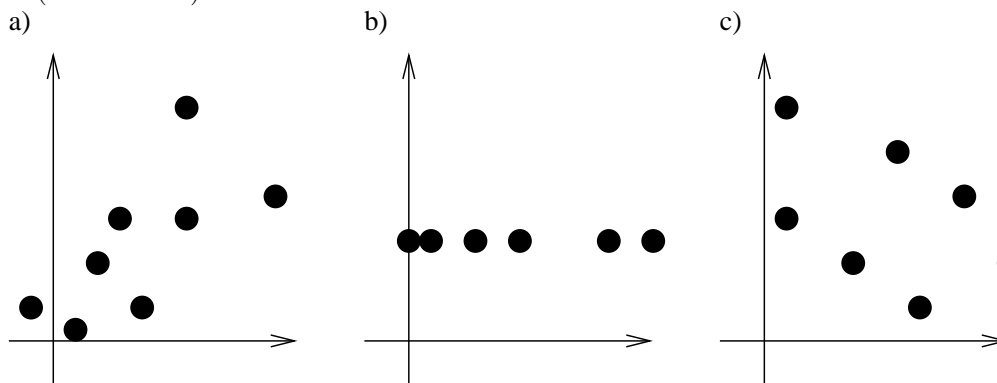




## 4. Übungsblatt zur „Statistik I für Human- und Sozialwissenschaft“

**Aufgabe 13** (Korrelation)

(3 Punkte)



Welche Aussage können Sie über die Größe der Korrelation der Datenmengen machen (z.B.  $r_{x,y} = -1$ ,  $-1 < r_{x,y} < 0$ ,  $r_{x,y} = 0$ ,  $0 < r_{x,y} < 1$  oder  $r_{x,y} = 1$ )? Begründen Sie Ihre Aussage!

**Aufgabe 14** (Lineare Regression)

(3 Punkte)

Fünf Versuchspersonen absolvieren zwei unterschiedliche psychologische Tests zur Kraftfahreignung. Dabei entstehen die folgenden Testergebnisse:

Versuchsperson-Nr.	Test x	Test y
1	31	15
2	128	95
3	67	35
4	46	40
5	180	80

- (a) Zeichnen Sie die Punkte (Testergebnis in Test x, Testergebnis in Test y) in ein Streudiagramm (Scatterplot).
- (b) Berechnen Sie die zugehörige Regressionsgerade und zeichnen Sie diese in das Diagramm ein. (*Hinweis:* Laut Vorlesung ist die Formel für die Regressionsgerade:

$$y = \hat{a} \cdot (x - \bar{x}) + \bar{y}$$

mit  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  und  $\hat{a} = \frac{s_{x,y}}{s_x^2} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ .

- (c) Interpretieren Sie den Verlauf der Regressionsgerade bzgl. eines Zusammenhanges zwischen Test x und Test y.

**Aufgabe 15**

(2 Punkte)

In der folgenden Tabelle ist der Schuldenstand der Länder und Gemeinden je Einwohner in den einzelnen Bundesländern am 31.12.2008 aufgelistet (Quelle: Statistische Bundesamt):

Bundesland	Schulden (Euro)	Bundesland	Schulden (Euro)
Baden-Württemberg	4439	Niedersachsen	7218
Bayern	2861	Nordrhein-Westfalen	7620
Berlin	16340	Rheinland-Pfalz	7904
Brandenburg	7408	Saarland	10182
Bremen	23084	Sachsen	3229
Hamburg	12223	Sachsen-Anhalt	9467
Hessen	6344	Schleswig-Holstein	8677
Mecklenburg-Vorpommern	6893	Thüringen	7803

- (a) Bestimmen Sie das empirische arithmetische Mittel.  
 (b) Warum stimmt es nicht mit der bundesweiten Verschuldung je Einwohner von 5866 Euro überein (die Schulden des Bundes auch hier nicht mitgerechnet)?

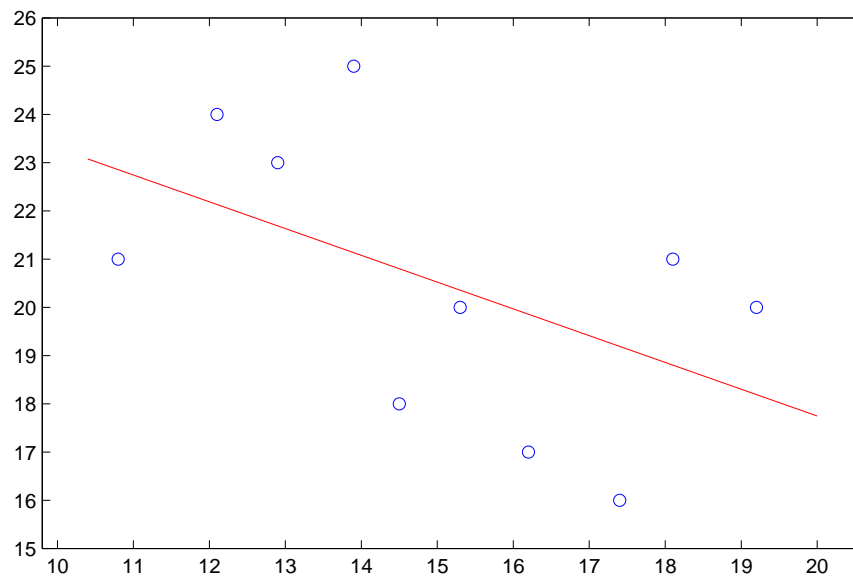
**Aufgabe 16** (Lokale Mittelung)

(4 Punkte)

In einer Fertigungsanlage kann eine der Maschinen durch eine Stellschraube justiert werden. Die Anzahl der Produktionsfehler lässt sich durch diese Schraube beeinflussen. Bei der Feinabstimmung wurden die folgenden Zahlen in Abhängigkeit von der Tiefe der Schraube beobachtet:

Tiefe ( $\mu\text{m}$ )	10,8	12,1	12,9	13,9	14,5	15,3	16,2	17,4	18,1	19,2
Fehlerzahl	21	24	23	25	18	20	17	16	21	20

Diese Daten sind in folgendem Scatterplot dargestellt, in dem auch schon die zugehörige Regressionsgerade eingezeichnet ist:



- (a) Wir wollen nun eine Schätzung für die Fehlerzahlen bei den Tiefen  $x = 11, x = 12, x = 13, \dots, x = 20$  mittels *lokaler Mittelung* bestimmen. Berechnen Sie dazu das (arithmetische) Mittel aller Punkte, deren Abstand vom jeweils betrachteten  $x$ -Wert kleiner als die Schranke  $h=1$  entfernt ist und tragen Sie die Werte in folgende Tabelle ein.

$x$ -Wert (Tiefe)	11	12	13	14	15	16	17	18	19	20
$y$ -Wert ( $h = 1$ )										

- (b) Tragen Sie alle in (a) berechneten Punkte in den Scatterplot ein und verbinden Sie diese.  
 (c) Vergleichen Sie das Ergebnis dieser *nichtparametrischen Regressionsschätzung* mit dem der linearen Regression (im Scatterplot).