Group F

D2 R Source Code

**R source code:**

**##Group F**

##Mohammad Obeidat

##Aysegul Yuksel

##Venkatesh Ramanathan

##Didier Bouba Ndengue

##Load the dataset

MYDATASET = read.csv("DFC\_SOCRATA\_FAC\_DATA.csv")

##Look at the contents of dataset

summary(MYDATASET)

##Response variable = Standardized.mortality.ratio

## there are 733 observations with NA value. Should it be zero or deleted from the analysis

## Change NAs to 0

MYDATASET$Standardized.mortality.ratio[is.na(MYDATASET$Standardized.mortality.ratio)] <- 0

##Important question to be answered is whether transform NAs to Zero or delete them from our

##dataset --- Listing the variables with lots of NAs:

##Variables Selection

##Run the logistic regression on each variable and look at the P-Value of each predictors so we ##can select variables with the lowest P-values and the availability of data (no N/A results)

> lm.fit = lm(Standardized.mortalization.ratio ~ Patient.survival.category.text , data = MYDATASET)

> lm.fit

> attach(MYDATASET)

> summary(lm.fit)

Number.of.adult.patients.included.in.arterial.venous.fistula.and.catheter.summaries)

1. summary(MYDATASET$Patient.survival.category.text)

sd(MYDATASET$Patient.survival.category.text)

x=MYDATASET$Patient.survival.category.text

Patient.survival.category.text=table(x)

barplot(Patient.hospitalization.category.text)

main=”Patient survival category code”, xlab=”Patient survival category code”

2. summary(MYDATASET$Rate.of.hospital.readmission.category.text)

sd(MYDATASET$Rate.of.hospital.readmission.category.text)

x=MYDATASET$Rate.of.hospital.readmission.category.text

Rate.of.hospital.readmission.category.text=table(x)

barplot(Rate.of.hospital.readmission.category.text)

main=”Rate of hospital readmission category code”, xlab=”Rate of hospital readmission category code”

3. summary(MYDATASET$Rate.of.hospital.readmission.data.availability.code)

sd(MYDATASET$Rate.of.hospital.readmission.data.availability.code)

x=MYDATASET$Rate.of.hospital.readmission.data.availability.code

Rate.of.hospital.readmission.data.availability.code=table(x)

hist(Rate.of.hospital.readmission.data.availability.code)

main=”Rate of hospital readmission data availability code”, xlab=”Rate of hospital

readmission data availability code”

4. summary(MYDATASET$Patient.hospitalization.category.text)

sd(MYDATASET$Patient.hospitalization.category.text)

x=MYDATASET$Patient.hospitalization.category.text

Patient.hospitalization.category.text=table(x)

barplot(Patient.hospitalization.category.text), main=”Patient hospitalization category

code”, xlab=”Patient hospitalization category code

5. summary(MYDATASET$Lists.the.facility.s.standardized.transfusion.ratio..facility.)

sd(MYDATASET$Lists.the.facility.s.standardized.transfusion.ratio..facility.)

x=MYDATASET$Lists.the.facility.s.standardized.transfusion.ratio..facility.

Lists.the.facility.s.standardized.transfusion.ratio..facility.=table(x)

Hist(Lists.the.facility.s.standardized.transfusion.ratio..facility, main=”Lists the facility s standardized transfusion ratio facility Frequency”,xlab=”Lists the facility s standardized transfusion ratio facility”

6. summary(MYDATASET$Offers.in.peritoneal.dialysis)

sd(MYDATASET$Offers.in.peritoneal.dialysis)

x=MYDATASET$Offers.in.peritoneal.dialysis

Offers.in.peritoneal.dialysis=table(x)

hist(Offers.in.peritoneal.dialysis, main=”Offers in peritoneal dialysis

Frequency”,xlab=” Offers In Peritoneal Dialysis ”

7. summary(MYDATASET$Offers.home.hemodialysis.training)

sd(MYDATASET$Offers.home.hemodialysis.training)

x=MYDATASET$Offers.home.hemodialysis.training)

Offers.home.hemodialysis.training=table(x)

hist(Offers.home.hemodialysis.training, main=”Offers home hemodialysis training

Frequency”,xlab=”Offers home hemodialysis training”

8. summary(MYDATASET$Standardized.readmission.ratio)

sd(MYDATASET$Standardized.readmission.ratio)

x=MYDATASET$Standardized.readmission.ratio

Standardized.readmission.ratio=table(x)

hist(Standardized.readmission.ratio, main=”Standardized readmission ratio

Frequency”, xlab=”Standardized readmission ratio”

9. summary(MYDATASET$Standardized.hospitalization.ratio)

sd(MYDATASET$Standardized.hospitalization.ratio)

x=MYDATASET$Standardized.hospitalization.ratio

Standardized.hospitalization.ratio=table(x)

hist(Standardized.hospitalization.ratio, main =”Standardized Hospitalization Ratio

Frequency”, xlab= “Standardized Hospitalization Ratio”

10. summary(MYDATASET$Hypercalcemia.data.availability.code)

sd(MYDATASET$Hypercalcemia.data.availability.code)

x=MYDATASET$Hypercalcemia.data.availability.code

Hypercalcemia.data.availability.code=table(x)

hist(Hypercalcemia.data.availability.code, main="Hypercalcemia Data Availability Code Frequency", xlab = "Hypercalcemia Data Availability Code"