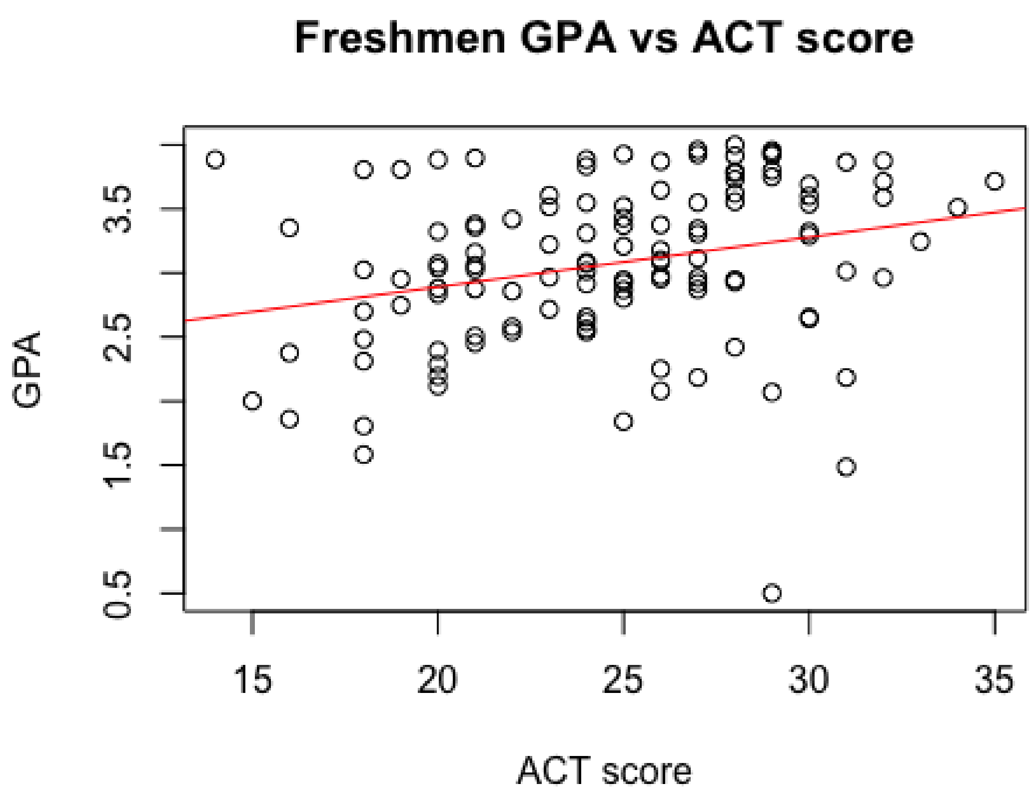
Confidence-Interval

First dataset:

```
file1 <- "gpa.csv"
data1 <- read.csv(file1,header = T)
# scatter plot of GPA vs ACT
act = data1$act
gpa = data1$gpa
plot(act, gpa, main="Freshmen GPA vs ACT score",
     xlab="ACT score", ylab="GPA")
abline(lm(gpa~act), col="red") # regression line
```




```
#lines(lowess(act,gpa), col="blue") # lowess line

library(boot)
p.npar <- function(x, i=c(1:n)) {
  boot.sample=x[i,]
  act = boot.sample$act
  gpa = boot.sample$gpa
  p = cor(act,gpa)
```

```r
  return(p)
}
p.npar.boot <- boot(data1, p.npar, R = 999, sim = "ordinary", stype = "i")
p.npar.boot
```

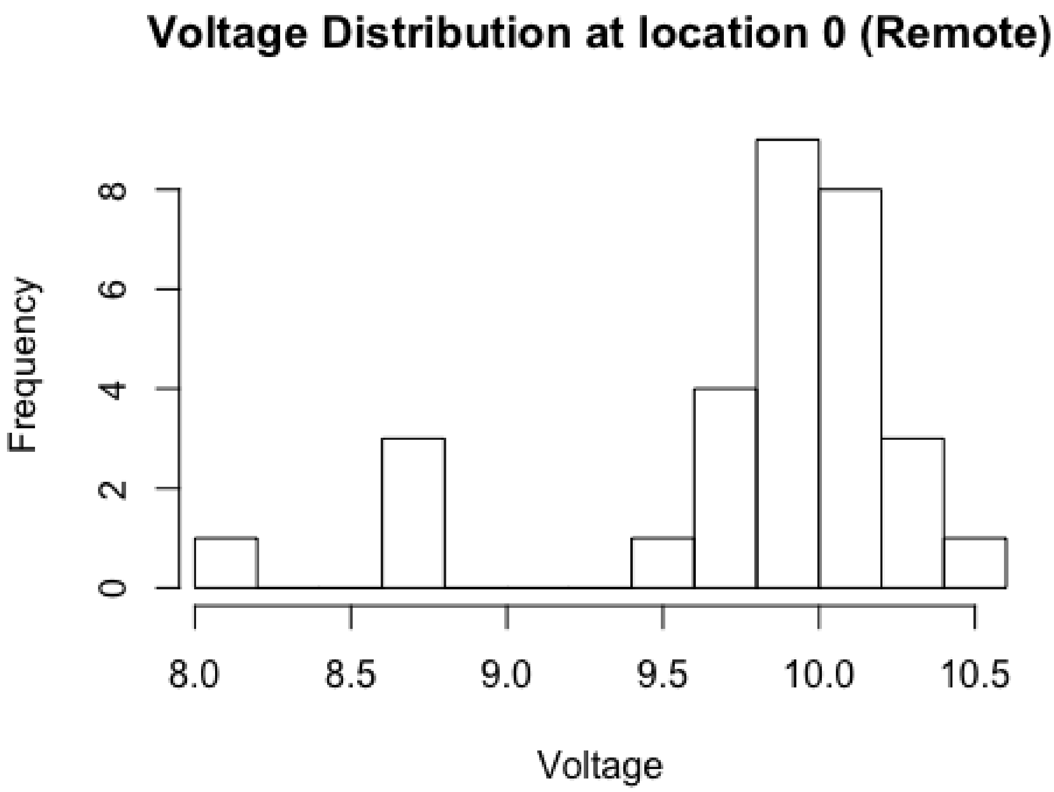
```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = data1, statistic = p.npar, R = 999, sim = "ordinary",
##     stype = "i")
##
##
## Bootstrap Statistics :
##      original       bias    std. error
## t1* 0.2694818 0.005299146   0.1040431
```

```r
# Get the 95% confidence interval
# Percentile bootstrap method
sort(p.npar.boot$t)[c(25, 975)]
```

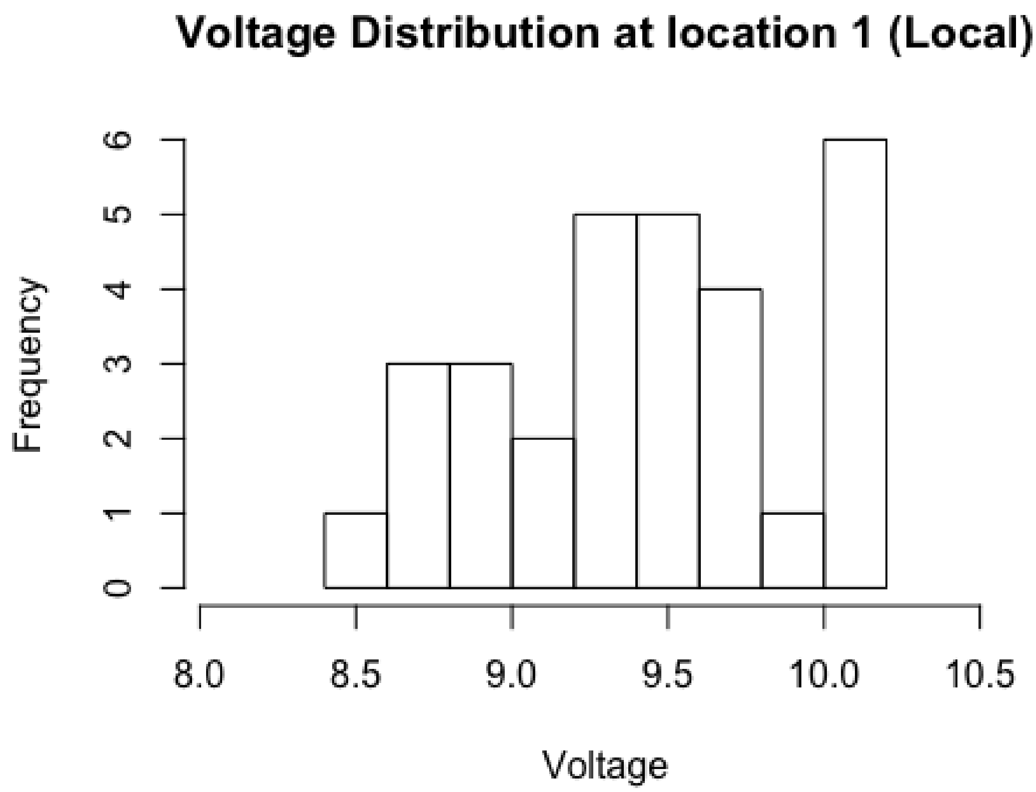
```
## [1] 0.07097203 0.48506268
```

Second dataset:

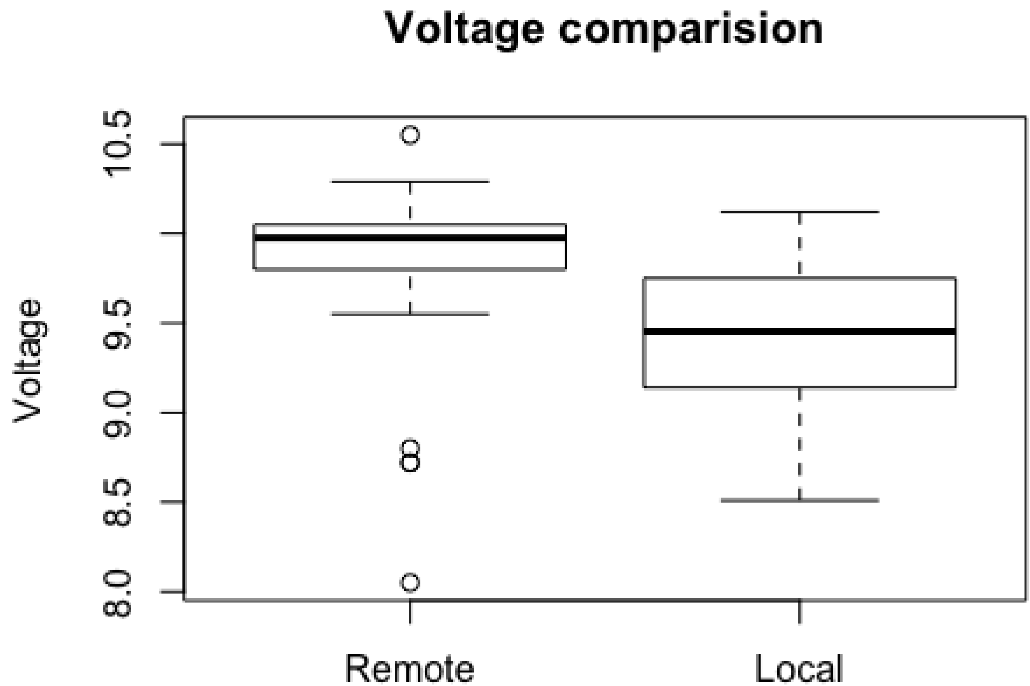
```r
file2 <- "VOLTAGE.csv"
data2 <- read.csv(file2,header = T)
loc0 <- data2[ which(data2$location==0), ]$voltage
loc1 <- data2[ which(data2$location==1), ]$voltage
# (a) explore data distribution
VoltRange = c(min(data2$voltage),max(data2$voltage))
hist(loc0, main="Voltage Distribution at location 0 (Remote)",xlim=VoltRange,
     xlab="Voltage",breaks=10)
```




```
hist(loc1, main="Voltage Distribution at location 1 (Local)",xlim=VoltRange,
     xlab="Voltage",breaks=10)
```




```r
boxplot(loc0,loc1, main="Voltage comparision", names=c("Remote","Local"),
        ylab="Voltage")
```




```
# (b) find 95% confidence interval of U0-U1
#n <- length(loc0)
#m <- length(loc1)
#xbar0 <- mean(loc0)
#xbar1 <- mean(loc1)
#svar0 <- var(loc0)
#svar1 <- var(loc1)
#df = n + m - 2
#pooled.var <- ((n - 1) * svar0 + (m - 1) * svar1)/(df)
#xbar0-xbar1 + c(-1,1) * qt(1-(1-0.95)/2, df=df) * sqrt(pooled.var*((1/n) +
(1/m)))
t.test(loc0, loc1, conf.level=0.95, var.equal=FALSE, mu=0,
alternative="two.sided")

##
##  Welch Two Sample t-test
##
## data:  loc0 and loc1
## t = 2.8911, df = 57.16, p-value = 0.005419
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1172284 0.6454382
## sample estimates:
```

```
## mean of x mean of y
##  9.803667  9.422333
```

Third dataset:

```r
file3 <- "VAPOR.csv"
data3 <- read.csv(file3,header = T)
theory = data3$theoretical
actual = data3$experimental

D <- theory-actual              # difference of the paired data
xd <- mean(D)                   # sample Xbar
numrow <- nrow(data3)           # n
xsd = sd(D)                     # sample standard deviation
tdf = (xd) / (xsd/sqrt(numrow)) # t value
p.val = 2*(1-pt(tdf, numrow-1)) # p value
```