PROJECT1 PDF

R Markdown

library(knitr) library(markdown) knit("PROJECT1 PDF.Rmd") 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:

library(ISwR) library(TTR) library(readr)

A2010 <- read.csv("BP Apprehensions 2010.csv", header = TRUE, stringsAsFactors = FALSE) A2017 <- read.csv("PB Apprehensions 2017.csv", header = TRUE, stringsAsFactors = FALSE) #par(mfrow=c(2,2))

Compare by sector

yearA <- as.data.frame(matrix(c(rowSums(A2010[,-1])),nrow = 1))

colnames(yearA) <- A2010[,1]

yearB<- as.data.frame(matrix(c(rowSums(A2017[,-1])),nrow = 1))

colnames(yearB) <- A2010[,1]

yearAB <- rbind(yearA, yearB)

row.names(yearAB) <- c("year2010", "year2017")

barplot(as.matrix(yearAB), beside = TRUE, col = c("red", "blue"), bty="n")

legend("topleft", c("year2010","year2017"), pch=15, col=c("red","blue"), bty="n")

Compare by month

yearA1 <- as.data.frame(matrix(c(colSums(A2010[,-1])),nrow = 1))

colnames(yearA1) <- c(10:12,1:9)

yearB1<- as.data.frame(matrix(c(colSums(A2017[,-1])),nrow = 1))

colnames(yearB1) <- c(10:12,1:9)

yearAB1 <- rbind(yearA1, yearB1)

row.names(yearAB1) <- c("year2010", "year2017")

barplot(as.matrix(yearAB1), beside = TRUE, col = c("red", "blue"), bty="n")

legend("topleft", c("year2010","year2017"), pch=15, col=c("red","blue"), bty="n")

t.test

a<-as.numeric(which.max(yearA)) b<-as.numeric(which.max(yearB)) t.test(as.numeric(A2010[a,-1]), as.numeric(A2017[b,-1]), paired = FALSE) #Assume 95% confidence level, the p-value 0.06346 is bigger than 0.05, #So, we do not reject H0 which states there is no change in the means # of the most apprehension sectors of 2010 and 2017. In conclusion, we # are 95% confident to say that there is a change between two means.

Find the three months with most apprehension of 2010 and 2017

sum2010 <- runSum(as.numeric(yearA1),n = 3, cumulative = FALSE) sum2017 <- runSum(as.numeric(yearB1),n = 3, cumulative = FALSE) max(sum2010,na.rm = T) max(sum2017,na.rm = T) which.max(sum2010) which.max(sum2017) #The three months period with the most apprehension of 2010 is March, #April, and May, with the sum 163643. #The three months period with the most apprehension of 2017 is October, #November, and December, with the sum 136646. #The sum of 2010 is bigger than the sum of 2017 #Use t test to compare the sums of March, April, and May in 2010 and #the sums of October, November, and December in 2017

t test

yearA2<-yearA1[(which.max(sum2010)-2):which.max(sum2010)] yearB2<-yearB1[(which.max(sum2017)-2):which.max(sum2017)] t.test(as.numeric(yearA2),as.numeric(yearB2), paired = FALSE) #Assuming 95% confidence level, the p-value is 0.1541 is bigger than 0.05, #So we do not reject H0 which states there is no change in the sums of three monthe #with the most apprehension of 2010 and 2017. In conclusion, we are 95% confident #to say that there is a change between two group of sums.

Time series

ts1 <- read_csv("NEW PB monthly summaries.csv", col_names = FALSE)

ts2 <- as.vector(t(ts1)) ts2 <- rev(ts2)

ts3 <- ts(ts2, start = c(2000,10), frequency=12)

ts.plot(ts3, gpars=list(xlab="year", ylab="Apprehensions", lty=1)) x <- 2000.5 y <- rev(rowMeans(ts1)) for(i in 1:18){ segments(x,y[i],x+1,y[i],col = 'red') text(x+1.5,y[i],as.character(x-.5),col = 'red',cex = 0.5) x <- x+1 }