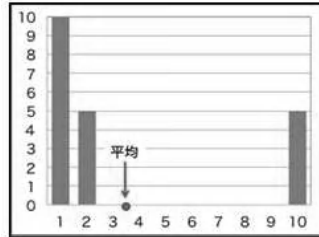


2. データ特性の表し方

平均のウソ

平均は、平均点、平均身長、平均所得などよく使われる特性値であるが、極端に大きな数値や小さな数値に影響を受けやすく、データの中心を的確に表さないことがある。たとえば、次の図のようにある試験で1点が10人、2点が5人、10点が5人の場合、平均点は3.5点であるが、これが回答者の分布特性を的確に示しているとは言い難い。



下の表のB社は、10年未満、10年以上ともにA社よりも平均月給が高い。しかし、社員全体の平均月給を計算すると、A社の方が高くなる。このことをもってA社の月収が高いというのは妥当ではない。

勤続年数	A社		B社	
	平均月給	人数	平均月給	人数
10年未満	26万円	50	30万円	100
10年以上	48万円	150	50万円	100
全体	42.5万円	200	40万円	200

(田中 [2009] pp.13-14)

このように、平均値が集団の誤った特性を示してしまうことを平均のウソという。平均値を使用したり、平均に関する記述を読むときは、平均のウソがないかどうか注意する必要がある。

中央値 (median)

量的データを小さい順に並べたとき、中央に位置するデータの値を中央値(メディアン)という。データの数が偶数の場合は、中央の2つの平均とする。たとえば、以下のような6つのデータの場合、平均値は外れ値である90の影響

B34

\$L\$34

2. データ特性の表し方

- ・ B3 に「地域別失業率」シートの 2015 年の北海道のデータと全国平均の差を 2 乗した値を計算する
このとき平均を絶対セル参照にする
B3 の式は「=(地域別失業率!~~L\$3~~ - 地域別失業率!~~B3~~)^2」となる
- ・ B3 の式を K3 までコピーする
- ・ L3 に B3 から K3 間での合計を 9 (= 10 地域 - 1) で割り、その平方根を計算する
平方根を計算するワークシート関数は「=SQRT(データ)」である
L3 の式は「=SQRT(SUM(B3:K3)/9)」となる

b) 関数を使った計算

- ・ Excel には標準偏差を計算する関数が組み込まれているので、これを使うと簡単に計算できる「標準偏差」シートの M3 に関数を使って標準偏差を計算し、L3 と同じ値になっているかどうか確認してみよう
標準偏差を計算するワークシート関数は「=STDEV.S(データ範囲)」である

豆知識

各データから平均を差し引き、標準偏差で割ると、データは平均 0、標準偏差 1 となる。このような変換をデータの標準化という。標準化されたデータは、測定単位に依存しないデータとなる。いわゆる偏差値は、平均 50、標準偏差 10 となるよう標準化した値である。

変動係数

標準偏差は、測定単位に応じて変化する。そのため、たとえば生産額を 1,000 万円単位で表示した場合は 1 億円単位の標準偏差の 10 倍となり、ばらつきが大ききような印象を与えてしまう。また、平均が大ききほど標準偏差は大きくなる。そこで、測定単位に依存しないよう変換した特性値として変動係数が使われる。ただし、変動係数は一般に平均が正のデータの場合に使われる。

$$\text{変動係数} = \frac{\text{標準偏差}}{\text{平均}}$$

2. データ特性の表し方

練習問題 2.7

「地域別失業率」のデータを使って、O 列に毎年の標準偏差を、P 列に毎年の変動係数を計算しよう。標準偏差と変動係数の推移を表す折れ線グラフを作成し、失業率のばらつきが 1985 年以降どのように変化したのかを観察してみよう。

消費支出

3) 第 2 章の課題

「家計調査」のデータを使って、支出項目別に年齢階層全体の平均、中央値、範囲、標準偏差、変動係数を計算しよう。

- ・ 数値の表示形式は桁区切りスタイルとし、小数点第 1 位まで表示させる
- ・ 「家計消費支出」シートを「課題 2」というファイル名で保存する
- ・ 分析結果から、どのようなことがわかるかを考察してみよう

4) 統計学からの補足

母集団 (population)

統計データが収集されたもとの集団を母集団という。これは統計分析の対象となる全体であり、通常非常に膨大なものである。その一部である統計データは母集団から任意に得られたもの(任意標本)で、母集団の情報を代表するものと捉えられる。その情報を精度高く取り出すことが統計分析の役割といえる。

母集団には質的なものと量的なものがあり、前者からは質的データが、後者からは量的データが収集される。本章で学んだ統計値は主に量的母集団に対するものである。量的母集団の中心的位置を表す代表的な特性量が母平均であり、ばらつきの程度を表す特性量が母分散と母標準偏差である。これらの値も含め、母集団の特性量を推測することが統計分析において求められる。

推定量 (estimator)

母集団の特性量に使われる統計値は不偏性と呼ばれる性質をもつことが必要である。不偏推定量は、その平均が対象となる特性量に一致するもので、この性質が推定として使われる根拠となる。

- ・ 図表の内容について必ず本文で説明する
- ・ 原則として本文の図表に関する説明の後に図表を配置する
- ・ グラフの下にデータの出所を以下のように記す
ex. 出所：内閣府「国民経済計算」より筆者作成

データの特徴を記述する際のポイント

グラフからデータの傾向を読み取る場合は、原則として以下のように記述するのが望ましい。

- ・ はじめにどのグラフについて説明しているのかが分かるようにし、グラフが何について示しているのかを記す~~全体の傾向から、部分的な特徴へ~~
ex. 「図 1 は 1994 年以降の日本の名目 GDP の推移を示している。……」
- ・ データの水準がどのくらいなのかを記す
ex. 「日本の GDP は 1990 年代半ばには約 520 兆円であったが ……」
- ・ 数字は半角で入力する
ex. × 「1 9 9 4 年」 → ○ 「1994 年」
× 「5 0 0 兆円」 → ○ 「500 兆円」
- ・ 掲載した図表については必ず本文で説明する（図表を掲載しながら説明しないことがないように）
- ・ 全体の傾向から、部分的な特徴へ
ex. 年齢階層別の雇用動向に関するグラフの場合、全体（年齢計）の特徴を述べてから、年齢階層別の特徴へ
- ・ 時系列データは変化前から変化後へ
ex. GDP の例では 90 年代半ばから 2000 年代初頭を経て 2008 年まで、そしてそれ以降へ
- ・ 上昇（増加）しているのか、下落（減少）しているのか、横ばいか

であるから、この間にユーロは0.11%上昇したのに対して、円は20.58%も下落したことが分かる。

指数には単純指数のほかに、複数のデータを総合した総合指数、複数の指数を組み合わせた複合指数がある。

消費者物価指数

よく使われる指数には、消費者物価指数や~~卸売~~物価指数がある。いずれもさまざまな商品の価格を総合して全体として物価がどう推移しているかを見るための総合指数である。消費者物価指数は、平均的な消費者が比較時に基準時と同じ商品を購入した場合、商品価格の変化により、購入総額がどのように変化するかを示す指標である。一般的には次式で定義される。 p は価格、 q は購入量、右下の添え字は基準時が0、比較時が t 、 $1 \sim n$ は個々の品目を示す。

$$\text{消費者物価指数} = \frac{(p_{t,1} \times q_{t,1}) + (p_{t,2} \times q_{t,2}) + \cdots + (p_{t,n} \times q_{t,n})}{(p_{0,1} \times q_{0,1}) + (p_{0,2} \times q_{0,2}) + \cdots + (p_{0,n} \times q_{0,n})} \times 100$$

消費者が購入する商品には、総務省統計局「家計調査」に基づいて約600品目が含まれているが、仮に米、味噌、醤油の3品目だけだったとして、基準年を2000年、比較年を2015年としたら消費者物価指数は次のように計算される。

$$\text{消費者物価指数}_{15} = \frac{\text{2015年の購入総額}}{\text{2000年の購入総額}} \times 100$$

ただし、

$$\begin{aligned} \text{2015年の購入総額} &= (\text{米の価格}_{15} \times \text{米の購入量}_{15}) \\ &\quad + (\text{味噌の価格}_{15} \times \text{味噌の購入量}_{15}) \\ &\quad + (\text{醤油の価格}_{15} \times \text{醤油の購入量}_{15}) \end{aligned}$$

$$\begin{aligned} \text{2000年の購入総額} &= (\text{米の価格}_{00} \times \text{米の購入量}_{00}) \\ &\quad + (\text{味噌の価格}_{00} \times \text{味噌の購入量}_{00}) \\ &\quad + (\text{醤油の価格}_{00} \times \text{醤油の購入量}_{00}) \end{aligned}$$

ここで、右下の添え字は00が基準年の2000年、15が比較年の2015年を示す。

3. 時系列データの分析

練習問題 3.3

「国内総生産」シートのC列以降は国内総生産の内訳を示している。まずは、GDPの内訳の推移を折れ線グラフにして、その変化を確認してみよう。次に、このデータを使って、GDP各項目の増加率と指数を計算しよう。

折れ線グラフの作成

国内総生産の内訳のグラフを作成する際には、A列に横(項目)軸ラベル、C列以降にデータがあり、データ範囲が連続していない。このような場合は、次のような手順で範囲指定をし、グラフを作成する。

- ・ A3 から A²⁵~~24~~ までの範囲を選択する
- ・ **[Ctrl]** キーを押しながら、マウスで C3 から H²⁵~~24~~ までをドラッグする
(A3の空白セルも選択範囲に入れることに注意する)
- ・ 折れ線グラフを作成する
- ・ 横(項目軸)がラベルと横軸と重なっているので修正するため、横軸で右クリックし、「軸の書式設定」から「ラベル」の「ラベルの位置」を「下端/左端」に変更
- ・ グラフタイトルは削除し、凡例の位置は右にする
- ・ グラフの左上が J17 の左上と重なるようグラフを移動させる


増加率と指数の計算

以下の手順でGDPとその内訳の増加率と指数(1995年を100とする)を計算する。

a) 増加率の計算

- ・ A31に「GDPの増加率」と入力する
- ・ データ系列名を入力するためA3からH3をコピーし、A32に貼り付ける

3. 時系列データの分析

- ・ 横(項目)軸ラベルを入力するため A⁵から A²⁵を選択し、セルの内容をコピーし、A33 へ貼り付ける (1994 年の行は作成しない)
- ・ B33 に「=(B5-B4)/B4」を入力する
- ・ B33 をコピーし、B33 から H⁵³を選択し、「ホーム」タブのペーストボタンの右の ▼ をクリックし、 をクリックして数式を貼り付ける

b) 指数の計算

- ・ A55 に「GDP 指数(1995 年=100)」と入力する
- ・ 横(項目)軸ラベルを入力する (GDP の表からコピーして貼り付ける)
- ・ 同様に縦軸ラベルも入力する
- ・ ~~データ系列名を入力する (GDP の表からコピーして貼り付ける)~~
- ・ B57 に「=B5/B\$5*100」を入力し、他のセルにコピーする

トル

c) グラフの作成

- ・ 民間最終消費支出から公的資本形成までの 4 項目について (在庫品増加と純輸出は変化が大きいため省略)、増加率のグラフ、指数のグラフをそれぞれ作成し、GDP の内訳のグラフと同じように見た目を整える

寄与度

時系列データ全体の変化に対して、その構成項目であるデータがどの程度影響したかを分析する際に、寄与度ないしは寄与率を使用する。

寄与度は、ある構成項目のデータの増減が、全体の増加率を何 % ポイント変化させているかを示している。たとえば、2010 年の GDP 増加率は 2.4 % であったが、このうち消費支出は 0.6 % ポイントであった。

$$\text{寄与度} = \frac{(\text{構成項目の当年データ} - \text{構成項目の前年データ})}{\text{前年の全体のデータ}} \times 100$$

3. 時系列データの分析

寄与率

寄与率は、全体の増加率を 100 とした時に、構成項目の寄与度がその何 % かを示したものである。上の例では、GDP の増加率 2.39 % のうち、消費支出の寄与度が 0.62 % ポイントであるから、寄与率は $0.62 \div 2.39 = 26.0$ (%) となる。

$$\begin{aligned} \text{寄与率} &= \frac{(\text{構成項目の当年データ} - \text{構成項目の前年データ})}{(\text{全体の当年データ} - \text{全体の前年データ})} \times 100 \\ &= \frac{\text{構成項目の寄与度}}{\text{全体の増加率}} \times 100 \end{aligned}$$

豆知識

パーセントとパーセントの差を表す時にはパーセント・ポイント、あるいは単にポイントと表現する。たとえば、GDP 増加率が 2.6 % から 3.0 % に変化した場合、0.4 % 増加したとするのは間違いで、「0.4 % ポイント増加」あるいは「0.4 ポイント増加」と表現する。

練習問題 3.4

「国内総生産」のデータを使って、以下の手順で GDP 各項目の寄与度と寄与率を計算してみよう。

寄与度の計算

- ・ A82 に「寄与度」と入力する
- ・ A32 から H5³までの GDP 増加率の表をコピーして A83 に貼り付ける
- ・ データ (B8⁴:H10⁴) を削除する
- ・ B8⁴に「=(B5-B4)/\$B4*100」を入力し、全体にコピーする

豆知識

数式をコピーするときに、罫線が引いてある表の中でフィル・ハンドルをドラッグすると、数式だけでなく罫線もコピーされてしまう。このとき、マウスの右ボタンでフィル・ハンドルをドラッグすると、メニューが表示されるので、「書式なしコピー」を選択すると、表示形式や罫線などはコピーされず、数式だけがコピーできる。

寄与率の計算

a) 定義式を使った計算

- ・ A106 に「寄与率」と入力する
- ・ 寄与度の表をコピーして A107 に貼り付ける
- ・ データ (B108:H127⁸) を削除する
- ・ B108 に「 $= (B5 - B4) / (B5 - B4) * 100$ 」を入力し、全体にコピーする

b) 寄与度の表を使った計算

- ・ すでに寄与度の表を作っているので、これを使うともう少し簡単に計算できる
- ・ B108 に「 $= B84 / B84 * 100$ 」を入力し、全体にコピーする

寄与度のグラフの作成

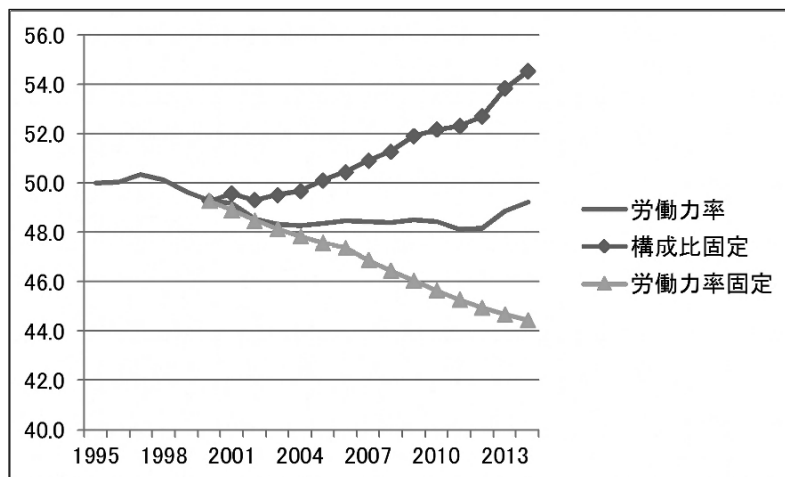
2000 年から 2014⁵ 年までの経済成長の寄与度を表すグラフを作成する。以下の手順で、下図のような構成項目の寄与度を積み上げ棒グラフとし、国内総生産は折れ線グラフとなるような図を作成してみよう。

- ・ データ範囲 (A83:H109⁴) を選択して「積み上げ縦棒グラフ」を作る
- ・ 「国内総生産」系列のグラフを右クリックし、「系列グラフの種類の変更」を選択して、「国内総生産」系列のグラフの種類を「マーカー付き折れ線」にして OK をクリック

(折れ線が太すぎてカッコ悪いので、線の太さを変更する)

- ・ A60 に「労働力率」と入力する
- ・ 上の年齢階層別の人口や労働力人口の表からコピーして表側と表頭を入力する。
- ・ 62 行に 1990 年の労働力率を計算するため、B62 に式「=B32/B3*100」を入力し、O61 までコピーする。さらに 87 行までコピーして、小数点第 2 位まで表示させる
- ・ A89 に「構成比」と入力する
- ・ 労働力率の表と同様に、表側と表頭を入力する。ただし、2000 年以降とする
B91 に式
- ・ 91 行に 2000 年の年齢別構成比を計算するため「=B13/\$B13」を入力し（分母を固定して横方向にコピーするので列を絶対参照とする）、表全体にコピーする
- ・ 労働力率の表の右端に構成比を 2000 年に固定した系列（高齢化が進展しなかった時の労働力率）と労働力率を 2000 年に固定した系列（各年齢階層で労働力率が変化しなかったと仮定した時の高齢化の影響だけを反映した労働力率）を作成するので、P61 に「構成比固定」、Q61 に「労働力率固定」と入力する
- ・ P72 に次式を入力し、P87 までコピーする
「=C72*\$C\$91+D72*\$D\$91+E72*\$E\$91+F72*\$F\$91+G72*\$G\$91
+H72*\$H\$91+I72*\$I\$91+J72*\$J\$91+K72*\$K\$91+L72*\$L\$91
+N72*\$N\$91+O72*\$O\$91」
- ・ Q72 に次式を入力し、Q77 までコピーする
「=\$C\$72*C91+\$D\$72*D91+\$E\$72*E91+\$F\$72*F91+\$G\$72*G91
+\$H\$72*H91+\$I\$72*I91+\$J\$72*J91+\$K\$72*K91+\$L\$72*L91
+\$N\$72*N91+\$O\$72*O91」
(式が長いが計算の意味をきちんと理解しよう)
- ・ 「総数」と「構成比固定」「労働力率固定」のデータで折れ線グラフを作成する

図 3-4 固定比率の労働力率



出所：総務省統計局「労働力調査」より作成

4) 第 3 章の課題

「輸出・輸入額」のデータを使って、2010 年以降の輸出額、輸入額それぞれについて、寄与度および寄与率を計算し、グラフを作成しよう。

- ・ 練習問題 3.4 のデータとは表頭・表側が逆なので注意する
- ・ 横軸を年次とする積み上げグラフとし、寄与率は総額を折れ線グラフとする
- ・ 「課題3」というファイル名で保存する トル
- ・ 分析結果から、どのようなことが分かるかを考察してみよう