---

title: "Exercise4"

author: "Chris Leonard"

date: "September 11, 2016"

output: pdf\_document

---

```{r setup, include=TRUE}

(echo = TRUE)

```

```{r, eval=FALSE, include=TRUE}

ddata <-read.csv("C:/Users/Chris/Desktop/Diabetiesresearch/diabetic\_data.csv", header=TRUE)

##Data downloaded 8/30/16

##ddata is the list of events in the hopsitals

discharge <-ddata[8]

dischargetable <-(table(discharge))

paste("Here is an example of calculating the probability of a givin discharge action.")

paste("We will be looking at patients transfered to a skilled nurse support facility (number 3)")

nursefac <-as.numeric(dischage[3])

total <-as.numeric(nrow(discharge))

percenttonurse <- round((nursefac/total)\*100, digits = 2)

paste("The percentage that a patient is going to be sent to a nursing facility is:",percenttonurse)

paste("The percentage of outputting location can be easiliy manipulated by changing the disarge rating in the equation above. The question then becomes, what factors could influence this calculation? With a total number of over 100,000 iterations of sampling, there is enough information to overcome a small sample size error. We are seeing a decent representation of the population in the sample size.")

null1 <-(as.numeric(dischargetable[18]))

null2 <-(as.numeric(dischargetable[25]))

null3 <-(as.numeric(dischargetable[26]))

nullall <-sum(null1,null2,null3)

nullall <-((nullall/total)\*100)

paste("One of the first steps in quantifying the value of data is identifying portions of data that were incomplete, unusable, or not relavent to the analitical process. Looking at some of these values in discharge, we see NULL or not filled, Not mapped, or unknown or invalid. What do these values mean? What cases brought about these outcomes? More explanation is required from the source of data.")

paste("About", nullall, "percent of the values listed in discharge indicate returns that we have no idea about."

paste("Accessability is fairly easly as this is a CSV file. In a standard form, this document can be manipulated in many programing languages")

paste("From this point I am going to look at areas from Table1 in the reading that could be areas of contension for this data.")

paste("Appropriate Amount of Data brings into question how many hospitals were served by this information and how said information can be applied. For a study of 100,000 cases of diabeties, an area of a state or a subregion of the US would be about the maximum applicable size of population to our N value. Even with 100,000 data points, this data may not be indicitive of an entire continent of Diabetic patients")

paste("Understandability of medical data can be a challange for those who do not have a solid background in medical or biological testing. Explanations on tests like metformin, what good concentrations should be and the mechanism of operation of the medication should have been included in the explanations.")

```