Statistische Methoden

Exakte statistische Verfahren bei kleinen Stichproben

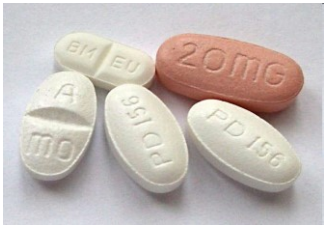
Uwe Menzel, 2010

uwe.menzel@matstat.de

www.matstat.org

Teil 1: Nebenwirkung von Arzneimitteln und Exakter Binomialtest

- "**Zerodin**": 40% der Patienten mit Nebenwirkungen
- "**Novodin**"–Reklame: "deutlich weniger Nebenwirkungen"
- (alle Arzneimittelnamen sind frei erfunden!)
- Stichprobe unter "**Novodin**"-Patienten:
  - 20 Patienten, zufällig ausgewählt
  - → 3 Patienten mit Nebenwirkungen



→ Schätzung der Proportion von **Novodin**-Nebenwirkungen:

$$p = \frac{3}{20} = 15\% \quad (!)$$

Die "entscheidende" Frage:

- Ist die Beobachtung - eine deutlich geringere Proportion von Nebenwirkungen bei **Novodin** - statistisch signifikant ?
- … oder ist die Beobachtung nur ein auf diese Stichprobe beschränkter Zufall ?

Verteilung der Teststatistik

- **Zufallsvariable** X : Anzahl der Patienten in der Stichprobe mit Nebenwirkungen
- X kann als Ergebnis einer Folge von **identischen** und **unabhängigen** Einzelversuchen angesehen werden
- Einzelversuch = Untersuchung **eines** Patienten :
  - 2 alternative Ergebnisse: Nebenwirkung / keine Nebenwirkung
  - Wahrscheinlichkeit von Nebenwirkungen im Einzelversuch: p
- → die stochastische Variable X ist **binomialverteilt**

Notation: $X \in Bin(n, p)$

- n = Anzahl der durchgeführten Einzelversuche
- p = "Erfolgswahrscheinlichkeit" im Einzelversuch

Anmerkung: Das Auffinden einer Nebenwirkung soll im folgenden als "Erfolg" bezeichnet werden, die Wahrscheinlichkeit p demzufolge als "**Erfolgs-wahrscheinlichkeit**". Natürlich kann man das Auftreten von Nebenwirkungen bei Patienten im ethischen Sinne nicht als Erfolg betrachten. Bezogen auf die Binomialverteilung ist es jedoch üblich, das mit der Wahrscheinlichkeit p verknüpfte Ereignis als Erfolg (engl. "success") zu bezeichnen.

Verteilung der Teststatistik $X \in Bin(n, p)$

Gemäß Binomialverteilung beträgt die Wahrscheinlichkeit, in einer Stichprobe der Größe n genau k "Erfolge" zu erzielen:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \qquad \binom{n}{k}: \text{Binomialkoeffizient}$$

Dies ist die **Wahrscheinlichkeitsfunktion der Binomialverteilung** (engl.: probability mass function, kurz **pmf**)

- $\boldsymbol{P(X = k)}$: Wahrscheinlichkeit, dass die Zufallsvariable X in der Stichprobe den Wert k annimmt. Man schreibt auch oft $p_X(k)$ anstelle von $P(X = k)$
- $\boldsymbol{p}$: Erfolgswahrscheinlichkeit (hier: W. des Auffindens einer Nebenwirkung)
- $\boldsymbol{1 - p}$: Wahrscheinlichkeit, dass das alternative Ereignis eintritt (hier: keine Nebenwirkung)

Die obige Formel gibt also die Wahrscheinlichkeit an, dass in der Stichprobe der Größe n genau k mal eine Nebenwirkung gefunden wird, vorausgesetzt, dass die Erfolgswahrscheinlichkeit im Einzelversuch gleich p ist. Wenn n Einzelversuche durchgeführt werden und k Erfolge eintreten, müssen zwangsläufig $(n - k)$ "Misserfolge" (= keine Nebenwirkung gefunden) eingetreten sein.

Hypothesen für den Signifikanztest, p-Wert

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

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

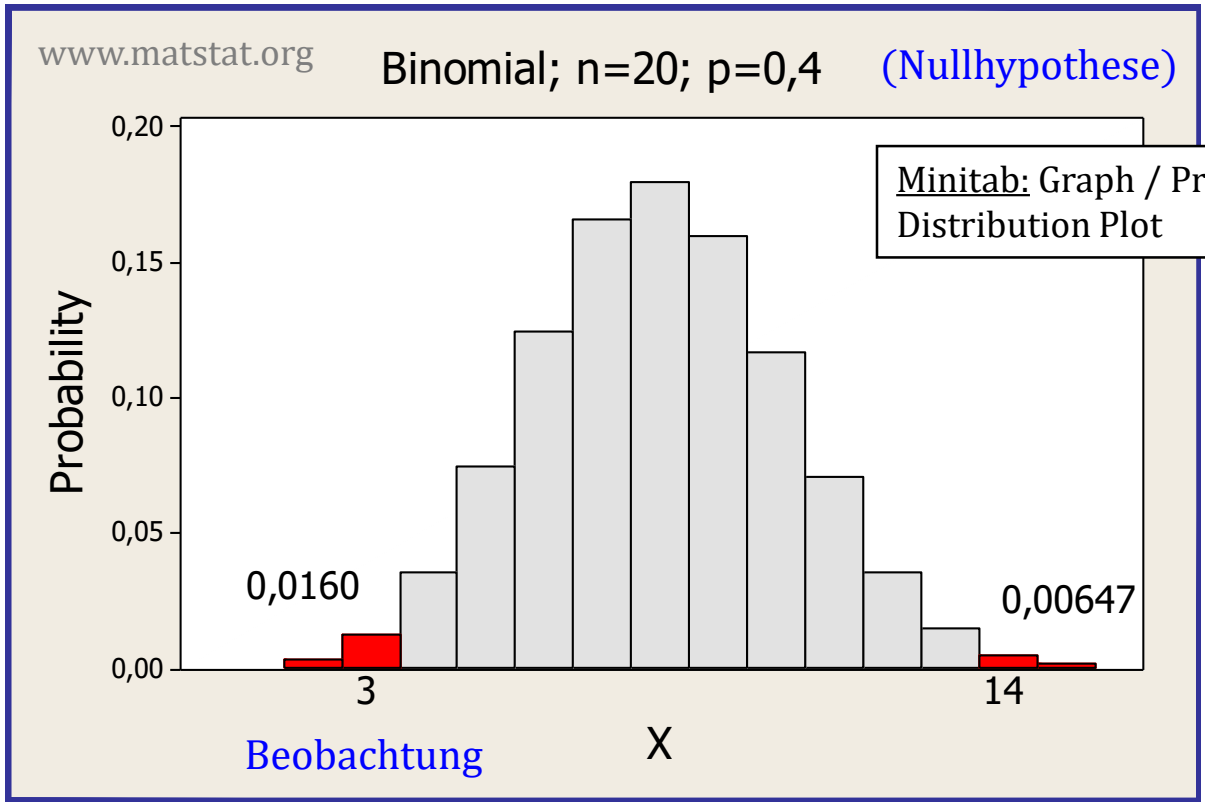
Nullhypothese: $p = 0.4$, also $X \in Bin(20, 0.4)$

Gesucht: **p-Wert**, also $P(X = 3)$ oder extremer/ebenso extrem
($X = 3$ war das beobachtete Ergebnis in der Stichprobe)

Anmerkung: das Ergebnis $X = 3$ ist schon ziemlich extrem, wenn $n = 20$ und $p = 0.4$. Mit diesen Parametern würde man "im Mittel" $20 \cdot 0.4 = 8$ Patienten mit Nebenwirkungen erwarten. (Dies ist der **Erwartungswert** für die Binomialverteilung.) Noch extremer wäre $X = \{0,1,2\}$, aber auch hohe Werte für X, z. B. $X = \{18,19,20\}$ - diese Ereignisse sind unter Annahme der Nullhypothese noch unwahrscheinlicher als $X = 3$.

Verteilung von X "unter H_0"

Wenn die Nullhypothese zutrifft, wenn also $X \in Bin(20, 0.4)$, dann hat die Zufallsvariable X die im Bild wiedergegebene Verteilung. Der beobachtete Wert war $X = 3$. Eine Anzahl von k-Werten hat eine geringere (oder gleiche) Wahrscheinlichkeit (rot dargestellt). Die Summe aller dieser Wahrscheinlichkeiten ist der p-Wert.



$$p.val = 0.0160 + 0.00647 = 0.02247$$

Dies ist also der p-Wert, **nicht** der Parameter der Binomialverteilung!

Verteilung von X "unter H_0"

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Wahrscheinlichkeitsfunktion für die Binomialverteiung

```
x = 0:20
dbinom(x, 20, 0.4)
cbind(x, dbinom(x, 20, 0.4))

[1,]  0 3.656158e-05
[2,]  1 4.874878e-04
[3,]  2 3.087423e-03
[4,]  3 1.234969e-02
[5,]  4 3.499079e-02
[6,]  5 7.464702e-02
[7,]  6 1.244117e-01
[8,]  7 1.658823e-01
[9,]  8 1.797058e-01
[10,]  9 1.597385e-01
[11,] 10 1.171416e-01
[12,] 11 7.099488e-02
[13,] 12 3.549744e-02
[14,] 13 1.456305e-02
[15,] 14 4.854351e-03
[16,] 15 1.294494e-03
[17,] 16 2.696862e-04
[18,] 17 4.230371e-05
[19,] 18 4.700412e-06
[20,] 19 3.298535e-07
[21,] 20 1.099512e-08
```

Hier noch eine Tabelle der 21 Werte, die X annehmen kann, sowie deren Wahrscheinlichkeit (R-Code). Es ist $P(X = 3) = 1.234$. Die Summe der Wahrscheinlichkeiten aller Ergebnisse, die gleich wahrscheinlich oder unwahrscheinlicher sind, ist der p-Wert:

```
p = sum(dbinom(0:3, 20, 0.4)) +
    sum(dbinom(14:20, 20, 0.4))
p   # 0.02242704
```

Dies ist also der p-Wert, **nicht** der Parameter in der Binomialverteilung!

Signifikanztest: Entscheidung

Wird die Nullhypothese in Anbetracht der beobachteten Stichprobe zurückgewiesen oder (zumindestens bis auf Weiteres) akzeptiert?

- wenn das beobachtete Ergebnis "unter H_0" - also unter Annahme der Richtigkeit der Nullhypothese – sehr unwahrscheinlich ist, wird die Nullhypothese verworfen (genau!, wir verwerfen die Hypothese, nicht etwa unsere Daten). Die Nullhypothese wird also verworfen, wenn der p-Wert klein ist. In der Praxis wird $p \leq 0.05$ meist als "klein" betrachet, oft auch $p \leq 0.01$.
- wenn der p-Wert größer als dieses Limit ist, wird die Nullhypothese akzeptiert. Dies bedeutet nicht unbedingt, dass die Nullhypothese auch richtig ist, es kann auch sein, dass die Stichprobe einfach zu klein war, um die Nullhypothese zu widerlegen (die "Power" des Tests hat nicht ausgereicht). Im Prinzip können Nullhypothesen nur widerlegt, nicht aber "bewiesen" werden. (Das Widerlegen der Nullhypothese ist auch meist das Ziel eines solchen Tests!).

Das oben genannte Limit wird Signifikanzniveau (α) genannt.

Zurück zu unserem Beispiel:

Sei $\alpha = 0.05 \rightarrow p = 0.022 < \alpha \rightarrow$ die Nullhypothese wird ("auf Signifikanzniveau 0.05") verworfen. Das heißt also, dass das neue Medikament tatsächlich eine (statistisch signifikant) andere Nebenwirkungsrate hat als das alte.

Berechnung des p-Wertes

Der p-Wert kann auch mit Hilfe der Verteilungsfunktion berechnet werden. Die Verteilungsfunktion für die Binomialverteilung ist an vielen Stellen für ausreichend viele Werte von n und p tabelliert (siehe nächste Seite).

$$p = P(X \leq 3 \cup X \geq 14) = P(X \leq 3) + P(X \geq 14)$$
$$= P(X \leq 3) + 1 - P(X \leq 13) = F_X(3) + 1 - F_X(13)$$
$$= 0.016 + 1 - 0.9935 \approx 0.022$$

F_X : Verteilungsfunktion der Zufallsvariablen X

Außerdem gibt es in vielen Software-Tools Funktionen, die den Test ausführen, z.B. in R:

```
binom.test(c(3,17), n = 20, p = 0.4)
# data:  c(3, 17)
# number of successes = 3, number of trials = 20, p-value = 0.02243
# alternative hypothesis: true probability of success is not equal to 0.4
```

Das erste Argument ist hier ein Vektor, der die Elemente "number of successes" und "number of failures" enthält. (Natürlich ist das Auffinden einer Nebenwirkung kein "success", wie bereits erwähnt ist diese Bezeichnung jedoch für die Binomialverteilung üblich ...)

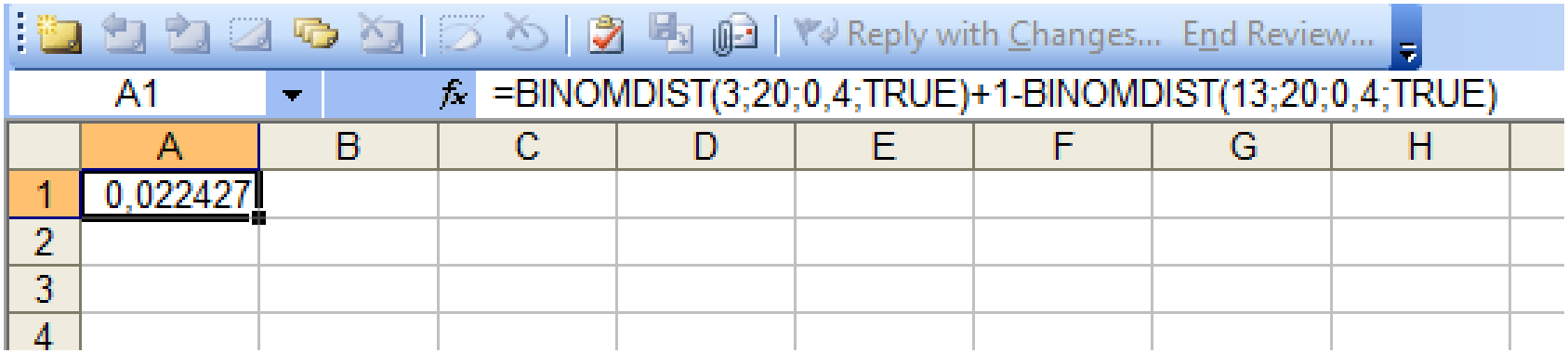
Verteilungsfunktion für die Binomialverteilung (Ausschnitt)

$X \in Bin(n,p)$ $\quad n = 20$ $\quad p = 0.4$

F_X		p									
n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
20	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.7358	.3917	.1756	.0692	.0243	.0076	.0021	.0005	.0001	.0000
	2	.9245	.6769	.4049	.2061	.0913	.0355	.0121	.0036	.0009	.0002
	3	.9841	.8670	.6477	.4114	.2252	.1071	.0444	.0160	.0049	.0013
	4	.9974	.9568	.8298	.6296	.4148	.2375	.1182	.0510	.0189	.0059
	5	.9997	.9887	.9327	.8042	.6172	.4164	.2454	.1256	.0553	.0207
	6	1.00	.9976	.9781	.9133	.7858	.6080	.4166	.2500	.1299	.0577
	7	1.00	.9996	.9941	.9679	.8982	.7723	.6010	.4159	.2520	.1316
	8	1.00	.9999	.9987	.9900	.9591	.8867	.7624	.5956	.4143	.2517
	9	1.00	1.00	.9998	.9974	.9861	.9520	.8782	.7553	.5914	.4119
	10	1.00	1.00	1.00	.9994	.9961	.9829	.9468	.8725	.7507	.5881
	11	1.00	1.00	1.00	.9999	.9991	.9949	.9804	.9435	.8692	.7483
	12	1.00	1.00	1.00	1.00	.9998	.9987	.9940	.9790	.9420	.8684
	13	1.00	1.00	1.00	1.00	1.00	.9997	.9985	.9935	.9786	.9423
	14	1.00	1.00	1.00	1.00	1.00	1.00	.9997	.9984	.9936	.9793
	15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.9997	.9985	.9941

Exakter Binomialtest – auch in Excel

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

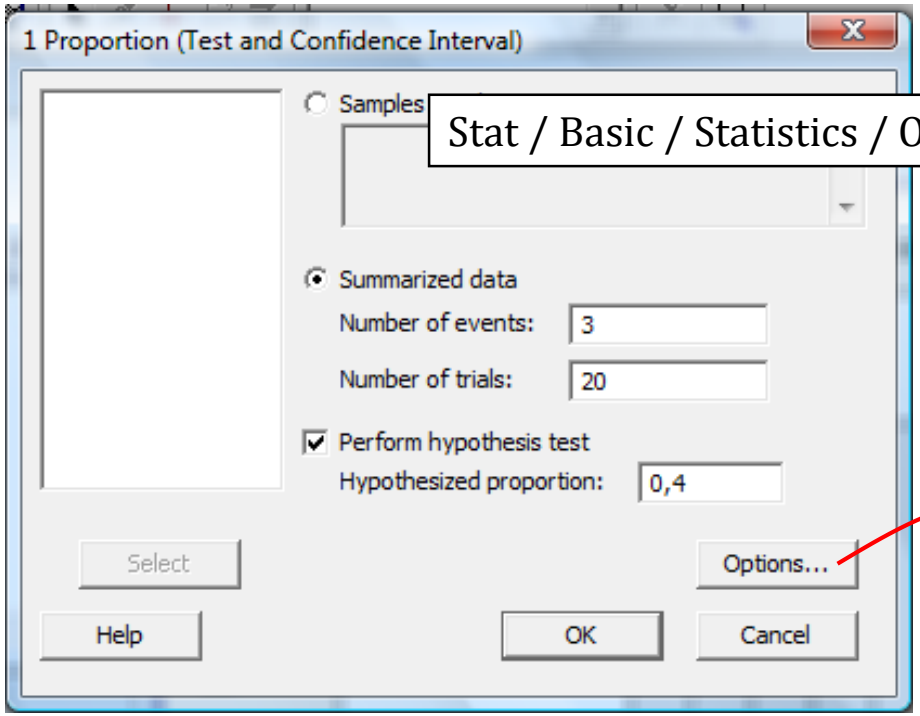


Dasselbe in R (pbinom ist die in R implementierte Verteilungsfunktion):

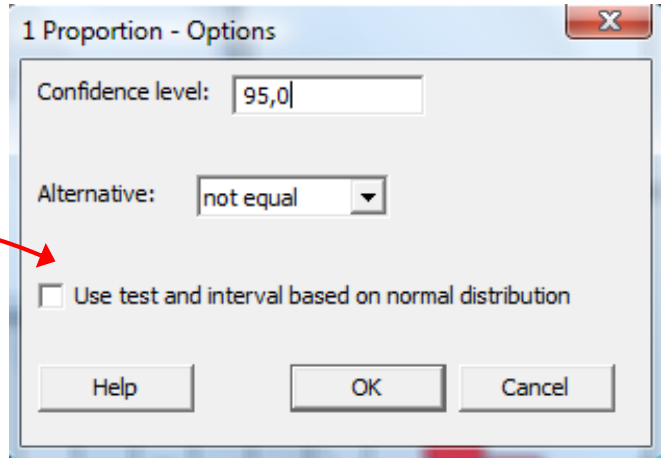
```
> p.value <- pbinom(3,20,0.4) + 1 - pbinom(13,20,0.4)
> p.value
[1] 0.02242704
```

Exakter Binomialtest

Minitab



Stat / Basic / Statistics / One proportion



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

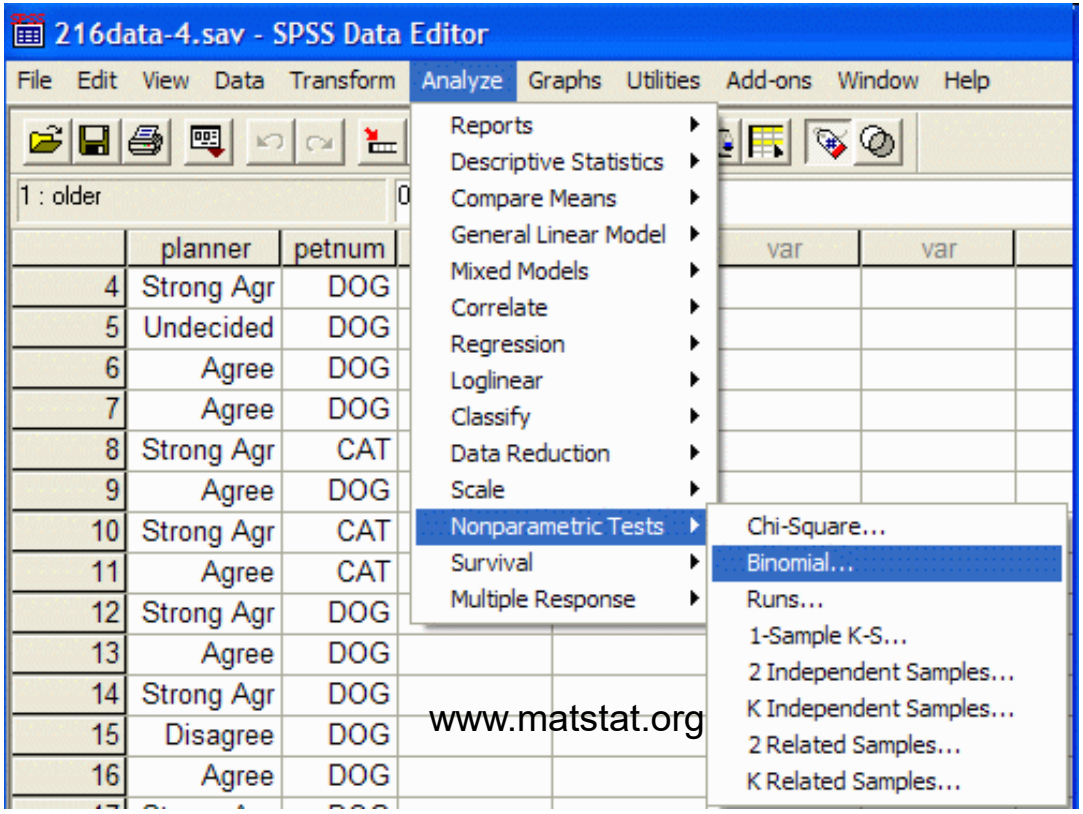
Sample	X	N	Sample p	95% CI	P-Value
1	3	20	0,150000	(0,032071; 0,378927)	**0,022**

Exakter Binomialtest: SAS

```
proc freq data = " C:\Users\Uwe\btst ";
  tables  medic  / binomial(p=.4);
  exact binomial;
run;
```

Nullhypothese

Exakter Binomialtest: SPSS



Vorsicht bei kleinen Stichproben!

Neue Stichprobe unter "**Novodin**"-Patienten:

20 Patienten, zufällig ausgewählt
→ 4 Patienten mit Nebenwirkungen (nur eine/r mehr!)

→ Schätzung der Proportion von **Novodin**-Nebenwirkungen:

$$p = \frac{4}{20} = 20\% \quad (!)$$

scheint auch viel besser zu sein als **Zerodin** mit 40%, aber:

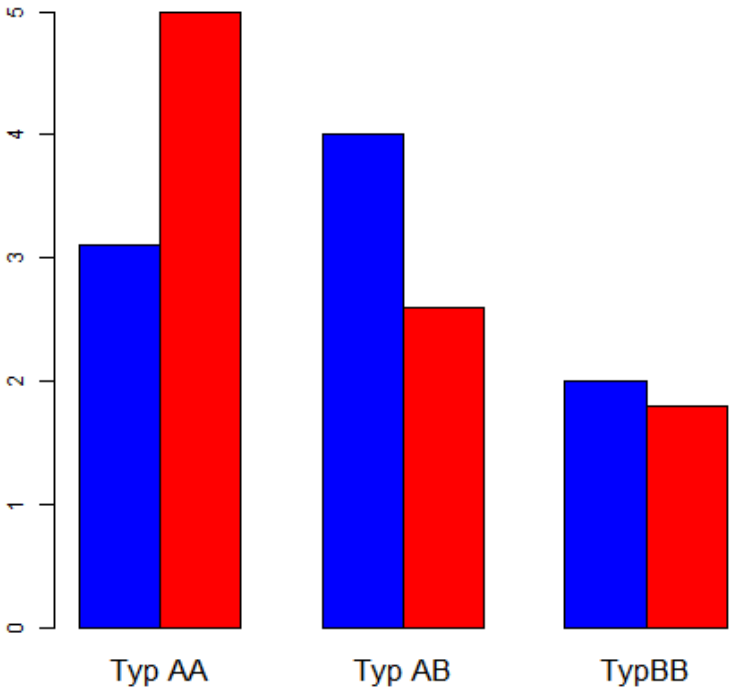
```
binom.test(c(4,16), n = 20, p = 0.4)
# data:  c(4, 16)
# number of successes = 4, number of trials = 20, p-value = 0.07198
# alternative hypothesis: true probability of success is not equal to 0.4
```

Nur ein Patient mit Nebenwirkung mehr → der Test ist nicht mehr signifikant (auf Signifikanzniveau 0.05). Daher müssen kleine Stichproben immer mit Vorsicht interpretiert werden!

[Pause]

Teil 2: Anpassungs-Test

- neudeutsch: ”Goodness-of-Fit Test”
- versucht die Frage zu beantworten, ob ein (statistisches) Modell zu einer Beobachtu



$K = 3$
(3 Kategorien)

Das Säulendiagramm stellt zwei Datengruppen, eingeteilt in 3 Kategorien, dar:

- die nach einem Modell erwarteten Werte (rot)
- die beobachteten Werte (blau)

Die Frage ist, ob man angesichts dieser Daten das Modell aufrechterhalten kann.

Mendelsche Experimente

Kreuzungs-Experiment bei Blütenpflanzen

Allele A (rot) und Allel B (grün) (zu gleichen Anteilen) in der Elterngeneration. In der 1. Folgegeneration erwarten wir folgende Verteilung der Genotypen: 25% AA; 50% AB, 25% BB. (Wikipedia: https://de.wikipedia.org/wiki/Mendelsche_Regeln)

	A	B
A	AA	AB
B	AB	BB

Das erwartete (aus dem **Modell** folgende) Farbverhältnis ist 1 : 2 : 1

Hypothesen für den Signifikanztest, p-Wert

	AA	AB	BB
”observed”	5	2	1
”expected”	2	4	2

$n = 8 ; K = 3$
kleine Stichprobe,
3 Kategorien

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

p-Wert: ”Wahrscheinlichkeit, bei Gültigkeit der Nullhypothese ein mindestens so extremes Ereignis zu erhalten wie das beobachtete”

Zur Berechnung des p-Wertes müssen wir:

1) die Wahrscheinlichkeiten der gemachten Beobachtung unter Annahme der Nullhypothese (”unter H_0”) berechnen
2) die Wahrscheinlichkeiten derjenigen Ereignisse addieren, die mindestens so extrem wie das beobachtete Ereignis sind → p-Wert

Wahrscheinlichkeit der Beobachtung "unter H_0"

Wir können annehmen, dass insgesamt n Individuen der Nachfolge-Generation beobachtet werden (Größe der Stichprobe). Die Zufallsvariable $\boldsymbol{X}$ wird nun zweckmäßigerweise als Vektor betrachet: $\boldsymbol{X} = (X_1, X_2, X_3)$

- X_1: Anzahl von AA-Genotypen in der Stichprobe (rot)
- X_2: Anzahl von AB-Genotypen in der Stichprobe (braun)
- X_3: Anzahl von BB-Genotypen in der Stichprobe (grün)

Die Zufallsvariable $\boldsymbol{X}$ kann wiederum als Ergebnis einer Folge von **identischen** und **unabhängigen** Einzelversuchen angesehen werden, wobei jetzt jedoch bei jedem Einzelversuch **drei Ereignisse** zur Auswahl stehen. Gemäß dem Modell sind die Wahrscheinlichkeiten dieser Ereignisse $p_1 = {}^1/_4$; $p_2 = {}^1/_2$ und $p_3 = {}^1/_4$

Da es im Einzelversuch nicht nur zwei ("bi-") sondern mehr ("multi-") Möglichkeiten gibt, wird die daraus resultierende statistische Verteilung **Multinomialverteilung** genannt. Die **Wahrscheinlichkeitsfunktion** der Multinomialverteiung soll im folgenden mit der Wahrscheinlichkeitsfunktion der Binomialverteiung verglichen werden.

Vergleich von Binomial- und Multinomialverteilung

Für die Binomialverteilung hatten wir folgende Wahrscheinlichkeitsfunktion:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad (1)$$

Dies kann auch folgendermaßen geschrieben werden:

$$P(X_1 = k_1, X_2 = k_2) = \frac{n!}{k_1! \cdot k_2!} \cdot p_1^{k_1} \cdot p_2^{k_2} \qquad (2)$$

Binomial

Hier gibt die Zufallsvariable X_1 die Anzahl der Erfolge an, während X_2 für die Anzahl der Misserfolge steht. Es muss gelten $k_1 + k_2 = n$ und $p_1 + p_2 = 1$.
Setzt man diese Beziehungen in Formel (2) ein, erhält man wieder Formel (1).

Formel (2) lässt sich nun leicht auf den Fall von 3 (und mehr) möglichen Ereignissen pro Einzelversuch erweitern:

$$P(k_1, k_2, k_3) = \frac{n!}{k_1! \cdot k_2! \cdot k_3!} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot p_3^{k_3}$$

Es muss gelten $\sum k_i = n$ und $\sum p_i = 1$

Multinomial

(Argument von $P(\ldots)$ abgekürzt)

Wahrscheinlichkeit der Beobachtung "unter H_0"

Wahrscheinlichkeitsfunktion der Multinomialverteiung:

$$P(k_1, k_2, k_3) = \frac{n!}{k_1! \cdot k_2! \cdot k_3!} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot p_3^{k_3}$$

Wenn H_0 (das Modell) stimmt, so gilt $p_1 = 0.25;\ p_2 = 0.5;\ p_3 = 0.25$. Die Stichprobe (Observation) ergab $k_1 = 5;\ k_2 = 2;\ k_3 = 1$ (→ Tabelle S. 17).
Die Wahrscheinlichkeit der Observation unter Annahme der Richtigkeit der Nullhypothese H_0 - Schreibweise $P(\boldsymbol{x}_{obs} \mid H_0)$ - ist dann:

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

Observation

Nullhypothese

In R steht dafür die Funktion dmultinom() zur Verfügung:

```
k = c(5,2,1)
n = 8
p = c(1/4, 1/2, 1/4)
dmultinom(k, size = n, prob = p)
# 0.01025391
```

p-Wert, Exakter Multinomialtest

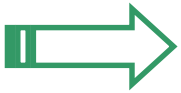
Der **p-Wert** kann berechnet werden, indem die Wahrscheinlichkeiten **aller** möglichen Ergebnisse "unter H_0" berechnet werden und dann alle Wahrscheinlichkeiten addiert werden, die kleiner (oder zumindestens nicht größer) sind als diejenige des beobachteten Ereignisses. Wir müssen also alle Tripel natürlicher Zahlen (k_1, k_2, k_3) finden, die der Bedingung $\sum k_i = n$ genügen – alle diese Ereignisse können eintreffen. Deren Anzahl beträgt:

$$M = \binom{n + K - 1}{K - 1}$$

n = Größe der Stichprobe
K = Anzahl der Kategorien

Für $n = 8$ und $K = 3$ ergibt dies $\binom{10}{2} = 45$ mögliche Ereignisse.

Hierfür kann eines kleines R-Script behilflich sein (nächste Seite). In einem doppelten Loop werden zunächst $k_1$und k_2 durchlaufen, k_3 ergibt sich dann zwangsläufig aus $\sum k_i = n$. Die Wahrscheinlichkeit für das aktuelle Ereignis $(p.x)$ wird berechnet und mit der Wahrscheinlichkeit der Observation $(p\_observed)$ verglichen (alle Wahrscheinlichkeiten unter H_0). Alle Wahrscheinlichkeiten mit $p.x \leq p\_observed$ werden aufaddiert – dies ergibt den p-Wert.



R-script, Exakter Multinomialtest

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magdeburg_Talk.R - R Editor

www.matstat.org

```
prob <- c(0.25, 0.5, 0.25)                  # nullhypothesis
xobs <- c(5, 2, 1)                          # observation
K = length(xobs)                            # number categories
n = sum(xobs)                               # sample size
p_observed = dmultinom(xobs, n, prob)       # 0.01025391
p.value = 0 ; counter = 0                                  # init
for (k1 in (n:0)) {
    for (k2 in 0:(n-k1)) {
        k3 = n - k1 - k2
        x <- c(k1, k2, k3)                  # possible outcome
        if(sum(x) != n) stop(" Not a valid outcome.\n")
        p.x = dmultinom(x, n, prob)         # prob. of outcome x under Ho
        if (p.x <= p_observed) p.value = p.value + p.x
        counter <- counter + 1
    }
}
if(counter != choose(n+K-1, K-1)) stop(" Invalid number of outcomes.\n")
p.value  # 0.07666
```

Es ergibt sich ein p-Wert von 0.077

library(EMT) → Exakter Multinomialtest

Einfacher geht es natürlich, wenn man ein fertiges R-Paket (library) benutzt:

```
library(EMT)
observed <- c(5,2,1)              # observed data
prob <- c(0.25, 0.5, 0.25)        # model: hypothetical prob.
out <- multinomial.test(observed, prob)
out$p.value  # 0.0767
```

Anzahl der möglichen Ereignisse (natürliche Zahlen mit $\sum k_i = n$):

$$M = \binom{n + K - 1}{K - 1}$$

n = Größe der Stichprobe
K = Anzahl der Kategorien

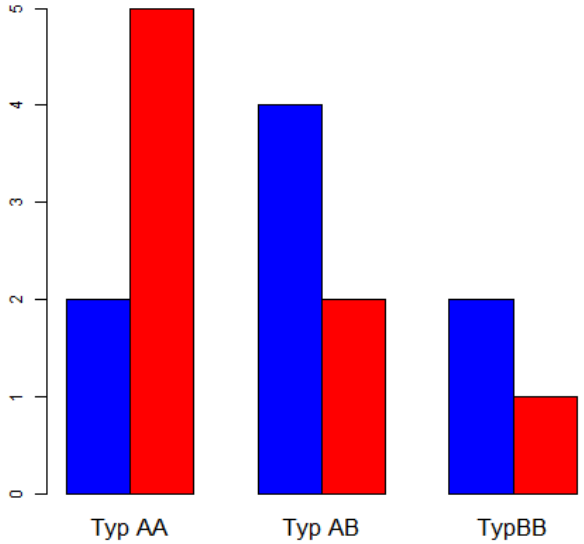
Ein Loop über alle möglichen Ereignisse kann sehr rechenaufwendig sein, besonders bei großen Stichproben und vielen Kategorien, daher soll weiter unten eine **Monte-Carlo-Methode** zur Berechnung des p-Wertes beschrieben werden.

Mendelsche Experimente

Kreuzungs-Experiment bei Blütenpflanzen

	AA	AB	BB
”observed”	5	2	1
”expected”	2	4	2

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



Der **Anpassungstest** ergab den p-Wert 0.076, d.h. es wurde keine statistische Signifikanz auf dem Signifikanzniveau 0.05 erreicht. Die Nullhypothese wird somit **nicht** verworfen und das Modell (bis auf Weiteres) akzeptiert. Dies verwundert, da die Abweichungen zwischen Modell und Beobachtung sehr groß sind. Ursache für die fehlende Signifikanz ist die zu kleine Stichprobe – die Trennschärfe (”**Power**”) des Tests ist aufgrund der kleinen Stichprobe nicht ausreichend.

Ermittlung der extremen Ereignisse mit χ^2

	AA	AB	BB
"observed"	5	2	1
"expected"	2	4	2

Oben wurde vorgeschlagen, solche Ereignisse als extremer als die Beobachtung anzusehen, die unter H_0 eine kleinere Wahrscheinlichkeit als die Beobachtung haben. Eine **alternative Methode** betrachtet diejenigen Ereignisse als extremer, die einen größeren "Abstand" zum unter H_0 erwarteten Ereignis als die Beobachtung haben. Dieser "Abstand" der beobachteten Werte von den (nach dem Modell) erwarteten Werten lässt sich mit Hilfe der Größe χ^2 quantifizieren:

$$\chi_0^2 = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} = \frac{(5-2)^2}{2} + \frac{(2-4)^2}{4} + \frac{(1-2)^2}{2} = 6$$

χ_0^2 =Abstand Modell ⇄ Beobachtung

O_k: observierte Werte in der Kategorie k

E_k: nach dem Modell erwartete Werte in der Kategorie k

In unserem Beispiel (Tabelle) beträgt der Abstand der Beobachtung vom erwarteten Ereignis genau 6. Folglich werden alle möglichen Ereignisse, deren χ^2 - Wert größer als 6 ist, als extrem betrachtet, so dass deren Wahrscheinlichkeiten zum p-Wert zusammenaddiert werden.

R-Script, Ermittlung der extremen Ereignisse mit χ^2

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magdeburg_Talk.R - R Editor

www.matstat.org

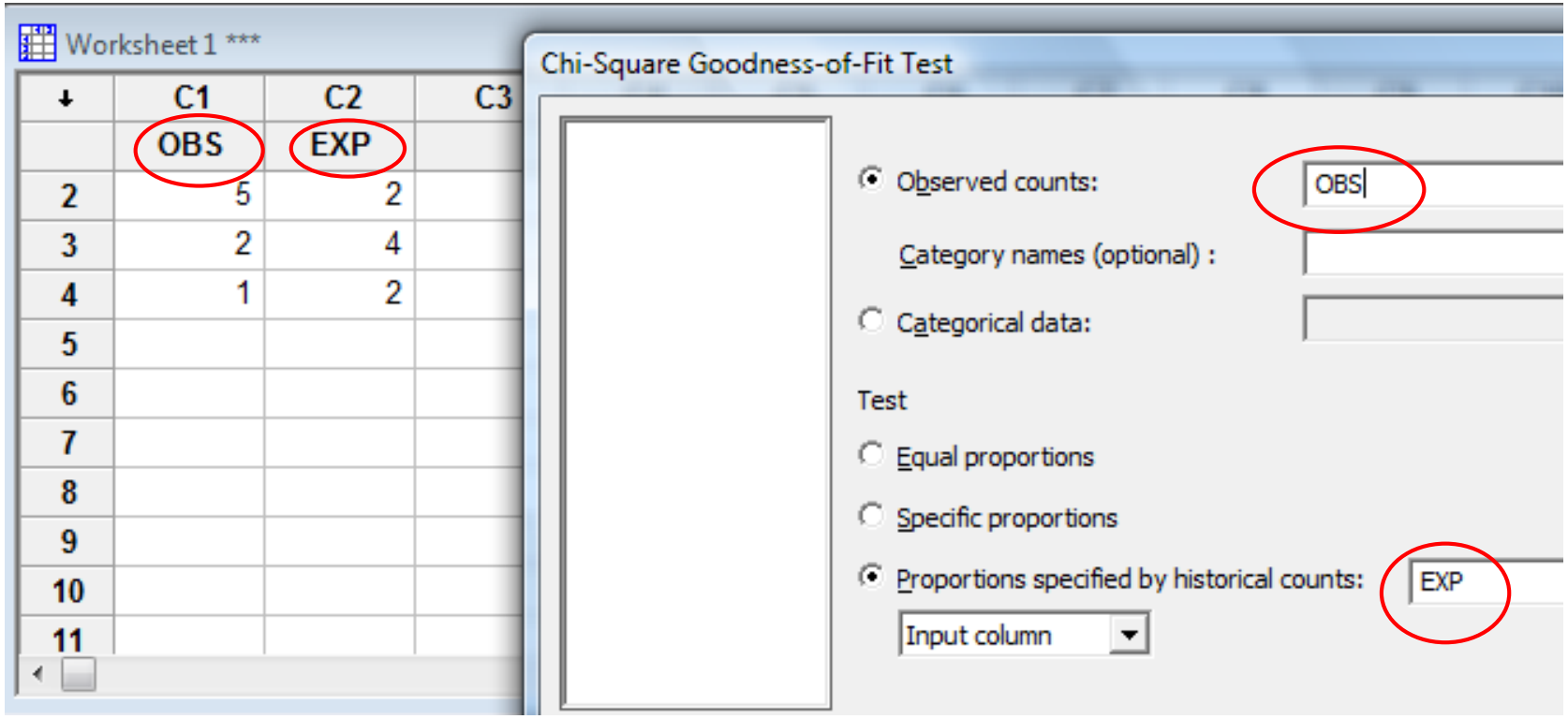
```
prob <- c(0.25, 0.5, 0.25)      # nullhypothesis
observed <- c(5, 2, 1)          # observed values
K = length(observed)            # number categories
n = sum(observed)               # sample size
expected = prob * n             # expected values
chi2_observed = sum((observed - expected)^2/expected)
p.value = 0 ; counter = 0
for (k1 in (n:0)) {
    for (k2 in 0:(n-k1)) {
        k3 = n - k1 - k2
        counter <- counter + 1
        x <- c(k1, k2, k3)                  # possible outcome
        if(sum(x) != n) stop(" Not a valid outcome.\n")
        p.x = dmultinom(x, n, prob)     # prob. of outcome x under Ho
        chisq = sum((x - expected)^2/expected)
        if (chisq >= chi2_observed) p.value = p.value + p.x
    }
}
if(counter != choose(n+K-1, K-1)) stop(" Invalid number of outcomes.\n")
p.value  # 0.05957031
```

Dies ergibt nicht genau denselben p-Wert (0.0596) . Etwas mehr dazu im Anhang.

χ^2 - Goodness-of-Fit

Minitab

Stat / Tables / Chi-square Goodness of Fit



χ^2 - Goodness-of-Fit

Minitab

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

Category	Historical Observed	Test Counts	Proportion	Contribution Expected	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

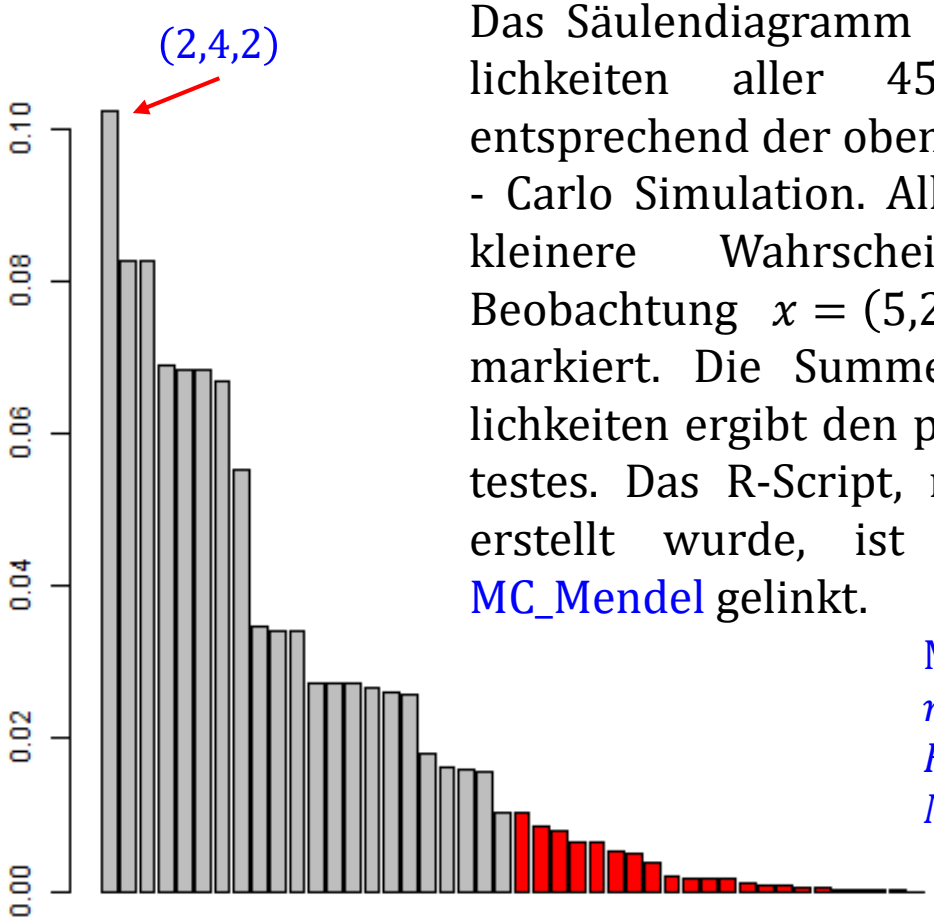
N	DF	Chi-Sq	P-Value
8	2	6	0,050

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

Berechnung des p-Wertes für den Anpassungstest mittels Monte-Carlo-Simulation

Mit Hilfe von MC können wir die Wahrscheinlichkeiten der Ereignisse unter H_0 sozusagen ”experimentell” ermitteln. Wir stellen uns eine große Urne vor, die rote/braune/grüne Kugeln im Verhältnis 1:2:1 enthält. Wenn wir nun eine Kugel ziehen, beträgt die Wahrscheinlichkeit für rot $^1/_4$, für braun $^1/_2$, und für grün $^1/_4$. Dies entspricht der **Nullhypothese**. Das Ziehen einer Kugel entspricht dem zuvor erwähnten Einzelversuch. Alle Einzelversuche müssen unabhängig sein, daher wird nach jedem Einzelversuch die gezogene Kugel wieder zurückgelegt. Es werden nun immer 8 Einzelversuche im Block durchgeführt, dies entspricht einer **Stichprobe der Größe 8**. Die Anzahlen von roten/ braunen/ grünen Kugeln ergeben also ein Zahlentripel, das sich zu 8 summiert. Es werden nun sehr viele solche 8-er Stichproben genommen und dabei jeweils notiert, wie oft jeder Zahlentripel vorkommt (was wir praktischerweise einem Computer überlassen). Die Wahrscheinlichkeit eines Zahlentripels ergibt sich dann aus dem Quotienten seiner Häufigkeit und der Gesamtanzahl der gezogenen Stichproben. Der **p-Wert** ist nun die Summe der Wahrscheinlichkeiten aller Tripel, die höchstens so oft vorkommen wie die tatsächliche Beobachtung $x = (5,2,1)$, denn diese sind ja (unter H_0) weniger wahrscheinlich als die Beobachtung und werden somit als extremer als diese angesehen.

Berechnung des p-Wertes für den Anpassungstest mittels Monte-Carlo-Simulation



Das Säulendiagramm zeigt die Wahrscheinlichkeiten aller 45 möglichen Tripel entsprechend der oben beschriebenen Monte - Carlo Simulation. Alle Ereignisse, die eine kleinere Wahrscheinlichkeit als die Beobachtung $x = (5,2,1)$ haben, sind rot markiert. Die Summe dieser Wahrscheinlichkeiten ergibt den p-Wert des Signifikanztestes. Das R-Script, mit dem diese Grafik erstellt wurde, ist unter dem Namen MC_Mendel gelinkt.

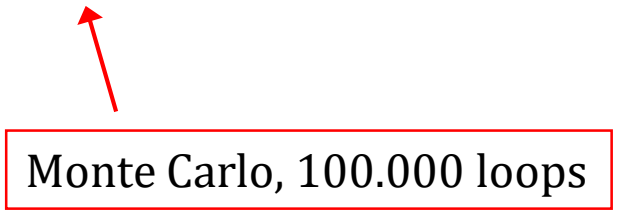
Monte Carlo
$n = 8$
$K = 3$
$N = 100000$

Uwe Menzel, 2010

Monte Carlo, SAS-Code

SAS ermittelt die extremeren Ereignisse mit der χ^2- Methode.
Hier ein Code-Snippet für eine Monte-Carlo Simulation:

```
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;
```



SAS – Resultat der Monte-Carlo Simulation

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

[Pause]

Teil 3: Unabhängigkeits-Test

Zwei Therapieformen, Untersuchung des Therapieerfolges:

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

Resultat:

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

Frage: Ist der beobachtete Unterschied in den Proportionen statistisch signifikant ?

Fishers Exakter Test auf Unabhängigkeit

Nullhypothese (H_0): $p_1 = p_2 = p$

- Die Wahrscheinlichkeit, gesund zu werden ist gleich groß für beide Therapieformen: $p = {}^2/_3$ (8 von 12, letzte Zeile)
- die beobachteten Unterschiede in den Proportionen kommen durch puren Zufall zustande (Unterschiede auf diese Stichprobe beschränkt)
- Therapieerfolg und Therapieform sind unabhängig.

Alternative Hypothese (H_a): $p_1 \neq p_2$

- Therapieerfolg und Therapieform sind abhängig.
- unterschiedliche Therapieformen ergeben unterschiedliche Erfolge
- Dieses Ergebnis will man eigentlich erreichen: beweisen, dass die neue Therapie besser ist !

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

Zur Berechnung des p-Wertes müssen wir:

1) die Wahrscheinlichkeiten der gemachten Beobachtung unter Annahme der Nullhypothese ("unter H_0") berechnen
2) die Wahrscheinlichkeiten derjenigen Ereignisse addieren, die mindestens so extrem wie das beobachtete Ereignis sind → p-Wert

Wahrscheinlichkeit der Beobachtung unter H_0

Patienten	gesund	krank	**gesamt**
Kontrollgr.	2	3	**5**
Neue Therapie	6	1	**7**
gesamt	**8**	**4**	**12**

Die Wahrscheinlichkeit des Zusandekommens dieser Tabelle "unter H_0" muss berechnet werden.

Gültigkeit der Nullhypothese bedeutet:

- die Wahrscheinlichkeit der Genesung ist $p = {}^2/_3$ (8 von 12)
- die beobachteten Proportionen kommen durch puren Zufall zustande

Die vorliegende Situation kann mit einem **Urnenmodell** verdeutlicht werden:

- Eine Urne enthält 8 weiße und 4 schwarze Kugeln: weiß steht für Genesung, schwarz für keine Genesung → das ergibt $P(weiß) = p = {}^2/_3$ (H_0)
- wir wählen 5 Kugeln **zufällig**, d.h. wir wählen 5 Patienten, die wir der Kontrollgruppe zuordnen - somit kommt das Ergebnis durch puren Zufall zustande (H_0)
- wir suchen nun die Wahrscheinlichkeit, dass unter diesen 5 zwei weiße (Gesunde) sind – dies ist die Wahrscheinlichkeit ("unter H_0"), dass sich die obenstehende Tabelle ergibt.

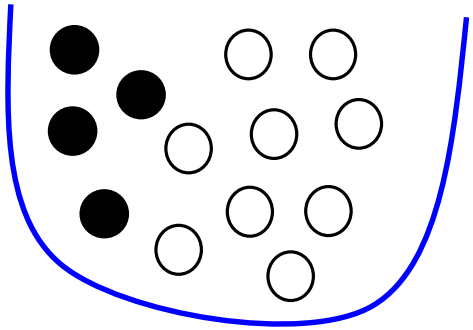
Wahrscheinlichkeit der Beobachtung unter H_0

Patienten	gesund	krank	**gesamt**
Kontrollgr.	2	3	**5**
Neue Therapie	6	1	**7**
gesamt	**8**	**4**	**12**

Anmerkung: Wenn die Zeilen- und Spaltensummen der Tabelle festgelegt sind (Gesamtzahl Gesunde/Kranke; Kontrolle/Neue) und nur eine einzige Zelle festgelegt wird, ergeben sich alle anderen Zellen zwangsläufig. Man hätte ebensogut 7 Kugeln wählen können und die Wahrscheinlichkeit berechnen, dass 6 weiße (Gesunde) darunter sind → es würde sich die gleiche Tabelle ergeben. Das bedeutet, dass die Tabelle nur einen **Freiheitsgrad** hat. Im Allgemeinen beträgt die Anzahl der Freiheitsgrade für Tabellen mit c Spalten und r Zeilen $f = (c-1) \cdot (r-1)$

Wahrscheinlichkeit der Beobachtung unter H_0

Wir suchen nun die Wahrscheinlichkeit, dass unter diesen 5 gezogenen Kugeln zwei weiße sind – dies ist die Wahrscheinlichkeit, dass sich die beobachtete Tabelle ergibt.



$w = 8$ weiße
$s = 4$ schwarze

$N = 12$ Gesamtanzahl der Kugeln
$n = 5$ Anzahl gezogene Kugeln

Zieht man 5 von 12 Kugeln (ohne Zurücklegen), so wird die Wahrscheinlichkeit, unter diesen 5 Kugeln genau 2 weiße zu finden, durch die **hypergeometrische Verteilung** beschrieben:

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

Wahrscheinlichkeit, dass unter den 5 Ausgewählten genau 2 Weiße sind.

Dies ist die Wahrscheinlichkeit der Beobachtung unter Annahme der Nullhypothese.

”Wahrscheinlichkeit der Tabelle” unter H_0

Beispiel:

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

allgemein:

	C_1	C_2	
R_1	a_{11}	a_{12}	Z_1
R_2	a_{21}	a_{22}	Z_2
	S_1	S_2	N

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

$$= P(\text{"}Tabelle\text{"})$$

$f = 1$ (ein Freiheitsgrad, schon eine einzige Zelle bestimmt alle anderen)

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

$$= \frac{S_1! \cdot S_2! \cdot Z_1! \cdot Z_2!}{N! \cdot a_{11}! \cdot a_{12}! \cdot a_{21}! \cdot a_{22}!}$$

gilt auch für größere Tabellen !

”Wahrscheinlichkeit der beobachteten Tabelle” unter der Nullhypothese

Patienten	gesund	krank	**gesamt**
Kontrollegr.	2	3	**5**
Neue Therapie	6	1	**7**
gesamt	**8**	**4**	**12**

Die Wahrscheinlichkeit des Zusandekommens dieser Tabelle ”unter H_0” muss berechnet werden.

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

hypergeometrische Verteilung

$$= \frac{8! \cdot 4! \cdot 5! \cdot 7!}{12! \cdot 2! \cdot 3! \cdot 6! \cdot 1!} = \frac{14}{99} = \underline{0.141414}$$

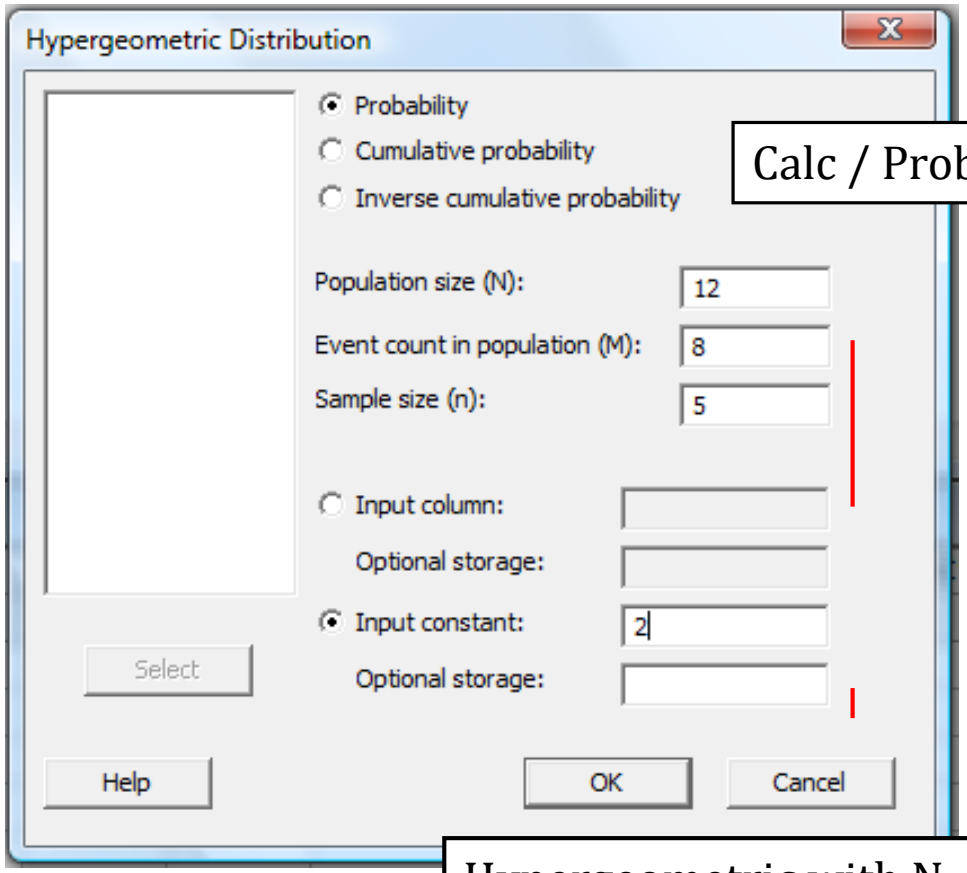
Die Wahrscheinlichkeit, dass sich diese Tabelle ergibt, ist 0.141414.

Voraussetzungen:

- festgelegte Zeilen- und Spaltensummen (festgelegte Anzahl von Patienten in beiden Gruppen, festgelegte Anzahl Gesunde/Kranke in beiden Gruppen)
- zufällige Auswahl

Hypergeometrische Verteilung

Minitab



Calc / Probability Distributions

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

x	P(X = x)
2	0,141414

Hypergeometrische Verteilung

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magdeburg_Talk.R - R Editor

```
## Hypergeometrische Verteilung

?dhyper

white_drawn = 2   # the number of white balls drawn
white_urn = 8     # the number of white balls in the urn
black_urn = 4     # the number of black balls in the urn
total_drawn = 5   # the number of balls drawn from the urn

dhyper(x = white_drawn, m = white_urn, n= black_urn, k = total_drawn)
# 0.1414141

# explicit calculation:
p = choose(8,2) * choose(4,3) / choose(12,5)
p   # 0.1414141
```

Fishers Exakter Test: p-Wert

Zur Berechnung des p-Wertes müssen wir:

1) die Wahrscheinlichkeiten der gemachten Beobachtung unter Annahme der Nullhypothese ("unter H_0") berechnen
2) die Wahrscheinlichkeiten derjenigen Ereignisse addieren, die mindestens so extrem wie das beobachtete Ereignis sind → p-Wert

Patienten	gesund	krank	**gesamt**
Kontrollegr.	a_{11}	a_{12}	**5**
Neue Therapie	a_{21}	a_{22}	**7**
gesamt	**8**	**4**	**12**

a_{12} kann am wenigsten variieren: zwischen 0 und 4 → 5 mögliche Tabellen (eine Zelle bestimmt automatisch alle anderen, da Zeilen- und Spaltensummen festliegen). Es ist also nicht sehr aufwendig, die Wahrscheinlichkeiten aller Tabellen zu berechnen, die vorkommen können →

”Wahrscheinlichkeit aller möglichen Tabellen” unter H_0

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

$a_{12} = 0$

5	0	5
3	4	7
8	4	12

$P = 0.070707$

$a_{12} = 1$

4	1	5
4	3	7
8	4	12

$P = 0.353535$

$a_{12} = 2$

3	2	5
5	2	7
8	4	12

$P = 0.424242$

$a_{12} = 3$

2	3	5
6	1	7
8	4	12

$\boldsymbol{P = 0.141414}$

Beobachtung

$a_{12} = 4$

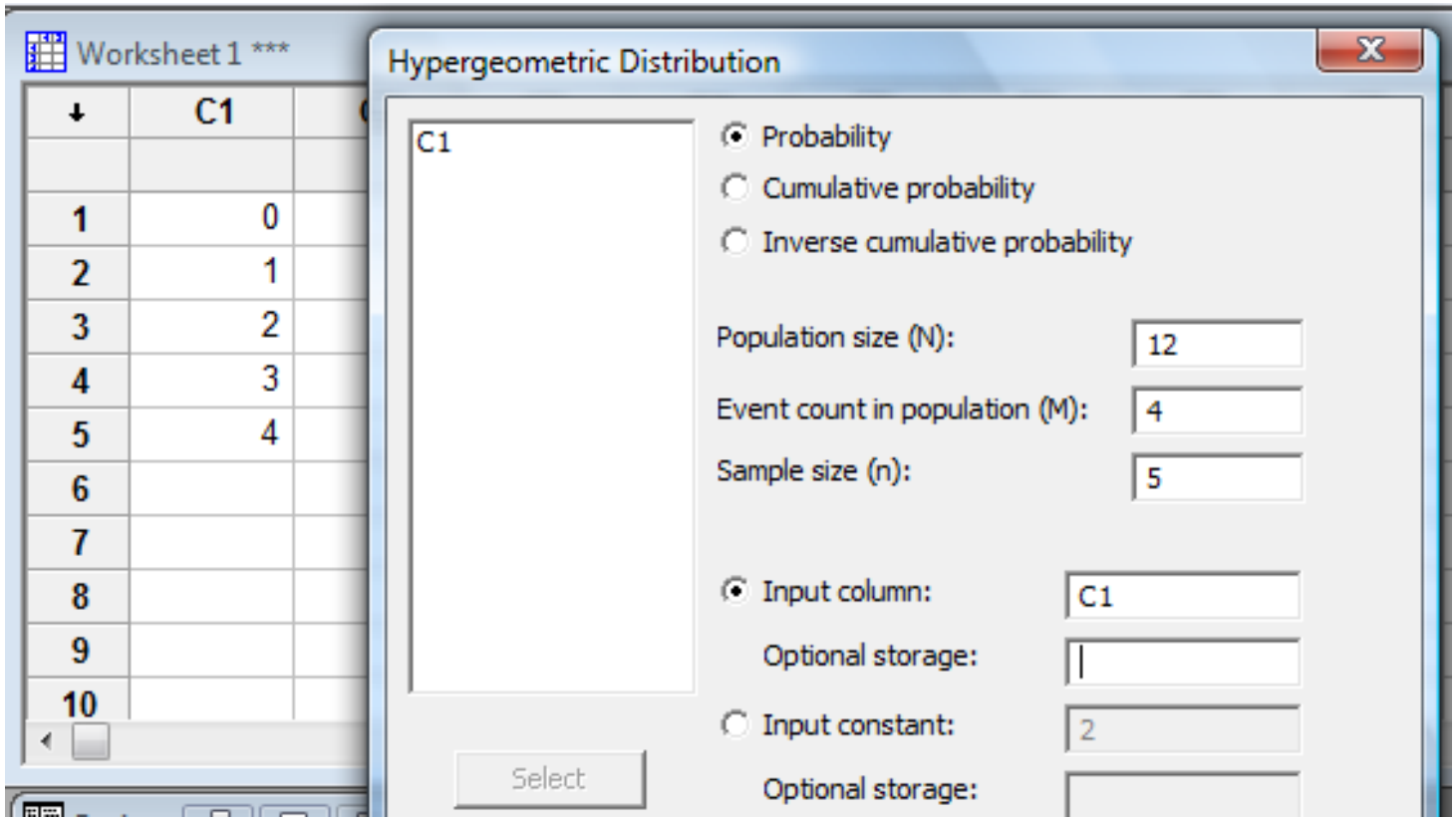
1	4	5
7	0	7
8	4	12

$P = 0.010101$

“… die Wahrscheinlichkeiten derjenigen Ereignisse addieren, die mindestens so extrem wie das beobachtete Ereignis sind”

$p = 0.1414 + 0.0707 + 0.0101$

$p \approx 0.222$



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

www.matstat.org

Minitab

Hypergeometrische Verteilung

a_{12} kann am wenigsten variieren: zwischen 0 und 4 → 5 mögliche Tabellen (eine Zelle bestimmt automatisch alle anderen, da Zeilen- und Spaltensummen festliegen). Es ist also nicht sehr aufwendig, die Wahrscheinlichkeiten aller Tabellen zu berechnen, die vorkommen können →

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magdeburg_Talk.R - R Editor

```
## Hypergeometrische Verteilung

white_urn = 8          # the number of white balls in the urn
black_urn = 4          # the number of black balls in the urn
total_drawn = 5        # the number of balls drawn from the urn

for (a12 in 0:4) {
  white_drawn = total_drawn - a12
  p = dhyper(x = white_drawn, m = white_urn, n = black_urn, k = total_drawn)
  p = round(p, 6)
  cat(paste("a12 =", a12, "  p =", p, "\n"))
}

# a12 = 0   p = 0.070707
# a12 = 1   p = 0.353535
# a12 = 2   p = 0.424242
# a12 = 3   p = 0.141414
# a12 = 4   p = 0.010101
```

Fishers Exakter Test: p-Wert

Patienten	gesund	krank	gesamt
Kontrollegr.	2	3	5
Neue Therapie	6	1	7
gesamt	8	4	12

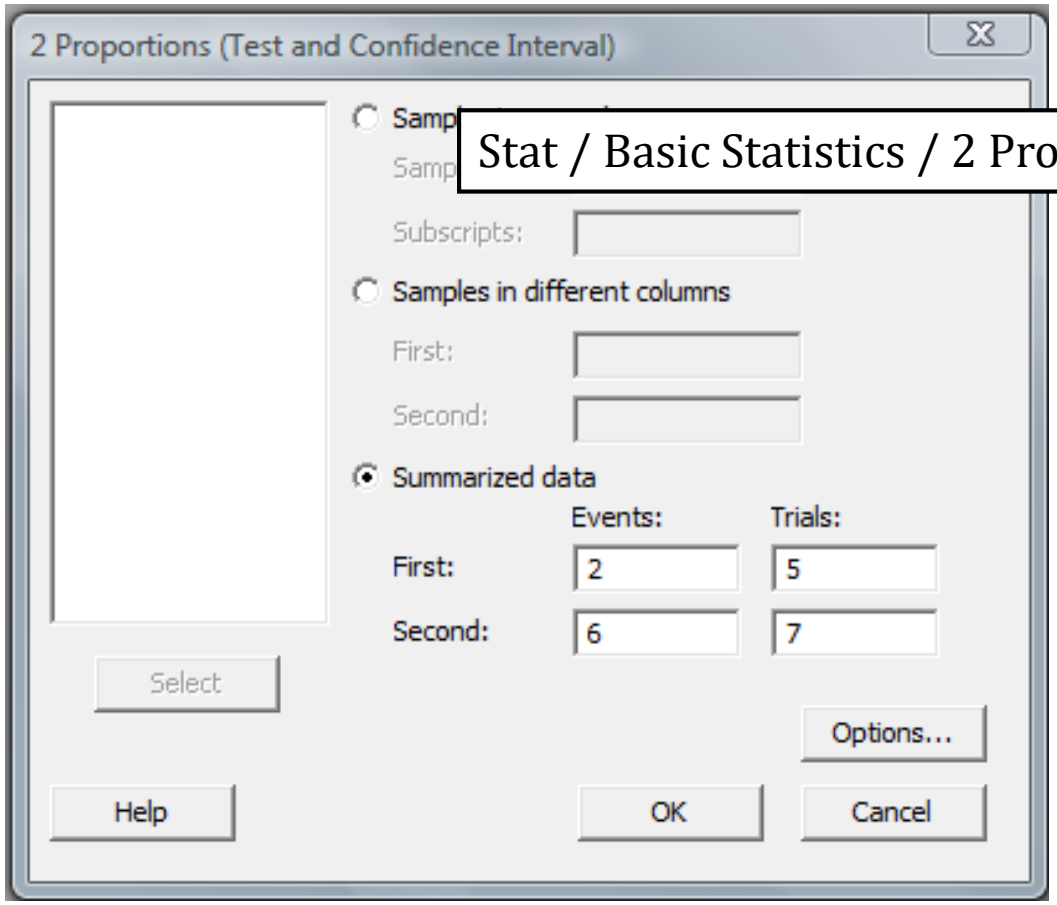
Uwe Menzel, 2010

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magdeburg_Talk.R - R Editor

```
## Fischers Exakter Test

A = matrix(c(2,3,6,1), nrow = 2, byrow = TRUE)
A
#      [,1] [,2]
# [1,]    2    3
# [2,]    6    1

fisher.test(A)

#         Fisher's Exact Test for Count Data
#
# data:  A
# p-value = 0.2222
# alternative hypothesis: true odds ratio is not equal to 1
# 95 percent confidence interval:
# | 0.001830735 2.834757137
```

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.

Vergleich zweier Therapieformen, p-Wert

Patienten	gesund	krank	**gesamt**
Kontrollegr.	2	3	**5**
Neue Therapie	6	1	**7**
gesamt	**8**	**4**	**12**

$p.value = 0.222$

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

Nullhypothese (H_0): $p_1 = p_2$ (gleiche Proportionen, gleiche Erfolgsrate)

Der p-Wert ist groß ($p > 0.05$). Die Nullhypothese kann daher **nicht** zurückgewiesen werden. Es konnte nicht gezeigt werden, dass der Unterschied in den beobachteten Proportionen statistisch sigifikant ist, d.h. wir müssen weiter davon ausgehen, dass beide Therapieformen gleich erfolgreich sind (Therapieform und Therapieerfolg sind **unabhängig**).
Dies ist in Anbetracht der beobachteten Proportionen erstaunlich. Die Ursache liegt in der schlechten Versuchsplanung – **die Stichprobe ist zu klein**.

p-Wert für eine große Stichprobe bei gleichen Proportionen

Uwe Menzel, 2010

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magdeburg_Talk.R - R Editor

```
## Fischers Exakter Test - grosse Stichprobe

A = matrix(c(200,300,600,100), nrow = 2, byrow = TRUE)
A
#      [,1] [,2]
# [1,]  200  300
# [2,]  600  100

fisher.test(A)

#         Fisher's Exact Test for Count Data

# data:  A
# p-value < 2.2e-16
# alternative hypothesis: true odds ratio is not equal to 1
# 95 percent confidence interval:
# | 0.08335038 0.14790301
```

Vergleich des p-Wertes für eine kleine und eine große Stichprobe bei gleichen Proportionen

	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

$p.value = 0.222$

$p.value < 10^{-16}$

Aufgrund der kleinen Stichprobe reicht die Trennschärfe (”Power”) des Testes nicht aus, um die Nullhypothese zu verwerfen. Dies hätte man eigentlich voraussehen können. Für dieses Beispiel hätte überhaupt nur die extremste Konfiguration einen p-Wert unter 0.05 ergeben können – für $a_{12} = 4$ (siehe nächste Seite).

Der "extremste" Fall

$a_{12} = 4$

	gesund	krank	
Kontr.	1	4	5
Ther.	7	0	7
	8	4	12

Nur diese eine Beobachtung hätte Signifikanz ergeben können:

- 20% Gesunde in Kontrollgruppe
- 100% Gesunde mit neuer Therapie

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magdeburg_Talk.R - R Editor

```
## Fischers Exakter Test - der "extremste" Fall

A = matrix(c(1,4,7,0), nrow = 2, byrow = TRUE)
A
#      [,1] [,2]
# [1,]    1    4
# [2,]    7    0

fisher.test(A)
#
#         Fisher's Exact Test for Count Data
#
# data:  A
# p-value = 0.0101
```

Ermittlung der extremen Ereignisse mit χ^2

Oben wurde vorgeschlagen, solche Ereignisse als extremer (als die Beobachtung) anzusehen, die unter H_0 eine kleinere Wahrscheinlichkeit als die Beobachtung haben. Eine **alternative Methode** betrachtet diejenigen Ereignisse als extremer, die einen größeren "Abstand" zum unter H_0 erwarteten Ereignis als die Beobachtung haben. Dieser "Abstand" der beobachteten Werte von den (nach dem Modell) erwarteten Werten lässt sich wie schon beim Anpassungstest mit Hilfe von χ^2 quantifizieren:

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

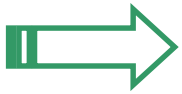
χ^2 = Abstand Modell $\rightleftarrows$ Beobachtung

O_{ij}: observierte Werte in der Zelle i, j

$E_{i,j}$: nach dem Modell erwartete Werte in der Zelle i, j

Es wird über alle Zellen der Tabelle summiert

Um die $E_{i,j}$ zu berechnen, müssen wir zunächst feststellen, welche Ergebnisse wir unter H_0 erwarten würden.



χ^2: Erwartete Frequenzen bei Unabhängigkeit

Um die $E_{i,j}$ zu berechnen, müssen wir zunächst feststellen, welche Ergebnisse wir unter H_0 erwarten würden.

Nullhypothese (H_0): $p_1 = p_2$ (gleiche Proportionen, gleiche Erfolgsrate)

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

Wenn der Anteil der Gesunden in beiden Gruppen gleich groß ist, muss gelten:

$$\frac{E_{11}}{5} = \frac{8}{12} \quad \text{und} \quad \frac{E_{21}}{7} = \frac{8}{12}$$

Dies bedeutet, dass der Anteil der gesund gewordenen sowohl in der Kontrollgruppe als auch in der Gruppe mit der neuen Therapie gleich $^2/_3$ ist.

χ^2: Erwartete Frequenzen bei Unabhängigkeit

Ebenso muss in beiden Gruppen der Anteil der Kranken bei $^1/_3$ liegen:

	gesund	krank	
Kontr.	…	E_{12}	5
Neu	…	E_{22}	7
	8	4	12

$$\frac{E_{11}}{5} = \frac{4}{12} \quad \text{und} \quad \frac{E_{22}}{7} = \frac{4}{12}$$

χ^2: Erwartete Frequenzen bei Unabhängigkeit

Bei Vorliegen von Unabhängigkeit (gleiche Proportionen) müssen also allgemein folgende Beziehungen gelten:

	gesund	krank	
Kontr.	E_{11}	E_{12}	5
Neu	E_{21}	E_{22}	7
	8	4	12

	gesund	krank	
Kontr.	E_{11}	E_{12}	Z_1
Neu	E_{21}	E_{22}	Z_2
	S_1	S_2	N

$$\frac{E_{11}}{Z_1} = \frac{S_1}{N} \qquad \frac{E_{21}}{Z_2} = \frac{S_1}{N} \qquad \frac{E_{12}}{Z_1} = \frac{S_2}{N} \qquad \frac{E_{22}}{Z_2} = \frac{S_2}{N} \qquad \text{oder:}$$

$$E_{11} = \frac{Z_1 \cdot S_1}{N} \qquad E_{21} = \frac{Z_2 \cdot S_1}{N} \qquad E_{12} = \frac{Z_1 \cdot S_2}{N} \qquad E_{22} = \frac{Z_2 \cdot S_2}{N}$$

Jedes unter H_0 erwartete Zelle ergibt sich also aus
Zeilensumme mal Spaltensumme geteilt durch Gesamtzahl:

$$E_{ij} = \frac{Z_i \cdot S_j}{N}$$

Berechnung des Abstandes der Beobachtung von dem unter H_0 erwarteten Ergebnis (χ_0^2)

Patienten	gesund	krank	gesamt
Kontrollegr.	2	3	5
Neue Therapie	6	1	7
gesamt	8	4	12

Tabelle, die beobachtet wurde

Patienten	gesund	krank	gesamt
Kontrollegr.	$^{10}/_{3}$	$^{5}/_{3}$	5
Neue Therapie	$^{14}/_{3}$	$^{7}/_{3}$	7
gesamt	8	4	12

die unter H_0 erwartete Tabelle, denn

$P(gesund) = {}^{8}/_{12} = {}^{2}/_{3}$

$P(krank) = {}^{4}/_{12} = {}^{1}/_{3}$

$$\chi_0^2 = \sum_{i,k=1}^{K} \frac{(O_{ik} - E_{ik})^2}{E_{ik}} = \frac{\left(2 - {}^{10}/_{3}\right)^2}{{}^{10}/_{3}} + \frac{\left(3 - {}^{5}/_{3}\right)^2}{{}^{5}/_{3}} + \frac{\left(6 - {}^{14}/_{3}\right)^2}{{}^{14}/_{3}} + \frac{\left(1 - {}^{7}/_{3}\right)^2}{{}^{7}/_{3}}$$

$\chi_0^2 = 2.743$ ”Abstand” Beobachtung ⇄ unter H_0 erwartete Tabelle

Berechnung des Abstandes der Beobachtung von dem unter H_0 erwarteten Ergebnis (χ_0^2)

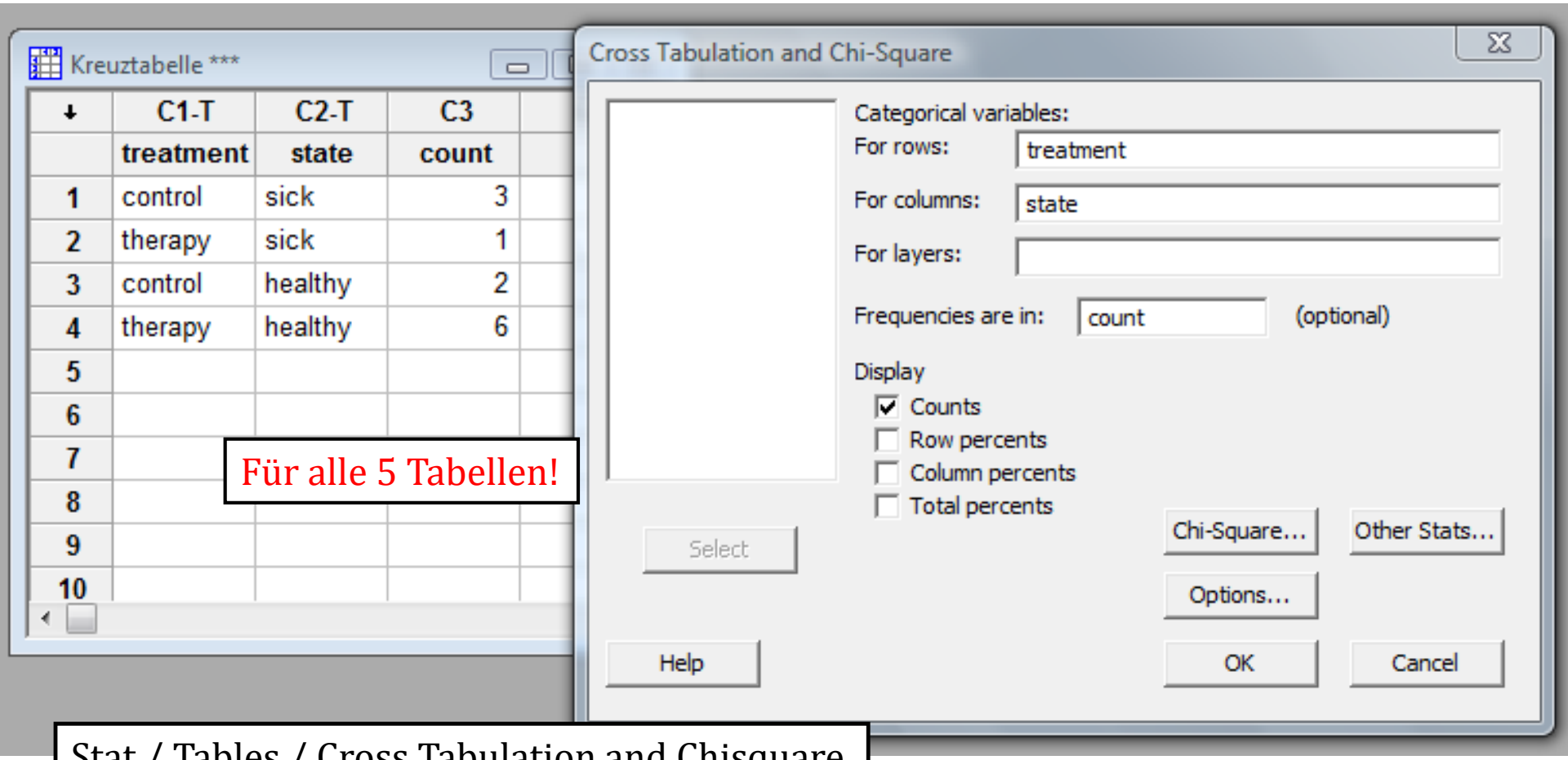
$$\chi_0^2 = \sum_{i,k=1}^{K} \frac{(O_{ik} - E_{ik})^2}{E_{ik}} = \frac{\left(2 - {}^{10}/_{3}\right)^2}{{}^{10}/_{3}} + \frac{\left(3 - {}^{5}/_{3}\right)^2}{{}^{5}/_{3}} + \frac{\left(6 - {}^{14}/_{3}\right)^2}{{}^{14}/_{3}} + \frac{\left(1 - {}^{7}/_{3}\right)^2}{{}^{7}/_{3}}$$

$$\chi_0^2 = 2.743$$

$\chi_0^2 =$ "**Abstand**" der beobachteten von der erwarteten Tabelle – es werden immer entsprechende Zellen subtrahiert. Es werden nun die χ^2 aller möglichen Tabellen berechnet und mit χ_0^2 verglichen. Ist $\chi^2 \geq \chi_0^2$ wird die zugehörige Tabelle als extrem angesehen, deren Wahrscheinlichkeit unter H_0 wird zum p-Wert dazuaddiert.

Berechnung des χ^2-"Abstandes" für alle Tabellen

Minitab



Stat / Tables / Cross Tabulation and Chisquare

Berechnung der χ^2-Statistik für alle Tabellen

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magdeburg_Talk.R - R Editor

Uwe Menzel, 2010

```
## Chi-squared for test of independence

S1 = 8   # Spalte 1
S2 = 4   # Spalte 2
Z1 = 5   # Zeile 1
Z2 = 7   # Zeile 2
N = 12   # total

E11 = Z1 * S1 / N
E12 = Z1 * S2 / N
E21 = Z2 * S1 / N
E22 = Z2 * S2 / N

E = matrix(c(E11, E12, E21, E22), nrow = 2, byrow = TRUE)

for (a12 in 0:4) {
  a11 = Z1 - a12
  a22 = S2 - a12
  a21 = S1 - a11
  A = matrix(c(a11, a12, a21, a22), nrow = 2, byrow = TRUE)
  chi2 = sum((E-A)^2/E)
  chi2 = round(chi2, 3)
  cat(paste("a12 =", a12, "  chi2 =", chi2, "\n"))
}
```

χ^2 und p für alle möglichen Tabellen

Tabelle	χ^2	p
5 | 0 | 5 3 | 4 | 7 8 | 4 | 12	4.286	0.070707
4 | 1 | 5 4 | 3 | 7 8 | 4 | 12	0.686	0.353535
3 | 2 | 5 5 | 2 | 7 8 | 4 | 12	0.171	0.424242
2 | 3 | 5 6 | 1 | 7 8 | 4 | 12	2.743	0.141414
1 | 4 | 5 7 | 0 | 7 8 | 4 | 12	8.4	0.010101

möglich sind alle Tabellen, bei denen die Zeilen- und Spaltensummen beibehalten werden

$$\chi^2 = \sum_{i,k=1}^{K} \frac{(O_{ik} - E_{ik})^2}{E_{ik}} \qquad E_{ik} = \frac{Z_i \cdot S_k}{N}$$

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

rot markiert: kleinere Wahrscheinlichkeit unter H_0 verglichen mit Beobachtung und/oder größeres χ^2 als Beobachtung. Die entsprechenden Wahrscheinlichkeiten tragen zum p-Wert bei. In diesem Fall stimmen beide Kriterien überein. Dies ist jedoch nicht immer der Fall (siehe Anhang).

$$p = 0.0707 + 0.1414 + 0.0101 \approx 0.222$$

Signifikanztests bei kleinen Stichproben

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

Anhang

Exakte statistische Verfahren bei kleinen Stichproben

Uwe Menzel, 2010

uwe.menzel@matstat.de

www.matstat.org

p-Wert-Berechnung:
Welche Ereignisse sind mindestens so extrem wie die Beobachtung ?

Variante 1: Das Ereignis x_i ist mindestens so extrem wie die Beobachtung x_{obs} wenn:

$$P(x_i \mid H_0) \leq P(x_{obs} \mid H_0)$$

Oder

Variante 2: Das Ereignis x ist mindestens so extrem wie die Beobachtung x_{obs} wenn:

$$\chi^2(x_i \mid H_0) \geq \chi^2(x_{obs} \mid H_0)$$

Welche Ereignisse sind extrem(er) ?

Beispiel: Anpassungstest, Kreuzungs-Experiment bei Blütenpflanzen

	AA	AB	BB
"observed"	5	2	1
"expected"	2	4	2

H_0 bedeutet:
$(p_1, p_2, p_3) = (1/4, 1/2, 1/4)$

Variante 1: das Ereignis x_i ist mindestens so extrem wie die Beobachtung x_{obs} wenn $P(x_i \mid H_0) \leq P(x_{obs} \mid H_0)$.

$$P(k_1, k_2, k_3 \mid H_0) = \frac{n!}{k_1! \cdot k_2! \cdot k_3!} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot p_3^{k_3}$$ Multinomialverteiung

$$P(5, 2, 1 \mid H_0) = \frac{8!}{5! \cdot 2! \cdot 1!} \cdot 0.25^5 \cdot 0.5^2 \cdot 0.25^1 = 0.01025$$ Beobachtung

$$P(4, 1, 3 \mid H_0) = \frac{8!}{4! \cdot 1! \cdot 3!} \cdot 0.25^4 \cdot 0.5^1 \cdot 0.25^3 = 0.00854$$ anderes Ereignis

⇨ Ereignis $(4, 1, 3)$ ist extremer als Beobachtung $(5, 2, 1)$

$$P(5,2,1 \mid H_0) = \frac{8!}{5! \cdot 2! \cdot 1!} \cdot 0.25^5 \cdot 0.5^2 \cdot 0.25^1 = 0.01025$$ Beobachtung

$$P(4,1,3 \mid H_0) = \frac{8!}{4! \cdot 1! \cdot 3!} \cdot 0.25^4 \cdot 0.5^1 \cdot 0.25^3 = 0.00854$$ anderes Ereignis

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magde...

```
## Anhang: zwei Kriterien

k = c(5,2,1)            # observation
n = sum(k)
p = c(1/4, 1/2, 1/4)    # null hypothesis
dmultinom(k, size = n, prob = p)
# 0.01025391

k = c(4,1,3)            # another outcome
n = sum(k)
p = c(1/4, 1/2, 1/4)    # null hypothesis
dmultinom(k, size = n, prob = p)
# 0.008544922
```

Welche Ereignisse sind extrem(er) ?

Beispiel: Anpassungstest, Kreuzungs-Experiment bei Blütenpflanzen

	AA	AB	BB
"observed"	5	2	1
"expected"	2	4	2

H_0 bedeutet:
$(E_1, E_2, E_3) = (2, 4, 2)$

Variante 2: das Ereignis x_i ist mindestens so extrem wie die Beobachtung x_{obs} wenn $\chi^2(x_i \mid H_0) \geq \chi^2(x_{obs} \mid H_0)$

$$\chi^2(5,2,1 \mid H_0) = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} = \frac{(5-2)^2}{2} + \frac{(2-4)^2}{4} + \frac{(1-2)^2}{2} = 6$$

$$\chi^2(4,1,3 \mid H_0) = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} = \frac{(4-2)^2}{2} + \frac{(1-4)^2}{4} + \frac{(3-2)^2}{2} = 4.75$$

⇨ Ereignis $(4, 1, 3)$ ist **nicht** extremer als Beobachtung $(5, 2, 1)$

$$\chi^2(5,2,1 \mid H_0) = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} = \frac{(5-2)^2}{2} + \frac{(2-4)^2}{4} + \frac{(1-2)^2}{2} = 6$$

$$\chi^2(4,1,3 \mid H_0) = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} = \frac{(4-2)^2}{2} + \frac{(1-4)^2}{4} + \frac{(3-2)^2}{2} = 4.75$$

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magde...

```
## chi-squared

O = c(5,2,1)            # observed
E = c(2,4,2)            # expected
chi2 = sum((O-E)^2/E)
chi2 # 6

O = c(4,1,3)            # outcome
E = c(2,4,2)            # expected
chi2 = sum((O-E)^2/E)
chi2 # 4.75
```

Welche Ereignisse sind mindestens so extrem wie die Beobachtung ?

Variante 1: Das Ereignis x_i ist mindestens so extrem wie die Beobachtung x_{obs} wenn:

$$P(x_i \mid H_0) \leq P(x_{obs} \mid H_0)$$

⇨ Ereignis (4, 1, 3) ist extremer als Beobachtung (5, 2, 1)

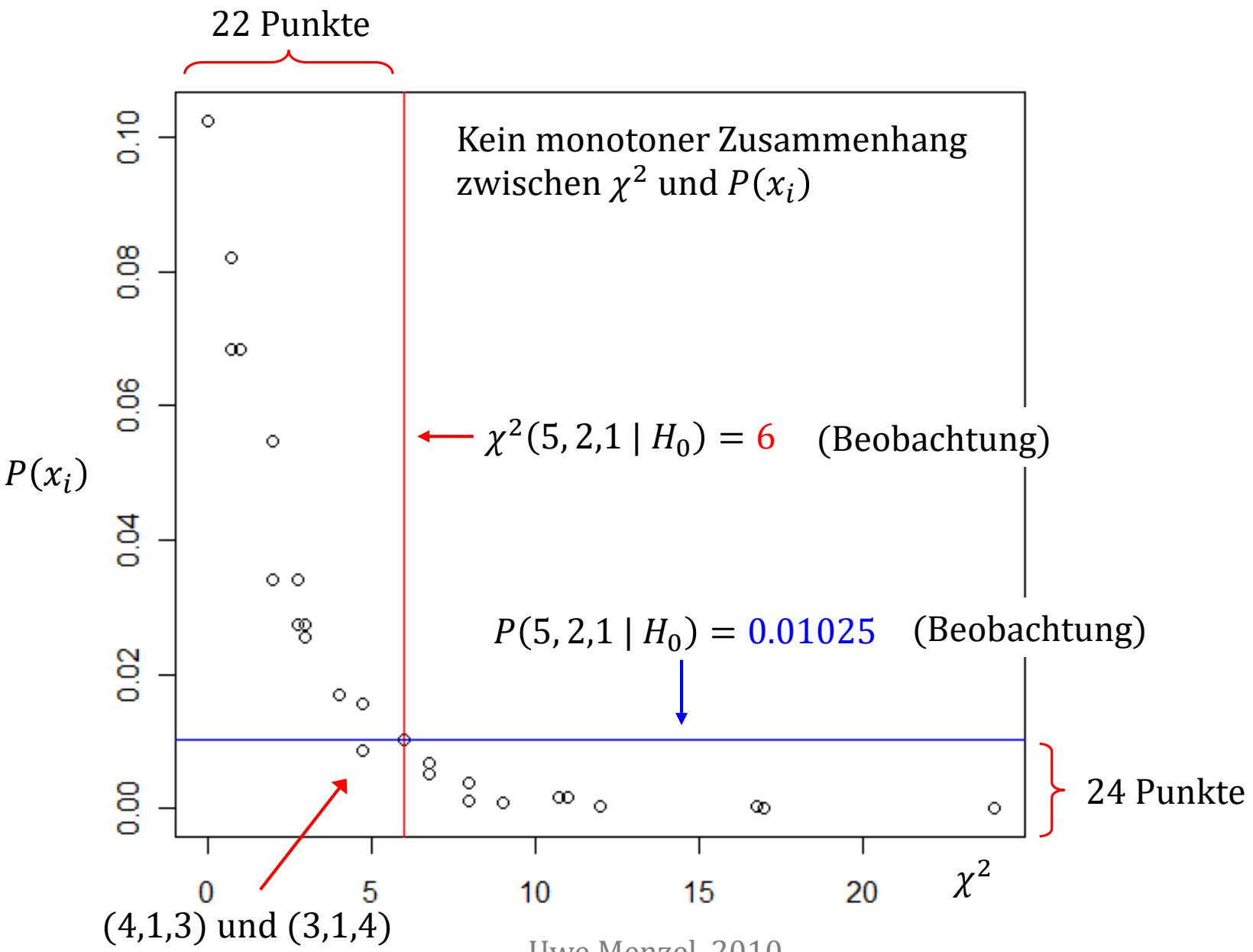
Variante 2: Das Ereignis x ist mindestens so extrem wie die Beobachtung x_{obs} wenn:

$$\chi^2(x_i \mid H_0) \geq \chi^2(x_{obs} \mid H_0)$$

⇨ Ereignis (4, 1, 3) ist **nicht** extremer als Beobachtung (5, 2, 1)

Wir werden für den p-Wert also je nach Methode unterschiedliche Ergebnisse erhalten.

p-Werte und χ^2-Werte aller möglichen Ereignisse unter H_0



Vergeich der Methoden: p-Werte

Variante 1: Das Ereignis x_i ist mindestens so extrem wie die Beobachtung x_{obs} wenn:

$$P(x_i \mid H_0) \leq P(x_{obs} \mid H_0) \quad \Rightarrow \quad p.value = 0.0767$$

Variante 2: Das Ereignis x ist mindestens so extrem wie die Beobachtung x_{obs} wenn:

$$\chi^2(x_i \mid H_0) \geq \chi^2(x_{obs} \mid H_0) \quad \Rightarrow \quad p.value = 0.0596$$

Wahrscheinlichkeit der beiden mit den Methoden unterschiedlich bewerteten Ereignisse:

$$P(4,1,3 \mid H_0) = P(3,1,4 \mid H_0) = 0.00854$$

$$0.05956 + 2 \cdot 0.00854 = 0.0767$$

→ die beiden Punkte erklären den Unterschied

$0.05956 + 2 \cdot 0.008545 = 0.0767$

→ die beiden Punkte erklären den Unterschied

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magdeburg_Talk.R - R Editor

Uwe Menzel, 2010

```
## Vergleich der Methoden:

library(EMT)
observed <- c(5,2,1)            # observed
prob <- c(0.25, 0.5, 0.25)      # null hyp.
out <- multinomial.test(observed, prob)
out$p.value  # 0.0767

observed <- c(5,2,1)            # observed
prob <- c(0.25, 0.5, 0.25)      # null hyp.
out <- multinomial.test(observed, prob, useChisq = TRUE)
out$p.value  # 0.0596

k = c(4,1,3)
n = sum(k)
p = c(1/4, 1/2, 1/4)
dmultinom(k, size = n, prob = p)
# 0.008544922

0.0596 + 2*0.008544922     # 0.07668984
```

Ein weiteres Abstands - Kriterium

$$A = \sum_i |O_i - E_i|$$

O_i - observed
E_i - expected

absolute Differenzen

$A_{obs} = |5-2| + |2-4| + |1-2| = 6$ 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 würden die beiden Punkte genauso extrem sein wie die Beobachtung, dieses Abstandsmaß würde damit den größeren der obigen p-Werte ergeben.

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;
```

χ^2 als Abstandsmaß

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
```

Große Stichproben

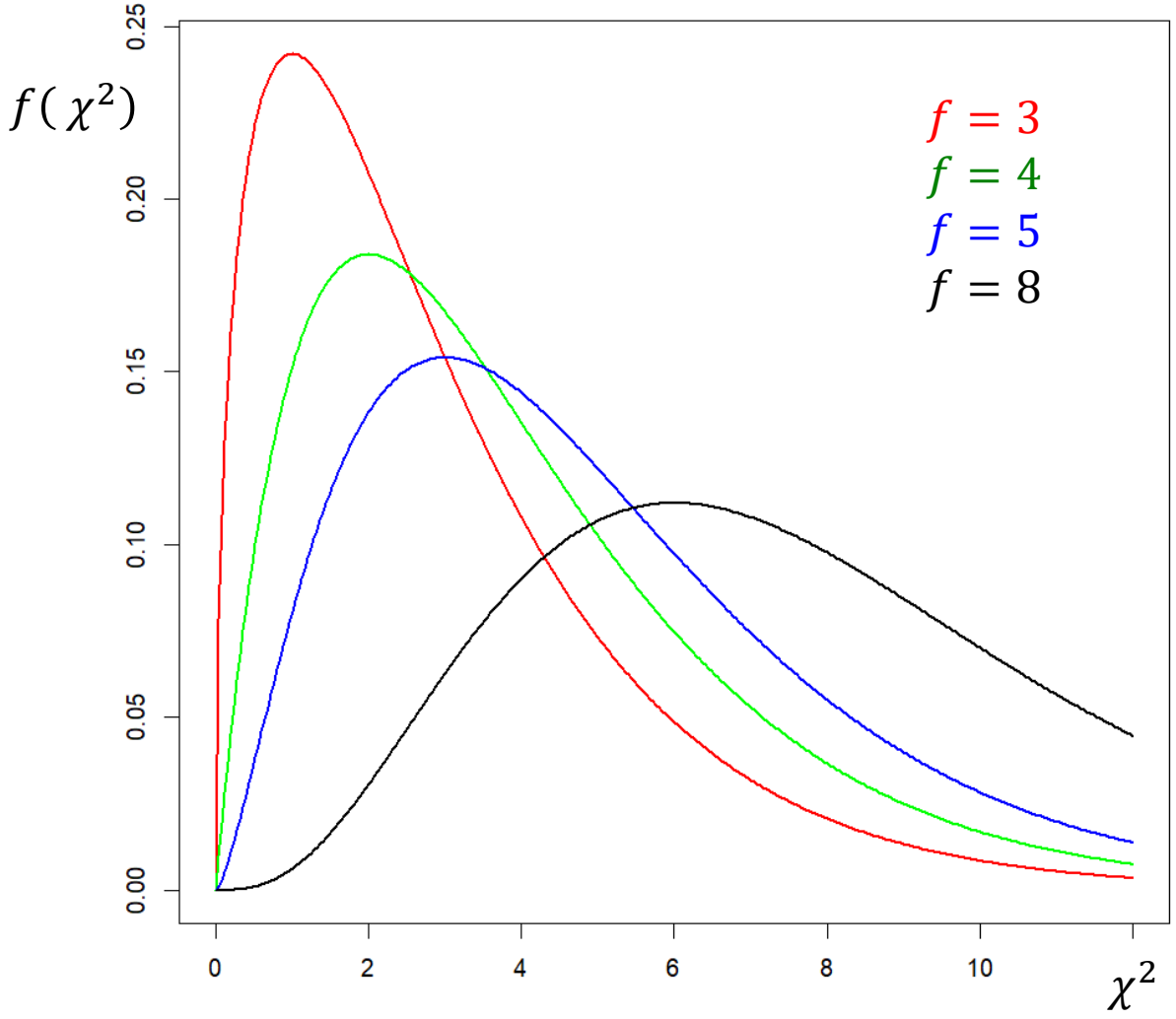
Bei großen Stichproben kann es aufwendig oder unmöglich werden, die Wahrscheinlichkeiten oder χ^2- Werte aller möglichen Ereigisse zu berechnen. Hier kann eine asymptotische Näherung helfen. Dabei wird die Verteilung der χ^2 -Werte durch eine kontinuierliche Verteilungsfunktion angenähert. Die Verteilungsfunktion heißt ebenfalls χ^2, einziges Argument dieser Funktion ist die Anzahl der Freiheitsgrade f. Als Faustregel für die Anwendbarkeit der asymptotischen Näherung gilt, dass die erwarteten Werte in jeder Kategorie oder Zelle mindestens 5 sein müssen: $\boldsymbol{E_i \geq 5}$ für alle i.

$$\chi^2 \sim \chi^2(f)$$

Statistik und deren Verteilung werden tatsächlich identisch bezeichnet!

$f =$ Anzahl der **Freiheitsgrade**

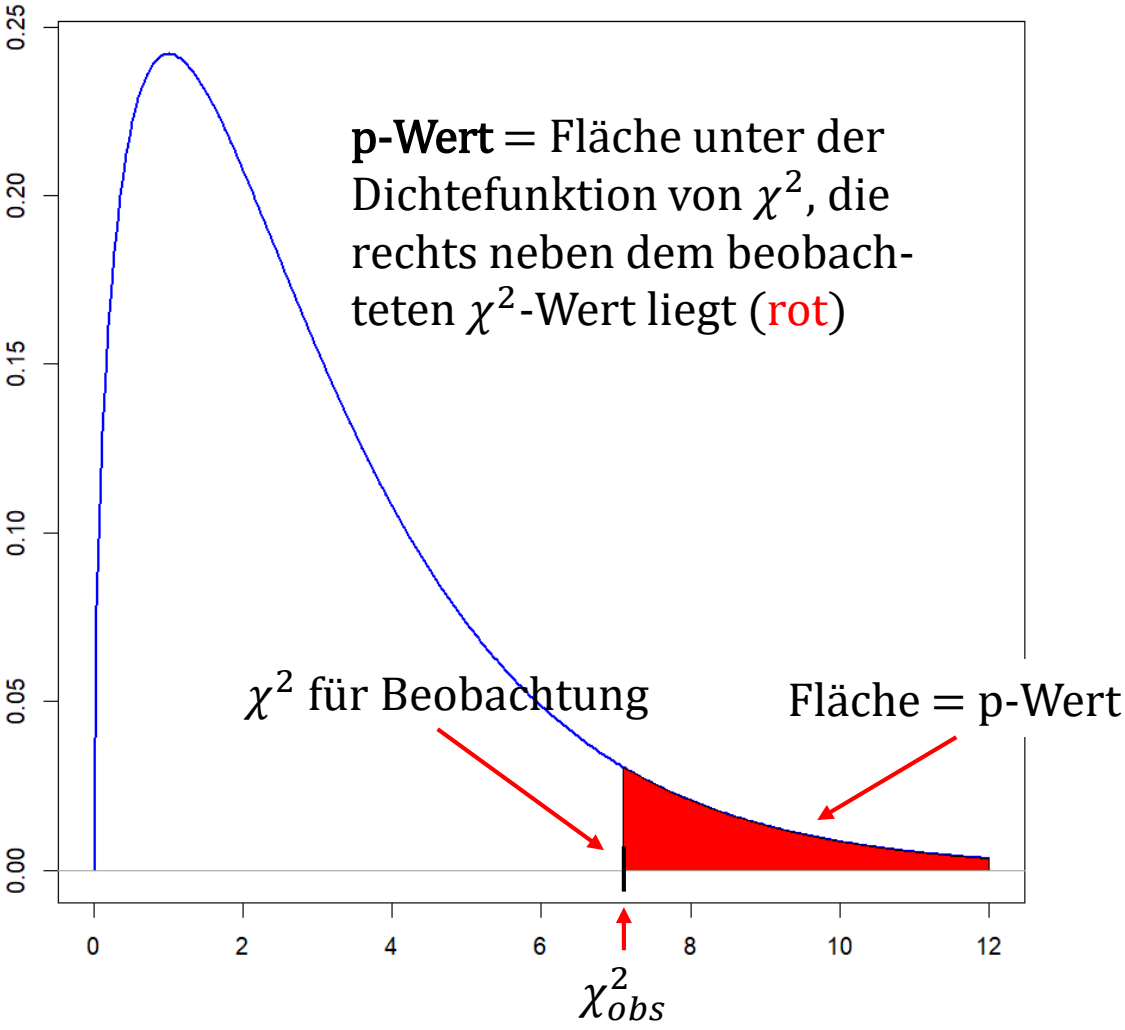
- **Anpassungstest**: $f = K - 1$
  - $K =$ Anzahl der Kategorien
- **Unabhängigkeitstest**: $f = (r-1) \cdot (c-1)$
  - r : Anzahl der Zeilen
  - c: Anzahl der Spalten


$f(\chi^2)$
0.25
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
10
χ^2
$f = 3$
$f = 4$
$f = 5$
$f = 8$

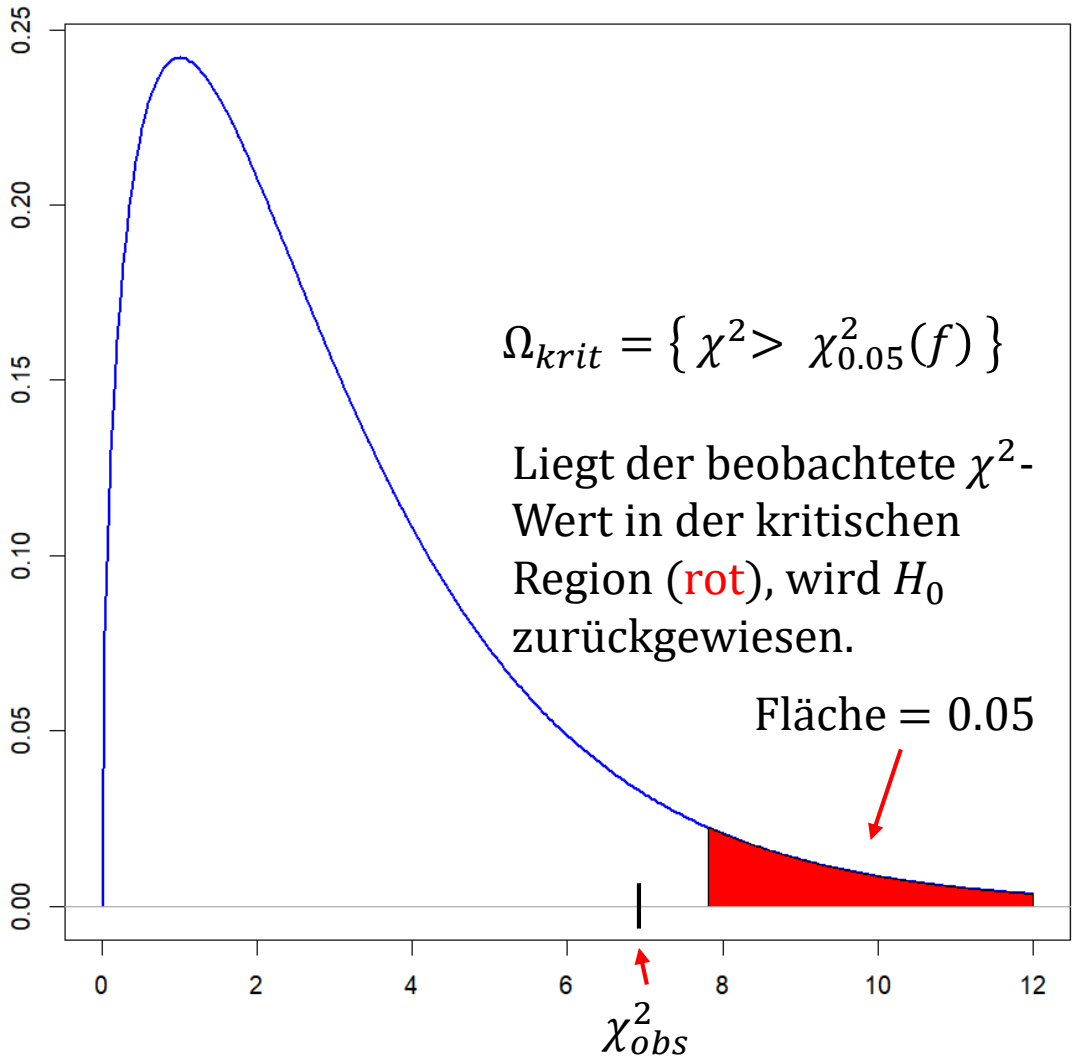
Asymptotische Näherung: p-Wert und kritische Region

Der p-Wert ist jetzt einfach die Fläche unter der Dichtefunktion von χ^2, die rechts neben dem beobachteten χ^2-Wert liegt (siehe Skizze). Ist der p-Wert kleiner als z. B. 0.05 (für ein Signifikanzniveau von 5%), wird die Nullhypothese zurückgewiesen.
Alternativ kann man im rechten Tail der Dichtefunktion eine Fläche von z.B. 0.05 abtrennen (für ein Signifikanzniveau von 5%). Der Wert auf der x-Achse, der eine Fläche dieser Größe abtrennt, ist das Quantil $\chi^2_{0.05}$. Die entstandene Fläche ist die sogenannte kritische Region. Liegt der beobachtete χ^2-Wert in dieser Region, wird die Nullhypothese zurückgewiesen.

p-Wert-Berechnung für χ^2 - Test



Kritische Region für χ^2 - Test



Uwe Menzel, 2010

Unabhängigkeitstest mit χ^2 nur für große Stichproben!

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magdeburg_Talk.R - R Editor

Uwe Menzel, 2010

Gesund	Krank
2	3
6	1

```
## Unabhängigkeitstest

A = matrix(c(2,3,6,1), nrow = 2, byrow = TRUE)
fisher.test(A)
# data:  A
# p-value = 0.2222    # korrekt

chisq.test(A)
# data:  A
# X-squared = 1.0714, df = 1, p-value = 0.3006  # inkorrekt
# Warning message:
# In chisq.test(A) : Chi-squared approximation may be incorrect

A = matrix(c(20,30,60,10), nrow = 2, byrow = TRUE)
chisq.test(A)
# data:  A
# X-squared = 25.41, df = 1, p-value = 4.635e-07  # korrekt
```

kleine Stichprobe

große Stichprobe

Anpassungstest mit χ^2 nur für große Stichproben!

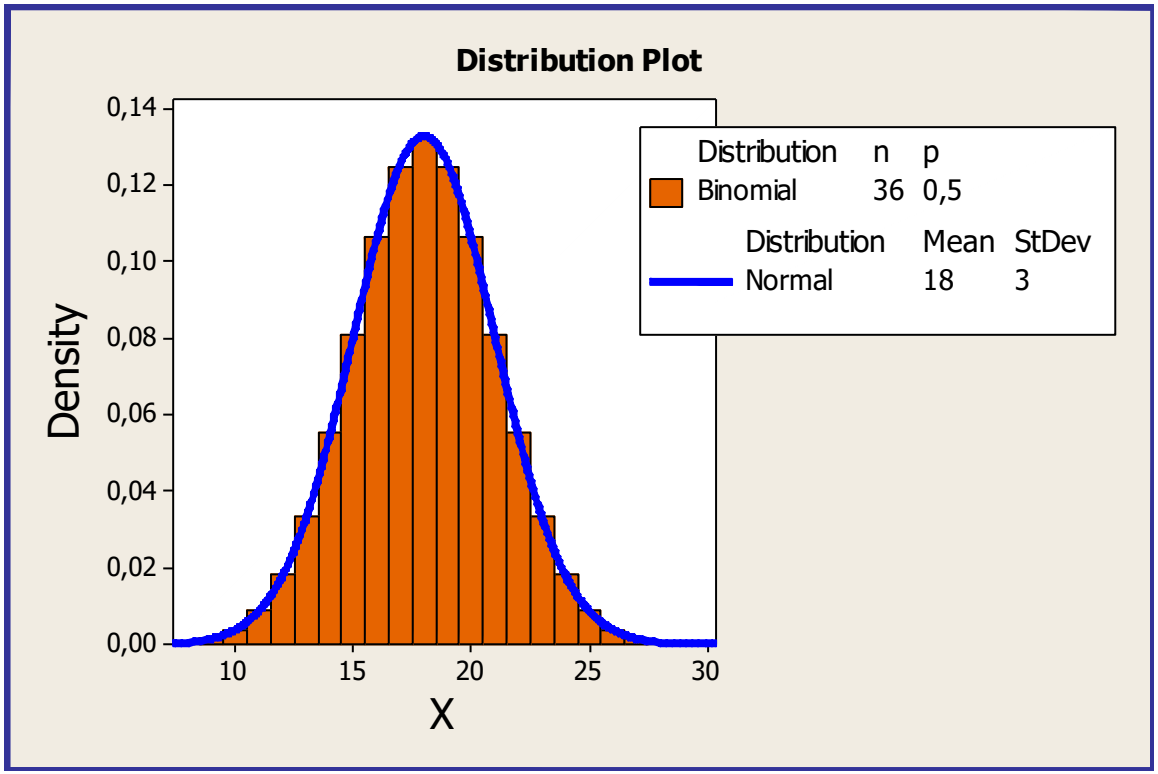
AA	AB	BB
5	2	1
2	4	2

C:\Users\Uwe\Desktop\TALKS_POSTERS\Magdeburg_Talk\R_scripts_for_Magdeburg_Talk....

```
## Anpassungstest
                    kleine Stichprobe
observed = c(5,2,1)              # observed counts
p.null = c(0.25, 0.5, 0.25)      # nullhyp.
chisq.test(observed, p = p.null)
# data:  observed
# X-squared = 6, df = 2, p-value = 0.04979
# Warning message:
# In chisq.test(observed, p = p.null) :
#   Chi-squared approximation may be incorrect

observed = c(50,20,10)    große Stichprobe
p.null = c(0.25, 0.5, 0.25)      # nullhyp.
chisq.test(observed, p = p.null)
# data:  observed
# X-squared = 60, df = 2, p-value = 9.358e-14
```

Normalapproximation für den Exakten Binomialtest

Vergleich von Binomialverteilung (orange) und Normalverteilung (blau)



$$X \sim Bin(n,p)$$

$$X \sim N\left(n\cdot p, \sqrt{n\cdot p\cdot(1-p)}\right) \quad \text{wenn} \quad V(X) = n\cdot p\cdot(1-p) > 5$$

noch besser wenn $p \cong 0.5$ (Symmetrie)

$Bin(n,p) \rightarrow N(np, \sqrt{np(1-p)})$

$$
\begin{aligned}
X &\sim Bin(n,p) \\
X &\sim N\left(n \cdot p, \sqrt{n \cdot p \cdot (1-p)}\right) \quad \text{om} \quad V(X) = n \cdot p \cdot (1-p) > 5
\end{aligned}
$$

große Varianz erforderlich

Für die Verteilungsfunktionen gilt dann:

$$F_X^{Bin}(a) \approx F_X^N(a) = \Phi\left(\frac{a - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}\right)$$

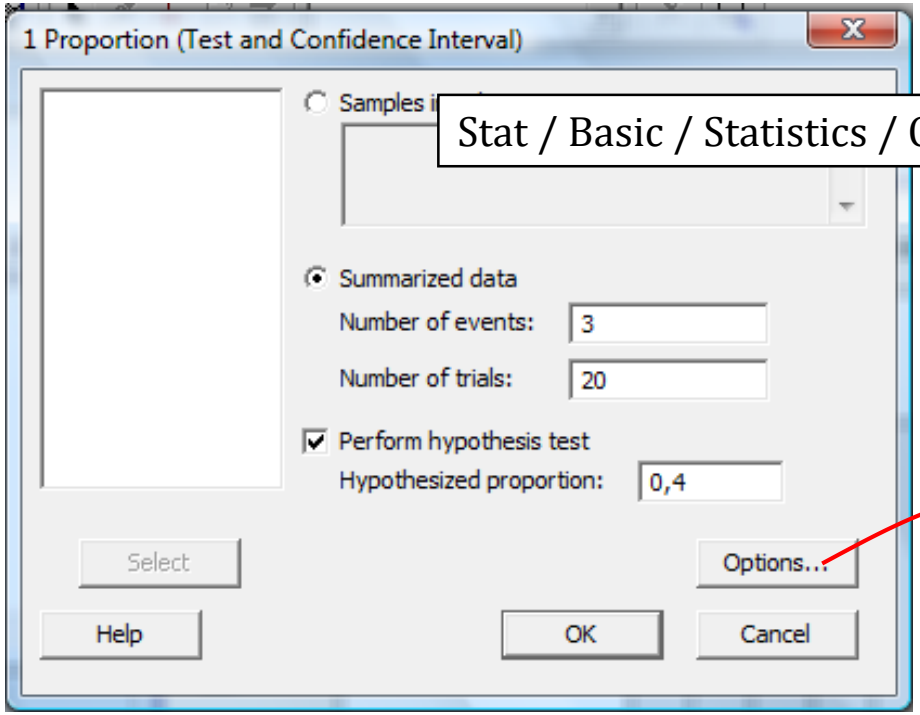
Noch besser mit Kontinuumskorrektur:

$$F_X^{Bin}(a) \approx \Phi\left(\frac{a + 0.5 - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}}\right)$$

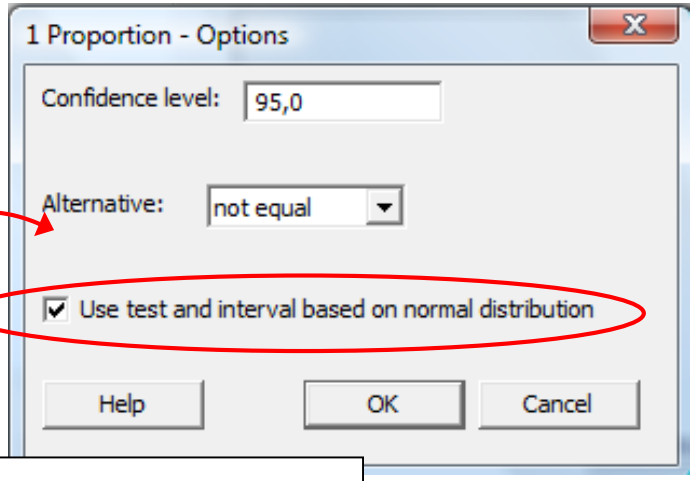
Eine Kontinuumskorektur sollte durchgeführt werden, wenn eine diskrete Verteilung durch eine kontinuierliche Verteilug angenähert wird.

Normalapproximation für den Exakten Binomialtest

Minitab



Stat / Basic / Statistics / One proportion



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.

Möglichkeiten zur Durchführung von Hypothesentests

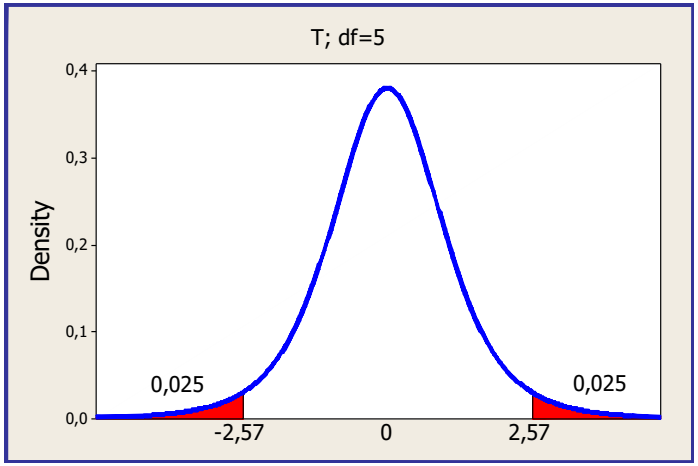
A) durch Festlegen eines kritischen Bereichs

- Verteilung der Teststatistik (”T”) unter H_0 ermitteln (Parameter: z.B. μ_0)
- **Signifikanzniveau** α festlegen ($5\% \rightarrow \alpha = 0.05$)
- kritischer Bereich = **Quantile** der Verteilung
- Stichprobe → realisierten Wert der Teststatistik (”t”) errechnen
- Verwerfen der Nullhypothese wenn t im kritischen Bereich

Beispiel:

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

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



Möglichkeiten zur Durchführung von Hypothesentests

B) durch Berechnung des p-Wertes

- Verteilung der Teststatistik (”T”) unter H_0 ermitteln (Parameter: z. B. μ_0)
- Stichprobe → $\bar{x}_{obs}$
- Berechnung des realisierten Wertes ”t” der Teststatistik
- Berechnung des p-Wertes auf Grundlage dieses Wertes
- Verwerfen der Nullhypothese wenn p-Wert klein ($p.val \leq \alpha$)

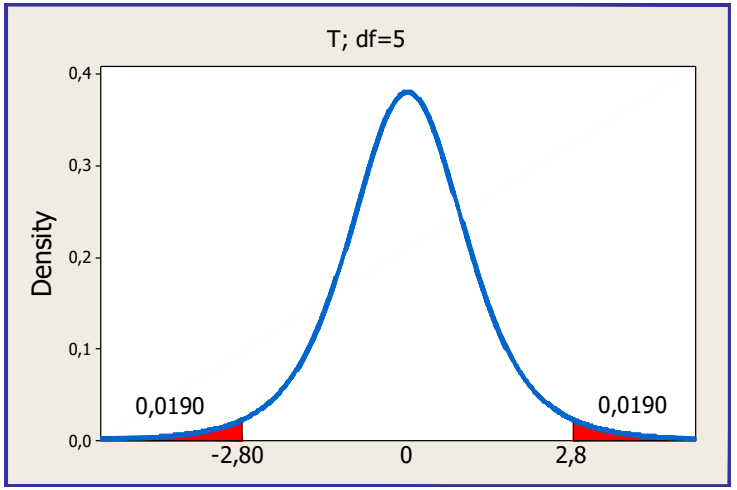
Beispiel:

$$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \; \epsilon \; t(n-1)$$

$$t = \frac{\bar{x}_{obs} - \mu_0}{s/\sqrt{n}}$$

$$p.val = P(|T| \geq t \mid H_0)$$

(für zweiseitigen Test)



Möglichkeiten zur Durchführung von Hypothesentests

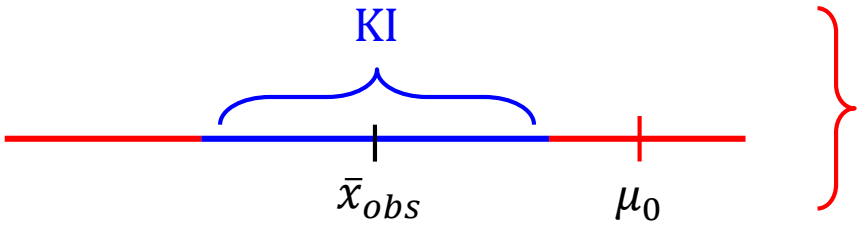
C) durch Berechnung eines Konfidenzintervalles (KI) für eine Parameterschätzung

- Nullhypothese: z. B. Parameter $\mu = \mu_0$
- Stichprobe $\bar{x}_{obs}$
- Schätzwert für Verteilungsparameter und KI berechnen
- H_0 wird verworfen wenn μ_0 außerhalb der Grenzen des KI liegt
- 95% Konfidenzintervall entspricht 5% Signifikanzniveau (α)

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

$$\Rightarrow \quad I_\mu = \bar{x} \pm t_{\alpha/2}(n-1) \cdot \frac{s}{\sqrt{n}}$$

KI für μ in $N(\mu, \sigma)$ mit Konfidenzgrad $1 - \alpha$ (σ unbekannt, kleine Stichprobe)



μ_0 außerhalb der Konfidenzgrenzen → Nullhypothese $\mu = \mu_0$ wird verworfen