

## 第 2 章: 因果関係 (2.6. 1 変数の記述統計量)

今井耕介 著

『社会科学のためのデータ分析入門 (QSS)』

2026-03-09

## 2.6 1 変数の記述統計量

## 記述統計の目的

- ▶ 大規模なデータの分布（散らばり具合や中心的な傾向）を、少数の数値で要約する。
- ▶ 主要な指標：
  1. **中心傾向の尺度**: 平均値 (Mean)、中央値 (Median)
  2. **散らばりの尺度**: 分位数 (Quantiles)、四分位範囲 (IQR)、標準偏差 (Standard Deviation)

## 2.6.1 分位数 (Quantiles)

## minwage データの読み込み (1)

▶ 最低賃金データの読み込みと準備を行います。

# 1. ローカルに保存した *minwage* データの読み込み (推奨)

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

# (参考) URL から直接読み込むことも可能

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

# 2. データの準備 (州別の抽出)

```
minwageNJ <- subset(minwage, subset = (location != "PA"))
```

```
minwagePA <- subset(minwage, subset = (location == "PA"))
```

## 中央値 (Median) (2)

- ▶ データを大きさの順に並べたときに、ちょうど真ん中にくる値。
- ▶ 外れ値 (極端に大きい・小さい値) の影響を受けにくい (ロバスト)。

```
# NJ州の引き上げ後のフルタイム雇用率の中央値を計算  
median(minwageNJ$fullAfter / (minwageNJ$fullAfter + minwageNJ$partAfter),  
        na.rm = TRUE)
```

```
## [1] 0.2916667
```

## 4 分位数と summary() 関数

- ▶ **4 分位数**: データを 4 等分する境界の値 (25%, 50%, 75%)。
- ▶ `summary()` を使うと、最小値、最大値、平均、中央値、4 分位数を一括で確認できる。

```
# NJ州の引き上げ前の賃金 (wageBefore) の要約  
summary(minwageNJ$wageBefore)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##      4.25   4.25   4.50   4.61   4.87   5.75
```

```
# 4 分位範囲 (Interquartile Range: IQR)  
# 75%点 と 25%点 の差 (真ん中の 50%のデータの広がり)  
IQR(minwageNJ$wageBefore)
```

```
## [1] 0.62
```

## 任意の分位数の計算 (quantile 関数)

▶ 10%刻み (10 分位数) などで分布を詳しく見る場合に便利。

```
# 0%から 100%まで 10%刻みの分位数を計算
```

```
quantile(minwageNJ$wageBefore, probs = seq(from = 0, to = 1, by = 0.1))
```

```
##      0%   10%   20%   30%   40%   50%   60%   70%   80%   90% 100%  
## 4.25 4.25 4.25 4.25 4.50 4.50 4.65 4.75 5.00 5.00 5.75
```

## 2.6.2 標準偏差 (Standard Deviation)

## 平均からのばらつき

- ▶ **標準偏差**: 各データが平均からどれくらい離れているかの平均的な距離。
- ▶ 値が大きいほど、データが平均の周りに広く散らばっていることを示す。

```
# NJ州の引き上げ前後の雇用率の変化を計算
```

```
minwageNJ$fullPropBefore <- minwageNJ$fullBefore /
```

```
  (minwageNJ$fullBefore + minwageNJ$partBefore)
```

```
minwageNJ$fullPropAfter <- minwageNJ$fullAfter /
```

```
  (minwageNJ$fullAfter + minwageNJ$partAfter)
```

```
minwageNJ$diff <- minwageNJ$fullPropAfter - minwageNJ$fullPropBefore
```

```
# 雇用率の変化の標準偏差を計算
```

```
sd(minwageNJ$diff, na.rm = TRUE)
```

```
## [1] 0.3010377
```

## 2.6.3 まとめ

## このセクションのまとめ

- ▶ **中央値と分位数**: データの順位に基づく要約。外れ値に強い (ロバスト)。
- ▶ **summary()**: データの分布を素早く把握するための最も基本的なコマンド。
- ▶ **標準偏差**: 平均の周りのばらつきを測定する。
- ▶ **R の操作**:
  - ▶ median(), quantile(), IQR() を用いた分位数の計算。
  - ▶ sd(), var() を用いたばらつきの計算。
  - ▶ summary() による一括集計。
  - ▶ na.rm = TRUE オプションによる欠損値の除外。