

PM-592 Final Project

Jaxon Abercrombie

12/10/2021

Functions

```
lrt.func <- function(model){
  model %>%
    anova(test = "LRT")
}

fp.func <- function(variable2){
  mfp(death ~ fp(variable2), data = heartfailure, family = binomial)
}

plot_resid_lev_logistic <- function(model)
{
  resp <- names(model.frame(model))[1]
  title <- paste("Outlier and Leverage Diagnostics for",
    resp)

  g <- dx(model, byCov=T) %>%
    rownames_to_column() %>%
    as_tibble()
  lthresh <- round(mean(g$h, na.rm=T)*2, 3)
  ann_label <- paste("Threshold:", lthresh)
  f <- g[,c("rowname", "h", "dChisq", "dDev", "dBhat", "sPr")]
  f$numinf <- as.integer(
    (f$dChisq > 4) + (f$dDev > 4) + (f$dBhat > 1)
  )
  f$obs <- case_when(
    (f$h > lthresh) & (abs(f$sPr) > 2) ~ "Influence",
    (f$h > lthresh) ~ "Leverage",
    (abs(f$sPr) > 2) ~ "Outlier",
    TRUE ~ "Normal"
  )
  f$txt <- ifelse(f$obs == "Normal", NA, f$rowname)

  ggplot(f, aes(h, sPr, label = txt)) +
    geom_point(shape = 1, aes(colour = obs, size = 0.5+0.25*numinf)) +
    scale_colour_manual(values=c("Influence" = "red", "Outlier" = "maroon",
      "Leverage" = "darkgreen", "Normal" = "blue")) +
    labs(colour = "Observation",
      x = "Leverage",
      y = "Pearson's Residual",
```

```

    title = title) +
  geom_hline(yintercept = c(2, -2), colour = "maroon") +
  geom_vline(xintercept = lthresh, colour = "maroon") +
  geom_text(vjust = -1, size = 3, family = "serif",
            fontface = "italic", colour = "darkred") +
  annotate("text", x = Inf, y = Inf, hjust = 1.2,
          vjust = 2, family = "serif", fontface = "italic",
          colour = "darkred", label = ann_label) +
  scale_size(guide = "none")
}

```

Data Section

```

# Load in Dataset
heartfailure <- read.csv("heartfailure.csv")

# Clean, Wrangle, and Explore Data
names(heartfailure)[13] <- "death"

names(heartfailure) # examine variable names and decide which are of interest

## [1] "age" "anaemia"
## [3] "creatinine_phosphokinase" "diabetes"
## [5] "ejection_fraction" "high_blood_pressure"
## [7] "platelets" "serum_creatinine"
## [9] "serum_sodium" "sex"
## [11] "smoking" "time"
## [13] "death"

dim(heartfailure) # check dimensions to make sure they match source website

## [1] 299 13

sum(is.na(heartfailure)) # ensure there are no cells with NA

## [1] 0

# Summary statistics found in Table 1 in later code chunk

```

Model Building Code and Output - Univariate

```

## MAIN INDEPENDENT CONTINUOUS VARIABLES FIRST

# Creatinine Phosphokinase
cp.m <- glm(death ~ creatinine_phosphokinase,

```

```

family = binomial, data = heartfailure)
summary(cp.m)

```

```

##
## Call:
## glm(formula = death ~ creatinine_phosphokinase, family = binomial,
##      data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1141  -0.8792  -0.8573   1.5084   1.5411
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.8265731   0.1447064  -5.712 1.12e-08 ***
## creatinine_phosphokinase  0.0001297  0.0001218   1.065   0.287
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 374.23  on 297  degrees of freedom
## AIC: 378.23
##
## Number of Fisher Scoring iterations: 4

```

```

mfp(death ~ fp(creatinine_phosphokinase),
    data = heartfailure, family = binomial)

```

```

## Call:
## mfp(formula = death ~ fp(creatinine_phosphokinase), data = heartfailure,
##      family = binomial)
##
##
## Deviance table:
##           Resid. Dev
## Null model    375.3488
## Linear model  374.2305
## Final model   374.2305
##
## Fractional polynomials:
##           df.initial select alpha df.final power1 power2
## creatinine_phosphokinase      4      1 0.05      1      1      .
##
##
## Transformations of covariates:
##                                     formula
## creatinine_phosphokinase I((creatinine_phosphokinase/1000)^1)
##
## Rescaled coefficients:
##           Intercept creatinine_phosphokinase.1
##           -0.8265731             0.0001297
##

```

```

## Degrees of Freedom: 298 Total (i.e. Null); 297 Residual
## Null Deviance: 375.3
## Residual Deviance: 374.2 AIC: 378.2

lrt.func(cp.m)

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)
##
##
##              Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                                298      375.35
## creatinine_phosphokinase 1    1.1182      297      374.23  0.2903

# Ejection Fraction
ef.m <- glm(death ~ ejection_fraction,
            family = binomial, data = heartfailure)
summary(ef.m)

##
## Call:
## glm(formula = death ~ ejection_fraction, family = binomial, data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3320  -0.9146  -0.7201   1.2173   2.3205
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.31169    0.46278   2.834  0.00459 **
## ejection_fraction -0.05620    0.01258  -4.468 7.88e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 351.97  on 297  degrees of freedom
## AIC: 355.97
##
## Number of Fisher Scoring iterations: 4

mfp(death ~ fp(ejection_fraction),
    data = heartfailure, family = binomial)

## Call:
## mfp(formula = death ~ fp(ejection_fraction), data = heartfailure,
##      family = binomial)
##
##
## Deviance table:

```

```
##          Resid. Dev
## Null model    375.3488
## Linear model  351.9682
## Final model   334.7072
##
## Fractional polynomials:
##          df.initial select alpha df.final power1 power2
## ejection_fraction      4      1 0.05      2      -2      .
##
## Transformations of covariates:
##                                formula
## ejection_fraction I((ejection_fraction/100)^-2)
##
## Rescaled coefficients:
##          Intercept  ejection_fraction.1
##          -1.969      1256.112
##
## Degrees of Freedom: 298 Total (i.e. Null);  297 Residual
## Null Deviance:      375.3
## Residual Deviance: 334.7      AIC: 338.7
heartfailure <-
  heartfailure %>%
  mutate(ef.fp = (1/sqrt(ejection_fraction))) # mutate based on fp output
glm(death ~ ef.fp, family = binomial, data = heartfailure) %>%
  anova(test = "LRT")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)
##
##          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                298      375.35
## ef.fp  1      34.605      297      340.74 4.038e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ef.m2 <- glm(death ~ ef.fp,
             family = binomial, data = heartfailure) # new model with transformation
lrt.func(ef.m2)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)
##
##
```

```

##           Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                298      375.35
## ef.fp  1    34.605          297      340.74 4.038e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Platelets
platelets.m <- glm(death ~ platelets,
                  family = binomial, data = heartfailure)
summary(platelets.m)

##
## Call:
## glm(formula = death ~ platelets, family = binomial, data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9794 -0.8908 -0.8644  1.4731  1.6882
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.573e-01  3.632e-01  -1.259   0.208
## platelets   -1.115e-06  1.316e-06  -0.847   0.397
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 374.61  on 297  degrees of freedom
## AIC: 378.61
##
## Number of Fisher Scoring iterations: 4

mfp(death ~ fp(platelets),
    data = heartfailure, family = binomial)

## Call:
## mfp(formula = death ~ fp(platelets), data = heartfailure, family = binomial)
##
##
## Deviance table:
##              Resid. Dev
## Null model      375.3488
## Linear model    374.611
## Final model     374.611
##
## Fractional polynomials:
##              df.initial select alpha df.final power1 power2
## platelets           4       1 0.05           1       1     .
##
##
## Transformations of covariates:
##              formula
## platelets I((platelets/1e+05)^1)
##
## Rescaled coefficients:

```

```

## Intercept platelets.1
## -4.573e-01 -1.115e-06
##
## Degrees of Freedom: 298 Total (i.e. Null); 297 Residual
## Null Deviance: 375.3
## Residual Deviance: 374.6 AIC: 378.6

lrt.func(platelets.m)

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)
##
##
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL 298 375.35
## platelets 1 0.73778 297 374.61 0.3904

# Serum Creatinine
sc.m <- glm(death ~ serum_creatinine,
            family = binomial, data = heartfailure)
summary(sc.m)

##
## Call:
## glm(formula = death ~ serum_creatinine, family = binomial, data = heartfailure)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.5213 -0.7966 -0.7417 1.2644 1.7990
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.8917 0.2939 -6.438 1.21e-10 ***
## serum_creatinine 0.8242 0.1972 4.180 2.91e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 375.35 on 298 degrees of freedom
## Residual deviance: 347.25 on 297 degrees of freedom
## AIC: 351.25
##
## Number of Fisher Scoring iterations: 5

mfp(death ~ fp(serum_creatinine),
    data = heartfailure, family = binomial)

## Call:
## mfp(formula = death ~ fp(serum_creatinine), data = heartfailure,
## family = binomial)

```

```
##
##
## Deviance table:
##      Resid. Dev
## Null model    375.3488
## Linear model  347.2521
## Final model   333.4361
##
## Fractional polynomials:
##      df.initial select alpha df.final power1 power2
## serum_creatinine      4      1 0.05      2      -1      .
##
##
## Transformations of covariates:
##                      formula
## serum_creatinine I(serum_creatinine^-1)
##
## Rescaled coefficients:
##      Intercept  serum_creatinine.1
##          1.642          -2.820
##
## Degrees of Freedom: 298 Total (i.e. Null); 297 Residual
## Null Deviance:      375.3
## Residual Deviance: 333.4      AIC: 337.4
```

```
heartfailure <-
  heartfailure %>%
  mutate(sc.fp = (1/serum_creatinine)) # mutate based on fp output
glm(death ~ sc.fp, family = binomial, data = heartfailure) %>%
  anova(test = "LRT")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)
##
##      Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                298      375.35
## sc.fp 1    41.913      297      333.44 9.544e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sc.m2 <- glm(death ~ sc.fp,
             family = binomial, data = heartfailure) # new model with transformation
lrt.func(sc.m)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
```



```

## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                      298      375.35
## serum_creatinine  1    28.097      297      347.25 1.154e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Serum Sodium
ss.m <- glm(death ~ serum_sodium,
            family = binomial, data = heartfailure)
summary(ss.m)

##
## Call:
## glm(formula = death ~ serum_sodium, family = binomial, data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8459  -0.8928  -0.7582   1.3197   1.9231
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  12.39442     4.07264   3.043  0.00234 **
## serum_sodium -0.09639     0.02989  -3.224  0.00126 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 364.02  on 297  degrees of freedom
## AIC: 368.02
##
## Number of Fisher Scoring iterations: 4

mfp(death ~ fp(serum_sodium),
    data = heartfailure, family = binomial)

## Call:
## mfp(formula = death ~ fp(serum_sodium), data = heartfailure,
##      family = binomial)
##
##
## Deviance table:
##           Resid. Dev
## Null model    375.3488
## Linear model  364.0221
## Final model   364.0221
##
## Fractional polynomials:
##           df.initial select alpha df.final power1 power2
## serum_sodium         4      1 0.05         1      1      .
##

```

```
##
## Transformations of covariates:
##               formula
## serum_sodium I((serum_sodium/100)^1)
##
## Rescaled coefficients:
##      Intercept  serum_sodium.1
##      12.39442      -0.09639
##
## Degrees of Freedom: 298 Total (i.e. Null); 297 Residual
## Null Deviance:      375.3
## Residual Deviance: 364  AIC: 368

lrt.func(ss.m)

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)
##
##
##              Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL              298      375.35
## serum_sodium  1    11.327      297      364.02 0.000764 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## BINOMIAL VARIABLES AFTER
# Anaemia
anaemia.m <- glm(death ~ anaemia,
                 family = binomial, data = heartfailure)
summary(anaemia.m)

##
## Call:
## glm(formula = death ~ anaemia, family = binomial, data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9391 -0.9391 -0.8346  1.4361  1.5645
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.8755     0.1683  -5.201 1.98e-07 ***
## anaemia       0.2853     0.2492   1.145  0.252
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 374.04  on 297  degrees of freedom
## AIC: 378.04
```

```

##
## Number of Fisher Scoring iterations: 4
lrt.func(anaemia.m)

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                298      375.35
## anaemia  1      1.3086      297      374.04  0.2527
# High Blood Pressure
hbp.m <- glm(death ~ high_blood_pressure,
             family = binomial, data = heartfailure)
summary(hbp.m)

##
## Call:
## glm(formula = death ~ high_blood_pressure, family = binomial,
##      data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9636  -0.8341  -0.8341   1.4074   1.5651
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.8769     0.1576  -5.564 2.64e-08 ***
## high_blood_pressure  0.3508     0.2562   1.369  0.171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 373.49  on 297  degrees of freedom
## AIC: 377.49
##
## Number of Fisher Scoring iterations: 4
lrt.func(hbp.m)

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)

```

```
##
##
##           Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                      298      375.35
## high_blood_pressure  1      1.863      297      373.49  0.1723
## POTENTIAL CONFOUNDERS AND EFFECT MODIFIERS LAST
# Age
age.m <- glm(death ~ age, family = binomial, data = heartfailure)
summary(age.m)

##
## Call:
## glm(formula = death ~ age, family = binomial, data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4276  -0.8993  -0.6922   1.2344   1.9251
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.65433    0.70662  -5.172 2.32e-07 ***
## age          0.04695    0.01107   4.241 2.23e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 355.99  on 297  degrees of freedom
## AIC: 359.99
##
## Number of Fisher Scoring iterations: 4
mfp(death ~ fp(age),
    data = heartfailure, family = binomial)

## Call:
## mfp(formula = death ~ fp(age), data = heartfailure, family = binomial)
##
##
## Deviance table:
##           Resid. Dev
## Null model    375.3488
## Linear model   355.9928
## Final model    355.9928
##
## Fractional polynomials:
##      df.initial select alpha df.final power1 power2
## age           4      1 0.05          1      1      .
##
##
## Transformations of covariates:
##           formula
## age I((age/100)^1)
```

```
##
## Rescaled coefficients:
## Intercept      age.1
## -3.65433      0.04695
##
## Degrees of Freedom: 298 Total (i.e. Null); 297 Residual
## Null Deviance:      375.3
## Residual Deviance: 356   AIC: 360
lrt.func(age.m)

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                298      375.35
## age  1   19.356      297      355.99 1.085e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Sex
sex.m <- glm(death ~ sex,
             family = binomial, data = heartfailure)
summary(sex.m)

##
## Call:
## glm(formula = death ~ sex, family = binomial, data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8846  -0.8776  -0.8776   1.5017   1.5105
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.73632    0.20856  -3.531 0.000415 ***
## sex         -0.01935    0.25923  -0.075 0.940504
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 375.34  on 297  degrees of freedom
## AIC: 379.34
##
## Number of Fisher Scoring iterations: 4
```

```
lrt.func(sex.m)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)
##
##
##      Df  Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                298    375.35
## sex   1 0.0055669    297    375.34  0.9405
```

```
# Diabetes
diabetes.m <- glm(death ~ diabetes,
                  family = binomial, data = heartfailure)
summary(diabetes.m)
```

```
##
## Call:
## glm(formula = death ~ diabetes, family = binomial, data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8813  -0.8813  -0.8782   1.5058   1.5096
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.745333   0.162270  -4.593 4.37e-06 ***
## diabetes    -0.008439   0.251190  -0.034  0.973
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 375.35  on 297  degrees of freedom
## AIC: 379.35
##
## Number of Fisher Scoring iterations: 4
```

```
lrt.func(diabetes.m)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)
##
##
##      Df  Deviance Resid. Df Resid. Dev Pr(>Chi)
```

```
## NULL                298      375.35
## diabetes  1 0.0011289      297      375.35      0.9732

# Smoking
smoke.m <- glm(death ~ smoking,
               family = binomial, data = heartfailure)
summary(smoke.m)

##
## Call:
## glm(formula = death ~ smoking, family = binomial, data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8868  -0.8868  -0.8657   1.4990   1.5252
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.73033    0.14984  -4.874 1.09e-06 ***
## smoking      -0.05813    0.26634  -0.218   0.827
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 375.30  on 297  degrees of freedom
## AIC: 379.3
##
## Number of Fisher Scoring iterations: 4

lrt.func(smoke.m)

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: death
##
## Terms added sequentially (first to last)
##
##              Df Deviance Resid. Df Resid. Dev Pr(>Chi)
## NULL                298      375.35
## smoking  1 0.047765      297      375.30      0.827
```

Model Building and Output - Multivariate

```
# Based on the univariate analyses and clinical significance of variables, the preliminary final model b

pfm1 <- glm(death ~ anaemia + creatinine_phosphokinase + ef.fp +
```

```

        high_blood_pressure + platelets + sc.fp + serum_sodium,
        family = binomial, data = heartfailure)
summary(pfm1)

```

```

##
## Call:
## glm(formula = death ~ anaemia + creatinine_phosphokinase + ef.fp +
##      high_blood_pressure + platelets + sc.fp + serum_sodium, family = binomial,
##      data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8833  -0.7409  -0.4846   0.8203   2.8623
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.349e+00  4.683e+00   0.288   0.7732
## anaemia         5.014e-01  2.980e-01   1.683   0.0924 .
## creatinine_phosphokinase 2.871e-04  1.404e-04   2.045   0.0408 *
## ef.fp          2.532e+01  5.500e+00  4.603 4.16e-06 ***
## high_blood_pressure    6.282e-01  3.039e-01   2.067   0.0388 *
## platelets        -4.790e-07  1.524e-06  -0.314   0.7533
## sc.fp           -2.753e+00  5.260e-01  -5.233 1.67e-07 ***
## serum_sodium     -3.370e-02  3.338e-02  -1.010   0.3126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 296.42  on 291  degrees of freedom
## AIC: 312.42
##
## Number of Fisher Scoring iterations: 4

```

The first preliminary final model indicates statistical significance with creat. phos., ef.fp, hbp, p

Adding diabetes (potential confounder)

```

pfm1.diabetes <- glm(death ~ anaemia + creatinine_phosphokinase + ef.fp +
        high_blood_pressure + platelets + sc.fp + serum_sodium + diabetes,
        family = binomial, data = heartfailure)
summary(pfm1)

```

```

##
## Call:
## glm(formula = death ~ anaemia + creatinine_phosphokinase + ef.fp +
##      high_blood_pressure + platelets + sc.fp + serum_sodium, family = binomial,
##      data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8833  -0.7409  -0.4846   0.8203   2.8623
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)

```



```
## (Intercept)          1.349e+00  4.683e+00  0.288  0.7732
## anaemia              5.014e-01  2.980e-01  1.683  0.0924 .
## creatinine_phosphokinase 2.871e-04  1.404e-04  2.045  0.0408 *
## ef.fp               2.532e+01  5.500e+00  4.603  4.16e-06 ***
## high_blood_pressure   6.282e-01  3.039e-01  2.067  0.0388 *
## platelets            -4.790e-07  1.524e-06 -0.314  0.7533
## sc.fp               -2.753e+00  5.260e-01 -5.233  1.67e-07 ***
## serum_sodium         -3.370e-02  3.338e-02 -1.010  0.3126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 375.35 on 298 degrees of freedom
## Residual deviance: 296.42 on 291 degrees of freedom
## AIC: 312.42
##
## Number of Fisher Scoring iterations: 4
```

```
summary(pfm1.diabetes)
```

```
##
## Call:
## glm(formula = death ~ anaemia + creatinine_phosphokinase + ef.fp +
##       high_blood_pressure + platelets + sc.fp + serum_sodium +
##       diabetes, family = binomial, data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8792  -0.7400  -0.4832   0.8196   2.8662
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.307e+00  4.709e+00  0.278  0.7813
## anaemia          5.017e-01  2.980e-01  1.684  0.0922 .
## creatinine_phosphokinase 2.875e-04  1.405e-04  2.045  0.0408 *
## ef.fp           2.533e+01  5.503e+00  4.603  4.16e-06 ***
## high_blood_pressure   6.291e-01  3.041e-01  2.068  0.0386 *
## platelets          -4.913e-07  1.529e-06 -0.321  0.7480
## sc.fp             -2.753e+00  5.260e-01 -5.234  1.66e-07 ***
## serum_sodium        -3.346e-02  3.349e-02 -0.999  0.3177
## diabetes           2.529e-02  2.926e-01  0.086  0.9311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 375.35 on 298 degrees of freedom
## Residual deviance: 296.42 on 290 degrees of freedom
## AIC: 314.42
##
## Number of Fisher Scoring iterations: 4
```

```
# Adding smoking (potential confounder)
pfm1.smoke <- glm(death ~ anaemia + creatinine_phosphokinase + ef.fp +
```

```

        high_blood_pressure + platelets + sc.fp + serum_sodium + smoking,
        family = binomial, data = heartfailure)
summary(pfm1)

##
## Call:
## glm(formula = death ~ anaemia + creatinine_phosphokinase + ef.fp +
##      high_blood_pressure + platelets + sc.fp + serum_sodium, family = binomial,
##      data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8833  -0.7409  -0.4846   0.8203   2.8623
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.349e+00  4.683e+00   0.288   0.7732
## anaemia           5.014e-01  2.980e-01   1.683   0.0924 .
## creatinine_phosphokinase 2.871e-04  1.404e-04   2.045   0.0408 *
## ef.fp            2.532e+01  5.500e+00   4.603 4.16e-06 ***
## high_blood_pressure  6.282e-01  3.039e-01   2.067   0.0388 *
## platelets        -4.790e-07  1.524e-06  -0.314   0.7533
## sc.fp            -2.753e+00  5.260e-01  -5.233 1.67e-07 ***
## serum_sodium      -3.370e-02  3.338e-02  -1.010   0.3126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 296.42  on 291  degrees of freedom
## AIC: 312.42
##
## Number of Fisher Scoring iterations: 4
summary(pfm1.smoke)

```

```

##
## Call:
## glm(formula = death ~ anaemia + creatinine_phosphokinase + ef.fp +
##      high_blood_pressure + platelets + sc.fp + serum_sodium +
##      smoking, family = binomial, data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8844  -0.7406  -0.4844   0.8205   2.8628
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.349e+00  4.683e+00   0.288   0.7732
## anaemia           5.017e-01  2.996e-01   1.675   0.0940 .
## creatinine_phosphokinase 2.871e-04  1.404e-04   2.045   0.0408 *
## ef.fp            2.532e+01  5.514e+00   4.591 4.41e-06 ***
## high_blood_pressure  6.284e-01  3.047e-01   2.062   0.0392 *

```

```
## platelets          -4.794e-07  1.524e-06  -0.315   0.7531
## sc.fp              -2.753e+00  5.267e-01  -5.227  1.72e-07 ***
## serum_sodium       -3.370e-02  3.338e-02  -1.010   0.3126
## smoking             3.547e-03  3.120e-01   0.011   0.9909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 296.42  on 290  degrees of freedom
## AIC: 314.42
##
## Number of Fisher Scoring iterations: 4
# Adding age (potential effect modifier)
median(heartfailure$age)

## [1] 60

age.young <-
  heartfailure %>%
  filter(age < 60)

age.older <-
  heartfailure %>%
  filter(age >= 60)

pfml.ageY <-
  glm(death ~ anaemia + creatinine_phosphokinase + ef.fp +
      high_blood_pressure + platelets + sc.fp + serum_sodium + smoking,
      family = binomial, data = age.young)
summary(pfml.ageY)

##
## Call:
## glm(formula = death ~ anaemia + creatinine_phosphokinase + ef.fp +
##      high_blood_pressure + platelets + sc.fp + serum_sodium +
##      smoking, family = binomial, data = age.young)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.97226  -0.55684  -0.27100  -0.08014   3.10817
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    9.825e+00  9.205e+00   1.067  0.285814
## anaemia        7.691e-01  5.954e-01   1.292  0.196464
## creatinine_phosphokinase 5.892e-04  2.571e-04   2.291  0.021940 *
## ef.fp          4.044e+01  1.116e+01   3.625  0.000289 ***
## high_blood_pressure  7.567e-01  5.745e-01   1.317  0.187793
## platelets      -2.199e-06  3.349e-06  -0.657  0.511352
## sc.fp          -3.061e+00  1.030e+00  -2.971  0.002966 **
## serum_sodium   -1.140e-01  6.831e-02  -1.668  0.095233 .
## smoking        -8.868e-01  6.441e-01  -1.377  0.168570
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 142.271  on 128  degrees of freedom
## Residual deviance:  88.957  on 120  degrees of freedom
## AIC: 106.96
##
## Number of Fisher Scoring iterations: 6
pfm1.age0 <-
  glm(death ~ anaemia + creatinine_phosphokinase + ef.fp +
        high_blood_pressure + platelets + sc.fp + serum_sodium + smoking,
        family = binomial, data = age.older)
summary(pfm1.age0)

##
## Call:
## glm(formula = death ~ anaemia + creatinine_phosphokinase + ef.fp +
##      high_blood_pressure + platelets + sc.fp + serum_sodium +
##      smoking, family = binomial, data = age.older)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9343  -0.8435  -0.5758   1.0160   2.1319
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.868e+00  5.506e+00  -0.339  0.734404
## anaemia         4.038e-01  3.713e-01   1.088  0.276778
## creatinine_phosphokinase 3.675e-05  2.300e-04   0.160  0.873044
## ef.fp          1.950e+01  6.765e+00   2.882  0.003947 **
## high_blood_pressure  3.719e-01  3.774e-01   0.985  0.324379
## platelets       1.334e-06  2.005e-06   0.665  0.506045
## sc.fp          -2.504e+00  6.674e-01  -3.752  0.000175 ***
## serum_sodium    -5.364e-03  3.826e-02  -0.140  0.888507
## smoking         4.598e-01  3.850e-01   1.194  0.232428
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 226.17  on 169  degrees of freedom
## Residual deviance: 191.61  on 161  degrees of freedom
## AIC: 209.61
##
## Number of Fisher Scoring iterations: 4
# Adding sex (potential effect modifier)
sex.male <-
  heartfailure %>%
  filter(sex == 1)

sex.female <-
```

```

heartfailure %>%
  filter(sex == 0)

pfm1.sexM <-
  glm(death ~ anaemia + creatinine_phosphokinase + ef.fp +
       high_blood_pressure + platelets + sc.fp + serum_sodium + smoking,
       family = binomial, data = sex.male)
summary(pfm1.sexM)

```

```

##
## Call:
## glm(formula = death ~ anaemia + creatinine_phosphokinase + ef.fp +
##      high_blood_pressure + platelets + sc.fp + serum_sodium +
##      smoking, family = binomial, data = sex.male)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8097  -0.7436  -0.4677   0.7344   2.2659
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.225e-01  6.153e+00  -0.020  0.98411
## anaemia       3.342e-01  3.756e-01   0.890  0.37350
## creatinine_phosphokinase 2.213e-04  1.516e-04   1.460  0.14432
## ef.fp        3.218e+01  7.107e+00   4.529 5.94e-06 ***
## high_blood_pressure  5.805e-01  3.934e-01   1.476  0.14003
## platelets     1.353e-06  1.868e-06   0.724  0.46888
## sc.fp        -2.119e+00  6.459e-01  -3.280  0.00104 **
## serum_sodium  -3.848e-02  4.384e-02  -0.878  0.38014
## smoking      -6.398e-02  3.630e-01  -0.176  0.86008
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 243.10  on 193  degrees of freedom
## Residual deviance: 192.44  on 185  degrees of freedom
## AIC: 210.44
##
## Number of Fisher Scoring iterations: 4

```

```

pfm1.sexF <-
  glm(death ~ anaemia + creatinine_phosphokinase + ef.fp +
       high_blood_pressure + platelets + sc.fp + serum_sodium + smoking,
       family = binomial, data = sex.female)
summary(pfm1.sexF)

```

```

##
## Call:
## glm(formula = death ~ anaemia + creatinine_phosphokinase + ef.fp +
##      high_blood_pressure + platelets + sc.fp + serum_sodium +
##      smoking, family = binomial, data = sex.female)
##
## Deviance Residuals:

```

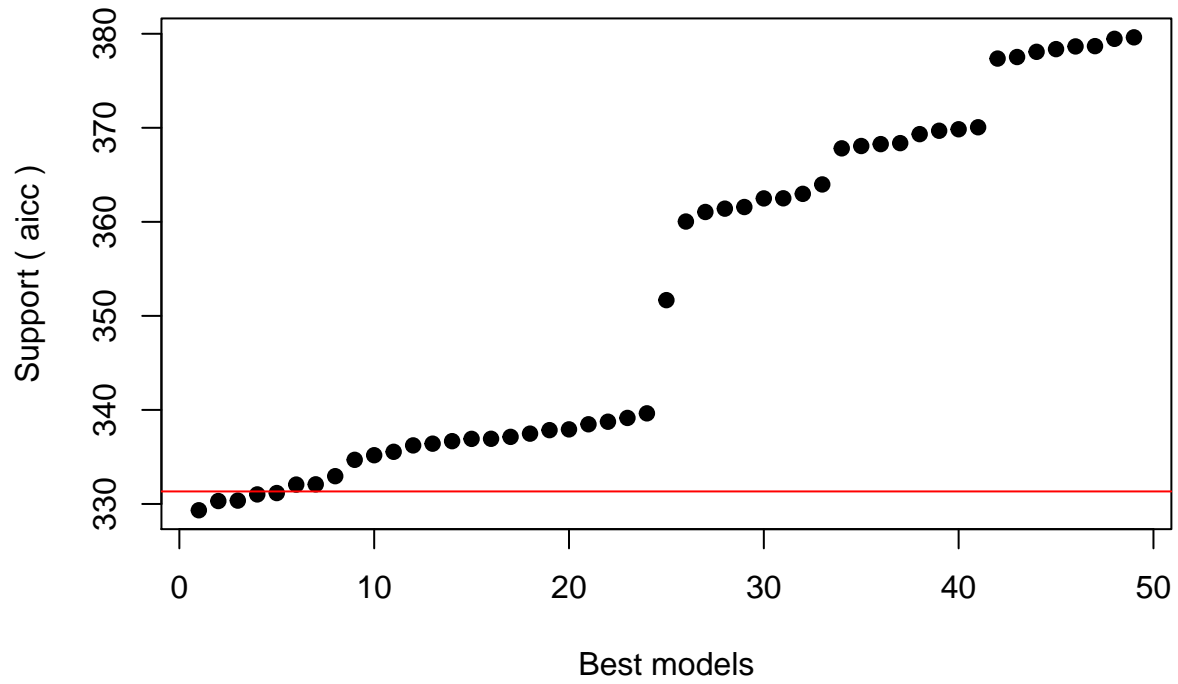
```
##      Min      1Q   Median      3Q      Max
## -1.9515 -0.6982 -0.3805  0.5188  3.1551
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      3.543e+00  8.334e+00  0.425 0.670758
## anaemia          1.046e+00  5.759e-01  1.816 0.069337 .
## creatinine_phosphokinase 7.583e-04  5.310e-04  1.428 0.153290
## ef.fp            1.471e+01  1.011e+01  1.456 0.145448
## high_blood_pressure 7.819e-01  5.640e-01  1.386 0.165604
## platelets        -5.195e-06  3.227e-06 -1.610 0.107391
## sc.fp            -4.503e+00  1.185e+00 -3.802 0.000144 ***
## serum_sodium     -2.041e-02  6.107e-02 -0.334 0.738296
## smoking           2.552e+00  1.739e+00  1.467 0.142311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 132.239  on 104  degrees of freedom
## Residual deviance:  90.573  on  96  degrees of freedom
## AIC: 108.57
##
## Number of Fisher Scoring iterations: 5
```

Assessing Final Model

```
# Run best subset with chosen independent variables
best_subset <- glmulti(death ~ anaemia +
  age:creatinine_phosphokinase + sex:ef.fp +
  high_blood_pressure + platelets +
  sc.fp + serum_sodium + age + sex,
  data = heartfailure, level = 1,
  family = binomial, crit="aicc",
  confsetsize=128)

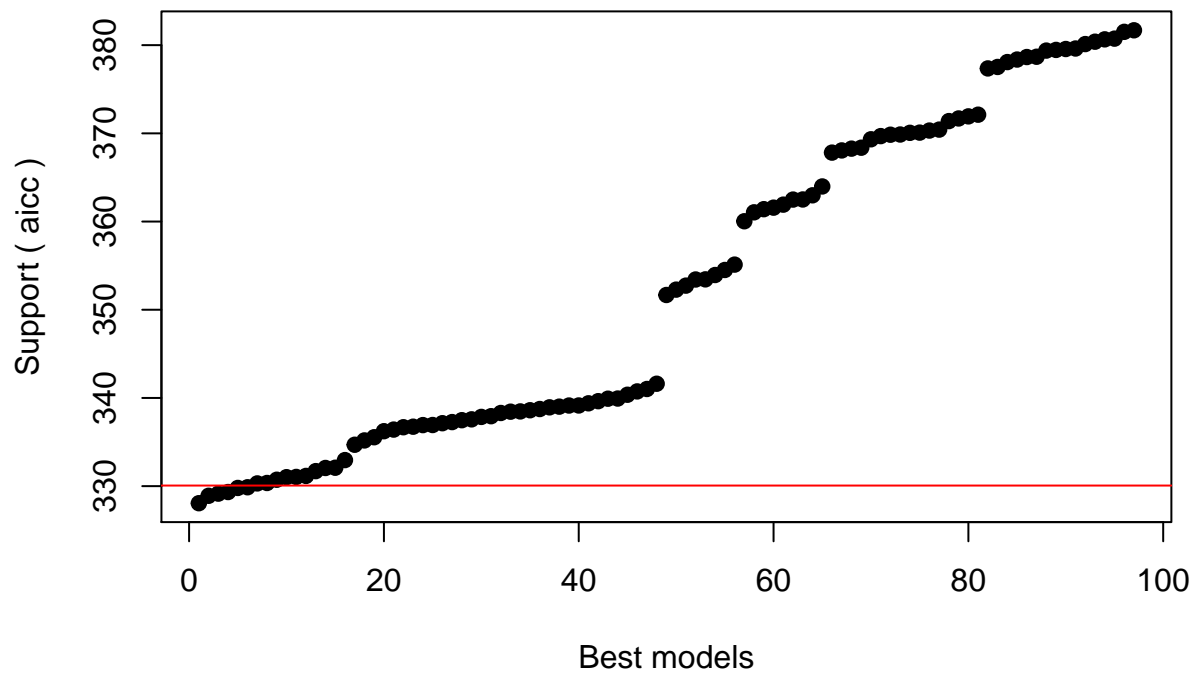
## Initialization...
## TASK: Exhaustive screening of candidate set.
## Fitting...
##
## After 50 models:
## Best model: death~1+high_blood_pressure+sc.fp+age
## Crit= 329.330641465575
## Mean crit= 352.47155567521
```

IC profile



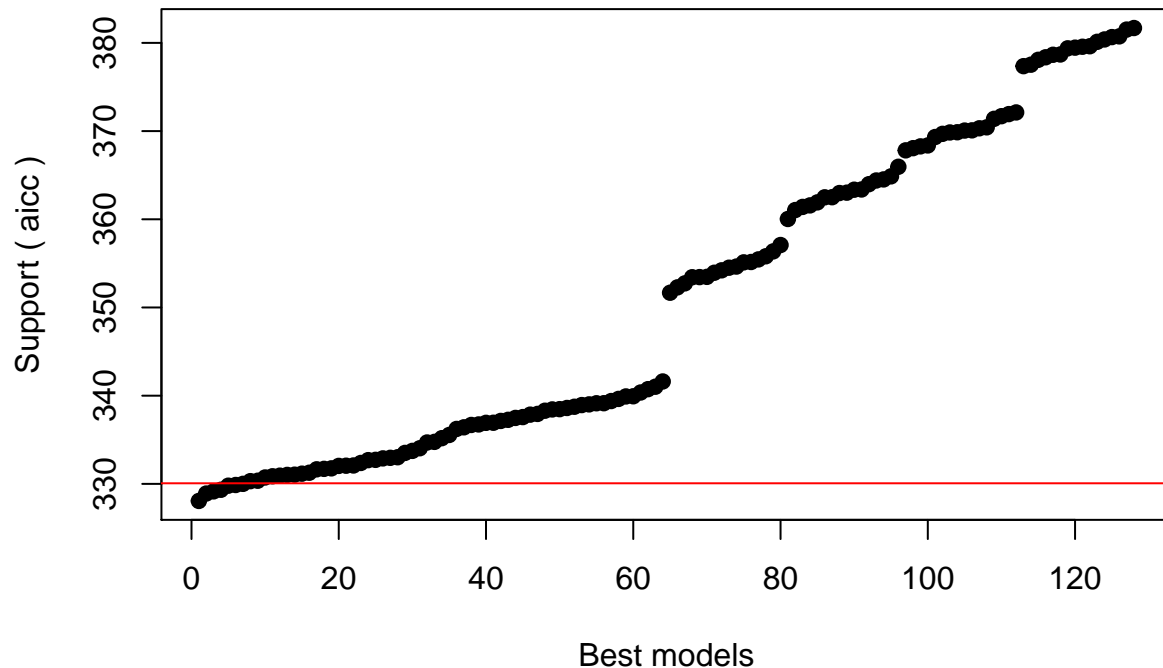
```
##
## After 100 models:
## Best model: death~1+high_blood_pressure+sc.fp+serum_sodium+age
## Crit= 328.065696808995
## Mean crit= 352.438739590334
```

IC profile



```
##
## After 150 models:
## Best model: death~1+high_blood_pressure+sc.fp+serum_sodium+age
## Crit= 328.065696808995
## Mean crit= 350.753830320583
```

IC profile



```
## Completed.
```

```
weightable(best_subset) %>%
  head()
```

```
##
## 1          death ~ 1 + high_blood_pressure + sc.fp + serum_sodium + age      model
## 2  death ~ 1 + anaemia + high_blood_pressure + sc.fp + serum_sodium + age
## 3          death ~ 1 + sc.fp + serum_sodium + age
## 4          death ~ 1 + high_blood_pressure + sc.fp + age
## 5          death ~ 1 + anaemia + sc.fp + serum_sodium + age
## 6 death ~ 1 + high_blood_pressure + platelets + sc.fp + serum_sodium + age
##      aicc      weights
## 1 328.0657 0.12190076
## 2 328.9038 0.08016977
## 3 329.1501 0.07087948
## 4 329.3306 0.06476303
## 5 329.7931 0.05139272
## 6 329.8717 0.04941342
```

```
# Based on output of 6 best subsets, the first one will be used in final model
# along with interactions indicated from analyzing potential effect modifiers
final.m <-
```

```
  glm(death ~ 1 + age + anaemia + creatinine_phosphokinase + ef.fp +
      high_blood_pressure + sc.fp + sex + age:creatinine_phosphokinase +
```



```

sex:ef.fp, data = heartfailure, family = binomial)
summary(final.m)

##
## Call:
## glm(formula = death ~ 1 + age + anaemia + creatinine_phosphokinase +
##      ef.fp + high_blood_pressure + sc.fp + sex + age:creatinine_phosphokinase +
##      sex:ef.fp, family = binomial, data = heartfailure)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0053  -0.7340  -0.3983   0.7656   2.9499
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -6.312e+00  1.995e+00  -3.164 0.001559 **
## age              6.032e-02  1.565e-02   3.855 0.000116 ***
## anaemia         4.733e-01  3.099e-01   1.527 0.126698
## creatinine_phosphokinase 1.029e-03  6.275e-04   1.640 0.100967
## ef.fp           2.092e+01  8.898e+00   2.351 0.018737 *
## high_blood_pressure  5.286e-01  3.230e-01   1.637 0.101682
## sc.fp           -2.559e+00  5.353e-01  -4.780 1.76e-06 ***
## sex             -3.408e+00  1.983e+00  -1.719 0.085598 .
## age:creatinine_phosphokinase -1.153e-05  1.003e-05  -1.149 0.250393
## ef.fp:sex        1.762e+01  1.149e+01   1.534 0.125040
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 375.35  on 298  degrees of freedom
## Residual deviance: 277.40  on 289  degrees of freedom
## AIC: 297.4
##
## Number of Fisher Scoring iterations: 5
# Goodness of Fit, Fit Statistics, and Pseudo-R2
ResourceSelection::hoslem.test(final.m$y, fitted(final.m), g=20)

##
## Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  final.m$y, fitted(final.m)
## X-squared = 11.804, df = 18, p-value = 0.8572
ResourceSelection::hoslem.test(final.m$y, fitted(final.m), g=20) %>%
  {cbind(
    .$observed,
    .$expected
  )}

##           y0 y1      yhat0      yhat1
## [0.00594,0.0325] 14  1 14.749429 0.2505709
## (0.0325,0.0463] 14  1 14.399831 0.6001693
## (0.0463,0.0568] 14  1 14.227118 0.7728816

```

```
## (0.0568,0.0789] 14 1 13.987733 1.0122670
## (0.0789,0.0992] 14 1 13.667628 1.3323717
## (0.0992,0.12] 15 0 13.342286 1.6577140
## (0.12,0.157] 13 2 12.874042 2.1259584
## (0.157,0.182] 13 2 12.399200 2.6007996
## (0.182,0.213] 13 2 12.036177 2.9638234
## (0.213,0.255] 9 6 11.434302 3.5656978
## (0.255,0.276] 12 2 10.306556 3.6934437
## (0.276,0.313] 11 4 10.615802 4.3841980
## (0.313,0.374] 10 5 9.866116 5.1338835
## (0.374,0.445] 7 8 8.852567 6.1474329
## (0.445,0.511] 7 8 7.834036 7.1659642
## (0.511,0.57] 9 6 7.014978 7.9850217
## (0.57,0.652] 4 11 5.708060 9.2919396
## (0.652,0.719] 4 11 4.639655 10.3603450
## (0.719,0.808] 4 11 3.437842 11.5621581
## (0.808,0.985] 2 13 1.606641 13.3933595
```

```
DescTools::PseudoR2(final.m)
```

```
## McFadden
## 0.2609575
```

```
# Assumptions, Diagnostics, and Influential Observations
```

```
mfp(death ~ 1 + fp(age) + fp(anaemia) + fp(creatinine_phosphokinase) +
    fp(ef.fp) + fp(high_blood_pressure) + fp(sc.fp) + fp(sex),
    data = heartfailure, family = binomial)
```

```
## Call:
```

```
## mfp(formula = death ~ 1 + fp(age) + fp(anaemia) + fp(creatinine_phosphokinase) +
##      fp(ef.fp) + fp(high_blood_pressure) + fp(sc.fp) + fp(sex),
##      data = heartfailure, family = binomial)
```

```
##
```

```
##
```

```
## Deviance table:
```

```
##          Resid. Dev
## Null model    375.3488
## Linear model  281.1596
## Final model   281.1596
##
```

```
## Fractional polynomials:
```

	df.initial	select	alpha	df.final	power1	power2
## ef.fp	4	1	0.05	1	1	.
## sc.fp	4	1	0.05	1	1	.
## age	4	1	0.05	1	1	.
## creatinine_phosphokinase	4	1	0.05	1	1	.
## high_blood_pressure	4	1	0.05	1	1	.
## anaemia	4	1	0.05	1	1	.
## sex	4	1	0.05	1	1	.

```
##
```

```
##
```

```
## Transformations of covariates:
```

	formula
## age	I((age/100)^1)
## anaemia	I((anaemia+1)^1)
## creatinine_phosphokinase	I((creatinine_phosphokinase/1000)^1)

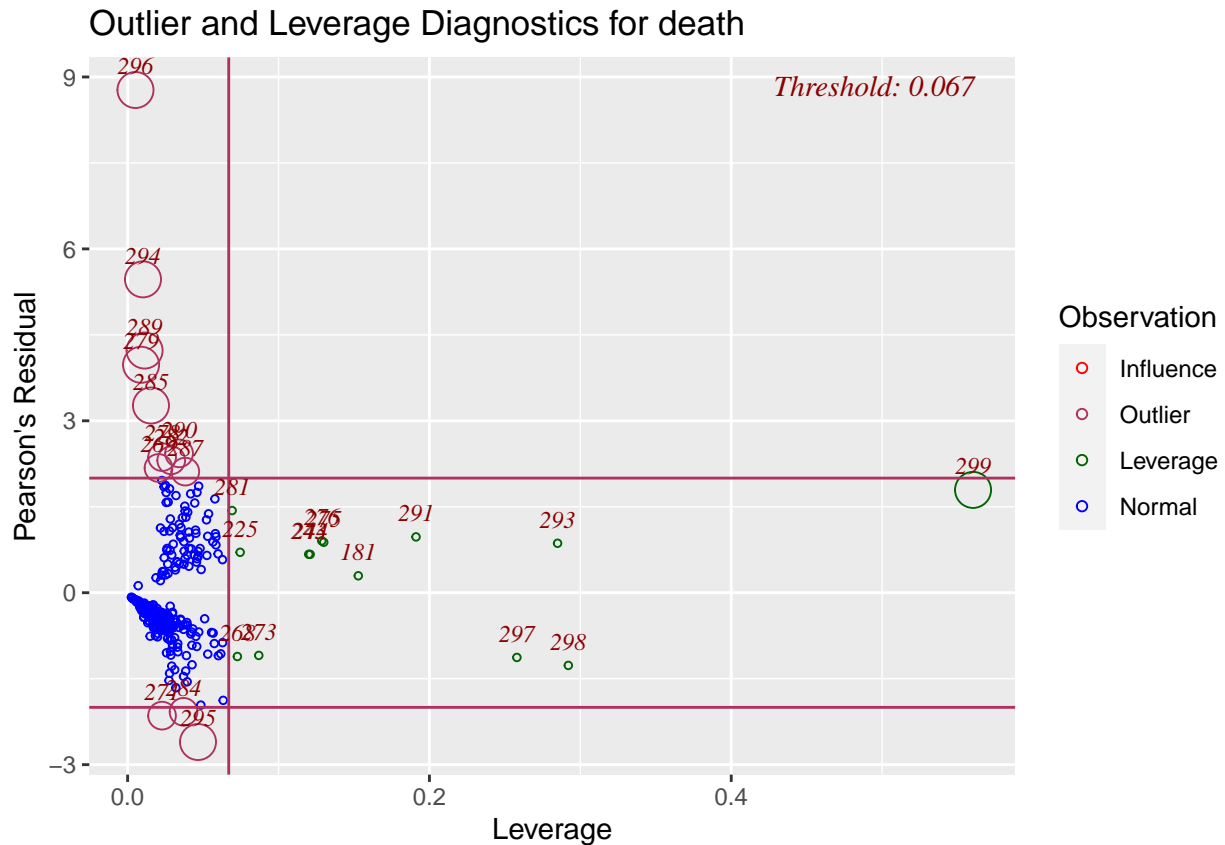
```

## ef.fp                                I((ef.fp/0.1)^1)
## high_blood_pressure                  I((high_blood_pressure+1)^1)
## sc.fp                                I(sc.fp^1)
## sex                                  I((sex+1)^1)
##
## Re-Scaling:
## Non-positive values in some of the covariates. No re-scaling was performed.
##
## Coefficients:
##             Intercept                ef.fp.1
##             -8.0402                 3.1507
##             sc.fp.1                 age.1
##             -2.5465                 5.0940
## creatinine_phosphokinase.1    high_blood_pressure.1
##             0.3407                 0.4994
##             anaemia.1             sex.1
##             0.4529                 -0.4008
##
## Degrees of Freedom: 298 Total (i.e. Null); 291 Residual
## Null Deviance: 375.3
## Residual Deviance: 281.2    AIC: 297.2
DescTools::VIF(final.m)

##             age                anaemia
##             1.486252            1.083453
## creatinine_phosphokinase        ef.fp
##             19.407158            2.602849
## high_blood_pressure             sc.fp
##             1.087496            1.064987
##             sex age:creatinine_phosphokinase
##             42.058269            19.149740
##             ef.fp:sex
##             46.272660
plot_resid_lev_logistic(final.m)

## Warning: Removed 272 rows containing missing values (geom_text).

```



Tables

```
# Table 1 - summary statistics for variables
options(scipen = 999)

sumTable <- heartfailure %>%
  select(age, anaemia, creatinine_phosphokinase, diabetes, ejection_fraction,
         high_blood_pressure, platelets, serum_creatinine, serum_sodium, sex,
         smoking, death) %>%
  psych::describe(quant = c(.1,.25,.5,.75,.90), skew = FALSE, trim = 0)

sumTable %>%
  kbl(digits = 2) %>%
  kable_classic_2(full_width = F)
```

```
# Table 2.1 - univariate analysis table
stargazer(age.m, anaemia.m, cp.m, diabetes.m, ef.m2, hbp.m,
          ci=TRUE, ci.level=0.95,
          type = "text", align=TRUE,
          title = "Univariate Regression Results for Outcome of Death",
          omit.stat=c("LL","f"), p.auto = TRUE)
```

```
##
```

	vars	n	mean	sd	min	max	range	se	Q0.1	
age	1	299	60.83	11.89	40.0	95.0	55.0	0.69	45.0	
anaemia	2	299	0.43	0.50	0.0	1.0	1.0	0.03	0.0	
creatinine_phosphokinase	3	299	581.84	970.29	23.0	7861.0	7838.0	56.11	67.6	
diabetes	4	299	0.42	0.49	0.0	1.0	1.0	0.03	0.0	
ejection_fraction	5	299	38.08	11.83	14.0	80.0	66.0	0.68	25.0	
high_blood_pressure	6	299	0.35	0.48	0.0	1.0	1.0	0.03	0.0	
platelets	7	299	263358.03	97804.24	25100.0	850000.0	824900.0	5656.17	153000.0	2
serum_creatinine	8	299	1.39	1.03	0.5	9.4	8.9	0.06	0.8	
serum_sodium	9	299	136.63	4.41	113.0	148.0	35.0	0.26	132.0	
sex	10	299	0.65	0.48	0.0	1.0	1.0	0.03	0.0	
smoking	11	299	0.32	0.47	0.0	1.0	1.0	0.03	0.0	
death	12	299	0.32	0.47	0.0	1.0	1.0	0.03	0.0	

Univariate Regression Results for Outcome of Death

```
## =====
##                                     Dependent variable:
##                                     -----
##                                     death
##                                     (1)          (2)          (3)          (4)
## -----
## age                                0.047***
##                                (0.025, 0.069)
##
## anaemia                          0.285
##                                (-0.203, 0.774)
##
## creatinine_phosphokinase         0.0001
##                                (-0.0001, 0.0004)
##
## diabetes                         -0.008
##                                (-0.501, 0.484)
##
## ef.fp                            27.9
##                                (17.87)
##
## high_blood_pressure
##
## Constant                        -3.654***      -0.875***      -0.827***      -0.745***      -5.1
##                                (-5.039, -2.269) (-1.205, -0.546) (-1.110, -0.543) (-1.063, -0.427) (-7.28)
##
## -----
## Observations                    299            299            299            299            2
## Akaike Inf. Crit.              359.993        378.040        378.231        379.348        34
## =====
```

Note:

Table 2.2 - univariate analysis table

```
stargazer(platelets.m, sc.m2,ss.m, sex.m, smoke.m,
```

```
  ci=TRUE, ci.level=0.95,
```

```
  type = "text", align=TRUE,
```

```
  title = "Univariate Regression Results for Outcome of Death Pt II",
```

```
omit.stat=c("LL","f"), p.auto = TRUE)
```

```
##
## Univariate Regression Results for Outcome of Death Pt II
## =====
##                               Dependent variable:
##                               -----
##                               death
##                               (3)
##                               (1)      (2)      (4)      (5)
## -----
## platelets      -0.00000
##                (-0.00000, 0.00000)
##
## sc.fp          -2.820***
##                (-3.763, -1.877)
##
## serum_sodium   -0.096***
##                (-0.155, -0.038)
##
## sex            -0.019
##                (-0.527, 0.489)
##
## smoking        -0.058
##                (-0.580, 0.464)
##
## Constant       -0.457      1.642***      12.394***      -0.736***      -0.730***
##                (-1.169, 0.255)  (0.829, 2.455)  (4.412, 20.377)  (-1.145, -0.328)  (-1.024, -0.436)
##
## -----
## Observations      299      299      299      299      299
## Akaike Inf. Crit. 378.611    337.436    368.022    379.343    379.301
## =====
## Note:                                                     *p<0.1; **p<0.05; ***p<0.001
```

```
# Table 3 - final model table
stargazer(final.m,
  ci=TRUE, ci.level=0.95,
  type = "text", align=TRUE,
  title = "Final Model",
  omit.stat=c("LL","f"), p.auto = TRUE)
```

```
##
## Final Model
## =====
##                               Dependent variable:
##                               -----
##                               death
##                               -----
## age              0.060***
##                  (0.030, 0.091)
##
## anaemia          0.473
##                  (-0.134, 1.081)
##
```

```

## creatinine_phosphokinase          0.001
##                                (-0.0002, 0.002)
##
## ef.fp                             20.917**
##                                (3.477, 38.357)
##
## high_blood_pressure                0.529
##                                (-0.104, 1.162)
##
## sc.fp                             -2.559***
##                                (-3.608, -1.509)
##
## sex                               -3.408*
##                                (-7.294, 0.478)
##
## age:creatinine_phosphokinase      -0.00001
##                                (-0.00003, 0.00001)
##
## ef.fp:sex                         17.621
##                                (-4.894, 40.136)
##
## Constant                         -6.312***
##                                (-10.222, -2.401)
##
## -----
## Observations                      299
## Akaike Inf. Crit.                 297.399
## =====
## Note:                            *p<0.1; **p<0.05; ***p<0.01

```