第２章－１　様々なデータの特徴を把握したい

例えば　３年１組の国語のテスト結果の特徴をつかみたい

３年１組の国語のテストの点数

国語のテストの点数一覧				
70	50	100	45	35
90	90	65	50	50
85	50	40	100	95

分析の仕方

１　集合データの特徴を表す数値（代表値）の種類

○平均値：各データを足し合わせて、データの数で割った数値。

全ての値を使用して計算するため、全データを踏まえた特徴を知ることができる。
単純平均、算術平均ともいう。

【計算式】　平均値　＝　（データの合計）　÷　（データの数）
(70＋50＋100……＋100＋95) ÷　15　＝　67.7　　平均…　67.7 点
【Excel の場合】　AVERAGE 関数

○中央値：データを小さい（大きい）順番に並べ、ちょうど中央に位置する数値。

100	100	95	90	90	85	70	65	50	50	50	50	45	40	35

※データの数が偶数の場合は、中央２つのデータを足して２で割ると中央値となる。
【Excel の場合】　MEDIAN 関数

○最頻値：データの中で最も頻繁に出現する数値。

100	100	95	90	90	85	70	65	50	50	50	50	45	40	35

最も数が多いのは４個ある 50 点なので、最頻値は 50 点となる。
【Excel の場合】　MODE 関数

２　集合データの特徴を表す数値（代表値）の選び方

代表値は、一般的には以下の手順で選ぶのがよいとされている。

＜選択の手順＞
①「平均値」と「中央値」を比較する。
　⇒大きくずれていなければ「平均値」を採用
　⇒大きくずれていれば②へ
②「中央値」と「最頻値」を比較する。
　⇒どちらが適切か検討　　より適切なものを選択
　⇒どれをとっても一長一短がある場合　③へ
③データの特徴を１つの数値で示していいかを検討する。
　⇒代表値を使わないという選択もありうる

分析結果からわかること

①それぞれの代表値を算出する。

平均値　　67.7 点　　　中央値　　65 点　　　最頻値　　50 点

②選択手順に沿って、１組の国語のテスト結果の特徴を表す数値（代表値）を選出する。

３年１組の国語のテスト結果の「平均値」と「中央値」を比較する。

⇒平均値 67.7 点と中央値 65 点を比較すると、大きくずれていない（差 2.7 点）

⇒平均値が３年１組の国語のテスト結果の特徴を表す数値（代表値）として適切

３年１組の国語のテスト結果の特徴を表す数値は、平均値の 67.7 点である。

＜代表値の選択が必要な例＞

以下の場合は代表値の選択について、検討が必要である。

①極端な数値（外れ値）がある場合

例えば　あるグループの年収をみてみると以下のようになっていた。５人の平均年収を算出してみると……

280 万　350 万　480 万　540 万　3000 万　⇒　平均値 930 万円／中央値 480 万円／最頻値　なし

最頻値がないため、平均値と中央値を比べると…

平均値　⇒　3000 万が平均値を引き上げているため、５人中４人が平均値以下。

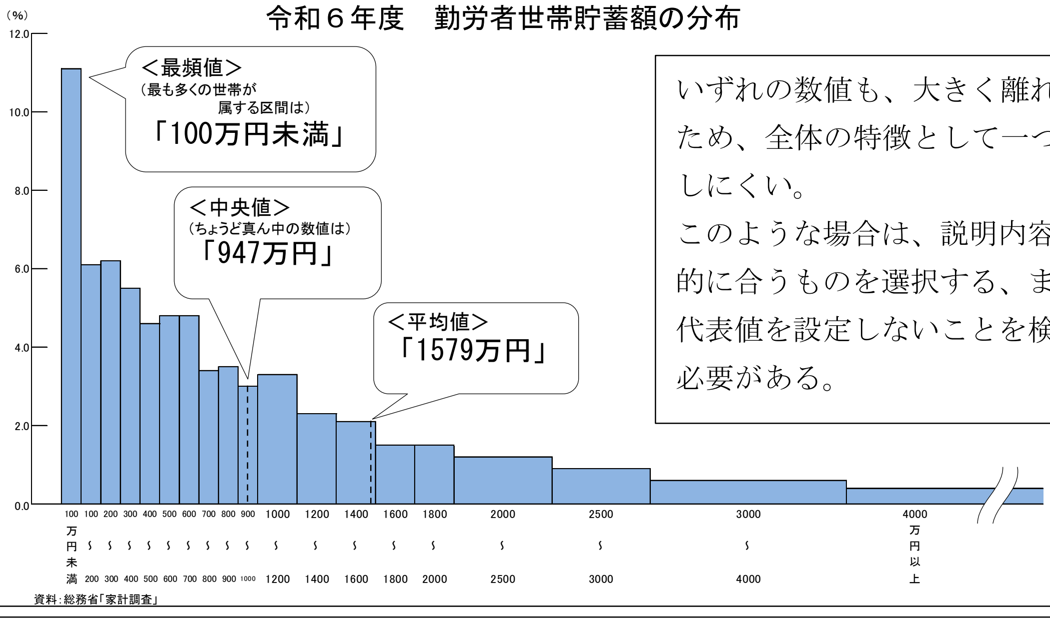
中央値　⇒　3000 万以外の４人の金額に近い。

代表値を設定するならば、中央値の方がより実情に近いと考えられる。

②データの分布が偏っている場合

例えば　令和６年度の勤労者世帯の貯蓄額の代表値を検討する場合…

令和６年度　勤労者世帯貯蓄額の分布



いずれの数値も、大きく離れているため、全体の特徴として一つを選択しにくい。
このような場合は、説明内容や、目的に合うものを選択する、または、代表値を設定しないことを検討する必要がある。

参考　＜知っておくとよい平均の算出方法＞

平均の算出方法にはいくつか種類があり、単純平均ではデータの特徴を表しきれない場合は、ほかの方法を検討する必要がある。

例えば　①人数や数量などの違い（ウエイト）を反映させた平均を知りたい（加重平均）

（例）A 中学の生徒 20 人の平均点は 70 点、B 中学の生徒 30 人の平均点は 90 点、中学生全体の平均点は？

単純平均の場合　$\dfrac{70(点)+90(点)}{2} = \boxed{80点}$

●上記の単純平均では、Ａ中学とＢ中学の生徒数の違いを反映していないため、中学生全体の正確な平均点とはいえない。

では、どうするか？

各クラスの人数をウエイトとして付加し、人数の違いを反映させる加重平均を用いる。

加重平均　＝　各データのウエイト（重み）を加味して計算する平均値

※この場合は、各中学の生徒数がウエイト（重み）となる

データ（ｘ）がｎ個、それぞれの重み（ｗ）もｎ個分あるとしたら…

【計算式】　加重平均　$= \dfrac{X_1W_1 + X_2W_2 + X_3W_3 + \cdots + X_nW_n}{W_1 + W_2 + W_3 + \cdots + W_n}$

計算結果

各中学の生徒数をウエイト（重み）として加重平均の式に当てはめてみると…

加重平均の場合　$\dfrac{70(点)\times 20(人)+90(点)\times 30(人)}{20(人)+30(人)} = \boxed{82点}$

Ａ中学とＢ中学の生徒数の差が反映された結果、全体の平均点は 82 点であった。

例えば　②増減率の平均を知りたい（幾何平均）

（例）（株）統計商事の1年目の売上は1000万、2年目は2160万、3年目は3240万であった。統計商事における3年間の売上増減率の平均は？

売上増減率　　1→2年目（+116%）、 2→3年目（+50%）

<u>単純平均の場合</u> $\dfrac{(100+116)+(100+50)}{2} - 100 = \boxed{83\%}$

＜検証してみると…＞

	（1年目）		（2年目）		（3年目）		
単純平均	1000万	×	(100 ＋ 83)％	×	(100 ＋ 83)％	＝	3348.9万

●上記の単純平均では、3年目の売上である3240万と一致しないことから、増減率の平均は単純平均では算出できないことがわかる

では、どうするか？

<u>増減率の平均を算出する場合は、幾何平均を用いる。</u>

幾何平均　＝　データを全てかけ合わせて、データ数の累乗根をとった値

データ（x）がn個あったとしたら…

【計算式】　幾何平均　＝　$\sqrt[n]{X_1 \times X_2 \times X_3 \times \cdots \times X_n}$

【Excelの場合】　GEOMEAN関数

計算結果

売上増減率を幾何平均の式に当てはめてみると…

<u>幾何平均の場合</u>　$\sqrt[2]{(100+116)\times(100+50)} - 100 = \boxed{80\%}$

＜検証してみると…＞

	（1年目）		（2年目）		（3年目）		
幾何平均	1000万	×	(100 ＋ 80)％	×	(100 ＋ 80)％	＝	3240万

幾何平均で算出した数値は、3240万と一致することから、
統計商事の売上増減率の平均は80%である。