OPEN ACCESS

Originalarbeit

Smoothed Bootstrap und seine Anwendung in parametrischen Testverfahren

The smoothed bootstrap and its applications in parametrical hypothesis testing

Abstract

In empirical research, the distribution of observations is usually unknown. This creates a problem if parametric methods are to be employed. The functionality of parametric methods relies on strong parametric assumptions. If these are violated the result of using classical parametric methods is questionable. Therefore, modifications of the parametric methods are required, if the appropriateness of their assumptions is in doubt. In this article, a modification of the smoothed bootstrap is presented (using the linear interpolation) to approximate the distribution law suggested by the data. The application of this modification to statistical parametric methods allows taking into account deviations of the observed data distributions from the classical distribution assumptions without changing to other hypotheses, which often is implicit in using nonparametric methods. The approach is based on Monte Carlo method and is presented using one-way ANOVA as an example. The original and the modified statistical methods lead to identical outcomes when the assumptions of the original method are satisfied. For strong violations of the distributional assumptions, the modified version of the method is generally preferable. All procedures have been implemented in SAS. Test characteristics (type 1 error, the operating characteristic curve) of the modified ANOVA are calculated.

Keywords: distributional assumption, resampling, inverse transform sampling, Monte Carlo method, ANOVA

Dmitri Handschuh[1]
Pavel Bobrov[2]

1 Universität Bremen, Kompetenzzentrum für Klinische Studien Bremen (KKSB), Abteilung Biometrie, Bremen, Deutschland

2 Universität Bremen, Institut für Statistik, Sonderforschungsbereich „Mikrokaltumformen“ (SFB747), Bremen, Deutschland

Zusammenfassung

In der empirischen Forschung ist die Verteilung der zu untersuchenden Daten häufig unbekannt. Daraus ergeben sich Schwierigkeiten, wenn aus inhaltlichen Gründen parametrische statistische Methoden zum Einsatz kommen sollen. Die Funktionalität von parametrische Methoden beruht in der klassischen Ausführung auf relativ strengen Verteilungsannahmen. Ist von einer Verletzung dieser Annahmen auszugehen, so ist ein Verfahren in seiner klassischen Form nicht anzuwenden. In diesem Artikel wird ein Prinzip vorgestellt, klassische Methoden so abzuwandeln, dass sie auf vorliegende Daten angewendet werden können, wobei die Form der Abwandlung aus den Daten abgeleitet wird. Diese Modifikation stellt eine spezielle Variante des smoothed bootstrap dar. Die Anwendung dieser Modifikation auf statistische parametrische Methoden erlaubt es, Abweichungen von den zunächst unterstellten Verteilungen zu berücksichtigen, ohne die Fragestellung inhaltlich zu ändern, wie es bei dem Übergang auf nichtparametrische Methoden häufig geschieht. Der Ansatz basiert auf Monte-Carlo-Simulation und wird am Beispiel einer einfaktoriellen Varianzanalyse illustriert. Am Beispiel realer Daten wird gezeigt, dass im regulären Fall (Verteilungen der Messwerte erfüllen alle Voraussetzungen des entsprechenden statistischen Verfahrens) die originale und die modifizierte statistische Methode identische Resultate liefern. Liegen aber grobe Verletzungen

der klassischen Verteilungsannahmen vor, dann kann die modifizierte Variante des Verfahrens auf diese adäquat reagieren. Für eine so modifizierte ANOVA werden die ein Testverfahren charakterisierenden Größen (Wahrscheinlichkeit für den Fehler 1. Art, Operationscharakteristik) angegeben. Die Vorgehensweise der vorgestellten MC-Modifikation (Monte-Carlo-Modifikation) wurde in einem SAS-Programm implementiert.

Schlüsselwörter: Verteilungsannahme, Resampling, Inversionsmethode, Monte-Carlo-Simulation, ANOVA

Einführung

In der üblichen Praxis klinischer Studien entsteht nicht selten die Notwendigkeit, Mittelwerte zu vergleichen. Ausgehend von der Datenlage bedient man sich in der Regel folgender statistischer Verfahren: ANOVA oder t-Test oder nichtparametrische Methoden wie z.B. Mann-Whitney U-Test, Kruskal-Wallis-Test u.a. Die Anwendung dieser Methoden ist an relativ strenge Voraussetzungen (wie z.B. Normalverteilung, gleiche Varianzen) gebunden. Als Minimum wird in den klassischen Versionen dieser Methoden verlangt, dass die Werte der interessierenden Variablen einer Normalverteilung folgen, bzw. gleiche Verteilungsformen in einzelnen Gruppen aufweisen. In der Arbeit [1] wurde die Robustheit des t-Tests auf einer breiten Verteilungsklasse (Fleishman [2]) gezeigt. Die Arbeit [3] zeigt dasselbe auch für Welch- und U-Test sowohl mit als auch ohne Pretesting. In der Realität entstehen dennoch Verteilungen aus einer relativ breiten Klasse (z.B. Mischverteilungen, andere Verteilungen, die den sogenannten medium-tailed Verteilungen ähnlich sind usw.), auf denen die Robustheit der oben erwähnten Methoden, wie es weiter gezeigt wird, verschwindet.
Bei den parametrischen Verfahren ist dies damit zu erklären, dass die empirische Verteilung der Teststatistik in solchen Fällen nicht die theoretische als ihren Grenzwert hat. Weiter unten wird gezeigt, dass dieses Problem sich durch eine Modellierung der empirischen Verteilung der Teststatistik lösen lässt.
Ein weiterer Ansatz, der sich bei dieser Problematik anbieten würde, sind die Permutationstests. In dieser Arbeit wird der studentisierte Zweistichproben-Permutationstest [4] betrachtet.
Die oben genannten nichtparametrischen Methoden, die für die Ausgangsfragestellung auch herangezogen werden können, hängen von den Annahmen über die Verteilungsklasse zwar weniger stark ab, aber auch sie haben ihre Anwendungsgrenzen (wie z.B. gleiche Verteilungsfamilien in den einzelnen zu vergleichenden Gruppen).
In dieser Arbeit wird eine spezielle Variante von smoothed bootstrap [5] vorgeschlagen, die es erlaubt, die Verteilung der empirischen Teststatistik entsprechend der Verteilung vorliegender Daten zu konstruieren und in den statistischen Test einzubeziehen.
In dem Programmpaket SPSS gibt es auch die Möglichkeit Mittelwerte mittels bootstrap-Konfidenzintervalle zu vergleichen. Es ist aber zu bemerken, dass dort die so genannte klassische Variante von bootstrap verwendet wird, die nur mit vorliegenden Stichprobenwerten operiert und, wie in [5] gezeigt wird, in einigen Fällen (besonders bei kleinen Stichproben) im Gegensatz zu der smoothed Variante einen größeren Standardfehler hat.

Methodenbeschreibung

Die Konstruktion eines Tests zielt auf eine Teststatistik mit bestimmten Eigenschaften. Eine davon ist, dass die Teststatistik unter der Nullhypothese eine bekannte Verteilung besitzt, aus der sich ein kritischer Wert als Quantil dieser Verteilung bestimmen lässt. Um das zu gewährleisten, müssen einige Bedingungen an die Verteilung der zu analysierenden Variablen gestellt werden. Weicht die Verteilungsform der Stichprobe von der in einem Verfahren unterstellten ab, so hat die Teststatistik nicht die gewünschte Verteilung und der kritische Wert kann sehr oft nicht bestimmt werden. Die in dieser Arbeit vorgeschlagene Test-Methodik beruht auf der Monte Carlo Methode, in der das Resampling nach der Inversionsmethode durchgeführt wird. Die empirische Verteilungsfunktion wird hier mit einer stückweise linearen Funktion approximiert.
Es werden nun die Einzelheiten dieses Ansatzes am Beispiel einer einfaktoriellen ANOVA für 2 Gruppen betrachtet und das so modifizierte Verfahren wird *MC-ANOVA* (Monte Carlo ANOVA) genannt.
Seien Y_1 und Y_2 stetige Zufallsvariablen, deren Realisierungen 2 Gruppen $(Y_{1,i_1})_{i_1=1,\dots,n_1}$ und $(Y_{2,i_2})_{i_2=1,\dots,n_2}$ innerhalb einer Stichprobe definieren. Die Gruppenerwartungswerte werden mit μ_1 und μ_2 bezeichnet. Zu klären ist, ob sich die Gruppen bezüglich dieser signifikant voneinander unterscheiden.
Zu prüfen ist die Nullhypothese
$H_0 : \mu_1 = \mu_2$
gegen die Alternativhypothese
$H_1 : \mu_1 \neq \mu_2$.
Tatsächlich wird bei der Hypothesenprüfung der beobachtete Wert der Teststatistik mit dem $(1-\alpha)$-Quantil ihrer Verteilung $(Q_{1-\alpha}^{H_0})$ unter H_0 (bei einem vorgegebenen α) verglichen. Sind Y_1 und Y_2 normalverteilt, dann ist dieses Quantil aus theoretischen Überlegungen bekannt. Interessant ist aber der Fall, wo die Verteilungsannahmen der Varianzanalyse (Normalverteilung, Varianzhomogenität) grob verletzt sind. In dieser Situation ist die Verteilung

der Teststatistik und in Folge das $(Q_{1-\alpha}^{H_0})$ im Allgemeinen nicht bekannt. Das Ziel ist also, die Verteilung der Teststatistik unter H_0 zu approximieren. Dazu sind einige Vorbereitungen notwendig. An dieser Stelle ist es wichtig, zu bemerken, dass sich die Nullhypothese H_0 nur auf die Gleichheit der Mittelwerte konzentriert und keine zusätzlichen Bedingungen (Normalverteilung, Varianzhomogenität usw.) an die Verteilung der Daten gestellt werden.
In einem ersten Schritt wird die Teststatistik T_{obs} zum F-Test der Varianzanalyse auf der Basis der beobachteten Daten berechnet. Es wird auch im Weiteren für die Teststatistik eine allgemeine Bezeichnung *T* bzw. T_n verwendet, um zum Einen das Lesen zu erleichtern und zum Anderen zu betonen, dass die weiter unten beschriebene MC-Modifikation ihre Anwendung nicht nur in der ANOVA, sondern auch in anderen Testverfahren finden kann.
Danach werden die Daten in den einzelnen Gruppen um ihren Mittelwert $\hat{\mu}$ zentriert, so dass sie der Nullhypothese H_0 entsprechen
$Y_{k,i_k}^z = Y_{k,i_k} - \hat{\mu}_k$, mit k=1,2 und i_k=1,...,n_k.
In jeder zentrierten Gruppe wird die empirische Verteilungsfunktion berechnet, bezeichnet mit F_n^1 und F_n^2. Seien $F_n^{1,int}$ und $F_n^{2,int}$ die mit Hilfe von stückweise linearen Funktionen approximierten F_n^1 und F_n^2.

Nun können aus den geglätteten Funktionen F_n^{int} mit Hilfe der Inversionsmethode *N* neue Stichproben erzeugt werden, deren Elemente wie die Elemente der zentrierten Stichprobe verteilt sind und der Nullhypothese entsprechen.
Im nächsten Schritt wird für jede der *N* neuen Stichproben die Teststatistik der Varianzanalyse berechnet, wodurch eine Folge $T_n^1, \ldots, T_n^N$ entsteht. Diese Folge beschreibt die Verteilung der Teststatistik T_n unter der Nullhypothese H_0 und liefert das empirische (1–α)-Quantil $Q_{1-\alpha}^{emp}$. Das war der letzte Baustein und jetzt kann der Test der Varianzanalyse wie folgt modifiziert werden:

H_0 wird zum Niveau α abgelehnt, wenn $T_{obs} > Q_{1-\alpha}^{emp}$.

Eigenschaften der MC-ANOVA

Nach der obigen Beschreibung werden im folgenden Abschnitt einige für die Testverfahren üblichen Charakterisierungen angegeben.
Zunächst liege der so genannte reguläre Fall vor, wo alle für ein gewähltes Testverfahren notwendigen Verteilungsannahmen gelten.
In diesem Fall hält das modifizierte Verfahren das vorgegebene Signifikanzniveau ein. In einer Simulationsstudie mit 10.000 Stichproben und einem nominellen α=5% betrug die Wahrscheinlichkeit für den Fehler 1. Art 4,29%.
Ein weiteres Beurteilungskriterium für einen Test ist die Operationscharakteristik (OC). Abbildung 1 und Abbildung 2 (mit schwächerer, Delta=4 bzw. stärkerer, Delta=2, Überlappung der Gruppen) zeigen die OC-Kurve der MC-ANOVA. Hier wurden Stichproben aus einer normalverteilten Grundgesamtheit zugrunde gelegt, um die theoretischen Werte der OC-Kurve der klassischen ANOVA als Referenz verwenden zu können. Diesen Abbildungen ist z.B. zu entnehmen, dass die Trennschärfe des MC-Tests mit wachsendem Stichprobenumfang größer wird.
Um zu veranschaulichen, dass die MC- und die klassische Variante eines Tests in einem regulären Fall (alle Verteilungsvoraussetzungen für das Testverfahren sind erfüllt) zu gleichen Resultaten führen, wird ein Datenbeispiel betrachtet.
Im Rahmen des Teilprojektes B2 „Verteilungsbasierte Simulation“ des SFB 747 „Mikrokaltumformen“ sollten Komponenten des Spannungstensors für ein RVE (repräsentatives Volumenelement) mit Kornstruktur für Stahl DC01 bezüglich ihrer Mittelwerte verglichen werden. Da die Hypothese einer Normalverteilung der Daten in beiden Gruppen (Abbildung 3) nach dem KS-Test nicht abgelehnt werden kann und die Gruppenumfänge groß sind, kann hier die Methode der klassischen ANOVA zur Beantwortung der Frage herangezogen werden.
Um die Varianzheterogenität zu berücksichtigen, wird eine Welch-ANOVA verwendet. Wie aus der Abbildung 4 ersichtlich wird, kann die Hypothese gleicher Mittelwerte nicht beibehalten werden. Außerdem zeigt diese Grafik, dass die beiden Methoden (Welch und MC-ANOVA) nicht nur in den p-Werten, sondern auch in den kompletten Verteilungsfunktionen der Teststatistik (p-Wert des KS-Tests ist gleich 0,89) identisch sind.
Dieses Beispiel unterstreicht, dass die beiden Testmethoden in einem regulären Fall zur gleichen Testentscheidung führen.
Nun ist zu dem Fall überzugehen, wo die Verteilungsannahmen für einen Test grob verletzt sind. Als Beispiel für solche Verteilungen dienen die Daten aus der folgenden präklinischen Laboruntersuchung.
Im Bereich der Thorax-Chirurgie wurden zwei Typen der Laserversiegelung von Lungengewebe am ex-vivo ventilierten Schweinelungenmodell: mit und ohne Wasserkühlung [6] untersucht. Es sollte überprüft werden, ob dadurch Unterschiede im Versiegelungsprozess entstehen. Hier wird das Verhalten der Wahrscheinlichkeit für den Fehler 1. Art verschiedener Testverfahren von Interesse sein. Dazu werden die Daten in den beiden Gruppen zentriert, so dass sie der Nullhypothese gleicher Erwartungswerte entsprechen.
Die Abbildung 5 gibt einen Überblick über die zentrierten Werte.
Danach werden 10.000 neue Stichproben erzeugt, deren Elemente dem Verteilungsgesetz der zentrierten Werte folgen. In jeder dieser Stichproben werden ANOVA, t-Test, studentisierter Zweistichproben-Permutationstest, MC-ANOVA und der MC-t-Test für den Gruppenvergleich herangezogen. Die beiden Gruppen wurden auch mit Hilfe von 95%-Konfidenzintervalle für Mittelwertdifferenz verglichen. Für die Konstruktion der Konfidenzintervalle wurden zwei Varianten von Bootstrap (smoothed und klassisch) verwendet. Die Abbildung 6 zeigt die Verläufe der dabei entstandenen Wahrscheinlichkeiten für den

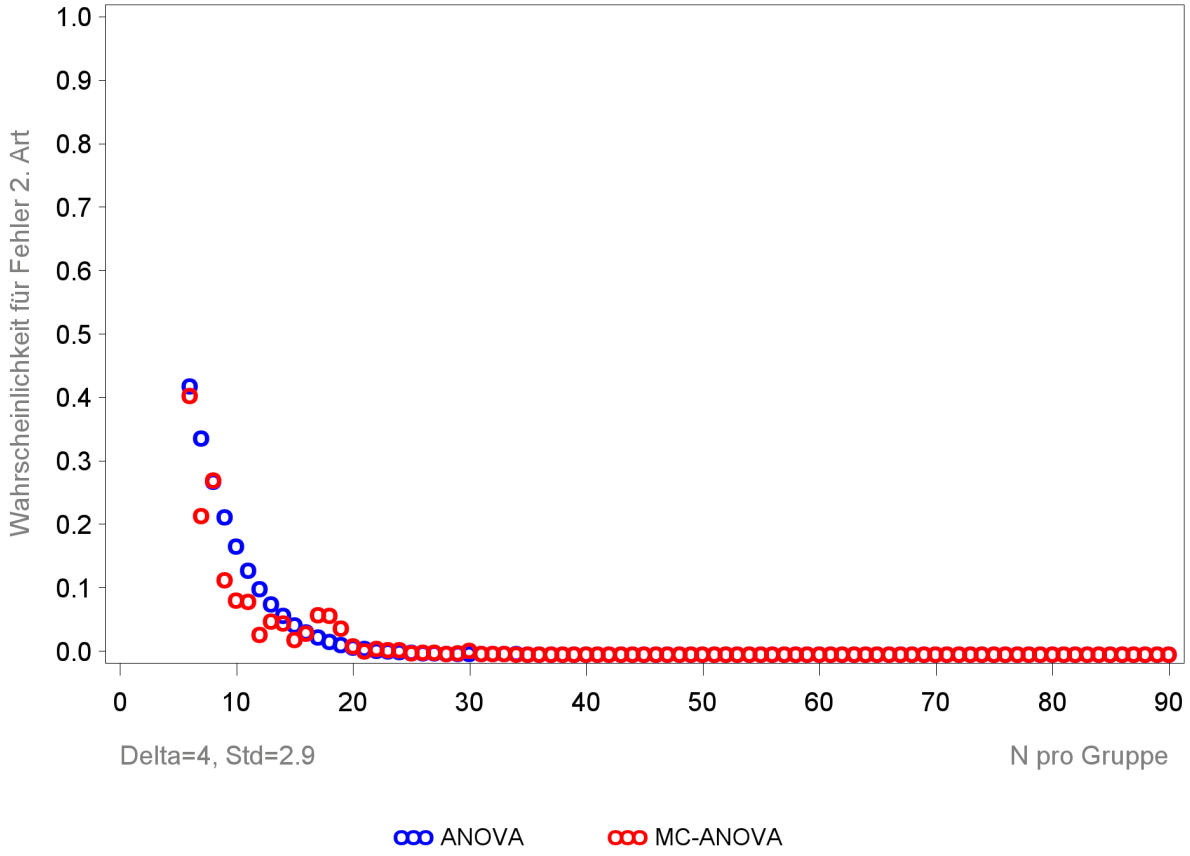

Abbildung 1: Operationscharakteristik der ANOVA und MC-ANOVA (schwache Überlappung der einzelnen Gruppen)

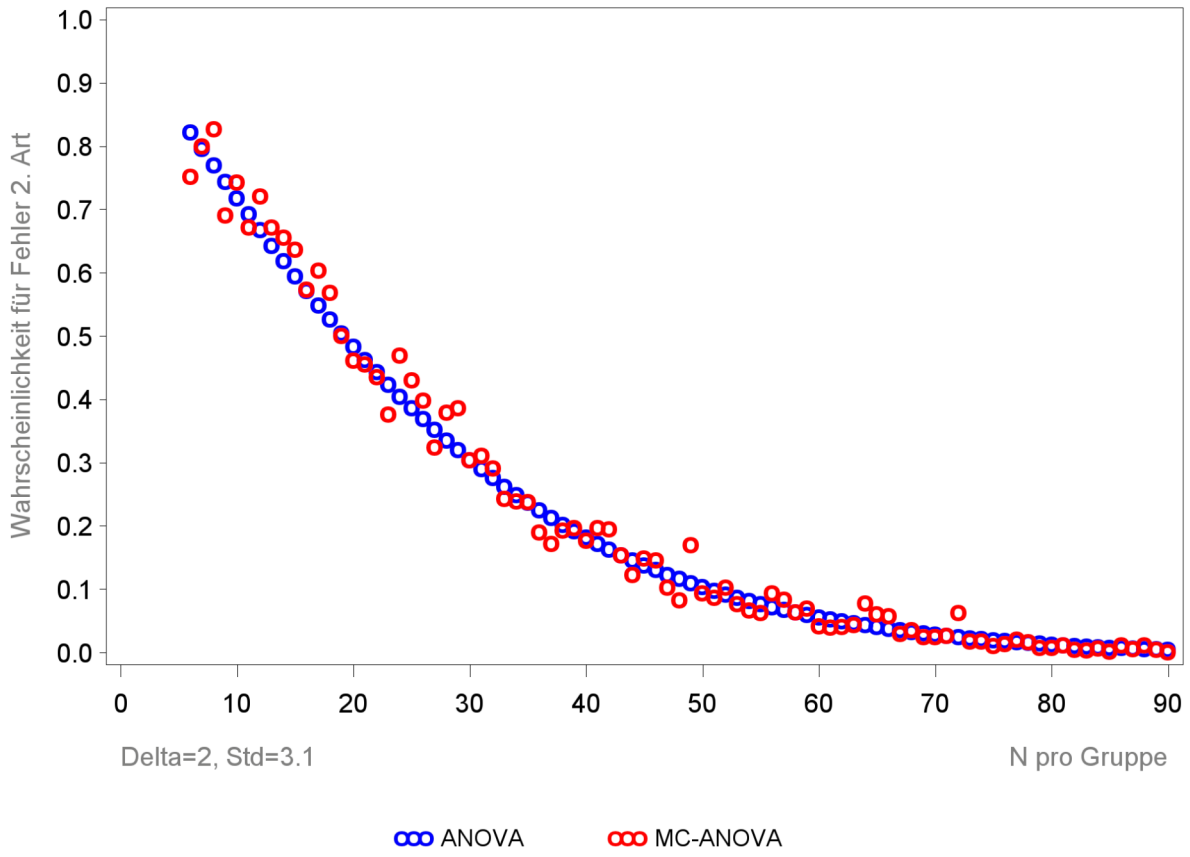

Abbildung 2: Operationscharakteristik der ANOVA und MC-ANOVA (starke Überlappung der einzelnen Gruppen)

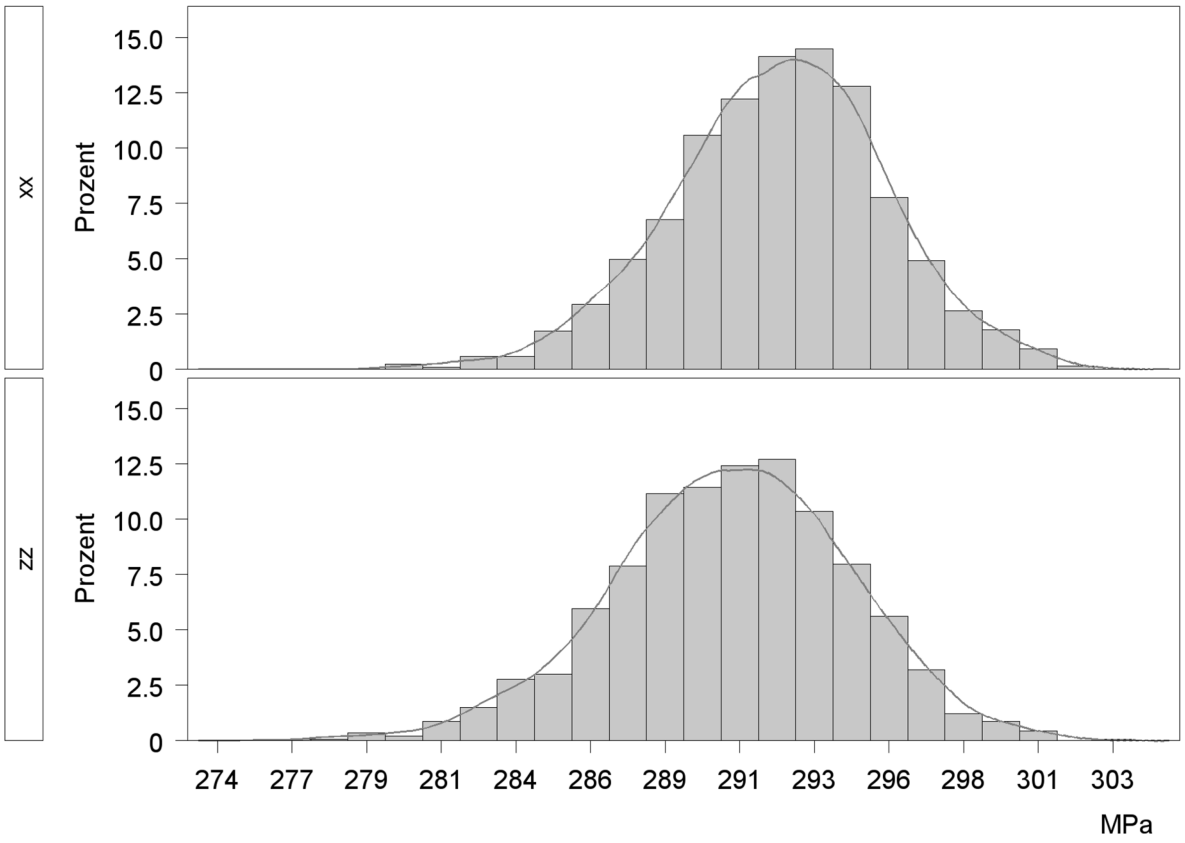

Abbildung 3: Histogramm der berechneten Werte σ_{xx} und σ_{zz} des Spannungstensors

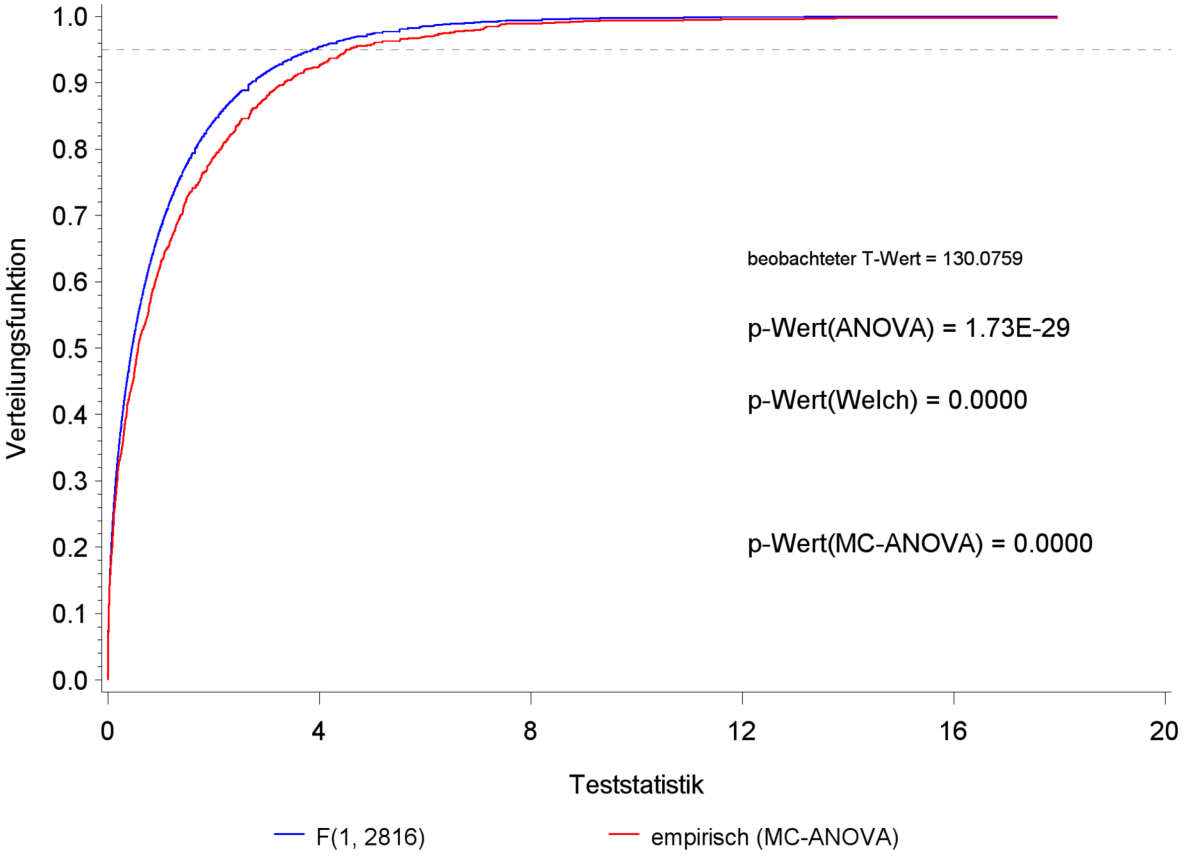

Abbildung 4: Verteilungsfunktion der Teststatistik unter H_0 und die Testergebnisse von ANOVA, Welch-ANOVA und MC-ANOVA

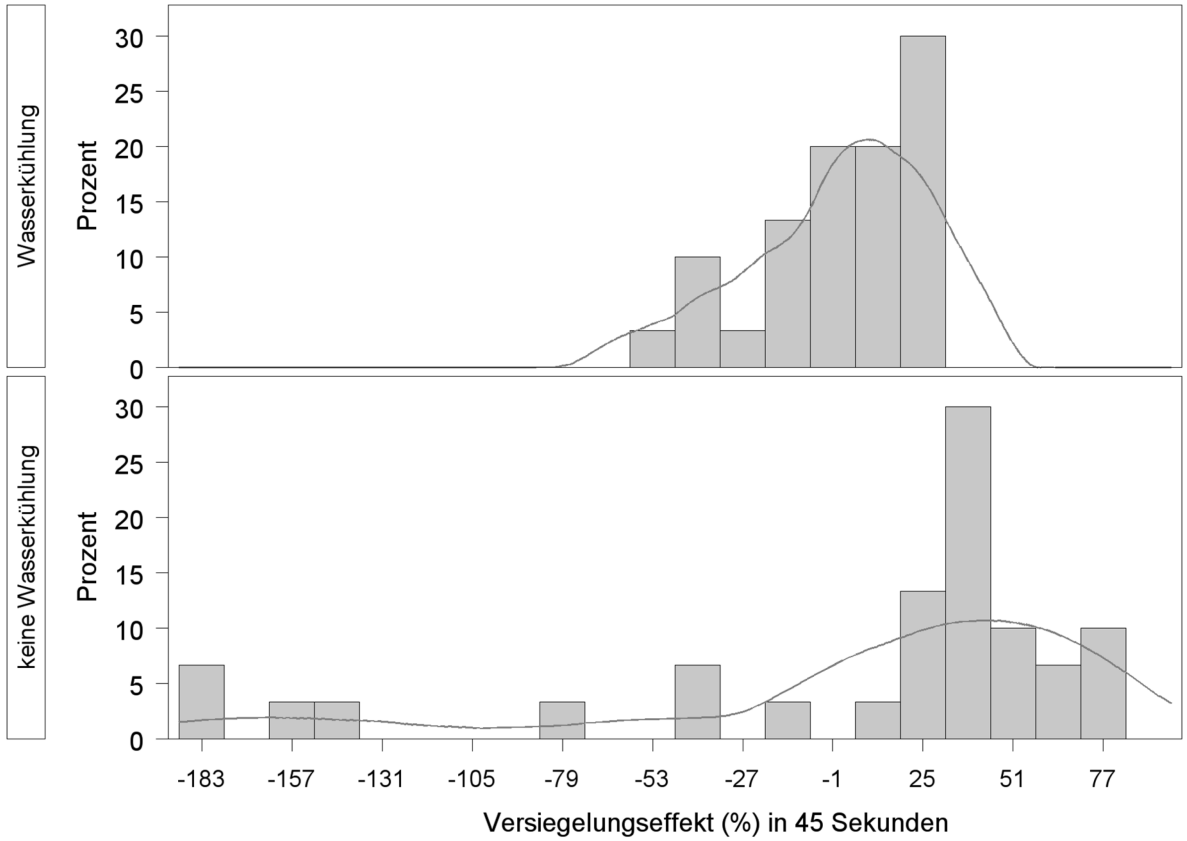

Abbildung 5: Histogramm der zentrierten Messwerte (Versiegelungseffekt)

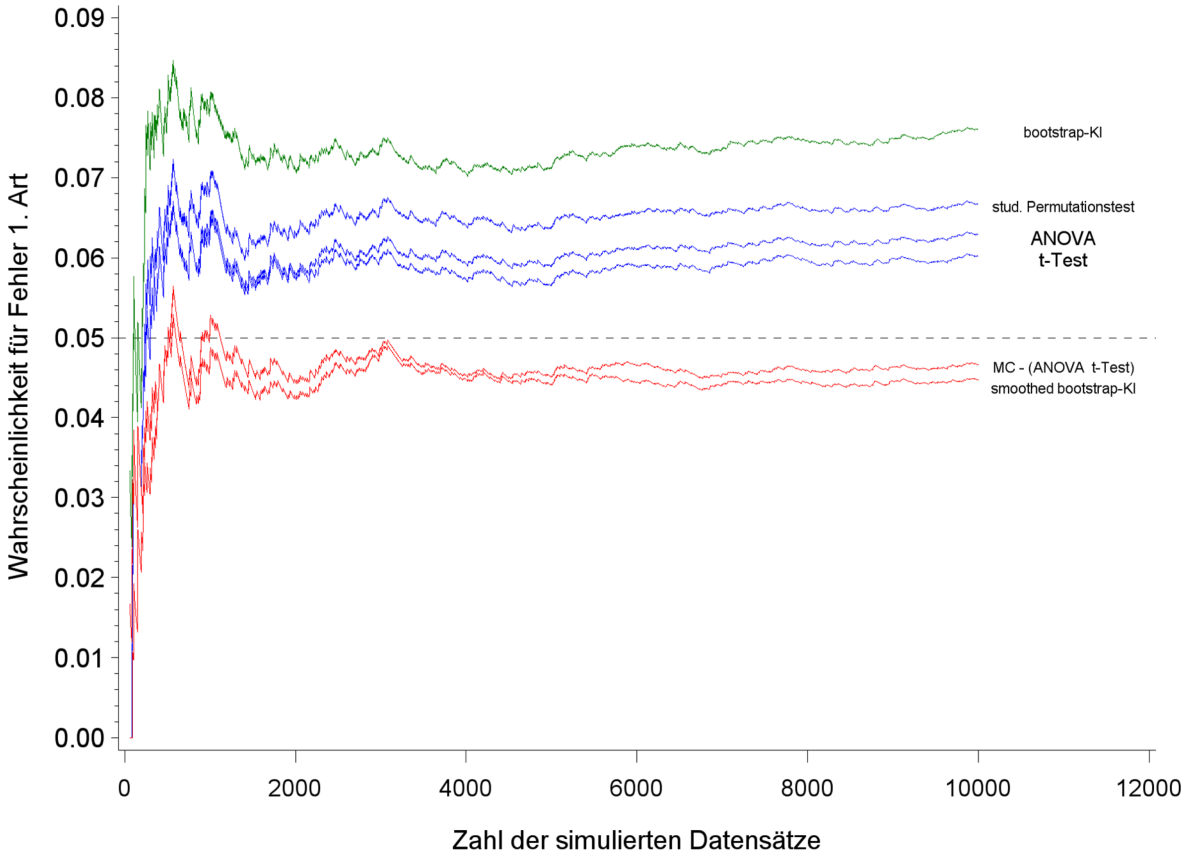

Abbildung 6: Verlauf der Wahrscheinlichkeit für den Fehler 1. Art für verschiedene Testverfahren

Fehler 1. Art. Die X-Achse zeigt die Anzahl der unter H_0 erzeugten Datensätze, auf deren Basis diese Wahrscheinlichkeit berechnet wurde.
Hier ist leicht zu sehen, dass Permutationstest, t-Test und das Verfahren der Varianzanalyse unter den Bedingungen der vorliegenden Daten zum Signifikanzniveau 5% nicht mehr robust sind. Unterschiedliche Verteilungsformen in einzelnen Gruppen sowie große Varianz der Daten sind die Ursache. Die beiden MC-Varianten der Tests halten jedoch das Signifikanzniveau ein, und die Verläufe ihrer Werte der Wahrscheinlichkeit für den Fehler 1. Art sind kaum voneinander zu unterscheiden.

Fazit

Die in dieser Arbeit betrachteten MC-Testvarianten sind zwar rechenaufwändiger als die klassischen Versionen der Testmethoden (ANOVA, t-Test), können aber für eine breitere Verteilungsklasse von zufälligen Fehlern angewendet werden. Die MC-modifizierten Verfahren verlangen keine Kenntnisse über die Form oder die Parameter der zugrundeliegenden Verteilung und benötigen auch kein Pretesting.
Im Fall gültiger Verteilungsannahmen zeigen die MC-Tests ein Resultat, welches mit dem der klassischen Testversionen vergleichbar ist. In der Verteilungsklasse, in der klassische Methoden ihre Robustheit verlieren, halten die MC-Verfahren das vorgegebene Signifikanzniveau (5%) ein.
Die vorgeschlagene MC-Modifikation eines Verfahrens kann ohne grundsätzliche Änderungen auf die Vergleiche von mehr als zwei Gruppen verallgemeinert werden. Auch der weitergehende Einsatz bei der Fallzahlberechnung zur Planung einer Studie ist unmittelbar möglich.

Anmerkungen

Interessenkonflikte

Die Autoren erklären, dass sie keine Interessenkonflikte in Zusammenhang mit diesem Artikel haben.

Danksagung

Die Autoren bedanken sich bei der Deutschen Forschungsgemeinschaft (DFG) für eine teilweise finanzielle Unterstützung dieser Arbeit im Rahmen des SFB 747 „Mikrokaltumformen: Prozesse, Charakterisierung, Optimierung“. Ein weiterer Dank geht an das Team der Klinik für Thoraxchirurgie Klinikum Bremen Ost für die zur Verfügung gestellten Daten [6] und an Dr. W. Wosniok für eine aktive nützliche Unterstützung.

Literatur

1. Rasch D, Guiard V. The robustness of parametric statistical methods. Psychol Sci. 2004;46(2):175-208.
2. Fleishman AI. A method for simulating non-normal distributions. Psychometrika. 1978;43(4):521-32. DOI: 10.1007/BF02293811
3. Rasch D, Kubinger KD, Moder K. The two-sample t-test: pre-testing its assumptions does not pay off. Stat Pap. 2011;52(1):219-31. DOI: 10.1007/s00362-009-0224-x
4. Janssen A. Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens-Fisher problem. Stat Probab Lett. 1997;36(1):9-21. DOI: 10.1016/S0167-7152(97)00043-6
5. Silverman BW, Young GA. The bootstrap: To smooth or not to smooth? Biometrika. 1987;74(3):469-79.
6. Tonoyan T, Prisadov G, Menges P, Herrmann K, Bobrov P, Linder A. Wassergekühlte Laserversiegelung von Lungengewebe am ex vivo ventilierten Schweinelungenmodell [Water-cooled laser sealing of lung tissue in an ex-vivo ventilated porcine lung model]. Zentralbl Chir. 2014 Jun;139(3):329-34. DOI: 10.1055/s-0033-1360282

Korrespondenzadresse:
Dmitri Handschuh
Universität Bremen, Kompetenzzentrum für Klinische Studien Bremen (KKSB), Abteilung Biometrie, Linzer Straße 4, 28359 Bremen, Deutschland, Tel.: +49 (0) 421 218 63795, Fax: +49 (0) 421 218 63799
handschuh@math.uni-bremen.de

Bitte zitieren als
Handschuh D, Bobrov P. Smoothed Bootstrap und seine Anwendung in parametrischen Testverfahren. GMS Med Inform Biom Epidemiol. 2015;11(1):Doc01.
DOI: 10.3205/mibe000157, URN: urn:nbn:de:0183-mibe0001578

Artikel online frei zugänglich unter
http://www.egms.de/en/journals/mibe/2015-11/mibe000157.shtml

Veröffentlicht: *30.03.2015*