

Dieses White Paper ist Teil einer Reihe von Veröffentlichungen, welche die Forschungsarbeiten der Minitab-Statistiker erläutern, in deren Rahmen die im Assistenten der Minitab Statistical Software verwendeten Methoden und Datenprüfungen entwickelt wurden.

Test für Prozentsatz fehlerhafter Einheiten bei einer Stichprobe

Übersicht

Mit einem Test von Anteilen bei einer Stichprobe wird bestimmt, ob ein Anteil von einem Sollwert abweicht. In der Qualitätsanalyse kommt der Test häufig zur Anwendung, wenn ein Produkt oder eine Dienstleistung als fehlerhaft oder nicht fehlerhaft eingestuft wird, um zu bestimmen, ob der Prozentsatz fehlerhafter Einheiten signifikant vom Sollprozentsatz fehlerhafter Einheiten abweicht.

Der Minitab-Assistent bietet einen Test für den Prozentsatz fehlerhafter Einheiten bei einer Stichprobe. Für die Daten des Tests wird die Anzahl der fehlerhaften Einheiten in einer Stichprobe erfasst. Es wird angenommen, dass es sich hierbei um die beobachteten Werte einer binomial verteilten Zufallsvariablen handelt. Der Assistent nutzt zum Berechnen der Ergebnisse des Hypothesentests und der Konfidenzintervalle exakte Methoden. Daher sollte die tatsächliche Wahrscheinlichkeit eines Fehlers 1. Art nahe dem für den Test angegebenen Signifikanzniveau (Alpha) liegen, so dass keine weitere Untersuchung erforderlich ist. Die Analyse der Trennschärfe und des Stichprobenumfangs für den Test für den Prozentsatz fehlerhafter Einheiten bei einer Stichprobe basiert jedoch auf einer Approximation, und ihre Genauigkeit muss ausgewertet werden.

Im vorliegenden White Paper wird die Methodologie zum Auswerten von Trennschärfe und Stichprobenumfang für den Test für den Prozentsatz fehlerhafter Einheiten bei einer Stichprobe untersucht. Dabei wird die theoretische Trennschärfe der Approximationsmethode mit der tatsächlichen Trennschärfe des exakten Tests verglichen.

Darüber hinaus wird beschrieben, wie wir eine Richtlinie erarbeitet haben, die Sie bei der Bewertung unterstützen soll, ob Ihr Stichprobenumfang ausreicht, um eine Abweichung des

Prozentsatzes fehlerhafter Einheiten von einem Sollprozentsatz fehlerhafter Einheiten feststellen zu können. Der Assistent führt automatisch eine Prüfung des Stichprobenumfangs durch und gibt die Ergebnisse in der Auswertung aus.

Der Test für den Prozentsatz fehlerhafter Einheiten bei einer Stichprobe hängt zudem von weiteren Annahmen ab. Weitere Informationen finden Sie in Anhang A.

Methode des Tests für den Prozentsatz fehlerhafter Einheiten bei einer Stichprobe

Leistung der theoretischen Trennschärfefunktion

Der Assistent führt den Hypothesentest für einen einzelnen Anteil einer Bernoulli-verteilten Grundgesamtheit (Prozentsatz fehlerhafter Einheiten) mit exakten Methoden (Likelihood-Quotient) durch. Da die Trennschärfefunktion dieses exakten Tests jedoch nicht auf einfache Weise abgeleitet werden kann, erfolgt eine Approximation der Trennschärfefunktion anhand der theoretischen Trennschärfefunktion des entsprechenden Tests auf Normal-Approximation.

Zielstellung

Wir wollten feststellen, ob die theoretische Trennschärfefunktion auf der Grundlage des Tests auf Normal-Approximation zum Auswerten der Anforderungen an Trennschärfe und Stichprobenumfang für den Test für den Prozentsatz fehlerhafter Einheiten bei einer Stichprobe im Assistenten verwendet werden kann. Hierfür musste untersucht werden, ob diese theoretische Trennschärfefunktion die tatsächliche Trennschärfe des exakten Tests (Likelihood-Quotienten-Test) genau abbildet.

Methode

Eine Definition der Teststatistik, also der p-Wert, und des Konfidenzintervalls für den exakten Test (Likelihood-Quotienten-Test) findet sich in Anhang B. Eine Definition der theoretischen Trennschärfefunktion auf der Grundlage der Normal-Approximation wird in Anhang C gegeben. Auf der Grundlage dieser Definitionen wurden Simulationen unter Verwendung des exakten Tests zum Schätzen der tatsächlichen Trennschärfen (die wir als simulierte Trennschärfen bezeichnen) durchgeführt.

Für die Simulationen wurden Zufallsstichproben mit unterschiedlichen Umfängen aus mehreren Bernoulli-verteilten Grundgesamtheiten gezogen. Für alle Bernoulli-verteilten Grundgesamtheiten wurde der exakte Test für jede der 10.000 Stichprobenreplikationen durchgeführt. Für jeden Stichprobenumfang wurde die simulierte Trennschärfe des Tests zum Erkennen einer gegebenen Differenz als Anteil der 10.000 Stichproben berechnet, bei denen der Test signifikant ist. Zum Vergleich wurde auch die entsprechende theoretische Trennschärfe auf der Grundlage des Tests auf Normal-Approximation berechnet. Wenn die Approximation gute Ergebnisse liefert, liegen die theoretischen und simulierten Trennschärfen nah beieinander. Weitere Informationen finden Sie in Anhang D.

Ergebnisse

Unsere Simulationen haben gezeigt, dass die theoretische Trennschärfefunktion des Tests auf Normal-Approximation und die simulierte Trennschärfefunktion des exakten Tests

(Likelihood-Quotienten-Test) im Allgemeinen annähernd gleich sind. Daher nutzt der Assistent die theoretische Trennschärfefunktion des Tests auf Normal-Approximation, um die Stichprobenumfänge zu schätzen, mit denen der exakte Test eine ausreichende Trennschärfe aufweist, um Differenzen mit praktischen Konsequenzen im Prozentsatz der fehlerhaften Einheiten zu erkennen.

Datenprüfungen

Stichprobenumfang

Normalerweise wird ein Hypothesentest durchgeführt, um einen Beleg für die Zurückweisung der Nullhypothese („keine Differenz“) zu erhalten. Wenn die Stichprobe zu klein ist, reicht die Trennschärfe des Tests u. U. nicht aus, um eine tatsächlich vorhandene Differenz zu erkennen; hierbei handelt es sich um einen Fehler 2. Art. Daher muss unbedingt sichergestellt werden, dass die Stichprobenumfänge ausreichend groß sind, um mit einer hohen Wahrscheinlichkeit Differenzen mit praktischen Konsequenzen zu erkennen.

Zielstellung

Wenn die Daten keine ausreichenden Hinweise zum Zurückweisen der Nullhypothese liefern, wollten wir ermitteln können, ob die Stichprobenumfänge groß genug für den Test sind, so dass dieser mit hoher Wahrscheinlichkeit Differenzen mit praktischen Konsequenzen erkennt. Bei der Planung der Stichprobenumfänge soll zwar sichergestellt werden, dass die Stichprobenumfänge ausreichend groß sind, um mit hoher Wahrscheinlichkeit wichtige Differenzen zu erkennen; andererseits dürfen sie aber nicht so groß sein, dass bedeutungslose Differenzen mit hoher Wahrscheinlichkeit statistisch signifikant werden.

Methode

Die Analyse von Trennschärfe und Stichprobenumfang für den Test für den Prozentsatz fehlerhafter Einheiten bei einer Stichprobe basiert auf der theoretischen Trennschärfefunktion unter Anwendung der Normal-Approximation, die einen guten Schätzwert der tatsächlichen Trennschärfe des exakten Tests liefert (siehe den vorausgegangenen Abschnitt „Methode des Tests für den Prozentsatz fehlerhafter Einheiten bei einer Stichprobe“). Wenn der Sollprozentsatz fehlerhafter Einheiten angegeben wird, hängt die theoretische Trennschärfefunktion vom Stichprobenumfang und der zu erkennenden Differenz ab.

Ergebnisse

Wenn die Daten keine ausreichenden Hinweise liefern, die gegen die Nullhypothese sprechen, berechnet der Assistent Differenzen mit praktischen Konsequenzen, die für den angegebenen Stichprobenumfang mit einer Wahrscheinlichkeit von 80 % und 90 % erkannt werden können. Wenn der Benutzer zudem eine konkrete Differenz mit praktischen Konsequenzen angibt, berechnet der Assistent die Stichprobenumfänge, bei denen die Differenz mit einer Wahrscheinlichkeit von 80 % und 90 % erkannt wird.

Um die Interpretation der Ergebnisse zu erleichtern, werden für die Prüfung auf die Trennschärfe und den Stichprobenumfang in der Auswertung des Assistenten für den Test für den Prozentsatz fehlerhafter Einheiten bei einer Stichprobe die folgenden Statusindikatoren angezeigt:

Status	Bedingung
	<p>Im Test wird eine Differenz zwischen dem Prozentsatz fehlerhafter Einheiten und dem Sollwert festgestellt, daher stellt die Trennschärfe kein Problem dar.</p> <p>ODER</p> <p>Die Trennschärfe ist ausreichend. Im Test wurde keine Differenz zum Sollwert festgestellt, die Stichprobe ist jedoch umfassend genug, dass die angegebene Differenz mit einer Wahrscheinlichkeit von mindestens 90 % ($\text{Trennschärfe} \geq 0,90$) erkannt wird.</p>
	<p>Die Trennschärfe ist möglicherweise ausreichend. Im Test wurde keine Differenz zum Sollwert festgestellt, die Stichprobe ist jedoch umfassend genug, dass die angegebene Differenz mit einer Wahrscheinlichkeit von 80 % bis 90 % ($0,80 \leq \text{Trennschärfe} < 0,90$) erkannt wird. Der erforderliche Stichprobenumfang zum Erzielen einer Trennschärfe von 90 % wird ausgegeben.</p>
	<p>Die Trennschärfe ist möglicherweise nicht ausreichend. Im Test wurde keine Differenz zum Sollwert festgestellt, die Stichprobe ist jedoch umfassend genug, dass die angegebene Differenz mit einer Wahrscheinlichkeit von 60 % bis 80 % ($0,60 \leq \text{Trennschärfe} < 0,80$) erkannt wird. Die erforderlichen Stichprobenumfänge zum Erzielen einer Trennschärfe von 80 % und 90 % werden ausgegeben.</p>
	<p>Die Trennschärfe ist nicht ausreichend. Im Test wurde keine Differenz zum Sollwert festgestellt, und die Stichprobe ist nicht groß genug, dass die angegebene Differenz mit einer Wahrscheinlichkeit von mindestens 60 % ($\text{Trennschärfe} < 0,60$) erkannt wird. Die erforderlichen Stichprobenumfänge zum Erzielen einer Trennschärfe von 80 % und 90 % werden ausgegeben.</p>
	<p>Im Test wurde keine Differenz zum Sollwert erkannt. Sie haben keine zu erkennende Differenz mit praktischen Konsequenzen angegeben; daher wird in der Auswertung die Differenz angegeben, die bei Ihrem Stichprobenumfang und Alpha mit einer Wahrscheinlichkeit von 80 % und 90 % erkannt wird.</p>

Literaturhinweise

Arnold, S.F. (1990). *Mathematical statistics*. Englewood Cliffs, NJ: Prentice Hall, Inc.

Casella, G. und Berger, R.L. (1990). *Statistical inference*. Pacific Grove, CA: Wadsworth, Inc.

Anhang A: Zusätzliche Annahmen für den Test für den Prozentsatz fehlerhafter Einheiten bei einer Stichprobe

Dem Test für den Prozentsatz fehlerhafter Einheiten bei einer Stichprobe liegen die folgenden Annahmen zugrunde:

- Die Daten bestehen aus n verschiedenen Einheiten, wobei jede Einheit als fehlerhaft oder nicht fehlerhaft klassifiziert ist.
- Die Wahrscheinlichkeit, dass eine Einheit fehlerhaft ist, ist für jede Einheit in einer Stichprobe gleich.
- Die Wahrscheinlichkeit, dass eine Einheit fehlerhaft ist, wird nicht dadurch beeinflusst, ob eine andere Einheit fehlerhaft ist.

Die Richtigkeit dieser Annahmen kann in den Datenprüfungen der Auswertung nicht bestätigt werden, da für diesen Test Zusammenfassungsdaten und keine Rohdaten erfasst werden.

Anhang B: Exakter Test (Likelihood-Quotienten-Test)

Angenommen, es liegt eine Zufallsstichprobe X_1, \dots, X_n aus einer Bernoulli-Verteilung mit einer Erfolgswahrscheinlichkeit von $p = \Pr(X_i = 1) = 1 - \Pr(X_i = 0)$ vor.

Im Folgenden werden die exakten Methoden zum Ziehen von Rückschlüssen über p beschrieben.

Formel B1: Exakter Test und p-Wert

Betrachten Sie einen Test der Nullhypothese $H_0: p = p_0$ gegen eine beliebige dieser Alternativhypothesen: $H_A: p > p_0$; $H_A: p < p_0$ oder $H_A: p \neq p_0$.

Sei $X = \sum_{i=1}^n X_i$

Dann ist X eine binomial verteilte Zufallsvariable mit n Versuchen und der Erfolgswahrscheinlichkeit p .

Ein einseitiger Test auf der Grundlage von X ist ein gleichmäßig trennschärfster Test (uniformly most powerful, UMP) und ein Likelihood-Quotienten-Test. Für beidseitige Tests basiert der Likelihood-Quotienten-Test ebenfalls auf X , und die Teststatistik lautet

$$\Lambda(X) = \left(\frac{\hat{p}}{p_0}\right)^X \left(\frac{1-\hat{p}}{1-p_0}\right)^{n-X}$$

(siehe Arnold, 1990).

Die p-Werte für einseitige Tests können direkt auf der Grundlage der genauen Verteilung von X bestimmt werden. Für beidseitige Tests werden die p-Werte als Wahrscheinlichkeit unter der Nullhypothese berechnet, dass ein Likelihood-Quotient (bzw. Log-Likelihood-Quotient) beobachtet wird, der mindestens so groß wie der tatsächlich beobachtete ist. Zum Berechnen dieser Wahrscheinlichkeit wird im Allgemeinen ein numerischer Algorithmus zur Wurzelbestimmung verwendet.

Formel B2: Exaktes Konfidenzintervall

Ein exaktes beidseitiges $100(1 - \alpha)\%$ -Konfidenzintervall für p ist

$$\frac{1}{1 + \frac{n-x+1}{x} F_{2(n-x+1), 2x, \alpha/2}} \leq p \leq \frac{\frac{x+1}{n-x} F_{2(x+1), 2(n-x), \alpha/2}}{1 + \frac{x+1}{n-x} F_{2(x+1), 2(n-x), \alpha/2}}$$

Hierbei sind x die beobachtete Anzahl der Erfolge und $F_{v_1, v_2, \alpha}$ der obere α . Perzentilpunkt der F -Verteilung mit v_1 und v_2 Freiheitsgraden (siehe Casella und Berger, 1990). Wir übernehmen die Konvention, dass die Untergrenze bei $x = 0$ gleich 0 und die Obergrenze bei $x = n$ gleich 1 ist.

Anhang C: Theoretische Trennschärfefunktion

Eine theoretische Trennschärfefunktion des exakten Tests ist zu komplex, um sie abzuleiten. Daher wird die Trennschärfefunktion des Tests anhand der theoretischen Trennschärfefunktion des Tests auf der Grundlage der Normal-Approximation approximiert. Dieser Approximationstest beruht auf der Tatsache, dass die Zufallsvariable

$$Z = \frac{n^{1/2}(\hat{p} - p)}{(p(1-p))^{1/2}}$$

asymptotisch als Standardnormalverteilung verteilt ist. Die theoretische Trennschärfefunktion dieses Tests ist hinreichend bekannt und dokumentiert. Für die beidseitige Alternativhypothese wird die Trennschärfefunktion wie folgt ausgedrückt:

$$\pi(n, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right) + \Phi\left(\frac{-\delta - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right)$$

Hierbei ist $p = \delta + p_0$; $\Phi(\cdot)$ die kumulative Verteilungsfunktion der Standardnormalverteilung und z_{α} das obere Perzentil der Standardnormalverteilung.

Für die einseitige Alternative $H_A: p > p_0$ kann die Trennschärfefunktionen ausgedrückt werden als

$$\pi(n, \delta) = 1 - \Phi\left(\frac{-\delta + z_{\alpha}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right)$$

Beim Testen gegen die einseitige Alternative $H_A: p < p_0$ kann die Trennschärfefunktion auch ausgedrückt werden als

$$\pi(n, \delta) = \Phi\left(\frac{-\delta - z_{\alpha}\sqrt{p_0(1-p_0)/n}}{\sqrt{p(1-p)/n}}\right)$$

Anhang D: Vergleich der tatsächlichen Trennschärfe mit der theoretischen Trennschärfe

Simulation D1: Schätzen der tatsächlichen Trennschärfe mit dem exakten Test

Wir haben eine Simulation zum Vergleichen der geschätzten tatsächlichen Trennschärfen (die als simulierte Trennschärfen bezeichnet werden) mit den theoretischen Trennschärfen auf der Grundlage der Trennschärfefunktion des Tests auf Normal-Approximation (die als approximierte Trennschärfen bezeichnet werden) konzipiert. In jedem Experiment wurden 10.000 Stichproben jeweils mit dem Umfang n aus einer Bernoulli-verteilten Grundgesamtheit mit der angegebenen Erfolgswahrscheinlichkeit p generiert. Wir haben zwei Fälle für die Erfolgswahrscheinlichkeit berücksichtigt: (1) eine mittlere Erfolgswahrscheinlichkeit mit einem p -Wert nahe bei 0,5 (genauer: $p = 0,45$) und (2) eine geringe bzw. große Erfolgswahrscheinlichkeit mit einem p -Wert nahe 0 bzw. 1 (genauer: $p = 0,85$). Diese beiden Fälle wurden berücksichtigt, weil die DeMoivre-Laplace-Normal-Approximation an die Binomialverteilung, von welcher der Test auf Normal-Approximation abgeleitet ist, nachweislich genau ist, wenn der Umfang der Bernoulli-verteilten Stichprobe größer als 10 ist und die Erfolgswahrscheinlichkeit nahe 0,5 liegt. Für kleinere oder größere Erfolgswahrscheinlichkeiten sind jedoch größere Bernoulli-verteilte Stichproben erforderlich, damit diese Approximation genau ist.

In jedem Experiment wurde der Stichprobenumfang auf einen einzigen Wert von n festgelegt, wobei $n = 10, 15, 20, 30, \dots, 100$. In allen Experimenten wurde die zu erkennende Differenz $\delta = p - p_0$ auf 0,2 festgelegt, um sicherzustellen, dass die erhaltenen Trennschärfewerte bei einer Vergrößerung des Stichprobenumfangs auf 100 nicht zu klein oder zu groß werden. Zum Schätzen der tatsächlichen Trennschärfe für den Test auf der Grundlage der Ergebnisse der einzelnen Simulationen wurde der Anteil der 10.000 Stichprobenreplikationen berechnet, für den der einseitige bzw. der beidseitige exakte Test beim Soll-Signifikanzniveau $\alpha = 0,05$ signifikant war. Schließlich wurden zu Vergleichszwecken die entsprechenden theoretischen Trennschärfen auf der Grundlage des Tests auf Normal-Approximation berechnet. Die Ergebnisse finden Sie in der nachfolgenden Tabelle 1.

Tabelle 1 Simulierte und approximierte (approx.) Trennschärfen des beidseitigen und des einseitigen exakten Tests. Das Soll-Signifikanzniveau ist $\alpha = 0,05$.

n	Beidseitiger Test				Einseitiger Test			
	p = 0,45		p = 0,85		p = 0,45		p = 0,85	
	Simulierte Trennschärfe	Approx. Trennschärfe	Simulierte Trennschärfe	Approx. Trennschärfe	Simulierte Trennschärfe	Approx. Trennschärfe	Simulierte Trennschärfe	Approx. Trennschärfe
10	0,101	0,333	0,200	0,199	0,257	0,436	0,200	0,335
15	0,339	0,441	0,322	0,327	0,339	0,550	0,322	0,489
20	0,406	0,537	0,409	0,455	0,590	0,643	0,648	0,621
30	0,632	0,690	0,708	0,674	0,632	0,779	0,708	0,808
40	0,781	0,799	0,863	0,822	0,781	0,867	0,863	0,911
50	0,877	0,872	0,874	0,910	0,877	0,921	0,933	0,961
60	0,878	0,920	0,942	0,957	0,922	0,954	0,969	0,984
70	0,925	0,951	0,972	0,981	0,953	0,973	0,987	0,994
80	0,954	0,971	0,986	0,992	0,986	0,985	0,993	0,998
90	0,971	0,982	0,993	0,996	0,991	0,991	0,996	0,999
100	0,989	0,990	0,998	0,999	0,994	0,995	0,999	1,000

Die Ergebnisse zeigen, dass die simulierten Trennschärfen und die approximierten Trennschärfen im Allgemeinen nahezu übereinstimmen. Diese Übereinstimmung wird noch deutlicher, wenn die Ergebnisse als Trennschärfekurven grafisch dargestellt werden, wie in den nachfolgenden Abbildungen 1 und 2.

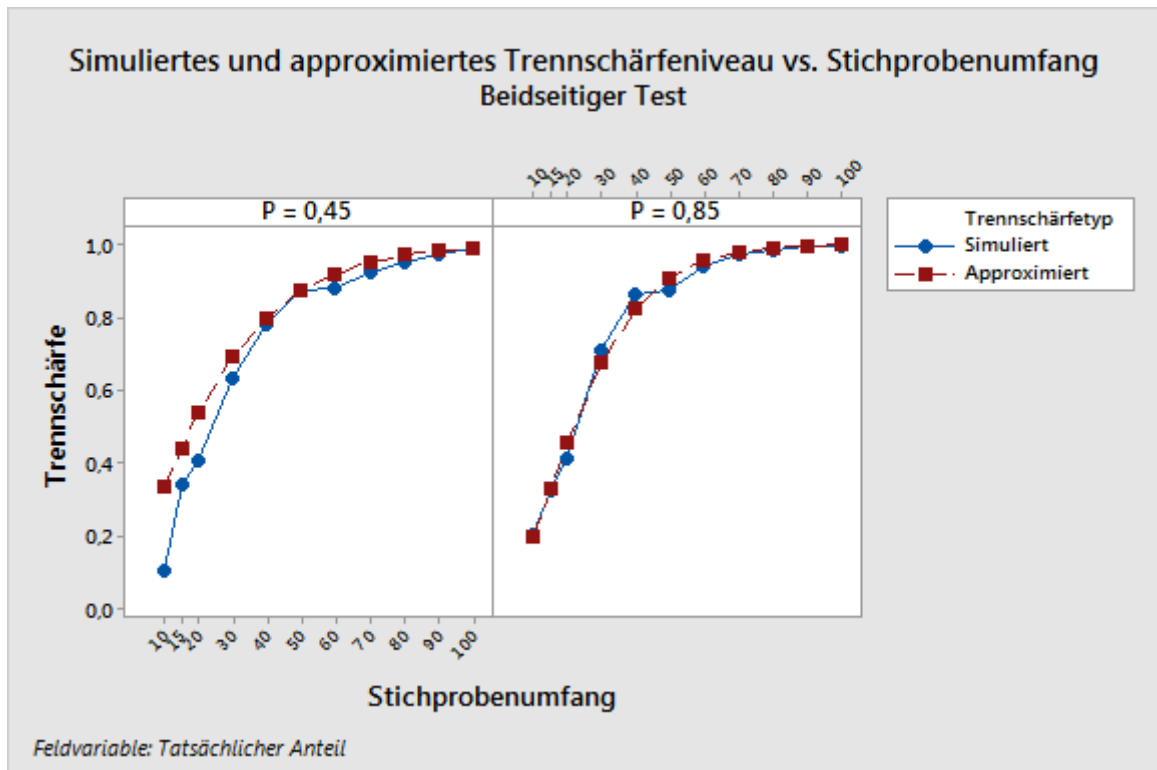


Abbildung 1 Darstellung der simulierten und approximierten Trennschärfen des beidseitigen exakten Tests im Vergleich zum Stichprobenumfang.

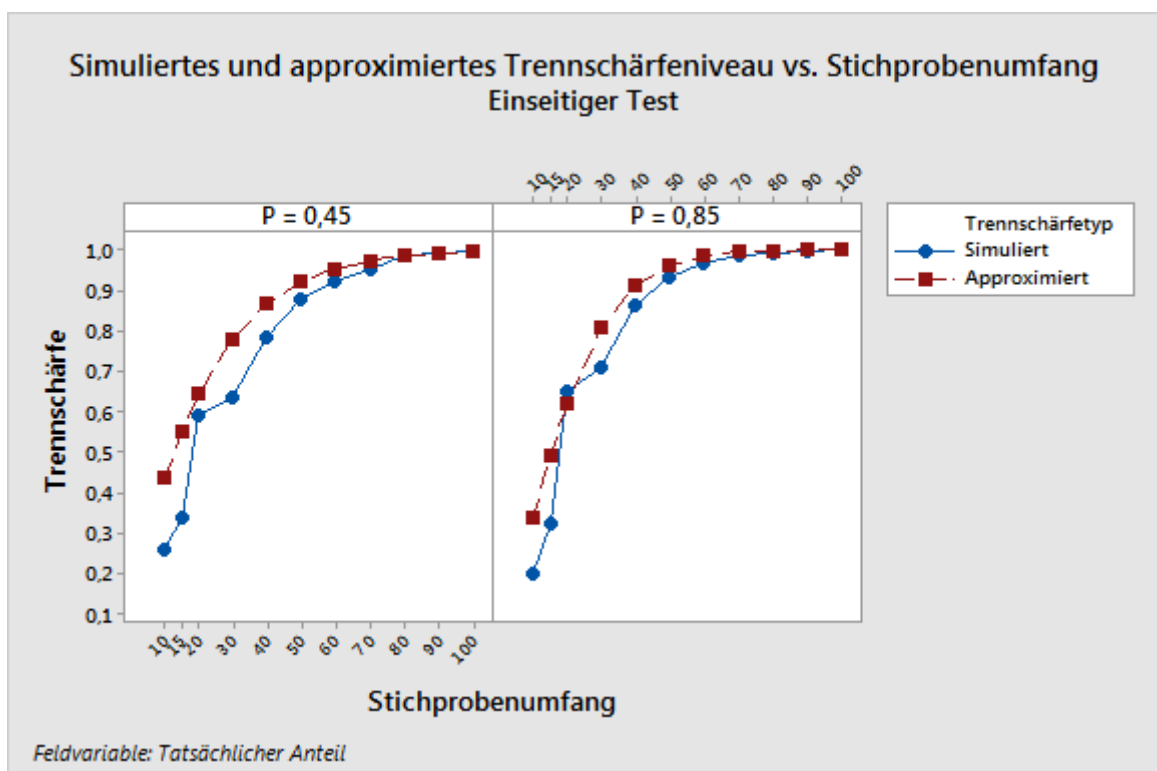


Abbildung 2 Darstellung der simulierten und approximierten Trennschärfen des einseitigen exakten Tests im Vergleich zum Stichprobenumfang.

Die zwei in den Feldern der Abbildungen 1 und 2 dargestellten Trennschärfekurven liegen dicht beieinander, außer bei einigen wenigen Datenpunkten für kleine Stichprobenumfänge. Der nahezu deckungsgleiche Verlauf der Kurven verweist darauf, dass die approximierte Trennschärfefunktion mit der simulierten Trennschärfe weitgehend übereinstimmt, wenn der exakte Test in der Praxis angewendet wird. Daher ist es angemessen, den Stichprobenumfang mit der approximierten Trennschärfefunktion zu schätzen.

Die Abbildungen 1 und 2 veranschaulichen außerdem, dass die theoretischen (approximierten) Trennschärfekurven generell höher als die simulierten Trennschärfekurven liegen. Die approximierten Trennschärfekurven liegen höher, weil die theoretischen Trennschärfen unter der Annahme eines genauen Werts für das Soll-Signifikanzniveau (0,05) berechnet werden. Im Vergleich ist der exakte Test insbesondere bei kleinen Stichproben tendenziell konservativ und liefert daher tatsächliche Signifikanzniveaus, die kleiner als das Sollniveau sind. Infolgedessen sind die simulierten Trennschärfen bei kleinen Stichprobenumfängen tendenziell kleiner.

Fazit: Unsere Simulationen zeigen, dass die theoretische Trennschärfefunktion des Tests auf Normal-Approximation nahezu mit der Trennschärfe des exakten Tests (Likelihood-Quotienten-Test) übereinstimmt. Somit bietet die theoretische Trennschärfefunktion des Tests auf Normal-Approximation eine robuste Grundlage zum Schätzen der Stichprobenumfänge, mit denen der exakte Test über eine ausreichende Trennschärfe verfügt, um Differenzen mit praktischen Konsequenzen erkennen zu können.

© 2015, 2017 Minitab Inc. All rights reserved.

Minitab®, Quality. Analysis. Results.® and the Minitab® logo are all registered trademarks of Minitab, Inc., in the United States and other countries. See minitab.com/legal/trademarks for more information.