Exercise – Low birth weight

- Download the data set from Moodle:

Variable	Description	Data type	Level
ID	Identification code	Integer	
LOW	Low birth weight (Outcome)	Binary	0: Birth weight >= 2500g 1: Birth weight < 2500g
AGE	Age of the mother in years	Continuous	
LWT	Weight in pounds at the last menstrual period	Continuous	
RACE	Race	Categorical	1: White, 2: Black, 3: Other
SMOKE	Smoking status during pregnancy	Categorical	1: Yes, 0: No
PTL	History of premature labor	Integer	
HT	History of hypertension	Categorical	1: Yes, 0: No
UI	Presence of uterine irritability	Categorical	1: Yes, 0: No
FTV	Number of physician visits during the 1st trimester	Integer	
BWT	Birth weight in grams (Outcome)	Continuous	

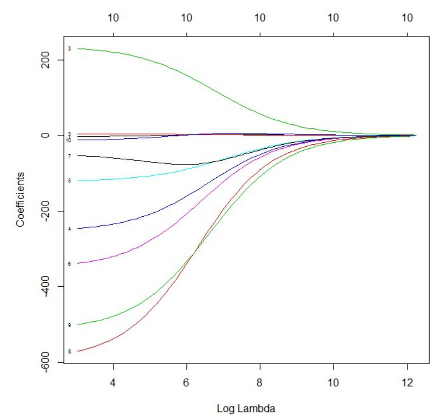
- Source: Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X. (2013) Applied Logistic Regression: Third Edition.

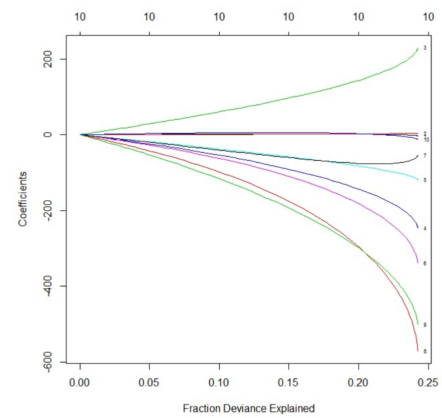
Prepare data sets

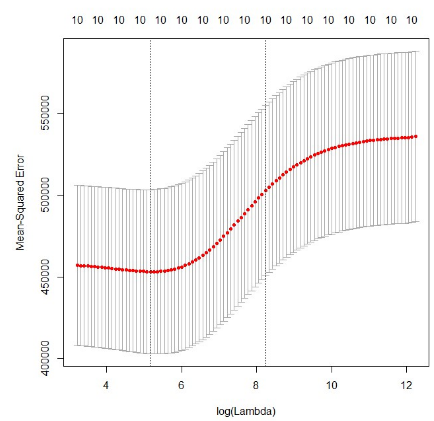
```
LowBWT <- read.csv("{Path}/lowbwt.csv")
LowBWT$LOW <- factor(LowBWT$LOW)
LowBWT$RACE <- factor(LowBWT$RACE)
LowBWT$SMOKE <- factor(LowBWT$SMOKE)
LowBWT$HT <- factor(LowBWT$HT)
LowBWT$UI <- factor(LowBWT$UI)
library(glmnet)

# Fitting continuous outcome, variable "BWT"
x.con = model.matrix(BWT ~ . - 1 - ID - LOW, data = LowBWT)
y.con = LowBWT$BWT

# Fitting binary outcome, variable "LOW"
x.bin = model.matrix(LOW ~ . - 1 - ID - BWT, data = LowBWT)
y.bin = LowBWT$LOW
```

Continuous outcome, Ridge regression

```
Ridge1 <- glmnet(x.con, y.con, alpha = 0)
plot(Ridge1, xvar="lambda", label=TRUE)
plot(Ridge1, xvar="dev", label=TRUE)

set.seed(56789)
Ridge1.cv = cv.glmnet(x.con, y.con, alpha = 0)
plot(Ridge1.cv)
coef(Ridge1, s = Ridge1.cv$lambda.1se)

                       1
(Intercept) 2847.7067859
AGE            1.6026551
LWT            0.7244883
RACE1         50.8523714
RACE2        -44.7463233
RACE3        -32.3425865
SMOKE1       -51.2953481
PTL          -34.7321317
HT1          -80.3629268
UI1          -96.8631099
FTV            4.7682106
```



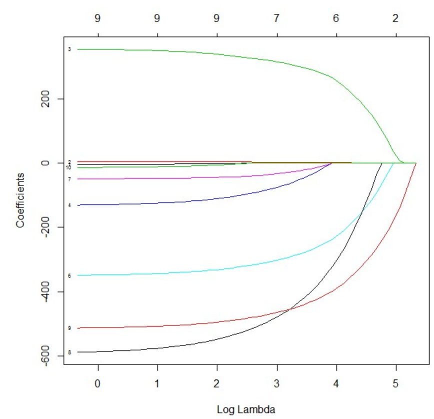


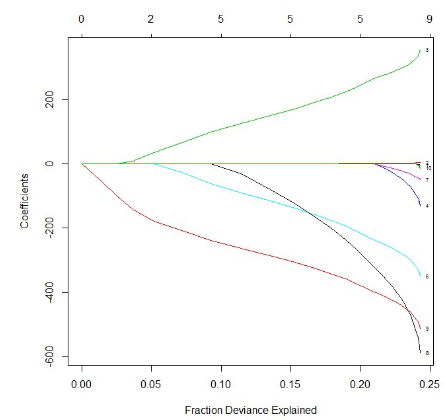


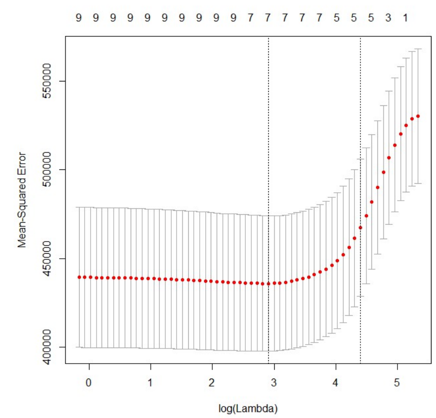
Continuous outcome, LASSO

```
LASSO1 <- glmnet(x.con, y.con, alpha = 1)
plot(LASSO1, xvar="lambda", label=TRUE)
plot(LASSO1, xvar="dev", label=TRUE)

set.seed(65432)
LASSO1.cv = cv.glmnet(x.con, y.con, alpha = 1)
plot(LASSO1.cv)
coef(LASSO1, s = LASSO1.cv$lambda.1se)
```

```
                       1
(Intercept) 2919.9406351
AGE            .
LWT            0.2770576
RACE1        126.8831125
RACE2          .
RACE3          .
SMOKE1       -89.3331247
PTL            .
HT1          -29.7483233
UI1         -262.1262524
FTV            .
```

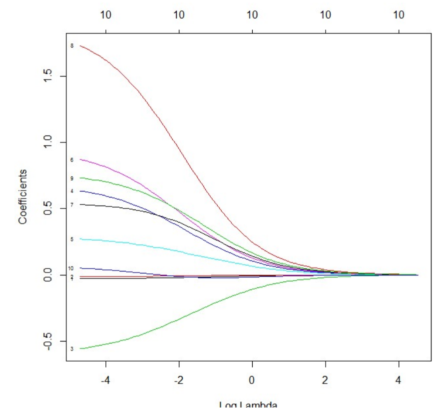


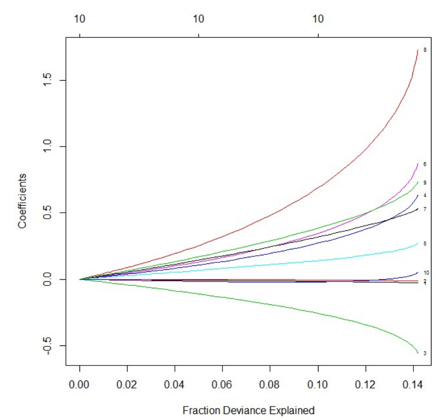


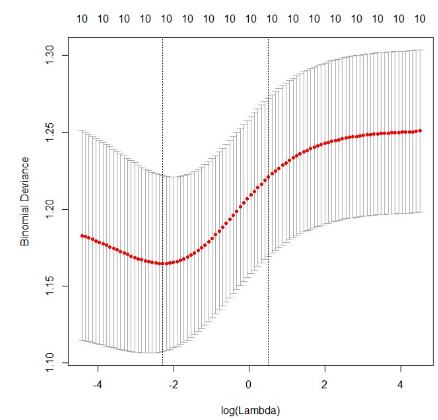


Binary outcome, Ridge regression

```
Ridge2 <- glmnet(x.bin, y.bin, alpha = 0, family = "binomial")
plot(Ridge2, xvar="lambda", label=TRUE)
plot(Ridge2, xvar="dev", label=TRUE)

set.seed(56789)
Ridge2.cv = cv.glmnet(x.bin, y.bin, alpha = 0, family = "binomial")
plot(Ridge2.cv)
coef(Ridge2, s = Ridge2.cv$lambda.1se)

                       1
(Intercept) -0.560026873
AGE         -0.005036865
LWT         -0.001320565
RACE1       -0.072877588
RACE2        0.067682690
RACE3        0.044508306
SMOKE1       0.082406149
PTL          0.093369022
HT1          0.158414315
UI1          0.110805589
FTV         -0.012006530
```







Binary outcome, LASSO

```
LASSO2 <- glmnet(x.bin, y.bin, alpha = 1, family = "binomial")
plot(LASSO2, xvar="lambda", label=TRUE)
plot(LASSO2, xvar="dev", label=TRUE)

set.seed(65432)
LASSO2.cv = cv.glmnet(x.bin, y.bin, alpha = 1, family = "binomial")
plot(LASSO2.cv)
coef(LASSO2, s = LASSO2.cv$lambda.1se)
```

```
                        1
(Intercept) -0.402500274
AGE          .
LWT         -0.003655509
RACE1       -0.248026319
RACE2        .
RACE3        .
SMOKE1       0.229056074
PTL          0.245427433
HT1          0.465913634
UI1          0.214356081
FTV          .
```