# Math 1180:Lab6

Kyle Gaffney

February 12th, 2014

## 1 Data Frames

A Data Frame is an object in R used for storing Data Tables. To see what a data.frame looks like in R lets enter the following commands:

```
A=c(1,2,3)
B=c(4,5,6)
C=c(7,8,9)
DF=data.frame(A,B,C,row.names=c('entry1','entry2','entry3'))
DF
```

Notice that the object DF is a matrix with headers A,B,C across the top and data rows entry1, entry2, entry3 related to how many entries are in each of the vectors. Furthermore the data lists are displayed as columns side by side to form a matrix. How can we get at this data?

```
DF[1,2]
DF[3,3]
DF[3,1]
```

When we wanted the 5th element of a list A, we could get it with the command A[5]. Similarly with a data frame if we want the entry in the 1st row and the 2nd column we use the command DF[1,2]. This will correspond to the 1st entry of B since B is the list that forms the second column of our data frame.

One of the benefits of a data.frame is that R recognizes the names of the rows and the columns. So a different way to call the 1st entry of the Data list B in the data.frame is with the following command.

```
DF['entry1','B']
```

## 1.1 Getting a row from a data.frame

Sometimes we don't want just a single entry but rather a full row from our data frame. There are a couple of commands that we can do which will do this. Suppose for example that we want the 3rd row (entry3). Note the commas are important, they are what signify that we want a row and not a column!!

```
DF[3,]
DF['entry3',]
```

Suppose we want the sum of the 2nd row of the data.frame. How can we do this?

```
sum(DF[2,])
```

## 1.2 Getting a column from a data.frame

Now suppose that we want to collect the data from column instead. The following commands will all extract the 2 column from the data.frame:

```
DF[2]
DF['B']
DF[[2]]
DF[['B']]
DF$B
DF[,'B']
DF[,2]
```

Do you see any differences between the outputs?

Similarly if we want to sum over the column we can do this with the following command:

```
sum(DF[[2]])
sum(DF[2])
```

# 2 Joint Distributions

A joint distribution relates 2 events to one another. For example let us consider the following situation for dart players which perform differently under pressure. Each player independent of the other will hit the board (score=1) with probability 0.8. However player 1 gets nervous under pressure and this causes him to only hit the board 40 percent of the time if his opponent has already hit the board. Player 2 on the other hand excels under pressure and increases his hit percentage to 90 percent when his opponent hits the board first. Let us construct data.frames describing the joint probabilities.

Case 1: Player 1 goes first.

$$Pr(S_1 = 1 \quad and \quad S_2 = 1) = Pr(S_1 = 1) * Pr(S_2 = 1|S_1 = 1) = (0.8) * (0.9) = 0.72$$
$$Pr(S_1 = 0 \quad and \quad S_2 = 1) = Pr(S_1 = 0) * Pr(S_2 = 1|S_1 = 0) = (0.2) * (0.8) = 0.16$$
$$Pr(S_1 = 1 \quad and \quad S_2 = 0) = Pr(S_1 = 1) * Pr(S_2 = 0|S_1 = 1) = (0.8) * (0.1) = 0.08$$
$$Pr(S_1 = 0 \quad and \quad S_2 = 0) = Pr(S_1 = 0) * Pr(S_2 = 0|S_1 = 0) = (0.2) * (0.2) = 0.04$$

Which we can construct in R as a data.frame with the following commands:

```
P1S1=c(0.72,0.08)
P1S0=c(0.16,0.04)
P1goesfirst=data.frame(P1S1,P1S0,row.names=c('P2S1','P2S0'))
```

## 3  Marginal Probability Distributions

The marginal probability distribution is the probability of a single result. For example what is the probability that Player 1 scores?

$$Pr(S_1 = 1) = Pr(S_1 = 1 \quad and \quad S_2 = 1) + Pr(S_1 = 1 \quad and \quad S_2 = 0) = 0.72 + 0.08 = 0.8$$

Notice that this makes sense. Since Player 1 went first his probability of hitting the target should be unchanged and should be 0.8 his normal hit percentage.

We can also get this information from the data.frame. Notice that the probability that player 1 scores is the sum of the column P1S1.

```
sum(P1goesfirst['P1S1'])
```

Find the other marginal probabilities from the data.frame.

## 4  Conditional Distributions

Let us consider a new problem dealing with the number of mites and lice on 2 different birds. We have collected the data in the following data frame:

```
BirdAL0=c(.2,.12,.08)
BirdAL1=c(.15,.09,.06)
BirdAL2=c(.15,.09,.06)
BirdBL0=c(.33,.05,.02)
BirdBL1=c(.08,.13,.09)
BirdBL2=c(.09,.12,.09)
A=data.frame(BirdAL0,BirdAL1,BirdAL2,row.names=c('Mites0','Mites1','Mites2'))
B=data.frame(BirdBL0,BirdBL1,BirdBL2,row.names=c('Mites0','Mites1','Mites2'))
```

First let us find the marginal probabilities for Bird A and enter them into the definitions to make a new data.frame:

```
BirdAL0[4]=sum(A['BirdAL0'])
BirdAL1[4]=sum(A['BirdAL1'])
BirdAL2[4]=sum(A['BirdAL2'])
MPA=c(sum(A['Mites0',]),sum(A['Mites1',]),sum(S['Mites2']),0)
AwM=data.frame(BirdAL0,BirdAL1,BirdAL2,MPA,row.names=c('Mites0','Mites1','Mites2','MP'))
```

Notice for Bird A the product of the marginal probabilities gives the joint distributions. So the 2 parasites are distributed independently for bird A.

What about for bird B?

Calculate the conditional probabilities for Bird A to have 2 lice supposing that it has 1 mite.

# 5 Assignment for the Week

Suppose you are given the following data.frame describing data collected on a third bird, Bird C.

```
BirdCL0=c(.1,.2,.1)
BirdCL1=c(.1,.05,.1)
BirdCL2=c(.15,.1,.1)
C=data.frame(BirdCL0,BirdCL1,BirdCL2,row.names=c('Mites0','Mites1','Mites2'))
```

1. Calculate the Marginal Probabilities and add them to the data.frame like we did for Bird A above.

2. Are the 2 parasites distributed independently for Bird B?

3. Calculate the conditional probabilities for Mites $= 0,1$, and 2 given that there is 1 louse. Plot the result in a bargraph using the command barplot:

   Suppose you named the conditional probabilies CPM0L1, CPM1L1, and CPM2L1 then the following command will draw a bargraph:

   ```
   barplot(c(CPM0L1,CPM1L1,CPM2L1))
   ```

   What does this tell you about the relationship between lice and mites?