

## 統計学(基礎)

第2回  
データの準備と出力の利用

1/35

## JASPで扱えるデータについて

2/35

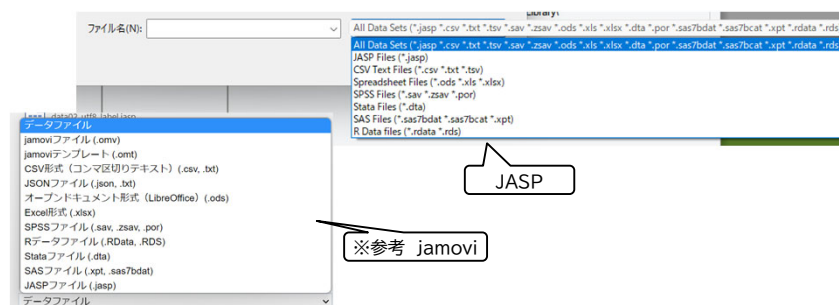
## データライブラリの場合

- データの場所
- C:\Program Files (x86)\JASP\resources\Data Sets\Data Library
- データ単体はCSV形式
- JASPで保存するとJASP形式になる
  - 出力や設定とセットのファイルになる

3/35

## JASPで読み込めるデータ形式

- 一応いろいろ読める



4/35

## 実際に読んでみると

- CSV  
- (Shift-JIS)
- 左から  
Excel  
JASP  
jamovi
- タイトルの日本語が読めない

5/35

## 実際に読んでみると

- CSV UTF-8
- ほぼ問題なし

6/35

## 実際に読んでみると

- Excel
- JASPはきついかも  
(Macだと問題ないことも多い)

7/35

## データに日本語がある場合

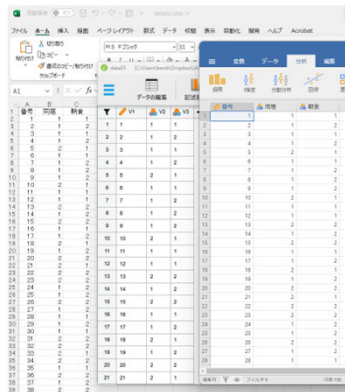
- CSV
- JASPIはアウト

8/35



## タイトルだけ日本語で、データが整数の場合

- Excel
- JASPIはタイトルが置き換わる



13/35

## データ入力形式

- ほぼ完璧なのはcsvのUTF-8
  - 普段はExcel形式で管理
  - JASPIに渡すときにCSV(UTF-8)にする
  - CSVファイルは複数シートを持ってないので注意
- 他のアプリケーションを使う場合も問題ない
  - jamoviiに渡しても問題ない
- Mac版のExcelはUTF-8が原則だけど、Windows版はそうでもないのが微妙なときがある

14/35

## ムダ知識

- アプリがデータファイルを読み込むとき
  - CSVファイルはテキストファイルなので、文字コード情報は原則持っていない
    - BOM (Byte Order Mark)を振ってあって、EF BB BFならUTF-8、FF FE(FF FF)ならUTF-16と判断するアプリもある
  - 全体から「文字コードはこれかな?」と推測している
  - 1行目はデータ名として判断する
    - 1行1列目から始まっていないとコケる
    - セル結合があるとコケる
    - 空白が入っていると空欄の変数を作る

15/35

## ムダ知識

- アプリがデータファイルを読み込むとき
  - 2行目以降からデータ本体と判断する
  - 2行目のデータを見て、その変数が文字なのか、数値なのか判断する。文字と整数なら名義か順序、小数点データならスケール(連続量)と判断する
    - いきなり欠損値だと、うまく行かないアプリがある
  - JASPIは「列の種類」(名義、順序、スケール)がデータ情報 jamoviiは「測定値型」(名義、順序、連続、ID)と「データ型」(整数、小数、文字)の2種類で情報を持つ

16/35

## 名義データをどう入力するか

- 文字でそのまま入力する方がわかりやすい
- 数値で入れた方がトラブルは少ない
  - 日本語でもExcelファイルの内容を結構渡してくれる
  - トラブルが起きる原因のほとんどはこれ

17/35

## 個人的な懸念

- アプリでの日本語の受け渡しはトラブルが多い
  - 昔は日本語が使えなかった
  - 文字コードセットの設定が複数ある(Shift-JIS UTF-8など)
  - 海外のソフトは、日本語の扱いに無頓着
    - 日本語話者が少ないから
    - 学術論文は英語がデファクトスタンダード

18/35

## 結論

- データ形式はCSVのUTF-8形式が最強
  - Webでも、WindowsでもMacでも大丈夫
- 名義データは、日本語で入れてもまあ大丈夫かな
  - 古いデータを持ってくる場合にトラブルになることがある
  - 古いテキストデータは、BOMが無いことが多いので、文字コード判定に難があることがある

19/35

面倒くさいけど便利かもしれない

## データラベル

20/35

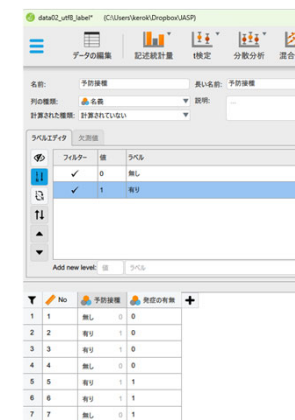
## ラベル

- 名義変数のデータを数値で入れると、その数値に対して表示上の文字を設定できる
- データ自体は元の数値のままなので、並び順は数字になる

21/35

## ラベル

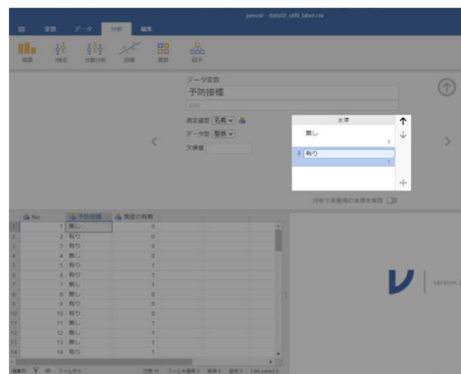
- ラベルをつけると保存はJASP形式(\*.jasp)  
- CSVではなくなる



22/35

## ラベル

- jamoviでもできる
- jamoviファイル(\*.omv)



/35

## ラベルの意義

- 文字データでも一応大丈夫なのに、わざわざ数値にしてラベルをつけるのはなぜか  
- 日本語でのデータ入力を信用していないから  
- ラベルにすると、表示したい順番をコントロールできるから

24/35

## ラベルの意義

- 表示順のコントロール
  - 文字のデータ(カテゴリ)だと、出力は文字コード順
    - それが、自分の望んだ順番とは限らない
  - 数値で入れてラベルにすると、数値の順で表示となる
    - 表示を望んだ順にできる

25/35

## データラベル

- 左 文字として入力 右 数値データにラベル

26/35

## 集計すると

- 左 文字として入力 右 数値データにラベル

度数分布表 ▼

居住の頻度 ▼

居住	頻度	パーセント	有効パーセント	累積パーセント
同居	40	40.0	40.0	40.0
独居	60	60.0	60.0	100.0
欠損値	0	0.0		
合計	100	100.0		

朝食の頻度

朝食	頻度	パーセント	有効パーセント	累積パーセント
食べない	50	50.0	50.0	50.0
食べる	50	50.0	50.0	100.0
欠損値	0	0.0		
合計	100	100.0		

度数分布表

同居の頻度

同居	頻度	パーセント	有効パーセント	累積パーセント
独居	60	60.0	60.0	60.0
同居	40	40.0	40.0	100.0
欠損値	0	0.0		
合計	100	100.0		

朝食の頻度

朝食	頻度	パーセント	有効パーセント	累積パーセント
食べる	50	50.0	50.0	50.0
食べない	50	50.0	50.0	100.0
欠損値	0	0.0		
合計	100	100.0		

27/35

## 出力の扱い

28/35

## JASPの出力

- コピー可能  
コピー時はHTMLが原則  
TeXも可能
- ファイル出力は  
HTML  
PDF
- グラフは原則pngかtiff

29/35

## 出力の利用

- 全体  
– HTMLかPDF



30/35

## 出力の利用

- 部分
- 個別

度数分布表

↑ 折りたたむ  
↓ コピー

同様の頻度	頻度	パーセント	有効パーセント	累積パーセント
同様の頻度				
独居	40	40.0	40.0	60.0
同居	0	0.0	40.0	100.0
欠損値	0	0.0		
合計	100	100.0		

朝食の頻度

朝食	頻度	パーセント	有効パーセント	累積パーセント
食べる	50	50.0	50.0	50.0
食べない	50	50.0	50.0	100.0
欠損値	0	0.0		
合計	100	100.0		

同様の頻度

↑ コピー  
↓ LaTeXにコピーする

同様の頻度	頻度	パーセント	有効パーセント	累積パーセント
同様の頻度				
独居	40	40.0	40.0	60.0
同居	0	0.0	40.0	100.0
欠損値	0	0.0		
合計	100	100.0		

31/35

## コピーをするとHTML形式

- Wordへの貼り付け

度数分布表

↑ 折りたたむ  
↓ コピー

同様の頻度	頻度	パーセント	有効パーセント	累積パーセント
同様の頻度				
独居	40	40.0	40.0	60.0
同居	0	0.0	40.0	100.0
欠損値	0	0.0		
合計	100	100.0		

朝食の頻度

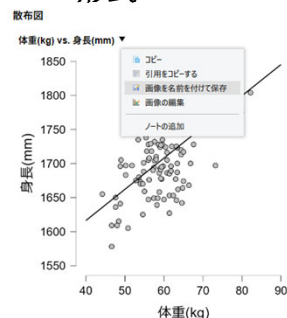
朝食	頻度	パーセント	有効パーセント	累積パーセント
食べる	50	50.0	50.0	50.0
食べない	50	50.0	50.0	100.0
欠損値	0	0.0		
合計	100	100.0		

32/35



## グラフ

### ・形式



ファイル名(N):

ファイルの種類(T):

フォルダの非表示

※全体の一括保存を選択したときにグラフがあると、png  
ファイルとなって、HTML内で埋め込まれる

33/35

## グラフ出力形式

- ・ png
- ・ tiff
- ・ pdf
- ・ pptx(PowerPoint)

ファイル名(N):

ファイルの種類(T):

フォルダの非表示

34/35

## まとめ

- ・ 入力データはExcelでCSVのUTF-8形式で作成する  
といい
- ・ 名義データは数値で入力して、JASPでラベルをつけた  
方がいい
- ・ 出力は、場合によってはExcel等で編集した方がいい  
- 棒グラフや円グラフ、ヒストグラム等の基本グラフは、Excel  
の方が編集しやすくていいかもしれない

35/35