

Elisabeth Svensson.
Bestämning av mellan-bedömar reliabilitet
(inter-rater agreement)
eller av test-retest reliabilitet
(stabilitet, intra-rater agreement)

- a. Åskådliggör frekvensfördelningen av parvisa bedömningar i kontingenstabell (huvuddiagonalens orientering analog med kontinuerliga spridningsdiagram)
- b. Bestäm procentuell enighet (percentage agreement) PA. Om $PA < 100\%$ så kan Du analysera orsaken till oenigheten:
- c. Studera marginalfördelningarna. Marginal-heterogenitet är tecken på systematisk oenighet, vilket kan illustreras med en ROC kurva (Q-Q-plot) Tolka resultatet
- d. Beräkna grad av systematisk oenighet i läge (RP) och grad av systematisk oenighet i koncentration (RC) mellan de båda bedömningarna.
- e. Det rang-transformerbara mönstret RTPA, är den fördelning av dataparen, som förväntas utifrån de två marginalfördelningarna. RTPA illustrerar alltså fördelningen av data i kontingenstabellen om all oenighet kunde förklaras av systematisk oenighet (bias) mellan de två bedömningarna.
- f. Jämför den observerade frekvensfördelningen med RTPA. En spridning av observationerna från RTPA betyder att all oenighet kan inte förklaras av den systematiska oenigheten. Ett enkelt mått på denna tillfälliga individuella oenighet är måttet för andelen oordnade par bland alla möjliga olika parkombinationer: "The measure of disorder" D"
- g. Relative Rank Variance (RV) är ett mått på den individuella variationen kring det rang-transformerbara mönstret, dvs den oenighet mellan två bedömningar som inte kan förklaras av en systematisk oenighet (bias).

Tolka:

Om RP och/eller $RC \neq 0$ och om det 95procentiga konfidensintervallet (95% CI) för måttet ($RP \pm 1,96 SE(RP)$) inte täcker värdet 0 så är detta tecken på en säkerställd systematisk oenighet mellan de två bedömningarna.

Om det är två bedömare, som jämförs, så betyder detta att de har olika kriterier för att bedöma. Detta kan tränas bort, diskuteras och alltså justeras så att denna bias inte upprepas. Om det är ett test-retest resultat så kan resultatet vara ett tecken på att något har systematiskt påverkat gruppen. Testsituationen är inte stabil.

Om RV är skilt från noll (beräkna 95% CI för RV på motsvarande sätt som för RP, RC) så är detta tecken på individvariationer som inte fångas upp av den systematiska oenigheten. Det kan finnas flera förklaringar: dåligt formulerad fråga/skalan dålig, tillfälligheter styr bedömningen. Det är svårt att reparera denna typ av brist på reliabilitet.

Lästips:

Svensson E, Holm S. **Separation of systematic and random differences in ordinal rating scales.** Statistics in Medicine 1994;13:2437-53

Svensson E. **A coefficient of agreement adjusted for bias in paired ordered categorical data.** Biometrical Journal 1997;39:643-57

Svensson, E. **Application of a rank-invariant method to evaluate reliability of ordered categorical assessments.** Journal of Epidemiology and Biostatistics, 1998; 3 (4):403-409

Svensson E, Starmark J-E, Ekholm S, von Essen C, Johansson A. **Analysis of interobserver disagreement in the assessment of subarachnoid blood and acute hydrocephalus on CT scans.** Neurological Research 1996;18: 487-94

Gummeson L, Jonsson I, Conner M T, Svensson E. **Assessing factors influencing food choice among 10-16-year-old school boys. A pilot study with a stacking box method.** Journal of Human Nutrition and Dietetics 1996;9(3):219-230

Jonsson I, Gummeson L, Conner M, Svensson E. **Assessing food choice: reliability and predictive validity of a method using food photographs in stacking boxes.** Appetite 1998;30:25-37

Berntson L, Svensson E. **Pain assessment in children with JCA; a matter of scale and rater.** Acta Paediatrica 2001;90:1131-6