Appendix

Q1-a

```
prob1_dt <- read.table('problem1.txt', header=T)
cal_dt <- prob1_dt$calcium
n <- length(cal_dt)

cal_resid <- (cal_dt- mean(cal_dt)) / (500 * sqrt(1 - 1/n))

plot(cal_resid, ylim=c(-4, 4), ylab='Residual', main='Calcium Residual Plot')
abline(h=0)

qqnorm(cal_resid)
abline(a=0, b=1)
```

Q1-b

```
qqnorm(cal_resid)
abline(a=0, b=1)
```

Q1-c

```
cal_r <- cal_dt - mean(cal_dt)
cal_df <- length(cal_r) - 1
D <- sum(cal_r^2) / 500^2
p_value <- pchisq(D, cal_df)
```

Q1-d

```
count <- c(length(cal_dt[cal_dt <= 600]),
           length(cal_dt[cal_dt > 600 & cal_dt <= 1200]),
           length(cal_dt[cal_dt > 1200 & cal_dt <= 1800]),
           length(cal_dt[cal_dt > 1800]))

sigma <- 500
likelifun <- function(x) {
  return(pnorm((600-x)/sigma)^9 * (pnorm((1200-x)/sigma) - pnorm((600-x)/sigma))^20
         * (pnorm((1800-x)/sigma) - pnorm((1200-x)/sigma))^7
         * (1 - pnorm((1800-x)/sigma))^2)
}

theta_seq <- seq(400, 2000)
likeli_value <- likelifun(theta_seq)
plot(x=theta_seq, y=likeli_value, xlab='Theta', ylab='Likelihood')

# find mle theta
```

```r
LL <- function(theta) {
  p1 <- pnorm((600-theta)/sigma)
  p2 <- pnorm((1200-theta)/sigma) - pnorm((600-theta)/sigma)
  p3 <- pnorm((1800-theta)/sigma) - pnorm((1200-theta)/sigma)
  p4 <- 1 - pnorm((1800-theta)/sigma)
  x1 <- 9
  x2 <- 20
  x3 <- 7
  x4 <- 2
  mloglik <- suppressWarnings(- x1 * log(p1) - x2 * log(p2) - x3 * log(p3) - x4 * log(p4))
}

library(stats4)
mle(LL, start=list(theta=1000))

# chi-square test
theta <- 914.2893
p1 <- pnorm((600-theta)/sigma)
p2 <- pnorm((1200-theta)/sigma) - pnorm((600-theta)/sigma)
p3 <- pnorm((1800-theta)/sigma) - pnorm((1200-theta)/sigma)
p4 <- 1 - pnorm((1800-theta)/sigma)
probs <- c(p1,p2,p3,p4)
chisq.test(count, probs)

expected <- probs * 38
D <- sum((count - expected)^2 / expected)
p_value <- 1- pchisq(D, df=2)
```

Q2
```r
# find mle of sigma theta
LL <- function(theta, sigma) {
  p1 <- pnorm((600-theta)/sigma)
  p2 <- pnorm((1200-theta)/sigma) - pnorm((600-theta)/sigma)
  p3 <- pnorm((1800-theta)/sigma) - pnorm((1200-theta)/sigma)
  p4 <- 1 - pnorm((1800-theta)/sigma)
  x1 <- 9
  x2 <- 20
  x3 <- 7
  x4 <- 2
  mloglik <- suppressWarnings(- x1 * log(p1) - x2 * log(p2) - x3 * log(p3) - x4 * log(p4))
}

mle(LL, start=list(theta=1000, sigma=500))

# chi-square test
```

```
theta <- 917.6799
sigma <- 477.9240
p1 <- pnorm((600-theta)/sigma)
p2 <- pnorm((1200-theta)/sigma) - pnorm((600-theta)/sigma)
p3 <- pnorm((1800-theta)/sigma) - pnorm((1200-theta)/sigma)
p4 <- 1 - pnorm((1800-theta)/sigma)
probs <- c(p1,p2,p3,p4)
expected <- probs * 38
D <- sum((count - expected)^2 / expected)
p_value <- 1- pchisq(D, df=1)
```

Q5

```
# problem 5-a
diet_pvalue <- read.table('problem4.txt', header=T, sep=',')

bf_pvalue <- p.adjust(diet_pvalue$P.value, 'bonferroni')
data.frame(Variable=diet_pvalue$Dietary.Variable, Bonferroni.P = bf_pvalue)

# problem 5-b
m <- nrow(diet_pvalue)
ordered_pvalue <- diet_pvalue[order(diet_pvalue$P.value),]
ben_pvalue <- rep(NA, m)
k_reject <- 0
for (ii in 1:m) {
  adj_pvalue <- ordered_pvalue$P.value[ii] * m / ii
  ben_pvalue[ii] <- adj_pvalue
  if (adj_pvalue < 0.05) {
    k_reject <- ii
  }
}
ordered_pvalue[1:k_reject,]

plot(x=1:m, y=ben_pvalue, xlab='K', ylab='Benjamini Adjusted', pch=16)
abline(h=0.05, lty=3, lwd=2)
legend('topleft', legend='Test Size', lty=3, lwd=2, cex=0.8)
```