**Mini Project 4**

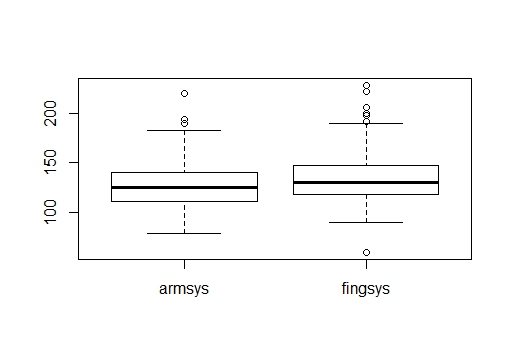
**Group members:** Arnav Sharma (axs144130), Divyanshu Paliwal (dxp151630)

**Contribution of each member:** Both members contributed equally to the following tasks:

* Writing the R code
* Generating graphs
* Deriving conclusions

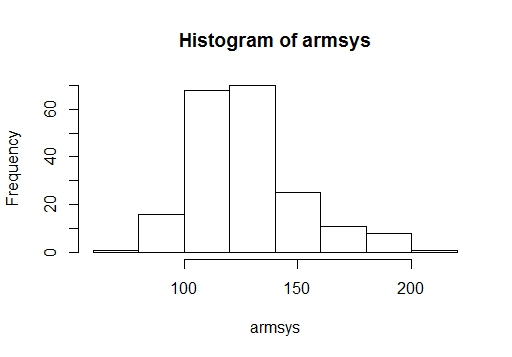
**Exercise 1**

1. The **box plots** for the two methods are below

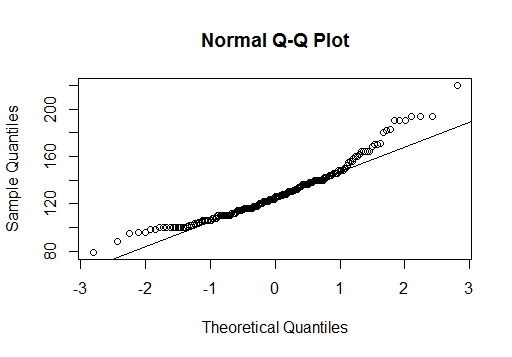
The two distributions seem almost similar except the fact that the finger method has more outliers than the arm method.

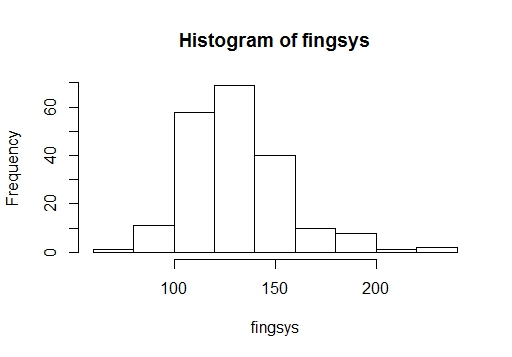
1. Yes, using central limit theorem, we can assume normality.

**Histogram for Arm method**

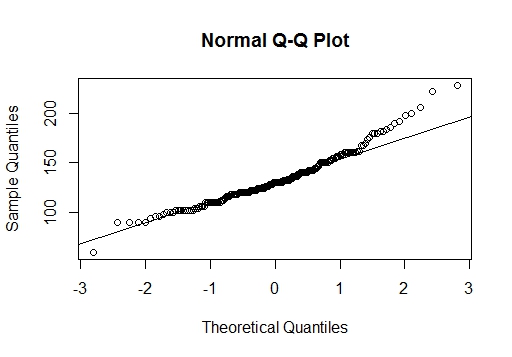


**Q-Q plot for the arm method**



**Histogram of the Finger method**

**Q-Q plot for the Finger method**



1. The two methods do not have identical means.

Mean for Arm method = 128.52 and

Mean for Finger method = 132.815

95% CI = [-7.60, -0.99]

During the construction of CI we assumed that the distributions for both methods are approximately normal and both methods have unequal standard deviations. Our assumptions seem to hold.

1. Null hypothesis: Ho : mean.Arm\_method = mean.Finger\_method

Alternative hypothesis: Hx : mean.Arm\_method != mean.Finger\_method

Given α = 0.05

Calculated P- value = 0.07

As P-value > α

We accept the null hypothesis that there is no difference between the two means.

We assumed approximately normal distribution using central limit theorem.

1. The results from C) and D) do not seem consistent as according to C) the means are unequal but according to D) the means are equal.

**Exercise 2**

1. Null hypothesis: Ho : Mean = 10

Alternative hypothesis: Hx : Mean != 10

1. We have used T-Test.

Test statistic = (mean x – mean.y)/(sd/sqrt(n))

Distribution of test statistic = T distribution

1. Observed value of test statistic = -1.97
2. P- value using usual way = 0.97. Therefore, We accept the null hypothesis.
3. P- value using Monte Carlo = 1. On 1000 trials, no. of times pval < 0.05 is 0. Therefore, using Monte Carlo method, we accept the null hypothesis.

Clearly P-value usual < P-value Monte

1. At 5% level of significance

We accept the null hypothesis as P-value > α

**Exercise 3**

1. 95% confidence interval for the difference in mean credit limits of all credit cards issued in January 2011 and in May 2011 is

95% CI = [201.1711 302.8289]

1. Null hypothesis: Ho : Mean. May = Mean. January

Hx : Mean. May > Mean.January

Using a Z- test Z statistic = 9.17 and P-value = 0

Thus we reject the null hypothesis and accept the alternative. Therefore, Mean.May > Mean.January.

Mean credit limit of all credit cards issued in May 2011 is greater than the mean credit limit in January 2011.

**APPENDIX**

**Exercise 1 R code**

#read the file

dat = read.csv("bp.csv", header = TRUE)

#boxplot

boxplot(dat)

x <- dat$armsys

y <- dat$fingsys

#histogram

hist(x)

hist(y)

#q plot

qqnorm(x)

qqline(x)

qqnorm(y)

qqline(y)

#standard deviation and mean of both methods

sx <- sd(x)

sy <- sd(y)

mx <- mean(x)

my <- mean(y)

#number of entries

n <- nrow(dat)

#standard error calculation

se <- sqrt((sx^2+sy^2)/n)

#confidence interval calculation

ci <- mx - my + c(-1,1)\*qnorm(0.975)\*se

#Z statistic

z <- (mx-my)/se

#p value

pval <- 2\*(1 - pnorm(abs(z)))

**Exercise 2 R code**

#given values of mean and standard deviation

m <- 10

mx <- 9.02

sd <- 2.22

n <- 20

#T statistic

tstat <- (mx - m)/(sd/sqrt(n))

#p value

pval2 <- (1 - pnorm(tstat))

#monte carlo simulation

#function for calculating p-value

test <- function(n, nsim, mu){

x = rnorm(n)

m <- mean(x)

s <- sd(x)

se <- s/sqrt(n)

# T stat for monte carlo

tstat2 <- (m – mu)/se

#p value for monte carlo

pval3 <- 1 - pt(tstat2, n-1)

return(pval3)

}

#function for simulating above function 1000 times and finding no. of times p-value is less than 0.05

mct <- function(n = 20, nsim = 1000, mu = 10){

pvals <- replicate(nsim, test(n, nsim, mu))

psim <- sum(pvals<0.05)/nsim

return (psim)

}

test(20,1000,10)

mct(n = 20, nsim = 1000, mu = 10)

**Exercise 3 R code**

#given values for mean standard deviation and number of samples

x <- 2635

y <- 2887

sd.x <- 365

sd.y <- 412

n.x <- 400

n.y<- 500

#standard error

se <- sqrt((sd.x^2/n.x)+(sd.y^2/n.y))

#confidence interval calculation

ci2 <- y-x +c(-1,1)\*qnorm(0.975)\*se

#Z stat calculation

z <- (y-x)/se

#p val calculation

pval4 <- (1 - pnorm(z))