Exakte Methoden bei kleinen Stichproben

[Signifikanztests]

Uwe Menzel, 22.01.10

Nebenwirkung von Arzneimitteln

- "Zerodin": 40% der Patienten mit Nebenwirkungen
- "Novodin"–Reklame: "deutlich weniger Nebenwirkungen"
- Stichprobe unter "Novodin"-Patienten:
  - 20 Patienten, zufällig ausgewählt
  - → 3 Patienten mit Nebenwirkungen

Schätzung: 3/20 = 15% (!)

Ist die Beobachtung (veränderte Proportion) statistisch signifikant?

Verteilung der Teststatistik

- X = Anzahl der Patienten in der Stichprobe mit Nebenwirkungen
- X kann als Ergebnis einer Folge von identischen und unabhängigen Einzelversuchen angesehen werden
- Einzelversuch = Befragung *eines* Patienten
  - 2 alternative Ereignisse
  - Wahrscheinlichkeit von Nebenwirkungen: p

$$\boxed{X \in Bin(n, p)}$$

$$p_X(x_1) = \binom{n}{x_1} \cdot p_1^{x_1} \cdot (1-p_1)^{n-x_1} = \frac{n!}{x_1! \cdot x_2!} \cdot p_1^{x_1} \cdot p_2^{x_2}$$

Signifikanztest: Hypothesen

- Nullhypothese (H_0): p=0.4
  - Novodin hat ebenfalls eine Nebenwirkungs-Rate von 40%
- Alternative Hypothese (H_a): p ≠ 0.4
  - Novodin hat tatsächlich eine andere Nebenwirkungs-Rate

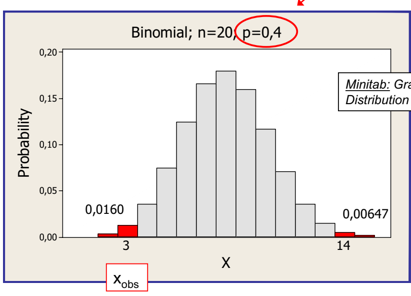
Signifikanztest: p-Wert

"Wahrscheinlichkeit, bei Gültigkeit der Nullhypothese ein mindestens so extremes Ereignis zu erhalten wie das beobachtete"

- Nullhypothese: X ∈ Bin(20, 0.4)
- Gesucht: P(X=3 oder extremer/ebenso extrem)

Graph in Minitab

Verteilung von X "unter H_0"



Minitab: Graph / Probability Distribution Plot

p-value = 0,0160 + 0,00647 = 0,02247

Signifikanztest: Entscheidung

- Signifikanzniveau α (üblich 0,05 oder 0,01):
- p-value = 0,022 < α
  - → H_0 wird verworfen
  - → Novodin hat eine *veränderte* Nebenwirkungsrate!

analytisch

Binomial Distribution: cdf

n = 20

k	0.01	0.05	0.1	0.2	0.3	0.4	0.5
0	0.818	0.358	0.122	0.012	0.001	0.000	0.000
1	0.983	0.736	0.392	0.069	0.008	0.001	0.000
2	0.999	0.925	0.677	0.206	0.035	0.004	0.000
3	1.000	0.984	0.867	0.411	0.107	0.016	0.001
4		0.997	0.957	0.630	0.238	0.051	0.006
5		1.000	0.989	0.804	0.416	0.126	0.021
6			0.998	0.913	0.608	0.250	0.058
7			1.000	0.968	0.772	0.416	0.132
8				0.990	0.887	0.596	0.252
9				0.997	0.952	0.755	0.412
10				0.999	0.983	0.872	0.588
11				1.000	0.995	0.943	0.748
12					0.999	0.979	0.868
13					1.000	0.994	0.942
14						0.998	0.979
15						1.000	0.994

Exakter Binomialtest

$X \in Bin(n, p) \qquad n = 20 \quad p = 0.4 \quad x_{obs} = 3$

$$p = P(X \le 3 \cup X \ge 14) = P(X \le 3) + P(X \ge 14)$$

$$= P(X \le 3) + 1 - P(X \le 13)$$

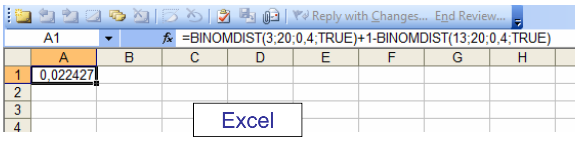
$$= F_X(3) + 1 - F_X(13) \qquad \textit{Verteilungsfunktion}$$

$$= 0.016 + 1 - 0.994$$

$$= 0.022$$

Exakter Binomialtest

$$p = P(X \le 3) + 1 - P(X \le 13) = F_X(3) + 1 - F_X(13)$$



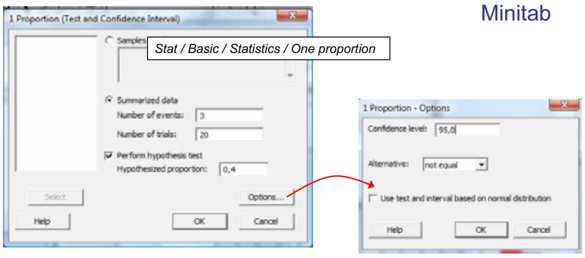
Excel

```
> p.value <- pbinom(3,20,0.4) + 1 - pbinom(13,20,0.4)
> p.value
[1] 0.02242704
```

Exakter Binomialtest

Exakter Binomialtest

Minitab



Stat / Basic / Statistics / One proportion

Test of p = 0,4 vs p not = 0,4   Exact

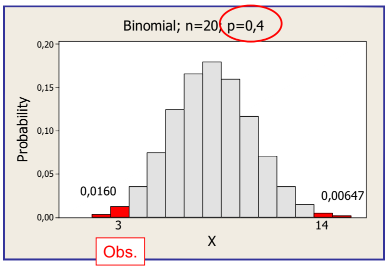
Sample	X	N	Sample p	95% CI	P-Value
1	3	20	0,150000	(0,032071; 0,378927)	**0,022**

Exakter Binomialtest in R, SAS, SPSS

```
> binom.test(3,20,p=0.4)

        Exact binomial test

data:  3 and 20
number of successes = 3, number of trials = 20, p-value = 0.02243
alternative hypothesis: true probability of success is not equal to 0.4
95 percent confidence interval:
 0.03207094 0.37892683
sample estimates:
probability of success
                 0.15
```

SAS, SPSS: $p_2 = 2 * p_1$??

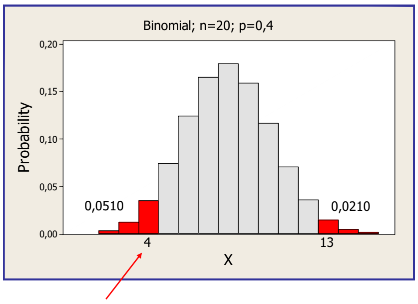
Exakter Binomialtest



Asymptotischer Test: $X \in Bin(n,p)$ $E(X) = n \cdot p$ $V(X) = n \cdot p \cdot (1-p)$

$X \in N\left(n \cdot p, \sqrt{n \cdot p \cdot (1-p)}\right)$ $n \cdot p \cdot (1-p) \geq 10$

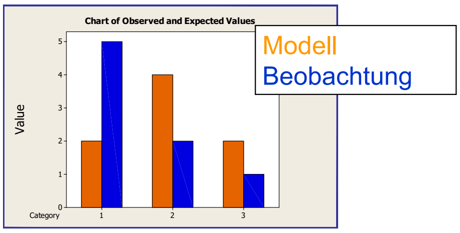
Verteilung von X "unter H_0"



Ergebnis "4 von 20" wäre nicht signifikant, bei einem Schätzwert von 20% – kleine Stichprobe !!

Anpassungs-Test

- "Goodness-of-Fit Test"
  - Passt ein Modell zu einer Beobachtung (Stichprobe)?



Mendelsche Experimente

Kreuzungs-Experiment bei Blütenpflanzen

	A	B
A	AA	AB
B	AB	BB

Erwartetes Farbverhältnis: **1** : **2** : **1**

Modell vs. Beobachtung

"observed"	5	2	1
"expected"	2	4	2

n = 8, k = 3
(kleine Stichprobe)

- Nullhypothese (H_0):
  - Farbverhältnis 1:2:1 (Modell wird *nicht* verworfen)
- Alternative Hypothese (H_a):
  - Farbverhältnis *nicht* 1:2:1 (Modell wird verworfen)

p-Wert

"Wahrscheinlichkeit, bei Gültigkeit der Nullhypothese ein mindestens so extremes Ereignis zu erhalten wie das beobachtete"

→ Wir müssen:

1. die Wahrscheinlichkeiten der Ereignisse unter H_0 berechnen
2. herausfinden, was "extrem" ist

1 Wahrscheinlichkeit der Beobachtung "unter H_0"

$$P(x_1, x_2) = \frac{n!}{x_1! \cdot x_2!} \cdot p_1^{x_1} \cdot p_2^{x_2} \quad x_1 + x_2 = n \quad p_1 + p_2 = 1 \quad \textit{Binomialverteilung}$$

$$P(x_1, x_2, x_3) = \frac{n!}{x_1! \cdot x_2! \cdot x_3!} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdot p_3^{x_3} \quad \sum x_i = n \quad \sum p_i = 1 \quad \textit{Multinomialvert.}$$

$$P(\mathbf{x}_{obs} \mid H_0) = P(x_1 = 5, x_2 = 2, x_3 = 1) = \frac{8!}{5! \cdot 2! \cdot 1!} \cdot 0.25^5 \cdot 0.5^2 \cdot 0.25^1 = 0.01025$$

Nullhypothese

```
> x = c(5,2,1)
> size = 8
> prob = c(0.25, 0.5, 0.25)
> dmultinom(x, size, prob)
[1] 0.01025391
```

"dmultinom"

2 Welche Ereignisse sind extrem?

"observed"	5	2	1
"expected"	2	4	2

$$\boxed{\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}} = \frac{(5-2)^2}{2} + \frac{(2-4)^2}{4} + \frac{(1-2)^2}{2} = 6$$

Ereignis *i* ist mindestens so extrem wie die Beobachtung wenn ♍$_i^2$ ¤ 6

Die Statistik ♍² ist bei kleinen Stichproben jedoch **nicht** ♍2 –verteilt !!

Exakter Multinomialtest

1. Finde alle möglichen Zahlentripel $\boldsymbol{x}=(x_1, x_2, x_3)$ und berechne deren Wahrscheinlichkeiten unter Annahme der Nullhypothese.
2. Finde alle Zahlentripel, die unter H_0 mindestens ebenso extrem wie die Beobachtung $\boldsymbol{x}=(5,2,1)$ sind und addiere deren Wahrscheinlichkeiten

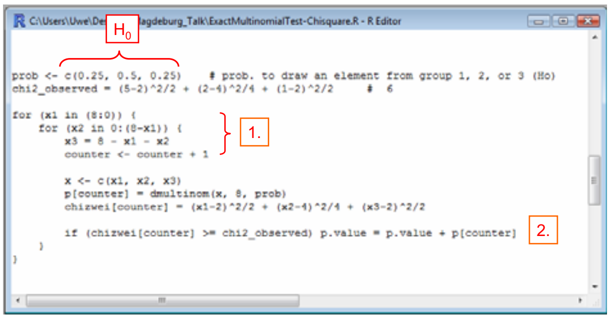
⟶ "p-Wert"

$$\boxed{M = \binom{n+k-1}{k-1}} \quad \textit{Kombinationen} \quad k = \textit{Anzahl der Kategorien}$$

$$M = 45 \quad bei \quad n = 8 \quad und \quad k = 3$$

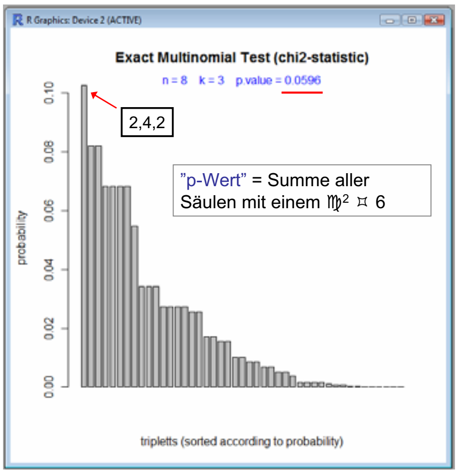
ExactMultinomialTest-Chisquare.R

Exakter Multinomialtest




```
prob <- c(0.25, 0.5, 0.25)     # prob. to draw an element from group 1, 2, or 3 (Ho)
chi2_observed = (5-2)^2/2 + (2-4)^2/4 + (1-2)^2/2      #  6

for (x1 in (8:0)) {
    for (x2 in 0:(8-x1)) {
        x3 = 8 - x1 - x2
        counter <- counter + 1

        x <- c(x1, x2, x3)
        p[counter] = dmultinom(x, 8, prob)
        chizwei[counter] = (x1-2)^2/2 + (x2-4)^2/4 + (x3-2)^2/2

        if (chizwei[counter] >= chi2_observed) p.value = p.value + p[counter]
    }
}
```

"ExactMultinomialTest-Chisquare.R"



Mendelsche Experimente

	A	B
A	AA	AB
B	AB	BB

Erwartetes Farbverhältnis: **2 : 4 : 2**
Beobachtetes Farbverhältnis: **5 : 2 : 1**

Anpassungstest: p-value = 0.06 > 0.05
→ Nullhypothese wird nicht verworfen
... obwohl Beobachtung große Abweichung vom Modell suggeriert
Kleine Stichprobe!

Monte-Carlo-Simulation

- Wahrscheinlichkeiten der Ereignisse unter H_0 "experimentell" ermitteln 
- Viele Stichproben mit n=8 Kugeln aus einer Urne mit Kugeln im Farb-Verhältnis 1:2:1
- Wie häufig kommen $\boldsymbol{x}$=(5,2,1) oder noch extremere/ebenso extreme Beobachtungen vor?

→ p-Wert

	A	B
A	AA	AB
B	AB	BB

RandMultinomialTest-Chisquare.R" 

"RandMultinomialTest-Chisquare.R"

MC-Methode

C:\Users\Uwe\Desktop\Magdeburg_Talk\RandMultinomialTest-Chisquare.R - R Editor

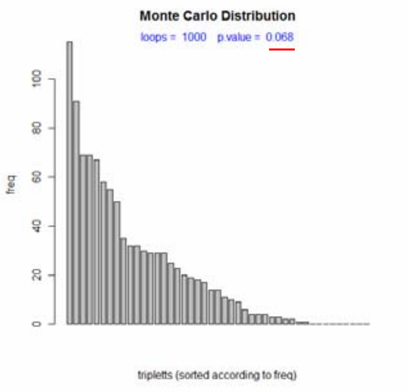
```
chisq_obs = (5-2)^2/2 + (2-4)^2/4 + (1-2)^2/2    #  6

h0 <- c("red","red","brown","brown","brown","brown","green","green")  # Ho

loops = 10000

for (n in 1:loops) {

    stichprobe <- sample(h0, size = 8, replace = TRUE)      # sample with replacement

    x1 = sum(stichprobe == "red")
    x2 = sum(stichprobe == "brown")
    x3 = sum(stichprobe == "green")

    # zähle Häufigkeit jeder der 45 Kombinationen von x1, x2 und x3
    counter_array[x1+1,x2+1,x3+1] = counter_array[x1+1,x2+1,x3+1] + 1
}

# Summiere die Häufigkeiten aller x-Tripel, zu denen ein chi2 >= 6 gehört:   (chi2_observed = 6)
```

H_0

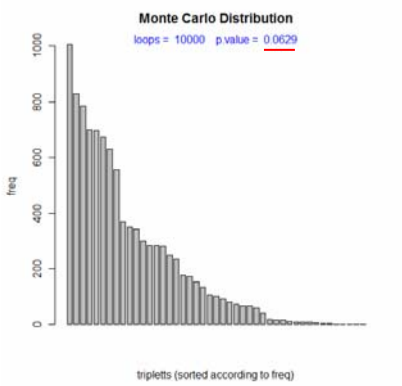
H_0

Ohne Benutzung der Multinomialverteilung für die x_i !

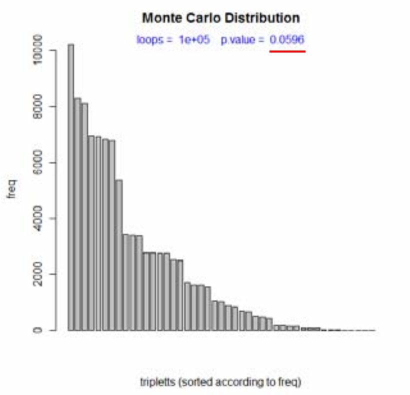
1.000 simulierte Stichproben



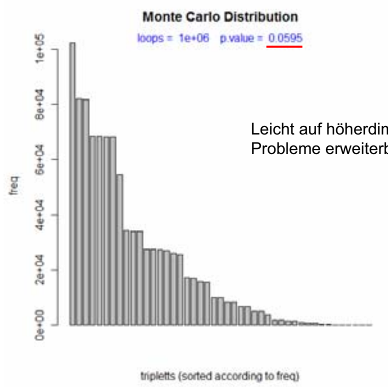
10.000 simulierte Stichproben



100.000 simulierte Stichproben



1.000.000 simulierte Stichproben



Leicht auf höherdimensionale Probleme erweiterbar!

SAS - Code

```
data snapdragons;
   input color $ observed;
   cards;
red   5
pink  2
white 1
;
proc freq data=snapdragons order=data;
   weight observed;
   tables color / chisq testp=(25 50 25);
   exact chisq / mc n=100000;
run;
```

Monte Carlo

SAS - Results

```
         Chi-Square Test
    for Specified Proportions
--------------------------------------
Chi-Square                      6.0000
DF                                   2
Asymptotic Pr > ChiSq           0.0498 Chi-square P-value

Monte Carlo Estimate for the Exact Test

 Pr >= ChiSq                  0.0594 Randomization P-value
 99% Lower Conf Limit         0.0575
 99% Upper Conf Limit         0.0613

 Number of Samples            100000
```

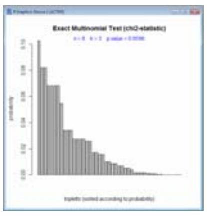
asymptotisch

Monte-Carlo

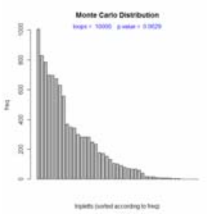
Anpassungstest

Summary

Exakter Multinomialtest



Monte-Carlo-Verfahren



Asymptotischer Test: $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$

$\chi^2 \in \chi^2(k-1) \qquad n \geq 6 \cdot r \cdot c$

[Pause]

Zwei Therapieformen: Untersuchung

	Patienten
Kontrollgruppe	5
Neue Therapie	7
Gesamt	12

Zweier Therapieformen: Resultat

	gesund	krank	Σ
Kontrollgruppe	2	3	5
Neue Therapie	6	1	7
Σ	8	4	12

- 40% Gesunde in Kontrollgruppe
- 85% Gesunde mit neuer Therapie
- 67% Gesunde insgesamt

Ist der beobachtete Unterschied in den Proportionen statistisch signifikant ?

Fishers Exakter Test auf Unabhängigkeit

- Nullhypothese: $p_1 = p_2 = p$
  - Die Wahrscheinlichkeit, gesund zu werden ist gleich groß für beide Therapieformen.
  - Gesundheitszustand und Therapieform sind unabhängig.
- Alternative Hypothese: $p_1 \neq p_2$

  $\longrightarrow$ Gesundheitszustand und Therapieform sind abhängig.

p-Wert

"Wahrscheinlichkeit, bei Gültigkeit der Nullhypothese ein mindestens so extremes Ereignis zu erhalten wie das beobachtete"

$\longrightarrow$ Wir müssen:

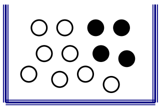
1. die Wahrscheinlichkeiten der Ereignisse unter H_0 berechnen
2. entscheiden, was "extrem" ist

	gesund	krank	
Kontr	2	3	5
Ther.	6	1	7
	8	4	12

Wahrscheinlichkeit dieser Beobachtung "unter H_0"

- **H_0:** Wahrscheinlichkeit der Genesung p = 2/3 (8 von den 12), unabhängig von Therapieform
- **H_0:** Die beobachteten Proportionen kommen durch puren Zufall zustande

$$\left.\begin{matrix} w = 8 \\ s = 4 \end{matrix}\right\} \quad N = 12 \qquad n = 5 \qquad P(w=2) = \frac{\binom{8}{2}\cdot\binom{4}{3}}{\binom{12}{5}}$$



Kontroll-gruppe

Wahrscheinlichkeit, dass unter den 5 Ausgewählten genau 2 Weiße sind

Wahrscheinlichkeit "der Tabelle" bei Annahme der Nullhypothese

	gesund	krank	
Kontr.	2	3	5
Ther.	6	1	7
	8	4	12

	gesund	krank	
Kontr.	a_{11}	a_{12}	Z_1
Ther.	a_{21}	a_{22}	Z_2
	S_1	S_2	N

$$P(a_{11}=2) = \frac{\binom{8}{2}\cdot\binom{4}{3}}{\binom{12}{5}}$$

$$\boxed{= P(\text{"}Tabelle\text{"})} \quad f = 1$$

$$\boxed{P(\text{"}Tabelle\text{"}) = \frac{\binom{S_1}{a_{11}}\cdot\binom{S_2}{a_{12}}}{\binom{N}{Z_1}} = \frac{S_1!\cdot S_2!\cdot Z_1!\cdot Z_2!}{N!\cdot a_{11}!\cdot a_{12}!\cdot a_{21}!\cdot a_{22}!}}$$

Auch für größere Tabellen !!

"Wahrscheinlichkeit der beobachteten Tabelle"

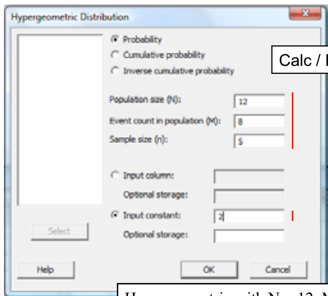
unter der Nullhypothese

	...	...	
...	2	3	5
...	6	1	7
	8	4	12

$$P(\text{"}Tabelle\text{"}) = \frac{S_1!\cdot S_2!\cdot Z_1!\cdot Z_2!}{N!\cdot a_{11}!\cdot a_{12}!\cdot a_{21}!\cdot a_{22}!} = \frac{8!\cdot 4!\cdot 5!\cdot 7!}{12!\cdot 2!\cdot 3!\cdot 6!\cdot 1!} = \frac{14}{99} = \underline{0.141414}$$

Hypergeometrische Verteilung

Minitab



Calc / Probability Distributions

Hypergeometric with N = 12, M = 8, and n = 5

x	P(X = x)
2	0,141414

Fisher's Exakter Test: p-value

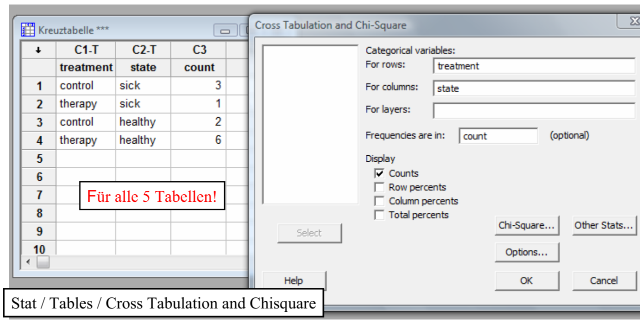
1. Finde alle "möglichen" Tabellen und berechne deren Wahrscheinlichkeiten unter Annahme der Nullhypothese.
2. Addiere die Wahrscheinlichkeiten *der* Tabellen, die mindestens so extrem von der Nullhypothese abweichen wie die beobachtete Tabelle.

→ "p-Wert"

	...	...	
...	2	3	5
...	6	1	7
	8	4	12

a_{12} kann zwischen 0 und 4 variieren → 5 mögliche Tabellen

Berechnung der χ^2-Statistik für alle Tabellen



Chi2_Fisher.MPJ

χ^2, Δp und p für alle Tabellen

Tabelle	χ^2	Δp	P
5 0 5 3 4 7 8 4 12	4,286	33%	0,070707
4 1 5 4 3 7 8 4 12	0,686	13%	0,353535
3 2 5 5 2 7 8 4 12	0,171	-7%	0,424242
2 3 5 6 1 7 8 4 12	2,743	-27%	0,141414
1 4 5 7 0 7 8 4 12	8,4	-47%	0,010101

$$\chi^2 = \sum_{i,j=1}^{2} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

$$E_{ij} = \frac{S_i \cdot Z_j}{n}$$

$$\Delta p = p_{obs} - p_{exp}$$

$$= p_{obs} - 0{,}67$$

$$P(\text{"table"}) = \frac{S_1! \cdot S_2! \cdot Z_1! \cdot Z_2!}{N! \cdot a_{11}! \cdot a_{12}! \cdot a_{21}! \cdot a_{22}!}$$

χ^2-Berechnung: Erwartete Frequenzen bei Unabhängigkeit

	gesund	krank	
Kontr.	E_{11}	...	5
Ther.	E_{21}	...	7
	8	4	12

$$\frac{E_{11}}{5} = \frac{8}{12}$$

$$\frac{E_{21}}{7} = \frac{8}{12}$$

	gesund	krank	
Kontr.	E_{11}	E_{12}	Z_1
Ther.	E_{21}	E_{22}	Z_2
	S_1	S_2	n

$$\frac{E_{11}}{Z_1} = \frac{S_1}{n} \Rightarrow E_{11} = \frac{S_1 \cdot Z_1}{n}$$

$$E_{ij} = \frac{S_i \cdot Z_j}{n}$$

"Wahrscheinlichkeit aller Tabellen" unter H_0

$$P(\text{"Tabelle"}) = \frac{S_1! \cdot S_2! \cdot Z_1! \cdot Z_2!}{N! \cdot a_{11}! \cdot a_{12}! \cdot a_{21}! \cdot a_{22}!}$$

5	0	5
3	4	7
8	4	12

P = 0,070707

4	1	5
4	3	7
8	4	12

P = 0,353535

3	2	5
5	2	7
8	4	12

P = 0,424242

2	3	5
6	1	7
8	4	12

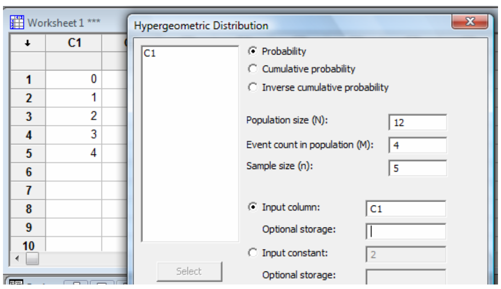
P = 0,141414

1	4	5
7	0	7
8	4	12

P = 0,010101

p-Wert = 0,141 + 0,071 + 0,01
= 0,222

Schlechte Versuchsplanung ?



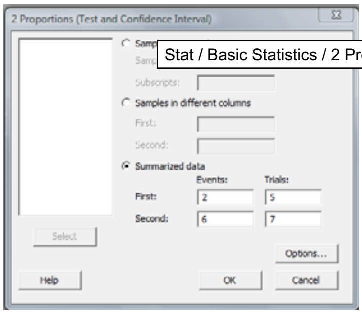
Hypergeometric with N = 12, M = 4, and n = 5

x	P(X = x)
0	0,070707
1	0,353535
2	0,424242
3	0,141414
4	0,010101

Minitab

Fishers Exakter Test

Minitab

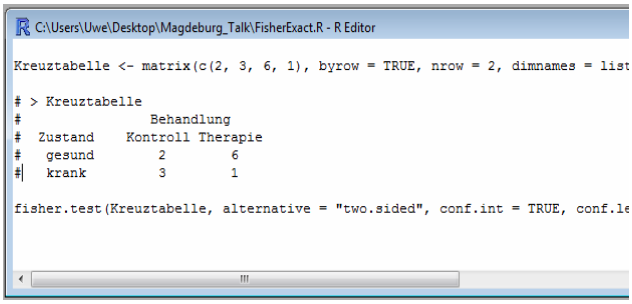


Stat / Basic Statistics / 2 Proportions

Normalapproximation ungeeignet

Test for difference = 0 (vs not = 0): Z = -1,79 P-Value = 0,074
Fisher's exact test: P-Value = 0,222
* NOTE * The normal approximation may be inaccurate for small samples.

Fishers Exakter Test




```
Kreuztabelle <- matrix(c(2, 3, 6, 1), byrow = TRUE, nrow = 2, dimnames = list

# > Kreuztabelle
#              Behandlung
#  Zustand    Kontroll Therapie
#    gesund        2        6
#    krank         3        1

fisher.test(Kreuztabelle, alternative = "two.sided", conf.int = TRUE, conf.le
```

Fisher's Exact Test for Count Data
data: Kreuztabelle
p-value = 0.2222

"FisherExact.R"

Vergleich zweier Therapieformen

	gesund	krank	
Kontrollgruppe	2	3	5
Neue Therapie	6	1	7
	8	4	12

p-value = 0,222

- 40% Gesunde in Kontrollgruppe
- 85% Gesunde mit neuer Therapie

- Die Nullhypothese wird nicht verworfen.
- Unterschied zwischen den Genesungsraten in beiden Gruppen nicht statistisch signifikant ... trotz der "überzeugenden" Prozentangaben
- Kleine Stichprobe !

Fisher Exact

Kleine / große Stichprobe

	gesund	krank	
Kontr.	2	3	5
Neu	6	1	7
	8	4	12

	gesund	krank	
Kontr.	200	300	500
Neu	600	100	700
	800	400	1200

- 40% Gesunde in Kontrollgruppe
- 85% Gesunde mit neuer Therapie

p-value = 0,222

p-value = 0,000

Signifikanztests bei kleinen Stichproben

- Intuition oft kein guter Ratgeber
  - Prozent- oder andere Angaben können irreführend sein
- Sorgfältige Versuchsplanung: – "Size does matter"

Anhang

1. Sign-Test (Wollen Mäuse einen eigenen Spiegel?)
2. Chisquare-Test
3. Binomialtest und Fishers Exakter Test in SAS, SPSS
4. Methoden zum Testen von Hypothesen
5. Normalapproximation für den binomischen Test
6. p-Wert: Welche Werte sind extremer als die Beobachtung?

Uwe Menzel, 22.1.2010

Sign-Test

John Arbuthnot
Born: April 1667 in Inverbervie, Kincardine, Scotland

Arbuthnot continued his scientific work submitting a paper to the Royal Society in 1710 discussing the slight excess of male births over female births in the years from 1629 to 1710. This paper, published in the *Philosophical Transactions,* is perhaps the first application of probability to social statistics and includes the first formal test of significance. In this paper Arbuthnot claims to demonstrate that divine providence, not chance, governs the sex ratio at birth.

Wollen Mäuse einen eigenen Spiegel?

- 16 Mäuse, alle Käfige mit 2 Räumen: einen mit Spiegel und einen ohne Spiegel
- Beobachtung: In welchen Raum hält sich die Maus am meisten auf?

Maus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
mit		x					x				x					
ohne	x		x	x	x	x		x	x	x		x	x	x	x	x

Sherwin, C.M. 2004. *Mirrors as potential environmental enrichment for individually housed laboratory mice. Appl. Anim. Behav. Sci. 87: 95-103.*

Wollen Mäuse einen eigenen Spiegel?

$\boxed{H_0 : p = 0.5}$ $\quad H_a : p \neq 0.5$

$Sticprobe: \quad M = 3$

$\boxed{M \in Bin(16, 0.5) \quad unter\ H_0}$

$$p = P(M \leq 3) + P(M \geq 13) \qquad symmetrisch\ (p = 0.5)$$

$$= P(M \leq 3) + 1 - P(M \leq 12)$$

$$= 0.0106 + 1 - 0.9894 = \underline{0.0212}$$

Die Nullhypothese (kein Raum bevorzugt, p=0.5) wird verworfen. Mäuse bevorzugen ... keinen Spiegel zu haben.

Sign-Test: Vergleich

Maus	Zeit mit Spiegel	ohne Spiegel
1	3	21
2	8	16
3	9	15
...	...	...

Bei Vorliegen konkreter Daten und unter der Voraussetzung, dass diese normalverteilt sind, ist der t-Test für gepaarte Stichproben besser.

Maus	Zeit mit Spiegel	ohne Spiegel
1	3	21
2	8	16
3	9	15
...	...	...

Bei Vorliegen konkreter Daten ist der verteilungsfreie Rangsummentest für gepaarte Stichproben besser.

Chisquare-Test

Unsere Stichprobe ist zu klein!

"observed"	5	2	1
"expected"	2	4	2

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \frac{(5-2)^2}{2} + \frac{(2-4)^2}{4} + \frac{(1-2)^2}{2} = 6$$

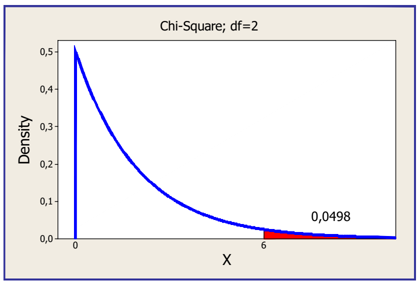
$$\chi^2 \in \chi^2(f) \quad mit \quad f = k - 1 \qquad wenn \quad E_i \geq 5$$

Nicht anwendbar ! (würde p = 0.0498 liefern)

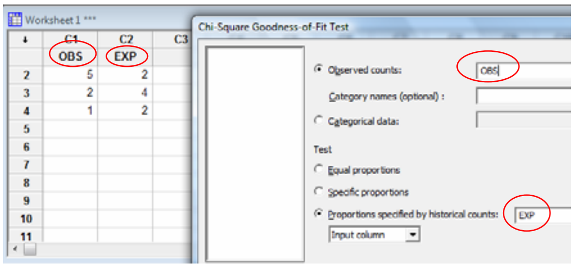
$$\Omega_{krit} = \left\{ \chi^2 > \chi^2_{0.05}(2) \right\} = \left\{ \chi^2 > 5.99 \right\}$$

Minitab: Stat / Tables / Chi-square Goodness of Fit

Chisquare Goodness-of-Fit



Chisquare Goodness-of-Fit



Minitab: *Stat / Tables / Chi-square Goodness of Fit*

Chisquare Goodness-of-Fit

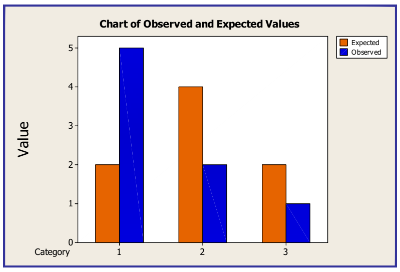
Chi-Square Goodness-of-Fit Test for Observed Counts in Variable: OBS

Category	Observed	Historical Counts	Test Proportion	Expected	Contribution to Chi-Sq
1	5	2	0,25	2	4,5
2	2	4	0,50	4	1,0
3	1	2	0,25	2	0,5

N	DF	Chi-Sq	P-Value
8	2	6	0,050

3 cell(s) (100,00%) with expected value(s) less than 5.

Chisquare Goodness-of-Fit



SAS: Exact binomial test

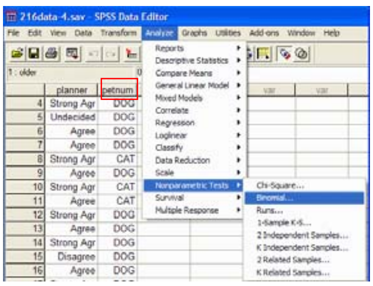
```
proc freq data = " C:\Users\Uwe\btst ";
  tables  female  / binomial(p=.5);
  exact binomial;
run;
```

Null hypotheses

Test whether the proportions of females ("*female*") differs significantly from 50%

	id	female	race	ses	schtyp	prog
1	70	0	4	1	1	1
2	121	1	4	2	1	3
3	86	0	4	3	1	1
4	141	0	4	3	1	3
5	172	0	4	2	1	2
6	113	0	4	2	1	2
7	50	0	3	2	1	1
8	11	0	1	2	1	2
9	84	0	4	2	1	1
10	48	0	3	2	1	2

SPSS: Exact binomial test



Binomial Test

		Category	N	Observed Prop.	Test Prop.	Asymp. Sig. (2-tailed)
Preferred Pet	Group 1	DOG	38	.83	.50	.000a
	Group 2	CAT	8	.17		
	Total		46	1.00		

a. Based on Z Approximation

n·p·(1-p)=11,5

Fishers Exakter Test in SAS

SAS automatically does Fisher's exact test for 2×2 tables. For greater numbers of rows or columns, you add a line saying `exact chisq;` . Here is an example using the data on heron and egret substrate use from above:

```
data birds;
   input bird $ substrate $ count;
   cards;
heron vegetation 15
heron shoreline  20
heron water      14
heron structures  6
egret vegetation  8
egret shoreline   5
egret water       7
egret structures  1
;
proc freq data=birds;
   weight count / zeros;
   tables bird*substrate / chisq;
   exact chisq;
run;
```

2 x 4 → f = 3

The results of the exact test are labelled "Exact Pr >= ChiSq"; in this case, P=0.5357.

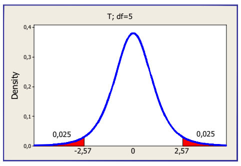
```
     Pearson Chi-Square Test
----------------------------------
Chi-Square                  2.2812
DF                               3
Asymptotic Pr >  ChiSq      0.5161
Exact      Pr >= ChiSq      0.5357
```

A) Möglichkeiten zur Durchführung von Hypothesentests

- A-priori-Festlegung des Signifikanzniveaus
  - Verteilung der Teststatistik ("T") unter H_0 ermitteln (Parameter: μ_0)
  - Signifikanzniveau α festlegen (5% $\rightarrow$ $\alpha = 0.05$)
  - Ablehnungsbereich (kritischen Bereich) ermitteln
  - Stichprobe $\rightarrow$ realisierten Wert der Teststatistik ("t") errechnen
  - Verwerfen der Nullhypothese wenn t im kritischen Bereich

$$\frac{\bar{X}-\mu_0}{s/\sqrt{n}} \in t(n-1)$$

$$\Omega_{krit} = \left\{ t : |t| \geq t_{\alpha/2}(n-1) \right\}$$

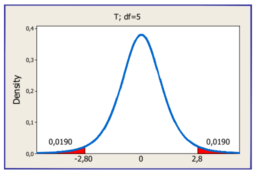


B) Möglichkeiten zur Durchführung von Hypothesentests

- Verwendung des p-Wertes
  - Verteilung der Teststatistik ("T") unter H_0 ermitteln (Parameter: μ_0)
  - Stichprobe
  - Berechnung des realisierten Wertes "t" der Teststatistik
  - Berechnung des p-Wertes auf Grundlage dieses Wertes
  - Verwerfen der Nullhypothese wenn p-Wert klein

$$\frac{\bar{X}-\mu_0}{s/\sqrt{n}} \in t(n-1)$$

$$t = \frac{\bar{x}_{obs}-\mu_0}{s/\sqrt{n}} \qquad p = P\left(abs(T) \geq t \mid H_0 \right)$$

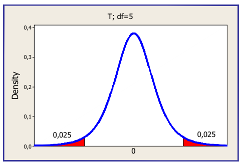


C) Möglichkeiten zur Durchführung von Hypothesentests

- Konfidenzintervall benutzen
  - Nullhypothese: Parameter μ_0
  - Stichprobe
  - Schätzwert für wahren Verteilungsparameter und KI berrechnen
  - H_0 wenn $\bigcirc_0$ außerhalb der Grenzen des KI liegt
  - (95% Konfidenzintervall entspricht 5% Signifikanzniveau)

$$P\left(-t_{\alpha/2}(n-1) < \frac{\bar{X}-\mu_0}{s/\sqrt{n}} \leq t_{\alpha/2}(n-1) \right) = 1-\alpha$$

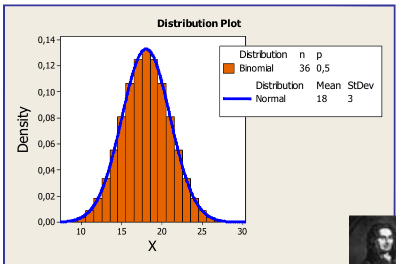
$$P\left(\bar{X} - t_{\alpha/2}(n-1)\cdot s/\sqrt{n} < \mu_0 \leq \bar{X} \quad t_{\alpha/2}(n-1)\cdot s/\sqrt{n} \right) = 1-\alpha$$



$\bigcirc_0$

Normalapproximation für den Exakten Binomialtest

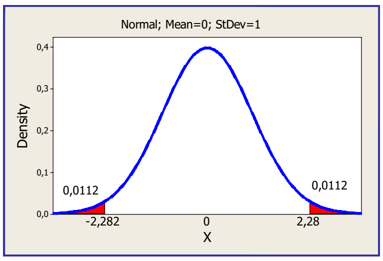
Vergleich mit Normalapproximation



Zentraler Grenzwertsatz

[Moivre-Laplace]

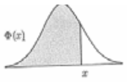
$$Z = \frac{X - n\cdot p}{\sqrt{n\cdot p\cdot(1-p)}} \in N(0,1)$$



p-value = 0,0112 + 0,0112 = 0,0224 < 0,05

Symmetrie ! (p=0,4)

Tabell 1. Standard normalfördelning



$\Phi(x) = P(X \le x)$, där $X \in N(0,1)$.
För negativa x, utnyttja att $\Phi(-x) = 1 - \Phi(x)$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.531
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.571
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.610
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.648
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.684
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.719
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.751
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.78
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.81
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.83
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.85
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.88
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.89
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.91
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.93
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.94
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.95
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.96
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.96
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.97
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99

$X \in Bin(n, p) \qquad E(X) = n \cdot p = 8 \qquad V(X) = n \cdot p \cdot (1-p)$

$X \in N\left(n \cdot p, \sqrt{n \cdot p \cdot (1-p)}\right) \qquad n \cdot p \cdot (1-p) > 10$

$p = 1 - P(3 < X \le 13) = 1 - [F_X(13) - F_X(3)]$

F_x für N statt für Bin

$$= 1 - \left[\Phi\left(\frac{13.5 - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}\right) - \Phi\left(\frac{3.5 - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}\right)\right]$$

Kontinuumskorrektur

$$= 1 - \left[\Phi\left(\frac{13.5 - 8}{2.191}\right) - \Phi\left(\frac{3.5 - 8}{2.191}\right)\right]$$

$$= 1 - [\Phi(2.51) - \Phi(-2.054)] = 1 - \Phi(2.51) + \Phi(-2.054)$$

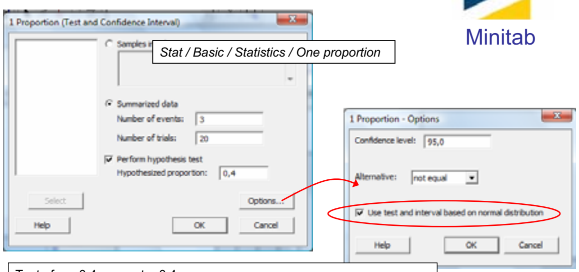
$\boxed{\Phi(-a) = 1 - \Phi(a)}$

$$= 1 - \Phi(2.51) + 1 - \Phi(2.054)$$

$$= 2 - 0.99396 - 0.98 = \underline{0.026}$$

Software

Minitab



Test of p = 0,4 vs p not = 0,4

Sample	X	N	Sample p	95% CI	Z-Value	P-Value
1	3	20	0,150000	(0,000000; 0,306491)	-2,28	0,022

Using the normal approximation.
The normal approximation may be inaccurate for small samples.

p-Wert-Berechnung: Welche Werte sind extremer als die Beobachtung ?

1. x_i ist extremer als die Beobachtung x_{obs}, wenn:

$$p(x_i) \le p_i(x_{obs})$$

ODER

2. x_i ist extremer als die Beobachtung x_{obs}, wenn:

$$\chi^2(x_i) \ge \chi^2(x_{obs})$$

p-Wert-Berechnung: Welche Werte sind extremer als die Beobachtung ?

1. Solche, die eine kleinere Wahrscheinlichkeit als die Beobachtung haben:

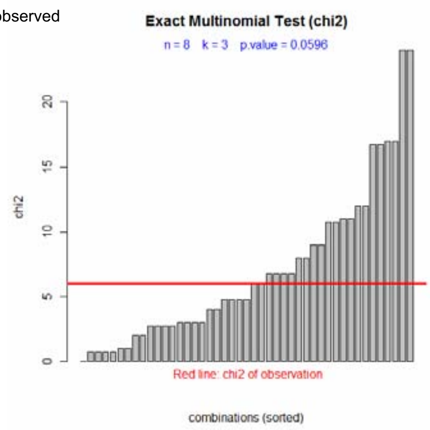
$$p - value = \sum_{p(x_i) \le p_i(x_{obs})} p(x_i)$$

ODER

2. Solche, die einen größeren χ2-Wert als die Beobachtung haben:

$$p - value = \sum_{\chi^2(x_i) \ge \chi^2(x_{obs})} p(x_i)$$

Mit chi2 > chi2_observed



p-Wert-Berechnung: Welche Werte sind extremer als die Beobachtung ?
Farbbeispiel 1:2:1

Beobachtung: (5,2,1)   Expectation: (2,4,2)

$p(x_i) \le p_i(x_{obs})$ *Kriterium, als extrem angesehen zu werden*

$p(5,2,1) = 0.01025$ $p(4,1,3) = 0.008544$ $p(3,1,4) = 0.008544$

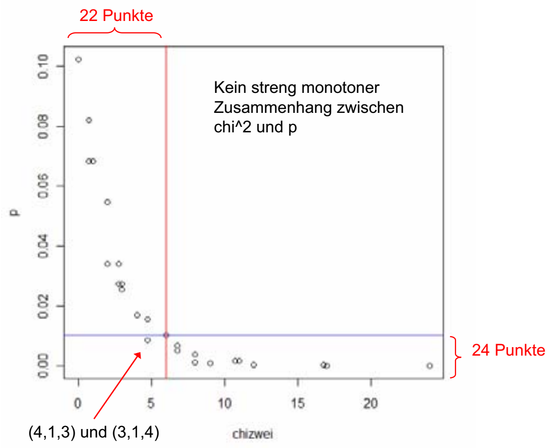
$(4,1,3)$ *und* $(3,1,4)$ *sind extremer!!*

ODER

$\chi^2(x_i) \ge \chi^2(x_{obs})$ *Kriterium, als extrem angesehen zu werden*

$\chi^2(5,2,1) = 6$ $\chi^2(4,1,3) = 4.75$ $\chi^2(3,1,4) = 4.75$

$\rightarrow$ $(4,1,3)$ *und* $(3,1,4)$ *sind nicht extremer!!*



Vergeich der Methoden

- *"Kleine Wahrscheinlichkeiten"*: p = 0.0767
- *"Chi-quadrat"*: p = 0.05957
- 0.05957 + 2 · 0.008544 = 0.0767
- Die beiden Punkte erklären den Unterschied

Anderes Kriterium: Abstände

$A = |O_1 - E_1| + |O_2 - E_2| + |O_3 - E_3|$ "Abstand"

$A_{obs} = |5-2| + |2-4| + |1-2| = 6$ *Abstand der Beobachtung*

$A_{4,1,3} = |4-2| + |1-4| + |3-2| = 6$ *Abstand von* $(4,1,3)$ *und* $(3,1,4)$

Nach diesem Kriterium werden die beiden Punkte genauso extrem, und würden damit dem größeren p-Wert entsprechen.