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.

Chi-Quadrat-Tests

Übersicht

In der Praxis müssen Qualitätsexperten gelegentlich kategoriale Daten erfassen, um einen Prozess auszuwerten, wenn es nur schwer oder gar nicht möglich ist, stetige Daten zu erfassen. Beispielsweise könnte ein Produkt in zwei Kategorien eingeordnet werden, z. B. „Fehlerhaft“/„Nicht fehlerhaft“, oder auch in mehr als zwei Kategorien, z. B. „Hervorragend“, „Gut“, „Mittel“ und „Schlecht“. Ein weiteres Beispiel ist eine Finanzabteilung, die mit Hilfe von Kategorien verfolgt, wie viele Tage Rechnungen überfällig sind: 15 Tage oder weniger, 16 bis 30 Tage, 31 bis 45 Tage oder 45 Tage und mehr. Die zu untersuchende Variable ist folglich die Anzahl der Elemente, die in die einzelnen Kategorien fallen.

Aufgrund ihrer Flexibilität werden Chi-Quadrat-Tests in vielen Anwendungsfällen eingesetzt, bei denen kategoriale Daten auftreten. Im Assistenten werden Chi-Quadrat-Tests für folgende Zwecke verwendet:

- Testen der Güte der Anpassung für eine multinomiale Verteilung

  Mit diesem Test können Sie bestimmen, ob die Daten derselben Verteilung wie in der Vergangenheit folgen. Die Verteilung ist als multinomiale Verteilung von historischen oder Soll-Prozentwerten definiert, die die Prozentsätze der Elemente in der jeweiligen Ergebniskategorie angeben. Der Chi-Quadrat-Test testet gleichzeitig, ob die Prozentsätze signifikant von ihren entsprechenden historischen oder Soll-Prozentwerten abweichen.

- Testen der Gleichheit der Prozentsätze fehlerhafter Einheiten bei mehr als 2 Gruppen

  Sie können diesen Test verwenden, um zu ermitteln, ob eine Differenz zwischen den Prozentsätzen fehlerhafter Einheiten in verschiedenen Gruppen vorliegt. Die Gruppen unterscheiden sich durch ein relevantes Merkmal, z. B. Produkte, die von verschiedenen Bedienern, in verschiedenen Werken oder zu unterschiedlichen Zeitpunkten produziert werden. Der Chi-Quadrat-Test testet gleichzeitig, ob die Prozentsätze fehlerhafter Einheiten signifikant von anderen Prozentsätzen fehlerhafter Einheiten abweichen.

- Testen der Assoziation zwischen zwei kategorialen Variablen

  Mit diesem Test können Sie ermitteln, ob eine kategoriale Ergebnisvariable (y) mit einer anderen kategorialen Prädiktorvariablen (x) in Beziehung steht (mit dieser assoziiert ist). Der Chi-Quadrat-Test testet gleichzeitig, ob eine Assoziation zwischen der Ergebnisvariablen und einer Prädiktorvariablen vorhanden ist. Im Assistenten können Sie einen Chi-Quadrat-Test auf Assoziation mit einer Prädiktorvariablen (x) durchführen, die zwei oder mehr eindeutige Werte (zwei oder mehr Stichproben) enthält.

Weitere Informationen zur Chi-Quadrat-Teststatistik finden Sie in Anhang A.

Bei Verfahren mit Hypothesentests empfiehlt es sich sicherzustellen, dass die Annahmen für den Test erfüllt sind, dass der Test eine ausreichende Trennschärfe aufweist und dass alle Approximationen, die zum Analysieren der Daten verwendet werden, gültige Ergebnisse liefern. Bei den Chi-Quadrat-Tests stellen die Annahmen einen inhärenten Teil der Datenerfassung dar und werden in den Datenprüfungen nicht behandelt.

Unser Schwerpunkt liegt vielmehr auf der Trennschärfe und der Gültigkeit der Approximationsmethoden. Der Assistent verwendet diese Approximationsmethoden, um die folgenden Prüfungen für die Daten durchzuführen, und führt die Ergebnisse anschließend in der Auswertung auf:

- Stichprobenumfang
- Gültigkeit des Tests
- Gültigkeit der Intervalle

In diesem Artikel wird untersucht, in welcher Beziehung diese Datenprüfungen zu Chi-Quadrat-Tests in der Praxis stehen. Außerdem wird erläutert, wie wir die Richtlinien für die Datenprüfungen im Assistenten festgelegt haben.

Datenprüfungen

Stichprobenumfang

Normalerweise wird ein statistischer Hypothesentest durchgeführt, um einen Beleg für die Zurückweisung der Nullhypothese („keine Differenz“) zu erhalten. Wenn die Stichproben zu klein sind, reicht die Trennschärfe des Tests u. U. nicht aus, um eine Differenz zwischen den tatsächlich vorhandenen Prozentsätzen der fehlerhaften Einheiten 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.

Die Datenprüfung für den Stichprobenumfang basiert auf der Trennschärfe des Tests. Für diese Berechnung muss der Benutzer eine bedeutsame Differenz zwischen einem tatsächlichen Parameter der Grundgesamtheit und dem entsprechenden Wert unter der Nullhypothese angeben. Da es zu schwierig war, diese praktische Differenz für den Chi-Quadrat-Anpassungstest und den Chi-Quadrat-Test auf Assoziation zu ermitteln und auszudrücken, prüft der Assistent lediglich den Stichprobenumfang für den Chi-Quadrat-Test für den Prozentsatz fehlerhafter Einheiten bei mehr als zwei Stichproben.

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 der Trennschärfe und des Stichprobenumfangs basiert auf den in Anhang B aufgeführten Formeln.

Ergebnisse

Wenn die Daten keine ausreichenden Hinweise liefern, die gegen die Nullhypothese sprechen, und Sie keine Differenz mit praktischen Konsequenzen angeben, berechnet der Assistent Differenzen mit praktischen Konsequenzen, die für die Stichprobenumfänge 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.

Für die Prüfung auf Trennschärfe und Stichprobenumfang werden in der Auswertung des Assistenten für den Chi-Quadrat-Test für den Prozentsatz fehlerhafter Einheiten bei mehr als zwei Stichproben die folgenden Statusindikatoren angezeigt:

Status	Bedingung
	Im Test wird eine Differenz zwischen den Prozentsätzen fehlerhafter Einheiten festgestellt, daher stellt die Trennschärfe kein Problem dar. ODER Die Trennschärfe ist ausreichend. Im Test wurde keine Differenz zwischen den Prozentsätzen fehlerhafter Einheiten festgestellt, die Stichprobe ist jedoch umfassend genug, dass die angegebene Differenz mit einer Wahrscheinlichkeit von mindestens 90 % erkannt wird.
	Die Trennschärfe ist möglicherweise ausreichend. Im Test wurde keine Differenz zwischen den Prozentsätzen fehlerhafter Einheiten festgestellt, die Stichprobe ist jedoch umfassend genug, dass die angegebene Differenz mit einer Wahrscheinlichkeit von 80 % bis 90 % erkannt wird. Der erforderliche Stichprobenumfang zum Erzielen einer Trennschärfe von 90 % wird ausgegeben..
	Die Trennschärfe ist möglicherweise nicht ausreichend. Im Test wurde keine Differenz zwischen den Prozentsätzen fehlerhafter Einheiten festgestellt, und die Stichprobe ist umfassend genug, dass die angegebene Differenz mit einer Wahrscheinlichkeit von 60 % bis 80 % erkannt wird. Die erforderlichen Stichprobenumfänge zum Erzielen einer Trennschärfe von 80 % und 90 % werden ausgegeben.
	Die Trennschärfe ist nicht ausreichend (< 60 %). Im Test wurde keine Differenz zwischen den Prozentsätzen fehlerhafter Einheiten festgestellt. Die erforderlichen Stichprobenumfänge zum Erzielen einer Trennschärfe von 80 % und 90 % werden ausgegeben.
	Im Test wurde keine Differenz zwischen den Prozentsätzen fehlerhafter Einheiten festgestellt. Sie haben keine zu erkennende Differenz zwischen den Prozentsätzen fehlerhafter Einheiten 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.

Gültigkeit des Tests

Die Teststatistik χ^2 folgt nur ungefähr einer Chi-Quadrat-Verteilung. Die Approximation verbessert sich mit größeren Stichprobenumfängen. In diesem Abschnitt wird die Approximation ausgewertet, mit der der mindestens erforderliche Stichprobenumfang ermittelt wird, um genaue Ergebnisse zu erhalten.

Die Chi-Quadrat-Approximation an die Teststatistik wird ausgewertet, indem die Auswirkung kleiner erwarteter Zellanzahlen auf die Wahrscheinlichkeit eines Fehlers 1. Art (Alpha) untersucht wird. Durch Verwendung des Fehlers 1. Art zum Auswerten der Gültigkeit des Tests wird eine Regel entwickelt, um Folgendes sicherzustellen:

- Die Wahrscheinlichkeit, dass die Nullhypothese zurückgewiesen wird, wenn sie tatsächlich wahr ist, ist gering und liegt nahe an der gewünschten Wahrscheinlichkeit eines Fehlers 1. Art.
- Der Randbereich der Nullverteilung kann hinreichend genau approximiert werden. Dies ist wichtig, um den p-Wert genau zu berechnen.

Eine kleine erwartete Zellenanzahl wurde gemäß der üblichen Festlegung als eine Zelle definiert, die eine erwartete Zellenanzahl kleiner oder gleich 5 aufweist.

Wir haben zwei Modelle entwickelt, um die Anteile unter der Nullhypothese zu definieren: das Modell mit gestörten Anteilen und das Modell mit gleichen Anteilen. Weitere Informationen finden Sie in Anhang C. Beide Modelle werden in den Simulationen verwendet, die in diesem Artikel an späterer Stelle behandelt werden. Die Modelle werden für jeden der Chi-Quadrat-Tests verwendet, mit einer Ausnahme: Das Modell mit gestörten Anteilen ist nicht auf den Chi-Quadrat-Test für den Prozentsatz fehlerhafter Einheiten bei mehr als zwei Stichproben anwendbar.

Die Datenprüfung auf die Gültigkeit des Tests gilt für alle Chi-Quadrat-Tests im Assistenten. Die einzelnen Datenprüfungen werden im Folgenden beschrieben.

Chi-Quadrat-Anpassungstest

Zielstellung

Wir haben die Chi-Quadrat-Approximation der Teststatistik ausgewertet, indem die Auswirkung von Größe und Häufigkeit der erwarteten kleinen Anzahlen auf die Wahrscheinlichkeit eines Fehlers 1. Art untersucht wurde.

Methode

Stichproben des Umfangs n wurden einer multinomialen Verteilung entnommen, deren Anteile durch das Modell mit gestörten Anteilen oder das Modell mit gleichen Anteilen beschrieben werden (siehe Anhang C). Für jede Bedingung wurden 10.000 Chi-Quadrat-Anpassungstests mit einem Ziel-Signifikanzniveau von 0,05 durchgeführt. Für jeden Test wurde der tatsächliche Fehler 1. Art als $\frac{\textit{Anzahl der zurückgewiesenen Tests}}{\textit{Anzahl der Replikationen}\ (10.000)}$ berechnet. Der Bereich für akzeptable Wahrscheinlichkeiten eines Fehlers 1. Art wurde als [0,03 – 0,07] definiert, und der Mindeststichprobenumfang mit einer Wahrscheinlichkeit eines Fehlers 1. Art in diesem Bereich wurde aufgezeichnet.

Ergebnisse

Die Simulationsergebnisse haben gezeigt, dass Soll-Zellanzahlen kleiner als 1,25 möglicherweise zu falschen p-Werten führen können, wenn der Prozentsatz kleiner erwarteter Zellanzahlen kleiner oder gleich 50 % ist. Außerdem können erwartete Zellanzahlen kleiner als 2,5 ebenfalls zu falschen p-Werten führen, wenn der Prozentsatz kleiner erwarteter Zellanzahlen größer als 50 % ist. Weitere Informationen finden Sie in Anhang D.

Bei der Prüfung der Gültigkeit des Chi-Quadrat-Anpassungstests werden in der Auswertung des Assistenten die folgenden Statusindikatoren angezeigt:

Status	Bedingung
	Die erwartete Mindestzellenanzahl ist größer oder gleich 1,25, wenn der Prozentsatz kleiner erwarteter Zellenanzahlen kleiner oder gleich 50 % ist. ODER Die erwartete Mindestzellenanzahl ist größer oder gleich 2,5, wenn der Prozentsatz kleiner erwarteter Zellenanzahlen größer als 50 % ist. Die Stichprobe ist groß genug, um ausreichende erwartete Anzahlen zu erhalten. Der p-Wert für den Test sollte genau sein.
	Wenn die oben aufgeführten Bedingungen nicht zutreffen.

Chi-Quadrat-Test auf Assoziation

Zielstellung

Wir haben die Chi-Quadrat-Approximation der Teststatistik ausgewertet, indem die Auswirkung von Größe und Häufigkeit der erwarteten kleinen Anzahlen auf die Wahrscheinlichkeit eines Fehlers 1. Art untersucht wurde.

Methode

Stichproben des Umfangs n_i wurden einer multinomialen Verteilung entnommen, deren Anteile durch das Modell mit gestörten Anteilen oder das Modell mit gleichen Anteilen beschrieben werden (siehe Anhang C). Zur Vereinfachung wurde $n_i = n \ \forall i$ gewählt. Für jede Bedingung wurden 10.000 Chi-Quadrat-Tests auf Assoziation mit einem Ziel-Signifikanzniveau von 0,05 durchgeführt. Für jeden Test wurde der tatsächliche Fehler 1. Art als $\frac{Anzahl\ der\ zurückgewiesenen\ Tests}{Anzahl\ der\ Replikationen\ (10.000)}$ berechnet. Der Bereich für akzeptable Wahrscheinlichkeiten eines Fehlers 1. Art wurde als [0,03 – 0,07] definiert, und der Mindeststichprobenumfang mit einer Wahrscheinlichkeit eines Fehlers 1. Art in diesem Bereich wurde aufgezeichnet.

Ergebnisse

Es wurde festgestellt, dass die erwartete Mindestzellenanzahl von der Anzahl der x-Werte und dem Prozentsatz kleiner erwarteter Zellenanzahlen abhängt.

- Für das Modell mit gestörten Anteilen gilt: Wenn der Prozentsatz kleiner erwarteter Zellenanzahlen kleiner oder gleich 50 % ist, sind die erwarteten Mindestzellenanzahlen ≤ 2 und ≤ 1 bei einer Anzahl von x-Werten gleich (2 oder 3) und (4, 5 oder 6). Wenn darüber hinaus der Prozentsatz kleiner erwarteter Zellenanzahlen > 50 % ist, sind die erwarteten Mindestzellenanzahlen ≤ 3 und ≤ 1,5 bei einer Anzahl von x-Werten gleich (2 oder 3) und (4, 5 oder 6).

- Für das Modell mit gleichen Anteilen gilt: Die erwartete Mindestzellenanzahl ist ≤ 2 bei einer Anzahl von x-Werten gleich (2 oder 3), und die erwartete Mindestzellenanzahl ist ≤ 1,5 bei einer Anzahl von x-Werten gleich (4, 5 oder 6).

Weitere Informationen finden Sie in Anhang E.

Bei der Prüfung der Gültigkeit des Chi-Quadrat-Tests auf Assoziation werden in der Auswertung des Assistenten die folgenden Statusindikatoren angezeigt:

Status	Anzahl der x-Variablenwerte	Bedingung
	2 oder 3	Die erwartete Mindestzellenanzahl ist größer oder gleich 2, wenn der Prozentsatz kleiner erwarteter Zellenanzahlen (kleiner oder gleich 5) kleiner oder gleich 50 % ist. Die erwartete Mindestzellenanzahl ist größer oder gleich 3, wenn der Prozentsatz kleiner erwarteter Zellenanzahlen (kleiner oder gleich 5) größer als 50 % ist.
	4, 5 oder 6	Die erwartete Mindestzellenanzahl ist größer oder gleich 1, wenn der Prozentsatz kleiner erwarteter Zellenanzahlen (kleiner oder gleich 5) kleiner oder gleich 50 % ist. Die erwartete Mindestzellenanzahl ist größer oder gleich 2 (1,5 wurde der Einfachheit halber auf 2 gerundet), wenn der Prozentsatz kleiner erwarteter Zellenanzahlen (kleiner oder gleich 5) größer als 50 % ist.
	Alle Fälle	Wenn die oben aufgeführten Bedingungen nicht zutreffen.

Chi-Quadrat-Test für den Prozentsatz fehlerhafter Einheiten bei mehr als zwei Stichproben

Zielstellung

Wir haben die Chi-Quadrat-Approximation der Teststatistik ausgewertet, indem die Auswirkung von Größe und Häufigkeit der erwarteten kleinen Anzahlen auf die Wahrscheinlichkeit eines Fehlers 1. Art untersucht wurde.

Methode

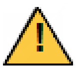
Es wurden die Modelle p = $p_i = p_j \; \forall i, j$ definiert, wobei p = 0,001; 0,005; 0,01; 0,025 und 0,25. Stichproben des Umfangs n_i wurden einer Binomialverteilung mit den oben beschriebenen Werten von p_i entnommen. Zur Vereinfachung wurde $n_i = n \; \forall i$ gewählt. Für jede Bedingung wurden 10.000 Chi-Quadrat-Tests für den Prozentsatz fehlerhafter Einheiten mit einem Ziel-Signifikanzniveau von 0,05 durchgeführt. Für jeden Test wurde der tatsächliche Fehler 1. Art als $\frac{Anzahl\ der\ zurückgewiesenen\ Tests}{Anzahl\ der\ Replikationen\ (10.000)}$ berechnet. Der Bereich für akzeptable Wahrscheinlichkeiten eines Fehlers 1. Art wurde als [0,03 – 0,07] definiert, und der Mindeststichprobenumfang mit einer Wahrscheinlichkeit eines Fehlers 1. Art in diesem Bereich wurde aufgezeichnet.

Ergebnisse

Wenn 3 bis 6 x-Werte vorliegen, ergibt eine erwartete Mindestanzahl fehlerhafter und nicht fehlerhafter Einheiten größer oder gleich 1,5 für den Test eine Wahrscheinlichkeit für einen Fehler 1. Art im Intervall [0,03; 0,07]. Wenn 7 bis 12 x-Werte vorliegen, ergibt eine erwartete Mindestanzahl fehlerhafter und nicht fehlerhafter Einheiten größer oder gleich 1 für den Test eine Wahrscheinlichkeit für einen Fehler 1. Art im Intervall [0,03; 0,07].

Weitere Informationen finden Sie in Anhang F.

Bei der Prüfung der Gültigkeit des Chi-Quadrat-Tests für den Prozentsatz fehlerhafter Einheiten bei mehr als zwei Stichproben werden in der Auswertung des Assistenten die folgenden Statusindikatoren angezeigt:

Status	Anzahl der x-Werte	Bedingung
	3 bis 6	Die erwartete Mindestanzahl fehlerhafter und nicht fehlerhafter Einheiten ist größer oder gleich 1,5.
	7 bis 12	Die erwartete Mindestanzahl fehlerhafter und nicht fehlerhafter Einheiten ist größer oder gleich 1.
	Alle Fälle	Wenn die oben aufgeführten Bedingungen nicht zutreffen.

Gültigkeit der Intervalle

Die Vergleichsintervalle im Chi-Quadrat-Test für den Prozentsatz fehlerhafter Einheiten bei mehr als zwei Stichproben und im Chi-Quadrat-Anpassungstest basieren auf der Normal-Approximation. Darüber hinaus basieren die einzelnen Konfidenzintervalle im Chi-Quadrat-Anpassungstest auf der Normal-Approximation. In diesem Abschnitt wird die Gültigkeit der Normal-Approximation ausgewertet. Gemäß der in den meisten statistischen Fachbüchern angeführten allgemeinen Regel ist das ungefähre Konfidenzintervall genau, wenn die beobachteten Anzahlen mindestens 5 betragen.

Die Datenprüfung auf die Gültigkeit der Intervalle gilt für den Chi-Quadrat-Test für den Prozentsatz fehlerhafter Einheiten bei mehr als zwei Stichproben und den Chi-Quadrat-Anpassungstest.

Chi-Quadrat-Test für den Prozentsatz fehlerhafter Einheiten bei mehr als zwei Stichproben

Zielstellung

Die allgemeine Regel für die Mindestanzahl der in jeder Stichprobe beobachteten fehlerhaften und nicht fehlerhaften Einheiten sollte ausgewertet werden, um sicherzustellen, dass die approximierten Konfidenzintervalle genau sind.

Methode

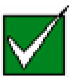
Zunächst definieren wir die Intervalle, die im Vergleichsdiagramm verwendet werden. Die Endpunkte der Intervalle werden so definiert, dass bei einer Gesamtfehlerquote von ungefähr $\propto$ jedes Intervall, das nicht überlappend ist, auf abweichende Prozentsätze fehlerhafter Einheiten der Grundgesamtheit hinweist. Die verwendeten Formeln finden Sie in Anhang G.

Die Vergleichsintervalle basieren auf gepaarten Konfidenzintervallen für Vergleiche. Weitere Informationen finden Sie im Abschnitt zu Vergleichsintervallen im White Paper des Assistenten für die einfache ANOVA. Für jedes Paar (p_i – p_j) wird ein Konfidenzintervall auf der Grundlage der Normal-Approximation verwendet, und dann wird die versuchsbezogene Gesamtfehlerquote mit Hilfe eines Mehrfachvergleichsverfahrens nach Bonferroni kontrolliert. Daher muss nur die Gültigkeit eines der Intervalle im paarweisen Vergleichsverfahren ausgewertet werden, um den Effekt der Normal-Approximation auf die Vergleichsintervalle zu verstehen.

Ergebnisse

Um die Gültigkeit der Normal-Approximation auszuwerten, muss lediglich untersucht werden, wie sich die Approximation auf ein Intervall für die Differenz zwischen Prozentsätzen fehlerhafter Einheiten auswirkt. Aus diesem Grund kann einfach die allgemeine Regel angewendet werden, die für den Fall eines Prozentsatzes fehlerhafter Einheiten bei zwei Stichproben entwickelt wurde. Weitere Informationen finden Sie im Abschnitt „Methoden für den Test für den Prozentsatz fehlerhafter Einheiten bei zwei Stichproben" im White Paper des Assistenten für den Test für den Prozentsatz fehlerhafter Einheiten bei zwei Stichproben. Die Simulationsergebnisse des Tests für den Prozentsatz fehlerhafter Einheiten bei zwei Stichproben deuten darauf hin, dass die Genauigkeit des approximierten Konfidenzintervalls für die Differenz zwischen Prozentsätzen fehlerhafter Einheiten im Allgemeinen zuverlässig ist, wenn die Stichproben ausreichend groß sind, d. h., wenn die beobachtete Anzahl von fehlerhaften Einheiten und die beobachtete Anzahl von nicht fehlerhaften Einheiten in jeder Stichprobe mindestens 5 beträgt.

Bei der Prüfung der Gültigkeit der Intervalle für den Chi-Quadrat-Test für den Prozentsatz fehlerhafte Einheiten bei mehr als zwei Stichproben werden in der Auswertung des Assistenten die folgenden Statusindikatoren angezeigt:

Status	Bedingung
	Alle Stichproben enthalten mindestens 5 fehlerhafte Einheiten und 5 nicht fehlerhafte Einheiten. Die Vergleichsintervalle sollten genau sein.
	Wenn die oben aufgeführte Bedingung nicht zutrifft.

Chi-Quadrat-Anpassungstest

Zielstellung

Die allgemeine Regel für die Mindestanzahl der in jeder Stichprobe beobachteten fehlerhaften und nicht fehlerhaften Einheiten sollte ausgewertet werden, um sicherzustellen, dass die approximierten Konfidenzintervalle genau sind.

Methode

Der Chi-Quadrat-Anpassungstest des Assistenten umfasst Vergleichs- und individuelle Konfidenzintervalle. Es werden die standardmäßigen Intervalle auf der Grundlage der Normal-Approximation für die Anteile verwendet, und die Korrektur für mehrere Intervalle erfolgt mit Hilfe der Korrektur nach Bonferroni (Goodman, 1965). Somit werden simultane Bonferroni-Intervalle wie folgt berechnet:

$$p_{iUntergrenze} = p_i - Z_{\alpha/2k}\sqrt{\frac{\mathrm{p_i}(1 - \mathrm{p_i})}{\mathrm{N}}}$$

$$p_{iObergrenze} = p_i + Z_{\alpha/2k}\sqrt{\frac{\mathrm{p_i}(1 - \mathrm{p_i})}{\mathrm{N}}}$$

Die Endpunkte der Intervalle werden so definiert, dass bei einer Gesamtfehlerquote von ungefähr ∝ jedes Intervall, das den Soll-Anteilswert nicht enthält, darauf hindeutet, dass der tatsächliche Anteil vom entsprechenden Zielanteil abweicht. Die individuellen Intervalle nutzen dieselbe Form wie die Intervalle nach Bonferroni, es erfolgt aber keine Korrektur für mehrere Intervalle mit Hilfe von $Z_{\alpha/2}$.

Ergebnisse

Beide der oben erläuterten Ansätze folgen einer Methodik, die der im Test für den Prozentsatz fehlerhafter Einheiten bei zwei Stichproben des Assistenten beschriebenen Methodik ähnelt. Daher können Regeln für die Gültigkeit der Normal-Approximation angewendet werden, die den für jenen Test entwickelten Regeln ähneln. Weitere Informationen finden Sie im Abschnitt „Methoden für den Test für den Prozentsatz fehlerhafter Einheiten bei zwei Stichproben" im White Paper des Assistenten für den Test für den Prozentsatz fehlerhafter Einheiten bei zwei Stichproben. In dem Dokument wurde geschlussfolgert, dass die Vergleichsintervalle und die individuellen Konfidenzintervalle möglicherweise nicht genau sind, wenn die Anzahlen in den Stichproben kleiner als 5 sind.

Bei der Prüfung der Gültigkeit der Intervalle für den Chi-Quadrat-Anpassungstest werden in der Auswertung des Assistenten die folgenden Statusindikatoren angezeigt:

Status	Bedingung
	Alle Stichprobenanzahlen betragen mindestens 5. Die Intervalle sollten genau sein.

Status	Bedingung
	Es liegen Stichprobenanzahlen vor, die kleiner als 5 sind.

Literaturhinweise

Agresti, A. (1996). An introduction to categorical data analysis. New York, NY: Wiley.

Read, T. und Cressie, N. (1988). Goodness-of-fit statistics for discrete multivariate data. New York, NY: Springer-Verlag.

Fienberg, S. (1980). The analysis of cross-classified categorical data. Cambridge, MA: MIT Press.

Goodman, L. (1965). On simultaneous confidence intervals for multinomial proportions. *Technometrics, 7*, 247-254.

Anhang A: Chi-Quadrat-Teststatistik

Der Assistent verwendet eine Chi-Quadrat-Teststatistik der folgenden Form:

$$x^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Dabei gilt Folgendes:

O_{ij} = beobachtete Anzahlen, wie in der folgenden Tabelle definiert:

Fall	$\boldsymbol{O_{ij}}$
Testen der Güte der Anpassung für eine multinomiale Verteilung	Die beobachtete Anzahl für das Ergebnis von i-ten ist definiert als O_{i1}.
Testen der Gleichheit von mehr als zwei Prozentsätzen fehlerhafter Einheiten	Die beobachtete Anzahl fehlerhafter Einheiten und nicht fehlerhafter Einheiten für die i-te Stichprobe ist jeweils als O_{i1} und als O_{i2} definiert.
Testen der Assoziation zwischen zwei kategorialen Variablen	Die beobachteten Anzahlen für den i-ten Wert der x-Variablen und den j-ten Wert der y-Variablen sind als O_{ij} definiert.

E_{ij} = erwartete Anzahl, wie in der folgenden Tabelle definiert:

Fall	$\boldsymbol{E_{ij}}$
Testen der Güte der Anpassung für eine multinomiale Verteilung	$E_{i1} = np_i$ $i = 1, \ldots, k$ (k = Anzahl der Ergebnisse) n = Stichprobenumfang p_i = historische Anteile $\sum_i p_i = 1$
Testen der Gleichheit von mehr als zwei Prozentsätzen fehlerhafter Einheiten	$E_{i1} = n_i p$ (für fehlerhafte Einheiten) $E_{i2} = n_i(1 - p)$ (für nicht fehlerhafte Einheiten) $i = 1, \ldots, k$ (k = Anzahl der Stichproben) n_i =. i-ten Stichprobenumfang p = Gesamtanteil fehlerhafte Einheiten

Fall	$\boldsymbol{E_{ij}}$
Testen der Assoziation zwischen zwei kategorialen Variablen	$E_{ij} = \frac{(n_{i.}n_{.j})}{n_{..}}$ $i = 1, \ldots, m$ (m = Anzahl der x-Werte) $j = 1, \ldots, k$ (k = Anzahl der y-Werte) $n_{i.}$ = Gesamtanzahl für den i-ten Wert der x-Variablen $n_{.j}$ = Gesamtanzahl für den j-ten Wert der y-Variablen $n_{..}$ = Gesamtstichprobenumfang

Anhang B: Trennschärfe für den Chi-Quadrat-Test für den Prozentsatz fehlerhafter Einheiten bei mehr als zwei Stichproben

Es wird eine nichtzentrale Chi-Quadrat-Verteilung verwendet, um die Trennschärfe des Tests zu berechnen, dass $p_i = p_j = p \; \forall i, j$. Der Nichtzentralitätsparameter hängt von n_i und $p_i \forall i$ ab.

Dabei gilt Folgendes:

n_i = der Stichprobenumfang für die i-te Stichprobe

Jedes p_i stellt einen alternativen Anteil dar (siehe den nächsten Abschnitt in diesem Anhang, „Berechnung alternativer Anteile"), der anhand der Anteilsdifferenz = δ berechnet wird.

Der Nichtzentralitätsparameter der Chi-Quadrat-Verteilung wird wie folgt berechnet:

$$\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Dabei gilt Folgendes:

$$O_{i1 =} n_i p_i$$

$$O_{i2 = n_i (1 - p_i)}$$

Die Trennschärfe des Tests wird wie folgt berechnet:

$$\text{Prob}(X \geq x_{1-\alpha} \mid \chi^2)$$

Dabei gilt Folgendes:

X = ist eine Zufallsvariable aus einer nichtzentralen Chi-Quadrat-Verteilung mit dem Nichtzentralitätsparameter χ^2.

$x_{1-\alpha}$ = inverse kumulative Verteilungsfunktion, ausgewertet bei $1 - \alpha$ für eine zentrale Chi-Quadrat-Verteilung.

Berechnung alternativer Anteile

Die alternativen Anteile wurden wie folgt definiert:

$$p_i = p_c + \frac{n_j}{n_i + n_j} \delta$$

$$p_j = p_c - \frac{n_i}{n_i + n_j} \delta$$

$$p_m = p_c \forall m \neq i, j$$

$$0 < \delta < 1$$

Dabei gilt Folgendes:

$$p_c = \frac{1}{N_T}\sum_{i=1}^{k} n_i \widehat{p_\iota}$$

$\widehat{p_\iota}$ = Anteil fehlerhafter Einheiten in der i-ten Stichprobe

N_T = Gesamtanzahl der Beobachtungen

n_i = Stichprobenumfang für die i-te Stichprobe

Für einige Differenzen δ ist $p_i > 1$ oder $p_j < 0$. Daher werden folgende Regeln aufgestellt:

Wenn $p_j < 0$	$p_i = \delta$ $p_j = 0$ $p_m = \frac{\delta}{2}\ \forall m \neq i, j$
Wenn $p_i > 1$	$p_i = 1$ $p_j = 1 - \delta$ $p_m = 1 - \frac{\delta}{2}\ \forall m \neq i, j$

Die Verwendung der zwei kleinsten Werte von n_i ergibt die minimale Trennschärfe, und die Verwendung der zwei größten Werte von n_i ergibt die maximale Trennschärfe.

Anhang C: Modell mit gestörten Anteilen und Modell mit gleichen Anteilen

Modell mit gestörten Anteilen

Nach Read und Cressie (1988) wird der Satz von Anteilen unter der Nullhypothese hier wie folgt definiert:

Wir haben δ nahe k - 1 gewählt (wobei k = Anzahl der Anteile für jede Stichprobe), und wir definieren einen Satz kleiner p_i als:

$$p_i = \frac{(1-\frac{\delta}{k-1})}{k} \text{ für } i = 1, \dots, r$$

und die verbleibenden p_i als:

$$p_i = \frac{(1-\sum_{i=1}^{r} p_i)}{(k-r)} \text{ für } i = r+1, \dots, k$$

Die in den Simulationen für δ verwendeten Werte werden in Tabelle 1 aufgeführt.

Tabelle 1 In den Simulationen verwendetes δ mit resultierenden kleinen p_i

k	δ	$p_{i=1,\dots,r}$
3	1,95	0,008
4	2,95	0,004
5	3,90	0,005
6	4,90	0,003

Für jedes k wurde r = 1,..., k – 1 variiert, um die Größe des Satzes kleiner p_i's. zu ändern. Zum Beispiel wurden für k = 3 die folgenden zwei in Tabelle 2 beschriebenen Modelle erhalten.

Tabelle 2 Die Werte von p_i für k = 3 unter Verwendung des Modells mit gestörten Anteilen

r	p1	p2	p3
1	0,008	0,496	0,496
2	0,008	0,008	0,984

Modell mit gleichen Anteilen

Um ein Modell zu erhalten, bei dem 100 % der erwarteten Zellenanzahlen klein sind, wird ein Modell mit gleichen Anteilen verwendet, definiert durch

$$p_i = \frac{1}{k} \forall\, i$$

Bei Verwendung dieses Modells und einem sehr kleinen Stichprobenumfang werden alle erwarteten Zellenanzahlen als klein betrachtet. Bei einem Modell mit gleichen Anteilen müssen die Stichprobenumfänge sehr klein sein, um eine kleine erwartete Zellenanzahl zu erzielen; dies tritt in der Praxis wahrscheinlich nicht auf.

Anhang D: Gültigkeit des Chi-Quadrat-Anpassungstests

Für das Modell mit gestörten Anteilen wurde die erwartete Mindestzellenanzahl, die erforderlich ist, um eine Wahrscheinlichkeit eines Fehlers 1. Art im Intervall [0,03; 0,07] zu erzielen, im Vergleich zum Prozentsatz kleiner erwarteter Zellenanzahlen grafisch dargestellt, wie in Abbildung 1 gezeigt.

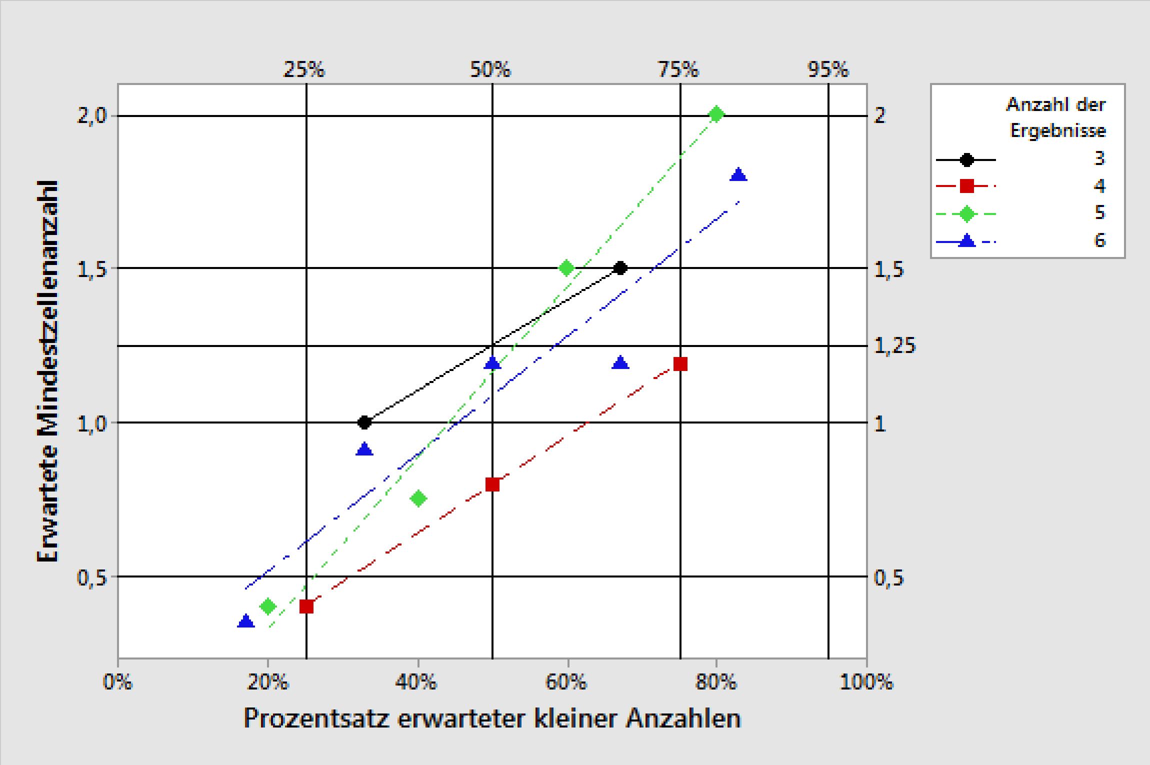


Abbildung 1 Erwartete Mindestzellenanzahlen, die erforderlich sind, um eine Wahrscheinlichkeit eines Fehlers 1. Art im Intervall [0,03; 0,07] zu erzielen, im Vergleich zum Prozentsatz kleiner erwarteter Zellenanzahlen

Abbildung 1 zeigt: Wenn der Prozentsatz kleiner erwarteter Zellenanzahlen kleiner als 50 % ist, sind die erwarteten Mindestzellenanzahlen kleiner oder gleich 1,25. Alle erwarteten Mindestzellenanzahlen sind kleiner oder gleich 2. Basierend auf diesen Simulationsergebnissen sind die in der Auswertung des Assistenten verwendeten Regeln konservativ.

Anschließend wurde dieselbe Simulation unter Verwendung des Modells mit gleichen Anteilen durchgeführt, um die Nullverteilung zu definieren. In Tabelle 4 werden die Ergebnisse der Simulation mit einem Modell mit gleichen Anteilen zusammengefasst.

Tabelle 4 Erwartete Mindestzellenanzahl, um eine Wahrscheinlichkeit eines Fehlers 1. Art im Intervall [0,03; 0,07] zu erzielen

k	Erwartete Mindestzellenanzahl
3	2,5
4	1,25
5	1
6	1,4

Wie oben erwähnt, führt das Modell mit gleichen Anteilen zu Fällen, in denen 100 % der Zellenanzahlen klein sind. Tabelle 4 zeigt, dass alle erwarteten Mindestzellenanzahlen kleiner oder gleich 2,5 sind. Dies untermauert die Regeln, die in der Auswertung des Assistenten verwendet werden.

Anhang E: Gültigkeit des Chi-Quadrat-Tests auf Assoziation

Für das Modell mit gestörten Anteilen wurde die erwartete Mindestzellenanzahl, die erforderlich ist, um eine Wahrscheinlichkeit eines Fehlers 1. Art im Intervall [0,03; 0,07] zu erzielen, im Vergleich zum Prozentsatz kleiner erwarteter Zellenanzahlen für jede Anzahl von x-Werten grafisch dargestellt, wie in Abbildung 2 gezeigt.

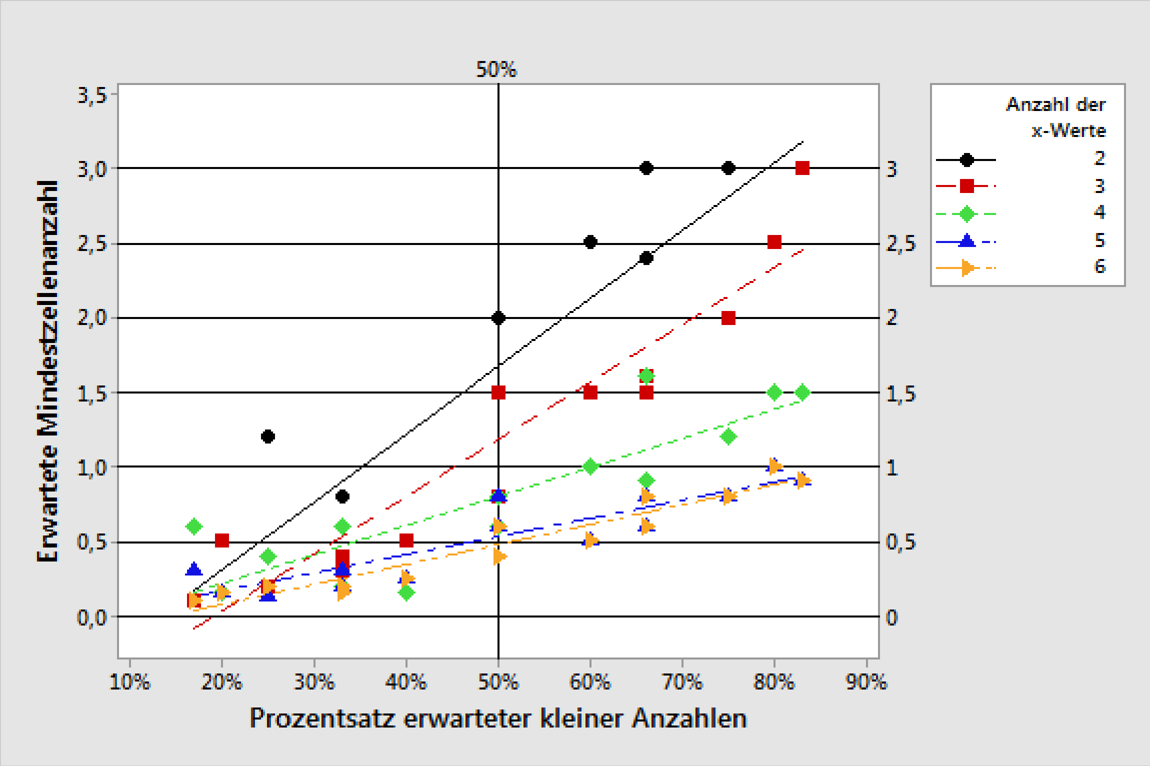


Abbildung 2 Erwartete Mindestzellenanzahlen, die erforderlich sind, um eine Wahrscheinlichkeit eines Fehlers 1. Art im Intervall [0,03; 0,07] zu erzielen, im Vergleich zum Prozentsatz kleiner erwarteter Zellenanzahlen

Abbildung 2 deutet darauf hin, dass die erwartete Mindestzellenanzahl von der Anzahl der x-Werte und dem Prozentsatz kleiner erwarteter Zellenanzahlen abhängt.

In Abbildung 2 wird deutlich: Wenn der Prozentsatz kleiner erwarteter Zellenanzahlen ≤ 50 %, ist, sind die erwarteten Mindestzellenanzahlen ≤ 2 und ≤ 1 bei einer Anzahl von x-Werten gleich 2 oder 3 bzw. 4, 5 oder 6. Wenn darüber hinaus der Prozentsatz kleiner erwarteter Zellenanzahlen > 50 % ist, sind die erwarteten Mindestzellenanzahlen ≤ 3 und ≤ 1,5 bei einer Anzahl von x-Werten gleich 2 oder 3 bzw. 4, 5 oder 6.

Für das Modell mit gleichen Anteilen wurde die erwartete Mindestzellenanzahl im Vergleich zur Anzahl der x-Werte (m) und zur Anzahl der y-Werte (k) grafisch dargestellt, wie in Abbildung 3 gezeigt.

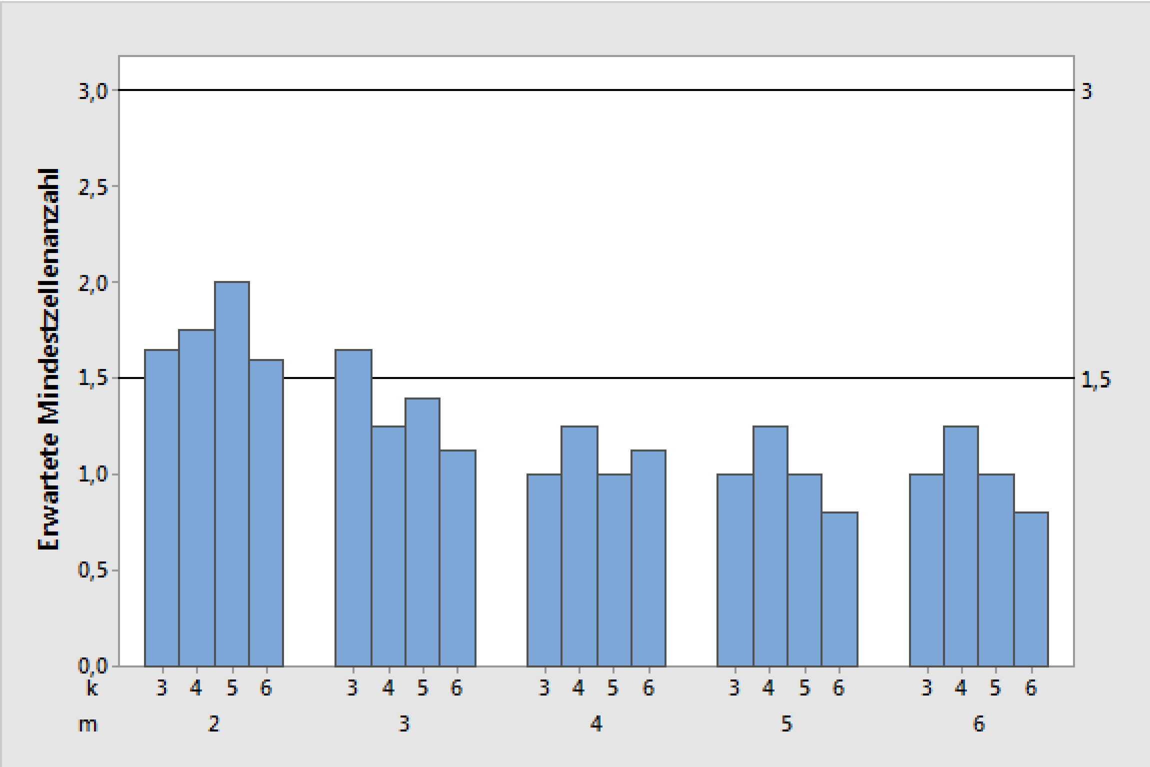


Abbildung 3 Erwartete Mindestzellenanzahlen, die erforderlich sind, um eine Wahrscheinlichkeit eines Fehlers 1. Art im Intervall [0,03; 0,07] zu erzielen, im Vergleich zu den x-Werten (m) und den y-Werten (k)

Abbildung 3 weist darauf hin, dass die erwartete Mindestzellenanzahl ≤ 2 ist, wenn die Anzahl der x-Werte gleich 2 oder 3 ist, und dass die erwartete Mindestzellenanzahl ≤ 1,5 ist, wenn die Anzahl der x-Werte gleich 4, 5 oder 6 ist. Basierend auf diesen Simulationsergebnissen sind die Regeln in der Auswertung des Assistenten konservativ.

Anhang F: Gültigkeit des Chi-Quadrat-Tests für den Prozentsatz fehlerhafter Einheiten bei mehr als zwei Stichproben

Für jedes p und jedes m = 3, 4, 5,..., 12 wurde die erwartete Mindestzellenanzahl grafisch dargestellt. Die Ergebnisse werden in den Abbildungen 4 und 5 gezeigt.

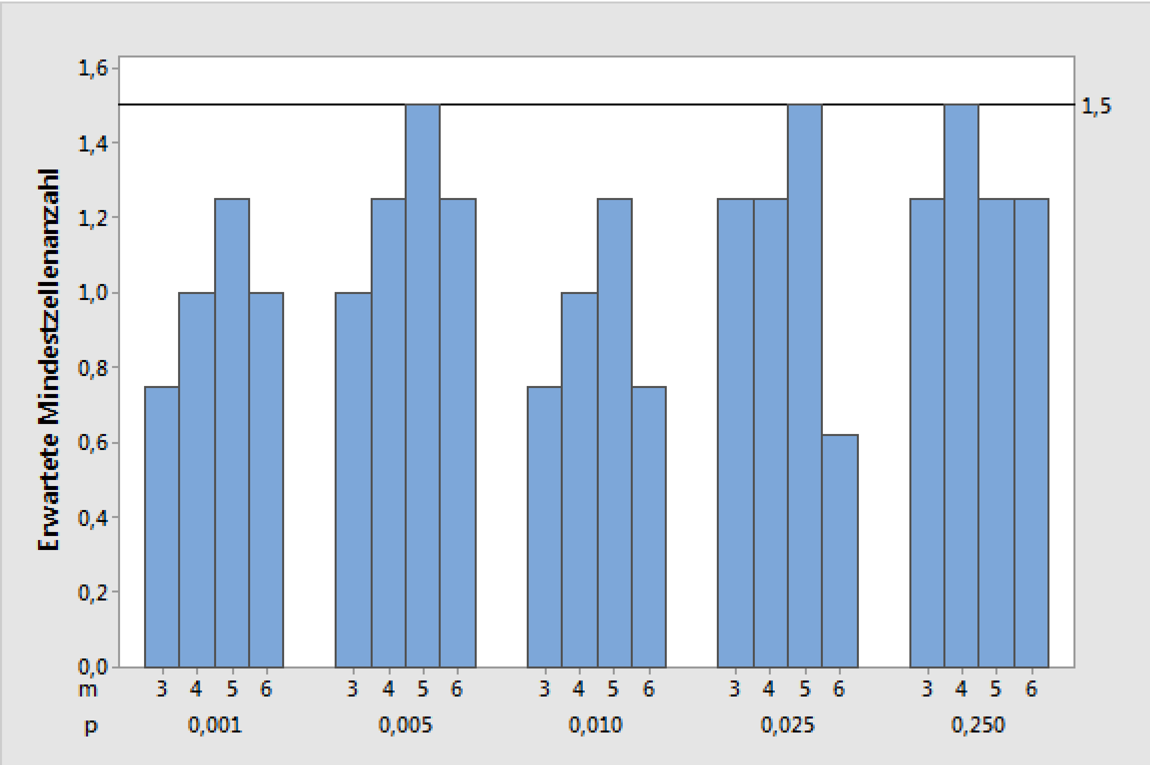


Abbildung 4 Erwartete Mindestzellenanzahlen, die erforderlich sind, um eine Wahrscheinlichkeit eines Fehlers 1. Art im Intervall [0,03; 0,07] zu erzielen, im Vergleich zur Anzahl der x-Werte (m = 3 bis 6)

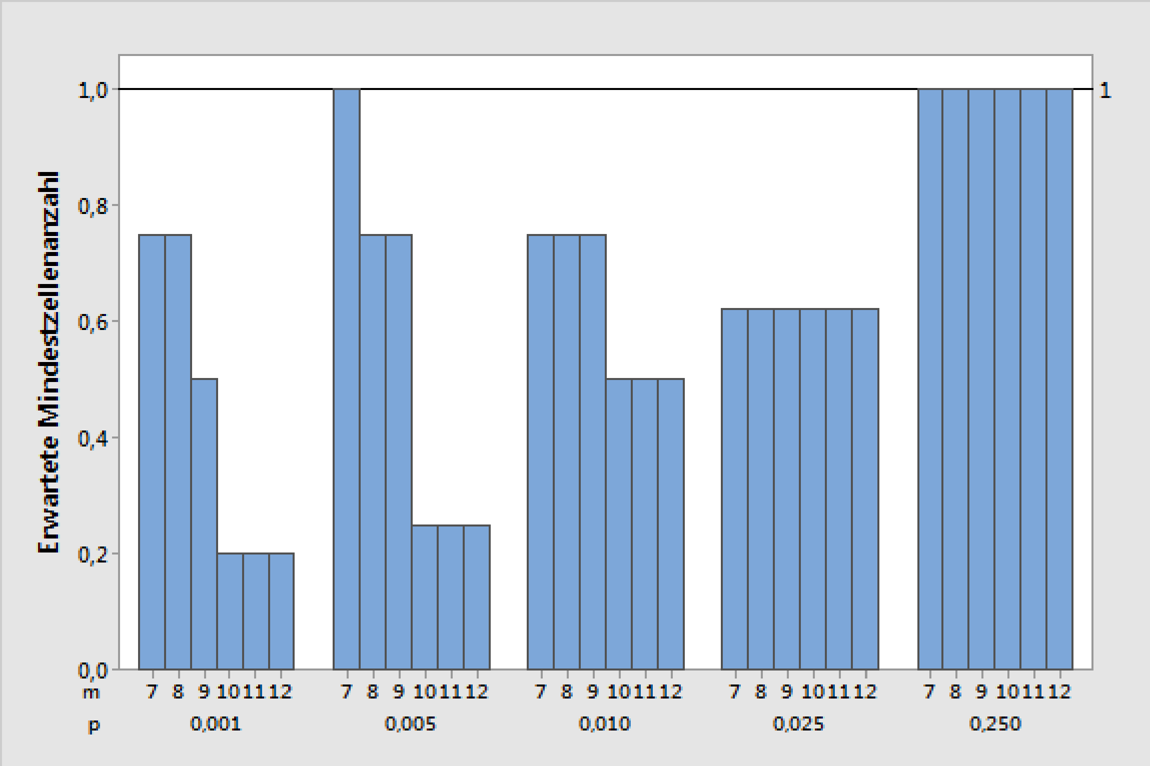


Abbildung 5 Erwartete Mindestzellenanzahlen, die erforderlich sind, um eine Wahrscheinlichkeit eines Fehlers 1. Art im Intervall [0,03; 0,07] zu erzielen, im Vergleich zur Anzahl der x-Werte (m = 7 bis 12)

Wenn die Anzahl der x-Werte gleich 3, 4, 5 oder 6 ist, ergibt eine erwartete Zellenanzahl größer oder gleich 1,5 für den Test eine Wahrscheinlichkeit eines Fehlers 1. Art im Intervall [0,03; 0,07]. Wenn die Anzahl der x-Werte gleich 7, 8, 9,..., 12 ist, ergibt eine erwartete Zellenanzahl größer oder gleich 1 für den Test eine Wahrscheinlichkeit eines Fehlers 1. Art im Intervall [0,03; 0,07].

Anhang G: Vergleichsintervalle für Chi-Quadrat-Tests für den Prozentsatz fehlerhafter Einheiten bei mehr als zwei Stichproben

Die Unter- und Obergrenze für p_i werden wie folgt definiert:

$$p_{iUntergrenze} = p_i - Z_{\alpha/c} X_i$$

$$p_{iObergrenze} = p_i + Z_{\alpha/c} X_i$$

Dabei gilt Folgendes:

c = Anzahl der Vergleiche = k (k - 1) /2

Hierbei ist k die Anzahl der Stichproben.

$Z_{\alpha/c}$ = $(1 - \frac{\alpha}{2c})$. Perzentil für eine Normalverteilung mit dem Mittelwert = 0 und der Standardabweichung = 1

$$X_i = ((k - 1)\sum_{j \neq i} b_{ij} - \sum\sum_{1 \leq j < l \leq k} b_{jl}) / ((k - 1)(k - 2))$$

Dabei gilt Folgendes:

$$b_{ij} = \sqrt{\frac{p_i(1 - p_i)}{n_i} + \frac{p_j(1 - p_j)}{n_j}}$$