

政治学 3（計量政治学）

4 回目

秦 正樹

京都府立大学公共政策学部 准教授

hatamasaki@kpu.ac.jp

2022/05/10

火 3 コース@一号館情報処理室

4 回目のテーマ

① 4 回目のテーマ

- 2 変量の変数から「関係」を把握する
 - 変数と変数の関係から細かく分析するために…
 - クロス表：質的変数*質的変数の関係を分析
 - 相関分析と平均値：量的変数*質的変数の関係を分析
- 複数の変数間の関係について，Rstudio を使いつつ勉強しましょう！

② 従業準備

- Online Rstudio はここからアクセスしてください。
- 上記システムにレジューメをあげてありますのでご確認を。
- 資料と online Rstudio を相互に動かしていくので，前回までのコードラインのご準備よろしくをお願いします

復習：データの種類

① 変数と尺度

- 量的変数（比例尺度）：0 を原点として、間隔と比率に意味があるもの
e.g. 体重：100kg は、50kg に比べて 2 倍重いといえる
- 質的変数（名義尺度）：数値とは無関係に、定義的に区別するためにつけられたもの
e.g. 性別（1. 男性，2. 女性，3. その他）→数字に性別の意味はない
- 量と質を併せ持つ変数（順序尺度）：数字の大小には意味があるが間隔が等価でないもの
e.g. 順位（1 位・2 位・3 位…）．世論調査の多くはこれ．

② 変数から大枠を把握する方法

- 量的変数：記述統計を見ながら、大きな傾向を把握
e.g. 日本人の平均身長，平均年齢（あとでやってみる）と分散など
- 質的変数：平均値に意味はないので，度数分布を見る
e.g. あるデータの都道府県平均が 17.2 でした→意味不明…

質的変数の間の関係を理解する

① 1 変量から 2 変量（そして多変量）の関係へ

- 1 変量: 1 つの変数から状態を把握するだけでは因果関係は不明
- 2 変量の関係: 説明変数とアウトカム変数の相関関係
e.g. 職業と年収の関係, 性別と投票参加率の関係など…

→ 分析するためには, 量/質的変数をまずは見極める

② クロス表の分析: 質的変数*質的変数

- クロス表: とくに名義尺度と名義尺度の関係を分析する際に利用
- あるカテゴリにおいて, なにかの状態に関する割合関係を見る
e.g. 性別（名義）による職業（名義）の関係
- R 上で使うコマンド → table 系のプログラム

質的変数の間の関係を理解する

① 質的変数間の関係を分析

- 一番単純なコマンドライン: `table(var1,var2)`
- 割合についても理解するために…
- ✓ 行パーセント: `prop.table(table(var1,var2),margin = 1)`
- ✓ 列パーセント: `prop.table(table(var1,var2),margin = 2)` → **ほぼ使わない!**

② 例題: 教育程度ごとの与党/野党/無党派の支持者の割合をクロス表で確認

- ✓ 欠損値処理を忘れないように!
- ✓ 教育程度を説明要因として, 行パーセントで表記.
- ✓ 途中, ラベル付けと丸め作業もやってみる

量的/質的変数の関係を理解する

① 量的/質的変数の関係を理解する

- 量的変数と質的変数でクロス表をやってもさほど明確な情報は得られない

e.g.1 年齢ごとの政党支持の関係をクロス表してみよう！

e.g.2 世代ごとの政党支持の関係をクロス表してみる

- しかし、量的変数は変数内容そのものに情報量があるので、これを使わない手はない

→ カテゴリ（名義尺度）ごとの平均値を見てみよう！

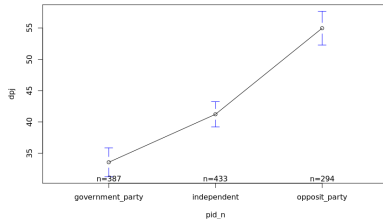
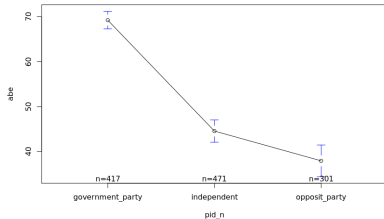
```
> #年齢とクロス
> round(prop.table(table(age,pid_n),1)*100,1)
      pid_n
age  government_party independent opposit_party
18      66.7          0.0          33.3
19      50.0          0.0          50.0
20      33.3          4.4          62.2
21      71.4          1.4          27.2
22      44.4          2.2          53.3
23      66.7          0.0          33.3
24      26.7          5.3          67.9
25      36.4          6.6          56.9
26      25.0          4.9          70.1
27      22.2          5.3          72.5
28      39.3          4.4          56.2
29      30.0          5.0          65.0
30      46.2          3.1          50.7
31      35.3          7.1          57.6
32      45.5          7.3          47.2
33      49.0          6.0          45.0
```

```
> round(prop.table(table(generation, pid_n),1)*100,1)
      pid_n
generation government_party independent opposit_party
20代          33.9          47.1          19.0
30代          40.3          42.3          17.3
40代          24.0          53.9          22.0
50代          33.3          46.7          20.0
60代          35.2          34.8          29.9
70代          34.3          29.5          36.2
```

Rstudio で分析

① Rstudio で分析

- 質的変数側がダミー変数（2 値の場合）→ t 検定
 - ✓ 用いる package1: `t.test(numeric_var1 ~ factor_var2, var.equal = TRUE)`
- 質的変数側がダミー変数（3 値以上の場合）→ 分散分析
 - ✓ `TukeyHSD(aov(numeric_var1 ~ factor_var2))`
 - ✓ `pairwise.t.test(numeric_var1, factor_var2, p.adj = "検定方法 (bonf)")`



宿題: 各政党への感情温度

① 宿題: 各政党への感情温度 (q4_1_x) の平均値を確認

- ✓ 無党派層に最も嫌われている政党はどこでしょうか
- ✓ 男女で最も好感度に差がある政党はどこでしょうか
- ✓ 20代で最も好感度に差がある政党はどこでしょうか
- 上の3つを分析して具体的な数値とともに答えを示しておいてください。