ETH zürich

Kategorielle Daten

Phase 3 Studie: Wirksamer als Placebo ?

	Medikament	Placebo	Total
Geheilt	15	9	24
Nicht geheilt	10	11	21
Total	25	20	45

Grundfrage: Sind “Heilung” und “Medikamentengabe” unabhängig ?

Statistische Tests für Tabellen

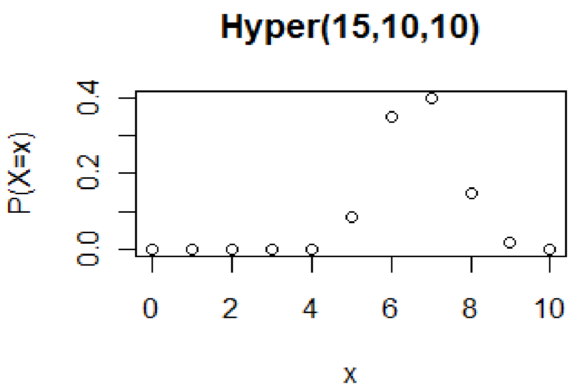
- **Fisher's Exact Test**: 2 x 2 Tabellen
  Verteilung der Teststatistik exakt

- **Chi-Quadrat Test**: m x n Tabellen
  Verteilung der Teststatistik asymptotisch bekannt

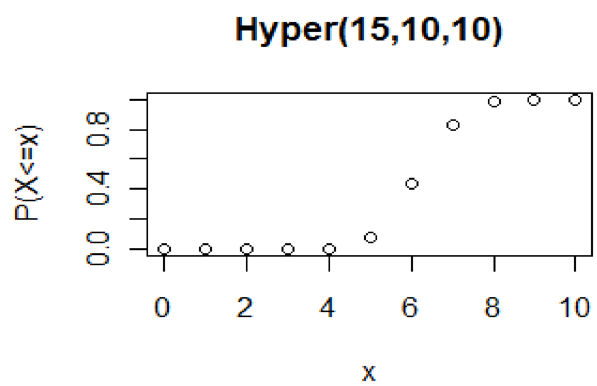
- **Logistische Regression**: 2 x m Tabellen;
  auch Mix aus mehreren kontinuierlichen und kategoriellen erklärenden Variablen möglich
  Verteilung der Teststatistik asymptotisch bekannt
  Multiple oder einfache ?

Wdh: Hypergeometrische Verteilung

- Situation: Urne mit N Kugeln; m sind markiert; ziehen n Kugeln ohne Zurücklegen; wie viele markierte Kugeln?
- ZV X: Anzahl markierter gezogener Kugeln
- $X \sim Hyper(N, n, m)$
  “X ist hypergeometrisch verteilt mit Paramtern N, n und m”
- $P(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$ ← ‘günstig’ (Zähler), ‘möglich’ (Nenner), $x \in \{0, 1, \ldots, \min(n, m)\}$
- $E(X) = \frac{n \cdot m}{N}$, $Var(X)$ kompliziert; siehe z.B. Wikipedia



pmf



cmf

Bsp: Phase 3 Studie – Wirksamer als Placebo?

- Doppel-blinde, randomisierte Studie

Markierte Bälle (m)

Gezogene und markierte Bälle

	Medikament	Placebo	Total
Geheilt	15	9	24
Nicht geheilt	10	11	21
Total	25	20	45

Gezogene Bälle (n)

Bälle in Urne (N)

- Falls Medikament keine Wirkung hat: Es gibt 24 Personen, bei denen unabhängig von der Gruppenzuteilung fest steht, dass sie gesund werden

➡ Urnenmodell

Fisher's Exact Test: Spalten und Zeilen unabhängig ?

	Medikament	Placebo	Total
Geheilt	15	9	24
Nicht geheilt	10	11	21
Total	25	20	45

- ZV X: Anzahl geheilter Patienten in Medikamenten-Gruppe
- Falls Medikament keine Wirkung hat:

$$X \sim Hyper(N = 45, m = 24, n = 25)$$

- Ist es dann plausibel in der Medikamenten-Gruppe 15 oder mehr geheilte Patienten zu beobachten?
- $P(X \geq 15) = 1 - P(X \leq 14) = 1 - 0.76 = \boxed{0.24}$ P-Wert

  R: phyper(14,24,21,25)
- Falls das Medikament nicht wirkt, ist es durchaus plausibel 15 oder mehr geheilte Patienten in der Medikamentengruppe zu beobachten

Wdh: Odds und Odds-Ratio

	Medikament	Placebo	Total
Geheilt	15	9	24
Nicht geheilt	10	11	21
Total	25	20	45

Alternative Beschreibung von Wahrscheinlichkeiten
- Odds(A) = P(A) / (1 – P(A))
- P(A) = Odds(A) / (1 + Odds(A))
- Log-Odds(A) = log(Odds(A))

Wirksamkeit von Medikament kann mit Odds-ratio ausgedrückt werden

Odds(Geheilt) = (24/45) / (21/45) = 24/21 = 1.14
Odds(Geheilt mit Medi) = (15/25) / (10/25) = 15/10 = 1.5
Odds(Geheilt ohne Medi) = (9/20) / (11/20) = 9/11 = 0.82
Odds-ratio: Odds(Geheilt mit Medi) / Odds(Geheilt ohne Medi) = 1.5 / 0.82 = 1.83

Fisher's Exact Test in R

```
> m  <- matrix(c(15,10,9,11),2,2)
> m
     [,1] [,2]
[1,]   15    9
[2,]   10   11
> fisher.test(m, alternative = "greater")

 Fisher's Exact Test for Count Data

data:  m
p-value = 0.2416
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 0.5753718       Inf
sample estimates:
odds ratio
  1.808415
```

Einseitige Alternative:
Grössere Macht ein wirksames Medikament zu finden
Blind für unwirksame Medikamente

P-Wert

Einseitiges 95%-Vertrauensintervall für Odds-ratio

Odds-ratio

R verwendet spezielle numerische Methoden um das Odds ratio zu bestimmen; es kann daher leichte Unstimmigkeiten zur Berechnung von Hand geben

Statistische Tests für Tabellen

- **Fisher's Exact Test**: 2 x 2 Tabellen
  Verteilung der Teststatistik exakt

- **Chi-Quadrat Test**: m x n Tabellen
  Verteilung der Teststatistik asymptotisch

- **Logistische Regression**: 2 x m Tabellen;
  auch Mix aus mehreren kontinuierlichen und kategoriellen erklärenden Variablen möglich
  Verteilung der Teststatistik asymptotisch
  Multiple oder einfache ?

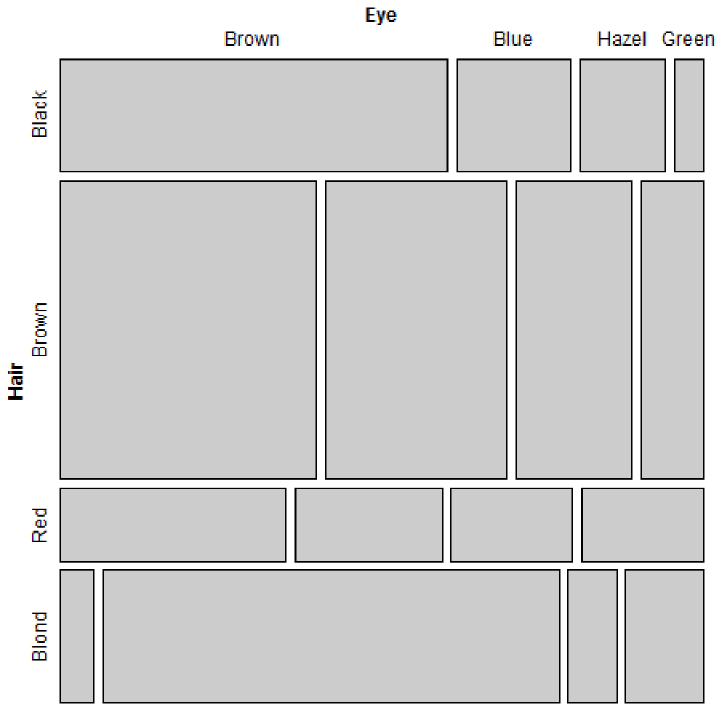
Chi-Quadrat Test: Spalten und Zeilen unabhängig?

- Haar- und Augenfarbe (R: ?HairEyeColor)

Hair / Eye	Brown	Blue	Hazel	Green	Total
Black	68	20	15	5	108
Brown	119	84	54	29	286
Red	26	17	14	14	71
Blond	7	94	10	16	127
Total	220	215	93	64	592

- Mögliche Fragen:
  - Visualisierung (v.a. wenn mehr als 2 Kategorien)
  - Abhängigkeit? Wo?

Visualisierung kategorischer Daten: Mosaic Plot



Hair / Eye	Brown	Blue	Hazel	Green	Total
Black	68	20	15	5	108
Brown	119	84	54	29	286
Red	26	17	14	14	71
Blond	7	94	10	16	127

Fläche proportional zu Tabelleneintrag

"observed values"
$O_{ij} = N_{ij}$

Chi-Quadrat Test

	A=1	…	A=n	Total
B=1	N_{11}		N_{1n}	N_{1*}
…				
B=m	N_{m1}		N_{mn}	N_{m*}
Total	N_{*1}		N_{*n}	N

H_0: A, B sind unabhängig

$$P(A = i \cap B = j) = P(A = i) * P(B = j) \approx \hat{P}(A = i) * \hat{P}(B = j) = \frac{N_{*i}}{N} * \frac{N_{j*}}{N}$$

Erwartungswert der Zelle falls H_0 stimmt: $E_{ij} = N * \frac{N_{*i}}{N} * \frac{N_{j*}}{N} = \frac{N_{*i} N_{j*}}{N}$

Chi-Quadrat Test

	A=1	…	A=n	Total
B=1	N_{11}		N_{1n}	N_{1*}
…				
B=m	N_{m1}		N_{mn}	N_{m*}
Total	N_{*1}		N_{*n}	N

Wie verschieden sind beobachtete und erwartete Werte?
Verbreitet: **Pearson Chi-Quadrat Statistik**

$$X^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i=1}^{n} \sum_{j=1}^{m} R_{ij}^2$$

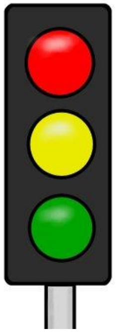
Pearson Residuen

$$R_{ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}}}$$

Beitrag jeder Zelle zur Modellabweichung

Falls H_0 stimmt, folgt X^2 einer Chi-Quadrat Verteilung mit (n-1)(m-1) Freiheitsgraden (*falls N gross – s. nächste slide*).
→ p-Werte

Chi-Quadrat Test: Wann ist Approximation gut ?



- Faustregel:
  $E_{ij} < 1$ für mind. ein Tabellenfeld → ungenügend
  $E_{ij} > 1$ für alle Tabellenfelder i,j → gerade noch OK
  $E_{ij} > 5$ für alle Tabellenfelder i,j → sehr gut

- Falls Faustregel nicht erfüllt:
  - Kategorien zusammenfassen
  - anderen Test verwenden (z.B. Fisher-Test mit Simulation)

Chi-Quadrat Test in R

```
> chisq.test(tab)

 Pearson's Chi-squared test

data:  tab
X-squared = 138.2898, df = 9, p-value < 2.2e-16
```

$$1 - pchisq(q = 138.2898, df = 9)$$

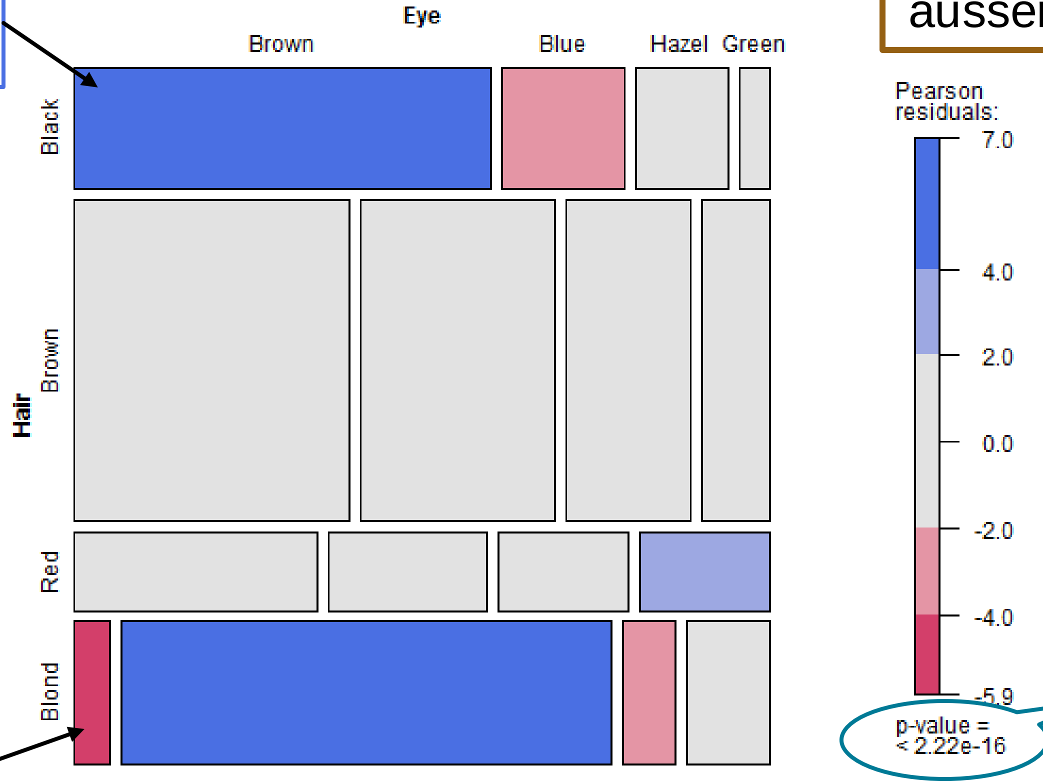
$$\sum_{i=1}^{n} \sum_{j=1}^{m} R_{ij}^2$$

$$(4-1) * (4-1) = 9$$

Mosaic plot mit Shading: Integrierter Chi-Quadrat Test

Sehr grosser Tabelleneintrag

Farbe falls Pearson Residuen ausserhalb [-2, 2]



Sehr kleiner Tabelleneintrag

p-Wert des Chi-Quadrat Tests

ETH zürich

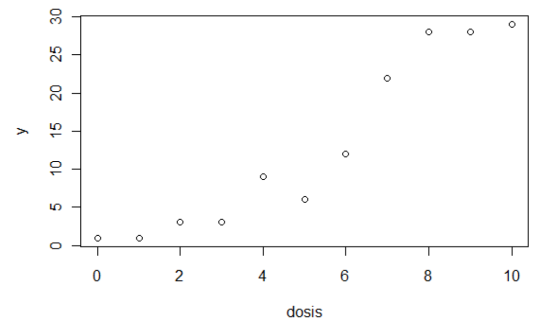
Statistische Tests für Tabellen

- **Fisher's Exact Test**: 2 x 2 Tabellen
  Verteilung der Teststatistik exakt

- **Chi-Quadrat Test**: m x n Tabellen
  Verteilung der Teststatistik asymptotisch

- **Logistische Regression**: 2 x m Tabellen;
  auch Mix aus mehreren kontinuierlichen und kategoriellen erklärenden Variablen möglich
  Verteilung der Teststatistik asymptotisch
  Multiple oder einfache ?

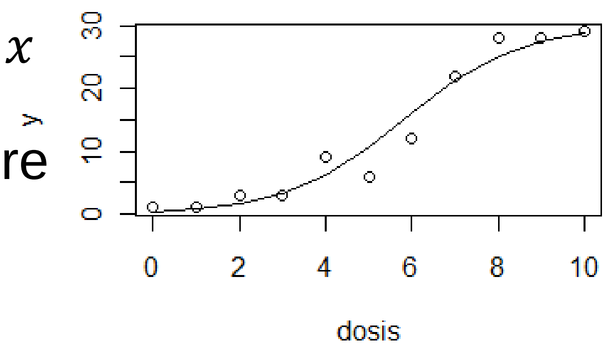
Logistische Regression



- $Y \sim Bin(n, p(x))$

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x$$

- Bsp: Dosis 0, 1, 2, …, 9, 10; je 30 kranke Tiere
  Y: Anzahl genesener Tiere




```
glm(cbind(y, n-y) ~ dosis, data = dat, family = binomial)
```

Dosis=0 → log-odds = $\beta_0 = -4.29$

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.29356    0.46067  -9.320   <2e-16
dosis        0.74103    0.07601   9.749   <2e-16
```

Wenn man Dosis um eine Einheit erhöht,
erhöhen sich die log-odds um $\beta_1 = 0.74$
95%-VI: 0.74 $\pm$ 2*0.076

Logistische Regression: Interpretation

- Wenn man Dosis um eine Einheit erhöht, erhöhen sich die log-odds um $\beta_1 = 0.74$
95%-VI: 0.74 $\pm$ 2*0.076, d.h. [0.588; 0.892]

oder äquivalent

- Wenn man Dosis um eine Einheit erhöht, erhöhen sich die odds um den Faktor $\exp(\beta_1) = 2.10$ (odds-ratio)
95%-VI: $[\exp(0.588)\,; \exp(0.892)] = [1.80; 2.44]$

Einfache oder Multiple Regression

(Gilt für alle GLMs; hier am Bsp der linearen Regression)

- Einfache Regression:
  "Totaler Effekt"
  y ~ x → "Wenn sich x um eine Einheit erhöht, erhöht sich y um β_1"
- Multiple Regression
  "Bereinigter Effekt"
  y ~ x1 + x2 → "Wenn sich x1 um eine Einheit erhöht und x2 konstant bleibt, erhöht sich y um β_1.
- Kein "richtig" oder "falsch"; eher zwei verschiedene Sichtweisen auf das gleiche Problem

Vorteil von Multipler Regression

- Andere Einflüsse werden ausgeschaltet, d.h., bereinigter Effekt ist möglich

Bsp: Diskriminierung

- Einfache Regression:
  Zulassung ~ Geschlecht
- Multiple Regression:
  Zulassung ~ Geschlecht + Qualifikation

  Berühmtes Beispiel: Simpson's Paradox

Simpson's Paradox
(Bsp: Aufgenommene Studenten an der UC Berkeley in 1973; nur 6 grösste Departemente)

	Angenommen	Abgelehnt
Männer	1198	1493
Frauen	557	1278

Werden Frauen benachteiligt?

Methodenvergleich

- Fisher Test
- Chi-Quadrat Test (mit Mosaikplot)
- (einfache) Logistische Regression

```
> head(dfUCB)
     Admit Gender Dept Freq
1 Admitted   Male    A  512
2 Rejected   Male    A  313
3 Admitted Female    A   89
4 Rejected Female    A   19
5 Admitted   Male    B  353
6 Rejected   Male    B  207
            ...
```

```
> fm1 <- glm(Admit ~ Gender, weights = Freq, family = binomial, data = dfUCB)
> summary(fm1)

Call:
glm(formula = Admit ~ Gender, family = binomial, data = dfUCB,
    weights = Freq)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-28.787  -14.662   -1.781   15.244   20.336

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.22013    0.03879   5.675 1.38e-08 ***
GenderFemale  0.61035    0.06389   9.553  < 2e-16 ***
```

```
> exp(confint(fm1))
Waiting for profiling to be done...
                2.5 %   97.5 %
(Intercept)  1.155124 1.344836
GenderFemale 1.624956 2.087499
```

Frauen benachteiligt?

Simpson's Paradox
(Bsp: Aufgenommene Studenten an der UC Berkeley in 1973)

Dept	Männer		Frauen	
	Bew.	Akz.	Bew.	Akz.
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	373	6%	341	7%

Simpson's Paradox
(Bsp: Aufgenommene Studenten an der UC Berkeley in 1973)

Dept	Männer		Frauen	
	Bew.	Akz.	Bew.	Akz.
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	373	6%	341	7%

Simpson's Paradox
(Bsp: Aufgenommene Studenten an der UC Berkeley in 1973)

Dept	Männer		Frauen	
	Bew.	Akz.	Bew.	Akz.
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	373	6%	341	7%

Nein: Frauen bewerben sich mehr bei „schwierigen“ Departments!

Bereinigter Effekt: Logistische Regression

```
glm(formula = Admit ~ Gender + Dept, family = binomial, data = dfUCB,
    weights = Freq)
```

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.58205    0.06899  -8.436   <2e-16
GenderFemale -0.09987    0.08085  -1.235    0.217
DeptB         0.04340    0.10984   0.395    0.693
DeptC         1.26260    0.10663  11.841   <2e-16
DeptD         1.29461    0.10582  12.234   <2e-16
DeptE         1.73931    0.12611  13.792   <2e-16
DeptF         3.30648    0.16998  19.452   <2e-16
```

Der bereinigte Geschlechtereffekt ist nicht signifikant

Zusammenfassung: Statistische Tests für Tabellen

- **Fisher's Exact Test**: 2 x 2 Tabellen
  Verteilung der Teststatistik exakt

- **Chi-Quadrat Test**: m x n Tabellen
  Verteilung der Teststatistik asymptotisch

- **Logistische Regression**: 2 x m Tabellen;
  auch Mix aus mehreren kontinuierlichen und kategoriellen erklärenden Variablen möglich
  Verteilung der Teststatistik asymptotisch
  Multiple oder einfache ?