

Data Dictionary

Steve Simon

Data dictionary example (1/5)

```
introduction-to-r-  
part1/doc/data_dictionary_fd.txt  
Written by Steve Simon
```

See introduction-to-r-part1/README.md for an
overview of everything.

This file was downloaded from

```
http://www.amstat.org/publications/jse/datasets/f  
at.dat.txt.
```

Data dictionary example (2/5)

A description of the file appears at

<https://ww2.amstat.org/publications/jse/datasets/fat.txt>.

It represents a study of body fat and body circumference measurements on 252 men. The data was first used in

Penrose, K., Nelson, A., and Fisher, A. (1985), "Generalized Body Composition Prediction Equation for Men Using Simple Measurement Techniques" (abstract), *Medicine and Science in Sports and Exercise*, 17(2), 189.

Data dictionary example (3/5)

and later described in a publication

Johnson (1996) "Fitting Percentage of Body Fat to Simple Body Measurements" Journal of Statistics Education.

and stored in the data archive for this journal. The data set is freely available for re-use for non-commercial purposes.

Data dictionary example (4/5)

There are 252 rows and 19 columns in this data set. It is stored as a text file with tab delimiters. There are no missing value codes in this data set.

case, Case Number from 1 to 252.
fat.b, Percent body fat using Brozek's equation
fat.s, Percent body fat using Siri's equation
dens, Density (gm/cm³)
age, Age (yrs)
wt, Weight (lbs)
ht, Height (inches)

Data dictionary example (5/5)

```
bmi, Adiposity index = Weight/Height^2 (kg/m^2)
ffw, Fat Free Weight = (1 - fraction of body fat)
neck, Neck circumference (cm)
chest, Chest circumference (cm)
abdomen, Abdomen circumference (cm)
hip, Hip circumference (cm)
thigh, Thigh circumference (cm)
knee, Knee circumference (cm)
ankle, Ankle circumference (cm)
biceps, Extended biceps circumference (cm)
forearm, Forearm circumference (cm)
wrist, Wrist circumference (cm)
```

Elements of a data dictionary (1/3)

- Where the data came from, including urls and references, if available
- A brief description of the data
- Licensing information for this data
- The format of the data, including information about delimiters, if appropriate
- The number of rows and columns in the data
- The codes for missing values

Elements of a data dictionary (2/3)

- The data dictionary also includes an entry for each variable with
 - the name of the variable,
 - a brief description,
 - range restrictions (if applicable), and
 - units of measurement (if appropriate).
- If the data is categorical
 - all possible values,
 - with labels for number codes/abbreviations

Elements of a data dictionary (3/3)

- A data dictionary does not have to include all of the things I listed above and it can include things that I did not include. Use your best judgement to decide how much to document in the data dictionary.

Further reading

- Kristin Briney. Data dictionaries. Data Ab Initio blog, August 5, 2014. Available at <http://dataabinitio.com/?p=454%C2%A0>.

About this presentation

- This Powerpoint presentation was created using an RMarkdown program. The program and presentation were written by Steve Simon on 2019-07-22.
- Both the program and the presentation are placed in the public domain (no copyright). If you use this program or the Powerpoint file, an acknowledgement, such as “Thanks to Steve Simon (www.pmean.com) for sharing this material.” would be appreciated, but it is not required.