Homework 4: Diffusion of Tetracycline

DingHuan, 3170102085

1.
```
ckm_nodes <- read.csv("ckm_nodes.csv")
ckm_network <- read.table(("ckm_network.dat"))
anum <- which(!is.na(ckm_nodes$adoption_date))
ckm_nodes <- ckm_nodes[anum, ]
ckm_network <- ckm_network[anum, anum]
rm(anum)
```

2. 6 columns contain 6 variables, and 2125 rows contains 125 doctors, each of whom is separated into 17 rows containing 17 months.

```
n <- max(ckm_nodes %>% filter(!is.infinite(adoption_date)) %>% dplyr::select(adoption_date))
ckm_new <- data.frame(doctor_num = rep(1:nrow(ckm_nodes),n),
                      date = rep(1:n,each = nrow(ckm_nodes))) %>%
  mutate(began_that_month = (ckm_nodes[doctor_num,"adoption_date"]==date),
         before_that_month = (ckm_nodes[doctor_num,"adoption_date"]<date))
before <- ckm_new %>% dplyr::select(doctor_num,date,before_that_month) %>%
  spread(date,before_that_month) %>%
  dplyr::select(-doctor_num)
began <- ckm_new %>% dplyr::select(doctor_num,date,began_that_month) %>%
  spread(date,began_that_month) %>%
  dplyr::select(-doctor_num) %>% apply(1,cumsum) %>% t()
temp <- t(as.matrix(before)) %*% as.matrix(ckm_network) %>% t() %>%
  as.data.frame() %>% gather()
ckm_new[,"contrast_before"] <- temp$value
temp <- t(as.matrix(began)) %*% as.matrix(ckm_network) %>% t() %>%
  as.data.frame() %>% gather()
ckm_new[,"contrast_that_monthnearlier"] <- temp$value
rm(n,temp,before,began)
```

3. a. Because the maximum number of contrasts of one doctor is only 20, which means that no one have more than 20 contrasts. We can estimate p_k and q_k when $k = 0, 1, ..., 20$, 21 values at most.

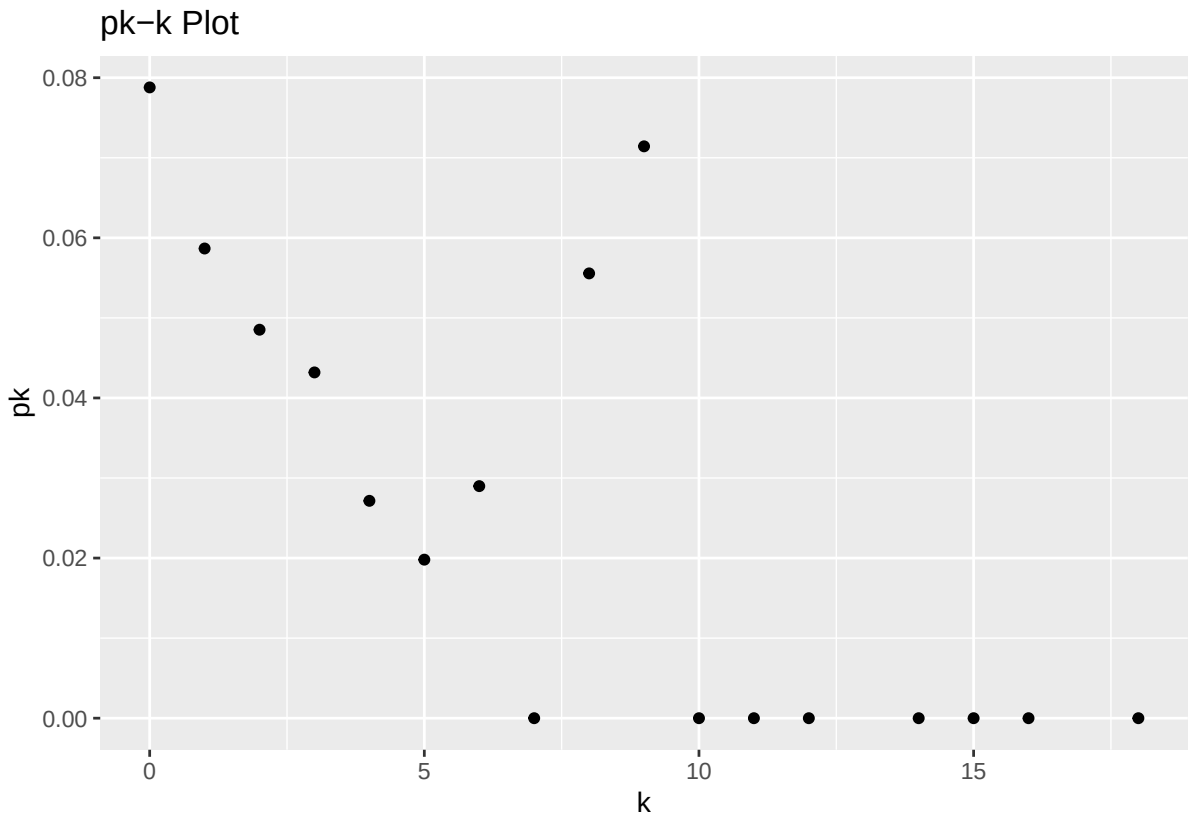
```
max(colSums(ckm_network))
```

```
## [1] 20
```

b.

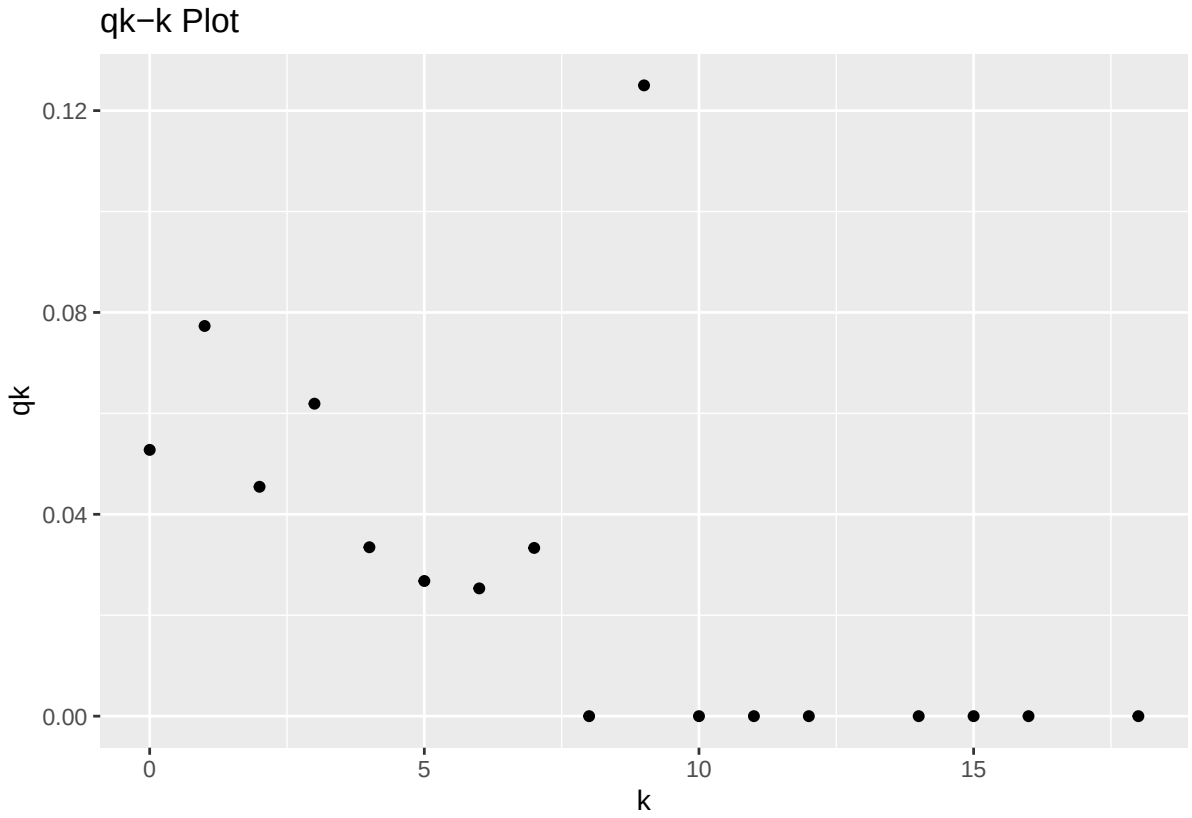
```
pk <- vector(length = 21)
for (i in 0:20){
  if (i %in% ckm_new$contrast_before){
    tempa <- ckm_new %>% filter(contrast_before==i) %>% nrow()
    tempb <- ckm_new %>% filter(contrast_before==i,
                                began_that_month==TRUE) %>% nrow()
    pk[i+1] <- tempb/tempa
  }else{
    pk[i+1] <- NA
```

```
  }
}
rm(tempa,tempb)
pk %>% data.frame() %>% mutate(k = 0:20) %>% filter(!is.na(.)) %>%
  ggplot() + geom_point(aes(x=k,y=.)) + labs(x="k",y="pk",title = "pk-k Plot")
```



c.

```
qk = vector(length = 21)
for (i in 0:20){
  if (i %in% ckm_new$contrast_before){
    tempa <- ckm_new %>% filter(contrast_that_monthnearlier==i) %>% nrow()
    tempb <- ckm_new %>% filter(contrast_that_monthnearlier==i,
                                began_that_month==TRUE) %>% nrow()
    qk[i+1] <- tempb/tempa
  }else{
    qk[i+1] <- NA
  }
}
rm(tempa,tempb)
qk %>% data.frame() %>% mutate(k = 0:20) %>% filter(!is.na(.)) %>%
  ggplot() + geom_point(aes(x=k,y=.)) + labs(x="k",y="qk",title = "qk-k Plot")
```



4. a.

```
est_a <- function(ab,k) ab[1] + ab[2] * k
lse_a <- function(ab){
  sum(na.omit((est_a(ab,c(0:20)) - pk)^2))
}
para_a <- nlm(lse_a,c(-1,1))
para_a <- para_a$estimate
para_a
```

```
## [1]  0.056944934 -0.003801307
```

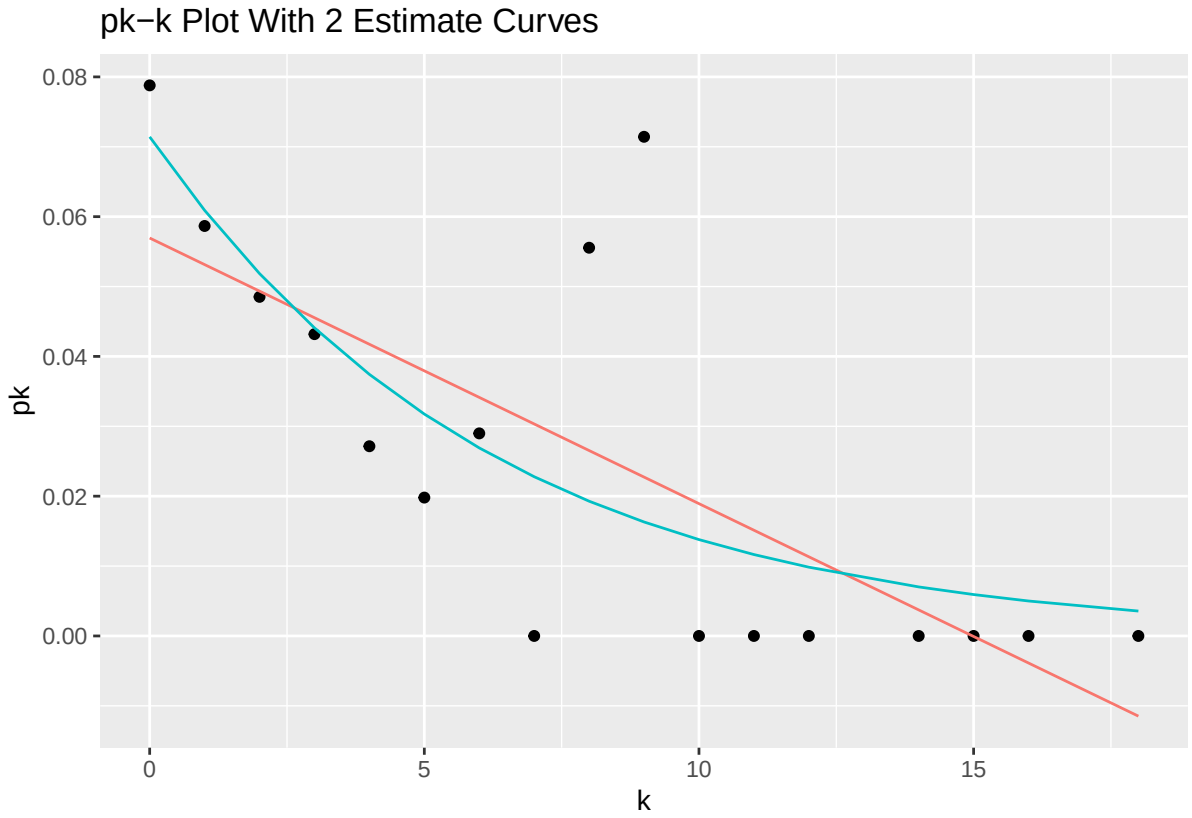
b.

```
est_b <- function(ab,k) exp(ab[1] + ab[2] * k)/(1+exp(ab[1] + ab[2] * k))
lse_b <- function(ab){
  sum(na.omit((est_b(ab,c(0:20)) - pk)^2))
}
para_b <- nlm(lse_b,c(1,1))
para_b <- para_b$estimate
para_b
```

```
## [1] -2.5650797 -0.1705084
```

c.

```
pk %>% data.frame() %>% mutate(k = 0:20) %>% filter(!is.na(.)) %>%
  ggplot() + geom_point(aes(x=k,y=.)) + labs(x="k",y="pk",title = "pk-k Plot With 2 Estimate Curves
  geom_line(aes(x=k,y=est_a(para_a,k)),color = "#F8766D") +
  geom_line(aes(x=k,y=est_b(para_b,k)),color="#00BFC4")
```


pk−k Plot With 2 Estimate Curves
0.08
0.06
0.04
0.02
0.00
pk
0
5
10
15
k