

第6回 統計的検定

A. 統計的検定

1. 統計的検定とは？

統計的検定：統計的に標本の統計量から母集団の母数に関する予測の真偽を検証することを統計的検定という。

統計的検定の例：

例1 さいころを5回振ったところ、5回とも奇数だった。二項分布から5回とも奇数になる確率は $1/32 (=0.03125)$ である。このような低い確率が出ることから、さいころは奇数と偶数が同じ確率で出ると考えるよりも、奇数が出やすいと結論した。

例2 同じM寸の卵でありながらスーパーSは、スーパーKよりも軽い卵を売っていると考え、両店の10個の卵についてそれぞれ調査したところ、スーパーSの卵は平均20g、標準偏差0.5g、スーパーKの卵は平均22g、標準偏差0.5gであった。したがって、スーパーSの卵の重さは信頼区間95%で $20 \pm 0.31g$ であり、スーパーKの卵の重さは信頼区間95%で $22 \pm 0.31g$ である。2つの信頼区間は重ならないこの結果から両店の卵の重さの母平均は違うと結論づけた。

2. 検定の手順

① 帰無仮説の設定

統計的検定によって否定したい仮説を立てる。これを**帰無仮説**という。例1ではさいころは奇数と偶数が同じ確率(0.5)で出る、例2ではスーパー2店の卵の重さが同じであるというのが帰無仮説に相当する。そして、帰無仮説に従えば、実際に得られたデータの出現する確率はいくらかを求め、その確率がきわめて低い(基準は必要によって定める)ならば、帰無仮説が間違っていると、対立仮説が正しいとするのが統計的検定である。

② 対立仮説の設定

帰無仮説が間違っていると判断した場合に採用する仮説を**対立仮説**という。

では、対立仮説は例1, 2ではどうなるか？

例1：

例2：

③ 帰無仮説が成り立つとしたら、今回得たデータが出現する確率を求める（実際には確率である有意水準を決めて、決められた数値を計算することが多い。エクセルなどの統計ソフトでは確率そのものを計算できる：(P値)。

④ 帰無仮説が成り立たないと判断したときに（これを「帰無仮説を棄却する」という）は対立仮説を採用する。

有意差がある：帰無仮説を棄却できるだけの反証が得られた。

5（あるいは1 or 0.1）%の有意水準で有意差がある。

⑤ 帰無仮説を棄却できないときは、有意差がないことになり、対立仮説を採用するだけの証拠がないことになる。

帰無仮説が棄却できないときは、帰無仮説を採用するのではなく、帰無仮説を棄却できるだけの証拠が不十分と考える

対立仮説を証明したいのに、帰無仮説のようなものを持ち出すのはまどろっこしく感じるかもしれない。しかし、統計的にできるのは帰無仮説の否定だけである。その理由は以下の2つである。

★ 「甲は乙より早い」といっても、どのくらい速いのか、甲が乙に勝つ確率はいくらなのか、がはっきりしない。これを明確に規定できなければ、数量的に取り扱えない。

★ 仮に、仮説を数量的に規定できても、それが真実であることを統計的には証明できない。統計的にできることは、仮説に対する「反証」を提示することだけである。

B. 二項分布を利用した統計的検定

例題：A君は対戦型ゲームでキャラクターBと互角に戦えると主張しているが、実際やってみると1勝9敗だった。A君の主張の真偽を統計的に検定せよ。

帰無仮説：A君はキャラクターBと互角に戦う。

A君がキャラクターBに勝つ確率は（ ）である。

対立仮説：A君はキャラクターBより弱い（互角でない）。

A君がキャラクターBに勝つ確率は（ ）ではない。

1. 二項分布に従うとしたらどうなるか？

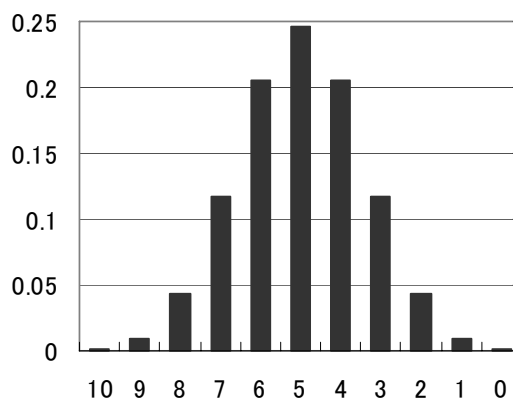
A君がキャラクターBと互角に戦う（勝つ確率0.5）としたら、二項分布に従って

A君が1勝以上しない確率を計算する

A君がちょうど1勝する確率

A君がちょうど0勝する確率

1勝だけの確率をみるのではなく、どれだけ中心から外れているかの確率を計算するので、0勝の場合も計算しなければならない



2. 有意水準と検定結果

① 有意水準

仮説が正しいにもかかわらず、帰無仮説を棄却する確率を（ ）という。

（ ）には5%、1%、0.1%などがよく使われる。

ここでは有意水準を5%としてみよう。

帰無仮説が正しいとすると今回のようなケース（A君が1勝あるいは0勝しかしないということ）が起こる確率は（ ）であった。

帰無仮説は棄却できるだろうか？

★ 5%の有意水準で帰無仮説は（ ），（ ）と判断できる。

★ 1%の有意水準で帰無仮説は（ ），（ ）と判断できる。

② 帰無仮説を検定する

帰無仮説が棄却される→対立仮説を採用する

帰無仮説が棄却されない→対立仮説の採用には証拠が不十分である

練習：あるチョコレートには0.2の割合で当たりくじが入っているとメーカーは主張する。10個買ったところ、1つも当たりくじが入っていなかった。メーカーの主張の真偽を有意水準5%で検定せよ。

帰無仮説：

対立仮説：

有意水準を5%とする。

計算：

検定結果：

有意水準5%で帰無仮説は（棄却される・棄却できない）。

したがって、

- 1) 対立仮説を採用する→
- 2) 対立仮説は採用できない→

2. 第1種の誤りと第2種の誤り（有意水準と検出力）

① 第1種の誤り

帰無仮説が正しいにもかかわらず、これを否定する過誤を第1種の誤りという。例1では本当はさいころは正しくできているのに、たまたま5回続けて奇数が出たなら、第1種の誤りを犯すだろう。

第1種の誤りを犯す危険率を**有意水準**といい、 α と表すことがある。有意水準を小さくして、検定すれば、第1種の誤りを犯す危険は小さくなる。さいころの例でいえば、10回続けて奇数が出なければ、帰無仮説を否定しないとすれば、有意水準0.001で検定したことになり、そういうことは1000回に1回しかないのだから、帰無仮説を否定できると結論できる。

② 第2種の誤り

しかし、5回続けて奇数の出るさいころが正しいさいころであるよりは、奇数の出やすいさいころと考える方がありそうである。このように帰無仮説が誤りであるにもかかわらず、これを否定しない過誤を第2種の誤りという。第1種の誤りを減らすために有意水準を小さくすれば、第2種の誤りの危険率(β)は高くなり、すなわち検出力($1 - \beta$)は低下する。2種類の誤りと検定結果には下の表のような関係がある。

表 検定における2種類の誤り

		本当に成り立っているのは	
		帰無仮説	対立仮説
検 定 結 果	帰無 仮説	正しい (その確率： $1 - \alpha$)	第2種の誤り (その確率： β)
	対立 仮説	第1種の誤り (その確率： α)	正しい (その確率： $1 - \beta$)

有意水準を小さくすれば、第1種の誤りをおかす危険性は(大きく・小さく)なる。しかし、第2種の誤りをおかす危険性は(大きく・小さく)なる。

統計的検定では有意水準を決めることによって、第()種の誤りの大きさを制御している。すなわち第()種の誤りを避けることに重点が置かれている。このことは統計的検定では貴重な発見を見落としてしまっても疑わしい結果を受け入れることを避けているのである。

C. 母分散が既知のときあるいは大標本の平均に関する統計的検定

1. ある決まった平均に対する検定

① 両側検定(通常検定)

標本平均は標本の大きさが十分大きければ正規分布に従う。そこで、正規分布の性質から母集団から得られた標本の平均がある値と異なるかを検定することができる。

例：A農場で出荷する桃の重さの標準偏差 σ は20gである。桃の重さの平均が150gとなるように出荷する。ある日の調査では100個の桃の重さを量ったところ、平均145gであった。桃の重さの母平均は150gではないのかを検定せよ。

帰無仮説 H_0 :

対立仮説 H_1 :

① 帰無仮説が成り立つとすると

母平均は () g であり, 母標準偏差は () g である. 標本数は () 個であるから, 標準誤差は () g である. したがって, 標本平均は平均 () g, 標準偏差 () g の正規分布に従う.

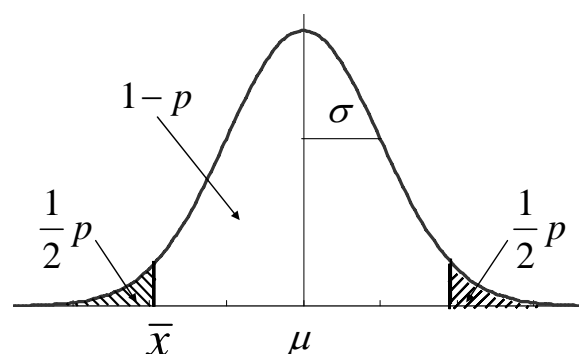
② 今回のデータが得られる確率を計算する

今回の標本平均は () g であるから, この標本平均が得られる確率 P 値はエクセルの **NORMSDIST** 関数で計算できる.

得られた標本の標本平均を \bar{x} , 帰無仮説でとりあげた母平均を μ , 標準誤差を SE とすると, 今回の標本が帰無仮説の下で得られる確率 (P 値) は以下のように計算できる.

$$P \text{ 値} = 2 \times \text{NORMSDIST}(-\text{ABS}(\bar{x} - \mu) / SE)$$

したがって, 今回の標本が得られる確率は,



帰無仮説 $\mu = 150$ が成り立つとき, 今回の標本 $\bar{x} = 145$, $n = 100$ が得られる確率は () である.

したがって, 有意水準 5% では帰無仮説 H_0 は棄却される (有意水準 5% で有意である). すなわち対立仮説 H_1 を採用し, 有意水準 5% でこの標本の母平均は 150g ではないと結論できる.

しかし有意水準 1% では帰無仮説 H_0 は棄却されない (有意水準 1% で有意でない). すなわち対立仮説は採用できないので, 有意水準 1% でこの標本の母平均は 150g ではないとはいえないと結論される.

練習: C村は塩分の濃い食事でも有名であり, 寿命が短いといわれる. 村民の平均寿命は 100 人調べたところ, 70.2 歳, 標準偏差は 0.9 歳だった. 平均寿命が 70 歳であるかを有意水準 5% で検定せよ.

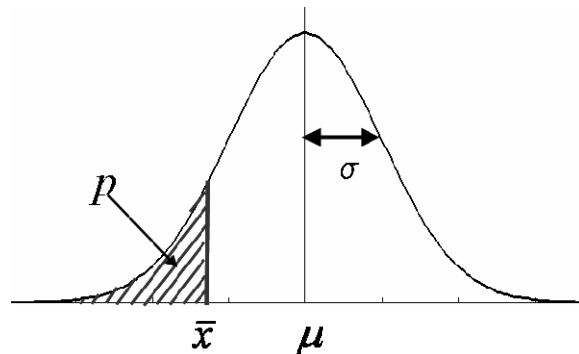
② 片側検定 どちらか一方に変化したことが予想できる場合

例：B牧場では牛のえさをF社からG社に変えた。G社のえさはより栄養価が高いので、F社
ときの泌乳量 5.0L, 標準偏差 0.8L より向上すると予想した。実際に 100 頭を調査した結果、
泌乳量は 5.2L となった。標準偏差は変わらない。泌乳量 (の平均) が増加したかを検定せよ。

帰無仮説 H_0 :

対立仮説 H_1 :

$$P \text{ 値} = \text{NORMSDIST}(-ABS(\bar{x} - \mu) / SE)$$



したがって、今回の標本が得られる確率は、

帰無仮説 $\mu = 5.0$ が成り立つとき、今回の標本 $\bar{x} = 5.2$, $n = 100$ が得られる確率は
() である。

したがって、有意水準 1% で帰無仮説 H_0 は棄却される (有意水準 1% で有意である)。すなわち対立仮説 H_1 を採用し、有意水準 1% でこの標本の母平均は 5.0L よりも増加したと結論できる。

このようにある一方向に対する検定では有意になりやすくなる。しかし、これはあらかじめ泌乳量が増加するという情報があるからである。たいていの場合、こういう片側検定を使うことは望ましくない。少なくとも実験を始める前にどういう検定をするかを決めた場合でないと使わない方がよい。

2. 2つの独立した母集団から得た2つの標本平均についての検定

それぞれ無作為に抽出した2つの標本から得た平均が異なっているかを検定する。

例：イネの1穂穎花数について100本ずつ調査し、品種Aは 150 ± 12 、品種Bは 155 ± 16 を得た (平均 \pm 標準偏差)。両品種の1穂穎花数 (の平均) は異なるのかを検定せよ。

この場合、2つの標本の平均の差が0であるという帰無仮説を検定する。2つの標本の平均の差も正規分布に従う。ここで2つの標本から得た標準偏差を一緒にする必要がある。2つの独立した母集団では1つの母集団にすると分散は両分散の和になるという性質がある（詳しい証明は略する）。このことから標本数 n_A の標本（標本平均 \bar{x}_A 、標準偏差 σ_A ）と標本数 n_B の標本（標本平均 \bar{x}_B 、標準偏差 σ_B ）の2つの独立した標本を得たとき、次のことがいえる。

$\bar{x}_A - \bar{x}_B$ の分布は正規分布で、 $N(\mu_A - \mu_B, \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B})$ になる。

帰無仮説：

対立仮説：

$$\begin{aligned} P \text{ 値} &= 2 \times \text{NORMSDIST}(-\text{ABS}((\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)) / \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}) \\ &= 2 \times \text{NORMSDIST}(-\text{ABS}((\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)) / \text{SQRT}(\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B})) \end{aligned}$$

計算結果

したがって、有意水準5%で帰無仮説を棄却できる（両側検定）。すなわち2つの品種の1穂穎花数（の平均）は異なると結論できる。

練習：20～24歳男子 2112人の身長は $170.5 \pm 5.9\text{cm}$ ，30～39歳男子 2776人の身長は $169.5 \pm 5.8\text{cm}$ （人間生活工学研究センター1992-1994年調査）であった。2つの母集団の間に平均身長は差があるかを検定せよ。

このようにサンプル数が十分あれば標準誤差が小さくなるので、平均に関する検定における検出力は高くなる。個々の標本については逆転することも多々ある（標準偏差の大きさからもわかるように分布はかなり重なる）が、平均については中心極限定理が保証する。

D. 統計的検定の手順のまとめ

1. 帰無仮説をたてる。
2. 対立仮説をたてる。ふつうは両側検定とする。
3. 有意水準を決める。

有意水準は統計的検定の計算を始める前に決める。可能ならば実験を始めたり、データを集めたりする前に決めるのが望ましい。とくに統計によって何らかの判断を検定結果から行うときは有意水準を事前に決めなければならない。P値をみて判断するというのでは、統計的な方法をつかわないで恣意的にカンで物事を決めるのとほとんど同じことである。

データの提示だけを行うときにはデータにP値あるいは有意差を示す記号を付ける。P値をつけるのがより明確だが、以前はP値の計算が面倒であったので、慣例として有意差を示す記号を付ける方が今でも一般的である。*は5%水準で、**は1%水準で、***は0.1%水準で有意差があることを示すことが一般的である。しかし、この記号は慣例で使われているだけなので、確認して使うべきである。

4. P値を計算する。
5. 統計的検定を行う。

P値が最初に決めた有意水準よりも小さければ、帰無仮説は棄却され、有意差があると判断できる。帰無仮説が棄却できないときは「差があるとはいえない」という結論になる。帰無仮説を採用して、「差がない」という結論にしてはいけない。

6. 現実的な意味での差があるかの判断

統計的に有意差があることと現実的な意味でその差に意味があるかというのは別問題である。すなわちP値が小さければ小さいほど、差があることを統計的に強く確信が持てるとはいえる。しかしP値が小さければ小さいほど、現実的な意味での価値が増すというわけではない。

E. 宿題

1. 第5回の宿題2. で調査したデータについて帰無仮説は2人の間に実力差はないとし、有意水準（危険率）を5%として、二項分布するという仮定の下で検定せよ。なお授業では片側検定として検定したが、ここでは両側検定を行うこと。
2. 第3回の宿題で調べたデータについて、今回返した前回の宿題の講評の最後に書いてあることについて検定を行え。
3. 太平洋に浮かぶ2つの島（X島，Y島）に共通する鳥であるP鳥は毎年、たくさんの卵を産む。しかし、Y島の近くに重油タンカーが座礁したため、卵の重さがX島と異なるかもしれない。両島で卵をそれぞれ無作為に100個ずつ採集したところ、X島では $35.0 \pm 1.2\text{g}$ 、Y島では $34.6 \pm 1.1\text{g}$ の値（平均±標準偏差）を得た。両島のP鳥の卵の重さの母平均は異なるのかを1%の有意水準で両側検定せよ。
4. 2つの別々のスーパーから10個入りの卵（M寸）を買い、両店の卵の重さは同じなのかを調べる。ここでは、卵の重さをはかり、平均と標準偏差を計算する。ここで得た平均と標準偏差から正規分布となる母集団のグラフをだいたい書いてみる。さらに95%信頼区間をつけた母集団の平均の推定値をそれぞれについて求める。そこから両店の卵の重さが同じかを考えて見よ。（同じようなデータであれば、卵でなくてもよい。下の例を見よ。2人一組で班をつくって、調査してもかまわない）
有効数字が3つ以上になるようにデータをとること。すなわち28.1g, 1.53cmのようにデータをとること。

例：2つの店のラーメンの重さ

2人のソフトボールの円筒距離

2人の50m走

補足 統計的検定の結果の表記について

1. 有意水準をかならず付けて表記すること

P値が0.05より小さく、0.01より大きい場合、5%の有意水準では有意であり、1%の有意水準では有意ではない、すなわち5%の有意水準では2つの標本の母平均に差がある、1%の有意水準では差があるとはいえないという結論になる。このときときどき差があるといいつつ、あるとはいえないともいうのかと混乱する人もいるかもしれない。あくまでも宣言した有意水準において差がある、差があるとはいえないと結論しているのであり、統計的検定においてはかならず有意水準を付けなければならない。

検定の結果を次の行動につなげる場合はかならず有意水準を決めてから検定しなければならない。この場合、有意水準 α %で検定せよということになり、もしP値が p %であれば、結論は以下ようになる。

$p < \alpha$ 有意水準 α %で有意差である。有意水準 α %で母平均に差がある。

$p > \alpha$ 有意水準 α %で有意差がない。有意水準 α %で母平均に差があるとはいえない。

とくに有意水準を指定しない場合は以下のように表記するのが一般的である。

$p > 5\%$ 有意水準5%で有意差がない。有意水準5%で母平均に差があるとはいえない。

$1\% < p < 5\%$ 有意水準5%で有意差がある。有意水準5%で母平均に差がある。

$0.1\% < p < 1\%$ 有意水準1%で有意差がある。有意水準1%で母平均に差がある。

$p < 0.1\%$ 有意水準0.1%で有意差がある。有意水準0.1%で母平均に差がある。

有意水準5%、1%、0.1%で有意差があるときに*、**、***をデータにつけることがある。

2. P値が小さいことと処理の効果が大きいことは別である

ときどきより小さい有意水準で有意である、すなわちP値が小さい方が効果が大きいと勘違いする人がいる。しかしP値が小さいことは、2つの標本の母平均が異なることをより強く確証できるだけであり、効果そのものの大きさとは関係がない。例えば、効果の強い薬を3人程度で比較試験してもかならずしも有意差を得られるとは限らない。すなわちP値はあまり小さくはならない。一方、効果のそれほど強くない薬でも数万人も使って、試験すれば、かなり小さなP値を得るかもしれない。

このようにP値の大きさは標本の大きさに大きく依存する。

さらに有意差がなくてもその薬に効果がないということを証明したわけではない。あくまでも帰無仮説が棄却されないということは、差があるとはいえないと結論しただけである。もし3人程度の実験で有意差を得られない薬でも10人程度で実験すれば有意差が得られるかもしれない。

このようにP値はあくまでも統計的に2つの標本の母平均に差があるかどうかの判断基準を確率によって与えるだけであり、その2つの母平均の差が大きい小さいかはP値からは判断できない。