Creating a New SPSS Dataset -- Variable Specification and Data Entry

When you open SPSS you will see the Data Editor.

- Each row holds the data from a participant or "case"
- Each column holds the data for a variable

After specifying the information for each variable using the Variable View of the editor, we will return to this Data View to enter the data from each case.

It is important to specify key aspects of each variable. These specifications will make the statistical analyses and output clearer and easier to read as well as provide a “reminder” of the decisions you made when collecting and entering the data.
The initial form of most datasets is a listing of the variable values for each participant. Usually each row is a different participant with a column for each variable that was collected.

Below are the data for 12 participants. Each “participant” is a different pet store. The researchers visited each store and recorded information about eight variables. Those variables and details about each are shown on the right:

1. Participant number -- usually consecutive numbers starting with 1, 001 or some similar numbering system
2. Type of reptile department -- “1” means the reptile displays are mixed in with other animal and product displays and “2” means that the store has a separate department
3. Rating of the quality of the reptiles -- using a 1-10 scale
4. Count of the number of reptiles in the store
5. Type if fish available -- “1” means only freshwater fish are available in the store and “2” means that both freshwater and saltwater fish are available
6. Rating of the quality of the fish -- using a 1-10 scale
7. Count of the number of fish in the store
8. The ownership of the store -- “1” means the store is a franchise of a pet store chain, “2” means the store is privately owned, and “3” means the store is part of a pet store cooperative
9. Count of the number of mammals in the store

Naming Variables

- Click a cell under Variable and type in the variable name.
- Variable names may have up to 8 characters -- letters, numbers, “-” and “_” work best -- the first character must be a letter.
- When you enter a name for a variable a set of “default” values automatically appears in the other columns -- some of these we will change later.
- The variable names for these data are shown to the right, along with the default values.

Type, Width & Decimals

- The default values for these specifications usually work very well.
- “Numerical” is the most common type for both qualitative & quantitative data.
- “Width” tells the number of digits in the largest variable value -- must be at least one more than the number of decimal values.
- “Decimals” tells how many digits of a variable value are decimal values.
Variable Labels
- Click a cell under Label and type up to a 40-character description of the variable -- this label will appear in the output making it easier to understand
- Giving a Label to every variable in the data set prevents forgetting what the 8-character variable names mean

Value Labels
- Specify Values for every qualitative variable -- these will appear in the output
- Click a cell under Values. Click the gray box that appears in the cell -- shown on the left for reptdept in row 2
- The Value Labels window will appear. In it type each “Value,” its “Value Label,” and click Add -- shown to the right for the value “2” of reptdept
- You can also highlight a specified value label and Change or Remove it later

Missing Values
- Leaving a cell in the dataset blank tells SPSS that value for that participant has a missing value for that variable -- a “.” will appear in that cell, signalling a “system missing value”
- You can also identify alternative “user-missing” values to indicate different types of missing data (e.g., -99 = did not respond, -98 = answer could not be scored -- up to 3 different missing values)
- Click the cell under Missing. Click the little gray box that appears in the cell -- shown on the left for reptdept in row 2
- Click the “Discrete missing values” button and enter the missing values you’ve chosen in the windows -- shown below for “-9” for reptdept
- Be sure to specify the meaning of each missing value in the “Values” column
Entering Data

- Click the “Data View” tab in the bottom-left corner of the screen.
- Enter each variable value -- in the correct column -- for each case or participant.
- If you leave a space blank a “.” will appear, indicating that the value is missing -- called a “system missing” value.

The completed data set for the petstores is shown below. These data will be used for many of the data analysis examples in this handbook.

Saving the Dataset

- File then Save As and proceed with the usual Windows “save” routine.
- SPSS data sets have the extension “.sav”.

![SPSS Data Editor](image-url)