Assignment 1  
  
Hello Professor,   
All the questions have been answered and explained with description in the commented-out codes. And the three graphs are attached below for your kind perusal.

# ASSIGNMENT 1 MAT 378 #

# Aditya Parey #

# aparey@oswego.edu#

#current\_timestamp <- Sys.time()

#print(current\_timestamp)

##[1] "2023-09-19 15:14:29 EDT"##

# read the data in the dataset#

house.data = read.csv("HouseData.txt", sep = ",")

names(house.data)

# display the "Pool" column

print(house.data$Pool)

print(house.data$Price)

#QUESTION 1#

pool\_count <- table(house.data$Pool)

# Print the counts

print(pool\_count)

#QUESTION 2#

#nrow is the total number of houses#

#sum function gives out number of houses with or without pool#

# Calculate the proportion of houses with a pool

prop\_with\_pool <- sum(house.data$Pool == 1) / nrow(house.data)

# Calculate the proportion of houses without a pool

prop\_without\_pool <- sum(house.data$Pool == 0) / nrow(house.data)

# Print the proportions

print(prop\_with\_pool)

print(prop\_without\_pool)

#QUESTION 3#

# bar plot #

barplot(pool\_count, ylim=c(0,100),

main = "Houses with and without a Pool",

xlab = "Pool (0 = No, 1 = Yes)",

ylab = "Count")

#QUESTION 4#

# histogram#

library(psych)

hist(house.data$Price,

xlab = "Price",

main = "Distribution of House Prices")

#QUESTION 5#

# a boxplot of the "Price" variable#

boxplot(house.data$Price,

ylab = "Price",

main = " House Prices")

#QUESTION 6#

#description and summary the Price variable#

describe(house.data$Price)

summary(house.data$Price)

#QUESTION 7#

#We will use the test of proportion for this question#

#We have a null hypothesis where p is equal or greater than 0.5 and alternative hypothesis where

# is less than 0.5, where alpha = 0.05#

#next up we will calculate the test statistic by inserting the formula#

#then we can get the p value by using pnorm function to derive conclusions#

sample\_prop = mean(house.data$Pool)

print(sample\_prop)

#Calculate the test stat#

sample\_size <- length(house.data$Pool)

population\_prop <- 0.5

test\_stat <- (sample\_prop - population\_prop) / sqrt((population\_prop \* (1 - population\_prop)) / sample\_size)

print(test\_stat)

## Calculate the p-value for a one-side test of proportion#

p\_value <- pnorm(test\_stat)

print(p\_value)

#QUESTION 8#

# the critical value from the standard normal distribution for a 98% confidence interval

critical\_value <- qnorm(0.99) # 0.01 on each tail for a 98% confidence level

# the margin of error

error <- critical\_value \* sqrt((sample\_prop \* (1 - sample\_prop)) / sample\_size)

# the lower and upper bounds of the confidence interval

lower\_bound <- sample\_prop - error

upper\_bound <- sample\_prop + error

#s the confidence interval

cat("98% Confidence Interval for Proportion of Houses with a Pool: (", round(lower\_bound, 4), ", ", round(upper\_bound, 4), ")\n")

#QUESTION 9#

# Create a Q-Q plot to check normality#

qqnorm(house.data$Price)

qqline(house.data$Price)

# the Shapiro-Wilk test to check for normality#

shapiro.test(house.data$Price)

# p-value being greater than 0.05, we fail to reject the null hypothesis (H0),

#which means that we do not have sufficient evidence

#to conclude that the "Price" variable significantly varies from a normal distribution.

#This suggests that the "Price" variable follows a normal distribution #

#QUESTION 10 #

# Set the hypothesized population mean

population\_mean <- 278000

# Perform a one-sample t-test

Price\_test <- t.test(house.data$Price, mu = population\_mean, alternative = "greater")

Price\_test

# we fail to reject the alternative hypothesis and conclude that the price is greater than $278,000

#the p-value of approximately 0.007889, which is less than the chosen significance level of alpha = 0.05, we would reject the null hypothesis #

#QUESTION 11#

# one-sample t-test to calculate the confidence interval

Price\_CI <- t.test(house.data$Price, conf.level = 0.90)

Price\_CI

# the confidence interval for Price data would range between ( 278406.3 to 280093.8) with a 90% confidence #

Q3) Bar Plot  
  
 A graph of a number of squares

Description automatically generated

Q4) Histogram A graph of a number

Description automatically generated

Q5) BoxPlot

A graph with numbers and lines

Description automatically generated