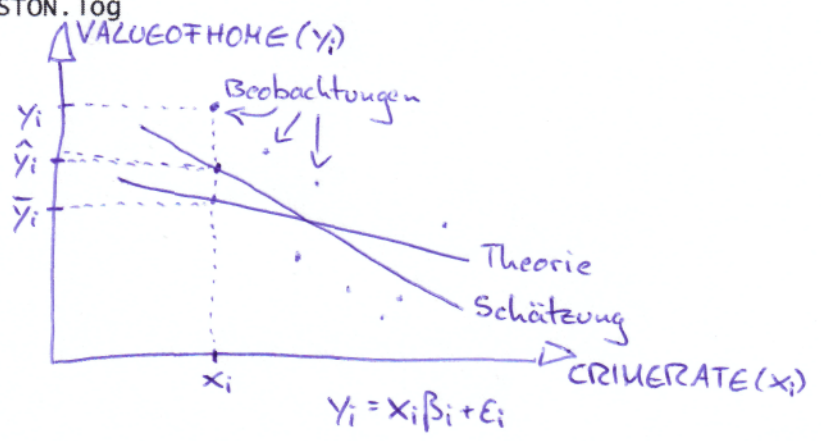


Mit der linearen Regression kann man die Stärke des Zusammenhangs zwischen einer gegebenen Variable Y und den restlichen Variablen X messen.

AufgabeBOSTON.log

```
log: U:\einfuehrung\AufgabeBOSTON.log
log type: text
opened on: 7 Jun 2010, 14:21:04
```

```
. * Oeffne Verzeichnis
. * cd U:\einfuehrung
. * Daten einlesen
. use "BOSTON.dta", clear
. describe
```



```
Contains data from BOSTON.dta
obs: 506
vars: 14
size: 30,360 (97.1% of memory free)
11 Jun 2007 14:34
```

506 Beobachtungen

variable name	storage type	display format	value label	variable label
VALUEOFHOME	float	%9.0g		median value of owner occupied
homes in this district in			\$1000	
CRIMRATE	float	%9.0g		per capita crime rate
RESIDENTIALZONE	float	%9.0g		proportion of residential land
zoned for large lots				
NONRETAIL	float	%9.0g		proportion of nonretail business
acres				
RIVER	float	%9.0g		1 if tract bounds river, 0
otherwise				
NOCONCENTRATION	float	%9.0g		nitric oxides concentration
NOOFROOMS	float	%9.0g		average number of rooms per
dwelling				
AGE	float	%9.0g		proportion of owner-occupied units
built prior to 1940				
DISTANCE	float	%9.0g		weighted distances to five Boston
employment centers				
ACCESSTOHIGHWAY	float	%9.0g		index of accessibility to radial
highways				
TAXRATE	float	%9.0g		full-value property tax rate per
\$10.000				
PUPILTEACHER	float	%9.0g		pupil/teacher ratio
BLACKPEOPLE	float	%9.0g		$I(B < 0.63) * 1000(B - 0.63)^2$ where $B =$
proportion of blacks				
LOWERSTATUS	float	%9.0g		% lower status of the population

sorted by:

```
. sum
```

variable	Obs	Mean	Std. Dev.	Min	Max
VALUEOFHOME	506	22.53281	9.197104	5	50

φ-Wert einer Wohnung = 22532\$

Die Einheiten der Regressoren sind wichtig für die spätere Interpretation!

$$VALUEOFHOME = -\text{cons} + \beta_1 \text{CRIMERATE} + \dots + \beta_{13} \text{LOWERSTATUS} + \epsilon$$

$\beta_1, \beta_2, \dots, \beta_{13}$ Koeffizienten

Aufgabe BOSTON.log					
CRIMERATE	506	3.613524	8.601545	.00632	88.9762
RESIDENTIALZONE	506	11.36364	23.32245	0	100
NONRETAIL	506	11.13678	6.860353	.46	27.74
RIVER	506	.06917	.253994	0	1
NOCONCENTRATION	506	.5546951	.1158777	.385	.871
NOOFFROOMS	506	6.284634	.7026172	3.561	8.78
AGE	506	68.5749	28.14886	2.9	100
DISTANCE	506	3.795043	2.10571	1.1296	12.1265
ACCESSSTOHIWAY	506	9.549407	8.707259	1	24
TAXRATE	506	408.2372	168.5371	187	711
PUPILTEACHER	506	18.45553	2.164946	12.6	22
BLACKPEOPLE	506	356.674	91.29486	.32	396.9
LOWERSTATUS	506	12.65306	7.141062	1.73	37.97

* a) OLS Regression

* Führe Regression durch und speichere Abweichungssumme ab

```
. regress VALUEOFHOME CRIMERATE RESIDENTIALZONE NONRETAIL RIVER NOCONCENTRATION
NOOFFROOMS AGE DISTANCE ACCESSSTOHIWAY TAXRATE PUPILTEACHER BLACKPEOPLE
LOWERSTATUS
```

Source	SS	df	MS	Number of obs =	506
Model	31637.511	13	2433.65469	F(13, 492) =	108.08
Residual	11078.7846	492	22.5178548	Prob > F =	0.0000
Total	42716.2956	505	84.586724	R-squared =	0.7406
				Adj R-squared =	0.7338
				Root MSE =	4.7453

VALUEOFHOME	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
CRIMERATE	-.1080114	.032865	-3.29	0.001	-.1725844 -.0434383
RESIDENTIALZONE	.0464205	.0137275	3.38	0.001	.0194488 .0733921
NONRETAIL	.0205587	.0614957	0.33	0.738	-.1002679 .1413852
RIVER	2.686734	.8615798	3.12	0.002	.9939043 4.379564
NOCONCENTRATION	-17.76661	3.819744	-4.65	0.000	-25.27164 -10.26159
NOOFFROOMS	3.809865	.4179252	9.12	0.000	2.988727 4.631003
AGE	.0006922	.0132098	0.05	0.958	-.0252623 .0266468
DISTANCE	-1.475567	.1994547	-7.40	0.000	-1.867455 -1.083679
ACCESSSTOHIWAY	.3060495	.0663464	4.61	0.000	.1756922 .4364068
TAXRATE	-.0123346	.0037605	-3.28	0.001	-.0197233 -.0049459
PUPILTEACHER	-.9527473	.1308267	-7.28	0.000	-1.209795 -.6956992
BLACKPEOPLE	.0093117	.002686	3.47	0.001	.0040343 .0145891
LOWERSTATUS	-.5247584	.0507153	-10.35	0.000	-.6244036 -.4251132
_cons	36.45949	5.103459	7.14	0.000	26.43223 46.48675

$0 \in [-0,1, 0,14]$

$0 \in [-0,75, 0,02]$

scalar rss_u = e(rss)

```
. display rss_u
11078.785
```

geschätztes $\hat{\beta}_j$

$SE(\hat{\beta}_j)$

$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$

* b) Test, ob alle Betas ausser Konstante gleich Null

* restringierte Regression (Regression nur auf Konstante)

```
. regress VALUEOFHOME
```

geschätzter Zusammenhang zw. VALUEOFHOME und den x:

Seite 2

$$VALUEOFHOME = 36,45 - 0,1080 \cdot \text{CRIMERATE} + \dots - 0,5247 \cdot \text{LOWERSTATUS}$$

Alle Regressoren außer NONRETAIL und AGE sind signifikant von Null verschieden
 $\Rightarrow$ kein Zusammenhang zw. VALUEOFHOME und NONRETAIL oder AGE festgestellt!

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon$$

$$\Rightarrow \frac{\partial y}{\partial x_j} = \beta_j$$

d.h. Änderung des j-ten Regressors x_j um eine Einheit geht mit einer Änderung von y um β_j Einheiten einher

CRIMRATE + 10 Prozentpunkte $\rightarrow$ VALUEOFHOME - 1080\$

RIVER (0 $\rightarrow$ 1) $\rightarrow$ VALUEOFHOME + 2686\$

DISTANCE + 1 $\rightarrow$ " - 1476\$

LOWERSTATUS + 10 Prozentpunkte $\rightarrow$ " - 5247\$

Hypothese H_0 : Zwischen Y und X besteht kein Zusammenhang

~~Ergebnis des Hypothesentests gesetzt H_0 ist wahr~~

$\Rightarrow$ Angenommen H_0 ist wahr, wie wahrscheinlich ist es die Daten (die wir bereits beobachtet haben) zufällig zu beobachten?

Antwort: P-Wert

$\Rightarrow$ Kein Zusammenhang zw. VALUEOFHOME & Nonretail oder Age.
Aber Zusammenhang zw. " & anderen Variablen

(Lehnt Hypothese ab, wenn p-Wert kleiner als das Signifikanzniveau (z.B. 5%))

CRIMRATE + 10 Prozentpunkte $\Rightarrow$ VALUEOFHOME - 1080\$

Das bedeutet, dass die Wohnungspreise in Distrikten mit einer 10 Prozentpunkte geringeren Kriminalitätsrate um ca. 1080\$ höher sind.

Es ist nicht klar, dass die Verringerung der CRIMRATE in ein und demselben Distrikt die Wohnungspreise um 1080\$ erhöhen würde.

(Es könnte sein, dass in den beobachteten Distrikten eine hohe

CRIMRATE mit anderen Wohnungswertmindernden Faktoren korreliert ist, wie z.B. schlechten Straßen. Eine Änderung der CRIMRATE bewirkt nicht unmittelbar eine Änderung des Straßenzustands.

$\Rightarrow$ VALUEOFHOME würde vermutlich um weniger als 1080\$ steigen.