

Chi-Quadrat und logistische Regression

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

1. (a) Prüfen Sie mit einem χ^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:

einen Trend vorliegt, d.h. ob die Proportionen der vokalisiert /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

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 = as.matrix(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? NB: Sie bekommen die F2-Werte, die als Reihennahmen gespeichert wurden, durch:

```
F2 = as.numeric(rownames(ui))
```

(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"

(Die ersten 12 dieser Werte sind von den Vpn aus Bayern).

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

5. Der Data-Frame `lost3` zeigt die Verteilung von /lost/ (high) und /lost/ (low) aufgeteilt nach Geschlecht (0, 1) und Alter (A, J).

```
pfad = "das Verzeichnis wo ich lost3.txt gespeichert habe"
```

```
lost3 = read.table(paste(pfad, "lost3.txt", sep="/"))
```

(a) Schätzen Sie anhand von einem Interaction-Plot ein, ob (i) Alter (ii) Geschlecht auf die high/low Verteilung einen signifikanten Einfluss hat.

(b) Was würde in Wörtern bedeuten: die Interaktion Alter x Geschlecht ist signifikant? Könnte laut dem Interaction-Plot eine Interaktion vorliegen?

(c) Prüfen Sie (a) und (b) anhand der logistischen Regression.

6. In einem Spracherkennungssystem wurden /b, d, g/ getrennt für Männer und Frauen klassifiziert. Das Ergebnis ist in dem Data-Frame `ver` vorhanden.

```
pfad = "das Verzeichnis wo ich lost3.txt gespeichert habe"
```

```
ver = read.table(paste(pfad, "ver.txt", sep="/"))
```

Die erste Spalte (K) enthält die vom Erkennungssystem klassifizierten Labels, die zweite Spalte (Lab) enthält die von Phonetikern manuell erstellten Etikettierungen, die dritte Spalte zeigt, ob der Konsonant von einem Mann oder Frau gesprochen wurde. Ein Konsonant wurde *richtig* klassifiziert, wenn die klassifizierten und manuell erstellten Etikettierungen denselben Wert haben (beide "b" oder beide "d" oder beide "g") sonst *falsch* (z.B. klassifiziert wurde "g" aber manuell erstellt wurde "b").

(a) Versuchen Sie anhand von einen Interaction-Plot einzuschätzen, ob die Proportion richtig klassifiziert (also die Erfolgsrate) durch (i) die Artikulationsstelle (ii) Geschlecht beeinflusst wird.

(b) Was würde in Wörtern bedeuten: die Interaktion Artikulationsstelle x Geschlecht ist signifikant? Könnte laut dem Interaction-Plot eine Interaktion vorliegen?

(c) Prüfen Sie (a) und (b) anhand der logistischen Regression.