

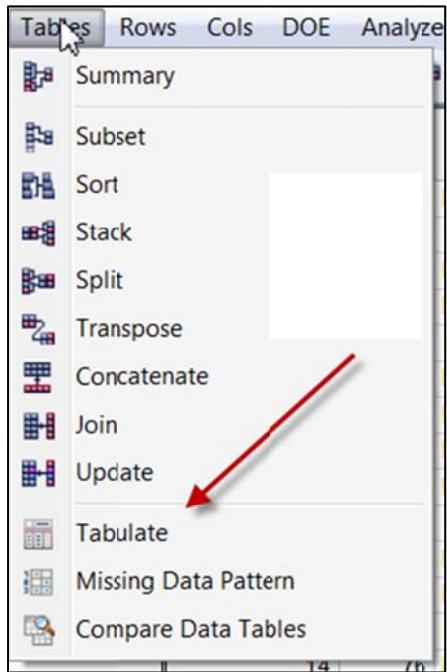
Converting individual-level data to summary data

Chong Ho Yu, Ph.D. (2013)

Azusa Pacific University

chonghoyu@gmail.com

<http://www.creative-wisdom.com/computer/sas/sas.html>



When a summary data set (e.g. by country, state...etc.) is used, the interpretation of the result is subject to the ecological fallacy. However, sometimes we have to use summary-level data. For example, the researcher may want to investigate the inter-relationships between GPA, SAT, assessment test scores, and National Assessment of Educational Progress (NAEP) scores. Even though the researcher has already collected individual-level data for each student, he has access to the NAEP data only (summary data by state). In this case, he needs to collapse student-level data into state-level data in order to match the NAEP data set.

He can make the data conversion in JMP easily. First, choose **Tabulate** from **Tables**. In the pop-up window, drag the variable "state" into **Drop zone for rows**. By default, the number of students (N) shows up in a column.

Two screenshots of the JMP Tabulate dialog box. The left screenshot shows the 'Control Panel' with 'state' selected and 'N' chosen. The right screenshot shows the resulting table with columns 'state' and 'N' containing data for various states.

state	N
AZ	6
CA	9
MN	6
NM	11
NY	15
OK	6
TX	11
UT	13
WA	13

Next, drag “scores”, “GPA”, and “SAT” into the **column zone**, which is occupied by N in this example. In the drop down menu, choose **Add Analysis Columns**. By default, the sum scores by state show up in three columns. But this is not what the researcher wants.

The figure consists of two side-by-side screenshots of the JMP Tabulate software interface. Both screens show a 'Control Panel' on the left and a data preview table on the right.

Left Screenshot: The 'Control Panel' shows the 'scores' column selected. A purple arrow points from the 'scores' entry in the list to the 'N' entry in the column zone below it. The data preview table shows the 'state' column and the 'N' column.

Right Screenshot: The 'Control Panel' shows the 'scores' column selected. A red arrow points from the 'Add Analysis Columns' button in the bottom-left corner of the panel to the 'Sum' entries in the 'scores', 'GPA', and 'SAT' columns of the data preview table. The data preview table now includes three new columns: 'Sum' under 'scores', 'Sum' under 'GPA', and 'Sum' under 'SAT'.

This screenshot shows the final state of the JMP Tabulate interface. The 'Control Panel' on the left has the 'Mean' option selected for the 'scores' column. A purple arrow points from the 'Mean' entry in the list to the 'Mean' entry in the column zone below it. The data preview table on the right now displays the mean scores for each state across the 'scores', 'GPA', and 'SAT' columns.

state	scores Mean	GPA Mean	SAT Mean
AZ	81.166666667	3.2166666667	1916.6666667
CA	65.111111111	2.7877777778	1700
MN	74.5	2.2166666667	1845
NM	58.544545454	3.1636363636	1817.2727273
NY	76.4	2.846	1806.6666667
OK	71.331333333	3.1666666667	1888.3333333
TX	74.81181818	3.2184618182	2000
UT	69.151846154	3.1184615385	2067.6923077
WA	69.61184615	2.5569230769	1674.6153846

To obtain the mean score of all the column variables, drag the **Mean** from the left panel into the **Mean** in the **column zone**. Now the sum scores are replaced by the mean scores.

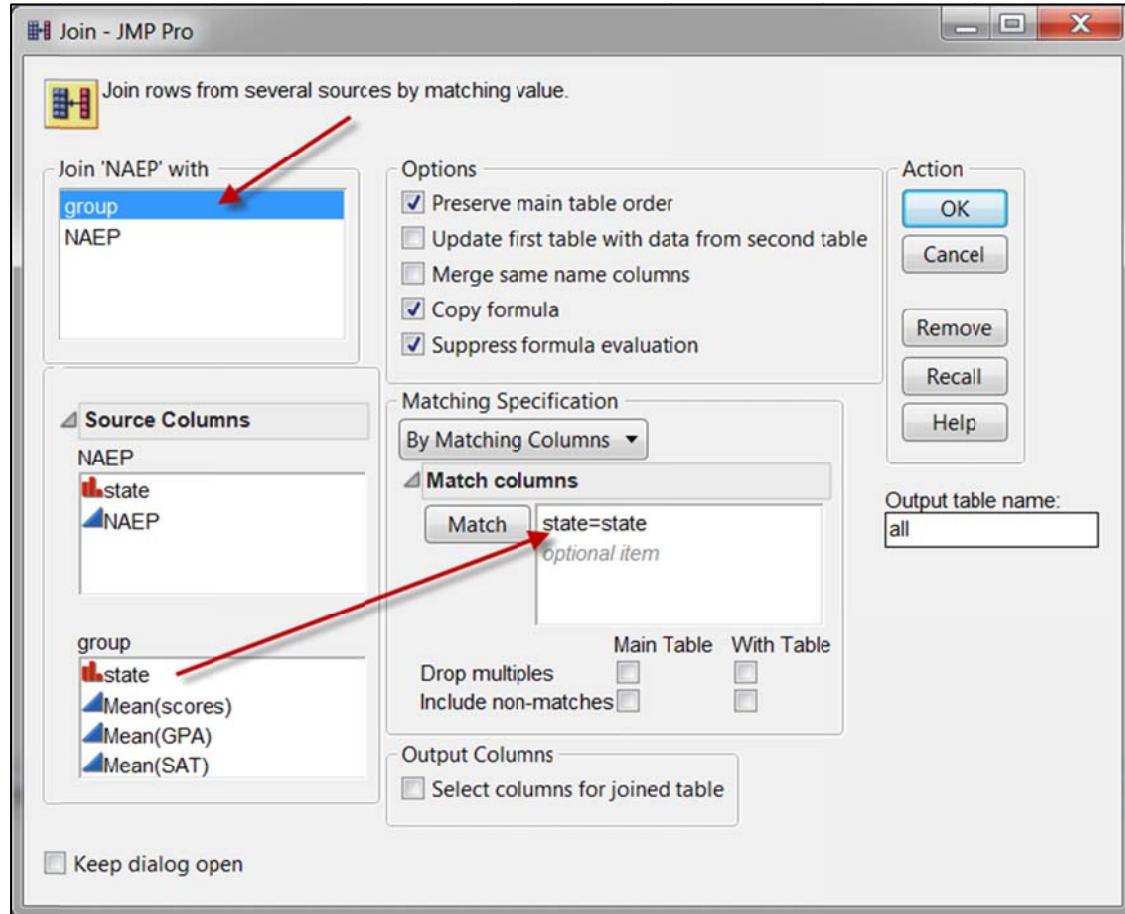
Afterwards, choose **Make Into Data Table** from the red triangle, and the summary data set is ready to go.

The screenshot shows the JMP software interface. On the left, a floating window displays the 'Tabulate' menu with several options: 'Show Table', 'Show Chart', 'Show Control Panel', 'Show shading', 'Show tool tip', 'Show Test Build Panel', 'Make Into Data Table' (which has a red arrow pointing to it), and 'Script'. The main workspace shows a data table titled 'group - JMP Pro' with columns 'state', 'Mear(scores)', 'Mean(GPA)', and 'Mean(SAT)'. The data includes rows for AZ, CA, MN, NM, NY, OK, TX, UT, and WA, along with summary statistics for each column.

The right panel below is the NAEP data set. To merge the NAEP data set with the newly created data set, choose **Join** from **Tables**.

The screenshot shows the JMP software interface. On the left, a floating window displays the 'Tables' menu with options: 'Summary', 'Subset', 'Sort', 'Stack', 'Split', 'Transpose', 'Concatenate', 'Join' (which has a red arrow pointing to it), 'Update', 'Tabulate', 'Missing Data Pattern', and 'Compare Data Tables'. The main workspace shows a data table titled 'NAEP - JMP Pro' with columns 'state' and 'NAEP'. The data includes rows for AZ, CA, MN, NM, NY, OK, TX, UT, and WA, along with summary statistics for each column.

Drag “state” into **Match** because both data sets have the state ID in common. You can merge the data into one of the original tables. But it is a good idea to keep the existing data intact, just in case any error happens. To join the two tables into a new data set, type a new name in **Output table name**. In this example the new data set is called “all”. Now the research has what he needs for state-level analysis.



The screenshot shows the 'all - JMP Pro' window displaying a data table. The table has columns: 'state of NAEP', 'NAEP', 'state of group', 'Mean(scores)', 'Mean(GPA)', and 'Mean(SAT)'. The data rows are:

	state of NAEP	NAEP	state of group	Mean(scores)	Mean(GPA)	Mean(SAT)
1	AZ	212	AZ	91.166666667	3.2166666667	1916.6666667
2	CA	220	CA	65.111111111	2.7877777778	1700
3	MN	233	MN		74.5	2.2166666667
4	NM	236	NM	58.545454545	3.1636363636	1817.2727273
5	NY	207	NY		76.4	2.846
6	OK	228	OK	71.333333333	3.1666666667	1888.3333333
7	TX	214	TX	74.818181818	3.2181818182	2000
8	UT	225	UT	69.153846154	3.1184615385	2067.6923077
9	WA	243	WA	69.615384615	2.5569230769	1674.6153846