Statistical Analysis of Telebix Data from 3/1/19 to 5/31/19

Telebix

June 03, 2019

Load and Review Data

```
#Load data
telebix <- read.csv("Telebix_Dataset.csv")
```

```
#Check structure of the data
telebix <- telebix %>% mutate(GENDER_NUM = if_else(GENDER == "Male", 0, 1),
                              CHEM_NUM = if_else(CHEM, 1, 0),
                              BIRTH_SITE_NUM = as.numeric(BIRTH_SITE),
                              CITY_NUM = as.numeric(CITY)
                              )
str(telebix)
```

```
## 'data.frame':    1000 obs. of  28 variables:
##  $ Ã¯..PATIENT_ID : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ CLINIC         : Factor w/ 4 levels "Biu","Jos","Kano",..: 2 4 1 3 3 1 1 1 2 2 ...
##  $ DATE           : Factor w/ 91 levels "2019-03-01","2019-03-02",..: 90 14 28 90 86 81 40 70 2 45 ...
##  $ BVALUE         : num  0.11 0.23 0.22 0.04 0.23 0.15 0.28 0.22 0.29 0.22 ...
##  $ AGE            : int  20 63 33 66 68 29 52 1 45 12 ...
##  $ GENDER         : Factor w/ 2 levels "Female","Male": 2 1 1 2 2 1 2 2 2 2 ...
##  $ WEIGHT         : num  5.8 5.3 5.9 7.4 6.4 4.6 6.5 4.1 7.3 5.4 ...
##  $ CHEM           : logi  TRUE FALSE FALSE TRUE TRUE TRUE ...
##  $ ANTENATAL      : int  2 2 1 2 2 1 0 2 2 2 ...
##  $ BIRTH_SITE     : Factor w/ 3 levels "Clinic","Home",..: 3 1 3 1 3 2 1 3 1 2 ...
##  $ CITY.ID        : int  8 18 3 11 12 2 2 3 6 7 ...
##  $ CITY           : Factor w/ 20 levels "Azare","Bauchi",..: 11 18 15 13 20 6 6 15 9 2 ...
##  $ LATITUDE       : num  10.5 12.6 11.8 12 11.1 ...
##  $ LONGITUDE      : num  7.43 4.99 13.14 8.53 7.69 ...
##  $ EST_DIST       : int  178 36 118 10 98 78 78 118 10 81 ...
##  $ ACTUAL_DIST    : int  171 32 126 1 97 80 85 110 4 86 ...
##  $ AGE_WT         : num  0.0392 0.0128 0.0154 0.0155 0.0171 ...
##  $ GENDER_WT      : num  0.0249 0.0136 0.013 0.0235 0.0285 ...
##  $ WEIGHT_WT      : num  0.0215 0.0442 0.0274 0.0235 0.0246 ...
##  $ CHEM_WT        : num  0.047 0.0137 0.0187 0.0418 0.0495 ...
##  $ ANTENATAL_WT   : num  0.00672 0.00951 0.01489 0.00783 0.00564 ...
##  $ BIRTH_SITE_WT  : num  0.0155 0.0182 0.0166 0.0176 0.0174 ...
##  $ X              : num  0.034 0.0124 0.0329 0.0103 0.0326 ...
##  $ BLEVEL         : num  0.212 0.149 0.166 0.166 0.196 ...
##  $ GENDER_NUM     : num  0 1 1 0 0 1 0 0 0 0 ...
##  $ CHEM_NUM       : num  1 0 0 1 1 1 0 0 1 0 ...
##  $ BIRTH_SITE_NUM: num  3 1 3 1 3 2 1 3 1 2 ...
##  $ CITY_NUM       : num  11 18 15 13 20 6 6 15 9 2 ...
```

Fitting data via multiple linear regression

```
all_predictors <- c("AGE", "GENDER_NUM", "WEIGHT", "CHEM_NUM", "ANTENATAL", "BIRTH_SITE_NUM", "CITY_NUM", "ACTUAL_DIST")

bval_fit <- lm(as.formula(paste("BLEVEL~",
        paste(all_predictors, collapse = "+"),
        sep = "")),
    data = telebix)

summary(bval_fit)
```

```
##
## Call:
## lm(formula = as.formula(paste("BLEVEL~", paste(all_predictors,
##     collapse = "+"), sep = "")), data = telebix)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.042622 -0.011488 -0.000353  0.010595  0.051617
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.290e-01  3.623e-03  63.210  < 2e-16 ***
## AGE            -4.070e-04  2.421e-05 -16.808  < 2e-16 ***
## GENDER_NUM     -8.364e-03  1.032e-03  -8.104 1.56e-15 ***
## WEIGHT         -6.648e-03  4.511e-04 -14.736  < 2e-16 ***
## CHEM_NUM        2.894e-02  1.027e-03  28.180  < 2e-16 ***
## ANTENATAL      -6.560e-03  6.231e-04 -10.528  < 2e-16 ***
## BIRTH_SITE_NUM  2.039e-04  6.423e-04   0.318    0.751
## CITY_NUM        1.134e-04  8.952e-05   1.267    0.206
## ACTUAL_DIST     1.425e-04  1.028e-05  13.865  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01621 on 991 degrees of freedom
## Multiple R-squared:  0.6181, Adjusted R-squared:  0.615
## F-statistic: 200.5 on 8 and 991 DF,  p-value: < 2.2e-16
```

Conclusions

Based on the statistical model obtained above, the following variables are considered most statistically significant to explaining the variations in the levels of bilirubin observed.

Risk Factors	Effects on Bilirubin Levels
Age	Every hour going by lowers bilirubin levels by about 0.04%
Gender	A baby girl will have lower bilirubin levels than a baby boy by about 0.8%
Weight	Every pound increase lowers bilirubin levels by about 0.7%
Chemical Exposure	Chemical exposure increases bilirubin levels by almost 3%
Antenatal visits	Every antenatal visit lowers bilirubin levels by about 0.7%
Distance traveled	Every mile traveled increases bilirubin levels by about 0.01%

Making Predictions

```
# c("AGE", "GENDER_NUM", "WEIGHT", "CHEM_NUM", "ANTENATAL", "BIRTH_SITE_NUM", "CITY_NUM", "ACTUAL_DIST")
size <- 10
testdata <- data.frame(
  AGE = round(runif(size, 10, 72), 0),
  GENDER_NUM = round(runif(size, 0, 1), 0),
  WEIGHT = round(runif(size, 4, 8), 2),
  CHEM_NUM = round(runif(size, 0, 1), 0),
  ANTENATAL = round(runif(size, 0, 2), 0),
  BIRTH_SITE_NUM = round(runif(size, 1, 3), 0),
  CITY_NUM = round(runif(size, 1, 20), 0),
  ACTUAL_DIST = round(runif(size, 50, 150), 0)
)

testdata <-
  testdata %>% mutate(BLEVEL = round(predict(bval_fit, newdata = testdata)*100, 0),
                      GENDER = if_else(GENDER_NUM == 1, "Girl", "Boy"),
                      CHEM = if_else(CHEM_NUM == 1, "Yes", "No"),
                      DIST = ACTUAL_DIST
)

testdata %>%
  select(AGE, GENDER, WEIGHT, CHEM, ANTENATAL, DIST, BLEVEL) %>%
  arrange(., desc(BLEVEL)) %>%
  kable(caption = "") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"))
```

AGE	GENDER	WEIGHT	CHEM	ANTENATAL	DIST	BLEVEL
28	Girl	4.19	Yes	0	106	23
52	Girl	6.89	Yes	1	140	20
45	Boy	7.63	Yes	2	148	20
14	Boy	5.41	No	1	89	19
42	Girl	6.33	No	1	136	18
33	Boy	7.43	No	2	117	17
70	Boy	6.68	No	1	96	17
53	Girl	7.00	No	1	120	16
58	Girl	6.04	No	2	126	16
38	Girl	7.46	No	2	54	15