



### 3. Übungsblatt zur „Einführung in die Stochastik“

#### Aufgabe 10

(3 Punkte)

Die folgende Messreihe beschreibt die Exportquote (prozentualer Anteil des Auslandsumsatzes am Gesamtumsatz) im Jahr 2002 für 23 Teilbereiche des Verarbeitenden Gewerbes in Deutschland:

12.7, 8.8, 37.3, 32.2, 30.4, 20.1, 36.1, 7.7, 3.6, 51.5, 33.0, 22.4, 38.6,  
24.8, 50.6, 38.6, 36.4, 54.8, 49.4, 59.6, 53.2, 24.0, 25.4

- (a) Bestimmen Sie das Mittel, den Median, die Spannweite, die Varianz, die Standardabweichung und den Interquartilabstand dieser Messreihe.
- (b) Zeichnen Sie ein Histogramm dieser Messreihe bzgl. der Partition

$\{[0, 10), [10, 20), [20, 30), [30, 40), [40, 50), [50, 70), [70, 100]\}$

des Intervalls  $[0, 100]$ .

#### Aufgabe 11

(3 Punkte)

Gegeben sei eine zweidimensionale Messreihe

$(x_1, y_1), \dots, (x_n, y_n)$

vom Umfang  $n$ . Anstelle einer Geraden (wie bei der linearen Regression) könnte man analog auch eine Parabel

$$y = a + b \cdot x + c \cdot x^2$$

durch Minimierung von

$$F(a, b, c) := \sum_{i=1}^n (y_i - (a + b \cdot x_i + c \cdot x_i^2))^2$$

an die Daten anpassen. Zeigen Sie (durch Nullsetzen geeigneter Ableitungen), dass die Werte  $a, b, c$ , für die  $F(a, b, c)$  minimal wird, Lösungen des linearen Gleichungssystems

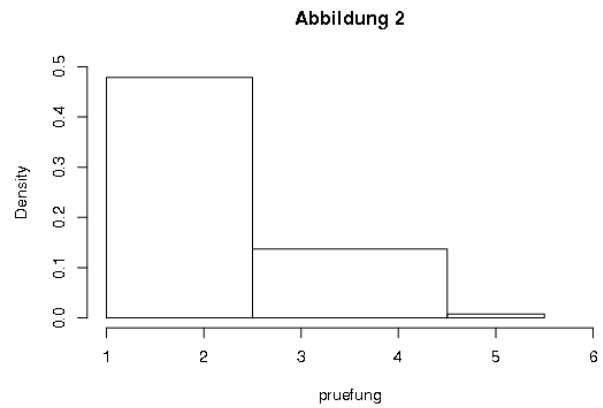
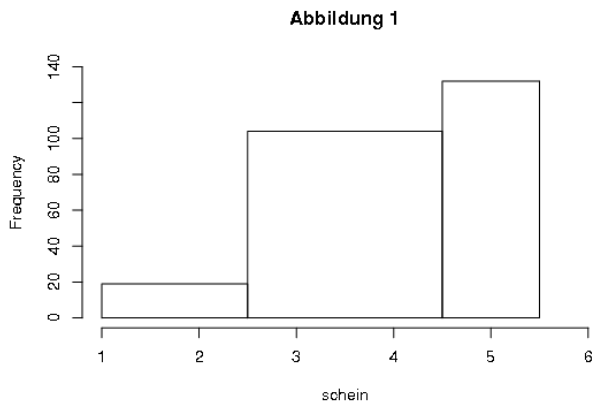
$$\begin{aligned} a + b \cdot \frac{1}{n} \sum_{i=1}^n x_i + c \cdot \frac{1}{n} \sum_{i=1}^n x_i^2 &= \frac{1}{n} \sum_{i=1}^n y_i \\ a \cdot \frac{1}{n} \sum_{i=1}^n x_i + b \cdot \frac{1}{n} \sum_{i=1}^n x_i^2 + c \cdot \frac{1}{n} \sum_{i=1}^n x_i^3 &= \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i \\ a \cdot \frac{1}{n} \sum_{i=1}^n x_i^2 + b \cdot \frac{1}{n} \sum_{i=1}^n x_i^3 + c \cdot \frac{1}{n} \sum_{i=1}^n x_i^4 &= \frac{1}{n} \sum_{i=1}^n x_i^2 \cdot y_i \end{aligned}$$

sind.

#### Aufgabe 12

(3 Punkte)

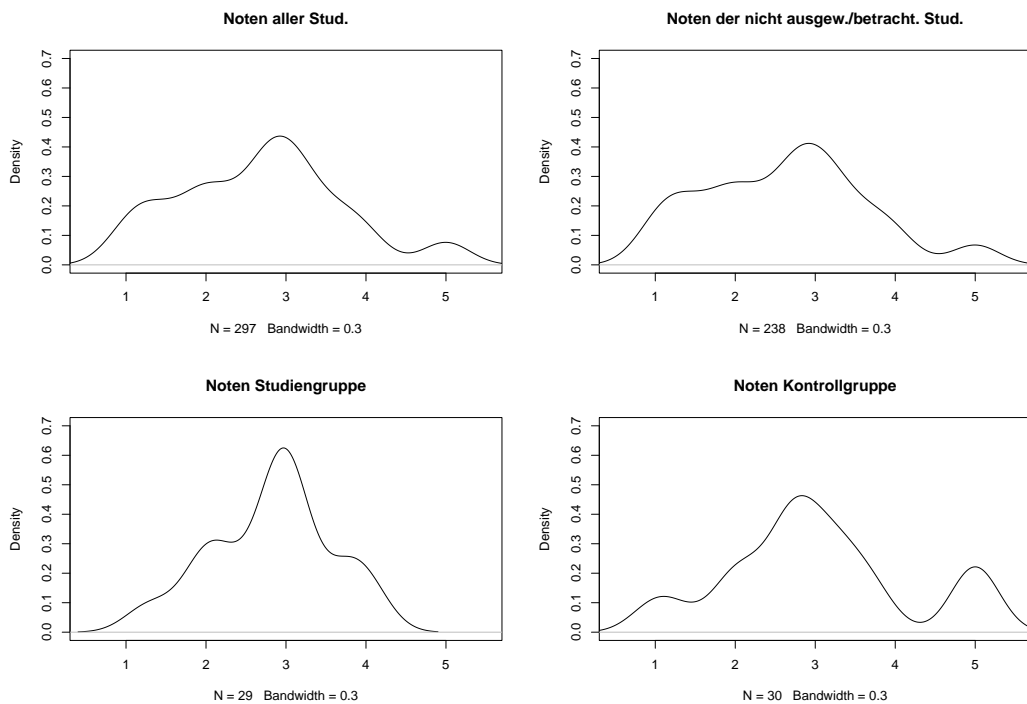
- a) Das Säulendiagramm in Abbildung 1 beschreibt die Noten von  $n = 255$  Studenten in der Scheinklausur zur Vorlesung "Statistik I für WiWi".
- a<sub>1</sub>) Inwieweit ist die Darstellung in diesem Säulendiagramm irreführend?
- a<sub>2</sub>) Stellen Sie die Daten in einem Histogramm so dar, dass die Flächeninhalte der einzelnen Balken (aufgrund von Problemen beim Ablesen der Werte in Abbildung 1 eventuell nur ungefähr) proportional zur Anzahl der Datenpunkte in den zugrundeliegenden Intervallen sind.
- b) Das Histogramm in Abbildung 2 beschreibt die Noten von  $n = 255$  Studenten in der Diplom-Vorprüfung zur Vorlesung "Statistik I für WiWi". Bestimmen Sie mit Hilfe dieses Histogramms (approximativ) die Anzahl der Studenten, die in der Prüfung eine Note besser als 2.5 hatten (d.h., die als Note eine 1.0, 1.3, 1.7, 2.0 oder 2.3 hatten).



**Aufgabe 13**

(3 Punkte)

Im Rahmen einer Diplomarbeit wurde an der Universität Stuttgart versucht, ein Verfahren zu entwickeln, welches in der Lage ist, StudentInnen zu identifizieren, die voraussichtlich Probleme mit dem Bestehen einer Statistik-Prüfung haben werden. Dieses Verfahren wurde im Rahmen der Vorlesung *Statistik II für WirtschaftswissenschaftlerInnen* überprüft. Dazu wurden durch Anwendung dieses Verfahrens von den ca. 300 Teilnehmern der Klausur 60 ausgewählt, und zufällig in zwei gleich große Gruppen, SG und KG, unterteilt. Die StudentInnen in der SG wurden vor der Prüfung schriftlich zu einem ca. sechsständigen Zusatzkurs zur Klausurvorbereitung eingeladen. In der unten stehenden Abbildung sind Kern-Dichteschätzer angewandt auf die Noten aller StudentInnen, bzw. der StudentInnen in der SG, bzw. der StudentInnen in der KG, bzw. der StudentInnen, die weder in der SG noch in der KG waren, abgebildet. Wie können Sie durch Vergleich dieser Kern-Dichteschätzer feststellen, ob



- (a) das Verfahren wirklich vor allem StudentInnen ausgewählt hat, welche die Klausur vermutlich nicht bestehen ?
- (b) das Anbieten des Zusatzkurses zu einer Verringerung der Nichtbestehensquote geführt hat ?