

Logistische Regression

1. (a) Die Tabelle zeigt die Anzahl der Sprecher aufgeteilt in 4 Altersgruppen die einen silbenfinalen /l/ vokalisieren.

Altersgruppe	18	30	45	70
vok	58	55	62	38
nicht-vok	34	49	84	59

(Die Altersgruppe zeigt jeweils den Altersmittelwert pro Gruppe in Jahren).

Prüfen Sie, ob die /l/-Vokalisierung von Altersgruppe beeinflusst wird.

2. Die folgenden Daten zeigen die Anzahl der Sprecher (P), die einen post-vokalisches Tap (in Wörtern wie water) in Australien produzieren oder nicht (Q) aufgeteilt nach Altersgruppe, Herkunftsland (Irland oder England), Geschlecht, und wie lange sie in Australien leben (Jahre). Prüfen Sie, inwiefern diese Faktoren auf die Produktion von /t/ (als Tap oder nicht) einen Einfluss haben.

	P	Q	Alt	Gen	Her	J
1	3	24	j	m	I	5
2	25	114	j	m	E	5
3	0	5	j	w	I	5
4	2	20	j	w	E	5
5	0	16	a	m	I	5
6	6	69	a	m	E	5
7	0	4	a	w	I	5
8	1	23	a	w	E	5
9	8	13	j	m	I	15
10	9	21	j	m	E	15
11	2	6	a	m	I	15
12	1	8	a	m	E	15
13	31	46	j	m	I	25
14	10	21	j	m	E	25
15	5	42	a	m	I	25
16	3	12	a	m	E	25

Die Spalten bekommen Sie wie folgt:

`P = c(3, 25, 0, 2, 0, 6, 0, 1, 8, 9, 2, 1, 31, 10, 5, 3)`

`n = c(27, 139, 5, 22, 16, 75, 4, 24, 21, 30, 8, 9, 77, 31, 47, 15)`

`Q = n - P`

`alt = c(rep("j", 4), rep("a", 4), rep("j", 2), rep("a", 2), rep("j", 2), rep("a", 2))`

`gen = c(rep("m", 2), rep("w", 2), rep("m", 2), rep("w", 2), rep("m", 8))`

`her = rep(c("I", "E"), 8)`

`jahre = c(rep(5, 8), rep(15, 4), rep(25, 4))`

3. In einem Spracherkennungssystem wurden /b, d, g/ getrennt für Männer und Frauen klassifiziert. Das Ergebnis ist als Data-Frame hier vorhanden:

```
ver = read.table(url("http://www.phonetik.uni-muenchen.de/~jmh/lehre/sem/ss11/statfort/ver.txt"))
```

Die Spalten Zeigen ob richtig klassifiziert wurde oder nicht (Erg), die Artikulationsstelle (Art), und Geschlecht (G). Inwiefern wird ob richtig erkannt wurde oder nicht von den anderen Faktoren beeinflusst?

4. Diese Daten:

```
auf <- read.csv(url("http://www.ats.ucla.edu/stat/r/dae/binary.csv"))
```

(aus UCLA) zeigen, ob Kandidaten in eine Universität aufgenommen wurden oder nicht (`admit`)¹. Sie zeigen zusätzlich den `gpa` (Grade-Point-Average) und Ranking ihrer Schule (auf einer Skala 1-5). Inwiefern wird die Aufnahme durch diese beiden Faktoren beeinflusst? (N.B. da `gpa` kontinuierlich ist, benötigen Sie zwei Histogramme, um die Daten für die 2 Stufen von `admit` abzubilden:

```
temp = auf$admit == 0  
hist(auf$gpa[temp])  
hist(auf$gpa[!temp])
```

Sie können das auch mit einem 'Back-to-Back' Histogramm abbilden, nachdem Sie `library(Hmisc)` installiert haben:

```
install.packages("Hmisc")  
library(Hmisc)  
histbackback(auf$gpa[temp], auf$gpa[!temp])
```

5. Diese Daten:

```
lei = read.table(url("http://www.phonetik.uni-muenchen.de/~jmh/lehre/sem/ss11/statfort/lei.txt"))
```

zeigen die Anzahl der Antworten in einem forced-choiced Test, in dem mit entweder *leiden* oder *leiten* in einem 7-stufigen Kontinuum (factor Stim) beantwortet werden musste. z.B. zeigt Reihe 3 zwei *leiden*-Antworten (P), und vier *leiten*-Antworten (Q) auf Stimulus-Nummer 3. Diese Ergebnisse sind für 2 Hörer vorhanden (Faktor `Dial`), einen Standard-Sprecher und einen (jungen) Sprecher aus Franken. Berechnen Sie die Umkipppunkte und zeigen Sie die psychometrischen Kurven getrennt für diese beiden Hörer.

¹ Siehe <http://www.ats.ucla.edu/stat/r/dae/logit.htm> für weitere Details