R script Code:

# Aadhithya Dinesh

# MIS 545 Section 02

# Lab05DineshA.R

# Import a dataset of spending by groups of people visiting a zoo and generate

# a multiple linear regression model by assigning data types, building a model

# and testing for model fit.

# install.packages("tidyverse")

# install.packages("dummies")

# install.packages("corrplot")

# install.packages("olsrr")

library(tidyverse)

library("corrplot")

library("olsrr")

# set the working directory

setwd("~/MIS/Classes/MIS545/Assignments/Lab05")

# read the csv file with column types specified

zooSpending <- read\_csv(file = "ZooVisitSpending.csv",

col\_types = "niil",

col\_names = TRUE)

# print the zooSpending tibble

print(zooSpending)

#print the structure of zooSpending

str(zooSpending)

#print the summary of zooSpending

print(summary(zooSpending))

# define the function to display all histograms

displayAllHistograms <- function(tibbleDataset) {

tibbleDataset %>%

keep(is.numeric) %>%

gather() %>%

ggplot() + geom\_histogram(mapping = aes(x=value, fill=key),

color = "black") +

facet\_wrap (~ key, scales = "free") +

theme\_minimal()

}

# call the function

displayAllHistograms(zooSpending)

# rounding the correlation to 2 decimal places

print(round(cor(zooSpending %>% keep(is.numeric)),2))

# displaying the correlation plot using number method

corrplot(cor(zooSpending),

method = "number",

type = "lower")

# generating the linear regression model by using all the independent variables

zooSpendingModel <- lm(data = zooSpending,

formula = VisitSpending ~ .)

# displaying the beta coefficients for the model

print(zooSpendingModel)

# displaying the linear regression model results using the summary function

print(summary(zooSpendingModel))

# testing for multicollinearity

print(ols\_vif\_tol(zooSpendingModel))

**Correlation Plot:**

**Chart

Description automatically generated**

**Model Sumary:**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.22141 6.49061 0.034 0.97284

PartySize 9.13619 1.01756 8.979 4.35e-15 \*\*\*

MilesFromZoo 0.88886 0.04865 18.272 < 2e-16 \*\*\*

MemberTRUE -14.90735 4.58300 -3.253 0.00148 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 24.46 on 121 degrees of freedom

Multiple R-squared: 0.765, Adjusted R-squared: 0.7592

F-statistic: 131.3 on 3 and 121 DF, p-value: < 2.2e-16

**Multicollinearity results:**

Variables Tolerance VIF

1 PartySize 0.9831086 1.017182

2 MilesFromZoo 0.9926983 1.007355

3 MemberTRUE 0.9890274 1.011094

**Rapid Miner Process:**

Graphical user interface, application

Description automatically generated

**Correlation Matrix:**

Graphical user interface, application, Word

Description automatically generated

**Linear Regression Model Results:**

A screenshot of a computer

Description automatically generated

Answers:

1. MilesFromZoo seems to be the most statistically significant variable with a correlation of 0.77 followed by PartySize with 0.32.
2. The multiple R squared value of 0.765 means that 76.5% of the zoo spending can be explained by the variance in all the other variables.
3. The zoo will be spending 9 units more (as the beta coefficient suggests) for every additional guest in a party.
4. The zoo will be spending about 14 units lesser on members compared to non-members as suggested by the coefficients.
5. The zoo will be spending 0.88 units for each additional mile travelled to visit the zoo.
6. As all the tolerance values of dependent variables are >0.2 and the VIF values are < 5, there seems to be no issues of multicollinearity.