

政治学 3 (計量政治学)

3 回目

秦 正樹

京都府立大学公共政策学部 准教授

hatamasaki@kpu.ac.jp

2022/04/26

火 3 コース@情報処理室

3 回目のテーマ

① 3 回目のテーマ

- 統計的推測の考え方を学ぶ
 - データの見方と処理について
 - 母集団と標本の考え方
 - 度数分布と記述統計
- データ分析の基盤となる技術について，Rstudio を使いつつ勉強しましょう！

② 参考資料など

- Teams にレジユメをあげてありますのでご確認を.
- 質問があれば，直接呼びかけてください.

分析するためのデータセット

① データセット

- 飯田健. (2013)『計量政治学』共立出版はちゃんと購入できましたか？
- 分析するために必要なもの→データ (csv が望ましい) + コードブック (質問票)
- **ここから**システムに入って、質問票とデータを見てみましょう！
- コードブックと画面のデータをちょっと見くらべてみてください

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	q4_1
id	sex	age_raw	pref	q2_1	q2_2	q2_3	q2_4	q3_1	q3_2	q3_3	q3_4	q4_1_1	q4_1_2	q4_1_3		
2	1	1	28	23	28	4	7	3					50	50	50	
3	2	1	25	4	25	2	10	3								
4	3	0	27	15	27	4	2	3								
5	4	0	29	20	29	3	3	4					50	50	50	
6	5	1	43	14	43	2	4	3								
7	6	1	47	12	47	6	21	2					50	50	50	
8	7	0	28	11	28	3	6	3				2	50	50	50	
9	8	1	38	13	38	4	5	1				1	10	10	0	
10	9	1	53	13	53	4	14	1		3			10	40	40	
11	10	1	58	42	58	4	17	3			3		70	20	10	
12	11	1	59	27	59	4	10	3			2		30	60	70	
13	12	0	70	1	70	2	1	2		2						
14	13	0	46	16	46	3	15	4			2		50	50	50	
15	14	0	45	9	45	2	1	4		1			50	50	50	
16	15	0	46	28	46	4	16	3		2			50	50	50	
17	16	0	73	13	70	4	1	1			1		70	40	40	
18	17	1	33	33	33	2	9	3			1		50	40	30	
19	18	0	43	17	43	3	13	3					60	50	50	
20	19	1	75	33	75	4	6	2			3		70	0	0	
21	20	1	27	13	27	2	1	1		5			70	20	30	
22	21	0	69	13	69	2	2	1				2	70	20	30	
23	22	0	48	27	48	4	3	3			2		30	60	70	
24	23	1	53	13	53	4	10	1								
25	24	0	63	33	63	2	5	3			6					
26	25	0	70	13	69	3	15	1				3	40	50	50	
27	26	0	45	13	45	3	9	1		2			20	0	0	
28	27	1	55	2	55	4	18	3		5						
29	28	0	66	9	66	4	1	3				1				
30	29	0	36	13	36	2	8	3					10	10	10	
31	30	1	57	28	57	4	14	3		2			70	0	10	
32	31	0	65	23	65	2	21	3				2	10	40	30	
33	32	1	61	14	61	4	21	2			1		10	10	10	
34	33	1	65	12	65	4	3	3			2		60	10	0	
35	34	0	29	28	29	4	9	3				3	60	0	0	
36	35	0	34	27	34	2	6	2					60	0	0	
37	36	0	43	13	43	4	5	1			3		30	20	10	
38	37	1	45	23	45	3	1	3				1	0	100	100	

【2019年3月実施】政治と社会に関する意識調査（陰謀論科研）

※ 調査時期：2019年3月6日～8日

※ 調査方法：調査補助フォーム Qualtrics を利用した WEB 調査

※ 調査対象：楽天インサイトのパネルモニター18歳から79歳の男女15**名
(2015年国勢調査を元に、人口分布に合う形で割付を行った上で調査を実施)

Q1_2 (イントロ) : 調査を開始する (1)

Q2_1 2019年1月31日現在のあなたの実年齢を、下記の欄に半角数字でご記入下さい。

Q2_2 あなたが最後に在籍した(あるいは、現在在籍している)学校はこの中のどれにあたりますか。

※ 中学校 (1)

※ 高校 (2)

※ 短大・高专・専門学校 (3)

※ 4年制大学 (4)

※ 大学院 (5)

※ 言いたくない (6)

変数と ID

① 一番上の行→変数のリスト

- id (ナンバリング)・sex (性別)・age_raw(年齢)・pref (都道府県)・q2_1…
- q2_1 以降は、質問票と対応。コードブックと見比べながらやります！
- 空白の部分→欠損値 (missing data)。一般的に分析から除外することが多い

② 列→回答者個人ごとのデータ

- id1 の人は、28 歳で、大卒で、年収 550-600 万、一般市在住で…という感じ
- このデータでは 1509 人分の個人データが所蔵されている
- 質問数はおよそ 145 問ほどあって、それぞれのデータは入っている

③ 分析すること

- これらのデータを使って、まずは、平均・分散・4 分位について分析してみよう！
- 欠損値の処理をしないとどうなるかを見てみよう！

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
- 間隔尺度：目盛りが等間隔になっていて、その間隔に意味があるもの
e.g. 気温：30 度の気温のとき、10 度よりも 3 倍暑いとはいえない。

② 変数と尺度：質的変数

- 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？
 - 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
- ✓ ただし、分析の便宜上、順序尺度は比例尺度（量的）に扱う場合も多い

尺度と確認方法

① 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…

② 順序尺度の取り扱い

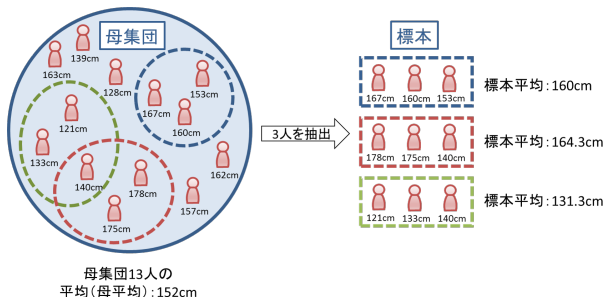
- 順序尺度：数字の大小には意味があるが間隔が等価でないもの
- * e.g. 政治関心：「あなたは政治に関心がありますか？」
1. 関心あり 2. やや関心あり 3. あまり関心なし 4. 関心なし
- 「やや関心あり」→「関心あり」の差（1）と「やや関心」と「あまり関心なし」の差（1）が同じであるというのは色々おかしい？
- ✓ この授業では順序尺度も「量的変数」として扱う（平均にも一定の意味はあるから）

サンプリングと調査方法

① 調査の種類とやり方

- 全数調査：存在するサンプルのすべてに行う調査（母集団）
- 標本調査：存在するサンプルから一部を取り出して（＝抽出/サンプリング）行う調査（標本）
- ランダムサンプリング調査：標本抽出の際の方法を無作為に行うこと

→ 十分な標本数であればランダム抽出による標本平均は、母平均に近似する傾向



一致性と「大数の法則」

① 一致性と「大数の法則」

- ときどきある話：なぜ、世論調査では、1.2 億人の意見を、たった 1500 人そこらの意見でわかるのか？あれは陰謀だ！
- 十分な標本数であればランダム抽出による標本平均は、母平均に近似する傾向
- 「大数の法則」と一致性の原則にもとづく
- 具体的にみなさんも Rstudio でやってみましょう！ → [ここからアクセス](#)

