

12 Korrelationskoeffizienten

| | |
|----------|--|
| 12 | Korrelationskoeffizienten |
| 12.1 | Maß-Korrelationskoeffizient r |
| 12.1.1 | Berechnung des Maß-Korrelationskoeffizienten |
| 12.1.1.1 | Grundlagen der Berechnung |
| 12.1.1.2 | Berechnung |
| 12.2 | Rang-Korrelationskoeffizient |
| 12.2.1 | Scoring und Rang-Transformation |
| 12.3 | Übungen |

Bemerkung: Die im Text verwendete Abkürzung TKS bedeutet Tabellenkalkulationssystem.

Korrelationskoeffizienten sind Maßzahlen (Schätzer) für die Stärke und Richtung des Zusammenhangs mehrerer Variabler, bei bivariaten Datensätzen zweier Variabler. Abhängig von der Struktur der zu untersuchenden Daten stehen verschiedene Korrelationskoeffizienten zur Verfügung. Wir werden den Maß-Korrelationskoeffizienten r nach Bravais und Pearson (für Messdaten) und den Rang-Korrelationskoeffizienten r_s nach Spearman (für Rangdaten) behandeln.

12.1 Der Maß-Korrelationskoeffizient r

(= Produkt-Moment-Korrelationskoeffizient, Erklärung des Begriffs siehe weiter unten)

12.1.1 Berechnung des Maß-Korrelationskoeffizienten

Voraussetzungen für die Berechnung:

1. Die Daten beider Variablen müssen stetig und mindestens intervallskaliert (Skala mit willkürlichem Nullpunkt, z.B. Celsiusskala) sein.
2. Beide Variablen müssen mindesten approximiert normalverteilt sein.
3. Die Variablen müssen mindesten approximiert linear korreliert sein.

Ist einer der Punkte nicht gegeben muss gegebenenfalls auf r_s ausgewichen werden. Siehe dort.

Beispiel 1

Wir werden die Daten der Hühnereier von Beispiel 1 Kap.11 untersuchen um festzustellen, ob Masse und Durchmesser stochastisch korreliert sind. Dazu berechnen wir den Maß-Korrelationskoeffizienten r . Die Voraussetzungen betrachten wir als gegeben.

| Nr. | X Masse (g) | Y Durch- messer (mm) | Nr. | X Masse (g) | Y Durch- messer (mm) | Nr. | X Masse (g) | Y Durch- messer (mm) | Nr. | X Masse (g) | Y Durch- messer (mm) |
|-----|-------------------|-------------------------------|-----|-------------------|-------------------------------|-----|-------------------|-------------------------------|-----|-------------------|-------------------------------|
| 1 | 53.3 | 41.2 | 11 | 57.9 | 43.4 | 21 | 60.6 | 43.2 | 31 | 61.9 | 43.9 |
| 2 | 56.0 | 42.0 | 12 | 58.3 | 43.2 | 22 | 60.6 | 43.3 | 32 | 62.0 | 43.5 |
| 3 | 56.1 | 42.7 | 13 | 58.4 | 42.6 | 23 | 60.7 | 43.1 | 33 | 62.1 | 43.1 |
| 4 | 56.3 | 42.2 | 14 | 59.1 | 43.5 | 24 | 61.2 | 43.9 | 34 | 62.2 | 43.9 |
| 5 | 56.6 | 43.0 | 15 | 59.4 | 42.5 | 25 | 61.2 | 44.1 | 35 | 62.3 | 42.6 |
| 6 | 57.0 | 42.9 | 16 | 60.0 | 43.8 | 26 | 61.3 | 44.2 | 36 | 62.3 | 43.6 |
| 7 | 57.2 | 42.0 | 17 | 60.0 | 42.8 | 27 | 61.5 | 43.5 | 37 | 62.4 | 44.3 |
| 8 | 57.4 | 42.3 | 18 | 60.2 | 43.2 | 28 | 61.6 | 42.9 | 38 | 64.0 | 45.0 |
| 9 | 57.7 | 42.1 | 19 | 60.4 | 43.3 | 29 | 61.6 | 43.8 | | | |
| 10 | 57.7 | 42.8 | 20 | 60.5 | 43.5 | 30 | 61.8 | 43.2 | | | |

Tabelle 1

12.1.1.1 Grundlagen der Berechnung

Die Variabilität der Messwerte

Grundlage für die Berechnung des Korrelationskoeffizienten sind die Varianzen der beiden Variablen. Diese, nämlich Durchmesser und Masse der Eier, sind als empirische Daten mit einer Messunsicherheit behaftet. Sie sind beeinflusst durch die Variabilität der Messgröße (interindividuelle Variabilität), die Qualität des Messgerätes und die Sorgfalt des Messenden und somit Zufallsvariable, die um ihren Mittelwert $\bar{x} = 59,8$ g und $\bar{y} = 43,2$ mm streuen. In Abb.1 weist die der Punktwolke angepasste Gerade auf Linearität hin.

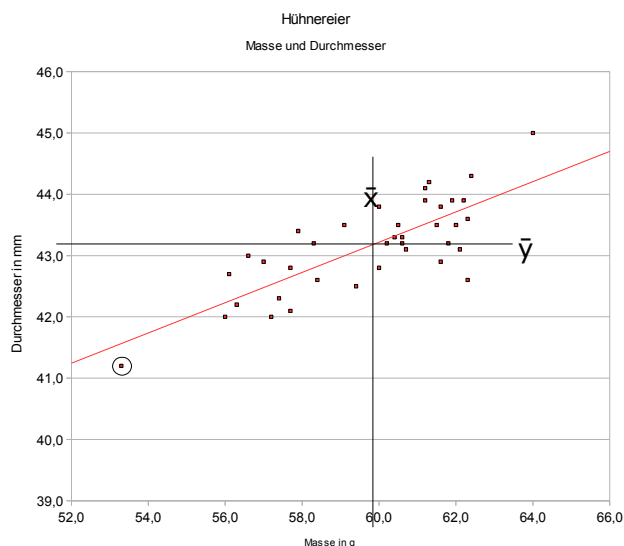


Abb.1

Die beiden Linien in Abb.1, die \bar{x} und \bar{y} mm kennzeichnen, bilden mit der Geraden einen Kreuzungspunkt, den sogenannten Schwerpunkt der Verteilung der bivariaten Daten. Um diesen Punkt streuen die Daten. Die Streuungen der beiden Variablen sind Faktoren, die die Stärke des gesuchten Zusammenhangs beeinflussen.

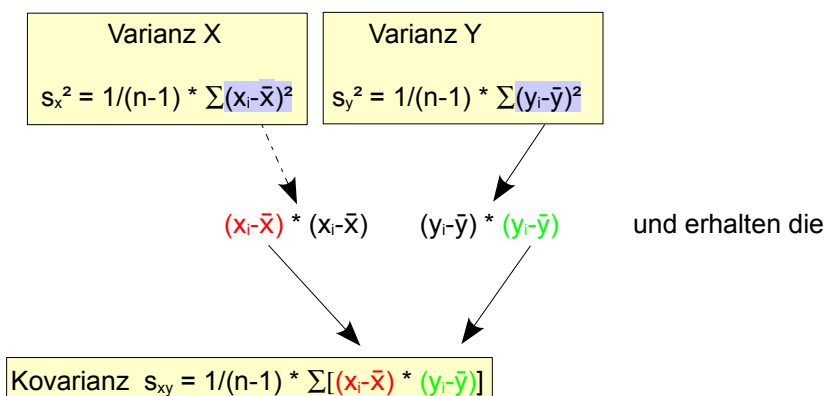
Varianz und Kovarianz

Als Maß für die Streuungen der beiden Variablen können wir deren Varianzen nach

$$s_x^2 = 1/(n-1) * \sum (x_i - \bar{x})^2$$

$$s_y^2 = 1/(n-1) * \sum (y_i - \bar{y})^2$$

berechnen. Die beiden Varianzen und die sogenannte Kovarianz spielen bei der Berechnung des Maß-Korrelationskoeffizienten r eine wesentliche Rolle. Während die Varianz ein Maß für die Streuung der jeweiligen Variablen ist, ist die Kovarianz ein Maß für die gemeinsame Streuung beider Variablen. Sie ist rechnerisch sozusagen ein „Gemisch“ aus den beiden Einzelvarianzen und wird über eine Gleichung berechnet, die Terme aus Gleichungen der beiden Einzelvarianzen enthält. Zur Veranschaulichung der Gleichung für die Kovarianz lösen wir die Abweichungsquadrate der Varianzen auf



Die Varianz wird mathematisch gelegentlich als ein sogenanntes zentrales Moment bezeichnet. (Sachs, Hedderich, Angewandte Statistik 13. Auflage, Springer Verlag, 2009). In diesem Zusammenhang ist der Begriff Produkt-Moment-Korrelationskoeffizient zu verstehen.

Zur Berechnung des Maß-Korrelationskoeffizienten r stehen mehrere äquivalente Gleichungen zur Verfügung. Wir benutzen die folgende.

Aus der Kovarianz und dem geometrischen Mittel der Einzelvarianzen $[\sqrt{(s_x^2 * s_y^2)}]$ berechnen wir den Maß-Korrelationskoeffizienten r nach

$$r = \frac{\text{Kovarianz}}{(\text{Varianz X} * \text{Varianz Y})} = \frac{s_{xy}}{\sqrt{(s_x^2 * s_y^2)}} = \frac{s_{xy}}{s_x * s_y}$$

$$r = \frac{1/(n-1) * \sum(x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{[1/(n-1) * \sum(x_i - \bar{x})^2 * 1/(n-1) * \sum(y_i - \bar{y})^2]}}$$

Mit dem Ziel, die Gleichung übersichtlicher zu gestalten, ändern wir die Reihenfolge der Faktoren des Nenners. Damit wird die Möglichkeit zum Kürzen des Terms $1/(n-1)$ deutlicher.

$$r = \frac{1/(n-1) * \sum(x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{[1/(n-1) * 1/(n-1) * \sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2]}}$$

$$r = \frac{1/(n-1) * \sum(x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{[1/(n-1)]^2 * \sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2}}$$

$$r = \frac{1/(n-1) * \sum(x_i - \bar{x}) * (y_i - \bar{y})}{1/(n-1) * \sqrt{[\sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2]}}$$

Durch Kürzen des Terms $1/(n-1)$ erhalten wir zwei äquivalente übersichtlichere Gleichung für r .

$$\text{Maß-Korrelationskoeffizient } r = \frac{\sum(x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{[\sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2]}} = \frac{\sum(x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} * \sqrt{\sum(y_i - \bar{y})^2}}$$

12.1.1.2 Berechnung des Maß-Korrelationskoeffizienten r

Arbeitstabelle

Es ist zweckmäßig, in einer Arbeitstabelle alle für die Berechnung benötigten Terme zusammenzustellen. Dies ist die eigentliche Rechenarbeit, die wir üblicherweise dem PC überlassen. Aus Gründen der Übersicht haben wir in Tabelle 2 die Daten nur für die Eier 1 bis 5 und 38 dargestellt. Die Zahlen in der letzten Zeile beziehen sich auf die Daten aller 38 Eier.

| Ei-Nr | x_i | y_i | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $y_i - \bar{y}$ | $(y_i - \bar{y})^2$ | $(x_i - \bar{x}) * (y_i - \bar{y})$ |
|----------------------|--------------|--------------|-------------------------|---------------------------|-------------------------|---------------------------|---|
| 1 | 53.3 | 41.2 | -6.46 | 41.70 | -1.96 | 3.84 | 12.66 |
| 2 | 56.0 | 42.0 | -3.76 | 14.12 | -1.16 | 1.35 | 4.36 |
| 3 | 56.1 | 42.7 | -3.66 | 13.38 | -0.46 | 0.21 | 1.68 |
| 4 | 56.3 | 42.2 | -3.46 | 11.96 | -0.96 | 0.92 | 3.32 |
| 5 | 56.6 | 43.0 | -3.16 | 9.97 | -0.16 | 0.03 | 0.51 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 38 | 64.0 | 45.0 | 4.24 | 18.00 | 1.84 | 3.38 | 51.88 |
| Für alle 38 Werte | \bar{x} | \bar{y} | $\Sigma(x_i - \bar{x})$ | $\Sigma(x_i - \bar{x})^2$ | $\Sigma(y_i - \bar{y})$ | $\Sigma(y_i - \bar{y})^2$ | $\Sigma(x_i - \bar{x}) * (y_i - \bar{y})$ |
| | 59.76 | 43.16 | 0 | 210.37 | 0 | 20.89 | 51,88 |

Tabelle 2

Nach Einsetzen der Zahlen für die Terme erhalten wir

$$r = \frac{\Sigma(x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{[\Sigma(x_i - \bar{x})^2 * \Sigma(y_i - \bar{y})^2]}}$$

$$r = \frac{51,88}{\sqrt{[210,37 * 20,89]}}$$

$$r = \frac{51,88}{\sqrt{4394,6293}}$$

$$r = \frac{51,88}{66,292}$$

$$r = +0,78259 \rightarrow +0,78$$

In der Regel wird auf zwei Nachkommastellen gerundet. Wenn Sie die Rechnung, ausgehend von den Versuchsergebnissen überprüfen wollen, dann müssen Sie die entsprechenden Terme für die Eier 6 bis 37 berechnen. Mit einem TKS erhalten wir $r = 0,78253$

Interpretation:

Mit der Zahl $r = +0,78$ haben wir eine quantitative Aussage zur Stärke und zur Richtung der Korrelation. Gelegentlich finden wir in der Literatur unverbindliche Zuordnungen ordinaler Prädikate zu Zahlenwerten für r . In Anlehnung an *Hagl, Schnelleinstieg Statistik, Haufe-Mediengruppe, 2008* nutzen wir folgende Zusammenstellung, die sich auf eine lineare Korrelation bezieht.

| | |
|------------------------|---------------------------|
| $r = 0$ | keine Korrelation |
| $0 < r \leq 0,2 $ | sehr schwache Korrelation |
| $ 0,2 < r \leq 0,5 $ | schwache Korrelation |
| $ 0,5 < r \leq 0,7 $ | mittlere Korrelation |
| $ 0,7 < r \leq 0,9 $ | starke Korrelation |
| $ 0,9 < r \leq 1 $ | sehr starke Korrelation |

$r = +0,78$ weist also darauf hin, dass Durchmesser und Masse der untersuchten Hühnereier stark positiv korreliert sind. Abb.1 weist auf Linearität hin.

Beachten Sie:

Der Korrelationskoeffizient sagt nichts über Kausalität aus!

Da es sich bei den untersuchten Daten um Stichprobenwerte handelt, haben wir mit r einen Schätzwert für den Korrelationskoeffizienten ρ der Grundgesamtheit ermittelt. Mit einem Signifikanztest kann überprüft werden, mit welcher Wahrscheinlichkeit sich der errechnete Wert von $r = 0$ unterscheidet. Einen solchen Test werden wir an dieser Stelle nicht behandeln.

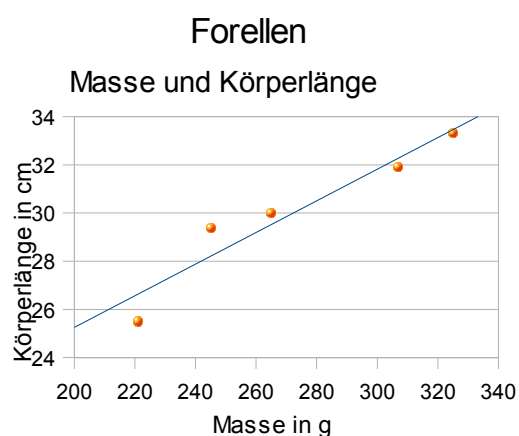
Beispiel 2

Im zoologischen Praktikum haben wir über Jahre bei Forellen die Masse und die Körperlänge gemessen. Damit die folgende Berechnung nicht zu umfangreich wird, wurden für dieses Beispiel aus der großen Zahl von Messwerten fünf Wertepaare at random ausgewählt. Das ist an sich eine sehr kleine Stichprobe, sie soll hier zur Darstellung des Rechenverfahrens aber genügen. Die Frage ist, ob ein stochastischer Zusammenhang zwischen Masse und Körperlänge besteht. Zu berechnen ist r .

Die Daten

| Forelle Nr. | Masse in g | Länge in cm |
|-------------|------------|-------------|
| 82 | 307 | 31,9 |
| 101 | 221 | 25,5 |
| 25 | 265 | 30,0 |
| 69 | 245 | 29,4 |
| 44 | 325 | 33,3 |

Tabelle 3

Abb.2 **Streudiagramm****Voraussetzungen**

1. Die Daten sind intervallskaliert
2. Die Gerade in Abb. 2 deutet auf eine positive lineare Korrelation hin.
3. Wir gehen von approximiert normalverteilten Daten beider Variablen aus.

Berechnung des Korrelationskoeffizienten r **Arbeitstabelle**

| x_i | y_i | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $y_i - \bar{y}$ | $(y_i - \bar{y})^2$ | $(x_i - \bar{x}) * (y_i - \bar{y})$ |
|--------------|--------------|-------------------------|---------------------------|-------------------------|---------------------------|---|
| 307 | 31,9 | 34,4 | 1183,36 | 1,88 | 3,5344 | 64,6720 |
| 221 | 25,5 | -51,6 | 2662,56 | -4,52 | 20,4304 | 233,2320 |
| 265 | 30,0 | -7,60 | 57,76 | -0,02 | 0,0004 | 0,152 |
| 245 | 29,4 | -27,60 | 761,76 | -0,62 | 0,3844 | 17,112 |
| 325 | 33,3 | 52,40 | 2745,76 | 3,28 | 10,7584 | 171,872 |
| \bar{x} | \bar{y} | $\Sigma(x_i - \bar{x})$ | $\Sigma(x_i - \bar{x})^2$ | $\Sigma(y_i - \bar{y})$ | $\Sigma(y_i - \bar{y})^2$ | $\Sigma(x_i - \bar{x}) * (y_i - \bar{y})$ |
| 272,6 | 30,02 | 0 | 7411,20 | 0 | 35,108 | 487,04 |

Tabelle 4

Berechnung

$$r = \frac{\sum(x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{(\sum(x_i - \bar{x})^2 * \sum(y_i - \bar{y})^2)}}$$

$$r = \frac{487,04}{\sqrt{[7411,20 * 35,108]}}$$

$$r = \frac{487,04}{\sqrt{260192,4096}}$$

$$r = \frac{487,04}{510,0906}$$

$$r = +0,9548 \rightarrow +0,95$$

Interpretation

Nach der Berechnung sind Körpermasse und Länge der untersuchten Forellen sehr stark positiv korreliert.

Beachten Sie:

Ein kausaler Zusammenhang ist damit nicht nachgewiesen.

12.2 Der Rang-Korrelationskoeffizient, r_s nach Spearman

Der Rang-Korrelationskoeffizient (r_s) nach Spearman ist wie r für den Bereich $-1 \leq r_s \leq +1$ definiert. Er wurde von Spearman als verteilungs- und parameterfreies Analogon zum Maß-Korrelationskoeffizienten entwickelt. Er wird angewendet bei ordinalskalierten Daten, also qualitativen Daten, die in eine logische Reihenfolge (Rangfolge) gebracht werden können. Gründe für das Vorliegen ordinalskalierter Daten können darin bestehen, dass es für die zu untersuchenden Merkmale keine Einheiten gibt, in denen sie messbar sind. Das trifft z.B. häufig zu bei psychologischen und pädagogischen Untersuchungen zu.

Voraussetzung

1. Die Daten beider Variablen müssen ordinalskaliert sein.
2. Den Daten müssen Rangzahlen zugeordnet werden.
Wenn eine Reihe eines bivariaten Datensatzes ordinalskaliert und die Daten der anderen Reihe quantitativ sind, dann müssen die quantitativen Daten rangtransformiert werden. (Siehe Beispiel 3)
3. Wenn quantitative Daten vorliegen, die nicht normalverteilt sind oder deren Verteilung nicht bekannt ist, dann müssen diese Daten - unter Informationsverlust – rangtransformiert werden.

Nicht vorausgesetzt sind eine bestimmte Verteilung und Linearität.

12.2.1 Scoring und Rangtransformation

Mit Ordinaldaten können wir zunächst nicht rechnen. Wir weisen ihnen in einem ersten Schritt Zahlen zu, die wir scores nennen. Die scores werden dann für die eigentliche Berechnung des Koeffizienten in Rangzahlen transformiert. Quantitative Daten werden direkt rangiert.

| | | | | |
|--------------------|--------|--------|------|-------|
| Ordinaldaten | ---> | scores | ---> | Ränge |
| Quantitative Daten | -----> | Ränge | | |

Erfahrungsgemäß verliert der Ungeübte bei scoring und Rangieren eines größeren Datenbestandes leicht den Überblick. Die Beachtung folgender Hinweise könnte da hilfreich sein.

Scoring

Scoring von Ordinaldaten

Ordinaldaten sind z.B. die Bewertungen des Gesundheitszustandes infizierter Mäuse nach einer Therapie: *leicht erkrankt, agonal, schwer erkrankt, gesund*.

Diesen Ordinaldaten müssen scores zugeordnet werden wozu wir sie **in eine logische Folge** bringen. (Zeile 1 Tab.5). Dann werden ihnen, ihrer Folge entsprechend, die scores zugeordnet. (Zeile 2 Tab.5).

| | | | | | |
|---|--------------|---------------|------------------------|------------------------|---------------|
| 1 | Ordinaldaten | <i>gesund</i> | <i>leicht erkrankt</i> | <i>schwer erkrankt</i> | <i>agonal</i> |
| 2 | scores | 1 | 2 | 3 | 4 |

Tabelle.5

Bei einem Versuch mit 10 Mäusen könnte das Ergebnis einer Untersuchung so aussehen:

| | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|----|
| Maus | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| score | 3 | 3 | 2 | 2 | 2 | 4 | 3 | 4 | 1 | 3 |

Tabelle.6

Die scores werden nun rangiert.

Rangtransformation

Rangieren von scores

Wir stellen die Tabelle 6 so um, dass sie **die scores in aufsteigender Folge** enthält. Vorsicht, in der neuen Tabelle müssen die Paarungen „score - Maus Nr.“ beibehalten werden (score 1 bleibt bei Maus 9). Wer Erfahrung im Rangieren hat, der kann das im Folgenden beschriebene Procedere aber auch an der nicht umgestellten Tabelle 6 durchführen.

| | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|----|---|---|
| Maus | 9 | 3 | 4 | 5 | 1 | 2 | 7 | 10 | 6 | 8 |
| score | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 4 |

Tabelle 7

Den sortierten scores werden vom niedrigsten bis zum höchsten Wert Ränge zugeordnet. Score 1 erhält Rang 1. Kommen gleiche scores mehrmals vor, wie score 2 (drei mal), so bezeichnen wir diese drei Werte als verbundene Daten, Bindungen oder **ties** (englisch tie für Kravatte, Binder). Der erste „score 2“ könnte Rang 2 erhalten, der zweite „score 2“ den Rang 3 und der dritte den Rang 4. Üblicherweise wird aber anders verfahren: Gleiche scores erhalten alle das arithmetische Mittel der ihnen zunächst zugeordneten Ränge. Alle drei „score 2“ erhalten also den Rang 3 $[(2+3+4)/3 = 3]$. Auf diesem Wege entsteht Zeile 3 der Tabelle 8.

| | | | | | | | | | | | |
|---|-------|---|------|---|---|------|---|---|----|------|-----|
| | | | ties | | | ties | | | | ties | |
| 1 | Maus | 9 | 3 | 4 | 5 | 1 | 2 | 7 | 10 | 6 | 8 |
| 2 | score | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 4 |
| 3 | Rang | 1 | 3 | 3 | 3 | 6 | 6 | 6 | 8 | 9,5 | 9,5 |

Tabelle.8

Rangieren quantitativer Daten

Wir denken uns eine Gruppe von 12 Mäusen mit folgenden Massen in g. 15, 20, 21, 16, 17, 19, 17, 22, 17, 18, 23, 19. Die quantitativen Daten werden in der gleichen Weise wie die scores rangtransformiert. Dazu

ordnen wir sie in aufsteigender Folge (Zeile 1 Tab.9) und weisen ihnen unter Berücksichtigung von ties wie beim Rangieren der scores beschrieben, Rangzahlen zu.

| | | | | | | | | | | | | | |
|---|-------|----|----|------|----|----|----|------|-----|----|----|----|----|
| | | | | ties | | | | ties | | | | | |
| 1 | Daten | 15 | 16 | 17 | 17 | 17 | 18 | 19 | 19 | 20 | 21 | 22 | 23 |
| 2 | Rang | 1 | 2 | 4 | 4 | 4 | 6 | 7,5 | 7,5 | 9 | 10 | 11 | 12 |

Tabelle 9

Wir wollen das Vorgehen zusammenhängend an Beispiel 3 zeigen.

Beispiel 3

Fünf Wasserlinsen (*Wolffia arrhiza*), mm-kleine Schwimmpflanzen, werden in Petrischalen auf Wasser mit unterschiedlichen Dosen eines Herbizids (Variable X) kultiviert. Die Dosierungen (mg/L) sind 10; 15; 23; 34; 51; 76. Zur Bewertung der Wirkung des Herbizids wird ermittelt, wie groß nach einer festgelegten Kulturzeit die bewachsene Fläche pro Schale (Variable Y) ist. Da die Fläche nicht exakt ausgemessen werden kann, wird der bewachsene Flächenanteil nach den folgenden Kriterien geschätzt. Wir wollen wissen, ob Dosis (X) und Wirkung (Y) korreliert sind. Dazu wird der Rang-Korrelationskoeffizient r_s nach Spearman berechnet.

Die Daten

| X Dosis mg/L | Y Effekt |
|--------------------|-------------------------------------|
| 10 | Platte ca. ganz bewachsen |
| 15 | Platte ca. $\frac{1}{2}$ bewachsen |
| 23 | Platte ca. $\frac{3}{4}$ bewachsene |
| 34 | Platte ca. $\frac{3}{4}$ bewachsen |
| 51 | Platte ca. $\frac{1}{4}$ bewachsene |
| 76 | Platte nicht bewachsen |

Tabelle.10

Die Daten der Variablen Y sind ordinalskaliert (ca. $\frac{1}{4}$ bewachsen), die der Variablen X quantitativ skaliert. Für die Berechnung von r_s müssen die Daten beider Variablen als Rangzahlen vorliegen. Diese erhalten wir wie folgt.

1. Rangtransformation der quantitativen Daten (Dosen, X)

Wir ordnen den Dosen, in steigender Folge sortiert, ihrem Wert entsprechend, Rangzahlen zu.

| X Dosis mg/L | Rang | Y Effekt |
|--------------------|------|------------------------------|
| 10 | 1 | ca. ganz bewachsen |
| 15 | 2 | ca. $\frac{1}{2}$ bewachsen |
| 23 | 3 | ca. $\frac{3}{4}$ bewachsene |
| 34 | 4 | ca. $\frac{3}{4}$ bewachsen |
| 51 | 5 | ca. $\frac{1}{4}$ bewachsene |
| 76 | 6 | nicht bewachsen |

Tabelle 11

2. Den Ordinaldaten (Effekte, Y) scores zuordnen

Dies geschieht auf folgendem Wege: Wir ordnen die Ordinaldaten vom schwächsten Wert zum stärksten Wert und weisen ihnen scores zu. (Spalte 4 Tab.12) (Wir haben die Ordinaldaten hier nur gedanklich geordnet um die Tabelle nicht umstellen zu müssen).

3. Rangtransformation der scores

Die scores werden nun (gedanklich) in aufsteigender Folge sortiert. Dann weisen wir ihnen, ihrer Größe entsprechend, Rangzahlen zu (Spalte 5). Dabei ist zu beachten, dass score **2** zwei mal vorkommt (ties). Diese beiden verbundenen scores erhalten den gleichen Rang (2,5). Für die Berechnung nach der unten stehenden Gleichung ermitteln wir die Werte der Spalten 6,7,8.

Arbeitstabelle

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------|--------|--------------------|-------|--------|-------------|-----------------|-----------------------------|
| X | | Y | | | | | |
| Dosis | Rang X | Effekt | Score | Rang Y | $R_X - R_Y$ | $(R_X - R_Y)^2$ | $\Sigma(R_X - R_Y)^2 = 6,5$ |
| 10 | 1 | ca. ganz bewachsen | 1 | 1 | 0 | 0 | |
| 15 | 2 | ca. ½ bewachsen | 3 | 4 | -2 | 4 | |
| 23 | 3 | ca. ¾ bewachsen | 2 | 2.5 | 0.5 | 0.25 | |
| 34 | 4 | ca. ¾ bewachsen | 2 | 2.5 | 1.5 | 2.25 | |
| 51 | 5 | ca. ¼ bewachsen | 4 | 5 | 0 | 0 | |
| 76 | 6 | nicht bewachsen | 5 | 6 | 0 | 0 | |

Tabelle 12

Berechnung

| |
|--|
| Rang-Korrelationskoeffizient $r_s = 1 - \frac{6 \sum (d_i)^2}{n^3 - n}$ |
|--|

Es bedeuten n = Anzahl der Wertepaare
 d_i = Differenz der Ränge ($R_X - R_Y$ oder $R_Y - R_X$). Das gewählte Rechenverfahren zur Differenzbildung muss für alle Dosen beibehalten werden.

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n}$$

$$r_s = 1 - \frac{6 * 6,5}{216 - 6} = 1 - \frac{39}{210} = 1 - 0,1857 = 0,8143$$

Interpretation

Da der Spearmansche Korrelationskoeffizient genauso zu interpretieren ist wie der Pearsonsche, besteht nach den empirischen Daten mit $r = +0,81$ eine starke Korrelation zwischen der Herbizidkonzentration und der Präparatewirkung (Wachstumshemmung). Eine Graphik mit den Rängen, die schnell erstellt ist, weist auf eine positive lineare Korrelation hin.

Bemerkung

Liegen mehr als 20% aller Ränge gebunden vor, so werden nach *Sachs, Hedderich, Angewandte Statistik, 13. Auflage, Springer Verlag, 2009*, andere Berechnungsverfahren angewendet. In Beispiel 3 liegt diese Einschränkung (>20% der Ränge als ties) vor. Dennoch

haben wir hier auf ein anderes Berechnungsverfahren verzichtet um die Grundform der Berechnung von r_s an diesem Beispiel zeigen zu können. Wenn >20% ties vorliegen, wird r_s durch die hier durchgeführte Rechnung nach *Sachs, Hedderich, Angewandte Statistik*, 13. Auflage, Springer Verlag, 2009 etwas zu hoch geschätzt.

Beispiel 4

In einer Schulklasse soll untersucht werden, ob ein Zusammenhang zwischen den Zeugnisnoten in Biologie und Mathematik besteht. Die folgenden Noten sind fiktiv.

Die Daten

| | | | | | | | | | | | | | | | | | | | | |
|---|------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| 1 | Schüler | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 2 | Bio-Note | 4 | 2 | 4 | 1 | 5 | 2 | 2 | 3 | 3 | 3 | 3 | 4 | 3 | 2 | 3 | 1 | 4 | 4 | 5 |
| 3 | Mathe-Note | 3 | 1 | 5 | 5 | 5 | 2 | 5 | 2 | 2 | 4 | 1 | 2 | 1 | 3 | 2 | 3 | 1 | 4 | 4 |

Tab.13

Arbeitstabelle

Den Noten beider Gruppen werden (gedanklich) rangiert. Dann werden ihnen mit dem kleinsten Wert (Note 1) beginnend unter Beachtung von ties Ränge zugeteilt. Auf diesem Wege erhalten wir die Einträge in Spalte 4 und 5 der Tabelle 14.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---------|----------|------------|----------|------------|-------------|-----------------|-------------------------------------|
| Schüler | Bio-Note | Mathe-Note | Bio-Rang | Mathe-Rang | $R_B - R_M$ | $(R_B - R_M)^2$ | $\Sigma(R_B - R_M)^2 = \Sigma di^2$ |
| 1 | 4 | 3 | 15 | 11 | 4 | 16 | 969.5 |
| 2 | 2 | 1 | 4.5 | 2.5 | 2 | 4 | |
| 3 | 4 | 5 | 15 | 17,5 | -2,5 | 6,25 | |
| 4 | 1 | 5 | 1,5 | 17,5 | -16 | 256 | |
| 5 | 5 | 5 | 18,5 | 17,5 | 1 | 1 | |
| 6 | 2 | 2 | 4,5 | 7 | -2,5 | 6,25 | |
| 7 | 2 | 5 | 4,5 | 17,5 | -13 | 169 | |
| 8 | 3 | 2 | 9,5 | 7 | 2,5 | 6,25 | |
| 9 | 3 | 2 | 9,5 | 7 | 2,5 | 6,25 | |
| 10 | 3 | 4 | 9,5 | 14 | -4,5 | 20,25 | |
| 11 | 3 | 1 | 9,5 | 2,5 | 7 | 49 | |
| 12 | 4 | 2 | 15 | 7 | 8 | 64 | |
| 13 | 3 | 1 | 9,5 | 2,5 | 7 | 49 | |
| 14 | 2 | 3 | 4,5 | 11 | -6,5 | 42,25 | |
| 15 | 3 | 2 | 9,5 | 7 | 2,5 | 6,25 | |
| 16 | 1 | 3 | 1,5 | 11 | -9,5 | 90,25 | |
| 17 | 4 | 1 | 15 | 2,5 | 12,5 | 156,25 | |
| 18 | 4 | 4 | 15 | 14 | 1 | 1 | |
| 19 | 5 | 4 | 18,5 | 14 | 4,5 | 20,25 | |

Tabelle 14

Zur Berechnung des Rangkorrelationskoeffizienten werden die Terme in den Spalten 6 bis 8 Tab.14 ermittelt.

Berechnung

$$r_s = 1 - \frac{6 \sum d_i^2}{n^3 - n} = 1 - \frac{6 \cdot 969,5}{19^3 - 19} = 1 - \frac{5817}{6859 - 19} = 1 - 0,8504 = 0,14956 \text{ runden zu } +0,15$$

Wir haben (bei den fiktiven Daten) eine sehr schwache positive Korrelation gefunden.

12.3 Übungen

Übung 1

Beim Rang-Korrelationskoeffizienten r_s wurde darauf hingewiesen, dass er ein Analogon zum Maß-Korrelationskoeffizienten r sei. Demzufolge kann man r_s auch nach einer Gleichung berechnen, die der in Beispiel 1 für r genannten analog ist. Diese Gleichung lautet

$$\text{Rang-Korrelationskoeffizient } r_s = \frac{\sum (R_{x_i} - \bar{R}_x) \cdot (R_{y_i} - \bar{R}_y)}{\sqrt{[\sum (R_{x_i} - \bar{R}_x)^2 \cdot \sum (R_{y_i} - \bar{R}_y)^2]}}$$

Darin bedeuten

- R_{x_i} = Rang x_i
- \bar{R}_x = arithmetisches Mittel der x-Ränge
- R_{y_i} = Rang y_i
- \bar{R}_y = arithmetisches Mittel der y-Ränge

Wenn Sie Lust haben, dann berechnen Sie r_s für Beispiel 3 nach dieser Gleichung.

Übung 2

Mit dem Verfahren zur Bestimmung kleiner Gasvolumina nach Otto Warburg haben wir bei 10 Larven des Mehlkäfers *Tenebrio molitor* nach deren Wägung den Sauerstoffverbrauch gemessen. Die Ergebnisse finden wir in der Tabelle 15. Berechnen Sie den Maß-Korrelationskoeffizienten r . Zeitlich effektiver ist die Berechnung mit Hilfe eines TKS. Effektiver für die Übung des Umgangs mit Gleichungen ist die Anwendung des Taschenrechners.

| Tenebrio-Larven Masse in mg | Sauerstoffverbrauch in $\mu\text{L/h}$ |
|--------------------------------|---|
| 319 | 214 |
| 220 | 201 |
| 315 | 188 |
| 175 | 158 |
| 270 | 160 |
| 258 | 220 |
| 219 | 195 |
| 135 | 111 |
| 187 | 200 |
| 201 | 185 |

Tabelle 15

Ergebnis Übung 1: $r = +0,81$. Ergebnis Übung 2: $r = +0,59$