Bedford Analysis

Vagmita Pabuwal

3/6/2021

```
knitr::opts_chunk$set(echo = TRUE, message =FALSE)
```

R Markdown

This is an R Markdown document to illustrate Bedofrd's law being applied to Census_2007 data.

```
library(ggplot2)
library(dplyr)
library(rmarkdown)
library(knitr)
library(png)




setwd("C:\\Users/vagmi/Downloads/")
Bedford_data <- read.delim("census_2009b", header=TRUE)
#is.na.data.frame(Bedford_data)
Bedford_data1<- na.omit(Bedford_data)

View(Bedford_data1)
```

##Extract Leading digit from 2007_9

```
Bedford_data1$X7_extract <- as.numeric(substr(Bedford_data1$X7_2009, 1,1))
```

##Compute Frequency, Relative Frequency and Bedford Law Log10

```
options(digits=5)
df_2007_frequency <- Bedford_data1 %>% group_by(X7_extract) %>% summarise(counts = n())
n1 <- sum(df_2007_frequency$counts)
Rel_Freq <- (df_2007_frequency$counts/n1)*100
BF_Law <- (log10(1+1/df_2007_frequency$X7_extract))*100
```

#Put all of them in a dataframe

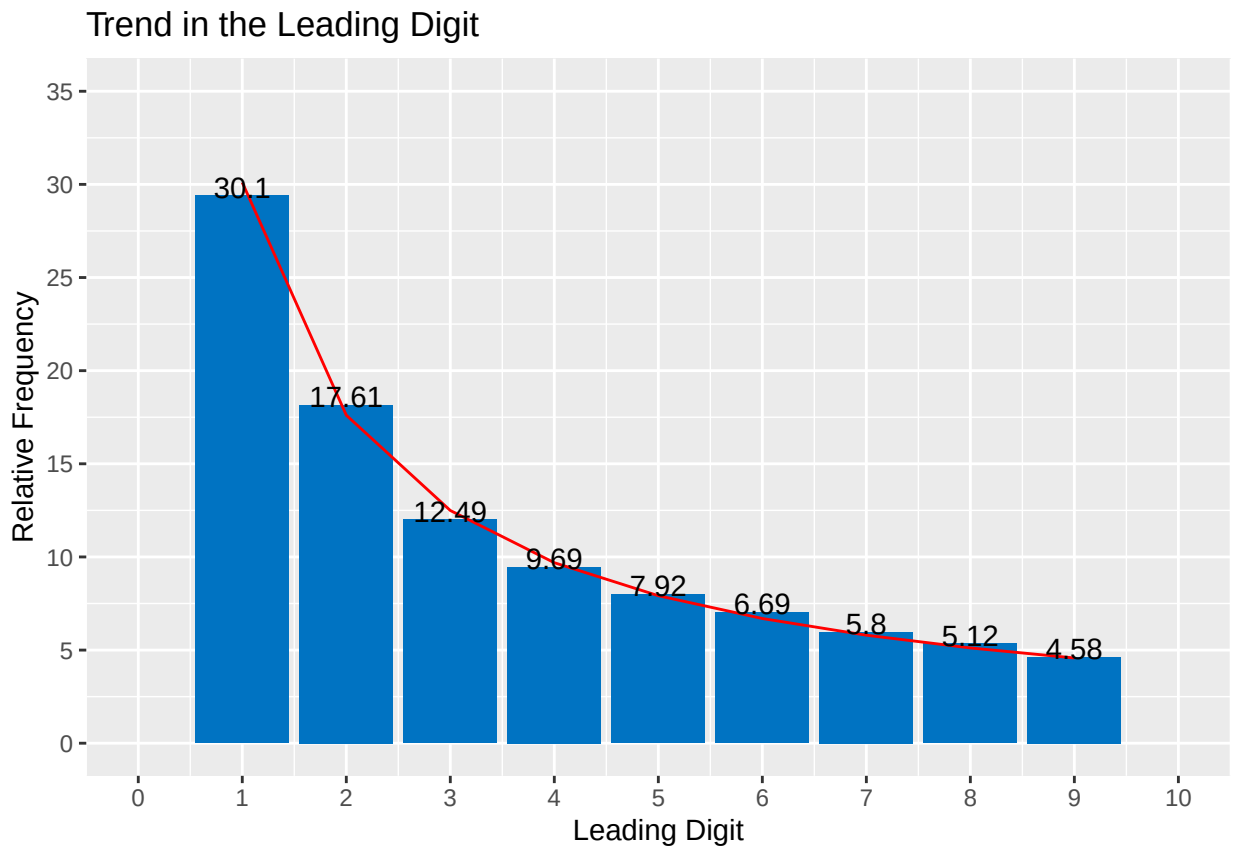
```
df_2007_frequency_f <- cbind(df_2007_frequency, Rel_Freq, BF_Law)
View(df_2007_frequency_f)

knitr::kable(x=df_2007_frequency_f, digits=2)
```

X7_extract	counts	Rel_Freq	BF_Law
1	5735	29.41	30.10
2	3540	18.15	17.61
3	2341	12.00	12.49
4	1847	9.47	9.69
5	1559	7.99	7.92
6	1370	7.02	6.69
7	1166	5.98	5.80
8	1042	5.34	5.12
9	903	4.63	4.58

##PLOT FREQ DISTRIBUTION and overlay with Bedford assumption

```
ggplot(data=df_2007_frequency_f, aes(x= X7_extract, y=Rel_Freq)) +
 geom_bar(fill = "#0073C2FF", stat="identity")+
  scale_x_continuous("Leading Digit", limits = c(0,10), breaks = seq(0, 10, 1))+
  scale_y_continuous("Relative Frequency", limits = c(0,35), breaks = seq(0, 35, 5))+
  ggtitle("Trend in the Leading Digit")+
  geom_line(data=df_2007_frequency_f, aes(y=BF_Law), color="red")+
  geom_text(aes(label=round(BF_Law,2)), vjust = .1) +
  theme()
```



#Clearly the observed data matched the expected distribution as seen from the red overlay of from Bedford prediction in red and frequency distribution plot.