

1. Motivation des Chi-Quadrat-Tests

Zur Herleitung, in welchen Ausgangssituationen der Chi-Quadrat-Test hilfreich ist und welche Fragestellungen mit dem Chi-Quadrat-Test beantwortet werden können, werden folgende Beispiele betrachtet:

i **Beispiel 1:** Auf dem Schulsommerfest treten die Klassen in verschiedenen Disziplinen gegeneinander an. In einer Disziplin muss ein Ball aus einer festgelegten Entfernung in einen Karton geworfen werden. Ziel ist es dabei möglichst viele Treffer zu erzielen. Besonders hierbei ist, dass alle Lernenden einer Klasse mit derselben Hand werfen müssen. Die Klasse muss gemeinsam entscheiden, ob alle mit der rechten oder mit der linken Hand werfen. Um dies entscheiden zu können, startet eine Klasse folgendes Experiment: Zunächst wirft jeder Lernende einen Ball mit der rechten Hand zwanzig Mal hintereinander und versucht dabei in einen Karton zu treffen. Anschließend wird aus gleicher Entfernung von jedem der Vorgang mit der linken Hand wiederholt. Die Treffer werden jeweils gezählt und die Zahlen notiert.

i **Beispiel 2:** Eine weitere Disziplin ist der Origami-Frosch-Weitsprung. Hierbei lässt man Origami-Frösche von der Startlinie aus einmal springen und versucht sie über eine Linie springen zu lassen. Auch hier gibt es die Besonderheit, dass alle Lernenden einer Klasse entweder einen kleinen oder einen großen Frosch benutzen. Zur Entscheidung wird auch hier ein Experiment herangezogen: Zunächst lässt jeder Lernende einen kleinen gebastelten Origami-Frosch zwanzig Mal hintereinander springen und zählt dabei, wie oft er über die Linie springt. Anschließend wird aus gleicher Entfernung der Vorgang von jedem Lernenden mit einem großen gebastelten Origami-Frosch wiederholt und die Anzahl der Frösche, die über die Linie springen (im Folgenden als Treffer bezeichnet), gezählt.

In beiden Beispielen muss die Klasse eine Entscheidung zwischen zwei möglichen Optionen treffen. Da möglichst viele Treffer erzielt werden sollen, sucht die Klasse nach einem Weg abzuschätzen, mit welcher Option voraussichtlich mehr Treffer erzielt werden. Dazu führt die Klasse die oben beschriebenen Experimente durch. Im Rahmen von diesem Skript werden zunächst deskriptive Möglichkeiten herangezogen, um die Beobachtungen zu analysieren. Darauf aufbauend, wird der Chi-Quadrat-Test, als mögliches inferenzstatistisches Verfahren, genutzt, um die Entscheidungsfindung zu unterstützen.

1.1. Beschaffenheit der Daten

In den beiden Beispielen werden Merkmale betrachtet, die sich auf kategoriale Zuordnungen beschränken. Diese Merkmale bezeichnet man als *nominalskalierte Daten*. Diese haben die Eigenschaft, dass sie keine Rangfolgen (wie es z.B. bei Schulnoten der Fall ist; ordinalskalierte Daten) und keine Abstände (wie es z.B. bei Zeitangaben der Fall ist; intervallskalierte Daten) aufweisen. Zudem kann diesen Merkmalen kein Nullpunkt (wie es z.B. bei Geschwindigkeiten der Fall ist; verhältnisskalierte Daten) zugeordnet werden. Die in den Beispielen erhobenen Daten erfüllen daher nicht die Anforderungen zur unmittelbaren Durchführung von (statistischen) Berechnungen. Allerdings können die Häufigkeiten der Beobachtungen (bspw. die Anzahl der Treffer) bestimmt

werden, welche wiederum verhältnisskaliert sind (es gibt einen absoluten Nullpunkt; die Anzahlen sind geordnete Mengen; die Differenz zwischen n und $n + 1$ ist stets konstant, so dass Aussagen über Verhältnisse getroffen werden können).

Beispiel 1: Die Merkmale Wurfhand („Links“/ „Rechts“) sowie Treffer („Ja“/ „Nein“) sind nominalskalierte Merkmale.

Beispiel 2: Die Merkmale Froschgröße („Klein“/ „Groß“) sowie Treffer („Ja“/ „Nein“) sind nominalskalierte Merkmale.



Im Lehrplan für den Vertiefungskurs Mathematik werden am Anfang des Moduls „Statistik“ die verschiedenen Skalenniveaus (nominalskaliert, ordinalskaliert und metrisch)¹ thematisiert.

1.2. Deskriptive Beschreibung und Annäherung über ein Assoziationsmaß

Um eine Entscheidung hinsichtlich der Wahl der Wurfhand bzw. der Größe des Froschs zu treffen, sollen die Beobachtungen zunächst deskriptiv beschrieben werden. Hierfür stehen unterschiedliche Möglichkeiten zur Verfügung, wobei in diesem Skript zunächst auf eine tabellarische Darstellung der Daten eingegangen wird. Dies dient zum einen der Vorbereitung auf den Chi-Quadrat-Test, zum anderen der Anfertigung einer grafischen Darstellung. Diese Darstellungen sollen insbesondere dafür genutzt werden, um Maße zu bestimmen, mit denen sich mögliche Abhängigkeiten in den Daten beschreiben lassen, sog. *Assoziationsmaße*. Anhand dieser können Überlegungen zur Entscheidungsfindung angestellt werden.

1.2.1. Tabellarische Darstellung der beobachteten Häufigkeiten

Die beobachteten Häufigkeiten der nominalskalierten Merkmale können u. a. in sogenannten Kontingenztafeln dargestellt werden. Kontingenztafeln sind Tabellen, in denen die Merkmale, die von Interesse sind, gegeneinander aufgetragen werden. Dabei werden einerseits die Häufigkeiten der Beobachtungen der einzelnen Merkmalsausprägungen als Randsummen dargestellt. Andererseits werden die Häufigkeiten, in welchen die jeweilige Kombination aus zwei Merkmalsausprägungen auftritt, in den inneren Feldern der Kontingenztafel dargestellt. Zuletzt kann auch die Gesamtzahl aller Beobachtungen aus der Kontingenztafel abgelesen werden.

Allgemein sieht eine 2×2 – Kontingenztafel für beobachtete Werte zweier Merkmale M_1 und M_2 mit je zwei Ausprägungen (M_{1_1} und M_{1_2} bzw. M_{2_1} und M_{2_2}) folgendermaßen aus:

	M_{1_1}	M_{1_2}	Σ
M_{2_1}	$w_b = H(M_{1_1} \cap M_{2_1})$	$x_b = H(M_{1_2} \cap M_{2_1})$	$H_b(M_{2_1})$
M_{2_2}	$y_b = H(M_{1_1} \cap M_{2_2})$	$z_b = H(M_{1_2} \cap M_{2_2})$	$H_b(M_{2_2})$
Σ	$H_b(M_{1_1})$	$H_b(M_{1_2})$	n

¹ <https://www.lehrplanplus.bayern.de/fachlehrplan/gymnasium/12/mathematik/vertieft>

Hierbei repräsentiert n die Anzahl aller Beobachtungen bzw. die Stichprobengröße. H wird stellvertretend für die absoluten Häufigkeiten genutzt. Der Index b wird genutzt, um anzuzeigen, dass es sich um beobachtete Werte handelt.

Beispiel 1: Es liegt beispielsweise folgende Kontingenztafel vor:

	Wurfhand „Links“	Wurfhand „Rechts“	Σ
Treffer „Ja“	$w_b = 20$	$x_b = 140$	160
Treffer „Nein“	$y_b = 280$	$z_b = 160$	440
Σ	300	300	600

Beispiel 2: Es liegt beispielsweise folgende Kontingenztafel vor:

	Origami-Frosch „Klein“	Origami-Frosch „Groß“	Σ
Treffer „Ja“	$w_b = 120$	$x_b = 140$	260
Treffer „Nein“	$y_b = 180$	$z_b = 160$	340
Σ	300	300	600

Anhand der Kontingenztafeln können verschiedene Maße entwickelt werden, die als Argumente für die Entscheidung herangezogen werden können. Eine Idee ist, die absoluten Trefferzahlen in den Kontingenztafeln zu vergleichen, um eine Entscheidung zu treffen. Aus dieser Idee kann bspw. die Differenz $d_1 = w_b - x_b$ abgeleitet werden. Gilt $d_1 > 0$ so würde das in unserem Beispiel für den kleinen Frosch bzw. die linke Hand als Wurfhand sprechen. In unseren Beispielen gilt jedoch $d_1 < 0$, sodass das Maß eher für den großen Frosch bzw. die rechte Wurfhand spricht. Je größer der Betrag von d_1 ist, desto gewichtiger ist das Argument für die zugehörige Entscheidung. Darüber hinaus lassen sich weitere Maße finden, die bspw. zum Vergleich der Trefferquoten führen.

Betrachtet werden hier zunächst absolute Zahlen. Dies geben keine unmittelbare Auskunft über Anteile an Treffern und liegen teilweise nah beieinander. Daher kann noch keine Aussage darüber getroffen werden, ob die Unterschiede zwischen den Treffern groß genug sind, um eine Entscheidung begründen zu können.

1.2.2. Graphische Visualisierungsmöglichkeiten

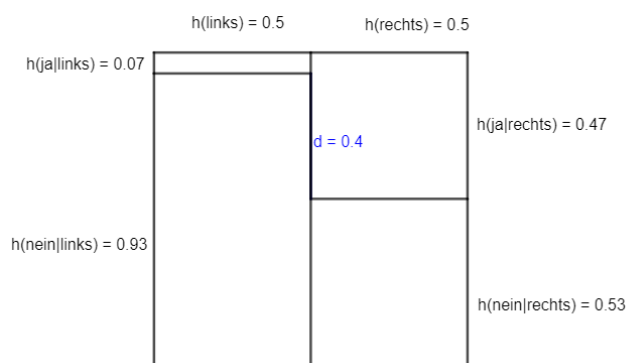
Die Untersuchung möglicher Zusammenhänge kann auf Basis einer grafischen Repräsentation der Daten erfolgen, indem sie bspw. in einem Einheitsquadrat² (s.u.) visualisiert werden. Die Flächeninhalte der Rechtecke entsprechen den relativen Anteilen der kombiniert aufgetretenen Merkmalsausprägungen $(\frac{w_b}{n}, \frac{x_b}{n}, \frac{y_b}{n}, \frac{z_b}{n})$.

Diese Darstellung erlaubt die Identifikation (weiterer) Assoziationsmaße, die auf den (bedingten) relativen Häufigkeiten basieren. Die Idee des Vergleichs der (Nicht-)Trefferquoten kann auch hier verfolgt werden. Hierzu können die waagerechten Linien innerhalb des Einheitsquadrates miteinander verglichen werden: Die Länge des vertikalen Streckenabschnittes dazwischen wird als Assoziationsmaß d bestimmt². Liegt ein großer Abstand d vor, so deutet das auf einen Zusammenhang zwischen den beiden Merkmalen hin. Ein Vorteil dieser Herangehensweise besteht darin, dass auf relativ einfache Weise eine größere Menge der verfügbaren Daten einbezogen werden.

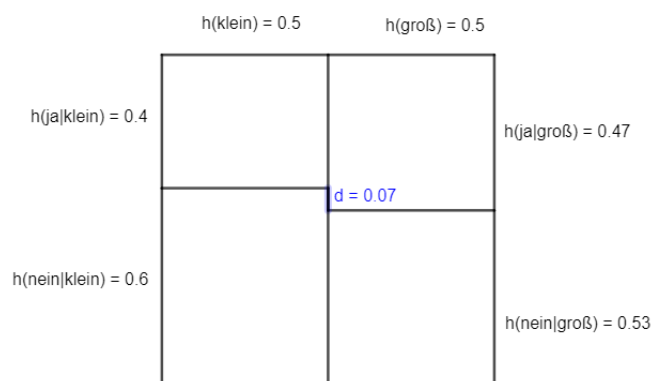
² Eichler, A., & Vogel, M. (2013). *Leitidee Daten und Zufall: Von konkreten Beispielen zur Didaktik der Stochastik*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-00118-6>

Da das Assoziationsmaß hier auf einer geometrischen Betrachtung beruht, ist dies stets positiv, so dass sich allein aus dem Maß zunächst keine Aussage hinsichtlich der Entscheidung treffen lässt. Zwar kann auf Basis eines großen d gesagt werden, dass wohl ein Zusammenhang zwischen Wurfhand bzw. Froschgröße und Treffer besteht, aber nicht unmittelbar daraus abgeleitet werden, für welche Wurfhand oder Froschgröße man sich entscheiden sollte.

Beispiel 1: Die visuelle Darstellung der deskriptiven Daten sieht folgendermaßen aus. Die senkrechte Linie liegt genau in der Mitte des Einheitsquadrats, da 50 % der Würfe mit links und 50 % mit rechts durchgeführt werden. Die beiden waagerechten Linien ergeben sich jeweils als relativer Anteil an Treffern (in Abhängigkeit der Hand), zum Beispiel: $h(ja|links) = \frac{20}{300} \approx 0,07$. Die beiden waagerechten Linien liegen relativ weit auseinander. Das Assoziationsmaß d beträgt hier $d = |h(nein|links) - h(nein|rechts)| = 0,4$.



Beispiel 2: Die visuelle Darstellung der deskriptiven Daten sieht folgendermaßen aus. Die senkrechte Linie liegt genau in der Mitte des Einheitsquadrats, da 50 % der Sprünge mit einem kleinen Frosch und 50 % mit einem großen durchgeführt werden. Die beiden waagerechten Linien ergeben sich jeweils als relativer Anteil an Treffern (in Abhängigkeit der Größe), zum Beispiel: $h(ja|klein) = \frac{120}{300} = 0,4$. Die beiden waagerechten Linien liegen nah zusammen. Das Assoziationsmaß d beträgt $d = |h(nein|klein) - h(nein|groß)| = 0,07$.



1.2.3. Assoziationsmaße – und die Entscheidung?

Mit den hier dargestellten Assoziationsmaßen können wir teilweise etwas dazu sagen, wie stark sich die Häufigkeiten unterscheiden. Je nach Maß kann zudem auch etwas darüber gesagt werden, welche Kombination an Merkmalsausprägungen stärker ausgeprägt ist. Mit Hilfe der Berechnungen und grafischen Aufbereitungen der Daten zu den beiden Beispielen, können somit

Vermutungen zu vorliegenden Zusammenhängen aufgestellt werden: Es scheint ein Zusammenhang zwischen dem Merkmal Wurfhand und dem Treffen zu bestehen.

Bei den Origami-Fröschen existiert zwar auch ein Unterschied in den Trefferquoten, aber reicht dieser wirklich aus, um von einem Zusammenhang zu sprechen? Immerhin könnten die Abweichungen auch dem Zufall geschuldet sein (bspw. auf Grund von Windstößen durch geöffnete Fenster / Türen).

Generell stellt sich die Frage, wann tatsächlich von einem Zusammenhang zwischen den Merkmalen gesprochen werden kann und wann dieser vernachlässigbar erscheint. Reicht bspw. $d = 0,4$ aus, um von einem Zusammenhang zwischen Wurfhand und Trefferquote zu sprechen? Um mehr Sicherheit hinsichtlich der Entscheidung zu gewinnen, wird daher der Chi-Quadrat-Unabhängigkeitstest als inferenzstatistisches Verfahren hinzugezogen. Mit diesem kann geprüft werden, inwieweit die Beobachtungen konform mit dem sind, was wir bei (stochastischer) Unabhängigkeit erwarten würden.

1.3. Grundidee des Chi-Quadrat-Tests – Zusammenhangsanalyse von nominalskalierten Daten

Der Chi-Quadrat-Unabhängigkeits-Test kann herangezogen werden, um Zusammenhänge zwischen nominalskalierten Daten mittels Hypothesentest zu prüfen. Dabei nutzen wir Anzahlen bzw. Häufigkeiten von beobachteten nominal skalierten Merkmalsausprägungen, die wiederum metrisch skaliert sind und für Berechnungen genutzt werden können.

i Beispiel: Man kann einen Zusammenhang zwischen der Wurfhand bzw. der Größe des Origami-Froschs und dem Erzielen eines Treffers vermuten. Ob diese Vermutungen stimmen, kann mit dem Chi-Quadrat-Unabhängigkeits-Test überprüft werden.

Als Grundlage für den Chi-Quadrat-Test gehen wir von zwei verschiedenen Hypothesen aus:

- Alternativhypothese: Ein Zusammenhang zwischen den Merkmalen besteht, diese sind also stochastisch abhängig.
- Nullhypothese: Es besteht kein Zusammenhang zwischen den Merkmalen, diese sind also stochastisch unabhängig.



Im Lehrplan für Mathematik in der 12. Jahrgangsstufe (erhöhtes Anforderungsniveau) werden ebenfalls Nullhypothesen (im Kontext von Binomialtests) thematisiert³.



Beispiel 1:

Nullhypothese: Die beiden Merkmale Wurfhand und Treffer sind stochastisch unabhängig voneinander.

³ <https://www.lehrplanplus.bayern.de/fachlehrplan/gymnasium/12/mathematik/regulaer>



Beispiel 2:

Nullhypothese: Die beiden Merkmale Größe des Origami-Froschs und Treffer sind stochastisch unabhängig voneinander.

Mit dem Chi-Quadrat-Test wird geprüft, ob die Nullhypothese auf der Datengrundlage haltbar ist, also ob die betrachteten Merkmale, tatsächlich unabhängig voneinander sind.

Um dieser Frage nachzugehen, werden die Abweichungen der in der Datenerhebung *beobachteten Häufigkeiten* mit den Werten, die man erwarten kann, wenn die Nullhypothese gilt, sie also unabhängig sind, bestimmt. Diese nennen wir *erwartete Häufigkeiten*. Entscheidend sind die Differenzen zwischen diesen beiden Werten und die Frage, ob die Differenzen zwischen beobachteten und erwarteten Häufigkeiten von Merkmalen klein sind und durch Zufall erklärt werden können (dann geht man von Unabhängigkeit aus) oder ob diese groß und signifikant sind (dann geht man von einem Zusammenhang aus). Es handelt sich bei dem Chi-Quadrat-Unabhängigkeits-Test um einen Signifikanztest.

Bei großen Abweichungen wird die Nullhypothese auf Grundlage der Datenauswertung mit dem Chi-Quadrat-Test verworfen, es ist von einem Zusammenhang zwischen zwei Merkmalen auszugehen. Die Alternativhypothese kann angenommen werden. Kann die Nullhypothese aufgrund der Ergebnisse des Chi-Quadrat-Tests jedoch nicht verworfen werden, so kann sie trotzdem nicht angenommen werden. Es ist keine Aussage möglich.



Neben dem Chi-Quadrat-Test werden in der 12. Jahrgangsstufe weitere Signifikanztests thematisiert: Im Lehrplan für Mathematik in der 12. Jahrgangsstufe (erhöhtes Anforderungsniveau) wird der Binomialtest als einseitiger Signifikanztest thematisiert⁴. Im Rahmen des Vertiefungskurses Mathematik wird der t-Test als weitere Möglichkeit für einen Signifikanztest behandelt⁵.

⁴ <https://www.lehrplanplus.bayern.de/fachlehrplan/gymnasium/12/mathematik/regulaer>

⁵ <https://www.lehrplanplus.bayern.de/fachlehrplan/gymnasium/12/mathematik/vertieft>

2. Chi-Quadrat-Test von Hand durchgeführt

Der Chi-Quadrat-Test stellt ein Instrument in der Forschung dar, welches verwendet wird, um Zusammenhangsanalysen durchzuführen und Hypothesen zu prüfen. Der Algorithmus des Tests läuft normalerweise im Hintergrund einer Statistiksoftware ab. Um diesen Prozess transparent zu machen, wird dieser anhand der beiden Beispiele von Hand durchgeführt. Hierfür wurde der Prozess in die folgenden vier Schritte aufgeteilt:

- 1) Ermittlung und tabellarische Darstellung der *beobachteten Häufigkeiten*
- 2) Ermittlung und tabellarische Darstellung der *erwarteten Häufigkeiten*
- 3) Bestimmen und normieren der Differenzen zwischen beobachteten und erwarteten Häufigkeiten
- 4) Prüfen der Differenz auf Signifikanz

2.1. Ermittlung und tabellarische Darstellung der beobachteten Häufigkeiten

Da im Rahmen des Chi-Quadrat-Tests die beobachteten mit erwarteten Häufigkeiten verglichen werden, werden zunächst die beobachteten Häufigkeiten aus vorliegenden Daten bestimmt. Diese können beispielsweise, wie in Kapitel 1.2.1 beschrieben, in einer Kontingenztafel dargestellt werden.

2.2. Ermittlung und tabellarische Darstellung der erwarteten Häufigkeiten

Die Frage, die nun beantwortet werden soll, ist, welche Häufigkeiten wir in der Stichprobe erwarten würden, wenn die Nullhypothese gilt, also kein Zusammenhang zwischen den Variablen besteht, sondern diese stochastisch unabhängig sind. Hierbei folgen wir der Annahme, dass sich die Merkmalsausprägungen unter Berücksichtigung der relativen Anteile an der Gesamtheit gleichmäßig verteilen.

2.2.1. Ermittlung der erwarteten Häufigkeiten mit Hilfe des visuellen Assoziationsmaß

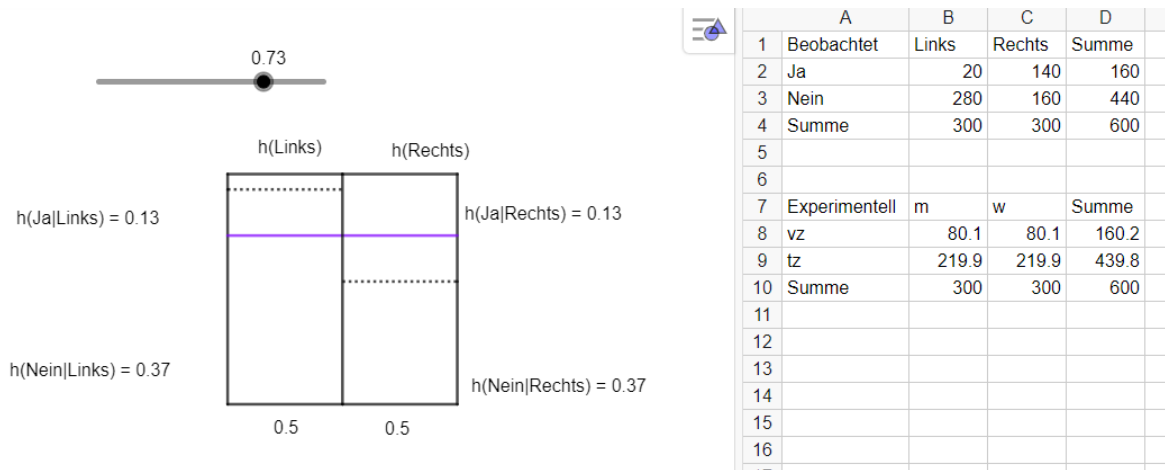
Aus den Überlegungen zur visuellen Bestimmung des Assoziationsmaßes (Abschnitt 1.2.2) kann abgeleitet werden, dass Unabhängigkeit besteht, wenn die beiden waagrechten Linien innerhalb des Einheitsquadrates eine Linie bilden. Denn dann werden anteilig gleich viele Treffer, unabhängig von der Voraussetzung (Wurfhand bzw. Größe), erzielt.

Die Bestimmung der erwarteten Häufigkeiten kann visuell durch Suchen nach ebendieser Linie erfolgen. Die Linie soll so gewählt werden, dass die Anzahl an beobachteten Treffern erhalten bleibt. Hintergrund ist, dass wir zur Bestimmung der erwarteten Häufigkeiten nicht die Ziehung unserer Stichprobe hinterfragen und verändern möchten, sondern nur die Ausprägungen der Merkmale innerhalb unserer Stichprobe.

Dies kann im Unterricht bspw. mithilfe eines [GeoGebra-Applets](#) realisiert werden (siehe Grafik). In dem Applet wird automatisch die Kontingenztafel in Abhängigkeit der Position der horizontalen

Strecke (lila in Grafik) berechnet. Diese experimentell ermittelte Kontingenztafel entspricht bei entsprechender Platzierung der Linie den erwarteten Häufigkeiten.

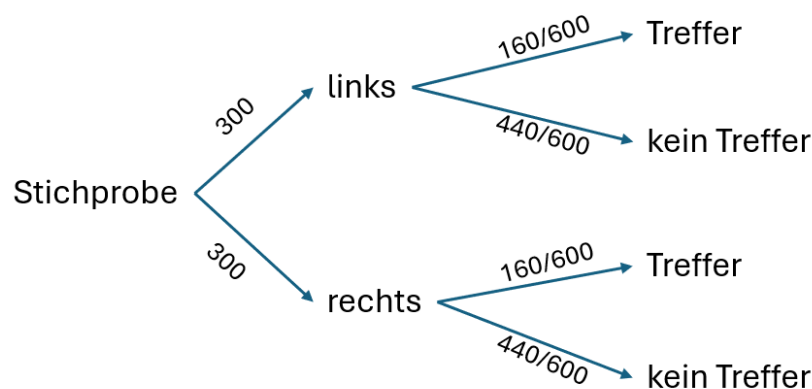
Beispiel 1: Visuelle Ermittlung der erwarteten Häufigkeiten:



Mit Hilfe der Simulation kann entdeckt werden, wie die erwarteten Häufigkeiten rechnerisch bestimmt werden. Dazu kann zunächst eine Vermutung aufgestellt werden und diese anschließend mit Hilfe der Darstellung in einem Baumdiagramm überprüft werden.

Hierbei wird zunächst hinsichtlich des ersten Merkmals unterteilt (z. B. Wurfhand). Im zweiten Schritt unterteilen wir nochmal jeweils hinsichtlich des zweiten Merkmals (z. B. Treffer). Dabei folgen wir der Annahme, dass die zweite Merkmalsausprägung unabhängig von der Ausprägung des ersten Merkmals ist.

Beispiel 1: Wenn $\frac{160}{600} = 0,267$, also 26,7 % aller Würfe Treffer sind, dann sind auch 26,7 % aller Würfe mit der linken bzw. rechten Hand Treffer: $\frac{160}{600} \cdot 300 = 80$.



Insgesamt lässt sich dann hieraus mittels der multiplikativen Pfadregel die formale Berechnung der erwarteten Häufigkeiten herleiten und begründen.

2.2.2. Rechnerische Ermittlung der erwarteten Häufigkeiten

Zur Bestimmung der erwarteten konjugierten Häufigkeiten, werden die Zelleninhalte der Tafel in Abhängigkeit der Randsummen und der Gesamtanzahl der Beobachtungen berechnet. Die jeweiligen erwarteten konjugierten Häufigkeiten ergeben sich aus dem Quotienten zwischen dem

Produkt der jeweiligen Spalten- und Zeilensumme (der beobachteten Kontingenztafel) sowie der Stichprobengröße.

Allgemein berechnet man die erwarteten Werte zweier Merkmale M1 und M2 mit je zwei Ausprägungen (M1₁ und M1₂ bzw. M2₁ und M2₂) in einer 2x2 Kontingenztafel folgendermaßen, wobei der Index e anzeigt, dass es sich um die erwarteten Werte handelt.

	M1 ₁	M1 ₂	Σ
M2 ₁	$w_e = \frac{H_b(M1_1) \cdot H_b(M2_1)}{n}$	$x_e = \frac{H_b(M1_2) \cdot H_b(M2_1)}{n}$	H _b (M2 ₁)
M2 ₂	$y_e = \frac{H_b(M1_1) \cdot H_b(M2_2)}{n}$	$z_e = \frac{H_b(M1_2) \cdot H_b(M2_2)}{n}$	H _b (M2 ₂)
Σ	H _b (M1 ₁)	H _b (M1 ₂)	n

i **Beispiel 1:** Die Berechnung der erwarteten Häufigkeiten ergibt folgende Kontingenztafel:

	Wurfhand „Links“	Wurfhand „Rechts“	Σ
Treffer „Ja“	$w_e = \frac{300 \cdot 160}{600} = 80$	$x_e = \frac{300 \cdot 160}{600} = 80$	160
Treffer „Nein“	$y_e = \frac{300 \cdot 440}{600} = 220$	$z_e = \frac{300 \cdot 440}{600} = 220$	440
Σ	300	300	600

i **Beispiel 2:** Die Berechnung der erwarteten Häufigkeiten ergibt folgende Kontingenztafel:

	Origami-Frosch „Klein“	Origami-Frosch „Groß“	Σ
Treffer „Ja“	$w_e = \frac{300 \cdot 260}{600} = 130$	$x_e = \frac{300 \cdot 260}{600} = 130$	260
Treffer „Nein“	$y_e = \frac{300 \cdot 340}{600} = 170$	$z_e = \frac{300 \cdot 340}{600} = 170$	340
Σ	300	300	600

2.3. Bestimmen und normieren der Differenzen zwischen beobachteten und erwarteten Häufigkeiten

Nachdem die beobachteten sowie die erwarteten Werte (tabellarisch) ermittelt und dargestellt wurden, gilt es diese in Beziehung zueinanderzusetzen und zu vergleichen. Hierzu wird jeweils paarweise die Differenz aus beobachtetem und erwartetem Wert gebildet. Anschließend wird diese Differenz quadriert. Im nächsten Schritt werden die Differenzen normiert, indem sie jeweils durch den zugehörigen erwarteten Wert dividiert werden. Die Summe aus den quadrierten und normierten Differenzen ergibt dann den empirischen Chi-Quadrat-Wert χ_{emp}^2 , den wir mit der Chi-Quadrat-Verteilung abgleichen, um zu bestimmen, ob der Unterschied zufällig oder signifikant ist.

Die Ermittlung des empirischen Chi-Quadrat-Wertes χ_{emp}^2 im 2x2-Fall erfolgt folgendermaßen:

	$M1_1$	$M1_2$	Σ
$M2_1$	$w = \frac{(w_e - w_b)^2}{w_e}$	$x = \frac{(x_e - x_b)^2}{x_e}$	$w + x$
$M2_2$	$y = \frac{(y_e - y_b)^2}{y_e}$	$z = \frac{(z_e - z_b)^2}{z_e}$	$y + z$
Σ	$w + y$	$x + z$	$\chi_{emp}^2 = w + x + y + z$

i **Beispiel 1:** Der empirische Chi-Quadrat-Wert ergibt sich folgendermaßen:

	Wurfhand „Links“	Wurfhand „Rechts“	Σ
Treffer „Ja“	$w = \frac{(80 - 20)^2}{80} = 45$	$x = \frac{(80 - 140)^2}{80} = 45$	90
Treffer „Nein“	$y = \frac{(220 - 280)^2}{220} = \frac{180}{11}$	$z = \frac{(220 - 160)^2}{220} = \frac{180}{11}$	$\frac{360}{11}$
Σ	$\frac{675}{11}$	$\frac{675}{11}$	$\chi_{emp}^2 \approx 122,73$

i **Beispiel 2:** Der empirische Chi-Quadrat-Wert ergibt sich folgendermaßen:

	Origami-Frosch „Klein“	Origami-Frosch „Groß“	Σ
Treffer „Ja“	$w = \frac{(130 - 120)^2}{130} = \frac{10}{13}$	$x = \frac{(130 - 140)^2}{130} = \frac{10}{13}$	$\frac{20}{13}$
Treffer „Nein“	$y = \frac{(170 - 180)^2}{170} = \frac{10}{17}$	$z = \frac{(170 - 160)^2}{170} = \frac{10}{17}$	$\frac{20}{17}$
Σ	$\frac{300}{221}$	$\frac{300}{221}$	$\chi_{emp}^2 \approx 2,71$

Der empirische Chi-Quadrat-Wert stellt ein weiteres Assoziationsmaß dar, mit dem der Unterschied zwischen beobachteten und erwarteten Werten beschrieben werden kann.

2.4. Prüfen des empirischen Chi-Quadrat-Wertes auf Signifikanz

Zur Entscheidung, ob die Nullhypothese verworfen werden kann, muss bestimmt werden, ob der durch den empirischen Chi-Quadrat-Wert χ_{emp}^2 berechnete Unterschied zwischen den beobachteten und erwarteten Merkmalen eher zufällig oder signifikant ist. Dazu wird ein Grenzwert χ_{krit}^2 festgelegt, bei dessen Überschreitung von einer Signifikanz der Unterschiede ausgegangen wird, woraus die Ablehnung der Nullhypothese erfolgt. Dieser Grenzwert wird durch die Werte der Chi-Quadrat-Verteilung festgelegt. Die Prüfung der Signifikanz erfolgt algorithmisch in drei Schritten:

1. Ermitteln und Festlegen des Parameters und des Signifikanzniveaus der Chi-Quadrat-Verteilung
2. Bestimmung des kritischen Wertes der Chi-Quadrat-Verteilung
3. Vergleich des empirischen Chi-Quadrat-Wertes mit dem Wert der Chi-Quadrat-Verteilung

2.4.1. Festlegung des Parameters und des Signifikanzniveaus der Chi-Quadrat-Verteilung

Die Chi-Quadrat-Verteilung ist eine Verteilung, deren Werte von der Anzahl der Merkmalsausprägungen abhängt. Diese werden in Form von Freiheitsgraden dF (engl. für *degrees of freedom*) berücksichtigt. Die Anzahl der Freiheitsgrade gibt dabei die Anzahl der Werte an, die frei verändert werden können, ohne dass sich der Wert der Chi-Quadrat-Verteilung ändert.

Bei zwei Merkmalen mit je zwei Merkmalsausprägungen, ergibt sich ein Freiheitsgrad von $(2 - 1) \cdot (2 - 1) = 1$, da bei der freien Wahl eines Wertes, der andere entsprechend angepasst werden muss, damit sich im Gesamten nichts verändert.

Weiterhin hängen die Werte der Chi-Quadrat-Verteilung von dem Signifikanzniveau α ab. Das Signifikanzniveau beschränkt die Wahrscheinlichkeit für einen Fehler 1. Art nach oben. Meist wird ein Signifikanzniveau von $\alpha = 5\%$ festgelegt, da der Fehler 1. Art möglichst klein sein soll.

Der Fehler 1. Art wird begangen, wenn die Nullhypothese verworfen wird, obwohl diese wahr ist. Im Fall des Chi-Quadrat-Tests ist der Fehler 1. Art, dass die Unabhängigkeit der Merkmale verworfen wird, also von einer Abhängigkeit ausgegangen wird, obwohl die Merkmale tatsächlich unabhängig sind.

2.4.2. Bestimmung des kritischen Wertes der Chi-Quadrat-Verteilung

Nachdem der Freiheitsgrad als Parameter der Chi-Quadrat-Verteilung und das Signifikanzniveau α festgelegt wurden, können die kritischen Werte der Chi-Quadrat-Verteilung bestimmt werden. Dazu gibt es drei Möglichkeiten.

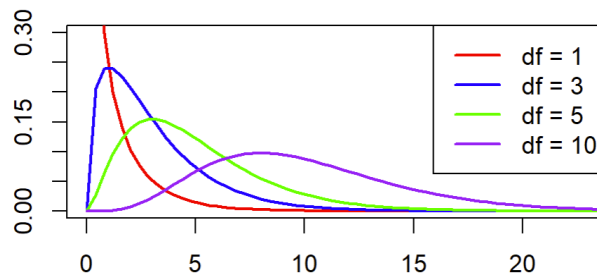
Die Werte der Chi-Quadrat-Verteilung können berechnet werden. Da zur Berechnung der Werte der Chi-Quadrat-Verteilung die Gamma-Funktion – eine Funktion, die uneigentliche Integrale beinhaltet und deren Werte nicht analytisch berechnet werden können – notwendig ist, können diese Werte nur numerisch und nicht per Hand berechnet werden. Dies stellt deshalb keine praktikable Möglichkeit für den Unterricht dar.

Als zweite Möglichkeiten können die Werte der Chi-Quadrat-Verteilung aus einem Tabellenwerk abgelesen werden. In den Tabellenwerken sind die kritischen Werte der Chi-Quadrat-Verteilung meist in Abhängigkeit des Freiheitsgrades und des Signifikanzniveaus aufgelistet, sodass der notwendige Wert abgelesen werden kann. Tabellenwerke findet man in vielen Statistik-Büchern, wie z.B. Bamberg et al., (2011)⁶. Alternativ dazu können die Werte der Chi-Quadrat-Verteilung auch mit Hilfe von Excel bestimmt werden. Dazu kann der Befehl `CHISQ.INV(1 - α ; dF)` oder alternativ `CHISQ.INV.RE(α ; dF)` genutzt werden.

Zuletzt kann der kritische Wert der Chi-Quadrat-Verteilung auch grafisch bestimmt werden. Dazu werden die Wahrscheinlichkeitsdichten der Chi-Quadrat-Verteilung in Abhängigkeit der Freiheitsgrade in einem Koordinatensystem dargestellt (siehe Grafik). Die Graphen dieser liegen komplett im 1. Quadranten. Die Fläche unter den jeweiligen Graphen der Wahrscheinlichkeitsverteilungen gibt die Wahrscheinlichkeit dafür an, dass ein Ergebnis eines Chi-Quadrat-verteilten

⁶ Bamberg, G., Baur, F. & Krapp, M. (2011). *Statistik*. München: Oldenbourg Wissenschaftsverlag.
<https://doi.org/10.1524/9783486714371>

Zufallsexperiments (zufällige Ziehung aus einer normalverteilten Stichprobe) in einem betrachteten Intervall liegt. Der Wert der Chi-Quadrat-Verteilung gibt folglich die rechte Grenze des Intervalls an, in dem mit einer Wahrscheinlichkeit von $1 - \alpha$ ein beobachtetes Ergebnis liegt. Somit kann man den kritischen Wert einer Chi-Quadrat-Verteilung χ^2_{krit} grafisch bestimmen, indem man die Stelle bestimmt, sodass die Fläche unter dem Graphen in dem Intervall $[0; \chi^2_{krit}[$ dem Anteil $1 - \alpha$ der Gesamtfläche unter dem Graphen entspricht. Die Fläche unter dem Graphen in dem Intervall $[\chi^2_{krit}; \infty[$ entspricht dem Anteil α der Gesamtfläche unter dem Graphen.



Der kritische Wert der Chi-Quadrat-Verteilung gibt die Grenze in Abhängigkeit des Signifikanzniveaus α an. Alle beobachteten Werte unterhalb der Grenze werden als vereinbar mit der Unabhängigkeit gesehen, alle oberhalb weichen zu stark von den erwarteten Werten ab.

i **Beispiel 1 und 2:** Der kritische Chi-Quadrat-Wert ist in beiden Beispielen identisch, weil jeweils nur ein Freiheitsgrad vorliegt. Er lautet für ein Signifikanzniveau von $\alpha = 5\%$:

$$\chi^2_{krit} = 3,84$$

2.4.3. Vergleich des empirischen Chi-Quadrat-Wertes mit dem Wert der Chi-Quadrat-Verteilung

Zuletzt wird der empirisch berechnete Chi-Quadrat-Wert χ^2_{emp} mit dem zugehörigen kritischen Wert der Chi-Quadrat-Verteilung χ^2_{krit} verglichen. Es können zwei Fälle auftreten:

- $\chi^2_{emp} \geq \chi^2_{krit}$: Der Unterschied zwischen beobachteten und erwarteten Häufigkeiten ist signifikant. Folglich wird die Nullhypothese, dass die Merkmale unabhängig voneinander sind, verworfen und eine Abhängigkeit der Merkmale kann angenommen werden.
- $\chi^2_{emp} < \chi^2_{krit}$: Der Unterschied zwischen beobachteten und erwarteten Häufigkeiten ist nicht signifikant. Folglich kann die Nullhypothese nicht verworfen werden. Sie kann aber auch nicht angenommen werden. Es ist insgesamt keine Aussage möglich, ob ein Zusammenhang zwischen den Merkmalen besteht oder nicht.

i **Beispiel 1:** Es gilt $122,73 = \chi^2_{emp}$ und $\chi^2_{krit} = 3,84$, woraus sich $\chi^2_{emp} > \chi^2_{krit}$ ergibt. Somit kann die Nullhypothese verworfen werden und ein Zusammenhang zwischen der Wurfbildung und dem Erzielen eines Treffers kann angenommen werden. Aus dieser Erkenntnis sollte sich die Klasse auf Grundlage des Experiments für das Werfen mit der rechten Hand entscheiden.



Beispiel 2: Es gilt $2,71 = \chi_{emp}^2$ und $\chi_{krit}^2 = 3,84$, woraus sich $\chi_{emp}^2 < \chi_{krit}^2$ ergibt. Somit kann die Nullhypothese nicht verworfen werden. Insgesamt kann hier keine Aussage über einen Zusammenhang zwischen den Merkmalen getroffen werden. Aus dieser Erkenntnis heraus kann die Entscheidung, mit welcher Froschgröße angetreten wird, nicht auf Grundlage der Daten getroffen werden.

Zusammengefasst ist der Wert der Chi-Quadrat-Verteilung χ_{krit}^2 ein kritischer Wert, der festlegt, ab wann eine Abweichung der beobachteten von erwarteten Werten nicht mehr als zufällig, sondern als signifikant angesehen wird. Anders gesagt, er kann als Grenzwert für das Assoziationsmaß in Form des empirisch ermittelten χ_{emp}^2 -Wertes gesehen werden. Dieser trägt als zusätzliches Argument zur Entscheidungsfindung bei, ob ein Zusammenhang zwischen zwei nominal skalierten Merkmalen besteht oder nicht. Denn als Folge einer Überschreitung des empirischen von dem kritischen Wert wird die Nullhypothese abgelehnt, weil die berechnete Abweichung zwischen beobachtetem Wert und bei Unabhängigkeit erwarteten Wert zu groß ist, um durch Zufall erklärt werden zu können. Sie stellt einen signifikanten Unterschied dar. Dabei nimmt man mit einer Wahrscheinlichkeit von α den Fehler 1. Art in Kauf, dass also die Annahme der Unabhängigkeit abgelehnt wird, obwohl die Merkmale unabhängig voneinander sind.

3. Chi-Quadrat-Test in der Praxis – mit Software durchgeführt

In der Praxis werden die Berechnungen beim Chi-Quadrat-Test häufig an Computerprogramme ausgelagert. In der Schule kann neben kostenfreier professioneller Statistiksoftware (bspw. JASP oder Jamovi) auch auf handelsübliche CAS-/Grafikfähige-Taschenrechner oder gängige Mathematiksoftware (GeoGebra, Excel) zurückgegriffen werden. Dabei ist zu beachten, dass die Daten je nach Tool in entsprechender Art zur Verarbeitung eingegeben werden müssen. Während typische Statistiksoftware im Allgemeinen die Rohdaten verarbeitet (eine Zeile entspricht einer Beobachtung, spaltenweise werden die Merkmalsausprägung der Beobachtung festgehalten) ist i.d.R. bei Taschenrechnern die Eingabe der Kontingenztafel in Form einer Matrix erforderlich. Bei den Ergebnissen ist zu beachten, dass teilweise automatisiert eine „Kontinuitätskorrektur“ (siehe z.B. Bühner & Ziegler, 2017⁷, S.334) vorgenommen wird, so dass die Chi-Quadrat-Werte von der händischen Berechnung leicht abweichen können.

⁷ Bühner, M., & Ziegler, M. (2017). *Statistik für Psychologen und Sozialwissenschaftler: Grundlagen und Umsetzung mit SPSS und R*. Pearson Education Deutschland GmbH.



Beispiel: Durchführung des Chi-Quadrat-Tests mit Jamovi auf Basis von Rohdaten

	A	B
1	grün	grün
2	gelb	grün
3	gelb	gelb
4	gelb	grün
5	grün	gelb
6	grün	gelb
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		

Ergebnisse

Kreuztabellen

Kreuztabellen

B	A		Insgesamt
	grün	gelb	
grün	1	2	3
gelb	2	1	3
Insgesamt	3	3	6

χ^2 -Tests

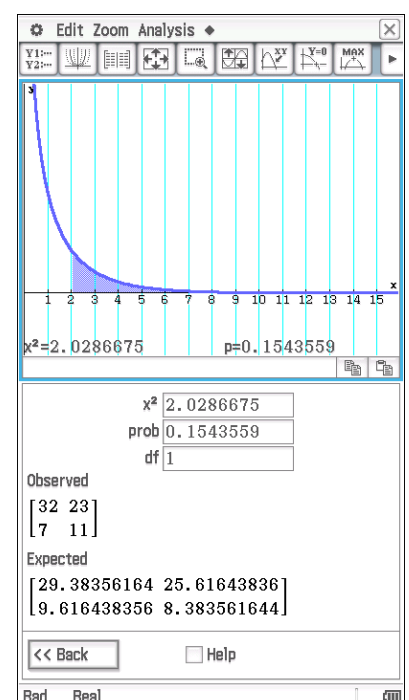
	Wert	df	p
χ^2	0.667	1	0.414
N	6		

Zusätzlich wird i.d.R. ein p-Wert angegeben, der der Fläche entspricht, die in dem Intervall $[\chi^2_{emp}; \infty[$ unterhalb der Chi-Quadrat-Dichtefunktion liegt. Wie bei anderen Signifikanztests auch, kann dieser Wert als Maß für die Vereinbarkeit der beobachteten mit den unter Annahme der Nullhypothese erwarteten Werten gesehen werden kann. Je kleiner der p-Wert, desto unwahrscheinlicher ist es, dass Abweichungen in den Beobachtungen auf Zufall zurückzuführen sind. Im Allgemeinen wird bei einem p-Wert der kleiner als das Signifikanzniveau α ist, die Nullhypothese abgelehnt und von einem signifikanten Unterschied gesprochen. Der p-Wert bietet den Vorteil, dass er unabhängig von Freiheitsgraden und der Dichtefunktion und analog zu anderen Signifikanztests interpretiert werden kann.

Hinweis: Der p-Wert kann u.a. mithilfe von Excel bestimmt werden, indem der Befehl CHIU.TEST(beobachtete Werte; erwartete Werte) verwendet wird.



Beispiel: Ergebnisse der Durchführung des Chi-Quadrat-Tests inkl. grafischer Aufbereitung mit dem Casio ClassPad II. Die markierte Fläche entspricht dem p-Wert.



4. Weiterführende Überlegungen zum Chi-Quadrat-Test

Bisher wurde der Chi-Quadrat-Test nur für zwei Merkmale mit je zwei Ausprägungen betrachtet, woraus sich beispielsweise der Freiheitsgrad 1 bei der Chi-Quadrat-Verteilung ergeben hat. Jedoch kann der Chi-Quadrat-Test auch auf mehr als zwei Merkmale mit mehr als zwei Merkmalsausprägungen angewendet werden. Dann erhöht sich die Anzahl der Freiheitsgrade der Chi-Quadrat-Verteilung. Die Anzahl der Freiheitsgrade kann im Allgemeinen durch $\prod_{i=1}^k (m_i - 1)$ berechnet werden, wobei k die Anzahl der verschiedenen Merkmale ist und m_i die Anzahl der jeweils möglichen Merkmalsausprägungen darstellt.

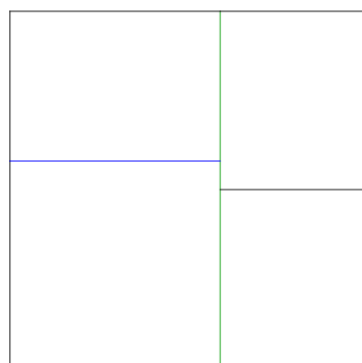
Es gibt also mehrere Freiheitsgrade (und somit auch weitere Dichtefunktionen) und mit Computerprogrammen kann der Chi-Quadrat-Test auch mit mehr als zwei Variablen, die auch mehr als zwei Ausprägungen aufweisen, durchgeführt werden. In Abhängigkeit der Merkmalsausprägungen sowie der Anzahl der Variablen ist der Freiheitsgrad mit der oben genannten Formel zu ermitteln und die zugehörige Dichtefunktion zu wählen. In Abhängigkeit von letzterem ist auch der kritische Chi-Quadrat-Wert zu ermitteln. Die Bestimmung der erwarteten Häufigkeiten erfolgt analog zu dem oben beschriebenen 2x2-Fall.

Wie bei jedem Testverfahren ist anzumerken, dass gewisse Voraussetzungen erfüllt werden müssen. So auch beim Chi-Quadrat-Test, für welchen die Beobachtungen unabhängig voneinander sein sollten (sind bspw. Beobachtungen zur Ernährung bei Befragungen beim Metzger unabhängig voneinander?) und jede Zelle in der Kontingenztafel mindestens den Wert 5 enthalten sollte⁸.

Zudem sind auch Grenzen des Test-Verfahrens zu diskutieren. Zwar kann mit ihm eine Aussage über den Zusammenhang getroffen werden, aber nicht unmittelbar auf die Richtung geschlossen werden. Auch die Frage nach der Bedeutsamkeit des Zusammenhangs bleibt offen. Wird bspw. systematisch die Stichprobengröße erhöht, so kann festgestellt werden, dass auch der empirische Chi-Quadrat-Wert steigt. Daher können auch kleine Unterschiede zur Signifikanz des Tests führen, wobei die Bedeutsamkeit in Frage gestellt werden kann. Es lohnt daher stets ein geeignetes deskriptives Assoziationsmaß heranzuziehen, um mehr über das gemeinsame Auftreten von Merkmalen sowie der potenziellen Bedeutsamkeit der Ergebnisse aussagen zu können.



Beispiel: Die Abhängigkeit des Chi-Quadrat-Wertes von der Stichprobengröße und unterschiedlicher bedingter Häufigkeiten kann bspw. mit der folgenden [Simulation](#) untersucht werden.



$$\chi^2 = 4$$

$$n = 640$$

	A	B	C	D
1	Beobachtet	M1	M2	Summe
2	N1	158.84	131.37	290.22
3	N2	218.76	131.03	349.78
4	Summe	377.6	262.4	640
5				
6	Erwartet	M1	M2	Summe
7	N1	171.23	118.99	290.22
8	N2	206.37	143.41	349.78
9	Summe	377.6	262.4	640
10				
11	Diff. std.	M1	M2	Summe
12	N1	0.9	1.29	2.19
13	N2	0.74	1.07	1.81
14	Summe	1.64	2.36	4
15				
16				

⁸ Bortz, J., & Schuster, C. (2011). *Statistik für Human- und Sozialwissenschaftler: Limitierte Sonderausgabe* (7. Aufl. 2010). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-12770-0>