Exploring Crime Rates in New South Wales

In this workshop, we will cover the following:

1. Explore the Dataset for Crime Rates by Local Government Areas in New South Wales.
2. How to move variables
3. How to sort cases
4. How to generate z-scores of a variable

This workshop introduces the Data for Crime Rates in Local Government Areas in New South Wales. The data include information on crime rates, population, economy and residents of local government areas (LGA) in New South Wales. The information is collected from various sources. For detailed description of variables, see “Codebook of Crime Rates of NSW LGAs”. Please download “Data File for Crime Rates of NSW LGAs” from the iLearn page (it is under the Dataset tab) and open it in SPSS. Then, you are ready to start the workshop.

# Moving (or Relocating) Variables

When you have many variables in your dataset, it is often hard to locate specific variables of interest when you browse the data. In that case, you can move a variable to a new location in the *Variable View*. Let’s say we want to move the *region2* variable next to the *name* variable.

1. **Click the row number of a variable you intend to move**. This will highlight the row of the selected variable.



<Figure 1>

1. **Drag and drop the variable to the new location** (in our case, just below the *name* variable).



<Figure 2>

1. You will see *region2* below *name*.



<Figure 3>

1. Double-click the *name* variable. It will automatically open the *data view* of the chosen variable. You see *region2* is nicely located next to *name.* (If you see values instead of labels for your region2 variable, go to **View** at the top menu and tick **Value Labels**



<Figure 4>

# Sorting Cases

Sorting cases may be necessary when you want to have a close look at the data. Sorting cases will rearrange the rows (cases) in the dataset by any variable or combination of variables you wish. By default, the cases will be sorted in ascending order (smallest to largest, or alphabetical), but sometimes it may be helpful to sort cases in descending order (largest to smallest, or reverse alphabetical).

## Sorting Cases by a Single Variable

Suppose that we want to know which local government areas (LGA) in NSW has the highest crime rates of robbery (*robbery*). Sorting all the NSW LGAs in descending order will make this job easier.

**Go to Data > Sort Cases**.



<Figure 5>

In the popped-up box, **select a variable by which cases will be sorted** and **move them to the *Sort by* box**. And **choose *Descending* in the *Sort Order* area**. Then, **click *OK*** at the bottom.



<Figure 6>

In the *Data View*, you see cases (LCAs) are sorted by robbery rate. The LGA with the highest rate of robbery is Moree Plains Shire, and the municipality of Woollahra has the lowest rate (excluding LCAs having missing values on robbery rates).



<Figure 7>

## Sorting Cases by Multiple Variables

It is often the case that we need to sort cases by more than one variable. Suppose now we would like to find LCAs with the highest robbery rates in the Greater Metropolitan Sydney, Murray and Hunter, respectively. To investigate it, it is necessary to sort all the LGAs by the combination of *region2* and *robbery* variables.

**Go to Data > Sort Cases**. In the popped-up box, **move *region2* first and then *robbery* to the *Sort by* box**. When you use two sorting variables, the order of variables matters. In our case, the cases will first be sorted by *region2*. Within each region, cases will be sorted by *robbery*. You can click and drag the variables to reorder them within the *Sort by* box.

Next, you can specify ascending or descending order for each variable. **Click on a variable in the *Sort by* box to highlight it, and then choose *Ascending* or *Descending***. We choose *Ascending* for *region2* and *Descending* for *robbery*. Then, click *OK*.

 

<Figure 8>

In the Data View, you will see all the cases sorted by region2 and robbery. LGAs in Greater Metropolitan Sydney (coded as 1) appear first, those in Sydney Surrounds (coded as 2) second and so forth. Within the same area, LGAs with the highest robbery rates appear first, and those with the lowest robbery rates appear last. Therefore, it is easy to find which LGA has the highest and lowest robbery rate in each area.



<Figure 9>

# Generating Z-scores of Variables

In this section, we will make a new variable which is the z-scores of *medage* (median age of residents). To generate z-scores, **Go to Analyze > Descriptive Statistics > Descriptives**.



<Figure 10>

Select a variable for which you want to generate z-scores and move it to the Variable(s) box. Then, tick the box of “Save standardized values as variables”. Then, click OK.



<Figure 11>

You will see a newly created variable, *Zmedage*, at the bottom of your *variable view*. Move this variable below to the original *medage* for easier comparison. Now you are ready to move on to the workshop activities



<Figure 12>

**Workshop Activities**

Using the Data for Crime Rates in Local Government Areas in New South Wales, answer the following questions.

Q1. What is the unit of analysis in this dataset?

Q2. Examine non-domestic violence rates. Which LGA is reported to have the highest non-domestic violence rate? And which LCA is reported to have the lowest rate? Use *astnondomviol* variable.

Q3. Find the LGA that is reported to have the highest crime rate of sexual offences (*sexoff*) in each sub-region of New South Wales (*region2*). Note that NSW has 13 sub-regions.

Q4. Compare the distribution of crime rates of sexual offences (*sexoff*) between urban and rural LGAs (*urban*).

1. Compare the central tendency of the distribution. On average, do urban LGAs have higher sexual offence rates than rural LGAs?
2. Compare the variability of the distribution. Do both distributions show a similar level of variability? Use all the available information to answer the question.

Q5. This question will ask you to compare the raw scores of average family size (*avgfamsize*) with its z-scores. Answer the following questions.

1. Check whether the distribution of *avgfamsize* is normal or skewed and justify your answer.
2. Have SPSS generate a new variable, which is the z-scores of *avgfamsize*. Report the z-score of avgfamsize for City of Sydney, City of Blacktown, Inner West Council, City of Ryde and Mosman Council. Note that all these LGAs are in the Greater Metropolitan Sydney.
3. Make histograms for *avgfamsize* and its z-score variable (*Zavgfamsize*). Do both histrograms look the same or not? And explain why.
4. Using the standard normal table, what is the percentage of LGAs having the average family size is equal to or less than 3.20? To answer this question, first find the equivalent z-score for LGA of which the average family size is 3.20, and then find the area below this z-score using the standard normal table.
5. Produce the frequency table of *avgfamsize*. And use the frequency table to find the percentage of NSW LGAs of which the average family size is equal to or less than 3.20. (Tip: the cumulative percentage is the most relevant information for this question)
6. Does the percentage you found from the standard normal table (Q5-Db.) correspond to the percentage from the frequency table (Q5-E.)? If the difference is equal to or smaller than 2%, you can say that they correspond to each other. If the difference is larger than 2%, we don’t say they correspond.