

A1\_231057008

2023-08-22

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
df_bats = read.csv('Data/bats.csv', row.names = 1) #importing the csv file

head(df_bats) #as data is large only 6 rows are printed

##   Gene.1 Gene.2 Gene.3 Gene.4 Gene.5 Ebola
## 1 FALSE  FALSE   TRUE   TRUE   TRUE  TRUE
## 2  TRUE  FALSE   TRUE  FALSE  FALSE FALSE
## 3  TRUE  FALSE   TRUE   TRUE   TRUE  FALSE
## 4 FALSE   TRUE   TRUE   TRUE   TRUE  TRUE
## 5 FALSE  FALSE  FALSE   TRUE  FALSE FALSE
## 6  TRUE  FALSE  FALSE   TRUE  FALSE FALSE

str(df_bats) #structure of the df_bats

## 'data.frame':   99999 obs. of  6 variables:
##  $ Gene.1: logi  FALSE TRUE TRUE FALSE FALSE TRUE ...
##  $ Gene.2: logi  FALSE FALSE FALSE TRUE FALSE FALSE ...
##  $ Gene.3: logi   TRUE TRUE TRUE TRUE FALSE FALSE ...
##  $ Gene.4: logi   TRUE FALSE TRUE TRUE TRUE TRUE ...
##  $ Gene.5: logi   TRUE FALSE TRUE TRUE FALSE FALSE ...
##  $ Ebola : logi   TRUE FALSE FALSE TRUE FALSE FALSE ...

ebola_chances = mean(df_bats$Ebola)
cat("Chance of a random bat carrying the Ebola virus =", ebola_chances, "\n")

## Chance of a random bat carrying the Ebola virus = 0.300793

gene_likelihoods <- colMeans(df_bats[, c("Gene.1", "Gene.2", "Gene.3", "Gene.4", "Gene.5")])
cat("Likelihood of each gene being expressed:\n")

## Likelihood of each gene being expressed:

cat("Gene 1:", gene_likelihoods[1], "\n")

## Gene 1: 0.702277

cat("Gene 2:", gene_likelihoods[2], "\n")

## Gene 2: 0.300763

cat("Gene 3:", gene_likelihoods[3], "\n")

## Gene 3: 0.500895
```

```

cat("Gene 4:", gene_likelihoods[4], "\n")

## Gene 4: 0.801618
cat("Gene 5:", gene_likelihoods[5], "\n")

## Gene 5: 0.3270533
genes_to_analyze <- c("Gene.1", "Gene.2", "Gene.3", "Gene.4", "Gene.5")

for (gene in genes_to_analyze) {
  # Calculate proportions
  presence_ebola_proportion <- mean(df_bats[df_bats[[gene]] == TRUE, "Ebola"])
  absence_ebola_proportion <- mean(df_bats[df_bats[[gene]] == FALSE, "Ebola"])

  # Print results
  cat("Gene:", gene, "\n")
  cat("Proportion of bats carrying Ebola with", gene, "present:", presence_ebola_proportion, "\n")
  cat("Proportion of bats carrying Ebola with", gene, "absent:", absence_ebola_proportion, "\n\n")
}

## Gene: Gene.1
## Proportion of bats carrying Ebola with Gene.1 present: 0.3020206
## Proportion of bats carrying Ebola with Gene.1 absent: 0.2978974
##
## Gene: Gene.2
## Proportion of bats carrying Ebola with Gene.2 present: 0.3022011
## Proportion of bats carrying Ebola with Gene.2 absent: 0.3001873
##
## Gene: Gene.3
## Proportion of bats carrying Ebola with Gene.3 present: 0.5832019
## Proportion of bats carrying Ebola with Gene.3 absent: 0.01737127
##
## Gene: Gene.4
## Proportion of bats carrying Ebola with Gene.4 present: 0.3705418
## Proportion of bats carrying Ebola with Gene.4 absent: 0.01895352
##
## Gene: Gene.5
## Proportion of bats carrying Ebola with Gene.5 present: 0.8999847
## Proportion of bats carrying Ebola with Gene.5 absent: 0.009584807

```