Plotting Categorical Variables using R

Ashique.org

`r format(Sys.time(), '%d %B, %Y')`

{r setup, include=FALSE} knitr::opts_chunk$set(echo = T) ## Reading Data file

To analyse categorical variables, let us load the data set containing information related to the survey on Lung capacities of smokers and non smokers in a state. The data is stored in *LungCapData.txt* .

{r loading LungCapData.txt} LungCapData=read.table(file="LungCapData.txt", header = TRUE,sep="\t") #LungCapData=read.table(file="LungCapData.txt", header = TRUE,sep="\t") LungCapData$Gender=as.factor(LungCapData$Gender) attach(LungCapData) str(LungCapData)# structure of data names(LungCapData) # name of variables class(LungCapData) # type of the variable "smoke"

Creating a Barplot

Barplot is used to analyse catagorical variables. It is important to note that bar plot can be generated only for frequency tables. In R, we use *table()* function on a categorical variable to generate frquency table in a single line code.

{r ,bar plot} freq_tab=table(Gender) #freq_tab barplot(freq_tab,las=1,col=3)

Proportionality Plot

A proportionality table can be genreated using the function *prop.table()* as shown bellow.

{r, proportionality table} prop_tab=prop.table(freq_tab) #prop_tab barplot(prop_tab,las=1,names.arg = c("Female","Male"))

Pie Chart in R

We can generate a pie chart in the same way as we do with bar chart. The smokers information will be displayed as a pie chart as follows.

{r} smoke_tab=table(Smoke) pie(smoke_tab,labels=c("No","Yes"),col=c("red","blue"),main = " Smoking habit")

Pie chart-II

In the similar way we can generate the pie chart of the Gender catagory as follows: {r} pie(freq_tab,labels = c("Female","Male"),col = c("green","yellow"),main="Gender distribution in the Data")

Box Plot-I

We can create a summary plot of the lungs capacity through Gender as follows:

{r} boxplot(LungCapData$LungCap~LungCapData$Gender,las=1,main="Lungs Capacity with Gender",notch=T)

Box plot- II

```
{r} boxplot(LungCapData$LungCap~Smoke,las=1,main="Lungs Capacity with Smoking Habit",notch=T)
```

Box plot-III

Let us plot the Lung Capacity through gender with smoke habit as follows:

```
{r} boxplot(LungCapData$Age~LungCapData$Smoke*LungCapData$Gender,las=1,main="Lungs Capacity
with  Gender and Smoking Habit",notch=T) ## Stripe Plot
```

Identify the distribution of data points, we can use stripe plot

```
{r} stripchart(LungCapData$Age~LungCapData$Smoke*LungCapData$Gender)
```

Testing significance of difference in smoking habit

A simple χ^2 test will help us to substantiate statistically that 'is variance in smoking habit significant?'.

```
{r} smoke_tab chisq.test(smoke_tab)
```

Cross Tabulation and dependancy of a catagorical variable on another one

```
{r} library(gmodels) CrossTable(Smoke,Caesarean,prop.t=FALSE, prop.r=TRUE, prop.c=TRUE,chisq
= TRUE,format = "SPSS")
```