



カフェインず(仮)

プロジェクト中間発表

# 日本人の生活データを用いた コーヒーの短期的及び長期的効果の統計的分析

武蔵野大学 工学部 数理工学科  
カフェインず(仮)

阿辻颯姫 白川桃子 田口冬佳 村田滉希

# 目次

## 1.統計データ分析コンペティション2023

概要

SSDSEについて

## 2.データ分析経過

分析の流れ

説明変数の選択

重回帰分析

分析結果のまとめ

結果の考察

## 3.今後の展望

今後の活動計画



# 統計データ分析コンペティション－概要

## 【概要】

総務省統計局、統計センター主催。

統計データを様々な手法で分析した結果を論文にまとめ、アイデアや解析力を競いつつ統計リテラシーを向上させることを目的としたコンペティション。

## 【論文提出締め切り】

令和5年8月31日

## 【表彰】

総務大臣賞、優秀賞、統計数理賞、統計活用奨励賞 など

## 【参加経緯】

データを信用できるソースから収集したり、適切な手法で分析したりすることができる能力を身につけたいと思う為。

# 統計データ分析コンペティション – SSDSE

Standardized Statistical Data Set for Education(教育用標準データセット)

独立行政法人統計センターが作成・公開している、**地域や性別ごとの様々な統計データ**

名称	収録項目
SSDSE-市区町村	社会・人口統計体系データを市区町村単位で抽出
SSDSE-県別推移	社会・人口統計体系データを都道府県単位で抽出
SSDSE-家計消費	食品の品目別年間支出金額データを 県庁所在地の1世帯ごとに収録
SSDSE-社会生活	生活の行動や時間に関するデータを 男女、都道区県別に収録
SSDSE-基本素材	他分野の都道府県別データを集めたデータセット

	A	B	C	D	E	F	G	H	I	J	K	
	SSDSE-A-202	Prefecture	Municipality	A1101	A110101	A110102	A1102	A110201	A110202	A1301	A130101	A130102
			年度	2020	2020	2020	2020	2020	2020	2020	2020	2020
	地域コード	都道府県	市区町村	総人口	総人口 (男)	総人口 (女)	日本人口	日本人口 (男)	日本人口 (女)	15歳未満人口	15歳未満人口 (男)	15歳未満人口 (女)
1	R01100	北海道	札幌市	1973395	918682	1054713	1933094	897727	1035367	215366	110196	
2	R01202	北海道	函館市	251084	113965	137119	248208	112718	135490	23560	11984	
3	R01203	北海道	小樽市	111299	50136	61163	109971	49441	60530	9169	4678	
4	R01204	北海道	旭川市	329306	152108	177198	325287	150318	174969	34691	17797	
5	R01205	北海道	室蘭市	82383	40390	41993	81658	39960	41698	7769	3916	
6	R01206	北海道	釧路市	165077	77506	87571	163026	76789	86237	16634	8495	
7	R01207	北海道	帯広市	166536	79623	86913	165759	79211	86548	19073	9762	
8	R01208	北海道	北見市	115480	54729	60751	114601	54346	60255	12014	6170	
9	R01209	北海道	夕張市	7334	3381	3953	7247	3373	3874	411	199	
10	R01210	北海道	岩見沢市	79306	37105	42201	79031	36970	42061	7859	3961	
11	R01211	北海道	網走市	35759	18060	17699	35429	17962	17467	3731	1932	
12	R01212	北海道	留萌市	20114	9649	10465	19940	9610	10330	1711	885	
13	R01213	北海道	苫小牧市	170113	83522	86591	168873	82739	86134	20426	10593	
14	R01214	北海道	稚内市	33563	16651	16912	33083	16509	16574	3484	1806	
15	R01215	北海道	美唄市	20413	9565	10848	20343	9527	10816	1539	813	
16	R01216	北海道	芦別市	12555	5762	6793	12536	5758	6778	845	435	
17	R01217	北海道	江別市	121056	57523	63533	119589	56702	62887	13400	6786	
18	R01218	北海道	赤平市	9698	4388	5310	9631	4355	5276	603	299	
19	R01219	北海道	紋別市	21215	9901	11314	20674	9803	10871	1971	992	
20	R01220	北海道	士別市	17858	8416	9442	17756	8349	9407	1652	855	
21	R01221	北海道	名寄市	27282	13322	13960	27014	13167	13847	3018	1561	
22	R01222	北海道	三笠市	8040	3631	4409	8025	3621	4404	655	315	
23	R01223	北海道	根室市	24636	11762	12874	24249	11666	12583	2396	1217	
24	R01224	北海道	千歳市	97950	49790	48160	97223	49450	47773	13012	6681	
25	R01225	北海道	滝川市	39490	18832	20658	39013	18548	20465	3995	2104	

# 統計データ分析コンペティション－テーマ

## 【分析テーマ】

日本人の生活データを用いた、  
コーヒーの短期的及び長期的効果の統計的分析

## 【今回の発表内容について】

本論文では生活習慣と病気(がん)について調査を行ったが、  
中間報告では**生活習慣の結果のみ**報告する

# 分析の流れ

		分析① 生活行動	分析② 趣味・娯楽	分析③ スポーツ
変数選択		5種 → 2種	12種 → 7種	14種 → 6種
重回帰分析	目的変数	コーヒー数量	コーヒー数量	コーヒー数量
	説明変数	学習・自己啓発訓練の総数 スポーツの総数 趣味・娯楽の総数 ボランティア活動の総数 旅行・娯楽の総数	演芸・演劇・舞踊 鑑賞、写真の撮 影・プリント、茶 道、陶芸・工芸、 編み物・手芸、音 楽鑑賞、将棋	器具を使ったト レーニング、ゲー トボール、ボウリ ング、剣道、野球、 バドミントン



# 変数選択（説明変数選択基準による変数選択）

i) 上田の説明変数選択

$$R_u = 1 - (1 - R^2) \times \frac{\text{データの数} + \text{説明変数の個数} + 1}{\text{データの数} + \text{説明変数の個数} - 1} \quad (R \text{ は重相関係数})$$

ii) 芳賀・竹内・奥野の説明変数選択規準

$$R_h = 1 - (1 - R^2) \times \frac{(\text{データの数} - 1)}{(\text{データの数} + 1)} \times \frac{(\text{データの数} + \text{説明変数の個数} + 1)}{(\text{データの数} - \text{説明変数の個数} - 1)} \quad (R \text{ は重相関係数})$$

iii) 赤池情報量規準 (AIC)

$$AIC = \text{データの数} \times \log_e(1 - R^2) + 2 \times \text{説明変数の個数} \quad (\log_e \text{ は自然対数、} R \text{ は重相関係数})$$

$R_u, R_h$ が**最大**、 $AIC$ が**最小**になる説明変数の個数を選択する

# 分析結果 生活行動 - 説明変数の選択

| 相関係数 |  $\leq 0.95$  であるので、すべての変数を調べる  
変数選択基準より、説明変数は2個が適切である

相関係数	学習・自己啓発・ 訓練の総数	スポーツの総数	趣味・娯楽の総数	ボランティア活動 の総数	旅行・行楽の総数
学習・自己啓発・ 訓練の総数	1.00	0.85	0.88	-0.18	0.68
スポーツの総数	0.85	1.00	0.82	-0.09	0.63
趣味・娯楽の総数	0.88	0.82	1.00	-0.13	0.77
ボランティア活動 の総数	-0.18	-0.09	-0.13	1.00	0.09
旅行・行楽の総数	0.68	0.63	0.77	0.09	1.00

図1 生活行動総数間の相関係数マトリクス

$r \leq 0.5$  : オレンジ  
 $0.0 \leq r \leq 0.09$  : 黄色  
 $0.09 < r < 1.00$  : 赤  
 $r < 0.0, r = 1.00$  : 白

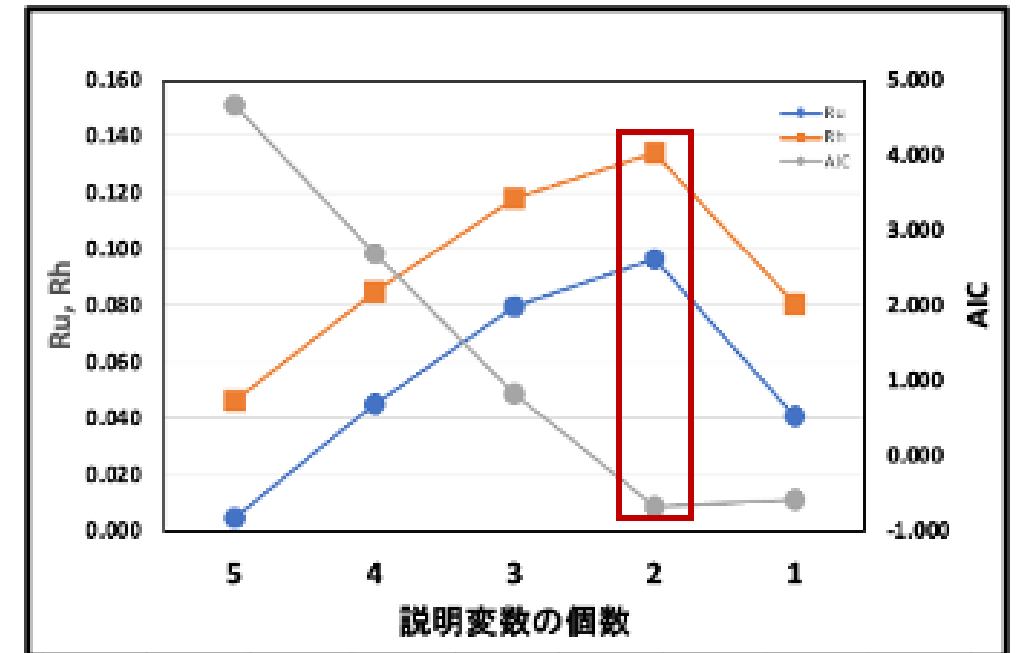
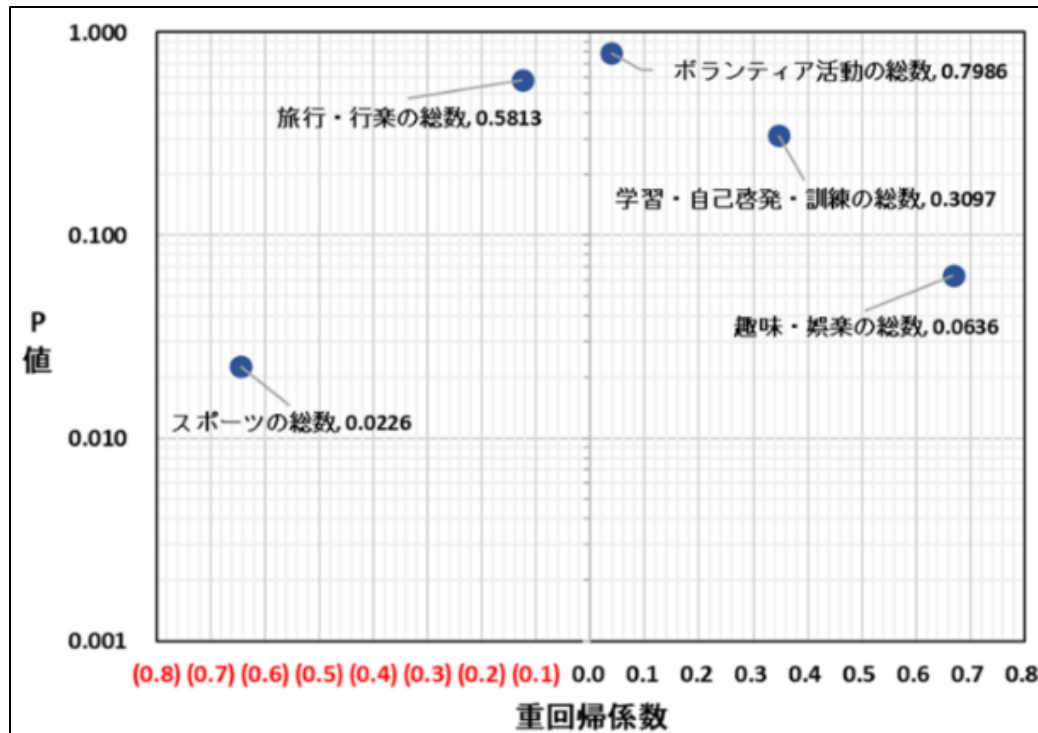


図2 生活習慣総数Ru,Rh,AICの説明変数の数依存性

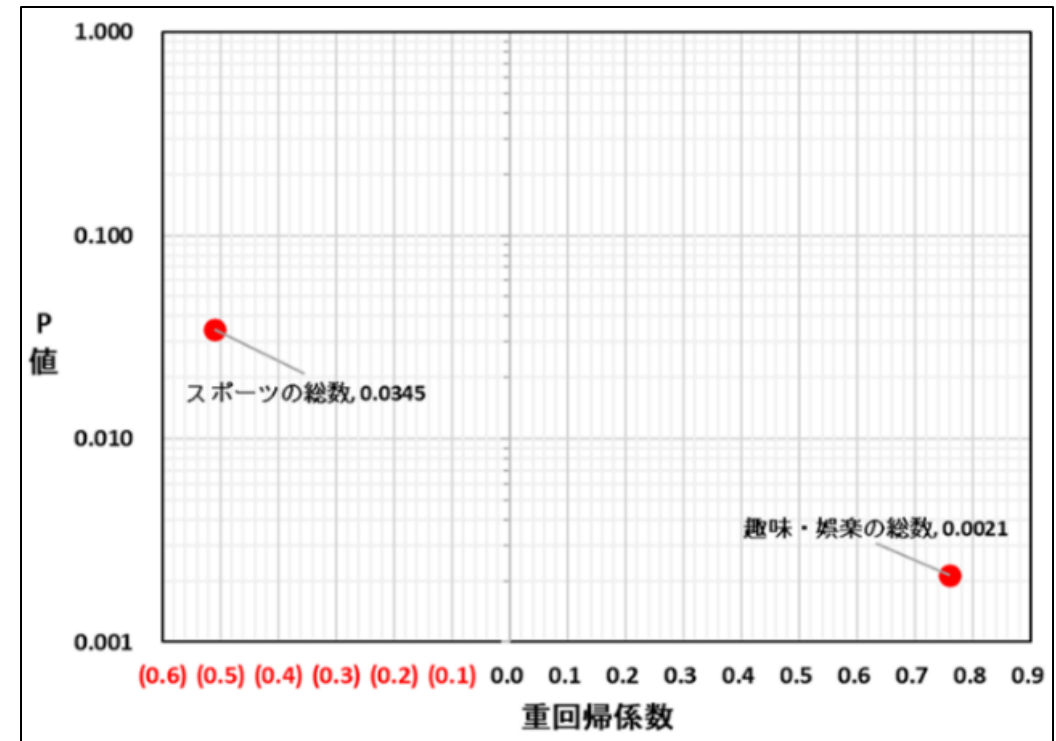


# 分析結果 生活行動 - 重回帰分析

趣味・娯楽の総数, スポーツの総数が選択された



生活行動の重回帰係数と p 値 (変数選択前, 5変数)



生活行動の重回帰係数と p 値 (変数選択後, 2変数)

重回帰係数: (0.5)  $\rightarrow$  -0.5

# 分析結果 趣味・娯楽 – 説明変数の選択

相関係数  $|r| \leq 0.95$  であるので、すべての変数を調べる  
変数選択基準より、説明変数は7個が適切である

相関係数	美術鑑賞 (テレビ・スマートフォン・パソコンなどは除く)	演芸・演劇・舞踊鑑賞(テレビ・スマートフォン・パソコンなどは除く)	CD・スマートフォンなどによる音楽鑑賞	楽器の演奏	茶道	編み物・手芸	陶芸・工芸	写真の撮影・プリント	趣味としての読書	囲碁	将棋	テレビゲーム・パソコンゲーム(家庭で行うもの、携帯用を含む)
美術鑑賞 (テレビ・スマートフォン・パソコンなどは除く)	1.00	0.80	0.72	0.68	0.55	0.45	0.67	0.72	0.82	0.47	0.46	0.66
演芸・演劇・舞踊鑑賞(テレビ・スマートフォン・パソコンなどは除く)	0.80	1.00	0.85	0.83	0.33	0.52	0.70	0.82	0.89	0.39	0.50	0.73
CD・スマートフォンなどによる音楽鑑賞	0.72	0.85	1.00	0.84	0.14	0.46	0.67	0.89	0.91	0.41	0.46	0.85
楽器の演奏	0.68	0.83	0.84	1.00	0.22	0.44	0.70	0.86	0.82	0.41	0.50	0.76
茶道	0.55	0.33	0.14	0.22	1.00	0.37	0.28	0.23	0.29	0.39	0.31	0.31
編み物・手芸	0.45	0.52	0.46	0.44	0.37	1.00	0.47	0.56	0.58	0.24	0.30	0.53
陶芸・工芸	0.67	0.70	0.67	0.70	0.28	0.47	1.00	0.72	0.75	0.37	0.26	0.60
写真の撮影・プリント	0.72	0.82	0.89	0.86	0.23	0.56	0.72	1.00	0.89	0.39	0.47	0.89
趣味としての読書	0.82	0.89	0.91	0.82	0.29	0.58	0.75	0.89	1.00	0.52	0.48	0.82
囲碁	0.47	0.39	0.41	0.41	0.39	0.24	0.37	0.39	0.52	1.00	0.26	0.27
将棋	0.46	0.50	0.46	0.50	0.31	0.30	0.26	0.47	0.48	0.26	1.00	0.49
テレビゲーム・パソコンゲーム(家庭で行うもの、携帯用を含む)	0.66	0.73	0.85	0.76	0.31	0.53	0.60	0.89	0.82	0.27	0.49	1.00

図5 趣味・娯楽の相関係数マトリクス

$r \leq 0.5$  : オレンジ  
 $0.0 \leq r \leq 0.09$  : 黄色  
 $0.09 < r < 1.00$  : 赤  
 $r < 0.0, r = 1.00$  : 白

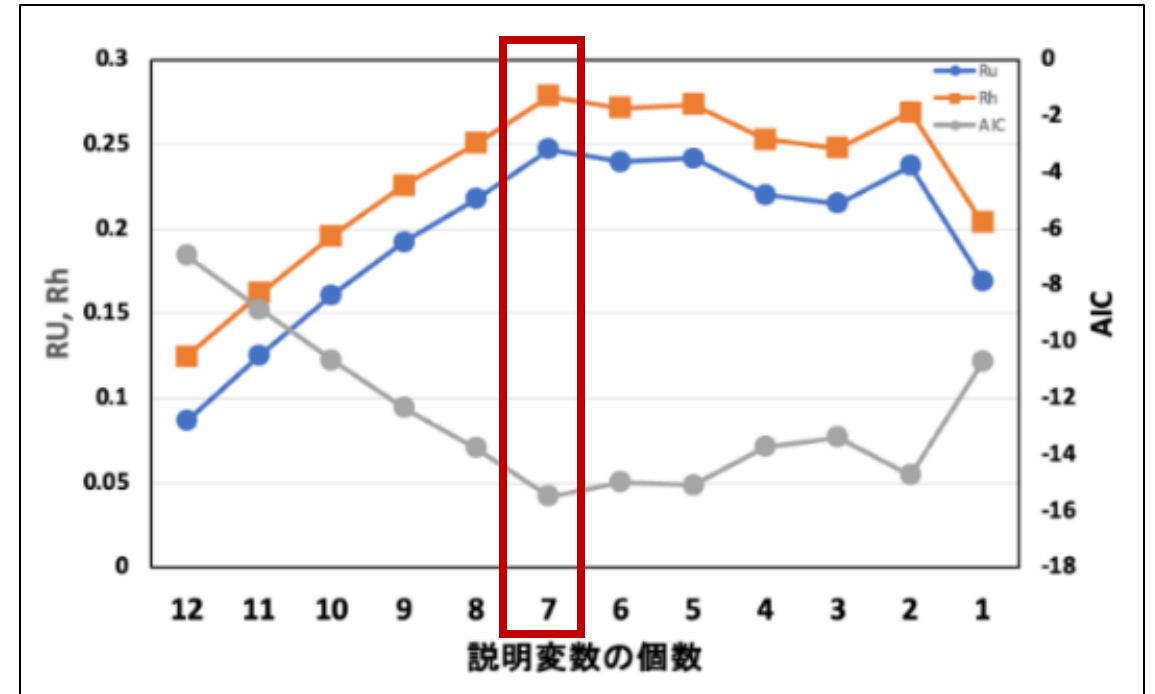
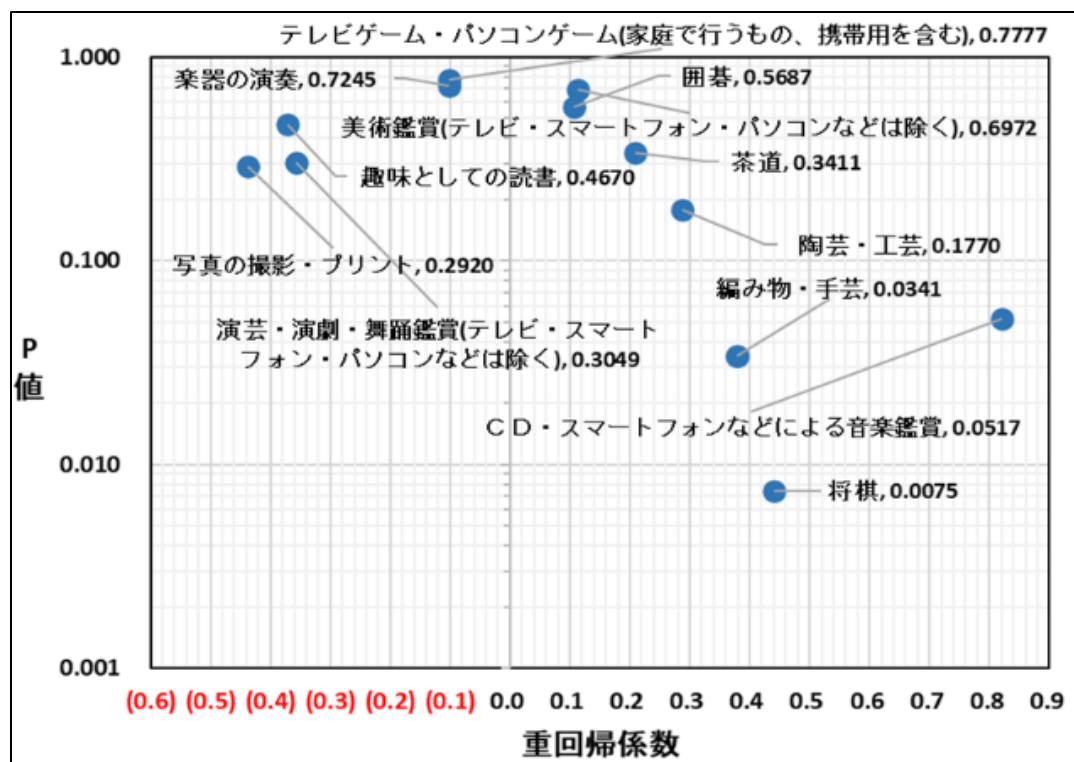


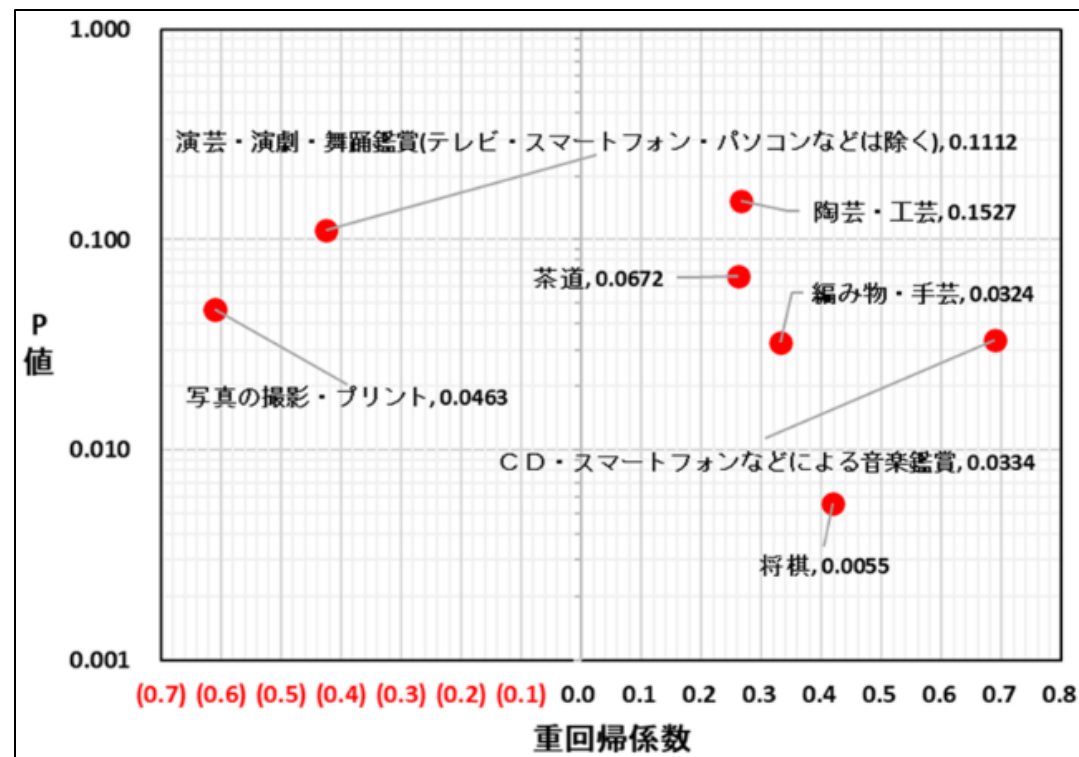
図6 趣味・娯楽のRu,Rh,AICの説明変数の数依存性

# 分析結果 趣味・娯楽 - 重回帰分析

将棋, 茶道, 音楽鑑賞, 編み物, 陶芸, 写真, 演芸が選択された



趣味・娯楽の重回帰係数と p 値 (変数選択前, 12 変数)



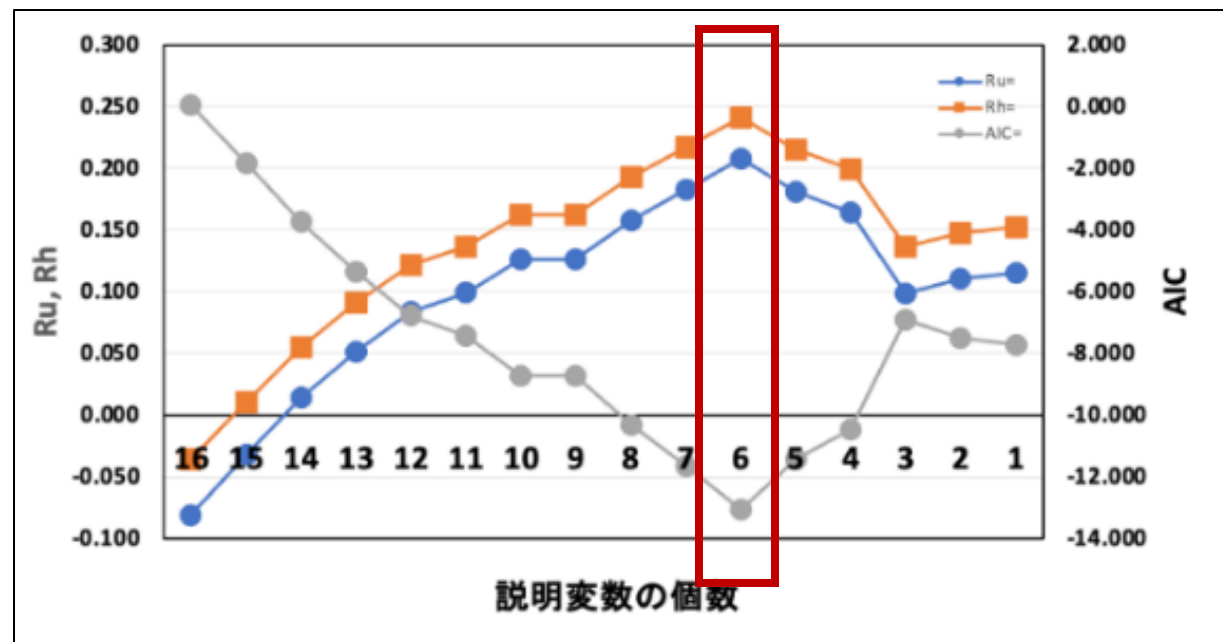
趣味・娯楽の重回帰係数と p 値 (変数選択後, 7 変数)

# 分析結果 スポーツ – 説明選択

| 相関係数 |  $\leq 0.95$ であるので、すべての変数を調べる  
変数選択基準より、説明変数は6個が適切である

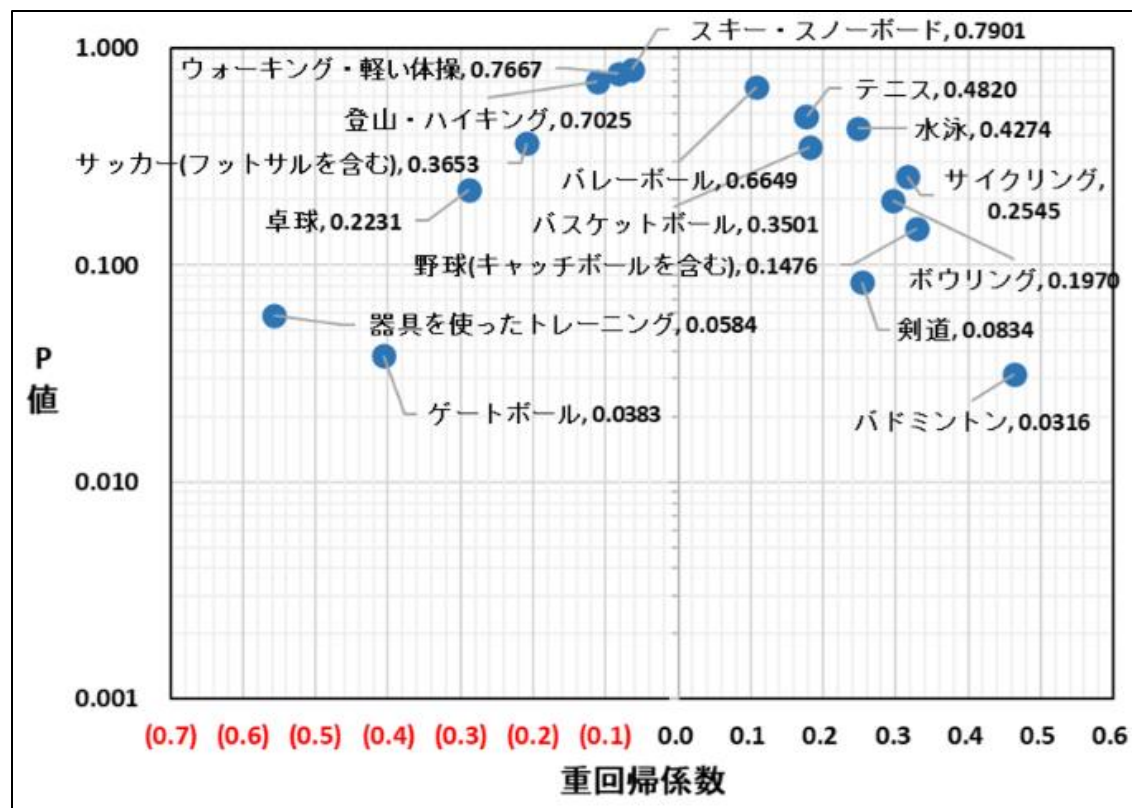
説明変数	野球(キャッチボールを含む)	バレーボール	バスケットボール	サッカー(フットサルを含む)	野球	テニス	バドミントン	剣道	ゲートボール	ボウリング	水泳	スキー・スノーボード	登山・ハイキング	サイタリング	ウォーキング・軽い体操	器具を使ったトレーニング
野球(キャッチボールを含む)	1.00	-0.08	0.63	0.40	0.34	0.26	0.40	0.01	0.13	0.63	0.31	0.21	0.12	0.23	0.20	0.57
バレーボール	-0.08	1.00	-0.03	-0.27	-0.30	-0.29	-0.16	-0.03	0.00	-0.05	-0.41	-0.57	-0.54	-0.29	-0.41	-0.22
バスケットボール	0.63	-0.03	1.00	0.43	0.34	0.12	0.39	-0.13	0.14	0.44	0.27	0.20	0.06	0.10	0.05	0.59
サッカー(フットサルを含む)	0.40	0.27	0.43	1.00	0.57	0.81	0.40	0.23	0.07	0.58	0.70	0.11	0.42	0.53	0.64	0.71
野球	0.34	-0.30	0.34	0.57	1.00	0.54	0.81	0.09	-0.12	0.38	0.65	0.45	0.54	0.67	0.56	0.63
テニス	0.26	-0.20	0.12	0.81	0.54	1.00	0.16	0.16	-0.17	0.44	0.73	0.14	0.59	0.66	0.76	0.61
バドミントン	0.40	-0.16	0.39	0.40	0.81	0.16	1.00	0.03	0.10	0.24	0.37	0.40	0.34	0.40	0.23	0.47
剣道	0.01	-0.03	-0.13	0.09	0.09	0.16	0.03	1.00	0.17	-0.03	0.04	-0.02	0.13	0.00	0.16	0.03
ゲートボール	0.13	0.00	0.14	-0.07	-0.12	-0.17	0.10	0.17	1.00	0.08	-0.16	-0.18	-0.20	-0.34	-0.21	-0.06
ボウリング	0.63	-0.05	0.44	0.58	0.38	0.44	0.24	-0.03	0.08	1.00	0.53	0.03	0.16	0.32	0.38	0.71
水泳	0.31	-0.41	0.27	0.70	0.65	0.73	0.37	0.04	-0.16	0.53	1.00	0.30	0.70	0.68	0.74	0.73
スキー・スノーボード	0.21	-0.57	0.20	0.11	0.45	0.14	0.40	-0.02	-0.18	0.03	0.30	1.00	0.69	0.51	0.13	0.21
登山・ハイキング	0.12	-0.54	0.06	0.42	0.54	0.59	0.34	0.13	-0.25	0.16	0.70	0.62	1.00	0.73	0.57	0.39
サイタリング	0.23	-0.29	0.10	0.53	0.67	0.66	0.40	0.08	-0.34	0.32	0.66	0.51	0.73	1.00	0.71	0.51
ウォーキング・軽い体操	0.20	-0.41	0.05	0.64	0.56	0.76	0.23	0.16	-0.21	0.38	0.74	0.13	0.57	0.71	1.00	0.62
器具を使ったトレーニング	0.57	-0.22	0.50	0.71	0.63	0.61	0.47	0.03	0.05	0.71	0.73	0.21	0.39	0.51	0.62	1.00

$r \leq 0.5$  : オレンジ  
 $0.0 \leq r \leq 0.09$  : 黄色  
 $0.09 < r < 1.00$  : 赤  
 $r < 0.0, r = 1.00$  : 白

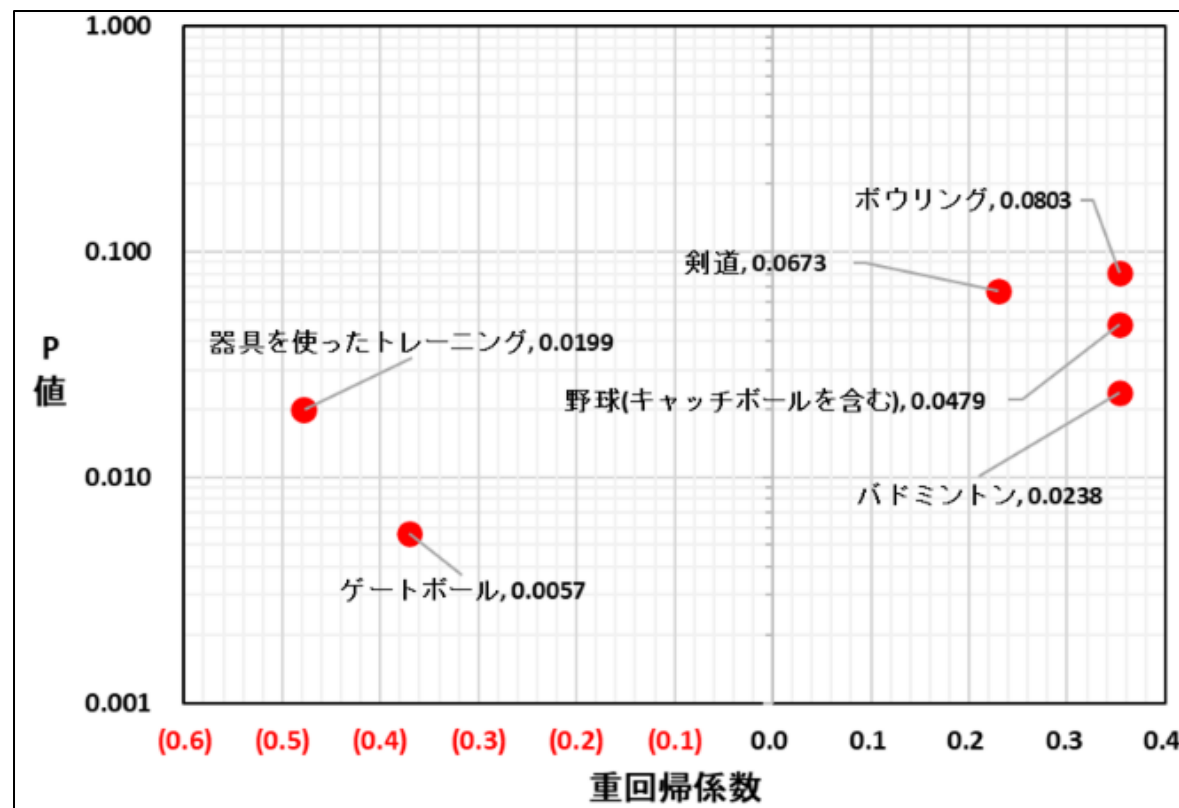


# 分析結果 スポーツ - 重回帰分析

バドミントン, 剣道, ボウリング, 野球, ゲートボール, 器具を使ったトレーニングが選択された



趣味・娯楽の重回帰係数と p 値 (変数選択前, 12変数)



趣味・娯楽の重回帰係数と p 値 (変数選択後, 7変数)

# 結果

- 生活行動 \* コーヒー
  - 「趣味・娯楽」 「スポーツ」
- 趣味・娯楽 \* コーヒー
  - 重回帰係数が**正**：「**将棋**」 「**茶道**」 「**音楽鑑賞**」 「**編み物**」 「**陶芸**」
  - 重回帰係数が**負**：「写真」 「演芸」
- スポーツ \* コーヒー
  - 重回帰係数が**正**：「**バドミントン**」 「**剣道**」 「**ボウリング**」 「**野球**」
  - 重回帰係数が**負**：「ゲートボール」 「器具を使ったトレーニング」





# 考察

- 趣味・娯楽の正の相関：「将棋」「茶道」「音楽鑑賞」「編み物」「陶芸」
- スポーツの正の相関：「バドミントン」「剣道」「ボウリング」「野球」

どの項目も「**集中力**」を要する項目

→カフェインには眠気の解消と集中力を高める効果があるからかも？

# 今後の展望

最終発表で論文のがんについてを発表予定  
コーヒーの摂取の効果は年単位の蓄積が影響しているのでは  
→**時系列を考慮して**分析する

- 11月～4学期末まで
  - データ分析の勉強(Python)
  - 参考書：東京大学データサイエンティスト育成講座

ありがとうございました

＼ ご清聴ありがとうございました ／