Descriptive

Mingxu Shan

3/10/2019

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.0.0 ✔ purrr 0.2.5  
## ✔ tibble 1.4.2 ✔ dplyr 0.7.6  
## ✔ tidyr 0.8.1 ✔ stringr 1.3.1  
## ✔ readr 1.1.1 ✔ forcats 0.3.0

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(readxl)

olddata <- read\_excel("data/Master worksheet CTICU extubation times .xls",   
 sheet = "Master data", col\_types = c("numeric",   
 "numeric", "text", "date", "date",   
 "date", "date", "text", "text", "numeric",   
 "numeric", "numeric", "numeric",   
 "text", "text", "text", "text", "text",   
 "text", "text", "text", "numeric",   
 "numeric", "numeric", "text", "text",   
 "text", "text", "text", "text", "text",   
 "text", "text", "text", "text", "text",   
 "numeric", "text", "numeric", "text",   
 "numeric", "numeric", "text", "text",   
 "text", "text", "date", "date",   
 "date", "date", "date",   
 "text", "numeric", "text", "numeric",   
 "numeric", "text", "numeric", "text",   
 "text", "numeric", "numeric", "text",   
 "text", "text", "text", "text", "numeric",   
 "numeric", "numeric", "numeric",   
 "numeric", "numeric", "numeric"))

rawdata <- read\_excel("data/new master data file 3 14 2019.xlsx",   
 col\_types = c("text", "text", "date",   
 "date", "date", "numeric", "text",   
 "text", "numeric", "numeric", "numeric",   
 "numeric", "text", "text", "text",   
 "text", "text", "text", "text", "text",   
 "numeric", "numeric", "numeric",   
 "text", "text", "text", "text", "text",   
 "text", "text", "text", "text", "text",   
 "text", "text", "numeric", "text",   
 "numeric", "text", "numeric", "numeric",   
 "text", "text", "text", "text", "text",   
 "numeric", "text", "numeric", "date",   
 "date", "date", "date",   
 "date", "text", "numeric", "text",   
 "numeric", "numeric", "text", "numeric",   
 "text", "text", "numeric", "numeric",   
 "text", "text", "text", "text", "text",   
 "numeric", "numeric", "numeric",   
 "numeric", "numeric", "numeric",   
 "numeric", "text", "numeric", "numeric",   
 "numeric", "numeric", "numeric",   
 "numeric", "numeric", "text", "numeric"))

desdata <- rawdata %>%   
 separate("Surgery duration", into = c("Sdate", "Stime"), sep= " ") %>%  
 separate("OR Exit Date And Time", into = c("OREdate", "OREtime"), sep= " ") %>%  
 separate("OREtime", into = c("OREhour", "OREmin","OREsec"), sep= ":") %>%  
 separate("Stime", into = c("Shour","Smin","Ssec"), sep= ":") %>%   
 mutate (Shour=as.numeric(Shour), Smin=as.numeric(Smin), Stime = Shour+ (Smin/60),  
 OREhour = as.numeric(OREhour) ,  
 period = ifelse (OREhour <18 & OREhour >=7, "P1", "P2"),  
 Ethnicity = NA,  
 Reintub = coalesce(`Re-intubated During Hospital Stay`, `Postop Intubation/Reintubation During Hospital Stay`),  
 Mor\_disc = coalesce(`Mort-DC Status`,`Discharge / Mortality Status`)  
   
 ) %>%  
  
 rename(age = "Patient Age",   
 sex = "Sex",   
 BMI = "BMI",   
 PMM = "Predicted Morbidity or Mortality",   
 PRM = "Predicted Risk of Mortality",   
 PPV = "Predicted Prolonged Ventilation",   
 Reop = "Incidence",  
 Urgent = "Status",  
 Sug\_Dur = "Stime",  
 Pri\_MI = "Prior MI",  
 HF\_w2w = "Heart Failure within 2 weeks",  
 NYHA = "Classification-NYHA",  
 smoker = "RF-Tobacco Use",  
 lung = "RF-Chronic Lung Disease",  
 PFT = "RF-Pulmonary Function Test",  
 FEV = "RF-Forced Expiratory Volume Predicted",  
 DLCOT = "DLCO Test Performed",  
 DLCO = "DLCO Predicted",  
 RFD = "RF-Renal Fail-Dialysis",  
 PAD = "RF-Peripheral Arterial Disease",  
 #Main outcome  
 IHV = "Initial Hours Ventilated",  
 #Secondary outcome  
 Post\_pneu = "Post-Op-Pulm-Pneumonia",   
 Post\_throm = "Post-Op-Venous Thromboembolism-VTE",  
 Post\_stroke = "Post-Op-Neuro-Stroke Perm" ,  
 Re\_ICU = "Readmission to ICU" ,  
 Hos\_LOS = "LOS-Admit-Discharge" ,   
 ICU\_LOS = "Total Hrs ICU",  
 #Post-operative blood product transfusion   
 POBPT = "Blood Prod" ,   
 #Total blood transfusion units  
 TBU = "Total units (all blood products)",   
 XCT = "Cross Clamp Time (min)",   
 CPBT = "Cardiopulmonary Bypass Time"   
 ) %>%   
 select(age,  
 sex,  
 Ethnicity,  
 BMI,   
 PMM ,   
 PRM ,   
 PPV,  
 Reop ,  
 Urgent ,  
 Sug\_Dur ,  
 Pri\_MI ,  
 HF\_w2w ,  
 NYHA ,  
 smoker,  
 lung,  
 PFT,  
 FEV,  
 DLCOT ,  
 DLCO,  
 RFD ,  
 PAD,  
 OREhour,  
 IHV,  
 period,  
 # secondary outcome  
 Reintub,   
 Post\_pneu,   
 Post\_throm ,  
 Post\_stroke ,  
 Re\_ICU ,   
 Hos\_LOS,   
 ICU\_LOS,   
 Mor\_disc,  
 # new variables added by new dataset   
 POBPT,  
 TBU,  
 XCT,  
 CPBT  
 ) %>%   
 mutate(IHV\_24 = IHV)  
  
   
  
  
# combine ethnicity  
  
desdata$Ethnicity[rawdata$`Race - White` == "Yes"] = "White"  
desdata$Ethnicity[rawdata$`Race - Black / African American` == "Yes"] = "Black"  
desdata$Ethnicity[rawdata$`Race - Asian` == "Yes"] = "Asian"  
desdata$Ethnicity[rawdata$`Race - American Indian / Alaskan Native` == "Yes"] = "American Indian"  
desdata$Ethnicity[rawdata$`Race - Native Hawaiian / Pacific Islander` == "Yes"] = "Hawaiian"  
desdata$Ethnicity[rawdata$`Race - Other` == "Yes"] = "Other"  
desdata$Ethnicity[rawdata$`Race Documented` == "No"] = "No document"

desdata <- desdata %>% mutate( IHV\_24 = IHV)  
desdata$IHV\_24[desdata$IHV>24] = NA   
  
glmdata <- desdata

names(rawdata)

## [1] "Medical Record Number"   
## [2] "Procedure Type"   
## [3] "Date of Admission"   
## [4] "Date of Surgery"   
## [5] "Date of Discharge"   
## [6] "Mort-Date"   
## [7] "Mort-DC Status"   
## [8] "Discharge / Mortality Status"   
## [9] "LOS-Admit-Surgery"   
## [10] "LOS-Surgery-Discharge"   
## [11] "LOS-Admit-Discharge"   
## [12] "Patient Age"   
## [13] "Sex"   
## [14] "Race Documented"   
## [15] "Race - White"   
## [16] "Race - Black / African American"   
## [17] "Race - Asian"   
## [18] "Race - American Indian / Alaskan Native"   
## [19] "Race - Native Hawaiian / Pacific Islander"   
## [20] "Race - Other"   
## [21] "Height (cm)"   
## [22] "Weight (kg)"   
## [23] "BMI"   
## [24] "RF-Diabetes"   
## [25] "RF-Diabetes-Control"   
## [26] "Heart Failure within 2 weeks"   
## [27] "Heart Failure"   
## [28] "Classification-NYHA"   
## [29] "Prior MI"   
## [30] "MI-When"   
## [31] "RF-Chronic Lung Disease"   
## [32] "RF-Chronic Lung Disease - Type"   
## [33] "RF-Tobacco Use"   
## [34] "RF-Peripheral Arterial Disease"   
## [35] "RF-Pulmonary Function Test"   
## [36] "RF-Forced Expiratory Volume Predicted"   
## [37] "DLCO Test Performed"   
## [38] "DLCO Predicted"   
## [39] "RF-Arterial Blood Gas"   
## [40] "RF-Carbon Dioxide Level"   
## [41] "RF-Oxygen Level"   
## [42] "RF-Renal Fail-Dialysis"   
## [43] "Incidence"   
## [44] "Status"   
## [45] "Urgent Or Emergent Reason"   
## [46] "IABP"   
## [47] "IABP-When Inserted"   
## [48] "Extracorporeal Membrane Oxygenation"   
## [49] "ECMO When Initiated"   
## [50] "OR Entry Date And Time"   
## [51] "OR Exit Date And Time"   
## [52] "Surgery duration"   
## [53] "Initial Intubation Date And Time"   
## [54] "Initial Extubation Date And Time"   
## [55] "ICU Visit"   
## [56] "Initial ICU hours"   
## [57] "Readmission to ICU"   
## [58] "Additional ICU Hours"   
## [59] "Total Hrs ICU"   
## [60] "Extubated In OR"   
## [61] "Initial Hours Ventilated"   
## [62] "Re-intubated During Hospital Stay"   
## [63] "Postop Intubation/Reintubation During Hospital Stay"  
## [64] "Additional Hours Ventilated"   
## [65] "Total Postoperative Ventilation Hours"   
## [66] "In Hospital Post-Op Events"   
## [67] "Post-Op-Venous Thromboembolism-VTE"   
## [68] "Post-Op-Pulmonary Thromboembolism"   
## [69] "Post-Op-Pulm-Pneumonia"   
## [70] "Post-Op-Neuro-Stroke Perm"   
## [71] "Predicted Risk of Mortality"   
## [72] "Predicted Morbidity or Mortality"   
## [73] "Predicted Permanent Stroke"   
## [74] "Predicted Prolonged Ventilation"   
## [75] "Predicted Renal Failure"   
## [76] "Predicted Reoperation"   
## [77] "Predicted Deep Sternal Wound Infx"   
## [78] "Blood Prod"   
## [79] "Blood Prod - Cryo Units"   
## [80] "Blood Prod - FFP Units"   
## [81] "Blood Prod - Platelet Units"   
## [82] "Blood Prod - RBC Units"   
## [83] "Total units (all blood products)"   
## [84] "Postoperative Hemoglobin"   
## [85] "Cross Clamp Time (min)"   
## [86] "CPB Utilization"   
## [87] "Cardiopulmonary Bypass Time"

names(olddata)

## [1] "STS Data Version"   
## [2] "Medical Record Number"   
## [3] "Procedure Type"   
## [4] "Date of Admission"   
## [5] "Date of Surgery"   
## [6] "Date of Discharge"   
## [7] "Mort-Date"   
## [8] "Mort-DC Status"   
## [9] "Discharge / Mortality Status"   
## [10] "LOS-Admit-Surgery"   
## [11] "LOS-Surgery-Discharge"   
## [12] "LOS-Admit-Discharge"   
## [13] "Patient Age"   
## [14] "Sex"   
## [15] "Race Documented"   
## [16] "Race - White"   
## [17] "Race - Black / African American"   
## [18] "Race - Asian"   
## [19] "Race - American Indian / Alaskan Native"   
## [20] "Race - Native Hawaiian / Pacific Islander"   
## [21] "Race - Other"   
## [22] "Height (cm)"   
## [23] "Weight (kg)"   
## [24] "BMI"   
## [25] "RF-Diabetes"   
## [26] "RF-Diabetes-Control"   
## [27] "Heart Failure within 2 weeks"   
## [28] "Heart Failure"   
## [29] "Classification-NYHA"   
## [30] "Prior MI"   
## [31] "MI-When"   
## [32] "RF-Chronic Lung Disease"   
## [33] "RF-Chronic Lung Disease - Type"   
## [34] "RF-Tobacco Use"   
## [35] "RF-Peripheral Arterial Disease"   
## [36] "RF-Pulmonary Function Test"   
## [37] "RF-Forced Expiratory Volume Predicted"   
## [38] "DLCO Test Performed"   
## [39] "DLCO Predicted"   
## [40] "RF-Arterial Blood Gas"   
## [41] "RF-Carbon Dioxide Level"   
## [42] "RF-Oxygen Level"   
## [43] "RF-Renal Fail-Dialysis"   
## [44] "Incidence"   
## [45] "Status"   
## [46] "Urgent Or Emergent Reason"   
## [47] "OR Entry Date And Time"   
## [48] "OR Exit Date And Time"   
## [49] "Surgery duration"   
## [50] "Initial Intubation Date And Time"   
## [51] "Initial Extubation Date And Time"   
## [52] "ICU Visit"   
## [53] "Initial ICU hours"   
## [54] "Readmission to ICU"   
## [55] "Additional ICU Hours"   
## [56] "Total Hrs ICU"   
## [57] "Extubated In OR"   
## [58] "Initial Hours Ventilated"   
## [59] "Re-intubated During Hospital Stay"   
## [60] "Postop Intubation/Reintubation During Hospital Stay"  
## [61] "Additional Hours Ventilated"   
## [62] "Total Postoperative Ventilation Hours"   
## [63] "In Hospital Post-Op Events"   
## [64] "Post-Op-Venous Thromboembolism-VTE"   
## [65] "Post-Op-Pulmonary Thromboembolism"   
## [66] "Post-Op-Pulm-Pneumonia"   
## [67] "Post-Op-Neuro-Stroke Perm"   
## [68] "Predicted Risk of Mortality"   
## [69] "Predicted Morbidity or Mortality"   
## [70] "Predicted Permanent Stroke"   
## [71] "Predicted Prolonged Ventilation"   
## [72] "Predicted Renal Failure"   
## [73] "Predicted Reoperation"   
## [74] "Predicted Deep Sternal Wound Infx"

names(desdata)

## [1] "age" "sex" "Ethnicity" "BMI" "PMM"   
## [6] "PRM" "PPV" "Reop" "Urgent" "Sug\_Dur"   
## [11] "Pri\_MI" "HF\_w2w" "NYHA" "smoker" "lung"   
## [16] "PFT" "FEV" "DLCOT" "DLCO" "RFD"   
## [21] "PAD" "OREhour" "IHV" "period" "Reintub"   
## [26] "Post\_pneu" "Post\_throm" "Post\_stroke" "Re\_ICU" "Hos\_LOS"   
## [31] "ICU\_LOS" "Mor\_disc" "POBPT" "TBU" "XCT"   
## [36] "CPBT" "IHV\_24"

# Categorical

#desdata$sex[desdata$sex == "Not documented"] <- NA  
length(desdata$sex)

## [1] 861

sum(is.na(desdata$sex))

## [1] 0

sum(is.na(desdata$sex)) / length(desdata$sex)

## [1] 0

table(desdata$sex)

##   
## Female Male   
## 209 652

#table(desdata$sex) / length(desdata$sex)  
paste(round(100\*(table(desdata$sex) / 861), 2), "%", sep="")

## [1] "24.27%" "75.73%"

length(desdata$sex[desdata$period == "P1"])

## [1] 567

length(desdata$sex[desdata$period == "P2"])

## [1] 294

sum(is.na(desdata$sex[desdata$period == "P1"]))

## [1] 0

sum(is.na(desdata$sex[desdata$period == "P2"]))

## [1] 0

table(desdata$sex[desdata$period == "P1"])

##   
## Female Male   
## 133 434

table(desdata$sex[desdata$period == "P2"])

##   
## Female Male   
## 76 218

#table(desdata$sex[desdata$period == "P1"]) / length(desdata$sex[desdata$period == "P1"])  
paste(round(100\*(table(desdata$sex[desdata$period == "P1"]) / length(desdata$sex[desdata$period == "P1"])), 2), "%", sep="")

## [1] "23.46%" "76.54%"

#table(desdata$sex[desdata$period == "P2"]) / length(desdata$sex[desdata$period == "P2"])  
paste(round(100\*(table(desdata$sex[desdata$period == "P2"]) / length(desdata$sex[desdata$period == "P2"])), 2), "%", sep="")

## [1] "25.85%" "74.15%"

chisq.test(desdata$sex,desdata$period, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: desdata$sex and desdata$period  
## X-squared = 0.60343, df = 1, p-value = 0.4373

glmdata$Ethnicity[glmdata$Ethnicity == "No document"] <- NA  
glmdata$Ethnicity <- as.factor(glmdata$Ethnicity) %>%  
 relevel(glmdata$Ethnicity, ref="White")  
  
length(glmdata$Ethnicity)

## [1] 861

sum(is.na(glmdata$Ethnicity))

## [1] 48

sum(is.na(glmdata$Ethnicity)) / length(glmdata$Ethnicity)

## [1] 0.05574913

table(glmdata$Ethnicity)

##   
## White American Indian Asian Black   
## 639 1 57 87   
## Hawaiian Other   
## 17 12

#table(glmdata$Ethnicity) / length(glmdata$Ethnicity)  
paste(round(100\*(table(glmdata$Ethnicity) / 861), 2), "%", sep="")

## [1] "74.22%" "0.12%" "6.62%" "10.1%" "1.97%" "1.39%"

length(glmdata$Ethnicity[glmdata$period == "P1"])

## [1] 567

length(glmdata$Ethnicity[glmdata$period == "P2"])

## [1] 294

sum(is.na(glmdata$Ethnicity[glmdata$period == "P1"]))

## [1] 32

sum(is.na(glmdata$Ethnicity[glmdata$period == "P2"]))

## [1] 16

table(glmdata$Ethnicity[glmdata$period == "P1"])

##   
## White American Indian Asian Black   
## 419 1 46 53   
## Hawaiian Other   
## 8 8

table(glmdata$Ethnicity[glmdata$period == "P2"])

##   
## White American Indian Asian Black   
## 220 0 11 34   
## Hawaiian Other   
## 9 4

#table(glmdata$Ethnicity[glmdata$period == "P1"]) / length(glmdata$Ethnicity[glmdata$period == "P1"])  
paste(round(100\*(table(glmdata$Ethnicity[glmdata$period == "P1"]) / length(glmdata$Ethnicity[glmdata$period == "P1"])), 2), "%", sep="")

## [1] "73.9%" "0.18%" "8.11%" "9.35%" "1.41%" "1.41%"

#table(glmdata$Ethnicity[glmdata$period == "P2"]) / length(glmdata$Ethnicity[glmdata$period == "P2"])  
paste(round(100\*(table(glmdata$Ethnicity[glmdata$period == "P2"]) / length(glmdata$Ethnicity[glmdata$period == "P2"])), 2), "%", sep="")

## [1] "74.83%" "0%" "3.74%" "11.56%" "3.06%" "1.36%"

chisq.test(glmdata$Ethnicity,glmdata$period, correct = FALSE)

## Warning in chisq.test(glmdata$Ethnicity, glmdata$period, correct = FALSE):  
## Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: glmdata$Ethnicity and glmdata$period  
## X-squared = 9.7382, df = 5, p-value = 0.083

glmdata <- glmdata %>%   
 mutate(Reop = ifelse(Reop == "First cardiovascular surgery", "Yes", "No"))  
  
length(glmdata$Reop)

## [1] 861

sum(is.na(glmdata$Reop))

## [1] 0

sum(is.na(glmdata$Reop)) / length(glmdata$Reop)

## [1] 0

table(glmdata$Reop)

##   
## No Yes   
## 7 854

#table(glmdata$Reop) / length(glmdata$Reop)  
paste(round(100\*(table(glmdata$Reop) / (length(na.omit(glmdata$Reop)))), 2), "%", sep="")

## [1] "0.81%" "99.19%"

length(na.omit(glmdata$Reop[glmdata$period == "P1"]))

## [1] 567

length(na.omit(glmdata$Reop[glmdata$period == "P2"]))

## [1] 294

sum(is.na(glmdata$Reop[glmdata$period == "P1"]))

## [1] 0

sum(is.na(glmdata$Reop[glmdata$period == "P2"]))

## [1] 0

table(glmdata$Reop[glmdata$period == "P1"])

##   
## No Yes   
## 4 563

table(glmdata$Reop[glmdata$period == "P2"])

##   
## No Yes   
## 3 291

#table(glmdata$Reop[glmdata$period == "P1"]) / length(glmdata$Reop[glmdata$period == "P1"])  
paste(round(100\*(table(glmdata$Reop[glmdata$period == "P1"]) / length(na.omit(glmdata$Reop[glmdata$period == "P1"]))), 2), "%", sep="")

## [1] "0.71%" "99.29%"

#table(glmdata$Reop[glmdata$period == "P2"]) / length(glmdata$Reop[glmdata$period == "P2"])  
paste(round(100\*(table(glmdata$Reop[glmdata$period == "P2"]) / length(na.omit(glmdata$Reop[glmdata$period == "P2"]))), 2), "%", sep="")

## [1] "1.02%" "98.98%"

fisher.test(glmdata$Reop,glmdata$period)

##   
## Fisher's Exact Test for Count Data  
##   
## data: glmdata$Reop and glmdata$period  
## p-value = 0.6957  
## alternative hypothesis: true odds ratio is not equal to 1  
## 95 percent confidence interval:  
## 0.1158052 4.7381389  
## sample estimates:  
## odds ratio   
## 0.6894534

#desdata$Urgent[desdata$Urgent == "Not documented"] <- NA  
length(desdata$Urgent)

## [1] 861

sum(is.na(desdata$Urgent))

## [1] 0

sum(is.na(desdata$Urgent)) / length(desdata$Urgent)

## [1] 0

table(desdata$Urgent)

##   
## Elective Urgent   
## 312 549

#table(desdata$Urgent) / length(desdata$Urgent)  
paste(round(100\*(table(desdata$Urgent) / 861), 2), "%", sep="")

## [1] "36.24%" "63.76%"

length(desdata$Urgent[desdata$period == "P1"])

## [1] 567

length(desdata$Urgent[desdata$period == "P2"])

## [1] 294

sum(is.na(desdata$Urgent[desdata$period == "P1"]))

## [1] 0

sum(is.na(desdata$Urgent[desdata$period == "P2"]))

## [1] 0

table(desdata$Urgent[desdata$period == "P1"])

##   
## Elective Urgent   
## 214 353

table(desdata$Urgent[desdata$period == "P2"])

##   
## Elective Urgent   
## 98 196

#table(desdata$Urgent[desdata$period == "P1"]) / length(desdata$Urgent[desdata$period == "P1"])  
paste(round(100\*(table(desdata$Urgent[desdata$period == "P1"]) / length(desdata$Urgent[desdata$period == "P1"])), 2), "%", sep="")

## [1] "37.74%" "62.26%"

#table(desdata$Urgent[desdata$period == "P2"]) / length(desdata$Urgent[desdata$period == "P2"])  
paste(round(100\*(table(desdata$Urgent[desdata$period == "P2"]) / length(desdata$Urgent[desdata$period == "P2"])), 2), "%", sep="")

## [1] "33.33%" "66.67%"

chisq.test(desdata$Urgent,desdata$period, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: desdata$Urgent and desdata$period  
## X-squared = 1.629, df = 1, p-value = 0.2018

glmdata$Pri\_MI[glmdata$Pri\_MI == "Unknown"] <- NA  
length(glmdata$Pri\_MI)

## [1] 861

sum(is.na(glmdata$Pri\_MI))

## [1] 5

sum(is.na(glmdata$Pri\_MI)) / length(glmdata$Pri\_MI)

## [1] 0.005807201

table(glmdata$Pri\_MI)

##   
## No Yes   
## 538 318

#table(glmdata$Pri\_MI) / length(glmdata$Pri\_MI)  
paste(round(100\*(table(glmdata$Pri\_MI) / (length(na.omit(glmdata$Pri\_MI)))), 2), "%", sep="")

## [1] "62.85%" "37.15%"

length(na.omit(glmdata$Pri\_MI[glmdata$period == "P1"]))

## [1] 563

length(na.omit(glmdata$Pri\_MI[glmdata$period == "P2"]))

## [1] 293

sum(is.na(glmdata$Pri\_MI[glmdata$period == "P1"]))

## [1] 4

sum(is.na(glmdata$Pri\_MI[glmdata$period == "P2"]))

## [1] 1

table(glmdata$Pri\_MI[glmdata$period == "P1"])

##   
## No Yes   
## 364 199

table(glmdata$Pri\_MI[glmdata$period == "P2"])

##   
## No Yes   
## 174 119

#table(glmdata$Pri\_MI[glmdata$period == "P1"]) / length(glmdata$Pri\_MI[glmdata$period == "P1"])  
paste(round(100\*(table(glmdata$Pri\_MI[glmdata$period == "P1"]) / length(na.omit(glmdata$Pri\_MI[glmdata$period == "P1"]))), 2), "%", sep="")

## [1] "64.65%" "35.35%"

#table(glmdata$Pri\_MI[glmdata$period == "P2"]) / length(glmdata$Pri\_MI[glmdata$period == "P2"])  
paste(round(100\*(table(glmdata$Pri\_MI[glmdata$period == "P2"]) / length(na.omit(glmdata$Pri\_MI[glmdata$period == "P2"]))), 2), "%", sep="")

## [1] "59.39%" "40.61%"

chisq.test(glmdata$Pri\_MI,glmdata$period, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: glmdata$Pri\_MI and glmdata$period  
## X-squared = 2.2905, df = 1, p-value = 0.1302

glmdata$HF\_w2w[desdata$HF\_w2w == "Unknown"] <- NA  
length(desdata$HF\_w2w)

## [1] 861

sum(is.na(desdata$HF\_w2w))

## [1] 520

sum(is.na(desdata$HF\_w2w)) / length(desdata$HF\_w2w)

## [1] 0.6039489

table(desdata$HF\_w2w)

##   
## No Unknown Yes   
## 163 12 166

#table(desdata$HF\_w2w) / length(desdata$HF\_w2w)  
paste(round(100\*(table(desdata$HF\_w2w) / 861), 2), "%", sep="")

## [1] "18.93%" "1.39%" "19.28%"

length(desdata$HF\_w2w[desdata$period == "P1"])

## [1] 567

length(desdata$HF\_w2w[desdata$period == "P2"])

## [1] 294

sum(is.na(desdata$HF\_w2w[desdata$period == "P1"]))

## [1] 341

sum(is.na(desdata$HF\_w2w[desdata$period == "P2"]))

## [1] 179

table(desdata$HF\_w2w[desdata$period == "P1"])

##   
## No Unknown Yes   
## 111 5 110

table(desdata$HF\_w2w[desdata$period == "P2"])

##   
## No Unknown Yes   
## 52 7 56

#table(desdata$HF\_w2w[desdata$period == "P1"]) / length(desdata$HF\_w2w[desdata$period == "P1"])  
paste(round(100\*(table(desdata$HF\_w2w[desdata$period == "P1"]) / length(desdata$HF\_w2w[desdata$period == "P1"])), 2), "%", sep="")

## [1] "19.58%" "0.88%" "19.4%"

#table(desdata$HF\_w2w[desdata$period == "P2"]) / length(desdata$HF\_w2w[desdata$period == "P2"])  
paste(round(100\*(table(desdata$HF\_w2w[desdata$period == "P2"]) / length(desdata$HF\_w2w[desdata$period == "P2"])), 2), "%", sep="")

## [1] "17.69%" "2.38%" "19.05%"

chisq.test(desdata$HF\_w2w,desdata$period, correct = FALSE)

## Warning in chisq.test(desdata$HF\_w2w, desdata$period, correct = FALSE):  
## Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: desdata$HF\_w2w and desdata$period  
## X-squared = 3.4936, df = 2, p-value = 0.1743

#desdata$NYHA[desdata$NYHA == "Not documented"] <- NA  
length(desdata$NYHA)

## [1] 861

sum(is.na(desdata$NYHA))

## [1] 508

sum(is.na(desdata$NYHA)) / length(desdata$NYHA)

## [1] 0.5900116

table(desdata$NYHA)

##   
## Class I Class II Class III Class IV Not documented   
## 59 139 106 21 28

#table(desdata$NYHA) / length(desdata$NYHA)  
paste(round(100\*(table(desdata$NYHA) / 861), 2), "%", sep="")

## [1] "6.85%" "16.14%" "12.31%" "2.44%" "3.25%"

length(desdata$NYHA[desdata$period == "P1"])

## [1] 567

length(desdata$NYHA[desdata$period == "P2"])

## [1] 294

sum(is.na(desdata$NYHA[desdata$period == "P1"]))

## [1] 330

sum(is.na(desdata$NYHA[desdata$period == "P2"]))

## [1] 178

table(desdata$NYHA[desdata$period == "P1"])

##   
## Class I Class II Class III Class IV Not documented   
## 41 92 76 14 14

table(desdata$NYHA[desdata$period == "P2"])

##   
## Class I Class II Class III Class IV Not documented   
## 18 47 30 7 14

#table(desdata$NYHA[desdata$period == "P1"]) / length(desdata$NYHA[desdata$period == "P1"])  
paste(round(100\*(table(desdata$NYHA[desdata$period == "P1"]) / length(desdata$NYHA[desdata$period == "P1"])), 2), "%", sep="")

## [1] "7.23%" "16.23%" "13.4%" "2.47%" "2.47%"

#table(desdata$NYHA[desdata$period == "P2"]) / length(desdata$NYHA[desdata$period == "P2"])  
paste(round(100\*(table(desdata$NYHA[desdata$period == "P2"]) / length(desdata$NYHA[desdata$period == "P2"])), 2), "%", sep="")

## [1] "6.12%" "15.99%" "10.2%" "2.38%" "4.76%"

glmdata$smoker <- "Yes"  
glmdata$smoker[desdata$smoker == "Smoking status unknown"] <- NA  
glmdata$smoker[desdata$smoker == "Never smoker"] <- "No"  
length(glmdata$smoker)

## [1] 861

sum(is.na(glmdata$smoker))

## [1] 28

sum(is.na(glmdata$smoker)) / length(glmdata$smoker)

## [1] 0.03252033

table(glmdata$smoker)

##   
## No Yes   
## 400 433

#table(glmdata$smoker) / length(glmdata$smoker)  
paste(round(100\*(table(glmdata$smoker) / (length(na.omit(glmdata$smoker)))), 2), "%", sep="")

## [1] "48.02%" "51.98%"

length(na.omit(glmdata$smoker[glmdata$period == "P1"]))

## [1] 550

length(na.omit(glmdata$smoker[glmdata$period == "P2"]))

## [1] 283

sum(is.na(glmdata$smoker[glmdata$period == "P1"]))

## [1] 17

sum(is.na(glmdata$smoker[glmdata$period == "P2"]))

## [1] 11

table(glmdata$smoker[glmdata$period == "P1"])

##   
## No Yes   
## 272 278

table(glmdata$smoker[glmdata$period == "P2"])

##   
## No Yes   
## 128 155

#table(glmdata$smoker[glmdata$period == "P1"]) / length(glmdata$smoker[glmdata$period == "P1"])  
paste(round(100\*(table(glmdata$smoker[glmdata$period == "P1"]) / length(na.omit(glmdata$smoker[glmdata$period == "P1"]))), 2), "%", sep="")

## [1] "49.45%" "50.55%"

#table(glmdata$smoker[glmdata$period == "P2"]) / length(glmdata$smoker[glmdata$period == "P2"])  
paste(round(100\*(table(glmdata$smoker[glmdata$period == "P2"]) / length(na.omit(glmdata$smoker[glmdata$period == "P2"]))), 2), "%", sep="")

## [1] "45.23%" "54.77%"

chisq.test(glmdata$smoker,glmdata$period, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: glmdata$smoker and glmdata$period  
## X-squared = 1.3362, df = 1, p-value = 0.2477

glmdata$lung <- "Yes"  
glmdata$lung[desdata$lung == "No"] <- "No"  
glmdata$lung[desdata$lung == "Unknown"] <- NA  
  
length(glmdata$lung)

## [1] 861

sum(is.na(glmdata$lung))

## [1] 2

sum(is.na(glmdata$lung)) / length(glmdata$lung)

## [1] 0.00232288

table(glmdata$lung)

##   
## No Yes   
## 693 166

#table(glmdata$lung) / length(glmdata$lung)  
paste(round(100\*(table(glmdata$lung) / 861), 2), "%", sep="")

## [1] "80.49%" "19.28%"

length(glmdata$lung[glmdata$period == "P1"])

## [1] 567

length(glmdata$lung[glmdata$period == "P2"])

## [1] 294

sum(is.na(glmdata$lung[glmdata$period == "P1"]))

## [1] 2

sum(is.na(glmdata$lung[glmdata$period == "P2"]))

## [1] 0

table(glmdata$lung[glmdata$period == "P1"])

##   
## No Yes   
## 458 107

table(glmdata$lung[glmdata$period == "P2"])

##   
## No Yes   
## 235 59

#table(glmdata$lung[glmdata$period == "P1"]) / length(glmdata$lung[glmdata$period == "P1"])  
paste(round(100\*(table(glmdata$lung[glmdata$period == "P1"]) / length(glmdata$lung[glmdata$period == "P1"])), 2), "%", sep="")

## [1] "80.78%" "18.87%"

#table(glmdata$lung[glmdata$period == "P2"]) / length(glmdata$lung[glmdata$period == "P2"])  
paste(round(100\*(table(glmdata$lung[glmdata$period == "P2"]) / length(glmdata$lung[glmdata$period == "P2"])), 2), "%", sep="")

## [1] "79.93%" "20.07%"

chisq.test(glmdata$lung,glmdata$period, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: glmdata$lung and glmdata$period  
## X-squared = 0.15837, df = 1, p-value = 0.6907

glmdata$PFT[desdata$PFT == "Unknown"] <- NA  
  
length(glmdata$PFT)

## [1] 861

sum(is.na(glmdata$PFT))

## [1] 0

sum(is.na(glmdata$PFT)) / length(glmdata$PFT)

## [1] 0

table(glmdata$PFT)

##   
## No Yes   
## 437 424

#table(glmdata$PFT) / length(glmdata$PFT)  
paste(round(100\*(table(glmdata$PFT) / 861), 2), "%", sep="")

## [1] "50.75%" "49.25%"

length(glmdata$PFT[glmdata$period == "P1"])

## [1] 567

length(glmdata$PFT[glmdata$period == "P2"])

## [1] 294

sum(is.na(glmdata$PFT[glmdata$period == "P1"]))

## [1] 0

sum(is.na(glmdata$PFT[glmdata$period == "P2"]))

## [1] 0

table(glmdata$PFT[glmdata$period == "P1"])

##   
## No Yes   
## 303 264

table(glmdata$PFT[glmdata$period == "P2"])

##   
## No Yes   
## 134 160

#table(glmdata$PFT[glmdata$period == "P1"]) / length(glmdata$PFT[glmdata$period == "P1"])  
paste(round(100\*(table(glmdata$PFT[glmdata$period == "P1"]) / length(glmdata$PFT[glmdata$period == "P1"])), 2), "%", sep="")

## [1] "53.44%" "46.56%"

#table(glmdata$PFT[glmdata$period == "P2"]) / length(glmdata$PFT[glmdata$period == "P2"])  
paste(round(100\*(table(glmdata$PFT[glmdata$period == "P2"]) / length(glmdata$PFT[glmdata$period == "P2"])), 2), "%", sep="")

## [1] "45.58%" "54.42%"

chisq.test(glmdata$PFT,glmdata$period, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: glmdata$PFT and glmdata$period  
## X-squared = 4.7867, df = 1, p-value = 0.02868

#desdata$DLCOT[desdata$DLCOT == "Not documented"] <- NA  
length(desdata$DLCOT)

## [1] 861

sum(is.na(desdata$DLCOT))

## [1] 437

sum(is.na(desdata$DLCOT)) / length(desdata$DLCOT)

## [1] 0.5075494

table(desdata$DLCOT)

##   
## No Yes   
## 248 176

#table(desdata$DLCOT) / length(desdata$DLCOT)  
paste(round(100\*(table(desdata$DLCOT) / 861), 2), "%", sep="")

## [1] "28.8%" "20.44%"

length(desdata$DLCOT[desdata$period == "P1"])

## [1] 567

length(desdata$DLCOT[desdata$period == "P2"])

## [1] 294

sum(is.na(desdata$DLCOT[desdata$period == "P1"]))

## [1] 303

sum(is.na(desdata$DLCOT[desdata$period == "P2"]))

## [1] 134

table(desdata$DLCOT[desdata$period == "P1"])

##   
## No Yes   
## 150 114

table(desdata$DLCOT[desdata$period == "P2"])

##   
## No Yes   
## 98 62

#table(desdata$DLCOT[desdata$period == "P1"]) / length(desdata$DLCOT[desdata$period == "P1"])  
paste(round(100\*(table(desdata$DLCOT[desdata$period == "P1"]) / length(desdata$DLCOT[desdata$period == "P1"])), 2), "%", sep="")

## [1] "26.46%" "20.11%"

#table(desdata$DLCOT[desdata$period == "P2"]) / length(desdata$DLCOT[desdata$period == "P2"])  
paste(round(100\*(table(desdata$DLCOT[desdata$period == "P2"]) / length(desdata$DLCOT[desdata$period == "P2"])), 2), "%", sep="")

## [1] "33.33%" "21.09%"

glmdata$RFD[desdata$RFD == "Unknown"] <- NA  
  
length(glmdata$RFD)

## [1] 861

sum(is.na(glmdata$RFD))

## [1] 2

sum(is.na(glmdata$RFD)) / length(glmdata$RFD)

## [1] 0.00232288

table(glmdata$RFD)

##   
## No Yes   
## 829 30

#table(glmdata$RFD) / length(glmdata$RFD)  
paste(round(100\*(table(glmdata$RFD) / 861), 2), "%", sep="")

## [1] "96.28%" "3.48%"

length(glmdata$RFD[glmdata$period == "P1"])

## [1] 567

length(glmdata$RFD[glmdata$period == "P2"])

## [1] 294

sum(is.na(glmdata$RFD[glmdata$period == "P1"]))

## [1] 2

sum(is.na(glmdata$RFD[glmdata$period == "P2"]))

## [1] 0

table(glmdata$RFD[glmdata$period == "P1"])

##   
## No Yes   
## 543 22

table(glmdata$RFD[glmdata$period == "P2"])

##   
## No Yes   
## 286 8

#table(glmdata$RFD[glmdata$period == "P1"]) / length(glmdata$RFD[glmdata$period == "P1"])  
paste(round(100\*(table(glmdata$RFD[glmdata$period == "P1"]) / length(glmdata$RFD[glmdata$period == "P1"])), 2), "%", sep="")

## [1] "95.77%" "3.88%"

#table(glmdata$RFD[glmdata$period == "P2"]) / length(glmdata$RFD[glmdata$period == "P2"])  
paste(round(100\*(table(glmdata$RFD[glmdata$period == "P2"]) / length(glmdata$RFD[glmdata$period == "P2"])), 2), "%", sep="")

## [1] "97.28%" "2.72%"

chisq.test(glmdata$RFD,glmdata$period, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: glmdata$RFD and glmdata$period  
## X-squared = 0.78904, df = 1, p-value = 0.3744

#desdata$PAD[desdata$PAD == "Not documented"] <- NA  
length(desdata$PAD)

## [1] 861

sum(is.na(desdata$PAD))

## [1] 0

sum(is.na(desdata$PAD)) / length(desdata$PAD)

## [1] 0

table(desdata$PAD)

##   
## No Yes   
## 775 86

#table(desdata$PAD) / length(desdata$PAD)  
paste(round(100\*(table(desdata$PAD) / 861), 2), "%", sep="")

## [1] "90.01%" "9.99%"

length(desdata$PAD[desdata$period == "P1"])

## [1] 567

length(desdata$PAD[desdata$period == "P2"])

## [1] 294

sum(is.na(desdata$PAD[desdata$period == "P1"]))

## [1] 0

sum(is.na(desdata$PAD[desdata$period == "P2"]))

## [1] 0

table(desdata$PAD[desdata$period == "P1"])

##   
## No Yes   
## 520 47

table(desdata$PAD[desdata$period == "P2"])

##   
## No Yes   
## 255 39

#table(desdata$PAD[desdata$period == "P1"]) / length(desdata$PAD[desdata$period == "P1"])  
paste(round(100\*(table(desdata$PAD[desdata$period == "P1"]) / length(desdata$PAD[desdata$period == "P1"])), 2), "%", sep="")

## [1] "91.71%" "8.29%"

#table(desdata$PAD[desdata$period == "P2"]) / length(desdata$PAD[desdata$period == "P2"])  
paste(round(100\*(table(desdata$PAD[desdata$period == "P2"]) / length(desdata$PAD[desdata$period == "P2"])), 2), "%", sep="")

## [1] "86.73%" "13.27%"

chisq.test(glmdata$PAD,glmdata$period, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: glmdata$PAD and glmdata$period  
## X-squared = 5.3322, df = 1, p-value = 0.02094

length(glmdata$POBPT )

## [1] 861

sum(is.na(glmdata$POBPT ))

## [1] 0

sum(is.na(glmdata$POBPT )) / length(glmdata$POBPT )

## [1] 0

table(glmdata$POBPT )

##   
## No Yes   
## 632 229

#table(glmdata$POBPT ) / length(glmdata$POBPT )  
paste(round(100\*(table(glmdata$POBPT ) / (length(na.omit(glmdata$POBPT )))), 2), "%", sep="")

## [1] "73.4%" "26.6%"

length(na.omit(glmdata$POBPT [glmdata$period == "P1"]))

## [1] 567

length(na.omit(glmdata$POBPT [glmdata$period == "P2"]))

## [1] 294

sum(is.na(glmdata$POBPT [glmdata$period == "P1"]))

## [1] 0

sum(is.na(glmdata$POBPT [glmdata$period == "P2"]))

## [1] 0

table(glmdata$POBPT [glmdata$period == "P1"])

##   
## No Yes   
## 426 141

table(glmdata$POBPT [glmdata$period == "P2"])

##   
## No Yes   
## 206 88

#table(glmdata$POBPT [glmdata$period == "P1"]) / length(glmdata$POBPT [glmdata$period == "P1"])  
paste(round(100\*(table(glmdata$POBPT [glmdata$period == "P1"]) / length(na.omit(glmdata$POBPT [glmdata$period == "P1"]))), 2), "%", sep="")

## [1] "75.13%" "24.87%"

#table(glmdata$POBPT [glmdata$period == "P2"]) / length(glmdata$POBPT [glmdata$period == "P2"])  
paste(round(100\*(table(glmdata$POBPT [glmdata$period == "P2"]) / length(na.omit(glmdata$POBPT [glmdata$period == "P2"]))), 2), "%", sep="")

## [1] "70.07%" "29.93%"

chisq.test(glmdata$POBPT ,glmdata$period, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: glmdata$POBPT and glmdata$period  
## X-squared = 2.5434, df = 1, p-value = 0.1108

# Continuous

desdata$age[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 567

summary(desdata$age[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 28.0 60.0 68.0 66.4 73.0 90.0

desdata$age[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 9.763268

desdata$age[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 294

summary(desdata$age[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 37.00 60.00 66.00 66.42 74.00 87.00

desdata$age[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 9.491814

desdata$age %>% na.omit() %>% length()

## [1] 861

summary(desdata$age)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 28.00 60.00 67.00 66.41 73.00 90.00

desdata$age %>% na.omit() %>% sd()

## [1] 9.665915

var.test(desdata$age[desdata$period == "P2"], desdata$age[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$age[desdata$period == "P2"] and desdata$age[desdata$period == "P1"]  
## F = 0.94517, num df = 293, denom df = 566, p-value = 0.5883  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.7764786 1.1579179  
## sample estimates:  
## ratio of variances   
## 0.9451658

t.test(x= desdata$age[desdata$period == "P1"], y = desdata$age[desdata$period == "P2"], paired = FALSE, var.equal = TRUE, conf.level = 0.95)

##   
## Two Sample t-test  
##   
## data: desdata$age[desdata$period == "P1"] and desdata$age[desdata$period == "P2"]  
## t = -0.033348, df = 859, p-value = 0.9734  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.387424 1.341065  
## sample estimates:  
## mean of x mean of y   
## 66.39859 66.42177

desdata$PMM[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 567

summary(desdata$PMM[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.02410 0.06575 0.09056 0.12285 0.14401 0.65547

desdata$PMM[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 0.09637599

desdata$PMM[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 294

summary(desdata$PMM[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.02529 0.06339 0.10855 0.13730 0.16008 0.66575

desdata$PMM[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 0.1083083

desdata$PMM %>% na.omit() %>% length()

## [1] 861

summary(desdata$PMM)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.02410 0.06486 0.09656 0.12778 0.15052 0.66575

desdata$PMM %>% na.omit() %>% sd()

## [1] 0.1007801

var.test(desdata$PMM[desdata$period == "P2"], desdata$PMM[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$PMM[desdata$period == "P2"] and desdata$PMM[desdata$period == "P1"]  
## F = 1.2629, num df = 293, denom df = 566, p-value = 0.01985  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 1.037545 1.547232  
## sample estimates:  
## ratio of variances   
## 1.262948

t.test(x= desdata$PMM[desdata$period == "P1"], y = desdata$PMM[desdata$period == "P2"], paired = FALSE, var.equal = FALSE, conf.level = 0.95)

##   
## Welch Two Sample t-test  
##   
## data: desdata$PMM[desdata$period == "P1"] and desdata$PMM[desdata$period == "P2"]  
## t = -1.9257, df = 536.19, p-value = 0.05468  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.0291836968 0.0002906079  
## sample estimates:  
## mean of x mean of y   
## 0.1228515 0.1372981

desdata$PRM[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 567

summary(desdata$PRM[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00229 0.00599 0.01009 0.01695 0.01840 0.30250

desdata$PRM[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 0.02527103

desdata$PRM[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 294

summary(desdata$PRM[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.002020 0.006443 0.011845 0.019615 0.022120 0.220700

desdata$PRM[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 0.02611441

desdata$PRM %>% na.omit() %>% length()

## [1] 861

summary(desdata$PRM)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00202 0.00615 0.01075 0.01786 0.01967 0.30250

desdata$PRM %>% na.omit() %>% sd()

## [1] 0.02557818

var.test(desdata$PRM[desdata$period == "P2"], desdata$PRM[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$PRM[desdata$period == "P2"] and desdata$PRM[desdata$period == "P1"]  
## F = 1.0679, num df = 293, denom df = 566, p-value = 0.5108  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.8772756 1.3082307  
## sample estimates:  
## ratio of variances   
## 1.067861

t.test(x= desdata$PRM[desdata$period == "P1"], y = desdata$PRM[desdata$period == "P2"], paired = FALSE, var.equal = TRUE, conf.level = 0.95)

##   
## Two Sample t-test  
##   
## data: desdata$PRM[desdata$period == "P1"] and desdata$PRM[desdata$period == "P2"]  
## t = -1.4494, df = 859, p-value = 0.1476  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.0062683536 0.0009430323  
## sample estimates:  
## mean of x mean of y   
## 0.01695217 0.01961483

desdata$PPV[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 567

summary(desdata$PPV[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.01156 0.03770 0.05597 0.07983 0.09177 0.55280

desdata$PPV[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 0.07441677

desdata$PPV[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 294

summary(desdata$PPV[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.01253 0.03785 0.06800 0.08784 0.09978 0.55772

desdata$PPV[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 0.07961393

desdata$PPV %>% na.omit() %>% length()

## [1] 861

summary(desdata$PPV)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.01156 0.03767 0.05900 0.08256 0.09539 0.55772

desdata$PPV %>% na.omit() %>% sd()

## [1] 0.07627977

var.test(desdata$PPV[desdata$period == "P2"], desdata$PPV[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$PPV[desdata$period == "P2"] and desdata$PPV[desdata$period == "P1"]  
## F = 1.1446, num df = 293, denom df = 566, p-value = 0.1784  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.9402817 1.4021880  
## sample estimates:  
## ratio of variances   
## 1.144555

t.test(x= desdata$PPV[desdata$period == "P1"], y = desdata$PPV[desdata$period == "P2"], paired = FALSE, var.equal = TRUE, conf.level = 0.95)

##   
## Two Sample t-test  
##   
## data: desdata$PPV[desdata$period == "P1"] and desdata$PPV[desdata$period == "P2"]  
## t = -1.4624, df = 859, p-value = 0.144  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.01876462 0.00274085  
## sample estimates:  
## mean of x mean of y   
## 0.07982727 0.08783915

desdata$Sug\_Dur[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 567

summary(desdata$Sug\_Dur[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.033 6.350 6.983 7.077 7.733 15.033

desdata$Sug\_Dur[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 1.133043

desdata$Sug\_Dur[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 294

summary(desdata$Sug\_Dur[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.317 6.125 6.783 7.006 7.562 16.933

desdata$Sug\_Dur[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 1.453969

desdata$Sug\_Dur %>% na.omit() %>% length()

## [1] 861

summary(desdata$Sug\_Dur)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.033 6.250 6.933 7.053 7.650 16.933

desdata$Sug\_Dur %>% na.omit() %>% sd()

## [1] 1.251513

var.test(desdata$Sug\_Dur[desdata$period == "P2"], desdata$Sug\_Dur[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$Sug\_Dur[desdata$period == "P2"] and desdata$Sug\_Dur[desdata$period == "P1"]  
## F = 1.6467, num df = 293, denom df = 566, p-value = 5.27e-07  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 1.352815 2.017376  
## sample estimates:  
## ratio of variances   
## 1.64671

t.test(x= desdata$Sug\_Dur[desdata$period == "P1"], y = desdata$Sug\_Dur[desdata$period == "P2"], paired = FALSE, var.equal = FALSE, conf.level = 0.95)

##   
## Welch Two Sample t-test  
##   
## data: desdata$Sug\_Dur[desdata$period == "P1"] and desdata$Sug\_Dur[desdata$period == "P2"]  
## t = 0.72883, df = 481.84, p-value = 0.4665  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.1201899 0.2619259  
## sample estimates:  
## mean of x mean of y   
## 7.077160 7.006293

desdata$FEV[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 264

summary(desdata$FEV[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 29.00 67.00 79.00 78.85 92.00 137.00 303

desdata$FEV[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 18.70239

desdata$FEV[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 160

summary(desdata$FEV[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 36.0 68.0 82.0 80.6 92.0 129.0 134

desdata$FEV[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 17.57614

desdata$FEV %>% na.omit() %>% length()

## [1] 424

summary(desdata$FEV)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 29.00 68.00 80.00 79.51 92.00 137.00 437

desdata$FEV %>% na.omit() %>% sd()

## [1] 18.28433

var.test(desdata$FEV[desdata$period == "P2"], desdata$FEV[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$FEV[desdata$period == "P2"] and desdata$FEV[desdata$period == "P1"]  
## F = 0.88319, num df = 159, denom df = 263, p-value = 0.3915  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.671235 1.173434  
## sample estimates:  
## ratio of variances   
## 0.8831878

t.test(x= desdata$FEV[desdata$period == "P1"], y = desdata$FEV[desdata$period == "P2"], paired = FALSE, var.equal = TRUE, conf.level = 0.95)

##   
## Two Sample t-test  
##   
## data: desdata$FEV[desdata$period == "P1"] and desdata$FEV[desdata$period == "P2"]  
## t = -0.95603, df = 422, p-value = 0.3396  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -5.352652 1.849622  
## sample estimates:  
## mean of x mean of y   
## 78.84848 80.60000

desdata$DLCO[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 114

summary(desdata$DLCO[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 14.00 50.00 60.00 59.82 70.75 97.00 453

desdata$DLCO[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 18.64269

desdata$DLCO[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 62

summary(desdata$DLCO[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 15.00 48.25 62.00 60.60 73.75 96.00 232

desdata$DLCO[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 19.72887

desdata$DLCO %>% na.omit() %>% length()

## [1] 176

summary(desdata$DLCO)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 14.00 49.00 60.00 60.09 73.00 97.00 685

desdata$DLCO %>% na.omit() %>% sd()

## [1] 18.97977

var.test(desdata$DLCO[desdata$period == "P2"], desdata$DLCO[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$DLCO[desdata$period == "P2"] and desdata$DLCO[desdata$period == "P1"]  
## F = 1.1199, num df = 61, denom df = 113, p-value = 0.5977  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.7293148 1.7716418  
## sample estimates:  
## ratio of variances   
## 1.11992

t.test(x= desdata$DLCO[desdata$period == "P1"], y = desdata$DLCO[desdata$period == "P2"], paired = FALSE, var.equal = TRUE, conf.level = 0.95)

##   
## Two Sample t-test  
##   
## data: desdata$DLCO[desdata$period == "P1"] and desdata$DLCO[desdata$period == "P2"]  
## t = -0.26007, df = 174, p-value = 0.7951  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -6.708031 5.146061  
## sample estimates:  
## mean of x mean of y   
## 59.81579 60.59677

desdata$TBU[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 567

summary(desdata$TBU[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0000 0.0000 0.5697 0.0000 20.0000

desdata$TBU[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 1.489064

desdata$TBU[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 294

summary(desdata$TBU[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0000 0.0000 0.6565 1.0000 14.0000

desdata$TBU[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 1.532696

desdata$TBU %>% na.omit() %>% length()

## [1] 861

summary(desdata$TBU)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0000 0.0000 0.5993 1.0000 20.0000

desdata$TBU %>% na.omit() %>% sd()

## [1] 1.503778

var.test(desdata$TBU[desdata$period == "P2"], desdata$TBU[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$TBU[desdata$period == "P2"] and desdata$TBU[desdata$period == "P1"]  
## F = 1.0595, num df = 293, denom df = 566, p-value = 0.5622  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.8703755 1.2979409  
## sample estimates:  
## ratio of variances   
## 1.059462

t.test(x= desdata$TBU[desdata$period == "P1"], y = desdata$TBU[desdata$period == "P2"], paired = FALSE, var.equal = TRUE, conf.level = 0.95)

##   
## Two Sample t-test  
##   
## data: desdata$TBU[desdata$period == "P1"] and desdata$TBU[desdata$period == "P2"]  
## t = -0.80297, df = 859, p-value = 0.4222  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.2989610 0.1253656  
## sample estimates:  
## mean of x mean of y   
## 0.5696649 0.6564626

desdata$XCT[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 541

summary(desdata$XCT[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 12.00 56.00 68.00 69.18 81.00 149.00 26

desdata$XCT[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 20.25255

desdata$XCT[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 277

summary(desdata$XCT[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 16.00 56.00 66.00 69.38 82.00 181.00 17

desdata$XCT[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 22.37339

desdata$XCT %>% na.omit() %>% length()

## [1] 818

summary(desdata$XCT)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 12.00 56.00 68.00 69.25 81.00 181.00 43

desdata$XCT %>% na.omit() %>% sd()

## [1] 20.98124

var.test(desdata$XCT[desdata$period == "P2"], desdata$XCT[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$XCT[desdata$period == "P2"] and desdata$XCT[desdata$period == "P1"]  
## F = 1.2204, num df = 276, denom df = 540, p-value = 0.05327  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.9972047 1.5039302  
## sample estimates:  
## ratio of variances   
## 1.220405

t.test(x= desdata$XCT[desdata$period == "P1"], y = desdata$XCT[desdata$period == "P2"], paired = FALSE, var.equal = TRUE, conf.level = 0.95)

##   
## Two Sample t-test  
##   
## data: desdata$XCT[desdata$period == "P1"] and desdata$XCT[desdata$period == "P2"]  
## t = -0.12646, df = 816, p-value = 0.8994  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -3.240708 2.848401  
## sample estimates:  
## mean of x mean of y   
## 69.17930 69.37545

desdata$CPBT[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 557

summary(desdata$CPBT[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 32.00 73.00 89.00 91.18 106.00 233.00 10

desdata$CPBT[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 27.90413

desdata$CPBT[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 289

summary(desdata$CPBT[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 34.00 72.00 87.00 92.16 106.00 232.00 5

desdata$CPBT[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 29.94644

desdata$CPBT %>% na.omit() %>% length()

## [1] 846

summary(desdata$CPBT)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 32.00 73.00 88.50 91.52 106.00 233.00 15

desdata$CPBT %>% na.omit() %>% sd()

## [1] 28.60428

var.test(desdata$CPBT[desdata$period == "P2"], desdata$CPBT[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$CPBT[desdata$period == "P2"] and desdata$CPBT[desdata$period == "P1"]  
## F = 1.1517, num df = 288, denom df = 556, p-value = 0.1628  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.9445758 1.4135405  
## sample estimates:  
## ratio of variances   
## 1.151738

t.test(x= desdata$CPBT[desdata$period == "P1"], y = desdata$CPBT[desdata$period == "P2"], paired = FALSE, var.equal = TRUE, conf.level = 0.95)

##   
## Two Sample t-test  
##   
## data: desdata$CPBT[desdata$period == "P1"] and desdata$CPBT[desdata$period == "P2"]  
## t = -0.47387, df = 844, p-value = 0.6357  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -5.055122 3.088929  
## sample estimates:  
## mean of x mean of y   
## 91.17953 92.16263

desdata$IHV[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 567

summary(desdata$IHV[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 4.27 6.20 10.14 9.70 260.42

desdata$IHV[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 17.8928

desdata$IHV[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 294

summary(desdata$IHV[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 4.853 6.835 9.457 11.197 78.570

desdata$IHV[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 9.301282

desdata$IHV %>% na.omit() %>% length()

## [1] 861

summary(desdata$IHV)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 4.450 6.370 9.904 10.370 260.420

desdata$IHV %>% na.omit() %>% sd()

## [1] 15.50109

var.test(desdata$IHV[desdata$period == "P2"], desdata$IHV[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$IHV[desdata$period == "P2"] and desdata$IHV[desdata$period == "P1"]  
## F = 0.27023, num df = 293, denom df = 566, p-value < 2.2e-16  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.2219988 0.3310540  
## sample estimates:  
## ratio of variances   
## 0.2702272

t.test(x= desdata$IHV[desdata$period == "P1"], y = desdata$IHV[desdata$period == "P2"], paired = FALSE, var.equal = FALSE, conf.level = 0.95)

##   
## Welch Two Sample t-test  
##   
## data: desdata$IHV[desdata$period == "P1"] and desdata$IHV[desdata$period == "P2"]  
## t = 0.73257, df = 858.99, p-value = 0.464  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.140078 2.497929  
## sample estimates:  
## mean of x mean of y   
## 10.135626 9.456701

desdata$IHV\_24[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 546

summary(desdata$IHV\_24[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.000 4.207 6.045 7.554 8.995 23.970 21

desdata$IHV\_24[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 4.719924

desdata$IHV\_24[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 284

summary(desdata$IHV\_24[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.000 4.815 6.615 7.992 10.520 21.920 10

desdata$IHV\_24[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 4.294708

desdata$IHV\_24 %>% na.omit() %>% length()

## [1] 830

summary(desdata$IHV\_24)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.000 4.350 6.240 7.704 9.607 23.970 31

desdata$IHV\_24 %>% na.omit() %>% sd()

## [1] 4.580988

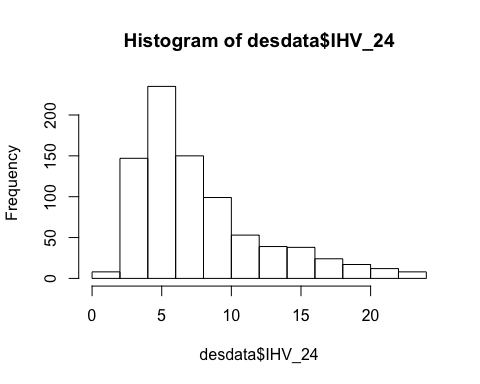
var.test(desdata$IHV\_24[desdata$period == "P2"], desdata$IHV\_24[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$IHV\_24[desdata$period == "P2"] and desdata$IHV\_24[desdata$period == "P1"]  
## F = 0.82794, num df = 283, denom df = 545, p-value = 0.07329  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.6777851 1.0180823  
## sample estimates:  
## ratio of variances   
## 0.8279368

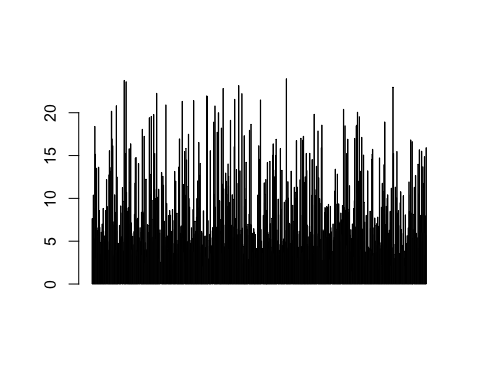
t.test(x= desdata$IHV\_24[desdata$period == "P1"], y = desdata$IHV\_24[desdata$period == "P2"], paired = FALSE, var.equal = TRUE, conf.level = 0.95)

##   
## Two Sample t-test  
##   
## data: desdata$IHV\_24[desdata$period == "P1"] and desdata$IHV\_24[desdata$period == "P2"]  
## t = -1.3067, df = 828, p-value = 0.1917  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.0953362 0.2198004  
## sample estimates:  
## mean of x mean of y   
## 7.553993 7.991761

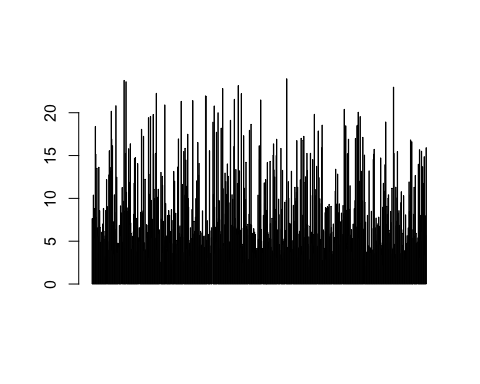
library(ggplot2)  
theme\_set(theme\_classic())  
  
hist(desdata$IHV\_24, breaks = 12, plot = TRUE, freq = TRUE)



barplot(desdata$IHV\_24,desdata$OREhour)

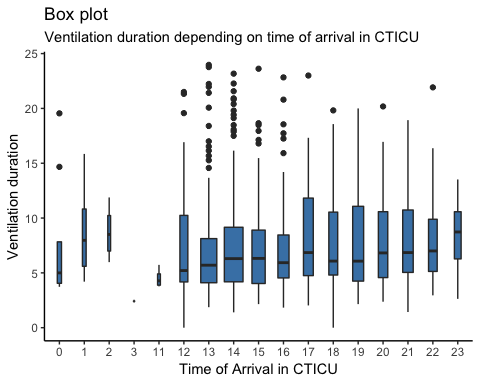


barplot(desdata$IHV\_24)



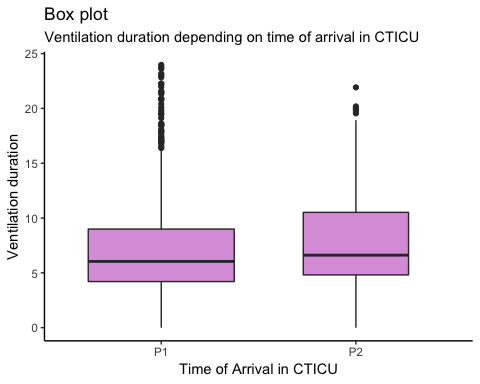
desdata$OREhour <- as.factor(desdata$OREhour)  
g <- ggplot(desdata, aes(desdata$OREhour, desdata$IHV\_24))  
g + geom\_boxplot(varwidth=T, fill="steelblue") +   
 labs(title="Box plot",   
 subtitle="Ventilation duration depending on time of arrival in CTICU",  
 x="Time of Arrival in CTICU",  
 y="Ventilation duration" )

## Warning: Removed 31 rows containing non-finite values (stat\_boxplot).

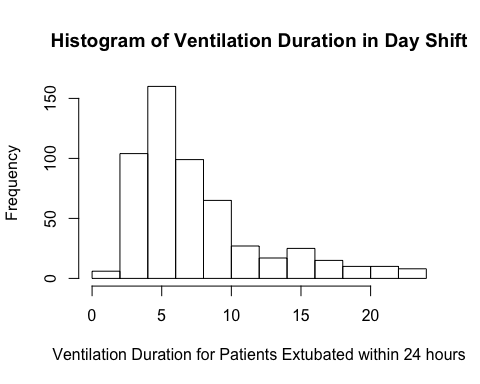


gg <- ggplot(desdata, aes(desdata$period, desdata$IHV\_24))  
gg + geom\_boxplot(varwidth=T, fill="plum") +   
 labs(title="Box plot",   
 subtitle="Ventilation duration depending on time of arrival in CTICU",  
 x="Time of Arrival in CTICU",  
 y="Ventilation duration" )

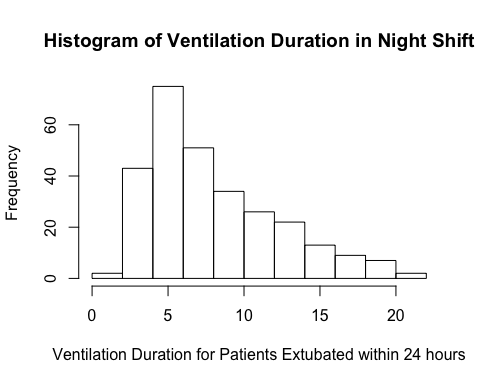
## Warning: Removed 31 rows containing non-finite values (stat\_boxplot).



par(mfrow=c(1,1))  
hist(desdata$IHV\_24[desdata$period == "P1"], main = "Histogram of Ventilation Duration in Day Shift", xlab = "Ventilation Duration for Patients Extubated within 24 hours")



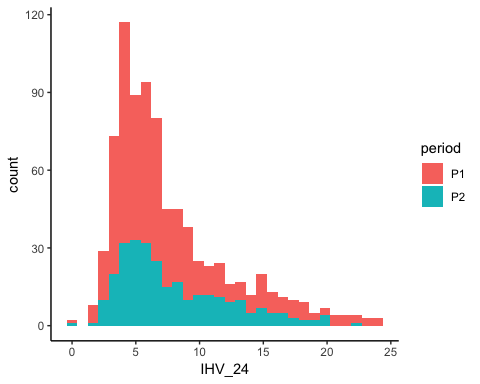
hist(desdata$IHV\_24[desdata$period == "P2"], main = "Histogram of Ventilation Duration in Night Shift", xlab = "Ventilation Duration for Patients Extubated within 24 hours")



g1 <- ggplot(desdata, aes(IHV\_24, fill=period))  
  
g1 + geom\_histogram()

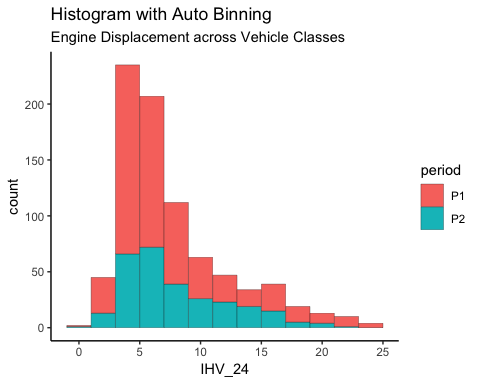
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 31 rows containing non-finite values (stat\_bin).



g1 + geom\_histogram(position = ,  
 binwidth = 2,   
 col="black",   
 size=.1) + # change binwidth  
 labs(title="Histogram with Auto Binning",   
 subtitle="Engine Displacement across Vehicle Classes")

## Warning: Removed 31 rows containing non-finite values (stat\_bin).



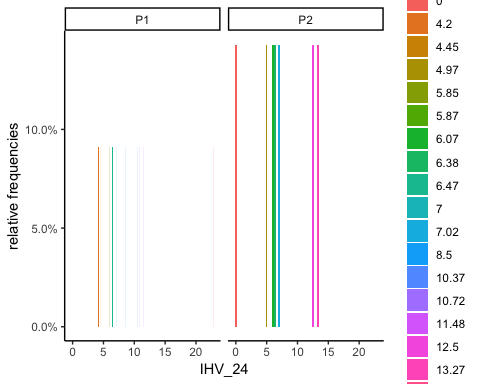
data <- desdata %>% na.omit(IHV\_24) %>% group\_by()  
library(scales)

##   
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':  
##   
## discard

## The following object is masked from 'package:readr':  
##   
## col\_factor

ggplot(data, aes(IHV\_24)) +   
 geom\_bar(aes(y = ..prop.., fill = factor(..x..)), stat="count") +   
 scale\_y\_continuous(labels=scales::percent) +  
 ylab("relative frequencies") +  
 facet\_grid(~period)



glmdata$period <- as.factor(glmdata$period) %>% relevel(glmdata$period, ref = "P1")  
simple\_logistic\_fit <- glm(formula= period ~ IHV\_24, family = binomial(link = "logit"),   
 data = glmdata)  
summary(simple\_logistic\_fit)

##   
## Call:  
## glm(formula = period ~ IHV\_24, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0446 -0.9095 -0.8879 1.4409 1.5364   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.81322 0.14336 -5.672 1.41e-08 \*\*\*  
## IHV\_24 0.02054 0.01574 1.305 0.192   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1066.5 on 829 degrees of freedom  
## Residual deviance: 1064.8 on 828 degrees of freedom  
## (31 observations deleted due to missingness)  
## AIC: 1068.8  
##   
## Number of Fisher Scoring iterations: 4

# Backward selection

## Full model

full\_logistic\_fit <- glm(formula= period ~ IHV\_24+ age+Ethnicity +BMI +PMM +PRM + PPV+ Reop+Urgent +Sug\_Dur +Pri\_MI +smoker+lung+ PFT +RFD + PAD+ POBPT+ TBU+ XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(full\_logistic\_fit)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + age + Ethnicity + BMI + PMM +   
## PRM + PPV + Reop + Urgent + Sug\_Dur + Pri\_MI + smoker + lung +   
## PFT + RFD + PAD + POBPT + TBU + XCT + CPBT, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7949 -0.9305 -0.8040 1.3127 2.1161   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.098e-02 1.328e+00 -0.038 0.9694   
## IHV\_24 1.842e-02 1.933e-02 0.953 0.3407   
## age 3.134e-05 9.300e-03 0.003 0.9973   
## EthnicityAmerican Indian -1.303e+01 5.354e+02 -0.024 0.9806   
## EthnicityAsian -1.062e+00 4.360e-01 -2.436 0.0148 \*  
## EthnicityBlack 9.645e-02 2.692e-01 0.358 0.7201   
## EthnicityHawaiian 3.069e-01 5.439e-01 0.564 0.5726   
## EthnicityOther -9.740e-02 7.139e-01 -0.136 0.8915   
## BMI -3.413e-03 1.765e-02 -0.193 0.8466   
## PMM 3.321e+00 2.616e+00 1.270 0.2042   
## PRM -1.942e+00 7.547e+00 -0.257 0.7969   
## PPV -3.977e+00 4.399e+00 -0.904 0.3660   
## ReopYes -2.567e-01 9.127e-01 -0.281 0.7786   
## UrgentUrgent 5.224e-02 1.832e-01 0.285 0.7754   
## Sug\_Dur -1.445e-01 8.430e-02 -1.713 0.0866 .  
## Pri\_MIYes 1.987e-01 1.787e-01 1.112 0.2660   
## smokerYes 1.132e-01 1.657e-01 0.683 0.4945   
## lungYes -2.653e-01 2.410e-01 -1.101 0.2710   
## PFTYes 3.070e-01 1.750e-01 1.754 0.0794 .  
## RFDYes -5.225e-01 6.484e-01 -0.806 0.4204   
## PADYes 2.578e-01 2.778e-01 0.928 0.3534   
## POBPTYes 2.600e-01 2.804e-01 0.928 0.3536   
## TBU -4.802e-02 1.070e-01 -0.449 0.6535   
## XCT -1.535e-02 9.253e-03 -1.659 0.0971 .  
## CPBT 1.449e-02 7.011e-03 2.067 0.0387 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 914.05 on 711 degrees of freedom  
## Residual deviance: 882.66 on 687 degrees of freedom  
## (149 observations deleted due to missingness)  
## AIC: 932.66  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_1 <- glm(formula= period ~ IHV\_24+Ethnicity +BMI +PMM +PRM + PPV+ Reop+Urgent +Sug\_Dur +Pri\_MI +smoker+lung+ PFT +RFD + PAD+ POBPT+ TBU+ XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_1)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + BMI + PMM + PRM +   
## PPV + Reop + Urgent + Sug\_Dur + Pri\_MI + smoker + lung +   
## PFT + RFD + PAD + POBPT + TBU + XCT + CPBT, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7951 -0.9306 -0.8040 1.3126 2.1161   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.048790 1.157036 -0.042 0.9664   
## IHV\_24 0.018424 0.019259 0.957 0.3387   
## EthnicityAmerican Indian -13.032451 535.411255 -0.024 0.9806   
## EthnicityAsian -1.062332 0.435663 -2.438 0.0148 \*  
## EthnicityBlack 0.096356 0.267563 0.360 0.7188   
## EthnicityHawaiian 0.306733 0.542561 0.565 0.5718   
## EthnicityOther -0.097319 0.713444 -0.136 0.8915   
## BMI -0.003425 0.017326 -0.198 0.8433   
## PMM 3.322392 2.597723 1.279 0.2009   
## PRM -1.939473 7.494042 -0.259 0.7958   
## PPV -3.977661 4.393352 -0.905 0.3653   
## ReopYes -0.256499 0.911548 -0.281 0.7784   
## UrgentUrgent 0.052214 0.182925 0.285 0.7753   
## Sug\_Dur -0.144449 0.084291 -1.714 0.0866 .  
## Pri\_MIYes 0.198646 0.177195 1.121 0.2623   
## smokerYes 0.113221 0.165743 0.683 0.4945   
## lungYes -0.265321 0.241010 -1.101 0.2710   
## PFTYes 0.307010 0.175008 1.754 0.0794 .  
## RFDYes -0.522808 0.640674 -0.816 0.4145   
## PADYes 0.257767 0.277141 0.930 0.3523   
## POBPTYes 0.260038 0.280344 0.928 0.3536   
## TBU -0.048014 0.106953 -0.449 0.6535   
## XCT -0.015351 0.009250 -1.660 0.0970 .  
## CPBT 0.014494 0.007001 2.070 0.0384 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 914.05 on 711 degrees of freedom  
## Residual deviance: 882.66 on 688 degrees of freedom  
## (149 observations deleted due to missingness)  
## AIC: 930.66  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_2 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM +PRM + PPV+ Reop+Urgent +Sug\_Dur +Pri\_MI +smoker+lung+ PFT +RFD + PAD+ POBPT+ TBU+ XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_2)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PRM + PPV +   
## Reop + Urgent + Sug\_Dur + Pri\_MI + smoker + lung + PFT +   
## RFD + PAD + POBPT + TBU + XCT + CPBT, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7951 -0.9315 -0.8059 1.3148 2.1210   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.128032 1.085540 -0.118 0.9061   
## IHV\_24 0.018195 0.019228 0.946 0.3440   
## EthnicityAmerican Indian -13.017495 535.411250 -0.024 0.9806   
## EthnicityAsian -1.050853 0.431680 -2.434 0.0149 \*  
## EthnicityBlack 0.100337 0.266780 0.376 0.7068   
## EthnicityHawaiian 0.316184 0.540386 0.585 0.5585   
## EthnicityOther -0.081772 0.709179 -0.115 0.9082   
## PMM 3.331520 2.599131 1.282 0.1999   
## PRM -1.378147 6.919564 -0.199 0.8421   
## PPV -4.186936 4.266415 -0.981 0.3264   
## ReopYes -0.258462 0.911494 -0.284 0.7767   
## UrgentUrgent 0.052437 0.182919 0.287 0.7744   
## Sug\_Dur -0.146692 0.083684 -1.753 0.0796 .  
## Pri\_MIYes 0.202189 0.176292 1.147 0.2514   
## smokerYes 0.113187 0.165744 0.683 0.4947   
## lungYes -0.262668 0.240604 -1.092 0.2750   
## PFTYes 0.307394 0.175002 1.757 0.0790 .  
## RFDYes -0.527162 0.640622 -0.823 0.4106   
## PADYes 0.257735 0.277099 0.930 0.3523   
## POBPTYes 0.258569 0.280159 0.923 0.3560   
## TBU -0.045862 0.106313 -0.431 0.6662   
## XCT -0.015361 0.009249 -1.661 0.0968 .  
## CPBT 0.014506 0.007001 2.072 0.0383 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 914.05 on 711 degrees of freedom  
## Residual deviance: 882.70 on 689 degrees of freedom  
## (149 observations deleted due to missingness)  
## AIC: 928.7  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_3 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV+ Reop+Urgent +Sug\_Dur +Pri\_MI +smoker+lung+ PFT +RFD + PAD+ POBPT+ TBU+ XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_3)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Reop +   
## Urgent + Sug\_Dur + Pri\_MI + smoker + lung + PFT + RFD + PAD +   
## POBPT + TBU + XCT + CPBT, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8005 -0.9338 -0.8053 1.3151 2.1168   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.128057 1.084792 -0.118 0.9060   
## IHV\_24 0.017916 0.019181 0.934 0.3503   
## EthnicityAmerican Indian -13.011194 535.411249 -0.024 0.9806   
## EthnicityAsian -1.043407 0.430034 -2.426 0.0153 \*  
## EthnicityBlack 0.111017 0.261353 0.425 0.6710   
## EthnicityHawaiian 0.317652 0.540248 0.588 0.5565   
## EthnicityOther -0.069131 0.706504 -0.098 0.9221   
## PMM 3.338672 2.600444 1.284 0.1992   
## PPV -4.606531 3.719870 -1.238 0.2156   
## ReopYes -0.251834 0.910195 -0.277 0.7820   
## UrgentUrgent 0.055205 0.182375 0.303 0.7621   
## Sug\_Dur -0.146940 0.083679 -1.756 0.0791 .  
## Pri\_MIYes 0.204821 0.175765 1.165 0.2439   
## smokerYes 0.113645 0.165729 0.686 0.4929   
## lungYes -0.260780 0.240389 -1.085 0.2780   
## PFTYes 0.308114 0.174968 1.761 0.0782 .  
## RFDYes -0.536909 0.641092 -0.837 0.4023   
## PADYes 0.254841 0.276683 0.921 0.3570   
## POBPTYes 0.258582 0.280110 0.923 0.3559   
## TBU -0.045509 0.106259 -0.428 0.6684   
## XCT -0.015184 0.009207 -1.649 0.0991 .  
## CPBT 0.014384 0.006975 2.062 0.0392 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 914.05 on 711 degrees of freedom  
## Residual deviance: 882.74 on 690 degrees of freedom  
## (149 observations deleted due to missingness)  
## AIC: 926.74  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_4 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV+Urgent +Sug\_Dur +Pri\_MI +smoker+lung+ PFT +RFD + PAD+ POBPT+ TBU+ XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_4)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Urgent +   
## Sug\_Dur + Pri\_MI + smoker + lung + PFT + RFD + PAD + POBPT +   
## TBU + XCT + CPBT, family = binomial(link = "logit"), data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7899 -0.9338 -0.8058 1.3141 2.1157   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.387649 0.541941 -0.715 0.4744   
## IHV\_24 0.017728 0.019167 0.925 0.3550   
## EthnicityAmerican Indian -13.010956 535.411249 -0.024 0.9806   
## EthnicityAsian -1.040192 0.429658 -2.421 0.0155 \*  
## EthnicityBlack 0.108615 0.261178 0.416 0.6775   
## EthnicityHawaiian 0.316473 0.540286 0.586 0.5580   
## EthnicityOther -0.072245 0.706416 -0.102 0.9185   
## PMM 3.295089 2.590479 1.272 0.2034   
## PPV -4.529104 3.708091 -1.221 0.2219   
## UrgentUrgent 0.053319 0.182265 0.293 0.7699   
## Sug\_Dur -0.143984 0.082771 -1.740 0.0819 .  
## Pri\_MIYes 0.203957 0.175701 1.161 0.2457   
## smokerYes 0.114316 0.165718 0.690 0.4903   
## lungYes -0.257990 0.240136 -1.074 0.2827   
## PFTYes 0.306613 0.174844 1.754 0.0795 .  
## RFDYes -0.545876 0.640770 -0.852 0.3943   
## PADYes 0.256338 0.276659 0.927 0.3542   
## POBPTYes 0.256465 0.280202 0.915 0.3600   
## TBU -0.044533 0.106347 -0.419 0.6754   
## XCT -0.015477 0.009151 -1.691 0.0908 .  
## CPBT 0.014514 0.006959 2.085 0.0370 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 914.05 on 711 degrees of freedom  
## Residual deviance: 882.81 on 691 degrees of freedom  
## (149 observations deleted due to missingness)  
## AIC: 924.81  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_5 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV +Sug\_Dur +Pri\_MI +smoker+lung+ PFT +RFD + PAD+ POBPT+ TBU+ XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_5)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Sug\_Dur +   
## Pri\_MI + smoker + lung + PFT + RFD + PAD + POBPT + TBU +   
## XCT + CPBT, family = binomial(link = "logit"), data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7902 -0.9314 -0.8011 1.3155 2.1248   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.350573 0.527399 -0.665 0.5062   
## IHV\_24 0.017764 0.019168 0.927 0.3541   
## EthnicityAmerican Indian -13.043270 535.411237 -0.024 0.9806   
## EthnicityAsian -1.036675 0.429528 -2.414 0.0158 \*  
## EthnicityBlack 0.107316 0.261171 0.411 0.6811   
## EthnicityHawaiian 0.315848 0.540015 0.585 0.5586   
## EthnicityOther -0.082879 0.705813 -0.117 0.9065   
## PMM 3.218781 2.574573 1.250 0.2112   
## PPV -4.324157 3.636112 -1.189 0.2344   
## Sug\_Dur -0.145044 0.082822 -1.751 0.0799 .  
## Pri\_MIYes 0.210520 0.174281 1.208 0.2271   
## smokerYes 0.110626 0.165209 0.670 0.5031   
## lungYes -0.268634 0.237389 -1.132 0.2578   
## PFTYes 0.313340 0.173340 1.808 0.0707 .  
## RFDYes -0.562581 0.638576 -0.881 0.3783   
## PADYes 0.252820 0.276308 0.915 0.3602   
## POBPTYes 0.257825 0.280055 0.921 0.3572   
## TBU -0.044499 0.106316 -0.419 0.6755   
## XCT -0.015430 0.009147 -1.687 0.0916 .  
## CPBT 0.014443 0.006953 2.077 0.0378 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 914.05 on 711 degrees of freedom  
## Residual deviance: 882.90 on 692 degrees of freedom  
## (149 observations deleted due to missingness)  
## AIC: 922.9  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_6 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV +Sug\_Dur +Pri\_MI +smoker+lung+ PFT +RFD + PAD+ POBPT+ XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_6)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Sug\_Dur +   
## Pri\_MI + smoker + lung + PFT + RFD + PAD + POBPT + XCT +   
## CPBT, family = binomial(link = "logit"), data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7656 -0.9314 -0.8012 1.3150 2.1258   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.331642 0.525816 -0.631 0.5282   
## IHV\_24 0.017543 0.019150 0.916 0.3596   
## EthnicityAmerican Indian -13.044242 535.411237 -0.024 0.9806   
## EthnicityAsian -1.039186 0.429816 -2.418 0.0156 \*  
## EthnicityBlack 0.110260 0.261020 0.422 0.6727   
## EthnicityHawaiian 0.312231 0.540335 0.578 0.5634   
## EthnicityOther -0.077227 0.705635 -0.109 0.9129   
## PMM 3.184497 2.567381 1.240 0.2148   
## PPV -4.297490 3.624697 -1.186 0.2358   
## Sug\_Dur -0.147851 0.082719 -1.787 0.0739 .  
## Pri\_MIYes 0.209168 0.174227 1.201 0.2299   
## smokerYes 0.109685 0.165168 0.664 0.5066   
## lungYes -0.268694 0.237270 -1.132 0.2574   
## PFTYes 0.311532 0.173225 1.798 0.0721 .  
## RFDYes -0.564724 0.637534 -0.886 0.3757   
## PADYes 0.258810 0.275825 0.938 0.3481   
## POBPTYes 0.171253 0.190535 0.899 0.3688   
## XCT -0.015566 0.009144 -1.702 0.0887 .  
## CPBT 0.014604 0.006946 2.103 0.0355 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 914.05 on 711 degrees of freedom  
## Residual deviance: 883.08 on 693 degrees of freedom  
## (149 observations deleted due to missingness)  
## AIC: 921.08  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_7 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV +Sug\_Dur +Pri\_MI +lung+ PFT +RFD + PAD+ POBPT+ XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_7)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Sug\_Dur +   
## Pri\_MI + lung + PFT + RFD + PAD + POBPT + XCT + CPBT, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7442 -0.9380 -0.8098 1.3123 2.1379   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.143286 0.512997 -0.279 0.7800   
## IHV\_24 0.009180 0.018623 0.493 0.6221   
## EthnicityAmerican Indian -13.031111 535.411228 -0.024 0.9806   
## EthnicityAsian -1.023792 0.404236 -2.533 0.0113 \*  
## EthnicityBlack 0.074516 0.256775 0.290 0.7717   
## EthnicityHawaiian 0.263932 0.539427 0.489 0.6246   
## EthnicityOther -0.139167 0.703105 -0.198 0.8431   
## PMM 3.021573 2.527936 1.195 0.2320   
## PPV -3.716716 3.529310 -1.053 0.2923   
## Sug\_Dur -0.159771 0.082985 -1.925 0.0542 .  
## Pri\_MIYes 0.267820 0.169524 1.580 0.1141   
## lungYes -0.310438 0.234914 -1.321 0.1863   
## PFTYes 0.304050 0.170343 1.785 0.0743 .  
## RFDYes -0.392657 0.581817 -0.675 0.4998   
## PADYes 0.218831 0.274157 0.798 0.4248   
## POBPTYes 0.210924 0.186767 1.129 0.2588   
## XCT -0.016704 0.008922 -1.872 0.0612 .  
## CPBT 0.015328 0.006786 2.259 0.0239 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 948.9 on 737 degrees of freedom  
## Residual deviance: 917.4 on 720 degrees of freedom  
## (123 observations deleted due to missingness)  
## AIC: 953.4  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_8 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV +Sug\_Dur +Pri\_MI +smoker+lung+ PFT + PAD+ POBPT+ XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_8)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Sug\_Dur +   
## Pri\_MI + smoker + lung + PFT + PAD + POBPT + XCT + CPBT,   
## family = binomial(link = "logit"), data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8250 -0.9298 -0.8016 1.3145 2.1273   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.352952 0.525074 -0.672 0.5015   
## IHV\_24 0.017791 0.019087 0.932 0.3513   
## EthnicityAmerican Indian -13.030540 535.411237 -0.024 0.9806   
## EthnicityAsian -1.045147 0.429592 -2.433 0.0150 \*  
## EthnicityBlack 0.119964 0.260932 0.460 0.6457   
## EthnicityHawaiian 0.333065 0.538852 0.618 0.5365   
## EthnicityOther -0.185013 0.692602 -0.267 0.7894   
## PMM 3.488194 2.588678 1.347 0.1778   
## PPV -5.074138 3.559182 -1.426 0.1540   
## Sug\_Dur -0.145067 0.082483 -1.759 0.0786 .  
## Pri\_MIYes 0.223952 0.173735 1.289 0.1974   
## smokerYes 0.112741 0.165018 0.683 0.4945   
## lungYes -0.286639 0.236678 -1.211 0.2259   
## PFTYes 0.315870 0.173179 1.824 0.0682 .  
## PADYes 0.256718 0.275592 0.932 0.3516   
## POBPTYes 0.169780 0.190311 0.892 0.3723   
## XCT -0.014725 0.009113 -1.616 0.1061   
## CPBT 0.013986 0.006918 2.022 0.0432 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 915.72 on 713 degrees of freedom  
## Residual deviance: 885.60 on 696 degrees of freedom  
## (147 observations deleted due to missingness)  
## AIC: 921.6  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_9 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV +Sug\_Dur +Pri\_MI +smoker+lung+ PFT + PAD+ XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_9)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Sug\_Dur +   
## Pri\_MI + smoker + lung + PFT + PAD + XCT + CPBT, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8193 -0.9319 -0.8103 1.3186 2.1106   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.380134 0.524402 -0.725 0.4685   
## IHV\_24 0.020157 0.018902 1.066 0.2862   
## EthnicityAmerican Indian -13.091181 535.411232 -0.024 0.9805   
## EthnicityAsian -1.038777 0.429255 -2.420 0.0155 \*  
## EthnicityBlack 0.121069 0.260854 0.464 0.6426   
## EthnicityHawaiian 0.350804 0.537382 0.653 0.5139   
## EthnicityOther -0.190761 0.692646 -0.275 0.7830   
## PMM 3.677025 2.570187 1.431 0.1525   
## PPV -5.142650 3.542675 -1.452 0.1466   
## Sug\_Dur -0.140969 0.082341 -1.712 0.0869 .  
## Pri\_MIYes 0.217791 0.173473 1.255 0.2093   
## smokerYes 0.117083 0.164864 0.710 0.4776   
## lungYes -0.294837 0.236701 -1.246 0.2129   
## PFTYes 0.319430 0.173048 1.846 0.0649 .  
## PADYes 0.266913 0.275346 0.969 0.3324   
## XCT -0.014907 0.009102 -1.638 0.1015   
## CPBT 0.014177 0.006911 2.051 0.0402 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 915.72 on 713 degrees of freedom  
## Residual deviance: 886.39 on 697 degrees of freedom  
## (147 observations deleted due to missingness)  
## AIC: 920.39  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_10 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV +Sug\_Dur +Pri\_MI +smoker+lung+ PFT + XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_10)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Sug\_Dur +   
## Pri\_MI + smoker + lung + PFT + XCT + CPBT, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7876 -0.9270 -0.8131 1.3245 2.1158   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.403499 0.523022 -0.771 0.4404   
## IHV\_24 0.020310 0.018897 1.075 0.2825   
## EthnicityAmerican Indian -13.118094 535.411231 -0.025 0.9805   
## EthnicityAsian -1.064953 0.428437 -2.486 0.0129 \*  
## EthnicityBlack 0.131788 0.260202 0.506 0.6125   
## EthnicityHawaiian 0.374704 0.536139 0.699 0.4846   
## EthnicityOther -0.221278 0.691870 -0.320 0.7491   
## PMM 3.984627 2.547679 1.564 0.1178   
## PPV -5.284224 3.536950 -1.494 0.1352   
## Sug\_Dur -0.138687 0.082102 -1.689 0.0912 .  
## Pri\_MIYes 0.207219 0.173026 1.198 0.2311   
## smokerYes 0.120054 0.164754 0.729 0.4662   
## lungYes -0.292733 0.236545 -1.238 0.2159   
## PFTYes 0.323128 0.172861 1.869 0.0616 .  
## XCT -0.014789 0.009094 -1.626 0.1039   
## CPBT 0.014156 0.006907 2.050 0.0404 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 915.72 on 713 degrees of freedom  
## Residual deviance: 887.32 on 698 degrees of freedom  
## (147 observations deleted due to missingness)  
## AIC: 919.32  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_11 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV +Sug\_Dur +smoker+lung+ PFT + XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_11)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Sug\_Dur +   
## smoker + lung + PFT + XCT + CPBT, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8635 -0.9300 -0.8207 1.3376 2.1146   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.396655 0.522299 -0.759 0.44759   
## IHV\_24 0.020043 0.018886 1.061 0.28857   
## EthnicityAmerican Indian -13.172710 535.411228 -0.025 0.98037   
## EthnicityAsian -1.121756 0.425962 -2.633 0.00845 \*\*  
## EthnicityBlack 0.122321 0.259496 0.471 0.63737   
## EthnicityHawaiian 0.404916 0.535844 0.756 0.44985   
## EthnicityOther -0.227599 0.691302 -0.329 0.74198   
## PMM 4.128494 2.560196 1.613 0.10684   
## PPV -5.036011 3.548307 -1.419 0.15582   
## Sug\_Dur -0.135423 0.082090 -1.650 0.09901 .   
## smokerYes 0.136097 0.163690 0.831 0.40573   
## lungYes -0.307793 0.235584 -1.307 0.19138   
## PFTYes 0.326106 0.172186 1.894 0.05824 .   
## XCT -0.013251 0.009017 -1.470 0.14168   
## CPBT 0.013036 0.006862 1.900 0.05746 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 921.2 on 718 degrees of freedom  
## Residual deviance: 893.9 on 704 degrees of freedom  
## (142 observations deleted due to missingness)  
## AIC: 923.9  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_12 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV +Sug\_Dur +smoker+ PFT + XCT+ CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_12)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Sug\_Dur +   
## smoker + PFT + XCT + CPBT, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.8569 -0.9248 -0.8276 1.3409 2.0953   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.417065 0.519313 -0.803 0.42191   
## IHV\_24 0.021698 0.018849 1.151 0.24967   
## EthnicityAmerican Indian -13.170870 535.411228 -0.025 0.98037   
## EthnicityAsian -1.096954 0.424945 -2.581 0.00984 \*\*  
## EthnicityBlack 0.145076 0.258633 0.561 0.57484   
## EthnicityHawaiian 0.433286 0.535384 0.809 0.41834   
## EthnicityOther -0.208213 0.690292 -0.302 0.76293   
## PMM 4.256882 2.595377 1.640 0.10097   
## PPV -5.835709 3.549611 -1.644 0.10017   
## Sug\_Dur -0.131882 0.081774 -1.613 0.10680   
## smokerYes 0.104311 0.162439 0.642 0.52077   
## PFTYes 0.266462 0.163494 1.630 0.10314   
## XCT -0.012303 0.008971 -1.371 0.17024   
## CPBT 0.012425 0.006825 1.820 0.06870 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 922.85 on 720 degrees of freedom  
## Residual deviance: 897.40 on 707 degrees of freedom  
## (140 observations deleted due to missingness)  
## AIC: 925.4  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_12 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV +Sug\_Dur +smoker+ PFT + CPBT, family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_12)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Sug\_Dur +   
## smoker + PFT + CPBT, family = binomial(link = "logit"), data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.7196 -0.9437 -0.8265 1.3626 1.9837   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.589467 0.503409 -1.171 0.2416   
## IHV\_24 0.017912 0.018264 0.981 0.3267   
## EthnicityAmerican Indian -13.011416 535.411224 -0.024 0.9806   
## EthnicityAsian -0.857839 0.382775 -2.241 0.0250 \*  
## EthnicityBlack 0.213153 0.251095 0.849 0.3959   
## EthnicityHawaiian 0.576248 0.514445 1.120 0.2627   
## EthnicityOther -0.234353 0.689384 -0.340 0.7339   
## PMM 3.416637 2.439685 1.400 0.1614   
## PPV -3.894482 3.243703 -1.201 0.2299   
## Sug\_Dur -0.109697 0.079577 -1.379 0.1680   
## smokerYes 0.139868 0.158748 0.881 0.3783   
## PFTYes 0.304739 0.159498 1.911 0.0561 .  
## CPBT 0.002602 0.003251 0.800 0.4235   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 954.40 on 742 degrees of freedom  
## Residual deviance: 932.48 on 730 degrees of freedom  
## (118 observations deleted due to missingness)  
## AIC: 958.48  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_13 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM + PPV +Sug\_Dur +smoker+ PFT , family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_13)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + PPV + Sug\_Dur +   
## smoker + PFT, family = binomial(link = "logit"), data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.5831 -0.9400 -0.8298 1.3702 1.9983   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.59499 0.48595 -1.224 0.2208   
## IHV\_24 0.01913 0.01797 1.065 0.2871   
## EthnicityAmerican Indian -13.02008 535.41122 -0.024 0.9806   
## EthnicityAsian -0.87965 0.38137 -2.307 0.0211 \*  
## EthnicityBlack 0.19830 0.24948 0.795 0.4267   
## EthnicityHawaiian 0.58731 0.51346 1.144 0.2527   
## EthnicityOther -0.25167 0.68837 -0.366 0.7147   
## PMM 3.05095 2.34060 1.303 0.1924   
## PPV -3.36541 3.13076 -1.075 0.2824   
## Sug\_Dur -0.07542 0.06469 -1.166 0.2437   
## smokerYes 0.16261 0.15744 1.033 0.3017   
## PFTYes 0.26360 0.15793 1.669 0.0951 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 969.19 on 755 degrees of freedom  
## Residual deviance: 948.54 on 744 degrees of freedom  
## (105 observations deleted due to missingness)  
## AIC: 972.54  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_13 <- glm(formula= period ~ IHV\_24+Ethnicity +PMM +Sug\_Dur +smoker+ PFT , family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_13)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + PMM + Sug\_Dur + smoker +   
## PFT, family = binomial(link = "logit"), data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.2312 -0.9420 -0.8331 1.3739 2.0025   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.55359 0.48380 -1.144 0.2525   
## IHV\_24 0.01752 0.01790 0.979 0.3276   
## EthnicityAmerican Indian -12.95917 535.41122 -0.024 0.9807   
## EthnicityAsian -0.90446 0.38147 -2.371 0.0177 \*  
## EthnicityBlack 0.19184 0.24923 0.770 0.4415   
## EthnicityHawaiian 0.59234 0.51339 1.154 0.2486   
## EthnicityOther -0.29369 0.68809 -0.427 0.6695   
## PMM 0.72203 0.83850 0.861 0.3892   
## Sug\_Dur -0.07531 0.06462 -1.165 0.2438   
## smokerYes 0.15571 0.15718 0.991 0.3219   
## PFTYes 0.25858 0.15777 1.639 0.1012   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 969.19 on 755 degrees of freedom  
## Residual deviance: 949.76 on 745 degrees of freedom  
## (105 observations deleted due to missingness)  
## AIC: 971.76  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_13 <- glm(formula= period ~ IHV\_24+Ethnicity +Sug\_Dur +smoker+ PFT , family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_13)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + Sug\_Dur + smoker +   
## PFT, family = binomial(link = "logit"), data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.2585 -0.9426 -0.8342 1.3722 1.9770   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.52491 0.48160 -1.090 0.2758   
## IHV\_24 0.02211 0.01707 1.295 0.1952   
## EthnicityAmerican Indian -12.98058 535.41122 -0.024 0.9807   
## EthnicityAsian -0.88303 0.37993 -2.324 0.0201 \*  
## EthnicityBlack 0.22867 0.24540 0.932 0.3514   
## EthnicityHawaiian 0.60751 0.51262 1.185 0.2360   
## EthnicityOther -0.28021 0.68772 -0.407 0.6837   
## Sug\_Dur -0.07430 0.06445 -1.153 0.2490   
## smokerYes 0.16378 0.15683 1.044 0.2963   
## PFTYes 0.27820 0.15606 1.783 0.0746 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 969.19 on 755 degrees of freedom  
## Residual deviance: 950.50 on 746 degrees of freedom  
## (105 observations deleted due to missingness)  
## AIC: 970.5  
##   
## Number of Fisher Scoring iterations: 12

back\_logistic\_fit\_13 <- glm(formula= period ~ IHV\_24+Ethnicity +smoker+ PFT , family = binomial(link = "logit"), data = glmdata)  
summary(back\_logistic\_fit\_13)

##   
## Call:  
## glm(formula = period ~ IHV\_24 + Ethnicity + smoker + PFT, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.2652 -0.9423 -0.8288 1.3810 1.9764   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.03533 0.19332 -5.355 8.54e-08 \*\*\*  
## IHV\_24 0.02095 0.01702 1.231 0.2184   
## EthnicityAmerican Indian -12.88518 535.41121 -0.024 0.9808   
## EthnicityAsian -0.88995 0.37952 -2.345 0.0190 \*   
## EthnicityBlack 0.23497 0.24524 0.958 0.3380   
## EthnicityHawaiian 0.62909 0.51221 1.228 0.2194   
## EthnicityOther -0.32861 0.68548 -0.479 0.6317   
## smokerYes 0.16463 0.15667 1.051 0.2934   
## PFTYes 0.27138 0.15579 1.742 0.0815 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 969.19 on 755 degrees of freedom  
## Residual deviance: 951.86 on 747 degrees of freedom  
## (105 observations deleted due to missingness)  
## AIC: 969.86  
##   
## Number of Fisher Scoring iterations: 12

exp(5\*0.02054 )

## [1] 1.108159

# Secondary outcome des

desdata$Hos\_LOS[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 567

summary(desdata$Hos\_LOS[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.00 6.00 8.00 10.32 12.00 120.00

desdata$Hos\_LOS[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 9.118375

desdata$Hos\_LOS[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 294

summary(desdata$Hos\_LOS[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.00 7.00 9.00 11.03 13.00 54.00

desdata$Hos\_LOS[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 6.913094

desdata$Hos\_LOS %>% na.omit() %>% length()

## [1] 861

summary(desdata$Hos\_LOS)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.00 6.00 8.00 10.56 12.00 120.00

desdata$Hos\_LOS %>% na.omit() %>% sd()

## [1] 8.433024

var.test(desdata$Hos\_LOS[desdata$period == "P2"], desdata$Hos\_LOS[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$Hos\_LOS[desdata$period == "P2"] and desdata$Hos\_LOS[desdata$period == "P1"]  
## F = 0.57479, num df = 293, denom df = 566, p-value = 1.571e-07  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.4722060 0.7041736  
## sample estimates:  
## ratio of variances   
## 0.5747911

t.test(x= desdata$Hos\_LOS[desdata$period == "P1"], y = desdata$Hos\_LOS[desdata$period == "P2"], paired = FALSE, var.equal = FALSE, conf.level = 0.95)

##   
## Welch Two Sample t-test  
##   
## data: desdata$Hos\_LOS[desdata$period == "P1"] and desdata$Hos\_LOS[desdata$period == "P2"]  
## t = -1.273, df = 745.86, p-value = 0.2034  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.7994741 0.3837523  
## sample estimates:  
## mean of x mean of y   
## 10.32275 11.03061

desdata$ICU\_LOS[desdata$period == "P1"] %>% na.omit() %>% length()

## [1] 566

summary(desdata$ICU\_LOS[desdata$period == "P1"])

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 4.00 25.00 46.75 70.78 75.00 1175.00 1

desdata$ICU\_LOS[desdata$period == "P1"] %>% na.omit() %>% sd()

## [1] 98.44843

desdata$ICU\_LOS[desdata$period == "P2"] %>% na.omit() %>% length()

## [1] 294

summary(desdata$ICU\_LOS[desdata$period == "P2"])

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 12.00 22.62 47.75 66.11 87.97 504.00

desdata$ICU\_LOS[desdata$period == "P2"] %>% na.omit() %>% sd()

## [1] 60.04385

desdata$ICU\_LOS %>% na.omit() %>% length()

## [1] 860

summary(desdata$ICU\_LOS)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 4.00 24.50 47.00 69.18 80.40 1175.00 1

desdata$ICU\_LOS %>% na.omit() %>% sd()

## [1] 87.23273

var.test(desdata$ICU\_LOS[desdata$period == "P2"], desdata$ICU\_LOS[desdata$period == "P1"])

##   
## F test to compare two variances  
##   
## data: desdata$ICU\_LOS[desdata$period == "P2"] and desdata$ICU\_LOS[desdata$period == "P1"]  
## F = 0.37198, num df = 293, denom df = 565, p-value < 2.2e-16  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.3055710 0.4557354  
## sample estimates:  
## ratio of variances   
## 0.3719799

t.test(x= desdata$ICU\_LOS[desdata$period == "P1"], y = desdata$ICU\_LOS[desdata$period == "P2"], paired = FALSE, var.equal = FALSE, conf.level = 0.95)

##   
## Welch Two Sample t-test  
##   
## data: desdata$ICU\_LOS[desdata$period == "P1"] and desdata$ICU\_LOS[desdata$period == "P2"]  
## t = 0.86221, df = 836.62, p-value = 0.3888  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -5.966284 15.314215  
## sample estimates:  
## mean of x mean of y   
## 70.78145 66.10748

logistic\_fit\_ICU\_LOS <- glm(formula= period ~ ICU\_LOS+Ethnicity +smoker+ PFT , family = binomial(link = "logit"), data = glmdata)  
summary(logistic\_fit\_ICU\_LOS)

##   
## Call:  
## glm(formula = period ~ ICU\_LOS + Ethnicity + smoker + PFT, family = binomial(link = "logit"),   
## data = glmdata)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.3291 -0.9389 -0.8415 1.3816 1.9150   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -8.307e-01 1.536e-01 -5.409 6.33e-08 \*\*\*  
## ICU\_LOS -8.974e-04 9.924e-04 -0.904 0.3659   
## EthnicityAmerican Indian -1.271e+01 5.354e+02 -0.024 0.9811   
## EthnicityAsian -7.501e-01 3.632e-01 -2.065 0.0389 \*   
## EthnicityBlack 2.184e-01 2.399e-01 0.910 0.3627   
## EthnicityHawaiian 7.959e-01 4.970e-01 1.602 0.1093   
## EthnicityOther -3.381e-02 6.198e-01 -0.055 0.9565   
## smokerYes 1.666e-01 1.544e-01 1.079 0.2805   
## PFTYes 2.826e-01 1.531e-01 1.846 0.0649 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 1005.73 on 783 degrees of freedom  
## Residual deviance: 989.98 on 775 degrees of freedom  
## (77 observations deleted due to missingness)  
## AIC: 1008  
##   
## Number of Fisher Scoring iterations: 12

#glmdata$Reintub[desdata$Reintub == "Unknown"] <- NA  
  
length(glmdata$Reintub)

## [1] 861

sum(is.na(glmdata$Reintub))

## [1] 0

sum(is.na(glmdata$Reintub)) / length(glmdata$Reintub)

## [1] 0

table(glmdata$Reintub)

##   
## No Yes   
## 837 24

#table(glmdata$Reintub) / length(glmdata$Reintub)  
paste(round(100\*(table(glmdata$Reintub) / 861), 2), "%", sep="")

## [1] "97.21%" "2.79%"

length(glmdata$Reintub[glmdata$period == "P1"])

## [1] 567

length(glmdata$Reintub[glmdata$period == "P2"])

## [1] 294

sum(is.na(glmdata$Reintub[glmdata$period == "P1"]))

## [1] 0

sum(is.na(glmdata$Reintub[glmdata$period == "P2"]))

## [1] 0

table(glmdata$Reintub[glmdata$period == "P1"])

##   
## No Yes   
## 549 18

table(glmdata$Reintub[glmdata$period == "P2"])

##   
## No Yes   
## 288 6

#table(glmdata$Reintub[glmdata$period == "P1"]) / length(glmdata$Reintub[glmdata$period == "P1"])  
paste(round(100\*(table(glmdata$Reintub[glmdata$period == "P1"]) / length(glmdata$Reintub[glmdata$period == "P1"])), 2), "%", sep="")

## [1] "96.83%" "3.17%"

#table(glmdata$Reintub[glmdata$period == "P2"]) / length(glmdata$Reintub[glmdata$period == "P2"])  
paste(round(100\*(table(glmdata$Reintub[glmdata$period == "P2"]) / length(glmdata$Reintub[glmdata$period == "P2"])), 2), "%", sep="")

## [1] "97.96%" "2.04%"

chisq.test(glmdata$Reintub,glmdata$period, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: glmdata$Reintub and glmdata$period  
## X-squared = 0.91846, df = 1, p-value = 0.3379

#glmdata$Re\_ICU [desdata$Re\_ICU != "No" | desdata$Re\_ICU !="Unknown"] <- "Yes"  
#glmdata$Re\_ICU [desdata$Re\_ICU == "No"] <- "No"  
#glmdata$Re\_ICU [desdata$Re\_ICU == "Unknown"] <- NA  
  
  
length(glmdata$Re\_ICU )

## [1] 861

sum(is.na(glmdata$Re\_ICU ))

## [1] 0

sum(is.na(glmdata$Re\_ICU )) / length(glmdata$Re\_ICU )

## [1] 0

table(glmdata$Re\_ICU )

##   
## No Yes   
## 843 18

#table(glmdata$Re\_ICU ) / length(glmdata$Re\_ICU )  
paste(round(100\*(table(glmdata$Re\_ICU ) / (length(na.omit(glmdata$Re\_ICU )))), 2), "%", sep="")

## [1] "97.91%" "2.09%"

length(na.omit(glmdata$Re\_ICU [glmdata$period == "P1"]))

## [1] 567

length(na.omit(glmdata$Re\_ICU [glmdata$period == "P2"]))

## [1] 294

sum(is.na(glmdata$Re\_ICU [glmdata$period == "P1"]))

## [1] 0

sum(is.na(glmdata$Re\_ICU [glmdata$period == "P2"]))

## [1] 0

table(glmdata$Re\_ICU [glmdata$period == "P1"])

##   
## No Yes   
## 556 11

table(glmdata$Re\_ICU [glmdata$period == "P2"])

##   
## No Yes   
## 287 7

#table(glmdata$Re\_ICU [glmdata$period == "P1"]) / length(glmdata$Re\_ICU [glmdata$period == "P1"])  
paste(round(100\*(table(glmdata$Re\_ICU [glmdata$period == "P1"]) / length(na.omit(glmdata$Re\_ICU [glmdata$period == "P1"]))), 2), "%", sep="")

## [1] "98.06%" "1.94%"

#table(glmdata$Re\_ICU [glmdata$period == "P2"]) / length(glmdata$Re\_ICU [glmdata$period == "P2"])  
paste(round(100\*(table(glmdata$Re\_ICU [glmdata$period == "P2"]) / length(na.omit(glmdata$Re\_ICU [glmdata$period == "P2"]))), 2), "%", sep="")

## [1] "97.62%" "2.38%"

chisq.test(glmdata$Re\_ICU ,glmdata$period, correct = FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: glmdata$Re\_ICU and glmdata$period  
## X-squared = 0.18389, df = 1, p-value = 0.6681

glmdata$Mor\_disc [desdata$Mor\_disc == "Dead"] <- "Dead"  
glmdata$Mor\_disc [desdata$Mor\_disc != "Dead"] <- "Alive"  
  
  
length(glmdata$Mor\_disc )

## [1] 861

sum(is.na(glmdata$Mor\_disc ))

## [1] 0

sum(is.na(glmdata$Mor\_disc )) / length(glmdata$Mor\_disc )

## [1] 0

table(glmdata$Mor\_disc )

##   
## Alive Dead   
## 860 1

#table(glmdata$Mor\_disc ) / length(glmdata$Mor\_disc )  
paste(round(100\*(table(glmdata$Mor\_disc ) / (length(na.omit(glmdata$Mor\_disc )))), 2), "%", sep="")

## [1] "99.88%" "0.12%"

length(na.omit(glmdata$Mor\_disc [glmdata$period == "P1"]))

## [1] 567

length(na.omit(glmdata$Mor\_disc [glmdata$period == "P2"]))

## [1] 294

sum(is.na(glmdata$Mor\_disc [glmdata$period == "P1"]))

## [1] 0

sum(is.na(glmdata$Mor\_disc [glmdata$period == "P2"]))

## [1] 0

table(glmdata$Mor\_disc [glmdata$period == "P1"])

##   
## Alive Dead   
## 566 1

table(glmdata$Mor\_disc [glmdata$period == "P2"])

##   
## Alive   
## 294

#table(glmdata$Mor\_disc [glmdata$period == "P1"]) / length(glmdata$Mor\_disc [glmdata$period == "P1"])  
paste(round(100\*(table(glmdata$Mor\_disc [glmdata$period == "P1"]) / length(na.omit(glmdata$Mor\_disc [glmdata$period == "P1"]))), 2), "%", sep="")

## [1] "99.82%" "0.18%"

#table(glmdata$Mor\_disc [glmdata$period == "P2"]) / length(glmdata$Mor\_disc [glmdata$period == "P2"])  
paste(round(100\*(table(glmdata$Mor\_disc [glmdata$period == "P2"]) / length(na.omit(glmdata$Mor\_disc [glmdata$period == "P2"]))), 2), "%", sep="")

## [1] "100%"

fisher.test(glmdata$Mor\_disc ,glmdata$period)

##   
## Fisher's Exact Test for Count Data  
##   
## data: glmdata$Mor\_disc and glmdata$period  
## p-value = 1  
## alternative hypothesis: true odds ratio is not equal to 1  
## 95 percent confidence interval:  
## 0.00000 75.12717  
## sample estimates:  
## odds ratio   
## 0

# {r} Reintub, Post\_pneu, Post\_throm , Post\_stroke , Re\_ICU , Hos\_LOS, ICU\_LOS, Mor\_disc, #