Chi-Quadrat Verfahren

Chi-Quadrat Verfahren werden bei nominalskalierten Daten verwendet. Die einzige Information, die wir bei Nominalskalenniveau zur Verfügung haben, sind Häufigkeiten. Die Quintessenz von Chi-Quadrattechniken ist der Vergleich von beobachteten (ausgezählten) Häufigkeiten mit theoretisch erwarteten Häufigkeiten.

Dazu zwei Anwendungen:

1) Der Vergleich einer empirischen Verteilung mit einer theoretischen Verteilung (= Goodness-of-fit-Test)

2) 2 x 2 Felder-Tafeln

Ad 1)

Denken Sie zurück an unsere ersten Beispiele, die wir in der Statistik I im vergangenen WS behandelt haben. Wir haben gleich zu Beginn des WS kontiuierliche, intervallskalierte Daten durch *Klassenbildung* auf ein Nominalskalenniveau reduziert und anschließend die Häufigkeiten der verschiedenen Klassen bzw. Kategorien ausgezählt.

Nehmen wir an, in einem konkreten Fall lägen folgende Altersklassen vor: Die Klasse 1 umfasse alle 0-10 jährigen, Klasse 2 alle 11-20 jährigen, Klasse 3 alle 21 - 30 jährigen, Klasse 4 alle 31 bis 40 jährigen, Klasse 5 alle 41 bis 50 jährigen, Klasse 6 alle 51 bis 60 jährigen, Klasse 7 alle 61 bis 70 jährigen, Klasse 8 alle 71 bis alle darüber (71 oder mehr).

Für die diese Altersklassen wurden die Häufigkeiten ausgezählt:

Intervall	Beobachtete Häufigkeiten
0-10 Jahre	13
11-20	7
21-30	12
31-40	17
41-50	15
51-60	13

61-70	7
71-darüber	6

Um nun aber irgendwelche inferenzstatistische Schlüsse ziehen zu können, müssen wir diese Häufigkeiten mit nach bestimmten Kriterien erwarteten Häufigkeiten vergleichen.

So wollen wir beispielsweise wissen, ob die vorliegende Häufigkeitsverteilung einer Normalverteilung entspricht. (Dies ein spezieller Goodness-of-fit-Test)

Man vergleicht hierzu die folgenden Häufigkeiten:

Intervall	Beobachtete Häufigkeit	Erwartete Häufigkeit	Erwartete Prozent (bei Normalverteilung)
0-10 Jahre	13	9,54	10,6%
11-20	7	9,99	11,1%
21-30	12	14,13	15,7%
31-40	17	16,29	18,1%
41-50	15	15,57	17,3%
51-60	13	11,52	12,8%
61-70	7	7,20	8%
71-darüber	6	5,49	6,1%

Die erwarteten Häufigkeiten werden berechnet, indem man die Kategoriengrenzen z-transformiert (Mittelwert und Standardabweichung werden aus den gruppierten Daten berechnet) und die Fläche zwischen den Grenzen mit Hilfe der Standardnormalverteilungstabelle ausrechnet. So sind beispielsweise in den Grenzen von 0-10 10,6% aller Fälle, wenn die Daten normalverteilt sind. Da die Gesamthäufigkeiten gleich 90 sind, erwarten wir in diesem Intervall 10,6% von 90. Dies sind 9,54.

Einen solchen Vergleich von beobachteten und erwarteten Häufigkeiten findet man oft in der Statistik. Stellen Sie sich ein Kollektiv vor mit 20% weiblichen und 80% männlichen Studenten (beispielsweise ein Kurs in Mathematik). Nun seien in der Gesamtbevölkerung 50% weiblichen und 50% männlichen Geschlechts. Auch in diesem Beispiel können Sie die erwarteten Häufigkeiten berechnen, indem Sie einfach ausrechnen, welche Häufigkeiten Sie in Ihrem Kollektiv erwarten, wenn 50% männlich und 50% weiblich sind. Besteht der Kurs aus beispielsweise 40 Stundenten, so erwarten wir aufgrund der Bevölkerungsstatistik 20 Frauen und 20 Männer.

In jedem Falle werden bei den Chi-Quadratverfahren die Abweichungen der beobachteten von den erwarteten Häufigkeiten untersucht.

Die gesuchte Prüfgröße ist das so genannte Chi-Quadrat. Wir erhalten diese Prüfgröße, indem wir zunächst die Abweichung jeder beobachteten Häufigkeit zur erwarteten Häufigkeit berechnen, diese Abweichung quadrieren und durch die erwartete Häufigkeit dividieren. Die Summe der quadrierten und jeweils durch die erwarteten Häufigkeiten dividierten Abweichungen ist der gesuchte Chi-Quadratwert.

$$\chi^2 = \sum \frac{(f_b - f_e)^2}{f_e}$$

Sind die beobachteten Häufigkeiten mit den erwarteten Häufigkeiten identisch, so ist unser Chi-Quadrat gleich Null.

Zu unserem Beispiel: $(13 – 9.54)^2 / 9.54 = 1.25$

$(7 – 9.99)^2 / 9.99 = 0.89$

usw.

Dies ergibt in der Summe: 2.77

Für die Signifikanzprüfung müssen wir in der entsprechenden Chi-Quadrat-Tabelle nachschlagen. Dazu benötigen wir noch die *Anzahl der Freiheitsgrade* für die Prüfgröße Chi-Quadrat:

Freiheitsgrade: Die Anzahl der Summanden, die in die Berechnung des Chi-Quadrat einfließen, ist in unserem Falle gleich 8. Da die Summe der beobachteten Häufigkeiten gleich sein muss der Summe der erwarteten Häufigkeiten, steht ein Summand von vornherein fest. Die Anzahl der Freiheitsgrade reduziert sich daher um eins.

Unser spezielles Beispiel weicht aber insofern von anderen Goodness-of-Fit-Verfahren noch in zwei Hinsichten ab:

a) Da wir in diesem speziellen Falle die erwarteten Häufigkeiten über die z-Transformation gewonnen haben, sind diese erwarteten Häufigkeiten zusätzlich noch durch den Mittelwert und die Streuung der beobachteten Häufigkeiten festgelegt. Wir haben die erwarteten Häufigkeiten geschätzt, indem wir Mittelwert und Streuung der beobachteten Häufigkeiten verwendet haben. Wir müssen in diesem speziellen Falle zusätzlich noch zwei Freiheitsgrade abziehen. Das ergibt insgesamt 8 -1-2 = 5 Freiheitsgrade.

b) Die Nullhypothese besagt, dass die beobachteten Häufigkeiten sich nicht von den erwarteten Häufigkeiten unterscheiden. Bei einem Test zur Überprüfung der Normalverteilung sind wir aber gerade an der Nullhypothese interessiert. Normalerweise ist es umgekehrt: Wir sind normalerweise an der Alternativhypothese interessiert, wir wollen beispielsweise wissen, ob sich zwei Stichproben unterscheiden. Um die Entscheidung zugunsten der Alternativhypothese sozusagen 'wasserdicht' zu machen, wählen wir einen möglichst kleinen α-Fehler. (p= 0,05 oder gar 0,01) Im speziellen Falle einer Normalverteilungsüberprüfung wollen wir dahingegen umgekehrt das Risiko, fälschlicherweise die Nullhypothese beizubehalten, möglichst klein halten. Wir müssen uns also gegen den ß-Fehler absichern. Dies tun wir, indem wir eine möglichst große Irrtumswahrscheinlichkeit wählen. Beispielsweise ein Signifikanzniveau von 0,1.

Wir entnehmen bei einer Fläche von 0,90 und bei 5 Freiheitsgraden einen kritischen Chi-Quadratwert von 9,24 (siehe unten aufgelistete Tabelle).

Da unser errechneter Chi-Quadratwert weit kleiner ist als der kritische Wert, können wir die Nullhypothese nicht verwerfen. Das bedeutet im vorliegenden Falle, dass wir von einer Normalverteilung ausgehen können.

Bei allen anderen Goodness-of-Fit Verfahren gilt unser Interesse aber der Alternativhypothese. Wir wollen beispielsweise wissen, ob sich die Altersklassen in einem Sample von den Altersklassen in der Population unterscheiden. Oft werden einfach beobachtete Häufigkeiten mit vorgegebenen Prozentanteilen verglichen. (Vergleich der Verteilung des Geschlechts in einem Sample im Unterschied zur Verteilung in der Population; Vergleich der Verteilung von Altersgruppen im Unterschied zur Population etc.)

Die Freiheitsgrade sind in all diesen Fällen einfach die Anzahl der in den Chi-quadratwert eingehenden Summanden - 1.

Das Signifikanzniveau ist in diesem Falle 5% (signifikant) oder 1% (h0ch signifikant).

Handelt es sich also bei dem oben verwendeten Beispiel um einen Vergleich von Altersklassen in einem Sample mit Altersklassen in einer Population, so bestimmen wir den kritischen Chi-quadrat-Wert wie folgt:

Der Tabelle entnehmen wir bei einer Fläche von 0,95 und bei 7 Freiheitsgraden einen kritischen Chi-Quadratwert von 14,07.

Da unser errechneter Chi-quadratwert weit kleiner ist als der kritische Wert, können wir die Nullhypothese nicht verwerfen.

2) **2 x 2 Felder Tafel**

Häufig werden Chi-Quadratverfahren bei so genannten zweidimensionalen Häufigkeitstabellen (auch Kreuztabellen genannt) verwendet.

Dies ist dann der Fall, wenn eine Häufigkeitsverteilung nach einem zweiten Kriterium nochmals untergliedert wird.

Betrachten wir hierzu als Beispiel eine typische 2 x 2 Feldertafel:

	Geschlecht			
	männlich	weiblich		
Ja	25	10	35	Brillenträger
Nein	35	30	65	
	60	40	100	

Zur Erklärung: 35; 65; 60; 40 sind die so genannten *Randsummen*. Wir haben in dieser Stichprobe also 35% Brillenträger und 65 % ohne Brille. Das Geschlecht verteilt sich zu 60 % männlich und 40 % weiblich.

Die *Nullhypothese* lautet in diesem Falle: Das Merkmal "Geschlecht" ist in seiner Verteilung unabhängig vom Merkmal "Brillenträger".

Alternativhypothese: Die beiden Merkmale sind nicht unabhängig.

Für die Verteilung der beiden Merkmale haben wir nur unser Sample. Welche Häufigkeiten können wir also erwarten, wenn die beiden Merkmale unabhängig sind, wenn also die Nullhypothese zutrifft? Wir schätzen diese erwarteten Häufigkeiten aus den Randsummen.

Beispiel: Sind Geschlecht und Fehlsichtigkeit voneinander unabhängig, so müssten von den insgesamt 60 % Männern 35 % Brillenträger sein. Für die Merkmalskombination Männlich/Brillenträger erwarten wir die *relative Häufigkeit*: $60/100 \cdot 35/100 = 0.21$.
$0.21 \cdot 100$ ergibt daraus dann die erwartete Häufigkeit, also 21.

Allgemein berechnen wir die erwartete Häufigkeit nach der Formel:

$$f_e = (Zeilensumme \cdot Spaltensumme) / Gesamtsumme$$

Fassen wir die beobachteten und erwarteten Häufigkeiten in dem gegebenen Beispiel wie folgt zusammen:

f_b	f_e
25	21
10	14
35	39
30	26

Quadrieren wir die Differenz von beobachteten und erwarteten Häufigkeiten, dividieren sie durch die erwarteten Häufigkeiten und bilden die Summe, so erhalten wir die Prüfgröße Chi-quadrat.

Dies ist: $(25 - 21)^2 / 21 + (10 - 14)^2 / 14 + (35\text{-}39)^2 / 39 + (30\text{-}26)^2 / 26 =$

2.930

Zur inferenzstatistischen Absicherung dieses Chiquadrat benötigen wir die Anzahl der Freiheitsgrade. Diese bestimmen wir wie folgt:

Da die Wahrscheinlichkeiten für die Merkmalskombinationen aus den Randsummen geschätzt werden, haben wir $(k-1)\cdot(l-1)$ Freiheitsgrade. k und l sind die Anzahl der Ausprägungen der beiden Merkmale. Diese Anzahl beträgt im Falle einer 4 - Felder-Tafel jeweils 2. Die Freiheitsgrade bei einer 4 - Felder-Tafel sind demnach insgesamt 1!

Chi-quadrat kritisch auf dem 5 % Signifikanzniveau ist gemäß unten stehender Tabelle 3.84.

Das Ergebnis ist demnach nicht signifikant. Das Merkmal Brillenträger ist in der Verteilung unabhängig vom Merkmal Geschlecht.

Voraussetzungen des Tests:

Die erwarteten Häufigkeiten pro Zelle sollten mindestens gleich 5 sein.

Tabelle der Chi-Quadrat-Verteilung

From Wikibooks

Quantile der Chi-Quadrat-Verteilung nach ausgewählten Wahrscheinlichkeiten p und Freiheitsgraden

	Wahrscheinlichkeit p										
Freiheitsgrade	**0,005**	**0,01**	**0,025**	**0,05**	**0,1**	**0,5**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
1	0,00	0,00	0,00	0,00	0,02	0,45	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,10	0,21	1,39	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,35	0,58	2,37	6,25	7,81	9,35	11,34	12,84
4	0,21	0,30	0,48	0,71	1,06	3,36	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,15	1,61	4,35	9,24	11,07	12,83	15,09	16,75
6	0,68	0,87	1,24	1,64	2,20	5,35	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09	21,95
9	1,73	2,09	2,70	3,33	4,17	8,34	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	9,34	15,99	18,31	20,48	23,21	25,19
p →	**0,005**	**0,01**	**0,025**	**0,05**	**0,1**	**0,5**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
11	2,60	3,05	3,82	4,57	5,58	10,34	17,28	19,68	21,92	24,73	26,76
12	3,07	3,57	4,40	5,23	6,30	11,34	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	5,89	7,04	12,34	19,81	22,36	24,74	27,69	29,82
14	4,07	4,66	5,63	6,57	7,79	13,34	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	7,26	8,55	14,34	22,31	25,00	27,49	30,58	32,80
16	5,14	5,81	6,91	7,96	9,31	15,34	23,54	26,30	28,85	32,00	34,27
17	5,70	6,41	7,56	8,67	10,09	16,34	24,77	27,59	30,19	33,41	35,72
18	6,26	7,01	8,23	9,39	10,86	17,34	25,99	28,87	31,53	34,81	37,16

	19	6,84	7,63	8,91	10,12	11,65	18,34	27,20	30,14	32,85	36,19	38,58
	20	7,43	8,26	9,59	10,85	12,44	19,34	28,41	31,41	34,17	37,57	40,00
p →		**0,005**	**0,01**	**0,025**	**0,05**	**0,1**	**0,5**	**0,9**	**0,95**	**0,975**	**0,99**	**0,995**
	21	8,03	8,90	10,28	11,59	13,24	20,34	29,62	32,67	35,48	38,93	41,40
	22	8,64	9,54	10,98	12,34	14,04	21,34	30,81	33,92	36,78	40,29	42,80
	23	9,26	10,20	11,69	13,09	14,85	22,34	32,01	35,17	38,08	41,64	44,18
	24	9,89	10,86	12,40	13,85	15,66	23,34	33,20	36,42	39,36	42,98	45,56
	25	10,52	11,52	13,12	14,61	16,47	24,34	34,38	37,65	40,65	44,31	46,93
	26	11,16	12,20	13,84	15,38	17,29	25,34	35,56	38,89	41,92	45,64	48,29
	27	11,81	12,88	14,57	16,15	18,11	26,34	36,74	40,11	43,19	46,96	49,65
	28	12,46	13,56	15,31	16,93	18,94	27,34	37,92	41,34	44,46	48,28	50,99
	29	13,12	14,26	16,05	17,71	19,77	28,34	39,09	42,56	45,72	49,59	52,34
	30	13,79	14,95	16,79	18,49	20,60	29,34	40,26	43,77	46,98	50,89	53,67