

Checking for Duplicate Cases¹

It is often necessary when cleaning data to check for duplicate cases. Following is SPSS syntax that will help you identify duplicate cases. First the cases are sorted by variable "ID". Then the **LAG** function is used to test if the case preceding each case has the same value for the variable "ID". (The variable "ID" is used as an example, but the same technique can be used with any variable.)

```
* dup sps .
* find duplicate cases .

DATA LIST FREE / ID * .
BEGIN DATA
1
2
1
3
4
5
END DATA .
sort cases by id .
compute dup = 0 .
if lag(id) = id dup = 1 .
freq var = dup .
list var = all .
```

Output Follows:

DUP

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	5	83.3	83.3	83.3
	1.00	1	16.7	16.7	100.0
	Total	6	100.0	100.0	

```
      ID      DUP
1.00      .00
1.00      1.00
2.00      .00
3.00      .00
4.00      .00
5.00      .00
```

Number of cases read: 6 Number of cases listed: 6

¹Prepared by Patty Glynn, University of Washington, C:\all\help\helpnew\dupsps.wpd, April 25, 2002

If you have a lot of cases, you may not want to list them all. So there are other ways that you find the duplicate cases. You may go to the Data Editor, go to the column that has the variable “dup” in it, and hold down the Ctrl Key and press F, for Find. Enter a 1 in the field, and then click on Find Next. You will then be able to see the duplicate case, and the case that precedes it.

The screenshot shows the SPSS Data Editor interface. The main window displays a data table with the following content:

	id	dup	var	var	var	var
1	1.00	.00				
2	1.00	1.00				
3	2.00	.00				
4	3.00	.00				
5	4.00	.00				
6	5.00	.00				
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

Overlaid on the data table is the "Find Data in Variable DUP" dialog box. The "Find" tab is active, and the "Find what" field contains the value "1". The "Match case" checkbox is unchecked. The "Find Next" button is highlighted, indicating it is the next action to be taken.