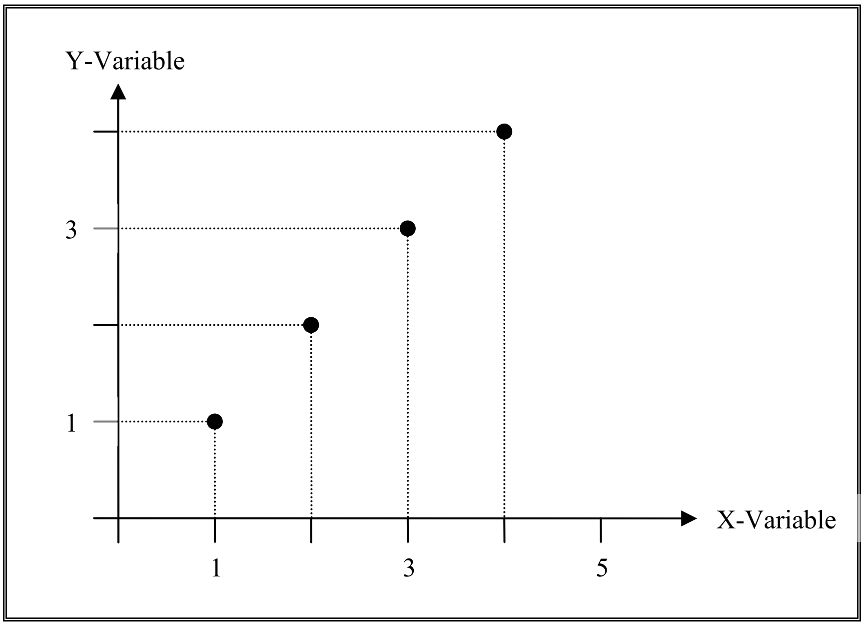
Streudiagramm und Korrelation

- für metrisches Messniveau

<u>Beispiel</u>:

Person	Haushaltsgröße (x_i)	Anzahl privat genutzter PKWs (y_i)
A	1	1
B	2	2
C	3	3
D	4	4

Graphische Umsetzung in ein Streudiagramm:

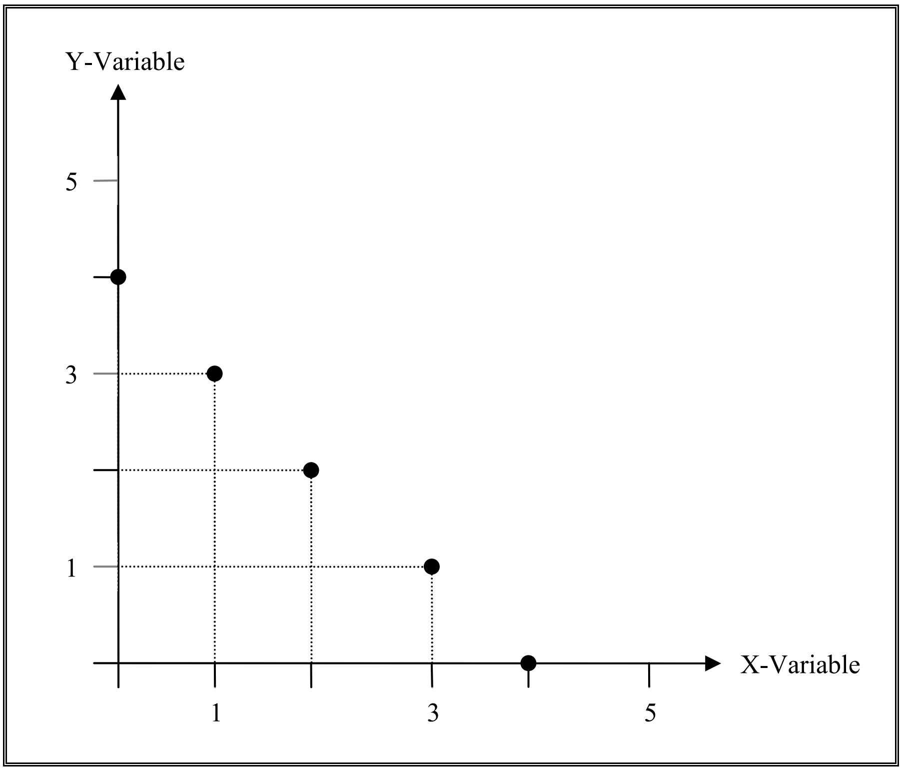


- Es besteht eine (perfekt) **positive lineare Beziehung (Korrelation)** zwischen den Variablen.

<u>Beispiel</u>:

Person	x_i	y_i
A	0	4
B	1	3
C	2	2
D	3	1
E	4	0

Graphische Umsetzung in ein Streudiagramm:



- Es besteht eine (perfekte) **negative lineare Beziehung (Korrelation)** zwischen den Variablen.

Kurvilineare Beziehung:

1) u-förmiger Verlauf der Punkte im Streudiagramm
2) j-förmiger Verlauf der Punkte im Streudiagramm
3) ……

Pearsonsche Produkt-Moment-Korrelationskoeffizient r:

1. Berechnungsmöglichkeit:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$$

2. Berechnungsmöglichkeit:

$$r = \frac{n \cdot \sum x \cdot y - \sum x \sum y}{\sqrt{(n \cdot \sum x^2 - (\sum x)^2) \cdot (n \cdot \sum y^2 - (\sum y)^2)}}$$

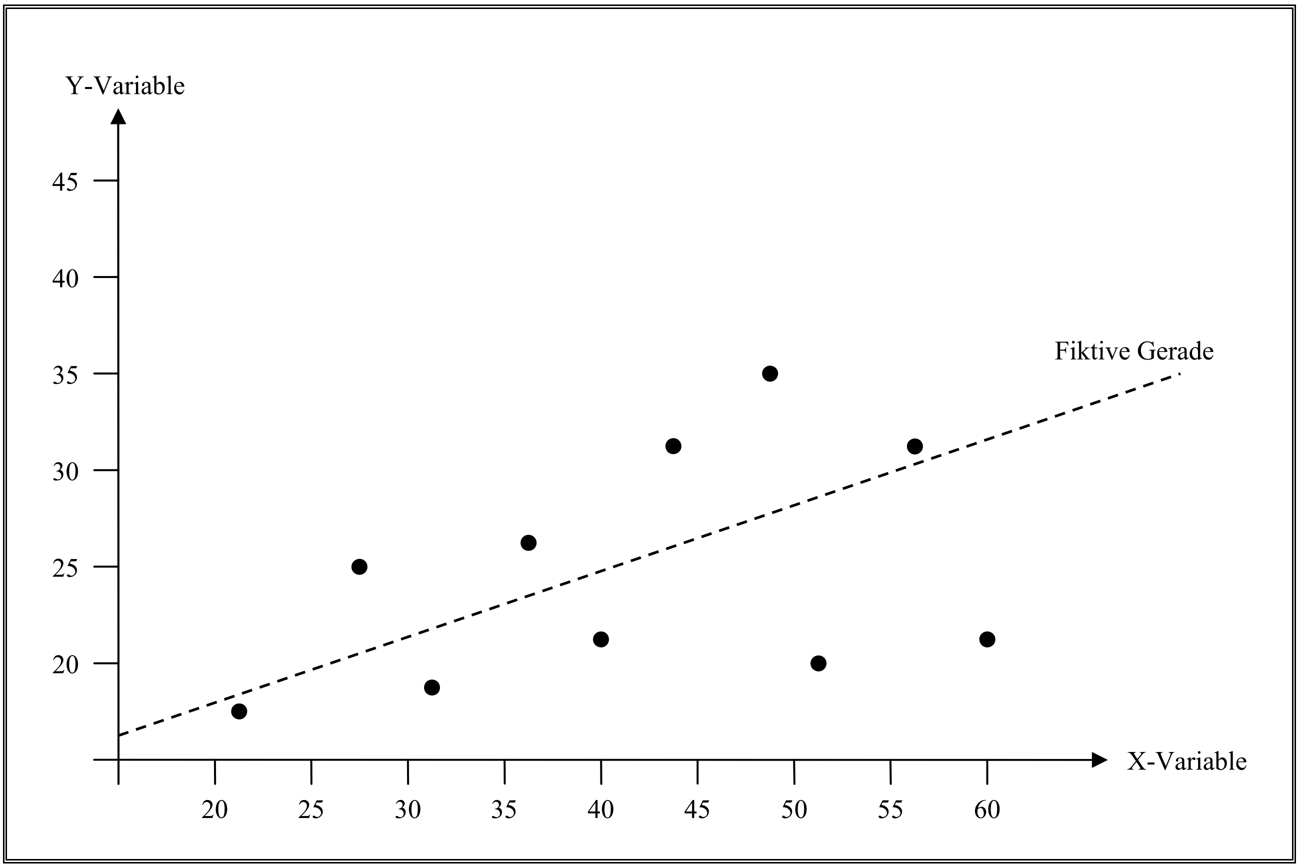
Wertebereich	
r = +1	• Es liegt ein **perfekt positive Beziehung** vor • Alle Punkte liegen auf einer **steigenden** Geraden
r nahe +1	• Es liegt eine **sehr starke positive Beziehung** vor • Fast alle Punkte liegen auf einer **steigenden** Geraden
r = 0	• Es liegt **keine statistische Beziehung** vor • Die Punkte ordnen sich kreisförmig an, so dass keine Gerade lokalisiert werde kann, die die Punkteverteilung am besten repräsentiert
r nahe -1	• Es liegt eine **sehr starke negative Beziehung** vor • Fast alle Punkte liegen auf einer **fallenden** Geraden
r = -1	• Es liegt eine **perfekt negative Beziehung** vor • Alle Punkte liegen auf einer **fallenden** Geraden

Zur Stärke der Beziehung	
bis 0,2	schwach
0,2 bis 0,4	niedrig
0,4 bis 0,7	mäßig
0,7 bis 0,9	hoch
über 0,9	sehr hoch

Anwendungsbeispiel:

Person	Alter (x_i)	Monatliches Einkommen (in 100 Euro) (y_i)
A	22	12
B	28	24
C	32	14
D	36	26
E	40	18
F	44	28
G	48	32
H	52	16
I	56	30
J	62	20

Graphische Umsetzung in ein Streudiagramm:



Pearsonsche Produkt-Moment-Korrelationskoeffizient r:

Person	x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x}) \cdot (y_i - \bar{y})$
A	22	12	-20	400	-10	100	200
B	28	24	-14	196	2	4	-28
C	32	14	-10	100	-8	64	80
D	36	26	-6	36	4	16	-24
E	40	18	-2	4	-4	16	8
F	44	28	2	4	6	36	12
G	48	32	6	36	10	100	60
H	52	16	10	100	-6	36	-60
I	56	30	14	196	8	64	112
J	62	20	20	400	-2	4	-40
∑	420	220	0	1.472	0	440	320

$$\bar{x} = \frac{420}{10} = 42 \qquad \bar{y} = \frac{220}{10} = 22$$

1. Berechnungsmöglichkeit

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}} = \frac{320}{\sqrt{1.472 \cdot 440}} = \frac{320}{\sqrt{647.680}} = \frac{320}{804{,}79} = 0{,}398$$

Person	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
A	22	12	264	484	144
B	28	24	672	784	576
C	32	14	448	1.024	196
D	36	26	936	1.296	676
E	40	18	720	1.600	324
F	44	28	1.232	1.936	784
G	48	32	1.536	2.304	1.024
H	52	16	832	2.704	256
I	56	30	1.680	3.136	900
J	62	20	1.240	3.844	400
∑	420	220	9.560	19.112	5.280

2. Berechnungsmöglichkeit:

$$r = \frac{n \cdot \sum x \cdot y - \sum x \sum y}{\sqrt{(n \cdot \sum x^2 - (\sum x)^2)(n \cdot \sum y^2 - (\sum y)^2)}} = \frac{10 \cdot 9.560 - 420 \cdot 220}{\sqrt{(10 \cdot 19.112 - 420^2) \cdot (10 \cdot 5.280 - 220^2)}}$$

$$= \frac{95.600 - 92.400}{\sqrt{(14.720) \cdot (4.400)}} = \frac{3.200}{\sqrt{64.768.000}} = 0{,}398$$