Statistics Notes

Detecting skewness from summary information

Douglas G Altman, J Martin Bland

As we have noted before, many statistical methods of analysis assume that the data have a normal distribution.[1] When the data do not they can often be transformed to make them more normal.[2] Readers of published papers may wish to be reassured that the authors have carried out an appropriate analysis. When authors present data in the form of a histogram or scatter diagram then readers can see at a glance whether the distributional assumption is met. If, however, only summary statistics are presented—as is often the case—this is much more difficult. If the summary statistics include the range of the data then some idea of the distribution may be gained. For example, a range from 7 to 41 around a mean of 15 suggests that the data have positive skewness. However, as the range is based on the two most extreme (and hence atypical) values this inference is not reliable. Similar asymmetry affecting the lower and upper quartiles[3] would be much more convincing evidence of a skewed distribution. Usually, however, the only summary statistics presented are the mean and either the standard deviation or standard error. Such information cannot show that the data are near to a normal distribution, but they can sometimes show that they are not.

There are two useful tricks. The normal distribution extends beyond two standard deviations either side of the mean. It follows that for measurements which must be positive (like most of those encountered in medicine) if the mean is smaller than twice the standard deviation the data are likely to be skewed. Table 1 shows urinary cotinine levels related to number of cigarettes smoked daily. Clearly the data must be highly skewed, as the mean is smaller than the standard deviation in each group. This aspect of the data was not apparent in the original paper, which gave just the means and standard errors. (We added the standard deviations, derived simply as standard error × √n.) As a conse-

ICRF Medical Statistics
Group, Centre for
Statistics in Medicine,
Institute of Health
Sciences, PO Box 777,
Oxford OX3 7LF
Douglas G Altman, head

Department of Public
Health Sciences,
St George's Hospital
Medical School, London
SW17 0RE
J Martin Bland, professor of
medical statistics

Correspondence to:
Mr Altman.

Table 1—Urinary cotinine excretion (µg/mg creatinine) related to number of cigarettes smoked daily[4]

Cigarettes smoked per day	No in group	Mean	SE	SD
1-9	25	0.31	0.08	0.40
10-19	57	0.42	0.10	0.75
20-29	99	0.87	0.19	1.89
30-39	38	1.03	0.25	1.54
>40	28	1.56	0.57	3.02
Unspecified	25	0.56	0.16	0.80

quence, the use of t tests was not easily seen to be incorrect.

The second indicator of skewness can be used when, as in table 1, there are data for several groups of individuals. As we have noted,[2] deviations from the normal distribution and a relation between the standard deviation and mean across groups often go together. If the standard deviation increases as the mean increases then this is a good indication that the data are positively skewed, and specifically that a log transformation may be needed.[2] There is a clear relation between mean and standard deviation for the cotinine data. As we have noted, log transformation often removes skewness and makes the standard deviations more similar.

In this example we can detect skewness from summary statistics, but we cannot tell what the effect of log transformation would have been. That requires the raw data.

1 Altman DG, Bland JM. The normal distribution. BMJ 1995;310:298.
2 Bland JM, Altman DG. Transforming data. BMJ 1996;312:770.
3 Altman DG, Bland JM. Quartiles, quintiles, centiles and other quantiles. BMJ 1994;309:996.
4 Matsukura S, Taminato T, Kitano N, Seino Y, Hamada H, Uchihashi M, et al. Effects of environmental tobacco smoke on urinary cotinine excretion in nonsmokers. N Engl J Med 1984;311:828-32.

Lesson of the Week

Importance of accurate assessment of fetomaternal haemorrhage after late abortions

Imelda Bromilow, J K M Duguid, S Walkinshaw

> Fetal blood after late abortion in RhD negative women should be RhD typed to ensure appropriate anti-D immunoglobulin dosage

Anti-D immunoglobulin is routinely administered to RhD negative women after an abortion. Before 20 weeks of gestation 25 IU of anti-D immunoglobulin should be given and after 20 weeks of gestation 500 IU. The Kleihauer stain test should be routinely performed to assess the adequacy of this dose.[1] RhD typing of the aborted fetus is not usually performed.

We report a case in which large doses of anti-D immunoglobulin were given inappropriately because the Kleihauer test detects only fetal haemoglobin; further testing with flow cytometry showed that the abortus was RhD negative.

Case report

A pregnant woman underwent a late termination of pregnancy for a lethal fetal malformation after 20 weeks of gestation. The procedure was carried out using mifepristone pessaries. There were no clinical problems: maternal blood loss was not excessive and the placenta was delivered normally. Anti-D immunoglobulin (500 IU) was given, and Kleihauer testing indicated that a fetal bleeding of about 100 ml had occurred. Additional anti-D immunoglobulin was given to a total dose of 10 500 U. Further Kleihauer testing 24 hours later showed that fetal cells were still evident in the maternal circulation.

Correspondence to:
Dr Duguid.