

# Sommersemester 2009, Statistik mit SPSS



# Überblick

1. Korrelation vs. Regression
2. Ziel der Regressionsanalyse
3. Syntax für den Regressionsbefehl
4. Beispiele

# Korrelation vs. Regression

## a) Korrelation:

„Größe“



„Gewicht“

## b) Regression (Einfachregression)

„Größe“ (UV)



„Gewicht“ (AV)

# Ziele der Regressionsanalyse

Die Regressionsanalyse ermöglicht Aussage zu folgenden Fragestellungen:

- Wie **stark ist der Einfluss** einer einzelnen unabhängigen Variablen auf die abhängige Variable?
- Wie **ändert sich die abhängige Variable** bei einer Veränderung der unabhängigen Variablen (Prognose)?
- Bei multipler Regression: Wie groß ist **die Erklärungskraft aller unabhängiger Variablen zusammen**?

# Voraussetzungen der linearen Regression

- **Metrische Variablen** (sowohl UV als auch AV), als UV können auch 0-1 codierte Variablen verwendet werden.
- Es liegt ein **linearer Zusammenhang** zwischen AV und UV vor
  - lineare Regression (Lässt sich mit Streudiagramm darstellen).

# Syntax für den Regressionsbefehl

```
REGRESSION [MATRIX=[IN(file)] [OUT(file)]]
  [/VARIABLES={varlist      }
             {(COLLECT)**} Command Syntax
             {ALL          }
  [/DESCRIPTIVES=[DEFAULTS] [MEAN] [STDDEV] [CORR] [COV]
                 [VARIANCE] [XPROD] [SIG] [N] [BADCORR]
                 [ALL] [NONE**]]
  [/SELECT={varname relation value}
  [/MISSING=[{LISTWISE**      } [INCLUDE]]
            {PAIRWISE        }
            {MEANSUBSTITUTION}
  [/REGWGT=varname]
  [/STATISTICS=[DEFAULTS**] [R**] [COEFF**]
               [ANOVA**] [OUTS**]
               [ZPP] [CHA] [CI] [F] [BCOV]
               [SES] [XTX] [COLLIN]
               [TOL] [SELECTION] [ALL]]
...
/DEPENDENT=varlist
[/METHOD=] {STEPWISE [varlist]      }
           {FORWARD  [varlist]      }
           {BACKWARD [varlist]      }
           {ENTER   [varlist]      }
           {REMOVE  varlist         }
           {TEST(varlist)(varlist)...}
[/RESIDUALS=[DEFAULTS] [ID(varname
```

# Beispiel 1: Einfachregression mit SPSS

→ Regression des Gewichtes (AV) auf die Körpergröße (UV).

```
***Lineare Regression, Beispiel 1, Allbus 2004.  
  
regr  
/dep v307  
/enter v305.
```

## Ausgabe Modellzusammenfassung: R

Modellzusammenfassung				
Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,513 <sup>a</sup>	,263	,263	13,288

a. Einflußvariable n : (Konstante), v305  
KOERPERGROESSE IN CM, BEFRAGTE<R>

R entspricht im bivariaten Modell dem Korrelationskoeffizienten (Pearson's r)

## Ausgabe Modellzusammenfassung: $R^2$

Modellzusammenfassung				
Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,513 <sup>a</sup>	,263	,263	13,288

a. Einflußvariable n : (Konstante), v305  
KOEPRERGROESSE IN CM, BEFRAGTE<R>

**$R^2$ :** Determinationskoeffizient (Bestimmtheitsmaß)

- er gibt Auskunft über die „Güte“ des Modells (Maß für die Erklärungskraft des Modells insgesamt), d.h.
- $R^2$  gibt an, wie viel Varianz der abhängigen Variablen durch die unabhängige Variable erklärt wird,
- hier wird 26,3 % der Varianz des Gewichtes (AV) durch die Varianz der Körpergröße (UV) „erklärt“.

## Ausgabe: $R^2$ als PRE-Maß

Modellzusammenfassung				
Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,513 <sup>a</sup>	,263	,263	13,288

a. Einflußvariable n : (Konstante), v305  
KOERPERGROESSE IN CM, BEFRAGTE<R>

### $R^2$ als PRE-Maß:

Bei Kenntnis der Var. „Körpergröße“ (UV) werden die Vorhersagefehler im Vergleich zur Vorhersage des Gewichtes (AV) ohne Kenntnis dieser Var. um 26,3% reduziert.

Standardfehler: bestimmt die Genauigkeit der Schätzung.

# Ausgabe: Anova-Tabelle, Varianzzerlegung

ANOVA <sup>b</sup>						
Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	181534,306	1	181534,306	1028,092	,000 <sup>a</sup>
	Residuen	509239,227	2884	176,574		
	Gesamt	690773,533	2885			

a. Einflußvariablen : (Konstante), v305 KOERPERGROESSE IN CM, BEFRAGTE<R>

b. Abhängige Variable: v307 GEWICHT IN KG, BEFRAGTE<R>

**Gesamt:** Gesamtstreuung

**Regression:** Anteil der Gesamtstreuung, der durch das Regressionsmodell erklärt wird

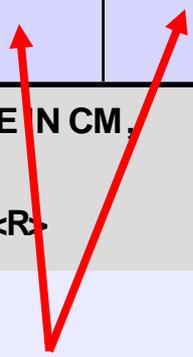
**Residuen:** Anteil der Gesamtstreuung, der durch das Regressionsmodell nicht erklärt wird

# Ausgabe: Anova-Tabelle, Varianzzerlegung

ANOVA <sup>b</sup>						
Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	181534,306	1	181534,306	1028,092	,000 <sup>a</sup>
	Residuen	509239,227	2884	176,574		
	Gesamt	690773,533	2885			

a. Einflußvariablen : (Konstante), v305 KOERPERGROESSE IN CM, BEFRAGTE<R>

b. Abhängige Variable: v307 GEWICHT IN KG, BEFRAGTE<R>



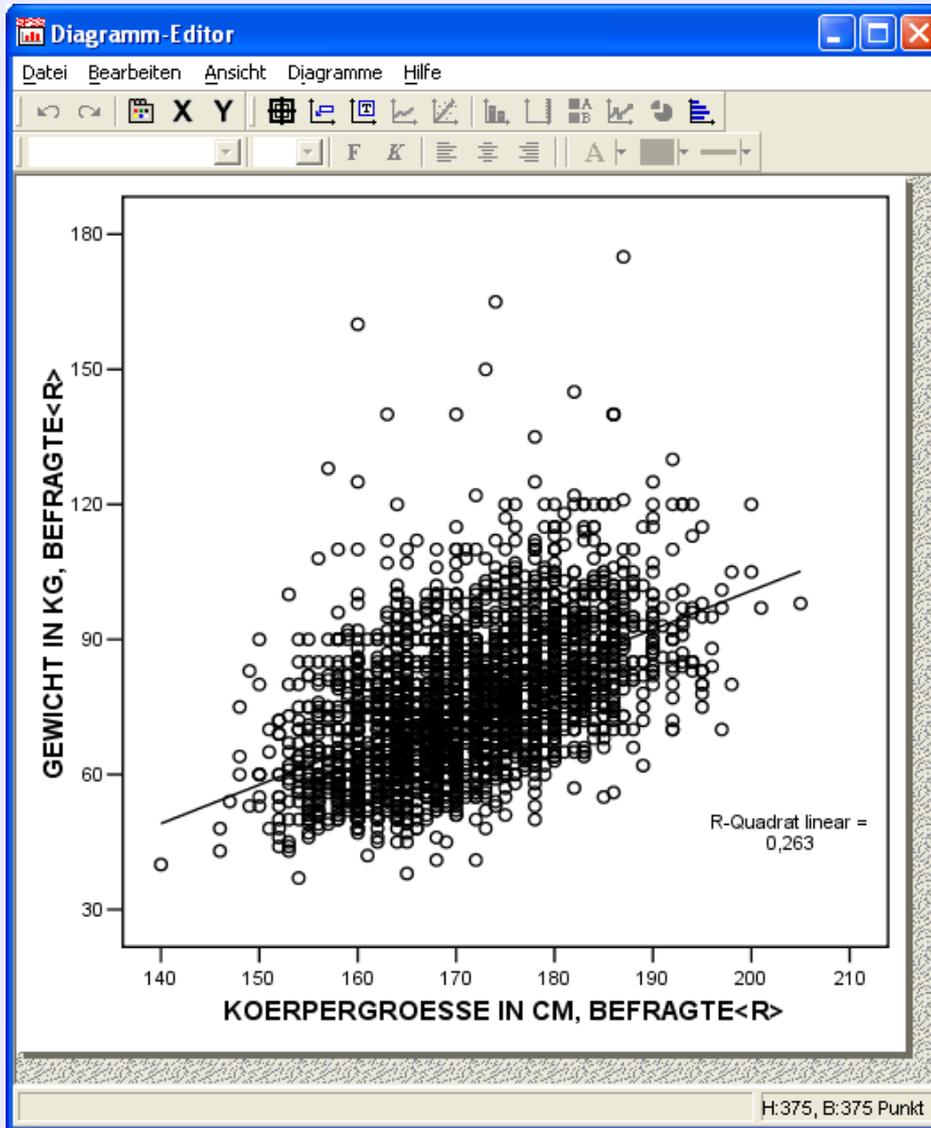
**F-Test:** Das Regressionsmodell ist hochsignifikant, d.h. der Determinationskoeffizient ist in der Grundgesamtheit mit einer Irrtumswahrscheinlichkeit von weniger als 0,1% ( $p < 0,001$ ) von null verschieden.

## Ausgabe: Koeffizientenblock

Koeffizienten <sup>a</sup>						
Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	-71,617	4,613		-15,526	,000
	v305 KOERPERGROESSE IN CM, BEFRAGTE<R>	,862	,027	,513	32,064	,000

a. Abhängige Variable: v307 GEWICHT IN KG, BEFRAGTE<R>

**Regressionskonstante:** Schnittpunkt der Regressionsgerade mit der y-Achse, wenn  $x=0$  (s. Streudiagramm).



Konstante: Schnittpunkt der Regressionsgerade mit der y-Achse, wenn  $x=0$  (s. Streudiagramm).

## Ausgabe: Koeffizientenblock

Koeffizienten <sup>a</sup>					
Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	Signifikanz
		B	Standardfehler	Beta	
1	(Konstante)	-71,617	4,613		,000
	v305 KOERPERGROESSE IN CM, BEFRAGTE<R>	,862	,027	,513	,000

a. Abhängige Variable: v307 GEWICHT IN KG, BEFRAGTE<R>

**Unstandardisierter Regressionskoeffizient** (Regressionsgewicht/Steigung):  
 Wenn X (UV) um eine Einheit ansteigt, dann steigt Y (AV) um 0,86 Einheiten an.

D.h. hier:

Wenn die Körpergröße um einen Zentimeter ansteigt (cm als Einheit), dann steigt das Körpergewicht im Durchschnitt um 0,86 kg an.

## Ausgabe: Koeffizientenblock

Koeffizienten <sup>a</sup>						
Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	-71,617	4,613		-15,526	,000
	v305 KOERPERGROESSE IN CM, BEFRAGTE<R>	,862	,027	,513	32,064	,000

a. Abhängige Variable: v307 GEWICHT IN KG, BEFRAGTE<R>

**T-Test:** Dient der Signifikanzprüfung des Regressionskoeffizienten.

**Signifikanz:** empirische Signifikanz, d.h. mit einer Irrtumswahrscheinlichkeit von <math><0,1\%</math> ( $p < 0,001$ ) ist der Regressionskoeffizient von Null verschieden (oder: der Koeffizient ist höchst signifikant).

# Zusammenfassung Einfachregression

- $R^2$  :** Anteil erklärter Varianz, PRE-Maß
- F-Test und Sign.:** testet Nullhypothese, dass der Determinationskoeffizient in der GG null ist; Signifikanz des Modells.
- Konstante:** Schnittpunkt der Regressionsgerade mit der y-Achse, wenn  $x=0$
- B:** Stärke des Effektes von X (UV) auf Y (AV),
- T-Test und Sign.:** testet Nullhypothese, dass der Regressionskoeffizient in der GG null ist; Signifikanz der Regressionskoeffizienten

# Beispiel 2 und Übungsaufgaben

# Klausurvorbereitung

In der Übung behandelte SPSS-Befehle (jeweils inkl. Unterbefehle)

FREQUENCIES

GRAPH

DISPLAY DICTIONARY

MISSING VALUES

RECODE

VARIABLE LABELS

VALUE LABELS

COMMENT

EXECUTE

COMPUTE

EXAMINE

IF

COUNT

SELECT IF

SORT CASES

SPLIT FILE

TEMPORARY

WEIGHT

T-TEST

CROSSTABS

CORRELATIONS

REGRESSION