

INTERPRETING LARGE DATA SETS – single variable

	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		www.stat.berkeley.edu/~statlabs/data/babies_data
4	3.52	284	17	0	0.1470		www.stat.berkeley.edu/~statlabs/labs.html
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 Source (VERY important)
Number

Mothers and Babies (1132 rows of data)

This set of data is available from

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.

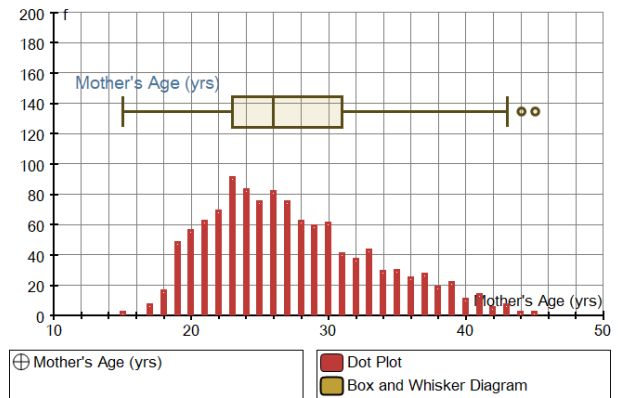
Single Variable analysis: Mother's Age

Select columns 'F', copy and paste into Autograph using "Enter Raw Data" on a Statistics page.

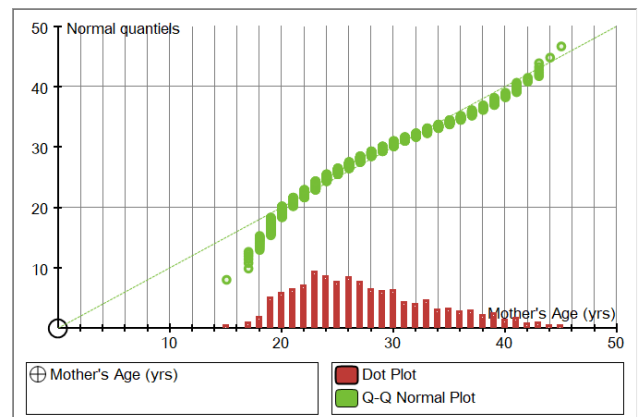
This dialogue can also be used for creating SAMPLE DATA from the many distributions in Autograph.

After pressing OK, a raw data set can be interpreted as a DOT PLOT (you can set the spacing)

or a BOX PLOT (with data outside 1.5 x IQR indicated as outliers)



You can position the box plot by dragging up/down, and edit the label through a right-click option.



A **QQ plot** indicates how 'normal' this data is (the dot plot has been rescaled to fit)