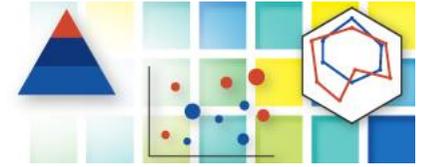




SSDSE を使ってみよう



— EXCEL を使った SSDSE データの使用例 —

第 1 回 生鮮肉の支出金額

◎ 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】	1	分析用のデータセットを作成する	-----	2
	2	データをプロットしてみる（データのばらつき）	-----	3
	3	平均値や四分位数などを計算する（平均、四分位数）	-----	4
	4	箱ひげ図を作成する（四分位数、箱ひげ図）	-----	5
	5	ヒストグラムを作成する（度数分布、ヒストグラム）	-----	6

1 分析用のデータセットを作成する

◎ SSDSE-家計消費から、今回の分析で使用するセルを残し、他の行・列をすべて削除します。

SSDSE-C-202	2021-2023年	City	LA03	LB03	LB031001	LB031002	LB031003	LB031004	LB122001
地域コード	都道府県	市	世帯人員	03 肉類	牛肉	豚肉	鶏肉	合いびき肉	学校給食
R00000	全国	全国	2.91	97296	22338	32644	17615	2994	9402
R01100	北海道	札幌市	2.78	89510	12463	32698	16903	1711	13392
R02201	青森県	青森市	2.83	91904	15301	33418	16536	1277	7373
R03201	岩手県	盛岡市	3.03	83671	11409	32621	14742	1333	11492
R04100	宮城県	仙台市	2.95	89212	13876	32951	16521	1653	10454
R05201	秋田県	秋田市	2.73	87663	14047	35183	16664	1407	12760
R06201	山形県	山形市	3.12	105439	26151	34770	16655	1617	14005
R07201	福島県	福島市	2.97	87918	12587	37301	14432	1171	14751
R08201	茨城県	水戸市	2.72	78897	13638	30248	14467	1751	9780
R41201	佐賀県	佐賀市	2.82	99473	26217	29569	19397	3959	8801
R42201	長崎県	長崎市	2.83	98043	23218	30844	18188	4704	7422
R43100	熊本県	熊本市	3.00	105801	26483	30612	20691	4854	15028
R44201	大分県	大分市	2.77	109792	30256	31832	21831	4529	11968
R45201	宮崎県	宮崎市	2.84	98237	24024	30620	20805	3556	11298
R46201	鹿児島県	鹿児島市	2.83	98602	23222	32165	20958	4084	10728
R47201	沖縄県	那覇市	2.97	81759	14644	28958	14550	2934	9624

◎ 使用するデータ項目は「LB031001 牛肉」「LB031002 豚肉」「LB031003 鶏肉」の3品目です。

◎ 網掛けの部分すべてを削除します。今回の分析では全国平均は使いませんので、3行目の全国のデータも削除します。

◎ 結果として、縦48行×横6列の表形式データになります。これが今回の分析用データセットになります。

2 データをプロットしてみる

- ◎ まず、牛肉、豚肉、鶏肉の各品目について、47地域の支出金額データを単純にプロットし、分布の様子を表示します。
- ◎ プロットした結果、データの分布は下図のようになります。

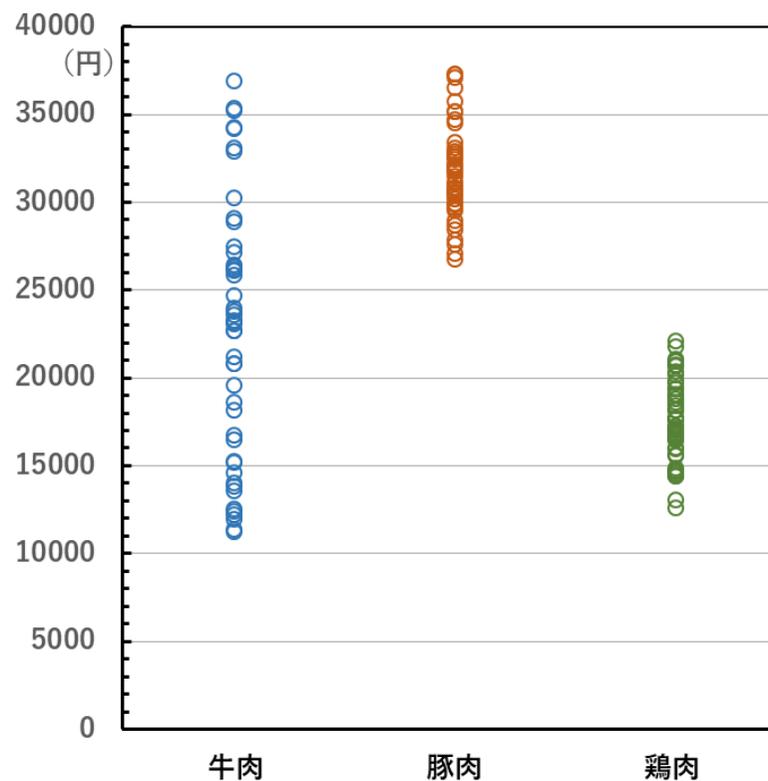


図1 生鮮肉3品目の支出金額の分布

3 平均値や四分位数などを計算する

◎ 分布を定量的にみるため、平均値や四分位数などの基本統計量を計算します。

表1 生鮮肉3品目の主な基本統計量

		牛肉	豚肉	鶏肉	使用EXCEL関数
最小値	a	11,293	26,765	12,656	MIN
最大値	b	36,937	37,344	22,143	MAX
範囲	b-a	25,644	10,579	9,487	
平均値	c	22,599	31,775	17,626	AVERAGE
第1四分位数	e	15,301	30,248	16,036	} QUARTILE.EXC
中央値	f	23,218	31,776	17,688	
第3四分位数	g	27,192	32,951	19,397	
四分位範囲	g-e	11,891	2,703	3,361	

◎ 平均値、中央値 …… 代表値 (データ全体を代表する値)

◎ 範囲、四分位範囲 …… 散布度 (データのばらつきを示す値)

※ このほかに、代表値としては最頻値が、散布度としては分散、標準偏差などがあります。

◎ MIN、MAX、AVERAGE は、各品目の47個のデータ範囲を指定して計算します。(例：=MIN(データ範囲))

◎ QUARTILE.EXC は、データ範囲と1～3の整数を指定して計算します。

=QUARTILE.EXC(データ範囲, R) ただし、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

◎ 四分位数を用いて、箱ひげ図を作成すると、次の図のようになります。

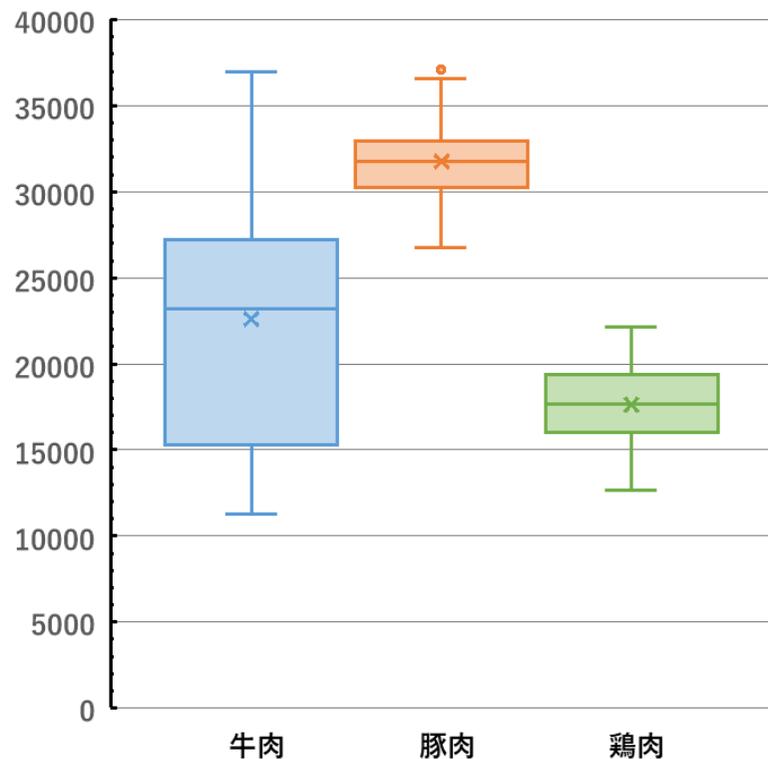


図2 生鮮肉3品目の支出金額の分布の箱ひげ図

5 ヒストグラムを作成する

◎ ヒストグラムを作成して分布の形を見ることにします。

表2 生鮮肉3品目の度数分布表

	牛肉	豚肉	鶏肉
10以上 11未満	0		
11以上 12未満	2		0
12以上 13未満	4		1
13以上 14未満	2		1
14以上 15未満	2		6
15以上 16未満	2		3
16以上 17未満	2		7
17以上 18未満	0		10
18以上 19未満	2		5
19以上 20未満	1		6
20以上 21未満	2		5
21以上 22未満	1		2
22以上 23未満	2		1
23以上 24未満	6		0
24以上 25未満	2		
25以上 26未満	1	0	
26以上 27未満	4	1	
27以上 28未満	2	3	
28以上 29未満	1	3	
29以上 30未満	1	4	
30以上 31未満	1	9	
31以上 32未満	0	7	
32以上 33未満	1	9	
33以上 34未満	1	2	
34以上 35未満	2	2	
35以上 36未満	2	3	
36以上 37未満	1	1	
37以上 38未満	0	3	
38以上 39未満		0	

(単位：千円)

◎ ヒストグラムは階級の分け方により形が変わりますが、ここでは暫定的に階級の幅を1000円に設定して**度数分布表**を作成し、この数値を用いてヒストグラムを作成してみます。

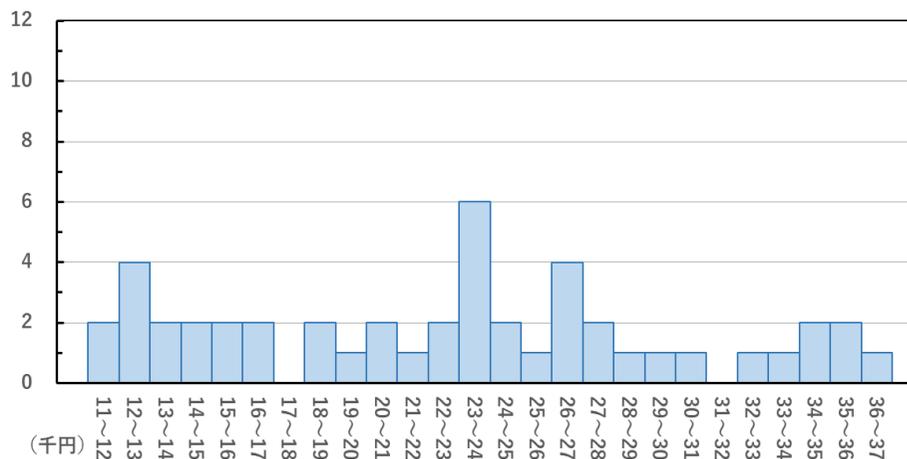


図3 牛肉の支出金額のヒストグラム

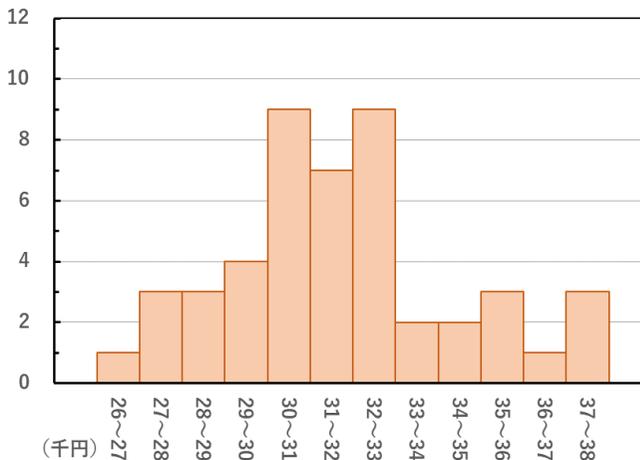
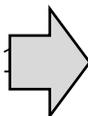


図4 豚肉の支出金額のヒストグラム

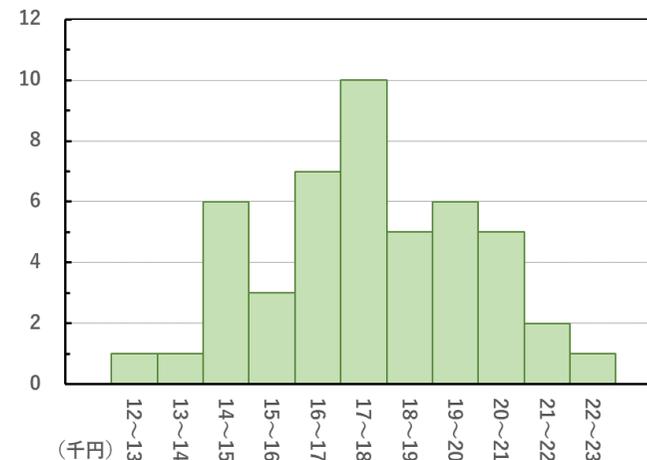


図5 鶏肉の支出金額のヒストグラム

◎ 前ページのヒストグラムのうち、豚肉と鶏肉は分布の特徴が把握できますが、牛肉は階級の数が多いため特徴が把握できません。

◎ そこで、牛肉の階級幅を2500円に設定し、
度数分布表とヒストグラムを再作成します。

表3 牛肉の度数分布表

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

(単位：千円)

◎ 比較のために、豚肉と鶏肉のヒストグラムを再掲します。

◎ これらの3つのヒストグラムは、階級（柱）の数がほぼ同数（11～12）なので、互いの分布の比較に適していると言えます。

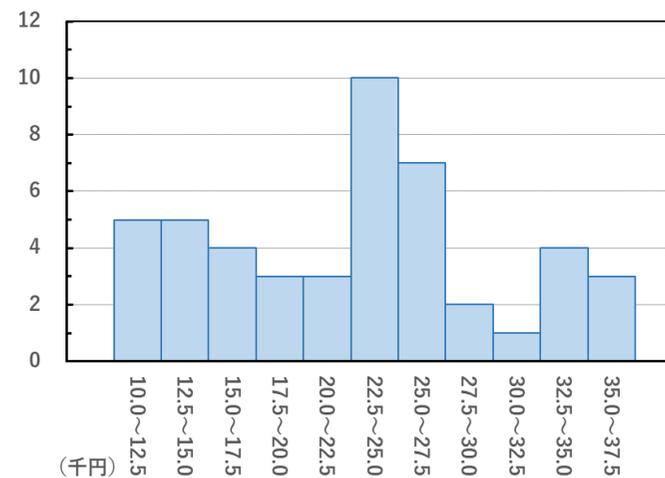
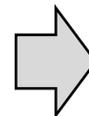


図6 牛肉の支出金額のヒストグラム

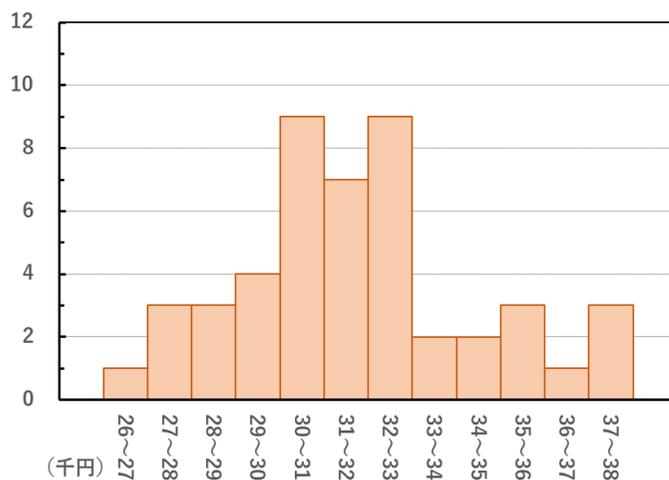


図4 豚肉の支出金額のヒストグラム

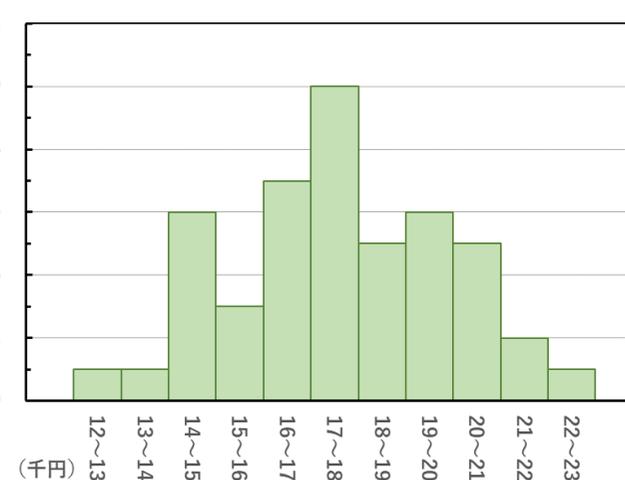


図5 鶏肉の支出金額のヒストグラム



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

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）

