

第 3 章: 測定
(3.3. 1 変量の分布をビジュアル化する)
今井耕介 著
『社会科学のためのデータ分析入門 (QSS)』

2026-03-09

3.3.1 変量の分布をビジュアル化する

データの可視化の重要性

- ▶ 数値の羅列（平均や標準偏差など）だけでは、データの全体像を把握しきれない。
- ▶ グラフを用いることで、以下を確認できる：
 - ▶ 分布の形状（対称性、偏り）
 - ▶ 外れ値の存在
 - ▶ グループ間の違い
- ▶ 変数の種類（カテゴリ変数か数値変数か）によって、適切なグラフを選択する必要がある。

3.3.1 棒グラフ (Bar Plot)

afghan データの読み込み (1)

- ▶ カテゴリ変数の可視化 (棒グラフ) のため、データを準備します。

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

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

```
# (参考) URL から直接読み込むことも可能
```

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

カテゴリ変数の集計 (2)

- ▶ ISAF による被害経験の比率表を作成します (欠損値 NA も含める)。

```
# ISAF による被害経験の比率表
```

```
isaf.ptable <- prop.table(table(ISAF = afghan$violent.exp.ISAF,  
                                exclude = NULL))
```

```
isaf.ptable
```

```
## ISAF
```

```
##           0           1          <NA>
```

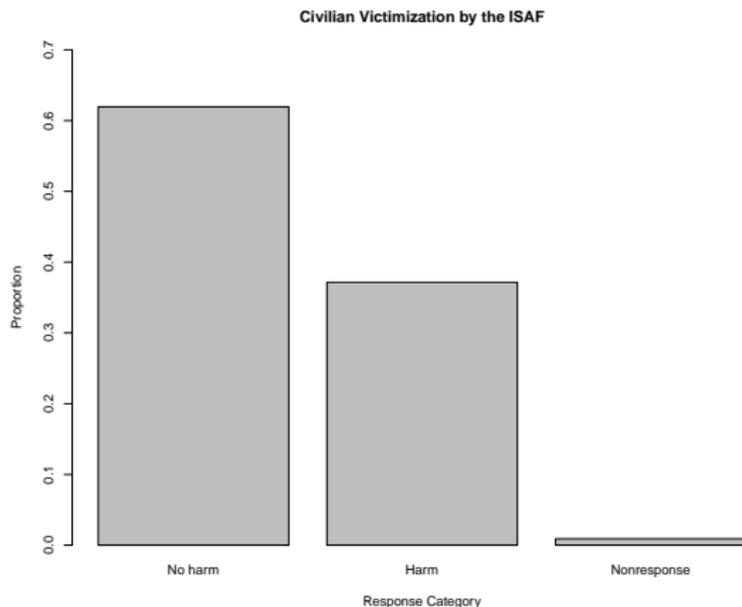
```
## 0.619462600 0.371459695 0.009077705
```

棒グラフの描画: コード

1. `barplot()` 関数で棒グラフを作成

```
barplot(isaf.ptable,  
        names.arg = c("No harm", "Harm", "Nonresponse"), # X 軸のラベル名  
        main = "Civilian Victimization by the ISAF", # タイトル  
        xlab = "Response Category", # X 軸ラベル  
        ylab = "Proportion", # Y 軸ラベル  
        ylim = c(0, 0.7)) # Y 軸の範囲
```

棒グラフの描画: 結果



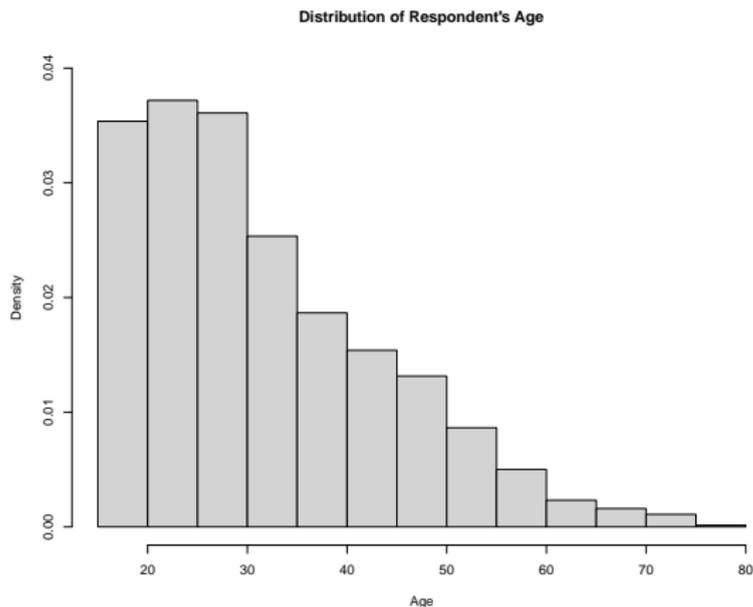
3.3.2 ヒストグラム (Histogram)

数値変数の密度を可視化する

- ▶ **ヒストグラム**: 連続的な数値変数の分布を、「ビン」と呼ばれる階級ごとに集計して表示する。

```
# 1. hist() 関数でヒストグラムを作成
# freq = FALSE とすることで、度数ではなく「密度 (面積の合計が 1)」で表示
hist(afghan$age,
     freq = FALSE,
     ylim = c(0, 0.04),
     xlab = "Age",
     main = "Distribution of Respondent's Age")
```

年齢の分布: 結果



ヒストグラムのカスタマイズ (教育年数)

- ▶ `breaks` 引数でビンの幅を細かく制御し、`abline()` で中央値を重ねます。

1. 教育年数のヒストグラム (1年刻みのビンを設定)

```
hist(afghan$educ.years, freq = FALSE,  
     breaks = seq(from = -0.5, to = 18.5, by = 1),  
     xlab = "Years of Education",  
     main = "Distribution of Education")
```

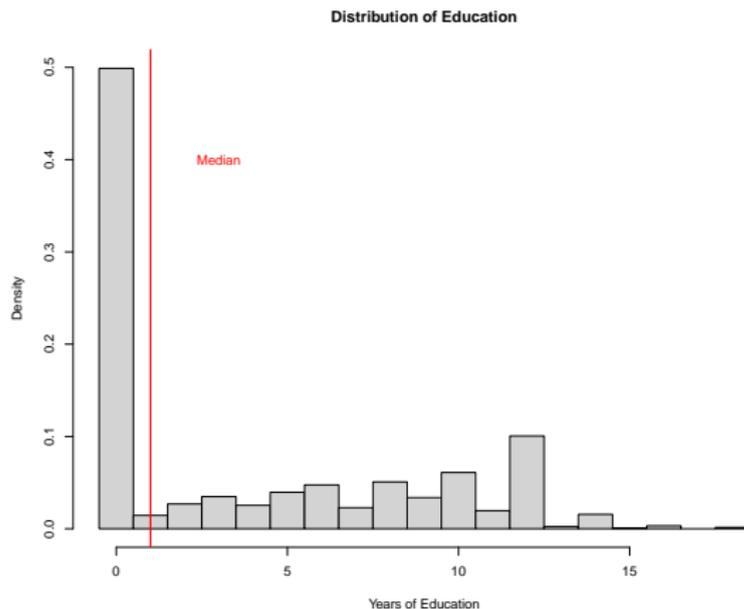
2. 中央値 (*median*) を示す垂直線 (*Vertical line*) を追加

```
abline(v = median(afghan$educ.years), col = "red", lwd = 2)
```

3. テキストラベルを追加

```
text(x = median(afghan$educ.years) + 2, y = 0.4, "Median", col = "red")
```

教育年数の分布: 結果



3.3.3 箱ひげ図 (Box Plot)

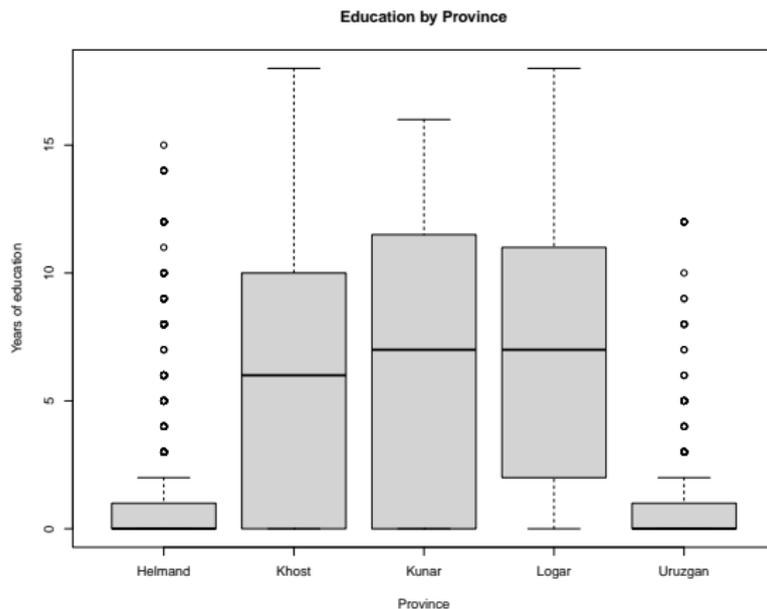
グループ間の分布の比較

- ▶ **箱ひげ図**: 5 数要約 (最小値、第 1 四分位数、中央値、第 3 四分位数、最大値) を可視化する。
- ▶ グループごとの分布の「位置」と「広がり」を比較するのに適している。

```
# 1. boxplot(数値変数 ~ 分類変数) という形式で描画  
# 州 (province) ごとの教育年数 (educ.years) を比較
```

```
boxplot(educ.years ~ province,  
        data = afghan,  
        main = "Education by Province",  
        ylab = "Years of Education",  
        xlab = "Province")
```

州別の教育年数: 結果



3.3.4 まとめ

このセクションのまとめ

- ▶ **棒グラフ**: カテゴリ変数に適しており、比率や度数を比較する。
- ▶ **ヒストグラム**: 数値変数の分布の形 (山がどこにあるかなど)を確認する。
- ▶ **箱ひげ図**: 複数のグループ間で分布を比較する際に非常に強力。
- ▶ **グラフの調整**:
 - ▶ タイトル (`main`) や軸ラベル (`xlab`, `ylab`) を適切につける。
 - ▶ 色や線 (`abline`)、テキスト (`text`) を追加して情報を強調する。
- ▶ **R の操作**: `barplot()`, `hist()`, `boxplot()`, `abline()`, `text()`。