Preston Robertson
Summer 2022: Data Analysis
Homework V
Due: TBA
Submit through Canvas

Instructions: Provided solutions to these questions using this template. Include graphics with your solutions. Put all code an appendix to this homework. Use the *verbatim* command to leave code unchanged.

1. Use *set.seed(123)* and generate 100 samples of IID $\{X_i\}_{i=1}^N$ where $X_i \sim N(0,1)$ and $N = 10$. Construct side-by-side 90% confidence intervals. Plot the results. How many intervals contained $\mu_X = 0$?

**Solution:**
In Figure (1) side by side 90% confidence intervals are depicted. The graphic depicts results for 100 total groups of size 10. The formula for the confidence interval is

$$\bar{x} \pm t_{0.95}^{df=9} s/\sqrt{9}.$$

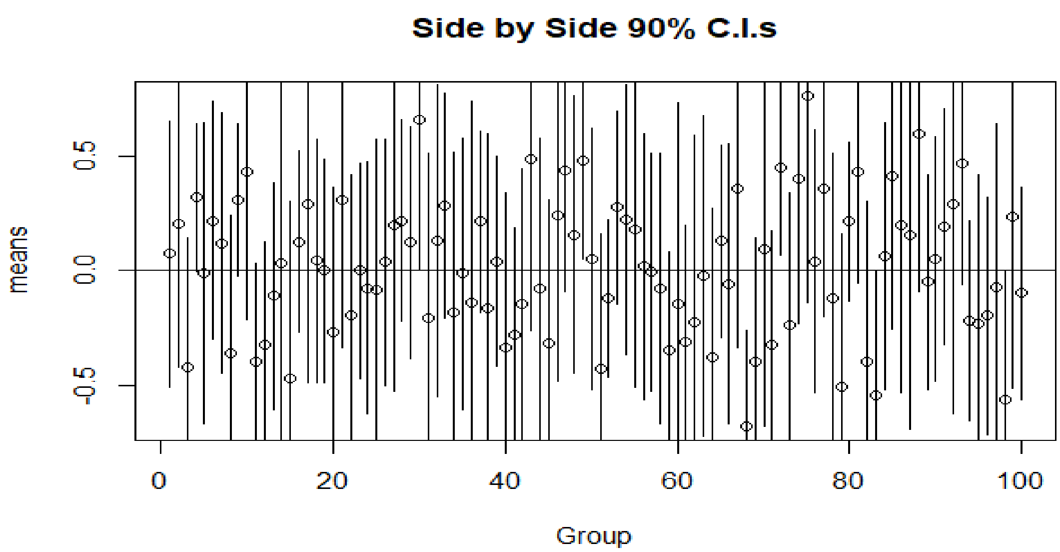


Figure 1:

Note there are a total of 6 groups whose interval does not contain zero.

2. Use *set.seed(123)* directly before simulating data for both parts of this problem.

a) Simulate an IID sample $\{X_i\}_{i=1}^N$ where $X_i \sim N(0,1)$ and $N = 100$. Make QQ-plot of the data. Interpret your results.

**Solution:**
Below, Figure (2) depicts the QQ normal probability plot for 100 normal random variables. Visually, the plot appears to follow a straight line, thus indicating the data is actually normally distributed.

---

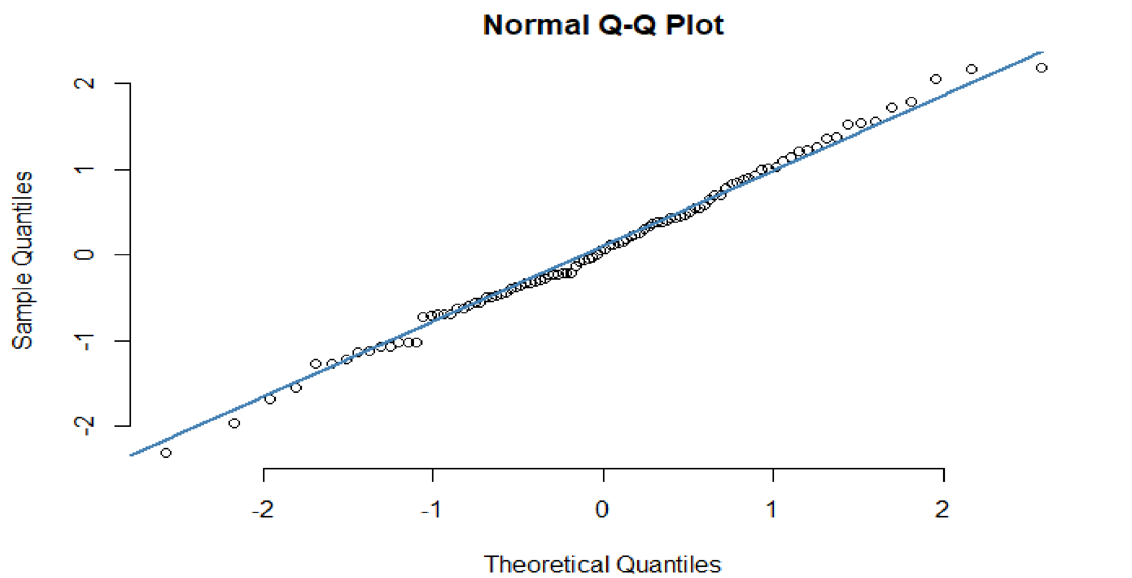


Figure 2: QQ plot for simulated $N(0,1)$ data.

b) Simulate an IID sample $\{Y_i\}_{i=1}^{N}$ where $Y_i \sim$ Chauchy$(0, 1)$. Make a QQ-plot of the data (where theoretical quantiles of the fitted Normal). Interpret your results.

**Solution:**
Figure (3) below depicts a qq normal probability plot applied to 100 simulated Cauchy(0,1) random variables. Note that the tails of distribution depart from the theoretical quantities associated with Gaussian random variables.

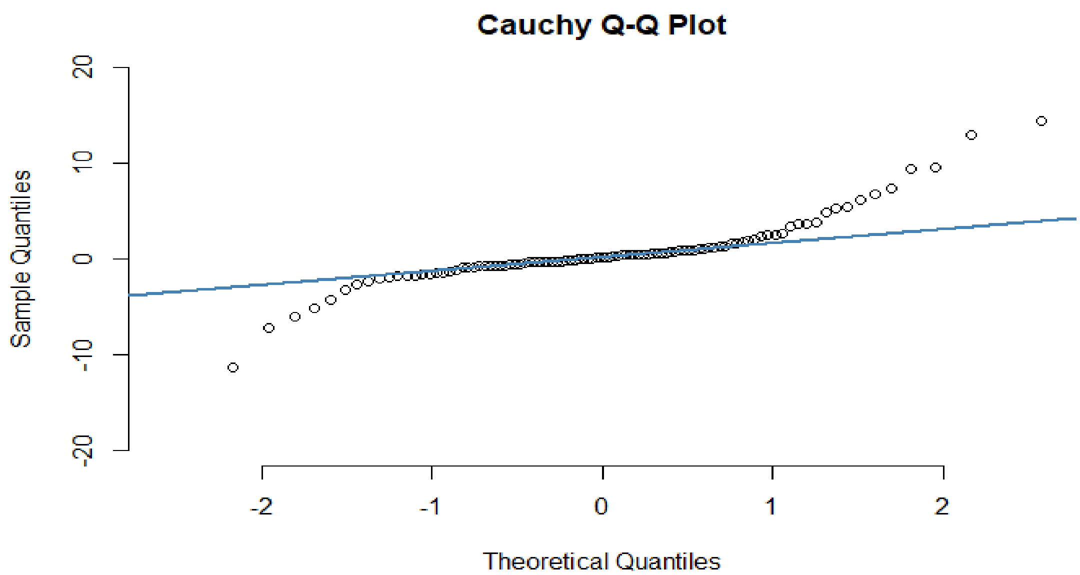


Figure 3: Chauchy(0,1) random variables are more likely to have very extreme outliers than Gaussian random variables.

## Appendix: Code

1. Below is the code used to produce the side by side confidence intervals.

~~~
```{r}



######## Question 1 ########

set.seed(123)
dat1 <- rep(0,1000)
dat2 <- rep(0,1000)
dat  <- cbind(dat1,dat2)
t    <- c(1:100)
s    <- c(1:10)

for (i in t) {
  for (j in s){

    dat[(i-1)*10+j,1] = i
    dat[(i-1)*10+j,2] = rnorm(1,0,1)


  }
}

```



```{r}

#### Quick Diagnostic ####

hist(dat[,2]) # Seems ok ##

```


```{r}
~~~

```
## Computing means and standard deviations ##

means <- rep(0,100)
sds   <- rep(0,100)

for (i in t){

  t1 = 10*(i-1)+1
  t2 = 10*i

  means[i] = mean(dat[t1:t2,2])
  sds[i]   = sd  (dat[t1:t2,2])

}

‘‘‘


‘‘‘{r}


# Plotting both as quick diagnostic #
plot(t,means)
hist(sds)

‘‘‘


‘‘‘{r}


## Generating box plots ##

boxplot(dat[,2]~dat[,1])
abline(0,0)

‘‘‘

################################


‘‘‘{r}


t_val = qt(0.95,9,lower.tail = TRUE,log.p = FALSE)
```

```
plot(t,means,ylim = c(-2,2), main= "Side by Side 90% C.I.s",
     xlab = "Group", ylab = "Means", family = "TNR")
abline(0,0)

for (i in t){

  lines(c(i,i),c(means[i]-t_val*sds[i]/3,means[i]+t_val*sds[i]/3))

}

‘‘‘


‘‘‘{r}


#### Counting number of CIs that don’t contain 0 ####

count = 0


for (i in t){

  if (abs(means[i]) > abs(t_val * sds[i]/3)) {

    count = count + 1
  }

}

count

‘‘‘

################################################################


‘‘‘{r}
######## Question 2a ########

set.seed(123)
normdat = rnorm (100,0,1)
qqnorm(normdat, pch = 1, frame = F, family = "TNR")
qqline(normdat, col = "steelblue", lwd = 2)
```

~~~
```

```{r}
######## Question 2b ########

set.seed(123)
cauchdat = rcauchy (100,0,1)
qqnorm(cauchdat, pch = 1, frame = F, family = "TNR", ylim = c(-20,20),  main = "Cauchy Q-Q
qqline(cauchdat, col = "steelblue", lwd = 2)

```
~~~