SSDSE の公開と統計リテラシー向上への取組

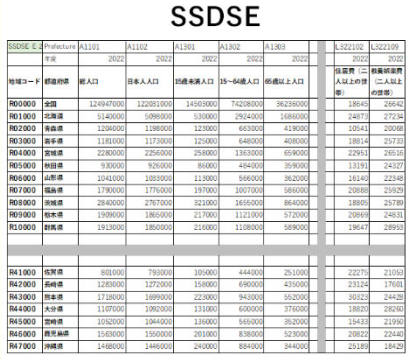
独立行政法人統計センター
深野淳一・白川和佳子

SSDSE（教育用標準データセット：Standardized Statistical Data Set for Education）

データ分析用素材として 独立行政法人統計センターが無償提供

- 政府統計を手軽に利用できるように編集・公開（欠データなし）
- 2018年に提供を開始し、**現在は6種類のSSDSEを提供中** →
- **SSDSE**は、行に地域、列にデータ項目が並ぶ表形式データ（EXCEL、CSV）
- 収録データの出典、利用上の注意などを掲載した**「SSDSEの解説」**も提供

SSDSE



SSDSEの解説

SSDSE-市区町村（SSDSE-A）
1741 市区町村×多分野 125 項目
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）
47 都道府県×12 年次×多分野 109 項目
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）
47 都道府県庁所在市×家計消費 226 項目
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）
47 都道府県×男女別×社会生活 121 項目
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）
47 都道府県×多分野 93 項目
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

SSDSE-気候値（SSDSE-F）
47 都道府県庁所在市×月・年×気象 42 項目
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 教育用標準データセット

SSDSEのページ
https://www.nstac.go.jp/use/literacy/ssdse/

SSDSE の分析事例の研究・提供

SSDSEの分析事例を研究し**利活用事例「SSDSEを使ってみよう」**を公開
- **SSDSE-家計消費**から**生鮮肉**(牛肉、豚肉、鶏肉)**の都道府県別支出金額**を分析 →
  - **第１回** 中学校の学習内容（四分位数、箱ひげ図、ヒストグラム）
  - **第２回** 高等学校の学習内容（標準偏差、相関係数、散布図）

→ 今後、第3回以降も公開

① 箱ひげ図

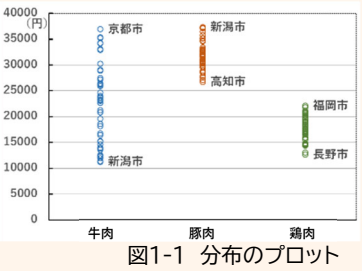


図1-1 分布のプロット

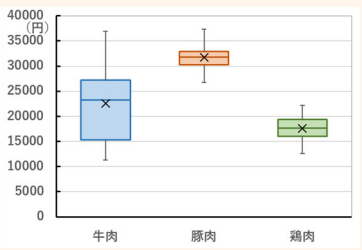


図1－2 箱ひげ図
（教科書の四分位数の計算法についても解説あり）

② ヒストグラム

階級幅を１千円に統一すると →

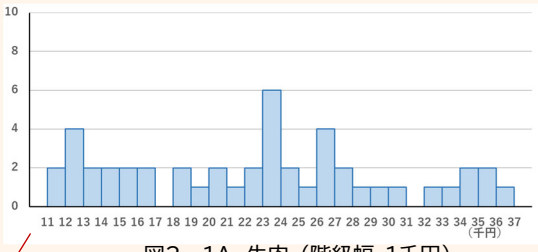


図2－1A 牛肉（階級幅 1千円）

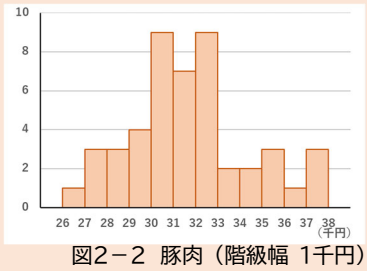


図2－2 豚肉（階級幅 1千円）

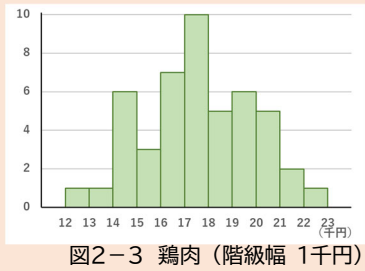


図2－3 鶏肉（階級幅 1千円）

分布の特徴が把握できない
↓
階級幅の変更 →

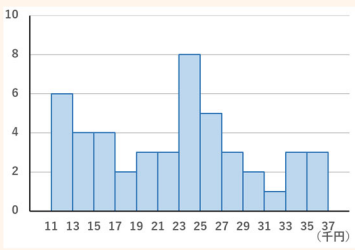


図2－1B 牛肉（階級幅 2千円）

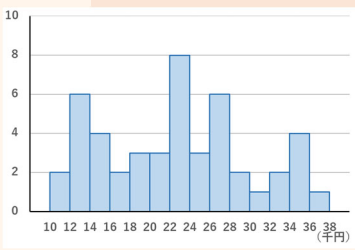


図2－1C 牛肉（階級幅 2千円）

（同じ階級幅でも形状が異なる）

階級幅の変更 →

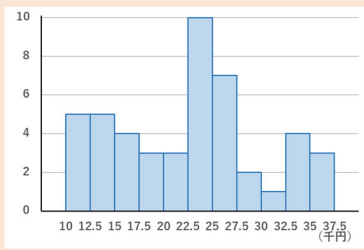


図2－1D 牛肉（**階級幅 2.5千円**）
（豚肉、鶏肉との比較に利用可）

③ 相関分析（仮説と検証）

(1) 都道府県別の支出金額

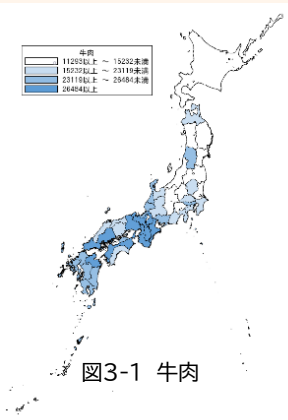


図3-1 牛肉

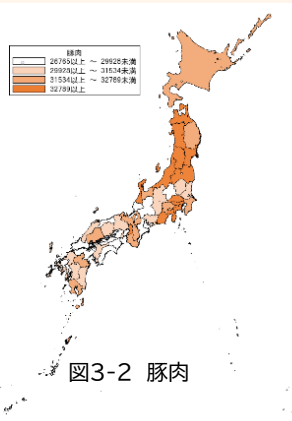


図3-2 豚肉

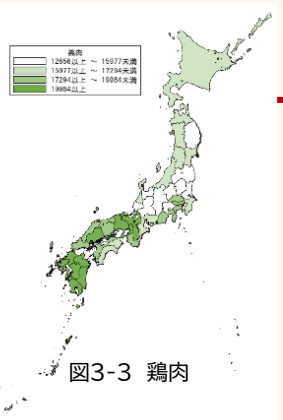
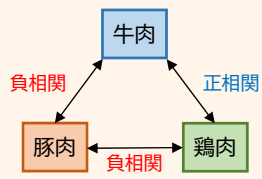
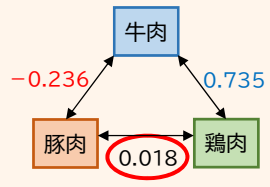


図3-3 鶏肉

(2) 仮説



(3) 相関係数の確認



(4) 豚肉と鶏肉の相関について検証

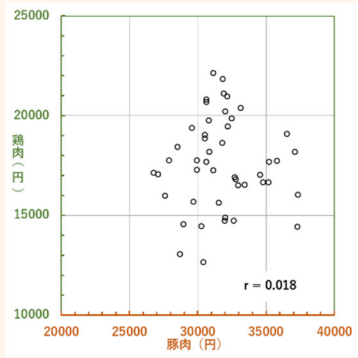


図3－4 豚肉と鶏肉の散布図

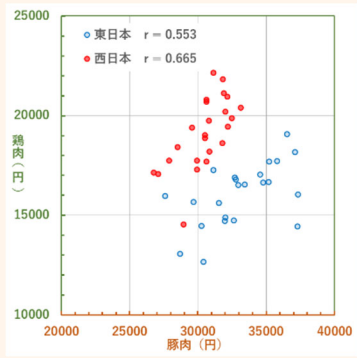


図3-5 豚肉と鶏肉の散布図（東西別）

（東日本と西日本の2地域に分割 → 地域内は正相関、地域間は負相関）

統計データ分析コンペティションの受賞論文の特徴

統計データ分析コンペティション の受賞論文について特徴等を研究
- 高校生・大学生等を対象に、SSDSEを用いた統計データ分析の論文を募集
- 分析力を活用した課題解決のアイデアを競い、優れた論文に総務大臣賞等の賞を授与

- **高校生の部**の過去3年間の**主要受賞論文12本を対象**に確認・検証

過去3年間の主要受賞論文

	2024年	2023年	2022年
総務大臣賞	食の外部化における地域特性 ◎ 雙葉高等学校 佐々木万悠子	生活の形態と女性の社会進出 愛知県立岡崎高等学校 杉山輝恵	体力が基礎学力に与える影響について フェリス女学院高等学校 太佐美結
優秀賞	福祉支援を通じた過疎化対策の提案 大分工業高等専門学校情報工学科 黒木喬士郎、井上和幸、髙山大綺、玉田章人	大腸がん罹患要因の探究と罹患しにくい生活の提案 慶應義塾大学湘南藤沢高等部 鈴木実由	ボランティア活動の決定要因 慶應義塾高等学校 今泉開
統計数理賞	子供の体力・運動能力 愛知県立一宮高等学校 大河内花音	地価に関する最適モデルの構築と手法提案 雲雀丘学園高等学校 柏原昊隼、田原睦己、大西裕貴	市区町村というミクロ的視点から投票率の実態を探る 茨城県立並木中等教育学校 林蔚欣
統計活用奨励賞	医療費削減に向けたスポーツ時間増加策のデータ分析 法政大学国際高等学校 過目今	独自指標作成による地方創生の方法論と兵庫県活性化の提案 兵庫県立姫路西高等学校 衣川凌太、中島琉士、穂積佑季、丸山晃平、盈優真	都市部と地方の教育格差の要因と課題～日本の教育現場において～ 愛知県立一宮高等学校 森下達也 ◎

◎：分析手法が高等学校の学習範囲内の論文

(1) 使用されたSSDSEの種類

SSDSE-E…7件、 **SSDSE-B**…5件、 **SSDSE-A**…4件、 **SSDSE-D**…4件、 **SSDSE-C**…2件

(2) その他の主な使用データ

総務省の基本統計のほか、**文部科学省**の学力・体力関係の統計、**厚生労働省**の健康・医療・福祉関係の統計など

(3) 使用された分析手法（高等学校の学習範囲を超えるもの）

重回帰分析…9件、 因子分析…2件、 主成分分析…1件、 クラスター分析…1件、 **（高等学校の範囲内…2件）**

(4) 散見される誤りの例

- **データの収集・加工**……複数のデータを取得した際の地域の照合ミス（データソースにより地域の並び順が異なる場合など）、 その他、データ取得や計算などの際の軽微なミス（半数以上の論文が該当）
- **グラフの作成**……ヒストグラムで階級設定が不適切なもの、 棒グラフで棒の起点が0でないもの
- **分析の考察**……相関関係と因果関係の混同、 重回帰分析で標準化していない偏回帰係数を用いて説明変数の影響度を判断するケース（多重共線性については大体留意できている）

統計データ分析コンペティション 現在、エントリー受付中です！
ホームページ： https://www.nstac.go.jp/statcompe/index.html