

FRAGE: Besteht ein Zusammenhang zwischen zwei unterschiedlichen Gruppen von Variablen
 → Antwort liefert kanonische Korrelationsanalyse

AufgabeCARS.log

```
log: U:\einfuehrung\AufgabeCARS.log
log type: text
opened on: 7 Jun 2010, 14:00:33
```

```
.
.
. * Daten einlesen
. use "CARS.dta", clear
.
. * Datendeskription
. describe
```

Contains data from CARS.dta

```
obs: 23 Anzahl an Beobachtungen
vars: 9
size: 1,219 (99.9% of memory free) 3 Jul 2007 11:56
```

variable name	storage type	display format	value label
CAR	str17	%17s	
ECONOMY	float	%9.0g	
SERVICE	float	%9.0g	
VALUE	float	%9.0g	Non-depreciation of value
PRICE	float	%9.0g	1 = cheap car, 6 = expensive car
DESIGN	float	%9.0g	
SPORT	float	%9.0g	Sporty car
SAFETY	float	%9.0g	
EASYINESS	float	%9.0g	Easy handling

sorted by:

. sum

Variable	Obs	Mean	Std. Dev.	Min	Max
CAR	0				
ECONOMY	23	3.291304	.8638777	2.1	5.3
SERVICE	23	3.06087	.809455	1.6	4.7
VALUE	23	3.152174	1.088731	1.6	5.5
PRICE	23	3.273913	1.188283	1.5	5.9
DESIGN	23	3.2	.8495989	1.3	4.8
SPORT	23	3.465217	1.026295	1.1	5.8
SAFETY	23	3.286957	1.122215	1.4	5.9
EASYINESS	23	2.786957	.5643213	1.8	4.3

```
. * Kanonische Korrelationsanalyse
. canon (PRICE VALUE) (ECONOMY SERVICE DESIGN SPORT SAFETY EASYINESS)
```

Linear combinations for canonical correlations Number of obs = 23
 Seite 1

erste Gruppe von Variablen zweite Gruppe von Variablen

$$X = \begin{bmatrix} X_1: \text{PRICE} \\ X_2: \text{VALUE} \end{bmatrix}, Y = \begin{bmatrix} Y_1: \text{ECONOMY} \\ Y_2: \text{SERVICE} \\ Y_3: \text{DESIGN} \\ Y_4: \text{SPORT} \\ Y_5: \text{SAFETY} \\ Y_6: \text{EASY} \end{bmatrix}$$

Indexvariablen: $I_x = a'X$
 $I_y = b'Y$

		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
u1	PRICE	-0.3332732	.0758159	-4.40	0.000	-.4905057	-.1760406
	VALUE	.5872981	.0827483	7.10	0.000	.4156885	.7589076
v1	ECONOMY	-.4325146	.1100989	-3.93	0.001	-.6608457	-.2041836
	SERVICE	.1911816	.2094525	0.91	0.371	-.2431963	.6255594
	DESIGN	.0047506	.1714016	0.03	0.978	-.3507147	.3602158
	SPORT	.4584596	.1260173	3.64	0.001	.1971157	.7198035
	SAFETY	.2226613	.1431398	1.56	0.134	-.0741925	.519515
	EASYINESS	.3761583	.203755	1.85	0.078	-.0464037	.7987203
u2	PRICE	1.601911	.1923573	8.33	0.000	1.202987	2.000836
	VALUE	1.68649	.209946	8.03	0.000	1.251089	2.121892
v2	ECONOMY	.5679744	.2793388	2.03	0.054	-.0113388	1.147288
	SERVICE	.544205	.5314151	1.02	0.317	-.5578825	1.646293
	DESIGN	-.0117983	.4348739	-0.03	0.979	-.9136717	.890075
	SPORT	-.0956752	.3197265	-0.30	0.768	-.7587474	.567397
	SAFETY	-.0141904	.363169	-0.04	0.969	-.7673568	.738976
	EASYINESS	.9147486	.5169597	1.77	0.091	-.1573603	1.986857

$$\Rightarrow I_x^{(u)} = a_1'X = -0.33 \cdot \text{PRICE} + 0.58 \cdot \text{VALUE}$$

$$I_y^{(u)} = b_1'Y = -0.43 \cdot \text{ECONOMY} + \dots + 0.37 \cdot \text{EASYINESS}$$

Canonical correlations: (Standard errors estimated conditionally)

0.9792 0.8851

zweite kanonische Korrelation = $\hat{\rho}_2 = \rho(a_2'X, b_2'Y)$
 erste kanonische Korrelation = $\hat{\rho}_1 = \rho(a_1'X, b_1'Y)$

Tests of significance of all canonical correlations

	Statistic	df1	df2	F	Prob>F
Wilks' lambda	.00891633	12	30	23.9757	0.0000 e
Pillai's trace	1.74227	12	32	18.0267	0.0000 a
Lawley-Hotelling trace	26.9055	12	28	31.3898	0.0000 a
Roy's largest root	23.2878	6	16	62.1009	0.0000 u

e = exact, a = approximate, u = upper bound on F

* schliesse log-file
 log close
 log: U:\einfuehrung\AufgabeCARS.log
 log type: text
 closed on: 7 Jun 2010, 14:00:33

Wichtigkeit, dass man diese Werte zufällig
 nicht gegebene Nullhypothese H_0 ist wahr
 (Es besteht kein Zusammenhang zw. X und Y)
 $\Rightarrow$ signifikanter Zusammenhang zw. X & Y , da
 Wichtigkeit ≈ 0

Beide Gruppen von Variablen enthalten in gewisser Weise ähnliche Information, da aus beiden jeweils eine Linearkombination konstruiert werden kann und die beiden sich ergebenden Linearkombinationen (0,9792) miteinander korreliert sind.

Die Sportlichkeit geht positiv in die Linearkombination ein, die maximal mit der Linearkombination aus Preis und Wertbeständigkeit korreliert ist. In letztere geht der Preis negativ und die Wertbeständigkeit positiv ein. Insofern sollte es einen negativen Zusammenhang zwischen Sportlichkeit und Preis und einen positiven zwischen Sportlichkeit und Wertbeständigkeit geben.