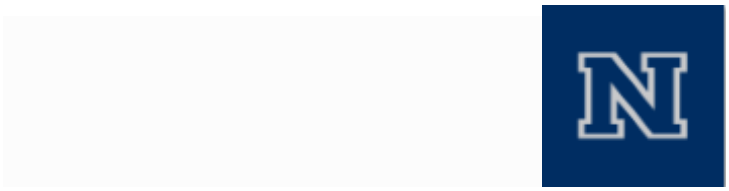
Assignment 8

Biplav Timalsina

April 23, 2018



STAT 757 Applied Regression Analysis

Instructions [20 points]

Modify this file to provide responses to the Ch.8 Exercises in @sheather2009. 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]

4. A number of authors have analyzed the following data on heart disease. Of key interest is the development of a model to determine whether a particular patient has heart disease (i.e., Heart Disease = 1), based on the following predictors:

x_1 = Systolic blood pressure
x_2 = A measure of cholesterol
x_3 = A dummy variable (= 1 for patients with a family history)
x_4 = A measure of obesity and
x_5 = Age.

We first consider the following logistic regression model with these five predictor variables:

$$\theta(\mathbf{x}) = \frac{1}{1+\exp\left(-\left\{\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_4 x_4+\beta_5 x_5\right\}\right)} \tag{8.6}$$

where

$$\theta(x) = E(Y \mid X = x) = P(Y = 1 \mid X = x)$$

Output for model (8.6) is given below along with associated plots (Figures 8.17 and 8.18). The data (HeartDiseare, CSV) can be found on the book web site.

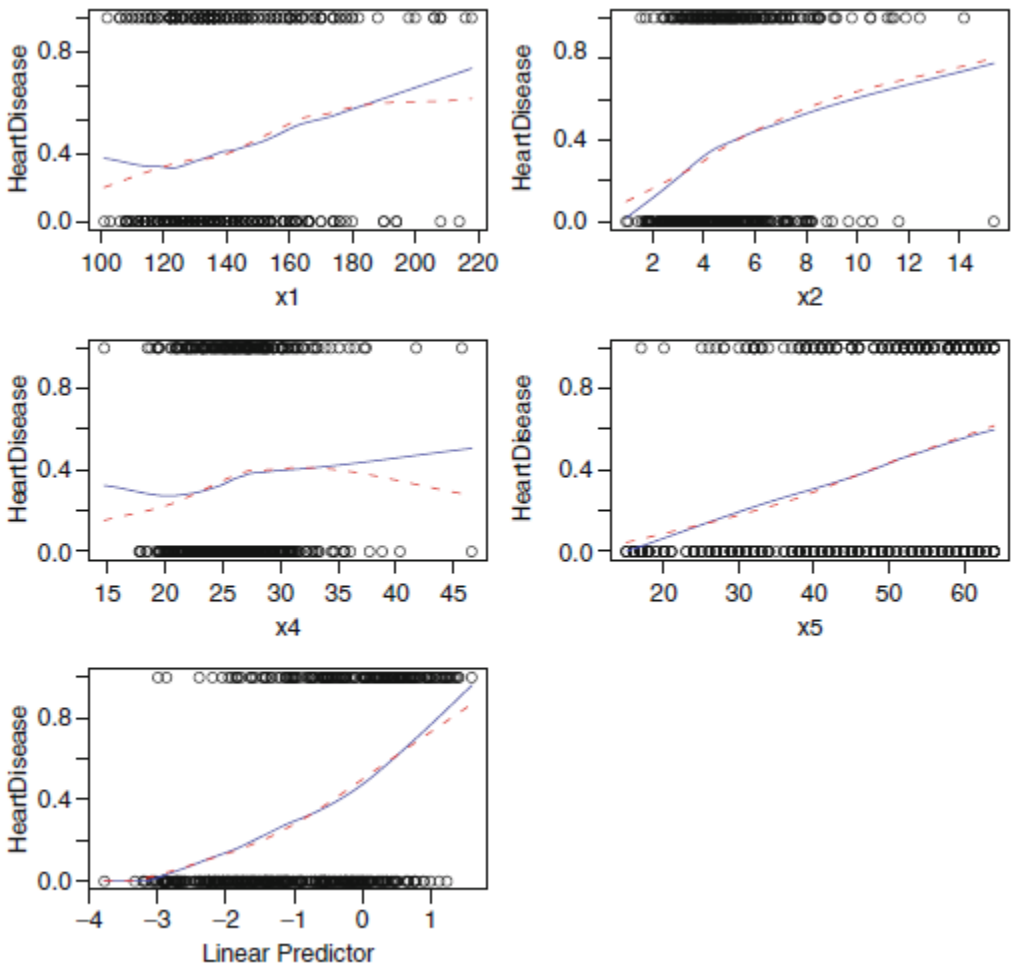


Figure 8.17 Marginal model plots for model (8.6)

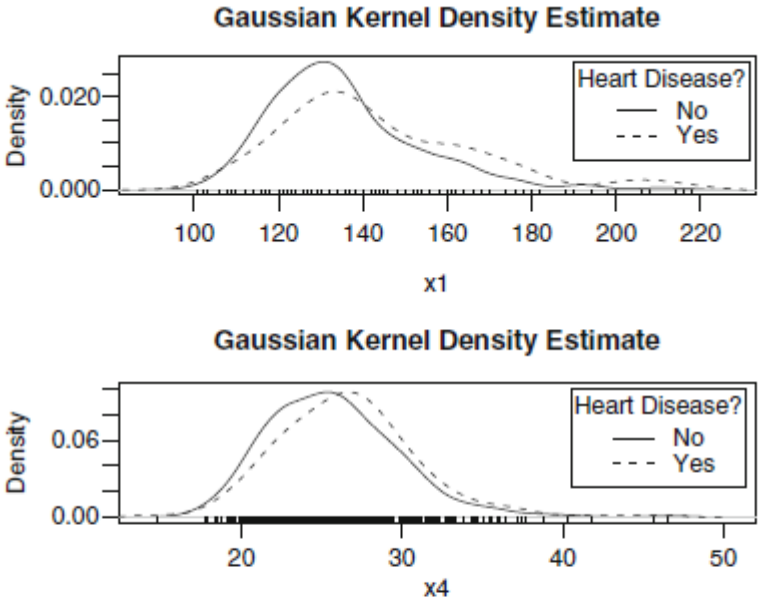


Figure 8.18 Kernel density estimates of x_1 and x_4

(a) Is model (8.6) a valid model for the data? Give reasons to support your answer.
(b) What extra predictor term or terms would you recommend be added to model (8.6) in order to improve it. Please give reasons to support each extra term.

(c) Following your advice in (b), extra predictor terms were added to model (8.6) to form model (8.7). We shall denote these extra predictors as $f_1(x_1)$ and $f_2(x_4)$ (so as not to give away the answer to (b)). Marginal model plots from model (8.7) are shown in Figure 8.19. Is model (8.7) a valid model for the data? Give reasons to support your answer.
(d) Interpret the estimated coefficient of x_3 in model (8.7).

Output from R for model (8.6)

```
Call:
glm(formula = HeartDisease ~ x1 + x2 + x3 + x4 + x5, family =
binomial(), data = HeartDisease)
Coefficients:
             Estimate   Std. Error   z value   Pr(>|z|)
(Intercept) -4.313426     0.943928    -4.570   4.89e-06   ***
x1           0.006435     0.005503     1.169    0.24223
x2           0.186163     0.056325     3.305    0.00095   ***
x3           0.903863     0.221009     4.090   4.32e-05
x4          -0.035640     0.028833    -1.236    0.21643
x5           0.052780     0.009512     5.549   2.88e-08   ***
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 596.11 on 461 degrees of freedom
Residual deviance: 493.62 on 456 degrees of freedom
AIC: 505.62
Number of Fisher Scoring iterations: 4
```

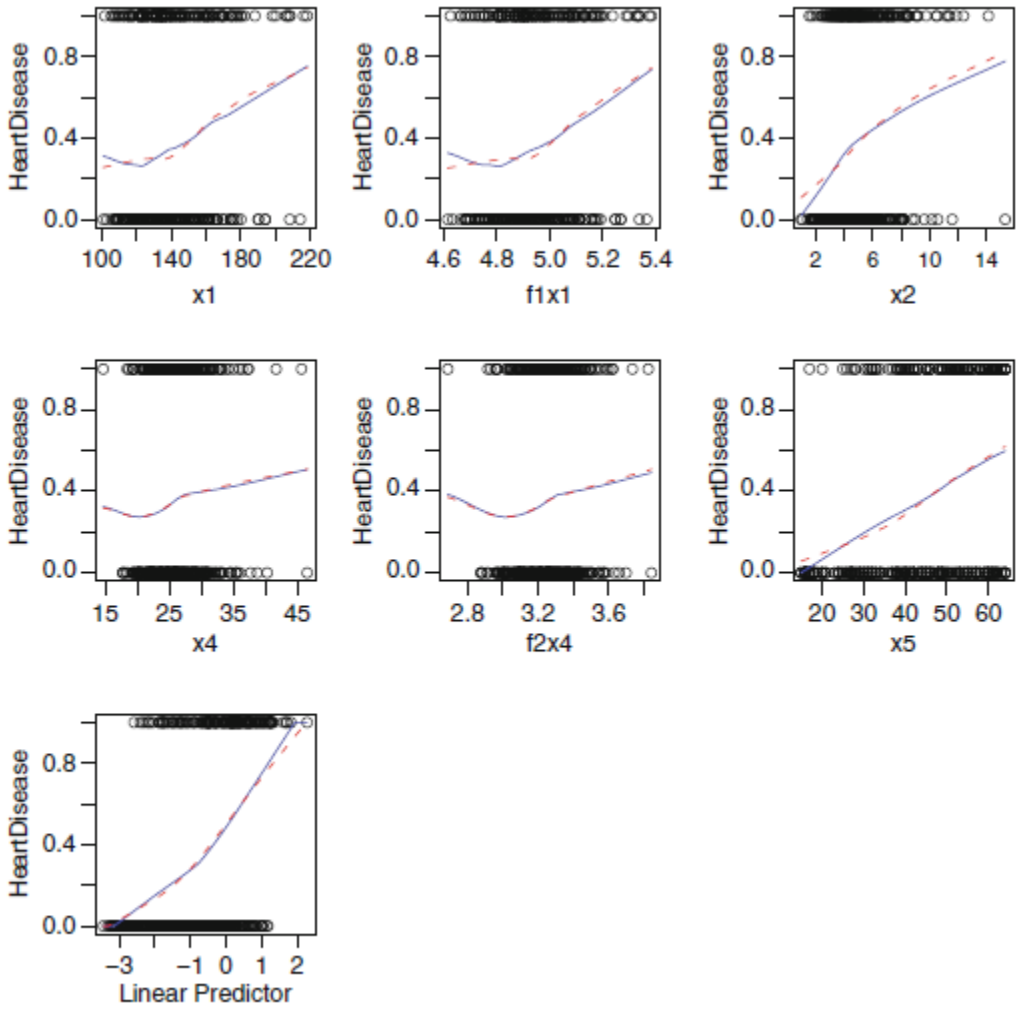


Figure 8.19 Marginal model plots for model (8.7)

Output from R for model (8.7)

```
Call:
glm(formula = HeartDisease ~ x1 + f1x1 + x2 + x3 + x4 + f2x4 +
x5, family = binomial(), data = HeartDisease)
Coefficients:
              Estimate   Std. Error  z value    Pr(>|z|)
(Intercept)   75.204768    33.830217    2.223    0.026215   *
x1             0.096894     0.052664    1.840    0.065792   .
f1x1         -13.426632     7.778559   -1.726    0.084328   .
x2             0.201285     0.057220    3.518    0.000435   ***
x3             0.941056     0.224274    4.196    2.72e-05   ***
x4             0.384608     0.208016    1.849    0.064467   .
f2x4         -11.443233     5.706058   -2.005    0.044915   *
x5             0.056111     0.009675    5.800    6.64e-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: 502.74
Number of Fisher Scoring iterations: 4
```

Project milestones [20 points]

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

References