

## Chi-Quadrat und logistische Regression

L-Vokalisierung wird signifikant von zunehmendem Alter beeinflusst (logistic regression,  $\chi^2[1] = 11.8$ ,  $p < 0.001$ ).

1. (a) Prüfen Sie mit einem  $\chi^2$  Test, ob in den diesen Daten:

|           | A20min | A20bis30 | A31bis40 | A41plus |
|-----------|--------|----------|----------|---------|
| vok       | 58     | 55       | 62       | 38      |
| nicht-vok | 34     | 49       | 84       | 59      |

die auch in der Webseite vorhanden ist:

```
pfad = "Verzeichnis wo Sie lvoc.txt gespeichert haben"
lvoc = read.table(paste(pfad, "lvoc.txt", sep="/"))
```

einen Trend vorliegt, d.h. ob die Proportionen der vokalisierten /l/s mit zunehmenden Alter abnimmt.

(b) Wiederholen Sie den Test mit logistischer Regression und überlagern Sie in einer Abbildung eine Regressionskurve auf Alter  $\times$  Proportionen.

2. In der Webseite ist ein Data-frame `glottal.txt` (/t/ Glottalisierung), der wie folgt in R eingelesen werden kann

```
pfad = "Verzeichnis wo Sie glottal.txt gespeichert haben"
glottal = read.table(paste(pfad, "glottal.txt", sep="/"))
```

Dieser Data-Frame enthält Informationen in der ersten Spalte, ob ein /t/ ohne Lösung ("t"), glottalisiert ("tQ"), oder mit einer Lösung "tgel" erzeugt wurde. Der Data-Frame enthält auch Informationen in der zweiten Spalte bezüglich des Geschlechtes der Versuchsperson ("m" für männlich, "w" fuer weiblich)

(a) Erstellen Sie eine Tabelle der Verteilung der /t/-Realisierungen getrennt für männlich und weiblich.

(b) Erstellen Sie einen Barplot für die Daten in (a).

(c) Prüfen Sie für diese Daten, ob die /t/-Realisierungen in Männern und Frauen unterschiedlich verteilt sind.

3. (Daten aus Harrington, Kleber, Reubold, *JASA*, 2008).

```
pfad = "Verzeichnis wo Sie ui.txt gespeichert haben"
ui = read.table(paste(pfad, "ui.txt", sep="/"))
```

Anhand der Sprachsynthese wurde ein F2-Kontinuum in 11 Schritten synthetisiert. 5 Vpn. (L1-Englisch) mussten zu jedem Stimulus mit "I" oder "U" antworten.

(a) Folgt die Proportion der /u/-Urteile als Funktion von zunehmendem F2 einem Trend?

(b) Der F2-Umkipppunkt,  $U$ , von /u/ auf /i/ wird durch

$$U = -k/m$$

gegeben. Hier sind  $k$  und  $m$  der Abschnitt (Intercept) und die Neigung der (log-odds) Regressionslinie. Bei welchem F2-Wert ist der Umkipppunkt? Überlagern Sie diesen Wert auf eine Abbildung mit der Regression in einer Abbildung von Proportion der /u/-Urteile als Funktion von F2.

4. Zwölf Versuchspersonen aus Bayern und acht aus Hessen produzierten einen initialen /z/ mit Stimmhaftigkeit (J) oder nicht (N) wie folgt:

"J", "J", "J", "J", "N", "N", "J", "N", "J", "J", "J", "N", "J", "N", "N", "N", "J", "J", "N", "J"

Prüfen Sie an hand der logistischen Regression, ob sich die Dialektgruppen in der Erzeugung von /z/ mit Stimmhaftigkeit unterscheiden.

5. Diese Matrix zeigt die Verteilung der Erzeugungen von /lost/ (high) und /lost/ (low) aufgeteilt nach Geschlecht (0, 1) und Alter (A, J).

|     | high | low |
|-----|------|-----|
| A.0 | 43   | 35  |
| A.1 | 30   | 15  |
| J.0 | 3    | 13  |
| J.1 | 6    | 75  |

```
pfad = "das Verzeichnis wo ich lost3.txt gespeichert habe"
lost3 = as.matrix(read.table(paste(pfad, "lost3.txt", sep="/")))
```

Verwenden Sie die logistische Regression, um zu prüfen, ob Alter und Geschlecht die high-low Verteilung signifikant beeinflussen.

6. In einem Spracherkennungssystem wurden /b, d, g/ klassifiziert. Die Verwechslungsmatrizen der Plosive für einen Sprecher und eine Sprecherinnen sind wie folgt:

Sprecher:

[ver.m](#)

|   | b  | d  | g  |
|---|----|----|----|
| b | 90 | 62 | 77 |
| d | 34 | 88 | 44 |
| g | 10 | 20 | 80 |

Sprecherin

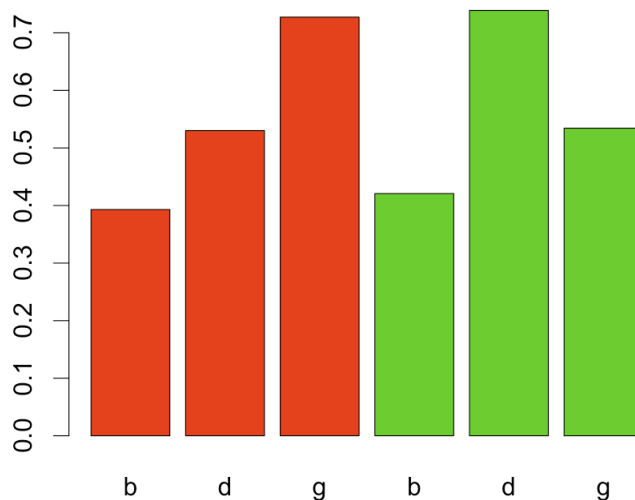
[ver.w](#)

```
ver.w
      b    d    g
b 85   52   65
d 25  150   28
g  0   61   70
```

Die richtig klassifizierten Laute sind in der Diagonale. Die erste Reihe von [ver.w](#) ist wie folgt zu interpretieren: 85 /b/ sind richtig als /b/ klassifiziert worden; 52 /b/ sind falsch als /d/, und 65 /b/ sind falsch als /g/ klassifiziert worden. Die Matrizen sind hier vorhanden:

```
pfad = "das Verzeichnis wo ich verm.txt und verw.txt gespeichert
habe"
ver.m = as.matrix(read.table(paste(pfad, "verm.txt", sep="/")))
ver.w = as.matrix(read.table(paste(pfad, "verw.txt", sep="/")))
```

6.1 Erstellen Sie einen Barplot wie unten der Proportionen der richtig klassifizierten Laute getrennt für männlich (rot) und weiblich (grün).



6.2 Schätzen Sie an hand des Barplots ein ob:

- (a) die Erfolgsrate von der Artikulationsstelle abhängig ist
- (b) die Erfolgsrate vom Geschlecht beeinflusst wird
- (c) eine Interaktion zwischen Artikulation und Geschlecht vorliegt.

6.3 Prüfen Sie (a), (b), (c) an hand der logistischen Regression.