

Aufgabe 1: Der Datensatz „Theater“ (verfügbar auf der Homepage) beinhaltet Informationen über eine Studie, die das Ausgabeverhalten der Bürger der Stadt Basel untersucht hat. Er enthält u.a. folgende Variablen:

- y_1 Ausgaben für das Theater 2008 in Schweizer Franken (SFR)
- y_2 Ausgaben für Kultur im Allgemeinen in SFR
- x_1 Alter in Jahren
- x_2 Jahresgehalt in 1000 SFR

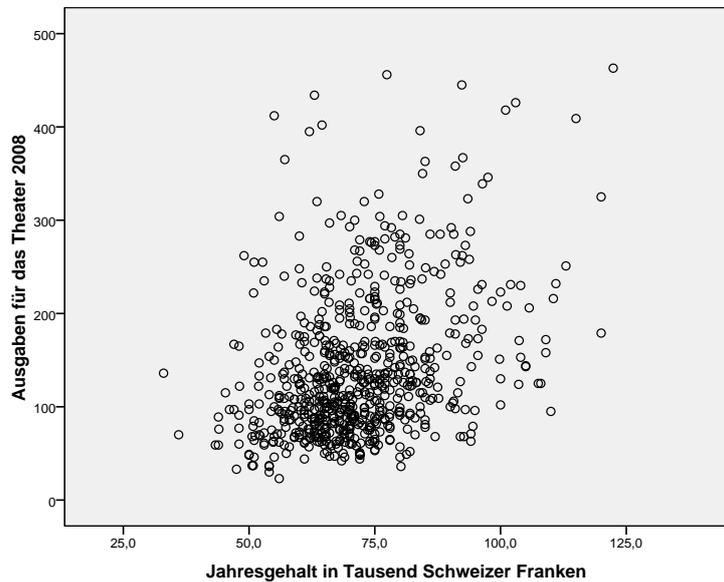
- a) Betrachten Sie die deskriptiven Statistiken zu ‘Ausgaben für das Theater 2008’. Wie sind die Werte zu interpretieren? Wie lautet die Spannweite? Wie groß ist der Variationskoeffizient?

Statistiken

Ausgaben für das Theater 2008

N	Gültig	699
	Fehlend	0
	Mittelwert	139,68
	Median	121,00
	Modus	135
	Standardabweichung	76,519
	Varianz	5855,083
	Minimum	23
	Maximum	463
	Perzentile	25
50		121,00
75		171,00

- b) Betrachten Sie das folgende Streudiagramm zwischen 'Ausgaben Theater' (y_1) und 'Jahresgehalt' (x_2). Ist eine Struktur zu erkennen? Ist diese stark ausgeprägt? Vermuten Sie, dass ein Zusammenhang zwischen den beiden Variablen besteht?



- c) Betrachten Sie den folgenden Regressionsoutput. Wie lauten die fehlenden Werte [1], [2] und [3]?

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	[3] ^a	[2]	,130	71,370

a. Einflußvariablen : (Konstante), Jahresgehalt in Tausend Schweizer Franken

ANOVA^b

Modell	Quadratsumme	df	Mittel der Quadrate	F	Sig. ^a
1 Regression	[1]	1	536549,675	105,336	,000 ^a
Nicht standardisierte Residuen	3550298,543	697	5093,685		
Gesamt	4086848,217	698			

a. Einflußvariablen : (Konstante), Jahresgehalt in Tausend Schweizer Franken

b. Abhängige Variable: Ausgaben für das Theater 2008

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressionskoeffizient B	Standardfehler	Beta	T	Sig.
1	(Konstante)	-11,408	14,967		-,762	,446
	Jahresgehalt in Tausend Schweizer Franken	2,108	,205	,362	10,263	,000

a. Abhängige Variable: Ausgaben für das Theater 2008

- d) Wie lautet das Regressionsmodell? Wie kann man es interpretieren?
- e) Berechnen Sie die in den folgenden Outputs mit '???' gekennzeichneten Korrelationen!

	Mittelwert	Standardabweichung	N
Ausgaben für Kultur	219,86	51,802	699
Alter	41,27	13,957	699
Jahresgehalt in Tausend Schweizer Franken	71,679	13,1535	699
Ausgaben für das Theater 2008	139,68	76,519	699

	Ausgaben für Kultur	Alter	Jahresgehalt in Tausend Schweizer Franken	Ausgaben für das Theater 2008
Ausgaben für Kultur	1	???	???	???
Korrelation nach Pearson				
Quadratsummen und Kreuzprodukte	1873048,406	177554,45	66474,005	1165936,634
Kovarianz	2683,450	254,376	95,235	1670,396
N	699	699	699	699

Wir gehen im Folgenden nun davon aus, dass die Ausgaben für Kultur im Allgemeinen (y_2) nur klassiert vorliegen ('keine Ausgaben', 'sehr wenig Ausgaben', ..., 'hohe Ausgaben' → 7 Kategorien) und untersuchen einen möglichen Zusammenhang zwischen den Ausgaben und der Alterskategorie ('0-20 Jahre', '21-30 Jahre', ..., 'über 60 Jahre' → 6 Kategorien).

Symmetrische Maße

	Wert	Asymptotischer Standardfehler ^a	Näherungsweise T ^b
Nominal- bzgl. Nominalmaß	Phi	,452	
	Cramer-V	,202	
	Kontingenzkoeffizient	,412	
Ordinal- bzgl. Ordinalmaß	Kendall-Tau-b	,300	,028
	Kendall-Tau-c	,294	,027
	Gamma	,364	,033
	Korrelation nach Spearman	,375	,034
Intervall- bzgl. Intervallmaß	Pearson-R	,369	,034
Anzahl der gültigen Fälle		699	

a. Die Null-Hyphothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

- f) Interpretieren Sie die Ergebnisse des SPSS-Outputs!
- g) Der Wert von χ^2 beträgt in Aufgabenteil f) 142,83. Bestimmen Sie die Werte von Cramers V, dem Kontingenzkoeffizienten und Phi von Hand! Erhalten Sie die gleichen Ergebnisse wie SPSS?

Aufgabe 2: In einer Studie soll der Zusammenhang zwischen der durchschnittlichen Monatstemperatur und der Hotelauslastung an drei Orten untersucht werden. Als typischer Wintersportort wurde Davos gewählt, für den Sommerurlaub Polenca auf Mallorca und als Stadt- und Geschäftsreiseziel Basel. Es wurden in den Monaten des Jahres 2002 die Durchschnittstemperaturen (X) tagsüber sowie die Hotelauslastungen (Y) erhoben.

Monat	Davos		Polenca		Basel	
	X	Y	X	Y	X	Y
Jan	-6	91	10	13	1	23
Feb	-5	89	10	21	0	82
Mar	2	76	14	42	5	40
Apr	4	52	17	64	9	45
May	7	42	22	79	14	39
Jun	15	36	24	81	20	43
Jul	17	37	26	86	23	50
Aug	19	39	27	92	24	95
Sep	13	26	22	36	21	64
Oct	9	27	19	23	14	78
Nov	4	68	14	13	9	9
Dec	0	92	12	41	4	12

- a) Berechnen Sie den Korrelationskoeffizienten $r(X,Y)$. Gibt es einen linearen Zusammenhang? (Hinweis: $\sum_{i=1}^{36} x_i y_i = 22776$, $\bar{x} = 12,22$, $\bar{y} = 51,28$, $S_{xx} = 2770,22$ und $S_{yy} = 25451,22$)
- b) Aufgesplittet nach den drei Orten ergeben sich (über SPSS) folgende drei Korrelationen.

$$r_D = -0,87 \quad r_P = 0,818 \quad r_B = 0,415$$

Wie interpretieren Sie dieses Ergebnis?

Wir wollen nun die kategoriale Variable 'Ort' in Dummykodierung in einem Regressionsmodell als Einflussgröße verwenden. Die Variable Z (also der 'Ort') hat $k = 3$ Ausprägungen, also benötigen wir $k - 1 = 2$ Dummies Z_1 und Z_2 . Wir wählen 'Basel' als Referenzkategorie, Z_1 steht für Davos, Z_2 für Polenca.

- c) Erläutern an Sie an diesem Beispiel kurz den Begriff der Dummycodierung!

- d) Wir wählen als abhängige Variable die 'Temperatur'. Mit SPSS erhalten wir folgendes Regressionsmodell:

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressionskoeffizient B	Standardfehler	Beta	T	Sig.
1	(Konstante)	12,000	2,234		5,372	,000
	Davos	-5,417	3,159	-,291	-1,715	,096
	Polenca	6,083	3,159	,327	1,926	,063

a. Abhängige Variable: Temperatur

Interpretieren Sie den Output: Wie lautet das Modell? Wie lauten die Durchschnittstemperaturen von Basel, Davos und Polenca?

- e) Wir wählen als abhängige Variable 'Hotelauslastung' und erhalten folgendes Regressionsmodell mit SPSS:

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		Regressionskoeffizient B	Standardfehler	Beta	T	Sig.
1	(Konstante)	48,333	7,946		6,083	,000
	Davos	7,917	11,237	,140	,705	,486
	Polenca	,917	11,237	,016	,082	,935

a. Abhängige Variable: Hotelauslastung

Interpretieren Sie den Output: Wie lautet das Modell? Wie lauten die durchschnittlichen Hotelauslastungen von Basel, Davos und Polenca?