究理IID テキストブック

1. はじめに

1-1 データサイエンス ≒ 統計学とは

「究理 I」では、1学期には科学的探究の進み方について、2・3学期には情報検索や専 門家からの情報収集法について学んだ。同時に、実験レポートや報告書のまとめ方や口頭発 表についても経験した。「究理 II」では「究理 I」で学んだ基礎的な体験をさらに深めて、 より質の高い探究活動にチャレンジする。

「究理II」の Data Science コース(「究理II D」)では、統計学的なアプローチによる探究 活動を行う。顕微鏡が「目に見えない小さなものを可視化するための道具」であるのに対し、 統計学は「目に見えない大きなものを可視化するためのほぼ唯一の科学的な道具」であると 言われている。現代では、統計学はすべての学問において重要なツールとなっており、多く の社会的なしくみは、統計学の成果をもとにして立案されるようになってきている。



これほどに強力なツールである統計学は、近年になってとくに注目を浴びている。その理 由は、21世紀に入ってからの急速な情報通信技術の発展にある。インターネットの普及に より、大量のデータが容易に入手できるようになった。また、コンピュータの高性能化と普 及により、誰でも簡単にデータの整理と分析を行えるようになった。現在、データ収集と分 析の能力は、特定の専門家だけでなく、より多くの職種において必要な力となってきている。

このような時代において、適切な分析により効率的に問題を解決するために、また、誰か のデータ分析に騙されないために、データを見極めるセンスや分析する力を磨く必要がある。 「究理 II D」の授業が、君たちにとってそのような場になることを期待している。



1-2 「究理 ID」の目標

1、データの収集・処理・分析・発信に関する知識とスキルを学び、活用する

データサイエンスの探究活動において強力なツールとなるのが、表計算ソフトである。 1学期は、その使い方を学ぶ。あわせて、統計学的な探究に必要な基本的な知識について も学習する。2学期には、1学期に学んだことを実際に活用して、探究活動を行う。

2、研究倫理や情報倫理に配慮して活動する

「究理 II D」では、科学的に誠実な態度を要求する。データの収集や分析に学問的誠実 性や科学的厳密性を欠いている発表は、よい発表とは見なさない。また、個人情報を雑に 取り扱ったり、他者の表現物をコピーしたりするような姿勢も厳禁とする。

3、自ら考え、自ら動いて統計的探究を進める

とくに2・3学期の探究活動においては、自ら考えて、自ら動く主体性が重要となる。 科学的な丁寧さを欠くことなく、粘り強く探究活動を行ってほしい。君たちの意欲的で地 道なデータ解析によって、新しい発見や面白い知見が見いだされることを期待している。

1 学期	オリエンテーション	
	情報処理実習	・表計算ソフトの使い方
		・統計処理の基本演習
		・グラフや図表の表し方
		・情報モラル、研究モラル
2 学期	統計的探究プロセスについて	・PPDACサイクル
		・各種データベースの紹介
	データベースの作成と整理	・データソースの取得
		・非数値データの数値化
		・データクリーニング
	探究活動の実施	・問いの設定
		・追加データの収集
3 学期		・データの整理と分析
	発表の準備	・資料の作成、発表の練習
	研究発表	 ・ポスター発表

「究理ID」授業の流れ

1-3 評価について

「究理 II D」では、他の普通教科と同じように成績評価 がある。定期考査だけでなく、情報処理の課題や探究活動 の成果物なども含めて総合的に評価する。積極性と主体性

評価の対象となるもの				
□定期考査	口情報処理課題	頁		
ロポスター	口各種提出物	等		

のある取り組み姿勢と、客観性と論理性を備えた発表やレポートを高く評価する。

2. 表計算ソフトの活用

表計算ソフトとは、数値データの集計や分析に用いられるソフトウェアである。代表的な ものに Microsoft 社の Excel がある。ここでは、Excel 2016の使い方を説明する。なお、Word や PowerPoint と共通の操作法については、究理 I のテキストを参照すること。

Excel には、以下の3つの機能がある。

 ① 表の作成……数字や文字、計算式を入力して表をつくることができる
 ② グラフの作成…入力したデータをもとにグラフをつくることができる
 ③ データベース…入力したデータは、用途に応じて検索、抽出、並べ替え、 集計などが可能であり、データベースとして利用できる

2-1 Excelの画面構成(Excel 2016)



1、データを入力する

次のいずれかの方法でデータを入力する。 ・セルをダブルクリックして入力する。 ・数式バーに入力する。 2 3

2、データ形式

<データ形式>

・Excel で扱うデータにはいくつかのデータ形式があり、目的に応じて適切なデータ形 式を選択する必要がある。

数値データ	・数字を入力すると数値データとして認識される。
	・計算の対象となる。
	・セルの右詰めに配置される。
文字列データ	・文字を入力すると文字列データとして認識される。
	・計算の対象とならない。
	・セルの左詰めに配置される。
	・数字を文字列データとして扱いたい場合は、数字の前に「'」(半
	角アポストロフィ)を添えて入力するとよい。
日付や時間	・日付や時間を入力するとシリアル値と呼ばれる数値データとして
	認識される。
	・計算の対象となる。
	・シリアル値は、1900年1月1日0時0分0秒を「1」とした実数
	値で、整数部は年月日を、小数部は時刻を表す。たとえば、もと
	の数に1が足されると、日付が1日進み、0.5が足されると、時
	刻が 12 時間進む。

<データの表示形式>

・ホームタブのボタンから、数値データの表示形式を変えることができる。

ファイル ホーム 挿入	ページ レイアウト	数式 データ	校閲	表示	ACROBAT	♀ 実行した	
	• 11 •	A A =	≡ ≫	Ē	日付	-	
貼り付け・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	- <u>A</u> - <u>A</u> -	· 🚡 • 🚍 🗄	= = = •	≣ ⊞ •	÷ % ,	€.0 .00 .00 →.0	
クリップボード ら	フォント	5	配置	5	数値	T _S	
パーセンテージで表示する	5 3桁ごとに	_ 「,」が入る	小	`数点以⁻	下の表	小数点以下	の表
例) 0.1 → 10%	例)1234	$4 \rightarrow 1,234$	示	術数を増	当やす :	示桁数を減	らす

3、数式を入力する

- ・「=」で始める
- ・セル番地を入力すると、セルの値が代入されて計算される。※ セルをクリックすることでも、セル番地を入力できる。
- ・四則計算などに使う記号は、数学で用いるものと一部異なるので注意する。

<四則演算の記号(算術演算子)>

	加算	減算	乗算	除算
数学	+	—	×	÷
Excel	+	—	*	/

D5のセルには、計算結果の「12」が表示される。

※()も使える。

<等号·不等号(比較演算子)>

	等しい	等しくない	より大きい	より小さい	以上	以下
数学	=	\neq	>	<	\geq	\leq
Excel	=	<>	>	<	>=	<=

2-3 セル・行・列の選択/削除/挿入

1、セル・行・列を選択する

- セルをクリックするとセルを選択できる
- ・行番号/列番号をクリックすると行/列を選択できる。

2、複数のセル・行・列を同時に選択する

<範囲選択>

- ・選択したい範囲の始点をクリックし、終点までドラッグする。
- ・選択したい範囲の始点をクリックし、shift キーを押しながら終点をクリックする。

<離れた複数の場所を同時に選択>

・選択したい箇所をクリックし、Ctrlキーを押しながら選択したい箇所をクリックする。



3、行/列を削除する

- ・削除したい行/列を選択して、「ホーム」の「セルの削除」をクリックする。
- ・削除したい行/列を選択して、<u>右クリックして「削除」を選択</u>する。

4、行/列を挿入する

- ・挿入したい場所の次の行/列を選択して、「ホーム」の「セルの挿入」をクリックする。
- ・削除したい行/列を選択して、<u>右クリックして「挿入」を選択</u>する。

2-4 データの移動/複製/修正/削除

1、データを移動する

- ・セルを選択し、「ホーム」の「切り取り」をクリックして、移動先で「ホーム」の「貼 り付け」をクリックする。
- ・セルを選択し、<u>右クリックして「切り取り」</u>を選択して、移動先で<u>右クリックして「貼</u> り付け」を選択する。
- ・セルを選択し、セルの囲み線をクリックして、移動先までドラッグ&ドロップする。
- ・セルを選択し、Ctrl キー+X キーを押して切り取り、移動先でCtrl キー+V キーを押して貼り付ける。

2、データを複製する

- ・セルを選択し、「ホーム」の「コピー」をクリックして、複製先で「ホーム」の「貼り
 付け」をクリックする。
- ・セルを選択し、<u>右クリックして「コピー」</u>を選択して、複製先で<u>右クリックして「貼り</u> 付け」を選択する。
- ・セルを選択し、セルの囲み線をクリックして、Ctrl キーを押しながら複製先までドラッグ&ドロップする。
- ・セルを選択し、<u>Ctrl キー+C キー</u>を押してコピーし、複製先で <u>Ctrl キー+V キー</u>を押し て貼り付ける。
- 3、データを修正する

<セルの内容の全体を修正する>

・セルを選択した状態でそのまま入力すると、セルの内容全体が置き換わる。

<セルの内容の一部を修正する>

- ・セルをダブルクリックして入力カーソルを出現させて修正する。
- ・数式バーをクリックして修正する。

4、データを削除する

・削除したいデータを含むセルを選択して Delete キーまたは Backspace キーを押す。

5、隣り合うセルに自動でデータをコピーしたり入力したりする

隣り合うセルに、自動的にコピーや連続データを入力してくれる機能のことを、**オートフ ィル**という。



<オートフィルのいろいろ>





連続性のあるデータは自動的に連続データ が入力される。

複数のセルを選択してからオートフィルす ると、自動で連続性を推定する。

)2 🔻 (f _x	=A2*B2	
A	А	В	С
1			
2	2	1	2
3	2	2	154
4	2	3	
5	2	4	
6	2	5	
141			

2 🗸 🤇	f _x	=A2*B2	
A	A	В	С
1			
2	2	1	2
3	2	2	4
4	2	3	6
5	2	4	8
6	2	5	10
000			1001

数式や書式(罫線や網掛け、文字の色等)もコピーされる。

D

C	D	E	F	右クリックでフィルハンドルをドラッグすると、コピーの 種類を選ぶことができる。
2				連続データでも単純にコピーする。
	่ชแตาห้-	-(C) •		連続データを入力する。
	連続データ	(<u>5)</u> パ <u>5)</u> パー (フィル)(I]•	書式(罫線・網掛け・文字色等)だけをコピーする。
	書式なしコ	ピ− (フィル)(<u>(</u>	2)	書式(罫線・網掛け・文字色等)はコピーせず、数値 や数式をコピーする。

セルの指定には、相対参照と絶対参照がある。

- ・相対参照……コピーをすると、コピー先を起点とした別のセルを参照する。通常のセル 番地は相対参照になっている。
- ・絶対参照……コピーをしても、同じセルを参照する。「\$」を付けることで、行番号や列
 番号を固定することができる。

具体例で説明しよう。

たとえば、下の図で B2 に「=A2*B1」という数式が入力されているとき、これを B6 ま でコピーすると結果は右下の図のようになる。このとき、B2 と B6 の数式を比べると、A2 が A6 に、B1 が B5 に変化している。これは、「1 つ左のセルの値×1 つ上のセルの値」とい う位置関係がコピーされたためである。これが**相対参照**である。



通常のセル番地を入力すると、位置関係がコピーされる

コピーしたときに、参照するセルがずれないようにするためには、ずれてほしくない行番 号または列番号の前に「\$」を付ける。たとえば、先ほどの数式を「=A2*B\$1」とした場 合、これを B6 までコピーすると結果は下の図のようになる。このとき、B2 と B6 の数式を 比べると、A2 は相対参照のままであったために A6 に変化しているが、B\$1 はそのままであ る。「1」という行番号を固定したために、下のセルへとコピーしても参照元のセルが変わら なかったのである。このような参照方法を**絶対参照**という。「\$」は行番号と列番号のそれぞ れに付けることができる。



記号の前に「\$」をつけると、その行/列が固定される

相対参照と絶対参照を上手に使い分けてコピーすることで、Excel による計算処理を効率 的に行うことができる。

1、セルの書式を設定する

・書式を設定したいセルや範囲を選択して、「ホーム」タブのボタンから設定する。



- 2、罫線を引く(詳細設定)
 - ・「ホーム」タブ「罫線」ボタンのドロップダウンリストから「その他の罫線」を選び、 セルの書式設定ウィンドウを開いて、詳細設定を行う。



- 3、行や列の幅を変える

 - ・行や列を複数選択した状態で上記の操作を行うと、複数の行や列の幅を同時に変えることができる。

1、関数とは

100 個のセルの値を合計したいときに、「=A1+A2+……」などと数式を入力するのはと ても手間がかかる。また、「平均点以上の得点を取った人は何人いるか」「最大値はいくつか」 といった値は数式では求めることができない。このように、数式で求められなかったり、数 式化することができなかったりする計算を可能とするのが**関数**である。

2、Excelにおける一般的な関数の表記法



※ データの範囲の表し方……「:」で区切って表す



3、関数の入力法

・セルや数式バーに直接入力する。

・「ホーム」タブの<u>「オート SUM」ボタンやその隣の「▼」</u>から関数を選択する。

ファイル ホー	ム挿入	ページ レイアウト	数式	データ	校閲	表示	ACRO	OBAT ♀ 操作アシスト	ل	242	R, #
時の付けで、	游ゴシック B I U ・	• 11 • ⊞ • <u>⇔</u> • <u>▲</u> •	A A A ⊈ •		= ₽ = ₽ ≫,•	標準 い。→00	•	留条件付き書式 → 要テーブルとして書式設定 → マレルのスタイル →	翻挿入・	Σ` ≝`	•ד: ۵۰
クリップボード ち		フォント	5	配置	i G	数値	G,	スタイル	セル	編	集

・「数式」タブの<u>「関数の挿入」ボタンや「関数ライブラリ」のボタン</u>から選択する。



4、いろいろな関数

※文字列として表示させるときには、"合格"のように""で囲む。 ※空白を表示させるときには""とする。

※SUMIF 関数や COUNTIF 関数などで検索条件を入れるときも""で囲む。 ※関数の中に別の関数を入れることもできる(これを関数の**ネスト**という)。

	r
SUM 関数	書式 =SUM(範囲)
	機能 [範囲]に含まれる数値の合計を表示する。
	例) =SUM(A2:A8)
	→A2~A8の範囲に含まれる数値の合計を表示する。
SUMIF 関数	書式 =SUMIF(範囲,檢索条件)
	 機能 [範囲]に含まれるデータで、[検索条件]に一致する値
	の合計を表示する。
	例) =SUMIF(A1:A100,">=70")
	→A1~A100 の範囲に含まれる値のうち、70 以上である値
	をすべて足し合わせた値を表示する。
MAX 関数	書式 =MAX(範囲)
	機能 [範囲]に含まれる数値の最大値を表示する。
	例) =MAX(A2:A8)
	→A2~A8の範囲に含まれる数値の最大値を表示する。
MIN 関数	書式 =MIN(範囲)
	機能 [範囲]に含まれる数値の最小値を表示する。
	例) =MIN(A2:A8)
	→A2~A8の範囲に含まれる数値の最小値を表示する。
AVERAGE 関数	書式 =AVERAGE(範囲)
	機能 [範囲]に含まれる数値の平均値を表示する。
	例) =AVERAGE(A2:A8)
	→A2~A8の範囲に含まれる数値の平均値を表示する。
MEDIAN 関数	書式 =MEDIAN(範囲)
	機能 [範囲]に含まれる数値の中央値を表示する。
	例)=MEDIAN(A2:A8)
	→A2~A8 の範囲に含まれる数値の中央値を表示する。

COUNT 関数	書式 =COUNT(範囲)
	機能 [範囲] に含まれる数値の件数の合計を表示する。
	例)=COUNT(A2:A8)
	→A2~A8の範囲に含まれる数値の件数の合計を表示する。

COUNTA 関数	書式 =COUNTA(範囲)								
	機能 [範囲]に含まれるすべてのデータ件数(空白以外)の								
	合計を表示する。								
	例)=COUNT(A2:A8)								
	→A2~A8の範囲に含まれるデータ件数(空白以外)の合計								
	を表示する。								

COUNTIF 関数	書式 =COUNTIF(範囲,検索条件)							
	機能 [範囲]に含まれるデータで、[検索条件]に一致するデ							
	ータの件数の合計を表示する。							
	例)=COUNTIF(A1:A100,">=70")							
	→A1~A100 の範囲に含まれる値のうち、70 以上であるデ							
	ータの個数の合計を表示する。							

QUARTILE.INC 関数	書式 =QUARTILE.INC(範囲,n)
	機能 [範囲]に含まれる数値の第n四分位数を表示する。
	例) =QUARTILE.INC(A2:A8,1)
	→A2~A8 の範囲に含まれる数値の第1四分位数を表示。

STDEV.P 関数	書式 =STDEV.P(範囲)						
	能 [範囲]に含まれる数値の標準偏差を表示する。						
	例)=STDEV.P(A2:A8)						
	→A2~A8の範囲に含まれる数値の標準偏差を表示する。						

STDEV.S 関数	書式 =STDEV.S(範囲)								
	機能 [範囲]の数値を標本としてとらえ、ここから母集団の								
	標準偏差の推定値を表示する。								
	列)=STDEV.S(A2:A8)								
	→A2~A8の範囲に含まれる数値の標準偏差を表示する。								

RANK.EQ 関数	書式 =RANK.EQ(数值,範囲,順序)									
	機能 [順序]で示すルールに従って、[範囲]内の [数値]を									
	並び替えたとき、[数値]の順位を表示する。[順序]は、									
	「0」で降順、「1」で昇順、省略すると降順になる。									
	例) =RANK.EQ(A3,A1:A100,0)									
	→A1 から A100 の範囲の数値のうち、A3 の数値が大きい									
	から数えて何位になるかを表示する。									
IF 関数	書式 =IF(条件式,真の場合,偽の場合)									
	機能 [条件式]が正しい(TRUE)のときは [真の場合]を、									
	正しくないときは [偽の場合] を実行する。									
	例)=IF(A2>=70,"合格","不合格")									
	→A2 の値が 70 以上であれば「合格」と表示し、そうでな									
	ければ「不合格」と表示する。									
AND 関数	書式 = AND(条件式1,条件式2,…)									
	機能 [条件式]が正しいときに「TRUE」(1)を、1つでも正									
	しくないときには「FALSE」(0)を返す。									
	例1) =AND(50<=B1,B1<=100)									
	→B1 の値が、50 以上 100 以下であれば「TRUE」となり、									
	50 未満または 100 より大きい場合は「FALSE」となる。									
	例 2) =IF(AND(50<=B1,B1<=100),"〇","×")									
	→B1の値が、50以上100以下であれば「〇」と表示し、50									
	未満または100より大きい場合は「×」と表示する。									
OR 関数	書式 =OR(条件式1,条件式2,…)									
	機能 [条件式]が1つでも正しいときに「TRUE」(1)、すべ									
	て正しくないときに「FALSE」(0)を返す。									
	例1) =AND(50<=B1,B1<=100)									
	→B1 の値が、50 以上 100 以下であれば「TRUE」となり、									
	50 未満または 100 より大きい場合は「FALSE」となる。									
	例 2) =IF(AND(50<=B1,B1<=100),"〇","×")									
	→B1の値が、50以上100以下であれば「〇」と表示し、50									
	未満または100より大きい場合は「×」と表示する。									

VLOOKUP 関数	書式 =VLOOKUP(検索値,範囲,左端列からの列数)										
	機能 [検索値]の値や文字列を、指定した [範囲]の左端列										
	の中で縦方向に検索して、見つかった行で、[左端列から の列数] の列にあるセルの値を表示する。										
	の列数]の列にあるセルの値を表示する。										
	例) =VLOOKUP(B2,\$C\$1:\$E\$100,3)										
	→B2 のセルにある値(または文字列)を、C1~E100 の左										
	端列(C列)の中から検索し、見つかった行で、左から3 列目の列(F列)にあるセルの値を表示する										
	列目の列(E 列)にあるセルの値を表示する。										
HLOOKUP 関数	書式 =HLOOKUP(検索値,範囲,上端行からの行数)										
	機能 [検索値]の値や文字列を、指定した [範囲]の上端行										
	の中で横方向に検索して、見つかった列で、[上端行から										
	の行数]の行にあるセルの値を表示する。										
	例) =HLOOKUP(B2,\$C\$1:\$S\$20,3)										
	→B2 のセルにある値(または文字列)を、C1~S20 の上端										
	行(第1行)の中から検索し、見つかった列で、上から3										
	行目の行(第3行)にあるセルの値を表示する。										
CORREL 関数	書式 =CORREL(範囲 1,範囲 2)										
	 機能 [範囲1]に含まれる数値と [範囲2] に含まれる数値										
	の相関係数を表示する。[範囲1]と[範囲2]の数値は										
	対応している必要がある。										
	例)=CORREL(A1:A100,B1:B200)										
	→A1 から A100 の範囲に含まれる数値と B1 から B100 の範										
	囲に含まれる数値の相関係数を表示する。										
RAND 関数	書式 =RAND()										
	機能 0以上1未満の数をランダムに発生させる。										
	例1) =RAND()*5										
	→0 から 5 の間の数をランダムに発生させる。										
	$\overline{m}(2) = \mathbf{P} \wedge \mathbf{N} \mathbf{D} () * 2 \cdot 10$										
	(712) -KAND()*5+10										

SQRT 関数	書式 =SQRT(数値)											
	機能 [数値]の正の平方根を表示する。											
	例1) = SQRT(16)											
	→16の正の平方根(4)を表示する。											
	例 2) = SQRT(B2)											
	→B2のセルにある値の正の平方根を表示する。											
FREQUENCY 関数	書式 =FREQUENCY(データ配列,区間配列)											
	機能 [データ配列]のデータについて、[区間配列]の区切り											
	値にしたがって、各階級に含まれるデータの個数を表示											
	する。この関数を使うときには、以下のことに注意する。											
	・[区間配列]の区切り値は「~以下」を示す値とする。											
	・[区間配列] より1つ多い範囲を選択して、関数の入力											
	を行う。1つ多く選択したセルには、[区間配列]の最											
	大値を超えた値の個数が表示される。											
	・関数の入力後には、Ctrl キーと Shift キーを押しながら											
	Enter キーを押して確定する。											
	例) =FREQUENCY(A1:A100,D1:D9)											
	→D1~D9の区切り値にしたがって、A1~A100のデータの											
	度数を表示する。区切り値の隣に表示させる場合は、E1											
	~E10の範囲を選択して、関数の入力を行う。											



1、データを並べ替える

- ① 作業するデータの範囲を選択する。
- ②「データ」タブの「並べ替え」ボタンを押す。
- ③「並べ替え」のウィンドウが開くので、条件を設定して OK ボタンを押す。

1	A	B	С	D		4	A	В	С	D
1		クラス	国語	数学		1		クラス	国語	数学
2	高月 観音	1	84	40	国語の点数が高い	2	虎姫 伊吹	1	52	95
3	河毛 八千代	2	76	66		3	長浜 城	3	63	72
4	木之本 治蔵	2	66	59	____ 順に並べ替える	4	高月 観音	1	84	40
5	長浜 城	3	63	72		5	余呉 湖子	2	34	40
6	醒ヶ井 あゆ	2	55	58		6	田村 童夢	3	49	41
7	楽市 平和	1	54	67		7	河毛 八千代	2	76	66
3	虎姫 伊吹	1	52	95		8	木之本 治蔵	2	66	59
Э	田村 童夢	3	49	41		9	楽市 平和	1	54	67
0	余呉 湖子	2	34	40		10	永原 マキノ	3	33	67
1	永原 マキノ	3	33	67		11	醒ヶ井 あゆ	2	55	58
2						12				
3						10				



※ 昇順……数が小さいものから大きなものへ並ぶ順(1、2、3、…)

※降順……数が大きいものから小さなものへ並ぶ順(10、9、8、…)

2、データを抽出する

- ① 作業するデータの範囲を選択する。
- ②「データ」タブの「フィルター」ボタンを押すと、上端の行にボタンが現れる。
- ③ 抽出する列のボタン(▼)をクリックし、表示されたメニューから抽出条件を選んで

OK ボタンを押す。



1、Excel におけるグラフの構成要素



- 2、グラフを挿入する
 - ① データの範囲を選択する。
 - ※ このとき、数値だけではなく要素名や項目名も含めて範囲を選択しておくと、凡例 や軸項目に名称が自動で表示される。

※ Ctrl キーを押しながらドラッグすると、離れた場所の範囲を同時に選択できる。 ②「挿入」タブの「グラフ」の中から適当なグラフの種類を選んでクリックする。

ファイル ホー	ム挿入	ページ レイアウト	数式 データ	校閲表示	ACROBAT	♀実行した	い作業を入力して
ビボット おすす テーブル ピボットテ	? ™ テーブル →ブル	■像 オンライン 画像 通像 ■	◇ 図形 ▼ ■ SmartArt ■ スクリーンショット ▼	 ストア マイ アドイン・ 	は おすすめ ガラフ ・ 上	■• 凸• b• 論• ⊵• 索•	■ ポットグラフ マ プ・
テープ	ΰı	B	3	アドイン	5	157	ra ツア-
A1	• : ×	√ fx					
A 1 2 3 4 5	< グラフ ・面グラ 後ろ フの表 ・3D円 手前 表現で	使用時の注意> フ いに隠れた部分は、 示には向かない。 グラフ 」の部分にある領 きない。	見えなくなるの , 域が大きく見え	で、データの大/ るので、データ0	小が逆転するよ の割合の大小を	うなグラ 正確には	
0 7 8 9	 ・円錐や 錯覚 ミッド 	ピラミッドのグラ により大小関係 での割合グラフは	ラフ を正しく認識し 見た目の印象と言	にくい場合がある 実際のデータが角	5。とくに、円 解離しやすいの	錐やピラ で注意。	

3、グラフを編集する

挿入したグラフをクリックすると、グラフツール「デザイン」「書式」の2つのタブが現 れる。これらのタブからボタンを選択することで、グラフの様々な要素を編集・変更するこ とができる。なお、グラフの中のある要素のみをクリックして選択し、グラフツールから編 集作業を行うと、その要素のみを編集することができる。

<グラフツール「デザイン」タブからの編集>



<グラフツール「書式」タブからの編集>



4. 近似線を追加する

移動平均

散布図や折れ線グラフなどには、 グラフの傾向を示すおおよその直 線や曲線(近似線)を追加するこ とができる。具体的には、 グラフ ツール「デザイン」タブの「グラ フ要素を追加」をクリックし、リ ストから「近似曲線」を選択する。 追加したい近似曲線を選択する。 「その他の近似曲線オプション」 を選ぶと詳細の設定ができる。



グラフに数式やR-2乗値を表示する。R-2乗値は決定係 数と呼ばれ、相関係数を2乗した値である。この値が1に近 いほど、データの集合がその近似線によく当てはまっている ことを示す。

<数式やR-2乗値の有効数字の変更>

- グラフの数式やR-2 乗値の部分をダブルクリックし て、「近似曲線ラベルの書式設定」ウィンドウを開く。
- ②「表示形式」の分類から、「数値」を選ぶ。
- ③「小数点以下の桁数」に適当な桁数を入力する。



5、エラーバーを追加する

棒グラフや折れ線グラフなどで、 平均値をデータとしてプロットし ている場合、**エラーバー**(誤差棒) を追加することができる。

- 平均値データ以外にも、エラ ーバーの範囲として設定した い値(範囲、標準偏差、標準 誤差、信頼区間等)をあらか じめ表にして作成しておく。
- ② 平均値でグラフを作成する。
- ③ グラフを選択し、グラフツー ル「デザイン」タブの「グラ フ要素を追加」をクリックし、 「誤差範囲」から「その他の 誤差範囲オプション」を選ぶ。
- ④ ユーザー設定の「値の指定」 ボタンをクリックすると、「ユ ーザー設定の誤差範囲」の小 さなウィンドウが現れる。
- ⑤ ①の表からエラーバーとして使用したい値のセルを選択して、正の誤差の値と負の誤差の値をそれぞれ決める。

範囲のエラーバーは、正と 負で別の値を設定する。標準 偏差、標準誤差、信頼区間の エラーバーは、正と負に同じ 値を入力する(下図)。





<Excel2016 の箱ひげ図>

Excel2016の箱ひげ図は高校数学で学習する箱ひげ図とは、"ひげ"の意味が異なる。下に Excel2016 で作成する箱ひげ図の各区切り値や点の意味を説明する。



A B

値

14

14

1 項日

A A A A A

<箱ひげ図の作成>

- ① データを選択する。なお、右図のように項目と値の2列の表をつくる こと。
- 「挿入」タブのグラフエリアから、箱ひげ図を選択する。
- ③ を選択する。

<データ系列の書式設定>

挿入されたグラフの部分をダブルクリックすると、データ系列の書式設 定ウィンドウが現れるので、ここから書式を設定する。



7、ヒストグラムを作成する

<度数分布表の作成>

・FREQUENCY 関数を使う方法



・COUNTIF 関数を使う方法

1	A	В	С	D	E	F		
1	名前	得点		階級	階級の上限	<u></u>		① F2には10 未満のテータ数を表
2	安曇川藤樹	83		0以上10未満	10			示させる。
З	<u>今津 雷太</u>	75		10以上20未満	20			1.C C 28
4	<u>小谷 茶々</u>	72		20以上30未満	30			
5	河毛 八千代	58		30以上40未満	40	\rightarrow		=COUNTIF(<u>B2·B21, < &E2</u>)
6	木之本 治蔵	65		40以上50未满	50	\rightarrow		
7	坂田 双葉	47		50以上60未满	60	<u> </u>	V	データ配列 条件式
8	醒ヶ井 あゆ	70		<u>60以上70未満</u>	70		Λ	
9	塩津 近江	50		70以上80未满	80			
10	関ヶ原豊	100		80以上90未满	90		_	② F3には10以上20未満のデータ
11	高月観音	68		90以上		-		粉を表示させる へまり 20 去
12	<u>竹生島 びわ</u>	81						数を衣小させる。 フェリ、20 木
13	田村 童夢	52						満のデータ数から 10 未満のデー
14	豊郷 稲枝	35						な 米ケナションナンギ トリン
15	虎姫伊吹	63					_	ク数を引いなよい。
16	長浜 城	62					_	
17	永原 マキノ	46						=COUNTIF(B\$2:B\$21,"<"&E3)
18	<u>比良 おろし</u>	76						- COUNTIF(B\$9.B\$91 "<"&F9)
19	余呉湖子	60					_	$OOONIII (D \varphi 2 \cdot D \varphi 2 1, \forall \alpha D 2)$
20	米原鳥彦	41						※ナートフィルナスためにゴーク
21	楽市 平和	54						
-00-	1						1	配列の範囲は絶対参照にする。
	Τ	D	E	F				
	\rightarrow	階級	階級の上限	度数				 ③ F 4 ~ F 1 0 には、F 3 の関数を
	確	0以上10未満	10	0				オートフィルでコピーする
	虎	<u>10以上20木満</u> 20以上20未満	20	0				
	必	20以上30不周 30以上40未満	40	1				
	D D	40以上50未満	50	3		E 1 1	17	はのドレのゴーク粉なまニキルマ
	シュー	50以上60未満	60	4	4	FII	1	は、90以上のケーク数を衣小させる。
	**	60以上70未満	70	5				
	小	<u>/U以上80木満</u> のIN Fooキ達	80	4		=COUN	NT	'IF'(B\$2:B\$21,">="&E10)
		<u>の成正的不満</u> 90以上	90	1				
	-		-	· · · · ·				

<ヒストグラムの作成>



- 「階級」(D2:D11)と「度数」(F2:F11)を選択する。(離れたセルを同時に選択する場合は、Ctrlキーを押しながら選択する)
- ② 「挿入」タブから棒グラフ(集合縦棒)を選択する。
- ③ 現れた棒グラフをダブルクリックして、データ要素の書式設定のウィンドウを開く。 要素の間隔を「0%」にする(通常、ヒストグラムは棒グラフの間隔をあけない)。
- ④ その他の書式設定をして見やすいグラフにする。
- ※ 階級値を文字列ではなく数値として入力している場合、階級値も棒グラフとなって 現れてしまうことがある(下図)。この場合は、階級値の棒グラフ(下図の灰色)を 選択して Delete すると、度数分布のみのグラフが表示される。



3. データの変換とグラフの活用

ここからは、統計的探究を進める上で必要となる基礎的な知識や考え方について学ぶ。第 3章では収集した生データを変換し、視覚化(グラフ化)する際の注意点や手法について紹 介する。データの収集や統計的解析法については、後の章で扱う。

3-1 データの変換

1、データ変換の目的

調査や測定によって得られたデータは**生データ**と呼ばれる。生データは、注目すべき特 徴や傾向をとらえやすくするために、**変換**する必要があることが多い。データの変換は一 定の意図をもって行われるため、すでに変換されているデータの意図を読み解くためにも、 データ処理について理解しておくことは重要である。データ変換の目的を以下に示す。

・データの特徴を指標化する

……合計值、平均值、最頻值、中央值、標準偏差 等

・グラフを直線化して傾向や関係性をとらえやすくする

……逆数、対数、n 乗值 等

基準をそろえてデータを比較できるようにする

……速度、割合、標準化(平均=0、分散=1になるように統計データを変換すること) 等

回数を記録する

……件数表(タリーチャート)、度数分布表 等

2、データ変換の例

合計値や統計的代表値(平均値、最頻値、中央値)、データの散らばり具合を表す値(分 散、標準偏差)などの求め方については、後の項目を参照。

n 乗する Xⁿ

・例えば、データ分布が二次関数(軸:X=0)である場合、元のデータを2乗するこ とで、直線的なグラフを得ることができる。逆に、2乗して直線的なグラフが得られ れば、データに二次関数の関係があるといえる。



② 逆数をとる 1/X

- ・データ分布が反比例である場合、元のデータの逆数をとることで、直線的なグラフを 得ることができる。逆に、逆数をとって直線的なグラフが得られれば、データに反比 例の関係があるといえる。
- ・異なる時間尺度で測定されたデータの逆数をとることで、時間当たりの量(速度など)
 を得ることができる。たとえば、ある工程を遂行するのにかかった時間の逆数をとる
 と、その工程に取り組む際の速度が求められる。



③ 対数をとる logX

- ・指数関数的に変化するデータについて、データの対数をとることで、直線的なグラフ を得ることができる。逆に、対数をとって直線的なグラフが得られれば、データは指 数関数的に変化しているといえる。
- あるデータが突出していて、そのせいで他のデータの差を比較しにくい場合に、データの対数をとることでデータが比較しやすくなる。
- ・対数をとると、グラフ上の距離が比率や割合の大きさを表すことになる。たとえば、
 10000万人から1000人に減少するのも、1000人から100人に減少するのも、同じ減少幅としてグラフ化される。そのため、割合の変化を見たい場合に適している。





④ 割合を計算する X/基準値

- ・データの量ではなく、それが占める比率を知りたいときに役に立つ。
- ・全体量が異なるデータを比べる時に有効である(例:市町村による高齢者が占める割合の比較)。時系列で比較すれば、ある年から次の年にかけて何%変化したかなどの傾向も知ることができる。



⑤ 速度を計算する X/時間

- ・時間をそろえて変化量を比較したい場合に有効。
- ・速度の計算には、変化量に加えて時間データが必要である。ただし、ある事象に要した時間(周期、作業時間等)がデータとして得られる場合は、その逆数をとることで 速度が求められる。



6 件数を数える

・データの中である数値が生じた回数を数えて、記録する。ヒストグラムの作成に必要。

・表計算ソフトの COUNT 関数を用いてデータを作成することができる。手作業で行う場合は、「正」の字を用いて数を記録すると見やすい。「正」の字や「 ⁺⁺⁺⁺ 」などを用いた数え上げ図のことをタリーチャートという。

表がデータをまんべんなくまとめたり一覧にしたりするときに用いられるのに対して、グ ラフはデータの特徴、変化、傾向、関係性を示したいときに有効である。グラフには様々な 種類があり、目的によって使い分けられる。

1、棒グラフ



<特徴>

- ・項目間で量の大小を比較するのに適している。
- ・項目を自由に並べ替えることができる。

<気をつけること>

- ・量の大小よりも変化を強調したい場合は、折れ線グラフが用いられる。
- ・項目の並び順を工夫してグラフが見やすくなるようにする。(例:値の大きい順、値の小さい順、北から南の順、50音順、時間順 等)





2、積み上げ棒グラフ



複数の棒グラフを積み重ねて1本の棒グ ラフにしたものを**積み上げ棒グラフ**という。

<特徴>

- ・全体量の大小とともに、内訳の大小を比較するのに適している。
- ・時系列の変化を、内訳の変化とともに示したいときによく用いられる。

3、帯グラフ



複数のデータについて同じ長さの横棒を 並べ、それを区切って構成比を示したグラフ を**帯グラフ**という。

<特徴>

・割合の大小を比較したり、割合の変化を示したりするのに適している。

<気をつけること>

- 一般的には、割合の大きなものから順に、帯の左端(または下端)から並べる。
- ・割合が小さい項目は、複数をまとめて「その他」として表す。「その他」の項目は割 合が大きくても最後におく。
- 4、円グラフ



円の中心角度の大きさで全体に対する各 項目の割合の大小を表現したグラフを**円グ** ラフという。

<特徴>

・全体に対する割合(構成比)の大小を示すのに適している。

<気をつけること>

- ・データの合計が必ず 100%になるようにする。合計が 100%にならない場合や 100% を超える場合(複数回答可能のアンケートなど)は、棒グラフで表現する方がよい。
- ・割合の大きなものから順に、12時の向きから時計回りに並べる。
- ・割合が小さい項目は、複数をまとめて「その他」として表す。「その他」の項目は割 合が大きくても最後におく。
- ・割合の大小を比較したり変化を見たりしたい場合は、帯グラフの方が適している。





データの点と点を直線で結んだグラフを **折れ線グラフ**という。

<特徴>

- ・横軸に対する変化を見るのに適している。折れ線の上昇・下降で増減の傾向を読み取れる。
- ・横軸の項目(値)は順番が決まっている。

<気をつけること>

- ・横軸が非等間隔で測定されたデータは、散布図を使う。
- ・変化よりも量の大小を表現したい場合は、棒グラフを使うこともある。





縦軸と横軸にそれぞれ別の量をとり、デー タがあてはまるところに点を打ってデータ の分布を示した図を**散布図**という。

<特徴>

- 縦軸と横軸の2つの量にどのような関係があるのかを見るのに適している。
- ・他のグラフは1つの棒や点で1つの量を表現するのに対し、散布図は1つの点で2つ の量を同時に示している。
- ・2つの量の相関が強いほど、一定の直線または曲線付近に多くの点が集まる。

<気をつけること>

- ・複数のデータが重なる場合も、1つの点として表示される。重なるデータ数が多いほ ど点の直径を大きく描く方法もある(**バブル図**)。
- ・相関関係の有無が分かるが、因果関係が分かるわけではないので注意。

7、ヒストグラム



データを一定の値ごとに区切り、その区間 (階級)ごとにデータの個数(度数)をまと めた表を度数分布表という。度数分布表をも とにして、一方の軸に階級の幅を、もう一方 の軸に度数をとってデータの分布を示した 図をヒストグラムという。

<特徴>

・データや頻度のばらつきや分布を表示するのに適している。

<気をつけること>

・ヒストグラムは縦に値が積み重なる棒
 グラフに似ているが、厳密には棒グラ
 フと異なって面積が度数を表している。
 したがって、階級の幅が異なってもグ
 ラフ化することができる。(階級幅が二
 倍になった場合は、度数の高さは半分
 にする。)



・階級の幅をどのように設定するかによって、分布の傾向が見えにくくなってしまう場合がある。

・異なる複数のデータのばらつきを比較したい場合は、箱ひげ図が適している。

<階級数を決める方法>

通常は、階級の数が5~10くらいになるように階級の幅を決めてヒストグラムを作成し、データの散らばりがあまり見られなければ階級数を変えてみる。階級数の設定に とくにルールはないが、データ数から階級数の目安を決定する公式として、スタージェ スの公式を紹介する。

スタージェスの公式 M = 1 + log₂ N M: 階級の数 N: データ数

【スタージェスの公式で計算したデータ数と階級数】

データ数 N	11	22	44	91	175	350	700	1399
階級数 M	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5



最大値、最小値、四分位数を用いてデータの散らば りを表したグラフを**箱ひげ図**という。四分位数とは、 データを小さい順に並べて4等分したときに、小さい 値から数えて全体の1/4(25%)、1/2(50%)、 3/4(75%)に該当する値のことである。小さい方 から1/4に当たる値を**第一四分位数、**1/2にあた る値を**第二四分位数、**3/4にあたる値を**第三四分位** 数という。第二四分位数は中央値と等しい。

<特徴>

・データのばらつきを表示するのに適している。とくに異なる複数のデータのばらつき
 を比較する場合は、ヒストグラフよりも箱ひげ図が適している。

<気をつけること>

- ・ひげの両端は最小値と最大値であることが多いが、上下10%(や15%等)の値を両端としているものあるので注意する。
- ・箱ひげ図の中央の値は平均値ではないので注意する。平均値は「+」印などで表示することがある。

9、レーダーチャート



各項目の座標軸として放射状に線を引き、それ ぞれの線上に各項目の値の点を打つ。これらの点 を結んでできた多角形で表現されるグラフをレー ダーチャートという。

<特徴>

・項目間のバランスがよいと正多角形に近くなり、値が大きいほど面積が大きくなるため、複数の項目のバランスや総合的な値の大小を直感的に把握しやすい。

<気をつけること>

- ・外側に広がるほど「良い」値となる。元のデータが小さいほど「良い」値となってい る場合には、逆数や速度をとるなどして大きいほど「良い」値になるよう変換する。
- ・似ている項目や相関のある項目を近くに配置すると分かりやすい。
- ・各項目の目盛のとり方によって印象が大きく変わるので注意が必要である。
- ・通常3~6項目で構成される。項目が多すぎると意図が不明確になる。

統計データを地図上に表現したものを**統計地図**という。



<特徴>

・地域ごとのデータの量や分布を把握するのに適している。

<気をつけること>

・情報量が多くなるので、見やすくするためにデザインを工夫したり、注目すべき点を 別に説明したりするとよい。

11、共通の注意事項

- ・棒グラフや折れ線グラフなどで、縦軸の原点をゼロとした場合に変化や大小が見えにく くなることがある。その場合は、途中を省略したり原点をゼロではない値にしたりする ことがある。ただし、印象が操作されることには十分留意する。
- ・複数のデータ系列を比較する際には、データを混合させないために線や点の色や形を変える。その際、白黒で印刷する場合にも違いが分かるような色の組合せを用いるとよい。
- ・他の値に比べてきわめて大きいまたは小さい値を同時に表示したい場合、主軸とは別の 第二軸を設定してもよい。

4. 統計量の活用

データ全体の特徴を1つの数値で表したものを統計量という。この章では統計分析の基礎 となる基本統計量について紹介する。具体的には、代表値と散らばりを表す統計量について 説明する。

4-1 代表値とその特徴

データ分布の中心的な位置を示し、そのデータを代表する値を**代表値**という。おもな代表 値には平均値、中央値、最頻値がある。

1、平均值

データを平らにならした値を**平均値**といい、データの総和をデータ数で割ることで求め られる。変量 x について n 個の値 x_1, x_2, \dots, x_n が得られたとき、平均値 \overline{x} は次 のような式で表される。

$\overline{x} = \frac{1}{n} \left(x_1 + x_2 + \dots + x_n \right)$

平均値は、計算のしやすさや、数学的に使い勝手のよい性質をもつことから、しばしば 代表値として用いられる。しかしながら、平均値は常に代表値として適切であるとは限ら ない。とくに、データの中に他のデータから極端に外れた値(**外れ値**という)がある場合 は、平均値はその値に引きずられて大きくなったり小さくなったりするため、使用に注意 が必要である。

<平均値が代表値として適している場合>

- ① 極端に大きな値や小さな値(外れ値)がない
- ② データの分布がおおむね左右対称である
- ③ すべてのデータ(外れ値も含め)の影響を反映する指標がほしい

2、中央值

データを大きさの順に並べたとき、中央の位置にくる値を中央値(メジアン)という。 データが偶数個あるときは、中央の2つの値の平均値を中央値とする。

平均値と異なり、中央値は、データの中に外れ値があっても、その影響をあまり受けな い。一方で、すべてのデータからの影響が反映されるわけではないため、データ全体の変 動と中央値の変動が一致しない場合もある。

<中央値が代表値として適している場合>

- ① 極端に大きな値や小さな値(異常値や外れ値)がある
- ② データの分布が左右非対称である

3、最頻値

データにおいて、最も個数の多い値を最頻値(モード)という。度数分布表においては、 最も度数の大きい階級の階級値を最頻値とする。

最頻値は、データの中に外れ値があっても、その影響を受けることがない。一方で、分 布のグラフに山が複数ある場合には、最頻値が1つに定まらないことがある。また、度数 分布表においては階級の幅をどうとるかによって、最頻値が大きく変わることがある。デ ータは、数値で表すことができる**量的データ**(測定したり計算したりできるデータ)と数 値では表せない**質的データ**(アンケート項目や順位などのデータ)に分類されるが、量的 データにおいて最頻値が用いられることはあまりない。

<最頻値が代表値として適している場合>

- ① データの種類が質的データである
- ② データの分布が左右非対称で、極端に偏っている
- 【問1】生徒100人の身長の平均が163.5cm だった。この結果から確実に正しいといえることには〇、 そうでないものは×をつけなさい。
 - (1) 伸長が 163.5cm より高い生徒と低い生徒はそれぞれ 50 人ずついる。
 - (2) 100 人全員の身長を足すと 16350cm になる。
 - (3) 身長を 10cm ごとに区分けすると、「160cm 以上 170cm 未満の生徒」が最も多い。



【問2】下のグラフから判断して、一般的な家庭の貯蓄額はいくらぐらいだといえそうか。

4-2 散らばりを表す値

データがどれだけ散らばっているかということを調べることはデータを語る上で重要で ある。データの散らばりを表す値として、範囲、偏差、分散、標準偏差を紹介する。

1、範囲

データの最大値と最小値の差を範囲(レンジ)という。範囲は、データの散らばり具合 を簡単に表現する値として便利であり、工場の品質管理のように、最大値や最小値が問題 になる場合には有用である。しかし、外れ値がある場合には、その値によって実際の散ら ばり具合以上に範囲が大きな値となってしまうことがある。

2、偏差

データの各値から平均値を引いた値を**偏差**という。<u>各データがどれだけ平均値から離れ</u> <u>ているか</u>を表す。n 個の値 x_1, x_2, \dots, x_n について、平均値を \bar{x} とすると、偏差は次 のような式で表される。

$x_1 - \overline{x}$, $x_2 - \overline{x}$,, $x_n - \overline{x}$

偏差は、あくまでも個々の値について、平均からの散らばり具合を示す統計量である。 偏差には正の値と負の値があり、すべて合計すると必ず0になる。そのため、偏差を合計 したり、偏差の平均を求めたりしても、データ全体の散らばり具合を表すことはできない。

3、分散

偏差を2乗した値の平均値を**分散**という。偏差を2乗することで、値が正になるため、 平均をとっても0にはならない。そのため、偏差の2乗の平均は、<u>データ全体の散らばり</u> 具合を表す指標として用いることができる。分散は s² で表し、次のように計算する。

$$s^{2} = \frac{1}{n} \left\{ (x_{1} - \bar{x})^{2} + (x_{2} - \bar{x})^{2} + \dots + (x_{n} - \bar{x})^{2} \right\}$$

分散は、データを2乗しているので、単位も2乗されたものとなる。

4、標準偏差

分散の平方根を**標準偏差**という。標準偏差も分散と同様に、<u>データ全体の散らばり具合</u> <u>を表す</u>。分散は2乗された値であるため、その<u>平方根を求めることで、測定単位と同じ単</u> <u>位にする</u>ことができる。標準偏差は s で表し、次のように計算する。(分散を s^2 と表す のは、分散が標準偏差を2乗した値であるためである。)

$$s = \sqrt{s^2}$$

この章で紹介した基本統計量は、Excel においては以下のような関数で求めることができる。

基本統計量	関数
平均值	AVERAGE 関数
中央値	MEDIAN 関数
分散	VAR 関数、VAR.S 関数、VAR.P 関数
標準偏差	STDEV 関数、STDEV.S 関数、STDEV.P 関数

これらの関数のうち、分散と標準偏差については使用時に注意が必要である。

分散と標準偏差の関数は、それぞれ3種類存在する。このうち、VAR 関数と STDEV 関数 は古いバージョンの Excel における関数表記であり、新しいバージョンではそれぞれ VAR.S 関数と STDEV.S 関数に変更された(新しいバージョンでも古いバージョンの関数は使用でき る)。問題となるのは、VAR.S 関数と VAR.P 関数、STDEV.S 関数と STDEV.P 関数の使い分け である。これらの関数を実際に使用してみると分かるが、「.S」で終わる関数と「.P」で終わ る関数では、同じデータを入力しても、異なる計算結果が返されてくる。

「.P」で終わる関数は、計算処理を行うデータを全データ(母集団データ)とみなして、その分散や標準偏差を求めたい場合に使う。一方、「.S」で終わる関数は、計算処理を行うデータを全データ(母集団のデータ)から取り出した一部のデータであるとみなして、その一部のデータから全データの分散や標準偏差を推定したい場合に使う。

たとえば、ある学校の生徒の座高について、その分散を知りたいとする。このとき、全て の生徒(これが母集団)の座高を測定することは大変なので、40人の生徒(これが標本) の座高を測定して、そこから全生徒の座高の分散を推定する。40人の生徒の座高データに ついて、「VAR.P 関数」を用いて分散を求めた場合、それは「40人の生徒データの分散」を 表す。一方、「VAR.S 関数」を用いて分散を求めた場合、それは「40人の生徒データから推 定した全生徒データの分散」を表す。この場合は、全生徒の座高の分散を推定したいのだか ら、「VAR.S 関数」を用いることになる。

母集団と標本については、次の章を参照のこと。

5. 平均の信頼性

平均値は代表値として大変便利な統計量である。しかしながら、平均値を比較したり、平 均値の変動を調べたりするときには、注意が必要である。なぜなら、その平均値が、調べた い対象の真の平均値を表しているとは限らないからである。この章では、データから得られ た平均値がどれだけ信頼性があるものなのかを判定する際の考え方について学習する。

5-1 母集団と標本

1、全数調査と標本調査

統計的探究を行うためには、調査対象を決めてデータを集める必要がある。調査対象を すべて調べる方法を全数調査(悉皆調査)といい、調査対象から一部を取り出して(これ を抽出という)調べる方法を標本調査(サンプル調査)という。たとえば、5年に一度行 われる国勢調査では日本人全員のデータを集めるが、これは全数調査である。一方、テレ ビの視聴率は、一部の世帯だけを統計的にランダムに選んで調査を行っているので、標本 調査である。

2、母集団と標本

調査対象が大規模なときには、全数調査は大きなコストがかかるため、実施が難しい。 また、製品の強度調査のように調査によって調査対象が破壊されるような場合も全数調査 は行えない。そのような場合は、標本調査が行われる。標本調査において、調べる対象と なる集団全体を**母集団**という。また、母集団から取り出された手段を**標本**といい、母集団 から標本を抜き出すことを**抽出**という。

母集団	統計的な分析を行おうとしている対象の集団全体。たとえば、日本人に
	ついてのデータを取ろうと考えている場合、母集団は日本人全員となる。
	物理現象についての実験データを取ろうと考えている場合は、母集団は
	真の値が現れるための実験回数ということになる(つまり無限大)。また、
	母集団を構成するデータの数を 母集団の大きさ という。
標本	実際に母集団全体を調査したり測定したりすることは不可能であるの
	で、母集団の中から一部を取り出して実験・調査することになる。この
	とき母集団から取り出したデータ集団を標本という。また標本を構成す
	るデータの数を 標本の大きさ という。たとえば、日本人についてのデー
	タを取ろうとして、1000 人からデータを集めた場合、標本の大きさは
	1000人となる。物理実験で 50回実験を行ってデータを得た場合、標本
	の大きさは 50 となる。

3、標本の抽出

母集団から標本を抽出するときには、母集団から偏りなく抽出することが必要である。 たとえば、日本人の身長を知りたいときには、性別や年齢の構成が日本人全体と同じよう になるように調査対象を決めなければならない。もしかしたら、住む地域も身長に影響し ているかもしれないから、それらの条件もそろえる必要がある。さらには、家庭の経済状 況も……このように、すべての条件をあぶりだしてそれらをそろえた集団を用意するのは 簡単なことではない。そこで、統計学的な調査においては、母集団の各個体がなるべくラ ンダムに選ばれるように標本を抽出する方法がとられる。これを**無作為抽出**(ランダムサ ンプリング)という。無作為抽出では、標本の性質が母集団からいくらかずれてしまう(誤 差が生じる)ことは避けられない。しかしながら、完全にランダムに選ぶことで「確率論」 が使えるようになり、そのような誤差も含めて推定できるようになる。一部のデータから 全体の傾向について述べたいときには、データは無作為抽出しなければならない。反対に 無作為抽出されていないデータから、安易に全体の傾向を断定することはできないので注 意が必要である。

4、母集団の推定

母集団と標本には、それぞれ平均値や分散、標準偏差などの統計量が存在する。しかし、 母集団の統計量を求めることは難しいため、多くの場合は標本の統計量から推定すること になる。



5-2 標本平均から母平均を推定する

1、標本平均は母集団の平均とは異なる

この章の冒頭で、「その平均値が、調べたい対象の真の平均値を表しているとは限らな い」ということを述べたが、それは「標本の平均値(**標本平均**という)は母集団の平均値 (**母平均**という)を表しているとは限らない」という意味である。たとえば、無作為抽出 された 100 人の日本人の身長の平均値が 160cm だったとしても、日本人全体の身長の平均 値が 160cm とは限らない。標本平均が母平均と必ずしも一致しないことから、標本平均の 大小関係もまた、母平均の大小関係と一致するとは限らないといえる。たとえば、100 人 の日本人の平均身長が160cm で、100 人のモンゴル人の平均身長が 162cm だったとしても、 モンゴル人の平均身長が日本人の平均身長よりも高いとは限らない。もしかしたら、たま たま背が低い日本人を抽出したのかもしれないし、たまたま背が高いモンゴル人を抽出し たのかもしれない。このようなときに、標本平均から母平均を推定することで、一部のデ ータの大小から、全体の大小関係を推測することができる。

2、標本平均は母平均を中心に分布する

母集団から標本を取り出して標本平均を求める、ということを何度も繰り返して標本平 均のデータを集める。標本の大きさが大きいとき、標本平均の分布は左右対称のベル形の 分布(正規分布)を示し、この分布の平均は母集団の平均と一致することが知られている。 言い換えると、標本平均を集めると、母平均(真の平均)を中心に分布することになる。



3、標準誤差

標本平均の分布がどれくらいばらついているかということは、標本平均の分散や標準偏 差を求めることで判断できる。

標本平均の分散のことを**誤差分散**といい、標本の大きさ n が大きいとき (n \geq 30)、誤差 分散は次の式で求められる。なお、 s^2 は標本の分散を表す。

誤差分散(標本平均の分散) = $\frac{s^2}{n}$

標本平均の標準偏差のことを標準誤差といい、次の式で求められる。

標準誤差(標本平均の標準偏差:SE) = $\frac{s}{\sqrt{s}}$

標準誤差は、誤差分散の平方根をとって、単位を測定単位と等しくした値である。標本 誤差は「データの平均値が真の平均値(母平均)からどれくらいずれているか」というこ とを示す指標となる。したがって、標準誤差を求めることで、標本平均からどれくらいず れた範囲に真の平均(母平均)が存在するのかを推定することが可能になる。

4、信頼区間から母平均を推定する

母集団から標本を100回取り出して標本平均を求め、「標本平均±標準誤差」の範囲内 に母平均が収まるかどうかを検討すると、100回中68回くらいの確率で母平均がその範囲 に含まれることが知られている。このような区間のことを、**信頼区間**といい、「標本平均 ±標準誤差」は68%信頼区間と呼ばれる。信頼区間を用いることで、標本平均から母平均 の位置を確率的に推定することができる。信頼度として68%はあまり高くないため、通常 は、95%信頼区間が用いられる。標本の大きさnがn≧30のとき、95%信頼区間は次のよ うにして求めることができる。

95%信頼区間:標本平均±1.96×標準誤差

5、信頼区間から平均の差を判定する

2つの母集団から取り出した標本について平均の大小を判断したい。このとき、95%信 頼区間が判断の指標となる。たとえば、2つの標本平均について、95%信頼区間が重なっ ていれば、母集団の平均に差がないとは言い切れないだろう。一方、95%信頼区間が重な っていなければ、かなりの高確率で母集団の平均に差があると言えそうだ。

標本調査においては、平均値の差や傾向を平均値だけで判断してはいけない。この章で 説明してきたようなステップを踏んで差の有無を判断する必要がある。

参考 誤差分散や標準誤差が n や √n で割られることの意味

母集団から取り出す標本の大きさによって標本平均の分布は変化し、小さい標本集団ほど 平均値の散らばり具合は大きくなる。たとえば、100匹のテントウムシ集団から5匹を取り出 して体長を測った場合、どの5匹を取り出すかによって、平均値がばらつくことが想像でき る。一方50匹を取り出して測った場合は、5匹のときよりも平均値のばらつきは小さくなる だろう。誤差分散や標準誤差の値が n や \sqrt{n} で割られていることは、上記のような効果を 表している。つまり、標本の大きさ (n) が大きくなると、平均値のばらつき(誤差分散、標 準誤差) は小さくなるのである。

参考 誤差分散の求め方

母集団の分散を
$$\sigma^2$$
、標本の分散を s^2 、データXの集団における分散をV(X)と表す。
 $V(\bar{x}) = V(\frac{1}{n}(x_1 + x_2 + \dots + x_n))$
 $= \frac{1}{n^2} \{V(x_1 + x_2 + \dots + x_n)\}$
 $\therefore x \& a \oplus bc \oplus b \oplus c \oplus b \oplus a^2 \oplus (\beta \oplus b \oplus c^{-} - \beta \& 2 \oplus c^{-} + b \oplus c^{-} \oplus c^{-}$

やすい。

①範囲………実際のデータの分布範囲を示す。データの最大値と最小値を上下の"ひげ" で表す。

②標準偏差……データの散らばり具合を示す。

③標準誤差………真の(母集団の)平均値が収まる範囲を知る指標となる。平均値の上下にそれぞれ標準誤差の分だけ伸ばした"ひげ"で表す。標準誤差のエラーバーの範囲内に真の平均値が収まる確率は、データ数>10であれば、おおよそ 68%である。エラーバーを2倍した範囲内に真の平均値が収まる確率は、データ数>10であれば、おおよそ 95%である。なお、データ数が 10以下の場合は、データ数が少なくなるほど、エラーバーの範囲内に平均値が収まる確率が減少する。データ数=3のときには、エラーバーを4倍に伸ばしてやっと、真の平均値がその範囲内に含まれる確率が 95%となる。

④信頼区間……真の平均値が収まる範囲を知る指標となる。信頼度の数値が大きいほど、
 その区間に平均値が収まる可能性が高い。信頼区間は標準誤差から求める
 ことができる。データ数によって、信頼区間の幅が変化するので、対応表
 (t分布表)を用いて計算する。なお、データ数が30以上のときは、値が
 ほとんど変化しないため、本文の説明ではその一定値を示してある。



分かる。



ーバー: D と F は「差があ る」と言えそうだが、D と E は「差がある」とは言えそう にない。

6. 関係性の分析

ここまでは調査項目が1つの要素(1変量)のデータについて分析する方法を紹介してき た。しかし、世の中で起こる様々な現象には、複数の要素が互いに関連しながら影響を与え ていることが多い。この章では、2つの要素(変量)の間の関係性を調べる方法について見 ていこう。

6-1 散布図と相関

1、相関とは

2つの変量において、「一方の値が大きくなると、他方も大きくなる」という関係を正 の相関(相関関係)があるという。逆に「一方の値が大きくなると、他方は小さくなる」 という関係を負の相関(相関関係)があるという。どちらの関係もないときには相関(相 関関係)がないという。2変量の関係を知るには、散布図に表すとよい。



2、共分散

相関関係を数値化する方法として、**共分散**がある。右の 表のようなn個のデータの組合せがあるとき、共分散 *s_{xy}* は*x*の偏差と*y*の偏差の積の平均値として表される。

データ番号	変量x	変量y
1	<i>x</i> ₁	y_1
2	<i>x</i> ₂	y_2
3	<i>x</i> ₃	y_3
	:	
n	x_n	y_n

共分散
$$S_{xy} = \frac{1}{n} \{ (x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}) \}$$

共分散が正のときは、xの偏差×yの偏差が正になるものの割合が大きいと考えられる。 つまり、xの偏差とyの偏差が両方とも正か、両方とも負であるものが多いときに、共分 散が正になる。このとき、散布図ではデータは右上がりの分布を示すので、共分散が正の ときには、正の相関があるといえる。反対に負の時には負の相関があるといえる。

共分散によって相関の正負を判断することはできるが、相関の強弱は判断できない。た とえば、身長と体重の相関をみるときに、体重データの単位を kg から g に変えると、同 じデータを見ているにも関わらず、共分散は 1000 倍の値となってしまう。

3、相関係数

共分散の大小はデータのスケールの大小の影響を受けるため、共分散だけでは相関の強 さは分からない。そこで、相関の強弱を見るための指標が必要となる。その1つが相関係 数である。相関係数rは共分散を、xの標準偏差 s_xとyの標準偏差 s_yの積で割ったもの である。これにより、どんなスケールのデータであっても、相関関係の強弱を同じ指標で 比較することができる。

相関係数
$$r = \frac{s_{xy}}{s_x s_y}$$

相関係数は $-1 \leq r \leq 1$ の値をとり、





相関係数がどの程度大きければ「強い相関」と言えるかについては、絶対的な基準は存 在しない。以下におおよその目安としての基準を挙げておく。

相関係数	$-1.0 \sim -0.7$	$-0.7 \sim -0.4$	$-0.4 \sim -0.2$	$-0.2 \sim 0.2$	0.2~0.4	0.4~0.7	0.7~1.0
目1)年	強い	中程度の	弱い	ほとんど	弱い	中程度の	強い
判理	負の相関	負の相関	負の相関	相関なし	正の相関	正の相関	正の相関

6-2 相関を見るときの注意

1、相関係数は外れ値の影響を受ける

相関係数は外れ値の影響を強くうける。一見相関 があるように見えても、外れ値を取り除くと相関が 消えることがある。

たとえば、右の図はある仮想の会社における年齢 と年収について示した図である。このデータ全体の



相関係数を計算すると、0.93となり、強い相関がみられる。しかし、右上の2つの点を取り除くと、相関係数は0.32となり、弱い相関しか見られなくなる。

外れ値を取り除くかどうかの判断は慎重に行う必要がある。正当な理由もなく外れ値を 除去すれば、それはデータの改ざんということになりかねない。外れ値が生じたときには、 その理由を考察することが大事だ。例えば、条件が大きく異なっていたために外れ値が生 じた可能性もある。上のグラフの右上の2点は、もしかしたら社長と副社長のデータだっ たのかもしれない。その場合、管理職のデータを取り除き、条件を「平社員のみ」にそろ えることで、より具体的に相関を考察することができる。あるいは、上のグラフでは、40 歳代から 50 歳代のデータや、25 歳以下の社員のデータが抜けているのかもしれない。そ の場合は、抜けているデータを補完して再び全体で相関を調べてみるべきだろう。

2、相関係数は直線的な関係しか見極められない

相関係数は直線的な関係しか見極めることができない。例え ば、右のデータではxとyの間に二次曲線の関係が見られる。 しかし、このデータの相関係数を計算すると、0.0となる。相関 係数は便利だが、それだけで関係性を判断せずに、必ず散布図 を描いて分布を確認した方がよい。なお、散布図は外れ値を見 極める上でも有効である。



参考 アンスコムの例

下の4つのデータは、x と y の平均値、分散、x と y の相関係数がいずれも等しい。しか し、散布図のようすは大きく異なっている。これは、統計学者のフランク・アンスコムが紹介 した事例で、「アンスコムの例」として知られている。これらの例は、散布図で分布の様子を確 認することの重要性と、外れ値が与える影響の大きさについて、分かりやすく物語っている。



3、グループに分けると相関が変わることがある

右の図Aは前に掲載した散布図(年齢と年収の関係 図)である。ただし、高収入の2人(右上の2点)は 除いてある。このデータ全体の相関係数は0.32となり、 相関関係は弱い。これを男女別のデータに分けてみて みたものが図Bである。■が男性のデータ、令が女性 のデータを示す。このように同様の条件を共有する集 団(ほぼ均質と考えられる)にデータを分けることを、 層別という。Aのデータを男女の層別にしてみてみる と、男性よりも女性の方が、年収が低いことが分かる。 また、男性のみのデータでは相関係数が0.56、女性の みのデータでは0.66となり、層別にすることで、年齢 と年収の相関関係は強まる。反対に、層別にすること で、相関関係が弱まることもある。データを分析する



ときには、集団の均質さに注目する必要がある。集団が均質ではなく、それが相関関係に 影響を与えているようであれば、より均質な集団にデータを分けて、それぞれについて関 係を調べてみる必要がある。

4、全体で見ると相関が変わることがある

右の図Cは大学入試の成績と大学入学後の成績の 散布図である(仮のデータ)。このデータの相関係数 は0.10であり、ほとんど相関はない。このことから、

「大学入試は入学後の学力の良し悪しを反映しない ので、無意味だ」と言えるだろうか。この理屈には大 きな見落としがある。このデータは合格者のみを対象 としている。入試による選抜の意味を考えるのであれ ば、不合格だった生徒も含めて相関を検討しなければ ならない。もし、不合格者も含めた全体のデータ分布 が図Dのようになったとしたら、どうだろう。図Dの データ全体の相関係数は 0.84 となり、今度は強い相 関が認められる。データ分析を行うときには、「一部 のデータしか見ていないために、相関が出ない(ある いは出る)のではないか」という可能性を頭に置いて おく必要がある。



5、相関があるからといって因果関係があるとは限らない

相関関係とは、一方が増えればもう一方が増える(または減る)という関係を意味する のにすぎない。したがって、それが**因果関係**(原因と結果の関係)を示すとは限らない。 例えば、「お菓子をよく食べる子どもほど、成績が悪い」という相関関係がデータから得 られたとする(あくまでも仮定である)。だからといって、「お菓子を食べると、成績が悪 くなる」とは言えない。成績が悪いのは、お菓子が直接の原因ではないかもしれない。経 済的条件や教育環境など別の要因があって、それが「成績」と「お菓子を食べる頻度」の 両方に影響を与えているのかもしれない。あるいは、因果関係が逆の可能性もある。成績 が悪い子どもはストレスがたまるので、それを発散するためにお菓子を食べるのかもしれ ない。

現象Aと現象Bに相関関係があるとき、その原因には次の3通りが考えられる。

A→B(因果関係あり:AがBの原因になっている)

2) B→A(逆の因果関係あり: BがAの原因になっている)

3) C→A、C→B(因果関係なし:別の要因CがAとBの両方の原因になっている)

相関係数からは、これらのうちのどの関係であるかは分からない。もし因果関係をはっ きりさせようと思ったら、時間軸を考慮したデータの検証をしたり、実験的なデータの集 め方をしたりする必要がある。

※ 時間軸を考慮したデータの検証

……Aが起こってからBが起これば、BがAの原因である可能性があるが、逆であ ればその可能性はない。

※ 実験的なデータの集め方

……A以外のすべての条件を同じにした集団をAがある場合とない場合の2グル ープに分けて、それぞれでBが起こるかどうかを調べる。

因果関係がないが、相関関係があるために、まるで因果関係があるかのように見える関係(上に挙げた3つのうち、2と3の場合)のことを、疑似相関(疑似相関関係)という。

1、回帰分析とは

ある変量が別の変量によってどのように説明されるかを分析することを、回帰分析という。端的にいうと、回帰分析とは、ある変量と別の変量の関係を数式で表すことである。 変量が2つの場合の回帰分析を単回帰分析と言い、変量が3つ以上の場合の回帰分析を重 回帰分析という。前の章では、2つの変量の相関について学習したが、xとyに相関があ るときに、yをxの式で表すのが単回帰分析である。

回帰分析によって得られた数式は、データの傾向を表した直線や曲線の式として表現される。回帰分析により得られた線を回帰線といい、直線の場合は**回帰直線**、曲線の場合は 回帰曲線という。これらはデータの分布を近似的に表した線であるため、近似線、近似直 線、近似曲線などとも呼ばれる。

ここでは2変量の関係を調べる単回帰分析のうち、とくによく用いられる回帰直線の求 め方を紹介する。

2、回帰直線の求め方

理科の実験ではデータに散らばりが生じる。これまで、実験データをグラフに表すとき には、「おおよそデータの中心を通るように直線(あるいは曲線)を引く」と習ってきた のではないだろうか。この直線が回帰直線である。ただし、厳密には「おおよそ中心を通 るように」引くのではなく、数学的な計算をしたうえで直線の式を決定する。

回帰線の式は、最小二乗法という方法 によって求められる。たとえば、右の図 のようなデータの回帰直線を求める場 合を考えてみよう。回帰直線と実際のデ ータとのy軸方向の差(残差という)に 注目する。データ全体にわたって、この 残差ができるだけ小さくなるような直 線が、データの傾向を最もよく表す直線 であると考える。しかしながら、残差に



は正と負があるため、残差を二乗する。そして、データ全体の残差の傾向を見るため、残 差の二乗を全て足し合わせる(この値を残差平方和という)。この残差平方和が最小にな るように直線の式を設定すれば、これが回帰直線ということになる。回帰直線の方程式を 求める過程では微分法による計算もあり、少々複雑である。そのため、ここでは詳細な説 明は省く。

3、回帰直線の精度を表現する:決定係数

回帰直線を求めたとして、本当にデータが求めた回帰直線に当てはまると言ってよいの かどうかは疑問が残る。たとえば、下図の(I)と(II)は回帰直線の式は同じであるが、 (I)の方がデータの「当てはまり」が良い。このような「当てはまり」の良さ、つまり 回帰線の精度を表す指標がある。それが、決定係数である。



決定係数の求め方の詳細は省くが、最小二乗法によって回帰直線を求めた場合、決定係数は相関係数の2乗と同じ値となる。そのため、決定係数は \mathbf{R}^2 と表され、最小二乗法による回帰直線においては、 $0 \leq \mathbf{R}^2 \leq 1$ の値をとる。 \mathbf{R}^2 が1に近いほど方程式がデータを「よく説明している」ことになり、0に近いと「ほとんど説明していない」ことになる。

Excel では、散布図から簡単に回帰直線を求めることができる。直線を引いたら、なん となく相関があるように見えるし、その直線に当てはまるように思えてくるが、その際に は、必ず R²値(決定係数)も表示させて、直線の「当てはまり」を確認するようにしよ う。

4、回帰線を求めるときの注意

回帰線は直線になるとは限らない。たとえば、右図のよ うなデータは直線で近似することもできるが、明らかに二 次曲線を描いているので、曲線で近似した方がよい。この ように、データの近似線をどのような線にするのかは、デ ータの分布を見て判断することがある。しかしながら、で



きれば、なぜ「直線を選んだのか」「二次関数を選んだのか」「指数関数を選んだのか」と いったことに答えられた方がよい。たとえば、落下時間と落下速度の関係であれば、一次 式で近似した方がよいだろう。あるいは、落下時間と落下距離の関係であれば、二次式で 近似した方がよいだろう。データに回帰線を追加する場合は、そのデータの関係性を予測 する理論があるかどうか調べてみると良い。もしよい理論があれば、その理論から予想さ れる関数で回帰線を作成するとよい。

参考決定係数の求め方

決定係数の求め方を少し詳しく見てみよう。 右の図で、i番目のデータ(yi)に着目 すると、aは、このデータの平均値から のずれを表す。一方、○は回帰線による 推定値を示しており、bは、その推定値 が平均値からどれくらいずれているか を表す。aやbは負の値をとることもあ るので、2乗する。この2乗値をi番目 だけでなく、すべてのデータについて合 計する(つまり、平方和を求める)。こ のとき、「bの平方和÷aの平方和」が 決定係数となる。



これは、データの値と推定値のそれぞれについて分散を計算して、その比をとっているのと 同じ計算になる。したがって、決定係数とは、実際のデータの分散に対して、回帰線で推定さ れるデータの分散が占める割合を表すことになる。言い換えれば、平均値からのずれのうち、 どれくらいの割合を回帰線で説明できるかを表している。たとえば、決定係数が 1.0 であれば、 回帰線による推定値と実際のデータで平均からのずれがすべて等しくなるので、「その回帰線は 実際のデータの分散を 100%説明できている」という理解になる。

参考文献

池田郁男(2015)『実験で使うとこだけ生物統計①』羊土社

市原清志(1990)『バイオサイエンスの統計学』南江堂

伊藤公一朗(2017)『データ分析の力 因果関係に迫る思考法』光文社新書

薩摩順吉(1989)『理工系の数学入門コース7 確率・統計』岩波書店

総務省政策統括官(統計基準担当)(2016)『生徒のための統計活用~基礎編~』日本統計協 会

総務省政策統括官(統計基準担当)(2017)『高校からの統計・データサイエンス活用〜上級 編〜』総務省

ダレル・ハフ著、高木秀玄訳(1968)『統計でウソをつく法』講談社ブルーバックス

坪井俊ほか(2014)『数学B』数研出版

涌井良幸・涌井貞美(2015)『統計学の図鑑』技術評論社

※ Excel は米国 Microsoft Corporation の米国およびその他の国における商標もしくは登録商標です。本 テキストでは、™および®マークの記載は省略しました。

究理ⅡD テキストブック

	2018年4月	発行	
	2019年4月	改訂	
	2020年4月	改訂	
発行者	滋賀県立虎姫高等学校		

〒529-0112 滋賀県長浜市宮部町 2410

TEL 0749-73-3055

年 組 番 名前