



Leonhard Held  
Kaspar Rufibach  
Burkhardt Seifert

# Medizinische Statistik

Konzepte, Methoden, Anwendungen



die transformierten Variablen stark unterschiedliche Varianzen auf, ist nicht klar, wie ein Gruppenunterschied beurteilt werden kann.

Bei abhängigen Beobachtungen, wenn wir beispielsweise lediglich in der Midazolam-Gruppe wissen möchten, ob sich die mittleren Angstwerte an mindestens zwei der insgesamt betrachteten sechs Zeitpunkte unterscheiden, greifen wir entweder auf die ANOVA für wiederholte Messungen bei normalverteilten Daten zurück oder deren nichtparametrisches Analogon, den Friedman-Test. Für die Analyse solcher longitudinaler Daten verweisen wir auf Abschnitt 6.4.

## ZUSAMMENFASSUNG

### Vergleich von mehr als zwei Gruppen

Möchten wir abklären, ob sich die Mittelwerte normalverteilter Daten in mehr als zwei Gruppen voneinander unterscheiden, verwenden wir die Varianzanalyse oder ANOVA. Finden wir Evidenz gegen die Nullhypothese von in allen Gruppen gleichen Mittelwerten, dürfen wir nur behaupten, dass mindestens zwei der Mittelwerte verschieden sind. Wir können jedoch mittels ANOVA nicht feststellen, welche. Dies leisten Post-hoc-Tests; hierbei ist bei Verwendung von Hypothesentests genau zu überlegen, wie für multiples Testen korrigiert werden muss.

Auch bei der Berechnung der Varianzanalyse treffen wir an die Daten einige Voraussetzungen. Diese Voraussetzungen müssen wir im erhaltenen Modell überprüfen, z. B. mittels Q-Q-Plots.

Das nichtparametrische Analogon zur ANOVA ist der Kruskal-Wallis-Test.

## Wahl des passenden Tests

## 5.8

Hat ein Forscher seine wissenschaftliche Fragestellung in eine statistische Nullhypothese übersetzt, muss er den geeigneten statistischen Test für seine Fragestellung bestimmen. Zur Wahl eines Tests müssen die folgenden Fragen beantwortet werden:

1. Welcher Datentyp liegt vor? **stetig**
2. Bei stetigen Daten: sind diese normalverteilt? **approximativ**
3. Wie viele Stichproben liegen vor? **zwei**
4. Bei mehr als einer Stichprobe – sind diese unabhängig oder gepaart?  
**unabhängig**

Wir haben die Fragen für die Hypothese (5.1) beantwortet, und dies führt uns auf den ungepaarten  $t$ -Test zur Beurteilung dieser Nullhypothese, wie in Abschnitt 5.2 beschrieben.

## Signifikanz- und Hypothesentest

## 5.9

Zum Schluss dieses Kapitels möchten wir noch einmal auf die Bezeichnungen „Signifikanztest“ und „Hypothesentest“ und die Unterschiede in den beiden Ansätzen eingehen. Wir nennen einen Test, bei dem eine explizite Entscheidung über Ablehnung oder Nicht-Ablehnung der Nullhypothese getroffen wird, unter Einhaltung der Wahrscheinlichkeit eines Fehlers 1. Art, einen „Hypothesentest“, da wir uns explizit zwischen zwei Hypothesen entscheiden. Der Signifikanztest ist für uns die Beurteilung der Evidenz gegen eine Nullhypothese mittels dem  $p$ -Wert. Diese Bezeichnungen orientieren sich an R.A. Fisher, der als Erster den  $p$ -Wert vorgeschlagen hatte und das Diskutieren der Evidenz gegen  $H_0$  mittels  $p$ -Wert „Signifikanztest“ nannte, siehe Biau et al. (2010, Tabelle 2). Ein „kleiner“  $p$ -Wert war für ihn „signifikant“, ohne dass er diesen jedoch mit einer beliebigen Schranke (z. B. 0,05) im Sinne eines Hypothesentests verglich. Fishers „Signifikanz“ ist folglich nicht in dem in Abschnitt 5.4 verwendeten Sinn zu verstehen!

Dass in Anwendungen der  $p$ -Wert oft kategorisiert oder sogar dichotomisiert wird, und damit die beiden Konzepte Signifikanz- und Hypothesentest vermischt werden, hat wohl auch historische Gründe: Als Fisher den  $p$ -Wert vorschlug, war für viele Tests nicht klar, wie ein solcher numerisch ohne sehr viel Aufwand berechnet werden kann. Die Forscher behelfen sich folglich mit Annäherungen und Abschätzungen. Heute ist dies dank der Verbreitung von Computern jedoch nicht mehr nötig, d. h., wir können leicht einen  $p$ -Wert numerisch bestimmen. Umso wichtiger ist die korrekte Interpretation und das Auseinanderhalten der zwei Testkonzepte.

## Literaturverzeichnis

- ALTMAN, D. G. und BLAND, J. M. (1996). Presentation of numerical data. *British Medical Journal* **312** 572.
- BIAU, D. J., JOLLES, B. M. und PORCHER, R. (2010). P-value and the theory of hypothesis testing: an explanation for new researchers. *Clinical Orthopaedics and Related Research* **468** 885–892.
- BLUME, J. und PEIPERT, J. F. (2003). What your statistician never told you about p-values. *Journal of the American Association of Gynecologic Laparoscopists* **10** 439–444.
- COOK, T. D. und DEMETS, L., DAVID (2008). *Introduction to Statistical Methods for Clinical Trials*. Chapman & Hall/CRC, Boca Raton, FL.
- GOODMAN, S. (2008). A dirty dozen: twelve p-value misconceptions. *Seminars in Hematology* **45** 135–140.
- GOODMAN, S. N. (1993). P values, hypothesis tests, and likelihood: Implications for epidemiology of a neglected historical debate (with commentary and response). *American Journal of Epidemiology* **137** 485–496.

- GOODMAN, S. N. (1999a). Toward evidence-based medical statistics. 1: The P value fallacy. *Annals of Internal Medicine* **130** 995–1004.
- GOODMAN, S. N. (1999b). Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of Internal Medicine* **130** 1005–1013.
- GOODMAN, S. N. (2005). P value. In *Encyclopedia of Biostatistics*, 2nd ed. Wiley, Chichester, 3921–3925.
- HEEREN, T. und D'AGOSTINO, R. (1987). Robustness of the two independent samples *t*-test when applied to ordinal scaled data. *Statistics in Medicine* **6** 79–90.
- HELD, L. (2010). A nomogram for P values. *BMC Medical Research Methodology* **10** 21.
- HÜSLER, J. und ZIMMERMANN, H. (2006). *Statistische Prinzipien für medizinische Projekte*. 4th ed. Hans Huber, Bern.
- KIRKWOOD, B. R. und STERNE, J. A. C. (2003). *Essential Medical Statistics*. Blackwell Publishing, Malden, MA. 2nd ed.
- LUDBROOK, J. und DUDLEY, H. (1998). Why permutation tests are superior to *t* and *F* tests in biomedical research. *American Statistician* **52** 127–132.
- MARTEAU, T. M. und BEKKER, H. (1992). The development of a six-item short-form of the state scale of the Spielberger state-trait anxiety inventory (STAI). *British Journal of Clinical Psychology* **31** 301–306.
- POCOCK, S. J. und WARE, J. H. (2009). Translating statistical findings into plain English. *Lancet* **373** 1926–1928.
- PUHAN, M. A., SUAREZ, A., CASCIO, C. L., ZAHN, A., HEITZ, M. und BRAENDLI, O. (2006). Didgeridoo playing as alternative treatment for obstructive sleep apnoea syndrome: randomised controlled trial. *British Medical Journal* **332** 266–270.
- SELLKE, T., BAYARRI, M. J. und BERGER, J. O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician* **55** 62–71.
- SHAPIRO, S. S. und WILK, M. B. (1965). An analysis of variance test for normality: Complete samples. *Biometrika* **52** 591–611.
- STERNE, J. A. und SMITH, G. D. (2001). Sifting the evidence – what's wrong with significance tests? *British Medical Journal* **322** 226–231.
- WACHOLDER, S., CHANOCK, S., GARCIA-CLOSAS, M., EL GHORMLI, L. und ROTHMAN, N. (2004). Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute* **96** 434–442.
- ZALUNARDO, M. P., IVLEVA-SAUERBORN, A., SEIFERT, B. und SPAHN, D. R. (2010). Quality of premedication and patient satisfaction after premedication with midazolam, clonidine or placebo : Randomized double-blind study with age-adjusted dosage. *Anaesthesist* **59** 410–418.

# Analyse von quantitativen Zielgrößen

6

6.1	Korrelation und linearer Zusammenhang .....	136
6.2	Einfache lineare Regression .....	145
6.3	Multiple lineare Regression .....	156
6.4	Analyse wiederholter Messungen .....	170

ÜBERBLICK

## LERNZIELE

### Analyse von quantitativen Zielgrößen

Die Stärke des Zusammenhangs zwischen stetigen Variablen wird mit dem Korrelationskoeffizienten quantifiziert. Meistens ist man aber an einem zielgerichteten Zusammenhang interessiert. Für diese Fragestellung ist die einfache lineare Regression geeignet. Liegen mehrere erklärende Variablen vor oder möchte man für Confounder adjustieren, kann man eine multiple lineare Regression durchführen. In diesem Kapitel lernt der Leser die entsprechenden Konzepte kennen. Er erfährt, wie die einzelnen Methoden angewandt werden, wie man die Voraussetzungen überprüfen kann und wie die Ergebnisse interpretiert und präsentiert werden.

## Korrelation und linearer Zusammenhang

## 6.1

Bisher diskutierten wir statistische Verfahren, um eine stetige Variable zwischen zwei oder mehreren Gruppen zu vergleichen, d. h., den Zusammenhang zwischen einer stetigen Zielvariablen und einer binären oder nominalen Einflussgröße zu beschreiben. Oft interessiert jedoch auch der Zusammenhang zwischen zwei stetigen Variablen: Wie verhält sich die eine Variable, wenn die andere steigt oder fällt? Ist der Zusammenhang monoton oder sogar linear? Und wie können wir die Stärke und Richtung des Zusammenhangs quantifizieren? Beispiele sind der Zusammenhang zwischen dem Körpergewicht eines Neugeborenen und dem Gewicht der Mutter oder der Zusammenhang von Körpergröße und Körpergewicht einer Person.

### BEISPIEL: HERZVERÄNDERUNGEN BEI FRÜHEREN RADRENNFAHRERN

Professioneller Radrennsport führt zu einem vergrößerten Sportlerherz. Es wird vermutet, dass die Vergrößerung des linken Vorhofs nicht vollständig reversibel ist. In einer prospektiven kontrollierten Studie (Luthi et al., 2008) wurden 62 ehemalige Teilnehmer der Tour de Suisse analysiert, die ihre professionelle Karriere vor 15–49 Jahren beendet haben. Die Größe des linken Vorhofs (in ml pro m<sup>2</sup> Körperoberfläche) wurde mit der von Golfern verglichen und mit der sportlichen Vergangenheit in Zusammenhang gebracht.

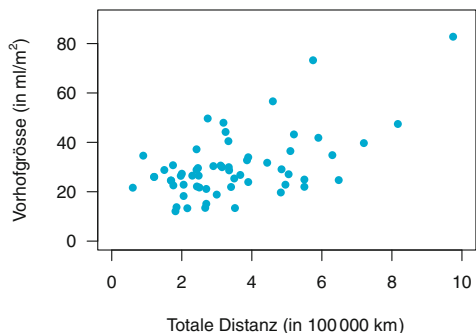


Abbildung 6.1: Streudiagramm der Größe des linken Vorhofs gegen die totale Distanz.

In der Studie der Radrennfahrer könnte man sich für den Zusammenhang zwischen den folgenden Variablen interessieren:

- der Größe des linken Vorhofs und der totalen Distanz, die der Rennfahrer in seinem Leben auf dem Fahrrad zurückgelegt hat,
- der Größe des linken Vorhofs und den beiden Distanzen als Rennfahrer (aktive Distanz) und nach dem Ende der Karriere (passive Distanz), die der Rennfahrer auf dem Fahrrad zurückgelegt hat.

Der **Korrelationskoeffizient** ist eine Möglichkeit, diesen Zusammenhang zu quantifizieren. Wir werden dies im Folgenden am Zusammenhang zwischen der Größe des linken Vorhofs (Vorhofgröße) und der totalen Distanz, die der Rennfahrer in seinem Leben auf dem Fahrrad zurückgelegt hat (totale Distanz), illustrieren.

Möchten wir den Zusammenhang zweier stetiger Variablen beurteilen, ist der erste Schritt das Betrachten eines **Streudiagramms** (engl. scatter plot), auch Punktwolke genannt, siehe Abbildung 6.1. Der Zusammenhang zwischen der Größe des linken Vorhofs und der totalen Distanz scheint nicht allzu stark, aber einigermaßen linear zu sein. Man beachte, dass beide Variablen nicht allzu symmetrisch verteilt sind. Die Beobachtung oben rechts beeinflusst die Korrelation stark. Wir werden diese Probleme später diskutieren.

In Tabelle 6.1 geben wir die wichtigsten deskriptiven Statistiken der beiden Variablen an. Es gilt zu beachten, dass wir Beobachtungen einer Person, für die ein Wert fehlt, ganz eliminieren müssen. Das bedeutet, dass wir zur Berechnung der Korrelation nur paarweise vollständige Beobachtungen verwenden. Da für zwei Studienteilnehmer jeweils ein Wert für die totale Distanz und die Vorhofgröße fehlt, enthält Abbildung 6.1 demnach 60 Punkte.

Variable	<i>n</i>	Min.	Median	$\bar{x}$	Max.	<i>s</i>
Totale Distanz (in 100 000 km)	61	0,6	3,1	3,4	9,8	1,8
Vorhofgröße (in $ml/m^2$ )	61	12,1	27,1	30,2	82,8	13,0

Tabelle 6.1: Deskriptive Statistiken in der Radrennfahrerstudie.



### 6.1.1 Linearer Zusammenhang zwischen zwei stetigen Variablen

Die Stärke des linearen Zusammenhangs zwischen zwei stetigen Variablen  $x$  und  $y$  misst man mit dem **Pearson-Korrelationskoeffizienten** (engl. correlation coefficient) bzw. der **Korrelation** (engl. correlation). Manchmal wird die Korrelation auch als **Produkt-Moment-Korrelation** bezeichnet.

#### DEFINITION 6.1 Pearson-Korrelation

Die Korrelation einer Stichprobe von  $n$  Paaren von Beobachtungen  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  ist

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{s_{xy}}{s_x \cdot s_y}, \quad (6.1)$$

wobei

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{und} \quad s_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (6.2)$$

die geschätzten Standardabweichungen der Variablen  $x$  und  $y$  bezeichnen, und

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

die geschätzte **Kovarianz** zwischen beiden Variablen.

Die Kovarianz  $s_{xy}$  setzt die Abstände der Beobachtungen  $(x_i, y_i)$  der beiden Variablen  $x$  und  $y$  vom **Schwerpunkt**  $(\bar{x}, \bar{y})$  des Datensatzes in Beziehung zueinander. Eine positive Kovarianz bedeutet, dass beide Abstände  $x_i - \bar{x}$  und  $y_i - \bar{y}$  überwiegend gleichsinnig sind, also beide positiv oder beide negativ; dann ist das Produkt  $(x_i - \bar{x})(y_i - \bar{y})$  jeweils positiv. Eine negative Kovarianz bedeutet entsprechend überwiegend gegensinnige Abstände. Der Nenner dient der Normierung. In unserem Beispiel aus Abbildung 6.1 beträgt die Korrelation  $r = 0,56$ .

### 6.1.2 Eigenschaften der Korrelation

Die Korrelation  $r$  schätzt die wahre Korrelation  $\rho$  zwischen den zugrunde liegenden Zufallsvariablen  $X$  und  $Y$ . Die Korrelation zwischen zwei Variablen ist symmetrisch, d. h., die Korrelation zwischen  $x$  und  $y$  ist gleich der Korrelation zwischen  $y$  und  $x$ .

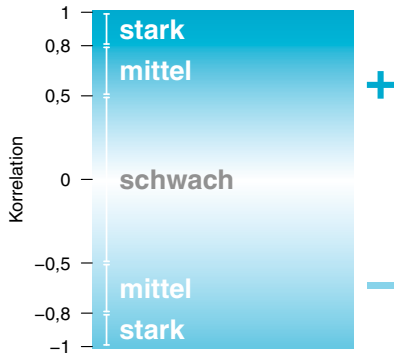


Abbildung 6.2: Die Bewertung von Korrelationskoeffizienten.

#### MERKE

Die Korrelation nimmt Werte zwischen  $-1$  und  $1$  an. Die absolute Größe von  $r$  gibt die Stärke des linearen Zusammenhangs wieder: Je näher  $r$  bei  $1$  oder  $-1$  liegt, desto stärker ist der lineare Zusammenhang zwischen den Variablen. Für  $r = 1$  und  $r = -1$  liegen die Punkte im Streudiagramm exakt auf einer Geraden, siehe dazu auch Abbildung 6.3.

Besteht kein linearer Zusammenhang zwischen  $x$  und  $y$ , so ist  $r$  nahe bei  $0$ . Man spricht dann von unkorrelierten Variablen. Ein Hypothesen- oder Signifikanztest der Nullhypothese einer Korrelation von Null ist aber kaum je von Interesse, da Variablen, die am selben Individuum erhoben werden, immer korrelieren. Bei großen Stichproben werden wir deshalb fast sicher die Nullhypothese einer wahren Korrelation von Null ablehnen (in einem Hypothesen-Test, siehe Abschnitt 5.4). Eine Einteilung in schwache, mittlere und starke Korrelationen ist wünschenswert, aber nicht allgemein möglich und variiert zwischen den Fachgebieten. In den Sozialwissenschaften gilt eine Korrelation von  $0,5$  bereits als stark. In der Medizin können Korrelationen mit Absolutwerten zwischen  $0,1$  und  $0,5$  meistens als schwach, solche zwischen  $0,5$  und  $0,8$  als mittel (moderat) und solche über  $0,8$  als stark bezeichnet werden, siehe Abbildung 6.2.

#### MERKE

Das Vorzeichen von  $r$  gibt die Richtung des Zusammenhangs an: Ist  $r$  positiv, gehen große  $x$ -Werte eher mit großen  $y$ -Werten einher:  $x$  und  $y$  sind **positiv korreliert**. Ist  $r$  negativ, gehen große  $x$ -Werte eher mit kleinen  $y$ -Werten einher, die Variablen sind **negativ korreliert**.

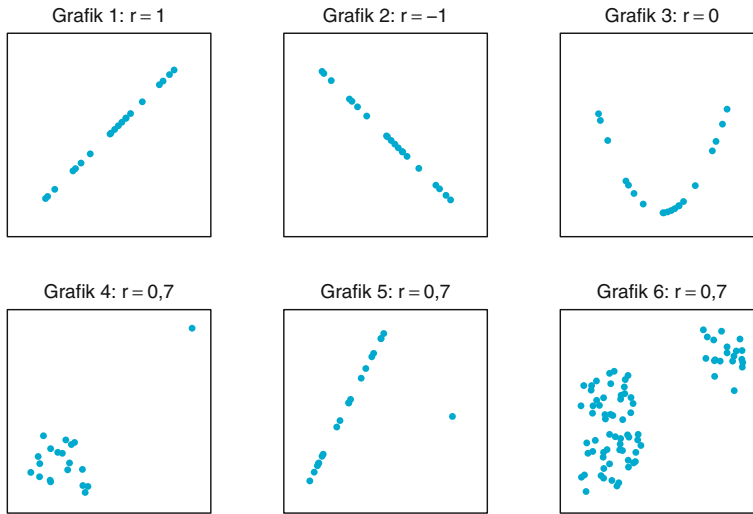


Abbildung 6.3: Illustration von speziellen Korrelationen.

Die Korrelation wird stark beeinflusst von extremen Werten. Sie beschreibt nur den linearen Zusammenhang, ist also bei nichtlinearen Zusammenhängen nicht sinnvoll, siehe dazu auch Abbildung 6.3. Aus diesen beiden Gründen dürfen Korrelationen immer nur im Zusammenhang mit Streudiagrammen der zugrunde liegenden Daten beurteilt werden. In unserem Beispiel aus Abbildung 6.1 beobachten wir mit  $r = 0,56$ , 95 %-Konfidenzintervall von 0,36 bis 0,71, einen moderaten positiven Zusammenhang.

### Vorsicht bei der Interpretation von Korrelationen

Die Korrelation misst den linearen Zusammenhang zwischen zwei stetigen Größen und nichts anderes. Abbildung 6.3 illustriert dies: Die ersten zwei Grafiken zeigen perfekte positive und negative Korrelationen. Die dritte Grafik demonstriert, dass von einer Korrelation von 0 nicht auf nicht-zusammenhängende Variablen geschlossen werden darf: Der Zusammenhang ist nur nicht linear, sondern in diesem Fall deterministisch quadratisch. Die vierte Grafik zeigt die Abhängigkeit von extremen Werten: Die Punktwolke links unten hätte eine Korrelation von praktisch 0, doch der einzelne Punkt rechts oben erhöht die Korrelation auf 0,7. Die fünfte Grafik ist ein anderes Beispiel für den Einfluss extremer Werte: Die perfekte Korrelation der Punkte links wird durch den einen Punkt rechts auf 0,7 gedrückt. In der 6. Grafik sieht man ein Beispiel für eine **Heterogenitätskorrelation**. Obwohl die Korrelation in den beiden Untergruppen links unten und rechts oben negativ ist, ist die gemeinsame Korrelation positiv. Man beachte schließlich, dass die drei völlig verschiedenen Situationen in der unteren Zeile von Abbildung 6.3 alle eine Korrelation von 0,7 liefern.

### 6.1.3 Konfidenzintervall und Test für die Korrelation

Die Voraussetzungen, damit die Aussagen in diesem Abschnitt erfüllt sind, lauten:

- Die Beobachtungen  $x_i$  und  $y_i$  sind jeweils gemeinsam normalverteilt.
- Die  $n$  Beobachtungspaare  $(x_i, y_i)$  sind unabhängig verteilt.

Da der Standardfehler des Korrelationskoeffizienten  $r$  vom wahren Korrelationskoeffizienten  $\rho$  abhängt, ist die direkte Konstruktion eines Wald-Konfidenzintervalls für den Korrelationskoeffizienten  $\rho$  gemäß Abschnitt 4.4 nicht empfehlenswert, speziell für Stichprobenumfänge unter 100. Deshalb empfehlen wir die Verwendung von **Fishers Z-Transformation**. Einer der Vorteile dieses Ansatzes ist, dass das resultierende Konfidenzintervall (6.3) im Intervall von  $-1$  bis  $1$  enthalten ist. Die Fisher-Transformierte eines Korrelationskoeffizienten  $r$  berechnet sich gemäß

$$Z(r) = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right).$$

Unter den oben genannten Voraussetzungen ist  $Z(r)$  annähernd normalverteilt mit Mittelwert  $Z(\rho)$  und Standardfehler  $1/\sqrt{n-3}$ . Der Standardfehler ist für den transformierten Korrelationskoeffizienten also unabhängig von der wahren Korrelation. Ein  $\gamma$ -Konfidenzintervall für  $Z(\rho)$  lautet folglich

$$Z(r) - z/\sqrt{n-3} \quad \text{bis} \quad Z(r) + z/\sqrt{n-3}. \quad (6.3)$$

Hierbei bezeichnet  $z$  das  $(\gamma + 1)/2$ -Quantil der Normalverteilung; bei  $\gamma = 0,95$  ist also  $z = z_{0,975} = 1,96$ . Rücktransformation der Endpunkte des Intervalls (6.3) mit der zu  $Z$  inversen Funktion

$$r(Z) = \frac{\exp(2 \cdot Z) - 1}{\exp(2 \cdot Z) + 1}$$

liefert schließlich ein Konfidenzintervall für  $\rho$ . Im Beispiel der Korrelation der Größe des linken Vorhofs mit der totalen gefahrenen Distanz erhalten wir die Werte in Tabelle 6.2.

Mithilfe der Fisher-Transformierten kann man auch approximative Hypothesentests für die Nullhypothesen  $H_0: r = r_0$  für beliebige  $r_0$  durchführen. Für den

Anzahl vollständige Paare	$n = 60$
Korrelationskoeffizient	$r = 0,56$
Transformierte Korrelation	$Z(0,56) = 0,63$
Standardfehler von $Z(0,56)$	$1/\sqrt{n-3} = 0,13$
Konfidenzintervall für $Z(\rho)$	von 0,37 bis 0,89
Konfidenzintervall für $\rho$	von 0,36 bis 0,71

**Tabelle 6.2:** Korrelationsanalyse der Größe des linken Vorhofs gegen die totale Distanz.

Spezialfall  $r_0 = 0$  gibt es einen exakten Test basierend auf der linearen Regression (siehe Abschnitt 6.2.3). Die Testgröße in dem Fall ist

$$t = r \sqrt{\frac{n-2}{1-r^2}}, \quad (6.4)$$

die bei Gültigkeit von  $H_0$  einer  $t$ -Verteilung mit  $n-2$  Freiheitsgraden folgt. In unserem Beispiel erhalten wir  $t = 5,15$ . In einem Signifikanztest ergibt dies einen  $p$ -Wert  $p < 0,0001$  und folglich starke Evidenz gegen  $H_0$ .

### 6.1.4 Spearman-Rangkorrelation

Die Korrelation wird durch extreme Werte stark beeinflusst, siehe Abbildung 6.3. Ein robustes Maß für die Korrelation ist, analog den Rangtests, die **Rangkorrelation**  $r_s$  (Spearman-Rangkorrelationskoeffizient). Zur Berechnung bildet man zuerst für jede Variable separat die Ränge der Daten und berechnet dann die Pearson-Korrelation dieser Ränge. Dadurch wird der Einfluss von Ausreißern reduziert. Zudem kann der Spearman-Korrelationskoeffizient auch für ordinale Variablen berechnet werden, was für die Pearson-Korrelation  $r$  nicht sinnvoll ist. Allerdings ist Spearmans  $r_s$  kein Maß mehr für den linearen, sondern nur noch für den monotonen Zusammenhang der Variablen. Sind die vorliegenden Daten normalverteilt, unterscheiden sich  $r$  und  $r_s$  nur wenig. Ein Konfidenzintervall kann deshalb mittels Fishers  $Z$ -Transformation der Ränge analog zur Pearson-Korrelation berechnet werden (Fieller et al., 1957). Dabei sollte man allerdings beachten, dass diese Approximation nur dann gerechtfertigt ist, wenn die Daten durch eine (unbekannte) monotone Transformation aus einer gemeinsamen Normalverteilung hervorgegangen sind. Ist dies nicht der Fall, wie z. B. in Abbildung 6.3, ist deshalb ein Bootstrap-Konfidenzintervall (siehe Abschnitt 4.4.6) vorzuziehen.

Der Spearman Rangkorrelationskoeffizient der Größe des linken Vorhofs mit der totalen Distanz in Abbildung 6.1 ist  $r_s = 0,40$  mit dem Konfidenzintervall von 0,16 bis 0,59. Dabei sind die Konfidenzintervalle nach Fishers  $Z$ -Transformation und mittels Bootstrap fast identisch. Die Korrelation  $r_s$  ist kleiner als die Pearson-Korrelation  $r$  in Tabelle 6.2. Das Beispiel zeigt analog zu Grafik 4 in Abbildung 6.3 die Abhängigkeit der Pearson-Korrelation von extremen Werten. Man sagt, dass die eine Beobachtung mit maximaler totaler Distanz von fast einer Million Kilometern (rechts oben in Abbildung 6.1) eine große **Hebelkraft** hat. Die Spearman-Rangkorrelation spiegelt deshalb den Gesamtzusammenhang geeigneter wider.

### 6.1.5 Confounding und Scheinkorrelation

Eine dritte Variable beeinflusst möglicherweise sowohl  $x$  als auch  $y$ , sodass ein tatsächlich nicht existierender Zusammenhang vorgetäuscht wird. Eine beobachtete Korrelation zwischen Schuhgröße und schulischen Leistungen bei Kindern ist wohl kaum kausal (d. h., die schulische Leistung von Kindern hat mit ihrer Schuhgröße direkt nichts zu tun). Der Zusammenhang ist durch einen dritten

# Copyright

Daten, Texte, Design und Grafiken dieses eBooks, sowie die eventuell angebotenen eBook-Zusatzdaten sind urheberrechtlich geschützt. Dieses eBook stellen wir lediglich als **persönliche Einzelplatz-Lizenz** zur Verfügung!

Jede andere Verwendung dieses eBooks oder zugehöriger Materialien und Informationen, einschließlich

- der Reproduktion,
- der Weitergabe,
- des Weitervertriebs,
- der Platzierung im Internet, in Intranets, in Extranets,
- der Veränderung,
- des Weiterverkaufs und
- der Veröffentlichung

bedarf der **schriftlichen Genehmigung** des Verlags. Insbesondere ist die Entfernung oder Änderung des vom Verlag vergebenen Passwortschutzes ausdrücklich untersagt!

Bei Fragen zu diesem Thema wenden Sie sich bitte an: [info@pearson.de](mailto:info@pearson.de)

## Zusatzdaten

Möglicherweise liegt dem gedruckten Buch eine CD-ROM mit Zusatzdaten bei. Die Zurverfügungstellung dieser Daten auf unseren Websites ist eine freiwillige Leistung des Verlags. **Der Rechtsweg ist ausgeschlossen.**

## Hinweis

Dieses und viele weitere eBooks können Sie rund um die Uhr und legal auf unserer Website herunterladen:

**<http://ebooks.pearson.de>**