

SPSSシンタックスの基本

立教大学社会学部 村瀬 洋一

シンタックスは、SPSSにおいて使う各種の命令文（プログラム）である。これを使うと、複雑な分析やデータ加工であっても、大量に一度に実行できる。また実行した記録が残るため、後で同じ作業を繰り返すことも容易である。とくにデータ加工を覚えることが大切。

SPSSを起動すると初めはデータ・ウィンドウのみ出る。分析結果を出すとき出力ウィンドウも出てくる。また、シンタックス・ウィンドウも作ることができる。つまり、SPSSには3つの画面がある。このことをまず理解すること。

データを読み込むと、データ・ウィンドウ下に「データビュー」と「変数ビュー」のタブ



があることで、確認すること。データビューは、1人のデータが横1行になっている。つまり500人分のデータならば500行ある。変数ビューでは、各変数（調査でいえば質問項目）の名前や値などを確認できる。詳しくは参考文献を読むこと。

図1. SPSSデータ画面

なお、パソコンのデータ記憶場所は、Cドライブ（ハードディスク）やDドライブ（DVDドライブ）などがあることをまず理解する。ドライブ内にフォルダやファイルを作る。

0. シンタックス・ウィンドウを開く

SPSSを起動し、画面上の「ファイル」をクリックし、新規作成→シンタックスを選ぶ

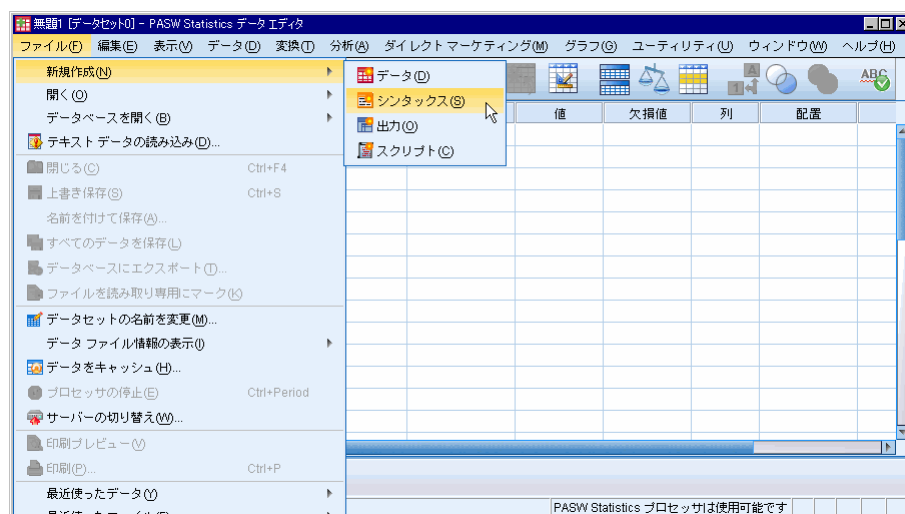


図2. SPSS 新規作成画面

すると、シンタックス・ウィンドウが出てくる。この後、以下の例1～7のように、シンタックスを書いてから実行する。シンタックスは、分析などの実行命令文である。SPSS

上で使う簡単なプログラムといってもよい。これを使いこなすことが重要。これにより、新変数の作成などが簡単にできる。

次の図の中は、欠損値処理MISSING VALUES、COMPUTE文による回答方向を逆転した新変数作成、合計得点作成、自営業ダミー変数作成の例。これらを書き、実行したい部分を選択し、CTRL+Rで実行（あるいは画面上の分析→実行）実行。村瀬研ホームページなどに各種のシンタックス見本ファイルがある。それらを書き換えると簡単。

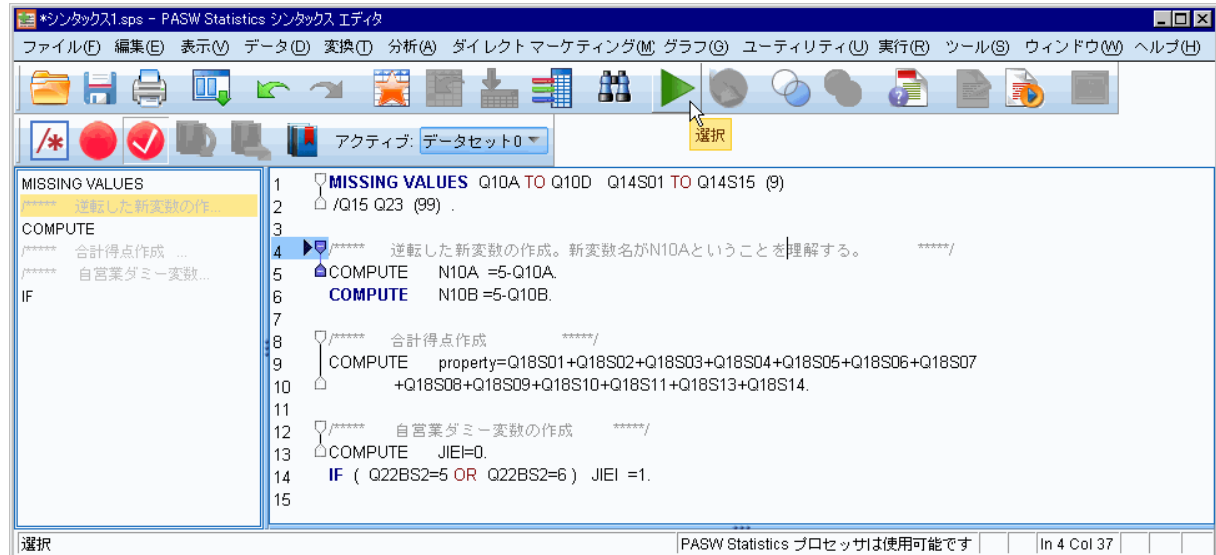


図3. シンタックス・ウィンドウの例

自分が作ったシンタックスは好きな名前でも保存する（拡張子sps）。シンタックスを作る時は、以下の点に注意する。

- ・半角英数字のみを用いる。全角スペースや全角のカッコなどは入れない。
- ・命令文は必ずピリオドと改行コードで終わる。
- ・各行末に余計なピリオドなどが入っていないか、確認する。
- ・保存時のファイル形式はワードなどで編集した場合は「書式なし(txt)」
「MS-DOSテキスト」とする。

1. シンタックスによるデータ読み込み

◆SPSS形式データの場合(拡張子sav)

シンタックスによってデータを読み込む必要はない

◆テキスト形式データの場合

まずCドライブなどにデータを置く。普通ハードディスクはCドライブ。

シンタックス内の最初にある、データファイルの位置指定を、以下のように書きかえる

初心者は、まずSPSS形式データファイルを開いた上で、以下にある4の度数分布表の命令文を実行してみると良い。ピリオドがある所までが1つの命令文。見本のシンタックスを書き替えると簡単。自分のPCにエディターソフトを入れ中身を確認すると良い。

分析の手順 **まずやってみよう！**

データを開く→シンタックスウィンドウに命令文を書く→実行→出てきた表を確認。
データを開いた状態で、シンタックスを実行すること。
シンタックスを書いたら名前を付けて保存する。

1.1. テキスト形式データファイルの読み込みシンタックス解説

シンタックス最初で、以下のようなDATA LIST FIL命令文を使う。データ場所の指定文として ' ' の間にデータファイル名を書く。CドライブのRENSYU という名前のフォルダの中に15fuku.txtというファイルが入っている場合は、以下ようになる。

```
DATA LIST FILE=' C:¥RENSYU¥15fuku.txt' RECORDS=2
```

自分のUSBメモリーなどがFドライブの時は F:¥ と書く。¥で区切ってフォルダを書く。RECORDS=2は1人分が2行のデータの場合。この後に桁指定文を書きピリオドを書く。

現在のフォルダー場所（ファイルが置いてあるアドレス）のことをパスという。パスを表示するには、Windowsでファイルを表示し、画面上のアドレスバーのどこかを右クリックし、アドレスの編集をクリック。個々のファイルの完全なパスをコピーするには、Shiftキーを押しながらファイルを右クリックして「パスのコピー」をクリック。

1.2. パスの書き方

コンピューターの「ライブラリ」内の、ドキュメント内のdata5というフォルダ内にデータ2019abc.txtがある場合は以下のように書く。

```
C:¥Users¥USERID¥Documents¥data5¥2019abc.txt
```

USERID の部分は、Windowsにログインした時のユーザー名(muraseなど)を書く。

フォルダやファイル名を正確に書くことが重要。全角空白や余計なピリオドを入れない。また、ユーザー名は半角英数字にすること。すべて書いてから、すべて選択しシンタックスを実行すればよい。実行しても、初めはデータウィンドウは空に見えるが、何らかの分析をするとデータに数字が入る。なおマックOSのパスの書き方は別途、自分で調べること。

1.3. S P S S 形式データファイルの読み込みシンタックス解説

GET FILE命令文を使って読み込むこともできる。例えば、Hドライブに abc77.sav というファイルが入っている場合、以下ようになる（フォルダは使っていない場合の例）。

```
GET FILE ='H:¥abc77.sav' .
```

テキスト形式データと違って、この後に桁指定文は不要。この後にリコード文や分析命令文などを書けばよい。

2. シンタックスの構造

普通、次の順で命令文が並ぶ。

- 1 データの場所指定文 — これによりデータファイルをSPSSに読み込む
- 2 データのケタ指定文
- 3 欠損値処理
- 4 データ加工 値の変換、新変数の作成、計算など
- 5 単純集計やクロス集計などの分析命令文

見本シンタックスをよく見る。まず、以下のデータ加工について理解することが大切。

欠損値処理はふつう、無回答は9か99を入力する。

```
DK/NA Don't know, No answer'
```

3. 分析命令文のシンタックス例

```
FRE ←度数分布表を出せ、という命令  
VAR =Q8 Q9A. ←使いたい変数名を正確に書く
```

CRO ←クロス集計をしる、という命令
 /TAB=Q8 BY Q9A
 /CEL= COL. ← COLでなくROWだと横%、COLだと実数になる

これは3行で一文なので、最後に1つだけピリオドがつく。

CRO
 /TAB=Q8 by Q9A by Q46SEX ←性別を入れて3重クロス集
 /CEL = COL
 /STA = CHISQ PHI CORR BTAU CTAU . ← χ^2 値やファイ係数など各種の統計量を出す

ONEWAY ←一元配置分散分析をせよ、という命令文
 Q6A Q6B_edu BY nendai ←下線部は好きな変数名を書く (nendaiごとの平均値の例)
 /PLOT MEANS ←平均値の折れ線グラフを出す
 /STA DES. ←**基本統計量**を出す。年代ごとの平均値など出すと便利
 BYの後は、説明変数になるものを1つだけ書く。前は、複数の変数を書いても良い。

4. クロス集計表の形式について

表内に人数と%が混ざったクロス集計表は分かりにくい。以下のような形式で作る(あるいは原・海野(2004)「エラボレイション」p. 84にあるbのような表を作る)こと。表内には縦%または横%の、どちらかのみを書く。混ざっていると分かりにくい。SPSS出力そのままでは分かりにくい。エクセル上で自分で表を作ることが大切。

無料ソフトのPSPPの場合は、出力をCSVファイルでまず保存する。その後、エクセルを開き、画面上「データ」をクリックして、CSVファイルを読み込む。その後、エクセルで表を作成する。ワードの文章に表を貼り付ける時は、ワード上で、挿入→オブジェクト→エクセルワークシートとしてから、表を貼り付けるときれいに貼ることができる。

以下は表形式の見本。合計して100%となることが分かるように、かつ表の外に、100%の人数を書くこと。表中には縦%のみを書く(あるいは横%のみ)。Yになる変数について、合計して100%になるようにすればよい。以下の表は、2×5表で10セルある。タイトルも適切に付ける。エクセル上で自分で書くこと。以下の図4のような横棒グラフ(横帯グラフ)にしても良い。その場合、男女で2本の棒がある。各棒の中で、合計で100になる%を書く。

表1. 幸福感和性別のクロス集計表 (%) 2015年仙台調査

	男性	女性
幸せだ	17.9	18.9
だいたい幸せだ	62.7	67.9
あまり幸せでない	16.0	10.2
幸せでない	3.1	2.6
DK/NA	0.3	0.3
計	100.0	100.0
(人数)	(582)	(607)

注意点

100%の人数を必ず書く。

小数点の位置をそろえること。右クリックして「セルの書式設定」。

小数点以下はなくてもよい。

出力画面で、ファイル→エクスポート で、エクセル形式で保存ができる。

PSPPの場合ファイル→書き出しで、ct1.csv のような名前をつけて保存。

エクセル上で分かりやすい表に直し、小数点の位置をそろえる。

100%の度数も人数として書くこと。

5. シンタックスによるデータ加工文

データを読み込んだ後に、カテゴリー合併や、新変数の作成、変数の計算などができる。
シンタックスで新変数を作ることは重要。クリックして処理すると、大量の処理を一度にすることはできない。また後で、分析時に再現することも難しい。

自分でシンタックスを書いて変数を作り、必ず、シンタックスを保存しておくこと。

5.1. リコード文による値の変換（テキストpp. 38-）

書式 RECODE 変数名 (条件式) ピリオド

例1 既存の変数Q10の4段階回答を2段階にする。

```
COMPUTE      N10 = Q10 .
RECODE       N10 (1, 2=1) (3, 4=2) .
MISSING VALUES N10 (9) .
```

```
/****** 上記の1行目は、新変数名（新しい質問項目）として      *****/
/****** N10を設定し、その中身をQ10と同じにしている。      *****/
/****** 新変数がN10ということを理解する。N10は好きな名前で良い *****/
/****** 2行目はリコード文でのカテゴリー合併                  *****/
/****** 3行目は欠損値処理。無回答9を分析から除く処理      *****/
```

これを実行しても何も起きないように見えるが、出力画面でエラーが出ていなければ、問題ない。その後、新変数N10を使って分析するとデータにN10が追加される。

何が既存変数名で、何が新変数名か理解する。既存変数名を間違わないことが重要。当然だがQ10はデータによって違う。Q10aはあるがQ10がない、というデータもあるので注意。

例2 Q46AGEという細かい変数（既存変数）をNENDAIという5段階の変数に変換する。クロス集計の時は、あまりカテゴリーが細かいと表が読みにくいので、カテゴリー合併をすることが多い。

この例では、NENDAIが新変数名だということを、まず理解すること。

```
COMPUTE      NENDAI =Q46AGE.
RECODE       NENDAI (10 THRU 29 =2) (30 THRU 39 =3) (40 THRU 49 =4)
              (50 THRU 59 =5) (60 THRU 98 =6) .
```

例3 学歴を教育年数に変換する時は、以下のように書く。

旧制の教育制度については原・海野(2004)p. 204参照。

```
COMPUTE      EDU=Q43.          ←Q45など調査によって変える
RECODE       EDU (1=6) (2=9) (3=12) (4=13) (5=14) (6=16) (7, 9=99).
MISSING VALUES EDU(99) .
```

5.2. IF文の例

書式 IF (条件式) 条件があった場合の命令文 ピリオド

例4 問33（2015福島市調査 従業上の地位）をもとに、自営業ダミー変数を作成
問が6か7の場合に、自営業ダミー変数の値を1にする

```
COMPUTE      JIEI =0.
IF (Q33=6)   JIEI =1.
IF (Q33=7)   JIEI =1.
```

例5 本人の従業上地位（無職を含め6カテゴリー）作成シンタックス

まず新変数jobstatsはQ33と同じ内容とする。その後RECODE文でjobstatsの内容を変更している。

次にIF文で、q34従業先規模が大きい場合、q36で4課長相当の人を管理職にする。
次にIF文で、q34従業先規模が大きい場合、q36で5部長相当の人を管理職にする。
次にIF文で、q34従業先規模が小さい場合、q33で1経営者・役員を自営業にする。

```

Compute      jobstats= q33.
Recode      jobstats (1=1) (2=3) (3, 4, 5=4) (6, 7=5) (else=9).
IF ((q34>5) and (q36 = 4)) jobstats=2.
IF ((q34>5) and (q36 = 5)) jobstats=2.
IF ((q34<6) and (q33 = 1)) jobstats=5.
VALUE LABELS jobstats 1 '経営者' 2 '管理職' 3 '常時雇用' 4 '非正規雇用'
5 '自営・家族従業者' 9 '無職' .

```

5.3. 変数の計算 足し算やかけ算など、計算ができる。+ - * / を用いる。

例6 問2Bの変数の回答を逆転。N2Bが逆転した新変数だということを理解する。

```
Compute N2B=5-Q2B.
```

例7 問7Aと問7Bの回答内容を足して、合計得点の新変数NEW7を作成する。

```
Compute NEW7=Q7A+Q7B.
```

6. おぼえておく便利なこと

- ・シンタックスでは、大文字と小文字は区別されない
- ・シンタックスでは、半角space、改行、tabは区別されない
- ・ctrl + A で全シンタックスを選択。ctrl + Rで実行
マックOSの場合はコマンド+A, コマンド+R(All, Run)
- ・変数名は、SPSS画面のデータウィンドウで確認できる。

7. シンタックス実行でエラーが出たとき (村瀬他, 2007:46を参照)

以下の点に注意してシンタックスを書き直せばよい。エラーが出た付近のシンタックスに何らかのミスがある。

- ・データのあるドライブ名、フォルダ名はあっているか確認。CドライブをEと書いた等ドキュメントフォルダに入れた場合、正式なアドレスはかなり長い。テキストデータファイルを、Cドライブのzemiフォルダなど、分かりやすい場所に入れるとよい。
- ・変数名はあっているか。Q6AをQ6と書いた等
- ・単純なスペルのミス RECODEと書くべきところを RECORDと書いた等
- ・最後にピリオドをつけるのを忘れた。あるいはピリオドをつけすぎた。
- ・シンタックスの中に全角スペースがあると動かないので注意!

なお、データを開いてから、データ加工文などのシンタックスを実行すること。

どれが既存の変数名で、どれが新変数名なのか理解して書くことがこつ。既存変数名はデータ中に存在する名前を正確に書く

8. クロス集計とエラボレイション

以下のようなシンタックスを書くと、三重クロス集計を行うことができる。エラボレイションとは、第三変数の導入による、因果関係の分析である。この考え方について詳しくは原・海野(2004)を参照。最初は3つの関連係数を出し、次に、第三変数Zでデータを分割した後の、XY間の関連係数を2つ出すので、5つの関連係数を出すことになる

まずカテゴリー合併をして、新変数を3つ作ってから、以下のような命令文により分析するとよい。結果を、原・海野(2004:91)にあるような%の表にまとめること。

```
CRO
```

```
/TAB=N10 by N2A
```

```
/STA=CHISQ PHI CORR BTAU CTAU.
```

```
CROSSTABS
```

```
/TABLES=N10 by N2A by N13
```

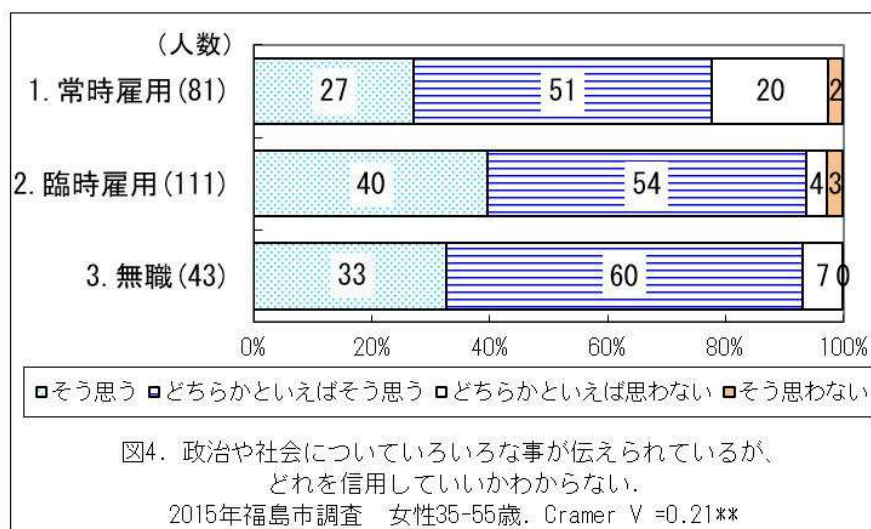
```
/STATISTICS=CHISQ PHI CORR BTAU CTAU
```

```
/CELLS=COLUMN.
```

CELLS行に =rowと書くと、縦ではなく横%が出る。どちらでも、分かりやすい方を使えばよい。Yとなる変数について%を出す。例えば、男女別のクロス集計ならば、男と女の中で、合計で100になる%を使う。STATISTICS行を書くと、各種の関連係数が出る。

9. 質的変数とクロス集計の関連係数

以下は、上記の例5で作った変数「従業上の地位」について、年齢を絞り女性のみとし「情報不信」の問とクロス集計をした結果である。「従業上の地位」カテゴリは人数が多いもの3つに絞った。



多いもの3つに絞った。

カテゴリに1, 2, 3と数字を付けたが、この数字には量的な意味はない。このようなカテゴリからなる変数を質的変数という。質的変数の関連係数は、2×2表(4セル)については、四分点相関係数rなど

を使う。この図は、回答が3カテゴリと4カテゴリだから、12セルある。質的変数(カテゴリ変数)で、2×3以上の場合はクラマーVを用いる。

10. 各種の関連係数について

量的変数(量的な意味がない)の関連を見るためにはピアソンの積率相関計数(r)などを使うが、2×2のクロス集計表(4セル)では、基本的に四分点相関係数rを使えば良い。完全関連でなく、最大関連の時は、Q係数を使う。rの値は、完全に関連がある時は±1か±1、無関連時は0になる(原・海野, 2004:85を参照)。

カイ二乗値とは、無関連時の表と、現実の表の2つを比べ、2つの距離を出した表を作ることができる。各セルにその距離を書いた新たな表を作り、その値を全て二乗してから、合計するとカイ二乗値となる。試しにやってみると良い。

変数に量的な意味がある場合、2×3セル以上の表では、タウbかタウcを用いる。タウb、タウcは、無関連が0、完全関連が1(あるいは-1)となる。

2×3以上の表の関連(詳しくは統計のテキストを見ること)

- ・変数に量的意味がない場合→ クラマーのV(無関連0、完全関連1)
- ・変数に量的意味がある場合→ τ_b (3×3など対象な表)、 τ_c (対象でない表)

11. ファイルの分割と結合

男女別の分析など、2回同じ分析結果を出したいときは、データファイルを開いた後に、SPSS画面上の「データ」をクリックして、ファイルの分割を選び、データを男女別等に分割すると良い。メニューが出るので、「グループごとの比較」などをクリックし、性別の変数(質問項目)を選びOKを押す。すると、データファイルが男女に並べ替えられる。

その後、分析を実行すると、男女別に分析結果が2つ表示される。

2つの調査結果など、データファイルを結合したい時は、SPSS画面上の「データ」をクリックして、ファイルの結合→ケースの追加を選ぶ。予め、2つのデータファイル内の変数名を同じにしておけば結合される。事前に同じ変数名にしておけば良い。

12. 重回帰分析など多変量解析をする時の注意点

どの変数が連続量で、どの変数がカテゴリーか注意する。量的な意味がある変数のみ、平均値を出したり、重回帰分析で用いることができる。

分析前に、使う変数の基本的な分布を見る。回答の偏りが大きい場合は注意。また、事前に欠損値処理をする。図2のようにMissing Valuesコマンドを書いて実行する。これを忘れていないか、分析時に基礎的な度数分布表などを見てよく確認することが重要。

そして、例6のように回答を逆転した新変数などを作り、変数の方向をそろえる。例7のように合計得点の新変数を作ってもよい。新変数を使ってから重回帰分析を行うと分かりやすい。重回帰のシンタックス例は以下の参考文献を参照。

13. SPSSでの文字化け

日本の調査データファイルは文字コードがシフトJISであることが多い。最近の分析ソフトは、文字コードがUTF-8(多言語対応のUNICODE)なので、変数ラベルが全角文字の場合は正常に表示されないことがある。SPSS「編集」→「設定」→「言語」で文字コードを変更するか、あるいはエディタソフトを使って、ラベルのシンタックスを書き、シンタックスをUTF-8で保存する。その後、SPSSでシンタックスを開いて実行すればよい。

参考文献

- 浅野正彦・矢内勇生. 2013. 『Stataによる計量政治学』オーム社.
足立浩平. 2006. 『多変量データ解析法 ―心理・教育・社会系のための入門』ナカニシヤ出版.
ボーンシュテット・ノーキ著＝海野道郎・中村隆監訳. 1990. 『社会統計学 ―社会調査のためのデータ分析入門』ハーベスト社.
原純輔・浅川達人. 2009. 『社会調査』放送大学教育振興会.
原純輔・海野道郎. 2004. 『社会調査演習 第2版』東京大学出版会.
林 拓也. 2012. 『社会統計学入門』放送大学教育振興会.
廣瀬毅士・寺島拓幸. 2010. 『社会調査のための統計データ分析』オーム社.
石村貞夫・石村光資郎. 2013. 『SPSSによる統計処理の手順』東京書籍.
岩井紀子・保田時男. 2007. 『調査データ分析の基礎 ―JGSSデータとオンライン集計の活用』有斐閣.
神林博史・三輪哲. 2011. 『社会調査のための統計学 ―生きた実例で理解する』技術評論社.
片瀬一男. 2007. 『社会統計学』放送大学教育振興会.
片瀬一男・阿部晃士・高橋征仁著. 2015. 『社会統計学ベシク』ミネルヴァ書房.
岸学. 2012. 『SPSSによるやさしい統計学 第2版』オーム社.
三輪哲・林雄亮編. 2014. 『SPSSによる応用多変量解析』オーム社.
村瀬洋一他編. 2007. 『SPSSによる多変量解析』オーム社.
小野寺孝義・山本嘉一郎編. 2014. 『SPSS事典 BASE編』ナカニシヤ出版.
盛山和夫. 2004. 『社会調査法入門』有斐閣.
杉野勇. 2017. 『入門・社会統計学：2ステップで基礎から[Rで]学ぶ』法律文化社.
轟亮・杉野勇編. 2017. 『入門・社会調査法〔第3版〕 ―2ステップで基礎から学ぶ』法律文化社.
土田昭司・山川栄樹. 2011. 『新・社会調査のためのデータ分析入門 ―実証科学への招待』有斐閣.
渡部洋. 1988. 『心理・教育のための多変量解析法入門 (基礎編)』福村出版.