

# **Chi-Quadrat**



# Inhaltsverzeichnis

Chi-Quadrat .....	2
Einführung .....	2
Lernhinweise .....	4
Theorie (1 - 8) .....	4
1. Erstellung einer Kreuztabelle .....	4
2. Beobachtete und erwartete Häufigkeiten .....	7
3. Berechnung von Chi-Quadrat .....	8
Vorgehen - Berechnung von Chi-Quadrat .....	
Übung 1 - Berechnung von Chi-Quadrat .....	
Übung 2 - Berechnung von Chi-Quadrat .....	
4. Verteilung von Chi-Quadrat .....	18
5. Interpretation .....	20
6. Freiheitsgrade, Signifikanz, kritischer Wert .....	20
7. Anzahl Fälle – Korrektur nach Yates .....	22
8. Zusammenfassung zum Lernschritt .....	22
SPSS-Kochbuch .....	22
Glossar .....	23

# Chi-Quadrat

## Einführung

Chi-Quadrat - nach dem griechischen Buchstaben Chi ( $\chi$ ) - ist eine vielfach verwendbare Test-Statistik, die vom amerikanischen Statistiker Karl Pearson entwickelt worden ist (deshalb auch "Pearson's Chi-Quadrat") und prüft, ob beobachtete Häufigkeiten ( $f_o$  für frequency observed) sich massgeblich ("signifikant") von solchen Häufigkeiten unterscheiden, die man aufgrund bestimmter Annahmen erwartet ( $f_e$  für frequency expected ). In den Sozialwissenschaften wird der Chi-Quadrat-Test oft dazu benutzt, Zusammenhänge zwischen nominalskalierten Variablen zu untersuchen.

**Beispiel:** Besteht bei Wählern ein Zusammenhang zwischen der Konfession und gewählter Partei?

Obschon Chi-Quadrat auch auf Variablen mit höherem Messniveau angewandt werden könnte, gibt es dafür angemessenere und bessere statistische Tests - eine Übersicht dazu finden Sie

hier.

**Statistische Kennzahlen für bivariate Zusammenhänge**

Messniveau	Dichotomie	Nominal	Ordinal	Intervall und Ratio
<b>Dichotomie</b>	Wie nominal x nominal, Yule's Q, Cohen's Kappa, Fisher's exact test, Risk estimate			
<b>Nominal</b>	Wie nominal x nominal	Chi-quadrat (# <sup>2</sup> ), Phi (#) und # <sup>2</sup> , Cramer's V, C (Kontingenzkoeff.), Lambda (#), Goodman und Kruskal's Tau (#), Uncertainty Coefficient, Cohen's Kappa (#), McNemar-Test		
<b>Ordinal</b>	Tests: Mann-Whitney, runs, Smirnov, signed-ranks.	Wie nominal x nominal, Varianzanalyse mit Rängen	Spearman's r, Kendall's Tau-b und Tau-c (#b, #c), Gamma (#), Somers' d	
<b>Intervall und Ratio</b>	Eta, Difference of means Test	Eta, Varianzanalyse, Interklassen-Korrelation		Korrelation und Regression

# Lernhinweise

### Lernziel

Sie lernen anhand von Übungen und Beispielen, wie der Chi-Quadrat-Test praktisch angewandt wird. Dazu sind folgende Schritte notwendig:

- Erstellung einer Kreuztabelle;
- Berechnung von Chi-Quadrat;
- Bestimmung des sog. kritischen Wertes von  $\chi^2$  auf einem gewählten Signifikanzniveau bei entsprechendem Freiheitsgrad;
- Interpretation der Resultate: Wird die zu prüfende Hypothese gestützt?

### Benötigte Vorkenntnisse

Für diesen Lernschritt sollten Sie wissen, was eine [Zufallsauswahl](#) ist, wie [Stichproben](#) zu Stande kommen, welche Bedeutung [Messniveaus](#) haben, was mit einer Sampling distribution (d. h. einer sog. Stichprobenkennwertverteilung) gemeint ist und wie und weshalb die Falsifikation von [Null-Hypothesen](#) in den Sozialwissenschaften eine besondere Bedeutung besitzt.

## Theorie (1 - 8)

### Inhaltsübersicht

- [1. Erstellung einer Kreuztabelle](#)
- [2. Beobachtete und erwartete Häufigkeiten](#)
- [3. Berechnung von Chi-Quadrat](#)
- [4. Verteilung von Chi-Quadrat](#)
- [5. Interpretation](#)
- [6. Freiheitsgrade, Signifikanz, kritischer Wert](#)
- [7. Anzahl Fälle – Korrektur nach Yates](#)
- [8. Zusammenfassung zum Lernschritt](#)

## 1. Erstellung einer Kreuztabelle

### Besteht ein Zusammenhang zwischen Konfession und politischer Präferenz?

In sozialwissenschaftlichen Untersuchungen stellt sich häufig die Frage nach dem Zusammenhang zwischen nominalskalierten Variablen. Im Rahmen einer Studie über das Wahlverhalten könnte man z. B. der Frage nachgehen, ob katholische WählerInnen einer christlichen Partei näher stehen als andere. Fragestellung: Gibt es einen Zusammenhang zwischen Konfession und politischer Präferenz? Hier kann der Chi-Quadrat-Test eingesetzt werden.

Brauchbare Daten für eine entsprechende Untersuchung liefert z. B. die UNIVOX-Befragung von 1997. Es handelt sich dabei um eine repräsentative Umfrage, die in regelmässigen Abständen vom [GfS-Forschungsinsitut](#) durchgeführt und zusammen mit den politikwissenschaftlichen Instituten der Universitäten Bern, Genf und Zürich ausgewertet wird.

### Operationalisierung

- "Konfession" wird als Antwort auf die Frage nach der Zugehörigkeit zu einer Glaubensgemeinschaft operationalisiert. Die entsprechenden Resultate wurden transformiert, um eine Variable mit 2 Ausprägungen "katholisch"/"nicht katholisch" zu erhalten.
- "Politische Präferenz" wird als Antwort auf die Frage nach der zuletzt gewählten Partei operationalisiert. Die entsprechenden Resultate wurden in einer weiteren Variablen mit 2 Ausprägungen "CVP-Wähler"/"andere" zusammengefasst.

50 Fälle wurden für dieses Beispiel willkürlich herausgegriffen:

ID	Partei	Konfession	ID	Partei	Konfession	ID	Partei	Konfession
1	CVP	k	18	andere	k	35	CVP	k
2	CVP	nk	19	andere	nk	36	CVP	nk
3	andere	k	20	CVP	k	37	andere	k
4	CVP	k	21	andere	k	38	andere	k
5	andere	nk	22	andere	nk	39	andere	nk
6	andere	k	23	CVP	k	40	andere	nk
7	andere	nk	24	CVP	nk	41	CVP	k
8	CVP	k	25	andere	k	42	andere	k
9	andere	nk	26	andere	nk	43	andere	nk
10	andere	k	27	CVP	k	44	CVP	nk
11	andere	nk	28	CVP	k	45	CVP	nk
12	CVP	k	29	andere	k	46	andere	k
13	CVP	k	30	andere	nk	47	andere	nk
14	CVP	k	31	CVP	nk	48	CVP	k
15	andere	k	32	andere	k	49	CVP	k
16	andere	nk	33	andere	nk	50	andere	nk
17	CVP	k	34	CVP	k			

Zählen Sie die Häufigkeiten aus und tragen Sie diese in die folgende Kreuztabelle ein:

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [\[link\]](#)**

	katholisch	nicht katholisch	Summe
<b>CVP-Wähler</b>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>
<b>andere</b>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>
<b>Summe</b>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>



Bisherige Forschungen zur sog. konfessionellen *Cleavage*<sup>1</sup> lassen vermuten, dass ein Zusammenhang zwischen Konfession und politischer Präferenz besteht. Im konkreten Fall lautet also die

- **Hypothese:** Es gibt einen positiven Zusammenhang zwischen der Konfession und der Wahlpräferenz.
- Forschungstechnisch geht es darum, die **Null-Hypothese** "es besteht kein Zusammenhang zwischen der Konfession und der Wahlpräferenz zurückzuweisen.

## 2. Beobachtete und erwartete Häufigkeiten

Wir können zur besseren Übersicht in die ermittelten Häufigkeiten auch als Prozentwerte darstellen und erhalten die folgende Kreuztabelle (%-Werte in Klammern):

	katholisch		nicht katholisch		Summe
CVP	a	16 (32%)	b	6 (12%)	22 (44%)
andere	c	13 (26%)	d	15 (30%)	28 (56%)
Summe	29 (58%)		21 (42%)		50 (100%)

Prozentzahlen erleichtern zwar die Interpretation von Kreuztabellen; wie das obige Beispiel zeigt, erlauben Prozentzahlen jedoch oft keine eindeutige Aussagen über mögliche Zusammenhänge: 32% der Befragten sind katholische und nur 12% nicht katholische CVP-Wähler. Aber von den 42% nicht katholischen Befragten wählen immerhin 6 ebenfalls die CVP. Je stärker die Zellen in der Diagonalen (von links oben nach rechts unten) belegt sind, umso eher würden wir einen Zusammenhang zwischen katholischer Konfession und Wahl der CVP vermuten... Um eine präzise Aussage zu machen, ist jedoch eine Teststatistik nötig - Chi-Quadrat ist eine solche.

Wie vorher schon betont wird bei der Berechnung Chi-Quadrat auf Unterschiede zwischen beobachteten Häufigkeiten ( $f_o$ ) und erwarteten Häufigkeiten ( $f_e$ ) abgestellt. Die obige Kreuztabelle enthält die durch die Befragung einer Stichprobe von Personen beobachteten (erfragten) Häufigkeiten. Es stellt sich nun die Frage, welche Häufigkeiten wir in den Zellen der Kreuztabellen erwarten würden, wenn die beobachteten Werte nicht bekannt wären, sondern lediglich die Zeilensummen und Spaltensummen. Beide kann man wahrscheinlichkeitstheoretisch interpretieren:

- Von den 50 Fällen liegen 44% in der ersten Zeile und 58% in der ersten Spalte.
- Die Wahrscheinlichkeit eines Falles, in der ersten Zeile zu liegen zu kommen (also einen CVP-Wähler zu repräsentieren), ist somit  $p=0.44$ .
- Die Wahrscheinlichkeit eines Falles, in der ersten Spalte zu liegen zu kommen (also einen katholischen Wähler zu repräsentieren), ist also  $p=0.58$ .
- Die Wahrscheinlichkeit eines Falles, im ersten Feld der Häufigkeitstabelle oben links (Feld a) zu liegen zu kommen, ist das Produkt aus den beiden:  $p_a = 0.44 \cdot 0.58 = 0.2552$ .

---

<sup>1</sup> Das Cleavage-Modell erklärt das Wahlverhalten anhand geschichtlich geprägter gesellschaftlicher Konfliktlinien und der Annahme, Parteiensysteme entsprechen diesen Konfliktlinien. Untersuchungen in Deutschland haben z. B. gezeigt, dass sich die Konfessionszugehörigkeit im Wahlverhalten niederschlägt. Auch wenn sie durch ihren Beruf oder ihr sonstiges Umfeld den unterschiedlichsten Einflüssen ausgesetzt sind, folgen viele Wähler konfessionellen Mustern, Katholiken wählen eher die CDU, Protestanten tendenziell eher andere Parteien.

## Chi-Quadrat

---

- Dies multiplizieren wir mit der Gesamtzahl der Fälle N (hier 50) und erhalten die erwartete Häufigkeit 12.76.

Eine einfachere Variante für die Berechnung der erwarteten Häufigkeiten ist diese:

**Spaltensumme \* Zeilensumme / N**

Nun können die erwarteten Häufigkeiten (in Klammern) in die Häufigkeitstabelle eingetragen werden:

	katholisch	nicht-katholisch	Summe
CVP-Wähler	16 (12.76)	6 (9.24)	22
andere	13 (16.24)	15 (11.76)	28
Summe	29	21	50

### 3. Berechnung von Chi-Quadrat

**Chi-Quadrat wird wie folgt berechnet:**

1. Für jede Zelle einer Kreuztabelle wird die Differenz von beobachtener Häufigkeit  $f_o$  und erwarteter Häufigkeit  $f_e$  gebildet,
2. diese quadriert (um negative Vorzeichen zu beseitigen),
3. durch die erwarteten Häufigkeiten  $f_e$  geteilt
4. und die Resultate über alle Zellen aufsummiert.

In mathematischer Schreibweise abgekürzt heisst dies:

$$\chi^2 = \sum [(f_o - f_e)^2 / f_e]$$

Chi-Quadrat lässt sich recht einfach von Hand berechnen. Wir verwenden dazu die Daten unseres Beispiels:

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]**

# Chi-Quadrat

	katholisch		nicht-katholisch		Summe		
<b>CVP</b>	a	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a>	b	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a>	e	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a>	
<b>andere</b>	c	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a>	d	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a>	f	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a>	
<b>http://mesosworld.ch - Stand vom: 20.1.2010</b>							<b>9</b>
<b>Summe</b>	g	Dieses Element	h	Dieses Element	N	Dieses Element	

## Chi-Quadrat

---

Zuerst berechnen wir die Spalten- und die Zeilensummen, also  $a+b=e$ ,  $c+d=f$  usw.:

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]**

Als nächstes erstellen wir eine Tabelle, in die wir die beobachteten Häufigkeiten ( $f_o$ ) der vier "Zellwerte" (a, b, c und d) übertragen:

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]**

**Chi-Quadrat**

	$f_o$	$f_e$	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
<b>a</b>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>
<b>b</b>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>
<b>b</b>					
<a href="http://mesosworld.com">http://mesosworld.com</a>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]</p>

## Chi-Quadrat

---

Dann berechnen wir die erwarteten Häufigkeiten ( $f_e$ ); für den Fall a (=ist katholisch und wählt CVP) entspricht die erwartete Häufigkeit der Formel: "alle CVP-Wähler" \* "alle Katholiken" / Anzahl Fälle, also  $e \cdot g / N$ ; dasselbe gilt natürlich auch für die restlichen Zeilen:

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]**

Jetzt können wir Chi-Quadrat berechnen, in dem wir für jede Zeile zuerst  $(f_o - f_e)$  berechnen, dann quadrieren  $(f_o - f_e)^2$ , und schliesslich durch  $f_e$  teilen  $(f_o - f_e)^2 / f_e$ . Das Chi-Quadrat ist die Summe der vier Zeilenwerte:

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]**

Notieren Sie sich bitte das Ergebnis, wir werden noch öfters darauf zurückkommen.

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]**

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]**

## Chi-Quadrat

Welche Werte kann Chi-Quadrat annehmen?

	A	B	N	Chi-Quadrat
<b>I</b>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>
<b>II</b>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>	<p>Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. <a href="#">[link]</a></p>		

Setzen Sie unterschiedliche Werte in die Kreuztabelle und lassen Sie die entsprechenden Chi-Quadrate berechnen.

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [\[link\]](#)

### Dabei sollten Sie beachten:

- Einzelne Häufigkeiten dürfen den Wert null annehmen; allerdings darf keine Zeilen- oder Spaltensumme den Wert null haben.
- **Warum?** Weil in diesem Falle eine erwartete Häufigkeit  $f_e$  null wäre und durch null dividiert werden müsste – dies ist verboten.
- Was ist der kleinstmögliche Wert von Chi-Quadrat (in einer 2x2-Kreuztabelle) und wann erhält man diesen?
- **Antwort:** null, wenn alle Zellen der Kreuztabelle gleichmässig besetzt sind.
- Was ist der grösstmögliche Wert von Chi-Quadrat (in einer 2x2-Kreuztabelle) und wann erhält man ihn?
- **Antwort:**  $N$ =Anzahl der Fälle, wenn nur die Zellen der Diagonale besetzt sind.
- **Hinweis:** Bei einer 3x3-Kreuztabelle ist der grösstmögliche Wert von Chi-Quadrat  $2N$ , bei einer 4x4-Kreuztabelle  $3N$ , usw. (bei asymmetrischen Kreuztabellen entsprechend reduziert).

**Merke:** Chi-Quadrat ist nicht auf einen festen Wertebereich normiert, sondern hängt u. a. von der Anzahl Fälle der Untersuchung, der Grösse der Kreuztabelle und der Verteilung der Daten in dieser ab.

Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [\[link\]](#)

In dieser Übung berechnen Sie das Chi-Quadrat für eine Kreuztabelle mit 6 Zellen:



**Chi-Quadrat**

	katholisch		nicht-katholisch		Summe	
<b>CVP</b>	<b>a</b>	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]	<b>b</b>	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]	<b>g</b>	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]
<b>SP</b>	<b>c</b>	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]	<b>d</b>	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]	<b>h</b>	Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [link]
<b>http://mesosworld.ch - Stand vom: 20.1.2010</b>					<b>15</b>	
<b>andere</b>	<b>e</b>	Dieses Element	<b>f</b>	Dieses Element	<b>i</b>	Dieses Element

## Chi-Quadrat

---

Beginnen Sie damit, die Werte in der Kreuztabelle zu addieren.

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [\[link\]](#)**

Übertragen Sie die Zellenwerte jetzt in die Hilfstabelle, berechnen Sie  $f_e$  usw.



## Chi-Quadrat

---

Hier können Sie Ihre Berechnungen Schritt für Schritt kontrollieren:

1. Stimmt meine Berechnung von  $f_o$ ?

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [\[link\]](#)**

2. Stimmt meine Berechnung von  $f_e$ ?

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [\[link\]](#)**

3. Stimmt meine Berechnung von Chi-Quadrat?

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [\[link\]](#)**

Möchten Sie das alles nochmals (mit eigenen Zahlen) durchrechnen?

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [\[link\]](#)**

Sie Ihre Eingaben (und überschreiben Sie die vorgegebenen Werte in der Kreuztabelle).

## 4. Verteilung von Chi-Quadrat

In Übung 1 wurde festgestellt, dass Chi-Quadrat einen Wert zwischen null und einem Vielfachen von N (Zahl der Fälle der Untersuchung) annehmen kann – in Abhängigkeit von N, von der Verteilung der Daten in der Kreuztabelle und der Grösse der Kreuztabelle. In unserem vorherigen Beispiel (Schritt 2 und 4) haben wir für Chi-Quadrat einen Wert von 3.4979 ermittelt. Ist dies ein grosser oder ein kleiner Wert? Was sagt er aus?

	katholisch	nicht katholisch
CVP-Wähler	16	6
andere	13	15
<b>Chi-Quadrat: 3.4979</b>		

- Bei vollkommen gleichmässiger Verteilung der Fälle in der Kreuztabelle wäre Chi-Quadrat=0.

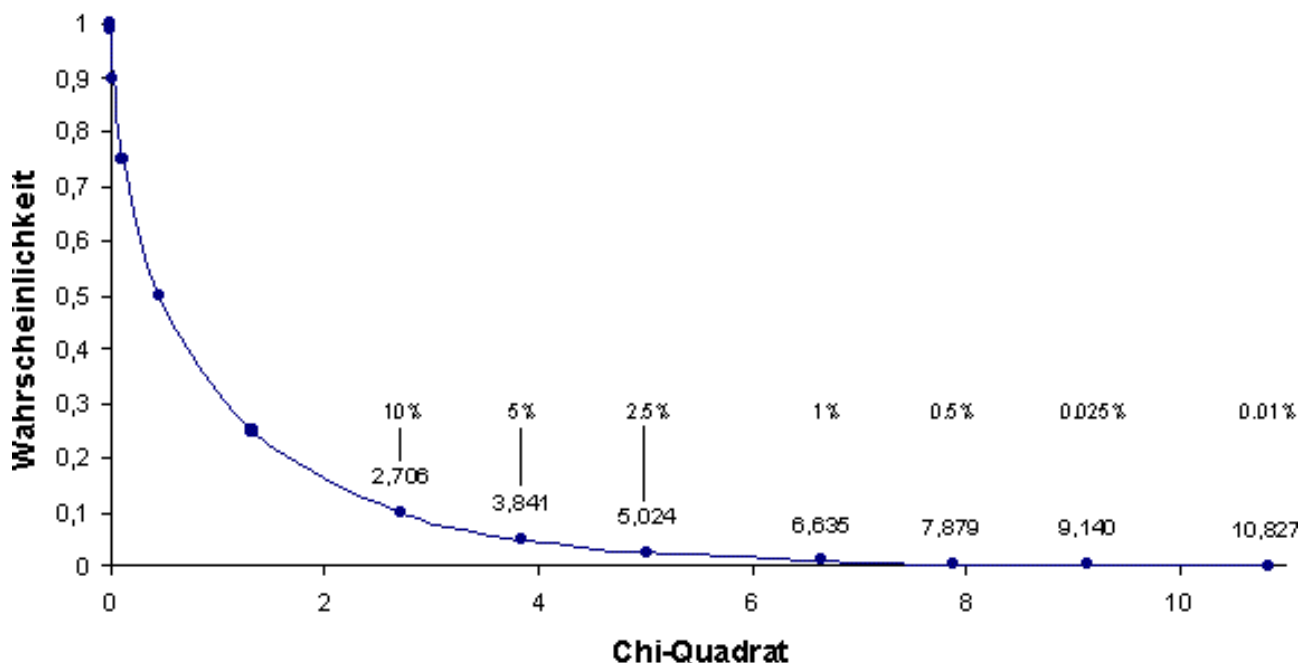
## Chi-Quadrat

---

- Der ermittelte Wert von Chi-Quadrat in unserem Beispiel ist jedoch 3.4979 und verweist auf eine Abweichung von einer vollkommen gleichmässigen Verteilung der Fälle (offenbar wählen Katholiken mehrheitlich tatsächlich die CVP – nicht-katholische Befragte wählen eher andere Parteien als die CVP).
- Es stellt sich aber dennoch die Frage, wie bedeutungsvoll die festgestellte Abweichung des berechneten Chi-Quadrat-Wertes von null ist. Können wir die Null-Hypothese "kein Zusammenhang" zurückweisen? Wir formulieren diese Frage nun in einer Weise, auf die uns die Statistik eine Antwort liefern kann:
- Wie wahrscheinlich ist es, einen Wert von 3.4979 für Chi-Quadrat (in einer 2x2-Kreuztabelle) durch schiere Zufälle bei der Stichprobenauswahl zu erhalten, wenn tatsächlich kein Zusammenhang zwischen den interessierenden Variablen besteht?

Glücklicherweise hat Chi-Quadrat wie andere statistische Kennzahlen auch eine typische sampling distribution, also eine Stichprobenkennwertverteilung. D.h. wir wissen, mit welcher Wahrscheinlichkeit entsprechende Werte von Chi-Quadrat rein zufällig zustande kommen, ohne dass ein Zusammenhang zwischen den interessierenden Variablen besteht.

Für den Fall einer 2x2-Kreuztabelle sieht diese Verteilung wie folgt aus (senkrecht die Wahrscheinlichkeit, waagerecht die entsprechenden Werte von Chi-Quadrat):



Aus dieser Standard-Verteilung für 2x2-Tabellen ist ersichtlich:

- Zunehmende Werte von Chi-Quadrat werden zunehmend unwahrscheinlicher.
- Ein Wert von Chi-Quadrat=3.84 tritt mit einer Wahrscheinlichkeit von 5% ( $p=0.05$ ) auf, auch wenn **kein Zusammenhang** zwischen den Variablen besteht. 3.84 ist der **kritische Wert** von Chi-Quadrat auf dem **Signifikanzniveau** von 5%.
- Ein Wert von Chi-Quadrat=6.63 tritt mit einer Wahrscheinlichkeit von 1% ( $p=0.01$ ) auf, auch wenn **kein Zusammenhang** zwischen den Variablen besteht. 6.63 ist der **kritische Wert** von Chi-Quadrat auf dem **Signifikanzniveau** von 1%.
- Ein Wert von Chi-Quadrat=7.88 tritt mit einer Wahrscheinlichkeit von 0.5% ( $p=0.005$ ) auf, auch wenn **kein Zusammenhang** zwischen den Variablen besteht. 7.88 ist der **kritische Wert** von Chi-Quadrat auf dem **Signifikanzniveau** von 0.5%.

## 5. Interpretation

Bei der Interpretation von Chi-Quadrat geht es um die Frage, ob die Null-Hypothese "es besteht kein Zusammenhang zwischen Parteienpräferenz und Konfession" im vorliegenden Beispiel zurückgewiesen werden kann.

	katholisch	nicht katholisch
CVP-Wähler	16	6
andere	13	15
<b>Chi-Quadrat: 3.4979</b>		

Nachdem wir Chi-Quadrat berechnet haben, sind noch folgende Schritte zu durchlaufen:

1. Wir wählen zunächst ein **Signifikanzniveau** oder eine **Irrtumswahrscheinlichkeit**. Dies ist jenes "Restrisiko", das man bei der Interpretation von statistischen Werten in Kauf nimmt. In den Sozialwissenschaften gilt 5% (oder  $p=0.05$ ) als jene Irrtumswahrscheinlichkeit, die allenfalls noch akzeptabel ist.
2. Wir kennen (siehe Schritt 9 oben) den **kritischen Wert** von Chi-Quadrat auf dem Signifikanzniveau von 5%. d. h. wir wissen, dass im vorliegenden Fall einer 2x2-Kreuztabelle bei zufälliger Auswahl von Fällen ein Wert von Chi-Quadrat=3.84 mit einer Wahrscheinlichkeit von 5% **auch dann** zu erwarten wäre, wenn in der Grundgesamtheit **kein Zusammenhang** zwischen den beiden Variablen "Konfession" und "Partei" besteht.
3. Wenn wir also die **Null-Hypothese** "kein Zusammenhang" **zurückweisen** wollen und ein Restrisiko von 5%, dass wir uns irren" in Kauf genommen werden soll, dann muss der berechnete Wert von Chi-Quadrat mindestens so gross wie der **kritische Wert** sein, d. h. 3.84 oder mehr.
4. Tatsächlich ist im vorliegenden Fall der berechnete Wert aber nur Chi-Quadrat=3.4979; wir können also die Null-Hypothese auf dem Signifikanzniveau von 5% **nicht zurückweisen**. Unsere ursprüngliche Hypothese, dass ein Zusammenhang zwischen "Konfession" und "Partei" besteht, wird durch unsere empirische Untersuchung **nicht gestützt**.

Bei dieser Interpretation ist wie überall in der Statistik ein **Irrtum** möglich – die Wahrscheinlichkeit eines Irrtums ist in unserem Falle sogar 5%. Prinzipiell sind bei der Interpretation von statistischen Resultaten 2 Formen von Irrtümern möglich:

1. **Fehler ersten Typs** (Type I error): Man akzeptiert die Hypothese (d. h. man weist die Null-Hypothese zurück), obschon sie tatsächlich falsch ist.
2. **Fehler zweiten Typs** (Type II error): Man verwirft die Hypothese (d. h. man kann die Null-Hypothese nicht zurückweisen), obschon sie tatsächlich richtig wäre.

## 6. Freiheitsgrade, Signifikanz, kritischer Wert

Es wurde vorher bereits darauf hingewiesen, dass der Wert von Chi-Quadrat auch von der Grösse der Kreuztabelle abhängt. Man gibt die Grösse der Kreuztabelle jedoch nicht mit der Zahl ihrer Felder an, sondern nennt ihre sog. Freiheitsgrade (df, degrees of freedom). Dabei ist df definiert als das Produkt der Anzahl Zeilen ( $r = \text{rows}$ ) der Kreuztabelle minus 1 und der Anzahl Spalten ( $c = \text{columns}$ ) der Kreuztabelle minus 1:

$$df = (r-1) * (c-1)$$

## Chi-Quadrat

---

Bei einer Kreuztabelle vom Format 2x2 haben wir also 1 Freiheitsgrad, bei einer Kreuztabelle vom Format 5x4 haben wir 12 Freiheitsgrade.

Die vorher schon gezeigte Stichprobenverteilung (sampling distribution) von Chi-Quadrat (siehe Schritt 9), d. h. die Wahrscheinlichkeit des schiefer zufälligen Auftretens entsprechender Chi-Quadrat-Werte, hängt nun von den Freiheitsgraden ab. Die Stichprobenverteilung hat je nach Freiheitsgrad eine andere Form.... Glücklicherweise gibt es Kreuztabellen, in denen man die entsprechenden Werte ablesen kann:

### Kritische Werte von Chi-Quadrat

df	p=0.1	p=0.05	p=0.025	p=0.01	p=0.001
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588
11	17.275	19.675	21.920	24.725	31.264
12	18.549	21.026	23.337	26.217	32.910
13	19.812	22.362	24.736	27.688	34.528
14	21.064	23.685	26.119	29.141	36.123
15	22.307	24.996	27.488	30.578	37.697
16	23.542	26.296	28.845	32.000	39.252
17	24.769	27.587	30.191	33.409	40.790
18	25.989	28.869	31.526	34.805	42.312
19	27.204	30.144	32.852	36.191	43.820
20	28.412	31.410	34.170	37.566	45.315

**Dieses Element (Animation, Video etc.) kann in der PDF version nicht dargestellt werden und ist nur in der online Version sichtbar. [\[link\]](#)**

### 7. Anzahl Fälle – Korrektur nach Yates

In der Statistik gilt generell: Je grösser die Stichprobe, um so verlässlicher die Resultate (das sog. Gesetz der grossen Zahl). In der Realität der Politikwissenschaft muss man jedoch oft mit einer geringen Fallzahl N auskommen.

#### Wie viele Fälle brauchen wir für die Berechnung von Chi-Quadrat?

- Bei kleineren Stichproben (weniger als 60 Fälle) sollte nach Meinung einiger (aber nicht aller) Autoren ein Chi-Quadrat-Wert mit sog. Yates-Korrektur verwendet werden.
- Wenn die erwartete Häufigkeit  $f_e$  für wenigstens eine Zelle der Kreuztabelle kleiner 5 ist (nach anderen Autoren: wenn mehr als 20 Prozent der erwarteten Häufigkeiten kleiner 5 sind), ist die Anwendbarkeit des Chi-Quadrat-Tests nicht mehr sinnvoll.
- Man wird je nach Datenlage entscheiden, ob Chi-Quadrat mit oder ohne Korrektur interpretiert wird, oder ob ein anderer statistischer Test (wie z. B. Fishers Exakt-Test) zur Anwendung kommen sollte.

Die Korrektur nach Yates "drückt" Chi-Quadrat nach unten und sorgt damit für eine vorsichtige, "konservative" Interpretation der Resultate. Sie wird hier nicht weiter besprochen. Beachten Sie aber, dass verschiedene Untersuchungen (z. B. Bradley et al. 1979) gezeigt haben, dass eine Korrektur deshalb nicht (immer) empfehlenswert ist.

### 8. Zusammenfassung zum Lernschritt

Merken Sie sich das **folgende Vorgehen** bei der **Anwendung von Chi-Quadrat**:

1. Kontrolle der Daten:
  - Handelt es sich um nominalskalierte Variablen?
  - Liegt eine genügende Anzahl von Fällen vor (am besten  $N > 60$ )?
2. **Auszählen der Häufigkeiten** und Erstellen einer **Kreuztabelle**.
3. Formulierung der **Hypothese** und der zugehörigen **Null-Hypothese**.
4. **Berechnung von Chi-Quadrat**; sind die erwarteten Häufigkeiten in allen Zellen grösser als 5?
5. Wahl des **Signifikanzniveaus** (Irrtumswahrscheinlichkeit) und Feststellen des kritischen Wertes von Chi-Quadrat.
6. **Vergleich** von Chi-Quadrat (bzw. korrigiertem Chi-Quadrat nach Yates) mit **kritischem Wert**: Kann die Null-Hypothese verworfen werden?
7. **Interpretation der Resultate** mit der **gebotenen Vorsicht**, denn **Achtung**: Chi-Quadrat gibt nur an, ob es in der untersuchten Kreuztabelle überzufällige Zellhäufigkeiten gibt, die auf mögliche Zusammenhänge zwischen den betreffenden Variablen hindeuten; welche diese sind, sagt Chi-Quadrat aber nicht.

## SPSS-Kochbuch

[Kochbuch: Chi-Quadrat in SPSS \[PDF, 84 KB\]](#)

#### Übungsdaten

- [im SPSS-eigenen \\*.sav Format, ready to use \(2 KB\)](#)
- [als Excel Tabelle, ohne Labels usw. \(17 KB\)](#)



# Glossar

### **Cleavage:**

Das Cleavage-Modell erklärt das Wahlverhalten anhand geschichtlich geprägter gesellschaftlicher Konfliktlinien und der Annahme, Parteiensysteme entsprächen diesen Konfliktlinien. Untersuchungen in Deutschland haben z. B. gezeigt, dass sich die Konfessionszugehörigkeit im Wahlverhalten niederschlägt. Auch wenn sie durch ihren Beruf oder ihr sonstiges Umfeld den unterschiedlichsten Einflüssen ausgesetzt sind, folgen viele Wähler konfessionellen Mustern, Katholiken wählen eher die CDU, Protestanten tendenziell eher andere Parteien.