

11. 統計的検定と平均値の比較

honocat

2025-12-23

統計的検定

有意水準の意味を理解する：帰無仮説が正しい場合の仮説検定シミュレーション

標準正規分布に従う変数 X について考える。つまり、

$$X \sim \text{Normal}(0, 1)$$

と表記される確率変数 X を考える。

私たちが母数を知らないとして、この集団からランダムに標本を抽出し、以下の仮説を検証する。

- 帰無仮説： $\mu = 0$
- 対立仮説： $\mu \neq 0$

有意水準 5% に設定し、検定を行う。実際には帰無仮説が正しいので、帰無仮説を棄却しないことが望ましい。母数を知らずに検定を行った場合、どのような結果が得られるだろうか。

■1 つの標本で検定を行う。

まず、標本を抽出する。標本サイズ $N = 20$ とする。

```
N <- 20
smp_1 <- rnorm(N, mean = 0, sd = 1)
```

この標本の標本平均は、

```
mean(smp_1)
```

```
[1] -0.1756794
```

である。この標本平均は 0 ではない。つまり $\bar{x} \neq 0$ であるが、ここから、 $\mu \neq 0$ と言えるだろうか。

標本平均の分布は自由度 $N - 1$ の t 分布に従うので、 t 分布を利用して検定を行う。

ある標本を利用して、「母平均が 0」という帰無仮説を検証したいときは、`t.test()` という関数を使う。対立仮説が「母平均は 0 ではない」のときは、次のようにする。

```
(res_1 <- t.test(smp_1))
```

One Sample t-test

```
data:  smp_1
t = -1.1779, df = 19, p-value = 0.2534
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.4878514  0.1364926
sample estimates:
mean of x
-0.1756794
```

この結果の読み方を説明する。まず、`data: smp_1` というところに、検定に利用した標本が示される。次の行を見ると、検定統計量 T の値が `t` で示されたあと、自由度(df)と p 値(p-value)が表示される。検定統計量の絶対値 $|T|$ が、自ら決めた有意水準における臨界値の絶対値 $|c|$ より大きいとき、帰無仮説を棄却する。検定統計量自体は、次のように取り出せる。

```
res_1$statistic
```

```
t
-1.17788
```

また、自由度 $N - 1$ の t 分布で、有意水準が 5% のときに必要な臨界値は、

```
qt(p = 0.05 / 2, df = N - 1)
```

```
[1] -2.093024
```

である。この 2 つの値の絶対値同士を比べると、

$$|T| < |c|$$

だから、帰無仮説は棄却されない。つまり、 $\mu = 0$ を否定する証拠はない。

実際、 $\mu = 0$ ということを知っているのです、これは妥当な結論である。

■仮説検定のシミュレーション

まず、標本抽出と仮説検定を 1 回行い、帰無仮説(null hypothesis)を棄却しないときは **null** を、帰無仮説を棄却して対立仮説(alternative hypothesis)を採用するときは **alt** を返す関数を作る。有意水準(level)と標本サイズは自分で設定できるようにしておく。

```
smp_test <- function(N = 10, level = 0.05) {  
  ## 標本を抽出し、「母平均 = 0」という仮説を検定する関数  
  ## 引数: N = 標本サイズ  
  ##       level = 有意水準  
  smp <- rnorm(N, mean = 0, sd = 1)  
  test <- t.test(smp)  
  judge <- abs(test$statistic) > abs(qt(p = level / 2, df = N - 1))  
  res <- ifelse(judge, 'alt', 'null')  
  return(res)  
}
```

試しにこの関数を使ってみると、

```
smp_test(N = 20, level = 0.05)
```

```
t  
"null"
```

というように、1 つの標本からどちらの仮説を採用したかがわかる。

標本サイズ $N = 20$ 、有意水準 5% で、この作業を 10,000 回繰り返してみよう。

```
n_sims <- 1e4  
sim_1 <- replicate(n_sims, smp_test(N = 20, level = 0.05))  
table(sim_1) / n_sims
```

```
sim_1  
  alt  null  
0.048 0.952
```

対立仮説が約 5%、帰無仮説が 95% の検定で採用されている。

上で $\mu = 0$ と設定しているのです、帰無仮説こそが正しい仮説である。しかし、帰無仮説が正しくても、誤って帰無仮説を棄却し、対立仮説を採用してしまうことがある。この誤りを「第 1 種の誤り (Type I error)」と呼ぶ。第 1 種の誤りの確率は、有意水準に一致する。私たちは有意水準 5% を選んだ。

危険率、検出力、標本サイズ : 対立仮説が正しい場合の仮説検定シミュレーション

平均 2、標準偏差 2 の正規分布に従う変数 Y について考える。

$$Y \sim \text{Normal}(\mu = 2, \sigma = 2)$$

- 帰無仮説: $\mu = 0$
- 対立仮説: $\mu \neq 0$

有意水準を 5% に設定し、検定を行う。対立仮説の方が正しいので、帰無仮説を棄却したい。

■ 1 つの標本で検定を行う

まず、標本を抽出する。標本サイズ $N = 20$ にする。

```
N <- 20
smp_2 <- rnorm(N, mean = 2, sd = 2)
mean(smp_2)
```

```
[1] 2.229284
```

```
(res_2 <- t.test(smp_2))
```

One Sample t-test

```
data:  smp_2
t = 4.7719, df = 19, p-value = 0.0001326
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 1.251486 3.207082
sample estimates:
mean of x
 2.229284
```

```
res_2$statistic
```

```
      t
4.771889
```

```
qt(p = 0.05 / 2, df = N - 1)
```

```
[1] -2.093024
```

検定統計量 T と臨界値 c の絶対値を比べると、

$$|T| > |c|$$

だから、帰無仮説は棄却される。

■仮説検定のシミュレーション

まず、標本抽出と仮説検定を 1 回行い、帰無仮説(null hypothesis)を棄却しないときは **null** を、帰無仮説を棄却して対立仮説(alternative hypothesis)を採用するときは **alt** を返す関数を作る。先程作った関数を改良し、母平均と母標準偏差も自分で設定できるようにする。

```
smp_test_2 <- function(N = 10, level = 0.05, mu = 0, sigma = 1) {  
  ## 標本を抽出し、「母平均 = 0」という仮説を検定する関数  
  ## 引数: N = 標本サイズ  
  ##      level = 有意水準  
  ##      mu = 母平均  
  ##      sigma = 母標準偏差  
  smp <- rnorm(N, mean = mu, sd = sigma)  
  test <- t.test(smp)  
  judge <- abs(test$statistic) > abs(qt(p = level / 2, df = N - 1))  
  res <- ifelse(judge, 'alt', 'null')  
  return(res)  
}
```

試しに使ってみる。

```
smp_test_2(N = 20, level = 0.05, mu = 2, sigma = 2)
```

```
      t  
"alt"
```

標本サイズ $N = 20$ 、有意水準 5% で、この作業を 10,000 回繰り返してみる。

```
n_sims <- 1e4  
sim_11 <- replicate(n_sims, smp_test_2(N = 20, level = 0.05,
```

```
mu = 2, sigma = 2))
table(sim_11) / n_sims
```

```
sim_11
  alt null
0.988 0.012
```

私たちは、上で $\mu = 2$ と設定しているので、対立仮説こそが正しい仮説である。しかし、対立仮説が正しくても、帰無仮説を棄却せず、対立仮説を採用できない場合がある。この誤りのことを「第 2 種の誤り (Type II error)」という。

第 2 種の誤りの確率は、有意水準と負の相関関係を持つ。第 2 種の誤りを小さくするには、有意水準を大きくすれば良い。しかし、優位種順は、それ自体が第 1 種の誤りの確率を表しているので、有意水準を大きくするということは、第 1 種の誤りの確率が大きくなるということ。

有意水準を変えずに、第 2 種の誤りの確率を小さくできるか。第 2 種の誤りを小さくするには、「0 との違い」を見出す力を強めたらいい。つまり、小さな違いでも違いとして見つけ出すことができれば、帰無仮説を棄却することができるようになり、第 2 種の誤りの確率は小さくなる。これを**検出力 (power)** という。検出力は標本サイズに依存する。

有意水準を 5% にして、標本サイズを 10, 20, 50, 100, 200 と変えて、第 2 種の誤りの割合を計算してみる。

$N = 10$ のとき。

```
sim_14 <- replicate(n_sims, smp_test_2(N = 10, level = 0.05, mu = 2, sigma = 2))
table(sim_14) / n_sims
```

```
sim_14
  alt null
0.8107 0.1893
```

$N = 20$ のとき。

```
sim_15 <- replicate(n_sims, smp_test_2(N = 20, level = 0.05, mu = 2, sigma = 2))
table(sim_15) / n_sims
```

```
sim_15
  alt null
0.9908 0.0092
```

$N = 50$ のとき。

```
sim_15 <- replicate(n_sims, smp_test_2(N = 50, level = 0.05, mu = 2, sigma = 2))
table(sim_15) / n_sims
```

```
sim_15
alt
1
```

$N = 100$ のとき。

```
sim_16 <- replicate(n_sims, smp_test_2(N = 100, level = 0.05, mu = 2, sigma = 2))
table(sim_16) / n_sims
```

```
sim_16
alt
1
```

$N = 200$ のとき。

```
sim_17 <- replicate(n_sims, smp_test_2(N = 200, level = 0.05, mu = 2, sigma = 2))
table(sim_17) / n_sims
```

```
sim_17
alt
1
```

このように、標本サイズを大きくすると、第 2 種の誤りの確率は下がる。したがって、統計的検定における誤りの確率を減らすためには、有意水準を小さくし、標本サイズを大きくすることが重要に成る(しかし、このことは、標本サイズさえ大きくすれば、些細な違いも検出されてしまうことも意味しているので注意が必要)。

平均値の差の検定

2 つのグループの平均値に差があるかどうか、統計的検定で判断しよう。これまで同様、`t.test()` を使って t 検定を行う。

対応のないデータの平均値の差の検定

例題：2 つのコーヒーチェーンで、どちらのコーヒーが美味しいか調べるため、無作為に選んだ 10 人に D 店のコーヒーを、無作為に選んだ別の 10 人に S 店のコーヒーを飲んでもらった。10 点満点で点数をつけてもらったところ、次のようなデータが得られた。

```
d <- c(8, 7, 8, 6, 4, 8, 9, 10, 7, 7)
s <- c(6, 10, 3, 10, 4, 4, 5, 7, 2, 6)
```

D 店のコーヒーを飲んだ 10 人と、S 店のコーヒーを飲んだ 10 人は異なるので、これは対応のないデータであると考えられる。

2 つの店のコーヒーの味に対する評価の平均値は、

```
mean(d)
```

```
[1] 7.4
```

```
mean(s)
```

```
[1] 5.7
```

であり、単純に比較すれば、D 店のコーヒーの方が美味しいということになりそうである。しかし、この結果は、1 つの標本から得られた結果に過ぎない。言い換えると、今回抽出された人たちは、たまたま D 店の味を好んだだけかもしれない。そこで、統計的検定の方法を使って、2 店の美味しさに違いがあるかどうか判断しよう。

各コーヒー店のコーヒーの味の評価の母数(パラメタ、真の値)をそれぞれ D と S とすると、今回検証する仮説は、

- 帰無仮説: $D = S$ (つまり、 $D - S = 0$)
- 対立仮説: $D \neq S$ (つまり、 $D - S \neq 0$)

ということになる。これを有意水準 5% で検定する。また、2 つのグループの分散は異なると仮定する。

以下のコマンドで Welch の t 検定を実行する。

```
(eg_1 <- t.test(d, s, var.equal = FALSE))
```

Welch Two Sample t-test

data: d and s

t = 1.6953, df = 14.848, p-value = 0.1109

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.4392736 3.8392736

sample estimates:

mean of x mean of y

7.4

5.7

検定統計量は、1.6952974 である。また、今検定に必要な t 分布の自由度(df)は、14.8476605 である。この自由度を使って検定に使う臨界値を求めると、

```
qt(p = 0.05 / 2, df = eg_1$parameter)
```

```
[1] -2.133356
```

である。 $|T| < |c|$ だから、帰無仮説は棄却されない。したがって D 店のコーヒーの方が美味しいとは言えない。

ちなみに、`t.test()` の結果として表示される 95% 信頼区間から、検定結果を出すこともできる。この例では、D と S の差の 95% 信頼区間が表示されており、その区間が 0 (つまり、2 つの美味しさに差がない) を含んでいるので有意水準 5% で帰無仮説を棄却できない。

対応のあるデータの平均値の差の検定

例題：2 つのコーヒーチェーンで、どちらのコーヒーが美味しいか調べるため、無作為に選んだ 10 人に D 店のコーヒーと S 店のコーヒーを 1 杯ずつ飲んでもらった。それぞれのコーヒーに 10 点満点で点数をつけてもらったところ、次のようなデータが得られた。

```
survey <- tibble(
  id = 1 : 10,
  d = c(8, 7, 8, 6, 4, 8, 9, 10, 7, 7),
  s = c(6, 10, 3, 10, 4, 4, 5, 7, 2, 6)
)
survey
```

```
# A tibble: 10 x 3
```

	id	d	s
	<int>	<dbl>	<dbl>
1	1	8	6
2	2	7	10
3	3	8	3
4	4	6	10
5	5	4	4
6	6	8	4
7	7	9	5
8	8	10	7
9	9	7	2
10	10	7	6

同じ 10 人が D 店のコーヒーと S 店のコーヒーの両方を飲んで評価しているので、これは対応のあるデータであると考えられる。

```
mean(survey$d)
```

```
[1] 7.4
```

```
mean(survey$s)
```

```
[1] 5.7
```

であり、単純に比較すれば、D 店のコーヒーの方が美味しいということになりそうである。しかし、この結果は、1 つの標本から得られた結果に過ぎない。言い換えると、今回抽出された人たちは、たまたま D 店の味を好んだだけかもしれない。そこで、統計的検定の方法を使って、2 店の美味しさに違いがあるかどうか判断しよう。

各コーヒー店のコーヒーの味の評価の母数(パラメタ、真の値)をそれぞれ D と S とすると、今回検証する仮説は、

- 帰無仮説: $D = S$ (つまり、 $D - S = 0$)
- 対立仮説: $D \neq S$ (つまり、 $D - S \neq 0$)

ということになる。これを有意水準 5% で検証する。対応のあるデータであることを伝えるため、`paired = TRUE` を指定する。

```
(eg_2 <- t.test(survey$d, survey$s, paired = TRUE))
```

Paired t-test

data: survey\$d and survey\$s

t = 1.6805, df = 9, p-value = 0.1272

alternative hypothesis: true mean difference is not equal to 0

95 percent confidence interval:

-0.5883968 3.9883968

sample estimates:

mean difference

1.7

検定統計量は、1.6805071 である。また、この検定に必要な t 分布の自由度(df)は 9 である。この自由度を使って検定に使う臨界値を求めると、

```
qt(p = 0.05 / 2, df = eg_2$parameter)
```

```
[1] -2.262157
```

である。 $|T| < |c|$ だから、帰無仮説は棄却されない。したがって、D 店のコーヒーの方が美味しいとは言えない。

95% 信頼区間を見ると、0（美味しさに差がない）を含んでいるので、帰無仮説は棄却されない。