Aufgabe 19

An 5 aufeinanderfolgenden Zeitpunkten wurden Preise p und Mengen q zweier Güter G_1 und G_2 festgestellt:

Jahr	1		2		3		4		5	
Gut	p_1	q_1	p_2	q_2	p_3	q_3	p_4	q_4	p_5	q_5
G_1	2,5	5	3	4	2	6	3,5	3	4	2
G_2	3	6	2	10	4	5	4	4	3	10

Betrachtet werden im Folgenden die Berichtszeit 5 und die Basiszeit 1.

a) Berechnen Sie jeweils den Preisindex von Laspeyres, Paasche, Fisher sowie den von Marshall-Edgeworth.

b) Bestimmen Sie jeweils zu Gut 1 bzw. Gut 2 die durchschnittlichen Mengen $\overline{q}^{G_1}, \overline{q}^{G_2}$ der konsumierten Mengen über alle 5 Jahre. Berechnen Sie damit den Preisindex von Lowe.

c) Angenommen die Mengen zu den Zeitpunkten 1 bzw. 5 sind nicht bekannt. Wie hoch ist der maximale bzw. minimale Wert für den Preisindex von Laspeyres?

Lösungshinweis:

a) $P^{\mathrm{L}}_{1,5} = \dfrac{4 \cdot 5 + 3 \cdot 6}{2,5 \cdot 5 + 3 \cdot 6} \approx 1,246\,, \qquad P^{\mathrm{P}}_{1,5} = \dfrac{4 \cdot 2 + 3 \cdot 10}{2,5 \cdot 2 + 3 \cdot 10} \approx 1,068$

$P^{\mathrm{F}}_{1,5} = \sqrt{P^{\mathrm{L}}_{1,5} \cdot P^{\mathrm{P}}_{1,5}} \approx 1,163\,, \qquad P^{\mathrm{ME}}_{1,5} = \dfrac{4 \cdot (5+2) + 3 \cdot (6+10)}{2,5 \cdot (5+2) + 3 \cdot (6+10)} \approx 1,16$

b) Durchschnittliche Mengen:

$$\overline{q}^{G_1} = \frac{1}{5}\sum_{i=1}^{5} q_i^{G_1} = \frac{1}{5}(5+4+6+3+2) = 4$$

$$\overline{q}^{G_2} = \frac{1}{5}\sum_{i=1}^{5} q_i^{G_2} = \frac{1}{5}(6+10+5+4+10) = 7$$

Damit ergibt sich für den Preisindex nach Lowe:

$$P^{\mathrm{LO}}_{1,5} = \frac{4 \cdot 4 + 3 \cdot 7}{2,5 \cdot 4 + 3 \cdot 7} \approx 1.194$$

c) „Extrem“-Warenkörbe:

- $q_1 = 1,\ q_2 = 0 \quad \Rightarrow \quad P^{\mathrm{L}}_{1,5} = \frac{4}{2,5} = 1,6$
- $q_1 = 0,\ q_2 = 1 \quad \Rightarrow \quad P^{\mathrm{L}}_{1,5} = \frac{3}{3} = 1,0$

Also: $1,0 \leqq P^{\mathrm{L}}_{1,5} \leqq 1,6$.

Aufgabe 20

Für zwei Güter G_1, G_2 sind in den Jahren $t = 1, \ldots, 5$ folgende Preise x_t (zu G_1) bzw. y_t (zu G_2) pro Mengeneinheit (= ME) registriert worden.

t	1	2	3	4	5	
x_t	2	2	3	3	5	(€)
y_t	4	5	8	7	8	(€)

Im Jahr 1 wurden 9 ME von G_1 und 8 ME von G_2 verbraucht, im Jahr 4 waren dies 12 ME von G_1 und 9 ME von G_2.

- Bei den im folgenden betrachteten Indexzahlen besteht der Warenkorb nur aus G_1 und G_2; als Basisperiode ist stets das Jahr 1 angesetzt.
  a) Offensichtlich sind die Verbrauchsmengen nicht für jedes Jahr bekannt. Für welche $t \in \{2, 3, 4, 5\}$ läßt sich dennoch ein Laspeyres–Preisindex P_{1t}^{L} ermitteln? (Kurze Begründung, aber keine Berechnung erforderlich!)
  b) Für welche $t \in \{2, 3, 4, 5\}$ läßt sich ein Paasche-Preisindex P_{1t}^{P} ermitteln? Berechnen Sie *einen* solchen Indexwert P_{1t}^{P}.
- Nun beschreibt $y = \hat{a} + \hat{b}x$ die Regressionsgerade mit dem Preis von G_1 als unabhängiger und dem Preis von G_2 als abhängiger Variablen.
  c) Berechnen Sie $\hat{a}$ und $\hat{b}$.
  d) Prognostizieren Sie auf Basis dieser Regression den Preis pro ME des Gutes G_2, wenn G_1 pro ME 6 € kostet.

Lösungshinweis:

a) Berechnung von $P^{\mathrm{L}}_{1,t}$ für $t \in \{2, 3, 4, 5\}$ möglich, denn die Mengen im Basisjahr $t = 1$ sind bekannt.

b) Nur $P^{\mathrm{P}}_{1,4}$ kann man berechnen: $P^{\mathrm{P}}_{1,4} = \dfrac{3 \cdot 12 + 7 \cdot 9}{2 \cdot 12 + 4 \cdot 9} \approx 1.65$

c) $y = 2.9 + 1.1666667 \cdot x$

d) $\hat{y} = 2.9 + 1.1666667 \cdot 6 = 9.9$

Aufgabe 27

Bei 8 zufällig ausgewählten Arbeitnehmerinnen und Arbeitnehmern wird die Schuhgröße S und das jährliche Einkommen E erfasst. Es ergeben sich folgende Werte:

i	1	2	3	4	5	6	7	8
S_i	36	44	40	49	38	34	34	45
E_i	41910	53860	37360	73450	46720	39560	20470	69040

a) Berechnen Sie einen geeigneten Korrelationskoeffizienten zwischen Schuhgröße und Einkommen. Interpretieren Sie das Ergebnis.

b) Bestimmen Sie die Regressionsparameter eines linearen Modells, in dem die Höhe des Einkommens in Abhängigkeit von der Schuhgröße beschrieben wird.

c) Wieviel Einkommen würden Sie gemäß diesem Modell bei einer Person wie zum Beispiel Dirk Nowitzki mit Schuhgröße 54 und bei jemanden mit Größe 35 (z.B. Kylie Minogue) erwarten?

d) Zeichnen Sie die Werte zusammen mit der Regressionsgerade in ein geeignetes Koordinatensystem ein.

e) Wie hoch ist der Determinationskoeffizient dieses Modells? Was sagt diese Größe aus?

f) Bewerten Sie das Modell bezüglich Kausalität versus Korrelation und geben Sie eine potentielle latente Variable an.

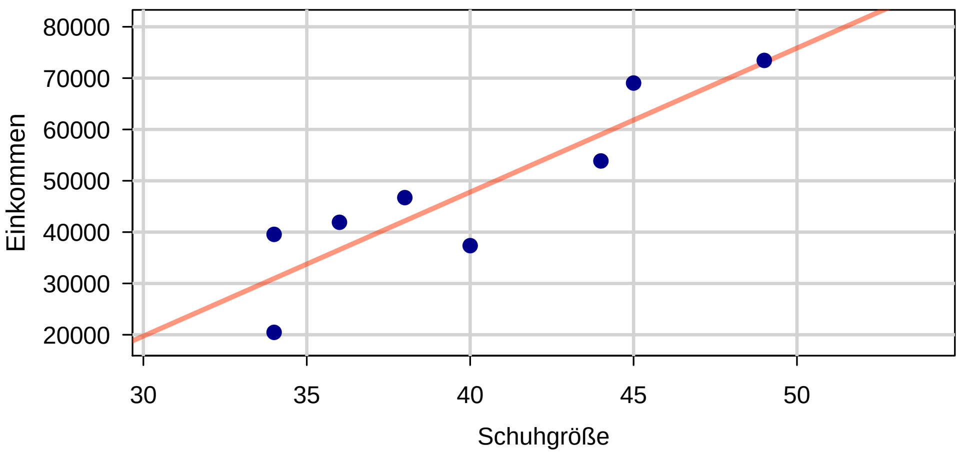
Lösungshinweis:

a) Bravais-Pearson: $r = 0{,}895$.

b) Modell: $E = 2805{,}75 \cdot S + -64400$

c) Prognose: $E(54) = 87076{,}72$, $E(35) = 33767{,}51$

d) Streuplot:



e) Determinationskoeffizient: $R^2 \approx 0{,}8$, also stecken ca. 80 % der Information aus den Daten im Modell.

f) Vermutlich Keine Kausalität im Modell; Eventuelle latente Variable: Geschlecht

Aufgabe 28

In einem Unternehmen fragt man sich, ob zwischen Umsatz und Marketingkosten ein Zusammenhang besteht. Folgende betrieblichen Daten (in 1000 €) liegen vor:

Marketingkosten/Kunde	Umsatz/Kunde
1,4	210
1,8	220
1,9	240
2,4	241
2,8	320
3,2	400
3,6	410
4,0	480

a) Erstellen Sie ein Streuungsdiagramm (y = Umsatz, x = Marketingkosten) und berechnen Sie den Bravais-Pearson- und den Rangkorrelationskoeffizienten.

b) Stellen Sie die Regressionsgerade $\hat{y} = \hat{a} + \hat{b}\,x$ auf und berechnen Sie den Determinationskoeffizienten.

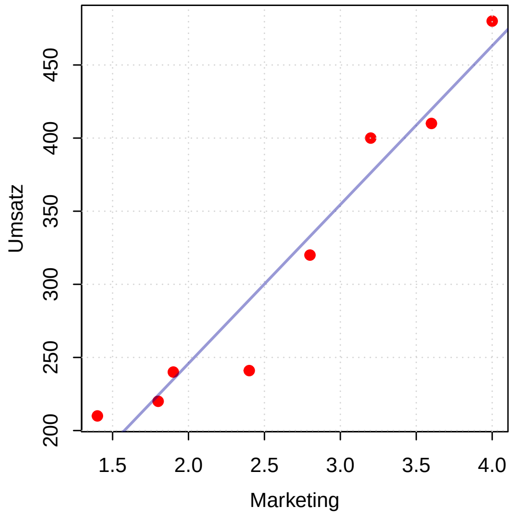
Lösungshinweis:

a)
```
Marketing <- c(1.4, 1.8, 1.9, 2.4, 2.8, 3.2, 3.6, 4)
Umsatz <- c(210, 220, 240, 241, 320, 400, 410, 480)
Regression <- lm(Umsatz ~ Marketing)
plot(Marketing, Umsatz, pch = 20, col = "red", cex = 1.5)
grid()
abline(Regression, col = rgb(0, 0, 0.6, 0.4), lwd = 2)
# Bravais-Pearson:
cor(Marketing, Umsatz)

## [1] 0.9704774

# Rangkorrelation:
cor(Marketing, Umsatz, method = "spearman")

## [1] 1
```



b)
```
a <- Regression$coefficients[1]
b <- Regression$coefficients[2]
```

Modell: $\hat{y} = 28.63 + 108.624 \cdot x$
Determinationskoeffizient: $R^2 \approx 0.9418263$

Aufgabe 29

Von einer Firma sind über mehrere Jahre hinweg die Umsätze und die Beschäftigtenzahlen bekannt:

Jahr t:	1	2	3	4	5	6
Umsatz x_t (in Millionen €):	60	55	57	61	65	62
Anzahl y_t der Beschäftigten:	1000	1100	960	840	800	700

a) Berechnen Sie den Variationskoeffizienten des Umsatzes.
b) Berechnen Sie den Rangkorrelationskoeffizienten von Spearman zwischen den beiden Merkmalen Umsatz und Beschäftigtenzahl.
c) Berechnen Sie die Regressionsgerade $\hat{y} = \hat{a} + \hat{b}t$ der Beschäftigtenzahl in Abhängigkeit von der Zeit. Mit welcher Anzahl der Beschäftigten ist im Jahr 8 zu rechnen?

Lösungshinweis:

a)

```
Umsatz = c(60,55,57,61,65,62)
Beschaeftigte = 100 * c(10, 11, 9.6, 8.4, 8, 7)
m = mean(Umsatz)            # arithm. Mittel
s = sqrt(mean((Umsatz-m)^2)) # Standardabweichung
V = s/m                     # Variationskoeffizient
V

## [1] 0.05443311
```

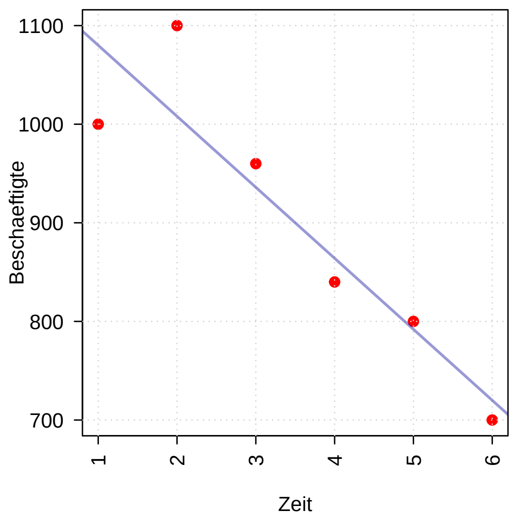
b) Rangkorrelationskoeffizient

```
# Rangkorrelation
cor(Umsatz, Beschaeftigte, method = "spearman")

## [1] -0.8857143
```

c)

```
Zeit <- 1:6
Regression <- lm(Beschaeftigte ~ Zeit)
plot(Zeit, Beschaeftigte, pch = 20, col = "red", cex = 1.5,
    las = 2)
grid()
abline(Regression, col = rgb(0, 0, 0.6, 0.4), lwd = 2)
a <- Regression$coefficients[1]
b <- Regression$coefficients[2]
```



Modell: $\hat{y} = 1152 + -72 \cdot x$
Prognose: $y(8) = 1152 + -72 \cdot 8 = 576$

Aufgabe 30

An 5 aufeinander folgenden Zeitpunkten wurden Preise p und Mengen q eines Gutes festgestellt:

Zeitpunkt	1	2	3	4	5
p	2,5	3	2	3,5	4
q	5	4	6	3	2

a) Berechnen Sie den Rangkorrelationskoeffizienten von Spearman. Welche Vermutung wird durch das Ergebnis nahe gelegt?
b) Bestimmen Sie die Regressionsgerade $\hat{q} = \hat{a} + \hat{b}\,p$.
c) Wie groß ist der Korrelationskoeffizient von Bravais-Pearson?
(Beachten Sie Ihr Ergebnis aus Teil b)!).

Lösungshinweis:

a) Rangkorrelationskoeffizient

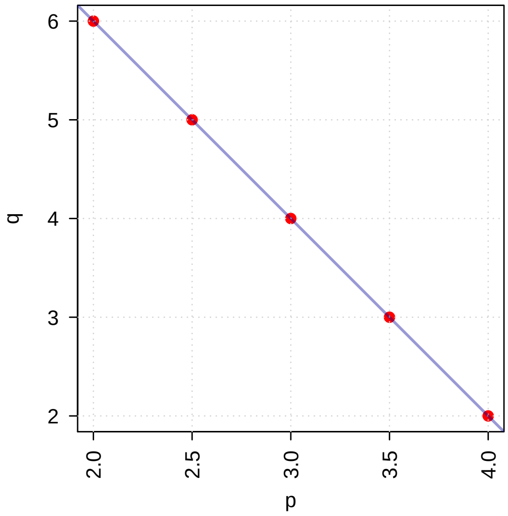
```
p = c(2.5, 3, 2, 3.5, 4)
q = c(5, 4, 6, 3, 2)

# Rangkorrelation
cor(p, q, method="spearman")

## [1] -1
```

Perfekte Rangkorrelation, die Preise und die Mengen haben eindeutig gegenläufige Rangnummern. Vermutlich ist auch der Korrelationskoeffizient von Bravais-Pearson nahe bei -1.

b)
```
Regression <- lm(q ~ p)
plot(p, q, pch = 20, col = "red", cex = 1.5, las = 2)
grid()
abline(Regression, col = rgb(0, 0, 0.6, 0.4), lwd = 2)
a <- Regression$coefficients[1]
b <- Regression$coefficients[2]
```



Modell: $\hat{y} = 10 + -2 \cdot x$
Bravais-Pearson ist $r = -1$.

Aufgabe 31

Zwischen der Anzahl der Besucher eines Freibades und der Tageshöchsttemperatur wird ein Zusammenhang vermutet. Es wurden folgende Daten erhoben:

Tag	Besucheranzahl	Höchsttemperatur (in Grad Celsius)
1	340	35
2	150	25
3	250	28
4	300	32
5	240	26
6	220	28

a) Berechnen Sie ein geeignetes Zusammenhangsmaß zwischen der Besucheranzahl und der Höchsttemperatur.
b) Berechnen Sie die Regressionskoeffizienten der linearen Regression, wenn die Höchsttemperatur als einzige Einflussgröße für die Besucheranzahl erachtet wird.
c) Mit welcher Besucheranzahl ist bei einer Höchsttemperatur von 30° zu rechnen?

Lösungshinweis:

a) Rangkorrelationskoeffizient

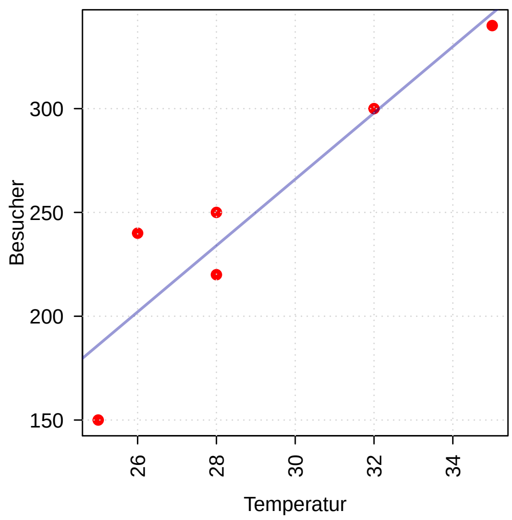
```
Besucher = c(340, 150, 250, 300, 240, 220)
Temperatur = c(35, 25, 28, 32, 26, 28)

# beide metrisch, also Bravais-Pearson
cor(Besucher, Temperatur)

## [1] 0.9221567
```

b)

```
Regression <- lm(Besucher ~ Temperatur)
plot(Temperatur, Besucher, pch = 20, col = "red", cex = 1.5,
    las = 2)
grid()
abline(Regression, col = rgb(0, 0, 0.6, 0.4), lwd = 2)
a <- Regression$coefficients[1]
b <- Regression$coefficients[2]
```



Modell: $\hat{y} = -213.194 + 15.972 \cdot x$

c) Prognose: $\hat{y}(30) = -213.194 + 15.972 \cdot 30 = 265.97$

Aufgabe 32

Das Jahreseinkommen einiger Fußballnationalspieler ist zusammen mit der Anzahl der Tore, die sie in Länderspielen für Deutschland erzielen konnten in folgender Tabelle dargestellt:

Spieler Nummer	Jahreseinkommen [in Mio. €]	Anzahl Tore
1	1,1	0
2	0,8	1
3	2,3	3
4	4,2	3
5	1,7	1
6	0,9	0
7	3,7	4
8	0,7	1
9	2,8	2

a) Stellen Sie ein lineares Regressionmodell der Toranzahl in Abhängigkeit vom Spielereinkommen auf.
b) Geben Sie den Determinationskoeffizienten an und interpretieren Sie ihn.
c) Wieviel Tore würden Sie nach diesem Modell bei einem Einkommen von 10 Mio. erwarten?
d) Wieviel Tore müsste nach diesem Modell ein Spieler mehr schießen, wenn er 1 Mio. € mehr verdient?

Lösungshinweis:

a) Rangkorrelationskoeffizient

```
Einkommen = c(1.1, 0.8, 2.3, 4.2, 1.7, 0.9, 3.7, 0.7, 2.8)
Tore = c(0, 1, 3, 3, 1, 0, 4, 1, 2)

Regression = lm(Tore ~ Einkommen)
a = Regression$coefficients[1]
b = Regression$coefficients[2]
```

Es ergibt sich: $\hat{y} = -0.218 + 0.932 \cdot x$

b)
```
R.Quadrat <- (cor(Tore, Einkommen))^2
```

Determinationskoeffizient: $R^2 = 0.7437987$. Damit sind ca. 74 % der Streuung (Informationsgehalt) der gegebenen Daten durch das Modell erklärbar.

c) Für ein Einkommen von 10 ergibt sich: $\hat{y}(10) \approx -0.218 + 0.932 \cdot 10 \approx 9.103$.

d) Pro Million zusätzlichem Einkommen erhöht sich laut Modell die Toranzahl um $b \approx 0.932$, also fast um 1 Tor.

Aufgabe 33

Zu verschiedenen Zeitpunkten wird der Wasserstand x_t der Isar gemessen.

t	1	2	3	4	5	6
Wasserstand in cm	100	110	?	?	125	120

Die Messwerte zum Zeitpunkt 3 und 4 sind leider verloren gegangen. Es ist jedoch Folgendes bekannt:

$$\sum_{t=1}^{6} x_t = 705 \qquad \text{und} \qquad \sum_{t=1}^{6} x_t^2 = 83425$$

Außerdem ist bekannt, dass der Messwert zum Zeitpunkt 4 größer ist als der Messwert zum Zeitpunkt 3.

a) Ermitteln Sie die fehlenden Messwerte zum Zeitpunkt 3 und 4.
b) Prognostizieren Sie den Wasserstand zum Zeitpunkt $t = 7$ mittels einer linearen Regression.
c) Ermitteln Sie den Determinationskoeffizienten der Regression.
d) Begründen Sie kurz, ob Sie das Vorgehen aus Teilaufgabe b) für sinnvoll erachten.

Lösungshinweis:

a) $$\sum_{i=1}^{6} x_t = 100 + 110 + x_3 + x_4 + 125 + 120 = 705 \quad \Leftrightarrow x_3 + x_4 = 250$$
$$\Leftrightarrow x_4 = 250 - x_3$$
$$\sum_{i=1}^{6} x_t^2 = 100^2 + 110^2 + x_3^2 + x_4^2 + 125^2 + 120^2 = 83425 \quad \Leftrightarrow x_3^2 + (250 - x_3)^2 = 31300$$

$$\Leftrightarrow\ 2x_3^2 - 500x_3 + 31200 = 0$$
$$\Leftrightarrow\ x_{3/4} = \frac{1}{4}\left(500 \pm \sqrt{25000 - 4 \cdot 2 \cdot 31200}\right) = \begin{cases} 120 \\ 130 \end{cases}$$
$$\Leftrightarrow\ x_3 = 120,\ x_4 = 130$$

b) $\hat{x}(t) \approx 102 + 4.429 \cdot x$. Damit ist $\hat{x}(7) \approx 102 + 4.429 \cdot 7 = 133$

c) $R^2 \approx 0.584$.

d) Ist vermutlich nicht sehr sinnvoll, nur den Trend als Grundlage der Prognose heranzuziehen; das Wasser könnte ja auch zyklisch steigen bzw. fallen.

(Lösung in R:)

```
Zeit = 1:6
Wasserstand = c(100,110,120,130,125,120)
Regression = lm(Wasserstand ~ Zeit)
a = Regression$coefficients[1]
b = Regression$coefficients[2]

# Prognose x(7):
x.7 = predict(Regression, data.frame(Zeit=7))

# Determinationskoeffizient:
R.Quadrat = (cor(Zeit, Wasserstand))^2
```

```
plot(Zeit, Wasserstand, pch=20, col="red", cex=1.5, las=2)
grid()
abline(Regression, col=rgb(0,0,0.6,0.4), lwd=2)
```