人口動態統計上巻から抜粋

I　人口動態調査の概要

PartI　Outline of Vital Statistics

第1章　調査の概要

1　調査の概要

我が国の人口動態統計は、市区町村長が作成する人口動態調査票に基づいて表章される。すなわち、出生・死亡・婚姻及び離婚については戸籍法（昭和22年法律第224号）による届書等から、死産については死産の届出に関する規定（昭和21年厚生省令第42号）による届書等から人口動態調査票が作成され、これを収集し集計した統計が人口動態統計である。

1］調査の目的

我が国の人口動態統計事象を把握し、人口及び厚生行政施策の基礎資料を得ることを目的とする。

2］調査の沿革

人口動態調査は、明治31年「戸籍法」が制定され、登録制度が法体系的にも整備されたのを機会に、同32年から人口動態調査票は1件につき1枚の個別票を作成し、中央集計をする近代的な人口動態統計制度が確立した。

さらに、昭和22年6月に「統計法」に基づき「指定統計第5号」として指定され、その事務の所管は同年9月1日に総理庁から厚生省に移管されて今日に至っている。

3］調査の対象

人口動態調査は、出生・死亡・婚姻・離婚及び死産の全数を対象としているが、本報告書は、日本において発生した日本人に関する事象を集計したものである。日本人の外国におけるもの及び外国人の日本におけるものについて、参考として掲載している。

4］調査の期間

調査該当年の1月1日から同年12月31日までに事件が発生したものであって、調査該当翌年の1月14日までに市区町村長に届け出られたものである。

なお、婚姻や協議離婚は、届書が市区町村長に受理されることによって事件が発生する。したがって、届出遅れの問題はないが、出生・死亡・死産や調停・審判・判決による離婚は、事件発生から届出までに相当の遅れのある場合がある。前年以前に発生した出生・死亡については、中巻（500～503ページ）に掲載してある。

5］調査票の種類及び調査の事項

調査票は、次の5種類である。その様式及び各届書は、別掲（40～49ページ）のとおりである。（次ページの注記参照）

人口動態調査出生票　人口動態調査死亡票　人口動態調査死産票

人口動態調査婚姻票　人口動態調査離婚票

調査の事項は、上記5種類の調査票を参照されたい。ただし、職業及び産業の事項については、国勢調査実施年の4月1日から翌年3月31日までについてのみ調査を行う。

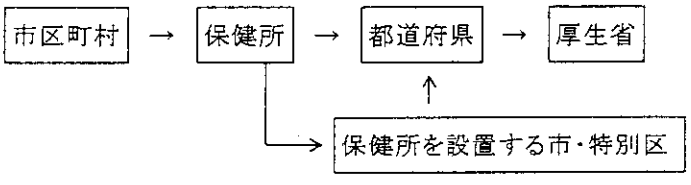
6］調査の方法及及び報告経路

届書の届出義務者及び届出期間は、次のとおりである。

種別	届出義務者	届出先	届出期間[1]
出生	1父又は母　2同居者　3出産に立ち会った医師・助産婦又はその他の者	市区町村長	14日
死亡	1同居の親族　2その他の同居者　3家主・地主又は家屋もしくは土地の管理人　4同居の親族以外の親族	市区町村長	7日
死産	1父又は母　2同居人　3死産に立ち会った医師　4死産に立ち会った助産婦　5その他の立会者	市区町村長	7日
婚姻	夫　妻	夫又は妻の本籍地もしくは所在地の市区町村長	規定なし
離婚	夫　妻	夫又は妻の本籍地もしくは所在地の市区町村長	協議離婚は規定なし 調停・審判・判決離婚は10日

注:1)出生・死亡及び裁判による離婚は届出事件発生の日から、死産はその日の翌日から起算。

市区町村長は、出生・死亡・死産・婚姻・離婚の届出を受けたときは、その届書等に基づいて人口動態調査票を作成し、これを保健所の管轄区域によって当該保健所長に送付する。
保健所長は、市区町村長から提出された調査票を取りまとめ、毎月、都道府県知事に送付する。
この場合、保健所を設置する市の保健所長は、当該市の市長を経由する。
都道府県知事は、保健所長から提出された調査票の内容を審査し、厚生大臣に送付する。



注:保健所を設置する市とは、地域保健法施行令(昭和23年4月2日政令第77号)第1条に規定する市をいう。

7] 関係法規

人口動態調査令(昭和21年9月30日勅令第447号)
人口動態調査令施行細則(昭和23年2月24日厚生省令第6号)
戸籍法(昭和22年12月22日法律第224号)
戸籍法施行規則(昭和22年12月29日司法省令第94号)
出生証明書の様式等を定める省令(昭和27年11月17日法務・厚生省令第1号)
国籍法(昭和25年5月4日法律第147号)
死産の届出に関する規程(昭和21年9月30日厚生省令第42号)
「ポツダム宣言の受諾に伴い発する命令に関する件」(昭和20年9月20日勅令第542号)に基づく厚生省関係諸命令の措置に関する法律(昭和27年4月28日法律第120号)第3条により法律としての効力を有する。
死産届書、死産証書及び死胎検案書に関する省令(昭和27年4月28日厚生省令第12号)

2　用語の解説

自然増加　出生数から死亡数を減じたものをいう。
乳児死亡　生後1年未満の死亡をいう。
新生児死亡　生後4週未満の死亡をいう。
早期新生児死亡　生後1週未満の死亡をいう。

妊娠期間　出生、死産及び周産期死亡の妊娠期間は満週数による。(昭和53年までは、数えによる妊娠月数)

早期:妊娠満37週未満(259日未満)

正期:妊娠満37週から満42週末満(259日から293日)

過期:妊娠満42週以上(294日以上)

死産　妊娠満12週(妊娠第4月)以後の死児の出産をいい、死児とは、出産後において心臓膊膊動、随意筋の運動及び呼吸のいずれも認めないものをいう。

自然死産と人工死産　人工死産とは、胎児の母体内生存が確実であるときに、人工的処置(胎児又は付属物に対する措置及び陣痛促進剤の使用)を加えたことにより死産に至った場合をいい、それ以外はすべて自然死産とする。

なお、人工的処置を加えた場合でも、次のものは自然死産とする。

(1)　胎児を出生させることを目的とした場合

(2)　母体内の胎児が生死不明か、又は死亡している場合

(参　考)

死産統計を観察する場合、次の沿革を考慮する必要がある。

昭和23年以降:優生保護法の施行(7月)により、人工妊娠中絶のなかの、妊娠第4月以降のものも人工死産に含むことになった。

昭和24年以降:優生保護法の改正(6月)により、人工妊娠中絶の理由に「経済的理由により母体の健康を著しく害するおそれのあるもの」も含むことになった。

昭和27年以降:優生保護法の改正(5月)により、優生保護審査会の審査を廃止するなど、その手続が簡素適正化され、優生保護法による指定医師は本人及び配偶者の同意を得て、要件に該当する者に対し、人工妊娠中絶を行うことができるようになった。

昭和43年以降:胎児を出生させる目的で人工的処置を加えたにもかかわらず死産をした場合、従来は人工死産であったが、自然死産として取り扱うこととなった。

昭和51年以降:優生保護法による人工妊娠中絶を実施することができる時期の基準を、従来の「通常妊娠8月未満」から「通常妊娠第7月末満」に改めた。
(昭和51年1月20日付け厚生省発衛第15号厚生事務次官通知)

昭和54年以降:優生保護法による人工妊娠中絶を実施することのできる時期の基準を、従来の「通常妊娠第7月未満」から「通常妊娠満23週以前」に表現を改めた。(昭和53年11月21日付け厚生省発衛第252号厚生事務次官通知)

平成3年以降:優生保護法による人工妊娠中絶を実施することのできる時期の基準を、従来の「通常妊娠満23週以前」から「通常妊娠満22週未満」に改めた。(平成2年3月20日付け厚生省発健医第55号厚生事務次官通知)

周産期死亡　妊娠満22週(154日)以後の死産に早期新生児死亡を加えたものをいう。

妊産婦死亡　妊娠中または妊娠終了後満42日未満[1]の女性の死亡で、妊娠の期間及び部位には関係しないが、妊娠もしくはその管理に関連した、又はそれらによって悪化した全ての原因によるものをいう。ただし、不慮又は偶発の原因によるものを除く。

その範囲は、直接産科的死亡(O00～O92)及び間接産科的死亡(O98～O99)に原因不明の産科的死亡(O95)、産科的破傷風(A34)及びヒト免疫不全ウイルス[HIV]病(B20～B24)を加えたものである[2]。

直接産科的死亡:妊娠時における産科的合併症が原因で死亡したもの。

間接産科的死亡:妊娠前から存在した疾患又は妊娠中に発症した疾患により死亡したもの。これらの疾患は、直接産科的原因によるものではないが、妊娠の生理的作用によって悪化したもの。

注1)　昭和53年までは「産後90日以内」とし、昭和54年から平成6年までは「分娩後42日以内」としている。

注2)　昭和53年までの範囲は、基本分類表「XI　妊娠、分娩および産じょくの合併症」には「間接産科的死亡」は含まれないので、「直接産科的死亡」がほぼ該当する。また、昭和54年から平成

6年までは、基本分類表「XI　妊娠、分娩および産じょくの合併症」(630～676)が該当する。

後発妊産婦死亡　ICD10で新たに定義されたものであり、妊娠終了後満42日以後1年末満における直接又は間接産科的原因による女性の死亡をいい、その範囲は、あらゆる産科的原因による母体死亡(O96)、産科的破傷風(A34)及びヒト免疫不全ウイルス[HIV病](B20～B24)である。

世帯の主な仕事

農家世帯　農業だけ又は農業とその他の仕事を持っている世帯

自営業者世帯　自由業・商工業・サービス業等を個人で経営している世帯

常用勤労者世帯(Ⅰ)　企業・個人商店等(官公庁は除く)の常用勤労者世帯で勤め先の従事者数が1人から99人までの世帯(日々又は1年末満の契約の雇用者はその他の世帯)

常用勤労者世帯(Ⅱ)　常用勤労者世帯(Ⅰ)にあてはまらない常用勤労者世帯及び会社団体の役員の世帯(日々又は1年末満の契約の雇用者はその他の世帯)

その他の世帯　上記にあてはまらないその他の仕事をしている世帯

無職の世帯　仕事をしている者のいない世帯

3　比率の解説

(注)　年次推移の表の昭和45年、50年及び55年については、10月1日現在日本人人口を国勢調査の確定数を用いて再計算したので、昭和45年、50年及び55年の報告書の数値と異なる場合がある。

(1)　総　覧

$$\text{出生率} = \frac{\text{年間出生数}}{\text{10月1日現在日本人人口}} \times 1,000$$

$$\text{死亡率} = \frac{\text{年間死亡数}}{\text{10月1日現在日本人人口}} \times 1,000$$

$$\text{乳児死亡率} = \frac{\text{年間乳児死亡数}}{\text{年間出生数}} \times 1,000$$

$$\text{新生児死亡率} = \frac{\text{年間新生児死亡数}}{\text{年間出生数}} \times 1,000$$

$$\text{自然増加率} = \frac{\text{自然増加数}}{\text{10月1日現在日本人人口}} \times 1,000$$

$$\text{死産率} = \frac{\text{年間死産数}}{\text{年間出産数(出生数＋死産数)}} \times 1,000$$

$$\text{自然死産率} = \frac{\text{年間自然死産数}}{\text{年間出産数(出生数＋死産数)}} \times 1,000$$

$$\text{人工死産率} = \frac{\text{年間人工死産数}}{\text{年間出産数(出生数＋死産数)}} \times 1,000$$

$$\text{周産期死亡率} = \frac{\text{年間周産期死亡数}}{\text{年間出産数(出生数＋妊娠満22週以後の死産数)}} \times 1,000$$

妊娠満22週以後の死産率(総数・自然・人工)

$$= \frac{\text{年間妊娠満22週以後の死産数(総数・自然・人工)}}{\text{年間出産数(出生数＋妊娠満22週以後の死産数)}} \times 1,000$$

$$\text{早期新生児死亡率} = \frac{\text{年間早期新生児死亡数}}{\text{年間出生数}} \times 1,000$$

$$\text{婚姻率} = \frac{\text{年間婚姻届出件数}}{\text{10月1日現在日本人人口}} \times 1,000$$

$$\text{離婚率} = \frac{\text{年間離婚届出件数}}{\text{10月1日現在日本人人口}} \times 1,000$$

(2)　出　生

$$\text{出生性比} = \frac{\text{年間の男子出生数}}{\text{年間の女子出生数}} \times 100$$

母の年齢(年齢階級)別出生率

$$= \frac{\text{ある年齢(年齢階級)の母が1年間に生んだ子の数}}{\text{10月1日現在における日本人女子のある年齢(年齢階級)の人口}} \times 1,000$$

月間出年率(年換算率)＝ $\frac{\text{月間出生数}}{\text{月初人口×年換算係数}}$ ×1,000

(注)年換算係数＝ $\frac{\text{月間日数(30, 31, 28又は29)}}{\text{年間日数(365又は366)}}$

すなわち1年の長さを1とした場合の各月の長さをいう。

合計特殊出生率＝ $\left\{\frac{\text{母の年齢別出生数}}{\text{年齢別女子人口}}\right\}$ 15歳から49歳までの合計

15歳から49歳までの女子の年齢別出生率を合計したもので、1人の女子が仮にその年次の年齢別出生率で一生の間に生むとした時の平均子ども数に相当する。

(3) 死　亡

死亡性比＝ $\frac{\text{年間の男子死亡数}}{\text{年間の女子死亡数}}$ ×100

年齢(年齢階級)別死亡率(総数・男・女),

＝ $\frac{\text{年間のある年齢(年齢階級)の死亡数(総数・男・女)}}{\text{10月1日現在における日本人(総数・男・女)のある年齢(年齢階級)人口}}$ ×1,000

年齢(年齢階級)別死亡率性比＝ $\frac{\text{ある年齢(年齢階級)の男子死亡率}}{\text{ある年齢(年齢階級)の女子死亡率}}$ ×100

月間死亡率(年換算率)＝ $\frac{\text{月間死亡数}}{\text{月初人口×年換算係数}}$ ×1,000

(注)年換算係数＝ $\frac{\text{月間日数(30,31,28又は29)}}{\text{年間日数(365又は366)}}$

すなわち1年の長さを1とした場合の各月の長さをいう。

死因別死亡率(年間)＝ $\frac{\text{年間の死因別死亡数}}{\text{10月1日現在日本人人口}}$ 100,000

年齢調整死亡率＝ $\frac{\left\{\left(\text{観察集団の各年齢(年齢階級)の死亡率}\right)\times\left(\text{基準人口集団のその年齢(年齢階級)の人口}\right)\right\}\text{の各年齢(年齢階級)の総和}}{\text{基準人口集団の総数}}$

(参　考)

死亡率は年齢によって異なるので、国際比較や年次推移の観察には人口構成の差異を取り除いて観察するために、年齢調整死亡率を使用する事が有用である。

年齢調整死亡率の基準人口については、平成元年までは昭和10年の性別総人口(都道府県は昭和35年総人口)を使用してきたが、現実の人口構成からかけ離れてきたため、平成2年からは昭和60年モデル人口(昭和60年国勢調査日本人人口をもとに、ベビーブーム等の極端な増減を補正し、1,000人単位で作成したもの)を使用している。

なお、計算式中の「観察集団の各年齢(年齢階級)の死亡率」は、1,000倍(死因の場合は100,000倍)されたものである。

基準人口(昭和60年モデル人口)

	基準人口
総数	120 287 000
0 ～ 4歳	8 180 000
5 ～ 9	8 338 000
10～14	8 497 000
15～19	8 655 000
20～24	8 814 000
25～29	8 972 000
30～34	9 130 000
35～39	9 289 000
40～44	9 400 000
45～49	8 651 000
50～54	7 616 000
55～59	6 581 000
60～64	5 546 000
65～69	4 511 000
70～74	3 476 000
75～79	2 441 000
80～84	1 406 000
85歳以上	784 000

(4)乳児死亡

$$乳児死亡性比 = \frac{年間の男子乳児死亡数}{年間の女子乳児死亡数} \times 100$$

月間乳児死亡率

$$= \frac{その月の月間乳児死亡数}{その月を含む過去1年間の出生数 \times \frac{その月の月間日数}{その月を含む過去1年間の日数}} \times 1,000又は100,000$$

$$死因別乳児死亡率 = \frac{年間の死因別乳児死亡数}{年間出生数} \times 100,000$$

$$死因別新生児死亡率 = \frac{年間の死因別新生児死亡数}{年間出生数} \times 100,000$$

(5)死産

$$死産性比 = \frac{年間の男子死産数}{年間の女子死産数} \times 100$$

$$月間死産率(総数・自然・人工) = \frac{月間死産数(総数・自然・人工)}{月間出産数(出生数+死産数)} \times 1,000$$

月間妊娠満22週以後の死産率(総数・自然・人工)

$$= \frac{月間妊娠満22週以後の死産数(総数・自然・人工)}{月間出産数(出生数+妊娠満22週以後の死産数)} \times 1,000$$

(6)周産期死亡

$$月間周産期死亡率 = \frac{月間周産期死亡数}{月間出産数(出生数+妊娠満22週以後の死産数)} \times 1,000$$

(7)妊産婦死亡

$$妊産婦死亡率 = \frac{妊産婦死亡数}{年間出産数(出生数+妊娠満12週以後の死産数)(又は年間出生数)} \times 100,000$$

$$後発妊産婦死亡率 = \frac{後発妊産婦死亡数}{年間出産数(出生数+妊娠満12週以後の死産数)} \times 100,000$$

注:妊産婦死亡については55頁を参照されたい。

《研修テキスト》

統計学の基礎１，２

目　次　　ﾍﾟｰｼﾞ

A．統計的方法
　１．統計的な考え方　1
　２．特性値　1
B．標本抽出法
　１．有意抽出法　2
　２．確率抽出法　2
C．確率分布
　１．二項分布　3
　２．ポアソン分布　3
　３．正規分布　3
　４．ｔ－分布　4
　５．カイ２乗分布　4
　６．Ｆ－分布　5
　７．二項分布の正規近似　5
D．推定と検定
　１．母数　6
　２．統計量　6
　３．点推定　6
　４．区間推定　6
　５．標本数の決定　7
　６．検定　7
　７．分布による仮説検定　8
E．相関分析
　１．相関係数ｒの検定　10
　２．偏相関係数の検定　11
　３．平行性検定法　11

統計学演習問題　1－2
統計学 *Exsample 1*（出題）

統計学の基礎１，２

～統計分布による推定・検定法を中心として～

福岡県保健環境研究所　篠原志郎

A．統計的方法

1．統計的な考え方

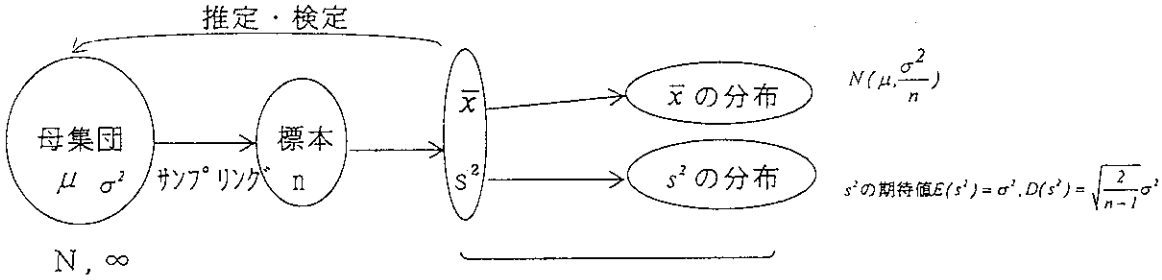
1）定値にはバラツキ（*Dispersion*）がつきもの（誤差）

2）そのバラツキには統計的な規則性がみられる（分布）

3）大数の法則：　xが平均μ，分散σ^2の分布に従うとき，xがどんな分布であっても，標本数nの無作為標本（*Random sampling*）によるxは正規分布$N(\mu, \sigma^2/n)$にnが大きくなるにつれて近づく（中心極限定理：一般に，$n>25$で適用）

4）計的推論に基づいて行動する（検定）

Fisher の３原則＋１ *(1935)*：反復 *(Replication)*, 局所管理 *(Local control)*, 無作為化 *(Randomization)*, 因子の組み合わせ *(Factorial combination)*



2．特性値

母集団を知る手がかりとして、母集団から取られた標本の度数分布表を作り、ヒストグラム等を描く．それによって分布の形や特徴がつかめるが、更に、数量化できれば、なお、的確な情報を得ることができる．この数量的に表現された指標を特性値とよぶ．

1）分布の中心を代表する指標

① 平均値・・・a．算術平均値　$\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n} = \dfrac{1}{n}\sum_{k=1}^{n} x_k = \sum x / n,$

b．幾何平均値　$\bar{y} = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} = (x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}},$

c．調和平均値　$\bar{z} = \dfrac{1}{n}\left(\dfrac{1}{x_1} + \dfrac{1}{x_2} + \cdots + \dfrac{1}{x_n}\right), x_1, x_2, \cdots, x_n \neq 0$

② 中央値・・・　Ｍｅ＝（n+1）／2番目の測定値　（ｎ：奇数odd number）

＝[n/2番目＋(n/2+1)番目]／2　（ｎ：偶数even number）

③ モード（最頻値）Ｍｏ＝最も度数の多い測定値

④ 最大値、最小値

2）分布の広がりの程度を代表する指標

$s^2 = \dfrac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2 = \dfrac{1}{n-1}\left(\sum_{k=1}^{n} x_k^2 - n\bar{x}^2\right), s = \sqrt{s^2}.$

① 分散（不偏分散）

② 標準偏差　$s = \sqrt{s^2}$,　$R = x_{max} - x_{min}$

③ 範囲

④ 変動係数　c.v. ＝ $s/\bar{x} \times 100(\%)$

3）正規性からのズレと異常　「 $\alpha_3 > 0$（対数正規型）

① ゆがみ具合（歪度：*Skewness*）　$\alpha_3 = \mu_3 / \sigma^3$, $\mu_k = E[(X-\mu)^k]$

② 尖り具合（尖度：*Kurtosis*）　$\alpha_4 = \mu_4 / \sigma^4 - 3$, $\alpha_4 > 0$（トガリ型）

4）データのスクリーニング

① 異常データの検出　*Grubbs,F.E.* 法

B．標本抽出法（*Sampling*）

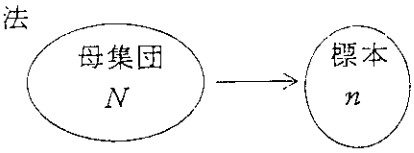
1．有意抽出法・・・意図的に標本を抽出する方法

1）典型法

母集団の特性値から母集団の相似形を意識して求めた抽出法

2）割当法

母集団の特性値が同じ比率となるように選ぶ方法



2．確率抽出法

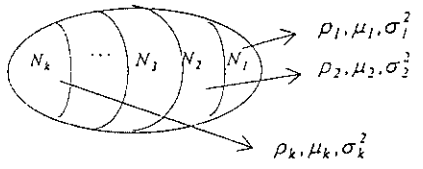
1）単純無作為抽出法（*Simple random sampling*）

N個の要素の中から大きさ n の標本 $x_1, x_2, \cdots, x_n$ を非復元抽出して標本平均 $\bar{x}$ を求める．

$$\bar{x} = \frac{1}{n}\sum x, \quad E(\bar{x}) = \mu, \quad V(\bar{x}) = \frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}$$ で与えられる．

2）層別抽出法（*Stratified sampling*）

母集団をｋ個の層に分け、各層から標本を抽出する方法



$$\text{全平均} \mu = \sum_{i=1}^{k} \rho_i \mu_i \ (\rho_1 + \rho_2 + \cdots + \rho_k = 1, \rho_i = \frac{N_i}{N})$$

$$\text{全分散} \sigma^2 = \sum \rho_i \sigma_i^2 + \sum \rho_i (\mu_i - \mu)^2 = \sigma_w^2 + \sigma_b^2$$

$$\sigma_w^2 = \bar{\sigma}^2 + \sum \rho_i (\sigma_i - \bar{\sigma})^2, \quad \bar{\sigma} = \sum \rho_i \sigma_i$$

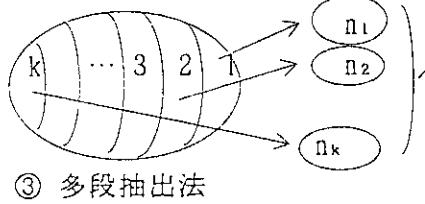
① 比例配分

ｋ個の層に分けられている母集団から、各層の大きさに比例した大きさの標本を抽出する方法、即ち、 $\rho_i = \frac{N_i}{N} = \frac{n_i}{n}$, 標本平均 $\bar{x}_s = \sum \rho_i \bar{x}_i$ （$\bar{x}_i$は第i層からの標本平均）

$$\text{分散} \quad V(\bar{x}_s) = \sum \rho_i \frac{N_i - n_i}{N_i - 1}\cdot\frac{\sigma_i^2}{n} = \sum \rho_i \cdot \frac{\sigma_i^2}{n}$$ 実は、$V(\bar{x}) \geq V(\bar{x}_s)$

が成り立つ．即ち、単純ﾗﾝﾀﾞﾑｻﾝﾌﾟﾘﾝｸﾞより層別抽出法が分散を小さくできる．

② 最適配分（*Neyman* 配分）



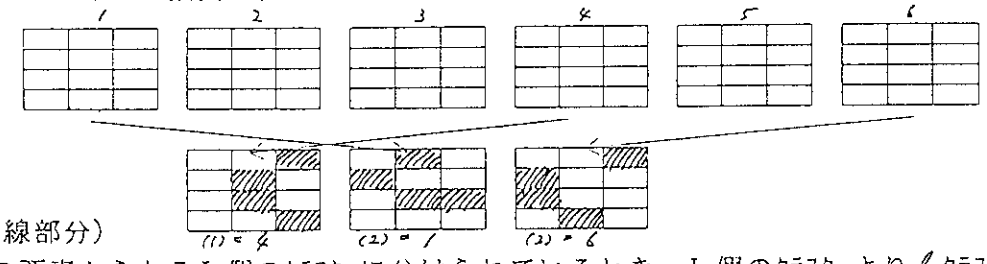
$$n_1 + n_2 + \cdots + n_k = n$$

$$n_i = \frac{\rho_i \sigma_i}{\rho_1\sigma_1 + \rho_2\sigma_2 + \cdots + \rho_k\sigma_k}\cdot n, (i = 1, 2, \cdots, k) \text{とおくと,}$$

$$V(\bar{x}_s) = \sum \rho_i^2 \frac{N_i - n_i}{N_i - 1}\cdot\frac{\sigma_i^2}{n_i} = \sum \rho_i^2 \cdot \frac{\sigma_i^2}{n_i} \quad \text{最小} \rightarrow \frac{(\sum \rho_i \sigma_i)^2}{n}$$

③ 多段抽出法

a．二段抽出法：第１段階として全母集団から適当な大きさの中間的抽出集落（ｸﾗｽﾀｰ）をいくつか抽出し、次に本来の調査個体を抽出する方法



母集団がＮ個の要素からなるＬ個のｸﾗｽﾀｰに分けられているとき、Ｌ個のｸﾗｽﾀｰよりℓｸﾗｽﾀｰ抽出し、抽出された ｸﾗｽﾀｰからそれぞれｎ個の標本を選び、標本平均 $\bar{\bar{x}}_{(1)}, \bar{\bar{x}}_{(2)}, \cdots, \bar{\bar{x}}_{(\ell)}$ を求め、

$$\bar{\bar{x}} = \frac{\bar{\bar{x}}_{(1)} + \bar{\bar{x}}_{(2)} + \cdots + \bar{\bar{x}}_{(\ell)}}{\ell}, \quad E(\bar{\bar{x}}) = \mu \ (\text{母平均})$$

$$V(\bar{\bar{x}}) = \frac{N-n}{N-1}\cdot\frac{\tilde{\sigma}_w^2}{\ell n} + \frac{\tilde{\sigma}_b^2}{\ell},$$

$$\tilde{\sigma}_w^2 = \frac{1}{L}\sum_j^L \sigma_j^2, \quad \tilde{\sigma}_b^2 = \frac{1}{L}\sum(\mu_j - \mu)^2$$

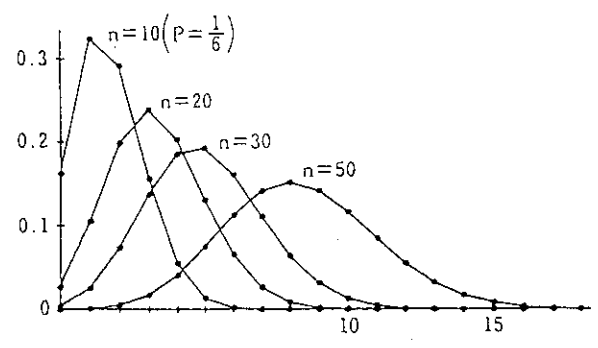
C．確率分布

1．二項分布（*Binomial distribution*）　B（np, npq），$q=1-p$

二項定理によって、$(p+q)^n = q^n + npq^{n-1} + \frac{n(n-1)}{2!}p^2q^{n-2} + \cdots + \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}p^kq^{n-k} + \cdots + np^{n-1}q + p^n$

二項分布とは離散型変数の分布で、表と裏、１と０のように２つのうちどちらかを選択する場合の確率分布である。この場合、ｐ＋ｑ＝１，ｐ，ｑ≧０である。例えば、ｎ回のうちｋ回は事象Ａが起き（確率ｐ）、（ｎ－ｋ）回は事象Ａ以外が起きる（確率：１－ｐ）場合の確率をＰ（ｋ）とすると、$P(k) = \binom{n}{k}p^kq^{n-k} = \frac{n!}{k!(n-k)!}p^kq^{n-k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}p^kq^{n-k}$

で示される。このとき、二項分布の平均値は　ｎｐ，分散は　ｎｐ（１－ｐ）＝ｎｐｑ．



サイコロの目の出る確率

n＝10の確率

x	${}_nC_xp^xq^{n-x}$
0	0.161506
1	0.323011
2	0.290710
3	0.155045
4	0.054266
5	0.013024
6	0.002171
7	0.000248
8	0.000019
9	0.000001
10	0.000000

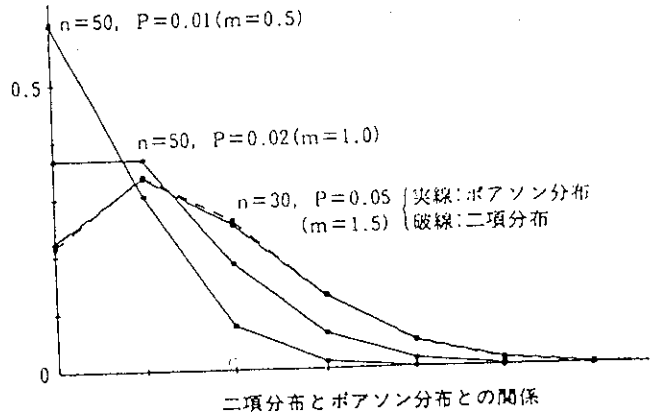
2．ポアソン分布（*Poisson distribution*）　Ｐ（λ）

ポアソン分布も離散型変数の分布で、例えば、交通事故や疾病による死亡数のような希に起こる事象の分布である。二項分布の平均値ｎｐにおいて、ｐが非常に小さく０に近く、ｎが大きいがｎｐ＜５の場合、このポアッソン分布によく当てはまる。ポアソン分布の確率をＰ（ｘ）とすると、$P(x) = e^{-\lambda}\frac{\lambda^x}{x!}, x=0,1,2,\cdots$

ここに、ｅ（自然対数の底）＝2.71828‥‥

ポアソン分布の平均値λ，分散λである。

二項分布で平均ｍ＝ｎＰを一定にし，ｎ→∞，Ｐ→０の極限型がポアソン分布であるが，ｎＰ＜10かつｎ≧50の場合，正規分布に近似できる。



二項分布とポアソン分布との関係

3．正規分布（*Normal distribution*）　Ｎ（μ，σ^2）

正規分布は連続型変数の分布である。最もよく使われる分布である。自然界の現象でこの正規分布が適用されるケースが非常に多い。正規分布の確率（この場合、確率密度関数とい

う） f (X)で示すと、 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty$

また、正規分布をN（μ，σ^2）で表す。正規分布の平均値はμ、分散はσ^2である。正規分布のグラフはｘ＝μで左右対称である。正規分布関数は

$$F(x)=\int_{-\infty}^{x}f(x)dx=P(-\infty<X<x)=1-P(x\leq X)$$

分布形の面積が確率を表している。全面積は１である。ｚ＝（ｘ－μ）／σの変換を施せば、N（μ，σ^2）はN（０，１）に変わる。即ち、平均０、分散１の正規分布である。これを規準（標準）正規分布という。

○正規確率紙

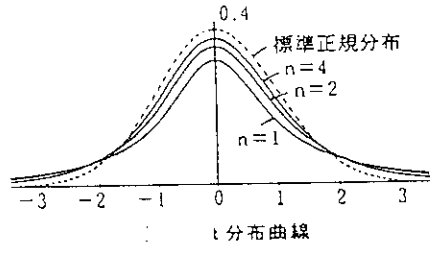
作成した分布が正規分布しているかどうかを調べるためのもので、正規分布に近ければ、それぞれの確率点を正規確率紙にプロットしたものが、ほぼ直線になる。より厳密に確かめるにはχ²－分布による検定を行う。

自由度がｎのｔ分布の基本的な性質は以下のとおりである。

①分布の全面積は１である
②ｔ＝０で極大値をとる
③分散は $\frac{n}{n-2}$ (n>2)

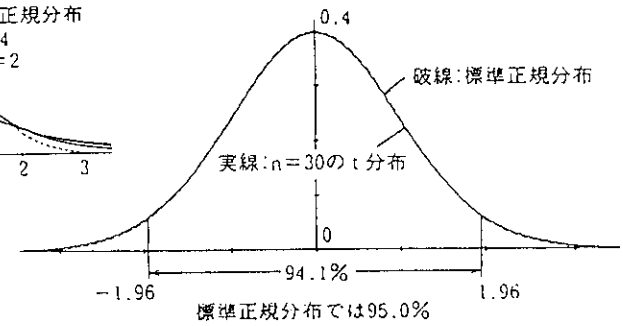
確率密度関数は、ガンマΓ関数(p. 36)を用いて，

$$y=\frac{\Gamma\left(\frac{n+1}{2}\right)}{n\pi\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}$$



ｔ分布曲線

（※ｎは自由度）



ｔ分布と標準正規分布との比較

4．ｔ－分布　TDIST

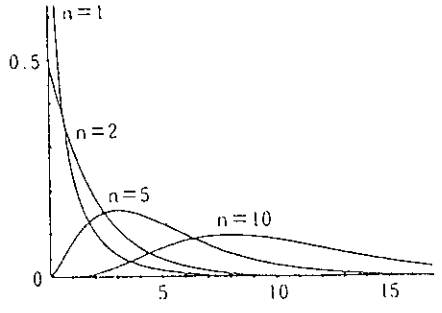
標本平均$\bar{x}$の分布はｎが大きいときはN（μ，σ^2／ｎ）に従う。σが未知のとき、σの代わりに標本分散s^2を使うことがよくある。しかし、標本数ｎ≦25の場合、$t=\frac{\bar{x}-\mu}{\sqrt{s^2/n}}$は、自由度$\nu$＝ｎ－１のｔ－分布に従う。ｔ－分布はｎ→大にすると、正規分布に近づいていく。ｔ－分布のグラフは左右対称の分布である。

5．カイ２乗分布（χ^2－分布）（*Chi−square distribution*）　CHIDIST, 逆 CHIINV

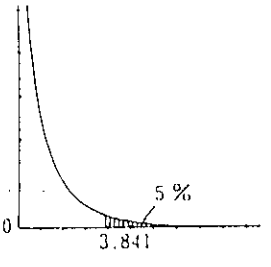
２つ以上の選択しがあるケース、あるいは正規分布に従う変数Ｘの２乗X^2はχ^2－分布に従う。χ^2－分布は自由度νをもち、ν（*0*、*1*、*2*、‥）の大きさで分布曲線が異なってくる。大きさｎの標本が正規母集団（正規分布に従う母集団）から抽出されたものであるなら、そこから求められる変数、$\frac{nS^2}{\sigma}=[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]/\sigma=\frac{\sum x^2-n\bar{x}^2}{\sigma}$ は自由度ν＝ｎ－１のχ^2分布に従う。

表の縦・横の項目に関連がない時，互いに「独立である」という。

平均m、分散S^2の母集団から大きさｎの標本を無作為抽出する時，$\chi^2_n=\frac{1}{S^2}\Sigma(x_i-m)^2$［≧0］は自由度ｎのカイ２乗分布に従う。①定義により上側確率しか存在しない。確率密度はガンマ関数を用いるが省略する。②ｎによって形が大きく異なり，ｎが大きくなると左右対称に近くなる。



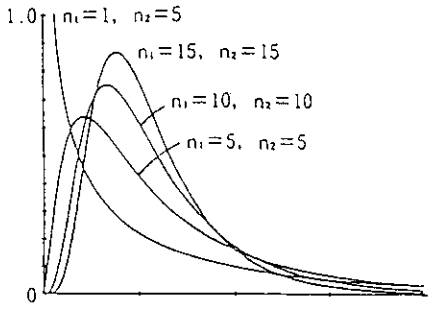
カイ２乗分布曲線



自由度１の棄却域

6．Ｆ－分布

標本分散s^2の検定によく用いられる分布である。例えば、分散がσ_1^2，σ_2^2の２つの正規母集団から、大きさｍとｎの標本を取ったとき、その標本分散をs_1^2,s_2^2とすると$F=\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$は自由度$\nu_1=m-1$、$\nu_2=n-1$のＦ－分布に従うことが分かっている。これを使えば、分散の大きさの違いを比較検定することができる。



Ｆ分布の自由度と形状

Ｆ分布は左右対称でないので，正規分布やｔ分布と異なり，棄却域も図に示すように，上側と下側で左右対称となっていない。

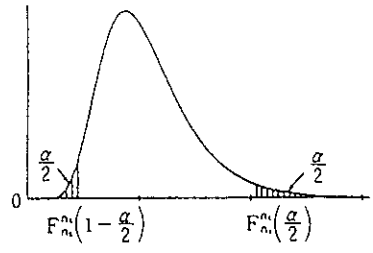
しかし，$F_{n2}^{n1}\left(1-\frac{\alpha}{2}\right)=\frac{1}{F_{n1}^{n2}\left(\frac{\alpha}{2}\right)}$

の関係が成立するので，等分散の検定の棄却域は，

$$\frac{u_1^2}{u_2^2}\geqq F_{n2}^{n1}\left(\frac{\alpha}{2}\right),\quad \frac{u_1^2}{u_2^2}\leqq F_{n2}^{n1}\left(1-\frac{\alpha}{2}\right)=\frac{1}{F_{n1}^{n2}\left(\frac{\alpha}{2}\right)}$$

つまり，$\frac{u_1^2}{u_2^2}\geqq F_{n2}^{n1}\left(\frac{\alpha}{2}\right)$，または$\frac{u_2^2}{u_1^2}\geqq F_{n1}^{n2}\left(\frac{\alpha}{2}\right)$の判定となるため，大きい方の不偏分散を分子として，上側確率だけで判定してよい。

また，ｔ分布との関係は，$\left\{t_n\left(\frac{\alpha}{2}\right)\right\}^2=F_n^1(\alpha)$である。



Ｆ検定の棄却域

二項確率をｆ(x)とすると，コンピュータでの計算は簡単である。

$f(0)=(1-P)^n$，$f(i+1)=\frac{n-i}{i+1}\cdot\frac{P}{1-P}\cdot f(i)[i=0\sim n-1]$

ただし，ｎ，Ｐがともに大きいと，計算精度が問題となる。手計算ではｎが大きいと階乗！の計算は無理なので，二項係数${}_nC_x$の数値表が作られているほか，他の分布へのさまざまな変換が試みられてきた。

① 正規分布への近似：二項分布でＰ＝Ｑ＝$\frac{1}{2}$でかつｎ→∞の連続型は正規分布となる。そこで，ｎＰ≧５（かつｎＱ≧５）の時，二項分布を正規分布に近似させてよいとされている。

② 逆正弦変換：多くの比率を扱う場合，θ（ラジアン）＝$\sin^{-1}\sqrt{P}$と変換すると分散は$\frac{1}{4n}$となり，確率とは無関係になる

7．二項分布の正規近似

○二項分布において、ｐ≦1/2のとき、ｎｐ＞５ならば、正規分布に近似する。また、ｐ＞1/2のときはｎｑ＞５ならば、正規分布に近似すると考えてよい。二項分布の平均値はｎｐ、分散はｎｐｑであるから、規準化して、$z=(x-np)/\sqrt{npq}$にすると、

ｚは正規分布$N(0,1)$に従う。即ち、平均０、分散１の正規分布に従う。

○ｚを更に、ｎで割って、$z=(\frac{x}{n}-p)/\sqrt{\frac{pq}{n}}$は、割合ｘ／ｎが平均ｐ、分散ｐｑ／ｎの正規分布が使えることになる。

○ｘがＮ(μ, σ^2)に従うなら、標本数ｎのランダムサンプリングによる$\bar{x}$は$\bar{x}\sim N(\mu,\frac{\sigma^2}{n})$であることがわかっている。これを$\bar{x}$の標本分布という。更に、ｎがかなり大きい数、例えば、ｎ≧50、実用的にはｎ≧25なら、ｘが正規分布していなくても平均μ、分散σ^2の分布でありさえすれば、$\bar{x}\sim N(\mu,\frac{\sigma^2}{n})$であることもわかっている。【中心極限定理といって 統計学では重要な法則の一つである。この法則があるため、あまり厳密な検証なしに$\bar{x}$の推定・検定ができるのである。】

D．推定と検定

1．母数（*Parameter)*

母集団を特徴づける変量をさす。例えば、母集団が正規分布Ｎ（μ，σ^2）に従うとき、この正規分布を決定づける変量μ、σを母数（ﾊﾟﾗﾒｰﾀ）という。μ、σが分かれば、その母集団は既知と考える。母集団が二項分布Ｂ（ｎ，ｐ）に従うなら母数はｐである。ｐが決まれば、試行ｎにより二項分布は決まる。

2．統計量（*Statistic)*

母数を推定するために標本から得られる任意の量のことである。例えば、母集団から抽出された$x_1, x_2, \cdots, x_n$は標本であるが、平均$\bar{x}=(x_1+x_2+\cdots+x_n)/n$、分散ｓ^2，標準偏差ｓなどは、標本から計算された統計量である。そして、この統計量の分布が考えられる。

3．点推定 *(Point estimation)*

一点で与えられる母数の推定を点推定という。例えば、母平均μの推定値として$\bar{x}$、Ｍｅ、Ｍｏ等を用いることができるが、どれがμの良い推定量であるかは、多数の繰り返し実験のなかで、どの統計量の値がμの代用値として適切であるかということである。良い推定量の基準として不偏推定量があげられる。不偏推定量はＥ（ｘ）＝μと記す。Ｅ（ｘ）は平均$\bar{x}$の標本分布の平均のことで、期待値とよび、これがμに等しいとき$\bar{x}$はμの不偏推定量であるという。そのような不偏推定量がいくつかある場合、最小の分散をもつものが最も有効であると考える。例えば、$\bar{x}$とＭｅは共に同じ母平均をもっている。しかし、$\bar{x}$の標本分布の分散はσ^2／ｎ、Ｍｅの標本分布の分散は$\pi\sigma^2/(2n)$で、σ^2／ｎの方が小さい。従って、$\bar{x}$の方がＭｅより有効な推定量といえる。また、分散$s'=(\sum x_i'-n\bar{x}')/(n-1)$は$E(s)=\sigma'$であることがわかっている。ｓ^{2}は$\sigma^2$の不偏推定量である。それゆえ、このｓ^{2}を不偏分散とよんでいる。

4．区間推定 *(Interval estimation)*

母数が与えられた２点の間に存在すると考える推定を区間推定という。一般に、推定の信頼度を９５％あるいは９９％に置く。その信頼度における推定区間を信頼区間という。この区間の両端を信頼限界値という。信頼度を高めようとすれば、推定区間は広くなる。

$\bar{x}\sim N(\mu, \frac{\sigma^2}{n})$のとき、母平均$\mu$の９５％信頼限界は$\bar{x}\pm 1.96\frac{\sigma}{\sqrt{n}}$、真の平均$\mu$は、９５％の信頼確率で、$\bar{x}-1.96\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+1.96\frac{\sigma}{\sqrt{n}}$ の間に存在するといえる。

1）正規分布によるμの区間推定

９５％の信頼区間は　$\bar{x}-1.96\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+1.96\frac{\sigma}{\sqrt{n}}$

９９％の信頼区間は　$\bar{x}-2.576\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+2.576\frac{\sigma}{\sqrt{n}}$

2）ｔ－分布によるμの区間推定

$\bar{x}-t_{n-1}(\alpha)\frac{s}{\sqrt{n}}<\mu<\bar{x}+t_{n-1}(\alpha)\frac{s}{\sqrt{n}}$，　$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$，$t_{n-1}(\alpha)$は自由度$\phi=n-1$，危険率$\alpha$のｔ－分布値

例えば、ｎ＝２１のとき、$n-1=20$、ｔ－分布９５％値は2.086、９９％値は2.845

3）割合ｐの区間推定

ｎ回の試行で事象Ａがｘ回起これば、割合の実現値　$\hat{p}=\frac{x}{n}$　、ｎ→大のとき、$\hat{p}$はＮ（ｐ．ｐｑ／ｎ）に従うので、９５％信頼区間は

$$\hat{p}-1.96\sqrt{\frac{pq}{n}}<p<\hat{p}+1.96\sqrt{\frac{pq}{n}}$$

、ｑ＝１－ｐ

９９％信頼区間の場合は、１．９６の代わりに２．５７６とすればよい。

5．標本数の決定

1）比率の場合

母集団の大きさが有限Nであるとき、比率pの標準偏差$\sigma_p = \sqrt{\frac{N-n}{N-1}\cdot\frac{p(1-p)}{n}}$より、$p$の９５％信頼限界は$p$の代わりに$\hat{p}$を用いて、$\hat{p} \pm 1.96\sqrt{\frac{N-n}{N-1}\cdot\frac{\hat{p}(1-\hat{p})}{n}}$、従って、母数の推定誤差を$d$とすれば、$\hat{p} \pm d$と書けるので、$n = \frac{N}{(\frac{d}{1.96})^2\cdot\frac{N-1}{\hat{p}(1-\hat{p})}+1} < \frac{N}{(\frac{d}{1.96})^2\cdot\frac{N-1}{0.5^2}+1}$

例えば、$d = \hat{p}/10$、即ち、母比率の１０％とおくと、$\hat{p} = 0.3$なら$d = 0.03$
信頼度９９％の場合は、１．９６の代わりに２．５７６とする。ｄは精度であるので適当に設定すればよい。

2）平均値の場合

大きさNの有限母集団の平均値$\bar{x}$の標準偏差は$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}}$であるから、同様に$d = 1.96\sqrt{\frac{N-n}{N-1}\cdot\frac{\sigma^2}{n}}$とおいて、$n = \frac{N}{(\frac{d}{1.96})^2\cdot\frac{N-1}{\sigma^2}+1}$

何らかの形でσの値を推定して、ｎを求める。精度ｄを小さくすれば、ｎは大きくなり、σが大きければ、ｎも大きくなることがわかる。

6．検定

1）仮説検定（*Test of hypothesis)*

仮説検定とは、ある仮説の下で母集団の母数、例えば、母平均μ、母分散σ^2、母比率ｐ等同士の比較あるいは母集団の母数と標本統計量、即ち、標本平均$\bar{x}$、標本分散s^2、標本比率$\hat{P}$等を比較し、確率分布による判定を行うことをさす。検定においては、まず、仮説を立て、設定したある確率の下でこの仮説が棄てられるかどうかを判定する。この設定した確率を有意水準または危険率という。

2）有意水準（*Significant level)*

仮説は棄てられることに有意な（偶然に起こったものとは認められない）意味を積極的に含むものである。その意味で、帰無仮説といい、判定する確率を有意水準とよぶ。危険率という言葉は、この仮説が実は正しいにも拘わらず、誤ってそれを棄てる危険を犯す場合があり、その確率をさす。この誤りを第１種の過誤（α）、生産者危険率、あわてものの誤りなどという。また、この仮説が実は正しくないにも拘わらず、誤って採用する危険も含んでいる。この誤りを第２種の過誤（β）、消費者危険率、ぼんやりものの誤りなどという。この２つの誤りα、βは共に小さいに越したことはないが、同時に最小にする方法はない。特に、βに対する検定法がない。そこで、一般に、α＝０．０５（５％）あるいは０．０１（１％）に設定した上で、βを小さくするには標本の大きさｎを大きくする。

3）両側検定と片側検定（*Two−tailed test, One−tailed test)*

この仮説検定において、仮説を否定した場合は何を採用するかが問題になる。これが対立仮説である。例えば、薬効試験の場合、仮説は薬効なしとすれば対立仮説は薬効ありとなる。血圧降下剤試験であれば、薬効ありなら投与後、明らかに血圧は下がっていることになるから、投与前と投与後の母数には大小関係、即ち、母数に関する事前情報があることになる。母数に関する事前情報がない場合は仮説検定が両側検定となり、事前情報がある場合は片側検定が考えられる。

4）検定手順

仮説検定の手順は次のように行う。

① 仮説H_0、対立仮説H_1を立てる。
② 仮説の下での標本統計量ｚを計算する。
③ 有意水準（または危険率）５％に対応する分布値ｚ*(0.05)*、有意水準１％に対応する分布値ｚ*(0.01)*を表より決定する。
④ ｚ＜ｚ*(0.05)*なら危険率５％で有意差は認められない。即ち、仮説はすてられない。
⑤ ｚ*(0.05)*≦ｚ＜ｚ*(0.01)*なら危険率５％で有意差ありとする。即ち、仮説は棄却される。
⑥ ｚ*(0.01)*≦ｚなら危険率１％で有意差ありとする。即ち、仮説ば棄却される。

7．分布による仮説検定

1）平均値の検定（分散σ^2未知、ｎ≧２５のとき）

① 仮説Ｈ$_0$：$\mu=\mu_0$、Ｈ$_1$：$\mu\neq\mu_0$

② 仮説の下では、 $z=\dfrac{|\bar{x}-\mu_0|}{\sqrt{\dfrac{s^2}{n}}}$ は規準正規分布$N(0,1)$に従う。

③ 正規分布Ｎ(0,1)の危険率５％点は

ｚ(0.05)=1.96、１％点ｚ(0.01)=2.576

④ ｚ＜*1.96*なら危険率５％で有意差は認められない。

⑤ *1.96*≦ｚ＜*2.576*なら危険率５％で有意差あり。

⑥ *2.576*≦ｚなら危険率１％で有意差あり。

2）ｔ－検定（分散σ^2未知、ｎ＜２５のとき）

① 仮説Ｈ$_0$：$\mu=\mu_0$、Ｈ$_1$：$\mu\neq\mu_0$

② 仮説の下では、 $t_0=\dfrac{|\bar{x}-\mu_0|}{\sqrt{\dfrac{s^2}{n}}}$ は自由度$(n-1)$のｔ－分布に従う。

③ 自由度（ｎ－１）のｔ－分布表より、

ｔ$_{n-1}$*(0.05)*、ｔ$_{n-1}$*(0.01)*を求める。

④ ｔ$_0$＜ｔ$_{n-1}$*(0.05)*なら危険率５％で有意差は認められない。

⑤ ｔ$_{n-1}$*(0.05)*≦ｔ$_0$＜ｔ$_{n-1}$*(0.01)*なら危険率５％で有意差あり。

⑥ ｔ$_{n-1}$*(0.01)*≦ｔ$_0$なら危険率１％で有意差あり。

3）対応のある場合の検定

２つの標本群が同一人の投薬前と投薬後のようにデータｘ$_1$とｙ$_1$、ｘ$_2$とｙ$_2$のように互い対応関係があるとき、それぞれの差ｚ$_1$＝ｘ$_1$－ｙ$_1$、ｚ$_2$＝ｘ$_2$－ｙ$_2$、・・・、ｚ$_n$＝ｘ$_n$－ｙ$_n$をつくる。そして、ｚ$_1$、ｚ$_2$、・・・、ｚ$_n$について $\bar{z}=\frac{1}{n}\sum_{i=1}^{n}z_i$、 $s^2=\frac{1}{n-1}(\sum_{i=1}^{n}z_i^2-n\bar{z}^2)$ を求める。以下は２）ｔ－検定を適用する。（ｎ≦２５のとき）

4）２つの割合の差の検定（$m\hat{p}_1\geq 5, n\hat{p}_2\geq 5$のとき）

母集団Ａ：母比率ｐ$_1$・・・・標本数ｍ、割合 $\hat{p}_1$

母集団Ｂ：母比率ｐ$_2$・・・・標本数ｎ、割合 $\hat{p}_2$

A,B全体の母比率ｐ＝$\dfrac{mp_1+np_2}{m+n}$で推定する。このとき、

分散s^2＝ｐｑ$(\dfrac{1}{m}+\dfrac{1}{n})$, $p+q=1, p,q\geq 0$である。

① 仮説Ｈ*0*：ｐ$_1$＝ｐ$_2$、対立仮説Ｈ$_1$：ｐ$_1$≠ｐ$_2$

② 仮説の下では、 $z=\dfrac{|p_1-p_2|}{\sqrt{\text{ｐｑ}(\dfrac{1}{m}+\dfrac{1}{n})}}$ は規準正規分布$N(0,1)$に従う。

③ ｚ＜*1.96*なら危険率５％で有意差は認められない。

④ *1.96*≦ｚ＜*2.576*なら危険率５％で有意差あり。

⑤ *2.576*≦ｚなら危険率１％で有意差あり。

5）２つの平均値の差の検定（ｎ＜２５のとき）

母集団Ａ：母平均μ_1、母分散$\sigma^2{}_1$・・・標本数ｍ、平均$\bar{x}_1$、分散s_1^2

母集団Ｂ：母平均μ_2、母分散$\sigma^2{}_2$・・・標本数ｎ、平均$\bar{x}_2$、分散s_2^2

① 仮説Ｈ$_0$：$\mu_1=\mu_2$、対立仮説Ｈ$_1$：$\mu_1\neq\mu_2$

② 仮説の下では、$\sigma^2{}_1=\sigma^2{}_1$のとき、

$$t=\frac{|\bar{x}_1-\bar{x}_2|}{\sqrt{\frac{(m-1)s_1^2+(n-1)s_2^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}}$$ は自由度$\phi=m+n-2$のｔ－分布に従う。

③ 自由度ｍ＋ｎ－２のｔ－分布表よりｔ(0.05)、ｔ(0.01)を求める。

④ ｔ＜ｔ(0.05)なら、危険率５％で有意差は認められない。

⑤ ｔ(0.05)≦ｔ＜ｔ(0.01)なら、危険率５％で有意差あり。

⑥ ｔ(0.01)≦ｔなら、危険率１％で有意差あり。

6）独立性の検定（H_0：A,B間に関連がない）適合度の検定

①$k>2$の場合

	B_1	B_2	B_3	$\cdots$	B_k	$Total$
A_1	n_{11}	n_{12}	n_{13}	$\cdots$	n_{1k}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	n_{23}	$\cdots$	n_{2k}	$n_{2\cdot}$
$Total$	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	$\cdots$	$n_{\cdot k}$	n

$$\chi_0^2=\frac{n^2}{n_1 n_2}\left(\frac{n_{11}^2}{n_{\cdot 1}}+\frac{n_{12}^2}{n_{\cdot 2}}+\frac{n_{13}^2}{n_{\cdot 3}}+\cdots+\frac{n_{1k}^2}{n_{\cdot k}}-\frac{n_{1\cdot}^2}{n}\right)$$

$>\chi_{k-1}^2(\alpha)$のとき　A,Bは独立ではない.

②ｋ＝２の場合

	B_1	B_2	$Total$
A_1	a	b	g
A_2	c	d	h
$Total$	e	f	n

$$\chi_0^2=\frac{(|ad-bc|-n/2)^2 n}{efgh}$$

≧3.84なら危険率５％で有意差あり。

≧6.63なら危険率１％で有意差あり。

E．相関分析

1．相関係数ｒの検定

1）標本が１組（Ｈo ： $\rho = 0$ の場合）

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \cdot \sum (y-\bar{y})^2}} = \frac{\sum xy - n\bar{x}\bar{y}}{\sqrt{(\sum x^2 - n\bar{x}^2)\cdot(\sum y^2 - n\bar{y}^2)}}$$

$F_0 = \frac{(n-2)r^2}{1-r^2}$ は自由度 $\nu_1 = 1, \nu_2 = n-2$ のＦ(α)分布に従う。ここに、αは有意水準100α%である。

2）標本が１組の場合（Ｈo ： $\rho = \rho_0$ の場合）

２次元正規母集団からの標本とみなす。

① Ｚ変換　$r^* = \frac{1}{2} ln \frac{1+r}{1-r}$,　$\rho^* = \frac{1}{2} ln \frac{1+\rho}{1-\rho}$　（lnはeを底とする自然対数）

② $Z = \sqrt{n-3}(r^* - \rho^*)$のとき$|Z| \sim N(0,1)$
即ち、$|Z| \geq 1.96$のとき有意水準5%で仮説を棄却する。

③ ｒの95％信頼限界は

下限値　$r_1^* = r^* - \frac{1.96}{\sqrt{n-3}}$, 上限値　$r_2^* = r^* + \frac{1.96}{\sqrt{n-3}}$ を求め、

$r_1^* = \frac{1}{2} ln \frac{1+r_1}{1-r_1}, r_2^* = \frac{1}{2} ln \frac{1+r_2}{1-r_2}$ を満たすr_1, r_2を信頼下限、信頼上限とする。

即ち、$r_1 = \frac{exp(2r_1^*) - 1}{exp(2r_1^*) + 1}$,　$r_2 = \frac{exp(2r_2^*) - 1}{exp(2r_2^*) + 1}$

3）標本が２組（Ｈo ： $\rho_1 = \rho_2$の場合）

① $r_1^* = \frac{1}{2} ln \frac{1+r_1}{1-r_1}$,　$r_2^* = \frac{1}{2} ln \frac{1+r_2}{1-r_2}$ を求める。

② $Z = \frac{r_1^* - r_2^*}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$　（n_iは第i標本のﾃﾞｰﾀ数, $i = 1,2$）

$|Z| \geq 1.96$ なら有意水準５％で仮説を棄却する。即ち、母相関係数 $\rho_1 \neq \rho_2$とする。
あるいは、

$Z^2 = \frac{(r_1^* - r_2^*)^2}{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \sim \chi_1^2(\alpha)$　即ち、Z^2は自由度１のχ^2分布に従う。

この$Z^2 \geqq 3.84$　（$= 1.96^2$）なら仮説を棄却する。

4）標本がｋ組（Ｈo ： $\rho_1 = \rho_2 = \cdots = \rho_k$の場合）

標本１	標本２	・・・・	標本ｋ	対立仮説Ｈ$_1$ ： $\rho_i \neq \rho_j$ $(i \neq j)$
r_1	r_2		r_k	
n_1個	n_2		n_k	

① $r_i^* = \frac{1}{2} ln \frac{1+r_i}{1-r_i} (i = 1,2,\cdots,k)$

② $\chi_0^2 = \sum_i^k (n_i - 3) r_i^{*2} - \frac{(\sum_i^k (n_i - 3) r_i^*)^2}{\sum_i^k (n_i - 3)}$

このχ$_0^2$はχ$^2_{k-1}(\alpha)$分布に従う。従って、自由度$(k-1)$のχ^{2}分布の5%点をχ$^2_{k-1}(0.05)$とすると、
χ$_0^2 \geq \chi^2_{k-1}(0.05)$のとき有意水準5%で仮説を棄却する。

2．偏相関係数の検定（p変量の場合）

1） 偏相関係数 $r_{12\cdot3}$ は変数3の値がすべて同一である個体の切断面内の変数1,2の間の相関である。

$$r_{12\cdot3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = r \text{とおく。}$$

①仮説H_o：$\rho_{12\cdot3} = 0$

$t = \frac{r\sqrt{n-3}}{\sqrt{1-r^2}}$ とおくと、$|t| \geq t(n-3, 0.05)$なら、5%で仮説を棄却する。

t(n-3, 0.05)は自由度n-3の両側5%点である。

注）変数x,y,zがあって、yのxに対する回帰式$y' = a_1 + b_1 x$を求める。残差$d_{yi} = y_i - y'_i$ $(i=1,2,..,n)$が得られる。同様に、zのxに対する回帰式$z' = a_2 + b_2 x$を求める。残差$d_{zj} = z_j - z'_j$ $(j=1,2,...,n)$が求められる。残差の組$(d_{y1}, d_{z1}), (d_{y2}, d_{z2}), \cdots, (d_{yn}, d_{zn})$について、単相関を求める。それを$r_{yz\cdot x}$で表すのである。

2）偏相関係数 $r_{12\cdot34}$（4変数の場合）

$$r_{12\cdot34} = \frac{r_{12\cdot4} - r_{13\cdot4} r_{23\cdot4}}{\sqrt{(1-r_{13\cdot4}^2)(1-r_{23\cdot4}^2)}} = \frac{r_{12\cdot3} - r_{14\cdot3} r_{24\cdot3}}{\sqrt{(1-r_{14\cdot3}^2)(1-r_{24\cdot3}^2)}} = r$$

$t = \frac{r\sqrt{n-4}}{\sqrt{1-r^2}}$ と置くと、$|t| \geq t(n-4, 0.05)$なら有意水準

5%で仮説を棄却する。

即ち、仮説Ho：$\rho = \rho_{12\cdot34} = 0$を捨て、$\rho_{12\cdot34} \neq 0$とするのである。

3．平行性検定法

2組のデータ列$(x_{s1}, x_{s2}, \cdots, x_{sn})$に対する$(y_{s1}, y_{s2}, \cdots, y_{sn})$と$(x_{t1}, x_{t2}, \cdots, x_{tm})$に対する$(y_{t1}, y_{t2}, \cdots, y_{tm})$を考える。2つの回帰直線

$$Ys = \bar{y}_s + b_s (x_s - \bar{x}_s)$$
$$Yt = \bar{y}_t + b_t (x_t - \bar{x}_t)$$

結論から述べよう。次の分散分析表を作成する。

要　因	SS	df	Ms
R（共有直線性）	S_R	1	V_R
Dp（非平行性）	S_{Dp}	1	V_{Dp}
R Σ（全直線性）	$S_{R\Sigma}$	2	$V_{R\Sigma}$
e　（残　差）	Se	n+m−4	Ve
Y　（全　体）	Syy	n+m−2	−

仮説Ｈo ： $b_s=b_t$ （平行である） に対して、

$F_{cal}=\dfrac{S_{Dp}}{V_e}\sim F^{1}_{n+m-4}(\alpha)$即ち、$F_{cal}$は自由度$(1,n+m-4)$のＦ分布に従う。

これにより、Ｆ $cal \geqq F(1,n+m-4:\ \alpha)$ なら $100\ \alpha$％で仮説を棄却する。
即ち、$b_s \neq b_t$
とする。２つの直線は平行ではないとする。Ｆ $cal<F(1,n+m-4:\ \alpha)$ なら平行とはいえない。

$Ys=\bar{y}_s+b_s(x_s-\bar{x}_s)$において、$Sxxs=\sum x_{si}^2-\dfrac{(\sum x_{si})^2}{n}$

$$Syys=\sum y_{si}^2-\frac{(\sum y_{si})^2}{n},\quad Sxys=\sum x_{si}y_{si}-\frac{(\sum x_{si})(\sum y_{sj})}{n},$$
$$b_s=\frac{Sxys}{Sxxs},\quad S_{Rs}=\frac{(Sxys)^2}{Sxxs}=b_s\cdot Sxys,\quad Ses=Syys-S_{Rs}\quad (s\text{の残差})$$

$Yt=\bar{y}_t+b_t(x_t-\bar{x}_t)$についても同様に、$Sxxt=\sum x_{ij}^2-\dfrac{(\sum x_{ij})^2}{m}$

$$Syyt=\sum y_{ti}^2-\frac{(\sum y_{ti})^2}{m},\quad Sxyt=\sum x_{ti}y_{ti}-\frac{(\sum x_{ti})(\sum y_{tj})}{m},$$
$$b_t=\frac{Sxyt}{Sxxt},\quad S_{Rt}=\frac{(Sxyt)^2}{Sxxt}=b_t\cdot Sxyt,\quad Set=Syyt-S_{Rt}\quad (t\text{の残差})$$

ｓ、ｔの全体のデータで考える。即ち、ｓとｔをプールする。

$$Sxx=Sxxs+Sxxt,\quad Syy=Syys+Syyt,\quad Sxy=Sxys+Sxyt$$
$$b=\frac{Sxy}{Sxx},\quad S_R=\frac{(Sxy)^2}{Sxx}=b\cdot Sxy$$
$$S_{R\Sigma}=S_{Rs}+S_{Rt},\quad S_{Dp}=S_{R\Sigma}-S_R\quad (\text{非平行性})$$
$$Se=Syy-S_{R\Sigma}=Ses+Set$$
$$Ve=\frac{Se}{n+m-4},\quad V_{Dp}=\frac{S_{Dp}}{1},\quad V_{R\Sigma}=\frac{S_{R\Sigma}}{2}$$

Ｆｃａｌで検定したものは、また、

$t_{cal}=\dfrac{|b_s-b_t|}{\sqrt{V_e\cdot(\dfrac{1}{Sxxs}+\dfrac{1}{Sxxt})}}$が$t_{n+m-4}(\alpha)$に従うことを利用することと同じになる。
$t_{n+m-4}(\alpha)$は自由度$n+m-4$のt分布のα点である。

以上

1998.11.20 & 12.9

統計学演習問題

問題１　小学生 5000人の都市で虫歯保有率 Pを標本比率 Pで推定するとして、その誤差 dを95%信頼度で５%以下にしたい。何人調査すればよいか？
＜ヒント＞　pが不明であれば、p=0.5として標本の大きさ nを求めればよい。

1) ある地方で小学生 200人を無作為に選び、虫歯調査をしたら 137人が虫歯であった。この地方の小学生の虫歯保有率 pの95%信頼区間を求めよ。

問題２　次表は1985年、F県在住女性について栄養調査を実施し、その時得られた脂質栄養素摂取量(g)のデータである。

年　齢	人口	調査数	平均摂取量	標準偏差	比例配分	最適配分
10代	70,000	8	56.9	11.5		
20	167,000	255	51.2	13.5		
30	140,000	1156	49.4	13.4		
40	139,000	1424	47.6	12.4		
50	110,000	1367	44.9	12.8		
60	85,000	532	43.6	12.2		
70	67,000	96	40.5	11.0		
Total	778,000	4838	46.9	13.0	4838	4838

1) 実際の調査では、母平均μの95%信頼区間はいくらか？

2) n=4838について、比例配分(Propotional allocation)で各年齢層の調査数を求めよ。

3) n=4838を最適配分(Neyman allocation)で各年齢層の調査数を求めよ。更に、実際の調査数の問題点を批評せよ。

問題３　ある電気ﾒｰｶｰは蛍光灯の新製法を開発した。新製品から 25本を無作為に抽出し測定したら、平均寿命 1160時間であった。従来の製品仕様書によれば、平均寿命は 1130時間、標準偏差 80時間であった。新製法によって寿命が延びたといえるか 有意水準５%で検定せよ？