

# Bonett-Methode

## ANMERKUNGEN ZUM ROBUSTEN KONFIDENZINTERVALL NACH BONETT FÜR EIN VERHÄLTNIS VON STANDARDABWEICHUNGEN

Senin J. Banga und Gregory D. Fox  
18. Juni 2013

### ZUSAMMENFASSUNG

Wir schlagen ein alternatives Verfahren vor, um eine geringfügige Unzulänglichkeit im Konfidenzintervall (KI) nach Bonett (2006) in Bezug auf das Verhältnis von zwei Standardabweichungen zu korrigieren. Der zusammengefasste Kurtosis-Schätzwert für die Layard-Teststatistik (1973), auf dem das Bonett-Intervall basiert, ist nur dann konsistent, wenn die Varianzen der Grundgesamtheiten gleich sind. Wir leiten einen alternativen Schätzwert ab, der sowohl bei Gleichheit als auch bei Ungleichheit der Varianzen der Grundgesamtheiten konsistent ist, und das korrekte KI wird mit diesem neuen Schätzwert berechnet. Simulationsstudien verdeutlichen, dass das neue KI im Allgemeinen genauer und präziser als das KI ist, das auf dem Levene/Browne-Forsythe-Test  $W_{50}$  und Pans (1999) Test  $L_{50}$  basiert. Übereinstimmend mit Pan beobachten wir, dass die auf Test  $W_{50}$  basierenden KIs durch einen Präzisionsverlust bei kleinen Stichproben gekennzeichnet sind, was häufig zu Intervallen mit unendlicher Breite führt. Auf Test  $L_{50}$  basierende KIs zeigen eine gute Leistung für symmetrische und nahezu symmetrische Verteilungen; ihre Leistung ist jedoch schlecht, wenn die Grundgesamtheiten schief sind.

*Indexbegriffe: Homogenität von Varianzen, Levenes Test, Brown-Forsythe-Test, Layard-Test, Konfidenzintervall (KI) für das Verhältnis von Varianzen*

## 1. Einführung

Es ist allgemein bekannt, dass der klassische F-Test und die entsprechenden Konfidenzintervalle (KIs) äußerst empfindlich gegenüber Abweichungen von der Normalverteilung sind. Die Empfindlichkeit ist sogar so stark, dass der klassische F-Test für die meisten praktischen Anwendungen ungeeignet ist. Daher wurden von vielen Seiten robustere Alternativen vorgeschlagen. Unter diesen Alternativen wird häufig der so genannte „Test  $W_{50}$ “ bevorzugt, da er äußerst gute Eigenschaften in Bezug auf Fehler 1. Art aufweist, aber dennoch einfach zu berechnen und leicht zu interpretieren ist. (Vergleichende Analysen finden sich in Conover et al. (1981), Balakrishnan und Ma (1990) sowie Lim und Loh (1996).)

Test  $W_{50}$  basiert auf einem Verfahren, das ursprünglich von Levene (1960) vorgeschlagen und später von Brown und Forsythe (1974) erweitert wurde. Test  $W_{50}$  wurde weithin übernommen und ist in den meisten gängigen statistischen Softwarepaketen wie der Minitab Statistical Software, SAS, R und JMP enthalten.

Die Eigenschaften von Test  $W_{50}$  in Bezug auf Fehler 2. Art sind etwas weniger bemerkenswert als seine Eigenschaften in Bezug auf Fehler 1. Art. Pan (1999) zeigt, dass die Trennschärfe von Test  $W_{50}$  bei einigen Verteilungen (u. a. für die Normalverteilung) und Fragestellungen mit zwei Stichproben eine Obergrenze aufweist, die potenziell weit unter 1 liegt. Darüber hinaus wird diese Obergrenze nicht durch die Größe der Differenz zwischen den Varianzen der Grundgesamtheiten beeinflusst. Diese Unzulänglichkeit gilt dementsprechend auch für KIs, die auf Test  $W_{50}$  basieren. Pan zeigt, dass eine nicht zu vernachlässigende Wahrscheinlichkeit besteht, dass ein KI für das Verhältnis der Varianzen der Grundgesamtheiten auf der Grundlage von Test  $W_{50}$  unendlich ( $0; +\infty$ ) ist und daher keine Aussagekraft besitzt. Pans Beobachtungen stimmen mit den Ergebnissen unserer eigenen Simulationen überein, die an späterer Stelle in dieser Veröffentlichung erörtert werden.

Pan schlägt ein alternatives Verfahren namens  $L_{50}$  vor, mit dem die Einschränkungen des  $W_{50}$ -Verfahrens korrigiert werden. Auf der Grundlage von Simulationsergebnissen kommt Pan zu dem Schluss, dass Test  $L_{50}$  trennschärfer als Test  $W_{50}$ , jedoch gleichermaßen robust ist und dessen erwünschte asymptotische Eigenschaften aufweist. Die Stichproben für die Simulationen von Pan wurden jedoch aus symmetrischen oder gemäßigt schiefen Verteilungen mit stärker bis schwächer besetzten Randbereichen gezogen. Die potenzielle Auswirkung der Schiefe auf die Leistung des  $L_{50}$ -Tests bei kleinen Stichproben wurde nicht gesondert dargelegt.

Pan behauptet zudem, dass das  $L_{50}$ -Verfahren ebenso trennscharf wie andere, ausgeprägt robuste Verfahren ist, z. B. der modifizierte Fligner-Killeen-Rangtest und der adaptive Hall-Padmanabhan-Test. In der Praxis sind jedoch der modifizierte Fligner-Killeen-Rangtest und der adaptive Hall-Padmanabhan-Test eher weniger nützlich als Test  $L_{50}$  und Test  $W_{50}$ , da sie umfangreiche Berechnungen voraussetzen.

Kürzlich wurde von Bonett (2006) ein alternatives KI-Verfahren vorgeschlagen, dem die Zwei-Stichproben-Version des Layard-Tests (1973) auf Homogenität der Varianzen zugrunde liegt. Bonett nahm mehrere Korrekturen vor, um die Leistung des Layard-Verfahrens in Bezug auf kleine Stichproben zu verbessern. So schlägt Bonett beispielsweise einen zusammengefassten Kurtosis-Schätzwert vor, der ein asymptotisches Äquivalent zu dem von Layard darstellt, jedoch bei kleinen Stichproben ein geringeres Ausmaß an Verzerrung aufweist.

Leider ist weder Layards ursprünglicher zusammengefasster Kurtosis-Schätzwert noch der von Bonett vorgeschlagene Ersatz konsistent, wenn die Varianzen der Grundgesamtheiten ungleich sind. Daher sind die von Bonett (2006) vorgeschlagenen Intervalle keine wahren KIs, sondern können besser als Annahmintervalle für den Test auf Gleichheit der Varianzen beschrieben werden. Somit ergeben sich die Wahrscheinlichkeiten eines Fehlers 1. Art des Tests auf Gleichheit der Varianzen, indem die von Bonett (2006) angegebenen simulierten Überdeckungswahrscheinlichkeiten von 1 subtrahiert werden. Ein Vergleich dieser Wahrscheinlichkeiten eines Fehlers 1. Art mit denen des ursprünglichen Tests von Layard bestätigt, dass Bonetts Korrekturen die Leistung des Layard-Tests in Bezug auf kleine

Stichproben tatsächlich verbessern. Das von Bonett vorgeschlagene KI für das Verhältnis der Varianzen bedarf jedoch einer Weiterentwicklung.

Beachten Sie auch, dass Bonett die vorgeschlagenen Intervalle mit KIs vergleicht, die auf dem F-Approximationstest nach Shoemaker (2003) basieren. Das KI für das Verhältnis der Varianzen für Shoemakers Test (wie kurz auf Seite 106 von Shoemakers Artikel erläutert) basiert jedoch ebenfalls auf dem zusammengefassten Kurtosis-Schätzwert von Layard. Daher können die in Abschnitt 7 von Shoemakers Artikel berechneten KIs ebenfalls am ehesten als Annahmintervalle für den Test auf Gleichheit der Varianzen beschrieben werden. Trotz dieser Fehler kann aus den Simulationsergebnissen von Bonett geschlossen werden, dass die Leistung des Layard-Tests auf Gleichheit der Varianzen in Bezug auf kleine Stichproben durch diese Korrektur verbessert wurde und dass der resultierende Test auf Gleichheit der Varianzen eine bessere Leistung als der Shoemaker-Test bietet.

Im vorliegenden Artikel korrigieren wir die Unzulänglichkeit in Bonett (2006) durch Erweitern der Zwei-Stichproben-Form des Layard-Tests auf einen Test von Nullhypothesen in Bezug auf das Verhältnis der Varianzen bzw. Standardabweichungen. Hierfür schlagen wir einen zusammengefassten Kurtosis-Schätzwert vor, der für beliebige hypothetische Verhältnisse konsistent ist. Anschließend wird die Teststatistik invertiert, um das KI für das Verhältnis zu erhalten. Schließlich werden Simulationsstudien durchgeführt, um die Eigenschaften des neuen KIs in Designs mit kleinen Stichproben in Bezug auf die Robustheit auszuwerten. Darüber hinaus wird die Leistung des neuen KI in Bezug auf kleine Stichproben mit der der KIs des klassischen F-Tests, von Test  $W_{50}$  sowie von Test  $L_{50}$  verglichen.

## 2. Layard-Test und eine Erweiterung

Angenommen,  $Y_{i1}, \dots, Y_{in_i}, \dots, Y_{k1}, \dots, Y_{kn_k}$  sind  $k$  unabhängige Stichproben, wobei jede Stichprobe unabhängig und mit dem Mittelwert  $E(Y_{ij}) = \mu_i$  und der Varianz  $\text{Var}(Y_{ij}) = \sigma_i^2 > 0$  identisch verteilt ist. Sei außerdem angenommen, dass die Stichproben aus Grundgesamtheiten mit der gemeinsamen Kurtosis  $\gamma = E(Y - \mu)^4 / \sigma^4 < \infty$  stammen. Wir stellen fest, dass Layard den Kurtosis-Exzess  $\gamma_e = \gamma - 3$  verwendet.

Seien  $\bar{Y}_i$  und  $S_i$  der Mittelwert und die Standardabweichung von Stichprobe  $i$ . Sei außerdem  $\tau^2 = 2 + (1 - 1/\bar{n})\gamma_e = 2 + (1 - 1/\bar{n})(\gamma - 3)$ , wobei  $\bar{n} = \sum n_i/k$ . Wie in Layard (1973) ausgeführt, gilt für große Stichproben  $\tau^2 \cong \text{Var}((n_i - 1)^{1/2} \ln S_i^2)$ .

Zum Testen der Nullhypothese der Gleichheit der Varianzen führt Layard eine orthogonale Transformation am Vektor durch, dessen Komponenten  $Z_i = (n_i - 1)^{1/2} \ln S_i^2 / \tau$  unter der Nullhypothese asymptotisch als Standardnormalverteilung verteilt sind. Anschließend weist er anhand der Abstandserhaltungseigenschaft orthogonaler Transformationen nach, dass die Teststatistik  $S'$  (unten angegeben) unter der Nullhypothese der Gleichheit der Varianzen asymptotisch als Chi-Quadrat-Verteilung mit  $k - 1$  Freiheitsgraden verteilt ist:

$$S' = \sum_{i=1}^k (n_i - 1) \left( \ln S_i^2 - \frac{\sum_{i=1}^k (n_i - 1) \ln S_i^2}{\sum_{i=1}^k (n_i - 1)} \right)^2 / \tau^2$$

Im Allgemeinen ist  $Z_i = (n_i - 1)^{\frac{1}{2}}(\ln S_i^2 - \ln \sigma_i^2)/\tau$  asymptotisch als Standardnormalverteilung verteilt. Daher können Layards Verfahren angewendet werden, um die stärker verallgemeinerte Teststatistik  $T'_k$  abzuleiten:

$$T'_k = \sum_{i=1}^k (n_i - 1) \frac{(\ln S_i^2 - \ln \sigma_i^2)^2}{\tau^2} - \left( \sum_{i=1}^k (n_i - 1) \frac{\ln S_i^2 - \ln \sigma_i^2}{\tau \sqrt{\sum_{i=1}^k (n_i - 1)}} \right)^2$$

$T'_k$  ist sowohl unter der Nullhypothese als auch unter der Alternativhypothese asymptotisch als Chi-Quadrat-Verteilung mit  $k - 1$  Freiheitsgraden verteilt.

$T'_k$  kann in einer Form ausgedrückt werden, die der von  $S'$  ähnlicher ist. Wird der quadrierte Term als doppelte Summe ausgedrückt, ergibt sich nach Umformen durch algebraische Operationen Folgendes:

$$T'_k = \sum_{i=1}^k (n_i - 1) \left( \ln S_i^2 - \ln \sigma_i^2 - \frac{\sum_{i=1}^k (n_i - 1) (\ln S_i^2 - \ln \sigma_i^2)}{\sum_{i=1}^k (n_i - 1)} \right)^2 / \tau^2$$

Wenn alle Varianzen gleich sind, dann ist  $T'_k = S'$ . Daher stellen  $S'$  und  $T'_k$  dieselbe Teststatistik dar, wenn die Nullhypothese der Gleichheit der Varianzen getestet wird.  $T'_k$  kann jedoch auch allgemeiner zum Testen beliebiger Hypothesen verwendet werden, die als Funktionen der Varianzen ausgedrückt werden. So kann beispielsweise mit  $T'_k$  eine beliebige Nullhypothese der Form  $H_0: \sigma_i = \sigma_{0i}$  für jedes gegebene  $\sigma_{0i} > 0, i = 1, \dots, k$  getestet werden.

Da  $\tau^2 = 2 + (1 - 1/\bar{n})(\gamma - 3)$  unbekannt ist, erfordert ein Test auf der Grundlage von  $S'$  oder  $T'_k$  einen Schätzwert für die gemeinsame Kurtosis der Grundgesamtheiten,  $\gamma$ . Zum Testen der Nullhypothese der Homogenität der Varianzen schlägt Layard z. B. den folgenden zusammengefassten Schätzwert für die gemeinsame Kurtosis vor:

$$\hat{\gamma} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^4}{\left[ \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right]^2} \sum_{i=1}^k n_i$$

Layard betont jedoch, dass  $\hat{\gamma}$  bei Ungleichheit der Varianzen nicht notwendigerweise ein konsistenter Schätzwert der gemeinsamen Kurtosis ist.

Im speziellen Fall von Designs mit zwei Stichproben kann die Größe der Differenz zwischen den Standardabweichungen durch Testen der Nullhypothese  $H_0: \sigma_1/\sigma_2 = \rho_0$  für ein gegebenes hypothetisches Verhältnis  $\rho_0 > 0$  beurteilt werden. Diese Differenz kann jedoch direkter beurteilt werden, indem das KI für das Verhältnis der Standardabweichungen berechnet wird.

Wenn  $\rho_0 = 1$ , dann stellt die Nullhypothese ein Äquivalent der Hypothese der Homogenität der Varianzen dar. Deshalb kann dem Test  $T'_2 = S'$  zugrunde gelegt werden, nachdem die Zwei-Stichproben-Version des Kurtosis-Schätzwerts von Layard im Ausdruck von  $\tau^2 = 2 + (1 - 1/\bar{n})(\gamma - 3)$  durch  $\gamma$  ersetzt wurde, um  $\hat{\tau}^2$  zu erhalten.

Wenn jedoch  $\rho_0 \neq 1$ , dann muss dem Test  $T'_2$  anstelle von  $S'$  zugrunde gelegt werden. Wenn  $\rho_0 \neq 1$ , dann ist der zusammengefasste Kurtosis-Schätzwert von Layard zudem nicht notwendigerweise konsistent, und er kann daher nicht verwendet werden, um die gemeinsame Kurtosis der Grundgesamtheiten zu schätzen. Daher wird ein alternativer

zusammengefasster Kurtosis-Schätzwert benötigt, der für jedes beliebige hypothetische Verhältnis  $\rho_0 > 0$  konsistent ist.

Ein solcher Schätzwert wird nun abgeleitet. Da es sich hierbei um eine Funktion von  $\rho_0$  handelt, wird der Schätzwert als  $\hat{\gamma}_P(\rho_0)$  angegeben. Zudem wird die Teststatistik  $T_2 = \tau^2 T_2' / \hat{\tau}^2$  definiert, wobei  $\hat{\tau}^2 = 2 + (1 - 1/\bar{n})(\hat{\gamma}_P(\rho_0) - 3)$ . Gemäß dem Theorem von Slutsky ist  $T_2$  asymptotisch als Chi-Quadrat-Verteilung mit 1 Freiheitsgrad verteilt. Schließlich wird  $T_2$  invertiert, um KIs für  $\rho = \sigma_1/\sigma_2$  zu erhalten.

### 3. KI für das Verhältnis der Standardabweichungen

Im vorhergehenden Abschnitt wird die Notwendigkeit eines alternativen Kurtosis-Schätzwertes beim Testen von Nullhypothesen erläutert, die als Verhältnis der Varianzen oder Standardabweichungen ausgedrückt werden. Das folgende Ergebnis liefert einen solchen Schätzwert.

#### ERGEBNIS 1

Für jedes gegebene  $\rho = \sigma_1/\sigma_2 > 0$  kann ein konsistenter zusammengefasster Kurtosis-Schätzwert der gemeinsamen Kurtosis der Grundgesamtheiten im Modell mit zwei Stichproben angegeben werden als

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2}$$

Den Beweis für dieses Ergebnis finden Sie in Anhang A.

Erwartungsgemäß ist  $\hat{\gamma}_P(1)$  identisch mit dem zusammengefassten Kurtosis-Schätzwert von Layard,  $\hat{\gamma}$ , da  $\sigma_1/\sigma_2 = 1$  impliziert, dass die Standardabweichungen (oder Varianzen) gleich sind.

Die Statistik  $T_2'$ , die die Zwei-Stichproben-Version der allgemeinen Statistik  $T_k'$  darstellt, wird angegeben als

$$T_2' = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}\right) \tau^2}$$

Hierbei ist  $\hat{\rho} = S_1/S_2$ ;  $\rho = \sigma_1/\sigma_2$  und  $\tau^2 = 2 + (1 - 1/\bar{n})\hat{\gamma}_e = 2 + (1 - 1/\bar{n})(\gamma - 3)$ .

Wie in Layard (1973) ausgeführt, gilt bei großen Stichproben  $\tau^2 \cong \text{Var}((n_i - 1)^{1/2} \ln S_i^2)$ . Bonett (2006) verwendet eine alternative Approximation, die auch von Shoemaker (2003) übernommen wird:  $\text{Var}((n_i - 1)^{1/2} \ln S_i^2) \cong \gamma - (n_i - 3)/n_i$ . Bei großen Stichproben sind diese Approximationen äquivalent. Shoemaker bemerkt jedoch, dass die letztere Version von Vorteil ist, wenn sein Test auf Gleichheit der Varianzen mit kleinen Stichproben verwendet wird. Mit dieser Korrektur kann die Statistik  $T_2'$  wie folgt abgewandelt werden:

$$T_2' = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\frac{\gamma - g_1}{n_1 - 1} + \frac{\gamma - g_2}{n_2 - 1}}$$

Hierbei ist  $g_i = (n_i - 3)/n_i$ .

Daraus folgt, dass die Teststatistik  $T_2 = \tau^2 T_2' / \hat{\tau}^2$  zum Testen der Nullhypothese  $H_0: \rho = \rho_0$  angegeben wird als

$$T_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho_0^2)^2}{\frac{\hat{\gamma}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho_0) - g_2}{n_2 - 1}}$$

In diesem Ausdruck von  $T_2$  kann die Quadratwurzel des Nenners als Schätzwert des Standardfehlers der zusammengefassten Kurtosis für große Stichproben angesehen werden.

Darüber verwendet Bonett (2006) im Ausdruck von  $\hat{\gamma}_P(1) \equiv \hat{\gamma}$  die getrimmten Stichprobenmittelwerte mit dem Trim-Anteil  $1/[2(n_i - 4)^{1/2}]$ . Demzufolge nehmen wir die gleiche Korrektur am zusammengefassten Kurtosis-Schätzwert vor:

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - m_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - m_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2}$$

Hierbei ist  $m_i$  der getrimmte Mittelwert für Stichprobe  $i$  mit dem Trim-Anteil  $1/[2(n_i - 4)^{1/2}]$ . Diese Version des zusammengefassten Kurtosis-Schätzwerts und die frühere Version sind asymptotisch äquivalent, da der getrimmte Mittelwert  $m_i$  ein konsistenter Schätzwert des Mittelwerts der Grundgesamtheit  $\mu_i$  ist. Diese alternative Version kann jedoch die Leistung des Tests auf der Grundlage von  $T_2$  in Bezug auf kleine Stichproben verbessern.

Die Teststatistik  $T_2$  kann nun invertiert werden, um ein approximiertes KI für das Verhältnis der Varianzen oder Standardabweichungen abzuleiten. Zunächst soll jedoch kurz die Unzulänglichkeit bei der Ableitung der Bonett-KIs (2006) für das Verhältnis der Standardabweichungen beschrieben werden.

### 3.1 Bonett-Intervalle

Anstatt  $T_2$  zu invertieren, um das KI zu erhalten, invertiert Bonett (2006) die folgende Statistik:

$$T = \frac{(\ln \hat{\rho}^2 - \ln \rho^2)^2}{\frac{\hat{\gamma}_P(1) - g_1}{n_1 - 1} + \frac{\hat{\gamma}_P(1) - g_2}{n_2 - 1}}$$

Daher ist das resultierende Intervall lediglich der Annahmebereich für den Test auf Gleichheit der Varianzen. Dies ist darauf zurückzuführen, dass der zusammengefasste Kurtosis-Schätzwert  $\hat{\rho}_P(1)$  nur dann konsistent ist, wenn die Varianzen gleich sind bzw. (und äquivalent) wenn das hypothetische Verhältnis 1 ist. Das resultierende Intervall wird in Bonett (2006) wie folgt angegeben:

$$\exp[\ln(c S_1^2/S_2^2) \pm z_{\alpha/2} se]$$

Dabei gilt Folgendes:

$$se^2 = \frac{\hat{\gamma}(1) - g_1}{n_1 - 1} + \frac{\hat{\gamma}(1) - g_2}{n_2 - 1}$$

Die Konstante  $c$  wird als Korrektur für kleine Stichproben eingebunden, um den Effekt ungleicher Fehlerwahrscheinlichkeiten in den Randbereichen von unbalancierten Designs zu mindern. Diese Konstante wird angegeben durch

$$c = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

Die Konstante verschwindet, wenn die Designs balanciert sind, und ihre Auswirkung wird bei zunehmenden Stichprobenumfängen vernachlässigbar.

Tabelle 1 veranschaulicht die Folgen einer Fehlinterpretation der obigen Intervalle als KIs. Diese Ergebnisse basieren auf einer kleinen Simulationsstudie, in der simulierte Überdeckungswahrscheinlichkeiten auf der Grundlage von Bonett-Intervallen (2006) berechnet werden. Für die Fälle mit gleichen Varianzen (linke Spalte) werden zwei unabhängige Stichproben aus der Standardnormalverteilung gezogen. Für die Fälle mit ungleichen Varianzen (rechte Spalte) werden die Beobachtungen der zweiten Stichprobe mit dem konstanten Faktor 4 skaliert. Die geschätzten Überdeckungswahrscheinlichkeiten basieren auf 100.000 Replikationen. Die nominale Sollüberdeckung ist 0,95.

**Tabelle 1** Auswirkung ungleicher Varianzen der Grundgesamtheiten auf KIs nach Bonett (2006) ( $\alpha = 0,05$ )

$n_1; n_2$	Simulierte Überdeckungswahrscheinlichkeiten	
	Gleiche Varianzen	Ungleiche Varianzen
10; 10	0,963	0,972
50; 50	0,952	0,991
100; 100	0,952	0,994

Wenn die Intervalle auf einem konsistenten zusammengefassten Kurtosis-Schätzwert basieren würden, könnte erwartet werden, dass die Überdeckungswahrscheinlichkeiten in den beiden Fällen identisch sind. Es lässt sich jedoch feststellen, dass die Intervalle bei ungleichen Varianzen durchgehend konservativer sind. Darüber hinaus nähern sich die Überdeckungswahrscheinlichkeiten bei zunehmenden Stichprobenumfängen 1 an. Beachten Sie, dass mit den approximierten KIs nach Shoemaker (2003) ähnliche Ergebnisse erzielt werden.

## 3.2 Berechnungen für das KI

Betrachten Sie die Fragestellung, bei der die Nullhypothese  $H_0: \rho = \rho_0$  gegen die Alternativhypothese  $H_A: \rho \neq \rho_0$  getestet werden soll, wobei  $\rho = \sigma_1/\sigma_2$  und  $\rho_0 > 0$ ; Grundlage ist die zuvor angegebene Teststatistik  $T_2$ . Unter der Nullhypothese ist die Teststatistik

$$T_2 = \frac{(\ln \hat{\rho}^2 - \ln \rho_0^2)^2}{\frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1}}$$

asymptotisch in einer Chi-Quadrat-Verteilung mit 1 Freiheitsgrad verteilt. Deshalb weist der Test die Nullhypothese beim Signifikanzniveau  $\alpha$  nur dann zurück, wenn

$$(\ln \hat{\rho}^2 - \ln \rho_0^2)^2 > z_{\alpha/2}^2 \left( \frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1} \right)$$

Hierbei ist  $z_{\alpha}$  der obere  $\alpha \times 100$ . Perzentilpunkt der Standardnormalverteilung. Beachten Sie, dass der obere  $\alpha \times 100$ . Perzentilpunkt der Chi-Quadrat-Verteilung mit 1 Freiheitsgrad,  $\chi_{1,\alpha}^2$ , die folgende Bedingung erfüllt:  $\chi_{1,\alpha}^2 = z_{\alpha/2}^2$ .

Die Simulationsergebnisse von Bonett (2006) zeigen, dass die Korrektur für kleine Stichprobenumfänge zum Mindern der Auswirkung ungleicher Fehlerwahrscheinlichkeiten in den Randbereichen von unbalancierten Designs gut funktioniert. Daher nehmen wir eine ähnliche Korrektur für den Test auf Grundlage von  $T_2$  vor. Bei dieser Korrektur weist der Test die Nullhypothese nur dann zurück, wenn

$$(\ln \rho_0^2 - \ln(c\hat{\rho}^2))^2 > z_{\alpha/2}^2 \left( \frac{\hat{Y}_P(\rho_0) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho_0) - g_2}{n_2 - 1} \right)$$

Hierbei ist  $c$  die Korrekturkonstante von Bonett, angegeben als

$$c = \frac{n_1}{n_1 - z_{\alpha/2}} \frac{n_2 - z_{\alpha/2}}{n_2}$$

Äquivalent wird eine approximierte Konfidenzmenge von  $(1 - \alpha)100$  Prozent für  $\rho = \sigma_1/\sigma_2$  auf der Grundlage von  $T_2$ , angegeben als

$$\left\{ \rho \in (0; \infty): (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left( \frac{\hat{Y}_P(\rho) - g_1}{n_1 - 1} + \frac{\hat{Y}_P(\rho) - g_2}{n_2 - 1} \right) \leq 0 \right\}$$

Beachten Sie, dass  $c$  in balancierten Designs keine Auswirkung hat und die Auswirkung in unbalancierten Designs mit großen Stichproben vernachlässigt werden kann.

Das nächste Ergebnis liefert einen alternativen Ausdruck der Konfidenzmenge, und zwar in einer Form, in der ihr Wesen anschaulicher beschrieben werden kann. In diesem Ausdruck wird der zusammengefasste Kurtosis-Schätzwert in Form der Kurtosis-Werte für die einzelnen Stichproben umgeschrieben, die angegeben werden als

$$\hat{\gamma}_i = n_i \frac{\sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{[(n_i - 1)S_i^2]^2}, i = 1; 2$$

## ERGEBNIS 2

Eine approximierte Konfidenzmenge von  $(1 - \alpha)100$  Prozent für  $\rho = \sigma_1/\sigma_2$  auf der Grundlage von  $T_2$  kann ausgedrückt werden als

$$\hat{\rho}\sqrt{c} \{r \in (0; \infty): H(r^2) \leq 0\}$$

Äquivalent kann die Konfidenzmenge für  $\rho^2 = \sigma_1^2/\sigma_2^2$  ausgedrückt werden als

$$c\hat{\rho}^2 \{r \in (0; \infty): H(r) \leq 0\}$$

Dabei gilt Folgendes:

$$H(x) = (\ln x)^2 - z_{\alpha/2}^2 se^2(cx), x > 0$$

$$se^2(x) = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 x^2/n_2}{(K + x)^2} - B$$



$$A = \frac{(n_1 + n_2)(n_1 + n_2 - 2)}{(n_1 - 1)(n_2 - 1)}, B = \frac{g_1}{n_1 - 1} + \frac{g_2}{n_2 - 1}, K = \frac{n_1 - 1}{n_2 - 1}$$

Einen Beweis dieses Ergebnisses finden Sie in Anhang B.

Es kann einfach nachgewiesen werden, dass die Funktion  $H(x)$  auf der positiven reellen Linie stetig ist, mit  $H(0) = H(+\infty) = +\infty$  und  $H(1) < 0$ . Nach dem Zwischenwertsatz weist die Funktion  $H(x)$  also mindestens eine Nullstelle im Intervall  $(0; 1)$  und mindestens eine Nullstelle im Intervall  $(0; +\infty)$  auf.

Im nächsten Ergebnis wird die Konfidenzmenge als Intervall oder Vereinigung disjunkter Intervalle beschrieben.

### ERGEBNIS 3

Wenn die Funktion  $H(x)$  genau zwei Nullstellen aufweist,  $x_L$  und  $x_U$ , dann gilt  $0 < x_L < 1 < x_U$ , und die Konfidenzmenge für  $\rho^2 = \sigma_1^2/\sigma_2^2$  ist das Intervall, das angegeben wird durch

$$[c\hat{\rho}^2 x_L, c\hat{\rho}^2 x_U]$$

Daraus folgt, dass das KI für  $\rho = \sigma_1/\sigma_2$  gleich diesem Intervall ist:

$$[\hat{\rho}\sqrt{cx_L}, \hat{\rho}\sqrt{cx_U}]$$

Wenn die Funktion  $H(x)$  hingegen über mehr als zwei Nullstellen verfügt, ist die Konfidenzmenge für  $\rho^2 = \sigma_1^2/\sigma_2^2$  die Vereinigung von nicht überlappenden Intervallen. Die Endpunkte der einzelnen Intervalle sind die aufeinanderfolgenden Nullstellen, bei denen sich die Funktion nach oben öffnet.

Einen Beweis dieses Ergebnisses finden Sie in Anhang C.

### ANMERKUNG

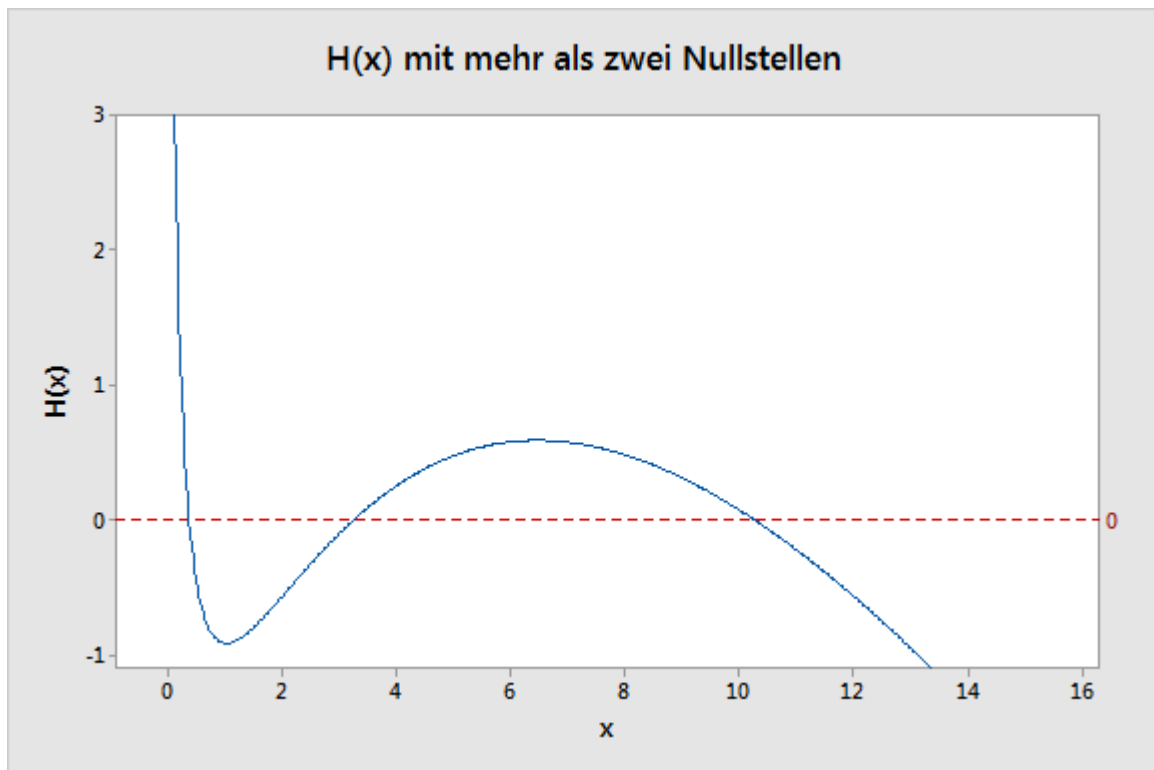
Zwar ist es mathematisch möglich, dass die Funktion  $H(x)$  zwei oder mehr Nullstellen aufweist, unsere Beobachtungen zeigen jedoch, dass dies nur bei extrem ungewöhnlichen und praktisch sinnlosen Designs auftritt, bei denen eine oder mehrere Stichproben entweder zu klein oder stark unbalanciert sind. Wir vermuten, dass  $H(x)$  entweder zwei oder vier Nullstellen aufweist.

Das folgende Beispiel basiert auf Daten, die gezielt so erstellt wurden, dass die Funktion  $H(x)$  über mehr als zwei Nullstellen verfügt. Die Daten werden wie folgt zusammengefasst:  $n_1 = 169$ ;  $n_2 = 7$ ;  $S_1 = 301,855$ ;  $S_2 = 4606,170$ ;  $\hat{\gamma}_1 = 1,877$ ;  $\hat{\gamma}_2 = 6,761$ ;  $c = 0,728$ ;  $A = 30,381$ ;  $B = 0,101$  und  $K = 28,000$ .

Für  $\alpha = 0,05$  wird die Funktion  $H(x)$  angegeben als

$$H(x) = (\ln x)^2 - 1,960^2 \left( 30,381 \frac{1,877 \times 28^2/169 + 6,761 \times (0,728x)^2/7}{(28,000 + 0,728x)^2} - 0,101 \right)$$

Die Funktion  $H(x)$  weist in diesem Fall vier Nullstellen auf. Das Diagramm der Funktion ist unten abgebildet. Beachten Sie, dass die vierte Nullstelle im Diagramm nicht sichtbar ist, da sie zu groß ist. Es ist jedoch bekannt, dass die vierte Nullstelle vorhanden ist, weil  $H(+\infty) = +\infty$ .



Die vier Nullstellen werden numerisch als  $x_1 = 0,389$ ;  $x_2 = 3,282$ ;  $x_3 = 10,194$  und  $x_4 = 39685,0$  berechnet. Das geschätzte Verhältnis der Standardabweichungen ist  $\hat{\rho} = S_1/S_2 = 0,066$ . Die Konfidenzmenge für  $\rho^2 = \sigma_1^2/\sigma_2^2$  kann ausgedrückt werden als

$$[c \hat{\rho}^2 x_1, c \hat{\rho}^2 x_2] \cup [c \hat{\rho}^2 x_3, c \hat{\rho}^2 x_4] = [0,001; 0,010] \cup [0,032; 124,072]$$

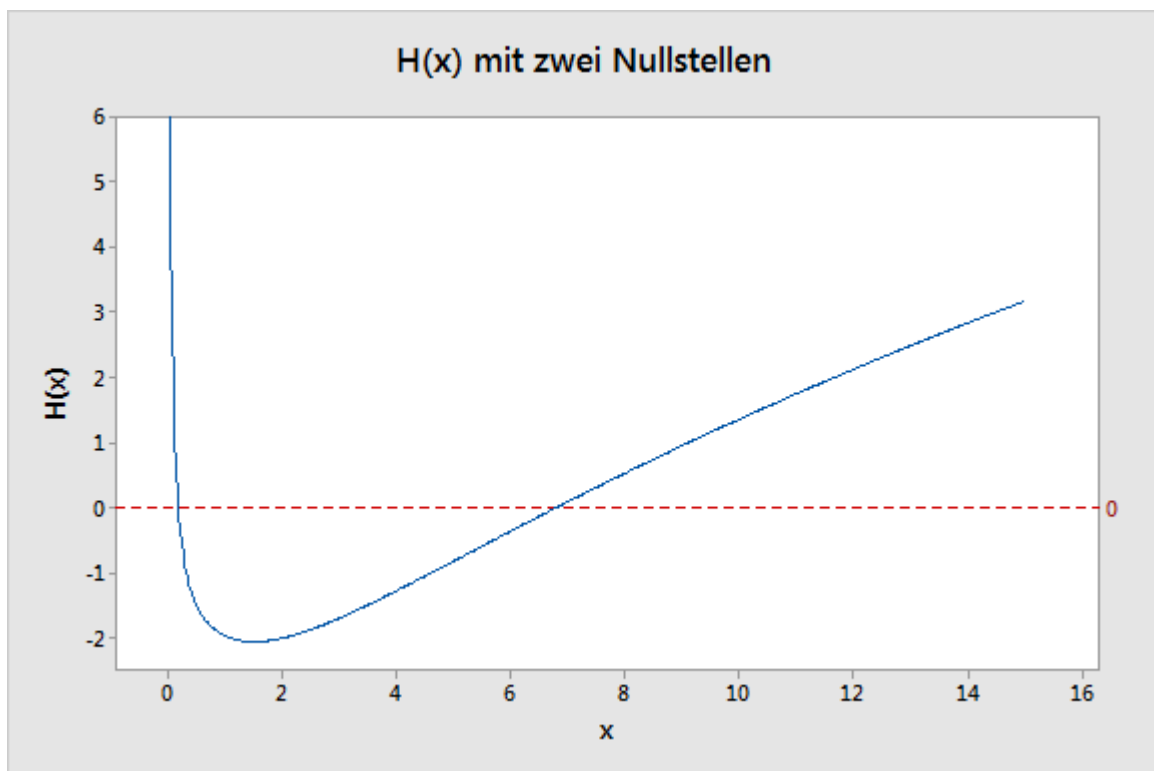
Die Konfidenzmenge für das Verhältnis der Standardabweichungen,  $\rho$ , wird durch Ziehen der Quadratwurzel der Endpunkte der Intervalle bestimmt.

Wenn die Stichproben nicht zu klein ( $n_i \geq 10$ ) und ihre Stichprobenumfänge nicht zu unterschiedlich sind, weist die Funktion  $H(x)$  normalerweise zwei Nullstellen auf. Eine Nullstelle liegt unter 1, und die andere Nullstelle liegt über 1, wie in Ergebnis 2 beschrieben. Das folgende Beispiel beruht auf zufällig erzeugten Daten. Die Daten können wie folgt zusammengefasst werden:  $n_1 = 10$ ;  $n_2 = 12$ ;  $S_1 = 1,150$ ;  $S_2 = 1,043$ ;  $\hat{\gamma}_1 = 2,704$ ;  $\hat{\gamma}_2 = 3,671$ ;  $c = 1,041$ ;  $A = 4,444$ ;  $B = 0,146$  und  $K = 0,818$ .

Für  $\alpha = 0.05$  wird die Funktion  $H(x)$  in diesem Fall angegeben als

$$H(x) = (\ln x)^2 - 1,960^2 \left( 4,444 \frac{2,704 \times 0,818^2/10 + 3,671 \times (1,041x)^2/12}{(0,818 + 1,041x)^2} - 0,146 \right)$$

Wie unten veranschaulicht, verfügt die Funktion  $H(x)$  über zwei Nullstellen:



Die zwei Nullstellen werden numerisch als  $x_1 = 0,200$  und  $x_2 = 6,824$  berechnet. Das geschätzte Verhältnis der Standardabweichungen ist  $\hat{\rho} = S_1/S_2 = 1,102$ . Die Konfidenzmenge für  $\rho^2 = \sigma_1^2/\sigma_2^2$  ist das Intervall, das angegeben wird als

$$[c \hat{\rho}^2 x_1, c \hat{\rho}^2 x_2] = [0,253; 8,634]$$

Das KI für das Verhältnis der Standardabweichungen,  $\rho$ , wird durch Ziehen der Quadratwurzel der Endpunkte des obigen Intervalls bestimmt.

Nun werden die zwei Algorithmen zum Bestimmen der Konfidenzgrenzen beschrieben.

Der erste Algorithmus besteht aus einem numerischen Verfahren für die Nullstellenbestimmung, mit dem die Nullstellen der Funktion  $H(x)$  bestimmt werden. Die Nullstelle, die einer unteren Konfidenzgrenze für das Verhältnis der Varianzen entspricht, ist im Intervall  $(0; 1)$  eingeschlossen. Wird diese Nullstelle mit  $x_L$  angegeben, wird die untere Konfidenzgrenze für das Verhältnis der Varianzen gemäß Ergebnis 3 als  $c \hat{\rho}^2 x_L$  berechnet, und die untere Konfidenzgrenze für das Verhältnis der Standardabweichungen wird als  $\hat{\rho} \sqrt{c x_L}$  berechnet. Dementsprechend ist die obere Konfidenzgrenze für das Verhältnis der Varianzen  $c \hat{\rho}^2 x_U$ , und die obere Konfidenzgrenze für das Verhältnis der Standardabweichungen ist  $\hat{\rho} \sqrt{c x_U}$ , wobei  $x_U > 1$  die andere Nullstelle von  $H(x)$  ist. Eine einfache Vorgehensweise beim Bestimmen der oberen Konfidenzgrenze besteht darin, den Umstand zu nutzen, dass die Untergrenze für  $1/\rho^2$  gleich der Obergrenze für  $\rho^2$  ist. Zunächst werden die Rollen der ersten Stichprobe und der zweiten Stichprobe im Ausdruck der Funktion  $H(x)$  ausgetauscht, als ob die Konfidenzgrenze für das Verhältnis  $1/\rho^2 = \sigma_2^2/\sigma_1^2$  berechnet werden würde. Zweitens wird der Algorithmus zum Bestimmen der Untergrenze auf die neue Funktion  $H(x)$  angewendet. Die resultierende Grenze schließlich wird invertiert, um die gewünschte obere Konfidenzgrenze zu bestimmen.

Ein alternativer Ansatz besteht darin, die untere Konfidenzgrenze für das Verhältnis der Varianzen rekursiv anhand der Rekursionsgleichung zu berechnen, die angegeben ist durch

$$\rho_0^2 = 1$$

$$\rho_{i+1}^2 = \exp \left[ \ln(c \hat{\rho}^2) - z_{\alpha/2} \sqrt{\frac{\hat{y}_P(\rho_i) - g_1}{n_1 - 1} + \frac{\hat{y}_P(\rho_i) - g_2}{n_2 - 1}} \right], i = 0; 1; 2; \dots$$

Die untere Konfidenzgrenze für das Verhältnis der Varianzen ist  $\rho_{j+1}^2$ , so dass  $|\rho_{j+1}^2 - \rho_j^2| < \varepsilon$ , wobei  $j > 0$  und ein kleines  $\varepsilon$  gewählt wird (z. B.  $\varepsilon = 10^{-6}$ ). Zum Bestimmen der oberen Konfidenzgrenze wird in der obigen Gleichung einfach  $-z_{\alpha/2}$  durch  $+z_{\alpha/2}$  ersetzt.

Offensichtlich sind die zwei Algorithmen zum Berechnen der Konfidenzgrenzen äquivalent, da das Rekursivverfahren seinem Wesen nach ein iteratives Verfahren zum Auflösen der Gleichung  $H(\rho^2/(c\hat{\rho}^2)) = 0$  für  $\rho^2$  ist. Der rekursive Algorithmus ist einfacher zu implementieren und ist daher eine nützliche Alternative, wenn kein Verfahren für die Nullstellenbestimmung verfügbar ist.

## 4. Simulationsstudien und Ergebnisse

In diesem White Paper wurde ein Verfahren zum Erweitern des Layard-Tests auf Gleichheit zweier Varianzen abgeleitet, um das Verhältnis der Varianzen zu prüfen. Wir bezeichnen dieses Verfahren als erweiterten Layard-Verhältnistest bzw. ELTR (Extended Layard's Test for the Ratio). In diesem Abschnitt werden die Eigenschaften von KIs auf der Grundlage des ELTR-Verfahrens in Bezug auf kleine Stichproben untersucht. Dabei wird der allgemeine Ansatz von Bonett (2006) verfolgt.

Wir vergleichen KIs, denen das ELTR-Verfahren zugrunde liegt, mit KIs, die auf Test  $L_{50}$  (Pan, 1999) und Test  $W_{50}$  (dem Levene/Brown-Forsythe-Test) basieren. Für Studie 1 berücksichtigen wir zu Vergleichszwecken auch KIs, die auf dem klassischen F-Test basieren. Es ist hinreichend bekannt, dass der klassische F-Test bei einer Normalverteilung der Daten optimal ist. Beachten Sie, dass die Berechnungen für KIs, die auf Test  $W_{50}$  und Test  $L_{50}$  basieren, in Pan (1999) enthalten sind. Die Berechnungen für KIs, die auf dem F-Test basieren, finden sich in vielen einführenden statistischen Fachbüchern. Eine Quelle ist auch Bonett (2006).

Wir haben drei Simulationsstudien durchgeführt, jeweils mit 100.000 Stichprobenreplikationen. Jede Replikation besteht aus zwei unabhängigen Stichproben mit kleinem bis gemäßigttem Stichprobenumfang. Jede Stichprobe wurde aus einer übergeordneten Grundgesamtheit mit bekannten Eigenschaften gezogen; hierzu zählen Symmetrie, Asymmetrie, stärker besetzte Randbereiche und schwächer besetzte Randbereiche. Der Standardfehler für jede Simulation beträgt annähernd 0,0009; 0,0007 und 0,0003 für nominale Konfidenzniveaus von 90 %, 95 % bzw. 99 %.

Um die Leistung der einzelnen Verfahren auszuwerten, geben wir die erzielte Überdeckungswahrscheinlichkeit und die mittlere Breite der simulierten Intervalle für das Verhältnis der Varianzen an. Einige der Intervalle für Test  $W_{50}$  wiesen eine unendliche Breite

auf (eine Möglichkeit, auf die von Pan (1999) hingewiesen wurde). In solchen Fällen geben wir sowohl die mittlere Breite der endlichen Intervalle als auch den Prozentsatz der Intervalle mit unendlicher Breite aus. Alle Simulationen wurden mit Version 8 des Mathematica-Softwarepakets durchgeführt.

## Studie 1: Vergleich der Überdeckungswahrscheinlichkeiten für normalverteilte Daten

In der ersten Studie werden zufällige Stichproben mit unterschiedlichem Stichprobenumfang aus der Normalverteilung generiert. Die Ergebnisse werden in Tabelle 2 aufgeführt.

**Tabelle 2** Vergleich der Überdeckungswahrscheinlichkeiten und durchschnittlichen Intervallbreiten

$1 - \alpha$	$n_1; n_2$	Maß	Verfahren			
			F	ELTR	$L_{50}$	$W_{50}$
0,90	10; 10	Überdeckung	0,898	0,918	0,913	0,921
		Breite	3,72	5,06	4,72	8,03 (0,01 %)
	30; 10	Überdeckung	0,90	0,909	0,897	0,911
		Breite	2,42	3,01	3,58	3,17
	25; 25	Überdeckung	0,902	0,907	0,914	0,916
		Breite	1,61	1,73	1,85	1,938
	50; 50	Überdeckung	0,90	0,901	0,906	0,907
		Breite	1,03	1,06	1,13	1,15
0,95	10; 10	Überdeckung	0,949	0,963	0,958	0,964
		Breite	4,90	7,72	6,52	497,24 (0,20 %)
	30; 10	Überdeckung	0,950	0,957	0,945	0,959
		Breite	2,98	4,91	4,67	4,07
	25; 25	Überdeckung	0,951	0,955	0,958	0,961
		Breite	1,99	2,24	2,31	2,49
	50; 50	Überdeckung	0,951	0,952	0,953	0,954
		Breite	1,25	1,31	1,38	1,41

$1 - \alpha$	$n_1; n_2$	Maß	Verfahren			
			F	ELTR	$L_{50}$	$W_{50}$
0,99	10; 10	Überdeckung	0,989	0,993	0,992	0,994
		Breite	8,29	17,76	12,52	$> 10^4$ (8,8 %)
	30; 10	Überdeckung	0,990	0,992	0,986	0,994
		Breite	4,26	15,76	8,26	6,77
	25; 25	Überdeckung	0,990	0,992	0,992	0,993
		Breite	2,86	3,66	3,43	4,03
	50; 50	Überdeckung	0,990	0,991	0,991	0,991
		Breite	1,71	1,89	1,92	2,02

In den dunkel unterlegten Zeilen werden die erzielten Überdeckungswahrscheinlichkeiten (Überdeckung) für die einzelnen Verfahren für jedes Konfidenzniveau ( $1 - \alpha$ ) und jede Kombination von Stichprobenumfängen ( $n_1, n_2$ ) angegeben. Der Mittelwert der Intervallbreiten (Breite) wird unter den einzelnen Überdeckungswahrscheinlichkeiten angegeben. Wenn Intervalle für eine Bedingung unendlich waren, werden sowohl der Mittelwert für die endlichen Intervalle als auch der Prozentsatz der Intervalle angegeben, die unendlich waren.

Die Ergebnisse zeigen erwartungsgemäß, dass die KIs für das F-Verfahren die größte Genauigkeit und Präzision aufweisen. Die mit dem F-Verfahren erzielten Überdeckungswahrscheinlichkeiten liegen näher an der Sollüberdeckung als diejenigen, die mit den anderen Verfahren erzielt wurden. Die durchschnittlichen Breiten der Intervalle für das F-Verfahren sind kleiner als diejenigen der anderen Verfahren. Die Tabelle zeigt jedoch auch, dass mit dem ELTR-Verfahren und dem  $L_{50}$ -Verfahren erstellte KIs beinahe so genau und präzise wie diejenigen sind, denen das F-Verfahren zugrunde liegt.

Die auf Test  $W_{50}$  basierenden Intervalle sind ebenfalls relativ genau.  $W_{50}$ -Intervalle können jedoch sehr breit sein und je nach Umfang der Stichproben sogar eine unendliche Breite aufweisen. Beachten Sie, dass wenn für beide Stichproben nur 10 Beobachtungen vorliegen, mindestens 0,01 % der vom  $W_{50}$ -Verfahren erzeugten Intervalle unendlich breit sind. Der Prozentsatz der unendlichen Intervalle steigt mit zunehmender Sollüberdeckung. Unter den meisten Bedingungen sind die mittleren Breiten der ELTR-Intervalle und der  $L_{50}$ -Intervalle kleiner als die mittleren Breiten der  $W_{50}$ -Intervalle.

## Studie 2: Vergleich der Überdeckungswahrscheinlichkeiten für nicht normalverteilte Daten

In der zweiten Studie sollen die Leistungen des ELTR-Verfahrens, des  $L_{50}$ -Verfahrens und des  $W_{50}$ -Verfahrens ausgewertet und miteinander verglichen werden, wenn die übergeordneten Grundgesamtheiten nicht normalverteilt sind. Außerdem wird zusätzlich eine kontaminierte

Normalverteilung betrachtet, um die Auswirkung von Ausreißern auf die Leistung der Verfahren zu beurteilen. Diese kontaminierte Verteilung wird als CN(0,1;3) bezeichnet, um anzugeben, dass 90 % der Beobachtungen aus der Standardnormalverteilung und die verbleibenden 10 % aus einer normalverteilten Grundgesamtheit mit dem Mittelwert 0 und der Standardabweichung 3 gezogen wurden. Die Ergebnisse werden in Tabelle 3 vorgestellt.

**Tabelle 3** Vergleich der Überdeckungswahrscheinlichkeiten und durchschnittlichen Intervallbreiten in einigen nicht normalverteilten Modellen; das nominale Konfidenzniveau ist  $1 - \alpha = 0,95$

Verteilung [ $\gamma$ ] $n_1; n_2$	ELTR	$L_{50}$	$W_{50}$	Verteilung [ $\gamma$ ] $n_1; n_2$	ELTR	$L_{50}$	$W_{50}$
<b>Gleich- verteilung</b>				$\chi^2(5)$			
[1,8]				[5,4]			
10; 10	0,971	0,971	0,966	10; 10	0,956	0,938	0,956
	5,27	4,87	42,08 (0,1 %)		11,61	8,78	> 10 <sup>4</sup> (2,6 %)
10; 30	0,964	0,961	0,957	10; 30	0,959	0,923	0,956
	2,51	2,4	2,89		6,25	4,14	190,645 (0,3 %)
25; 25	0,967	0,972	0,968	25; 25	0,956	0,944	0,954
	1,43	1,79	1,88		3,66	2,92	3,26
50; 50	0,959	0,962	0,959	50; 50	0,959	0,946	0,952
	0,83	1,06	1,08		2,07	1,7	1,77
<b>Beta (3; 3)</b>				<b>Exponential</b>			
[2,5]				[9]			
10; 10	0,968	0,966	0,966	10; 10	0,947	0,916	0,950
	6,26	5,59	254,62 (0,1 %)		20,99	14,47	> 10 <sup>4</sup> (9,1 %)
10; 30	0,960	0,954	0,960	10; 30	0,954	0,896	0,953
	3,14	2,76	3,71		10,46	6,19	> 10 <sup>4</sup> (4,1 %)
25; 25	0,959	0,966	0,965	25; 25	0,956	0,931	0,951
	1,81	2,06	2,18		6,09	4,13	5,48 (0,008 %)
50; 50	0,957	0,959	0,958	50; 50	0,962	0,942	0,952
	1,06	1,23	1,26		3,18	2,24	2,38

Verteilung [ $\gamma$ ] $n_1; n_2$	ELTR	$L_{50}$	$W_{50}$	Verteilung [ $\gamma$ ] $n_1; n_2$	ELTR	$L_{50}$	$W_{50}$
<b>Laplace</b>				$\chi^2(1)$			
[6]				[15]			
10; 10	0,946	0,935	0,961	10; 10	0,928	0,889	0,947
	13,47	10,45	$> 10^4$ (3,0 %)		55,09	37,4	$> 10^5$ (25,1 %)
10; 30	0,947	0,919	0,957	10; 30	0,943	0,882	0,956
	6,78	4,82	$> 10^4$ (0,4 %)		18,71	11,14	$> 10^6$ (25,7 %)
25; 25	0,945	0,940	0,952	25; 25	0,952	0,925	0,954
	4,00	3,372	3,86		10,97	6,84	$> 10^4$ (0,4 %)
50; 50	0,952	0,949	0,955	50; 50	0,958	0,936	0,951
	2,19	1,91	1,99		5,08	3,31	3,75 (0,001 %)
$t(5)$				<b>Lognormal</b>			
[9]				[113,9]			
10; 10	0,957	0,946	0,965	10; 10	0,923	0,876	0,955
	11,07	8,81	$> 10^3$ (2,0 %)		59,22	46,15	$> 10^5$ (23,0 %)
10; 30	0,957	0,930	0,959	10; 30	0,949	0,866	0,958
	6,06	4,24	$> 10^3$ (0,7 %)		29,13	17,67	$> 10^6$ (31,6 %)
25; 25	0,954	0,948	0,960	25; 25	0,947	0,917	0,965
	3,54	2,93	4,86 (0,01 %)		16,21	8,73	$> 10^4$ (2,4 %)
50; 50	0,954	0,947	0,954	50; 50	0,955	0,928	0,960
	2,10	1,71	1,77 (0,003 %)		8,62	4,11	164,38 (0,2 %)



Verteilung [ $\gamma$ ] $n_1; n_2$	ELTR	$L_{50}$	$W_{50}$	Verteilung [ $\gamma$ ] $n_1; n_2$	ELTR	$L_{50}$	$W_{50}$
<b>Halbnormal</b>				<b>CN(0,1; 3)</b>			
[3,9]				[8,3]			
10; 10	0,956	0,942	0,954	10; 10	0,977	0,965	0,979
	10,41	7,89	> 10 <sup>4</sup> (1,5 %)		12,64	9,52	> 10 <sup>4</sup> (4,9 %)
10; 30	0,959	0,930	0,954	10; 30	0,981	0,952	0,979
	5,18	3,64	13,00 (0,02 %)		7,82	4,71	944,68 (1,1 %)
25; 25	0,959	0,952	0,959	25; 25	0,982	0,972	0,981
	3,01	2,62	2,88		4,63	3,22	3,71
50; 50	0,960	0,951	0,954	50; 50	0,983	0,972	0,978
	1,69	1,54	1,59		2,64	1,83	1,91

In den dunkel unterlegten Zeilen werden die erzielten Überdeckungswahrscheinlichkeiten für die einzelnen Verfahren, übergeordnete Verteilung und Kombinationen von Stichprobenumfängen angegeben. Der Mittelwert der Intervallbreiten wird unter den einzelnen Überdeckungswahrscheinlichkeiten angegeben. Wenn Intervalle für eine Bedingung unendlich waren, werden sowohl der Mittelwert für die endlichen Intervalle als auch der Prozentsatz der Intervalle angegeben, die unendlich waren. Die Kurtosis ( $\gamma$ ) jeder übergeordneten Verteilung wird in Klammern angezeigt.

Bei symmetrischen Verteilungen mit schwächer besetzten Randbereichen weisen die Ergebnisse darauf hin, dass alle drei Methoden ähnlich konservative Überdeckungswahrscheinlichkeiten ergeben. Die ELTR- und  $L_{50}$ -Intervalle sind jedoch für kleine Stichproben präziser als die  $W_{50}$ -Intervalle. Wenn Stichproben z. B. aus einer Beta-Verteilung mit den Parametern (3; 3) gezogen werden, sind die erzielten Überdeckungswahrscheinlichkeiten für die ELTR-Intervalle und die  $L_{50}$ -Intervalle mindestens so genau wie die der  $W_{50}$ -Intervalle, die  $W_{50}$ -Intervalle sind jedoch durchgehend breiter.

Die ELTR-Intervalle und die  $W_{50}$ -Intervalle sind zudem für symmetrische Verteilungen mit stärker besetzten Randbereichen etwas konservativ, während die  $L_{50}$ -Intervalle liberal sind. Bei nicht balancierten Designs sind die  $L_{50}$ -Intervalle sogar noch liberaler. Wenn z. B. Stichproben mit den Umfängen 10 und 30 aus der Laplace-Verteilung gezogen werden, beträgt die erzielte Überdeckungswahrscheinlichkeit für die  $L_{50}$ -Intervalle 0,919. Wenn Stichproben mit dem gleichen Umfang aus einer t-Verteilung mit 5 Freiheitsgraden gezogen werden, beträgt die erzielte Überdeckungswahrscheinlichkeit für die  $L_{50}$ -Intervalle 0,930.

Die  $L_{50}$ -Intervalle sind ebenfalls relativ liberal, wenn kleine Stichproben aus stark schiefen Verteilungen mit stärker besetzten Randbereichen gezogen werden. Wenn Stichproben beispielsweise aus einer Lognormalverteilung gezogen werden, beträgt die erzielte Überdeckungswahrscheinlichkeit u. U. nur 0,866. Für diese Verteilungen ist die  $W_{50}$ -Methode

die am wenigsten liberale der drei Methoden. Zu viele der  $W_{50}$ -Intervalle weisen jedoch eine unendliche Breite auf. Wenn beispielsweise Stichproben aus der Chi-Quadrat-Verteilung mit 1 Freiheitsgrad ( $\chi^2(1)$ ) gezogen werden, können mehr als 25 % der  $W_{50}$ -Intervalle eine unendliche Breite aufweisen. Die ELTR-Intervalle sind etwas weniger genau, jedoch viel schmaler und damit aussagekräftiger als die  $W_{50}$ -Intervalle.

Schließlich stellen wir fest, dass alle Verfahren durch Ausreißer beeinträchtigt werden. Die  $L_{50}$ -Methode ist davon am wenigsten betroffen; dies war jedoch zu erwarten, da die  $L_{50}$ -Methode abgeleitet wurde, um die Auswirkung von Ausreißern auf Test  $W_{50}$  zu mindern (Pan, 1999). Wenn Stichproben aus der kontaminierten Normalverteilung  $CN(0,1; 3)$  gezogen werden, beträgt das Minimum der erzielten Überdeckungswahrscheinlichkeiten für das ELTR-Verfahren und das  $W_{50}$ -Verfahren 0,977. Weitere (hier nicht angegebene) Simulationsergebnisse verweisen darauf, dass sich die Intervalle bei zunehmenden Stichprobenumfängen nur langsam verbessern.

## Studie 3: Empfindlichkeit gegenüber der Annahme gleicher Kurtosiswerte

In unserer letzten Studie wird die Empfindlichkeit des ELTR-Verfahrens gegenüber der Annahme gleicher Kurtosiswerte untersucht, unter der es abgeleitet wurde. Wir untersuchen die Leistung des ELTR-Verfahrens, wenn die Kurtosiswerte der übergeordneten Grundgesamtheiten ungleich sind, d. h. wenn  $\gamma_1 \neq \gamma_2$ . Dabei werden auch das  $L_{50}$ -Verfahren und das  $W_{50}$ -Verfahren aufgeführt, da sie unter der Annahme abgeleitet wurden, dass die Grundgesamtheiten einander ähneln. Diese Annahme der Ähnlichkeit wird untergraben, wenn die Kurtosiswerte der übergeordneten Grundgesamtheiten ungleich sind. Die Ergebnisse werden in Tabelle 4 aufgeführt.

**Tabelle 4** Empfindlichkeit des ELTR-Verfahrens gegenüber der Annahme gleicher Kurtosiswerte; das nominale Konfidenzniveau ist  $1 - \alpha = 0,95$

Vert. 1, Vert. 2 [ $\gamma_1; \gamma_2$ ] $n_1; n_2$	ELTR	$L_{50}$	$W_{50}$	Vert. 1, Vert. 2 [ $\gamma_1; \gamma_2$ ] $n_1; n_2$	ELTR	$L_{50}$	$W_{50}$
<b>Beta (3; 3), Normal [2,5; 3]</b>				<b>Normal, CN (0,9; 3) [3; 8,3]</b>			
10; 10	0,964	0,961	0,964	10; 10	0,955	0,948	0,951
	0,27	0,23	204,50 (0,20 %)		6,88	5,16	> 10 <sup>4</sup> (4,89 %)
30; 10	0,946	0,939	0,946	30; 10	0,941	0,910	0,942
	0,16	0,17	0,15		5,26	3,77	3,20
10; 30	0,966	0,956	0,967	10; 30	0,961	0,950	0,958
	0,14	0,11	0,17		4,26	2,40	630,42 (1,10 %)

<b>Vert. 1, Vert. 2</b> [ $\gamma_1; \gamma_2$ ] $n_1; n_2$	<b>ELTR</b>	$L_{50}$	$W_{50}$	<b>Vert. 1, Vert. 2</b> [ $\gamma_1; \gamma_2$ ] $n_1; n_2$	<b>ELTR</b>	$L_{50}$	$W_{50}$
50; 50	0,951	0,950	0,949	50; 50	0,936	0,910	0,907
	0,04	0,05	0,05		1,27	1,11	1,19
<b>Normal, Laplace</b> [3; 6]				<b>Halbnormal, <math>\chi^2(5)</math></b> [3,9; 5,4]			
10; 10	0,941	0,935	0,947	10; 10	0,956	0,940	0,954
	6,67	5,17	$> 10^6$ (2,90 %)		0,42	0,32	304,41 (2,60 %)
30; 10	0,912	0,888	0,914	30; 10	0,954	0,918	0,949
	5,06	3,85	3,21		0,33	0,22	0,20
10; 30	0,963	0,943	0,955	10; 30	0,962	0,934	0,958
	3,33	2,25	$> 10^3$ (0,40 %)		0,23	0,15	3,28 (0,30 %)
50; 50	0,935	0,894	0,889	50; 50	0,955	0,941	0,945
	0,98	1,04	1,12		0,07	0,06	0,07
<b>Normal, Halbnormal</b> [3; 3,9]				$\chi^2(5),$ <b>Exponential</b> [5,4; 9]			
10; 10	0,956	0,948	0,957	10; 10	0,938	0,914	0,940
	28,16	20,65	$> 10^4$ (1,50 %)		211,17	137,88	$> 10^6$ (9,10 %)
30; 10	0,946	0,924	0,947	30; 10	0,928	0,875	0,929
	20,59	14,83	12,78		194,70	93,02	83,02
10; 30	0,961	0,946	0,962	10; 30	0,968	0,930	0,954
	14,06	9,37	49,11 (0,02 %)		102,35	55,29	$> 10^5$ (3,90 %)
50; 50	0,953	0,950	0,952	50; 50	0,950	0,920	0,923
	4,32	4,16	4,33		29,64	23,37	25,54

In den dunkel unterlegten Zeilen werden die erzielten Überdeckungswahrscheinlichkeiten für die einzelnen Verfahren, Kombinationen von übergeordneten Verteilungen (Vert. 1, Vert. 2) und Kombinationen von Stichprobenumfängen angegeben. Der Mittelwert der Intervallbreiten wird unter den einzelnen Überdeckungswahrscheinlichkeiten angegeben. Wenn Intervalle für eine

*Bedingung unendlich sind, werden sowohl der Mittelwert für die endlichen Intervalle als auch der Prozentsatz der unendlichen Intervalle angegeben. Die Kurtosis jeder übergeordneten Verteilung ( $\gamma_1; \gamma_2$ ) wird in Klammern angezeigt.*

Im Allgemeinen scheint die Leistung des ELTR-Verfahrens bei ausreichend großen Stichprobenumfängen durch ungleiche Kurtosiswerte nicht beeinträchtigt zu werden. Wenn Designs jedoch unbalanciert sind und die kleinere Stichprobe aus der Verteilung mit stärker besetzten Randbereichen gezogen wird, sind die erzielten Überdeckungswahrscheinlichkeiten liberal. Die erzielten Überdeckungswahrscheinlichkeiten sind besser, wenn die größere Stichprobe aus der Verteilung mit stärker besetzten Randbereichen gezogen wurde.

Bei ausreichend großen Stichprobenumfängen scheinen die  $L_{50}$ -Intervalle und die  $W_{50}$ -Intervalle ebenfalls generell robust gegenüber der Unähnlichkeit von Verteilungen zu sein, die auf ungleiche Kurtosiswerte zurückzuführen ist. Wenn Stichproben jedoch aus einer Normalverteilung und einer Laplace-Verteilung oder aus einer Normalverteilung und einer kontaminierten Normalverteilung gezogen werden, sind die Überdeckungswahrscheinlichkeiten für die  $L_{50}$ -Intervalle und die  $W_{50}$ -Intervalle instabil, selbst bei Stichproben mit einem großen Stichprobenumfang von 50.

Die  $L_{50}$ -Intervalle sind generell liberaler als die ELTR-Intervalle und die  $W_{50}$ -Intervalle. In drei Fällen sind die erzielten Überdeckungswahrscheinlichkeiten für die  $L_{50}$ -Intervalle kleiner als 0,90. Im Gegensatz dazu ist nur eine der erzielten Überdeckungswahrscheinlichkeiten für die  $W_{50}$ -Intervalle kleiner als 0,90. Die kleinste erzielte Überdeckungswahrscheinlichkeit für die ELTR-Intervalle beträgt 0,912.

Die vorherige Studie (Tabelle 3) zeigt, dass alle drei Verfahren Intervalle erzeugen, die merklich konservativer sind, wenn beide Stichproben aus der kontaminierten Normalverteilung  $CN(0,1; 3)$  gezogen werden. Die aktuelle Studie zeigt, dass alle drei Verfahren eine spürbar bessere Leistung aufweisen, wenn nur eine Stichprobe aus  $CN(0,1; 3)$  gezogen wird. Beachten Sie jedoch, dass sich die Leistung der  $L_{50}$ -Intervalle und der  $W_{50}$ -Intervalle offenbar wesentlich verschlechtert, wenn der Stichprobenumfang auf 50 ansteigt.

## 5. Beispiel

In diesem Abschnitt werden alle vier Verfahren – F, ELTR,  $L_{50}$  und  $W_{50}$  – auf einen Datensatz angewendet, der von Pan (1999) übernommen wurde. Ott (1993, Seite 352) beschreibt die Daten wie folgt:

Ein Chemiker in einer Eisenerzgrube vermutet, dass sich die Varianz im Gehalt (Gewicht, in Unzen) an Eisenoxid pro Pfund Erz mit Zunahme des mittleren Gehalts an Eisenoxid pro Pfund tendenziell vergrößert. Zum Testen dieser Theorie werden zehn Eisenerzproben zu je 1 Pfund von jeweils zwei Teststellen ausgewählt. Dabei ist Teststelle 1 durch einen viel höheren mittleren Gehalt an Eisenoxid als Teststelle 2 gekennzeichnet. Die Werte des Eisenoxid-Gehalts in den Erzproben werden unten angegeben:

<b>Teststelle 1</b>	8,1	7,4	9,3	7,5	7,1	8,7	9,1	7,9	8,4	8,8
<b>Teststelle 2</b>	3,9	4,4	4,7	3,6	4,1	3,9	4,6	3,5	4,0	4,2

Die 95%-KIs für das mit den vier unterschiedlichen Methoden berechnete  $\sigma_2/\sigma_1 = 1/\rho$  werden in der folgenden Tabelle angegeben:

<b>Verfahren</b>	<b>95%-KI</b>
F	(0,262; 1,055)
ELTR	(0,277; 0,924)
$L_{50}$ (Pan)	(0,295; 0,938)
$W_{50}$ (Levene/Brown-Forsythe)	(0,237; 0,908)

## 6. Schlussfolgerung

Unsere Studien zeigen, dass die auf dem ELTR-Verfahren basierenden KIs im Allgemeinen so genau wie die KIs sind, die vom  $L_{50}$ -Test und vom  $W_{50}$ -Test abgeleitet wurden. Die ELTR-Intervalle und die  $L_{50}$ -Intervalle sind jedoch für die meisten Verteilungen präziser als die  $W_{50}$ -Intervalle. Die  $W_{50}$ -Intervalle sind tendenziell genauer als die ELTR-Intervalle und die  $L_{50}$ -Intervalle, wenn kleine Stichproben aus stark schiefen Verteilungen mit stärker besetzten Randbereichen gezogen werden. Dieser Vorteil wird jedoch üblicherweise durch einen erheblichen Präzisionsverlust wieder aufgehoben. Die erhaltenen  $W_{50}$ -Intervalle sind i. d. R. zu breit und weisen wahrscheinlich eine unendliche Breite auf.

Die  $L_{50}$ -Intervalle verbessern entsprechend ihrer Konzeption die Präzision der  $W_{50}$ -Intervalle. Bei schiefen Verteilungen sind jedoch die  $L_{50}$ -Intervalle so kurz, dass sie übermäßig liberale Überdeckungswahrscheinlichkeiten liefern. Im Gegensatz dazu sind die ELTR-Intervalle im Allgemeinen stabiler. Die ELTR-Intervalle sind üblicherweise weder zu lang noch zu kurz, so dass die Überdeckungswahrscheinlichkeiten i. d. R. nicht zu konservativ oder zu liberal sind. Daher scheint das ELTR-Verfahren für die meisten praktischen Anwendungen am besten geeignet zu sein.

Die Berechnung von ELTR-Intervallen ist etwas aufwändiger als die von Intervallen, die auf Test  $L_{50}$  oder Test  $W_{50}$  basieren. Im Allgemeinen wiegen aber die höhere Präzision (im Vergleich mit den  $W_{50}$ -Intervallen) und die höhere Genauigkeit (im Vergleich mit den  $L_{50}$ -Intervallen) den Zusatzaufwand beim Berechnen mehr als auf. Das ELTR-Verfahren wurde als Teil der Analyse „Test auf Varianzen bei zwei Stichproben“ in Release 17 der Minitab Statistical Software eingeführt, in der es als Bonett-Verfahren bezeichnet wird.

Bei künftigen Studien ist zu erwägen, die Eigenschaften des Layard-Tests in Bezug auf kleine Stichproben in Designs mit mehreren Stichproben zu untersuchen und den zusammengefassten Kurtosis-Schätzwert nach Layard dabei durch die robustere Version von Bonett zu ersetzen, der angegeben wird als

$$\hat{\gamma}_A = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{\left[ \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - m_i)^2 \right]^2} \sum_{i=1}^k n_i$$

Hierbei ist  $m_i$  der getrimmte Mittelwert für Stichprobe  $i$  mit dem Trim-Anteil  $1/[2(n_i - 4)^{1/2}]$  und  $i = 1; \dots; k$ .

Darüber hinaus könnte es von Vorteil sein, Shoemakers Approximation der asymptotischen Varianz der log-transformierten Stichprobenvarianz zu verwenden.

Schließlich ist anzumerken, dass die von Bonett (2006) vorgeschlagenen Intervalle zwar nicht als KIs geeignet, für die meisten Verteilungen aber nichtsdestoweniger genau und präzise sind, wenn sie als Annahmebereiche für den Test auf Gleichheit zweier Varianzen interpretiert werden. Diese Annahmebereiche sind hervorragend als Grundlage für ein grafisches Verfahren zum Vergleichen mehrerer Varianzen geeignet. Hochberg, Weiss und Hart (1982) haben ein ähnliches Verfahren zum Testen der Gleichheit von Mittelwerten vorgeschlagen. Ein solches Verfahren wurde als Teil der Analyse „Test auf gleiche Varianzen“ in Release 17 der Minitab Statistical Software eingeführt, in der es als Mehrfachvergleichsverfahren bezeichnet wird.

## 7. Anhänge

### Anhang A: Beweis von Ergebnis 1

Entsprechend den Notationen und Annahmen von Abschnitt 2 sei  $X_j = \rho Y_{2j}$  für ein gegebenes  $\rho = \sigma_1/\sigma_2$ . Dann ist

$$\text{Var}(X_j) = \rho^2 \text{Var}(Y_{2j}) = \rho^2 \sigma_2^2 = \sigma_1^2 = \text{Var}(Y_{1j})$$

und

$$E\left(X_j - \mu_{X_j}\right)^4 / \sigma_{X_j}^4 = \rho^4 E(Y_{2j} - \mu_2) / (\rho^4 \sigma_2^4) = E(Y_{2j} - \mu_2) / \sigma_2^4 = \gamma$$

Da gemäß Annahme  $E(Y_{1j} - \mu_1)^4 / \sigma_1^4 = \gamma$  gilt, folgt daraus, dass die übergeordneten Grundgesamtheiten der zwei Stichproben  $Y_{1j}$  und  $X_j = \rho Y_{2j}$  die gleiche Varianz  $\sigma_1^2$  und die gleiche Kurtosis  $\gamma$  aufweisen. Nach Layard (1973) wird ein konsistenter zusammengefasster Kurtosis-Schätzwert von  $\gamma$  auf der Grundlage der zwei Stichproben  $Y_{1j}$  und  $X_j$  angegeben als

$$\begin{aligned} \hat{\gamma}' &= (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \sum_{j=1}^{n_2} (X_j - \bar{X})^4}{[(n_1 - 1)S_1^2 + (n_2 - 1)S_X^2]^2} \\ &= (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - \bar{Y}_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - \bar{Y}_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2} = \hat{\gamma}_P(\rho) \end{aligned}$$

wie gefordert.

## Anhang B: Beweis von Ergebnis 2

Es wurde bereits festgestellt, dass eine approximierte Konfidenzmenge von  $(1 - \alpha)100$  Prozent für  $\rho = \sigma_1/\sigma_2$  auf der Grundlage von  $T_2$  angegeben wird durch

$$\left\{ \rho \in (0; \infty): (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left( \frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} \right) \leq 0 \right\}$$

Der zusammengefasste Kurtosis-Schätzwert kann anhand der Kurtosis-Schätzwerte für die einzelnen Stichproben ausgedrückt werden, die angegeben werden als

$$\hat{\gamma}_i = n_i \frac{\sum_{j=1}^{n_i} (Y_{ij} - m_i)^4}{[(n_i - 1)S_i^2]^2}, i = 1; 2$$

Konkreter heißt dies: Wenn  $t = \rho/\hat{\rho}$ , dann

$$\hat{\gamma}_P(\rho) = (n_1 + n_2) \frac{\sum_{j=1}^{n_1} (Y_{1j} - m_1)^4 + \rho^4 \sum_{j=1}^{n_2} (Y_{2j} - m_2)^4}{[(n_1 - 1)S_1^2 + \rho^2(n_2 - 1)S_2^2]^2} = (n_1 + n_2) \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 t^4/n_2}{(K + t^2)^2}$$

Hierbei ist  $K = (n_1 - 1)/(n_2 - 1)$ .

Folglich kann der quadrierte Standardfehlerterm ausgedrückt werden als

$$\frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 t^4/n_2}{(K + t^2)^2} - B$$

Dabei gilt Folgendes:

$$A = \frac{(n_1 + n_2)(n_1 + n_2 - 2)}{(n_1 - 1)(n_2 - 1)}, B = \frac{k_1}{n_1 - 1} + \frac{k_2}{n_2 - 1}$$

Wenn also  $r^2 = \rho^2/(c\hat{\rho}^2)$  gilt, wird problemlos ersichtlich, dass

$$\begin{aligned} & (\ln \rho^2 - \ln(c\hat{\rho}^2))^2 - z_{\alpha/2}^2 \left( \frac{\hat{\gamma}_P(\rho) - k_1}{n_1 - 1} + \frac{\hat{\gamma}_P(\rho) - k_2}{n_2 - 1} \right) \\ &= (\ln r^2)^2 - z_{\alpha/2}^2 \left( A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 c^2 r^4/n_2}{(K + c r^2)^2} - B \right) \end{aligned}$$

Daraus folgt, dass eine approximierte Konfidenzmenge von  $(1 - \alpha)100$  Prozent für  $\rho = \sigma_1/\sigma_2$  auf der Grundlage von  $T_2$  angegeben werden kann als

$$\hat{\rho}\sqrt{c} \{r \in (0; \infty): H(r^2) \leq 0\}$$

Äquivalent kann die Konfidenzmenge für  $\rho^2 = \sigma_1^2/\sigma_2^2$  ausgedrückt werden als

$$c\hat{\rho}^2 \{r \in (0; \infty): H(r) \leq 0\}$$

Dabei gilt Folgendes:

$$H(x) = (\ln x)^2 - z_{\alpha/2}^2 se^2(cx), x > 0$$

und

$$se^2(x) = A \frac{\hat{\gamma}_1 K^2/n_1 + \hat{\gamma}_2 x^2/n_2}{(K + x)^2} - B$$

## Anhang C: Beweis von Ergebnis 3

Es kann einfach nachgewiesen werden, dass  $H(x)$  auf der positiven reellen Linie stetig ist, mit  $H(0) = H(+\infty) = +\infty$  und  $H(1) < 0$ . Nach dem Zwischenwertsatz weist die Funktion  $H(x)$  mindestens eine Nullstelle im Intervall  $(0; 1)$  und mindestens eine Nullstelle im Intervall  $(0; +\infty)$  auf. Wenn also die Funktion  $H(x)$  genau zwei Nullstellen aufweist, liegt eine Nullstelle unter 1 und die andere über 1. Da sich diese Funktion nach oben öffnet, ist die Bedingung der Ungleichheit  $H(r) \leq 0$  erfüllt, wenn sich  $r$  zwischen den Nullstellen befindet. Diese Nullstellen definieren die Endpunkte des KIs für  $\rho^2/(c\hat{\rho}^2)$ . Werden die zwei Nullstellen also mit  $x_L < 1 < x_U$  angegeben, wird laut Ergebnis 2 die untere Konfidenzgrenze für das Verhältnis der Varianzen,  $\rho^2$ , als  $c\hat{\rho}^2 x_L$  berechnet, und die untere Konfidenzgrenze für das Verhältnis der Standardabweichungen wird als  $\hat{\rho}\sqrt{cx_L}$  berechnet. Dementsprechend ist die obere Konfidenzgrenze für das Verhältnis der Varianzen  $c\hat{\rho}^2 x_U$ , und die obere Konfidenzgrenze für das Verhältnis der Standardabweichungen ist  $\hat{\rho}\sqrt{cx_U}$ .

Wenn die Funktion  $H(x)$  andererseits mehr als zwei Nullstellen aufweist, ist die Bedingung der Ungleichheit  $H(r) \leq 0$  erfüllt, wenn sich  $r$  zwischen aufeinanderfolgenden Nullstellen befindet, bei denen sich die Funktion nach oben öffnet. Damit ist die Konfidenzmenge eine Vereinigung von nicht überlappenden Intervallen.

## 8. Literaturhinweise

Balakrishnan, N. und Ma, C. W. (1990). A Comparative Study of Various Tests for the Equality of Two Population Variances. *Journal of Statistical Computation and Simulation*, 35, 41–89.

Bonett D. G. (2006). Robust Confidence Interval for a Ratio of Standard Deviations. *Applied Psychological Measurements*, 30, 432–439.

Boos, D. D. und Brownie, C. (1989). Bootstrap Methods for Testing Homogeneity of Variances. *Technometrics*, 31, 69–82.

Brown, M. B. und Forsythe A. B. (1974). Robust Tests for the Equality of Variances. *Journal of the American Statistical Association*, 69, 364–367.

Conover, W. J., Johnson, M. E. und Johnson, M. M. (1981). A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data. *Technometrics*, 23, 351–361.

Fligner, M. A. und Killeen, T. J. (1976). Distribution-Free Two-Sample Tests for Scale. *Journal of the American Statistical Association*, 71, 210–213.

Hall, P. und Padmanabhan, A. R. (1997). Adaptive Inference for the Two-Sample Scale Problem. *Technometrics*, 39, 412–422.

Hochberg, Y., Weiss, G. und Hart S., (1982). On Graphical Procedures for Multiple Comparisons. *Journal of the American Statistical Association*, 77, 767–772.

Layard, M. W. J. (1973). Robust Large-Sample Tests for Homogeneity of Variances. *Journal of the American Statistical Association*, 68, 195–198.

Levene, H. (1960). "Robust Tests for Equality of Variances," in I. Olkin, Hrsg., *Contributions to Probability and Statistics*, Palo Alto, CA: Stanford University Press, 278–292.



- Lim, T.-S. und Loh, W.-Y. (1996). A Comparison of Tests of Equality of Variances. *Computational Statistics and Data Analysis*, 22, 287–301.
- Ott, L. (1993). *An Introduction to Statistical Methods and Data Analysis*, Belmont, CA: Duxbury Press.
- Pan, G. (1999). On a Levene Type Test for Equality of Two Variances. *Journal of Statistical Computation and Simulation*, 63, 59–71.
- Shoemaker, L. H. (2003). Fixing the F Test for Equal Variances. *The American Statistician*, 57, 105–114.
- Wolfram, S. (1999). *The Mathematica Book*, 4th ed. Wolfram Media/Cambridge University Press.

© 2020 Minitab, LLC. All rights reserved. Minitab®, Minitab Workspace™, Companion by Minitab®, Salford Predictive Modeler®, SPM®, and the Minitab® logo are all registered trademarks of Minitab, LLC, in the United States and other countries. Additional trademarks of Minitab, LLC can be found at [www.minitab.com](http://www.minitab.com). All other marks referenced remain the property of their respective owners.