R functions for exam 2

Chi-square
chisq.test(dataframe[,i:j], correct =FALSE)

McNemar test (two-tailed test)
x<-matrix(c(A,B,C,D),2,2)
mcnemar.test(x, correct = FALSE)

Wilcoxon Signed-Rank Test
wilcox.test(Growth_Value$‘Growth‘, mu=5, alternative="greater")

Wilcoxon Signed-Rank Test for Matched-Pairs
wilcox.test(Growth_Value$'Growth‘, Growth_Value$'Value‘, alternative="two.sided“, paired=TRUE)

Wilcoxon Rank-Sum Test for Independent Samples
wilcox.test(Undergrad_Salaries$'Computer Science‘, Undergrad_Salaries$'Finance‘,
alternative="two.sided“, paired=FALSE)

Kruskal-Wallis Rank Test
Stacked2<- melt(KWexample)
colnames(Stacked2)<- c("Major", "Size")
Stacked2
kruskal.test(Stacked2$'Size',Stacked2$'Major')

Calculate and interpret the correlation coefficient between debt payments and income.
cor(Debt_Payments$'Income', Debt_Payments$'Debt’)
cor(Debt_Payments[2:4], use = "all.obs")

Create plot
install.packages("tidyverse")
library(tidyverse)
ggplot (data = Debt_Payments) + geom_point (mapping = aes (x = Income, y = Debt))

Simple Linear Regression
Simple <– lm(Debt~Income, data=Debt_Payments)
summary(Simple)
anova(Simple)

Multiple Linear Regression
Multiple2 <- lm(pie$'pie sales'~pie$'price ($)'+pie$'advertising ($100s)')
summary(Multiple2)
MR1 <- lm(pie$'pie sales'~ 1)
anova(MR1, Multiple2)
confint(Multiple2, 'pie$"price ($)"', level=0.95)

Model with dummy variable
GNV<- ifelse(GNV_JAX_Jan2022$city == "Gainesville, Florida", 1, 0)
mlr2 <- lm(GNV_JAX_Jan2022$PRICE ~ GNV_JAX_Jan2022$`SQUARE FOOTAGE` + GNV)
summary(mlr2)

Model with interaction variable

```
GNV_Int<- ifelse(GNV_JAX_Jan2022$city == "Gainesville, Florida", GNV_JAX_Jan2022$`SQUARE FOOTAGE`, 0)
mlr2_b <- lm(GNV_JAX_Jan2022$PRICE ~ GNV_JAX_Jan2022$`SQUARE FOOTAGE` + GNV+GNV_Int)
summary(mlr2_b)
```

Create residual plot

```
plot(mlr2_b)
```

Calculate VIF

```
install.packages("car")
library(car)
lmobject1 <- lm(PRICE ~ BEDS+ SQFT + BEDSANDBATHS+ LOTSIZE, data = Tampa2022)
summary(lmobject1)
vif(lmobject1)
```

Logistic regression

```
logmod = glm(Purchase~Age, family = binomial, data = MacysPurchases)
summary(logmod)
```

Build model using stepwise method

```
none <-lm(price ~1, data = GainesvilleHomes_Sp2019_Quant)
full <- lm(price ~ beds_baths + square_footage + lot_size+commute + year_built + es_dist + ms_dist + hs_dist, data = GainesvilleHomes_Sp2019_Quant)
MSE <- (summary(full)$sigma)^2
step(none, scope=list(upper= full), scale=MSE) (#by default, it uses stepwise method)
```