

Homework 09 - STAT440

Joseph Sepich (jps6444)

11/08/2020

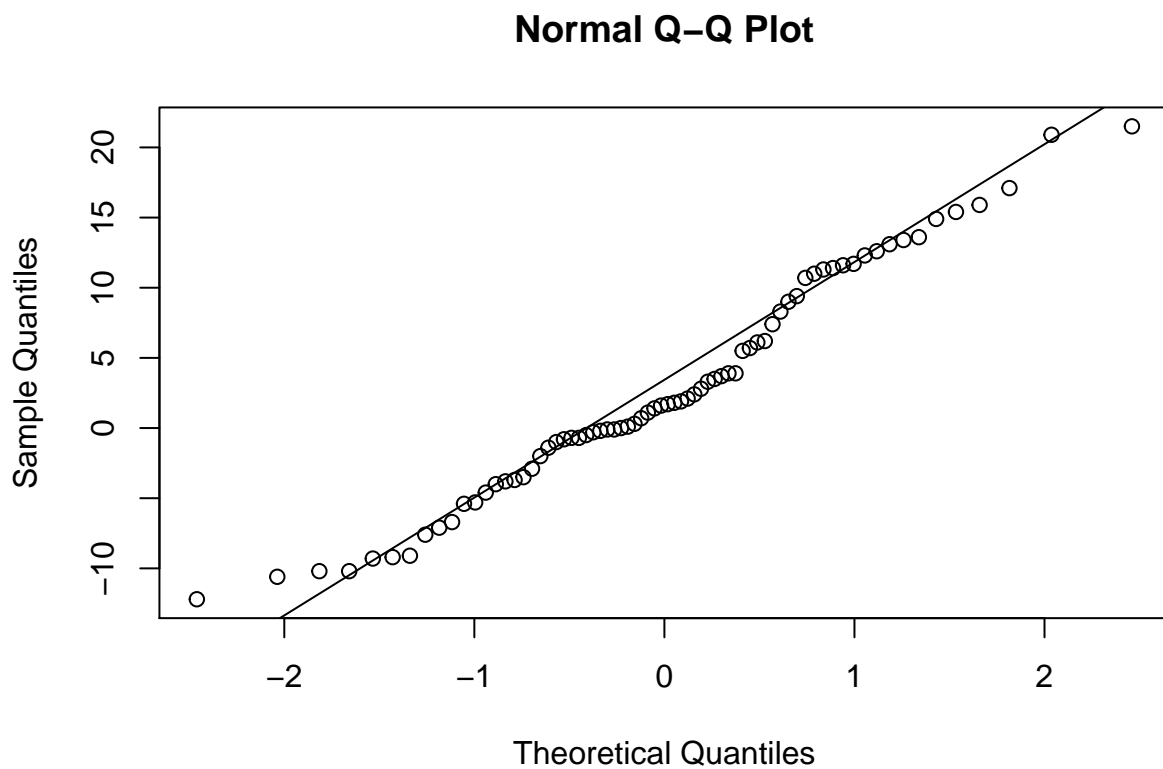
```
set.seed(42)
```

Problem 1

```
df$Weight_Increase <- df$Postwt - df$Prewt
```

Part a

```
qqnorm(df$Weight_Increase)  
qqline(df$Weight_Increase)
```

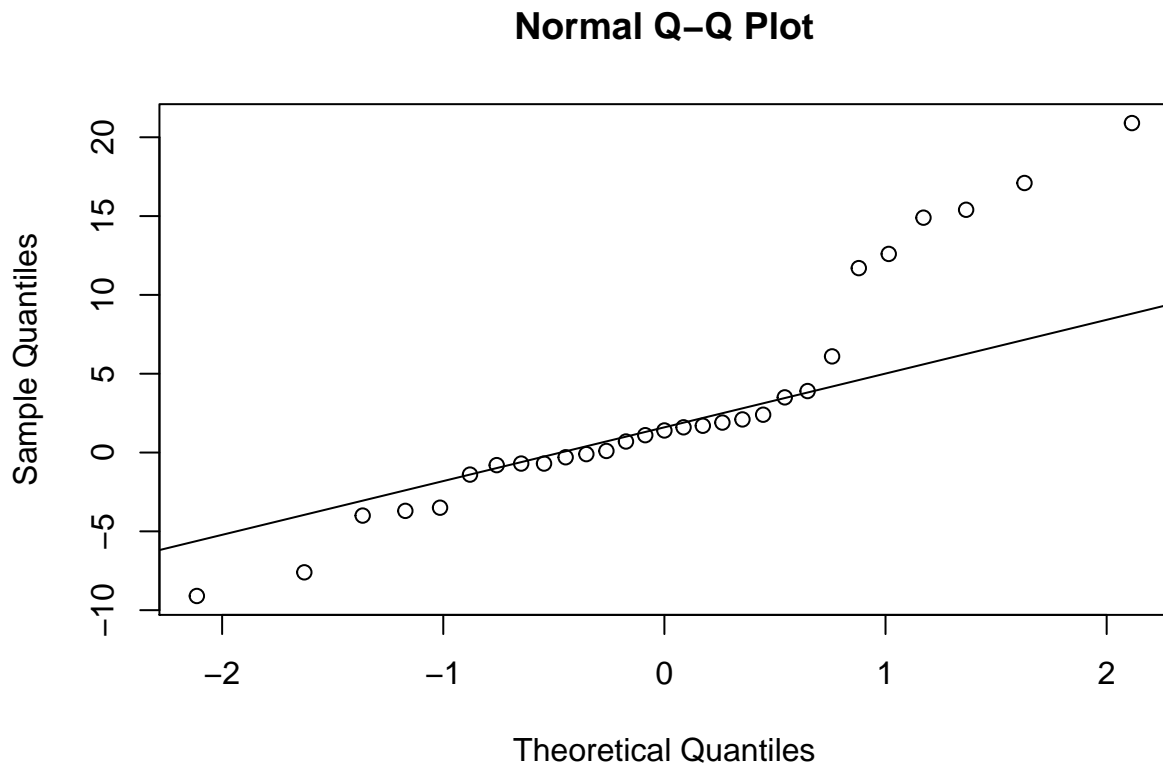


```
#qqplot(y, rt(300, df = 5))
```

The results follow the diagonal line (theoretical quantile match) fairly closely, but the tail to the left is steeper than the tail on the right compared to the normal, since the data points lie under the normal line of quantiles.

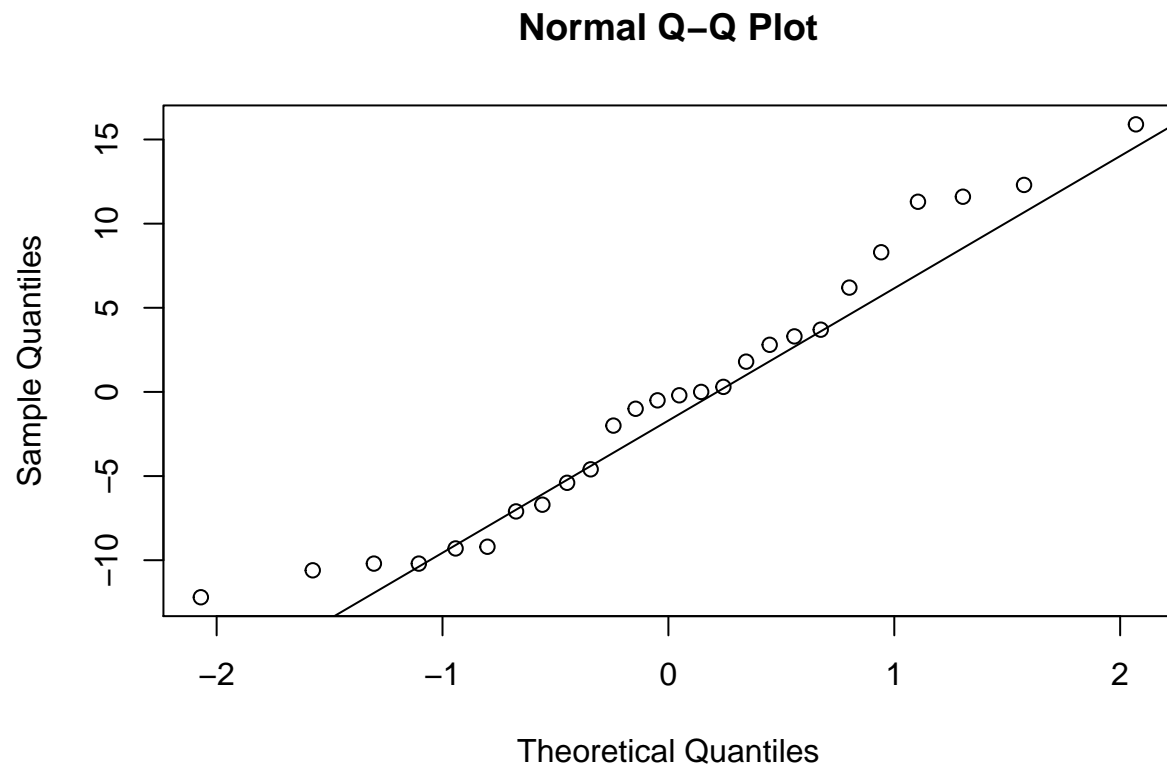
Part b

```
# CBT
df %>%
  filter(Treat == 'CBT') %>%
  .$Weight_Increase %>%
  qqnorm()
df %>%
  filter(Treat == 'CBT') %>%
  .$Weight_Increase %>%
  qqline()
```

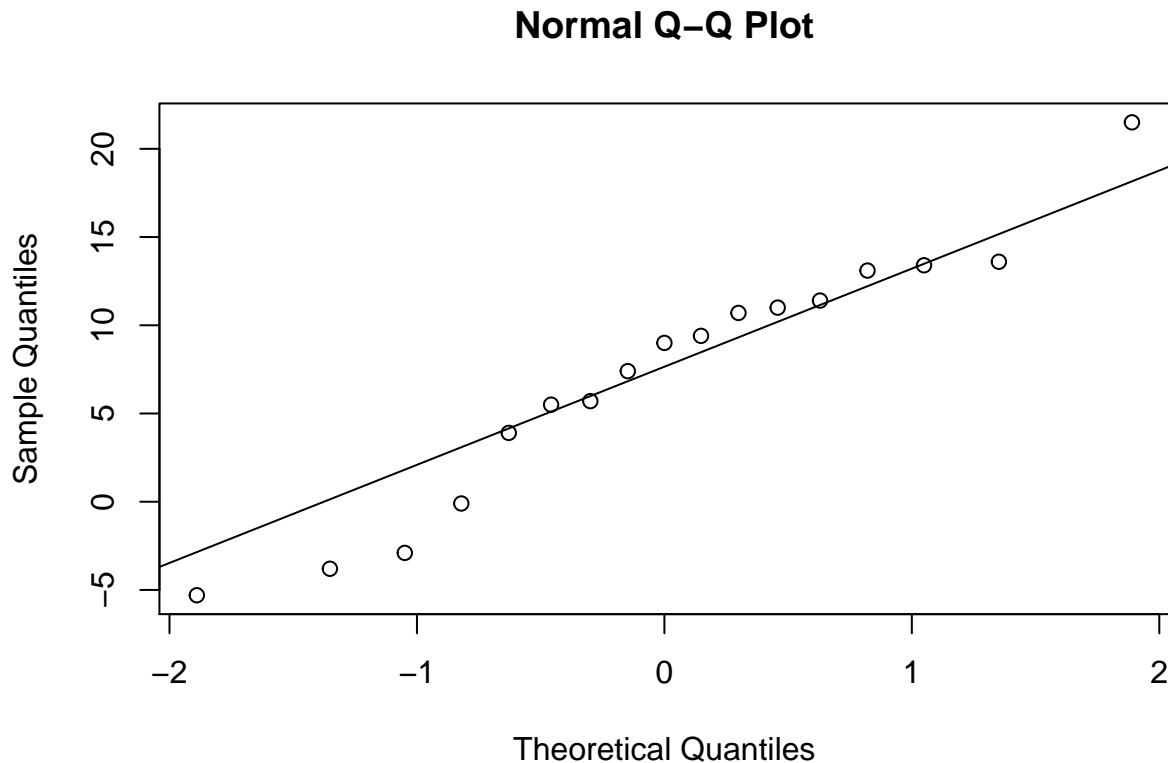


```
# Cont
df %>%
  filter(Treat == 'Cont') %>%
  .$Weight_Increase %>%
  qqnorm()
```

```
df %>%
  filter(Treat == 'Cont') %>%
  .$Weight_Increase %>%
  qqline()
```



```
# FT
df %>%
  filter(Treat == 'FT') %>%
  .$Weight_Increase %>%
  qqnorm()
df %>%
  filter(Treat == 'FT') %>%
  .$Weight_Increase %>%
  qqline()
```



Each of the three groups roughly follow a gaussian distribution. Actually a lot of the middle quantiles match up for the CBT group, which is promising and this trend of having the middle quantiles match up also happens with the other two treatment groups, so we likely can approximate the distribution using a gaussian distribution.

Problem 2

Part a

The conjugate prior to the likelihood for the normal distribution is the normal distribution. Our posterior is then also a normal distribution.

Part b

The mean of the prior is essentially our initial assumption about where the mean is for the true distribution. I will therefore set the prior parameter $\mu_0 = 0$ which would operate under the assumption that the average weight increase after treatment is zero. This would essentially say we expect there to be no weight increase after treatment. Using this as our prior parameter can help us to see how much our evidence shows of an actual weight increase by looking at the posterior parameter we receive from the evidence. (We do not need to decide on a prior variance since we are operating on the assumption that our sample variance is the fixed know variance.)

Part c

$$\sigma_0^2 \rightarrow \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1}$$

$$\mu_0 \rightarrow \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1} \left(\frac{n\bar{X}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)$$

```
update_sigma <- function(x, prior_sig, sig=1) {  
  (length(x) / sig^2 + 1 / prior_sig^2) ^ -1  
}  
  
update_mu <- function(x, prior_mu, prior_sig, sig=1) {  
  update_sigma(prior_sig, sig) * (length(x) * mean(x) / sig^2 + prior_mu / prior_sig^2)  
}
```

```
df_cbt <- df %>%  
  filter(Treat == 'CBT')  
df_cont <- df %>%  
  filter(Treat == 'Cont')  
df_ft <- df %>%  
  filter(Treat == 'FT')  
  
prior_sigma_cbt <- sd(df_cbt$Weight_Increase)  
prior_mu_cbt <- 0  
prior_sigma_cont <- sd(df_cont$Weight_Increase)  
prior_mu_cont <- 0  
prior_sigma_ft <- sd(df_ft$Weight_Increase)  
prior_mu_ft <- 0
```

```
print(paste0('Prior CBT: ', prior_sigma_cbt, ' ', prior_mu_cbt))
```

```
## [1] "Prior CBT: 7.30850439236387 0"
```

```
print(paste0('Prior Cont: ', prior_sigma_cont, ' ', prior_mu_cont))
```

```
## [1] "Prior Cont: 7.98870452576637 0"
```

```
print(paste0('Prior FT: ', prior_sigma_ft, ' ', prior_mu_ft))
```

```
## [1] "Prior FT: 7.15742107679772 0"
```

```
post_sigma_cbt <- update_sigma(df_cbt$Weight_Increase, prior_sigma_cbt, prior_sigma_cbt)  
post_mu_cbt <- update_mu(df_cbt$Weight_Increase, prior_mu_cbt, prior_sigma_cbt, prior_sigma_cbt)  
post_sigma_cont <- update_sigma(df_cont$Weight_Increase, prior_sigma_cont, prior_sigma_cont)  
post_mu_cont <- update_mu(df_cont$Weight_Increase, prior_mu_cont, prior_sigma_cont, prior_sigma_cont)  
post_sigma_ft <- update_sigma(df_ft$Weight_Increase, prior_sigma_ft, prior_sigma_ft)  
post_mu_ft <- update_mu(df_ft$Weight_Increase, prior_mu_ft, prior_sigma_ft, prior_sigma_ft)
```

```
print(paste0('Posterior CBT: ', post_sigma_cbt, ' ', post_mu_cbt))
```

```
## [1] "Posterior CBT: 1.78047454844007 1.60252179730565"
```

```
print(paste0('Posterior Cont: ', post_sigma_cont, ' ', post_mu_cont))
```

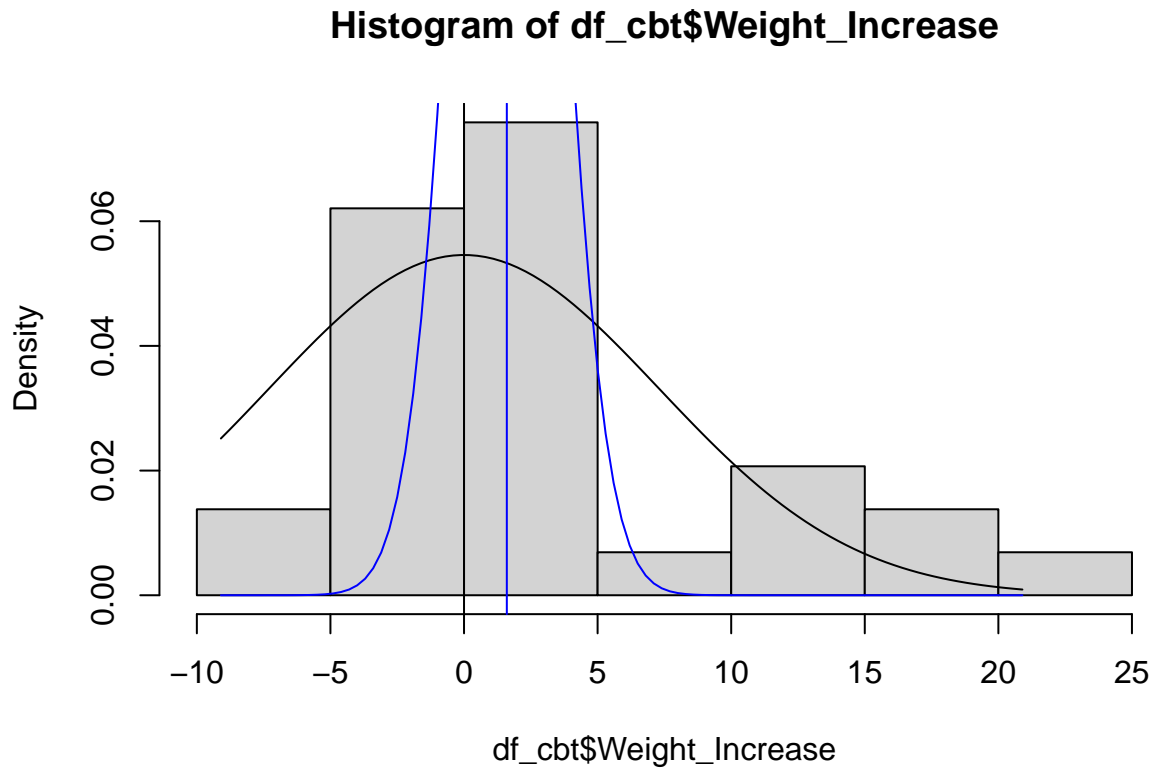
```
## [1] "Posterior Cont: 2.36368148148148 -0.180501516521288"
```

```
print(paste0('Posterior FT: ', post_sigma_ft, ' ', post_mu_ft))
```

```
## [1] "Posterior FT: 2.84603758169935 2.36460137123228"
```

Part d

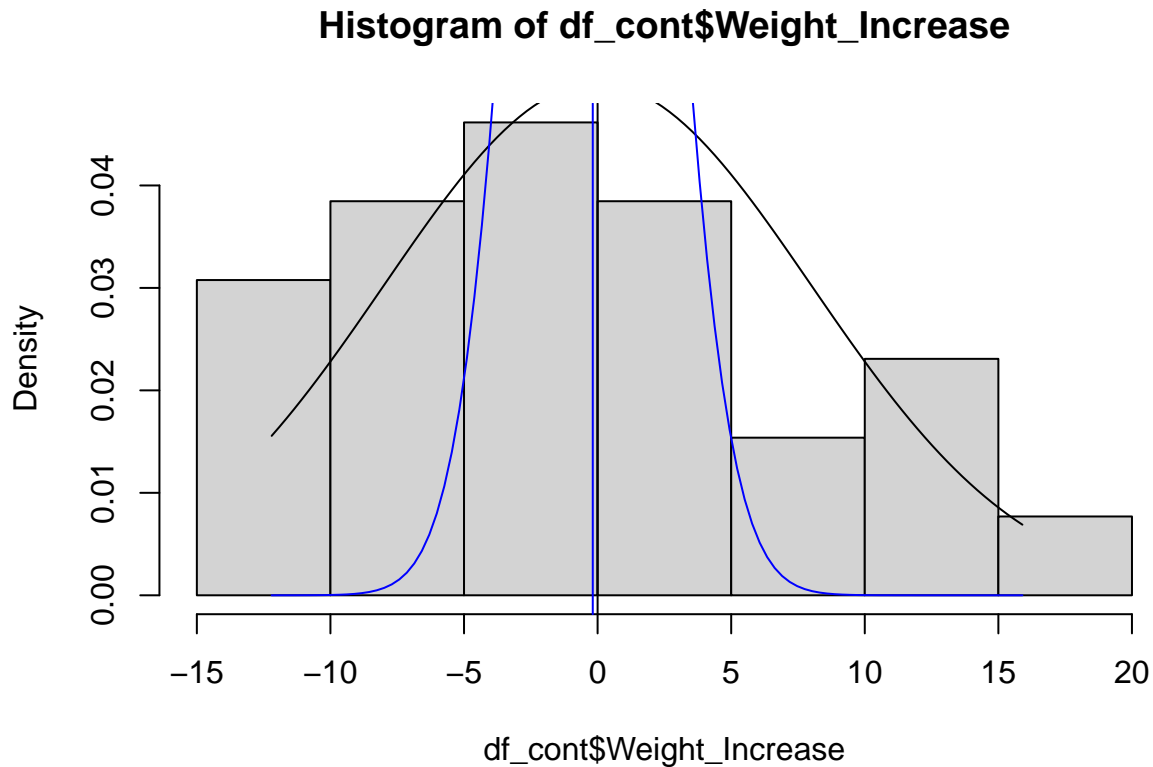
```
min_x <- min(df_cbt$Weight_Increase)
max_x <- max(df_cbt$Weight_Increase)
hist(df_cbt$Weight_Increase, freq=F)
curve(dnorm(x,mean=prior_mu_cbt,sd=prior_sigma_cbt), min_x, max_x, add=T, col='black')
abline(v=prior_mu_cbt,col='black')
curve(dnorm(x,mean=post_mu_cbt,sd=post_sigma_cbt), min_x, max_x, add=T, col='blue')
abline(v=post_mu_cbt,col='blue')
```



```

min_x <- min(df_cont$Weight_Increase)
max_x <- max(df_cont$Weight_Increase)
hist(df_cont$Weight_Increase, freq=F)
curve(dnorm(x,mean=prior_mu_cont,sd=prior_sigma_cont), min_x, max_x, add=T, col='black')
abline(v=prior_mu_cont,col='black')
curve(dnorm(x,mean=post_mu_cont,sd=post_sigma_cont), min_x, max_x, add=T, col='blue')
abline(v=post_mu_cont,col='blue')

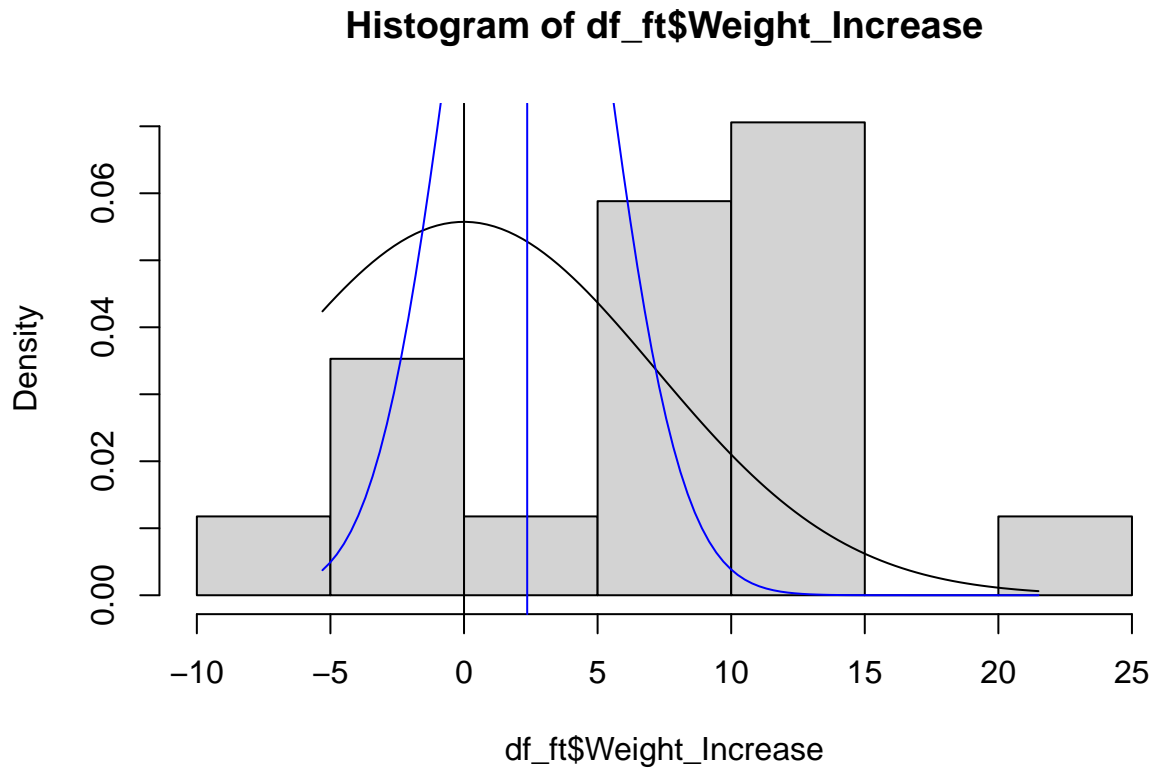
```



```

min_x <- min(df_ft$Weight_Increase)
max_x <- max(df_ft$Weight_Increase)
hist(df_ft$Weight_Increase, freq=F)
curve(dnorm(x,mean=prior_mu_ft,sd=prior_sigma_ft), min_x, max_x, add=T, col='black')
abline(v=prior_mu_ft,col='black')
curve(dnorm(x,mean=post_mu_ft,sd=post_sigma_ft), min_x, max_x, add=T, col='blue')
abline(v=post_mu_ft,col='blue')

```



The above charts have the actual data plotted a histogram, the black curve is the prior distribution, and the posterior distribution is the blue curve. We can see that in each of the three plots the posterior mean moved from zero toward the mean of the data points, or evidence, we have in our data set. It is also important to note that the variance of the posterior distribution also decreased for each plot. This decrease indicates more confidence in our posterior estimate of μ , since the posterior has more evidence behind it than the prior. Along the same line of thinking the group with the most data point had the smallest posterior variance.

Part e

```
M <- 1000
c_val <- 0.05

samples_cbt <- rnorm(M, post_mu_cbt, post_sigma_cbt)
samples_cont <- rnorm(M, post_mu_cont, post_sigma_cont)
samples_ft <- rnorm(M, post_mu_ft, post_sigma_ft)

prob_vec <- c(c_val, 1-c_val)
ci_cbt <- quantile(samples_cbt, probs=prob_vec)
ci_cont <- quantile(samples_cont, probs=prob_vec)
ci_ft <- quantile(samples_ft, probs=prob_vec)
```

```
print(paste0('CBT Cred Int: ', ci_cbt))
```

```
## [1] "CBT Cred Int: -1.32704024556603" "CBT Cred Int: 4.33285652875773"
```



```
print(paste0('Cont Cred Int: ', ci_cont))
```

```
## [1] "Cont Cred Int: -4.06134687688708" "Cont Cred Int: 3.64911468805187"
```

```
print(paste0('FT Cred Int: ', ci_ft))
```

```
## [1] "FT Cred Int: -2.68692425774453" "FT Cred Int: 7.14890869430698"
```

According to this analysis our evidence is not credible enough to determine that any of the treatments had an effect that raised (or lowered) weight in the anorexic patients.

Problem 3

Problem 4