ISSN: 2754-4737

Journal of Neurology Research Reviews & Reports

SCIENTIFIC Research and Community

Research Article

Open Access

Better Use of Scales as Measuring Instruments in Mental Disorders

Satyendra Nath Chakrabartty

Consultant, Indian Ports Association, New Delhi, India

ABSTRACT

Objective: Linearity implies high correlation but the converse is not true. For meaningful application of correlation and other descriptive and inferential analysis, checking of linearity, and assumptions of correlation including normality are needed The paper describes method of converting ordinal scores from Likert/ Rating scale to continuous, monotonic, equi-distant scores with fixed zero point and following normal distribution.

Method: The method involves selection of weights for different response categories of different items so that weighted item score forms an arithmetic progression. Normalization of such score followed by further weights to items to ensure equal item-total correlation justifies addition of such converted scores.

Results: Converted scores satisfying many desired properties can assess progress/deterioration of a patient over time and facilitate comparison, ranking, classification and assessing effectiveness of treatment plan. It also helps in computing reliability of Likert or Rating scale avoiding assumptions of Cronbach's alpha.

Conclusions: Converted scores will help the researchers and practitioners to find improved content validity and meaningfully undertake correlation and other analysis under parametric set up for deriving useful and valid conclusions about the sample, population and test parameters.

***Corresponding author**

Satyendra Nath Chakrabartty, Consultant, Indian Ports Association, New Delhi, India. Tel: 917838333652; E-mail: snc12@rediffmail.com

Received: May 26, 2020; **Accepted:** July 17, 2020; **Published:** August 31, 2020

Keywords: Equi-Distant, Normality, Correlation, Validity, Reliability

Introduction

Various scales are being used as instruments for assessment and diagnostic purposes in mental disorders. Studies on mental disorder used methods like standardized questionnaires, clinical examinations, observational methods, registers, secondary data, and qualitative interviews. Most researchers agree that well-validated standardized questionnaires are necessary tools in research, as well as in practical prevention.

Evaluation may include measurement of variables which are objective like goniometric measurement, manual muscle testing, sensory evaluations, etc or subjective often consisting of self-report outcome measures on perception of symptoms, disabilities, emotional function, participation in daily activities, etc. Assessments of such variables are often done using Likert scale or Rating scale. A 5-point Likert scale has ordered response levels marked as 1, 2, 3, 4, 5 or 0 to 4 or -2,-1, 0, 1, 2. Response categories of a typical 5-point Likert item are: (1) (1) Strongly Disagree, (2) Disagree, (3) Neither agree nor disagree (undecided or neutral), (4) Agree, and (5) Strongly agree. For accurate and consistent evaluation of variables, descriptive and inferential statistics including concepts of validity, reliability, responsiveness are used for inter-individual or inter-group comparisons.

Subjective reports of sleep in Likert scale may report good sleep despite disturbed sleep pattern when monitored physiologically. Similarly, there are people who report disturbed sleep, but show normal sleep patterns when objectively monitored [1]. The most reliable indicator of sleep disturbance is self-reported global sleep dissatisfaction [2]. However, self-reported disorders may not capture isolated, short-lived cases or mild cases requiring early treatments [3].

Outcomes of analysis of data emerging from such scales influence highly sensitive and important areas like patient care, policy issues, etc. Thus, it is necessary to investigate methodological issues of scoring scales for mental disorder and properties of such scores like monotonically increasing continuous variables along with their responsiveness, discriminating power (ability to discriminate persons with problems from those without such problems), reliability, sensitivity (accuracy of the tool in identifying a problem), specificity (identification of individuals who do not have a problem), etc. Analysis of data emerging from such scales depend significantly on nature of data, types of variables being assessed, admissibility of operations and hence, type of analysis are different if the measurement scales are nominal or ordinal or interval/ratio levels. Consideration of ordinal discrete data as interval or ratio data and application of techniques like correlation, regression, reliability analysis and inferences, without verification of associated assumptions of the techniques may lead to invalid and inconsistent findings.

Major limitations of Likert scale and Rating scale are levels are not equidistant and addition is not admissible. Equal psychological distance between categories will provide exact measurements of the psychological trait being assessed [4]. To define meaningfully the operation "addition", one needs to ensure that distance between (Strongly Disagree and Disagree) (d_{12}) = distance between (Disagree and Neither Agree nor Disagree) (d_{23}) = distance between (Neither Agree nor Disagree and Agree) (d_{34}) = distance between (Agree and Strongly Agree) (d_{45}) and also $d_{13}=2d_{12}$ and so on. This is possible if item scores are taken as weighted sum so that 1W1, 2W1,3W1,4W1 and 5W_1 are in arithmetic progression. The distance measure for the ordinal data cannot be defined unless the ordinal to interval variable conversion is used [5]. For ordinal data, mean and standard deviation (SD) are inappropriate measures [6].

Self-completed LANSS (S-LANSS) scale to identify pain of predominantly neuropathic origin (POPNO) on the basis of the patient's current symptoms and signs and a cutoff score, rather than perform as a measurement scale. The scale consists of 7 binary items, each with two response categories (yes or no) to the presence of symptoms (5 items) or clinical signs (2 items) [7]. S-LANSS score is a summative score and assumes that the 7 items are of different importance. If importance of i-th item is denoted by I_i, then the scale assumes $I_1 = I_2 = I_6 = 5\ I_5$; $I_3 = I_7 = 3I_5$; $I_4 = 2I_5$; which implies $I_4 + I_5 = I_3 = I_7$; $I_3 + I_4 = I_1 = I_2 = I_6$.

While feeling of burning sensations in the painful areas or a sudden temperature change – is considered to have least importance, unpleasant sensations such as tingling, pricking or pins and needles - is having maximum importance. Importance of the latter is equal to the same of experience of different skin aspect in the painful areas, i.e. skin redder than usual or appearing mottled (Item 2). All these need to be verified and tested before generalization.

Score of 12 can be obtained by several ways as can be seen from the following hypothetical example at Table – 1.

Table 1: S-LANSS score of 12 by different ways

Patient	Item-1	Item-2	Item-3	Item-4	Item-5	Item-6	Item-7	Total
1	5	5	0	2	0	0	0	12
2	5	0	0	2	0	5	0	12
3	0	5	0	2	0	5	0	12
4	0	5	3	0	1	0	3	12
5	0	0	3	0	1	5	3	12
6	5	0	3	0	1	0	3	12
Mean	2.5	2.5	1.5	1	0.5	2.5	1.5	12
Variance	7.5	7.5	2.7	1.2	0.3	7.5	2.7	0

Thus, S-LANSS score fails to discriminate patients with a particular total score. In addition, distribution of item score is different for different items and Cronbach's alpha is not defined for the group of patients with equal S-LANSS score.

Kurtzke's expanded disability scale (EDSS) is a popular rating scale in treatment trials of multiple sclerosis (MS), generates ordinal scores is less able to differentiate among individuals and detect change in disability over time in comparison to other ''disability'' scales. Like EDSS, Ashworth scale for spasticity Modified Rankin Scale (mRS) as a functional outcome measure for post-stroke patients are single item scales and not well represent well the broad scope of the domain in consideration. Multi-item rating scale like Multiple Sclerosis Walking scale (MSWS-12) creates difficulties in allowing each item to act as a rating scale and combining the items cancels out the random error of each single item, giving a high value of reliability [8-10].

For rating scale, often inter-rater agreement (IRA) and inter-rater reliability (IRR) are not distinguished [11, 12]. IRR are computed by different approaches like percentages of agreement (PA), kappa, weighted kappa and naturally results in different values and conflicting results of IRR. IRA values differed for different levels. Strong evidence of systematic difference among the raters was observed [13, 14].

The above motivates need to identify major limitations of Likert type scale, rating scale with usual summative scoring and statistical analysis being reported without verification of associated assumptions along with remedial measures.

Rest of the paper is organized as follows. Issues relating to correlations and regression along with remedial action for meaningful use of the same are discussed in the following section. This is followed by method of converting ordinal Likert/Rating score to continuous, eqi-distant monotonic score following normal distribution and properties thereof. The paper is rounded up in Section 3 by recalling the salient outcomes and emerging suggestions.

Problem areas

Correlations and regressions

Analysis involving correlations are frequent to find relationship between a pair of variables or to investigate cause and effect pattern or to know the functional architecture of neuronal networks etc. Correlation coefficient or product moment correlations are frequently used in many studies, including multivariate statistical procedures—such as multiple regression, principal component analysis (PCA), factor analysis (FA), path analysis, structural equation modeling, etc. Cardiovascular disorders (CVDs) have associations with neurological disorders (NDs) like Alzheimer's disease (AD) and Parkinson disease (PD) and suggested further works to find the exact linking point among these diseases and hence, verifying assumptions and interpretations of correlations are important. 13 ways of interpreting a correlation was described. Another interpretation of correlation as the proportion of matches was suggested [15, 17].

Reliability of rating scales is usually expressed as inter-observer and intra-observer agreement, primarily to reflect consistency of different observers and computed by kappa (κ) statistic where $\kappa \in [-1,1]$. Value of $\kappa = -1$ imply complete disagreement and +1 for perfect agreement. However, agreement among raters and scale reliability are different concepts. Assignment of a particular level/

rank to each patient by all the raters indicates perfect agreement and zero scale reliability since between–patients variance is zero. Other measures of agreement among raters are weighted kappa, percentages of agreement (PA), standard error of measurement for an individual ($SEMX_i$). They differ in properties and result in different values and conflicting results of Inter Raters Reliability. Values of Inter-rater agreement (IRA) differ for different levels. Coefficient of variation (*CV*) was propose as a measure of agreement among raters [18].

Reliability of Likert scales are popularly given in terms of Cronbach's α. For a scale consisting of n-items,

$$\alpha = \left(\frac{n}{n-1}\right)\left(1 - \frac{Sum\ of\ item\ variances)}{Test\ variance}\right)$$

Assumptions of Cronbach's α include among others uni-dimensionality. Despite presence of several factors, many studies used Cronbach's alpha as the measure of test reliability ignoring assumptions of Cronbach's alpha, which increases with increase in number of items.

Validity is reported according to criterion, construct, and content. Criterion validity is the correlation of test score with the criterion score (score of another test developed for similar purposes or a standard, often referred as "Gold Standard"). However, there is no unique gold standard for stroke. Neuropathology criterion is rarely available, since stroke is seldom lethal in the acute phase. Correlation between test score and criterion score (r_{XC}) could also be interpreted as validity of the criterion score. A high value of r_{XC} may not justify need of the test. Moreover, such correlations may be influenced by many factors including shape of the distributions of the two scores. Validity of Pittsburgh Sleep Quality Index (PSQI) in terms of correlation with polysomnography (PSG) measures were low, presumably due to non- satisfaction of assumptions of correlation and different distribution of test score and criterion score [19-21].

Correlation coefficient is taken as degree of linearity between two variables. If correlation between X and Y (r_{XY}) is high i.e.$|r_{XY}|\approx 1$, the variables are usually taken as linearly related and attempt is made to establish linear regression of the form $Y=\alpha 1+\beta 1X$ or $X=\alpha_2+\beta_2 Y$ for prediction of dependent variable. Despite high correlation coefficient ($r > 0.997$), among the investigated curves the straight-line model was rejected at the 95% confidence level on the basis of the Lack-of-fit and Mandel's fitting test [22]. In fact, $|r_{XY}|\approx 1$ even if X and Y are related by a non-linear fashion. For example, if X takes integer values from 1 to 30, correlation between X and several non-linear function of X are shown in Table – 2.

Table 2: Correlation between X and non-linear function of X

	X	X^2	$1/X$	X^3	$\log^X_{10}$	**Cos X**	**Sin X**
X	1	0.97029	-0.64789	0.92011	0.92064	-0.97156	0.99982
X^2		1	-0.50445	0.98629	0.81179	-0.99998	0.96559
1/X			1	-0.42219	- 0.87699	0.50689	-0.65623
X^3				1	0.72716	-0.98529	0.91251
$\log X_{10}$					1	-0.81425	0.92630
Cos X						1	-0.96696
Sin X							1

Observations

- If X is increasing and Y is decreasing, r_{XY} is negative.
- High correlation does not imply linearity. In other words, correlation may not be a useful indicator of linearity. Scatter plot may throw more light on linearity and validity of linear regression line.
- Residual plots give useful information on the chosen regression model. The residual plot can also be used to check whether the underlying assumptions, like normality of the residuals and homo-scedasticity are satisfied for evaluating the goodness of fit of the regression model. The U-shaped residual plot implies better fit for a curvilinear regression model.
- To ascertain, linearity, testing of significance of standard error of prediction could be better than testing significance of correlation coefficient.

Take another example where X~ N (0, 1) and

$Y = \frac{1}{\sqrt{2\pi}}\, e^{\frac{-1}{2}X^2}$ i.e. Y is the ordinate of N (0, 1). Clearly, X and Y are not linear. Consider:

Case – 1: Here, $0 \le X \le 3.9$ then $r_{XY} = -0.93302$.
Case – 2: Here, $-3.9 \le X \le 3.9$ results in $r_{XY} = 0.00036$.

Interpretation of r_{XY} from Case – 1 is X and Y are highly correlated but correlation is negative i.e. increase of one unit in X will result in decrease of Y and vice versa. However, interpretation of r_{XY} from the Case – 2 will be just reverse. Low value of $r_{XY}=0.00036$ tends to indicate that X and Y are independent, which is not the case in reality. In Case – 1, r_{XY} increased due to consideration of restricted range of values of X. In other words, truncated values of one or more variables may not give true relationships between two variables. However, many studies in social science in general and neurological studies in particular, involve variable taking positive values only and assuming that it follows normal distribution and investigate relationship of X with other variables. PCA will ignore r_{XY} in Case – 2 unlike the same in Case – 1.

Summated score of PSQI with seven components and 19 items had different structure validity for different age groups, which require different models to assess sleep disturbance across age groups [23]. This could happen due to various reasons including different score range of PSQI for different age groups.

The problem of truncated values may also occur if we want to find correlation between height and weight of students of say Class V. Here, r_{XY} will be poor, primarily due to range restriction of both the variables and also due to high homogeneity of the sample. Similarly, correlation between SAT scores and undergraduate grade point average (GPA) could be as low as 0.20. This is due to small range of SAT scores of students admitted to the colleges and universities. Similarly, validity of selection test as a correlation between test scores and job performance is poor since range of test score is small for the persons selected through the test. Other

factors being equal, a restricted range usually yields a smaller correlation. In other words, in case of high positive skew of the test score(X) and/or criterion score(C), validity as r_{XC} may be lower since the data contains predominantly high performers [24].

Other issues that can affect correlation are the amount of variability in the data, presence of outliers, characteristics of the sample, measurement error, etc.

The above can be summarized as "Linearity implies high correlation but the converse is not true". Question therefore arises on how to know linearity between two variables.

A simple way to know linearity between Y and X is to check whether $\frac{Resulting\ change\ in\ Y}{Small\ change\ in\ X}$ = constant for all values of X. In other words, one may check for constant slope of the straight line connecting X and Y by considering $\frac{Y_i - Y_{i+1}}{X_i - X_{i+1}}$. and checking whether the ratio is constant for all values of i. If yes, $\frac{Y_i - Y_{i+1}}{X_i - X_{i+1}}$ can be taken as slope of the straight line (β). Computation of $\frac{Y_i - Y_{i+1}}{X_i - X_{i+1}}$ for different functions of X are shown in the Table – 3

Table 3: Checking of $\frac{Y_i - Y_{i+1}}{X_i - X_{i+1}} = k$ for different $Y=f(X)$

$1 \le X \le 30$		$Y=X^2$		$\frac{Y_i - Y_{i+1}}{X_i - X_{i+1}}$	Observation
Xi	$Xi+1$	Yi	$Yi+1$		
1	2	1	4	3	$Y = X^2$ is not linear ($r_{XY} = 0.97$)
10	11	100	121	21	
	Y= 1/X				
1	2	1	0.5	0.5	$Y = 1/X$ is not linear ($r_{XY} = -0.65$)
10	11	0.1	0.090909	0.01389	
	$Y = log X_{10}$				
1	2	0	0.30103	0.30103	$Y = log_{10}^{X}$ is not linear ($r_{XY} = 0.92$)
10	11	1.0	1.04139	0.04139	
	$Y = \text{Cos}X$				
1	2	0.99985	0.99939	0.00046	$Y = CosX$ is not linear($r_{XY} = -0.97$)
10	11	0.98481	0.98163	0.00318	
$0 \le X \le 3.9$		$Y = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}X^2}$			$Y = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}X^2}$ is not linear($r_{XY} = 0.93$)
0.1	0.2	0.397	0.391	- 0.06	
1.1	1.2	0.2179	0.1942	- 0.237	

Alternatively, linearity can be tested by first fitting a linear regression line of the form (say) $Y = \alpha + \beta X + \epsilon$ followed by finding predicted values of Y as $\hat{Y}$ and then testing significance of standard error $S_E = \sqrt{\frac{1}{n}\Sigma(Y_i - \hat{Y}_i)^2}$ where n denotes number of observations.

For visual purpose, residual plots may help.

Suggested action points

- Convert item score (X) and also test score (Y) so that each becomes continuous satisfying (i) equidistant property (ii) normality condition to ensure that scores are not skewed (iii) no outliers i.e.no value beyond [*mean* ±3.29(SD)].
- Test for linearity between X and Y by testing H_0: S_E=0 against H_1: $S_E \neq 0$.
- For criterion validity, ensure above mentioned two conditions before calculating r_{XC}.

Conversion of Likert scores

Method of converting Likert items with equal number of response categories to ratio scale suggested by is briefly reproduced below [25].

1. Assign 1, 2, 3, 4, 5, and so on to the levels of Likert items avoiding zero.
2. Convert each negative item to be positively related to the test score.
   Let the discrete variable X_{ij} denote the raw score of the i-th person in the j-th item of a Likert scale, for i=1,2,….,n and j=1,2,……,m. For 5-point scale, X_{ij} takes value 1, 2, 3, 4, 5.
3. Find weights W_ij's for different levels for different items so that Wij>0, $\sum_{j=1}^{5} W_{ij} = 1$ 1 for each item so that $W_1, 2W_2, 3W_3, 4W_4, 5W_5$ forms an Arithmetic Progression. A positive value of the common difference will ensure $5W_5 > 4W_4 > 3W_3 > 2W_2 > W_1$

One way to find such weights are:

Let f_{ij} be the frequency of i-th item for the j-th level. For each item, find maximum (f_{max}) and minimum frequency (f_{min}).

Find proportions: $\omega_{ij} = \frac{f_{ij}}{n}$.

b. Put initial weights: $W_{i1} = \omega_{i1} = \frac{f_{min}}{n}$. Find the common difference $= \frac{5f_{max} - f_{min}}{4n}$.

c. Define $W_{i2} = \frac{\omega_{i1} + \alpha}{2}$; $W_{i3} = \frac{\omega_{i1} + 2\alpha}{3}$; $W_{i4} = \frac{\omega_{i1} + 3\alpha}{4}$ and $W_{i5} = \frac{\omega_{i1} + 4\alpha}{5}$

Here, $W_{ij} > 0$ and $\sum_{j=1}^{5} W_j \neq 1$.

d. Get final weights $W_{ij(Final)} = \frac{W_{ij}}{\sum_{j=1}^{5} W_j}$ so that $\sum W_{ij(Final)} = 1$

4. Normalize the scores obtained by above mentioned fashion using $Z = \frac{X - \bar{X}}{SD(X)}$ so that the normalized scores follow $N(0, 1)$.

5. Take further weights to items to satisfy additional property of making the test scores equi-correlated with the items i.e. equal item reliability and thus justify addition of such converted item scores.

6. Note that if frequency of a particular level of an item is zero, the method may fail and can be taken as zero value for scoring items as weighted sum

Benefits

Converted scores are continuous satisfying equidistant property with a fixed zero point and following normal distribution have the following advantages:

1. Generate monotonic scores since choice of j-th level will result in higher score than the choice of (j-1)-th level for any item for j=2, 3, 4, 5
2. Facilitate ranking group of patients uniquely avoiding ties unlike ranks from usual summative Likert scores. Possible to find sample mean and SD for a group of patients.
3. Assess progress/deterioration of a patient over time points. If X_{it} denotes severity of the i-th patient in t-th time period, then $\frac{X_{it} - X_{i(t-1)}}{X_{i(t-1)}} \times 100$ will indicate percentage of progress or deterioration made by the i-th patient in t-th time in comparison to the previous time period. Thus, the ratio $\frac{X_{it} - X_{i(t-1)}}{X_{i(t-1)}}$ reflects responsiveness of the scale and evaluate the effectiveness of a treatment plan.
4. Path of improvement/decline of one or a group of patients during the time point 1, 2, and so on can be drawn which may facilitate drawing useful conclusions including better prognostication.
5. Test reliability as a function of item reliabilities, avoiding assumptions of Cronbach alpha can be obtained as follows:

$$r_{tt} = \frac{\sum_{i=1}^{m} r_{tt(i)} S_{Xi} + \sum_{i=1, i \neq jk}^{m} \sum_{j=1}^{m} 2COV(X_i, X_j)}{\sum_{i=1}^{m} S_{Xi} + \sum_{i=1, i \neq jk}^{m} \sum_{j=1}^{m} 2COV(X_i, X_j)}$$

6. Where $r_{tt(i)}$ and S_{xi} denote respectively reliability and sample SD of the i-th item.
7. Denoting rank received by the i-th person from the j-th rater in a ranking scale by X_{ij} and following the procedure given in 2.2 above, it is possible to convert raw ranks to interval scale and find reliability of the scale. In addition, agreement among the raters could be found by coefficient of variation (CV) where higher $CV \Longrightarrow$ more variability. CV is preferred than the SEM in assessing IRA [26].

Clustering and threshold values

Scales for the purpose of diagnosis, decide a cut-off score (threshold value) (X_0) such that persons with scores less than X_0 are normal and persons with scores $\geq X_0$ are taken to be suffering from the disorder. Example: S-LANSS score with cut-off score of 12. For the Insomnia Severity Index (ISI], persons scoring < 14 are considered as Normal and those scoring ≥ 14 are considered as having insomnia. In such cases, scores of normal persons should be similar to each other (homogeneity) but persons belonging to other cluster may have higher variability depending on intensity of the concerned disorder.

However, scales for assessing degree of disorderness find a few cut-off scores and cluster the persons in K number of mutually exclusive classes. There are a number of measures to reflect efficiency of such clustering. Measures of dissimilarity (distance) can be easily defined for the proposed method which converts ordinal data to interval/ratio variables. The Euclidean, Manhattan, Maximum, Minkowski, Mahalonobis, Average, Chord, Canberra, Czekanowski distances, etc. could be used in this case. Davies-Bouldin Index (DBI) was found best among other cluster validity indices [27, 28]. Computation of DBI index for K-classes requires calculation of mean, variance, maximum and minimum value for each class and uses them to the definition of DBI as

$DB_K = \frac{1}{K}\sum_{=1}^{K} j = 1.2 \ldots\ldots k, i \neq j \; Max[\frac{diam(C_i)+diam(C_j)}{\|C_i - C_j\|}]$ where diameter of a cluster/class is defined as: $diam\ (C_i) = \sqrt{\frac{(\sum_{s \in C_i} \|x - C_i\|^2)}{n_i}}$ where

n_i : Number of members in the *i*-th class and C_i: Centroid (or mean) of i-th cluster Upper limit of DBI is 1 and lower value implies better efficiency

Conclusions

The paper proposes method of converting ordinal Likert scales or rating scales to ratio scales. Scores generated by the method were continuous, monotonic, satisfying equi-distant property and normality. Such scores also assess progress/deterioration of a patient over time and facilitate comparison, ranking, classification and assessing paths of improvement and effectiveness of treatment plan. They can be better used to find mental health scores (Y) of an individual or a group of individuals for a uni-dimensional tool or domain scores for multi-dimensional tool. It also helps in computing reliability of Likert or Rating scale avoiding assumptions of Cronbach's alpha.

With checking of linearity between two such variables, sample characteristics and associated score ranges, the researchers and practitioners can find improved content validity of the scale and meaningfully undertake correlation analysis and subsequent descriptive and inferential statistics under parametric set up for deriving useful and valid conclusions about the sample, population and test parameters.

Empirical investigations with multi data sets are suggested for future studies to explore properties of the proposed measure along with invariance of factor structure [17].

Highlights

- Linearity implies high correlation but the converse is not true. Hence, checking of linearity and assumptions of correlation including normal distribution of variables for meaningful application of correlation and other descriptive and inferential analysis are needed.
- Describes method of converting ordinal scores from Likert/ Rating scale to continuous, monotonic, equi-distant scores with fixed zero point and following normal distribution.
- Converted scores satisfying many desired properties can assess progress/deterioration of a patient over time and facilitate comparison, ranking, classification and assessing effectiveness of treatment plan. It also helps in computing reliability of Likert or Rating scale avoiding assumptions of Cronbach's alpha.
- Converted scores to help researchers and practitioners to find improved content validity and meaningfully undertake correlation and other analysis under parametric set up for deriving useful and valid conclusions about the sample, population and test parameters.

References

1. Edinger JD, Fins AI, Glenn DM, Sullivan RJ Jr, Bastian LA et al. (2000) Insomnia and the eye of the beholder: Are there clinical markers of objective sleep disturbances among adults with and without insomnia complaints? J Consult Clin Psychol 68: 586-593.
2. Ohayon M M, Riemann D, Morin C, Reynolds III C F (2012) Hierarchy of insomnia criteria based on daytime consequences. Sleep Medicine 13: 52-57.
3. Tannenbaum C, Lexchin J, Tamblyn R, Romans S (2009) Indicators for Measuring Mental Health: Towards Better Surveillance, Healthc Policy 5: 177-186
4. Wakita T, Ueshima N, Noguchi H (2012) Psychological Distance between categories in the Likert Scale: Comparing Different Numbers of Options, Educational and Psychological Measurement 72: 533-546.
5. Marateb HR, Mansourian M, Adibi P, Farina D (2014) Manipulating measurement scales in medical statistical analysis and data mining: A review of methodologies, J Res Med Sci. 19: 47-56.
6. Jamieson S. (2004): Likert scales: how to abuse them. Blackwell Publishing Ltd Medical Education 38: 1212-1218.
7. Bennett M I, Smith B H, Torrance N, Potter J (2005) The S-LANSS Score for Identifying Pain of Predominantly Neuropathic Origin: Validation for Use in Clinical and Postal Research, The Journal of Pain 6: 149-158.
8. Hobart J (2003) Rating scales for Neurologists, J Neurol Neurosurg Psychiatry 74: 22-26.
9. Urban PP, Wolf T, Uebele M, Marx JJ, Vogt T et al (2010) Occurrence and clinical predictors of spasticity after ischemic stroke. Stroke 41: 2016-2020.
10. Banks JL, Charles A Marotta CA (2007) Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials: a literature review and synthesis, Stroke 38: 1091-1096
11. Bruno A, Shah N, Lin C, Close B, Hess DC, et al. (2010) Improving modified Rankin Scale assessment with a simplified questionnaire. Stroke 41:1048–1050.
12. Ghandehari, K (2013) Challenging comparison of Stroke Scales, Journal of Research in Medical Sciences 18: 906-910.
13. Wilson JT, Hareendran, A Hendry, A Potter, I an Bone J, Bone et al. (2005) Reliability of the modified Rankin Scale across multiple raters: benefits of a structured interview, Stroke 36: 777-781.
14. Quinn TJ, Dawson J, Walters MR, Lees KR (2009) Reliability of the modified Rankin Scale: a systematic review, Stroke 40: 3393-3395.
15. Firoz C K, Jabir N R, Khan M S, Mahmoud M, Shakil S et al. (2015) An overview on the correlation of neurological disorders with cardiovascular disease; Saudi Journal of Biological Sciences 22: 19-23.
16. Rodgers J L, Nicewander W L (1988) Thirteen ways to look at the correlation coefficient. The American Statistician 42: 59-66.
17. Rovine, M J, von Eye A (1997) A 14th way to look at a correlation coefficient: Correlation as the proportion of matches. The American Statistician 51: 42-46.
18. Petersson H, Gill H, Åhlfeldt H (2002) A variance-based measure of inter-rater agreement in medical databases, Journal of Biomedical Informatics 35: 331-342
19. Bamford J, Dennis M, Sandercock P, Burn J, Warlow CP (1990) The frequency, causes, and timing of death within 30 days of a first stroke: the Oxford shire Community Stroke Project. J Neurol Neurosurg Psychiatry 53: 824-829.

20. Silver FL, Norris JW, Lewis AJ, Hachinski VC (1984) Early mortality following stroke: a prospective review. Stroke 15: 492-496.
21. Manzar MD, Moiz JA, Zannat W, Spence D W, Pandi-Perumal SR (2015) Validity of the Pittsburgh Sleep Quality Index in Indian University Students, Oman Med J 30: 193-202.
22. Loco J V, Elskens M, Croux C, Beernaert H (2002) Linearity of calibration curves: use and misuse of the correlation coefficient, Accreditation and Quality Assurance 7: 281-285.
23. Jia Y, Chen S, Deutz NEP, Bukkapatnam STS, Woltering S (2019) Examining the structure validity of the Pittsburgh Sleep Quality Index, Sleep and Biological Rhythms DOI 10.1007/s41105-018-00201-0.
24. Vaughan E D (1998) Statistics: Tools for understanding data in the behavioral sciences. Upper Saddle River, NJ: Prentice-Hall.
25. Chakrabartty, Satyendra Nath (2019) Limitations of Insomnia Severity Index and possible remedies, JSM Neurological Disorders and Stroke 5: 1-9.
26. Chakrabartty, Satyendra Nath (2019b) Alternate measure of disability intensity: Modified Rankin Scale, Journal of Stroke Medicine 1-8.
27. Davies D, Bouldin W (1979) A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence 1: 224-227.
28. Kim, M, Ramakrishna R (2005) New indices for cluster validity assessment. Pattern Recognition Letters 26: 2353-2363.