

Analyse der Übereinstimmung zweier stetiger Messmethoden bei korrelierten Daten

Thomas Bruckner
Universität Heidelberg
Im Neuenheimer Feld 305
69120 Heidelberg
bruckner@imbi.uni-heidelberg.de

Andreas Deckert
Universität Heidelberg
Im Neuenheimer Feld 364
69120 Heidelberg
a.deckert@uni-heidelberg.de

Lorenz Uhlmann
Universität Heidelberg
Im Neuenheimer Feld 305
69120 Heidelberg
uhlmann@imbi.uni-heidelberg.de

Zusammenfassung

Soll eine neue Messmethode eingeführt werden, so ist eine Bedingung, dass diese eine sehr gute Übereinstimmung mit einer bereits etablierten aufweist. In der Praxis werden typischerweise beide Methoden an denselben Objekten angewendet, um anschließend über statistische Verfahren die Messwerte zu vergleichen. Bei stetigen Daten werden neben Regressions- und Korrelationsanalysen vor allem die bereits etablierten Bland-Altman-Plots [1] verwendet. Über die „Limits of agreement“ kann dann eine Abschätzung der Übereinstimmung anschaulich dargestellt werden. Selbstverständlich ist dieses methodisch relativ einfache Verfahren sehr allgemein anwendbar, sprich in allen Situationen, in denen zwei Messverfahren miteinander verglichen werden sollen.

Problematischer und technisch anspruchsvoller wird die Analyse bei korrelierten Daten, wie z.B. bei Messwiederholungen. Hierbei ist sowohl die Varianz innerhalb der Wiederholungen, als auch die Varianz zwischen den Versuchsobjekten geeignet zu berücksichtigen. Auch für diese Datensituation wurden bereits Verfahren entwickelt [2,4]. In unserem Poster wollen wir die Idee, den theoretischen Hintergrund, sowie die Umsetzung mit SAS und R der für korrelierte Daten angepassten Verfahren präsentieren. Der Fokus liegt dabei auf der praktischen Anwendung. Der Programmcode wird so aufbereitet, dass er möglichst allgemein verwendet werden kann.

Schlüsselwörter: Bland-Altman Plot, korrelierte Daten, Makro %BA_REP

1 Einleitung

Die Güte der Übereinstimmung von zwei unterschiedlichen Messmethoden bei stetigen Daten kann man quantifizieren, indem man die Unterschiede von Messungen beider Methoden, die in einer Stichprobe an derselben Versuchseinheit gemessen wurde, analysiert. Die klassischen Methoden der Korrelations- und Regressionsanalyse sind dabei nicht geeignet, die Güte der Übereinstimmung zu zeigen, denn sie versagen, wenn systematische Verzerrungen vorhanden sind (Abbildung 1).

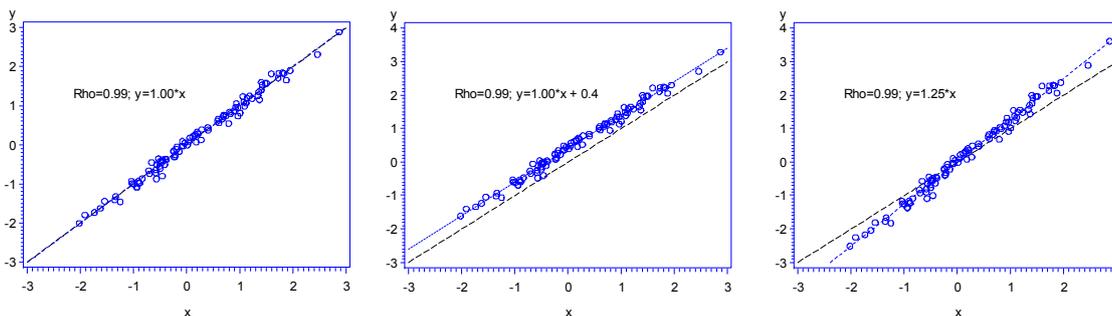


Abbildung 1: Simulierte Daten ohne und mit Verzerrungen

Bland und Altman [1] haben in dieser Situation vorgeschlagen, die Differenz der beiden Messwerte (als Maß der Verzerrung) über den Mittelwert der beiden Messungen (als bestmöglicher Schätzer des wahren Wertes) zu plotten und dabei den Mittelwert aller Differenzen als horizontale Linie sowie als Grenzen der Übereinstimmung („limits of agreement“) $\pm 1.96 *x$ die Standardabweichung der Differenzen ebenfalls als horizontale Linien einzuzichnen.

Dieses Verfahren ist mittlerweile ein Standardverfahren beim Methodenvergleich stetiger Messwerte, und die Anerkennung in der Fachwelt zeigt, dass diese Methode in einem renommierten medizinischen Fachjournal [1] publiziert wurde. Diese Arbeit zählt zu den zehn meist zitierten statistischen Arbeiten [2].

Etwas komplizierter gestaltet sich die Berechnung der „limits of agreement“, wenn es sich um wiederholte Messungen an gleichen Patienten, z.B. im zeitlichen Verlauf, handelt. In diesem Fall hat man eine Mischung aus unabhängigen und abhängigen Beobachtungen. Die Schätzung der Verzerrung kann genauso erfolgen wie im Falle unabhängiger Beobachtungen. Zur Berechnung der Streuung benötigt man allerdings Verfahren der Varianzanalyse [3], denn die Nichtbeachtung der speziellen Varianzstruktur kann zu einer Unterschätzung der empirischen Streuung und damit zu Grenzen der Übereinstimmung kommen, die zu eng sind.

2 Material und Methoden

Das Hauptprinzip der Methode von Bland und Altman ist es, sowohl die mittlere Differenz zwischen den beiden Messmethoden als auch die Variabilität dieser Differenzen zu betrachten. Eine Annahme, die der Methode zugrunde liegt, ist, dass die Differenzen der beiden Methoden stabil über die gesamte Spannweite der Messungen sind.

Hat man korrelierte Beobachtungen, also mehrere Messungen pro Versuchseinheit, so ist die Variabilität nicht mehr einfach als empirische Varianz der Stichprobe zu schätzen. Man muss zwei verschiedene Varianzen bestimmen: Die Varianz der Differenzen der wiederholten Messungen innerhalb einer Versuchseinheit (within) und die Varianz der Differenzen der Mittelwerte der beiden Methoden zwischen den Versuchseinheiten (between). Die gesuchte Gesamtvarianz ist die Summe dieser beiden Varianzen [3]. Die Hauptidee ist, die Varianzen mit Hilfe einer einfaktoriellen Varianzanalyse mit den Differenzen der Messungen als abhängige Variable und die Beobachtungseinheiten als Faktor zu schätzen und daraus die korrigierten Varianzen zu berechnen.

Als Beispieldaten werden die Daten aus [3] genommen, 60 Messungen der Ejektionsfraktion in % an 12 Patienten (3-7 Messwiederholungen) mit zwei unterschiedlichen Messmethoden (Radionuklid Ventikulographie (RV) und Impedanz Kardiographie (IC)).

```
data test;
    input Sub RV IC @@;
cards;
1 7.83 6.57 5 3.13 3.03 9 4.48 3.17
1 7.42 5.62 5 2.98 2.86 9 4.92 3.12
1 7.89 6.90 5 2.85 2.77 9 3.97 2.96
1 7.12 6.57 5 3.17 2.46 10 4.22 4.35
1 7.88 6.35 5 3.09 2.32 10 4.65 4.62
2 6.16 4.06 5 3.12 2.43 10 4.74 3.16
2 7.26 4.29 6 5.92 5.90 10 4.44 3.53
2 6.71 4.26 6 6.42 5.81 10 4.50 3.53
2 6.54 4.09 6 5.92 5.70 11 6.78 7.20
3 4.75 4.71 6 6.27 5.76 11 6.07 6.09
3 5.24 5.50 7 7.13 5.09 11 6.52 7.00
3 4.86 5.08 7 6.62 4.63 11 6.42 7.10
3 4.78 5.02 7 6.58 4.61 11 6.41 7.40
3 6.05 6.01 7 6.93 5.09 11 5.76 6.80
3 5.42 5.67 8 4.54 4.72 12 5.06 4.50
4 4.21 4.14 8 4.81 4.61 12 4.72 4.20
4 3.61 4.20 8 5.11 4.36 12 4.90 3.80
4 3.72 4.61 8 5.29 4.20 12 4.80 3.80
4 3.87 4.68 8 5.39 4.36 12 4.90 4.20
4 3.92 5.04 8 5.57 4.20 12 5.10 4.50
;
```

Abbildung 2 zeigt die Verteilung der beiden Messmethoden als Punktdiagramm.

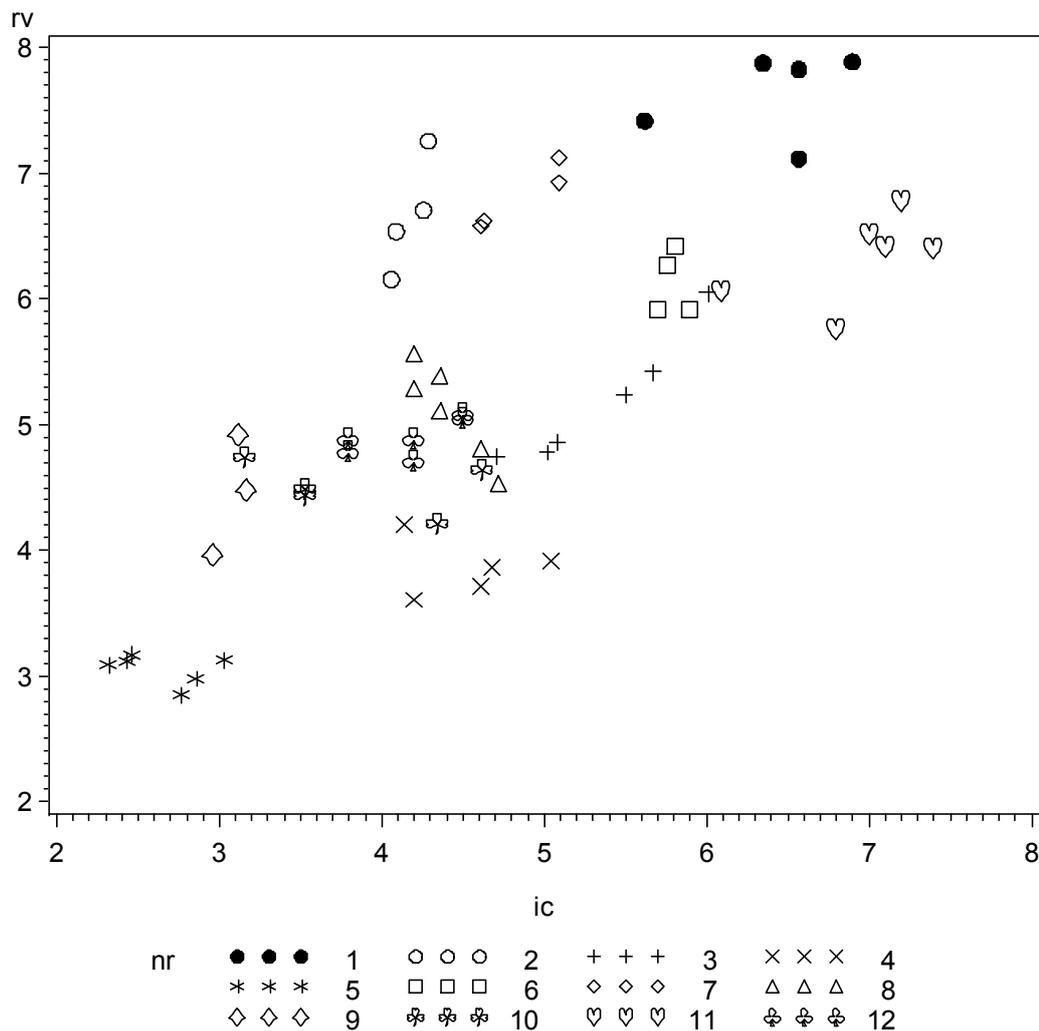


Abbildung 2: Scatterplot der beiden Messungen

Der SAS-Code für die Varianzanalyse ist im Folgenden skizziert, die ODS Output-Anweisung dient dazu, die Ergebnisse der Varianzschätzungen in temporäre SAS-Tabellen zu speichern, um damit weiter rechnen zu können.

```
proc glm data=b outstat=glm;
  ods output modelanova=m fitstatistics=f overallanova=o;
  class id;
  model diff=id ;
run;
quit;
```

Das Ergebnis der Varianzanalyse zeigt folgende Tabelle:

Tabelle 1: Varianzanalysetabelle (SAS 9.1)

Source	D F	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	46.29994667	4.20908606	24.66	<.0001
Error	48	8.19427167	0.17071399		
Corrected Total	59	54.49421833			

Die geschätzte Varianz innerhalb der Versuchseinheiten (within) ist unter der Annahme der Gleichheit über alle Versuchseinheiten gleich dem „Mean square error“, im Beispiel also 0,17071399. Die zweite Varianzkomponente ist die Differenz der mittleren Quadratsumme des Modells und des Fehlerterm, also $4,20908606 - 0,17071399 = 4,03837207$. Dieser Wert muss durch eine Zahl F dividiert werden, die von der Anzahl der Beobachtungen pro Versuchseinheit abhängt. Sei m_i die Anzahl der Beobachtungen von Versuchseinheit i , dann ist F:

$$F = \frac{(\sum m_i)^2 - \sum m_i^2}{(n-1) \sum m_i}$$

In unserem Beispiel ist die Anzahl der Versuchseinheiten $n=12$, die Anzahl der Beobachtungen insgesamt ist $\sum m_i = 60$, und F berechnet sich zu

$$F = 60^2 - 312 / ((12-1) * 60) = 4,9818182.$$

Die geschätzte Varianzkomponente der Varianz zwischen den Versuchseinheiten ist damit $4,03837207 / 4,9818182 = 0,81062213$. Die Gesamtvarianz ist die Summe der beiden Varianzkomponenten, $0,17071399 + 0,81062213 = 0,98133612$, die Standardabweichung errechnet sich zu 0,9906241. Die geschätzte Verzerrung ist der Mittelwert der individuellen Differenzen, 0,60217. Die Grenzen der Übereinstimmung ergeben sich damit zu $0,602 \pm 1,96 * 0,991 = [-1,34; 2,54]$. Diese sind in unserem Beispiel nur wenig größer als die Werte, die man erhält, wenn man die 60 Beobachtungen als unabhängig betrachtet (-1,29 und 2,49).

Diese Analyse wurde, ebenso wie der Fall ohne Messwiederholung von den Autoren als Makro in SAS realisiert. Der Makro-Code ist im Anhang zu finden und auch im deutschsprachigen SAS-Wiki (de.saswiki.org) als SAS-Programm abgelegt.

Auch in R [5] lassen sich diese Ergebnisse über eine Kombination der Funktionen „glm()“ und „aov()“ replizieren und über die plot()-Funktion darstellen.

3 Ergebnisse

Das SAS-Makro %BA_REP verlangt als Eingabedaten die SAS-Datei mit den auszuwertenden Daten, die eindeutige Zuordnung zu den Versuchseinheiten, die beiden Variablen der Ergebnisse der Messmethoden und Minimum, Maximum und Schrittweite der Skalierung der y-Achse (auf der die Differenzen der beiden Methoden abgebildet werden).

Die Ausgabe erfolgt sowohl als Graphik als auch als Listenoutput mit den wichtigsten Ergebnissen. Will man eine permanente Graphik als externe Datei speichern, kann man das z. B. mit der folgenden Goptions-Anweisung und der zugehörigen Filename-Anweisung tun.

```
Goptions reset=all
    device=EMF
    rotate=landscape
    ctext=black
    ftext='Arial'
    htext=0.4 cm
    gsfname=fileout
    hsize=10 cm vsize=9 cm colors=(blue red black )
;

Filename fileout ".\figure5.emf" ;
%ba_rep(_data=a, _id=nr, _xvar=rv, _yvar=ic, _min=-3, _max=+3,
_step=1);
```

Tabellarische Ausgabe im Standard-Output Fenster:

Simple Statistic

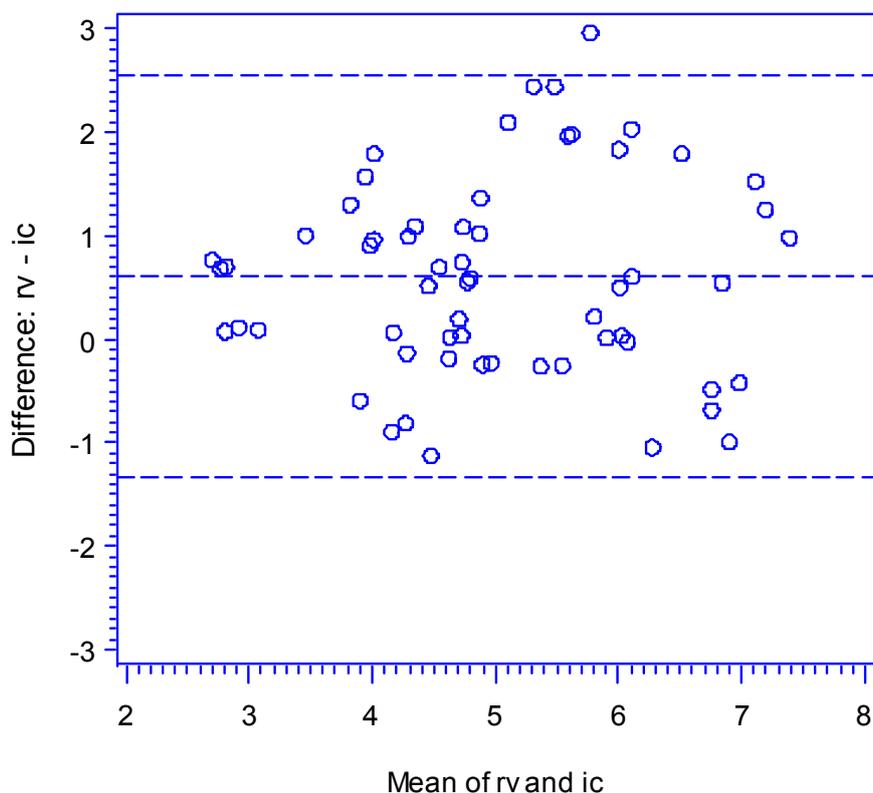
The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Median	Maximum
ic	60	4.72	1.29	2.32	4.61	7.40
rv	60	5.32	1.33	2.85	5.11	7.89
diff	60	0.60	0.96	-1.12	0.58	2.97

Mean Difference, Variance estimates and Limits of Agreement

Obs	Mean Difference	Between Subject Variance	Within Subject Variance	Total Variance	Lower Limit	Upper Limit
1	0.60217	0.81062	0.17071	0.98134	-1.33946	2.54379

Graphische Ausgabe des SAS-Makros



4 Diskussion

Das Makro %BA_REP ist eine sinnvolle Ergänzung zu dem oben genannten Makro für den Fall ohne Messwiederholung (Der Code ist beim Erstautor erhältlich.), das die Bland-Altman Methode bei unabhängigen Beobachtungen mit dem gleichen Graphik-Output auswertet.

Eine weitere Anwendungsmöglichkeit bietet sich bei „geclusterten“ Daten, also bei hierarchischen Datenstrukturen (s. z.B. [6]). Ein Beispiel hierfür wäre die Situation, in der die Messwerte an mehreren Familienmitgliedern erhoben werden. Die Werte der Personen aus derselben Familie sind sich der Erwartung nach ähnlicher, als Personen aus verschiedenen Familien. Diese Korrelationsstruktur kann mit den hier gezeigten Verfahren ebenfalls berücksichtigt werden.

Literatur

- [1] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; i:307-310
- [2] Ryan TP, Woodall WH. The most-cited statistical papers. *J Appl Stat* 2005; 32:461-474
- [3] Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Meth Med Res* 1999; 8:135-160

- [4] Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat.* 2007; 17(4): 571-82.
- [5] R Core Team (2014). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.
URL <http://www.R-project.org/>.
- [6] Snijders T, Bosker R. *Multilevel Analysis – An introduction to basic and advanced multilevel modeling.* London: Sage Publications, 1999.