Exercise 3

Nicole Navarro

September 1, 2016

```
url<-"https://archive.ics.uci.edu/ml/machine-learning-databases/00296/dataset_diabetes.zip"

diabetesData<-read.csv("~/School/fall 16/EDA/data/diabetic_data.csv")
diabetesCodebook<-read.csv("~/School/fall 16/EDA/data/IDs_mapping.csv")

diabetesData[diabetesData == "?"]<- NA
completeCases<-complete.cases(diabetesData)
sum(completeCases)
```

```
## [1] 1043
```

```
ad_missing1<-(length(which(diabetesData$admission_type_id == 5)) + length(which(diabetesData$admission_t

ad_missing2<-sum(is.na(diabetesData$admission_type_id))

dis_missing1<-(length(which(diabetesData$discharge_disposition_id == 18)) + length(which(diabetesData$di

dis_missing2<-sum(is.na(diabetesData$discharge_disposition_id))

p_ER_adm<-100*length(which(diabetesData$admission_type_id == 1))/length(diabetesData$admission_type_id)

p_ER_Exp<-100*length(which((diabetesData$admission_type_id == 1) & (diabetesData$discharge_disposition_i

exp_prob<-p_ER_Exp/p_ER_adm
```

2. Document any missing values in the data.

For this exercise we are concerned with the data on Admission and Discharge type

Admission Type:

- There are 0 rows with no entry
- There are 10396 rows with non-response entries (Not Available, NULL, Not Mapped)

Discharge Type:

- There are 0 rows with no entry
- There are 4680 rows with non-response entries(NULL, Not Mapped, Unknown/Invalid)

3. What percentage of patients are admitted from the emergency room? Given a patient is admitted from the emergency room, what is the probability that their discharge status will be "expired"?

- 53.0530825619559% of patients are admitted from the emergency room.
- The probability of a patient's discharge status being "expired" if they were admitted from the emergency room is 0.0204111872568994

4. What is the most frequent admission status? What is the most frequent discharge status? For the most frequent admission status, what is the most frequent discharge status?

```
Mode <- function(x, na.rm = FALSE) {
  if(na.rm){
    x = x[!is.na(x)]
  }

  ux <- unique(x)
  return(ux[which.max(tabulate(match(x, ux)))])
}

freq<-apply(diabetesData[,c('admission_type_id','discharge_disposition_id')],2,Mode)

ad_types<-diabetesCodebook[2:9,2]
dis_types<-diabetesCodebook[12:41,2]

emergency_ad_data<-diabetesData[which(diabetesData$admission_type_id == 1),]
emergency_dis_mode<-apply(emergency_ad_data[,c('admission_type_id','discharge_disposition_id')],2,Mode)
```

- The most frequent admission status is Urgent
- The most frequent discharge status is Discharged/transferred to another short term hospital
- The most frequent discharge status, with Urgent as the admission status is 1