5. 統計的推定と仮説検定の基礎

honocat

2025-12-18

例題：表が出る確率が 0.5 のコインを N 回投げたところ、10 回表が出た。コインを投げた回数 N はいくつか？

統計的仮説検定の基礎

上で出された例題について、以下の 2 つの仮説を考える。

- 仮説 1：$N = 16$
- 仮説 2：$N = 36$

仮説 1（$N = 16$）が正しいと仮定する

仮説 1 が正しいとすると、1 回のコイン投げ実験は `rbiom(n = 1, size = 16, prob = 0.5)` で表されるはず。例題では、たまたま 10 回表が出たということになる。

```
rbinom(n = 1, size = 16, prob = 0.5)
```

```
[1] 12
```

表は何回出ただろうか？

結果を `result1` に保存しよう。

```
result1 <- rbinom(n = 10000, size = 16, prob = 0.5)
```

まず、表が出る回数の平均値(mean)はいくらか。

```
mean(result1)
```

```
[1] 7.9793
```

理論値である $16 \cdot 0.5 = 8$ に近い値である。分散（variance）と標準偏差（standard deviation）はいくつだろうか。

```
var(result1)
```

```
[1] 3.972469
```

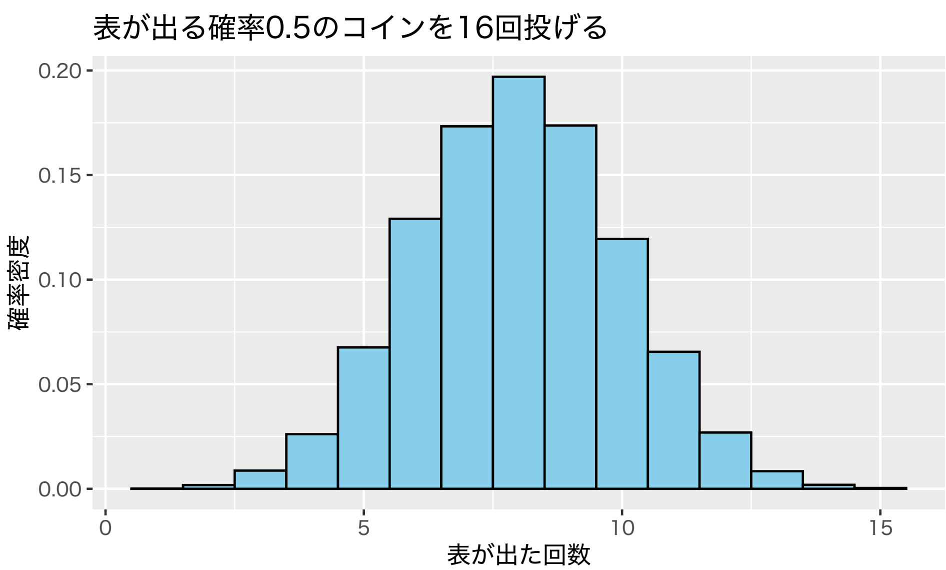
```
sd(result1)
```

```
[1] 1.993105
```

分散の理論値は $16 \cdot 0.5(1-0.5) = 4$、標準偏差の理論値は $\sqrt{4} = 2$ なので、どちらもほぼ理論値通り。

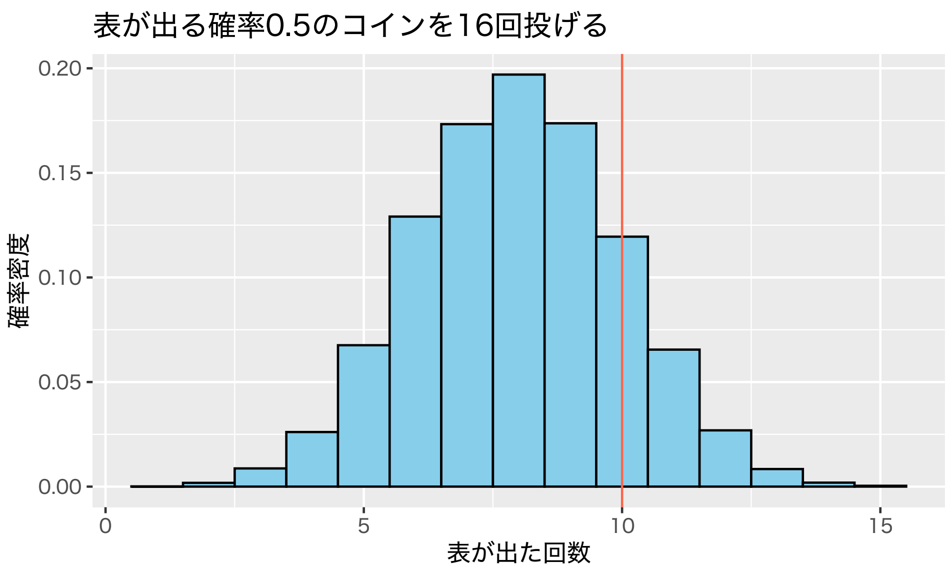
結果をヒストグラムにしてみる。

```
df1 <- tibble(x = result1)
hist1 <- ggplot(df1, aes(x = x, y = after_stat(density))) +
  geom_histogram(color = 'black',
                 fill  = 'skyblue',
                 binwidth = 1) +
  labs(x = '表が出た回数',
       y = '確率密度',
       title = '表が出る確率 0.5 のコインを 16 回投げる')
plot(hist1)
```



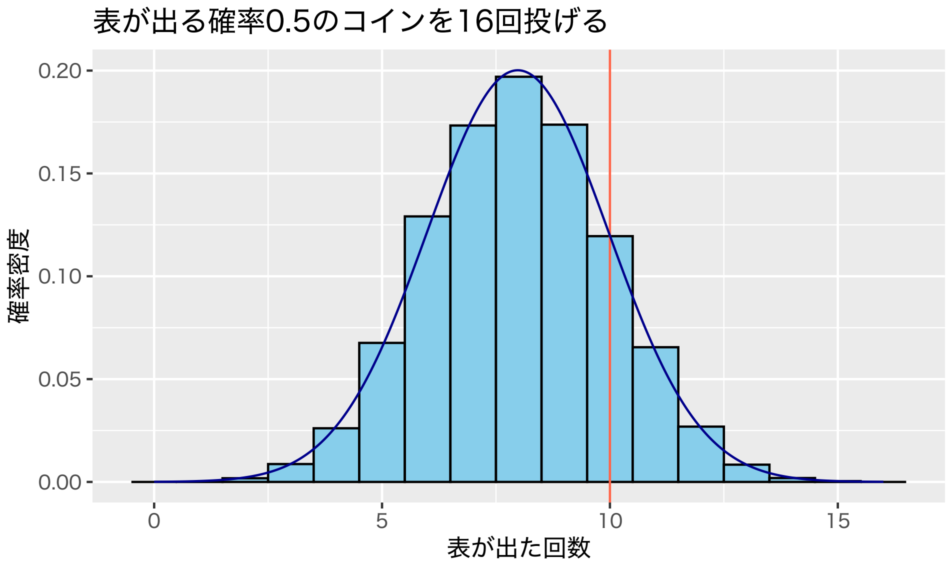
例題で得られた「表が 10 回」という結果を図示してみる。

```
hist2 <- hist1 +
  geom_vline(xintercept = 10,
             color      = 'tomato')
plot(hist2)
```



このヒストグラムを見ると、なんとなく正規分布に見える。先程計算した平均値と標準偏差を利用して、平均 7.9793、標準偏差 1.9931053 の正規分布を重ねてみる。

```
df2 <- tibble(x = seq(from = 0, 16, length.out = 1000)) |>
  mutate(y = dnorm(x, mean = mean(result1), sd = sd(result1)))
hist3 <- hist2 +
  geom_line(data = df2, aes(x = x, y = y),
            color = 'darkblue')
plot(hist3)
```



この図からわかるように、二項分布から得られたヒストグラムの形は正規分布によく似ている(こういう状況を「正規分布で近似できる」という)。

正規分布の特徴から、平均 ± 2 標準偏差の間にデータの約 95% があるはず。したがって仮説 1 ($N = 16$)が正しいとすれば、表が出る回数の 95% は、

```
mean(result1) - 2 * sd(result1)
```

```
[1] 3.993089
```

と

```
mean(result1) + 2 * sd(result1)
```

```
[1] 11.96551
```

の間にあるはず。

念の為、正規分布に頼らずに私たちの実験結果からも同じことを確かめる。私たちは実験を 10,000 回繰り返した。小さい順に並べる。

```
result1_sorted <- sort(result1)
```

この並べ替えたデータのうち、小さい方から 2.5% と大きい方から 2.5% を取り除くと、平均に近い 95% のデータを残すことができる。

```r
result1_sorted[251]
```

```
[1] 4
```

```r
result1_sorted[9750]
```

```
[1] 12
```

つまりデータの 95% は、4 と 12 の間にある。これは正規分布を利用して求めた値とほぼ同じである。

```r
quantile(result1, probs = c(0.025, 0.975))
```

```
 2.5% 97.5%
    4    12
```

いずれの方法を使っても、仮にコイン投げの回数が 16 回(仮説 1)だとすれば、例題の観測値として得られた「10 回」という回数は、平均周りの 95% の範囲に含まれている。つまり、「16 回コインを投げて 10 回表が出る」という現象は、特に珍しい訳では無い。よって、$N = 16$ という**仮説を否定するような証拠はない**と考えられる。

仮説 2 ($N = 36$)が正しいと仮定する

仮説 2 が正しいとすると、

```r
rbinom(n = 1, size = 36, prob = 0.5)
```

```
[1] 10
```

である。表は何回出ただろうか？

先ほどと同様に、`result2` に保存しよう。

```r
result2 <- rbinom(n = 10000, size = 36, prob = 0.5)
```

平均値は、

```r
mean(result2)
```

```
[1] 17.9909
```

分散と標準偏差は、

```r
var(result2)
```

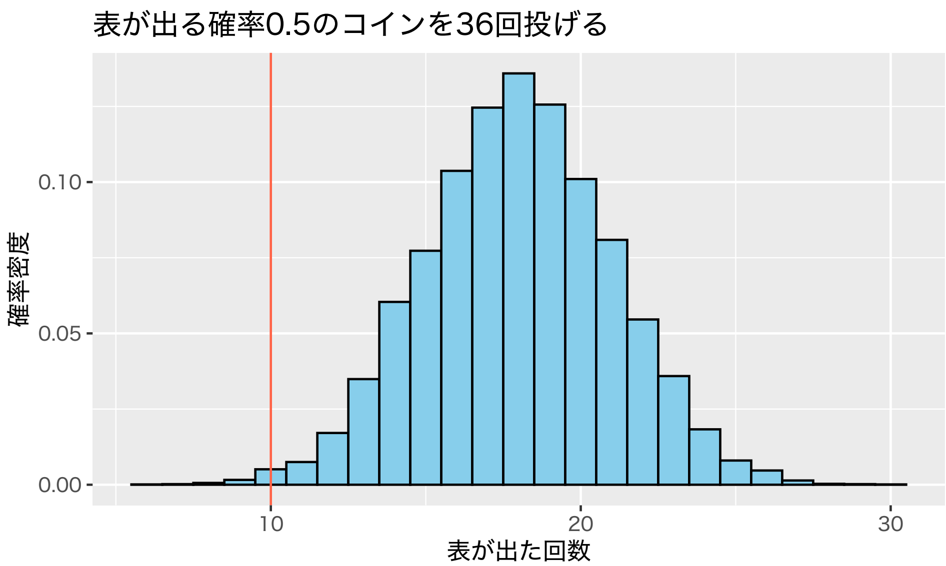
```
[1] 9.174935
```

```r
sd(result2)
```

```
[1] 3.029015
```

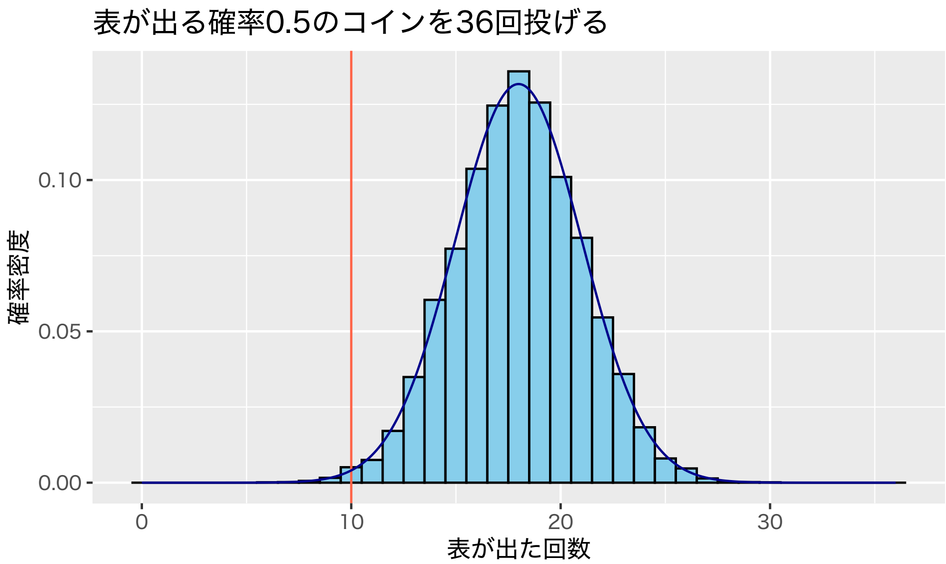
結果をヒストグラムにしてみる。

```r
df3 <- tibble(x = result2)
hist4 <- ggplot(df3, aes(x = x, y = after_stat(density))) +
  geom_histogram(color = 'black',
                 fill   = 'skyblue',
                 binwidth = 1) +
  geom_vline(xintercept = 10,
             color      = 'tomato') +
  labs(x = '表が出た回数',
       y = '確率密度',
       title = '表が出る確率 0.5 のコインを 36 回投げる')
plot(hist4)
```



正規分布で近似できそうなので、重ねて描いてみる。

```
df4 <- tibble(x = seq(from = 0, to = 36, length.out = 1000)) |>
  mutate(y = dnorm(x, mean = mean(result2), sd = sd(result2)))
hist5 <- hist4 + geom_line(data = df4, aes(x = x, y = y), color = 'darkblue')
plot(hist5)
```



正規分布で近似できそうなので、正規分布の特徴が当てはまる。したがって、仮説 2 が正しいとすれば、表が出る回数の 95% は、

```
mean(result2) - 2 * sd(result2)
```

```
[1] 11.93287
```

と

```
mean(result2) + 2 * sd(result2)
```

```
[1] 24.04893
```

の間にあるはず。

念の為、実験結果からも同じことを確かめる。

```
quantile(result2, probs = c(0.025, 0.975))
```

```
 2.5% 97.5% 
   12    24 
```

いずれの方法を使っても、仮にコイン投げの回数が 36 回(仮説 2)だとすれば、例題の観測値として得られた「10 回」という回数は、平均周りの 95% に含まれない。つまり、「36 回コインを投げて 10 回表が出る」という現象は、とても珍しい現象であり、あまり起きることは期待されない。これは、$N = 36$ という**仮説を否定するような証拠**と考えられるので、$N = 36$ という**仮説は棄却**(reject)する。

統計的推定の基礎

仮説検定の方法を使えば、1 つひとつの仮説について、その説がおかしいと言える証拠があるかどうか確かめることができる。しかし、理論的に可能な仮説はたくさんあり、すべてを確かめるのは面倒だ。

そこで、そもそも N はいくつなのかということ自体を直接考えるのが、統計的推定(statistical estimation)だ。

点推定(point estimation)

二項分布では平均値が最も起こりやすいと考えられる。そうすると、表が出る確率 0.5 のコインを N 回投げて 10 回表が出るのは、

$$N \cdot 0.5 = 10$$

この式を解いて、$N = 20$ となる。つまり「試行回数 20 で成功確率 0.5 の二項分布の平均値が 10 である」ということがわかる。この平均値が私たちのデータと合致するので、私たちは「$N = 20$」と推定する。このようにして得られた、1 つの数をピンポイントで示す推定法を点推定(point estimation)といい、得られた値を点推定値(point estimate)という。

区間推定(interval estimation)

上で 1 つの値を推定したが、その推定にどのくらい自信があるか(より正確には、どれくらい自信がないか)は問題によって違う。推定値が同じ 10 でも、「絶対 10 だ」「10 かもしれない」というので答えの意味は異なる。そこで、この「推定に対する自信のなさ(uncertainty)」を推定の方法に取り込んだのが、**区間推定(interval estimation)**である。

より具体的には、統計的仮説検定で棄却されない仮説の集合を、区間推定に使う。私たちは、$N = 36$ という仮説を棄却した。したがって、$N = 36$ は区間推定には使わない。他方、$N = 16$ という仮説は棄却しなかったので、$N = 16$ は区間推定の一部として使う。$N = 36$ が仮説として妥当でないことがわかっているので、$N = 10, 11, \dots, 35$ について、仮説が棄却できるかどうか確かめれば良い。

```
bin_exp <- function(h, trials = 10000) {
  # h: 仮説(hypothesis)、N = h を検証する
  # trials: 実験の繰り返し回数、既定値は 10,000

  res <- rbinom(n = trials, size = h, prob = 0.5)
```

```
  return(quantile(res, prob = c(0.025, 0.975)))
}
```

これを使ってみる。

```
bin_exp(16)
```

```
 2.5% 97.5% 
    4    12 
```

先ほどとほぼ同じ結果が得られた。データである 10 はこの 2 つの数値の間にあるので、仮説 1 は保留する。

同様に、仮説 2（$N = 36$）は、

```
bin_exp(36)
```

```
 2.5% 97.5% 
   12    24 
```

となり、棄却される。

1 つ 1 つの値を順番に検証してもいいが、一気に実行するには次のようにする。

```
set.seed(2025-12-18)
x <- 10 : 35
names(x) <- 10 : 35
sapply(x, bin_exp)
```

```
      10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26     27 28 29 30 31 32
2.5%   2  2  3  3  3  4  4  4  5  5  6  6  6  7  7  8  8  8.975  9  9 10 10 11
97.5%  8  9  9 10 11 11 12 12 13 14 14 15 16 16 17 17 18 19.000 19 20 20 21 22
      33 34 35
2.5%  11 11 12
97.5% 22 23 23
```

for ループを使ってもいい。

```
set.seed(2025-12-18)
for (x in 10:35) {
  res <- bin_exp(x)
  cat(x, ':', res, '\n')
```

```
}
```

```
10 : 2 8
11 : 2 9
12 : 3 9
13 : 3 10
14 : 3 11
15 : 4 11
16 : 4 12
17 : 4 12
18 : 5 13
19 : 5 14
20 : 6 14
21 : 6 15
22 : 6 16
23 : 7 16
24 : 7 17
25 : 8 17
26 : 8 18
27 : 8.975 19
28 : 9 19
29 : 9 20
30 : 10 20
31 : 10 21
32 : 11 22
33 : 11 22
34 : 11 23
35 : 12 23
```

N が 10 から 12 の間と、32 以上の場合にはデータである 10 が 2 つの数値の間にないことがわかるので、これらの仮説は棄却する。よって、「N は 13 以上 31 以下の整数である」と答えるのが区間推定である。