

# INTERPRETING LARGE DATA SETS - bivariate analysis

	D	E	F	O	P	Q	R
1	Birth Weight (kg)	Gestation (days)	Mother's Age (yrs)	Smoked?	Random		SOURCE: BERKLEY UNIVERSITY OF CALIFORNIA STAT LAB
2	3.23	283	15	1	0.9325		
3	3.60	242	17	1	0.0187		<a href="http://www.stat.berkeley.edu/~statlabs/data/babies_data">www.stat.berkeley.edu/~statlabs/data/babies_data</a>
4	3.52	284	17	0	0.1470		<a href="http://www.stat.berkeley.edu/~statlabs/labs.html">www.stat.berkeley.edu/~statlabs/labs.html</a>
5	4.08	289	17	1	0.3756		
6	3.40	271	17	1	0.4576		
7	4.00	284	17	0	0.8333		

|< ----- DATA (1132 items) ----- >| Random Number      Source (VERY important)

## Mothers and Babies (1132 rows of data)

This set of data is available from

[www.tsm-resources.com/useful-files.html](http://www.tsm-resources.com/useful-files.html)

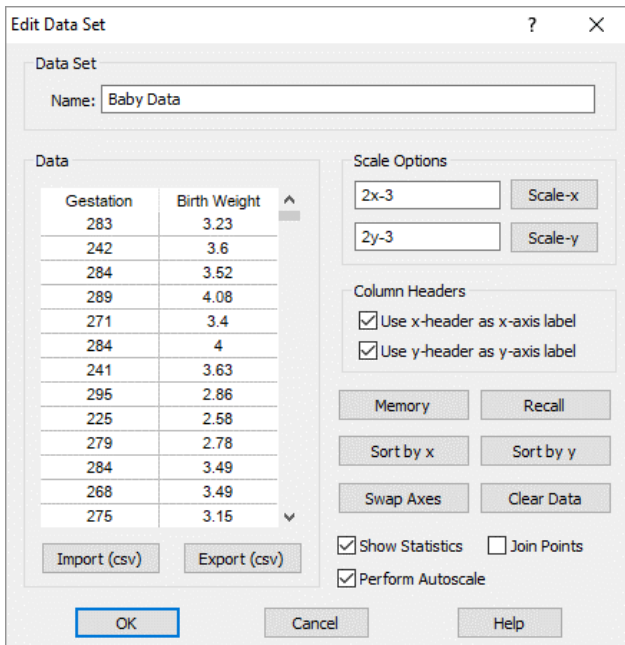
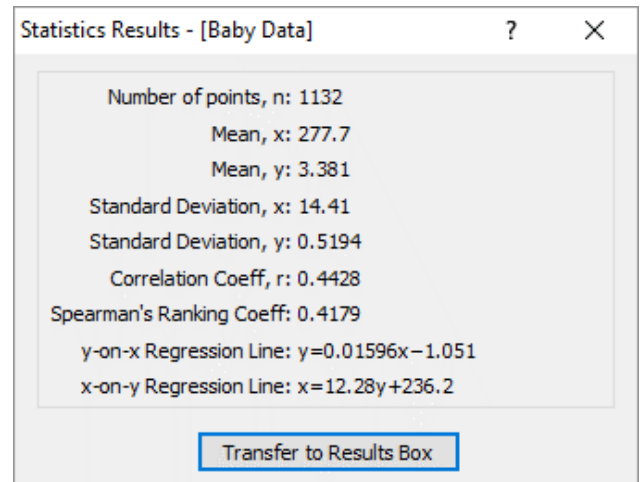
The "RANDOM" column contains random number in the range 0 – 1. If you sort the data by this column, you can then select a random sample.

The SOURCE information gives important links to the data from Berkeley University, San Francisco.

## Bivariate data: Weight vs Gestation

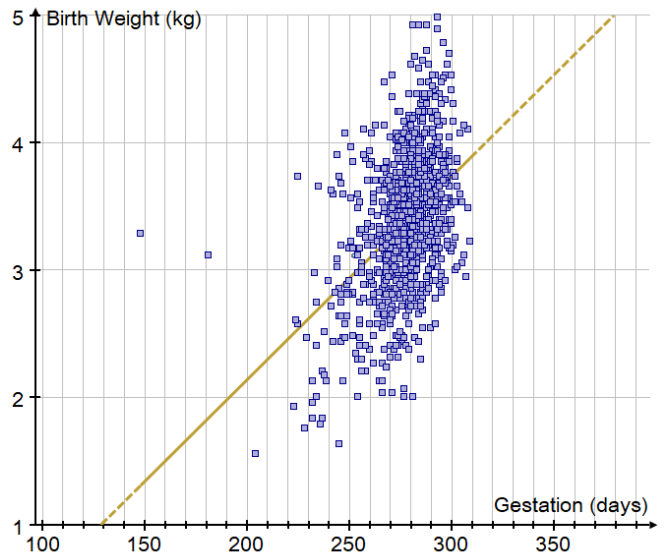
Select columns 'D' and 'E', copy and paste into Autograph using "Enter XY Data" on a 2D page.

Ticking SHOW STATISTICS is also useful:



Note that the two column headers are used to label the 'x' and 'y' axes.

The usual convention is for the INDEPENDENT VARIABLE to be on the x-axis. In this case that is the Mother's Age, so you will want to use the option to "SWAP AXES". to put this right.



In Excel, you can use the DATA -> FILTER tool to analyse separately Mothers who do and do not smoke. The effect on the weights of the babies is alarming.

Individual plotted points can be selected and moved by pressing CTRL-Click. For white-board use, HOLD and DRAG has the same effect.