

第一部

調査の概要、標本特性、基礎集計

I 調査の概要と標本の特性

1 調査概要

(1) 背景と目的

令和2年度の「人口変動要因等に関するアンケート調査」、令和3年度の「新型コロナウイルス感染症による港区の在住者、滞在人口の動向と都市機能への影響の実態に関する研究」において、区民と区への来訪者の、区に対する評価における買い物環境の重要性が明らかとなった。具体的には、区民の区に対する総合満足度の規定要因の一つとして、買い物環境満足度が重要であること、また、港区への来訪者についてもコロナ禍後の居住地選択要因として買い物環境の重要性が増していることが明らかとなった(ウェブモニター調査データによる)。さらに、コロナ禍前後の比較に基づくと、滞在人口が減少している地区は飲食業営業許可件数が減少していることが明らかとなった(業務記録統計及びモバイル空間統計データによる)。

これらの結果を受け、買い物環境の内実や飲食サービス業・食料品小売業を取り巻く環境を詳細に把握することが区民福祉の向上のために必要となると考えられることから、令和4年度の調査・研究では区民と事業者の両者を対象としたアンケート調査により、買い物環境の実態を捉えることとした。

上述のような背景を踏まえ、本調査では、買い物環境の改善に向けた対応策の検討材料を得ることを目的とし、飲食サービス業・食料品小売業を営む事業所及び区民の各種実態を把握する。本調査結果を踏まえ、買い物環境の改善のための課題と対応策を検討していく。

(2) 調査対象及び方法

区内の飲食サービス業・食料品小売業及び区民を対象に質問紙調査を実施した(それぞれ「港区の飲食業・小売業の実態と事業展開の方向性に関する調査」、「港区における買い物環境・食事行動に関する実態調査」)。調査票は郵送法により令和4年7月中旬に配布し、8月初旬を期日に回収した(郵送配布及び郵送回収)。調査対象の区民は令和4年5月の住民基本台帳の情報をもとに抽出し、事業所は調査委託事業者の保有する事業者リスト、ぐるなびウェブページ及びタウンページ記載の情報をもとに抽出された。調査票の有効配布数は事業所が4,791件(回収率:16.45%)、区民が4,947件(回収率:22.35%)となっている。調査の概要は図表1-1のとおりである。

図表1-1 標本調査の設計・回収情報

調査主体	調査対象の範囲	標本抽出法	調査法	有効配布調査票数	調査票配布期間	有効回収調査票数	回収率
事業所	港区内の飲食業、小売業の事業所(令和4年5月現在)	調査委託事業者が保有するリスト、ぐるなびウェブページ、タウンページ掲載の飲食・小売業から無作為抽出	郵送法(郵送配布、郵送回収)による、質問紙調査	4,791 ※宛先不明返送分、閉業等での非該当が無効扱い。	7月15日～8月5日	788	16.45%
個人	港区民(令和4年5月13日現在)	住民基本台帳から無作為抽出(地区と年代による層化抽出)	郵送法(郵送配布、郵送回収)による、質問紙調査	4,948 ※宛先不明返送分が無効扱い。	7月15日～8月5日	1,106	22.35%

(3) 調査票の概要と構成

事業所調査と区民調査の調査票の構成概要は次の図表のとおりである。なお、調査票（単純集計値付き）については第Ⅲ部の資料を参照されたい。

図表1-2 事業所調査の調査票の概要

大項目	問番号	質問概要
基本情報	Q1	所在地
	Q2	業種
	Q3	区内での現在の業種としての営業年数
	Q4	事業所規模(面積)
	Q5	従業員規模(人数)
事業の状況	Q6	飲食業のコロナ禍前からの項目別売上の変化
	Q7	食料品小売業のコロナ禍前からの品目別売上の変化
事業展開	Q8	コロナ禍前後の事業展開の方向性
	Q9	飲食業のコロナ禍収束後に力をいれることを考えている部門
	Q10	食料品小売業のコロナ禍収束後に力をいれることを考えている販売形態
	Q11	コロナ禍収束後に力をいれることを考えている商品
	Q12	コロナ禍前後の顧客層の属性
	Q13	今後の事業展開の課題
	Q14	コロナ禍前と比べた物価の変化
	公的サービス	Q15
Q16		コロナ禍の公的支援策の利用の有無(終了した支援策)
Q17		コロナ禍の公的支援策の利用の有無(継続中の支援策)
Q18		制度の利用予定がない理由
Q19		新たに希望する公的支援策の内容
地域とのかかわり	Q20	商店街内への立地の有無
	Q21	商店街連合会への加盟の有無
	Q22	地域からの要望があった場合に対応したい内容
	Q23	本調査、区政への意見

図表1-3 区民調査の調査票の概要

大項目	問番号	質問概要
基本事項	Q1	性別、年齢
	Q2	世帯主の別
買い物環境への満足度	Q3	食料品の買い物環境への満足度
	Q4	日用品・消耗品の買い物環境への満足度
普段の買い物	Q5	品目別購入先・利用頻度、所在地(各上位3つ)
	Q6	最も利用する食料品の買い物先への移手段・所要時間・距離
	Q7	店舗での食料品・日用品・消耗品の買い物時に重視すること
	Q8	店舗での食料品・日用品・消耗品の買い物を主に行う人
	Q9	区による買い物支援サービスの導入希望
食事行動・生活時間	Q10	現在の食事行動
	Q11	コロナ禍前の食事行動
	Q12	生活時間のコロナ禍前後の変化
	Q13	インターネット通販での買い物行動のコロナ禍前後の変化
家計	Q14	世帯収入・支出のコロナ禍前後の変化
	Q15	昨年の世帯総年収
基本属性・世帯情報・働き方	Q16	世帯人数
	Q17	婚姻状況
	Q18	同居者続柄・年齢・性別・職業
	Q19	国籍
	Q20	居住地区
	Q21	居住年数
	Q22	港区転入前の居住地
	Q23	住宅の保有形態
	Q24	情報機器の保有内容
	Q25	職業
	Q26	勤務先の事業内容
	Q27	週労働時間(本人・配偶者)
Q28	最終学歴(本人・配偶者)	
考え・意識	Q29	コロナ禍前と比べた暮らし向きの変化
	Q30	コロナ禍前と比べた物価の変化
	Q31	近隣商店街への期待
	Q32	5年後の希望居住地
	Q33	1万円をもらった時の貯蓄額
	Q34	政策志向
	Q35	税金に対する考え
スマイル商品券	Q36	スマイル商品券の利用の有無
	Q37	利用したスマイル商品券の媒体
	Q38	スマイル商品券の購入前後の近隣商店街の利用頻度
	Q39	スマイル商品券を利用して購入した品目
	Q40	調査・区政に対する意見

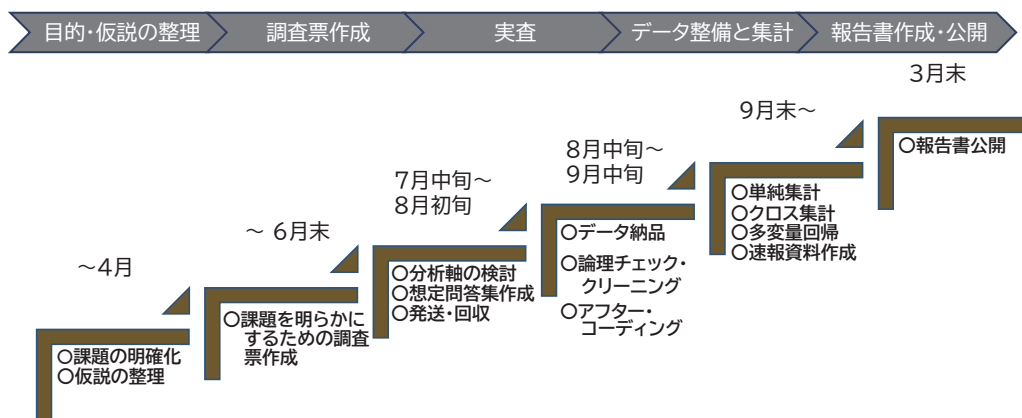
2 調査研究のプロセス

(1) スケジューリング・活動記録

本調査研究の企画、実査、集計、報告書作成は図表1-4の工程のとおりに進められた。まず、4月に課題の明確化と仮説の整理を行ったのちに調査票の作成を進めていった。次に、調査対象の抽出を行い、7月中旬に対象者宛に調査票を配布し、8月初旬に回収した。9月以降に、後述の値の再割当て、論理チェックによるデータクリーニングを行い、基礎集計を進め、報告書執筆に取り組んだ。

個別のテーマの分析については、相互に意見交換をする研究会を12月から各月で開催し、計3回開催した。個別テーマの分析結果については、第Ⅱ部の各章を参照されたい。

図表1-4 調査企画から報告書作成までの経過



(2) 想定問答集と問合せ記録

本調査では、調査票配布後の調査対象者からの問合せに備え、想定問答集を作成し、問合せに対応した。想定問答集及び実際の問合せ記録については第Ⅲ部の資料を参照されたい。

3 回答データのクリーニング

質問紙調査の回答データは、次のような理由でデータの編集やクリーニングが必要となる。

- (イ) 元来の調査票の設計（コード化が必要な自由記入の設問等）
- (ロ) 回答者の誤回答や回答漏れ
- (ハ) 調査票情報の入力委託業者の誤入力、入力漏れ

そのため、分析の前に回答データの点検を丁寧に行い、必要に応じて回答原票を確認しながらデータの編集を行った。

(1) アフターコーディング

ア 自由記入回答の「港区への転入前の居住地」と「5年後の希望居住地」のコード化

区民調査の設問には、港区への転入前の居住地及び5年後の居住希望地を自由記入形式で尋ねるものがある。転入前の居住地を尋ねる設問は以下のとおりである。

他の地域から港区へ転入してこられた方に伺います。

あなたが港区に転入する直前に居住していた地域を教えてください。（場所を記入）なお、海外の場合は国名を教えてください。

都道府県名		市区町村名		海外	国名：	
-------	--	-------	--	----	-----	--

転入前の居住地及び5年後居住希望地の設問のいずれも、上の例のとおり、都道府県、市区町村、国名（海外の場合）を記入する欄が3つ並ぶ自由回答形式となっている。これらの自由記入回答のテキスト情報をコード化する作業を行った。

具体的には、自由記入回答の地域に対してコードを割り振り数値化した。第III部の資料の単純集計値付きの調査票では、「東京都港区」、「東京都特別区部（港区以外）」、「東京都特別区部以外の関東4都県」、「その他」に区分けした集計値を掲載している。

イ 選択式設問の「その他」の回答への値の再割当て

データの精度を高めるため、選択式設問における「その他」の選択肢の自由記入欄の記述内容を精査し、より妥当な選択肢が存在している場合は事後的に同値を割当てる処置を施した（アフターコーディング/リコーディング）。

より具体的には、選択式設問における「その他」の選択肢の自由記入欄の記述内容には、すでに選択肢にある内容を記入している場合が多くみられた。

このことは、

（イ）回答者の不注意により選択肢を見過ごした場合

（ロ）調査票の選択肢が回答者の認識する概念に該当しない場合

（ハ）調査票の文言がわかりづらかったり、類似する選択肢が複数含まれることから回答者が判断に迷い、その他の自由記入欄に記入した場合（調査票設計の問題）

などの理由で生じることが考えられる。

そのため、所内の3名の担当者により、選択式の設問の「その他」の自由記入回答の内容を吟味し、必要に応じて値の修正を行った。なお、3名の解釈が分かれた場合は合議により割当て後の値を決定した。

本調査データにおける値の再割当ての例及び実際に再割当てを行った設問・件数については第III部の資料を参照されたい。

（2）複数選択式の設問における無回答の処理

回答選択肢の中の当てはまるもの全てを選択させる形式で質問をする複数選択式の設問については、いずれの選択肢も選んでいない場合に無回答として処理した。本調査の複数選択式の設問には、無回答と非該当が識別できるよう、全てに「その他」の選択肢が含まれている（項目によっては「該当しない」に相当する選択肢が含まれている）。そのため、複数選択式の設問で選択肢が一つも選択されていない場合に無回答として処理し、集計時の分母から除いた（複数選択式の設問は、無回答を識別するか否かで集計時の分母が異なる）。

（3）論理の整合性チェック

次に、回答データの論理整合についてのクリーニングを行った。

事業所調査については、図表1-5のとおり、営業年数の回答を用いてコロナ禍前の状況についての回答に制約を設けた。具体的には、営業年数が「2年未満」の場合は、コロナ禍前の事業の状況については「非該当」となるはずであるため、「非該当」の値を割当てた。

区民調査については、図表1-6のとおり、婚姻状況、就業状況、商品券利用の設問の回答を用いて、回答の対象か否かの制約を受ける設問については制限を設け、回答対象でない設問は「非該当」の値を割り当てた。例えば、婚姻状況が「既婚（事実婚含む）」以外の場合は、Q12の配偶者の生活時間のコロナ禍前後の変化についての設問では「非該当」（相当）の選択肢を選ぶはずであるため、同項目を選択していなくても「非該当」（相当）の値を割当てた。

図表1-5 事業所調査データの論理制約

制限の条件	対象の設問
Q3. 営業年数が「2年未満」	Q4_2. コロナ禍前の延床面積、Q5. コロナ禍前の従業員数、Q6. コロナ禍前からの飲食業の売上変化、Q7. コロナ禍前からの食料品小売業の売上変化、Q8. コロナ禍前の事業展開の方向性、Q12. コロナ禍前の顧客層・属性を「非該当」に

図表1-6 区民調査データの論理制約

制限の条件	対象の設問
Q17. 婚姻状況が「既婚(事実婚含む)」以外(「未婚」「離別・死別」)	Q12_イ. 配偶者が行う家事、Q12_オ. 配偶者が行う在宅勤務・テレワーク、Q27b. 配偶者の労働時間、Q28b. 配偶者が最後に卒業した学校を「非該当」に
Q26. 事業内容が「働いていない」	Q25. 何らかの職業を選択していても「11. 無職」に
Q36. スマイル商品券の利用が「いいえ」	Q37～Q39の商品券利用者対象の設問は「非該当」に

(4) その他

世帯所得の設問について、値の再割り当てを行った。具体的には、Q15の世帯所得についての設問の選択肢には「100万円以上200万円未満」と「800万円以上1000万円未満」の選択肢はないが（調査票設計の誤り）、同金額を調査票の欄外に記載する回答がみられた。そのため、欄外記載のあった回答については、アフターコーディングにより同2区分の選択肢の値を割り当てた。

そのほか、世帯人数が10名以上の場合、親の年齢よりも本人の年齢が高い場合、夫婦の性別が同一の場合などの回答については特異値としてデータから除いた。

4 標本の位置づけと特性

本節では、母集団である港区の特徴を特別区部との比較から示すとともに、各種統計・業務記録統計と本調査の標本との分布の比較を通して、標本の偏りの傾向をみていく。

なお、事業所の個別回答を所在地区別に集計した上で、母集団との地域分布の違いを確認する（個別データと集計データの違いについてはコラム①を、公開されている母集団データへのアクセス方法についてはコラム②を参照）。母集団と本調査の標本の分布を比較するのは、第Ⅱ章以降において標本から母集団の情報を捉える目的で、統計的推定や検定を行うためである（推定と検定については、コラム③を参照）。

(1) 事業所調査

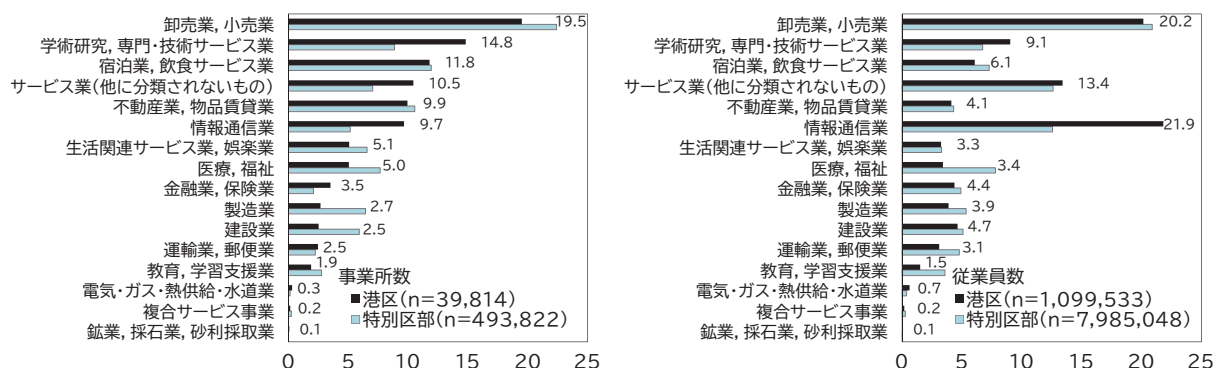
まず、産業特性に係る統計（事業所数・従業員数）から、港区における卸売・小売業と宿泊・飲食サービス業の位置づけを把握したのちに、区内の飲食サービス業・小売業の地域分布の母集団（業務記

録統計)と標本情報の比較を通して、事業所調査の標本の特性をみていく。

ア 港区の産業特性

- ・事業所数をみると「卸売業、小売業」と「宿泊業、飲食サービス業」は港区の主要産業
- ・「宿泊業、飲食サービス業」は事業所数、従業員数ともにコロナ禍前からの減少幅が大きい

図表1-7 港区と特別区部の業種別事業所数と従業員数の構成比 (2021年)



出所：総務省「令和3年経済センサス-活動調査 速報集計」より作成。

注：図中の値は港区の数値。

図表1-7は、経済センサス統計より、港区と特別区部の事業所と従業員の業種別構成比を示したものである。左の図では、港区は約3万9800カ所、特別区部は約49万3800カ所の事業所に占める各業種の割合を、右の図では、港区は約100万人、特別区部は約800万人の従業員数に占める各業種の割合を示している。

港区は、事業所の構成比で見ると、「卸売業、小売業 (19.50%)」、「学術研究、専門・技術サービス業 (14.84%)」、「宿泊業、飲食サービス業 (11.80%)」が多く、従業員の構成比で見ると、「情報通信業 (21.85%)」、「卸売業、小売業 (20.18%)」、「サービス業 (他に分類されないもの) (13.42%)」が多くなっている。

本調査の対象となる、飲食サービス業と小売業 (統計区分上は、それぞれ「宿泊業、飲食サービス業」、「卸売業、小売業」) は事業所数で見ると両者とも上位に位置し、従業員数で見ると前者は中位、後者は上位に位置している。両業種は港区において重要な位置を占めている。

図表1-8 港区と特別区部の業種別事業所数と従業員数の変化 (2016年~2021年)

事業所数の増減	特別区部		港区		従業員数の増減	特別区部		港区	
	増減	割合	増減	割合		増減	割合	増減	割合
全体	-	-0.1	-	6.8	-	5.5	-	10.0	
卸売業、小売業	-9.3		-1.9		-1.2		-2.3		
学術研究、専門・技術サービス業	-	19.2	-	18.6	-	19.6	-	17.7	
宿泊業、飲食サービス業	-21.1		-24.8		-20.5		-21.8		
サービス業(他に分類されないもの)	-	10.5	-	10.2	-	9.1	-	23.5	
不動産業、物品賃貸業	-	13.6	-	24.7	-	11.3	-	11.6	
情報通信業	-	21.9	-	20.0	-	19.3	-	20.9	
生活関連サービス業、娯楽業	-7.4		-	0.8	-	1.2	-	29.5	
医療、福祉	-	7.0	-	9.7	-	10.5	-	6.9	
金融業、保険業	-	13.3	-	30.3	-	4.2	-	21.6	
製造業	-15.0		-0.8		-2.7		-	10.0	
建設業	-	3.8	-	3.3	-	6.9	-	7.4	
運輸業、郵便業	-15.9		-5.4		-	2.9	-	-0.4	
教育、学習支援業	-	6.6	-	8.7	-	8.1	-64.8		
電気・ガス・熱供給・水道業	-	53.5	-	54.7	-	38.4	-	26.5	
複合サービス事業	-	1.1	-	3.1	-		-	5.2	
鉱業、採石業、砂利採取業	-8.3		-34.5		-4.5		-16.3		

出所：総務省「令和3年経済センサス-活動調査 速報集計」「平成28年経済センサス-活動調査」より作成。

注：「全体」は図表中の業種の総計。

続いて、業種別の事業所数と従業員数の2時点間の変化をみていく。図表1-8は、港区と特別区部の業種別事業所数と従業員数の実数の、2016年から2021年にかけての変化率を示したものである。左は事業所数、右は従業員数の実数の変化率を示している。

事業所数でみると、港区は「鉱業、採石業、砂利採取業」(-34.5%)と「宿泊業、飲食サービス業」(-24.8%)の減少幅が顕著である。従業員数でみると、「教育、学習支援業」(-64.8%)と「宿泊業、飲食サービス業」(-21.8%)の減少幅が顕著である。

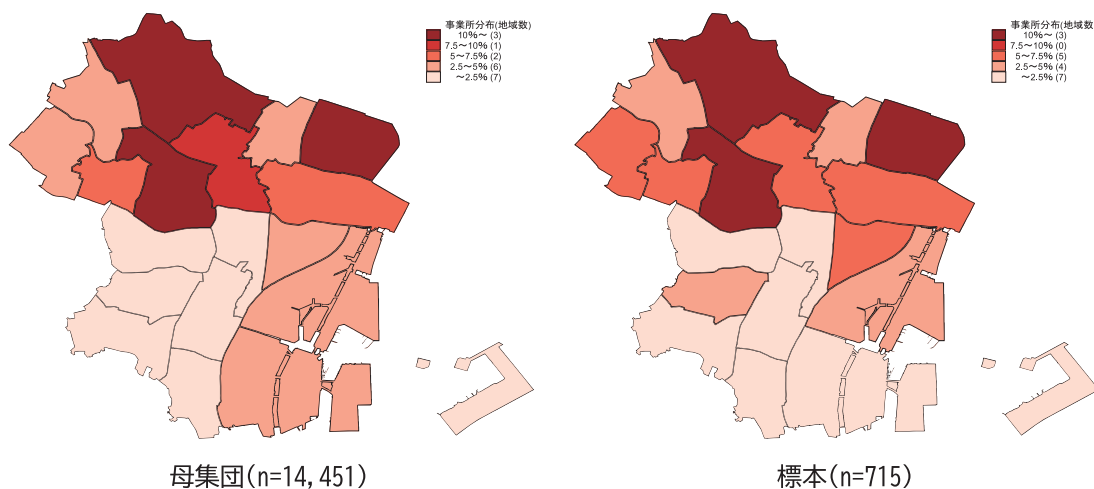
本調査の調査対象となる飲食サービス業と小売業の2時点間の変化の傾向はそれぞれ異なるが、業種内でみると、事業所数、従業員数のいずれも連動する傾向が示されている。具体的には、図表1-8の「卸売業、小売業」は事業所数と従業員数の増減が少なく、「宿泊業、飲食サービス業」は両値の大幅な減少が示されている。港区において重要な位置を占める飲食産業は、コロナ禍により大きく影響を受けたことが示されている。

次に、港区の飲食業営業許可の業務記録情報と本調査の事業所標本の地区別の分布比較から標本の偏りの傾向を確認する。

イ 区内事業所の母集団と標本の分布の相違

飲食サービス業・小売業事業所の地域分布は母集団と標本とで酷似している

図表1-9 港区の母集団と標本の飲食サービス業・小売業の事業所の地域分布



出所：母集団は飲食業営業許可の業務記録統計（2023年3月現在）より作成。

注：（1）飲食業営業許可統計は、同一事業所で複数の営業種別の登録がある場合に複数カウントとならないよう、事業所名、住所、営業種別の情報をもとに固有事業所を識別し、各事業所につき1事業所のみを残した。ただし、廃業したにもかかわらず廃業届を提出していない場合は統計に含まれる。

（2）次に該当する事業所は母集団データより除かれている。食品製造・加工業の専業事業所、自動販売機、移動販売車、小学校・中学校・高等学校の学生食堂、企業の社員食堂。

続いて、本調査の事業所標本の偏りを地域分布の傾向からみていく。図表1-9は、母集団と本調査の事業所標本データにおける、飲食サービス業・小売業の事業所の区内19地域の分布を示したものである。色の濃い地域ほど、飲食サービス業・小売業の事業所の分布割合が高いことが示されている。

母集団と標本のいずれでも、飲食サービス業・小売業事業所の分布割合が最も高い地域は「新橋」、「麻布十番・元麻布・六本木」、「赤坂」の3地域となっている。なお、全事業所のうち飲食サービス業の事業数が多くを占めることから、同3地域に飲食サービス業が多いということが反映されている。

本調査の標本については、母集団の地域分布と相似の傾向を示している。具体的には、区内19地域の事業所分布の傾向は母集団と標本とで同様の傾向を示しており、相関係数を求めると強い正の有意な相関を示している (r=0.9714)。

(2) 区民調査

次に、性別・年齢別人口の基本属性情報に基づいて、本調査の区民標本の特性をみていく。

ア 区民調査の母集団と標本の性別・年齢分布の相違

- ・港区は特別区部に比べ高齢者の割合が低い
- ・区民標本は母集団に比べ高齢者層がやや多く、中でも中高年女性の構成割合が高い

図表1-10 母集団と回収標本の性別・年齢の分布

	(参考)特別区部				港区母集団				標本				標本-母集団構成比 差分	
	人数		構成比		人数		構成比		人数		構成比			
	男性	女性	男性	女性	男性	女性	男性	女性	男性	女性	男性	女性		
18歳~29歳	704,754	709,151	0.084	0.085	15,489	15,940	0.071	0.073	52	52	0.049	0.049	-0.022	-0.024
30歳~39歳	709,115	692,153	0.085	0.083	19,612	22,464	0.089	0.102	57	96	0.054	0.090	-0.036	-0.012
40歳~49歳	821,869	756,851	0.098	0.090	22,618	25,443	0.103	0.116	66	141	0.062	0.133	-0.041	0.017
50歳~59歳	760,481	691,987	0.091	0.083	19,636	21,020	0.089	0.096	69	129	0.065	0.121	-0.025	0.025
60歳~69歳	460,326	453,472	0.055	0.054	10,804	11,716	0.049	0.053	64	97	0.060	0.091	0.011	0.038
70歳~79歳	423,789	491,798	0.051	0.059	8,630	11,216	0.039	0.051	64	101	0.060	0.095	0.021	0.044
80歳以上	248,877	457,036	0.030	0.055	5,017	9,913	0.023	0.045	35	41	0.033	0.039	0.010	-0.007
合計	8,381,659		1		219,518		1		1,064		1			

出所：人口統計は各区のホームページの公開統計より作成。

注：特別区部と港区は令和4年5月1日時点の年齢別人口。ただし、特別区部の人口うち、荒川区、北区は同年1月時点、台東区、太田区、文京区は同年4月時点の数値。千代田区は公開されている年齢区分の制約により、15歳~19歳人口の半数を18歳~19歳人口とみなして集計。

図表1-10は、母集団である港区人口と本調査の区民標本データにおける、性別・年齢別の分布情報を示したものである（特別区部は参考値）。標本の特性をみるための母集団と区民標本の性別・年齢別構成比の差分ポイントは、右の列に示した。

まず、特別区部の年齢分布との比較から港区の特徴をみると、港区は特別区部に比べ高齢者層が少なく働き盛り層が多い傾向にある（男女計60歳以上人口の構成比：港区は26.10%、特別区部は30.25%）。

本調査の区民標本については、母集団よりも高齢者層の構成比が高い傾向にある（男女計60歳以上人口の構成比：港区は30.25%、区民標本は37.78%）。右の列をみると、母集団人口よりも若年・中年男性の構成比がやや低く、中高年女性の構成比がやや高い傾向にある。

この傾向と関連する背景として、回収標本は、

(イ) 就業等で在宅時間が少ない男性が比較的に多いこと

(ロ) 専業主婦等の在宅時間が比較的に長い女性の回答が多いこと

(ハ) 世帯主男性宛に送られた調査票を配偶者が回答していることが多いこと（本調査では宛名本人の回答に限るというご案内はしていない）

などの理由が考えられる。

イ 区民調査の母集団と標本の所得分布の相違

- ・港区は特別区部に比べ「1000万円以上」（個人所得）の構成比が高い
- ・区民標本は母集団に比べ「200万円未満」の構成比が低い（個人所得と等価所得の比較）

図表1-11 港区の母集団と標本の所得の分布

	(参考)特別区部	港区母集団	標本	標本-母集団構成比 差分
人数	5,221,927	143,239	979	
200万円未満	52.63	35.65	11.24	-24.41
200～700万円未満	38.59	39.95	46.88	6.93
700～1000万円	3.98	7.96	15.63	7.67
1000万円以上	4.80	16.44	26.25	9.81

出所：母集団は東京都総務局行政部政課(2022)「令和3年度市町村税課税状況等の調(特別区関係)」より作成。

注：母集団は住民税の算定の基礎となる令和3年分(2021年1月～12月)の個人所得の課税標準額、標本は2021年分(2021年1月～12月)の額面の世帯収入を世帯人数の平方根で除した値(等価所得額)に基づく。

続いて、所得情報に基づいて標本の特性をみていく。図表1-11は、母集団である区民と本調査の区民標本データの額面所得の分布情報を示したものである。大きな注意点としては、母集団の統計は税務記録によって把握した正確な「個人所得」情報に基づくものであるが、標本の統計は「世帯所得」を階級値で尋ねた質問に基づくものであるため、厳密な比較は難しい点があげられる。加えて、母集団の指標は個人所得であるのに対し、標本は世帯所得を世帯人数の平方根で除した「等価所得」となっている。等価所得は消費行動を左右する家計の実態を反映している指標であり、個人所得とは厳密な対応関係がない点に留意されたい。

参考値である特別区部の所得分布との比較の視点から母集団の特徴をみると、港区の200万円未満の構成比は35.65%と特別区部の52.63%に比べ小さく、1000万円以上の構成比は16.44%と特別区部の4.80%に比べて大きい(23区中、200万円未満の構成比は千代田区と中央区に次ぐ値の小ささであり、1000万円以上の構成比は最大の値)。

200万円未満の所得層には年金生活者が多く含まれると考えられることから、港区において200万円未満の構成比が小さいことには、現役世代人口が比較的多い区の特徴が反映されていると考えられる。さらに、配偶者控除等の税制への対応として就労を調整している専業主婦・主夫などの所得の制約を受けている層の存在も想定が可能である。これらの点から、港区の200万円未満の層には、所得の低い現役世代労働者や年金生活の高齢者、税制への対応として就労調整を行っている層が含まれていると推察される。

本調査の区民標本については、比較可能性についての留意は必要だが、図表1-11の右の列のとおり、母集団よりも200万円未満の所得層が大幅に少なく200万円以上の層の構成比が全体としてやや大きい傾向にあり、なかでも1000万円以上の層の構成比が大きくなっている。

なお、次章以降、とりわけ第V章で扱うデータの要約の方法については、コラム④を参照されたい。次章以降では、事業所調査、区民調査のデータを用いて集計及び分析を行っていく。

本報告書の①～⑩のコラムは、各章の補足情報及び実務に役立つ情報をまとめている。データ分析の知識が浅い方でも読み進められるよう、詳細の説明ではなく、基礎的な事柄の概要や実践手順の紹介に努めている。各コラムで示している文献の選定に際しては、一部の専門書を除き、知識の浅い方でも理解を深められる工夫がなされたものに限っており、各種コラムの概説や例示は参考文献に多く依拠している。いずれの文献も実務レベルで直観的な理解を促す有用なものであるため、必要に応じて各種文献を入手・参照されたい。

コラム① 個票データと集計データ

本報告書では、アンケート調査を基に収集した、区民や事業所の行動や意識についてのデータを集計及び分析しており、集計及び分析には個票データと集計データを利用している。

個票データは、1人の区民や1つの事業所の調査主体単位のままの回答データを指す。個票データを基に、設問ごとの単純集計を行ったり、基本属性と各設問とのクロス集計を行ったりするのが、第二章と第三章である。なお、本調査の個票データは、秘匿性を高める処置を施した上で、オープンデータとして公開している（URL：https://opendata.city.minato.tokyo.jp/dataset/kaimonokaizen_chosa）。

集計データは、1人の区民や1つの事業所の回答データを地域単位で集計したデータを指す。区民や事業所の回答を区内19地域別に集計したデータを利用して、地域単位の各種指標間の相関関係を分析したのが第V章である。

補足をすると、行政が公開する代表的な経済統計の多くがアンケート調査の個票データを集計して構築した統計となっている。例えば、景気に遅行する指標として知られる（完全）失業率は、個人を対象としたアンケート調査である、総務省統計局の労働力調査を用いて地域別及び国全体の指標が構築されている。指標の大まかな算出概要は以下のとおりであり、分子は「失業者」、分母は「労働力」に該当する（百分率にする際には100を乗ずる）。

$$\text{失業率} = \frac{\text{現在無職であり、かつ求職活動をしている人口}}{\text{現在無職であり、かつ求職活動をしている人口} + \text{現在就業している人口}}$$

アンケート調査に対する個人の回答である個票データを基に、上の式によって地域や国全体の失業率が算出され、基幹統計の1つとして公開される。個票データと集計データは、それぞれ英語のマイクロデータ（micro-data）とマクロデータ（macro-data）に対応しており、この語の方が直観的に理解しやすいかもしれない。

コラム図表1 個票データと集計データの実例

区民調査の個票データ

A	B	C	D	E	F	G	H	I	J	
1	回答ID	性別	年齢階層	世帯主	価格	品揃え	距離	営業時間	情報発信	電子決済
2	id	q1.1	q1.2	q2	q3.1	q3.2	q3.3	q3.4	q3.5	q3.6
3	1	1	4	1	1	3	1	1	1	3
4	2	2	6	2	3	2	3	1	2	1
5	3	2	5	2	3	3	1	1	3	2
6	4	1	11	1	3	2	3	3	3	3
7	5	2	4	2	2	2	1	1	2	1
8	6	1	8	1	2	3	2	2	2	1
9	7	1	9	1	2	4	4	3	3	2
10	8	2	9	1	3	3	3	2	3	2
11	9	2	7	2	3	4	2	1	1	1
12	10	2	10	2	3	3	2	2	3	3
13	11	2	11	1	4	2	1	3	3	3
14	12	1	11	1	3	3	2	3	3	3
15	13	2	6	2	2	2	2	2	2	2
16	14	1	10	1	3	3	2	3	2	2
17	15	1	10	1	3	2	1	1	1	1
18	16	2	4	1	4	2	1	1	1	1
19	:	:	:	:	:	:	:	:	:	:

区民及び事業所調査の19地域別の集計データ

A	B	C	D	E	F	G	H	I	J	K	L	
1	地域ID	total_sale	food_sale	date_sale	shiftdays	shiftdays	namein	inshoku_dome	dateofsy_yp	emp_dome	premium	segin
2	A	41.94	12.23	11.07	3.70	26.13	12.90	93.40	9.30	32.00	38.71	503.69
3	B	47.02	12.36	10.97	13.64	29.41	8.33	90.08	36.84	43.33	36.29	616.53
4	C	54.84	11.56	11.44	0.00	24.26	5.19	89.47	12.50	39.22	46.29	545.38
5	D	41.67	11.32	12.00	1.58	32.13	9.69	100.00	32.14	21.78	34.55	498.30
6	E	47.69	16.80	16.97	13.21	32.88	15.94	100.00	20.00	38.46	43.48	707.30
7	F	39.22	11.49	10.73	1.55	30.89	13.73	73.33	30.77	58.00	32.69	685.15
8	G	52.70	12.28	11.42	8.20	29.07	5.88	87.72	26.00	34.85	51.76	764.64
9	H	56.10	9.26	9.59	2.19	41.07	8.51	82.33	33.33	14.29	26.42	645.78
10	I	33.58	8.18	11.00	9.52	20.00	18.00	88.34	19.23	18.42	50.00	509.97
11	J	43.55	16.52	16.42	13.85	29.23	9.38	86.75	19.44	35.42	36.62	704.41
12	K	46.72	16.20	16.28	10.00	27.21	9.52	95.91	55.00	40.00	42.86	678.41
13	L	65.85	12.22	11.93	6.38	36.30	15.72	84.00	28.57	29.03	36.30	829.36
14	M	47.27	11.63	11.35	15.29	41.94	8.42	100.00	12.50	25.00	47.46	810.42
15	N	63.04	11.90	12.02	5.56	22.64	5.80	100.00	100.00	40.00	41.51	684.96
16	O	45.65	11.69	11.75	1.14	30.19	1.55	76.47	14.29	27.27	54.72	778.77
17	P	60.00	16.94	16.57	1.27	27.27	9.23	71.43	14.67	4.00	52.77	720.15
18	Q	41.67	11.42	10.75	13.24	32.00	12.50	85.71	30.77	33.33	40.00	718.57
19	R	45.45	9.86	9.15	9.52	22.11	4.55	83.33	80.00	50.00	25.27	734.64
20	S	36.77	16.36	11.08	12.50	29.63	11.11	66.67	6.00	65.71	30.77	701.52

コラム図表1の左は区民調査の個票データ（A列：ID16番目の行までの表示）、右は区民及び事業所調査の19地域別の集計データ（A列：地域IDのA～Sの全行表示）の実例である。

どのようなタイプのデータを利用するべきかについては、分析者の問いや捉えたい内容に依存する。調査の個体レベルでの意識や行動の詳細のパターンや関係性を明らかにしたい場合は個票データが、地域等の単位の集計指標レベルの現象間の関係性等を捉えたい場合は集計データが適している。

なお、実務的な視点からすると、区民を対象とした政策の効果高精度高く析出したい場合は、個票データが必要となる。詳細は割愛するが、必要となるデータは、政策の対象となった群とならなかった群の政策実施前後の複数時点の追跡データである。日本国内では利活用が限定的であるが、調査に限らず、業務の中で蓄積される行政記録統計（転居記録、税・所得情報、事業所の各種届出等）の利活用により構築された区民や事業所の同一主体の追跡データが、政策効果の検証のための有用な資源となる。政策効果の検証の難しさについての具体例を交えた概説は伊藤（2017）を、政策効果の識別のための技法のわかりやすい概説は山口（2020）を参照されたい。

参考文献

- 伊藤公一朗（2017）「なぜデータから因果関係を導くのは難しいのか」『データ分析の力 因果関係に迫る思考法』光文社、第1章、pp.27-52
- 山口慎太郎（2020）「データ分析で社会を変える《実証ミクロ経済学》」市村英彦・岡崎哲二・佐藤泰裕・松井彰彦編『経済学を味わう』日本評論社、第4章、pp.73-92
- ➡両文献ともに本コラムの内容を超える、そして本報告書では扱えていない、正確な政策評価のための技法が一般向けにわかりやすく概説されている。

コラム② 各種データへのアクセス

(1) 各種データサイト

① e-Stat

全国を対象とした公的統計調査の集計データは、e-Stat 政府統計の総合窓口のウェブサイトからアクセスすることができる（URL: <https://www.e-stat.go.jp/>）。コラム図表2のe-Stat トップページから、分野別に統計の検索を進めることや、キーワードを入力し統計を検索することができ、統計はCSV等の形式でダウンロードすることができる。

市区町村別統計や時系列統計が公開されているものも多くある。例えば、労働力調査に基づく地域別失業率の長期時系列の統計が入手できる。第I章では、港区と特別区部の業種別の事業所数と従業員数の統計の比較を行っているが、同データはe-Statより入手した「令和3年経済センサス 活動調査」の速報集計値に基づいている。

コラム図表1 e-Statのトップページ



公的統計調査の個票データについては、統計作成という本来の目的以外の利用・提供は制限されているが（統計法第40条）、個票データ（調査票情報）の提供が例外的になされる（自治体職員は統計法第33条第1項第1号に該当）。各種手続きののちに、調査票情報の提供（実質的には貸与）を受ける形となる。

全国の自治体の独自調査についても、統計（集計データ）については公開が進んでいるものの、個票データへのアクセスは限定的となっている。

② SSJ データアーカイブ

コラム図表3 SSJデータアーカイブのデータ検索ページ

調査番号	調査名	寄託者 (寄託者名称)	トピック	Nesstar
1560	グローバル就業支援・成長豊潤調査 (18ヶ国・地域データ) , 2022	パーソル総合研究所	経済・産業・経営 国際比較・外文 雇用・労働	--
1475	勤労者の仕事と暮らしについてのアンケート, 2021.10	連合総合生活開発研究所	雇用・労働	--
1355	勤労者の仕事と暮らしについてのアンケート, 2020.4	連合総合生活開発研究所	雇用・労働 社会・文化	--
1443	男女共同参画の視点からの新型コロナウイルス感染症拡大の影響等に関する調査, 2020	内閣府男女共同参画局	雇用・労働	--
1429	全国就業実態パネル調査, 2021	リクルートワークス研究所	雇用・労働 社会・文化	--

一方で、民間組織の調査や学術調査及び一部の行政機関の調査の個票データは、東京大学社会科学研究所 SSJ データアーカイブのページからオンライン上で、所定の手続を経ることでアクセスが可能となっている。コラム図表3は、SSJ データアーカイブのデータ検索ページで実際にデータ検索をした結果となっている（URL：<https://ssjda.iss.u-tokyo.ac.jp/Direct/datasearch.php>）。キーワードを入力し、調査年の新しい調査の順に検索結果を表示させている。表示される調査名をクリックすることで調査概要のページにつながり、調査のデザインの情報に加え、調査票や基礎集計結果へのアクセスができるようになっている。

個票データの利用は、大学又は公的機関の研究者等に限られるものの、上述のとおり、調査概要、基礎集計結果及び調査票等の情報についてはオープンアクセスとなっている。同データアーカイブに登録されているのは、個人・世帯や事業所・団体を対象としたアンケート調査のデータであり、約2,450件のデータへのアクセスが可能となっている（令和5年3月現在）。

③港区オープンデータカタログサイト

港区の各種集計データ（港区行政資料集に掲載しているデータ等）については、オープンデータのウェブサイトからアクセスすることができ（URL：<https://opendata.city.minato.tokyo.jp/dataset>）、データはCSV形式でダウンロードすることができる。

集計データのみならず、本調査を含め、一部の調査の個票データのオープンデータとしての公開が進められている。

(2) 調査設計に際しての既存調査情報の活用

自治体等でアンケート調査を行う際には、SSJ データアーカイブの活用が有効となる。既存調査の設計デザインや調査票を参考にすることにより、調査実務の効率と質を高めることができるためである。本調査の調査票設計に際しても、複数の既存調査の調査票を参照している。

調査実務の改善のため方法として、具体的には、調査票の作成・改訂時に、SSJ データアーカイブに格納されている既存調査の調査目的と調査対象の選定方法、標本の抽出方法、設問項目の概要、主たる目的を明らかにするためにデザインされた質問の文言等を参照し、参考になる点を取り入れることができる。新規の調査の設計時のみならず、所管課が継続的に実施している調査についても批判的に点検し、調査の効率と収集するデータの質の改善を意識することが必要となる。なお、SSJ データアーカイブに登録されている調査の情報だけではなく、必要に応じて、e-Stat で公開されている全国を対象とした公的統計調査のアンケート票情報を併せて参照されたい。

既存調査の調査票の活用の利点は他にもある。アンケート調査の設計において、既存の調査と同様の設問を取り入れることで、得られた回答データから他の地域と比較した、自治体の特性を示すことが可能となる。

参考文献

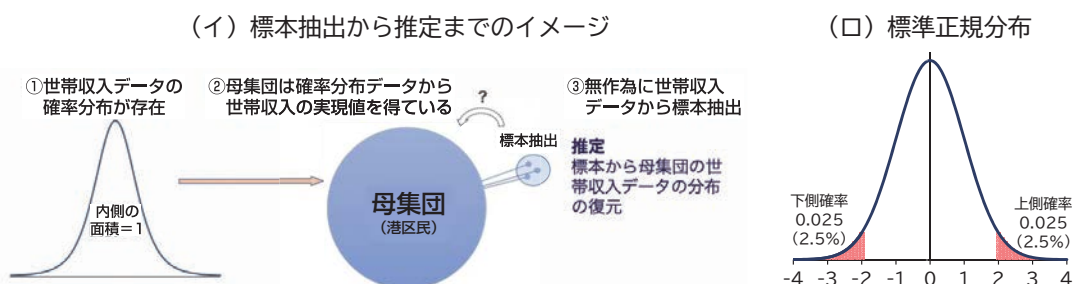
- 畑農鋭矢・水落正明 (2022) 「データに親しむ」『データ分析をマスターする12のレッスン 新版』有斐閣、第2章、pp.45-73
- ➡第2章の第3節以降で各種データの探し方について、実用的な情報とともに解説されている。

コラム③ 推定と検定

(1) 推定

本調査においては、母集団である港区の人口から、5,000人の調査対象の標本を住民基本台帳に基づいて無作為に抽出（厳密には地区や年代で層化無作為抽出）している。データ分析のプロセスとしては、コラム図表4（イ）のとおり、標本データを用いて母集団の値（平均や分散）を推し定めたり、仮説の検証を行っている。

コラム図表4 推定のイメージと正規分布



ただし、詳細の理解は必要ないが、標本の捉え方はデータの分布を基にしていることに注意が必要となる。発想の手順としては、例えば、以下のように考える。

①まず、「世帯収入」という釣鐘型の正規分布にしたがう「確率的情報」が存在していると考え（曲線の内部の面積が1となる確率分布。コラム図表4（ロ）は最もシンプルな、平均0、標準偏差1の標準正規分布の例）。特定の母集団を設定すると時代や場所が異なると変動するため、世帯収入の分布を抽象化した「大元の母集団」を想定する。

②母集団（全港区民）の各世帯が、大元の母集団である世帯収入の正規分布データから世帯収入の実現値を得ていると考える。

③標本調査では、母集団の世帯収入の実現値から無作為にデータを抜き出しており、①大元の母集団から③世帯収入の標本データを無作為抽出していることと同義だと考える。データの集計や分析作業においては、③標本データから①大元の母集団の値の復元を行う。

①大元の母集団であるデータの分布の存在を前提に、②母集団（全港区民）を經由して③標本を抽出するという発想に基づくと、仮に港区民の全数調査から得られた（測定誤差のない）完全なデータがあったとしても、観測されている状態はあくまでも実現した1つの状態であると考えられることから、後述の母集団の値についての推定や検定が有効となる。

標本データにおける世帯収入の平均値は「確率的情報」であり、例えば1回あたり1,000人程度を対象としたアンケート調査を100回実施し、毎回800人程度の有効回答がある100標本を得た場合に、標本平均は毎回異なり、平均が555万円の標本もあれば700万円の標本も存在することになる。同平均値を小さい順に並べ出現頻度を図で表すと左右対称の釣鐘型の正規分布となる。横軸の金額を確率に変換すると、コラム図表4（ロ）のような確率分布を示すことができ、0近辺の確率が最も高く両端に行くにつれて出現確率が下がることが示される。こうすることで、先の例でいえば、標本平均555万円が母集団の平均において出現する確率を求めることができる（▶データの分布範囲については、コラム④（2）の標準偏差の項目を参照）。

そして、標本から得られた回答データから、母集団全体の確率的情報の分布を復元するような作業を「推定」という（統計的推測ともいう）。区民の全数調査でなくとも、標本データのサイズ（データの人数）が大きければ、確率的情報の標本平均が母集団の平均に近づくという性質が知られている（大数の法則）。また、標本のサイズが大きければ、母集団の分布がどのようなものであっても、確率的情報が標準正規分布に近づくことが知られている（中心極限定理）。なお、正規分布は、平均値、中央値、最頻値が一致する性質を持っており（▶記述統計については、コラム④を参照）、現実社会で多くみられる分布の形である。

なお、推定や検定の作業において、各種の「誤差」がその正確性を左右する。まず、無作為抽出した標本の特性が母集団からずれる「標本誤差」がある。関連するものとして、住民基本台帳の非網羅性によって生じる誤差がある。標本誤差については、標本のサイズを大きくすることで小さくすることができる。一方で、「非標本誤差」も存在し、調査への非協力や特定の質問項目の無回答によっても標本データの質が左右される。特定の項目の無回答により発生する欠測値への対応には、何らかの値を代入すべきだという考え方もある。他にも、回答者の読み飛ばしや誤解、集中力の不足による回答誤差や、調査者の設計した設問の選択

肢や文言の不備に起因する回答誤差、紙媒体の調査の場合のデータ入力時に発生する入力誤差、調査方法に起因する測定誤差（詳細は割愛するが、質問紙調査とウェブ調査とで異なる測定誤差など）がある。

（2）仮説の検定

標本データから統計的に仮説を検定する論理はやや回りくどい。実務的な例でいえば、標本データからある政策の効果があるということを証明したい場合、政策の効果がないということ否定することで、政策の効果があるといえそう、という論理過程を経る必要がある。実務的には、エクセルや統計ソフトが算出する統計的検定の有意確率（p 値）が0.05（5%）よりも小さければ、効果がありそう（又は要因間に関係があるといえそう）という判断ができれば問題はない。以下では、仮説検定の大まかな過程をみていく（▶仮説検定の実際の例は、第II章末尾に掲載のコラム⑦を参照）。

大別して、次のような過程が必要となる。以下で登場する仮説は、「棄却したい帰無仮説」と「採択したい対立仮説」と考えると理解がしやすい。まず、①採択したい仮説と真逆の「政策の効果がない」という帰無仮説を設定し、②帰無仮説が正しい前提での観測データが実現する確率の算出後、③同条件下であまり得られない（棄却域に入る）観測データの値ならば帰無仮説を誤りとみなし、もともと立証したいと考えていた「政策の効果がある」という対立仮説を採択する。

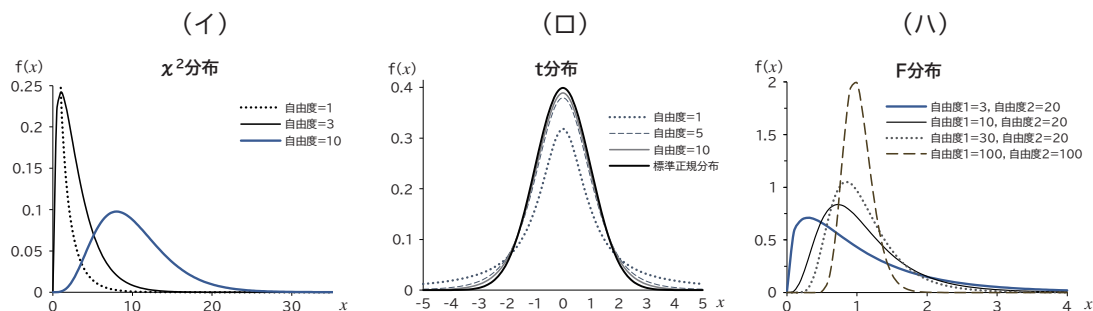
別の表現をすると、帰無仮説（「効果がない」「関係がない」「差がない」）の下では、偶然（一般に5%の確率を基準とする）にしか生じないような推定値が標本でみられたのは、母集団において帰無仮説が成り立っておらず、むしろ同効果があるからだといえそう、と考えることになる。

そのために、母集団における値を中心に誤差の幅を設定し、観察された値が設定幅の中に収まる確率を求める。コラムの図表4（ロ）の正規分布の両端には棄却域の例を示しており、100回中95回は入る区間（横軸で見ると±1.96の区間）に含まれない稀な現象を示している。絶対的な基準値ではなくあくまでも慣習値であるが、5%の稀な確率の値が出現した場合に帰無仮説（政策の効果がない）を誤りとみなし、対立仮説（政策の効果がある）を採択することとなる。

統計的検定に用いる代表的な確率分布には、コラム図表5の（イ） χ^2 （カイ二乗）分布、（ロ）t 分布及び（ハ）F 分布がある。実務に際しては詳細の理解は必要ないが、各種確率分布を仮定して各種検定が行われているという事実と、データの尺度によって仮定する分布及び検定方法が異なるという点は把握しておくとう用である（▶データの尺度と集計方法については、第II章末尾に掲載のコラム⑤とコラム⑥を参照）。

対応関係としては、クロス集計では要因間の独立性の検定にカイ二乗検定を用い、相関係数や2群間の平均値の差の検定ではt 検定を用い、3群以上の平均値の差の検定には分散分析を行いF 検定を用いる。政策効果の検証の基礎となる重回帰分析でもt 検定を利用する（▶第VI章末尾に掲載のコラム⑨（3）を参照）。

コラム図表5 代表的な確率分布の確率密度



検定の際には、「自由に動かせる値の数」である自由度が重要な情報となる。実務的には詳細の理解は必要ないが、検定に利用する計算式において、回答者数から1を引くこと ($n-1$) がよくあるという事実把握があると有用である。

具体例を挙げると、母集団における分散がわからない場合に（わからないことが一般的）、標本分散の推定において標本平均 (\bar{x}) を利用することから自由度が一つ減るが、このことを考慮に入れ標本のサイズから1を引く ($n-1$) 操作を行う。データにおける偏差 ($\bar{x} - x_i$) の計算により、確率情報1つ分の動きを落としてしまうためである。他の推定の計算でも自由度が変化することがあり、標本のサイズ-1で割ることがあるが、このような理由に基づいている。

参考文献

- 川出真清 (2011) 『コンパクト経済学ライブラリ コンパクト統計学』新世社
- ➡見開きの図解によるわかりやすい解説が多く、本書全体を通して標本抽出、各種指統計用語、確率論、推定、仮説検定などについてコンパクトに学ぶことができる仕様となっている。
- 西山慶彦 (2019) 「統計理論の基礎」西山慶彦・新谷元嗣・川口大司・奥井亮『計量経済学』有斐閣、第3章、pp.79-102
- ➡一般向けではなく専門的文献であるが、データ分析のための各種理論の丁寧な解説や実データの例示が豊富であり、深く理解したい場合に有用な文献である。本書をとおして、上級レベルを含む分析技法がカバーされており、充実した専門書となっている。
- 三好大悟・堅田洋資 (2021) 「仮説が正しいかどうか仮説検定で結論を出す」『統計学の基礎から学ぶExcelデータ分析の全知識』インプレス、第4章、pp.121-176
- ➡エクセルの操作の図解を含め、実践的な手順で仮説検定の方法が解説されている。

コラム④ 記述統計

本コラムでは、データ特性を数値的に整理・要約する、代表値について概説していく。
 コラム①のコラム図表1のように、数値が並んだ状態のデータを眺めても得られる情報は少ない。データ分析の基本として、情報を効率的に整理・可視化するため、度数分布表や

ヒストグラムを確認したり、代表値を確認する作業が必要となる（▶度数分布表については、第二章末尾に掲載のコラム⑦（1）と第七章末尾に掲載のコラム⑩（1）を参照）。この作業により、データの分布の形状や異常値を確認することができる。

ここでは、データの特性や散らばり度合いを数値的に捉えるための基礎的な指標を概説する。

（1）データの特性の要約（範囲、最頻値、平均値、中央値）

ここでは、以下の仮想的なデータを想定する。

7名分のテストの得点データ {74点、79点、79点、84点、89点、93点、100点}

▶ 範囲（range）

最小値（下限値）と最大値（上限値）の差。上記の数値例だと100-74=26点となる。

▶ 最頻値（mode）

データにおいて最も出現頻度が高い値。上記の数値例だと79点となる。

▶ 平均値（mean）

データの総和をデータの数で割った値で、データの分布の重心の値。上記の数値例だと、次のとおりとなる。

$$\text{平均} = \frac{74 + 79 + 79 + 84 + 89 + 93 + 100}{7} = 85.43 \text{ 点}$$

▶ 中央値（median）

数値を大きさの順に並べたときに中間の位置にある値であり、上記の7つの数値例だと、4番目にあたる84点が該当する。なお、データの数が偶数の場合には中央に並ぶ2つの数値の平均を中央値として扱う。中央値は、全ての数値を基に重心を計算する平均値とは異なり、極端に大きな数値が含まれていてもその影響を受けにくいという特性を持つ。

（2）平均からのデータの散らばり（分散、標準偏差、変動係数）

データの散らばりを表す分散及び分散から派生した標準偏差並びに変動係数について概説する。いずれの指標も、値が大きいほどデータの散らばりが大きいことを表す。

▶ 分散（variance）

データの分布が平均からどの程度散らばっているかを表す値。各値の平均からの乖離を捉えた「偏差」を二乗した値の平均であり、元のデータとは異なった単位となる。

ここでは、以下の仮想的なデータを想定する。

	2つのデータ例								
データA	4	4	9	9	12	13	14	15	平均=10
データB	8	9	10	10	10	10	11	12	平均=10

上の2つのデータの例だと、まず以下のように、各値の平均（両データともに10）との差である絶対偏差及び偏差の二乗を取ることができる。

	2つのデータの絶対偏差						2つのデータの偏差の二乗									
データA	6	6	1	1	2	3	4	5	36	36	1	1	4	9	16	25
データB	2	1	0	0	0	0	1	2	4	1	0	0	0	0	1	4

上の数値例だと、分散は次のとおり、データAの分散の値は16、データBは1.25となる。

分散の値はデータA>データBとなっており、相対的に、データBは平均の近辺にデータが集中し、データAは平均から離れたところにまでデータが散らばっていることを意味している。

$$\text{データ Aの分散} = \frac{36 + 36 + 1 + 1 + 4 + 9 + 16 + 25}{8} = 16$$

$$\text{データ Bの分散} = \frac{4 + 1 + 0 + 0 + 0 + 0 + 1 + 4}{8} = 1.25$$

➤ 標準偏差 (standard deviation)

分散の正の平方根の値で、元のデータと同じ単位でデータの散らばりを表すことができる指標。先のデータAの場合は $\sqrt{16} = 4$ 、Bの場合は $\sqrt{1.25} = 1.12$ となる。

データが左右対称で釣鐘型の分布（正規分布）をしている場合、以下のとおり、平均からどの程度の範囲にデータの何割が含まれるかがわかる。

- 平均 - 1 標準偏差以上、平均 + 1 標準偏差以下の範囲（1シグマ範囲）に全データの約68.26%
- 平均 - 2 標準偏差以上、平均 + 2 標準偏差以下の範囲（2シグマ範囲）に全データの約95.44%
- 平均 - 3 標準偏差以上、平均 + 3 標準偏差以下の範囲（3シグマ範囲）に全データの約99.74%

ちなみに、一般的な基準としてしばしば利用される区切りの良い「95%」の場合は、2シグマの値に近い、平均 ± 1.96 標準偏差の範囲となる（▶コラム③のコラム図表4（ロ）を参照）。

例えば、第VI章の食料品買い物環境満足度スコアが、釣鐘型の分布（正規分布）をしている前提で、平均11.27点、標準偏差3.11であった場合以下がわかる。

- 8.16～14.38点の満足度の範囲に約68.26%の区民が入る
- 5.05～17.49点の満足度の範囲に約95.44%の区民が入る
- 1.94～20.60点の満足度の範囲に約99.74%の区民が入る

➤ 変動係数 (coefficient of variation)

標準偏差÷平均値で求める値で、平均を基準値1とした場合の散らばりを表す指標。変動係数が0の場合、データの散らばりが小さいことを表す。

参考文献

- 倉田博史（2019）『図解 大学4年間の統計学が10時間でざっと学べる』KADOKAWA
- ➔図解と計算例で各種指標がわかりやすく解説されており、統計学の知識が乏しい方でも理解が深められる仕様となっている。
- 佐々木弾（2019）「記述統計—数値でみるデータの性質」『算数からはじめて一生使える確率・統計』河出書房新社、第2章、pp.55-76
- ➔図解と具体的例及びクイズにより各種指標が解説されており、統計学の知識が乏しい方でも理解が深められる仕様となっている。第2章のみならず、本書全体をとおして、同スタイルが貫かれている。