



独立行政法人
統計センター

SSDSEと統計データ分析コンペティション の概要

信頼に応えてつくる確かな統計

2024年3月3日

佐藤 美和

独立行政法人統計センター

第 21 回 統計・データサイエンス教育の方法論
ワークショップ

セッション IV：統計センターのSSDSE（教育用標準データセット）の利活用

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

- ・ データ分析のための汎用素材として、統計センターが提供しているデータセット
- ・ 政府統計の地域別データを手軽に利用できるよう編集して提供
- ・ 縦に地域、横にデータ項目を並べた、2次元の表形式データ

○ 2018年に提供を開始し、現在は6種類のSSDSEを無償提供

- ・ SSDSE-市区町村 (SSDSE-A) 市区町村別、多分野データ
- ・ SSDSE-県別推移 (SSDSE-B) 都道府県別、時系列、多分野データ
- ・ SSDSE-家計消費 (SSDSE-C) 都道府県庁所在市別、家計消費データ
- ・ SSDSE-社会生活 (SSDSE-D) 都道府県別、男女別、社会生活データ
- ・ SSDSE-基本素材 (SSDSE-E) 都道府県別、多分野データ
- ・ SSDSE-気候値 (SSDSE-F) 都道府県庁所在市別、月・年別、気象データ **NEW**

⇒ 毎年、最新データに更新

⇒ 今後も内容を充実 (新たな SSDSE の作成・提供など)

○ SSDSE を用いた統計データ分析コンペティションを毎年度開催

SSDSEの4つの特徴

簡便性

- ・ 無償で容易に使うことができる
 - ⇒ 一括ダウンロードし必要なデータを取捨選択
- ・ 特別な前処理が不要
 - ⇒ 全セルに数値が入っている
 - ⇒ 欠測値がない「完備データ」

親近性

- ・ 自分に関する地域のデータが含まれる
 - ⇒ 誰でもなじみがある「自分たちのデータ」

具体性

- ・ 個々のデータの意味が分かり議論できる
 - ⇒ 政府統計から抽出・編集した「実データ」

多様性

- ・ 幅広い分野のデータが含まれている
 - ⇒ 様々な課題抽出、多様な分析が可能
 - ⇒ 自由度の高い「標準データ」

1. SSDSEの概要

SSDSE

・ EXCEL と CSV で 提供



統計を活かす

- 公的統計の二次的利用サービス
- 調査票情報のオンサイト利用
- データの利用
- モード集計の利用
- 機関等との連携
- 情報
- マイクロデータの利用
- データベース利用
- データ向上のために
- (教育用標準データセット)
- データの利活用事例
- 公開したSSDSE
- データ分析コンペティション

SSDSE-市区町村 (SSDSE-A)

様々な分野の市区町村別データを集めたデータセットです。

(1741市区町村×多分野125項目)

(出典)総務省統計局「統計でみる市区町村のすがた(社会・人口統計体系)2023」

	SSDSE (EXCEL)	SSDSE (CSV)	SSDSEの解説
2023年版 2023年6月29日公開	SSDSE-A-2023 (EXCEL:1,283KB)	SSDSE-A-2023 (CSV:920KB)	SSDSE-市区町村の解説 (PDF:1,740KB)

SSDSE-県別推移 (SSDSE-B)

様々な分野の都道府県別・時系列データを集めたデータセットです。

(47都道府県×12年次×多分野109項目)

(出典)総務省統計局「統計でみる都道府県のすがた(社会・人口統計体系)2023」

	SSDSE (EXCEL)	SSDSE (CSV)	SSDSEの解説
2023年版 2023年4月26日公開	SSDSE-B-2023 (EXCEL:424KB)	SSDSE-B-2023 (CSV:353KB)	SSDSE-県別推移の解説 (PDF:1,665KB)

SSDSE (教育用標準データセット)



2023年10月12日 気象データの平均値を集めたSSDSE-気候値(SSDSE-F)を新規公開しましたので、ご活用ください。 **NEW**

→ [最新版のSSDSE](#)

→ [参考資料](#)

→ [過去に公開したSSDSE \(別ページへ\)](#)

→ [SSDSEの利活用事例 \(別ページへ\)](#)

SSDSEとは

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作成・公開している統計データです。主要な公的統計を地域別に一覧できる表形式のデータセットで、直ちにデータ分析に利用することができます。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 の解説



・ 利用する際の留意点などを記載

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

1. SSDSEの概要

(例) SSDSE-県別推移の解説



- データの出典、単位、年次や、利用する際の留意点など、分析する上で必要な情報を掲載

SSDSE-県別推移の解説

SSDSE-B-2023

- SSDSE (教育用標準データセット: Standardized Statistical Data Set for Education) データサイエンス教育のための汎用素材として公開している統計データです。
- 主要な公的統計の地域別データを表形式に編集したもので、欠測データがないので、ダウンロード直ちにデータ分析に使用できます。
- SSDSE-県別推移 (SSDSE-B) は、様々な分野の都道府県別・時系列データを集めたデータセットです。
(出典) 総務省統計局「統計でみる都道府県・市区町村のすがた (社会・人口統計体系)」

SSDSEを授業や講義でお使いになる先生方へ

- SSDSEを配布する際には、この「解説」も一緒に配布していただき、データの出典や単位などについても意識してお使いいただけますよう、お願いいたします。

データのレイアウト

縦 (行の数) 項目情報 (2) + 都道府県 (47) × 年次 (12) = 566 行
 横 (列の数) 年次情報 (1) + 地域情報 (2) + データ項目 (109) = 112 列
 ※ 前年に比べ、データ項目が2項目増加しました (3ページ参照)

SSDSEのID情報 (SSDSE-B-2023)

項目情報	項目コード	項目名	地域情報		データ項目 (109)						
			SSDSE-B-2 Code	Prefecture	A1101	A110101	A110102	A1102			L322109
		年度	地域コード	都道府県	総人口	総人口 (男)	総人口 (女)	日本人人口			教養類消費 (二人以上の世帯)
12年次		2020	R01000	北海道	5224614	2465088	2759526	5151366			26539
		2019	R01000	北海道	5259000	2480000	2780000	5223000			29335

別表 SSDSE-県別推移のデータ一覧

◎項目コードをクリックすると「社会・人口統計体系」の項目定義のサイトが表示されます。
 ・本SSDSEに収録されていない関連データの入手には、出典の名称を参考に検索してください。

項目コード	項目名	単位	出典	e-Stat「地域」で入手できる年次の範囲
A 人口・世帯				
A1101	総人口	人	・国勢調査実施年(5年毎:西暦下1桁0又は5)	1975~2021
A110101	総人口 (男)	人	総務省統計局	
A110102	総人口 (女)	人	「国勢調査」人口等基本集計	
A1102	日本人人口	人	・国勢調査実施年以外	
A110201	日本人人口 (男)	人	総務省統計局	
A110202	日本人人口 (女)	人	「人口推計」	
A1301	15歳未満人口	人		
A130101	15歳未満人口 (男)	人		
A130102	15歳未満人口 (女)	人		
A1302	15~64歳人口	人		
A130201	15~64歳人口 (男)	人		
A130202	15~64歳人口 (女)	人		
A1303	65歳以上人口	人		
A130301	65歳以上人口 (男)	人		
A130302	65歳以上人口 (女)	人		
A4101	出生数	人	厚生労働省政策統括官 (統計・情報政策担当)	1980~2020
A410101	出生数 (男)	人	「人口動態統計」	
A410102	出生数 (女)	人		
A4103	合計特殊出生率	-	※ 合計特殊出生率の2020年度のデータは、社会・人口統計体系には未掲載ですが、人口動態統計から収集・補足しています。	1980~2019 ※
A4200	死亡数	人		1975~2020
A420001	死亡数 (男)	人		
A420002	死亡数 (女)	人		
A5101	転入者数 (日本人移動者)	人	総務省統計局	1975~2021
A510101	転入者数 (日本人移動者) (男)	人	「住民基本台帳人口移動報告」	
A510102	転入者数 (日本人移動者) (女)	人		
A5102	転出者数 (日本人移動者)	人		
A510201	転出者数 (日本人移動者) (男)	人		
A510202	転出者数 (日本人移動者) (女)	人		
A9101	婚姻件数	組	厚生労働省政策統括官 (統計・情報政策担当)	1975~2020
A9201	離婚件数	組	「人口動態統計」	
B 自然環境				
B4101	年平均気温	℃	気象庁観測部	1975~2021
B4102	最高気温 (日最高気温の月平均の最高値)	℃	「過去の気象データ」	
B4103	最低気温 (日最低気温の月平均の最低値)	℃	数値は、各都道府県庁所在地の観測値。 ※ 観測所が異なる場合があります。	

1. SSDSEの概要

SSDSE-県別推移：都道府県別、時系列、多分野データ



SSDSE-B-2	Code	Prefecture	A1101	A110101	A110102	A1102	A110201	A110202	A1301	A130101	A130102
年度	地域コード	都道府県	総人口	総人口 (男)	総人口 (女)	日本人人口	日本人人口 (男)	日本人人口 (女)	15歳未満人口	15歳未満人口 (男)	15歳未満人口 (女)
2020	R01000	北海道	5224614	2465088	2759526	5151366	2429697	2721669	555804	284510	271294
2019	R01000	北海道	5259000	2480000	2780000	5223000	2464000	2759000	565000	289000	276000
2018	R01000	北海道	5293000	2495000	2798000	5262000	2482000	2780000	577000	295000	282000
2017	R01000	北海道	5325000	2510000	2815000	5298000	2499000	2799000	588000	301000	288000
2016	R01000	北海道	5355000	2523000	2831000	5330000	2514000	2817000	600000	306000	294000
2015	R01000	北海道	5381733	2537089	2844644	5348768	2522980	2825788	608296	310387	297909
2014	R01000	北海道	5410000	2551000	2859000	5390000	2543000	2847000	621000	316000	304000
2013	R01000	北海道	5438000	2565000	2873000	5419000	2558000	2861000	630000	321000	309000
2012	R01000	北海道	5465000	2580000	2886000	5446000	2572000	2874000	640000	326000	314000
2011	R01000	北海道	5488000	2593000	2896000	5470000	2585000	2884000	650000	331000	318000
2010	R01000	北海道	5506419	2603345	2903074	5482650	2593193	2889457	657312	335353	321959
2009	R01000	北海道	5524000	2612000	2911000	5506000	2605000	2901000	663000	338000	325000
2020	R02000	青森県	1237984	583402	654582	1224334	577003	647331	129112	65898	63214
2019	R02000	青森県	1253000	590000	663000	1247000	588000	659000	133000	68000	65000
2018	R02000	青森県	1268000	597000	671000	1263000	595000	668000	137000	70000	67000
2017	R02000	青森県	1282000	603000	679000	1277000	601000	676000	141000	72000	69000
2016	R02000	青森県	1295000	609000	686000	1291000	607000	684000	145000	74000	71000
2015	R02000	青森県	1308265	614694	693571	1302132	612113	690019	148208	75661	72547
2014	R02000	青森県	1323000	622000	701000	1320000	620000	699000	155000	79000	76000
2013	R02000	青森県	1337000	628000	708000	1334000	627000	706000	159000	81000	78000
2012	R02000	青森県	1350000	635000	716000	1347000	634000	714000	164000	83000	80000
2011	R02000	青森県	1363000	641000	722000	1360000	640000	720000	168000	86000	83000
2010	R02000	青森県	1373339	646141	727198	1367057	643407	723650	171842	87585	84257
2009	R02000	青森県	1383000	651000	732000	1380000	650000	730000	173000	88000	85000
2020	R03000	岩手県	1210534	582952	627582	1194745	575610	619135	132447	67763	64684
2019	R03000	岩手県	1226000	590000	636000	1218000	587000	631000	137000	70000	67000
2018	R03000	岩手県	1240000	597000	643000	1234000	594000	639000	140000	72000	68000

1. SSDSEの概要

◎ 現行の SSDSE の比較表

名称	SSDSE-市区町村	SSDSE-県別推移	SSDSE-家計消費	SSDSE-社会生活	SSDSE-基本素材	SSDSE-気候値
内容	市区町村別、多分野データ	都道府県別、時系列、多分野データ	都道府県庁所在市別、家計消費データ	都道府県別、男女別、社会生活データ	都道府県別、他分野データ	都道府県庁所在市別、月・年別、気象データ
地域数	1741市区町村	47都道府県	全国 + 47県庁所在市	全国 + 47都道府県	全国 + 47都道府県	47県庁所在市
元になる統計	総務省統計局「統計でみる都道府県・市区町村のすがた（社会・人口統計体系）」		総務省「家計調査」	総務省「社会生活基本調査」	総務省「統計でみる都道府県のすがた（社会・人口統計体系）」	気象庁「地上気象観測統計」
データ項目数	125	109	226	121	90	42
データの年次	各項目、最新の1年次	2009～2020年（12年次）	2020～2022年の平均値	2021年	各項目、最新の1年次	1991～2020年の平均値
提供開始年	2018年	2019年	2020年	2021年	2022年	2023年

※ **毎年、最新データに更新**

今後も内容の充実を図る（新たな SSDSE の作成・提供など）

2. 統計データ分析コンペティションの概要

- 我が国の次代を担う高校生、大学生等の統計データの利活用マインドと分析技術のかん養を図るとともに、公的統計の利活用を推進するため、2018年度に初めて開催し、2023年度に6回目を実施。
- 地域別の統計をまとめたSSDSE（教育用標準データセット：Standardized Statistical Data Set for Education）を用いて、統計データ分析力を活用した課題解決のアイデアを競うコンペティションであり、成績優秀者には、総務大臣賞等の表彰を行う。

2. 統計データ分析コンペティションの概要

共催：総務省統計局・独立行政法人 統計センター・

大学共同利用機関法人 情報・システム研究機構統計数理研究所・

一般財団法人 日本統計協会

後援：文部科学省・国立研究開発法人 科学技術振興機構（JST）・

全国高等学校長協会・一般社団法人 日本統計学会・

全国統計教育研究協議会

募集部門：高校生の部…高校、高専（1～3年次）の生徒

大学生・一般の部…短大、高専（4、5年次、専攻科）、

大学、大学院の学生、一般

2. 統計データ分析コンペティションの概要

2023年の主な日程

- 5月10日 エントリー及び論文募集開始
- 8月9日 エントリー締切
- 8月31日 論文提出締切 大学生・一般の部
- 9月8日 論文提出締切 高校生の部
- 10月18日 (統計の日) 受賞論文の発表
- 12月6日 (全国統計大会) 受賞者表彰式



総務省統計局 | 独立行政法人 統計センター | 大学共同利用機関法人 情報・システム研究機構 統計数理研究所 | 一般財団法人 日本統計協会
 The Institute of Statistical Mathematics | Japan Statistical Association

統計データ分析 コンペティション 2023

高校生、大学生等の皆さんの統計分析における
アイデアと技術を競います

論文募集

エントリー期間
令和5年5月10日(水)
～8月9日(水)

論文締切
[大学生・一般の部]
令和5年8月31日(木)
[高校生の部]
令和5年9月8日(金)

最優秀作品には
総務大臣賞と副賞
が贈られます
このほか、受賞論文が
専門誌等に掲載されます

第6回統計データ分析コンペティション
詳細は、下記ウェブサイトをご覧ください。
<https://www.nstac.go.jp/use/literacy/statcompe/>



共催 総務省統計局・独立行政法人 統計センター
大学共同利用機関法人 情報・システム研究機構 統計数理研究所・一般財団法人 日本統計協会

後援 文部科学省・国立研究開発法人 科学技術振興機構 (JST)
全国高等学校長協会・一般社団法人 日本統計学会・全国統計教育研究協議会

2. 統計データ分析コンペティションの概要

報道資料



MIC Ministry of Internal Affairs and Communications

令和5年3月22日

「統計データ分析コンペティション2023」を開催します！

総務省は、高校生、大学生等の統計の有用性への理解と統計データの利活用拡大のため、独立行政法人統計センター等との共催により、「統計データ分析コンペティション2023」を開催します。

○ コンペティションの概要

高校生、大学生等を対象に、地域別の統計をまとめた「教育用標準データセット (SSDSE)」を用いた統計データ分析の論文を募集し、そのアイデアと解析力を競うコンペティションです。平成30年度から開催しており、令和5年度は6回目となります。

(注) SSDSE は、データサイエンス教育のための汎用素材として、作成・公開しているデータセットです。

○ 応募資格

【高校生の部】

高等学校、中等教育学校（後期課程）及び高等専門学校（1～3年次）の生徒又は学生

【大学生・一般の部】

短期大学、高等専門学校（4、5年次、専攻科）、大学及び大学院の学生並びに一般

（統計分析等の学習を目的とする方）

○ 募集期間

エントリー期間 令和5年5月10日（水）～8月9日（水）

論文の提出期限 8月31日（木）（大学生・一般の部）

9月8日（金）（高校生の部）

○ 表彰

審査の結果、優秀な論文には表彰状及び副賞を授与します。

(別添資料)

別紙 「統計データ分析コンペティション2023」の開催

(連絡先)

【報道発表の内容について】

統計局統計情報利用推進課

担当：和田課長補佐 山田主査 石田主査

電話：03-5273-1023（直通）

E-mail: y-senryaku_atmark_soumu.go.jp

【応募方法と SSDSE 等について】

(独) 統計センター技術開発課

担当：田中統括統計職 仲西統計専門職 堀川統計専門職

電話：03-5273-1368（直通）

E-mail: statcompe_atmark_nstac.go.jp

※スパムメール防止のため、「@」を「_atmark_」と表示しています。
メールを送る際に「_atmark_」を「@」に直して入力してください。



「統計データ分析コンペティション2023」の開催

【別紙】

高校生、大学生等が統計の有用性を理解し、統計データを適切に活用できるよう、統計データの分析に関するコンペティションを令和5年度も開催します。奮って御応募ください！！

▶ 概要

教育用標準データセット (SSDSE) ※を用いた統計分析の論文を募集し、アイデアと解析力を競うことで、高校生、大学生等の統計リテラシーの向上を図ります。

▶ エントリー期間

令和5年5月10日（水）から8月9日（水）まで

▶ 論文締切

[大学生・一般の部] 令和5年8月31日（木）

[高校生の部] 令和5年9月8日（金）

▶ 表彰

優秀な論文には、総務大臣賞、優秀賞、統計数理賞、統計活用奨励賞と各副賞が授与されます。また、受賞論文が統計専門誌等に掲載されます。

▶ 共催

- ・総務省統計局
- ・独立行政法人 統計センター
- ・大学共同利用機関法人 情報・システム研究機構 統計数理研究所
- ・一般財団法人 日本統計協会

※教育用標準データセット (SSDSE) とは

SSDSE: Standardized Statistical Data Set for Education

データサイエンス教育のための汎用素材として作成・公開しているデータで、様々な分野（人口・世帯、経済基盤、教育、健康・医療、福祉・社会保障など）の公的統計を、市区町村別又は都道府県別にまとめた、表形式のデータセットです。

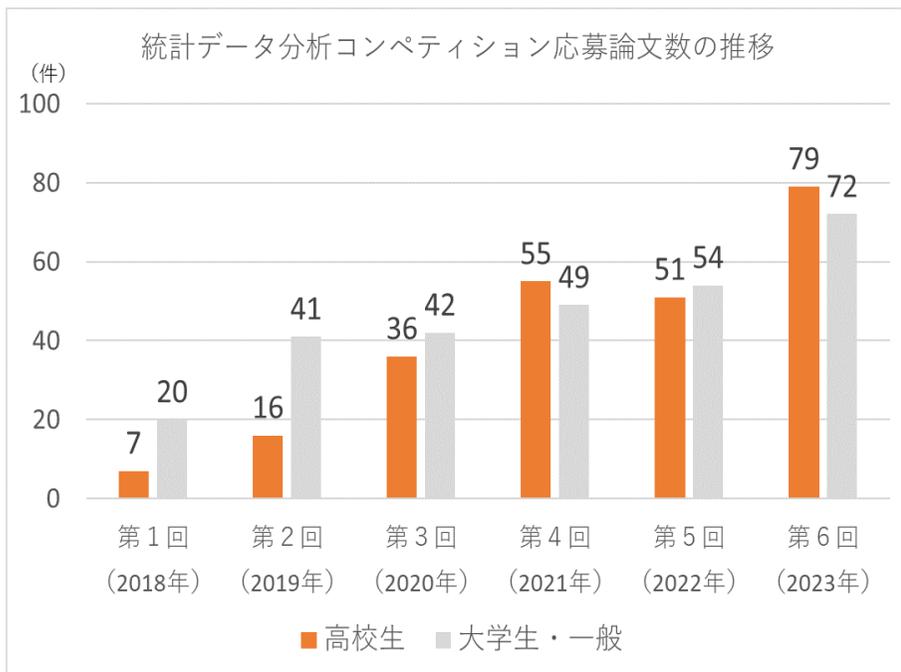
(データセットは、A.市区町村別データ、B.都道府県別時系列データ、C.家計消費データ、D.自由時間活動・生活時間データ、E.都道府県別小サイズデータの5種類を提供中)

詳細については、当コンペティションのウェブサイトを御覧ください。

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

3. 統計データ分析コンペティションの応募状況

応募論文数は増加傾向



応募者還元策としての個別講評の試行

応募者からの要望に応え、 高校生の全応募論文に個別講評を実施

論文一つひとつの内容に即した講評を
高校生の全応募者（79件）に対して
作成し、個別に還元

3. 統計データ分析コンペティションの応募状況

受賞論文の傾向

これまで（2018～2023）の受賞論文中で用いられた分析手法について集計

高校生の部：

57本中45本が相関分析（他の分析手法との併用含む）
回帰分析等高度な分析はそれほど多くない

大学生・一般の部：

63本中で用いられた分析手法は多岐に渡る
相関分析、回帰分析、主成分分析、クラスター分析、
パネルデータ分析、因子分析、ベイズモデリング…等

4. 受賞論文の概要等

受賞者及び受賞論文発表（10月18日）

「統計データ分析コンペティション2023」受賞者及び受賞論文 ～ 高校生の部 ～

受賞者	受賞論文(タイトル及び概要)
<p>【総務大臣賞】</p> <p>杉山 輝恵 (愛知県立岡崎高等学校)</p>	<p>生活の形態と女性の社会進出</p> <p>女性の社会進出に影響を与える要因を勤続年数、正規雇用率、管理職有職率の3つの目的変数それぞれから分析し、雇用環境や住居環境ではなく、育児支援や介護支援などの公的生活要因に属する社会制度が大きく関わっていることを示した。</p>
<p>【優秀賞】</p> <p>鈴木 実由 (慶應義塾湘南藤沢高等部)</p>	<p>大腸がん罹患要因の探究と罹患しにくい生活の提案</p> <p>男女の総罹患数が最も多い大腸がんに着目し、生活や環境に関するデータを用いて相関分析や重回帰分析を行うことで、どのような要素が大腸がん罹患に起因しているかを明らかにし、その結果を基に罹患しにくい生活を提案した。</p>
<p>【統計数理賞】</p> <p>柏原 昊隼、田原 睦己、大西 裕貴 (雲雀丘学園高等学校)</p>	<p>地価に関する最適モデルの構築と手法提案</p> <p>都市部への人口集中や地方の過疎化の問題に着目し、地価データを用いて因子分析及び重回帰分析を行い、地価の高い地域に共通する因子の特徴を推定することで、地方の人口増加に貢献する要因を見出した。</p>
<p>【統計活用奨励賞】</p> <p>衣川 凌太、中島 琉士、穂積 佑季、 丸山 晃平、盈 優真 (兵庫県立姫路西高等学校)</p>	<p>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</p> <p>自身が生活する兵庫県を題材に、地方創生・活性化を目指し、様々なデータを用いて新たな指標を作成しクラスター分析や主成分分析を行うことで、地域の課題を発見し、兵庫県活性化の施策提言を行った。</p>

4. 受賞論文の概要等

「統計データ分析コンペティション2023」受賞者及び受賞論文 ～ 大学生・一般の部 ～

受賞者	受賞論文(タイトル及び概要)
<p>【総務大臣賞】 廣野 准貴、藤井 優菜、 山下 航、吉本 正崇 (南山大学総合政策学部総合政策学科)</p>	<p>小中学生の不登校率における環境要因分析</p> <p>小中学生の不登校の要因について家庭環境、学校環境、地域環境から分析を行い、小中学生ともに家計支出に占める教育費割合と高校進学率が低いほど不登校率の上昇に影響があることや、中学生では転校が影響を与えることを示した。</p>
<p>【優秀賞】 菊地原 守 (名古屋大学大学院教育発達科学研究科)</p>	<p>市町村費負担教員任用の規定要因 —ハードルモデルを用いた多変量解析から—</p> <p>市町村費負担教員制度に着目し、市町村別のデータを用いて回帰分析や検定を行うことで、自治体が市町村費負担教員の任用を規定する要因について明らかにするとともに、これからの教育政策のあり方を示唆した。</p>
<p>【統計数理賞】 宮部 美月、戴 士淵 (日本電子専門学校AIシステム科)</p>	<p>CO₂排出特性と地域特性の関係 —2050年カーボンニュートラルの実現に向けて—</p> <p>カーボンニュートラル社会に向けた取組が進む中、地域特性ごとのCO₂排出量について、市区町村のCO₂排出特性に基づいた市区町村の類型化を行い様々な分析を行うことで、地域特性との相互関係やCO₂排出特性の規定要因を明らかにした。</p>
<p>【統計活用奨励賞】 井手 健太 (法政大学大学院経済学研究科/ 東京都総務局統計部調整課) 山口 真菜 (東京都総務局統計部調整課)</p>	<p>行動制限下における家計消費の変化に伴う経済波及効果の算出</p> <p>新型コロナウイルス感染症に伴う行動制限下で人々の消費にどのような変化があったか、またどのような経済波及効果があったか家計調査のデータを用いて分析を行い、家計消費支出や生産誘発額が大きく減少したことを示した。</p>

- **統計データ分析コンペティション**

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

- **SSDSE（教育用標準データセット）**

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

- **総務省統計局報道資料**

<https://www.stat.go.jp/info/guide/public/houdou/pdf/ho231018.pdf>