

## 2 データの整理

テーマ「年間平均気温と関係がある項目は？」

年間平均気温と家庭の消費で関係がある項目を調べましょう。

### 学習活動

2022年の年間平均気温と家計調査の「アイスクリーム・シャーベットの消費額」、「手袋の消費額」、「サンダルの消費額」、「灯油の消費額」のデータを一つにまとめた。このデータのうち、「灯油の消費額」の外れ値を求める。

収集したデータをそのまま分析ができるわけではない。データの中には、数値が欠けている欠損値や他の値から大きく外れた外れ値が含まれているかもしれない。分析の行う前に、集めたデータを確認しておくことが重要である。

### 外れ値

今回は、「第1四分位数 $-1.5 \times$ 四分位範囲」より小さい値、「第3四分位数 $+1.5 \times$ 四分位範囲」より大きい値を「外れ値」として扱う。

※四分位範囲 = 第3四分位数 - 第1四分位数

### 外れ値の求め方

(方法1) 関数から外れ値を求める

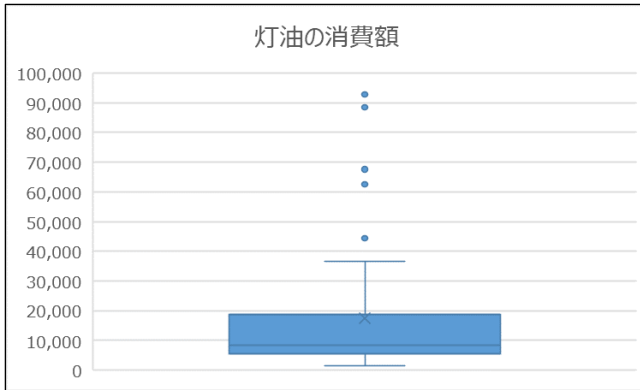
	灯油
第3四分位数 $+1.5 \times$ 四分位範囲	38680.75
第3四分位数	18727
中央値	8287
第1四分位数	5424.5
第1四分位数 $-1.5 \times$ 四分位範囲	-28090.5

中央値 (第2四分位数)、第1四分位数、第3四分位数を求めるための数式例

中央値
=median(範囲) または =QUARTILE. INC(配列, 2)
第1四分位数
=QUARTILE. INC(配列, 1)
第3四分位数
=QUARTILE. INC(配列, 3)

※範囲、配列は、外れ値を求めるための該当データ

## (方法2) 箱ひげ図から外れ値を求める



表計算ソフト (Excel 2019) で箱ひげ図を作成する

### 1. データを選択する

	F
1	灯油
2	88,359
3	92,855
4	67,609
5	22,533
6	62,541
7	44,489

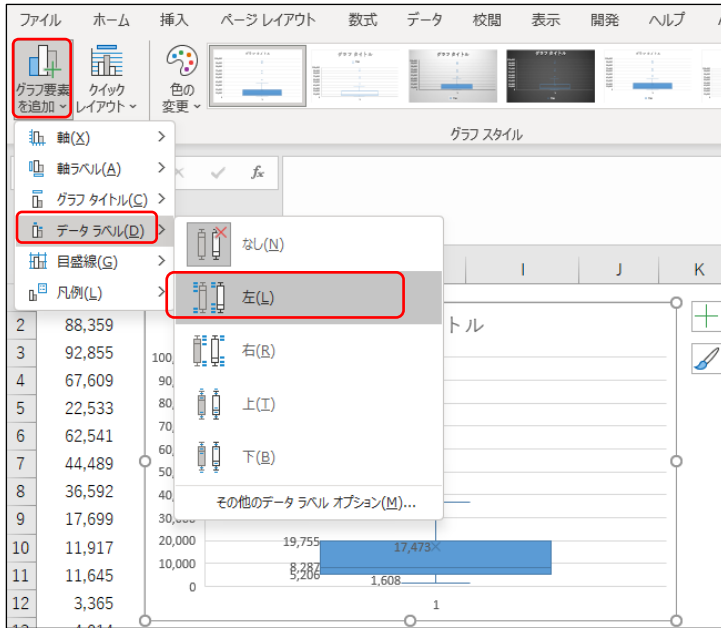
### 2. グラフを挿入する

([挿入]-[おすすめグラフ]-[すべてのグラフ]-[箱ひげ図])

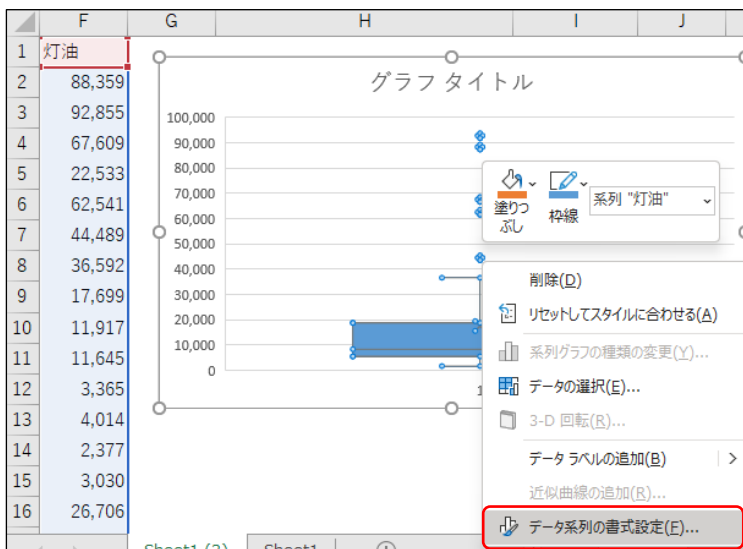
Excel 2019 の「挿入」タブの「おすすめグラフ」グループで「すべてのグラフ」を選択し、「箱ひげ図」を選択する様子。

消費額
88,359
92,855
67,609
22,533
62,541
44,489

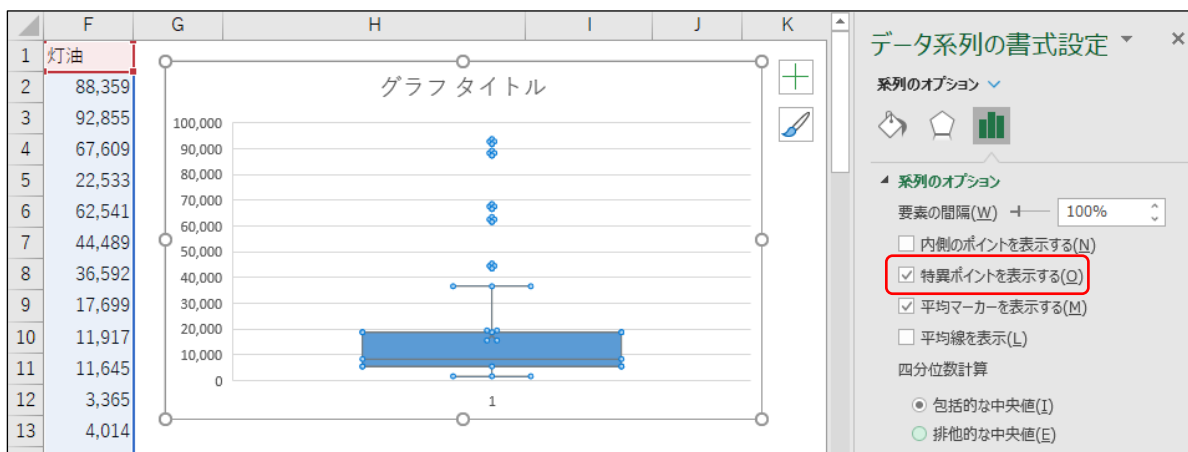
### 3. データラベルを表示する ([グラフ要素を追加]-[データラベル])



### 4. 外れ値 (特異ポイント) を表示する (該当グラフを右クリック-[データ系列の書式設定])



「特異ポイントを表示する」にチェックを入れる



(留意点)

外れ値を除いて統計処理する機会が多いが、外れ値が入力ミス等による異常値か、ある条件が重なったときに現れる値かは、分析結果を比較した上で判断する必要がある。  
外れ値を求める他の方法として、クラスター分析などがある。

学習評価例

項目	ねらい・学習活動	重点	備考
1	○データの整理 データに含まれる外れ値の扱いやデータを整理する必要性を理解するとともに技能を身につけられるようにする。	知	

重点…重点的に生徒の学習状況を見取る観点を示している。