

Statistik für Biologen

Korrelation

Wie messen wir
die Stärke
der Abhängigkeit
zwischen Variablen?

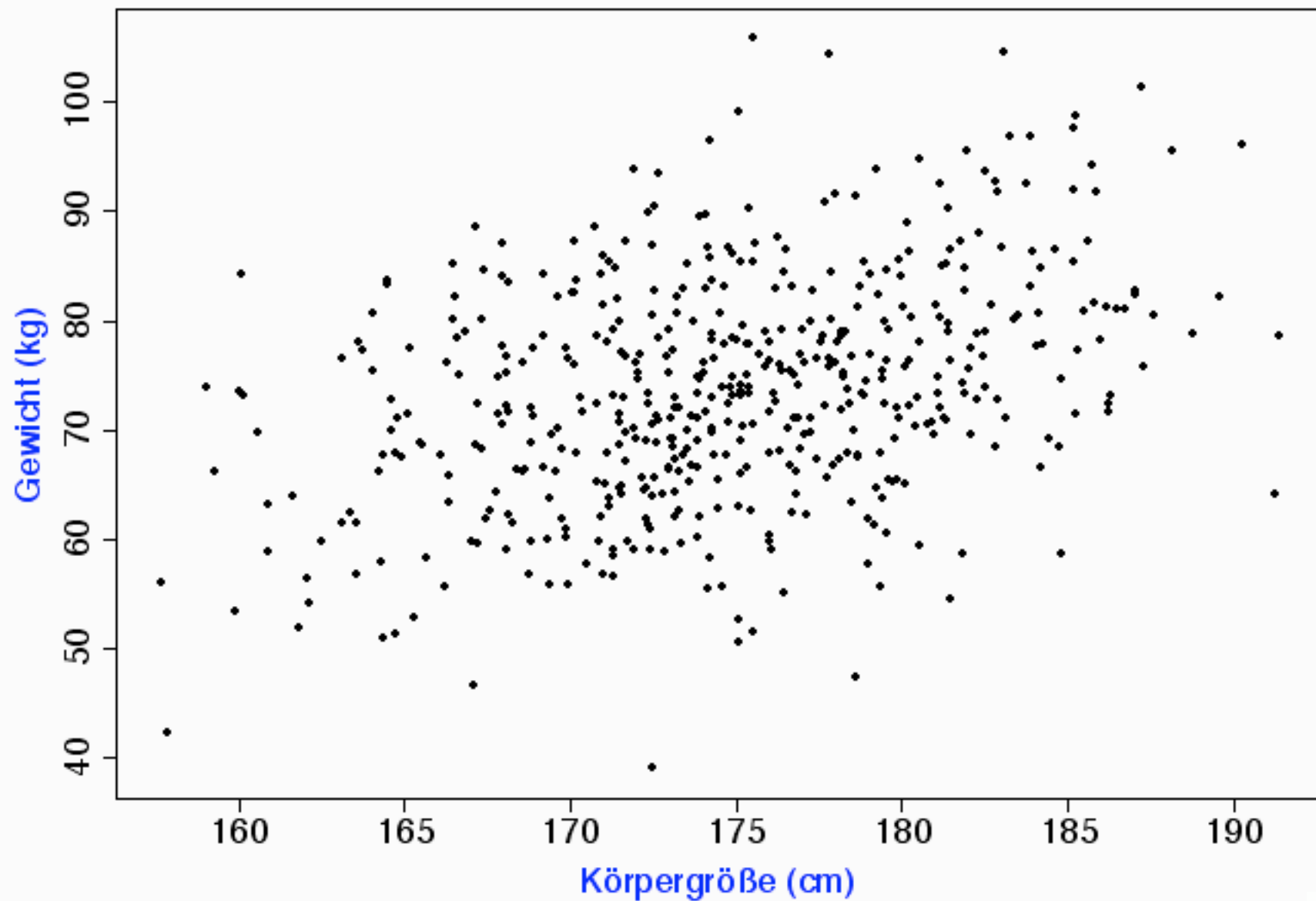
Beispiel:

Körpergröße

und

Gewicht

Männliche US-Amerikaner 18–24 J. (n=500)



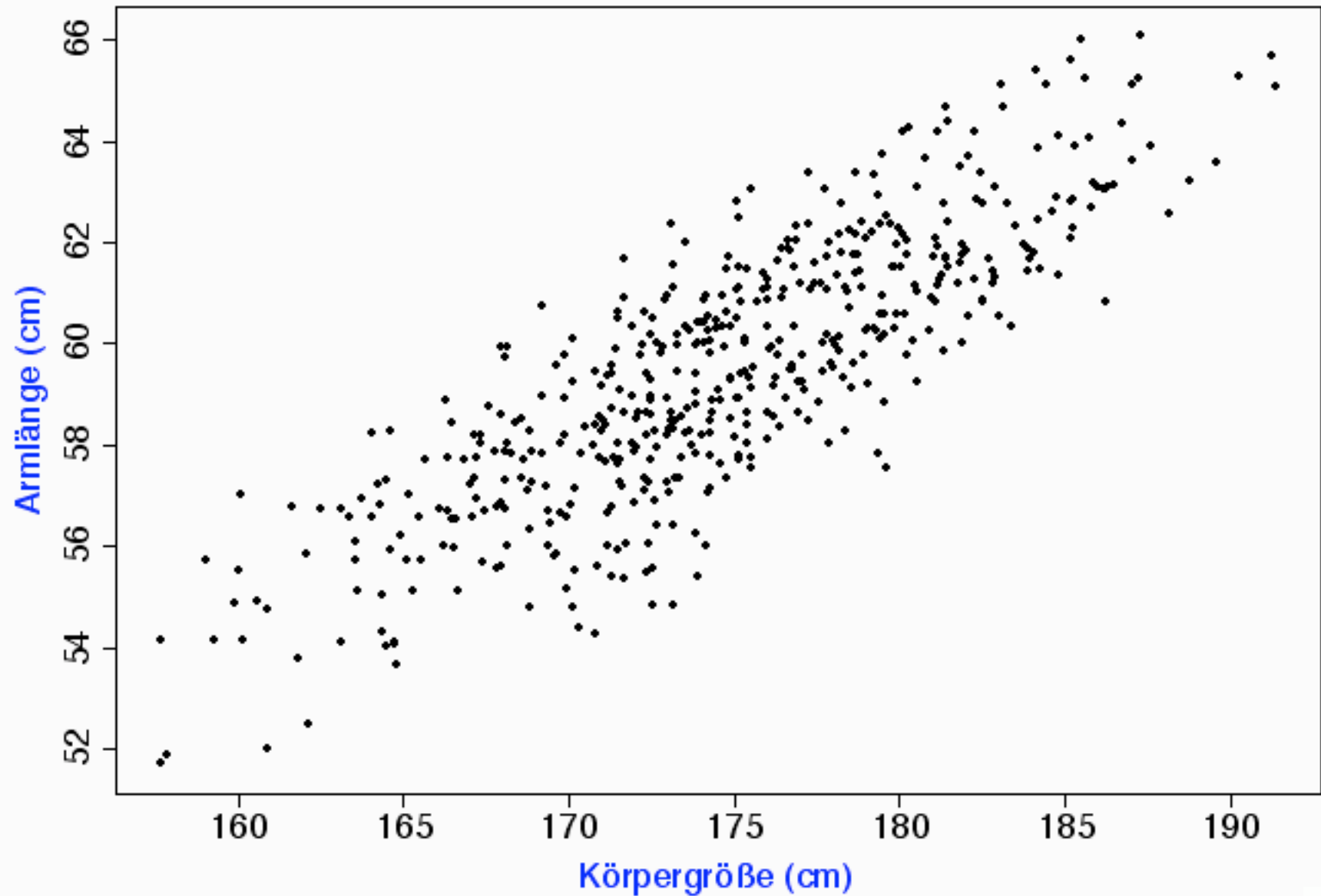
Ein klarer Zusammenhang
aber nicht sehr stark.

Körpergröße

und

Armlänge

Männliche US-Amerikaner 18–24 J. (n=500)

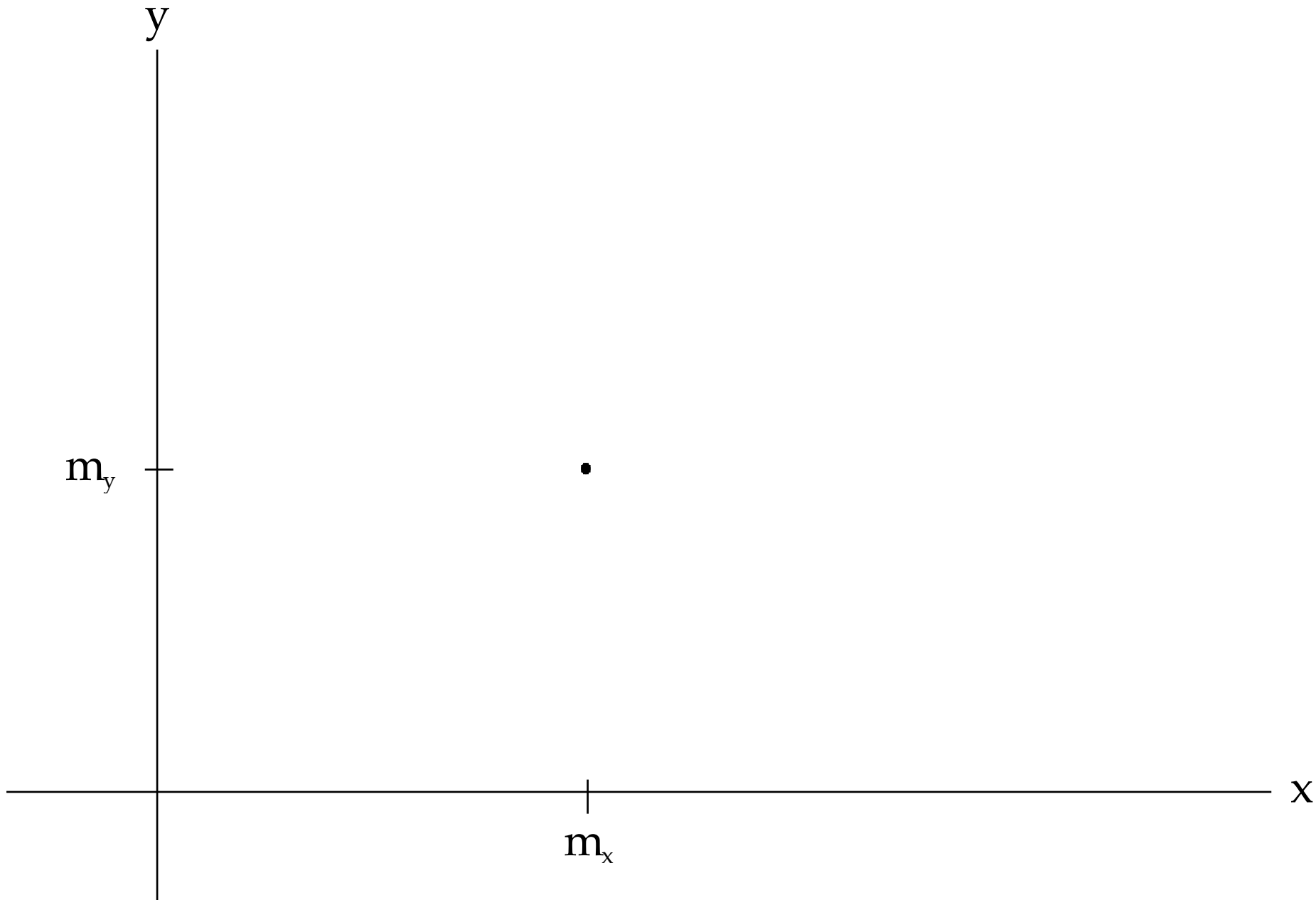


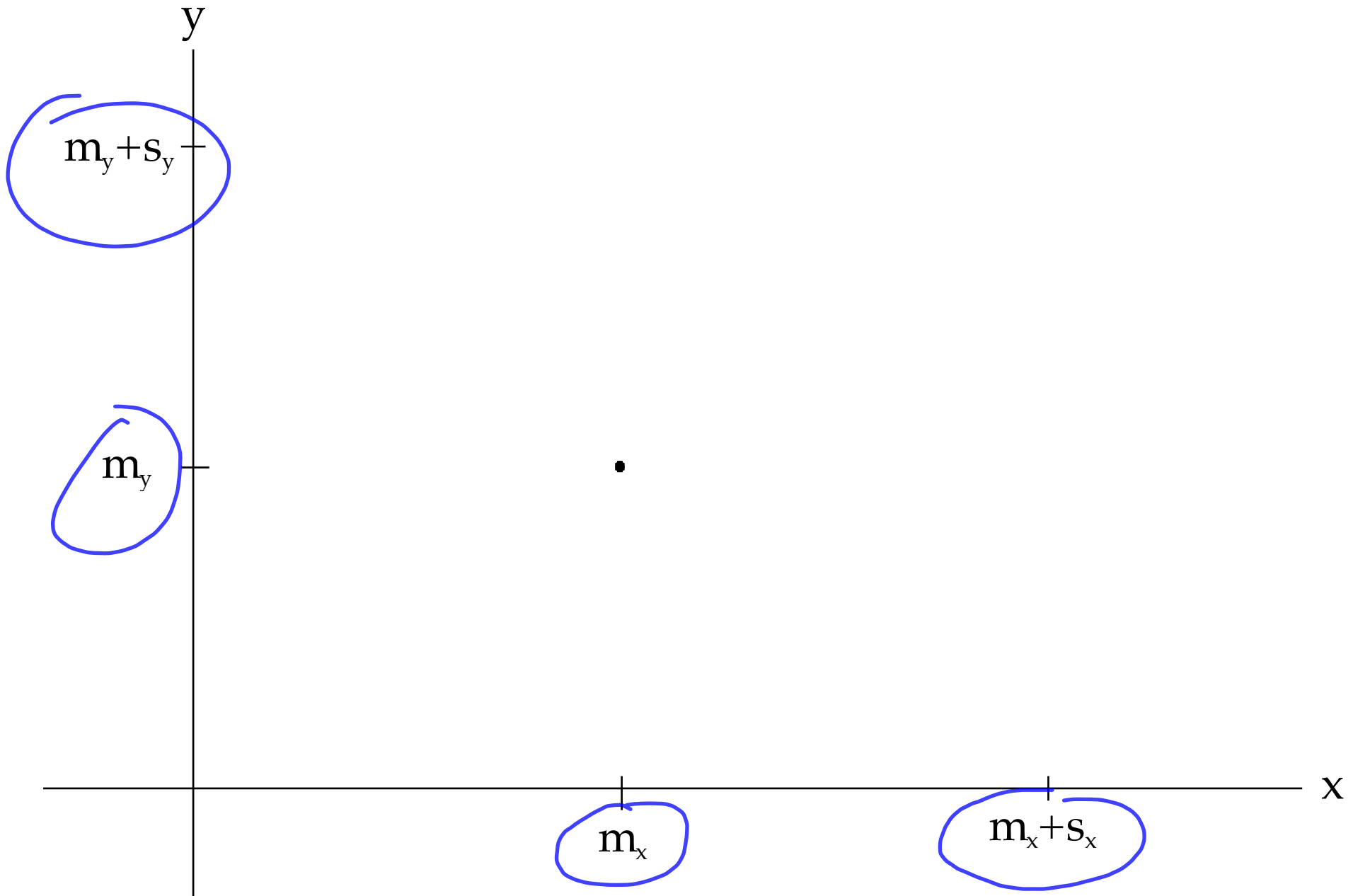
Hier ist der Zusammenhang
viel stärker.

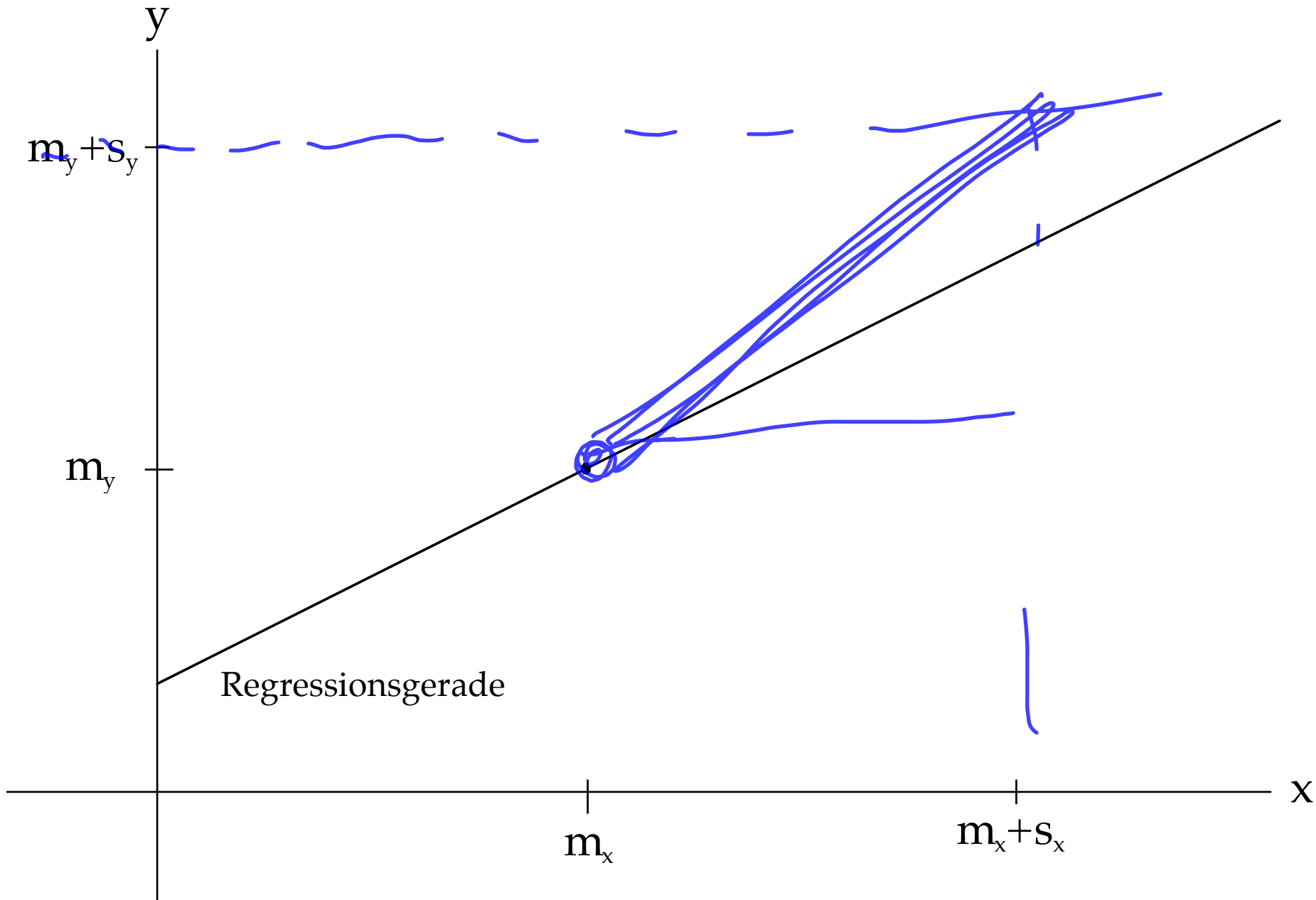
Wie können wir
die Stärke des Zusammenhangs
quantitativ
erfassen?

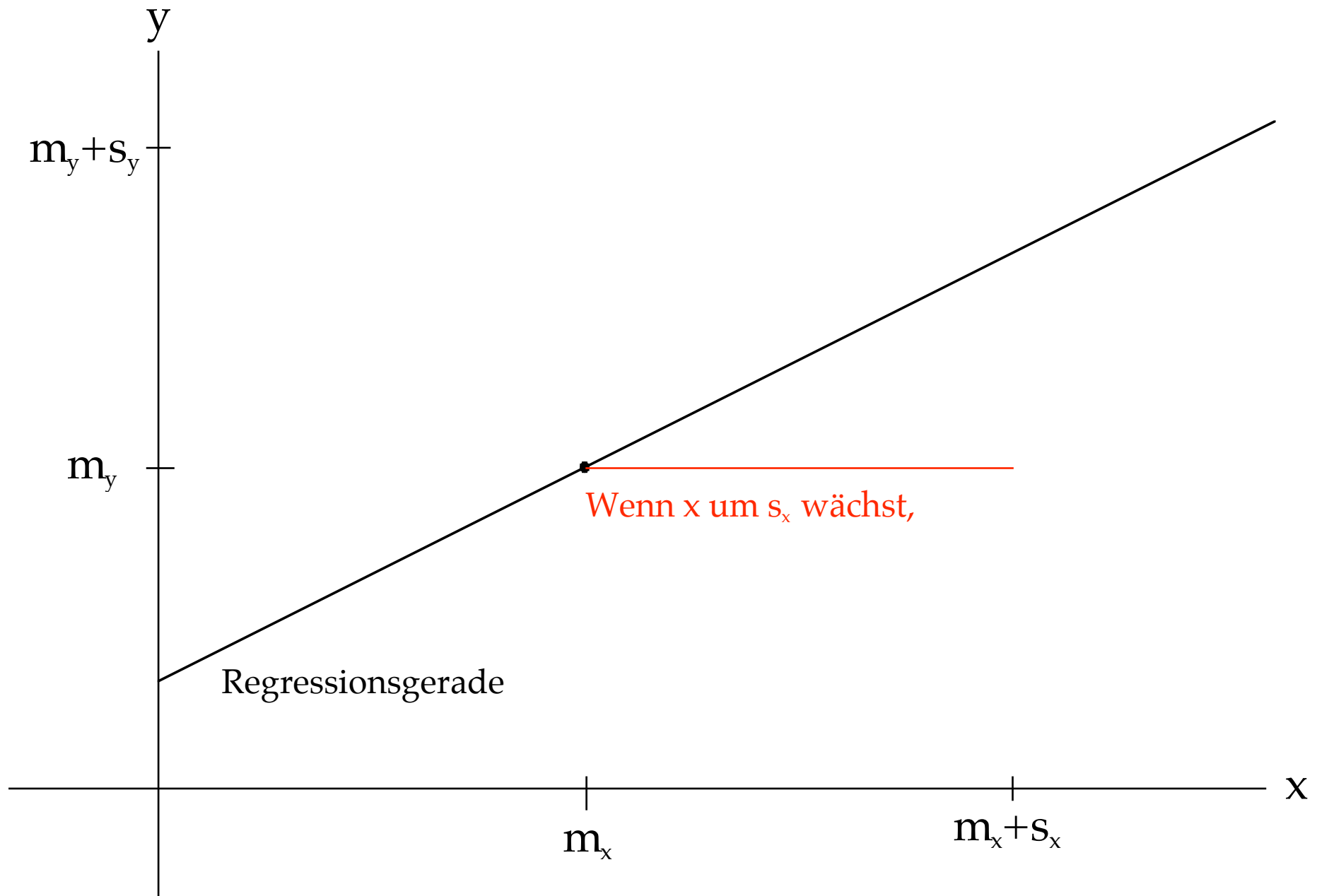
Der
Korrelationskoeffizient

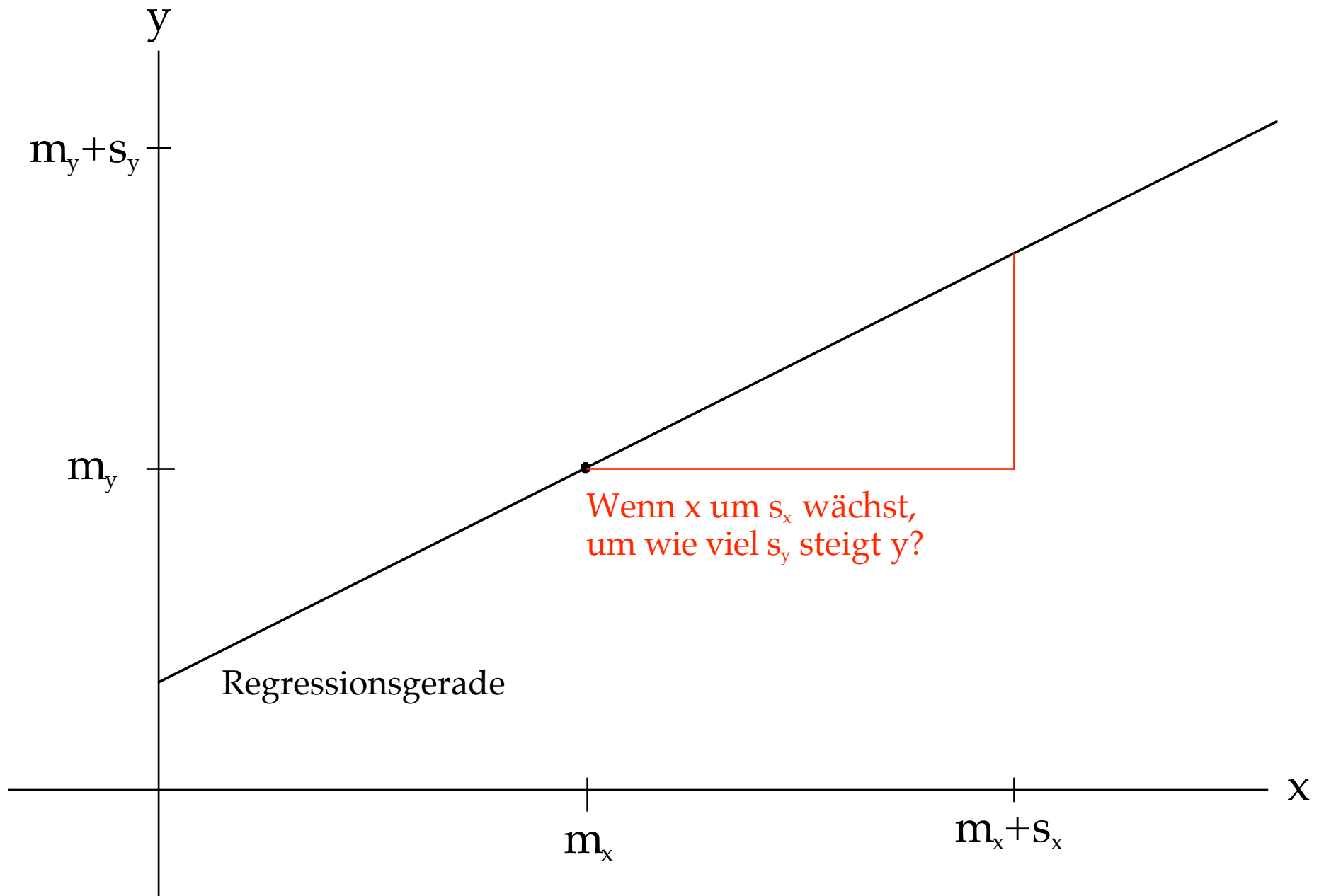
r

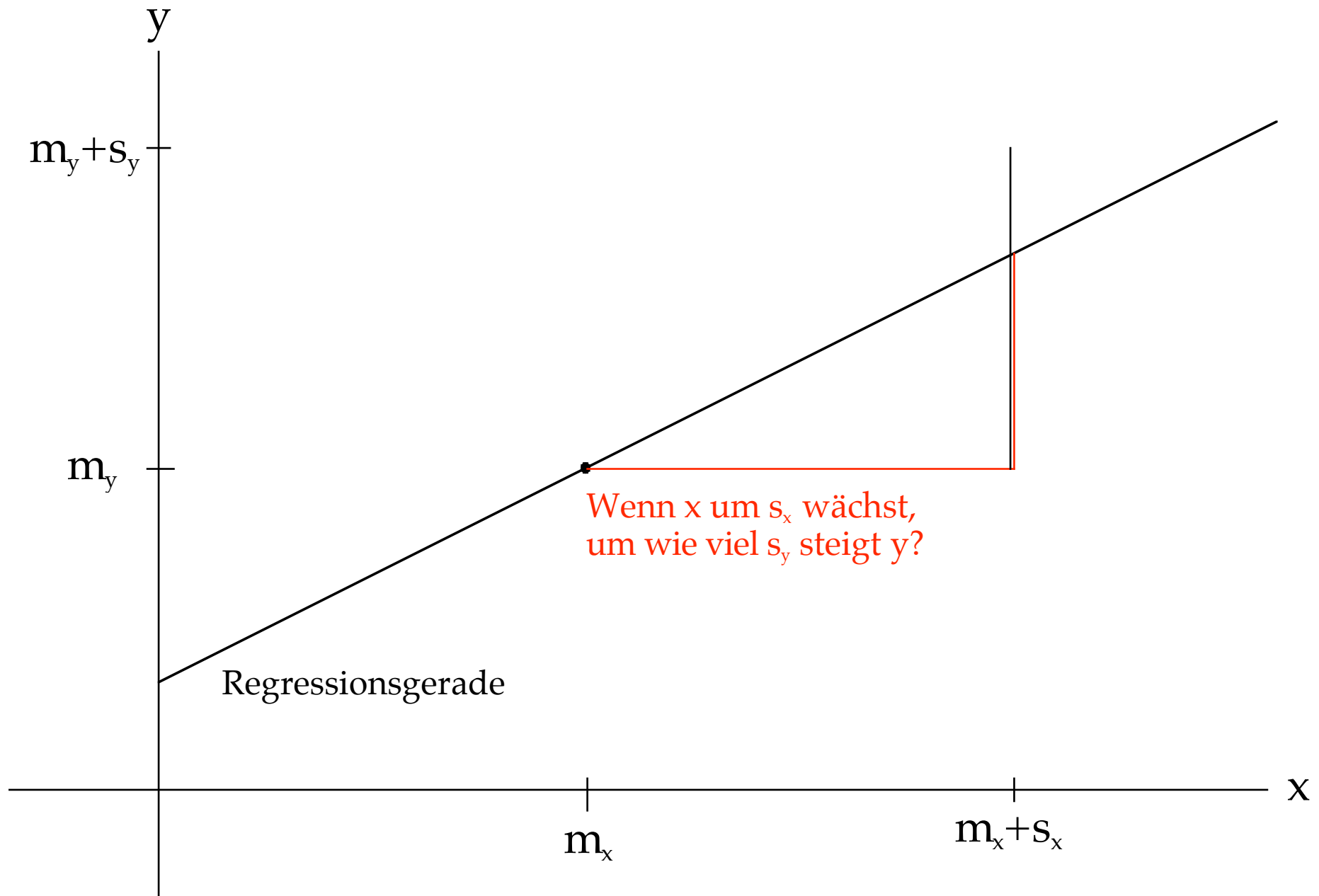


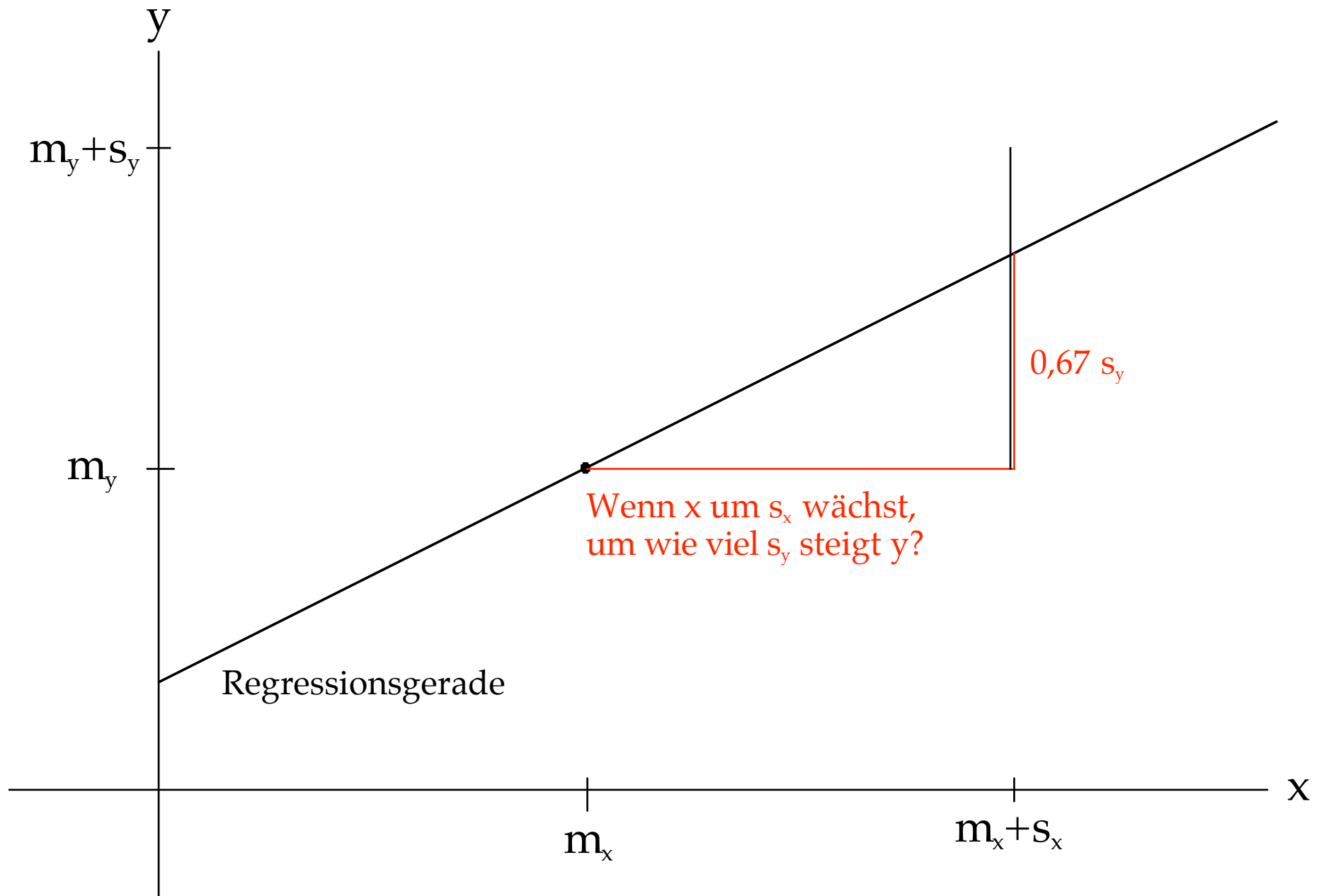


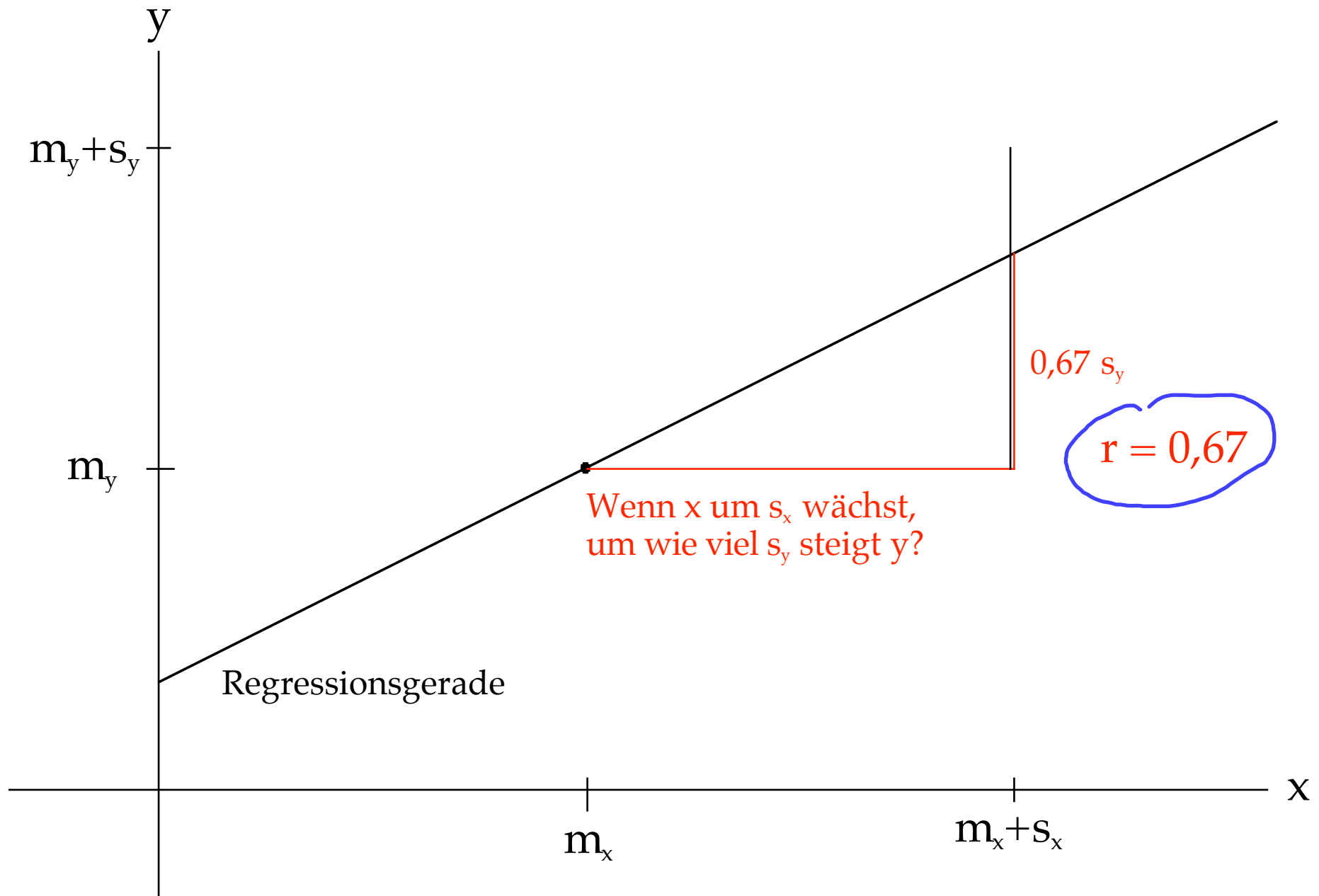












Wenn x
um s_x größer wird,
wird y
im Schnitt um
 $r \cdot s_y$
größer.

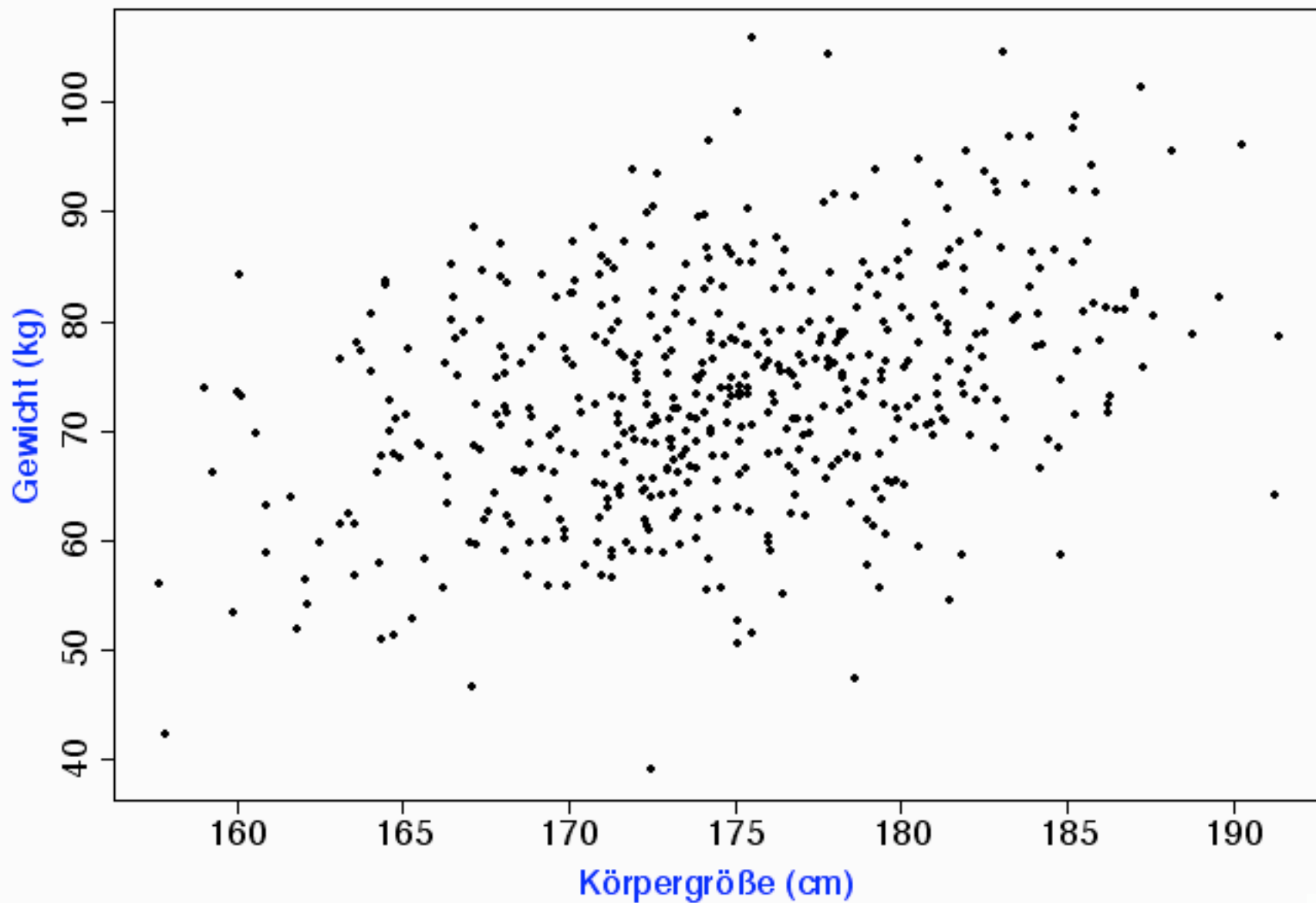
$r = \frac{\Delta y}{\Delta x}$

Körpergröße

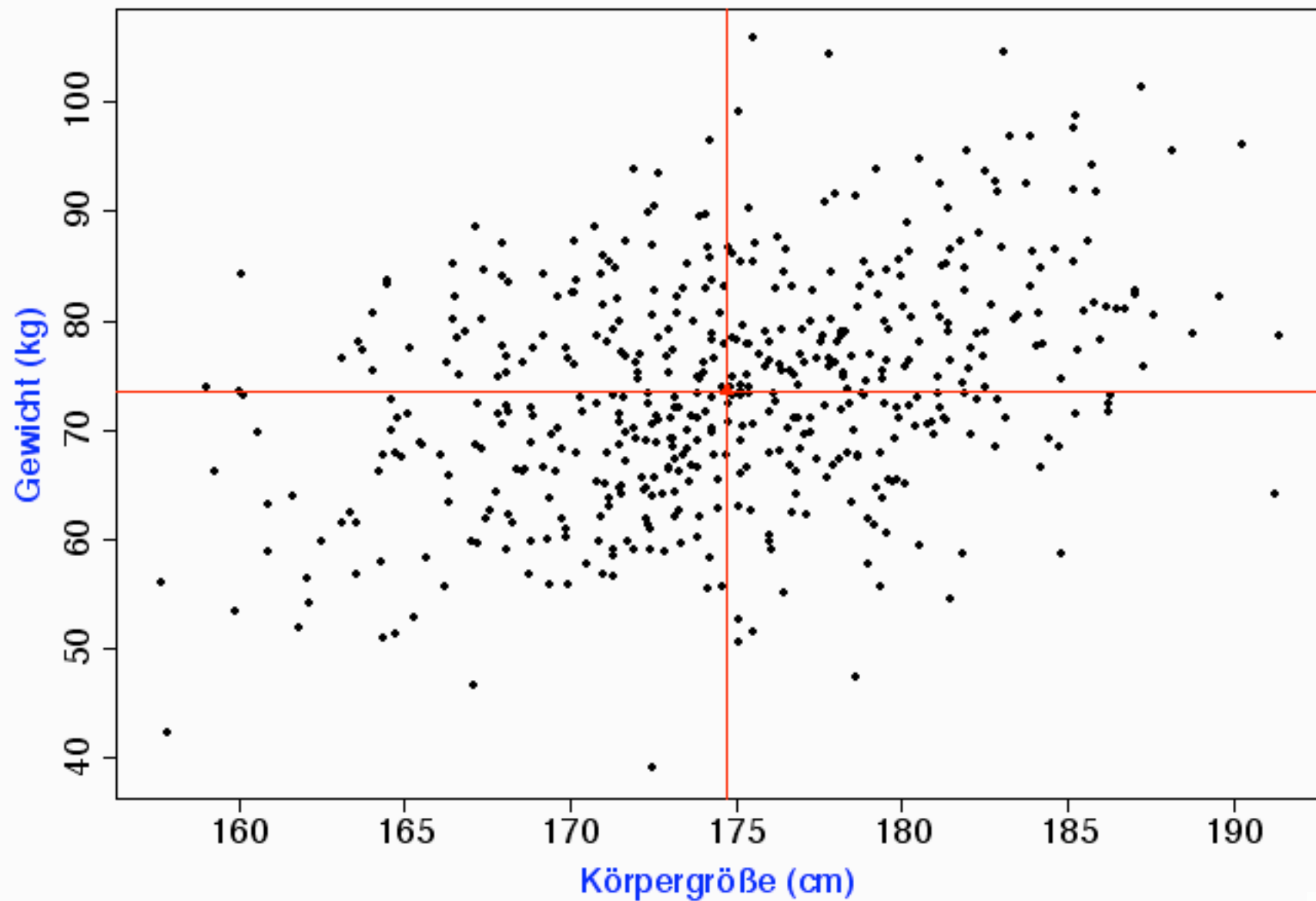
und

Gewicht

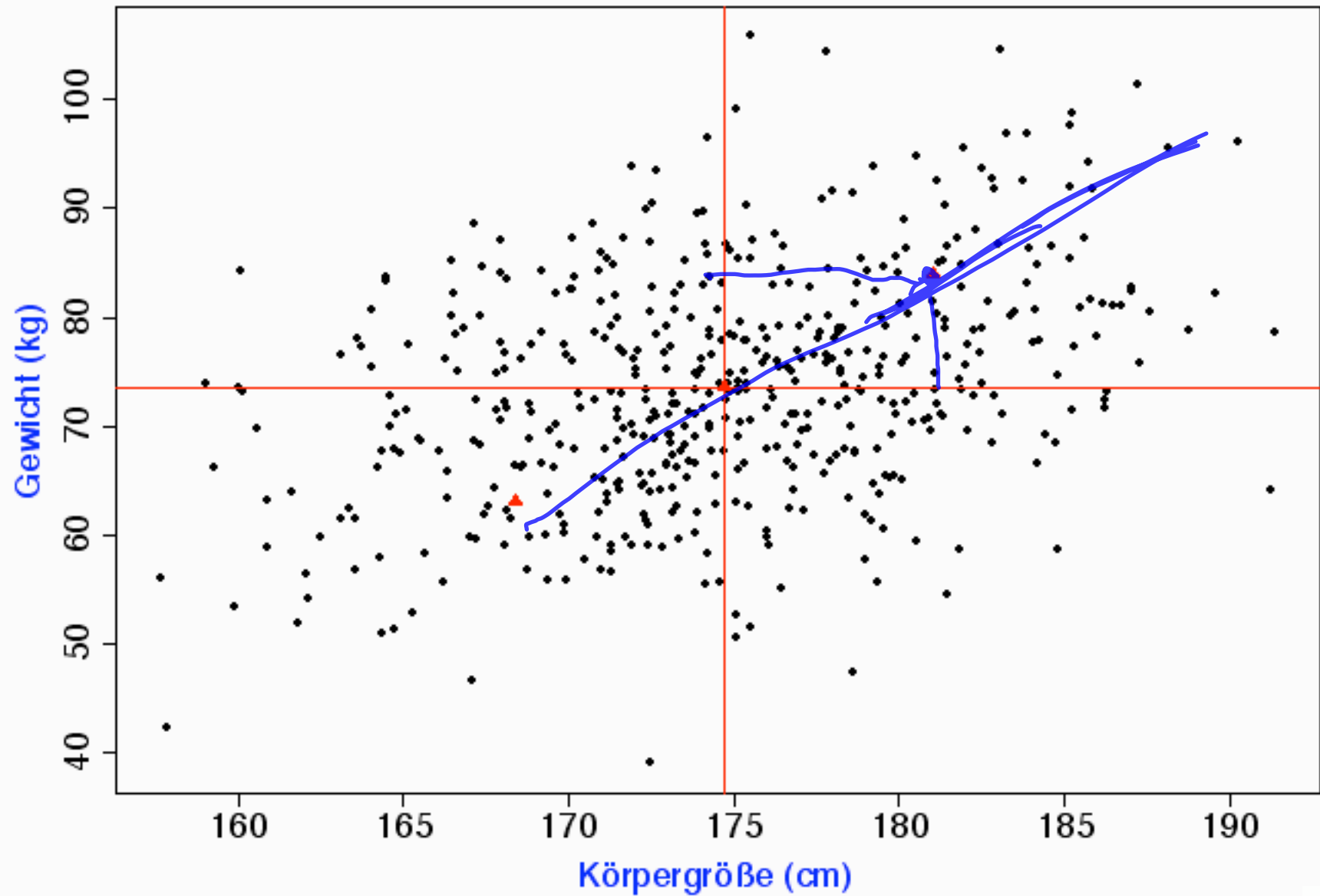
Männliche US-Amerikaner 18–24 J. (n=500)



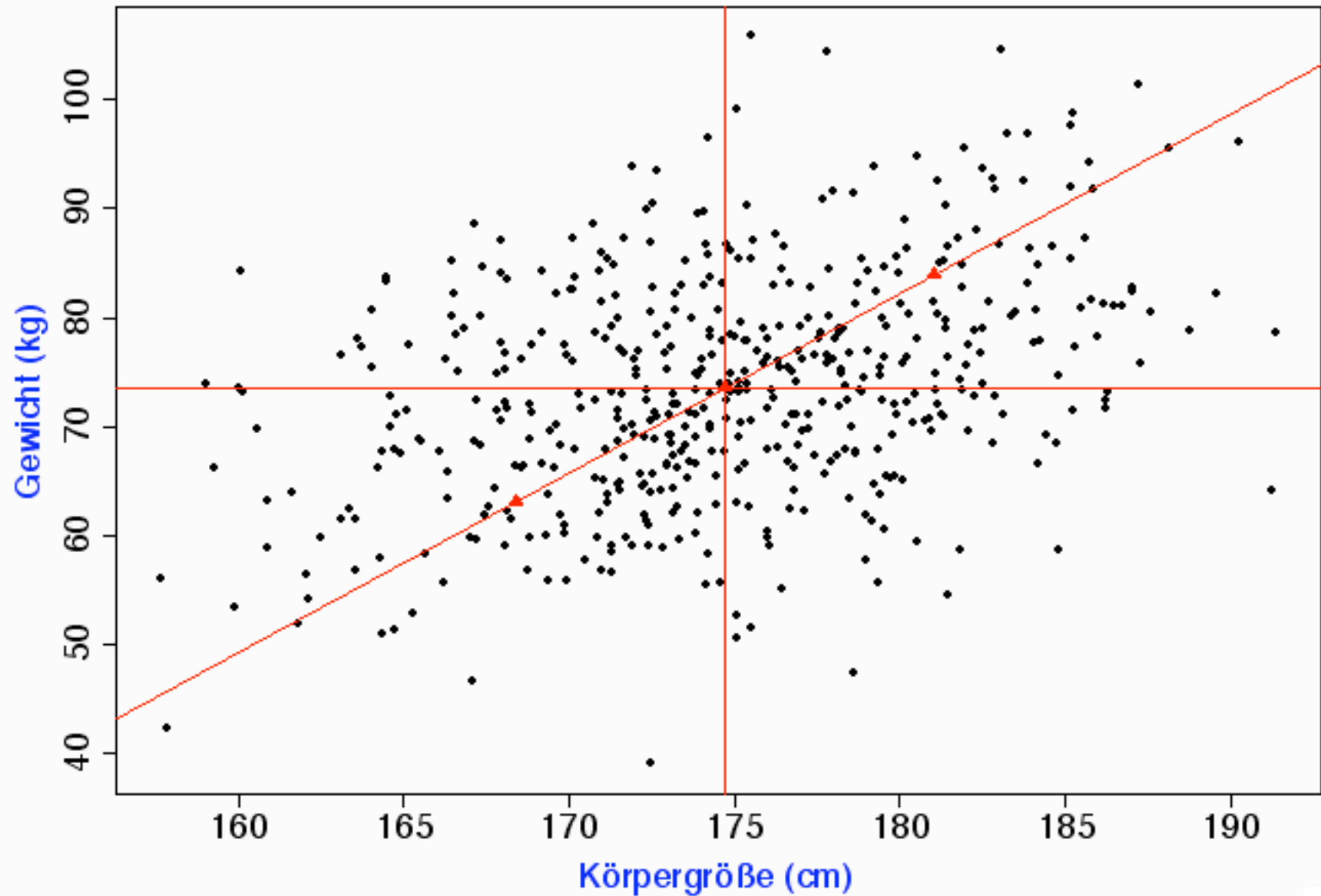
Mittelwerte



Mittelwerte \pm Standardabweichungen



Mittelwerte \pm Standardabweichungen



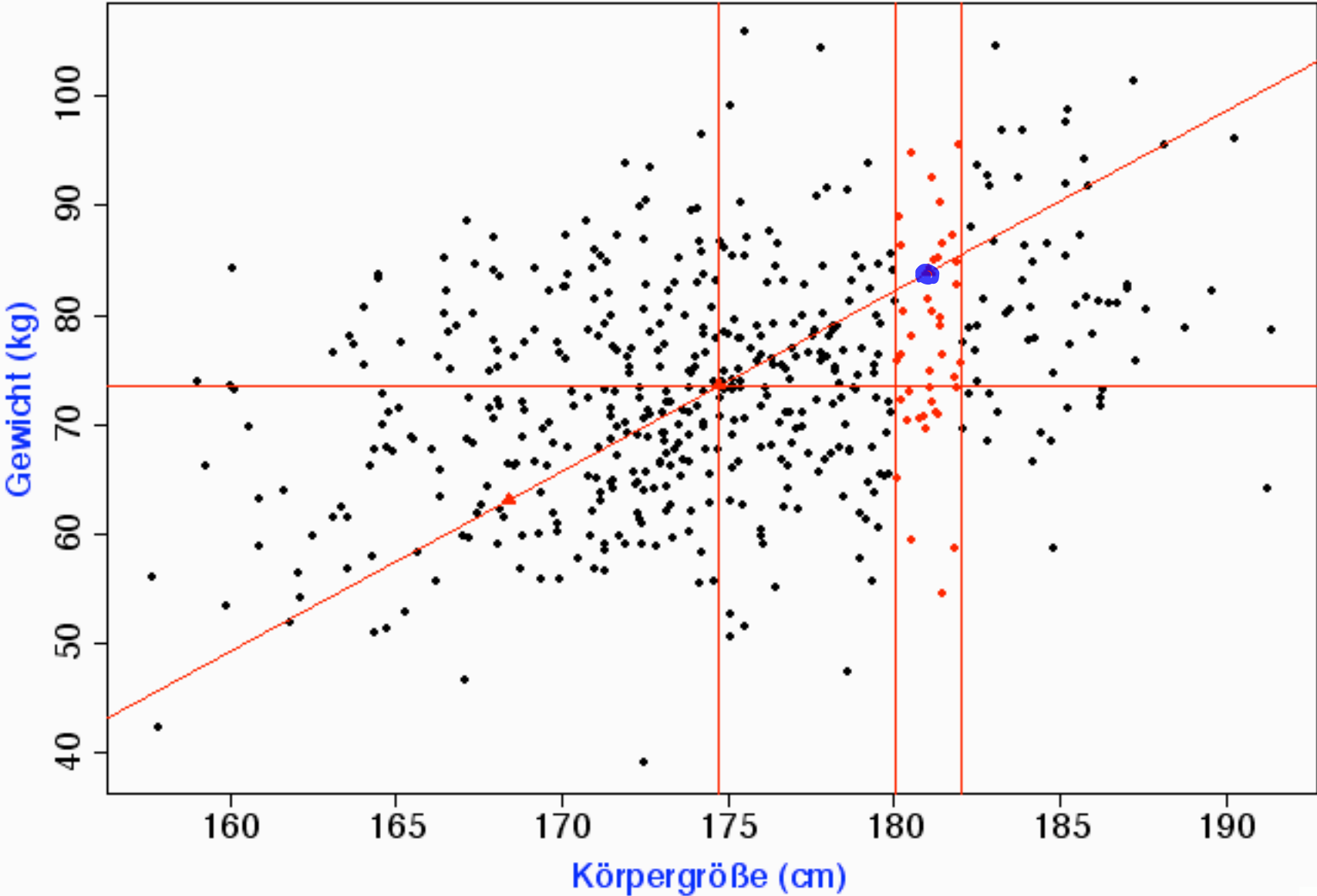
Ein Mann ist
eine Standardabweichung größer
als der Durchschnitt.

Was wiegt er?

(Im Schnitt)

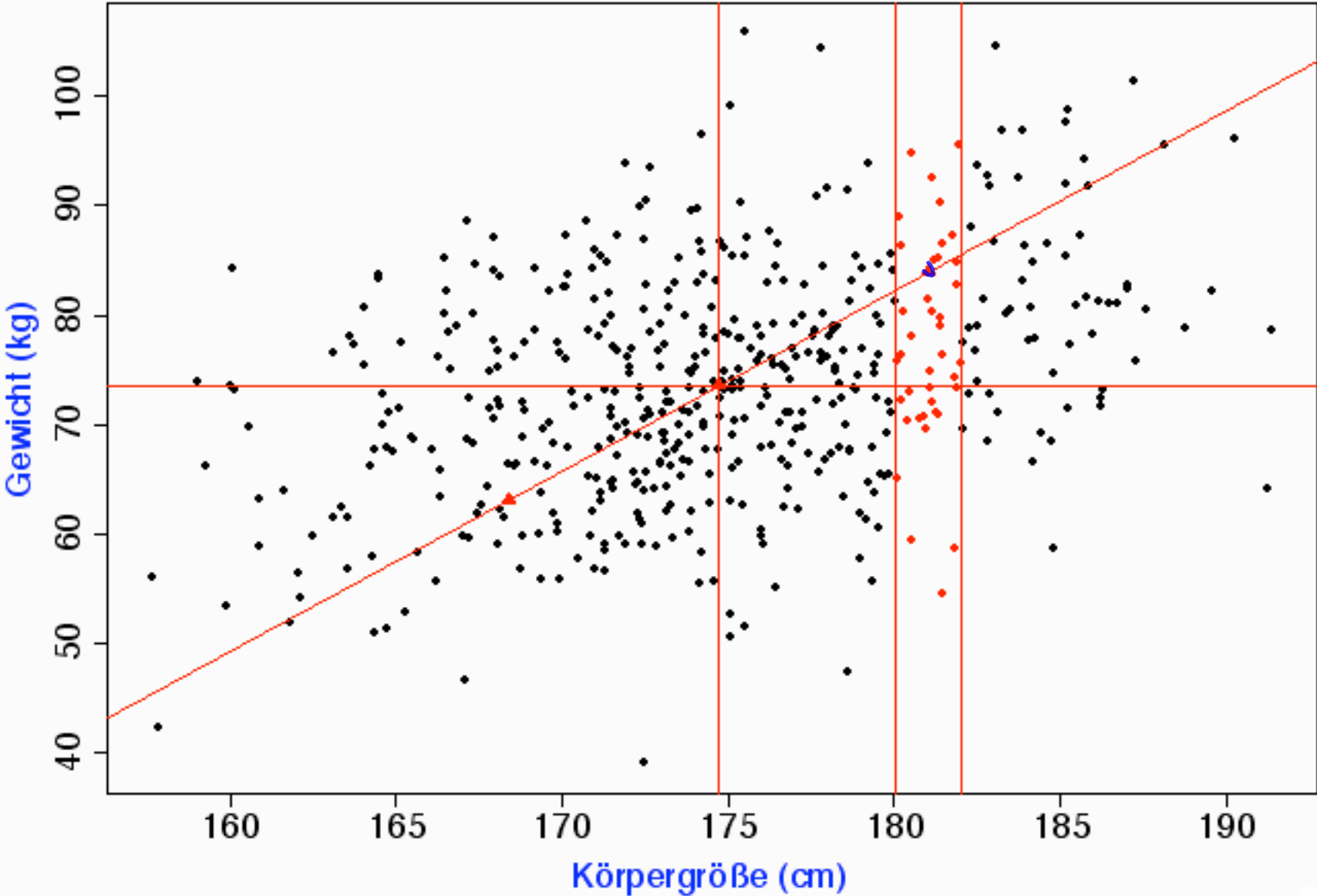
(Wie viele Standardabweichungen
schwerer als der Durchschnitt?)

Körpergröße 1 StdAbw oberhalb des Durchschnitts (± 1 cm)



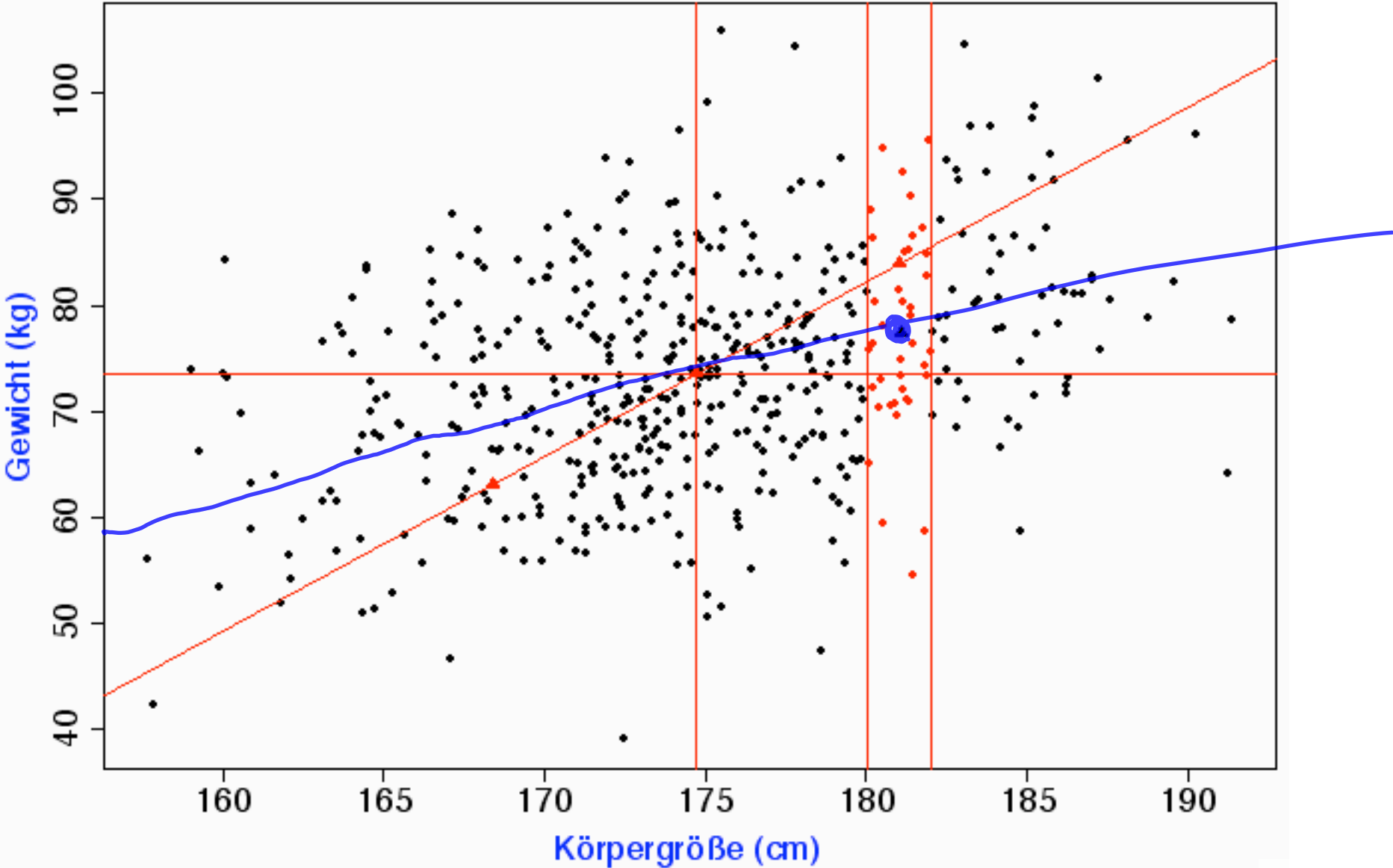
Die meisten Punkte
liegen unterhalb der „Diagonale“.

Körpergröße 1 StdAbw oberhalb des Durchschnitts (± 1 cm)



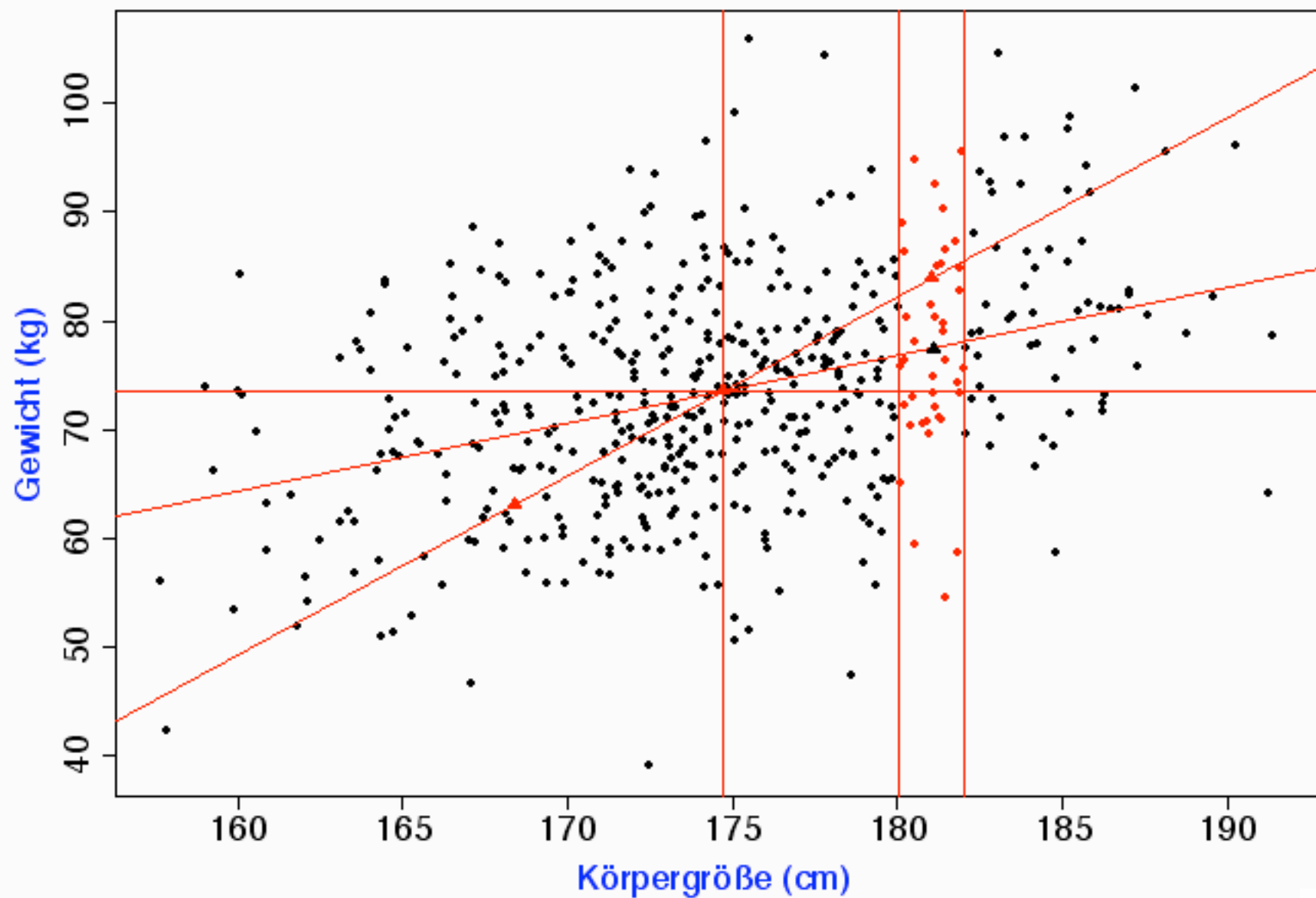
Der Mittelwert auch.

Körpergröße 1 StdAbw oberhalb des Durchschnitts (± 1 cm)

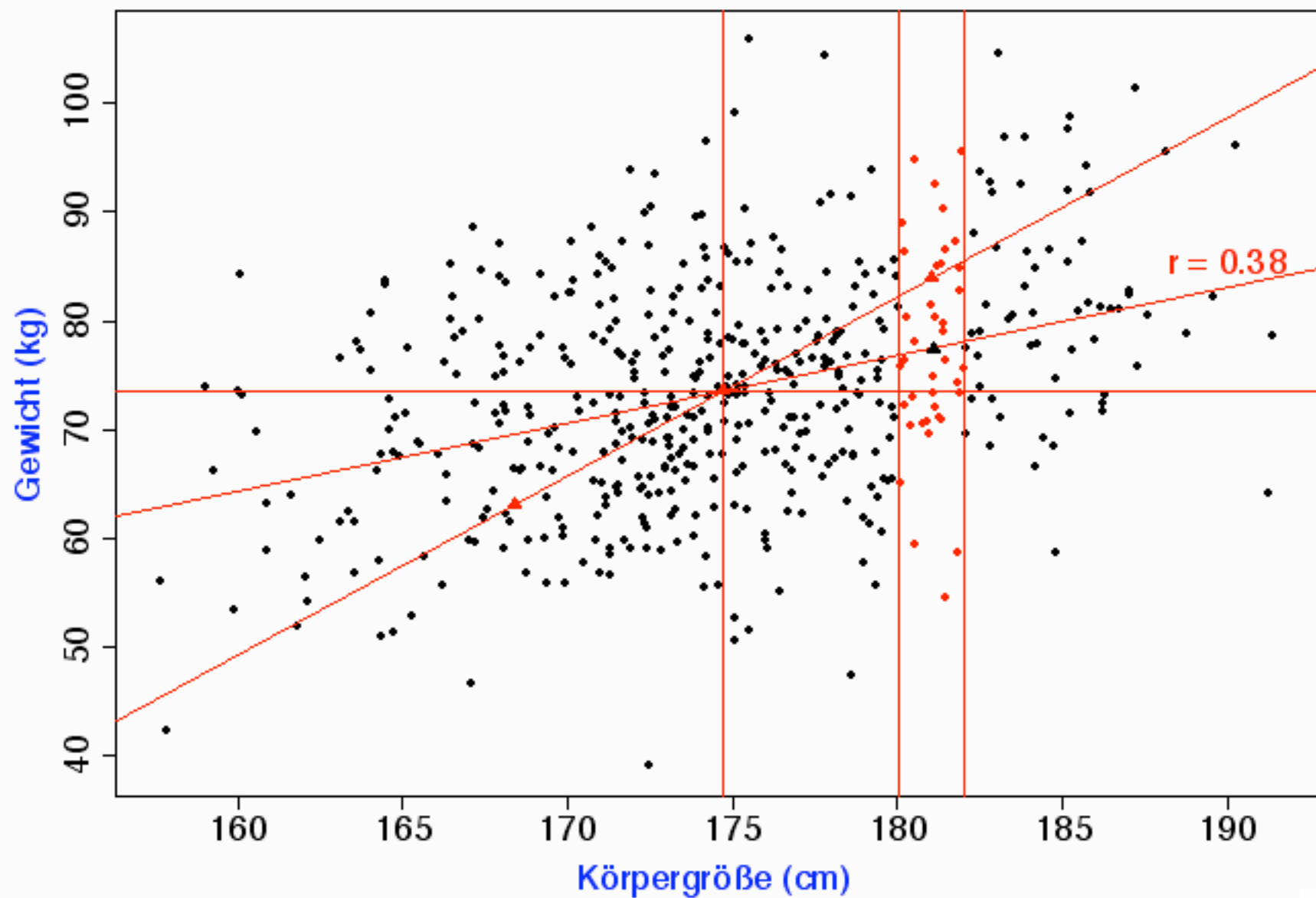


Die Regressionsgerade

Die Regressionsgerade



Die Regressionsgerade



$$r = 0,38$$

INTERPRETATION

Wenn x um s_x größer wird,

wird y im Schnitt

um $0,38 s_y$ größer.

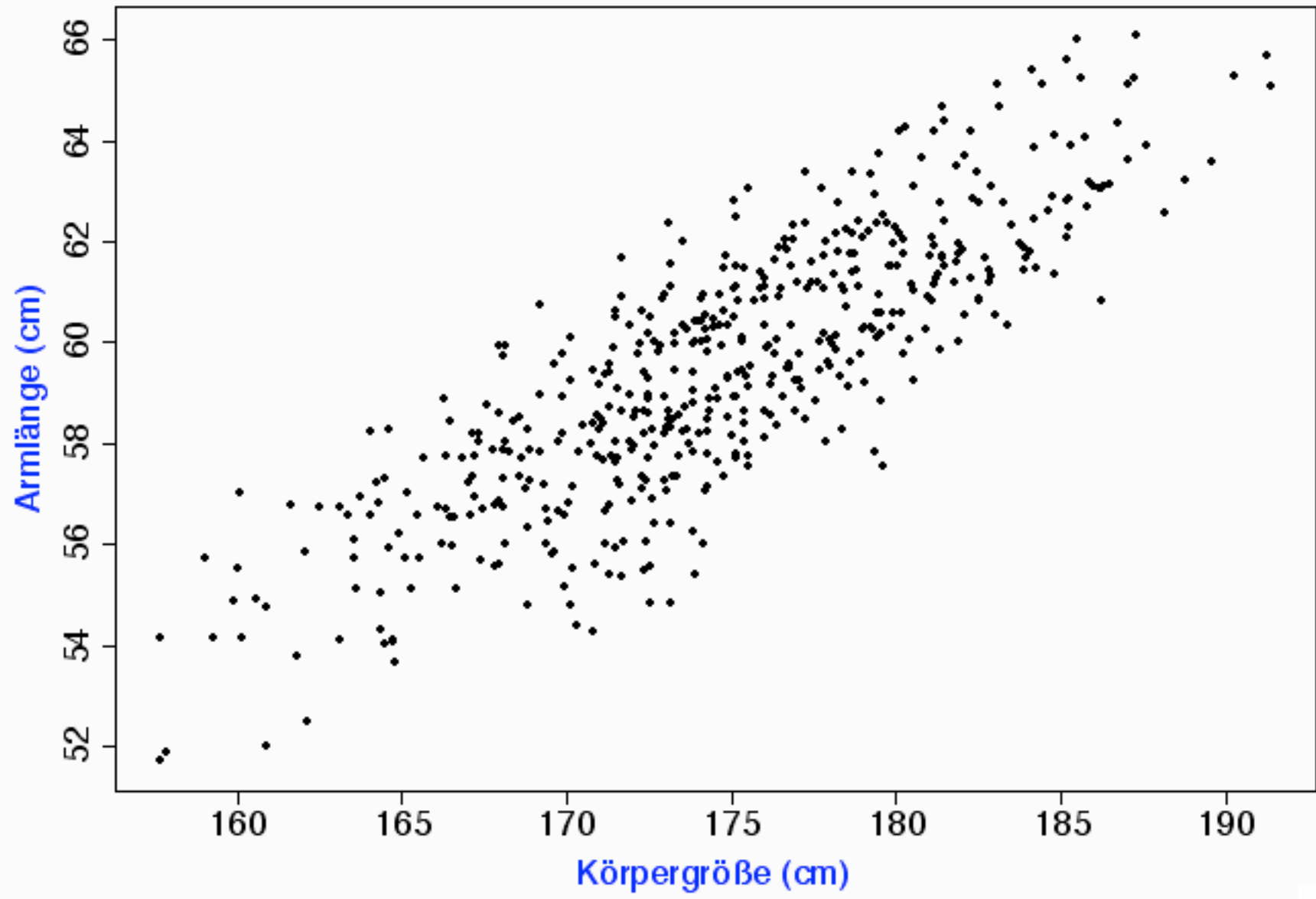
Zum Vergleich:

Körpergröße

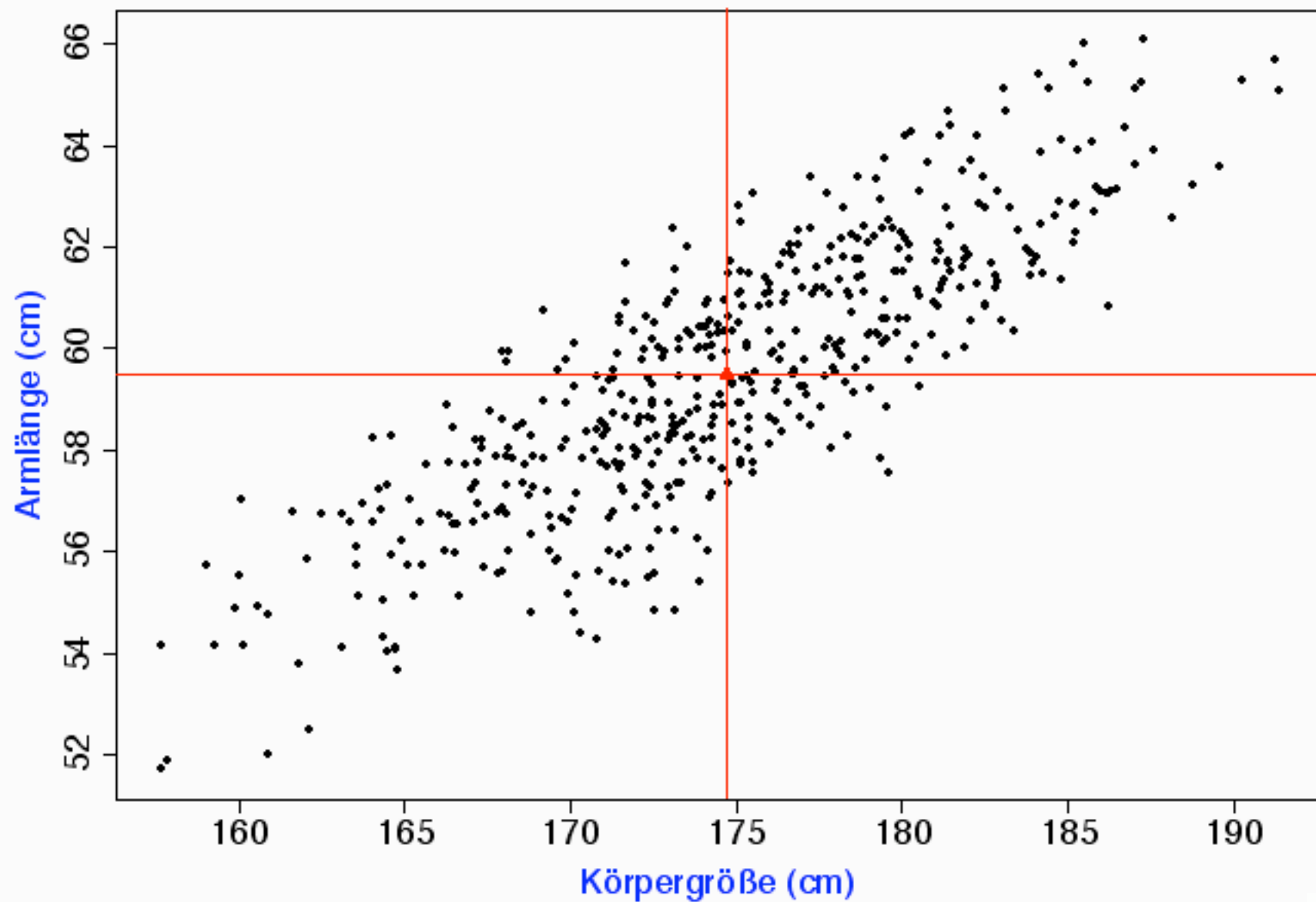
und

Armlänge

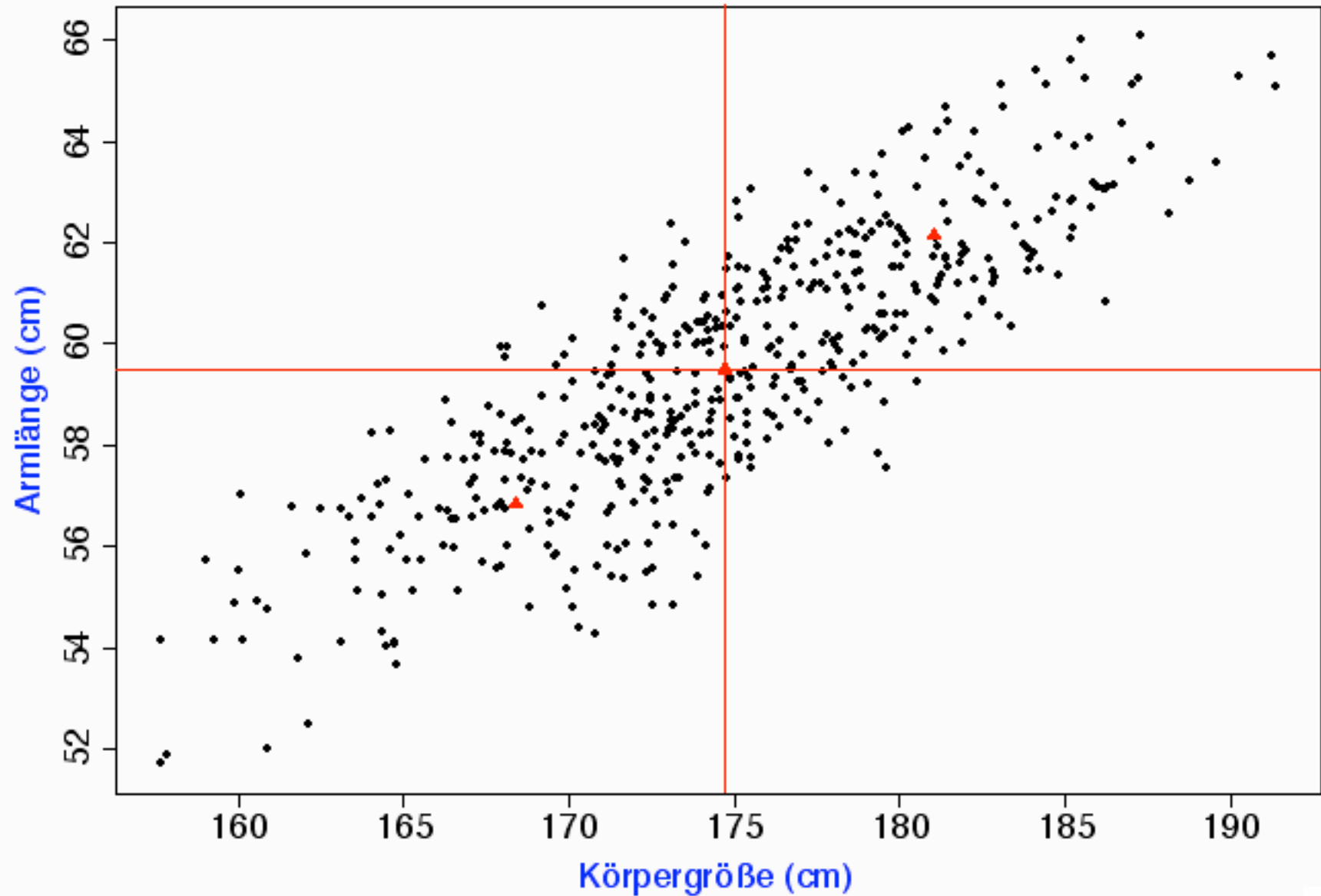
Männliche US-Amerikaner 18–24 J. (n=500)



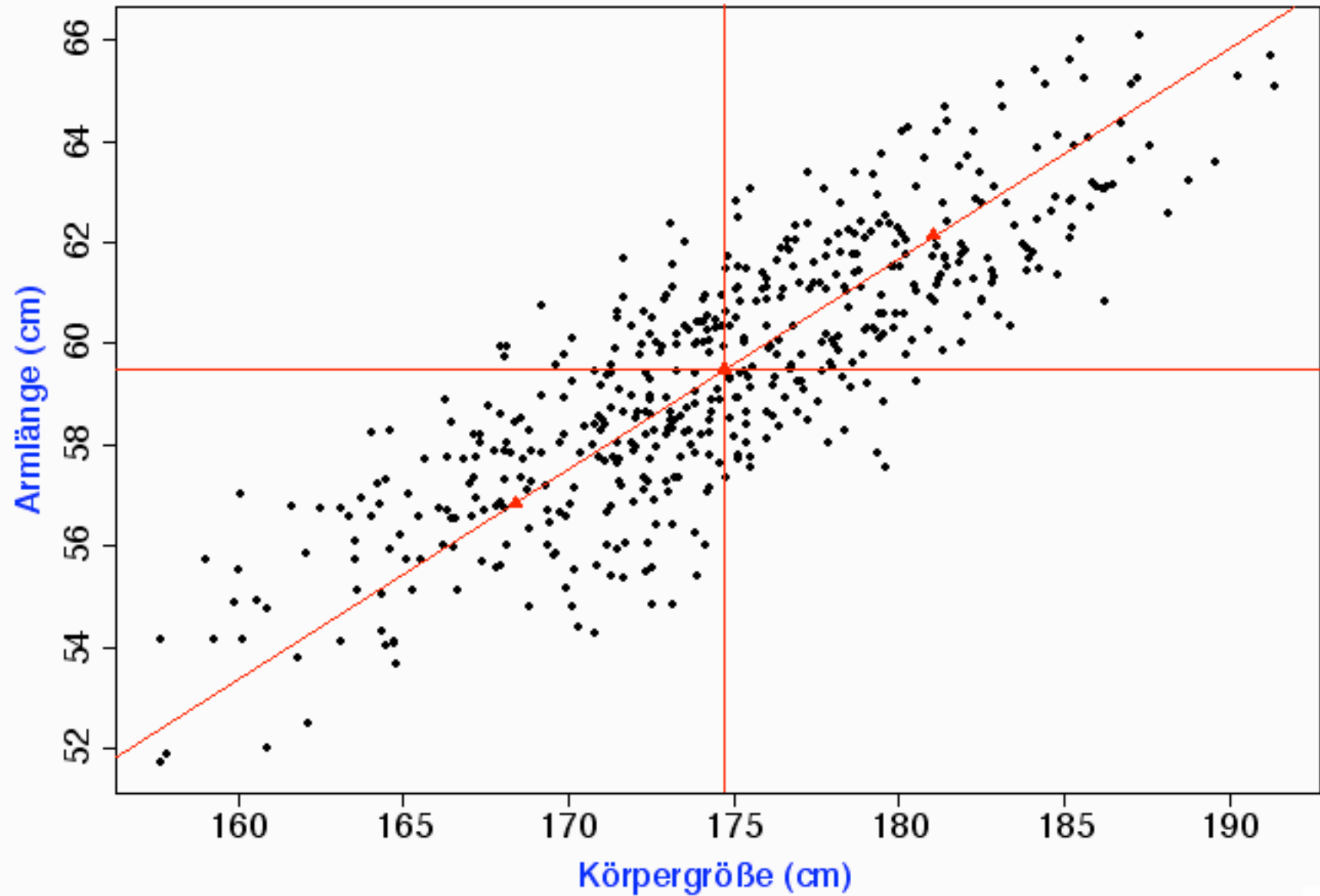
Mittelwerte



Mittelwerte \pm Standardabweichungen

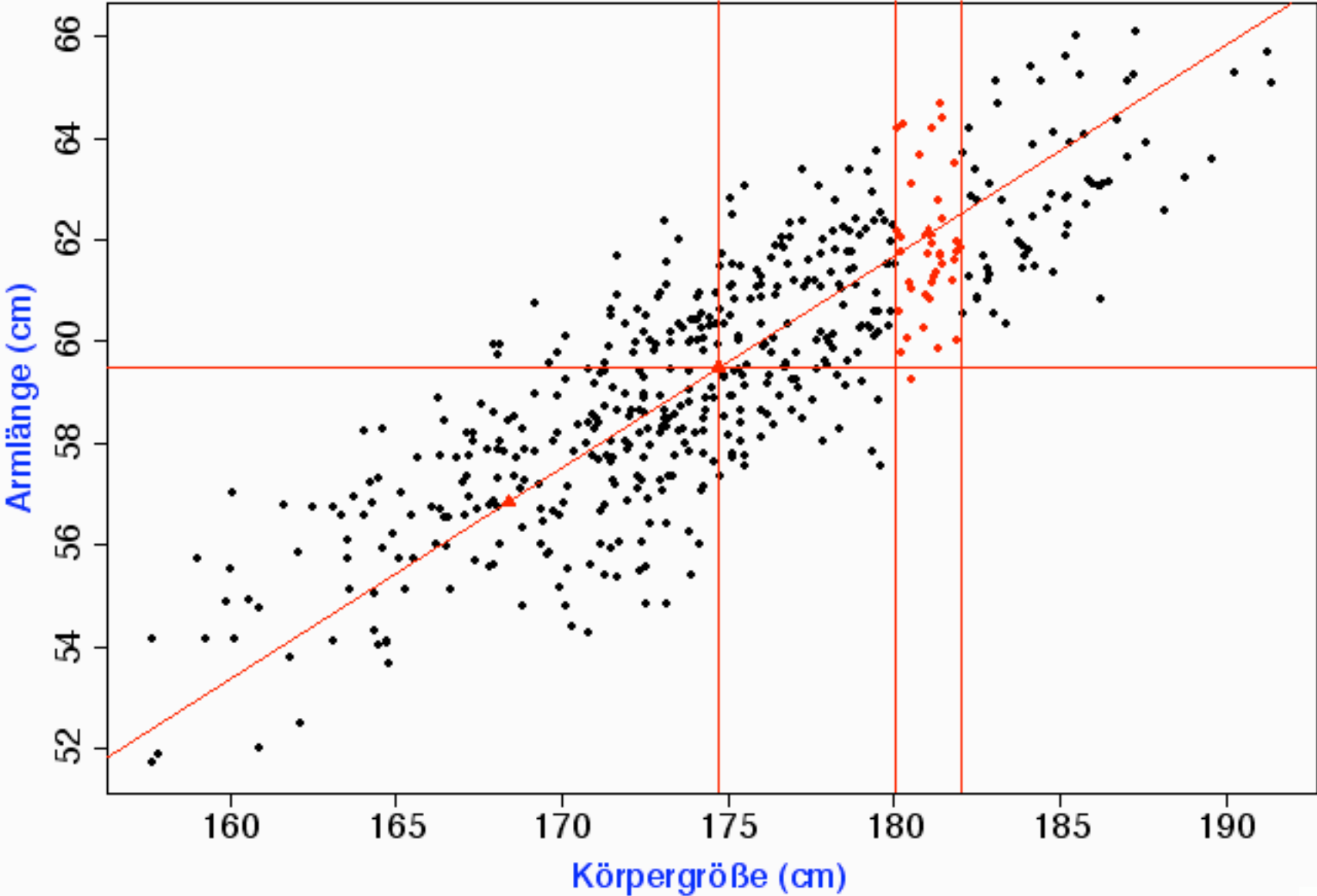


Mittelwerte \pm Standardabweichungen

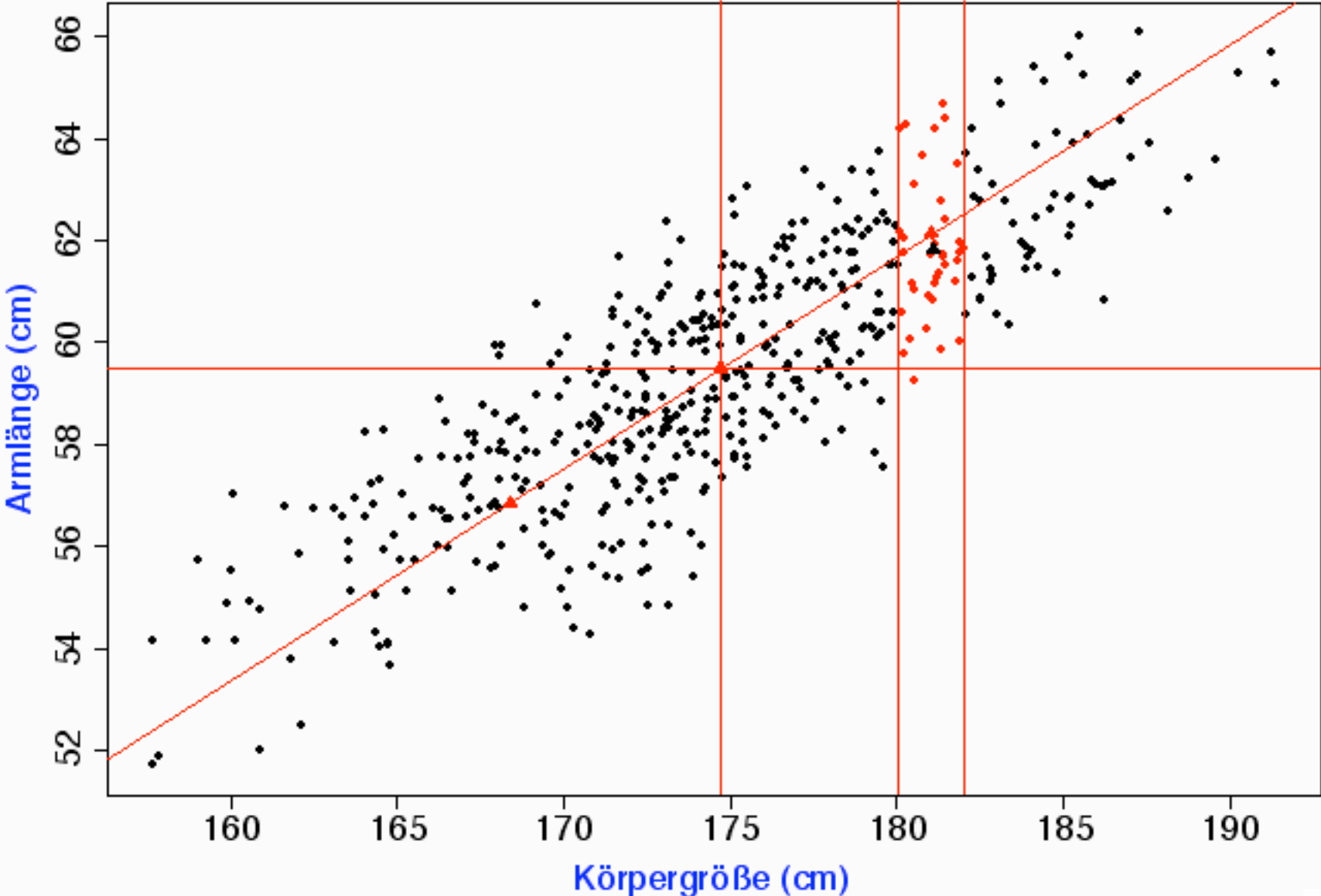


Wie groß im Schnitt
ist die Armlänge
eines Mannes,
dessen Körpergröße
um eine Standardabweichung
größer als der Durchschnitt ist?

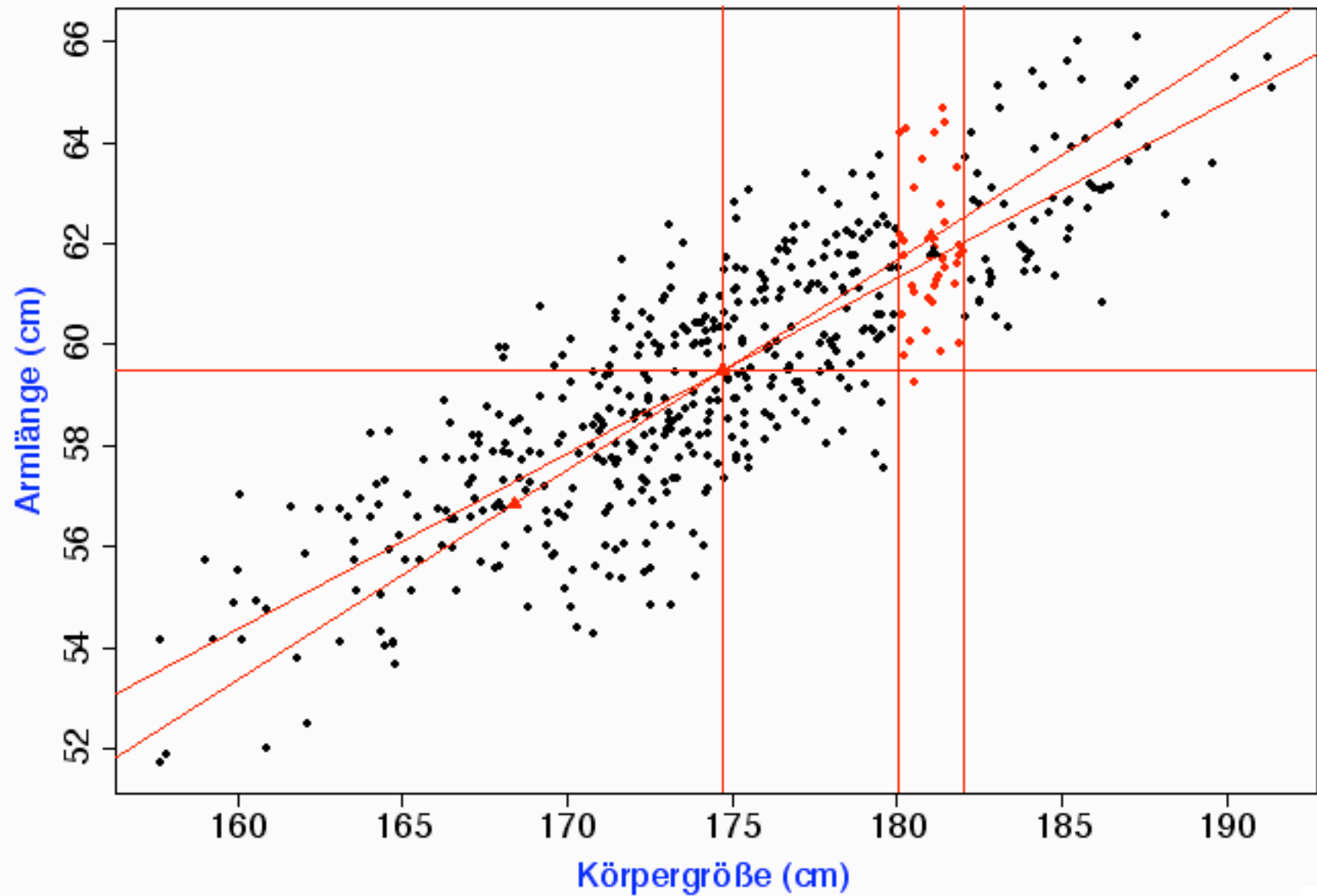
Körpergröße 1 StdAbw oberhalb des Durchschnitts (± 1 cm)



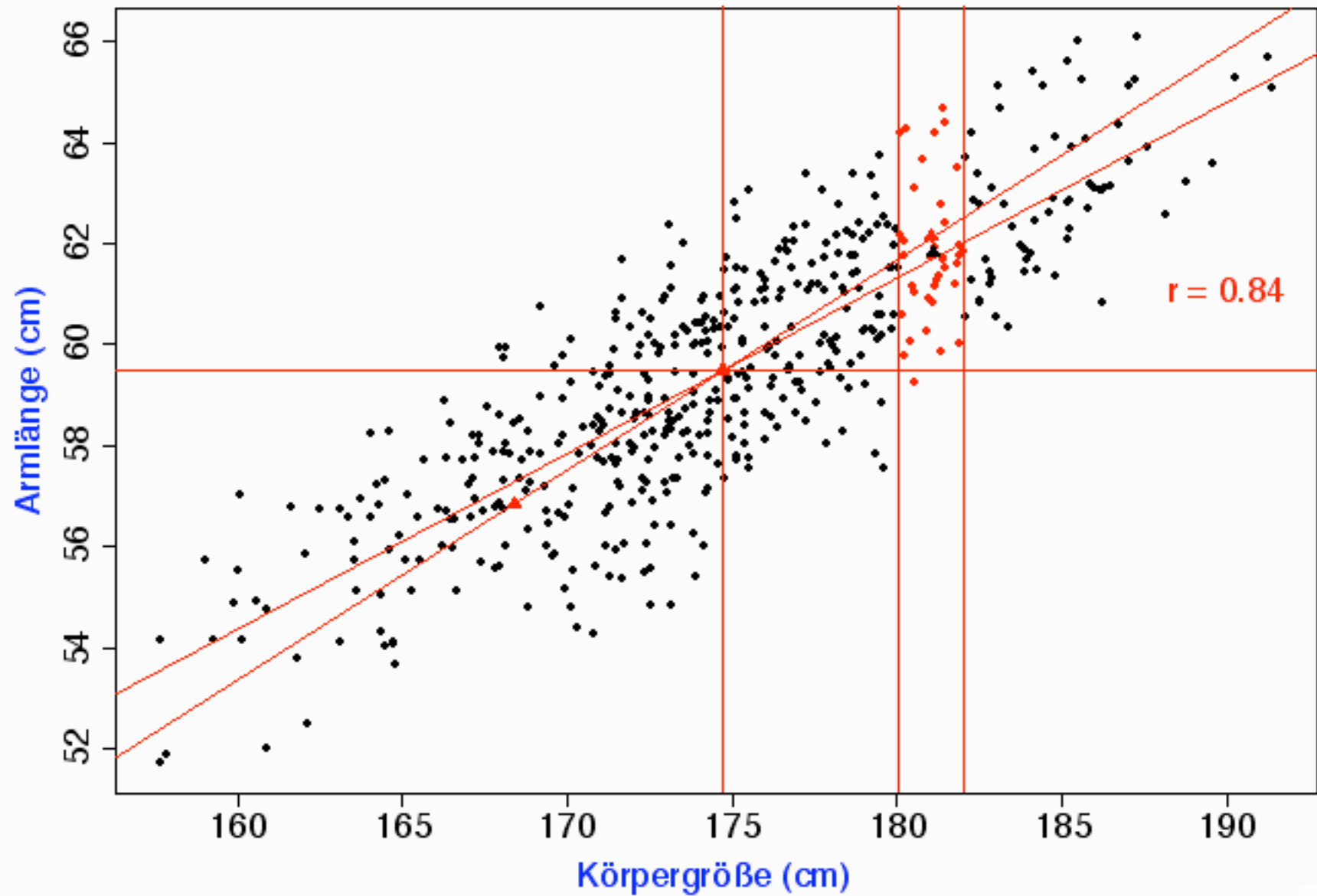
Körpergröße 1 StdAbw oberhalb des Durchschnitts (± 1 cm)



Die Regressionsgerade



Die Regressionsgerade



$$r = 0,84$$

Wenn x
um s_x größer wird,
wird y
im Schnitt
um $0,84 s_y$ größer.

Eigenschaften

von

r

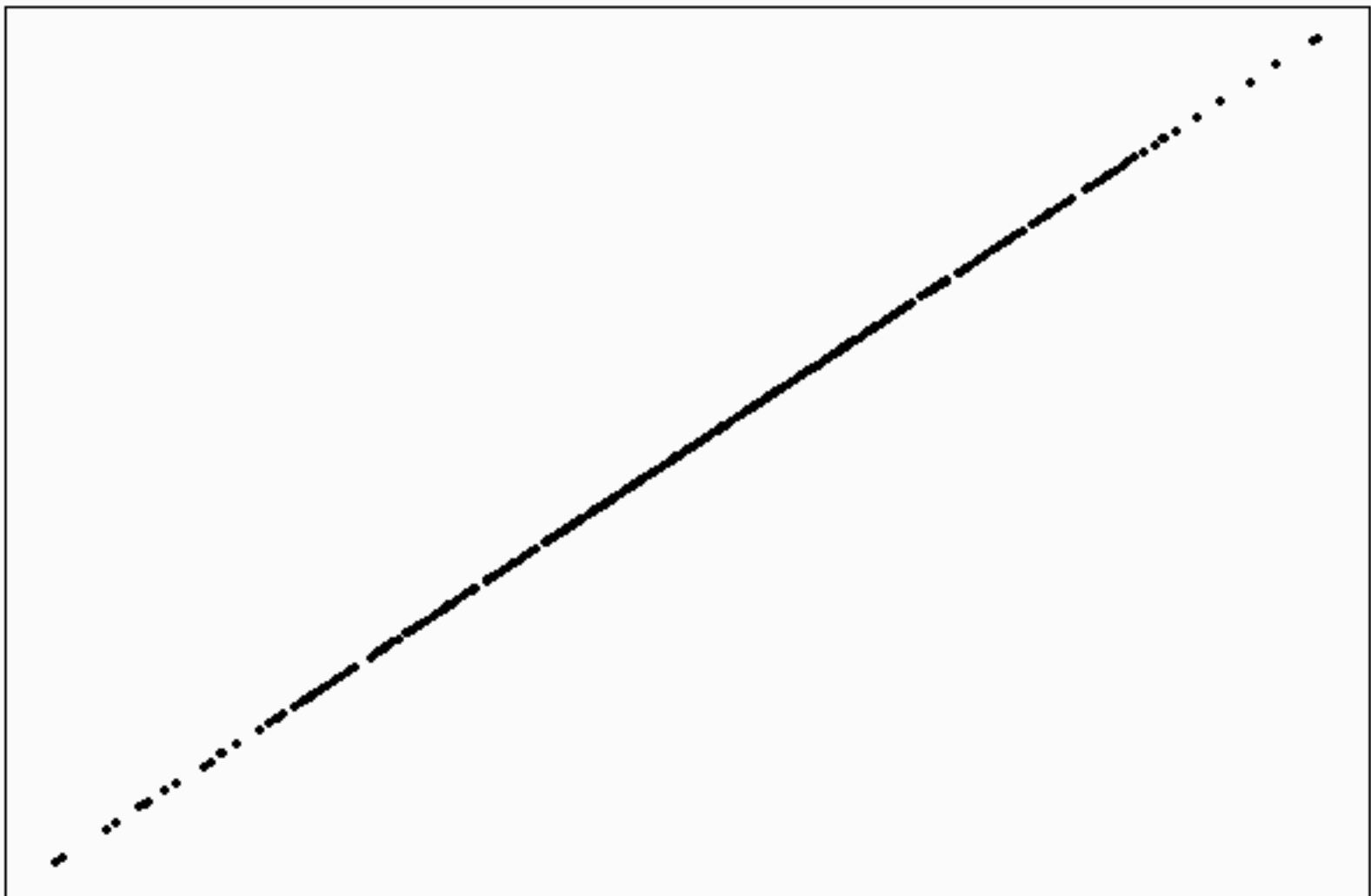
$$-1 \leq r \leq 1$$

$$r = 1:$$

Alle Punkte

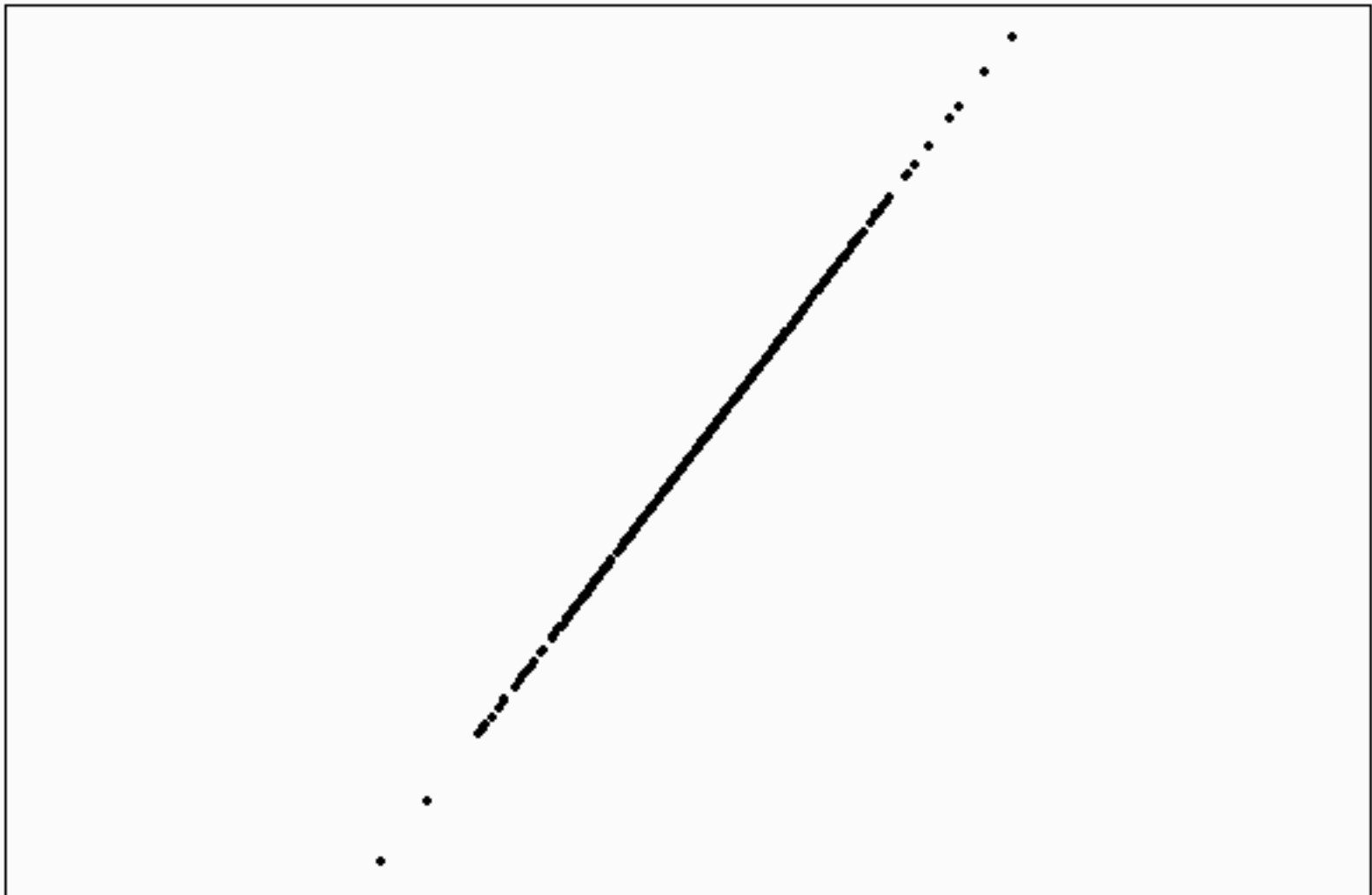
liegen auf einer steigenden Gerade.

$r = 1$

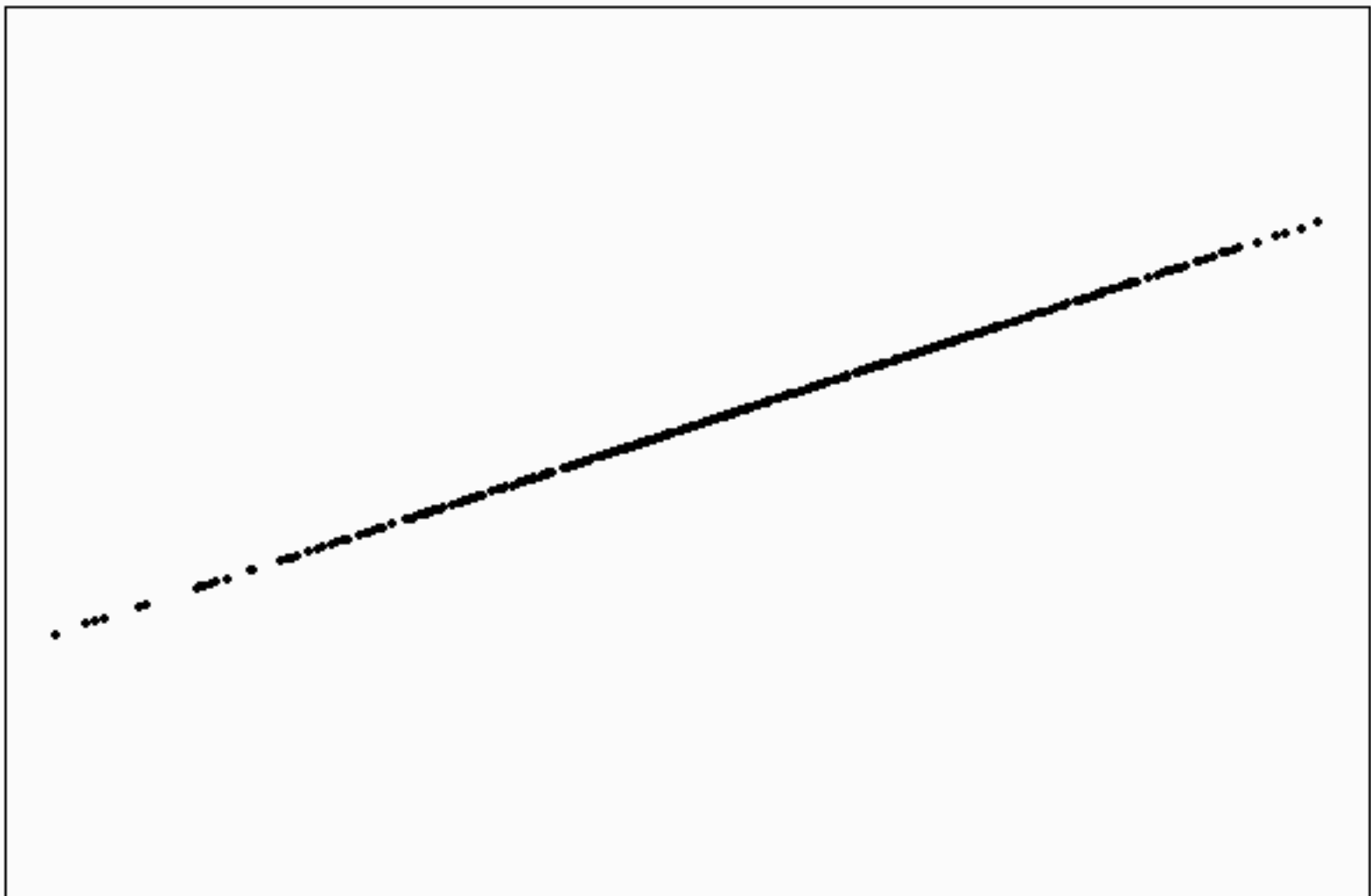


Die Steigung der Gerade,
solange sie positiv ist,
ist egal.

$r = 1$



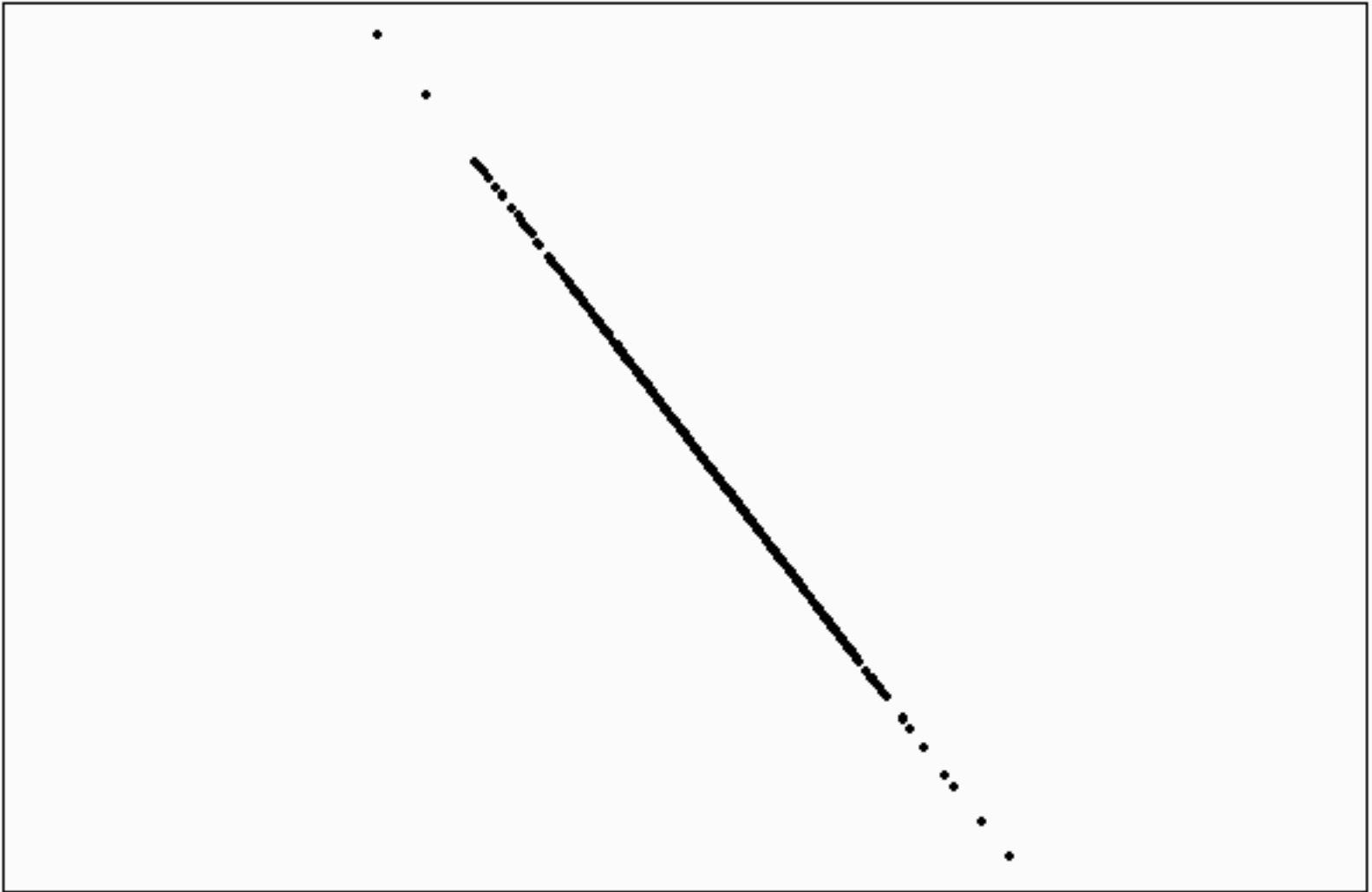
$r = 1$



$$r = -1:$$

Alle Punkte
liegen auf einer
fallenden
Gerade.

$$r = -1$$



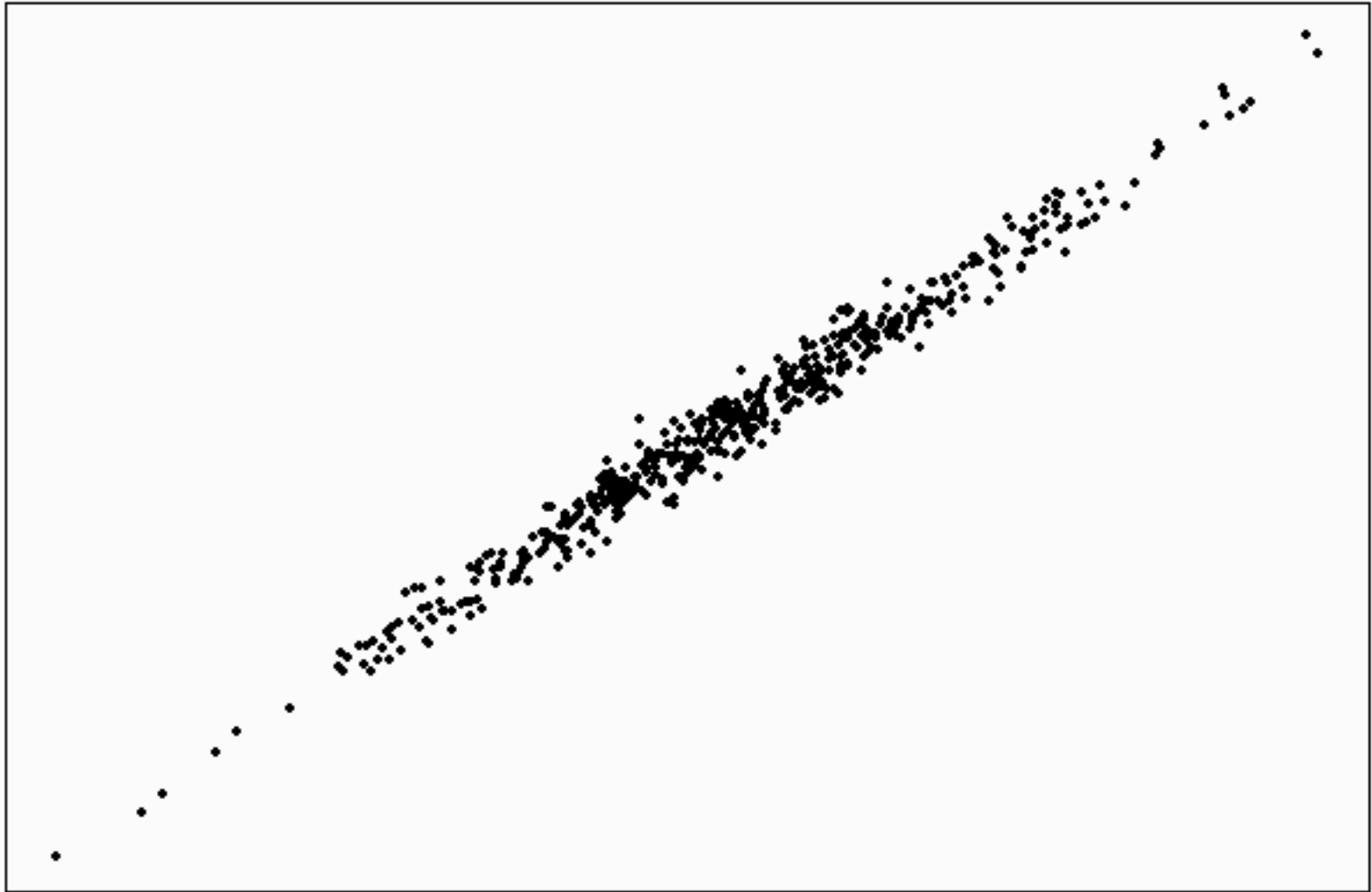
$$r \approx 1$$

Die Punktwolke

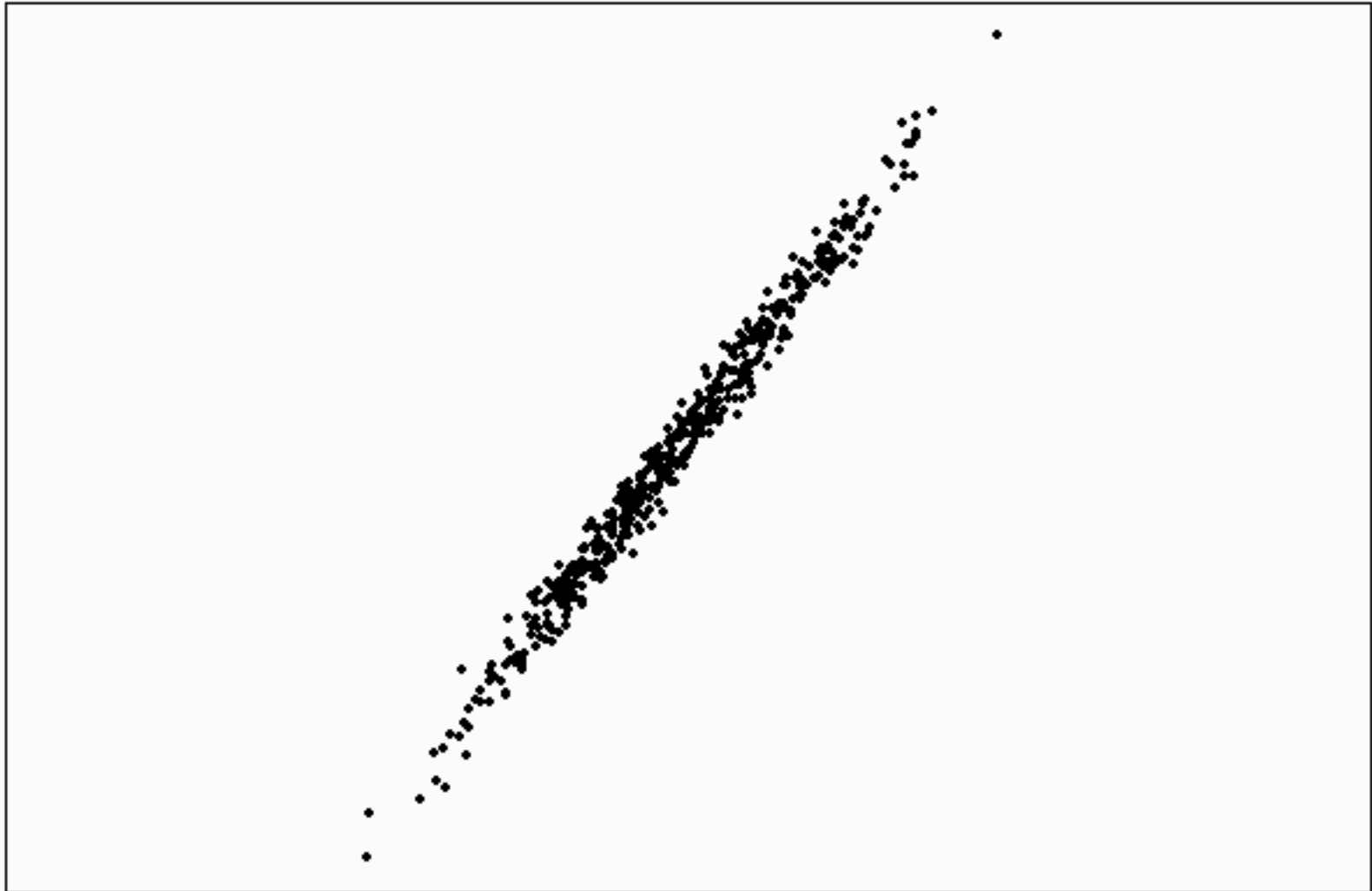
liegt eng

um eine steigende Gerade.

$r = 0.99$

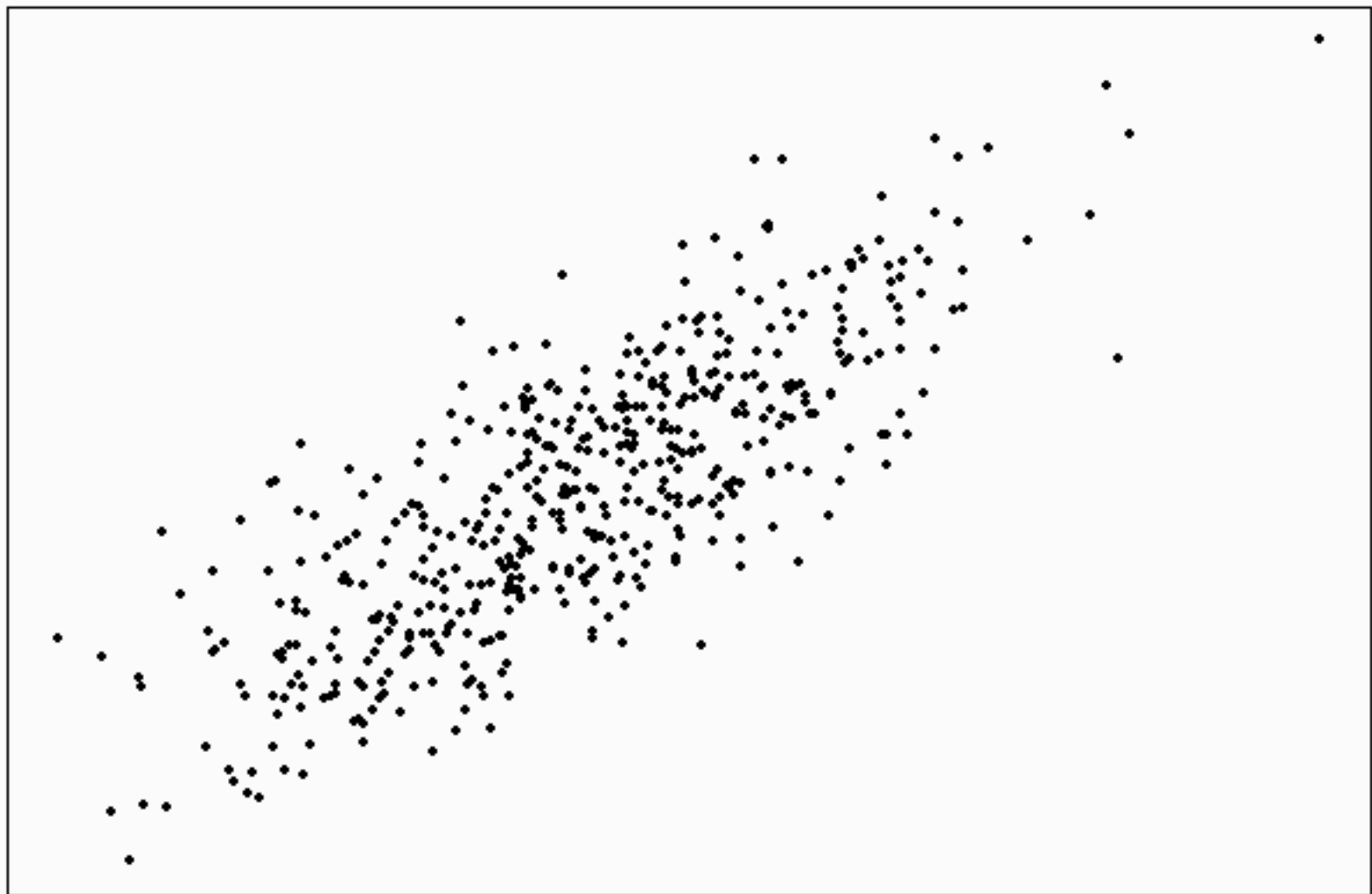


$r = 0.99$

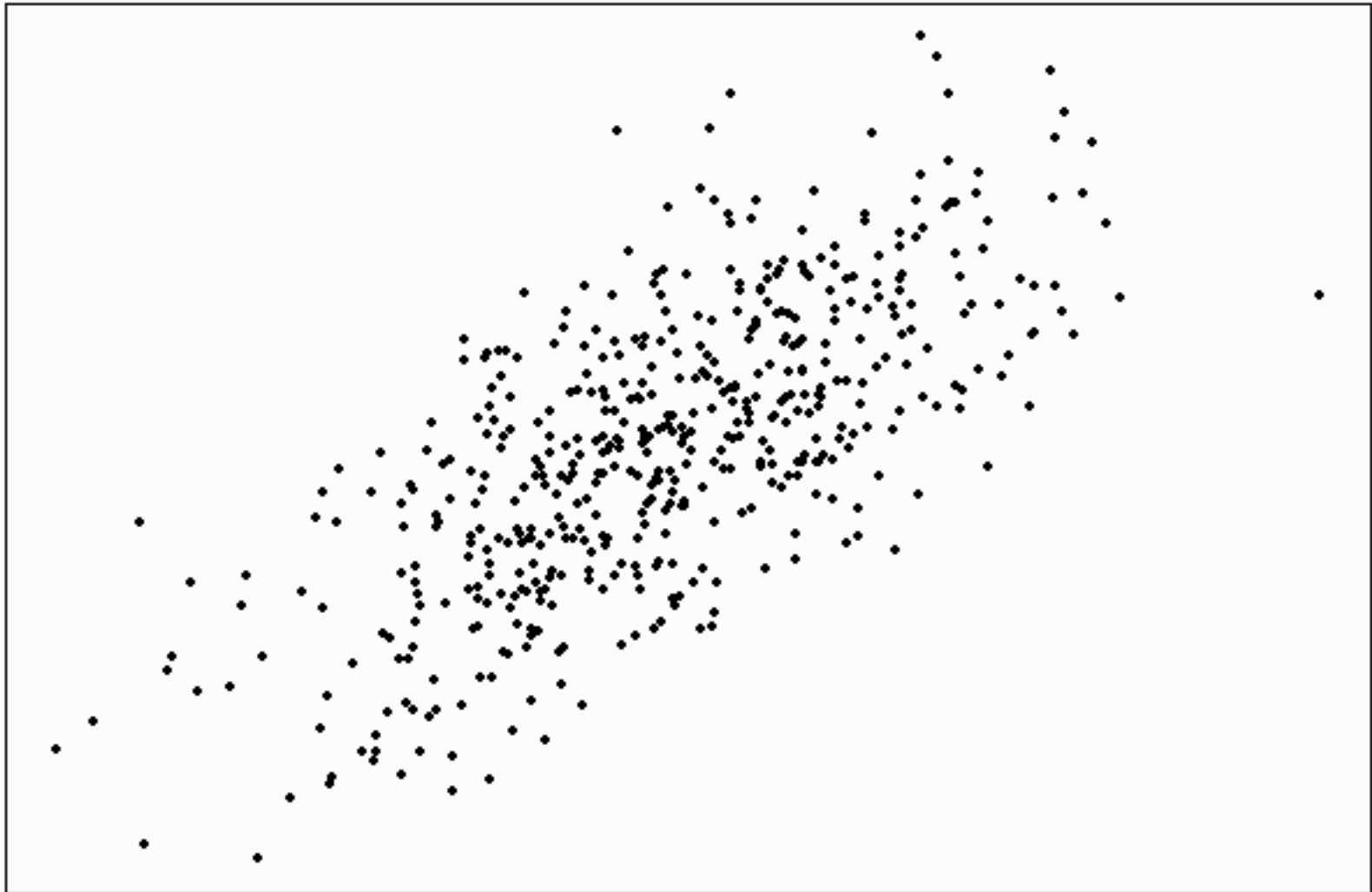


Wenn r kleiner wird,
streut die Punktwolke
immer mehr
um die Regressionsgerade.

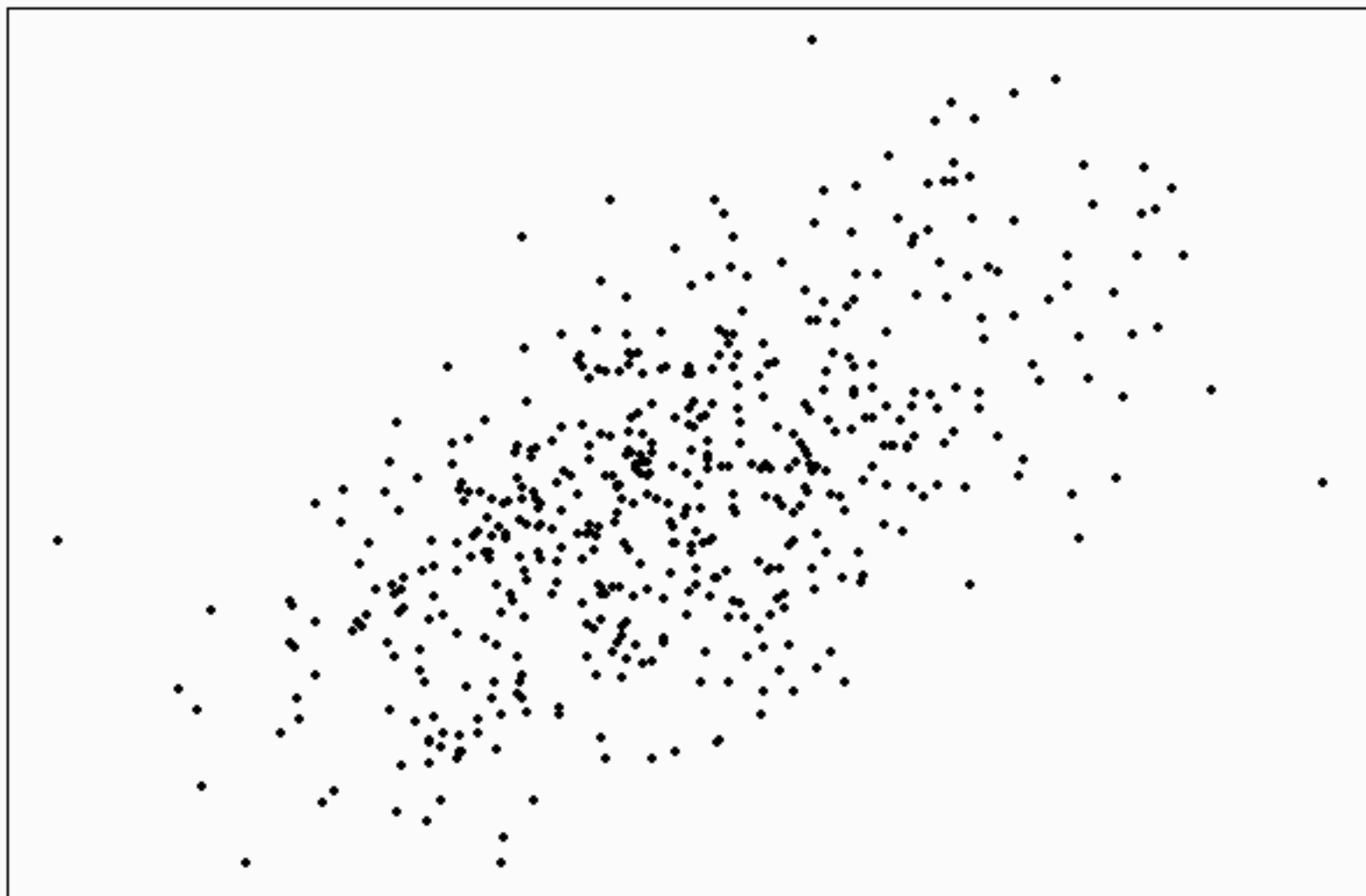
$r = 0.8$



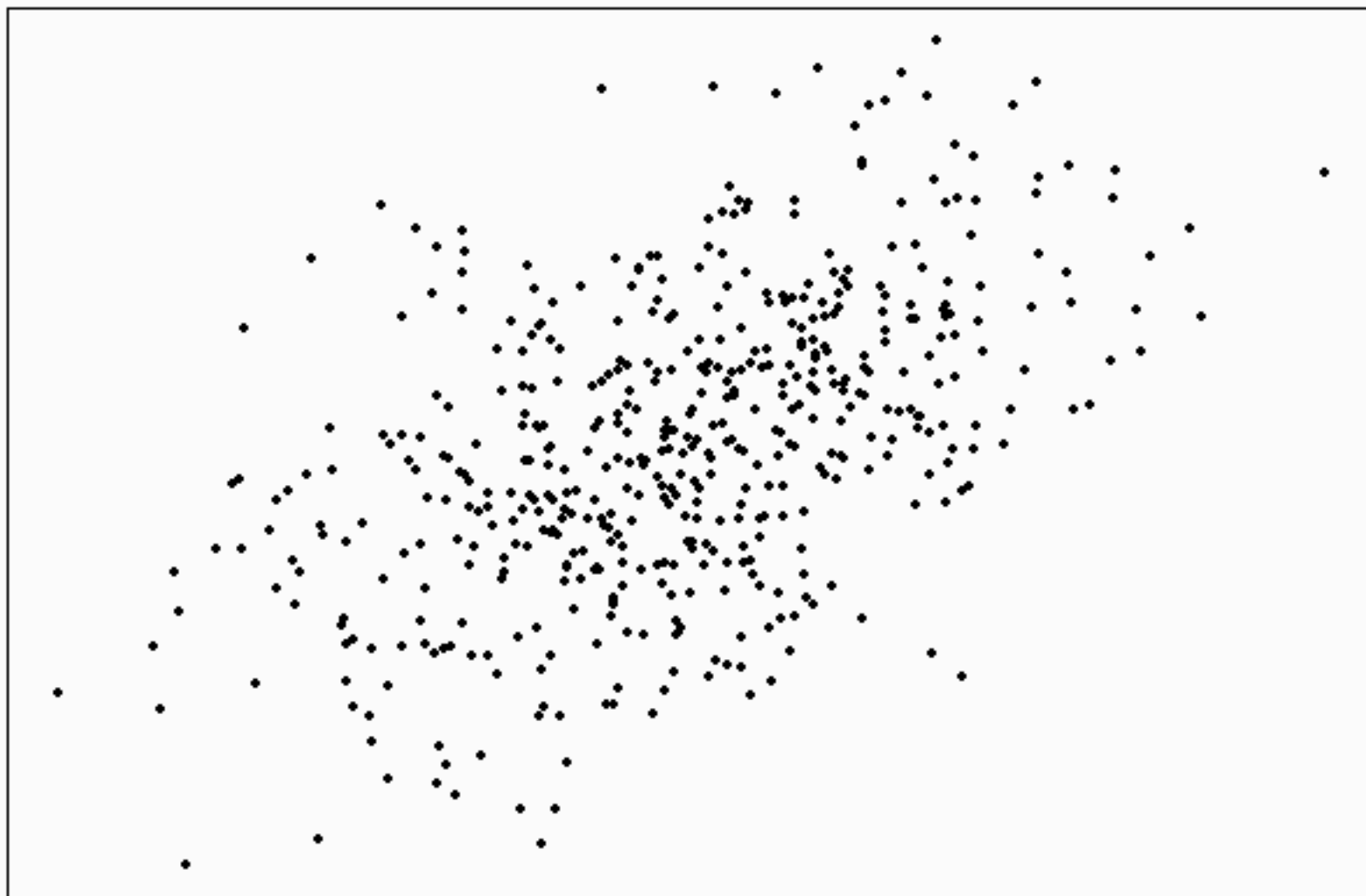
$r = 0.7$



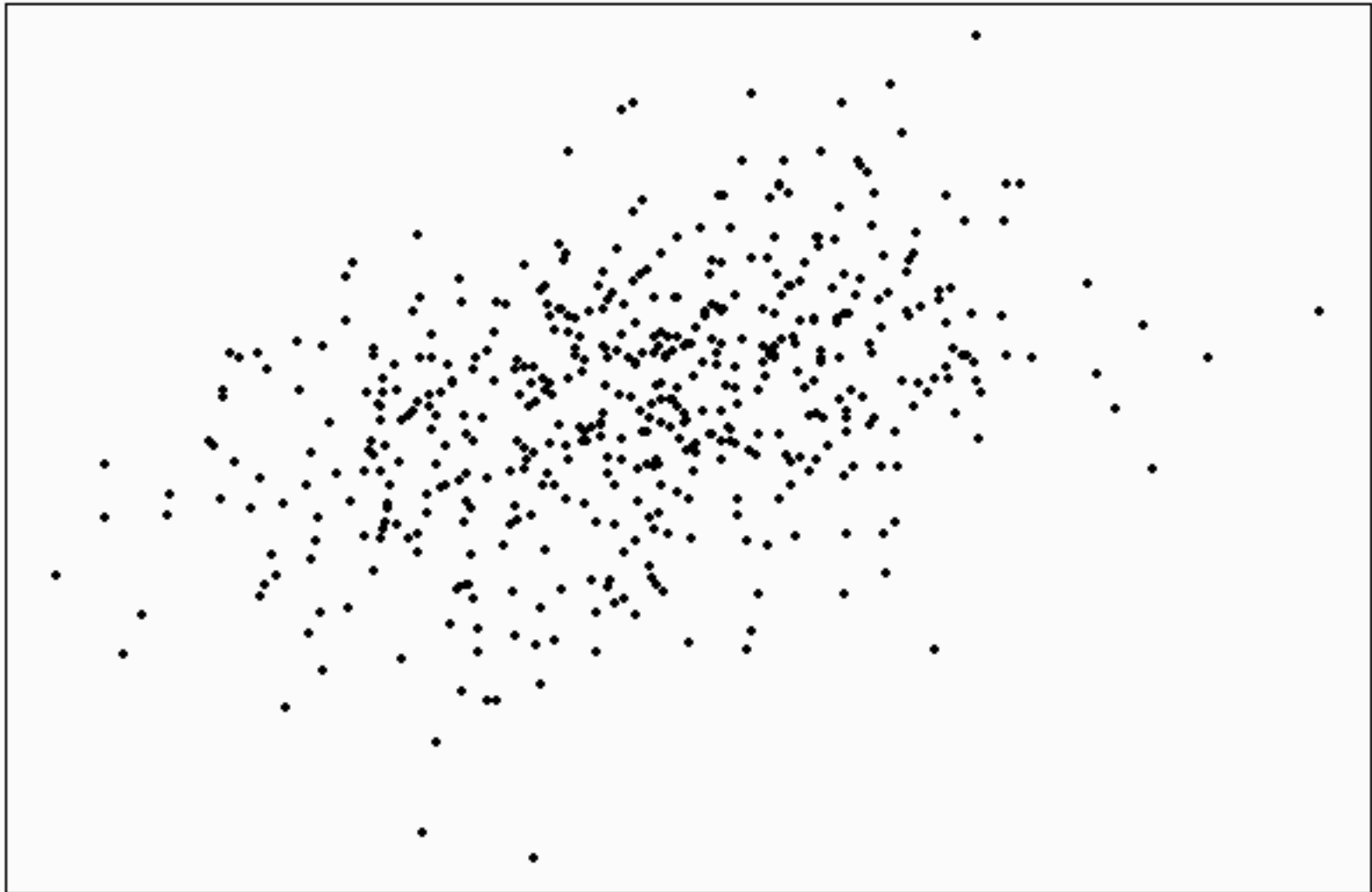
$r = 0.6$



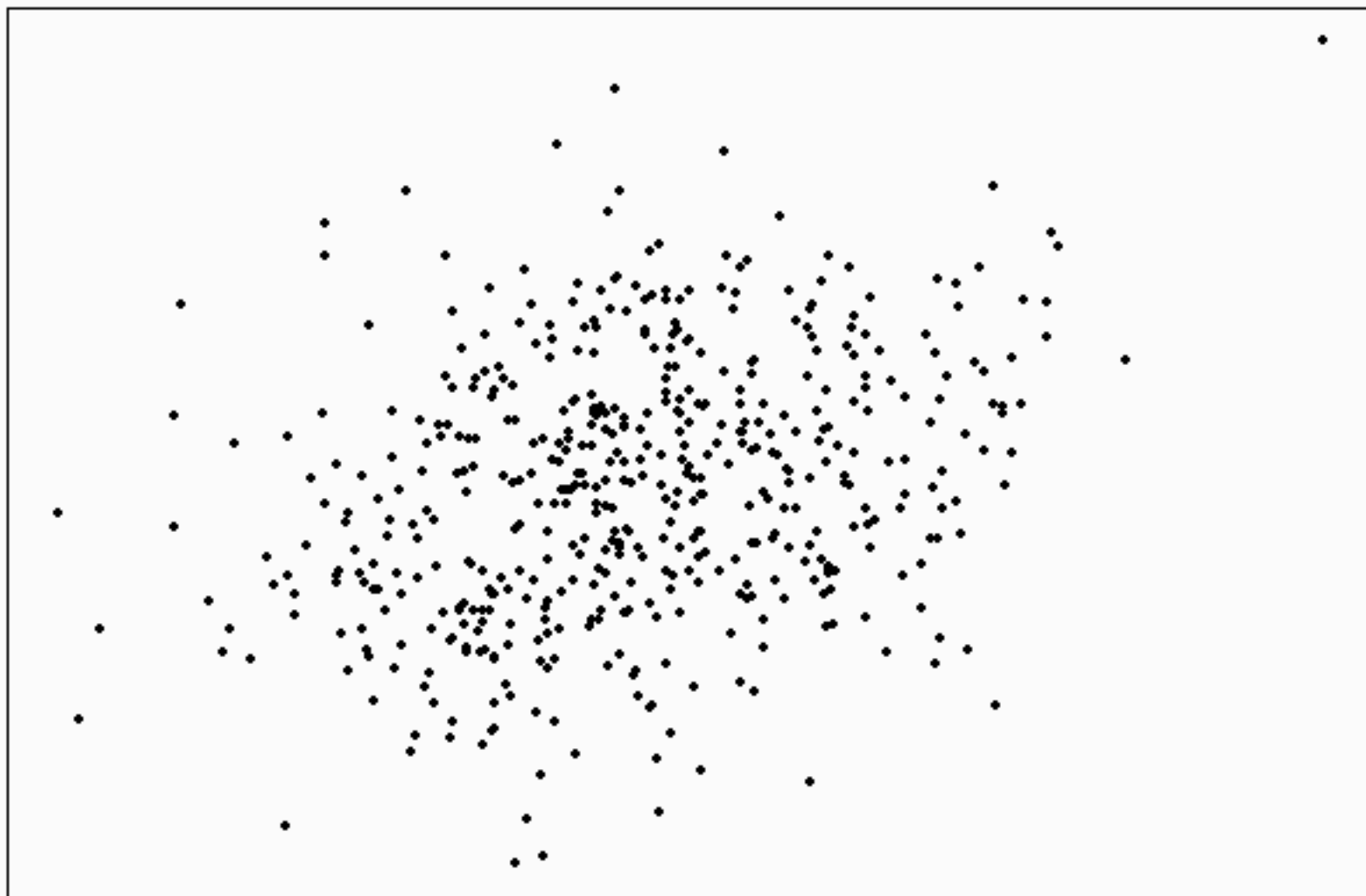
$r = 0.5$



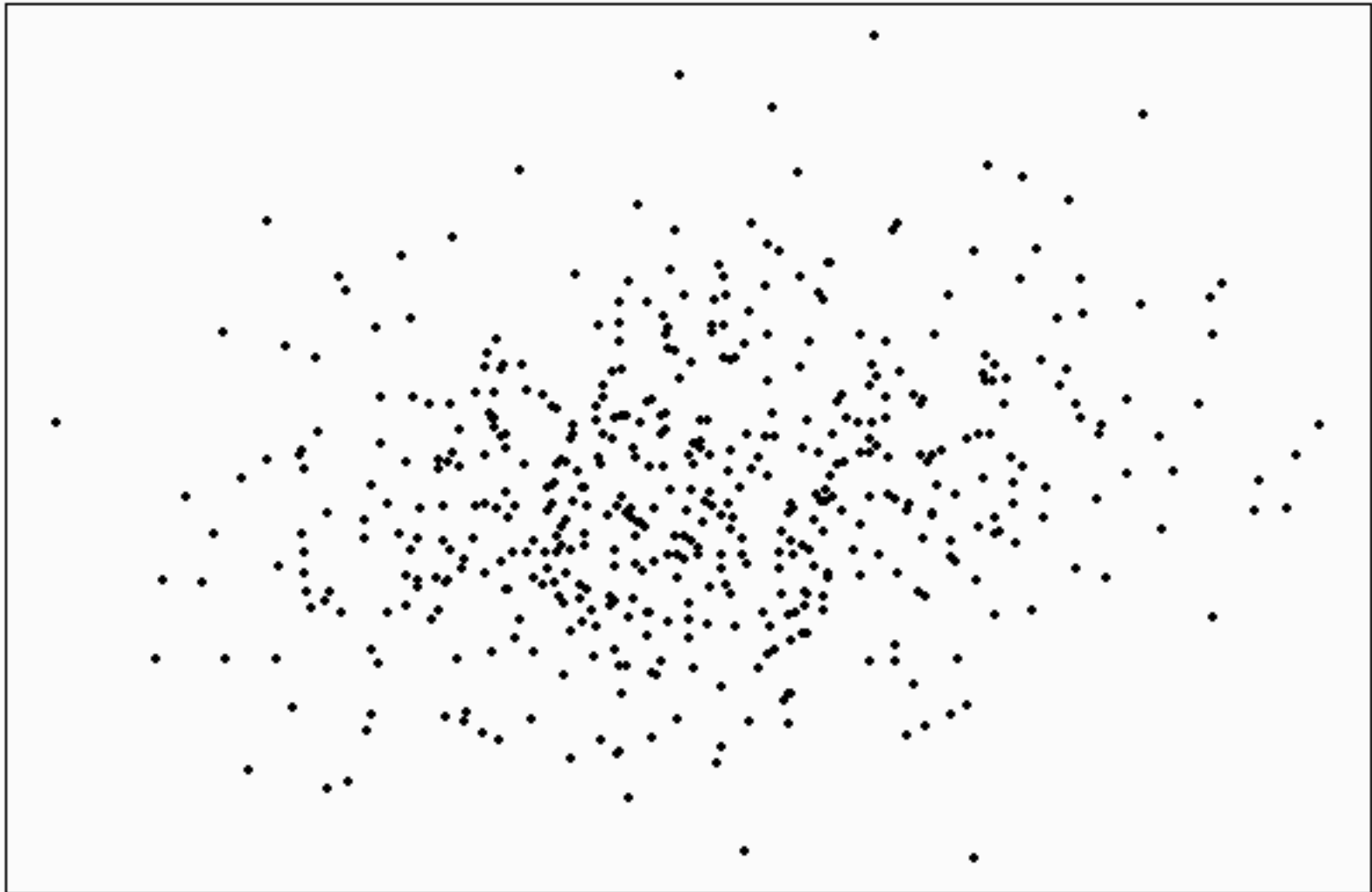
$r = 0.4$



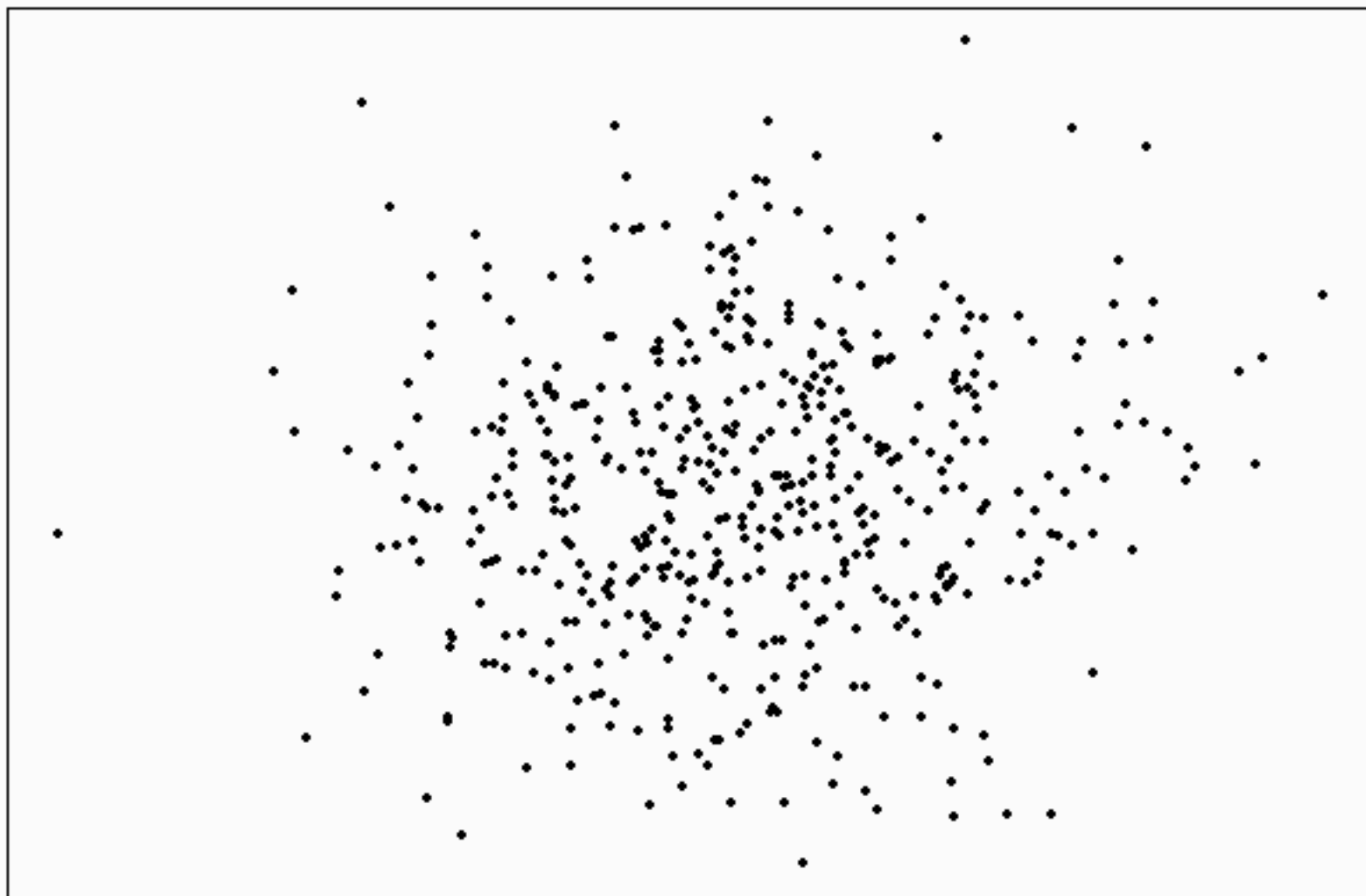
$r = 0.3$



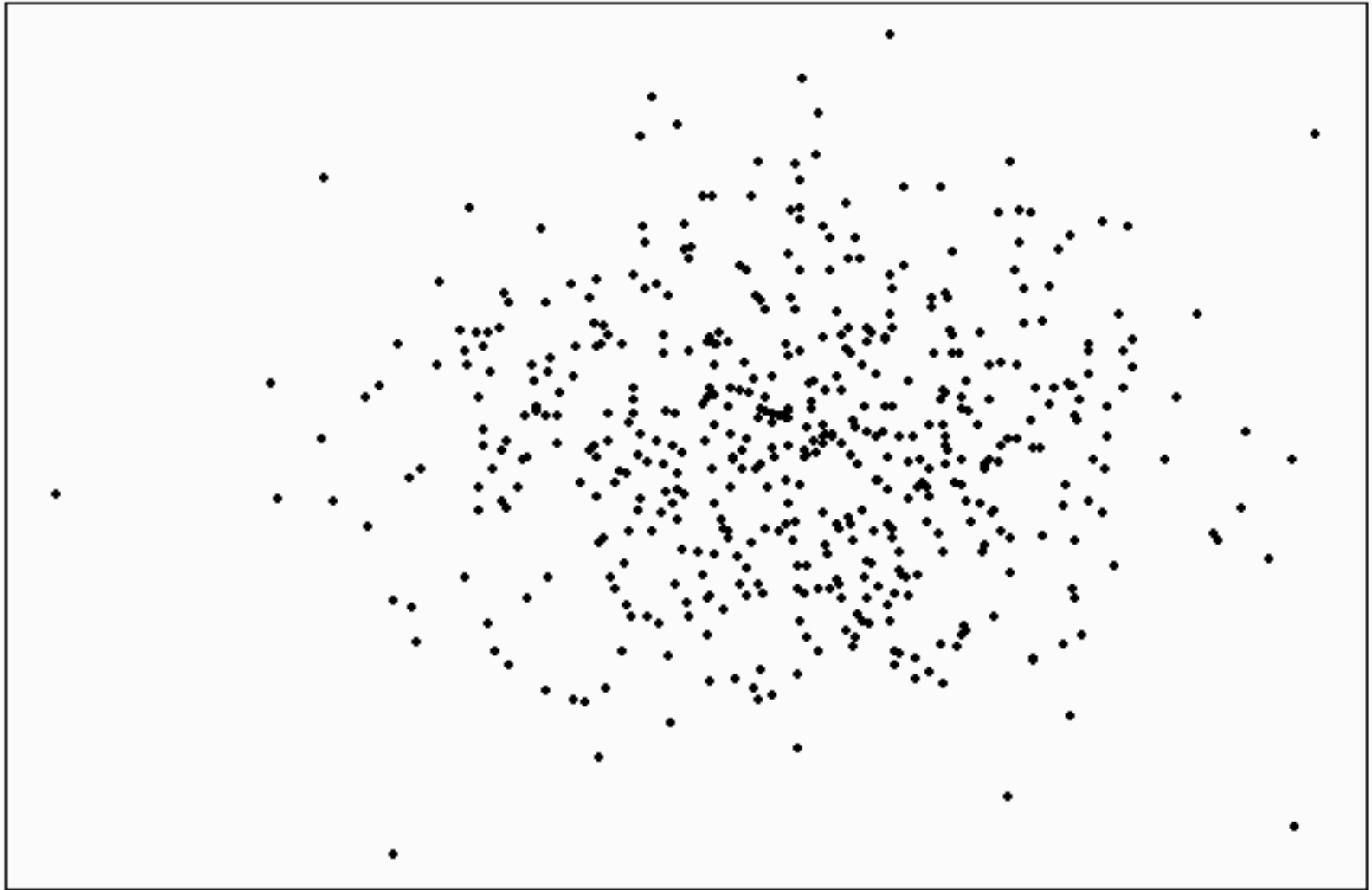
$r = 0.2$



$r = 0.1$



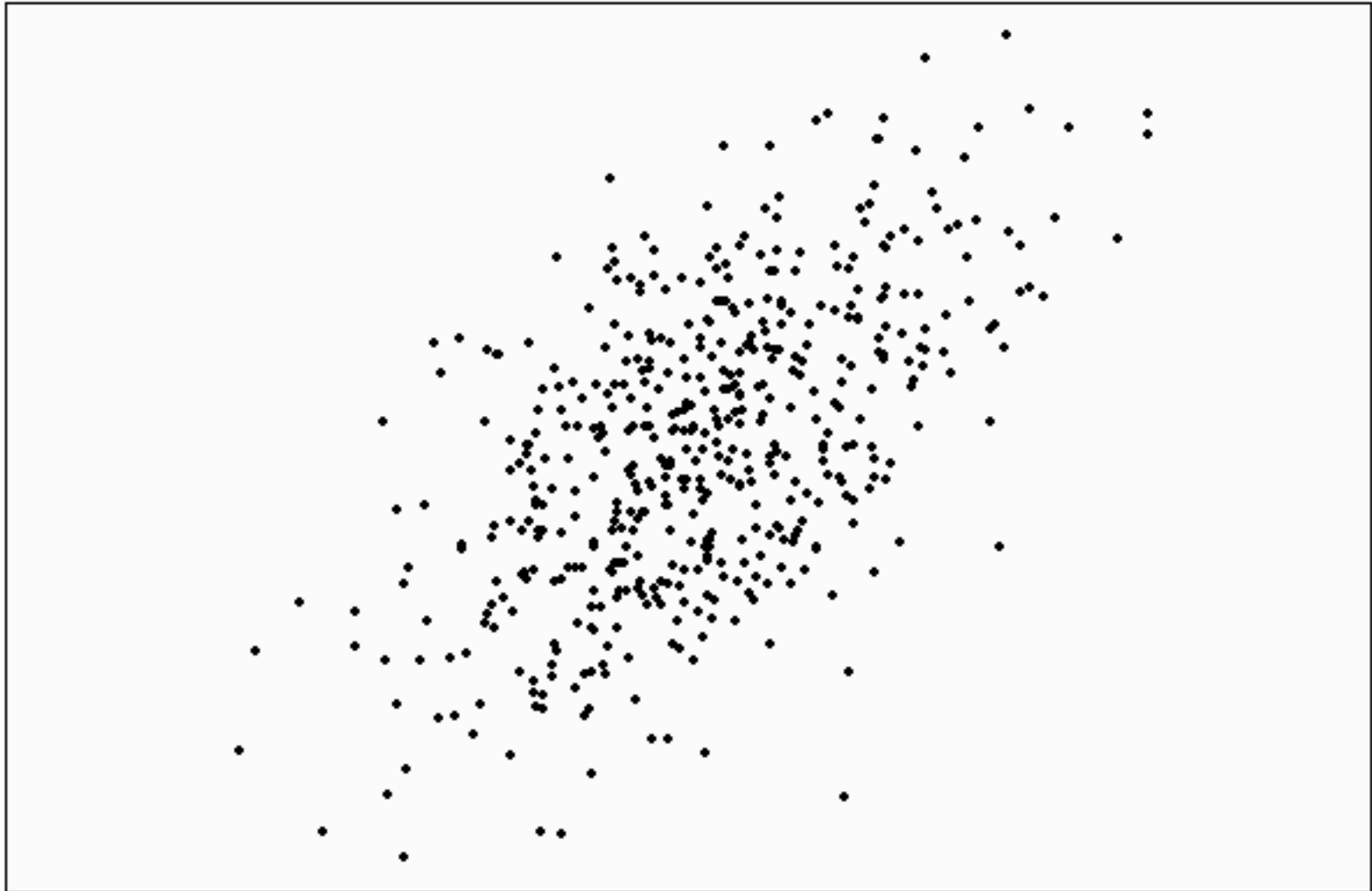
$r = 0$



$$r > 0:$$

Wenn x größer wird,
wird y größer.

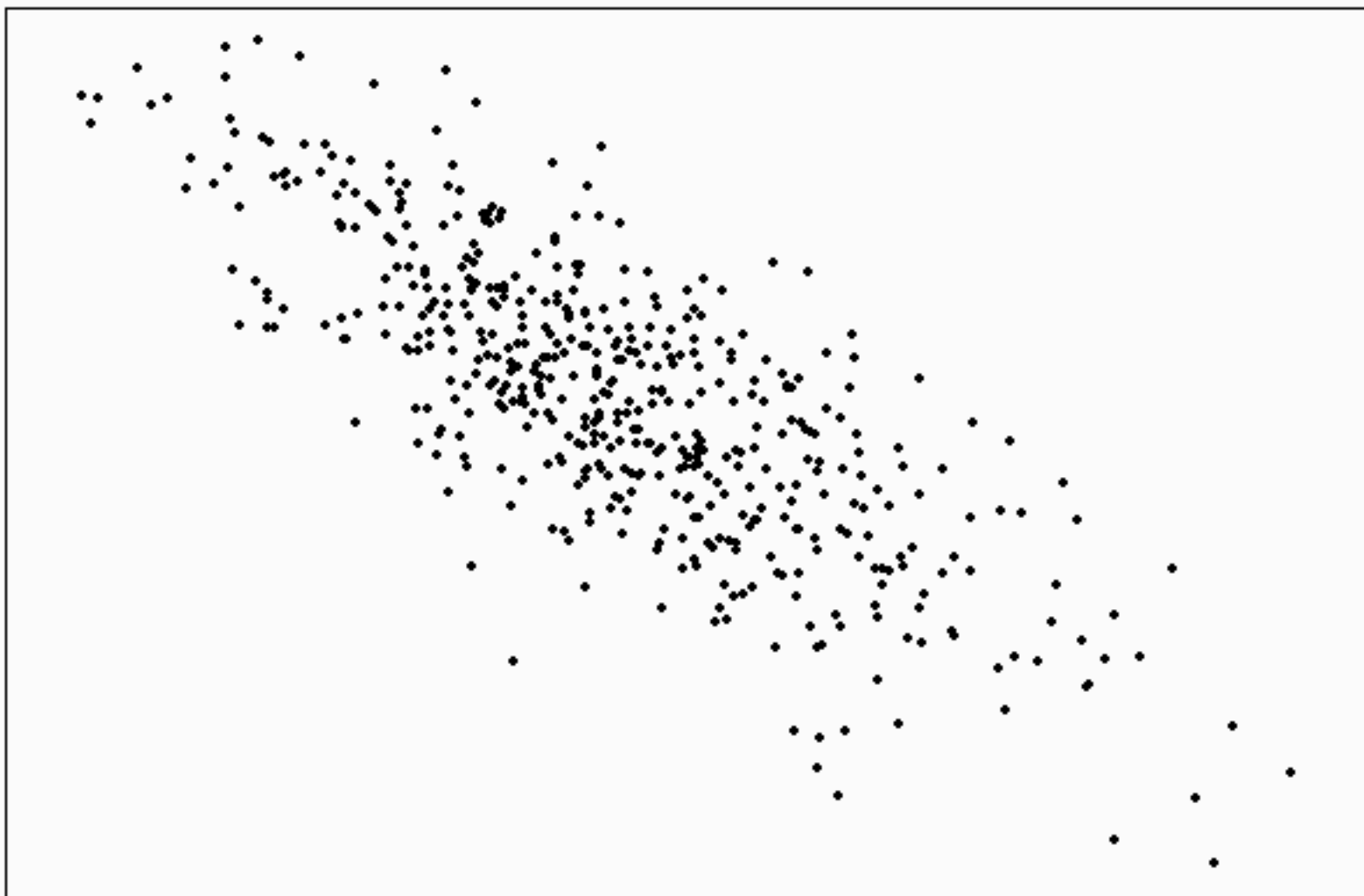
$r = 0.58$



$$r < 0$$

Wenn x größer wird,
wird y kleiner.

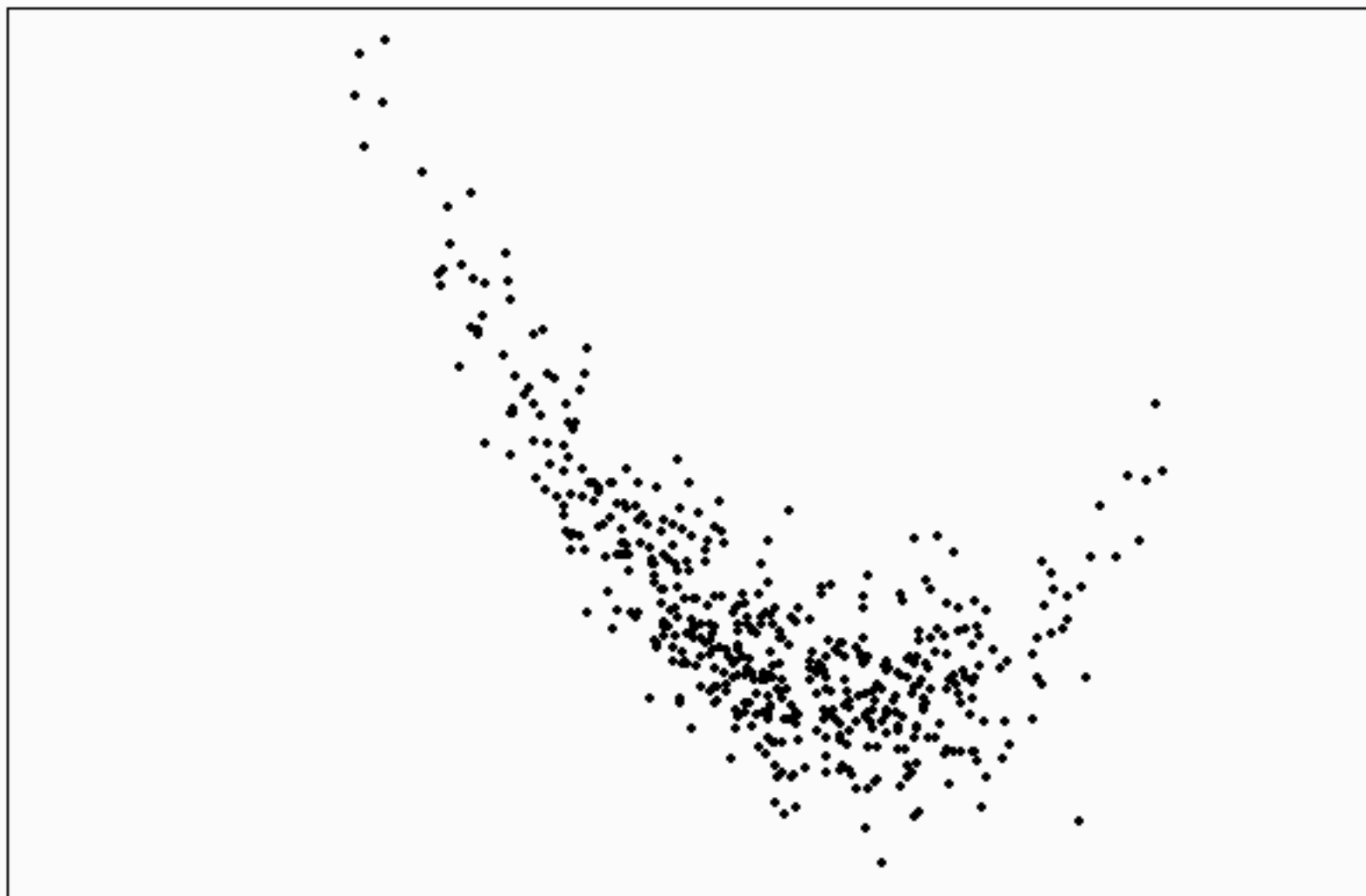
$r = -0.78$



r misst die Stärke
des **linearen** Zusammenhangs
zwischen X und Y .

Wenn der Zusammenhang
nicht ungefähr linear ist,
ist r irreführend.

$r = -0.67$



$$r < 0$$

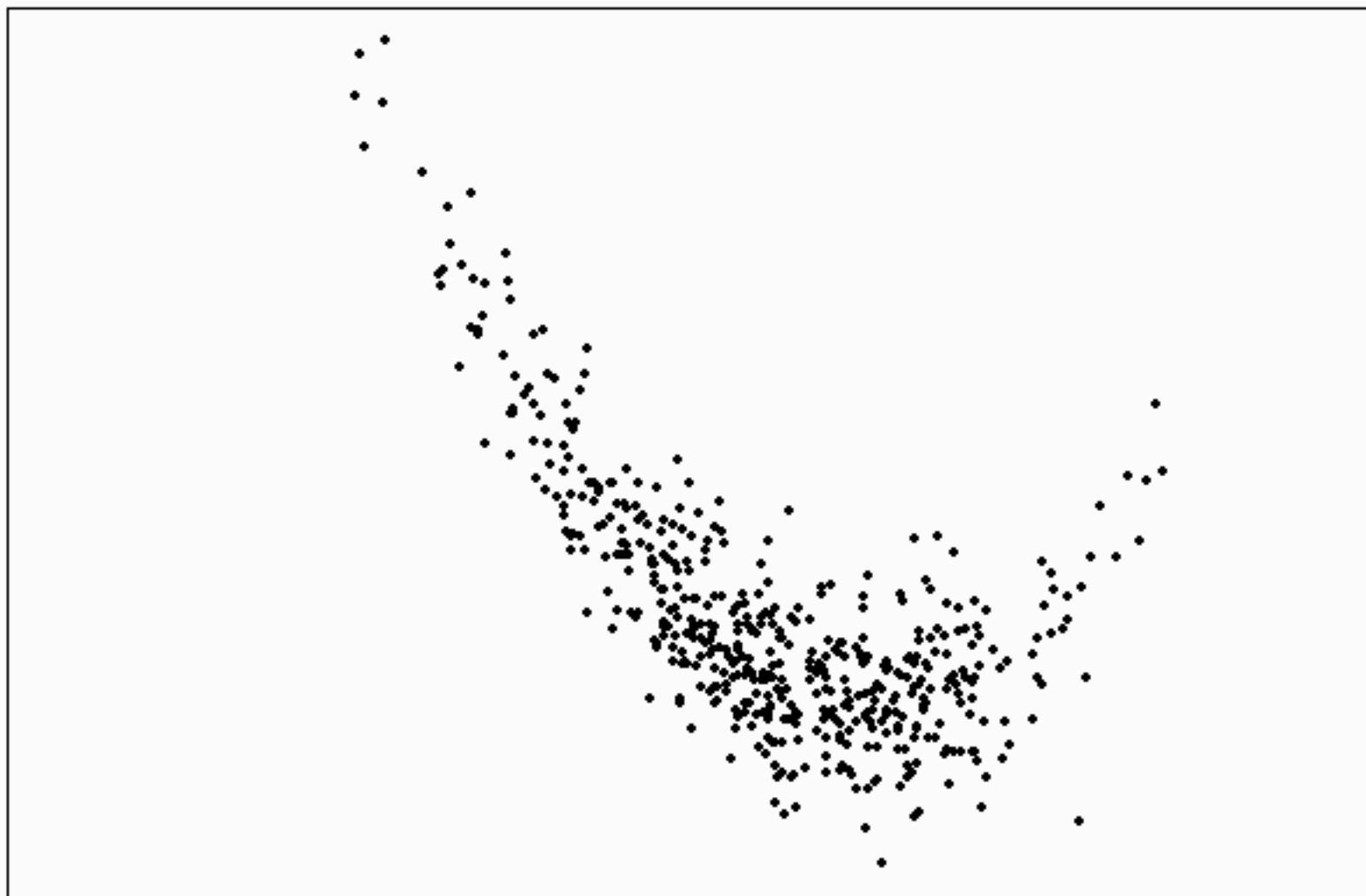
$$r < 0$$

aber

für große X

steigt Y .

$r = -0.67$



In solchen Fällen

sollte man

r **nicht** benutzen.