2024 年度 統計データ分析コンペティション

統計活用奨励賞 「高校生の部】

医療費削減に向けたスポーツ時間増加策のデータ分析

過 目今(法政大学国際高等学校)

論文の概要

医療費削減に向けスポーツ時間を阻害する要因を見つけるため、スポーツ行動者率とスポーツ行動者平均時間の二つについて分析し、青少年期では、睡眠時間とスポーツ時間の確保が重要であり、現役世代では、健康への危機感はスポーツをする強い動機になることや、飲みニケーションは阻害要因であることを示し、シニア世代においては、寝すぎを避けることとスポーツ仲間作りが重要であることを示した。

論文審査会コメント

医療費削減に向けたスポーツ時間の増加という視点は興味深く、データの性質をよく 理解した上で分析している。各県の特徴を示すグラフ利用などは十分ではないものの、 政策提言についてはよい探究的活動ができている。

医療費削減に向けたスポーツ時間増加策のデータ分析

過目今 (法政大学国際高等学校)

1. 研究のテーマと目的

1.1 背景

国民医療費は、毎年一貫して増加しており、20 年で 43%増加している(1)。医療費削減は急務であり、スポーツが有効だと報告されている。国土交通省(2)によると、一人一日一歩歩くだけで最小 0.065 円の医療費も節約できる。それを半分の人口が年間平日(261 日)毎日 1000 歩歩くとしたら、医療費は約 1 兆円(約 2.4%)削減できると計算できる。しかし、スポーツが十分に行われていないのが現状。どうしたらスポーツ時間を増やし、医療費削減に貢献できるのか。

1.2 目的

国民健康・栄養調査⁽³⁾結果によると、スポーツの一番の阻害要因は「仕事(家事・育児等)が忙しくて時間がないこと」。では、仕事時間を減らしたら、スポーツ時間が増えるだろうか。過去データを確認したところ(図 1)、仕事時間が年々減少し続けてきたが、スポーツ時間に有意な増減がなかった(仕事時間の減少分を身の回りの用事や休養・くつろぎ時間に当てた)。**医療費削減**するため、データ分析を通してスポーツ阻害の真因を見つけ、スポーツ時間を増やす解決策を提案することを本研究の目的とする。

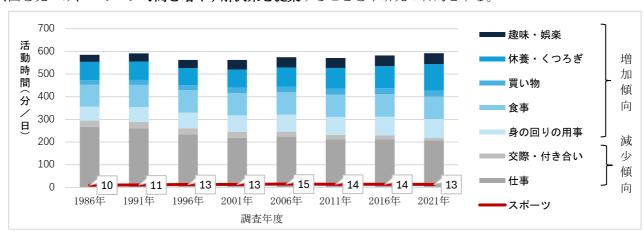


図1 仕事時間が減り続けてきたが、スポーツ時間に有意 (トレンド P 値<0.05) な増減が確認できなかった

2. 研究の方法と手順

2.1 研究方法

スポーツ時間を増やすには、スポーツ行動者数を増やすと、スポーツ行動者一人当たりのスポーツ時間を増やす二つの方向が考えられる。それぞれの解決策が違うため、スポーツ時間をスポーツ行動者率とスポーツ行動者平均時間の二つに分割して研究する(図2)。それぞれに対して、相関分析や重回帰分析を行い、影響の大きい要因を特定し、解決策を提案する。

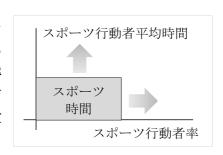


図2 スポーツ時間を行動者率と行動者平均時間に二分割

2.2 研究手順

以下(表1)の手順に基づいて研究を進める。

表1 研究手順

段階		手順	手順目的	手順説明
データ 準備	1.	影響要因仮説 を立てる	分析目的と範囲を明確 にする。	スポーツ行動者率とスポーツ行動者平均時間に影響を 与えそうな要因仮説を立てる。
	2.	分析データを 収集と加工	分析用データを集める。	欠損値がなく使いやすい SSDSE を主に、不足分は e-stat 等から収集し、適切に加工する。
	3.	欠損値分析	データの完全性を確保。	欠損値を特定する。特定後に発生原因を考察して、処置 を検討する。
	4.	外れ値分析	分析結果の歪みを防ぐ。	外れ値を特定する。特定後に発生原因を考察して、処置 を検討する。
	5.	説明変数間で の相関分析	回帰分析の多重共線性 問題を回避。	相関関係の強い説明変数ペアを特定し、処置を検討する。
	6.	VIF(分散膨張 因子)分析	回帰分析の多重共線性 問題を回避。	VIF 値の高い説明変数を特定し、処置を検討する。 補足:手順5は二つの説明変数間の相関関係分析に対し、 VIF は各説明変数が他全ての説明変数との共線性を持つ かを評価するので、より複雑な相関関係を検出可能。
データ 分析	7.	目的変数と説 明変数の相関 分析	重回帰モデルの解釈性 の向上と単純化を図る。	目的変数と相関関係のない・弱い説明変数を特定し、重 回帰分析用説明変数から除外する。
	8.	重回帰分析	変数間の影響力比較、重 要要因特定。	説明変数と目的変数の重回帰分析を行い、説明変数の 影響を比較分析する。
結論 提案	9.	考察と提案	スポーツ時間を増やす 解決策を提案。	データ分析の結果を考察し、解決策を提案する。

3. データセットの加工

3.1 影響要因仮説を立てる

影響要因を**社会環境、自然環境、スポーツ環境、個人生活**に分類して仮説を立てる(図3)。

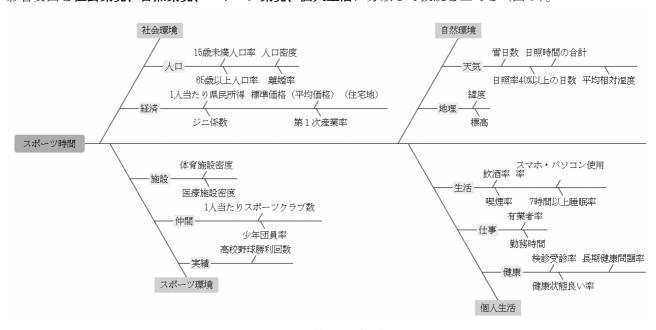


図3 影響要因仮説

3.2 分析データを収集と加工

影響要因仮説を基に、取得できるデータを収集&加工(表2)。

表2 分析用データの収集と加工

分類	変数	出典	期間	加工変換
目的	スポーツ行動者率	社会生活基本調査(4)	2021	-
変数	スポーツ行動者平均時間	社会生活基本調査(4)	2021	-
	15 歳未満人口率	SSDSE-E	2021	15 歳未満人口/総人口
	65 歳以上人口率	SSDSE-E	2021	65 歳以上人口/総人口
	人口密度	SSDSE-E	2021	総人口/可住地面積
	離婚率	SSDSE-E	2020	離婚件数/総人口
	1人当たり県民所得	SSDSE-E	2018	1人当たり県民所得(平成 23 年基準)
	ジニ係数	全国家計構造調査(5)	2019	等価可処分所得 (OECD新基準準拠)
	標準価格(平均価格)(住宅地)	SSDSE-B	2020	-
	笠 1	SSDSE-A	2020	第1次産業就業者数/(第1次産業就業者数+第2
	第1次産業率	22D2E-4	2020	次産業就業者数+第3次産業就業者数)
	雪日数、日照率 40%以上の日数、 日照時間の合計、平均相対湿度、 緯度、標高	SSDSE-F	1991- 2020	県庁所在地の年平均値を採用
국사 미디	体育施設密度	SSDSE-E	2018	(社会体育施設数+民間体育施設数)/可住地面積
説明	医療施設密度	SSDSE-E	2020	(一般病院数+一般診療所数)/可住地面積
変数	1人当たりスポーツクラブ数	文部科学省(7)	2020	育成クラブ総数/総人口
	少年団員率	日本スポーツ協会 ⁽⁸⁾	2021	少年団員数/15 歳未満人口
	高校野球勝利回数	Wikipedia ⁽⁹⁾	2024	-
	飲酒率	国民生活基礎調査(6)	2019	(毎日+週 5∼6 日+週 3∼4 日+週 1∼2 日)/総数
	喫煙率	国民生活基礎調査(6)	2019	(毎日吸っている+時々吸う日がある) /総数
	スマホ・パソコン使用率	社会生活基本調査(4)	2021	使用した/総数
	7 時間以上睡眠率	国民生活基礎調査(6)	2019	(7 時間以上 8 時間未満+8 時間以上 9 時間未満+9 時間以上)/総数
	有業者率	社会生活基本調査(4)	2021	有業者/総数
	勤務時間	SSDSE-D	2021	仕事からの帰宅時間-出勤
	検診受診率	国民生活基礎調査(6)	2019	検診等を受けた/総数
	健康状態良い率	社会生活基本調査(4)	2021	(良い+まあ良い) /総数
	長期健康問題率	社会生活基本調査(4)	2021	2_慢性的な病気や長期的な健康問題がある/総数

3.3 目的変数のグループ分け

スポーツ行動者率とスポーツ行動者平均時間のデータ分布を確認すると、年齢階層と性別によるパターンが確認できる(図 4)。より適切な影響要因を抽出するため、目的変数を年齢階層と性別の組み合わせでグループ分けして分析を行う。

年齢階層をパターンの見える青少年期:10~19歳、現役世代:20~64歳、シニア世代:65歳以上と設定。

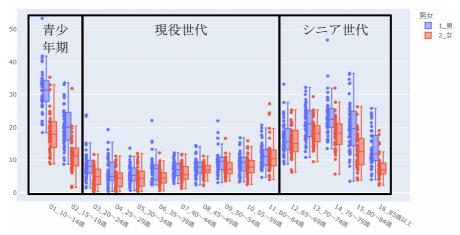


図4 年齢別・性別のスポーツ行動者率分布 補足:スポーツ行動者平均時間にも類似分布がある

3.4 説明変数の欠損値、外れ値分析

(表 3) 欠損値処置する後、(図 5) 箱ひげ図を利用して外れ値を特定し、(表 4) 外れ値処置する。

表3 欠損値処置

変数	欠損値	欠損値処置				
スポーツ行動者率	'-'がある	同年齢同性別の平均値を設定する。				
スポーツ行動者平均時間	'-' がある	同年齢同性別の平均値を設定する。				
飲酒率、喫煙率	10~19 歳データ欠損	未成年で飲酒しないため、0を埋める。				
スマホ・パソコン使用率	人数に'-','…'がある	計数のないあるいは計数不明のため、0 に置換える。				
7 時間以上睡眠率	10~11 歳データ欠損	同年齢帯は類似のため、12~14歳データを10~14歳として使う。				
有業者率	10~14 歳データ欠損	10~14 歳は有業とは考えづらいため、0 を埋める。				
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	75 歳以上段階分欠損	75 歳以上のデータを 5 歳段階に分割(各段階に同じ値使用)する。				
検診受診率	10~19 歳データ欠損	年齢的に一番近い20~24歳のデータを設定する。				
	人数に'-','…'がある	計数のないあるいは計数不明のため、0に置換える。				
健康状態良い率	10 歳段階	10 歳段階を 5 歳段階に分割(各段階に同じ値使用)する。				
	10~14 歳データ欠損	10~14 歳は年齢的に一番近い 15~24 歳のデータを設定する。				
長期健康問題率	10 歳段階	10 歳段階を 5 歳段階に分割(各段階に同じ値使用)する。				



図5 箱ひげ図(例示) 補足:上下限は四分位範囲から上下2倍

表4 外れ値処置

変数	外れ値処置			
スポーツ行動者率、スポーツ行動者平均時間、スマホ・パソコン使用率、7 時間以	年齢別でチェックしたところ、外れ値で			
上睡眠率、検診受診率	はなかったため、処置不要。			
ジニ係数、緯度、高校野球勝利回数	突出して外れてはなく、特性として残し			
ン一体数、輝度、前仪野郊勝利四数	たいため、処置不要。			
15 歳未満人口率、人口密度、離婚率、1 人当たり県民所得、標準価格(平均価格)	一般傾向把握したいため、外れ値を上下			
(住宅地)、標高、体育施設密度、医療施設密度、1人当たりスポーツクラブ数	限に合わせるようにクリップ処置。			

3.5 説明変数間での相関分析

説明変数間相関関係のヒートマップを使い、相関関係の強い(相関係数が 0.7 以上、あるいは-0.7 以下)説明変数同士を特定する。因果関係から、よりモデル解釈性高い説明変数を後続分析に使用する。(図 6、図 7)

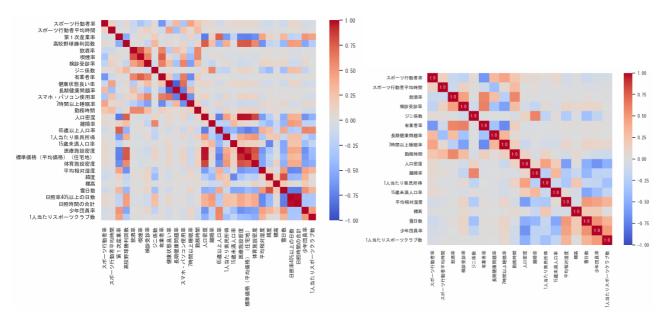


図6 説明変数間相関係数(除外前)

図 7 説明変数間相関係数 (除外後)

補足:0.7以上・-0,7以下の係数がないことを確認

3.6 VIF (分散膨張因子) 分析

VIF 値を出力して、VIF 値の高い順から説明変数を除外する。全ての説明変数が 10 未満になるまで繰り返し実施する (表 5)。残った説明変数を後続分析に使用する。

表 5 VIF 分析

説明変数	VIF 値(除外前)	除外変数	VIF 値(除外後)
15 歳未満人口率	484. 9	•	
人口密度	9. 4		4.2
離婚率	286. 9	•	
1人当たり県民所得	120. 0	•	
ジニ係数	476. 4	•	
雪日数	7. 1		3. 5
平均相対湿度	775. 3	•	
標高	2. 5	•	
1人当たりスポーツクラブ数	9. 0	•	
少年団員率	10. 2		6.6
飲酒率	19. 6		8.0
7 時間以上睡眠率	22. 2		7.3
有業者率	20. 9		8.6
勤務時間	230. 3	•	
検診受診率	116. 9	•	
長期健康問題率	11. 1		6. 2

背景色ない 変数は後続 分析に使用 する

4. データ分析の結果

4.1 目的変数と説明変数の相関分析

分析の結果(表6)を以下にまとめる。

- ① 説明変数の分類毎に見ると、社会環境、自然環境、スポーツ環境等外部環境とも相関関係があるが、**個人生活との相関**関係が最も強い。
- ② 目的変数毎に見ると、選定した説明変数は、スポーツ行動者平均時間より、**スポーツ行動者率との相 関**関係が強い。
- ③ 個別変数に注目していくと、想定外の相関関係がいくつかある。
 - (ア) 「少年団員率」は「スポーツ行動者率」と負の相関。(「少年団員率」が高まると、少年の保護者達にも影響を与え、彼らのスポーツ行動者率も高まると考えていた。)
 - (イ)「飲酒率」は「スポーツ行動率」、「スポーツ行動者平均時間」(の一部)と正の相関。(飲酒者は 比較的にスポーツに関心が薄く、スポーツを行動しないかと考えていた。)
 - (ウ) 「7時間以上睡眠率」と「有業者率」は青少年期とシニア世代の場合に相関関係が強い一方、「長期健康問題率」は現役世代の場合に相関関係が強い。(年齢階層に大きな違いを想定しなかった。)

目的変数	年齢階	男	有意 (P 値<0.05) な相関係数						
	層	女	人口密度	雪日数	少年団員	飲酒率	7 時間以	有業者率	長期健康
					率		上睡眠率		問題率
	青少年	男	-0. 21				0.62	-0.66	
	期	女					0.48	-0.46	
スポーツ	現役	男		-0. 19	-0.11	0.20		-0. 15	0.35
行動者率	世代	女	0.15	-0. 15	-0. 17	0. 24	-0. 41	-0.33	0. 54
	シニア	男	0. 28	-0. 22	-0. 17	0.14	-0. 27	-0.15	0.14
	世代	女	0.25	-0.30	-0.23	0.49	-0.64	0. 24	
	青少年	男					-0. 43	0.42	
20	期	女					-0.36	0. 24	
スポーツ	現役	男				-0.34		-0. 19	-0. 24
平均時間	世代	女				-0. 15	0.16		-0. 19
	シニア	男	0.14	-0. 13		0.36	-0.38	0.30	-0. 22
	世代	女	0.15	-0. 16	-0.20		-0. 27		

表6 目的変数と説明変数の相関分析の結果

4.2 重回帰分析

4.1 の相関分析で、目的変数と**有意な相関関係**のある説明変数に対し、重回帰分析を実施。また、説明変数の単位違いを吸収するため、回帰分析前に説明変数の**標準化**を実施。分析の結果(表 7)を以下にまとめる。

- ① 全ての回帰モデルが**有意**である。(F P値 < 0.05)
- ② スポーツ行動者率のモデルのほうは説明変数の**説明可能範囲**が大きい。(調整 R²の値比較)
- ③ 個別変数の標準化係数を注目していくと
 - (ア) 「スポーツ行動者率」は複数要因による影響を受けているが、「スポーツ行動者平均時間」は単一要因になっている。(小さい調整 R²とも関連していると考えられる。)
 - (イ)複数要因の影響度合い(標準化係数)について、相関の強いものは影響も大きいのが確認できる。

表 7 重回帰分析の結果

目的変数	年齢	男	有意(P値<0.05)な標準化係数							モデル評価	
	階層	女	人口密度	雪日数	少年団員	飲酒率	7 時間以	有業者率	長期健康	F_P 値	調整
					率		上睡眠率		問題率		R ²
	青少年	男	-1.4				2.5	-3. 1		1. E-12	0.46
	期	女					2			2. E-06	0.23
スポーツ	現役	男		-0.7				-0.7	1.7	2. E-19	0.20
行動者率	世代	女			-0. 4		-0.6	-0.4	1.6	2. E-36	0.35
	シニア	男	1.5			1. 3	-2.8	-2.6	1.5	2. E-14	0.27
	世代	女		-1.1		1. 9	-3.6	-2. 1		2. E-31	0.48
	青少年	男								5. E-05	0.18
2 to 10	期	女					-20.2			2. E-03	0.11
スポーツ 行動者 平均時間	現役	男				-17.8				1. E-11	0.11
	世代	女							-5. 1	5. E-04	0.03
	シニア	男				5. 1				9. E-08	0.15
	世代	女					-4.1			1. E-04	0.08

5. 結果の解釈

5.1 影響要因分析結果の考察

年齢階層別に分析結果を考察する内容を以下にまとめる。記述を簡潔にするため、「スポーツ行動者率」を 【率】、「スポーツ行動者平均時間」を【時】と略して記述する。

① 青少年期

- (ア) **睡眠時間とスポーツ時間の確保が重要。**「7 時間以上睡眠率」は【率】と正の相関。十分な睡眠時間を取ることで元気になり、スポーツしたくなるだろう。一方で、【時】と負の相関である。睡眠時間を長くとることでスポーツに割ける時間が短くなるだろう。また、「有業者率」が【率】と負の相関があるのは、バイトに時間がとられ、スポーツする気がなくなったからだろう。
- (イ) スポーツ少年団の効果に疑問符。「少年団員率」は【率】【時】と弱い相関関係のみである。且つ、 無相関の帰無仮説を棄却できなかった。スポーツ少年団の取り組みはスポーツ促進を目的として いるため、正の相関かと思ったが、そうではなかった。

② 現役世代

- (ア) **健康への危機感は強い動機になる。**「長期健康問題率」は【率】と正の相関、スポーツ行動者平均時間と負の相関である。長期健康問題を抱えていることで、健康に危機感を抱き、スポーツで健康状態を改善しようとしているだろう。だが、健康問題で体力的に長時間スポーツできないのだ。
- (イ) **飲みニケーションは阻害要因。**「飲酒率」は【時】と負の相関である。特に男性現役世代は、仕事で各種飲み会の付き合いがあり、なかなかスポーツ時間を取れない状態だろう。
- (ウ) スポーツ促進策の副作用要注意。「少年団員率」は女性の【率】と負の相関である。それは、おそらく子供のスポーツ活動に付き合うため、自分がスポーツできなくなってしまっただろう。このように、ある年齢階層に対して解決策を打つことによって、ほかの年齢階層に影響が出てしまう可能性があるので、要注意。

③ シニア世代

(ア) **寝すぎを避けることが重要。**「7 時間以上睡眠率」は【率】【時】と負の相関である。シニア世代

は空き時間が比較的多いため、過剰に睡眠時間を取ってしまう傾向がある。結果、体力低下や疲 労感を感じやすい生理的な影響や、社会的な孤立感とやる気を失う精神面の問題があるだろう。

(イ) スポーツ仲間作りが重要。「飲酒率」は【率】【時】と正の相関である。飲酒すると、スポーツしやすくなるとは考えにくい。飲酒を通してスポーツする仲間を作れ、仲間と一緒にスポーツできるからスポーツしやすいということで間接的に繋がっているのだろう。また、「人口密度」が【率】と正の相関であることも、仲間を作りやすい観点からで説明できる。

5.2 結論と解決策

スポーツ時間増やすには、仕事時間を減らすような安易な解決策では効かず、有効な解決策を策定するには、5.1 で考察したそれぞれの**年齢階層の特性**を踏まえて、**個別対策**を講じる必要がある。私が勘案した解決策を以下にまとめる。

- ① **青少年期**:(ア)学校あるいは保護者から**睡眠時間**を確保するように**指導**する。(イ)スポーツをバイト、勉強等と両立できるように、**時間管理**の教育を実施する。
- ② 現役世代:(ア)健康診断等を機に、健康状態を観測できるスマートウォッチを配布/貸出、健康状態と健康トレンドを可視化させて、危機感を醸成する。(イ)コミュニケーション文化を飲み会からスポーツに誘導する。例:団体でコミュニケーション取れる、且つ気軽にできるスポーツの場を用意する。
- ③ **シニア世代**: (ア)話し相手になる**見守りロボット**を支給し、孤独感を低減、生活リズムを保つように 提言する。(イ)社交の場の提供、イベントの企画等で、**仲間を作れる環境**を整える。

5.3 今後のアクション

今後は、データ分析モデルの改善と解決策効果の検証の同時進行していきたいと考えている。

- ① 分析モデル改善 (ア)説明変数追加:特にスポーツ行動者平均時間のモデルの説明可能範囲が小さい問題を要改善。(イ)多重共線性問題追加検証:相関性分析時強い相関関係があったが、回帰分析時は有意な影響が確認できない変数がある。そこには多重共線性の可能性があり、要検証。(ウ)交互作用項検証。問題解決には、複数変数の交互作用を考慮し、より複雑な影響関係を要理解。
- ② **解決策効果検証** (ア)既存解決策効果検証:スポーツ少年団の効果確認できなかったが、長年存在していた仕組みであり価値あるはずだ。本当に効果ないかをデータを変えて再確認。(イ)提案解決策効果検証:今回の解決策は因果関係含めて各種仮説を元に推理されたものもあり、効果要検証。家族内で年齢階層別の解決策を試行と検証を実施。そしてフィードバックを受けて、改善していきたい。

参考文献

- (1) 厚生労働省: "令和2 (2020) 年度 国民医療費の概況"、pp. 3 (2020).
- (2) 国土交通省: "<u>資料 5</u>"、まちづくりにおける健康増進効果を把握するための歩行量(歩数)調査のガイドライン、pp. 5 (2017).
- (3) 厚生労働省: "<u>令和元年国民健康・栄養調査報告</u>"、国民健康・栄養調査報告、pp. 45 (2019).
- (4) 総務省統計局: "生活時間に関する結果"、"生活行動に関する結果"、令和 3 年社会生活基本調査の結果 (2021).
- (5) 総務省統計局: "全国・都市階級・地方・4大都市圏・都道府県別等価所得のジニ係数"、全国家計構造調査 (2019).
- (6) 厚生労働省:"令和元年国民生活基礎調查"、国民生活基礎調查(2019).
- (7) 文部科学省: "<u>令和2年度総合型地域スポーツクラブ育成状況(一覧表)</u>"、令和2年度総合型地域スポーツクラブ育成状況調査(2020).
- (8) 日本スポーツ協会: "●令和3年度スポーツ少年団登録状況(都道府県別 性別 競技別 団員数)"(2021).
- (9) Wikipedia: "高校野球の都道府県別記録一覧" (2024 春時点).