

カラーイメージで学ぶ

〈新版〉

基礎統計学の

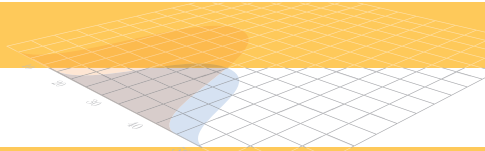
第2版

市原清志
佐藤正一
山下哲平

山口大学名誉教授

国際医療福祉大学准教授

滋慶医療科学大学院大学講師



CONTENTS

本書で使用する主な記号と意味	3
検定法と検定統計量	4
本書で取り扱う統計処理法の分類	5
検定法の使い分け	6

第1章 統計学とは 9

■ 統計学とは	10
統計学は、バラツキを伴う情報を客観的に分析・評価する学問	10
我々は、日常いろいろな数値を読み経験的に判断している	10
サイエンスでは、統計がなぜ必要とされるか？	11
経験的な“統計処理”は主観的であいまい	11
サイエンスでは、未知の情報を扱うため、客観的な“統計処理”が不可欠となる	11
バラツキがなければ、統計処理はいらない	13
母集団と標本	13
標本はどの程度ばらつくのか？	14
■ 統計の2つの機能	16
機能1：少ない情報から全体像を把握する（推計学）	16
統計学的仮説検定と有意確率	16
機能2：データを要約する（記述統計学）	17

第2章 記述統計 19

■ 標本の分布	20
度数分布の概念と作り方	20
統計量 ～標本の特徴を表す要約値～	21
分布の中心位置の表し方	21
分布の広がり（散布度、変動）の表し方	21
探究：標準偏差の計算で、 n ではなく $n-1$ で割る理由	23
参考：他の箱ひげ図作成法	27
探究：分布の形状を示す統計量（歪度と尖度）	28
■ 正規分布の特徴と使い方	32
正規分布とは	32
正規分布と信頼区間確率	32
正規分布となる計測値の例	33
正規分布とならない計測値の例	33
よく見られる分布型	35

正規分布の標準化と z スコア	36
標準正規分布表の見方	37
■ 分布の特徴点の表し方	38
z スコアによる、分布の中での相対位置の表し方	38
片側有意確率 P による分布中の相対位置の表し方	38
参考：パーセンタイル（百分位数）による相対位置の表し方	41
探究：p パーセンタイルの値 x の求め方	43
参考：計測尺度と統計処理方式	44

第3章 検定の原理 45

■ 統計的仮説検定の目的と理論	46
有意差検定では反証の論理が使われる	47
■ 検定の原理を考えよう（平均値の検定）	50
標本平均の理論分布と標準誤差（SE）	52
シミュレーションで考えよう 標本平均の分布	54
実行手順	54
実行例	55

第4章 関連2群の差の検定 57

■ 1 標本 t 検定（パラメトリック法）	58
1 標本 t 検定の概念	58
検定の手順	59
1 標本 t 検定における t 分布とは	64
正規分布と t 分布の違い	65
t 分布表の見方	66
1 標本から求めた各種統計量の理論分布	67
シミュレーションで考えよう 標本の t 値の分布	68
■ 統計学的推定（平均値の検定の場合）	70
探究：母平均の 95% 信頼区間の推定	72
探究：自由度（df：degree of freedom）	73
■ 一標本 Wilcoxon 検定（ノンパラメトリック法）	74
検定の概念	74
検定の手順	75
注意：統計量 T の求め方：チェックポイント	76
Wilcoxon 検定における統計量 T の理論分布	78
Wilcoxon T 検定表の見方	78

- 2標本t検定(パラメトリック法) 82
 - 2標本t検定の概念 82
 - 検定の手順 83
 - シミュレーションで考えよう 平均値の差の標準化値tの分布 88
 - 実行手順 88
 - 実行例 89
 - 2標本から求めた各種統計量の理論分布 90
 - 2標本t検定におけるt分布とは 91

- 統計学的推定(平均値の差の検定の場合) 92

- 等分散性の検定(F検定) 94
 - 検定の手順 94
 - F分布表の見方 96
 - 参考:F分布の形状と期待値E(F) 97
 - 参考:正規分布の加法定理 98

- Mann-Whitney検定(ノンパラメトリック法) 102
 - 検定の手順 102
 - 検定の概念 103
 - Mann-Whitney検定と統計量Uについて 103
 - 同順位があるとき 103
 - Mann-Whitney検定の統計量Uの理論分布 108
 - $n_1=2, n_2=2$ の場合 108
 - $n_1=3, n_2=2$ の場合 108
 - $n_1=3, n_2=3$ の場合 109
 - $n_1=n_2 \geq 4$ の場合 109
 - 2標本t検定の制約とMann-Whitney検定との使い分け 111
 - シミュレーションで考えよう 検出力の比較 112
 - 実行手順 112
 - 実行例 母集団が正規分布の場合と対数正規分布の場合の比較 113

- 感度・特異度・ROC解析 116
 - 感度と特異度 116

- ROC分析による2群の判別度の分析 119
 - ROC分析と曲線下面積 120

- カットオフ値の設定法 121
 - 感度・特異度曲線の利用 121

有病率によるカットオフ値の調整	122
偽陽性や偽陰性の重要性を考慮した調整	123

第7章 出現度数に関する検定 125

■ 一要因の場合	126
比率の検定 (二項検定)	126
二項検定の概念	126
検定の手順	128
注意：個別確率と有意確率の区別	131
シミュレーションで考えよう 出現度数の分布	132
実行手順	132
実行例	133
母比率 p と試行回数 n によって変わる二項分布の形状	134
母比率 p 既知：出現度数 r の分布と出現比率 r/n の分布の関係	135
母比率 p 未知：観察した比率 p_0 から母比率 p を推定するには	135
探究：二項分布を正規分布に近似するための連続補正	137
探究：正規近似できない場合の出現度数 (比率) の区間	138
χ^2 適合度検定 (多項分布の検定)	140
検定の手順	140
χ^2 分布表の見方	142
■ 2要因の場合	144
χ^2 独立性検定 (2×2 分割表検定)	144
検定の手順	144
参考： 2×2 分割表検定は、比率の差の検定と同じ	147
Fisher の直接確率計算法	148

第8章 独立多群間の比較 153

■ 多群間の同時比較が必要な場合	154
■ 一元配置分散分析 (one-way ANOVA)	155
検定の概念	155
検定の手順	156
分散分析を利用した日間 CV、日内 CV の求め方	161
■ Kruskal-Wallis検定	164
検定の概念	164
検定の手順	165
Kruskal-Wallis 検定の統計量 H の理論分布	168
分散の均一性の検定 (Bartlett 検定)	170
一元配置分散分析法の制約と Kruskal-Wallis 検定との使い分け	172

■ 相関係数	174
相関係数の定義	175
探究：偏差積和 ⇒ 共分散 ⇒ 相関係数	176
参考：相関係数の別の定義	176
偏差積和（共変動）の意味を $n=4$ の場合で考えてみよう	177
偏差積和（共変動）の意味を $n=10$ について考えてみよう	178
正の相関の場合の偏差積和	178
負の相関の場合の偏差積和	179
無相関の場合の偏差積和	179
単相関係数の検定	180
標本相関係数の理論分布	185
シミュレーションで考えよう 相関係数の分布	186
実行手順	186
実行例	187
■ スピアマン順位相関係数	188
スピアマン順位相関係数 r_s の概念	188
スピアマンの順位相関係数の特性	189
検定の手順	190
■ 回帰直線（linear regression）	195
何を独立変数にするかで回帰式が変わる（回帰の方向性）	198
探究：最小二乗法の原理と回帰直線	199
探究：回帰直線による予測の求心性と線形関係式	200
回帰直線を計算するときに注意すべき点	202

Q1. パラメトリック検定では、分布の正規性を検定で確認する必要がありますか？	204
Q2. 検定法によって判定が異なる場合、どう対処すればいいですか？	206
Q3. 片側検定と両側検定をどのように使い分けるのですか？	207
Q4. 有意差検定の有意水準は常に $P=0.05$ でいいのですか？	208
Q5. データ数が大きく、有意差検定が無意味な場合、研究結果をどのようにまとめればいいですか？	209
Q6. 臨床試験などの介入研究では、計画段階でデータ数の設定が要求されるのはなぜですか？	
また何を目安に設定するのですか？	211
参考：有意差検定における α エラーと β エラー	215
Q7. 多群間で 2 群ずつ検定したところ、査読者から検定回数による有意確率の補正を要求されました。なぜ補正が必要なのですか？	217
Q8. 観察研究では群間比較に有意差検定を使えないって本当ですか？	220

■ 実験してみよう	226
実験結果記録用紙 (実験 4a) $k = 4, n_1 = n_2 = n_3 = n_4 = 4$ の場合	233
実験結果記録用紙 (実験 4b) $k = 3, n_1 = n_2 = n_3 = 5$ の場合	234
計算例	235
標本相関係数記録用紙 ($n=6$)	238
標本相関係数記録用紙 ($n=12$)	239

解答集

241

統計表

269

付 録

281

■ StatFlexのインストール	282
インストール手順	282
StatFlex 製品版との違い	282
StatFlex 本体のインストール	283
■ StatFlexの機能と基本的な使い方	284
基本事項	284
データ形式と視点	284
変数型と自動グラフ	286
データの取り込み (新規作成)	288
貼り付け時の自動確認と修正	289
演習データの読込み	290
StatFlex の基本的な使い方	290
グラフの調整法	290
データの並べ替え	292
値の一時除外	292
統計表機能	293
統計量→確率の計算機能	294

索 引

301

はじめに

情報社会の急速な進展により、日常的なデータであれ、科学的に得られたデータであれ、それらを正しく分析し、有効に活用するニーズがますます高まっている。データ分析を的確に行うには、統計処理は必須であり、統計学を理解した上での利用が求められる。

統計処理を正しく理解するには、統計処理がどのようなプロセスで進められていくのかについて、明確なイメージを描いて学習していくことが重要となる。そのために本書では、様々な視点から多数のシエマを用意し、統計理論をできるだけ分かりやすく理解していただけるように努めた。

特に、統計学の理論は確率論に基づいており、標本抽出に伴うデータの確率論的な揺らぎの特性の理解が求められる。そこで、常に図式的に提示することで、統計学の背景にある確率論の意味をイメージできるようにした。

本書は、好評いただいている1990年に出版した「バイオサイエンスの統計学」(南江堂)を発展させた形で構成した。各種統計手法の基本的な概念と処理手順の解説に加えて、より深い理解のために、「探求」、「参考」、「注意」の欄を設けて解説した。また、図中に示した「ここがポイント」では、その重要点を明瞭に述べ、理解を促すようにしている。

なお巻頭では、多様な統計処理法を分類・整理し、それらの全体像を把握できるようにした。さらに検定法と統計処理法のまとめ図表から、検定法間の関連性と使い分けについても、すぐに確認できるよう工夫している。また、10章の「適切な統計処理法に必要な考え方：よくある質問に答えて」では、これまで曖昧にされてきた、誰もが抱く統計処理上の様々な疑問点を取り上げて、その本質的な部分に触れて回答し、適切な統計処理の指針とした。

本書では、実際に統計処理を行えるようにすべく、著者らが開発した統計ソフト(機能限定版)を付けている。これは単なる学問としての統計学ではなく、実データを用いて読者自身が統計処理を体験的に学んでいただくことを目指している。このソフトが持つ統計シミュレーション機能により、統計学の根本にある、標本抽出理論(母集団と標本統計量の関係)を、乱数データを発生させて視覚的に理解できるようにした。

本書を利用することで、統計学をより身近なものに感じてもらえればと願っている。なお、本書作成にあたっては、StatFlex 開発チームのメンバーである佐藤和孝氏、川野伶緒氏に、付属統計ソフトの開発のみならず、編集作業においても多大な援助をいただいたことに、改めて感謝の意を表す。

2014年9月 市原清志
佐藤正一

第2版によせて

市原による「バイオサイエンスの統計学」(1990年南江堂)は、基礎統計学についての知識を広く、深く取り扱っており、これまで多くの読者を得ている。しかし、やや難解な内容も含んでおり、初学者には敷居の高い内容となっていた。そこで、大学1年の教養課程における統計学の授業での利用を想定して、内容を簡略化した「カラーイメージで学ぶ統計学の基礎」を2006年に刊行し、2011年に同書の改訂版を出版した。しかし、教養課程で利用するには、さらなる簡略化や補足説明の必要な箇所が多々あった。

そこで2014年に全面的な改定を行い、統計データの解析を行う際に多くの方々が疑問をもつ点に関して、より明快な解説を試みた「カラーイメージで学ぶ新版統計学の基礎」を刊行した。幸い、教養課程の教科書としての価値が認められ、より多くの大学で採用されるようになった。ただ、解説の流れや説明文の明解さなどに問題があったことや、誤字・誤植もいくつか存在した。

今回の増刷にあたり、細部にわたり記述内容を見直し、かつ授業でより使いやすくするため、頁の配置を全体的に整理した。特に、統計学の理解を助けるシミュレーションを付録から本文の該当箇所に再配置し、標本のばらつきや統計量への理解を高める工夫を行った。また、類似した内容の記述をまとめ、主要な図を作り直した。さらに、その要点を把握しやすくするため「ここがポイント」の数を増やすとともに、表現をより簡潔なものとした。新たな検定法としては、分布の正規性を確認するための歪度と尖度について解説を追加した。

この改訂により、統計学の理論とデータ処理法に関し、よりスムーズな理解が得られ、授業ばかりではなく、独学でも本書から統計学の基本を十分習得していただけることを、執筆者一同願っている次第である。

2016年8月 市原清志
佐藤正一
山下哲平

本書で使用する主な記号と意味

記号	意味
AUC	曲線下面積
α	有意水準
μ	母平均
μ_T	Wilcoxon T 検定統計量の期待値
μ_U	Mann-Whitney U 検定統計量の期待値
σ^2	母分散
σ	母標準偏差
Σ	累和
σ_T	Wilcoxon T 検定統計量の標準誤差
σ_U	Mann-Whitney U 検定統計量の標準誤差
χ^2	χ^2 統計量
χ^2_α	χ^2 統計量の有意水準 α に対する有意点
a	回帰直線の y の切片
b	回帰直線の y の傾き
CV	変動係数 (Coefficient of variation)
${}_n C_r$	n 個から r 個取り出す 組み合わせの数
CI	信頼区間 (Confidence Interval)
d	偏差 (特に対をなす観測値の差)
\bar{d}	対をなす観測値の差の平均値
df	自由度
E	期待度数
$E(x)$	確率変数 x の期待値
F	F 統計量 (分散比)
H	Kruskal-Wallis の検定統計量
H_0	帰無仮説
H_1	対立仮説
IQR	四分位範囲
N	総データ数
n	標本データ数

記号	意味
$N(0, 1^2)$	標準正規分布
$N(\mu, \sigma^2)$	平均値 μ , 分散 σ^2 の正規分布
NS	not significant 「有意でない」
O	観察度数
P	有意確率
p	母比率
p_o	出現比率
QD	四分位偏差
r	実現度数
τ	標本相関係数
ρ	母相関係数
r_S	スピアマンの順位相関係数
s	標本標準偏差
S	偏差平方和
SD	標準偏差
SE	標準誤差
S_{xx}	x の偏差平方和
S_{yy}	y の偏差平方和
S_{xy}	x, y の偏差積和
s_x	x の標準偏差
s_y	y の標準偏差
s^2	標本分散
T	Wilcoxon の検定統計量
t	t 統計量
t_α	t 統計量の有意水準 α に対する有意点
U	Mann-Whitney の検定統計量
$Var(x)$	確率変数 x の分散期待値
\bar{x}	標本平均
$\bar{\bar{x}}$	総平均
Z	Z 変換値
z	標準化値
z_α	標準正規分布の 有意水準 α に対する有意点



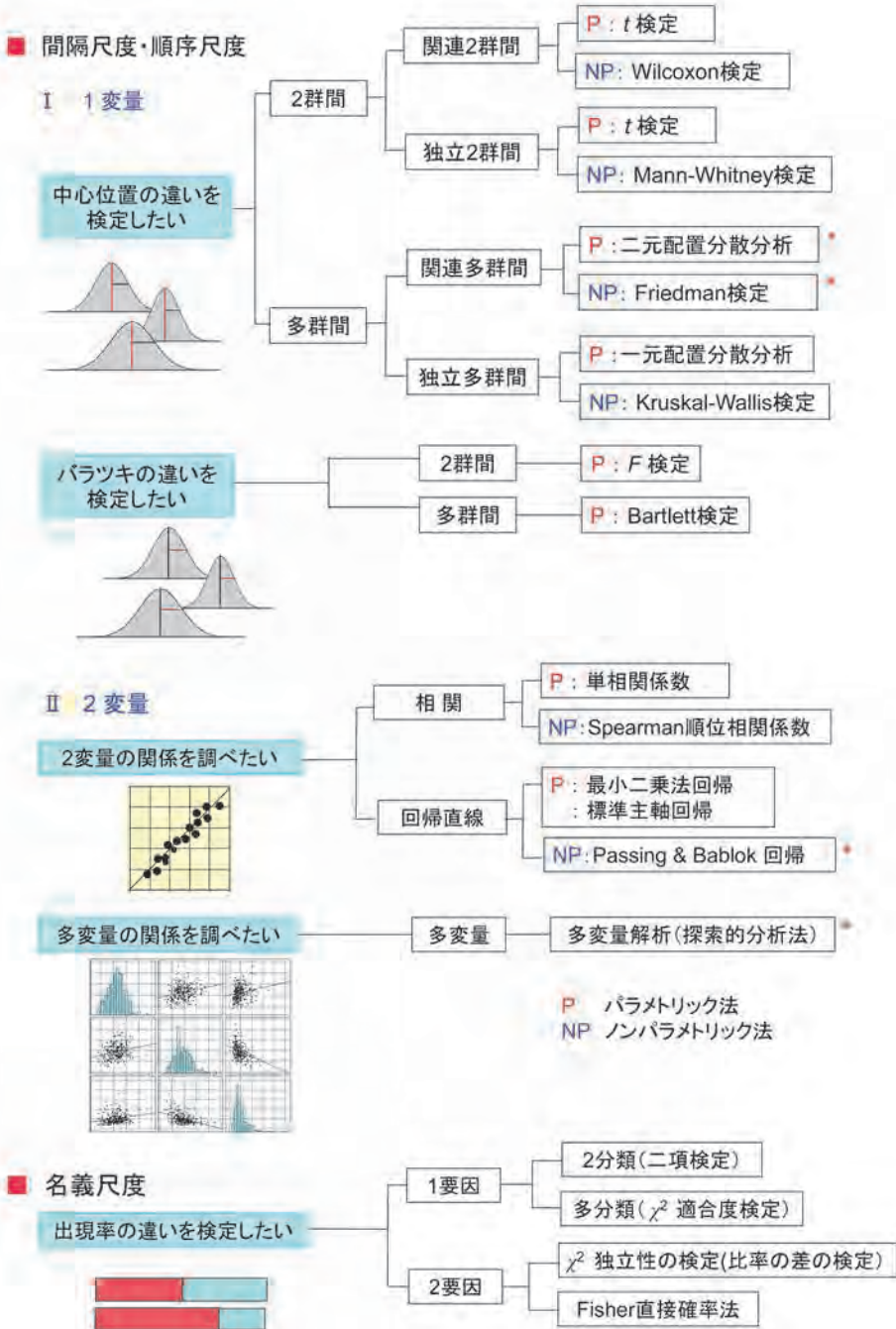
検定法と検定統計量

データ形式	検定法	P/NP	検定統計量	自由度 (df), データ数 (n)	統計表
	1 標本 t 検定	P	$\bar{d} \Rightarrow t$	$df = n - 1$	
n 組	Wilcoxon 検定	NP	$T \Rightarrow z$	n	$n \leq 25$ は Wilcoxon 検定表
	2 標本 t 検定	P	$\bar{x}_1 - \bar{x}_2 \Rightarrow t$	$df =$ $n_1 + n_2 - 2$	
	Mann-Whitney 検定	NP	$U \Rightarrow z$	n_1, n_2	$n_1, n_2 \leq 20$ のとき Mann-Whitney 検定表
	一元配置分散分析	P	F	$df_A = k - 1$ $df_E = N - k$	k は群数 N はデータ総数
k 群	Kruskal-Wallis 検定	NP	$H \doteq \chi^2$	$df = k - 1$	k は群数、但し $k = 3, n \leq 17$ は Kruskal-Wallis 検定表
	χ^2 適合度検定	NP	χ^2	$df = k - 1$	k は分類数
k 分類					
	2×2 分割表	NP	χ^2	$df = 1$	
	単相関係数の検定	P	$r \Rightarrow t$	$df = n - 2$	n はデータ組数
	スピアマン順位相関係数	NP	$r_s \Rightarrow t$	$df = n - 2$	n はデータ組数

P : パラメトリック法 NP : ノンパラメトリック法 \Rightarrow : 標準化

本書で取り扱う統計処理法の分類

統計処理は、以下のような内容について調べたいときに利用される。
 * は本書では取り扱っていない



検定法の使い分け

●適用要件による使い分け

データ形式	検定法	方式	分布型の制約	計測尺度の制約	分散の制約
	1 標本 t 検定	P	分布の正規性 *1	間隔尺度 *3	—
	1 標本 Wilcoxon 検定	NP	なし	なし	—
	2 標本 t 検定	P	分布の正規性 *2	間隔尺度 *3	2 群の等分散性
	Mann-Whitney 検定	NP	なし	なし	なし *4
	一元配置分散分析	P	分布の正規性 *2	間隔尺度 *2	分散の均一性
	Kruskal-Wallis 検定	NP	なし	なし	なし *4
	単相関係数	P	分布の正規性 *2	間隔尺度 *3	—
	スピアマン順位相関係数	NP	なし	なし	—

P = パラメトリック法 NP = ノンパラメトリック法

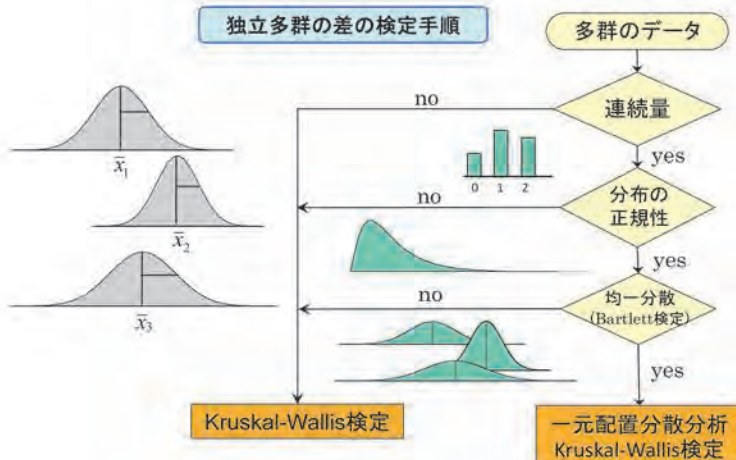
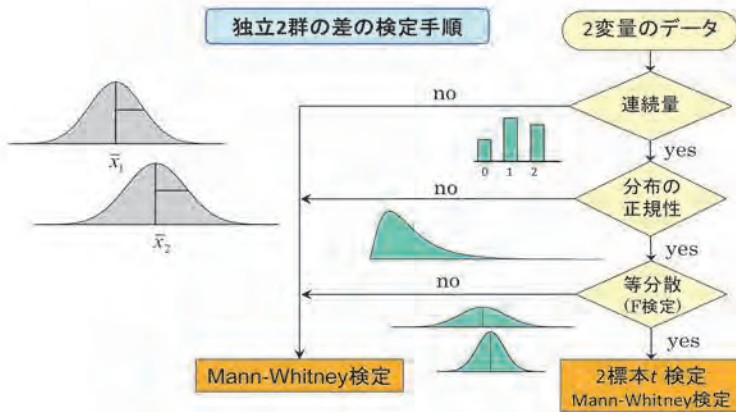
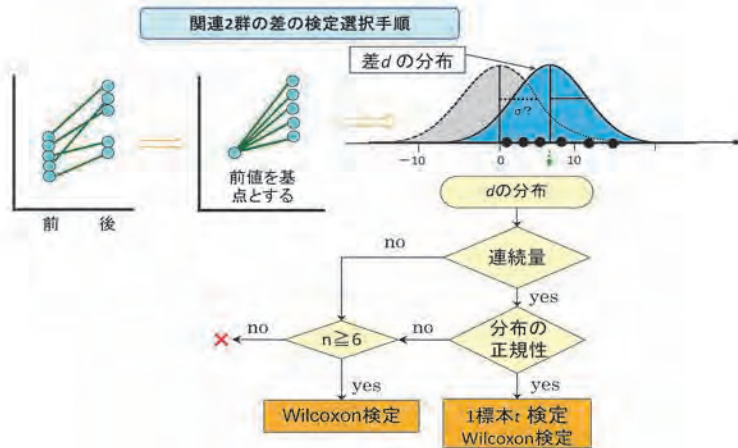
1. 差 d の分布の正規性。データ数が大きくなると制約なし。
2. 各群の分布の正規性。データ数が大きくなると制約なし。
3. 順序尺度でも、細かく段階分けされている場合には制約なし。
4. ノンパラメトリック法では、基本的に群間での分散の不一致を問わないが、分散が有意に異なる場合には、仮に位置関係のずれが有意でないとしても、分散が異なるという意味で、群間差があると解釈する必要がある。

●検出力による使い分け

データ形式	検定法	方式	正規分布	一様分布	片側に歪んだ分布	両裾広がり分布
	2 標本 t 検定	P	◎	◎	×	×
	Mann-Whitney 検定	NP	○	◎	◎	◎
	一元配置分散分析	P	◎	◎	×	×
	Kruskal-Wallis 検定	NP	○	◎	◎	◎
	単相関係数	P	◎	◎	×	×
	スピアマン順位相関係数	NP	○	◎	◎	◎

◎良好 ○可能 ×不可

母集団が正規分布と仮定できる場合には、パラメトリック法は、有意確率によりの確な判定を行え、かつノンパラメトリック法よりも、少しだけ差の検出力が高くなる。一様分布の場合には、一般にどちらの方法でも、差の検出力に大差はない。しかし、分布の非対称性が強い場合や極端値がある場合など、明らかに正規分布とみなせない場合には、ノンパラメトリック法のほうが常に差の検出力が高く、最適な検定法となる。

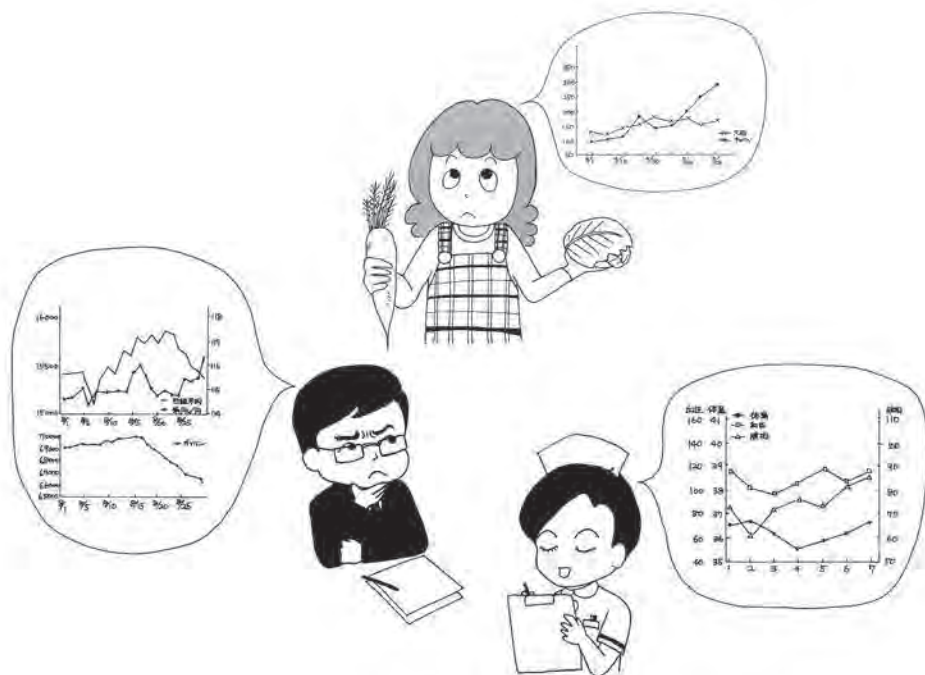


得られるデータの性質から、どの統計処理法が適切であるのかを示すフロー図である。これは検定法の使い分けの目安であるが、絶対的なものではない。一般に厳密な実験の場合には、データを見てからその特性によって検定法を選ぶことは不適切で、実施前に用いる検定方法を決めておくことが要求される(ランダム化比較試験)。よって、データ分布の特性が不明な場合には、分布に依存しないノンパラメトリック法がふさわしいと言える。

統計学は、バラツキを伴う情報を客観的に分析・評価する学問

我々は、日常いろいろな数値を読み経験的に判断している

体重であれ、野菜や果物の値段であれ、日常取り扱う数値を、繰り返し調べる（計測する）と、常にゆらぎ（バラツキ）があり、一定ではない。我々は、その数値が大きいのか小さいか、等しいか異なっているかを、経験的に処理して判断している。



日常我々は数値を無意識に“統計処理”している

この意味で、我々は日常出くわす数値に対して、“常識”に従って無意識のうちに“統計処理”を行っているのだが、その判断は主観的なため曖昧であり、また経験の程度に依存するので、人によって異なる。

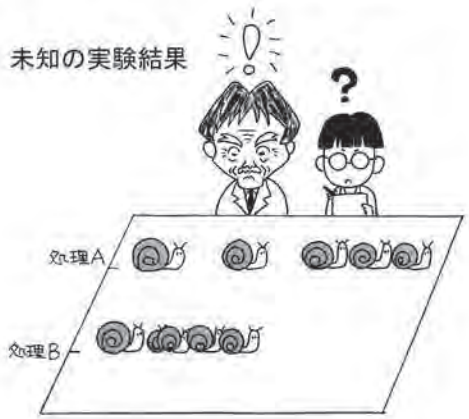
サイエンスでは、統計がなぜ必要とされるか？

経験的な“統計処理”は主観的でありまい

統計学を用いると、バラツキを伴う数値が、確率の理論に基づいて処理されるので、判断をより客観的に行えるようになる。

サイエンスでは、未知の情報を扱うため、客観的な“統計処理”が不可欠となる

サイエンス（科学）の分野でも、様々な計測が行われ、それに基づいていろいろな評価や判断をする。しかし、新しい実験をしたり、調査をするとき、多くの場合、未知の情報であるが故に、その計測値がどの程度ばらつくのかを予測できない。このため、統計による客観的な判断が求められる。



複雑な調査データ

性別	年齢	BMI	TCHO	TG	HDL-C	LDL-C	TP	BS	HbA1c
0	67	21.41	258	169	46	178.2	6.9	109	5.2
1	82	22.73	216	107	44	150.6	7.2	110	5.1
1	51	21.39	186	88	55	113.4	7.2	99	5.0
0	54	23.93	207	136	43	136.6	7.2	281	5.6
1	51	20.92	203	125	66	112	7.6	109	5.1
0	59	25.69	227	288	53	116.4	7.3	118	5.0
0	51	24.22	169	146	45	94.6	7.2	70	4.8
0	59	23.83	220	112	60	137.6	7.7	89	4.9
0	52	24.51	190	85	61	112	8.1	90	4.7
1	57	24.56	205	120	55	126	7.2	141	7.0
1	52	22.41	264	97	90	154.6	8.0	124	6.1
1	50	30.09	223	246	45	128.8	7.8	99	5.3
1	30	23.24	186	168	54	98.4	7.4	67	4.9
1	66	24.20	205	104	48	138.2	7.1	95	5.1
0	73	17.33	164	87	70	76.6	7.0	85	4.6
0	75	26.02	285	156	65	188.8	7.9	110	5.0
1	24	28.63	265	158	42	191.4	7.1	96	4.8
0	35	24.75	201	132	47	127.6	7.6	98	4.9

人は元々数値に弱い。情報量が多く、たくさんのパラメータからなる新しい研究データを眺めてもわからない。その基本構造を把握するために統計処理が必要となる。



従って、医学、行動科学、理工学など、新しいものを探求するサイエンスでは、いかなる計測を行っても、常に、計測した情報を統計学に基づいて処理し、解釈することが要求される。

Key word キーワード

サイエンス（科学）と統計

「科学」 science とは、「知識や経験を実証可能な形で体系化してゆく知的活動（学問）」である。実際には、研究の成果（観察、調査、実験により得た結果とその解釈）を、実証（再現）可能な形で科学論文の形にまとめて報告する。そして、その論文を共有することにより、知識や経験が体系化され、この一連の活動（科学）により世の中が進歩・発展してゆくことになる。

一般に、科学は、**自然科学**（物理学、化学、生物学、医学）、**社会科学**（経済学、法学）、**人文科学**（心理学、言語学）などに分類される。

しかし、どの分野の科学においても個々の研究成果は、過去の報告や仮説との違い（新規点）を明らかにした上で公表することが要求される。ここで科学研究では、もともと新規で未知な情報を取り扱うことが多く、「差の程度」についての「常識」が働かないので、「客観的」な判定の手段として統計学が必要となる。

また、統計学は複雑な研究データを要約し、様々な角度から分析する手段を提供するので、研究結果の解釈を的確に行う上でも重要となる。



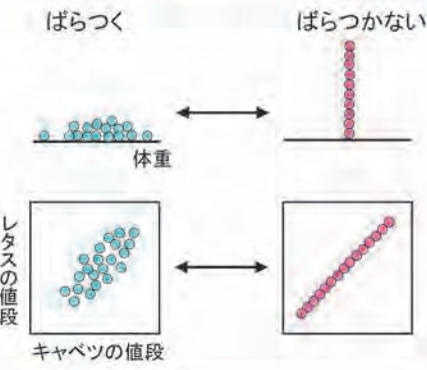
科学論文では新規性や従来との違いが重要視され、統計学はその判定で必要となる。また、複雑な調査データを的確に読み解く手段を提供する。

バラツキがなければ、統計処理はいらない

バラツキがなければ、1種類の計測値の時は1回測れば済み、2変量の関係を求めるときには、2回測れば関係が決まる。しかし、実際には計測するたびに値がばらつくため、繰り返し測定が必要になる。

統計学では繰り返し計測の情報から、バラツキの傾向を把握し、確率的に様々な推論を行う。すなわち統計学は、計測値のバラツキを科学的に解析し、それに基づいて最も妥当な判断を行う手段を提供する。

“バラツキ”とは



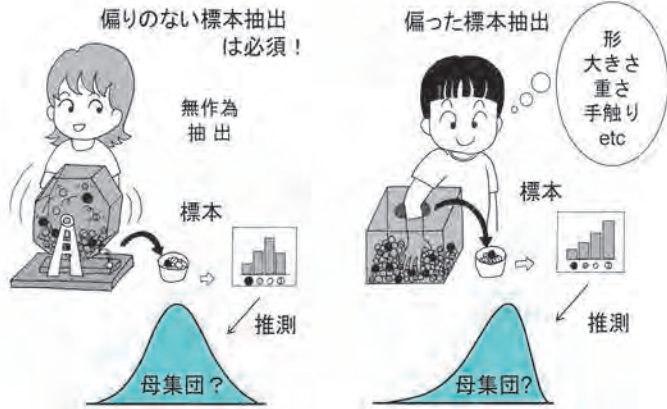
統計学では、反復測定で“バラツキ”を調べ、それを基準に変化を判断する。バラツキ情報なしでは統計処理できない。



母集団と標本

分析対象となる全事象の計測値の集合を**母集団** population と呼び、そこから取り出した計測値の部分集合を**標本** sample と呼ぶ。母集団は無限のことが多く、有限でも計測には多大な時間と労力を要する。そこで統計学では、母集団から適当なサイズの標本を取り出して計測し、その結果から全体（母集団）の状態について様々な推論を行う。

この場合、できるだけ偏りなく標本を選ばないと、母集団の状態を正しく把握できない。この偏りなく標本を選び出す方法を**無作為抽出** random sampling と呼び、ランダムに対象を選ぶ工夫が必要となる。



統計では一部の標本から全体を推測するので、標本の“無作為”抽出が求められる

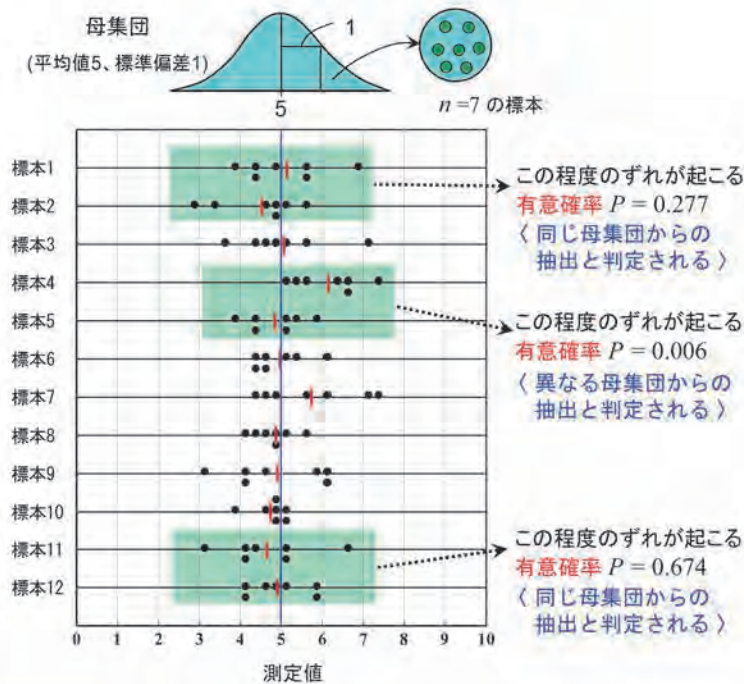
標本はどの程度ばらつくのか？

下図は、ある母集団から、7つのデータを取り出し分布を調べる実験を、計12回繰り返した結果である。

標本の分布は、その広がりが大きい場合（標本7、標本11など）や、逆にかなり小さい場合（標本8、標本10など）がある。また、その中心位置が、母集団の中心とほぼ一致している場合（標本3）や、かなり離れている場合（標本4、標本7）が見られる。実際には、このゆらぎの程度は、標本サイズ（データ数）に依存して変化する（15頁の図を参照）。

個々の標本のゆらぐ範囲は、母集団の分布についてある仮定を置き、標本サイズを指定すれば、統計学的に予測可能となる。

また、任意の2つの標本について、その平均値のずれが、どの程度の確率で生じるかを推定できる。下の例では、標本サイズが小さいので、偶然とはいえ、任意の2つの標本の間には様々なずれが見られるが、それぞれについて、確率的にどの程度起こりうるかを後述の**有意確率**の形で計算できる。



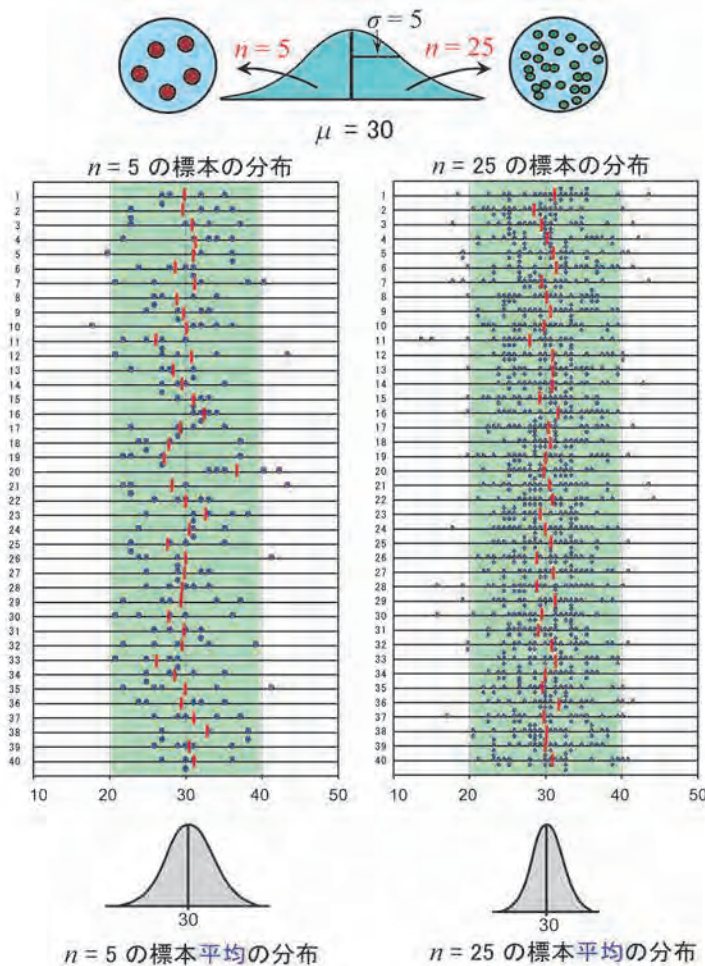
同じ母集団からの標本でも、取り出すたびにその分布が揺らぐ。偶然極端な標本も生じうるので要注意。

統計学では、**標本サイズ**によって、どの程度のゆらぎが生じるかを理論的に説明できるため、標本や母集団について様々な分析や推論を行える。

例えば、正規母集団（母平均= 30、母標準偏差= 5）の中から、データ数（標本サイズ） $n = 5$ と $n = 25$ の標本を計40回取り出した場合に、標本の値の広がり（分布）がどの程度ゆらぐかを調べてみると、次の図のような結果が得られた。

標本サイズが $n = 5$ のとき、赤い線で示した中心位置（平均値）は大きくゆらいているが、これに対して $n = 25$ のときには、平均値のゆらぎは小さい。この情報を利用して、観察した平均値の偏りについて確率的な推論が可能となる。（標本平均のシミュレーション 54 頁を参照）

標本数による標本平均分布の変化



統計理論により、データ数に応じた標本平均の変動範囲を推定できる

～ 統計学をその機能から分類すると、大きく 推計学 と 記述統計学 に分けられる ～

機能1：少ない情報から全体像を把握する（推計学）

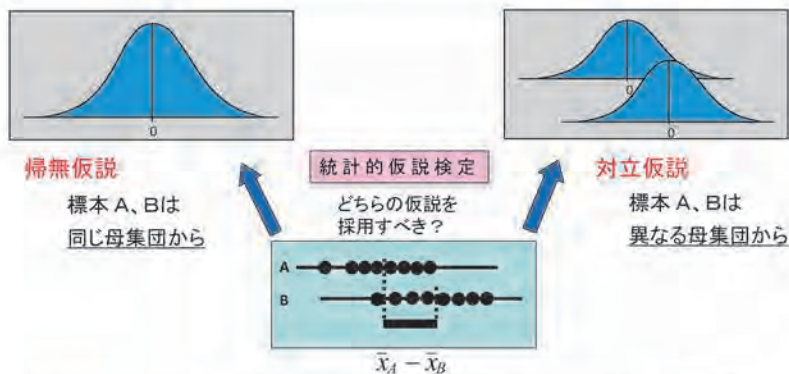
統計学的仮説検定と有意確率

統計学を使うと、全てのデータ（母集団）を網羅的に調べなくても、そこから取り出した少数のデータ（標本）から、母集団についていろいろな推測を行えるようになる。この機能に焦点をあてた統計学の分野は、特に推計学 *inferential statistics* と呼ばれる。

例えば、2つの標本の平均値を比較する場合、平均値の差が偶然のものか、それとも、偶然では起こりにくい、意味のある差（統計では“**有意差**”と呼ぶ）であるかどうかを判定したいときに、統計学の検定という機能を利用する。

具体的には、2つの標本の平均値には“**差がない**”という「**帰無仮説**」（どちらも同じ母集団から得た標本で、平均値のずれは、偶然の範囲内であるとする仮説）をまず設定し、それが確率的にどの程度起こりうるかを計算する。もしその差が確率的に起こりにくい場合には、“**差がある**”とする逆の仮説「**対立仮説**」を採用する。一般にこの証明方法を、**統計学的仮説検定** *statistical hypothesis testing*（一般的には、**有意差検定**）と呼ぶ。そして、対立仮説が採用されたとき、2つの標本の平均値には、“**統計的に有意差がある**”と判断する。

有意差検定の考え方



- 1) 帰無仮説が正しいとして、観察された平均値の差が生じる有意確率 P を求める
- 2) P の値からどちらの仮説がより妥当かを判定



有意差検定は、観察された差が偶然のずれか、偶然を超えたずれであるかを客観的に判定する手段

機能2：データを要約する（記述統計学）

統計学のもう1つの機能は、データ（母集団）の把握に役立つ**要約値**を求めることにある。すなわち、母集団から標本を取り出せば、そのデータの要約値として平均値、標準偏差、相関係数、回帰直線などの数値が計算され、それらの数値から母集団の状態を推定できる。一般にデータを要約した数値のことを、**統計量 (statistic)** と呼ぶ。

この機能は、データが単純な**種類**の情報からなっている場合だけでなく、**多種類**の情報が複雑に絡み合っている場合にも適用でき、それぞれに相応しい要約値を求める多様な機能が用意されている。後者は、特に**多変量解析**と呼ばれ、複雑で大規模な情報を的確に分類・整理でき、情報の骨格を把握できるようになる。

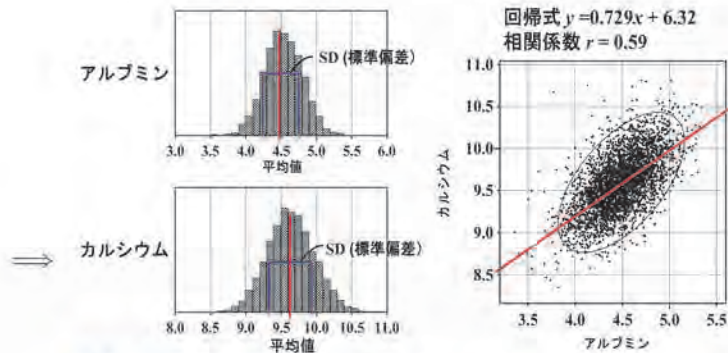
このような要約値を求めるという機能に焦点をあてた統計学の分野は、特に**記述統計学 descriptive statistics** と呼ばれる。しかし、推計学の中でも記述統計学による要約値が常に使われ、両者は密接に関連しており、決して独立したものではない。また、統計学そのものが大きく発展し、多様化している現在、統計学を推計学と記述統計学に区別して考える意味は乏しい。

記述統計学の機能を具体的に述べると、例えば、下記の健常者から得られた、血中のアルブミンとカルシウムの濃度の数値を眺めただけでは、何も見えない。

そこで、記述統計学の機能を使って、それぞれの測定値の**度数分布図**を作ってみると、各測定値がどのように広がっている（分布している）のかが見えてくる。また、2つの測定値の間にある関係も、**相関図**を描いてみると、数値表の上からは全く見えていなかった2変量間の密な関連が、明瞭に浮かび上がってくる。

ID	Alb	Ca
1	4.1	10.8
2	5.0	10.8
3	4.7	10.8
4	4.8	10.8
5	3.8	10.7
6	4.9	10.7
7	5.1	10.7
8	4.6	10.7
9	5.0	10.6
10	4.9	10.6
11	4.3	10.6
12	4.7	10.6
...

数値を眺めても何も見えてこないが...



度数分布図だと数値の広がりが分かる。記述統計では、その数値を要約した値（平均値や標準偏差）が求まるので、その傾向を把握しやすくなる。

相関図を描いてみてはじめて、アルブミンとカルシウムの関連の強さが分かる。

この例では、測定値別に見たときの統計量（要約値）として、平均値や標準偏差を求めている。また2つの測定値を同時に調べたときには、回帰直線の傾き、切片、相関係数などを統計量として求めている。



人は数値に弱いが、作図と統計量によりデータの特徴をつかめる

Key word

キーワード

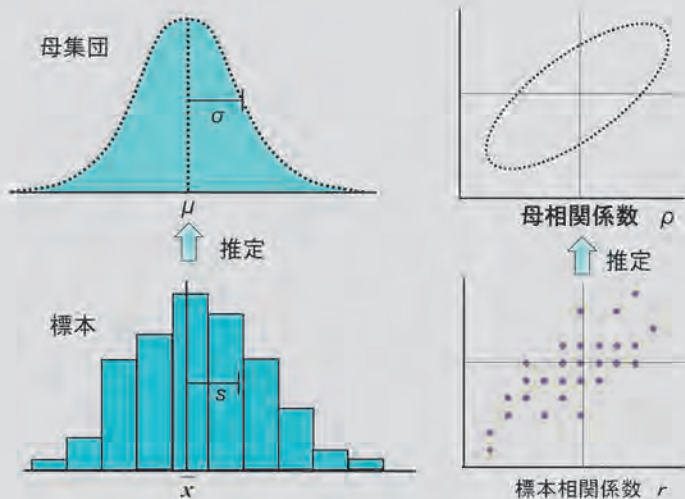
統計量とは

～ 標本から算出された要約値 ～

統計量とは標本の中の数値を要約した値で、一種類の計測値に対する平均値 (\bar{x}) や、標準偏差 (s) などがそれに当たる。また、計測値が二種類ある場合は、相関係数 (r) や回帰直線を求めたときの傾きや切片といった数値も統計量である。

統計量という用語の英語訳は、**statistic** であるが、通常複数形の **statistics** で用いる。一方、統計学という学問の英語訳も **statistics** である。もともと、統計学は、集団の状態を特徴づける複数の数値（統計量）を調べ、解釈するのが目的であったので、統計量と統計学の単語が一致しているのもうなずける。

一般に統計量は、標本についての要約値であり、それを母集団全体について求めた数値を**母数**と呼ぶ。例えば、母集団の平均値は**母平均** (μ)、標準偏差は**母標準偏差** (σ) と呼ばれるが、それが母数に相当する。母数と区別するため、標本から求めた平均値、標準偏差などの統計量を**標本平均** (\bar{x})、**標本標準偏差** (s) などと呼び、母数のそれと区別して扱う。



統計量には、このように母数の推定を目的としたものと、それとは別に「検定統計量」として、有意差検定のための1つの差の代表値として求めるものがある。

正規分布とは

生物現象など、自然界で観察される多くの計測値は、何であれ、平均値に近いほどその出現率が高く、平均値からその両側に値が遠ざかるに従って出現頻度は少なくなる。このうち、同じものを何度も繰り返し計測し、平均値からのずれ（誤差）の大きさを求め、その出現度数を描いてみると、平均値を中心とした左右対称の釣り鐘状の分布型となることが多い。1812年に数学者ガウスは、純粋な条件で繰り返し計測した時に、一貫して現れる計測値の分布形状を数理的に導き、正規分布 **normal distribution** と名付けた。発見者の名前をとって、ガウス分布 **Gaussian distribution** とも呼ばれる。

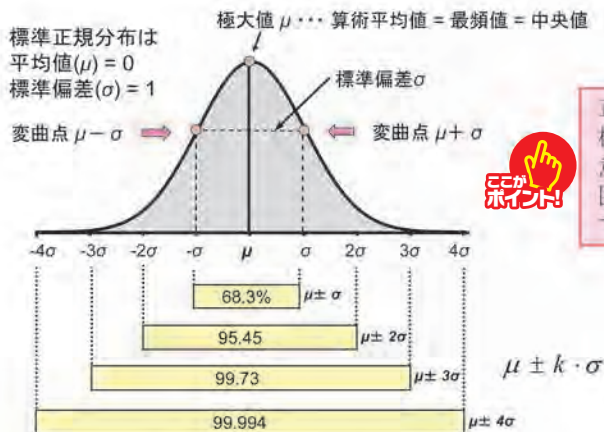
正規分布と信頼区間確率

正規分布は理論上次の3つの特性を持つ。

- (1) 分布曲線の極大値は、平均値 μ であり、 μ を中心に左右対称である。
- (2) 分布曲線には2つの変曲点が存在し、中心 μ から $\pm\sigma$ 偏位した位置にある。
- (3) その平均 μ と標準偏差 σ が分かれば、任意の値 k を指定して、 $\mu \pm k\sigma$ の範囲内に入るデータの割合を的確に推定できる。

(3) については、例えば $k = 1.0$ すなわち $\mu \pm 1\sigma$ 内に 68.3% のデータが含まれる。同様に、 $k = 2.0, 3.0$ では、 $\mu \pm 2\sigma$ ($\bar{x} \pm 2s$) に 95.45%、 $\mu \pm 3\sigma$ ($\bar{x} \pm 3s$) に 99.73% のデータが含まれる。

一方統計でよく使われる中央95%のデータが含まれる区間は、 $\mu \pm 1.96\sigma$ ($\bar{x} \pm 1.96s$) として求まる。なお、この標準偏差の倍率 k を本書では“**信頼区間限界指数**”と呼ぶ。後述の標準正規分布で k は、 z スコアに相当する。

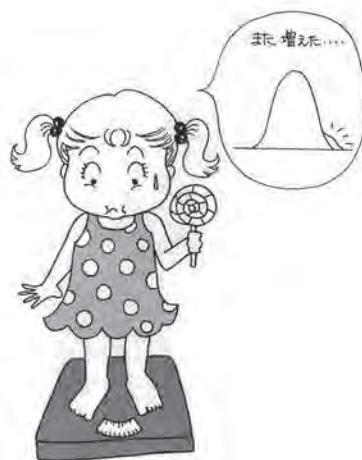


正規分布の場合、平均値 μ 、標準偏差 σ が分かれば、任意の k に対して $\mu \pm k\sigma$ の範囲内に入るデータの割合を推定できる

正規分布の確率密度関数は次のように表される： $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

正規分布となる計測値の例

1. 毎朝飲んでいる牛乳1パックの重量を、精密な秤（はかり）で、測って記録したときの重量の分布
2. 同じ患者さんの血清試料を、同じ日に何度も繰り返し測定したときの測定値の分布
3. 毎朝同じ条件で、何日も繰り返して精密に測った体重の分布



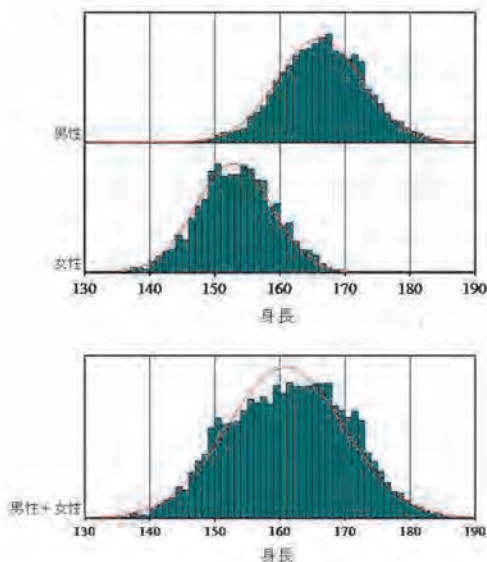
このように、同じものを、同じ条件で、繰り返して 測った場合、その計測値は正規分布となる。

正規分布とならない計測値の例

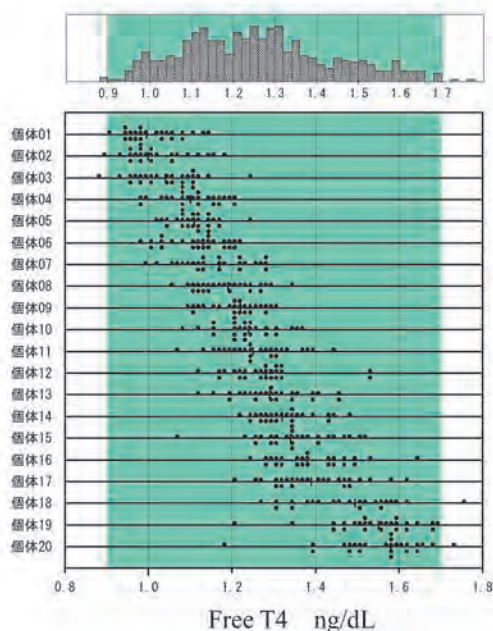
異質な集団の計測値が組み合わさった分布は正規分布とならないことが多い。例えば、右下図のように男性の身長も女性の身長も、それぞれほぼ正規分布とみなせるが、男女を合わせた混成分布を作ると、正規分布ではなくなる。

一方、35頁で示した臨床検査の測定値のほとんどは、厳密な意味では正規分布ではない。しかし、個人個人の計測値の分布を見ると、ほぼ正規分布を示していることが多い。以下に示す図は、健常成人20例を対象に、1年間にわたり毎月2回、2つの血中ホルモンFree T4（遊離サイロキシン）とTSH（甲状腺刺激ホルモン）を繰り返し測定したときの、検査値の分布を個人別に分類して示している。

Free T4の個体内の分布は、いずれもほぼ正規分布と考えられるが、値が大きいほど分布幅がやや広く、上段に示す全体を合成した分布はやや右裾広がりとなっている。



TSH の分布は、個体内の分布をみるとやはりほぼ正規分布に近似しているが、値が大きくなるほど分布幅が広がり、全体を合成した分布は、明らかに右裾広がり強く、歪んだ分布（対数正規分布）となっている。

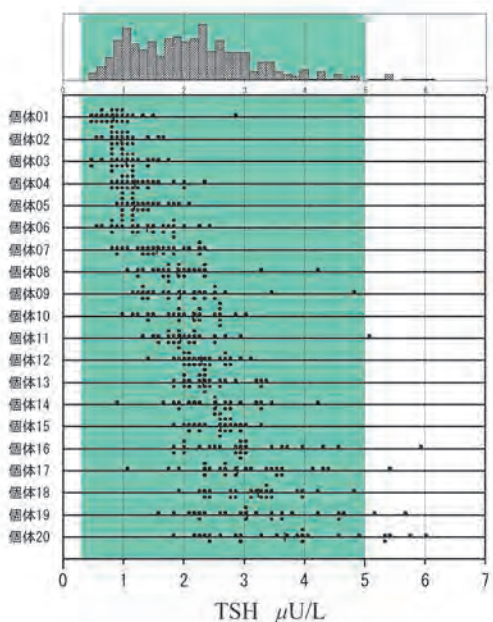


集団の分布

正規分布よりやや右裾広がり分布

個別別分布

個別別の平均値は大きく異なる(個体差大)がどの個体の値もほぼ正規分布とみなせ標準偏差も同程度



集団の分布

高値側への値の広がりが大きく、**対数正規分布**とみなせる分布

個別別分布

個別別の平均値は大きく異なる(個体差大)がどの個体の値もほぼ正規分布と考えられる。ただし、平均値が大きいほど、標準偏差が大きい。

この事例に示されるように、一般に繰り返し測定された血中物質の個体内変動は、正規分布を示すことが多い。一方、測定値の散布度は個体によって異なり、値が大きい個体ほど散布度が大きくなる傾向が見られる。このため、各個体の分布を合成した集団の分布は、右裾広りの形をとる検査項目が多い。

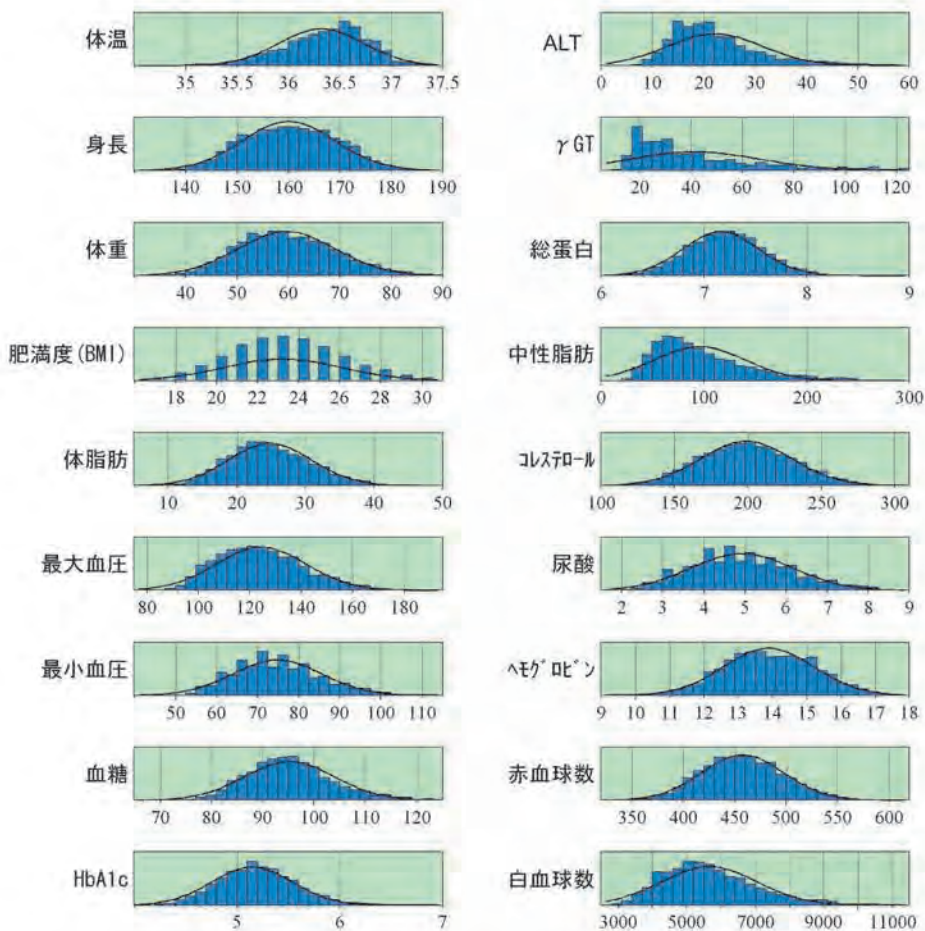
よく見られる分布型

計測する項目によって、いろいろな分布の形がある。下図では、健診時に健常男性から得られた主な身体計測値と血液検査値の分布を示す。

臨床検査のデータでは、正規分布型となるものもあるが、値の大きい方に広がる傾向の分布（右裾広がり分布）を示すものが多い。

健常男性の身体計測値と血液検査データ

1. 正規分布型→身長、肥満度(BMI)、最小血圧、HbA1c、総蛋白、コレステロール、尿酸、ヘモグロビン、赤血球数
2. 軽度の左裾広がり型→体温
3. 軽度の右裾広がり型→体重、体脂肪、最大血圧
4. 強い右裾広がり型→ALT、 γ GT、中性脂肪、白血球数



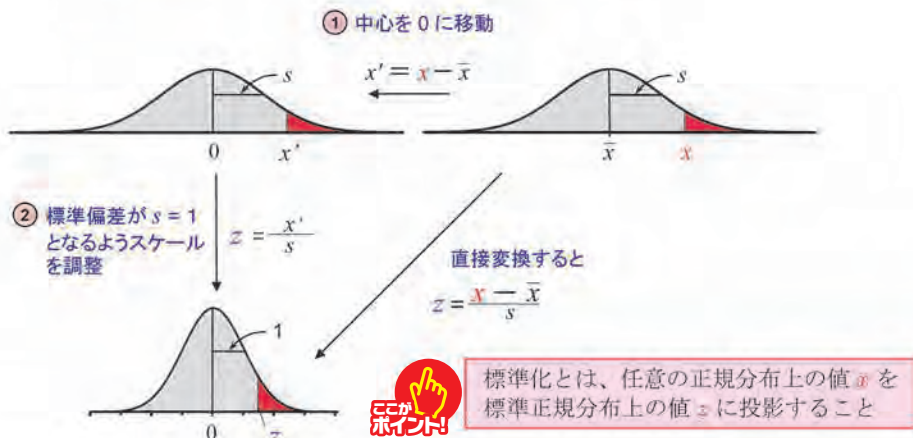
曲線は平均値、標準偏差から描いた理論正規分布曲線。青色の度数分布図との合致度で正規分布からの偏りを評価できる。

正規分布の標準化とzスコア

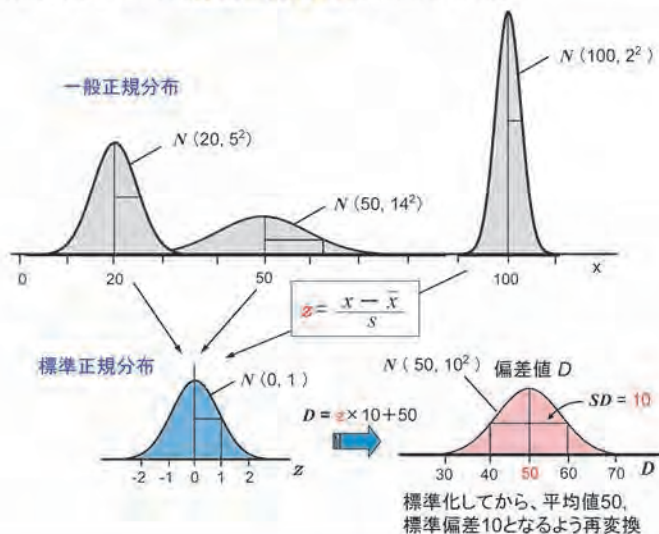
正規分布の標準化とは、平均値が0.0、標準偏差が1.0となるように、値を変換することである。これには、標本から得た標本平均と標本標準偏差を使って次式により変換する。

$$z = \frac{x - \text{標本平均 } \bar{x}}{\text{標本標準偏差 } s}$$

この標準化操作を **z変換** と呼び、変換後の値を **zスコア** という。zには単位が存在しない。



正規分布は、その形（平均値、標準偏差）は異なるが、その値をz変換することにより、平均値0、標準偏差1である**標準正規分布**に変換できる。



どんな正規分布も標準化（z変換）により標準正規分布に変換できる。さらに、zを平均50、標準偏差10となるよう再変換すると、成績評価で用いられる偏差値となる。

標準正規分布表の見方

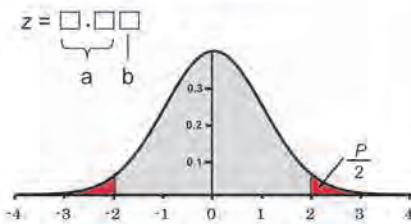
標準正規分布表には、任意の z スコアに対する、両側確率 ($|z|$ の値がそれよりも大きくなる合計の確率) が示されている。任意の正規分布上の値 x に対する両側確率を求めるには、まず平均値と標準偏差を使って、 z スコアに変換しておく。

表の見方は、まず z スコアの小数第1位までの値から縦軸の位置を決め、次に小数第2位の値から横軸の値を選ぶ。そして、その交点の値を読むと、それがその z に対する両側確率になっている。

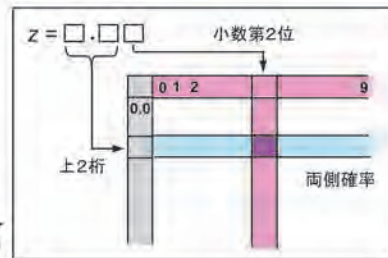
例えば、 $z = 1.96$ の場合、交点の値が 0.0500 となっているので、その z スコアの両側の赤い部分の合計面積 (両側確率) が全体の 5% であることが分かる。

逆に、両側確率が 0.2000 となる z スコアを求めるには、表から、 $z = 1.28$ に対する両側確率が、0.2006 で、 $z = 1.29$ に対して 0.1970 であるので、 z 値は両者の間にあることが分かる。

標準正規分布表 (両側確率)



正規分布表の見方



z	(b) .00	.01	.02	.03	.04	.05	.06	.07	.08	.09
(a) 0.0	1.0000	.9920	.9840	.9760	.9680	.9602	.9522	.9442	.9362	.9282
0.1	.9204	.9124	.9044	.8966	.8886	.8808	.8728	.8650	.8572	.8494
0.2	.8414	.8336	.8258	.8180	.8104	.8026	.7948	.7872	.7794	.7718
0.3	.7642	.7566	.7490	.7414	.7338	.7264	.7188	.7114	.7040	.6966
0.4	.6892	.6818	.6744	.6672	.6600	.6528	.6456	.6384	.6312	.6242
0.5	.6170	.6100	.6030	.5962	.5892	.5824	.5754	.5686	.5620	.5552
0.6	.5486	.5418	.5352	.5286	.5222	.5156	.5092	.5028	.4964	.4902
0.7	.4840	.4776	.4716	.4654	.4592	.4532	.4472	.4412	.4354	.4296
0.8	.4238	.4180	.4122	.4066	.4010	.3954	.3898	.3842	.3788	.3734
0.9	.3682	.3628	.3576	.3524	.3472	.3422	.3370	.3320	.3270	.3222
1.0	.3174	.3124	.3078	.3030	.2984	.2938	.2892	.2846	.2802	.2758
1.1	.2714	.2670	.2628	.2584	.2542	.2502	.2460	.2420	.2380	.2340
1.2	.2301	.2262	.2224	.2186	.2150	.2112	.2076	.2040	.2006	.1970
1.3	.1936	.1902	.1868	.1836	.1802	.1770	.1738	.1706	.1676	.1646
1.4	.1616	.1586	.1556	.1528	.1498	.1470	.1442	.1416	.1388	.1362
1.5	.1336	.1310	.1286	.1260	.1236	.1212	.1188	.1164	.1142	.1118
1.6	.1096	.1074	.1052	.1030	.1010	.0990	.0970	.0950	.0930	.0910
1.7	.0892	.0872	.0854	.0836	.0818	.0802	.0784	.0768	.0750	.0734
1.8	.0718	.0702	.0688	.0672	.0658	.0644	.0628	.0614	.0602	.0588
1.9	.0574	.0562	.0548	.0536	.0524	.0512	.0500	.0488	.0478	.0466
...
...

正規分布表は、270 頁参照

StatFlex では、任意の z 値に対して両側や片側の確率を計算ができる。

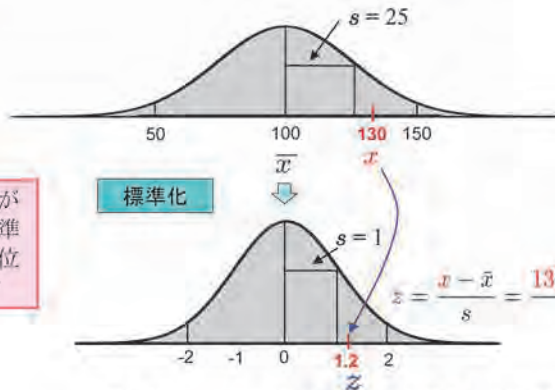
zスコアによる、分布の中での相対位置の表し方

前述のように、任意の値 x を、標本平均 \bar{x} と標本標準偏差 s を使って、次式により標準化した値を zスコアと呼ぶ。もとの分布が、正規分布のときには、z は、平均値 0.0、標準偏差 1.0 の標準正規分布の形に標準化されたことになるので、zスコアがわかれば、 x がどの程度偏った位置にあるかを、標準正規分布表から両側確率の形で求めることができる。

$$z \text{ スコア} = \frac{x - \bar{x}}{s}$$



z スコアは、 x が分布の中心から標準偏差 s の何倍分偏位しているかを表す



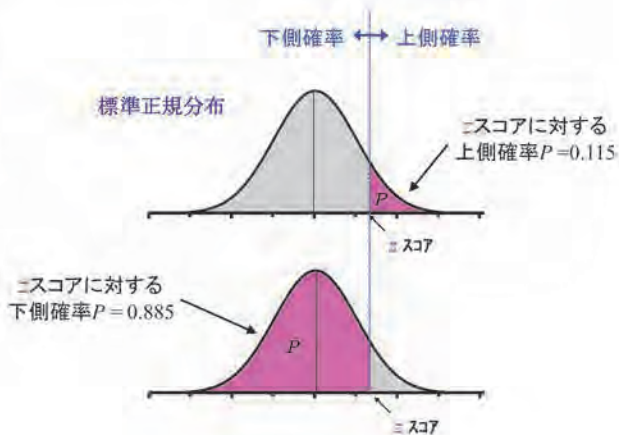
$$z = \frac{x - \bar{x}}{s} = \frac{130 - 100}{25} = 1.20$$

平均値 $\bar{x}=100$ 、標準偏差 $s=25$ のとき 130 は zスコアにすると 1.2 となる。

片側有意確率 P による分布中の相対位置の表し方

zスコアから、その上側（または下側）の確率 P が標準正規分布表から求まる。

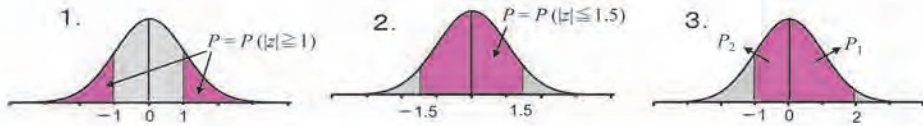
値の小さい方から順に累積した確率を下側確率、大きい方から累積したものを上側確率と呼ぶ。 P が小さいほど x は極端な値と判定される。



例題
3

標準正規分布表から、次の5つの値を求めよ。

- $|z| \geq 1.0$ となる確率 $P = 0.3174$
- $|z| \leq 1.5$ となる確率 $P = 1 - P(|z| \geq 1.5) = 1 - 0.1336 = 0.8664$
- $-1 \leq z \leq 2$ となる確率 $P = \frac{1 - P_1(|z| \geq 2.0)}{2} + \frac{1 - P_2(|z| \geq 1.0)}{2} = 0.819$
- 両側（両裾）の確率 P が $0.001(0.1\%)$ となる $|z|$ 値 $= 3.291$
- 内側の確率 P が $0.50(50\%)$ となる $|z|$ 値 $= 0.674$


例題
4

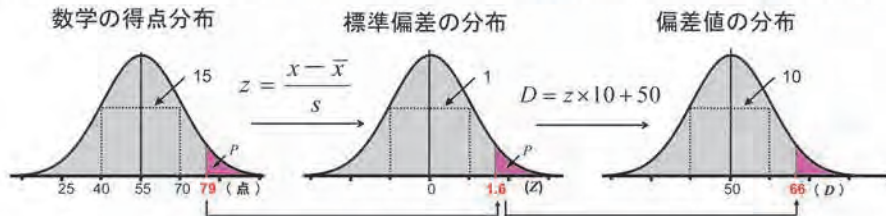
高校生400人の数学の成績が、平均値55点、標準偏差15の正規分布に従う。点数が79点の学生について次の問に答えよ。

- z スコアを求めよ。
- 偏差値 (D) になおせ。
- 79点以上は何人いると推測されるか。

- 標準化値 z は下記の式で求まる。

$$z = \frac{\text{点数} - \text{標本平均}}{\text{標本標準偏差}} = \frac{79 - 55}{15} = 1.6$$

- この成績を偏差値で表すと、 $z \times 10 + 50 = 1.6 \times 10 + 50 = 66.0$ となる。



- 標準正規分布表から、 z 値が 1.6 以上 (79 点以上) の生徒の割合 (確率) は、

$$P = \frac{0.1096}{2} = 0.0548 \text{ と予測される。}$$

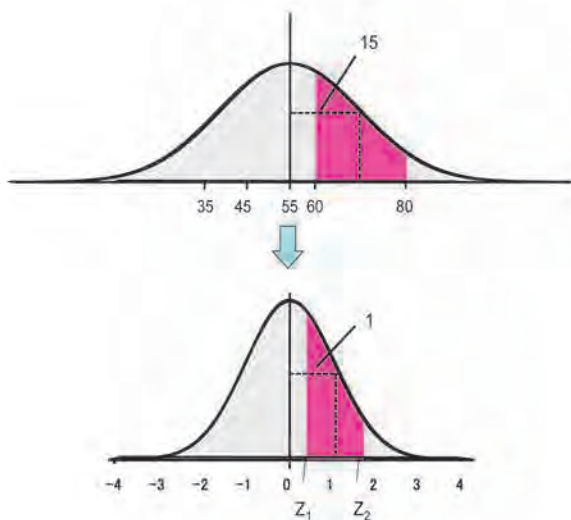
よって、400人中79点以上は、 $400 \times 0.0548 = 21.92$ で、約22人である。
 (標準正規分布表の値は、両側の確率であることに注意!)



演習:2

例題4の数学の成績分布において、数学の得点が60~80点の間に何人の生徒がいるかを予測せよ。(解答243頁)

1. 60点の標準化値 $z_1 =$
2. 80点の標準化値 $z_2 =$
3. $z_1 \leq z$ となる片側確率 $P_1 =$
4. $z_2 \leq z$ となる片側確率 $P_2 =$
5. 求める確率は $P = P_1 - P_2 =$
6. 60~80点に含まれる生徒数は、 P に生徒数をかけて 人として求まる。
7. 得点60~80点を、偏差値の形で表すと ~ となる



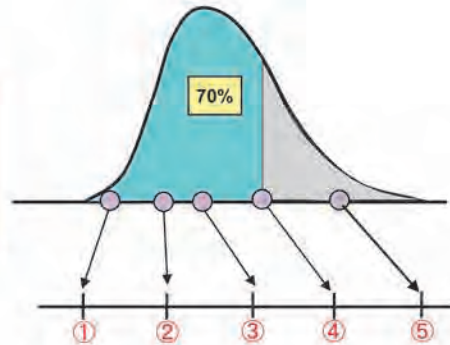
参考 パーセンタイル(百分位数)による相対位置の表し方

データを大きさの順に並べ、個々の点が下から何番目の位置にあるかを調べ、その相対位置を百分率で表した数値を**パーセンタイル (percentile)** または百分位数と呼び、本書では%点とも記す。パーセンタイルは、前項のように分布上の特定の点の極端さを片側確率で表したものと全く同じで、それを100倍した値である。

パーセンタイルの算出法には主に**4つの方法**がある。いま n 個 (5個) よりなる標本のデータを大きさの順に並べたとき、第 i 番目の点のパーセンタイルは次のように求める。



パーセンタイルには複数の計算方法がある。
パーセンタイルの表記として対称補正法2による割り付けのバランスが最も良い。



• 単純なパーセンタイル	$P = \frac{i}{n} \times 100$	
• 対称補正法 1	$P = \frac{i-1}{n-1} \times 100$	
• 対称補正法 2	$P = \frac{i-0.5}{n} \times 100$	
• 対称補正法 3	$P = \frac{i}{n+1} \times 100$	

単純なパーセンタイルの問題点は、例えば、 $n = 5$ で1番目の値は20パーセンタイルであるが、5番目は100パーセンタイルとなり、全体が右側にシフトし、左右非対称となっている点である。それを補正するため、パーセンタイルの対称補正が行われる。

対称補正法1では両端にそれ以上のデータが存在しないことを仮定したことになる。逆に対称補正法3では、まだ両端にデータがあることを仮定した形になっており、全体としてデータが中央に集まった形となっている。これに対して対称補正法2では、データが他の2法の間中間的な位置に配置され、最も相応しいと考えられる。

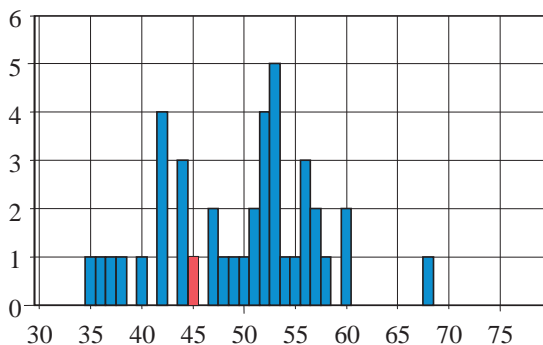
なお、Microsoft Excel では対称補正法1を採用しているが、StatFlex ではバランスのとれた対称補正法2を採用している。

別の例として、次の例題で示した図で考えると、 $n = 40$ の場合、①単純なパーセントイル式で計算すると、1番は2.5%、40番は100%になり、中心が100%の方向へずれ、対称ではない。しかし③対象補正法2式で計算すると、1番は1.25%、40番は98.8%になり、対称性のある百分位が求まる。

例題
5

$n=40$ の次のデータの分布について、 $x = 45$ (13番目)の値が何パーセントイルに相当するかを、4つの計算法で求め比較せよ。

35	36	37	38	40	42	42	42	42	44
44	44	45	47	47	48	49	50	51	51
52	52	52	52	53	53	53	53	53	54
55	56	56	56	57	57	58	60	60	68



① 単純なパーセントイル:

$$p = \frac{i}{n} \times 100 = \frac{13}{40} \times 100 = \mathbf{32.5}$$

② 対称補正法1によるパーセントイル:

$$p = \frac{i - 1.0}{n - 1} \times 100 = \frac{12.0}{39} \times 100 = \mathbf{30.8}$$

③ 対称補正法2によるパーセントイル:

$$p = \frac{i - 0.5}{n} \times 100 = \frac{12.5}{40} \times 100 = \mathbf{31.3}$$

④ 対称補正法3によるパーセントイル:

$$p = \frac{i}{n + 1} \times 100 = \frac{13}{41} \times 100 = \mathbf{31.7}$$

n 数が大きいと各パーセントイルに大きな違いは認められないが、 n 数が小さい場合には大きな違いを生じるので注意が必要である。

探究 p パーセンタイルの値 x の求め方

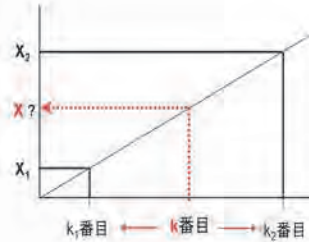
パーセンタイルの数値 p を指定して、それに対応する測定値 x を換算する場合、通常完全に一致した数値は無いことが多く、次の補間法を利用して x を推定する。

ここで対照補正法として、 p パーセンタイルの値 x は、まずそれに最も近い順位

$$k = p \times \frac{n}{100} + 0.5 \text{ を求める。}$$

いま、 k が整数で求まっていれば、 k 番目の値が求める p パーセンタイルになる。 k が整数でないときは、 k より小さく、それに最も近い整数を k_1 とし、 k_1 番目の測定値 x_1 を p_1 パーセンタイルとする。また、 $k_1 + 1$ (k_2) 番目の測定値 x_2 を p_2 パーセンタイルとして、次式から補間法で p パーセンタイルの値 x を求める。

最も近い下の整数 k_1 番目 $\rightarrow x_1 = 49$
 \uparrow
 p パーセンタイル k 番目 $\rightarrow x$
 \downarrow
 最も近い上の整数 k_2 番目 $\rightarrow x_2 = 50$



よって求める値 x は、次の式から計算できる。

$$x = x_1 + \frac{k - k_1}{k_2 - k_1} \times (x_2 - x_1)$$

例えば、例題 5(42 頁) の $n=40$ のデータの分布について、42 パーセンタイルの値を求める場合は以下ようになる。

$$k = p \times \frac{n}{100} + 0.5 = 42 \times \frac{40}{100} + 0.5 = 17.3$$

$k_1=17$ 番目の測定値 x_1 と 18 番目の測定値 x_2 から次の補間法で p パーセンタイルの値 x を求める。

最も近い下の整数 17 番目 $\rightarrow x_1 = 49$
 \uparrow
 p パーセンタイル $k = 17.3$ 番目 $\rightarrow x$
 \downarrow
 最も近い上の整数 18 番目 $\rightarrow x_2 = 50$

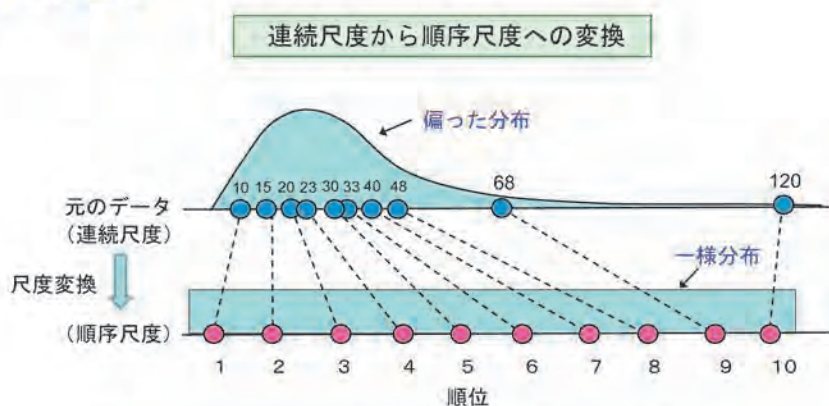
$$x = x_1 + \frac{k - k_1}{k_2 - k_1} \times (x_2 - x_1) = 49 + \frac{17.3 - 17}{18 - 17} \times (50 - 49) = 49.3$$

参考 計測尺度と統計処理方式


● 計測尺度

統計処理を行う場合、分布の形状に依存する統計量（平均値や標準偏差などのパラメータ）を用いる方法を**パラメトリック法**と呼ぶ。一方、分布の形状に依存しない統計量（順位、中央値、パーセンタイルなど）を用いる統計手法を**ノンパラトリック法**と呼ぶ。パラメトリック法による検定（**パラメトリック検定**）を用いる場合には、標本の分布型が重要となる。通常は、正規分布を仮定して検定法が組み立てられているため、正規分布から大きく偏ったデータを扱う場合には、正しい判断ができない（有意確率が歪む）ことになる。実際には、データは正規分布に従わないケースが多いので、パラメトリック法を用いる場合には、先に、データの**正規分布への変換**が必要になることがある。これに対し、**ノンパラメトリック検定**は、順序尺度や名義尺度の形でデータを取り扱うため、標本の分布型を考慮する必要がない。また、データが間隔尺度（連続量）で計測されている場合でも、それを順序尺度に変換して用いるので、分布型に依存せず汎用的な検定法と言える（検定法の使い分け 6 頁, 111 頁を参照）。

● 尺度変換



どんな分布も、順序尺度に変換すると一様分布となり、分布の形状や極端値の影響を受けずに統計処理できる



第3章



検定の原理

統計的仮説検定の目的と理論

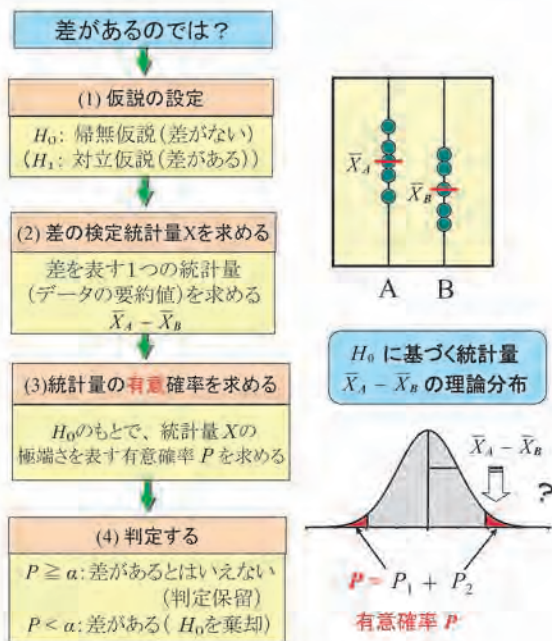
「差がある」という仮説を検定する場合、差の程度が不明のため、そのままでは検定できない。そこで、その逆の「差がない」という仮説を検定し、それに何らかの矛盾が見つければ、もとの「差がある」という仮説を採用する。逆に、明らかな矛盾がないときは判定を保留する、という論法で検定を行う。

ここで、「差がない」という仮説は本来無に帰すべきものとして、「帰無仮説」 null hypothesis と呼び、 H_0 と略す。また、もとの「差がある」という仮説は、「対立仮説」 alternative hypothesis とよび、 H_1 と略す。

下図に、「2群に差がある」という仮説検定のフローを示す。

- (1) **仮説の設定**：まず「2群に差がない」($A = B$)という仮定(帰無仮説 H_0)をおき、本来の「差がある」($A \neq B$)という仮説(対立仮説 H_1)は、いったん伏せておく。
- (2) **検定統計量の計算**：検定しやすいように、データを一つの数値に要約する。この要約値を**検定統計量 test statistic**と呼ぶ。例えば、この例では2群の平均値 \bar{X}_A 、 \bar{X}_B を求め、その差 $\bar{X}_A - \bar{X}_B$ を検定統計量 X とする。

- (3) **検定統計量の有意確率を求める**：帰無仮説が正しい場合に検定統計量 X が生じる**有意確率 P** を求める。この例では、帰無仮説から、 X (平均値の差)の期待値は0であるが、実際にはある大きさをもっている。そこで確率的に、その差が十分起こりうるかを調べる。これには、統計数理に基づいた検定統計量の理論分布を利用する。



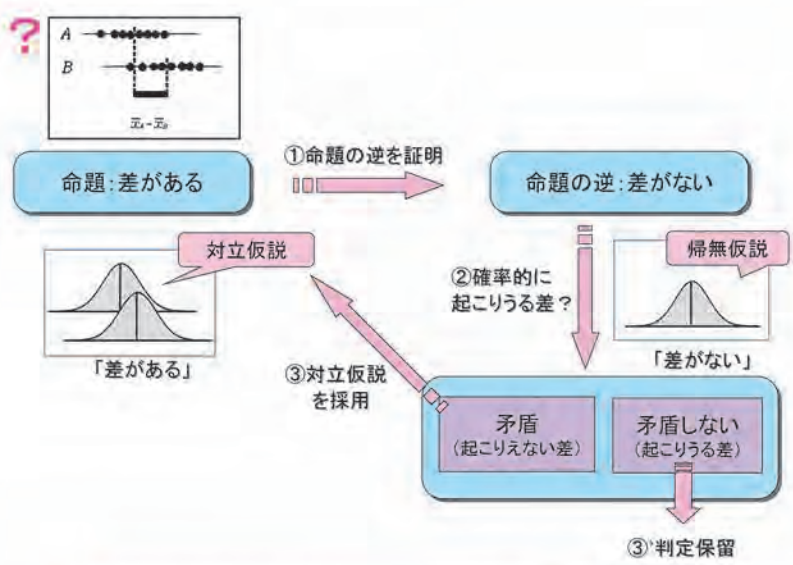
- (4) 判定：検定統計量 X の極端さを示す有意確率 P が、有意水準 α よりも大きいとき、その程度の差は確率的に十分起こりうるので、帰無仮説を棄却できない（判定保留）。逆に、 P が有意水準 α より小さいときは、「差がない」という帰無仮説 H_0 の方がおかしいと判断する。よって H_0 を棄却し、「差がある」とする対立仮説 H_1 を採用する。ここで、有意水準 α には通常 0.05(5%) または 0.01(1%) が使われることが多い。

有意差検定では反証の論理が使われる

上記のフローで示したように有意差検定では、反証の論理（背理法）を使用する。すなわち「差がある」という仮説は証明が困難なので、その逆の「差がない」という仮説（帰無仮説 H_0 ）を証明する形が取られる。

そして、観察した差が H_0 のもとで、ある程度以上の確率で起こる場合には H_0 を棄却できず「判定保留」とする。一方、観察された差が H_0 の下では、確率的に稀な場合には H_0 を棄て、対立仮説 H_1 すなわち「差がある」という仮説を採択する。

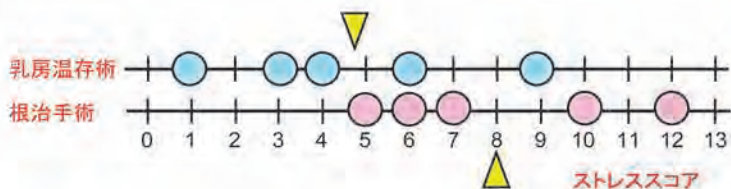
有意差検定は反証の論理



有意差検定では、「差がない」という仮説 H_0 （帰無仮説）を否定することで、本来言いたい「差がある」という仮説 H_1 （対立仮説）を立証する

有意差検定の利用例 1

乳癌患者に対して、乳房温存術と根治手術を施行し、退院時に患者のストレス度を13段階で比較し、下図のような結果を得た。治療法で、術後の生活の質(QOL)は異なると言えるか？



有意差検定では、平均値のずれに対して、差がないと仮定したときに、このずれがどの程度起こりうるのかを、**有意確率 P** という形で求める（キーワード参照）。

そして、その有意確率 P がある**有意水準 α** より大きければ、そのずれは偶然的範囲内のものと考え、逆に有意確率 P が有意水準より小さければ、偶然ではないと判断する。一般に有意水準 α は0.05(5%)や0.01(1%)が用いられるので、上の例では、有意確率は統計理論から、 $P = 0.109$ と計算され、偶然起こりうる範囲内の差であると判定される。

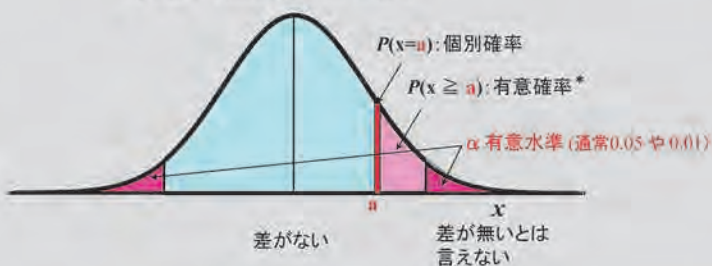
Key word キーワード

有意確率 P と有意水準 α の意味

統計における有意確率とは、ある現象が生じる**個別確率**ではなく、その現象およびそれ以上の極端な現象の個別確率を累和した確率をさす。有意確率が小さいほど、確率的に起こりにくい現象であることを表す。

一方、**有意水準 α** とは、**有意確率 P** がどの程度小さければ対立仮説を採択するかを決定する基準となる確率である。通常、有意水準 α には、0.05や0.01が用いられる。有意差検定では、標本から求めた検定統計量 a に対する有意確率 P と有意水準 α を比較し判定を下す（131頁参照）。

H_0 に基づく検定統計量 x の分布



標本の極端さは、個別確率ではなく**有意確率**（個別確率の累和）で表す

有意差検定の利用例 2
 肺癌の診断後1年目の生存率は $\frac{1}{6}$ であった。

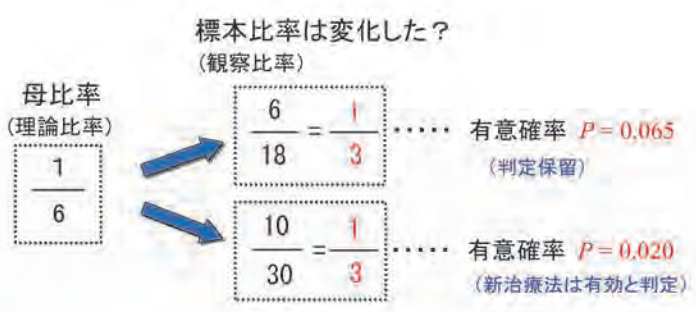
1) 新しい化学療法剤で18名を治療したところ生存例は6例であった。
 この治療は有効と言えるか？

2) さらに症例を増やし、30名治療したところ、生存例は10例であった。
 この場合、治療の有効性はどうか判定されるか？

1) の場合、理論的な有効率が $\frac{1}{6}$ の条件下で、18名の少ない例数ではあるが、有効率が $\frac{6}{18} = \frac{1}{3}$ となっており、それが確率的に十分ありうるかが問題となっている。第6章で述べる出現度数に関する検定を利用して、このような現象が起こる有意確率を求めてみると、0.065 と計算される。

一方、2) のように、例数を増やすと、同じ有効率 $\frac{10}{30} = \frac{1}{3}$ に対する有意確率は0.020となる。

上述のように、一般に有意水準として $\alpha = 0.05$ が用いられるので、1) の例は偶然の範囲内の現象（新治療法は有効とは言えない）と考えて、判定を保留し、2) の例では偶然を超えた現象と考えて新治療法は有効と判定する。観察比率は同じであっても例数が増加することによって研究結果の信頼性が高まり、有意な差があると判定された例である。



観察比率は同じでも、例数によって有意確率が異なる

検定の原理を考えよう(平均値の検定)

この検定は、母集団の平均値と標準偏差がわかっている場合に、観察した標本がそれから得られた標本として矛盾しないかを、平均値の偏りを目安に判定する方法である。

例題
6

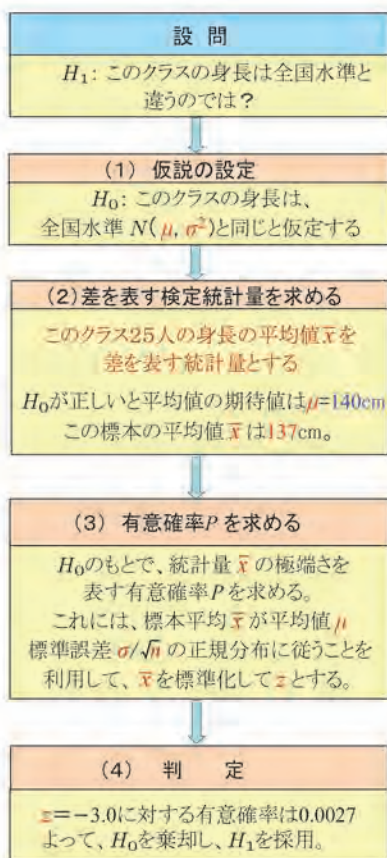
全国10歳女子の身長(μ)と標準偏差(σ)は、それぞれ $\mu=140\text{cm}$ 、 $\sigma=5\text{cm}$ の正規分布をする。いま、ある10歳のクラス25人の身長の分布は下図のようになり、その平均値 \bar{x} は、137cmであった。このクラスの身長は全国水準と違うと言えるか。

次頁の図から、このクラスの身長の分布は、全国水準(母平均 μ) からずれているようにみえる。しかし、客観的にもそうと言えるかを検定する。

- 仮説の設定**：「全国水準と違う」という仮説は、違いの程度を特定できないので論証できない。そこで、その逆の「全国水準と同じ」という仮説(帰無仮説 $H_0: \mu=140$)を採用し、もとの仮説(対立仮説 $H_1: \mu \neq 140$)はいったん伏せておく。ここで、右の解説図中の $N(\mu, \sigma^2)$ 表記法は、平均値 $\mu=140$ 、標準偏差 $\sigma=5$ の正規分布であることを示す。
- 検定統計量を求める**：データを要約して、その偏りを表す検定統計量を求める。この場合25人の身長の平均値 \bar{x} を検定統計量とするのが一般的。 H_0 が正しいとすると、 \bar{x} の期待値は $\mu=140\text{cm}$ であるが、この標本の \bar{x} は **137cm** である。

- 有意確率 P を求める**：この偏り(平均値が3cmずれること)が、どの程度極端な現象かを有意確率の形で求める。これには、 H_0 が正しい場合、標本平均 \bar{x} は平均値 μ 、標準誤差 σ/\sqrt{n} の正規分布に従うことを利用して、まず \bar{x} を z スコアで表す(52頁 標本平均の理論分布と標準誤差(SE)を参照)。

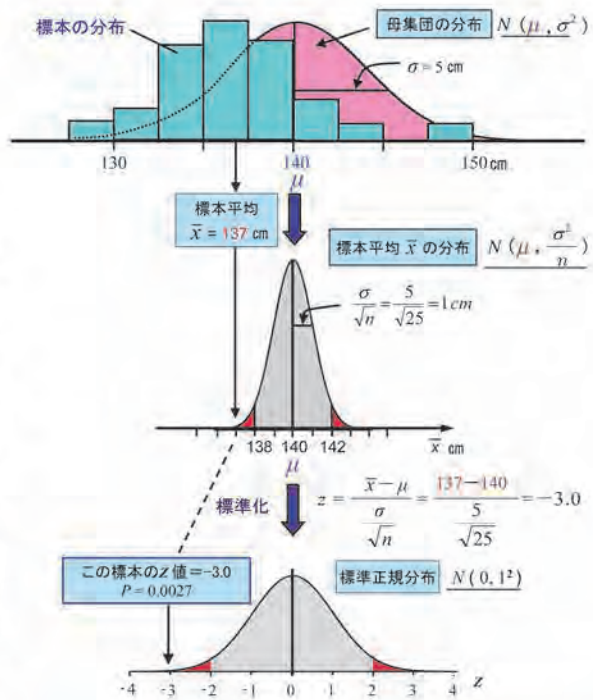
$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{137 - 140}{\frac{5}{\sqrt{25}}} = -3.0$$



次に、 $z = -3.0$ に対する有意確率（両側確率）を標準正規分布表から調べると、 $P = 0.0027$ となる。

- (4) 判定： z の極端さを表す有意確率は、 $P = 0.0027$ と小さい。このような稀な現象が実際に起こったと考えるよりは、 H_0 が正しくなかったと考える方が妥当である。従って、 H_0 を棄却して、対立仮説 H_1 を採用する。

すなわち、クラスの身長は全国平均と比べて、確率 $P = 0.0027$ で、有意水準 $\alpha = 0.05$ より低いと判定する。



標本平均の理論分布に照らして、観察した \bar{x} の極端さを z スコアで表しその有意確率を求める

母集団
(全国10才女子の身長分布)

検定すべき標本

標本の理論抽出
データ数 n

平均値 \bar{x}

標本平均の理論分布
 \bar{x}

この標本の平均値

理論分布のどこにくるか

σ/\sqrt{n}

μ

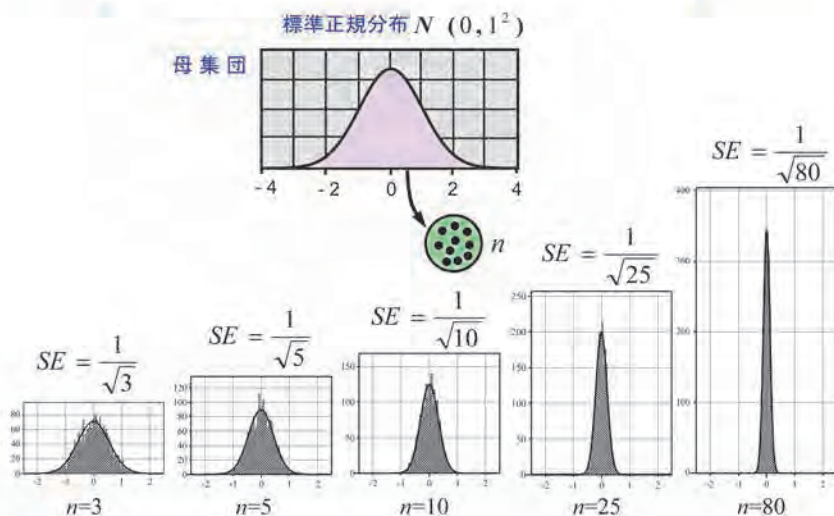
検定の概念：
観察した標本が、既知の母集団 $N(\mu, \sigma^2)$ から得られたものかを標本平均 \bar{x} を目安に判定する。これには標本平均の理論分布に照らして \bar{x} の偏り度を z スコアで表して、有意確率 P を求めて判断する。

標本平均の理論分布と標準誤差 (SE)

正規分布(平均値 $\mu = 0$ 、分散 $\sigma = 1^2$) から、データ数 n の標本を抽出すると、その標本平均の理論分布は正規分布に従い、平均値の期待値は母集団のそれと同じ μ である。しかし、平均値の標準偏差は下図のように、標本のデータ数 n が大きくなるにつれて小さくなり、一般に σ/\sqrt{n} となる(98 頁の正規分布の加法定理による)。

ここで、標本平均 \bar{x} の分布の標準偏差 σ/\sqrt{n} を平均値の標準誤差 (SE: standard error of mean) と呼ぶ。一般に、平均値などの統計量の標準偏差を特に**標準誤差 (SE)** と呼ぶ(54 頁 シミュレーションで考えよう 標本平均の分布を参照)。

標本平均の理論分布に関するシミュレーション



ポイント!

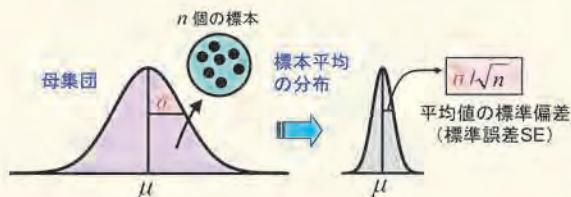
平均値の標準誤差 (SE) は標本データ数 (n) が増えると小さくなる

標本データ数を $n = 3, 5, 10, 25, 80$ として、母平均 0、母標準偏差 1 の正規分布 $N(0, 1^2)$ から 2000 回の繰り返し標本抽出を行い、その標本平均の分布を調べた。標本平均の分布の中心は、母平均と同じ 0 になるが、平均値の標準偏差(標準誤差 SE) は、 n の大きさにより $1/\sqrt{3}, 1/\sqrt{5}, 1/\sqrt{10}, 1/\sqrt{25}, 1/\sqrt{80}$ と小さくなっていく。

キーポイント

標本平均の理論分布

正規母集団から得たデータ数 n の標本の平均値は正規分布となり、その標準誤差 (SE) は、母標準偏差 σ の $1/\sqrt{n}$ 倍となる。



例題
7

40代の男性の血中ヘモグロビン(Hb)濃度は、平均値 $\mu = 14.5$ g/dL、母標準偏差 $\sigma = 1.2$ g/dLの正規分布と仮定する。40代喫煙男性の36名について、Hb濃度を測定したところ平均値は15.0 g/dLであった。

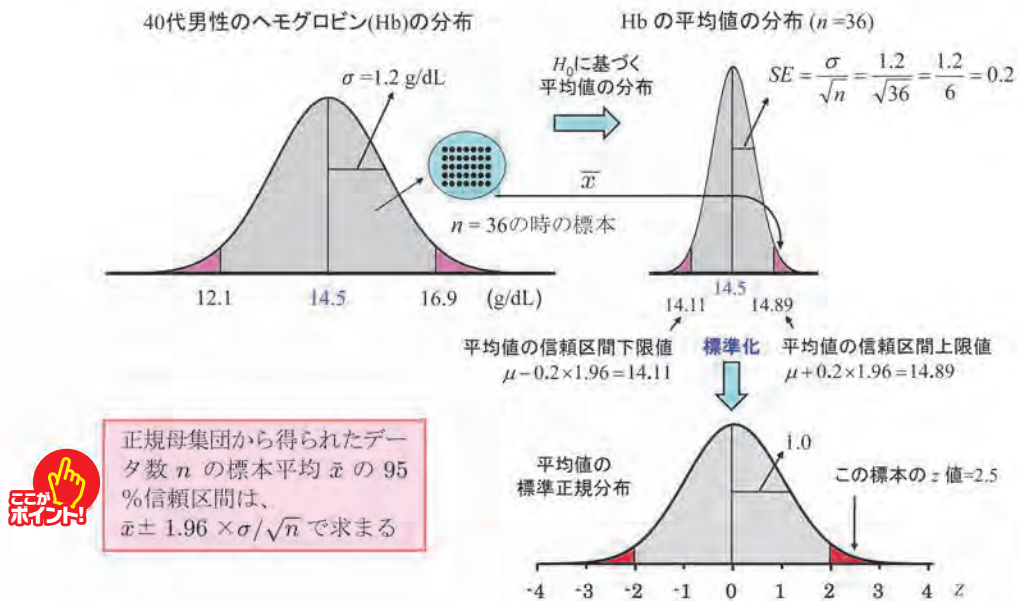
(1) 平均値の偏りは、有意に異なると言えるか？

(2) 有意な偏りがある場合、喫煙群の母平均 μ' の95%信頼区間を求めよ。ただし、標準偏差は元の母集団のそれと同じとして求めよ。

- (1) 36名のヘモグロビン(Hb)濃度は母平均に一致するという仮説 (H_0) が正しい場合、標本平均 \bar{x} は平均値 μ 、標準誤差 σ/\sqrt{n} の正規分布に従うことを利用して、まず \bar{x} を標準化して z スコアで表す。

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{15.0 - 14.5}{\frac{1.2}{\sqrt{36}}} = 2.5$$

次に、 $z = 2.5$ に対する有意確率（両側確率）を標準正規分布表から調べると、 $P = 0.012$ となる。この z の極端さを表す有意確率は、有意水準 0.05 より小さい。従って、 H_0 を棄却し、 H_1 採用する。よって、40代喫煙男性のHb濃度は全国平均と比べて異なると言える。



- (2) 40代喫煙群の母平均 μ' の95%の信頼区間は、母標準偏差 σ には変化がないと仮定すれば次の範囲に推定される。

$\alpha = 0.05$ に対する z スコア z_α は 1.96 であり、

$$\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu' \leq \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$15.0 - 1.96 \times \frac{1.2}{\sqrt{36}} \leq \mu' \leq 15.0 + 1.96 \times \frac{1.2}{\sqrt{36}}$$

$$14.61 \leq \mu' \leq 15.39$$

よって、母平均の95%信頼区間は14.61 g/dL から15.39 g/dL と推定される。

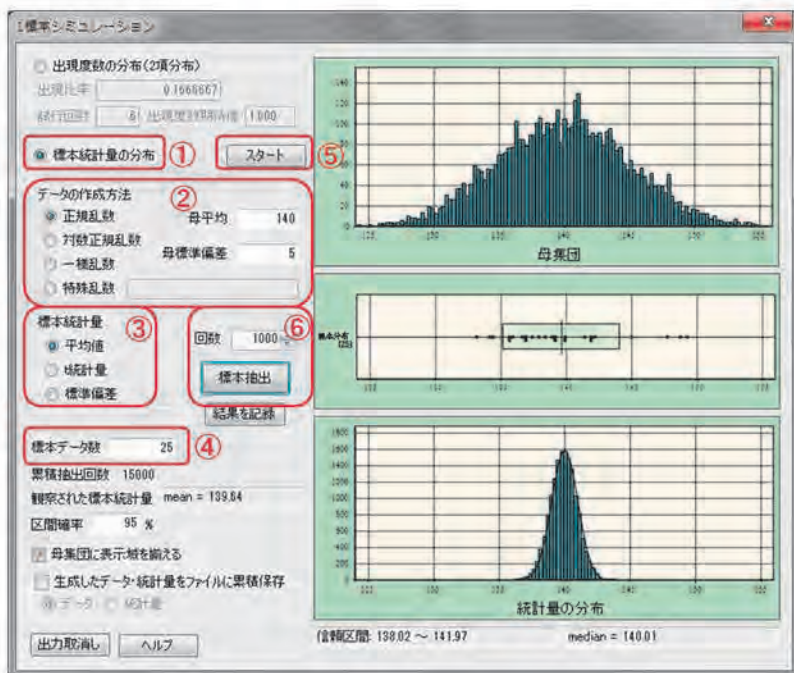


ポイント

シミュレーションで考えよう 標本平均の分布

StatFlex のシミュレーション機能を用いて、母集団から標本を抽出し、その標本平均の分布を調べてみよう。

実行手順



「統計」メニューの「シミュレーション」の「1 標本」を選択する。

1. シミュレーションモード選択①

[標本統計量の分布] を選択する。*1

2. 母集団を指定②

「データの作成方法」で母集団の分布型、母平均、母標準偏差を指定する。

3. 標本統計量を指定③

分布特性を調べるべき標本統計量を選択する。ここでは平均値を選ぶ。

4. 標本データ数を指定④

標本データ数（抽出標本のサイズ）を指定する。

5. 母集団を作成⑤

[スタート] ボタンを押すと、それまでの実験内容がリセットされる。指定した条件で母集団が作成され、最上段にそのグラフが表示される。

*1[出現度数の分布] モードは、比率に関するシミュレーションで利用

6. 標本抽出を実行⑥

標本抽出 ボタンを押す毎に、標本が抽出され標本統計量（平均値）が計算される。標本の分布が中段のグラフに表示され、標本統計量が [観察された標本統計量] に表示される。

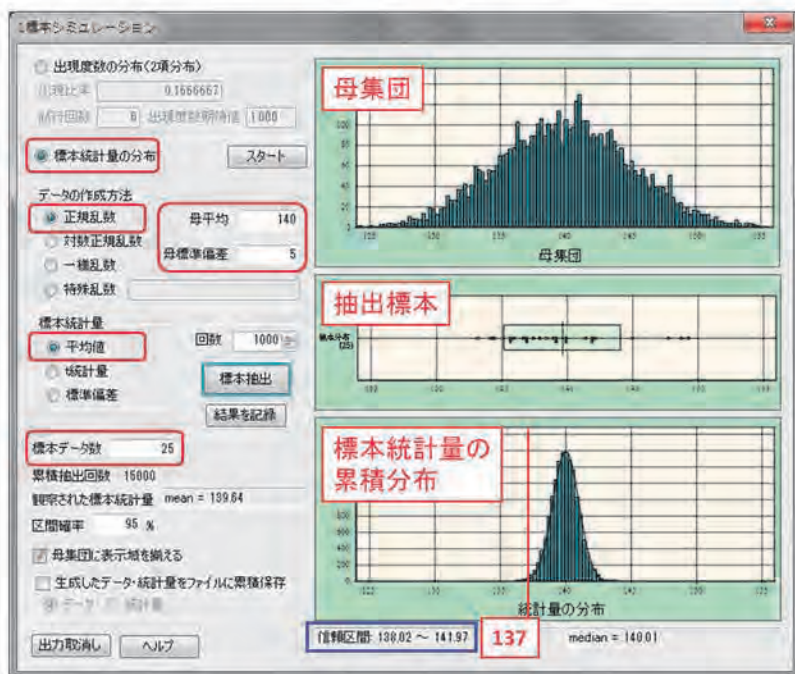
[回数] を変更することで、1クリック当たりの抽出回数を調整できる。この場合、中段のグラフには最終抽出結果のみが表示される。

7. 標本統計量の分布を確認

[累積抽出回数] に表示されている回数分の標本統計量の度数分布図が最下段に表示される。その形状は、抽出回数が多くなると標本統計量の理論分布に近似する。


実行例

全国10歳女子の身長の例題（50頁）について、シミュレーション機能を用いて、標本平均の理論分布を作成し、観察された標本平均137cmの有意性を調べてみよう。



上図は赤枠の部分を設定し、標本抽出を15,000回行った結果である。15,000回分の平均値が最下段の度数分布図に表示されている。これから、観察された標本平均137cmは、極めて稀な値であることが分かる。標本平均の標準誤差 SE は、95%信頼区間が138.02~141.97（青枠）であり、その幅 $141.97 - 138.02 = 3.95$ を $1.96 \times 2 = 3.92$ で割ると1.0077と計算され、ほぼ母集団の $SD = 5.0$ の $1/\sqrt{25} = 1/5$ となっている。

次に、標本データ数を変化させて、標本平均の信頼区間がどのように変化するか調べて、標準誤差との関係を確認してみよう。



第5章



独立2群の差の検定

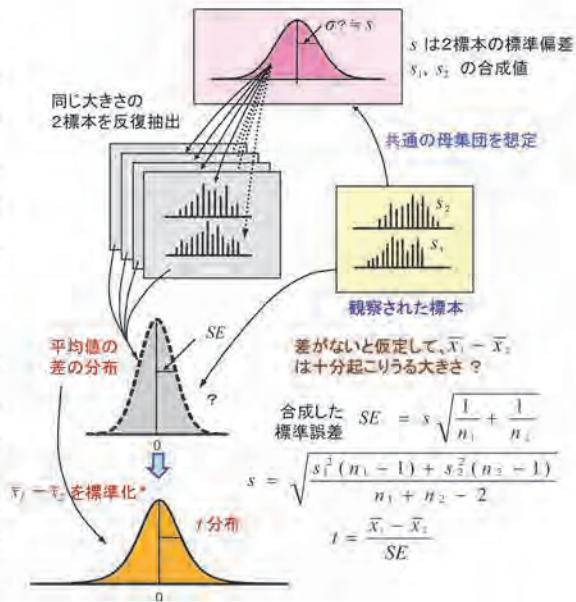
データ数 n_1, n_2 からなる2つの独立した標本の計測値に差があるか、標本平均値の差を目安に検定する方法である。考案者のウィリアム・ゴセットが Student というペンネームで論文を発表したことから、Student's t test とも呼ばれる。計測値が正規分布に従うことを前提に作られた検定法であり、平均値や標準偏差など分布の形に依存する検定統計量が使われる。

2標本t検定の概念

2つの独立した標本 (独立2群) の計測値 (データ数: n_1, n_2) が同じとみなせるかを、各々の平均値 \bar{x}_1, \bar{x}_2 の差 $\bar{x}_1 - \bar{x}_2$ を検定統計量として、それが0と有意に異なるかどうかで判断する。この検定法の意味を理解するには、観察された2標本を合成して共通の母集団を想定すると分かり易い。

いま、共通の母集団から n_1, n_2 の2つの標本を抽出したと仮定 (帰無仮説) して、その平均値の差 $\bar{x}_1 - \bar{x}_2$ を求める。この標本抽出操作を無限に繰り返すと、平均値の差の理論分布が得られる。観察した標本の $\bar{x}_1 - \bar{x}_2$ がどの程度極端な値であるかは、その理論分布上の位置から判定すればよい。

実際の判定には、検定統計量 $\bar{x}_1 - \bar{x}_2$ をその標準誤差 SE で割って標準化すると、その値 t が t 分布に従うことを利用し、 t 値の有意確率から判定する。ここで、真の母集団の標準偏差は分からないので、2つの標本の標準偏差 s_1, s_2 を合成した値 s で代用する。



検定の手順

図のように、異なる条件で求めた2群の計測値があり、2群間に差があると言いたい。しかし真の差が不明なため、差があるという仮定は、そのままでは検定できない。

(1) 仮説の設定：

そこで、いったん「差がない」(帰無仮説 H_0 : 2標本は同一の正規母集団から得られた)と仮定し、「差がある」とする仮定(対立仮説 H_1 : 2標本は異なる母集団から得られた)は保留しておく。

(2) 検定統計量を求める：

両群の平均値を \bar{x}_1 、 \bar{x}_2 とすると、2群の差が平均値の差 $\bar{x}_1 - \bar{x}_2$ に要約されていると考え、 $\bar{x}_1 - \bar{x}_2$ を検定統計量とする。 H_0 より、その期待値は0。

(3) 有意確率 P を求める：

H_0 のもとで、 $\bar{x}_1 - \bar{x}_2$ の有意確率を求める。

これには、 $\bar{x}_1 - \bar{x}_2$ をその標準誤差 SE で割って標準化すると、その値 t が自由度 $df = n_1 + n_2 - 2$ の t 分布に従うことを利用して、 t 分布表から有意確率を調べる。ここに、

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE}$$

$$SE = s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (98 \text{ 頁を参照})$$

ここで標準偏差の合成値 s は2群の分散 s_1^2 、 s_2^2 から、

$$s = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}} \quad \text{で求まる。}$$

(4) 判定：

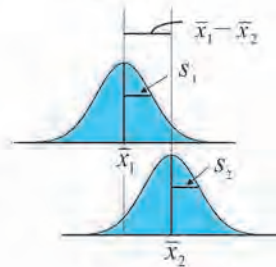
t 分布表より自由度 $n_1 + n_2 - 2$ 、有意水準 α の t 値 (t_α) を調べ、標本の t 値と比較する。

$|t| \leq t_\alpha$ のとき、 $P \geq \alpha$ となり差はない

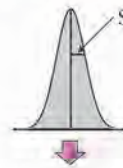
(H_0 を棄却できず、判定保留)。

$|t| > t_\alpha$ のとき、 $P < \alpha$ となり帰無仮説 H_0 を棄却し対立仮説 H_1 を採用。

観察した2標本の分布



$\bar{x}_1 - \bar{x}_2$ の理論分布



SE: $\bar{x}_1 - \bar{x}_2$ の標準誤差

母集団の標準偏差 σ 未知なので、標本の標準偏差合成値 s で代用する

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE} \quad \bar{x}_1 - \bar{x}_2 \text{ を標準誤差 } SE \text{ で割って標準化して } t \text{ 値から判定する}$$

例題
 12

ウサギにある抗原 A を注射して免疫し、その抗体価を調べた。この際、抗原にある添加物 X を加えることで、抗体価が高まるかを調べ、次の結果を得た。A+X 群と A 群の抗体価に有意差があると言えるか？

考え方

差がないという仮説（帰無仮説： H_0 ）より、2群の平均値の差 $\bar{x}_1 - \bar{x}_2$ の期待値は 0 になる。この例では、2 標本から観察された $\bar{x}_1 - \bar{x}_2$ は $16.0 - 9.0 = 7.0$ である。 $\bar{x}_1 - \bar{x}_2$ が 7.0 以上となる有意確率を求め、仮説 H_0 の妥当性を判定する。

ここで、通常は A+X 群と A 群の分散（標準偏差の 2 乗）に差があるか否かをまず F 検定で調べ（94 頁を参照）、2 群間の分散が等しいと見なせる場合に次のように検定する。

① A+X 群と A 群は、同じ正規母集団からの標本と仮定する。母集団は未知であるが、考え方として、観察した 2 標本、A+X 群と A 群のデータを合成した共通の母集団を想定する。

② 共通母集団の平均値は 2 群の総平均 12.8 で代用するが、その標準偏差は不明なので、A+X 群と A 群の標準偏差 $s_1 = 5.83$, $s_2 = 4.52$ を合成し、その平方根 $s = 5.27$ （合成標準偏差）を母標準偏差の推定値とする。

③ この母集団から、データ数 $n_1 = 7$, $n_2 = 6$ の標本を抽出し、 $\bar{x}_1 - \bar{x}_2$ を求める。

④ 理論的に、標準誤差は母集団の標準偏差より

$$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{1}{7} + \frac{1}{6}} = 0.56$$

倍分だけ小さくなる。（98 頁を参照）

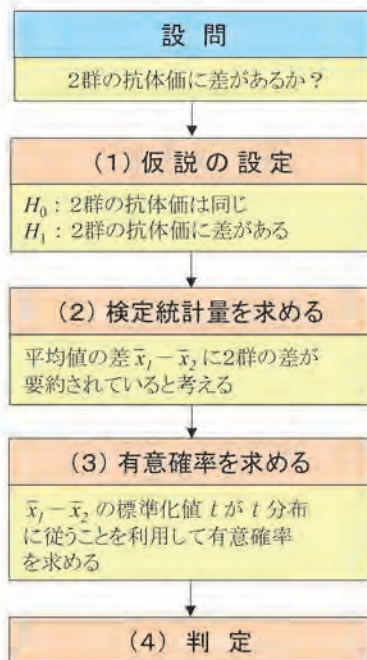
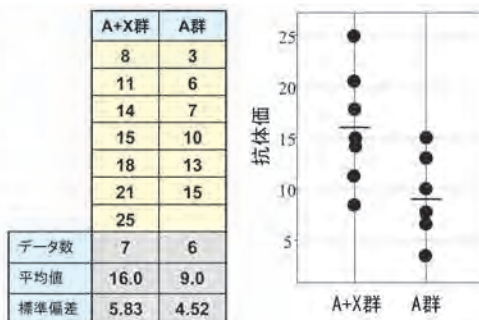
ここでは、合成標準偏差 $s = 5.27$ より標準誤差の推定値は 2.95 となる。

⑤ この理論分布に照らして、観察された平均値の差 7.0 は、分布のかなり偏った位置にある。しかし、真の母集団は未知であり、観察した平均値の差に対する真の有意確率は求まらない。

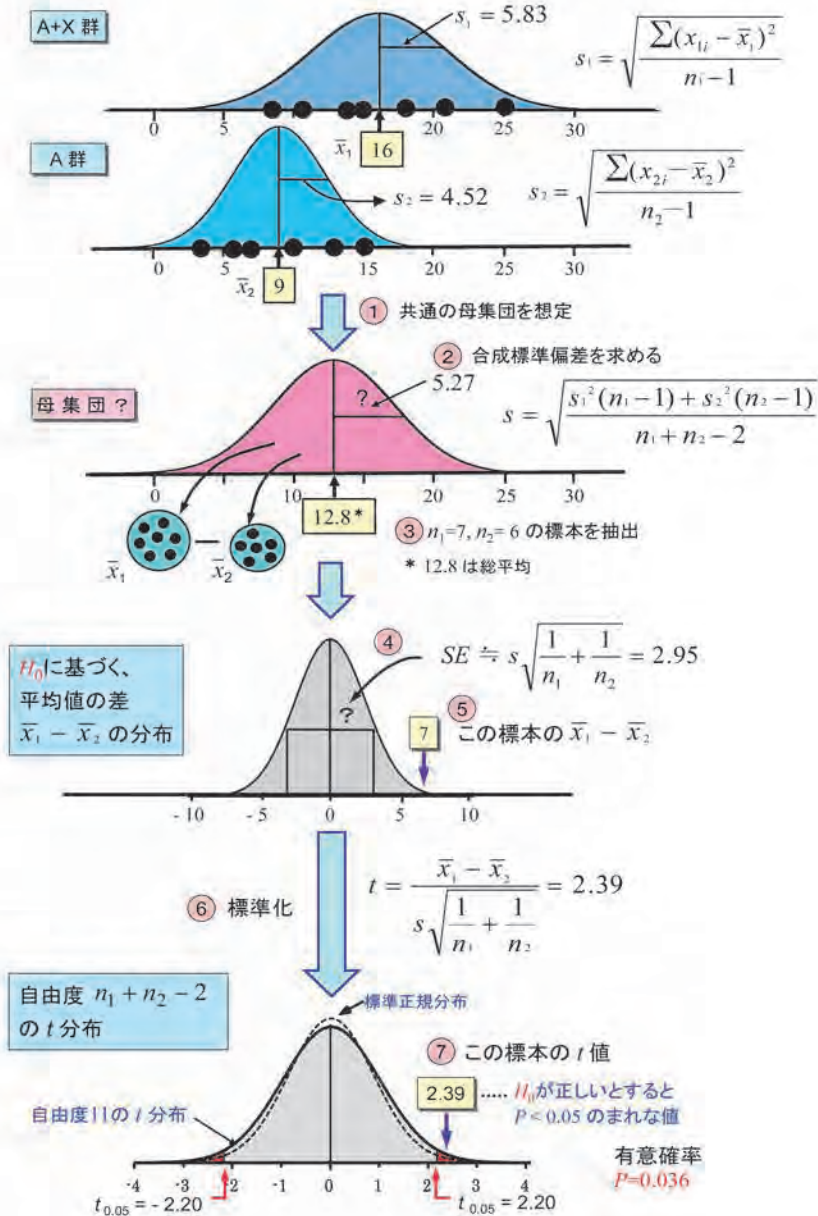
⑥ しかし、 $\bar{x}_1 - \bar{x}_2$ をその標準誤差の推定値で標準化した値、

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{16.0 - 9.0}{5.27 \sqrt{\frac{1}{7} + \frac{1}{6}}}$$

は、理論的に自由度 $n_1 + n_2 - 2$ の t 分布に従うことが分かっているので、 t 値から $\bar{x}_1 - \bar{x}_2$ の有意確率 P を求めることができる。



⑦ $\bar{x}_1 - \bar{x}_2$ の標準化値 $t = 2.39$ は、有意水準 0.05 (両側確率) に対応する自由度 $n_1 + n_2 - 2 = 11$ の t の有意点 $t(11, 0.05) = 2.20$ よりも大きな値である。よって $\bar{x}_1 - \bar{x}_2$ の有意確率は $P < 0.05$ と判定、帰無仮説 H_0 を棄却して対立仮説 H_1 を採用する。すなわち 2 群の標本平均は有意に異なると判定し、「添加物 X は抗体価を高めた」と解釈する (両側検定では変化の方向を考えないが、差が有意な場合は、変化の方向を述べてよい。この例では、抗体価上昇と解釈。) (88 頁 シミュレーションで考えよう 平均値の差の標準化値 t の分布を参照)。



t 分布は分母にある s の曖昧さのため、正規分布 (点線) より両裾広がりとなる。



観察した 2 標本から共通の母集団を想定し、平均値の差 $\bar{x}_1 - \bar{x}_2$ の分布を考える。
 $\bar{x}_1 - \bar{x}_2$ を標準化した値 t が t 分布に従うことを利用し有意確率を求める。

StatFlex での計算

手順:

1. サンプルファイル「例題 12.二標本 t 検定①.SFD6」を開く。
2. 「統計」メニューの「独立群間の比較」の「2 群間検定」を選択する。
3. 統計処理パネルが出るので、検定法の「2 標本 t 検定」にチェックを入れ、「実行」ボタンをクリックする。

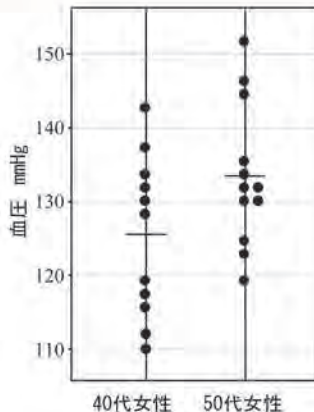
計算結果:

```
<< 独立多群 2 群間比較 >>
< 2 標本 t 検定 >
頁 = [変数 1] A+X 群 vs. A 群
平均値 1 = 16.0000 SD1 = 5.8310 n1 = 7
平均値 2 = 9.0000 SD2 = 4.5166 n2 = 6
平均値の差 = 7.0000 合成標準偏差 = 5.2743
平均値の差の 95 %信頼区間 = 0.5415 ~ 13.4585
t 値 = 2.386
自由度 = 11
有意確率 P = 0.03615
```

例題
13

40代女性 11名と 50代女性 12名の血圧を測定したところ次の結果を得た。40代女性と 50代女性の血圧の平均値に差があるか検定せよ。

	40代女性	50代女性
	130	122
	116	135
	128	132
	110	134
	138	152
	143	145
	118	120
	134	146
	132	130
	112	132
	120	124
		130
データ数	11	12
平均値	125.5	133.5
標準偏差	10.97	9.85



解

- (1) 仮説を設定する。
- (2) 統計量を求める: 2群の合成標準偏差 s と検定統計量 t を下記の式にて求める。

$$\begin{aligned} \text{合成標準偏差 } s &= \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}} = \sqrt{\frac{10.97^2(11 - 1) + 9.85^2(12 - 1)}{11 + 12 - 2}} \\ &= 10.398 \end{aligned}$$

$$\begin{aligned} \text{検定統計量 } t &= \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{125.5 - 133.5}{10.398 \sqrt{\frac{1}{11} + \frac{1}{12}}} = -1.833 \end{aligned}$$

- (3) 判定： $\bar{x}_1 - \bar{x}_2$ の標準化値 $t = 1.833$ は、有意水準 0.05(両側確率)に対応する自由度 $n_1 + n_2 - 2 = 21$ の t の有意点 $t(21, 0.05) = 2.080$ よりも小さい値である。
よって $\bar{x}_1 - \bar{x}_2$ の有意確率は $P > 0.05$ となり、差があるとは言えない (H_0 を棄却できず、判定保留)。

StatFlex での計算

手順：

1. サンプルファイル「例題 13_二標本 t 検定②.SFD6」を開く。
2. 「統計」メニューの「独立群間の比較」の「2 群間検定」を選択する。
3. 統計処理パネルが出るので、検定法の「2 標本 t 検定」にチェックを入れ、「出力」ボタンをクリックする。

計算結果：

```
<< 独立多群 2 群間比較 >>
< 2 標本 t 検定 >
頁 = [変数 1] 40 代女性 vs. 50 代女性

平均値 1 = 125.5455   SD1 = 10.9669   n1 = 11
平均値 2 = 133.5000   SD2 = 9.8489   n2 = 12
平均値の差 = -7.9545   合成標準偏差 = 10.3963
平均値の差の 95 %信頼区間 = -16.9793 ~ 1.0702

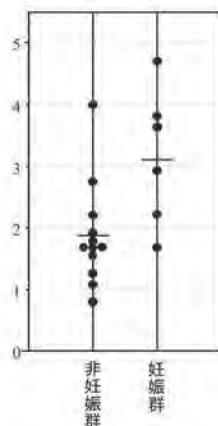
t 値 = -1.833
自由度 = 21
有意確率 P = 0.08102
```



演習:5

25～30歳の女性を対象に調査し、ホルモン値Hを非妊娠群12例、妊娠群6例につき測定し、次のデータを得た。妊娠群と非妊娠群の間に差があると判断してよいか？(解答 246 頁)

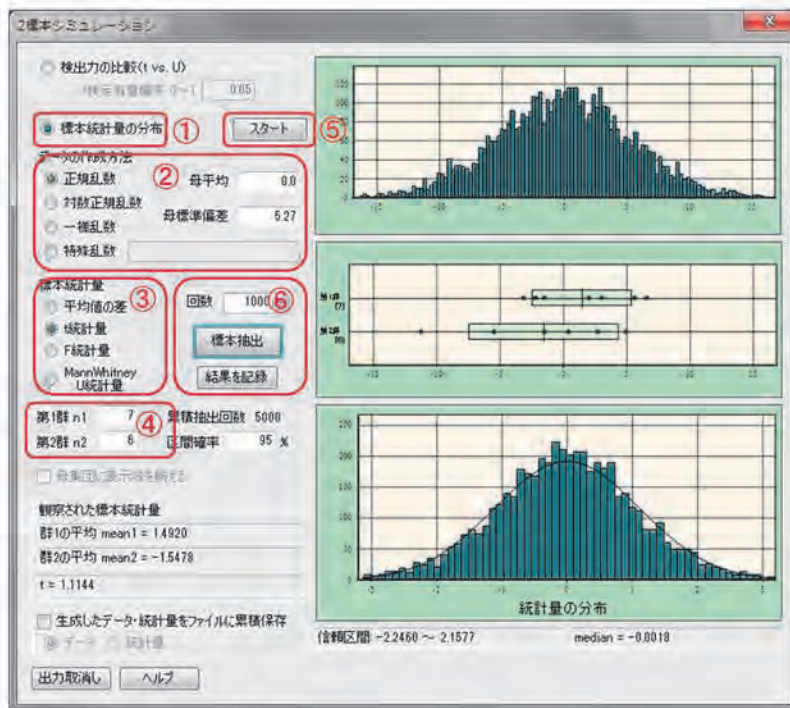
非妊娠群	妊娠群
4.0	4.7
2.7	3.8
2.2	3.6
1.9	2.9
1.8	2.2
1.7	1.7
1.7	
1.7	
1.6	
1.3	
1.1	
0.8	



シミュレーションで考えよう 平均値の差の標準化値 t の分布

ここでは、2標本シミュレーション機能を使って、同一の母集団から2つの標本を取り出し、それから計算される平均値の差の標準化値 t の理論分布を調べてみよう。

実行手順



1. シミュレーションモード選択①

[標本統計量の分布] を選択する*4。

2. 母集団を指定②

母集団の分布型、母平均、母標準偏差を指定する。

※ここで指定した値は、観察する標本統計量が [平均値の差] の場合にのみ、その分布に影響する。

3. 標本統計量を指定③

分布特性を調べるべき標本統計量を選択する。ここでは t 統計量を選ぶ。

4. 標本データ数を指定④

二つの標本の標本データ数(抽出標本のサイズ)を指定する。

5. 母集団を作成⑤

[スタート] ボタンを押すと、指定した条件で母集団が作成され、最上段にそのグラフが表示される。

*4[出現度数の分布] モードは、比率に関するシミュレーションで利用

6. 標本抽出を実行⑥

標本抽出 ボタンを押す毎に、抽出された標本の t 値が計算される。標本の分布が中段のグラフに表示され、 t 値が [観察された標本統計量] に表示される。

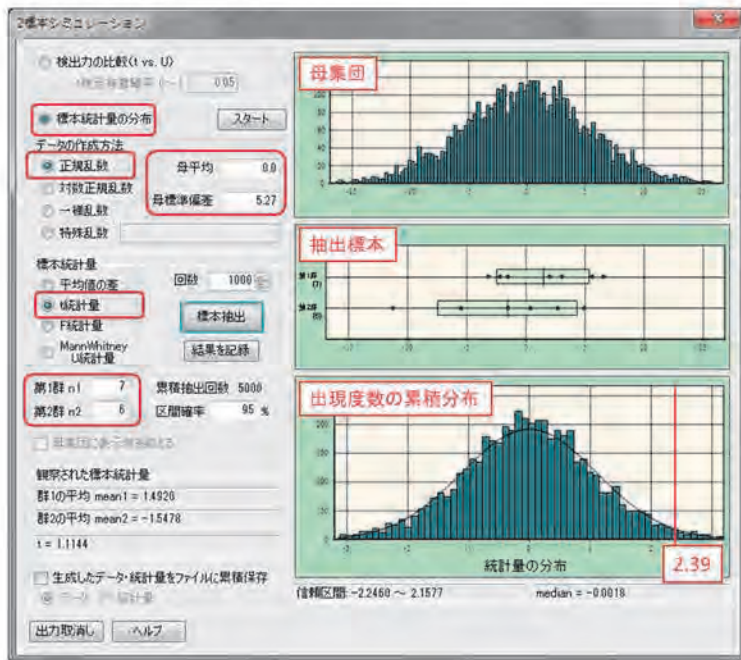
[回数] を変更することで、1 クリック当たりの抽出回数を調整できる。この場合、中段のグラフには最終抽出結果のみが表示される。

7. 標本統計量の分布を確認

[累積抽出回数] に表示されている回数分の標本統計量が最下段の度数分布図に表示される。

実行例

84 頁の例題に対して、2 標本シミュレーション機能を用いてみよう。



上図は以下のように設定し、5,000 回標本を抽出した様子である。

- データの作成法
 - － 正規乱数
 - － 母平均 = 0 (2 群の抗体価に差がないと仮定)
 - － 母標準偏差 = 5.27 (2 標本の合成標準偏差を仮に用いる)
- 標本統計量 = t 統計量
- 標本データ数
 - － 第 1 群 $n_1 = 7$
 - － 第 2 群 $n_2 = 6$

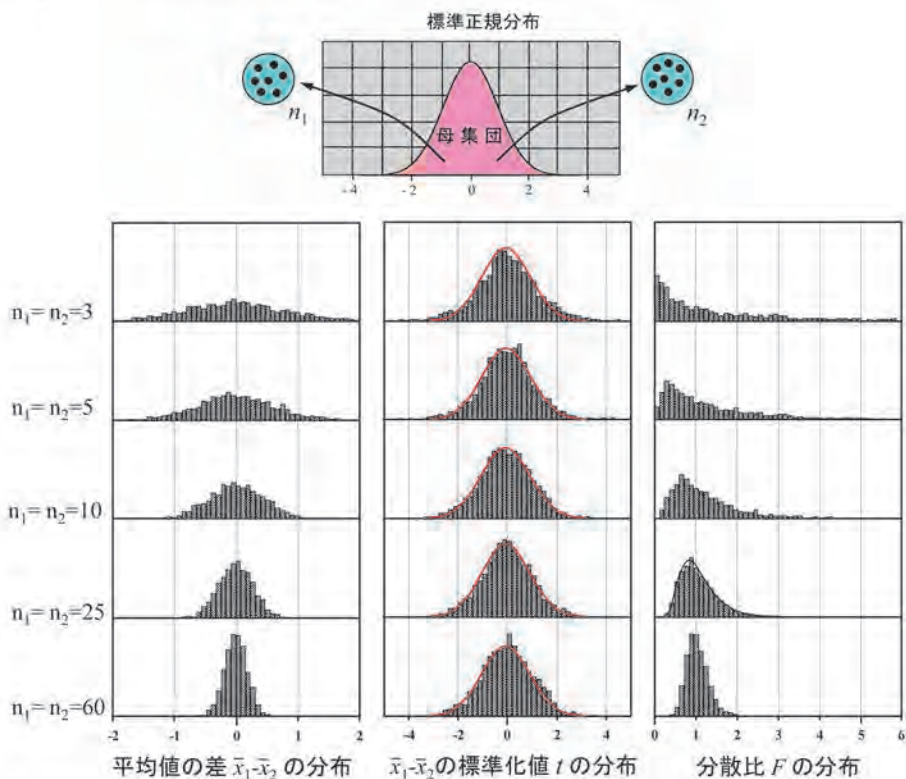
例題で求められた平均値の差は $\bar{x}_1 - \bar{x}_2 = 7.0$ であり、それを標準化した t 値は 2.39 となる。最下段の度数分布図より、比較的稀な値であることが分かる。

ここでは母標準偏差として 2 標本の合成標準偏差で便宜上代用したが、 t 分布は標本データ数にのみ依存する。従って、標本データ数が同じである限り母平均と母標準偏差をどのような値に設定しても、 t 分布の形状は変化しない。

続いて、標本データ数をいろいろ変化させて t 分布の形状を調べてみよう。

2標本から求めた各種統計量の理論分布

シミュレーションで、標準正規分布の母集団(平均値=0、標準偏差=1)から、同じデータ数($n_1 = n_2 = 3, 5, 10, 25, 60$)の標本を2組取り出す操作を2000回繰り返した。各2標本について、平均値の差($\bar{x}_1 - \bar{x}_2$)、平均値の差の標準化値(t 値)、分散比 $F = \frac{s_1^2}{s_2^2}$ を求め、その分布を調べると下図の結果を得た。



この実験から、以下のことがわかる。

- 標本平均の差 $\bar{x}_1 - \bar{x}_2$ は、常に正規分布で、その標準誤差は標本のデータ数が増えると小さくなる。
- 標本平均の差の標準化値(t 値)は、データ数が小さいと、両裾広がり分布(t 分布)になるが、データ数が増えると、標準正規分布に近似する。
- データ数が等しい場合、分散比 F の期待値は1.0となる。データ数が小さいと単調な減衰曲線であるが、 n が増加すると、期待値付近にピークを持つやや右裾広がり分布となる。



正規分布からの2標本について求めた統計量の理論分布

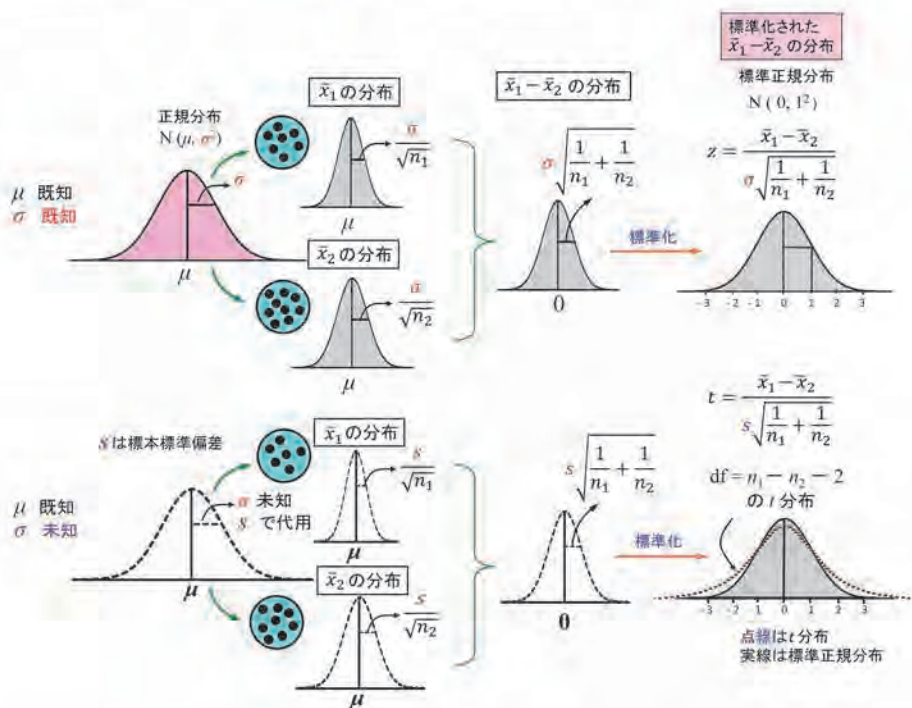
標本平均の差 $\bar{x}_1 - \bar{x}_2$ は、**正規分布**

標本平均の差 $\bar{x}_1 - \bar{x}_2$ の標準化値(t 値)は、自由度 $n_1 + n_2 - 2$ の **t 分布**
(n が大きいと、標準正規分布に近似)

分散比 s_1^2/s_2^2 の分布は、自由度 $df_1 = n_1 - 1$ 、 $df_2 = n_2 - 1$ の **F 分布**

2 標本 t 検定における t 分布とは

2 標本 t 検定では母集団から 2 つの標本を抽出し、その平均値の差が、その理論分布に照らして、どの程度極端であるかを調べる。



上段の図は、母集団の平均値 μ と標準偏差 σ が**既知**で、そこから 2 つの標本（データ数 n_1, n_2 ）を抽出した場合を示す。標本平均の分布は、平均値 μ 、標準誤差はそれぞれ σ の $1/\sqrt{n_1}$ 倍、 $1/\sqrt{n_2}$ 倍となる。また、2 標本の平均値の差 $\bar{x}_1 - \bar{x}_2$ の分布は、平均値 0、標準誤差 SE は**加法定理**（98 頁を参照）により、もとの標準偏差 σ の $\sqrt{1/n_1 + 1/n_2}$ 倍になる。この平均値と SE を用いて平均値の差 $\bar{x}_1 - \bar{x}_2$ を標準化（ z 変換）し、 z 値からどの程度極端な値であるかを判定する。

下段の図は、母集団の標準偏差 σ が**未知**で、そこから 2 つの標本（データ数 n_1, n_2 ）を抽出した場合を示す。標本平均の分布の平均値は母集団のそれと同じであるが、標準誤差は σ が未知なので、2 つの標本の標準偏差の合成値 s で代用すると、それぞれ $s/\sqrt{n_1}$ 、 $s/\sqrt{n_2}$ となる。また、2 標本の平均値の差 $\bar{x}_1 - \bar{x}_2$ の分布は、平均値 = 0、標準誤差 $SE = s\sqrt{1/n_1 + 1/n_2}$ になる。この平均値と SE を用いて平均値の差 $\bar{x}_1 - \bar{x}_2$ を標準化し t と置く。この t 値は、未知の σ を変動を伴う s で代用するため、正規分布よりも両裾の広がった t 分布に従う。 t 値から、平均値の差がどの程度極端な値であるかを判定する。

統計学的推定 (平均値の差の検定の場合)

検定結果が有意な場合、用いた検定統計量 $\bar{x}_1 - \bar{x}_2$ の標準誤差をもとに、その信頼区間を推定する。同じ差でもデータ数によってその信頼性（再現性）が異なるため、学術発表ではこの信頼区間の提示が重要となる。

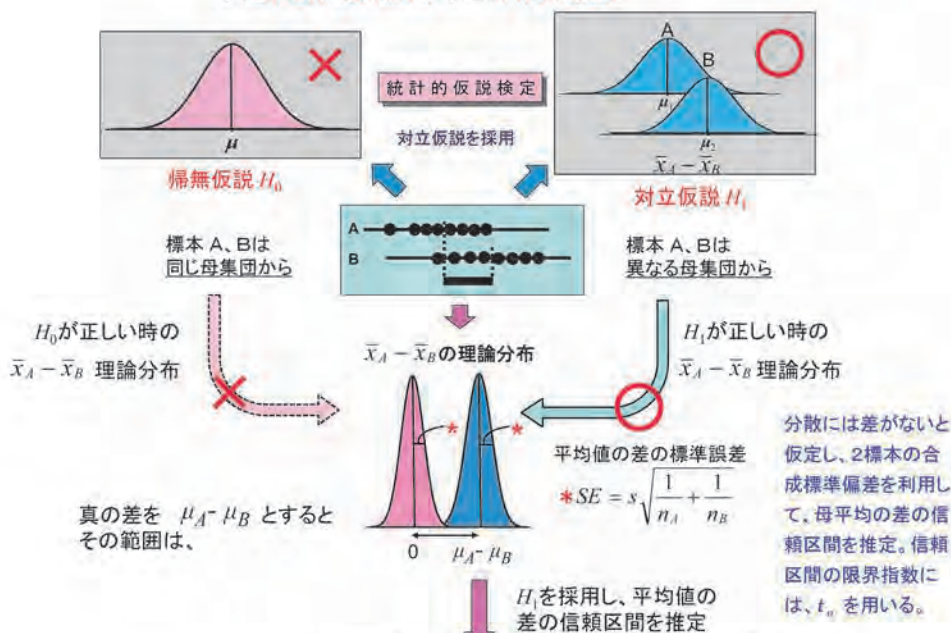
母平均の差の区間推定：観察された2標本の平均値の差 $\bar{x}_1 - \bar{x}_2$ が0ではないと判定した場合（通常は標本平均の差が有意の場合）、母平均の差がどの範囲にあるかを推定する。有意水準を α とし、中央の $100 \times (1 - \alpha) \%$ の信頼区間は次式で求まる。

$$(\bar{x}_1 - \bar{x}_2) - t_\alpha \cdot s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + t_\alpha \cdot s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

ここに、 n_1, n_2 は2群のデータ数を、 s は2群の合成標準偏差を表す。また、 t_α は、有意水準 α 、自由度 ($n_1 + n_2 - 2$) の t 値を表し、データ数に依存する。ただデータ数が十分大きいとき、 t 値は標準正規分布に近似し、通常用いられる 95% 信頼区間の計算では、 $t_\alpha \approx 1.96$ となる。

有意差検定後の推定

観察された差 $\bar{x}_A - \bar{x}_B$ が、“有意”であると判定されたとき(○)、その差にどの程度の揺らぎ(誤差)があるかを推定



$$(\bar{x}_A - \bar{x}_B) - t_\alpha \cdot s \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \leq \mu_A - \mu_B \leq (\bar{x}_A - \bar{x}_B) + t_\alpha \cdot s \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

例題

14

84 頁のウサギと抗体価の例題について、A+X 群と A 群の抗体価の平均値の差の信頼区間を求めよ。

例題 12 より、 $n_1 = 7$ 、 $\bar{x}_1 = 16.0$ 、 $n_2 = 6$ 、 $\bar{x}_2 = 9.0$ 、
2 群の合成標準偏差は $s = 5.28$ である

このとき、自由度 11、有意水準 0.05 の t 値は 2.201 であり、信頼率 95 % の信頼区間は下記の式で求まる。

$$(16.0 - 9.0) - t \times 5.28 \sqrt{\frac{1}{7} + \frac{1}{6}} \leq \mu_1 - \mu_2 \leq (16.0 - 9.0) + t \times 5.28 \sqrt{\frac{1}{7} + \frac{1}{6}}$$

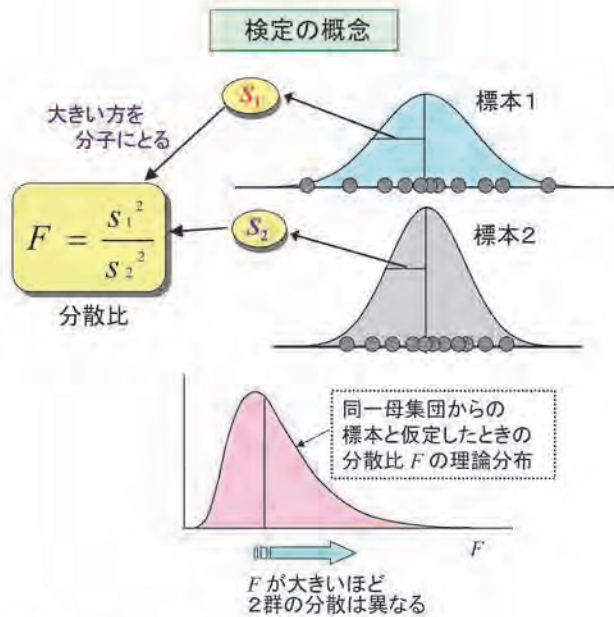
$$7.0 - 2.201 \times 5.28 \times 0.556 \leq \mu_1 - \mu_2 \leq 7.0 + 2.201 \times 5.28 \times 0.556$$

$$4.24 \leq \mu_1 - \mu_2 \leq 11.76$$

A+X 群と A 群の抗体価の平均値の差の 95 % 信頼区間は、4.24 ~ 11.76 である。

等分散性の検定 (F検定)

2標本t検定では、2標本の分散を均等とみなせることが必要条件となっている。いま母集団が正規分布の時、分散比の分布は、2群のデータ数(n_1, n_2)に依存するF分布に従う。このことを利用した2群の等分散性の検定をF検定と呼ぶ。なお検定では、常に分散の大きい方を分子に、分散の小さい方を分母にとって、その比が常に1.0以上になるように調整して検定する。判定は有意水準 α に相当するF値(F_α)の有意点と比較して行う。



検定の手順

上図のように、2群の計測されたデータがあり、2群の計測値の分散に差があると言いたい。しかし差の程度は不明である。

(1) 仮説の設定:

2標本は同一の正規母集団から得られたと仮定(帰無仮説 H_0 : 2群の分散は同等)し、2群標本は異なる母集団から得られたとする逆の仮説(対立仮説 H_1 : 2群の分散は異なる)は保留にしておく。

(2) 検定統計量を求める:

それぞれの分散 s_1^2, s_2^2 を求め、その分散比 F 値を統計量とする。

$$F = \frac{s_1^2}{s_2^2} \quad (\text{大きい方の値を分子において統計量を求める})$$

(3) 有意確率 P を求める :

観察された F 値が、自由度 $df_1 = n_1 - 1$ 、自由度 $df_2 = n_2 - 1$ の F 分布に従うことを利用して有意確率を求めるが、実際には F 分布表から有意水準 α に対する F 値 (F_α) を調べて F 値と比較する。 H_0 が正しいとすると、 F 値の期待値は 1.0 に近い値となる。 F 値の期待値は $df_1 = df_2$ のとき 1.0 となるが、 $df_1 > df_2$ のときは 1.0 よりやや大きな値、 $df_1 < df_2$ のとき 1.0 よりやや小さな値となる (F 分布の形状と期待値 97 頁参照)。

(4) 判定 :

F_α と観察した F 値と比較する (F は常に 1.0 以上となるように計算するので、 F 値の有意確率を 上側の片側確率 として判定する)。

$F \leq F_\alpha$ のとき、分散に差はなく、 H_0 を棄却できない (判定保留)

$F > F_\alpha$ のとき、分散に差があるとして、 H_0 を棄却し対立仮説 H_1 を採用

例題
15

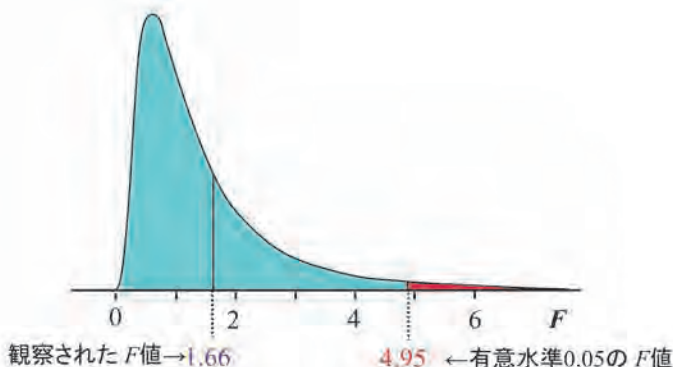
2 標本 t 検定の例題 12 (84 頁) について、2 標本の分散が異なるかを調べよ。

	標準偏差	データ数
A+X 群	$s_1 = 5.83$	$n_1 = 7$
A 群	$s_2 = 4.52$	$n_2 = 6$

$$\text{分散比 } F = \frac{s_1^2}{s_2^2} = \frac{5.83^2}{4.52^2} = 1.67$$

この F 値は、有意水準 $\alpha = 0.05$ 、自由度 $df_1 = 6$ 、 $df_2 = 5$ の F 値 4.95 よりも小さく、「2 群の分散が異なるとは言えない」と判定する。従って等分散と考えて矛盾しないので、2 標本 t 検定を利用してよいことになる。

自由度 $df_1 = 6, df_2 = 5$ の F 分布



有意水準 0.05 の F 値よりも観測された F 値が小さい値であるため「分散に差があるとはいえない」と判断。

StatFlex での計算

手順:

1. サンプルファイル「例題 12_二標本 t 検定①.SFD6」を開く。
2. 「統計」メニューの「独立群間の比較」の「2 群間検定」を選択する。
3. 統計処理パネルが出るので、検定法の「F 検定」にチェックを入れ、「実行」ボタンをクリックする。

計算結果:

```
<< 独立多群 2 群間比較 >>
< F 検定 >
頁 = [変数 1] A+X 群 vs. A 群
F 値 = 1.66667
自由度 df1, df2 = 6, 5
有意確率 P = 0.29580
```

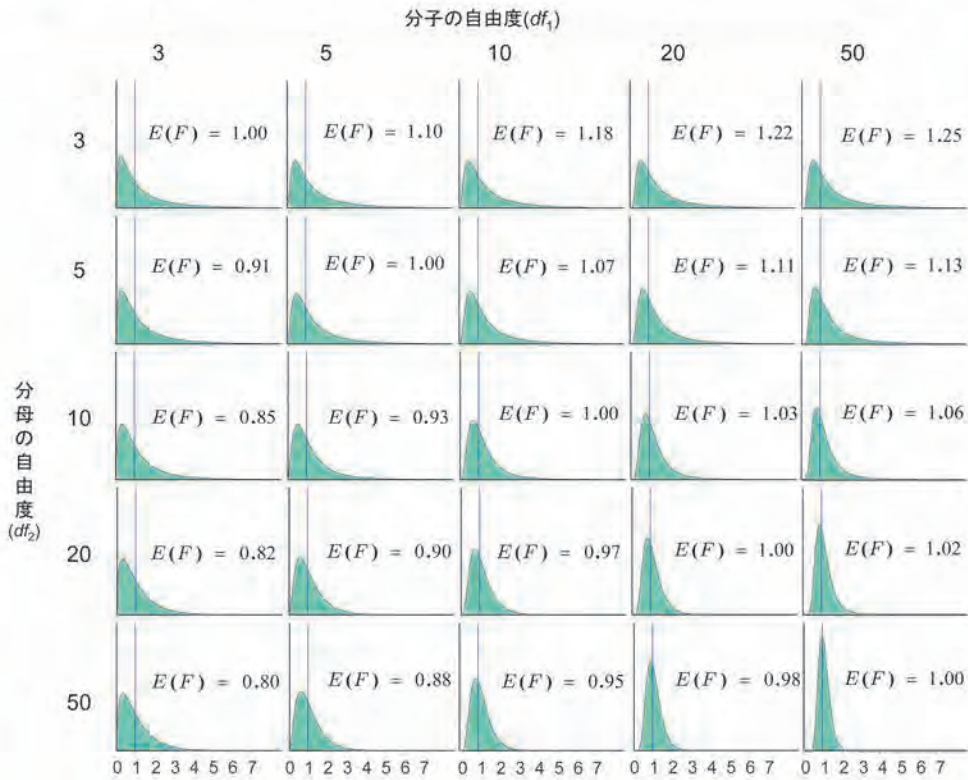
F 分布表の見方

F 分布には、分子側と分母側の 2 つの自由度がある。このため、分布表では代表的な有意水準である $P = 0.05$ と $P = 0.01$ など、有意水準ごとに 1 つの表として構成される。表では、列 (横) 方向に分子の自由度をとり、行 (縦) 方向に分母の自由度をとり、計算された F 値が、表中の値を超える場合に有意水準以下となる。

表3-1: F分布表 $\alpha = 0.05$ 列: 分子の自由度 df_1

	0	1	2	3	4	5	6	7	8
行 分母の 自由度 df_2	1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.86	8.84
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.01
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.83	4.80
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.16	4.13
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.74	3.71
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.45	3.42
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.24	3.21
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.09	3.06
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.96	2.93

参考 F 分布の形状とその期待値 $E(F)$



分散比較 F の分布形状は、分子、分母の分散の自由度 (df_1, df_2) で変化する。

$df_1 = df_2$ のとき F 分布の期待値は 1.0

$df_1 > df_2$ のとき F 分布の期待値は < 1.0

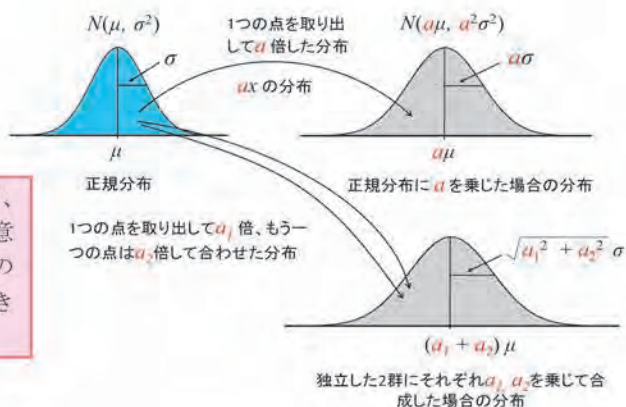
$df_1 < df_2$ のとき F 分布の期待値は > 1.0

参考 正規分布の加法定理

正規分布 $N(\mu, \sigma^2)$ に従う値を正規変数 x としたとき、 x を様々な形で合成すると、どのような分布形状になるかは、加法定理により簡単に導ける。

1. 正規変数 x を a 倍した値 ax の分布

$$\begin{aligned} \text{期待値 } E(x) &= \mu &\Rightarrow E(ax) &= aE(x) = a\mu \\ \text{分散 } Var(x) &= \sigma^2 &\Rightarrow Var(ax) &= a^2 Var(x) = a^2 \sigma^2 \\ \text{標準偏差 } SD(x) &= \sigma &\Rightarrow SD(ax) &= \sqrt{Var(ax)} = a\sigma \end{aligned}$$



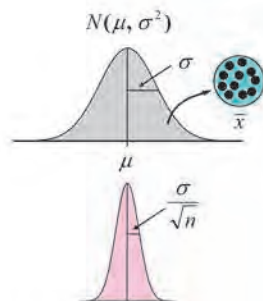
加法定理を使えば、正規分布に従う任意の計測値の合成値の分布形状を予測できる

2. 正規変数 x を a_1, a_2 倍して合成した値 $a_1x_1 + a_2x_2$ の分布 (上図)

$$\begin{aligned} E(a_1x_1 + a_2x_2) &= a_1E(x_1) + a_2E(x_2) = a_1\mu + a_2\mu = (a_1 + a_2)\mu \\ Var(a_1x_1 + a_2x_2) &= a_1^2 Var(x_1) + a_2^2 Var(x_2) = a_1^2 \sigma^2 + a_2^2 \sigma^2 = (a_1^2 + a_2^2) \sigma^2 \\ SD(a_1x_1 + a_2x_2) &= \sqrt{a_1^2 + a_2^2} \sigma \end{aligned}$$

3. 正規変数の平均値の分布 (n 個の正規変数を累和して n で割った値の分布)

$$\begin{aligned} E(\bar{x}) &= E\left(\frac{x_1}{n} + \frac{x_2}{n} + \cdots + \frac{x_n}{n}\right) \\ &= \frac{1}{n} E(x_1 + x_2 + \cdots + x_n) = \frac{n}{n} E(x) = \mu \\ Var(\bar{x}) &= Var\left(\frac{x_1}{n} + \frac{x_2}{n} + \cdots + \frac{x_n}{n}\right) \\ &= \frac{1}{n^2} [Var(x_1) + Var(x_2) + \cdots + Var(x_n)] \\ &= \frac{1}{n^2} (\sigma^2 + \sigma^2 + \cdots + \sigma^2) \\ &= \frac{\sigma^2}{n} \end{aligned}$$



正規分布の平均値の期待値は μ であるが、平均値の分散は σ^2/n となる。

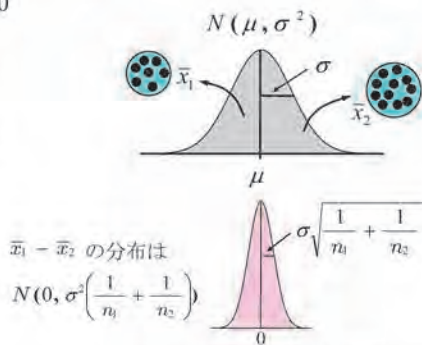
$$SD(\bar{x}) = \sqrt{Var(\bar{x})} = \frac{\sigma}{\sqrt{n}}$$

4. 正規母集団から得た二標本の平均値の差 $\bar{x}_1 - \bar{x}_2$ の分布

$$E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu_1 - \mu_2 = 0$$

$$\begin{aligned} \text{Var}(\bar{x}_1 - \bar{x}_2) &= 1^2 \text{Var}(\bar{x}_1) + (-1)^2 \text{Var}(\bar{x}_2) \\ &= \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \end{aligned}$$

$$SD(\bar{x}_1 - \bar{x}_2) = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$



同一の正規母集団から抽出された2群の平均値の期待値は0で、その分散は $\sigma^2(1/n_1 + 1/n_2)$

例題 16 お中元商品として、500gのマスカット1房と、250gの白桃5個の詰め合わせを作る。それぞれの実際の重量は正規分布とみなせ、下表の平均値と標準偏差をもつ。商品の内容量の平均値および標準偏差の期待値を求めよ。また、実際の商品の内容量の95%信頼区間を求めよ。

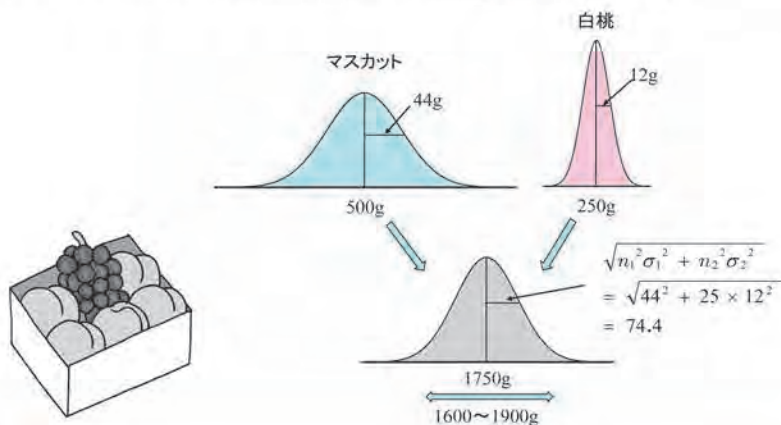
	商品	平均値	標準偏差
x_1	マスカット	500g	44
x_2	白桃	250g	12

期待値： $E(x_1 + 5x_2) = E(x_1) + 5E(x_2) = 500 + 5 \times 250 = 1750g$

分散の期待値： $\text{Var}(x_1 + 5x_2) = \text{Var}(x_1) + 5^2 \text{Var}(x_2) = 44^2 + 25 \times 12^2 = 5536$

標準偏差の期待値： $\sqrt{5536} = 74.4g$

重量の95%信頼区間： $1750 \pm 1.96 \times 74.4g = 1600 \sim 1900g$



正規分布の加法定理から、異なる正規変量の合成値の分布を推定できる



演習:6

和菓子と饅頭を詰め合わせた「おみやげ」を作る。それぞれ下表のような正規分布に従うとしたとき、以下の問いに答えよ。(解答 247 頁)

	商品	平均値	標準偏差
x_1	和菓子	85g	2g
x_2	饅頭	40g	1g

- (1) 和菓子 9 個と饅頭 25 個を詰め合わせる場合、総重量の分布はどうなるか。

平均値 = g 標準偏差 = g

95 % 信頼区間 = g ~ g

- (2) 総重量が 1720g 以下を不良品とみなす場合、全体の何%が不良品となるか。

- (3) 不良品を 3 % 以下にしたい場合は、何 g 以下を不良品とすれば良いか。ただし、不良品かどうかは片側で判断するものとする。



演習:7

ある大学の学生の男女別の体重の分布は下表のとおりである。以下の問いに答えよ。(解答 247 頁)

	性別	平均	標準偏差
x_1	男性	70kg	10kg
x_2	女性	55kg	9kg

- (1) エレベータに男子 4 名、女子 3 名が同時に乗る場合、7 人の合計体重の分布は、平均値 、標準偏差 の _____ 分布をなす。

- (2) 上の問題で、500kg を超えると、ブザーが鳴るとすると、7 名が同時に乗って、ブザーが鳴る確率は、 と計算される。

memo

基本的に、2標本t検定では(1)計測値が連続量で、(2)正規分布に従い、(3)2群は等分散とみなせることを前提としている。しかし、現実のデータでは、それらの仮定が成立することはむしろ少ない。実際上、分布の非対称度が強かったり、極端値が存在すると、t検定は平均値や標準偏差をベースに2群の差を比較しているの、その影響を受けて差の検出力が低下することがある。

一方Mann-Whitney検定では、2群の計測値の順序関係を調べ、その食い違いに着目して群間差の有意性を判定するので、計測値の分布型に依存せず、極端値の影響を受けにくいだけでなく、2群の分散が異なる場合や計測値の尺度が数段階の簡単な場合にも適用できる汎用的な方法である。

検定の手順

(1) 仮説の設定：

2群の点の配置(順序関係)に偏りが無いという帰無仮説 H_0 をおく。対立仮説 H_1 (2群の配置に偏りがある)は保留にしておく。

(2) 検定統計量を求める：

どちらか一方の群に注目し、各点より大きい他方の群のデータ数を調べていく。その合計 U は、群間の差を表す。この例では、 $U = 0 + 0 + 1 + 3$ となる。 U 値が小さいほど、群間の差が大きいことになる。

(3) 有意確率 P を求める：

データ数により U 値の有意確率の求め方が異なる。

・ $n_1 \leq 20$ かつ $n_2 \leq 20$ の場合 Mann-Whitney 検定表をみる。(実際には、 $n_1, n_2 \geq 5$ で $n_1 \times n_2 \geq 100$ の場合には正規分布とみなせる)

・ n_1, n_2 の一方が 20 以上の場合、 U は近似的に正規分布に従う。

$$\text{平均値 } \mu_U = \frac{n_1 n_2}{2}$$

$$\text{標準誤差 } \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

このことを利用して、観察した U 値を次式により標準化し、

その値 z を標準正規分布に照らし、有意確率 P を求める。

$$z = \frac{U - \mu_U}{\sigma_U}$$

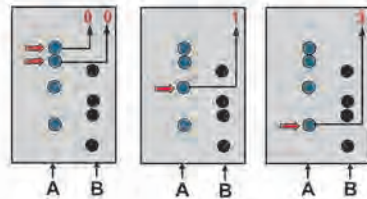
(4) 判定：

有意水準を α とすると、

$P \geq \alpha$ のとき、 H_0 を棄却できず判定保留。差はない。

$P < \alpha$ のとき、 H_0 を棄却し、対立仮説 H_1 を採用。有意水準 α で差があると判定する。

検定統計量 U の求め方



$$U = 0 + 0 + 1 + 3 = 4$$

一方の群(この例ではA群)に注目して個々の値よりも大きい他群のデータの個数を求め、それを累和して U 統計量とする

検定概念

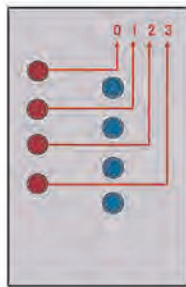
Mann-Whitney 検定と統計量 U について

U をどちらの群に注目して求めるかで、 U 、 U' の2通りに求まる。検定では、通常小さい方の値から判定するので、逆に求めたときは、 $U + U' = n_1 \times n_2$ の公式で、小さい方に直してから判定する。

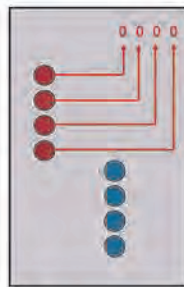
一般に、2群に差がない場合、 U は大きい値となり、差があると言える場合には U は小さい値となる。期待値は $(n_1 \times n_2)/2$ である。

U 値の取りうる値の範囲は、 $0 \sim n_1 \times n_2$ であり、2群が完全に分離しているとき U 値は、 0 または $n_1 \times n_2$ となる。

差があるとはいえない



差がある



統計量 U は2通り計算できるが、検定では小さい方の U を用いる。 U が小さいほど群間差が大きい。

● について $\Rightarrow U = 0 + 1 + 2 + 3 = 6$
 ● について $\Rightarrow U' = 1 + 2 + 3 + 4 = 10$

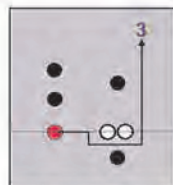
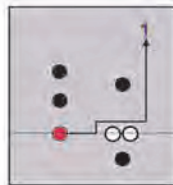
● について $\Rightarrow U = 0 + 0 + 0 + 0 = 0$
 ● について $\Rightarrow U' = 4 + 4 + 4 + 4 = 16$

$$U + U' = n_1 \times n_2$$

同順位があるとき

同順位があるとき^{*5}、次の2つの場合を想定しその平均をとる。

すなわち自分が相手より少しだけ大きい場合（左図）と、相手より少しだけ小さい場合（右図）を想定し、その平均個数を用いて U 値を計算する。



同順位があるとき、2つの場合を想定しその平均をとる。

本例では、赤白の点について、相手と同順位のもの2つあり、 $\frac{1+3}{2} = 2$ を自分より大きな相手の個数とする。

^{*5} 同順位が多いとき、厳密には U 値の補正計算が必要である。詳しくは市原清志著「バイオサイエンスの統計学」p95. 南江堂 1990. を参照。なお、StatFlex の計算は同順位補正に対応している。

例題
17

小標本の場合 ($n_1, n_2 \leq 20$ の場合) :

動物実験で騒音のストレス実験を試みた。ランダムにラットを2群に分け、音に曝露しない群 (C 群)、曝露した群 (E 群) の血漿中コルチゾールを測定した。両群間に差があると考えてよいか。

C 群	13	10	8	6		n=4
E 群	40	26	23	17	13	n=5

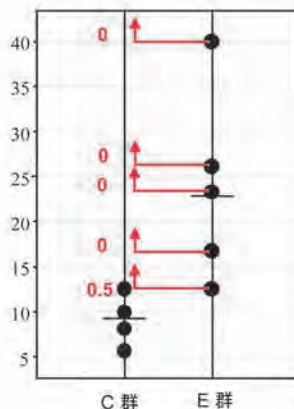
(単位 : $\mu\text{g/dL}$)

考え方

(1) 仮説の設定 :

帰無仮説 H_0 をおく。すなわち C 群と E 群には差はない (2 群の点の順序関係には差がない) という仮説をたてる。

この場合の対立仮説 H_1 は C 群と E 群の測定値に差があるとなる。



(2) 検定統計量 U を求める :

E 群に注目すると、 $U = 0 + 0 + 0 + 0 + 0.5 = 0.5$

逆に C 群に注目すると、 $U' = 4.5 + 5 + 5 + 5 = 19.5$

ここで、 $U + U' = n_1 n_2$ が成立し、 $0.5 + 19.5 = 4 \times 5 = 20$ となっている。

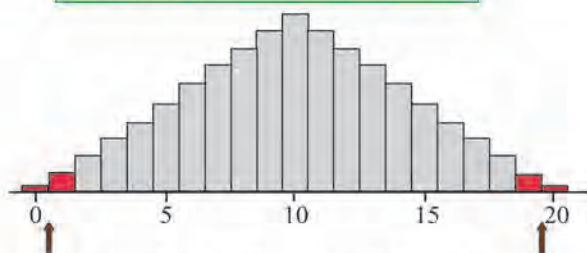
(3) 有意確率を求める :

U 値の有意性をみるため、 U 値がそれ以上に極端となる確率 P 、すなわち、 $P (U \leq 0.5 \text{ or } 19.5 \leq U)$ を調べる。付表より両側確率 $P < 0.05$ となる U 値の下側有意点は 1 (上側点は $U' = 20 - 1 = 19$) で、この標本の U 値はそれより小さい。

(4) 判定 :

従って、帰無仮説 H_0 を棄却し、対立仮説 H_1 を採用する。

U の理論分布 ($n_1 = 5, n_2 = 4$)



0.5 ← この標本のU値 この標本のU'値 → 19.5

両側の有意確率 $P < 0.05$ の部分を赤色で示す。

StatFlex での計算

手順：

1. サンプルファイル「例題 17_Mann Whitney 検定①.SFD6」を開く。
2. 「統計」メニューの「独立群間の比較」の「2 群間検定」を選択する。
3. 統計処理パネルが出るので、検定法の「Mann-Whitney 検定」にチェックを入れ、「実行」ボタンをクリックする。

計算結果：

```
<< 独立多群 2 群間比較 >>

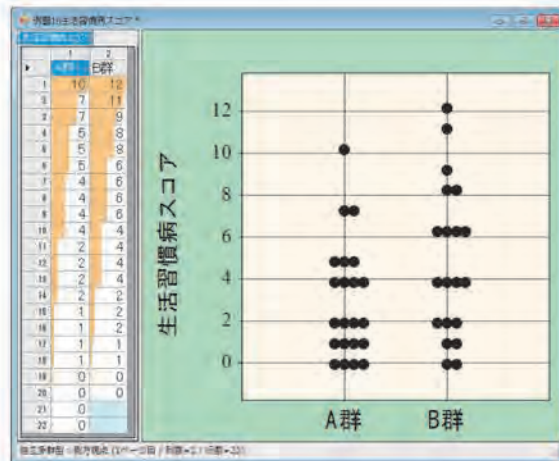
< Mann-Whitney U 検定 >
頁 = [変数 1] C 群 vs. E 群
U 値 = 0.5 (P < 0.05 : 統計表より)
n1, n2 = 4, 5
有意確率に対する U 値
P < 0.05 : U ≤ 1
```

例題
18

n が大きい場合: (U を正規近似可)

A, B 2 つの会社で、40~50 歳の管理職それぞれ 22 人、20 人を対象に、血圧、食習慣、運動習慣、喫煙習慣、ウエスト径の 5 項目を調べ、各項目を 0~3 の 4 段階で評価し、個人ごとに生活習慣病スコアを求めると表のようになった。両社のスコアに差があると言えるか。

A 群	10	7	7	5	5	5	4	4	4	4	2	2	2	2	1	1	1	1	0	0	0	0	n=22
B 群	12	11	9	8	8	6	6	6	6	4	4	4	4	2	2	2	1	1	0	0			n=20
B より大きい A の数→			1	1	1	3	3	3	3	8	8	8	8	12	12	12	16	16	20	20			$U=155$



解

(1) 仮説の設定 :

帰無仮説 H_0 をおく。すなわち A 群と B 群には差はない。
 2 群の点の配置 (順序関係) には差がないという仮説をたてる。
 (この場合の対立仮説 H_1 は A 群と B 群の測定値に差がある。)

(2) 検定統計量 U を求める :

B 群の各データに注目して、それより大きい A 群のデータの数を順に調べると、表の下欄に示す値となる。その合計が 2 群の差を表す $U = 155$ である。

(3) 有意確率を求める :

データ数がともに 20 以上なので、 H_0 が正しいとすると、

$$\text{平均値 } \mu_U = \frac{n_1 \times n_2}{2} = \frac{22 \times 20}{2} = 220$$

$$\text{標準誤差 } \sigma_U = \sqrt{\frac{n_1 \times n_2 \times (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{22 \times 20 \times (22 + 20 + 1)}{12}}$$

= $\sqrt{1576.7} = 39.7$ の正規分布に従う。

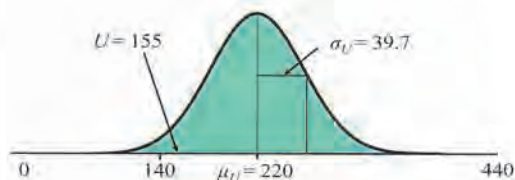
これによって、標本の U 値を標準化すると、 $z = \frac{155 - 220}{39.7} = -1.65$

標準正規分布表より、 $P(|z| \geq 1.64) = 0.099$

すなわち、 $P(U \leq 155) = P(|z| \geq 1.65) > 0.05$

(4) 判定 :

H_0 を棄却できない、すなわち 2 つの会社職員間に生活習慣病スコアの分布に差があるとは言えない (判定保留)。



StatFlex での計算

手順 :

1. サンプルファイル「例題 18.生活習慣病スコア.SFD6」を開く。
2. 「統計」メニューの「独立群間の比較」の「2 群間検定」を選択する。
3. 統計処理パネルが出るので、検定法の「Mann-Whitney 検定」にチェックを入れ、「実行」ボタンをクリックする。

計算結果 :

<< 独立多群 2 群間比較 >>

< Mann-Whitney U 検定 >>

頁 = [生活習慣病スコア] A 群 vs. B 群

U 値 = 155.0 (NS ; 正規近似より)

n1, n2 = 22, 20

U 値の正規近似 $z = 1.65$

有意確率 $P = 0.09850$

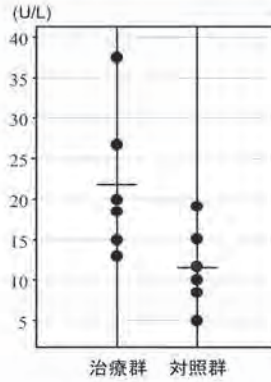


演習:8

ある実験で、12匹の純系動物を無作為に6匹ずつ2群に分け、一方を治療群としてある薬物Aを、他方を対照群としてプラシボ（偽薬）を投与した。一定期間後、肝障害検査であるALT(GPT)の値を測定した。両群に差があると言えるか。（解答 248 頁）

治療群	38	27	20	18	15	13
対照群	19	15	12	10	8	5

(単位 ; U/L)



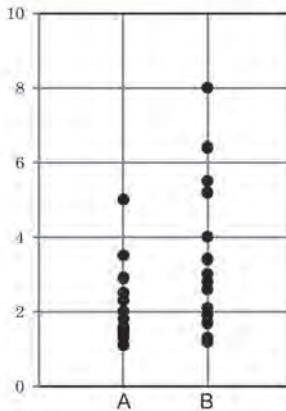
演習:9

腫瘍マーカーの一つである癌胎児性抗原 (CEA) 値について、老人の喫煙者と非喫煙者で比較したところ、次のデータを得た。両者に差があると言えるか。（解答 249 頁）

(単位 ; ng/mL)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
A:非喫煙群	5.0	3.5	2.9	2.5	2.3	2.0	1.8	1.8	1.6	1.5	1.5	1.4	1.3	1.1	< 1	< 1	< 1	< 1
B:喫煙群	8.0	6.4	5.5	5.2	4.0	3.4	3.0	2.8	2.6	2.1	1.9	1.7	1.3	1.2				
Bの前にあるAの数															-	-	-	-

< 1 は測定感度以下であることを示す。



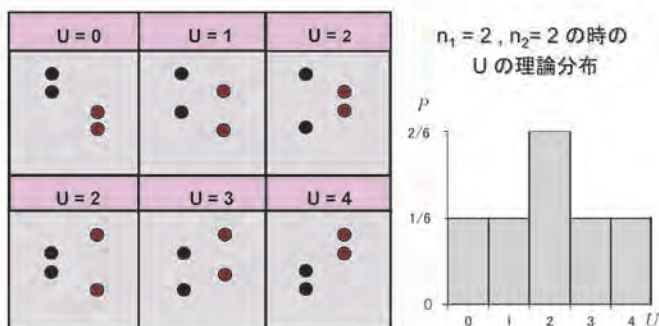
Mann-Whitney 検定の統計量 U の理論分布

帰無仮説 H_0 が正しく、2つの標本が同じ母集団から得られたと仮定し、2群の位置関係(配置)が、偶然どの程度ずれるかを調べ、それぞれの配置について U 値を求めると、 U 値の理論分布が得られる。ここでは、2標本のデータ数が小さな3つの場合について、 H_0 が正しい場合の U 値の理論分布を調べてみると次のようになる。

$n_1=2, n_2=2$ の場合

2群に差がないと仮定すると、両群のデータの相互関係(順序関係)の組み合わせは6通り (${}_4C_2 = 6$) である。

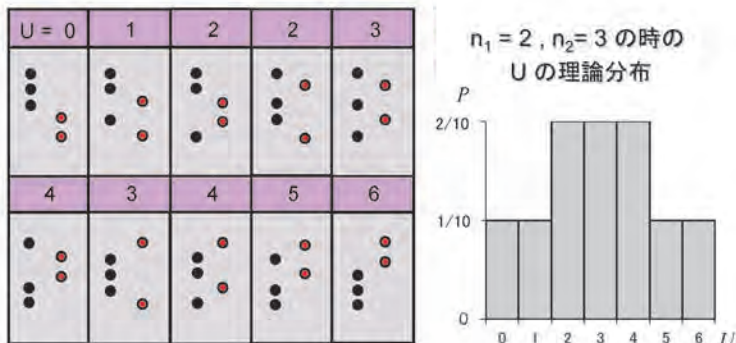
それぞれにつき、 U 値を求めると、0から4の範囲に分布する。



$n_1=3, n_2=2$ の場合

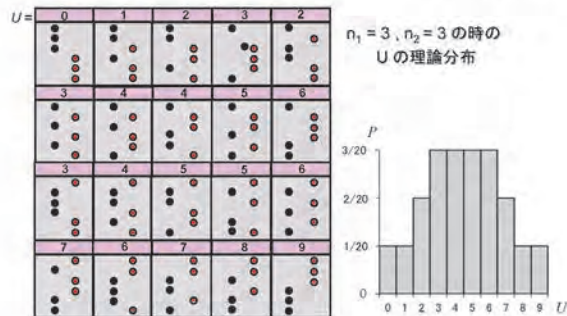
2群に差がないと仮定すると、両群のデータの相互関係(順序関係)の組み合わせは10通り (${}_5C_2 = 10$) である。

それぞれにつき、 U 値を求めると、0から6の範囲に分布し、群間のずれが大きいかほど、 U 値は分布の両端に来る。



$n_1=3, n_2=3$ の場合

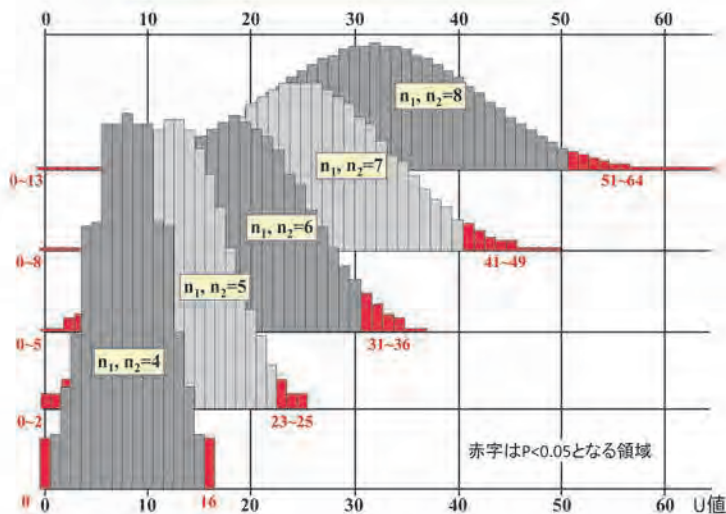
2群に差がないと仮定すると、両群のデータの相互関係（順序関係）の組み合わせは計 20 通り (${}_6C_3 = 20$) である。それぞれにつき、 U 値を求めると、0 から 9 の範囲に分布する。これから、両群の位置関係に差がないとき U 値は分布の中央に、群間のずれが大きいと、 U 値は分布の両端に来る。しかし、両端でもその有意確率は片側で、 $0.05(1/20)$ であり、両側確率で $P < 0.05$ となる U 値は存在せず、このデータサイズでは、検定は成立しないことになる。



$n_1 = n_2 \geq 4$ の場合

2群のデータ数が増えると、 H_0 のもとでの U 値の理論分布は、次第に左右対称の滑らかな分布となる。一般に、 n_1, n_2 が大きいとき ($n_1, n_2 \geq 5$ かつ $n_1 \times n_2 \geq 100$ のとき)、 U は下記の正規分布に近似する。各理論分布両裾の赤い部分は、 U 値に対する有意確率（両側確率）が 0.05 以下となる領域である。すなわち、本来 2 群の配置に差がなくても、 U 値が統計的には有意と判定される領域を表す。

データ数と統計量 U の理論分布



キーポイント

Mann-Whitney の統計量 U の理論分布

データ数 n_1, n_2 が共に大きい場合、 U は近似的に次の正規分布に従う

$$\text{平均値 } \mu_U = \frac{n_1 n_2}{2} \quad \text{標準誤差 } \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

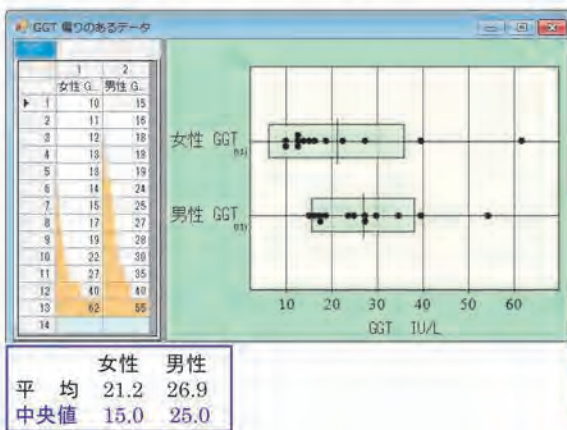
例題
19

2標本t検定とMann-Whitney検定との比較:

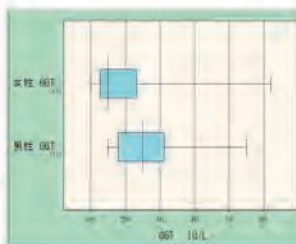
30代の成人男女各13人のGGTの性差について調査をしたところ次のような結果であった。性差があると言えるか。2標本t検定とMann-Whitney検定結果を比較せよ。

(サンプルファイル「例題19_GGT 偏りのあるデータ.SFD6」)

女性	10	11	12	13	13	14	15	17	19	22	27	40	62	n=13
男性	15	16	18	18	19	24	25	27	28	30	35	40	55	n=13



散布図を箱ひげ図に変えると分布の歪みがより明確になる。



考え方:

2標本t検定では、データの分布が正規分布に従うことを前提としている。このデータの分布は全体に低値側に集中する一方で、高値側に極端値の出やすい形となっている。このような場合には、t検定の差の検出力は低下する。一方、Mann-Whitney検定は、分布の形状や極端値の影響を受けず、このような例では検出力が高くなるので、独立2群の差の検定として有効である。

$$\text{平均値 } \mu_U = \frac{n_1 n_2}{2} = \frac{13 \times 13}{2} = 84.5$$

$$\text{標準誤差 } \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

$$= \sqrt{\frac{13 \times 13 (13 + 13 + 1)}{12}} = 19.5$$

$$U = 44 \quad z = \frac{U - \mu_U}{\sigma_U} = \frac{44 - 84.5}{19.5} = -2.079$$



<< 独立多群2群間比較 >>

< 2標本t検定 >

頁 = [GGT] 女性 GGT vs. 男性 GGT
 平均値 1 = 21.1538 SD1 = 14.7583 n1 = 13
 平均値 2 = 26.9231 SD2 = 11.2950 n2 = 13
 平均値の差 = -5.7692 合成標準偏差 = 13.1412
 平均値の差の95%信頼区間 = -16.4074 ~ 4.8690
 t値 = -1.119 自由度 = 24
有意確率 P = 0.27410

< Mann-Whitney U検定 >

頁 = [GGT] 女性 GGT vs. 男性 GGT
 U値 = 44.0 (P < 0.05: 統計表より)
 n1, n2 = 13, 13
 U値の正規近似 z = 2.079
有意確率 P = 0.03761
 有意確率に対する U 値
 P < 0.05 : U ≤ 45
 P < 0.01 : U ≤ 34
 P < 0.001 : U ≤ 23



2 標本 t 検定の制約と Mann-Whitney 検定との使い分け

<2 標本 t 検定の制約>

1) 連続尺度

名義尺度や順序尺度で計測されたデータは分布の正規性という観点から相応しくない¹⁾。

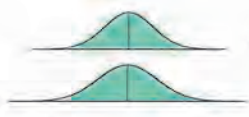
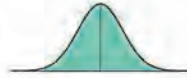
2) 分布の正規性

検定統計量 t に対する有意確率 P は計測値の分布が正規分布であることを前提としている²⁾。

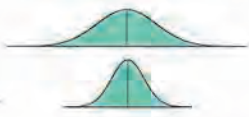
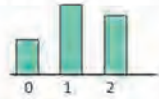
3) 等分散性 (F 検定)

2 群の標本分散が有意に異なれば、同じく統計量 t の有意確率 P の妥当性が失われる。通常 F 検定統計量を用いて等分散性を検定する。

2 標本 t 検定

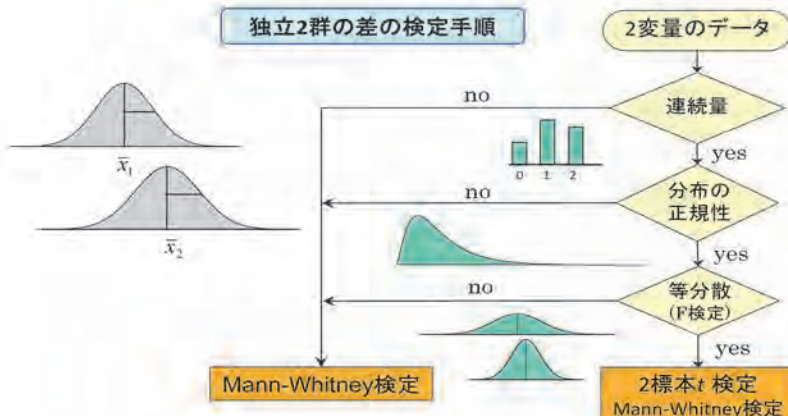


Mann-Whitney 検定



Mann-Whitney 検定の場合、データを順序尺度で分析するので、分布型に依存しない。しかも、正規分布でない場合には、一般に 2 標本 t 検定よりも検出力が高い³⁾。また、等分散性を考慮せずに利用できる⁴⁾。これらの特性から、分布形状が不明な場合には、Mann-Whitney 検定が第一選択となる。

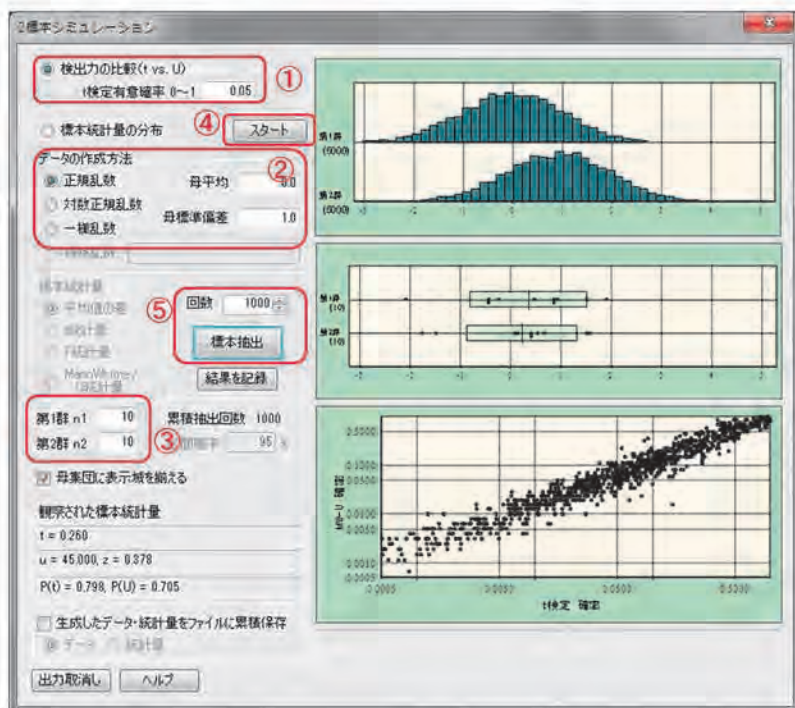
- 注 1 同じ順序尺度でも、多段階で計測されておれば、検定統計量 t の有意確率に偏りを生じにくく、2 標本 t 検定を利用しうる。
- 注 2 正規性の仮定は名目上のことで、通常厳密に検定（歪度検定、尖度検定、 χ^2 適合度検定など）することはない。なぜなら、データ数が少ないと正規分布から偏っているように見えても、正規分布でないと判定できることは少ない。逆にデータ数が多いと正規分布でなくても、中心極限定理により、平均値の差は正規分布に従うので、 t 検定による有意確率は妥当となるからである（204 頁を参照）。ただし、正規分布でない場合に、2 標本 t 検定を利用すると、一般に差の検出力が悪くなるので使うべきでない。
- 注 3 検出力とは、有意差検定において、同じ大きさの差を有意と判定するのに必要なデータ数をさす。より少ないデータ数で差を検出できる検定法のほうが、検出力が高い。
- 注 4 通常 F 検定を行って、等分散性が否定されず、分布の正規性を仮定できる場合には 2 標本 t 検定を利用する。逆に、等分散性が否定された場合、Mann-Whitney 検定を利用する。この場合、2 群の位置関係に有意差はないと判定されたとしても、2 群は少なくとも分布の広がり（形状）という観点で異なる母集団から取られたと判定する必要がある。



シミュレーションで考えよう 検出力の比較

2標本シミュレーション機能のうち検出力の比較機能を使うと、指定した差(2標本 t 検定による有意確率で指定)を持つ2つの母集団を発生させ、それぞれから標本を抽出し、2標本 t 検定とMann-Whitney検定の差の検出力を比較できる。

実行手順



「統計」メニューの「シミュレーション」の「2標本」を選択する。

1. 母集団の差を指定①
[t 検定有意確率]に、2つの母集団のずれを P 値(2つの母集団でから抽出した標本の差に対して、2標本 t 検定で求まる有意確率)で指定する。
2. 母集団を作成②
一方の母集団(第1群)の分布型、母平均、母標準偏差を指定する。他方の母集団(第2群)の位置は、[t 検定有意確率]で指定した有意差の程度に基づき自動的に設定される。
3. 標本データ数を指定③
2つの標本の標本データ数(抽出標本のサイズ)を指定する。
4. 母集団を作成④
[スタート]ボタンを押すと、指定した条件で二つの母集団が作成され、母集団のグラフが表示される。

5. 標本抽出を実行⑤

[標本抽出] ボタンをクリックする毎に、各母集団から標本が抽出されその分布が中段に表示される。そして2標本 t 検定による有意確率 P_t 、および Mann-Whitney 検定による有意確率 P_U が計算され、[観察された標本統計量] に表示される。

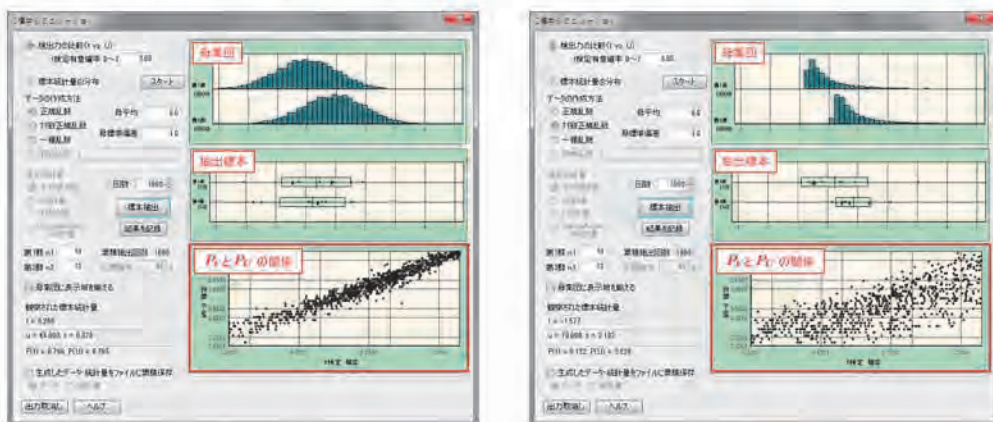
[回数] を変更することで、1クリック当たりの抽出回数を調整できる。この場合、中段のグラフには最終抽出結果のみが表示される。

6. 標本統計量の分布を確認

[累積抽出回数] に表示されている回数分の P_t と P_U の相関図が最下段に表示される。

実行例 母集団が正規分布の場合と対数正規分布の場合の比較

下図は、2標本 t 検定を行った時に有意確率5%となるように2つの母集団をセットし、各々から指定された標本サイズの標本を抽出し、2標本 t 検定および Mann-Whitney 検定を行った結果である。前者の有意確率 P_t を x 軸に、後者の有意確率 P_U を y 軸に取った相関図から2つの検定法の検出力を比較できる。左図は母集団が正規分布の場合、右図は対数正規分布の場合である。



母集団が正規分布の場合には、 P_t と P_U の相関図上の点対角線 ($y = x$) に対してほぼ対称に分布している。これは、 P_t と P_U が平均的には等しいことを表し、 t 検定と Mann-Whitney 検定の検出力は同程度ということになる（厳密には、 P_t は P_U よりやや小さく、前者の方が検出力が少しだけ高い）。

一方、母集団が対数正規分布の場合には、相関図上の点は対角線の下側に偏っている。これは、ほとんどの場合 $P_t \gg P_U$ であることを示し、Mann-Whitney 検定の差の検出力ははるかに高いことを表している。

この結果は、対数正規分布では中央値から大きく偏位した値が出やすく、 t 検定はその影響を受ける (P_t が大きくなる) が、Mann-Whitney 検定は計測値を順位に変換して検定を行うため極端値の影響を受けないことを示している。



第6章



判断分析

感度・特異度・ROC解析

検査法の診断的有用性を評価する方法として感度と特異度、尤度比などの指標算出と ROC 分析 などがある。それぞれの特徴について紹介する。

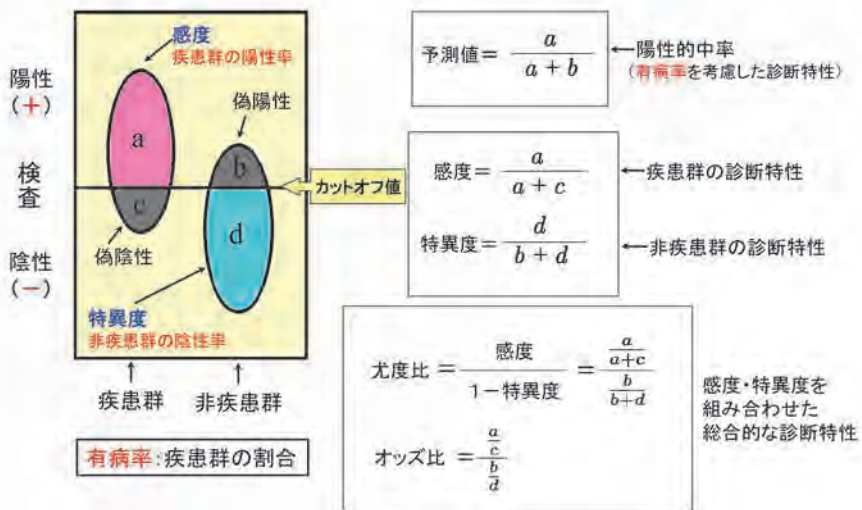
感度と特異度

感度とは、疾患群における検査の陽性率（**真陽性率**）で、感度が高い検査では疾患の見逃し（偽陰性）は少なくなり、逆に感度が低いと偽陰性率が増え、検査の有用性が低下する。

特異度とは、非疾患における検査の陰性率（**真陰性率**）で、特異度が高い検査では偽陽性は起こりにくくなる。

ただし、感度と特異度は相反関係にあり、**カットオフ値**（陽性と判断する境界値）により変化し、感度を上げる（カットオフ値の境界線を下に移動する）と偽陰性は減少するが、偽陽性は増加する。逆に感度を下げる（カットオフ値の境界線を上側へ移動する）と、偽陽性は減るが偽陰性が増えることになる。そこで感度と特異度を組み合わせた検査法の評価指標が必要となる。これには**尤度比**、**オッズ比**、**ROC 曲線下面積**が用いられる。

検査の診断的有用性に関する基本特性



感度と特異度は 2×2 の分割表の形にまとめると求めやすい。これは基本的に χ^2 の独立性の検定で用いる 2×2 の分割表と同じものである。

		疾患・症状		合 計	
		あり	なし		
検査結果	陽性	a (真陽性)	b (偽陽性)	a+b	陽性的中率 = $a/(a+b)$
	陰性	c (偽陰性)	d (真陰性)	c+d	陰性的中率 = $d/(c+d)$
合 計		a+c	b+d	a+b+c+d	
		感度 = $a/(a+c)$	特異度 = $d/(b+d)$		

■ 判別特性値の計算法

(1) 感度 (sensitivity) = $\frac{a}{a+c}$

疾患群における真陽性の割合。感度が高い検査で陰性であれば疾患である可能性が低いと解釈される。

(2) 特異度 (specificity) = $\frac{d}{b+d}$

非疾患群における真陰性の割合。特異度が高い検査で陽性であれば疾患である可能性が高いと解釈される。

(3) 尤度比 (ゆうどひ) (likelihood ratio: LR) = $\frac{\text{真陽性率}}{\text{偽陽性率}} = \frac{\text{感度}}{(1 - \text{特異度})}$

疾患群が非疾患群の何倍陽性になりやすいかを表し、感度と特異度をまとめた指標である。尤度比=1.0 は検査に診断力がないことを表し、1.0 から離れるほど有用な検査と判断される。

(4) オッズ比 (odds ratio) = $\frac{a/c}{b/d}$

オッズは、検査が陽性となる場合と陰性となる場合の比で、オッズが高いほど陽性率（感度）が高い。オッズ比は疾患群が陽性となるオッズと非疾患群が偽陽性となるオッズの比である。

オッズ比が 1.0 の場合、検査に診断力がないことを表し、1.0 から離れるほど検査の有用性が高いと判断される。

(5) 有病率 prevalence (検査前確率) = $\frac{(a+c)}{(a+b+c+d)}$

疾患群の数と非疾患群の数の比。

(6) 陽性的中率 (positive predictive value: PPV) = $\frac{a}{(a+b)}$

真の陽性の割合 (検査後確率とも言われる)。有病率が大きくなると陽性的中率が上がる。

(7) 陰性的中率 (negative predictive value: NPV) = $\frac{d}{(c+d)}$

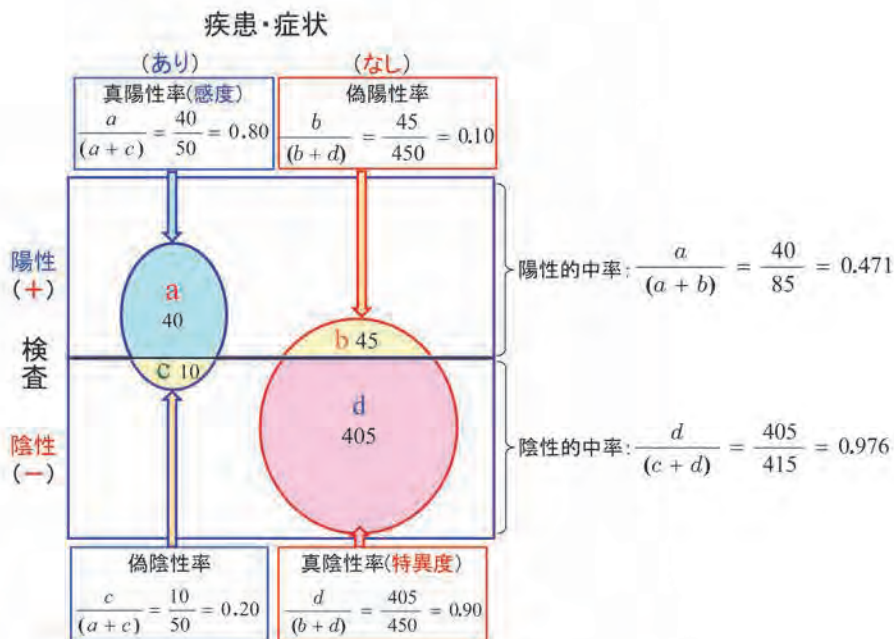
真の陰性の割合

例題
20

ある病院の消化器外来で、初診患者全員に、消化器癌Aで陽性度が高いとされる腫瘍マーカーTを、その疑いのある患者 500 例に対し測定した。確定診断のついた時点でデータを分析すると、下の表のようになった。この腫瘍マーカー検査の診断的有用性を定量的に評価せよ。

		消化器癌A		合計
		あり	なし	
腫瘍マーカーT	陽性	40	45	85
	陰性	10	405	415
合計		50	450	500

感度 (%) = $40/50 \times 100 = 80.0$
 特異性 (%) = $405/450 \times 100 = 90.0$
 陽性的中率 (%) = $40/85 \times 100 = 47.1$
 陰性的中率 (%) = $405/415 \times 100 = 97.6$
 オッズ比 = $(40/10)/(45/405) = 36.0$
 尤度比 = $0.8/(1 - 0.9) = 8.0$



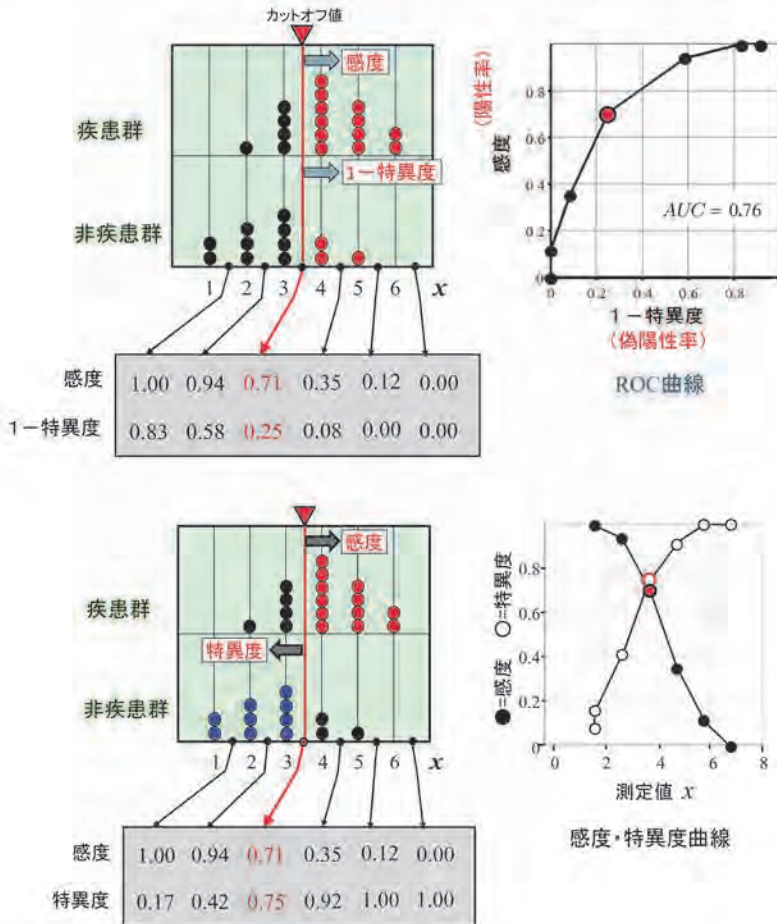
オッズ比 = $\frac{\text{真陽性Odds}}{\text{偽陽性Odds}} = \frac{40/10}{45/405} = 36.0$
 尤度比 LR = $\frac{\text{真陽性率}}{\text{偽陽性率}} = \frac{0.8}{0.1} = 8.0$

ROC分析による2群の判別度の分析

感度と特異度は、カットオフ値をどこに設定するかで変化することから、検査法の有用性を総合的に判断するには不適切である。ROC 曲線 (receiver operating characteristic curve) は^{*6}、カットオフ値を変えながら、縦軸に感度、横軸に偽陽性率 (1 - 特異度) をプロットしそれを順に繋いで作成した曲線である。ROC 曲線の形状や右下部の面積から、該当検査の疾患に対する診断的有用性が評価される。

下図上段は ROC 分析による検査法の評価の具体例で、判別すべき疾患群と非疾患群の検査値の分布と ROC 曲線を示す。また下段右には、カットオフ値を決定する際に参考となる感度・特異度曲線を示す。

例えばカットオフ値を 3 とすると、感度は $12/17 = 0.71$ 、特異度は $9/12 = 0.75$ となる。これをカットオフ値の場所を順に変えて右図のようにプロットして線で結ぶと ROC 曲線が得られる。その曲線の右下部分の曲線下面積 (AUC: Area Under the Curve) の大きさが検査の分別能を表す。この例で StatFlex で計算すると曲線下面積は $AUC = 0.76$ となる。

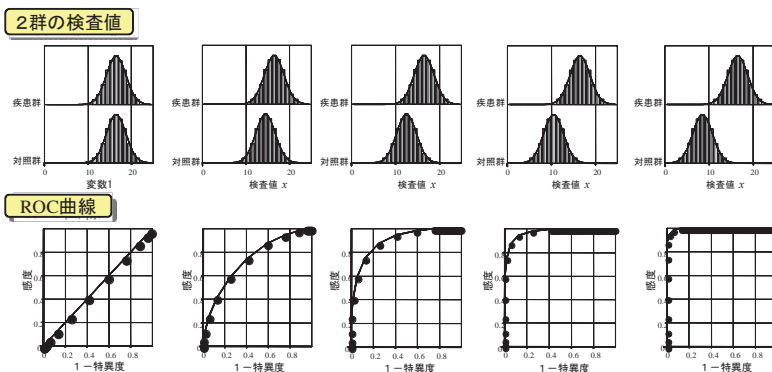


^{*6}ROC 曲線は、第二次世界大戦時にレーダー装置の飛行物体検出特性評価するために考察された装置の操作特性曲線である。飛行物体 (+) のときのレーダー反応シグナルの分布と (-) のときの反応シグナルの分布から装置の評価が行われた。R は受信機 (receiver)、OC は操作特性 (operating characteristic) を表す。

ROC 分析と曲線下面積

一般に 2 群の重なりが少ないほど検査の性能が良く、重なりが多いと検査の性能が悪い。下図に示すごとく左側の図は、疾患群と非疾患群が完全に重なっており、 $AUC = 0.5$ で最も悪い。一方、右の図にいくにしたがって重なりが少なくなり、 AUC は順に大きくなり、一番右の図では $AUC = 1.0$ に近く、非常に優れた検査である。

2 群の検査値の分布と ROC 曲線



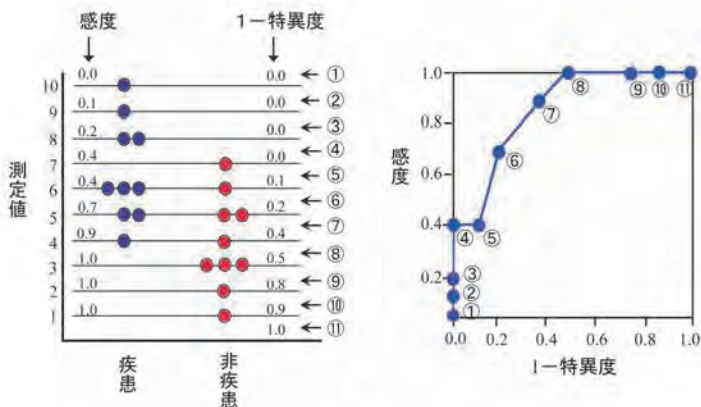
なお、ROC 解析は患者対照研究の中で 2 群（疾患群と非疾患群）の分離度を見る目的で使用されるため、検定を行っているわけではない。疾患群と非疾患群での検査値の群間差を検定し、統計的に有意な差が見られても、曲線下面積 (AUC) が大きくないと検査値の臨床的な有用性は乏しい。

例題 21

検査 A を疾患群と非疾患群について施行し、次のデータを得た。検査 A の疾患分別度を表す ROC 曲線を描け。

疾患群	4	5	5	6	6	6	8	8	9	10
非疾患群	1	2	3	3	3	4	5	5	6	7

カットオフ値を 11 段階に設定し、それぞれについて感度、偽陽性率 (1 - 特異度) を求め、ROC 曲線を作成する。



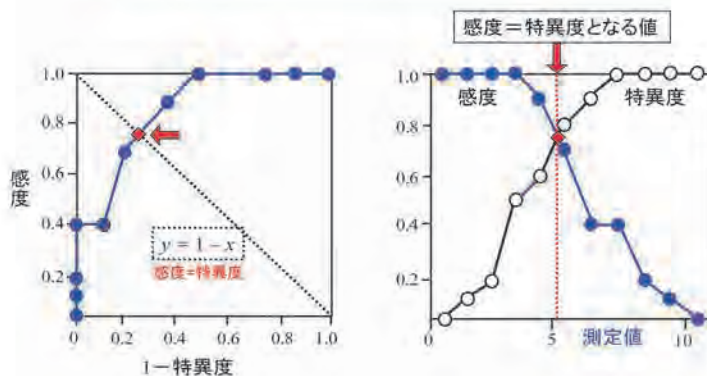
カットオフ値の設定法

カットオフ値は、臨床検査が施行される状況(有病率)や偽陰性、偽陽性の臨床的意義により変わる。従ってそれを一義的に決めることはできず、検査を施行する施設のポリシーにより設定値が変化する。

感度・特異度曲線の利用

有病率に依存する、陽性的中率、陰性的中率を無視して、単に検査の感度と特異度から機械的にカットオフ値を決めるとすると、感度曲線と特異度曲線を下図右のように描き、感度=特異度となる点を採用する事になる。なお、下図左側のROC曲線では、対角線($y = 1 - x$)と、曲線との交点が感度=特異度に相当するが、それに対するカットオフ値はもとの度数分布図上から選択することになる。

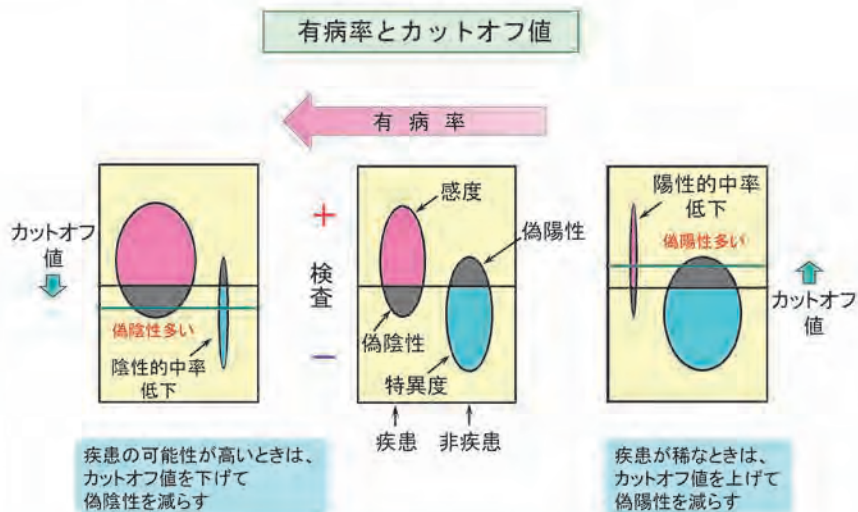
感度・特異度曲線とカットオフ値の決め方



ただし、実際には次項で記すごとく、有病率(疾患群の割合)を考慮して、陽性的中率や陰性的中率が高まるように、カットオフ値を調整する必要がある。また、偽陰性を減らすことが重要な場合や偽陽性を抑える必要がある場合にも、カットオフ値を調整する必要がある。

有病率によるカットオフ値の調整

感度と特異度が同じでも、有病率によって検査的中率は大きく変わる。一般に有病率が高いとわかっている状況では、見逃し（偽陰性）を減らすためカットオフ値を下げることになる（図左）。一方、有病率が低いときは、偽陽性の症例が増える中率が低下するため、カットオフ値を高めることになる（図右）。それをどう調整するかは、検査を施行する側のポリシーで変化し、偽陽性、偽陰性をどの程度重視するかで変化することになる。



ここがポイント!

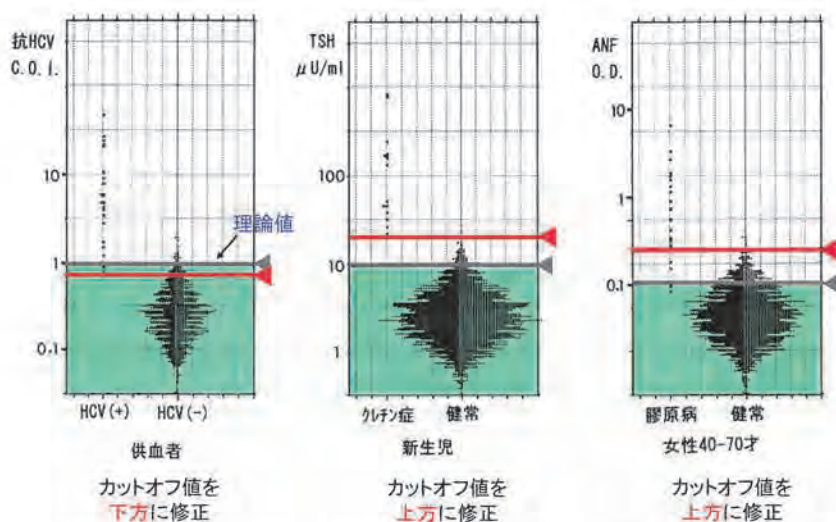
有病率に応じてカットオフ値の補正が必要

偽陽性や偽陰性の重要性を考慮した調整

カットオフ値は検査の偽陰性率や偽陽性率をどの程度重視するかで調整する。図は3つのケースを想定したモデルデータであるが、C型肝炎検査（HCV抗体の検査）では、見逃しが許されないとすると、偽陰性率をほぼゼロにできるカットオフ値が求められる。この場合、特異度や有病率によらず、カットオフ値を下方修正することになる。

一方、血中TSHでクレチン症のマススクリーニングを行う場合、患児のTSHの分布は健常児のそれより明らかに高く、かつ有病率は低い。そこで偽陽性を減らすため、健常児の基準範囲上限よりもかなり高いところにカットオフ値を設定することになる。

また、抗核抗体(ANA)で40~70歳の健常女性を対象として膠原病のスクリーニングを試みたとする。一般にANAは偽陽性率が極めて高く、かつ仮に無症候の偽陰性例もある程度見逃したとしても大きな損失にはつながらないと予測される。従ってこの場合もカットオフ値を高めに設定する必要がある。



C型肝炎検査(HCV抗体)は、輸血のチェックに重要で偽陰性は許容されず、カットオフ値を下げる。

クレチン児は、健常児よりも明瞭にTSHが高く、有病率も低いので、通常よりカットオフ値を上げる。

膠原病の診断に有用な抗核抗体検査(ANA)は、健常者での偽陽性率が高く、多少の見逃しは許容されるため、通常よりカットオフ値を上げる。



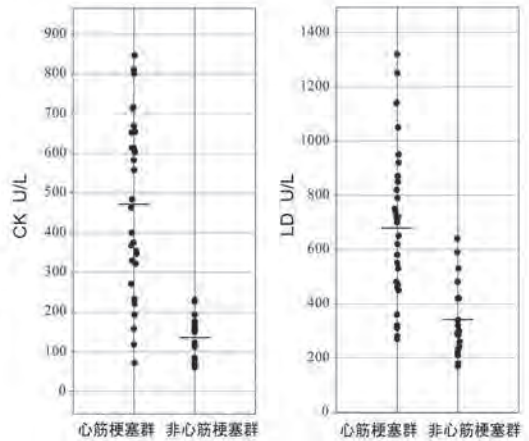
偽陰性、偽陽性をどの程度重視するかで、カットオフ値を調整



演習:10

胸痛で受診した症例 50 に、心筋マーカーである CK と LD を測定した。最終的に心筋梗塞と診断されたのは 30 例で、次のデータを得た。CK と LD の診断特性を比較せよ。(解答 250 頁)

CK		LD	
心筋梗塞群	非心筋梗塞群	心筋梗塞群	非心筋梗塞群
847	230	1250	640
811	226	950	340
801	194	1320	590
716	177	1140	480
714	174	720	420
669	171	1050	320
654	164	850	220
652	160	700	240
614	152	450	284
614	150	820	290
610	123	550	300
603	122	790	170
583	113	870	290
558	112	740	420
484	84	920	530
463	82	720	230
400	80	480	180
375	70	650	420
368	66	580	210
354	61	470	260
347		750	
330		620	
322		530	
271		360	
233		280	
221		450	
193		270	
158		310	
118		460	
72		320	



著者略歴

市原清志 (いち はら きよし)

1975年 山口大学医学部卒業
1979年 大阪大学大学院 医学研究科博士課程修了
1981年 大阪大学講師 医学部臨床検査診断学講座
1992年 川崎医科大学助教授 検査診断学講座
2002年 山口大学教授 医学部保健学科・病態検査学
2005年 山口大学大学院教授 医学研究科保健学専攻・生体情報検査学
2016年 山口大学大学院特命教授 同上、山口大学名誉教授
現在に至る

専門：臨床検査医学、医学統計学、情報科学、内分泌学

主な著作

「バイオサイエンスの統計学」(株南江堂、1990年)
ビジュアル統計ソフト「StatFlex for Windows Ver.6」(株アーテック、2009年)
エビデンスに基づく検査診断実践マニュアル (日本教育研究センター、2011年)

佐藤正一 (さとう しょういち)

1980年 大東医学技術専門学校卒業
2008年 千葉県循環器病センター 医療局 検査部 検査科長
2010年 千葉県救急医療センター 医療局 検査部 検査科長
2013年 千葉県救急医療センター 医療局 検査部 部長
2016年 山口大学大学院 生体情報検査学 博士課程修了
2017年 国際医療福祉大学 成田保健医療学部 医学検査学科 准教授
現在に至る

専門：臨床検査医学、医学統計学

主な著作

病院で使える「臨床統計解析」(株東広社、2009年)

山下哲平 (やましたてっぺい)

2006年 山口大学大学院 理工学研究科 博士課程修了
2011年 滋慶医療科学大学院大学 助手
2013年 滋慶医療科学大学院大学 助教授
2015年 山口大学大学院 理工学研究科 博士課程修了
2019年 滋慶医療科学大学院大学 講師
現在に至る

専門：情報工学、医学統計学

主な著作

ビジュアル統計ソフト「StatFlex for Windows Ver.6」(株アーテック、2009年)

カラーイメージで学ぶ

〈新版〉統計学の基礎 第2版

2016年8月31日 第1刷発行
2020年5月30日 第2刷発行 (定価はカバーに表示してあります)

著 者：市原清志 佐藤正一 山下哲平
発 行：株式会社日本教育研究センター
〒540-0026 大阪市中央区内本町2-3-8
ダイアバレスビル本町1010
TEL:06-6937-8000 FAX:06-6937-8004
制 作：有限会社ユーミット
プログラム開発：市原清志 山下哲平 小柳祐二 佐藤和孝
イラスト：清宮朋子
印 刷：有限会社三共印刷

乱丁・落丁本はお取り替えいたします。
ISBN978-4-89026-180-2

ISBN978-4-89026-180-2
C3041 ¥4500E



9784890261802

定価：本体**4,500円**(税別)



1923041045009

カラーイメージで学ぶ

〈新版〉統計学の基礎 第2版

統計ソフト付属