

## Übungsblatt 4

### Aufgabe 1:

Berechnen Sie für die statistischen Reihen die Varianzen, Kovarianzen und Korrelationskoeffizienten

- a)  $s_X^2, s_Y^2, s_Z^2, s_U^2, s_V^2, s_W^2, s_T^2$
- b)  $c_{XY}, c_{YZ}$
- c)  $c_{ZU}, c_{VT}, c_{UV}$
- d)  $r_{UW}, r_{ZU}, r_{VT}$

Gegeben sind die folgenden statistischen Reihen

X:	0, 0, 0, 0, 0, 0, 0, 0
Y:	1, 1, 1, 1, 1, 1, 1, 1
Z:	a, a, a, a, a, a, a, a
U:	0, 1, 0, 1, 0, 1, 0, 1
V:	1, 0, 1, 0, 1, 0, 1, 0, 1
W:	3, 3, 3, 7, 5, 3, 5, 4
T:	1, 2, 3, 4, 5, 6, 7, 8, 9

(Aufgabe aus Schira, Kapitel 2; Aufg. 2.10)

### Aufgabe 2:

Gegeben ist die statistische Reihe X. Sie hat den Mittelwert 240 und die Varianz 81. Die statistische Reihe Y errechnet sich aus X, indem man jedes Element der Reihe X mit dem konstanten Faktor  $b > 0$  multipliziert, also

$$y_i := b \cdot x_i \text{ für } i = 1, \dots, n.$$

- a) Berechnen Sie die Kovarianz zwischen X und Y und den Korrelationskoeffizienten.
- a) Welchen Wert hat  $r_{XY}$ , wenn der Faktor  $b$  negativ ist?

(Aufgabe aus Schira, Kapitel 3; Aufg. 3.3)

### Aufgabe 3:

**Erwerbstätige.** In der amtlichen Statistik finden Sie folgende Verteilung der Erwerbstätigen in der Bundesrepublik Deutschland für April 1990 (in 1.000 Personen):

Altersgruppe		Selbständige u. mithelfende	
von ...	bis unter ...	Familienangehörige	abhängige Beschäftigte
15-25		99	5002
25-35		531	7009
35-45		1243	5731
45-55		937	6051
55-65		595	2284
65-75		160	63
75-95		42	16

*Quelle: Statistisches Jahrbuch 1992*

- Was sind die statistischen Einheiten, Grundgesamtheiten und Merkmale?
- Zeichnen Sie ein Histogramm der Randverteilung des Merkmals Alter und der beiden bedingten Verteilungen des Merkmals Alter.
- Zeichnen Sie beide bedingten Verteilungsfunktionen in ein Koordinatensystem. Geben Sie die bedingten Mediane an.
- Berechnen Sie die beiden bedingten Mittelwerte.
- Müssen die Selbständigen länger arbeiten? Welcher Anteil der Selbständigen und welcher Anteil der Unselbständigen ist 55 Jahre und älter? Welcher Anteil der über 65jährigen Erwerbstätigen ist selbständig? Kann man aus diesen Daten die durchschnittliche "Lebensarbeitszeit" berechnen?

Hinweis: Gehen Sie von der Annahme einer gleichmäßigen Verteilung innerhalb der Altersgruppe aus. (Aufgabe aus Schira, Kapitel 3; Aufg. 3.5)

### Aufgabe 4:

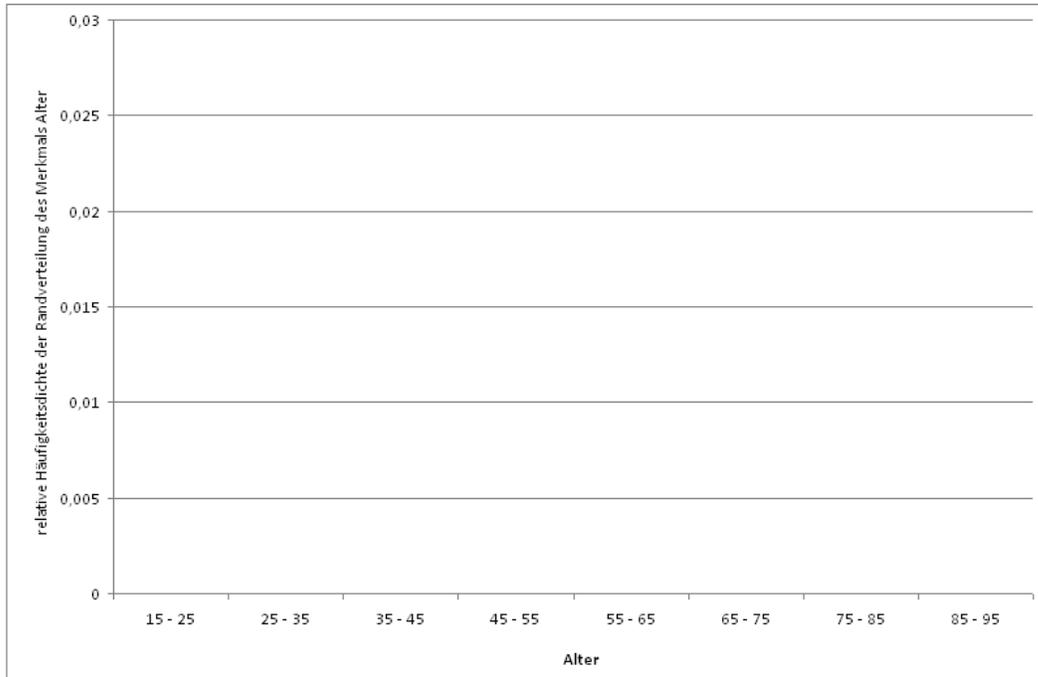
Für die Größe und das Gewicht von 100 Studenten sei folgende Kontingenzstabelle gegeben:

Gewicht (X) in kg  (von ... bis unter)	Größe (Y) in cm (von ... bis unter...)						
	$\hat{y}_j$	145 - 155	155 - 165	165 - 175	175 - 185	185 - 195	195 - 205
$\hat{x}_i$		150	160	170	180	190	200
45 - 55	50	4	2	4			
55 - 65	60	3	3	2	4		
65 - 75	70		5	10	4	1	
75 - 85	80			10	12	8	
85 - 95	90				5	8	5
95 - 105	100					3	7

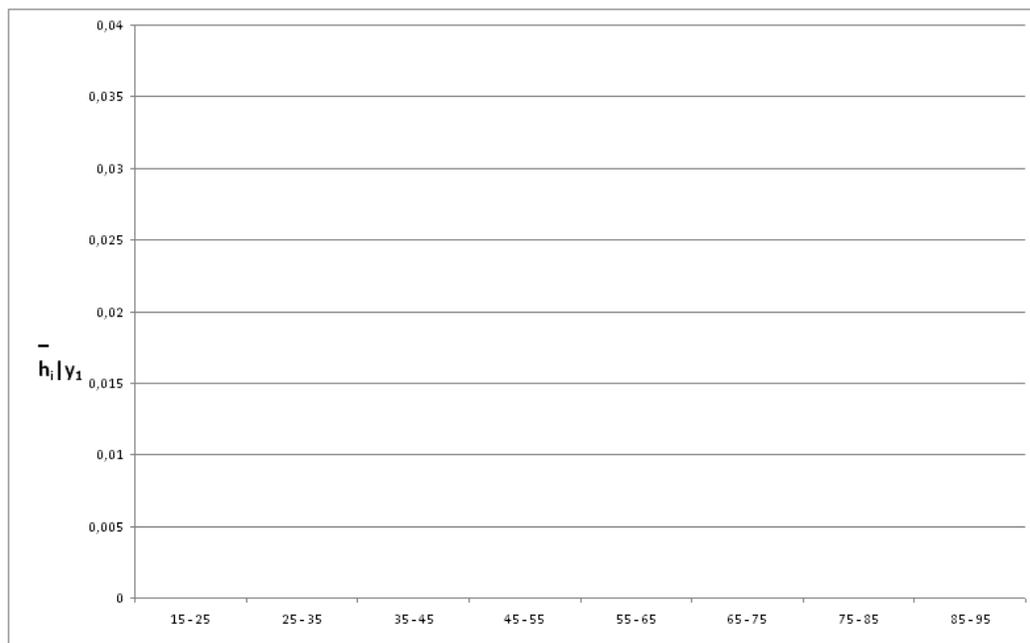
- Bestimmen Sie die Randverteilungen.
- Berechnen Sie  $\bar{x}$ ,  $\bar{y}$ ,  $s_X^2$ ,  $s_Y^2$ .
- Berechnen Sie die bedingten Mittelwerte  $\bar{x}|y_j$  und  $\bar{y}|x_i$ .
- Bestimmen Sie die Kovarianz zwischen X und Y und interpretieren Sie das Ergebnis.

## Anhang

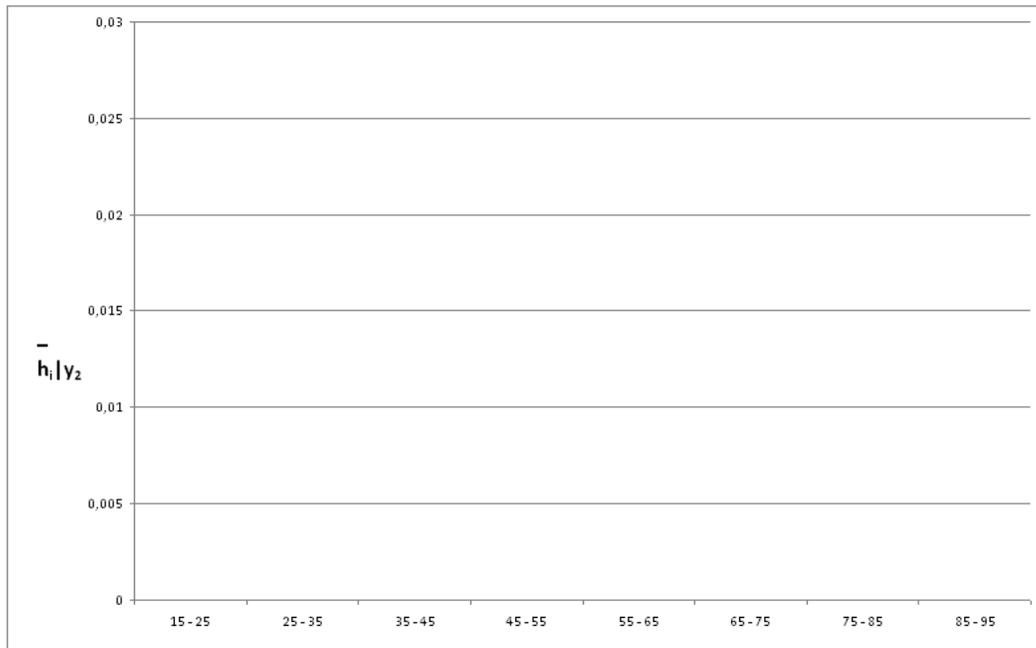
Koordinatensystem für Aufgabe 3b) Randverteilung des Merkmals Alters



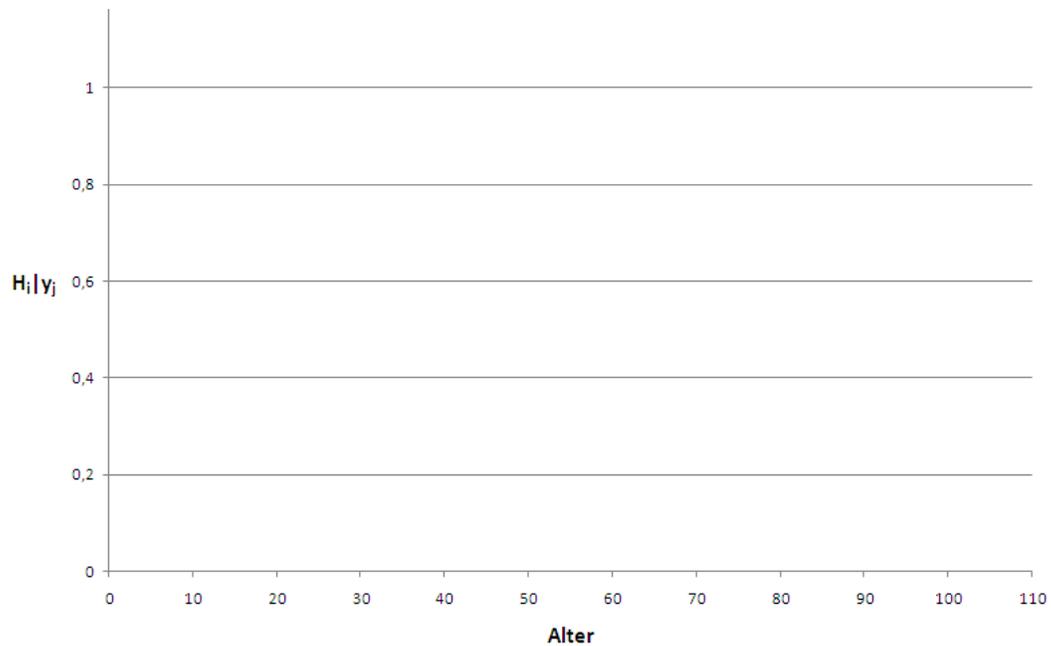
Koordinatensystem für Aufgabe 3b) bed. Verteilung des Alters für Selbständige,...



**Koordinatensystem für Aufgabe 3b) bed. Verteilung des Alters für abhäng.  
Beschäftigte**



**Koordinatensystem für Aufgabe 3c) bedingte Verteilungsfunktionen**



## Zusatzaufgaben

### Aufgabe Z 1:

- a) Unterstellen Sie, dass Sie eine Stichprobe mit  $n > 3$  Beobachtungen der Variable X in drei disjunkte Teilstichproben (Gruppen) zerlegen können. Beweisen Sie die Zerlegung der Gesamtvarianz in die innere Varianz und die äußere Varianz entsprechend Kapitel 2.6 im Lehrbuch von Schira.

**Hinweis:** Erläutern Sie jeden Ihrer Beweisschritte genau.

- b) Die Variable X messe die Eigenkapitalrentabilität eines Unternehmens und es liegen Ihnen Beobachtungen für X in drei Wirtschaftszweigen vor. Erläutern Sie die Varianzzerlegung in Aufgabenteil a) für dieses konkrete Beispiel und interpretieren Sie die Varianzzerlegung ökonomisch.

### Aufgabe Z 2:

Bei einem Experiment in der Vorlesung Statistik I wurden 30 Bonbons, darunter 15 gelbe und 15 orange Bonbons, zufällig an 30 Studierende verteilt. Dabei ergab sich folgende Kontingenztabelle der absoluten Häufigkeiten für den Zusammenhang zwischen der Farbe des gewählten Bonbons und dem Geschlecht des/der Studierenden:

	Farbe Gelb	Farbe Orange
Männlich	8	3
Weiblich	7	12

- a) Wie hoch ist der Anteil weiblicher Studierender?
- b) Stellen Sie den Zusammenhang zwischen Farbwahl und Geschlecht graphisch in geeigneter Form dar. Erläutern Sie Ihre Darstellung.
- c) Berechnen Sie die relativen Häufigkeiten der Farbwahl sowohl unbedingt als auch bedingt auf das Geschlecht.  
Sind Farbwahl und Geschlecht statistisch unabhängig?
- d) Berechnen Sie den Kontingenzkoeffizienten und den korrigierten Kontingenzkoeffizienten.  
Interpretieren Sie das Ergebnis im Hinblick auf die Frage nach der statistischen Unabhängigkeit.

### Aufgabe Z 3:

Für Bargeldbestand (B) und Geschlecht (G) von 10 Studierenden ergaben sich folgende Beobachtungswerte:

(m, €10), (m, €15), (m, €15), (m, €15), (m, €50),  
(m, €10), (w, €10), (w, €10), (w, €100), (w, €10),

wobei m  $\hat{=}$  männlich und w  $\hat{=}$  weiblich.

**Hinweis:** Stellen Sie die Kontingenztabelle auf.

- A) Frauen haben im Durchschnitt mehr Bargeld als Männer.
- B) Es gilt:  $h(G = m | B = €10) = 0,4$ , da es drei Frauen aber nur zwei Männer mit einem Bargeldbestand von €10 gibt.

- C) Bargeld und Geschlecht sind statistisch unabhängig.
- D) Die relative Häufigkeitsdichte an der Stelle  $B = €15$  in der Stichprobe der Männer entspricht  $\bar{h}(B = €15 | G = m) = \frac{1}{12}$  für das Intervall von €10 bis unter €20.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>

Geben Sie an, welche Antwortvorschläge richtig sind. Beachten Sie, dass mindestens ein Antwortvorschlag richtig ist. Schreiben Sie Ihre Berechnungen zur Lösung jeder Teilaufgabe als Teil der Lösung auf und erläutern Sie kurz verbal Ihre Vorgehensweise.  
(Aufgabe 8 der Zwischenklausur vom 25.05.09)