第 3 章: 測定
(3.4. 標本調査)

今井耕介 著

『社会科学のためのデータ分析入門 (QSS)』

2026-03-09

3.4 標本調査 (Survey Sampling)

標本調査の目的

- **母集団** (Population) 全体を調べるのが困難な場合、その一部である**標本** (Sample) を抽出して調査する。
- **無作為抽出** (Random Sampling): 母集団の全ての個体が選ばれる確率を等しくすることで、標本から母集団の特徴を正確に推測（推定）できるようにする。
- 社会科学における課題：
  - 抽出の偏り
  - 無回答バイアス
  - 虚偽の回答

3.4.1 無作為化の役割

afghan-village データの読み込み (1)

- アフガニスタンの村レベルデータ (afghan-village.csv) 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）
afghan.village <- read.csv("afghan-village.csv")

#（参考）URL から直接読み込むことも可能
# afghan.village <- read.csv("https://ayumu-tanaka.github.io/QSS/QSS_Data/afghan
```

データの構造を確認 (2)

- 実際に調査対象となった村 (`village.surveyed == 1`) と、そうでない村を確認します。

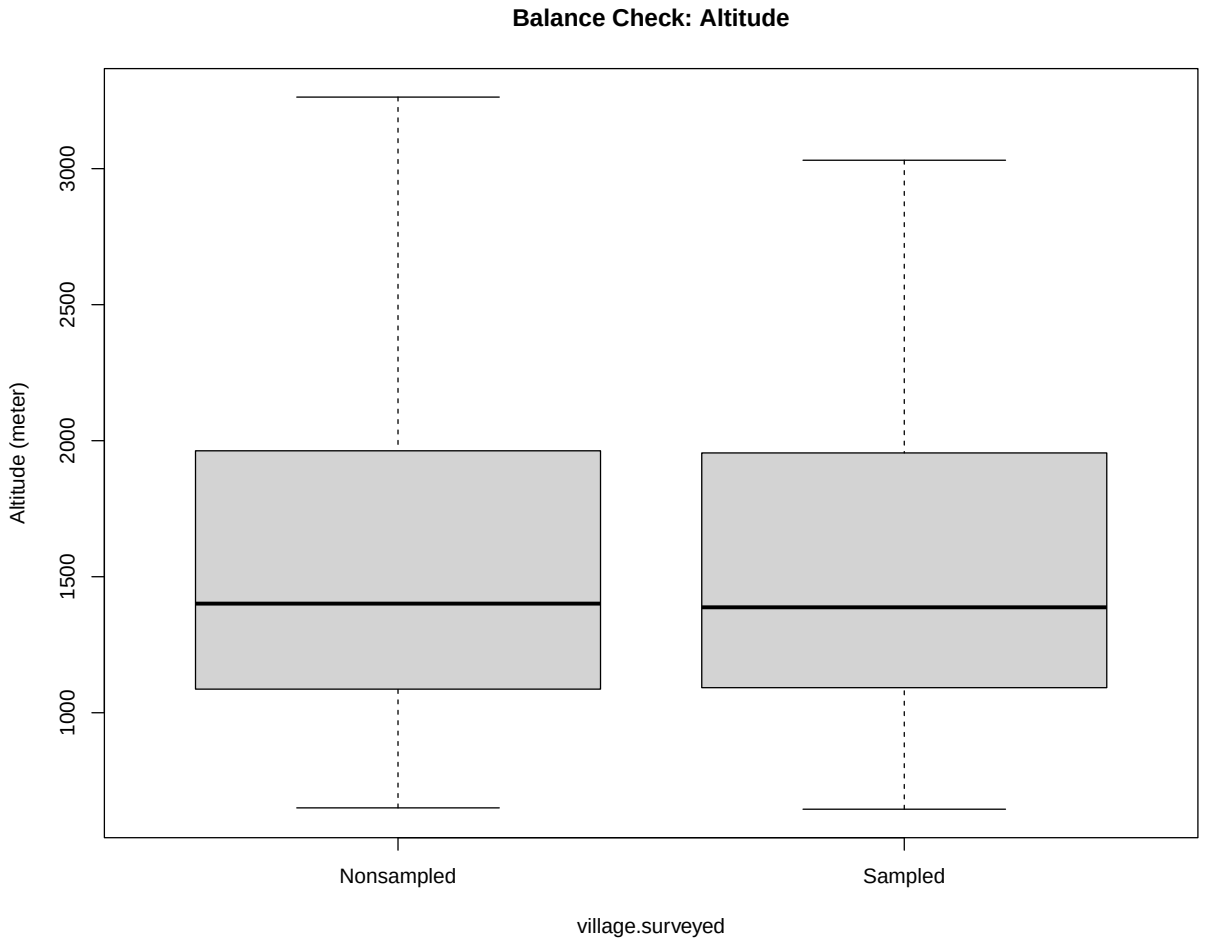
```
# village.surveyed: 調査対象になったか (1=Yes, 0=No)
# altitude: 標高, population: 人口
head(afghan.village, n = 3)
```

```
##   altitude population village.surveyed
## 1  1959.08        197                1
## 2  2425.88        744                0
## 3  2236.60        179                1
```

標高の比較: プロットのコード

```
# boxplot(数値変数 ~ 分類変数)
# 調査対象の有無によって標高 (altitude) の分布を比較
boxplot(altitude ~ village.surveyed,
        data = afghan.village,
        ylab = "Altitude (meter)",
        names = c("Nonsampled", "Sampled"),
        main = "Balance Check: Altitude")
```

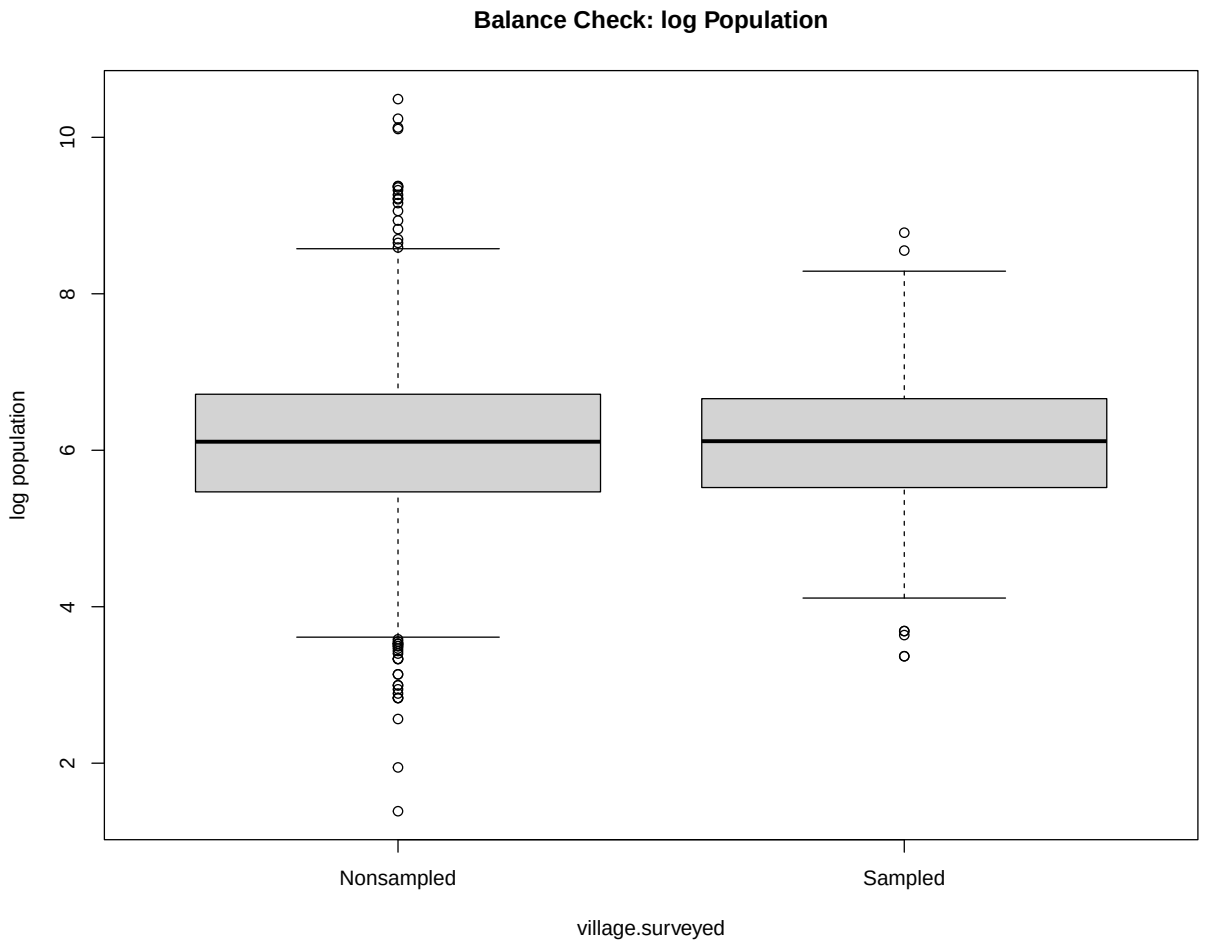
標高の比較: 描画結果


Balance Check: Altitude
Altitude (meter)
1000
1500
2000
2500
3000
Nonsampled
Sampled
village.surveyed

人口の比較（対数）：プロットのコード

```
# 人口 (population) は分布が偏っているため、対数 (log) をとって比較
boxplot(log(population) ~ village.surveyed,
        data = afghan.village,
        ylab = "log population",
        names = c("Nonsampled", "Sampled"),
        main = "Balance Check: log Population")
```

人口の比較（対数）：描画結果



3.4.2 無回答とバイアス

afghan データの読み込み (1)

- 個人の調査データを読み込み、欠損値を確認します。

```
# 1. ローカルに保存したデータの読み込み（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")
```

項目無回答 (Item Nonresponse) (2)

- 特定の質問（例：タリバンによる被害）に答えない人が、特定の地域に偏っていないか確認します。

```
# 州 (province) ごとの「タリバン被害」の欠損値 (NA) の割合を計算
# is.na() は欠損なら TRUE(1), そうでなければ FALSE(0) を返す
tapply(is.na(afghan$violent.exp.taliban), afghan$province, mean)
```

```
##     Helmand       Khost       Kunar       Logar     Uruzgan
## 0.030409357 0.006349206 0.000000000 0.000000000 0.062015504
```

敏感な質問への対処 (List Experiment)

- 直接聞きにくい質問（例：武装勢力への支持）に対し、間接的に答えてもらう手法。
- **リスト実験**:
  - 統制群：無害な項目のリストを渡し、当てはまる「数」だけを答えてもらう。
  - 処置群：無害な項目に「敏感な項目」を 1 つ加えたリストを渡し、同じく「数」を答えてもらう。
  - **差**が、敏感な項目に「はい」と答えた人の割合の推定値になる。

```
# 敏感な項目（ISAF への支持）に関するリスト実験の結果
# list.group: "ISAF"（処置群）または "control"（統制群）
# list.response: 当てはまる項目の回答数
mean(afghan$list.response[afghan$list.group == "ISAF"]) -
    mean(afghan$list.response[afghan$list.group == "control"])
```

```
## [1] 0.04901961
```

3.4.3 まとめ

このセクションのまとめ

- **無作為抽出**: 母集団を代表する標本を得るための強力な方法。
- **バランス確認**: 抽出されたグループとそうでないグループで、背景属性に差がないか（共変量のバランス）を箱ひげ図などで確認することが重要。
- **バイアスへの警戒**:
  - **無回答**: 欠損値がランダムに発生しているか、特定のグループに偏っていないかを確認する。
  - **リスト実験**: 直接質問では得られない「本音」を統計的に推計する工夫。
- **R の操作**: `is.na()` による欠損値の判定、`log()` による対数変換。