基礎統計学 I

統計学入門

東京大学教養学部統計学教室 編

東京大学出版会

基礎統計学 I

統計学入門

東京大学教養学部統計学教室 編

東京大学出版会

ISBN978-4-13-042065-5

C3333 ¥2800E

9784130420655

1923333028000

定価(本体価格2800円＋税)

11.8 〈二項母集団〉 母数 p の二項母集団 $Bi(1,p)$ から，$n=50$ の大きさの標本を抽出して，0 が 23，1 が 27 であった．p の信頼係数 95% の信頼区間を求めよ．

11.9 〈ポアソン母集団〉 1時間毎の受信電話数を記録したところ

$$4,\ 3,\ 5,\ 4,\ 8,\ 2,\ 5,\ 9,\ 3,\ 5$$

であった．ポアソン母集団 $Po(\lambda)$ を仮定して，λ の信頼係数 99% の信頼区間を求めよ．

第12章 仮説検定

「仮説検定」は，統計的仮説の「有意性」の検定である．仮説の下でわれわれが期待するものと，観測した結果との違いを，これらの差が単に「偶然」によって起ったものか否かという見地から，確率の基準で評価する．

仮説検定は推定とならんで統計的推測の理論の双壁であるとともに，他方で，統計的判断の論理学，科学方法論という意味ももつ．

12.1 検定の考え方

有意性検定　推定の考え方は数学的には単純なものであるが，仮説検定も人間の論証感覚を定式化したもので，ごく自然で理解しやすい．

仮説検定 hypothesis testing の目的は，母集団について仮定された命題を，標本にもとづいて，検証することである．たとえば，エンドウ豆の型の度数分布が，理論上の仮説に合致しているかどうかの検証がそれである．

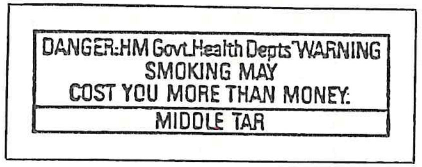
度数の比が厳密に 9：3：3：1 になっていないことはいうまでもない．重要なことは，理論比からのずれが誤差の範囲内か，それ以上の何かの**意味のある**ものか，ということである．後者の場合，統計学では，仮説からのずれ(簡単に，仮説)は**有意** significant であるという．ここで立てられた仮説を**統計的仮**

表 12.1[†] エンドウ豆についてのメンデルの有名な実験データ(メンデルの法則)

型	黄色・丸い	黄色・しわがある	緑色・丸い	緑色・しわがある	計
度　数	315	101	108	32	556
理論比	9	3	3	1	(16)

1865 年のものである．「植物の雑種に関する実験」というものであったが，長く生物学者の関心を引かなかった．この論文の意味と価値が認められたのは，1900 年になってからである．

説 statistical hypothesis または簡単に**仮説** hypothesis という．したがって，仮説検定とは統計的仮説の**有意性の検定** test of significance にほかならない．



イギリスにおける喫煙の警告

Her Majesties' Government Health Department (英国政府保健省)の警告として，「喫煙はタバコ代よりも高くつくかもしれない」と，箱の外側に印刷してあるものである．

統計的仮説は，このように直接的に数理的な命題ばかりではない．たとえば，タバコと肺ガンの因果関係の有無の問題も考えられる．この場合も，ある確率論的関係式が成立することを，統計的仮説にとればよい(練習問題 12.6 参照)．

有意性は標本が有意なずれを示す確率で表される．すなわち，標本分布がそこで役割を果たす．たとえば，コインを 20 回投げたときに 14 回表が出たなら，コインに歪みがないという仮説は支持できるか否か．これを，表の回数 X という標本の確率分布から見よう．これは二項分布 $Bi(20, 1/2)$ である．

もし，「歪みがない」という仮説を $p=1/2$ とすると，「コイン」という母集団分布に対する仮説である．もし，仮にこの仮説が正しいなら，表の回数 X について，表 12.2 の二項分布の計算から

$$\mathrm{P}(X \geqq 14)=1-0.9423=0.0577$$

であるから，$X=14$ という標本は仮説からすれば，出るはずのないかなりはずれた(ずれが有意な)値である．したがって，$p=1/2$ という仮説は「誤っている」と判断せざるを得ない．このとき，仮説は**棄却** reject される，という．つまり，仮説検定とは，仮説が有意であるか否かに応じて，それを棄却するかあるいは棄却しないかを決定することである．

ここで，0.0577 を“稀”と考えたのであるが，一般に，あらかじめどの程度の稀少確率を考えるかにより，有意か否かが変りうる．この基準の確率を**有意水準** significance level といい*，α で表す．たとえば，$\alpha=0.1$ と約束するとき 0.0577 は稀と判断されるが，$\alpha=0.01$ ならこの確率は“十分に起りうる”，“あってもおかしくない”ということだから，$X=14$ は有意にずれている，と

*) 「危険率」という語も用いられるが，理論上は「有意水準」の方がよい．

は考えられない．

帰無仮説と対立仮説　コインを 20 回投げる実験において，コインに歪みがないという仮説 $p=1/2$ は(たとえば)有意水準 $\alpha=0.1$ 棄却された．

棄却されたことで判断が終るという考え方(フィッシャーの立場)もあるが，p について何かの積極的判断をしたいならば，あらかじめ，もう一つの仮説を $p \neq 1/2$ と立てておき，(あるいは $p>1/2$ としてもよい)もう一つの仮説が**採択** accept されたとしよう．

このとき，もとの仮説 $p=1/2$ を**帰無仮説** null hypothesis，これと対立する仮説を**対立仮説** alternative hypothesis という．「帰無」は，もと(最初)の仮説が，棄却されるかされないかの判断にさらすために立てられたもので，‘無に帰することも予定して’くらいの意味である．深い意味はないので，「帰無」の語を無視しても，後の展開に差し支えない．

表 12.2† 表が出た回数の確率分布 N=20, $p=1/2$ である二項分布 $Bi(20, 1/2)$

x	$f(x)$	累積確率分布関数
0	$.9537\times10^{-6}$	.0000
1	$.1907\times10^{-4}$	.0000
2	$.1811\times10^{-3}$	.0002
3	$.1087\times10^{-2}$	.0013
4	$.4621\times10^{-2}$	.0059
5	$.1478\times10^{-1}$	.0207
6	$.3696\times10^{-1}$	.0577
7	$.7392\times10^{-1}$	.1316
8	.1201	.2517
9	.1602	.4119
10	.1762	.5881
11	.1602	.7483
12	.1201	.8684
13	$.7392\times10^{-1}$	.9423
14	$.3696\times10^{-1}$	.9793
15	$.1478\times10^{-1}$	.9941
16	$.4621\times10^{-2}$	.9987
17	$.1087\times10^{-2}$	.9998
18	$.1811\times10^{-3}$	.9999
19	$.1907\times10^{-4}$	.9999
20	$.9537\times10^{-6}$	1.0000

(出典：ヘンケル『有意性検定』)
コインに歪みがないということは，$p=1/2$ と表現されている．

帰無仮説と対立仮説は互いに否定の関係にある．それぞれ H_0, H_1，あるいは H_1, H_2 と記す．H は Hypothesis(仮説*)の意味である．ただし，$p=1/2$，$p>1/2$ 以外の第三の可能性($p<1/2$)もあるが考慮しないこともあり，否定は完全なものではない．

そのことを別にすれば，一応は，帰無仮説を棄却

表 12.3 帰無仮説を棄却する，しないの決定に関しての四つの場合

		真実 H_0 が正しい	H_0 が誤り(H_1 正しい)
決定	H_0 を棄却しない(採択する)	①	②
	H_0 を棄却する	③	④

①〜④の場合が考えられ，①，④は正しいが②，③は誤り．

*) 「仮設」というあて字もある．‘仮に設ける’の意なら，全く誤った語とはいえないであろう．

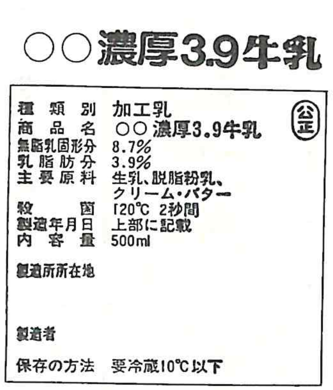
表 12.4　2種類の誤りと品質管理のための抜取検査における例(林)

第一種の誤り	第二種の誤り
H_0 が正しいときにこれを棄却する誤り 〈例〉抜取検査にあたって当然合格するはずの良製品に‘不合格’の判定を下してしまう誤り (生産者のリスク producer's risk)	H_0 が正しくないときにこれを採択する誤り 〈例〉抜取検査にあたって，当然不合格であるはずの不良製品に‘合格’の判定を下してしまう誤り(消費者のリスク consumer's risk)

することは対立仮説を採択することを意味している．帰無仮説を棄却するか，しないかの決定に関しては表 12.3 の四つの場合が考えられる．いま，「棄却しない」を「採択する」といいかえれば，つぎの二つの誤り，(a)帰無仮説が正しいのに，それを棄却する**第一種の誤り** error of the first kind，および(b)帰無仮説が誤っているのに，それを採択する**第二種の誤り** error of the second kind，が考えられる．この考え方は，2×2=4 通りの判断の結果のうち，正しい判断 2 通り，誤った判断 2 通りが生じるという一般的なものであるが，表 12.4 のごとく大量生産の**品質管理**の場面に適用されるほか，刑事訴訟で，無

❖きびしい品質管理—牛乳の例❖

牛乳には，品質が構成成分，処理方式で数量表示されている．とくに，「乳脂肪分」が品目名の一部を構成しており，価格の他に，きびしい数量的品質管理の徹底が商品戦略のかなめであることがわかる．



A 社製品(148 円)

B 社製品(128 円)

(箱の一面から，企業名，POS バーコードを除いて示す．A 社の栄養成分表は他の面に印刷してある)

罪を有罪とする誤り，逆に有罪を無罪とする誤りなど，判断の誤りに対する一般的思考基準を与える．

帰無仮説の採択について　有意性検定は，(帰無)仮説の下でわれわれが期待する結果が生じなかったことを根拠として，仮説を棄却，否定することが，おもな内容である．これは論理学では，**背理法**といわれているものである．あくまで，棄却されることが中心であって，仮説が棄却されなかったからといって，積極的に支持されたわけではない．単に，結果が帰無仮説と「矛盾はしない」ことがいわれただけである．仮説を採択したからといって，仮説が真であることを積極的に「証明」したわけではない．

棄却域と両側・片側検定　仮説検定の考え方と具体的な計算をよりよく理解するために，後にも述べる **t 検定** t-test の例をあげよう．

——正規母集団 $\mathrm{N}(\mu, \sigma^2)$ から大きさ $n=25$ の標本を抽出して，$\bar{X}=13.7$，$s=2.3$ を得た(たとえば，ある飲料中の某成分の%を考えよ)．これをもとに，帰無仮説

$$H_0 : \mu=15$$

を対立仮説

$$H_1 : \mu\neq15$$

に対して，有意水準 $\alpha=0.05$ で検定する．——

検定に用いる統計量(**検定統計量** test statistic)として，(10.15)のスチュー

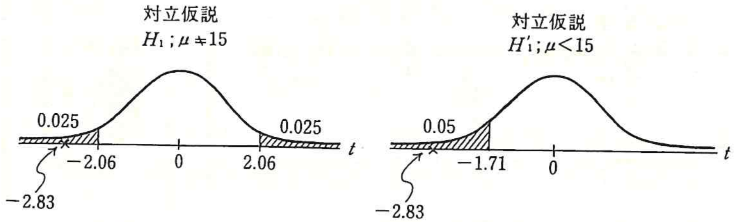


図 12.1　仮説検定のやり方

帰無仮説をまず設け，対立仮説 H_1 を考慮して，あらかじめ選んだ有意水準(5%)に等しい確率で，0 から遠い所に棄却域(斜線を付してある数直線の範囲)を作る．計算された t 統計量の値(−2.83)がその棄却域の中に入るとき，その帰無仮説は棄却される．

デントの t 統計量を計算すると，

$$t=\frac{13.7-15}{2.3/\sqrt{25}}=-2.83$$

となる．有意水準 $\alpha=0.05$ にちょうど対応する t 分布 $t(24)$ のパーセント点は，$t_{0.025}(24)=2.06$ であるから，$|-2.83|>2.06$ より，有意水準 5% で帰無仮説を棄却する．

たしかに $\bar{X}=13.7$，$\mu=15$ であり，その差は 1.3 である．$\bar{X}$ の分散の推定値は s^2/n，標準偏差は $s/\sqrt{n}=2.3/\sqrt{25}=0.46$ であるから，それを基準とすれば差 1.3 はこの標準偏差の 3 倍近くで（$t=2.83$），大きいという他はない．なぜなら，±2.06 以上の差は確率 0.05 という稀なものであって，2.83 はそれ以上に異常な差だからである．したがって，$\mu=15$ と考えることに無理がある．これがいま用いている t 統計量の具体的な意味である．

いま，帰無仮説はそのままとし，対立仮説を

$$H_1':\mu<15$$

とすると，$\bar{X}$ が $\mu=15$ より相当に小さくなった場合にだけ帰無仮説を棄却すればよいから，

(12.1)　$$t=(\bar{X}-15)/(s/\sqrt{n})$$

において，t が十分負になったときにだけ，帰無仮説を棄却すればよい．いま $t_{0.05}(24)=1.71$ であるから，$-2.83<-1.71$．よって，この場合も，有意水準 5% で帰無仮説を棄却する．

帰無仮説を棄却すべき統計量の値の集合を**棄却域** rejection region といい，棄却しない領域を**採択域** acceptance region という．採択域は，$t=0$ の周辺の領域（15 に近い $\bar{X}$ の値に対応）になるが，棄却域は対立仮説が H_1 のような**両側対立仮説** two-sided alternative hypothesis のときは，t の値が著しく（これは有意水準で決める）0 から左右にはずれた領域

(12.2)　$$|t|>t_{\alpha/2}(n-1)$$

となる．これを**両側検定** two-sided test という．確率 α を左右（著しく大きい所と小さい所）に按分している．H_1' のような**片側対立仮説** one-sided alternative hypothesis に対しては，$\bar{X}$ が十分小さい所に棄却域が定められ，**片側検定** one-sided test

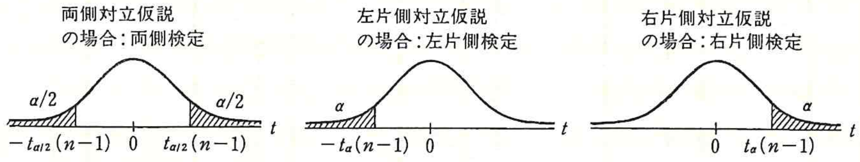


図 12.2　検定の手続例：母平均は帰無仮説の通りか

スチューデントの t 統計量とは $t=(\bar{X}-\mu_0)/(s/\sqrt{n})$ で定義される（μ_0 はあらかじめ指定された母平均値）．関数は自由度 $n-1$ の t 分布 $t(n-1)$ の密度関数．$t=0$ は標本平均 $\bar{X}$ と帰無仮説の μ_0 の一致を意味する．斜線の棄却域は $\bar{X}$ と μ_0 の十分なずれがある領域であり，分布の端部に有意水準で決められる．片側対立仮説および片側検定は，右と左の場合がある．

(12.3)　$$t<-t_\alpha(n-1)$$

となる．また，$\mu>c$ の形の片側対立仮説もあるが，そのときは

(12.4)　$$t>t_\alpha(n-1)$$

が棄却域となる．これも片側検定である．

なお，片側対立仮説か両側対立仮説かは，問題の現実の内容によってわれわれが選ぶものである．

両側か片側か　一般に，両側検定は，母数 θ の値がある目標値 θ_0 と等しいかどうかだけを調べる場合に用いられる．たとえば，工場で新しく機械を購入したとしよう．機械が正しく働いていれば，材料や運転条件によるばらつきがあるにしても，製品は目標値の近くのものであるはずである．製品が目標値から（いずれの方向においても）大きく異なることは，機械が正しく働いていないことを意味する．このような場合には，両側検定となる．

片側検定は母数の大きさが理論的・経験的に予測される場合に使われる．たとえば，英語の特別授業の効果を調べる場合を考えよう．英語の特別授業の前後での英語の試験の点数の平均を比べる場合，特別授業に効果があれば試験後の点数がよくなっているはずである．このような場合，われわれが知りたいのは授業前後の得点が異なっていることだけではなくて，授業後の得点が向上したかどうかである．このような場合には，対立仮説を不等号で与える片側検定を用いる．

12.2　正規母集団に対する仮説検定

もっとも広く使われている検定の例として，母集団の分布が正規分布である場合の仮説検定について述べよう．実用上重要なものとして母平均に関する両