

【態度目標】しゃべる、質問する、説明する、動く、協力する、貢献する

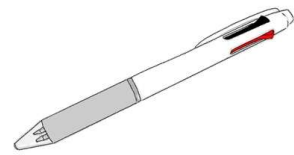
【内容目標】推測が正しいかどうかを判断する方法を学ぼう

□ 仮説検定の考え方

集団に対して調査を行う場合、調べたい集団の全体のデータを集めることは困難な場合が多い。そのようなときに、調べたい集団から一部を抜き出して、そのデータから集団全体の状況を推測することがある。

ここでは、その推測が正しいかどうかを判断する1つの考え方について学ぼう。

ボールペンを製造している会社が、すでに販売しているボールペン A を改良して新製品 B を開発した。B が A よりも書きやすいと消費者に評価されるかを調査したいと考えたが、すべての消費者を調査するのは不可能である。



そこで、無作為に選んだ 30 人に 2 つのボールペン A, B を使ってもらい、どちらが書きやすいと感じるかを回答してもらった。

その結果を集計したところ、70 % にあたる 21 人が B と回答した。この回答のデータから、

[1] Bの方が書きやすいと評価される

と判断できるだろうか。

この問題を解決するために、[1]の主張に反する次の仮定を立てよう。

[2] A, B のどちらの回答も全くの偶然で起こる

すなわち、A, B のどちらの回答の起こる確率も $\frac{1}{2} = 0.5$ である、という仮定を立てる。

その仮定のもとで、30 人中 21 人以上が B と回答する確率がどれくらいかを考察しよう。

[2]の仮定は、公正な 1 枚のコインを投げる実験にあてはめることができる。ここでは、コインの表が出る場合を、B と回答する場合とする。

そして、コイン投げを 30 回行うことを 1 セットとし、1 セットで表の出た回数を記録していく。

この実験を 200 セット繰り返したところ、次の表のような結果となった。

表の回数	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	計
度数	2	3	3	12	16	22	22	31	31	22	14	12	6	2	1	1	200

<補足> この実験の代わりに、コンピュータでシミュレーションを行ってもよい。

上の表から、21 回以上表が出たのは、200 セットのうち $2 + 1 + 1 = 4$ セットであり、相対度数は

$\frac{4}{200} = 0.02$ である。

つまり、A、Bのどちらの回答も全くの偶然で起こるとした[2]の仮定のもとでは、21人以上がBと回答する確率は0.02程度であると考えられる。

これは見方を変えると、0.02程度という確率の小さいことが起こったのだから、そもそも[2]の仮定が正しくなかったと考えられる。そう考えると、[1]の主張は正しい、つまりBの方が書きやすいと評価されると判断してよさそうである。

得られたデータをもとに、ある主張が正しいかどうかを判断する、右のような手法を**仮説検定**という。

また、上では0.02を確率が小さいとしたが、仮説検定では基準となる確率（有意水準（危険率））をあらかじめ決めておき、それより小さければ確率が小さいと判断する。

「有意水準が5%」とは
「仮説が正しいのに誤って棄却してしまう確率が5%」ということ

例11) ボールペンの調査で、30人中19人がBと回答したとする。

主張[1] Bの方が書きやすいと評価されるが正しいと判断できるか。基準となる確率を0.05として考察してみよう。

コイン投げの実験結果を利用すると、19回以上表が出る場合の相対度数は

表の回数	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	計
度数	2	3	3	12	16	22	22	31	31	22	14	12	6	2	1	1	200

$$\frac{12+6+2+1+1}{200} = \frac{22}{200} = 0.11$$

これは0.05より大きいから、

主張[2] A、Bのどちらの回答も全くの偶然で起こるは否定できない。

したがって、Bの方が書きやすいと評価されるとは判断できない。終

<注意> 例11について、「主張[2]が正しい」という判断ができるわけではない。

なお、前ページや上の例11ではコイン投げの実験結果を利用しているが、通常は計算で確率を求め、それを利用する。

主張[1]が正しいと判断できるか

対立仮説 (Alternative hypothesis, H_a や H_1 で表す) という

主張[2] (主張[1]と反する仮定) を立てる

帰無仮説 (きむかせつ null hypothesis H_0 で表す) という

主張[2]のもとで、実際に起こった出来事が起こる確率を調べる

実際に起こった出来事が起こる確率はかなり小さい

そもそも、主張[2]の仮定が正しくなかった

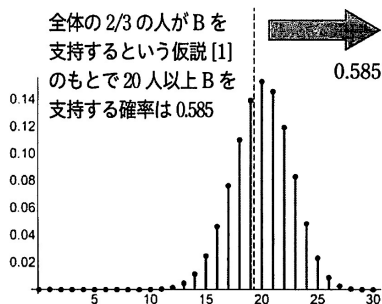
主張[1]は正しいと判断してもよいと考えられる

棄却域に入らなかった場合に「帰無仮説が正しい」とはならない。これは「検出力が十分ではなかった」という可能性があるためである。

検定の精度の基準として有意水準があるが、もう1つ重要な基準として検出力というものがある。
 検出力とは、対立仮説が正しいという仮定のもとできちんと正しく対立仮説を採用する確率のことである。

例えば仮説 [1] が正しいとし、Bの方を支持する人がAの2倍
 いるとする。すなわち、全体の $\frac{2}{3}$ の人がBの方が書きやすいと
 判断するとする。このとき、前述の検定ルールのもと、仮説 [1]
 が正しいと判断できる確率は 0.585 である。すなわち、Bの方
 を支持する人がAの2倍もいるにも関わらず正しく判断できる
 確率は 60 %未満である。

この検出力を高めるには調査する人数を増やすことで、検出
 力を高めることができる。例えば、100 人に調査をして 59 人
 以上が支持すれば仮説 [2] を棄却するとすると、有意水準は
 0.0443、検出力は 0.957 となる。



発展 仮説検定と反復試行の確率

ボールペンの書きやすさの調査に関する仮説検定において、「A、B のどちらの回答も全くの偶然で起こる」という仮定のもとで、30 人中 21 人以上が B と回答する確率を、コイン投げの実験を通して考えた。この確率は、数学 A で学習する次の「反復試行の確率」を用いると、計算することができる。

その結果が偶然によって決まる実験や観測を試行という。また、試行の結果として起こる事柄を **事象** という。

反復試行の確率（数学 A のおさらい）
 1 回の試行で事象 A の起こる確率を p とする。
 この試行を n 回繰り返して行うとき、事象 A がちょうど r 回起こる確率は

$${}_n C_r p^r (1-p)^{n-r}$$

<補足> ${}_n C_r$ は異なる n 個のものから r 個取り出して作る組合せの総数を表す。

A、B どちらの回答の起こる確率も $\frac{1}{2}$ であるという仮定が正しいとする。このとき、30 人中 21 人以上が B と回答する確率は

$${}_{30} C_{21} \left(\frac{1}{2}\right)^{21} \left(\frac{1}{2}\right)^{30-21} + {}_{30} C_{22} \left(\frac{1}{2}\right)^{22} \left(\frac{1}{2}\right)^{30-22} + \dots + {}_{30} C_{29} \left(\frac{1}{2}\right)^{29} \left(\frac{1}{2}\right)^{30-29} + \left(\frac{1}{2}\right)^{30}$$

となる。これをコンピュータで計算すると、 $\frac{22964087}{1073741824} = 0.0213\dots$

となる。コイン投げの実験で求めた相対度数 0.02 は、この確率に近い値である。

補足 コインを投げたときの表の枚数は二項分布に従い、更に投げる回数
 十分に大きい場合は正規分布に近似できる。このような正規分布を
 利用した仮説検定は数学 B の統計で行う。

