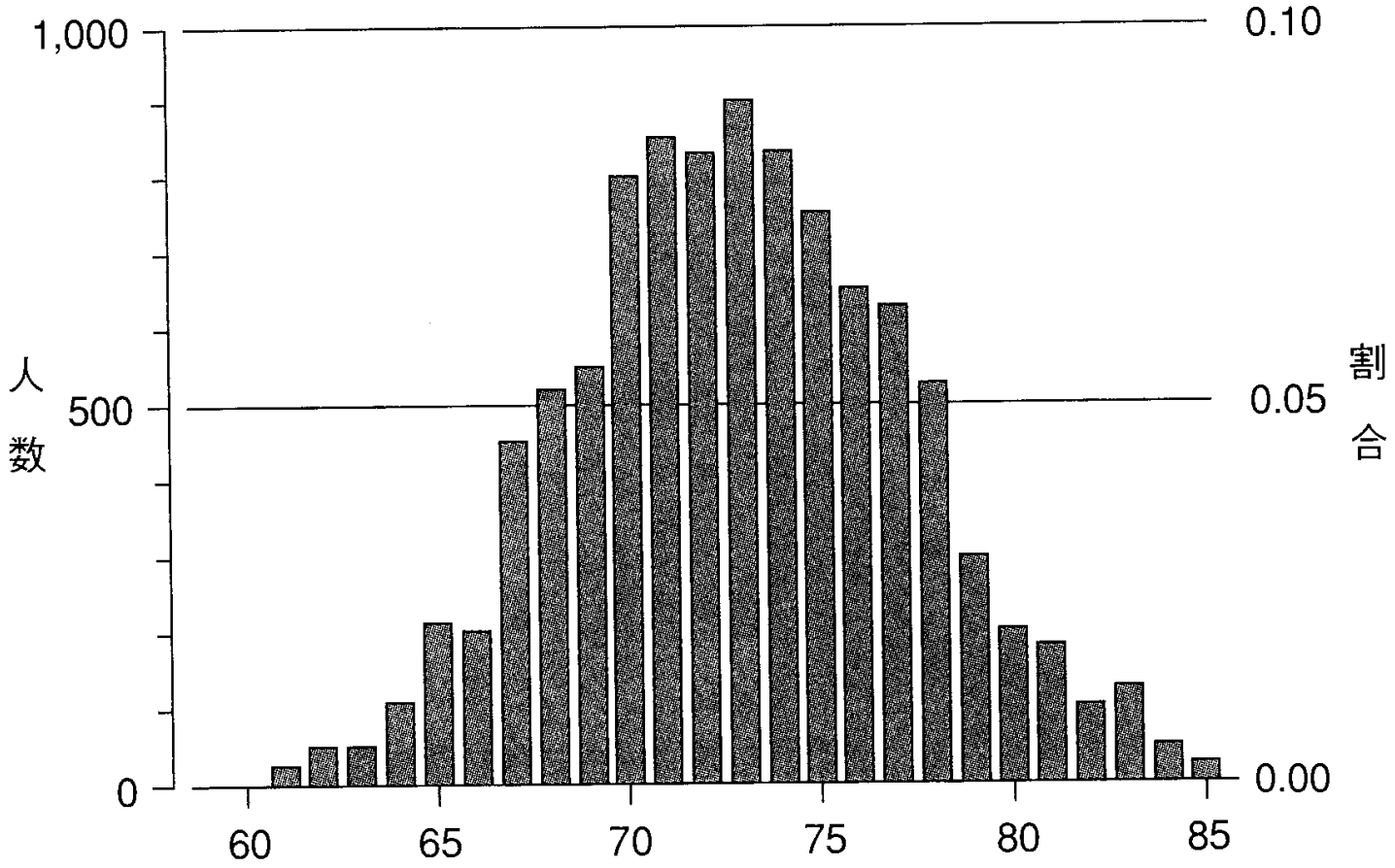



図 2-2　高齢者ローラー競技の参加者 10,000 人の年齢分布

＊訳注：本来，連続量はヒストグラムで表し，柱の間隔は開けません．

# 平均値と中央値

- 質問
  従業員1001人の会社が2社ある。A社は平均給与が50万円、B社は中央値が50万円で同額であった。あなたは、どちらの会社を希望しますか？

- サブ質問　A社の500番目の人の給与？
  　　B社の500番目の人の給与？
- 分布が左右対称の場合は
  分布が左右非対称の場合は
  ※左右の非対称性を示すものさし：歪度（skewness）

# 中央の測定
# ;平均値･中央値･最頻値

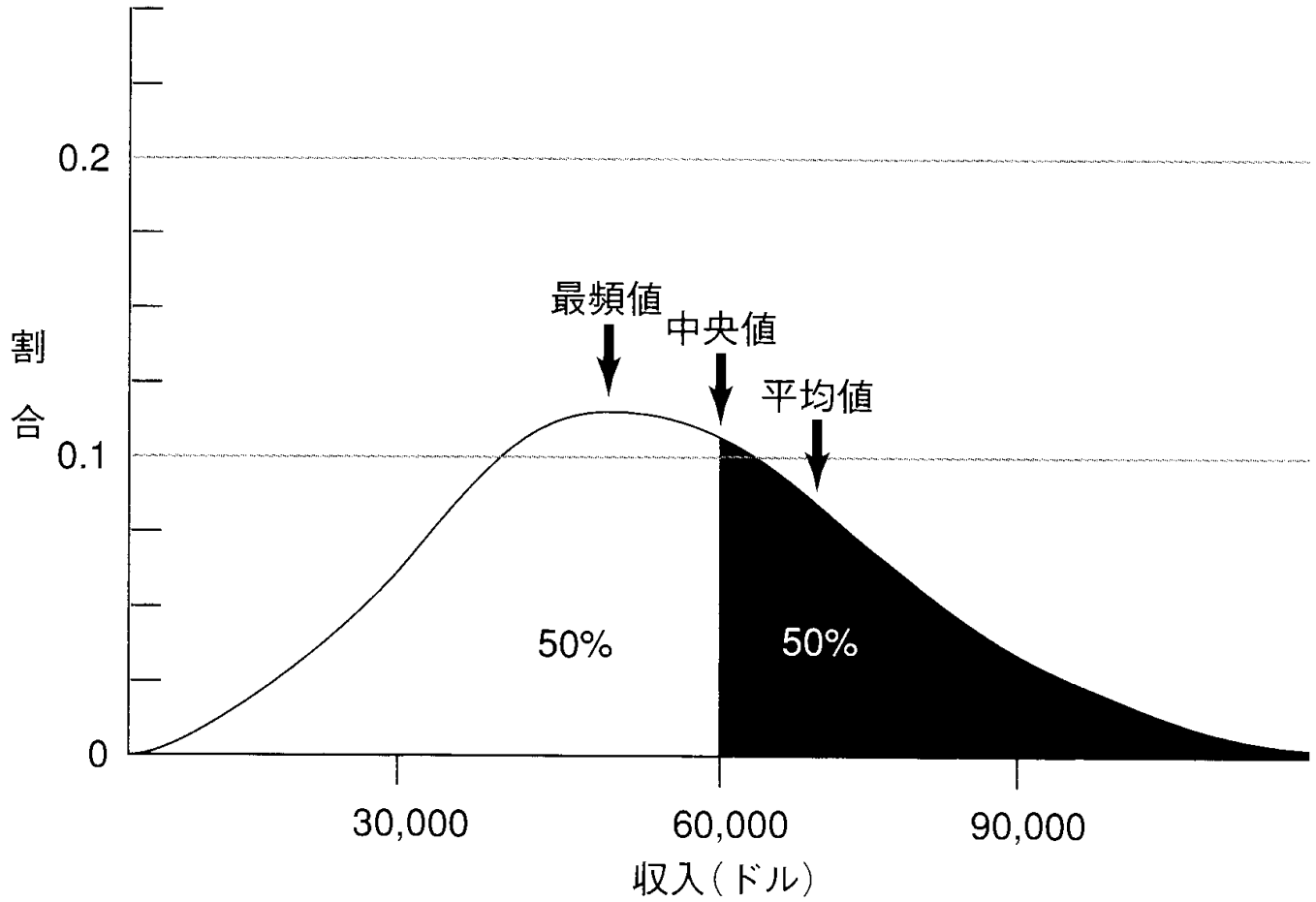



図 2-3　医師の収入の分布

# 2峰性の分布

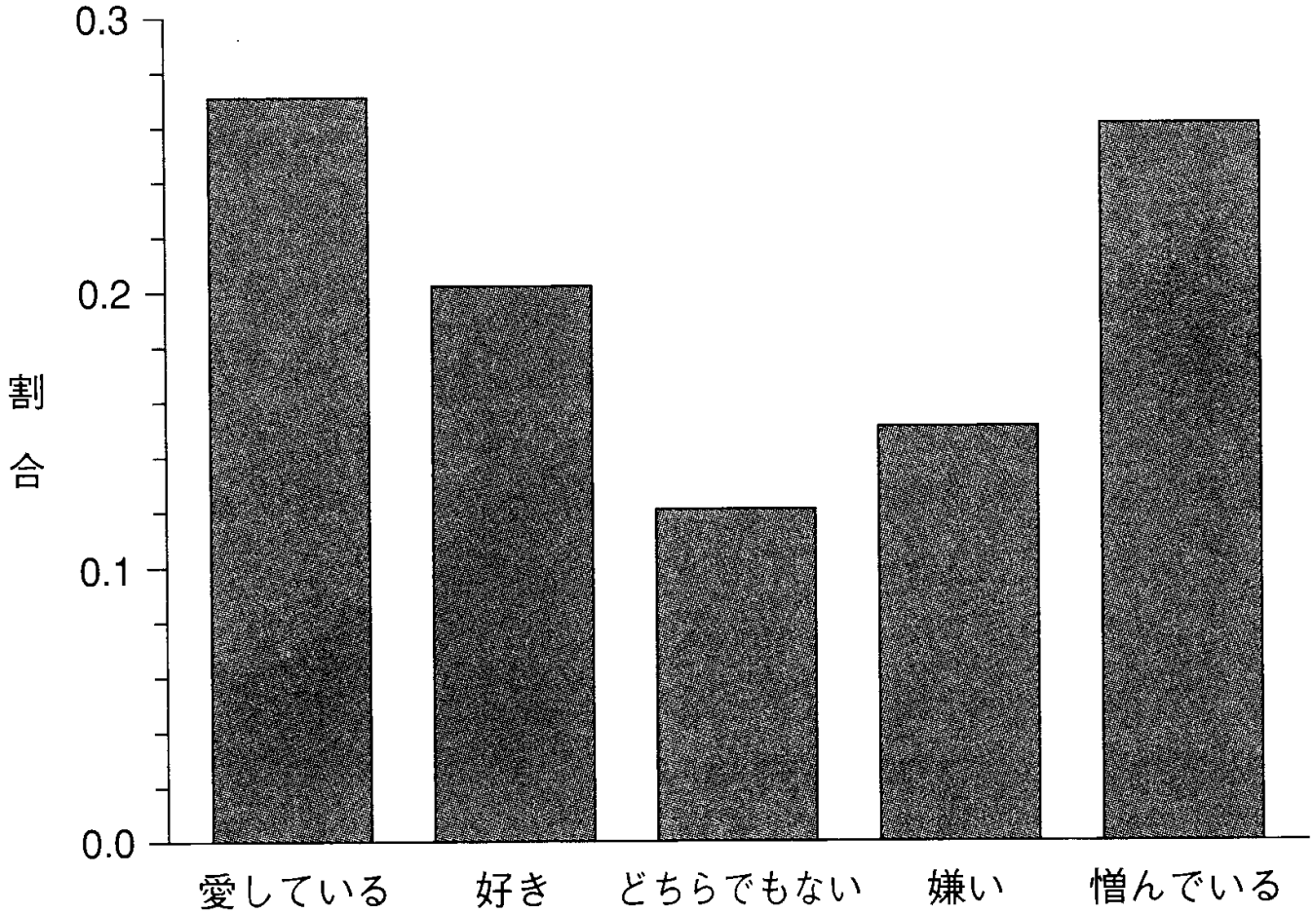
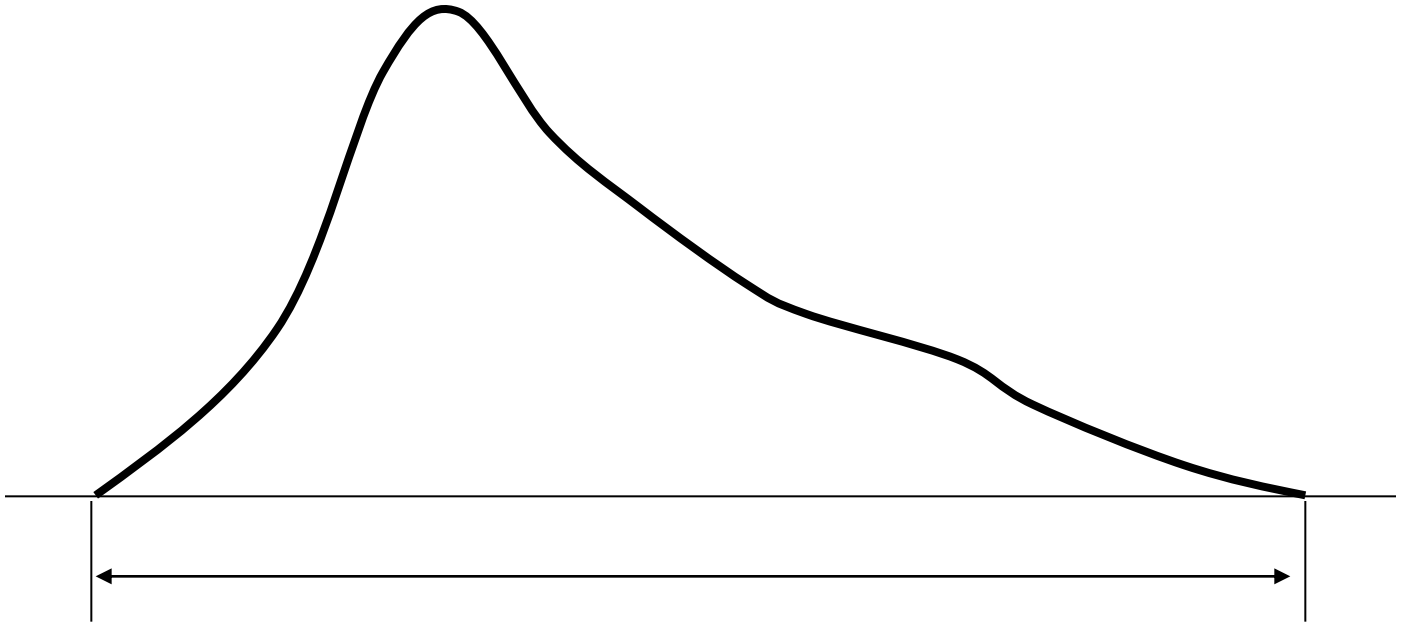



図 2-4　100 人の 10 歳代の両親に対する気持ちの分布

# 変動の測定：範囲

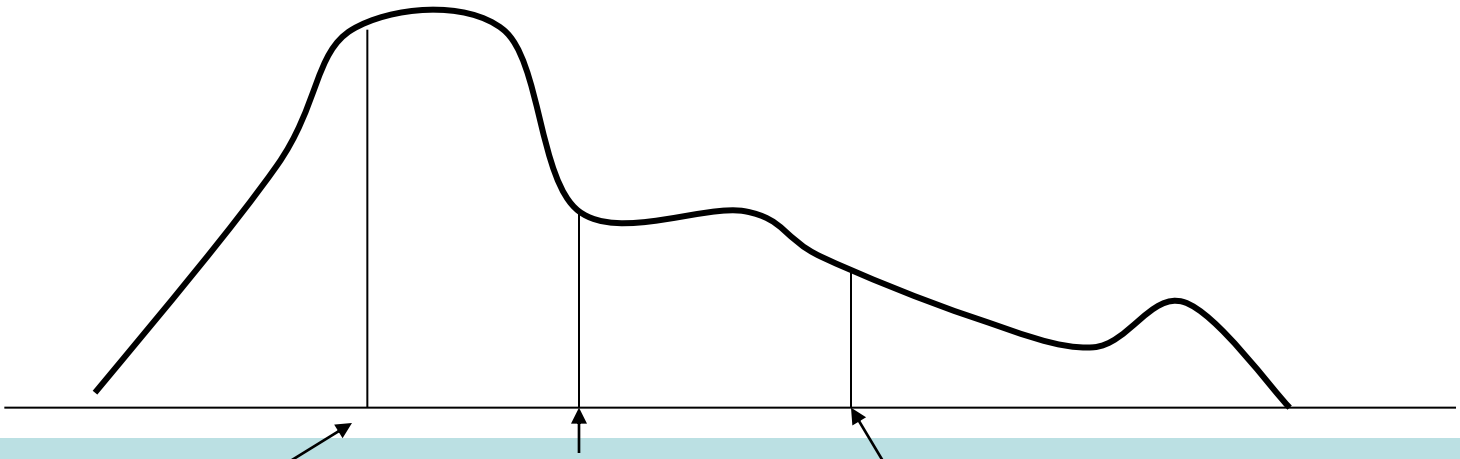
範囲＝最大値－最小値



# 変動の測定：
# 四分位数、百分位数、四分位範囲

**四分位範囲　＝　第3四分位数　－　第1四分位数**

**四分位偏差　＝　四分位範囲/2**



第1四分位数　25ﾊﾟｰｾﾝﾀｲﾙ

中央値　第2四分位数　50ﾊﾟｰｾﾝﾀｲﾙ　→面積を2等分

第3四分位数　75ﾊﾟｰｾﾝﾀｲﾙ　→面積を4等分

→面積を100等分

# 分散と標準偏差

偏差　＝　個々の値　－　平均値

↓

平均偏差　＝｜個々の値　－　平均値｜の合計/数値の数

↓

分散　　＝（個々の値　－　平均値）$^2$ の合計/数値の数

↓

$$\text{標準偏差} = \sqrt{\frac{(\text{個々の数値} - \text{平均値})^2 \text{の合計}}{\text{数値の数}}}$$

# 分散と標準偏差について

- なぜ標準偏差は平方根をとるのか？
- 標準偏差には2種類有
  母標準偏差：nで割る（Excel STDEVP)
  標本標準偏差：n-1で割る（Excel STDEV）
- どちらを用いたらよいか？
- 標準偏差には2つの意味がある。
  ＠相対的尺度として・・・変動（ばらつき）の大きさを示す
- ＠絶対的な尺度として
  ・・・区間（平均ー2×標準偏差、平均＋2×標準偏差）
  に含まれるデータの割り合い＝95%

# 正規分布

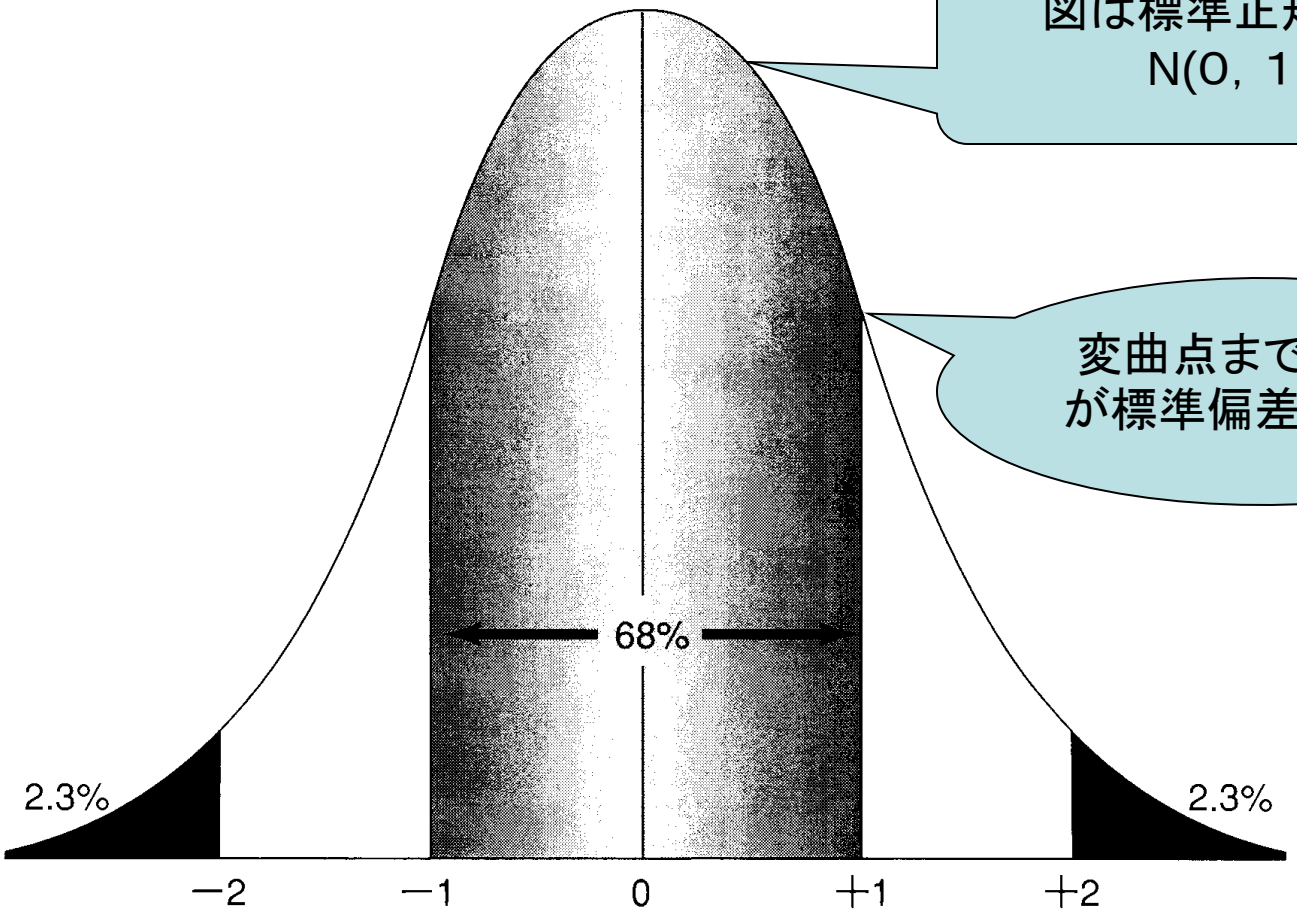



図 2-5　正規分布(釣り鐘型曲線)

# 正規分布についての注意

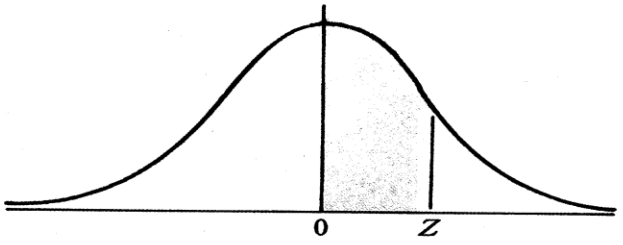
- 正規分布は平均μ、分散$\sigma^2$（標準偏差σ）に依存して決まる。$N(\mu, \sigma^2)$

  ※確率密度関数を知ると上の概念は容易。

  　積分すると1になる。
- 平均＝0、σ＝1の正規分布を標準正規分布$N(0, 1)$と呼び、教科書などにはこの表が掲載。
- 標準化$Z=(X-\mu)/\sigma$により、すべての正規分布は$N(0, 1)$に変換される。

# 付表１

## 標準正規分布表:N(0, 1)




**標準正規分布**

表中の数字は，全体の面積を1.0としたときの，Z＝0からZまでの面積を表わす．

| Z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .0000 | .0040 | .0080 | .0120 | .0160 | .0199 | .0239 | .0279 | .0319 | .0359 |
| 0.1 | .0398 | .0438 | .0478 | .0517 | .0557 | .0596 | .0636 | .0675 | .0714 | .0754 |
| 0.2 | .0793 | .0832 | .0871 | .0910 | .0948 | .0987 | .1026 | .1064 | .1103 | .1141 |
| 0.3 | .1179 | .1217 | .1255 | .1293 | .1331 | .1368 | .1406 | .1443 | .1480 | .1517 |
| 0.4 | .1554 | .1591 | .1628 | .1664 | .1700 | .1736 | .1772 | .1808 | .1844 | .1879 |
| 0.5 | .1915 | .1950 | .1985 | .2019 | .2054 | .2088 | .2123 | .2157 | .2190 | .2224 |
| 0.6 | .2258 | .2291 | .2324 | .2357 | .2389 | .2422 | .2454 | .2486 | .2518 | .2549 |
| 0.7 | .2580 | .2612 | .2642 | .2673 | .2704 | .2734 | .2764 | .2794 | .2823 | .2852 |
| 0.8 | .2881 | .2910 | .2939 | .2967 | .2996 | .3023 | .3051 | .3078 | .3106 | .3133 |
| 0.9 | .3159 | .3186 | .3212 | .3238 | .3264 | .3289 | .3315 | .3340 | .3365 | .3389 |
| 1.0 | .3413 | .3438 | .3461 | .3485 | .3508 | .3531 | .3554 | .3577 | .3599 | .3621 |
| 1.1 | .3643 | .3665 | .3686 | .3708 | .3729 | .3749 | .3770 | .3790 | .3810 | .3830 |
| 1.2 | .3849 | .3869 | .3888 | .3907 | .3925 | .3944 | .3962 | .3980 | .3997 | .4015 |
| 1.3 | .4032 | .4049 | .4066 | .4082 | .4099 | .4115 | .4131 | .4147 | .4162 | .4177 |
| 1.4 | .4192 | .4207 | .4222 | .4236 | .4251 | .4265 | .4279 | .4292 | .4306 | .4319 |
| 1.5 | .4332 | .4345 | .4357 | .4370 | .4382 | .4394 | .4406 | .4418 | .4429 | .4441 |
| 1.6 | .4452 | .4463 | .4474 | .4484 | .4495 | .4505 | .4515 | .4525 | .4535 | .4545 |
| 1.7 | .4554 | .4564 | .4573 | .4582 | .4591 | .4599 | .4608 | .4616 | .4625 | .4633 |
| 1.8 | .4641 | .4649 | .4656 | .4664 | .4671 | .4678 | .4686 | .4693 | .4699 | .4706 |
| 1.9 | .4713 | .4719 | .4726 | .4732 | .4738 | .4744 | .4750 | .4756 | .4761 | .4767 |
| 2.0 | .4772 | .4778 | .4783 | .4788 | .4793 | .4798 | .4803 | .4808 | .4812 | .4817 |
| 2.1 | .4821 | .4826 | .4830 | .4834 | .4838 | .4842 | .4846 | .4850 | .4854 | .4857 |
| 2.2 | .4861 | .4864 | .4868 | .4871 | .4875 | .4878 | .4881 | .4884 | .4887 | .4890 |
| 2.3 | .4893 | .4896 | .4898 | .4901 | .4904 | .4906 | .4909 | .4911 | .4913 | .4916 |
| 2.4 | .4918 | .4920 | .4922 | .4925 | .4927 | .4929 | .4931 | .4932 | .4934 | .4936 |
| 2.5 | .4938 | .4940 | .4941 | .4943 | .4945 | .4946 | .4948 | .4949 | .4951 | .4952 |
| 2.6 | .4953 | .4955 | .4956 | .4957 | .4959 | .4960 | .4961 | .4962 | .4963 | .4964 |
| 2.7 | .4965 | .4966 | .4967 | .4968 | .4969 | .4970 | .4971 | .4972 | .4973 | .4974 |
| 2.8 | .4974 | .4975 | .4976 | .4977 | .4977 | .4978 | .4979 | .4979 | .4980 | .4981 |
| 2.9 | .4981 | .4982 | .4982 | .4983 | .4984 | .4984 | .4985 | .4985 | .4986 | .4986 |
| 3.0 | .4987 | .4987 | .4987 | .4988 | .4988 | .4989 | .4989 | .4989 | .4990 | .4990 |
| 3.1 | .4990 | .4991 | .4991 | .4991 | .4992 | .4992 | .4992 | .4992 | .4993 | .4993 |
| 3.2 | .4993 | .4993 | .4994 | .4994 | .4994 | .4994 | .4994 | .4995 | .4995 | .4995 |
| 3.3 | .4995 | .4995 | .4995 | .4996 | .4996 | .4996 | .4996 | .4996 | .4996 | .4997 |
| 3.4 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4998 |

# 標準化と標準得点

PDQ

$$\text{標準得点}(z) = \frac{(\text{素得点} - \text{平均値})}{\text{標準偏差}}$$

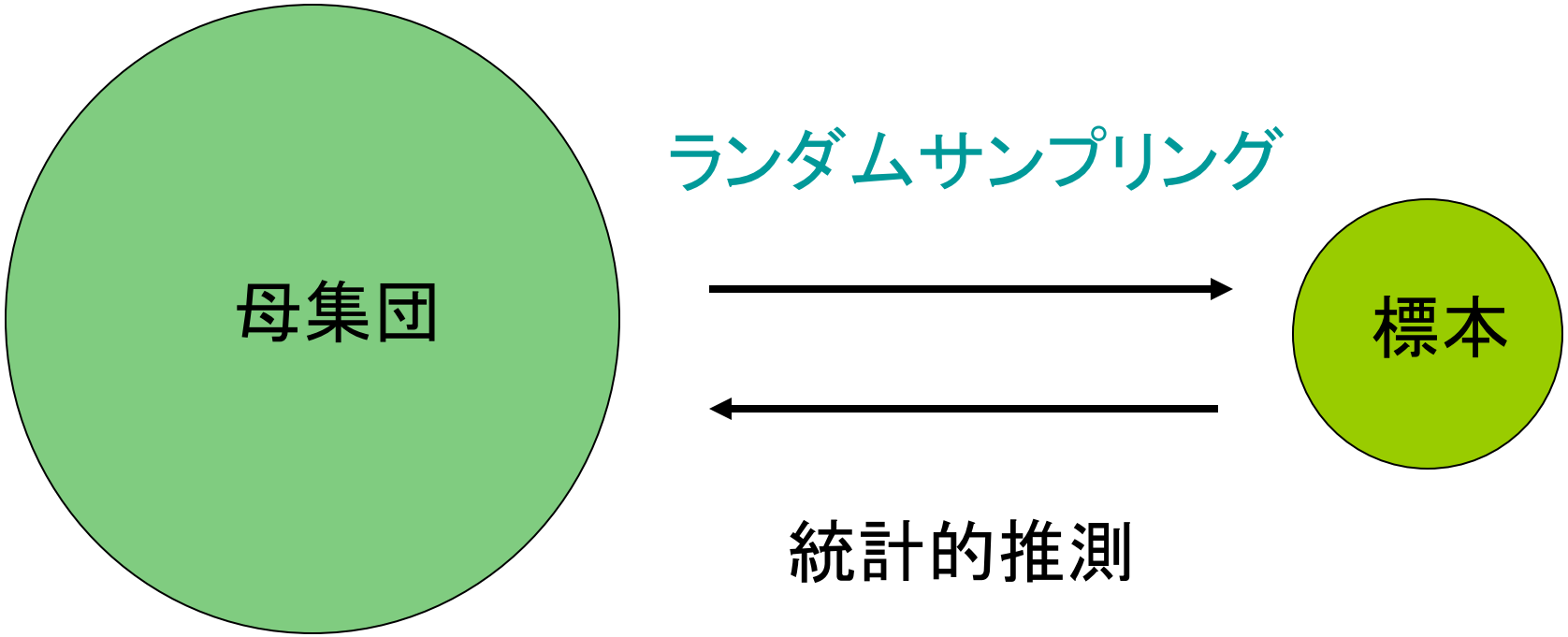
# 偏差値について

- 問題　大学入試センター試験で偏差値が80点の人は上位何番目か？受験生を10万人とする。
- 偏差値とは、素点を標準得点に変換し、それを10倍し50を加えたものさし。

  偏差値＝[(X－μ)／σ]＊10＋50
- もし、素点分布が$N(\mu, \sigma^2)$ であれば、偏差値の分布は$N(50,10^2)$となる。

# 統計的推測

PDQ



# 標準偏差SDと標準誤差SE

PDQ

$$SE = \frac{SD}{\sqrt{標本の大きさ}}$$

**※標準誤差は、一般的には統計量の標準偏差を指す用語で、上の式は平均の標準偏差を示している。英語では、Standard Error of Mean である。**

# 標本の大きさと標本平均の分布

PDQ

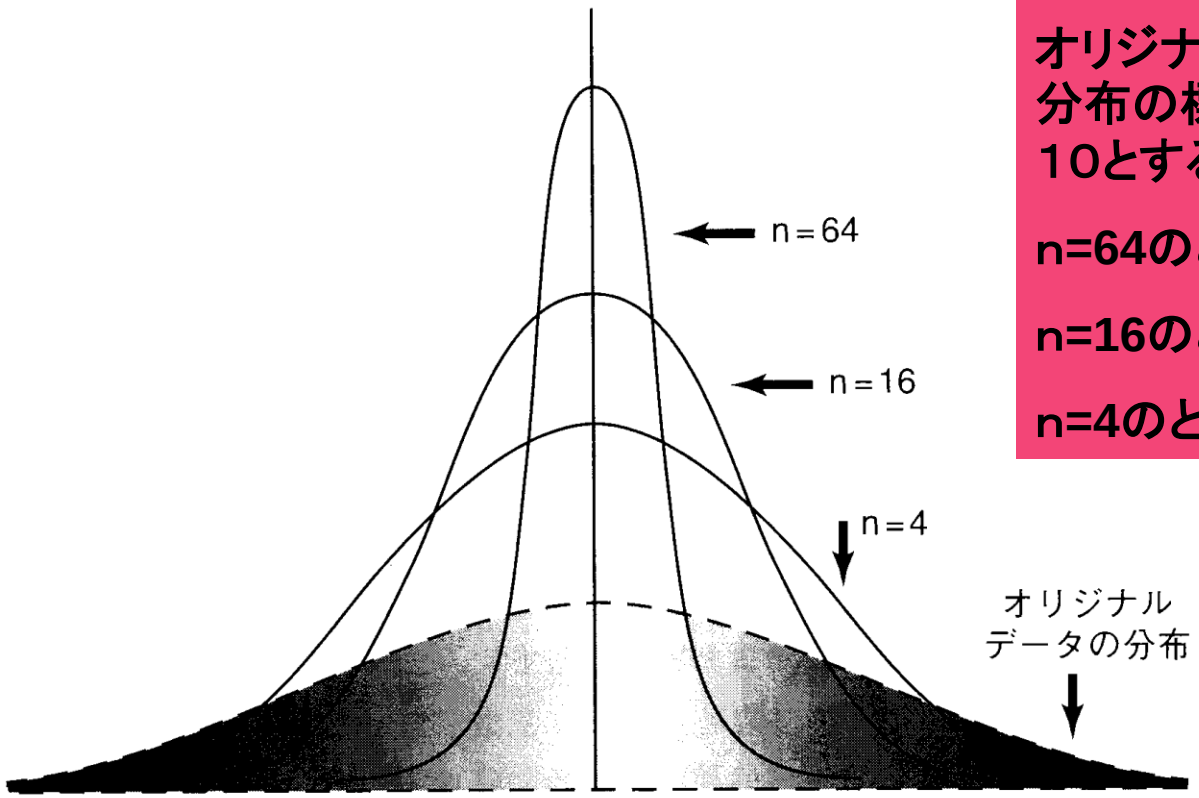



図 3-1　オリジナルデータの分布と標本の大きさの変化に伴う標本平均の分布

**質問1：**

**オリジナルデータの分布の標準偏差を10とすると**

**n=64のときのSE？**

**n=16のときのSE？**

**n=4のときのSE？**

**質問2：**

**正規分布$N(\mu,\sigma^2)$とすると**

**n=64のときN(μ,？）**

**n=16のときN(μ,？）**

**n=4のとき N(μ,？）**

# 母集団の分布

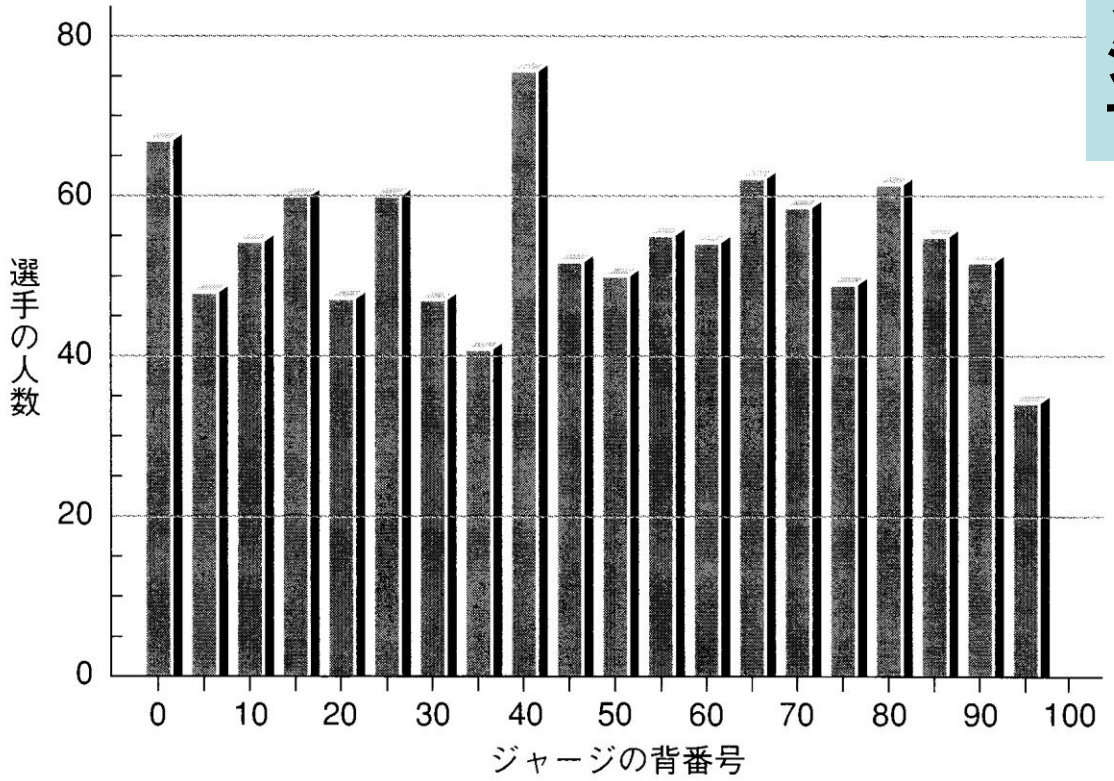
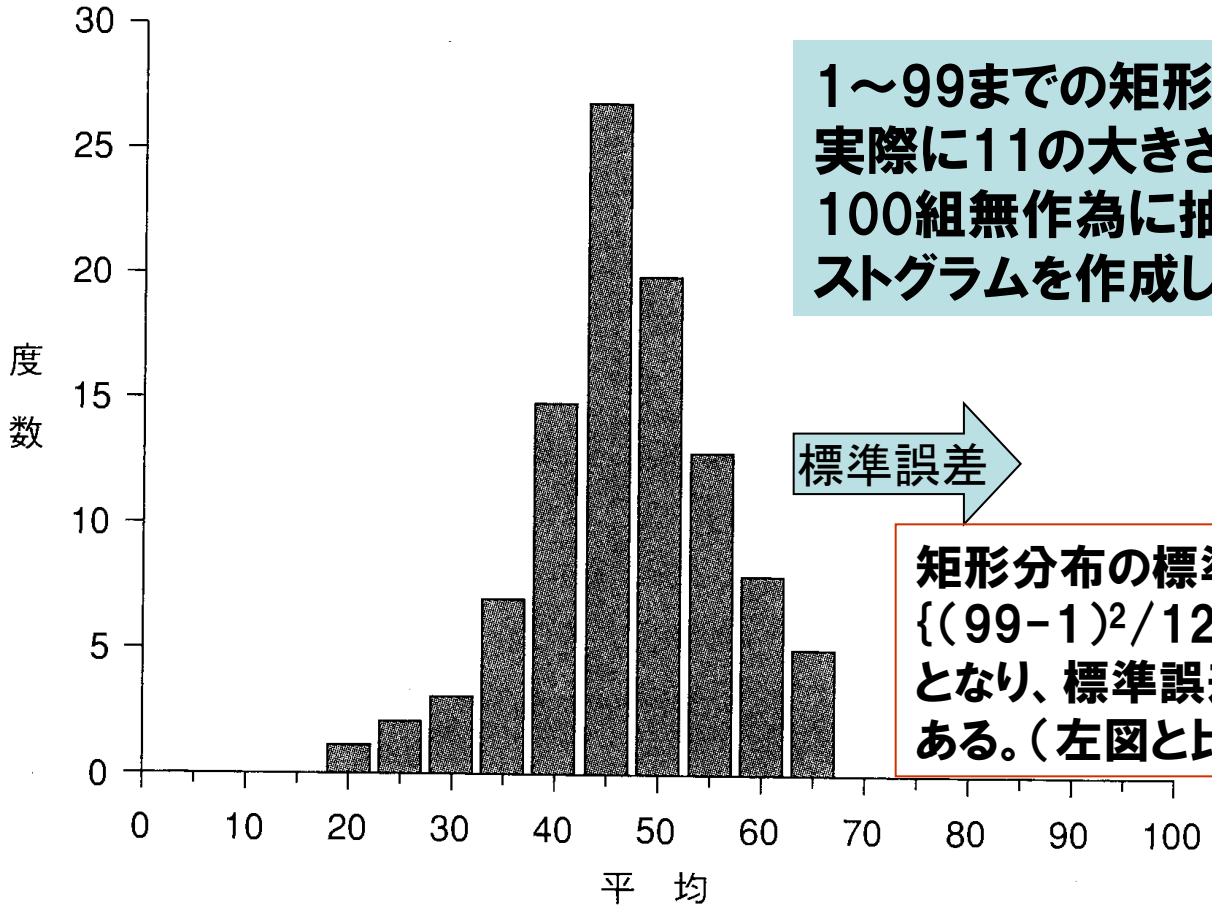



図 3-2　ジャージの背番号の分布

フットボールのゲームで、ジャージの背番号に関する分布の例

オリジナルの背番号は基本的には矩形分布（1～99の任意の数字がほかのどの数字とも同様に確からしいという性質をもつ分布）と考える

# 標本平均の分布




1～99までの矩形分布から、実際に11の大きさの標本を100組無作為に抽出してヒストグラムを作成したもの

標準誤差

矩形分布の標準偏差は、SD=√{(99−1)²/12}=49/√3=28となり、標準誤差は28/√11=8.4である。（左図と比べてください）

図 3-3　標本の大きさが 11 の場合の標本平均の分布

PDQ

# 中心極限定理

そこには，統計学において最も忘れられている，1つの魔法のような定理が存在します．――それは，**中心極限定理** Central Limit Theorem です．標本の大きさが十分大きいとき(大きいとは，5または10以上であることをいいます)，平均値は**オリジナルのデータの分布形にかかわらず**正規分布に従うということを述べた定理です．つまり，平均値に関して推論を行う場合，オリジナルのデータが正規分布をしていようがいまいが，パラメトリックな統計を用いて計算することができます．

母集団分布の推測

標本平均の平均　→　母集団の平均

標本平均の標準偏差(標準誤差)×√標本の大きさ　→　母集団の標準偏差

# 分布の推定：平均値と標準偏差

平均値と標準偏差は、

（1）相対度数分布の位置と広がりを表す

（2）分布がひとつ山で左右対称のときに使われることが多い

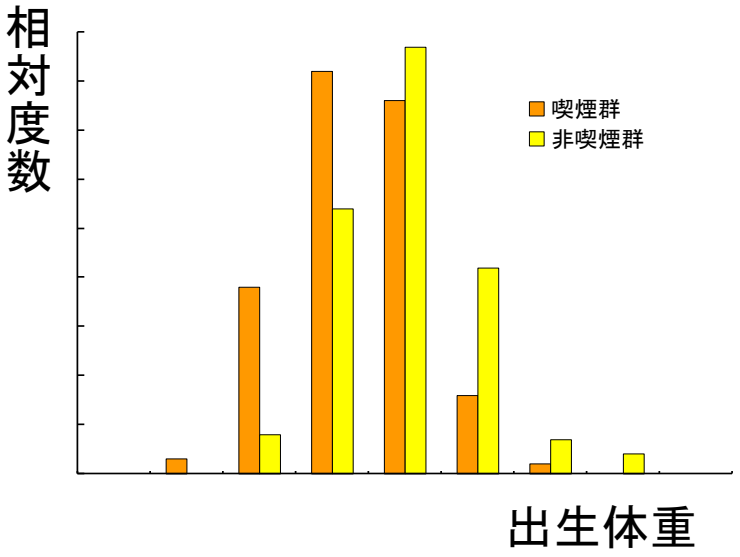
**平均値　：　分布の中心位置**

**標準偏差：　分布の広がり**

(例)

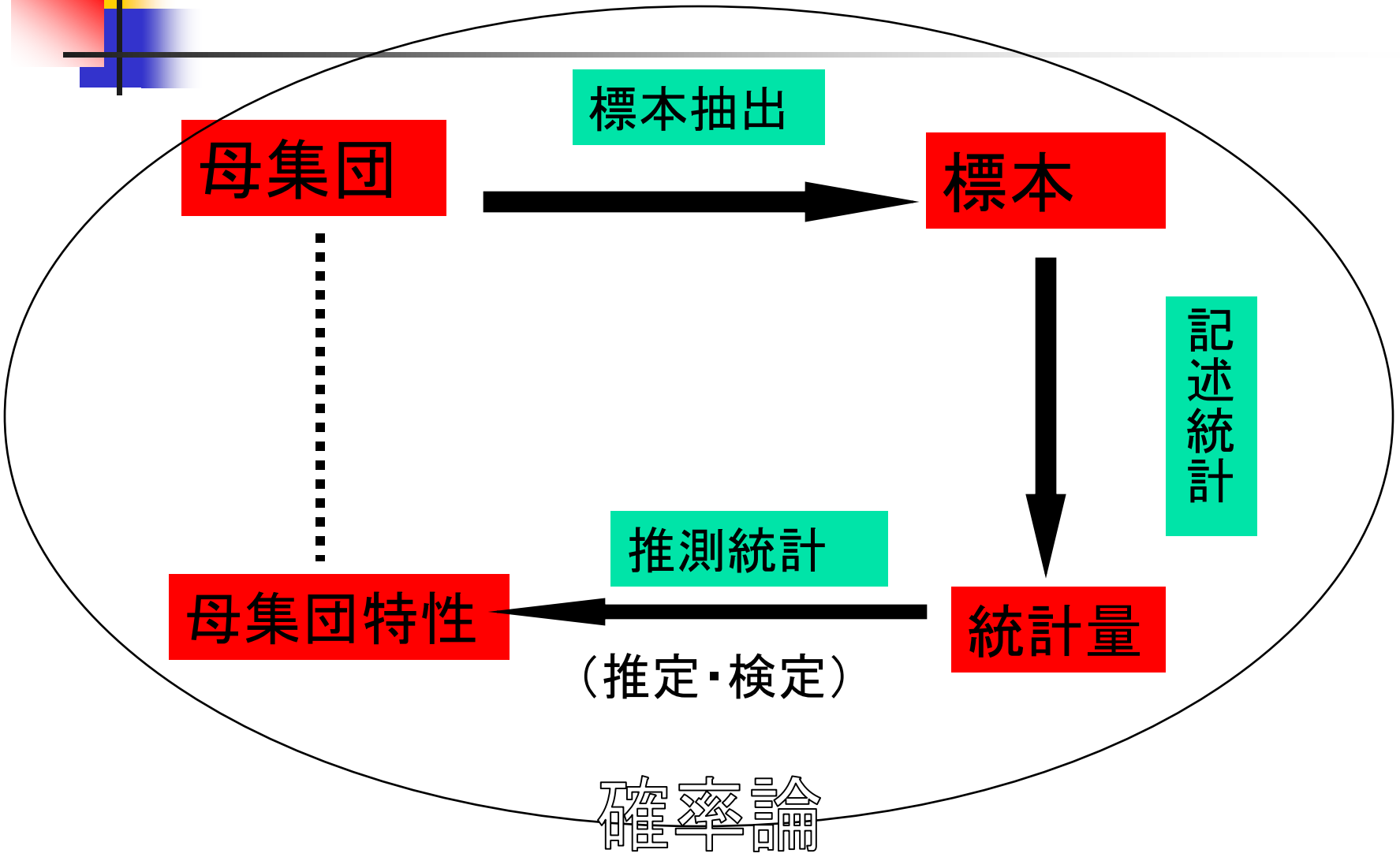
妊娠中も喫煙した母親から生まれた児の出生体重　　**2,960±380(g)**

妊娠前から非喫煙の母親から生まれた児の出生体重　　**3,067±370(g)**

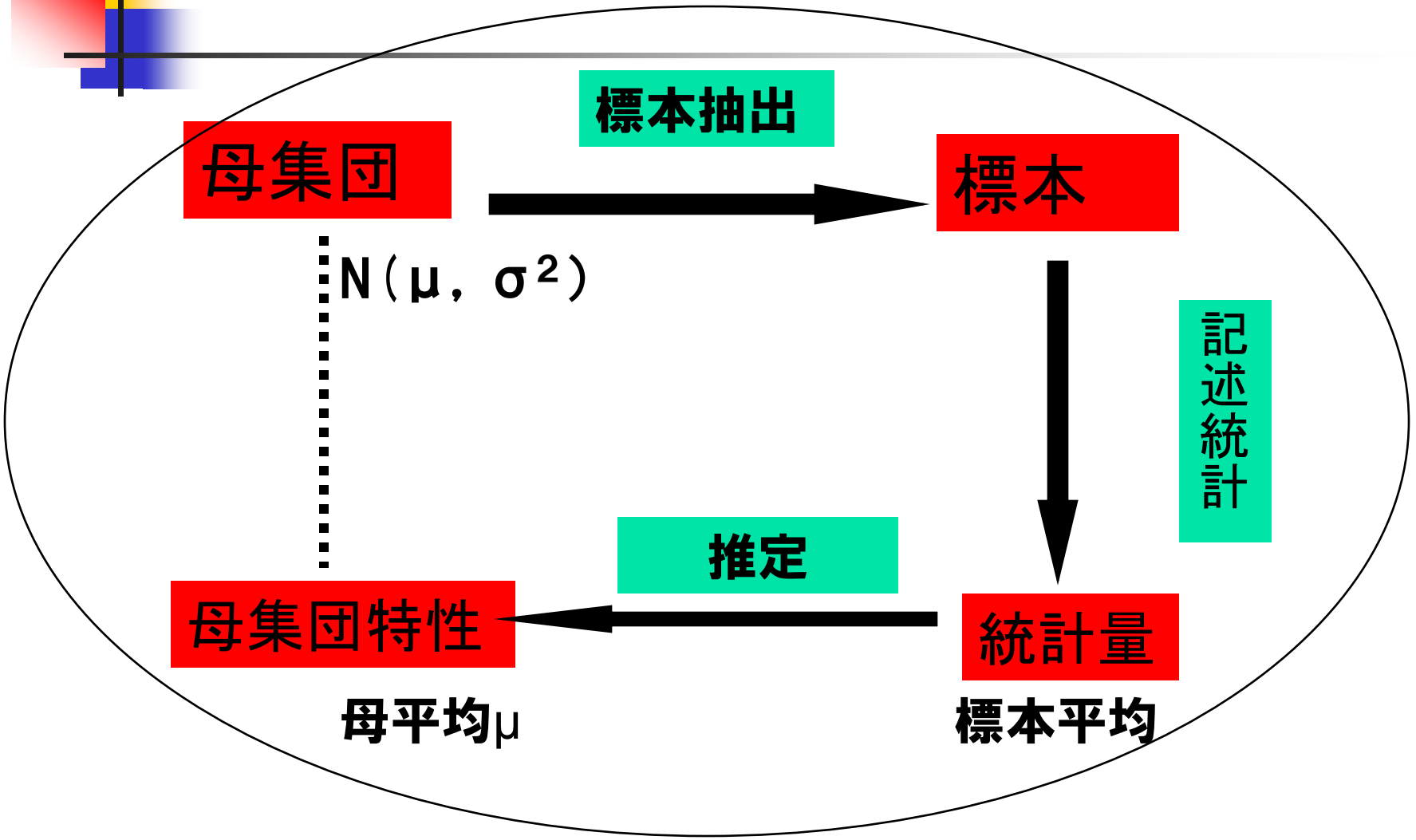



**喫煙群は非喫煙群に比べて平均出生体重が100gも少ない。**

# 統計的推測のモデル



# 区間推定－信頼区間の構築(1)



# 区間推定－信頼区間の構築(2)

**目的：未知の母平均μがある区間に含まれるという形式で推定を行う。**

**区間の形式：** $(\bar{x}-\delta, \bar{x}+\delta), \bar{x}-\delta < \mu < \bar{x}+\delta$

**区間の特徴：区間の長さが長くなると信頼度（区間の中にμが含まれる確率）は増すが、推定の精度（情報の利用価値）は小さくなる。**
**また逆に区間の長さが短くなると信頼度は減少するが推定の精度は良くなる。**

| | |
|---|---|
| 区間の長さ | 長 <――> 短 |
| 信頼度 | 高 <――> 低 |
| 精度（利用価値） | 悪 <――> 良 |

# 区間推定(3)

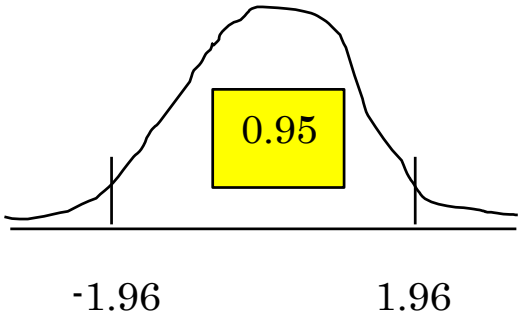
## 母平均の推定（1 標本問題）：信頼区間

標本平均の分布　$\bar{x} \sim N(\mu, \frac{\sigma^2}{\sqrt{n}})$

標準化　$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

信頼度 95%　$\Pr\{-1.96 < z < 1.96\} = 0.95$

$$-1.96 < \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < 1.96$$

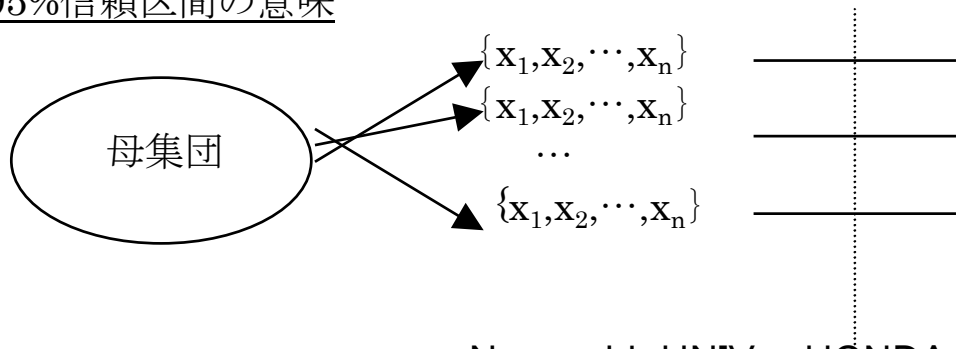



95%信頼区間：

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

（注：母標準偏差 σ は現実的には未知である）

95%信頼区間の意味




100 回中、95 回の割合で $\mu$ が含まれる

# 区間推定(4)

母標準偏差σ：未知の場合（現実的状況）

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1) \rightarrow\rightarrow\rightarrow\rightarrow \qquad T = \frac{\bar{x} - \mu}{s / \sqrt{n}} \sim t_{n-1}$$

σをsで置換

$$s = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}$$ ：標本標準偏差

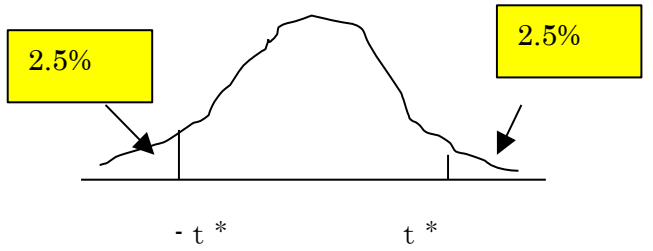
$t_{n-1}$：自由度n－１のｔ分布（スチューデントのｔ分布）

例：身長のデータが正規分布N($\mu$, $\sigma^2$)に従うと仮定する。今、10人を無作為に抽出し、10人の平均身長ｘ＝１５７㎝、標準偏差ｓ＝５cmであった。母平均$\mu$の９５％信頼区間を求める。

(i) σ：既知（σ＝５）のとき

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

$$157 - 1.96\frac{5}{\sqrt{10}} < \mu < 157 + 1.96\frac{5}{\sqrt{10}}$$




(ii) σ：未知のとき

$$\bar{x} - t^*\frac{s}{\sqrt{n}} < \mu < \bar{x} + t^*\frac{s}{\sqrt{n}}$$

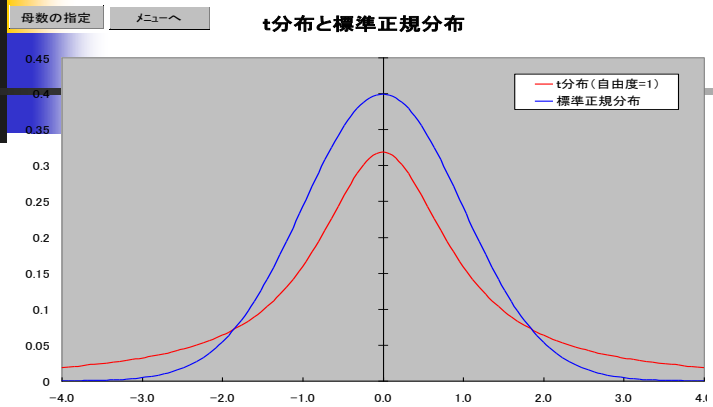
$$157 - 2.262\frac{5}{\sqrt{10}} < \mu < 157 + 2.262\frac{5}{\sqrt{10}}$$

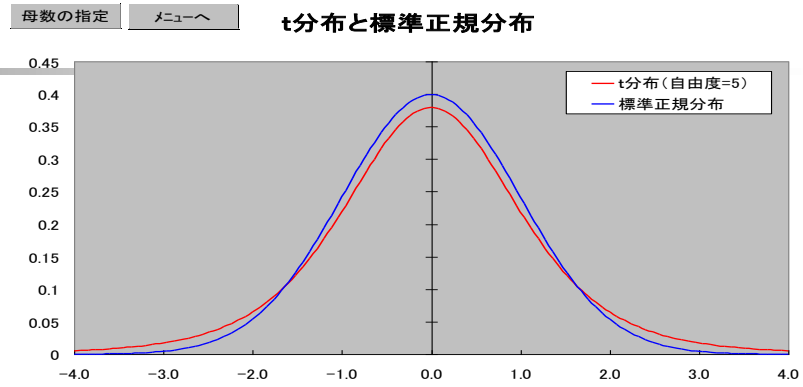
ｔ分布表において、
自由度＝９、α＝0.05　より
ｔ*＝2.262

注）・ｔ分布は自由度を大きく（つまり、標本の大きさｎを大きく）すると、
標準正規分布 N(0,1)に近づく。
・標本の大きさｎがある程度大きい場合は、σ：未知のときであってもｔ*≒1.96
と考えて、次のようでもよい。 $\bar{x} - 1.96\frac{s}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{s}{\sqrt{n}}$

# 区間推定(5)

標準正規分布N(0,1)とｔ分布について
：自由度が大きくなるに連れてｔ分布は、N(0,1)に近づく。




問　生後４日目の新生児１０人の血色素量（ヘモグロビン）のデータである。
データ：{25.0, 22.0, 28.4, 23.2, 20.8, 20.0, 16.4, 17.2, 22.4, 18.4}
新生児の血色素量の平均$\mu$の95%信頼区間、99%信頼区間を求めよ。

まとめ：信頼度１－$\alpha$の信頼区間の求め方
データ：$\{x_1, x_2, \cdots, x_n\}$
（１）　標本平均、標本標準偏差を電卓等を用いて計算

$$\overline{x}, s = \sqrt{\frac{1}{n-1}\sum (x_i - \overline{x})^2}$$

（２）　ｔ分布表よりｔ*を求める。
自由度ｎ－１，信頼度１－$\alpha$
（３）　信頼区間の計算

$$\overline{x} - t^* \frac{s}{\sqrt{n}} < \mu < \overline{x} + t^* \frac{s}{\sqrt{n}}$$

# 統計的仮説検定の例（1）

例：入浴後と安静時の最高血圧に差があるか？
標本 {20,4,10,2,10,−10,4,24,10,−6,14,10,16}　(n=13)
　　:（入浴後の最高血圧）ー（安静時の最高血圧）
　　より、統計的に検証

［仮説検定のステップ］
①「最高血圧には差がない」…仮説（帰無仮説）をたてる
②標本抽出（無作為標本抽出）
③帰無仮説の下で、標本の出現確率（出現頻度）を調べる
④出現確率：大　→　そのまま保留（結論を下さない）
　　：小　→　・仮説が正しくて、希にしか起きない
　　ことが偶然起きた
　　・**仮説が正しくなかった**

仮説検定ではこの
結論を採用

# 統計的仮説検定の例(2)

［仮説検定における2種類の過誤］

| 判定<br>母集団の状態 | H0:棄却しない | H0:棄却する |
|---|---|---|
| H0:真 | 正 | 誤(第1種の過誤) |
| H1:真 | 誤(第2種の過誤) | 正 |

※第1種の過誤の確率＝有意水準(α)
※β:第2種の過誤の確率
※検出力(power)＝1－β

検出力最大の検定が望ましい

# 統計的仮説検定の例(3)

**例題の解法**

帰無仮説　$H_0$：μ=0(「差はない」)　　**両側検定**
対立仮説　$H_1$：μ≠0(「差はある」)

有意水準1％　　$H_0$の下で $T = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{\bar{x} - 0}{s/\sqrt{n}} \sim t_{n-1}$

n=13より、自由度n−1＝12のt分布表から
棄却域：{T＞3.055, T＜−3.055}
データより、x＝8.31、s＝9.59
∴t＝8.31／(9.59／√13)=3.12　　は棄却域に含まれる
よって、帰無仮説$H_0$を棄却する。
つまり血圧値には差があると言える。

# 統計的仮説検定の例(4)

**例題の解法**

**片側検定**

帰無仮説　$H_0$：μ=0(「差はない」)
対立仮説　$H_1$：μ＞0(「差はある」)

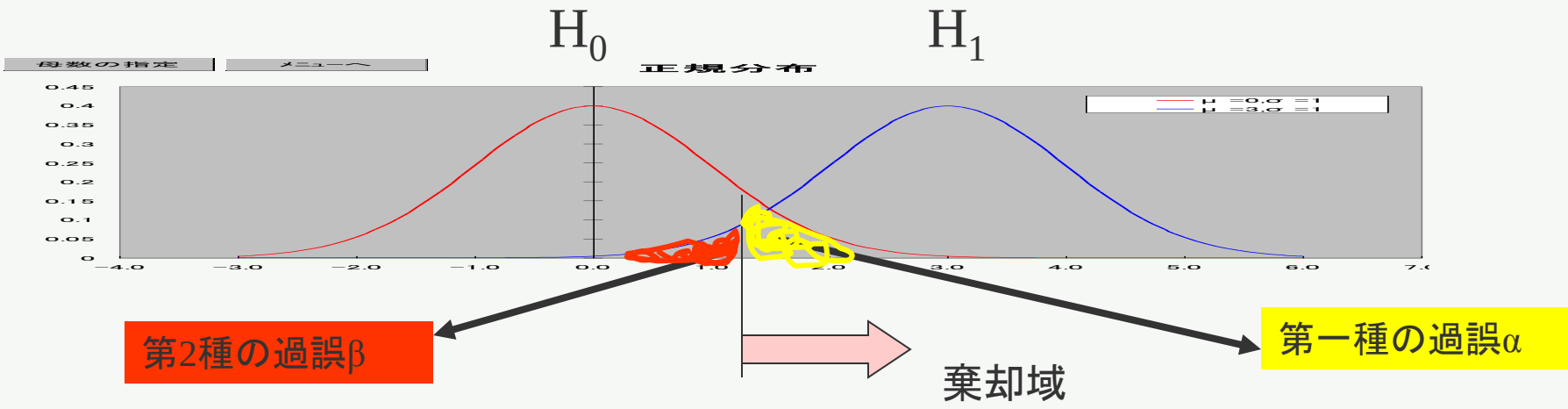
事前情報(常識)より入浴後の方が安静時より最高血圧は高い

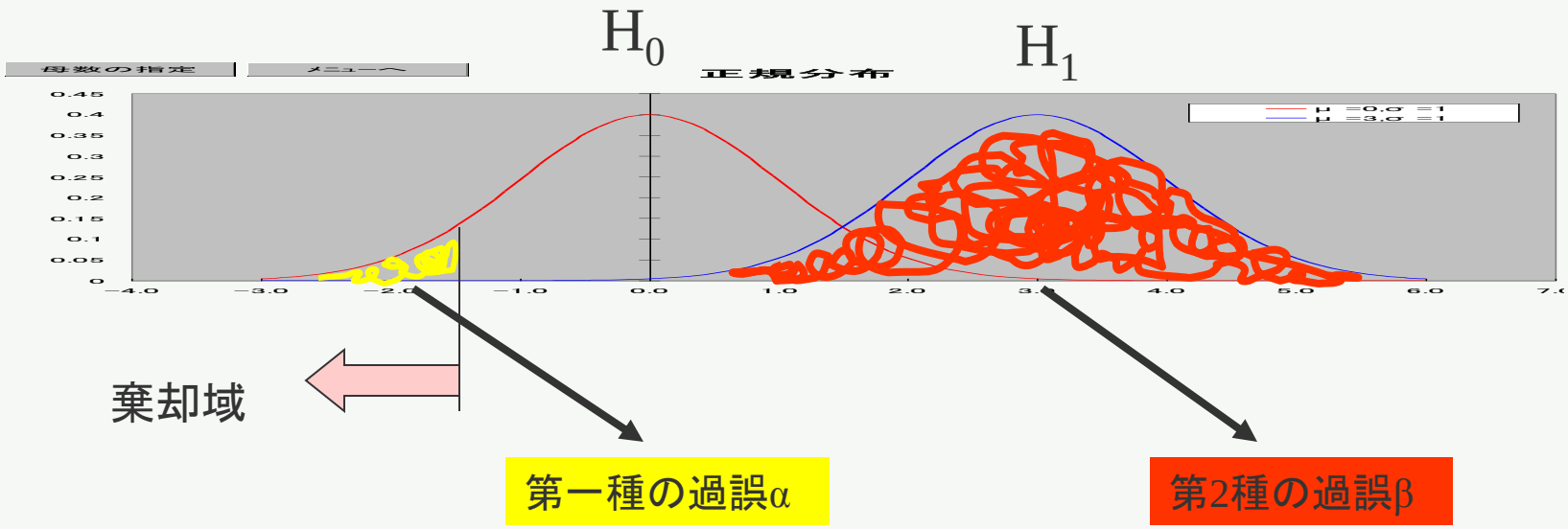
有意水準1％　　$H_0$の下で $T = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{\bar{x} - 0}{s/\sqrt{n}} \sim t_{n-1}$

n=13より、自由度n−1＝12のt分布表から
棄却域：{T＞2.681}
データより、t＝3.12　　より棄却域に含まれる
よって、帰無仮説$H_0$を棄却する。
つまり血圧値には差があると言える。

**問：片側検定において、本例で何故右側に棄却域を取るのか？**

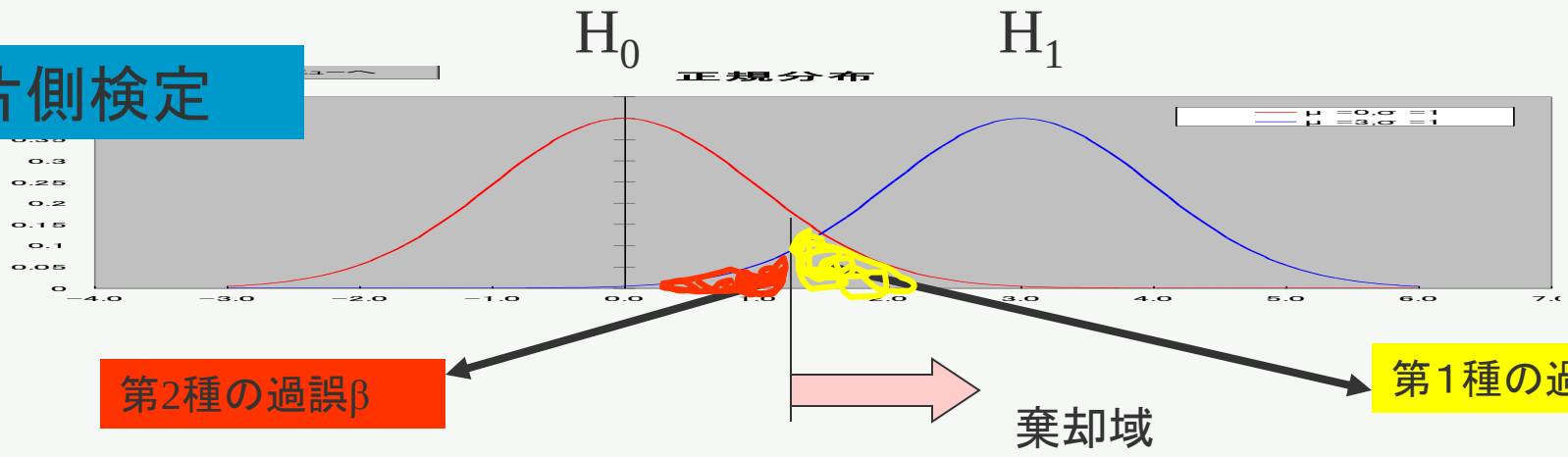
# 棄却域と第2種の過誤の確率



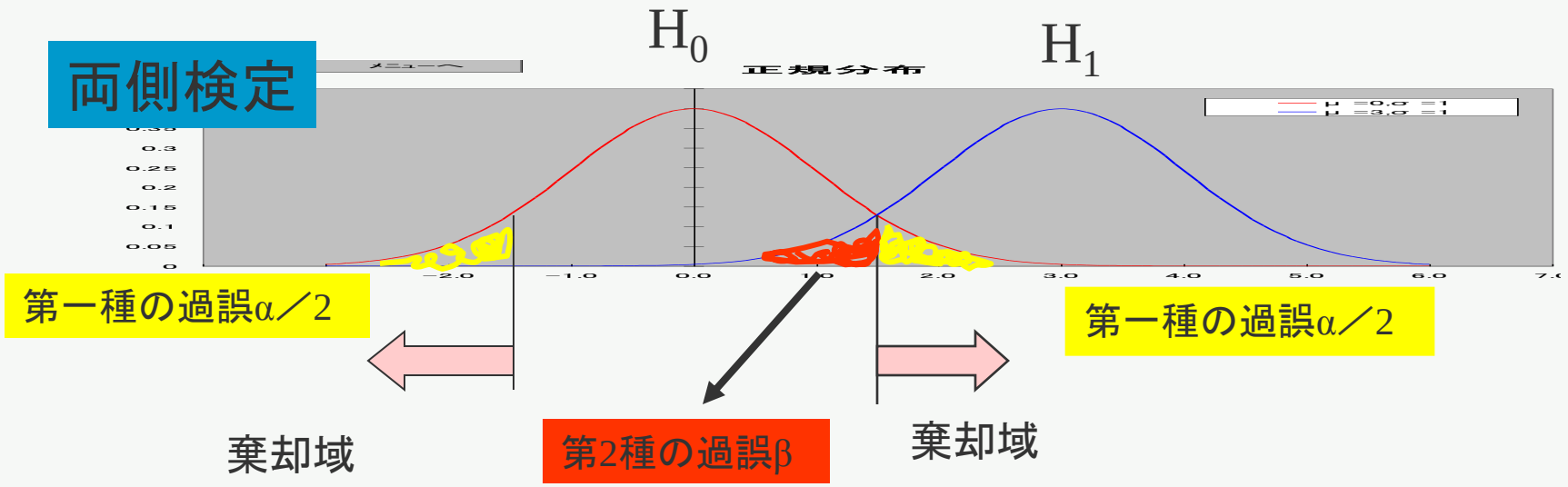



# 片側検定と両側検定における第2種の過誤の確率β

## 片側検定




## 両側検定



# 統計的分析：仮説検定（概説-1）

- 有意性検定とは、差があることをデータを用いて統計的アプローチにより検証する方法である。
- 差があることを主張したいときに、差がないと仮定（**帰無仮設**）し、その仮定の下でデータの出現頻度がある基準（**有意水準**）より小さいときに、その仮定が間違っているとして否定（**棄却**）することにより、差があることを出張（**対立仮設**を採択）する方法である。
- 仮説検定は、2重否定の論理である。
- **第1種の過誤**（差がないときに、差があると結論する誤り）の確率を有意水準と呼び、通常5%ないし1%として設定する。
- ※赤字は統計的仮説検定で用いられる用語

# 統計的分析：仮説検定（概説-2）

- **第2種の過誤**（差があるときに、差がないと結論する誤り）の確率（β ）を、第１種の過誤の確率（α ）を固定した上（例えば有意水準を5%とした条件下）で、最小にする検定が望ましい。**検出力**（Power)とは、１－β のことで、差があるときに差がある事を正しく結論する確率からである。つまり、検出力最大となる仮設検定が良い検定となる。
- AとBに差があることを検出する場合に、対立仮説としてA≠B（A>BかA<Bかはわからないがとにかく違っているということを主張したい）とした場合、帰無仮説の棄却域を確率分布に両側に設定するので、**両側検定**と呼ぶ。また、対立仮説としてA＜B（事前情報からA>Bであることは想定外であるか、」積極的にA<Bを出張したい）とした場合、帰無仮説の棄却域を確率分布に片側に設定するので、**片側検定**と呼ぶ。

※赤字は統計的仮説検定で用いられる用語

# 統計的分析：仮説検定（例説）

（例）**医学医療の問題**

妊娠中も喫煙した母親から生まれる児の出生体重は、非喫煙の母親の児の出生体重に比べて少ないか？

↓ おきかえる

**統計学での仮説**

（帰無仮説）喫煙群の出生体重の平均値＝非喫煙群の出生体重の平均値

（対立仮説）喫煙群の出生体重の平均値＜非喫煙群の出生体重の平均値

注意１：非喫煙の母親の児の出生体重に比べて少ないことを出張したい。
あえて、等しいと仮定して、その仮定の下でデータの出現頻度を調べる。
注意２：有意水準は５％か１％かを決定する。

注意３：対立仮説が不等号なので、片側検定となる。

# 統計的分析：仮説検定（例説）ー続き
# 正しい検定手法の選定とその実行

帰無仮説：喫煙群の出生体重の平均値＝非喫煙群の出生体重の平均値
対立仮説：喫煙群の出生体重の平均値＜非喫煙群の出生体重の平均値

↓

喫煙者、非喫煙者100例ずつデータ抽出

↓

分布が左右対称、S.D.も等しいのでStudentのt-検定
統計ソフトウエアSAS,SPSS等で検定を実行する

**解析結果　検定統計量　t=-2.06,　P=0.041 (<0.05)**
**有意水準5%を棄却域とする検定ではP<0.05なので**
**対立仮説を採択する。**

**結論**　喫煙群の出生体重は非喫煙群のそれに比べて有意に少ない(P<0.05)。

# 1標本における平均値に関する推測(１)

PDQ

**帰無仮説：** $H_0$：読者の平均IQ ＝ 一般の母集団の平均IQ

IQは平均値100、標準偏差15の正規分布する

↓

本書の読者もこの母集団からの無作為標本であるとする

**対立仮説：** $H_1$：読者の平均IQ ＞ 一般の母集団の平均IQ

：読者の平均IQ ＝ 一般の母集団の平均IQ

# 1標本における平均値に関する推測（２）
# 母平均と標本平均の分布

PDQ

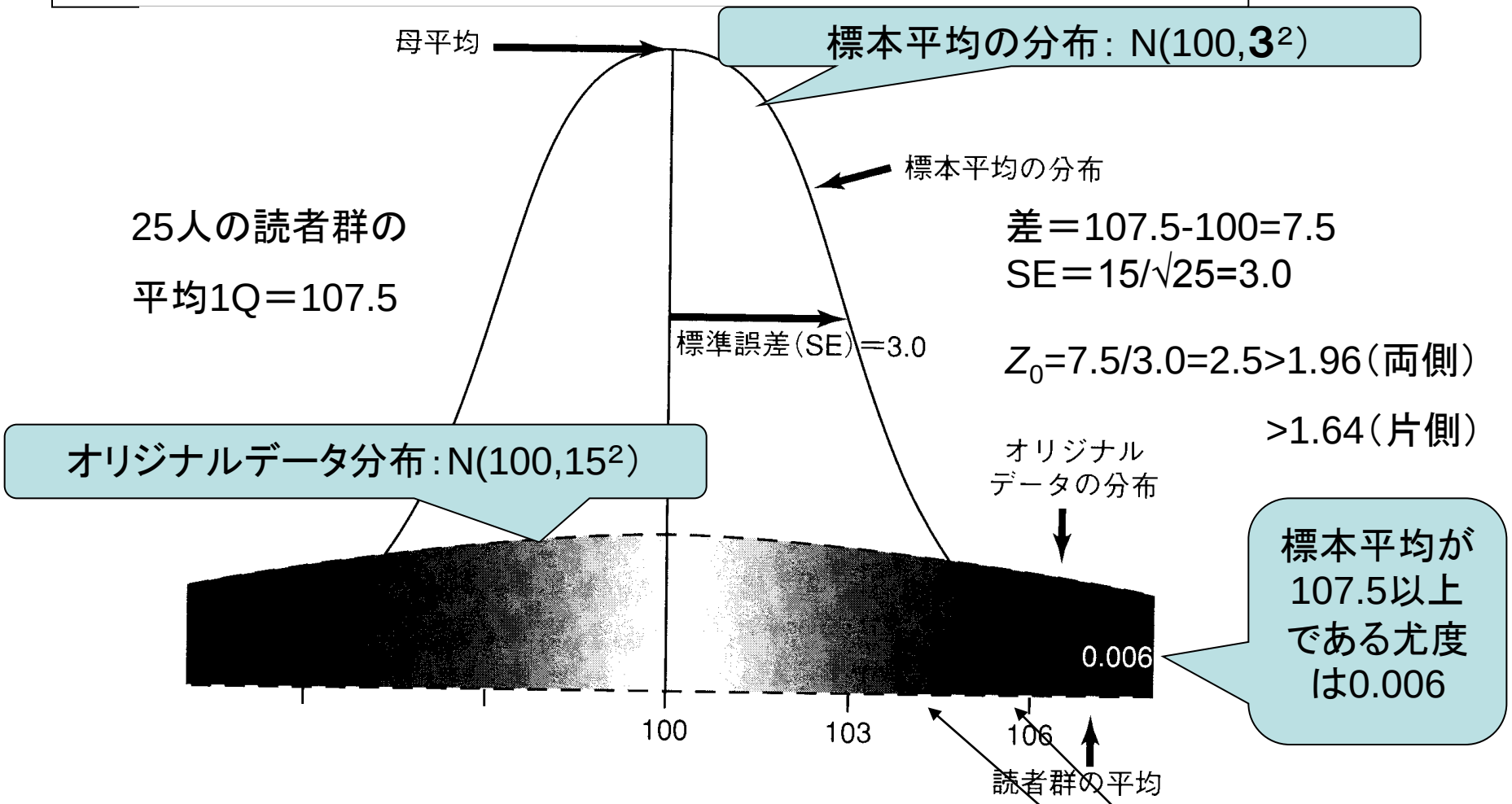



図 3-4　標本の大きさが 25 の場合の標本平均の分布

# 1標本における平均値に関する推測(３)

PDQ

**Z検定：** この母平均と標本平均の差と、標準誤差nとの比によって両者を比較する方法

差＝107.5−100=7.5

SE＝15/√25=3.0

$Z_0$=7.5/3.0=2.5>1.96(両側)

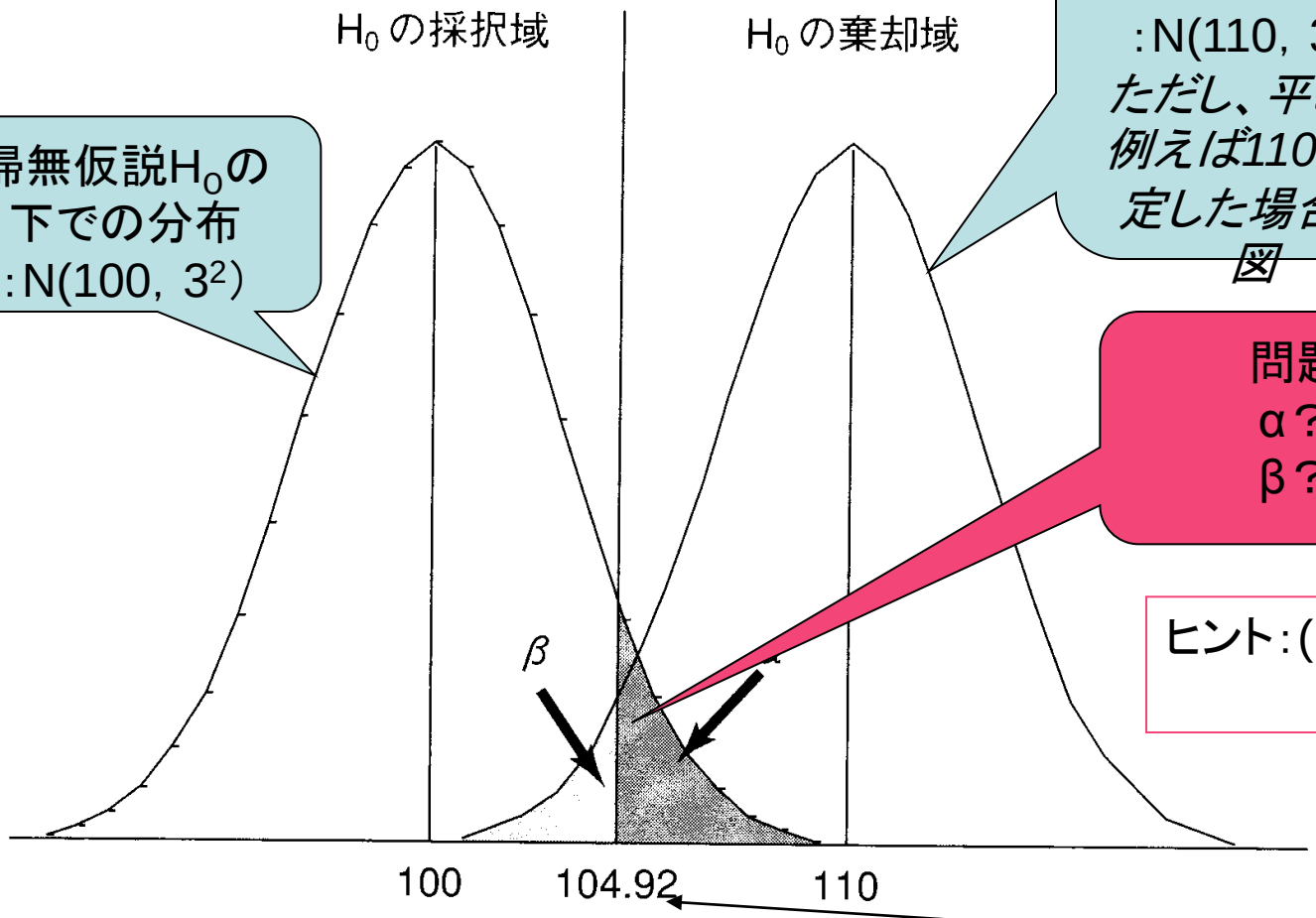
>1.64(片側)

(有意水準は5％とした場合)

標本平均が107.5以上である尤度、つまりZが2.5以上である尤度は0.006

# αとβ

PDQ




ヒント:(104.92-110)/3
=-1.69

100+1.64×3.0
片側検定

図 3-5　第 1 種の過誤と第 2 種の過誤の確率
$\alpha$ ：第 1 種の過誤の確率，$\beta$ ：第 2 種の過誤の確率．

# 第一種の過誤α(有意水準)について

**問題**

**この例ではα＝0．05が使われているが、コイン投げで表が連続して何回出ればその確率は0．05以下となるか？**

# 検出力(1-β)と標本の大きさ

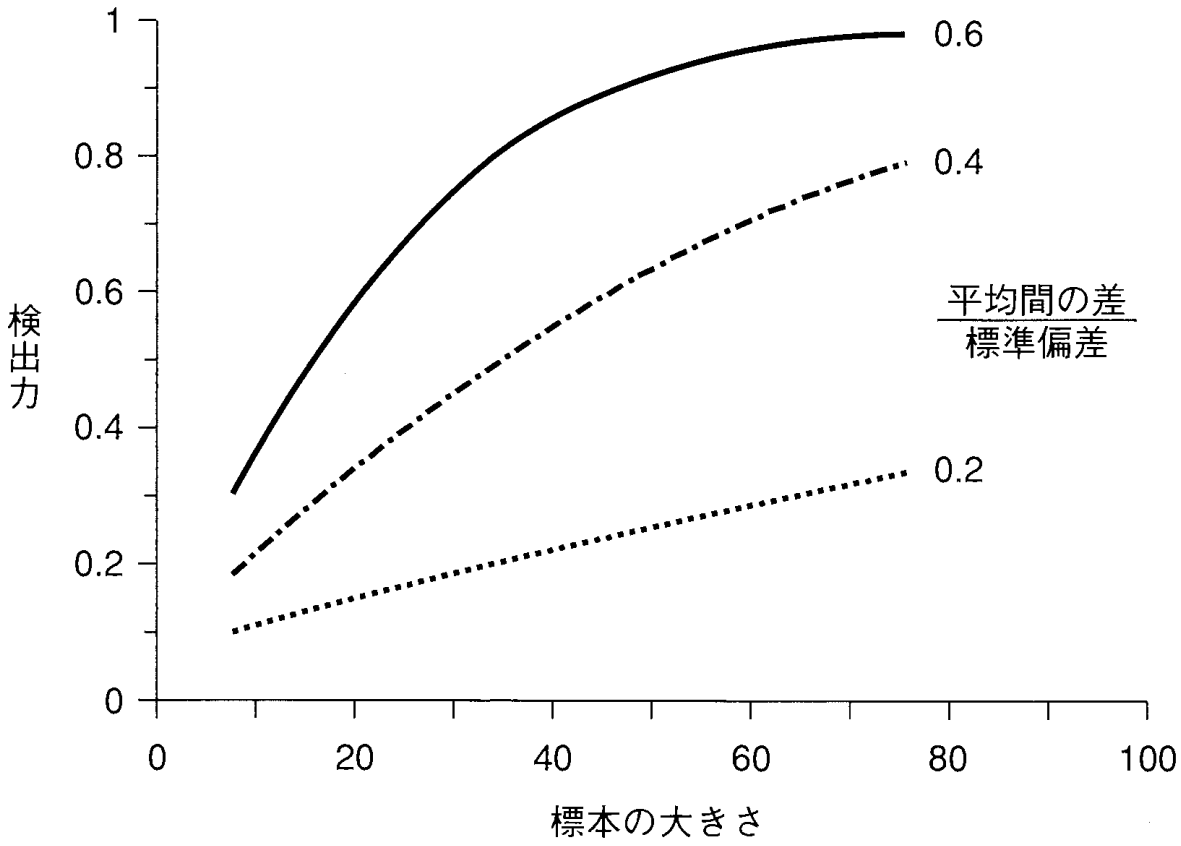



図 3-6　平均間の差と標準偏差(SD)の比をパラメータにした場合の，検出力と標本の大きさの関係

# 片側検定と両側検定

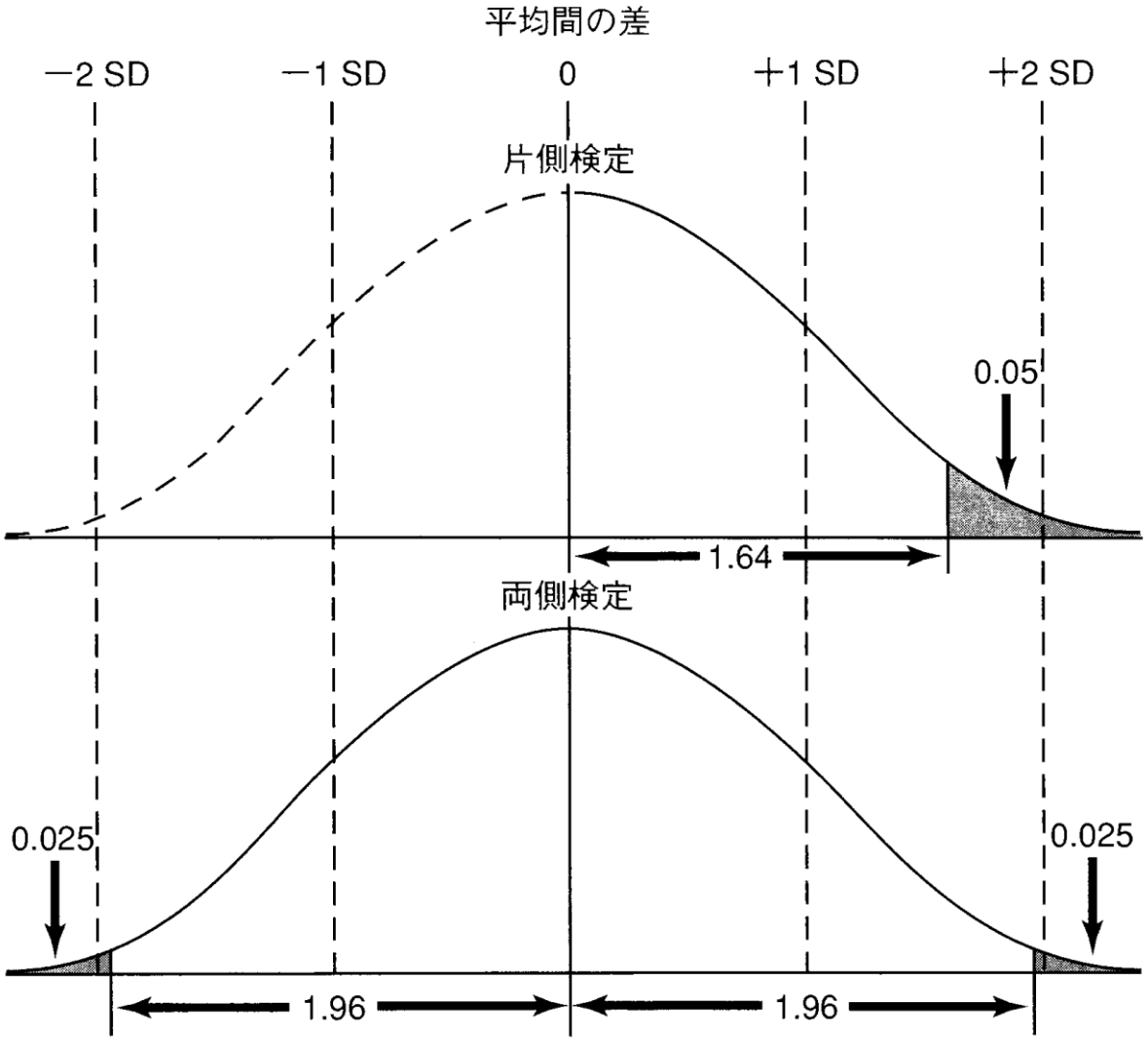



図 3-8　片側検定と両側検定における有意水準の比較

追加スライド1あり

# 両側検定のすすめ(教科書からの引用)



があれば5%の有意水準で棄却され，両側検定では1.96倍の差が要求されます(図3-8).

実際には，ほとんどすべての統計解析では，両側検定が利用されています．ある薬とプラセボ(偽薬)を比較するときでさえ，人々はその薬がプラセボに比べて有意に優れているか劣っているかを半分半分で見つけたいと思っています．この奇妙な行動は，薬が優れているという論理的領域に根差しているのではなく，平等に評価しようとする科学的領域に根差しているのです．研究者が片側検定で薬とプラセボを比較したいとしましょう．理由はわかりませんが，薬を投与した群の死亡者がプラセボを投与した群より多かった場合(笑わないでください．それは起こりうることです．クロフィブレート clofibrate を思い出してください)，研究者らは，その原因は偶然性によるものであると強いて結論するはずです．

# 95%信頼区間

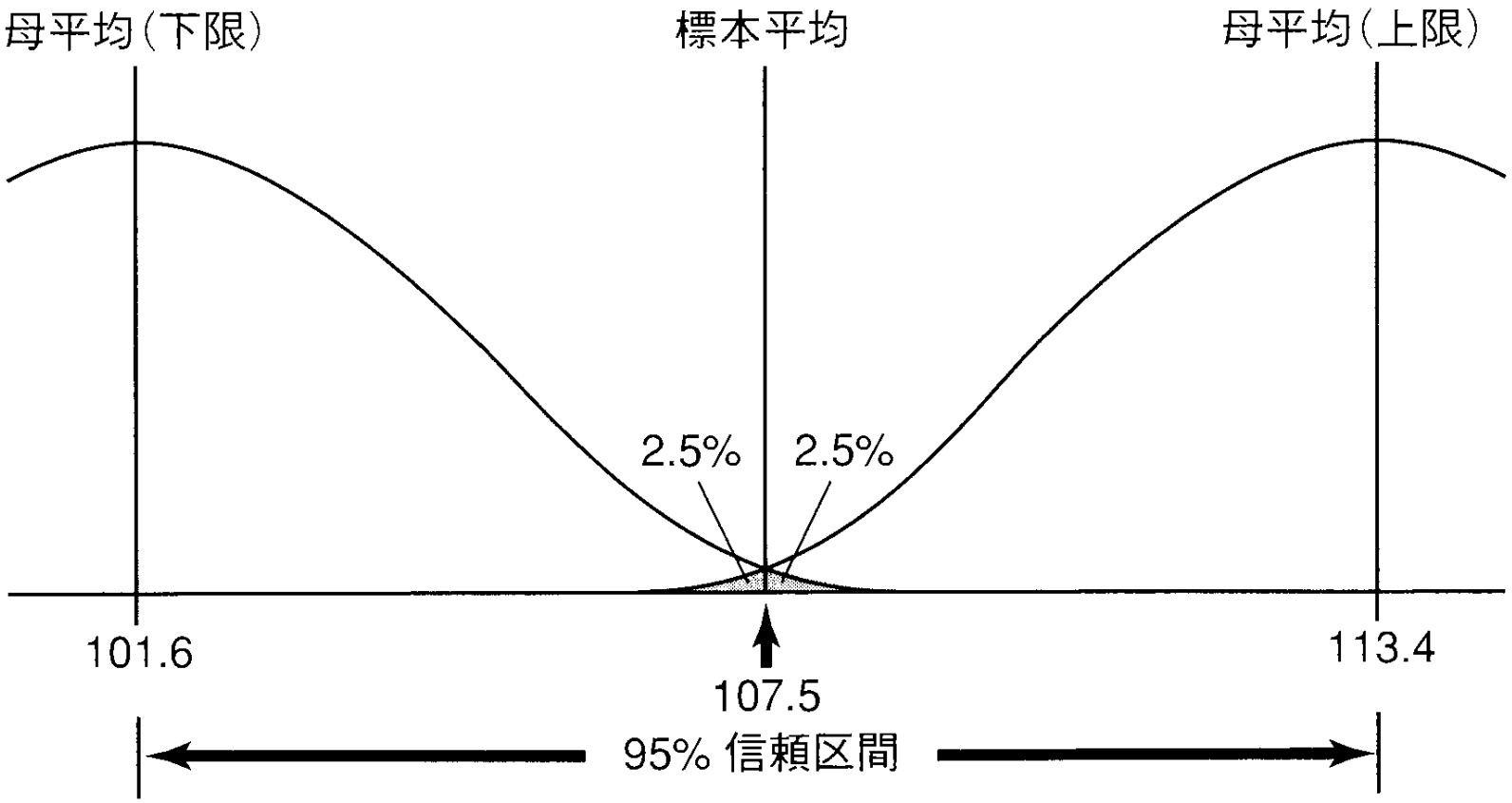



図 3-9　標本平均および母平均と信頼区間の関係

95%信頼区間＝標本平均±1.96×SE

# 標本の大きさの決定（1）

PDQ

①有意水準（$\alpha$）：通常 $\alpha=0.05$

⇨ 154 頁

②検出力（$1-\beta$）：　〃　$\beta=0.2$

③予測される比較群の差（平均間の差：d）

④バラツキの大きさ　（標準誤差：s／√n）

**図3．5(「αとβ」のスライド)において、5つの量が曲線の形を決定：**

**α、β、d、s、n**

**nは、α、β、d、s が分かれば計算できる**

# 標本の大きさの計算(2)
# (平均値の差の検定の場合)

PDQ

α＝0.05(両側検定)、β＝0.2の場合

$$n = 16\frac{s^2}{d^2}$$

**(例)α＝0.05、β＝0.2、差(d)＝10、標準偏差(s)＝15**

$$n = 16\frac{s^2}{d^2} = 16 \times \left(\frac{15}{10}\right)^2 = 36$$

**問:この式から何が言えるか？**

# 標本の大きさの計算(3)
# 比率の差の検定の場合(参考)

$$n = \left\{ \frac{4\sqrt{p(1-p)}}{|Pp - Pd|} \right\}^2 = \frac{16[p(1-p)]}{(Pp - Pd)^2}$$

**先ほどの式において、$S^2 = pq$、**
**ただし、$p = (Pp+Pd)/2$、$q = 1-p$、**
**$d = |Pp-Pd|$**

# 標本の大きさの決定（対話）ー原文

**That's how sample size calculations proceed. We take guesses (occasionally educated, more commonly wild) at the difference between means and the SD, fix the alpha and beta levels at some arbitrary figures, and then crank out the sample size. The dialogue between the statistician and a client may go something like this:**

***Statistician:*** **So,Dr. Doolittle, how big a treatment effect do you think you'll get?**

***Dr. Doolittle:*** **I dunno, stupid. That's why I'm doing the study.**

***S:*** **Well, how do 10 points sound?**

***D:*** **Sure, why not?**

***S:*** **And what is the standard deviation of your patients?**

***D:*** **Excuse me. I'm an internist, not a shrink. None of my patients are deviant.**

***S:*** **Oh well, let's say 15 points. What alpha level would you like?**

***D:*** **Beats me.**

***S:*** **0.05 as always. Now please tell me about your beta?**

***D:*** **I bought a VHS machine years ago.**

***S:*** **Well, we'll say point 2. (Furiously punches calculator.)**

**You need a sample size of 36.**

***D:*** **Oh, that's much too big! I only see five of those a year.**

***S:*** **Well, let's see. We could make the difference bigger.**

**Or the standard deviation smaller. Or the alpha level higher.**

**Or the beta level higher. Where do you want to start fudging?**

# 標本の大きさの決定（対話）ー翻訳

これが標本の大きさを計算する方法です．私たちは平均間の距離，標準偏差，αレベルとβレベルに関して，時には慎重に，通常は大胆に適当な数値を設定し，そして標本の大きさを算出します．以下は統計家と依頼主の会話です．

統計家 statistician(S)：ドーリトル先生，処理の効果についてはどの程度の大きさであると思われますか？

ドーリトル Doolittle 先生(D)：そんなばかな質問はわからないね．だからこれから研究するのだからね．

S：そうですね．10 ポイントというのはどうでしょう？

D：もちろん，いいんじゃないかね．

S：患者のデータの標準偏差はいくつぐらいでしょう？

D：ごめん，わたしは精神科医ではなく，内科医だよ．私の患者には偏りはないよ．

S：わかりました．15 ポイントではどうでしょう．それではαレベルはどうでしょう？

D：参った参った．

S：いつもどおり，0.05 にしましょう．今度はβレベルですが？

D：以前 VHS の VTR は買ったのだが．

S：わかりました．ポイント 2 としましょうか．(猛烈に電卓をたたいた後)必要な標本の大きさは 36 と出ました．

D：なんと，それでは大きすぎる．わたしは 1 年に対象患者を 5 人しか診てないんだから．

S：そうですか．えーと．私たちは平均間の差を大きくすることができます．または標準偏差を小さくするか，αレベルを大きくするか，βレベルを大きくするかですが，どこから手をつけましょうか？

# 統計的有意性と医学的有意性（1）

PDQ

- 観測された差（読者群と一般群との差；d）が偶然変動のみで生じる確率とを結びつけることで、その確率が十分に小さい場合（α以下である場合）に、その差は偶然変動のみで起こされたものではない、と結論付ける。
- 「読者の群は一般群に比べて有意に賢いようである」
  統計的有意性（statistical　significance）
- 有意性≠重要性
  「統計的にはゼロでない効果」「統計的に示された効果」
- 同等性を示すには？

# 統計的有意性と医学的有意性（2）
# 標本の大きさと平均との関係

PDQ

表 3-1　ある一定の有意水準を満たすような標本の大きさと平均との関係

| 標本の大きさ | 読者群の IQ の平均 | 一般の母集団の IQ の平均 | $p$：有意水準 |
|---|---|---|---|
| 4 | 110.0 | 100.0 | 0.05 |
| 25 | 104.0 | 100.0 | 0.05 |
| 64 | 102.5 | 100.0 | 0.05 |
| 400 | 101.0 | 100.0 | 0.05 |
| 2,500 | 100.4 | 100.0 | 0.05 |
| 10,000 | 100.2 | 100.0 | 0.05 |

問:この表から
言えることは？

# 統計的有意性と医学的有意性（3）

PDQ

## ここでのまとめと注意

- n＝4のとき、IQに10点の差があれば統計的に有意であり、n＝10000のとき、0.2点の差で有意になる。
- 任意の差dに対して、nを大きくしていけば必ず統計的有意性が示される。
- 任意の差dに対して、nを大きくしていけば、βも小さくなる。つまり、検出力1-βは大きくなる。
- nが小さい場合、真に効果があってもそれを検出できない可能性がある。
- 適切な試験とは、予備試験であらかじめdを予測し、αとβに対する適切なnを想定することであるが、・・・
- 同等性の検定

# 2つの標本平均の比較:t検定

PDQ

（t検定とは、連続変量に関する2つの平均値を比較する際に用いられる検定）

- 対応がない場合の*t*検定
（non-paired t-test）
  2つの集団の平均値を比較する

- 対応がある場合の*t*検定(paired t-test)
  処置前後の差の平均値を比較する

# 対応がない場合の*t*検定（1）

PDQ

・高血圧症に対する薬の無作為化臨床試験

・50人の高血圧症⇒　無作為割付
⇒　25人：薬投与群（処理群）
⇒　25人：プラセボの群（対照群）

・表4．1より、統計的な質問：
「処理群と対照群の平均血圧の差（4mmHg）が偶然生じた確率はどれくらいか？」

↓

仮説検定のステップを実行

・もしこの確率が十分に小さい場合、この4mmHgの差は偶然生じたものではないと考える。
つまり、この薬が有効に作用していると考える。

# 対応がない場合のt検定（2）

PDQ

**•統計的仮説検定の手順**

**①帰無仮説：$H_0$　2群の母集団の平均値は異なっていない**
**　対立仮説：$H_1$　2群の母集団の平均値は異なっている**

**②平均値の差の統計量に対する標準誤差を求める。**
**　ここでは、標準偏差の2乗を加え（$6^2+8^2=100$）、**
**　標本の大きさで割り、平方根をとること。**

$$SE = \sqrt{100/25} = 2.0$$

表 4-1　典型的な無作為化比較試験の結果

| | 処理群 | 対照群 |
|---|---|---|
| 標準の大きさ | 25 | 25 |
| 平均 | 98 mmHg | 102 mmHg |
| 標準偏差 | 6 mmHg | 8 mmHg |

# 対応がない場合の*t*検定（3）

PDQ

$$t = \frac{\text{平均値の差}}{\text{(差の)標準誤差}} = \frac{4.0}{2.0} = 2.0$$

**•統計的仮説検定の手順**
**③上の式のように、t統計量の値を求め、自由度48のt-分布から仮説の棄却域を求める。棄却域と上の数値を比べて、**

**•棄却域に含まれた場合：差が偶然生じたものではなく、薬は有効。**

**•棄却域に含まれない場合：差が偶然生じたものではないとは言えず、従って薬は有効とは言えない。**

**積極的な結論は得られない。**

**※棄却域は、両側検定、片側検定、有意水準によって、決定される**

# 対応がない場合の*t*検定の図（4）

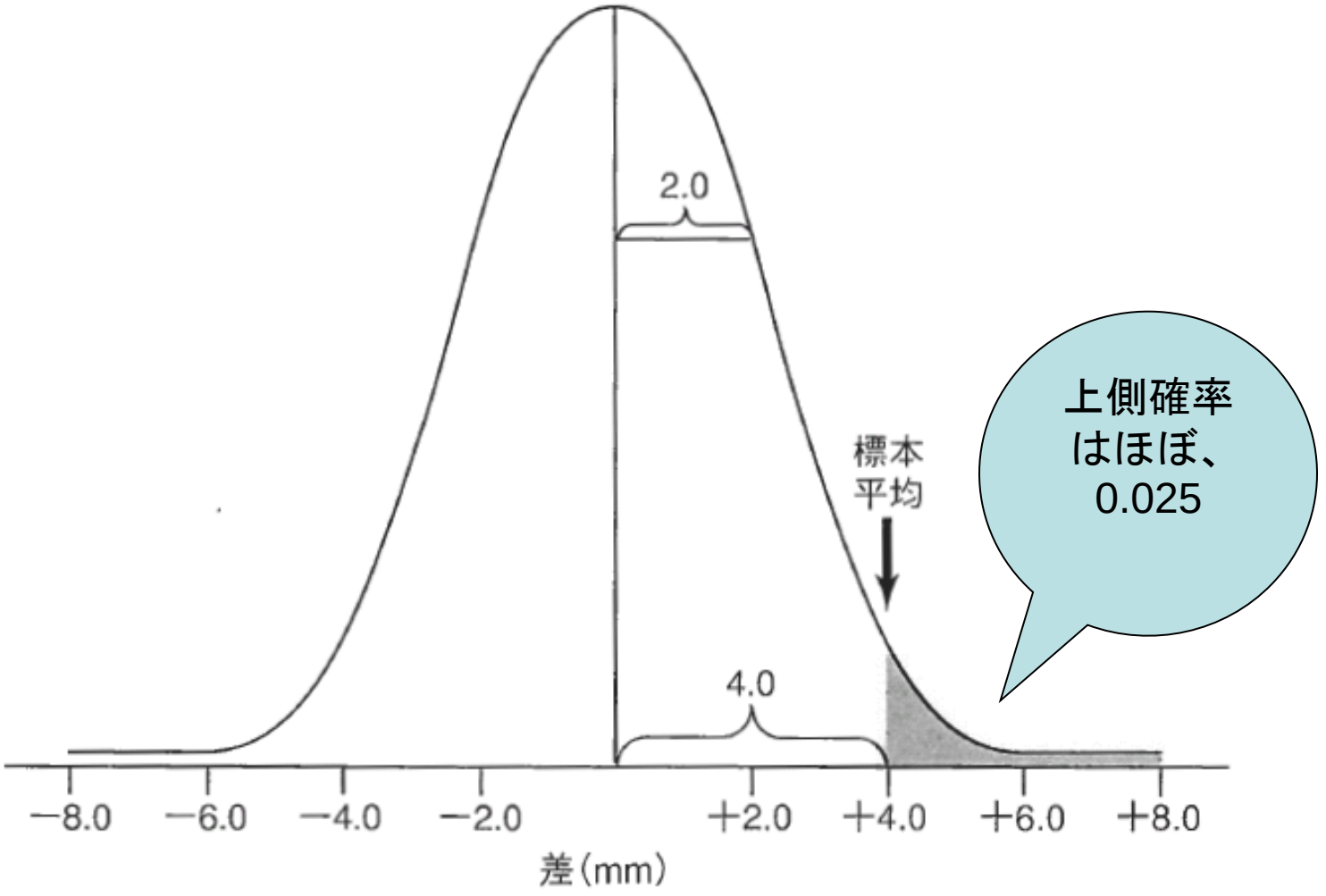



図 4-1　*t* 検定の図による説明

# 対応がない場合のt検定(5)

PDQ

- 一般的な条件での標準誤差は？

- 群1からの標本の大きさ=$n_1$、平均＝$x_1$、標準偏差＝$S_1$
  群2からの標本の大きさ=$n_2$、平均＝$x_2$、標準偏差＝$S_2$

・標準偏差　$S^2=\{(n_1-1)*S_1^2+(n_2-1)*S_2^2\}/(n_1+n_2-2)$
- t統計量　$T=(x_1-x_2)/\underline{S*\sqrt{\{(1/n_1)+(1/n_2)\}}}$

・標準偏差　$S^2=(n-1)*\{S_1^2+S_2^2\}/(2*(n-1))$
$=\{S_1^2+S_2^2\}/2$

・t統計量　$T=(x_1-x_2)/\underline{S*\sqrt{(2/n)}}$

下線部は標準誤差

下線部＝$\sqrt{\{(S_1^2+S_2^2)/n\}}$
＝$\sqrt{\{6^2+8^2)/25\}}$

# 対応がない場合のt検定（6）

PDQ

・スチューデントのt統計量（Student‘s　t）について

・統計学者ウィリアム・ゴセット（William Gossett）が開発
・・・ダブリンのギネス醸造所の品質責任者をしていたため、彼の発表を真に受けないと考え、Studentというペンネームで発表
・標本の大きさnが小さいときは、分布の大きさが正規分布と大きく異なる。

・2つの母集団の分散が等しい場合は、スチューデントのt検定を用いることができるが、等しくない場合はウェルチの検定を用いる。
（ウェルチの検定では、t統計量の分母の標準偏差の推定量が少し異なることと、自由度が調整されている点である。）

# 対応がある場合のt検定（1）

PDQ

**•高血圧症に関する臨床試験で、今度は、患者群に対して収縮期血圧を測定し、薬を投与する。**

**•そして、1，2月経過後再度血圧を測定する。**
**下の表は5人の患者を被験者として考えた場合。**

表 4-2　拡張期血圧に対する高血圧症の薬の効果

| 患　者 | 投与前血圧 | 投与後血圧 | 差 |
|---|---|---|---|
| 1 | 120.0 | 117.0 | −3 |
| 2 | 100.0 | 96.0 | −4 |
| 3 | 110.0 | 105.0 | −5 |
| 4 | 90.0 | 84.0 | −6 |
| 5 | 130.0 | 123.0 | −7 |
| 平均 | 110.0 | 105.0 | −5.0 |
| 標準偏差 | 15.8 | 15.7 | 1.58 |

# 対応がある場合のt検定(2)

PDQ

- **薬の投与前と投与後ということは忘れて、独立2群として先の対応がない場合のt検定を用いた場合**
  **SE**＝√{(15.8²＋15.7²)／5.0}＝9.96
  **t統計量の値**＝-5.0／9.96＝-0.502

有意とはならない

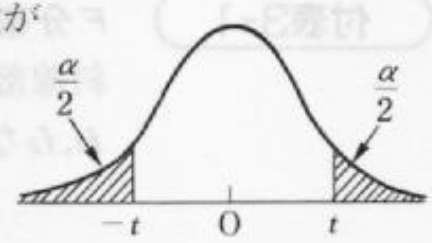
- **しかし、独立2標本の場合とは異なり、標準偏差が違ってくる。**
  **差のデータ**{-3,-4,-5,-6,-7}**の標準偏差**＝1.58
- **標準誤差**＝1.58／√5＝0.71
- **t(ペアード)＝差の平均値／(差)の標準誤差**
  ＝5.0／0.71＝7.04

有意水準0.00001で有意

**※検定量としては、1標本平均の検定と同様**

## 付表2

$t$ 分布表：$t$ と $-t$ との外側を合計した面積が $\alpha$ である．$d.f.$ は自由度を表わす．




| d.f. \ α | 0.500 | 0.400 | 0.300 | 0.200 | 0.100 | 0.050 | 0.020 | 0.010 | 0.001 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 | 636.619 |
| 2 | .816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 31.599 |
| 3 | .765 | .978 | 1.250 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 12.924 |
| 4 | .741 | .941 | 1.190 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 8.610 |
| 5 | .727 | .920 | 1.156 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 6.869 |
| 6 | .718 | .906 | 1.134 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.959 |
| 7 | .711 | .896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 5.408 |
| 8 | .706 | .889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 5.041 |
| 9 | .703 | .883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.781 |
| 10 | .700 | .879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.587 |
| 11 | .697 | .876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.437 |
| 12 | .695 | .873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 4.318 |
| 13 | .694 | .870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 4.221 |
| 14 | .692 | .868 | 1.076 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 4.140 |
| 15 | .691 | .866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 4.073 |
| 16 | .690 | .865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 4.015 |
| 17 | .689 | .863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.965 |
| 18 | .688 | .862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.922 |
| 19 | .688 | .861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.883 |
| 20 | .687 | .860 | 1.064 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.850 |
| 27 | ..... | | | | | 2.052 | ..... | | |
| 60 | .679 | .848 | 1.045 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.460 |
| 80 | .678 | .846 | 1.043 | 1.292 | 1.664 | 1.990 | 2.374 | 2.639 | 3.416 |
| 120 | .677 | .845 | 1.041 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.373 |
| ∞ | .674 | .842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.291 |

# 多群の平均値の比較：分散分析（１）

PDQ

**分散分析ANOVAとは、3群以上の平均値間比較を行う方法です。一元配置分散分析は、単一のカテゴリカルな独立変数（または独立因子）を取り扱う方法です。要因分散分析はさまざまな状況における複数因子に関して行う方法です。**

**（PDQにおける5章の頭書）**

# 多群の平均値の比較：分散分析（2）

PDQ

・アセチルサリチル酸系の多くの薬について、鎮痛作用に有意な差が存在するかどうかに興味があるとする。

・問題：　今、A薬からB,C,・・・F薬までの6種類の医薬品間の比較をすべて行う場合、${}_6C_2$=15通りの比較を行うが、有意水準5%で繰り返し仮説検定を行うとする。15通りの比較の中で少なくとも1回以上有意となる確率は？

# 多群の平均値の比較：分散分析（３）

PDQ

・解答　　約54％

・理由　　$1-(1-0.05)^{15}=0.5367$

・2群間」の比較に用いるt検定を繰り返し用いることは不適切となる。なぜなら、上に示したように有意水準が上昇してしまうから。

・「検定の多重性」と呼び、これを回避するために全体として有意水準をたとえば5％にするような、多重比較の方法が開発されている。

# 多群の平均値の比較:分散分析(4)

PDQ

表 5-1　6 種類のアセチルサリチル酸系の鎮痛薬に対する患者の評点

| 患者 | 鎮痛薬 A 薬 | B 薬 | C 薬 | D 薬 | E 薬 | F 薬 |
|---|---|---|---|---|---|---|
| 1 | 5.0 | 6.0 | 7.0 | 10.0 | 5.0 | 9.0 |
| 2 | 6.0 | 8.0 | 8.0 | 11.0 | 8.0 | 8.0 |
| 3 | 7.0 | 7.0 | 9.0 | 13.0 | 6.0 | 7.0 |
| 4 | 8.0 | 9.0 | 11.0 | 12.0 | 4.0 | 5.0 |
| 5 | 9.0 | 10.0 | 10.0 | 9.0 | 7.0 | 6.0 |
| 平均 | 7.0 | 8.0 | 9.0 | 11.0 | 6.0 | 7.0 |
| | | | 全平均＝8.0 | | | |

$H_0$＝すべての平均値は等しい

$H_1$＝すべての平均値は等しくなると限らない

# 多群の平均値の比較：分散分析（5）

PDQ

## 群間変動と群内変動

群間変動を表す平方和＝（群平均－全平均）$^2$ の和×N

$$(7-8)^2+(8-8)^2+(9-8)^2+(11-8)^2+(6-8)^2+(7-8)^2=16.0$$

$$16.0\times 5=80$$

群内変動を表す平方和＝（個体の観測値－群平均）$^2$ の和

$$=(5-7)^2+(6-7)^2+(7-7)^2+(8-7)^2+(9-7)^2+\cdots+(6-7)^2$$

$$F=\text{群間平均平方}/\text{群内平均平方}$$

# 多群の平均値の比較:分散分析(6)

PDQ

# 分散分析(ANOVA)

表 5-2　分散分析

| 要因 | 平方和 | 自由度 | 平均平方 | $F$ 比 |
|---|---|---|---|---|
| 群間 | 80.0 | 5 | 16.0 | 6.4 |
| 群内 | 60.0 | 24 | 2.5 | — |
| 全体 | 140.0 | 29 | — | — |

**一元配置分散分析(one-wayANOVA)**

# 多群の平均値の比較：分散分析(７)

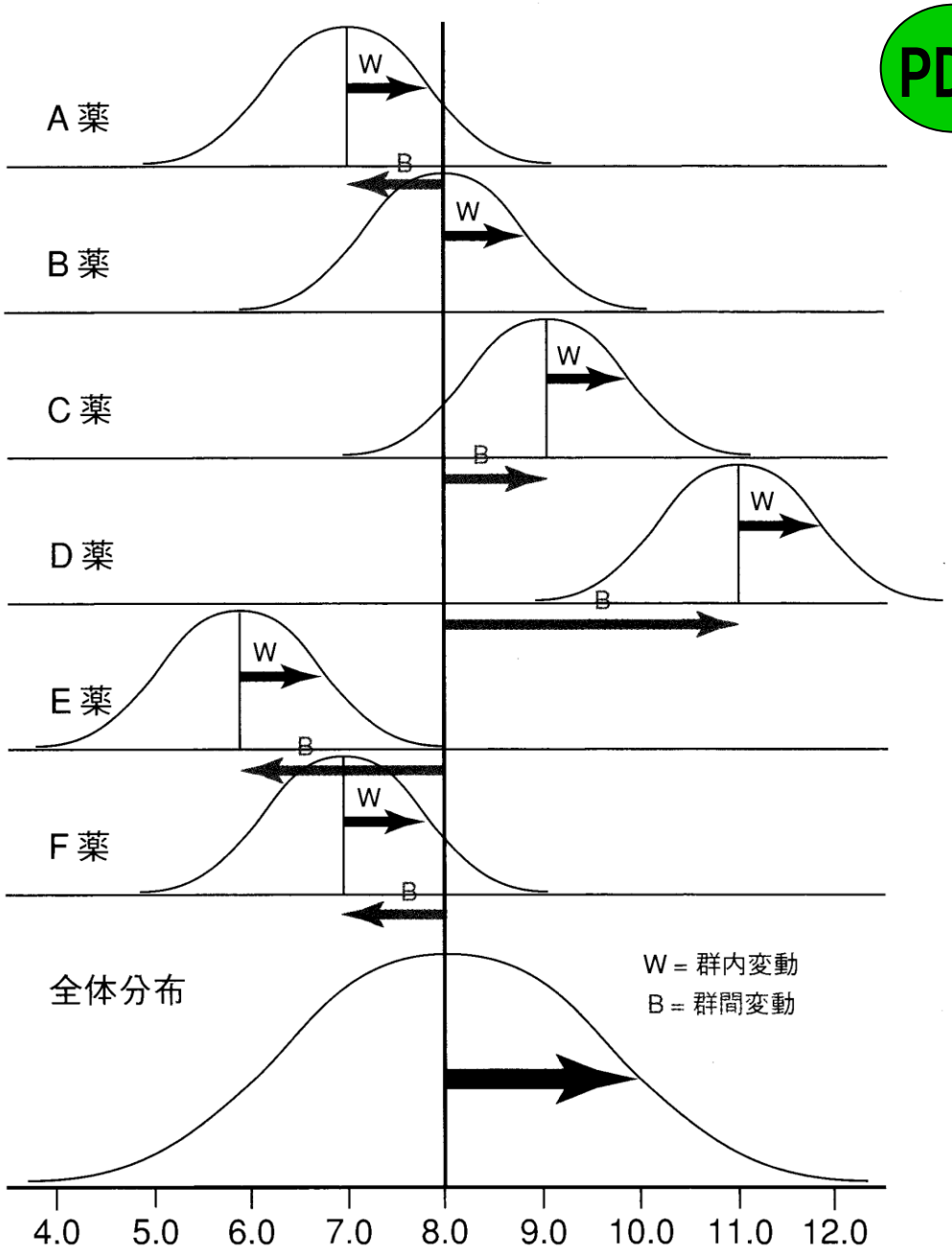



図 5-1　一元配置分散分析の構造

# CRAP　Detectors：例3－1

LipidsResearchCIinics（LRC）の臨床試験では，30万人の男性を検査し，コレステロール値が上位1%の3，000人の中で，心臓病がなく，処置に十分に協力してくれた人を選びました。本当の薬を投与する群とプラセボの群に標本を無作為に割り付け，10年後，心臓病で亡くなった人が対照群，つまりプラセボ群では38人，薬の投与群では30人でした。横定の結果は，p＜0･05，つまり5%で有意となりました。そこでマーガリンの会社は，バターを使わずにマーガリンを使えば同様の効果があるため，心臓病の予防につながると訴えました.

問：勉強した知識から，皆さんはマーガリンに変えますか？

# 例3-1に対する解答と手引き

これは大標本に関する誤解です．最初に，薬の効果は統計的には有意であったかもしれませんが、臨床的には些細なものです。2番目に，この結果は男性で高コレステロール値の人々には適用できますが，次の人には適用できません。(1)女性，(2)コレステロール値が低い人，(3)マーガリンを使っている人。残念ながら，この例は実際にあったものです。

手引き1　大標本に要注意。統計的には有意でも，臨床的にはとるに足らないこともありうるのです。

手引き2　結果の一般化に要注意。研究の結果は、その研究で用いた標本と類似する母集団にのみ適用可能です。

# CRAP　Detectors：例3－2

リウマチ学者が，関節炎の患者10人に対し銅製のブレスレットの効果について研究を行いました。彼は26種類の項目を測定し，リウマチ因子の血中レベルに顕著な減少傾向を見つけました。しかし，痛みや機能については効果を見いだせませんでした。彼はリウマチ因子の血中のレベルに銅が影響を与え，それ以外は何も影響を与えないと結論づけました.

問：彼の結論を借用しますか？

# 例3-2に対する解答と手引き

これは小標本の例です。第1に、10人の患者では何かを示すのに十分な検出力をもちえないでしょう。つまり，どのリウマチ学者であってもその他の25の項目について間違いなくいえるのは，差を示すことができなかったということであり，差がなかったということではありません。実際には帰無仮説を証明することはできないでしょう．第2に，26項日も検定すれば，1つぐらい偶然に有意になるかもしれません，つまり何もいえないということです。

手引き1　小標本に要注意。有意な差を検出することは困難であり，差があることがいえなければそれは何もいえないということになります。

手引き2　多重比較の検討。この種のデータを処理する特別な方法があります。

# CRAP　Detectors：例4－1

奇妙な話ですが，本当のことです・数年前まで，狭心症に対する一般的な処置は，乳房の動脈を外科的に結紮することでした。患者は手術の前後で痛みの程度を評価しました。t検定を行ったところ，手術によって痛みの程度が有意に軽減されたと結論されました。引き出された結論は，その治療が有効であるということでした.

問：皆さんは信じますか？

# 例4-1に対する解答と手引き

小さな点ですが，この計画では対応があるt検定を行う必要があります。特に断りがない場合、対応がないt検定を指しますが，この場合，それではまずいわけです。のちに臨床試験のところで示されることですが，大きな点として，患者の半分は動脈の結紮を行わない，つまりプラセボの影響をみる必要があるということです。因果関係究明のために、常に対照群を置く必要があるということが，倫理的な問題としてあるのです。

手引き

2群間の差や，ある群の中の差を検出しようとした場合，2つの標本において独立変数にのみ違いが出て，その他の因子には違いが出ないようにする必要があります。

# CRAP　Detectors：例4－2

ベネットBennttらは，診断学と治療学の2つの領域で，比較評価に関する知識の増大の効果をみるために無作為化試験を行いました．彼らは，比較評価に関する訓練を受けた処理群と受けなかった対照群に属する学生に対し，事前テストと事後テストを施しました。対応があるt検定の結果，処理群において診断学に関してはp＜0．001で有意になり，治療学に関してはp＜0.01で有意になりました。対照群では両方とも有意になりませんでした。そこで調査者らは，比較評価の訓練は効果があったと結論づけました。

問：皆さんはこれと同じ方法で解析しますか？

# 例4-2に対する解答と手引き

この方法では，本質的には対照群を無視した解析となっています。処理群と対照群の対比を鮮明にするために，彼らは両群の差の得点に対して対応がないt検定を行うべきです。実際，その調査者らは，両方の解析結果を示しました.

手引き

1番目に、複数の検定から結論を導くことは，第1種の過誤のαレベルを満たさなくなってしまい，さらにはその結果の解釈も非常に難しいものになります。2番目に、処理群と対照群で別々に対応がある方検定を行うことは，統計的には不合理なものです。

# 医学関係論文における統計解析関係の記述のガイドライン

- International Committee of Medical Journal Editors:Uniform Requirements for Manuscripts submitted to biomedical Journals
  Ann. Intern. Med. 1988, 108: 258-265

# 「統計ガイドライン」の解説論文

- Bailer JC, Mosteller F: Guidelines for Statistical Reporting in Articles for Medical Journals - amplications and explanations. Ann. Intern. Med. 1988, 108: 266-273
- 丹後敏郎：消化器病学に関する研究論文での統計的方法について―これだけは知っておきたい基本的指針―第3回：統計ガイドライン，日消誌　1992，1：75-82
- 医学統計ハンドブック（宮原英夫、丹後敏郎編）、23章統計手法のまとめ方と英語論文における表現法、pp.668-677，朝倉書店

# 統計解析に関する記述を明確に

1）Statistical　Analysis

・・・Methodsの中に「Statistical　Analysis」のsub-sectionを設定

(1)データ要約方法を明記する:

・mean±SD　or　mean±SE

・肝機能検査値のように高値に裾を引く場合、percentile　median　25%、75%、range

（2）解析目的（仮説検定）を詳述し、適用する仮説検定名の明記

（3）検定の場合、片側検定（one-tailed）か両側検定（two-tailed）

（4）事前に設定した有意水準の明記

# 統計解析に関する記述を明確に（続）

２）Results
解析結果を示すときには、利用した統計手法を明記。
多くの統計手法を使用した場合、統計手法の併記

３）Tables and Figures
図表において、解析(特に検定)結果を表現する場合は本文とは独立に、解析内容と使用した統計手法名（検定の場合は片側、両側の区別）を明記

原文1）Describe Statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results.
12)Put general description of methods in the Methods section, specify the statistical methods used to analyze them.
15)Define statistical terms, abbreviations, and most symbols.

# 信頼区間も報告しよう

- 検定結果として、p valueだけを報告する傾向があるが、これは正しくない。探索的名報告ではどこに差があるかが問題となるので検定が重視されるが、検証的な報告では検定で有意性を評価するだけでなく、平均値（の差）の大きさ、有効率(の差)の大きさなどの評価指標の差、比を推定することに意味がある。
- この場合には、そのバラツキの大きさを表現する信頼区間を併記することが必要不可欠となる。
- 原文2）When possible, quality findings and present them with appropriate indicators of measurement error or uncertainty(such as confidence intervals)
  3)Avoid sole reliance on statistical hypothesis testing, such as the use of P values, which fails to convey important quantitative information.

# 研究対象、母集団を明確に

- どのような患者、または動物を対象にしているかを明確にしなければならない
- 患者特性に関する情報が不明：例“Thirty consecutive patients with liver cirrhosis who visited our outpatients clinic were recruited for this study...”
- 特に臨床試験の中では、少なくともstudy protocolの中で、患者（標本）のeligibility criteria (inclusion & exclusion)について明確に記述
- 原文4)Discussion eligibility of experimental subjects.

# ランダマイゼーション

- 統計的推測が可能な基本原則「標本が無作為に母集団から抽出されること」・・・ある病院のデータをまとめるとき、研究目的とする対象疾患の縮図となりうるか？
- 処置の比較研究・・・異なる処置を施された患者間の背景（交絡）因子の比較可能性が重要、「患者の無作為割付」が必須条件
  : : 方法（単純無作為化、マッチング、層化法）や実施法（電話法、封筒法）の明記
  : : 多施設臨床試験：施設間差解消のための施設内バランスをとるための患者割付情報の明示
- 原文5)Give details about randomization

# 標本の大きさを明記しよう

- 調査、実験で何例の患者、動物を利用したかは、統計処理の基本
- 比較的大きな患者を要する臨床試験では、想定する仮説を検証するために必要な標本の大きさを試算しておくことが重要（患者保護、効率的な研究遂行）
- 治療効果の評価：研究期間中の中止、脱落、除外による評価時点ごとの例数の明記
- 原文8)Give numbers of observations.
  9)Report losses to observation (such as dropouts from a clinical trial).

# その他

- 統計手法の引用は標準テキストを
  ・原著論文より、わかりやすく例題豊富なtextbookや総説を
  ・原文10)References for study design and statistical methods should be to standard works (with pages stated) when possible, rather than to papers where designs or methods were originally reported.
- 利用した統計パッケージを明記する・・・プログラム名も
  ・原文11)Special any general-use computer programs used
- 統計言語を明確に区別しよう
  ・原文14)Avoid non-technical uses of technical terms in statistics, such as ”random”(which implies a randomizing device) ”normal”,”significant”,”correlation” and ”sample”.