

DATASNOOP 1.2

For PalmOS® 3.5 or higher

Compiled in CASL 3.3, by Galen Currah, 25 August 2002

DataSnoop will read a Memo Pad file, accept pasted data, take graffiti input, or open a file created by SocSurvey.prc, and analyze a set of data, obtaining some 26 descriptive and inferential data, generating two graphical plots.

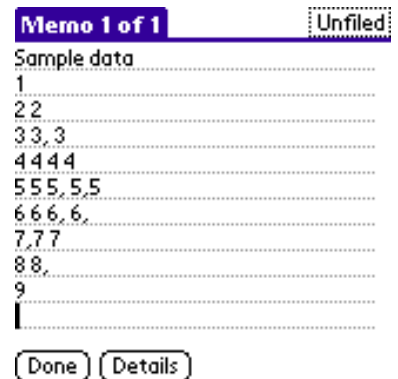
The input must consist of numerical, monovariate data, that is, all numbers that are measures of several instances of the same variable. Data may be normal in their distribution, linear or highly skewed. Non-numerical text will be ignored.

STEP 1

Generate your data. Alpha text will be ignored.

Data may be delimited by tabs, spaces or commas, and may be written on one or many lines.

Or, copy data from another application into the system copy buffer using the Edit menu or /C.



STEP 2

Launch Data Snoop.

DataSnoop requires two other Palm programs:

- (a) CASLrt.prc version 3.3, available free from many sources, and
- (b) MathLib.prc, available free from many sources.

View the "About" and "Help" screens.

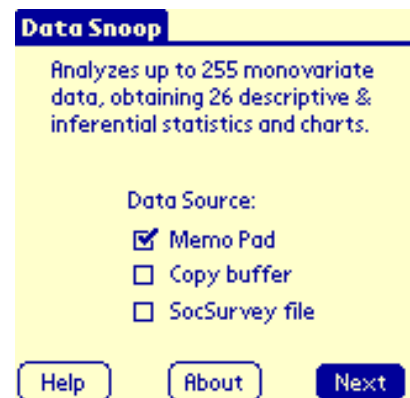


STEP 3

At the first screen, tap on the source of the data to be analyzed:

- a) Memo file = Data found in the Palm Memo Pad.
- b) Copy butter = Data copied from another application.
- c) SocSurvey file = Data gathered through SocSurvey.prc

Data can also be entered later as graffiti at the Edit screen (Step 5).



STEP 4

If you chose to take data from the Memo Pad or from SocSurvey, then the file names will be listed out. Scroll down the list and tap on the name of the file holding the data. Then tap on the Next button.

If you chose to take data from the copy buffer, then this screen will not appear.

Data Source

Tap on memo or SocSurvey file name:

Sample data

File name: Sample data

Help Back Next

STEP 5

At the Edit screen, you may edit data through with graffiti feature.

If you wish to limit the range of data, then enter the lowest and highest values acceptable in the analysis.

If there is a default value in the data set representing missing data, then enter that, too.

These data will not be included in calculations.

Edit Sample data

Ignore data greater than:

Ignore data less than:

Ignore 'missing' data: -999

1
2
3
4
5
6
7

Help End Back Next

STEP 6

DataSnoop will then filter the data per the criteria set at step 5, and will sort the data in numerical order.

This may take several seconds of time, but is required by some statistics and graphical plots.

You may edit data by clicking on them and using the graffiti feature, or delete a datum by tapping on it and on the Del button.

Sample data

6
6
6
7
7
7
8
8
9

Edit: 9

Help Del Back Next

STEP 7

The first set of statistics consists of the data set size (data count), their sum, mean average, population variance, sample variance, population standard deviation, sample standard deviation, standard error of the mean, and a measure of their skewness. See the Appendix.

You must know whether your data represent a population or a sample.

Tap on the Plot button.

(1) Sample data

Size (n)	25
Sum (S)	125.0000
Mean (m)	5.0000
Pop. var. (S ²)	4.0000
Samp. var. (s ²)	4.1667
Pop. St. Dev. (SD)	2.0000
Samp. St. Dev. (sd)	2.0412
Std. Error (SE)	0.4082
Skewness (b)	0.0000

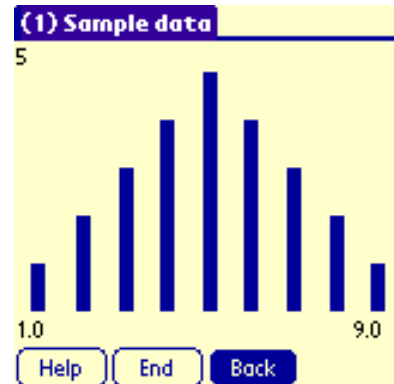
Help Plot Back Next

STEP 8

The first plot presents a vertical bar graph of an approximate data distribution in numerical order.

Some adjacent data may be merged into a single bar.

From this graph, you can make a tentative judgement about the normality of the data set.

**STEP 9**

The second set of statistics consist of the lowest datum, the highest datum, their range, first quartile, median, third quartile, inter-quartile range (IQR), and a count of low and high outliers.

Outliers are data whose values are more than 1.5 IQR above or below the quartiles. Outliers can cause a distortion of the other statistics.

(2) Sample data

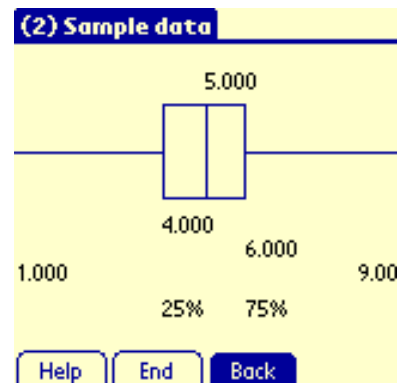
Lowest (min)	1.0000
Highest (max)	9.0000
Range (max-min)	8.0000
1st quartile (Q1)	4.0000
Median (Q2)	5.0000
3rd quartile (Q3)	6.0000
Interq. range (IQR)	1.0000
Low outliers	3
High outliers	3

Buttons: Help, Plot, Back, Next

STEP 10

The second graphical plot represents an approximation of the data set's low, high, range, quartiles, 25th and 75th percentiles, and their median.

From this chart you can make a tentative judgement about the skewness of the data set.

**STEP 11**

Enter any number to find how it relates to the data set as a percentile, and obtain its z score as a single value, its z score as a mean value, its t statistic as a mean value relative to the data mean, and the confidence interval around it as a frequency count.

(The z score of the mean assumes a similar variance.)

(3) Sample data

Enter a number: 7.456

Percentile:	80.70
z score (x):	1.23
z score (m):	6.140
t statistic:	6.016
Conf. Interval:	7.05 - 7.86

Buttons: Help, Clear, Back, End

APPENDIX
The Meanings of These Statistics

Size (n)	Then count of data in the set. Fewer than 32 data are often not statistically useful in social sciences.
Sum (Σ)	The data added together.
Mean (μ)	The sum divided by the size (Σ / n).
Population variance (S^2)	The mean average of squared differences of all data from their mean. This value serves as an element of several other statistics.
Sample variance (σ^2)	Similar to the previous but calculated with $n - 1$ to compensate for the likelihood of sampling error.
Population standard deviation (SD)	The square root of the variance. This allow comparison of the means of populations of different sizes.
Sample standard deviation (σ)	Similar to the previous. This allows comparison of the sample mean with a population mean or between the means of two samples. A difference of 1.96 standard deviations between means is statistically significant at a 95% significance level.
Standard error (SE)	The probable error introduced by chance when drawing a sample from a larger population.
Skewness (b)	A value of 1.96 or more indicates a probability of 95% that a sample data distribution is different from a normal one.
Low	The smallest datum in the set.
High	The biggest datum in the set.
Range	The difference between the high and low values.
First quartile (Q1)	The highest of the lowest 25% of the data in the set.
Median	The highest of the lowest 50% of the data in the set.
Third quartile (Q3)	The highest of the lowest 75% of the data in the set.
Inter-Quartile Range (IQR)	The difference between the first and third quartiles. This range include the middle half of all the data in the set.
Low and high outliers	Data whose values are less than 1.5 IQR below the first quartile or more than 1.5 IQR more than the third quartile.
Percentile	The percentage of the range represented by a single value.
z score of a value	The number of standard deviations at which a score lies from the mean of a population.
z score of a mean	The number of standard deviations at which a sample mean lies from the mean of a population.
t statistic	The number of standard deviations at which a sample mean lies from the mean of another sample.
Confidence interval	The range of values included in 1.96 standard deviations on either side of a value, representing the range of probability that a sample value represents a population value.