BST 222 - Homework 5

Xiaowei Zeng

2023-11-07

The addicts data set is from a study by Caplehorn et al. ("Methadone Dosage and Retention of Patients in Maintenance Treatment," Med. J. Aust., 1991) in Sydney, Australia and contains a cohort of 238 heroin addicts who entered maintenance programs between February 1986 and August 1987. The time in the clinic was determined by dates of entry into and exit or transfer from the methadone maintenance program to which they had been assigned. There are two further covariates, namely, prison record and methadone dose, believed to affect the survival times.

The data set and R input code are on the website. The variables are as follows:

- id: Subject ID
- clinic: Clinic (1 or 2)
- status: Survival status (1 = left the clinic, 0 = censored, i.e. we do not know when they left)
- time: Survival time in days
- prison: Prison record (0 = none, 1 = any)
- methadone: Methadone dose (mg/day)

Load the R packages and the dataset.

```
library(survival)
library(KMsurv)
library(MASS)

vars <- c("id","clinic","status","time","prison","methadone")
addicts <- read.table("https://dmrocke.ucdavis.edu/Class/BST222.2023.Fall/addicts.txt",
                      header=F, col.names=vars)

#change variables to factors to be used in Cox PH
addicts$clinic <- factor(addicts$clinic,labels=c("Clinic1","Clinic2"))
addicts$prison <- factor(addicts$prison,labels=c("No","Yes"))

head(addicts)
```

```
##   id  clinic status time prison methadone
## 1  1 Clinic1      1  428     No        50
## 2  2 Clinic1      1  275    Yes        55
## 3  3 Clinic1      1  262     No        55
## 4  4 Clinic1      1  183     No        30
## 5  5 Clinic1      1  259    Yes        65
## 6  6 Clinic1      1  714     No        55
```

```
# Create a survival object
dfsurv <- Surv(addicts$time, addicts$status)
```

1. Fit the Cox model with clinics as strata, prison and methadone as variables.

1.1. Separate Models.

Let's check each strata separately: separate by the clinic variable that has issues with proportionality and fit the same Cox model in each group.

```
cox1 <- coxph(dfsurv ~ methadone + prison, data = addicts, sub = (clinic == 'Clinic1'))
summary(cox1)
```

```
## Call:
## coxph(formula = dfsurv ~ methadone + prison, data = addicts,
##     subset = (clinic == "Clinic1"))
##
##   n= 163, number of events= 122
##
##                coef exp(coef)  se(coef)      z Pr(>|z|)
## methadone -0.035799  0.964834  0.007738 -4.626 3.72e-06 ***
## prisonYes  0.502846  1.653421  0.188706  2.665  0.00771 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## methadone    0.9648     1.0364    0.9503    0.9796
## prisonYes    1.6534     0.6048    1.1422    2.3934
##
## Concordance= 0.647  (se = 0.028 )
## Likelihood ratio test= 26.05  on 2 df,   p=2e-06
## Wald test            = 25.11  on 2 df,   p=4e-06
## Score (logrank) test = 25.57  on 2 df,   p=3e-06
```

```
cox2 <- coxph(dfsurv ~ methadone + prison, data = addicts, sub = (clinic == 'Clinic2'))
summary(cox2)
```

```
## Call:
## coxph(formula = dfsurv ~ methadone + prison, data = addicts,
##     subset = (clinic == "Clinic2"))
##
##   n= 75, number of events= 28
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## methadone -0.03696   0.96371  0.01235 -2.994  0.00275 **
## prisonYes -0.08014   0.92298  0.38430 -0.209  0.83481
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## methadone    0.9637      1.038    0.9407    0.9873
## prisonYes    0.9230      1.083    0.4346    1.9603
##
## Concordance= 0.663  (se = 0.057 )
## Likelihood ratio test= 9.72  on 2 df,   p=0.008
## Wald test            = 8.98  on 2 df,   p=0.01
## Score (logrank) test = 9.41  on 2 df,   p=0.009
```

The coefficients of methadone in the models are close, while the coefficients of prison look different. We want to test whether they are statistically different.

1.2. Interaction Model.

We can test the assumption of the same regression coefficients for each stratum by testing the interaction between the stratification variable and the covariates.

```
inter.cox <- coxph(dfsurv ~ (methadone + prison) * strata(clinic), data = addicts)
summary(inter.cox)
```

```
## Call:
## coxph(formula = dfsurv ~ (methadone + prison) * strata(clinic),
##     data = addicts)
##
##   n= 238, number of events= 150
##
##                                      coef exp(coef)  se(coef)      z Pr(>|z|)
## methadone                        -0.035799  0.964834  0.007738 -4.626 3.72e-06
## prisonYes                         0.502846  1.653421  0.188706  2.665  0.00771
## methadone:strata(clinic)Clinic2 -0.001164  0.998837  0.014570 -0.080  0.93632
## prisonYes:strata(clinic)Clinic2 -0.582989  0.558227  0.428135 -1.362  0.17329
##
## methadone                        ***
## prisonYes                        **
## methadone:strata(clinic)Clinic2
## prisonYes:strata(clinic)Clinic2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                  exp(coef) exp(-coef) lower .95 upper .95
## methadone                           0.9648     1.0364    0.9503    0.9796
## prisonYes                           1.6534     0.6048    1.1422    2.3934
## methadone:strata(clinic)Clinic2     0.9988     1.0012    0.9707    1.0278
## prisonYes:strata(clinic)Clinic2     0.5582     1.7914    0.2412    1.2919
##
## Concordance= 0.649  (se = 0.026 )
## Likelihood ratio test= 35.77  on 4 df,   p=3e-07
## Wald test            = 34.09  on 4 df,   p=7e-07
## Score (logrank) test = 34.97  on 4 df,   p=5e-07
```

The coefficients of the interaction terms are not statistically significant at the significance level of 0.05, so we can assume that the coefficients are the same across strata.

1.3. Stratified Model.

```
cox <- coxph(dfsurv ~ strata(clinic) + methadone + prison, data = addicts)
summary(cox)
```

```
## Call:
```

```
## coxph(formula = dfsurv ~ strata(clinic) + methadone + prison,
##     data = addicts)
##
##   n= 238, number of events= 150
##
##                coef exp(coef)  se(coef)      z Pr(>|z|)
## methadone -0.035115  0.965495  0.006465 -5.432 5.59e-08 ***
## prisonYes  0.389605  1.476397  0.168930  2.306   0.0211 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## methadone    0.9655     1.0357    0.9533    0.9778
## prisonYes    1.4764     0.6773    1.0603    2.0559
##
## Concordance= 0.651  (se = 0.026 )
## Likelihood ratio test= 33.91  on 2 df,   p=4e-08
## Wald test            = 32.66  on 2 df,   p=8e-08
## Score (logrank) test = 33.33  on 2 df,   p=6e-08
```

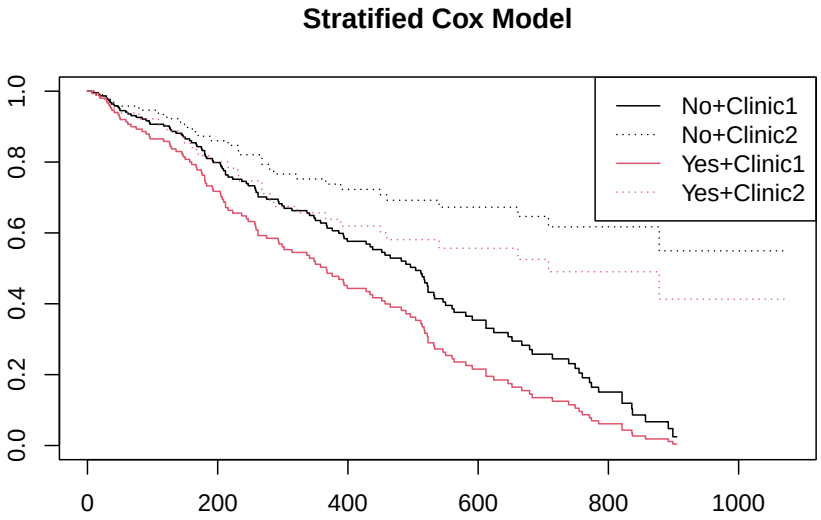
Check for a difference between the interaction model and the stratified model.

```
anova(cox, inter.cox, test = 'Chisq')
```

```
## Analysis of Deviance Table
##  Cox model: response is  dfsurv
##  Model 1: ~ strata(clinic) + methadone + prison
##  Model 2: ~ (methadone + prison) * strata(clinic)
##    loglik Chisq Df P(>|Chi|)
## 1 -597.52
## 2 -596.59 1.859  2    0.3948
```

We get a p-value of 0.3948 which means that there is no significant difference between the stratified model with and without interaction terms. It shows that fitting interactions with the strata variable cannot improve the model (via the likelihood ratio test). We can conclude that the coefficients are not significantly different between the stratified models and the only difference is in the baseline hazard. So it is appropriate to use a stratified model with coefficients that are the same across strata.

```
plot(survfit(cox, data.frame(methadone = mean(addicts$methadone), prison = c('No', 'Yes'))),
     col = c(1, 1, 2, 2), lty = c(1, 3, 1, 3), main = 'Stratified Cox Model')
legend('topright',c('No+Clinic1', 'No+Clinic2', 'Yes+Clinic1', 'Yes+Clinic2'),
       col = c(1, 1, 2, 2), lty = c(1, 3, 1, 3))
```



We can see that they are proportional within each strata, the two Clinic1 curves are proportional and the two Clinic2 curves are proportional. We will use this model for the remaining questions. Suppose the significance level is 0.05.

2. Test proportionality of hazards.

2.1. Use `cox.zph` to test PH assumption.

```
addicts.zph <- cox.zph(cox)
print(addicts.zph)
```

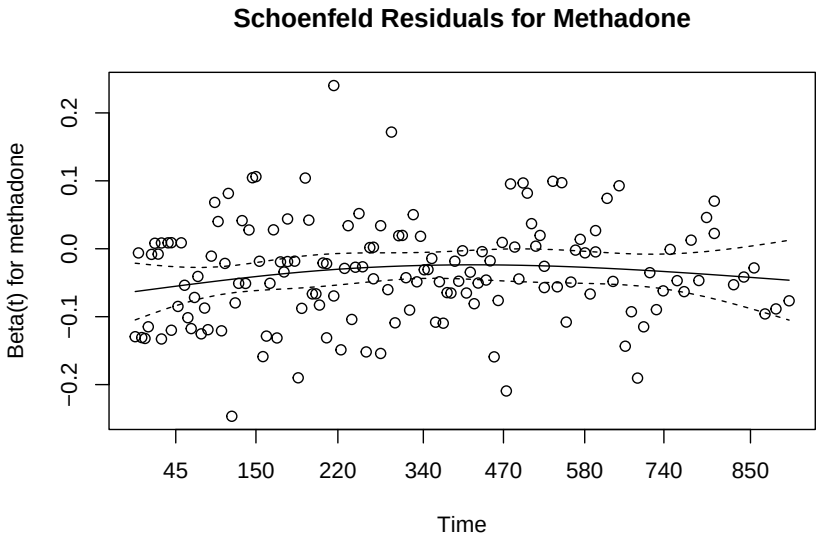
```
##            chisq df    p
## methadone 1.0552  1 0.30
## prison    0.0264  1 0.87
## GLOBAL    1.1302  2 0.57
```

All the p-values in the `zph` table are larger than 0.05, so we can assume that these covariates meet the PH assumption.

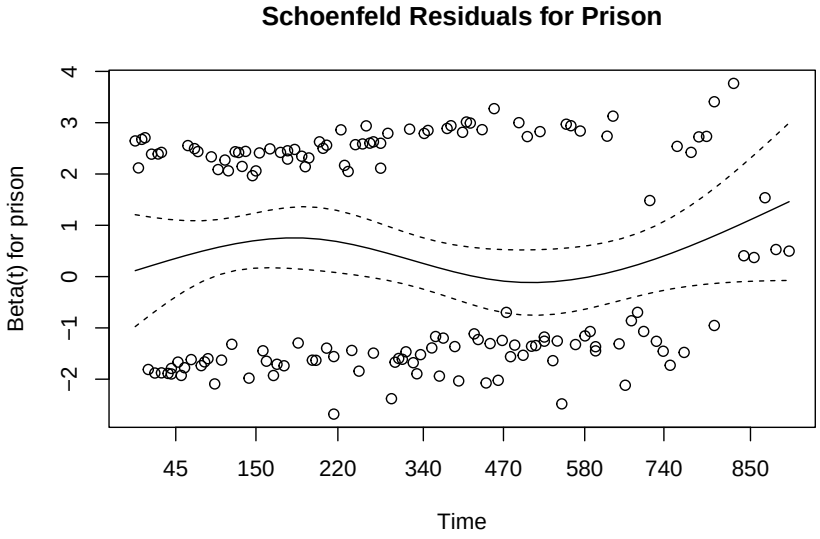
2.2. Make the plots vs. time of the Schoenfeld residuals.

There are two non-stratified covariates in the model, so we will have two plots.

```
plot(addicts.zph[1], main = 'Schoenfeld Residuals for Methadone')
```




```
plot(addicts.zph[2], main = 'Schoenfeld Residuals for Prison')
```



The residuals for methadone and prison are roughly uncorrelated with time and the pattern of residuals vs time is nearly random, so we can assume that these covariates meet the PH assumption.

Then have a quick look at their correlation with time.

```
cor.spear <- function(x) {cor.test(x, addicts.zph$x, method = "spearman")}
correlations <- apply(X = addicts.zph$y, MARGIN = 2, FUN = cor.spear); correlations
```

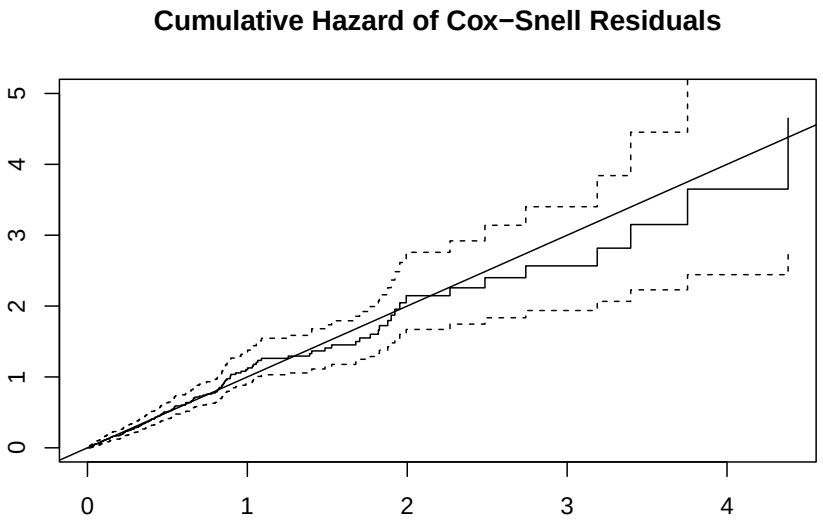
```
## $methadone
##
##  Spearman's rank correlation rho
##
## data:  x and addicts.zph$x
## S = 505263, p-value = 0.2155
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
```

```
##       rho
## 0.1017139
##
##
## $prison
##
##  Spearman's rank correlation rho
##
## data:  x and addicts.zph$x
## S = 480627, p-value = 0.07561
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.1455135
```

The p-values in the correlation test are all larger than 0.05, so we can assume that there's no time dependence.

3. Plot the cumulative hazard of the Cox-Snell residuals.

```
addicts.mart <- residuals(cox, type = 'martingale')
addicts.cs <- addicts$status - addicts.mart
surv.csr <- survfit(Surv(addicts.cs, addicts$status) ~ 1, type = 'fleming-harrington')
plot(surv.csr, fun = 'cumhaz', ylim = c(0, 5)); abline(0, 1)
title('Cumulative Hazard of Cox-Snell Residuals')
```

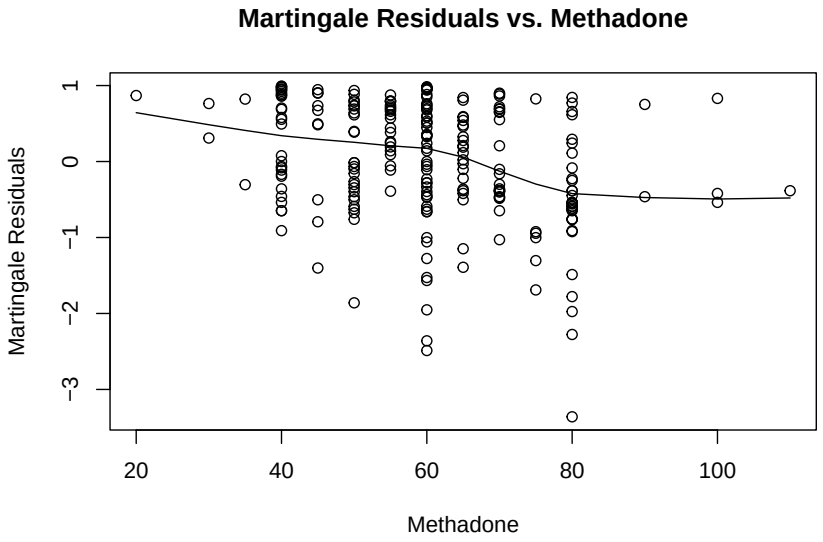


The line with slope 1 and intercept 0 fits the curve relatively well, so we don't see lack of goodness of fit based on these residuals.

4. Examine the functional form of methadone.

Plot the martingale residuals (omitting methadone) vs. methadone.

```
mres <- residuals(coxph(dfsurv ~ strata(clinic) + prison, data = addicts), type = 'martingale')
plot(addicts$methadone, mres, xlab = 'Methadone', ylab = 'Martingale Residuals')
lines(lowess(addicts$methadone, mres))
title('Martingale Residuals vs. Methadone')
```

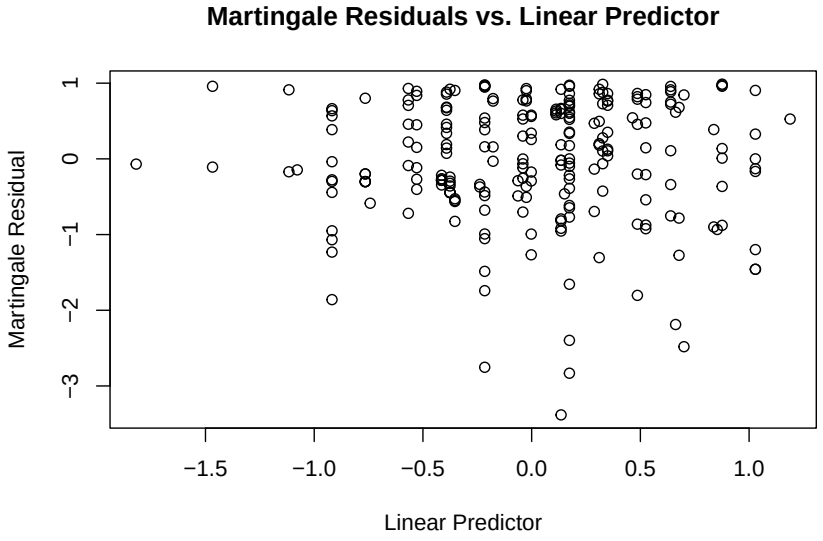


The slope of the line changes dramatically between methadone = 60 and 80. It suggests that some transformation of Methadone, e.g. treating it as a categorical variable, may help improve the goodness of fit.

5. Other model diagnostics.

Plot the martingale residuals vs. the linear predictor.

```
addicts.preds <- predict(cox) # linear predictor
plot(addicts.preds, addicts.mart, xlab = 'Linear Predictor', ylab = 'Martingale Residual')
title('Martingale Residuals vs. Linear Predictor')
```



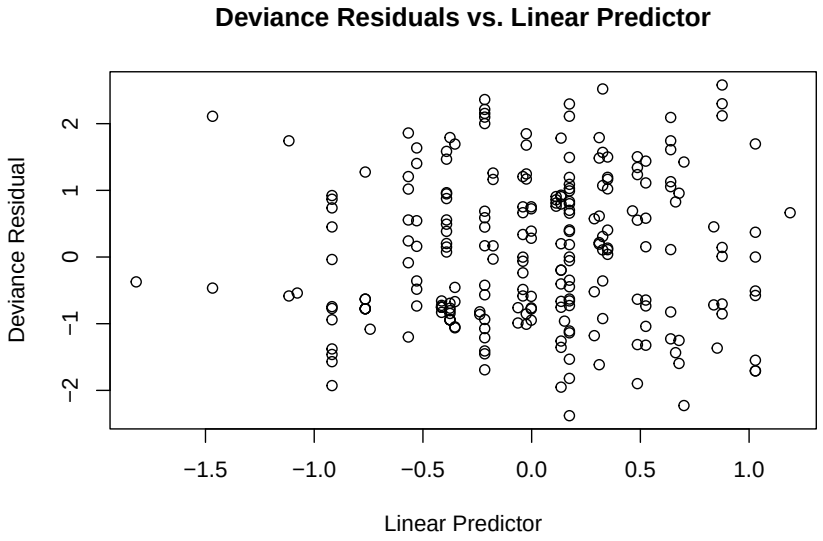
```
addicts.mart[addicts.mart < - 2.2]
```

```
##         8         9        12        26        54
## -2.831624 -3.382002 -2.397699 -2.481308 -2.753101
```

The smallest five martingale residuals in order are observations 8, 9, 12, 26 and 54.

Plot the deviance residuals vs. the linear predictor.

```
addicts.dev <- residuals(cox, type = 'deviance')
plot(addicts.preds, addicts.dev, xlab = 'Linear Predictor', ylab = 'Deviance Residual')
title('Deviance Residuals vs. Linear Predictor')
```



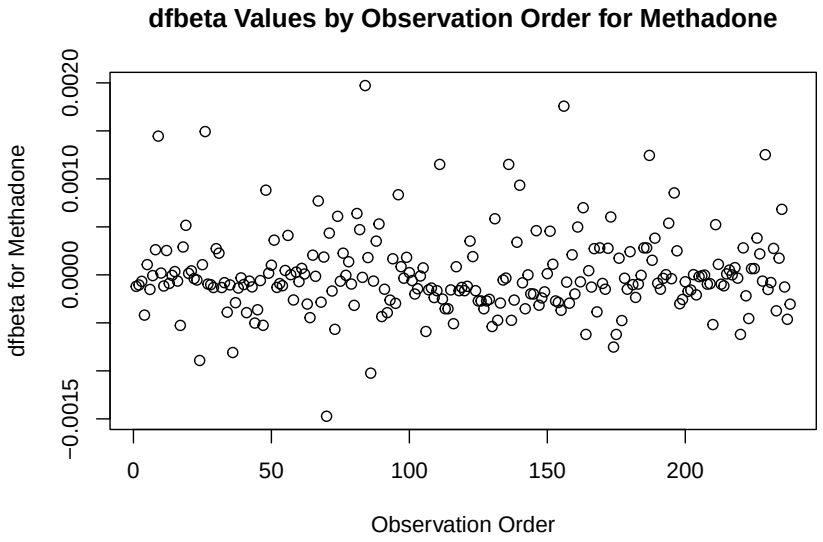

```
addicts.dev[abs(addicts.dev) > 2.3]
```

```
##         8       123       175       185
## -2.379758  2.520641  2.580119  2.361883
```

The four largest absolute deviance residuals are observations 8, 123, 175 and 185.

Plot the two dfbeta values by observation order.

```
addicts.dfb <- residuals(cox, type = 'dfbeta')
plot(addicts.dfb[, 1], xlab = 'Observation Order', ylab = 'dfbeta for Methadone')
title('dfbeta Values by Observation Order for Methadone')
```

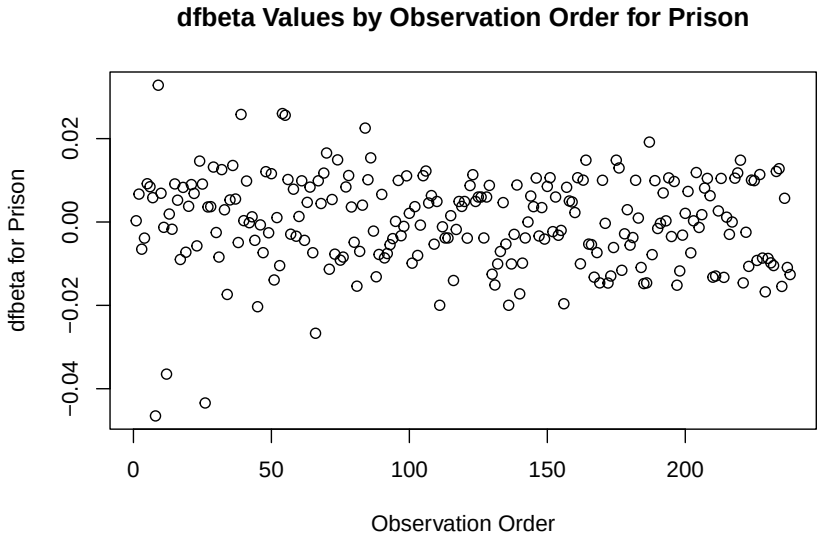



```
addicts.dfb[abs(addicts.dfb[, 1]) > 0.0014, 1]
```

```
##           9          26          70          84         156
##  0.001445286  0.001492789 -0.001472983  0.001971976  0.001757395
```

The five largest absolute dfbetas for methadone are observations 9, 26, 70, 84 and 156.

```
plot(addicts.dfb[, 2], xlab = 'Observation Order', ylab = 'dfbeta for Prison')
title('dfbeta Values by Observation Order for Prison')
```




```
addicts.dfb[abs(addicts.dfb[, 2]) > 0.03, 2]
```

```
##           8           9          12          26
## -0.04652025  0.03281088 -0.03649003 -0.04341864
```

The four largest absolute dfbetas for prison are observations 8, 9, 12 and 26.

Table: Observations to Examine by Residuals and Influence

Model Diagnostics	Observations
Martingale Residuals	8, 9, 12, 26, 54
Deviance Residuals	8, 123, 175, 185
Methadone Influence	9, 26, 70, 84, 156
Prison Influence	8, 9, 12, 26

So the most important observations to examine seem to be 8, 9, 12 and 26.

```
with(addicts, summary(time[status==1]))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     7.0   157.8   283.0   345.5   521.0   899.0
```

```
with(addicts, summary(methadone))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    20.0    50.0    60.0    60.4    70.0   110.0
```

```
with(addicts, summary(prison))
```

```
##  No Yes
## 127 111
```

```
addicts[c(8, 9, 12, 26), -1]
```

```
##     clinic status time prison methadone
## 8  Clinic1      0  796    Yes        60
## 9  Clinic1      1  892     No        50
## 12 Clinic1      1  836    Yes        60
## 26 Clinic1      0  566    Yes        45
```

- 8: Clinic1, keeping staying in clinic, with prison record, median methadone doze.
- 9: Clinic1, leaving clinic late, without prison record, low methadone doze.
- 12: Clinic1, leaving clinic late, with prison record, median methadone doze.
- 26: Clinic1, keeping staying in clinic, with prison record, low methadone doze.

It turns out that nearly all the unusual observations are from Clinic 1, have features of leaving the clinic late or not leaving, having at least one prison record and methadone doze below average. But this can actually happen in real case.