Aufgabe 21

Für den Aktienkurs und den Optionspreis einer deutschen Aktie ergaben sich folgende Daten:

Kurs	Optionspreis	Kurs	Optionspreis
240,3	16,00	226	11,20
252,5	15,40	202	10,50
238	17,40	208	13,10
228	12,60	177	14,50
223	11,80	190	14,80
238	11,00	180,5	13,70

Zeichnen Sie für diesen Datensatz das Streuungsdiagramm und berechnen Sie den Bravais-Pearson-Korrelationskoeffizienten.

Lösungshinweis:

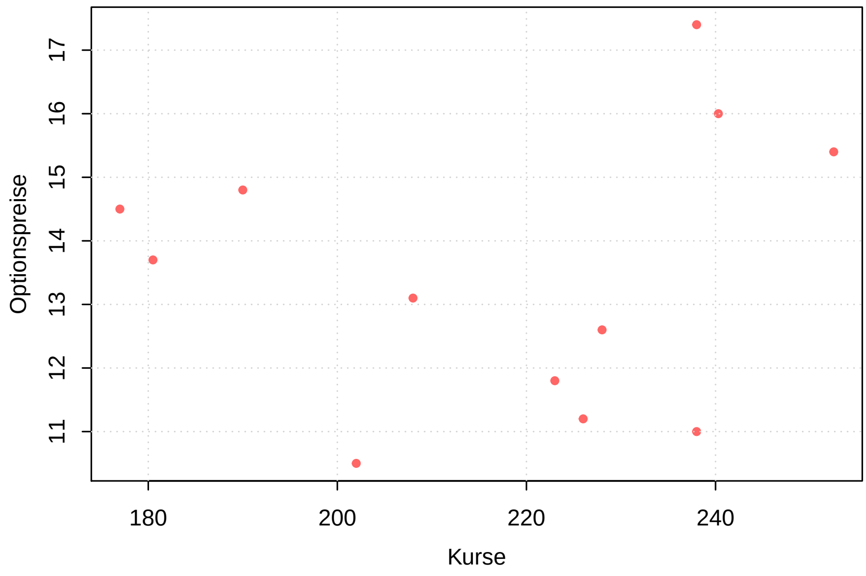
```
# Eingabe der Daten
Kurse = c(240.3, 252.5, 238, 228, 223, 238,  226, 202, 208, 177, 190, 180.5)
Optionspreise = c(16, 15.4, 17.4, 12.6, 11.8, 11, 11.2, 10.5, 13.1, 14.5, 14.8, 13.7)

# Bravais-Pearson-Korrelationskoeffizient:
cor(Kurse, Optionspreise)

## [1] 0.134227
```

Die beiden Merkmale sind anscheinend fast nicht korreliert.

```
# Streuplot:
plot(Kurse, Optionspreise)
```



Aufgabe 22

Zwei Personen sollen fünf verschiedene Produkte A bis E durch Angabe einer Reihenfolge beurteilen. Die Befragung ergab folgende Ergebnisse:

Produkt	Person I	Person II
A	5	3
B	2	1
C	3	4
D	4	2
E	1	5

Berechnen Sie den Korrelationskoeffizienten von Spearman.

Lösungshinweis:

Sowohl mit als auch mit Formel bzw. Taschenrechner ergibt sich $r_{SP} = -0.3$.

```
Person1 <- c(5, 2, 3, 4, 1)
Person2 <- c(3, 1, 4, 2, 5)
```

```
# method='spearman' ist der Rangkorrelationskoeffizient
# (Schalter ist hier eigentlich überflüssig, da sowieso
# schon Ränge vorliegen)
cor(Person1, Person2, method = "spearman")

## [1] -0.3
```

Aufgabe 23

Ein Betrieb hat im Kalenderjahr 2004 zwölf neue Mitarbeiter eingestellt. Von diesen sind unter anderem folgende Daten bekannt:

Mitarbeiter Nr.	Geschlecht	Ausbildungsdauer (in Jahren)	Abschlussnote
1	männlich	9	4
2	weiblich	10	2
3	weiblich	10	4
4	männlich	11	4
5	weiblich	12	2
6	weiblich	13	2
7	weiblich	14	1
8	männlich	15	3
9	männlich	16	2
10	männlich	17	3
11	weiblich	19	3
12	männlich	22	2

a) Geben Sie die Skalierung der drei Merkmale Geschlecht, Ausbildungsdauer und Abschlussnote an.

b) Ermitteln Sie für jedes der drei Merkmale die folgenden Größen, soweit diese aufgrund des jeweiligen Skalenniveaus sinnvollerweise berechnet werden können:
(1) Modus
(2) Median
(3) Arithmetisches Mittel
(4) Mittlere quadratische Abweichung
(5) Variationskoeffizient

c) Geben Sie für jedes der zwei Merkmalspaare
(1) Geschlecht – Abschlussnote
(2) Ausbildungsdauer – Abschlussnote
einen statistisch sinnvollen Korrelationskoeffizienten an.
(Die Korrelationskoeffizienten müssen nicht berechnet werden.)

Lösungshinweis:

a) Geschlecht: *nominal*,
Ausbildungsdauer: *kardinal*,
Abschlussnote: *ordinal*

b)

	Geschlecht	Ausbildungsdauer	Abschlussnote
Modus	(nicht eindeutig)	10	2
Median	–	13,5	2,5
Arithmetisches Mittel	–	14	–
Mittlere quadratische Abweichung	–	14,5	–
Variationskoeffizient	–	0,27	–

c) Geeignete Korrelationskoeffizienten:
(1) Geschlecht – Abschlussnote: *Kontingenzkoeffizient*
(2) Ausbildungsdauer – Abschlussnote: *Rangkorrelationskoeffizient*

Aufgabe 24

Die Aufgliederung einer Population nach Arbeitslosigkeit und Schulbildung liefere folgende Kontingenztabelle:

arbeitslos / Bildung	ja	nein
Volksschule	770.000	13.375.000
mittlere Reife	140.000	4.000.000
Abitur	90.000	1.625.000

Bildung \ arbeitslos	ja	nein	Sum
Volksschule	770000	13375000	14145000
mittlere Reife	140000	4000000	4140000
Abitur	90000	1625000	1715000
Sum	1000000	19000000	20000000

$\frac{1\text{ Mio} \cdot 14{,}145\text{ Mio}}{20\text{ Mio}}$

	ja	nein
VS	707250	.
mR	.	.
Abi	.	.

$$\chi^2 = \frac{(770000 - 707250)^2}{707250} + \ldots = 28909{,}53515$$

$$\Rightarrow k = \sqrt{\frac{\chi^2}{\chi^2 + 20\text{ Mio}}} \approx 0.03799$$

Berechnen Sie den Kontingenzkoeffizienten und den normierten Kontingenzkoeffizienten. Interpretieren Sie das Ergebnis.

Lösungshinweis:

Es ergibt sich: $K = 0.037992$ sowie mit $K_{\max} = 0.7071068$ die normierte Variante

$$K^* = \frac{K}{K_{\max}} = 0.0537288\,.$$

Die Korrelation ist ziemlich klein, also vermutlich kaum Abhängigkeit.

(Lösung in R zu dieser Aufgabe nicht klausurrelevant, nur für Interessierte)

```
Bildung.Arbeitslosigkeit = 1000 * matrix(c(770,13375,140,4000,90,1625),
                                         nrow=3, byrow=T)
dimnames(Bildung.Arbeitslosigkeit) =
  list(Bildung=c("Volksschule", "mittlere Reife", "Abitur"),
       arbeitslos=c("ja", "nein"))

# wird noch nicht als Kontingenztabelle erkannt
Bildung.Arbeitslosigkeit = as.table(Bildung.Arbeitslosigkeit)

# Paket vcd: "visualizing categorical data"
library(vcd)
K = assocstats(Bildung.Arbeitslosigkeit)$cont
```

```
Ergebnisse <- assocstats(Bildung.Arbeitslosigkeit)
chi.quadrat <- Ergebnisse$chisq_tests[2, 1]
chi.quadrat

## [1] 28909.54
```

```r
n <- sum(Bildung.Arbeitslosigkeit)
K <- sqrt(chi.quadrat/(n + chi.quadrat))
K

## [1] 0.03799198

K.max <- sqrt(1/2)
K.normiert <- K/K.max
K.normiert

## [1] 0.05372877

assocstats(Bildung.Arbeitslosigkeit)$cont

## [1] 0.03799198
```

Aufgabe 25

Bei einer Befragung von Passanten in einer Fußgängerzone bezüglich ihres Bierkonsums in Litern pro Woche und ihrer Selbsteinschätzung als Fußballfan ergaben sich folgende Daten:

20 Fußballfans und 120 Nichtfußballfans gaben einen Bierkonsum von höchstens 1 Liter pro Woche an. Zwischen 1 und 3 Liter pro Woche trinken 210 Fußballfans und 200 Nichtfußballfans. 150 Fußballfans und 90 Nichtfußballfans gaben einen Bierkonsum von mindestens 7 Liter an. 145 Fußballfans und 65 Nichtfußballfans lagen in der verbleibenden Zwischengruppe.

a) Stellen Sie die zugehörige Kontingenztabelle auf.
b) Errechnen Sie die Randhäufigkeiten.
c) Berechnen Sie die bedingte Verteilung des Bierkonsums für Fußballfans.
d) Sind Bierkonsum und die Fußballaffinität unabhängig?
e) Berechnen Sie ein geeignetes Zusammenhangsmaß für die beiden Merkmale.

Lösungshinweis:

(Lösung in R nicht prüfungsrelevant, mit Bleistift und Papier schon)

a)
```
Fussball.Bier

##            Fussball
## Bier        Fan kein Fan
##    [0; 1]    20      120
##    (1; 3)   210      200
##    [3; 7)   145       65
##    [7; inf) 150       90
```

b)
```
addmargins(Fussball.Bier)

##            Fussball
## Bier         Fan kein Fan  Sum
##    [0; 1]     20      120  140
##    (1; 3)    210      200  410
##    [3; 7)    145       65  210
##    [7; inf)  150       90  240
##    Sum       525      475 1000
```

c) Bedingte Verteilung der Spalte *Fan* bezogen auf alle Fans (das sind 525):

```
round(Fussball.Bier[, 1]/sum(Fussball.Bier[, 1]), 3)

##    [0; 1]    (1; 3)    [3; 7) [7; inf)
##     0.038     0.400     0.276    0.286
```

20/525   210/525

d) zum Beispiel Bierkonsum [0; 1], Fussball *Fan*: Erwartete Häufigkeit bei Unabhängigkeit:

$$\frac{140 \cdot 525}{1000} = 73.5 \neq 20,$$

also nicht unabhängig.

e) Kontingenzkoeffizient: $K = 0.3210725$

Aufgabe 26

An einer Hochschule sollen die Studierenden ihre Mensa bezüglich der Qualität des Essens beurteilen. In einer Voruntersuchung haben 50 Studenten aus vier Studienjahren befragt ein bestimmtes Gericht bezüglich des Geschmacks als schlecht, mittel bzw. gut bewertet. Folgende Häufigkeitstabelle fasst die Ergebnisse zusammen:

	Studienjahr			
Bewertung	1	2	3	4
schlecht	0	6	3	1
mittel	5	9	1	0
gut	5	5	6	9

Berechnen Sie den normierten Kontingenzkoeffizient zwischen der Zugehörigkeit zum Studienjahr und der vergebenen Bewertung.

Lösungshinweis:

(Lösung in R nicht prüfungsrelevant, mit Bleistift und Papier schon)

```
Mensa = matrix(c(0, 6, 3, 1,
                 5, 9, 1, 0,
                 5, 5, 6, 9), nrow=3, byrow=T)
dimnames(Mensa) =
  list(Bewertung=c("schlecht", "mittel", "gut"),
       Studienjahr=c("1", "2", "3", "4"))

# wird noch nicht als Kontingenztabelle erkannt
Mensa = as.table(Mensa)

# Paket vcd: "visualizing categorical data"
library(vcd)
K = assocstats(Mensa)$cont
K_max = sqrt(2/3)
K_normiert = K/K_max
```

Mit den Randhäufigkeiten

```
addmargins(Mensa)

##           Studienjahr
## Bewertung   1  2  3  4 Sum
##   schlecht  0  6  3  1  10
##   mittel    5  9  1  0  15
##   gut       5  5  6  9  25
##   Sum      10 20 10 10  50
```

ergibt sich $K = 0.5044533$ und die normierte Variante $K^* = 0.6178266$.