

List of Symbols & Formulas used on this site

<p>About the symbols used in these formulas</p>	<p>X, Y, etc = raw scores on tests X, Y etc</p> $M_x = \bar{X} = \frac{\sum X}{N}$ <p>x, y, etc = deviation scores = (X - M_x) , (Y - M_y) , etc</p> <p><i>r</i> = a correlation coefficient</p> <p>\hat{X}, \hat{Y}, etc = predicted X. predicted Y, etc</p> <p>Z score =</p> $Z_x = \frac{X - M_x}{s_x}$ <p>(Z scores have a mean of zero and a standard deviation of 1.)</p> <p>Z_{nd} = the value required for significance at the desired level in the normal distribution tables.</p>
<p>Base rate estimation</p>	$BR = (AS - FP) / (TP - FP)$ <p>Where:</p> <p>BR = estimated base rate AS = proportion in your population of assessed patients who obtain an abnormal score FP = false positive rate (as proportion) TP - true positive rate (as proportion)</p>

Bayes Theorem – adapted for use with test scores.
 Gives the probability that somebody who obtains an abnormal score on a test is in fact abnormal.

$$p_{D/T} = (p_D \times p_{T/D}) / ((p_D \times p_{T/D}) + (p_N \times p_{T/N}))$$

Where:

X/Y = probability of X **given that** Y is known to exist etc.

D is the condition

N is the absence of the condition

T is a test result indicating the presence of the condition

p_D is the base-rate for the condition

p_N equals (1 - p_D)

$p_{T/D}$ is the true positive rate

$p_{T/N}$ is the false positive rate

Coefficient alpha

$$r_{kk} = \frac{k}{k-1} \left(1 - \frac{\sum \mathbf{s}_i^2}{\mathbf{s}_t^2} \right)$$

or

$$r_{kk} = \frac{k \bar{r}_{ij}}{1 + (k-1) \bar{r}_{ij}}$$

Where:

r_{kk} = the reliability of a k item test

k = the number of items

\bar{r}_{ij} = the mean inter-item correlation coefficient

\mathbf{s}_i^2 = the variance of an item

\mathbf{s}_t^2 = the variance of the whole test

<p>Converting scores from one scale to another – raw scores</p>	$X_{new} = \frac{S_{new}}{S_{old}} \times (X_{old} - \bar{X}_{old}) + \bar{X}_{new}$
<p>Converting scores from one scale to another – using Z scores to find new raw score</p>	$X_{new} = \bar{X}_{new} + (Z_{old} \times S_{new})$ <p>Note that $(Z_{old} \times S_{new})$ can be a positive OR a negative value</p>
<p>Converting scores from one scale to another – Z-scores</p>	$Z_{new} = Z_{old}$
<p>Correlation coefficient: biserial r</p>	$\frac{M_1 - M_0}{S_t} \times \frac{pq}{y_o}$ <p>where:</p> <p>M_1 = the mean continuous variable score (test score) of the higher group on the dichotomised variable (in this case the mean test score of the older group)</p> <p>M_u = the mean continuous variable score of the lower group on the dichotomised variable</p> <p>p = the proportion of all the cases who are in the higher group</p> <p>q = the proportion of cases who are in the lower group.</p> <p>y – the ordinate of the normal curve at the point which divides p from q</p> <p>S_t = the standard deviation of all the scores on the continuous variable</p>

<p>Correlation coefficient: Pearson Product Moment – raw scores</p>	$r_{xy} = \frac{\sum (X - M_x)(Y - M_y)}{N s_x s_y}$
<p>Correlation coefficient (Pearson Product Moment) – Z scores</p>	$r_{xy} = \frac{\sum Z_x Z_y}{N}$
<p>Correlation coefficient: phi coefficient</p>	$r_{phi} = \sqrt{\frac{c^2}{N}}$ <p>or</p> $\frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$
<p>Correlation coefficient: point-biserial r</p>	$r_{p.bis} = \frac{M_1 - M_0}{s_t} \times \sqrt{pq}$ <p>where:</p> <p>M_1 = the mean score of those in one category of the dichotomised variable</p> <p>M_0 = the mean score of those scoring in the other category</p> <p>p = the proportion scoring in the first category</p> <p>q = the proportion scoring in the other category.</p> <p>s_t = the standard deviation of all the scores on the continuous variable</p>

<p>Correlation ratio – useful as a measure of relationship when regression is non-linear</p>	$h = \sqrt{\frac{BetweenSS}{TotalSS}}$
<p>Correlation: Coefficient of alienation – just as r_{xy} is an index of the degree of relationship, the coefficient of alienation k_{xy} is an index of the lack of relationship between two variables</p>	$k = \sqrt{1 - r^2_{xy}}$
<p>Correlation: Coefficient of determination = proportion of variance in one variable accounted for by its correlation with another variable</p>	r^2_{xy}
<p>Covariance</p>	$s_{xy} = \frac{\sum xy}{N}$ <p>or</p> $r_{xy} s_x s_y = \frac{\sum xy}{N}$
<p>Differences between test scores: abnormality of a difference between scores on different tests</p>	$Z_{diff} = \frac{Z_x - Z_y}{\sqrt{(2 - 2r_{xy})}}$

<p>Differences between test scores: abnormality of a difference between a predicted and an obtained score</p>	$Z_{diff} = \frac{Z_y - \hat{Z}_y}{\sqrt{(1 - r_{xy}^2)}}$ <p>where:</p> $\hat{Z}_y = r_{xy} Z_x$
<p>Differences between test scores: reliability of a difference between two scores on different tests</p>	$Z_{diff} = \frac{Z_x - Z_y}{\sqrt{2 - (r_{xx} + r_{yy})}}$
<p>Differences between test scores: reliability of difference between 2 individuals on the same test</p>	$Z_{diff} = \frac{Z_{x_1} - Z_{x_2}}{\sqrt{2 - 2r_{xx}}}$
<p>Differences between test scores: the Crawford, , Howell, and Garthwaite modification of the Payne and Jones formula. Use when comparing an individual's score with that of a smallish control group. Use tables for t to estimate the probability that the observed difference is abnormal.</p>	$t = \frac{Z_x - Z_y}{\sqrt{(2 - 2r_{xy}) \left(\frac{N_2 + 1}{N_2} \right)}}$ <p>where:</p> <p>N_2 is the number of cases in the group with whom the individual is being compared.</p>

Differences between test scores: the Crawford and , Howell modification of the Payne and Jones formula for the significance of a difference between a predicted and an obtained score. Use when the regression equation is based on small N, Use tables for t to estimate the probability that the observed difference is abnormal.

$$t = \frac{Z_y - \hat{Z}_y}{\sqrt{1 - r_{xy}^2} \times \left(\sqrt{1 + \frac{(1 + Z_o^2)}{N}} \right)}$$

X_0 signifies the new score that we are using to predict Y from X.

Differences between test scores: Grubbs test for an outlier. Use Grubbs tables for T to assess the probability that the outlying score differs from that of the larger group.

$$T = \frac{(X - M_{x+})}{S_{x+}}$$

The + subscript in the case of both the mean and the standard deviation is there to indicate that both should be based on the original group plus the suspect score. In addition the standard deviation should be an estimate of the population standard deviation (i.e. divided by N - 1, not just N)

<p>Kappa coefficient: measure of agreement between two diagnosticians or other raters.</p>	$\frac{\textit{observed agreement} - \textit{chance agreement}}{1 - \textit{chance agreement}}$ <p>or</p> $\mathbf{k} = \frac{p_o - p_c}{1 - p_c}$ <p>where:</p> <p>p_o = observed agreement</p> <p>and</p> <p>p_c = chance agreement</p>
<p>Kurtosis</p>	$ku = \frac{\Sigma x^4 / N}{s_x^4} - 3$
<p>Normal distribution</p>	$y = \frac{1}{s\sqrt{2\pi}} e^{-(X-M)^2 / 2s^2}$
<p>Regression equation – raw score</p>	$\hat{Y} = r_{xy} \frac{s_y}{s_x} (X - M_x) + M_y$
<p>Regression equation – Z score</p>	$\hat{Z}_y = r_{xy} Z_x$

Reliability coefficient	$r_{xx} = \frac{s_t^2}{s_x^2}$
Reliability: correlation between true (T) and obtained (X) scores	$r_{tx} = \sqrt{r_{xx}}$
Reliability: effect of shortening or lengthening test.	$\hat{r}_{xx} = \frac{mr_{xx}}{1 + (m-1)r_{xx}}$ <p>where: m = (length of modified test) divided by (length of original test)</p>
Reliability: effects of a change in the reliability of both predictor and criterion on validity	$\hat{r}_{xy} = r_{xy} \sqrt{\frac{r'_{xx} r'_{yy}}{r_{xx} r_{yy}}}$ <p>where: r'_{xx} and r'_{yy} are the changed reliability coefficients</p>
Reliability: effects of a change in the reliability of both predictor and criterion on validity if only one reliability coefficient changes	$\hat{r}_{xy} = r_{xy} \sqrt{\frac{r'_{xx}}{r_{xx}}}$

<p>Reliability: effects of reliability on validity, 1. Highest possible validity coefficient obtainable with tests of given reliabilities</p>	$r_{xy}(\text{max}) = \sqrt{r_{xx}} \sqrt{r_{yy}}$ <p>where</p> $r_{xy}(\text{max}) = \text{highest possible correlation between X and Y}$
<p>Reliability: effects of reliability on validity, 2. What would the correlation between a predictor and a criterion be if both were error free.</p>	$\hat{r}_{XY} = \frac{r_{xy}}{\sqrt{r_{xx}} \sqrt{r_{yy}}}$ <p>where \hat{r}_{XY} is the error free correlation.</p>
<p>Reliability: predicting a re-test score – raw scores</p>	$r_{xx}x + M_x$ <p>Standard error = $s_x \sqrt{1 - r_{xx}^2}$</p>
<p>Reliability: predicting a re-test score – Z scores</p>	$r_{xx}Z_{x1}$ <p>Standard error = $\sqrt{1 - r_{xx}^2}$</p>
<p>Reliability: predicting the obtained score from the true score – raw scores</p>	$\hat{X} = T$ <p>Standard error = $s_x \sqrt{1 - r_{xx}}$</p>

<p>Reliability: predicting the obtained score from the true score – Z scores</p>	$Z_x \sqrt{r_{xx}}$ <p>Standard error = $\sqrt{1 - r_{xx}}$</p>
<p>Reliability: predicting the true score from the obtained score – raw scores</p>	$\hat{T} = r_{xx} (X - M_x) + M_x$ <p>Standard error = $s_x \sqrt{r_{xx} (1 - r_{xx})}$</p>
<p>Reliability: predicting the true score from the obtained score – Z scores</p>	$Z_t = Z_x \sqrt{r_{xx}}$ <p>Standard error = $\sqrt{(1 - r_{xx})}$</p>
<p>Reliability: the range within which the true score will lie. 1. Traditional method.</p>	$Range = X \pm z_{nd} s_x \sqrt{(1 - r_{xx})}$ <p>Where: Z_{nd} is value required for significance at the desired level in the normal distribution tables.</p>
<p>Reliability: the range within which the true score will lie. 2. Modern method</p>	$(r_{xx} x + M_t) \pm z_{nd} (s_x \sqrt{r_{xx} (1 - r_{xx})})$ <p>Where: Z_{nd} is value required for significance at the desired level in the normal distribution tables.</p>

Skew	$Sk = \frac{\Sigma x^3 / N}{s_x^3}$
Standard deviation	$s_x = \sqrt{\frac{\Sigma (X - M_x)^2}{N}}$
Standard error of estimate/prediction. (This is the standard deviation of the Y scores around the predicted Y score.) Raw score formula	$s_y \sqrt{(1 - r^2_{xy})}$
Tchebycheff's inequality	$probability(Z_{score} \geq k) \leq \frac{1}{k^2}$ <p>This translates into:</p> <p>The probability of getting a Z score of equal to or greater than a stated value (k), is less than or equal to 1 divided by k^2.</p>