1 Einleitung

Es soll anhand des `UCBAdmissions` Datensatzes die Frage beantwortet werden, ob bei der Zulassung von Studenten Frauen benachteiligt werden. Die Stichprobe lautet zunächst

```
# Daten
(UCB <- margin.table(UCBAdmissions, 2:1))

        Admit
Gender   Admitted Rejected
  Male       1198     1493
  Female      557     1278
```

Man bemerkt, dass ein Anteil von $\frac{1198}{1198+1493} = 0.445$, also annähernd die Hälfte aller Männer zugelassen wurde, während es nur ein Anteil von $\frac{557}{557+1278} = 0.304$ der Frauen geschafft hat. Dies würde die Hypothese, dass Frauen schwerer zugelassen werden, also benachteiligt werden, nahelegen.

Das allgemeine Aussehen einer solchen **Vierfeldertafel** ist folgendermaßen: man hat 2 diskrete Merkmale (X, hier das Geschlecht und Y, hier der Aufnahmestatus) mit jeweils 2 Ausprägungen a_i bzw. b_j (männlich/weiblich bzw. zugelassen/abgelehnt). Für jede sich bewerbende Person bestimmt man ihre konkrete Ausprägung x und y und damit die Gesamtsummen n_{11} (zugelassene Männer), n_{12} (abgelehnte Männer), n_{21} (zugelassene Frauen) und n_{22} (abgelehnte Fauen). Die Zeilensummen $n_{1.}$ (Anzahl Männer), $n_{2.}$ (Anzahl Frauen) bzw. die Spaltensummen $n_{.1}$ (Anzahl zugelassene Personen) und $n_{.2}$ (Anzahl abgewiesene Personen) sind die Randhäufigkeiten (Punkte bezeichnen den Index, über den summiert wird). Der gesamte Stichprobenumfang ergibt sich als Summe der 4 Einträge $n = n_{11} + n_{12} + n_{21} + n_{22}$ oder als Summe der Randhäufigkeiten $n = n_{1.} + n_{2.} = n_{.1} + n_{.2}$.

	Y		
X	b_1	b_2	Summe
a_1	n_{11}	n_{12}	$n_{1.}$
a_2	n_{21}	n_{22}	$n_{2.}$
Summe	$n_{.1}$	$n_{.2}$	n

2 Odds und Odds Ratio

Eine den obigen Anteilen ähnliche und aus der Kontingenztafel leicht zu berechnende Größe sind die **odds** (**Chancen**). So ergibt sich die Chance für einen Mann, zugelassen zu werden, als

$$\gamma(b_1, b_2 | X = a_1) = \frac{n_{11}}{n_{12}} = \frac{1198}{1493} = 0.80$$

während die Chancen für eine Frau nur rund halb so groß sind

$$\gamma(b_1, b_2 | X = a_2) = \frac{n_{21}}{n_{22}} = \frac{557}{1278} = 0.44$$

Eine Chance von 1 bedeutet also ein $50\,\%$ige Wahrscheinlichkeit, aufgenommen zu werden.

Um das Ergebnis auf eine Zahl zu komprimieren, berechnet man das **Chancenverhältnis** (*odds ratio*) als Quotient der beiden obigen Werte (hier aus Sicht des Mannes, aus Sicht der Frau wäre es der Kehrwert):

$$\gamma = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{21}n_{12}} = \frac{0.80}{0.44} = 1.84$$

Männer haben also eine fast doppelt so hohe Chance, zugelassen zu werden wie Frauen. In **R** erhält man diese Größen mittels der Funktionen `odds()` und `oddsratio()` im package `vcd`:

```
# odds fuer Maenner und Frauen, zugelassen zu werden
odds(t(UCB), log = FALSE)

 odds for Admit by Gender

    Male    Female
0.8024113 0.4358372

# odds ratio
oddsratio(t(UCB), log = FALSE)

 odds ratios for Admit and Gender

[1] 1.84108
```

3 χ^2-Test

Ein statistischer Test, um festzustellen, ob die (zunächst nur vermutete) Bevorzugung von Männern signifikant ist oder noch mit dem Zufall der Stichprobennahme erklärt werden kann, ist der χ^2-Test. Ausgangspunkt sind die 2 kategoriellen Merkmale X (Geschlecht) und Y (Zulassungsstatus) mit r bzw. c möglichen Ausprägungen (hier: $r = 2$ und $c = 2$). Wir fragen, ob zwischen X und Y ein Zusammenhang besteht, ob also die Tatsache, dass jemand zugelassen wird (oder auch nicht) vom Geschlecht abhängt. Als **Nullhypothese** $\mathcal{H}_0$ wählt man:

$$\mathcal{H}_0: \; X \text{ und } Y \text{ sind unabhängig}$$

Für unabhängige Zufallsgrößen und 2 Ereignisse A und B gilt

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Definiert man als Ereignisse

$$\begin{aligned} \text{A} &: X \text{ nimmt den Wert } i \text{ an} \\ \text{B} &: Y \text{ nimmt den Wert } j \text{ an} \end{aligned}$$

so gilt im Falle der **Unabhängigkeit** (wenn also $\mathcal{H}_0$ zutrifft)

$$p_{ij} = \mathbb{P}(X = i \wedge Y = j) = \underbrace{\mathbb{P}(X = i)}_{p_{i.}} \cdot \underbrace{\mathbb{P}(Y = j)}_{p_{.j}} = p_{i.} \cdot p_{.j}$$

p_{ij} ist die **gemeinsame Wahrscheinlichkeit**, dass gleichzeitig ein Wert i für X und j für Y beobachtet wird. $p_{i.}$ bezeichnet die sogenannte **Randwahrscheinlichkeit**, dass für X ein Wert i beobachtet wird (über das Ergebnis von Y wird keine Aussage gemacht). Analog bezeichnet $p_{.j}$ die Randwahrscheinlichkeit, dass für Y ein Wert j beobachtet wird (ohne Aussage über X). Im Falle der Unabhängigkeit lassen sich also gemeinsame Wahrscheinlichkeiten durch die Randwahrscheinlichkeiten berechnen:

$$p_{ij} = p_{i.} \cdot p_{.j}$$

In der Praxis sind alle diese Wahrscheinlichkeiten unbekannt und man geht zu den **relativen Häufigkeiten** der Stichprobe über. Wir beobachten nun die absolute Häufigkeit n_{ij} von i für X und j für Y. Daraus berechnet man zunächst die Randhäufigkeiten $n_{i.}$ und $n_{.j}$ für X bzw. Y:

$$n_{i.} = \sum_{j=1}^{c} n_{ij} \quad \text{bzw.} \quad n_{.j} = \sum_{i=1}^{r} n_{ij}$$

```r
# Randhaeufigkeiten
(UCB2 <- addmargins(UCB, margin = 1:2))
```

```
        Admit
Gender   Admitted Rejected  Sum
  Male       1198     1493 2691
  Female      557     1278 1835
  Sum        1755     2771 4526
```

Mit diesen Randhäufigkeiten schätzt man die Randwahrscheinlichkeiten für X und Y gemäß

$$\hat{p}_{i.} = n_{i.}/n \quad \text{bzw.} \quad \hat{p}_{.j} = n_{.j}/n$$

Ein Schätzwert für z.B. das Eintreten des Ereignisses *Die Person ist ein Mann*, $\hat{p}_{1.}$, lautet daher

$$\frac{2691}{4526} = 0.59 = \frac{\text{Anzahl Maenner}}{\text{Gesamtanzahl Bewerber}}$$

Unter $\mathcal{H}_0$ (Unabhängigkeit) lautet die **erwartete Häufigkeit** für das Auftreten des Paares (i, j):

$$e_{ij} = \mathbb{E}(N_{ij}) = n \cdot p_{ij} = n \cdot p_{i.} \cdot p_{.j} \approx n \cdot \frac{n_{i.}}{n} \cdot \frac{n_{.j}}{n} = \frac{n_{i.} \cdot n_{.j}}{n}$$

Für jedes der 4 Zellenfelder ergibt sich somit eine unter der Nullhypothese erwartete Häufigkeit e_{ij} (gerundete Werte):

X – Geschlecht	Y – Status zugelassen	abgelehnt	Summe
m	$e_{11} = \frac{2691 \cdot 1755}{4526} = 1043.461$	$e_{12} = \frac{2691 \cdot 2771}{4526} = 1647.539$	2691
w	$e_{21} = \frac{1835 \cdot 1755}{4526} = 711.539$	$e_{22} = \frac{1835 \cdot 2771}{4526} = 1123.461$	1835
Summe	1755	2771	4526

Intuitiv wird man erwarten, dass bei Gültigkeit der Nullhypothese $\mathcal{H}_0$ (Unabhängigkeit von Geschlecht und Zulassungsstatus bzw. keine Bevorzugung/Benachteiligung eines bestimmten Geschlechts) die beobachteten n_{ij} nur wenig von den unter $\mathcal{H}_0$ erwarteten Häufigkeiten e_{ij} abweichen. Die Testgröße t lautet

$$t = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

wodurch man im vorliegenden Fall erhält

$$t = \frac{(1198 - 1043.461)^2}{1043.461} + \frac{(1493 - 1647.539)^2}{1647.539} + \frac{(557 - 711.539)^2}{711.539} + \frac{(1278 - 1123.461)^2}{1123.461} =$$

$$= 22.888 + 14.496 + 33.564 + 21.258 = 92.2$$

Ein großer Wert der Testgröße spricht für große Abweichungen der n_{ij} von den e_{ij} und daher für Abweichungen von der $\mathcal{H}_0$ und einen Einfluß des Geschlechts, zugelassen zu werden oder nicht. Unter $\mathcal{H}_0$ ist die Testgröße χ^2-verteilt mit df $= (r-1) \cdot (c-1) = 1$ Freiheitsgrad(en). Das 0.95er Quantil der χ^2_1 Verteilung lautet

```r
# kritischer Wert
qchisq(p = 0.95, df = 1)

[1] 3.841459
```

wodurch die **Nullhypothese** $\mathcal{H}_0$ deutlich **abgelehnt** wird, da $t > \chi^2_{1,0.95}$. Man kann also – auf Basis dieser Daten – auf einen Einfluß des Geschlechts, zum Studium zugelassen zu werden, schließen. In **R** erfolgt der Test einfach mittels

```r
# Chi^2 - Test
chisq.test(UCB, correct = FALSE)


Pearson's Chi-squared test

data:  UCB
X-squared = 92.205, df = 1, p-value < 2.2e-16
```

4 Berücksichtigung des Departments

Die eben behandelten Daten sind in Wirklichkeit die Summe über 6 Departments `A` bis `F`, an denen die Bewerbungen erfolgten. Man hat somit eine $2 \times 2 \times 6$ Kontingenztafel.

```r
# Originaldaten
UCBAdmissions


, , Dept = A

          Gender
Admit      Male Female
  Admitted  512     89
  Rejected  313     19

, , Dept = B

          Gender
Admit      Male Female
  Admitted  353     17
  Rejected  207      8

, , Dept = C

          Gender
Admit      Male Female
  Admitted  120    202
  Rejected  205    391

, , Dept = D

          Gender
Admit      Male Female
  Admitted  138    131
  Rejected  279    244

, , Dept = E

          Gender
Admit      Male Female
  Admitted   53     94
  Rejected  138    299

, , Dept = F

          Gender
Admit      Male Female
  Admitted   22     24
  Rejected  351    317
```

Zunächst wäre die Frage, ob in jedem Department das Chancenverhältnis zugelassen zu werden ähnlich/gleich ist. Nur in einem solchen Fall wäre eine Auswertung der kumulierten Daten (wie in den vorigen Abschnitten) sinnvoll. Andernfalls (wenn das Maß der Bevorzugung von Männern oder Frauen in den 6 Departments unterschiedlich ist) wäre eine getrennte Betrachtung/Auswertung sinnvoll. Man sieht sich zunächst die **Chancenverhältnisse** in den 6 Departments an. Ein oddsratio von 1 zeigt gleiche Chancen für Männer und Frauen an.

```
# oddsratios nach department
oddsratio(UCBAdmissions, log = FALSE)

 odds ratios for Admit and Gender by Dept

        A         B         C         D         E         F
0.3492120 0.8025007 1.1330596 0.9212838 1.2216312 0.8278727
```

Man beachte, dass in den Departments `B`, `D` und `F` die oddsratios leicht unter 1 sind (Frauen werden dort also leicht bevorzugt) und in den Departments `C` und `E` die oddsratios leicht über 1 liegen (Männer werden dort leicht bevorzugt). Lediglich in Department `A` hat man ein deutlich von 1 verschiedenes oddsratio von 0.35. Männer werden dort deutlich benachteiligt – die Chance einer Frau, zugelassen zu werden, ist dort etwa 3 mal so hoch wie für einen Mann. Man kann sich Konfidenzintervalle für die oddsratios für jedes Department ansehen

```
# CI fuer oddsratios nach department
confint(oddsratio(UCBAdmissions, log = FALSE))

      2.5 %    97.5 %
A 0.2086756 0.5843954
B 0.3403815 1.8920166
C 0.8545328 1.5023696
D 0.6863345 1.2366620
E 0.8250748 1.8087848
F 0.4552059 1.5056335
```

und erkennt, dass in allen Departments außer `A` der Wert 1 im Konfidenzintervall liegt, also keine signifikante Benachteiligung von Frauen vorliegt. Hingegen sind im Department `A` Männer signifikant schlechter gestellt. Mit Hilfe des **Woolf Tests** kann die Gleichheit der oddsratios getestet werden

```
# Gleichheit der oddsratios
woolf_test(UCBAdmissions)


Woolf-test on Homogeneity of Odds Ratios (no 3-Way assoc.)

data:  UCBAdmissions
X-squared = 17.902, df = 5, p-value = 0.003072
```

Als Ergebnis erhält man signifikante Unterschiede in den oddsratios. Die scheinbare Benachteiligung von Frauen im kumulierten Datensatz ergeben sich aufgrund der unterschiedlichen Zulassungsraten und der ungleichen Bewerbung der Frauen in den 6 Departments. Man erkennt zunächst, dass in den Departments `A` und `B` deutlich höhere Chancen bestehen angenommen zu werden – ca. $2/3$ aller Bewerber werden angenommen. Die Chance in den Departments `C`, `D` und `E` ist hingegen nur rund halb so groß und in Department `F` mit ca. $6\,\%$ nochmals deutlich geringer.

```
# Prozentsatz der zugelassenen Studenten
round(apply(UCBAdmissions[1,,,drop = FALSE], 3, sum) / apply(UCBAdmissions, 3, sum), 2)

   A    B    C    D    E    F
0.64 0.63 0.35 0.34 0.25 0.06
```

Insgesamt sind unter allen Bewerbern rund $40\,\%$ Frauen.

```
# Anteil Frauen
sum(UCBAdmissions[,2,]) / sum(UCBAdmissions)

[1] 0.4054353
```

Sieht man sich nun den Frauenanteil in den Bewerbungen an den einzelnen Departments an, so bemerkt man, dass sich Frauen genau in den Departments (`C` bis `F`) beworben haben, die generell eine geringere Chance auf Zulassung haben, während 88 % bzw. 96 % aller Bewerber in den Departments `A` und `B` (mit hohen Zulassungsraten) Männer sind.

```
# Frauen Bewerbungen unter allen Bewerbungen nach Departments
round(apply(UCBAdmissions[,2,,drop = FALSE], 3, sum) / apply(UCBAdmissions, 3, sum), 2)

   A    B    C    D    E    F
0.12 0.04 0.65 0.47 0.67 0.48
```

5 Schlußfolgerung

Statistisch bedeuten die Ergebnisse im letzten Abschnitt, dass die 6 Departments gleichsam unterschiedliche Populationen (hinsichtlich der oddsratios) darstellen. Eine Auswertung der Daten über alle Departments hinweg bzw. die Berechnung von Chancen/Chancenverhältnissen über alle Departments ist daher nicht sinnvoll.