

Inhaltsverzeichnis

1 Einleitung	13
1.1 Verschiedene Arten der Reliabilität.....	17
1.2 Hinweise für den Leser zur Bearbeitung des Textes	19
2 Grundlegende Hinweise zum Problem der Interraterreliabilität und -übereinstimmung	23
2.1 Zur Interpretation und Bedeutung der Ausprägung von Gütemaßen der Übereinstimmung und der Reliabilität.....	25
2.2 Ursachen mangelnder Reliabilität und Übereinstimmung.....	30
3 Unterscheidungskriterien für die Interpretation und die Anwendung der verschiedenen Reliabilitäts- und Übereinstimmungskoeffizienten	33
3.1 Übersicht über die Entscheidungskriterien und die in den Kapiteln 4 bis 6 behandelten Koeffizienten.....	40
4 Die Übereinstimmung zwischen Ratern: Maße für mindestens nominalskalierte Daten	45
4.1 Maße für polytome Kategoriensysteme	47
4.1.1 Die prozentuale Übereinstimmung (PÜ)	47
4.1.1.1 Die PÜ für 2 Rater, wenn Nichtkategorisierungen vorliegen	49
4.1.2 Signifikanztestung und zufallskorrigierte Assoziationsmaße, die auf dem X^2 -Wert basieren.....	50
4.1.2.1 Die Signifikanzbestimmung: Der X^2 -Wert für mehr als zwei Kategorien.....	51
4.1.2.2 Die Stärke des Zusammenhangs: Der Kontingenzkoeffizient C und Cramers V.....	54
4.1.3 Zufallskorrigierte Übereinstimmungsmaße: Cohens κ - und Scotts π -Koeffizient	55
4.1.3.1 Cohens κ - und Scotts π für 2 Rater.....	56
4.1.3.2 Orientierungshilfen zur Beurteilung der Güte von κ und π	59
4.1.3.3 Wann sollte Cohens κ und wann sollte Scotts π berechnet werden?.....	60
4.1.3.4 Die Überprüfung der Gleichheit von Randverteilungen	64
4.1.3.5 κ und π für mehrere Rater, wenn alle Objekte von allen Ratern beurteilt wurden.....	67

4.1.4	Berechnung der Übereinstimmungsmaße mit Hilfe von SPSS	67
4.1.4.1	Dateneingabe	68
4.1.4.2	Definition der angeforderten Berechnungen.....	69
4.1.4.3	Interpretation des Auswertungsbogens.....	73
4.1.5	Vertiefung: Modifizierte κ -Koeffizienten	75
4.1.5.1	κ für mehrere Rater, wenn dieselben Objekte von gleich vielen, aber nicht denselben Ratern beurteilt wurden.....	75
4.1.5.2	Übereinstimmung in Bezug auf eine bestimmte Kategorie.....	77
4.1.5.3	Die Übereinstimmung in Bezug auf ein bestimmtes Objekt	78
4.1.5.4	Der gewichtete κ_w -Koeffizient.....	78
4.1.6	Maße der ‚ungefähren‘ Übereinstimmung: Der X^2_{LL} -Wert und T.....	82
4.2	Maße für dichotome Daten.....	87
4.2.1	Die Berücksichtigung von zufälligen Übereinstimmungen bei der PÜ und die kategorienspezifische PÜ	88
4.2.2	Signifikanz und Ausmaß des Zusammenhangs zwischen zwei Ratern in Bezug auf zwei Kategorien: Der X^2 -Wert und der ϕ (Phi)-Koeffizient...	93
4.2.2.1	Korrektur von ϕ und von Cohens κ gemäß der Verteilung der Randsummen	91
4.2.3	Das Problem ungleicher Randverteilungen: Lösungsansätze vor dem Hintergrund der Signalentdeckungstheorie (Signal detection theory)	98
4.2.4	Test auf Gleichheit der Schwellenwerte: Die zusammengefasste PÜ und der McNemar-Test.....	103
4.2.5	Maße, die den Effekt ungleicher Randverteilungen korrigieren	105
4.2.5.1	Odds Ratio und Yules Y	105
4.2.5.2	Die Reliabilität latenter Merkmalsausprägungen: Der tetrachorische Korrelationskoeffizient r_{tet}	108
4.2.6	Berechnung der Maße für dichotome Daten mit Hilfe von SPSS	114
4.3	Zusammenfassung zur Übereinstimmung zwischen Ratern.....	119
5	Maße der Zuverlässigkeit für mindestens ordinalskalierte Ratings	123
5.1	Das Problem des Skalenniveaus von Ratingskalen	123
5.2	Maße der Zuverlässigkeit für ordinalskalierte Ratings.....	127
5.2.1	Test auf Homogenität der Randverteilungen.....	129
5.2.1.1	Der Kolmogorov-Smirnov-Test für zwei Stichproben	129
5.2.1.2	Der Wilcoxon-Test für zwei abhängige Stichproben	132
5.2.2	Die durchschnittliche Spearman-Rangkorrelation ρ (rho) und Kendalls Konkordanzkoeffizient W	133
5.2.2.1	Spearman's ρ	133
5.2.2.2	Kendalls Konkordanzkoeffizient (W).....	135
5.2.3	Vertiefung: Kendalls τ (tau)	137
5.2.3.1	Kendalls τ , wenn keine Rangbindungen vorliegen.....	137
5.2.3.2	Kendalls τ , wenn Rangbindungen vorliegen	139

5.2.3.3	Das Intraklassen- τ	141
5.2.4	Die polychorische Korrelation und Goodman und Kruskals Γ (gamma) ...	143
5.2.5	Die Berechnung der Maße für ordinale Daten mit Hilfe von SPSS	146
5.2.5.1	Wilcoxon-Test und Friedmans Rangvarianzanalyse.....	146
5.2.5.2	Spearman's ρ und Kendalls τ	149
5.2.5.3	Kendalls W	150
5.2.5.4	Goodman und Kruskals Γ	152
5.3.	Zusammenfassung zur Zuverlässigkeit ordinalskalierte Urteile	154
6	Maße der Reliabilität intervallskalierte Ratings	157
6.1	Die Intraklassenkorrelation (ICC)	163
6.1.1	Vertiefung: Die Reliabilität in der klassischen Testtheorie	164
6.1.2	Vertiefung: Das ein- und zweifaktorielle Modell der Reliabilität von Raterurteilen	168
6.1.2.1	Definition des Faktors ‚Rater‘ als ‚fixed‘ oder ‚random‘	172
6.1.2.2	Vertiefung: Zur Herleitung der Berechnungsformeln.....	173
6.1.3	Die Berechnung der verschiedenen Intraklassenkorrelationen.....	174
6.1.3.1	Die Berechnung der ICC _{unjust,einfakt} im einfaktoriellen Modell	174
6.1.3.1.1	Zum Problem negativer Intraklassenkorrelationen.....	177
6.1.3.1.2	Voraussetzungen im einfaktoriellen Modell und Überprüfung der Voraussetzungen	179
6.1.3.2	Die Berechnung der Intraklassenkorrelationen im zweifaktoriellen Modell	180
6.1.3.2.1	Die Berechnung der ICC _{unjust}	182
6.1.3.2.2	Die Berechnung der ICC _{just}	185
6.1.3.2.3	Die Homogenität der Ratermittelwerte.....	187
6.1.3.3	Zusammenfassung der Anwendungskriterien für die Intraklassenkorrelation.....	189
6.1.4	Die ICC des Mittelwerts mehrerer Rater	192
6.1.4.1	Die ICC _{unjust,einfakt,MW}	194
6.1.4.2	Die ICC _{unjust,MW}	195
6.1.4.3	Die ICC _{just,MW}	195
6.2	Zusammenhang der Intraklassenkorrelation mit Schätzeigenschaften von Ratingdaten	198
6.2.1	Der Standardmessfehler, das Vertrauensintervall und die kritische Differenz	198
6.2.2	Der Standardschätzfehler und die Minderungskorrektur	200
6.3	Das Problem des eingeschränkten Variationsbereichs der ICC und die Fisher-z-Transformation	204
6.4	Der Zusammenhang zwischen der ICC und der internen Konsistenz (Cronbachs α) psychometrischer Tests	208
6.5	Die korrigierte Trennschärfe ($r_{\Sigma,korr}$) der einzelnen Rater	210

6.6	Empfohlene Vorgehensweise bei der Anwendung des zweifaktoriellen Modells.....	212
6.7	Die Berechnung der Kennwerte mit Hilfe von SPSS.....	213
6.8	Anwendungsbeispiele.....	218
6.8.1	Beispiel 1: Die Einschätzung der psychischen Belastung von Klienten	218
6.8.2	Beispiel 2: Stephen Spurrier's Weinstudie aus dem Jahre 1976.....	224
6.9	Weitere Maße der Reliabilität intervallskaliert Ratings.....	229
6.10	Zusammenfassung zur Reliabilität intervallskaliert Ratings.....	232
7	Spezielle Probleme bei der Analyse der Interraterreliabilität	237
7.1	Einleitung.....	237
7.2	Von ihrer Natur her schwer zu erfassende Merkmale	237
7.3	Seltene Ereignisse.....	237
7.4	Berücksichtigung der Zeit	239
8	Hinweise zum Verbessern von Ratings	241
9	Anhang	249
9.1	Anhang A: Die Bestimmung der Signifikanz der κ -Werte (Ergänzung zu Kapitel 4).....	249
9.1.1	Signifikanztest für κ bei zwei Ratern (Ergänzung zu 4.1.3.1)	249
9.1.2	Signifikanztest für κ_v (Ergänzung zu 4.1.5.2)	251
9.1.3	Signifikanztest für κ_m (Ergänzung zu 4.1.5.3).....	251
9.2	Anhang B: Weitere Zuverlässigkeitsmaße für mindestens ordinalskalierte Ratings (Ergänzung zu Kapitel 5)	252
9.2.1	Finns r	252
9.2.2	Lawlis und Lus A	253
9.3	Anhang C: Tukeys Additivitätstest: Überprüfung der Voraussetzung $\sigma^2_{\alpha\beta} = 0$ im zweifaktoriellen varianzanalytischen Modell (Ergänzung zu Abschnitt 6.1.3).....	255
9.4	Anhang D: Berechnung der varianzanalytischen Kenngrößen zur Bestimmung der ICC's (Ergänzung zu Abschnitt 6.1.3).....	257
9.4.1	Die Berechnung der Kenngrößen aus der einfaktoriellen Varianzanalyse für die ICC _{unjust} in Beispiel 6.5	257
9.4.2	Die Berechnung der Kenngrößen aus der zweifaktoriellen Varianzanalyse für die ICC _{unjust,random} und die ICC _{just,fixed} in Beispiel 6.7	258
9.5	Anhang E: Vertrauensintervalle für die ICC's (Ergänzung zu Abschnitt 6.1.3).....	259
9.5.1	Das Vertrauensintervall für die ICC _{unjust} (Ergänzung zu Abschnitt 6.1.3.1).....	259
9.5.2	Das Vertrauensintervall für die ICC _{unjust,random} (Ergänzung zu Abschnitt 6.1.3.2.1).....	260

9.5.3 Das Vertrauensintervall für die $ICC_{\text{just, fixed}}$ (Ergänzung zu Abschnitt 6.1.3.2).....	261
9.6 Anhang F: Berechnung der Produkt-Momentkorrelation und der durchschnittlichen Interkorrelation für 2 Rater	262
9.6.1 Pearsons Produkt-Moment-Korrelation (r_{pm}).....	262
9.6.1.1 Die Testung der Signifikanz der Produkt-Moment-Korrelation.....	263
9.6.2 Die durchschnittliche Interkorrelation (r_{int}).....	264
9.7 Anhang G: Kritische Werte der χ^2 -Verteilung	266
9.8 Anhang H: Kritische Werte der t-Verteilung und für die Produkt-Moment-Korrelation.....	267
9.9 Anhang I: Kritische Werte der F-Verteilung für $\alpha = 0.25$ und $\alpha = 0.05$	268
9.10 Anhang J: Kritische Werte für den Wilcoxon-Test für $\alpha = 0.1$ und 0.05	271
Literaturverzeichnis	273
Index	283