

GESStabs

Spaltenweise Signifikanztests



Gesellschaft für Software
in der Sozialforschung mbH

Waterloohain 6 - 8
22769 Hamburg
Tel.: 040 - 853 753 - 0
Fax: 040 - 853 753 - 33
www.gessgroup.de

Inhaltsübersicht

| | |
|---|----|
| Überblick | 3 |
| Signifikanz und Hypothese..... | 3 |
| Signifikanz und Gewichtung | 3 |
| Auswahl von Verfahren für Signifikanztests bei Gewichtung | 4 |
| Standard: Summe der Gewichte | 4 |
| Ungewichtet..... | 4 |
| Doppelter Test: Crosscheck..... | 5 |
| Effective Sample Size | 5 |
| Hybride Signifikanztests | 5 |
| | |
| Systematik der GESStabs CELLEMENTS für Signifikanztests zwischen Spalten ... | 6 |
| Anmerkungen zur Verwendung von ESS in Signifikanztests bei GESStabs..... | 8 |
| „Lokale“ Berechnung vs. „Globale“ Berechnung von ESS..... | 8 |
| Rechenbeispiel für χ^2 und ESS | 8 |
| Rechenbeispiel für McNemar und ESS..... | 10 |
| | |
| Anmerkungen zu den Rechenverfahren in hybriden Signifikanztests bei GESStabs. | 12 |
| Unabhängiger hybrider t-Test für zwei Stichproben (HYTTTEST)..... | 12 |
| Abhängiger hybrider t-Test (HYCOLDEPTTEST)..... | 12 |
| Unabhängiger hybrider Test auf Anteilsunterschieden (HYCOLCHIQU) | 12 |
| Abhängiger hybrider Test auf Anteilsunterschiede (HYMCNEMAR)..... | 13 |
| | |
| Weitere Formen von Signifikanztests in GESStabs | 13 |

Überblick

Die Ausweisung der Signifikanz der Unterschiede zwischen den Werten in verschiedenen Spalten derselben Zeile von Tabellen wird bei Tabellen oft verlangt. Es hat sich eingebürgert, die Spalten mit Kennbuchstaben zu versehen, und in jeder Spalte die übrigen Spalten zu benennen, soweit die Unterschiede auf einem gegebenen „Level“ signifikant sind. Entsprechende CELLELEMENTS gibt es in GESStabs seit Langem, z.B. COLCHIU oder TTEST. Die Menge der Testverfahren wurde immer wieder erweitert. Es war an der Zeit, dies Angebot etwas systematischer zu erweitern.

Signifikanz und Hypothese

Nach der statistischen Logik steht hinter jeder ausgewiesenen Signifikanz ein Hypothesentest. Hierfür gibt es formal immer eine Hypothese (h_1), z.B. die Mittelwerte der Variablen X in den Gruppen A und B sind unterschiedlich. Als logisches Gegenstück zu h_1 gibt es die sogenannte Nullhypothese (h_0), in unserem Beispiel wäre das: Der Mittelwert der Variablen X in der Gruppe A unterscheidet sich nicht von dem in der Gruppe B. Man spricht dann von einer Signifikanz des Unterschieds, wenn der Versuch scheitert, die Nullhypothese zu widerlegen. Und der Versuch gilt dann als gescheitert, wenn es hinreichend unwahrscheinlich ist, dass ein in der Stichprobe vorgefundener Unterschied nur darauf beruht, dass die Stichprobe eben nur ein Teil der Grundgesamtheit ist. Was als hinreichend unterschiedlich gelten soll, legt der Betrachter selbst fest: Wenn ein Signifikanzniveau von 5% verwendet wird, bedeutet das, dass es einem ausreicht, dass die Wahrscheinlichkeit kleiner als 5% ist, dass man die Existenz eines Unterschieds behauptet, der in der Realität gar nicht besteht. Anders formuliert: Wenn ich viele Zufallsstichproben aus einer Grundgesamtheit ziehen würde, in der dieser Unterschied nicht existiert, würden nur 5% dieser Stichproben einen Unterschied in der Größe ausweisen, wie die Stichprobe ausweist. Umgekehrt darf ich dann mit einer Wahrscheinlichkeit von 95% davon ausgehen, dass es den behaupteten Unterschied gibt.

Den theoretischen Überlegungen hierfür liegt ein Urnenmodell der Zufallsstichprobe zugrunde. Je mehr zufällige Ziehungen eine Stichprobe ausmachen, desto geringer wird die Verzerrung, die von der Auswahl ausgeht. Die verwendeten Testgrößen und Verteilungsfunktionen (z.B. t-Verteilung oder Chiquadratverteilung) geben grob gesprochen die Wahrscheinlichkeit an, bei welcher Anzahl von Zufallsziehungen aus der Urne (=Stichprobengröße) ein beobachteter Unterschied durch Zufallsprozesse entstehen kann.

Signifikanz und Gewichtung

Die Frage nach der Signifikanz eines in der Stichprobe beobachteten Unterschieds z.B. zwischen zwei Mittelwerten ist deshalb bei gewichteten Daten nicht ganz so einfach zu beantworten. Inhaltlich ist die Frage nach der Signifikanz eines Unterschieds eines Messwerts in zwei Stichproben nach dem oben gesagten die Frage danach, wie (un)wahrscheinlich es ist, dass man aus der Grundgesamtheit eine Zufallsstichprobe mit einem Unterschied (z.B. zwischen den Männern und den Frauen) zieht, obwohl dieser in der Grundgesamtheit nicht existiert. Grundlage aller Wahrscheinlichkeitsabschätzungen ist beim Ziehungsmodell aus der theoretischen Urne, dass jedem Element ein (zufälliger) Ziehungsvorgang entspricht; in der tabellarischen Auswertung ist das aber nur gegeben, wenn alle Fälle das Gewicht 1.0 haben. Das aber ist bei gewichteter

Tabellierung offensichtlich nicht der Fall: wenn ein Befragter das Gewicht von z.B. 2,34 hat, konnte er trotzdem nur 1.0-mal zufällig aus der Grundgesamtheit ausgewählt werden, und die Anzahl dieser Auswahlvorgänge ist Ausgangspunkt der Überlegungen zur Signifikanz. Die Ausweisung von Signifikanzen in gewichteten Tabellen ist also nicht unproblematisch und die verschiedenen möglichen Vorgehensweisen sind mit Vor- und Nachteilen behaftet, die man gegeneinander abwägen muss.

An dieser Stelle ist zu betonen, dass die Signifikanztests logisch alle vom Modell der einfachen Zufallsstichprobe ausgehen. Diesem Ideal der Urne kann man sich in der Praxis nur annähern. Bei allen Stichproben, die nicht nach diesem Modell gezogen sind, ist die Frage nach der Signifikanz eigentlich gar nicht beantwortbar.

Auswahl von Verfahren für Signifikanztests bei Gewichtung

GESStabs bietet für die gängigsten Fragestellungen durchgängig vier Verfahren an, wie man mit diesem Problem umgehen kann, die weiter unten die vier Spalten der Matrix bilden.

Standard: Summe der Gewichte

Im Standardfall verwendet GESStabs die Summe der Gewichte als Grundlage der Signifikanzberechnungen. Das ist nach dem oben Ausgeführten bei gewichteten Datensätzen nicht wirklich korrekt. Es ist als Näherung vertretbar, solange die Gewichte sich nicht erheblich von 1.0 unterscheiden. Da die Gesamtsumme der Gewichte der Anzahl der Fälle entspricht, kommt es zu einer gewissen Nivellierung der Effekte, da Gewichte größer als 1 eben auch Gewichte kleiner als 1 bedingen. Der Vorteil dieser einfachen Methode ist, dass die Berechnungen zu Anteilwerten und zur Signifikanz auf denselben Daten beruhen: Ergebnisse und Signifikanzdarstellung sind in diesem Sinne kongruent.

Man verwendet im Falle der gewichteten Darstellung einfach dieselben CELLELEMENTS wie auch bei der ungewichteten: TTEST, COLCHIQU etc.

Ungewichtet

Eine Alternative ist, für die Signifikanzberechnung die ungewichteten Werte heranzuziehen. Man weist also dann die Signifikanz für einen Unterschied aus, wenn dieser Zusammenhang in den ungewichteten Rohdaten besteht. Damit ist die Bedingung erfüllt, dass jeder Fall mit derselben Chance in die Stichprobe gekommen ist. Das ist der Vorteil; der Nachteil ist, dass ein Unterschied getestet wird, der so in den Tabellen nicht ausgewiesen ist und numerisch davon abweichen kann.

Den entsprechenden CELLELEMENTS ist zur Kennzeichnung der ungewichteten Vorgehensweise die Kennung „PHYS“ wie PHYSICALRECORD vorangestellt: PHYSTTEST, PHYSCOLCHIQU etc.

Doppelter Test: Crosscheck

Es bietet sich unmittelbar an, diese Methoden miteinander zu verknüpfen. GESStabs berechnet die Signifikanz nach der Methode (1) auf der Basis der Fallgewichte, und im Anschluss wird überprüft, ob dieser Unterschied auf der Basis der ungewichteten Fälle (2) ebenfalls besteht. GESStabs weist dann Signifikanz nur in dem Maße aus, wie beide Methoden dies unterstützen. Hiermit werden solche Signifikanzausweisungen unterdrückt, die lediglich auf der Basis der gewichteten Daten bestehen, also in gewissem Sinne erst durch die Gewichtung produziert wurden.

Die hierfür zu verwendenden CELLELEMENTS sind durch ein vorangestelltes „X“ gekennzeichnet: XTTEST, XCOLCHIQU usw.

Effective Sample Size

Es wird deutlich, dass es bei der Betrachtung von Signifikanzen bei gewichteter Darstellung vor allem darum geht, Artefakte auszuschließen. Eine andere Möglichkeit (4) hierzu ist es, bei der Analyse des Unterschieds statt der Summe der Gewichte eine „effektive Basis“ zu verwenden, in der die Abweichung der Gewichte von 1.0 berücksichtigt wird. Die hierfür üblicherweise¹ verwendete „effective sample size“² heißt in GESStabs ESS und steht auch als eigenes CELLELEMENT zur Verfügung. In diese Größe fließen die Summe der Gewichte (G) und die Summe der Quadrate der Gewichte (SQ) ein:

$$ESS = G * \frac{G}{SQ}$$

Es ist leicht zu sehen, dass ESS immer dann gleich der Anzahl der Fälle ist, wenn alle Gewichte die Größe 1.0 haben. Je weiter sich die Gewichte von 1.0 entfernen, um so kleiner wird die effektive Stichprobengröße ESS, da die Summe der Quadrate wächst. Daraus folgt, dass eine Signifikanzberechnung bei einem gegebenen Unterschied immer seltener eine kleine Fehlerwahrscheinlichkeit hat, je stärker die Gewichte sich von 1.0 unterscheiden.

Die entsprechenden CELLELEMENTS in GESStabs sind durch die Vorsilbe „ESS“ gekennzeichnet.

Hybride Signifikanztests

Dank des Vorschlags eines guten Kunden ist noch eine weitere Schicht von Signifikanztests für gewichtete Daten hinzugekommen, in die die ungewichteten und gewichteten Zählergebnisse in neuartiger „hybrider“ Art und Weise miteinander verrechnet werden. Grob kann man diese hybride Form dadurch kennzeichnen, dass die gewichteten Ergebnisse betrachtet werden (z.B. wird die Mittelwertdifferenz auf der Basis der Gewichtung errechnet), bei den statistischen Tests werden aber die absoluten ungewichteten Häufigkeiten herangezogen um die Fehlerwahrscheinlichkeit zu bestimmen. Als Größe der Stichproben wird also immer die tatsächliche Fallzahl verwendet.

Die entsprechenden CELLELEMENTS in GESStabs sind durch die Vorsilbe „HY“ gekennzeichnet.

¹ So geht auch Quantum vor, siehe User Manual, Vol. 3

² Diese wird in der Literatur oft dem Statistiker Leslie Kish zugeschrieben. (Vgl. Leslie Kish, Survey Sampling, New York, London, Sydney 1965)

Systematik der GESStabs CELLEMENTS für Signifikanztests zwischen Spalten

Für die Auswahl eines Verfahrens muss man inhaltlich folgende Fragen entscheiden:

Was soll verglichen werden: **Anteilswerte** oder **Mittelwerte** von Variablen?

Abhängige oder **unabhängige** Fragestellung: Geht es im Vergleich um verschiedene Variablen in einem Sample oder um gleiche Variablen in verschiedenen Samples?

Wie sollen **Gewichte** berücksichtigt werden: gewichtet, ungewichtet, Crosscheck, ESS oder hybrid?

In einigen Konstellationen bietet GESStabs mehr als ein statistisches Verfahren an. Dann ist noch die Frage zu beantworten: Welche **Teststatistik** soll verwendet werden, z.B. t-Test oder χ^2 ?

Die folgende Tabelle ist anhand dieser Fragen gegliedert:

| Gewichtet | Ungewichtet | ESS | Crosscheck | Hybride Tests |
|------------------|--------------------|------------|-------------------|----------------------|
|------------------|--------------------|------------|-------------------|----------------------|

Anteilswerte (%):

unabhängige Variablen (Vergleich jeweils zweier Samples)

| | Gewichtet | Ungewichtet | ESS | Crosscheck | Hybride Tests |
|------------------|------------------|--------------------|----------------------------|-------------------|----------------------|
| χ^2 | COLCHIQU | PHYSCOLCHIQU | ESSCOLCHIQU | XCOLCHIQU | HYCOLCHIQU |
| kombiniert mit % | colpercandchiqu | | | | colpercandhychiqu |
| z-Test | COLPERCZ | | | | |
| t-Test | | | COLPERCT colpercandsign | | |

abhängige Variablen (Vergleich zweier Variablen im selben Sample)

| | Gewichtet | Ungewichtet | ESS | Crosscheck | Hybride Tests |
|------------------|---------------------------|--------------------|------------|-------------------|-------------------------------|
| kombiniert mit % | MCNEMAR colpercmcnemar | PHYSMCNEMAR | ESSMCNEMAR | XMCNEMAR | HYMCNEMAR colperchymcnemar |

Mittelwerte:

unabhängige Variablen (Vergleich jeweils zweier Samples)

| | Gewichtet | Ungewichtet | ESS | Crosscheck | Hybride Tests |
|---------------------|------------------|--------------------|-------------|-------------------|----------------------|
| t-Test | TTEST | PHYSTTEST | ESSTTEST | XTTEST | HYTTEST |
| kombiniert mit MEAN | meantest | physmeantest | essmeantest | xmeantest | hymeantest |
| t-Test für MEANCUT | TTESTCUT | | | | |

abhängige Variablen (Vergleich zweier Variablen im selben Sample)

| | Gewichtet | Ungewichtet | ESS | Crosscheck | Hybride Tests |
|---------------------|------------------|--------------------|----------------|-------------------|----------------------|
| t-test | COLDEPTTEST | PHYSCOLDEPTTEST | ESSCOLDEPTTEST | XCOLDEPTTEST | HYCOLDEPTTEST |
| kombiniert mit MEAN | meancoldept | physmeancoldept | essmeancoldept | xmeancoldept | Hymeancoldept |

In den Spalten der Tabelle sind die CELLEMENTS anhand des Vorgehens bei gewichteten Daten dargestellt. Bei ungewichteten Daten ist das Ergebnis in den Spalten identisch.

In den Zeilen der Tabelle sind die Verfahren anhand der Fragestellung aufgelistet. Für die abhängigen und unabhängigen Test auf Mittelwertsunterschiede und Prozentwerte ist jeweils eine Zeile komplett gefüllt. Generell gibt es oft neben dem Signifikanztest als atomarem CELLELEMENT auch die Zusammenfassung mit dem Zahlenwert, also zusätzlich zum TTEST der MEANTEST.

Anmerkungen zur Verwendung von ESS in Signifikanztests bei GESStabs

Für den Test auf Mittelwertsunterschiede bei unabhängiger Fragestellung (also i.d.R. dieselbe Variable in verschiedenen Teilgruppen) sind als neue Behandlung der Gewichtung jetzt also ESSTEST, HYTTEST und XTTEST hinzugekommen. Der Crosscheck bei XTTEST ist rechnerisch einfach: es werden nacheinander beide Tests durchgeführt, das Resultat ist das Niveau, das von beiden Verfahren „gehalten“ wird.

„Lokale“ Berechnung vs. „Globale“ Berechnung von ESS

Das Vorgehen beim ESSTEST ist etwas komplexer. Grundsätzlich gibt es zwei Wege, die effektive Stichprobengröße in die Berechnung einzuführen: global und lokal. Beim globalen Vorgehen ergibt sich als Quotient aus dem ESS für die gesamte Stichprobe und der Summe der Gewichte ein konstanter Faktor ≤ 1 , mit dem alle Fallgewichte multipliziert werden. Die Summe der Gewichte aus einer Stichprobe von 1000 Befragten ergibt dann für alle Fälle ein ESS von z.B. 789. Auf dieser Basis wird dann sowohl tabelliert als auch die Signifikanz berechnet.

In GESStabs wird bei den mit ESS gekennzeichneten Verfahren immer das lokale Vorgehen angewendet. Beim Vergleich von Mittelwerten in zwei Teilgruppen werden die ESS für jeweils diese Teilgruppen berechnet, bei denen die Abweichung des Gewichts von 1.0 unterschiedlich sein kann. Die Anzahl der Freiheitsgrade für die Beurteilung des Unterschieds wird dann auf der Basis der Summe der einzelnen ESS bestimmt³. Im Prinzip ist das beim abhängigen t-Test ebenso wie beim unabhängigen: hier sind nicht zwei sondern nur eine Teilgruppe betroffen, deren ESS nach derselben Formel berechnet wird; der Unterschied beruht darauf, dass beim abhängigen Test nicht die zwei Mittelwerte verglichen werden, sondern der Mittelwert der Differenz der Einzelwerte wird auf Unterschied zu 0.0 getestet.

Soll die effektive Basis global in die Tabellierung einfließen, dann ist das relativ einfach zu realisieren: GESStabs liefert die ESS auch als isoliertes CELLELEMENT unter dem Namen ESS. Man kann dann den Quotienten aus ESS und der Summe der Gewichte bilden. In einem zweiten Schritt sind dann die Fallgewichte mit diesem konstanten Faktor zu multiplizieren.

Rechenbeispiel für χ^2 und ESS

Etwas anders sieht es beim Signifikanztest auf Prozentwertunterschiede aus. Wenn zwei Anteilswerte (Prozent) über Chiquadrat miteinander verglichen werden, handelt es sich intern immer um Tabellen mit einem Freiheitsgrad ($df=1$). Die Größe der Stichprobe geht über die Zellenbesetzung in die Abschätzung der Signifikanz ein. Bei ungewichteten Tabellen ist das die Zahl der Fälle, bei gewichteter Betrachtung ist das die Summe der Gewichte, und wenn die effektive Basis benutzt werden soll, ist ESS zu berechnen.

³ Das kann im Einzelnen zu unterschiedlichen Abschätzungen führen: Wenn z.B. alle Gewichte in einer Teilgruppe < 1.0 sind, dann wird dabei ein ESS herauskommen, das größer ist als die Summe der Gewichte dieser Teilgruppe, da die Summe der Quadrate der Gewichte dann kleiner ist als die Summe der Gewichte, denn für alle Werte von x zwischen 0 und 1 ist das Quadrat von x kleiner als x selbst. Bei der Ausweisung von ESS für Teilgruppen ist also ein Zahlenwert größer als ABSOLUTE möglich.

Für diese häufigkeitsbasierten Signifikanztests werden also die Summen der Gewichte der zu betrachtenden Tabellenspalten proportional auf ihre effektive Basis ESS umgerechnet. Für den unabhängigen Test (Prozentwerte in verschiedenen Teilstichproben) geht das so. Ein Beispiel:

| | Sample 1 | Sample 2 |
|----------------------------|----------|----------|
| Testzeile | 27% | 35% |
| Summe der Gewichte | 600 | 400 |
| Summe der Gewichtsquadrate | 700 | 450 |

In einer Studie mit einer 1000-er Stichprobe in zwei Teilgruppen mit einem n von 600 bzw. 400 nach Gewichtung seien die Prozentwerte 27% und 35% gegeneinander zu testen. Die Summen der Gewichtsquadrate seien 700 respektive 450. Die Berechnung geht jetzt folgende Schritte:

| | Sample 1 | Sample 2 |
|-------------------------------------|----------|----------|
| Testzeile | 27% | 35% |
| n : Summe der Gewichte in der Zelle | 162 | 140 |
| n*ESS-Faktor | 138,9 | 124,4 |
| Summe der Gewichte | 600 | 400 |
| Summe der Gewichtsquadrate | 700 | 450 |
| ESS | 514,29 | 355,56 |

Zunächst wird (siehe in der letzten Zeile) die effektive Stichprobengröße für die Teilpopulationen (Spalten) berechnet. In der linken Spalte ist das z.B. $600 * 600 / 700 = 514,29$. Daraus wiederum lässt sich die Summe der Gewichte für die Zellen in der Testzeile berechnen: 27% von 514,29 sind 138,9.

| | Sample 1 | Sample 2 |
|-----------|----------|----------|
| Testzeile | 138,9 | 124,4 |
| Restzeile | 375,4 | 231,1 |
| Basis | 514,29 | 355,56 |

Hieraus lässt sich nach der Berechnung auch der rechten Spalte die komplette 4-Felder-Tabelle für den Signifikanztest konstruieren. Letztendlich werden 27% auf der Basis von 514,29 Befragten gegen 35% auf der Basis von 355,56 Befragten getestet. Das ergibt einen χ^2 von 6,37, das ist knapp unter der Grenze von 6.64 für 1%, damit ist die Signifikanz auf dem 5%-Niveau gegeben. Der Wert von 7,28 für die dieselben Prozentwerte auf der Basis der Summe der Gewichte hätte eine Signifikanz auf dem 1% Niveau ergeben.

Rechenbeispiel für McNemar und ESS

Beim abhängigen Signifikanztest für Anteilswerte, in GESStabs als Test nach McNemar realisiert, liegen die Verhältnisse ähnlich. In den Test nach McNemar geht allerdings nur die Summe der Gewichte der Befragten ein, bei denen sich die Messwerte in beiden Variablen unterscheiden. Diese Gewichtssumme wird dann ebenfalls in eine ESS umgerechnet und diese dann proportional auf die relevanten Zellen für den Test nach McNemar verteilt.

In einem Beispiel sieht das so aus:

| | Variable 1 | Variable 2 |
|----------------------------|------------|------------|
| Testzeile | 27% | 35% |
| Summe der Gewichte | 1000 | 1000 |
| Summe der Gewichtsquadrate | 1150 | 1150 |

Anhand von 1000 Befragten werden zwei Variablen, Variable 1 und Variable 2 erhoben. Bei der Variable 1 gibt es eine Zustimmung von 27%, bei der anderen von 35%. Wenn wir nun die Signifikanz des Unterschieds testen wollen, ist ein abhängiger Test angemessener. Für einen abhängigen Test müssen die beiden Variablen auf der Ebene der Befragten miteinander verglichen werden. Wir schauen uns zu diesem Zweck eine 4-Felder-Kreuztabelle⁴ der beiden Variablen an.

| | | Variable 1 | | |
|------------|------|------------|-----|------|
| | | nein | ja | |
| Variable 2 | nein | 530 | 120 | 650 |
| | ja | 200 | 150 | 350 |
| | | 730 | 270 | 1000 |

⁴ Innerhalb von GESStabs ist an dieser Stelle ein erheblicher Aufwand nötig. Zwei unterschiedliche Variablen paarweise auf der Basis der Fälle kann man nur über TABLE ADD tabellieren. Es muss also eine beliebige Menge von TABLE und TABLE ADD daraufhin untersucht werden, welche Zellen von einem Fall betroffen sind. Hieraus werden dann alle möglichen Paare gebildet.

Die Prozentwerte 27% und 35% finden wir als absolute Häufigkeiten in den Randverteilungen wieder. Die horizontal tabellierte Variable 1 weist in der Randverteilung 270 Befragte aus, die zustimmen, die vertikal tabellierte Variable 2 350. Wir sehen weiterhin, dass 150 der Befragten beiden Variablen zugestimmt haben. Diese Information konnten wir den einfachen Prozentzahlen noch nicht entnehmen. Die übrigen Zellenbesetzungen ergeben sich aus den Differenzen. Die Tabelle hat nur einen Freiheitsgrad.

Der Test nach McNemar interessiert sich nur für die Zellen, bei denen ein Wechsel auftritt. Die einzelnen Felder der Matrix werden üblicherweise mit a bis d bezeichnet, es geht also um die Besetzung der Felder b und c:

| | | |
|------------|------------|----|
| | Variable 1 | |
| Variable 2 | nein | ja |
| nein | a | b |
| ja | c | d |

Die Testgröße für den Test nach McNemar wird nach folgender Formel berechnet:

$$\chi^2 = \frac{(|b-c|-0,5)^2}{b+c}$$

Die Summe der entsprechenden Gewichte ist $120 + 200 = 320$. Um ESS für diese Fälle zu bestimmen, benötigen wir noch die Summe der Quadrate der Gewichte in diesen Zellen. GESStabs hat diese Größe beim Zählen auch bestimmt, sie sei 432,7. Damit sind alle Größen bekannt, ESS zu berechnen. $ESS = G * \frac{G}{SQ} = 320 * 320 / 423,7 = 236,7$. Hieraus ergibt sich ein Faktor von 0,740, um den die 4-Felder-Tafel verkleinert wird. Die resultierende Tafel hat folgenden Inhalt:

| | | | |
|------------|------------|-------|-------|
| | Variable 1 | | |
| Variable 2 | nein | ja | |
| nein | 392,0 | 88,7 | 480,7 |
| ja | 147,9 | 110,9 | 258,8 |
| | 539,9 | 199,7 | 739,5 |

Als numerischer Wert ergibt sich eine χ^2 -Testgröße von 14,5. Der Unterschied (siehe χ^2 -Verteilung mit einem Freiheitsgrad) ist auf dem 0,1%-Niveau signifikant. Trotzdem ist dieser Testwert kleiner, als wenn man die gewichteten Daten ohne die ESS-Korrektur verwendet hätte, in diesem Fall hätte sich 19,8 ergeben.

Anmerkungen zu den Rechenverfahren in hybriden Signifikanztests bei GESStabs

Unabhängiger hybrider t-Test für zwei Stichproben (HYTTEST)

In die Berechnung des t-Wertes gehen die Stichprobengrößen mehrfach ein. In die Berechnung des zu untersuchenden Mittelwertunterschieds, bei der Schätzung der gemeinsamen Varianz bzw. Standardabweichung aus den Varianzen der Einzelstichproben, und in die Berechnung des t-Wertes selbst.

Beim hybriden t-Test werden die zu testenden Mittelwerte und die Varianzen der Einzelstichproben (V_1 und V_2) auf der Basis der gewichteten Daten ermittelt. Hieraus ergibt sich auch das zu bewertende Δ der beiden Mittelwerte. Die Einzelvarianzen beider Stichproben werden gewichtet ermittelt.

In die dann folgenden Berechnungen zur Schätzung der gemeinsamen der Berechnung der Varianz, des t-Werts und der Freiheitsgrade gehen hingegen die ungewichteten Stichprobengrößen (u_1 , u_2) ein:

$$V = \frac{V_1(u_1-1)+V_2(u_2-1)}{u_1+u_2-2}$$

$$SD = \sqrt{V}$$

$$t = \frac{\Delta}{SD * \sqrt{\left(\frac{1}{u_1} + \frac{1}{u_2}\right)}}$$

$$df = u_1 + u_2 - 2$$

Abhängiger hybrider t-Test (HYCOLDEPTTEST)

Analog hierzu wird auch beim abhängigen t-Test verfahren. Der Mittelwert und die Varianz der Abweichung ist das Ergebnis einer gewichteten Betrachtungsweise. In die Berechnung des t-Tests und der Anzahl der Freiheitsgrade geht die ungewichtete Fallzahl ein.

Unabhängiger hybrider Test auf Anteilsunterschieden (HYCOLCHIQU)

Basis des hybriden χ^2 Tests sind die gewichteten gegeneinander zu testenden prozentualen Anteilswerte (p_1 , p_2). Für jeden Anteilswert ist die ungewichtete Basis bekannt. Aus diesen Größen lässt sich eine fiktive ungewichtete Zellenbesetzung errechnen, die diesem Prozentwert entspräche. Diese Tabelle lässt sich zur vollständigen 4-Felder-Tafel erweitern, die dann Basis des χ^2 Tests wird. Das Vorgehen hat also sehr große Ähnlichkeit zum Verfahren bei ESSCOLCHIQU, nur dass anstelle der ESS-Werte zur Bestimmung der Basenzeile die ungewichteten Fallzahlen herangezogen werden.

Abhängiger hybrider Test auf Anteilsunterschiede (HYMCNEMAR)

In der Veränderungsmatrix geht es (s.o.) um die Felder b und c. Diese werden zunächst gewichtet und ungewichtet ermittelt. Für die Schätzung der relativen Anteile der Fälle, bei denen eine Veränderung zu messen ist (eben b und c), ist die gewichtete Zählung maßgebend, da die Gewichtung ja idealerweise eine Verbesserung der Strukturanpassung bewirkt. Die absolute Größe der Zellenbesetzungen, die ja entscheidend in den McNemar Test eingeht, soll aber anhand der ungewichteten Zahl der Befragten errechnet werden. Seien bw und cw die gewichteten Besetzungen von b und c, und bu und cu die ungewichteten, dann ergibt sich die angepasste Besetzung b zu

$$b = bw * \frac{bu + cu}{bw + cw} \quad (\text{Berechnung von c analog})$$

Aus b und c ergibt sich die Testgröße (χ^2 , $df = 1$) nach dem Standard-Berechnungsverfahren (s.o.)

Weitere Formen von Signifikanztests in GESStabs

Die meisten Signifikanztests werden verwendet, um die Verallgemeinerbarkeit von beobachteten Unterschieden zu prüfen. In der Branche hat es sich auch eingebürgert, dass die zu testenden Größen in denselben Zeilen benachbarter Spalten dargestellt werden. Hier sind nun die GESStabs CELLELEMENTS zum Ausweisen Signifikanztests in Tabellen dieser gebräuchlichsten Form. Die Darstellung orientiert sich zuerst daran, was getestet werden soll: Mittelwerte oder Anteilswerte. Eine weitere Unterscheidung ist, ob die Unterschiede an paarweise erhobenen Daten geprüft werden sollen (z.B. Zustimmung vor und nach einer Präsentation), oder an unabhängig erhobenen Daten, (z.B. Zustimmung bei Teilgruppen wie Männer und Frauen). Es sind dann jeweils unterschiedliche statistische Testverfahren angemessen, die bei GESStabs als CELLELEMENTS aufgerufen werden.

Zu erwähnen ist noch das CELLELEMENT COLPERCT, das in der ESS-Spalte der Tabelle steht. Hierbei handelt es sich um einen Test auf Unterschiede von Prozentwerten mittels eines t-Tests, bei dem der Effekt von Überlappungen zwischen den Spalten kontrolliert wird. Im Rahmen dieses Tests wird die effektive Basis ebenfalls nach der Formel $ESS = G * \frac{G}{SQ}$ berechnet.

GESStabs bietet daneben auch Signifikanztests, die Unterschiede zwischen den Zeilen eine Tabelle abbilden. Das beruht oft auf Tabellendarstellungen, die gegenüber dem Üblichen um 90° gedreht sind. Darüber hinaus gibt es Signifikanztests, in denen die H1-Hypothese nicht einen Unterschied, sondern Gleichheit behauptet (siehe z.B. COLPERCEQUAL). Noch etwas spezieller sind die Signifikanztests, die sogenannten „Benchmark“ Tabellen unterliegen; bei diesen wird auf Unterschiede getestet, aber nicht zwischen den Spalten, sondern gegenüber von außen eingeführten in die Tabellierung eingeführten Benchmark-Werten.