

Ilmenauer Beiträge zur Wirtschaftsinformatik

Herausgegeben von U. Bankhofer, V. Nissen
D. Stelzer und S. Straßburger

Dieter W. Joenssen, Jürgen Vogel

**Zur Güte von Signifikanztests auf
mehrdimensionale Normalverteiltheit**

Arbeitsbericht Nr. 2010-02, November 2010



Technische Universität Ilmenau
Fakultät für Wirtschaftswissenschaften
Institut für Wirtschaftsinformatik

Autor: Dieter W. Joenssen, Jürgen Vogel

Titel: Zur Güte von Signifikanztests auf mehrdimensionale Normalverteiltheit

Ilmenauer Beiträge zur Wirtschaftsinformatik Nr. 2, Technische Universität Ilmenau, 2010

ISSN 1861-9223

ISBN 978-3-938940-30-3

urn:nbn:de:gbv:ilm1-2010200316

© 2010 Institut für Wirtschaftsinformatik, TU Ilmenau

Anschrift: Technische Universität Ilmenau, Fakultät für Wirtschaftswissenschaften,
Institut für Wirtschaftsinformatik, PF 100565, D-98684 Ilmenau.
<http://www.tu-ilmenau.de/wid/forschung/ilmenauer-beitraege-zur-wirtschaftsinformatik/>

Gliederung

Abbildungsverzeichnis	iii
1 Einführung.....	1
2 Methodik.....	2
2.1 Ausgewählte Tests	2
2.2 Vorgehensweise	3
2.3 Erzeugung normalverteilter Stichproben	4
2.4 Erzeugung nicht normalverteilter Stichproben	5
3 Ergebnisse.....	8
4 Schlussfolgerungen.....	13
Literaturverzeichnis.....	14

Abbildungsverzeichnis

Abbildung 1 Dichte: Verteilung mit Spitzen.....	6
Abbildung 3 Dichte: Verteilung mit Spitzen.....	6
Abbildung 4 Höhenlinien: Verteilung mit Spitzen.....	6
Abbildung 5 Geschätzte Güte des Shapiro-Wilk-Test	8
Abbildung 6 Geschätzte Güte des Shapiro-Francia-Test	8
Abbildung 7 Geschätzte Irrtumswahrscheinlichkeit des Kurtosis-Test von Kankainen, Taskinen und Oja	9
Abbildung 8 Geschätzte Güte des Schiefe-Test von Kankainen, Taskinen und Oja	9
Abbildung 9 Geschätzte Irrtumswahrscheinlichkeit des Mardia-Kurtosis-Test aus dem Paket QRMLib	10
Abbildung 10 Geschätzte Güte des Mardia-Kurtosis-Test aus dem Paket QuantPsyc	10
Abbildung 11 Geschätzte Güte des Doornik-Hansen-Omnibustest	10
Abbildung 12 Geschätzte Irrtumswahrscheinlichkeit des Doornik-Hansen-Omnibustest angepasst gemäß Lobato und Velasco	10
Abbildung 13 Geschätzte Güte der verallgemeinerten Shapiro-Wilk-Test gemäß Villasenor, José und Estrada	11
Abbildung 14 Geschätzte Irrtumswahrscheinlichkeit des „Multivariate Shapiro-Wilk- Test“	12
Abbildung 15 Geschätzte Irrtumswahrscheinlichkeit des „Multivariate Shapiro-Francia- Test“	12
Abbildung 16 Geschätzte Güte des Energy-Statistic-Test	12
Abbildung 17 Geschätzte Güte für den auf die quadrierten Mahalanobisdistanzen angewandten Kolmogorov-Smirnov-Test (β -Verteiltheit)	12

Zusammenfassung: Die meisten multivariaten statistischen Analyseverfahren erfordern mehrdimensional normalverteilte Beobachtungen. Zur Prüfung, ob eine Stichprobe aus einer normalverteilten Grundgesamtheit stammen könnte, sind zahlreiche Signifikanztests entwickelt worden. Diese Tests weisen nicht immer die gleiche Sensitivität hinsichtlich Abweichungen von der Normalverteilung auf, daher ist die Auswahl des geeigneten Testverfahrens wichtig. Diese Arbeit untersucht simulativ die Güte der in dem Statistikpaket R implementierten Signifikanztests auf mehrdimensionale Normalverteiltheit und stellt sie der einfachen Variante, jede eindimensionale Beobachtungsreihe einzeln auf Normalverteiltheit zu prüfen, gegenüber. Die Untersuchungen zweidimensionaler Verteilungen zu einer Irrtumswahrscheinlichkeit von 5 % ergeben, dass fast ein Drittel der in R implementierten Tests die geforderte Wahrscheinlichkeit für den Fehler erster Art überschreiten. Die anderen Tests sind mächtiger als die einfache Variante, zum Teil sogar bei Verteilungen mit nicht normalen eindimensionalen Randverteilungen. Zwar erweist sich kein Test in allen Belangen als der beste, jedoch ist der „Energy-Statistic-Test“ stets eine gute Wahl, insbesondere bei großen Stichprobenumfängen.

Abstract: Most multivariate statistical analysis procedures require data to be multivariate normally distributed. Many tests have been developed to verify if a sample could indeed have come from a normally distributed population. These tests do not all have the same sensitivity for detecting departures from normality, and thus a choice of test is of central importance. The following study investigates through simulated data the power of those tests for multivariate normality implemented in the statistic software R and pits them against the naive variant of testing each marginal distribution for normality. The results of testing two dimensional data at a level of significance $\alpha=5\%$ show that almost a third of those tests implemented in R do not have a type I error below this. Other tests outperformed the naive variant in terms of power even when the marginals were not normally distributed. Even though no test was consistently better than all alternatives with every alternative distribution, the „Energy-Statistic-Test“ always showed relatively good power across all tested sample sizes.

Schlüsselworte: multivariate Normalverteilung, Anpassungstests, Güte

1 Einführung

Die meisten multivariaten statistischen Verfahren für stetige Variablen setzen mehrdimensional normalverteilte Variablen voraus. Sicherlich sind viele Abweichungen von der Normalität in höheren Dimensionen auch durch entsprechende univariate Verfahren zu erkennen, jedoch besteht ein Bedarf, die multivariate Natur der Daten auszunutzen, um eine größere Sensitivität für Abweichungen zu erlangen. Die höhere Sensitivität ist auch erforderlich, da durch die verbundene Betrachtung von Daten die „Datendichte“ pro Grundeinheit der jeweiligen Dimension (vom eindimensionalen Intervall bis hin zum multidimensionalen Hyperwürfel) exponentiell abnimmt. Dieses auch als „Flucht in die Dimensionen“ bezeichnete Problem ist der primäre Grund, warum die klassischen Verfahren, wie der χ^2 -Anpassungstest, keine Anwendung in höheren Dimensionen finden. Ferner gibt es auch Abweichungen von der Normalität in höheren Dimensionen, die sich nicht in den Randverteilungen widerspiegeln oder gar in einer vergleichbaren Form im Eindimensionalen existieren (vgl. Kowalski, 1973).

Während jedoch professionelle Statistiksoftware in der Regel grafische und numerische Möglichkeiten zum Testen auf univariate Normalverteiltheit bietet, sind multivariate Verfahren so gut wie nicht implementiert. In der Statistiksoftware R werden dagegen zahlreiche Signifikanztests auf mehrdimensionale Normalverteiltheit angeboten, auf die mittlerweile auch durch andere Statistiksoftware über eine Integration von R zugegriffen werden kann.

Ziel der Untersuchungen ist es, auf simulativem Wege die Güte der in R implementierten Signifikanztests auf multivariate Normalverteiltheit zu ermitteln. Insbesondere wird auf die Frage eingegangen, inwiefern die verfügbaren multivariaten Tests einen Vorteil gegenüber der Anwendung einfacher univariater Signifikanztests auf jeder Komponente der Beobachtungen einzeln bieten.

2 Methodik

Ausgewählt für die Überprüfung ihrer Güte wurden 15 für R in der Version 2.10.1 bereitgestellte Tests auf mehrdimensionale Normalverteiltheit, die sich ohne Neuprogrammierung automatisiert prüfen ließen. Die Ergebnisse wurden mit denen verglichen, die der Shapiro-Wilk- (Shapiro, Wilk, 1965) und der Shapiro-Francia-Tests (Shapiro, Francia, 1972) bei Anwendung allein auf die eindimensionalen Randverteilungen lieferten. Diese beiden Tests zählen bekanntermaßen zu den mächtigsten Normalverteilungstests im univariaten Fall (vgl. Shapiro, Wilk, Chen 1968; Shapiro, Francia, 1972). Die Einzelirrtumswahrscheinlichkeiten für jede Randverteilung wurden hier als gleich groß angenommen und mit Hilfe der Bonferroni-Ungleichung bestimmt. Zur Vereinfachung wurde die Studie auf den Fall zweidimensionaler Beobachtungen beschränkt und ausschließlich das Signifikanzniveau $\alpha = 0,05$ benutzt.

2.1 Ausgewählte Tests

Die ausgewählten Tests können wie folgt klassifiziert werden:

- Methoden mittels Randverteilungsprüfung

Die eindimensionalen Randverteilungen wurden einzeln mit univariaten Tests zum Signifikanzniveau $\alpha/2$ auf Normalverteiltheit geprüft. Dabei wurde ausgenutzt, dass eine mehrdimensionale Normalverteilung auch in ihren Randverteilungen normal sein muss.

- Methoden mittels Schiefe und Exzess

Es wurden 6 R-Funktionen untersucht, die eine direkte Bewertung der Normalität mittels Maßen für Schiefe und Exzess vornehmen. In diese Klasse von Tests fallen unter anderem der Kurtosis- und der Schiefe-Test von Kankainen, Taskinen und Oja (2007). Auch der Doornik-Hansen-Omnibustest (1994), welcher auf den quadrierten Summen von Schiefe und Exzess jeder einzelnen Randverteilung basiert, sowie eine Anpassung dessen durch Lobato und Velasco (2004), der eine schwache Abhängigkeit der Beobachtungen erlaubt, passen in diese Klasse. Weitere Tests in dieser Kategorie waren jeweils zwei Implementierungen des Mardia-Schiefe- und des Mardia-Kurtosis-Tests (Mardia, 1970; Mardia, 1974).

- Erweiterungen der Shapiro-Wilk-Statistik

Als einziger Vertreter von Tests, die auf einer Erweiterung der Shapiro-Wilk-Statistik

basieren, wurde die Verallgemeinerung des Shapiro-Wilk-Tests nach Villasenor, José und Estrada (2009) untersucht. Dieser Test basiert auf einer Prüfgröße, die nach einer Transformation der Daten durch Mittelung der Shapiro-Wilk-Statistiken aller Randverteilungen gebildet wird. Außerdem wurde der Vollständigkeit halber der „Multivariate Shapiro-Wilk-Test“ und der „Multivariate Shapiro-Francia-Test“ (Domanski, 1998) in die Untersuchungen mit einbezogen.

- Methoden der Datendimensionsreduktion

Der Energy-Statistic-Test basiert auf dem Vergleich zwischen der hypothetischen und tatsächlichen Verteilung der euklidischen Distanzen (vgl. Székely, Rizzo, 2005). Auch betrachtet wurde eine Implementierung, bei der die Verteilung der quadrierten Mahalanobisdistanzen mit dem Kolmogorov-Smirnov-Test sowohl auf χ^2 - als auch auf β -Verteiltheit geprüft werden kann.

2.2 Vorgehensweise

Das Anliegen der Arbeit ist es, die Güte der ausgewählten Tests durch ihre Anwendung auf viele simulierte Stichproben abzuschätzen. Die Güte oder Macht eines Tests ist die Wahrscheinlichkeit, die Nullhypothese abzulehnen, wenn sie tatsächlich falsch ist. Da die Nullhypothese „Die Merkmale sind normalverteilt.“ auf mannigfaltige Weise falsch sein kann, ist die Güte immer mit Bezug auf das gewählte Signifikanzniveau, den Stichprobenumfang und die Art der Abweichung, sprich der vorhandenen Alternativverteilung der Merkmale, zu sehen. Deshalb wurden 5 verschiedene zweidimensionale Verteilungen zur Bestimmung der Güte ausgewählt, die von der Normalverteilung zum Teil erheblich abweichen. Das Hauptaugenmerk lag dabei auf 3 Verteilungen, deren Randverteilungen normalverteilt sind, deren gemeinsame Verteilung jedoch nicht normalverteilt ist, um zu bestimmen ob die multivariaten Testverfahren gegenüber den univariaten Testverfahren einen Vorteil bieten. Diese Alternativverteilungen sowie die Erzeugung von Stichproben mit einer solchen Verteilung werden im Abschnitt 2.4 beschrieben.

Zur Güteuntersuchung gehört auch eine Abschätzung der Wahrscheinlichkeit für die Fehlentscheidung, dass die Nullhypothese abgelehnt wird, obwohl sie wahr ist. Für die Wahrscheinlichkeit dieses so genannten Fehlers 1. Art ist mit dem Signifikanzniveau α eine obere Grenze vorgegeben. Ein wichtiges Qualitätskriterium besteht also darin, dass die relative Häufigkeit des Auftretens eines Fehlers 1. Art das vorgegebene α nicht überschreitet. Andererseits

sollte diese Fehlerwahrscheinlichkeit auch nicht wesentlich kleiner als α sein, weil ansonsten womöglich Macht verschenkt wird.

Zu jeder ausgewählten Verteilung wurden jeweils 1000 zweidimensionale Zufallsvektoren generiert. Diese Vektoren wurden mit den Stichprobenumfängen $n = 15, 20, 25, 30, 35, 40, 50, 75, 100, 125, 150, 175, 250, 500$ und 1000 von jedem Test zum Signifikanzniveau $\alpha = 0,05$ geprüft. Als Abschätzung für die Güte und die Wahrscheinlichkeit des Fehlers 1. Art diente dann die relative Häufigkeit der Verwerfungen der Nullhypothese. Zur Erzeugung von Stichproben einer bestimmten Verteilung werden zunächst gleichmäßig verteilten Pseudozufallszahlen benötigt. Zu deren Erzeugung wurde in dieser Studie das Mersenne-Twister-Generator (Matsumoto, Nishimura, 1998) gewählt. Beim Fehler 1. Art wurden zusätzlich auch die Verfahren Wichmann-Hill (1982), Marsaglia-Multicarry (Marsaglia, 1997), Super-Duper (Marsaglia, Ananthanarayanan, Paul, 1973) und Knuth-TAOCP-2002 (Knuth, 2002) herangezogen, um eine Abhängigkeit des Analyseergebnisses vom Zufallszahlengenerator auszuschließen. Die Ergebnisse waren konsistent mit denen unter Verwendung des Mersenne-Twister, so dass auf diese Differenzierung nicht weiter eingegangen wird. Ferner ist anzumerken, dass zur Bestimmung der kritischen Schranken bei dem Energy-Statistic-Test wegen des ansonsten zu hohen Rechenaufwands die Anzahl der Bootstrap-replicates auf 199 begrenzt wurde.

2.3 Erzeugung normalverteilter Stichproben

Um die Wahrscheinlichkeit des Fehlers 1. Art abzuschätzen, wurden zweidimensional normalverteilte Pseudozufallszahlen mittels des „Multinormal-Cholesky-Algorithmus“ (vgl. Hörmann, Leydold, Derflinger, 2004, S. 250) mit den Korrelationskoeffizienten $\rho = -0,99; -0,66; 0; 0,25; 0,5$ und $0,75$ generiert, jeweils einmal mit konstanten einheitlichen Varianzen $\sigma_1^2 = \sigma_2^2 = 1$ und einmal mit unterschiedlichen Varianzen σ_1^2 und σ_2^2 zufällig aus dem Bereich zwischen 0 und 1. Die Dichte dieser Verteilung ist

$$f(x) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}x \cdot \Sigma^{-1} \cdot x^T\right) \text{ mit der Kovarianzmatrix } \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

2.4 Erzeugung nicht normalverteilter Stichproben

Um die Wahrscheinlichkeit des Nichtbegehens eines Fehlers 2. Art bei einer bestimmten Alternativverteilung, abzuschätzen, müssen Stichproben gerade dieser Alternativverteilungen generiert werden. Folgend werden die 5 verwendeten Alternativverteilungen dargestellt.

Quaderverteilung

Die Verteilung, mit gleichmäßiger Verteilung der Variaten auf 0 bis 1, hat die Dichte

$$f(x, y) = \begin{cases} 1 & \text{für } x, y \in (0, 1), \\ 0 & \text{sonst.} \end{cases}.$$

Halbkugelverteilung

Die Verteilung hat die Dichte

$$f(x, y) = \begin{cases} \sqrt{P^2 - x^2 - y^2} & \text{für } P^2 \geq x^2 + y^2, \\ 0 & \text{sonst,} \end{cases}.$$

mit dem Radius $P = \sqrt[3]{6/4\pi}$. Die Erzeugung von Zufallszahlen mit dieser Verteilung ist nicht trivial und wird unter anderem von Johnson (1987, S. 125-127) thematisiert.

Verteilung mit den Spitzen

Erster Vertreter der zweidimensionalen Verteilungen, die normalverteilte Randverteilungen haben, jedoch keine gemeinsame Normalverteilung besitzen, stammt von Feller (1971, S. 121). Die Dichte entspricht

$$f(x, y) = \begin{cases} \frac{1}{2\pi} e^{-0,5(x^2+y^2)} + \frac{1}{2\pi e} x^3 y^3 & \text{für } x, y \in [-1, 1], \\ \frac{1}{2\pi} e^{-0,5(x^2+y^2)} & \text{sonst.} \end{cases}$$

Die Verteilungsglocke hat hier im 1. und 3. Quadranten jeweils eine spitze Aus-, im 2. und 4. Quadranten dagegen jeweils eine spitze Einbuchtung (vgl. Abbildung 1).

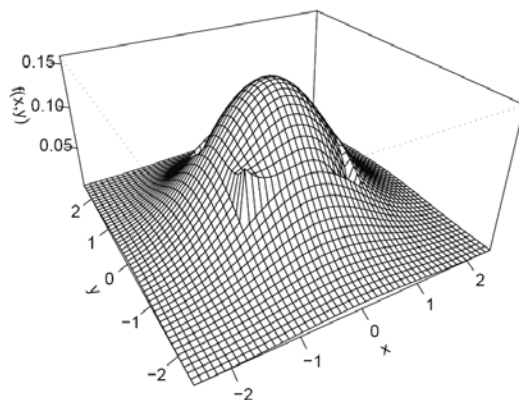


Abbildung 1: Dichte: Verteilung mit Spitzen

Sternförmige Verteilung

Ebenfalls bei Feller (1971) findet man die Dichte

$$f(x, y) = \frac{1}{4\pi} \left(\frac{\exp\left(-\frac{x^2 - 2\rho_1 xy + y^2}{2(1-\rho_1^2)}\right)}{\sqrt{1-\rho_1^2}} + \frac{\exp\left(-\frac{x^2 - 2\rho_2 xy + y^2}{2(1-\rho_2^2)}\right)}{\sqrt{1-\rho_2^2}} \right).$$

Sie ist das arithmetische Mittel zweier Normalverteilungsdichten mit unterschiedlichen Korrelationskoeffizienten ρ_1 und ρ_2 . Sie beschreibt eine glockenförmige Fläche mit sternförmigen Höhenlinien (Abbildungen 2 und 3). Auch diese Verteilung ist zweidimensional nicht normal und hat normalverteilte Randverteilungen. Zur Güteanalyse wurden Zufallsvektoren gemäß dieser Verteilung mit den Werten $\rho_1 = 0,75$ und $\rho_2 = -0,75$ erzeugt. Hier, wie auch in dem anderen Feller-Beispiel, wurden die Pseudo-Zufalls-Stichproben mit Hilfe der „Multivariate Native Ratio-of-Uniforms Method“ realisiert (vgl. Hörmann et. al. 2004).

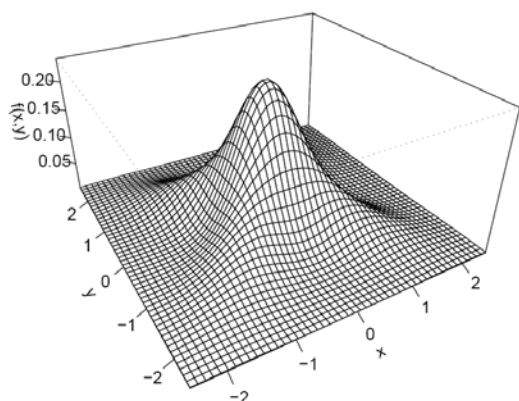


Abbildung 2: Dichte: Verteilung mit Spitzen

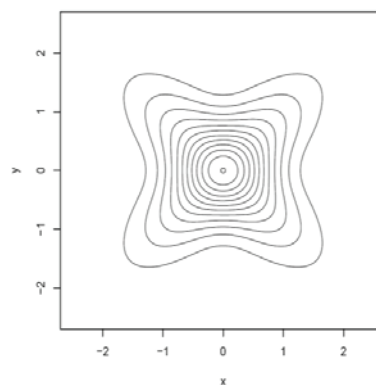


Abbildung 3: Höhenlinien: Verteilung mit Spitzen

Verteilung mit leeren Quadranten korreliert

Als letzte nicht normale Verteilung wurde ein Beispiel von Stoyanov (1987) gewählt. Diese Verteilung mit dem Erwartungswert auf dem Koordinatenursprung kann in dem 2. und 4. Quadranten keine Werte annehmen. Sie weist von den 3 Alternativverteilungen mit normalen Randverteilungen die deutlichste Abweichung von der Normalverteilung auf. Die Dichte dieser Verteilung lautet

$$f(x, y) = \begin{cases} \frac{\exp\left(\frac{-0,5(x^2 - 2\rho xy + y^2)}{1 - \rho^2}\right) + \exp\left(\frac{-0,5(x^2 - 2\rho xy + y^2)}{1 - \rho^2}\right)}{2\pi\sqrt{1 - \rho^2}} & \text{für } xy \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

Zu dieser Verteilung sind Beobachtungsvektoren mit den Korrelationskoeffizienten $\rho = 0$; 0,33; 0,50; 0,75 und 0,90 erzeugt worden. Zur Generierung der Stichproben wurden lediglich bei normalverteilten Stichproben entsprechende Vorzeichenwechsel vorgenommen.

3 Ergebnisse

Es zeigte sich zunächst, dass die Varianzen der eindimensionalen Randverteilungen keinen erheblichen Einfluss auf die Untersuchungsergebnisse haben, so dass hier bei den normalverteilten Beobachtungen nicht zwischen gleich und unterschiedlich großen Varianzen in den Rändern unterschieden werden braucht. Die Verteilung mit den Spitzen wurde von keinem Test als nicht normal erkannt, sie wird deshalb in der Diskussion über die Güte der Tests keine Rolle spielen. Auf die Wahrscheinlichkeit des Fehlers 1. Art wird nur eingegangen, wenn er nicht eingehalten wurde.

Methoden mittels Randverteilungsprüfung

Der Shapiro-Wilk-Test (R-Paket: stats, Funktion: shapiro.test) (Shapiro, Wilk, 1965) und der Shapiro-Francia-Test (R-Paket: nortest, Funktion: sf.test) (Shapiro, Francia, 1972), die nur für das Prüfen auf univariate Normalverteilung geeignet sind, wurden jeweils mit dem Signifikanzniveau $\alpha/2$ einzeln auf die beiden eindimensionalen Beobachtungsreihen angewandt und die Nullhypothese genau dann abgelehnt, wenn der Test wenigstens eine der beiden Randverteilungen als nicht normal erkannte. Bei diesem Vorgehen erwies sich der Shapiro-Wilk- dem Shapiro-Francia-Test als überlegen. Beide Tests hielten das vorgegebene Signifikanzniveau α ein. Für hoch korrelierte Merkmale büßt diese Vorgehensweise jedoch wegen der verwendeten Bonferroni-Ungleichung deutlich an Signifikanz ein. Bei der Güte liegen die beiden Tests im Vergleich mit den anderen im unteren Mittelfeld, sofern die Randverteilungen nicht normalverteilt sind. Verteilungen mit normalen Rändern werden erwartungsgemäß nicht erkannt (vgl. Abbildungen 4 und 5).

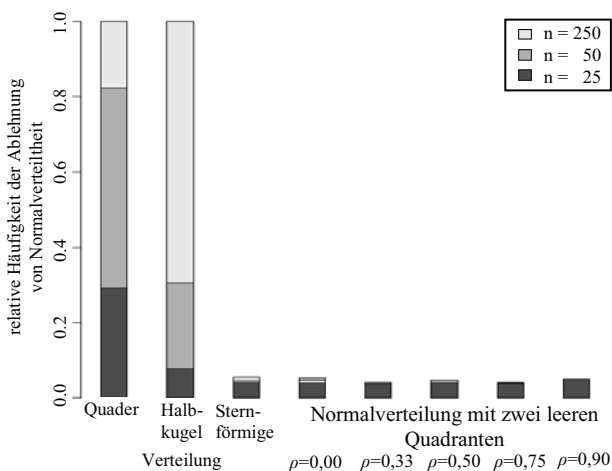


Abbildung 4: Geschätzte Güte des Shapiro-Wilk-Test

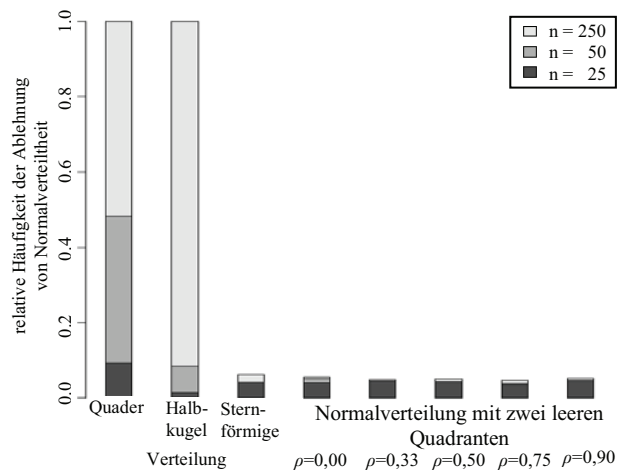


Abbildung 5: Geschätzte Güte des Shapiro-Francia-Test

Methoden mittels Schiefe und Exzess

Nicht geeignet zum Prüfen auf Vorliegen von multivariater Normalverteilung ist der Kurtosis-Test (R-Paket: ICS, Funktion: mvnorm.kur.test) und nur bedingt geeignet ist der Schiefe-Test (R-Paket: ICS, Funktion: mvnorm.skew.test) nach Kankainen, Taskinen und Oja (2007). Der Kurtosis-Test weist einen Fehler 1. Art auf, der stark von der Korrelation der Merkmale abhängt (Abbildung 6). Er verwirft fälschlicherweise die Nullhypothese in bis zu 80 % der Fälle. Lediglich bei einer Standardnormalverteilung mit einheitlichen Varianzen werden die vorgegebenen 5 % durchgängig eingehalten. Der Schiefe-Test kommt bei den verwendeten Alternativverteilungen selbst bei großen Stichprobenumfängen nicht über 30 % Güte hinaus, vgl. Abbildung 7.

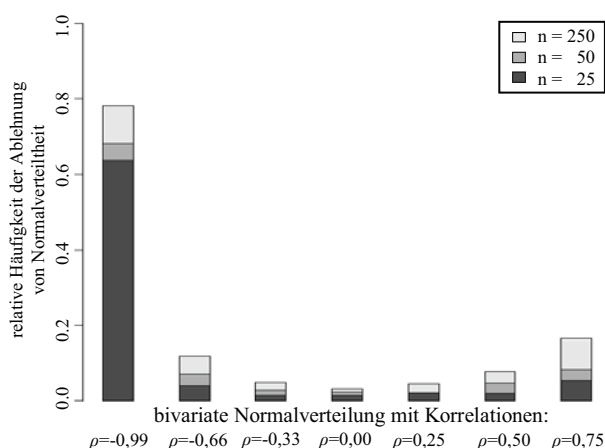


Abbildung 6: Geschätzte Irrtumswahrscheinlichkeit des Kurtosis-Test von Kankainen, Taskinen und Oja

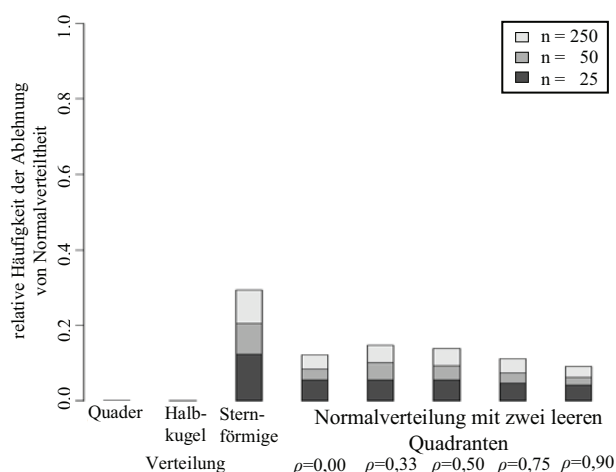


Abbildung 7: Geschätzte Güte des Schiefe-Test von Kankainen, Taskinen und Oja

Auch nicht geeignet zum Testen auf Vorliegen einer multivariaten Normalverteilung ist der Mardia-Kurtosis-Test (Mardia, 1970; Mardia, 1974) aus dem R-Paket QRMLib (Funktion: MardiaTest). Er ist offenbar falsch implementiert, da er mitunter mehr als doppelt so oft die Nullhypothese fälschlicherweise verwirft, als er dürfte (Abbildung 8). Der Mardia-Kurtosis-Test (Mardia, 1970; Mardia, 1974) aus dem Paket QuantPsyc (Funktion: mult.norm) hält die Irrtumswahrscheinlichkeit 1. Art dagegen ein. Tatsächlich ist er einer der besten Tests dieser Studie. Die Halbkugelverteilung erkannte er am häufigsten als nicht normal, vgl. Abbildung 9. Ab einem Stichprobenumfang von 25 kam dieser Test bei jeder verwendeten Alternativverteilung auf über 50 % Macht. Die Mardia-Schiefe-Tests (Mardia, 1970; Mardia, 1974) aus den beiden Paketen weisen die gleichen Schwächen auf wie der von Kankainen,

Taskinen und Oja (2007), was sicher dem Umstand geschuldet ist, dass alle verwendeten Verteilungen gar nicht oder nur unwesentlich schief sind.

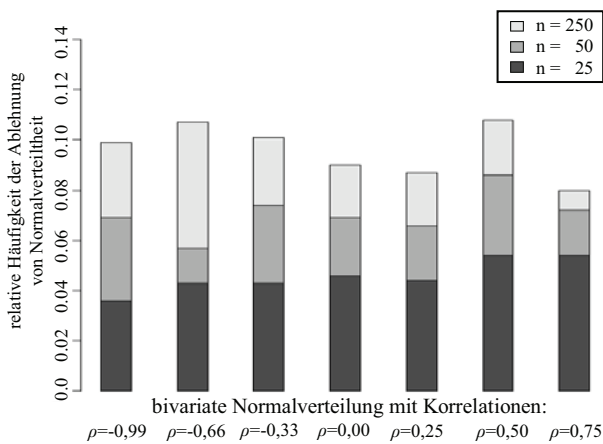


Abbildung 8: Geschätzte Irrtumswahrscheinlichkeit des Mardia-Kurtosis-Test aus dem Paket QRMLib

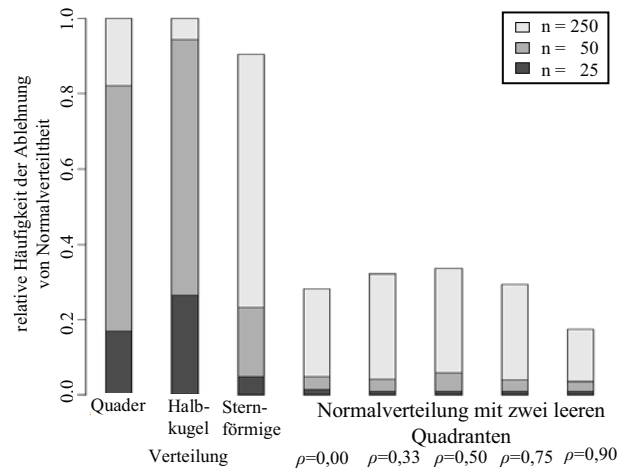


Abbildung 9: Geschätzte Güte des Mardia-Kurtosis-Test aus dem Paket QuantPsync

Der Doornik-Hansen-Omnibustest (1994) (R-Paket: normwn.test, Funktion: normality.test1) überzeugt mit relativ guter Macht bei allen Alternativverteilungen bis auf der sternförmigen. Hier erzielt der Test keine Güte, die besser ist als das vorgegebene Signifikanzniveau, vgl. Abbildung 10. Die Anpassung des Tests nach Lobato und Velasco (2004) verbessert die Situation in diesem Falle auch nicht. Bei dieser angepassten Version (R-Paket: normwn.test, Funktion: normality.test2) wird das vorgegebene Signifikanzniveau bei kleinen Stichproben ($n \leq 50$) nicht eingehalten, und bei größeren Stichproben ist die Güte meist etwas schlechter als die der unangepassten Version, vgl. Abbildung 11.

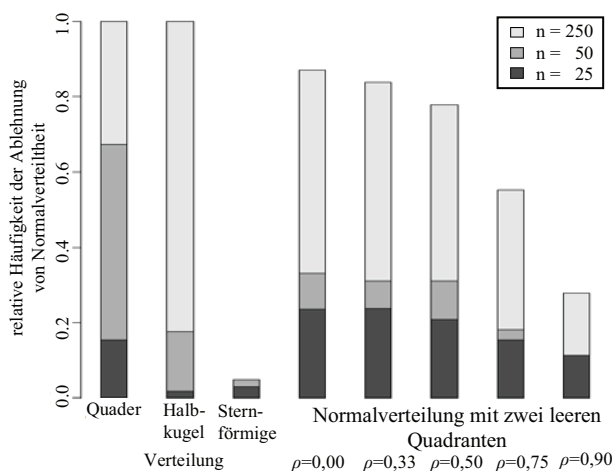


Abbildung 10: Geschätzte Güte des Doornik-Hansen-Omnibustest

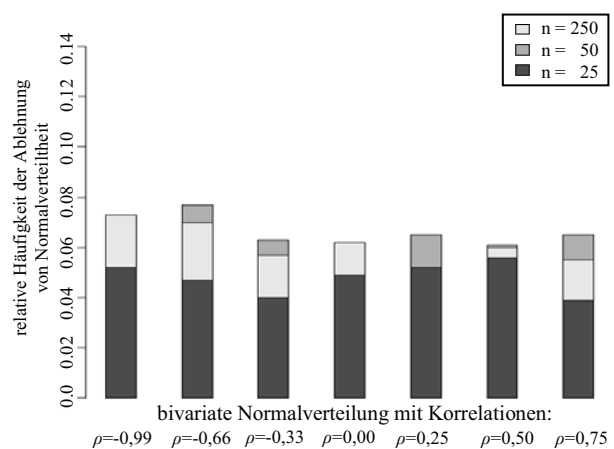


Abbildung 11: Geschätzte Irrtumswahrscheinlichkeit des Doornik-Hansen-Omnibustest angepasst gemäß Lobato und Velasco

Erweiterungen der Shapiro-Wilk-Statistik

Die Verallgemeinerung des Shapiro-Wilk-Test gemäß Villasenor, José und Estrada (2009) (R-Paket: mvShapiroTest, Funktion: mvShapiro.Test) ist bezüglich der Güte mit dem Doornik-Hansen-Omnibustest vergleichbar. Er ist bei fast allen Alternativverteilungen einer der mächtigsten Tests im Vergleich, bei der Quaderverteilung sogar der mächtigste. Ausnahme ist auch hier die sternförmigen Verteilung, bei der die Ablehnungswahrscheinlichkeit nur das vorgegebene Signifikanzniveau erreicht. Dies verdeutlicht Abbildung 12.

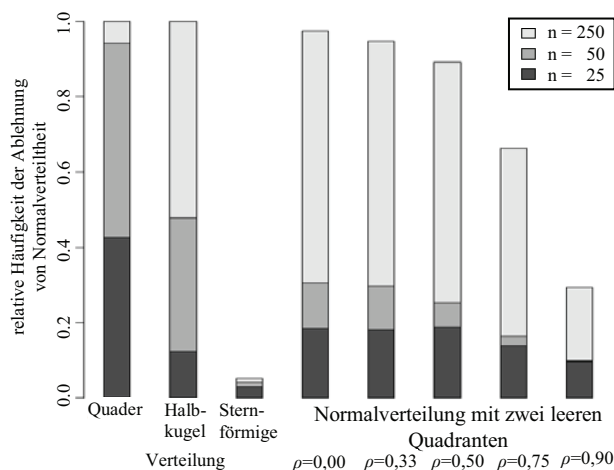


Abbildung 12: Geschätzte Güte der verallgemeinerten Shapiro-Wilk-Test gemäß Villasenor, José und Estrada

Der „Multivariate Shapiro-Wilk-Test“ (Domanski, 1998) (R-Paket: mvnormtest, Funktion: mshapiro.test) und der „Multivariate Shapiro-Francia-Test“ (Domanski, 1998) (R-Paket: mvssf, Funktion: mvssf) sind beide völlig ungeeignet zum Testen auf das Vorliegen von mehrdimensionaler Normalverteilung. Bei beiden Tests liegt die relative Häufigkeit des Fehlers 1. Art weit über dem gewählten Signifikanzniveau, und das für jeden Stichprobenumfang (Abbildung 13). Die Shapiro-Francia-Variante ist dabei noch deutlich schlechter als die Shapiro-Wilk-Variante (Abbildung 14).

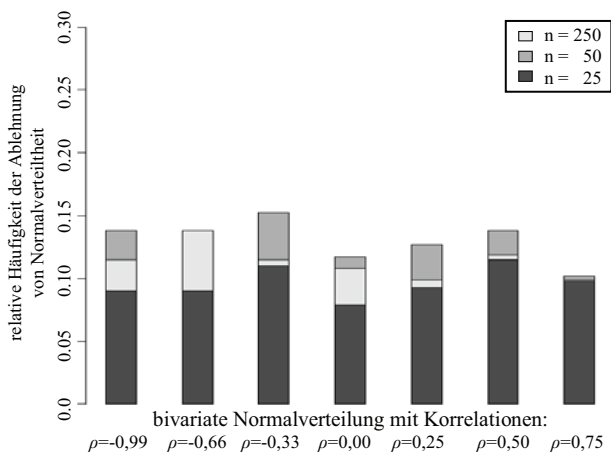


Abbildung 13: Geschätzte Irrtumswahrscheinlichkeit des „Multivariate Shapiro-Wilk-Test“

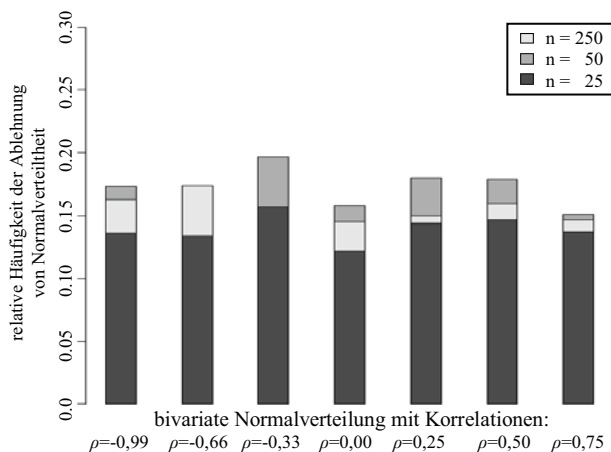


Abbildung 14: Geschätzte Irrtumswahrscheinlichkeit des „Multivariate Shapiro-Francia-Test“

Methoden der Datendimensionsreduktion

Der Energy-Statistic-Test (Székely, Rizzo, 2005) (R-Paket: energy, Funktion: mvnorm.etest) hebt sich hervor, weil er auf jede Alternativverteilung reagiert. Für geringe Stichprobenumfänge ist er bei den Verteilungen mit leeren Quadranten und bei der sternförmigen Verteilung durchgängig besser als alle anderen Tests, aber auch bei den anderen Verteilungen gehört er mit zu den mächtigsten Vertretern wie Abbildung 15 verdeutlicht.

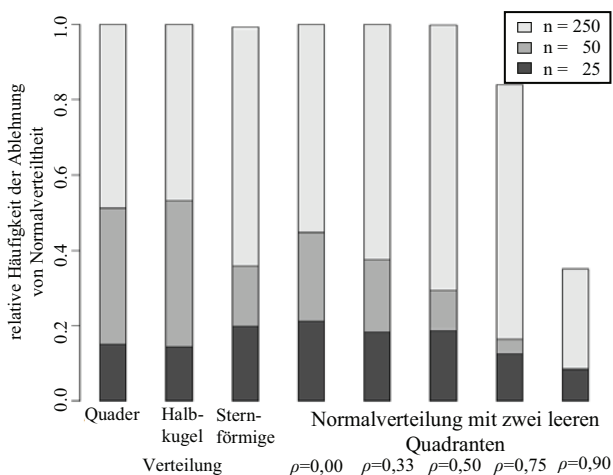


Abbildung 15: Geschätzte Güte des Energy-Statistic-Test

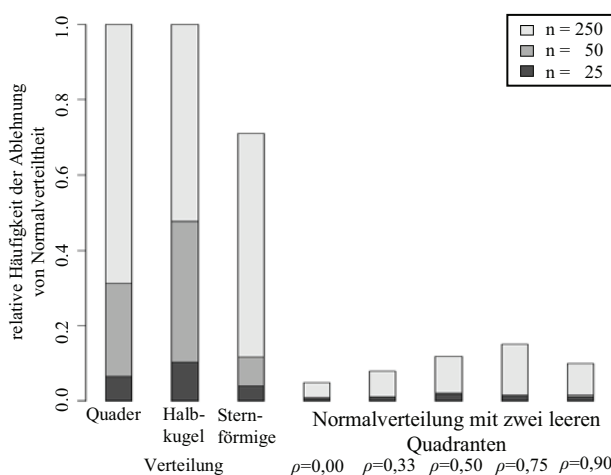


Abbildung 16: Geschätzte Güte für den auf die quadrierten Mahalanobisdistanzen angewandten Kolmogorov-Smirnov-Test (β -Verteiltheit)

Für den auf die quadrierten Mahalanobisdistanzen angewandten Kolmogorov-Smirnov-Test (R-Paket: QRMLib, Funktion: jointnormalTest) spricht, dass er bei jeder Verteilung eine Güte hat, die größer ist als das vorgegebene Signifikanzniveau (Abbildung 16). Aber er ist bei keiner Verteilung der beste. Es macht so gut wie keinen Unterschied, ob die quadrierten Mahalanobisdistanzen auf χ^2 - oder β -Verteiltheit geprüft werden. Keine Methode ist durchgängig oder bedeutend besser als die andere.

4 Schlussfolgerungen

In dieser Studie wurden 15 Anpassungstests auf multivariate Normalverteilung, alles frei verfügbaren R-Funktionen, auf ihre Güte geprüft. Dabei zeigte sich, dass es keinen Test gibt, der in jeder Situation allen anderen überlegen ist. 4 Tests, nämlich der Energy-Statistic-Test aus dem Paket ‚energy‘, der Mardia-Kurtosis-Test aus dem Paket ‚QuantPsyc‘ und die beiden Varianten des auf die quadrierten Mahalanobisdistanzen angewandten Kolmogorov-Smirnov-Tests aus dem Paket ‚QRMLib‘, lehnten bei jeder verwendeten Alternativverteilung und für alle betrachteten Stichprobenumfänge die falsche Nullhypothese mit mindestens 50 % Wahrscheinlichkeit ab. Insgesamt gesehen zeigte sich der Energy-Statistic-Test als bester Test bei der vorliegenden Auswahl. Er gehört bei jeder Verteilung und bei jedem Stichprobenumfang zu den mächtigsten Tests.

4 der untersuchten Tests sind nicht zu empfehlen. Der „Multivariate Shapiro-Wilk-Test“ aus dem Paket ‚mvnormtest‘, der „Multivariate-Shapiro-Francia-Test“ aus dem Paket ‚mvsf‘ und der Kurtosis-Test nach Kankainen-Taskinen-Oja aus dem Paket ‚ICS‘ halten das vorgegebene Signifikanzniveau nicht ein, der Mardia-Kurtosis-Test aus dem R-Paket QRMLib ist wahrscheinlich falsch programmiert.

Schließlich ist anzumerken, dass die alleinige Verwendung univariaten Tests auf die Randverteilungen nicht ausreicht ist, um auf multivariate Normalverteiltheit zu testen. Die Häufigkeit der Ablehnung bei nicht normalen Verteilungen mit normalen Randverteilungen war nicht höher als das gewählte Signifikanzniveau. Auch bei der Quader- und der Halbkugelverteilung stellte diese Vorgehensweise kein gutes Testverfahren dar.

Literaturverzeichnis

- D'Agostino, R. B. und Stephens, M. A. (1986):** *Goodness-of-fit techniques*. New York: Marcel Dekker.
- Doornik, J. A. and Hansen, H. (1994):** *An Omnibus Test for Univariate and Multivariate Normality. Working Paper, Nuffield College*.
- Domanski C. (1998):** Własności testu wielowymiarowej normalności Shapiro-Wilka i jego zastosowanie. *Cracow University of Economics Rector's Lectures*, 37.
- Feller, W. (1971):** *An Introduction to Probability Theory and Its Applications, Volume 2*. New York: John Wiley & Sons.
- Hörmann, W., Leydold, J. and Derflinger, G. (2004):** *Automatic Nonuniform Random variate Generation*. Berlin: Springer.
- Johnson, M. E. (1987):** *Multivariate Statistical Simulation*. New York: John Wiley & Sons.
- Kankainen, A., Taskinen, S. and Oja, H. (2007):** *Tests of multinormality based on location vectors and scatter matrices. Statistical Methods and Applications*, 16, S. 357-379.
- Knuth, D. E. (2002):** *The Art of Computer Programming. Volume 2, third edition*.
- Kowalski, C. J. (1973):** *Non-Normal Bivariate Distributions with Normal Marginals. The American Statistician*, 27, S. 103-106.
- Lobato, I. and Velasco, C. (2004):** *A Simple Test of Normality of Time Series. Econometric Theory*, 20, S. 671-689.
- Mardia, K. V. (1970):** *Measures of multivariate skewness and kurtosis with applications. Biometrika*, 57, S. 519-530.
- Mardia, K. V. (1974):** *Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. Sankhyā, Ser B*, 36, S. 115-128
- Marsaglia, G. (1997):** *A random number generator for C. [online Dokument] URL: <http://mathforum.org/kb/thread.jspa?messageID=1607565> [15.10.2010]*.

- Marsaglia, G., Ananthanarayanan, K. and Paul, N.J. (1973):** *How to use the McGill Random Number Package 'Super-Duper'*. Montreal: School of Computer Science, McGill University.
- Matsumoto, M. and Nishimura, T. (1998):** *Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator*, *ACM Transactions on Modeling and Computer Simulation*, 8, pp. 3-30.
- Shapiro, S. S. and Francia, R. S. (1972):** *Approximate analysis of variance test for normality*. *Journal of the American Statistical Association*, 67, pp. 215-216.
- Shapiro, S. S. and Wilk, M. B. (1965):** *An analysis variance test for normality (complete samples)*. *Biometrika*, 52, pp. 591-611.
- Shapiro, S. S., Wilk, M. B. and Chen, H. J. (1968):** *A comparative study of various tests for normality*. *Journal of the American Statistical Association*, 70, pp. 1343-1372.
- Stoyanov, J. M. (1987):** *Counterexamples in Probability*. New York: John Wiley & Sons.
- Székely, G. J. and Rizzo, M. L. (2005):** *A New Test for Multivariate Normality*. *Journal of Multivariate Analysis*, 93, pp. 58-80.
- Villasenor A., José A. and Estrada, E. G. (2009):** *A Generalization of Shapiro-Wilk's Test for Multivariate Normality*. *Communications in Statistics - Theory and Methods*, 38, pp. 1870-1883.
- Wichmann, B. A. and Hill, I. D. (1982):** *Algorithm AS 183: An Efficient and Portable Pseudo-random Number Generator*. *Journal of the Royal Statistical Society: Series C- Applied Statistics*, 31, pp.188-190