

Einführung in die Statistik für SozialwissenschaftlerInnen

Übungen/Fragen zur fünften Veranstaltung

1. Worüber geben uns bei einer Kreuztabelle die erwarteten Häufigkeiten Auskunft? Was bedeutet es, wenn für den ganz überwiegenden Teil der Zellen beobachtete und erwartete Häufigkeiten sehr nah beieinander liegen?
2. Was wissen Sie auf der Basis des χ^2 -Koeffizienten über die bivariate Häufigkeitsverteilung, was wissen Sie nicht? Welche Probleme bestehen bei der Interpretation des Wertes für χ^2 ?
3. Berechnen Sie für die nachfolgende Kreuztabelle den Wert für χ^2 :

Zeitaufwand für die Statistik-Veranstaltung nach Geschlecht

Anzahl		Geschlecht		Gesamt
		männlich	weiblich	
Zeitaufwand	bis zu einer Stunde	20	33	53
dichotomisiert	mehr als eine Stunde	18	38	56
Gesamt		38	71	109

4. Zu den wichtigsten, auf χ^2 basierenden Kontingenzkoeffizienten zur Beschreibung des Zusammenhangs zweier nominal-skalierten Variablen gehören phi, Pearsons Kontingenzkoeffizient C und Cramers V. Bitte überlegen Sie bei jedem dieser Koeffizienten, wann er sinnvoll berechnet werden kann.
5. Für die bivariate Häufigkeitstabelle aus Aufgabe 3 wird die Höhe der Koeffizienten mit 0,06 berechnet (siehe nachfolgende Tab.). Was sagt dieser Wert über den Zusammenhang der beiden Variablen aus? Was gilt generell bei der Interpretation der Höhe von Kontingenzkoeffizienten?

		Wert
Zeitaufwand x	Phi	,059
Geschlecht	Cramer-V	,059
	Kontingenzkoeffizient C	,059
Anzahl der gültigen Fälle		109

6. Neben den auf χ^2 basierenden Kontingenzkoeffizienten können sog. PRE-Maße für die Beschreibung des Zusammenhangs zweier auf nominalem Messniveau erfasster Variablen herangezogen werden. Welches ist die zentrale Idee dieser Maße? Warum werden sie als asymmetrische Maße bezeichnet?

7. Was haben die beiden PRE-Maße lambda und Goodman and Kruskalls tau gemeinsam, worin unterscheiden sie sich?
8. Verdeutlichen Sie sich anhand der folgenden Tabelle, was unter konkordanten, was unter diskordanten Paaren zu verstehen ist. Welche Formen von Paarbindung können unterschieden werden und mit welcher Häufigkeit treten sie in der Tabelle auf?

Zeitaufwand für Statistik (dichotomisiert) * Anzahl der Semester

Anzahl		Anzahl der Semester		Gesamt
		1-3	4-9	
Zeitaufwand	bis zu einer Stunde	27	25	52
dichotomisiert	mehr als eine Stunde	31	24	55
Gesamt		58	49	107

9. Es werden drei verschiedene tau-Koeffizienten unterschieden: Kendall-tau_a, Kendall-tau_b und Kendall-tau_c. Was muss bei der Anwendung der ersten beiden Koeffizienten bedacht werden?
10. Welche zentrale Idee liegt der Berechnung des Rangkorrelationskoeffizienten nach Spearman zugrunde? Was geht als wichtigste Größe in seine Berechnung ein?
11. Warum können bei Koeffizienten für die Bestimmung de Zusammenhangs zweier ordinaler Variablen negative Werte auftreten, während dies bei Kontingenzkoeffizienten für den Zusammenhang zweier auf nominalem Messniveau erfasster Variablen nicht möglich ist

