DATA SIMULATION PROJECT FOR HS-616

Vaishali Chaudhuri

May 23, 2015

Title : A simulated study that shows the association of Intelligence Quotience and Maternal and Infant Factors

Reference Links :http://jama.jamanetwork.com/article.aspx?articleid=194901

Study design , Setting and Participants described in paper

This population study is a prospective longitudinal sub-sample derived from the main Copenhagen Perinatal Cohort comprising of 9125 individuals born at the Copenhagen University Hospital between October 1959 and December 1961.This sub -cohort consists of a sample of 973 men and women When the cohort was established, demographic, socioeconomic, prenatal, and postnatal medical data were recorded prospectively during pregnancy, at delivery, and at a 1-year examination. Information on duration of breastfeeding was collected by a physician who interviewed the mothers at the 1-year examination.

Introduction: This is an Data Simulation Project that references the above link and builds the story based on the paper. Breastfeeding has clear short-term benefits for child survival through reduction of morbidity and mortality from infectious diseases.The paper concludes that certain other parameters (parental and infant) determine Intelligence during adult stage of life determined by WAIS scores Analytics on this simulated dataset is aimed to first generate the data set and then find the answer which mystery parameter has long term benefits on IQ and what's the relation between IQ and the mystery parameter.

Participant's data

- Sex : 976 singletons (490 males and 486 females)
- Age : Mean assessment age of 27.2 years (SD = 4.4; range, 20-34 years)

Main Outcome Measure Intelligence was assessed using the Wechsler Adult Intelligence Scale (WAIS) at a mean age of 27.2 years in the mixed-sex sample

Factors that affect the outcome There is a main factor (not revealed but left for the analyst to come up with) on which the out come depended however thirteen potential confounders were included as covariates: It is upto the analyst to predict which is the primary variable on which the WAISscore is dependant .

```
generateTable <- function(N){

  ## Statistical Data  for the Parents ##

  MA <- runif(N, min=(29.3-6.6), max=(29.3+6.6))   # Maternal Age at time of pregnancy
  MA[1] <- 45
  PSS <-runif(N, min=(4.6-1.9), max=(4.6+1.9)) # Social_Status
  BE <- runif(N, min=(2.6-0.8), max=(2.6+0.8)) # Breadwinners_Education
  MH <- runif(N, min=(163.3-5.4), max=(163.3+5.4)) # Mother's Height (cm)
  MW <- runif(N, min=(4.2-2.5), max=(4.2+2.5)) # Mother's weight gain during pregnancy (kg)
```

```r
SM <- sample(c("SMOKER", "NON_SMOKER"), N, replace=TRUE, prob=c(.4, .6))#smokers & nonsmokers
CC <- ifelse(SM=="SMOKER", runif(N *(0.4),min=(3.7-1.2), max=(3.7+1.2)),0)
NP <- runif(N, min=(2.0-1.2), max=(2.0+1.2)) # No. of pregnancies
PC <- runif(N, min=(70.6-37.6), max=(70.6+37.6)) # Pregnancy Complications
DC <- runif(N, min=(71.6-40.5), max=(71.6+40.5)) # Delivery Complications

##### Infant Characteristics
#Intelligence scores were also affected by 3 factors defined as infant characteristics  at the   ti
GA <-runif(N, min=(39.2-2.0), max=(39.2+2.0)) # Estimated gestational age(GA) (wk)
BW <-runif(N, min=(3251-562), max=(3251+562)) # Birth weight(BW) (g)
BL <-runif(N, min=(51.1-2.6), max=(51.1+2.6)) # Birth height(BL) (cm)

DBF<- DBF <- (
             10^(-0.3) * (MA) +
             10^(-1) * (PSS) +
             10^(-1.2) * (BE) +
            -10^(-0.4) * (CC) -6)
# Finally the output is in the form of IQ score of the participants which is WAIS score of the partic

WAISscore <- 20*DBF - DBF^2 + rnorm(N, sd=2)

#Generating data frame based on parental and  infant characteristics
dataframe1<- data.frame(MA,PSS,BE,MH,MW,CC,NP,PC,DC,GA,BW,BL,DBF,WAISscore)

}

P_dataset<-generateTable(10e3)

head(P_dataset)
```

```
##         MA      PSS       BE       MH       MW       CC       NP       PC
## 1 45.00000 4.035046 3.130615 168.3200 4.313808 0.000000 1.693592 64.73760
## 2 32.11494 4.257518 2.399431 168.6662 5.713586 2.677617 1.531503 72.26266
## 3 32.79926 4.232255 1.849517 164.1501 4.452190 0.000000 2.893343 73.18973
## 4 22.90177 4.753393 2.296244 164.4860 4.112760 0.000000 2.547453 76.34482
## 5 27.83386 2.767922 3.278440 163.9492 5.456723 4.133118 1.062726 95.16074
## 6 25.30465 2.809573 2.055902 168.2602 4.556884 0.000000 2.956251 57.33462
##         DC       GA       BW       BL       DBF WAISscore
## 1 55.17401 38.69677 3059.162 49.01443 17.154459  45.17005
## 2 47.49712 40.99378 3595.789 53.54193  9.606764  98.17784
## 3 60.96090 38.55678 2925.080 48.71771 10.978491  99.96636
## 4 37.96714 37.67042 3334.617 49.59476  6.098297  87.66417
## 5 70.22805 38.65870 3551.774 49.94395  6.788200  91.53762
## 6 52.91664 37.96200 3643.594 48.64721  7.093045  91.71313
```

```r
# Adding a few outliers to the simulated data as is the case in actual world

  P_dataset$MA[1] <- 43
  P_dataset$PSS[1] <- 2.0
  P_dataset$BE[1] <- 3.2
  P_dataset$BL[1] <- 52
  P_dataset$DBF[1] <- (
               10^(-0.3) * (P_dataset$MA[1]) +
```

```r
              10^(-1) * (P_dataset$PSS[1]) +
              10^(-1.2) * (P_dataset$BE[1]) +
             -10^(-0.4) * (P_dataset$CC[1]) -6)
  P_dataset$WAISscore[1] <- 20*(P_dataset$DBF[1]) - (P_dataset$DBF[1])^2


 P_dataset$MA[6] <- 47
 P_dataset$PSS[6] <- 2.0
 P_dataset$BE[6] <- 3.2
 P_dataset$BL[6] <- 52
 P_dataset$DBF[6] <- (
              10^(-0.3) * (P_dataset$MA[6]) +
              10^(-1) * (P_dataset$PSS[6]) +
              10^(-1.2) * (P_dataset$BE[6]) +
             -10^(-0.4) * (P_dataset$CC[6]) -6)
  P_dataset$WAISscore[1] <- 20*(P_dataset$DBF[6]) - (P_dataset$DBF[6])^2
head(P_dataset)
```

```
##         MA      PSS       BE       MH       MW       CC       NP       PC
## 1 43.00000 2.000000 3.200000 168.3200 4.313808 0.000000 1.693592 64.73760
## 2 32.11494 4.257518 2.399431 168.6662 5.713586 2.677617 1.531503 72.26266
## 3 32.79926 4.232255 1.849517 164.1501 4.452190 0.000000 2.893343 73.18973
## 4 22.90177 4.753393 2.296244 164.4860 4.112760 0.000000 2.547453 76.34482
## 5 27.83386 2.767922 3.278440 163.9492 5.456723 4.133118 1.062726 95.16074
## 6 47.00000 2.000000 3.200000 168.2602 4.556884 0.000000 2.956251 57.33462
##         DC       GA       BW       BL       DBF WAISscore
## 1 55.17401 38.69677 3059.162 52.00000 15.952957  36.67491
## 2 47.49712 40.99378 3595.789 53.54193  9.606764  98.17784
## 3 60.96090 38.55678 2925.080 48.71771 10.978491  99.96636
## 4 37.96714 37.67042 3334.617 49.59476  6.098297  87.66417
## 5 70.22805 38.65870 3551.774 49.94395  6.788200  91.53762
## 6 52.91664 37.96200 3643.594 52.00000 17.957706  91.71313
```

Duration of breastfeeding was positively associated with mother's age, social status, education, birth weight and negatively associated with cigarette consumption. So we form an equation giving weightage to these parameters according to their association To choose the correct coefficient values to the different parameters responsible for the "Duration of Breast feeding" equation . The manipulate function is used to set the parameters

```r
logistic <- function(t) 1 / (1 + exp(-t))
library(manipulate)
manipulate({
 DBF <- with (P_dataset ,
              10^a * (MA) +
              10^b * (PSS) +
              10^c * (BE) +
             -10^e * (CC) -6)
             hist(DBF, breaks=50)
},
a=slider(-9, 9, step=0.1, initial = -1),
b=slider(-9, 9, step=0.1, initial = -1),
c=slider(-9, 9, step=0.1, initial = -1),
d=slider(-9, 9, step=0.1, initial = -1),
```

```
e=slider(-9, 9, step=0.1, initial = -1))

#a=-0.3,b=-1,c=-1.2,e=-0.4
```

The next steps is to change the abbreviated column names to more descriptive names and create a simulated Data Table which describes the IQ level of an adult at 27.2 years of age that might could have been influenced by a bunch of factors in infancy or by the characteristics determined by the parents . Our job is to back run analytics and find out which one .

```
colnames(P_dataset) <- c("MaternaL_Age","Social_Status","Parent_Educn","Mothers_ht","Mothers_wt_gain","C
IQ_cohort <-P_dataset
names(IQ_cohort) <- tolower(names(IQ_cohort))
head(IQ_cohort)
```

```
##   maternal_age social_status parent_educn mothers_ht mothers_wt_gain
## 1     43.00000      2.000000     3.200000   168.3200        4.313808
## 2     32.11494      4.257518     2.399431   168.6662        5.713586
## 3     32.79926      4.232255     1.849517   164.1501        4.452190
## 4     22.90177      4.753393     2.296244   164.4860        4.112760
## 5     27.83386      2.767922     3.278440   163.9492        5.456723
## 6     47.00000      2.000000     3.200000   168.2602        4.556884
##   cig_cons no._of_pregn preg_compl delivery_compl gestational_age birth_wt
## 1 0.000000     1.693592   64.73760       55.17401        38.69677 3059.162
## 2 2.677617     1.531503   72.26266       47.49712        40.99378 3595.789
## 3 0.000000     2.893343   73.18973       60.96090        38.55678 2925.080
## 4 0.000000     2.547453   76.34482       37.96714        37.67042 3334.617
## 5 4.133118     1.062726   95.16074       70.22805        38.65870 3551.774
## 6 0.000000     2.956251   57.33462       52.91664        37.96200 3643.594
##   birth_len durn_breast_feed iq_level
## 1  52.00000        15.952957 36.67491
## 2  53.54193         9.606764 98.17784
## 3  48.71771        10.978491 99.96636
## 4  49.59476         6.098297 87.66417
## 5  49.94395         6.788200 91.53762
## 6  52.00000        17.957706 91.71313
```

```
write.csv(IQ_cohort,file="/Users/vchaudhuri/Desktop/HS-616/IQ_data.csv")
```

Analytics

Exploratory Analysis (To make sense of this data table and predict which variable should be considered in prediction of the outcome) I used the correlogram package to generate a correlation matrix of the different vraiables. It is very useful to highlight the most correlated variables in a data table. In this plot, correlation coefficients is colored according to the value. Correlation matrix can be also reordered according to the degree of association between variables. The R corrplot package is used here.

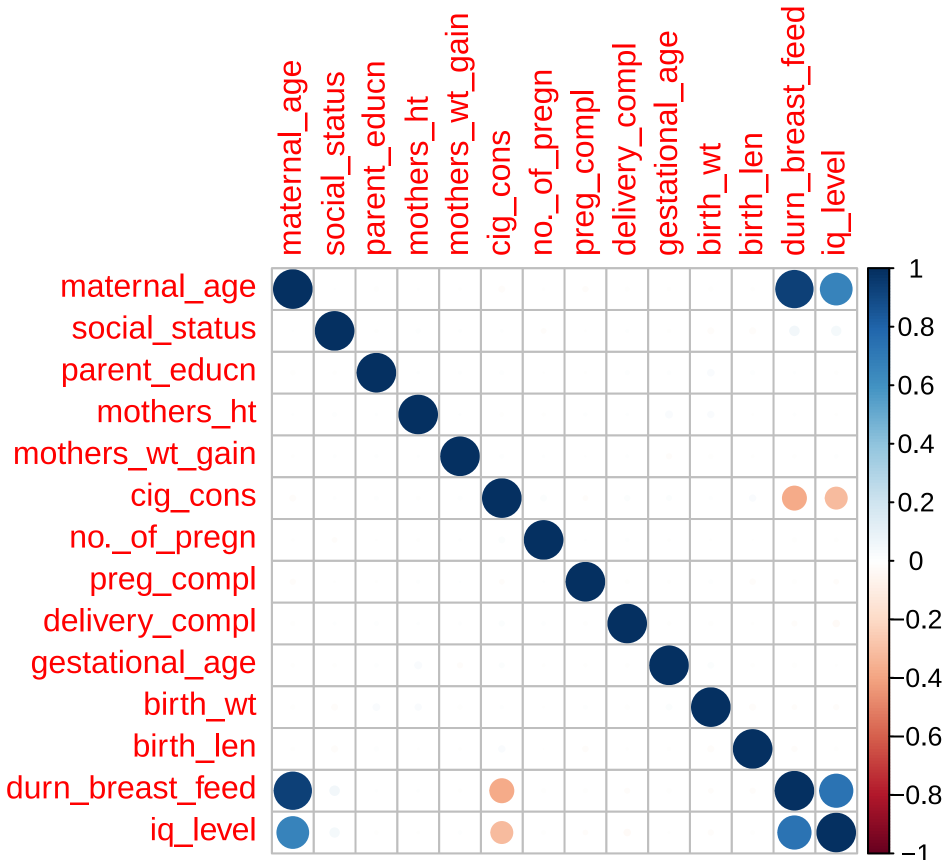
```
library("corrplot")
M<-cor(IQ_cohort)
round(M,2)
```

```
##                    maternal_age social_status parent_educn mothers_ht
```

```
## maternal_age            1.00          0.00        -0.01       0.00
## social_status           0.00          1.00         0.00       0.01
## parent_educn           -0.01          0.00         1.00       0.00
## mothers_ht              0.00          0.01         0.00       1.00
## mothers_wt_gain         0.00          0.00         0.00       0.00
## cig_cons               -0.02          0.00         0.01      -0.01
## no._of_pregn            0.00         -0.01         0.00       0.00
## preg_compl             -0.01          0.00         0.00       0.00
## delivery_compl         -0.01          0.00         0.01       0.00
## gestational_age        -0.01         -0.01         0.01       0.03
## birth_wt               -0.01         -0.02         0.02       0.02
## birth_len              -0.01         -0.02         0.01       0.00
## durn_breast_feed        0.93          0.05         0.00       0.00
## iq_level                0.67          0.05         0.00       0.00
##                  mothers_wt_gain cig_cons no._of_pregn preg_compl
## maternal_age                0.00    -0.02         0.00      -0.01
## social_status               0.00     0.00        -0.01       0.00
## parent_educn                0.00     0.01         0.00       0.00
## mothers_ht                  0.00    -0.01         0.00       0.00
## mothers_wt_gain             1.00     0.01         0.00       0.00
## cig_cons                    0.01     1.00         0.02      -0.01
## no._of_pregn                0.00     0.02         1.00       0.00
## preg_compl                  0.00    -0.01         0.00       1.00
## delivery_compl              0.00     0.01         0.01      -0.01
## gestational_age            -0.01     0.01         0.00       0.00
## birth_wt                   -0.01     0.00         0.00       0.01
## birth_len                   0.00     0.02         0.00      -0.02
## durn_breast_feed            0.00    -0.37        -0.01      -0.01
## iq_level                    0.01    -0.31        -0.01      -0.01
##                  delivery_compl gestational_age birth_wt birth_len
## maternal_age              -0.01           -0.01    -0.01     -0.01
## social_status              0.00           -0.01    -0.02     -0.02
## parent_educn               0.01            0.01     0.02      0.01
## mothers_ht                 0.00            0.03     0.02      0.00
## mothers_wt_gain            0.00           -0.01    -0.01      0.00
## cig_cons                   0.01            0.01     0.00      0.02
## no._of_pregn               0.01            0.00     0.00      0.00
## preg_compl                -0.01            0.00     0.01     -0.02
## delivery_compl             1.00           -0.01    -0.01      0.00
## gestational_age           -0.01            1.00     0.02     -0.01
## birth_wt                  -0.01            0.02     1.00     -0.02
## birth_len                  0.00           -0.01    -0.02      1.00
## durn_breast_feed          -0.01           -0.01    -0.01     -0.01
## iq_level                  -0.02            0.00    -0.01     -0.01
##                  durn_breast_feed iq_level
## maternal_age                 0.93     0.67
## social_status                0.05     0.05
## parent_educn                 0.00     0.00
## mothers_ht                   0.00     0.00
## mothers_wt_gain              0.00     0.01
## cig_cons                    -0.37    -0.31
## no._of_pregn                -0.01    -0.01
## preg_compl                  -0.01    -0.01
## delivery_compl              -0.01    -0.02
```

```
## gestational_age              -0.01     0.00
## birth_wt                     -0.01    -0.01
## birth_len                    -0.01    -0.01
## durn_breast_feed              1.00     0.73
## iq_level                      0.73     1.00
```

```
corrplot(M, method="circle")
```




```
## the corrplot package narRows down the Independant varaiable to Duration of Breast Feeding(DBF). Howe
```

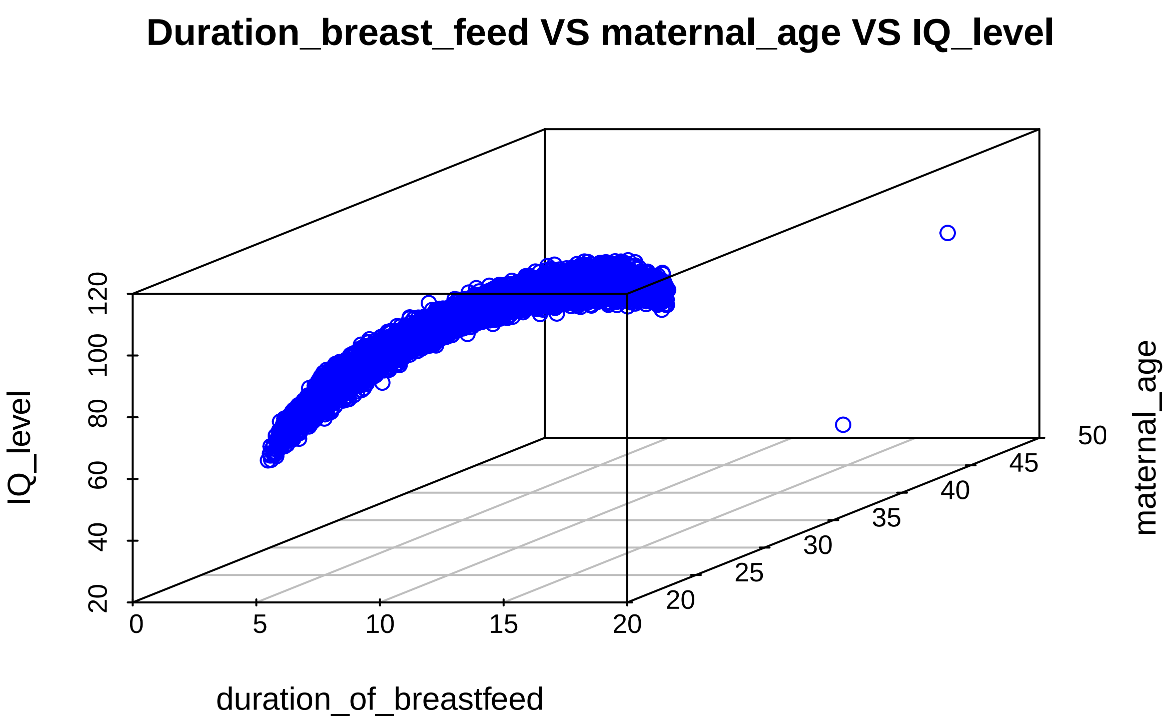
Analyzed each parameter vs iq_level as that gives some idea on which parameter is involved

```
library("ggplot2")
ggplot(IQ_cohort)+geom_point(aes(x=maternal_age,y=iq_level,color='red'))
ggplot(IQ_cohort)+geom_point(aes(x=social_status,y=iq_level))
ggplot(IQ_cohort)+geom_point(aes(x=parent_educn,y=iq_level))
ggplot(IQ_cohort)+geom_point(aes(x=mothers_ht,y=iq_level))
ggplot(IQ_cohort)+geom_point(aes(x=mothers_wt_gain,y=iq_level))
ggplot(IQ_cohort)+geom_point(aes(x=cig_cons,y=iq_level))
ggplot(IQ_cohort)+geom_point(aes(x=no._of_pregn,y=iq_level))
ggplot(IQ_cohort)+geom_point(aes(x=preg_compl,y=iq_level))
ggplot(IQ_cohort)+geom_point(aes(x=delivery_compl,y=iq_level))
```

```
ggplot(IQ_cohort)+geom_point(aes(x=gestational_age,y=iq_level))
ggplot(IQ_cohort)+geom_point(aes(x=birth_wt,y=iq_level))
ggplot(IQ_cohort)+geom_point(aes(x=birth_len,y=iq_level))
ggplot(IQ_cohort)+geom_point(aes(x=durn_breast_feed,y=iq_level))
```

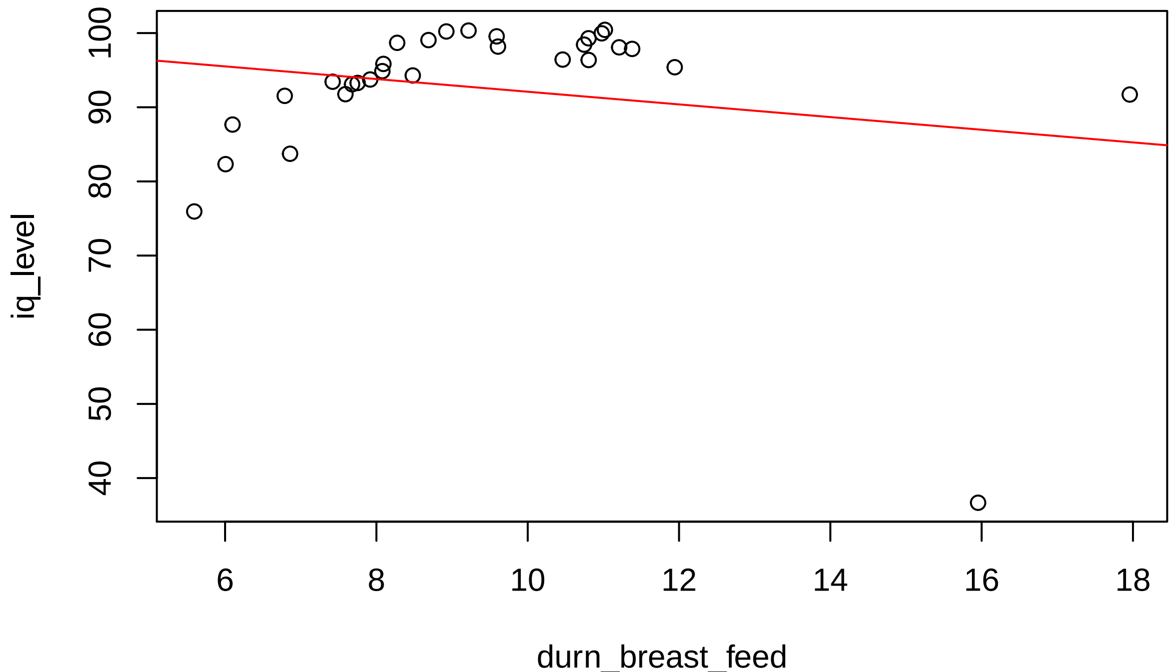
The two variables that seem to affect IQ level seems to be maternal age and Durtaion of breast feeding. So all three are plotted into a 3D scatter plot

```
library("scatterplot3d")

IQ3d <-scatterplot3d( IQ_cohort$durn_breast_feed,IQ_cohort$maternal_age,IQ_cohort$iq_level,
                    color = "blue",
              xlab= "duration_of_breastfeed",ylab= "maternal_age",zlab= "IQ_level",
              main="Duration_breast_feed VS maternal_age VS IQ_level")
```



Checking which parameter and which function gives the best fit

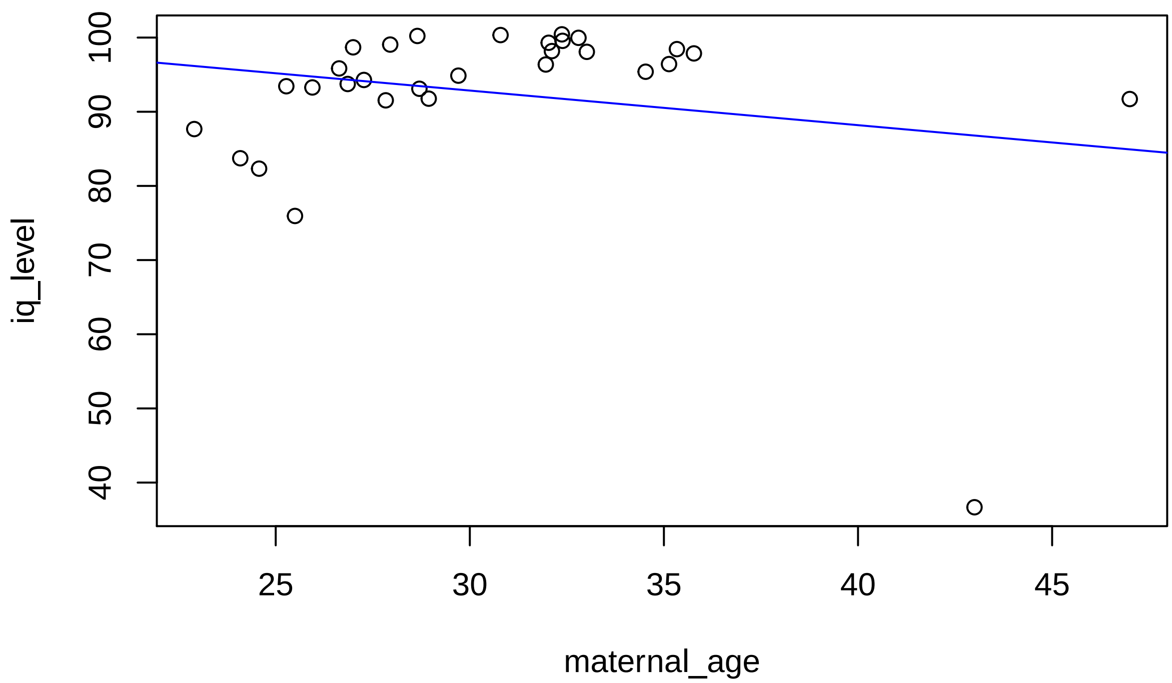
```
#Fitting IQ_level as a function of Duration of Breast feeding in a linear model with a small sample siz

Population_dataset<-IQ_cohort[1:30,]
with (Population_dataset, plot(durn_breast_feed, iq_level))
fit1 <- lm(iq_level ~ durn_breast_feed, Population_dataset)
abline(fit1, col="red")
```




```r
summary(fit1)
```

```
##
## Call:
## lm(formula = iq_level ~ durn_breast_feed, data = Population_dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.332  -0.948   4.846   6.980   9.214
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      100.6345     8.0371  12.521 5.43e-13 ***
## durn_breast_feed  -0.8542     0.8227  -1.038    0.308
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.02 on 28 degrees of freedom
## Multiple R-squared:  0.03707,    Adjusted R-squared:  0.00268
## F-statistic: 1.078 on 1 and 28 DF,  p-value: 0.308
```

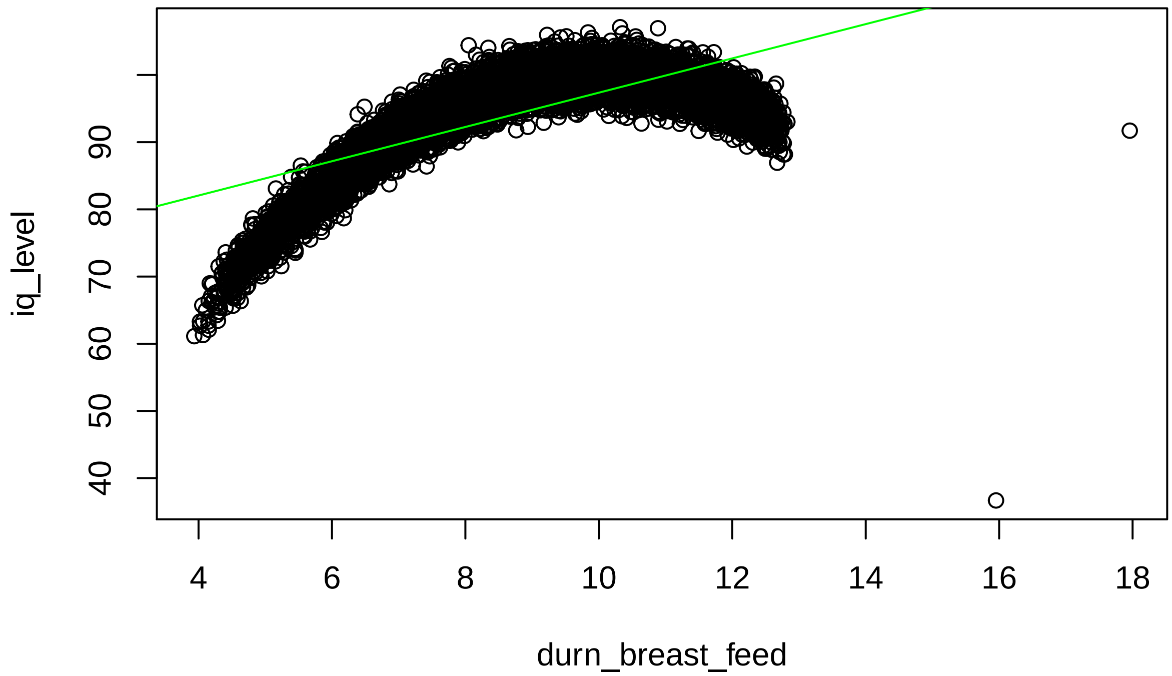
```r
#Fitting IQ_level as a function of Maternal Age in a linear model with a small sample size

Population_dataset<-IQ_cohort[1:30,]
with (Population_dataset, plot(maternal_age, iq_level))
fit2 <- lm(iq_level ~ maternal_age, Population_dataset)
abline(fit2, col="blue")
```




```r
summary(fit2)
```

```
##
## Call:
## lm(formula = iq_level ~ maternal_age, data = Population_dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.118  -1.558   4.426   6.773   8.685
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  106.8510    12.9330   8.262 5.44e-09 ***
## maternal_age  -0.4665     0.4174  -1.117    0.273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.99 on 28 degrees of freedom
## Multiple R-squared:  0.04269,    Adjusted R-squared:  0.008503
## F-statistic: 1.249 on 1 and 28 DF,  p-value: 0.2733
```

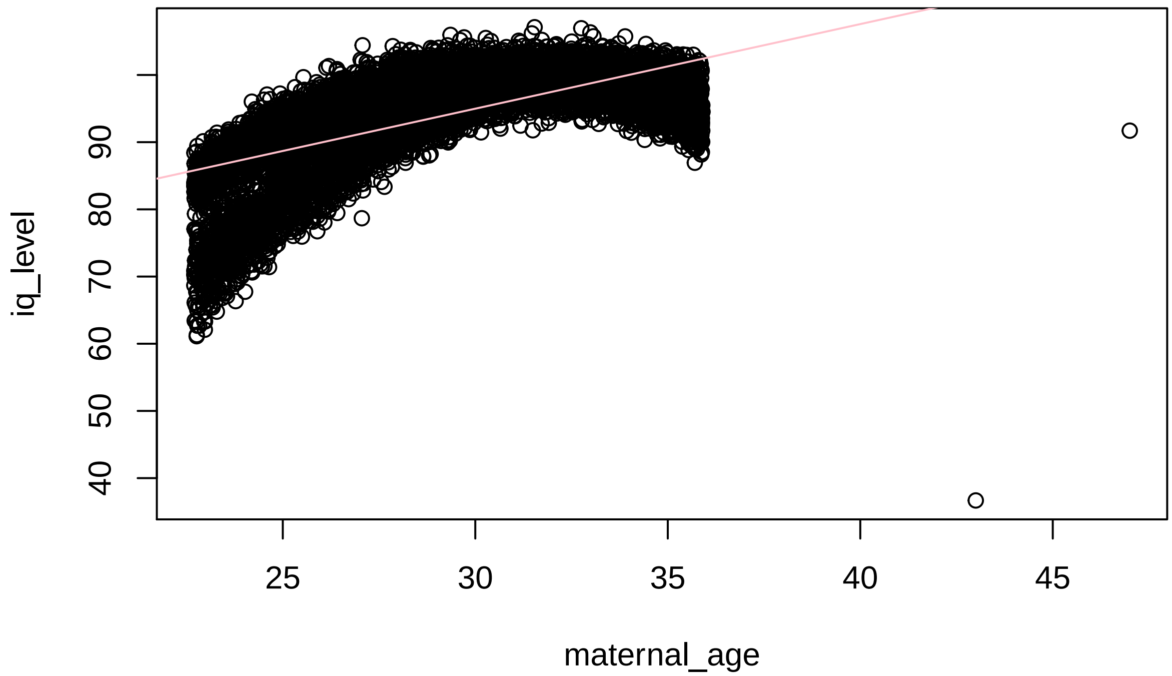
```r
#Fitting IQ_level as a function of Duration of Breast feeding in a linear model with a the entire sampl

Population_dataset<-IQ_cohort
with (Population_dataset, plot(durn_breast_feed, iq_level))
fit3 <- lm(iq_level ~ durn_breast_feed, Population_dataset)
abline(fit3, col="green")
```




```r
summary(fit3)
```

```
##
## Call:
## lm(formula = iq_level ~ durn_breast_feed, data = Population_dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -75.877  -2.540   1.129   3.476  12.037
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      71.85234    0.21230   338.4   <2e-16 ***
## durn_breast_feed  2.55121    0.02368   107.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 9998 degrees of freedom
## Multiple R-squared:  0.5373, Adjusted R-squared:  0.5372
## F-statistic: 1.161e+04 on 1 and 9998 DF,  p-value: < 2.2e-16
```

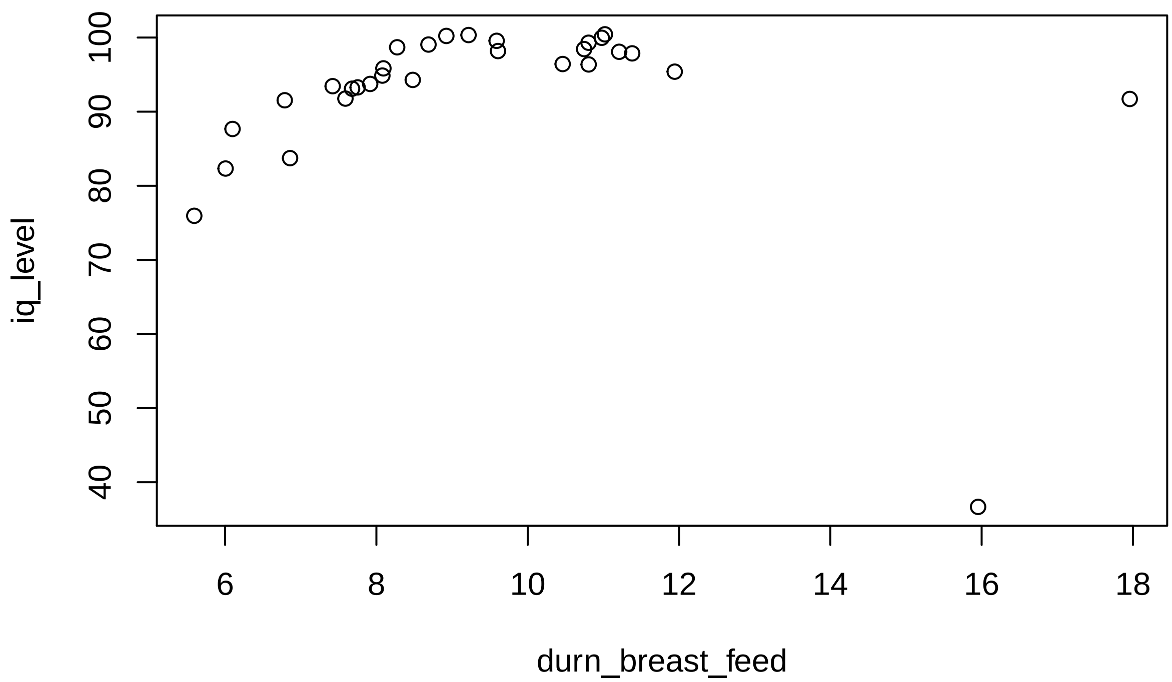
```r
#Fitting IQ_level as a function of Maternal_Age in a linear model with a the entire sample size IQ_coho

Population_dataset<-IQ_cohort
with (Population_dataset, plot(maternal_age, iq_level))
fit4 <- lm(iq_level ~ maternal_age, Population_dataset)
abline(fit4, col="pink")
```




```r
summary(fit4)
```

```
##
## Call:
## lm(formula = iq_level ~ maternal_age, data = Population_dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -74.676  -2.703   0.951   3.772  13.122
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  57.22079    0.41414  138.17   <2e-16 ***
## maternal_age  1.25883    0.01401   89.83   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.358 on 9998 degrees of freedom
## Multiple R-squared:  0.4466, Adjusted R-squared:  0.4466
## F-statistic:  8069 on 1 and 9998 DF,  p-value: < 2.2e-16
```

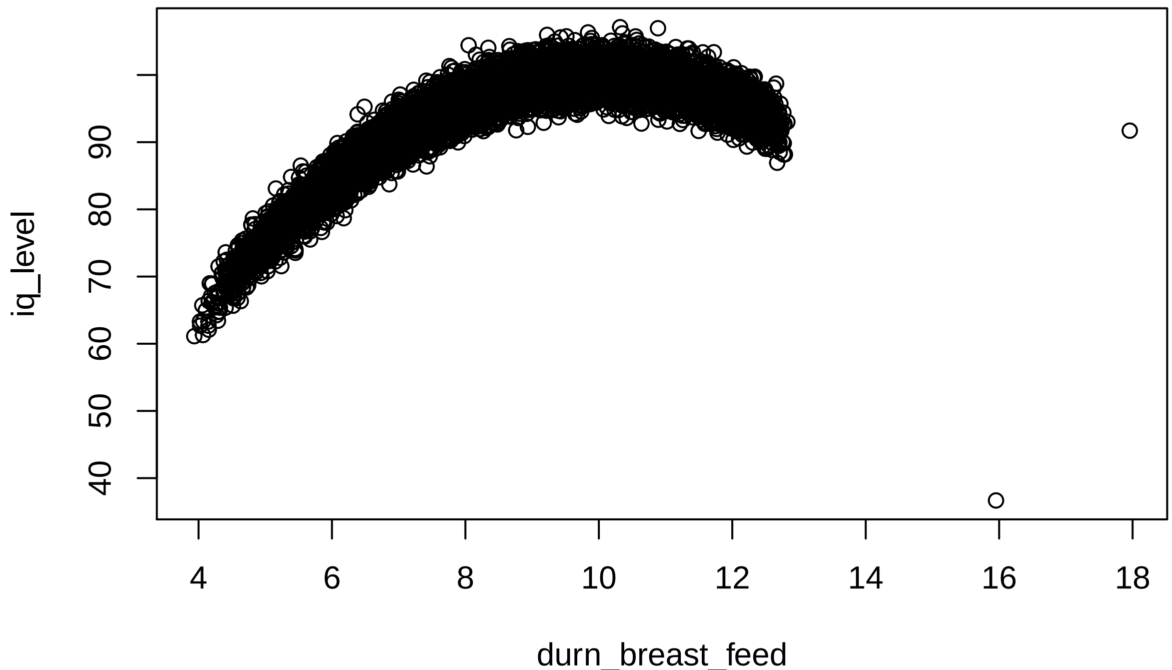
```r
#Fitting IQ_level as a function of duration of Breast feeding in a Quadratic model with a small sample 

Population_dataset<-IQ_cohort[1:30,]
with (Population_dataset, plot(durn_breast_feed, iq_level))
```




```
fit1_quad <- lm(iq_level ~ durn_breast_feed + I((durn_breast_feed)^2), data=Population_dataset)
summary(fit1_quad)
```

```
##
## Call:
## lm(formula = iq_level ~ durn_breast_feed + I((durn_breast_feed)^2),
##     data = Population_dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -43.076  -0.798   1.114   2.478  27.075
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              34.3708    20.4108   1.684  0.10372
## durn_breast_feed         12.0685     3.8160   3.163  0.00384 **
## I((durn_breast_feed)^2)  -0.5782     0.1679  -3.445  0.00188 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.2 on 27 degrees of freedom
## Multiple R-squared:  0.331,  Adjusted R-squared:  0.2815
## F-statistic: 6.681 on 2 and 27 DF,  p-value: 0.004394
```

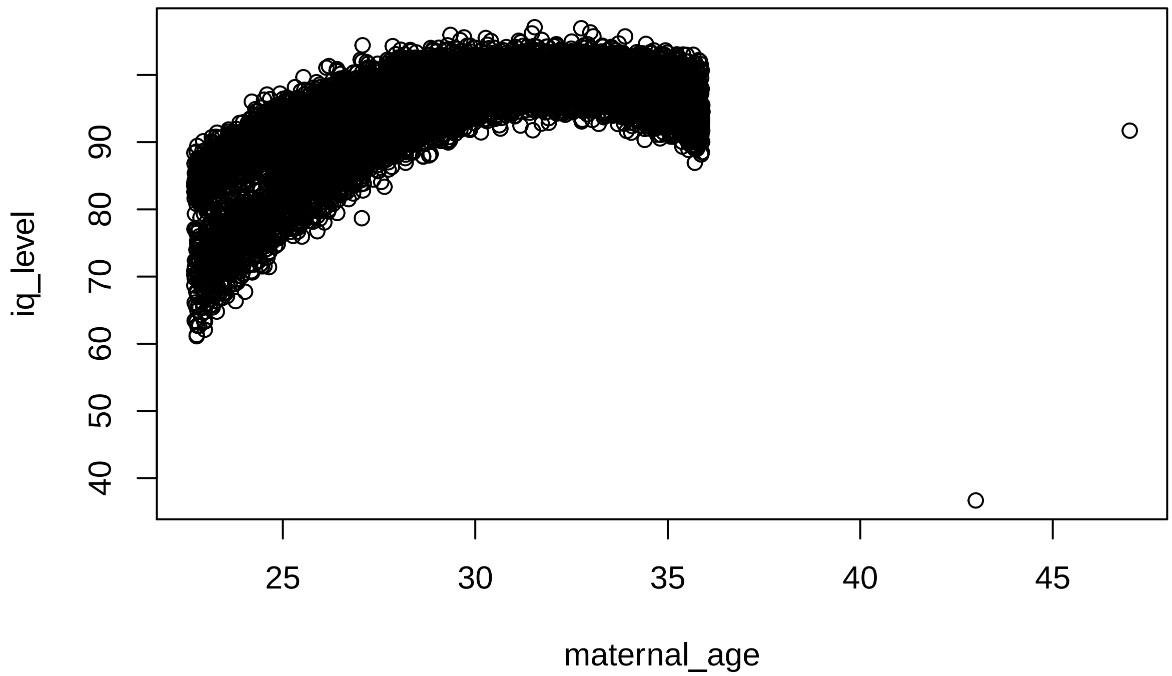
```
#Fitting IQ_level as a function of duration of Breast feeding in a Quadratic model with a large sample

Population_dataset<-IQ_cohort
with (IQ_cohort, plot(durn_breast_feed, iq_level))
```




```r
fit1_quad <- lm(iq_level ~ durn_breast_feed + I((durn_breast_feed)^2), data=Population_dataset)
summary(fit1_quad)
```

```
##
## Call:
## lm(formula = iq_level ~ durn_breast_feed + I((durn_breast_feed)^2),
##     data = Population_dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.875  -1.350  -0.024   1.377  53.390
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              1.435994   0.344396    4.17 3.08e-05 ***
## durn_breast_feed        19.653908   0.081331  241.65  < 2e-16 ***
## I((durn_breast_feed)^2) -0.980069   0.004625 -211.93  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.091 on 9997 degrees of freedom
## Multiple R-squared:  0.9158, Adjusted R-squared:  0.9157
## F-statistic: 5.433e+04 on 2 and 9997 DF,  p-value: < 2.2e-16
```

```r
#Fitting IQ_level as a function of maternal_age in a Quadratic model with a large sample size

Population_dataset<-IQ_cohort
with (IQ_cohort, plot(maternal_age, iq_level))
```




```
fit2_quad <- lm(iq_level ~ maternal_age + I((maternal_age)^2), data=Population_dataset)
summary(fit1)
```

```
##
## Call:
## lm(formula = iq_level ~ durn_breast_feed, data = Population_dataset)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -50.332  -0.948   4.846   6.980   9.214
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      100.6345     8.0371  12.521 5.43e-13 ***
## durn_breast_feed  -0.8542     0.8227  -1.038    0.308
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.02 on 28 degrees of freedom
## Multiple R-squared:  0.03707,    Adjusted R-squared:  0.00268
## F-statistic: 1.078 on 1 and 28 DF,  p-value: 0.308
```

CONCLUSION

R-squared goes from essentially 0 to close to 1 when Iqlevel is a quadratic function of Duration of Breast Feeding