

解 説

統計解析における統計手法の選択方法 —高いお金を出して統計ソフトを買わなくても、 ホームページを使って統計処理ができる—

病理センター、病理科；病理医

いからし とし ひこ
五十嵐 俊 彦

病理センターにおいて、臨床検査技師の書き上げた投稿論文の査読を依頼される機会が多いが、ときどき統計解析のところでつまずくようである。実際のところ、われわれの扱う統計処理は、マイクロソフト・エクセル（Microsoft Excel）分析ツールの中に収まっている検定ソフトのみで十分作業できるのであろうが、どの検定法を使ったらよいのかわからないようである。エクセル・分析ツール操作の為の市販の「あんちょこ」を読みながら孫引き論文と同様の検定を実施しどうにか統計の結果だけは弾き出されるのであるが、「検定法の選択が正しかったのか？最終的結果の $p=0.123$ とは一体どういう意味があるの？」等々の不安がつきまとうのかどうも座り心地が悪いようである。また、発表会場においても、「これらの2群における5年生存率には有意差は認められない。が、半年目の10才台男児群と2年目における70才台女性群は $p=0.01$ で有意であった。」など多重検定の誤りが後をたない。これらは基本的に日常業務では統計処理が無縁の為に不慣れなことが原因であろう。学校で習ってきた統計講義が実践では全く役に立たないことや、データ収集に苦労した割には満足のゆく統計学的意義を検出できず、また、相談する相手も無い為に、仕方なく、調べたデータ数値だけでもそのまま表にして終了するか、適当に検定を実施することとなる。

私は統計屋ではないが、素人ながら自分流の4つの検定法選択基準に基づいて最適な検定方法を選び、検定処理し、論文として採用してもらってきた。それゆえ読者はこの文章を読むことにより幾分かでも統計処理に自信を持てただけなのではないかと考え、統計処理の手順について、「電子レンジによる簡単な調理法」を比喻にして、解説させてもらうことにした。本来、比喻は科学的文章にはそぐわないのであるが・・・医学用語と同様に統計用語はそれ自体に専門性がある為に日常語で代用できずにとつき難く、それゆえに、統計解析を理解する以前に統計用語自体にアレルギーを示し易く、私の提唱する「猫でもわかる医学ABC」の統計章をより身近なものとしていたかくためにも、あえて、私淑する廣田忠雄・猫田泰敏・雪本卓諸氏のホームページや文献を参考に比喻解説することにした。

以下に、統計処理の基本的作業過程を示す：

(1)「調理レシピ集」を調理用の「電子レンジ」に組み

込む：

「種々の検定ソフトをまとめたもの」たとえば、エクセル添付ソフトから分析ツール・ソフトを「コンピューター」に組み込む（アドイン add-in）。さらに、高価な検定ソフト集を買ってきても良いが、ホームページから集めた4つの無料ソフト（岡本の分散分析・多変量解析、廣田のノンパラ、雪本の「入れボン」、高木の「入れボン」）による解析も愛着が沸いて捨てがたい（アドレスは表1,2を参照、注意：「入れボン」とは数値を入れると、自動的に結果がボンと音はしないが出てくるソフト）。

- (2)「料理の材料」を買い、「冷蔵庫」に保存する：「データを回収」する（表3,4）。統計解析自体はエクセル・ソフト上での検定ソフトによるコンピューター自動処理とし、まず、エクセル上で、苦労して集めたデータをコンピューターに入力する。
- (3)「料理の材料の吟味」：材料の良し悪しや種類を吟味することは、その後の料理の出来を左右する：統計処理においても、検定前に、元データがどのような傾向のものであるかオリエンテーションをつける必要がある。すなわち、データ入力後、まず最初に、基本統計量・度数分布表（クロス表）・ヒストグラム（図1）を求める。これらはほとんどボタン1つで自動的に計算される。実は、ここまでできると統計解析の90%は出来上がっているのである。すなわち、悩みの原因であったはずの狭い意味での検定自体の操作は極めて自動化されており悩む必要が無く、元データの入力とデータ数値の検定前の前処理が本当は大変であり重要であることがわかる。また、基本的な疫学確率も自動的に算定する（表2、5）。
- (4)「どんな料理を作るか、調理法を決める」：「レシピ」集から、材料に最適な「レシピ」を選ぶ：どんな「検定方法」を選択すべきかは、以下の通りである：

(イ)「第1選択基準」は、「自分が何をやりたいか」を明確にすることである。そのことにより、およそその検定の方針が見えてくる（表8）。一般には、以下の5項目に分類される：(i) データの要約と母集団の推定（精密度が不明のため平均値だけでは意味がなく、精密度としての標準偏差また

- は信頼区間を併記することが重要である)、(ii) 比較・区別の検定(群内のデータのバラつきを前提とした群間の平均値の差の有意差を求める)、(iii) 類似性(相関や回帰)、(iv) 打ち切り例のある場合の検討(生存時間解析 Kaplan-Meier 生存曲線⁽¹⁾)、(v) 多変量解析(変数が多項目にわたる場合の検討⁽²⁾)によって構成される。
- (ロ)「第2 選択基準」は材料であるデータ変数の性質を確認することで、以下のように分類される：目的変数(基準変数)の有無、変数の数、変数が量的か質的ななど。
- 以上の(イ)(ロ) 2つの組み合わせで、検定レシビが選択される(表6、9)。以下の2つの基準は、選択された複数の検定法から、更に、最適な検定法を選択する為の基準である。
- (ハ)「第3 選択基準」は検定を妨げる統計誤差(バイアス)の1つである「交絡因子」の有無を考慮することである(表4)。この誤差を避ける為には無作為割付しかないで、われわれの通常の統計処理には常にとまわりつく。「これから調理しようとする材料にまぎれ込んだ新鮮でない材料」の様なものである。この材料の新鮮さに関しての対処の仕方は、(i) 最初から加熱料理(ノンパラメトリックなレシビ集など)のみに限定するか、(ii) 材料(元データ)の新鮮度を確認(2群の分析変数以外の条件を同一にする、すなわち、マッチング)したレシビ集を選択するかのいずれかである。後者の新鮮度の評価方法と対策は表8-3と8-4を参照していただきたい。
- (ニ)「第4 選択基準」は、「データの正規性・等分散であるか否か」である。量的変数のデータにおいて、データが正規性・等分散性であるかまたはデータ数が25~30以上の多数の場合であればパラメトリックな検定法群(エクセル分析ツール)を選択し、正規性でないかまたは不明またはデータ数が少数または変数が質的な変数(順序変数や分類変数)等の場合であればノンパラメトリックな検定法群を選択すべきである(表2)。すなわち、パラメトリック検定は元データそのまの量的変数(間隔尺度)を使うので検出力の高い優れた検定法であり、ノンパラメトリック検定は質的な変数(順序変数または分類尺度)すなわち相対的な順位番号(表6)を使う為に検出力が劣る検定法で、検定数値もその結果としての検出力も全く異なるのである。例えてみれば、この正規性の確認は、調理法上での和食レシビ集と洋食レシビ集のどちらを選択すればよいのかの判断基準に比喻される。すなわち、和風調理集を選択しその調理集からアンキモの調理レシビを選択すべきであるのに、間違ってフランス調理集を選択すると、そのフランス調理集の中から似て非なるフォアグラの調理レシビを選択することになる。材料(データ)が似ているからと、調理集を取り違えてはアンコウも浮かばれないであろう。データの正規性確認は、ヒストグラムの裾が狭く(データのバラつきが少ない)か、グラフが左右対称(データが正規分布)である。正確には、(i) 前述で計算された基本統計量表中の尖度(富士山のようなグラフの尖がり度)のことで、裾広がり程度を示す)が3に近いこと(およそ2.5~3.5)、(ii) 歪度(左右

対称性のことで、右や左に中心がズレていないかを示す)が0に近いこと(およそ-0.5~+0.5)。

(iii) 正規確率紙、(iv) Shapiro-Wilk 検定法、(v) Kolmogorov-Smirnov 検定法、(vi) 2群の正規検定 $z = |(\text{平均値の差}) / \sqrt{[(\text{分散}/n_1) + (\text{分散}/n_2)]}| \leq |z_{0.05}|$ で、2群の等分散性は (vii) $F = \text{大きい方の分散} / \text{小さい方の分散} \leq F(n_1 + n_2 - 2, 0.05)$ で確認される。

しかしながら、著者はいたってアバウトの性格で、検出力が劣っても、統計解析にはバラツキがつきもので、灰色の検定結果が出て仕方がないと思っているので、正規性・等分散性の基準を厳格に規定せずに変数変換してパラメトリック検定したり、更に、ズボラに検出力の劣るノンパラメトリックで有意差がでたら充分くらいに検定している。

以上が、統計解析法上の最適の調理レシビ(検定ソフト)の選択時に気をつけなければならない「イロハ」である。

- (5)「調理」：選択したレシビで、調理用電子レンジに、実際に調理させる：
選択された検定ソフト上で、コンピューターに統計解析をやらせる。これもボタン1つで結果を得ることが出来るので極めて容易である。
検定結果は、有意差が認められる場合、おおよそ、 $p < 0.05$ (または*) で表記されることが多い(表7)。有意差が認められた場合、データのバラつきに対して平均値の差が顕著であっただけで、症例データ数を多くすれば微小な差であっても統計学的有意差は認めやすくなるので、検定後は常に検定結果の実用性を検討し、非現実的な統計の為の検定だけは避けるべきである。逆に、有意差が認められなかった場合、検定された両群が同一のものであることとイコールではなく、症例数を増やして再度検討することが必要であろう(表4b)。

また、ホームページ・参考書・無料統計ソフトのダウンロード・アドレスを紹介した。無料統計ソフトはいずれも常に短し擧に長して、包括的なソフトを提供しているホームページは少ないが、比較的纏まったソフトの提供アドレスを明示したつもりである。自分に最適のレシビを揃えることをお勧めする(表1,2)。

逆に、よりわかり易い統計のホームページや無料ソフトがあれば、教えていただき、会員に紹介したいので、よろしくご教授してください。

キーワード：臨床試験、生物統計学、検定、無料統計ソフト、ホーム・ページ

蛇足。検定手技自体よりも重要な「証明力の強さと実験計画手技」—とどのつまり統計はバラつき(正規性の有無)とバイアスがriskyという話—：

日常診療において統計は馴染みの無いものゆえに、統計処理の手技自体が難しいようである。個人的には、統計処理よりも、それ以前のデータ収集の段階での「偏り、バイアス(bias)」や「交絡因子(広義のバイアスの1つ)」を取り除くことが一層難しく思う。純粋数学としての確率論とは異なり、統計学では random sampling は不可能である

為に、常に、群内でのデータのバラツキの存在下での精密さ（precision、バラツキの少ない状態（標準偏差が小さい））と正確さ（accuracy、平均値が真値に近い）が求められる。実験計画は、Fisher（技術評価のための）三原則が満足されて乱塊法となりうる。すなわち、(1)無作為化：偏りを避ける目的、(2)反復：検定力を高める、(3)局所管理：なるべく条件を均一にする為の努力が重要とされる。故に、信頼度の低い統計素人のわれわれの検討している観察的研究の case control study においてはバイアスが多く、まず、データ結果を常に疑うことが重要であろう。

老婆心ながら（お節介？）：

なお、新潟県厚生連医誌編集・投稿支援目的に病理センターとして市販の安価な統計解析ソフトを購入しました。看護研究の統計学的解析（EBN）等の研究発表時に相談していただければわかる範囲で協力したいと存じますので、元データを入力したフロッピーディスクを添付して相談してください。

謝 辞

東京都立保健科学大学看護学科猫田泰敏先生には、ホームページの紹介の掲載についてご賛同をいただき感謝いたします。

文 献

1. 五十嵐俊彦他. 子宮頸癌の臨床統計—特に子宮頸部腺癌をめぐる問題点—. 日産婦新潟地方会誌

1987；44：1-5.

2. 統計データ処理. 五十嵐の猫でもわかる医学 ABC—病態の理解とその診療指針—. 五十嵐俊彦著. 第37版, ビーマン出版.
長岡.2003. (Hard Disc, 33.0GB. softs: Windows Me, Word 2000, 一太郎 Ver. 12, Exel 2000, DocuWorks Ver. 4. 0, Photoshop Ver. 6)
3. Ikarashi T et al. Histopathological statistical discrimination between well differentiated type of endometrial adenocarcinoma and endometrial adenomatous hyperplasia. Acta Medica et Biologica 1989；37：129-33.

英 文 抄 録

Commentary

Clinical (experimental) examination and biostatistics

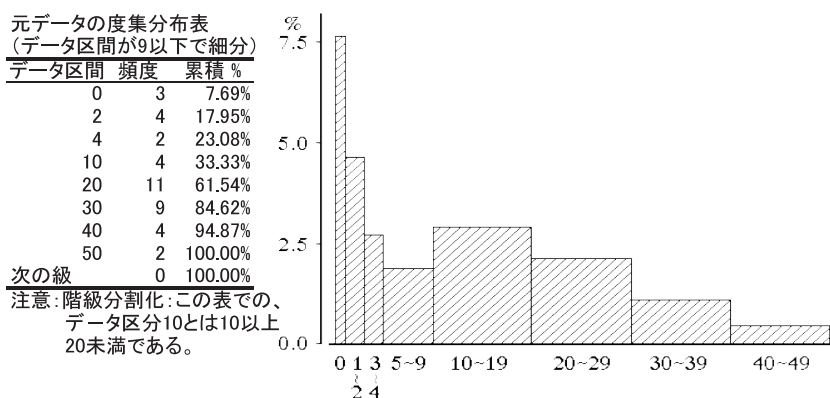
Pathology Department, Pathology Center；Pathologist
Toshihiko Ikarashi

There were several opportunities to check articles written by medical technologists in Pathology Center, holding troubles in a statistical analysis. Though most data were automatically analyzed by statistical softs with the Exel Analysis Tool (Microsoft), it seemed a problem that the present case should be analyzed by what kind of method. Therefore we introduced several references and figures to show which method to be chosen.

Key Words: clinical examination, biostatistics, statistical analysis, statistical softs, free of charge, home-page

図1. ヒストグラム—グラフ表示に神経質な研究者がいるので、注意が必要—

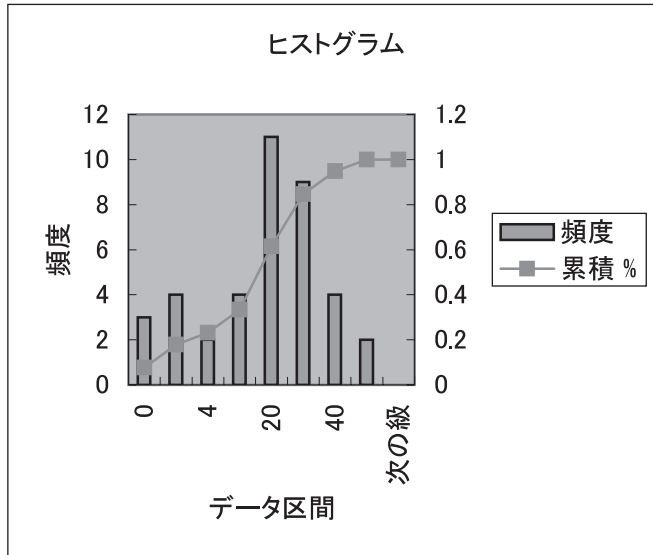
- (1) 左下の度数分布表を正確なヒストグラムにすると：



正しいヒストグラム：x軸は等間隔で、細分化された部位では長方形面積で頻度（度数）が表現される。x軸の値は区間分類なので当然隙間が無い。

統計解析における統計手法の選択方法 —高いお金を出して統計ソフトを買わなくても、
ホームページを使って統計処理ができる—

(2) Excel でヒストグラムを自動作成させると：

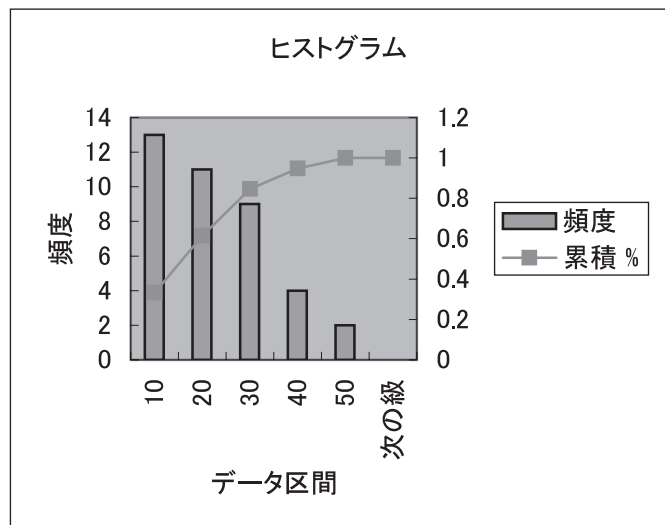


正しくは正確でないヒストグラム：Excel では x 軸が区間で等間隔となり、隙間が開く。x 軸値は境界値未満の区間値 (A ~B) である。

(3) Excel ヒストグラムにおいて等間隔のデータ区間に区切ると：

元データの度集分布表
(データ区間が等間隔)

データ区間	頻度	累積 %
10	13	33.33%
20	11	61.54%
30	9	84.62%
40	4	94.87%
50	2	100.00%
次の級	0	100.00%



正しくは正確でないヒストグラム：Excel では x 軸区間を等間隔指定にしかできない！当然、x 軸値は境界値未満の区間値 (A ~B) である。が、隙間はどうしても開くのです。

表1. ホームページ、文献、ソフト集

分類		商用ソフトウェア	開発元・販売元 (賛助会員・ あいうえお順)	ホームページ・アドレス	コ メ ン ト
統計ソフト	統計ソフト・データに関する情報			http://www.jscs.or.jp/etc/softlink.html	統計解析の市販のソフトやフリーソフト
			ウータンの統計講座、準備編2	http://www.takenet.or.jp/~hayakawa/u-tan%20index.htm	
	無料	Exel, Mcirosoft. ツール→分析ツール一覧.			エクセルにアドインして利便性に優れている。統計基本3ソフトの1つ。パラメトリック検定の基本ソフトのみであるが、生存率曲線が不要の場合、これのみでも充分である。安上がりで、最も一般的。
		メールマガジン「独学をすすめ(MM版) 統計の研究法編」について、雪本草編		http://fresheye.oem.melma.com/mag/39/m00049439/a00000021.html	一押し！使いやすいソフトが1つになっている。著者の理解力によって、記載内容は読者の為に漫談風とし、判りやすい。分散分析の部分だけでもプリントして読むことをお勧めします。
		統計学とHALBAU/HALWINの世界、わかる統計学、簡単統計計算		http://www.clg.niigata-u.ac.jp/~takagi/takagi.html	2押し！わかる統計学で勉強し、簡単統計計算に数個データを入力すると、暗箱より、ボトリと答えだけが出てくる。
		心を測る。岡本安晴 (日本女子大心理学)		http://www.ikuta.jwu.ac.jp/~yokamoto/openwww/	3押し！大半の統計ソフトがダウンロード可能。分散分析と分散分析用ノンパラメトリック解析に優れている。
		お手軽統計マクロ集 Stat macros for Exel. 廣田忠雄 (国際キリスト大)		http://www.tuat.ac.jp/~ethology/Columbo/Stat/Data_Trans/	基本3ソフトの1つ。ノンパラメトリック検定の基本ソフトを含む。無料ソフト！インターネット上よりダウンロード可能で、Exelのツールアドインとして Windows Excel 2000では、"Stat macros.xla" でソフト91kbの登録が可能で、自分のExelの分析ツールの下に Nonparametric Statとして使用可能。例題がついており、バイオサイエンスの統計学の例題の数値を入力して、操作と表の見方を検討することが必要である。
		正規性検定・変数変換マクロ Data Transformation for Excel. 廣田忠雄 (国際キリスト大)		http://www.tuat.ac.jp/~ethology/Columbo/Stat/Data_Trans/	基本3ソフトの1つ。検定をする前に、基本統計処理の時点で、検出力・再現性の為に、常に、検討しなければならない基本ソフト。
		統計解析のためのプログラム		http://hoel.med.kyushu-u.ac.jp/~nutahara/stat/	
		生物統計のためのプログラム (Java アプレット&サープレット)		http://chiryo.phar.nagoya-cu.ac.jp/javastat/JavaStat-j.htm	
	有料汎用統計パッケージ	SAS シリーズ	株式会社 SAS Institute Japan, 年間契約レンタル50万円以上	http://www.sas.com/	大学・研究所・企業用、最高クラス、英語仕様
		JMP 4	SAS Institute Japan, 125,000円	http://www.impdiscvery.co/japan/	教材用良いとこ取りセット
		Dr. SPSS II for Windows 日本語版. CD-R (Win)	SPSS Japan, 南江堂, 98,000円	http://www.spss.co.jp/	基本ソフト、勧めの1つ、教材用良いとこ取りセット
		SPSS 10.0	SPSS Japan, 168,000円	http://www.spss.co.jp/	個人用として最も一般用の1つ。
		Statistica 2000	StatSoft Japan, 129,500円	http://www.statsoft.co.jp/	

統計解析における統計手法の選択方法 ―高いお金を出して統計ソフトを買わなくても、
ホームページを使って統計処理ができる―

分類		商用ソフトウェア	開発元・販売元 (賛助会員・ あいうえお順)	ホームページ・アドレス	コ メ ン ト
		ASSUM	日本電子計算 1,000,000円 より	http://www.jip.co.jp/index.html	
		S version 4	ISAC, Inc. , 180,000円より	http://www.isac.co.jp/	
		S-PLUS	株式会社数理シ ステム、Insight- ful (S-PLUS の 開発元)、 320,000円より	http://www.msico.jp/splus/	
		SYSTAT 10.2	株式会社 ヒューリンクス 198,000円	http://www.hulinks.co.jp/	
		JMP 5	株式会社 ヒューリンクス 135,000円	http://www.hulinks.co.jp/	
		Halwin Halbau	現代数学社 75,000円	http://www.gensu.co.jp/halwin.html	国産、安価、良いとこ取りセット。日本語仕様 がお勧めの理由。
		臨床統計解析のための ソフトウェア FISHER for Win- dows. 開発 東京大 学医学部コンピュ ータ医療研究会 (指導 開原成充、大江和 彦)	中山書店 95,000円		基本ソフト、解説書が丁寧で順序だっている。 Excel で物足りないと感じた時のお勧めの CD。愛 国者向き。残念ながら多変量解析が無い。
		統計解析ハンドブック for Win 多変量 解 析 Ver. 2.0, Windows 版. 開 発 垂水共之、田中豊	共立出版 65,000円		多変量解析のみ。必ず別売りの解説書 (参考書) を購入すること！本の例題を見ながら操作だけ は直ぐにできる。蛇足ながら、掲載された理論 数式を理解しようなどと決して思わないほうが 良い。
	有 料 Excel アドイン可 能ソフト	エクセル統計2000、 秀 吉2000、秀 吉 Pro 2000、即断2000	エスミ 130,000円	http://www.esumi.co.jp/	Excel に慣れた人はアドイン仕様がお勧め！
		エクセル統計解析 Ver 5	エスミ 38,000円	http://www.esumi.co.jp/	Excel に慣れた人はアドイン仕様がお勧め！基本 統計ソフトで、多変量統計ソフトを追加購入す ることが必要。
		エクセル多変量解析 Ver 4	エスミ 38,000円	http://www.esumi.co.jp/	Excel に慣れた人はアドイン仕様がお勧め！多変 量統計ソフトのみで、基本統計ソフトを追加購 入することが必要。
		エクセル統計2000	社会情報サービス 32,000円	http://www.ssri.com/	Excel に慣れた人はアドイン仕様がお勧め！
		Startartner, Ver 4	オーハ 44,800円	http://www.o-ha.co.jp/package/statpartner.html	Excel に慣れた人はアドイン仕様がお勧め！
文献		Excel 関数活用 Super Master、工藤雅 俊著	エクスメディア 出版 1999. 3,000円		書店から買ってきたら、その時から、理解せず に、見ながら反射的に「分析ツール」について のみの操作だけではできるアンチョコ。
		Excelでわかる統計入門、Windows対応サン プルデータマクロ プログラム収録CD -ROM付、清水理著	ナツメ 2000. 2,300円		データ入力が必要ないように CD が添付され、今 様のバーチャル仕様となっている。
		文系の学生のための Excelによる統計解 析、二宮智子・恵良 和代著	弘学出版 2000. 1,900円		抜群にわかり易い！見開き4頁1事項となっ ており、表を参考にして、すぐに Excel が使える。 データのまとめ方が判りやすい。できれば、超 図解 Excel 基礎編と応用編 (エクスメディア、 各巻1,300円) を買い揃えた方がよいが。

分類		商用ソフトウェア	開発元・販売元 (賛助会員・ あいうえお順)	ホームページ・アドレス	コ メ ン ト
		Excel によるデータ解析. 成田滋・筱更治 (しのこうじ) 著	エコノミスト 2001. 3,600円		ノンパラメトリックについての操作時に使用。 ただし、分析ツールと異なり、1つの検定において、その項の「Excel による検定」を参考にし て、自分で各セルに Excel 式を挿入してゆかなければならない・・・が、操作は比較的容易で、 子供の頃のプラモデル作製を彷彿とさせる。注 意：誤字、索引の頁のミスプリントが多いので、 怒って壁に投げつけないこと。
		バイオサイエンスの 統計学. 市原清志著	南江堂 1990. 4,800円		グラフによる説明が多い。わかり易さ抜群。Excel 分析ツールと Excel 関数活用 Super Master と 併せて、廉価な統計の三種の神器
		医学統計データを読む、医学・医療に必 要な統計学活用法 (Basic & clinical biostatistics), 第 2 版、森田茂穂監訳 (eds. Dawson-Saun- ders,B & Trapp, RG)	メディカル・サ イエンス・イン ターナショナル 1997. 3,800円		翻訳が忠実直訳の最悪で5回読んでも理解でき ない最悪の本であるが、困ると必ず読み返し、 ゴミ箱に捨てられないホームタウンの様な教科 書。また、個人用に Kaplan-Meier 生存率曲線の 計算式表を Excel で作成するのに役立った。
		実践統計学入門. 足立堅一著	篠原出版.東京. 2001. 2,800円		3 回読んだが良くわからない。
		EBM 実践のための統 計学的 Q&A. 足立堅 一訳	篠原出版.東京. 2002. 3,200円		(Statistical Questions in Evidence-based Medi- cine. eds. Martin Bland, Janet Peacock. Oxford Uni- versity Press.2000.) (検定手技に漸くなれた頃、 実験計画・解析の難解さの本体を垣間見せてく れる本。身近に相談できる本物の統計専門家が いないことに漠然とした不安を抱かせる。残念 ながら、姉妹本 An Introduction to Medical Statis- tics,3rd. ed. Bland.2000.はまだ訳本が出版されて ません。)
		多変量解析のはなし. 大村平著	日科技連出版 1985. 1,550円		横になって読める唯一の入門書。
		エクセル実用医学統 計テキスト スタッ トメイト・ミニ 高 橋泰生(天理よろず)	ATMS, 4,800円		StatMate III ソフト ミニ版添付
		すぐわかる統計処 理、石村貞夫著	東京図書 1994. 2,200円		医療データ、数式なし、安直。
		すぐわかる統計用 語、石村貞夫著	東京図書 1997. 2,800円		統計用語の解説書。
		医師のためのロジス チック・Cox 回帰入 門、高橋善弥太	日本医学館 1995. 3,000円		多変量解析の理解用。
		らくらく生物統計 学、足立堅一著	中山書店 3,200円		基礎統計学の理解用。
		オンコロジストのた めの統計学。大橋康 雄著,2000.			趣味の範疇。国立がんセンターで研修できない 人の自習用。
		誰も教えてくれな かった因子分析-数 式が絶対に出てこ ない因子分析入門-、 松尾太加志、中村知 靖著	北大路書房 2002, 2,500円		SPSS, SAS 使用目的。確かに、数式は出てきま せん。3 回読むと、操作はわかるようになるが、 因子分析は理解できない。
		Crowly J. Handbook of Statistics in Oncol- ogy, Marcel,2001			本棚を飾るのに最適。
	ホームペー ジ	探索的因子分析リン ク集 (日本語中心)、 香川大経済学部、堀 啓造		http://www.ec.kagawa-u.ac.jp/~hori/spss/factor-link.html#kaisetsu	

統計解析における統計手法の選択方法 ―高いお金を出して統計ソフトを買わなくても、
ホームページを使って統計処理ができる―

分類		商用ソフトウェア	開発元・販売元 (賛助会員・ あいうえお順)	ホームページ・アドレス	コ メ ン ト
		統計学と HALBAU/ HALWIN の世界、わ かる統計学、簡単統 計計算		http://www.clg.niigata-u.ac.jp/~takagi/takagi.html	わかる統計学で勉強し、簡単統計計算に数個データを入力すると、暗箱より、ボトリと答えだけが出てくる。
		Welcome to Nekoda's Homepage, 保健医療 領域におけるデー タ解析支援システム、 東京都立保健科学大 学保健科学部看護学 科猫田泰敏		http://www.metro-hs.ac.jp/~nekoda/index.html	実用的。わかり易い。医師用で無く、看護師用だからこそ、わかり易い。
		データ解析にかかわ る基本的事項を確認 したい、東京都立保 健科学大学保健科学 部看護学科猫田泰敏		http://www.metro-hs.ac.jp/~nekoda/project/kihon.html	実用的。わかり易い。医師用で無く、看護師用だからこそ、わかり易い。
		統計手法アラカル ト：ノンパラメトリ ック法、大阪大林 賢一		http://koko15.hus.osaka-u.ac.jp/~hayashi/study/pdf/npar.pdf	ノンパラメトリック検定全般の解説集
		医学統計学（神奈川 歯科大学、森貫敏夫、 医用統計・保健統計 学の講義ノート）		http://www.kdcnet.ac.jp/toukei/index.htm	講義として聞くのには最高、お奨め品
		看護研究		http://www.geocities.jp/kazu_hiro/nurse/research.htm	判りやすいオリエンテーションで、一読することをお勧めします。
		ナースに役立つ種類 のサイトとは？ Nurse's Soul selection of useful links		http://www.geocities.jp/kazu_hiro/index.html	わかり易い。常に更新してくれるので役に立つ。流行語を噛み砕いて説明している。
		統計の初歩（兵庫教 育大学、成田滋、心 理・教育のための統 計の初歩、SAS、 SPSS、Excel による 統計解析の実際など		http://grace.ceser.hyogo-u.ac.jp/sta/index.html	少数精鋭の項目ながら、それぞれがわかり易い。
		医療統計学、茨城県 立こども病院、新井 順一、医師のための 統計学、SPSS Ver. 9 対応		http://www.h5.dion.ne.jp/~ge3j-ari/stat/mokuji.html	統計解析の総論的教科書。わかり易さでは 1 番。
		WWW で統計を学習 しよう（統計に関す るリンク集）香川大、 堀啓造		http://www.ec.kagawa-u.ac.jp/~hori/statedu.html	網羅的で細かすぎる。困ったときに開くとヒントが見つかる
		EBM のための臨床疫 学入門講座、日本大 公衆衛生学教室		http://www.med.nihon-u.ac.jp/departement/public_health/ebm/index.html	わかり易い。
		心を測る、日本女子 大人間社会学部心理 学科、岡本安晴		http://www.ikuta.jwu.ac.jp/~yokamoto/openwww/	ソフトを使用する為に必然的に目を通さなければならぬ。わかり易い。数式つき。
		実験計画法、島根大 学生物資源科学部作 物学研究室、小林和 弘		http://homepage2.nifty.com/crop_shimane-u/sub4.htm	説明とその Excel 操作が対になっておりわかり易い。分散分析はわかり易く書けている。
		ノンパラメトリック 法、大阪大、林賢一		http://koko15.hus.osaka-u.ac.jp/~hayashi/study/pdf/npar.pdf	SPSS 使用目的、ノンパラメトリック索引一覧とパワーポイント形式の簡易解説。廣田のノンパラメトリック一覧表と対照のこと。
		心理データ解析、虫 部大学人文学部心理 学科、小塩真司		http://psy.isc.chubu.ac.jp/~oshiolab/teaching_folder/datakaiseki_folder/top_kaiseki.html	SPSS 使用の解説であるが、わかり易いので、一読することが薦められる。分散分析は必読。共分散構造分析・Amos・パス解析。
		パス解析、奈良大社 会学部社会統計学問 淵領吾		http://www.nara-u.ac.jp/soc/staffs/mabuchi/lectures/path_analysis.html	パス解析

表2. 無料ソフト集、搭載レシピの内容一覧

[illegible]

表 3. 研究デザイン study design による証明力の強さの順番

Agency for Health Care Policy and Research (AHCPR) 分類に基づく、evidence based management (EBM) の為の Level Ia より IV までの分類 (精度順)

実験的研究 experimental study、介入研究 intervention study

Ia 少数複数の大規模な無作為割付臨床試験 (無作為対照化試験、無作為割付比較試験、ランダム化対照試験、randomized controlled trial (RCT)), メタアナリシス

IIa よくデザインされた無作為割付のない比較試験 nonrandomized controlled trial

IIb よくデザインされた非実験的研究

観察的研究 observational study

III 前向き研究 prospective study: (前向き) コホート研究 cohort study、コホート内症例対照研究

III 地域相関研究、後向き研究 retrospective study: 症例対照研究 case control study (retrospective)、横断研究、行政における時系列研究

IV 専門委員会の報告・意見、記述的研究 descriptive study、権威者の臨床経験: 専門家の見解 expert opinion 「権威者の経験談は、1 篇の研究論文に勝てない」と言い切られている!

表 4. 作業工程: 研究計画の立て方と検定のフローチャート

ステップ	内容と注意事項
1	介入研究 (実験計画) または観察研究 (横断研究 Crosssectional、前向き研究 Cohort、後向き研究 Case-Control) を立案(1)
2	偏り無くデータを収集(2)
3	データの分布型図示化 (ヒストグラム、分散図) し、実質的に (感覚的に) 有意であることの確認を感覚的に把握する、発表時に説得力が出る(3)
4	変数(4)
5	正規 (分布) 性の検定 (パラメトリック、ノンパラメトリックの区別) (5)
6	ITT(6)
7	検定法を選択
8	鑑別: 有意差判定(7)、類似性・因果: 相関・連関(8)
9	意義の確認: 有意差には、1. 真の有意差・関連(9)、2. バイアス(10)、3. 交絡(11)、4. 偶然 (誤差) の可能性があり、類似性には擬似相関(12)、因果連鎖(13)、逆向因果(14)の可能性もある。更に、現実的な意義(15)が考察されなければならない。

注意

- 避けなければならないこと: バイアス、交絡因子。
方法論: A. 実験的研究: 1. 無作為化、2. メタアナリシス、B. 観察的研究: 1. コホート、2. ケース・コントロール、3. 横断研究。
エンドポイントを選択する。
必要なサンプルサイズを計算。
- 実験手技の偏りを避ける。実施者の個人的判断を避ける。欠落値による偏りを避ける。
- 作図: 度数分布図、ヒストグラム (密度関数)、箱ひげ図、正規確率紙、相関・回帰、診断精度に関する ROC 曲線は必須、(外れ値に対するスミルノフの棄却検定)。
- 変数: 表 2 参照
- 統計は、正規性を原則としたパラメトリック解析であることをまず念頭に置く。パラメトリック検定 parametric test: 正規分布、間隔尺度を前提とする。ノンパラメトリック検定 non-parametric test: 正規分布を前提とする必要が無い、分布に依存しない検定 distribution free test。但し、基本的には、検出力の程度の差であって、パラメトリックで代用も OK。
サンプル数が 25 以上と多いと、中心極限定理により、正規分布に近似する。
正規性の検定: 確率紙、(分散比の) F 検定、(分散それ自身の) χ^2 乗検定、(図示としての) 左

右対称性マーカーの歪度 (正規性 = 0 ± 0.5) や裾広がり度 = 尖がり度マーカーの尖度 (正規性 = 3 ± 0.5)、変数変換後の 正規性近似 (Excel 乱数発生、変数変換マクロ Data Transformation for Excel (廣田忠雄))。注意: 対数変換: 値 = 0 の変数は、(次に大きい変数) $\times 1/2$ として、対数変換すること。

分布図: z: (標準) 正規分布

分布図: t: 少ないデータ数からの統計的検定法上、平均値の検定用修正分布

分布図: χ^2 , F: 少ないデータ数からの統計的検定法上、分散の検定用修正分布

- 生存率、相関、検定 (独立、関連)、多変量解析より選択。変更・脱落・打ち切り症例は群間の比較可能性を維持するために本来的 original intention to treat (ITT、本来の治療意図) 定義に従い、当初割り付け群で扱うこと!

1. 介入研究 (実験計画): 分散分析、2. 観察研究 (横断研究 Crosssectional、前向き研究 Cohort: リスク比、層別解析 + Mantel-Haenszel、リスク比 + ロジスティック回帰、重回帰、Cox 回帰などによる交絡調整、後向き研究 Case-Control: オッズ比、Mantel-Haenszel、ロジスティック回帰などによる交絡調整)

- χ^2 分布と F 分布は上側確率 (右片側検定) のみであるが、正規分布と t 分布は両側確率 (両側検定と片側検定の両方が可能) である。

両側検定は2群が異なるか、片側検定は片方が大きいのかまたは小さいかを調べる。有意水準 $\alpha=5\%$ で、有意差棄却域境界値は、両側検定で下限0.06、上限1.96であり、片側検定で下限0.03、上限1.64と異なる。すなわち、有意水準5%において、両側検定は正規分布グラフの両端の和が5%で、片側検定はグラフの左側または右側の片側のみで5%である。

結果的には、両側検定がより厳密な検定である。なぜならば、群内には消去不能な誤差的バラつきがあるからである。しかしながら、厳密ならば良いというものでもなく、実験計画の期待される精度を考慮して判断されるべきである。

図示説明： 参考： 表7

- 8 検定は基本的に（片側検定より厳密な）両側検定とし、F分布と χ^2 分布は片側確率。データ数が少ない時は偶然が否定できない。逆に、症例数が多くなると、実用的な有意差を認めないものでも統計的有意差が出現しやすくなることに注意。

相関：一般に、 $r \leq 0.2$ は相関無し、 ≤ 0.4 は軽度、 ≤ 0.7 は中等度、 $0.7 < r$ は強い相関とする。

- 9 サンプル数が多くなると正規分布となり、実用的でない微小な差でも統計学的には顕著な有意差を示しうる。統計学的有意差が、実質的な意義を示さないこともあるので注意すべきである。

- 10 精度の向上：精度は、好ましくないことをいかに回避できたかに依り、好ましくない事象はバイアスと交絡因子で代表される：

バイアス, bias: 偏り (系統誤差) のことで、因果関係 $X \rightarrow Y$ において、原因変数であるデータ X における人為的な誤謬のことである。原因として、大きくは、A. 対象の選択の方法 (選択バイアス) と、B. 情報の収集・ならびに評価における研究者の思い入れ等 (測定バイアス) の二種類が指摘されている。A. 選択バイアスは、以下を含む：1. 標本抽出 (全構成員からの無作為抽出でない)、2. 自己選択 (自発的な参加と不参加の特性の差)、3. 脱落 (持病の発生と関連した脱落)、4. Neyman (早期曝露の後向き研究で、早期死亡・軽症・無症状が脱落し、罹患率と有病率が異なる)、5. 持続性 (有病患者を対象とした時、片方の群に長期生存例が多い)、6. 時間差 (早期診断可能と不可能の差)、6. 所属集団 (対照集団の健康意識の差)、7. Berkson (カルテ検討では無作為で無く患者の意思の受診の反映)。B. 測定バイアスは、以下を含む：1. 情報 (調査精度が異なる)、2. 観察者 (観察者の思い入れ)、3. 診断 (診断基準の不安定)、4. 数字 (好みの順序変数にまとめられた時)、5. 想起 (健康意識度により被曝記憶がうる覚えとなる)、6. 発見 (確

の選択基準・診断手技が相違)、7. 思い巡らし (症例群は自己の疾患に深い考察があり、対照群の思い出し方が異なる)、8. 記憶上 (症例群は質問を反復されていることが多く、対照とは質問回数が増え、思い出し方が異なる)、9. 家族歴 (対照は記憶が少ない)、10. 報告 (性情報は隠蔽されたり表面的になりやすい)、11. 追従 (調査者に迎合)、12. 面接者 (面接者の一方的選択)、13. 気遣い (症例群に対して、ケースの悩みを過剰に判定)、14. 非認容 (話し辛いことの拒否・はぐらかし)、15. 要因予知 (患者の症状に関する予備知識がある)。

これらのバイアスに対する対策は、原因変数への研究者が意図しない剰余変数の影響を統制 (blocking) することである。選択バイアスには、恒常化 (剰余変数を統制する為、対象者を限定したり、マッチングしたりすること) と無作為化 (剰余変数をランダムな確率に任せる RCT) を実施しなければならない。鑑別すべきものに、誤差的 (偶然的) バラツキがある。すなわち、本来何か理由があってバラツいた結果となるのであるが、現時点ではその原因を確定できない状態の結果のバラツキである。例えば、同様の抗癌剤治療検討において、A 施設は有効率100%、B 施設は有効率50%、C 施設は有効率20%であるという結果がでて、それらの平均値は真値に近く、この結果は正確である。このバラツキは、転移部位の相違や PS 不良症例の混入数により予後が不良となったのであろう。このバラツキの減少 (精密さ) は、転移部位を一定にしたり、PS 良好に限定するしかなく、実験計画の段階が問題であったので、一般には、症例数を多くすることが勧められる。

- 11 交絡因子 confounding factor: 因果関係 $X \rightarrow Y$ において、 $Z \rightarrow X$ かつ $Z \rightarrow Y$ の作用をする Z 因子のことで、 $X \rightarrow Y$ の因果を交絡 (ごちゃごちゃに混乱) させる。交絡因子の難しさは、交絡因子 Z の存在を知っていないと解決できない点であり、永遠に解消される可能性は無い。交絡を疑う要因は、性別、年齢、無作為割付以外の方法での群の割付、群分けの基準がある。交絡の防止には、介入的研究においては Fisher の (技術評価のための) 実験計画法における厳密な無作為化割付が重要であり、また、解析時における層別化 (亜分類化) + Mantel-Haenszel 検定、(条件付) ロジスティック回帰分析、重回帰分析、Cox 回帰など多変量解析によって調整することが妥当とされる所以である。

- 12 相関 $A \leftrightarrow B$ において、 $X \Rightarrow A$ かつ $X \Rightarrow B$

- 13 因果 $A \Rightarrow B$ において、 $A \Rightarrow X \Rightarrow Y \Rightarrow Z \Rightarrow B$

- 14 実際の因果が $A \Rightarrow B$ において、 $B \Rightarrow A$ と誤認

- 15 その結果が一般化できるか、または、その統計的判断が診療において個別の症例に応用できるか否かは難しい問題である。

表 5. 疫学確率

発症率 (%), 事前有病率		20
A スクリーニング検査 (%)	感度	34
	特異度	86

感度 = $\{(\text{疾患}+) \text{ で } (A \text{ 検査}+)\} / (\text{疾患}+) \text{ 合計}$
 特異度 = $\{(\text{疾患}-) \text{ で } (A \text{ 検査}-)\} / (\text{疾患}-) \text{ 合計}$

↓
2 × 2 表作成の為に実数を計算する

TD「(疾患+) で (A 検査+)」= (疾患+) × 感度
 tD「(疾患+) で (A 検査-)」= (疾患+) × (1 - 感度)
 Td「(疾患-) で (A 検査+)」= $\{1 - (\text{疾患}+) \times (1 - \text{特異度})\}$
 td「(疾患-) で (A 検査-)」= $\{1 - (\text{疾患}+)\} \times \text{特異度}$

↓
表を作成

男

	A 検査- (t) 曝露- (e)	A 検査+ (T) 曝露+ (E)	合計
疾患+ (D)	13.2	6.8	20
疾患- (d)	68.8	11.2	80
合計	82	18	100

女

	曝露- (e)	曝露+ (E)	
疾患+ (D)	10	5	15
疾患- (d)	50	8	58
合計	60	13	73

↓
統計量の計算

RCT、コホート研究：

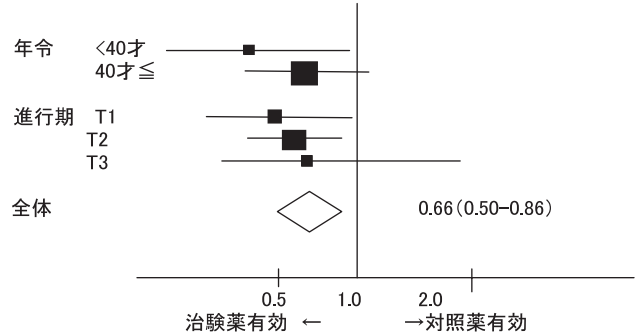
曝露群のリスク (累積罹患率)	DE/E	0.377777778	
非曝露群のリスク (累積罹患率)	De/e	0.16097561	
曝露の発症に対するリスク比、累積罹患率比、 相対危険：個人の危険度に使用、ハザード比 (打ち切り例のある場合、ある 1 時点での 2 群間の死亡確率比で、D=death と読み換えると相対危険と類似する) ▼	(DE/E) / (De/e)	2.346801347	95%信頼区間：m±1.96 SE で、1 を含まない場合 5 % の危険率で有意
曝露の発症に対するリスク差	DE/E - De/e	0.216802168	
寄与危険：集団の危険度に使用	(DE/E) - (De/e)	0.216802168	
* 多因子：重回帰、多重ロジスティック、対数線形			
* 潜伏期脱落例有り：Cox 回帰 (比例ハザードモデル)			

症例対照研究、断面研究：

* オッズ、患者群の曝露オッズ	DE/De	0.515151515	
* オッズ、対照群の曝露オッズ	dE/de	0.162790698	
* オッズ比 (相対危険、罹患率比の近似値として)、曝露オッズ比	(DE/de) / (dE/De), (DE/De) / (dE/de)、セルの値が 0 のある場合、すべてのセルに 0.5 を加算する修正を行う	3.164502165	95%信頼区間：m±1.96 SE で、1 を含まない場合 5 % の危険率で有意
* 対数オッズ比の分散	$1 / DE + 1 / De + 1 / dE + 1 / de$	0.326636997	
* 男女共通オッズ比	$\{(\text{男 } DE \times \text{男 } de) / \Sigma \text{男} + (\text{女 } DE \times \text{女 } de) / \Sigma \text{女}\} / \{(\text{男 } dE \times \text{男 } dE) / \Sigma \text{男} + (\text{女 } dE \times \text{女 } dE) / \Sigma \text{女}\}$	3.147685863	

事前確率、有病率 prevalence	D/Σ	0.2
事前確率、無病率	d/Σ	0.8
事後確率としての陽性反応適中度◆, PPV, positive predictive value、「問診時に疑われた病名が正しい診断である場合に、陽性検査結果の期待できる確率」	DT/T	0.37777778
事後確率としての偽陰性反応適中度, 「問診時に疑われた病名が正しい診断である場合に、陰性検査結果である確率」	Dt/t	0.16097561
*尤度グラフより（下図）		
事後確率としての偽陽性反応適中度, 「問診時に否定的とされた病名がまさしく妥当である場合に、陽性検査結果である確率」	dT/T	0.62222222
事後確率としての陰性反応適中度, NPV, negative predictive value、「問診時に否定的とされた病名がまさしく妥当である場合に、陰性検査結果の期待できる確率」	dt/t	0.83902439
陽性率 test positive rate、要精査率	T/Σ	0.18
正診率、有効度 efficiency	(DT+dt) / Σ	0.756
感度 sensitivity、有病正診率	DT/D	0.34
特異度 specificity、無病正診率	dt/d	0.86
診断効率	感度 x 特異度	0.2924
第1種過誤★、α 過誤、偽陽性率、無病誤診率	dT/d	0.14
	1－特異度	0.14
第2種過誤★、β 過誤、偽陰性率、有病誤診率	Dt/D	0.66
	1－感度	0.66
検査前オッズ	事前確率／（1－事前確率）	0.25
検査後オッズ	検査前オッズ x 尤度比	0.607142857
陽性尤度比	感度／偽陽性率	2.428571429
判定：<0.1：除外診断に利用、0.1～10：無効、10<確定診断に利用		
陰性尤度比	特異度／偽陰性率	1.303030303

▼ hazard 比
「治療薬と対照薬による再発の比較」のシェーマ



∴判断：1未満（1を含まない）ので、治療薬の方が有効
ひげは95%CI
メタアナリシスでも可（例、The Cochran Library）

◆：グラフから求める：事前確率→尤度比→陽性適中率

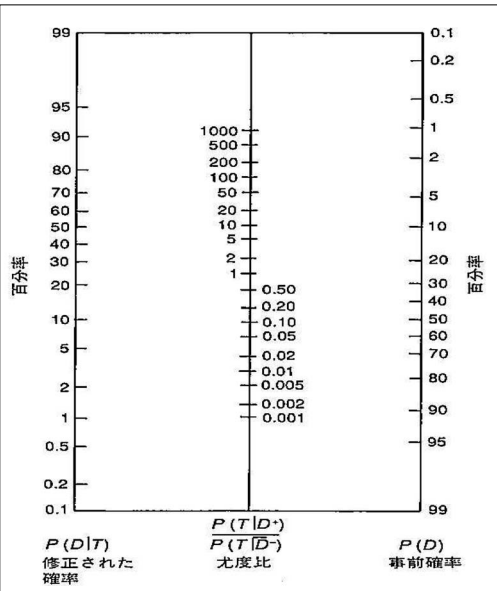


図7-11 Bayes 理論を使用する際のノモグラム
Fagan TJ:Nomogram for Bayes theorem
(Letter). *N Engl J Med* 1975 ; 293 : 257より許可を得て改変。

- ★ 第1種の α 過誤：
検査では、偽に対する検査の偽陽性のこと。
検定では、差がない（帰無仮説 H_0 ）を異なる（対立仮説 H_1 ）とする誤り。
- 第2種の β 過誤：
検査では、真に対する検査の偽陰性のこと。
検定では、差がある（対立仮説 H_1 ）を差がない（帰無仮説 H_0 ）とする誤り。

表 6．変数

(1) 変数の呼称

応用	因、x	果、y
多変量解析	説明変数	基準変数
因果モデル	原因	結果
生産デザイン	仕様、スベック	総合評価
実験計画	因子	観測データ
数量化理論	要因	外的基準
数学	独立変数	従属変数
経済	外生変数	内生変数
多属性決定モデル	属性、アトリビュート	選好

(2) 変数変換

正規性の検定
変数変換マクロ：乱数発生ツール (Excel)、Data Transformation for Excel（廣田忠雄）
注意：対数変換：値＝0の変数は、（次に大きい変数） $\times 1/2$ として、対数変換すること。

(3) ノンパラメトリック検定（分布によらない検定）の順位検定

タイのある大小だけの順位データの順位のつけ方：
応用： Wilcoxon 順位和検定、Kruskal-Wallis 検定

例： 元データ

	治験薬 (1)	対照薬 (2)	合計	累積
著効	4	0	4	4
有効	3	2	5	9
不変	2	4	6	15
悪化	1	6	7	22
合計	10	12	22	

処理(1) 順位の振り方：

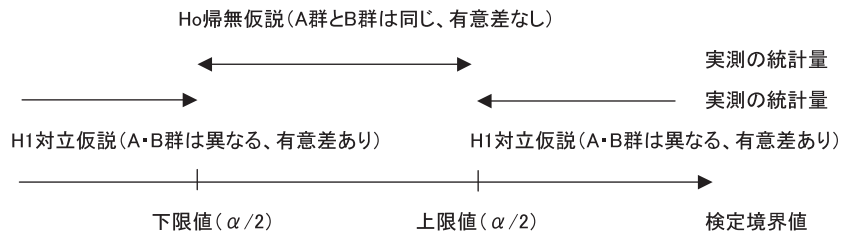
著効の中間順位 $(4 + 1) / 2 = 2.5$
有効の中間順位 $4 + (5 + 1) / 2 = 7$
不変の中間順位 $9 + (6 + 1) / 2 = 12.5$
悪化の中間順位 $15 + (7 + 1) / 2 = 19$

処理(2) 実際のデータ入力：

グループ変数の指定：症例群	分析用変数：治療効果の順位
1	2.5
1	2.5
1	2.5
1	2.5
1	7
1	7
1	7
1	12.5
1	12.5
1	19
2	7
2	7
2	12.5
2	12.5
2	12.5
2	12.5
2	19
2	19
2	19
2	19
2	19
2	19
2	19

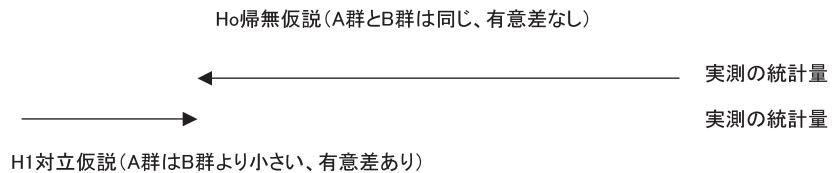
表 7. 両側検定、片側検定

両側検定

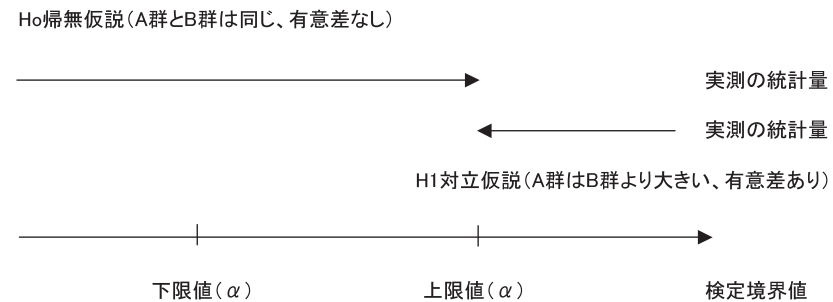


片側検定

左片側検定



右片側検定



- ・両側検定・・・ $\mu A \neq \mu B$ を見る場合 (変化したかどうか、または違うかどうかを見る場合)
- ・片側検定・・・ $\mu A > \mu B$ などを見る場合 (単に大小関係を見る場合) (両側検定よりも甘い検定)
- * 明らかにどちらかが大きいと判断される場合のみ片側検定を使う方がよい

表 8. 関係（相関と関連）

2 変数	両方が量的データのみ	まず最初に検討すべきこと	基礎統計量 ヒストグラム 散布図			
		検定	パラメトリック（正規性）	直線検討	Peason 積率相関係数（単相関係数） ＋無相関検定＋Fisher z 変換による母相関係数の信頼区間の推定	
				予測	回帰分析	
				曲線、単調関係	順位データは Spearman 順位相関係数、大小関係のみのデータは Kendall 順位相関係数（タイは修正する）、 ＋無相関検定＋Fisher z 変換による母相関係数の信頼区間の推定	
		ノンパラメトリック（正規性で無い）	曲線、単調関係	順位データは Spearman 順位相関係数、大小関係のみのデータは Kendall 順位相関係数（タイは修正する）、 ＋無相関検定＋Fisher z 変換による母相関係数の信頼区間の推定		
	1 方が量的データで、他方が質的データ	質的データの各カテゴリーを群と考え				群間の比較を参照実施
		量的データをカテゴリーし、順序尺度と考え				両者とも質的データとしての群間の比較を参照実施
	両方とも質的データのみ	順序尺度	順位データは Spearman 順位相関係数、大小関係のみのデータは Kendall 順位相関係数（タイは修正する）、 ＋無相関検定＋Fisher z 変換による母相関係数の信頼区間の推定			
			名義尺度			χ^2 検定、残差分析
		2 x 2 クロス表			ϕ 係数	
一方が順序尺度でカテゴリー数が 3 以上		群間の比較で、質的データと比較したいを参照				
1 つの目的変数と複数の説明変数	目的変数	量的データ	説明変数	量的データ	重回帰分析	
				質的データ混在	数量化 1 類	
		質的データ		質的データ	数量化 2 類	
				量的データ混在	Mantel－Haenszel、条件付多重 Logistic モデル、判別分析、数量化 2 類	
				交絡因子を調整	例： 変数（曝露など）と、カテゴリー化された交絡因子（素因あり：1、素因なし：0）除外による、カテゴリー化された目的変数（発症：1、未発症：0）への影響を検討	Logistic 回帰
				交絡因子を調整（除外）して、相関を調べる	偏相関係数	
		交絡因子をそのまま加えて、分散分析する		共分散分析		
複数の変数間の関連を一度に観察したい	量的データ		量的データ	正準相関分析		
	複数の変数を少数の複合指標に縮約				主成分分析、数量化 3 類	
	変数に潜む少数の因子の構造を把握				因子分析、潜在構造分析	
	変数を似たもの同士に分類				主成分分析、Cluster 分析、数量化 3 類、4 類	
複数の変数間の因果モデル	変数の因果関係や相互関係を図示				共分散構造分析：多変量解析（重回帰分析、因子分析、判別分析）による Amos プログラム（SPSS ソフト）によるパス図（path diagram）描写	

注意

相関
関連・連関
相関関係の強さ

correlation：変数が間隔変数・順序変数の場合の呼称。
association：変数が分類尺度の場合の呼称。
相関係数
0.2～0.4
0.4～0.7
0.7～1.0
軽度
中等度
高度

統計解析における統計手法の選択方法 —高いお金を出して統計ソフトを買わなくても、
ホームページを使って統計処理ができる—

表 9. 変数から見た検定法の選択

y	x	数	選択すべき検定法
無	量	1	平均値、分散
		2	相関係数
		3 ≤	Cluster 分析、主成分分析、因子分析、数量化 4 類
	質	1	比率
		2	クロス表（独立性の検定）
		3 ≤	数量化 3 類
量	量	1	単回帰分析（相関係数）
		複数	重回帰分析、正準相関分析
	質	1	一元性分散分析、差の検定
		複数	数量化 1 類
質	量	1	差の検定、分散分析
		複数	判別分析、多重 Logistic モデル、条件付多重 Logistic モデル、Cox 比例 hazard モデル
	質	1	クロス表（一様性の検定）
		複数	数量化 2 類、Logistic モデル

注意： x： 独立変数、説明変数
y： 従属変数、基準変数