統計的仮説検定とは

瀬尾 隆（理学部 第一部応用数学科 数理データサイエンス部門）

1　はじめに

統計科学は，数理データサイエンス分野の中核であり，自然科学や社会科学を問わず多くの分野で用いられるデータ解析の手法を開発する研究分野です．特に現実社会では，多次元データを取り扱う統計解析である多変量解析（重回帰分析，判別分析，主成分分析，相関分析，クラスター分析，統計的多重比較法，データマイニングなど）がコンピュータの急速な発展とともに注目を浴びています．日常生活の中で AI や機械学習といった言葉を耳にしたことがあるかと思いますが，それらにも深く関係する分野となっています．このような統計解析では，パソコン上で大量のデータを整理し，そのデータを統計解析ソフトを用いて解析したり，自分でプログラムを組むなどして解析していきます．最近はデータを入力すると解析結果が分かりやすく出力され，手軽に分析できるようになってきました．しかしながら，統計データ解析は単にデータを統計的手法にあてはめて計算するだけでなく，パソコンから出てきた数値的結果についての考察やその解釈を考えなければなりません．また，分析しようとするデータが分析法の条件に沿ったものであるか吟味した上で，適した統計解析ソフトを使用しないと間違った解析結果や誤った結論を導くことになります．間違った使い方や誤った解釈などしないようにするためには，統計的解析手法の理論を十分に理解していることが不可欠です．またプログラム作成についても同様です．ここでは，その統計理論の重要な考え方のひとつである「統計的仮説検定」についてお話します．この統計的仮説検定は，高校数学の新しい数学 I や数学 B に含まれている内容となっています．

2　統計的仮説検定

ここでは，仮説検定とは何かを説明するために杉山，藤越 (2009) [1] にある平均の検定の例を使って紹介したいと思います．例えば，新しく開発されたある薬の薬効持続時間が従来の薬と比べて長くなったかどうかを調べるため，10 人の患者に対して臨床実験を行ったところ，次のデータを得たとします．

表 1: データ

100, 107, 98, 103, 111, 95, 100, 105, 103, 108

この表 1 のデータから，平均値，標本分散，標準偏差をそれぞれ計算すると

$$\overline{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 103, s^2 = \frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x})^2 = 24,\ s = 4.9$$

となります．一方，従来の薬の薬効時間は，平均 100，標準偏差 5 の正規分布 $N(100, 5^2)$ に従うことがわかっているとします．このとき，新薬の平均薬効時間は従来の薬の平均薬効時間と比べ

[1] 参考書：「統計データ解析入門」，杉山高一，藤越康祝 編著，医学評論社，2009

て長くなったと結論づけられるでしょうか？観測誤差の範囲内ではないのか，それとも本当に差があってそのような結論が得られたのではないのか，ということです．このような問題に対して，客観的な基準に基づいて効果の有無を判定する方法に「統計的仮説検定」があります．つまり，「新しく開発された薬の薬効時間は従来のものと同じである」という仮説をたて，その仮説を否定するか否かを確率を用いて判断する方法を統計的仮説検定といいます．新薬と従来の薬効時間は同じであるという仮説は，平均を μ とすると，$\mu = 100$ とおくことができ，これが検定したい仮説となります．通常，仮説は当事者が疑わしいと思われる内容をおくもので，初めから捨てるつもりでたてる仮説を帰無仮説といい，H_0 で表します．一方，望んでいる仮説 $\mu > 100$ を対立仮説と呼び，H_1 で表します．したがって，この仮説検定では

$$\text{帰無仮説 } H_0 : \mu = 100 \qquad \text{vs.} \qquad \text{対立仮説 } H_1 : \mu > 100$$

の 2 種類の仮説についての真偽を判定する問題に帰着されます．

そこで，帰無仮説 H_0 が真であるとき，与えられた標本がどの程度の確率で出現するか考えてみます．いま，標準偏差に関しては従来のものと新薬で変化がないと仮定すると, 新薬に関する母集団分布は $N(100, 5^2)$ となることから，大きさ 10 に基づく標本平均 $\overline{X}$ の分布は $N(100, 5^2/10)$ となります．このとき，標本平均 $\overline{X}$ がデータの平均値 103 以上となる確率は，Z を標準正規分布 $N(0, 1)$ に従う確率変数とすると

$$\begin{aligned}\Pr(\overline{X} \geq 103) &= \Pr\left(\frac{\overline{X} - 100}{5/\sqrt{10}} \geq \frac{103 - 100}{5/\sqrt{10}}\right)\\ &= \Pr(Z \geq 1.90)\end{aligned}$$

と書くことができます．ここで，正規分布表をみるとこの値は $0.0287 (= 2.87\%)$ となり，確率は非常に小さいことがわかります（図 1 参照）.

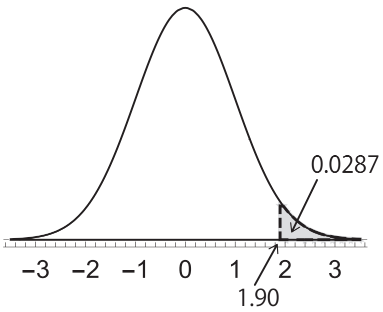

図 1：標準正規分布と $\Pr(Z \geq 1.90) = 0.0287$ の説明

つまり，帰無仮説 H_0 が真のときに，同じような実験を 100 回繰り返したとすると，平均値が 103 以上となることが平均約 3 回しかおこらないことを意味しています．このことがたった 1 回の実験で起こった．つまりめったに起こらないことが起こったということになります．この場合，帰無仮説 H_0 が正しくて，めったに起こらないことがたまたま偶然に起こったと考えるより，帰無仮説 H_0 は正しくなかったと判定して，仮説を否定します．よって，仮説を棄却します．まとめると，母集団に関してたてた仮説 $\mu = 100$ は正しくなかったとみなし，この仮説を棄却して $\mu > 100$ であると結論づけます．この考え方が統計的仮説検定の考え方です．

ちなみに，非常に小さい確率で生じたとは考えられないとき，どのように判断したらよいでしょうか？このときは，帰無仮説 H_0 を棄却できないので，結論を保留あるいは帰無仮説をとりあえず

は容認しておきます．このことは帰無仮説 H_0 が正しいことを意味するものではありません．なぜなら帰無仮説が真でないときに，帰無仮説を採択する誤りの確率を評価していないからです．したがって，統計的仮説検定の特徴は，帰無仮説が棄却されたときに初めて積極的に何かを主張することができるということになります．帰無仮説を棄却する基準としては，通例 0.05 あるいは 0.01 が用いられ，これを有意水準あるいは危険率と呼んでいます．

3　まとめ

母集団分布が正規分布に従う場合の母平均の仮説検定を例にして，統計的仮説検定の考え方を紹介しました．今回は（1 次元の）正規分布でしたが，多次元データになると多次元正規分布（多変量正規分布（図 2 参照））というものがあり，平均ベクトルの検定を行うことができます．詳細は大学に入ってから学ぶことになるかと思います．また，本学では AI・データサイエンス教育として，大学 1 年次より「AI・データサイエンス概論」を学ぶことができるとともに「データサイエンス教育プログラム [基礎]」という認証書を学長名で取得できる制度もあります．応用数学科は統計学やデータサイエンスに関する科目が数多く開講されており，応用数学科で開講されている科目のみで修了要件を満たせば認証書を取得できるようになっています．

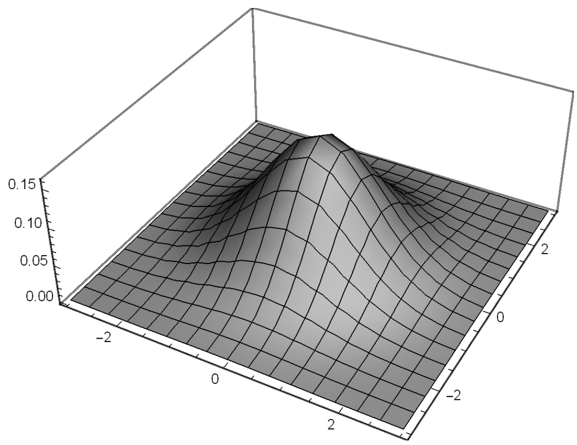

図 2：多変量正規分布（2 次元の場合）