University of York Department of Health Sciences

M.Sc. in Evidence Based Practice

Measurement in Health and Disease

Suggested Answers to Specimen Assessment, June 2005

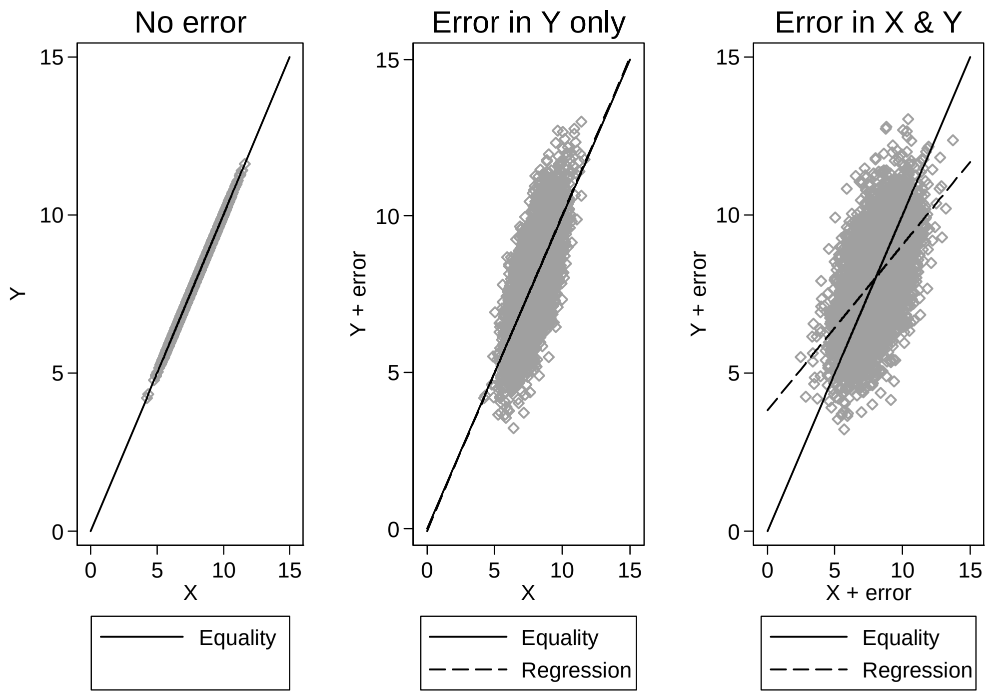
Question 1.

a) *What is meant by 'coefficients of variation <5%'?* A coefficient of variation is a standard deviation divided by a mean. Here it means the standard deviation of repeated measurements on the same subject divided by the mean value of the measurements on that subject. A single value has been estimated for all subjects. This is done because the standard deviation for the subject is proportional to the subject mean. It has been multiplied by 100 to turn the ratio into a percentage and for both HbA(1c) and HbF this was less than 5%.

b) *What are the disadvantages of correlation as a measure of agreement?* The correlation coefficient between two methods of measurement depends on the variability between the subjects. If subjects are chosen to have a wide range of measurements, by selecting extra subjects with extreme values, the correlation coefficient will be increased. If they are selected to be very similar it will be reduced. It is meaningless unless we have a representative sample. The correlation coefficient ignores bias. If one method of measurement gives readings consistently higher than another, this will not affect the correlation.

c) *What is meant by 'limits of agreement'?* The limits of agreement are a range of values within which we expect 95% of differences between measurements by the two methods to lie. The limits are calculated from the mean, $\bar{d}$, and standard deviation, *s*, of differences between the two methods. The limits are then $\bar{d} - 1.96s$ to $\bar{d} + 1.96s$.

d) *Why would we expect a positive intercept and slope to be less than 1.0?* If there is no bias between the two methods of measurements and the average values for a subject obtained by the two methods are the same, we would expect the observations in a scatter diagram to be distributed about the line of equality. This is the line observations would lie on if the two methods always gave identical measurements. It has intercept = 0 and slope = 1. However, the regression line does not estimate this line if there is any error in the X variable. The error in X pulls the points further apart horizontally. This makes the slope of the line less. This in turn increases the intercept, because the line has to go through the point defined by the means of X and Y. The following graph illustrates this. It shows the regression line which we get for 5000 points when Y = X exactly and there is no error. This is the line of equality. It then shows what happens when Y is measured with some error, but X is measured exactly with no error. The regression line again estimates the line of equality. It then shows what happens when there is error in X. The spread of X increases and the line is less steep.



(Of course, this is more than you could do in an exam. I include it to make the point clear.)

e) *'At low concentrations of HbF, the DCA 2000(TM) immunologic method tended to underestimate and at higher concentrations tended to overestimate HbA(1c) when compared with Diamat(TM).' Is this a valid conclusion from the regression analysis?* It is not, because it is based on the regression line having slope less than one and intercept greater than zero. But we would expect this to happen even if there were no bias and the DCA 2000(TM) method always estimated HbA(1c) without any bias when compared with Diamat(TM).

Question 2.

a) *What was meant by 'kappa coefficient = 0.75'?* Kappa is a measure of agreement, here between observations by different observers. It is adjusted for the agreement which would expected if the two sets of observations were not related, called the chance agreement. We divide the proportion of observations for which agreement is observed minus the expected agreement, by one minus the expected agreement. Thus kappa = one if agreement is perfect and zero if agreement is what would be expected by chance. Kappa =0.75 is usually interpreted as meaning 'good' agreement.

b) *What was meant by 'weighted kappa = 0.85'? Under what circumstances would we use weighted kappa?* Weighted kappa takes into account that disagreement may be more severe when observations are in one pair of categories than when they are in another. We usually apply it to variables with ordered categories, such as 'poor', 'fair', 'good', 'excellent'. Disagreement when the categories are close is given less weight than when categories are at opposite ends of the scale. 0.85 would be classified as 'very good' agreement.

c) *What was meant by 'sensitivity (37.5%)'? What does it tell us about using square wave as a test for low ejection fraction?* Sensitivity is used to describe a property of diagnostic tests. Sensitivity is the proportion of those subjects who have the disease who are positive on the test, i.e. the proportion of patients with a low ejection fraction who have a square wave. It tells us how good the test is at detecting the disease. Here a square wave would detect only 37.5% of cases so would not be a good test for detecting low ejection fraction. It is not sensitive.

d) *What was meant by 'specificity (92.7%)'? What does it tell us about using square wave as a test for low ejection fraction?* Specificity describes a property of diagnostic tests. Specificity is the proportion of subjects who do not have the disease who are negatives on the test, i.e. the proportion of patients ***without*** a low ejection fraction who ***do not*** have a square wave. It tells us how good the test is at excluding people who do not have the disease. Here a square wave would be not be found in 92.7% of people without a low ejection fraction cases so would be quite a good test for excluding patients who do not have a low ejection fraction. It is fairly specific.

e) *Why did sensitivity increase but specificity decrease when any abnormal response was considered, compared to a square wave response only?* Any abnormal response is a less stringent test than is a square wave response. Of those patients with a low ejection fraction, more will be positive to the test of any abnormality than to a square wave response, because this will include all those with a square wave response and other patients with other abnormalities. Hence sensitivity goes up. Of those patients ***without*** a low ejection fraction, more will be positive to the test of any abnormality in the same way, because this will include all those with a square wave response and other patients with other abnormalities. Hence fewer of these patients will be negative to the test and specificity will go down.

Question 3.

a) *What is a test-retest correlation at 4 weeks and what does a value of 0.94 tell us?* This is the correlation coefficient between the scale measurement recorded on the same subjects on two separate occasions four weeks apart. This is usually done using an intraclass correlation coefficient, which ignores the order in which measurements were made. It is the ratio of the variance of the subjects true scores to the variance of their measured scores. Correlation coefficients are for a particular population and the sample must be a representative sample of this population. A test-retest correlation equal to 0.94 tells us that, in this population, measurement error is low relative to the variability in the construct being measured.

b) *What is the 'alpha coefficient of internal consistency' (Cronbach's alpha) and what does a value for alpha of 0.90 tell us?* This is a measure of how closely related the items in the scale are to one another. If the items are closely related, the scale will be consistent and will be closely related to the construct it measures. The alpha coefficient estimates the correlation between the measurement and the 'true' value of the construct in this population.. 0.90 is quite a high correlation and tells us the scale has what would usually be regarded as high internal consistency.

c) *What is factor analysis and how do its practitioners decide on the names and meaning of their factors?* Factor analysis is a method of reducing the dimensions of a set of data. We usually have a lot of variables, some of which are correlated with one another. We ask whether these could be explained as the result of a smaller number of underlying variables, the factors. We use a method to find out how many factors we will need, which gives us a series of Eigenvalues. The number of Eigenvalues greater than one is one estimate of how many factors we need. We then find the best combinations of our variables to estimate these factors (by rotation). We decide on the names and meaning of a factor from the variables which contribute substantially to its estimation.

d) *What does it mean when the authors say that 'these factors explained 54.60% of the total variance'?* The total variance is the sum of the variances of all the variables after they have been standardised, i.e. divided by their standard deviations, so it is equal to the number of variables. The sum of the variances of the five factors is equal to the sum of their Eigenvalues. This divided by the total variance is 54.60%.

e) *Why do the authors give us the correlation with Templer's Death Anxiety Scale?* They are testing the validity of their scale, the Death Anxiety Inventory, by checking that measurements made using it are related to another scale which measures a similar or closely related thing. They are trying to establish the validity of their scale. This would be described as establishing convergent validity, the scale is correlated to something with which the construct it measures should be related. This could be seen as establishing criterion validity, as both scales set out to measure similar things.