

第 4 講

目次

2. 課題解決のためのデータ分析	31
(1) 教育用標準データセット (SSDSE) の利用	31
(2) SSDSE を用いた課題解決事例	33

2. 課題解決のためのデータ分析

本章の Point

SSDSE に関する理解を深め、利用方法、利用事例を学びます。また、実際に SSDSE を用いた課題解決事例について理解を深めます。また、SSDSE を用いた統計データ分析コンペティションの事例や実際に統計データ分析コンペティションに出場する場合の指導スケジュールについて一例を提案し、指導の参考としてご活用いただけます。

(1) 教育用標準データセット (SSDSE) の利用

本項における 参考情報

- ◇ SSDSE (教育用標準データセット)
 - <https://www.nstac.go.jp/SSDSE/>

① 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・公開している統計データで、公的統計の主要な 111 項目のデータを全国 1741 市区町村別に一覧できるデータセットです。



出典 : 2019 年版 SSDSE (教育用標準データセット) の解説

<https://www.nstac.go.jp/SSDSE/data/kaisetsu-2019.pdf> (2020 年 3 月 18 日アクセス確認) より整理

データ名	集計単位	収録地域	収録項目	収録年次
SSDSE-2019A	市区町村	<ul style="list-style-type: none"> ● 全国の市区町村、1741 地域（2018 年 3 月 31 日時点） ● 市：791、町：744、村：183、東京都特別区：23（政令指定都市は、区に分割せず、市全体のデータのみ収録、東京都特別区は、23 区別のデータのみ収録、区部計のデータは含まない ● 1741 地域を単純合計すると全国計となる（実際は一致しない場合もある） 	<ul style="list-style-type: none"> ● 総務省統計局「社会・人口統計体系」の市区町村別データのうち、125 項目 ● 各項目の最新年次のデータを収録 ● 項目により単位や収録年次が異なる場合がある 	<ul style="list-style-type: none"> ● 1 時点
SSDSE-2019B	都道府県	<ul style="list-style-type: none"> ● 47 都道府県 	<ul style="list-style-type: none"> ● 総務省統計局「社会・人口統計体系」の都道府県別データのうち、107 項目 ● 項目により単位が異なる場合がある 	<ul style="list-style-type: none"> ● 全項目、2005 年～2016 年の 12 年分の年次データ ● 2017 年のデータがある項目も 2016 年までを収録。項目によっては、途中で定義等が変わっている場合があり、年次間の比較に注意が必要。

② SSDSE 利用の仕方

SSDSE を利用した学校向け教材として、日本統計協会の月刊誌「統計」¹に連載された「授業に使える SSDSE の統計教材（中学・高校編）」があります。箱ひげ図、外れ値、相関分析の落とし穴、インターネットを用いた情報収集と地域特性の分類等のトピックを扱う教材となっています。

教材のタイトル一覧

「SSDSE（教育用標準データセット）の利活用事例 SSDSE を利用した学校向け教材」

<https://www.nstac.go.jp/SSDSE/utility.html>

日本統計協会月刊誌「統計」

https://www.jstat.or.jp/monthly_toukei/

¹ 日本統計協会月刊誌「統計」のページはこちらになります。（https://www.jstat.or.jp/monthly_toukei/）

(2) SSDSE を用いた課題解決事例

本項における 参考情報

- ◇ 統計局 HP 統計 Today No.131
 - <https://www.stat.go.jp/info/today/131.html>
- ◇ Tableau Desktop (分析ツール)
- ◇ SSDSE (教育用標準データセット)
 - <https://www.nstac.go.jp/SSDSE/>
- ◇ 統計データ分析コンペティション
 - <https://www.nstac.go.jp/statcompe/>
- ◇ 日本統計協会／月刊誌「統計」
 - https://www.jstat.or.jp/monthly_toukei/

① 地域分析事例 統計 Today No.131 より

地域分析事例として、「統計 Today No.131」(平成 30 年 7 月 19 日)の記事を紹介します。

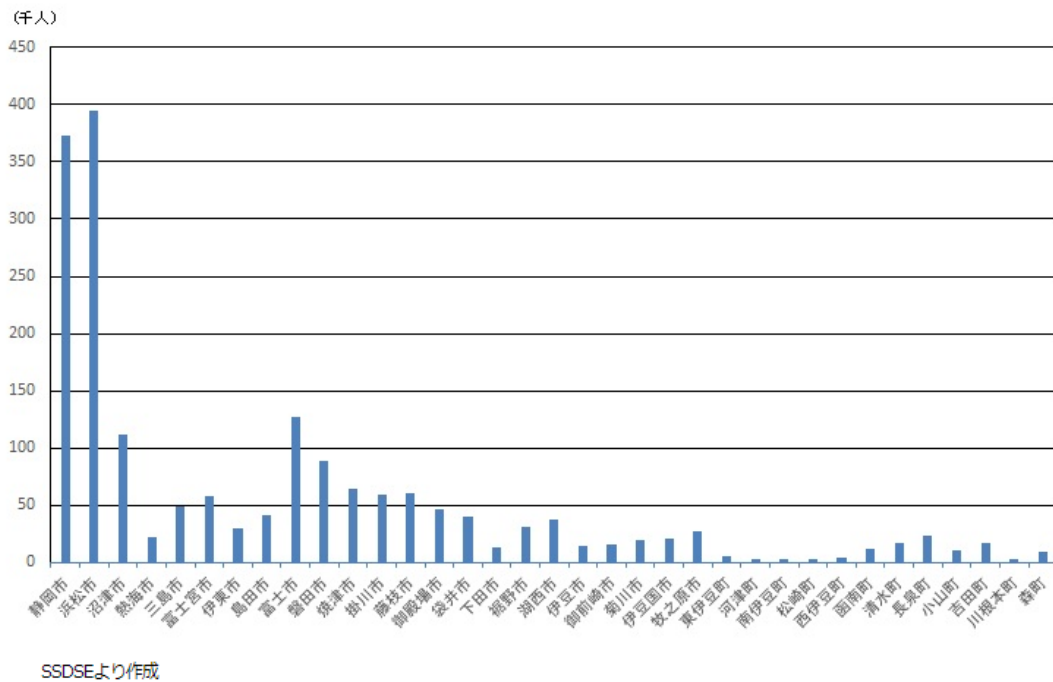
はじめに

SSDSE は、我が国の社会・経済情勢を正確に捉えることができ、かつ、できる限り簡易に分析が可能となるよう、公的統計を整備している総務省統計局の「都道府県・市区町村のすがた(社会・人口統計体系)」の市町村別データから作成しました。EBPM(証拠に基づく政策立案)などを推進していく上で、公的統計の役割がますます重視されており、統計データを的確に分析する能力が以前にも増して重要となってきています。今回は、基本的な統計モデルによる分析の事例を紹介します。

記述統計による考察

まず、記述統計による分析事例を紹介します。記述統計は、データを分析する方法の一つで、収集したデータの平均や分散などを計算して分布を明らかにし、データの示す傾向や性質を把握する手法です。下の図は静岡県(SSDSE 952 行目～986 行目)の従業者総数(同 60 列目)を示したものです。静岡市と浜松市の従業者数が他の市町に比べて相当多いことが分かります。次いで、富士市、沼津市、磐田市の順に多くなっています。

図1 静岡県の市町村別従業者数



次に、この市町別従業者数を更に産業別に見ることで、市町ごとの特長を分析します。表1は、産業別従業者数(61列目～77列目)を従業者総数(60列目)で除して計算しました。なお、データセットにある産業大分類17産業のうち、主な12産業だけを表章しました。同様に、静岡県の35市町のうち、23市のみを表章しました。

上述のとおり、静岡市と浜松市は従業者総数では他の市町に比べて相当数多いことが分かりますが、この両市を比較すると、産業構造は顕著に異なることがみてとれます。浜松市は製造業の従業者の割合が高く、静岡市は、卸・小売業、金融・保険業などの従業者数の割合が多いことが特長です。

また、他の市についてみると、全国平均に比べて製造業の割合が高く、特に、湖西市60.8%、磐田市42.8%、牧之原市42.7%など製造業の従業者割合が高いことが特長と思われます。また、宿泊・飲食業についても、全国平均に比べてその割合が高い市が多数あり、特に熱海市29.8%、伊東市22.4%、下田市22.2%、伊豆市20.2%などとなっています。

さらに、特長的な市をみると、裾野市では、学術・専門・技術サービスで16.0%、御前崎市では、建設業が12.6%、電気・ガス・水道業が5.1%などとなっています。

表 1 静岡県内の市別・産業別従業者の割合

(単位: %)

	建設業	製造業	電気・ガス・水道業	情報通信業	運輸・郵便業	卸・小売業	金融・保険業	不動産・賃貸業	学術、専門・技術サービス	宿泊・飲食業	生活関連・娯楽業	医療・福祉
静岡市	6.7	14.9	0.5	2.2	5.7	20.8	3.1	2.3	2.6	7.9	3.9	11.3
浜松市	6.0	22.6	0.3	1.2	4.9	18.3	2.0	1.9	2.5	8.9	4.1	11.6
沼津市	5.3	21.4	0.6	2.0	5.3	18.6	3.3	2.1	2.9	8.7	3.9	9.7
熱海市	5.7	2.2	0.2	0.3	4.5	15.0	0.9	5.4	1.6	29.8	4.4	14.5
三島市	6.5	15.6	0.5	1.4	4.4	16.1	1.7	3.1	3.6	12.0	4.5	11.5
富士宮市	6.1	34.4	0.1	0.3	2.7	16.3	1.4	0.9	1.6	7.8	4.0	11.1
伊東市	7.0	3.4	0.7	0.4	3.8	20.5	1.5	4.1	1.2	22.4	7.0	15.5
島田市	6.7	29.2	0.7	0.6	4.2	17.1	2.0	1.2	2.9	6.2	4.3	12.2
富士市	6.0	28.0	0.5	0.6	7.6	16.9	1.6	1.8	2.1	7.8	3.8	10.6
磐田市	3.8	42.8	0.3	0.5	4.1	13.1	1.3	1.1	3.7	6.1	2.9	9.0
焼津市	5.5	28.5	0.1	0.6	7.6	18.0	1.5	1.3	2.0	8.2	3.6	9.9
掛川市	5.3	35.1	0.5	0.4	4.9	14.2	1.4	0.9	1.4	7.8	3.6	10.7
藤枝市	5.8	24.4	0.3	0.3	4.6	18.8	1.9	1.8	1.9	8.2	4.0	12.1
御殿場市	4.0	21.6	0.2	0.1	4.6	17.4	1.2	1.7	2.8	10.9	5.4	10.4
袋井市	5.4	34.3	0.1	0.1	8.7	15.4	1.2	1.7	1.7	7.7	4.2	7.4
下田市	6.4	2.6	0.7	0.4	2.7	21.0	1.8	2.8	2.4	22.2	4.4	13.0
裾野市	4.1	33.5	0.1	0.3	5.3	12.3	0.8	1.3	16.0	5.1	4.3	6.3
湖西市	2.6	60.8	0.1	0.2	3.1	8.9	0.7	1.1	1.1	4.2	2.8	5.7
伊豆市	7.1	10.1	0.2	0.2	3.0	16.3	1.3	1.9	0.9	20.2	8.4	16.1
御前崎市	12.6	24.4	5.1	0.2	4.2	13.7	0.9	1.0	2.7	8.9	4.1	8.0
菊川市	3.9	38.6	0.1	0.0	3.7	15.2	1.1	1.2	1.9	6.1	3.2	11.1
伊豆国市	6.0	19.9	0.9	1.0	1.2	15.7	1.2	2.3	1.3	14.6	4.5	16.6
牧之原市	4.9	42.7	0.0	0.0	5.6	16.7	1.1	1.2	2.3	4.9	2.8	8.4
全国平均	6.1	14.9	0.5	2.6	5.3	19.5	2.4	2.4	3.1	8.9	4.1	12.8

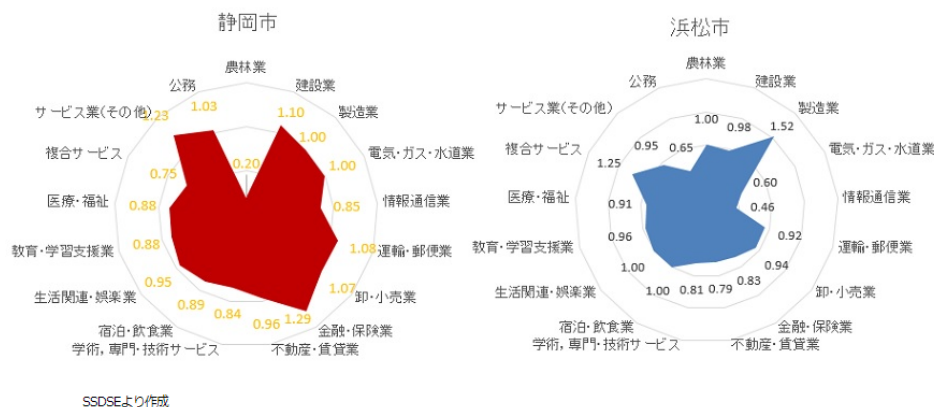
SSDSEより作成

さらに、この市町別従業者数について、「産業別の特化係数を(各市町の産業別従業者割合) ÷ (全国の産業別従業者割合)」で計算します。静岡市と浜松市の特化係数を比較すると、静岡市では、金融・保険業、運輸・郵便業、卸・小売業などサービス関連の産業で高く、浜松市では、製造業が高いことが分かります。

上記の2市以外を産業別でみると、表2では上述したとおり、製造業と宿泊・飲食業で高い係数がみられる市が多く、静岡県の産業の特長となっていると思われます。また、裾野市の学術、専門・技術サービスで5.16、御前崎市では電気・ガス・水道業で10.20と非常に高くなっており、特長的な性格を持つ都市があることが分かります。

このように特化係数が1よりも大きい産業を丁寧にみることで、該当する市の特長が分かります。

図1 静岡市と浜松市の産業別特化係数



SSDSEより作成

表 2 静岡県内の市別・産業別従業者割合の特化係数

	建設業	製造業	電気・ガス・水道業	情報通信業	運輸・郵便業	卸・小売業	金融・保険業	不動産・賃貸業	学術、専門・技術サービス	宿泊・飲食業	生活関連・娯楽業	医療・福祉
静岡市	1.10	1.00	1.00	0.85	1.08	1.07	1.29	0.96	0.84	0.89	0.95	0.88
浜松市	0.98	1.52	0.60	0.46	0.92	0.94	0.83	0.79	0.81	1.00	1.00	0.91
沼津市	0.87	1.44	1.20	0.77	1.00	0.95	1.38	0.88	0.94	0.98	0.95	0.76
熱海市	0.93	0.15	0.40	0.12	0.85	0.77	0.38	2.25	0.52	3.35	1.07	1.13
三島市	1.07	1.05	1.00	0.54	0.83	0.83	0.71	1.29	1.16	1.35	1.10	0.90
富士宮市	1.00	2.31	0.20	0.12	0.51	0.84	0.58	0.38	0.52	0.88	0.98	0.87
伊東市	1.15	0.23	1.40	0.15	0.72	1.05	0.63	1.71	0.39	2.52	1.71	1.21
島田市	1.10	1.96	1.40	0.23	0.79	0.88	0.83	0.50	0.94	0.70	1.05	0.95
富士市	0.98	1.88	1.00	0.23	1.43	0.87	0.67	0.75	0.68	0.88	0.93	0.83
磐田市	0.62	2.67	0.60	0.19	0.77	0.67	0.54	0.46	1.19	0.69	0.71	0.70
焼津市	0.90	1.91	0.20	0.23	1.43	0.92	0.63	0.54	0.65	0.92	0.88	0.77
掛川市	0.87	2.36	1.00	0.15	0.92	0.73	0.58	0.38	0.45	0.88	0.88	0.84
森枝市	0.95	1.64	0.60	0.12	0.87	0.96	0.79	0.75	0.61	0.92	0.98	0.95
御殿場市	0.66	1.45	0.40	0.04	0.87	0.89	0.50	0.71	0.90	1.22	1.32	0.81
袋井市	0.89	2.30	0.20	0.04	1.64	0.79	0.50	0.71	0.55	0.87	1.02	0.58
下田市	1.05	0.17	1.40	0.15	0.51	1.08	0.75	1.17	0.77	2.49	1.07	1.02
裾野市	0.67	2.25	0.20	0.12	1.00	0.63	0.33	0.54	5.16	0.57	1.05	0.49
湖西市	0.43	4.08	0.20	0.08	0.58	0.46	0.29	0.46	0.35	0.47	0.68	0.45
伊豆市	1.16	0.68	0.40	0.08	0.57	0.84	0.54	0.79	0.29	2.27	2.05	1.26
御前崎市	2.07	1.64	10.20	0.08	0.79	0.70	0.38	0.42	0.87	1.00	1.00	0.63
菊川市	0.64	2.59	0.20	0.00	0.70	0.78	0.46	0.50	0.61	0.69	0.78	0.87
伊豆国市	0.98	1.34	1.80	0.38	0.23	0.81	0.50	0.96	0.42	1.64	1.10	1.30
牧之原市	0.80	2.87	0.00	0.00	1.06	0.86	0.46	0.50	0.74	0.55	0.68	0.66

SSDSEより作成

主成分分析による考察

この静岡県の産業別従業者数のデータから主成分分析の手法を用いた分析を紹介します。主成分分析は、多くの説明変数を要約してその特長をみることができるというメリットがありますが、今回の分析では、17の産業の特長を2つの主成分に要約して分析しました。産業別従業者数のデータを標準化した相関係数行列から計算しています。主成分負荷量は参考表4にまとめていますが、第1主成分は、全ての産業でマイナスとなっており、かつ、農林業と電気・ガス・水道業以外は0.9以上となっていることから、これは従業者数を大きくする総合的な特性（都市特性）を示すと考えられます。また、第2主成分は、農林業、製造業、複合サービスでプラス、電気・ガス・水道業、情報通信業、金融・保険業でマイナスとなっています。

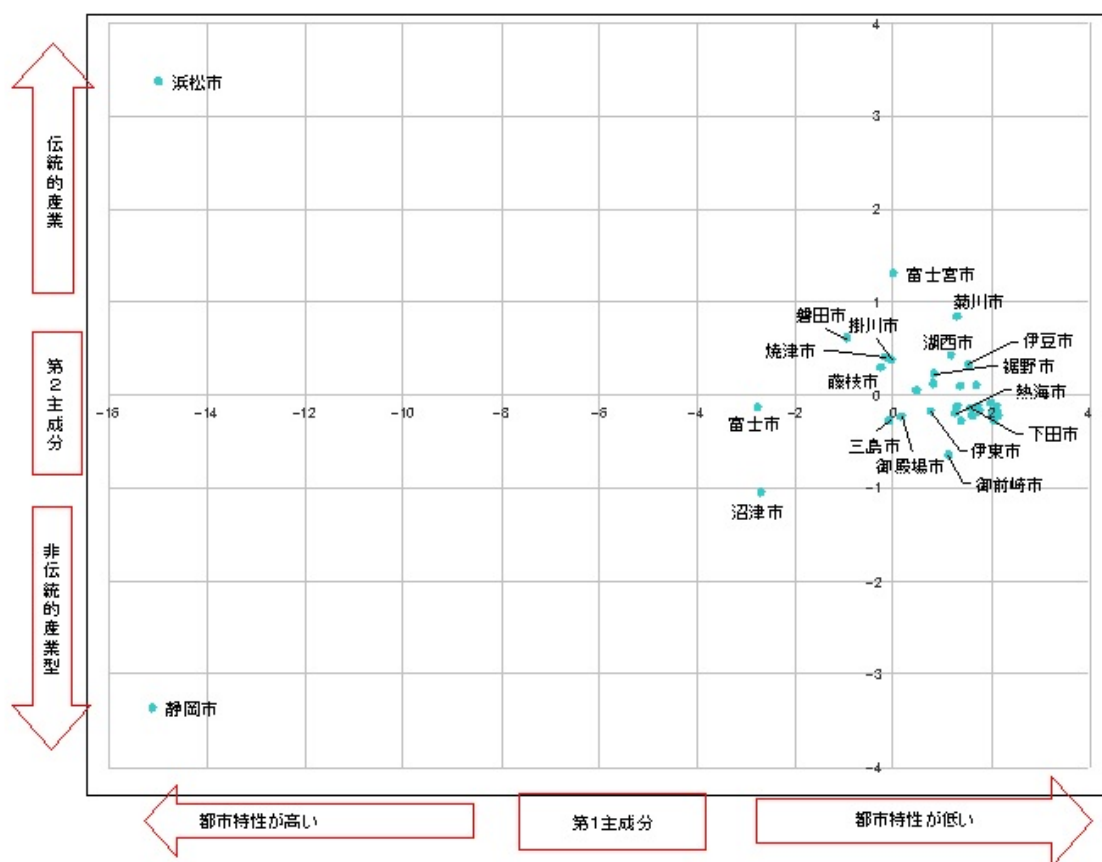
上述のとおり、第1主成分は都市特性の度合いを表していると考えられます。

記述統計による考察部分で説明した通り、静岡市と浜松市では他の市町に比べて大きな従業者数となっています。次いで、富士市、沼津市、磐田市と続いて、都市特性が高いという結果が現れています。

また、第2主成分として、静岡県の伝統的な産業、製造業、農・林業が高い特性を持つか、あるいは電気・ガス・水道業、情報通信業、金融・保険業といった第3次産業に含まれる非伝統的な産業特性を持っているかを表していると考えられます。静岡市と浜松市は対極に位置することがわかります。また、浜松市、富士宮市、菊川市、磐田市、掛川市、湖西市などでプラスの方向になっており、一方、マイナス方向に大きいのは、静岡市、沼津市、御前崎市などとなっています。第

以上、幾つかの統計的手法を用いた研究事例を紹介しました。研究の目的に合わせ、適切な統計データを選択し、適切な統計モデルを選択することが重要となります。本データセットはさまざまな研究を実施することが可能となっています。

図 2 静岡県の従業者数に関する主成分得点



SSDSEより作成

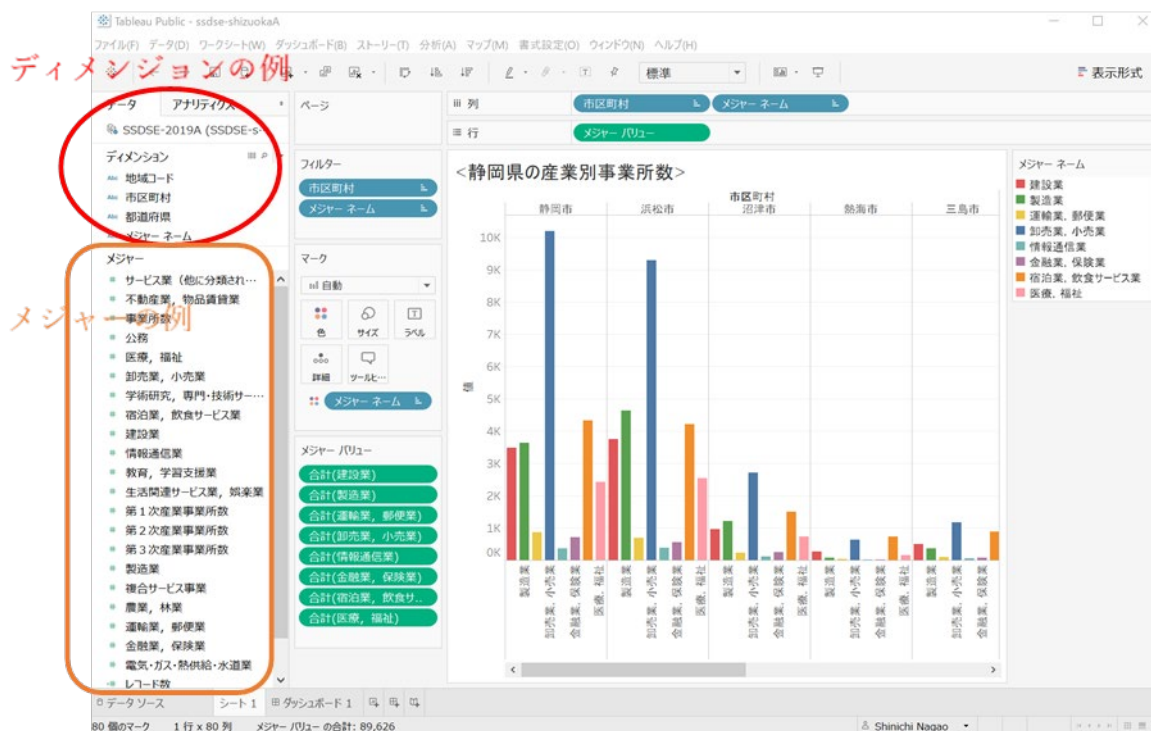
<参考文献>

- [1]総務省政策統括官（統計基準担当）編「高校からの統計・データサイエンス活用～上級編～」2017年3月発行
[2]加藤久和，「高校生からの統計入門」ちくまプリマー新書，2016年5月発行

② Tableau Desktop を活用した SSDSE データの可視化

独立行政法人統計センターが提供している SSDSE は行列形式で欠測値がなく、統計分析を容易に学習するために開発された統計教育用データセットです。Tableau（タブロー）Desktop は、一般的な表計算ソフトよりも簡単にグラフを作成することを可能にする等データの可視化に特徴があります。Tableau Desktop は、Tableau 社が提供しているソフトウェアであり、企業の経営意思決定に使える本格的なツールとしても有名な BI ツール（Business intelligence tool）ですが、教育機関での授業の利用や学生が使用する場合は、無償のライセンスを提供する Tableau Academic プログラムもあり、今後データ分析の重要性が高まる教育機関での活用が増加していくものと思われます。EXCEL、CSV 等様々な形式のデータに対応しており、データを読み込んだ後は、ドラッグ&ドロップ等の操作で手軽に可視化することができます。また、ダッシュボード機能を使うことで複数のグラフ等を 1 つのレポートとしてまとめることも可能です。

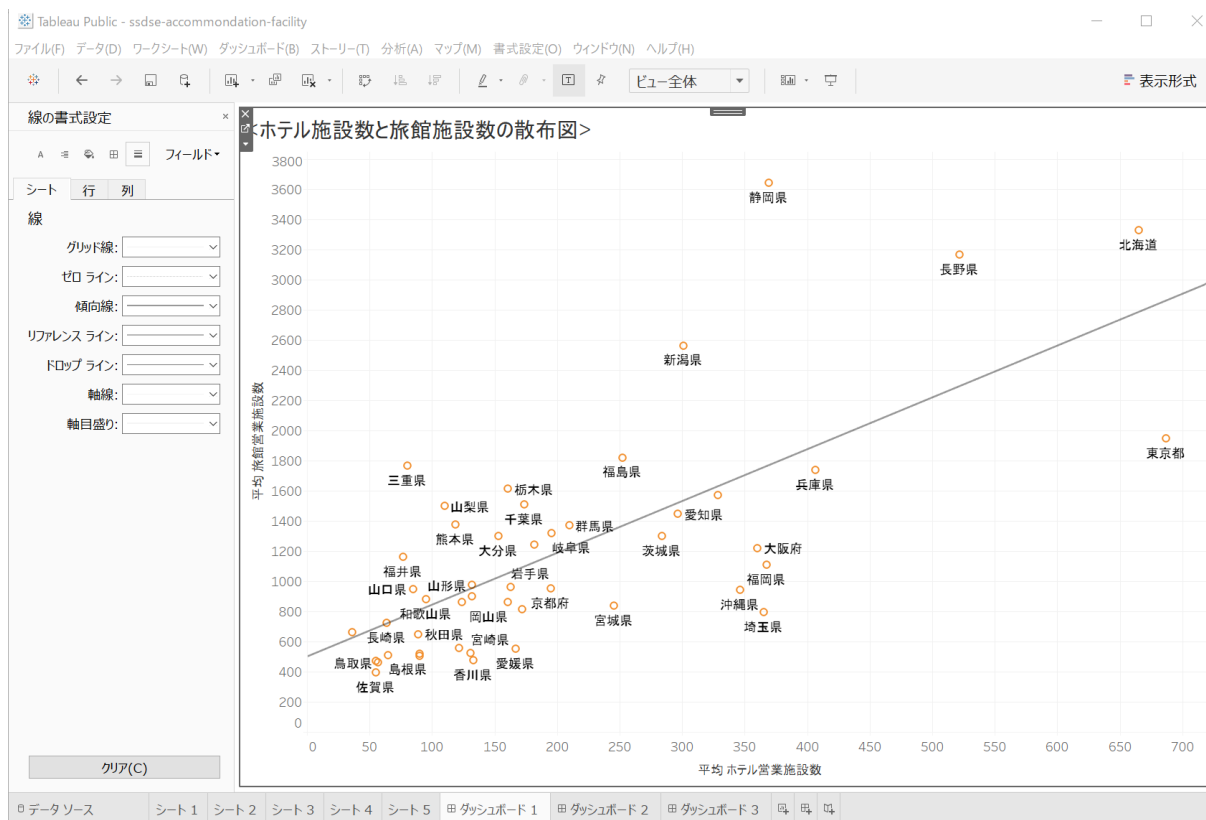
このソフトウェアは、すでに世界中で利用されていることから、チュートリアルビデオ等も豊富にあり、使い方を自学自習で学ぶこともできます。EXCEL 等で、データの加工やクリーニングを行い、Tableau Desktop で可視化するというような使い方もでき、EXCEL 等のデータを読み込んだ時点で、都道府県名や市町村名等の文字属性のデータをディメンション（定性的値）として、数値データはメジャー（測定可能な数値、又は定量的な値）として分けてくれます。これは、質的データと量的データの判別を自動的にしてくれる機能とも考えられます。



さらに、一般的な表計算ソフトと異なる点として、地図グラフが簡単に作成できることです。以下に SSDSE を加工した上で、Tableau Desktop を利用して作成した幾つかのグラフの事例を紹介します。

(1)SSDSE-2019B(EXCEL 形式)の「旅館営業施設数」と「ホテル営業施設数」からの分析事例

下図は、SSDSE-2019B から「旅館営業施設数」【AO 列】と「ホテル営業施設数」【AQ 列】の時系列データを抽出し、Tableau Desktop のグラフ機能を用いて散布図を描いてみたものです。傾向線は、5 種類用意されており、[線形]、[対数]、[指数]、[多項]、[累乗]の中から選択できます。今回は[線形]を選択しています。散布図は本来、2 つの変数の相関関係を見る際によく利用されるグラフですが、このように 2 つの変数をプロットしたグラフに傾向線を引くことで、それぞれの施設の数的大小が分かるだけでなく、日本の伝統的な宿泊施設である旅館の多い地域と都市部に多く立地するホテルの多い地域に分かれることも容易に分かると思います。



さらに、Tableau Desktop のグラフ機能を用いて、時系列データの折れ線グラフを描いてみました。下図は 47 都道府県のうち、ホテル施設及び旅館施設のそれぞれの数の多い上位 8 都道府県の 2005 年～2016 年までの時系列データをプロットしたものです。

ホテル施設は、北海道、兵庫県、大阪府、福岡県、沖縄県、静岡県等でこの12年間の間で上昇傾向にあることが分かります。

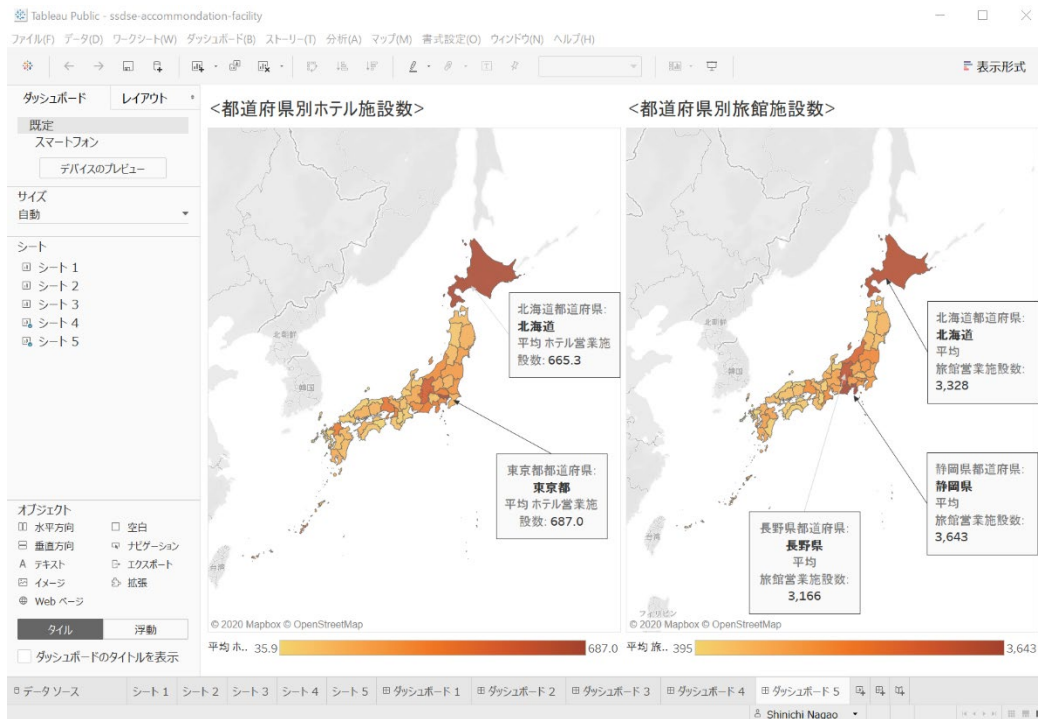
また、旅館施設は、数の多い静岡県、北海道、長野県、新潟県のいずれの地域もこの12年間で減少傾向にあり、東京都だけは近年上昇傾向となっていることが分かります。

近年、インバウンド旅行者の増加が話題になることが多いですが、このことは、旅館施設数の増加には貢献していないということがこの結果から伺えます。

下図のように Tableau Desktop ではグラフの特徴的なところに注釈をつけることも可能です。



また、Tableau Desktop では、地図グラフも容易に描けます。下図は2005年から2016年までのホテル施設数と旅館施設数の平均で色分けして地図に表したものです。ホテル施設は、東京都、北海道、次いで長野県の順に多く、旅館施設は、静岡県、北海道、次いで長野県の順となっています。このように Tableau Desktop を活用することで比較的簡単にデータと地図を可視化して表すことが可能になります。

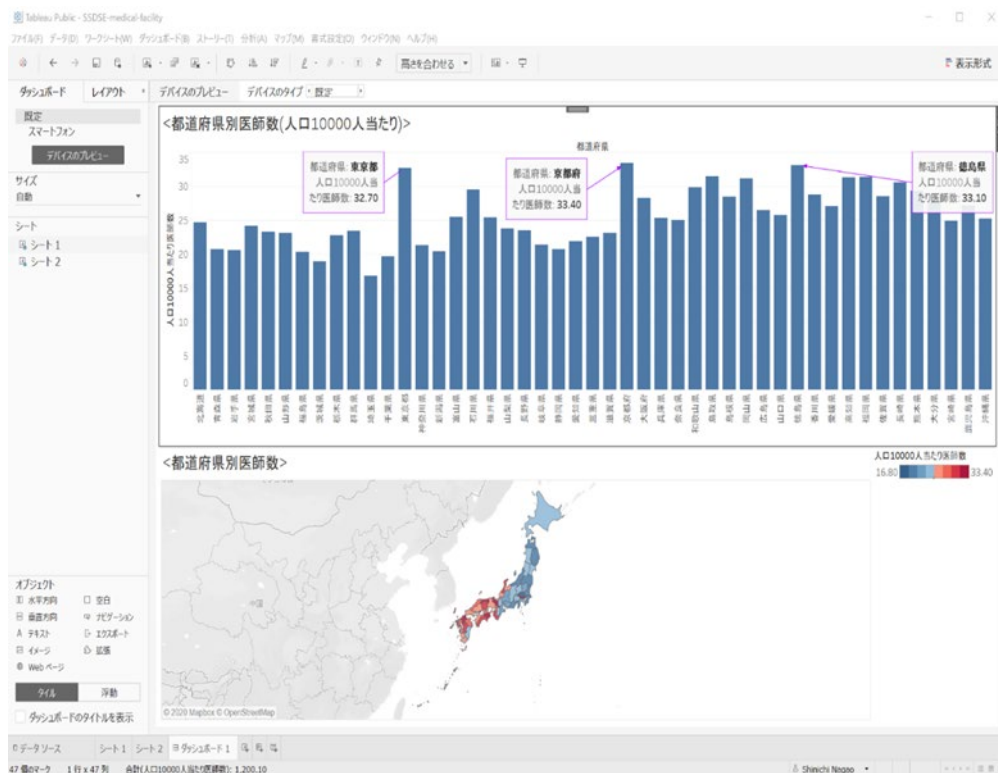


(2)SSDSE-2019A(EXCEL 形式)の「総人口」と「医師数」からの分析事例

SSDSE-2019A から「総人口」【D 列】と「医師数」【DT 列】のデータを抽出し、EXCEL で「人口 1 万人当たり医師数」を算出し、それを Tableau Desktop で描いたグラフが下図になります。「人口 1 万人当たり医師数」の最も多い都道府県は京都府で、次いで徳島県、東京都の順となっています。下段の地図グラフを見ると容易に分かりますが、日本では、医師数は西側の地域の方が総じて高いことが特徴²となっています。東京を含む首都圏にあらゆる機能が集中しているように感じることも多いですが、医師数に関しては西側の方が充実しているようです。このようにデータを地図に合わせて見ることで意外な事実を発見することもあります。

データを数値として見るだけでなく、グラフを用いて分析することや地図等で可視化することは、現在の社会情勢や地域の特性を把握する上で、いかに有効な手段かが分かります。

² SSDSE のデータを用いた日本の医師数の偏在についての先行研究としては大段(2019)があり、医療機関と自治体の経済力の関係等詳細な分析を行っています。



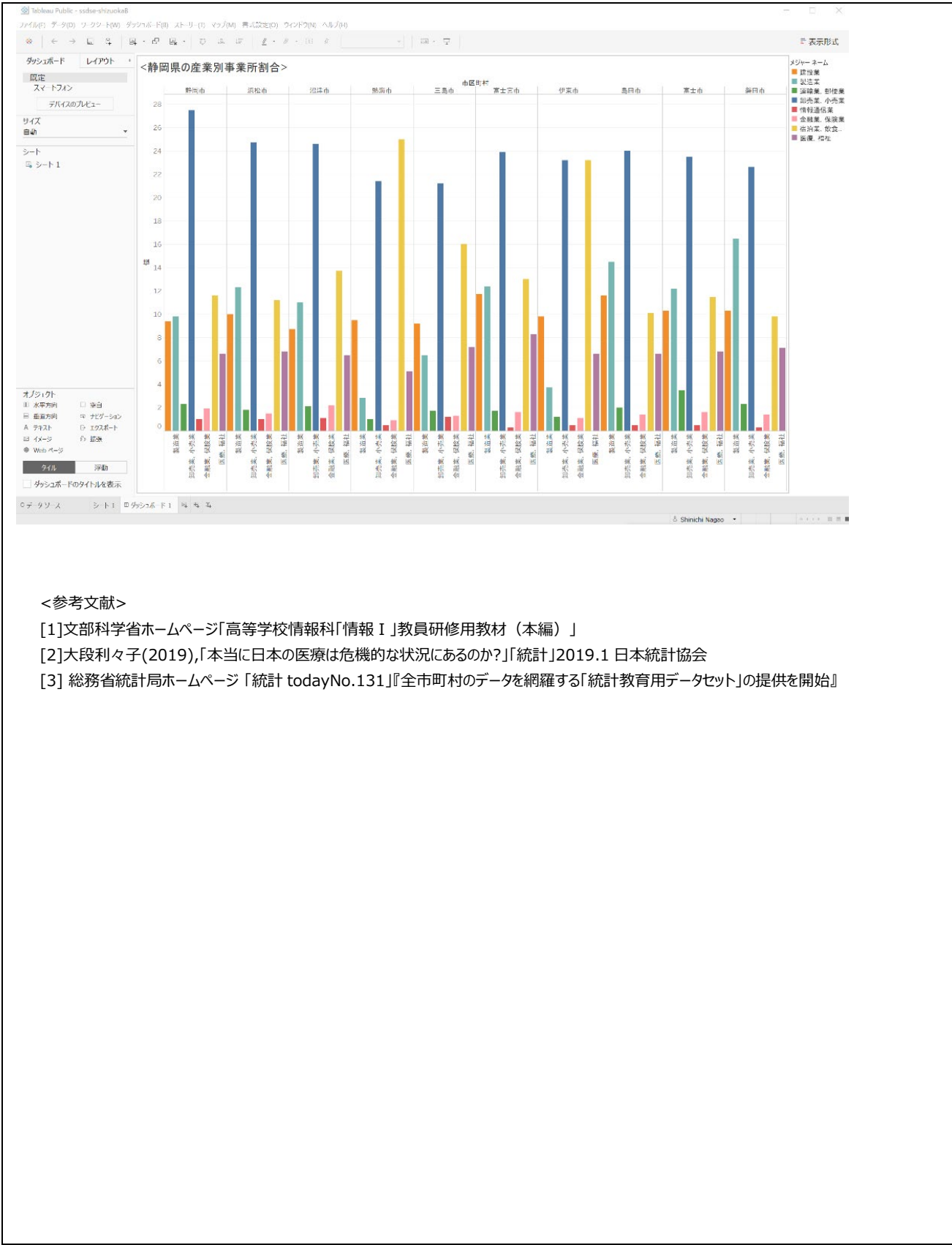
(3) SSDSE-2019A(EXCEL 形式)の静岡県の「事業所数」に係るデータからの分析事例

SSDSE-2019A で提供されているデータセットには、日本のすべての市町村別結果が収録されており、全都道府県について市町村別の分析することが可能となっています。下の事例は、静岡県の事業所数を抜き出して、各産業別「事業所数」を総「事業所数」で割ることで市町村別に産業別の事業所割合を計算できます。そのうちの主要な産業別事業所割合を分析したものです。

ここでは静岡県内の地域コードの若い順に 10 市を選択しています。このグラフは静岡県の経済の特徴をよく表していて、県庁所在市の静岡市は、「卸売業、小売業」の事業所割合が高く、本田技研工業株式会社、スズキ株式会社等多くのメーカーの事業所が立地する浜松市は製造業の割合が高くなっています。同様に、磐田市、島田市、富士宮市、富士市等でも多くのメーカーが立地していることから、製造業の割合が高くなっています。

また、温泉旅館が多数立地する熱海市、伊東市等で「宿泊業、飲食サービス」の割合が高いのも静岡県の特徴をよく表していると思われます。

SSDSE のデータには、産業別の事業所数や従業者数が市町村別にあります。これらのデータを可視化することで地域経済の特徴³を的確に把握することができます。



<参考文献>

- [1] 文部科学省ホームページ「高等学校情報科「情報Ⅰ」教員研修用教材（本編）」
 [2] 大段利々子(2019),「本当に日本の医療は危機的な状況にあるのか?」統計2019.1 日本統計協会
 [3] 総務省統計局ホームページ「統計 todayNo.131」『全市町村のデータを網羅する「統計教育用データセット」の提供を開始』

³ SSDSE のデータを用いた静岡県の産業別従業者数についての事例としては「統計 today No.131」があり、特化係数等を用いた産業別の特徴を分析しています。



Column

統計データ分析コンペティションの取組みについて

独立行政法人 統計センター

はじめに

統計データ分析コンペティションは、これからのデータサイエンス時代を担う高校生や大学生等を対象として、統計データ分析の論文を広く募集し、その分析のアイデアと技術を競うコンペティションです。

2018年度から開始し、2020年度には第3回のコンペティションを、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、及び一般財団法人日本統計協会の共催により開催する予定です。

募集する論文

このコンペティションでは、SSDSE（教育用標準データセット）を用いて分析した論文、又はSSDSEに他のデータを加えて分析した論文を募集の対象としています。

論文の内容は、高度な統計的分析手法を用いたものでなくても、SSDSEなどから発見した興味深いデータについて、新たな観点で可視化（グラフ化）するなどして、取りまとめたものでも応募できます。

論文執筆のガイドラインとして、論文原稿の様式（テンプレート）を提供し、論文の構成や図表の記載方法などを解説しています。論文を執筆した経験がなくても、この様式（テンプレート）に従えば、整理された構成の論文を書くことができます。

SSDSEとは

SSDSE（教育用標準データセット）は、このコンペティションの開始に合わせて2018年から統計センターで提供しているデータセットで、

Standardized Statistical Data Set for Education の頭文字を呼称にしています。

政府統計のデータ入手する場合、「政府統計の総合窓口（e-Stat）」から統計名、地域、項目、年次などを指定することで、任意にカスタマイズしてダウンロードできますが、この手順を省き、統計教育の現場で手軽に分析素材として使えるように整備したものがSSDSEです。

現在は、SSDSE-AとSSDSE-Bの2種類のデータセットを提供しています。いずれも縦に地域、横にデータ項目を並べた表形式のデータセットで、エクセルとCSVの形式で提供しています。

	縦（行）	横（列）
SSDSE-2019A	1741市区町村	125項目
SSDSE-2019B	47都道府県×12年 = 564行	107項目

比較的大きなサイズの表になっていますが、ダウンロード後に不要な行と列を削除すれば、そのまま分析に使用できます。なお、SSDSEのデータには欠測値がなく、全セルに数値（0の場合もあります）が入っています。

今後も毎年データを更新して提供していきます。さらに、新たな内容のSSDSEも整備・提供していく予定です。

論文の審査・表彰

統計データ分析コンペティションは、高校生の部と大学生・一般の部の2部門に分けて論文を募集します（短大、4年次以上の高専、大学院等は大学生・一般の部）。

応募のあった論文から、1次審査を経て、共催4者及び有識者で構成される論文審査会により、2つの部門ごとにそれぞれ、総務大臣賞、優秀賞、

統計数理賞、及び統計活用奨励賞が1件ずつ選定されます。さらに数件の特別賞が選定される場合があります。

各賞受賞者（特別賞を除く）には副賞として以下のとおり図書カードが贈られます。

	高校生の部	大学生・一般の部
総務大臣賞	2万円相当	5万円相当
優秀賞	1万円相当	3万円相当
統計数理賞	1万円相当	3万円相当
統計活用奨励賞	1万円相当	3万円相当

また、論文を提出していただいた応募者には、参加証を発行します。

主な日程

論文の募集開始から表彰式までの主な日程は、2019年度の場合、以下のとおりでした。

2020年度の日程は近日公表しますが、論文の募集開始から提出締切までの日程は、若干早くなる予定です。

2019年 7月1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
9月9日	大学生・一般の部 論文提出締切
9月17日	高校生の部 論文提出締切
10月18日	受賞論文の発表（統計の日）
11月18日	表彰式（全国統計大会）

統計専門誌への掲載

受賞論文は、日本統計協会の統計専門誌である、月刊誌「統計」に、論文審査会による添削・編集を経た上で掲載されます。2019年度の受賞論文の掲載実績及び予定は、以下のとおりです。

	高校生の部	大学生・一般の部
総務大臣賞	2020年1月号	2020年3月号

優秀賞	2月号	*	4月号
統計数理賞	*	5月号	* 6月号
統計活用奨励賞	*	7月号	* 8月号

（*は今後の掲載予定）

おわりに

統計データ分析コンペティションは、統計教育の場で活用できるSSDSEを実際に使ったデータ分析の論文を募集するコンペティションとして、2018年度、2019年度と2か年開催してきました。

応募論文の件数は、2018年度が27件、2019年度が57件と着実に増加し、力作・力作の論文の応募をいただきました。特別賞も含めた受賞論文の件数は、2018年度が8件、2019年度が14件となっています。

2020年度の第3回も、学生の皆様の積極的なご参加を期待しております。

<参考>

- ・統計データ分析コンペティション
(<https://www.nstac.go.jp/statcompe/>)
- ・SSDSE（教育用標準データセット）
(<https://www.nstac.go.jp/SSDSE/>)
- ・日本統計協会／月刊誌「統計」
(https://www.jstat.or.jp/monthly_toukei/)

2019年度 統計データ分析コンペティション 受賞者及び受賞論文

	高校生の部	大学生・一般の部
総務大臣賞	ワンオペ育児から見る離婚 お茶の水女子大学附属高等学校 竹内 通、江本 もえ、木下 舞、永井 あゆる	地方創生目標指標に関する変化要因ネットワークの 推定とそれに基づく地域間連携策の提案 筑波大学大学院システム情報工学研究科 張 瀧天、白鳥 友風
優秀賞	南海トラフ地震に備えて ～指定避難所に3人に1人が避難できず、災害時 の医療体制は本当に十分か？～ 愛媛県立松山南高等学校 渡邊 璃里香、吉田 美咲	我が国における人口増減の決定要因 大阪大学医学部 竹内 太郎
統計数理賞	過疎地域の現状分析と発展に重要な視点 福島工業高等専門学校 猪狩 信人	マルチレベル分析を用いた市町村大学等進学率の決 定要因分析 一橋大学経済学部 松本 洋輔
統計活用奨励賞	日本で暮らす外国人の動向から見た多民族化 広島大学附属高等学校 大段 利々子	市区町村別でみる合計特殊出生率推移の特徴分析 武蔵野大学工学部 村松 波、熊野 翔、川田 瑛貴
特別賞	旅館及びホテルにおける日本人・外国人宿泊客の都 道府県別増減から考える旅館の復活 —岡山県湯原温泉の視点からインバウンド需要を 旅館に取り込む方策— 岡山県立岡山操山高等学校 池田 雅子 香川県の交通事故発生率の要因を交通違反件数を基に 分析する 香川県立観音寺第一高等学校 宇川 昇吾、宮本 紫苑、山地 悠介	潜在患者数に対する医師偏在の可視化 東北大学大学院文学研究科 眞田 英毅 株式会社社会情報サービス 三浦 萌実
		外国人人口と市区町村の特性との関係性 関西学院大学経済学部 西尾 春香
		「広域連携の政策検証」 —空間計量経済学的手法による実証分析— 早稲田大学政治経済学部、商学部、社会科学部 原 康熙、福田 和生、柳田 はづき 地方創生実現のロジック —地域経済活性化のメカニズムを解明する— 早稲田大学大学院人間科学研究科 平原 幸輝