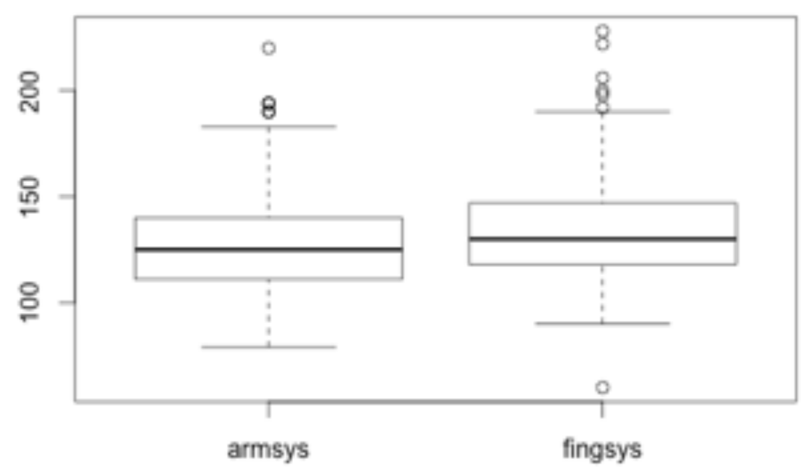
Name: Ajay Vembu
Mini Project: #4

Exercise - 1:

Part - a:

The box plot between the two variables is given as below,



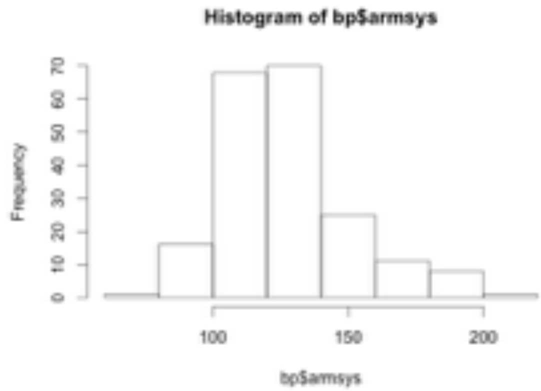
Below are the observations from the above box plot

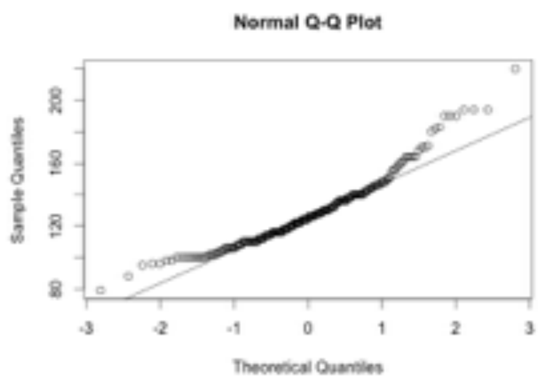
- The quantiles of the fingsys method (Q1,M,Q3) is slightly greater than the quantiles of the armies method.
- The distribution of the data of armsys method seems normal but the distribution of the data of the fingsys method seems slightly right skewed as | Q3 - M | > | Q1 - M |

Part - b:

Histogram of the plot:

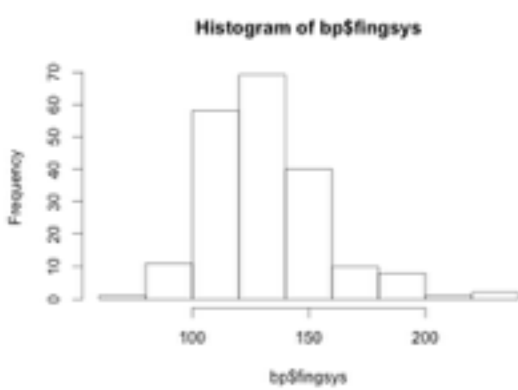
armsys - method:

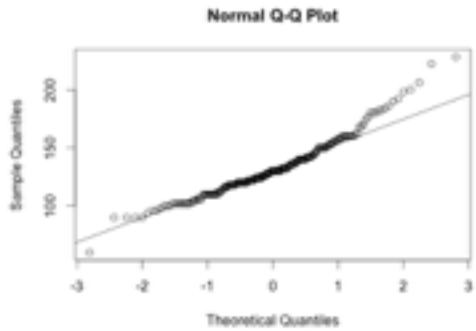




The histogram shows that the plot is slightly right skewed. The QQ plot also confirms to it by a showing a small right tail at the leading end of the ramp.
Also for the above data,
Mean = 128.52 and Median = 125 i.e Mean > Median indicates that the data is right skewed.

fingsys - Method:





The histogram shows that the plot is slightly right skewed. The QQ plot also confirms to it by a showing a small right tail at the leading end of the ramp.
Also for the above data,
Mean = 132.815 and Median = 130 i.e Mean > Median indicates that the data is right skewed.

Part - c:

The confidence interval between the two data is given as,
(MUarmsys - MUfingsys) = (-6.316529, -2.273471)

This shows that there is a difference between the mean of the armies method and the fingsys method. i.e the mean of the fingsys method is clearly greater than the armsys method as there is no 0 in the obtained confidence interval.
The assumptions made were

- The distributions are not normal and large sample case, and
- The quantiles of fingsys method were greater than the quantiles of the armsys method.

The assumptions seems to hold as there a difference in mean between the two methods.

Part - d:

Below are the hypothesis made,

H0: xBar - yBar = 0
H1: xBar - yBar != 0

xBar - mean of armsys
yBar - mean of fingsys

The obtained pValue in R is 3.124491e-05 < 0.05 (the level of significane)

This shows that the null hypothesis is not true that is the assumption that the two means were equal is not true.

Part - e:

The results from the part - c & d is consistent as part - c indicates the difference in mean (i.e no 0 in the confidence interval) and part - d indicates the same i.e. the null hypothesis is not true.

Exercise 2:

Part - a:

The hypothesis is given as,

H0: mu0 <= 10 (NULL) or mu0 = 10 (NULL)
H1: mu0 > 10 (ALTERNATIVE)

Part - b:

- We will use the single sample T test.
- The null distribution is T - distribution and right sided.

Part - c:

The test statistic is given as tOBS = -1.9742

Part - d:

The pValue is computed to be 0.9684606

Part - e:

The pValue computed by the monte carlo simulation is 0.915422

The pValue computed in part d & e are almost equal and found to be greater than 0.90

Part - f:

alpha = 0.05 (5% level of significance)

and pValue > 0.05 in both the parts both d & e.

This shows that the data obtained is strongly in conjunction with the NULL hypothesis hence we accept the null hypothesis and reject the alternative hypothesis.

Exercise - 3:

Part - a:

The confidence interval is constructed for the non normal large sample case and found out to be, ci (JanMonth - MayMonth) = (-302.8289,-201.1711)

This shows that the two means are not equal (No 0 in the interval). The mean of the credits of May month is greater than the mean of the credits of Jan month.

Part - b:

The test hypothesis is given as

H0: may.mean - jan.mean = 0 (NULL) or may.mean - jan.mean <= 0 (NULL)
H1: may.mean - jan.mean > 0 (ALTERNATIVE)

Non normal large sample case is used and the choice of statistic is t-statistic or z-statistic and right sided.

The pValue is found out to be 0

and pValue < alpha (at 5 % level of significance)

This shows that NULL hypothesis is not true and we reject the NULL and accept the ALTERNATIVE hypothesis.

R - code :

Exercise - 1

Part - a and b:

```
# to find the box plot of the two data's
boxplot(bp$armsys,bp$fingsys)

# to find the histogram of the data and qqplot
hist(bp$armsys)
qqplot(bp$armsys);qqline()

hist(bp$fingsys)
qqnorm(bp$fingsys);qqline(bp$fingsys)
```

Part - c:

```
# confidence interval for difference in two methods means

diffMethod <- bp$armsys - bp$fingsys
diffMethod.mean <- mean(diffMethod)
diffMethod.var <- var(diffMethod)
n <- nrow(bp)
alpha <- 1 - 0.95

# confidence interval for non-normal large sample case
```

```
diffMethod.ci <- diffMethod.mean + c (-1,+1) * qnorm(1 - alpha / 2) * sqrt(diffMethod.var / n)
print(diffMethod.ci)
```

Part - d:

```
#Hypothesis test

# H0: xBar - yBar = 0
# H1: xBar - yBar != 0

alpha <- 0.05
diffMethod <- bp$armsys - bp$fingsys

# hypothesis test for non-normal large sample case

thetaHat <- mean(diffMethod)
thetaDelta <- 0
n <- nrow(bp)

se <- sd (diffMethod) / sqrt(n)

zObs <- (thetaHat - thetaDelta) / se

# two sided case

pValue <- 2 * (1 - pnorm(abs(zObs)))

print(pValue < alpha)

# the answer is TRUE which shows that the NULL hypothesis that there is not diff is false.
```

Exercise - 2

Part - c & d:

```
# part - c

thetaHat <- 9.02
theta <- 10
sd <- 2.22
n <- 20

tObs <- ( thetaHat - theta ) / ( sd / sqrt (n) )

print(tObs)

# part - d
```

```
# right sided T distribution

pValue <- 1 - pt ( tObs, n - 1 )

print(pValue)
```

Part - e:

```
# part - e
n <- 20
montecarloReplications <- 10000

montecarloResult <- replicate(montecarloReplications, {
  twentyObs <- rnorm(n, mean = 9.02, sd = 2.22 )
  thetaHat <- sum(twentyObs) / n
  sd <- sd(twentyObs)
  tObs <- ( thetaHat - theta ) / ( sd / sqrt (n) )

  pValue <- 1 - pt ( tObs, n - 1 )
  return (pValue)
})
# find the Pvalue by proportion of the monte carlo result
pValueByProptions <- sum(montecarloResult) / length(montecarloResult)
print(pValueByProptions)
```

Exercise - 3:

Part - a:

```
# difference in means

Jan.n <- 400
Jan.mean <- 2635
Jan.sd <- 365

May.n <- 500
May.mean <- 2887
May.sd <- 412

# confidence interval for non normal large sample case

alpha <- 1 - 0.95

ci.diff.mean <- ( Jan.mean - May.mean ) + c (-1,+1) * qnorm(1 - alpha / 2) * sqrt(Jan.sd^2 / Jan.n
+ May.sd^2 / May.n)

print(ci.diff.mean)
```

Part - b:

```
#part - b

Jan.n <- 400
Jan.mean <- 2635
Jan.sd <- 365

May.n <- 500
May.mean <- 2887
May.sd <- 412

alpha <- 0.05

# H0: may.mean - jan.mean = 0
# H1: may.mean - jan.mean > 0

delta <- 0

# non normal two sample large case

# tStat or Zstat to be used

zObs <- ( ( May.mean - Jan.mean ) - delta) / sqrt (Jan.sd^2 / Jan.n + May.sd^2 / May.n)

print(zObs)

# right sided T distribution

pValue <- 1 - pnorm (zObs)
alpha <- 0.05

print(pValue < alpha)
```