6. 文献にみる統計解析

文献には統計解析した結果をまとめた表が数多く記載されている。その一例から有意差検定を確かめてみよう（下の表2）。また、半年間に提出され、査読されたある論文雑誌のうち、不採用となった24件の内訳が記載されている（下の表4）。データに関するもの、統計解析に関するものが多いことがわかる。

身体介護の支援活動	ホームヘルパーの経験年数 1年未満	1～3年	3～5年	5～7年	7年以上	計
なし	21	49	33	17	11	131
あり(ai)	4	41	105	84	34	268
計(ni)	25	90	138	101	45	399(N)
pi=ai/ni	0.16	0.456	0.761	0.832	0.756	0.6717(p)
スコア(Xi)	−2	−1	0	1	2	

表2　支援活動状況（%）

	合計(人)	要介護高齢者宅での主な支援活動内容 身体介護	家事一般	相談助言	服薬援助
年齢					
30～39歳	30	63.3	100.0	76.7	20.0
40～49歳	158	66.5	98.7	73.4	15.8
50～59歳	146	67.1	97.3	69.2	24.0
60歳以上	65	66.2	92.3	73.8	15.4
		ns	ns	ns	ns
経験年数					
1年未満	25	16.0	100.0	44.0	8.0
1年以上3年未満	90	45.6	100.0	77.8	7.8
3年以上5年未満	138	76.3	97.8	74.1	20.7
5年以上7年未満	101	83.0	98.0	76.0	28.0
7年以上	45	75.6	95.6	68.9	24.4
		***	ns	ns	**
受け持ち高齢者人数					
1人	44	59.1	93.2	61.4	18.2
2人	108	62.0	97.2	70.4	17.6
3人	142	69.7	97.2	68.3	16.9
4人以上	105	69.5	99.0	83.8	23.8
		ns	ns	*	ns
合計	399	66.4	97.2	72.2	19.0

有意差検定　***：$P<0.001$，**：$P<0.01$，*：$P<0.05$，ns：not significant

第47巻　日本公衛誌　第1号

公衛誌　第1号　　平成12年1月15日

表4　不採用となった主な理由（1999年4～10月審査分24件について）

1. 研究の意義・目的に関して
   - a) 普遍的な意義がない。
   - b) 評価法の紹介と介入研究のデザイン紹介にとどまっている。
   - c) 目的・仮説が不明確。(2)
   - d) 概念の定義や基準があいまい。(2)
   - e) 疾病の疫学についての知識不足。
   - f) 独創性が乏しい。（単なる実態調査あるいは教科書の紹介）(7)
   - g) 公衆衛生学的意義に乏しい。
2. データの収集，観察方法等に関して
   - a) データの質についての検証不足。
   - b) データの代表性に疑問。(2)
   - c) 例数が少ない。(5)
   - d) 研究方法・観察方法が不適当。(10)
   - e) 対照群の設定法が誤まっている。(4)
3. 統計解析方法に関して
   - a) 統計解析法が不適切。(8)
4. 研究結果の提示に関して
   - a) データの説明不足。(3)
   - b) 研究内容が不明確。(2)
   - c) 本文と図表との間の整合性が乏しい。(2)
   - d) 図表の出来が悪く枚数が多すぎる。(2)
5. 結果の解釈・考察に関して
   - a) 客観的根拠に乏しい。(3)
   - b) 論理の統一性欠如。
   - c) 他要因との交絡性についての検証不足。(5)
   - d) 解析結果の一般性に疑問。
   - e) 考察の論理展開不十分。(3)
   - f) 調査結果について考察していない。

（注）同様の指摘が2件以上あったものはカッコ内に件数を示した。

用語解説

資料：医学・生物学のための統計学，1977，和田　攻監修訳（Basic Medical Statistics, 1972）

母集団 population (universe)　情報を必要とする集団のすべて

標本（試料） sample　情報が正確に得られる母集団中の一部

偏り bias　ある母集団の値とその母集団から抽出した標本の対応する値との間の系統的な差．偏りというのは必ずしも偏見という意味でない．

均一母集団 homogeneous population　ある母集団をそれ以上グループ分けしなくても適切な分析がおこなわれうるような母集団

異質母集団 heterogeneous population　現在考察中の性状以外のある性状にもとづいて，実測中の事項に関して相互にかなり異なる亜群に分けられる母集団

質的（定性的，計数的）データ qualitative data　各個人が有するか有しないかどちらかの性状によって各個人を分類，計数したデータ．例えば，生・死などである．また，常に不連続であるデータである．

量的（定量的，計量的）データ quantitative data　身長，体重など測定することにより得られるデータ．典型例では連続的であるが，ときに不連続のデータとなる．

連続データ continuous data　連続体上の可能性ある無数の点を表わすデータ

離散（不連続）データ discrete data　不連続のデータ

推論 inference　標本について得られた値から母集団の値を推定すること．

任意（無作為）抽出法 random sampling　母集団の各要素に標本として抽出される平等の機会を与えられるような標本の抽出法

単純任意抽出法 simple random sampling　各要素に平等な抽出の可能性を与える抽出法

帰無仮説 null hypothesis　差が全くない，すなわち差は０であると仮定すること

有意差検定 significance test　帰無仮説の検定法．もし帰無仮説が否認されると，指定されたもう一つの仮定が容認される．

対照群 control group　実験処理の効果を評価するための比較群

機会 chance　できごとの偶発性．物事のおこり方

精度 precision　同一物をくり返し測定したときの一致性

正確さ accuracy (validity)　測定値が真の値と一致する厳密度

追跡調査 prospective study　ある因子の有無によって分けられた群について，その因子とある事象の発生頻度との関係を将来にわたって追跡していく調査法．例えば，ある病気の出現率など

履歴調査 retrospective study (case control study)　症例と対照を，過去のある因子について比較し，その因子と病気の発生の危険度との関係を決定する研究法

率 rate　$\dfrac{\text{ある期間内の事象（症例）数}}{\text{その事象の危険性にさらされている集団中の数}}$

標本誤差 sampling error　標本抽出過程での機会による誤差で，標本の結果と，同様の方法で行われた全数の結果との間にある差を示すことになる．

非標本誤差 non-sampling (systematic) error　統計調査を行う際の機会以外に基づく誤差．例えば測定の誤り，面接法，質問計画法，解釈などによる偏りで，この誤差は標本のみならず全数調査によってもみられるものである．

標本分布 sampling distribution　可能性あるすべての標本抽出による結果の度数分布で，各々の相対的な比率を示すもの．

確率分布 probability distribution　可能性あるすべての結果の確率を示す分布．曲線内のすべての部分を加えると<u>1</u>になる．標本分布と同じになり得るものである．

ある事象の確率 probability of an event　可能性あるすべての事象のうち，ある特定の事象の比率

比率 proportion　小数で表わされた率

二項型母集団 binomial population　互いに独立して排他的な二つのカテゴリーからなる母集団．例えばある特定の属性 a をもったものともたないものからなる集団．

二項分布 binomial distribution　二項型母集団から標本抽出を行ったときの可能性あるすべての結果の度数分布（確率分布）．

順列 permutation　項目（対象）を一定の順にしたがって選ぶこと．

組み合わせ combination　項目（対象）を順を無視して選ぶこと．

パラメーター（指標）parameter　ある母集団を特徴づける特定の象徴

同時（複合）確率 joint (compound) probability　二つないしそれ以上の事象 $a, b, c......n$ が同時に起きる確率

条件つき確率 conditional probability　事象 a が起きたと仮定したときの事象 b の確率

独立事象 independent events　ある事象の発生が他の事象の発生の確率と全く無関係な事象

試行 trial　特定数 n について実験したり，標本化すること

非対称分布 skewed distribution　片方に長く尾を引いた形の分布．

正の非対称 positively skewed　尾を右方（正の方向または大きい数の方向）へ引いたもの．

負の非対称 negatively skewed　尾を左方（負の方向または小さい数の方向）へ引いたもの．

均一（方形）分布 uniform (rectangular) distribution　すべての x 軸の値に対して同じ度数のもの．

パラメーター検定 parametric test　母集団がある特定の分布のパラメーターないしは形を有しているものと仮定して，そのパラメーターあるいは形を調べてゆく方法．

非パラメーター検定 non-parametric test　母集団の形やパラメーターを仮定しないで調べてゆく方法．

変換 transformation　データを数学的に操作してその幾何学的形態を変えること．

対数正規分布 lognormal distribution　$\log x$ をとることによって正規分布に直しうる非対称性分布．

集中傾向の表現 measure of central tendency　平均値

平均 mean　相加平均または算術平均

最頻値 mode　最も度数の多い値

中央値 median　低から高に並んだデータの中央の値

幾何学的平均 geometric mean　対数データの平均値の真数

度数 (f) frequency (f)　x の値（または x 番目のグループ）に属する要素（標本）の数．計算上重さを決定する因子でもある．

分散 variance　平均からの隔たり（偏差）を二乗したものの平均値で散らばりを表現する．分散 $\frac{\Sigma(x-\mu)^2}{N}$

標準偏差 standard deviation　分散の二乗根で，散らばりを表現する．標準偏差 $\sqrt{\frac{\Sigma(x-\mu)^2}{N}}$

分散係数 coefficient of variation　平均に対する標準偏差の大きさを％で表わした値．分散係数 $\frac{\sigma}{\mu}\times100$

階級（グループ）間隔 class interval　生のデータをグループ分けする際に任意に選ぶその間隔．

度数分布 frequency distribution　ある x の値に対する度数を表すようデータを並べかえたもの．

度数折線 frequency polygon　多面形をなし，x に対する度数を表わす点を直線で結んだもの．グループ分けをした資料については，各グループの中点の位置に度数を示す．

ヒストグラム histogram　グループ分けした資料を棒グラフで表わしたもの．等間隔に分けた場合，各グループの度数はその高さで，間隔はその幅で表わされる．

累積分布 cumulative distribution　大きさの順に並んだデータの累積和の分布をさす．累積百分率は，母集団中ある x 値以下の値をもつ部分の全体に対する割合を表している．

シグモイド曲線 sigmoid curve　S字状の曲線で，中央部で直線に近づき，両端は曲線である．これは，累積正規分布に特徴的である．

確率紙 probability paper　累積正規分布の曲線の上・下端をひきのばし，直線になるようにするグラフ用紙．

生物検定 bioassay　生きている生物の示す反応により，薬物その他の効力や性質を検定する方法．

用量-反応曲線 dose responce curve　各 log (用量) 以下に反応する百分率を示す曲線．この曲線は，ある薬物に反応する動物の数の増加の%とか，特定の動物や組織の示す反応の大きさの増加などを表わしている．

線の傾き slope of a line　x 値の単位変化 (横軸) に対する y 値の変化 (縦軸)

百分位数 percentiles　大きさの順に並んだデータを 100 等分した値．

十分位数 deciles　大きさの順に並んだデータを 10 等分した値．

四分位数 quartiles　大きさの順に並んだデータを 4 等分した値．

補間法 interpolation　中間の値を求める作業をいい，普通直線比例方法で行う．

多項型母集団 multinomial population　二つ以上の相互に排反する事象の母集団．

両側検定 two-tailed test　帰無仮説に対する選択が無方向性の差 (A≠B) である．それゆえ，偶然の起こる確率は，両側の，ある値より極端な方向の全領域を含む．

一側検定 one-tailed test　帰無仮説に対する選択がAがBよりも小さい (あるいは大きい) ということである．それゆえに，偶然の起こる確率は，一側の，ある値より極端な方向の領域のみとなる．

正規分布 normal distribution　二つのパラメーター，μ (平均値)，σ (標準偏差) による'つりがね状' (ガウス) 分布．その性質：

(1) 連続で対称的，両側が無限に延長している．(2) 算術平均，中央値，最頻値が同一である．(3) 特定な面積は μ からの距離 x (σ 単位での) によって計算できる．(4) μ から変曲点までの距離は σ で表される．

標準正規分布 standard normal distribution　平均 (μ) が 0，標準偏差 (σ) が 1 であるような正規分布

変数 z variate z　標準正規曲線の尺度としての新しい単位．観察値 x の正規曲線の平均 μ からの距離を，その曲線の標準偏差 σ で除して変換したもの．$x-\mu$ が 10 で，σ が 2 であれば，x は μ から σ の 5 倍の距離にある．

$$z=\frac{x-\mu}{\sigma}=\frac{10}{2}=5$$

標準誤差 standard error

SE　標準誤差．標本抽出分布の標準偏差に対する特別の語 (標本抽出ないし偶然によるばらつきを反映する)．

SE_p　比率の標準誤差．真の比率の周囲における標本の比率の分布の標準偏差に対する特別の語

$$\sqrt{\frac{pq}{n}}$$

棄却比 critical ratio　標本の結果から作られた検定のための統計量で，標本の結果の有意性を示す．

標準正規偏差 (z) 検定 standard normal deviate (z) test　どんな有意検定でも標準正規曲線を用いる．棄却比は

$$z=\frac{\overset{\text{観察された結果}}{x}-\underset{\text{正規曲線の中心}}{\mu}}{\underset{\text{正規曲線の標準偏差}}{\sigma}}$$

z, μ, 0 の特別な解釈は，特別な正規分布と検定による．

正規差分検定 normal difference test　同じ二項分布母集団から無作為抽出した二つの標本間の偶然による比率の差を表す正規曲線に基づいた z 検定である．検定は帰無仮説が正しい(差が 0 であることが存在)と仮定する．

$$z=\frac{(\hat{p}_1-\hat{p}_2)-0}{SE_{\hat{p}_1-\hat{p}_2}}$$

p_1　二項分布母集団 1 で a という性質を持つものの真の比率

p_2　二項分布母集団 2 で a という性質を持つものの真の比率

q_1　$1-p_1$，すなわち二項分布母集団 1 で a という性質を持たないものの真の比率

q_2　$1-p_2$，すなわち二項分布母集団 2 で a という性質を持たないものの真の比率

$\hat{p}_1$　標本 1 で a という性質を持つものの比率

$\hat{p}_2$　標本 2 で a という性質を持つものの比率

$\hat{q}_1$　$1-\hat{p}_1$，すなわち標本1でaという性質を持たないものの比率

$\hat{q}_2$　$1-\hat{p}_2$，すなわち標本2でaという性質を持たないものの比率

p'　二項分布母集団でaという性質を持つものの真の比率pの共通の見積もり．n_1とn_2を重みとして，標本の比率$\hat{p}_1$と$\hat{p}_2$の重みつき平均から得られるか，あるいは合併した標本の全体の比率から得られる．

q'　$1-p'$ すなわち二項分布母集団でaという性質を持たないものの真の比率 $1-p$ の共通の見積もり

$SE_{p_1-p_2}$　二つの標本の比率の差の標準誤差．同じ二項分布母集団から無作為抽出した二つの標本の間の差の分布の標準偏差に対する特別な用語である．

$$=\sqrt{\frac{p'q'}{n_1}+\frac{p'q'}{n_2}}.\quad または\quad =\sqrt{p'q'\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

差の標準誤差 standard error of the difference　　記号の $SE_{p_1-p_2}$

重みづけ weighting　ある総合値を得る際に適当な重要性を与えるために，標本の大きさnといったような要素で値を乗ずること．

記述統計学 descriptive statistics　より大きなデータへの一般化を含まない一組のデータの記述，ないしは要約に用いる方法．

推測統計学 inferential statistics　比較的小さな一組のデータ（すなわち標本）をまとめて，さらに大きな一組の可能データ（つまり，母集団）に関して一般化をおこなうのに用いる方法．

偏りのない推定値 unbiased estimate　特定の方法で求められる可能なすべての標本推定値の総平均が，母数に全く等しいという性質をもった標本推定値．

偏差平方和 sum of squares　母平均のまわり，または標本平均を中心とした偏差平方和，$\Sigma(x-\mu)^2$，または $\Sigma(x-\bar{x})^2$．分散の分子．

t-分布 t-distibution　変量 $t=\frac{\bar{x}-\mu}{s/\sqrt{n}}$ の分布．有意臨界比において，sをσのかわりにしていることに注意 $\left(z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\text{と比較せよ}\right)$．分布の形は自由度の数（$df$）に依存する：$df\to\infty$ に伴って，正規分布に近づく．母分散が未知の場合に，小さな標本（df が 30 かそれ以下）の検定に用いる．

$\bar{x}$　標本平均$=\frac{\Sigma x}{n}$．母平均 μの偏りのない推定値

s^2　標本分散$=\frac{\Sigma(x-\bar{x})^2}{n-1}$．母分散 σ^2 の偏りのない推定値．

s　標本標準偏差$=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$．母集団の標準偏差 σの偏りのない推定値．

t　変量または臨界比 $\frac{\bar{x}-\mu}{s/\sqrt{n}}$．

$\sigma_{\bar{x}}$ ($SE_{\bar{x}}$)　標本平均の分布の標準偏差．すなわち平均の標準誤差．母集団の標準偏差を標本の大きさの平方根で割ったもの $\left(\frac{\sigma}{\sqrt{n}}\right)$ に等しい．$\frac{s}{\sqrt{n}}$ によって推定される．

中心極限定理 central limit theorem　非正規型母集団からの任意標本の平均の分布は，標本の大きさ(n)がじゅうぶんに大きくなるにつれて，正規分布に近づく（上の t-分布の項参照）．

第1型 (α) 過誤 type I (α) error　帰無仮説は真である（真の差は0）が，観測結果が，帰無仮説曲線のα（アルファ）または棄却域にはいるために，棄却される．ゆえに，誤りの危険は任意の有意水準であるαであり，前もって決められる．

第2型 (β) 過誤 type II (β) error　帰無仮説は偽である（真の差が存在する）が，観測結果が，帰無仮説曲線の $1-\alpha$ の範囲（受容域）にはいるので，受容される．誤りの危険は，すぐには決められない．β 過誤ともいわれる

H_0　帰無仮説（差は 0）

H_0: $\mu_1=\mu_2$ すなわち $\mu_1-\mu_2=0$

H_a　対立仮説

H_a: $\mu_1\neq\mu_2$（方向づけなし，両側検定）ないしは

H_a: $\mu_1 > \mu_2$ ないしは $\mu_1 < \mu_2$（方向づけあり，片側検定）

α　アルファ，すなわち有意性について前もって任意にきめた値（通常 0.05 ないし 0.01）．第Ⅰ型過誤すなわち真の差があると誤って主張する危険率を示す．

β　ベータ，第Ⅱ型過誤すなわち真の差を検出するのに失敗する危険率と同じ．α 値，真の差，標本の大きさおよび分散の大きさに負の相関を示す．

受容域 acceptance region　帰無仮説曲線の $1-\alpha$ の範囲．

棄却域 rejection region　帰無仮説曲線の，α の範囲または（両）端．

帰無仮説（有意）検定 null hypothesis (significance) test　帰無仮説（差が 0）の検定．もし，帰無仮説が棄却されるならば，特定の対立仮説が受容される．

対立仮説 alternate hypothesis　帰無仮説が棄却されるときに受容される仮説．

帰無仮説曲線 null hypothesis curve　真の差はないという仮定のもとでの理論的標本抽出分布．

カイ二乗（χ^2）分布 chi square (χ^2) distribution　質的なデータに関する有意性の検定に用いられる分布である．標本の分布あるいは理論上の分布をなす標本分布を比べるのに用いられる．χ^2 に対する近似式は

$$\sum\left[\frac{(\text{観察値}-\text{期待値})^2}{\text{期待値}}\right]=\sum\frac{(O-E)^2}{E}$$

である．分布の形は，自由度 df の数によって決まる．

カイ二乗検定は

……二つの標本の比率の間の差に対するもう一つの検定法である．

……要因と結果との間の関連性の有無に関する検定である．

……帰無仮説に基づいた期待値の計算を必要とする．

……表の各々の欄に対する項の和 $\sum\left[\frac{(\text{観察値}-\text{期待値})^2}{\text{期待値}}\right]$ に等しい．

縦２列・横２列の表のカイ二乗検定は

……自由度が１である．

……χ^2 の値は，標準正規偏差 z の二乗に等しい．

Yates の修正 Yates' correction　自由度１のカイ二乗検定で用いられる，離散的（質的）データに連続的分布を適用するための連続性に関する修正である．修正は，各々の差（観察値－期待値）を二乗する前に，差から ½ を引くことによって行われる．$\chi_c^2=\sum\frac{(|O-E|-½)^2}{E}$

自由度（df） degree of freedom　理論上の標本分布に独立して影響を与える数（例えば，χ^2, t, F 分布などに）

分割表 contingency table　縦 r 列・横 c 列への質的データの分類表（その分類に関連性についてのカイ二乗検定が適用できる）

期待値 expected number　帰無仮説（真の差が存在しない）の仮定のもとで，標本の特別な範囲において期待された数．

χ^2 分割検定　chi square contingency test

二つ，あるいはそれ以上の標本分布を互いに比較する定性的資料に対する χ^2 検定（分割表）．

$(r-1)\times(c-1)$ 個の df がある．

適合度検定　goodness of fit test

標本分布と理論的母集団分布とを比較する定性的資料に対する χ^2 検定．$(r-1)$ 個の df がある．

標本から推定される理論分布のそれぞれのパラメーターに対して自由度は一つ減ぜられる．

集約された分散 pooled variance　標本分散推定値（s^2）をプールして求めた母分散（σ^2）の推定値．それぞれの s^2 は，計算にはいる $n-1$ の項で加重されている．

s^2_{pooled}　集約（プール）された分散．公式は，$\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$

差の標準誤差 standard error of the difference　偶然による２標本平均の差（$\bar{x}_1-\bar{x}_2$）が示す分布の標準偏差．すなわち差の標準誤差．公式は，(a) 母分散（σ^2）がわかっているかどうか，(b) 標本の大きさが等しいか否かによって変わる．

$\sigma_{\bar{x}_1-\bar{x}_2}(SE_{\bar{x}_1-\bar{x}_2})$　2標本平均の差の標準誤差．独立な平均に対して公式は，——

σ^2 (母分散) が既知のとき：

$$\sqrt{(SE_{\bar{x}_1})^2+(SE_{\bar{x}_2})^2}=\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}$$

σ^2 が未知のとき

$$n_1=n_2:\ \sqrt{(\text{推定}\,SE_{\bar{x}_1})^2+(\text{推定}\,SE_{\bar{x}_2})^2}=\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

$$n_1\neq n_2:\ \sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)},\quad s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$$

F-検定 *F*-test　2分散の同等性の検定．検定にあたって，F分布 (すなわち，母分散に関しての二つの独立な標本からの推定値の比率の分布) と比較されるF値 ($F=s_1^2/s_2^2$) を用いる

独立な標本 independent sample　全く別の要素からの測定値を表すか，さもなければ，対をなさないかあるいは無関係で，その値が相互に独立な標本．

非独立な (相関する) 標本 non-independent (correlated) sample　その値が同一か関連しあう要素に基づく測定値を表わし，よって独立でない相互関係のある標本．

対のデータ Paired data　各 x 値に対し対応する y 値をもつ1組の観測結果．

共分散 covariance　各対のデータ x, y について，x の平均からの x 偏差 (すなわち $x-\bar{x}$) と，y の平均からの y 偏差 (すなわち $y-\bar{y}$) の積をとる．その積を全部の対について合計する．すなわち，

$$\text{標本の共分散}=\frac{\Sigma(x-\bar{x})(y-\bar{y})}{n-1}$$

[これを，x, y に関する分散項，$\dfrac{\Sigma(x-\bar{x})^2}{n-1}$，$\dfrac{\Sigma(y-\bar{y})^2}{n-1}$ と対照せよ]．

相関係数 correlation coefficient　r．これは2変量間の関連の強さ (線型関係の程度) を表わす数字である．r は -1 (完全な負の相関) と $+1$ (完全な正の相関) の間の値をとる．

r　相関係数 (Pearson の r または積能率係数とも呼ばれる)

$$=\frac{\Sigma(x-\bar{x})(y-\bar{y})/(n-1)}{\sqrt{\Sigma(x-\bar{x})^2\Sigma(y-\bar{y})^2}/(n-1)}$$

$$=\frac{x,\ y\ \text{の共分散}}{(x\ \text{の標準偏差})(y\ \text{の標準偏差})}$$

r が，母集団の相関係数 ρ (ロー) を推定する標本相関係数であることに注意．

対のデータの *t*-検定 (階差法) *t*-test for paired data (difference method)　1対のデータ間の差，d (すなわち $x-y$) を変数として扱う *t*-検定．

t-検定 (対のデータ用)

これは対の間の差 (d) をとり扱うことにより簡単化される．すなわち，

——平均差 $=\bar{d}=\Sigma d/$対の数

——差の標準偏差 $=s_d=\sqrt{\dfrac{\Sigma(d-\bar{d})^2}{\text{対の数}-1}}$

——差の推定標準誤差 $=$ 推定 $SE_{\bar{d}}=\dfrac{s_d}{\sqrt{\text{対の数}}}$

——そのとき，*t*-検定は

$$t=\frac{\bar{d}-0}{\text{推定}\,SE_{\bar{d}}}$$

非パラメーター検定 non-parametric test　分布のパラメーターについて何の仮定もしない検定 (分布によらない検定)．たとえば符号検定．

符号検定 sign test　対間の正の差，負の差の数に基づく非パラメーター検定．帰無仮説のもとでは，+，− 符号の数は $p=0.5$ で2項分布している．

McNemar の検定 McNemar's test　非独立な (相関している) 比にする検定．

$$\chi^2_{1df}=\frac{(\text{正の対の数}-\text{負の対の数})^2}{(\text{正の対の数})+(\text{負の対の数})}$$

信頼区間推定 confidence interval estimation　(母集団パラメーターの) 真の値がある信頼度 (たとえば, 100 回中 95 あるいは 99 回) でその内に存在する区間の決定

信頼限界 (区間) confidence limits (interval)　(母集団パラメーターの) 真の値を推定するための信頼区間の (上, 下の) 境界.

C.L. または C.I.　母集団の (真の) パラメーターの信頼限界または信頼区間.

(95% または 99%)　式は, 標本値±[(因子)×(統計量の標準誤差)] である. 因子の大きさは, (1) 適当な df に対する標本統計固有の分布 (たとえば z または t) と (2) 信頼水準 (たとえば 95% または 99%) に依存する. 標準誤差の大きさは, (1) 標本の大きさと (2) 母集団分散に依存する.

回帰直線 regression line　平均の y (従属変数) を x (独立変数) の関連した値と関係づける式. 線形回帰では, その式は

$$y_c=a+bx$$

である.

b　回帰係数で, 回帰直線の勾配とも呼ばれる.

$$=\frac{\Sigma(x-\bar{x})(y-\bar{y})/(n-1)}{\Sigma(x-\bar{x})^2/(n-1)}=\frac{x, y \text{ の共分散}}{x \text{ の分散}}$$

b は母集団回帰係数 B を推定する標本回帰係数であることに注意されたい.

y_c　回帰式 $y_c=a+bx$ から予測される y の値. したがって, 常に回帰直線上に乗る.

回帰係数 regression coefficient　回帰直線の勾配 (b). x の単位変化当たりの y の平均変化を表わす.

回帰からの平均二乗偏差 mean square deviation from regression　回帰直線からの y の値の平均の垂直二乗偏差:

$$(s^2{}_{y.x})$$

回帰からの標準偏差 standard deviation from regression　回帰直線からの y の値の平均の垂直 2 乗偏差の平方根:

$$(s_{y.x})$$

最小二乗法 least square method　データの組に対し, 回帰直線のまわりで y 値の垂直方向の偏差の平方和が最小になるように回帰直線を適合させる方法.

決定係数 coefficient of determination　標本相関係数の 2 乗, r^2. r 同様, 2 変数間の関連度の尺度である. x の分散に伴うかまたは x の分散で説明される y の分散の比率を示す.

分散分析 analysis of variance(ANOVA)　母集団分散の独立な推定値を生ずるように, 分散を二つかそれ以上の処理グループ間と処理グループ内とのように, 分散部分に分割する方法. その場合, これらの独立な推定値の比は, F 分布を用いて検定される. これは, 標本平均値の有意性検定を与えるものである.

F 分布 F distribution　母集団分散の二つの独立な標本推定値の比にしたがう分布. すなわち, $F=s_1^2/s_2^2$. 分布の形は, それぞれ s_1^2, s_2^2 の二つの自由度の値に依存する.

ポアソン分布 Poisson distribution　離散分布 (0 か正の整数のみが可能) であり, p が非常に小さいが n が, $n\times p$ が小さな有限定数になるくらい大きい場合の 2 項分布の極限の形として求められる. 平均値と分散を表わす 1 個のパラメーター λ (ラムダ) がある.

ベイズの定理 Bayes' theorem　条件付き確率を得るのに有用な定理である. 一つの応用は, 症候群 (または陽性検査) を与える病気の確率 P(D/S) を決定することである. これは, その病気を与える症候群 (または陽性検査) の確率 P(S/D) の知識を考慮に入れるものである. 簡単化した表現は:

$$P(D/S)=\frac{P(D)\cdot P(S/D)}{[P(D)\cdot P(S/D)]+[P(\bar{D})(S/\bar{D})]}$$