Statistik, WS 2019/20

Exkurs: χ^2-Test mit angepassten Parametern

Matthias Birkner

http://www.staff.uni-mainz.de/birkner/

5.2.2020

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

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

$\leadsto$ Genotyp-Häufigkeiten

AA	AB	BB
p^2	$2 \cdot p \cdot (1-p)$	$(1-p)^2$

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

beobachtet:

MM	MN	NN
1787	3037	1305

Geschätzte Allelhäufigkeit von M:

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

(Dies ist der Anteil beobachteter M-Chromosomen in der Stichprobe,

denn in der Stichprobe sind 6129 diploide Individuen, also $2 \cdot 6129$ Chromosomenkopien,

davon sind $2 \cdot 1787 + 3037$ vom Typ M.)

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

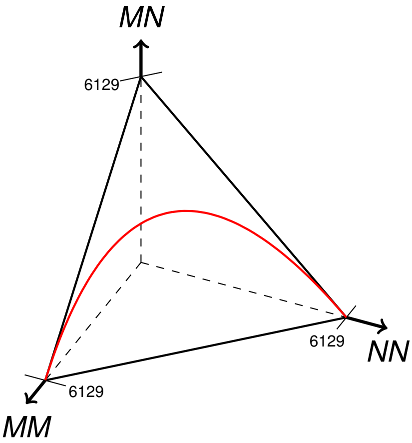
beobachtet:	MM	MN	NN
	1787	3037	1305

Geschätzte Allelhäufigkeit von M:

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

$\rightsquigarrow$ Erwartete Werte (anhand der Schätzung):

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



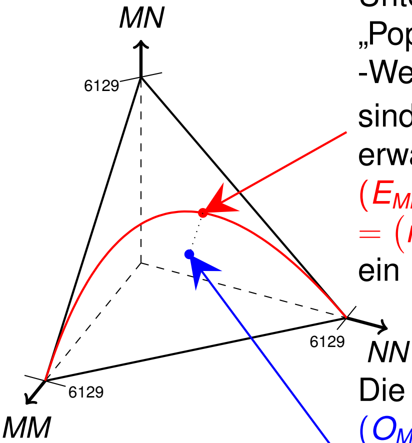
Die möglichen Beobachtungen (O_{MM}, O_{MN}, O_{NN}) liegen auf dem Dreieck („Simplex") mit Eckpunkten $(6129, 0, 0)$, $(0, 6129, 0)$ und $(0, 0, 6129)$.

Wenn die Population im Hardy-Weinberg-Gleichgewicht ist, so liegen die erwarteten Häufigkeiten

$$(E_{MM}, E_{MN}, E_{NN}) = \left(n \cdot p^2, n \cdot 2p(1-p), n \cdot (1-p)^2\right)$$

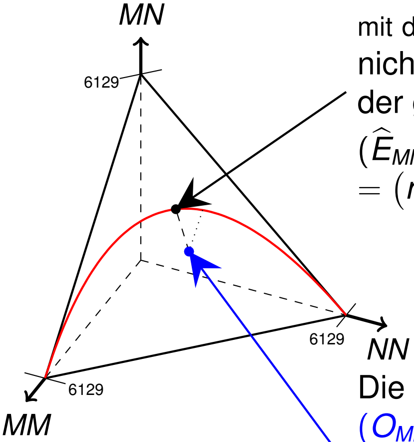
auf der roten Kurve.

(Das wahre p könnte irgend ein Wert aus $[0, 1]$ sein.)



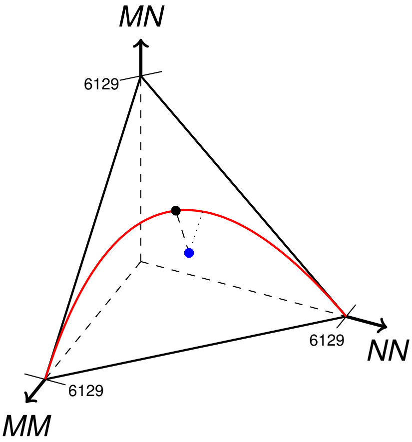
Unter der Nullhypothese „Population ist im Hardy -Weinberg-Gleichgewicht“ sind die tatsächlichen erwarteten Häufigkeiten (E_{MM}, E_{MN}, E_{NN}) $= (n \cdot p^2, n \cdot 2p(1-p), n \cdot (1-p)^2)$ ein Punkt auf der roten Kurve.

Die beobachteten Häufigkeiten (O_{MM}, O_{MN}, O_{NN}) liegen (aufgrund von Zufallsschwankungen typischerweise) *nicht* auf der roten Kurve.



Wir kennen das wahre p (und damit die wahren erwarteten Häufigkeiten) nicht und ersetzen es durch $\widehat{p}$, der geschätzte Punkt ist
$(\widehat{E}_{MM}, \widehat{E}_{MN}, \widehat{E}_{NN})$
$= \left(n\cdot\widehat{p}^2, n\cdot 2\widehat{p}(1-\widehat{p}), n\cdot(1-\widehat{p})^2\right)$

Die beobachteten Häufigkeiten (O_{MM}, O_{MN}, O_{NN}) liegen (aufgrund von Zufallsschwankungen typischerweise) *nicht* auf der roten Kurve.



Es gibt also „effektiv“ nur noch eine Richtung, in der die Beobachtungen von der Nullhypothese abweichen können:

Senkrecht zur roten Kurve.

Daher bleibt (hier)
nur 1 Freiheitsgrad übrig.

Für die Anzahl Freiheitsgrade im χ^2-Test mit angepassten Parametern gilt

$$\mathrm{df} = k - 1 - m$$

mit

$k =$ Anzahl Gruppen (hier k=3 Genotypen)
$m =$ Anzahl Modellparameter (hier m=1, der Parameter p)

im Blutgruppenbeispiel also:

$$\mathrm{df} = 3 - 1 - 1 = 1$$

Der Wert der χ^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 χ^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).

Simpson-Paradoxon

Durch Zusammenfassen von Gruppen können sich (scheinbare) statistische Trends in ihr Gegenteil verkehren.

Dieses Phänomen heißt Simpson-Paradoxon oder Yule-Simpson-Effekt.

(nach Edward H. Simpson, *1922 und George Udny Yule, 1871–1951)

Simpson-Paradoxon

Beispiel: Zulassungsstatistik der UC Berkeley 1973

Im Herbst 1973 haben sich an der Universität Berkeley 12763 Kandidaten für ein Studium beworben, davon 8442 Männer und 4321 Frauen. Es kam zu folgenden Zulassungszahlen:

	Aufgenommen	Abgelehnt
Männer	3738	4704
Frauen	1494	2827

Demnach betrug die Zulassungsquote
bei den Männern $\frac{3738}{8442} \approx 44\%$, bei den Frauen nur $\frac{1494}{4321} \approx 35\%$.

Ein χ^2-Test auf Homogenität (z.B. mit R) zeigt, dass eine solche Unverhältnismäßigkeit nur mit verschwindend kleiner Wahrscheinlichkeit durch „reinen Zufall“ entsteht:

```
> berkeley <- matrix(c(3738,1494,4704,2827),nrow=2)
> berkeley
     [,1] [,2]
[1,] 3738 4704
[2,] 1494 2827
> chisq.test(berkeley,correct=FALSE)

Pearson's Chi-squared test

data:  berkeley
X-squared = 111.2497, df = 1, p-value < 2.2e-16
```

Dieser Fall hat einiges Aufsehen erregt, s.a. P.J. Bickel, E.A. Hammel, J.W. O'Connell, Sex Bias in Graduate Admissions: Data from Berkeley, *Science*, **187**, no. 4175, 398–404 (1975).

Das Ungleichgewicht verschwindet, wenn man die Zulassungszahlen nach Departments aufspaltet:

Es stellt sich heraus, dass innerhalb der Departments die Aufnahmewahrscheinlichkeiten nicht signifikant vom Geschlecht abhängen, aber sich Frauen häufiger bei Departments mit (absolut) niedriger Aufnahmequote beworben haben als Männer – dies ist ein Beispiel für das *Simpson-Paradox*.

Die genauen nach Departments aufgeschlüsselten Bewerber- und Zulassungszahlen sind leider nicht öffentlich zugänglich (siehe aber Abb. 1 in Bickel et. al, loc. cit., für eine grafische Aufbereitung der Daten, die den Simpson-Effekt zeigt).

Bickel et. al demonstrieren das Phänomen mittels eines hypothetischen Beispiels:

	Aufgenommen	Abgelehnt
	Department of machismathics	
Männer	200	200
Frauen	100	100
	Department of social warfare	
Männer	50	100
Frauen	150	300
	Gesamt	
Männer	250	300
Frauen	250	400