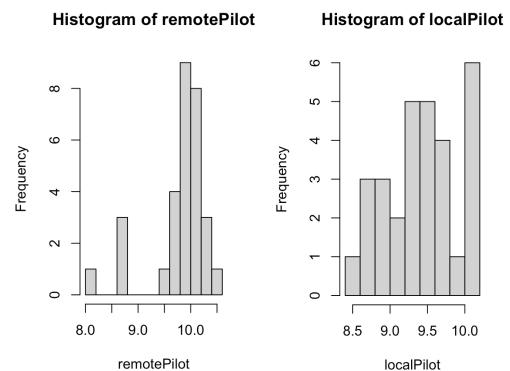Mini Project 4
Kaden Chan

Section 1:

1. a) The remote pilot has a more scattered
   distribution, left skew, with a normal distribution
   towards the right, centered at 10. The local pilot
   has a more uniform distribution if not increasing
   in frequency as voltage increases, so it is likely
   that these distributions at each location is different.

   b) null hypothesis u_0: R - L = 0 | Se = Var(R - L) = sqrt($s_r^2/n_1^2 + s_l^2/n_2^2$) = .132
   Test Statistic, T = (.3813 - 0)/Se = 2.89
   Assuming using alpha = .05
   Calculating cutoff value for a .05 significance level, df = ($n_1 + n_2 - 2$) → 2 (used R's qt)
   Since T > cutoff, the null hypothesis is rejected, so it appears the manufacturing process
   cannot be established locally.
   A confidence interval for the actual difference in mean voltage between locations:
   (.3813 - 0) +- $t_{.05/2, n1 + n2 - 2}$ (.132) = ..3183 += .264 = [.0543, .5823], so 0 is out of the 95%
   confidence interval… there is not a 95% chance the manufacturing process can be
   established locally.

   c) The fact that there is a significant difference between the means of the remote and
   local locations, in part b, strengthens the argument that the distributions of each location
   are different from part a.

2. Null Hypothesis: $u_{Exp} - u_{The}$ = 0, there is no difference between experimental and
   calculated values of vapor pressure for dibenzothiophene.
   Since, experimental and theoretical data points are correlated, we can treat it as a
   pooled variable. Using the difference, D, as a single variable.
   Finding the mean of D: .00069
   SE of D: sqrt (Var(D)) = .00355
   Calculate T: .00069/.00355 = .1934
   Calculate cutoff for .05 significance and df = n - 1 →2.13
   Since T < cutoff, by a lot, it fails to reject the Null Hypothesis, so it is likely the theoretical
   model for vapor pressure is a good model of reality

```r
# Mini Project 4
# Name: Kaden Chan
# Section 1. Answers to the specific questions asked
# Section 2: R code. Your code must be annotated.
# No points may be given if a brief look at the code
# does not tell us what it is doing.

# 1
data <- read.csv("VOLTAGE.csv")

# 1a distributions at each location
remotePilot <- data[data$location == 0, ]$voltage
localPilot <- data[data$location == 1, ]$voltage

hist(remotePilot, breaks = 9)
hist(localPilot, breaks = 9)

# 1b calc mean, se, T cutoff and confidence interval
remoteMean <- mean(remotePilot)
localMean <- mean(localPilot)
cat("difference in sample location means", remoteMean - localMean, "\n")

sampleSDRemote <- 0 #Standard Error (SE) squared
for (r in remotePilot) {
 sampleSDRemote <- sampleSDRemote + (r - remoteMean) * (r - remoteMean)
}
sampleSDRemote <- sampleSDRemote / (length(remotePilot) - 1)

sampleSDLocal <- 0 #Standard Error (SE) squared
for (r in localPilot) {
 sampleSDLocal <- sampleSDLocal + (r - localMean) * (r - localMean)
}
sampleSDLocal <- sampleSDLocal / (length(localPilot) - 1)

# SE of two variables
diffSE <- sqrt(
 sampleSDRemote / length(remotePilot) + sampleSDLocal / length(localPilot)
)

cat("mean diff SE:", diffSE)
# calculate test statistic and see if it's in rejection range
T <- (remoteMean - localMean) / diffSE
```

```r
cat(T)
cutoff <- qt(1 - .05 / 2, length(remotePilot) + length(localPilot) - 2)
if (T > cutoff) {
 cat("T is high, reject null hypothesis")
} else {
 cat("T is low, fail to reject null hypothesis")
}

# 2 analysis of if experimental = theoretical
data <- read.csv("VAPOR.csv")
theData <- data$theoretical
expData <- data$experimental

theMean <- mean(theData)
expMean <- mean(expData)

# diff in sample mean
cat("difference in sample vapor means:", theMean - expMean)

# sampleSDThe <- 0 #Standard Error (SE) squared
# for (r in theData) {
#    sampleSDThe <- sampleSDThe + (r - theMean) * (r - theMean)
# }
# sampleSDThe <- sampleSDThe / (length(theData) - 1)

# sampleSDExp <- 0 #Standard Error (SE) squared
# for (r in expData) {
#    sampleSDExp <- sampleSDExp + (r - expMean) * (r - expMean)
# }
# sampleSDExp <- sampleSDExp / (length(expData) - 1)

diffData <- theData - expData
diffMean <- mean(diffData)
cat(diffMean) # the same as the difference in means

sampleSDDiff <- 0 # Standard Error (SE) squared
for (r in diffData) {
 sampleSDDiff <- sampleSDDiff + (r - diffMean) * (r - diffMean)
}
#calc SE of pooled variables
sampleSDDiff <- sampleSDDiff / (length(diffData) - 1) # s^2
diffSE <- sqrt(sampleSDDiff / length(diffData)) # s/sqrt(n)
```

```r
cat("SE of D: ", diffSE)

# calculate test statistic and see if it's in rejection range
T <- (diffMean) / diffSE
cat(T)
cutoff <- qt(1 - .05 / 2, length(diffData) - 1)
cat(cutoff)
if (T > cutoff) {
 cat("T is high, reject null hypothesis")
} else {
 cat("T is low, fail to reject null hypothesis")
}
```