IDS 702 - Homework #3

Ana Belen Barcenas J.

9/24/2018

Maternal Smoking and Birth Weights

Summary of the data available

After looking carefully to each variable and taking into account the intuition of the problem, I believe that the father's height and weight could be strongly related with the baby's weight. Thus, I'll create a data frame including these variables to compare models with father's information and without it. The database with father's data has 508 observations vs. 869 obs. in the database without father's data (dataset cleaned provided by the professor).

Moreover, I will take the mother's education and income variables as continuos since both of them take values that has a specific order, they're ordinal variables. I will interpret the coefficients accordingly to the values that each predictor takes.

Cleaning data:

1) In both datasets there are not rows with missig values.
2) Mrace is equal to 10 in some cases. Those rows will be deleted since that variable should take values between 0 and 9 and 99 for unknown cases. Unfortunately, we can not talk with the people who collect the data to understand these values.

```
smoking_dad <- na.omit(smoking_comp[c(-11, -12, -13, -16, -19,-20,-21)])
smoking_dad2 <- smoking_dad[which(smoking_dad$mrace<=9), ]

summary(smoking_dad2)
```

```
##        id             date        gestation         bwt.oz
##  Min.   :  15   Min.   :1350   Min.   :148.0   Min.   : 55.0
##  1st Qu.:5620   1st Qu.:1467   1st Qu.:273.0   1st Qu.:108.0
##  Median :6928   Median :1567   Median :280.0   Median :119.0
##  Mean   :6137   Mean   :1557   Mean   :278.9   Mean   :118.5
##  3rd Qu.:7802   3rd Qu.:1651   3rd Qu.:287.0   3rd Qu.:129.0
##  Max.   :9263   Max.   :1714   Max.   :338.0   Max.   :174.0
##      parity           mrace            mage            med
##  Min.   : 0.000   Min.   :0.000   Min.   :15.00   Min.   :0.000
##  1st Qu.: 1.000   1st Qu.:0.000   1st Qu.:23.00   1st Qu.:2.000
##  Median : 2.000   Median :3.000   Median :27.00   Median :2.000
##  Mean   : 2.045   Mean   :3.189   Mean   :27.55   Mean   :2.902
##  3rd Qu.: 3.000   3rd Qu.:7.000   3rd Qu.:31.00   3rd Qu.:4.000
##  Max.   :11.000   Max.   :9.000   Max.   :43.00   Max.   :5.000
##       mht           mpregwt            dht             dwt
##  Min.   :54.00   Min.   : 87.0   Min.   :60.00   Min.   :110.0
##  1st Qu.:62.00   1st Qu.:115.0   1st Qu.:68.00   1st Qu.:155.0
##  Median :64.00   Median :125.0   Median :71.00   Median :170.0
##  Mean   :64.05   Mean   :128.7   Mean   :70.27   Mean   :170.5
##  3rd Qu.:66.00   3rd Qu.:140.0   3rd Qu.:72.00   3rd Qu.:185.0
```

```
##  Max.   :72.00   Max.   :220.0   Max.   :78.00   Max.   :260.0
##       inc             smoke
##  Min.   :0.000   Min.   :0.0000
##  1st Qu.:2.000   1st Qu.:0.0000
##  Median :3.000   Median :0.0000
##  Mean   :3.801   Mean   :0.4311
##  3rd Qu.:6.000   3rd Qu.:1.0000
##  Max.   :9.000   Max.   :1.0000
```

3) I will collapse race categories from 0 to 5 in "white" and create dummy vars for each category:

```
n = nrow(smoking)
smoking$white = rep(0, n)
smoking$white[smoking$mrace == "1" | smoking$mrace == "2" | smoking$mrace == "3" | smoking$mrace == "4"
smoking$mexican = rep(0, n)
smoking$mexican[smoking$mrace == "6"] = 1
smoking$black = rep(0, n)
smoking$black[smoking$mrace == "7"] = 1
smoking$asian = rep(0, n)
smoking$asian[smoking$mrace == "8"] = 1
smoking$mix = rep(0, n)
smoking$mix[smoking$mrace == "9"] = 1

n = nrow(smoking_dad2)
smoking_dad2$white = rep(0, n)
smoking_dad2$white[smoking_dad2$mrace == "1" | smoking_dad2$mrace == "2" | smoking_dad2$mrace == "3" | s
smoking_dad2$mexican = rep(0, n)
smoking_dad2$mexican[smoking_dad2$mrace == "6"] = 1
smoking_dad2$black = rep(0, n)
smoking_dad2$black[smoking_dad2$mrace == "7"] = 1
smoking_dad2$asian = rep(0, n)
smoking_dad2$asian[smoking_dad2$mrace == "8"] = 1
smoking_dad2$mix = rep(0, n)
smoking_dad2$mix[smoking_dad2$mrace == "9"] = 1
```

4) To obtain more accurate interpretations, I'll substract the mean of the mother's age, height and weight as well as father's height and weight in the second database:

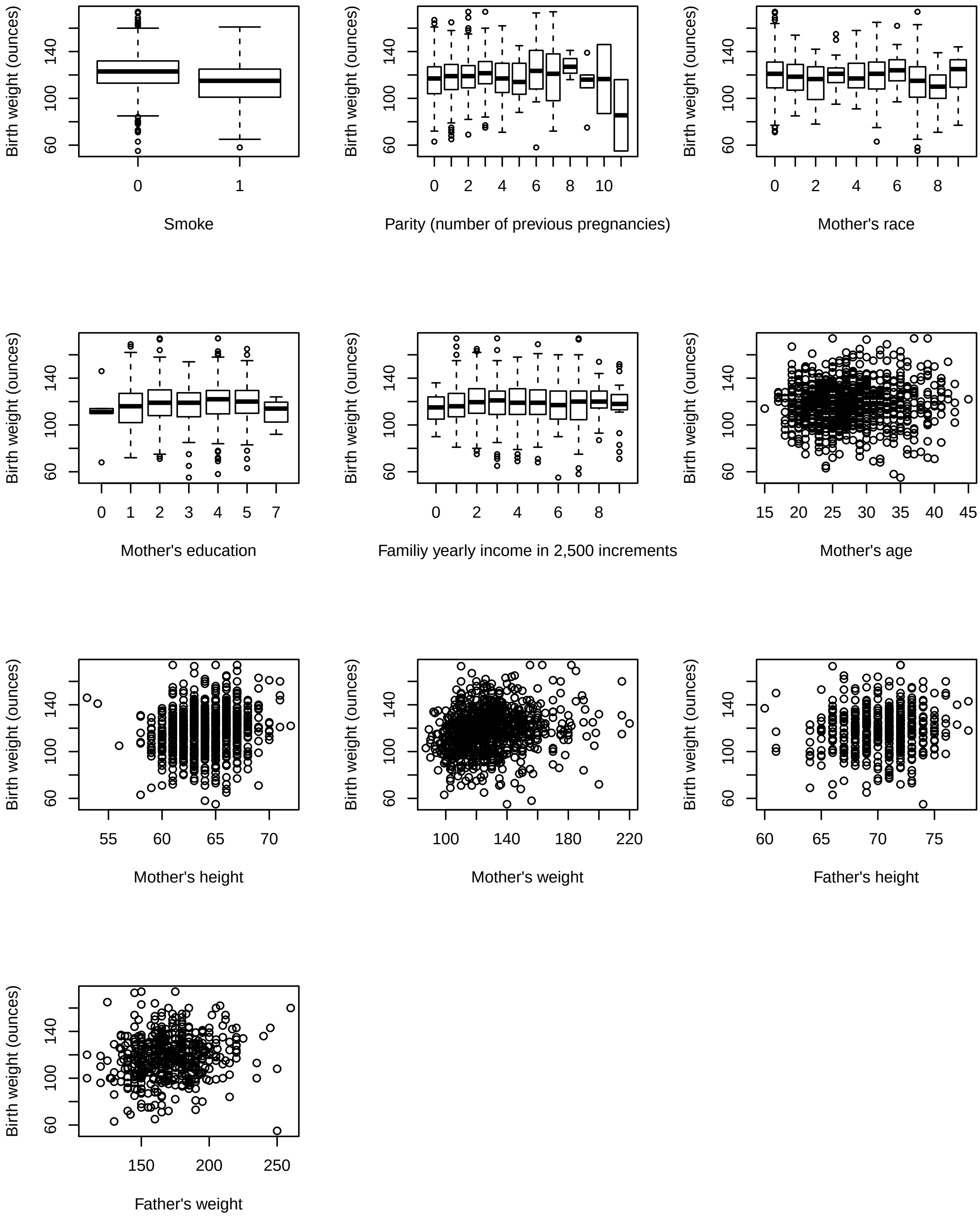
```
smoking$mage_cent = smoking$mage - mean(smoking$mage)
smoking$mht_cent = smoking$mht - mean(smoking$mht)
smoking$mpregwt_cent = smoking$mpregwt - mean(smoking$mpregwt)

smoking_dad2$mage_cent = smoking_dad2$mage - mean(smoking_dad2$mage)
smoking_dad2$mht_cent = smoking_dad2$mht - mean(smoking_dad2$mht)
smoking_dad2$mpregwt_cent = smoking_dad2$mpregwt - mean(smoking_dad2$mpregwt)
smoking_dad2$dht_cent = smoking_dad2$dht - mean(smoking_dad2$dht)
smoking_dad2$dwt_cent = smoking_dad2$dwt - mean(smoking_dad2$dwt)
```

Exploratoy analysis

Let's see the relation of each explanatory variable with birth weight to incorporate transformations if necessary:

```
par(mfrow=c(4,3))

boxplot(bwt.oz~smoke, data = smoking, ylab = "Birth weight (ounces)", xlab = "Smoke")
```

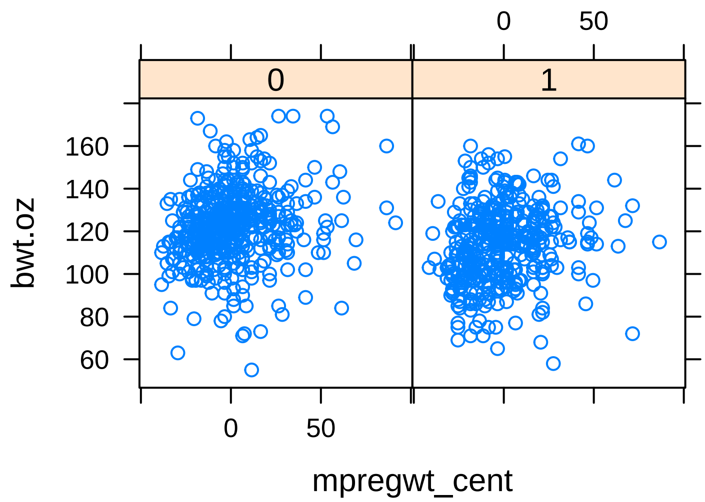
```r
boxplot(bwt.oz~parity, data = smoking, ylab = "Birth weight (ounces)", xlab = "Parity (number of previou
boxplot(bwt.oz~mrace, data = smoking, ylab = "Birth weight (ounces)", xlab = "Mother's race")
boxplot(bwt.oz~med, data = smoking, ylab = "Birth weight (ounces)", xlab = "Mother's education")
boxplot(bwt.oz~inc, data = smoking, ylab = "Birth weight (ounces)", xlab = "Familiy yearly income in 2,5

plot(y = smoking$bwt.oz, x = smoking$mage, xlab = "Mother's age", ylab = "Birth weight (ounces)")
plot(y = smoking$bwt.oz, x = smoking$mht, xlab = "Mother's height", ylab = "Birth weight (ounces)")
plot(y = smoking$bwt.oz, x = smoking$mpregwt, xlab = "Mother's weight", ylab = "Birth weight (ounces)")

plot(y = smoking_dad2$bwt.oz, x = smoking_dad2$dht, xlab = "Father's height", ylab = "Birth weight (ounc
plot(y = smoking_dad2$bwt.oz, x = smoking_dad2$dwt, xlab = "Father's weight", ylab = "Birth weight (ounc

par(mfrow=c(1,1))
```



There is no evidence of any pattern between birth weight and the predictor variables. This suggests that is not necessary to transform neither the birth weight nor the explanatory variables. Also, there is no evidence of important outliers in the predictor variables. A more detailed analysis will be performed later to determine if there are important outliers.

I will proceed to analyze if there should be interactions between predictive variables (the plots are not included but they has been analyzed).

```
par(mfrow=c(4,2))
#xyplot(bwt.oz~mage_cent | as.factor(smoke), data = smoking)
#xyplot(bwt.oz~mht_cent | as.factor(smoke), data = smoking)
xyplot(bwt.oz~mpregwt_cent | as.factor(smoke), data = smoking)
```




```
#xyplot(bwt.oz~dht_cent | as.factor(smoke), data = smoking_dad2)
#xyplot(bwt.oz~dwt_cent | as.factor(smoke), data = smoking_dad2)
#bwplot(bwt.oz~as.factor(smoke) | as.factor(mrace), data = smoking)
#bwplot(bwt.oz~as.factor(smoke) | as.factor(med), data = smoking)
```

There's no evidence of any pattern between the predictive variables. Except for the case of smoking and mother's weight (shown above). There seems to be different patterns between birth weight and mother's weight depending on whether the mother smokes or not. Thus, I'll create an interaction between smoke and mother's weight to fit a model using that interaction and evaluate if it makes a big change.

```
smoking$smoke_mwt = smoking$smoke * smoking$mpregwt_cent
smoking_dad2$smoke_mwt = smoking_dad2$smoke * smoking_dad2$mpregwt_cent
```

Fitting regresion models

Let's define if it is appropiate to include father's data:

- Without father's data

```
reg_weight = lm(bwt.oz~as.factor(smoke) + smoke_mwt + parity + med + mage_cent + mht_cent + mpregwt_cen
summary(reg_weight)
```

```
##
## Call:
## lm(formula = bwt.oz ~ as.factor(smoke) + smoke_mwt + parity +
##     med + mage_cent + mht_cent + mpregwt_cent + inc + mexican +
##     black + asian + mix, data = smoking)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -68.811  -9.284  -0.231  10.183  49.945
##
```

```
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)    
## (Intercept)         123.760927   2.105472  58.781  < 2e-16 ***
## as.factor(smoke)1    -9.344526   1.168623  -7.996 4.15e-15 ***
## smoke_mwt             0.007408   0.055450   0.134 0.893759    
## parity                0.793038   0.388447   2.042 0.041501 *  
## med                   0.133412   0.438856   0.304 0.761202    
## mage_cent            -0.054823   0.130262  -0.421 0.673955    
## mht_cent              0.936936   0.265210   3.533 0.000433 ***
## mpregwt_cent          0.106898   0.041656   2.566 0.010451 *  
## inc                  -0.262057   0.271419  -0.966 0.334563    
## mexican               3.093743   3.482152   0.888 0.374544    
## black                -9.182420   1.552091  -5.916 4.76e-09 ***
## asian                -7.791050   3.090119  -2.521 0.011874 *  
## mix                  -2.145492   4.402249  -0.487 0.626126    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 16.73 on 856 degrees of freedom
## Multiple R-squared:  0.1528, Adjusted R-squared:  0.141 
## F-statistic: 12.87 on 12 and 856 DF,  p-value: < 2.2e-16
```

- With father's data

```r
reg_weight_d = lm(bwt.oz~as.factor(smoke) + smoke_mwt + parity + med + mage_cent + mht_cent + mpregwt_c
summary(reg_weight_d)
```

```
## 
## Call:
## lm(formula = bwt.oz ~ as.factor(smoke) + smoke_mwt + parity + 
##     med + mage_cent + mht_cent + mpregwt_cent + dht_cent + dwt_cent + 
##     inc + mexican + black + asian + mix, data = smoking_dad2)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -72.936  -9.055  -0.286   9.659  52.510 
## 
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)    
## (Intercept)        123.67295    2.96638  41.692  < 2e-16 ***
## as.factor(smoke)1   -9.09121    1.59412  -5.703 2.03e-08 ***
## smoke_mwt            0.04239    0.07470   0.567  0.57064    
## parity               0.46528    0.51749   0.899  0.36903    
## med                  0.33136    0.61881   0.535  0.59256    
## mage_cent            0.02825    0.17306   0.163  0.87038    
## mht_cent             0.90739    0.34904   2.600  0.00961 ** 
## mpregwt_cent         0.08991    0.05276   1.704  0.08898 .  
## dht_cent            -0.03761    0.33265  -0.113  0.91003    
## dwt_cent             0.08022    0.04127   1.944  0.05250 .  
## inc                 -0.31206    0.35593  -0.877  0.38105    
## mexican              5.96215    4.59116   1.299  0.19468    
## black               -8.29862    1.97102  -4.210 3.03e-05 ***
## asian               -7.13855    3.95096  -1.807  0.07140 .  
## mix                 -0.51982    5.50314  -0.094  0.92478    
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.97 on 493 degrees of freedom
## Multiple R-squared:  0.1653, Adjusted R-squared:  0.1416
## F-statistic: 6.975 on 14 and 493 DF,  p-value: 3.542e-13
```

Given the previous results of the regressions with and without father's height and weight, I conclude that including father's information do not worth it. The reasons are the following:

1) First of all, if I include father's data in the analysis, I will lose almost half of the observations of the original data.

2) It could be the case that losing those observations worth it if the model including father's data improve a lot the predictive power of the model. This is not the case since the difference in the R-squared is minimal (15.28% vs. 16.53%).

3) On the other hand, including those variables produces an increase in the standard errors of each coefficient, what suggests that there is a strong correlation between mother's and father's height and weight. After calculating pearson correlation, I can conclude that those variables are in fact strongly correlated.

On the other hand, after modeling the previous equations with and without the interaction between smoking and mother's weight, I can conclude that the interaction is not adding value to the prediction of birth weight: the R-squared does not change at all and the interaction is not statistically significant. Thus, I will exclude the interaction in the final model to avoid overfitting.

```r
reg_weight_f = lm(bwt.oz~as.factor(smoke) + parity + med + mage_cent + mht_cent +
                    mpregwt_cent + inc + mexican + black + asian + mix , data = smoking)
```

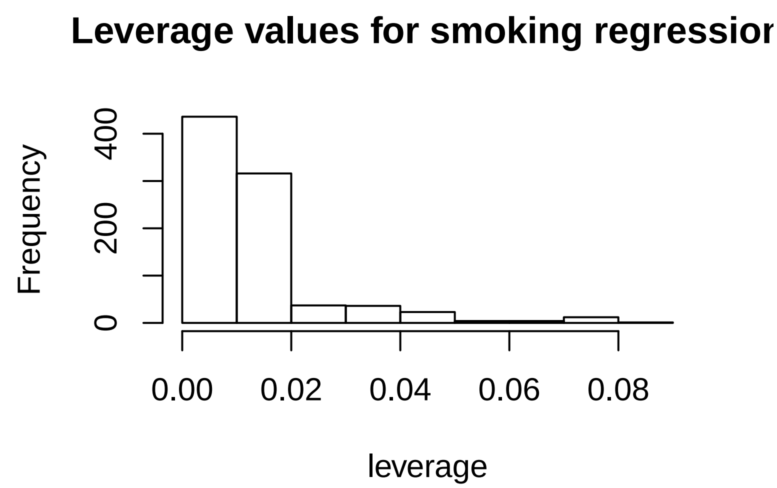
Outliers, leverage, and/or influential points.

Now that I have decided about what predictors, interactions, and transformations I will include in the model, I'll double check if there is any outlier, leverage points, and/or influential points to pay attention in.

```r
library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```r
leverage = hatvalues(reg_weight_f)
cooks = cooks.distance(reg_weight_f)
new_dataset = cbind(smoking, leverage, cooks)

hist(leverage, main = "Leverage values for smoking regression")
```



It seems that if I set leverage > 0.07 I could find some leverage points. Let's see some examples to determine if there is something weird in the data:

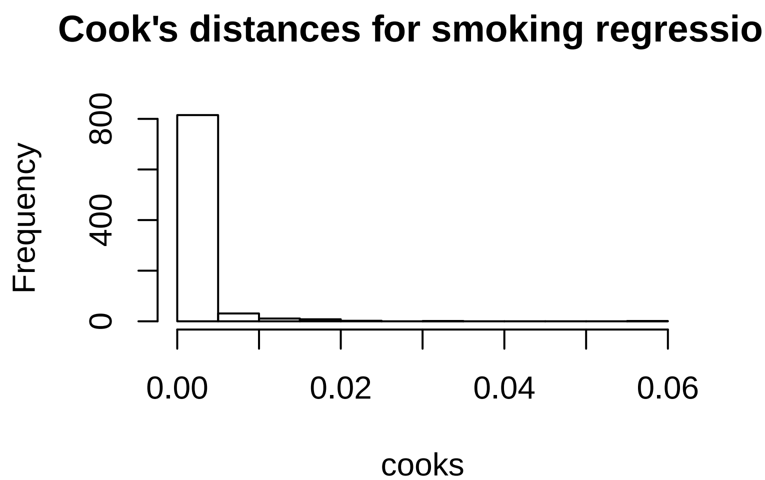
```r
new_dataset[new_dataset$leverage > .07,]
```

```
##       id date gestation bwt.oz parity mrace mage med mht mpregwt inc smoke
## 83   253 1553       270    105      3     9   22   2  56      93   3     0
## 114 7762 1616       273    101      5     9   39   2  60     113   3     0
## 137 8130 1694       274    140      5     9   41   4  63     122   8     0
## 193 2794 1533       278    136      1     9   23   4  61     105   3     0
## 212 7097 1414       279    125      0     9   21   4  66     126   5     0
## 222 6210 1441       280    130      2     9   29   1  66     135   2     0
## 259 8123 1692       282    128      0     9   19   4  66     118   4     0
## 334 4496 1690       286    125      1     9   21   2  64     139   6     0
## 376 4509 1401       290    114      3     9   30   2  66     160   1     0
## 473 7333 1574       238     77      0     9   23   4  63     103   7     1
## 510 6122 1713       263    146     10     6   39   0  53     110   3     1
## 559 6813 1354       270     93      5     9   25   1  64     125   2     1
## 867 3917 1703       329    144      3     9   22   2  65     190   2     1
##     white mexican black asian mix mage_cent     mht_cent mpregwt_cent
## 83      0       0     0     0   1 -5.294591  -8.06904488   -35.478711
## 114     0       0     0     0   1 11.705409  -4.06904488   -15.478711
## 137     0       0     0     0   1 13.705409  -1.06904488    -6.478711
## 193     0       0     0     0   1 -4.294591  -3.06904488   -23.478711
## 212     0       0     0     0   1 -6.294591   1.93095512    -2.478711
## 222     0       0     0     0   1  1.705409   1.93095512     6.521289
## 259     0       0     0     0   1 -8.294591   1.93095512   -10.478711
## 334     0       0     0     0   1 -6.294591  -0.06904488    10.521289
## 376     0       0     0     0   1  2.705409   1.93095512    31.521289
## 473     0       0     0     0   1 -4.294591  -1.06904488   -25.478711
## 510     0       1     0     0   0 11.705409 -11.06904488   -18.478711
## 559     0       0     0     0   1 -2.294591  -0.06904488    -3.478711
## 867     0       0     0     0   1 -5.294591   0.93095512    61.521289
##       smoke_mwt   leverage        cooks
## 83     0.000000 0.07924952 0.0014863164
## 114    0.000000 0.07697654 0.0086261604
## 137    0.000000 0.07699791 0.0091635011
## 193    0.000000 0.07021936 0.0087926315
## 212    0.000000 0.07096315 0.0001249640
```

```
## 222   0.000000 0.07345981 0.0005753107
## 259   0.000000 0.07191372 0.0008280021
## 334   0.000000 0.07131357 0.0001464320
## 376   0.000000 0.07297409 0.0058050206
## 473 -25.478711 0.07608897 0.0245770669
## 510 -18.478711 0.07787550 0.0323166205
## 559  -3.478711 0.07458996 0.0131809760
## 867  61.521289 0.08462318 0.0140970077
```

There is nothing that really stands out in these cases. I'll proceed to check for cooks distance.

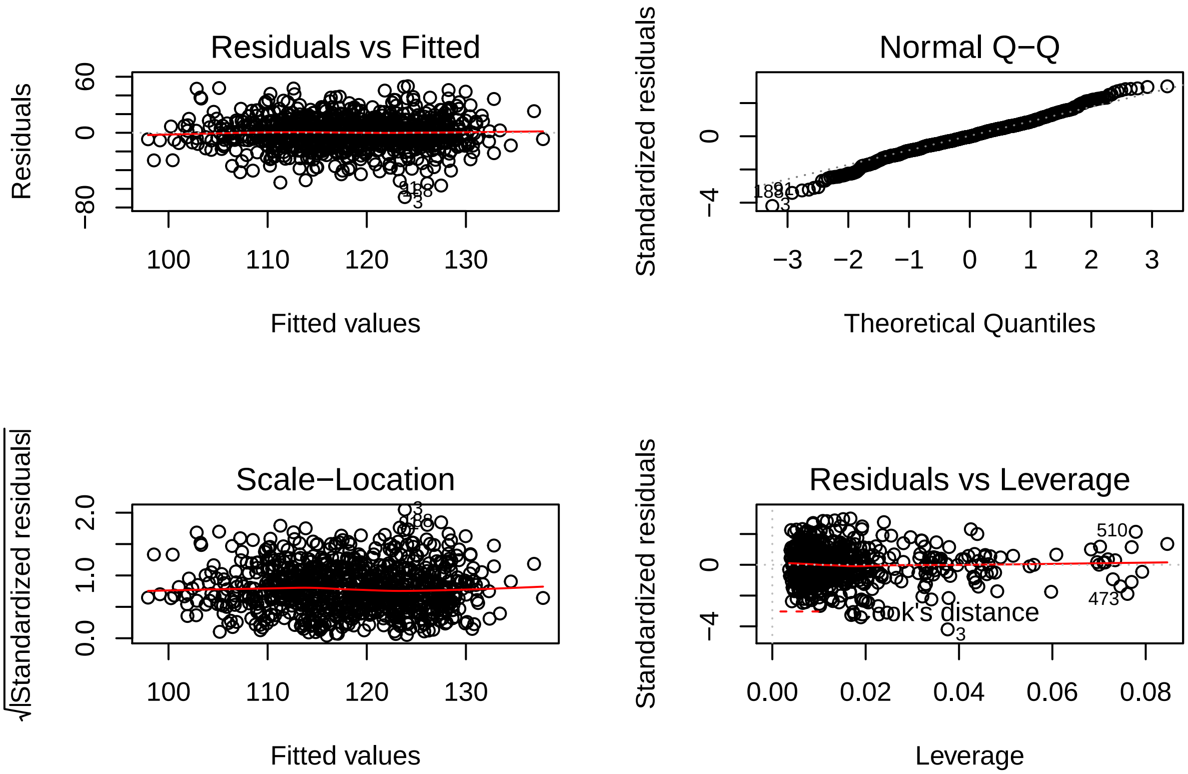
```r
hist(cooks, main = "Cook's distances for smoking regression")
```



The bar graph above suggests that there is no evidence of outliers in the data. If there would be an outlier or leverage point, the next step will be to fit the model again without that atypical observation to check if something changes dramatically. Since I did not find neither outliers nor leverage points, I will proceed to check the regression assumptions.

Regression assumptions

```r
par(mfrow=c(2,2))
plot(reg_weight_f)
```




```r
par(mfrow=c(1,1))
```

These graphs looks pretty well. The linearity, constant variance, and normality assumptions seems to be met. Now, I feel confident about not including transformations and/or interactions in the final model. Besides that the assumptions are met, I do not have to worry about being overfitting the model.

Are smoking and mother's race significant predictors of birth weight?

Let's do a nested-F test to determine if smoking is a useful and significant predictor:

```r
# Model excluding mother's race:
reg_weight_f2 = lm(bwt.oz~parity + med + mage_cent + mht_cent + mpregwt_cent + inc + mexican + black + a

anova(reg_weight_f2, reg_weight_f)
```

```
## Analysis of Variance Table
##
## Model 1: bwt.oz ~ parity + med + mage_cent + mht_cent + mpregwt_cent +
##     inc + mexican + black + asian + mix
## Model 2: bwt.oz ~ as.factor(smoke) + parity + med + mage_cent + mht_cent +
##     mpregwt_cent + inc + mexican + black + asian + mix
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    858 257496
## 2    857 239602  1     17895 64.005 4.016e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Given the p-value we can reject the null hypotheses: it seems that whether the mother smokes or not is a really significant predictor of birth weight.

Let's do a nested-F test to determine if the mother's race a useful and significant predictor:

```r
# Model excluding mother's race:
reg_weight_f3 = lm(bwt.oz~as.factor(smoke) + parity + med + mage_cent + mht_cent + mpregwt_cent + inc, d

anova(reg_weight_f3, reg_weight_f)
```

```
## Analysis of Variance Table
##
## Model 1: bwt.oz ~ as.factor(smoke) + parity + med + mage_cent + mht_cent +
##     mpregwt_cent + inc
## Model 2: bwt.oz ~ as.factor(smoke) + parity + med + mage_cent + mht_cent +
##     mpregwt_cent + inc + mexican + black + asian + mix
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    861 251569
## 2    857 239602  4     11968 10.701 1.819e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value suggests that the mother's race is also a significant predictor of birth weight.

Interpretation of the coefficients and confidence intervals

```r
summary(reg_weight_f)
```

```
##
## Call:
## lm(formula = bwt.oz ~ as.factor(smoke) + parity + med + mage_cent +
##     mht_cent + mpregwt_cent + inc + mexican + black + asian +
##     mix, data = smoking)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -68.837  -9.290  -0.277  10.222  49.852
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       123.74460    2.10072  58.906  < 2e-16 ***
## as.factor(smoke)1  -9.34393    1.16795  -8.000 4.02e-15 ***
## parity              0.79350    0.38821   2.044 0.041259 *
## med                 0.13608    0.43815   0.311 0.756193
## mage_cent          -0.05582    0.12997  -0.430 0.667658
## mht_cent            0.93711    0.26506   3.536 0.000429 ***
## mpregwt_cent        0.11036    0.03260   3.385 0.000744 ***
## inc                -0.26114    0.27118  -0.963 0.335831
## mexican             3.10633    3.47888   0.893 0.372155
## black              -9.19076    1.54995  -5.930 4.40e-09 ***
## asian              -7.76078    3.08004  -2.520 0.011926 *
## mix                -2.12876    4.39794  -0.484 0.628485
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.72 on 857 degrees of freedom
## Multiple R-squared:  0.1528, Adjusted R-squared:  0.1419
```

```
## F-statistic: 14.05 on 11 and 857 DF,  p-value: < 2.2e-16
```

```r
confint(reg_weight_f)
```

```
##                          2.5 %      97.5 %
## (Intercept)       119.62144705 127.8677526
## as.factor(smoke)1 -11.63630214  -7.0515666
## parity              0.03154485   1.5554489
## med                -0.72388985   0.9960536
## mage_cent          -0.31092384   0.1992750
## mht_cent            0.41687206   1.4573392
## mpregwt_cent        0.04636970   0.1743476
## inc                -0.79338345   0.2711108
## mexican            -3.72179283   9.9344580
## black             -12.23289280  -6.1486179
## asian             -13.80608032  -1.7154859
## mix               -10.76075936   6.5032461
```

– Intercept: The average birth weight of babies whose mother do not smoke, is white, has average height, weight, and age, zero previous pregnancies, zero years of education and income, is 123.7 ounces. We are 95% confident that the average birth weight when the person has the characteristics mentioned above falls between 119.6 and 127.9.

– Smoke: The average birth weight of babies with white mothers with average height, weight, and age (and zero in the secondary predictors) is 9.34 ounces less if the mother smoke than if the mother does not smoke. We are 95% confident that the average birth weight when the mother smoke and has the characteristics mentioned above decrease around 11.6 and 7.1 ounces.

– Mother's race: Assuming a white mother with average height, weight, and age (and zero in the secondary predictors), the average birth weight of her baby will be 3.1 ounces higher if the mother is Mexican instead of white (95% CI: -3.7 - 9.9); 9.2 ounces lower if the mother is black instead of white (95% CI: -12.2 - -6.1); 7.8 ounces lower if the mother is asian instead of white (95% CI: -13.8 - 1.7); and 2.1 lower if the mother's race is a mix instead of white (95% CI: -10.8 - 6.5).

– Besides smoking and the mother's race, mother's height and weight seems to be strong predictors of birth weight. This could be interpreted as an inheritance from mothers to babies and/or as the health status of the mother that is shaping the birth weight of the baby. For each additional inch in the mother's height, the average birth weight increase 0.94 ounces (95% CI: 0.42 - 1.46). For each additional pound in the mother's weight, the average birth weight increase by 0.11 ounces (95% CI: 0.05 - 0.17).

Conclusion and limitations of the model

According with the findings previously shown, mothers who smoke tend to have babies whose birth weight is lower compared with mothers that do not smoke. The CI of the effect of smoking on birth weight is narrow, what provides conclusive evidence of the negative relation between smoking and birth weight. Moreover, the mother's race seems to be a strong predictor of babies birth weight. Finally, there are another interesting associations between birth weight and mother's height and weight. This could be associated with inheritance and/or mother's health condition.

The model explains 15% of the variance, which is not that bad for human related analysis. Nevertheless, it could be useful to collect more observations to obtain more accurate predictions. Also, including predictors releated with parent's general health condition could be useful to build a better model.