

検定の考え方, 独立性の検定

儀我真理子

日本医科大学基礎科学数学

Statistical Tests, Test of Independence

Mariko Giga

Department of mathematics, Nippon Medical School

Abstract

In this paper we describe statistical test. This is the most important statistical field for examining the mean, variance and other characteristics of population by using sample. First, we propose a null hypothesis, for example, that the population mean is a certain value. With statistical analysis, if we conclude that the null hypothesis is to be rejected, we judge that the hypothesis is probably not true. When we conclude that the null hypothesis is to be accepted, we can only say that we cannot reject it. In this paper, we also describe a test of independence. This is a method of testing, for example, whether we can recognize a difference in the numbers of male and female patients with certain diseases. The main method of testing for independence is not applicable to small samples, for which other methods must be used.

(日本医科大学医学会雑誌 2014; 10: 115-119)

Key words: statistical test, null hypothesis, reject, accept, test of independence

1. 検定の考え方

検定 (test) の考え方の基本は推定のそれと変わらない。前回の母平均の推定[1]において, “信頼度 95% の信頼区間 ($\alpha=0.05$)” の正確な意味が, 「標本をいくつもとったとすると, そこから計算できる信頼区間はその都度異なる。その中で信頼区間が真の母平均を含んでいる確率が, 95% である。」ということ述べた。またこれは, 信頼区間を導いたのと同じ式

$$-A_\alpha \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq A_\alpha \quad (1)$$

(記号 A_α の意味は [1] の (6) 式参照) において, 母平均 μ を $\mu = \mu_0$ (μ_0 は値) と仮定したとき (これを帰無仮説 (null hypothesis) と言い, H_0 で表す), 不等式

$$-A_\alpha \leq \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq A_\alpha \quad (2)$$

が満たされる確率は $1 - \alpha$ ($=0.95$) であることを意味する。言い換えれば, 帰無仮説 $\mu = \mu_0$ が成り立つのに (2) が成り立たない確率は α ($=0.05$) しかない。したがって (2) が成り立っていないときには帰無仮

説はどれも成り立たないらしいということで“棄却”し、それ以外なら採択する。これが検定の基本的な考え方である。今述べた値 α は有意水準 (significance level) と呼ばれる。

公式1 母分散既知の場合の母平均の検定 (両側検定)

正規母集団 $N(\mu, \sigma^2)$ において母分散 σ^2 は既知とし、そこから抽出された大きさ n の標本の標本平均を \bar{x} とする。

$$\text{帰無仮説 } H_0: \mu = \mu_0$$

とするとき、有意水準 α での両側検定における棄却域は次で与えられる：

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < A_\alpha \quad \text{または} \quad z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > A_\alpha. \quad (3)$$

帰無仮説の「帰無」とは“無に帰する”という意味である。検定をするとき、多くの場合棄却を期待している。例えば、新しい治療法が既成のそれより効果があるか否かを検定したいときは、帰無仮説 $H_0: \mu = \mu_0$ (μ は新たな治療法による調べている量の平均、 μ_0 は既成の治療法による対応する量の平均の値) は“新たな治療法と既存のそれとの間に効果の差はない”であるから、まさに棄却を期待している。逆に帰無仮説が採択された場合の意味するところは(“既存の治療法による効果と変わらない”というよりは)“既存の治療法とは異なる効果があったという結論は導けなかった”である。このように、棄却と採択は同等ではない。検定の考え方は公式1の前にも述べたが、棄却、採択についてもう一度考えてみる。

公式1の z を確率変数とみたものを Z とする。 Z のなす正規分布は帰無仮説が成り立つという仮定のもとでの分布である。したがって、たまたま偏った標本になり、 Z の値がグラフの端の方 (μ_0 から遠いところ) になることは帰無仮説が正しくても起こるが、めったに起こるはずのないことなのだから、標本から計算した Z の値がグラフ端の方の値をとったら棄却とするのである。しかし、 $\mu = \mu_0$ が成り立っていても z の値がグラフの端の方の値をとることがときにはあり得るので、“棄却”と結論するのは実は誤りであったという危険もある。これを第1種の過誤 (error of first kind) という。その危険の確率は、意味から考えて有意水準 α に等しい。この意味で有意水準のことを危険率ともいう。棄却できないときは採択になるが、こちらに関しても、本当は帰無仮説が成り立っていないのに採択としてしまう可能性もあるはずである。これ

を第2種の過誤 (error of second kind) という。

検定の一番基本的な考え方では、第2種の過誤についてはあまり深く追求しない。棄却は“帰無仮説はうそらしいから棄却しよう”という意味である。しかし採択は上の治療法の例でも述べたように、“これくらいでは棄却することができない”という意味である。たとえそのときの標本平均が μ_0 からある程度ずれていても、“標本の具合によりそれくらいずれることはある。”という感じである。

今まで述べてきた手法は両側検定であるが、上側検定、下側検定もある。両側検定は、母平均とある値 μ_0 との間に差があるか否かに関心のある場合に、上側(下側)検定は、母平均がある値 μ_0 より大きい(小さい)か否かに関心のある場合に用いられる。また検定の種類によっては、帰無仮説が成り立たない場合には、統計検定量の値が必ず、上側にずれるというものもある。

ここまでは検定の基本概念として、帰無仮説の棄却の意味について述べ、採択については「棄却できない」というきわめて消極的な言い方をしてきた。これはまずは第1種の過誤に対して慎重であるべきであるという検定の基本的立場である。しかし採択の方にもある程度の意味を持たせる、つまり第2種の過誤もなるべく少なくするという立場で考える検定力という理論もある ([2] 参照)。

標本平均 \bar{x} から計算した (2) の中辺の値を z_1 とするとき、確率 $P(z \geq |z_1|)$ の値を p 値 (p value) という。つまり p 値とは、帰無仮説の下で実際に標本から得られた値 z_1 より極端な(つまり帰無仮説の値から離れている)値となる確率はどれだけかを表しており、SPSSなどの統計ソフトでも用いられている。なお、結果が棄却の場合には、 p 値は帰無仮説を棄却するための最少の有意水準であることもわかる。

多くの検定手法が知られているが、以下では適合度検定、独立性の検定について述べる。

2. 適合度検定

例2 メンデルの実験

メンデルは1865年、エンドウ豆の色と形状についての遺伝形質を調べた実験において、次の上の段の数値結果を得た。この観測結果より、4つの表現型は9:3:3:1の比率で現れると彼は考えた。これは統計的に正しいであろうか。下の段の数値は、556個のエンドウ豆を、理論比9:3:3:1にしたがって分けた場合の理論度数を表す。

	黄丸	黄しわ	緑丸	緑しわ	計
観測	315	101	108	32	556
理論	312.8	104.3	104.3	34.75	556

メンデルの観測結果が理論比に合っているかどうかの検定をする際に使われるのが、次に述べる適合度検定 (χ^2 -test of goodness of fit) である。

公式 3 適合度検定

事象 $E_i (i=1, \dots, k)$ が n 回中 x_i 回起こったとすると、帰無仮説 H_0 : 観測結果は理論値 np_i と合うとして有意水準 α で検定する。ただし、 $np_i \geq 5$ とする。棄却域は次で与えられる：

$$\chi^2 = \sum_{i=1}^k \frac{(x_i - np_i)^2}{np_i} > \chi^2(k-1; \alpha).$$

ここで $\chi^2(k-1; \alpha)$ は χ^2 分布表における値である。

この公式を使って例 2 の適合度検定を行う。

帰無仮説 H_0 は「 $p_1 : p_2 : p_3 : p_4$ は $9 : 3 : 3 : 1$ 」である。これを有意水準 0.05 で検定する。 $n=556$, $p_1 = \frac{9}{16}$, $p_2 = \frac{3}{16}$, $p_3 = \frac{3}{16}$, $p_4 = \frac{1}{16}$ であるから、検定統計量の値は次のようになる。

$$\begin{aligned} \chi^2 = & \frac{(315 - 312.8)^2}{312.8} + \frac{(101 - 104.3)^2}{104.3} \\ & + \frac{(108 - 104.3)^2}{104.3} + \frac{(32 - 34.75)^2}{34.75} = 0.470 \end{aligned}$$

χ^2 分布表より $\chi^2(4-1; 0.05) = 7.815$ であるから、 $\chi^2 < \chi^2(4-1; 0.05)$ が成り立ち棄却できない、つまりメンデルの実験結果は理論比 $9 : 3 : 3 : 1$ に合っていると言ってよい。ここで、適合度検定の結果“適合している”という結論を得たということの意味は、「理論比に適合していると積極的に判断できた」ではなく、「帰無仮説を棄却することはできない」である。 χ^2 の計算式を見るとわかるように、統計量 χ^2 の値が大きい場合に、理論度数と観測度数の数値には開きがある。したがって適合度検定の場合には、棄却域は上側だけになる。つまり必ず上側検定となる。

適合度検定を使って、データがある分布に従っているか否かを調べることもできる。

3. 独立性の検定

母集団からランダムに選んだ n 個の標本を、2 種類の属性 A, B により、 A_1, A_2, \dots, A_r と B_1, B_2, \dots, B_s に分類したとき、 A_i と B_j をともに持つ標本の個数を x_{ij} 、属性 $A_i (i=1, \dots, r)$ を持つ標本の個数の和

を a_i 、属性 $B_j (j=1, \dots, s)$ を持つ標本の個数の和を b_j として、表 4 (A) のような分割表を作る。ここで $a_i (i=1, \dots, r)$, $b_j (j=1, \dots, s)$ を周辺度数という。「属性 A, B は互いに独立であるか否か」を検定したい。この手法を独立性の χ^2 検定 (χ^2 -test of independence) という。独立性が成り立たない度合が大きいほど検定統計量 χ^2 の値は大きくなるので、棄却域は大きい方のみにとる (上側検定)。

観測度数と A, B が独立としたときの理論度数を次のようにおく。理論度数は $y_{ij} = a_i b_j / n$ である。

表 4 (A) 観測度数表

	B_1	B_2	...	B_s	計
A_1	x_{11}	x_{12}	...	x_{1s}	a_1
A_2	x_{21}	x_{22}	...	x_{2s}	a_2
\vdots	\vdots
A_r	x_{r1}	x_{r2}	...	x_{rs}	a_r
計	b_1	b_2	...	b_s	n

表 4 (B) 理論度数表

	B_1	B_2	...	B_s	計
A_1	y_{11}	y_{12}	...	y_{1s}	a_1
A_2	y_{21}	y_{22}	...	y_{2s}	a_2
\vdots	\vdots
A_r	y_{r1}	y_{r2}	...	y_{rs}	a_r
計	b_1	b_2	...	b_s	n

公式 5 独立性の検定

帰無仮説 H_0 : A, B は互いに独立であるとして有意水準 α で検定する。ただし、理論度数は 5 以上とする。棄却域は次で与えられる：

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(x_{ij} - y_{ij})^2}{y_{ij}} > \chi^2((r-1)(s-1); \alpha). \quad (4)$$

なお 2×2 分割表の場合は、検定統計量は次のようになることが直接計算でわかる。ゆえに、この場合には理論度数表を作る必要はない。

	B_1	B_2	計
A_1	a	b	$a+b$
A_2	c	d	$c+d$
計	$a+c$	$b+d$	n

$$\chi^2 = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} \quad (6)$$

例 6 50 人の児童のうち 35 人はインフルエンザの予防接種を受けた。インフルエンザに罹患した者と罹患しなかった者の数は表のようであった。この予防接種はインフルエンザの発症を抑えるのに効果があった

といえるかを、有意水準5%で検定する。

	罹患	罹患せず	計
受けた	3	32	35
受けなかった	7	8	15
計	10	40	50

帰無仮説 H_0 は、「予防接種を受けたこととインフルエンザに罹患したことは独立」である。

ここではまず、公式5を用いて考えてみる。理論度数表は次のようになる。

	罹患	罹患せず	計
受けた	7	28	35
受けなかった	3	12	15
計	10	40	50

実際の数 x_{ij} 、理論度数 y_{ij} として、(4)式を使うと、

$$x^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(x_{ij} - y_{ij})^2}{y_{ij}} = 9.524 > 3.842 \quad (= x^2((2-1)(2-1); 0.05))$$

となり棄却、つまり予防接種は効果があったといえる。なお、例6は2×2分割表の場合なので、検定統計量の値は理論度数表を作らず(6)式で直接計算することもできる。

公式3、公式5を見ると、これが x^2 分布としてうまく機能するのは n が十分大きいときであることがわかる。事実、適合度検定における np_i や分割表の理論度数が5以下の値を含む場合、上の x^2 分布による近似はとても悪くなる。このようなときは以下の(I)～(III)の方法を用いるとよい。

(I) 2つのクラスのデータをまとめる。たとえば表4(A)において、 B_{s-1} と B_s の値が小さい場合には、これら2つの値を合わせて B_{s-1} の値とする。

(II) イエーツ (Yates) の補正をする。すなわち、統計量として次のものを使う。

$$x^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(|x_{ij} - y_{ij}| - \frac{1}{2})^2}{y_{ij}} \quad (7)$$

特に2×2分割表の場合には次のようになる。

$$x^2 = \frac{n(|ad - bc| - \frac{n}{2})^2}{(a+b)(c+d)(a+c)(b+d)} \quad (8)$$

x^2 の値が小さくなるように補正する理由は、もともとは離散的な分布を連続分布である x^2 分布で近似して解析するからである。なお例6ではイエーツの補正

をしても結論は変わらない。

(III) 理論度数が5以下の値を含み、どの値もあまり大きくないときは、次の4節で述べるフィッシャーの直接法を用いる。

適合度検定における公式3の np_i は、独立性の検定における公式5の y_{ij} に相当する。このことからわかるように、適合度検定と独立性の検定は統計理論としては同じである。適合度検定は理論度数の割合が初めから定めてある場合の検定と言える。

独立性の検定をする場合、「独立性が成り立たなくなるほど x^2 の値は大きくなるので、棄却域は大きい方のみにとる。」とした。この考えはどこもまちがってはいないのであるが、 x^2 分布の確率密度関数のグラフを見ると、自由度が3以上のときは、 $x=0$ のときは確率は低く、 x が少し大きくなるあたりで確率が最大になっていることに気づく。真に独立で理想的な状態ならば $x=0$ のとき確率最大になるはずであろう。しかし独立性の検定に使う確率変数の分布がこのようになる理由は、統計の理論がすべて正規分布によるばらつきを持つことを仮定してできているからである。もし標本から独立性の検定をおこなった結果、検定統計量が0に非常に近くなったら、それは“合いすぎ”なのであり何らかの作為を疑われる結果であると考えられることもできる。(メンデルのエンドウ豆の実験結果は“合いすぎ”であると言われている。しかしそれはメンデルの功績を損なうものではないであろう。) この考え方で独立性の検定をする流儀もある。その場合は両側検定になる。

ついでながら説明を付け足す。例えば母平均の検定で使う検定統計量の標準化する前のグラフは、帰無仮説 $\mu = \mu_0$ を中心とする正規分布である。これは左右対称であるからばらつきの影響も左右対称に働き合い、 $\mu = \mu_0$ のところがグラフの値が一番大きくなる。しかし独立性の検定の場合は、検定統計量は正の値のみをとる x^2 分布に従うので、グラフの最大値は $x=0$ からずれるのである。

4. フィッシャーの直接法による独立性の検定

前節で述べたフィッシャーの直接法またはフィッシャーの正確確率検定 (Fisher's exact test) とは、データの数が非常に少なく、公式5を使つての検定が適用できない場合 (目安としては、理論度数が5以下の値を含みどの値もあまり大きくないとき) の属性に関する独立性の検定法であり、両側検定、片側検定どちらも可能である。この手法は、母平均が特定の分布

に従うことを仮定していないのでノンパラメトリック検定である。2×2分割表で使うことが多い。

例7

肉と魚の嗜好の違いによって、コレステロールが基準より高いか低いかを調べたところ、左側の表のような結果を得た。なお、右側の表は説明のために記号を当てたものである。

表8

	高め	低め	計		B_1	B_2	計
肉	9	9	14	A_1	a	b	r
魚	3	10	13	A_2	c	d	s
計	12	15	27	計	t	u	n

検定したいことは、肉と魚の嗜好の違いによって、コレステロール値の高低に差があるか、すなわち A_1 か A_2 によって、 B_1 か B_2 かに差があるか否か、である。今周辺度数 r, s, t, u は固定して考える。このとき a が決まれば b, c, d の値は自然に決まることに注意する ($0 \leq a \leq \min(r, t)$)。したがって、帰無仮説 H_0 「因子 A と B とは互いに独立である」の下で、(A_1, B_1) に対応する値 a を調べることによって、独立性を判断することができる。

(A_1, B_1) の値が a となる確率、すなわち a, b, c, d が表8のように決まる確率は次のようになる。

$$P_a = \frac{{}_r C_a \times {}_s C_c}{{}_n C_t} = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{n! a! b! c! d!} \quad (9)$$

ゆえに、その出現確率は図9のようになる。

“肉が好きか魚が好きか”が、“コレステロールの高い低い”とは無関係、つまり完全に独立であったとすると、

$$\frac{a}{b} = \frac{c}{d} \text{ すなわち } ad - bc = 0$$

となるはずである。この $ad - bc$ の値の絶対値が小さ

ければ関連は低いし、大きければ関連は強いと考えてよい。このことから $|ad - bc|$ は独立性の指標と考えることができる。

有意水準5%で A, B の独立性を検定したい場合は、 a の値が小さい方からと大きい方から、それぞれ図9の累積確率が0.025になるところまでが棄却域である。それは a の値で言うと、例7の場合3以下および10以上である。この例における実際の値は $a=9$ であったから、帰無仮説は棄却されない。つまり“肉が好きか魚が好きかで、コレステロール値が基準より高いか低いかに違いがあるとは言えない”ことがわかった。

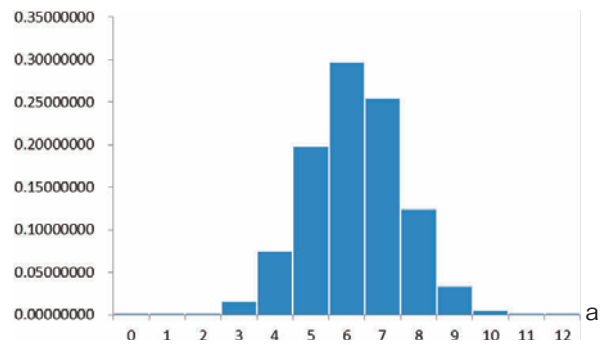


図9

文献

1. 儀我真理子：正規母集団、推定の考え方。日本医科大学医学会雑誌 2014; 10: 16-20.
2. 儀我真理子：確率・統計の基礎。2014, ムイスリ出版.
3. 薩摩順吉：確率・統計。1989, 岩波書店.
4. 澤田 昇, 田澤新成：統計学の基礎と演習。2005, 共立出版.
5. 篠原昌彦：確率・統計。1989, 朝倉書店.
6. 丹後俊郎：新版医学への統計学。1993, 朝倉書店.
7. 東京大学教養学部統計学教室：自然科学の統計学。1993, 東京大学出版会.
8. 服部哲也：理工系の確率・統計入門。2005, 学術図書.
9. 宮原英夫, 白鷹増男：医学統計学。1992, 朝倉書店.

(受付：2014年1月16日)

(受理：2014年3月17日)