

令和元年版SSDSEの公開と統計リテラシー向上への取組み

—中高向け統計教材とコンペティションの結果報告—

独立行政法人統計センター 山下 雅代・飯島 信也

(独) 統計センターは、統計作成の専門機関として総務省と協働し、統計リテラシー向上への取組みを行っています。

令和元年版SSDSE(教育用標準データセット)

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

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す。欠測のない完備なデータセットの形 (エクセル及びCSV) で、無料ですぐに分析可能です。

DLはこちらから



SSDSE-2019A : 市区町村別 (2018年度版の更新)

データレイアウト

縦 : 1741市区町村データ × 横 : 125項目

縦 : 1744行 = 項目情報 (3) + 市区町村 (1741)
横 : 128列 = 地域情報 (3) + データ項目 (125)

code	prefecture	municipality	A1101	A110101	A110102	...	J2503	J2506
year	年度	2015	2015	2015	...	2016	2016	2016
R01100	北海道	札幌市	1952356	910614	1041742	...	258	25037
R01202	北海道	函館市	265979	120376	145603	...	45	3494
R01203	北海道	小樽市	121924	54985	66939	...	21	1394
R47375	沖縄県	多良間町	1194	639	555	...	1	35
R47381	沖縄県	竹富町	3998	2057	1941	...	0	0
R47382	沖縄県	与那国町	1843	1147	696	...	1	35

縦横項目の一覧表 (125項目) を含む。

SSDSE-2019B : 都道府県別・時系列

データレイアウト

縦 : (47都道府県 × 12年) × 横 : 107項目

縦 : 566行 = 項目情報 (2) + 都道府県 (47) × 年次 (12)
横 : 110列 = 年次情報 (1) + 地域情報 (2) + データ項目 (107)

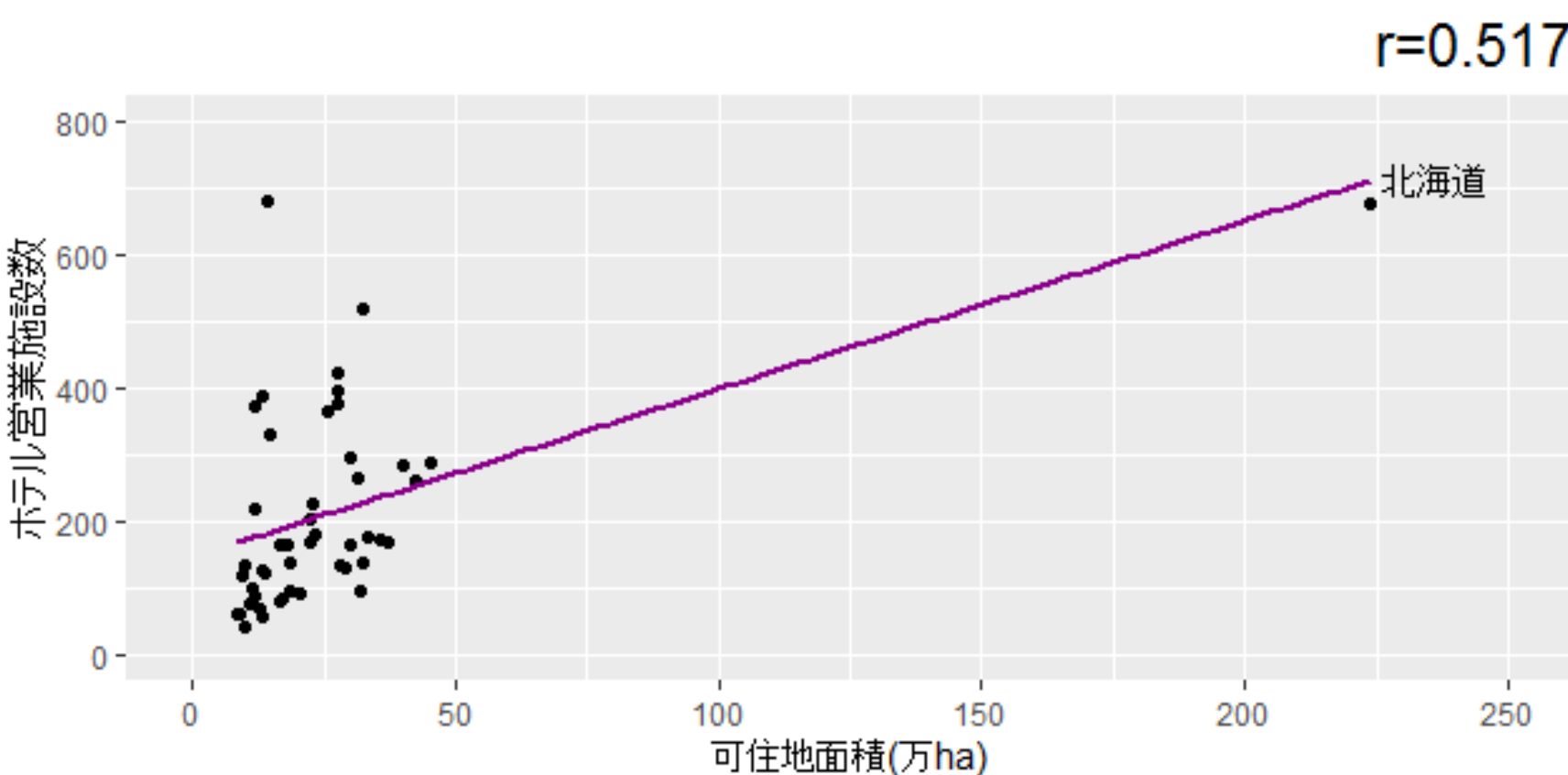
year	code	prefecture	A1101	A110101	A110102	...	L322109	L322110
2016	R01000	北海道	5352000	2521000	2830000	...	29305	59564
2015	R01000	北海道	5381733	2537089	2844644	...	26930	52197
2005	R01000	北海道	5605000	2675033	2952704	...	28072	62998
2016	R02000	青森県	1293000	688000	686000	...	18255	49302
2005	R46000	鹿児島県	1753179	819646	933533	...	28767	80787
2016	R47000	沖縄県	1439000	708000	732000	...	17320	48748
2005	R47000	沖縄県	1361594	668502	693092	...	16698	48159

縦横項目の一覧表 (107項目) を含む。

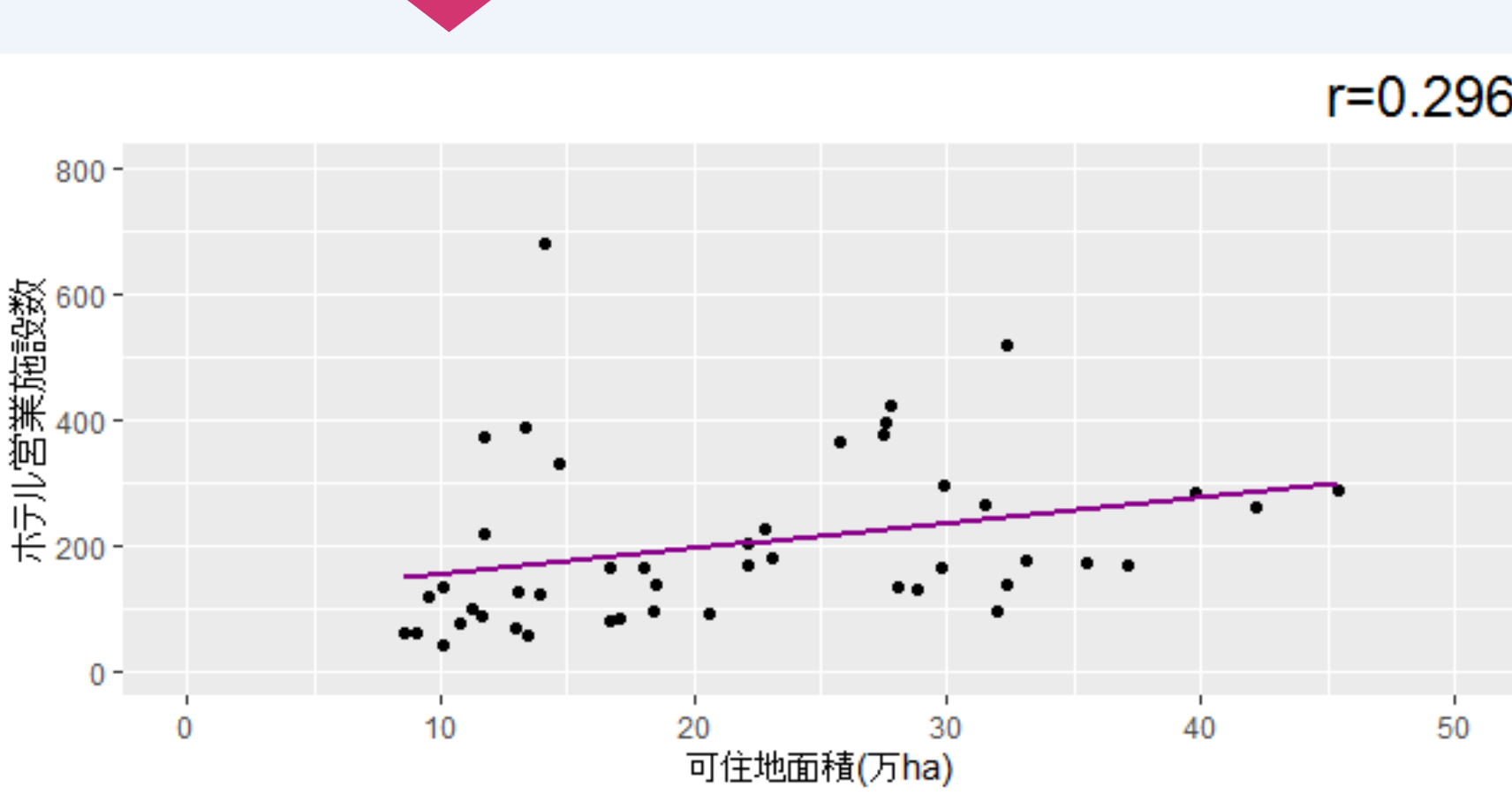
ニーズを踏まえ、来年も更新・拡張の予定です。ご意見・ご要望をお待ちしております。

SSDSEの活用事例 : 相関分析の例

下記の散布図には相関があると言えますか?



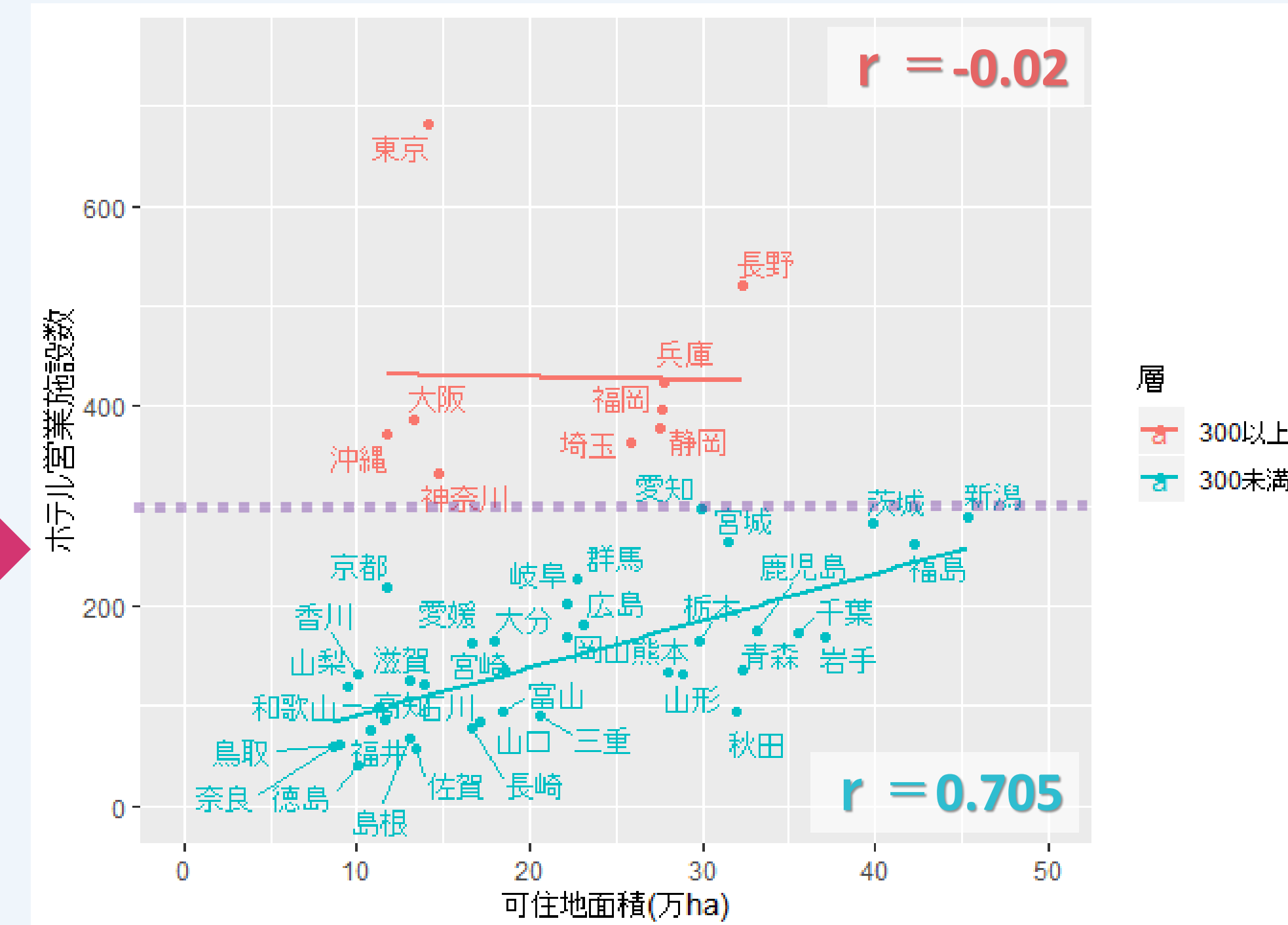
外れ値除去



外れ値を除くと、相関係数が大きく変化します。相関はないと言えますか?

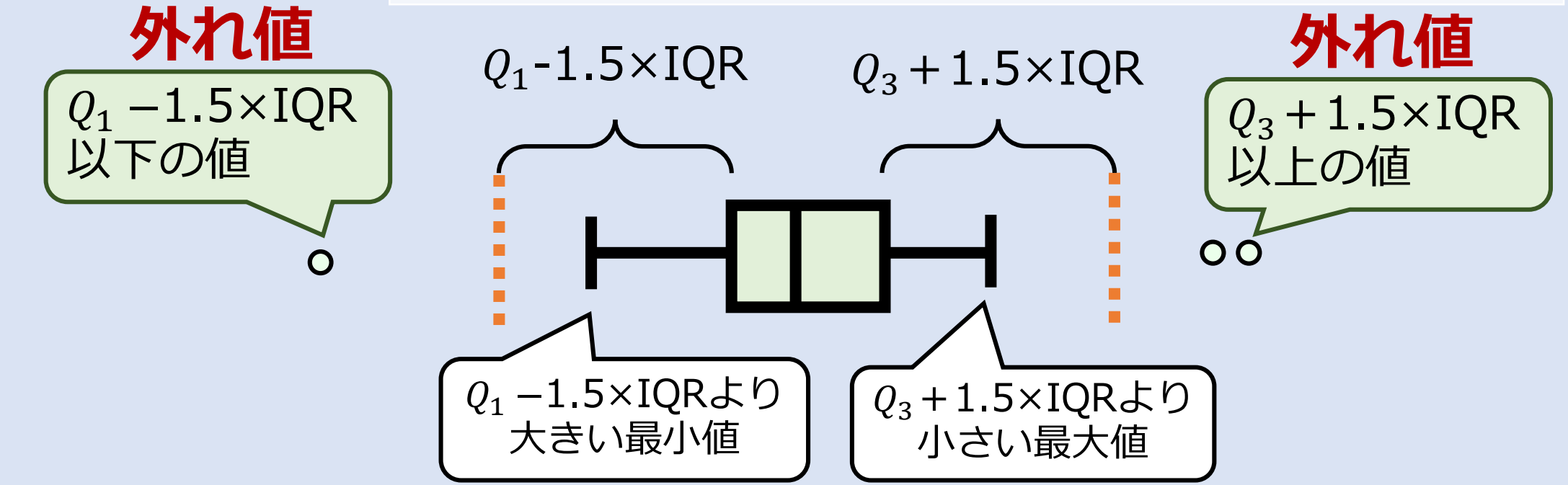
層別 : 違いが大きく出るように分類すること

ホテル営業施設数 ≥ 300、ホテル営業施設数 < 300 で層別すると、ホテル営業施設数 < 300 のグループに、相関関係が観察されます。



注 : 可住地面積とホテル営業施設数との間に因果関係があることを示すわけではないことに注意が必要です。

外れ値 : 散らばり方 (分布) の中心から離れた値
第1四分位数 (Q_1) - 1.5 × 四分位範囲 (IQR) 以下の値、または
第3四分位数 (Q_3) + 1.5 × 四分位範囲 (IQR) 以上の値、を外れ値と呼ぶ。



少子化をテーマにした問題解決教材は、日本統計協会発行月刊誌「統計」に連載

探究型統計教材一覧

学習内容	教材名	概要
① 四分位数、箱ひげ図	散らばり方を比較しよう	四分位数の導出と、箱ひげ図による複数の分布の比較
② 箱ひげ図、ヒストグラム	外れ値に着目しよう	外れ値に着目した問題発見と、層別と指数化による特徴の抽出
③ 相関係数、散布図	相関分析をしてみよう	因果関係で考えることの重要性 (疑似相関の確認) と、外れ値の処理
④ 層別	必要な汗はかこう	インターネットを用いた情報収集と外れ値市区町村の分類による対策立案
⑤ ドットプロット、代表値	散らばり方で白地図を塗り分けよう	散らばり方による分類と、地図グラフの作成

統計データ分析コンペティション 2019年度開催報告

☆趣旨 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。

☆賞及び副賞等 表彰優秀作品には、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、受賞論文が月刊誌「統計」に掲載されます。

2019年度実施概要

主催 : 総務省統計局、独立行政法人統計センター、統計数理研究所、一般財団法人日本統計協会
後援 : 国立研究開発法人科学技術振興機構 (JST)、一般社団法人日本統計学会、全国統計教育研究協議会、全国高等学校長協会
主な日程 : 2019年5月14日 (火) 統計データ分析コンペティションの開催発表
7月1日 (月) エントリー及び論文募集開始
8月9日 (金) エントリー締切
9月9日 (月) 論文締切 (大学生・一般の部)
9月17日 (火) 論文締切 (高校生の部)
10月18日 (金) 受賞論文の発表 (統計の日)
11月18日 (月) 受賞者の表彰式 (全国統計大会)
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)、大学、大学院の学生、一般の方

2019年度 受賞論文一覧

各賞	高校生の部	大学生・一般の部
総務大臣賞	ワンオベ育児から見る離婚 お茶の水女子大学附属高等学校 竹内遥、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の提案 筑波大学大学院システム情報工学研究科 張 瀚天、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

この他に 特別賞 があります。2019年度は、高校生の部2本、大学生・一般の部4本が受賞しました。

来年も開催予定です。ご応募、お待ちしております。2019年度版SSDSEを活用した作品もOKです。



受賞論文の詳細は、QRコードにアクセス下さい