

```

library(lattice)
library(ez)
library(lme4)
source(file.path(pfadu, "phoc.txt"))

zweit = read.table(file.path(pfadu, "zweit.df.txt"))
bil = read.table(file.path(pfadu, "bil.df.txt"))
fremd = read.table(file.path(pfadu, "fremd.df.txt"))
tv = read.table(file.path(pfadu, "tv.df.txt"))
phr = read.table(file.path(pfadu, "phr.df.txt"))
elwi = read.table(file.path(pfadu, "elwi.df.txt"))
votspan = read.table(file.path(pfadu, "votspan.txt"))
stimm = read.table(file.path(pfadu, "votalle2.txt"))
franken = read.table(file.path(pfadu, "franken.txt"))

sig = function(k, m, add = T, ...)
{
  # Funktion um Sigmoiden auf Proportionen zu  $\frac{1}{4}$ berlagern.
  curve(exp(m * x + k)/(1 + exp(m * x + k)), add = add, ...)
}

# 1. In diesen Daten:
dim(franken)
# haben Hörer Wörter richtig oder falsch erkannt (Faktor Correct). Erstellen
# Sie eine Abbildung OHNE einen statistischen Test durchzuführen, um
# einzuschätzen, inwiefern die Erkennung (Faktor Correct) vom Alter des
# Hörers (Alter) und von der phonologischen Kategorie (Faktor Phon: /d/
# oder /t/) beeinflusst wurde. Erklären Sie in 1–2 Zeilen, ob Correct von
# diesen Faktoren beeinflusst wird.

tab = with(franken, table(Alter, Phon, Correct))
p = prop.table(tab, 1:2)
barchart(p, auto.key=T, horizontal=F)
# Alter hat einen Einfluss: mehr FALSE im Vgl. zu jung. Phon hat einen
# Einfluss: mehr FALSE für /t/ im Vgl. zu /d/. Es gibt eine Interaktion:
# die Altersunterschiede sind ausgeprägter in /t/ im Vgl. zu /d/.

# 2. Diese Daten:
dim(tv)
# zeigen eine Messung der Sprachkompetenz (cdi) für 80 Schüler (Vpn) sowie die
# Anzahl der Stunden am Tag, die sie fernsehen (tv.hours). Prüfen Sie
# durch eine Abbildung und statistischen Test, ob eine Verbindung zwischen
# Sprachkompetenz und fernsehen vorliegt.
plot(cdi ~ tv.hours, data = tv)
reg = lm(cdi ~ tv.hours, data = tv)
abline(reg)
summary(reg)
# Es gibt eine lineare Beziehung zwischen cdi und tv.hours ( $R^2 = 0.23$ ,
#  $F[1,78] = 23.1$ ,  $p < 0.001$ ).
shapiro.test(resid(reg))
# OK
plot(resid(reg))
# OK
acf(resid(reg))
# OK

# 3. Für diese Daten:
dim(zweit)
# nahmen Versuchspersonen (Vpn) an einem Test in einer zweiten Sprache teil
# (l2score). Prüfen Sie durch eine Abbildung und statistischen Test, ob

```

l2score durch Geschlecht (G) beeinflusst wird.

```
bwplot(l2score ~ G, data = zweit)
with(zweit, table(Vpn, G))
table(zweit$G)
t.test(l2score ~ G, data = zweit)
# Geschlecht hat einen signifikanten Einfluss auf l2score (t[117.6] = 2.1, p <
  0.05).
with(zweit, tapply(l2score, G, shapiro.test))
```

4. Für diese Daten:

```
dim(elwi)
# wurde ein 13-stufiges F2-Kontinuum erstellt (F2) und eine Versuchsperson
  musste pro Stimulus entscheiden, ob 'will' oder 'wool' wahrgenommen
  wurde (Urteil). Prüfen Sie durch eine Abbildung und statistischen Test,
  ob die will/wool Entscheidung durch F2 beeinflusst wird. Berechnen Sie
  den F2-Umkipppunkt, zu dem die Entscheidung von 'will' nach 'wool'
  kippt, und überlagern Sie eine Sigmoid mit der sig() Funktion auf die
  proportionalen Werte.
```

```
levels(elwi$Urteil)
P = elwi$Urteil == "wool"
Q = !P
e.m = aggregate(cbind(P, Q) ~ F2, sum, data = elwi)
p = with(e.m, P/(P+Q))
e.m = cbind(e.m, p)
plot(p ~ F2, data = e.m)
wool.glm = glm(Urteil ~ F2, binomial, data = elwi)
# oder
wool.glm = glm(cbind(P, Q) ~ F2, binomial, data = e.m)
summary(wool.glm)
# Das Urteil wird von F2 beeinflusst (z = 3.2, p < 0.01)
# oder
ohne = update(wool.glm, ~ . -F2)
anova(wool.glm, ohne, test="Chisq")
# oder
anova(wool.glm, test="Chisq")
# Das Urteil wird von F2 beeinflusst (X^2[1] = 65.1, p < 0.001)
```

```
cf = coef(wool.glm)
k = cf[1]; m = cf[2]
sig(k, m, col = "blue")
u = -k/m
abline(v=u, lty=2)
```

5. Prüfen Sie durch eine Abbildung und statistischen Test für diese Daten:

```
dim(votspan)
# inwiefern VOT vom Alter und von der Stadt beeinflusst wird.
```

```
bwplot(VOT ~ Alter | Stadt, data = votspan)
with(votspan, table(Vpn, interaction(Alter, Stadt)))
ezANOVA(votspan, .(VOT), .(Vpn), between = .(Alter, Stadt))
# VOT wird signifikant vom Alter (F[1,44] = 46.2, p < 0.001) und von der Stadt
  (F[1,44] = 17.5, p < 0.001) signifikant beeinflusst, und es gibt eine
  signifikante (F[1, 44] = 5.7, p < 0.05) Interaktion zwischen diesen
  Faktoren.
```

```
p = phoc(votspan, .(VOT), .(Vpn), .(Alter, Stadt))
round(phsel(p$res), 3)
round(phsel(p$res, 2), 3)
```

```
# Post-hoc t-tests bestätigten die signifikanten Unterschiede zwischen alt und
jung in Granada ( $p < 0.05$ ) und in Sevilla ( $p < 0.001$ ). Der Unterschied
zwischen den Städten war für jung ( $p < 0.01$ ) jedoch nicht für alt
signifikant.
```

```
# 6. Für diese Stichproben:
```

```
d = c(17, 6, 12, 11, 8, 8, 13, 3, 11, 8, 13, 5, 15, 5, 8, 14, 14, 8,
      13, 3)
```

```
# erstellen Sie einen Densityplot und berechnen Sie ein 95% Konfidenzintervall
für den Mittelwert.
```

```
densityplot(~d)
m = mean(d)
n = length(d)
SE = sd(d)/sqrt(n)
```

```
unten = m + SE * qt(0.025, n-1)
oben = m + SE * qt(0.975, n-1)
```

```
# oder ganz einfach:
```

```
t.test(d)
# 95 percent confidence interval:
# 7.824065 11.675935
```

```
# 7. Diese Daten:
```

```
dim(phr)
# zeigen eine Messung der Sprechgeschwindigkeit (tempo) in verschiedenen
Kontexten (Faktor Kontext). Die Sprechgeschwindigkeitsmessungen wurden
von 30 Sprechern (Vpn) in 10 verschiedenen Äußerungen (Ag) erhoben.
Prüfen Sie durch eine Abbildung und statistischen Test, inwiefern die
Sprechgeschwindigkeit vom Kontext beeinflusst wird.
```

```
bwplot(tempo ~ Kontext, data = phr)
with(phr, table(Vpn, Kontext))
with(phr, table(Ag, Kontext))
kon.lmer = lmer(tempo ~ Kontext + (1+Kontext|Vpn) + (1+Kontext|Ag), data =
phr)
kon.ohne = update(kon.lmer, ~ . -Kontext)
anova(kon.lmer, kon.ohne)
# Sprechgeschwindigkeit wurde signifikant vom Kontext beeinflusst ( $X^2[2] =$ 
40.6,  $p < 0.001$ ).
# Der Grund für Signifikanz: pp > ip in fast allen Wörtern
bwplot(tempo ~ Kontext | Ag, data = phr)
```

```
# 8. Für diese Daten:
```

```
dim(zweit)
# erstellen Sie eine Abbildung mit l2score auf der y-Achse und l1score auf der
x-Achse getrennt für die drei Stufen von 'Ses' und mit einer
Differenzierung (durch Farbkodierung) pro Abbildung für Gender (Faktor
G).
```

```
xyplot(l2score ~ l1score | Ses, groups = G, data = zweit, auto.key=T)
```

```
# 9. Für diese Daten:
```

```
dim(zweit)
# prüfen Sie durch eine Abbildung und statistischen Test, inwiefern l2score
aus l1score vorhergesagt werden kann. Was wäre aufgrund von Ihrem
statistischen Modell der vorhergesagte l2score für einen l1score von 65?
```

```

plot(l2score ~ l1score, data = zweit)
reg = lm(l2score ~ l1score, data = zweit)
abline(reg)
summary(reg)
# Es gibt eine signifikante lineare Beziehung zwischen l2score und l1score
  (R^2 = 0.33, F[1,118] = 56.9, p < 0.001)
# OK
shapiro.test(resid(reg))
# OK
plot(resid(reg))
# OK
acf(resid(reg))

# 10. Für diese Daten:
dim(zweit)
# prüfen Sie durch eine Abbildung und statistischen Test inwiefern Gender aus
  gpa (grade-point-average) vorhergesagt werden kann. Zu welchem gpa-Wert
  kommt der Umkipppunkt zwischen M und F vor?

levels(zweit$G)
P = zweit$G == "M"
Q = !P
zweit.m = aggregate(cbind(P, Q) ~ gpa, sum, data = zweit)
p = with(zweit.m, P/(P+Q))
zweit.m = cbind(zweit.m, p)
plot(p ~ gpa, data = zweit.m, ylab = "Proportion M")
gpa.glm = glm(G ~ gpa, binomial, data = zweit)
summary(gpa.glm)
# Gender kann aus GPA vorhergesagt werden (z = 3.0, p < 0.01)

cf = coef(gpa.glm)
# Umkipppunkt
-cf[1]/cf[2]
sig(cf[1], cf[2], col = "blue")

# 11. 8 Würfel wurden zusammengeworfen und der Mittelwert davon wurde
  berechnet. Erstellen Sie ein 95% Konfidenzintervall für den Mittelwert.
  Berechnen Sie die Wahrscheinlichkeit, dass der Mittelwert über 4 liegt.
mu = 3.5
SE = sd(1:6) * sqrt(5/6) / sqrt(8)
# Konfidenzintervall
unten = qnorm(0.025, mu, SE)
oben = qnorm(0.975, mu, SE)
# Pr(m > 4)
1 - pnorm(4, mu, SE)

# 12. Diese Daten:
dim(bil)
# zeigen die durchschnittliche Äußerungslänge (mlu) für 20 Kinder (Vpn), die
  die zweite Sprache entweder in der Schule oder zu Hause gelernt haben
  (Faktor Sprache) und die sich in drei verschiedenen Schulklassen (Faktor
  Klasse) befinden. Prüfen Sie durch eine Abbildung und statistischen Test
  inwiefern die Zweitsprachkompetenz von der Sprache und/oder Klasse
  beeinflusst wird.

bwplot(mlu ~ Sprache | Klasse, data = bil)
bwplot(mlu ~ Klasse | Sprache, data = bil)
# Beide Faktoren sind within

```

```

with(bil, table(Vpn, interaction(Sprache, Klasse)))
ezANOVA(bil, .(mlu), .(Vpn), .(Sprache, Klasse))
# mlu wird von der Sprache (F[1,19] = 7.1, p < 0.05) und von der Klasse
  (F[2,38] = 17.8, p < 0.001) signifikant beeinflusst und es gab keine
  Interaktion zwischen diesen Faktoren.

# 13. Diese Daten:
dim(fremd)
# zeigen für 30 Sprecher (Vpn) eine Messung der Sprechgeschwindigkeit (tempo),
  wenn sie in ihrer Muttersprache oder ihrer Zweitsprache (Sprache) reden.
  Prüfen Sie durch eine Abbildung und statistischen Test, inwiefern die
  Sprechgeschwindigkeit von der Sprache beeinflusst wird.

bwplot(tempo ~ Sprache, data = fremd)
with(fremd, table(Vpn, Sprache))
# Sprache ist within
ezANOVA(fremd, .(tempo), .(Vpn), .(Sprache))
# tempo wurde signifikant von der Sprache beeinflusst (F[1,29] = 28.2, p <
  0.001).
# oder als gepaarter t-test
d = aggregate(tempo ~ Vpn, diff, data = fremd)
bwplot(d$tempo)
t.test(d$tempo)
# tempo wurde signifikant von der Sprache beeinflusst (t = 5.3, p < 0.001).

# 14. Diesen Daten:
dim(stimm)
# zeigen VOT-Werte von zwei verschiedenen Konsonanten (Faktor K). Die
  Konsonanten stammen aus verschiedenen Wörtern (Wort) und wurden von 9
  Versuchspersonen (Vpn) produziert. Prüfen Sie durch eine Abbildung und
  statistischen Test, ob VOT von den Konsonanten beeinflusst wird.

bwplot(VOT ~ K, data = stimm)
with(stimm, table(Vpn, K))
with(stimm, table(Wort, K))
k.lmer = lmer(VOT ~ K + (1+K|Vpn) + (1|Wort), data = stimm)
k.ohne = update(k.lmer, ~ . -K)
# VOT wurde nicht signifikant von K beeinflusst
anova(k.lmer, k.ohne)
# VOT wurde signifikant von K beeinflusst (X^2[1] = 19.4, p < 0.001).

# 15. Für dieselben Daten:
dim(stimm)
# erstellen Sie einen neuen Vektor mit aufgerundeten VOT-Werten (z.B. 12.711 -
  > 12). Prüfen Sie mit einer Abbildung und statistischen Test inwiefern
  der Unterschied zwischen /d/ und /t/ (Faktor Kons) durch diese
  aufgerundeten VOT-Werte vorhergesagt werden kann. Überlagern Sie eine
  Sigmoid-Funktion auf die Abbildung. Zu welchem VOT-Wert kommt der
  Umkipppunkt zwischen /d/ und /t/ vor?

votr = round(stimm$VOT)
levels(stimm$K)
P = stimm$K == "t"
Q = !P
vot.m = aggregate(cbind(P, Q) ~ votr, sum, data = stimm)
p = with(vot.m, P/(P+Q))
vot.m = cbind(vot.m, p)
plot(p ~ votr, main = "Proportion /t/", data = vot.m)
vot.glm = glm(K ~ votr, binomial, data = stimm)

```

```
# Die Wahl zwischen /d/ und /t/ wurde signifikant von VOT beeinflusst (z =  
  5.4, p < 0.001)  
# oder  
anova(vot.glm, test="Chisq")  
# Die Wahl zwischen /d/ und /t/ wurde signifikant von VOT beeinflusst (X^2[1]  
  = 144.4, p < 0.001)  
cf = coef(vot.glm)  
k = cf[1]; m = cf[2]  
sig(k, m, col = "blue")  
u = -k/m  
abline(v=u, lty=2)
```