

Sommersemester 2009

Statistik mit SPSS



Überblick

1. Korrelationsanalysen

- Kovariation und Kovarianz
- **Korrelation:**
 - Interpretation
 - Stärke des Zusammenhangs

2. Streudiagramme

1. Korrelationsanalyse

Kovariation und Kovarianz

- **Symmetrische Zusammenhangsmaßen**, die das Miteinander-Variieren zweier metrischen Variablen abbilden.
- **Kovariation** ist die Summe der Produkte der mittelwertbereinigten Realisationen von X und Y.
 - Sie gibt die Beziehung eines Zusammenhangs an (positive/ negative Je-desto-Beziehung),
 - ist aber kein geeignetes Maß für die Zusammenhangsstärke, da ihre Größe durch die Fallzahl beeinflusst wird

- **Kovarianz**

$$s_{XY} = \frac{SP_{XY}}{n} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$$

- Kovarianz ist nicht geeignet für Zusammenhangsstärke, da die Werte von den Maßeinheiten der beteiligten Variablen abhängen.

1. Korrelationsanalyse

Korrelation

- Wünschenswert ist ein Zusammenhangsmaß, das nicht von den Maßeinheiten der beteiligten Variablen abhängt, sondern das nur Werte zwischen +1 und -1 annehmen kann.
- Dieses Zusammenhangsmaß ergibt sich, wenn die Kovarianz durch das Produkt der Standardabweichungen dividiert wird.
- Diese Größe wird als Pearsons Korrelation bezeichnet.

$$\begin{aligned}
 r_{XY} &= \frac{SP_{XY}}{\sqrt{SS_X \cdot SS_Y}} = \frac{s_{XY}}{\sqrt{s_X^2 \cdot s_Y^2}} = \frac{s_{XY}}{s_X \cdot s_Y} = \frac{\frac{1}{n} \sum_{i=1}^n x_i \cdot y_i - \bar{x} \cdot \bar{y}}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right) \cdot \left(\frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2 \right)}} \\
 &= \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\sum_{i=1}^n x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right) \cdot \left(\sum_{i=1}^n y_i^2 - n \cdot \bar{y}^2 \right)}}
 \end{aligned}$$

1. Korrelationsanalyse

- Korrelationskoeffizienten werden berechnet, um die Stärke des Zusammenhangs zweier metrischen Variablen zu bestimmen.
- Dieser Korrelationskoeffizient kann nur lineare Zusammenhänge erkennen.
- Korrelationskoeffizient ist ein symmetrisches Zusammenhangsmaß
- Korrelation misst die Schlankheit der Punktwolke, sie misst die Annäherung der Punktwolke an eine Gerade.

1. Korrelationsanalyse

Korrelation: Interpretation

- Wert null > kein **linearer Zusammenhang** zwischen den Variablen oder statistische Unabhängigkeit.
- Maximalwerte von +1 bzw. -1 werden nur dann erreicht, wenn alle Datenpunkte im Streudiagramm auf einer Geraden mit positiver bzw. negativer Steigung liegen.
- Positive Korrelationen deuten auf einen positiven linearen Zusammenhang. Negative Korrelationen kennzeichnen einen negativen linearen Zusammenhang.

1. Korrelationsanalyse

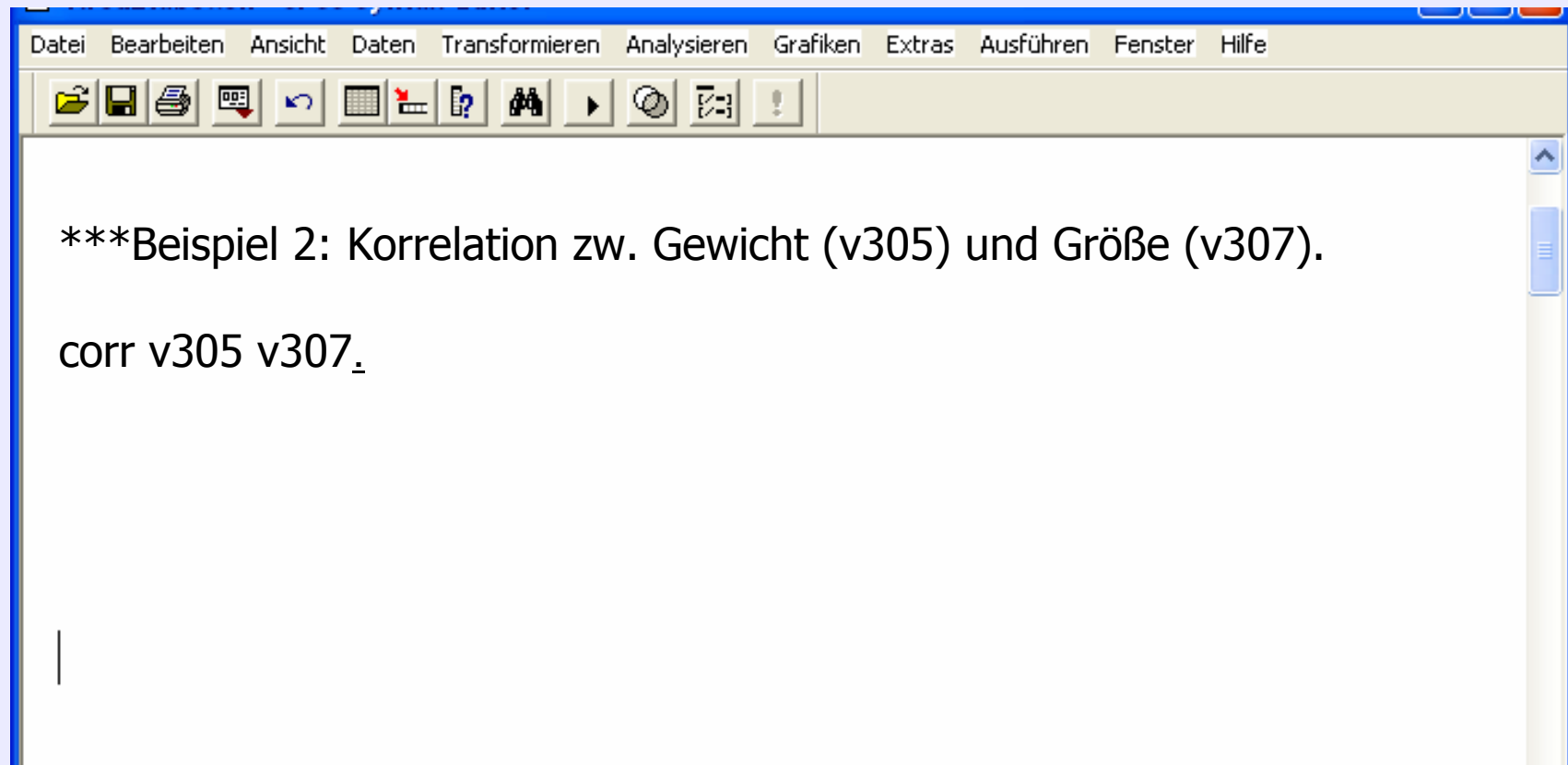
Stärke des Zusammenhangs

Faustregeln für Korrelationen bei sozialwiss. Datenanalysen

$\leq 0,05$	zu vernachlässigen
$> 0,05$ und $< 0,2$	gering
$> 0,2$ und $< 0,5$	mittel
$> 0,5$ und $< 0,7$	hoch
$\geq 0,7$	sehr hoch

➤ Werte gelten für den positiven und negativen Bereich

1. Korrelationsanalyse



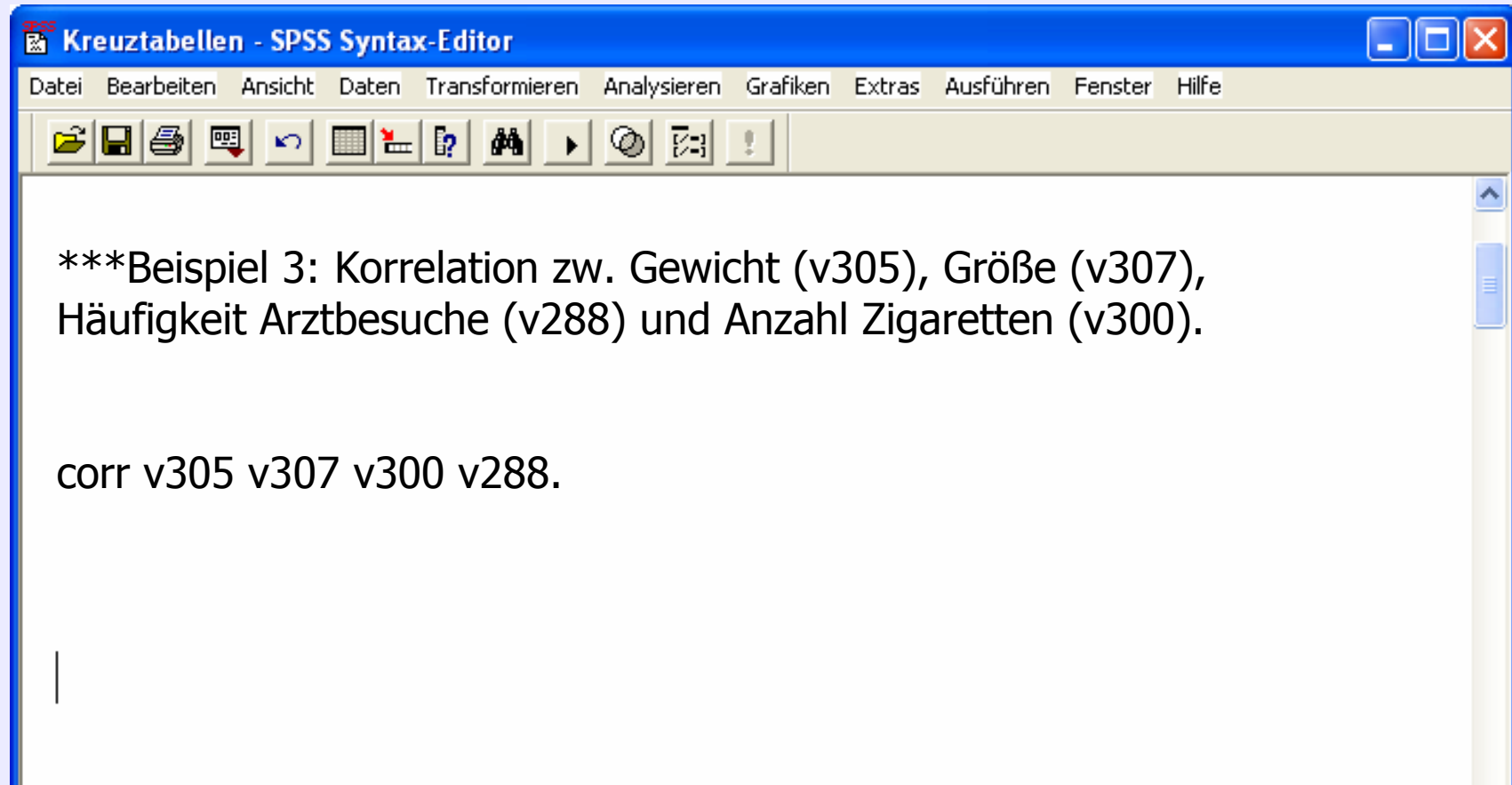
Korrelationen

		v305 KOERPER GROESSE IN CM, BEFRAGT E<R>	v307 GEWICH T IN KG, BEFRAGT E<R>
v305 KOERPERGROESSE IN CM, BEFRAGTE<R>	Korrelation nach Pearson	1	,513
	Signifikanz (2-seitig)		,000
	N	2931	2886
v307 GEWICHT IN KG, BEFRAGTE<R>	Korrelation nach Pearson	,513	1
	Signifikanz (2-seitig)	,000	
	N	2886	2887

Interpretation:

Es besteht eine hohe positive Beziehung, d.h. je größer man ist, desto schwerer ist man (bzw. umgekehrt). Die Koeffizienten sind bei einer Irrtumswahrscheinlichkeit von kleiner als 0,1% höchst signifikant.

1. Korrelationsanalyse



Korrelationen

		v305 KOERPER GROESSE IN CM, BEFRAGT E<R>	v307 GEWICH T IN KG, BEFRAGT E<R>	v300 ANZAHL GERAUCHTE R ZIGARETTEN PRO TAG	v288 ARZTBESUC HE IN DEN LETZTEN 3 MONATEN
v305 KOERPERGROESSE IN CM, BEFRAGTE<R>	Korrelation nach Pearson Signifikanz (2-seitig) N	1 2931	,513 ,000 2886	,074 ,034 817	-,035 ,127 1878
v307 GEWICHT IN KG, BEFRAGTE<R>	Korrelation nach Pearson Signifikanz (2-seitig) N	,513 ,000 2886	1 2887	,161 ,000 808	,015 ,512 1856
v300 ANZAHL GERAUCHTER ZIGARETTEN PRO TAG	Korrelation nach Pearson Signifikanz (2-seitig) N	,074 ,034 817	,161 ,000 808	1 818	,050 ,289 459
v288 ARZTBESUCHE IN DEN LETZTEN 3 MONATEN	Korrelation nach Pearson Signifikanz (2-seitig) N	-,035 ,127 1878	,015 ,512 1856	,050 ,289 459	1 1884

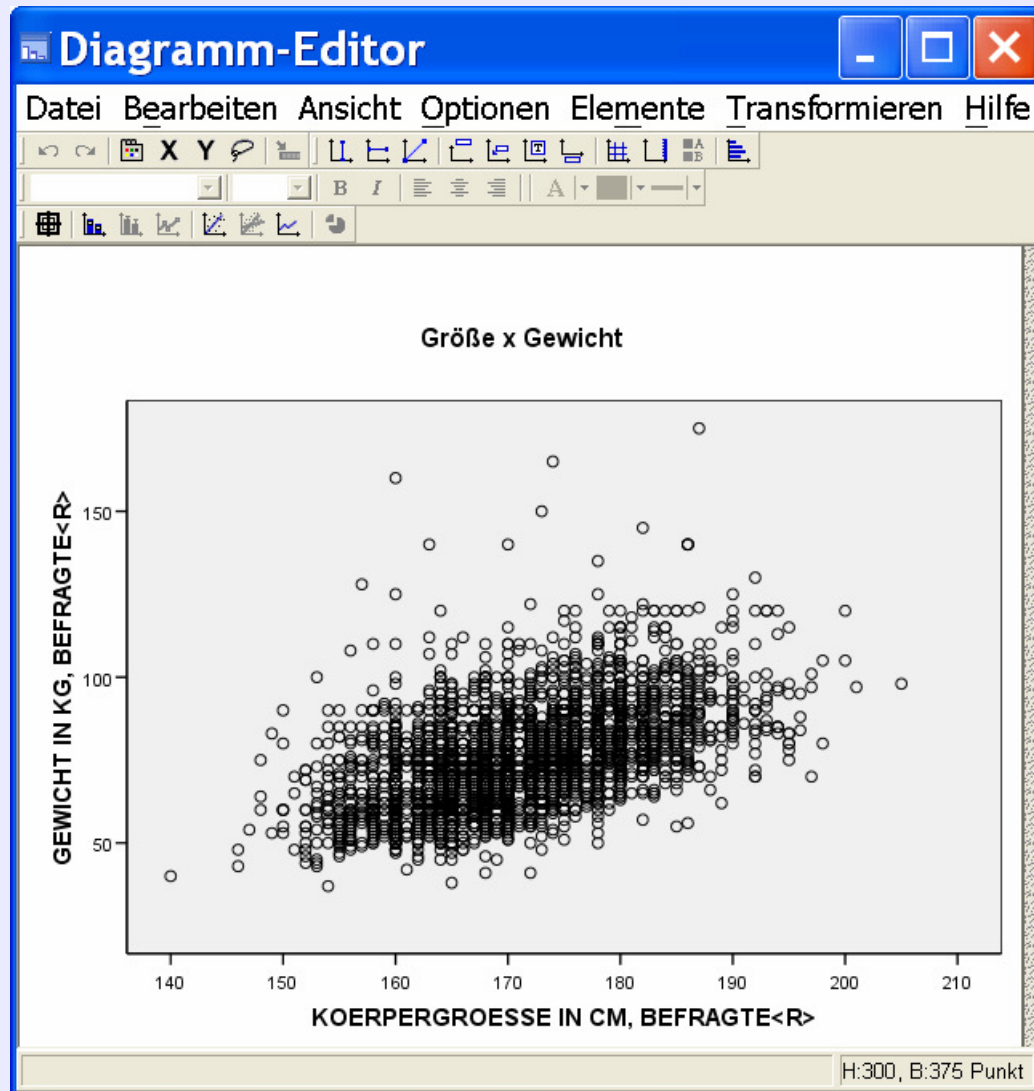
2. Streudiagramm

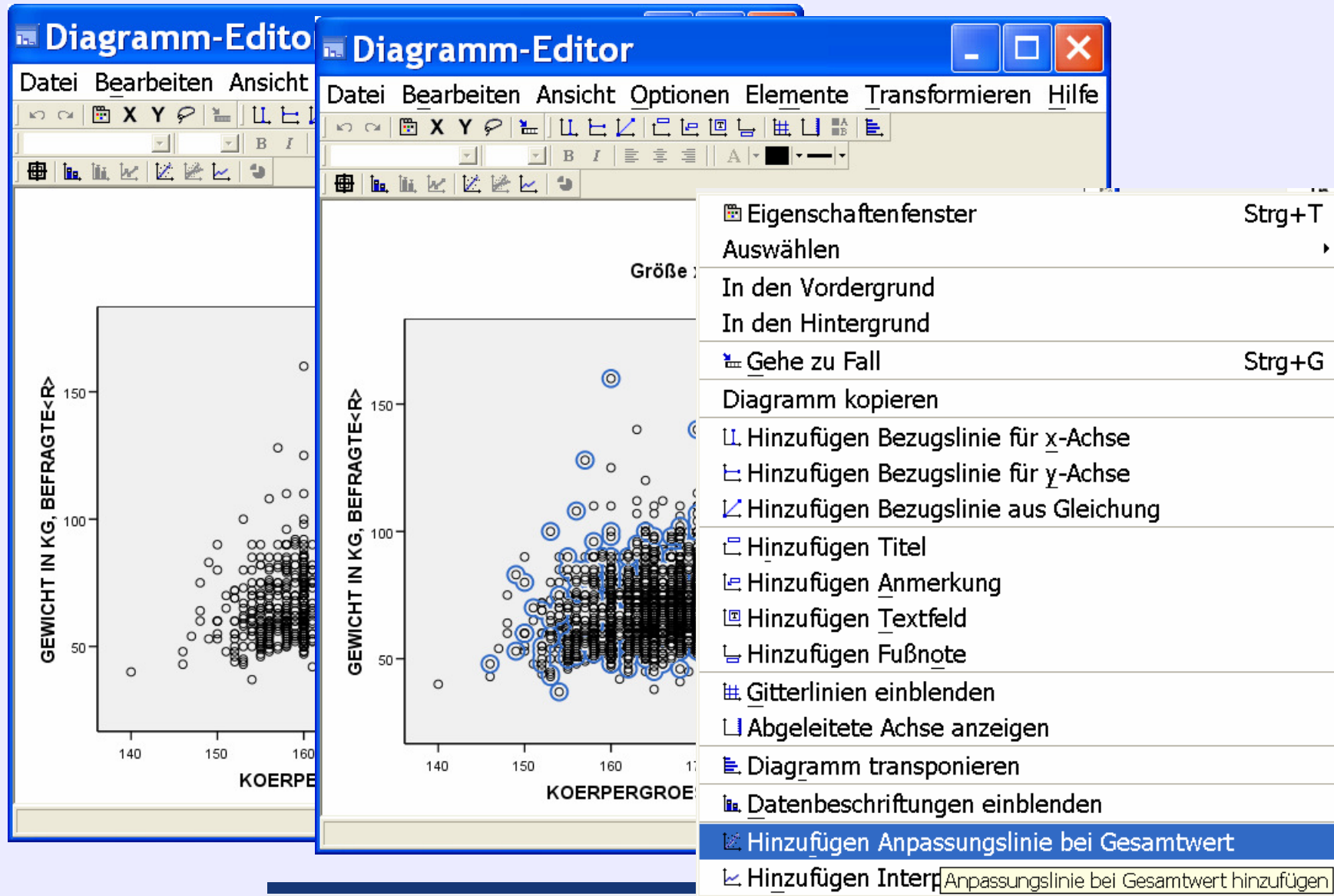
- Da Korrelationskoeffizienten nur lineare Zusammenhänge identifizieren können, ist es empfehlenswert, die Werte der Variablen in einem Diagramm abzubilden.
- In einem Streudiagramm lässt sich häufig die Stärke und die Form eines möglichen Zusammenhangs beobachten.

2. Streudiagramm

***Beispiel 1: Streudiagramm, Allbus 2004.

```
graph scatterplot v305 with v307  
/title 'Größe x Gewicht'.
```





Eigenschaften

Diagrammgröße Linien Anpassungslinie

☐ Projektionslinien anzeigen

Anpassungsmethode



☐ Mittelwert von Y



☒ Linear



☐ Loess

% der Punkte für die Anpassung: 50

Kern: Epanechnikov

Konfidenzintervalle

☒ Keine

☐ Mittelwert

☐ Individuell

%, 95

Zuweisen

Abbrechen

Diagramm-Editor

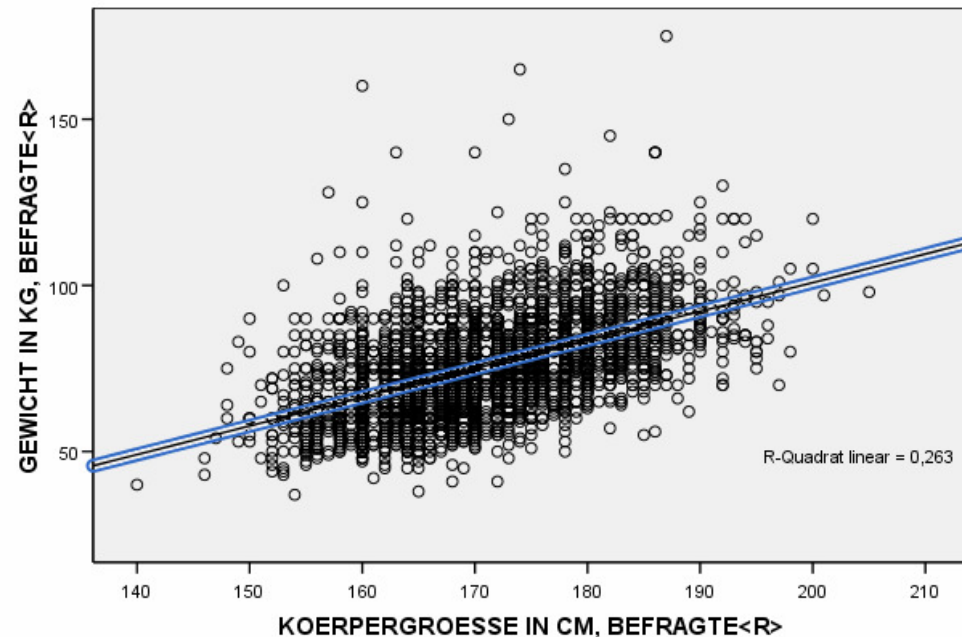
Datei Bearbeiten Ansicht Optionen Elemente Transformieren Hilfe

X Y

B I

A

Größe x Gewicht



H:300, B:375 Punkt