データの分析【仮説検定】　p.194～197

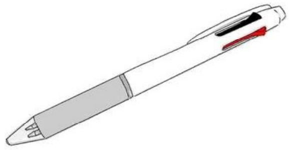
1 学習内容の説明 ⇒ 2 問題演習 ⇒ 3 振り返り（確認テスト・相互採点・リフレクションの記入）

【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。

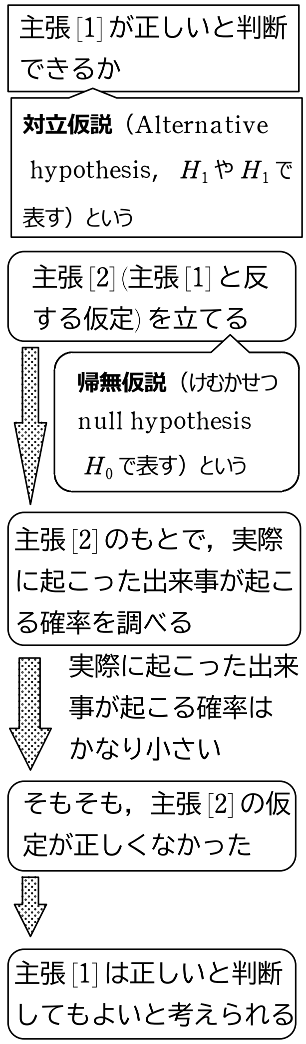
上の表から，21 回以上表が出たのは，200 セットのうち $2+1+1=4$ セットであり，相対度数は $\frac{4}{200}=0.02$ である。

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。



例１３） 194 ページの調査で，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。

前ページの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１３ 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１３ 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。