9. Correlation Test and Plots

Shawn Behrend

2022-05-07

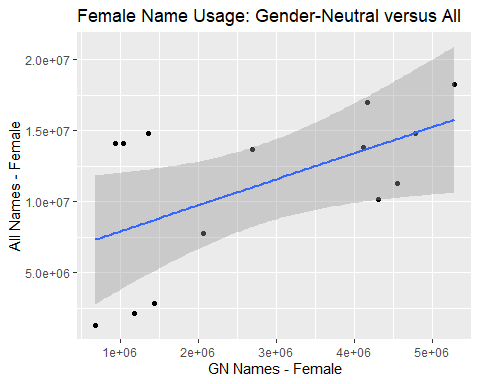
Using the All Decades Summary Table created in a previous step, correlation tests were run for the occurrences of gender neutral names versus the total occurrences of any name, for both males and females.

# Load ggplot2  
library("ggplot2")

## Registered S3 methods overwritten by 'tibble':  
## method from   
## format.tbl pillar  
## print.tbl pillar

sum\_all\_decades\_df<-read.csv("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/all\_decades\_summary\_table.csv")  
  
#correlation for female names  
f\_test\_set<-data.frame(sum\_all\_decades\_df$Years,sum\_all\_decades\_df$F.occurrances.of.GN.names,sum\_all\_decades\_df$Total.F.occurrances)  
f\_gn\_names<-as.numeric(unlist(f\_test\_set[2]))  
f\_all\_names<-as.numeric(unlist(f\_test\_set[3]))  
f\_gn\_all<-data.frame(f\_gn\_names,f\_all\_names)  
# correlation tests  
f\_gn\_scatter\_plot <- ggplot(f\_gn\_all, aes(f\_gn\_names, f\_all\_names))  
f\_gn\_scatter\_plot + geom\_point() + labs(title="Female Name Usage: Gender-Neutral versus All",x = "GN Names - Female", y = "All Names - Female") + geom\_smooth(method="lm")

## `geom\_smooth()` using formula 'y ~ x'

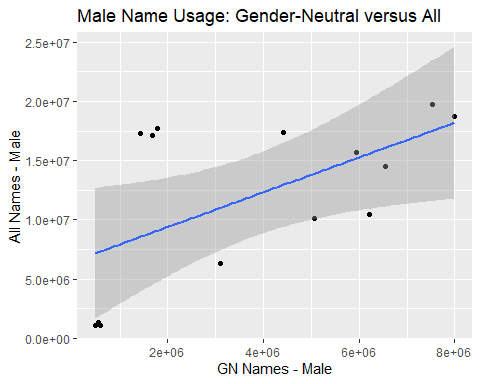


cortest\_f<-cor.test(f\_gn\_names, f\_all\_names, method = "pearson")  
  
sink("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/f\_gn\_names\_corr.txt")  
print(cortest\_f)

##   
## Pearson's product-moment correlation  
##   
## data: f\_gn\_names and f\_all\_names  
## t = 2.3283, df = 12, p-value = 0.03819  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.03870847 0.83984009  
## sample estimates:  
## cor   
## 0.5578313

sink() # returns output to the console  
#----  
#correlation for male names  
m\_test\_set<-data.frame(sum\_all\_decades\_df$Years,sum\_all\_decades\_df$M.occurrances.of.GN.names,sum\_all\_decades\_df$Total.M.occurrances)  
m\_gn\_names<-as.numeric(unlist(m\_test\_set[2]))  
m\_all\_names<-as.numeric(unlist(m\_test\_set[3]))  
m\_gn\_all<-data.frame(m\_gn\_names,m\_all\_names)  
# correlation tests  
m\_gn\_scatter\_plot <- ggplot(m\_gn\_all, aes(m\_gn\_names, m\_all\_names))  
m\_gn\_scatter\_plot + geom\_point() + labs(title="Male Name Usage: Gender-Neutral versus All",x = "GN Names - Male", y = "All Names - Male") + geom\_smooth(method="lm")

## `geom\_smooth()` using formula 'y ~ x'



cortest\_m<-cor.test(m\_gn\_names, m\_all\_names, method = "pearson")  
#---  
sink("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/f\_gn\_names\_corr.txt")  
print(cortest\_f)

##   
## Pearson's product-moment correlation  
##   
## data: f\_gn\_names and f\_all\_names  
## t = 2.3283, df = 12, p-value = 0.03819  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.03870847 0.83984009  
## sample estimates:  
## cor   
## 0.5578313

sink() # returns output to the console  
  
sink("C:/Users/shawn/OneDrive/Shawn/CSU\_global/MIS581/project\_r\_code/name\_files/m\_gn\_names\_corr.txt")  
print(cortest\_m)

##   
## Pearson's product-moment correlation  
##   
## data: m\_gn\_names and m\_all\_names  
## t = 2.4599, df = 12, p-value = 0.03004  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.06986763 0.84881121  
## sample estimates:  
## cor   
## 0.578984

sink() # returns output to the console