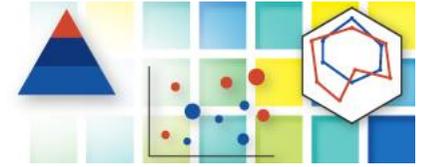




SSDSE を使ってみよう



— EXCEL を使った SSDSE データの使用例 —

第 1 回 生鮮肉の支出金額【解説編】

◎ SSDSE-家計消費から、都道府県庁所在市別の生鮮肉（牛肉、豚肉、鶏肉）の支出金額を抜き出して使用します。

使用データ SSDSE-家計消費 2024年版（SSDSE-C-2024）

分析の内容 データのばらつき、平均、四分位数、箱ひげ図、度数分布、ヒストグラム

【目次】	2	データをプロットしてみる ▶ EXCELでの作図方法	-----	2
	3	平均値や四分位数などを計算する ▶ 四分位数についての補足	-----	5
		▶ EXCELでの計算方法	-----	6
	4	箱ひげ図を作成する ▶ 箱ひげ図についての補足	-----	7
		▶ EXCELでの作図方法	-----	8
	5	ヒストグラムを作成する ▶ EXCELでの度数分布表の計算方法	-----	10
		▶ EXCELでのヒストグラムの作図方法	-----	11

※ この資料では、EXCEL2016 を使用しています。

2 データをプロットしてみる ▶ EXCELでの作図方法

◎ 各品目を1系列とする**散布図**として作図します。

(**散布図**：XとYの2つの値を持つデータの集まりを、横軸位置がX、縦軸位置がYの点の集まりとして表示したグラフ)

1 各品目の**横軸位置を示す列** (右図の**①**、**③**、**⑤**) を追加し、それぞれ同じ数値を入れておきます。

(ここでは、牛肉 = 1、豚肉 = 2、鶏肉 = 3としています。)

2 まず、牛肉のデータの散布図を作成します。

①と**②** (太線で囲んだ範囲) を選択した状態で、

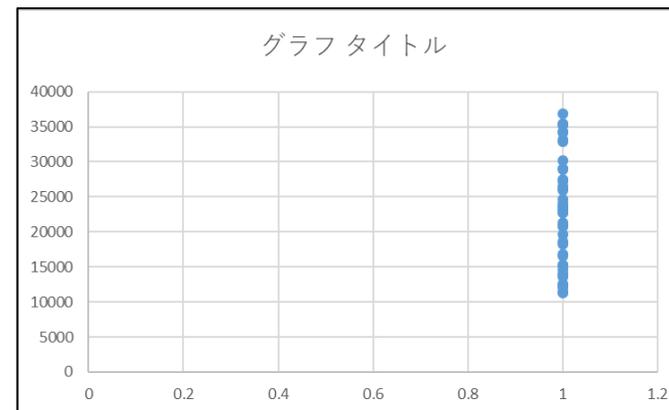
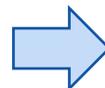
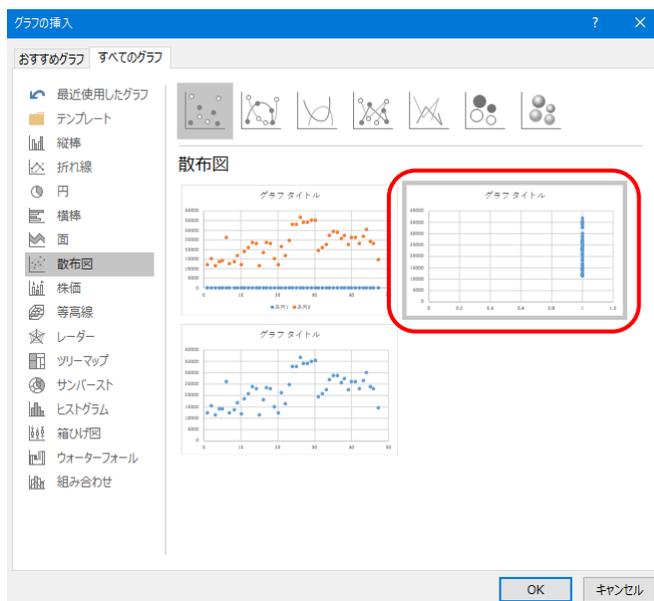
挿入 > グラフ → **グラフの挿入** を表示

すべてのグラフ > 散布図 を選択

下図のように選択し、**OK**をクリックすると、

右下図のように散布図が表示されます。

地域コード	都道府県	市	① X 1	② 牛肉	③ X 2	④ 豚肉	⑤ X 3	⑥ 鶏肉
R01100	北海道	札幌市	1	12463	2	32698	3	16903
R02201	青森県	青森市	1	15301	2	33418	3	16536
R03201	岩手県	盛岡市	1	11409	2	32621	3	14742
R04100	宮城県	仙台市	1	13876	2	32951	3	16521
R05201	秋田県	秋田市	1	14047	2	35183	3	16664
R06201	山形県	山形市	1	26151	2	34770	3	16655
R07201	福島県	福島市	1	12587	2	37301	3	14432
R41201	佐賀県	佐賀市	1	26217	2	29569	3	19397
R42201	長崎県	長崎市	1	23218	2	30844	3	18188
R43100	熊本県	熊本市	1	26483	2	30612	3	20691
R44201	大分県	大分市	1	30256	2	31832	3	21831
R45201	宮崎県	宮崎市	1	24024	2	30620	3	20805
R46201	鹿児島県	鹿児島市	1	23222	2	32165	3	20958
R47201	沖縄県	那覇市	1	14644	2	28958	3	14550



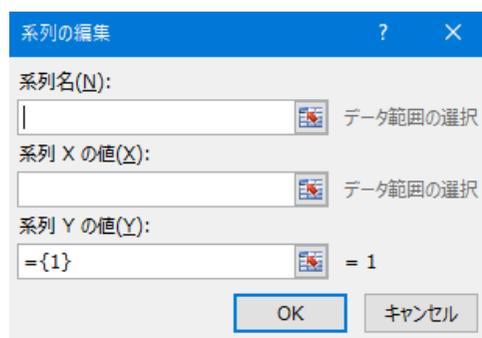
3 作成した散布図の上にカーソルを置き、右クリックして、**データの選択** を選択します。

右図の**データソースの選択** で、**系列1** を選択して **編集** をクリックし、**系列名** に「牛肉」と入力しておきます。

(前ページのデータの**B**の最初のセルを指定しても構いません。)

4 次に、2つ目の系列として、豚肉のデータを追加します。

右図の**データソースの選択** で、**追加** をクリックすると、



左図が表示されるので、

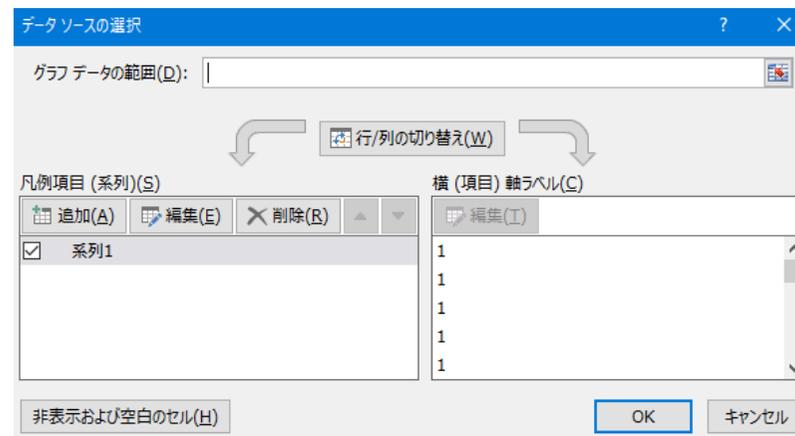
系列名 を「豚肉」、

系列Xの値 を**C** (47個のセル)、

系列Yの値 を**D** (47個のセル)、

と指定して**OK**をクリックします。

(をクリックして、セルのアドレスを指定します。)



5 同様に、3つ目の系列として、鶏肉のデータを追加します。

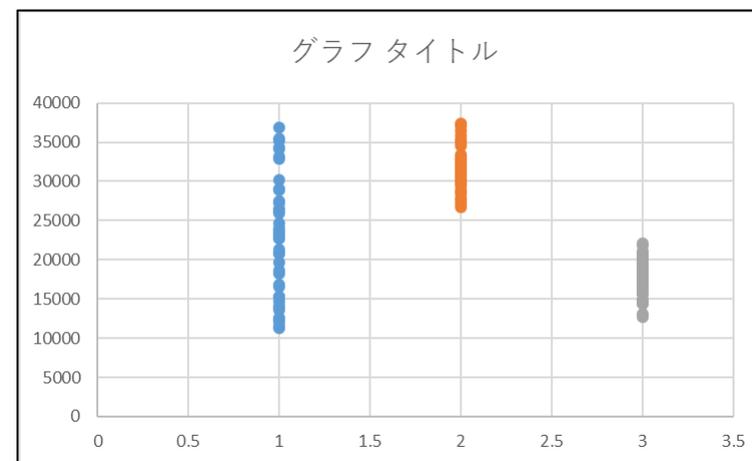
データソースの選択 で、再度 **追加** をクリックし、

系列名 を「鶏肉」、**系列Xの値** を**E**、**系列Yの値** を**F**

と指定します。

6 右下図のような散布図が出来上がります。

このグラフをもとに、各グラフ要素の書式を調整していきます。



7 各グラフ要素の書式を適宜調整します。

各要素にカーソルを置き、右クリックして表示される  の書式設定 を選択すると、それぞれの書式設定画面が表示されます。

① データ系列（各品目ごとに設定） ▶ （塗りつぶしと線） > マーカー を表示し、

- ・ マーカーのオプション＝組み込み：種類＝○、サイズは適宜
- ・ 塗りつぶし＝塗りつぶしなし
- ・ 枠線＝線（単色）：色は品目ごとに変えて指定、幅は適宜

※マーカーは、内側が塗りつぶされていないリング状の形にすると、データの集積している状況が分かりやすくなります。

② 横軸 ▶ （軸のオプション）を表示し、

- ・ 軸のオプション：最小値、最大値は適宜（この資料では、最小値＝0.5、最大値＝3.5）
- ・ 目盛：目盛の種類＝なし、補助目盛の種類＝なし
- ・ ラベル：ラベルの位置＝なし

※このグラフでは、横軸に品目名のラベルを表示できないので、品目名を入れたテキストボックスを作成し張り付けています。

③ 縦軸 ▶ （軸のオプション）を表示し、

- ・ 軸のオプション：この資料では、最小値＝0、最大値＝40000、目盛＝5000、補助目盛＝1000
- ・ 目盛：この資料では、目盛の種類＝内向き、補助目盛の種類＝内向き

④ タイトル・ラベル

- ・ この資料では、グラフの表題は別途作成したため、EXCELのグラフタイトルは削除しています。
- ・ この資料では、縦軸の単位（円）は、縦軸ラベルを追加して作成しています。

3 平均値や四分位数などを計算する ▶ 四分位数についての補足

◎ **四分位数**は、全データを大きさの順に並べて4分割する際の境界値です。

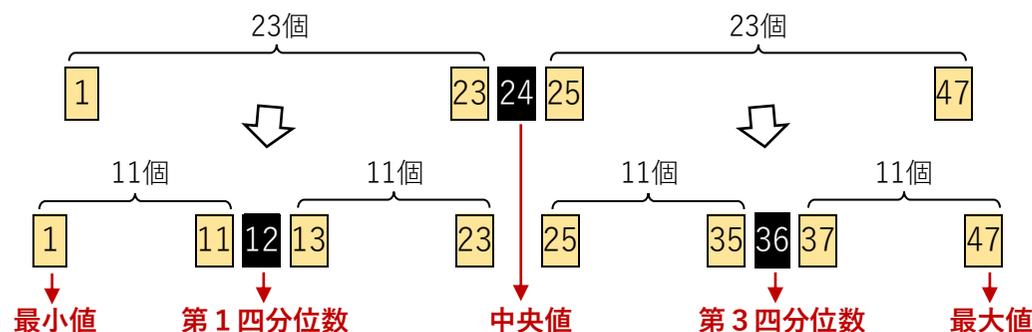
※ 四分位数の計算には幾つかの方法がありますが、このページでは、教科書に掲載されている方法について説明します。

1 全データを大きさの順に並べる ▶ 中央のデータの値が**中央値**（データ数が**偶数の場合**は、中央の2つのデータの平均値）

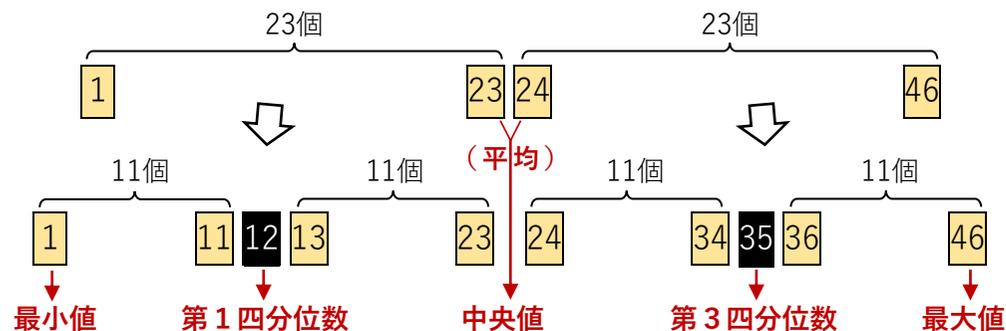
2 全データを中央で2グループに分割する（データ数が**奇数の場合**は、中央値のデータを除外して2分割）

- ▶ 値が小さいグループの**中央値**が**第1四分位数**
 - ▶ 値が大きいグループの**中央値**が**第3四分位数**
- }（グループのデータ数が**偶数の場合**は、中央の2つのデータの平均値）

【例1】この分析ではデータ数が**47**で**奇数**なので、四分位数は以下のようになります。



【例2】データ数が1つ少ない**46**の場合は**偶数**になるので、四分位数は以下のようになります。



3 平均値や四分位数などを計算する ▶ EXCELでの計算方法

◎ EXCEL の **QUARTILE.EXC 関数**は、データ数の奇数・偶数にかかわらず中央値を除外して四分位数を計算するので、データ数が奇数の場合は教科書の方法と同じですが、**データ数が偶数の場合は教科書の方法とは必ずしも同じにはなりません。**

★ **データ数の奇数・偶数にかかわらず教科書の方法と同じ結果を得るには、**少し複雑になりますが、**PERCENTILE.EXC 関数を使う方法があります。**

=PERCENTILE.EXC (データ範囲, **割合**)

データ数を **N**としたとき、**割合**は**右表**の値を使います。

(割合)	Nが奇数	Nが偶数
第1四分位数	0.25	$(0.25 \times N + 0.5) \div (N + 1)$
中央値	0.5	0.5
第3四分位数	0.75	$(0.75 \times N + 0.5) \div (N + 1)$

【参考】 データ数**偶数**の **QUARTILE.EXC** の計算方法は、以下のようなイメージになります。(データ数 **46** の例)



- Ⓐ $22.5 = (46 - 1) \div 2$
(中央のデータ 1 個分を除外して 2 分割)
- Ⓑ **中央値** : 23 番目と 24 番目の平均値
= 23 番目 \times 0.5 + 24 番目 \times 0.5
- Ⓒ $10.75 = (22.5 - 1) \div 2$
- Ⓓ **第 1 四分位数** :
= 11 番目 \times 0.25 + 12 番目 \times 0.75
- Ⓔ **第 3 四分位数** :
= 35 番目 \times 0.75 + 36 番目 \times 0.25

◎ データ数が**偶数**の場合、**QUARTILE.EXC** の計算結果は、**教科書の方法**に比べ、**第 1 四分位数は 0.25 個分だけ最小値寄りに、第 3 四分位数は 0.25 個分だけ最大値寄り**になります。(この 0.25 個分の差は、データ数の大きさにかかわらず一定です。)

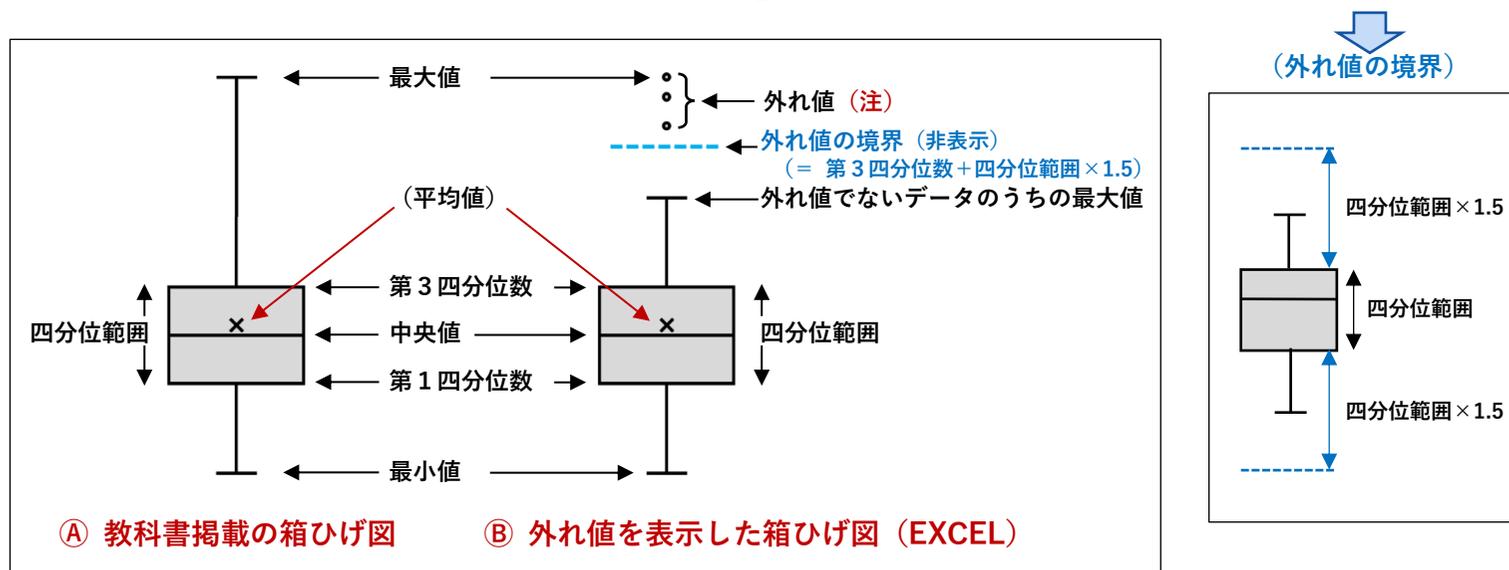
★ EXCEL を使うなら、以上のように**データ数が偶数**の場合は**教科書の方法とは一定の差があることも理解した上で、**データ数の奇数・偶数にかかわらず **QUARTILE.EXC** を**利用する方法が現実的**かもしれません。

=QUARTILE.EXC (データ範囲, **R**) **R** は、1 = 第 1 四分位数、2 = 中央値、3 = 第 3 四分位数

4 箱ひげ図を作成する ▶ 箱ひげ図についての補足

- ◎ EXCEL のグラフ機能にある箱ひげ図により作成したグラフです。
- ◎ EXCELのグラフ機能にある箱ひげ図は、教科書掲載の箱ひげ図と比較すると、以下の点で相違点があります。
 - ① EXCELの箱ひげ図で使う四分位数はQUARTILE関数の値なので、教科書の四分位数とは必ずしも一致しない場合がある。
(この資料では、データ数が奇数(47)なので、QUARTILE.EXCと教科書の四分位数は一致)
 - ② EXCELの箱ひげ図では、外れ値がある場合は「ひげ」の外側に表示される。(下図参照)

外れ値：他の値から極端に離れた値（よく使われる基準は、箱ひげ図の「箱」の上下に四分位範囲の1.5倍を超える位置にある値）



(注) EXCELのグラフでは、値の近い複数の外れ値データがある場合、○の数が1つだけになる場合がある。
すなわち、○の数が実際の外れ値のデータ数より少ない場合がある。

★ EXCELでグラフを作成する場合は、

- ① データ数が偶数の場合は、教科書の方法に比べて四分位数に若干のずれがあること
- ② 外れ値がある場合は、上図の②の外れ値の最大値、最小値を、①の最大値、最小値とみなすことに気を付けて、EXCELのグラフ機能を使って作成することが現実的と考えられます。

4 箱ひげ図を作成する ▶ EXCELでの作図方法

- 3つの品目のデータすべて（右図の太枠で囲んだ範囲）を選択した状態で、
挿入 > グラフ → グラフの挿入 を表示
すべてのグラフ > 箱ひげ図 を選択し、**OK**をクリックすると、
右下のように箱ひげ図が表示されます。

- 各グラフ要素の書式を適宜調整します。

各要素にカーソルを置き、右クリックして表示される ○○の**書式設定** を選択

① データ系列（各品目ごとに設定）

- ▶ （**系列のオプション**）を表示し、
 - ・ **要素の間隔**：数値が小さいほど箱の横幅が拡大し左右の空白が狭まる
 - ・ **特異ポイントを表示** と **平均マーカを表示** をチェック
 - ・ **四分位計算 = 排他的な中央値**
- ▶ （**塗りつぶしと線**）を表示し、
 - ・ **塗りつぶし = 塗りつぶし（単色）**：色は品目ごとに変えて指定
 - ・ **枠線 = 線（単色）**：色は品目ごとに変えて指定、幅は適宜

② 縦軸 ▶ （**軸のオプション**）を表示し、

- ・ **軸のオプション**：この資料では、最小値 = 0、最大値 = 40000
- ・ **目盛**：この資料では、目盛の種類 = 内向き、補助目盛の種類 = 内向き

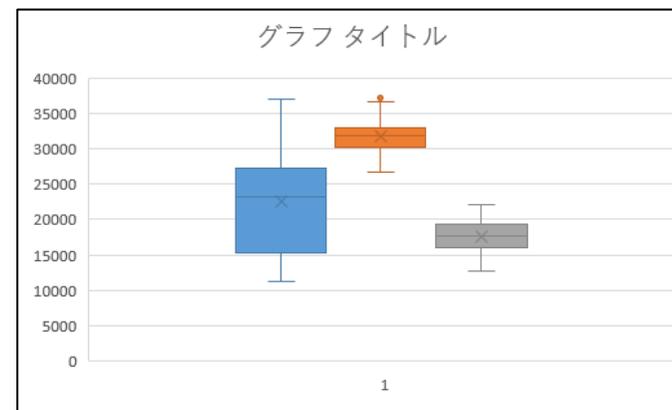
③ タイトル・ラベル

- ・ 横軸ラベル「1」は削除します。

※このグラフでは、横軸に品目名のラベルを表示できないので、品目名を入れたテキストボックスを張り付けています。

- ・ この資料では、グラフの表題は別途作成したため、グラフタイトルは削除しています。

地域コード	都道府県	市	牛肉	豚肉	鶏肉
R01100	北海道	札幌市	12463	32698	16903
R02201	青森県	青森市	15301	33418	16536
R03201	岩手県	盛岡市	11409	32621	14742
R04100	宮城県	仙台市	13876	32951	16521
R05201	秋田県	秋田市	14047	35183	16664
R06201	山形県	山形市	26151	34770	16655
R07201	福島県	福島市	12587	37301	14432
R41201	佐賀県	佐賀市	26217	29569	19397
R42201	長崎県	長崎市	23218	30844	18188
R43100	熊本県	熊本市	26483	30612	20691
R44201	大分県	大分市	30256	31832	21831
R45201	宮崎県	宮崎市	24024	30620	20805
R46201	鹿児島県	鹿児島市	23222	32165	20958
R47201	沖縄県	那覇市	14644	28958	14550

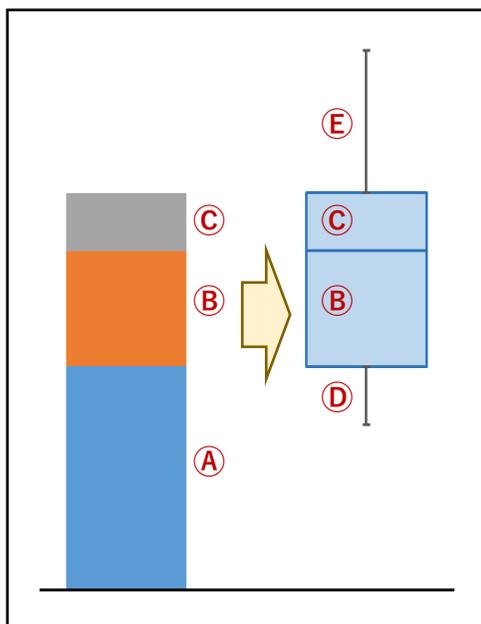


★ 教科書の箱ひげ図と全く同じ形式で作成したい場合は、EXCELの**積み上げ棒グラフ**を応用して作成できます。作成手順の概要は以下のとおりです。

- 1 中央値、四分位数、最大・最小値から**右表**を準備します。
- 2 **赤太枠範囲**を選択した状態で、**積み上げ縦棒グラフ**を作成します。
下から**①**、**②**、**③**の順に並ぶ形の積み上げ縦棒グラフが作成されます。

	牛肉	豚肉	鶏肉
① 第1四分位数	15,301	30,248	16,036
② 中央値 - 第1四分位数	7,917	1,528	1,652
③ 第3四分位数 - 中央値	3,974	1,175	1,709
④ 第1四分位数 - 最小値	4,008	3,483	3,380
⑤ 最大値 - 第3四分位数	9,745	4,393	2,746

- 3 積み上げ縦棒グラフの一番下の棒を消去し、上下に**ひげ**を付けるため、次のように「**誤差範囲**」の機能を利用して、各グラフ要素の書式を調整します。



① 系列1**①** ▶ **塗りつぶしなし、線なし**（非表示にする）

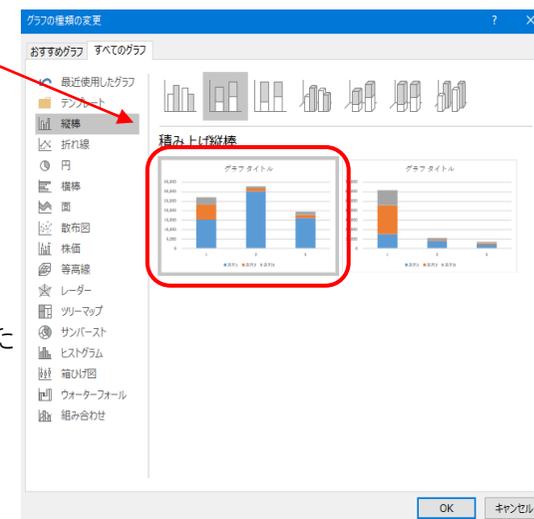
② 系列1**①**を選択した状態で、右上の「**+**」から**誤差範囲** ▶ **その他のオプション** をチェックし、表示された**誤差範囲の書式設定** ▶ （**縦軸誤差範囲**）において、

- ・ 方向 = **負方向**
- ・ 終点のスタイル = **キャップあり**
- ・ 誤差範囲 = **ユーザー設定**、値の指定で**負**の誤差の値として**④**を指定

③ 系列2**②**と系列3**③** ▶ **塗りつぶし** と **線** は品目ごとに同じ色を指定

④ ②と同様に系列3**③**を選択した状態で「**+**」から**誤差範囲** ▶ **その他のオプション** をチェックし、表示された**誤差範囲の書式設定** ▶ （**縦軸誤差範囲**）において

- ・ 方向 = **正方向**
- ・ 終点のスタイル = **キャップあり**
- ・ 誤差範囲 = **ユーザー設定**、値の指定で**正**の誤差の値として**⑤**を指定



5 ヒストグラムを作成する ▶ EXCELでの度数分布表の計算方法

◎ EXCELの度数分布計算の関数 **FREQUENCY** は、「◇より大きく ◆以下」という条件で度数が算出されますので、教科書の表記に合わせた「◇以上 ◆未満」という条件の度数とは、厳密には境界値の扱いが異なります。

(分析するデータの中に、境界値と一致するデータがある場合には、度数が変わります。)

◎ 度数分布表は「◇より大きく ◆以下」と「◇以上 ◆未満」のどちらの条件を使っても問題ないと考えられますが、ここでは、**COUNTIF**関数を使い、教科書の表記に合わせて「◇以上 ◆未満」の条件で度数分布表を作成しています。

◎ 最初に**累積度数分布表**を作成し、次に**度数分布表**を作成する、という手順になります。(例として、本編の表3(牛肉)の作成手順)

1 右の表の(A)、(B)、(D)を準備しておきます。

2 **COUNTIF**関数を用いて、(C)に、条件式(B)を満たすデータの数を
入れます。使用する数式は、**=COUNTIF(データ範囲, 条件式)**

右上の表の赤枠の範囲」 (B)

3 (C)の上下2つのセルの差分を計算し、(E)の各セルに入れます。

※ (C)が累積度数、(E)が度数になります。

◎ なお、条件式(B)を「<=10000」のように設定すると、
「◇より大きく ◆以下」の条件で度数を算出できます。

地域コード	都道府県	市	牛肉
R01100	北海道	札幌市	12463
R02201	青森県	青森市	15301
R03201	岩手県	盛岡市	11409
R04100	宮城県	仙台市	13876
R05201	秋田県	秋田市	14047
R06201	山形県	山形市	26151
R07201	福島県	福島市	12587
R41201	佐賀県	佐賀市	26217
R42201	長崎県	長崎市	23218
R43100	熊本県	熊本市	26483
R44201	大分県	大分市	30256
R45201	宮崎県	宮崎市	24024
R46201	鹿児島県	鹿児島市	23222
R47201	沖縄県	那覇市	14644

(A)	(B)	(C)	(D)	(E)
	条件式	牛肉		牛肉
10000 未満	<10000	0	7.5 以上 10.0 未満	0
12500 未満	<12500	5	10.0 以上 12.5 未満	5
15000 未満	<15000	10	12.5 以上 15.0 未満	5
17500 未満	<17500	14	15.0 以上 17.5 未満	4
20000 未満	<20000	17	17.5 以上 20.0 未満	3
22500 未満	<22500	20	20.0 以上 22.5 未満	3
25000 未満	<25000	30	22.5 以上 25.0 未満	10
27500 未満	<27500	37	25.0 以上 27.5 未満	7
30000 未満	<30000	39	27.5 以上 30.0 未満	2
32500 未満	<32500	40	30.0 以上 32.5 未満	1
35000 未満	<35000	44	32.5 以上 35.0 未満	4
37500 未満	<37500	47	35.0 以上 37.5 未満	3
40000 未満	<40000	47	37.5 以上 40.0 未満	0

(単位：円) (単位：千円)

【累積度数分布表】 (47 - 44 = 3) 【度数分布表】

5 ヒストグラムを作成する ▶ EXCELでのヒストグラムの作図方法

◎ ここでは、前ページで作成した**度数分布表**を使用して**ヒストグラム**を作成します。

- 1 度数分布表に、グラフで表示するラベルを追加しておきます（**右表**の中央の列）。
- 2 **赤太枠範囲**を選択した状態で、**縦棒グラフ**を作成します（**右下図**を参照）。
- 3 各グラフ要素の書式を適宜調整します。

① データ系列

- ▶ （**系列のオプション**）を表示し、**要素の間隔 = 0%**
- ▶ （**塗りつぶしと線**）を表示し、**塗りつぶし**と**枠線**を適宜指定

② 横軸 ▶ **文字のオプション**を表示し、

- ・  **テキストボックス**：この資料では、**文字列の方向 = 右へ90度**
(ラベルの文字数が多いため)

③ 縦軸 ▶ （**軸のオプション**）を表示し、

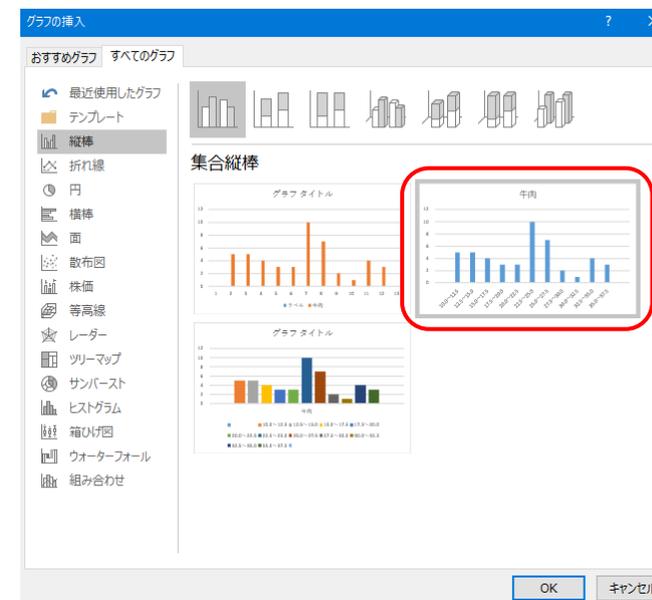
- ・ **軸のオプション**：**最小値 = 0**、最大値等は適宜
- ・ **目盛**：目盛の種類、補助目盛の種類等は適宜

④ タイトル・ラベル

- ・ この資料では、横軸の単位（千円）は横軸ラベルを追加して作成

	ラベル	牛肉
7.5 以上 10.0 未満		0
10.0 以上 12.5 未満	10.0～12.5	5
12.5 以上 15.0 未満	12.5～15.0	5
15.0 以上 17.5 未満	15.0～17.5	4
17.5 以上 20.0 未満	17.5～20.0	3
20.0 以上 22.5 未満	20.0～22.5	3
22.5 以上 25.0 未満	22.5～25.0	10
25.0 以上 27.5 未満	25.0～27.5	7
27.5 以上 30.0 未満	27.5～30.0	2
30.0 以上 32.5 未満	30.0～32.5	1
32.5 以上 35.0 未満	32.5～35.0	4
35.0 以上 37.5 未満	35.0～37.5	3
37.5 以上 40.0 未満		0

(単位：千円)





SSDSE（教育用標準データセット：Standardized Statistical Data Set for Education）は、
データ分析のための汎用素材として、独立行政法人統計センターが無償で提供しています。

<https://www.nstac.go.jp/use/literacy/SSDSE/>

- ・ 主要な統計を地域別に一覧できる表形式のデータセットです。
- ・ 6種類のSSDSEを公開しています。（令和6年12月現在）
- ・ データサイエンス演習、統計教育などにご活用ください。

SSDSE-市区町村（SSDSE-A）	全国の全市区町村の、人口、経済、教育、労働、医療、福祉など、様々な分野の統計データを収録
SSDSE-県別推移（SSDSE-B）	人口、経済、教育、労働、医療、福祉など、様々な分野の統計データを、12年分の時系列で収録
SSDSE-家計消費（SSDSE-C）	1世帯当たりの食料の年間支出金額（消費額）を、魚介、肉、野菜、果物、菓子、飲料などに分類し、それぞれ詳細な品目別にデータを収録
SSDSE-社会生活（SSDSE-D）	男女別に、スポーツ・趣味・娯楽・ボランティアなどの詳細な活動データや、1日の睡眠・食事・学業・家事・仕事・趣味・娯楽などの時間配分データを収録
SSDSE-基本素材（SSDSE-E）	人口、経済、教育、文化、医療、福祉など、様々な分野の統計データを収録し、初学者にも扱いやすいデータセットです
SSDSE-気候値（SSDSE-F）	気温、気圧、風速、日照、降水量、雪など、様々な気象データについて、月・年別の平年値を収録

SSDSE を使ってみよう — EXCEL を使った SSDSE データの使用例 —

第1回 生鮮肉の支出金額【解説編】

公開：令和6年12月3日

作成：独立行政法人 統計センター <https://www.nstac.go.jp/>

お問い合わせ先： SSDSE 担当

〒162-8668 東京都新宿区若松町19-1

統計センター 統計技術・提供部 技術研究開発課

ssdse_atmark_nstac.go.jp （「_atmark_」を「@」に置き換えてください）

