

Ein nominal- und ein intervallskaliertes Merkmal: η bzw. η^2

- Vergleich von Gruppen (Männer/Frauen, Ost-/Westdeutsche) bezüglich eines intervall- oder ratioskalierten Merkmals (Einkommen, Testwerte etc.)
- Auch η^2 ist ein PRE-Maß
- Beste Prognose für intervallskalierte Variable bei einem beliebigen Befragten ist nicht die Modalkategorie, sondern das arithmetische Mittel
 - Summe der absoluten Abweichung = 0
 - Summe der quadrierten Abweichungen = min.
- Läßt sich der Vorhersagefehler (SAQ) reduzieren, wenn die Zugehörigkeit zu einer Gruppe bekannt ist?

Berechnung von η^2

$$\begin{aligned} h^2 &= \frac{Fehler_1 - Fehler_2}{Fehler_1} \\ &= \frac{SAQ_{gesamt} - SAQ_{Kategorien}}{SAQ_{gesamt}} \end{aligned}$$

- Maß für den Vorhersagefehler: SAQ
- SAQ_{gesamt} = Summe der Abweichungsquadrate vom gemeinsamen Durchschnitt
- $SAQ_{Kategorien}$ = Summe der Summe der Abweichungsquadrate in den Gruppen vom jeweiligen Gruppendurchschnitt

Beispiel: Geschlecht und Alter von Kursteilnehmern

	Kurs A					Kurs B			
	Frauen	Männer	Gesamt			Frauen	Männer	Gesamt	
Alter	19	20			Alter	19	20		
	19	20				20	19		
	19	20				19	20		
	19	20				20	19		
	19	20				19	20		
	19	20				20	19		
	19	20				19	20		
	19	20				20	19		
	19	20				19	20		
	19	20				20	19		
A.M.	19	20	19,5		A.M.	19,5	19,5	19,5	
SAQ	0	0	5		SAQ	2,5	2,5	5	
		5 - (0+0)				5 - (2,5+2,5)			
	eta ² =	-----	=1			eta ² =	-----	=0	
		5					5		

Beispiel: Geschlecht und Alter von Kursteilnehmern II

	Kurs C		
	Frauen	Männer	Gesamt
Alter	19	20	
	19	20	
	20	19	
	20	19	
	20	19	
	20	19	
	19	20	
	19	20	
	19	20	
	19	20	
A.M.	19,4	19,6	19,5
SAQ	2,4	2,4	5
	5 - (2,4 + 2,4)		
	eta ² = $\frac{\text{-----}}{5}$ = 0,04		
		5	

Eigenschaften von η^2

- η^2 gibt an, um wieviel Prozent sich der Vorhersagefehler (ausgedrückt durch SAQ_{gesamt}) reduzieren läßt, wenn die Gruppenzugehörigkeit bekannt ist
- η^2 kann Werte zwischen 0 % (keine Erklärungsleistung) und 100% (die Varianz in der intervallskalierten Variablen kann vollständig auf die Gruppenzugehörigkeit zurückgeführt werden) annehmen
- Das Maß η ist einfach die Wurzel aus η^2 . Sein Wertebereich liegt ebenfalls zwischen 0 und 1

Zwei intervallskalierte Variablen:

Pearson's r (Korrelationskoeffizient)

- Zusammenhang zwischen zwei intervallskalierten Merkmalen kann durch Streudiagramm visualisiert werden
- Konvention: unabhängige Variable auf der x-, abhängige Variable auf der y-Achse abtragen
- Aber: r ist ein symmetrisches Maß
- Positiver Zusammenhang: überdurchschnittliche x-Werte fallen mit überdurchschnittlichen y-Werten zusammen und umgekehrt
- Trägt man die beiden arithmetischen Mittel als Linien in das Diagramm ein, ist mit dem bloßen Auge zu erkennen, was über- und was unterdurchschnittlich ist

Berechnung von r

- Ist ein Fall hinsichtlich eines Merkmals überdurchschnittlich? ➔ einfache Abweichung vom Durchschnitt: $x_i - \bar{x}$
- Ist ein Fall hinsichtlich beider Merkmale überdurchschnittlich? ➔ Abweichungsprodukt (Fläche im Diagramm): $(x_i - \bar{x}) \times (y_i - \bar{y})$
- Teilt man die Summe aller Abweichungsprodukte durch die Zahl der Fälle, erhält man die Kovarianz (gemeinsame Streuung zweier Merkmale)
- Die Kovarianz hat einen Wertebereich von $-\infty$ bis $+\infty$

Berechnung von r II

- Die Kovarianz ist ähnlich wie χ^2 ein Absolutwert
- Deshalb hängt sie vom Maßstab beider Merkmale ab
- Um aus der Kovarianz ein Zusammenhangsmaß (r) zu berechnen, wird sie durch das Produkt *beider* Standardabweichungen geteilt
- r hat einen Wertebereich von -1 bis +1

$$r = \frac{\text{COV}_{xy}}{s_x \times s_y} = \frac{\frac{SAP}{n}}{\sqrt{\frac{SAQ_x}{n}} \times \sqrt{\frac{SAQ_y}{n}}} = \frac{SAP}{\sqrt{SAQ_x \times SAQ_y}}$$