

目次

- 仮説検定のキモと流れ (数学Ⅰ・データの分析)
- 確率分布と仮説検定 (数学Ⅱ・統計)

誘惑のない動画や公式検索アプリ **okke**



仮説検定のキモと流れ (数学Ⅰ・データの分析)

「あるサイコロを6回投げたところ、6の目が4回出た。」

→ 直観的にはさすがに6の目に偏ってそう。

これを統計的に示したい!

仮説検定では、以下の流れで背理法的に検証していく。

- ① サイコロに偏りが無いとする。
- ② このとき、6回中6が4回以上出る確率を求めろ。
- ③ それが小さければサイコロは偏っていると判断する。

基準となる確率はあらかじめ決めておく! (有意水準)
0.05が多い。厳しいと0.01なども。

つまりこんな感じ。

判断したいこと。

6の目が出やすい
偏ったサイコロだ！

いやいや。

偏ってない！

これがホントだとして、今回のように
「6が4回以上出る」ってありえんの??

計算！

十分起こりにくい
($<$ 有意水準)

(そんなだけ起こりにくいことが
起こったということは…)

じゃあ確かに6の目が出やすいサイコロだ！

割と起こる
($\geq$ 有意水準)

たまたま多く出ただけ
でした。残食

実際に有意水準 = 0.05 として検証してみる。

偏りがない、つまり 6 の目が出る確率が $\frac{1}{6}$ と仮定すると、

6 回投げて 6 が 4 回以上出る確率は

$$\underbrace{{}_6C_4 \times \left(\frac{1}{6}\right)^4 \times \left(\frac{5}{6}\right)^2}_{4\text{回}} + \underbrace{{}_6C_5 \times \left(\frac{1}{6}\right)^5 \times \left(\frac{1}{6}\right)}_{5\text{回}} + \underbrace{\left(\frac{1}{6}\right)^6}_{6\text{回出る}} \quad \leftarrow \text{単なる確率の問題}$$

$$= \dots = \frac{406}{6^6} = \underline{0.0087\dots} \quad \leftarrow \text{めちゃくちゃ起こりにくいことが起こっていることになる! (「有意に」起こりにくい)}$$

となり、有意水準 0.05 より小さいので、

偏りのあるサイコロだと判断できる!

言葉をおさえよう!

対立仮説 (H_1)

6の目が出やすい
偏ったサイコロだ!

★判断したいこと.

^{きむ}帰無仮説 (H_0)

偏ってない!

「無」に「帰」することを
予定している.

いやいや.

これがホントだとして、今回のように
「6が4回以上出る」確率を計算

<有意水準

H_0 は^{ききやく}棄却され、
 H_1 を採択する.

(じゃあ確かに6の目が出やすいサイコロだ!)

≥有意水準

H_1 は^{ききやく}棄却される

(たまたま多く出ただけ)
でした. 残念

大事なこと.

あくまでも判断するのは対立仮説を棄却するか採択するか.

仮説検定の結論は

- ① 6の目が出やすい偏ったサイコロだ (対立仮説を採択)
 - ② 6の目が出やすい偏ったサイコロとは言えない (対立仮説を棄却)
- のどちらか!

→ 帰無仮説はその判断のために置かれたものなので、

③ 偏っていないサイコロだ (← 帰無仮説)

ということも証明したわけではないことに注意!! 矛盾しない. <どい>の感じ

確率分布と仮説検定 (数学B・統計)

流れは全く同じ!!

「500mLと書かれているペットボトルについて、
400本選んで内容量を調べたところ、
平均値は498mL、標準偏差は25mLだった」

標本平均

標本標準偏差



「内容量の平均値は500mLではない」を示すには??

有意水準 = 0.05 とする。 ← あらかじめ決めとく。

対立仮説 (H_1)

内容量の平均値は
500mLではない!

帰無仮説 (H_0)

500mLだ!

いやいや.

これがホントだとして、今回の値を
もとに、統計量を計算

←ここがメイン!!

★判断したいこと.

有意水準を $\rightarrow$ 棄却域に含まれる
もとに求める.

H_0 は棄却され.

H_1 を採択する.

(内容量の平均値は
500mLではない!)

棄却域に含まれない

H_1 は棄却される

(たまたま少なかっただけ
でした. 残念)

という流れ! 確率よりも棄却域の方が考えやすい.

実際に計算してみる。 → この流れは定番!!

十分に標本が大きいので、標本標準偏差 $s = 25$ (ml)

を母標準偏差とみなすと、400本の標本平均 $\bar{X}$ の

期待値と標準偏差は、

← 確率変数!

$$\begin{cases} E(\bar{X}) = 500 & \leftarrow \text{帰無仮説を仮定しているため} \\ \sigma(\bar{X}) = \frac{25}{\sqrt{400}} = 1.25 \end{cases}$$

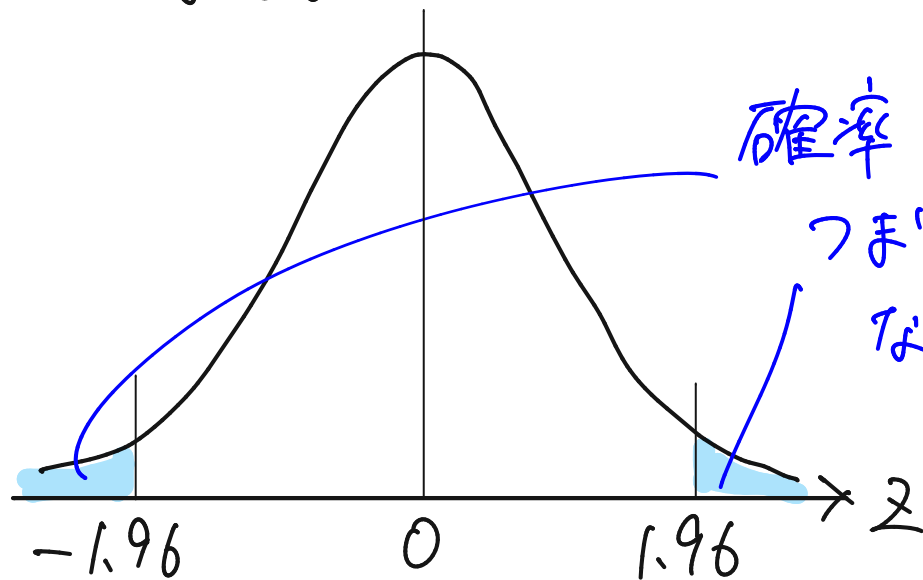
となる。十分に標本が大きいので、 $\bar{X}$ は正規分布 $N(500, 1.25^2)$ に従う。とみなせる。(中心極限定理!)

→ あとは、正規分布 → 標準化 → 標準正規分布
で定番の確率計算!

確率変数 Z を

$$Z = \frac{\bar{X} - 500}{1.25} \quad \leftarrow \text{標準化}$$

で定めると、 Z は標準正規分布 $N(0, 1)$ に従うことになる。



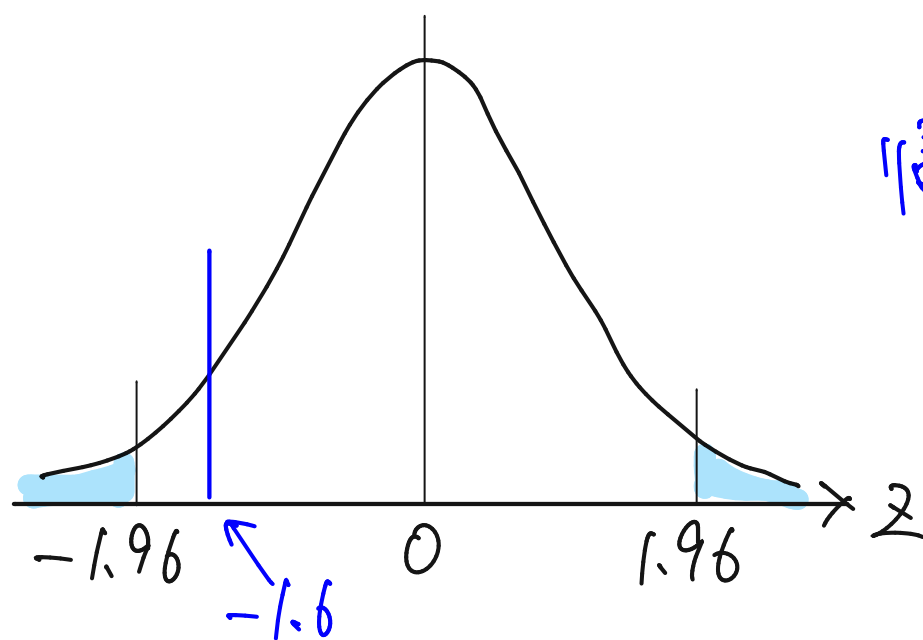
確率 0.05 になる！

つまり、統計量 Z がこの範囲の値になれば「帰無仮説を棄却する」。
(これを 棄却域 という！)

今回の観測された標本平均の値 $\bar{X} = 498$ を用いると、
 Z の値は

$$Z = \frac{498 - 500}{1.25}$$

$= -1.6$ なので、棄却域に含まれない。



↓
帰無仮説は
棄却されない!

よって 対立仮説は棄却され、

「内容量の平均値は 500 mL ではない」とは言えないと判断される。

今回は両側検定で行ったが、

ペットボトルのサイズ的に、

内容量の平均値が 500 mL を超えることはないと

わかっているときなどは、両側に棄却域を設定すると

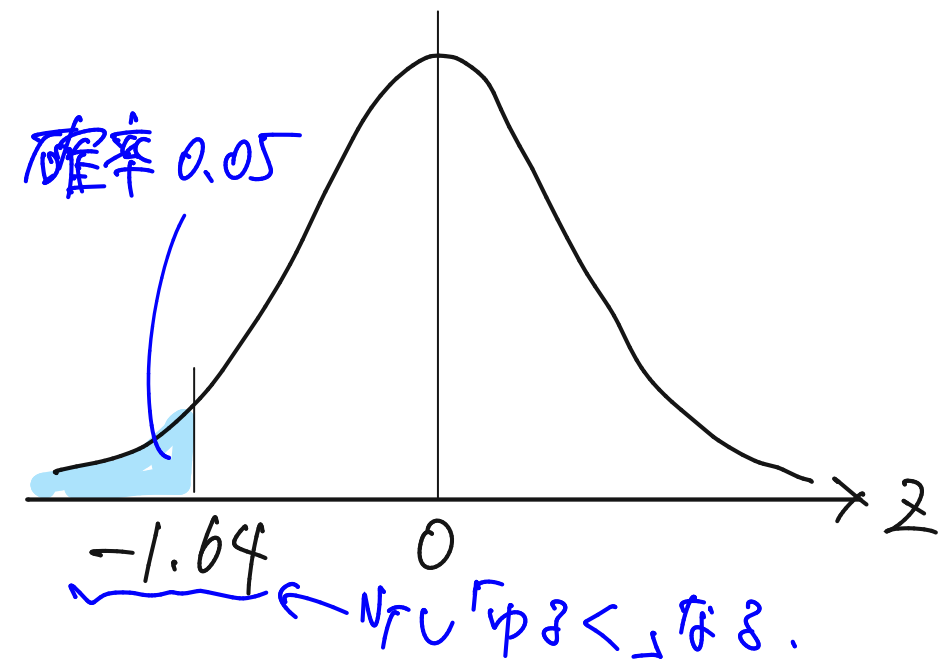
必要以上に厳しく検定してしまうことになる。そのときは、

(対立仮説「内容量の平均値は 500 mL 未満」

帰無仮説「内容量の平均値は 500 mL」

として検定を行えばよい。

このとき、それに合わせて棄却域も片側のみに
設定することになるので、片側検定という。



(補足) 動画では説明していません…!

仮説検定はあくまでも統計的な判断なので、'帰無仮説'や対立仮説を棄却したが「本当は」正しかった、めたいなこともある。

神のみぞ知るケースも多い。

① '帰無仮説'を棄却したが、「本当は」正しかった → 第1種の過誤

② 対立仮説を棄却したが、「本当は」正しかった → 第2種の過誤

という。

△ この2つはトレードオフ (共に確率を下げることはできない)なので、状況に応じてどこまで許容するかを決めて、有意水準を定めることになる。

たとえば薬の効果を判断したいときは、効果ないものを、認可したくない(= 第1種の過誤の確率を下げるたい)ので、有意水準を低くする!