

Was tun, wenn die Daten abhängig sind?

Analyse bei wiederholten Messungen (Repeated Measurements) und Zeitreihen

Ulrike Held

Horten-Zentrum, UniversitätsSpital, Zürich

Einer statistischen Datenanalyse liegt meist die Annahme zugrunde, dass die einzelnen Beobachtungen unabhängig voneinander sind. Falls die Daten aber abhängig sind, ist deren Variabilität in der Regel geringer, als unter der Unabhängigkeitsannahme zu erwarten wäre. Dieser Tatsache muss eine Analysemethode für abhängige Daten Rechnung tragen. Abhängige Daten entstehen z.B. bei wiederholten Messungen eines bestimmten Parameters an ein und demselben Individuum (Gewicht, EEG, Blutdruck ...) unter verschiedenen Versuchsbedingungen, bei der Aufzeichnung von wöchentlichen Eintrittszahlen in ein Krankenhaus, monatlichen Sterberaten an einer Krankheit, täglichen Messungen eines Umweltschadstoffs, aber auch bei Aktienkursen oder der Betrachtung von Arbeitslosenzahlen. Wenn zusätzlich zu der Abhängigkeit noch ein zeitlicher Bezug vorliegt, dann spricht man von Zeitreihendaten. Wir möchten in diesem Artikel erklären, welche graphischen Darstellungsarten und Analysemethoden für abhängige Daten zur Verfügung stehen.

Wiederholte Messungen oder «Repeated Measurements»

Meistens entstehen sog. Repeated-Measurements-Daten aus wiederholten Messungen einer Zielgrösse an einem Individuum, z.B. unter verschiedenen Versuchsbedingungen. Die Variabilität in der Zielgrösse zwischen Individuen ist dann viel grösser als wiederholte Messungen über die Zeit oder über verschiedene Versuchsbedingungen am gleichen Individuum.

Als Beispiel betrachten wir folgende kleine *Cross-over-Studie* (siehe auch: U. Held. Welche Arten von Studiendesigns gibt es und wie werden sie korrekt eingesetzt?¹). Wir haben an 10 Patienten randomisiert zwei verschiedene blutdrucksenkende Medikamente getestet (mit einer genügend langen «Auswaschzeit») und auch die Baseline-Messung der Patienten vor Studienbeginn notiert. Das bedeutet also, unser Datensatz könnte wie in Tabelle 1 dargestellt aussehen.

Wir schauen uns die Daten zunächst an, wie in Abbildung 1 dargestellt.

Möchte man nun den Effekt der Behandlungen gegenüber der Baseline-Messung schätzen, so verwendet man ein lineares Regressionsmodell, mit dem systolischen Blutdruck als Zielgrösse und der Behandlung als Einflussgrösse. Das Ergebnis ist in Tabelle 2 dargestellt.

Aufgrund dieses Resultats würde man nun bei einem vorgegebenen Niveau von 5% einen signifikanten Be-

handlungseffekt von A versus Baseline postulieren, dass der zugehörige p-Wert 0,037 beträgt. Was haben wir allerdings hierbei nicht berücksichtigt? Wir haben ignoriert, dass die Beobachtungen nicht unabhängig waren und jeweils drei Messungen an jedem Patienten durchgeführt wurden. Wir haben also für die ganze Studie anstatt 30 nur 10 Patienten gebraucht. Wenn aber die Daten abhängig sind, dann ist ihre Variabilität kleiner, als bei Unabhängigkeit zu erwarten wäre. Dies kann dann zu unterschätzten *Standardfehlern* und eventuell sogar zu fälschlicherweise signifikanten Effekten führen. Wählen wir nun für dieses Beispiel eine andere, robuste Analysemethode, welche die Schätzung der Standardfehler dafür korrigiert, dass die Daten innerhalb eines Patienten abhängig sind, so erhalten wir folgendes Ergebnis (Tab. 3).

Man sieht nun, dass es keine Evidenz mehr für das Vorliegen eines Behandlungseffekts von A versus Baseline gibt.

Zeitreihenanalyse

Zeitreihen und deren Analyse sind wohl die typischste Variante von abhängigen Daten. Normalerweise sind die Daten über einen (langen) Zeitraum erhoben worden, und die einzelnen Beobachtungen sind regelmässig, also in festen Zeitintervallen, gemessen worden. Wie erwähnt, braucht man spezielle Analysemethoden, da zwei z.B. aufeinander folgende Messungen sich vermutlich ähnlicher sind als zwei zeitlich weit auseinander liegende Messungen. Zusätzlich zu dieser sog. seriellen Korrelation können noch weitere zeitliche Trends und/oder saisonale Muster in den Daten stecken. Das Ziel einer Zeitreihenanalyse kann also sein, die regelmässige oder systematische Variation in den Daten zu beschreiben. Falls man ein geeignetes statistisches (Regressions-)Modell findet, können auch z.B. zeitliche Trends oder andere Effekte geschätzt werden, unter anderem mit dem Ziel, Prognosen für die Zukunft in Abhängigkeit von der Zeit und/oder anderen Einflussgrössen machen zu können. Zunächst jedoch sollte jede Zeitreihe graphisch dargestellt werden, indem man die Beobachtungen gegen die Zeit abträgt. Dieses Vorgehen ermöglicht eine rein deskriptive Annäherung an die Daten.

Wir möchten einen Datensatz von monatlichen Beobachtungen aus Grossbritannien im Zeitraum 1974 bis



Ulrike Held

Die Autorin erklärt, dass er keine Interessenkonflikte im Zusammenhang mit diesem Beitrag hat.

1 Diese Publikation wird im Rahmen der Biostatistik-Reihe zu einem späteren Zeitpunkt im Schweiz Med Forum 2010;10(41) erscheinen.

Tabelle 1. Systolischer Blutdruck in mm Hg.

Patientennummer	Baseline	Behandlung A	Behandlung B
1	144,04	145,92	132,78
2	130,55	102,32	103,93
3	153,03	129,80	155,22
4	165,91	114,47	107,18
5	140,36	111,01	127,86
6	145,52	120,82	107,45
7	120,36	145,62	124,43
8	151,08	103,58	160,31
9	140,67	115,66	145,90
10	124,27	153,11	116,60

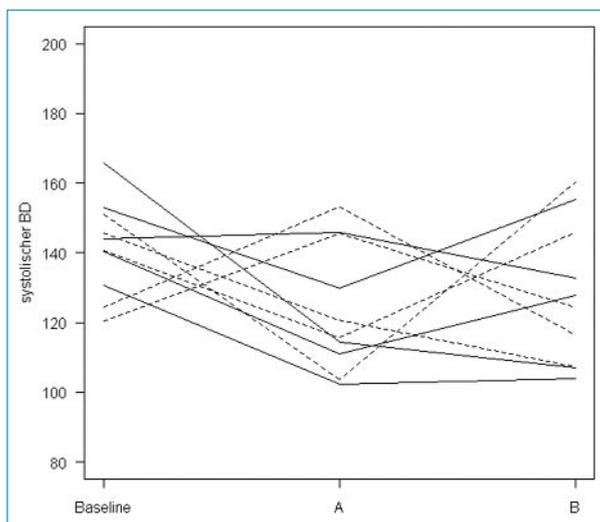


Abbildung 1
Messungen des systolischen Blutdrucks (BD) bei Baseline und unter Behandlung A und B. Durchgezogene Linien haben diejenigen Patienten mit der Behandlungsabfolge (A, B). Bei Behandlungsabfolge (B, A) wurden gestrichelte Linien verwendet. Die Zuteilung war randomisiert (d.h. zufällige Wahl der Behandlungsabfolge).

Tabelle 2. Ergebnis des linearen Regressionsmodells: Einfluss von Behandlung A und B auf den systolischen Blutdruck.

	Effekt	Standardfehler	p-Wert	95%-KI
Behandlung A	-17,35	7,92	0,037	[-33,60; -1,10]
Behandlung B	-13,41	7,92	0,102	[-29,66; 2,84]
Intercept	141,58	5,60	<0,001	[130,09; 153,07]

Tabelle 3. Ergebnis des linearen Regressionsmodells mit robusten Standardfehlern (Sandwich-Varianzschätzer).

	Effekt	Standardfehler (robust)	p-Wert	95%-KI
Behandlung A	-17,35	9,01	0,086	[-37,73; 3,03]
Behandlung B	-13,41	7,15	0,093	[-29,59; 2,76]
Intercept	141,58	4,52	<0,001	[131,36; 151,79]

1979 als Beispiel verwenden, in dem die Todesfälle bei Männern und Frauen an Bronchitis, Emphysem und Asthma gezählt wurden (Diggle, 2000). Man bemerke hier, dass in Abbildung 2 jede Beobachtung die aufaddierten Todesfälle innerhalb eines Monats repräsentiert und nicht eine Einzelmessung zu einem bestimmten Zeitpunkt.

Was ist nun an diesen Zeitreihen zu erkennen? Zunächst sieht man, dass zwischen der Reihe für Männer und der für Frauen ein deutlicher Unterschied in der absoluten Anzahl besteht. Zusätzlich scheint auch eine jährliche (saisonale) Komponente vorzuliegen, die sich bei den Männern und Frauen ähnelt: Im Winter gibt es mehr Todesfälle als im Sommer. Im Winter 1975/76 gab es auffällig viele Fälle in beiden Zeitreihen (Monate 24–26). Um die spezifischen Eigenschaften einer Zeitreihe noch deutlicher zu erkennen, kann man einen gleitenden Durchschnitt verwenden, indem man jeden Punkt durch den Mittelwert seiner Beobachtungen einen Zeitpunkt zuvor, währenddessen und einen Zeitpunkt später repräsentiert. Diese Methode zur Glättung einer Zeitreihe nennt man auch Dreipunktgleitenden Durchschnitt. In Bezug auf unser Beispiel stellt Abbildung 3 die Zeitreihe vollständig geglättet dar.

An dieser geglätteten Darstellung ist nun das saisonale Muster ohne die zufällige Variation der Daten besser zu erkennen. Die Punkte repräsentieren die stärker streuenden Originaldaten.

Nachdem wir nun die saisonale Komponente in den Daten verdeutlicht haben, stellt sich die Frage, ob darüber hinausgehend noch ein zeitlicher Trend vorliegt. Man kann diesen darstellen, indem man einen gleitenden Durchschnitt wählt, in dem jeder Monat des Jahres den gleichen Beitrag liefert. Dadurch wird die jährliche Variation herausgefiltert. Wir haben einen Dreizehnpunktgleitenden Durchschnitt gewählt, bei dem jeder Monat das Gewicht $w_j = 1/12$ erhält, bis auf $w_6 = w_{-6} = 1/24$.

In Abbildung 4 sind die zeitlichen Trends der Männer und Frauen jetzt deutlich zu erkennen, man sieht, dass die absolute Anzahl Todesfälle über die Zeit abnimmt.

Über die graphische Darstellung von Zeitreihen hinausgehend sind die bekanntesten Zeitreihenmodelle die sog. ARMA-Modelle, bei denen man davon ausgeht, dass eine Zeitreihe aus einer Linearkombination früherer Beobachtungen (autoregressiver Teil [AR]) und einer Linearkombination von früheren Zufallsfehlern (moving average [MA]) besteht. Die Todesfalldaten zu Bronchitis, Emphysem und Asthma könnten beispielsweise mit Hilfe eines ARMA-Modells beschrieben werden.

Zeitreihenanalysen haben heutzutage eine grosse praktische Relevanz bei der Beurteilung von Infektionsdaten, die starken saisonalen Schwankungen unterworfen sind. So gibt es komplexe statistische Modelle zur Ausbruchserkennung von z.B. Grippe-Epidemien oder anderen meldepflichtigen Infektionskrankheiten. In diesem Zusammenhang ist eine gute Prognose der Anzahl neuer Fälle von grosser Bedeutung, um z.B. den potentiellen Nutzen von Impfstoffen besser abschätzen zu können.

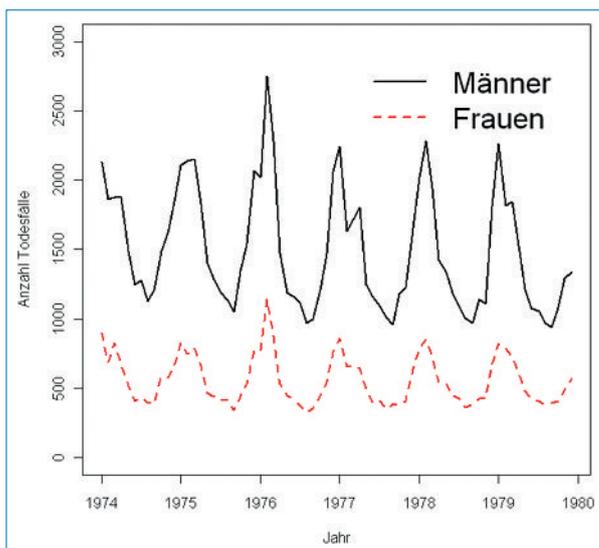


Abbildung 2
Zeitreihen der monatlichen Todesfälle an Bronchitis, Emphysem und Asthma von 1974 bis 1979 in Grossbritannien.

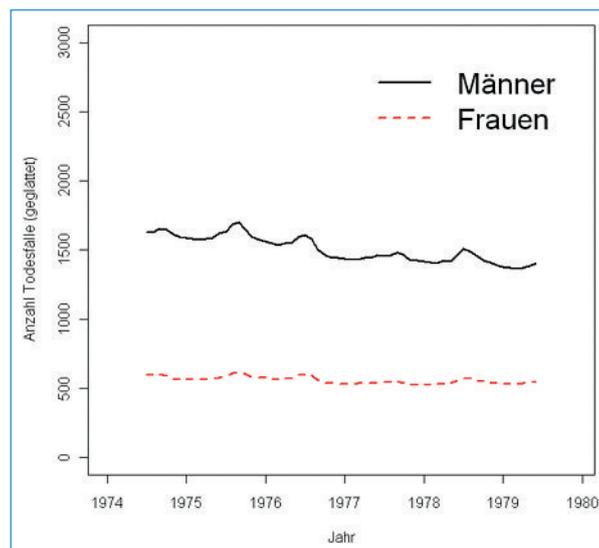


Abbildung 4
Geglättete Zeitreihen: Dreizehnpunkt-gleitender Durchschnitt über die Zeitreihen der Männer und Frauen.

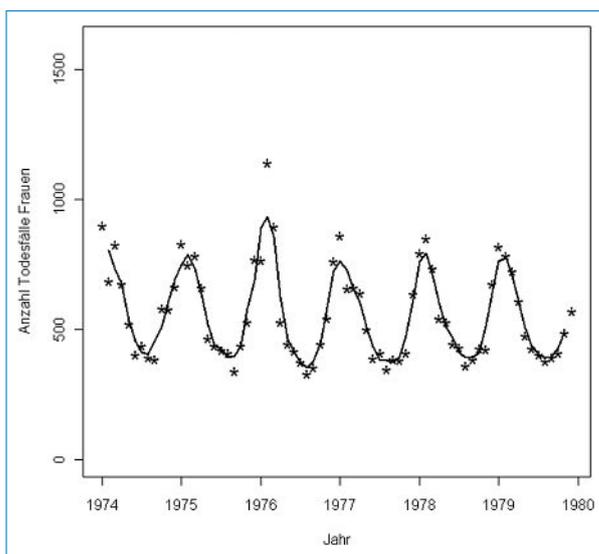


Abbildung 3
Geglättete Zeitreihe: Dreipunkt-gleitender Durchschnitt der Todesfälle bei Frauen.

Glossar

Cross-over-Studie

Studientyp, bei dem jeder Proband mehrere Behandlungen hintereinander erhält. Diese dürfen erst nach genügend langer Wash-out-Phase verabreicht werden, um eine anhaltende Wirkung einer früheren Behandlung auszuschliessen. Die Reihenfolge der Behandlungen sollte randomisiert werden.

Standardfehler

Mit dem Standardfehler bezeichnet man die Standardabweichung des Mittelwerts, und er ist ein Mass für die Streuung einer Stichprobenverteilung. Er wird berechnet als $\sigma_x = \frac{\sigma}{\sqrt{n}}$, wobei n die Stichprobengrösse ist und σ die Varianz der Grundgesamtheit.

Mit Hilfe des Standardfehlers kann die gemessene Standardabweichung zweier Datensätze mit unterschiedlichen Stichprobenumfängen verglichen werden, da die Standardabweichung auf Anzahl Beobachtungen normiert wird.

Korrespondenz:

Dr. rer. nat. Ulrike Held
Horten-Zentrum
UniversitätsSpital Zürich
Postfach Nord
CH-8091 Zürich
ulrike.held@usz.ch

Empfohlene Literatur

- Diggle P. Time Series – A Biostatistical Introduction. Oxford: Oxford University Press; 2000.
- Everitt B. Time Series. In: Encyclopedia of Biostatistics. P. Armitage P, Colton T, Hrsg. 2. Auflage. New York: Wiley; 2005.
- Held L, Rufibach C, Seifert B. Einführung in die Biostatistik. 4. Auflage. Zürich: Abteilung Biostatistik, Institut für Sozial- und Präventivmedizin der Universität Zürich; Juli 2009. <http://www.biostat.usz.ch>.
- Hüsler J, Zimmermann H. Statistische Prinzipien für medizinische Projekte. 4. Auflage. Bern: Huber-Verlag; 2006.
- Kreienbrock L, Schach S. Epidemiologische Methoden. 4. Auflage. München: Elsevier-Verlag; 2005.
- R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008. ISBN 3-900051-07-0. URL <http://www.R-project.org>.
- Williams RL. A note on robust variance estimation for cluster-correlated data. Biometrics. 2000;56:645-6.