

金曜 3,4 限 プロジェクト 最終報告書

武蔵野大学工学部数理工学科

3 年 白川桃子 阿辻颯姫

2 年 村田滉希 田口冬佳

チーム名

カフェインず（仮）

タイトル

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

メンバー

リーダー 村田滉希

副リーダー 白川桃子

メンバー 阿辻颯姫 田口冬佳

目的

1. 統計的手法を用いて日本人のコーヒーと生活行動やがん疾患による死亡率の関係性を明らかにする
2. 分析を進めていく課程で、チームメンバー各々が自発的に学び発言することで、統計学に関するスキルを向上する。
3. 発表する際の情報整理やスライド作成、伝えたいことを適切に伝える手法を考える。
4. 統計コンペティションに参加し、統計リテラシーの向上につとめる。

コンテスト概要

統計データ分析コンペティションとは

総務省統計局、統計センター主催。高校生、大学生等を対象に、地域別の統計をまとめた SSDSE（教育用標準データセット）を用いた統計データ分析の論文を募集し、そのアイデアと解析力を競うコンペティションである。令和5年度第六回大会は、エントリー期間が令和5年5月10日（水）～8月9日（水）、論文提出期間は令和5年5月10日（水）～8月31日（木）（大学生・一般の部の場合）であった。部門ごとに「総務大臣賞」、「優秀賞」、「統計数理賞」、「統計活用奨励賞」等の受賞作品が選ばれる。我々は大学生・一般の部に参加したが、惜しくも入賞することはできなかった。

スケジュール

日付	内容
5/26～6/2	分析テーマ設定
6/9～6/30	データ収集（コーヒー購入量、生活行動） コーヒーと生活行動の重回帰分析 企画発表会へ向けた発表準備
7/26～8/13	データ収集（がん死亡率、コホート研究） コーヒーとがん疾患による死亡率の重回帰分析
8/14～8/31	コンペティション提出用の論文作成
9/1～9/29	中間発表会へ向けた発表準備 本を用いた統計学の勉強
10/1～12/1	python を用いた統計的分析手法の実装
12/2～12/25	最終発表会へ向けた発表準備

成果物

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

村田滉希 白川桃子 田口冬佳 阿辻颯姫
(武蔵野大学工学部数理工学科)

要約

コーヒーの持つ短期的効果（覚醒効果や集中度を高める効果）と長期的効果（がんなどの疾患の予防効果）について、それぞれが及ぼす日本人の生活行動への影響、及び疾患による死亡率への影響を統計的に明らかにすることを目的として、コーヒー購入量と、スポーツや趣味などの生活行動との関係、及びがんによる死亡率との関係を、重回帰分析によって求めた。その結果、スポーツと趣味・娯楽は、コーヒー購入量と強い相関があり、その中でも、より集中力を要するスポーツや趣味との相関が大きいことが分かった。また、コホート研究で実証されたコーヒー摂取による発症リスク低減の有意性を支持する結果が、有意性は高くはなかったが、いくつか得られた。

1. 導入

日本では約 70%もの人が毎日コーヒーを飲んでおり、私たちの生活とコーヒーは密接な関係にある[1]。唯一無二の香りや味だけではなく、短期的な効果として、カフェインによる覚醒作用[2]や、香りによる集中度を高める作用・リラックス効果などが知られており[3]、[3]では、コーヒー摂取によって作業成績が上昇すると共に、ストレス負荷によるポジティブ気分の軽減を抑制する可能性が示唆されている。コーヒーにはこのような効果があるので、高い集中力を必要とするスポーツや将棋などの趣味の分野で、コーヒーが良く飲まれていることが、多くのサイトで指摘されている[4][5]。

また、病気の発症や死亡率を低減する長期的な効果も知られている。多目的コホート研究によって、コーヒーの摂取が日本人の総死亡率やがん罹患、心疾患、脳血管疾患、呼吸器疾患低下と有意な関連があることが示唆されている。これらのメカニズムについては、まだ解明されていないが、コーヒーの炎症を和らげる作用やカフェインの抗酸化作用などががん化を防御している可能性が議論されている[6][7]。

ここでは、コーヒーの持つ短期的効果と長期的効果について、それぞれが及ぼす日本人の生活行動への影響と、疾患による死亡率への影響を統計的に明らかにすることを目的とする。前者の目的のために、都道府県毎のコーヒー購入量を目的変数とし、SSDSE-社会生活の「生活行動に関する結果」の各項目のデータを説明変数として、重回帰分析を行った。また、後者の目的のために、がん種毎の都道府県毎 75 歳未満年齢調整死亡率を目的変数として、都道府県ごとの要因データ（野菜摂取量、食塩摂取量など）にコーヒー購入量を加えて、重回帰分析を行った。

2. 研究方法と手順

2.1 データの取得

本研究で使用したデータについて以下に記載する。

表1 使用したデータ名と出典一覧

タイトル	年次	出典
コーヒー数量 *1	2007 - 2021	(e-Stat 政府統計の総合窓口から取得)
社会生活(生活行動に関する結果) *2	2021 版(収録年次は 2016)	独立行政法人 統計センター SSDSE-社会生活
国民健康・栄養調査 *3	2006-2010	(e-Stat 政府統計の総合窓口から取得)
人口動態統計(厚生労働省大臣官房統計情報部) 都道府県別がん死亡データ*4	2020	国立がん研究センターがん情報サービス「がん統計」(人口動態統計)

*1 コーヒー数量は二人以上の世帯を対象とした家計調査から県庁所在市別 1 年間のコーヒーの総購入数量(単位: g)の 2007 年～2021 年のデータを e-Stat から取得した。なお、県庁所在市を県として扱った(例: 札幌市→北海道)。

*2 社会生活は SSDSE-社会生活に収録されている生活行動に関する結果を使用した。生活行動に関する結果は、自由時間等での 5 種類の活動(学習・自己啓発・訓練、スポーツ、趣味・娯楽、ボランティア活動、旅行・行楽)について、それぞれの内訳別の、過去 1 年間に活動した人の割合(行動者率)を示している。

*3 国民健康・栄養調査から、都道府県別の野菜摂取量、食塩摂取量、肥満率、飲酒率、喫煙率のデータを用いた。野菜摂取量と食塩摂取量は、20 歳以上の人の摂取量の平均値であり、単位は g/日である。肥満率は、20～69 歳の肥満者の割合(%)である。飲酒率は、20 歳以上の飲酒習慣者の割合(%)である。喫煙率は、20 歳以上の現在習慣的に喫煙をする者の割合(%)である。それぞれのデータは年齢調整を行っている。

*4 人口動態統計(厚生労働省大臣官房統計情報部)の部位別 75 歳未満年齢調整死亡率(人口 10 万人対)から都道府県別がん死亡データをエクセルファイルとして取得した。

2.2 分析方法

最初に、コーヒーの都道府県ごとの購入量が、どの程度ばらついているかを知るために、ヒストグラムを描いて分布の特徴を明らかにする。次に、注目している変数間の重回帰分析を順に行う。まず、生活行動に関する結果を説明変数、コーヒーの購入数を目的変数として、重回帰分析を行う。生活行動に関する結果は階層的構造となっている。5 種類の活動に分けら、それぞれの活動が、さらに数十種類の活動から成っている。そこで、まず、5 種類の活動毎の総数を説明変数とする(重回帰分析(1))。その結果有意であった活動について、その中に含まれる活動を説明変数として重回帰分析を行う(重回帰分析(2))。さらに、各種がん死亡率を目的変数、各種要因データにコーヒー購入数を加えたものを説明変数として、重回帰分析を行う(重回帰分析(3))。分析の最初に、すべての変数の標準化を行う。分析には、Microsoft Excel と Python を用いた。

2.2.1 重回帰分析(1)の手順

①多重共線性の検討

説明変数間の相関係数が 1 に近いと多重共線性の問題が生じて結果が不安定になる。相関係数が 0.95 以下であれば影響は大きくないといわれているので[8]、5 種類の変数間の相関係数マトリクスを計算し、相関係数 ≤ 0.95 が成り立っているかどうかを確認する。

②変数選択(説明変数選択基準による変数選択)

3 つの説明変数選択基準を用いて、減少法によって最適な変数の数を決定する。有意な説明変数の個数の選択について、Ru[9](上田、1997)、Rh[10](芳賀ら、1976)、AIC[11](赤池、1973)を算出して検討する。Ru と Rh は最大かつ AIC は最小となる値が最適な説明変数の個数であり、最適な回帰式であると判断する。以下に Ru、Rh 及び AIC の計算式を示す。

i) 上田の説明変数選択

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

ii) 芳賀・竹内・奥野の説明変数選択規準

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

iii) 赤池情報量規準 (AIC)

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

③重回帰分析

②で選択された変数を説明変数として重回帰分析を行う。

2.2.2 重回帰分析(2)

手順は、以下の通りである。

①変数選択-1

考慮すべき変数の数が 20 個以上あるので、ある程度相関がある 10 数個の変数に絞り込む。各説明変数と目的変数の間で単回帰分析を行い、各変数の p 値を取得し、p 値がある一定の値以下の変数を選択する。

②多重共線性の検討

選択した変数間の相関係数マトリクスを計算し、相関係数<0.95 であることを確認する。

③変数選択-2

上述した説明変数選択基準を用いて、減少法によって、最適な変数の数を決定する。重回帰係数の p 値が最大の説明変数を 1 個ずつ削除していき、説明変数が 1 個になるまで重回帰分析を繰り返す。

④重回帰分析

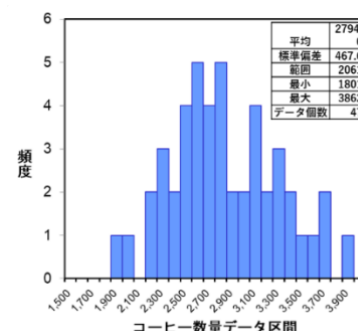
③で選択された変数を説明変数として重回帰分析を行う。

2.2.3 重回帰分析(3)

男性のがんの場合は、肥満、野菜、喫煙、飲酒、食塩、コーヒーの 6 種類の変数を説明変数とした。女性と男女総数のがんの場合は、喫煙、飲酒、肥満のデータが女性の場合に取られていないので、野菜、食塩、コーヒーを説明変数とした。その後、重回帰分析を行った。

2.2.4 重回帰分析の結果の可視化

得られた重回帰分析の結果より、変数ごとの重回帰係数と p 値を、一つの散布図上にプロットする。このように可視化することで、重回帰係数の符号と大きさ、及 p 値の大きさを同時に見ることができ、また複数の変数についてのプロットを同時に見て比較考察することが容易になる。



3. 結果と考察

3.1 コーヒー購入量の分布

コーヒー購入量のヒストグラムを図 1 に示す。図 1 に示すように、若干右に偏った分布を示しており、平均約 2800、最大 3863、最小 1801 で、標準偏差 467.7 であった。標準偏差×2 は、平均値に対して、約 33%の大きさがある。

図 1 コーヒー購入量のヒストグラム

3.2 重回帰分析(1)

3.2.1 多重共線性の検討

5 種類の変数（生活行動総数）間の相関係数マトリクスを図 2 に示す。このように、全ての組み合わせで相関係数が 0.95 以下であった為、5 種類全ての変数を用いることとした。

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・行事の総数
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・行事の総数	0.68	0.63	0.77	0.09	1.00

図 2 生活行動総数間の相関係数マトリクス

3.2.2 変数選択

説明変数の数を 5 個から、1 個ずつ減らしながら Ru、Rh 及び AIC を計算した（図 3）。Ru と Rh が最大

かつ AIC が最小となる回帰式は、説明変数が 2 個の場合（スポーツと趣味・娯楽）であった。

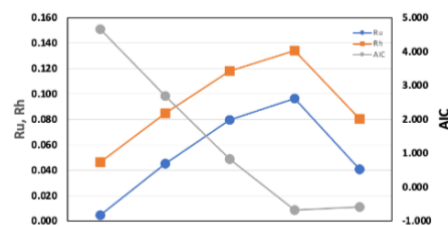


図3 社会生活総数の \bar{R}_u 、 \bar{R}_h 及び \bar{AIC} の説明変数の数依存性

3.2.3 重回帰分析

説明変数を 2 個（スポーツと趣味・娯楽）にして、目的変数をコーヒー購入量として重回帰分析を行った。x 軸に重回帰係数、y 軸に p 値をプロットしたものを示す(図 4,図 5)。図 4 が説明変数の選択前の分析結果を、図 5 が説明変数の選択後の分析結果をプロットしたものである。また、x 軸について赤文字括弧のものは負の値を示す。(例：(0.5) → -0.5)。各プロットの右に変数名と p 値を示す。説明変数の選択前の図から、釣鐘型にプロットされていることが分かる。重回帰係数の絶対値が大きくなると、p 値が小さくなる傾向にある。5 つの変数の内、スポーツの総数と趣味・娯楽総数のみが、選択前後に p 値が 0.1 を下回っており、選択後には p 値が 0.05 以下に減少していることから、コーヒー購入量との関係が統計学的に有意であることが分かった。コーヒーを飲む人はスポーツをプレイしない傾向、趣味・娯楽をする傾向にあるものとなった。

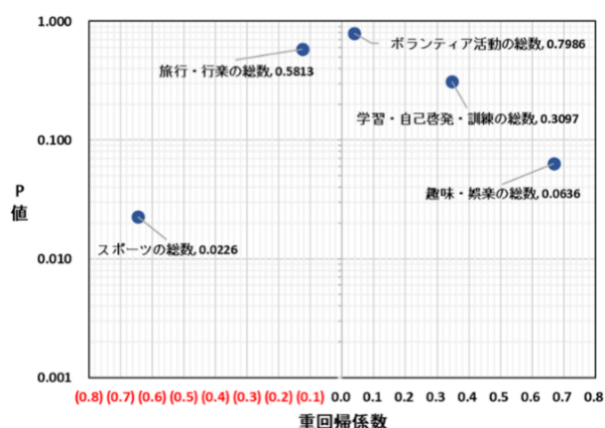


図 4 各種社会生活の重回帰係数と p 値（変数選択前,5 変数）

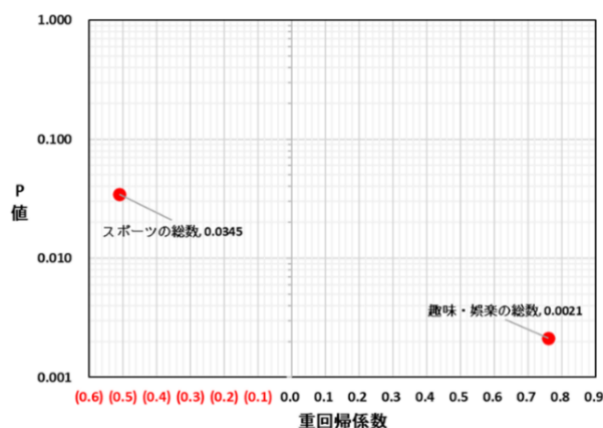


図 5 各種社会生活の重回帰係数と p 値（変数選択後,2 変数）

3.3 重回帰分析(2)

3.3.1 趣味・娯楽に属する変数の利用

1) 変数選択-1

趣味・娯楽に属する 34 の変数と目的変数の間で単回帰分析を行い、各変数の p 値を取得し、p 値が 0.05 以下の 12 個の変数を選択した。（美術鑑賞、演芸・演劇・舞踊鑑賞、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）

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 などによる 音楽鑑賞)	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 用も含む)
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	0.65	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	0.88	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	0.85	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	0.65	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	0.83
陶芸・工芸	0.67	0.70	0.67	0.70	0.28	0.47	1.00	0.72	0.73	0.37	0.26	0.60
写真の撮影・プリント	0.72	0.82	0.85	0.86	0.23	0.56	0.72	1.00	0.88	0.39	0.47	0.82
趣味としての読書	0.82	0.88	0.91	0.82	0.29	0.58	0.75	0.88	1.00	0.62	0.48	0.82
囲碁	0.47	0.39	0.41	0.41	0.39	0.24	0.37	0.39	0.62	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、携帯用も含む)	0.66	0.73	0.85	0.76	0.31	0.53	0.60	0.85	0.82	0.27	0.49	1.00

図 6 趣味・娯楽の詳細項目間の相関係数マトリクス

2) 多重共線性の検討

選択した変数間の相関係数マトリクスを計算した(図6)。最も高い値が 0.91 であり、0.95 以下であることが確認できたので、これら 12 個の変数を出発点として、変数選択を実施することとした。

3) 変数選択-2

説明変数の数を 12 個から 1 個ずつ減らしながら Ru、Rh 及び AIC を計算した(図7)。Ru と Rh が最大かつ AIC が最小となる回帰式は、説明変数が 7 個の場合(将棋、茶道、写真の撮影・プリント、CD・スマートフォンなどによる音楽鑑賞、編み物・手芸、演芸・演劇・舞踊鑑賞、陶芸・工芸)であることが分かった。

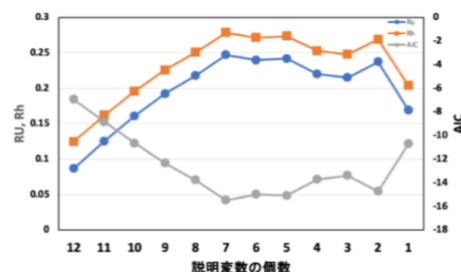


図 7 趣味・娯楽の各項目の Ru、Rh 及び AIC の説明変数の数依存性

4) 重回帰分析

説明変数を 7 個にして、目的変数をコーヒー購入量として重回帰分析を行った。x 軸に重回帰係数、y 軸

に p 値をプロットしたものを示す(図 8, 図 9)。図 8 が説明変数の選択前の分析結果をプロットしたもので、図 9 が選択後のプロットである。選択前後の図から、重回帰分析(1)と同様に、釣鐘型のプロット、及び重回帰係数の絶対値が大きくなると p 値が小さくなる傾向が観察された。選択後は、コーヒー購入量と写真の撮影・プリント、演芸・演劇・舞踊鑑賞が負の相関にある一方、コーヒー購入量と将棋、CD・スマートフォンなどによる音楽鑑賞、編み物・手芸、茶道、陶芸・工芸は正の相関にある。また、選択後の p 値は、多くの変数で 0.1 以下、将棋、CD・スマートフォンなどによる音楽鑑賞、編み物・手芸、写真の撮影・プリントで 0.05 以下であり、多くが統計的に有意であることが分かる。正の相関にある項目は、将棋や茶道など、より「集中力」を求める趣味・娯楽となっているように見える。また、コーヒーに多く含まれる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[2]こともこれらを裏付けるものと考えられる。また、負の相関にある項目では、演芸・演劇・舞踊鑑賞など、すぐに飲み物を摂取できないような状況が多いと思われるが、正の相関にある項目は基本的に室内で行われるものが多く、コーヒーの摂取がしやすい環境であると考えられる。

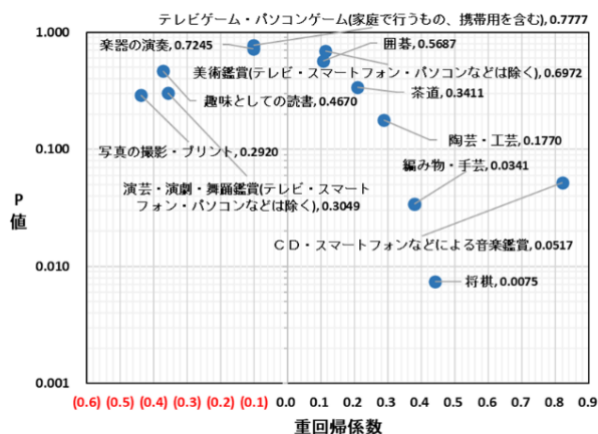


図 8 各種趣味・娯楽の重回帰係数と p 値 (変数選択前, 12 変数)

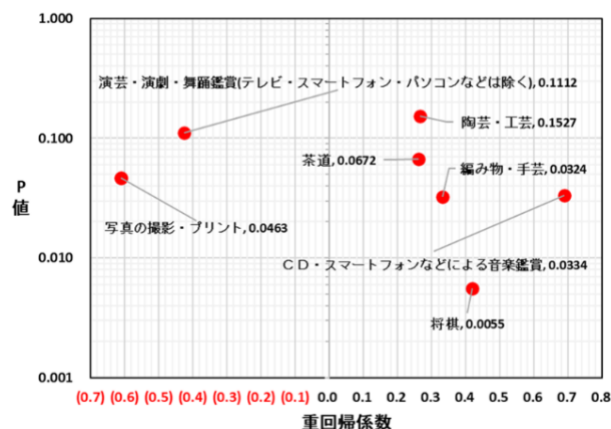


図 9 各種趣味・娯楽の重回帰係数と p 値 (変数選択後, 7 変数)

3.3.2 スポーツに属する変数の利用

1) 変数選択-1

スポーツに属する 22 の変数と目的変数の間で単回帰分析を行い、各変数の p 値を取得し、p 値が 0.3 以下の次に示す 16 個の変数を選択した。(m)

2) 多重共線性の検出

選択した変数間の相関係数マトリクスを計算した(図 10)。最も高い値が 0.74 であり、0.95 以下であることが確認できたので、これら 16 個の変数を出発点として、変数選択を実施することとした。

変数	野球	サッカー	バスケット	テニス	卓球	バドミントン	剣道	ゲートボール	ボウリング	器具	その他	合計
野球	1.00	0.65	0.45	0.35	0.25	0.15	0.10	0.05	0.02	0.01	0.00	0.00
サッカー	0.65	1.00	0.55	0.40	0.30	0.20	0.15	0.10	0.05	0.02	0.01	0.00
バスケット	0.45	0.55	1.00	0.45	0.35	0.25	0.20	0.15	0.10	0.05	0.02	0.01
テニス	0.35	0.40	0.45	1.00	0.50	0.40	0.30	0.25	0.20	0.15	0.10	0.05
卓球	0.25	0.30	0.35	0.50	1.00	0.60	0.50	0.40	0.35	0.30	0.25	0.20
バドミントン	0.15	0.20	0.25	0.40	0.60	1.00	0.70	0.60	0.50	0.45	0.40	0.35
剣道	0.10	0.15	0.20	0.30	0.50	0.70	1.00	0.80	0.70	0.65	0.60	0.55
ゲートボール	0.05	0.10	0.15	0.25	0.40	0.60	0.80	1.00	0.90	0.85	0.80	0.75
ボウリング	0.02	0.05	0.10	0.15	0.35	0.50	0.70	0.90	1.00	0.95	0.90	0.85
器具	0.01	0.02	0.05	0.10	0.30	0.45	0.65	0.85	0.95	1.00	0.95	0.90
その他	0.00	0.00	0.00	0.00	0.20	0.35	0.55	0.80	0.90	0.95	1.00	0.95
合計	0.00	0.00	0.00	0.00	0.20	0.35	0.55	0.80	0.90	0.95	0.95	0.95

図 10 スポーツの詳細項目間の相関係数マトリクス

3) 変数選択-2

説明変数の数を 16 個から始めて、1 個ずつ減らしながら Ru、Rh 及び AIC を計算した(図 11)。Ru と Rh が最大かつ AIC が最小となる回帰式は、説明変数が 6 個の場合(野球、バドミントン、剣道、ゲートボール、ボウリング、器具を使ったトレーニング)であることが分かった。

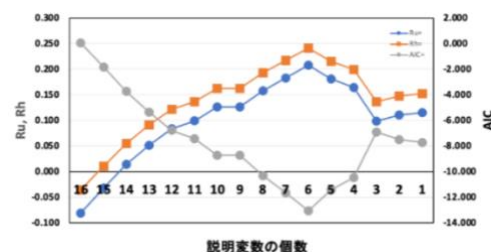


図 11 スポーツの各項目の Ru、Rh 及び AIC の説明変数の数依存性

4)重回帰分析

説明変数を 6 個にして、目的変数をコーヒー購入量として重回帰分析を行った。x 軸に重回帰係数、y 軸に p 値をプロットしたものを示す(図 13,図 14)。図 13 が説明変数の選択前の分析結果を、図 14 が選択後の分析結果をプロットしたものである。選択前後の図から、重回帰分析(1)(2)と同様に、釣鐘型のプロット、及び重回帰係数の絶対値が大きくなると p 値が小さくなる傾向が観察された。選択後は、コーヒー購入量とゲートボール、器具を使ったトレーニングが負の相関にある一方、コーヒー購入量と野球、バドミントン、剣道、ボウリングは正の相関にある。また、選択後の p 値は、全ての変数で 0.1 以下、ボウリングと剣道を除いて 0.05 以下であり、統計的に有意であることが分かる。

正の相関にある項目は、趣味・娯楽の場合と同様に、バドミントン、野球、剣道など、勝利のために、より「集中力」が必要なスポーツとなっており、カフェインの覚醒効果と集中力を高める効果が関連しているものと考えられる。このように、趣味・娯楽と同様に、「集中力」を求められるものがコーヒー購入量と正の相関があることが分かった。

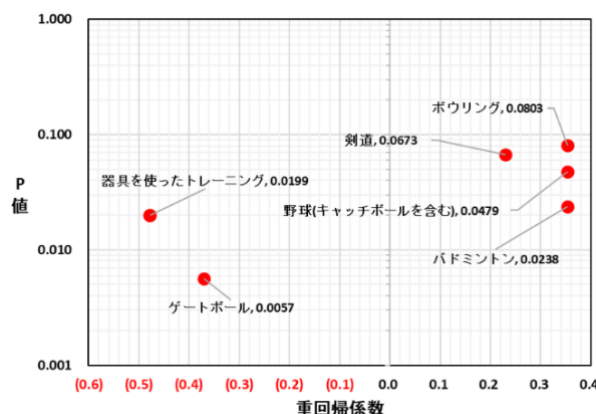
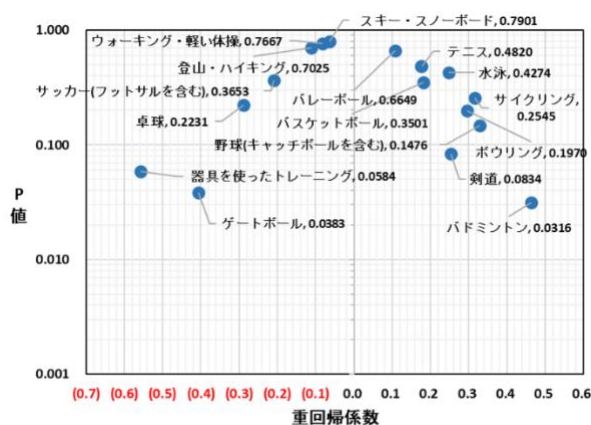


図 13 各種スポーツの重回帰係数と p 値(変数選択前,16 変数)

図 14 各種スポーツの重回帰係数と p 値(変数選択後,6 変数)

3.4 がんの死亡率の重回帰分析

男性と女性と男女総数に分け重回帰分析を行った。目的変数と説明変数として、以下のデータを用いた。

3.4.1 目的変数と説明変数

目的変数を、2020 年の都道府県別の、男性、女性のがん種毎の 75 歳未満年齢調整死亡率(人口 10 万人対)とした。説明変数として、「国民健康・栄養調査」からの 2006 年から 2010 年の都道府県ごとの要因データとコーヒー購入量を用いた。コーヒー購入量として「総務省統計局 家計調査」からのデータを用いた。

①男性では、説明変数として、野菜摂取量、食塩摂取量、肥満率、飲酒率、喫煙率、コーヒー購入量を用いた。

②女性では、説明変数として、野菜摂取量、食塩摂取量、コーヒー購入量を用いた。

③男女総数では、目的変数を、2020 年の都道府県別の男性と女性のがん種毎の 75 歳未満年齢調整死亡率(人口 10 万人対)の平均とした。説明変数として、野菜摂取量、食塩摂取量、コーヒー購入量を用いた。

3.4.2 重回帰分析

x 軸に重回帰係数、y 軸に p 値をプロットしたものを示す(図 15~図 24)。がん名称の後に、男または女と記していない場合は、男女総数を示す。x 軸の目盛りのラベルの赤字括弧のものは負の値、野菜、食塩、喫煙、飲

酒のプロットは青色で、コーヒー購入量のプロットは赤色で示す。各プロットの右に、変数名と p 値を、図右上に、コホート研究で示されたがん罹患率への優位性（表 2 [12]）の有無を示す。

図 15~図 24 から、重回帰分析によって、コーヒー購入量と、子宮がん、膀胱がん、全がん、大腸がん（女性）、膵がん（男性）、胆道がん、肺がん死亡率との関係が負の相関であり、膵がん（女性）の死亡率との関係が正の相関であることが分かった。コホート研究でリスク低下が示唆されている肺がん、子宮がん、大腸がん（男性）においては、研究の結果を裏付けるものとなり、リスク増大が示唆されている膀胱がんはコホート研究とは異なる結果を示した。また、コホート研究で関連なしとされているものでも全がん、胆道がん、大腸がん（男性）はリスク低下の可能性があるし、膵がん（女性）はリスク増大の可能性があるものとなった。

ここで、コーヒー購入量の p 値が 0.2~0.5 であり、統計学的に優位な結果になっていない原因について考察してみる。コホート研究によると、コーヒーの摂取のリスク低減効果は、一日に飲む回数に応じて、急激に増大することが知られている[7]。一方、日本人でコーヒーを一日に 4~5 回以上飲む人は 10%程度しかいないので[1]、重回帰分析で使ったコーヒーの購入量が増大しても、そのうち 1 日たくさんのコーヒーを飲む人は多くおらず、コーヒー摂取の効果が見えにくくなっているのではないかと考えられる。また、コーヒーの摂取の効果は、数年以上の蓄積があって現れるものと考えられるので、コーヒー購入量の年次変動があった場合は、やはり効果が見えにくくなるものと考えられる。



図 15 子宮がん_死亡率

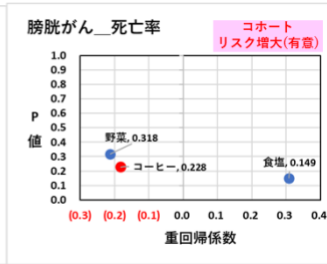


図 16 膀胱がん_死亡率

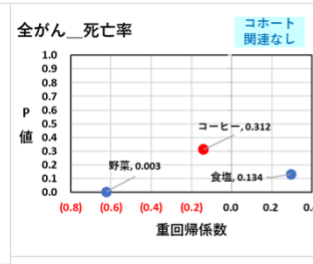


図 17 全がん_死亡率

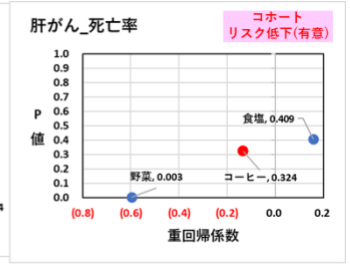


図 18 肝がん_死亡率

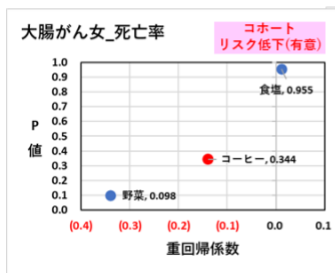


図 19 大腸がん女_死亡率

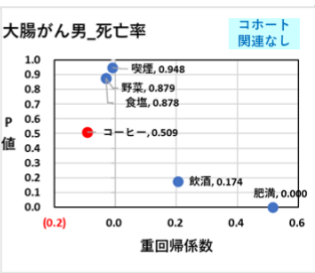


図 20 大腸がん男_死亡率

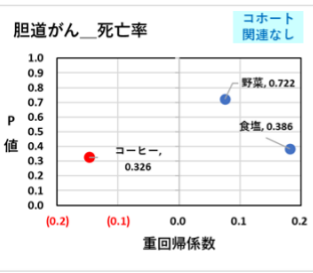


図 21 胆道がん_死亡率

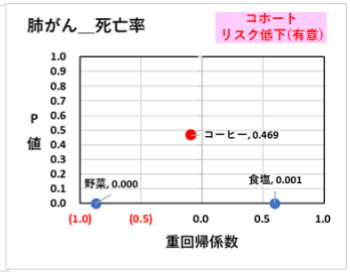


図 22 肺がん_死亡率

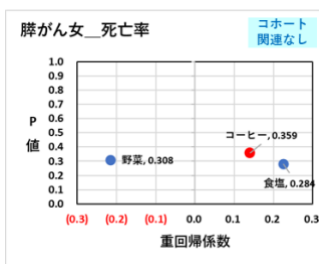


図 23 膵がん女_死亡率

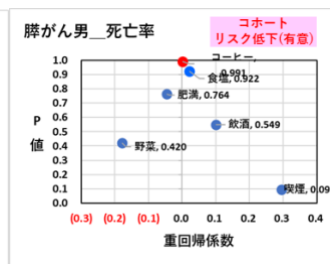


図 24 膵がん男_死亡率

表 2 がんの罹患率に関するコホート研究の結果

記事	期間	対象	人数	疾患	有意性	重回帰分析実施
緑茶・コーヒー摂取と膵がんとの関連について	1990-2003	40-69歳男女	約13万人	膵がん(男性)	リスク低下(有意)	○
				膵がん(女性)	関連なし	○
喫煙、コーヒー、緑茶、カフェイン摂取と膀胱がん発生率との関係について	1993-2005	40-69歳男女	約10万	膀胱がん(非喫煙者)	リスク増大(有意)	○
コーヒー摂取と肺がんリスクとの関連について	1990-2011	40-69歳男女	87,000	肺がん	関連なし	○
緑茶・コーヒー摂取と胆道がん罹患率との関連について	1995-2010	45-74歳男女	約9万	胆道がん	関連なし	○
コーヒー摂取と大腸がんとの関連について	1990-2002	40-69歳男女	約10万人	大腸がん(男性)	関連なし	○
				大腸がん(女性)	リスク低下(有意)	○
コーヒー摂取と全死亡・主要死因死亡との関連について	1990-2011	40-69歳男女	約9万	全死亡	リスク低下(有意)	○
				がん死亡	関連なし	○
コーヒーと子宮体がんの発生率との関係について	1990-2005	40-69歳男女	54,000	子宮体がん	リスク低下(有意)	○
コーヒー摂取と肝がんの発生率との関係について	1990-2001	40-69歳男女	約9万人	肝がん	リスク低下(有意)	○

4. 結論と今後の課題

コーヒーの持つ短期的効果と長期的効果について、それぞれが及ぼす日本人の生活行動への影響、及び疾患による死亡率への影響を統計的に明らかにすることを目的として、コーヒー購入量と、スポーツや趣味などの生活行動との関係、及びがんによる死亡率との関係を、重回帰分析によって求めた。その結果、以下の

ことが明らかとなった。

- 1) 生活行動の 5 つの変数の内、スポーツの総数と趣味・娯楽総数のみが、p 値が 0.05 以下になっていることから、コーヒー購入量との間の関係が統計学的に有意であることが分かった。
- 2) 趣味・娯楽に属する 34 の変数では、12 の変数に選択した後、多くの変数で 0.1 以下、将棋、CD・スマートフォンなどによる音楽鑑賞、編み物・手芸、写真の撮影・プリントで 0.05 以下であり、多くの変数が統計的に有意であることが分かった。
- 3) スポーツに属する 22 の変数では、16 の変数に選択した後、野球、バドミントン、ゲートボール、器具を使ったトレーニングで p 値が 0.05 以下であり、統計的に有意であることが分かった。
- 4) がん死亡率との関係は、p 値は有意でなかったものの、コホート研究でリスク低下が示唆されている肺がん、子宮がん、大腸がん(男性)においては、研究の結果を裏付けるものとなり、リスク増大が示唆されている膀胱がんはコホート研究とは異なる結果を示した。また、コホート研究で関連なしとされているものでも全がん、胆道がん、大腸がん(男性)はリスク低下の可能性があり、膵がん(女性)はリスク増大の可能性のあるものとなった。

今後の課題として、コーヒーの摂取の効果は、数年以上の蓄積があつて現れるものと考えられるので、コーヒー購入量の年次変動を考慮した分析を行う必要がある。

5. 参考文献

- [1] マイボイスコム株式会社(2020)日常生活とコーヒーに関するアンケート結果(第 7 回)
(https://myel.myvoice.jp/products/detail.php?product_id=28802)最終閲覧日：2023-08-31
- [2] Sridhar Ramakrishnan(2014) Dose-dependent model of caffeine effects on human vigilance during total sleep deprivation
(<https://www.sciencedirect.com/science/article/abs/pii/S0022519314002884>)最終閲覧日：2023-08-30
- [3] 矢島潤平, 賀二郎 (2014) コーヒー摂取による作業成績の向上とストレス反応の軽減
(http://repo.beppu-u.ac.jp/modules/xoonips/download.php/gk01608.pdf?file_id=7085)
最終閲覧日：2023-08-30
- [4] チャンバーコーヒー(2021)プロ棋士中村太地様インタビュー
(<https://chamber-coffee.com/column01>)最終閲覧日：2023-08-31
- [5] 全日本コーヒー協会(2016)田中壮【元プロ野球選手】
(<https://coffee.ajca.or.jp/webmagazine/interview/84taguchi/>)最終閲覧日：2023-08-31
- [6] 木村俊博, 伏脇裕一(2019) コーヒーの成分と発がん抑制作用
(https://www.jstage.jst.go.jp/article/safety/58/5/58_310/_pdf/-char/ja)最終閲覧日：2023-08-30
- [7] 斎藤栄子, 津金祥一郎(2015) Association of coffee intake with total and cause-specific mortality in a Japanese population: the Japan Public Health Center-based Prospective Study
(<https://www.sciencedirect.com/science/article/pii/S0002916523273876?via%3Dihub>)最終閲覧日 2023-08-31
- [8] 福山平成大学 多重共線性の意味について
(<https://www.heisei-u.ac.jp/ba/fukui/tips/tip006.pdf>) 最終閲覧日：2023-08-31
- [9] 上田太一郎(1997) 相関があるかを見つける簡便法
(https://orsj.org/wp-content/or-archives50/pdf/bul/Vol.42_07_493.pdf) 最終閲覧日：2023-08-31
- [10] 芳賀敏郎(1976) 重回帰分析における変数選択の新しい規準
(https://www.jstage.jst.go.jp/article/quality/6/2/6_KJ0000320666/_article/-char/ja/) 最終閲覧日：2023-08-31
- [11] 赤池 弘次(1973)."Information theory and an extension of the maximum likelihood principle" (1973).
(https://link.springer.com/chapter/10.1007/978-1-4612-1694-0_15)
- [12] 現在までの成果 | 多目的コホート研究 | 国立がん研究センター がん対策研究所
(<https://epi.ncc.go.jp/cgi-bin/cms/public/index.cgi/nccepi/jphc/outcome/index>)最終閲覧日：2023-08-31