

第11回：仮説検定

尚 晋
大学院経済学研究科 助教

2025年6月17日

今日のポイント

1. 仮説検定

1 仮説検定	1	1.2.4 母分散の比の検定	10
1.1 検定の考え方	1	1.3 いろいろの χ^2 検定	10
1.2 正規母集団に対する仮説検定 .	5	1.3.1 適合度の検定	10
1.2.1 母平均に関する検定 .	5	1.3.2 分割表と独立性の検定	11
1.2.2 母分散に関する検定 .	7	1.4 中心極限定理を用いる検定	12
1.2.3 母平均の差の検定	8	1.5 検出力	13

1 仮説検定

「仮説検定」とは、統計的仮説の「有意性」の検定すること。仮説に基づく期待値と観測結果の差が偶然によるものかを確率的基準で評価する。仮説検定は推定と並ぶ統計的推測理論の双壁であり、統計的判断の論理学・科学方法論としての意味も持つ。

1.1 検定の考え方

有意性検定 仮説検定(hypothesis testing)は人間の論証感覚を定式化したものであり、仮説検定の目的は、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。

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)

たとえば、エンドウ豆の型の度数分布が理論上の仮説(比率が 9 : 3 : 3 : 1)に合致しているかどうかを検証する問題がある。このとき、度数の比が厳密に 9 : 3 : 3 : 1 になっていない。重要なことは、理論比からのずれが誤差の範囲内か、それともそれ以上何らかの意味のあるものかという点である。

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)であるといふ。ここで立てられた仮説を 統計的仮説(statistical hypothesis)，あるいは単に仮

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)という。

統計的仮説の範囲は数理的命題に限らず、タバコと肺ガンの因果関係のような実際的問題も扱える。確率論的関係式として定式化することで統計的仮説として検証可能。

有意性は「標本がどの程度有意なずれを示すか」の確率で表される。このとき、標本分布が重要な役割を果たす。

- たとえば、コインを20回投げて14回表が出たとする。
- コインに歪みがないという仮説 $p = 1/2$ が妥当かどうか？
- この場合、 X : 表が出た回数は $Bi(20, 1/2)$ に従う。
- 表12.2より、 $P(X \geq 14) = 1 - 0.9423 = 0.0577$
- これは「コインに歪みがない」と仮定したときに、 $X = 14$ が出る確率が約 5.77% という意味である。

このように、 $X = 14$ は仮説からすれば、出るはずのないかなりはずれた値で、すなわち「出にくい」値なので、 $p = 1/2$ という仮説は棄却される(=誤っていると判断する)。このとき、「仮説は棄却された(reject)」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、棄却するかしないかを決定する手続きである。

ここで 0.0577 を「稀」と考えたが、一般に、あらかじめ「どの程度の確率を稀とするか」は決めておく。この基準値を 有意水準(significance level)といい、通常 α で表す。

- たとえば $\alpha = 0.10$ と決めた場合、0.0577 は稀と判断される(棄却される)。
- しかし $\alpha = 0.01$ ならば、0.0577 は十分に起こりうる値であるため、仮説は棄却されない。

帰無仮説と対立仮説 たとえば、コインを20回投げて、表の出る確率 $p = 1/2$ という仮説が有意水準 $\alpha = 0.1$ で棄却されたとする。このように仮説が棄却されること自体で判断を終える立場もあるが、より積極的な判断をしたい場合には、あらかじめもう一つの仮説 $p \neq 1/2$ を立てておく。このとき、

- もとの仮説 : $p = 1/2$ を 帰無仮説(null hypothesis, H_0)
- 対立する仮説 : $p \neq 1/2$ を 対立仮説(alternative hypothesis, H_1)

帰無仮説と対立仮説は互いに否定関係にあるとみなされる。ただし、 $p < 1/2$ のような第三の可能性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。したがって、一応の理解として、帰無仮説を棄却することは、対立仮説を採択することを意味する。

検定における4通りの結果と誤りの分類 帰無仮説 H_0 を「棄却する」か「棄却しない」かという判断に対して、次の4通りの結果がある(表12.3)。

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

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

正しいのは①、④の場合；誤りのは②(第一種の誤り)、③(第二種の誤り)の場合。

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	$.9537 \times 10^{-6}$.0000
1	$.1907 \times 10^{-4}$.0000
2	$.1811 \times 10^{-3}$.0002
3	$.1087 \times 10^{-2}$.0013
4	$.4621 \times 10^{-2}$.0059
5	$.1478 \times 10^{-1}$.0207
6	$.3696 \times 10^{-1}$.0577
7	$.7392 \times 10^{-1}$.1316
8	.1201	.2517
9	.1602	.4119
10	.1762	.5881
11	.1602	.7483
12	.1201	.8684
13	$.7392 \times 10^{-1}$.9423
14	$.3696 \times 10^{-1}$.9793
15	$.1478 \times 10^{-1}$.9941
16	$.4621 \times 10^{-2}$.9987
17	$.1087 \times 10^{-2}$.9998
18	$.1811 \times 10^{-3}$.9999
19	$.1907 \times 10^{-4}$.9999
20	$.9537 \times 10^{-6}$	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。

誤りには以下の2種類がある：

- (a) 第1種の誤り(Type I Error) : H_0 が正しいのに、それを棄却してしまう誤り \Rightarrow ②
- (b) 第2種の誤り(Type II Error) : H_0 が誤っているのに、それを棄却しない誤り \Rightarrow ③

この分類は、 $2 \times 2 = 4$ 通りの判断の中で、正しい判断2通り、誤った判断2通りがあるという一般的な構造に基づく。この考え方は、たとえば大量生産における品質管理、刑事訴訟での無罪を有罪と判断する誤り／逆の誤りなど、幅広い判断の誤りへの思考枠組みとして応用される(表12.4)。

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	H_0 が正しくないときにこれを採択する誤り
(例) 抽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)	(例) 抽取検査にあたって、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)

帰無仮説の採択とその意味 有意性検定は、帰無仮説のもとで、期待されるような結果が観察されなかつたことを根拠として、仮説を棄却(否定)することが、主な内容である。

これは論理学における背理法に相当する。

重要なのは、仮説が棄却されなかつたからといって、仮説が積極的に正しい(真であることが証明された)というわけではないという点である。

仮説を採択したということは、あくまでも「帰無仮説と矛盾しなかつた」ことを意味するに過ぎない。

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 $N(\mu, \sigma^2)$ 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

$$H_1 : \mu \neq 15$$

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、スチューデントの t 統計量を計算すると、

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

有意水準 $\alpha = 0.05$ に対応する t 分布 $t(24)$ のパーセント点は

$$t_{0.025}(24) = 2.06$$

であるから、 $|t| = 2.83 > 2.06$ 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 はこの標準偏差の約 3 倍で($t = 2.83$)、大きいというほかはない。

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、 $\mu = 15$ と考えにくい。

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。したがって、 t 統計量

$$t = \frac{\bar{X} - 15}{s/\sqrt{n}} \quad (1)$$

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る。

今、 $t_{0.05}(24) = 1.71$ であるから、 $-2.83 < -1.71$ より、この場合も有意水準5%で帰無仮説を棄却する。

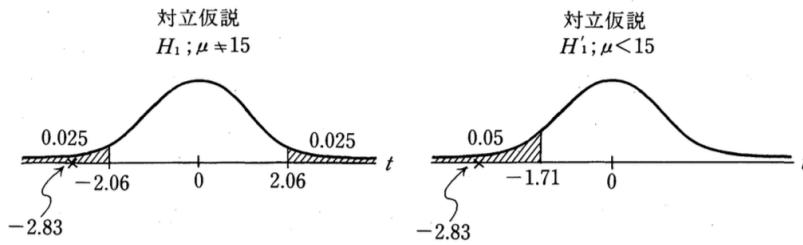


図 12.1 仮説検定のやり方

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、その帰無仮説は棄却される。

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($\mu = 15$ に近い \bar{X} の値に対応)になるが、棄却域は対立仮説が次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

$$|t| > t_{\alpha/2}(n - 1) \quad (2)$$

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)

$$t < -t_{\alpha}(n - 1) \quad (3)$$

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

$$t > t_{\alpha}(n - 1) \quad (4)$$

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 μ の値がある目標値 μ_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。

一方、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とき、特別授業に効果があれば試験後の点数が良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、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、対立仮説を片側検定を用いる。

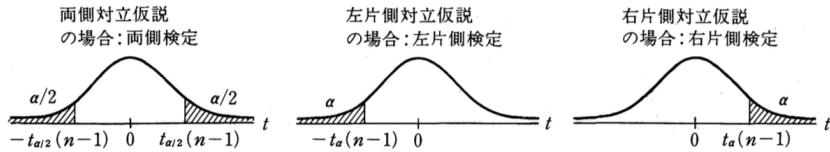


図 12.2 検定の手続例：母平均は帰無仮説の通りか

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。

1.2 正規母集団に対する仮説検定

正規分布における仮説検定が最も広く使われる検定手法であり、母平均の両側・片側検定や二つの母集団の平均差の検定が実用上重要である。また、ここで考える検定は、中心極限定理により正規分布以外でもサンプルサイズが大きければ近似的に適用可能で、すべて有意水準 α で行う。

1.2.1 母平均に関する検定

まず、もっともよく用いられる母平均 μ 、母分散 σ^2 の正規母集団に関する母平均 μ の検定について、両側検定と片側検定に分けて説明する。

両側検定 両側検定では、帰無仮説と対立仮説は次のように与えられる：

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

母平均 μ に関する検定は、標本平均 \bar{X} が μ_0 からどれくらい離れているかを基準にして行われる。このとき、その差を判断するための尺度は、母分散 σ^2 が既知か未知かによって異なる。

母分散 σ^2 が既知のとき

このときの検定統計量 Z は、次式で与えられる標準化変数である：

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \tag{5}$$

帰無仮説が正しければ、 $\mu = \mu_0$ のもとで Z は標準正規分布 $N(0, 1)$ に従う。また、標準正規分布は原点に対して対称であるため、両側検定では標準正規分布表から得られるパーセント点 $z_{\alpha/2}$ と比較して、

$$|Z| > z_{\alpha/2} \text{ ならば } H_0 \text{ を棄却する}$$

$$|Z| \leq z_{\alpha/2} \text{ ならば } H_0 \text{ を棄却しない}$$

母分散 σ^2 が未知のとき

実際には母分散 σ^2 が既知であることはほとんどないため、標本分散 s^2 を用いるスチューデントの t 統計量を使う：

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \quad (6)$$

このとき t は自由度 $n - 1$ の t 分布に従い、パーセント点 $t_{\alpha/2}(n - 1)$ と比較して判断する： $|t| > t_{\alpha/2}(n - 1)$ ならば H_0 を棄却する。 $|t| \leq t_{\alpha/2}(n - 1)$ ならば H_0 を棄却しない。これらの検定はスチューデントの t 検定(Student's t -test)と呼ばれる。

例：空調システムの検定 空調システムの作動状況を調べるために、設定温度を 25°C として、7日間の室温データを以下のように得た：

24.2, 25.3, 26.2, 25.7, 24.4, 25.1, 25.6

このシステムが正しく働いているかどうかを有意水準 5% で検定する。

帰無仮説： $H_0 : \mu = 25.0$ 対立仮説： $H_1 : \mu \neq 25.0$

標本平均 $\bar{X} = 25.21$ 、標本標準偏差 $s = 0.715$ より、

$$t = \frac{25.21 - 25.0}{0.715/\sqrt{7}} = 0.777$$

$t_{0.025}(6) = 2.447$ より、

$$|t| = 0.777 < 2.447$$

よって、有意水準 5% では帰無仮説は棄却されない。

片側検定 帰無仮説と対立仮説が次のように与えられるとき(右片側検定)：

$$H_0 : \mu = \mu_0, \quad H_1 : \mu > \mu_0$$

帰無仮説が正しいときの検定統計量の分布は両側検定と同様であるが、対立仮説が異なっているので、棄却域が異なる。

- σ^2 が既知のとき： $Z > z_\alpha$ で H_0 を棄却
- σ^2 が未知のとき： $t > t_\alpha(n - 1)$ で H_0 を棄却

左片側検定($H_1 : \mu < \mu_0$)の場合は、以下のように判断する：

- σ^2 が既知のとき： $Z < -z_\alpha$ で H_0 を棄却
- σ^2 が未知のとき： $t < -t_\alpha(n - 1)$ で H_0 を棄却

例：講義の効果検定 英語の特別講義の効果を調べるため、10名の受講者に講義の前後で英語の試験を実施し、その得点差(後-前)を得た：

-1, 3, 4, 5, 3, 0, 7, 4, 2, -2

得点差が正規分布 $N(\mu, \sigma^2)$ に従うとし、講義の効果を次の仮説で検定する：

$$H_0 : \mu = 0, \quad H_1 : \mu > 0$$

このとき、標本平均 $\bar{X} = 2.5$ 、標本標準偏差 $s = 2.50$ より、

$$t = \frac{2.5 - 0}{2.5/\sqrt{10}} = 2.82$$

$$t_{0.05}(9) = 1.833 \text{ より,}$$

$$t = 2.82 > 1.833$$

よって、有意水準 5%において H_0 は棄却され、講義に効果があったと判断される。

1.2.2 母分散に関する検定

母分散 σ^2 に対する帰無仮説 $H_0 : \sigma^2 = \sigma_0^2$ の検定は、標本分散 s^2 に基づいて行われる。検定統計量は次式で与えられる：

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \quad (7)$$

この統計量は、帰無仮説が正しいとき、自由度 $n-1$ のカイ二乗分布 $\chi^2(n-1)$ に従う。有意水準を α としたとき、カイ二乗分布表(付表3)から自由度 $(n-1)$ のパーセント点 $\chi_{1-\alpha/2}^2(n-1)$ および $\chi_{\alpha/2}^2(n-1)$ などを用いて検定する。

(a) 両側検定 : $H_1 : \sigma^2 \neq \sigma_0^2$

この場合、次のように判定する：

- $\chi_{\alpha/2}^2(n-1) < \chi^2 < \chi_{1-\alpha/2}^2(n-1)$ のとき、 H_0 を棄却しない。
- それ以外の場合は、 H_0 を棄却する。

(b) 片側検定

- 対立仮説 $H_1 : \sigma^2 > \sigma_0^2$ (右片側検定)のとき：

$$\chi^2 > \chi_{1-\alpha}^2(n-1) \Rightarrow H_0 \text{ を棄却}$$

- 対立仮説 $H_1 : \sigma^2 < \sigma_0^2$ (左片側検定)のとき：

$$\chi^2 < \chi_{\alpha}^2(n-1) \Rightarrow H_0 \text{ を棄却}$$

これらの検定を正規母集団の母分散に対する カイ二乗検定(χ^2 -test) という。

例：能力のばらつきの検定 ある小学校では入学時に知能テストを行っており、従来は平均 50、分散 36 であった。本年度の児童について25人をランダムに抽出し、例年と同じ条件でテストしたところ、平均 53、分散 48 を得た。本年度の児童の揃い方が例年と異なるといえるかを検定する。(肥田野)

帰無仮説 : $H_0 : \sigma^2 = 36$ 対立仮説 : $H_1 : \sigma^2 \neq 36$ 有意水準 : 10%

検定統計量 :

$$\chi^2 = \frac{24 \cdot 48}{36} = 32 \quad (8)$$

自由度 $n-1 = 24$ におけるパーセント点 :

$$\chi_{0.05}^2(24) = 13.848, \quad \chi_{0.95}^2(24) = 36.415$$

これにより、

$$13.848 < 32 < 36.415$$

したがって、有意水準10%では H_0 は棄却されない。よって、本年度の児童の質の揃い方がとくに例年と異なるとはいえない。

なお、分散が大きいことが特別な対応を要する場合は、対立仮説を $H_1 : \sigma^2 > 36$ として右片側検定を行うのが適切である。

1.2.3 母平均の差の検定

二つの正規母集団の母平均の差の検定は、実用上きわめて重要である。

たとえば、新しい治療法の効果を調べる場合、患者を二つのグループに分け、一方のみに新しい治療法を行い、その効果に差があるかを検定することがある。これは**2標本検定**(two-sample test)と呼ばれる。治療を施すグループを**処理群**(treatment group)、比較のため治療を行わないグループを**対照群**(contrast group)または**制御群**(control group)という。

二つの正規母集団 $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ からそれぞれ大きさ m, n の標本

$$X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n$$

を抽出したとする。このとき、帰無仮説と対立仮説は次のようになる：

- 帰無仮説 : $H_0 : \mu_1 = \mu_2$
- 対立仮説 :
 - (a) 両側検定 : $H_1 : \mu_1 \neq \mu_2$
 - (b) 片側検定 : $H_1 : \mu_1 > \mu_2$ または $H_1 : \mu_1 < \mu_2$

対立仮説が両側か片側かは、目的に応じて決定される。

分散が等しい場合(等分散)

分散が等しい($\sigma_1^2 = \sigma_2^2 = \sigma^2$)と仮定できる場合、合併分散(pooled variance)を用いる：

$$s^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2} \quad (12.8)$$

このとき、標本平均の差に基づく検定統計量は

$$t = \frac{\bar{X} - \bar{Y}}{s \sqrt{\frac{1}{m} + \frac{1}{n}}} \quad (12.9)$$

これは、自由度 $m+n-2$ の t 分布 $t(m+n-2)$ に従う。

- (a) 両側検定 : $|t| > t_{\alpha/2}(m+n-2)$ のとき H_0 を棄却。
- (b) 片側検定 :

$$\begin{aligned} t > t_{\alpha}(m+n-2) &\text{なら } H_0 \text{ を棄却(右片側).} \\ t < -t_{\alpha}(m+n-2) &\text{なら } H_0 \text{ を棄却(左片側).} \end{aligned}$$

分散が等しくない場合(異分散)

分散が等しくない場合、ウェルチの検定(Welch's test)を用いる。

標本分散 s_1^2, s_2^2 に基づき、検定統計量は以下のように定義される：

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} \quad (12.10)$$

この統計量は、以下の近似自由度をもつ t 分布に従う：

$$\nu = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}} \quad (12.11)$$

ν^* に最も近い整数を自由度として、 $t(\nu^*)$ 分布に基づいて検定を行う。

- (a) 両側検定 : $|t| > t_{\alpha/2}(\nu^*)$ のとき H_0 を棄却。
- (b) 片側検定 :

$$\begin{aligned} t > t_{\alpha}(\nu^*) &\text{のとき } H_0 \text{ を棄却(右片側).} \\ t < -t_{\alpha}(\nu^*) &\text{のとき } H_0 \text{ を棄却(左片側).} \end{aligned}$$

この検定は**ウェルチの検定**(Welch's test)と呼ばれる。

例：投薬群と対照群の平均の差の検定 次のデータは、20匹のラットを10匹ずつ2群に分け、一方にはふつうの食餌を与え、他方には血中の赤血球数を減らすと考えられている薬を混入した食餌を与えた場合の、血液 1mm^3 中の赤血球数である。有意水準 α は **0.05** とする。投薬群と対照群の平均の差の検定せよ。(単位：100万個)。

投薬群	7.97	7.66	7.59	8.44	8.05	8.08	8.35	7.77	7.98	8.15
対照群	8.06	8.27	8.45	8.05	8.51	8.14	8.09	8.15	8.16	8.42

(出典：スネデカー、コクラン『統計的方法』)

帰無仮説 (H_0) と対立仮説 (H_1) の設定

薬が「赤血球数を減らす」と明記されているため、その方向性で片側検定(左側検定)を設定する。

- 投薬群の母平均を μ_T 、対照群の母平均を μ_C とする。
- 帰無仮説 (H_0)：薬に赤血球数を減らす効果はない。

$$H_0 : \mu_T = \mu_C$$

- 対立仮説 (H_1)：薬には赤血球数を減らす効果がある。

$$H_1 : \mu_T < \mu_C$$

母分散が等しい場合

検定手法の選択：2つの独立した標本の平均を比較し、母分散が等しいと仮定するため、合併した分散を用いた2標本t検定を用いる。

データの整理と記述統計量

- 投薬群: $n_T = 10$, $\bar{x}_T = 8.004$, $s_T^2 \approx 0.082944$
- 対照群: $n_C = 10$, $\bar{x}_C = 8.230$, $s_C^2 \approx 0.0289$

合併した分散 (s_p^2) の計算

$$s_p^2 = \frac{(n_T - 1)s_T^2 + (n_C - 1)s_C^2}{(n_T - 1) + (n_C - 1)} = \frac{9 \times 0.082944 + 9 \times 0.0289}{9 + 9} \approx 0.055922$$

合併した標準偏差 $s_p \approx 0.236$

検定統計量 (t値) の計算

$$t = \frac{\bar{x}_T - \bar{x}_C}{s_p \sqrt{\frac{1}{n_T} + \frac{1}{n_C}}} = \frac{8.004 - 8.230}{0.236 \times \sqrt{\frac{1}{10} + \frac{1}{10}}} \approx -2.141$$

結論 片側検定(左側検定)であり、自由度 (df) は18で、 $t_{0.05,18} \approx 1.734$ より、 $-2.141 < -1.734$ なので、帰無仮説を棄却する。薬はラットの血液中の赤血球数を減少させる効果がある、と統計的に結論づけられる。

母分散が等しくない場合

先ほどの例：問題文を少し変えて、「20匹のラットは、それぞれ二つの異なる環境から10匹ずつ無作為抽出した」とする。

検定手法の選択：「20匹のラットは、それぞれ二つの異なる環境から10匹ずつ無作為抽出した」と仮定したので、母分散は等しいとは限らない。ここでは母分散が等しくない場合のWelchのt検定を用いる。

検定統計量 (t値) の計算 (Welchのt検定)

$$t = \frac{\bar{x}_T - \bar{x}_C}{\sqrt{\frac{s_T^2}{n_T} + \frac{s_C^2}{n_C}}} \approx -2.137$$

自由度 (ν) の計算 (Welch-Satterthwaiteの近似)

$$\nu = \frac{\left(\frac{s_T^2}{n_T} + \frac{s_C^2}{n_C}\right)^2}{\frac{\left(\frac{s_T^2}{n_T}\right)^2}{n_T-1} + \frac{\left(\frac{s_C^2}{n_C}\right)^2}{n_C-1}} \approx 14.59$$

計算された自由度: $\nu \approx 14.59$ (切り下げて $\nu^* = 14$)

結論 片側検定(左側検定)であり、自由度 $\nu^* = 14$ で、 $t_{0.05,14} \approx 1.761$ より、 $-2.137 < -1.761$ なので、帰無仮説を棄却する。この結果は、統計的に有意な差があり、薬はラットの血液中の赤血球数を減少させる効果があると結論づけられる。

1.2.4 母分散の比の検定

二つの正規母集団の母平均が等しいか否かの検定方法は、母分散が等しい(母分散の比 = 1)かどうかによって異なるため、分散が等しいかどうかの検定が必要である。また、母分散が等しいことそれ自体に意味があることもある(例えば、小学校の2クラスの成績のばらつきが等しいかどうか)。

帰無仮説 :

$$H_0 : \sigma_1^2 = \sigma_2^2$$

対立仮説 :

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

このとき、フィッシャーの分散比は以下のように定義される :

$$F = \frac{s_1^2}{s_2^2} \tag{9}$$

ただし,

$$s_1^2 = \frac{\sum(X_i - \bar{X})^2}{m-1}, \quad s_2^2 = \frac{\sum(Y_j - \bar{Y})^2}{n-1}$$

帰無仮説が正しい場合、 F は自由度 $(m-1, n-1)$ の F 分布に従う : $F \sim F(m-1, n-1)$
検定の基準 :

$$F_{1-\alpha/2}(m-1, n-1) \leq F \leq F_{\alpha/2}(m-1, n-1) \Rightarrow H_0 \text{ を棄却する}$$

例 : $F = 0.0761/0.0294 = 2.59 < F_{0.025}(9, 9) = 4.026$ より、有意水準5%では、母分散が等しいという仮説は棄却されない。

1.3 いろいろの χ^2 検定

χ^2 分布は母分散についての検定以外にも、ばらつきの検定の基準として、近似的によく用いられる。そのうち、適合度の検定と独立性の検定は特に重要なものである。

1.3.1 適合度の検定

仮定された理論上の確率分布に対して、標本から求められた度数が適合するか否かを検証するのが、適合度の χ^2 検定 χ^2 -test of goodness of fit である。

一般に、ある属性 A によって、 n 個の個体(人、もの、データなどの数字)が k 種のカテゴリ(分類項目) A_1, A_2, \dots, A_k へ分類されたとし、各カテゴリ各カテゴリへ属する観測度数 f_1, f_2, \dots, f_k が、理論的な確率分布 p_1, p_2, \dots, p_k に基づく理論度数(或いは期待度数) np_1, np_2, \dots, np_k に適合しているかどうかを検定する。

K.ピアソンの適合度基準の検定統計量は

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - np_i)^2}{np_i} \tag{10}$$

自由度 $k-1$ の $\chi^2(k-1)$ に従う。

自由度は $k - 1$ のは, f_1, f_2, \dots, f_k は確率変数だが, 度数の合計 $\sum f_i = n$ から, 変数1個分だけ自由度が減るからである.

帰無仮説 :

$$H_0 : P(A_1) = p_1, \dots, P(A_k) = p_k$$

臨界値 :

$$\chi^2 > \chi_{\alpha}^2(k - 1) \Rightarrow H_0 \text{ を棄却} \quad (11)$$

例(メンデルの法則) :

$$\chi^2 = \frac{(315 - 312.75)^2}{312.75} + \frac{(101 - 104.25)^2}{104.25} + \frac{(108 - 104.25)^2}{104.25} + \frac{(32 - 34.75)^2}{34.75} = 0.470$$

自由度3, $\chi_{0.05}^2(3) = 7.815$ より, H_0 は棄却されず, 理論に適合しているといえる.

なお, ここで適合度の検定の原理を

$$\chi^2 = \sum (O - E)^2 / E \quad (12)$$

と要約しておこう. ここに O は ‘観測された’ Observed’, E は理論によって ‘予測(期待)された’ Expected’ の意である. ここで, $(O - E)^2 / E$ は ‘相対誤差’ であり, また, \sum は, 一般に, すべてのカテゴリーにわたる和である.

1.3.2 分割表と独立性の検定

表 12.7 分割表の一般的な形

属性 B		B_1	B_2	…	B_c	計
属性 A	A_1	f_{11}	f_{12}	…	f_{1c}	$f_{1\cdot}$
	A_2	f_{21}	f_{22}	…	f_{2c}	$f_{2\cdot}$
•	•	•	•	•	•	•
•	•	•	•	•	•	•
•	•	•	•	•	•	•
A_r	f_r	f_r	…	f_r	$f_{r\cdot}$	
計	$f_{\cdot 1}$	$f_{\cdot 2}$	…	$f_{\cdot c}$	n	

3章でも見たごとく, ‘分割表’は日常よく出あう統計で, 統計学への親しみ深い入門の一つである. これに独立性の検定を適用することが多い. 点線は後述.

n 個の個体に対して二つの異なる属性 A, B (たとえば A として性別, B として車を所有しているかどうかなど)を同時に測定したとする. A は A_1, A_2, \dots, A_r のカテゴリーに, B は B_1, B_2, \dots, B_c のカテゴリーに分割されているとする. A は男, 女, B は車を所有している, していないなどである.

この二つの属性について度数を集計することにより, 表 12.7 のような分割表を得ることができる.

この分割表において, 独立とは, $A_i \cap B_j$ の各確率に対し, 帰無仮説

$$H_0 : \text{すべての } i, j \text{ に対し } P(A_i \cap B_j) = P(A_i)P(B_j) \quad (13)$$

であるこという. 言い換えれば, A_1, \dots, A_r の条件付確率がすべての B_j によらないことである. つまり, (4.15), (4.16) で見ただ

$$P(A_i | B_j) = P(A_i), \quad i = 1, \dots, r$$

であって(右辺には j なし), 要するに, A_1, A_2, \dots, A_r の起りの割合はどの B に対しても共通で

ある。このとき, BはAの起りに影響していない。いま,

$$P(A_i \cap B_j) = p_{ij}, \quad (14)$$

$$P(A_i) = p_{i\cdot} = \sum_j p_{ij}, \quad (15)$$

$$P(B_j) = p_{\cdot j} = \sum_i p_{ij} \quad (16)$$

と表せば, $p_{i\cdot}, p_{\cdot j}$ は p_{ij} の周辺確率分布で, (式13)の独立とは

すべての i, j に対し $p_{ij} = p_{i\cdot}p_{\cdot j}$

が成り立つことを意味する。 $p_{i\cdot}, p_{\cdot j}$ に対応するから, 相対度数による推定値

$$\hat{p}_{i\cdot} = f_{i\cdot}/n, \quad \hat{p}_{\cdot j} = f_{\cdot j}/n \quad (17)$$

によっておきかえれば, H_0 が成立しているときには, $\hat{p}_{ij} = \hat{p}_{i\cdot}\hat{p}_{\cdot j}$ は確率 $P(A_i \cap B_j)$ の推定値, その理論度数は, (式17)を用いて

$$E_{ij} = n\hat{p}_{i\cdot}\hat{p}_{\cdot j} = n \frac{f_{i\cdot}}{n} \frac{f_{\cdot j}}{n} = \frac{f_{i\cdot}f_{\cdot j}}{n} \quad (18)$$

となる(表 12.10)。他方, 観測度数を

$$O_{ij} = f_{ij}$$

とおいて, n が大きいとき, 適合度検定(今回は, 独立性への適合)の原理 (式12)を用いれば, 独立性の χ^2 検定 χ^2 test for independenceの基準

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - f_{i\cdot}f_{\cdot j}/n)^2}{f_{i\cdot}f_{\cdot j}/n} \quad (19)$$

を得る。 χ^2 分布の自由度は, 表 12.7の点線内が自由変数であるから $(r-1)(c-1)$ となる。

表 12.9 ある大学の工学部の期末試験の成績(広津)

代数 \ 解析	優	良	可	計
優	4	2	3	9
良	8	4	6	18
可	6	3	6	15
計	18	9	15	42

対角線上は特に大きいわけではなく、代数のよくできる学生は解析もよくできるという様子は見えない。むしろ、独立性(両者は無関係)に近い結果がうかがわれる。

表 12.9 の χ^2 の値は,

$$\chi^2 = \frac{(4 - 9 \cdot 18/42)^2}{9 \cdot 18/42} + \frac{(2 - 9 \cdot 9/42)^2}{9 \cdot 9/42} + \dots + \frac{(6 - 15 \cdot 15/42)^2}{15 \cdot 15/42} = 0.19$$

表 12.10 表 12.9 の理論度数(広津)

代数 \ 幾何	優	良	可	計
優	3.86	1.93	3.21	9
良	7.71	3.86	6.43	18
可	6.43	3.21	5.36	15
計	18	9	15	42

独立性を仮定した場合の度数はこのようになる。たとえば, $E_{11}=9 \times (18/42)=3.86$ となる。前の表 12.9 の度数の 4 に近く、独立性が成立することがわかる。

1.4 中心極限定理を用いる検定

母集団分布が正規分布でなくとも、検定統計量が近似的に正規分布に従うならば、正規分布に対する検定の手続きをそのまま用いることができる。中心極限定理が成立するときがその場合である。

例：ある人が、カードのマークが黒か赤かを言いあてる能力を確かめるため、カード52枚をあてさせたところ40枚をあてることに成功した。この結果をどう評価するか。

あたりの成功率を $p = 1/2$ とし、帰無仮説 $H_0 : p = 1/2$ が棄却されるかどうかを検討する。

52枚のカードをランダムに黒か赤かと予測するならば、これがあたる確率は $1/2$ である。

これが帰無仮説となる。あたり、はざれを $X_i = 1, 0$ ($i = 1, 2, \dots, 52$) で表すと、各 X_i は独立で二項分布 $B(1, 1/2)$ に従う。

一般に、 n 枚通したあたりの枚数は、 $S_n = X_1 + X_2 + \dots + X_n$ であり、これは二項分布 $B(n, p)$ に従う。中心極限定理により、

$$Z = \frac{S_n - np}{\sqrt{np(1-p)}} \quad (20)$$

は、 n が大きいとき標準正規分布 $N(0, 1)$ に近似的に従う。今の場合、 $n = 52$, $S_n = 40$, $p = 0.5$ を代入すると、

$$Z = \frac{40 - 52 \cdot 0.5}{\sqrt{52 \cdot 0.5 \cdot 0.5}} = \frac{40 - 26}{\sqrt{13}} = \frac{14}{\sqrt{13}} \approx \frac{14}{3.6056} \approx 3.883$$

となる。両側検定を仮定し、有意水準5% (0.05) の場合、 $z_{0.025} = 1.96$ である。計算されたZ値 (3.883) は1.96よりも大きいため、帰無仮説は棄却される。すなわち、あたりはずれはランダムに生じているものとは考えられない(偶然の一一致とは言えないくらいに小さい確率)。

ここで、 $\hat{p} = \frac{S_n}{n}$ は標本における成功の比率(相対頻度)であり、これは母比率の検定と考えることができる。すなわち、式(20)の分子分母それぞれを n で割れば、

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \quad (21)$$

を得る。これが母比率(母集団の比率)が p であることの検定統計量である。

1.5 検出力

有意性検定の理論は、帰無仮説について下す決定が誤っていることを前提としている。もしも帰無仮説 H_0 が正しければとうてい出そうもない（例として、確率 $\alpha = 0.05$ ）検定量の値が出れば、帰無仮説を偽として棄却する。したがって、逆にいえば、帰無仮説が正しくても、出そうもない棄却域の値がたまたま出てしまい（確率 0.05），その結果、帰無仮説の棄却に帰することがある。これは第一種の誤りであり、その確率は検定の有意水準 ($\alpha = 0.05$) に等しい。

これは有意性検定の考え方そのものに必然的に伴うものであり、この確率を0にすることはできない。

一方、帰無仮説 H_0 が偽であるにもかかわらず、たまたま（この確率を β とする）統計量の値が棄却域に入らなかったために、 H_0 を棄却しない誤りが生じる。これは、表12.4で説明した第二種の誤りである。

β は、対立仮説の各値（たとえば $\theta > 0$ なら正の θ ごとに）が正しいという条件のもとに棄却域に入らないという確率を求めればよい。なお、一般に、仮説が1個の値のとき 単純仮説 (simple hypothesis), 2個以上のとき 複合仮説 (composite hypothesis) という。

α , β は検定方法、つまり棄却域のとり方によるが、棄却域の範囲を狭くすれば α は小さくなるが、 β は大きくなる（図12.2）。広くとれば、 β は小さくなり、 α は大きくなる。標本の大きさが一定というもとでは、 α , β をともに小さくすることはできない。

有意性検定では、 α を先に固定している。その条件で β をなるべく小さく、すなわち第二種の誤りをおかさない確率 $1 - \beta$ をなるべく大きくする。この確率を 検出力 (power) という。検出力は、帰無仮説 H_0 が真でないとき、それを正しく棄却する確率である。検出力は検定方法の良さの評価基準であり、検出力の大きいものほど、そのような誤りをおかさない厳しい検定である。