Thorax 1991;**46**:536–538

Statistics in respiratory medicine · 3

Scale, parametric methods, and transformations

Susan Chinn

Types of data and measurement scales

The first two articles tacitly assumed that we were dealing with a *continuous quantitative* measurement—that is, one like height, blood pressure, or FEV_1 that (within limits compatible with life) can take any value; only the restrictions of the measuring instrument lead to a finite rather than infinite number of possible values. Most continuous quantitative measurements are on a *ratio* scale—that is, one on which the units can change, as from inches to centimetres for height, but on which the zero point is fixed. A difference of, say, two units has the same meaning at each point of the scale. If the latter is satisfied but the scale has no fixed zero point, then it is known as an *interval* scale; temperature measured in degrees Fahrenheit or Celsius is the most obvious example. We shall not need to distinguish between interval and ratio scales for most purposes and will refer to variables of each type as continuous quantitative variables.

A second type of quantitative measurement is *discrete*—that is, it can take only predetermined values, such as integers. Examples are radioactive counts, the number of asthma attacks experienced by a patient in the last 12 months, and the number of siblings with hayfever.

More restricted still are variables and resulting data that are *qualitative* or *categorical.* These come in two types: *ordered,* such as severity of disease (none, mild, moderate, severe) or *unordered* or *nominal,* such as diagnosis. Ordered categorical variables are often given numerical values, but caution should be displayed in using such numerical values as if the variables were discrete quantitative ones. Although symptom scores, derived by adding such numbers over several symptoms, may give as much information as a more complicated weighting of each symptom's severity, this should be recognised for what it is, a discrete quantitative variable derived from an arbitrarily weighted combination of ordinal data. In general, the four types of variable are distinct, and different summary statistics, methods for calculating confidence intervals, and significance tests are appropriate for each.

Parametric methods for quantitative variables

If considerable effort has been put into obtaining a quantitative measurement, then it is sensible to make best use of it in a statistical analysis. In general, this has the implication that, wherever possible, *parametric* methods, which are based on assumptions about the population distribution or distributions, should be used rather than *non-parametric* or *distribution free* methods. The latter (often based on the rank order of the data, which thus converts a continuous quantitative variable into an ordered categorical one) give limited information on the size of differences between groups, put emphasis on significance levels rather than on estimates, and are harder to adapt to complex data sets. This does not rule out their very legitimate use for categorical data, or when other options for quantitative data have been ruled out, but these options should be explored first.

Most commonly used parametric methods are based on two assumptions. The first is that in the population or populations we have sampled the measurement has a normal, or Gaussian, distribution. The second assumption is that any "residual" (that is, unexplained) variation has the same variance, or standard deviation, throughout. Thus for the one sample or paired t test the paired differences must come from a normal distribution, and for the two sample or unpaired t test each group should be a sample from a normal distribution and the two distributions should have the same variance; for testing the significance of the slope of a regression line the variation about the line should be normally distributed with equal variance for each x value.

This is required also for the related confidence intervals. Most methods, however, are reasonably "robust" to non-normality, though obviously unequal variances should not be ignored. In practice this means that we should not worry about non-normality unless we can actually demonstrate it. If samples are too small for this to be demonstrated they are too small for it to matter much which significance test is used; an unpaired t test and Mann-Whitney U test will give similar p values, but the former leads to an estimate and confidence interval for the mean difference between the groups.

Normality can be investigated for larger groups, or for error variation from fitted means or regression lines. A histogram will show any major problem, but a "normal plot"[1] gives some information on the reason for non-normality. If the distribution is Gaussian the normal plot will show a straight line. Departure from linearity can be tested on the

Department of Public Health Medicine, United Medical and Dental Schools of Guy's and St Thomas's Hospitals, St Thomas's Campus, London SE1 7EH
S Chinn

Reprint requests to:
Miss Chinn

Thorax: first published as 10.1136/thx.46.7.536 on 1 July 1991. Downloaded from http://thorax.bmj.com/ on April 17, 2024 by guest. Protected by copyright.

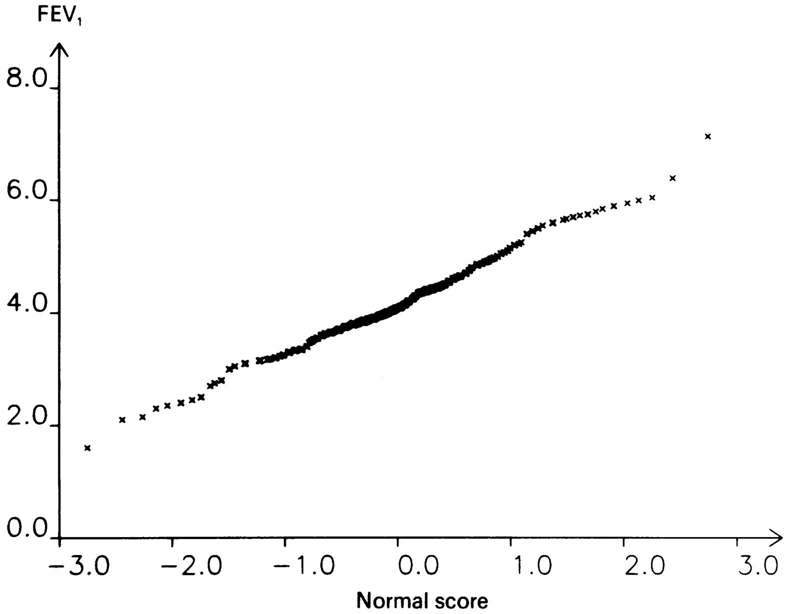

Figure 1 FEV_1 plotted against the normal score that each measurement would have, according to its rank order, if FEV_1 is normally distributed (normal plot) for 220 men aged 18–64 years.

basis of the correlation between the data and the corresponding normal scores, which will have the value 1·0 for perfect normality. This is known as the Shapiro-Wilk statistic.[1] Outliers will deviate from the line, and curvature, uniformly convex or concave, indicates skewness. Other departures from a straight line arise from a mixture of distributions. Figure 1 shows the normal plot for baseline FEV_1 before histamine challenge for a random sample of 220 men aged 18–64 years in a community survey,[2] for which the Shapiro-Wilk statistic is 0·996. Even without height or age adjustment it is reasonably normally distributed, with one or two outliers. The constant variance assumption should be investigated by plotting within subject or within group (as appropriate) standard deviation against mean value.[3] If just two measurements per subject are obtained the absolute value of the difference between them, divided by the square root of 2, plotted against their mean value, is the short cut method for achieving this. Figure 2 shows the plot for 111 repeat values[4] of baseline FEV_1, showing little relation between standard deviation and mean value. Thus these FEV_1 values satisfy both criteria for parametric analysis and would require no transformation.

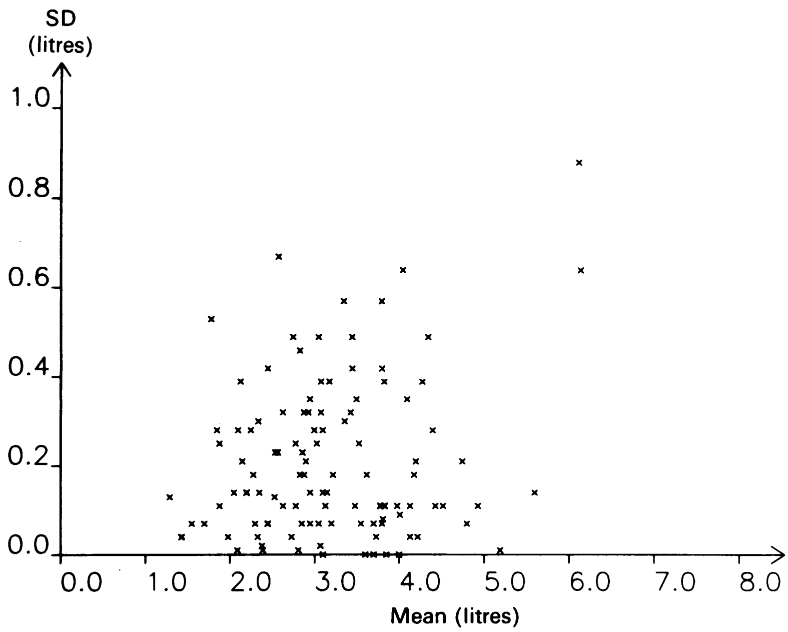

Figure 2 Within subject standard deviation of FEV_1 plotted against the subject mean value for 112 subjects aged 18–64 years.

Transformations for quantitative variables

If a measurement violates the normality and the constant variance assumptions, or even just the latter, the possibility of *transforming* the data should be considered before non-parametric methods are used. The logarithmic transformation is familiar to most research workers, and its use in connection with bronchial challenge tests was described in the last article (June, p454). Taking the log values of data on dose, for example, changes a ratio scale into an interval scale because log(1) = zero and log(0) = minus infinity. From the standpoint of statistical methods it does not matter whether logarithms to base 10, to base e (natural logarithm, denoted 1n), or some other base are used. There are reasons for using base e on occasions, one of which will be mentioned later; but when these do not apply base 10 is preferred for ease of approximate mental arithmetic.

Many variables, once log transformed, satisfy the assumption of normality and constant variance. As described in article 2, antilogged results are easy to interpret. Bland and Altman[5] recommended that no other transformation should be used in the context of method comparison. Certainly when standard deviation increases with mean value the log transformation should be tried before anything more complicated is considered. Many measurements—for example, specific airway conductance[3]—do not show a close relation between standard deviation and mean, but log transformation adequately stabilises the variance. This will not suffice, however, for all measurements. If a graph of standard deviation against the mean value produces a straight line *through the origin* then a logarithmic transformation must be used. If it is a straight line of the form

$$SD = a + b.\text{mean} \quad \text{with } a \neq 0$$

then log $(a + bx)$, where x represents the original measurements, will stabilise variance. This requires $(a + bx)$ to be above zero, so $x > -a/b$; if zero or negative values of x occur then a must have a large enough positive value to satisfy this even if it is greater than that suggested by the relation between standard deviation and mean. A plot (fig 3) and regression line of standard deviation against mean for the slope of the dose-response curve[6] from histamine challenge tests suggested that log($-3+0{\cdot}6$ slope) would improve stability of

Thorax: first published as 10.1136/thx.46.7.536 on 1 July 1991. Downloaded from http://thorax.bmj.com/ on April 17, 2024 by guest. Protected by copyright.

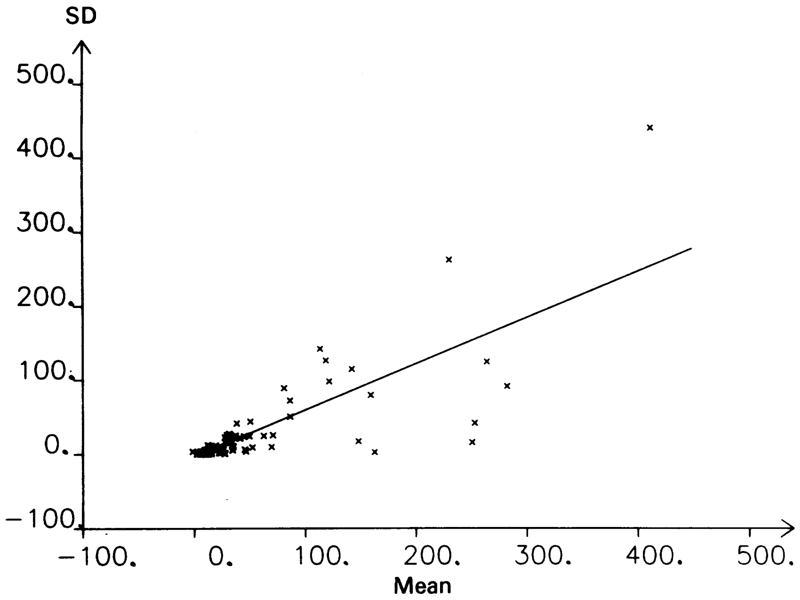

Figure 3 Within subject standard deviation of slope of dose-response curve from histamine challenge tests plotted against the subject mean value for 104 subjects aged 18–64 years, with the regression line of standard deviation on mean.

variance, but in practice log(slope + 10) was used because the slope was negative for some subjects.

More sophisiticated transformations can be investigated by plotting log (SD) against log(mean)[3]; it is recommended that a statistician is consulted about the necessity of this before the analysis proceeds as interpretation of the results may be difficult. If there is more than one component of variation, and they imply or may imply different transformations of the data, a statistician should definitely be consulted, preferably during the design of the study. Usually one component will be of much greater magnitude than the other (or the others), between subject variation almost always swamping any measurement or within subject variation. Provided that this is the source of variation that is being studied the corresponding transformation is the appropriate one. Plotting standard deviation against mean value enables the dependence of standard deviation on mean value to be investigated. A more stringent test of the constancy of the standard deviation is provided by the index of heterogeneity,[3] but satisfying the weaker independence criterion will usually suffice. When a reference range for change is calculated from data from several of many subjects, which it should be, this index can be used to assess the validity of the reference range.

Note on the coefficient of variation

Calculation of the coefficient of variation, the standard deviation divided by the mean value, implies that the value is constant for different mean values—that is, that the standard deviation is proportional to the mean. This in turn, as we have seen above, implies that a log transformation should be used before analysis. The coefficient of variation, as calculated on the original scale, is therefore not valid except when log transformation is appropriate. For variables requiring some other transformation or none at all the coefficient of variation should not be used. The appropriate dimensionless index is the intraclass correlation coefficient, as described in the previous article, and it is preferable to use this in all circumstances as it relates the size of the error variation to the size of the variation of interest. In addition, the intraclass correlation coefficient is calculated on the scale used for analysis—that is, after transformation when this is necessary—whereas the coefficient of variation has to be calculated on the original scale, which must be a ratio scale, with analysis performed on the log scale. Because the log transformation is so often appropriate and research workers wish to compare their repeatability with published values, it is inevitable that the use of the coefficient of variation will be replaced only gradually by the intraclass correlation coefficient. When, as is frequently required, a coefficient of variation has to be estimated from several subjects or samples, an approximate value is given by the within group standard deviation from a one way between subject or sample analysis of variance of $\log_e$ (measurements).[3] In this case logarithms to base e have to be used.

If a coefficient of variation is quoted the source of variation represented by the standard deviation must be made explicit, especially when several components of variation exist, such as within day and between day variation.

1 Minitab Inc. *Minitab reference manual.* Release 7. Pennsylvania: Minitab Inc, 1989:4–8.
2 Burney PGJ, Britton JR, Chinn S, *et al.* Descriptive epidemiology of bronchial reactivity in an adult population: results from a community study. *Thorax* 1987;**42**: 38–44.
3 Chinn S. The assessment of methods of measurement. *Statistics in Medicine* 1990;**9**:351–62.
4 Chinn S, Britton JR, Burney PGJ, Tattersfield AE, Papacosta AO. Estimation and repeatability of the response to inhaled histamine in a community survey. *Thorax* 1987;**42**:45–52.
5 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307–10.
6 O'Connor G, Sparrow D, Taylor D, Segal M, Weiss S. Analysis of dose-response curves to methacholine. *Am Rev Respir Dis* 1987;**136**:1412–7.

Thorax: first published as 10.1136/thx.46.7.536 on 1 July 1991. Downloaded from http://thorax.bmj.com/ on April 17, 2024 by guest. Protected by copyright.