

公的統計ミクロデータ用変換ツール

操作マニュアル

第 1.0 版
2017 年 11 月 20 日

一橋大学経済研究所

Copyright (C) 2017 一橋大学経済研究所

更新履歴

項目番号	版数	更新内容	更新日	更新者
1	1.0	初版制定	2017/11/20	

1. はじめに	1
1.1. 本書の位置づけ	1
1.2. 本ツールの説明	1
2. 前提条件・注意事項	2
2.1. 前提条件	2
2.2. 注意事項	2
3. 本ツールの操作手順	3
3.1. 本ツールの配置	3
3.2. 設定ファイルの編集	4
3.3. 本ツールの実行	5
4. 正常に動作していない場合の確認・対処方法	13
4.1. ログメッセージからの確認方法	13
5. ツールの処理時間がかかる場合の対処方法	17
6. (参考)本ツールの処理時間	26

表 1 フォルダ構成	3
表 2 設定ファイル項目	4
表 3 ツール画面項目	6
表 4 ツール出力ファイル	9
表 5 コンバートファイル項目	9
表 6 マシンスペック	26
表 7 ツール処理時間	26

1. はじめに

1.1. 本書の位置づけ

本書は、公的統計ミクロデータ用変換ツール（以下、「本ツール」という）を操作する手順について記載した資料です。

1.2. 本ツールの説明

本ツールは、提供されている符号表、個票データ、項目名一覧を取り込み、分析者が使用する統計解析ツール（SPSS、STATA、R、CSV）のファイル形式に変換することができます。

変換を行う際、取り込んだ符号表の内容に則り、文字列補足などを行い、文字列項目はダブルクオーテーションで囲みます。

また、符号表に記載されている項目の「型」が「1：数値」、かつ、個票データの該当項目に文字列が存在した場合、統計解析ツールは、個票データの該当項目を文字列として認識してしまいます。本ツールは、上記のような項目が存在した場合、統計解析ツールで数値として認識させるため、個票データの当該項目列から文字列を削除した内容を、個票データの各行の最終列尾に追加します。

2. 前提条件・注意事項

2.1. 前提条件

本ツールの操作説明の前提条件について、以下に記述します。

- 本書は、本ツールが動作するために必要となるソフトウェアがインストール済であること、OSS ライブラリのダウンロードが完了していることを前提に記述しています。（ソフトウェアのインストール、OSS ライブラリのダウンロード手順については、別途「インストールマニュアル」を参照ください。）
- 本書は、本ツール一式のダウンロードが完了していることを前提に記述しています。
- 本ツールは、以下の OS で動作することを前提としております。
OS : Windows 10 (64bit)

2.2. 注意事項

本ツールの操作における注意事項について、以下に記述します。

- 本書で記述するツール操作手順以外の方法で操作した場合、正常に動作しない可能性があります。（動作した場合でも動作保障の範囲外となります。）
- 入力ファイル、出力ファイルのファイルパスは、ドライブレター～ファイル拡張子までの「文字数」が 256 文字以内に収める必要があります。
ファイル名が長い、または、フォルダ階層が深い場合は、Windows OS の制約を受ける可能性があります。
- 個票データ、符号表、項目一覧(※1)が格納されているフォルダ名、またはファイル名や出力ファイル名には、「¥ / : * ? " < > |」の文字は使用しないで下さい。
- 出力ファイル名に使用できる全角文字は、JIS 第 1 水準(※2)の文字となります。
- コンバートデータファイル(※3)は、出力予定フォルダに、同じファイル名がある場合は、確認などなく上書きを行います。
- 項目一覧の変数名に、全角文字列を使用した場合は、文字化け等が起こってしまう可能性があるため、半角のみを使用して下さい。
- 項目一覧に記載する項目番号は、昇順となるように記載して下さい。
- 正常に変換処理が行われない可能性があるため、入力ファイルとなる個票データには、ダブルクオーテーションで囲まれたデータがないこととします。
- 符号表に記載されている内容に則っていない個票データを含む行は、読み飛ばし行として変換後データファイルには出力されません。その際、読み飛ばし行の内容をログに出力します。

※1：項目一覧は、項目番号、個票データの 1 行名に付与する変数名などが記載されたファイル。

※2：JIS 第 1 水準の文字は、下記 URL を参照して下さい。

http://jp.idec.com/temp_faq/PLC/faq_H0A010.pdf

※3：コンバートデータファイルは、符号表に記載されている項目の「型」が「1：数値」、かつ、個票データの該当項目に文字列が存在した場合に出力されるファイル。

3. 本ツールの操作手順

3.1. 本ツールの配置

(1) ダウンロードした本ツール一式を任意のフォルダに配置します。

フォルダ構成は以下とします。

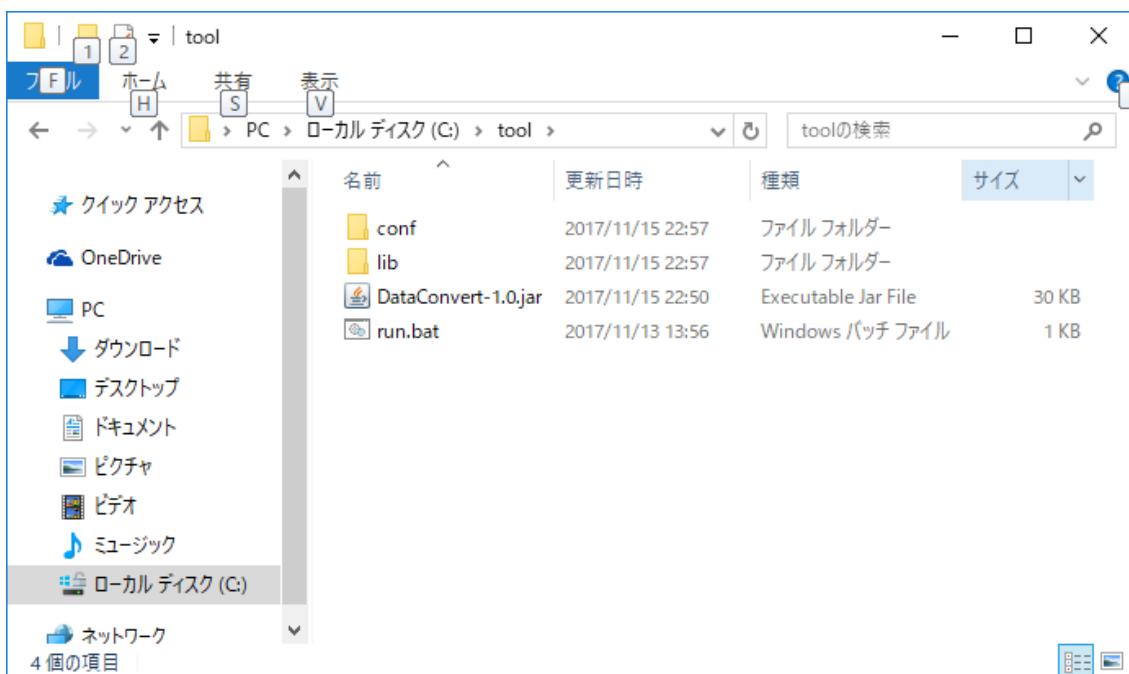


表 1 フォルダ構成

conf	本ツールで使用する設定ファイルを格納するフォルダ
lib	本ツールが利用する OSS ライブライアリを格納するフォルダ
DataConvert-1.0.jar	本ツールの実行ファイル
run.bat	本ツールの実行ファイルを呼び出すバッチファイル

(2) 「lib」フォルダ配下に、ダウンロードした下記 OSS ライブライアリを配置します。

- commons-codec-1.10.jar
- log4j-1.2.17.jar
- poi-3.14.jar
- poi-ooxml-3.14.jar
- poi-ooxml-schemas-3.14.jar
- stax-api-1.0.1.jar
- xmlbeans-2.6.0.jar

※上記 OSS ライブライアリは、「公的統計ミクロデータ用変換ツール_インストールマニュアル.docx」を参照し、ダウンロードしてください。

3.2. 設定ファイルの編集

(1) 使用する符号表の内容に合わせて「conf」配下の「setting.properties」の編集を行います。

設定ファイルの項目を以下となります。

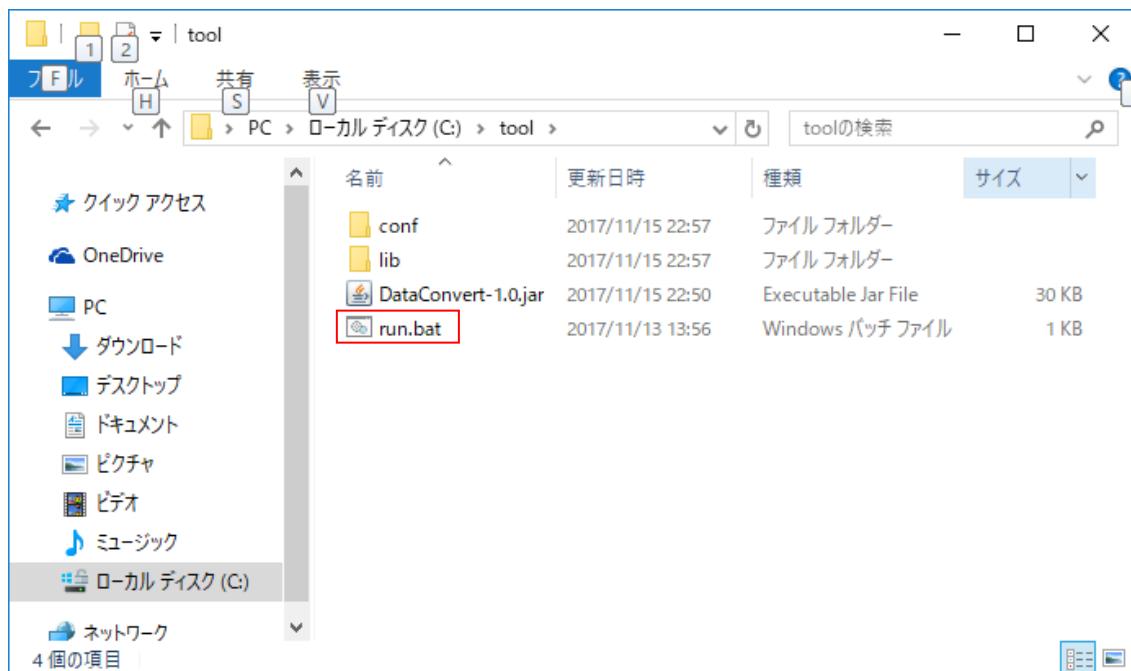
表 2 設定ファイル項目

項目名	説明
codeTable.lineNumber	符号表に記載されている「行番号」の列番号 (*1)
codeTable.hierarchy	符号表に記載されている「階層」の列番号 (*1)
codeTable.position	符号表に記載されている「位置」の列番号 (*1)
codeTable.itemNumber	符号表に記載されている「項目番号」の列番号 (*1)
codeTable.byte	符号表に記載されている「バイト数」の列番号 (*1)
codeTable.repetition	符号表に記載されている「繰返し」の列番号 (*1)
codeTable.arrangement	符号表に記載されている「配置」の列番号 (*1)
codeTable.type	符号表に記載されている「型」の列番号 (*1)

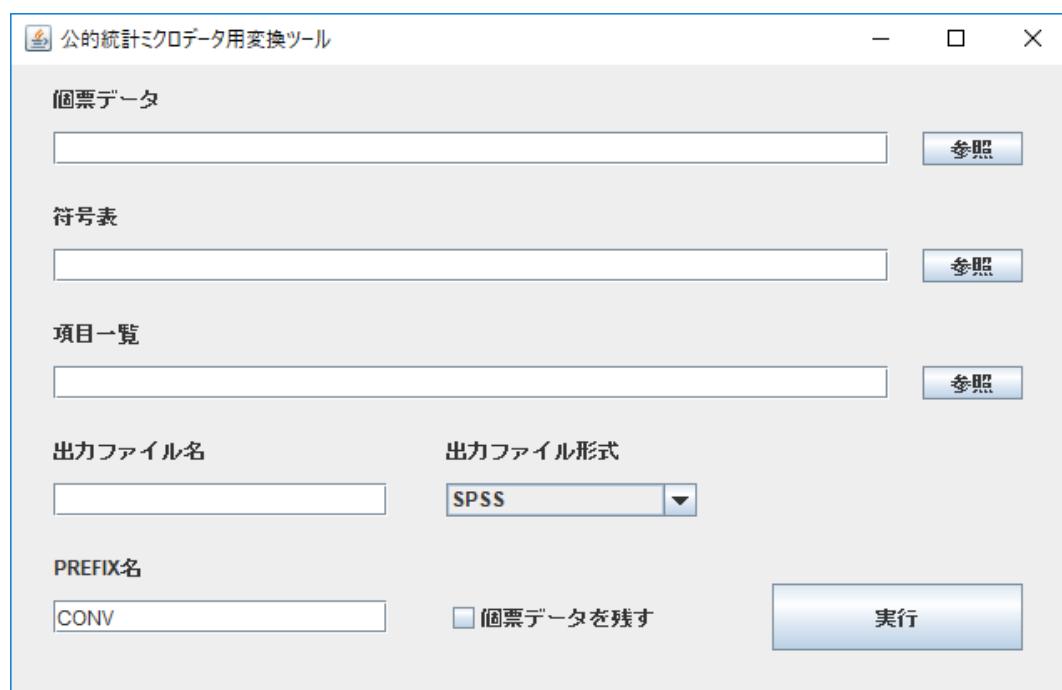
*1：列番号は、Excel の A 列であれば 1、B 列であれば 2、C 列であれば 3…となります。

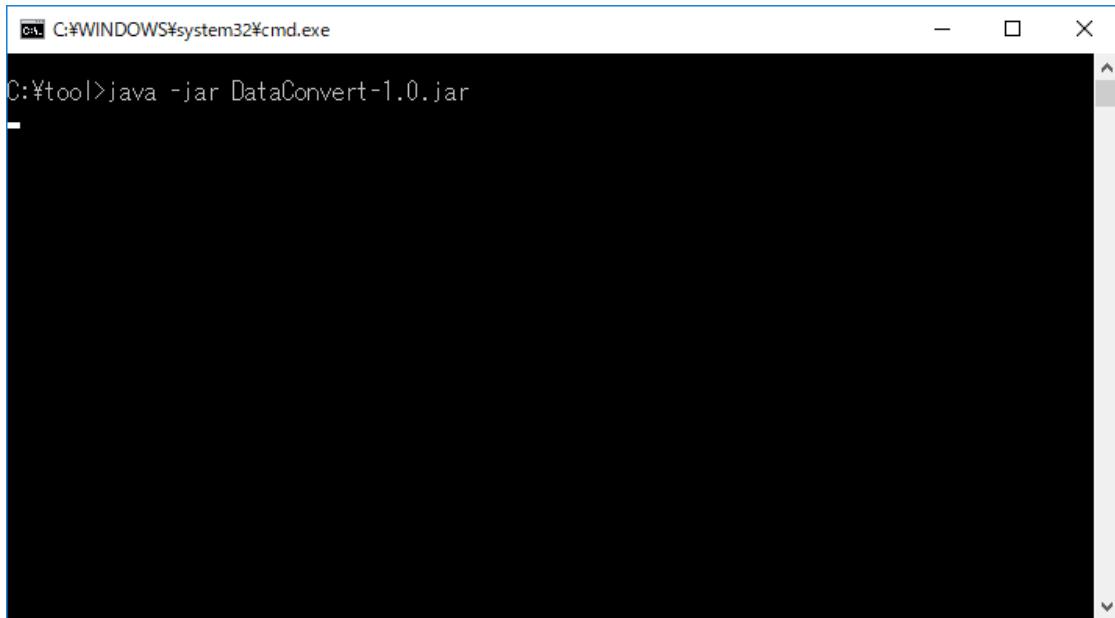
3.3. 本ツールの実行

(1) 3.1 (1)で配置した「run.bat」をダブルクリックし、本ツールを起動します。



ツールの画面とコマンドプロンプトが表示されます。



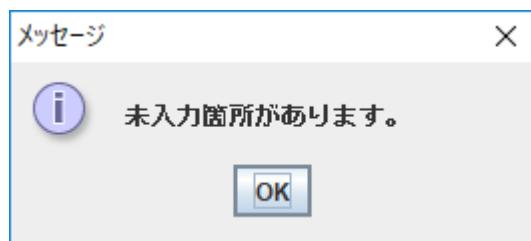


以下、画面の各項目について説明します。

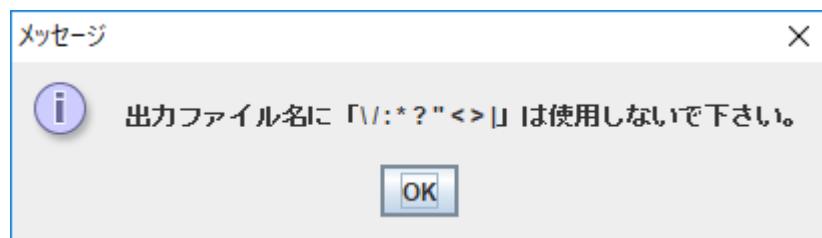
表 3 ツール画面項目

個票データ	データ変換を行う個票データを指定してください。 「参照」ボタンを押下し、ファイルを選択できます。
符号表	個票データに紐づく符号表のファイルを指定してください。 「参照」ボタンを押下し、ファイルを選択できます。
項目一覧	変換後の個票データの1行目に付与するヘッダ情報のファイルを指定してください。 「参照」ボタンを押下し、ファイルを選択できます。
出力ファイル名	変換後データのファイル名を設定してください。
出力ファイル形式	変換後データのファイル形式を選択してください。 デフォルトは「SPSS」となります。
PREFIX名	符号表の型が「1：数値」、かつ、当該項目の個票データに文字列が含まれる場合、個票データ最終列に追加する項目名の付与する内容を設定してください。 デフォルトは「CONV」となります。
個票データを残す	データ変換を行う際、変換元の個票データファイルを残す場合は、チェックを付けてください。チェックが付いていない場合は、変換元の個票データファイルは削除されます。 デフォルトは「✓」が付いていない状態となります。

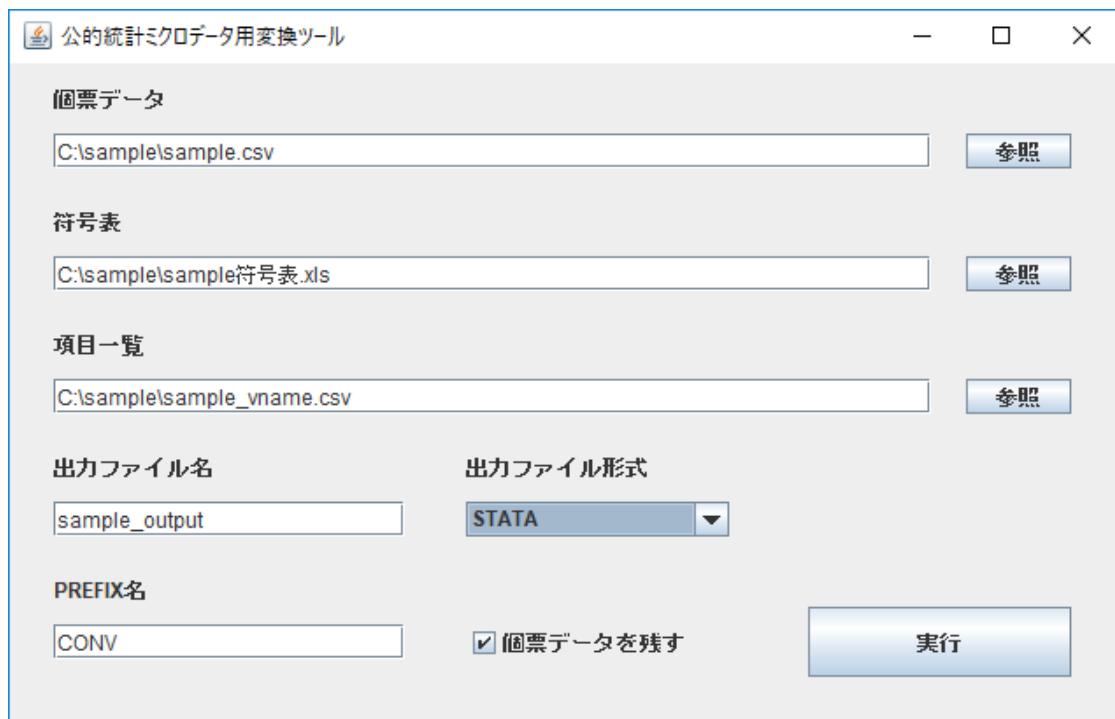
「個票データを残す」チェックボックス以外の項目はすべて必須項目となるため、未入力項目がある場合は、以下のようなメッセージが表示されます。



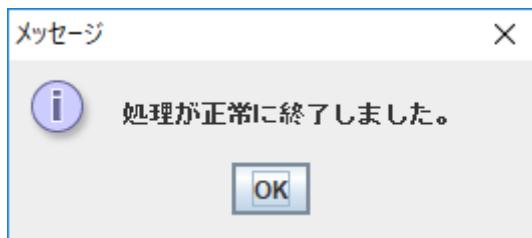
出力ファイル名に、「¥ / : * ? " < > |」のいずれかの文字が使用されている場合は、以下のようなメッセージが表示されます。



(2) 画面の各項目を入力し、「実行」ボタンを押下します。



(3) 正常に処理が終了した場合は、下記のように表示されます。



コマンドプロンプトにはツール実行時のログが表示されます。

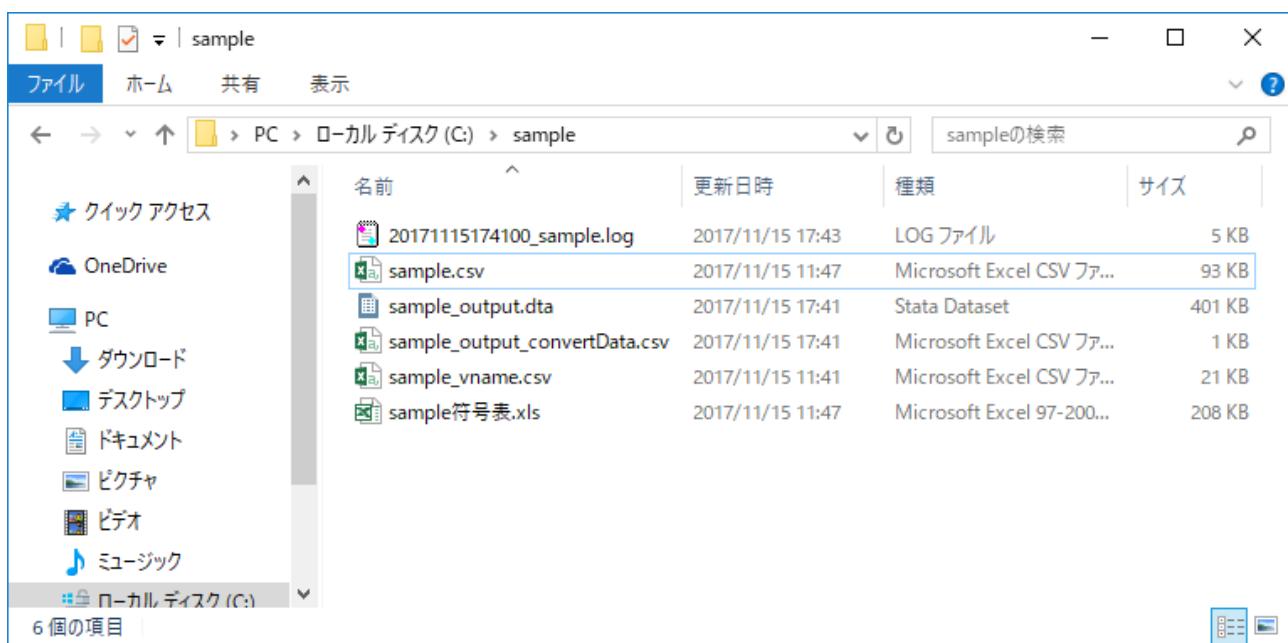
```
C:\WINDOWS\system32\cmd.exe
ConvertDataFile(PeiceDataManager.java:600) - コンバートデータファイル出力処理が終了しました。
2017-11-15 23:03:34,536 INFO jp.ac.hit_u.iер.dataconvert.manager.PeiceDataManager.readPeiceData(PeiceDataManager.java:389) - 変換元データファイルの作成が完了しました。
2017-11-15 23:03:34,536 INFO jp.ac.hit_u.iер.dataconvert.manager.PeiceDataManager.readPeiceData(PeiceDataManager.java:390) - 変換元データファイルパス : C:\sample\sample_output.csv
2017-11-15 23:03:34,537 INFO jp.ac.hit_u.iер.dataconvert.DataConvertMain.actionPerformed(DataConvertMain.java:332) - 個票データ読み込み処理が正常に終了しました。
2017-11-15 23:03:34,539 INFO jp.ac.hit_u.iер.dataconvert.DataConvertMain.actionPerformed(DataConvertMain.java:340) - データ変換処理を開始します。
2017-11-15 23:03:34,539 INFO jp.ac.hit_u.iер.dataconvert.manager.DataConvertManager.convertData(DataConvertManager.java:67) - 変換後ファイルパス : C:\sample\sample_output.dta
2017-11-15 23:03:34,540 INFO jp.ac.hit_u.iер.dataconvert.manager.DataConvertManager.convertData(DataConvertManager.java:74) - R データ変換コマンド : convert("C:/sample/sample_output.csv", "C:/sample/sample_output.dta")
2017-11-15 23:03:36,500 INFO jp.ac.hit_u.iер.dataconvert.DataConvertMain.actionPerformed(DataConvertMain.java:353) - データ変換処理が正常に終了しました。
2017-11-15 23:03:36,502 INFO jp.ac.hit_u.iер.dataconvert.DataConvertMain.actionPerformed(DataConvertMain.java:363) - 処理が正常に終了しました。
```

データ変換が正常に行えた場合、個票データと同じフォルダに変換後データファイルとログファイルが出力されます。

また、符号表の内容で「型」が「1：数値」の項目があり、個票データでその項目が文字列であった場合は、コンバートファイルが出力されます。

表 4 ツール出力ファイル

変換後データファイル	「出力ファイル名」、「出力ファイル形式」で指定された拡張子 例。 出力ファイル名 : sample_output 出力ファイル形式 : STATA (.dta) 変換後データファイル : sample_output.dta
ログファイル	「ツール実行時の日付時分秒」_「個票データファイル名」.log 例。 ツール実行日時 : 2017年11月15日 17時41分00秒 個票データファイル名 : sample ログファイル : 20171115174100_sample.log
コンバートファイル	「出力ファイル名」_convertData.csv 例。 出力ファイル名 : sample_output 変換後データファイル : sample_output_convertData.csv

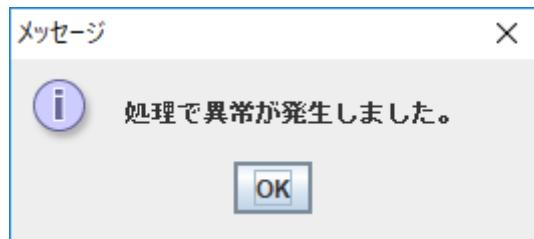


コンバートファイルは、以下内容が記載されます。

表 5 コンバートファイル項目

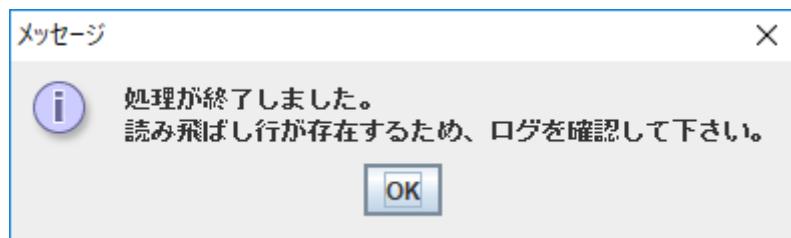
項目	説明
基データカラム名	追加基となったデータ列のカラム名
変換データカラム名	追加された変換データ列のカラム名
基データカラム位置	追加基となったデータ列のカラム列の位置
変換データカラム位置	追加された変換データ列のカラム列の位置

(4) 処理で異常が発生した場合は、下記のように表示されます。



各エラーメッセージが出力された場合の確認と対処方法は、「4.1 ログメッセージからの確認方法」を確認して下さい。

(5) 読み飛ばし行が発生した場合は、下記のように表示されます。



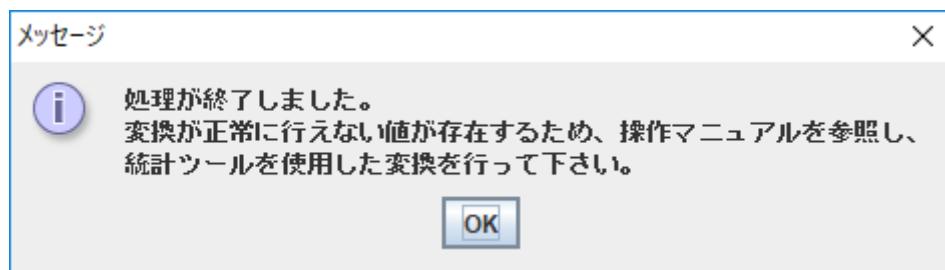
読み飛ばし行の詳細を確認するためには、個票データと同じフォルダに出力されたログファイルの内容を確認して下さい。

メッセージ内容は、以下のいずれかが出力されます。

- ・ XX 行目 : 項目数が符号表に記載されている繰返しを含む項目数と異なるため、読み飛ばします。個票データ : YY、符号表 : ZZ
※XX : 個票データの行数、YY : 個票データの1行あたりの項目数、ZZ : 符号表の1行あたりの項目数
- ・ XX 行目 : 、YY 項目目 : 全角半角文字が混在しているため、読み飛ばします。個票データ : ZZ
※XX : 個票データの行数、YY : 個票データの項目の位置、ZZ : 個票データの内容
- ・ XX 行目 : 、YY 項目目 : 全角文字チェックで NG のため、読み飛ばします。個票データ : ZZ
※XX : 個票データの行数、YY : 個票データの項目の位置、ZZ : 個票データの内容
- ・ XX 行目 : 、YY 項目目 : バイト数チェックで NG のため、読み飛ばします。個票データ : ZZ、符号表バイト数 : WW
※XX : 個票データの行数、YY : 個票データの項目の位置、ZZ : 個票データの内容、WW : 符号表に記載されているバイト数

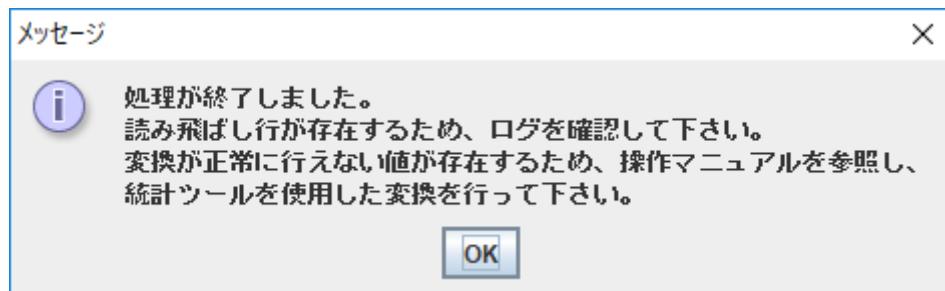
- XX 行目 :、YY 項目目 : バイト数チェックで NG のため、読み飛ばします。個票データ : ZZ、個票データ バイト数 : VV、符号表バイト数 : WW
※XX : 個票データの行数、YY : 個票データの項目の位置、ZZ : 個票データの内容、VV : 個票データ のバイト数、WW : 符号表に記載されているバイト数

(6) 個票データに、データ変換が正常に行えない値が存在した場合は、下記のように表示されます。



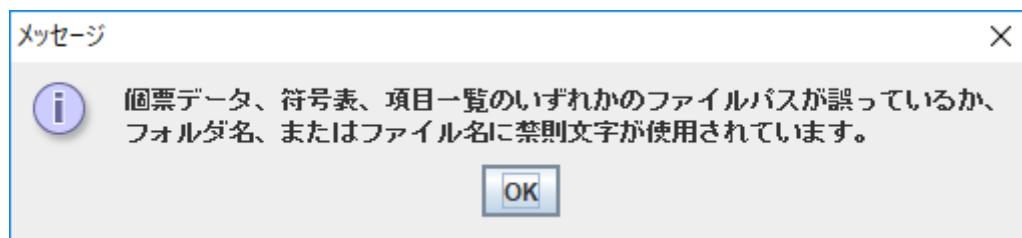
本メッセージが表示された場合は、個票データと同じフォルダに、「出力ファイル名」.csv ファイルが出力されているため、「5 ツールの処理時間がかかる場合の対処方法」の(3)以降を参照し、変換を行って下さい。

(7) 読み飛ばし行が発生した、かつ、個票データにデータ変換が正常に行えない値が存在した場合は、下記のように表示されます。



(5)、(6)に記載されている対処を行って下さい。

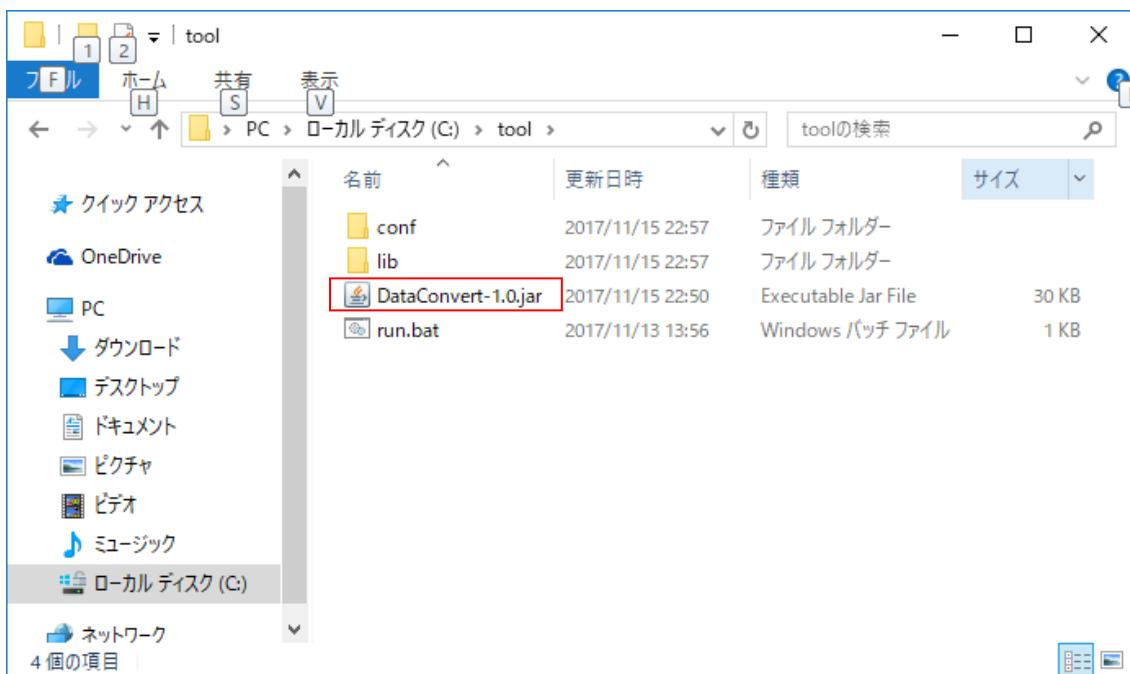
- (8) 個票データ、符号表、項目一覧のファイルパスが誤っている、または、フォルダ名、ファイル名に禁則文字（¥ / : * ? " < > |）が使用されている場合は、下記のように表示されます。



本メッセージが表示された場合は、個票データ、符号表、項目一覧のファイルパスを確認し、誤っている場合は修正して下さい。

フォルダ名、またはファイル名に禁則文字（¥ / : * ? " < > |）のいずれかが使用されている場合は、本ツールでは禁則文字を含むファイルパスは扱えないため、フォルダ名、ファイル名から禁則文字を削除して下さい。

- (9) コマンドプロンプトを表示させずにツールを起動する場合は、「DataConvert-1.0.jar」をダブルクリックします。
この場合、ツールの画面のみが表示されます。



4. 正常に動作していない場合の確認・対処方法

4.1. ログメッセージからの確認方法

異常が発生した場合は、個票データと同じフォルダに出力されたログファイルの内容を確認してください。

ログに出力されているエラー内容によって、エラー原因とその対処方法が異なります。

(1) メッセージ内容：符号表読み込みに失敗しました。

原因	対処方法
指定した符号表のファイルパス誤り。	ツール画面の「符号表」で指定したファイルパスの内容が正しいか確認してください。
指定した符号表のファイルが存在しない。	ツール画面の「符号表」で指定したファイルが存在するか確認してください。

(2) メッセージ内容：項目一覧ファイルの読み込みに失敗しました。

原因	対処方法
指定した項目一覧のファイルパス誤り。	ツール画面の「項目一覧」で指定したファイルパスの内容が正しいか確認してください。
指定した項目一覧のファイルが存在しない。	ツール画面の「項目一覧」で指定したファイルが存在するか確認してください。

(3) メッセージ内容：個票データファイルの操作に失敗しました。

原因	対処方法
指定した個票データのファイルパス誤り。	ツール画面の「個票データ」で指定したファイルパスの内容が正しいか確認してください。
指定した個票データのファイルが存在しない。	ツール画面の「個票データ」で指定したファイルが存在するか確認してください。

(4) メッセージ内容：拡張子が「xls」「xlsx」以外の符号表が指定されています。

原因	対処方法
指定した符号表のファイル拡張子が「xls」「xlsx」以外であった。	指定する符号表のファイル拡張子は「xls」または「xlsx」にしてください。

- (5) メッセージ内容：符号表の項目数と項目一覧の項目数が一致しません。符号表項目数：XXX、項目一覧項目数：YYY
※XXX：符号表の項目数、YYY：項目一覧の項目数

原因	対処方法
指定した符号表の繰返しを含む項目数と項目一覧の項目数が一致していない。	指定する符号表の繰返しを含む項目数と項目一覧の項目数を一致させて下さい。 ※抽象項目の行は除きます。

- (6) メッセージ内容：XXX 行目の項目番号に数値以外の値が入っています。内容：YYY
※XXX：符号表の行数、YYY：値の内容

原因	対処方法
指定した符号表の XXX 行目の「項目番号」に数値以外が入っている。	指定する符号表の「項目番号」には、数値のみが入っていることを確認してください。 ※抽象項目の行は除きます。

- (7) メッセージ内容：XXX 行目の階層に数値以外の値が入っています。内容：YYY
※XXX：符号表の行数、YYY：値の内容

原因	対処方法
指定した符号表の XXX 行目の「階層」に数値以外が入っている。	指定する符号表の「階層」には、数値のみが入っていることを確認してください。 ※抽象項目の行は除きます。

- (8) メッセージ内容：XXX 行目のバイト数に数値以外の値が入っています。内容：YYY
※XXX：符号表の行数、YYY：値の内容

原因	対処方法
指定した符号表の XXX 行目の「バイト数」に数値以外が入っている。	指定する符号表の「バイト数」には、数値のみが入っていることを確認してください。 ※抽象項目の行は除きます。

- (9) メッセージ内容：XXX 行目の繰返しに数値以外の値が入っています。内容：YYY
※XXX：符号表の行数、YYY：値の内容

原因	対処方法
指定した符号表の XXX 行目の「繰返し」に数値以外が入っている。	指定する符号表の「繰返し」には、数値のみが入っていることを確認してください。 ※抽象項目の行は除きます。

(10) メッセージ内容：XXX 行目の配置に数値以外の値が入っています。内容：YYY

※XXX：符号表の行数、YYY：値の内容

原因	対処方法
指定した符号表の XXX 行目の「配置」に数値以外が入っている。	指定する符号表の「配置」には、数値のみが入っていることを確認してください。 ※抽象項目の行は除きます。

(11) メッセージ内容：XXX 行目の配置に 1～3 以外の数値が入っています。内容：YYY

※XXX：符号表の行数、YYY：値の内容

原因	対処方法
指定した符号表の XXX 行目の「配置」に 1～3 以外の数値が入っている。	指定する符号表の「配置」には、1～3 が入っていることを確認してください。 ※抽象項目の行は除きます。

(12) メッセージ内容：XXX 行目の型に 1～4、空白以外の値が入っています。内容：YYY

※XXX：符号表の行数、YYY：値の内容

原因	対処方法
指定した符号表の XXX 行目の「型」に 1～4、空白以外が入っている。	指定する符号表の「型」には、1～4、空白が入っていることを確認してください。

(13) メッセージ内容：項目番号：「XXXX」が重複しています。

※XXXX：項目一覧の項目番号の内容（重複が複数ある場合は、複数出力されます。）

原因	対処方法
指定した項目一覧の項目番号「XXXX」が重複して指定されている。	指定する項目一覧の項目番号は、項目一覧ファイル内で重複しない項目番号を指定してください。

(14) メッセージ内容：変数名：「XXXX」が重複しています。

※XXXX：項目一覧の変数名の内容（重複が複数ある場合は、複数出力されます。）

原因	対処方法
指定した項目一覧の変数名「XXXX」が重複して指定されている。	指定する項目一覧の変数名は、項目一覧ファイル内で重複しない変数名を指定してください。

(15) メッセージ内容：R スクリプトファイルの作成に失敗しました。

原因	対処方法
端末に R がインストールされていない。	「インストールマニュアル」を参照し、端末に R をインストールしてください。
個票データを格納しているディスクサイズの空き容量がないため、R スクリプトファイルを作成できない。	不要なファイルを削除し、空き容量を増やして下さい。

(16) メッセージ内容：R スクリプトファイル内に使用禁止文字が存在しました。

原因	対処方法
出力ファイル名に、JIS 第 1 水準に記載されていない文字が指定された。	出力ファイル名は、半角文字、または JIS 第 1 水準の文字を指定して下さい。

(17) メッセージ内容：変換後データファイルが存在しません。出力ファイルパス：XXX

※XXX：変換後データファイルのパス

原因	対処方法
端末に R の rio パッケージがインストールされていない。	「インストールマニュアル」を参照し、端末に R の rio パッケージをインストールしてください。
端末のシステム環境変数の Path に R が通っていない。	端末のシステム環境変数の Path に R の bin 配下のパスを設定してください。

5. ツールの処理時間がかかる場合の対処方法

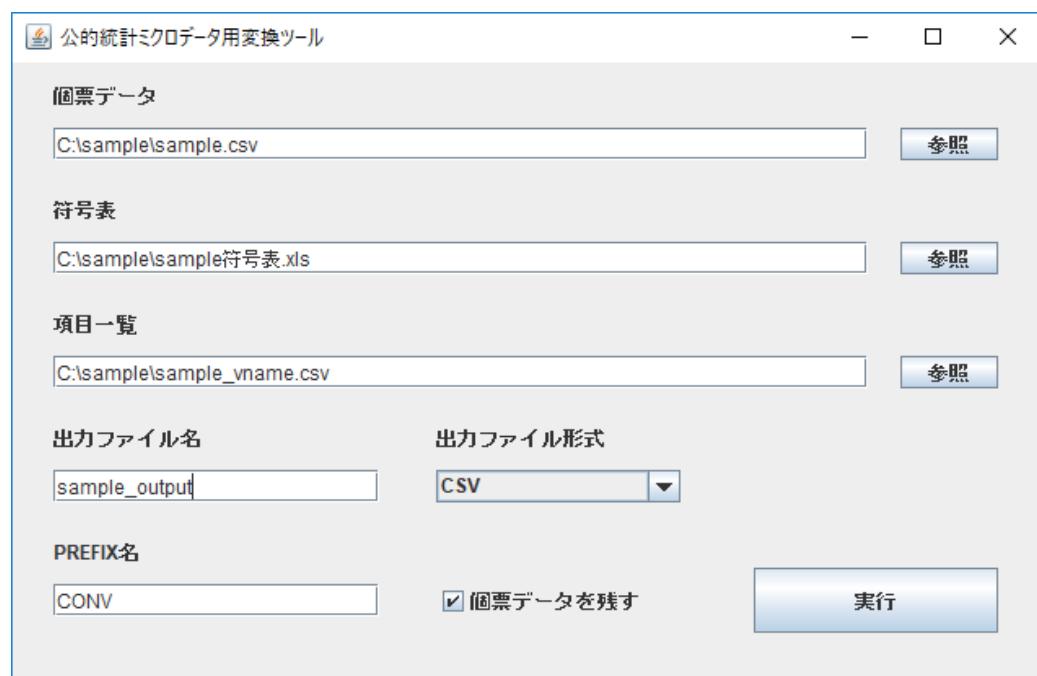
個票データのファイルサイズが 100MByte 前後の場合、本ツールの処理時間は約 5 分程度となります。

そのため、個票データのファイルサイズが 100Mbyte を超える場合、本ツールでの処理時間は、5 分以上かかる可能性があります。

本ツールと統計ツールを併用することで、データ変換時間を短縮することが可能です。

以下、データ変換手順を記載します。

- (1) 本ツールを使用し、「CSV」ファイルに変換する。

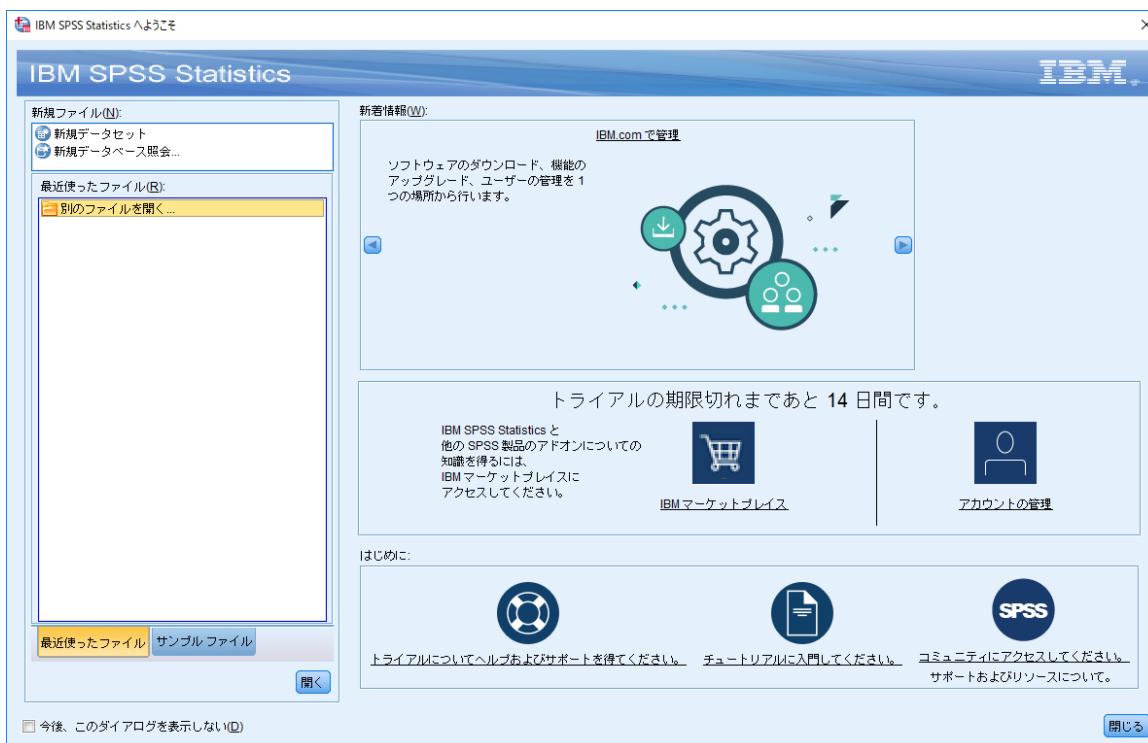


- (2) 個票データと同じフォルダに、出力ファイル名で指定したファイル名の CSV が作成されます。

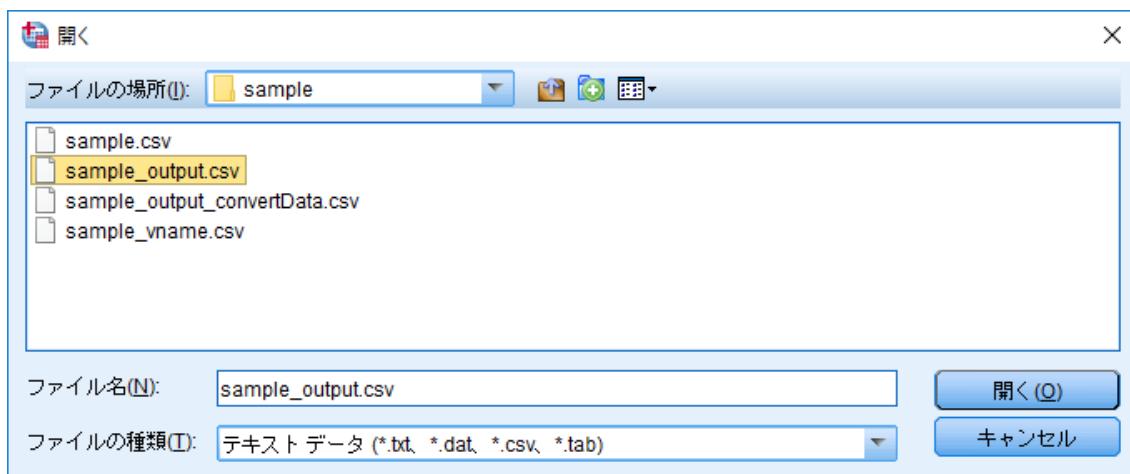
(3) 変換したい形式の統計ツールを起動し、CSV ファイル読み込み～ファイル保存までを行います。

➤ SPSS の場合

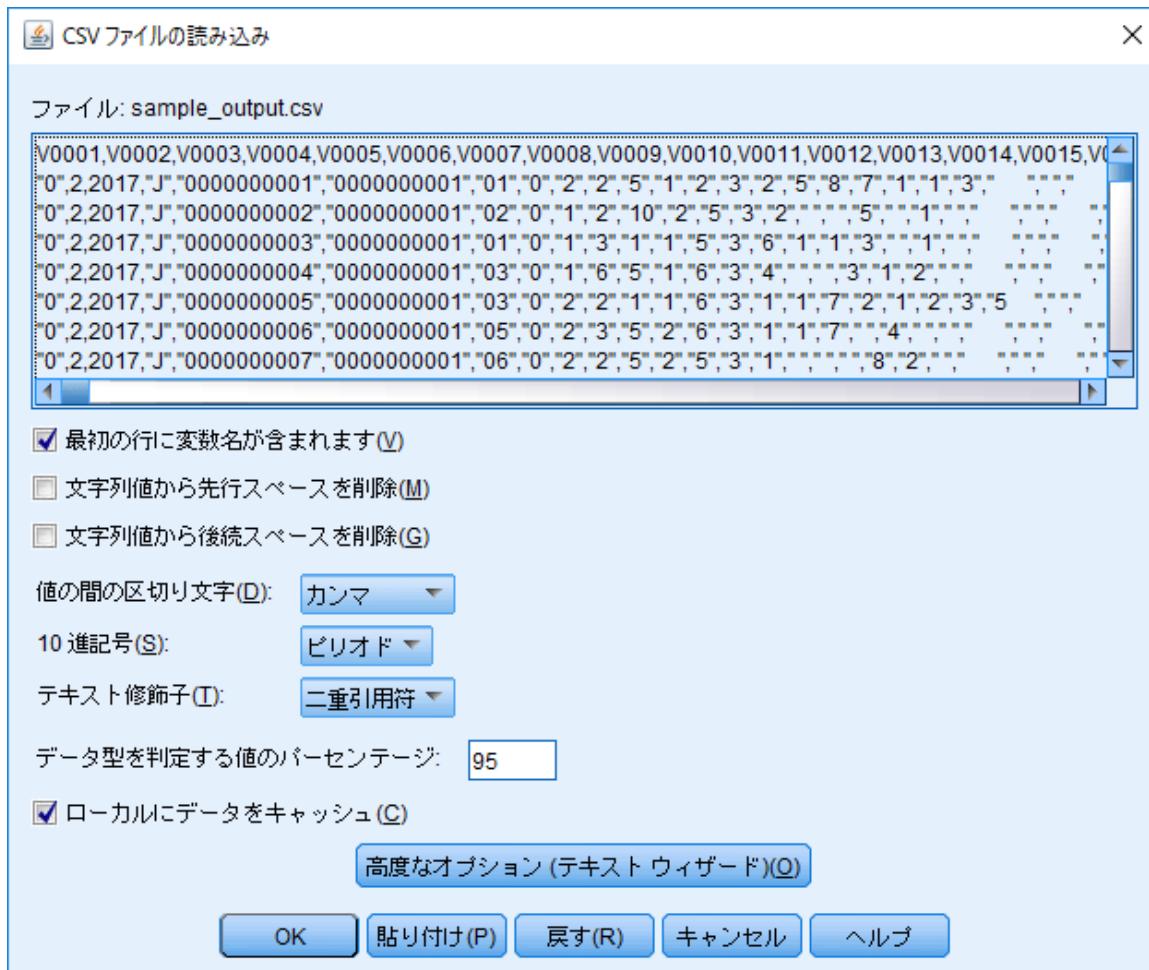
- ① 「別のファイルを開く」をダブルクリック。



- ② 「ファイルの種類」は「テキストデータ(*.txt, *.dat, *.csv, *.tab)」を選択し、(2)で出力した CSV ファイルを選択し、「開く」を押下します。



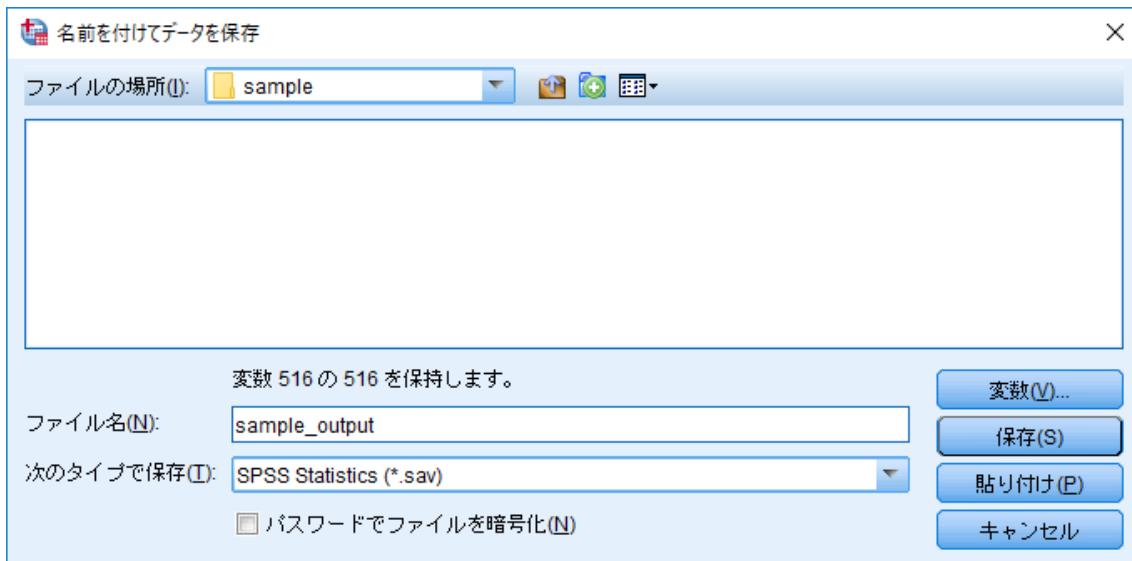
③ 「CSV ファイルの読み込み」画面が表示されるので、「OK」を押下します。



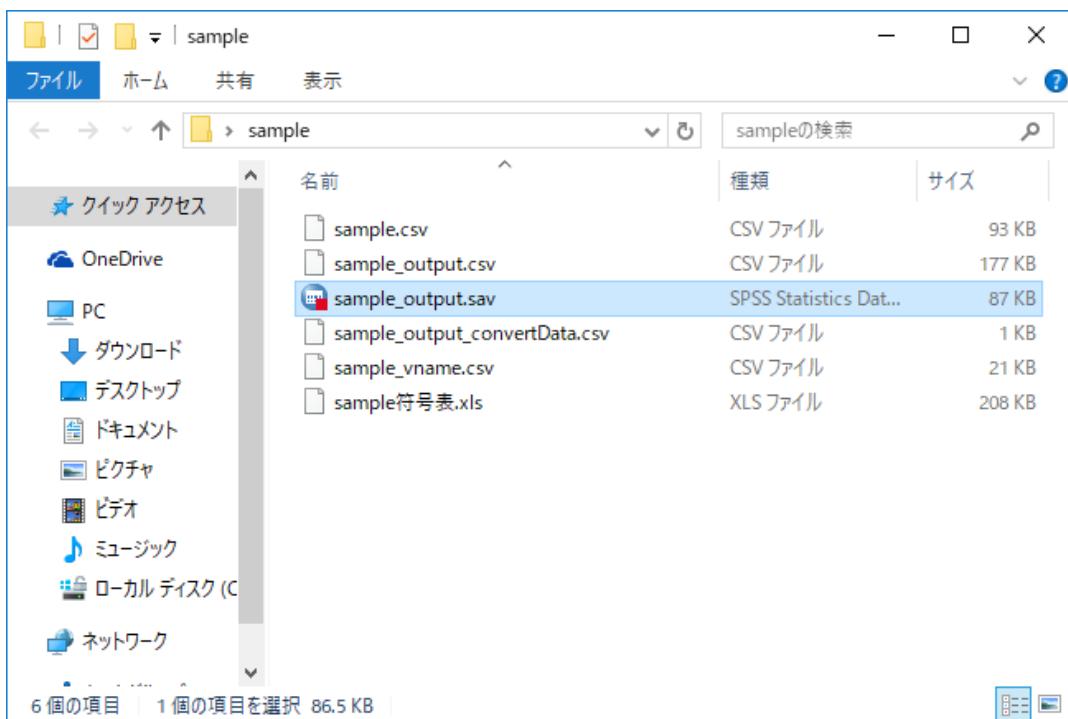
④ SPSS で CSV ファイルが読み込まれます。

	V0001	V0002	V0003	V0004	V0005	V0006	V0007	V0008	V0009	V0010
1	0	2	2017	J		1	1	1	0	2
2	0	2	2017	J		2	1	2	0	1
3	0	2	2017	J		3	1	1	0	1
4	0	2	2017	J		4	1	3	0	1
5	0	2	2017	J		5	1	3	0	2
6	0	2	2017	J		6	1	5	0	2
7	0	2	2017	J		7	1	6	0	2
8	0	2	2017	J		8	1	2	0	1
9	0	2	2017	J		9	1	1	1	1
10	0	2	2017	J		10	1	2	1	2
11	0	2	2017	J		11	1	2	1	2

⑤ 「ファイル - すべてのデータを保存」を選択し、ファイル名を入力して保存します。

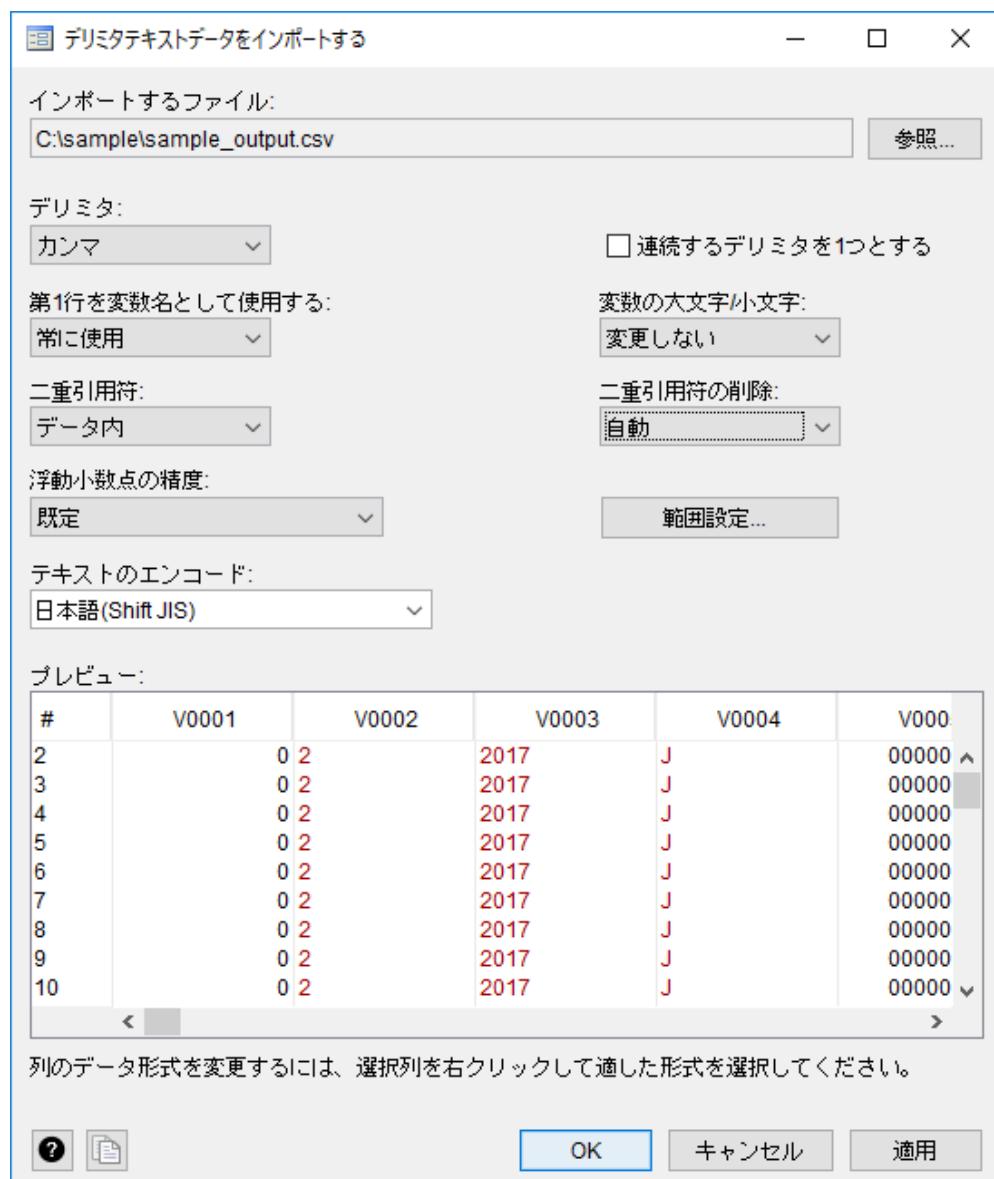


⑥ 拡張子に「*.sav」が付いたファイルが作成されます。

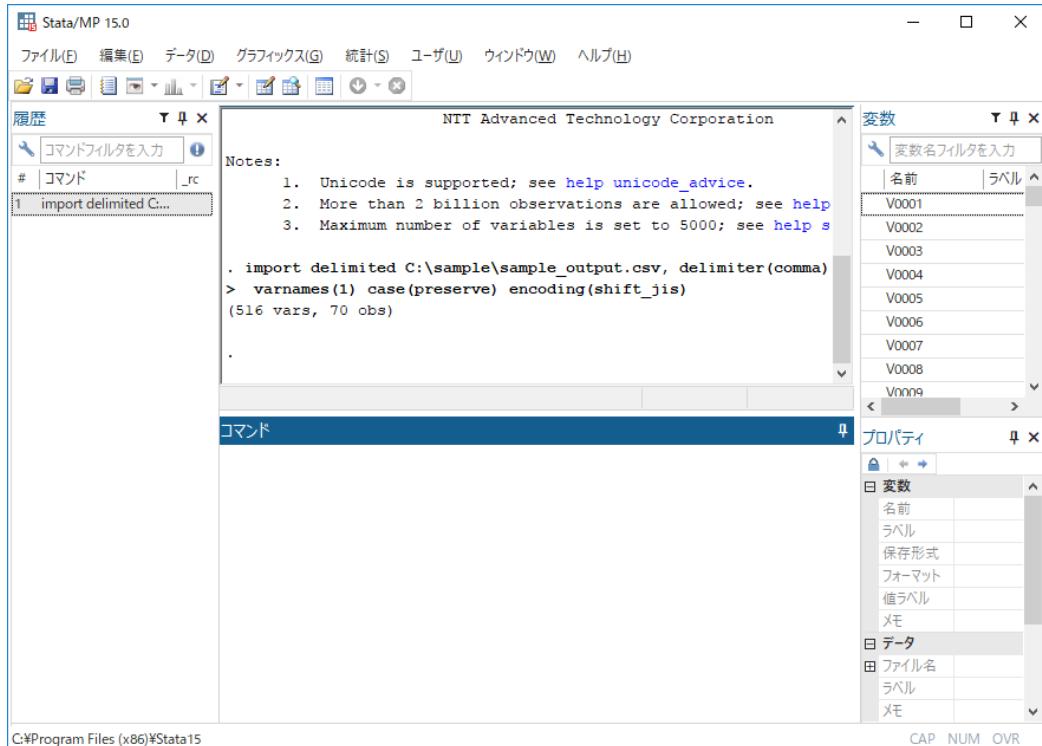


➤ STATA の場合

- ① 「ファイル - インポート - テキストデータ(デリミタ、.csv 等)を選択します。
 - ② 「デミリタテキストデータをインポートする」画面が表示されるので、「インポートするファイル」に(2)で出力した CSV ファイルを指定し、下記項目の指定を行い、「OK」を押下します。
 - ・ 「デミリタ」：カンマ
 - ・ 「第 1 行を変数名として使用する」：常に使用
 - ・ 「変数の大文字/小文字」：変更しない
 - ・ 「テキストのエンコード」：日本語(Shift JIS)
- ※上記以外の項目はデフォルトのまま



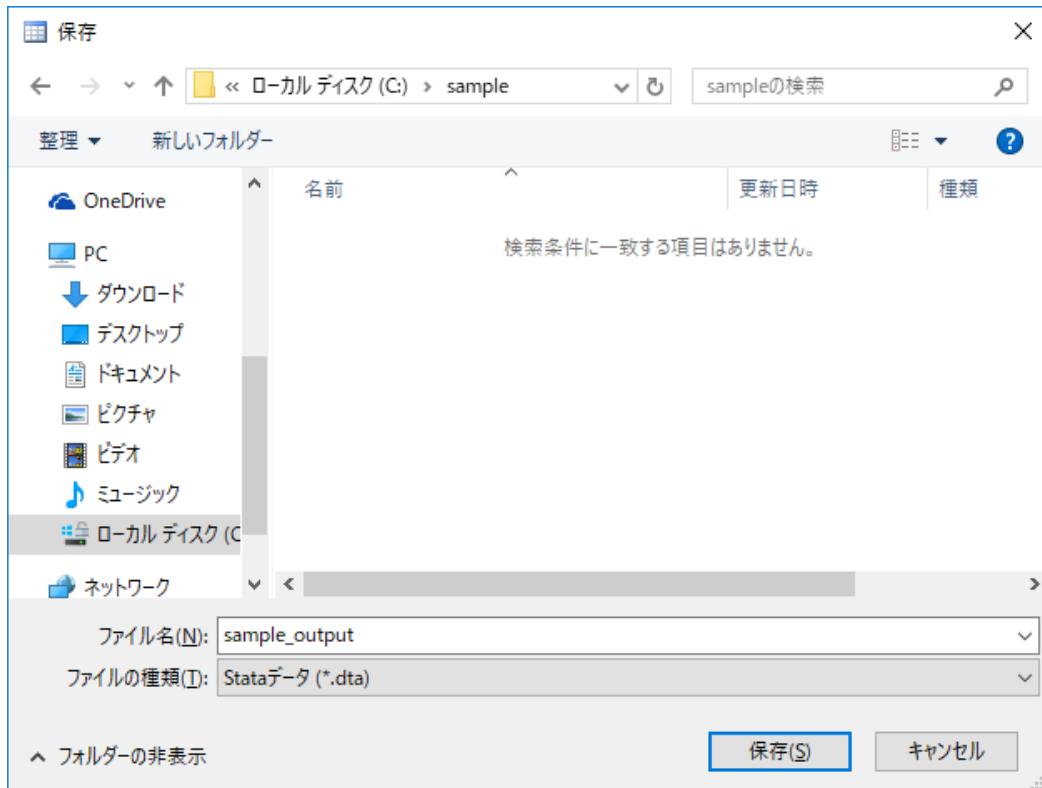
- ③ 正常に読み込まれたら、「データ - データエディタ - データエディタ(編集)」を選択し、読み込んだデータを表示します。



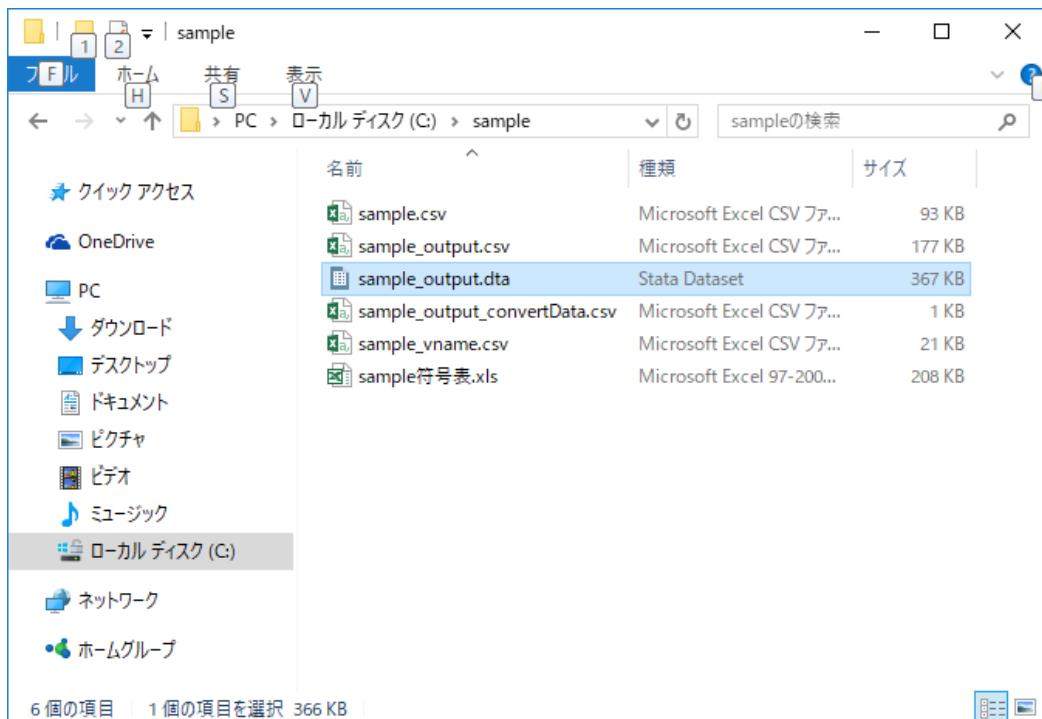
データエディタ(編集) - [無題]						
V0001[1] 0						
	V0001	V0002	V0003	V0004	V0005	V0006
1	0	2	2017	J	1	1
2	0	2	2017	J	2	1
3	0	2	2017	J	3	1
4	0	2	2017	J	4	1
5	0	2	2017	J	5	1
6	0	2	2017	J	6	1
7	0	2	2017	J	7	1
8	0	2	2017	J	8	1
9	0	2	2017	J	9	1
10	0	2	2017	J	10	1
11	0	2	2017	J	11	1
12	0	2	2017	J	12	1
13	0	2	2017	J	13	1
14	0	2	2017	J	14	1
15	0	2	2017	J	15	1
16	0	2	2017	J	16	1

準備完了 変数: 516 列順: データセット 観測値: 70 フィルタ: オフ モード: 編集 CAP NUM

④ 「ファイル - 名前を付けて保存…(A)」を選択し、ファイル名を入力し「*.dta」形式で保存します。



⑤ 拡張子に「*.dta」が付いたファイルが作成されます。



➤ R の場合

- ① (2)で出力した CSV ファイルを読み込みます。

※読み込んだ CSV ファイルデータを格納する変数名は、任意です。

R version 3.4.2 (2017-09-28) -- "Short Summer"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

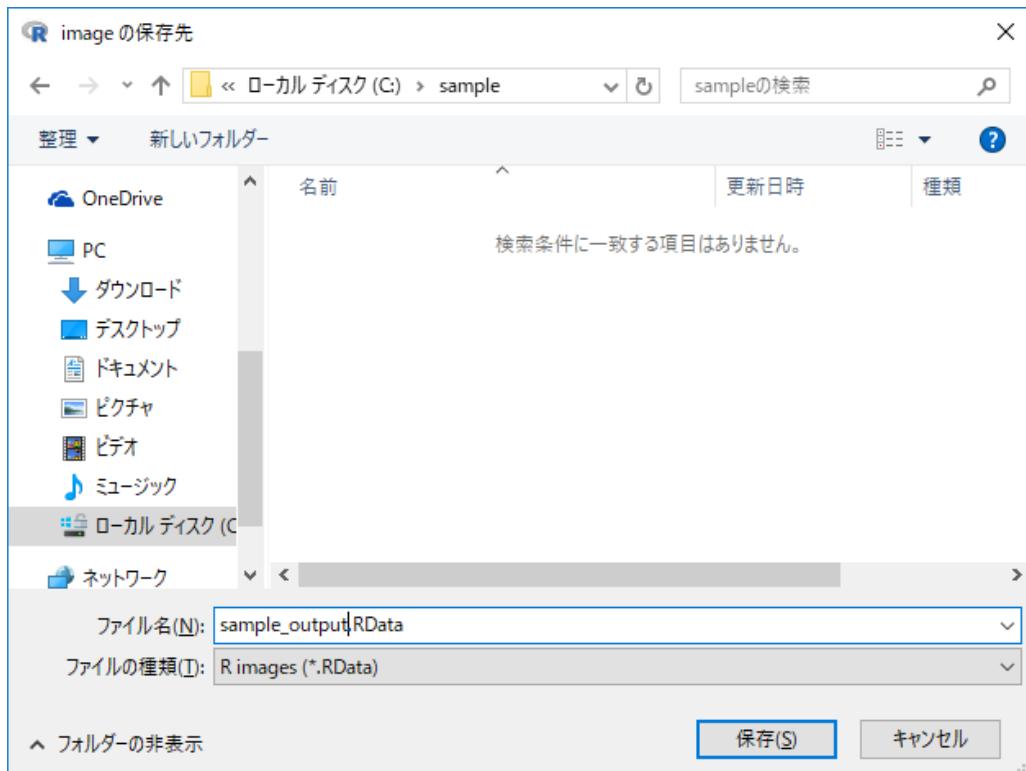
R は、自由なソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、'license()' あるいは 'licence()' と入力してください。

R は多くの貢献者による共同プロジェクトです。
詳しくは 'contributors()' と入力してください。
また、R や R のパッケージを出版物で引用する際の形式については
'citation()' と入力してください。

'demo()' と入力すればデモをみることができます。
'help()' とすればオンラインヘルプが出ます。
'help.start()' で HTML ブラウザによるヘルプがみられます。
'q()' と入力すれば R を終了します。

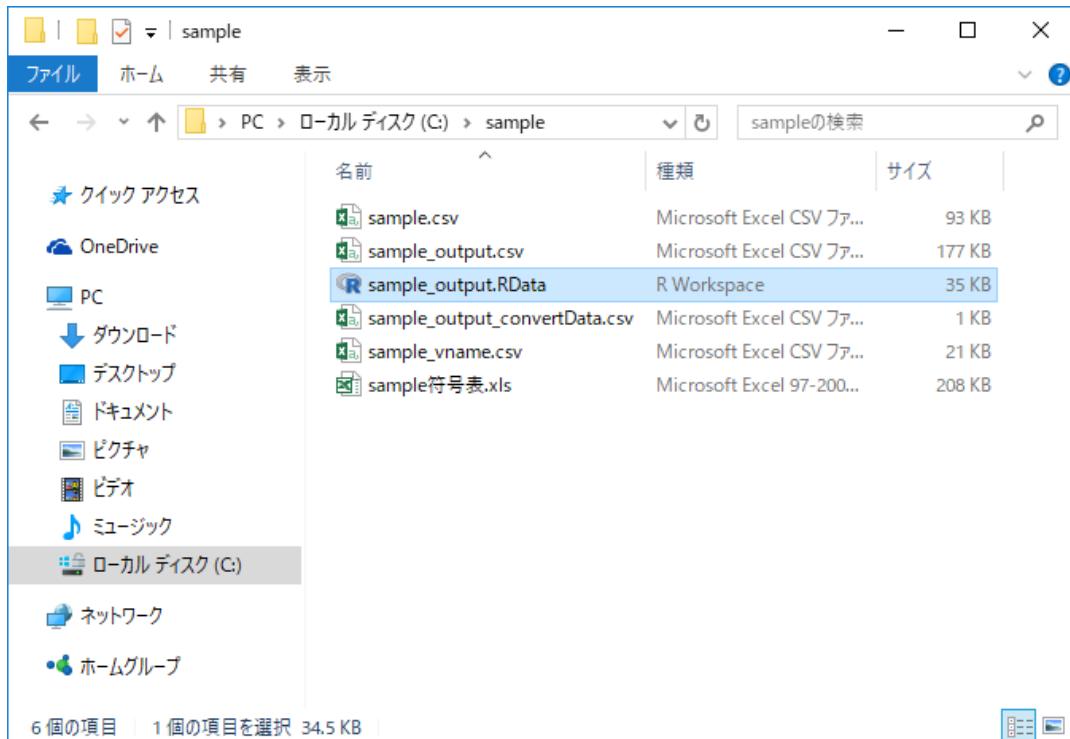
```
> sample <- read.csv("C:\\sample\\sample_output.csv")
```

- ② 「ファイル - 作業スペースの保存…」を選択し、ファイル名を入力して保存します。



③ 拡張子に「*.Rdata」が付いたファイルが作成されます。

※本ツールは、「*.R(R スクリプト)」ファイルが作成されますが、本手順では「*.Rdata(作業スペース)」ファイルの作成となります。



6. (参考) 本ツールの処理時間

本ツールは、指定された出力ファイル形式（CSV を除く）への変換は、R を使用して行っており、データ変換の主な処理時間は、R を使用した変換部分が占めています。

サンプルデータを使用して計測を行った PC のマシンスペックは以下となります。

表 6 マシンスペック

OS	Windows 10 64bit
プロセッサ	Intel(R) Core(TM) i3-2350M CPU @ 2.30GHz 2.30 GHz
コア数	2 コア
論理プロセッサ数	4
メモリ	8.0 GB
コア数	2 コア

以下、サンプルデータを用いた場合の処理時間を記載しますので、ツール実行時の処理完了時間の参考にしてください。

表 7 ツール処理時間

個票データレコード数	1 レコードあたりの項目数	個票データファイルサイズ (KB)	処理時間
500,000	510	825,984	55 分 48 秒
1,000,000	255	826,609	1 時間 21 分 28 秒
1,250,000	255	1,033,396	2 時間 3 分 39 秒