STA302A2

## ISHAAN NAGI

## 1002452525

SETUP

setwd("~/Google Drive/Uni/Winter 2018/STA302/A2")  
dat <- read.table("Census.txt", sep = "" ,header=T)  
LIFE <- dat$LIFE  
MALE <- dat$MALE  
BIRTH <- dat$BIRTH  
DIVO <- dat$DIVO  
BEDS <- dat$BEDS  
EDUC <- dat$EDUC  
INCO <- dat$INCO  
y <- LIFE

x = as.matrix(dat)  
x <-x[1:51, 2:7]  
x<- cbind(1, x)  
mode(x) = 'numeric'  
n <- 51  
p <-6  
X<-x  
Xt <- t(X)  
XtX <- Xt %\*% X  
XtXinv <-solve(XtX)  
Xty <- Xt %\*% y  
bhat <- XtXinv %\*% Xty

1a)

y\_hat <- X %\*% bhat  
y\_hat

## [,1]  
## [1,] 71.32712  
## [2,] 69.44152  
## [3,] 70.61014  
## [4,] 70.29776  
## [5,] 71.65898  
## [6,] 71.97295  
## [7,] 72.28677  
## [8,] 66.87019  
## [9,] 70.41924  
## [10,] 71.38047  
## [11,] 69.42376  
## [12,] 71.53120  
## [13,] 71.42614  
## [14,] 71.04053  
## [15,] 70.16589  
## [16,] 70.29139  
## [17,] 71.64995  
## [18,] 70.78642  
## [19,] 70.31884  
## [20,] 70.92242  
## [21,] 72.03916  
## [22,] 70.25335  
## [23,] 70.24288  
## [24,] 71.16816  
## [25,] 70.77073  
## [26,] 68.72947  
## [27,] 72.04422  
## [28,] 70.81770  
## [29,] 71.67050  
## [30,] 71.22929  
## [31,] 71.04496  
## [32,] 72.13043  
## [33,] 70.80619  
## [34,] 68.76111  
## [35,] 70.46957  
## [36,] 70.31486  
## [37,] 71.13894  
## [38,] 72.51628  
## [39,] 70.91515  
## [40,] 71.43818  
## [41,] 70.51633  
## [42,] 71.11649  
## [43,] 70.03452  
## [44,] 70.18014  
## [45,] 69.57848  
## [46,] 71.56217  
## [47,] 70.51127  
## [48,] 72.50224  
## [49,] 71.48537  
## [50,] 69.92352  
## [51,] 70.45665

e\_hat = y - y\_hat  
e\_hat

## [,1]  
## [1,] -2.01712273  
## [2,] -0.39152270  
## [3,] 0.04985782  
## [4,] 0.25224335  
## [5,] 0.05102320  
## [6,] 0.08704553  
## [7,] 0.19323411  
## [8,] -1.16019333  
## [9,] -0.35923989  
## [10,] -0.72046876  
## [11,] -0.88376476  
## [12,] 2.06880340  
## [13,] 1.13386172  
## [14,] 0.82946681  
## [15,] -0.02589499  
## [16,] 0.58861254  
## [17,] 0.93004669  
## [18,] -0.68642174  
## [19,] -1.55884433  
## [20,] 0.90758262  
## [21,] -1.81915643  
## [22,] 0.67665009  
## [23,] 0.38711962  
## [24,] 1.79184021  
## [25,] -0.08073436  
## [26,] -0.63947073  
## [27,] -1.48421996  
## [28,] -1.60770155  
## [29,] 1.11949999  
## [30,] 1.37071164  
## [31,] 0.18503566  
## [32,] -1.20043418  
## [33,] -0.48618775  
## [34,] 0.26888640  
## [35,] 0.08043332  
## [36,] 0.50514224  
## [37,] 0.28105954  
## [38,] -0.38627710  
## [39,] -0.48514830  
## [40,] 0.46181996  
## [41,] -2.55632758  
## [42,] 0.96350641  
## [43,] 0.07548104  
## [44,] 0.71986346  
## [45,] 3.32151698  
## [46,] -1.48217033  
## [47,] 1.12872551  
## [48,] -0.78223737  
## [49,] 0.99463448  
## [50,] -0.44351769  
## [51,] -0.16664779

bhat

## [,1]  
## 70.5577812705  
## MALE 0.1261018758  
## BIRTH -0.5160557876  
## DIVO -0.1965375074  
## BEDS -0.0033392036  
## EDUC 0.2368222541  
## INCO -0.0003612011

Equation: LIFE = 70.5577812705 + 0.1261018758 (MALE) - 0.5160557876 (BIRTH) - 0.1965375074 (DIVO) - 0.0033392036 (BEDS) + 0.2368222541 (EDUC) - 0.0003612011 (INCO)

1b) - MALE

bhat[2] #b\_1

## [1] 0.1261019

b\_1 (MALE) corresponds to the Expected change (+ 0.1261019) in the Averafe Lifespan with 1 unit increase in the proportion of Males to Female.

* BIRTH

bhat[3] #b\_2

## [1] -0.5160558

b\_2 (BIRTH) corresponds to the Expected change (-0.5160558) in the Averafe Lifespan with a unit increase in the birth rate per 1,000 people.

1c)

RSS <- (t(e\_hat) %\*% e\_hat)  
sigma\_sq\_hat = RSS / (n)  
sigma\_sq\_hat

## [,1]  
## [1,] 1.192215

s\_sq = RSS / (n-p-1)  
s\_sq

## [,1]  
## [1,] 1.381885

1d)

s <-c(s\_sq^(1/2))  
se\_bs = diag(s \* (XtXinv)^(1/2))  
se\_bs

## MALE BIRTH DIVO BEDS   
## 4.2897471299 0.0472317551 0.1172774621 0.0739532971 0.0009795303   
## EDUC INCO   
## 0.1110224835 0.0004597943

#Corresponding to B\_0, B\_1 ... B\_6

1e)

y\_bar = c(sum(y\_hat)/ n)  
SST <- sum((LIFE - y\_bar)^(2))  
R\_sq <- 1 - (RSS/SST)  
R\_sq

## [,1]  
## [1,] 0.4684927

Explains the proportion of variation in the Average lifespan explained by the regression, which is 46.849%.

1. 2a)

MLR <- lm(formula = LIFE~ MALE + BIRTH + DIVO + BEDS + EDUC + INCO, data=dat)   
summary(MLR)

##   
## Call:  
## lm(formula = LIFE ~ MALE + BIRTH + DIVO + BEDS + EDUC + INCO,   
## data = dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.5563 -0.6629 0.0755 0.6983 3.3215   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 70.5577813 4.2897471 16.448 < 2e-16 \*\*\*  
## MALE 0.1261019 0.0472318 2.670 0.01059 \*   
## BIRTH -0.5160558 0.1172775 -4.400 6.78e-05 \*\*\*  
## DIVO -0.1965375 0.0739533 -2.658 0.01093 \*   
## BEDS -0.0033392 0.0009795 -3.409 0.00141 \*\*   
## EDUC 0.2368223 0.1110225 2.133 0.03853 \*   
## INCO -0.0003612 0.0004598 -0.786 0.43633   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.176 on 44 degrees of freedom  
## Multiple R-squared: 0.4685, Adjusted R-squared: 0.396   
## F-statistic: 6.464 on 6 and 44 DF, p-value: 6.112e-05

2b) [any i in range 0 to 6]

SSReg = SST-RSS  
F\_val = (SSReg/(p)/(RSS/(n-p-1)))   
F\_val

## [,1]  
## [1,] 6.463905

F\_crt = qf(.95, df1=p, df2=n-p-1)  
F\_crt

## [1] 2.313264

F\_val > F\_crit, hence we fail to accept . Then atleast one of the B\_i’s [i = 1, … 6] is not Zero. Our Model is significant.

2c)

# MALE

t\_value1 =2.670  
t\_crit = qt(c(.005, .995), df=n-p-1)  
#t\_crit = [-2.692278, 2.692278]

Since |t\_crit| > |t\_value1|, we accept . Then and we can remove the predictor (MALE) from the model.

# BIRTH

t\_value2 = 4.400  
t\_crit = qt(c(.005, .995), df=n-p-1)  
#t\_crit = [-2.692278, 2.692278]

Since |t\_value2| > |t\_crit|, we fail to accept . Then then we can’t remove the predictor (BIRTH) variable from the model.

# DIVO

t\_value3 =2.658  
t\_crit = qt(c(.005, .995), df=n-p-1)  
#t\_crit = [-2.692278, 2.692278]

Since |t\_crit| > |t\_value3|, we accept . Then and we can remove the predictor (DIVO) from the model.

# BEDS

t\_value4 =3.409  
t\_crit = qt(c(.005, .995), df=n-p-1)  
#t\_crit = [-2.692278, 2.692278]

Since |t\_value4| > |t\_crit|, we fail to accept . Then then we can’t remove the predictor (BEDS) variable from the model.

# EDUC

t\_value5 = 2.133  
t\_crit = qt(c(.005, .995), df=n-p-1)  
#t\_crit = [-2.692278, 2.692278]

Since |t\_crit| > |t\_value5|, we accept . Then and we can remove the predictor (EDUC) from the model.

# INCO

t\_value6 =0.786  
t\_crit = qt(c(.005, .995), df=n-p-1)  
#t\_crit = [-2.692278, 2.692278]

Since |t\_crit| > |t\_value6|, we accept . Then and we can remove the predictor (INCO) from the model.

Yes, the results indicate that variables MALE (x1), DIVO (x3), EDUC (x5) and INCO x(6) should be removed.

2d)

MLR\_reduced <- lm(formula = LIFE~ BIRTH + BEDS, data=dat)   
summary(MLR\_reduced)

##   
## Call:  
## lm(formula = LIFE ~ BIRTH + BEDS, data = dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.5627 -0.8180 -0.0819 0.9261 3.6202   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 79.1473186 2.2717401 34.840 < 2e-16 \*\*\*  
## BIRTH -0.3281679 0.1026214 -3.198 0.00245 \*\*   
## BEDS -0.0027415 0.0009388 -2.920 0.00531 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.352 on 48 degrees of freedom  
## Multiple R-squared: 0.2329, Adjusted R-squared: 0.2009   
## F-statistic: 7.286 on 2 and 48 DF, p-value: 0.001725

LIFE = 79.1473186 - 0.3281679 (BIRTH) - 0.0027415 (BEDS)

2e) both not zero.

MLR\_red\_MALE\_INCO <- lm(formula = LIFE~ BIRTH + BEDS + DIVO + EDUC, data=dat)   
anova(MLR, MLR\_red\_MALE\_INCO)

## Analysis of Variance Table  
##   
## Model 1: LIFE ~ MALE + BIRTH + DIVO + BEDS + EDUC + INCO  
## Model 2: LIFE ~ BIRTH + BEDS + DIVO + EDUC  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 44 60.803   
## 2 46 70.654 -2 -9.8507 3.5642 0.03676 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

F\_val = 3.5642  
F\_Crit = qf(.99, df1=2, df2=n-p-1)  
#F\_Crit = 5.122628

|F\_val| > |F\_crit| we fail to reject , hence the predictors MALE and INCO can be removed from the model.

2f)

MLR\_MALE <- lm(LIFE ~ MALE, data=dat)   
anova(MLR, MLR\_MALE)

## Analysis of Variance Table  
##   
## Model 1: LIFE ~ MALE + BIRTH + DIVO + BEDS + EDUC + INCO  
## Model 2: LIFE ~ MALE  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 44 60.803   
## 2 49 109.834 -5 -49.031 7.0963 6.099e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

F\_val = 7.0963

We fail to accept , hence we infer that predictors BIRTH (x2), DIVO (x3), BEDS (x4), EDUC (x5) and INCO (x6) are significant after taking the predictor MALE (x1) into consideration.

2g) [i = 1, 2] , both not zero.

MLR\_B0 = lm(LIFE ~ 1, data=dat)   
MLR\_MALE\_BIRTH <- lm(LIFE ~ MALE + BIRTH, data=dat)   
anova(MLR\_B0, MLR\_MALE\_BIRTH)

## Analysis of Variance Table  
##   
## Model 1: LIFE ~ 1  
## Model 2: LIFE ~ MALE + BIRTH  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 50 114.397   
## 2 48 85.424 2 28.973 8.14 0.0009036 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

F\_val = 8.14

We fail to accept the , we infer that (MALE) x1 and(BIRTH) x2 are significant/useful in predicting the response.

2h)

MLR\_B0 = lm(LIFE ~ 1, data=dat)   
MLR\_B0B3 = lm(LIFE ~ DIVO)  
MLR\_B0B2B3 = lm(LIFE ~ BIRTH + DIVO)  
MLR\_MALE\_BIRTH\_DIVO <- lm(LIFE~ MALE + BIRTH + DIVO, data=dat)   
MLR\_BIRTH\_DIVO <- lm(LIFE ~ BIRTH + DIVO, data=dat)

SSRegB3B0 <- anova(MLR\_B0)["Residuals", "Sum Sq"] - anova(MLR\_B0B3)["Residuals", "Sum Sq"]  
SSRegB3B0

## [1] 3.307311

SSRegB2\_B0B3 <- anova(MLR\_B0B3)["Residuals", "Sum Sq"] - anova(MLR\_B0B2B3)["Residuals", "Sum Sq"]  
SSRegB2\_B0B3

## [1] 8.914548

RSSB0B1B2B3 <- anova(MLR\_MALE\_BIRTH\_DIVO)["Residuals", "Sum Sq"]  
  
RSS\_B0B2B3 <- anova(MLR\_BIRTH\_DIVO)["Residuals", "Sum Sq"]   
  
SSRegB1\_B0B2B3 <- RSS\_B0B2B3 - RSSB0B1B2B3  
SSRegB1\_B0B2B3

## [1] 21.42422

SSRegB1B2B3\_B0 = SSRegB3B0 + SSRegB2\_B0B3 + SSRegB1\_B0B2B3  
SSRegB1B2B3\_B0

## [1] 33.64607

2i) - 1

#MLR\_B0  
#MLR\_MALE  
MLR\_BIRTH <- lm(LIFE ~ BIRTH, data=dat)   
MLR\_DIVO <- lm(LIFE ~ DIVO, data=dat)   
MLR\_INCO <- lm(LIFE ~ INCO, data=dat)   
#MLR\_MALE\_BIRTH  
MLR\_MALE\_DIVO <- lm(LIFE~ MALE + DIVO, data=dat)   
MLR\_MALE\_INCO <- lm(LIFE~ MALE + INCO, data=dat)   
#MLR\_BIRTH\_DIVO  
MLR\_BIRTH\_INCO <- lm(LIFE ~ BIRTH + INCO, data=dat)   
MLR\_DIVO\_INCO <- lm(LIFE ~ DIVO + INCO, data=dat)   
MLR\_MALE\_BIRTH\_DIVO

##   
## Call:  
## lm(formula = LIFE ~ MALE + BIRTH + DIVO, data = dat)  
##   
## Coefficients:  
## (Intercept) MALE BIRTH DIVO   
## 62.3656 0.1689 -0.3912 -0.1272

MLR\_MALE\_BIRTH\_INCO <- lm(LIFE~ MALE + BIRTH + INCO, data=dat)   
MLR\_BIRTH\_DIVO\_INCO <- lm(LIFE ~ BIRTH + DIVO + INCO, data=dat)   
MLR\_MALE\_DIVO\_INCO <- lm(LIFE ~ MALE + DIVO + INCO, data=dat)   
MLR\_mbdi <- lm(LIFE ~ MALE + BIRTH + DIVO + INCO, data=dat)   
  
#AIC  
aic\_b0 <- AIC(MLR\_B0)   
aic\_b0

## [1] 189.9321

aic\_m <- AIC(MLR\_MALE)   
aic\_m

## [1] 189.8561

aic\_d <- AIC(MLR\_DIVO)  
aic\_d

## [1] 190.4359

aic\_i <- AIC(MLR\_INCO)  
aic\_i

## [1] 191.2477

aic\_b <- AIC(MLR\_BIRTH)  
aic\_b

## [1] 186.7525

aic\_mb <- AIC(MLR\_MALE\_BIRTH)  
aic\_mb

## [1] 179.0377

aic\_md <- AIC(MLR\_MALE\_DIVO)  
aic\_md

## [1] 188.5535

aic\_mi <- AIC(MLR\_MALE\_INCO)  
aic\_mi

## [1] 191.4429

aic\_bd <- AIC(MLR\_BIRTH\_DIVO)  
aic\_bd

## [1] 188.1698

aic\_bi <- AIC(MLR\_BIRTH\_INCO)  
aic\_bi

## [1] 188.5226

aic\_di <- AIC(MLR\_DIVO\_INCO)  
aic\_di

## [1] 191.4755

aic\_mbd <- AIC(MLR\_MALE\_BIRTH\_DIVO)  
aic\_mbd

## [1] 178.1686

aic\_mbi <- AIC(MLR\_MALE\_BIRTH\_INCO)  
aic\_mbi

## [1] 180.932

aic\_bdi <- AIC(MLR\_BIRTH\_DIVO\_INCO)  
aic\_bdi

## [1] 189.8012

aic\_mdi <- AIC(MLR\_MALE\_DIVO\_INCO)  
aic\_mdi

## [1] 189.9273

aic\_mbdi <- AIC(MLR\_mbdi)  
aic\_mbdi

## [1] 180.1406

Lowest AIC is with model LIFE ~ BIRTH + MALE + DIVO

- 2

null<- MLR\_B0  
full<- MLR\_mbdi  
forwdAIC=step(null, scope=list(lower=null, upper=full), direction="forward")

## Start: AIC=43.2  
## LIFE ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + BIRTH 1 11.0479 103.35 40.021  
## + MALE 1 4.5632 109.83 43.124  
## <none> 114.40 43.200  
## + DIVO 1 3.3073 111.09 43.704  
## + INCO 1 1.5249 112.87 44.516  
##   
## Step: AIC=40.02  
## LIFE ~ BIRTH  
##   
## Df Sum of Sq RSS AIC  
## + MALE 1 17.9252 85.424 32.306  
## <none> 103.349 40.021  
## + DIVO 1 1.1739 102.175 41.438  
## + INCO 1 0.4647 102.885 41.791  
##   
## Step: AIC=32.31  
## LIFE ~ BIRTH + MALE  
##   
## Df Sum of Sq RSS AIC  
## + DIVO 1 4.6730 80.751 31.437  
## <none> 85.424 32.306  
## + INCO 1 0.1768 85.247 34.200  
##   
## Step: AIC=31.44  
## LIFE ~ BIRTH + MALE + DIVO  
##   
## Df Sum of Sq RSS AIC  
## <none> 80.751 31.437  
## + INCO 1 0.044334 80.707 33.409

Best Model according to forwardAIC is LIFE ~ BIRTH + MALE + DIVO

- 3

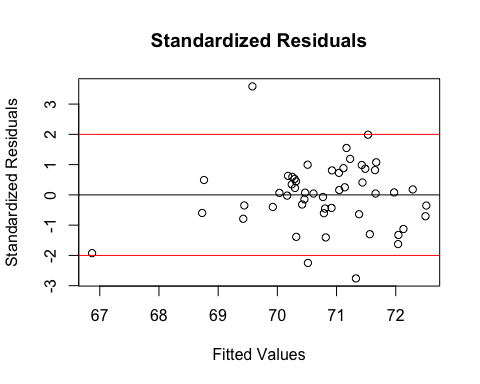
backAIC=step(full, direction="backward", data=dat)

## Start: AIC=33.41  
## LIFE ~ MALE + BIRTH + DIVO + INCO  
##   
## Df Sum of Sq RSS AIC  
## - INCO 1 0.0443 80.751 31.437  
## <none> 80.707 33.409  
## - DIVO 1 4.5405 85.247 34.200  
## - MALE 1 20.7328 101.440 43.070  
## - BIRTH 1 20.9838 101.691 43.196  
##   
## Step: AIC=31.44  
## LIFE ~ MALE + BIRTH + DIVO  
##   
## Df Sum of Sq RSS AIC  
## <none> 80.751 31.437  
## - DIVO 1 4.673 85.424 32.306  
## - MALE 1 21.424 102.175 41.438  
## - BIRTH 1 22.196 102.947 41.822

Best Model according to backwardAIC is LIFE ~ BIRTH + MALE + DIVO

3a)

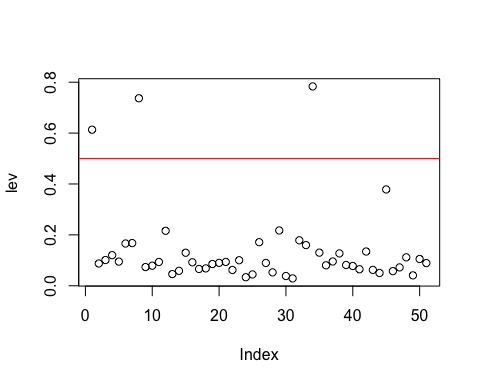
std\_res <- rstandard(MLR)  
plot(MLR$fitted.values, std\_res, ylab="Standardized Residuals", xlab="Fitted Values", main="Standardized Residuals")  
abline(0, 0)  
abline(2, 0, col=c("red"))  
abline(-2, 0, col=c("red"))



Yes, there are 3 points with residuals greater than |2| and an outlier that may be influential with the lowest fitted value (~67).

3b)

lev = hat(model.matrix(MLR))  
plot(lev)  
abline(0.5, 0, col=c("red"))



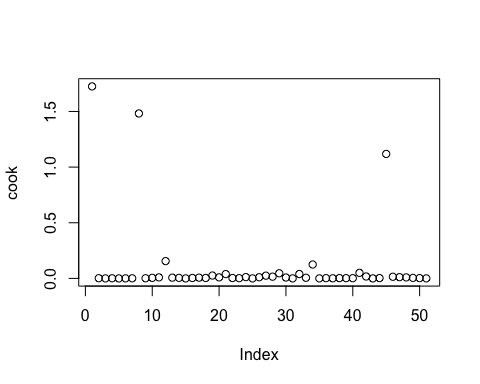
dat[lev >0.5,]

## STATE MALE BIRTH DIVO BEDS EDUC INCO LIFE  
## 1 AK 119.1 24.8 5.6 603.3 14.1 4638 69.31  
## 8 DC 86.8 20.1 3.0 1859.4 17.8 4644 65.71  
## 34 NV 102.8 19.6 18.7 560.7 10.8 4583 69.03

3 points have lev > 0.5. Indices 1, 8 and 34. Corresponding to States: AK, DC and NV.

3c)

cook = cooks.distance(MLR)  
plot(cook)



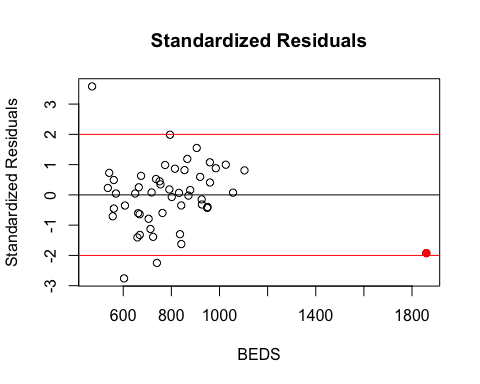
dat[cook > 1,]

## STATE MALE BIRTH DIVO BEDS EDUC INCO LIFE  
## 1 AK 119.1 24.8 5.6 603.3 14.1 4638 69.31  
## 8 DC 86.8 20.1 3.0 1859.4 17.8 4644 65.71  
## 45 UT 97.6 25.5 3.7 470.5 14.0 3169 72.90

No, not all Observations are the same. We now see data corresponding to UT instead of NV.

3d)

plot(BEDS, std\_res, ylab="Standardized Residuals",   
 xlab="BEDS", main="Standardized Residuals")  
abline(0, 0)  
abline(2, 0, col=c("red"))  
abline(-2, 0, col=c("red"))  
points(BEDS[8], std\_res[8] , col='red', pch = 19)



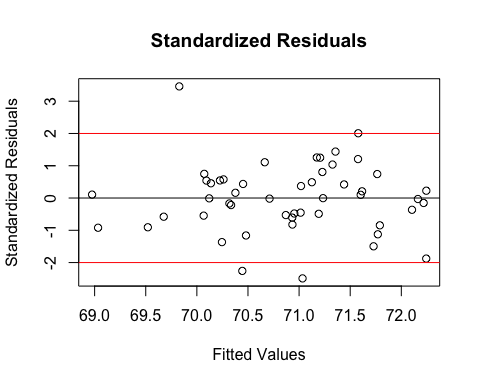
There are 3 points that are outliers based on std\_res (>2) and the outlier red point corresponding to DC (RED)

3e)

dat2 <- subset(dat, STATE != 'DC')  
LIFE2 <- dat2$LIFE  
MALE2 <- dat2$MALE  
BIRTH2 <- dat2$BIRTH  
DIVO2 <- dat2$DIVO  
BEDS2 <- dat2$BEDS  
EDUC2 <- dat2$EDUC  
INCO2 <- dat2$INCO  
MLR\_NoDC <- lm(formula = LIFE2 ~ MALE2 + BIRTH2 + DIVO2 + BEDS2 + EDUC2 + INCO2, data=dat2)  
summary(MLR\_NoDC)

##   
## Call:  
## lm(formula = LIFE2 ~ MALE2 + BIRTH2 + DIVO2 + BEDS2 + EDUC2 +   
## INCO2, data = dat2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.48448 -0.61603 -0.00768 0.58701 3.07199   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 69.8207566 4.1692564 16.747 < 2e-16 \*\*\*  
## MALE2 0.0922054 0.0487996 1.889 0.06558 .   
## BIRTH2 -0.4261398 0.1222113 -3.487 0.00114 \*\*   
## DIVO2 -0.1377833 0.0774530 -1.779 0.08232 .   
## BEDS2 -0.0011637 0.0014481 -0.804 0.42604   
## EDUC2 0.3157699 0.1145801 2.756 0.00855 \*\*   
## INCO2 -0.0004698 0.0004485 -1.048 0.30064   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.138 on 43 degrees of freedom  
## Multiple R-squared: 0.3679, Adjusted R-squared: 0.2797   
## F-statistic: 4.171 on 6 and 43 DF, p-value: 0.002171

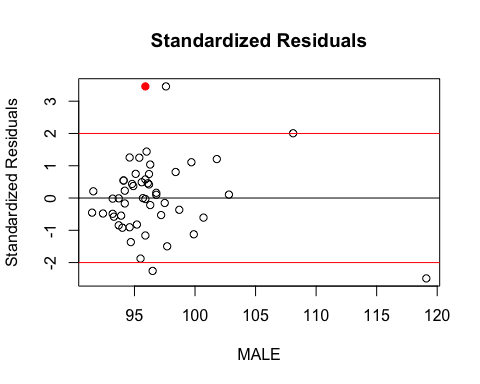
std\_res\_NoDC <- rstandard(MLR\_NoDC)  
plot(MLR\_NoDC$fitted.values, std\_res\_NoDC, ylab="Standardized Residuals", xlab="Fitted Values", main="Standardized Residuals")  
abline(0, 0)  
abline(2, 0, col=c("red"))  
abline(-2, 0, col=c("red"))



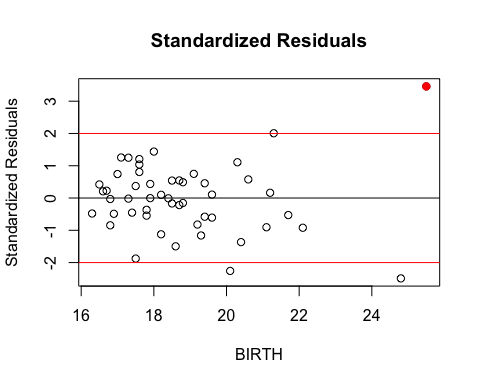
Yes there is a noticeable change in the Standardised residuals. The fitted line has changed. NEW: - LIFE = 69.8207566 + 0.0922054 (MALE2) - 0.4261398 (BIRTH2) - 0.1377833 (DIVO2) - 0.0011637 (BEDS2) + 0.3157699 (EDUC2) - 0.0004698 (INCO2) OLD: - LIFE = 70.5577813 + 0.1261019 (MALE2) - 0.5160558 (BIRTH2) - 0.1965375 (DIVO2) - 0.0033392 (BEDS2) + 0.2368223 (EDUC2) - 0.0003612 (INCO2)

Also, s changed: New: Multiple R-squared: 0.3679 Old: Multiple R-squared: 0.4685

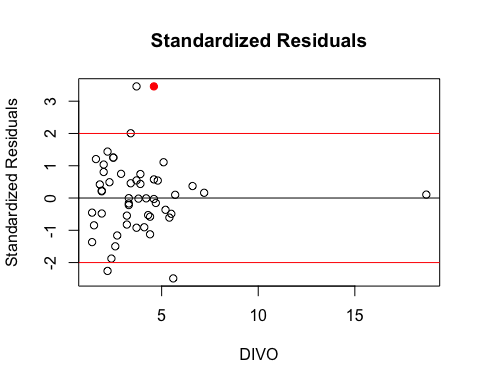
#MALE  
plot(MALE2, std\_res\_NoDC, ylab="Standardized Residuals", xlab="MALE", main="Standardized Residuals")  
abline(0, 0)  
abline(2, 0, col=c("red"))  
abline(-2, 0, col=c("red"))  
points(MALE[44], std\_res\_NoDC[44] , col='red', pch = 19)



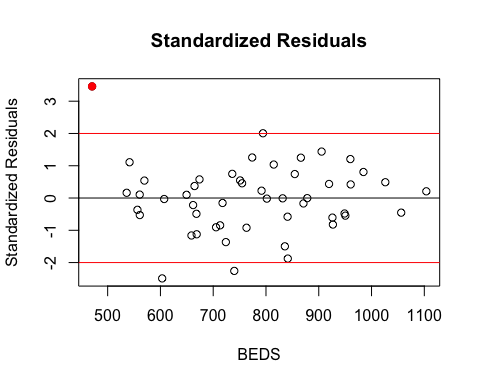
#BEDS  
plot(BIRTH2, std\_res\_NoDC, ylab="Standardized Residuals", xlab="BIRTH", main="Standardized Residuals")  
abline(0, 0)  
abline(2, 0, col=c("red"))  
abline(-2, 0, col=c("red"))  
points(BIRTH2[44], std\_res\_NoDC[44] , col='red', pch = 19)



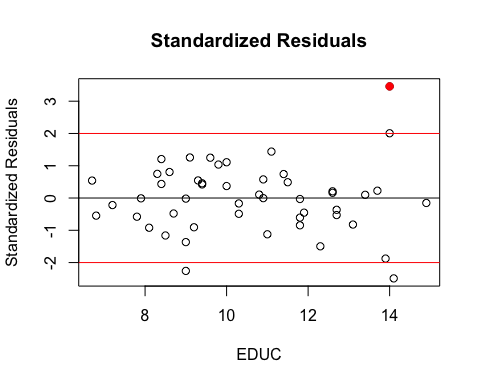
#DIVO  
plot(DIVO2, std\_res\_NoDC, ylab="Standardized Residuals", xlab="DIVO", main="Standardized Residuals")  
abline(0, 0)  
abline(2, 0, col=c("red"))  
abline(-2, 0, col=c("red"))  
points(DIVO[44], std\_res\_NoDC[44] , col='red', pch = 19)



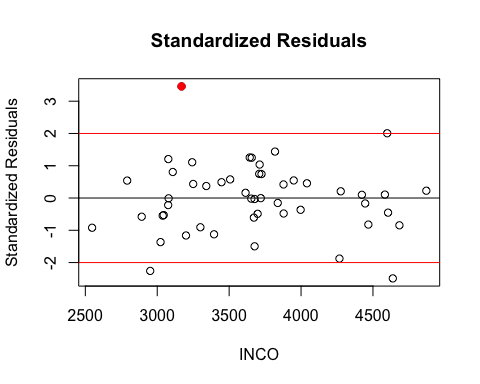
#BEDS  
plot(BEDS2, std\_res\_NoDC, ylab="Standardized Residuals", xlab="BEDS", main="Standardized Residuals")  
abline(0, 0)  
abline(2, 0, col=c("red"))  
abline(-2, 0, col=c("red"))  
points(BEDS2[44], std\_res\_NoDC[44] , col='red', pch = 19)



#EDUC  
plot(EDUC2, std\_res\_NoDC, ylab="Standardized Residuals", xlab="EDUC", main="Standardized Residuals")  
abline(0, 0)  
abline(2, 0, col=c("red"))  
abline(-2, 0, col=c("red"))  
points(EDUC2[44], std\_res\_NoDC[44] , col='red', pch = 19)



#INCO  
plot(INCO2, std\_res\_NoDC, ylab="Standardized Residuals", xlab="INCO", main="Standardized Residuals")  
abline(0, 0)  
abline(2, 0, col=c("red"))  
abline(-2, 0, col=c("red"))  
points(INCO2[44], std\_res\_NoDC[44] , col='red', pch = 19)

 ALL Data corresponding to this STATE has very standardized residual values. Difference between observations and expected values is large.

3g)

dat3 <- subset(dat2, STATE != 'UT')  
LIFE3 <- dat3$LIFE  
MALE3 <- dat3$MALE  
BIRTH3 <- dat3$BIRTH  
DIVO3 <- dat3$DIVO  
BEDS3 <- dat3$BEDS  
EDUC3 <- dat3$EDUC  
INCO3 <- dat3$INCO  
MLR\_NoDC\_NoUT <- lm(formula = LIFE3 ~ MALE3+ BIRTH3 + DIVO3 + BEDS3 + EDUC3 + INCO3, data=dat3)

is higher without UT data. Therefore this model explains the proportion of variability explained by the regression, better. Predictor have coefficients chnaged:

MLR\_NoDC\_NoUT$coefficients

## (Intercept) MALE3 BIRTH3 DIVO3 BEDS3   
## 68.2344178920 0.1494306784 -0.6334939159 -0.1116666651 -0.0007407427   
## EDUC3 INCO3   
## 0.2163469795 -0.0003390279

Vs

MLR$coefficients

## (Intercept) MALE BIRTH DIVO BEDS   
## 70.5577812704 0.1261018758 -0.5160557876 -0.1965375074 -0.0033392036   
## EDUC INCO   
## 0.2368222541 -0.0003612011