MODELING MALARIA INCIDENCE IN KENYA USING MACHINE LEARNING ALRORTHMS

Lumumba Wandera Victor

2024-07-08

```r
library(tidyverse)
library(haven)
malaria_survey<- read_dta("KNBS Malaria Data set2.dta")
#head(malaria_survey)
```

View the Variables

```r
#View(malaria_survey)
```

Exctract the Most appropriate variables

```r
data <-malaria_survey%>%
  select(hv024,shcounty, hv227, hv228, hml1, hml2, hc57,hc61,hml12, hml32a, hml32b, hml32c, hml32)
#head(data,10)
```

Variable	Description
hv024	Region
shcounty	County of Residence
hv227	Has mosquito bed net for sleeping
hv228	Children under 5 who sleped under mosquito net last night
hml1	Number of Mosquito bed nets
hml2	Number of children under mosquito bet nets last night
hc57	Anaemia Level
hc61	Mothers highest educational level
hml12	The type of mosquito the person slept under last night
hml32a	Presence of species: Falciparum
hml32b	Presence of species: Malariae
hml32c	Presence of species: Ovale
hml32	Final results of malaria from blood smear test (Target variable)

View the data

```
#View(data)
```

Remove columns with Missing values

```
data <-na.omit(data)
```

Check Columns with Missing values in the data set

```
library(kableExtra)
missing_values <- colSums(is.na(data))
kable(missing_values)
```

	x
hv024	0
shcounty	0
hv227	0
hv228	0
hml1	0
hml2	0
hc57	0
hc61	0
hml12	0
hml32a	0
hml32b	0
hml32c	0
hml32	0

Load the following libraries for malaria modeling

```
library(summarytools)
library(sjmisc)
```

Frequency Tables

```
frq(data, shcounty)
```

```
county (shcounty) <numeric>
# total N=3280 valid N=3280 mean=567.25 sd=190.92

Value |        Label |  N | Raw % | Valid % | Cum. %
----------------------------------------------------------
  101 |       nairobi | 27 |  0.82 |    0.82 |   0.82
  201 |     nyandarua | 34 |  1.04 |    1.04 |   1.86
```

```
 202 |          nyeri |  36 | 1.10 |  1.10 |   2.96
 203 |      kirinyaga |  28 | 0.85 |  0.85 |   3.81
 204 |        muranga |  47 | 1.43 |  1.43 |   5.24
 205 |         kiambu |  30 | 0.91 |  0.91 |   6.16
 301 |        mombasa |  69 | 2.10 |  2.10 |   8.26
 302 |          kwale | 113 | 3.45 |  3.45 |  11.71
 303 |         kilifi |  99 | 3.02 |  3.02 |  14.73
 304 |     tana river |  59 | 1.80 |  1.80 |  16.52
 305 |           lamu |  29 | 0.88 |  0.88 |  17.41
 306 |   taita taveta |  57 | 1.74 |  1.74 |  19.15
 401 |       marsabit |  76 | 2.32 |  2.32 |  21.46
 402 |         isiolo |  55 | 1.68 |  1.68 |  23.14
 403 |           meru |  74 | 2.26 |  2.26 |  25.40
 404 |        tharaka |  66 | 2.01 |  2.01 |  27.41
 405 |           embu |  72 | 2.20 |  2.20 |  29.60
 406 |          kitui |  40 | 1.22 |  1.22 |  30.82
 407 |       machakos |  30 | 0.91 |  0.91 |  31.74
 408 |        makueni |  19 | 0.58 |  0.58 |  32.32
 501 |        garissa |  50 | 1.52 |  1.52 |  33.84
 502 |          wajir |  68 | 2.07 |  2.07 |  35.91
 503 |        mandera |  40 | 1.22 |  1.22 |  37.13
 601 |          siaya | 135 | 4.12 |  4.12 |  41.25
 602 |         kisumu | 107 | 3.26 |  3.26 |  44.51
 603 |         migori | 195 | 5.95 |  5.95 |  50.46
 604 |       homa bay | 222 | 6.77 |  6.77 |  57.23
 605 |          kisii |  26 | 0.79 |  0.79 |  58.02
 606 |        nyamira |  42 | 1.28 |  1.28 |  59.30
 701 |        turkana |  45 | 1.37 |  1.37 |  60.67
 702 |     west pokot | 142 | 4.33 |  4.33 |  65.00
 703 |        samburu |  51 | 1.55 |  1.55 |  66.55
 704 |     trans-nzoia |  47 | 1.43 |  1.43 |  67.99
 705 |        baringo |  39 | 1.19 |  1.19 |  69.18
 706 |    uasin gishu |  40 | 1.22 |  1.22 |  70.40
 707 | elgeyo marakwet |  62 | 1.89 |  1.89 |  72.29
 708 |          nandi |  49 | 1.49 |  1.49 |  73.78
 709 |       laikipia |  34 | 1.04 |  1.04 |  74.82
 710 |         nakuru |  45 | 1.37 |  1.37 |  76.19
 711 |          narok |  37 | 1.13 |  1.13 |  77.32
 712 |        kajiado |  60 | 1.83 |  1.83 |  79.15
 713 |        kericho |  19 | 0.58 |  0.58 |  79.73
 714 |          bomet |  64 | 1.95 |  1.95 |  81.68
 801 |       kakamega | 167 | 5.09 |  5.09 |  86.77
 802 |         vihiga | 101 | 3.08 |  3.08 |  89.85
 803 |        bungoma | 195 | 5.95 |  5.95 |  95.79
 804 |          busia | 138 | 4.21 |  4.21 | 100.00
<NA> |           <NA> |   0 | 0.00 |  <NA> |  <NA>
```

frq(data)

```
region (hv024) <numeric>
# total N=3280 valid N=3280 mean=5.08 sd=2.44

Value |          Label |  N | Raw % | Valid % | Cum. %
-------------------------------------------------------
```

```
     1 |          coast | 426 | 12.99 |  12.99 |  12.99
     2 | north eastern | 158 |  4.82 |   4.82 |  17.80
     3 |        eastern | 432 | 13.17 |  13.17 |  30.98
     4 |        central | 175 |  5.34 |   5.34 |  36.31
     5 |    rift valley | 734 | 22.38 |  22.38 |  58.69
     7 |        western | 601 | 18.32 |  18.32 |  77.01
     8 |         nyanza | 727 | 22.16 |  22.16 |  99.18
     9 |        nairobi |  27 |  0.82 |   0.82 | 100.00
  <NA> |          <NA> |   0 |  0.00 |   <NA> |   <NA>
```

county (shcounty) <numeric>
total N=3280 valid N=3280 mean=567.25 sd=190.92

```
Value |            Label |   N | Raw % | Valid % | Cum. %
-------------------------------------------------------------
  101 |          nairobi |  27 |  0.82 |    0.82 |   0.82
  201 |        nyandarua |  34 |  1.04 |    1.04 |   1.86
  202 |            nyeri |  36 |  1.10 |    1.10 |   2.96
  203 |        kirinyaga |  28 |  0.85 |    0.85 |   3.81
  204 |          muranga |  47 |  1.43 |    1.43 |   5.24
  205 |           kiambu |  30 |  0.91 |    0.91 |   6.16
  301 |          mombasa |  69 |  2.10 |    2.10 |   8.26
  302 |            kwale | 113 |  3.45 |    3.45 |  11.71
  303 |           kilifi |  99 |  3.02 |    3.02 |  14.73
  304 |        tana river |  59 |  1.80 |    1.80 |  16.52
  305 |             lamu |  29 |  0.88 |    0.88 |  17.41
  306 |     taita taveta |  57 |  1.74 |    1.74 |  19.15
  401 |         marsabit |  76 |  2.32 |    2.32 |  21.46
  402 |           isiolo |  55 |  1.68 |    1.68 |  23.14
  403 |             meru |  74 |  2.26 |    2.26 |  25.40
  404 |          tharaka |  66 |  2.01 |    2.01 |  27.41
  405 |             embu |  72 |  2.20 |    2.20 |  29.60
  406 |            kitui |  40 |  1.22 |    1.22 |  30.82
  407 |         machakos |  30 |  0.91 |    0.91 |  31.74
  408 |          makueni |  19 |  0.58 |    0.58 |  32.32
  501 |          garissa |  50 |  1.52 |    1.52 |  33.84
  502 |            wajir |  68 |  2.07 |    2.07 |  35.91
  503 |          mandera |  40 |  1.22 |    1.22 |  37.13
  601 |            siaya | 135 |  4.12 |    4.12 |  41.25
  602 |           kisumu | 107 |  3.26 |    3.26 |  44.51
  603 |           migori | 195 |  5.95 |    5.95 |  50.46
  604 |         homa bay | 222 |  6.77 |    6.77 |  57.23
  605 |            kisii |  26 |  0.79 |    0.79 |  58.02
  606 |          nyamira |  42 |  1.28 |    1.28 |  59.30
  701 |          turkana |  45 |  1.37 |    1.37 |  60.67
  702 |        west pokot | 142 |  4.33 |    4.33 |  65.00
  703 |          samburu |  51 |  1.55 |    1.55 |  66.55
  704 |       trans-nzoia |  47 |  1.43 |    1.43 |  67.99
  705 |          baringo |  39 |  1.19 |    1.19 |  69.18
  706 |       uasin gishu |  40 |  1.22 |    1.22 |  70.40
  707 | elgeyo marakwet |  62 |  1.89 |    1.89 |  72.29
  708 |            nandi |  49 |  1.49 |    1.49 |  73.78
  709 |          laikipia |  34 |  1.04 |    1.04 |  74.82
  710 |           nakuru |  45 |  1.37 |    1.37 |  76.19
```

```
711 |       narok |  37 | 1.13 |   1.13 |  77.32
712 |     kajiado |  60 | 1.83 |   1.83 |  79.15
713 |     kericho |  19 | 0.58 |   0.58 |  79.73
714 |       bomet |  64 | 1.95 |   1.95 |  81.68
801 |    kakamega | 167 | 5.09 |   5.09 |  86.77
802 |      vihiga | 101 | 3.08 |   3.08 |  89.85
803 |     bungoma | 195 | 5.95 |   5.95 |  95.79
804 |       busia | 138 | 4.21 |   4.21 | 100.00
<NA> |       <NA> |   0 | 0.00 |   <NA> |   <NA>
```

has mosquito bed net for sleeping (hv227) <numeric>
total N=3280 valid N=3280 mean=0.73 sd=0.44

```
Value | Label |    N | Raw % | Valid % | Cum. %
-------------------------------------------------
    0 |    no |  883 | 26.92 |   26.92 |  26.92
    1 |   yes | 2397 | 73.08 |   73.08 | 100.00
 <NA> |  <NA> |    0 |  0.00 |    <NA> |   <NA>
```

children under 5 slept under mosquito bed net last night (hv228) <numeric>
total N=3280 valid N=3280 mean=1.52 sd=1.00

```
Value |               Label |    N | Raw % | Valid % | Cum. %
----------------------------------------------------------------
    0 |                  no |  359 | 10.95 |   10.95 |  10.95
    1 |        all children | 1724 | 52.56 |   52.56 |  63.51
    2 |       some children |  314 |  9.57 |    9.57 |  73.08
    3 | no net in household |  883 | 26.92 |   26.92 | 100.00
 <NA> |                <NA> |    0 |  0.00 |    <NA> |   <NA>
```

number of mosquito bed nets (hml1) <numeric>
total N=3280 valid N=3280 mean=1.54 sd=1.36

```
Value |      Label |   N | Raw % | Valid % | Cum. %
------------------------------------------------------
    0 |          0 | 883 | 26.92 |   26.92 |  26.92
    1 |          1 | 827 | 25.21 |   25.21 |  52.13
    2 |          2 | 891 | 27.16 |   27.16 |  79.30
    3 |          3 | 440 | 13.41 |   13.41 |  92.71
    4 |          4 | 143 |  4.36 |    4.36 |  97.07
    5 |          5 |  49 |  1.49 |    1.49 |  98.57
    6 |          6 |  29 |  0.88 |    0.88 |  99.45
    7 |          7 |  18 |  0.55 |    0.55 | 100.00
   98 | don't know |   0 |  0.00 |    0.00 | 100.00
 <NA> |       <NA> |   0 |  0.00 |    <NA> |   <NA>
```

number of children under mosquito bed net previous night (hml2) <numeric>
total N=3280 valid N=3280 mean=0.87 sd=0.83

```
Value |    N | Raw % | Valid % | Cum. %
------------------------------------------
    0 | 1242 | 37.87 |   37.87 |  37.87
    1 | 1339 | 40.82 |   40.82 |  78.69
    2 |  602 | 18.35 |   18.35 |  97.04
```

```
    3 |     84 |  2.56 |     2.56 |  99.60
    4 |     13 |  0.40 |     0.40 | 100.00
 <NA> |      0 |  0.00 |     <NA> |    <NA>
```

anemia level (hc57) <numeric>
total N=3280 valid N=3280 mean=3.18 sd=0.89

```
Value |       Label |    N | Raw % | Valid % | Cum. %
-----------------------------------------------------------
    1 |      severe |   77 |  2.35 |    2.35 |   2.35
    2 |    moderate |  840 | 25.61 |   25.61 |  27.96
    3 |        mild |  781 | 23.81 |   23.81 |  51.77
    4 | not anemic  | 1582 | 48.23 |   48.23 | 100.00
 <NA> |        <NA> |    0 |  0.00 |    <NA> |    <NA>
```

mother's highest educational level (hc61) <numeric>
total N=3280 valid N=3280 mean=1.31 sd=0.86

```
Value |         Label |    N | Raw % | Valid % | Cum. %
------------------------------------------------------------
    0 | no education  |  535 | 16.31 |   16.31 |  16.31
    1 |       primary | 1517 | 46.25 |   46.25 |  62.56
    2 |     secondary |  892 | 27.20 |   27.20 |  89.76
    3 |        higher |  336 | 10.24 |   10.24 | 100.00
    8 |    don't know |    0 |  0.00 |    0.00 | 100.00
 <NA> |          <NA> |    0 |  0.00 |    <NA> |    <NA>
```

type of mosquito bed net(s) person slept under last night (hml12) <numeric>
total N=3280 valid N=3280 mean=0.70 sd=0.78

```
Value |                                  Label |    N | Raw % | Valid % | Cum. %
-------------------------------------------------------------------------------------
    0 |                did not sleep under a net | 1411 | 43.02 |   43.02 |  43.02
    1 |                 only treated (itn) nets  | 1650 | 50.30 |   50.30 |  93.32
    2 | both treated (itn) and untreated nets    |    1 |  0.03 |    0.03 |  93.35
    3 |                     only untreated nets   |  218 |  6.65 |    6.65 | 100.00
 <NA> |                                    <NA>  |    0 |  0.00 |    <NA> |    <NA>
```

presence of species: falciparum (pf) (hml32a) <numeric>
total N=3280 valid N=3280 mean=0.04 sd=0.19

```
Value | Label |    N | Raw % | Valid % | Cum. %
-------------------------------------------------
    0 |    no | 3161 | 96.37 |   96.37 |  96.37
    1 |   yes |  119 |  3.63 |    3.63 | 100.00
 <NA> |  <NA> |    0 |  0.00 |    <NA> |    <NA>
```

presence of species: malariae (pm) (hml32b) <numeric>
total N=3280 valid N=3280 mean=0.01 sd=0.11

```
Value | Label |    N | Raw % | Valid % | Cum. %
-------------------------------------------------
    0 |    no | 3242 | 98.84 |   98.84 |  98.84
    1 |   yes |   38 |  1.16 |    1.16 | 100.00
```

```
 <NA> |  <NA> |    0 |  0.00 |    <NA> |    <NA>
```

presence of species: ovale (po) (hml32c) <numeric>
total N=3280 valid N=3280 mean=0.00 sd=0.03

```
Value | Label |    N | Raw % | Valid % | Cum. %
-------------------------------------------------
    0 |    no | 3276 | 99.88 |   99.88 |  99.88
    1 |   yes |    4 |  0.12 |    0.12 | 100.00
 <NA> |  <NA> |    0 |  0.00 |    <NA> |    <NA>
```

final result of malaria from blood smear test (hml32) <numeric>
total N=3280 valid N=3280 mean=0.04 sd=0.20

```
Value |                                 Label |    N | Raw % | Valid % | Cum. %
-------------------------------------------------------------------------------
    0 |                              negative | 3149 | 96.01 |   96.01 |  96.01
    1 |                              positive |  131 |  3.99 |    3.99 | 100.00
    6 |                    test undetermined |    0 |  0.00 |    0.00 | 100.00
    7 | sample not found in lab database |    0 |  0.00 |    0.00 | 100.00
 <NA> |                                 <NA> |    0 |  0.00 |    <NA> |    <NA>
```

Export the dataset with the required variables

```
#write.csv(data,"malaria_survey_data.csv")
```