Development And Validation Of Risk Prediction Model For Adverse Birth Outcomes Among Mothers Conceiving Through Assisted Reproductive Technology In Saint Paul Specialized Hospital, Addis Ababa, Central Ethiopia

Sydney Getinet (MPH in Epidemiology)

2025-06-12

Contents

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
################################################################################
################################################################################
######        Development and Validation of a Clinical          ######
######        Prognostic Risk Score to Predict Early Neonatal    ######
######        Mortality, Ethiopia: A Receiver Operating          ######
######        Characteristic Curve AnalysisDevelopment           ######
######        and validation of dynamic risk prediction  model   ######
######        weibull AFT survival model                         ######
################################################################################
################################################################################

library(haven)
finaldata = read_dta("C:/Users/hp/Documents/Materials/Thesis file/Data/fin/updated/420/New folder/New fo

# Load necessary libraries
library(glmnet)
library(caret)
names(finaldata)
```

```
## [1] "id"          "apo"         "lbw"         "prebirth"
## [5] "stillbirth"  "conganomal"  "lapscore"    "age"
```

```
##  [9] "bmi"             "healthinsu"      "placeresid"      "edustat"
## [13] "parity"          "twin"            "childgender"     "hxcs"
## [17] "matcomo"         "prliveb"         "fullbirth"       "preARTattempt"
## [21] "yrstayedinfert"  "iugr"            "hxabortion"      "bleedepisode"
## [25] "prevutrinesurg"  "prom"            "infcause"        "alcoholdri"
## [29] "cigsmoke"        "afc"             "cycletype"       "hyperstimuepi"
## [33] "embryotype"      "embryostage"     "numembryotrans"  "assistedhatch"
## [37] "hdp"             "dmhx"            "predmhx"         "reparity"
## [41] "ga"              "bmireclass"      "infcausereclass" "agegrp"
## [45] "bmicont"         "gestage"
```

```r
library(caret)

set.seed(123)
train_index <- createDataPartition(finaldata$apo, p = 0.8, list = FALSE)

train_data <- finaldata[train_index, ]
test_data  <- finaldata[-train_index, ]

# Check proportions
prop.table(table(train_data$apo))
```

```
##
##         0         1
## 0.3584906 0.6415094
```

```r
prop.table(table(test_data$apo))
```

```
##
##         0         1
## 0.3417722 0.6582278
```

```r
# Your original code with minor improvements
exclude_vars <- c("id", "apo", "ga","lbw", "prebirth", "stillbirth", "conganomal", "lapscore")
predictor_names <- setdiff(names(train_data), exclude_vars)

# Convert to factors (safer approach)
train_data[predictor_names] <- lapply(train_data[predictor_names], function(x) {
  if(is.numeric(x) & length(unique(x)) <= 5) as.factor(x) else x
})


# Convert to factors (safer approach)
test_data[predictor_names] <- lapply(test_data[predictor_names], function(x) {
  if(is.numeric(x) & length(unique(x)) <= 5) as.factor(x) else x
})

# Ensure outcome is binary factor with proper level names
train_data$apo <- factor(train_data$apo, levels = c(0,1), labels = c("No", "Yes"))
table(test_data$apo)
```

```
##
```

```
## 0  1
## 27 52
```

```r
# Handle missing data (critical for stepwise)
data_complete <- na.omit(train_data[, c("apo", predictor_names)])

data_complete$apo = as.factor(data_complete$apo)

table(data_complete$apo)
```

```
##
## No Yes
## 114 204
```

```r
################################################################################
################################################################################
##                        Model Building Process                            ##
################################################################################
################################################################################

# Fit full model
full_model <- glm(apo ~ agegrp + bmi + healthinsu + placeresid + edustat +
                  reparity + twin + childgender + hxcs + matcomo + prliveb +
                  fullbirth + preARTattempt + yrstayedinfert + iugr + hxabortion +
                  bleedepisode + prevutrinesurg + prom + infcause + alcoholdri +
                  cigsmoke + afc + cycletype + embryotype + embryostage +
                  numembryotrans + assistedhatch + hdp + dmhx + predmhx,
              data = data_complete,
              family = binomial(link = "logit"))

# Stepwise selection
reduced_model <- step(full_model,
                 direction = "backward",
                 trace = 1)  # Keep trace=1 to see
```

```
## Start:  AIC=230.84
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + hxcs + matcomo + prliveb + fullbirth +
##     preARTattempt + yrstayedinfert + iugr + hxabortion + bleedepisode +
##     prevutrinesurg + prom + infcause + alcoholdri + cigsmoke +
##     afc + cycletype + embryotype + embryostage + numembryotrans +
##     assistedhatch + hdp + dmhx + predmhx
##
##                  Df Deviance    AIC
## - prom            2   153.72 227.72
## - bleedepisode    2   153.82 227.82
## - matcomo         1   152.85 228.85
## - hdp             1   152.93 228.93
## - prevutrinesurg  1   152.96 228.96
## - preARTattempt   1   153.04 229.04
## - predmhx         1   153.22 229.22
## - prliveb         1   153.44 229.44
## - fullbirth       1   153.67 229.67
```

3

```
## - hxabortion        1    153.71 229.71
## - bmi               3    157.96 229.96
## - cycletype         1    154.02 230.02
## - reparity          1    154.18 230.18
## - numembryotrans    1    154.71 230.71
## - hxcs              1    154.78 230.78
## <none>                   152.84 230.84
## - assistedhatch     1    155.49 231.49
## - agegrp            1    156.01 232.01
## - healthinsu        1    156.17 232.17
## - iugr              1    156.23 232.23
## - alcoholdri        1    156.94 232.94
## - cigsmoke          1    157.01 233.01
## - edustat           2    159.51 233.51
## - childgender       1    160.46 236.46
## - yrstayedinfert    2    163.01 237.01
## - placeresid        1    162.07 238.07
## - dmhx              1    165.84 241.84
## - twin              1    166.97 242.97
## - embryotype        1    167.62 243.62
## - infcause          2    170.03 244.03
## - afc               1    179.38 255.38
## - embryostage       1    186.76 262.76
##
## Step:  AIC=227.73
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + hxcs + matcomo + prliveb + fullbirth +
##     preARTattempt + yrstayedinfert + iugr + hxabortion + bleedepisode +
##     prevutrinesurg + infcause + alcoholdri + cigsmoke + afc +
##     cycletype + embryotype + embryostage + numembryotrans + assistedhatch +
##     hdp + dmhx + predmhx
##
##                     Df Deviance    AIC
## - bleedepisode       2    154.95 224.95
## - matcomo            1    153.75 225.75
## - hdp                1    153.81 225.81
## - preARTattempt      1    153.89 225.89
## - prevutrinesurg     1    153.92 225.92
## - predmhx            1    154.00 226.00
## - prliveb            1    154.40 226.40
## - fullbirth          1    154.46 226.46
## - hxabortion         1    154.64 226.64
## - bmi                3    158.78 226.78
## - cycletype          1    154.82 226.82
## - reparity           1    154.94 226.94
## - numembryotrans     1    155.60 227.60
## <none>                   153.72 227.72
## - hxcs               1    155.78 227.78
## - assistedhatch      1    156.32 228.32
## - iugr               1    156.86 228.86
## - agegrp             1    157.18 229.18
## - healthinsu         1    157.31 229.31
## - alcoholdri         1    157.34 229.34
## - cigsmoke           1    157.50 229.50
```

4

```
## - edustat          2   160.91 230.91
## - childgender       1   161.22 233.22
## - yrstayedinfert    2   163.53 233.53
## - placeresid        1   162.76 234.76
## - dmhx              1   166.40 238.40
## - twin              1   167.35 239.35
## - embryotype        1   168.67 240.67
## - infcause          2   171.60 241.60
## - afc               1   180.61 252.61
## - embryostage       1   187.52 259.52
##
## Step:  AIC=224.95
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + hxcs + matcomo + prliveb + fullbirth +
##     preARTattempt + yrstayedinfert + iugr + hxabortion + prevutrinesurg +
##     infcause + alcoholdri + cigsmoke + afc + cycletype + embryotype +
##     embryostage + numembryotrans + assistedhatch + hdp + dmhx +
##     predmhx
##
##                    Df Deviance    AIC
## - prevutrinesurg   1   154.97 222.97
## - matcomo          1   154.97 222.97
## - hdp              1   154.99 222.99
## - preARTattempt    1   155.06 223.06
## - predmhx          1   155.18 223.18
## - hxabortion       1   155.57 223.57
## - prliveb          1   155.58 223.58
## - cycletype        1   155.76 223.76
## - fullbirth        1   155.92 223.92
## - bmi              3   160.34 224.34
## - reparity         1   156.40 224.40
## - hxcs             1   156.55 224.55
## - numembryotrans   1   156.75 224.75
## <none>                 154.95 224.95
## - assistedhatch    1   157.77 225.77
## - iugr             1   158.09 226.09
## - alcoholdri       1   158.37 226.37
## - healthinsu       1   158.39 226.39
## - cigsmoke         1   158.50 226.50
## - agegrp           1   159.14 227.14
## - edustat          2   161.84 227.84
## - yrstayedinfert   2   164.32 230.32
## - childgender      1   162.47 230.47
## - placeresid       1   163.18 231.18
## - dmhx             1   167.59 235.59
## - twin             1   168.05 236.05
## - embryotype       1   169.00 237.00
## - infcause         2   172.72 238.72
## - afc              1   180.90 248.90
## - embryostage      1   188.66 256.66
##
## Step:  AIC=222.96
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + hxcs + matcomo + prliveb + fullbirth +
```

```
##       preARTattempt + yrstayedinfert + iugr + hxabortion + infcause +
##       alcoholdri + cigsmoke + afc + cycletype + embryotype + embryostage +
##       numembryotrans + assistedhatch + hdp + dmhx + predmhx
##
##                       Df Deviance    AIC
## - matcomo             1   154.98 220.98
## - hdp                 1   155.00 221.00
## - preARTattempt       1   155.06 221.06
## - predmhx             1   155.20 221.20
## - hxabortion          1   155.59 221.59
## - prliveb             1   155.60 221.60
## - cycletype           1   155.79 221.79
## - fullbirth           1   155.92 221.92
## - bmi                 3   160.35 222.35
## - reparity            1   156.40 222.40
## - hxcs                1   156.55 222.55
## - numembryotrans      1   156.76 222.76
## <none>                    154.97 222.97
## - assistedhatch       1   157.90 223.90
## - iugr                1   158.09 224.09
## - healthinsu          1   158.40 224.40
## - alcoholdri          1   158.49 224.49
## - cigsmoke            1   158.56 224.56
## - agegrp              1   159.14 225.14
## - edustat             2   161.88 225.88
## - childgender         1   162.48 228.48
## - yrstayedinfert      2   164.48 228.48
## - placeresid          1   163.32 229.32
## - dmhx                1   167.66 233.66
## - twin                1   168.14 234.14
## - embryotype          1   169.01 235.01
## - infcause            2   172.74 236.74
## - afc                 1   180.97 246.97
## - embryostage         1   188.78 254.78
##
## Step:  AIC=220.98
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##       twin + childgender + hxcs + prliveb + fullbirth + preARTattempt +
##       yrstayedinfert + iugr + hxabortion + infcause + alcoholdri +
##       cigsmoke + afc + cycletype + embryotype + embryostage + numembryotrans +
##       assistedhatch + hdp + dmhx + predmhx
##
##                       Df Deviance    AIC
## - hdp                 1   155.02 219.02
## - preARTattempt       1   155.08 219.08
## - predmhx             1   155.22 219.22
## - hxabortion          1   155.61 219.61
## - prliveb             1   155.63 219.63
## - cycletype           1   155.80 219.80
## - fullbirth           1   155.92 219.92
## - bmi                 3   160.35 220.35
## - reparity            1   156.41 220.41
## - hxcs                1   156.55 220.55
## - numembryotrans      1   156.76 220.76
```

```
## <none>                    154.98 220.98
## - assistedhatch  1    157.91 221.91
## - iugr           1    158.14 222.14
## - alcoholdri     1    158.52 222.52
## - healthinsu     1    158.56 222.56
## - cigsmoke       1    158.58 222.58
## - agegrp         1    159.21 223.21
## - edustat        2    161.92 223.92
## - yrstayedinfert 2    164.48 226.48
## - childgender    1    162.80 226.80
## - placeresid     1    163.32 227.32
## - dmhx           1    167.68 231.68
## - twin           1    168.41 232.41
## - embryotype     1    169.15 233.15
## - infcause       2    172.75 234.75
## - afc            1    181.46 245.46
## - embryostage    1    188.79 252.79
##
## Step:  AIC=219.02
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##      twin + childgender + hxcs + prliveb + fullbirth + preARTattempt +
##      yrstayedinfert + iugr + hxabortion + infcause + alcoholdri +
##      cigsmoke + afc + cycletype + embryotype + embryostage + numembryotrans +
##      assistedhatch + dmhx + predmhx
##
##                  Df Deviance    AIC
## - preARTattempt  1    155.11 217.11
## - predmhx        1    155.25 217.25
## - prliveb        1    155.63 217.63
## - hxabortion     1    155.64 217.64
## - cycletype      1    155.81 217.81
## - fullbirth      1    156.01 218.01
## - bmi            3    160.44 218.44
## - hxcs           1    156.56 218.56
## - reparity       1    156.60 218.60
## - numembryotrans 1    156.79 218.79
## <none>                155.02 219.02
## - assistedhatch  1    157.91 219.91
## - iugr           1    158.18 220.18
## - alcoholdri     1    158.53 220.53
## - cigsmoke       1    158.58 220.58
## - healthinsu     1    158.62 220.62
## - agegrp         1    159.25 221.25
## - edustat        2    161.95 221.95
## - yrstayedinfert 2    164.54 224.54
## - childgender    1    163.02 225.02
## - placeresid     1    163.43 225.43
## - dmhx           1    167.78 229.78
## - twin           1    168.48 230.48
## - embryotype     1    169.20 231.20
## - infcause       2    172.81 232.81
## - afc            1    181.69 243.69
## - embryostage    1    189.07 251.07
##
```

```
## Step:  AIC=217.11
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + hxcs + prliveb + fullbirth + yrstayedinfert +
##     iugr + hxabortion + infcause + alcoholdri + cigsmoke + afc +
##     cycletype + embryotype + embryostage + numembryotrans + assistedhatch +
##     dmhx + predmhx
##
##                  Df Deviance    AIC
## - predmhx         1   155.34 215.34
## - prliveb         1   155.72 215.72
## - hxabortion      1   155.74 215.74
## - cycletype       1   155.86 215.86
## - fullbirth       1   156.14 216.14
## - bmi             3   160.44 216.44
## - hxcs            1   156.62 216.62
## - reparity        1   156.71 216.71
## - numembryotrans  1   156.85 216.85
## <none>               155.11 217.11
## - assistedhatch   1   158.11 218.11
## - iugr            1   158.40 218.40
## - healthinsu      1   158.62 218.62
## - alcoholdri      1   158.66 218.66
## - cigsmoke        1   158.74 218.74
## - agegrp          1   159.25 219.25
## - edustat         2   162.01 220.01
## - yrstayedinfert  2   164.60 222.60
## - childgender     1   163.18 223.18
## - placeresid      1   163.44 223.44
## - dmhx            1   167.87 227.87
## - embryotype      1   169.22 229.22
## - twin            1   169.69 229.69
## - infcause        2   172.97 230.97
## - afc             1   181.70 241.70
## - embryostage     1   190.04 250.04
##
## Step:  AIC=215.34
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + hxcs + prliveb + fullbirth + yrstayedinfert +
##     iugr + hxabortion + infcause + alcoholdri + cigsmoke + afc +
##     cycletype + embryotype + embryostage + numembryotrans + assistedhatch +
##     dmhx
##
##                  Df Deviance    AIC
## - hxabortion      1   155.92 213.92
## - prliveb         1   155.93 213.93
## - cycletype       1   156.19 214.19
## - fullbirth       1   156.25 214.25
## - hxcs            1   156.66 214.66
## - reparity        1   156.84 214.84
## - bmi             3   160.97 214.97
## - numembryotrans  1   157.15 215.15
## <none>               155.34 215.34
## - assistedhatch   1   158.21 216.21
## - iugr            1   158.75 216.75
```

```
## - healthinsu       1   158.88 216.88
## - alcoholdri       1   158.93 216.93
## - cigsmoke         1   159.18 217.18
## - agegrp           1   159.37 217.37
## - edustat          2   162.32 218.32
## - childgender      1   163.29 221.29
## - yrstayedinfert   2   165.34 221.34
## - placeresid       1   163.47 221.47
## - dmhx             1   167.90 225.90
## - embryotype       1   169.22 227.22
## - twin             1   169.70 227.70
## - infcause         2   173.06 229.06
## - afc              1   181.84 239.84
## - embryostage      1   190.16 248.16
##
## Step:  AIC=213.92
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + hxcs + prliveb + fullbirth + yrstayedinfert +
##     iugr + infcause + alcoholdri + cigsmoke + afc + cycletype +
##     embryotype + embryostage + numembryotrans + assistedhatch +
##     dmhx
##
##                     Df Deviance    AIC
## - prliveb            1   156.19 212.19
## - cycletype          1   156.63 212.63
## - fullbirth          1   156.64 212.64
## - hxcs               1   157.19 213.19
## - bmi                3   161.37 213.37
## - reparity           1   157.55 213.55
## - numembryotrans     1   157.67 213.67
## <none>                   155.92 213.92
## - assistedhatch      1   158.62 214.62
## - iugr               1   159.14 215.14
## - healthinsu         1   159.67 215.67
## - agegrp             1   159.84 215.84
## - cigsmoke           1   159.94 215.94
## - alcoholdri         1   160.05 216.05
## - edustat            2   163.05 217.05
## - placeresid         1   163.58 219.58
## - childgender        1   163.69 219.69
## - yrstayedinfert     2   166.66 220.66
## - dmhx               1   168.95 224.95
## - embryotype         1   169.52 225.52
## - twin               1   171.20 227.20
## - infcause           2   174.39 228.39
## - afc                1   183.59 239.59
## - embryostage        1   190.97 246.97
##
## Step:  AIC=212.19
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + hxcs + fullbirth + yrstayedinfert +
##     iugr + infcause + alcoholdri + cigsmoke + afc + cycletype +
##     embryotype + embryostage + numembryotrans + assistedhatch +
##     dmhx
```

9

```
##
##                   Df Deviance    AIC
## - cycletype        1   156.85 210.85
## - fullbirth        1   157.03 211.03
## - bmi              3   161.44 211.44
## - hxcs             1   157.47 211.47
## - numembryotrans   1   157.80 211.80
## <none>                 156.19 212.19
## - assistedhatch    1   158.62 212.62
## - iugr             1   159.62 213.62
## - agegrp           1   159.93 213.93
## - cigsmoke         1   159.97 213.97
## - alcoholdri       1   160.16 214.16
## - healthinsu       1   160.72 214.72
## - edustat          2   163.14 215.14
## - reparity         1   161.91 215.91
## - placeresid       1   163.86 217.86
## - childgender      1   164.05 218.05
## - yrstayedinfert   2   166.95 218.95
## - dmhx             1   169.12 223.12
## - embryotype       1   169.53 223.53
## - twin             1   171.88 225.88
## - infcause         2   174.47 226.47
## - afc              1   183.61 237.61
## - embryostage      1   191.36 245.36
##
## Step:  AIC=210.85
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##      twin + childgender + hxcs + fullbirth + yrstayedinfert +
##      iugr + infcause + alcoholdri + cigsmoke + afc + embryotype +
##      embryostage + numembryotrans + assistedhatch + dmhx
##
##                   Df Deviance    AIC
## - fullbirth        1   157.88 209.88
## - bmi              3   162.04 210.04
## - hxcs             1   158.16 210.16
## - numembryotrans   1   158.24 210.24
## <none>                 156.85 210.85
## - assistedhatch    1   159.71 211.71
## - iugr             1   160.29 212.29
## - cigsmoke         1   160.65 212.65
## - agegrp           1   160.93 212.93
## - alcoholdri       1   160.96 212.96
## - edustat          2   163.54 213.54
## - healthinsu       1   161.83 213.83
## - reparity         1   162.83 214.83
## - placeresid       1   164.48 216.48
## - childgender      1   164.51 216.51
## - yrstayedinfert   2   167.37 217.37
## - embryotype       1   169.77 221.77
## - dmhx             1   170.60 222.60
## - twin             1   172.76 224.76
## - infcause         2   175.08 225.08
## - afc              1   183.66 235.66
```

```
## - embryostage     1   191.80 243.80
##
## Step:  AIC=209.88
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + hxcs + yrstayedinfert + iugr + infcause +
##     alcoholdri + cigsmoke + afc + embryotype + embryostage +
##     numembryotrans + assistedhatch + dmhx
##
##                   Df Deviance    AIC
## - hxcs             1   158.88 208.88
## - numembryotrans   1   159.03 209.03
## - bmi              3   163.67 209.67
## <none>                 157.88 209.88
## - assistedhatch    1   160.61 210.61
## - cigsmoke         1   162.03 212.03
## - agegrp           1   162.05 212.05
## - iugr             1   162.13 212.13
## - alcoholdri       1   162.49 212.49
## - edustat          2   164.51 212.51
## - healthinsu       1   162.89 212.89
## - childgender      1   165.12 215.12
## - yrstayedinfert   2   167.73 215.73
## - placeresid       1   165.92 215.92
## - reparity         1   166.40 216.40
## - dmhx             1   172.08 222.08
## - embryotype       1   172.12 222.12
## - twin             1   172.99 222.99
## - infcause         2   175.58 223.58
## - afc              1   184.44 234.44
## - embryostage      1   191.98 241.98
##
## Step:  AIC=208.88
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + yrstayedinfert + iugr + infcause + alcoholdri +
##     cigsmoke + afc + embryotype + embryostage + numembryotrans +
##     assistedhatch + dmhx
##
##                   Df Deviance    AIC
## - numembryotrans   1   160.01 208.01
## - bmi              3   164.66 208.66
## <none>                 158.88 208.88
## - assistedhatch    1   161.61 209.61
## - agegrp           1   162.98 210.98
## - iugr             1   163.20 211.20
## - edustat          2   165.28 211.28
## - healthinsu       1   163.72 211.72
## - cigsmoke         1   164.27 212.27
## - alcoholdri       1   165.16 213.16
## - placeresid       1   166.76 214.76
## - yrstayedinfert   2   168.80 214.80
## - childgender      1   167.09 215.09
## - reparity         1   167.25 215.25
## - dmhx             1   172.70 220.70
## - embryotype       1   173.04 221.04
```

```
## - twin            1    173.62 221.62
## - infcause        2    177.08 223.08
## - afc             1    185.83 233.83
## - embryostage     1    192.04 240.04
##
## Step:  AIC=208.01
## apo ~ agegrp + bmi + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + yrstayedinfert + iugr + infcause + alcoholdri +
##     cigsmoke + afc + embryotype + embryostage + assistedhatch +
##     dmhx
##
##                   Df Deviance    AIC
## - bmi             3    165.22 207.22
## <none>                 160.01 208.01
## - assistedhatch   1    162.78 208.78
## - agegrp          1    163.87 209.87
## - edustat         2    166.90 210.90
## - cigsmoke        1    165.28 211.28
## - iugr            1    165.46 211.46
## - alcoholdri      1    166.39 212.39
## - healthinsu      1    167.28 213.28
## - placeresid      1    167.86 213.86
## - yrstayedinfert  2    169.90 213.90
## - childgender     1    168.36 214.36
## - reparity        1    169.24 215.24
## - embryotype      1    174.49 220.49
## - twin            1    174.53 220.53
## - dmhx            1    175.58 221.58
## - infcause        2    177.64 221.64
## - afc             1    188.60 234.60
## - embryostage     1    194.66 240.66
##
## Step:  AIC=207.22
## apo ~ agegrp + healthinsu + placeresid + edustat + reparity +
##     twin + childgender + yrstayedinfert + iugr + infcause + alcoholdri +
##     cigsmoke + afc + embryotype + embryostage + assistedhatch +
##     dmhx
##
##                   Df Deviance    AIC
## - agegrp          1    166.10 206.10
## <none>                 165.22 207.22
## - assistedhatch   1    167.33 207.33
## - cigsmoke        1    169.77 209.77
## - edustat         2    172.50 210.50
## - iugr            1    170.72 210.72
## - childgender     1    171.97 211.97
## - healthinsu      1    172.02 212.02
## - alcoholdri      1    172.04 212.04
## - placeresid      1    172.34 212.34
## - yrstayedinfert  2    175.05 213.05
## - reparity        1    174.71 214.71
## - embryotype      1    179.06 219.06
## - twin            1    180.91 220.91
## - dmhx            1    181.98 221.98
```

12

```
## - infcause        2    185.58 223.58
## - afc             1    195.35 235.35
## - embryostage     1    198.53 238.53
##
## Step:  AIC=206.1
## apo ~ healthinsu + placeresid + edustat + reparity + twin + childgender +
##     yrstayedinfert + iugr + infcause + alcoholdri + cigsmoke +
##     afc + embryotype + embryostage + assistedhatch + dmhx
##
##                   Df Deviance    AIC
## - assistedhatch    1    168.06 206.06
## <none>                  166.10 206.10
## - cigsmoke         1    170.80 208.80
## - iugr             1    171.75 209.75
## - edustat          2    174.27 210.27
## - healthinsu       1    172.94 210.94
## - alcoholdri       1    173.14 211.14
## - placeresid       1    173.73 211.73
## - yrstayedinfert   2    175.87 211.87
## - childgender      1    175.75 213.75
## - reparity         1    176.91 214.91
## - embryotype       1    181.23 219.23
## - twin             1    182.11 220.11
## - dmhx             1    183.29 221.29
## - infcause         2    186.41 222.41
## - afc              1    198.24 236.24
## - embryostage      1    200.06 238.06
##
## Step:  AIC=206.06
## apo ~ healthinsu + placeresid + edustat + reparity + twin + childgender +
##     yrstayedinfert + iugr + infcause + alcoholdri + cigsmoke +
##     afc + embryotype + embryostage + dmhx
##
##                   Df Deviance    AIC
## <none>                  168.06 206.06
## - cigsmoke         1    173.06 209.06
## - edustat          2    176.62 210.62
## - iugr             1    174.84 210.84
## - placeresid       1    175.12 211.12
## - alcoholdri       1    175.56 211.56
## - yrstayedinfert   2    178.07 212.07
## - healthinsu       1    176.34 212.34
## - childgender      1    177.16 213.16
## - reparity         1    179.56 215.56
## - dmhx             1    184.56 220.56
## - twin             1    184.73 220.73
## - embryotype       1    184.77 220.77
## - infcause         2    187.57 221.57
## - afc              1    203.31 239.31
## - embryostage      1    203.85 239.85
```

```r
candidatemodel <- glm(apo ~ healthinsu + placeresid + edustat + reparity +
                      twin + childgender + yrstayedinfert + iugr + infcause +
                      alcoholdri + cigsmoke + afc + embryotype + embryostage + dmhx,
```

```
                    data = data_complete,
                    family = binomial(link = "logit"))

summary(candidatemodel)
```

```
##
## Call:
## glm(formula = apo ~ healthinsu + placeresid + edustat + reparity +
##     twin + childgender + yrstayedinfert + iugr + infcause + alcoholdri +
##     cigsmoke + afc + embryotype + embryostage + dmhx, family = binomial(link = "logit"),
##     data = data_complete)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -3.0632     0.9541  -3.210 0.001325 **
## healthinsu2       1.7326     0.6332   2.736 0.006213 **
## placeresid2      -1.2145     0.4742  -2.561 0.010436 *
## edustat2          1.3158     0.4788   2.748 0.005991 **
## edustat3          0.2873     0.5702   0.504 0.614391
## reparity2        -1.4005     0.4331  -3.233 0.001223 **
## twin2             1.7758     0.4712   3.768 0.000164 ***
## childgender2      1.2281     0.4209   2.918 0.003521 **
## yrstayedinfert2   0.9629     0.5569   1.729 0.083833 .
## yrstayedinfert3   1.6673     0.5489   3.037 0.002386 **
## iugr2             1.3341     0.5359   2.489 0.012803 *
## infcause2         0.3267     0.4868   0.671 0.502174
## infcause3         2.1765     0.5586   3.896 9.76e-05 ***
## alcoholdri2       3.3755     1.2841   2.629 0.008573 **
## cigsmoke2        -2.4888     1.1542  -2.156 0.031064 *
## afc2              2.4237     0.4613   5.254 1.49e-07 ***
## embryotype2      -1.7569     0.4550  -3.862 0.000113 ***
## embryostage2     -2.7171     0.5203  -5.222 1.77e-07 ***
## dmhx2             2.5090     0.6933   3.619 0.000296 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 415.02  on 317  degrees of freedom
## Residual deviance: 168.06  on 299  degrees of freedom
## AIC: 206.06
##
## Number of Fisher Scoring iterations: 7
```

```
################################################################################
################################################################################
###                    Further model reduction by LRT                       ###
################################################################################
################################################################################

drop1_results <- drop1(candidatemodel, test = "LRT")
print(drop1_results)
```

```
## Single term deletions
##
## Model:
## apo ~ healthinsu + placeresid + edustat + reparity + twin + childgender +
##     yrstayedinfert + iugr + infcause + alcoholdri + cigsmoke +
##     afc + embryotype + embryostage + dmhx
##                Df Deviance    AIC    LRT  Pr(>Chi)
## <none>             168.06 206.06
## healthinsu      1  176.34 212.34  8.281 0.0040063 **
## placeresid      1  175.12 211.12  7.060 0.0078803 **
## edustat         2  176.62 210.62  8.559 0.0138510 *
## reparity        1  179.56 215.56 11.503 0.0006947 ***
## twin            1  184.73 220.73 16.678 4.429e-05 ***
## childgender     1  177.16 213.16  9.103 0.0025519 **
## yrstayedinfert  2  178.07 212.07 10.010 0.0067035 **
## iugr            1  174.84 210.84  6.789 0.0091728 **
## infcause        2  187.57 221.57 19.511 5.798e-05 ***
## alcoholdri      1  175.56 211.56  7.505 0.0061525 **
## cigsmoke        1  173.06 209.06  5.003 0.0252997 *
## afc             1  203.31 239.31 35.258 2.888e-09 ***
## embryotype      1  184.77 220.77 16.711 4.352e-05 ***
## embryostage     1  203.85 239.85 35.792 2.196e-09 ***
## dmhx            1  184.56 220.56 16.506 4.849e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
candidatemodel2 <- glm(apo ~ healthinsu + placeresid + reparity +
                        twin + childgender + yrstayedinfert + iugr + infcause +
                        alcoholdri + afc + embryotype + embryostage + dmhx,
                   data = data_complete,
                   family = binomial(link = "logit"))

summary(candidatemodel2)
```

```
##
## Call:
## glm(formula = apo ~ healthinsu + placeresid + reparity + twin +
##     childgender + yrstayedinfert + iugr + infcause + alcoholdri +
##     afc + embryotype + embryostage + dmhx, family = binomial(link = "logit"),
##     data = data_complete)
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -2.4075     0.8644  -2.785 0.005350 **
## healthinsu2      1.7183     0.5837   2.944 0.003240 **
## placeresid2     -1.1349     0.4426  -2.564 0.010335 *
## reparity2       -1.4236     0.4119  -3.457 0.000547 ***
## twin2            1.5009     0.4236   3.543 0.000396 ***
## childgender2     1.3430     0.4031   3.332 0.000864 ***
## yrstayedinfert2  0.6376     0.5164   1.235 0.216896
## yrstayedinfert3  1.1806     0.4905   2.407 0.016093 *
## iugr2            1.3383     0.5278   2.536 0.011221 *
## infcause2        0.4610     0.4750   0.971 0.331768
## infcause3        1.8698     0.5205   3.592 0.000328 ***
```

```
## alcoholdri2        1.1750      0.6019    1.952 0.050930 .
## afc2               2.3207      0.4249    5.462 4.70e-08 ***
## embryotype2       -1.5278      0.4247   -3.597 0.000322 ***
## embryostage2      -2.3074      0.4669   -4.942 7.72e-07 ***
## dmhx2              2.2756      0.6593    3.452 0.000557 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 415.02  on 317  degrees of freedom
## Residual deviance: 181.53  on 302  degrees of freedom
## AIC: 213.53
##
## Number of Fisher Scoring iterations: 6
```

```
drop1_results <- drop1(candidatemodel2, test = "LRT")
print(drop1_results)
```

```
## Single term deletions
##
## Model:
## apo ~ healthinsu + placeresid + reparity + twin + childgender +
##     yrstayedinfert + iugr + infcause + alcoholdri + afc + embryotype +
##     embryostage + dmhx
##                Df Deviance    AIC    LRT  Pr(>Chi)
## <none>             181.53 213.53
## healthinsu      1   191.00 221.00  9.467 0.0020919 **
## placeresid      1   188.51 218.51  6.985 0.0082188 **
## reparity        1   194.68 224.68 13.152 0.0002872 ***
## twin            1   195.52 225.52 13.994 0.0001834 ***
## childgender     1   193.56 223.56 12.031 0.0005233 ***
## yrstayedinfert  2   187.49 215.49  5.966 0.0506352 .
## iugr            1   188.55 218.55  7.023 0.0080484 **
## infcause        2   196.84 224.84 15.315 0.0004725 ***
## alcoholdri      1   185.32 215.32  3.795 0.0514210 .
## afc             1   217.90 247.90 36.372 1.630e-09 ***
## embryotype      1   195.46 225.46 13.936 0.0001892 ***
## embryostage     1   211.63 241.63 30.104 4.095e-08 ***
## dmhx            1   196.28 226.28 14.756 0.0001223 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
candidatemodel3 = glm(apo ~ healthinsu + placeresid + reparity + twin + childgender +
                      iugr + infcause + afc + embryotype + embryostage + dmhx,
                   data = data_complete,
                   family = binomial)

summary(candidatemodel3)
```

```
##
## Call:
## glm(formula = apo ~ healthinsu + placeresid + reparity + twin +
```

```
##      childgender + iugr + infcause + afc + embryotype + embryostage +
##      dmhx, family = binomial, data = data_complete)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.7847     0.5474  -1.434 0.151714
## healthinsu2     1.5812     0.5684   2.782 0.005402 **
## placeresid2    -1.1472     0.4287  -2.676 0.007450 **
## reparity2      -1.2546     0.3907  -3.211 0.001321 **
## twin2           1.3516     0.3989   3.388 0.000704 ***
## childgender2    1.3716     0.3936   3.485 0.000493 ***
## iugr2           1.5546     0.5227   2.974 0.002937 **
## infcause2       0.6935     0.4466   1.553 0.120486
## infcause3       1.8286     0.5004   3.654 0.000258 ***
## afc2            2.3517     0.4110   5.722 1.06e-08 ***
## embryotype2    -1.7081     0.4030  -4.238 2.26e-05 ***
## embryostage2   -2.2753     0.4507  -5.049 4.45e-07 ***
## dmhx2           2.3253     0.6296   3.693 0.000221 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 415.02  on 317  degrees of freedom
## Residual deviance: 190.97  on 305  degrees of freedom
## AIC: 216.97
##
## Number of Fisher Scoring iterations: 6
```

```
drop1_results <- drop1(candidatemodel3, test = "LRT")
print(drop1_results)
```

```
## Single term deletions
##
## Model:
## apo ~ healthinsu + placeresid + reparity + twin + childgender +
##      iugr + infcause + afc + embryotype + embryostage + dmhx
##               Df Deviance    AIC     LRT  Pr(>Chi)
## <none>             190.97 216.97
## healthinsu     1   199.37 223.37   8.400 0.0037514 **
## placeresid     1   198.62 222.62   7.648 0.0056823 **
## reparity       1   202.02 226.02  11.047 0.0008881 ***
## twin           1   203.46 227.46  12.483 0.0004107 ***
## childgender    1   204.22 228.22  13.243 0.0002735 ***
## iugr           1   200.79 224.79   9.816 0.0017301 **
## infcause       2   205.90 227.90  14.923 0.0005747 ***
## afc            1   231.26 255.26  40.290 2.189e-10 ***
## embryotype     1   210.91 234.91  19.935 8.013e-06 ***
## embryostage    1   222.39 246.39  31.413 2.085e-08 ***
## dmhx           1   207.44 231.44  16.467 4.950e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
finalmodel = glm(apo ~ healthinsu + placeresid + reparity + twin + childgender +
                     iugr + infcause + afc + embryotype + embryostage + dmhx,
                 data = data_complete,
                 family = binomial)

summary(finalmodel)
```

```
##
## Call:
## glm(formula = apo ~ healthinsu + placeresid + reparity + twin +
##     childgender + iugr + infcause + afc + embryotype + embryostage +
##     dmhx, family = binomial, data = data_complete)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.7847     0.5474  -1.434 0.151714
## healthinsu2   1.5812     0.5684   2.782 0.005402 **
## placeresid2  -1.1472     0.4287  -2.676 0.007450 **
## reparity2    -1.2546     0.3907  -3.211 0.001321 **
## twin2         1.3516     0.3989   3.388 0.000704 ***
## childgender2  1.3716     0.3936   3.485 0.000493 ***
## iugr2         1.5546     0.5227   2.974 0.002937 **
## infcause2     0.6935     0.4466   1.553 0.120486
## infcause3     1.8286     0.5004   3.654 0.000258 ***
## afc2          2.3517     0.4110   5.722 1.06e-08 ***
## embryotype2  -1.7081     0.4030  -4.238 2.26e-05 ***
## embryostage2 -2.2753     0.4507  -5.049 4.45e-07 ***
## dmhx2         2.3253     0.6296   3.693 0.000221 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 415.02  on 317  degrees of freedom
## Residual deviance: 190.97  on 305  degrees of freedom
## AIC: 216.97
##
## Number of Fisher Scoring iterations: 6
```

```r
################################################################################
################################################################################
##                      Model Assumption Check                               ##
################################################################################
################################################################################


# 1. Influential Observations

# Average leverage
p <- length(coef(finalmodel))  # Number of parameters
n <- nrow(data_complete)
avg_leverage <- (3*p) / n

avg_leverage
```
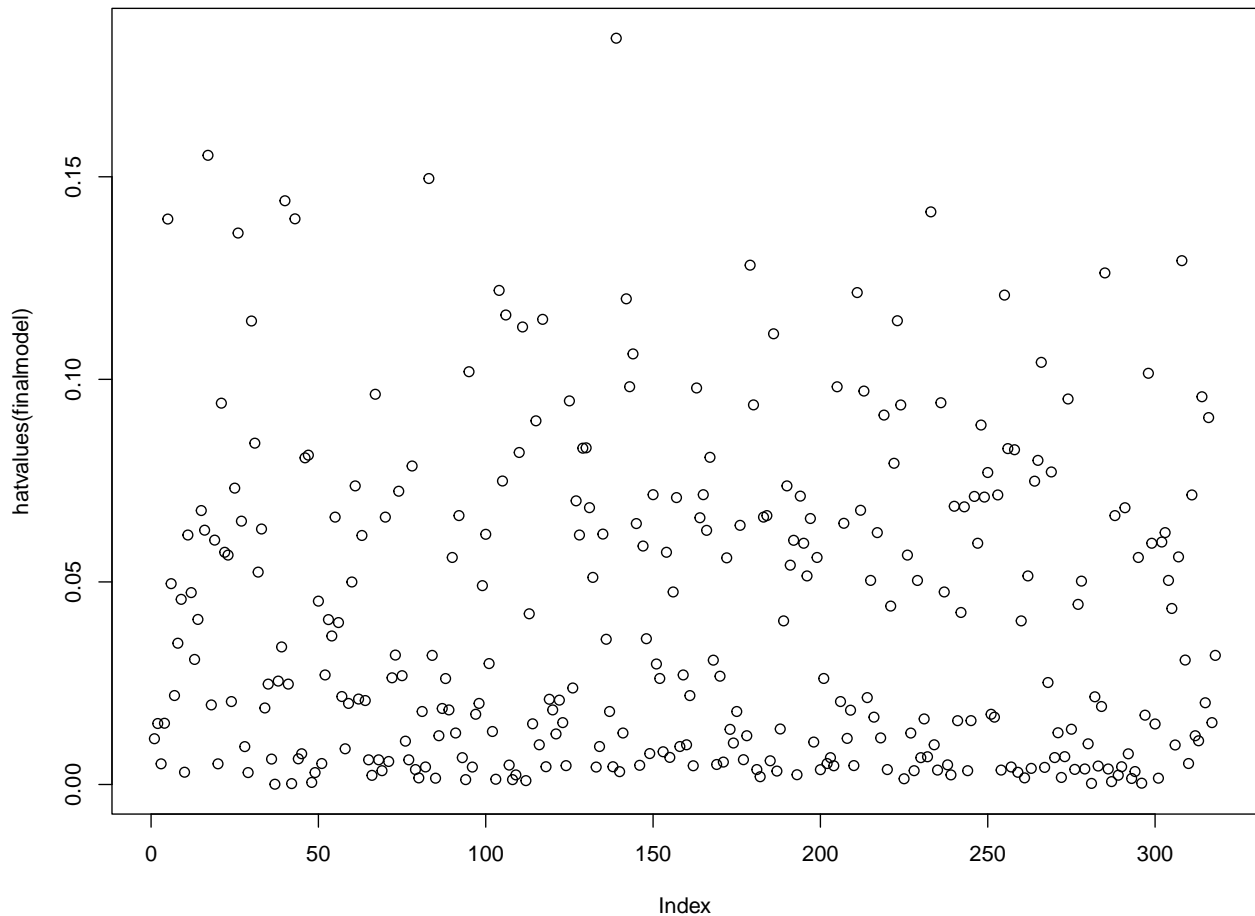
```
## [1] 0.1226415
```

```
plot(hatvalues(finalmodel))
```



```
#There is no observation with huge deviation from average hat values


#2. Multicollinearity

# Variance Inflation Factor (VIF)
car::vif(finalmodel)   # Values >5-10 indicate problems
```
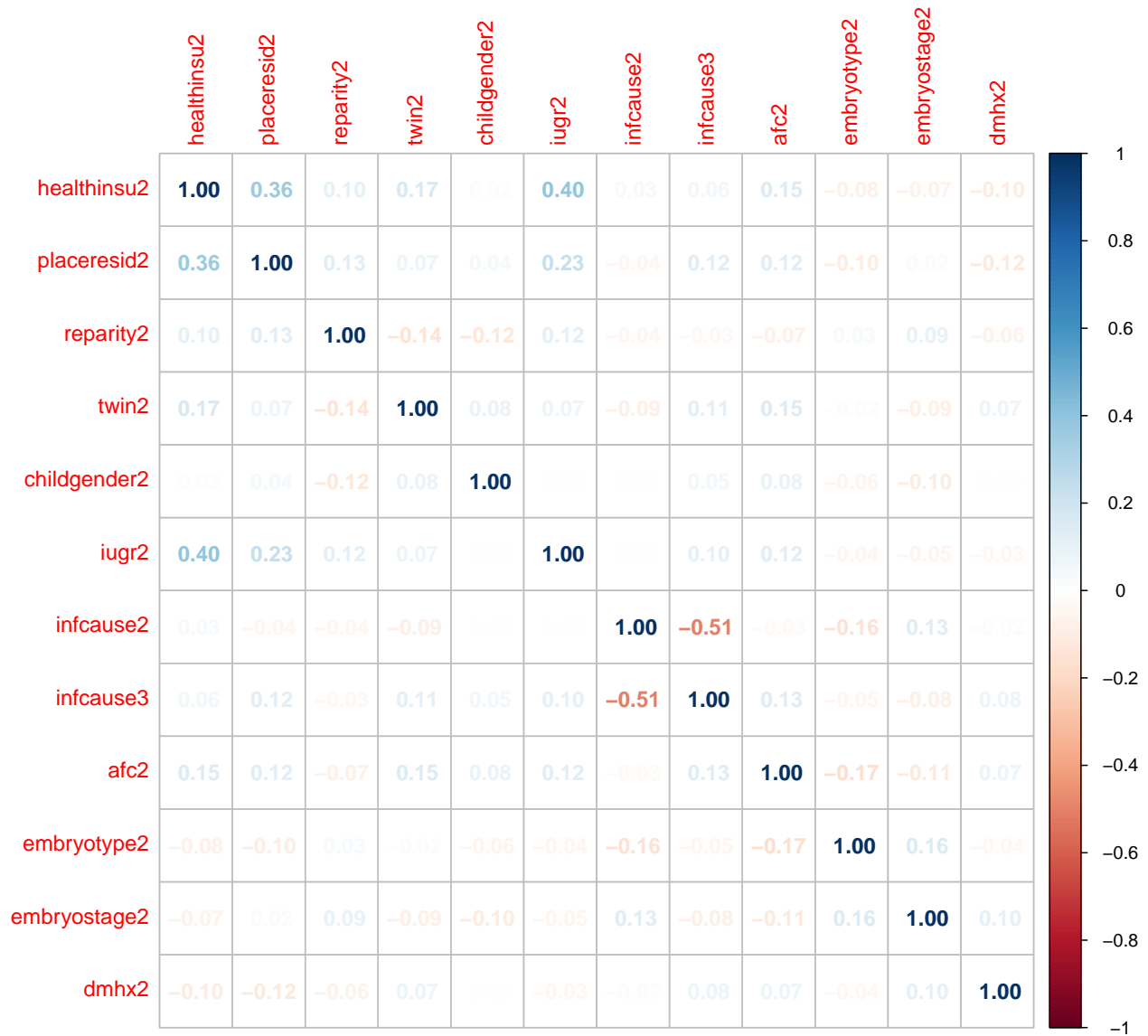
```
##                GVIF Df GVIF^(1/(2*Df))
## healthinsu 1.339633  1        1.157425
## placeresid 1.328595  1        1.152647
## reparity   1.130756  1        1.063370
## twin       1.186154  1        1.089107
## childgender 1.136771 1        1.066194
## iugr       1.352032  1        1.162769
## infcause   1.243973  2        1.056094
## afc        1.258852  1        1.121986
## embryotype 1.216758  1        1.103067
## embryostage 1.427680 1        1.194856
## dmhx       1.241370  1        1.114168
```

```r
# Correlation matrix
library(corrplot)
corrplot(cor(model.matrix(finalmodel)[,-1]), method = "number")
```

|	healthinsu2	placeresid2	reparity2	twin2	childgender2	iugr2	infcause2	infcause3	afc2	embryotype2	embryostage2	dmhx2
healthinsu2	1.00	0.36	0.10	0.17	0.02	0.40	0.03	0.06	0.15	−0.08	−0.07	−0.10
placeresid2	0.36	1.00	0.13	0.07	0.04	0.23	−0.04	0.12	0.12	−0.10	0.02	−0.12
reparity2	0.10	0.13	1.00	−0.14	−0.12	0.12	−0.04	−0.03	−0.07	0.03	0.09	−0.06
twin2	0.17	0.07	−0.14	1.00	0.08	0.07	−0.09	0.11	0.15	−0.02	−0.09	0.07
childgender2	0.02	0.04	−0.12	0.08	1.00			0.05	0.08	−0.06	−0.10	
iugr2	0.40	0.23	0.12	0.07		1.00		0.10	0.12	−0.04	−0.05	−0.03
infcause2	0.03	−0.04	−0.04	−0.09			1.00	−0.51	−0.03	−0.16	0.13	−0.02
infcause3	0.06	0.12	−0.03	0.11	0.05	0.10	−0.51	1.00	0.13	−0.05	−0.08	0.08
afc2	0.15	0.12	−0.07	0.15	0.08	0.12	−0.03	0.13	1.00	−0.17	−0.11	0.07
embryotype2	−0.08	−0.10	0.03	−0.02	−0.06	−0.04	−0.16	−0.05	−0.17	1.00	0.16	−0.04
embryostage2	−0.07	0.02	0.09	−0.09	−0.10	−0.05	0.13	−0.08	−0.11	0.16	1.00	0.10
dmhx2	−0.10	−0.12	−0.06	0.07		−0.03	−0.02	0.08	0.07	−0.04	0.10	1.00

```r
# The vif shows no multicolinearity that have an issue

#3. Outliers

# Standardized residuals
std_resid <- rstandard(finalmodel)
plot(std_resid, ylab = "Standardized Residuals")
abline(h = c(-2.5, 2.5), col = "red")  # Flagging potential outliers
```
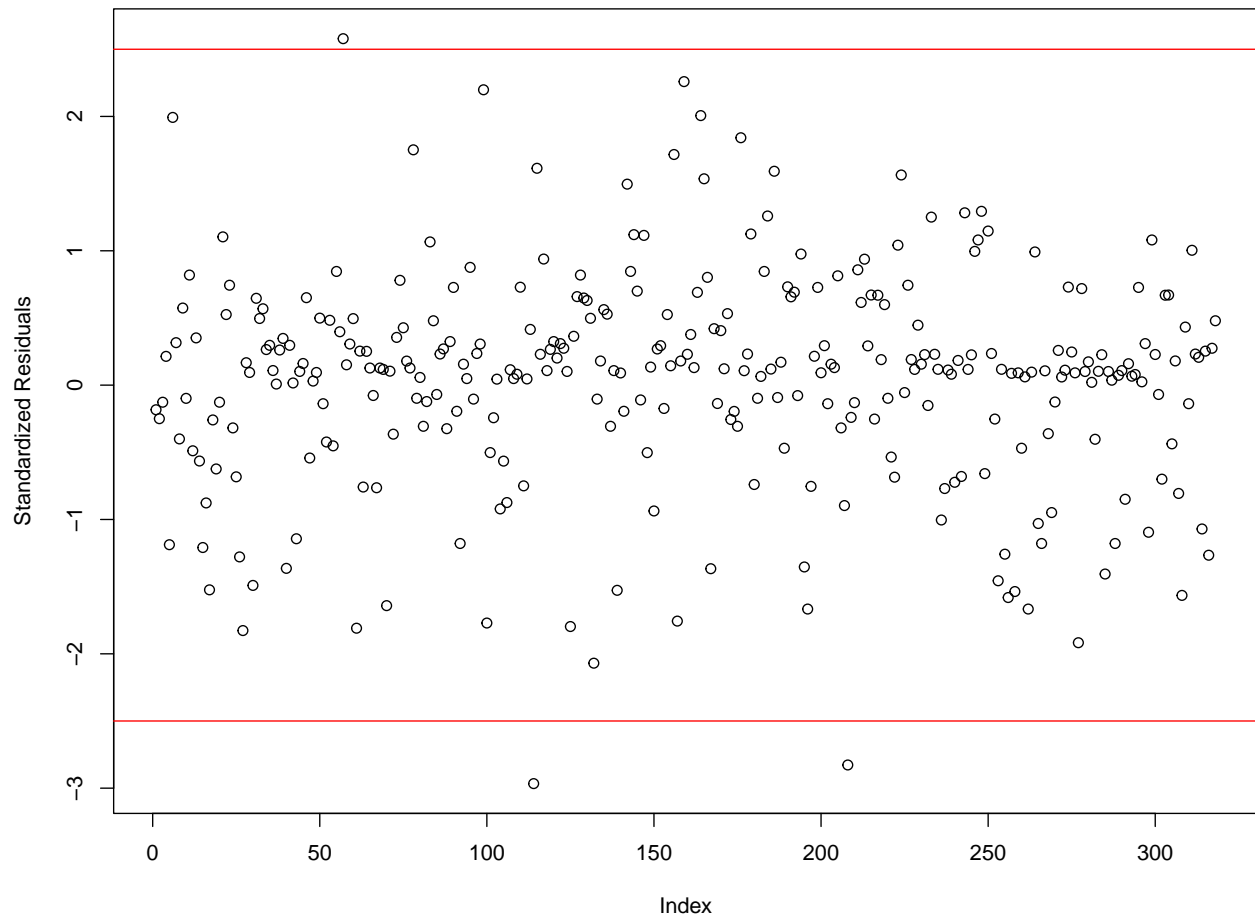
```r
# no observation here also out of the demarcated area.

# Goodness-of-Fit

# Hosmer-Lemeshow Test
ResourceSelection::hoslem.test(finalmodel$y, fitted(finalmodel), g = 10)
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  finalmodel$y, fitted(finalmodel)
## X-squared = 8.2596, df = 8, p-value = 0.4085
```

```r
# The Model fit the data

# Pearson residuals
pearson_res <- residuals(finalmodel, type = "pearson")
plot(pearson_res, ylab = "Pearson Residuals")
abline(h = 0, col = "red")
```
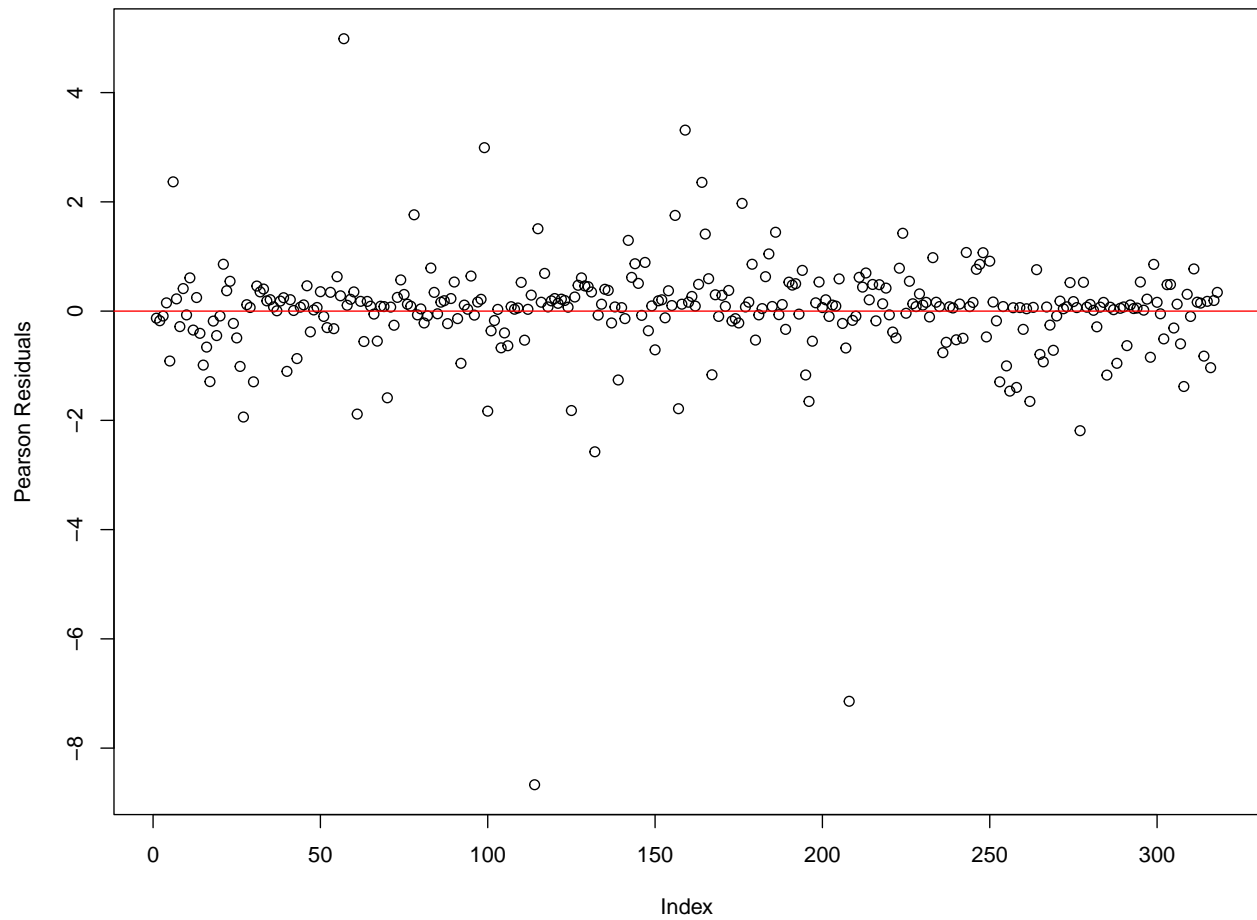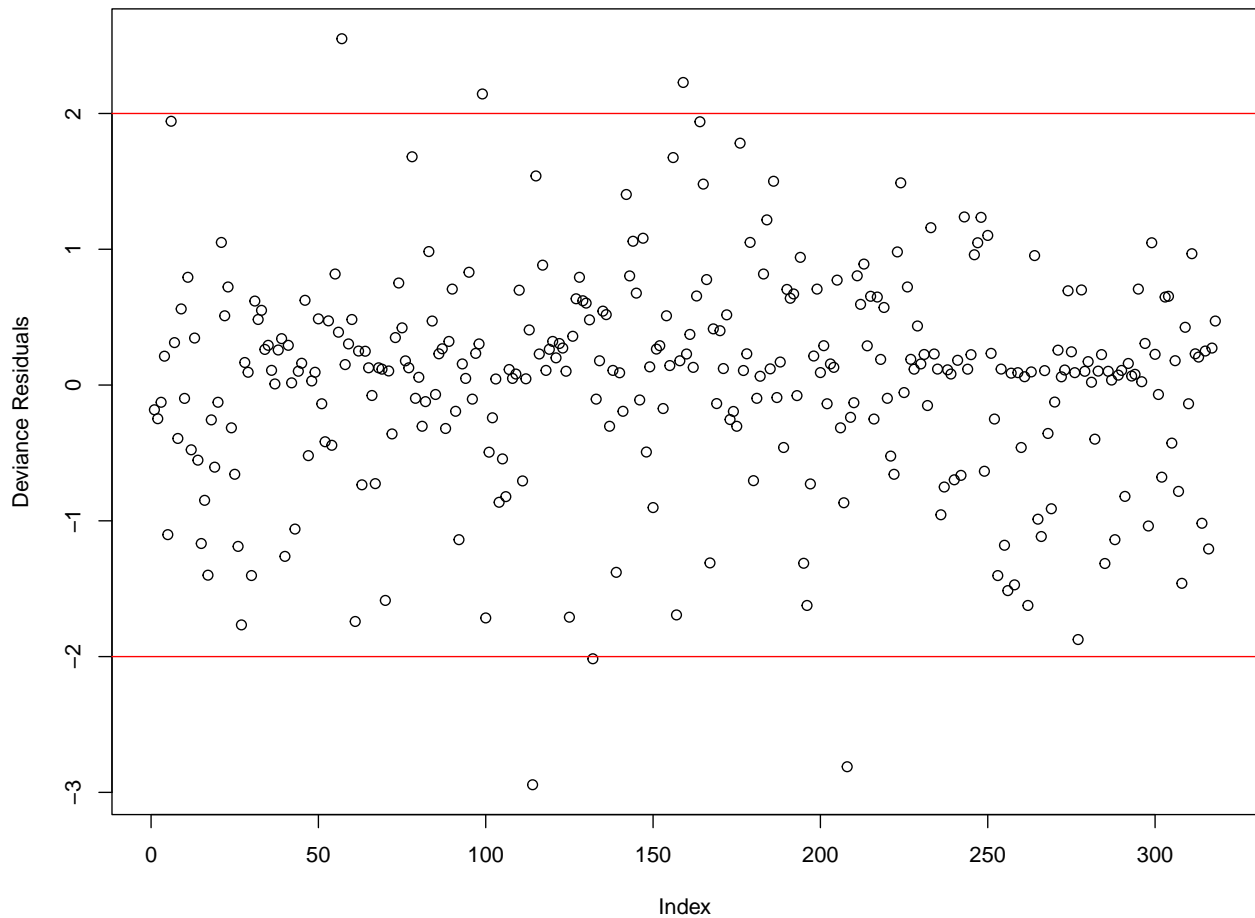
```r
# Deviance residuals
dev_res <- residuals(finalmodel, type = "deviance")
plot(dev_res, ylab = "Deviance Residuals")
abline(h = c(-2, 2), col = "red")
```

```
###############################################################################
###############################################################################
##                        Risk Score Development                            ##
###############################################################################
###############################################################################


# Step 1: Extract coefficients and prepare scoring
coefs <- coef(finalmodel)
score_coefs <- round(coefs / min(abs(coefs[-1])))   # Exclude intercept when scaling

# Step 2: Create model matrix (X) that matches what was used in the model
X <- model.matrix(finalmodel)[, -1]   # Remove intercept column

# Step 3: Convert to data frame and add variable names that match coef names
X_df <- as.data.frame(X)
colnames(X_df)   # These should match names(score_coefs)[-1]
```

```
##  [1] "healthinsu2"  "placeresid2"  "reparity2"    "twin2"        "childgender2"
##  [6] "iugr2"        "infcause2"    "infcause3"    "afc2"         "embryotype2"
## [11] "embryostage2" "dmhx2"
```

```r
# Step 4: Define a risk scoring function
calculate_risk_score <- function(row) {
  score <- 0
  for (var in names(score_coefs)) {
    if (var != "(Intercept)" && var %in% names(row)) {
      score <- score + row[[var]] * score_coefs[[var]]
    }
  }
  return(score)
}

# Step 5: Apply function row by row
X_df$risk_score <- apply(X_df, 1, function(row) {
  calculate_risk_score(as.list(row))
})

# Step 6: Combine back with original patient IDs or outcomes if needed
data_complete$risk_score <- X_df$risk_score

# Optional: Calculate predicted probability from logistic model
intercept <- coefs["(Intercept)"]
data_complete$predicted_prob <- plogis(intercept + data_complete$risk_score)


################################################################################
################################################################################
###                      Develop a Nomogram                                 ###
################################################################################
################################################################################


# Install required packages
# install.packages(c("rms", "pROC", "caret", "ResourceSelection", "dcurves"))

# Load libraries
library(rms)
library(pROC)
library(caret)
library(ResourceSelection)
library(dcurves)


table(data_complete$healthinsu)


## 
##   1   2
## 249  69


data_complete$healthinsu <-factor(data_complete$healthinsu,
                    levels= c(1,2),
                    labels= c("no", "yes"))
```

```r
data_complete$placeresid <-factor(data_complete$placeresid,
                            levels= c(1,2),
                            labels= c("urban", "rural"))




data_complete$reparity <-factor(data_complete$reparity,
                                   levels= c(1,2),
                                   labels= c("multiparous", "primiparous"))




data_complete$childgender <-factor(data_complete$childgender,
                                   levels= c(1,2),
                                   labels= c("male", "female"))




data_complete$iugr <-factor(data_complete$iugr,
                            levels= c(1,2),
                            labels= c("no", "yes"))




data_complete$infcause <-factor(data_complete$infcause,
                            levels= c(1,2,3),
                            labels= c("male factor", "femae factor", "unexplained"))




data_complete$afc <-factor(data_complete$afc,
                            levels= c(1,2),
                            labels= c(">= 5", "<5"))




data_complete$embryotype <-factor(data_complete$embryotype,
                            levels= c(1,2),
                            labels= c("fresh", "frozen"))




data_complete$embryostage <-factor(data_complete$embryostage,
                                   levels= c(1,2),
                                   labels= c("blastocyte", "cleavage"))




data_complete$dmhx <-factor(data_complete$dmhx,
```

```
                                    levels= c(1,2),
                                    labels= c("no", "yes"))


# Nomogram

# Convert data to datadist object for rms package
dd <- datadist(data_complete)
options(datadist = "dd")

# Fit model using rms::lrm
finalmodel_rms <- lrm(apo ~ healthinsu + placeresid + reparity + twin + childgender +
                   iugr + infcause + afc + embryotype + embryostage + dmhx,
                data = data_complete, x=TRUE,y=TRUE )

# Draw Nomogram
nom <- nomogram(finalmodel_rms, fun = plogis, fun.at = c(0.1, 0.3, 0.5, 0.7, 0.9))
plot(nom)
```
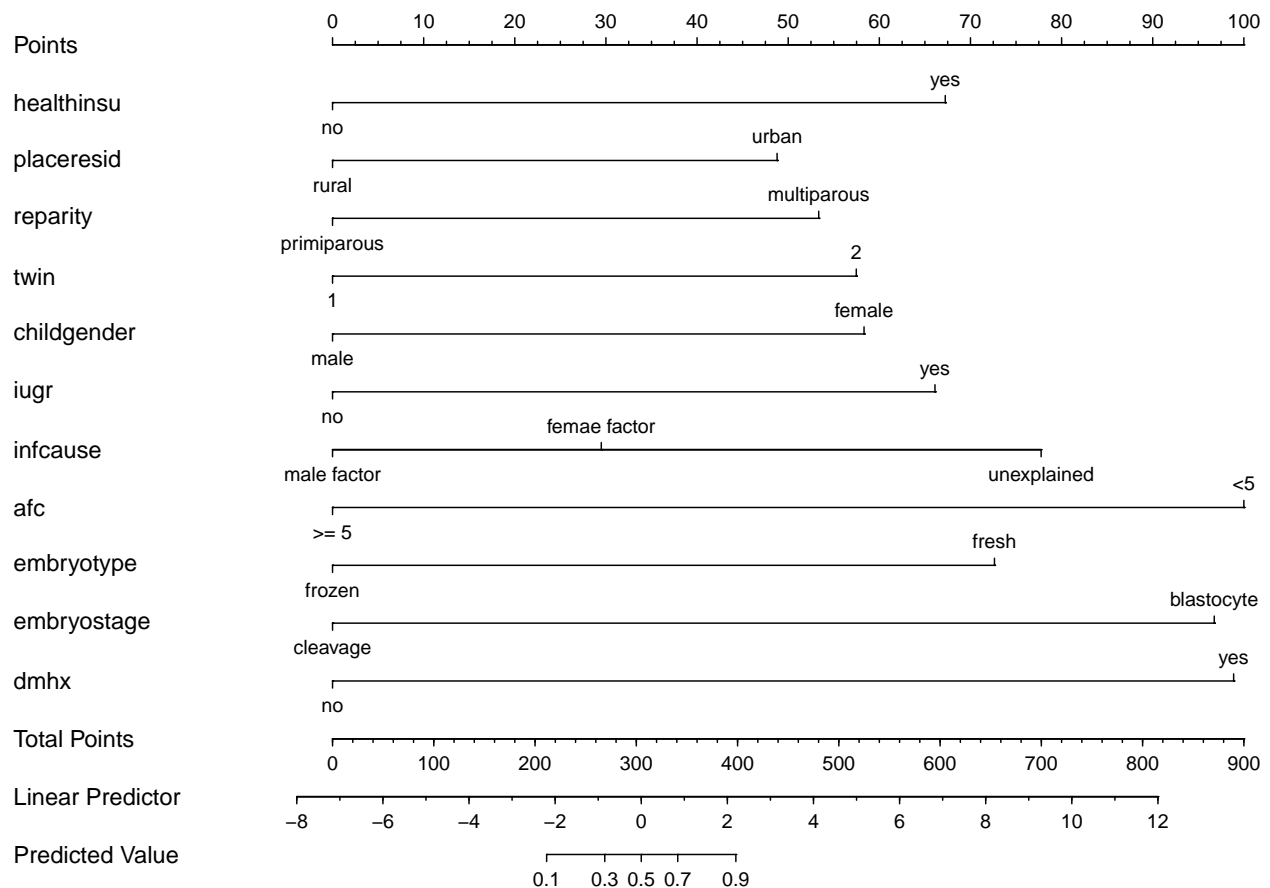


```
##############################################################################
#                 Risk classification using Youden Index                     #
##############################################################################


# Predict probabilities
data_complete$pred_prob <- predict(finalmodel, type = "response")
```

```r
# ROC and Youden's index
roc_obj <- roc(data_complete$apo, data_complete$pred_prob)
coords_obj <- coords(roc_obj, "best", best.method = "youden", ret = c("threshold", "sensitivity", "spec

# Print cutoff and metrics
print(coords_obj)
```

```
##   threshold sensitivity specificity
## 1 0.7703686   0.7941176   0.9473684
```

```r
# Apply cutoff to classify risk
cutoff <- coords_obj["threshold"]

# Classify into 2 risk categories: Low and High
data_complete$risk_class <- cut(data_complete$pred_prob,
                                breaks = c(0, cutoff, 1),
                                labels = c("Low", "High"),
                                include.lowest = TRUE)

# See the counts for each group
table(data_complete$risk_class)
```

```
##
## Low High
## 150  168
```

```r
#Calculate sensitivity, specificity, PPV, NPV, accuracy

# Predicted class
data_complete$pred_class <- ifelse(data_complete$pred_prob >= 0.7703686, 1, 0)
table(data_complete$pred_class)
```

```
##
##   0   1
## 150 168
```

```r
table(data_complete$apo)
```

```
##
##  No Yes
## 114 204
```

```r
# First recode apo to 0/1
data_complete$apo_bin <- ifelse(data_complete$apo == "Yes", 1, 0)

# Now make both factors
conf_mat <- confusionMatrix(factor(data_complete$pred_class),
                            factor(data_complete$apo_bin),
                            positive = "1")

print(conf_mat)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0    1
##          0 108   42
##          1   6  162
##
##                Accuracy : 0.8491
##                  95% CI : (0.8049, 0.8866)
##     No Information Rate : 0.6415
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 0.6932
##
##  Mcnemar's Test P-Value : 4.376e-07
##
##             Sensitivity : 0.7941
##             Specificity : 0.9474
##          Pos Pred Value : 0.9643
##          Neg Pred Value : 0.7200
##              Prevalence : 0.6415
##          Detection Rate : 0.5094
##    Detection Prevalence : 0.5283
##       Balanced Accuracy : 0.8707
##
##        'Positive' Class : 1
##
```

```r
################################################################################
#           Model performance Discrimination and Calibration                   #
################################################################################

# Load packages
library(pROC)

# Plot the ROC curve
plot(roc_obj, col = "red", main = "ROC Curve")


# Get optimal threshold (Youden's index)
opt_coords<- coords(roc_obj, "best", best.method = "youden",
                        ret = c("threshold", "sensitivity", "specificity"))
print(opt_coords)
```

```
##   threshold sensitivity specificity
## 1 0.7703686   0.7941176   0.9473684
```
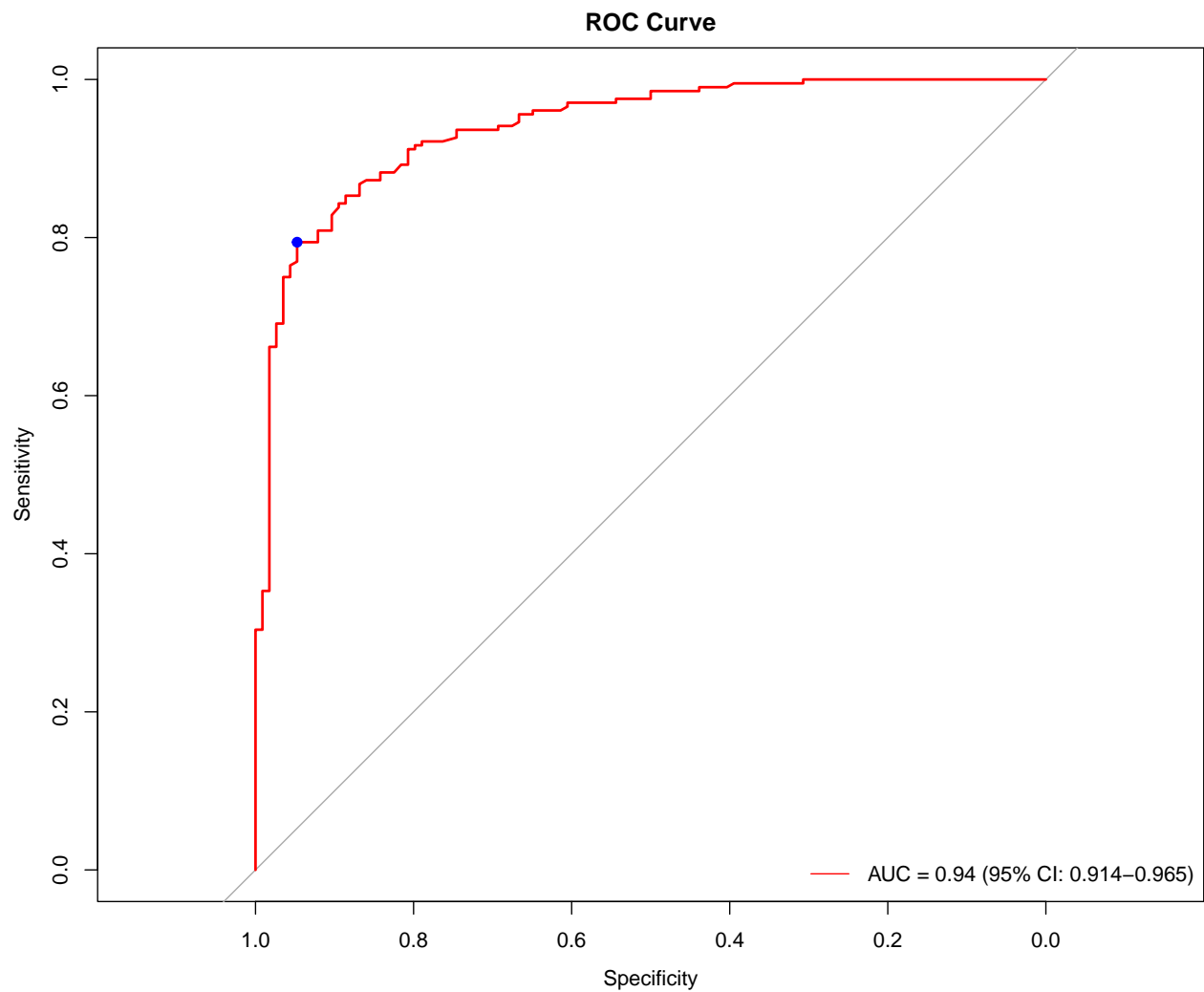
```r
# Add the optimal threshold point
points(opt_coords["specificity"], opt_coords["sensitivity"], col = "blue", pch = 19)

# Calculate AUC with CI
ci_auc <- ci.auc(roc_obj)
```
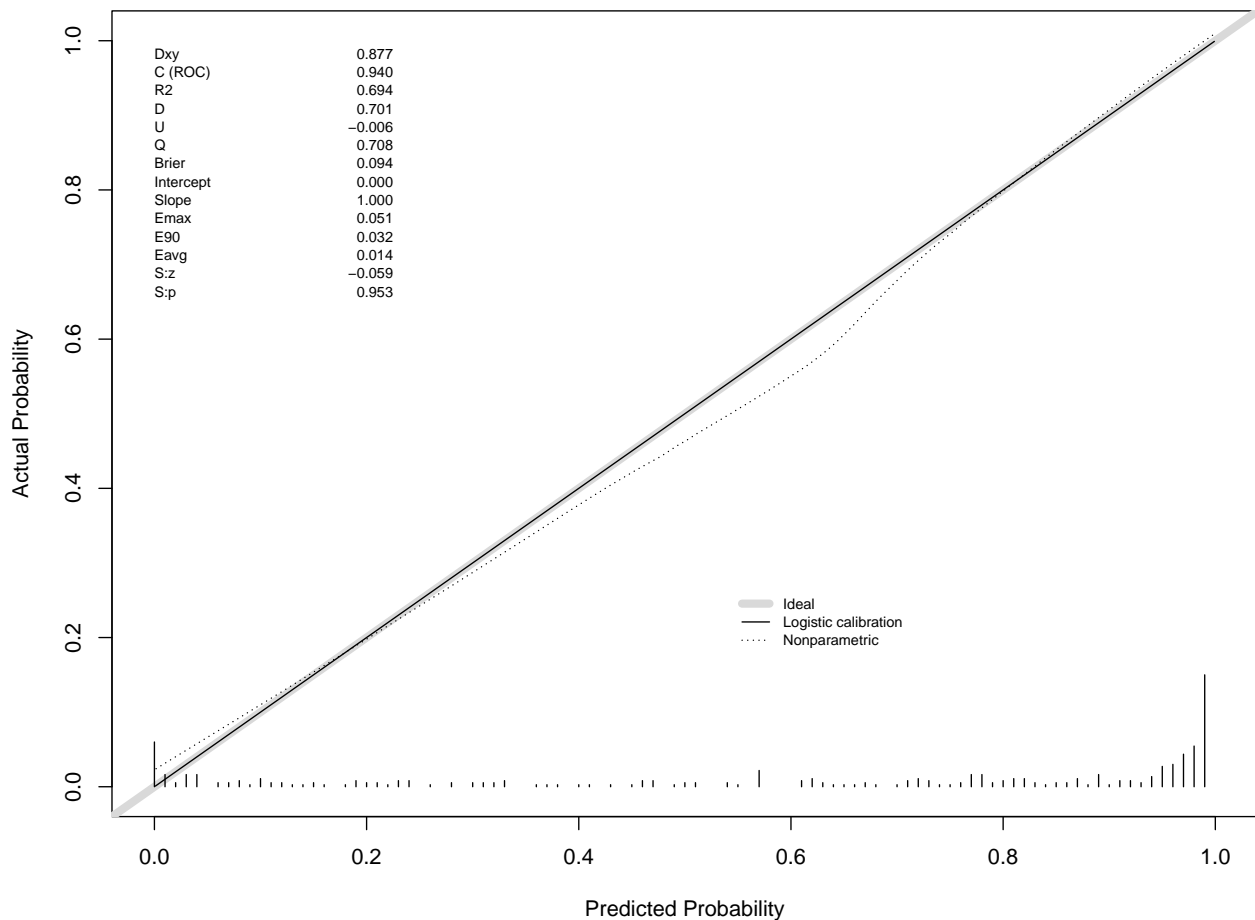
```r
# Add a legend with AUC and 95% CI
legend("bottomright",
       legend = paste0("AUC = ", round(auc(roc_obj), 3),
                       " (95% CI: ", round(ci_auc[1], 3), "-",
                       round(ci_auc[3], 3), ")"),
       col = "red",
       lty = 1,
       bty = "n")
```

ROC Curve


```r
# Calibration plot

library(rms)
val.prob(predict(finalmodel, type = "response"), finalmodel$y)
```

Actual Probability

Dxy	0.877
C (ROC)	0.940
R2	0.694
D	0.701
U	−0.006
Q	0.708
Brier	0.094
Intercept	0.000
Slope	1.000
Emax	0.051
E90	0.032
Eavg	0.014
S:z	−0.059
S:p	0.953

Ideal
Logistic calibration
Nonparametric

Predicted Probability

```
##         Dxy       C (ROC)            R2             D        D:Chi-sq
## 8.774661e-01  9.398650e-01  6.937865e-01  7.014022e-01  2.240459e+02
##         D:p             U       U:Chi-sq           U:p             Q
## 0.000000e+00 -6.289308e-03  1.705303e-13  1.000000e+00  7.076915e-01
##       Brier     Intercept         Slope          Emax           E90
## 9.375293e-02 -2.651423e-10  1.000000e+00  5.064930e-02  3.245855e-02
##        Eavg           S:z           S:p
## 1.443283e-02 -5.942306e-02  9.526151e-01
```

```r
################################################################################
#                     Internal validation (80/20 split)                       #
################################################################################

# Load library (if not already loaded)
library(pROC)

# Predict on test data
test_data$pred_prob <- predict(finalmodel, newdata = test_data, type = "response")

# ROC on test data
roc_test <- roc(test_data$apo, test_data$pred_prob)

# Compute AUC with 95% CI
ci_auc_test <- ci.auc(roc_test)
```
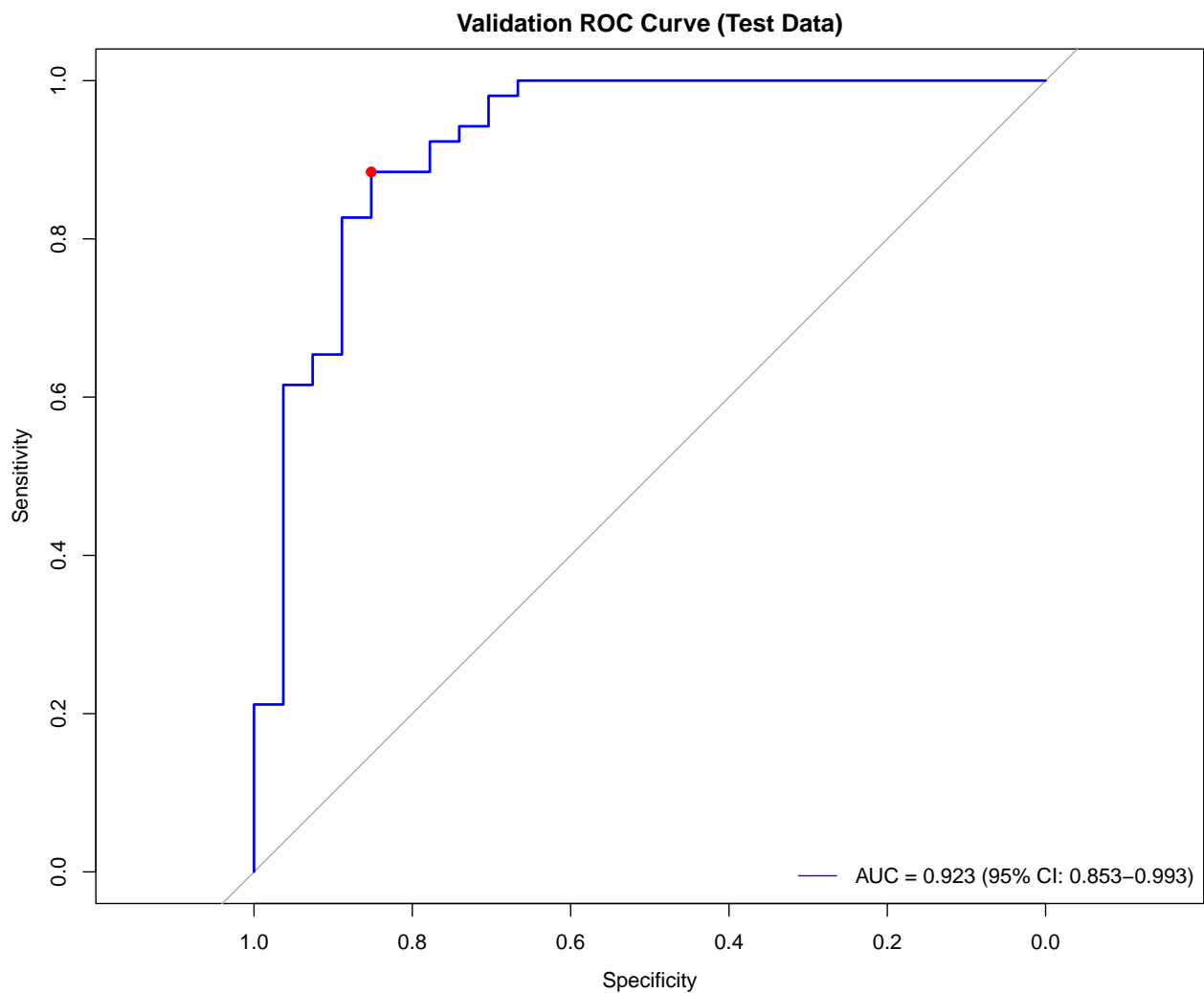
```r
# Get optimal threshold (Youden's index)
opt_coords_test <- coords(roc_test, "best", best.method = "youden",
                          ret = c("threshold", "sensitivity", "specificity"))

# Plot ROC
plot(roc_test, col = "blue", main = "Validation ROC Curve (Test Data)")

# Add optimal threshold point
points(opt_coords_test["specificity"], opt_coords_test["sensitivity"],
       col = "red", pch = 19)

# Add legend with AUC and CI
legend("bottomright",
       legend = paste0("AUC = ", round(auc(roc_test), 3),
                       " (95% CI: ", round(ci_auc_test[1], 3), "-",
                       round(ci_auc_test[3], 3), ")"),
       col = "blue",
       lty = 1,
       bty = "n")
```


Validation ROC Curve (Test Data)

```r
################################################################################
##----------------------------------------------------------------------------##
##                    Decision Curve Analysis (DCA) for APO                   ##
##----------------------------------------------------------------------------##
################################################################################

library(ggplot2)
library(reshape2)
library(rmda)  # Make sure this is loaded

table(data_complete$apo_numeric)
```

```
## < table of extent 0 >
```

```r
# Convert outcome variable apo to numeric: 1 for "Yes", 0 for "No"
data_complete$apo_numeric <- ifelse(data_complete$apo == "Yes", 1, 0)

library(rmda)

# Step 1: Prepare predicted risk scores from the logistic model
# Assume 'finalmodel' is your fitted logistic model
data_complete$pred_risk <- predict(finalmodel, type = "response")  # predicted probabilities

# Step 2: Run DCA
dca_result <- decision_curve(
  apo_numeric ~ pred_risk,
  data = data_complete,
  thresholds = seq(0.01, 0.99, by = 0.01),
  fitted.risk = TRUE)

# Warning may appear for missing data - that's okay if only a few are removed

# Step 3: Plot with cost:benefit axis and custom legend
plot_decision_curve(
  dca_result,
  curve.names = "Logistic Model",
  cost.benefit.axis = TRUE,
  confidence.intervals = FALSE,
  standardize = TRUE,
  col = c("#1F77B4", "darkred", "darkgreen"),  # Blue, Red, Green
  lty = c(1, 2, 3),  # Solid, Dashed, Dotted
  lwd = 2,
  legend.position = "none",  # Remove default legend
  xlab = "Risk Threshold (Probability)",
  ylab = "Standardized Net Benefit",
  main = "Decision Curve Analysis for Adverse Birth Outcome")

# Step 4: Add custom legend
legend(
  x = 0.95, y = 0.85,
  legend = c("Logistic Model", "Treat All", "Treat None"),
  col = c("#1F77B4", "darkred", "darkgreen"),
  lty = c(1, 2, 3),
```
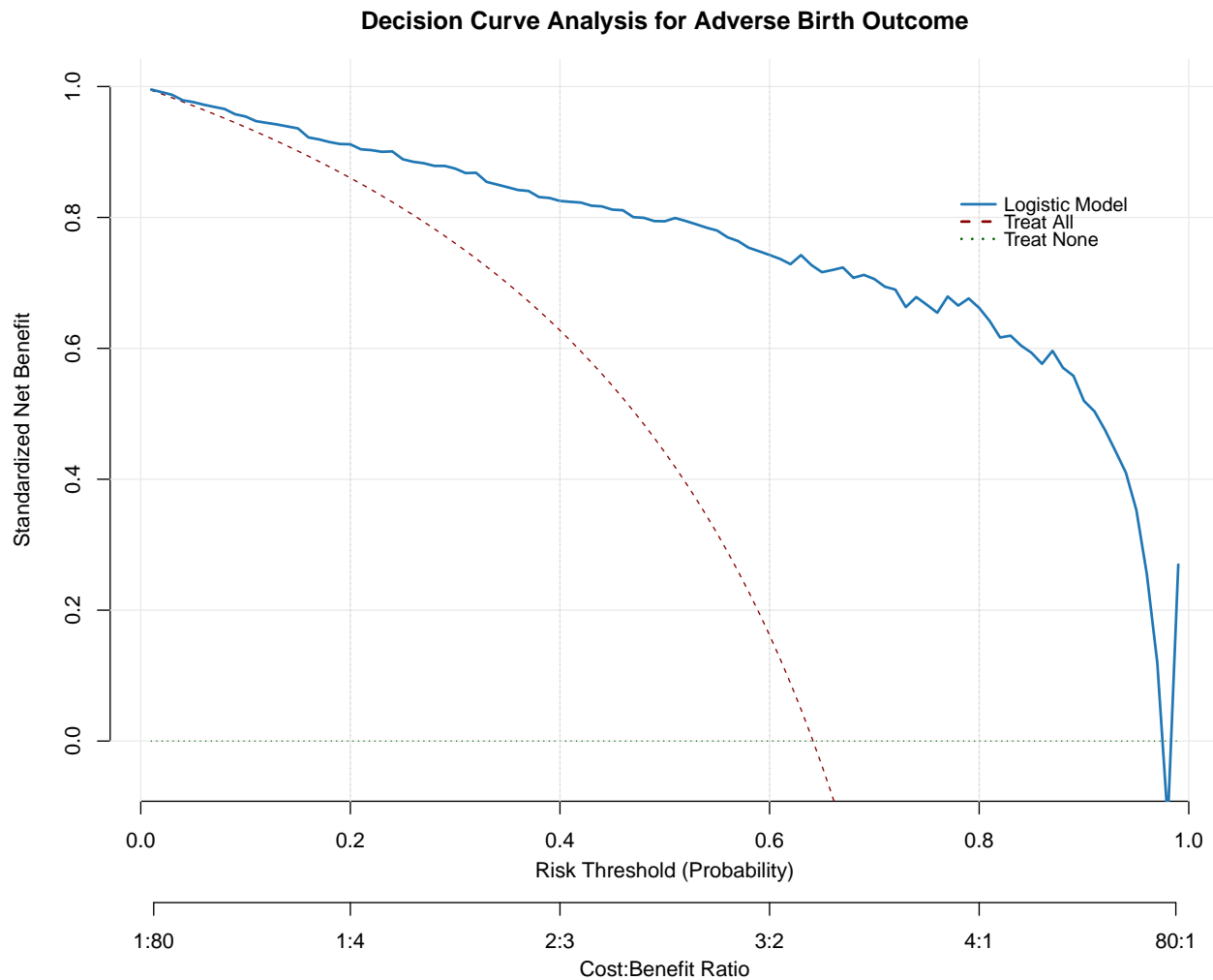
```
  lwd = 2,
  bty = "n",        # No box around the legend
  xjust = 1,        # Right-align x
  yjust = 1,        # Top-align y
  cex = 0.9,        # Slightly smaller text
  x.intersp = 0.5,
  y.intersp = 0.8)

# Optional: Add vertical reference lines at selected thresholds
abline(v = c(0.2, 0.4, 0.6, 0.8), col = "gray80", lty = 3)
```

Decision Curve Analysis for Adverse Birth Outcome


```
summary(finalmodel)
```

```
##
## Call:
## glm(formula = apo ~ healthinsu + placeresid + reparity + twin +
##     childgender + iugr + infcause + afc + embryotype + embryostage +
##     dmhx, family = binomial, data = data_complete)
##
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.7847     0.5474  -1.434 0.151714
## healthinsu2    1.5812     0.5684   2.782 0.005402 **
## placeresid2   -1.1472     0.4287  -2.676 0.007450 **
## reparity2     -1.2546     0.3907  -3.211 0.001321 **
## twin2          1.3516     0.3989   3.388 0.000704 ***
## childgender2   1.3716     0.3936   3.485 0.000493 ***
## iugr2          1.5546     0.5227   2.974 0.002937 **
## infcause2      0.6935     0.4466   1.553 0.120486
## infcause3      1.8286     0.5004   3.654 0.000258 ***
## afc2           2.3517     0.4110   5.722 1.06e-08 ***
## embryotype2   -1.7081     0.4030  -4.238 2.26e-05 ***
## embryostage2  -2.2753     0.4507  -5.049 4.45e-07 ***
## dmhx2          2.3253     0.6296   3.693 0.000221 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 415.02  on 317  degrees of freedom
## Residual deviance: 190.97  on 305  degrees of freedom
## AIC: 216.97
##
## Number of Fisher Scoring iterations: 6


################################################################################
##----------------------------------------------------------------------------##
##                      Web Appliaction Development                          ##
##----------------------------------------------------------------------------##
################################################################################

#library(shiny)
#
#ui <- fluidPage(
#  titlePanel("Risk of Adverse Birth Outcome (APO) following Assisted Reproductive Technology Pregnancy
#
#  sidebarLayout(
#    sidebarPanel(
#      h4("Patient Information"),
#
#      selectInput("healthinsu", "Health Insurance:",
#                  choices = c("No" = 1, "Yes" = 2)),
#
#      selectInput("placeresid", "Place of Residence:",
#                  choices = c("Urban" = 1, "Rural" = 2)),
#
#      selectInput("reparity", "Parity:",
#                  choices = c("Multiparous" = 1, "Primiparous" = 2)),
#
#      selectInput("twin", "Twin Pregnancy:",
#                  choices = c("No" = 1, "Yes" = 2)),
#
#      selectInput("childgender", "Child Gender:",
#                  choices = c("Male" = 1, "Female" = 2)),
```

```r
#
#      selectInput("iugr", "IUGR:",
#                  choices = c("No" = 1, "Yes" = 2)),
#
#      selectInput("infcause", "Infertility Cause:",
#                  choices = c("Male Factor" = 1, "Female Factor" = 2, "Unexplained" = 3)),
#
#      selectInput("afc", "Antral Follicle Count (AFC):",
#                  choices = c(" 5" = 1, "<5" = 2)),
#
#    selectInput("embryotype", "Embryo Type:",
#                  choices = c("Fresh" = 1, "Frozen" = 2)),
#
#      selectInput("embryostage", "Embryo Stage:",
#                  choices = c("Blastocyte" = 1, "Cleavage" = 2)),
#
#      selectInput("dmhx", "History of Diabetes Mellitus:",
#                  choices = c("No" = 1, "Yes" = 2)),
#
#      actionButton("predict", "Predict Risk")
#    ),
#
#    mainPanel(
#      h3("Prediction Result"),
#      verbatimTextOutput("riskOutput")
#    )
#  )
#)

#server <- function(input, output) {

#  Logistic regression coefficients (from your final model)
#  coefs <- c(
#    `(Intercept)` = -0.7847,
#    healthinsu2 = 1.5812,
#    placeresid2 = -1.1472,
#    reparity2 = -1.2546,
#    twin2 = 1.3516,
#    childgender2 = 1.3716,
#    iugr2 = 1.5546,
#    infcause2 = 0.6935,
#    infcause3 = 1.8286,
#    afc2 = 2.3517,
#    embryotype2 = -1.7081,
#    embryostage2 = -2.2753,
#    dmhx2 = 2.3253
#  )

  # Set optimal cutoff value (replace 0.65 by your real threshold)
#  cutoff <- 0.7703686

#  predict_risk <- function(input) {
#    x <- c(
```

```
#      1,
#      as.numeric(input$healthinsu) == 2,
#      as.numeric(input$placeresid) == 2,
#      as.numeric(input$reparity) == 2,
#      as.numeric(input$twin) == 2,
#      as.numeric(input$childgender) == 2,
#      as.numeric(input$iugr) == 2,
 #     as.numeric(input$infcause) == 2,
#      as.numeric(input$infcause) == 3,
#      as.numeric(input$afc) == 2,
#      as.numeric(input$embryotype) == 2,
#      as.numeric(input$embryostage) == 2,
#      as.numeric(input$dmhx) == 2
#    )
#
#    names(x) <- names(coefs)
#
#   lp <- sum(coefs * x)
#    prob <- 1 / (1 + exp(-lp))
#   return(prob)
#  }
#
#  observeEvent(input$predict, {
#    prob <- predict_risk(input)
#    risk_class <- ifelse(prob >= cutoff, "HIGH RISK", "LOW RISK")
#
#    output$riskOutput <- renderText({
#      paste0("Predicted Risk: ", round(prob * 100, 2), "%\nRisk Classification: ", risk_class)
#    })
#  })
#}
#
#shinyApp(ui = ui, server = server)
```

Including Plots

You can also embed plots, for example:

The End of Analysis