

第 7 章: 不確実性

今井耕介 著

『社会科学のためのデータ分析入門 (QSS)』

2026-03-09

7.1 推定と不確実性

STAR データの読み込み (1)

- ▶ 推定の不確実性を評価するため、STAR プロジェクトのデータを読み込みます。

```
# 1. ローカルに保存したデータの読み込み (推奨)
```

```
star <- read.csv("STAR.csv")
```

```
# (参考) URL から直接読み込むことも可能
```

```
# star <- read.csv("https://ayumu-tanaka.github.io/QSS/QSS_Data/STAR.csv")
```

標準誤差 (Standard Error) (2)

▶ 推定値のばらつき (標準偏差) を標準誤差と呼びます。

```
# 小人数学級 (1) の読解力スコアの標準誤差を計算
```

```
n.small <- sum(star$classtype == 1 & !is.na(star$g4reading))
```

```
est.small <- mean(star$g4reading[star$classtype == 1], na.rm = TRUE)
```

```
se.small <- sd(star$g4reading[star$classtype == 1], na.rm = TRUE) / sqrt(n.small)
```

```
est.small
```

```
## [1] 723.3912
```

```
se.small
```

```
## [1] 1.913012
```

7.2 信頼区間 (Confidence Intervals)

95%信頼区間の算出

- ▶ 母集団の真の値が一定の確率 (例: 95%) で含まれると期待される範囲。
- ▶ 公式: 推定値 $\pm 1.96 \times$ 標準誤差

```
# 95% 信頼区間 (小人数学級の平均)
```

```
ci.small <- c(est.small - 1.96 * se.small, est.small + 1.96 * se.small)
ci.small
```

```
## [1] 719.6417 727.1407
```

```
# 平均の差 (ATE) の信頼区間
```

```
n.reg <- sum(star$classtype == 2 & !is.na(star$g4reading))
est.reg <- mean(star$g4reading[star$classtype == 2], na.rm = TRUE)
se.reg <- sd(star$g4reading[star$classtype == 2], na.rm = TRUE) / sqrt(n.reg)
ate.est <- est.small - est.reg
ate.se <- sqrt(se.small^2 + se.reg^2)
c(ate.est - 1.96 * ate.se, ate.est + 1.96 * ate.se)
```

```
## [1] -1.699599 8.702064
```

7.3 仮説検定 (Hypothesis Testing)

p 値 (p-value)

- ▶ 帰無仮説 (H_0): 「効果がない (差はゼロである)」という仮定。
- ▶ p 値: H_0 が正しい場合に、観測された結果 (またはそれ以上に極端な結果) が得られる確率。
- ▶ p 値 < 0.05 であれば、「統計的に有意」と判断し、帰無仮説を棄却する。

```
# 平均の差の検定 (Z 検定)
```

```
z.score <- ate.est / ate.se
```

```
2 * pnorm(-abs(z.score)) # 2 面 p 値
```

```
## [1] 0.1870072
```

```
# t.test 関数を用いた検定
```

```
t.test(star$g4reading[star$classtype == 1],  
       star$g4reading[star$classtype == 2])
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: star$g4reading[star$classtype == 1] and star$g4reading[star$classtype == 2]  
## t = 1.3195, df = 1541.2, p-value = 0.1872
```

7.4 回帰分析における不確実性

resume データの読み込み (1)

- ▶ 回帰係数の有意性を確認するため、履歴書実験データを読み込みます。

```
# 1. ローカルに保存したデータの読み込み (推奨)
```

```
resume <- read.csv("resume.csv")
```

```
# (参考) URL から直接読み込むことも可能
```

```
# resume <- read.csv("https://ayumu-tanaka.github.io/QSS/QSS_Data/resume.csv")
```

回帰係数の有意性 (2)

▶ `summary()` で各係数の標準誤差、t 値、p 値を確認できます。

```
# 人種とコールバック率の回帰
fit <- lm(call ~ race, data = resume)
summary(fit)$coefficients
```

##	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	0.06447639	0.005504805	11.712747	2.871674e-31
## racewhite	0.03203285	0.007784969	4.114705	3.940803e-05

選挙データの読み込みと誤差計算 (1)

- ▶ 予測の不確実性を視覚化するため、選挙結果と世論調査データを読み込みます。

```
# 1. ローカルに保存したデータの読み込み (推奨)
```

```
pres08 <- read.csv("pres08.csv")
```

```
polls08 <- read.csv("polls08.csv")
```

```
# (参考) URL から直接読み込むことも可能
```

```
# pres08 <- read.csv("https://ayumu-tanaka.github.io/QSS/QSS_Data/pres08.csv")
```

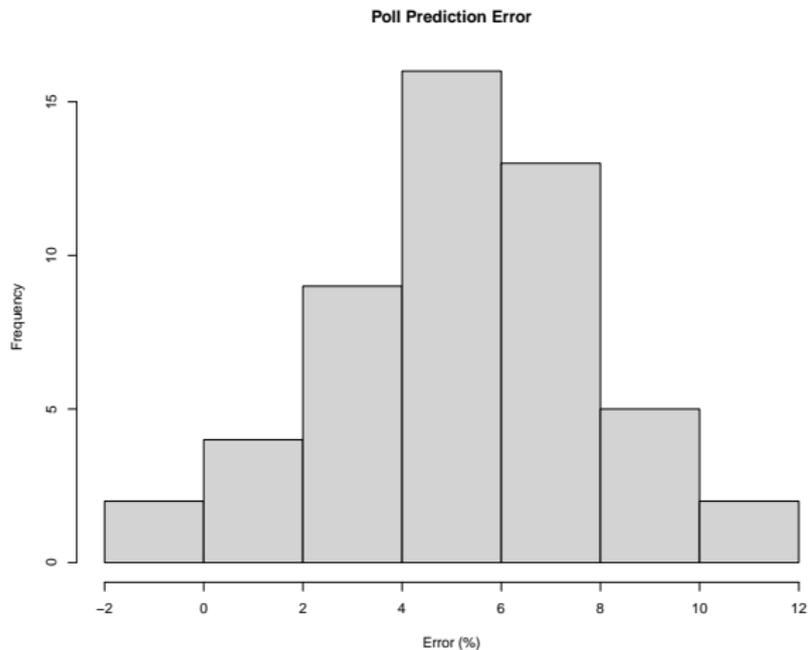
```
# polls08 <- read.csv("https://ayumu-tanaka.github.io/QSS/QSS_Data/polls08.csv")
```

```
# 2. 予測誤差の計算
```

```
errors <- pres08$Obama -
```

```
  apply(polls08$Obama, polls08$state, mean)[pres08$state]
```

信頼区間のプロット (2)



7.5 まとめ

第 7 章のまとめ

- ▶ **不確実性:** 標本から母集団を推測する際には、常に誤差が伴う。
- ▶ **標準誤差:** 推定値の精度を測る尺度。
- ▶ **信頼区間:** 真の値が含まれる可能性が高い範囲を提示する。
- ▶ **仮説検定:** 観測された結果が偶然によるものかどうかを確率的に判断する。
- ▶ **統計的有意性:** p 値が基準 (0.05 など) を下回るかどうかで判断するが、その解釈には注意が必要。
- ▶ **R の操作:** `sd()`, `qt()`, `qnorm()`, `t.test()`, `confint()` などを習得した。