

A-novel-face-scale.R

r439596

2022-08-30

```
# Title: qVIEWS

# Author: Jeronimo Moreno Cuesta
# Date: 4.5.2022

# Read data
data <- read.table("data_final.txt",header=T,sep="\t")
head(data)

##   study_number  sreen_date  group  sex  age  ethnicity  consent  type_of_consent  Total
## 1             1 23/07/2021     1    0   57           2         0             2         5
## 2             2 22/07/2021     1    0   26           2         1             1         5
## 3             3 25/07/2021     1    0   73           2         1             1        19
## 4             4 23/07/2021     1    0   63           3         0             2         7
## 5             5 02/08/2021     1    0   70           4         0             2        12
## 6             6 02/08/2021     1    1   55           3         0             2         8
##   rr  sat  air_oxygen  sbp  hr  avpu  temp  assessor1  assessor2  itu_admission
## 1 24  89           1 172 100     1 36.2           2         NA             1
## 2 24  95           1 114  78     1 37.0           1         NA             0
## 3 26  90           1  80 118     2 34.0           3         NA             1
## 4 26  98           1 139 100     1 39.0           2         NA             0
## 5 26  95           1  87 110     1 38.0           1         2             0
## 6 26  91           1 150  81     1 37.0           2         2             1
##   time_after_review  outcome  X  X.1
## 1                 0         0  NA  NA
## 2                <NA>         0  NA  NA
## 3                 0         1  NA  NA
## 4                <NA>         0  NA  NA
## 5                 1         0  NA  NA
## 6                 0         0  NA  NA

data <-data[1:398,-c(22,23)]
names(data)

##   [1] "study_number"      "sreen_date"        "group"
##   [4] "sex"              "age"               "ethnicity"
##   [7] "consent"          "type_of_consent"   "Total"
##  [10] "rr"              "sat"               "air_oxygen"
##  [13] "sbp"            "hr"                "avpu"
##  [16] "temp"           "assessor1"         "assessor2"
##  [19] "itu_admission"   "time_after_review" "outcome"

# Prepare data
str(data)
```

```
## 'data.frame': 398 obs. of 21 variables:
## $ study_number : int 1 2 3 4 5 6 7 8 9 10 ...
## $ sreen_date : chr "23/07/2021" "22/07/2021" "25/07/2021" "23/07/2021" ...
## $ group : int 1 1 1 1 1 1 0 1 1 0 ...
## $ sex : int 0 0 0 0 0 1 0 0 1 1 ...
## $ age : int 57 26 73 63 70 55 33 67 63 55 ...
## $ ethnicity : int 2 2 2 3 4 3 2 3 1 3 ...
## $ consent : int 0 1 1 0 0 0 0 1 1 0 ...
## $ type_of_consent : int 2 1 1 2 2 2 2 1 1 2 ...
## $ Total : int 5 5 19 7 12 8 0 7 5 2 ...
## $ rr : int 24 24 26 26 26 26 16 25 19 20 ...
## $ sat : int 89 95 90 98 95 91 98 96 96 96 ...
## $ air_oxygen : chr "1" "1" "1" "1" ...
## $ sbp : int 172 114 80 139 87 150 150 122 170 139 ...
## $ hr : int 100 78 118 100 110 81 88 80 95 87 ...
## $ avpu : int 1 1 2 1 1 1 1 1 1 1 ...
## $ temp : num 36.2 37 34 39 38 37 36.5 36.7 37 36.7 ...
## $ assessor1 : int 2 1 3 2 1 2 0 2 2 1 ...
## $ assessor2 : int NA NA NA NA 2 2 0 2 2 1 ...
## $ itu_admission : int 1 0 1 0 0 1 1 0 1 1 ...
## $ time_after_review: chr "0" NA "0" NA ...
## $ outcome : chr "0" "0" "1" "0" ...
```

```
Rdate1 <- strptime(as.character(data$sreen_date), "%d/%m/%Y")
data <- data.frame(data, Rdate1)
data$group <- as.factor(data$group)
data$sex <- as.factor(data$sex)
data$ethnicity <- as.factor(data$ethnicity)
data$consent <- as.factor(data$consent)
data$type_of_consent <- as.factor(data$type_of_consent)
data$air_oxygen <- as.factor(data$air_oxygen)
data$avpu <- as.factor(data$avpu)
data$itu_admission <- as.factor(data$itu_admission)
data$time_after_review <- as.factor(data$time_after_review)
data$outcome <- as.factor(data$outcome)
summary(data)
```

```
## study_number sreen_date group sex age
## Min. : 1.0 Length:398 0 :242 0:182 Min. :18.00
## 1st Qu.:100.2 Class :character 1 :154 1:216 1st Qu.:51.00
## Median :199.5 Mode :character 2 : 1 Median :63.00
## Mean :199.5 NA's: 1 Mean :60.79
## 3rd Qu.:298.8 3rd Qu.:74.00
## Max. :398.0 Max. :95.00
##
## ethnicity consent type_of_consent Total rr
## 1 : 26 0: 65 1 :294 Min. : 0.000 Min. :15.00
## 2 : 77 1:333 2 : 31 1st Qu.: 1.000 1st Qu.:18.00
## 3 :250 3 : 4 Median : 3.000 Median :20.00
## 4 : 38 4 : 14 Mean : 3.947 Mean :20.66
## NA's: 7 5 : 8 3rd Qu.: 6.000 3rd Qu.:22.00
## 6 : 42 Max. :19.000 Max. :45.00
## NA's: 5 NA's :1
## sat air_oxygen sbp hr avpu
## Min. : 9.00 0 :128 Min. : 64.0 Min. : 18.00 1:383
```

```
## 1st Qu.: 93.00 1 :222 1st Qu.:110.0 1st Qu.: 78.00 2: 10
## Median : 95.00 1% : 2 Median :120.0 Median : 89.00 3: 3
## Mean : 94.08 2 : 44 Mean :124.4 Mean : 90.58 4: 2
## 3rd Qu.: 96.00 NA's: 2 3rd Qu.:140.0 3rd Qu.:102.00
## Max. :100.00 Max. :197.0 Max. :192.00
## NA's :4 NA's :1
## temp assessor1 assessor2 itu_admission
## Min. :34.00 Min. :0.000 Min. :0.000 0 :204
## 1st Qu.:36.20 1st Qu.:1.000 1st Qu.:1.000 1 :180
## Median :36.50 Median :1.000 Median :1.000 NA's: 14
## Mean :36.69 Mean :1.367 Mean :1.372
## 3rd Qu.:37.00 3rd Qu.:2.000 3rd Qu.:2.000
## Max. :63.20 Max. :3.000 Max. :3.000
## NA's :3 NA's :43
## time_after_review outcome Rdate1
## : 2 : 1 Min. :2021-07-15 00:00:00.00
## 0 : 93 0 :316 1st Qu.:2022-01-04 00:00:00.00
## 1 :168 1 : 52 Median :2022-03-31 00:00:00.00
## N/A : 1 sill : 1 Mean :2022-03-04 12:19:56.98
## NA's:134 still: 28 3rd Qu.:2022-05-17 00:00:00.00
## Max. :2022-06-27 00:00:00.00
## NA's :1
```

```
data$air_oxygen[data$air_oxygen=="1%"] <- 1
summary(data)
```

```
## study_number sreen_date group sex age
## Min. : 1.0 Length:398 0 :242 0:182 Min. :18.00
## 1st Qu.:100.2 Class :character 1 :154 1:216 1st Qu.:51.00
## Median :199.5 Mode :character 2 : 1 Median :63.00
## Mean :199.5 NA's: 1 Mean :60.79
## 3rd Qu.:298.8 3rd Qu.:74.00
## Max. :398.0 Max. :95.00
##
## ethnicity consent type_of_consent Total rr
## 1 : 26 0: 65 1 :294 Min. : 0.000 Min. :15.00
## 2 : 77 1:333 2 : 31 1st Qu.: 1.000 1st Qu.:18.00
## 3 :250 3 : 4 Median : 3.000 Median :20.00
## 4 : 38 4 : 14 Mean : 3.947 Mean :20.66
## NA's: 7 5 : 8 3rd Qu.: 6.000 3rd Qu.:22.00
## 6 : 42 Max. :19.000 Max. :45.00
## NA's: 5 NA's :1
## sat air_oxygen sbp hr avpu
## Min. : 9.00 0 :128 Min. : 64.0 Min. : 18.00 1:383
## 1st Qu.: 93.00 1 :224 1st Qu.:110.0 1st Qu.: 78.00 2: 10
## Median : 95.00 1% : 0 Median :120.0 Median : 89.00 3: 3
## Mean : 94.08 2 : 44 Mean :124.4 Mean : 90.58 4: 2
## 3rd Qu.: 96.00 NA's: 2 3rd Qu.:140.0 3rd Qu.:102.00
## Max. :100.00 Max. :197.0 Max. :192.00
## NA's :4 NA's :1
## temp assessor1 assessor2 itu_admission
## Min. :34.00 Min. :0.000 Min. :0.000 0 :204
## 1st Qu.:36.20 1st Qu.:1.000 1st Qu.:1.000 1 :180
## Median :36.50 Median :1.000 Median :1.000 NA's: 14
## Mean :36.69 Mean :1.367 Mean :1.372
```

```

## 3rd Qu.:37.00 3rd Qu.:2.000 3rd Qu.:2.000
## Max. :63.20 Max. :3.000 Max. :3.000
## NA's :3 NA's :43
## time_after_review outcome Rdate1
## : 2 : 1 Min. :2021-07-15 00:00:00.00
## 0 : 93 0 :316 1st Qu.:2022-01-04 00:00:00.00
## 1 :168 1 : 52 Median :2022-03-31 00:00:00.00
## N/A : 1 sill : 1 Mean :2022-03-04 12:19:56.98
## NA's:134 still: 28 3rd Qu.:2022-05-17 00:00:00.00
## Max. :2022-06-27 00:00:00.00
## NA's :1

data_consent <- data[data$consent=="1",]

# Inter-rater variability
class(data_consent$assessor1)

## [1] "integer"

class(data_consent$assessor2)

## [1] "integer"

db_agree <- as.data.frame (cbind(data_consent$assessor1,data_consent$assessor2))
library(irr)

## Loading required package: lpSolve
library(psych)
cohen.kappa(db_agree)# we use weighted estimates for nominal scales such as assessor

## Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels)
##
## Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundaries
## lower estimate upper
## unweighted kappa 0.93 0.96 0.99
## weighted kappa 0.98 0.98 0.98
##
## Number of subjects = 310

# Demography high-risk patients
data_consent_highrisk<- data_consent[data_consent$Total >= 5,]
summary(data_consent_highrisk)

## study_number sreen_date group sex age ethnicity
## Min. : 2.00 Length:120 0: 2 0:47 Min. :19.00 1 :10
## 1st Qu.: 86.25 Class :character 1:118 1:73 1st Qu.:51.00 2 :24
## Median :228.00 Mode :character 2: 0 Median :64.00 3 :74
## Mean :210.98 Mean :61.53 4 :10
## 3rd Qu.:325.50 3rd Qu.:74.00 NA's: 2
## Max. :396.00 Max. :95.00
##
## consent type_of_consent Total rr sat
## 0: 0 1:98 Min. : 5.000 Min. :16.00 Min. :79.00
## 1:120 2: 0 1st Qu.: 5.000 1st Qu.:20.00 1st Qu.:93.00
## 3: 0 Median : 7.000 Median :24.00 Median :95.00
## 4: 1 Mean : 7.225 Mean :23.71 Mean :93.77
## 5: 0 3rd Qu.: 9.000 3rd Qu.:26.00 3rd Qu.:96.00

```

```
##           6:21           Max.   :19.000   Max.   :45.00   Max.   :99.00
##                                     NA's   :1       NA's   :1
##  air_oxygen      sbp           hr           avpu           temp
## 0   : 5      Min.   : 64.0   Min.   : 64.0   1:115   Min.   :34.00
## 1   :87     1st Qu.:103.8   1st Qu.: 88.0   2: 5     1st Qu.:36.40
## 1%  : 0     Median :118.0   Median :100.0   3: 0     Median :36.50
## 2   :27     Mean   :119.8   Mean   :101.1   4: 0     Mean   :37.05
## NA's: 1     3rd Qu.:137.0   3rd Qu.:112.0           3rd Qu.:37.00
##           Max.   :197.0   Max.   :144.0           Max.   :63.20
##                                     NA's   :1
##  assessor1      assessor2      itu_admission time_after_review outcome
## Min.   :0.000   Min.   :0.000   0   :68           : 1           : 1
## 1st Qu.:1.000   1st Qu.:1.000   1   :49           0   :34           0   :85
## Median :2.000   Median :2.000   NA's: 3           1   :44           1   :23
## Mean   :1.617   Mean   :1.613           N/A : 0           sill : 1
## 3rd Qu.:2.000   3rd Qu.:2.000           NA's:41          still:10
## Max.   :3.000   Max.   :3.000
##           NA's   :14
##      Rdate1
## Min.   :2021-07-15 00:00:00.00
## 1st Qu.:2021-12-09 00:00:00.00
## Median :2022-04-14 00:00:00.00
## Mean   :2022-03-06 18:21:10.58
## 3rd Qu.:2022-05-27 00:00:00.00
## Max.   :2022-06-25 00:00:00.00
## NA's   :1
```

```
sd(data_consent_highrisk$age)
```

```
## [1] 17.00269
```

```
sd(data_consent_highrisk$Total)
```

```
## [1] 2.342384
```

```
sd(data_consent_highrisk$rr,na.rm=T)
```

```
## [1] 5.459175
```

```
sd(data_consent_highrisk$sat,na.rm=T)
```

```
## [1] 3.880497
```

```
sd(data_consent_highrisk$sbp,na.rm=T)
```

```
## [1] 25.82192
```

```
sd(data_consent_highrisk$hr,na.rm=T)
```

```
## [1] 18.75389
```

```
sd(data_consent_highrisk$temp,na.rm=T)
```

```
## [1] 2.654693
```

```
sd(data_consent_highrisk$assessor1,na.rm=T)
```

```
## [1] 0.7467952
```

```
air_oxygen_factor <- as.factor(data_consent_highrisk$air_oxygen)
summary(air_oxygen_factor)
```

```
##      0      1      1%      2 NA's
##      5     87      0     27      1
```

```
avpu_numeric <- as.numeric(data_consent_highrisk$avpu)
summary(avpu_numeric)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      1.000  1.000   1.000   1.042  1.000   2.000
```

```
sd(avpu_numeric,na.rm=T)
```

```
## [1] 0.2006642
```

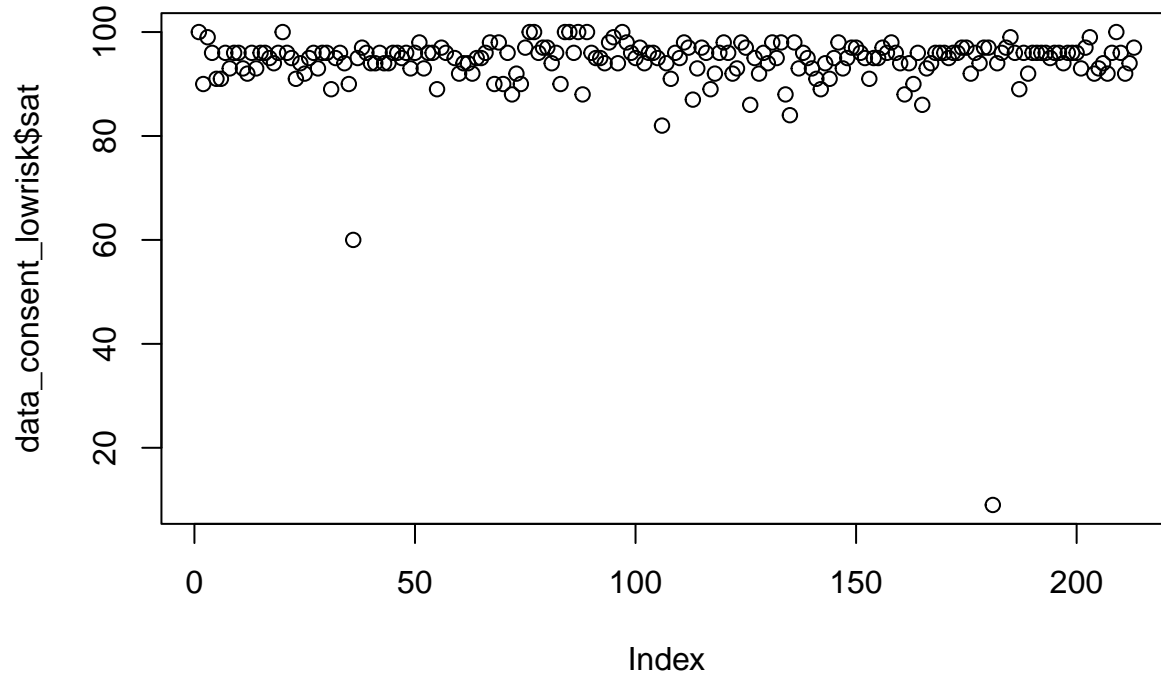
```
# Demography low-risk patients
```

```
data_consent_lowrisk<- data_consent[data_consent$Total <= 4,]
summary(data_consent_lowrisk)
```

```
##      study_number      sreen_date      group      sex      age
##      Min.       : 13.0      Length:213      0      :208      0: 96      Min.       :20.00
##      1st Qu.:158.0      Class :character      1      : 3      1:117      1st Qu.:51.00
##      Median :235.0      Mode  :character      2      : 1              Median :62.00
##      Mean      :227.9              NA's: 1              Mean      :61.24
##      3rd Qu.:306.0              3rd Qu.:74.00
##      Max.       :398.0              Max.       :94.00
##
##      ethnicity  consent type_of_consent      Total      rr
##      1      : 14      0: 0      1:191      Min.       :0.000      Min.       :15.00
##      2      : 38      1:213      2: 0      1st Qu.:1.000      1st Qu.:18.00
##      3      :144              3: 1      Median :2.000      Median :18.00
##      4      : 14              4: 0      Mean      :1.831      Mean      :18.64
##      NA's: 3              5: 0      3rd Qu.:3.000      3rd Qu.:20.00
##              6: 21      Max.       :4.000      Max.       :24.00
##
##      sat      air_oxygen      sbp      hr      avpu
##      Min.       : 9.00      0      :114      Min.       : 80.0      Min.       : 18.0      1:210
##      1st Qu.: 93.00      1      : 84      1st Qu.:110.0      1st Qu.: 72.0      2: 1
##      Median : 96.00      1%      : 0      Median :120.0      Median : 81.0      3: 0
##      Mean      : 94.16      2      : 14      Mean      :125.3      Mean      : 82.8      4: 2
##      3rd Qu.: 96.00      NA's: 1      3rd Qu.:139.0      3rd Qu.: 90.0
##      Max.       :100.00              Max.       :184.0      Max.       :130.0
##      NA's       :1
##      temp      assessor1      assessor2      itu_admission time_after_review
##      Min.       :36.00      Min.       :0.000      Min.       :0.00      0      :104              : 1
##      1st Qu.:36.20      1st Qu.:1.000      1st Qu.:1.00      1      :102              0      :54
##      Median :36.50      Median :1.000      Median :1.00      NA's: 7              1      :98
##      Mean      :36.49      Mean      :1.246      Mean      :1.26              N/A : 1
##      3rd Qu.:36.60      3rd Qu.:2.000      3rd Qu.:2.00              NA's:59
##      Max.       :39.00      Max.       :3.000      Max.       :3.00
##      NA's       :2      NA's       :9
##      outcome      Rdate1
##      : 0      Min.       :2021-09-13 00:00:00.00
##      0      :178      1st Qu.:2022-02-23 00:00:00.00
##      1      : 17      Median :2022-04-18 00:00:00.00
```

```
## sill : 0    Mean   :2022-03-31 08:47:19.44
## still: 18   3rd Qu.:2022-05-21 00:00:00.00
##          Max.    :2022-06-27 00:00:00.00
##
```

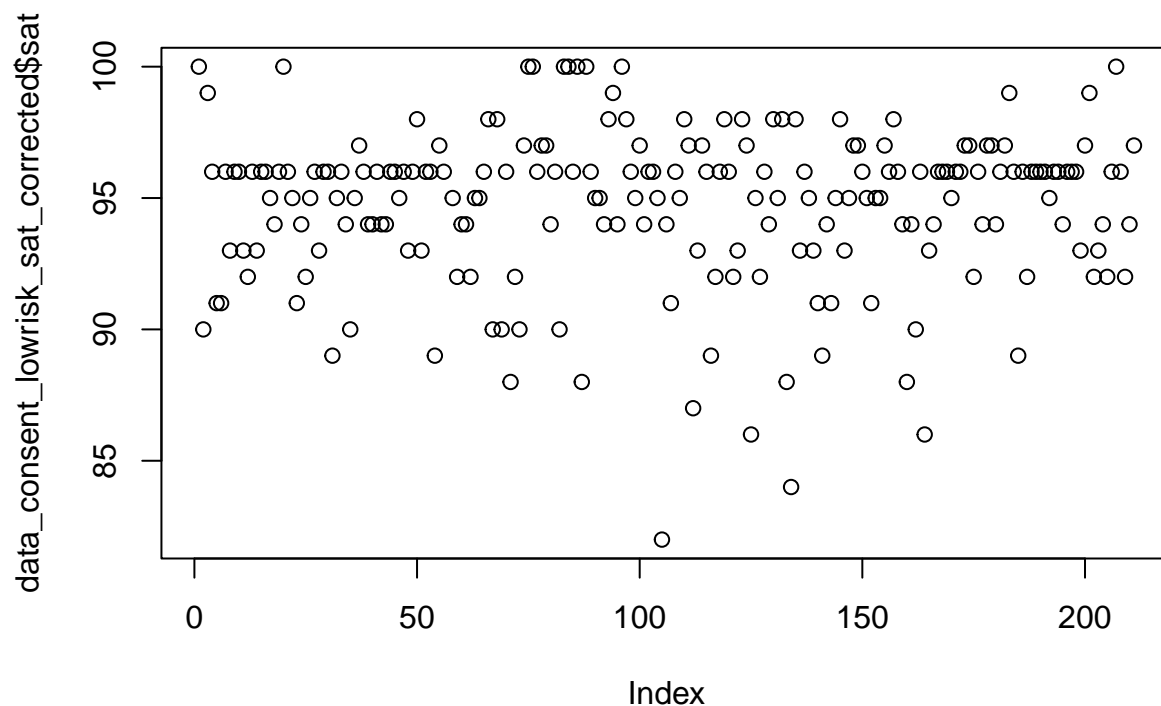
```
plot(data_consent_lowrisk$sat) # There are outliers / errors
```



```
which(data_consent_lowrisk$sat < 70)
```

```
## [1] 36 181
```

```
data_consent_lowrisk_sat_corrected <- data_consent_lowrisk[-c(36,181),]
plot(data_consent_lowrisk_sat_corrected$sat) # We checked rows 36 and 86 removed
```



```
summary(data_consented_lowrisk_sat_corrected$sat)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      82.00  93.00   96.00   94.72  96.00   100.00         1
```

```
sd(data_consented_lowrisk_sat_corrected$sat,na.rm=T)
```

```
## [1] 3.037234
```

```
sd(data_consented_lowrisk$age)
```

```
## [1] 16.97825
```

```
sd(data_consented_lowrisk$Total)
```

```
## [1] 1.414073
```

```
sd(data_consented_lowrisk$rr,na.rm=T)
```

```
## [1] 1.493512
```

```
sd(data_consented_lowrisk$sbp,na.rm=T)
```

```
## [1] 19.94701
```

```
sd(data_consented_lowrisk$hr,na.rm=T)
```

```
## [1] 15.63766
```

```
sd(data_consented_lowrisk$temp,na.rm=T)
```

```
## [1] 0.3832165
```

```
sd(data_consented_lowrisk$assessor1,na.rm=T)
```

```
## [1] 0.7341494
```



```
air_oxygen_factor_lr <- as.factor(data_consent_lowrisk$air_oxygen)
summary(air_oxygen_factor_lr)
```

```
##      0      1      1%      2 NA's
## 114     84      0     14      1
```

```
avpu_numeric_lr <- as.numeric(data_consent_lowrisk$avpu)
summary(avpu_numeric_lr)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  1.000   1.000   1.000   1.033   1.000   4.000
```

```
sd(avpu_numeric_lr,na.rm=T)
```

```
## [1] 0.2975525
```

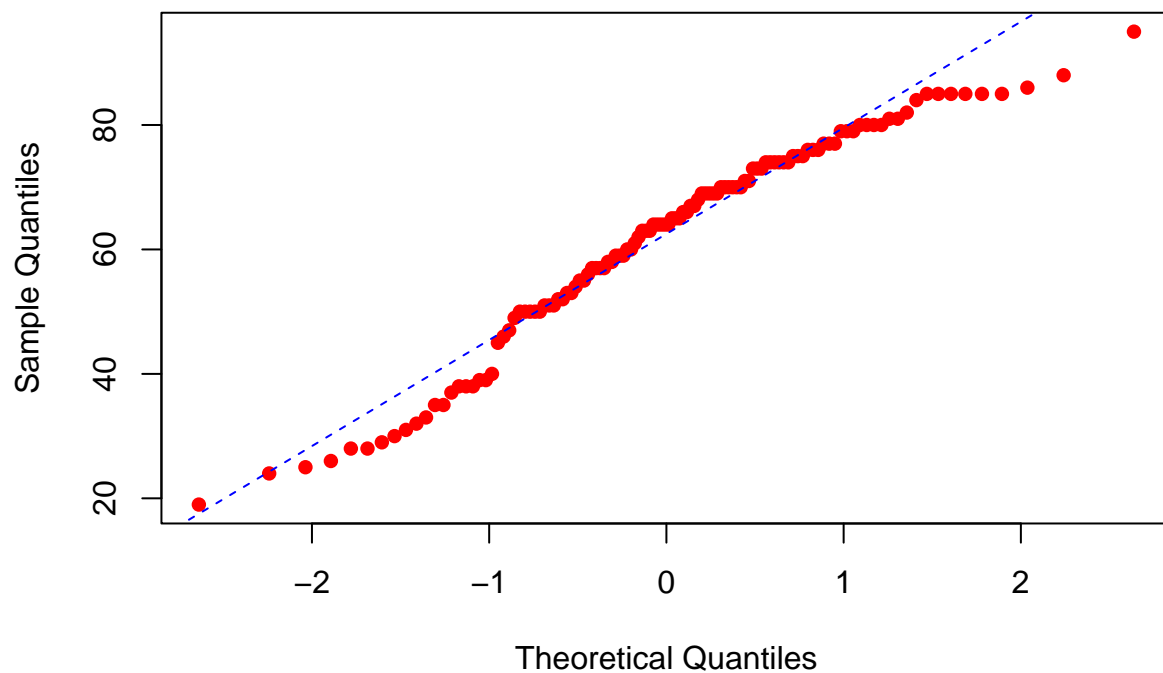
```
# Comparison demography high-risk and low-risk: FIRST ASSESSING NORMALITY OF VARIABLES
```

```
##Age
```

```
qqnorm(data_consent_highrisk$age,pch=16,col="red")
```

```
qqline(data_consent_highrisk$age,lty=2,col="blue")
```

Normal Q-Q Plot



```
shapiro.test(data_consent_highrisk$age)# Age not normal, then we do Wilcoxon
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: data_consent_highrisk$age
```

```
## W = 0.95895, p-value = 0.001042
```

```
wilcox.test(data_consent_highrisk$age,data_consent_lowrisk$age,paired=FALSE)
```

```
##
```

```
## Wilcoxon rank sum test with continuity correction
##
## data: data_consent_highrisk$age and data_consent_lowrisk$age
## W = 13038, p-value = 0.7596
## alternative hypothesis: true location shift is not equal to 0
#For table 1 then we use median and IQR
summary(data_consent_highrisk$age)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      19.00  51.00   64.00   61.53   74.00   95.00

summary(data_consent_lowrisk$age)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      20.00  51.00   62.00   61.24   74.00   94.00
##Sex: you have to create a group to compare vectors of different length
names(data_consent)

## [1] "study_number"      "screen_date"        "group"
## [4] "sex"                "age"                "ethnicity"
## [7] "consent"            "type_of_consent"    "Total"
## [10] "rr"                 "sat"                "air_oxygen"
## [13] "sbp"                "hr"                 "avpu"
## [16] "temp"               "assessor1"          "assessor2"
## [19] "itu_admission"      "time_after_review"  "outcome"
## [22] "Rdate1"

group_corrected <- ifelse(data_consent$Total >= 5,1,0) # 1 = high-risk, 0=low-risk
t1 <- table(group_corrected,data_consent$sex)
chisq.test(t1)

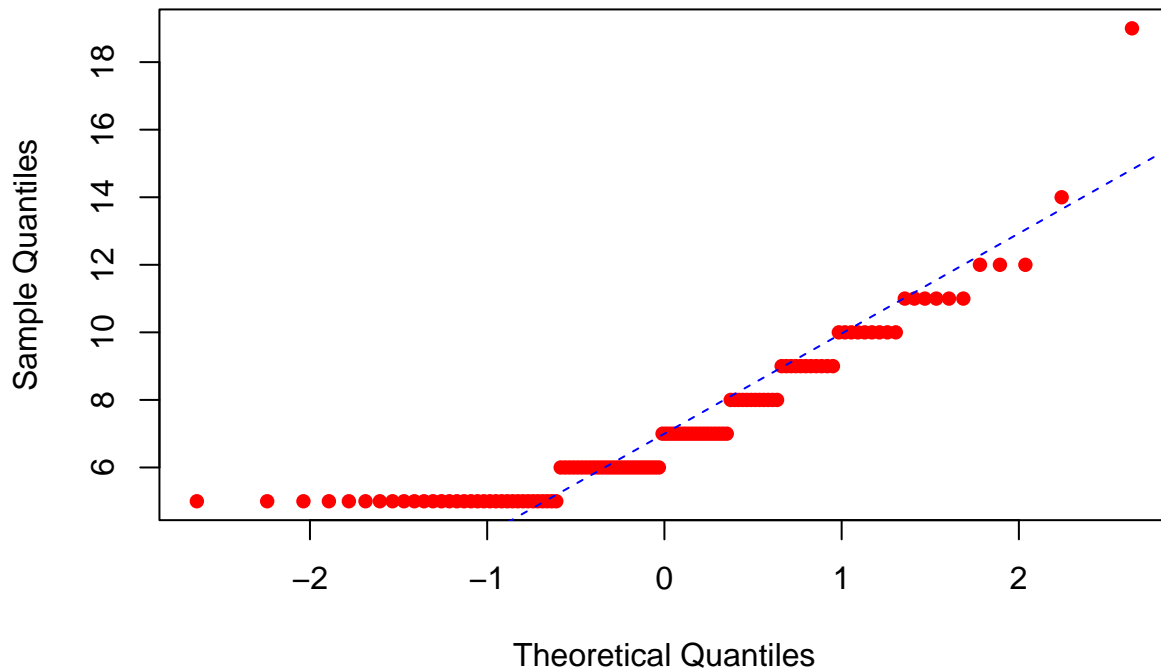
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: t1
## X-squared = 0.86422, df = 1, p-value = 0.3526
##Ethnicity
t2 <- table(group_corrected,data_consent$ethnicity);t2

##
## group_corrected    1    2    3    4
##                0  14  38 144  14
##                1  10  24  74  10

chisq.test(t2)

##
## Pearson's Chi-squared test
##
## data: t2
## X-squared = 1.2664, df = 3, p-value = 0.7371
##NEWS2 score
qqnorm(data_consent_highrisk$Total,pch=16,col="red")
qqline(data_consent_highrisk$Total,lty=2,col="blue")
```

Normal Q-Q Plot



```
shapiro.test(data_consent_highrisk$Total)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_consent_highrisk$Total
## W = 0.8362, p-value = 3.241e-10
```

```
wilcox.test(data_consent_highrisk$Total,data_consent_lowrisk$Total,paired=FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  data_consent_highrisk$Total and data_consent_lowrisk$Total
## W = 25560, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
summary(data_consent_highrisk$Total)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  5.000   5.000   7.000   7.225   9.000  19.000
```

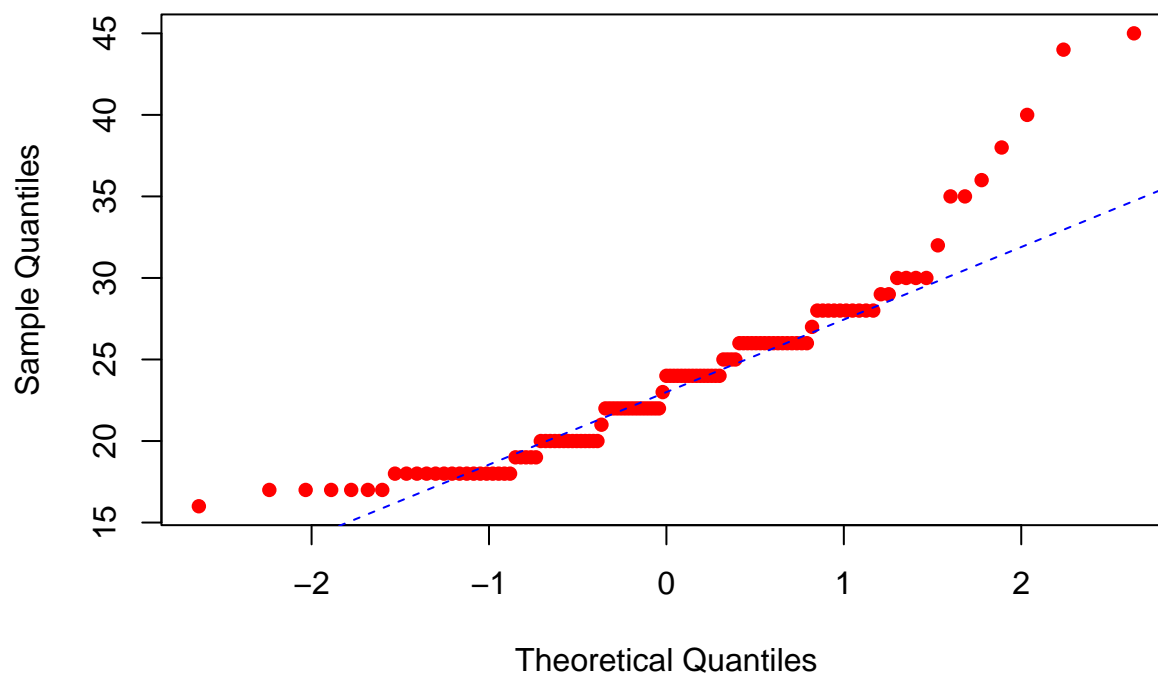
```
summary(data_consent_lowrisk$Total)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000   1.000   2.000   1.831   3.000   4.000
```

```
## Respiratory rate
```

```
qqnorm(data_consent_highrisk$rr,pch=16,col="red")
qqline(data_consent_highrisk$rr,lty=2,col="blue")
```

Normal Q-Q Plot



```
shapiro.test(data_consent_highrisk$rr)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_consent_highrisk$rr
## W = 0.88752, p-value = 5.251e-08
```

```
wilcox.test(data_consent_highrisk$rr,data_consent_lowrisk$rr,paired=FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  data_consent_highrisk$rr and data_consent_lowrisk$rr
## W = 20789, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

```
summary(data_consent_highrisk$rr)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##      16.00  20.00   24.00   23.71  26.00   45.00         1
```

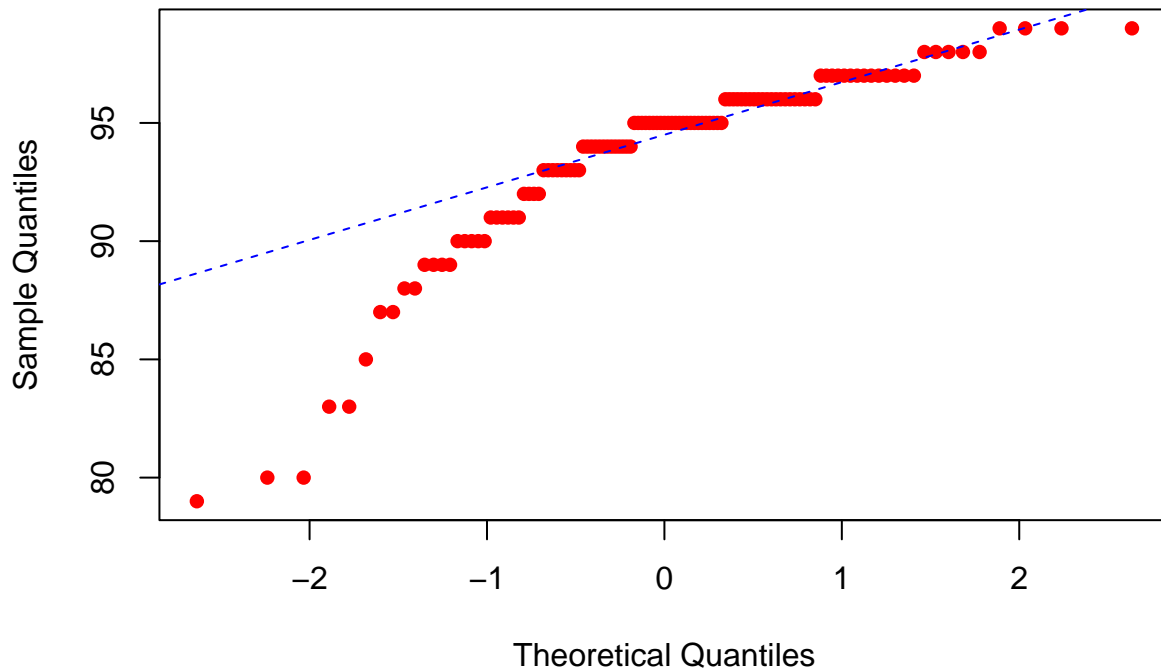
```
summary(data_consent_lowrisk$rr)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      15.00  18.00   18.00   18.64  20.00   24.00
```

```
## Oxygen saturation
```

```
qqnorm(data_consent_highrisk$sat,pch=16,col="red")
qqline(data_consent_highrisk$sat,lty=2,col="blue")
```

Normal Q-Q Plot



```
shapiro.test(data_consent_highrisk$sat)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_consent_highrisk$sat
## W = 0.83648, p-value = 3.736e-10
```

```
wilcox.test(data_consent_highrisk$sat,data_consent_lowrisk$sat,paired=FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  data_consent_highrisk$sat and data_consent_lowrisk$sat
## W = 11142, p-value = 0.07509
## alternative hypothesis: true location shift is not equal to 0
```

```
summary(data_consent_highrisk$sat)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
## 79.00   93.00   95.00   93.77   96.00   99.00         1
```

```
summary(data_consent_lowrisk$sat)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's
##   9.00   93.00   96.00   94.16   96.00  100.00         1
```

```
##Oxygen administration. There is a misterious level called "1%" that has not value!
levels(data_consent$air_oxygen)[levels(data_consent$air_oxygen)== '1%'] <- NA # to remove it
t3 <- table(group_corrected,data_consent$air_oxygen);t3
```

```
##
## group_corrected    0    1    2
```

```
##           0 114  84  14
##           1   5  87  27
```

```
chisq.test(t3)
```

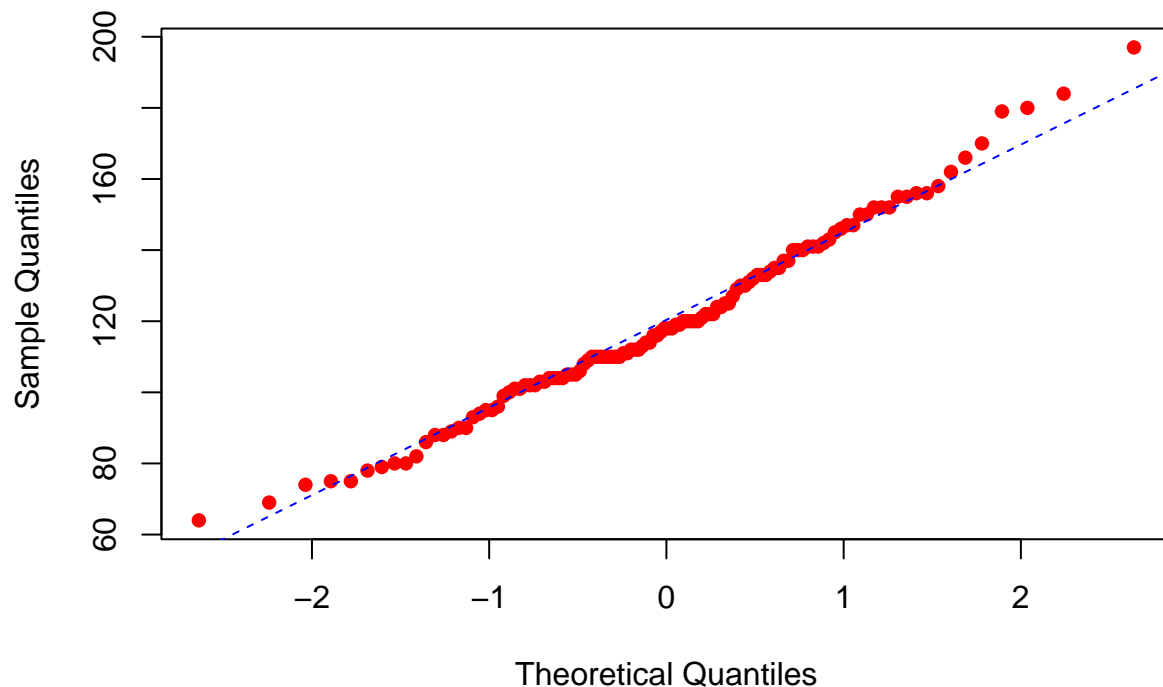
```
##
##  Pearson's Chi-squared test
##
## data:  t3
## X-squared = 84.56, df = 2, p-value < 2.2e-16
```

```
## Systolic blood pressure
```

```
qqnorm(data_consent_highrisk$sbp,pch=16,col="red")
```

```
qqline(data_consent_highrisk$sbp,lty=2,col="blue")
```

Normal Q-Q Plot



```
shapiro.test(data_consent_highrisk$sbp)# It is normal, hence t-test
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_consent_highrisk$sbp
## W = 0.98656, p-value = 0.282
```

```
var.test(data_consent_highrisk$sbp,data_consent_lowrisk$sbp)# variance are different
```

```
##
##  F test to compare two variances
##
## data:  data_consent_highrisk$sbp and data_consent_lowrisk$sbp
## F = 1.6758, num df = 119, denom df = 212, p-value = 0.00112
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
```

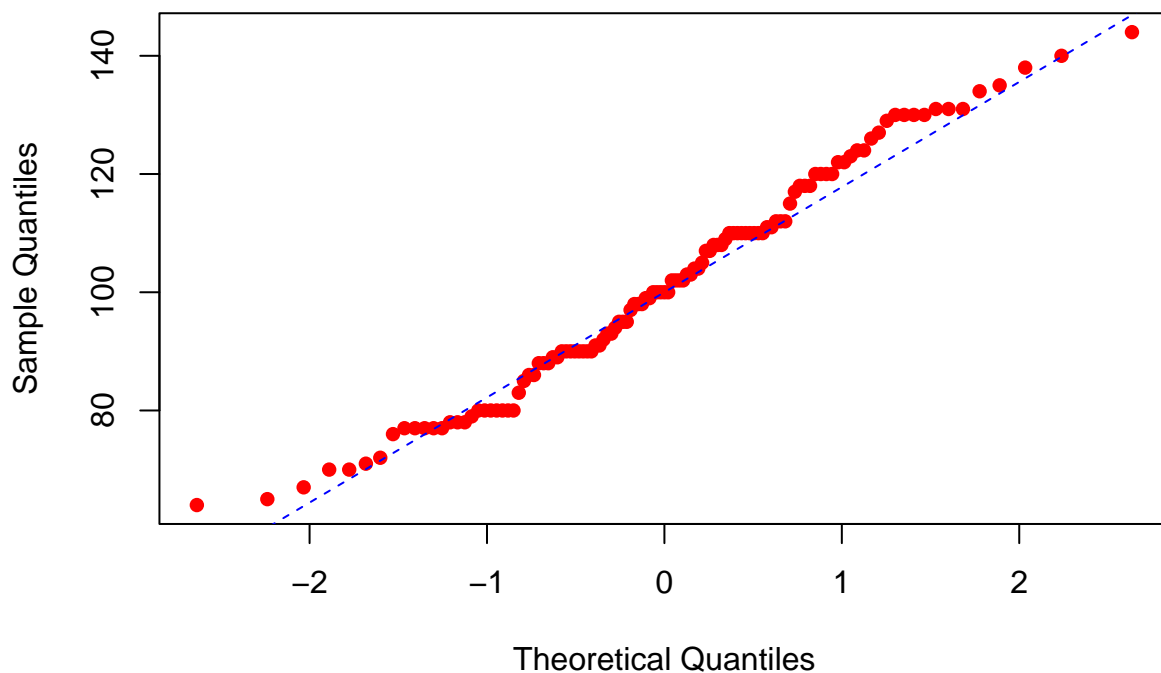
```
## 1.22723 2.32162
## sample estimates:
## ratio of variances
## 1.675796

t.test(data_consent_highrisk$sbp,data_consent_lowrisk$sbp,paired=FALSE,var.equal=FALSE,
conf.level =0.95,alternative="two.sided")

##
## Welch Two Sample t-test
##
## data: data_consent_highrisk$sbp and data_consent_lowrisk$sbp
## t = -2.008, df = 199.79, p-value = 0.04599
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -10.84426201 -0.09822625
## sample estimates:
## mean of x mean of y
## 119.7917 125.2629

##Heart rate
qqnorm(data_consent_highrisk$hr,pch=16,col="red")
qqline(data_consent_highrisk$hr,lty=2,col="blue")
```

Normal Q-Q Plot



```
shapiro.test(data_consent_highrisk$hr)

##
## Shapiro-Wilk normality test
##
## data: data_consent_highrisk$hr
## W = 0.98046, p-value = 0.08097
```

```

var.test(data_consent_highrisk$hr,data_consent_lowrisk$hr)# Var are different

##
## F test to compare two variances
##
## data: data_consent_highrisk$hr and data_consent_lowrisk$hr
## F = 1.4383, num df = 118, denom df = 212, p-value = 0.02245
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 1.052508 1.994568
## sample estimates:
## ratio of variances
## 1.438266

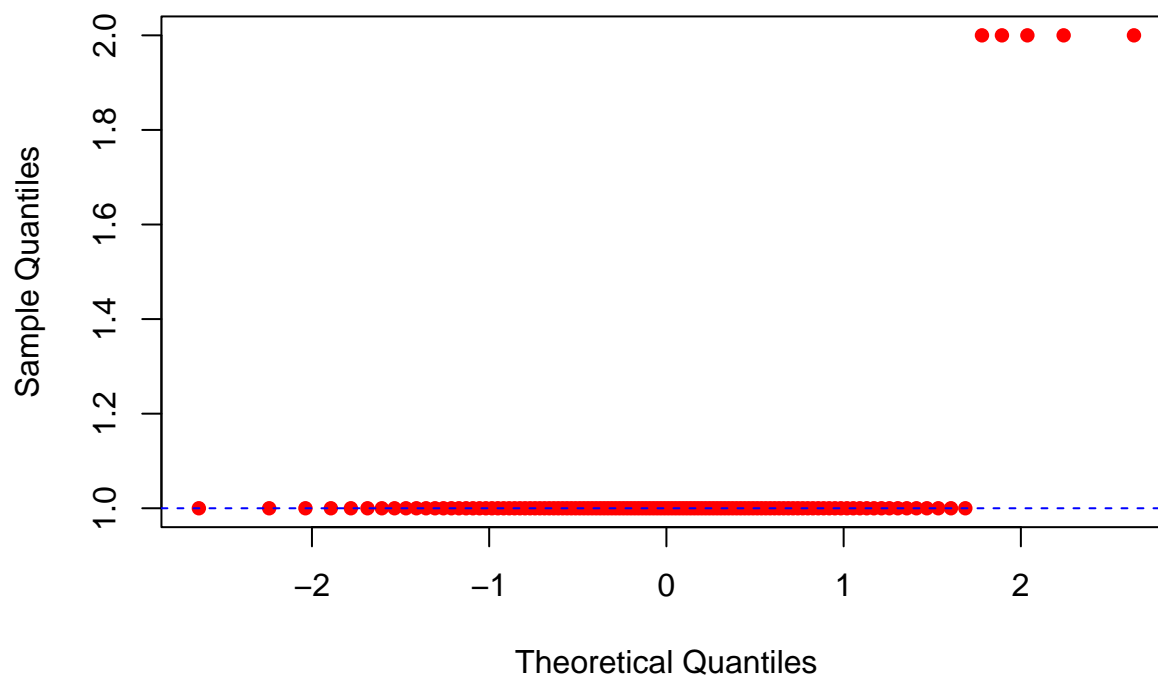
t.test(data_consent_highrisk$hr,data_consent_lowrisk$hr,paired=FALSE,var.equal=FALSE,
conf.level =0.95,alternative="two.sided")

##
## Welch Two Sample t-test
##
## data: data_consent_highrisk$hr and data_consent_lowrisk$hr
## t = 9.0369, df = 209.85, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 14.31303 22.29982
## sample estimates:
## mean of x mean of y
## 101.10924 82.80282

##AVPU
avpu_numeric_highrisk <- as.numeric(data_consent_highrisk$avpu)
avpu_numeric_lowrisk <- as.numeric(data_consent_lowrisk$avpu)
qqnorm(avpu_numeric_highrisk,pch=16,col="red")
qqline(avpu_numeric_highrisk,lty=2,col="blue")

```


Normal Q-Q Plot



```
shapiro.test(avpu_numeric_highrisk)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  avpu_numeric_highrisk
## W = 0.19971, p-value < 2.2e-16
```

```
wilcox.test(avpu_numeric_highrisk,avpu_numeric_lowrisk,paired=FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  avpu_numeric_highrisk and avpu_numeric_lowrisk
## W = 13128, p-value = 0.1209
## alternative hypothesis: true location shift is not equal to 0
```

```
summary(avpu_numeric_highrisk)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000  1.000   1.000   1.042  1.000   2.000
```

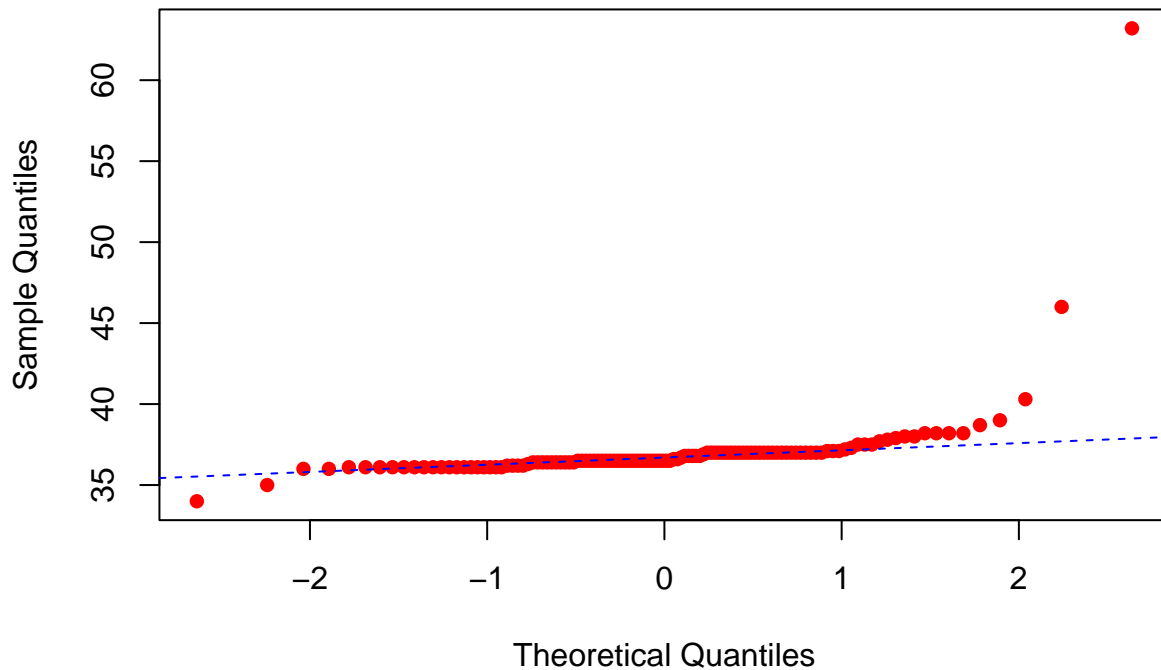
```
summary(avpu_numeric_lowrisk)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000  1.000   1.000   1.033  1.000   4.000
```

```
##Temperature
```

```
qqnorm(data_consent_highrisk$temp,pch=16,col="red")
qqline(data_consent_highrisk$temp,lty=2,col="blue")
```

Normal Q-Q Plot



```
shapiro.test(data_consent_highrisk$temp)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_consent_highrisk$temp
## W = 0.28557, p-value < 2.2e-16
```

```
wilcox.test(data_consent_highrisk$temp,data_consent_lowrisk$temp,paired=FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  data_consent_highrisk$temp and data_consent_lowrisk$temp
## W = 16010, p-value = 0.0001056
## alternative hypothesis: true location shift is not equal to 0
```

```
summary(data_consent_highrisk$temp)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  34.00  36.40   36.50   37.05  37.00   63.20
```

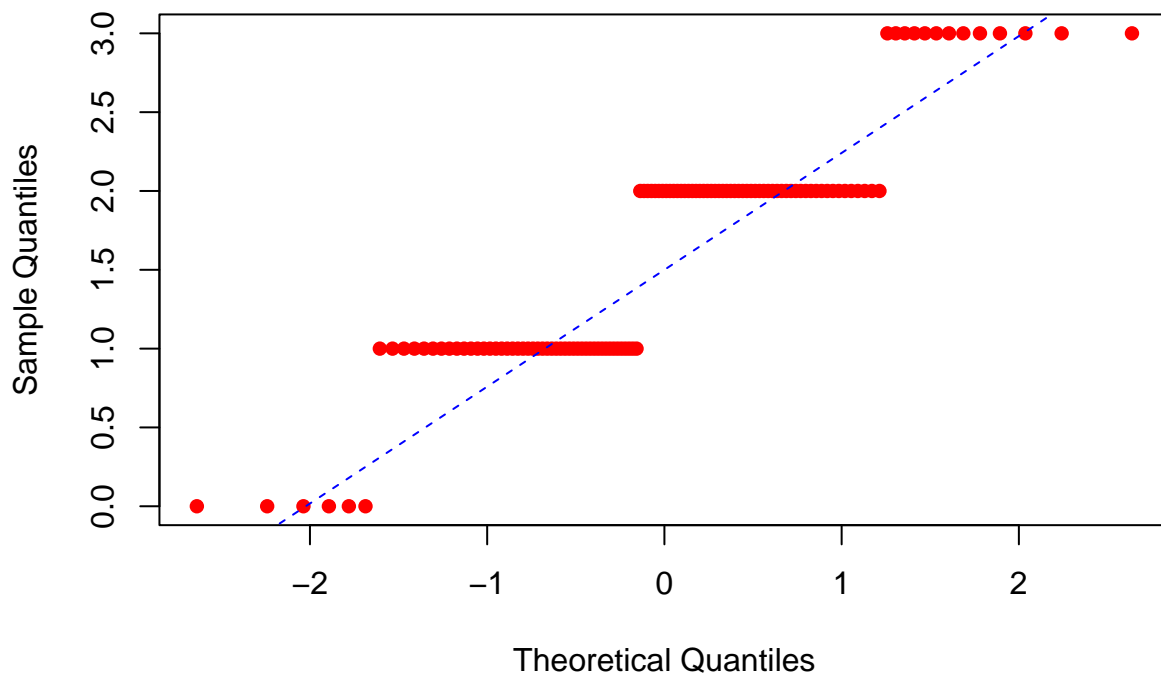
```
summary(data_consent_lowrisk$temp)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  36.00  36.20   36.50   36.49  36.60   39.00
```

```
#qVIEWS
```

```
qqnorm(data_consent_highrisk$assessor1,pch=16,col="red")
qqline(data_consent_highrisk$assessor1,lty=2,col="blue")
```

Normal Q-Q Plot



```
shapiro.test(data_consent_highrisk$assessor1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data_consent_highrisk$assessor1
## W = 0.84567, p-value = 7.483e-10
```

```
wilcox.test(data_consent_highrisk$assessor1,data_consent_lowrisk$assessor1,paired=FALSE)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  data_consent_highrisk$assessor1 and data_consent_lowrisk$assessor1
## W = 15828, p-value = 4.121e-05
## alternative hypothesis: true location shift is not equal to 0
```

```
summary(data_consent_highrisk$assessor1)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000   1.000   2.000   1.617   2.000   3.000
```

```
summary(data_consent_lowrisk$assessor1)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##  0.000   1.000   1.000   1.246   2.000   3.000     2
```

```
## Preparing for Rigde regression
```

```
#Removing NA and selecting variables
```

```
data_consent <- data[data$consent=="1",]
names(data_consent)
```

```
## [1] "study_number"      "screen_date"       "group"
```

```
## [4] "sex"          "age"          "ethnicity"
## [7] "consent"      "type_of_consent" "Total"
## [10] "rr"          "sat"          "air_oxygen"
## [13] "sbp"         "hr"           "avpu"
## [16] "temp"        "assessor1"    "assessor2"
## [19] "itu_admission" "time_after_review" "outcome"
## [22] "Rdate1"
```

```
newdata <- data_consent[,c(4:6,10:17,19)]
names(newdata)
```

```
## [1] "sex"          "age"          "ethnicity"    "rr"
## [5] "sat"          "air_oxygen"   "sbp"          "hr"
## [9] "avpu"         "temp"        "assessor1"    "itu_admission"
```

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

```
#Looking for correlation between NEWS and qVIEWS all numerical variables
names(data_consent)
```

```
## [1] "study_number" "sreen_date"   "group"
## [4] "sex"          "age"          "ethnicity"
## [7] "consent"      "type_of_consent" "Total"
## [10] "rr"          "sat"          "air_oxygen"
## [13] "sbp"         "hr"           "avpu"
## [16] "temp"        "assessor1"    "assessor2"
## [19] "itu_admission" "time_after_review" "outcome"
## [22] "Rdate1"
```

```
str(data_consent)
```

```
## 'data.frame': 333 obs. of 22 variables:
## $ study_number : int 2 3 8 9 13 14 15 16 17 18 ...
## $ sreen_date : chr "22/07/2021" "25/07/2021" "07/08/2021" "03/08/2021" ...
## $ group : Factor w/ 3 levels "0","1","2": 2 2 2 2 1 2 2 1 1 2 ...
## $ sex : Factor w/ 2 levels "0","1": 1 1 1 2 1 1 2 2 1 2 ...
## $ age : int 26 73 67 63 91 85 81 76 25 70 ...
## $ ethnicity : Factor w/ 4 levels "1","2","3","4": 2 2 3 1 3 3 3 3 2 3 ...
## $ consent : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
## $ type_of_consent : Factor w/ 6 levels "1","2","3","4",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Total : int 5 19 7 5 2 7 4 0 5 7 ...
## $ rr : int 24 26 25 19 18 24 24 15 18 24 ...
## $ sat : int 95 90 96 96 100 95 90 99 94 92 ...
## $ air_oxygen : Factor w/ 4 levels "0","1","1%", "2": 2 2 4 2 2 2 4 1 2 2 ...
## $ sbp : int 114 80 122 170 133 120 140 111 102 112 ...
## $ hr : int 78 118 80 95 70 111 78 90 88 102 ...
## $ avpu : Factor w/ 4 levels "1","2","3","4": 1 2 1 1 1 1 1 1 1 1 ...
## $ temp : num 37 34 36.7 37 37.1 37 36.4 37 37 37 ...
## $ assessor1 : int 1 3 2 2 1 1 1 1 2 3 ...
## $ assessor2 : int NA NA 2 2 NA NA NA 1 NA 3 ...
## $ itu_admission : Factor w/ 2 levels "0","1": 1 2 1 2 1 1 2 2 2 2 ...
## $ time_after_review: Factor w/ 4 levels "", "0", "1", "N/A": NA 2 2 2 NA NA NA 3 3 NA ...
## $ outcome : Factor w/ 5 levels "", "0", "1", "sill",...: 2 3 2 2 2 3 2 2 2 3 ...
## $ Rdate1 : POSIXct, format: "2021-07-22" "2021-07-25" ...
```

```
data_consent$avpu <- as.numeric(data_consent$avpu)
names(data_consent)
```

```
## [1] "study_number"      "screen_date"      "group"
## [4] "sex"                 "age"              "ethnicity"
## [7] "consent"            "type_of_consent"  "Total"
## [10] "rr"                 "sat"              "air_oxygen"
## [13] "sbp"                "hr"               "avpu"
## [16] "temp"              "assessor1"        "assessor2"
## [19] "itu_admission"      "time_after_review" "outcome"
## [22] "Rdate1"

data_consent_numerical <- na.omit(data_consent)
XVars <- data_consent_numerical[,c(9,10,11,13,14,15,16,17)] # X numeric variables
round(cor(XVars), 2) # Correlation
```

```
##          Total    rr    sat    sbp    hr avpu    temp assessor1
## Total      1.00  0.54 -0.03 -0.18  0.54 0.10  0.19      0.27
## rr         0.54  1.00 -0.05  0.21  0.30 0.03  0.01      0.13
## sat        -0.03 -0.05  1.00 -0.18 -0.03 0.03  0.03     -0.02
## sbp        -0.18  0.21 -0.18  1.00  0.06 0.11 -0.05     -0.09
## hr         0.54  0.30 -0.03  0.06  1.00 0.15  0.11      0.12
## avpu        0.10  0.03  0.03  0.11  0.15 1.00  0.00      0.01
## temp        0.19  0.01  0.03 -0.05  0.11 0.00  1.00      0.21
## assessor1  0.27  0.13 -0.02 -0.09  0.12 0.01  0.21      1.00
```

```
## # to check if cor between Total News and qVIEWS is significant
cor.test(data_consent_numerical$Total,data_consent_numerical$assessor1)
```

```
##
## Pearson's product-moment correlation
##
## data: data_consent_numerical$Total and data_consent_numerical$assessor1
## t = 4.1868, df = 215, p-value = 4.126e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.146748 0.393366
## sample estimates:
##      cor
## 0.2745662
```

```
#RIDGE REGRESSION FOR NEWS variables alone
news_data1 <- newdata[,c(4:10,12)]
names(news_data1)
```

```
## [1] "rr"          "sat"          "air_oxygen"   "sbp"
## [5] "hr"          "avpu"         "temp"         "itu_admission"
```

```
##Splitting the sample (internal validation)
```

```
x1 <- model.matrix(news_data1$itu_admission~.,news_data1)[, -1]
y1 <- news_data1$itu_admission
set.seed(1)
train1 <- sample(1:nrow(x1),nrow(x1)/2)
test1 <- (-train1)
y.test1 <-y1[test1]
library(glmnet)
```

```
## Loading required package: Matrix
## Loaded glmnet 4.1-4
```

```

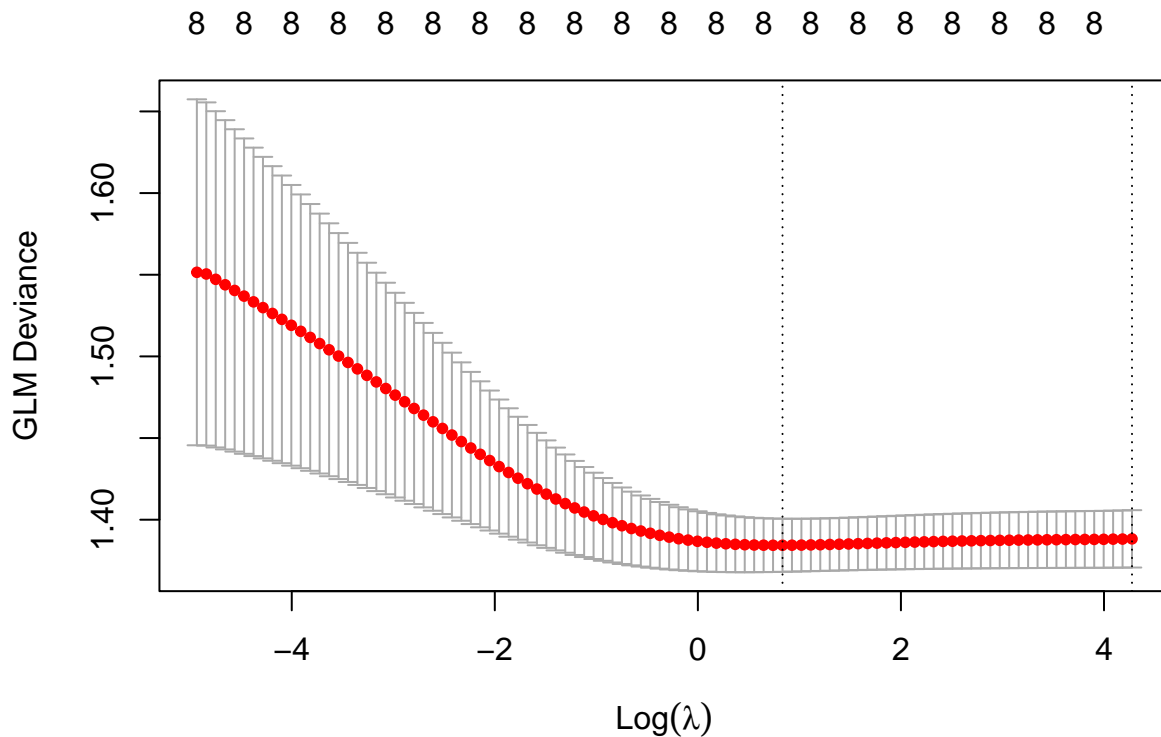
grid <- 10^seq(10,-2,length=100)

## Calculating best lambda
set.seed(1)
cv.out1 <- cv.glmnet(x1[train1,],y1[train1],alpha=0,family=binomial,type.measure="class")

## Warning: Only deviance, mse, mae available as type.measure for GLM models;
## deviance used instead

plot(cv.out1)

```



```

bestlam1 <- cv.out1$lambda.min
bestlam1

## [1] 2.301567

##Doing a ridge regression
out1 <- glmnet(x1[train1,],y1[train1],alpha=0,lamda=grid,family=binomial)
ridge.coeff1 <- predict(out1,s = bestlam1,newx = x1[test1 , ],type="coefficients")
ridge.coeff1

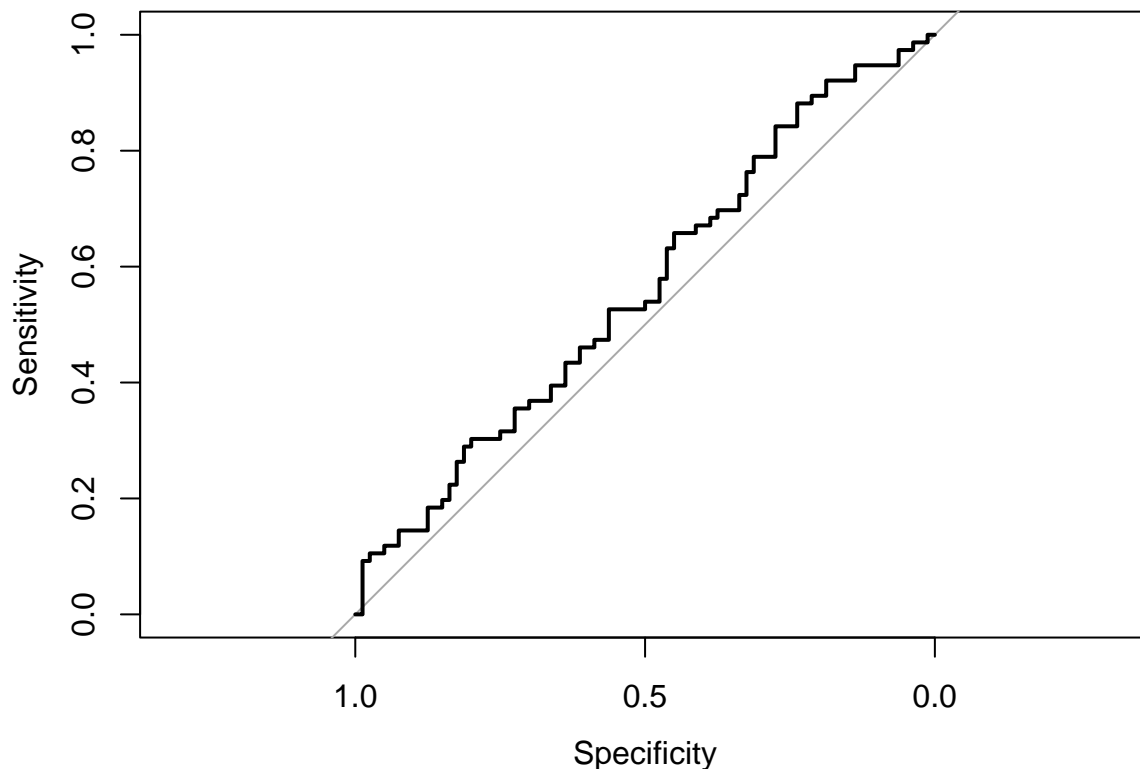
## 12 x 1 sparse Matrix of class "dgCMatrix"
##              s1
## (Intercept)  1.1250538976
## rr          -0.0029238414
## sat         -0.0090857200
## air_oxygen1 -0.0133241501
## air_oxygen1% .
## air_oxygen2  0.0338682851
## sbp          0.0004122020
## hr          -0.0007236309
## avpu2       -0.0466924945

```

```
## avpu3      .
## avpu4      .
## temp      -0.0104529102
## Doing AUC for NEWS
ridge.coeff1 <- predict(out1,s = bestlam1,newx = x1[test1 , ],type="response")
library(pROC)

## Type 'citation("pROC")' for a citation.
##
## Attaching package: 'pROC'
## The following objects are masked from 'package:stats':
##
##      cov, smooth, var
roccurve1 <- roc(y1[test1]~ ridge.coeff1)

## Setting levels: control = 0, case = 1
## Warning in roc.default(response, predictors[, 1], ...): Deprecated use a matrix
## as predictor. Unexpected results may be produced, please pass a numeric vector.
## Setting direction: controls < cases
plot(roccurve1)
```



```
auc(roccurve1)

## Area under the curve: 0.561
ci.auc(roccurve1)

## 95% CI: 0.4707-0.6514 (DeLong)
```

```

#RIDGE REGRESSION FOR NEWS variables + qVIEWS
news_data2 <- newdata[,c(4:12)] # Adding assessor1 which is qVIEWS
names(news_data2)

## [1] "rr"          "sat"          "air_oxygen"   "sbp"
## [5] "hr"          "avpu"         "temp"         "assessor1"
## [9] "itu_admission"

##Splitting the sample (internal validation)

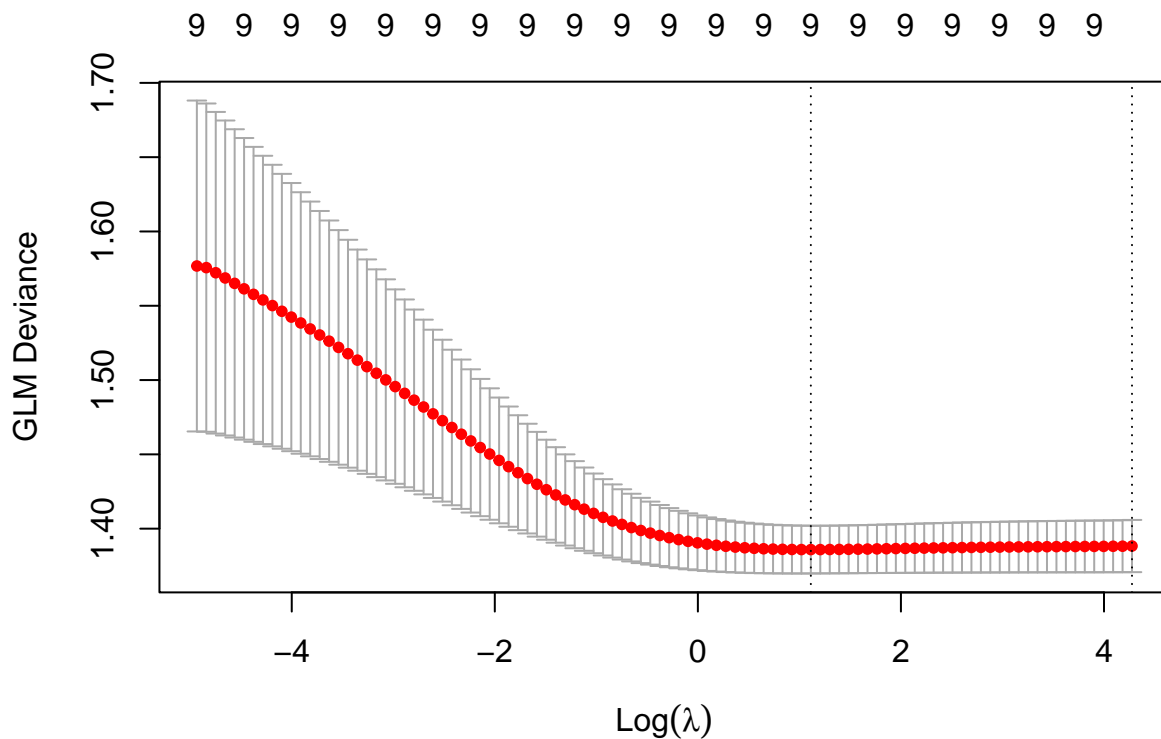
x2 <- model.matrix(news_data2$itu_admission~.,news_data2)[, -1]
y2 <- news_data2$itu_admission
set.seed(1)
train2 <- sample(1:nrow(x2),nrow(x2)/2)
test2 <- (-train2)
y.test2 <- y2[test2]
library(glmnet)
grid <- 10^seq(10,-2,length=100)

## Calculating best lambda
set.seed(1)
cv.out2 <- cv.glmnet(x2[train2,],y2[train2],alpha=0,family=binomial,type.measure="class")

## Warning: Only deviance, mse, mae available as type.measure for GLM models;
## deviance used instead

plot(cv.out2)

```



```

bestlam2 <- cv.out2$lambda.min
bestlam2

```

```
## [1] 3.042536
```



```

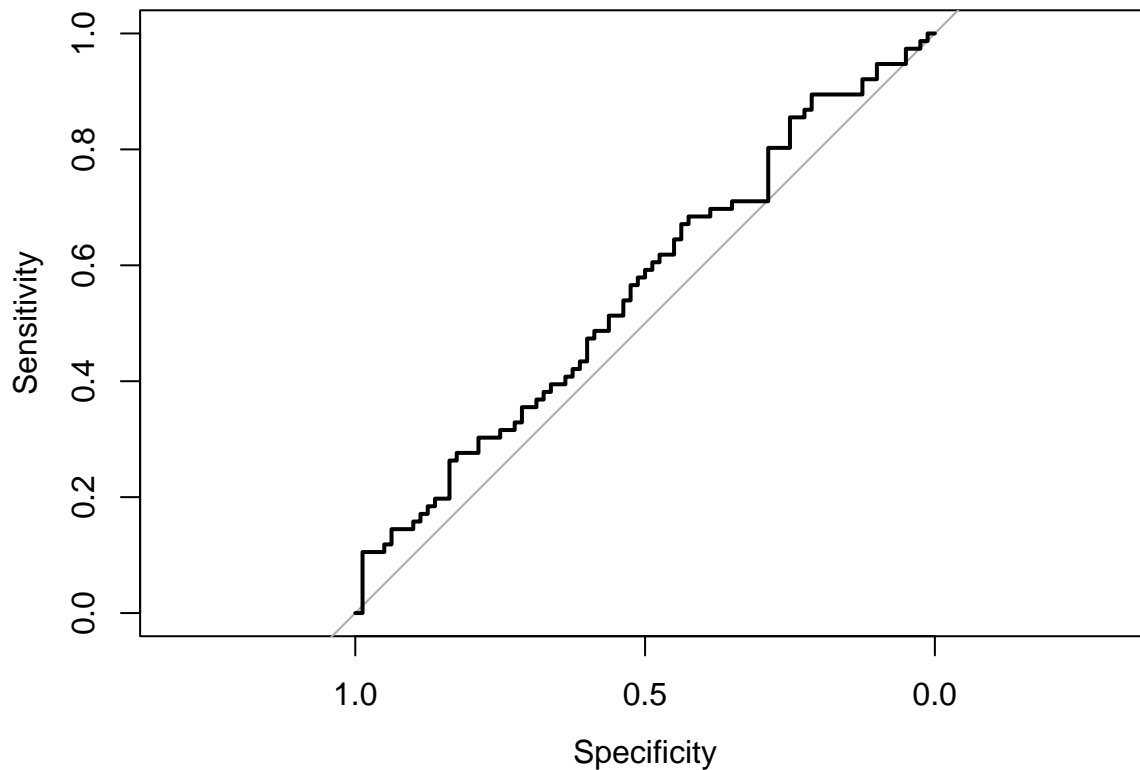
##Doing a ridge regression
out2 <- glmnet(x2[train2,],y2[train2],alpha=0,lamda=grid,family=binomial)
ridge.coeff2 <- predict(out2,s = bestlam2,newx = x2[test2 , ],type="coefficients")
ridge.coeff2

## 13 x 1 sparse Matrix of class "dgCMatrix"
##              s1
## (Intercept)  0.8360483349
## rr          -0.0022403986
## sat         -0.0070619480
## air_oxygen1 -0.0107636679
## air_oxygen1% .
## air_oxygen2  0.0265976062
## sbp          0.0003219603
## hr          -0.0005645104
## avpu2       -0.0360369242
## avpu3        .
## avpu4        .
## temp        -0.0081027133
## assessor1   -0.0045762624

## Doing AUC for NEWS+qVIEWS
ridge.coeff2 <- predict(out2,s = bestlam2,newx = x2[test2, ],type="response")
library(pROC)
roccurve2 <- roc(y2[test2]~ ridge.coeff2)

## Setting levels: control = 0, case = 1
## Warning in roc.default(response, predictors[, 1], ...): Deprecated use a matrix
## as predictor. Unexpected results may be produced, please pass a numeric vector.
## Setting direction: controls < cases
plot(roccurve2)

```



```

auc(roccurve2)

## Area under the curve: 0.5584
ci.auc(roccurve2)

## 95% CI: 0.4679-0.6489 (DeLong)
#GRAPH THAT COMBINES BOTH AUROC
plot.roc(y2[test2]~ ridge.coeff2,percentage=T,col="red",smooth=T)# for NEWS+qVIEWS

## Setting levels: control = 0, case = 1

## Warning in roc.default(response, predictor, plot = TRUE, ...): Deprecated use a
## matrix as predictor. Unexpected results may be produced, please pass a numeric
## vector.

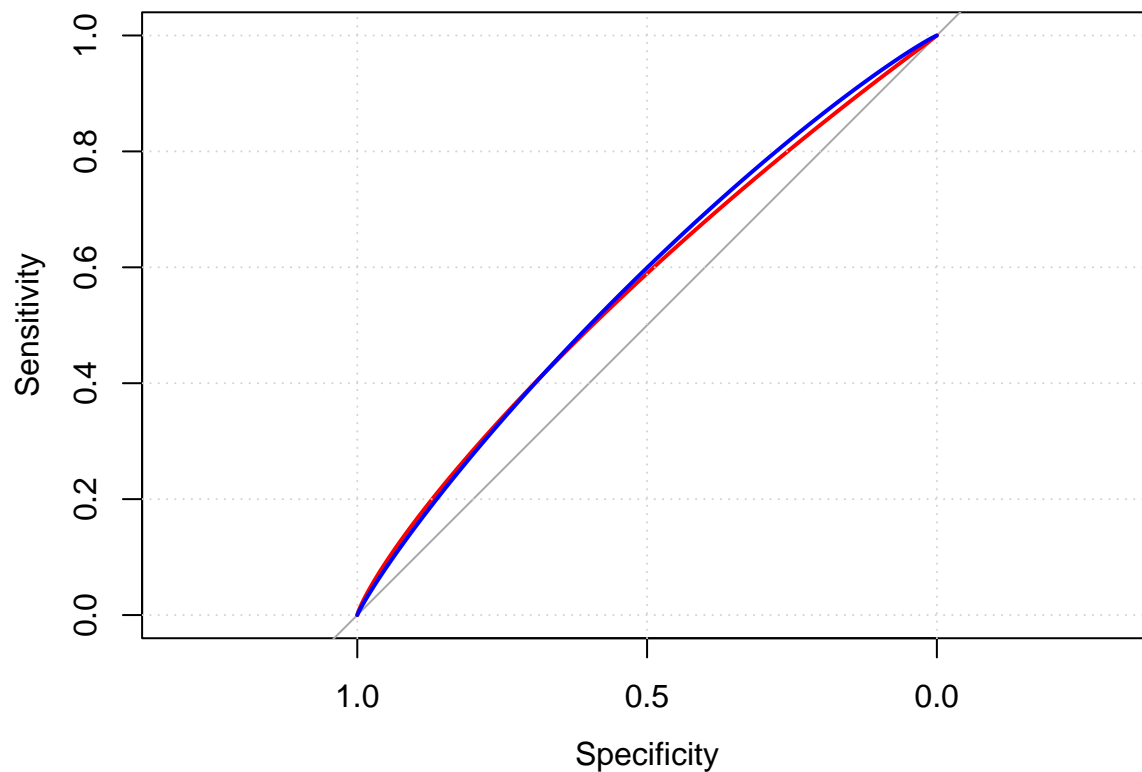
## Setting direction: controls < cases
grid(NULL,NULL,col = "lightgray", lty = "dotted",lwd = par("lwd"), equilogs = TRUE)
lines.roc(y1[test1]~ ridge.coeff1,percentage=T,col="blue",smooth=T)# For NEWS only

## Setting levels: control = 0, case = 1

## Warning in roc.default(response, predictor, ...): Deprecated use a matrix as
## predictor. Unexpected results may be produced, please pass a numeric vector.

## Setting direction: controls < cases

```



```
## RANDOM FOREST
```

```
data_consent <- data[data$consent=="1",]
names(data_consent)
```

```
## [1] "study_number"      "screen_date"       "group"
## [4] "sex"               "age"              "ethnicity"
## [7] "consent"           "type_of_consent"   "Total"
## [10] "rr"                "sat"              "air_oxygen"
## [13] "sbp"               "hr"               "avpu"
## [16] "temp"              "assessor1"         "assessor2"
## [19] "itu_admission"     "time_after_review" "outcome"
## [22] "Rdate1"
```

```
names(newdata)
```

```
## [1] "sex"      "age"      "ethnicity" "rr"
## [5] "sat"      "air_oxygen" "sbp"      "hr"
## [9] "avpu"     "temp"     "assessor1" "itu_admission"
```

```
newdata <- data_consent[,c(10:17,19)]
```

```
newdata <- na.omit(newdata)
names(newdata)
```

```
## [1] "rr"      "sat"      "air_oxygen" "sbp"
## [5] "hr"      "avpu"     "temp"       "assessor1"
## [9] "itu_admission"
```

```
library(randomForest)
```

```
## randomForest 4.7-1.1
```

```

## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
## The following object is masked from 'package:psych':
##
## outlier
# Separate train and test sample
set.seed(1)
library(caTools)
sample <- sample.split(newdata$itu_admission, SplitRatio = 0.5)
train <- subset(newdata, sample == TRUE)
testdata <- subset(newdata, sample == FALSE)
names(train)

## [1] "rr"          "sat"          "air_oxygen"   "sbp"
## [5] "hr"          "avpu"         "temp"         "assessor1"
## [9] "itu_admission"

names(testdata)

## [1] "rr"          "sat"          "air_oxygen"   "sbp"
## [5] "hr"          "avpu"         "temp"         "assessor1"
## [9] "itu_admission"

dim(train)

## [1] 158 9

dim(testdata)

## [1] 158 9

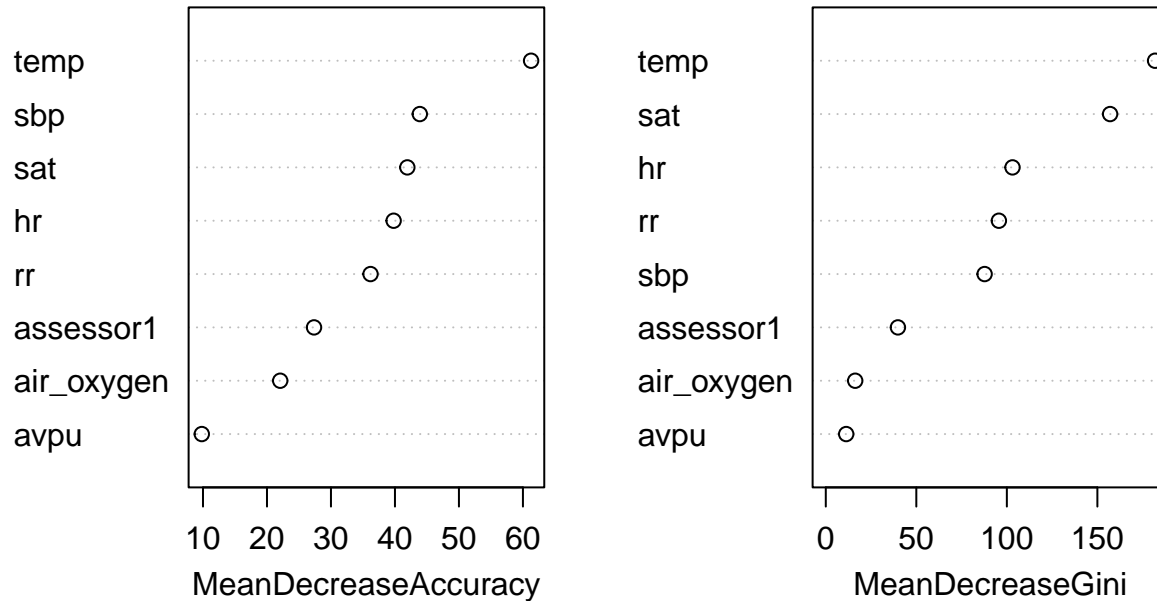
# Random Forest
set.seed(1)
library(randomForest)
rf.newdata <- randomForest(newdata$itu_admission~., data= newdata, subset=unlist(train),
mtry =3, importance = TRUE)
importance(rf.newdata)

##              0              1 MeanDecreaseAccuracy MeanDecreaseGini
## rr          28.052253 31.944769              36.183033              95.59939
## sat          35.940974 31.147522              41.936336             157.01804
## air_oxygen   16.174831 20.221729              22.069342              16.24112
## sbp          34.589565 34.976893              43.887144              87.71821
## hr           31.947055 31.131106              39.793760             103.11545
## avpu          9.476563  6.726149              9.813107              11.27454
## temp         50.874434 45.062224              61.277081             181.86569
## assessor1    25.569541 18.791137              27.352904              39.84012

varImpPlot(rf.newdata)

```

rf.newdata



```
# A better way to present the importance value as barplot
var.1 <- c("RR", "SAT", "AIR/OX", "SBP", "HR", "AVPU",
"TEMP", "qVIEWS")
Gini <- c(97.3, 160.2, 22.3, 81.0, 92.0, 11.9, 185.6, 36.9)
Gini_per <- (Gini*100)/185.6
dt1 <- data.frame(var.1, Gini_per)
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
## The following object is masked from 'package:randomForest':
##
##     margin
## The following objects are masked from 'package:psych':
##
##     %+%, alpha
```

```
library(reshape2)
library(tidyverse)
```

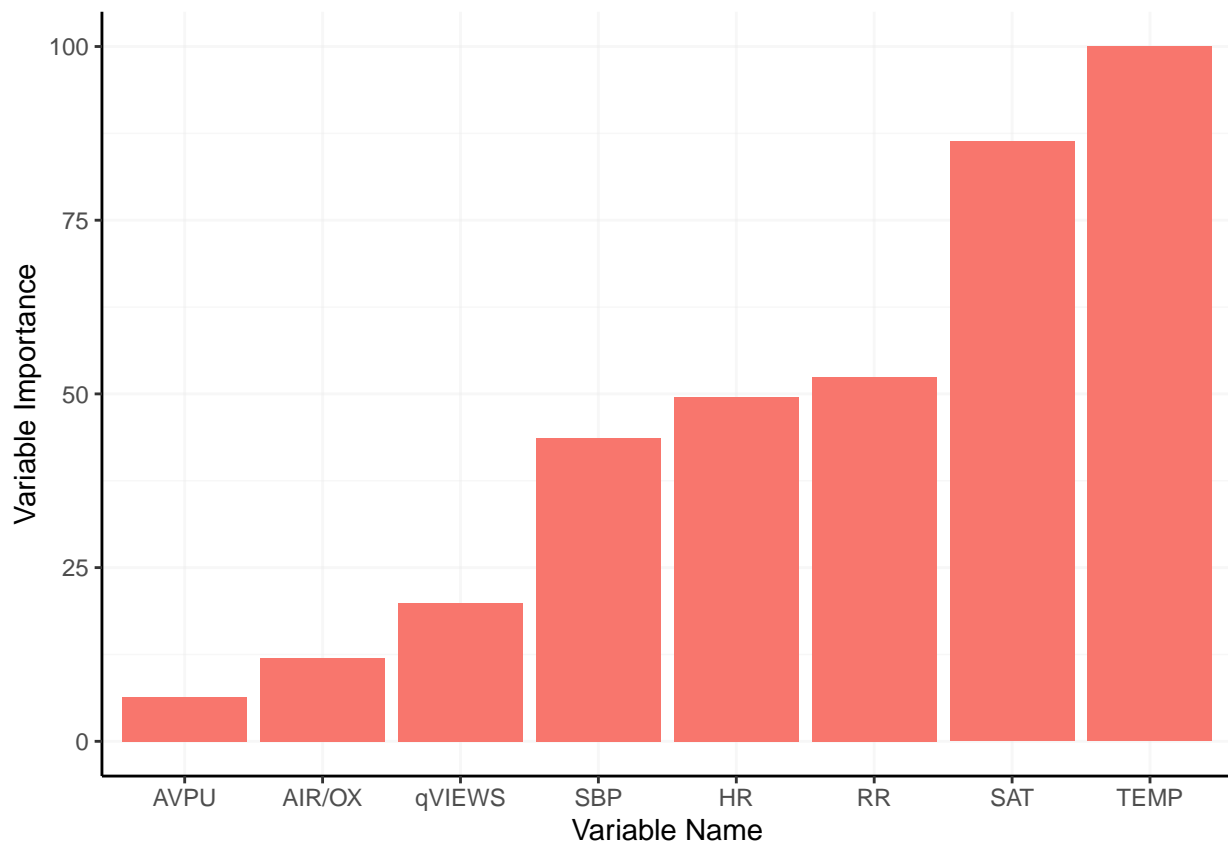
```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v tibble  3.1.8      v dplyr    1.0.9
## v tidyr   1.2.0      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## v purrr   0.3.4
## -- Conflicts ----- tidyverse_conflicts() --
## x ggplot2::%+%( ) masks psych::%+%( )
## x ggplot2::alpha( ) masks psych::alpha( )
```

```
## x dplyr::combine() masks randomForest::combine()
## x tidyr::expand() masks Matrix::expand()
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
## x ggplot2::margin() masks randomForest::margin()
## x tidyr::pack() masks Matrix::pack()
## x tidyr::unpack() masks Matrix::unpack()

dt2 <- melt(dt1,id.vars='var.1')

col_grid <- rgb(235, 235, 235, 100, maxColorValue = 255)

ggplot(dt2, aes(fct_reorder(var.1,value),value, fill=variable))+
  labs(y= "Variable Importance", x = "Variable Name")+
  theme_bw()+
  theme(panel.grid = element_line(color = col_grid))+
  theme(legend.position="none")+
  theme(panel.border = element_blank(), axis.line = element_line())+
  geom_bar(stat='identity',position='dodge')
```



```
pred1 <- predict(rf.newdata,newdata=testdata,type="class")
table(pred1,testdata$itu_admission)
```

```
##
## pred1  0  1
##      0 43  6
##      1 41 68
```

```

(40+68)/158 # Correct predictions for around 68% of the locations in the test data set

## [1] 0.6835443
#Sensitivity, specificity of random forest prediction versus itu_admission in test sample
mx <- matrix(c(68,44,6,40),byrow=T,nrow=2)# Order is modified to allow itu_admission (1) as "1"
mx

##      [,1] [,2]
## [1,]   68   44
## [2,]    6   40

library(caret)

## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##      lift

confusionMatrix(mx)

## Confusion Matrix and Statistics
##
##      A  B
## A 68 44
## B  6 40
##
##              Accuracy : 0.6835
##              95% CI : (0.6049, 0.7551)
##      No Information Rate : 0.5316
##      P-Value [Acc > NIR] : 7.332e-05
##
##              Kappa : 0.3834
##
##  Mcnemar's Test P-Value : 1.672e-07
##
##      Sensitivity : 0.9189
##      Specificity : 0.4762
##      Pos Pred Value : 0.6071
##      Neg Pred Value : 0.8696
##      Prevalence : 0.4684
##      Detection Rate : 0.4304
##      Detection Prevalence : 0.7089
##      Balanced Accuracy : 0.6976
##
##      'Positive' Class : A
##
#End of data analysis and end of script

```