

STAT 757 Assignment 8

DUE 4/22/2018 11:59PM

AG Schissler

2/14/2018

Instructions [20 points]

Modify this file to provide responses to the Ch.8 Exercises in Sheather (2009). You can find some helpful code here: <http://www.stat.tamu.edu/~sheather/book/docs/rcode/Chapter8.R>. Also address the project milestones indicated below. Please email **both** your .Rmd (or roxygen .R) and one of the following either .HTML, .PDF, or .DOCX using the format SURNAME-FIRSTNAME-Assignment8.Rmd and SURNAME-FIRSTNAME-Assignment8.pdf.

Exercise 8.3.4 [60 points]

Part A

The marginal model plots show nice agreement for x_2, x_5, \hat{y} . However, x_1, x_4 show poor agreement, especially in the tails. Therefore, the model is invalid for these data as specified.

Part B

Both x_1, x_4 are right-skewed. Following the guidelines by Sheather, I would include $\log(x_1), \log(x_4)$ in the model to account for this.

Part C

All the marginal model plots show nice agreement; I conclude the model is valid.

Part D

When a patient has a family history of heart disease ($x_3 = 1$), the log-odds of heart disease is increased by $\hat{\beta}_3 = 0.941$. This corresponds to an odds ratio of $\exp(\hat{\beta}_3) = 2.5625$.

Appendix

The appendix contains the diagnostics for the model posed in Part B above.

```
myDir <- "~/OneDrive - University of Nevada, Reno/Teaching/STAT_757/Sheather_data/Data/"
dat <- read.delim(file.path(myDir, "HeartDisease.csv"), sep = ",")
str(dat)
```

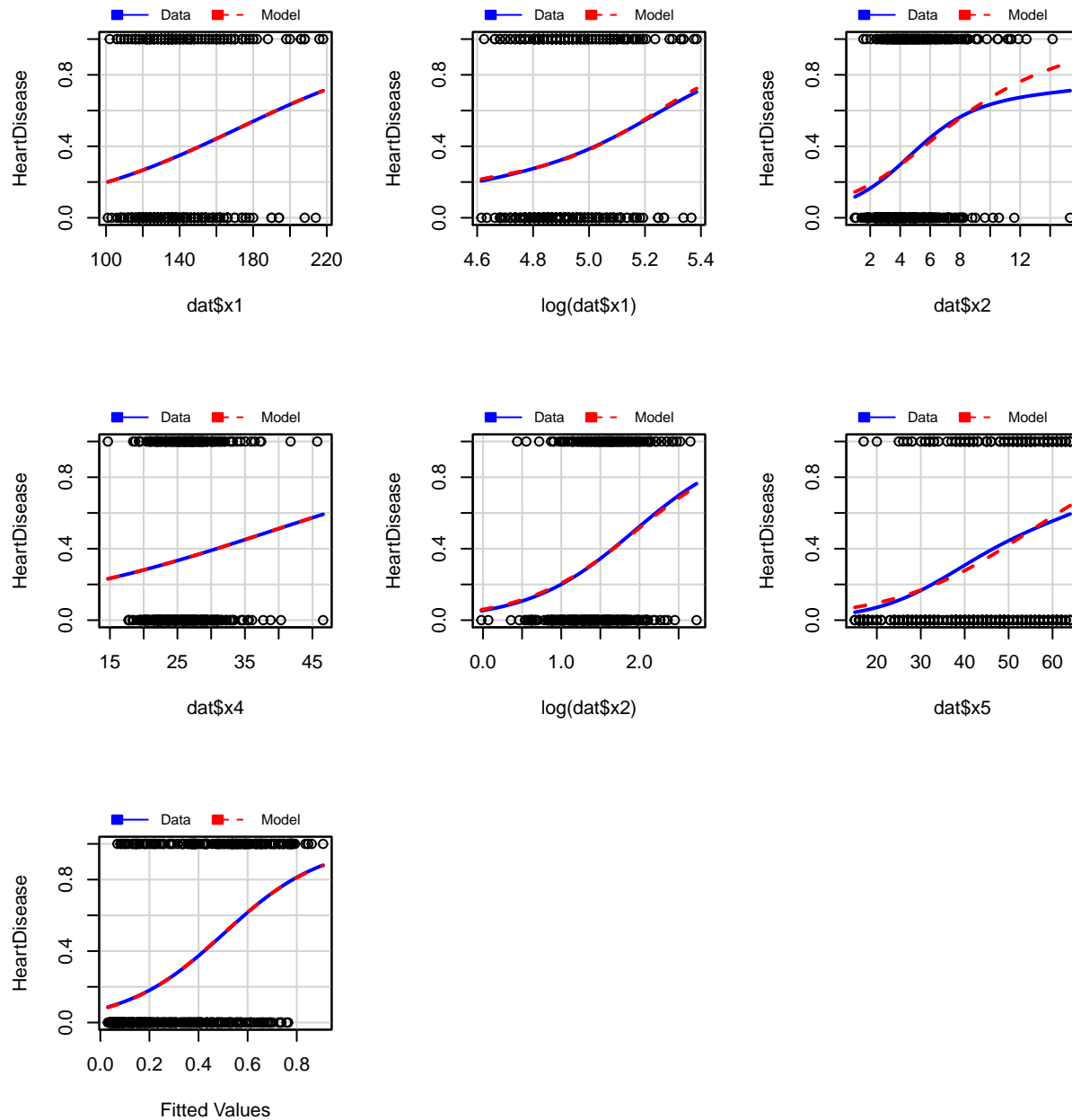
```
## 'data.frame':   462 obs. of  7 variables:
## $ Case         : int  1 2 3 4 5 6 7 8 9 10 ...
## $ x1           : int  160 144 118 170 134 132 142 114 114 132 ...
## $ x2           : num  5.73 4.41 3.48 6.41 3.5 6.47 3.38 4.59 3.83 5.8 ...
## $ x3           : int  1 0 1 1 1 1 0 1 1 1 ...
```

```
## $ x4          : num  25.3 28.9 29.1 32 26 ...
## $ x5          : int   52 63 46 58 49 45 38 58 29 53 ...
## $ HeartDisease: int    1 1 0 1 1 0 0 1 0 1 ...

m2 <- glm(HeartDisease ~ x1 + log(x1) + x2 + x3 + x4 + log(x4) + x5,
          family=binomial() , data=dat )
summary(m2)

##
## Call:
## glm(formula = HeartDisease ~ x1 + log(x1) + x2 + x3 + x4 + log(x4) +
##      x5, family = binomial(), data = dat)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.702   -0.835   -0.455    0.944    2.307
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  75.20477   33.83022    2.22 0.02622
## x1           0.09689    0.05266    1.84 0.06579
## log(x1)     -13.42663    7.77856   -1.73 0.08433
## x2           0.20129    0.05722    3.52 0.00044
## x3           0.94106    0.22427    4.20 2.7e-05
## x4           0.38461    0.20802    1.85 0.06447
## log(x4)     -11.44323    5.70606   -2.01 0.04491
## x5           0.05611    0.00967    5.80 6.6e-09
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 596.11  on 461  degrees of freedom
## Residual deviance: 486.74  on 454  degrees of freedom
## AIC: 502.7
##
## Number of Fisher Scoring iterations: 4

library(alr3)
par(mfrow=c(3,3))
mmp(m2,dat$x1)
mmp(m2,log(dat$x1))
mmp(m2,dat$x2)
mmp(m2,dat$x4)
mmp(m2,log(dat$x2))
mmp(m2,dat$x5)
mmp(m2,m2$fitted.values,xlab="Fitted Values")
```



Project milestones [20 points]

1. Finalize your analysis.
2. Draft your methods section of the written report.

References

Sheather, Simon. 2009. *A Modern Approach to Regression with R*. Springer Science & Business Media.