

---

# ウェアラブル端末を用いた 健康向上のための分析

---

はじまりを、つくる

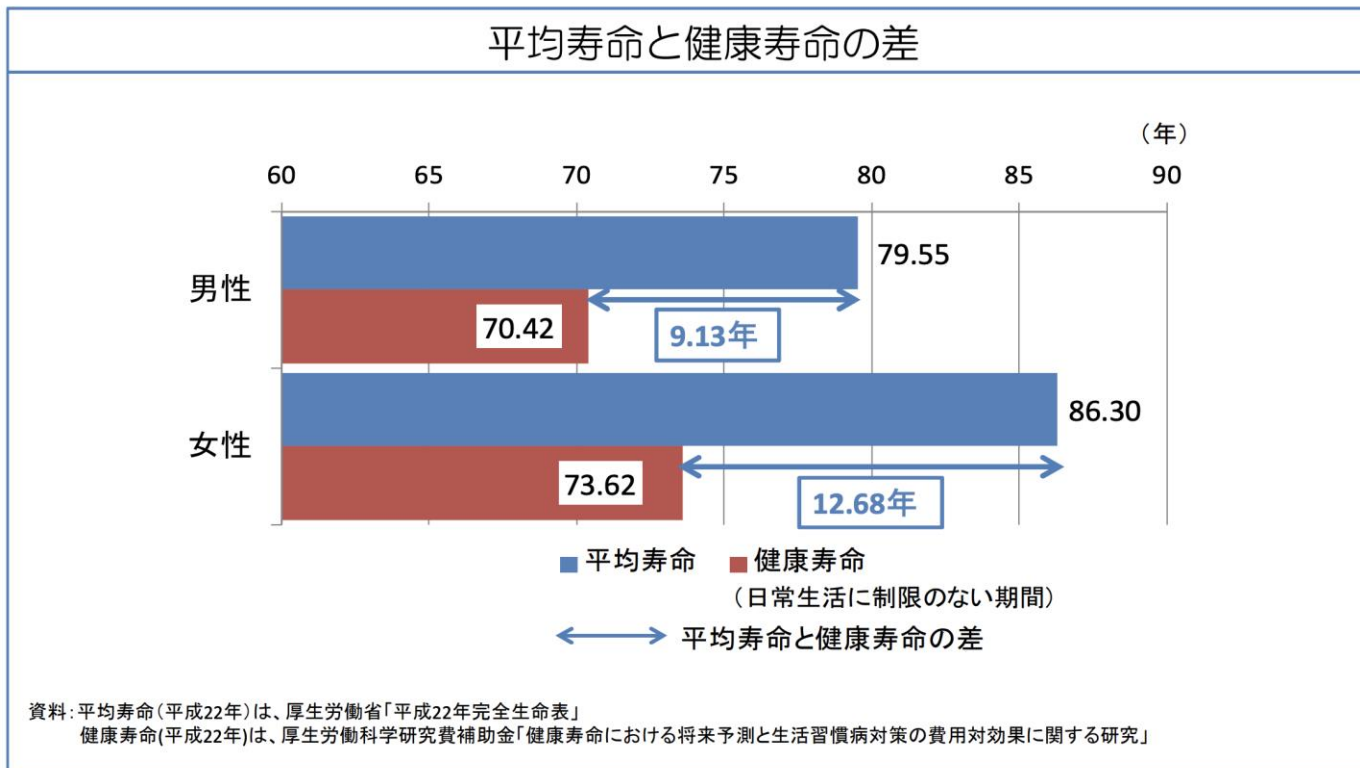


静岡理工科大学 伊藤太一、大場春佳、水野 信也  
宮城大学 藤澤由和  
静岡県立大学 合田敏尚



NTTデータ数理システム学生研究奨励賞  
2020/12/18

日本では寿命が伸びているが、健康寿命は介護状態の期間が長い



[出典] 厚生科学審議会地域保健健康増進栄養部会・次期国民健康づくり運動プラン策定専門委員会  
「健康日本21(第二次)の推進に関する参考資料」p25

健康寿命を延伸することで、生活の質向上と社会保障費の負担減が可能

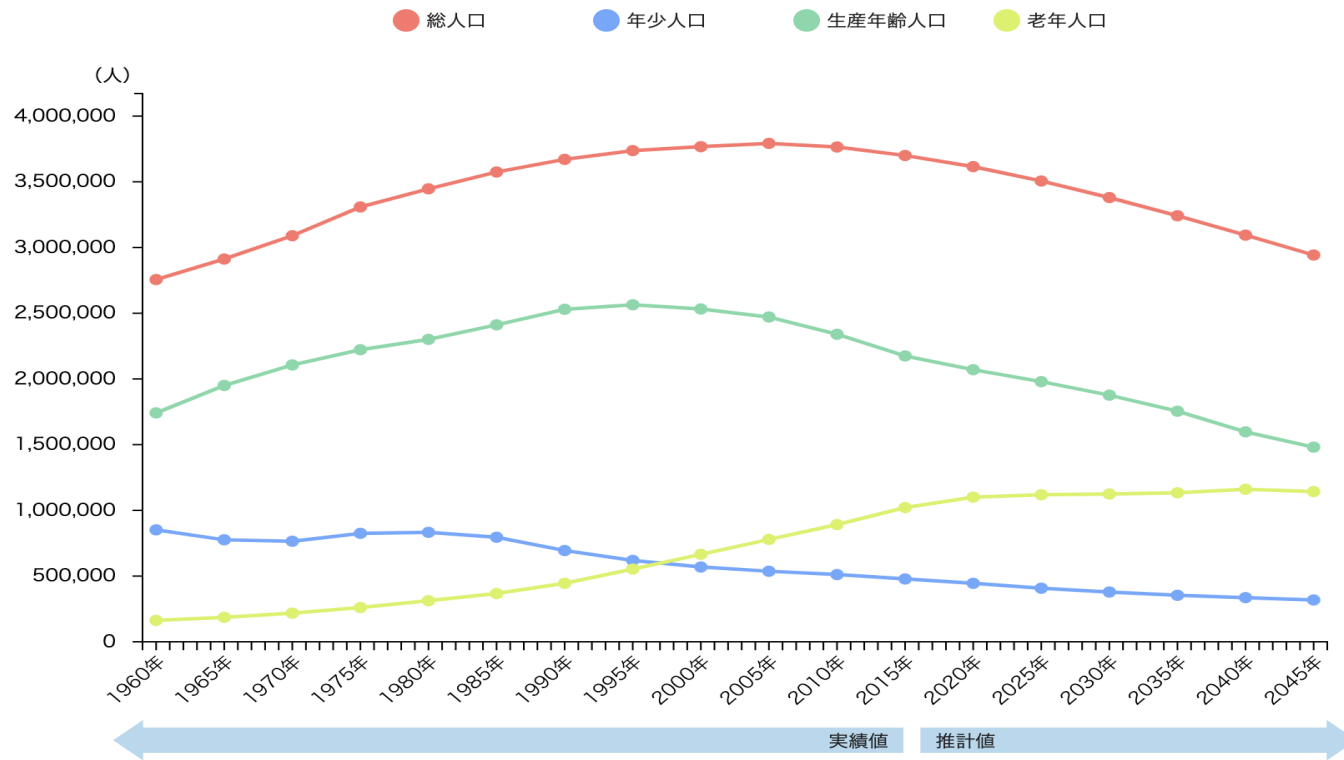
# 人口減少と老年人口の増加



静岡県内では総人口、年少人口、生産年齢ともに減少が予想されているのに対して、老年人口は増加が予想されている。

## 人口推移

静岡県

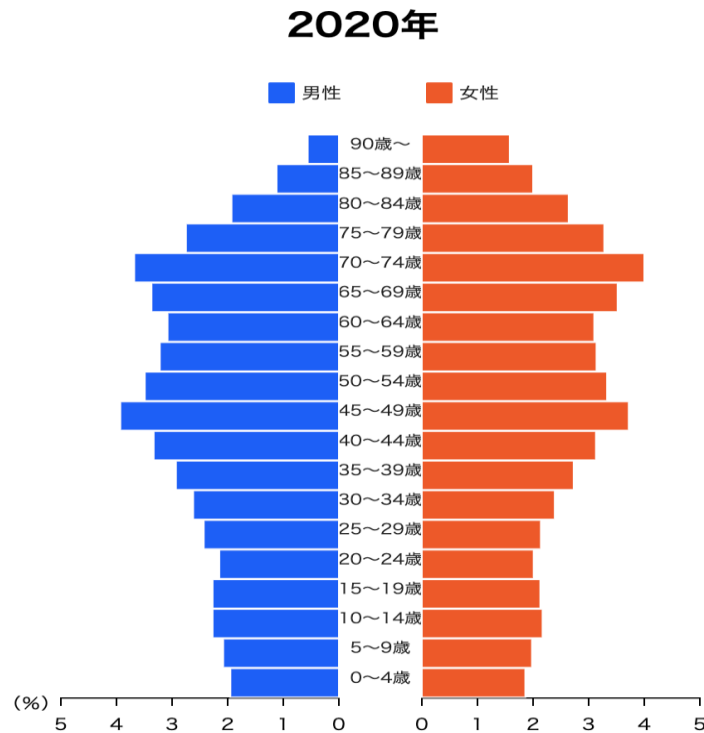


# 人口減少と老年人口の増加

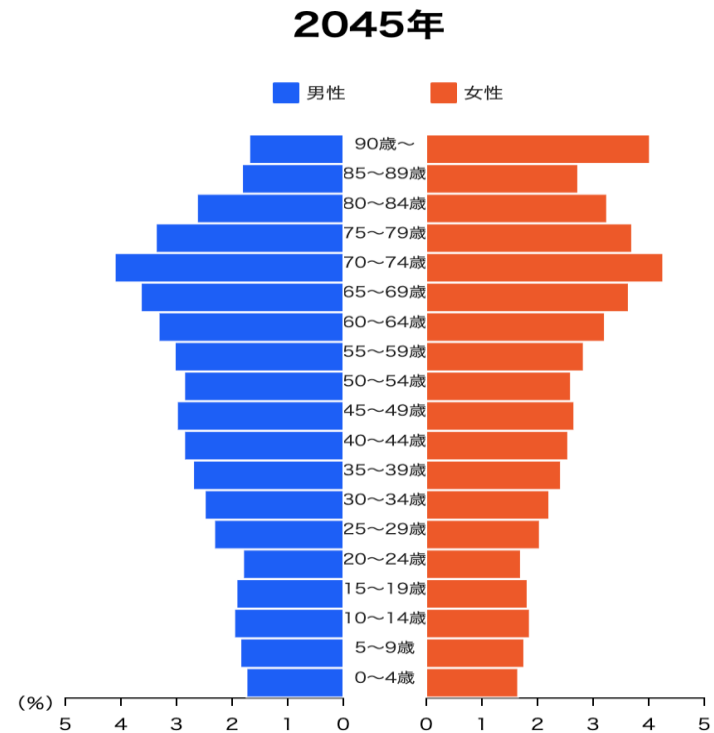


静岡県内では総人口、年少人口、生産年齢ともに減少が予想されているのに対して、老年人口は増加が予想されている。

静岡県



老年人口（65歳以上）：1,100,280人（30.43%）  
生産年齢人口（15歳～64歳）：2,070,481人（57.27%）  
年少人口（0歳～14歳）：444,825人（12.3%）



老年人口（65歳以上）：1,143,423人（38.85%）  
生産年齢人口（15歳～64歳）：1,481,066人（50.33%）  
年少人口（0歳～14歳）：318,376人（10.82%）

# 健康寿命延伸のための取り組み例

はじまりを、つくる



- 健康日本21 (平成12年)
- 健康増進法の制定 (平成14年)
- スマート・ライフ・プロジェクト (平成23年)
  - 「健康寿命をのばしましょう」をスローガンにした国民運動
- 健康寿命延伸プラン(令和元年)
  - 2040年までに健康寿命を男女ともに3年以上延伸(2016年比)し、75歳以上とすることを目指す取り組み
- 静岡社会健康医学大学院大学(2021年開学)
  - 公衆衛生学と新領域の融合

## [テクノロジーの進歩]

### AI・データサイエンス領域と他分野との連携

- IoT環境の広がり：データ取得環境の向上(APIなど)
- クラウド環境の広がり：データ蓄積の容易さ・コスト削減
- 計算処理速度の向上：ビッグデータを分析可能に



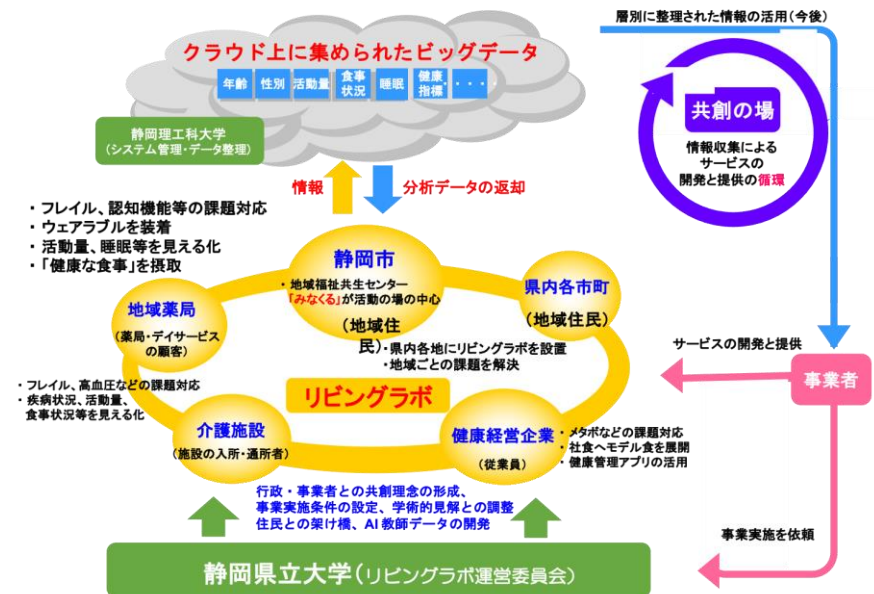
# 研究目的

- 今後の少子高齢化社会を支えるためには、健康寿命を延伸することが不可欠
- 進歩の著しいAI・データサイエンス領域と連携して、健康寿命を延伸するための手法を検討する。

## [実施事項]

1. ウェアラブル端末を用いた生活活動量を取得して、日常的な健康活動に対する要素を分析する
2. コロナ禍の社会的な変化と健康データとの関係を分析する

静岡県立大学を中心とするリビングラボの概要 — セルフケアの推進



# 使用データ



今回使用したデータは以下である

- コロナに関するデータ

- ・ オープンデータ | 厚生労働省 より 陽性者数

- <https://www.mhlw.go.jp/stf/covid-19/open-data.html>

- ウェアラブル端末から取得されるデータ

- ウェアラブル端末として、今回はFitbit Charge3で取得できるデータを利用した

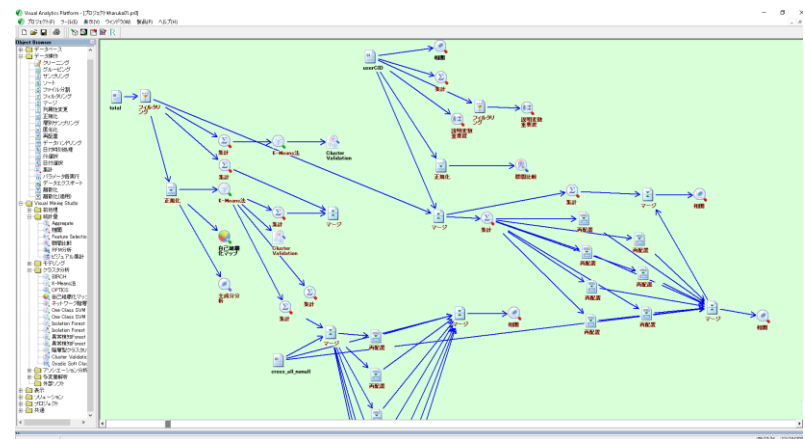
- 期間：2020/02/01~2020/10/23までの265日間

- 参加者数：79人

取得項目については次頁に記載

- 使用ソフト

- Visual Mining Studio



# 使用データ



## ● ウェアラブル端末から取得されるデータ

ウェアラブル端末として、今回はFitbit Charge3で取得できるデータを利用した  
使用データ： ユーザID、日付、15分ごとの心拍データ（本研究では使用せず）のうち記録されている件数

### 運動に関するデータ

activityCalories	運動での消費カロリー
caloriesBMR	基礎代謝
caloriesOut	消費カロリー
elevation	上方向の移動距離
fairlyActiveMinutes	アクティブな運動をしている時間（分）
floors	登った段数（elevationの階数表現）
lightlyActiveMinutes	軽い運動をしている時間（分）
marginalCalories	吸収されるカロリーの限界
restingHeartRate	安静時の脈拍
sedentaryMinutes	座っている時間（分）
steps	歩数
veryActiveMinutes	とてもアクティブな時間（分）

### 運動種類ごとのsummary(対象外、脂肪燃焼、有酸素運動、ピーク)

name	運動種類名
caloriesOut	消費カロリー
max	最高心拍
min	最低心拍
minutes	時間

### 睡眠に関するデータ

awakeCount	起床回数
awakeDuration	起床の持続時間(分)
awakeningsCount	起床回数(awakeCountと同じ)
duration	ベッドに入っていた時間（ミリ秒）
efficiency	睡眠効率( $\text{minutesAsleep} \div \text{timeInBed} \times 100$ )
endTime	起床時間
isMainSleep	その睡眠がメインの睡眠かどうか
logId	ログデータのID情報
minutesAfterWakeup	起きてから起き上がるまでの時間（分）
minutesAsleep	睡眠時間（分）
minutesAwake	ベッドで起きていた時間（分）
minutesToFallAsleep	眠りに落ちるまでの時間（分）
restlessCount	寝返り回数
restlessDuration	寝返りの持続時間(分)
startTime	睡眠開始時間
timeInBed	ベッドに入っていた時間（分）

重点項目として設定



# 分析の流れ

## ① ウェアラブル端末から取得できるすべてのデータでの分析

すべてのデータ間の関係性の把握

k-means法によりデータをグルーピング、特徴把握

## ② コロナによる影響の確認

コロナの感染者数のデータを用いて、コロナの影響があるか把握

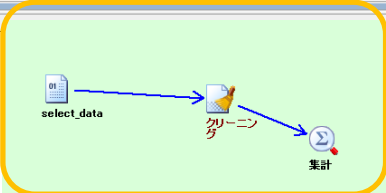
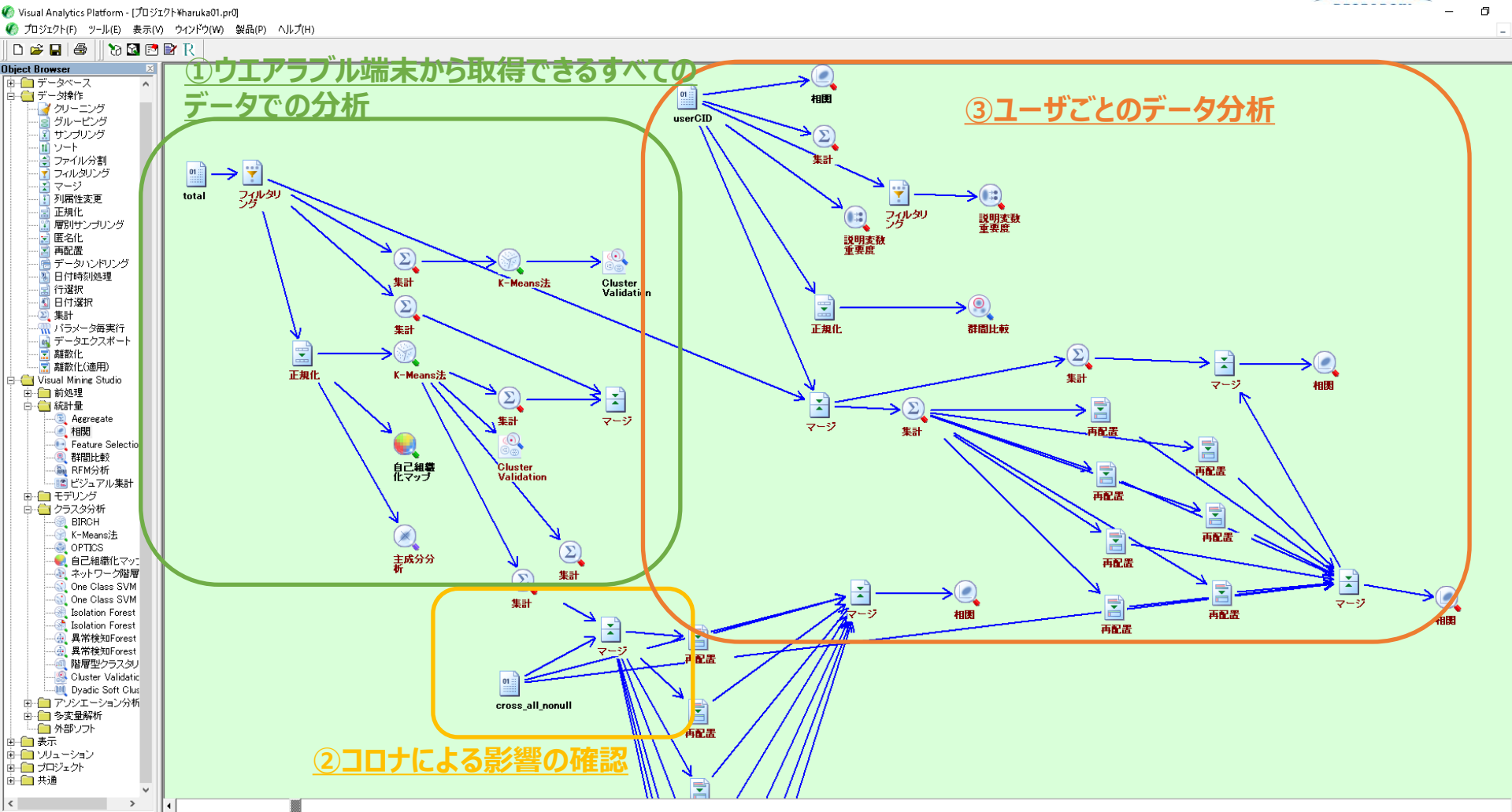
## ③ ユーザごとのデータ分析

ユーザごとに①のデータからクラスタ判定

クラスタ別にコロナ感染者数との影響把握

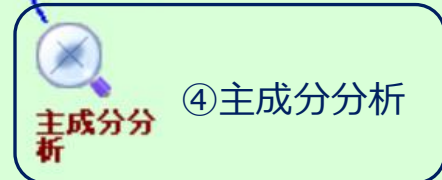
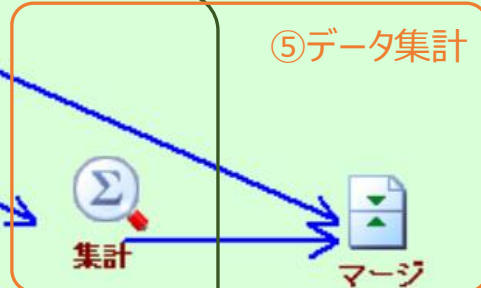
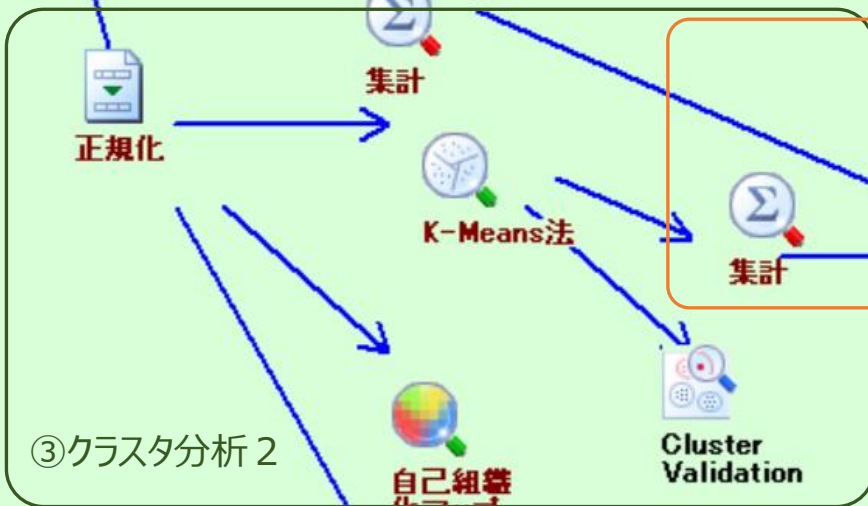
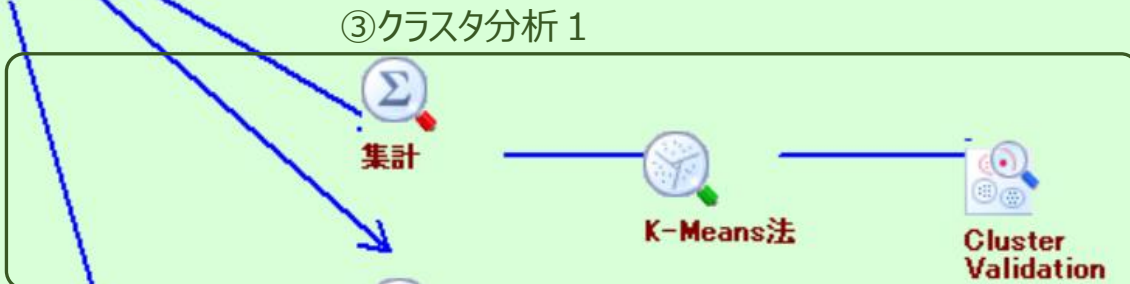
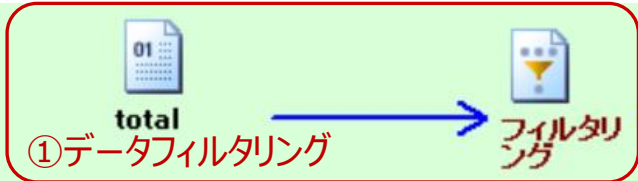
説明変数重要度分析

# 分析の流れ

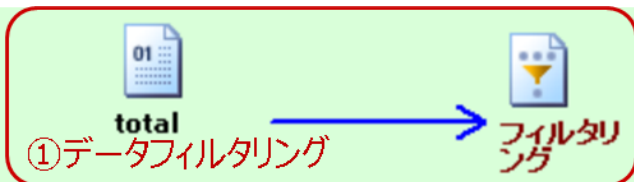


別プロジェクトで実行

# ①すべてのデータでの分析



# ①すべてのデータでの分析



全てのデータ（期間、ユーザ）には欠損値が含まれているため、

フィルタリングにより欠損値を削除する

全てのデータ（48,412件）

fitbitusers	dates	activeS	activity	calories	calories	elevation	fairlyAc	floors	lightlyA	margin	resting	sedent	steps	veryAct
24190	f034	2020-02--	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)
24191	f034	2020-02--	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)
24192	f034	2020-02--	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)
24193	f034	2020-02--	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)
24194	f034	2020-02--	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)
24195	f034	2020-02--	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)
24196	f034	2020-02--	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)	(NA)
24197	f017	2020-08--	-1.00	3.00	1498.00	1500.00	0.00	0.00	0.00	1.00	1.00	(NA)	1439.00	15.00
24198	f017	2020-07--	-1.00	0.00	1501.00	2005.00	20.00	0.00	2.00	0.00	0.00	(NA)	1440.00	3158.00
24199	f028	2020-08--	-1.00	7.00	1194.00	1375.00	0.00	0.00	0.00	3.00	4.00	(NA)	1437.00	30.00
24200	f028	2020-08--	-1.00	0.00	1193.00	1371.00	10.00	0.00	1.00	0.00	0.00	(NA)	1440.00	292.00
24201	f028	2020-08--	-1.00	2.00	1193.00	1370.00	0.00	0.00	0.00	1.00	1.00	(NA)	1439.00	60.00
24202	f028	2020-02--	-1.00	2.00	1197.00	1374.00	0.00	0.00	0.00	1.00	1.00	(NA)	1439.00	11.00
24203	f065	2020-08--	-1.00	2.00	1279.00	1280.00	0.00	0.00	0.00	1.00	1.00	(NA)	1439.00	5.00
24204	f104	2020-02--	-1.00	14.00	1193.00	1202.00	0.00	0.00	0.00	6.00	6.00	(NA)	1434.00	93.00
24205	f064	2020-05--	-1.00	3.00	1546.00	1550.00	0.00	0.00	0.00	1.00	1.00	(NA)	1439.00	19.00
24206	f115	2020-09--	-1.00	6.00	1092.00	1121.00	40.00	0.00	4.00	3.00	2.00	(NA)	1437.00	797.00
24207	f115	2020-07--	-1.00	6.00	1092.00	1130.00	50.00	0.00	5.00	3.00	3.00	(NA)	1437.00	610.00
24208	f061	2020-02--	-1.00	7.00	1252.00	1489.00	0.00	0.00	0.00	3.00	4.00	(NA)	1437.00	66.00
24209	f116	2020-03--	-1.00	2.00	959.00	960.00	0.00	0.00	0.00	1.00	0.00	(NA)	1439.00	10.00
24210	f053	2020-10--	-1.00	4.00	972.00	975.00	0.00	0.00	0.00	2.00	1.00	(NA)	1438.00	21.00
24211	f053	2020-08--	-1.00	6.00	972.00	977.00	0.00	0.00	0.00	3.00	3.00	(NA)	1437.00	76.00
24212	f053	2020-08--	-1.00	6.00	972.00	976.00	0.00	0.00	0.00	3.00	2.00	(NA)	1437.00	57.00
24213	f041	2020-10--	-1.00	0.00	1072.00	1076.00	0.00	0.00	0.00	0.00	0.00	(NA)	1440.00	57.00
24214	f108	2020-09--	-1.00	4.00	1197.00	1200.00	0.00	0.00	0.00	2.00	2.00	(NA)	1438.00	44.00
24215	f108	2020-07--	-1.00	6.00	1197.00	1201.00	10.00	0.00	1.00	2.00	3.00	(NA)	1438.00	61.00
24216	f108	2020-07--	-1.00	5.00	1197.00	1200.00	10.00	0.00	1.00	2.00	2.00	(NA)	1438.00	39.00
24217	f108	2020-06--	-1.00	2.00	1197.00	1198.00	0.00	0.00	0.00	1.00	1.00	(NA)	1439.00	15.00
24218	f017	2020-08--	-1.00	3.00	1498.00	1500.00	0.00	0.00	0.00	1.00	1.00	(NA)	1439.00	15.00
24219	f017	2020-07--	-1.00	0.00	1501.00	2005.00	20.00	0.00	2.00	0.00	0.00	(NA)	1440.00	3158.00
24220	f028	2020-08--	-1.00	7.00	1194.00	1375.00	0.00	0.00	0.00	3.00	4.00	(NA)	1437.00	30.00
24221	f028	2020-08--	-1.00	0.00	1193.00	1371.00	10.00	0.00	1.00	0.00	0.00	(NA)	1440.00	292.00
24222	f028	2020-08--	-1.00	2.00	1193.00	1370.00	0.00	0.00	0.00	1.00	1.00	(NA)	1439.00	60.00
24223	f028	2020-02--	-1.00	2.00	1197.00	1374.00	0.00	0.00	0.00	1.00	1.00	(NA)	1439.00	11.00
24224	f065	2020-08--	-1.00	2.00	1279.00	1280.00	0.00	0.00	0.00	1.00	1.00	(NA)	1439.00	5.00
24225	f104	2020-02--	-1.00	14.00	1193.00	1202.00	0.00	0.00	0.00	6.00	6.00	(NA)	1434.00	93.00

## フィルタリング設定

- 運動、睡眠に関するデータが含まれている
- 15分ごとのデータ（本研究では利用せず）のうち70%以上取得されている（データ有効性のため）もののみ抽出

## ④主成分分析

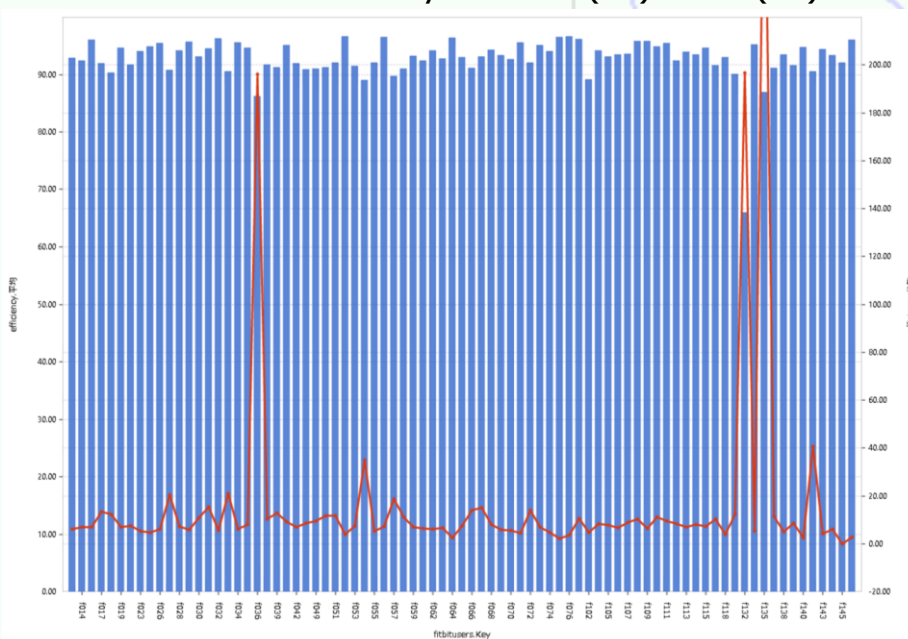
15,830件へ

# ①すべてのデータでの分析



集計：ユーザーごとと睡眠効率に関する情報を取得

睡眠効率 (efficiency) の平均(棒)と分散(線)

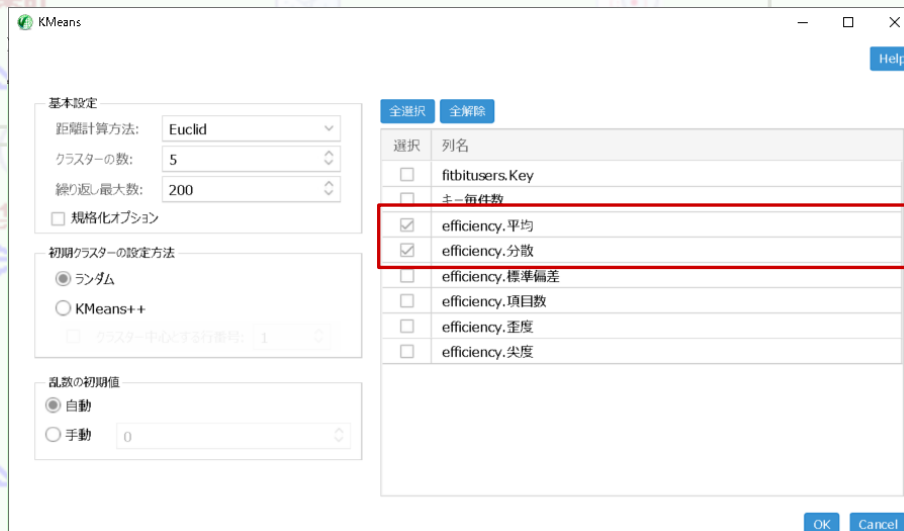


分散が大きい人がある

③クラスタ分析 1

K-means法：

ユーザーごとと睡眠効率に関する情報を利用する





# ①すべてのデータでの分析



K-means法 :

ユーザごと睡眠効率に関する情報を利用する

cluster info (5行/5列)

	id	size	efficiency.平均	efficiency.分散	residual
1	1	5	91.24	22.77	38.65
2	2	19	91.64	10.73	40.86
3	3	30	94.67	7.62	51.42
4	4	3	80.03	227.01	145.30
5	5	24	93.97	4.61	49.67

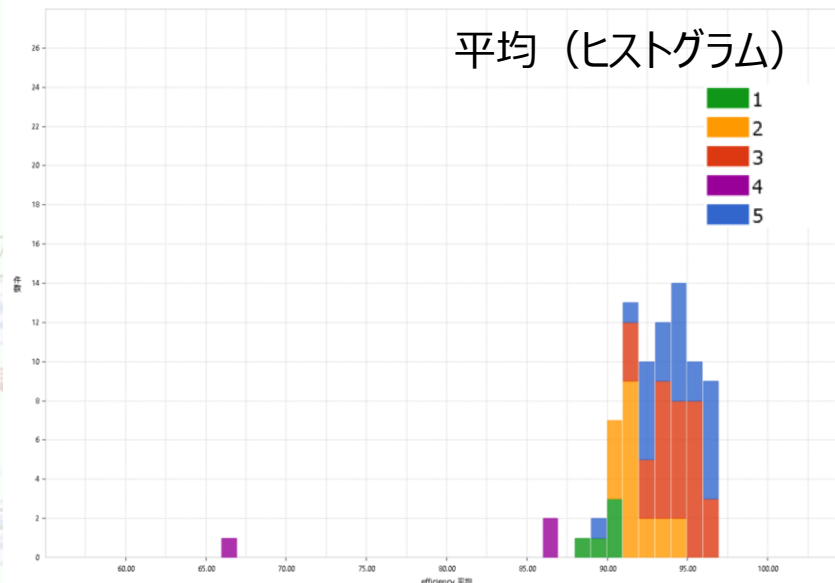
正規化

[特徴的なクラスタ]

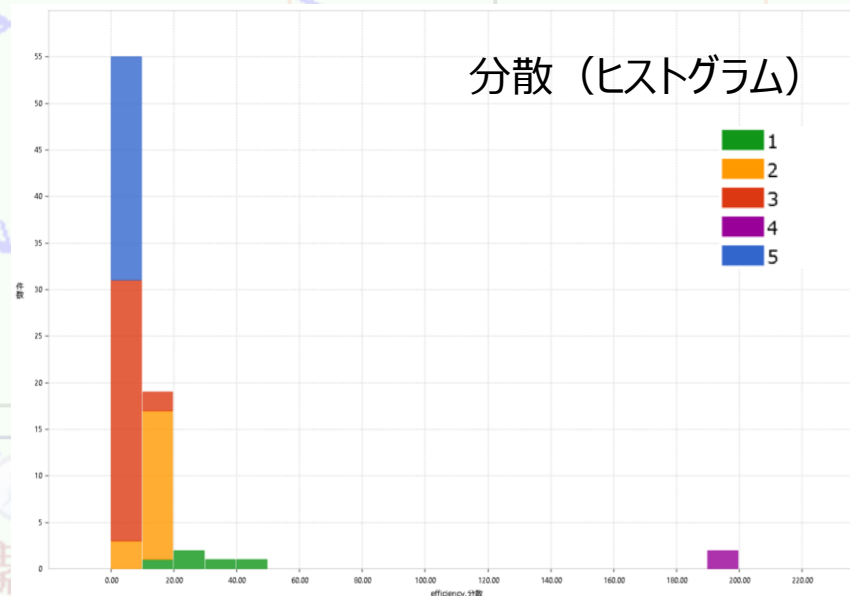
- 1 : 睡眠効率は低め、分散も大きい
- 2 : 睡眠効率はやや低め、分散は大きめ
- 3 : 睡眠効率は高い、やや安定的**
- 4 : 睡眠効率が低い (ばらつきがおおきい)
- 5 : 睡眠効率は高め、安定的**

3と5のクラスタの人は安定的に良い睡眠をとっている

平均 (ヒストグラム)

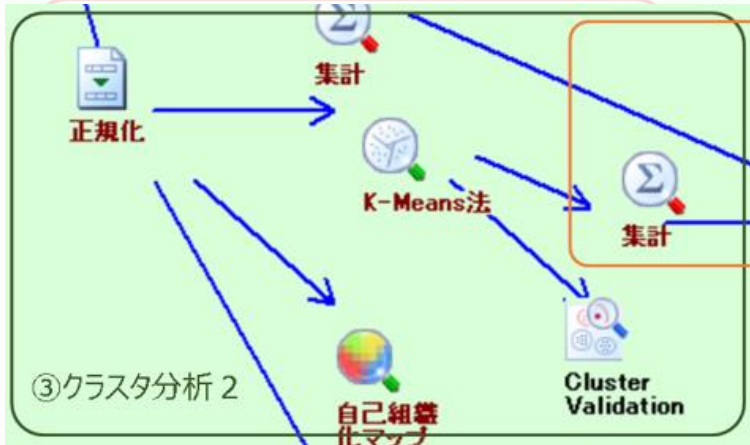


分散 (ヒストグラム)





# ①すべてのデータでの分析



正規化：K-means法を実施するにあたり正規化を行う  
全てのデータを標準化（標準偏差）

列名	規格化方法
fitbitusers	
dates	
activeScore	標準化(標準偏差)
activityCalories	標準化(標準偏差)
caloriesBMR	標準化(標準偏差)
caloriesOut	標準化(標準偏差)
elevation	標準化(標準偏差)
fairlyActiveMinute:	標準化(標準偏差)
floors	標準化(標準偏差)
lightlyActiveMinute:	標準化(標準偏差)
marginalCalories	標準化(標準偏差)

K-means法：  
全てのデータの重点項目に関する情報を利用する

**KMeans**

基本設定

距離計算方法: Euclid

クラスターの数: 7

繰り返し最大数: 100

規格化オプション

初期クラスターの設定方法

ランダム

KMeans++

乱数の初期値

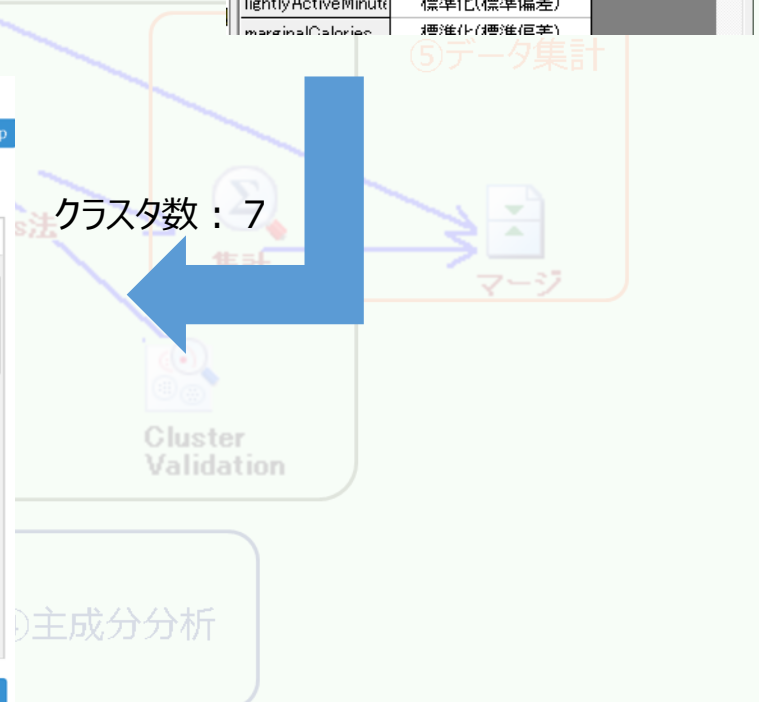
自動

手動 0

全選択 全解除

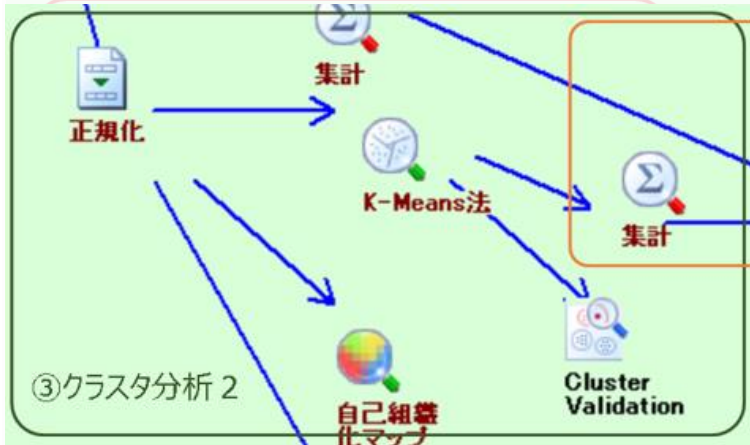
選択	列名
<input checked="" type="checkbox"/>	steps
<input checked="" type="checkbox"/>	Out of range_minutes
<input checked="" type="checkbox"/>	Fat burning_minutes
<input checked="" type="checkbox"/>	aerobic exercise_minutes
<input checked="" type="checkbox"/>	peak_minutes
<input checked="" type="checkbox"/>	duration
<input checked="" type="checkbox"/>	efficiency
<input type="checkbox"/>	fitbitusers
<input type="checkbox"/>	dates
<input type="checkbox"/>	activeScore
<input type="checkbox"/>	activityCalories
<input type="checkbox"/>	caloriesBMR
<input type="checkbox"/>	caloriesOut

OK Cancel





# ①すべてのデータでの分析

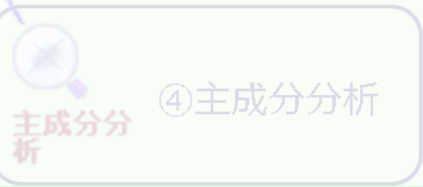


## K-means法 :

- ID1 : 良く運動している (有酸素運動) が睡眠時間が短い
- ID2 : 良く歩いており、特に心拍数が高い運動をする
- ID3 : あまり運動をしておらず睡眠時間は短い、睡眠スコアは良い
- ID4 : あまり歩いていないが、有酸素運動を特にしており、睡眠時間は短め
- ID5 : 歩いてはいないが有酸素運動を行い、睡眠スコアもよい
- ID6 : 運動は少なめだが睡眠時間が長い
- ID7 : 歩いており睡眠効率も良い

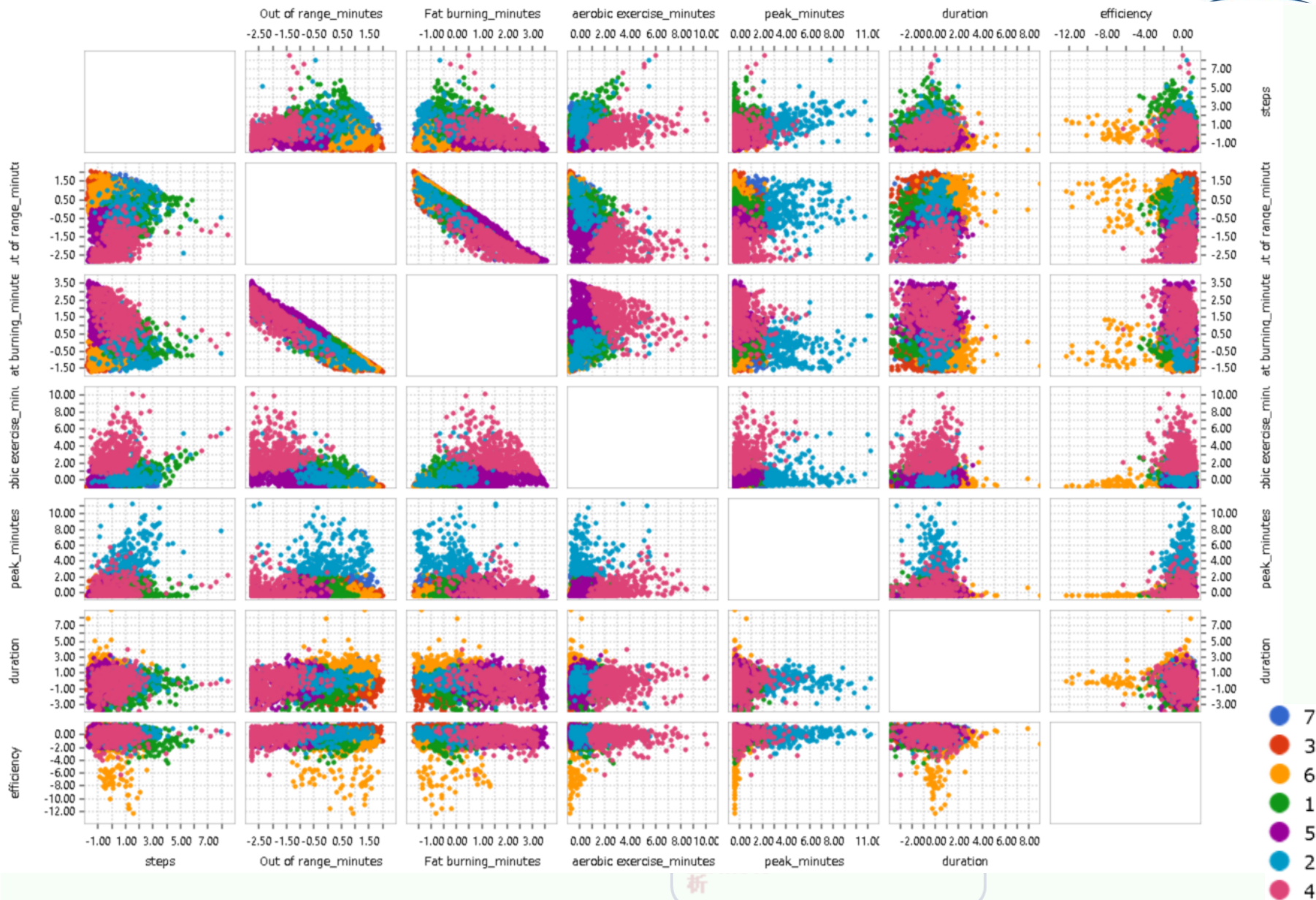
cluster info (7行/10列)

	id	size	steps	Out of ...	Fat bur...	aerobic...	peak_...	duration	efficiency	residual
1	1	2888	0.92	-0.37	0.30	0.40	0.11	-0.47	-0.14	5133.30
2	2	510	1.32	-0.01	-0.27	0.53	4.35	0.09	-0.01	1365.00
3	3	3906	-0.75	0.66	-0.67	-0.54	-0.32	-0.23	0.35	5008.83
4	4	814	0.30	-1.97	1.66	2.92	0.30	-0.29	-0.22	2237.37
5	5	3110	-0.56	-0.98	1.14	-0.07	-0.22	0.04	0.17	4939.50
6	6	1926	-0.38	0.33	-0.33	-0.27	-0.22	0.99	-1.04	3560.70
7	7	2676	0.67	0.99	-0.93	-0.42	-0.17	0.13	0.32	3532.52

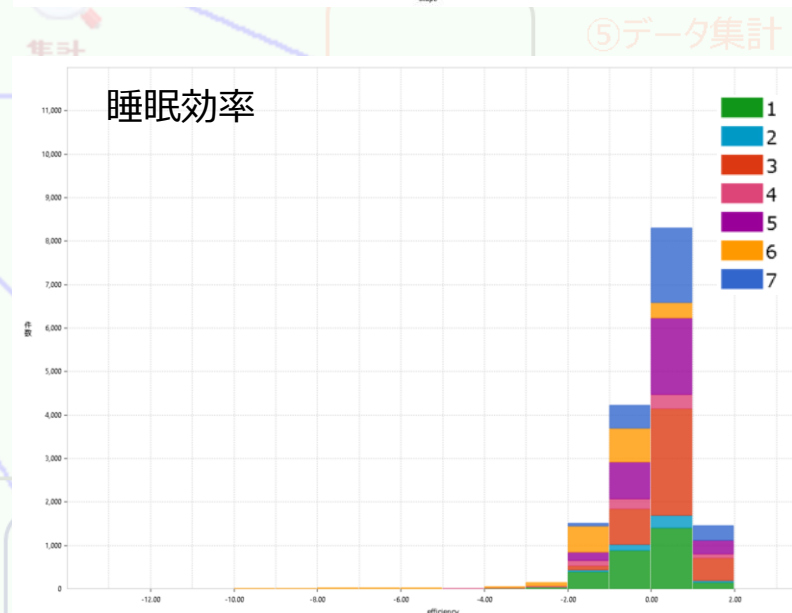
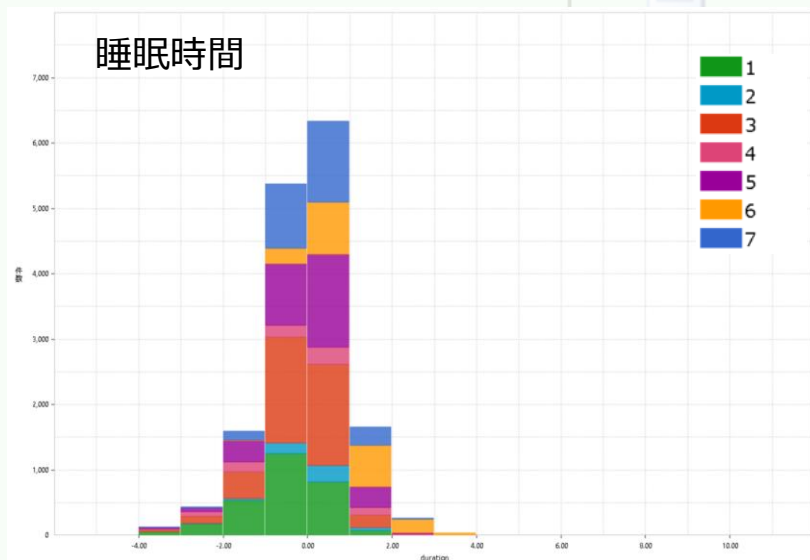
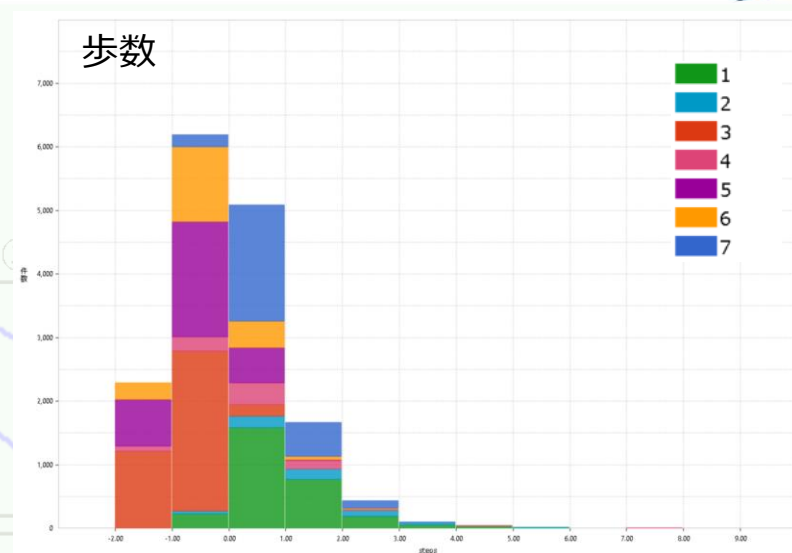
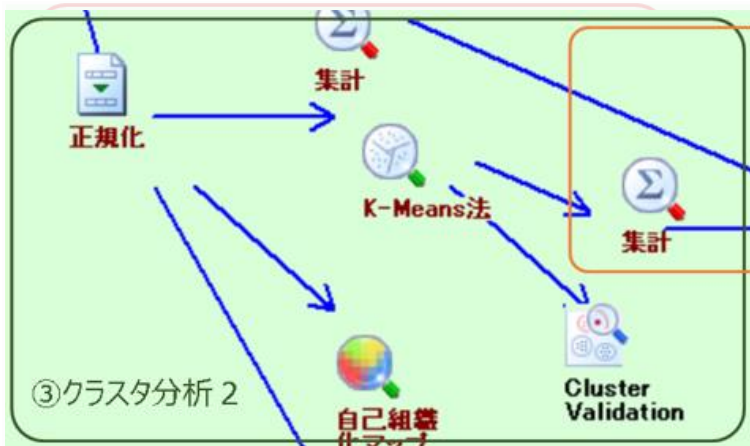




# ①すべてのデータでの分析



# ①すべてのデータでの分析



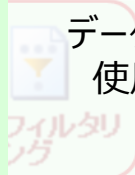
主成分分析

# ①すべてのデータでの分析



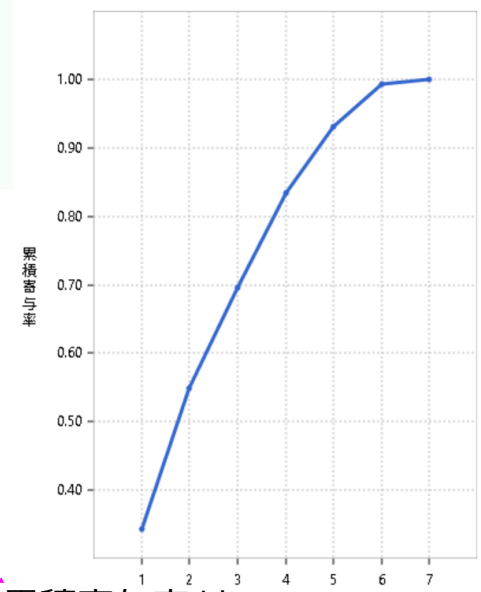
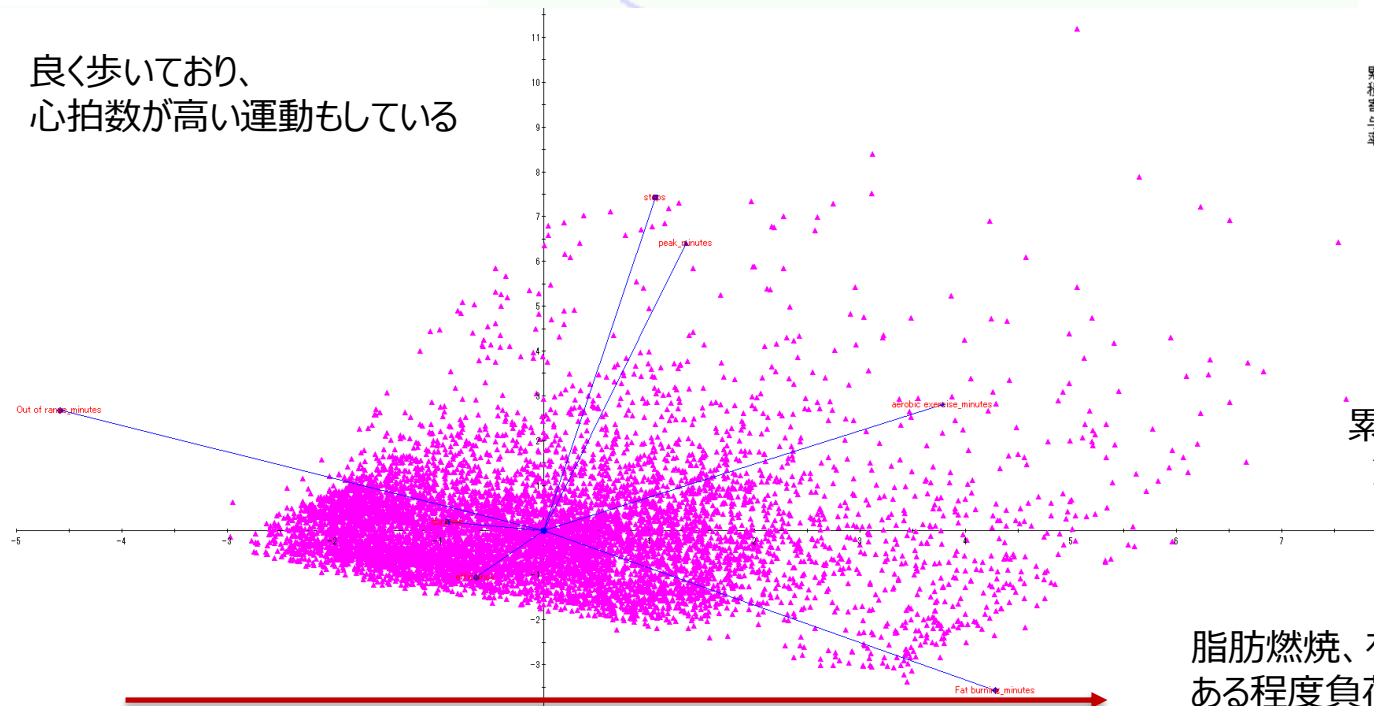
主成分分析

## ④主成分分析



データ項目間の関係性を確認する  
使用データは重点項目データとする

良く歩いており、  
心拍数が高い運動もしている



累積寄与率より  
第四主成分まで見る必要がある



脂肪燃焼、有酸素運動など  
ある程度負荷のかかる運動をしている

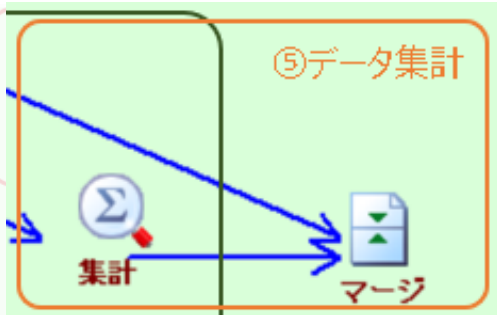
負荷量 (7行/8列)

	列名	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分	第6主成分	第7主成分
1	steps	0.14	0.66	-0.07	0.16	0.54	0.47	0.04
2	Out of range_minutes	-0.60	0.24	-0.01	0.05	0.04	-0.16	-0.74
3	Fat burning_minutes	0.56	-0.32	0.02	-0.05	-0.02	0.39	-0.65
4	aerobic exercise_minutes	0.50	0.25	0.07	-0.10	0.27	-0.76	-0.15
5	peak_minutes	0.18	0.57	-0.07	-0.28	-0.74	0.08	-0.04
6	duration	-0.12	0.02	0.70	-0.67	0.18	0.14	0.03
7	efficiency	-0.08	-0.09	-0.71	-0.66	0.22	0.02	0.01



# ①すべてのデータでの分析

全てのデータを利用したK-means法でできたフラグから、ユーザのクラスタを決定するために、データ集約を行う



## ③クラスタ分析 1

ユーザIDごとに、クラスタIDの個数をカウント

result (81行/8列)	fitbitusers.Key	ClusterID.1 ↑	ClusterID.2 ↑	ClusterID.3 ↑	ClusterID.4	ClusterID.5
1	f145	0	0	0	0	0
2	f118	0	0	0	40	0
3	f023	0	0	0	0	0
4	f075	0	0	2	0	0
5	f051	0	0	2	0	0
6	f028	0	0	2	0	0
7	f102	0	0	4	0	0
8	f070	0	0	4	0	0
9	f031	0	0	6	0	0
10	f032	0	0	10	0	0
11	f135	0	0	12	0	0
12	f138	0	0	14	0	0
13	f144	0	0	16	0	0
14	f034	0	0	22	0	0
15	f140	0	0	46	0	0
16	f030	0	0	50	0	0
17	f056	0	0	76	0	0
18	f053	0	0	96	0	0
19	f068	0	0	160	0	12

ユーザIDごとに、すべての項目の平均値を計算

集計

集計方法: 項目カウンタ

結果形式:  リスト  マトリクス

集計キー列名: fitbitusers<文字>

集計対象列名: ClusterID<文字>

集計

集計方法: サマリ

サマリ項目:  平均

集計キー列名: fitbitusers<文字>

集計対象列名: Fat burning\_minutes<実数>, aerobic\_exercise\_calorie<実数>, etc.

## ③クラスタ分析 2

マージして保存

マージ

マージ方法: 列

オプション: 単一参照

主テーブル選択:  集計result(1)

キー列指定 (重複=セル選択+削除ボタン):

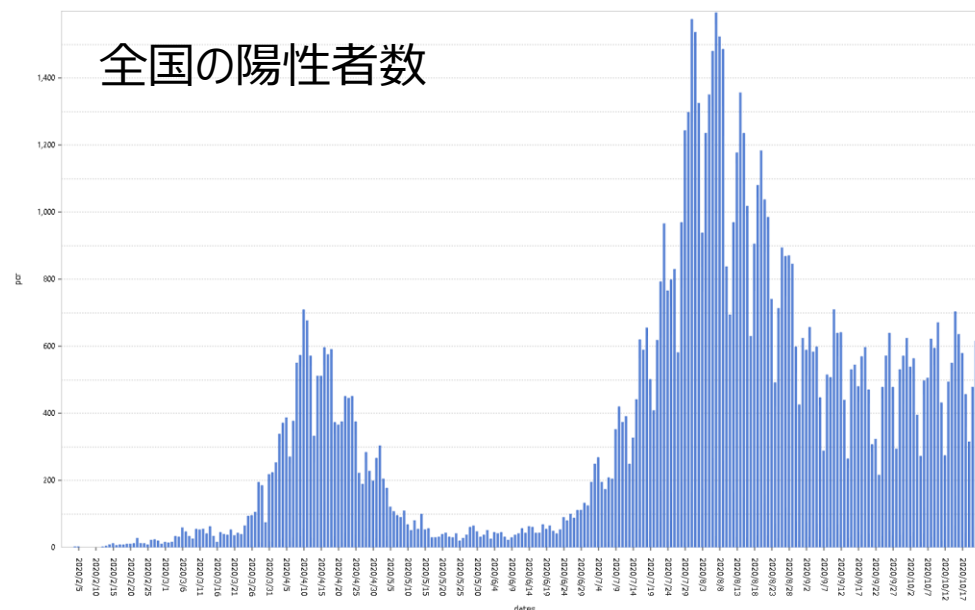
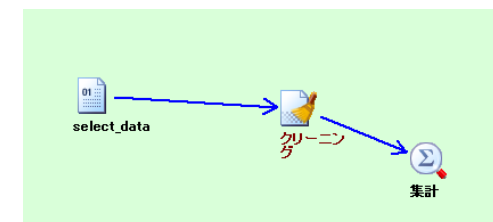
集計result(1)	集計result(2)
fitbitusers.Key	fitbitusers.Key

## ②コロナによる影響の確認



### 新型コロナウイルスの影響期間

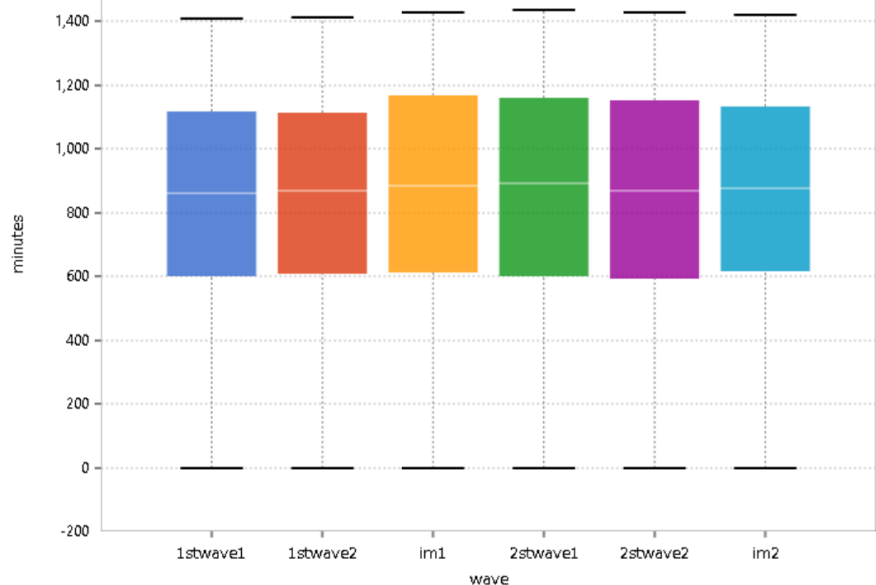
- 第一波(前半：～4/15まで、後半：～5/15まで)
  - 学校休校措置(2/28～5/10)
  - 非常事態宣言(4/16～5/25)
- 中間期1(5/16～6/30まで)
- 第二波(前半：7/1～7/31、後半：8/1～8/31)
  - 7/22 Go to Travel開始(東京都を除く)
- 中間期2(9/1～10/31)
  - 10/1 Go to Eat開始
  - 10/1 Go to Travel東京都追加
- 第三波(11/1～現在)



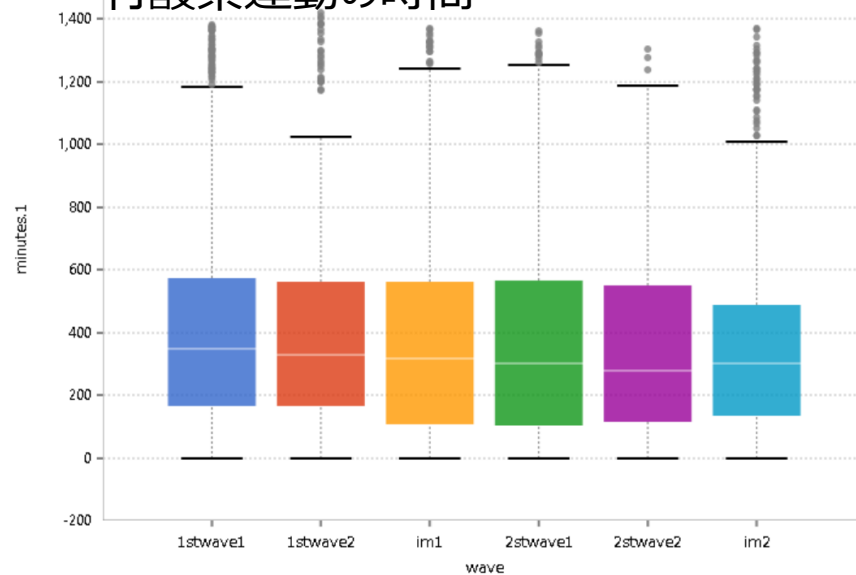
# ② コロナによる影響の確認



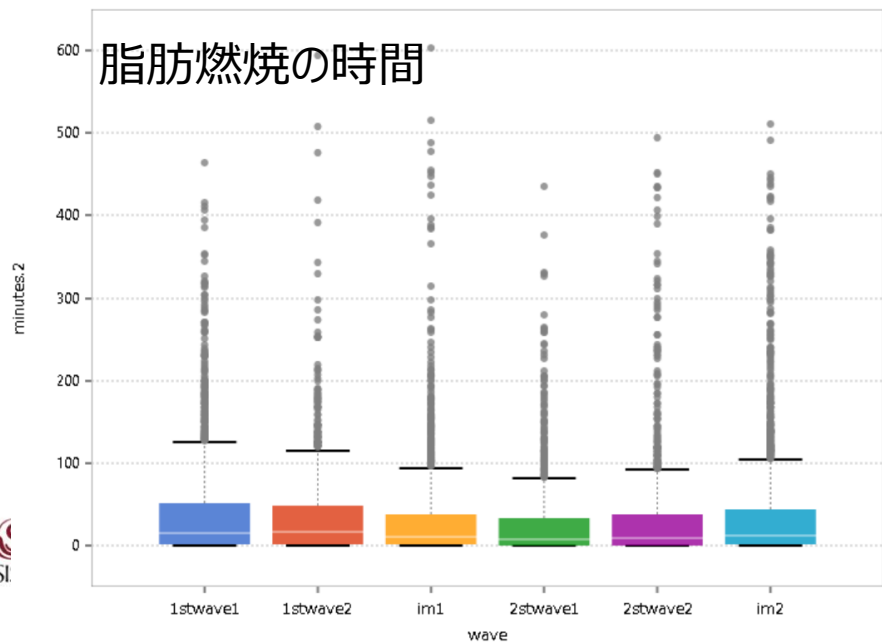
## 運動対象外の時間



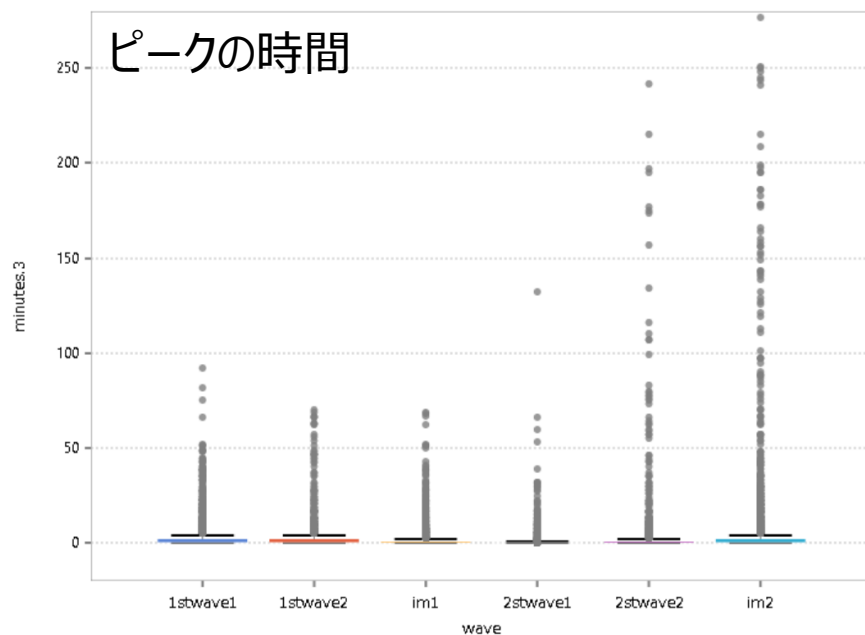
## 有酸素運動の時間



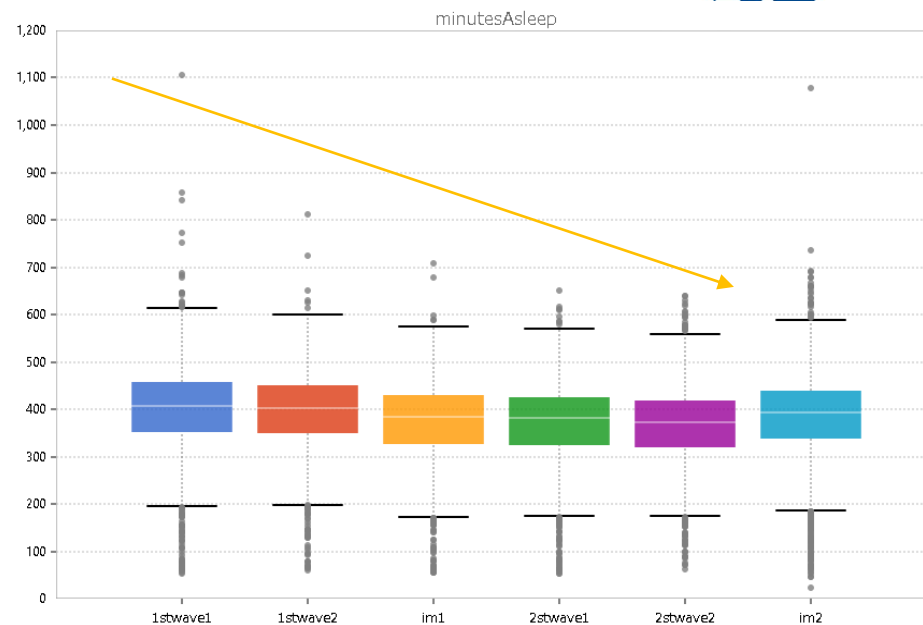
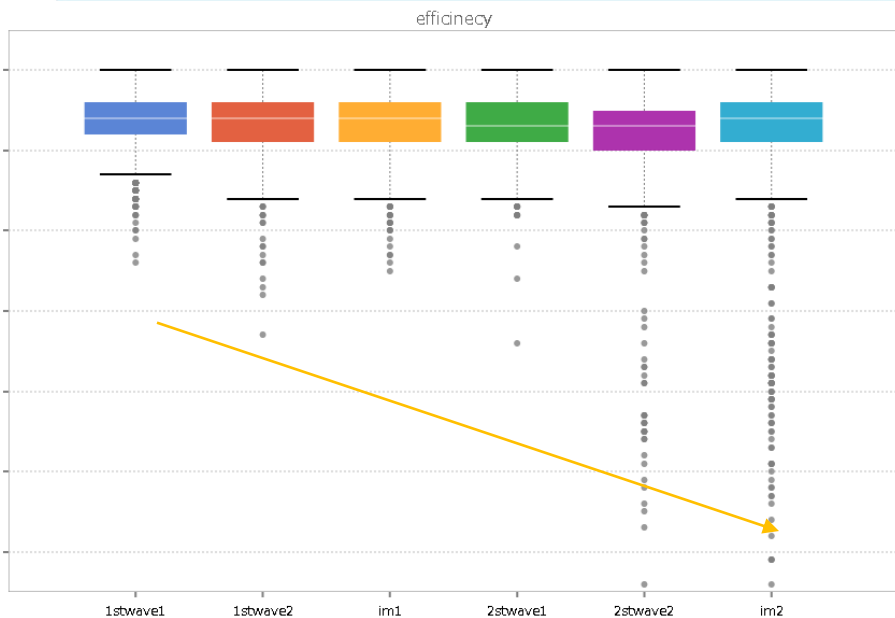
## 脂肪燃焼の時間



## ピークの時間



# ② コロナによる影響の確認



## 睡眠効率

睡眠効率が悪い人が徐々に増えている傾向

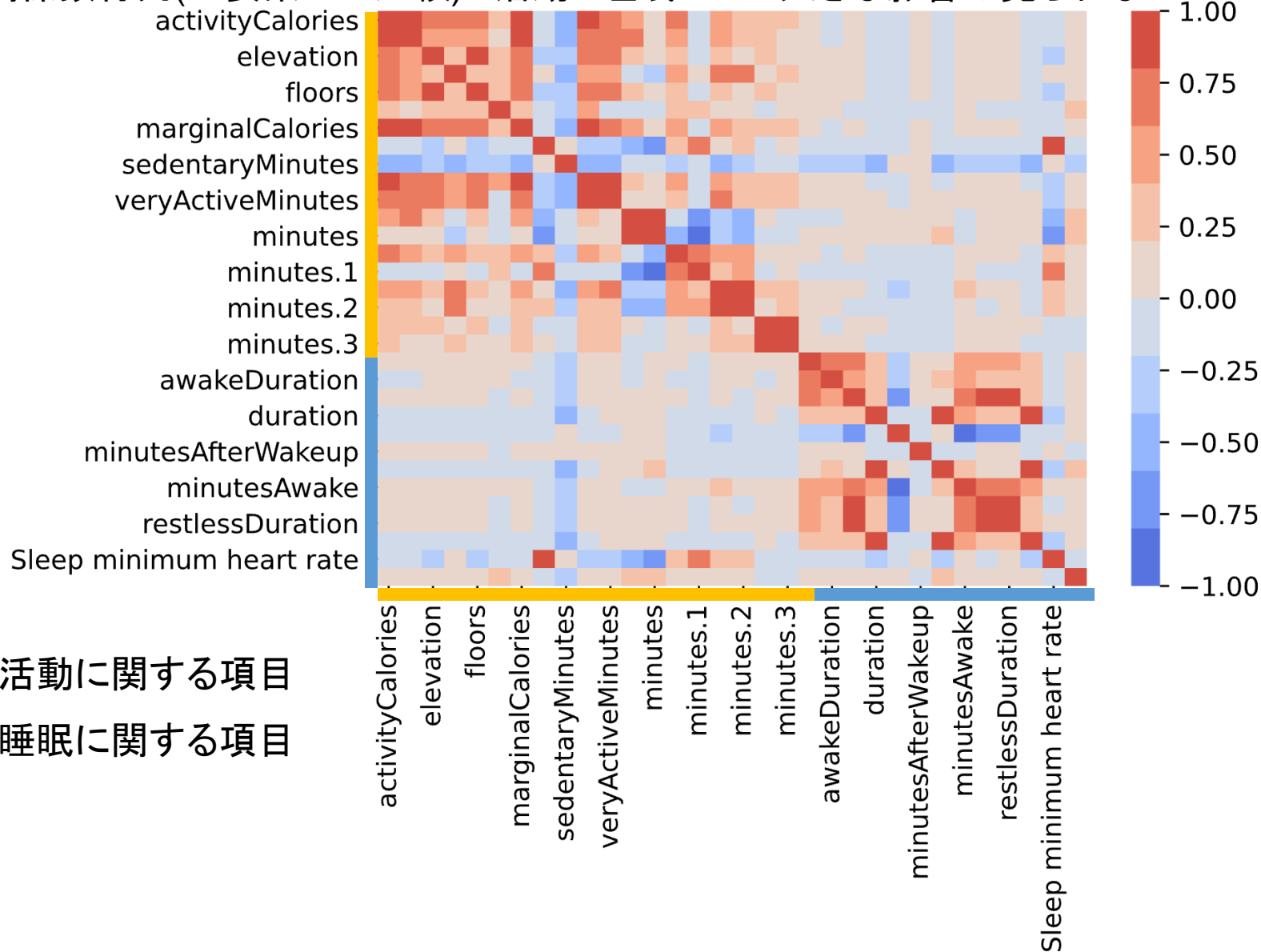
## 睡眠時間

睡眠時間は短くなっている傾向

## ②コロナによる影響の確認



相関係数行列(全要素での比較) : 活動と睡眠とでの大きな影響は見られない







# ③ ユーザごとのデータ分析

①で作成した  
ユーザーごとにクラスタ番号を  
付与したデータ



userCID



相関



集計

集計



説明変数  
重要度

項目間の関係性の把握



正規化



群間比較

コロナの影響度確認

コロナの全期間のデータ  
(欠損値削除済み)



total



マージ



集計

データ整形



再配置



再配置



再配置



再配置



再配置



再配置



再配置



再配置



再配置



マージ



相関

コロナの陽性者数  
(全国、静岡県)



cross\_all\_nonnull

# ③ ユーザごとのデータ分析



今回使用するデータは、ユーザごとにクラスタIDを付与したデータ  
 全体でのデータで出たクラスタIDの中で一番多く当てはまっていたクラスタIDを付与している

①で作成した  
 ユーザごとにクラスタ番号を  
 付与したデータ



table (81行/56列)	Cluster...	Cluster...	Cluster...	Cluster...	Cluster...	Cluster...	Cluster...	Cluster...	Cluster...	MaxCID	activeS...	activity...	calories...	calories...	elevation	fairlyAc...
1	3	6	2	130	8	144	156	4	ClusterID.6	-1	749.74	1294.44	1952.12	40.44	9.26	
2	4	106	4	142	0	16	30	0	ClusterID.3	-1	723.37	896.00	1453.68	68.52	9.70	
3	6	4	0	12	0	0	0	146	ClusterID.7	-1	1732.64	1583.94	3051.06	279.14	34.16	
4	7	70	2	50	2	96	84	20	ClusterID.5	-1	1008.57	1518.02	2411.87	81.60	8.69	
5	8	108	76	56	0	22	216	38	ClusterID.6	-1	885.92	927.00	1609.28	84.50	9.83	
6	9	294	78	26	22	26	6	64	ClusterID.1	-1	1192.17	957.66	1890.19	231.67	27.29	
7	11	124	12	102	10	126	112	2	ClusterID.5	-1	686.59	871.29	1412.81	76.80	21.99	
8	3	0	0	0	0	0	0	30	ClusterID.7	-1	1850.07	1586.00	3123.00	517.33	36.13	
9	5	0	2	8	10	0	0	50	ClusterID.7	-1	1198.89	976.83	1892.57	56.57	39.71	
10	6	0	2	118	0	4	18	350	ClusterID.7	-1	583.94	814.30	1270.97	156.10	7.43	
11	7	0	2	0	202	196	0	0	ClusterID.4	-1	468.63	927.42	1322.99	23.55	6.99	
12	8	0	0	2	0	0	0	12	ClusterID.7	-1	792.43	1130.00	1746.86	115.71	16.86	
13	9	6	4	0	22	406	0	0	ClusterID.5	-1	411.16	730.87	1065.39	75.84	5.02	
14	10	0	0	50	0	0	36	4	ClusterID.3	-1	452.16	1275.00	1674.16	37.33	11.36	
15	11	0	0	6	0	0	0	2	ClusterID.3	-1	1051.50	1601.00	2494.50	20.00	3.50	
16	12	0	0	10	0	0	0	36	ClusterID.7	-1	898.13	1130.61	1831.65	69.57	20.39	
17	13	12	0	26	0	2	9	0	ClusterID.6	-1	570.74	919.57	1374.66	105.43	14.40	
18	14	0	0	22	0	0	0	98	ClusterID.7	-1	1118.05	1573.37	2471.65	266.83	9.36	
19	15	40	8	16	0	2	16	70	ClusterID.7	-1	821.51	883.53	1502.82	226.71	13.00	
20	16	8	0	38	0	94	52	0	ClusterID.5	-1	877.35	1089.00	1769.81	151.46	7.13	

クラスタID

確認

そのほかの要素は平均値



集計

期間のデータ  
 (削除済み)



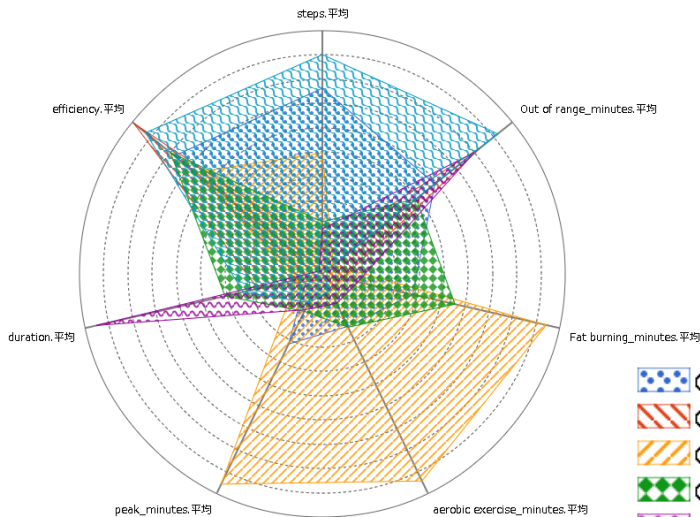
total

クラスタごとに平均値計算

データ整形

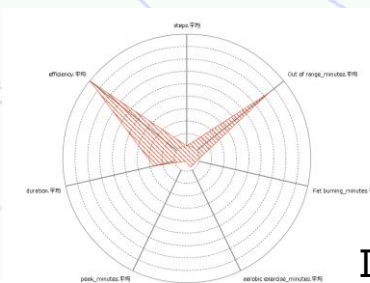
[クラスタ特徴](再掲)

- ID1 : 良く運動している (有酸素運動) が睡眠時間が短い
- ID3 : あまり運動をしておらず睡眠時間は短い、睡眠効率は良い
- ID4 : あまり歩いていないが、有酸素運動を特にしており、睡眠時間は短め
- ID5 : 歩いてはいないが有酸素運動を行い、睡眠効率もよい
- ID6 : 運動は少なめだが睡眠時間が長い
- ID7 : 歩いており睡眠効率も良い



- ClusterID.1
- ClusterID.3
- ClusterID.4
- ClusterID.5
- ClusterID.6
- ClusterID.7

nonull



ID3のみ

再配置

再配置

再配置

再配置

再配置

再配置

再配置

再配置

再配置

相関





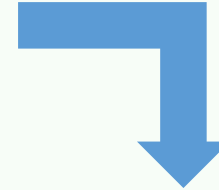
# ③ ユーザごとのデータ分析

データ整形により次のようなデータが生成された

日付、クラスIDごとの重要項目の平均値

result	dates	pcr	shizuoka	steps...	steps...	steps...	steps...	steps...	steps...	steps...	Out of ...	Out of ...	Out of ...	Out of ...	Out of ...
1	2020/2/1	0	0	8457.00	6356.00	11050.00	9262.33	0.00	12563.33	802.00	902.33	201.00	606.67	0.00	109.00
2	2020/2/2	0	0	14782.00	5998.00	0.00	8487.14	12950.67	10055.50	580.50	928.00	0.00	693.86	889.00	119.00
3	2020/2/3	0	0	10743.00	6125.33	16450.00	8995.71	9843.00	12735.33	726.00	821.33	316.00	635.14	764.33	111.00
4	2020/2/4	2	0	17904.00	4759.50	13961.00	7646.86	11230.50	11441.67	680.00	992.00	302.00	658.43	745.50	107.00
5	2020/2/5	2	0	16977.50	5456.33	9395.00	8148.43	11593.50	8187.00	590.50	962.00	405.00	637.71	781.50	110.00
6	2020/2/6	0	0	12865.00	4951.75	6821.00	8633.57	8996.67	13251.00	765.50	878.75	158.50	626.86	783.67	115.00
7	2020/2/7	0	0	14529.00	5219.60	12588.00	8279.71	9508.33	14514.75	760.00	1020.40	136.50	588.43	849.67	109.00
8	2020/2/8	0	0	11957.00	5533.67	8274.50	9121.57	7728.00	16667.33	692.00	949.00	155.00	606.57	691.67	116.00
9	2020/2/9	0	0	7714.50	4817.25	16611.50	6822.43	8372.33	11984.67	614.50	949.50	671.14	626.00	929.00	121.00
10	2020/2/10	0	0	11099.00	7297.80	8865.50	7494.83	8746.67	12210.40	660.50	976.20	147.50	626.00	904.00	113.00
11	2020/2/11	0	0	25341.00	5948.80	12048.00	11379.33	11289.67	9614.33	507.00	856.80	391.00	658.83	710.33	123.00
12	2020/2/12	1	0	11794.00	4847.50	7391.00	9105.14	12108.50	11585.75	752.00	758.00	198.00	589.71	858.00	112.00
13	2020/2/13	4	0	11402.50	7033.00	6956.50	7886.60	8444.00	9551.50	575.00	873.67	50.00	495.40	899.00	112.00
14	2020/2/14	7	0	12723.00	7279.50	8838.00	9345.00	8062.33	10083.67	622.50	1076.50	100.50	626.43	833.33	116.00
15	2020/2/15	12	0	9691.00	5212.25	7818.50	9182.57	9296.00	10726.00	695.00	902.75	52.50	726.86	645.33	113.00
16	2020/2/16	6	0	7814.50	4958.50	5714.50	4992.17	6897.50	12069.00	697.00	919.50	211.00	675.17	1035.00	122.00
17	2020/2/17	7	0	11784.50	5038.50	12627.00	8447.83	8071.50	12224.67	754.00	919.25	296.00	696.17	873.00	112.00
18	2020/2/18	7	0	13462.00	6753.17	6758.00	8589.43	9532.00	10842.50	694.00	1031.83	88.50	606.29	727.00	113.00
19	2020/2/19	10	0	12315.00	6037.50	9541.50	6263.83	11644.00	10635.50	567.00	975.75	209.00	698.00	759.67	114.00
20	2020/2/20	9	0	9518.50	6149.40	9011.50	9883.00	10277.00	13367.00	507.00	989.60	192.50	608.50	812.67	113.00
21	2020/2/21	11	0	13835.00	6724.43	8632.00	7950.50	8387.67	11588.00	474.00	956.00	175.00	669.83	716.33	109.00
22	2020/2/22	27	0	12813.50	5524.17	10172.50	6539.50	6428.00	11233.00	492.00	1042.83	173.50	723.67	827.33	107.00
23	2020/2/23	12	0	9788.00	7200.89	6202.00	8886.40	7443.33	14004.40	712.50	999.89	180.00	646.60	960.00	114.00
24	2020/2/24	13	0	13068.00	7580.89	6736.40	6087.75	8685.67	13181.75	620.50	865.67	163.50	505.00	926.00	117.00

このデータをもとに  
相関分析を行う



相関

計算

- 相関係数
- 独立係数
- 相関比
- X2値
- F値

出力形式

- マトリックス形式
- リスト形式

対象列選択

列名	Type	変数1	変数2
dates	カテゴリ		
pcr	数値	X	X
shizuoka	数値	X	X
steps平均1	数値	X	X
steps平均3	数値	X	X
steps平均4	数値	X	X
steps平均5	数値	X	X
steps平均6	数値	X	X
steps平均7	数値	X	X
Out of range_minuties平均1	数値	X	X
Out of range_minuties平均3	数値	X	X
Out of range_minuties平均4	数値	X	X
Out of range_minuties平均5	数値	X	X
Out of range_minuties平均6	数値	X	X
Out of range_minuties平均7	数値	X	X
Fat burning_minuties平均1	数値	X	X
Fat burning_minuties平均3	数値	X	X
Fat burning_minuties平均4	数値	X	X
Fat burning_minuties平均5	数値	X	X
Fat burning_minuties平均6	数値	X	X
Fat burning_minuties平均7	数値	X	X

OK Cancel

result (266行/45列)

result	dates	pcr	shizuoka	steps...	steps...	steps...	steps...	steps...	steps...	Out of ...	Out of ...	Out of ...	Out of ...	Out of ...	Out of ...
1	2020/2/1	0	0	8457.00	6356.00	11050.00	9262.33	0.00	12563.33	802.00	902.33	201.00	606.67	0.00	109.00
2	2020/2/2	0	0	14782.00	5998.00	0.00	8487.14	12950.67	10055.50	580.50	928.00	0.00	693.86	889.00	119.00
3	2020/2/3	0	0	10743.00	6125.33	16450.00	8995.71	9843.00	12735.33	726.00	821.33	316.00	635.14	764.33	111.00
4	2020/2/4	2	0	17904.00	4759.50	13961.00	7646.86	11230.50	11441.67	680.00	992.00	302.00	658.43	745.50	107.00
5	2020/2/5	2	0	16977.50	5456.33	9395.00	8148.43	11593.50	8187.00	590.50	962.00	405.00	637.71	781.50	110.00
6	2020/2/6	0	0	12865.00	4951.75	6821.00	8633.57	8996.67	13251.00	765.50	878.75	158.50	626.86	783.67	115.00
7	2020/2/7	0	0	14529.00	5219.60	12588.00	8279.71	9508.33	14514.75	760.00	1020.40	136.50	588.43	849.67	109.00
8	2020/2/8	0	0	11957.00	5533.67	8274.50	9121.57	7728.00	16667.33	692.00	949.00	155.00	606.57	691.67	116.00
9	2020/2/9	0	0	7714.50	4817.25	16611.50	6822.43	8372.33	11984.67	614.50	949.50	671.14	626.00	929.00	121.00
10	2020/2/10	0	0	11099.00	7297.80	8865.50	7494.83	8746.67	12210.40	660.50	976.20	147.50	626.00	904.00	113.00
11	2020/2/11	0	0	25341.00	5948.80	12048.00	11379.33	11289.67	9614.33	507.00	856.80	391.00	658.83	710.33	123.00
12	2020/2/12	1	0	11794.00	4847.50	7391.00	9105.14	12108.50	11585.75	752.00	758.00	198.00	589.71	858.00	112.00
13	2020/2/13	4	0	11402.50	7033.00	6956.50	7886.60	8444.00	9551.50	575.00	873.67	50.00	495.40	899.00	112.00
14	2020/2/14	7	0	12723.00	7279.50	8838.00	9345.00	8062.33	10083.67	622.50	1076.50	100.50	626.43	833.33	116.00
15	2020/2/15	12	0	9691.00	5212.25	7818.50	9182.57	9296.00	10726.00	695.00	902.75	52.50	726.86	645.33	113.00
16	2020/2/16	6	0	7814.50	4958.50	5714.50	4992.17	6897.50	12069.00	697.00	919.50	211.00	675.17	1035.00	122.00
17	2020/2/17	7	0	11784.50	5038.50	12627.00	8447.83	8071.50	12224.67	754.00	919.25	296.00	696.17	873.00	112.00
18	2020/2/18	7	0	13462.00	6753.17	6758.00	8589.43	9532.00	10842.50	694.00	1031.83	88.50	606.29	727.00	113.00
19	2020/2/19	10	0	12315.00	6037.50	9541.50	6263.83	11644.00	10635.50	567.00	975.75	209.00	698.00	759.67	114.00
20	2020/2/20	9	0	9518.50	6149.40	9011.50	9883.00	10277.00	13367.00	507.00	989.60	192.50	608.50	812.67	113.00
21	2020/2/21	11	0	13835.00	6724.43	8632.00	7950.50	8387.67	11588.00	474.00	956.00	175.00	669.83	716.33	109.00
22	2020/2/22	27	0	12813.50	5524.17	10172.50	6539.50	6428.00	11233.00	492.00	1042.83	173.50	723.67	827.33	107.00
23	2020/2/23	12	0	9788.00	7200.89	6202.00	8886.40	7443.33	14004.40	712.50	999.89	180.00	646.60	960.00	114.00
24	2020/2/24	13	0	13068.00	7580.89	6736.40	6087.75	8685.67	13181.75	620.50	865.67	163.50	505.00	926.00	117.00

相関

# ③ ユーザごとのデータ分析



\* クラスタID2については、ユーザごとの集計時に対象者なし

静岡県よりも全国の感染者数に影響を受けている

## 相関係数行列より、コロナの感染者を抽出

### 全国の感染者数

クラスタID	1	3	4	5	6	7
steps	0.022087	0.072265	-0.19595	-0.26151	-0.01687	-0.09325
Out of range minutes	-0.04629	0.090605	-0.21984	-0.06258	0.148005	-0.12272
Fat burning minutes	-0.08834	-0.1985	-0.21094	0.107304	-0.10189	-0.05568
aerobic exercise minutes	0.020624	-0.01669	0.039008	-0.31492	-0.15898	0.04943
peak_minutes	0.103914	0.019451	0.252479	-0.19322	-0.14826	-0.07434
duration	0.121938	-0.44602	-0.20891	-0.13711	-0.17611	0.230874
efficiency	-0.47732	-0.29101	-0.28428	-0.20111	-0.15455	0.084013

睡眠に関する項目との相関が多い

### 静岡県の感染者数

クラスタID	1	3	4	5	6	7
steps	0.011697	0.064919	-0.07531	-0.20363	-0.01958	-0.06888
Out of range minutes	-0.02281	0.102264	-0.04087	-0.09523	0.132897	-0.01438
Fat burning minutes	-0.0171	-0.11543	-0.13499	0.124129	-0.09421	-0.10718
aerobic exercise minutes	-0.06001	-0.0053	0.021466	-0.19962	-0.18152	-0.01414
peak_minutes	-0.04734	-0.01416	0.166381	-0.15879	-0.12502	-0.10107
duration	0.019087	-0.2862	-0.11714	-0.15591	-0.19199	0.132277
efficiency	-0.29009	-0.1773	-0.17146	-0.06747	-0.06656	0.12508

### ③ ユーザごとのデータ分析

全体データ（日毎の平均）とコロナ感染者数の相関係数についても求めた

	全国の感染者数	静岡県の感染者数
steps	-0.03377	-0.0365
Out of range_minutes	0.142363	0.085696
Fat burning_minutes	-0.26593	-0.11595
aerobic exercise_minutes	-0.2297	-0.19145
peak_minutes	0.01329	-0.12376
duration	-0.23113	-0.21821
efficiency	-0.50779	-0.25827

この結果から、睡眠効率がコロナ感染者数（全国）と弱い相関があることが確認された

#### **特に睡眠に関する項目に影響**

感染者数が増えると**睡眠効率が下がる**（負の相関）



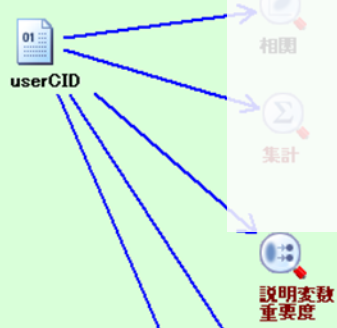
**睡眠効率を中心として解析を進める**



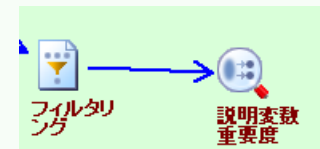
# ③ ユーザごとのデータ分析

睡眠効率の影響が、他の項目に与える影響度を、説明変数重要度を使用し確認

①で作成した  
ユーザごとにクラスタ番号を  
付与したデータ



全てのデータ、クラスタごとのデータで  
重要な変数を確認



項目間の関係性の把握

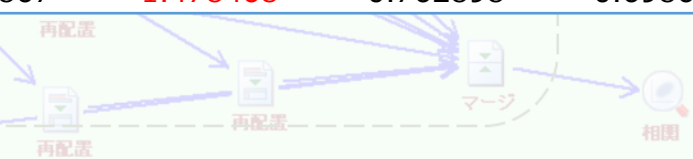
重要度

col_names	全体	クラスタ1	クラスタ3	クラスタ4	クラスタ5	クラスタ6	クラスタ7
steps	<b>1.447209</b>	0.630473	0.713159	0.36197	0.846277	<b>1.618289</b>	<b>1.446846</b>
Out of range minutes	<b>1.399048</b>	0.662222	0.559572	0.854525	1.170099	<b>1.376271</b>	<b>1.225118</b>
Fat burning minutes	0.868166	0.919715	<b>1.262519</b>	<b>1.862543</b>	1.095498	0.691739	0.659538
aerobic exercise minutes	0.832654	0.61875	2.027431	<b>1.592211</b>	0.55402	0.901017	0.965683
peak_minutes	0.521262	<b>1.905696</b>	0.529639	0.636884	0.855638	0.649788	1.004188
duration	0.93166	<b>1.263144</b>	0.90768	0.691867	<b>1.478468</b>	0.762898	0.698628

クラスタごとの特徴 (再掲)

- ID1 : 良く運動している (有酸素運動) が睡眠時間が短い
- ID3 : あまり運動をしておらず睡眠時間は短い、睡眠効率は良い**
- ID4 : あまり歩いていないが、有酸素運動を特にしており、睡眠時間は短め
- ID5 : 歩いてはいないが有酸素運動を行い、睡眠効率もよい**
- ID6 : 運動は少なめだが睡眠時間が長い
- ID7 : 歩いており睡眠効率も良い**

cross\_all\_nonnull





- コロナに関するオープンデータとウェアラブル端末(fitbit)から得られた健康データとの関係を、VMSを用いて分析した
- 睡眠効率が全国陽性者数と相関が確認された
  - メディアでの報道が睡眠効率に影響していると考えられる
  - 説明変数重要度分析により、睡眠効率に影響する項目の算出が可能
  - ユーザの特徴に合わせたアドバイスが可能になる