

4 Form des Zusammenhangs: Regression

4.1 Einführung Regression

Untersuchungen stehen häufig in einem kausalen Abhängigkeitsverhältnis:

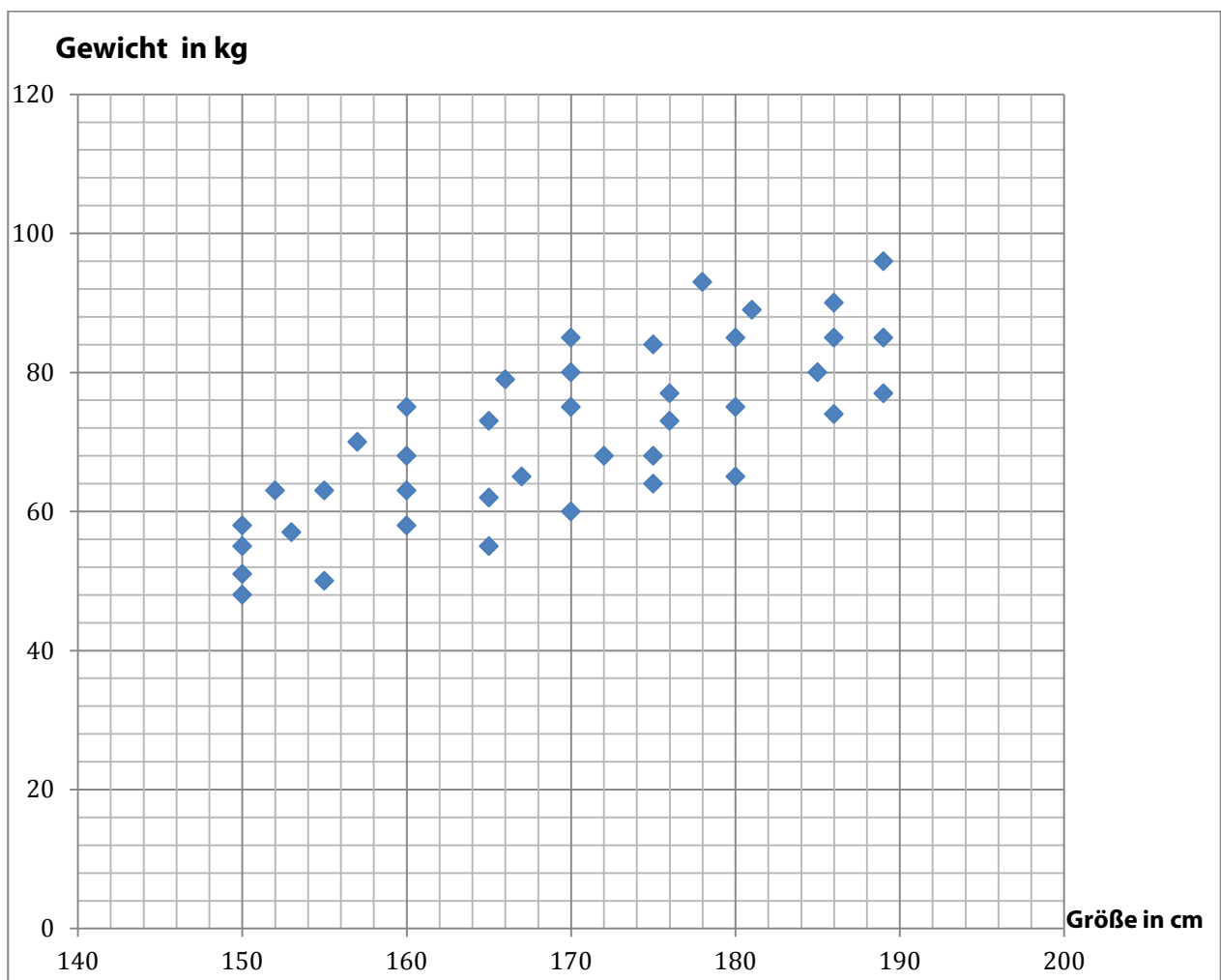
Preis eines Gutes von der Nachfrage

Ernteertrag von Düngerleistung

4.1.1 Regressionskurven (Kurven der bedingten Mittelwerte)

Es wurden folgende Beobachtungswerte ermittelt:

Größe der x_i in cm	Gewicht y_i in kg	x_i	y_i	x_i	y_i	x_i	y_i
150	48	160	63	170	80	180	75
150	51	160	68	170	85	180	85
150	55	160	75	172	68	181	89
150	58	165	55	175	64	185	80
152	63	165	62	175	68	186	74
153	57	165	73	175	84	186	85
155	50	166	79	176	73	186	90
155	63	167	65	176	77	189	77
157	70	170	60	178	93	189	85
160	58	170	75	180	65	189	96



Übung, Kurven der bedingten Mittelwerte

Der untenstehenden Kontingenztabelle liegt eine Untersuchung eines Heiratsinstituts über das Heiratsalter von Männern und Frauen zugrunde.

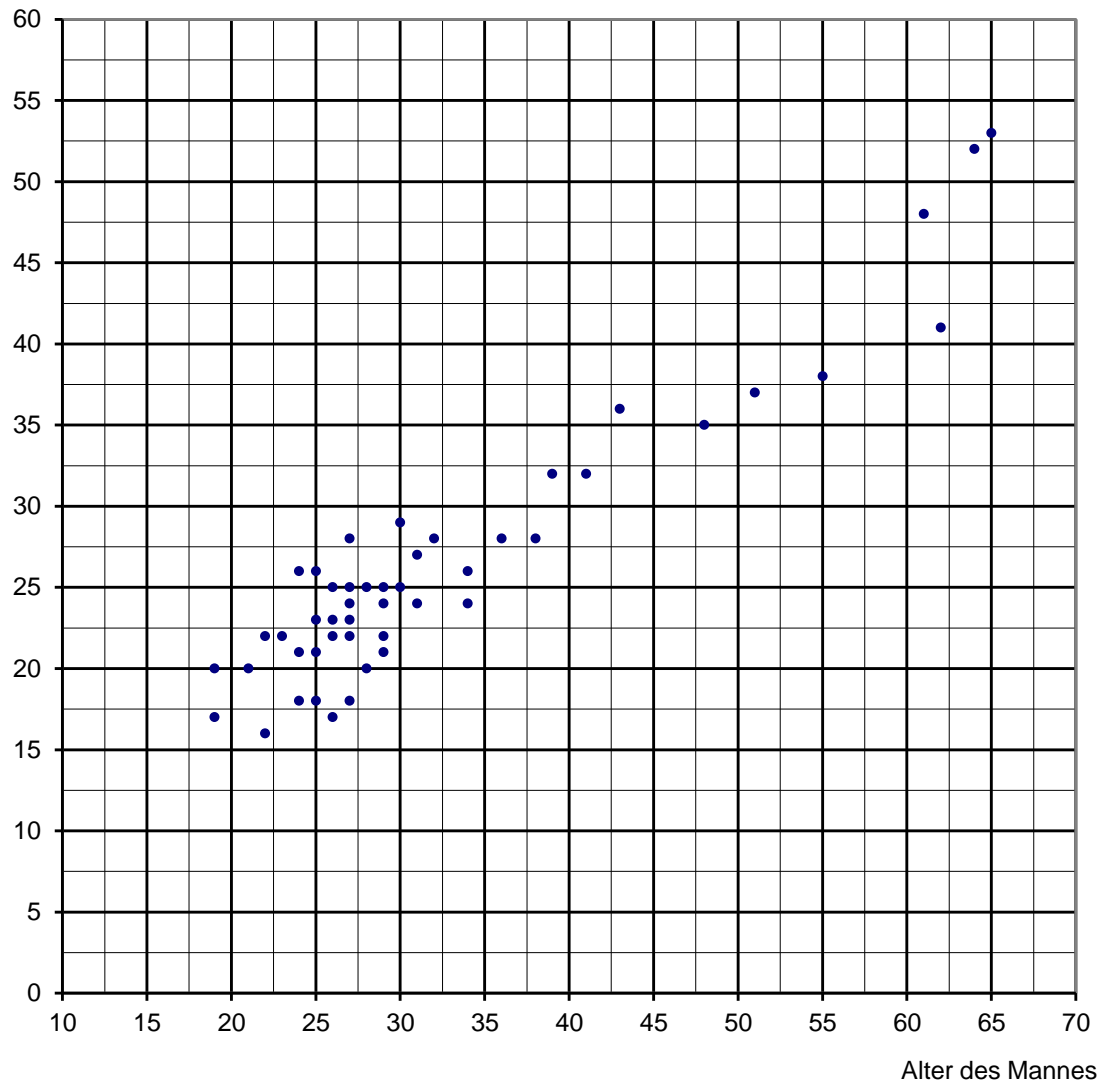
Y Frau	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	X fest Bed. MW
X Mann									
15-20	1	1	0	0	0	0	0	0	
20-25	2	4	1	0	0	0	0	0	
25-30	3	11	8	0	0	0	0	0	
30-35	0	2	5	0	0	0	0	0	
35-40	0	0	2	1	0	0	0	0	
40-45	0	0	0	1	1	0	0	0	
45-50	0	0	0	0	1	0	0	0	
50-55	0	0	0	0	1	0	0	0	
55-60	0	0	0	0	1	0	0	0	
60-65	0	0	0	0	0	1	1	2	
Y fest bed. MW									

Interpretation:

Männer im Alter von 40-45 Jahren heiraten Frauen mit dem Durchschnittsalter von _____.

Frauen im Alter von 25-30 Jahren heiraten Männer mit dem Durchschnittsalter von _____.

Alter der Frau

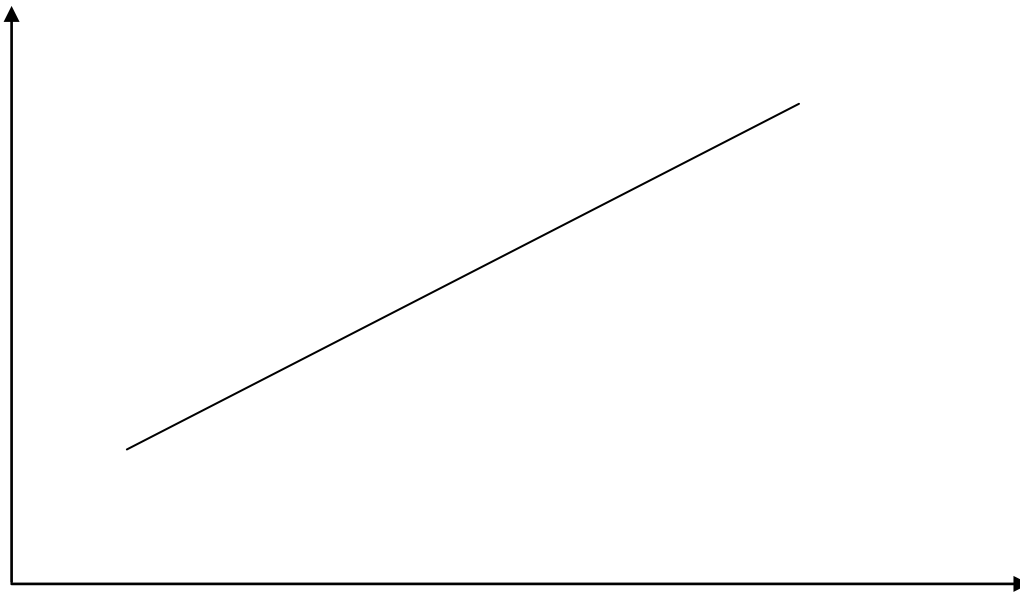


4.1.2 Lineare Regressionsfunktion

Ziel ist es, eine Geradengleichung zu ermitteln, um Probleme rein rechnerisch lösen zu können.

Forderung: Die Summe der quadrierten Abweichung der Beobachtungswerte von der Regressionsgeraden soll minimal sein.

Bestimmung der ersten Regressionsgeraden mit der Methode der kleinsten Quadrate.

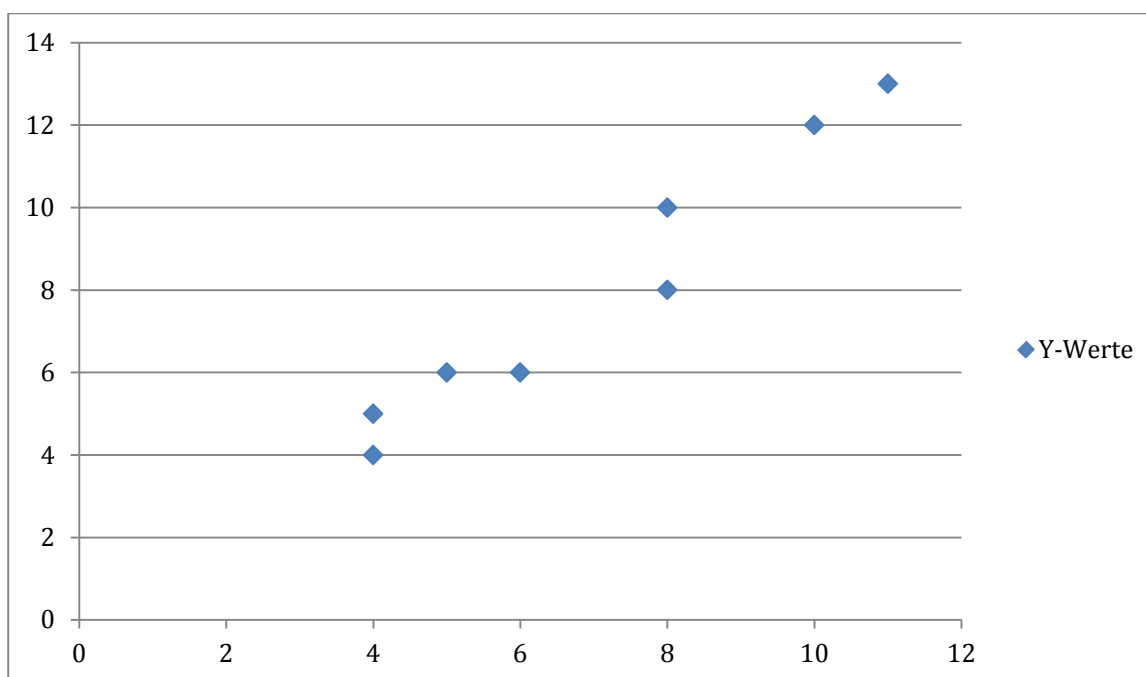


4.2 Lineare Regressionsfunktion Beispiel 1

Es soll die Abhängigkeit des Umsatzes Y eines Unternehmens von der Höhe seiner Werbungskosten X mit Hilfe einer Regressionsgeraden dargestellt werden.

Für die Jahre 2008 – 2015 liegen folgende Werte vor:

Jahr	2008	2009	2009	2010	2011	2012	2014	2015
x_i (in Tsd. Euro)	4	4	5	6	8	8	10	11
y_i (in Mio. Euro)	4	5	6	6	8	10	12	13



Berechnen Sie die Regressionsgerade $y = a + bx$ und zeichnen Sie diese in das oben stehende Koordinatensystem ein.

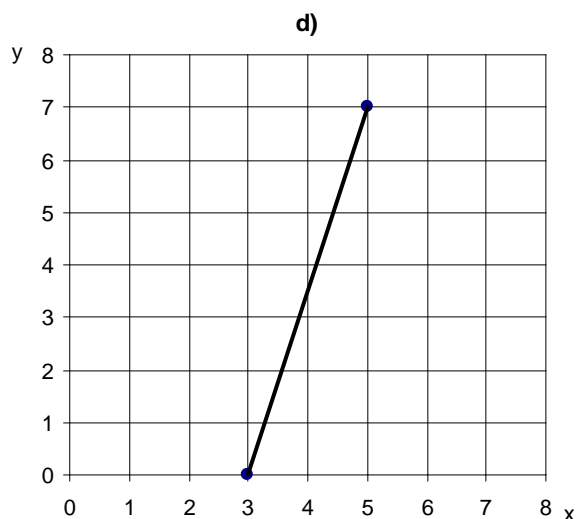
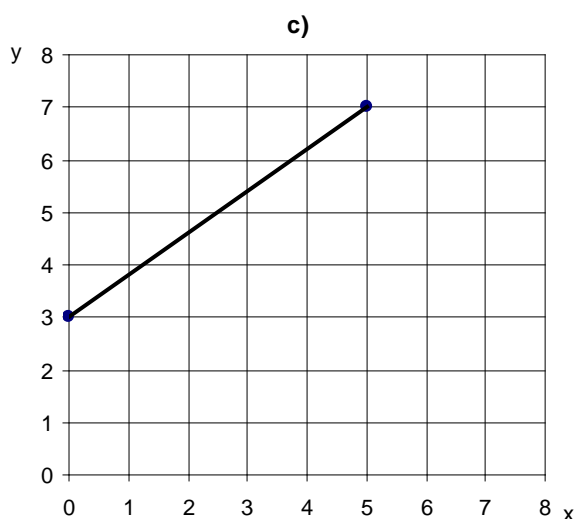
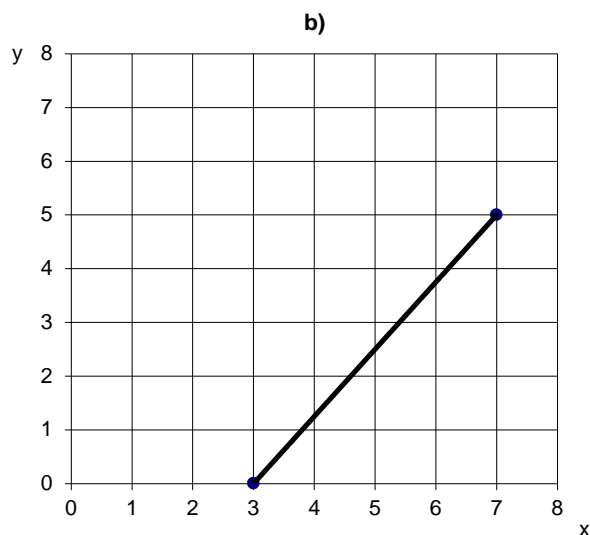
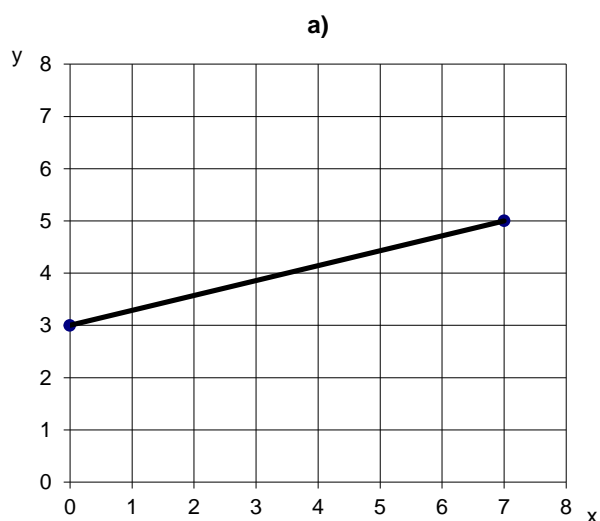
x_i	y_i				
4	4				

4.3 Übung lineare Regressionsfunktion

Welche der abgebildeten Geraden ist die nach der Methode der kleinsten Quadratsumme berechnete Regressionsgerade $y = a + bx$ für die 5 Wertepaare?

x_i	2	4	5	6	8
y_i	5	6	7	7	10

x_i	y_i		



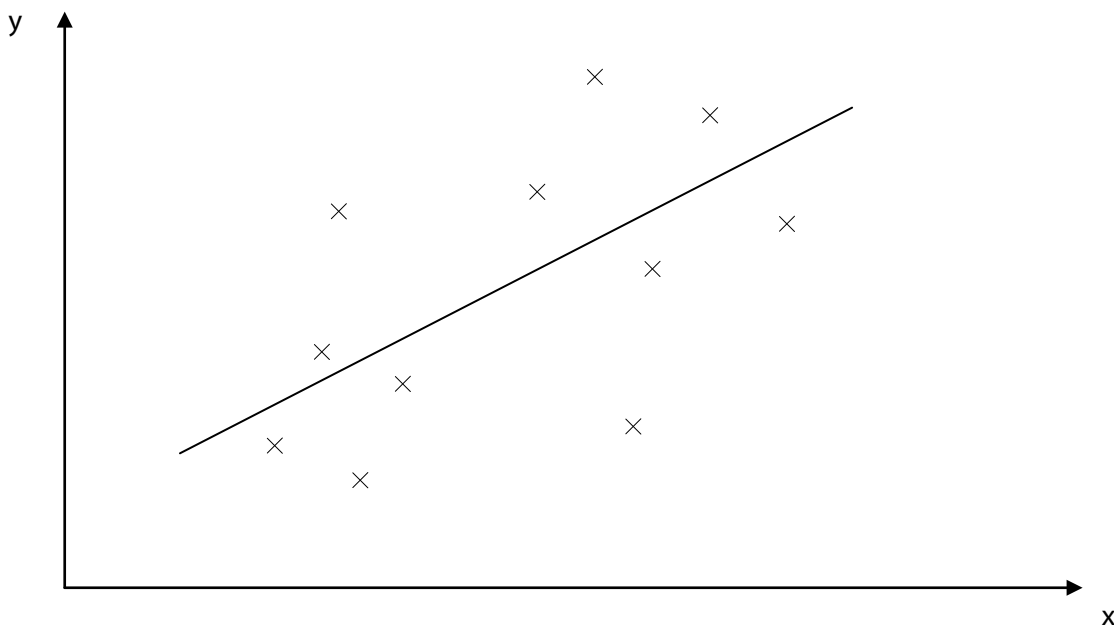
4.4 Zweite Regressionsgerade

Wiederholung: Die Gleichung der **ersten Regressionsgeraden** lautet:

Schlussfolgerung:

Herleitung der Formeln durch Minimierung von:

Analog für die **zweite Regressionsgerade**:



4.5 Übungen lineare Regressionsfunktionen

4.5.1 Aufgabe 1 – Miete und Nettoeinkommen

Die 15 Bewohner eines Hauses werden nach ihrem monatlichen Nettoeinkommen (X) und ihren monatlichen Ausgaben für Miete (Y) gefragt.

Man erhält folgende Angaben:

Nettoeinkommen x_i	Mietausgaben y_i					
450	150					
1600	300					
1000	230					
1200	225					
1850	350					
1300	300					
1400	280					
1200	250					
2000	350					
1800	400					
500	125					
800	220					
1400	250					
1000	200					
500	120					

Ermitteln Sie die Gleichungen beider Regressionsgeraden und zeichnen Sie diese in ein Koordinatensystem ein.

4.5.2 Aufgabe 2 – Auslagen und Anzahl Kunden

In einer Boutique soll untersucht werden, ob zwischen der Zahl der Kunden, die den Laden betreten (Merkmal Y) und der Zahl der Kunden, die dort vor dem Schaufenster stehen (Merkmal X) ein Zusammenhang besteht, der sich durch eine lineare Regressionsfunktion ausdrücken lässt. In verschiedenen Stundenintervallen wurden die folgenden Beobachtungen gemacht:

i	Zahl der Auslagen- betrachter x_i	Zahl der Kunden, die den Laden betraten y_i
1	5	3
2	7	5
3	6	4
4	4	2
5	5	4
6	9	6

Ermitteln Sie die Gleichungen beider Regressionsgeraden.

4.5.3 Aufgabe 3 – Kunstdünger und Ernteergebnis

Auf gleich großen Flächeneinheiten eines homogenen Bodens wurden unterschiedliche Mengen eines Kunstdüngers eingesetzt. Für die Mengen des eingesetzten Kunstdüngers in dz (=X) und das Ernteergebnis in dz (=Y) wurden folgende Werte beobachtet:

i	Kunstdünger x_i	Ernteergebnis y_i
1	1	24
2	2	32
3	3	32
4	5	47
5	7	58
6	9	63

- Berechnen Sie die Gleichung der 1. Regressionsgeraden.
- Bestimmen Sie das zu erwartende Ernteergebnis bei einem Kunstdüngereinsatz von 11 dz.