Verwaltungs- und Wirtschafts-Potsdam e. V

Dipl.-Kfm. Thomas Rochow

Musteraufgaben zur Korrelationsrechnung

Aufgabe 1

Im Rahmen einer gesundheitsstatistischen Untersuchung wurden fünf Kinder bezüglich ihres Alters (in Jahren) und ihres Körpergewichts (in kg) befragt. Das Ergebnis der statistischen Untersuchung ist in nachstehender Tabelle zusammengefasst:

Kind	1	2	3	4	5
X: Alter (in Jahren)	5	9	7	9	7
Y: Körpergewicht (in kg)	30	34	34	32	30

1. Beschreiben Sie den Vorgang der Datenerfassung, d. h. wie zum Zwecke der Daten-erhebung die oben stehenden Daten ermittelt gesammelt wurden!
2. Tragen Sie die erhobenen Daten in ein Streuungsdiagramm ein! Halten Sie den linearen Ansatz für gerechtfertigt? Beurteile die Datenlage
3. Wählen Sie einen geeigneten Korrelationskoeffizienten und berechnen Sie diesen! Interpretieren Sie den erhaltenen Wert!

Bravais-Pearson

Aufgabe 2

Im Rahmen einer psychologischen Untersuchung wurden sieben Personen bezüglich ihres analytischen Fähigkeiten und ihres Sozialverhaltens beurteilt und je Merkmal in ein eindeutiges Ranking gebracht. Das Ergebnis (das Ranking der beiden Merkmale) der psychologischen Untersuchung ist in nachstehender Tabelle zusammengefasst:

Person	A	B	C	D	E	F	G
X: Analytische Fähigkeiten	3	1	4	2	5	7	6
Y: Sozialverhalten	6	5	2	7	4	3	1

1. Beschreiben Sie den Vorgang der Datenerfassung, d. h. wie zum Zwecke der Daten-erhebung die oben stehenden Daten ermittelt gesammelt wurden!
2. Wählen Sie einen geeigneten Korrelationskoeffizienten und berechnen Sie diesen!

Aufgabe 3

Bei neun Studierenden wurden die Merkmale: X: “Schriftliche Note in VWL“ und Y: „Mündliche Note in VWL“ erhoben, wobei nur ganze Noten von 1,0 bis 5,0 vergeben wurden. Das Ergebnis der Ergebung ist in nachstehender Tabelle für Sie zusammengestellt:

Studierende(r)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
X: Schriftliche Note	1	2	2	2	3	3	3	4	5
Y: Mündliche Note	1	1	2	2	1	4	5	4	4

1. Weisen Sie den Merkmalen X und Y Ränge zu!

2. Berechnen und interpretieren Sie den Rangkorrelationskoeffizient von Spearman! r_{SP}

Aufgabe 4
Die Fa. "sauber-chemie" hat es geschafft: In ihrem gentechnischen Labor konnte sie die "A-Bakterie" entwickeln, die erste ölfressende Bakterienart der Welt. Diese will die Firma demnächst bei Tankerunfällen, wie sie leider immer wieder vorkommen, einsetzen. Bedauerlicherweise haben die "A-Bakterien" jedoch einen Feind, der überall dort auftaucht, wo sich die "A-Bakterien" blicken lassen: die "B-Bakterie".

Anhand eines Laborversuchs will die Firma feststellen, wie sich diese beiden Bakterienarten zahlenmäßig fortentwickeln. Dazu setzt sie eine Zahl von "A-Bakterien" in einer Glasschüssel mit Öl aus und registriert mittels einer Messeinrichtung die Zahl der "A-Bakterien" und die Zahl der "B-Bakterien" an den folgenden 10 Tagen:

Sei X: "Anzahl der 'A-Bakterien' (in 1000) in der Glasschüssel am Tag t" und
Y: "Anzahl der 'B-Bakterien' (in 1000) in der Glasschüssel am Tag t"

Tag t	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A-Bakterien	1	2	4	5	6	8	9	10	12	13
B-Bakterien	23	1	2	4	5	6	8	9	10	12

Die Geschäftsleitung will nun wissen, ob ein linearer Zusammenhang zwischen der Anzahl der "A-Bakterien" und der Anzahl der "B-Bakterien" besteht. Sie beauftragt ihre Chefchemiker Dr. Schlau und Dr. Blau.

1. Zeichnen Sie die Daten in ein Streudiagramm ein!
2. Dr. Schlau setzt für derartige Untersuchungen immer den Korrelationskoeffizienten r nach Bravais-Pearson! Berechnen Sie diesen Koeffizienten und interpretieren Sie den erhaltenen Wert!
3. Dr. Blau verlässt sich bei derartigen Aufgabenstellungen meistens auf den Rang-Korrelationskoeffizienten r_{SP} nach Spearman. Berechnen Sie nun diesen Koeffizienten und interpretieren Sie den erhaltenen Wert!
4. Vergleichen Sie die Ergebnisse von Dr. Schlau und Dr. Blau. Welches Ergebnis würden Sie an die Firma "sauber-chemie" weiterleiten. Begründen Sie Ihre Entscheidung auch unter Zuhilfenahme des unter dem ersten Aufgabenpunkt gezeichneten Streudiagramms!

Aufgabenstellung 5
30 Biologen, 35 Chemiker und 35 Physiker, die sich in einem naturwissenschaftlich orientierten Unternehmen beworben haben, wurden nach einem Eignungstest in die Kategorien „geeignet“ und „nicht geeignet“ eingeordnet. Das Ergebnis ist in nachstehender Tabelle dargestellt (Merkmal „Eignung“, Merkmal „Studienabschluss“):

Studien-abschluss / Eignung	Biologen	Chemiker	Physiker	$\sum$
geeignet	14	10	16	**40**
ungeeignet	16	25	19	**60**
$\sum$	**30**	**35**	**35**	**100**

Überprüfen Sie, ob zwischen X und Y ein Zusammenhang besteht. Berechnen Sie dazu X^2 (Chi^2), C *und* C_{korr} ! Interpretieren Sie die beiden letzten Werte!

Aufgabenstellung 6
In einer Untersuchung wurden 600 Frauen und 400 Männer danach befragt, wie sie es mit der Mülltrennung (Kategorien: nie, gelegentlich, immer) halten. Das Ergebnis ist in nachstehender Tabelle für Sie festgehalten:

Müll-trennung / Geschlecht	nie	gelegent-lich	immer	$\sum$
weiblich	125	250	225	**600**
männlich	125	200	75	**400**
$\sum$	**250**	**450**	**300**	**1.000**

Überprüfen Sie, ob zwischen X und Y ein Zusammenhang besteht. Berechnen Sie dazu X^2 (Chi^2), C *und* C_{korr} ! Interpretieren Sie die beiden letzten Werte!

Aufgabenstellung 7

Seien X : "Schlusskurs der Aktie X (in €)" und
Y : "Schlusskurs der Aktie Y (in €)"

Tag	1	2	3	4	5	6	7	8	9
X: Kurs der Aktie X (in €)	7	8	13	10	15	10	12	18	15
Y: Kurs der Aktie Y (in €)	10	9	11	12	13	12	13	14	14

Die Schlusskurse der Aktien X und Y wurden an 9 aufeinander folgenden Börsentagen notiert. Sie wollen nun wissen, ob ein linearer Zusammenhang zwischen dem Schlusskurs der Aktie X und dem Schlusskurs der Aktie y besteht.

1. Zeichnen Sie die Daten in ein Streuungsdiagramm! Beurteilen Sie die Datenlage!
2. Berechnen und interpretieren Sie den Korrelationskoeffizient nach Bravais-Pearson!

Ziel der Korrelationsanalyse

Feststellen der Stärke und der Richtung des linearen Zusammenhangs zweier Größen X und Y.

Interpretationen bestimmter Ausprägungen des Korrelationskoeffizienten

- $r_{xy} = +1$: in der Stichprobe herrscht ein perfekt, positiver linearer Zusammenhang zwischen den Größen X und Y, d. h. alle Punkte der Punktwolke liegen auf **einer gedachten Gerade** mit **positiver** Steigung.

- $r_{xy} = -1$: in der Stichprobe herrscht ein perfekt, negativer linearer Zusammenhang zwischen den Größen X und Y, d. h. alle Punkte der Punktwolke liegen auf **einer gedachten Gerade** mit **negativer** Steigung.

- $r_{xy} = 0$: in der Stichprobe besteht kein linearer Zusammenhang zwischen den Größen X und Y.

- $-1 < r_{xy} < +1$: in der Stichprobe liegt kein **eindeutiger** linearer Zusammenhang vor.

Aber hierfür gibt es durchaus Klassifikationsversuche, durchaus in den Nuancen verschieden, aber in der Tendenz gleich:

- $-1 < r_{xy} < -0{,}7 \;\; bzw. \;\; 0{,}7 < r_{xy} < 1$

  in der Stichprobe herrscht ein starker, negativer linearer bzw. ein starker positiver linearer Zusammenhang zwischen den Größen X und Y,

- $-0{,}7 < r_{xy} < -0{,}3 \;\; bzw. \;\; 0{,}3 < r_{xy} < 0{,}7$

  in der Stichprobe herrscht ein mäßiger, negativer linearer bzw. ein mäßiger positiver linearer Zusammenhang zwischen den Größen X und Y,

- $-0{,}3 < r_{xy} < 0 \;\; bzw. \;\; 0 < r_{xy} < 0{,}3$

  in der Stichprobe herrscht ein schwacher, negativer linearer bzw. ein schwacher positiver linearer Zusammenhang zwischen den Größen X und Y,

- Der Korrelationskoeffizient sollte nur dann berechnet werden, wenn der untersuchte Zusammenhang sinnvoll durch eine lineare (Regressions-)Funktion beschrieben werden kann. Also man sollte sich – falls sinnvoll – zunächst das Streuungsdiagramm anschauen.

- **Für unabhängige Merkmale gilt $r_{xy} = 0$. Aus $r_{xy} = 0$ kann aber nicht auf eine Unabhängigkeit der Merkmale geschlossen werden.**

- Der Korrelationskoeffizient sagt nichts über die **Art des Zusammenhangs** aus. Ergibt sich beispielsweise $r_{xy} = -1$, so können alle Punkte auf einer Gerade liegen, deren Steigung „fast" 0 bzw. deren Steigung „fast" unendlich beträgt.

- Das Vorzeichen des Korrelationskoeffizienten gibt in jedem Fall die Richtung des linearen Zusammenhangs an.

4.2 Korrelation

$\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}$

4.2.1 Korrelationskoeffizient nach Bravais-Pearson

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_i x_i^2 - n\bar{x}^2\right)\left(\sum_i y_i^2 - n\bar{y}^2\right)}}$$

4.2.2 Korrelationskoeffizient nach Spearman

4.2.2.1 ohne Bindungskorrektur

$$r_{Sp} = 1 - \frac{6 \cdot \sum_{i=1}^{n} (R_{x_i} - R_{y_i})^2}{n \cdot (n^2 - 1)},$$

wobei R_{x_i} und R_{y_i} Rangziffern von X und Y darstellen.

4.2.2.2 mit Bindungskorrektur

$$r_{Sp}^* = 1 - \frac{6 \cdot \sum_{i=1}^{n} (R_{x_i} - R_{y_i})^2}{n \cdot (n^2 - 1) - \frac{1}{2} \cdot (G_S + H_S)},$$

wobei R_{x_i} und R_{y_i} Rangziffern von X und Y darstellen;

$G_S = \sum \left(g_i^3 - g_i\right)$ *und* $H_S = \sum \left(h_i^3 - h_i\right)$, wobei g_i die Länge der i-ten Bindung des Merkmals X und h_i die Länge der i-ten Bindung des Merkmals Y.

4.2.3 Kontingenzkoeffizient

nicht normiert

$$C = \sqrt{\frac{\mathrm{X}^2}{\mathrm{X}^2 + n}} \qquad \text{mit} \qquad \mathrm{X}^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(h_{ij}^{\;o} - h_{ij}^{\;e}\right)^2}{h_{ij}^{\;e}} \qquad \text{und} \qquad h_{ij}^{\;e} = \frac{h_{i\bullet} \cdot h_{\bullet j}}{n}$$

und in korrigiert-normierter Form:

$$C_{korr} = C \cdot \sqrt{\frac{C_0}{C_0 - 1}}$$

$$C_0 = \min(k, l),$$

wobei: k : *Anzahl derAusprägungen bei Merkmal X*

l : *Anzahl derAusprägungen bei Merkmal Y*

Aufgabe 1

Im Rahmen einer gesundheitsstatistischen Untersuchung wurden fünf Kinder bezüglich ihres Alters (in Jahren) und ihres Körpergewichts (in kg) befragt. Das Ergebnis der statistischen Untersuchung ist in nachstehender Tabelle zusammengefasst:

Kind	1	2	3	4	5
X: Alter (in Jahren)	5	9	7	9	7
Y: Körpergewicht (in kg)	30	34	34	32	30

1. Beschreiben Sie den Vorgang der Datenerfassung, d. h. wie zum Zwecke der Datenerhebung die oben stehenden Daten ermittelt gesammelt wurden!
2. Tragen Sie die erhobenen Daten in ein Streuungsdiagramm ein! Halten Sie den linearen Ansatz für gerechtfertigt? Beurteile die Datenlage
3. Wählen Sie einen geeigneten Korrelationskoeffizienten und berechnen Sie diesen! Interpretieren Sie den erhaltenen Wert!

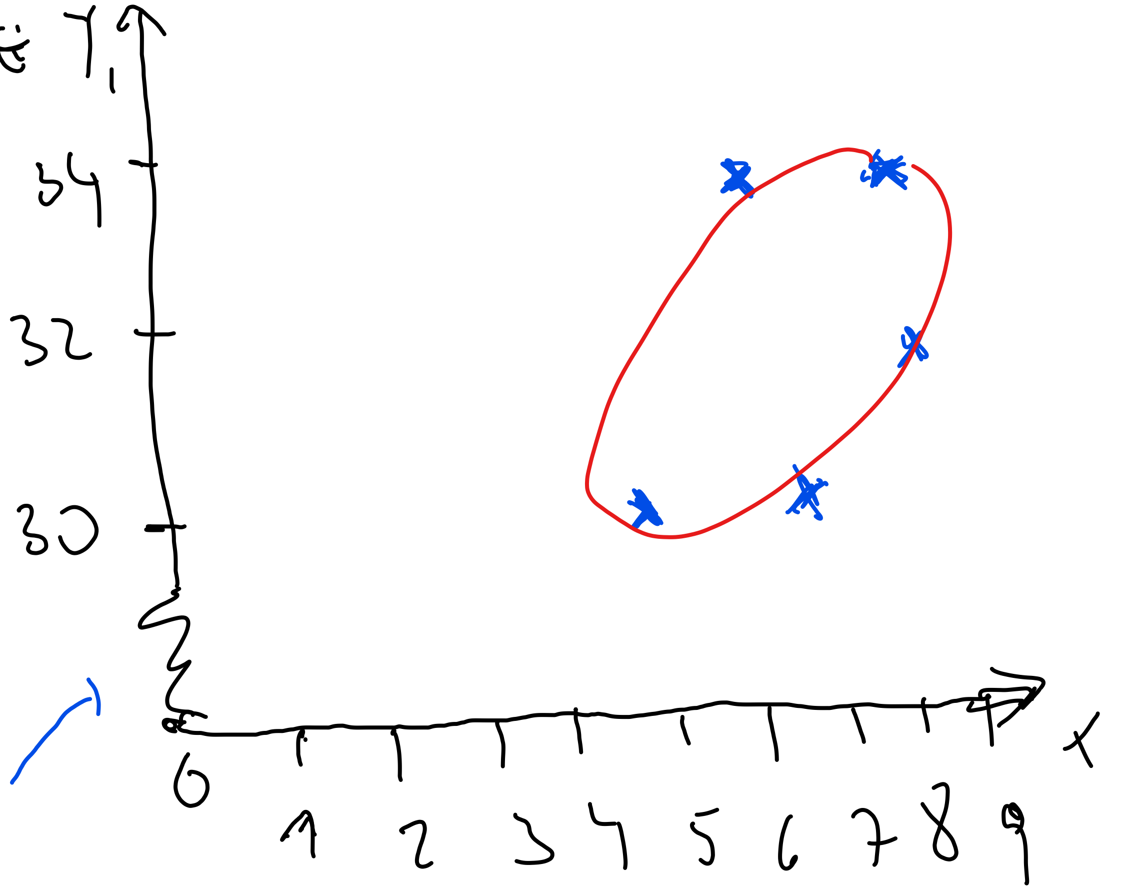
1.) X: Alter (in Jahren)

Y: Gewicht (in kg)

(x_i, y_i) 5 Kinder wurden befragt bezgl. Alter und Gewicht erfasst.

2)

Alter x_i	Gewicht y_i
5	30
9	34
7	34
9	32
7	30



Sägezahn

Achsenunter-
drückung

- linearer Anstieg
  [illegible]
- zu wenige Daten-
  paare

3.) $n = 5$ $\quad \bar{x} = \frac{1}{n} \sum x_i \quad \bar{y} = \frac{1}{n} \sum y_i$

$$\boxed{r = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\left(\sum x_i^2 - n \bar{x}^2\right)\left(\sum y_i^2 - n \bar{y}^2\right)}}}$$

Standardtabelle

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
5	30	150	25	900
9	34	306	81	1156
7	34	238	49	1156
9	32	288	81	1024
7	30	210	49	900
37	160	1.192	285	5.134

$$\bar{x} = \frac{1}{n} \cdot \sum x_i \qquad \bar{y} = \frac{1}{n} \sum y_i$$

$$\bar{x} = \frac{1}{5} \cdot 37 = 7{,}4 \qquad \bar{y} = \frac{1}{5} \cdot 160 = 32$$

$$r = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{\left(\sum x_i^2 - n\bar{x}^2\right)\left(\boxed{\sum y_i^2} - n\bar{y}^2\right)}}$$

$$r = \frac{1192 - 5 \cdot 7{,}4 \cdot 32}{\sqrt{\left(285 - 5 \cdot 7{,}4^2\right)\left(5136 - 5 \cdot 32^2\right)}}$$

$$r = \frac{8}{\sqrt{179{,}2}} = \underline{\underline{0{,}5976}}$$

In der Stichprobe liegt ein mäßiger positiver linearer Zusammenhang zwischen X und Y vor.

Aufgabe 3
Bei neun Studierenden wurden die Merkmale: X: "Schriftliche Note in VWL" und Y: „Mündliche Note in VWL" erhoben, wobei nur ganze Noten von 1,0 bis 5,0 vergeben wurden. Das Ergebnis der Ergebung ist in nachstehender Tabelle für Sie zusammengestellt:

Studierende(r)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
X: Schriftliche Note	1	2	2	2	3	3	3	4	5
Y: Mündliche Note	1	1	2	2	1	4	5	4	4

verbundene Stichprobe

1. Weisen Sie den Merkmalen X und Y Ränge zu! Y nicht kommentieren.
2. Berechnen und interpretieren Sie den Rangkorrelationskoeffizient von Spearman!

r_{sp}: Rangkorrelationskoeffizient nach Spearman

$$r_{sp} = 1 - \frac{6 \cdot \sum (R_{x_i} - R_{y_i})^2}{n \cdot (n^2 - 1)}$$

n: Anzahl der Datenpaare
Stichprobenumfang
hier $n = 9$

↓

<u>Standardtabelle</u>

x_i	y_i	R_{x_i}	R_{y_i}	$(R_{x_i}-R_{y_i})$	$(R_{x_i}-R_{y_i})^2$
1	1	1	2	-1	1
2	1	3	2	1	1
2	2	3	4,5	-1,5	2,25
2	2	3	4,5	-1,5	2,25
3	1	6	2	4	16
3	4	6	7	-1	1
3	5	6	9	-3	9
4	4	8	7	1	1
5	4	9	7	2	4
		$\sum R_{x_i} =$	$\sum R_{y_i}$	0	37,5

NIE NEGATIV

$$\left[\sum R_{x_i} = \sum R_{y_i} = \frac{n}{2}\cdot(n+1)\right] \quad \frac{9}{2}(9+1) = \underline{\underline{45}}$$

Ränge für X

i	1	2	3	4	5	6	7	8	9
x_i sort	1	2	2	2	3	3	3	4	5

R_{x_i}:

$\underline{\underline{1}}$

$$\frac{2+3+4}{3} = \underline{\underline{3}}$$

$$\frac{5+6+7}{3} = \underline{\underline{6}}$$

$$\frac{8}{1} = \underline{\underline{8}}$$

$$\frac{9}{1} = \underline{\underline{9}}$$

Range für y

i	1	2	3	4	5	6	7	8	9
y_i sort	1	1	1	2	2	4	4	4	5

R_{y_i}:

$$\frac{1+2+3}{3} = \underline{\underline{2}}$$

$$\frac{4+5}{2} = \underline{\underline{4{,}5}}$$

$$\frac{6+7+8}{3} = \underline{\underline{7}}$$

$$\frac{9}{1} = \underline{\underline{9}}$$

$$r_{sp} = 1 - \frac{6 \cdot \sum (R_{x_i} - R_{y_i})^2}{n \cdot (n^2 - 1)}$$

$$-1 \leq r_{sp} \leq +1$$

Kategorien:

$r_{sp} = +1$ perfekt

$0{,}7 < r_{sp} < 1$ stark

$0{,}3 \leq r_{sp} \leq 0{,}7$ mäßig

$0 < r_{sp} \leq 0{,}3$ schwach

$r_{sp} = 0$ kein

erstelle linear durch

monoton

$$r_{sp} = 1 - \frac{6 \cdot 37{,}5}{9 \cdot (9^2 - 1)} = \underline{\underline{0{,}6875}}$$

In der Stichprobe liegt ein

mäßiger positiver monotoner

Zusammenhang zwischen X und Y

vor..

Aufgabenstellung 5

30 Biologen, 35 Chemiker und 35 Physiker, die sich in einem naturwissenschaftlich orientierten Unternehmen beworben haben, wurden nach einem Eignungstest in die Kategorien „geeignet“ und „nicht geeignet“ eingeordnet. Das Ergebnis ist in nachstehender Tabelle dargestellt (Merkmal „Eignung“, Merkmal „Studienabschluss“):

Studien-abschluss / Eignung	**Biologen**	**Chemiker**	**Physiker**	$\sum$
geeignet	14	10	16	**40**
ungeeignet	16	25	19	**60**
$\sum$	**30**	**35**	**35**	**100**

Überprüfen Sie, ob zwischen X und Y ein Zusammenhang besteht. Berechnen Sie dazu X^2 (Chi^2), C *und* C_{korr} ! Interpretieren Sie die beiden letzten Werte!