

# Biostatistik, SS 2016

## Rangsummen-Test, Kontingenztafeln und Chi-Quadrat-Test

Matthias Birkner

<http://www.staff.uni-mainz.de/birkner/Biostatistik16/>

17.6.2016

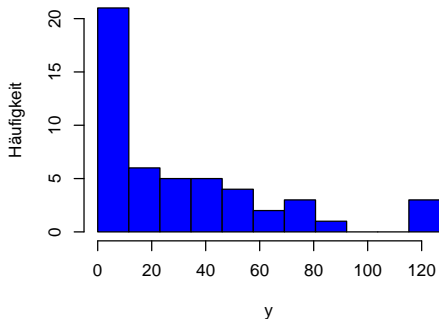
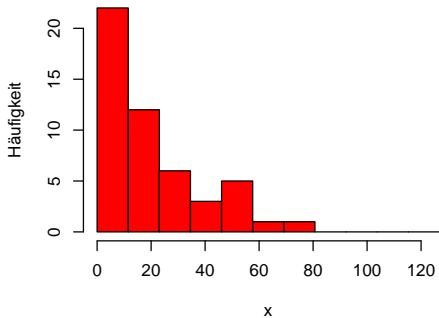


JOHANNES GUTENBERG  
UNIVERSITÄT MAINZ

Bei (ungefähr) glockenförmigen und symmetrisch verteilten Beobachtungen oder wenn die Stichprobenumfänge genügend groß sind können wir den  $t$ -Test benutzen, um die Nullhypothese  $\mu_1 = \mu_2$  zu testen: Die  $t$ -Statistik ist (annähernd) Student-verteilt.

Besonders bei sehr asymmetrischen und langschwänzigen Verteilungen kann das anders sein

Nehmen wir an, wir sollten folgende Verteilungen vergleichen:



## Beispiele

- Wartezeiten
- Ausbreitungsentfernungen
- Zelltypenhäufigkeiten

## Gesucht:

ein „verteilungsfreier“ Test,  
mit dem man die Lage zweier Verteilungen  
zueinander testen kann

Beobachtungen: Zwei Stichproben

$$X : x_1, x_2, \dots, x_m$$

$$Y : y_1, y_2, \dots, y_n$$

Wir möchten die **Nullhypothese**:  
 $X$  und  $Y$  aus derselben Population  
( $X$  und  $Y$  haben **diesselbe Verteilung**)  
testen

gegen die **Alternative**:

Die beiden Verteilungen sind gegeneinander verschoben.

Wir sind also in einer Situation, die wir schon beim  $t$ -Test getroffen haben: Die zwei Verteilungen sind möglicherweise gegeneinander verschoben (haben insbesondere möglicherweise unterschiedliche Mittelwerte), aber wir möchten *nicht* die implizite Annahme treffen, dass es sich dabei (wenigstens ungefähr) um Normalverteilungen handelt.

# Idee

Beobachtungen:

$X : x_1, x_2, \dots, x_m$

$Y : y_1, y_2, \dots, y_n$

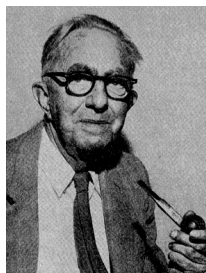
- Sortiere alle Beobachtungen der Größe nach.
- Bestimme die Ränge der  $m$   $X$ -Werte unter allen  $m + n$  Beobachtungen.
- Wenn die Nullhypothese zutrifft, sind die  $m$   $X$ -Ränge eine rein zufällige Wahl aus  $\{1, 2, \dots, m + n\}$ .
- Berechne die Summe der  $X$ -Ränge, prüfe, ob dieser Wert untypisch groß oder klein.

# Wilcoxons Rangsummenstatistik

Beobachtungen:

$X : x_1, x_2, \dots, x_m$

$Y : y_1, y_2, \dots, y_n$



Frank Wilcoxon,  
1892–1965

$U =$  Summe der  $X$ -Ränge  $- (1 + 2 + \dots + m)$   
heißt

**Wilcoxons Rangsummenstatistik**

Die Normierung ist so gewählt, dass  $0 \leq U \leq mn$ .

# Wilcoxon's Rangsummenstatistik

Bemerkung 1:

$$U = \text{Summe der } X\text{-Ränge} - (1 + 2 + \dots + m)$$

Wir könnten auch die Summe der  $Y$ -Ränge benutzen, denn

$$\begin{aligned} & \text{Summe der } X\text{-Ränge} + \text{Summe der } Y\text{-Ränge} \\ &= \text{Summe aller Ränge} \\ &= 1 + 2 + \dots + (m + n) = \frac{(m + n)(m + n + 1)}{2} \end{aligned}$$



# Wilcoxons Rangsummenstatistik

Bemerkung 2:

Der Wilcoxon-Test heißt auch Mann-Whitney-Test, die Rangsummenstatistik auch Mann-Whitney-Statistik, in der Literatur werden verschiedene Normierungskonventionen betrachtet (speziell verwenden manche Autoren nur die Summe der  $X$ -Ränge =  $U + m(m + 1)/2$  als Teststatistik).

(In der Literatur sind beide Bezeichnungen üblich, man prüfe vor Verwendung von Tabellen, etc. die verwendete Konvention.)

# Ein kleines Beispiel

- Beobachtungen:

$X$  : 1,5; 5,6; 35,2

$Y$  : 7,9; 38,1; 41,0; 56,7; 112,1; 197,4; 381,8

- Lege Beobachtungen zusammen und sortiere:

1,5; 5,6; 7,9; 35,2; 38,1; 41,0; 56,7; 112,1; 197,4; 381,8

- Bestimme Ränge:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

- Rangsummenstatistik hier:  $U = 1 + 2 + 4 - (1 + 2 + 3) = 1$

# Interpretation von $U$

$X$ -Population kleiner  $\implies U$  klein:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10  $U = 0$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10  $U = 1$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10  $U = 2$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10  $U = 2$

$X$ -Population größer  $\implies U$  groß:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10  $U = 21$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10  $U = 20$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10  $U = 19$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10  $U = 19$

# Signifikanz

Nullhypothese:  
 X-Stichprobe und Y-Stichprobe  
 stammen aus  
 derselben Verteilung

Die 3 Ränge der X-Stichprobe

1 2 3 4 5 6 7 8 9 10

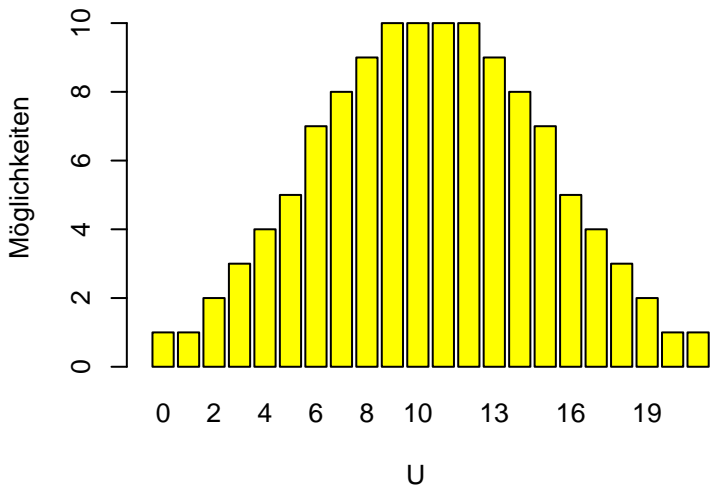
hätten genausogut irgendwelche 3 Ränge

1 2 3 4 5 6 7 8 9 10

sein können.

Es gibt  $\frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120$  Möglichkeiten.

(Allgemein:  $\frac{(m+n)(m+n-1)\dots(n+1)}{m(m-1)\dots 1} = \frac{(m+n)!}{n!m!} = \binom{m+n}{m}$  Möglichkeiten)

Verteilung der Wilcoxon-Statistik ( $m = 3, n = 7$ )

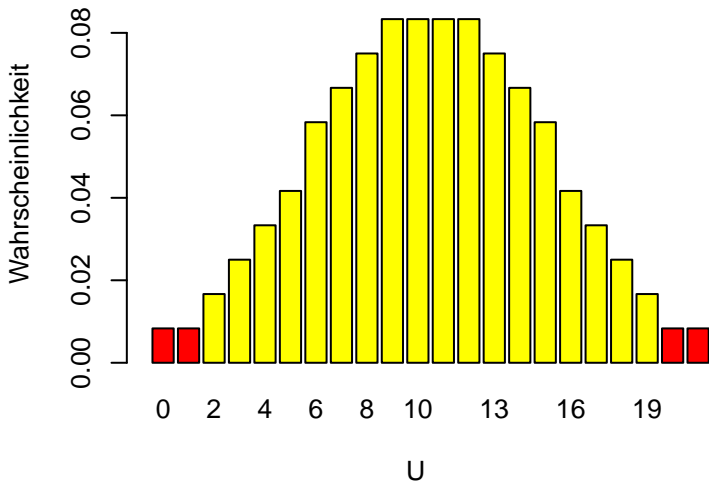
Unter der Nullhypothese sind alle Rangbelegungen gleich wahrscheinlich, also

$$\mathbb{P}(U = u) = \frac{\text{Anz. Möglichkeiten mit Rangsummenstatistik } u}{120}$$

Wir beobachten in unserem Beispiel:

1,5, 5,6; 7,9; 35,2; 38,1; 41,0; 56,7; 112,1; 197,4; 381,8  
somit  $U = 1$

$$\begin{aligned} & \mathbb{P}(U \leq 1) + \mathbb{P}(U \geq 20) \\ = & \mathbb{P}(U = 0) + \mathbb{P}(U = 1) + \mathbb{P}(U = 20) + \mathbb{P}(U = 21) \\ = & \frac{1+1+1+1}{120} \doteq 0,033 \end{aligned}$$

Verteilung der Wilcoxon-Statistik ( $m = 3, n = 7$ )

Für unser Beispiel ( $U = 1$ ) also:

$$p\text{-Wert} = \mathbb{P}(\text{ein so extremes } U) = 4/120 = 0,033$$

Wir **lehnen** die **Nullhypothese**,  
dass die Verteilungen  
von  $X$  und  $Y$   
identisch sind,  
auf dem 5%-Niveau **ab**.



# Wilcoxon-Test: Vorgehen allgemein

Wähle  $0 < \alpha < 1$  (z.B.  $\alpha = 0.05$ ), sei  $q_{U_{m,n},\alpha/2}$  das  $\alpha/2$ -Quantil,  $q_{U_{m,n},1-\alpha/2}$  das  $(1 - \alpha/2)$ -Quantil der  $(m, n)$ -Wilcoxon-Rangsummenverteilung.

Seien  $m$   $x$ -Werte und  $n$   $y$ -Werte beobachtet worden, der Wert von  $U_{m,n} =$  Summe der  $x$ -Ränge  $-\frac{m(m+1)}{2}$  sei  $u$  ( $u$  ist eine ganze Zahl zwischen 0 und  $n \cdot m$ ).

Wir lehnen die Nullhypothese

$H_0$  : „die  $x$ -Werte und die  $y$ -Werte stammen aus derselben Verteilung“

zum Signifikanzniveau  $\alpha$  ab, wenn  $u > q_{U_{m,n},1-\alpha/2}$  oder  $u < q_{U_{m,n},\alpha/2}$  gilt.

# Wilcoxon-Test: Anmerkungen

- Falls in den Daten Werte mehrfach vorkommen (sog. „Bindungen“, engl. “ties”), vergibt man geteilte Ränge (wenn beispielsweise der zweit- und der drittgrößte Wert gleich sind, bekommen beiden den Rang 2.5)

Der Test hält das geforderte Niveau dann zwar nur approximativ ein, in der Praxis allerdings meist unproblematisch, wenn es nicht zu viele Bindungen gibt.

- Einseitiger Test: Die Nullhypothese

$H_0$  : „die  $x$ -Verteilung ist gegenüber der  $y$ -Verteilung nach links verschoben“

lehnen wir zum Signifikanzniveau  $\alpha$  ab, wenn der beobachtete Wert  $u > q_{W_{m,n}, 1-\alpha}$  ist.

(analog, wenn wir eine Verschiebung nach rechts prüfen wollen, oder vertausche die Rollen von  $x$  und  $y$ )

# Wilcoxon-Test: Anmerkungen, 2

- Die Quantile von  $U_{m,n}$  entnimmt man einer Tabelle oder einem Computerprogramm, beachte:

$$q_{U_{m,n},\alpha/2} = n \cdot m - q_{U_{m,n},1-\alpha/2} \text{ wegen Symmetrie.}$$

(Für kleine  $m$  und  $n$  kann man die Verteilung auch durch Auszählen per Hand bestimmen, für den allgemeinen Fall gibt es eine Rekursionsformel;

$$\text{es gilt } \mathbb{E}_{H_0}[U_{m,n}] = \frac{mn}{2} \text{ und } \text{Var}_{H_0}[U_{m,n}] = \frac{mn(m+n+1)}{12},$$

für große  $m$  und  $n$  ist  $U_{m,n}$  unter  $H_0$  ungefähr normalverteilt.)

# Hinweis

R kennt den Wilcoxon-Test mittels `wilcox.test`:

```
> x
[1] 1.5 5.6 35.2
> y
[1] 7.9 38.1 41.0 56.7 112.1 197.4 381.8
> wilcox.test(x,y)
```

Wilcoxon rank sum test

data: x and y

$W = 1$ , p-value = 0.03333

alternative hypothesis: true location shift is  
not equal to 0

## Hinweis

Wenn der Wilcoxon-Test Signifikanz anzeigt, so kann das daran liegen, dass die zu grunde liegenden Verteilungen verschiedene Formen haben.

Der Wilcoxon-Test kann beispielsweise Signifikanz anzeigen, **selbst wenn die Stichproben-Mittelwerte übereinstimmen!**

# Vergleich von $t$ -Test und Wilcoxon-Test

Beachte:

Sowohl der (zwei-Stichproben-)  $t$ -Test als auch der Wilcoxon-Test können verwendet werden, um eine vermutete Verschiebung der Verteilung zu stützen.

Der  $t$ -Test testet „nur“ auf Gleichheit der Erwartungswerte.  
Der Wilcoxon-Test dagegen testet auf Gleichheit der gesamten Verteilungen.

**In den meisten Fällen liefern beide Tests dasselbe Ergebnis.**  
Im Allgemeinen ist für Lagetests der  $t$ -Test empfehlenswerter.

In besonderen Fällen

- Verteilungen sind asymmetrisch
- Stichprobenlänge ist klein

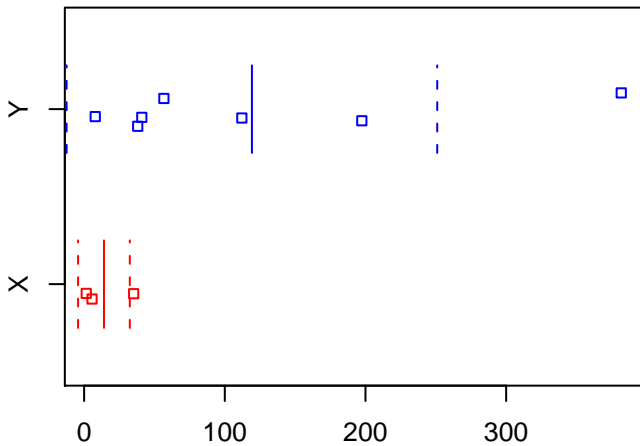
hat der Wilcoxon-Test eine höhere Testpower.

Vergleichen wir (spaßeshalber) mit dem  $t$ -Test  
(hier mit R ausgeführt):

```
> x
[1]  1.5  5.6 35.2
> y
[1]  7.9 38.1 41.0 56.7 112.1 197.4 381.8
> t.test(x,y,var.equal=TRUE)
```

Two Sample t-test

```
data:  x and y
t = -1.3319, df = 8, p-value = 0.2196
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -287.30632  76.93489
sample estimates:
mean of x mean of y
 14.1000 119.2857
```





# Mendels Erbsenexperiment

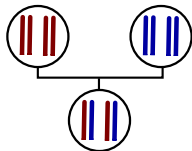
Betrachte zwei Merkmale:

Farbe: grün (rezessiv) vs. gelb (dominant)

Form: rund (dominant) vs. runzlig (rezessiv)

Beim Kreuzen von Doppelhybriden erwarten wir folgende Phänotypwahrscheinlichkeiten unter Mendelscher Segregation:

	grün	gelb
runzlig	$\frac{1}{16}$	$\frac{3}{16}$
rund	$\frac{3}{16}$	$\frac{9}{16}$



Im Experiment beobachtet ( $n = 556$  Versuche):

	grün	gelb
runzlig	32	101
rund	108	315

**Frage:**

Passen die Beobachtungen zu den theoretischen Erwartungen?

Relative Häufigkeiten:

	grün/runzlig	gelb/runzlig	grün/rund	gelb/rund
erwartet	0,0625	0,1875	0,1875	0,5625
beobachtet	0,0576	0,1942	0,1816	0,5665

bzw. in absoluten Häufigkeiten ( $n = 556$ ):

	grün/runzlig	gelb/runzlig	grün/rund	gelb/rund
erwartet	34,75	104,25	104,25	312,75
beobachtet	32	101	108	315

Können diese Abweichungen plausibel durch  
Zufallsschwankungen erklärt werden?

Wir messen die Abweichungen durch die  $\chi^2$ -Statistik:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

wobei  $E_i$  = erwartete Anzahl in Klasse  $i$  und  $O_i$  = beobachtete (engl. *observed*) Anzahl in Klasse  $i$ .

(im Beispiel durchläuft  $i$  die vier möglichen Klassen grün/runzlig, gelb/runzlig, grün/rund, gelb/rund.)

Wieso teilen wir dabei  $(O_i - E_i)^2$  durch  $E_i = \mathbb{E}O_i$ ?

Sei  $n$  die Gesamtzahl und  $p_i$  die Wahrscheinlichkeit (unter der Nullhypothese) jeder Beobachtung, zu  $O_i$  beizutragen.

Unter der Nullhypothese ist  $O_i$  binomialverteilt:

$$\mathbb{P}(O_i = k) = \binom{n}{k} p_i^k (1 - p_i)^{n-k}.$$

Also

$$\mathbb{E} O_i = n p_i, \quad \mathbb{E}(O_i - E_i)^2 = \text{Var}(O_i) = n p_i (1 - p_i)$$

und

$$\mathbb{E} \left[ \frac{(O_i - E_i)^2}{E_i} \right] = \frac{\text{Var}(O_i)}{\mathbb{E} O_i} = 1 - p_i$$

(was gar nicht von  $n$  abhängt).

Für das Erbsenbeispiel finden wir:

	gr/runz	ge/runz	gr/rund	ge/rund	Summe
theor. Ant.	0.0625	0.1875	0.1875	0.5625	
erw. ( $E$ )	34.75	104.25	104.25	312.75	556
beob. ( $O$ )	32	101	108	315	556
$O - E$	-2.75	-3.25	3.75	2.25	
$(O - E)^2$	7.56	10.56	14.06	5.06	
$\frac{(O-E)^2}{E}$	0.22	0.10	0.13	0.02	0.47

$$\chi^2 = 0.47$$

Ist ein Wert von  $\chi^2 = 0.47$  ungewöhnlich?

Die (asymptotische) Verteilung von  $\chi^2$  hängt ab von der sog. Anzahl der Freiheitsgrade **df** (eng. *degrees of freedom*), anschaulich gesprochen die Anzahl der Dimensionen, in denen man von der Erwartung abweichen kann.

In diesem Fall: Es gibt vier Klassen, die Summe der Beobachtungen muss die Gesamtzahl  $n = 556$  ergeben.

↪ wenn die ersten Zahlen 32, 101, 108 gegeben sind, ist die letzte bestimmt durch

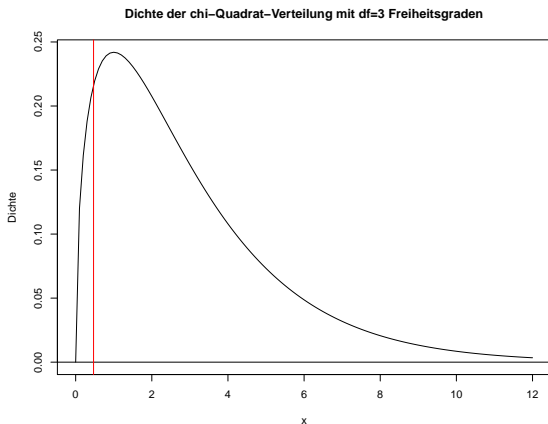
$$315 = 556 - 32 - 101 - 108.$$

$$\Rightarrow \text{df} = 3$$

### Merkregel

*Allgemein gilt beim Chi-Quadrat-Anpassungstest mit  $k$  Klassen (wenn das Modell voll spezifiziert ist, d.h. keine Parameter geschätzt werden)*

$$\text{df} = k - 1.$$



Wir hatten im Erbsenbeispiel gesehen:  $\chi^2 = 0.47$  mit  $df=3$  Freiheitsgraden.

Für eine  $\chi^2$  mit 3 Freiheitsgraden-verteilte ZV  $X$  (man schreibt oft auch  $\chi_3^2$ -verteilt) gilt

$\mathbb{P}(X \leq 0.47) \doteq 0,075$  (und somit ist der  $p$ -Wert  $\mathbb{P}(X \geq 0.47) = 0.93$ ), demnach zeigt der  $\chi^2$ -Test keine signifikante Abweichung.

Der Kuhstärling ist ein Brutparasit des Oropendola.



photo (c) by J. Oldenettel

Wir verwenden simulierte Daten in Anlehnung an



N.G. Smith (1968) The advantage of being parasitized.  
*Nature*, **219(5155)**:690-4



- Kuhstärling-Eier sehen Oropendola-Eiern meist sehr ähnlich.
- Normalerweise entfernen Oropendolas alles aus ihrem Nest, was nicht genau nach ihren Eiern aussieht.
- In einigen Gegenden sind Kuhstärling-Eier gut von Oropendola-Eiern zu unterscheiden und werden trotzdem nicht aus den Nestern entfernt.
- Wieso?
- Mögliche Erklärung: Junge Oropendolas sterben häufig am Befall durch Dasselfliegenlarven.
- Nester mit Kuhstärling-Eier sind möglicherweise besser vor Dasselfliegenlarven geschützt.

Anzahlen von Nestern, die von Dasselfliegenlarven befallen sind

Anzahl Kuhstärling-Eier	0	1	2
befallen	16	2	1
nicht befallen	2	11	16

		Anzahl Kuhstärling-Eier	0	1	2
In Prozent:	befallen		89%	15%	6%
	nicht befallen		11%	85%	94%

- Anscheinend ist der Befall mit Dasselfliegenlarven reduziert, wenn die Nester Kuhstärlingeier enthalten.
- statistisch signifikant?
- Nullhypothese: Die Wahrscheinlichkeit eines Nests, mit Dasselfliegenlarven befallen zu sein hängt nicht davon ab, ob oder wieviele Kuhstärlingeier in dem Nest liegen.

Anzahlen der von Dasselfliegenlarven befallenen Nester

Anzahl Kuhstärling-Eier	0	1	2	$\Sigma$
befallen	16	2	1	19
nicht befallen	2	11	16	29
$\Sigma$	18	13	17	48

Welche Anzahlen würden wir unter der Nullhypothese erwarten?

Das selbe Verhältnis  $19/48$  in jeder Gruppe.

Erwartete Anzahlen von Dasselfliegenlarven befallener Nester, bedingt auf die Zeilen- und Spaltensummen:

Anzahl Kuhstärling-Eier	0	1	2	$\Sigma$
befallen	7.13	5.15	6.72	19
nicht befallen	10.87	7.85	10.28	29
$\Sigma$	18	13	17	48

$$18 \cdot \frac{19}{48} = 7.13 \quad 13 \cdot \frac{19}{48} = 5.15$$

Alle anderen Werte sind nun festgelegt durch die **Summen**.

beobachtet (O, observed):	befallen	16	2	1	19
	nicht befallen	2	11	16	29
	$\Sigma$	18	13	17	48

erwartet: (E):	befallen	7.13	5.15	6.72	19
	nicht befallen	10.87	7.85	10.28	29
	$\Sigma$	18	13	17	48

O-E:	befallen	8.87	-3.15	-5.72	0
	nicht befallen	-8.87	-3.15	5.72	0
	$\Sigma$	0	0	0	0

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = 29.5$$

- Wenn die Zeilen- und Spaltensummen gegeben sind, bestimmen bereits 2 Werte in der Tabelle alle anderen Werte
- $\Rightarrow df=2$  für Kontingenztafeln mit zwei Zeilen und drei Spalten.
- Allgemein gilt für  $n$  Zeilen und  $m$  Spalten:

$$df = (n - 1) \cdot (m - 1)$$

Wir haben den Wert  $\chi^2 = 29.5$  beobachtet.

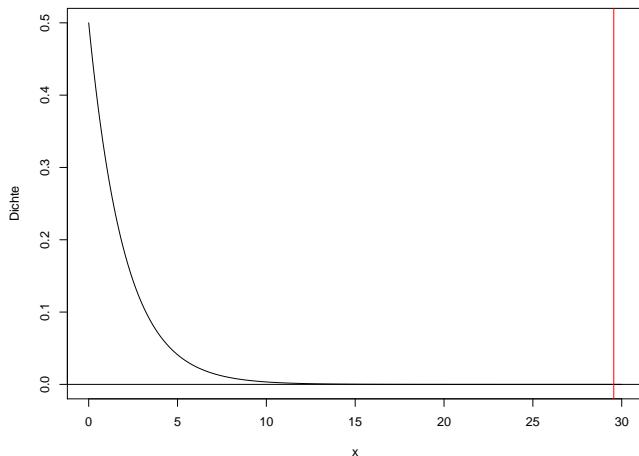
Unter der Nullhypothese „die Wahrscheinlichkeit, mit der ein Nest von Dasseliegenlarven befallen wird, hängt nicht von der Anzahl Kuhstärling-Eier ab“ ist die Teststatistik (approximativ)  $\chi^2$ -verteilt mit  $2 = (2 - 1) \cdot (3 - 1)$  Freiheitsgraden.

Das 99%-Quantil der  $\chi^2$ -Verteilung mit  $df=2$  ist 9.21 ( $<29.5$ ), wir können also die Nullhypothese zum Signifikanzniveau 1% ablehnen.

(Denn wenn die Nullhypothese zutrifft, so würden wir in weniger als 1% der Fälle einen so extremen Wert der  $\chi^2$ -Statistik beobachten.)

Faustregel: Die  $\chi^2$ -Approximation ist akzeptabel, wenn alle Erwartungswerte  $E_i \geq 5$  erfüllen, was in dem Beispiel erfüllt ist. (Siehe die folgenden Folien für die mit dem Computer bestimmten exakten  $p$ -Werte.)



Dichte der chi-Quadrat-Verteilung mit  $df=2$  Freiheitsgraden

**Bemerkung 1:** Genauere Rechnung ergibt: Für ein  $\chi_2^2$ -verteiltes  $X$  gilt  $\mathbb{P}(X \geq 29.6) = 3.74 \cdot 10^{-7}$  (was hier wörtlich der  $p$ -Wert des  $\chi^2$ -Tests auf Unabhängigkeit wäre, in dieser Genauigkeit für statistische Zwecke allerdings sinnlos ist).

Bemerkung 2: Um die Gültigkeit der  $\chi^2$ -Approximation (und der Faustregel) in diesem Beispiel einzuschätzen, könnten wir einen Computer beauftragen, durch vielfach wiederholte Simulation den  $p$ -Wert zu schätzen.

Mit **R** funktionierte das beispielsweise folgendermaßen:

```
> M <- matrix(c(16,2,2,11,1,16),nrow=2)
> M
      [,1] [,2] [,3]
[1,]   16    2    1
[2,]    2   11   16
> chisq.test(M,simulate.p.value=TRUE,B=50000)
```

```
  Pearson's Chi-squared test with simulated p-value
  (based on 50000 replicates)
```

```
data:  M
X-squared = 29.5544, df = NA, p-value = 2e-05
```

Wir sehen: Der empirisch geschätzte  $p$ -Wert  $2 \cdot 10^{-5}$  stimmt zwar nicht mit dem aus der  $\chi^2$ -Approximation überein, aber beide sind hochsignifikant klein (und in einem Bereich, in dem der exakte Wert sowieso statistisch „sinnlos“ ist). Insoweit ist die Faustregel hier bestätigt.

Gegeben sei eine Population im *Hardy-Weinberg-Gleichgewicht* und ein Gen-Locus mit zwei möglichen Allelen A und B mit Häufigkeiten  $p$  und  $1 - p$ .

↪ Genotyp-Häufigkeiten

$$\begin{array}{c|c|c} \text{AA} & \text{AB} & \text{BB} \\ \hline p^2 & 2 \cdot p \cdot (1 - p) & (1 - p)^2 \end{array}$$

Beispiel: M/N Blutgruppen; Stichprobe:  $n = 6129$  Amerikaner  
europäischer Abstammung

beobachtet:

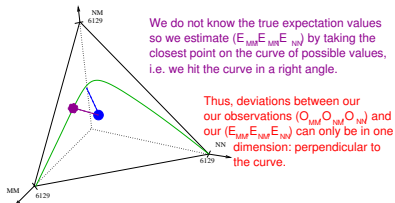
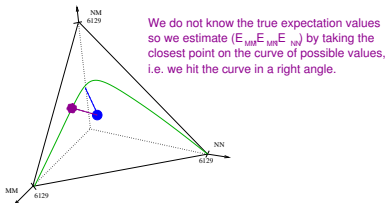
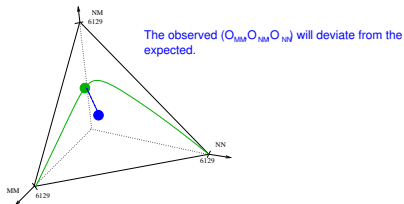
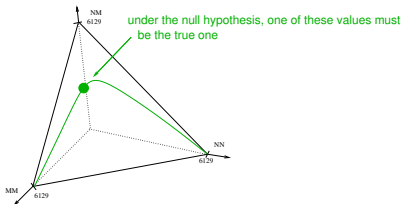
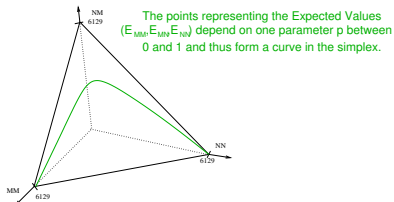
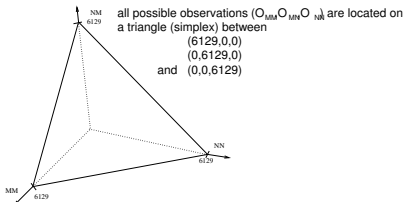
MM	MN	NN
1787	3037	1305

Geschätzte Allelhäufigkeit von M:

$$\hat{p} = \frac{2 \cdot 1787 + 3037}{2 \cdot 6129} = 0.5393$$

↪ Erwartete Werte:

MM	MN	NN	
$\hat{p}^2$	$2 \cdot \hat{p} \cdot (1 - \hat{p})$	$(1 - \hat{p})^2$	
0.291	0.497	0.212	(Anteile)
$n \cdot \hat{p}^2$	$n \cdot 2 \cdot \hat{p} \cdot (1 - \hat{p})$	$n \cdot (1 - \hat{p})^2$	
1782.7	3045.6	1300.7	(Häufigkeiten)



Thus, deviations between our observations ( $O_{MM}, O_{MN}, O_{NN}$ ) and our ( $E_{MM}, E_{MN}, E_{NN}$ ) can only be in one dimension: perpendicular to the curve.

Für die Anzahl Freiheitsgrade im  $\chi^2$ -Test mit angepassten Parametern gilt

$$df = k - 1 - m$$

$k$  = Anzahl Gruppen ( $k=3$  Genotypen)

$m$  = Anzahl Modellparameter ( $m=1$  Parameter  $p$ ) im

Blutgruppenbeispiel also:

$$df = 3 - 1 - 1 = 1$$

Der Wert der  $\chi^2$ -Statistik ist

$$\frac{(1787 - 1782.7)^2}{1782.7} + \frac{(3037 - 3045.6)^2}{3045.6} + \frac{(1305 - 1300.7)^2}{1300.7} = 0.049.$$

Dieser Wert gibt keinen Anlass, an der Nullhypothese „die Population ist bezüglich des M/N-Blutgruppensystems im HW-Gleichgewicht“ zu zweifeln: 0.049 liegt zwischen dem 10%- und dem 30%-Quantil der  $\chi^2$ -Vert. mit einem Freiheitsgrad, wir könnten also eine solche oder noch größere Abweichung zwischen Beobachtung und Erwartung in ca. 80% der Fälle erwarten (der  $p$ -Wert ist 0.83).