

SPDA の使い方 — 誰でもできる二次分析用公開 — by 加藤 厚 (Jan. 2018)

二次分析 = (他の個人・組織などが収集した) 二次 data の分析



「人が事実を用いて科学を作るのは、石を用いて家を造るようなものである。事実の集積が科学でないことは、石の集積が家でないのと同様である。」
ポアンカレ『科学と仮説』



I 問題と目的

多くの労力・時間などを費やして収集した資料は、研究者の当初の問題意識・観点から分析され、得られた知見は論文などとして報告されます。しかし、この過程のみでは、同じ資料を異なる問題意識・観点から分析すれば得られたかもしれない知見は得られず、また論文などに含まれない結果などを知りたい場合、研究者に連絡して分析・報告を依頼する、素 data を入手して自ら分析するといった方法に頼るしかありません。逆に、折角の資料を異なる問題意識・観点から分析してほしい、論文などに含め得なかった結果についても関係者に伝えたいなどと研究者自身が考えても、それを簡便※に実現する手段は存在しないのが現状です(※例:学会などの組織に頼らず数日程度で・)。

この状況の改善方策の1つに、**集計・分析などを誰もがネット経由で行える形での資料公開を支援・促進する tool の作成と公表**があります。本マニュアルの目的は、そのような tool の一例となることを目指して作成された SPDA を紹介し、その「使い方」の要点を解説することです。

名前	更新日時	サイズ
spda.htm	2017/12/...	12 KB
data.js	2017/12/...	5 KB
10-2.htm	2016/03/...	1 KB

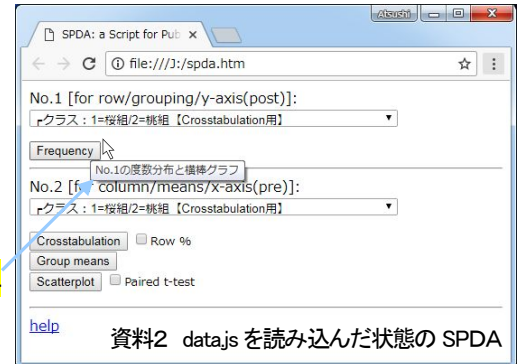
資料1 SPDAの本体・資料file例・変換tool

II SPDAの起動と機能確認

資料1中の spda.htm が本体で、data.js は資料fileの一例です※。

※data.jsには各機能の例示用の data (KogoLab 2003&2004) が保存してあります。SPDAは起動時に同一 folder 内の data.js を読み込むので、ネット公開には各自の資料を保存した data.js と spda.htm の同一 folder へのアップロードが必要です(資料9参照)。

ローカル(PCなど)では spda.htm のダブルクリック(ネット上では spda.htm への link のクリック)で資料2(資料9)の頁が開き、**ボタンなどに矢印を重ねると、簡単な説明が表示されます**(この仕組みについては V まとめなどの最終段落参照)。



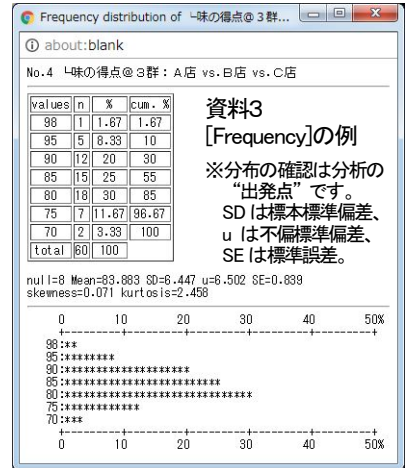
資料2 data.jsを読み込んだ状態のSPDA

No.1とNo.2の一覧から変数を選択し、変数の特性(カテゴリか数量か)を考慮して最適なボタンをクリックすると処理結果が新頁に出力されます。

① 各変数の“全体像”を掴む [Frequency] :

クリックで No.1 の変数の各値の度数と分布の特徴の指標が出力されます(資料3参照)。

平均値が有意なら数量(含人数、得点、重さなど)、無意味ならカテゴリ。



資料3 [Frequency]の例

※分布の確認は分析の“出发点”です。SDは標本標準偏差、uは不偏標準偏差、SEは標準誤差。

② カテゴリ変数間の関連(=連関)を検討する [Crosstabulation] :

No.1のカテゴリ変数を表側(row=行)に、No.2のカテゴリ変数を表頭(c1m=列)に置いた各セルの該当件数などを出力します(☐ Row%をチェックすると行%を追加)。資料4の例では、クラスと好きな科目に弱い連関が認められます(クラメールの V=.257)。

	1	2	total
1	18	18	36
2	9	24	32
total	26	42	68

資料4 [Crosstabulation]の例

③ カテゴリ間で数量を比較する [Group means] :

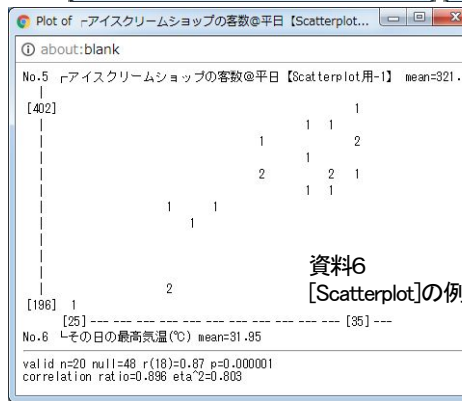
No.1のカテゴリ変数の値に基づく群ごとに、No.2の数量変数を集計します。後者に対する前者の影響を検討したい時に適切な分析です。資料5の例では店の要因と味の得点との間に中程度の相関比が認められます($\eta = .548$)。

values	val id	n	mean	SD	u
3	20	88.15	5.341	5.40	
2	20	79.5	6.454	5.596	
1	20	84	5.385	5.525	
total	60	83.883	6.447	6.502	

資料5 [Group means]の例

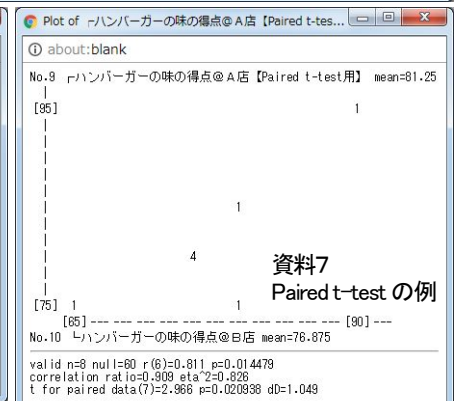
④ 数量変数間の関連(=相関)などを検討する [Scatterplot] :

No.1の数量変数を縦軸、No.2の数量変数を横軸に置いた散布図を描き、相関係数 r などを出力します。資料6の例では、最高気温と客数@平日の間に強い相関が認められます($r=.87$)。



資料6 [Scatterplot]の例

また、☐ Paired t-test をチェックすると、対応のある数量変数(例:同じ人による2店の味の評価)間の差のt検定が追加されます。資料7の例では、A店とB店の評価に有意な差が認められます($p<.05$)。



資料7 Paired t-test の例

SPDAの統計処理機能はSQP(加藤2009)のサブセット(部分集合)です。各出力のより詳細な意味などについては「SQPの使い方」を参照してください。

有効な値は-99999~999999の整数と小数
(クロス表や群化のカテゴリとしては 0以上)

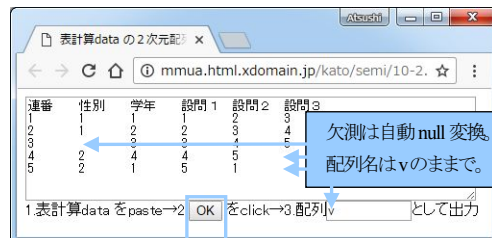
III 資料 file の作成

SPDA の機能が把握できたら、10-2.htm で各自の資料 file を作成します※。
※SPDA には読み替え機能などはありません。有意義な新変数(例:数量に基づくカテゴリ)は事前に表計算ソフトや SQL で作成・追加しておきましょう。

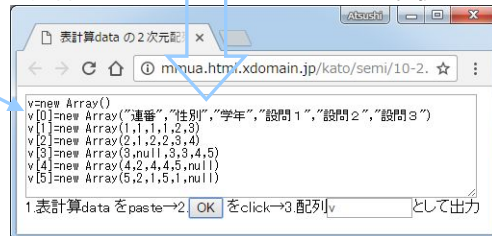
SPDA 用の資料 file は2次元配列(資料8の出力参照)を保存した外部ファイルです。10-2.htm※の入出力欄に表計算ソフトの data★をコピーして変換し、出力をエディタにコピーしてファイル名 data.js(文字コードは UTF-8)で保存します#。

※「素朴に記述する JavaScript」(加藤 2015-2016)第 10 回の内容その2なので。
★(無回答などの)欠測は空欄のままにしておけば自動的に null に変換します。
出力の末尾に(不要な改行に由来する)無効な行があったら要削除です。

新しい data.js と spda.htm を同じ folder に置き、後者をダブルクリックで起動して No.1・2 の一覧から変数を選択の上、処理を行って結果を確認しましょう。



資料8 10-2.htm による2次元配列への変換



IV 独自資料の公開

公開にはアップロードと link 紹介が必要です。まず、無料レンタルサーバー(例: Xdomain)などにスペースを確保し、作成した folder に spda.htm と各自の data.js をアップロードしましょう。次に、その spda.htm への link※を説明とともに関係者への mail、各自の web 頁・blog などで紹介・提示すれば公開完了です。

※資料9の例では <http://mmua.html.xdomain.jp/kato/spda/spda.htm>



V まとめなど

SPDA が処理できる data は「カテゴリ」と「数量」に対応する数字のみ(変数名は除く)で、値を読み替える、足し上げて尺度化する、といった機能はありません(SQP はこれらの機能付)。

他方、表計算ソフトの data から容易に作成できる2次元配列の外部ファイルとセットにして、ネットにアクセスできる誰もが2変数までの基本的な集計・分析を行える形で、つまり二次分析可能な形で資料を公開できる tool です。

そもそも、SPDA を作るキッカケは、2005 年の夏休みに中別府温和先生から「インドのグジャラートで行った調査※の結果を、回答者を含む現地の人たちに報告書よりも詳しく示す方法はないだろうか?」と相談されたことでした。

※<http://mmua.html.xdomain.jp/bak/gujarati.pdf> 【←教示文中の All the results are sure to be given back to the Parsi community.に注目】

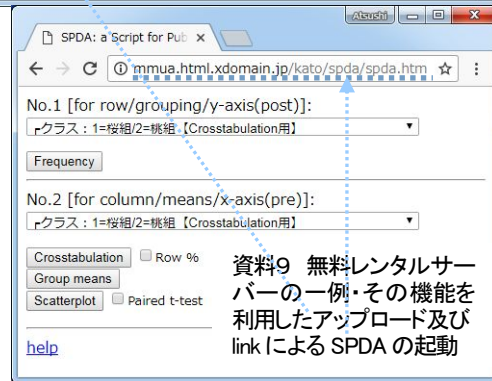
当時、プログラミング言語としての JavaScript の可能性に注目していたボクは、それなりの script を書き上げて先生に差し上げるとともに、関連情報をまとめて「二次分析機能を備えた online data archive の利点とその要件に関する検討」(加藤 2006)として発表しました。その script をあれこれ書き改めたのがこの SPDA version 1.0.9 です。

SPDA で使われていて「素朴に記述する JavaScript」(加藤 2015-2016)で言及・解説されていない内容は確率関数の function と「簡単な説明」のポップアップ(資料2内の「No.1 の度数分布と…」)なので、本資料の末尾で後者を説明しておきます。

資料 10 は[Frequency]ボタンと help リンクの script です。前者にポインタが重なると、1行目の title に続く内容(「No.1 の度数分布…」)がボタンのそばに提示されます。また、後者にポインタが重なると onmouseover 状態となるため関数「inf()」が実行されます。これらの機能の利用によって、表記が主に英語の SPDA の「使いやすさ」の向上を目指しました。

```
<input type='button' value='Frequency' onClick='fiq()' title='No.1 の度数分布と棒グラフ'>
(省略)
<a href='javascript:void(0)' onmouseover='inf()'>help</a>
```

資料 10 ポップアップは title="説明"と onmouseover="関数"



資料9 無料レンタルサーバーの一例・その機能を利用したアップロード及び link による SPDA の起動

文献

加藤 厚 2006 二次分析機能を備えた online data archive の利点とその要件に関する検討

<http://mmua.html.xdomain.jp/kato/study/2006oda.pdf>

加藤 厚 2009 SQP(a Scripted Questionnaire Processor)

<http://www.vector.co.jp/soft/other/java/se478769.html>

加藤 厚 2009 SQP の使い方

<http://hp.vector.co.jp/authors/VA021211/beta/sqp.pdf>

加藤 厚 2015-2016 素朴に記述する JavaScript

<http://mmua.html.xdomain.jp/kato/semi/javascript.htm>

KogoLab 2003 アイスクリーム屋さんで学ぶ楽しい統計学

<http://kogolab.chillout.jp/learn/icecream/>

KogoLab 2004 ハンバーガーショップで学ぶ楽しい統計学

<http://kogolab.chillout.jp/learn/hamburger/>