sheet 2 solutions (coding part)

carsten stahl

2023-10-17

Loading the data

Before starting this exercise we are going to load the data and requirements

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(nlme)
```

```
##
## Attache Paket: 'nlme'
##
## Das folgende Objekt ist maskiert 'package:dplyr':
##
##     collapse
```

```
load("./data/SimulatedTreatmentEffect.RData")
```

a)

```
gnls_treatment <- gnls(resp ~ th1+(th4-th1)/(1+(exp((conc-th2)*th3))),
                       data = conc.resp.df,
                       params=list(th1+th2+th3+th4~1),
                       control=gnlsControl(nlsTol=0.1),
                       start=c(0, 2, 1, 100))

summary(gnls_treatment)
```

```
## Generalized nonlinear least squares fit
##   Model: resp ~ th1 + (th4 - th1)/(1 + (exp((conc - th2) * th3)))
##   Data: conc.resp.df
##        AIC      BIC   logLik
##   313.2579 322.6139 -151.629
##
## Coefficients:
##        Value Std.Error  t-value p-value
## th1  2.73936 2.8128950  0.97386  0.3355
## th2  1.82248 0.0866156 21.04101  0.0000
## th3  1.36932 0.1414898  9.67787  0.0000
## th4 98.13863 1.4919776 65.77755  0.0000
##
##  Correlation:
##     th1    th2    th3
## th2 -0.643
## th3  0.668 -0.299
## th4 -0.228 -0.229 -0.478
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -1.8609042 -0.6092146 -0.2138574  0.4650925  3.1081192
##
## Residual standard error: 5.950669
## Degrees of freedom: 48 total; 44 residual
```

- The slope is positive -> upward slope
- The response ranges from 2 to almost 100
- The alert-dosage (logscaled) is at around 2 (a little smaller than 1)

b)

```
# creating dummy variables
conc.resp.df$in_T1 <- conc.resp.df$treat == "T1"
conc.resp.df$in_T2 <- conc.resp.df$treat == "T2"

# applying model
gnls_treatment_with_dummy <- gnls(resp~(th1+(th4-th1)/(1+(exp((conc-th2)*th3))))*in_T1 +
                                    (th1+(th4-th1)/(1+(exp((conc-th2)*th3))))*in_T2,
                       data = conc.resp.df,
                       params=list(th1+th2+th3+th4~1),
                       control=gnlsControl(nlsTol=0.1),
                       start=c(0, 2, 1, 100))

summary(gnls_treatment_with_dummy)
```

```
## Generalized nonlinear least squares fit
##   Model: resp ~ (th1 + (th4 - th1)/(1 + (exp((conc - th2) * th3)))) *      in_T1 + (th1 + (th4 - th1
##   Data: conc.resp.df
##        AIC      BIC   logLik
##   313.2579 322.6139 -151.629
```

```
##
## Coefficients:
##        Value Std.Error  t-value p-value
## th1  2.73936 2.8128950  0.97386  0.3355
## th2  1.82248 0.0866156 21.04101  0.0000
## th3  1.36932 0.1414898  9.67787  0.0000
## th4 98.13863 1.4919776 65.77755  0.0000
##
##  Correlation:
##     th1    th2    th3
## th2 -0.643
## th3  0.668 -0.299
## th4 -0.228 -0.229 -0.478
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -1.8609042 -0.6092146 -0.2138574  0.4650925  3.1081192
##
## Residual standard error: 5.950669
## Degrees of freedom: 48 total; 44 residual
```

Nothing really changed

c)

```r
gnls_treatment_with_dummy_diff <- gnls(resp~(th1+(th4-th1)/(1+(exp((conc-th21)*th31))))*in_T1 +
                                  (th1+(th4-th1)/(1+(exp((conc-th22)*th32))))*in_T2,
                                 data = conc.resp.df,
                                 params=list(th1+th21+th22+th31+th32+th4~1),
                                 control=gnlsControl(nlsTol=0.1),
                                 start=c(0, 2, 2, 1, 1, 100))

summary(gnls_treatment_with_dummy_diff)
```

```
## Generalized nonlinear least squares fit
##   Model: resp ~ (th1 + (th4 - th1)/(1 + (exp((conc - th21) * th31)))) *      in_T1 + (th1 + (th4 - t
##   Data: conc.resp.df
##        AIC      BIC    logLik
##   306.0199 319.1183 -146.0099
##
## Coefficients:
##         Value Std.Error  t-value p-value
## th1   3.00020 2.5133967  1.19368  0.2393
## th21  1.65162 0.1010026 16.35222  0.0000
## th22  1.94849 0.0865419 22.51499  0.0000
## th31  1.23366 0.1339484  9.20998  0.0000
## th32  1.54790 0.1817802  8.51524  0.0000
## th4  98.34782 1.3582122 72.40977  0.0000
##
##  Correlation:
##      th1    th21   th22   th31   th32
```

```
## th21 -0.546
## th22 -0.506  0.381
## th31  0.560 -0.213 -0.194
## th32  0.533 -0.200 -0.172  0.377
## th4  -0.222 -0.202 -0.195 -0.402 -0.387
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -1.8025729 -0.5448639 -0.1250426  0.4162767  3.3504426
##
## Residual standard error: 5.417861
## Degrees of freedom: 48 total; 42 residual
```

Both alert-concentration and slope have changed after treatment. The model is expected to have a better fit.

d)

```r
anova(gnls_treatment_with_dummy_diff, gnls_treatment)
```

```
##                                Model df      AIC      BIC   logLik   Test
## gnls_treatment_with_dummy_diff     1  7 306.0199 319.1183 -146.010
## gnls_treatment                     2  5 313.2579 322.6139 -151.629 1 vs 2
##                                 L.Ratio p-value
## gnls_treatment_with_dummy_diff
## gnls_treatment                 11.23801  0.0036
```

Yes, the p value is suffiently low to say, that the models are different.

Exercise 6

a)

The explanatory variable is the plant-weight (unit of measure not known). Because no sandwich estimator can be seen, the distribution is assumed to have variance homogeniety.

b)

Because we use a sigmoid-curve, it is assumed that there is a sudden and rapid change in the response.

c)

$c > 0$ implies a positive slope.