

Übungsaufgaben VI: Korrelation und lineare Regression

1) 14 Haushalte wurden nach ihrem monatlichen Einkommen und ihren monatlichen Ausgaben gefragt. Man erhielt folgendes Ergebnis:

Haushalt	Einkommen	Ausgaben	
1	1000	1000	
2	2500	2000	
3	2000	1800	
4	2500	2200	
5	1500	1200	
6	1000	800	
7	2000	1600	
8	1500	1400	
9	1200	1060	
10	1800	1540	
11	1700	1460	
12	1300	1140	
13	2300	1940	
14	2200	1860	

- Berechnen Sie für das Einkommen und die Ausgaben die arithmetischen Mittelwerte und die Standardabweichungen.
- Bestimmen Sie die Kovarianz und den Korrelationskoeffizienten.
- Lässt sich für die Grundgesamtheit ein signifikanter Zusammenhang zwischen Einkommen und Ausgaben annehmen ($\alpha = 0.05$)? Formulieren Sie bitte zunächst eine Alternativ- und eine Nullhypothese.

2) Berechnen Sie die lineare Regression der Konsumausgaben auf das Einkommen (d.h. Sie führen die Ausgaben auf das Einkommen zurück). Geben Sie an:

- Steigungskoeffizienten und Regressionskonstante.
- Determinationskoeffizient (Anteil erklärter Varianz).
- Zeichnen Sie ein Streudiagramm und die Regressionsgerade.
- Ist der Zusammenhang zwischen den Einkommen und Ausgaben auf 5%-Niveau signifikant?
- Welche Konsumausgaben schätzen Sie mit Hilfe der linearen Regression für einen Haushalt mit einem monatlichen Einkommen von €3000,--?

3) Anhand von Daten des Allbus 2000 sollen Einflussfaktoren geklärt werden, die Personen dazu bewegen, der Aussage „Man sollte den in Deutschland lebenden Ausländern jede politische Betätigung untersagen“ zuzustimmen. (1: Stimme überhaupt nicht zu, 7: Stimme vollkommen zu.) Die mittlere Zustimmung zu der Frage liegt bei 3,58, die Varianz bei 4,33. Als unabhängige Variablen dienen:

1) Alter	(Mittelwert:47,58	Standardabweichung: 17,15)
2) Einkommen	(Mittelwert: 2538 DM	Standardabweichung; 1908,35)
3) Nationalstolz	(Mittelwert: 2,89	Standardabweichung: 0,8243)
4) Religiosität	(Mittelwert: 5,11	Standardabweichung: 2,95)

Da für alle Variablen ein metrisches Skalenniveau angenommen wird, konnte eine multivariate Regression durchgeführt werden. Das Ergebnis der Regression ist in folgender Tabelle abgedruckt:

	b	Standardfehler
(Konstante)	1,4875	
Alter	0,0175	0,0038
Einkommen	-0,00013	0,000034
Nationalstolz	0,5408	0,078
Religiosität	-0,01161	0,02211

$R^2 = 0,093$ $N = 986$

- Wie lautet die Regressionsgleichung?
- Für welche der unabhängigen Variablen kann in der Grundgesamtheit ein Einfluss auf die Zustimmung zu der Aussage angenommen werden? (ungerichtete Hypothesen, $\alpha = 0,05$).
- Welche der signifikanten Unabhängigen hat den stärksten Einfluss auf die abhängige Variable? Welche den schwächsten? Bitte begründen Sie die Antwort.