Stat 502 HW4

1

(a)

```
zinc <- readRDS("zinc.RDS")
summary(zinc)
```

```
##    ZINC      DIVERSITY
##  BACK:8   Min.   :0.630
##  HIGH:8   1st Qu.:1.393
##  LOW :8   Median :1.880
##  MED :8   Mean   :1.710
##           3rd Qu.:2.070
##           Max.   :2.830
```

```
n <- 5
m <- 4
N <- m*n
lambda <- n*(1^2 + 5^2 +3^2)/20^2
lambda
```

```
## [1] 0.4375
```

```
alpha <- 0.05
powerF <- 1 - pf(qf(1-alpha, m-1, N-m), m-1,N-m, ncp = lambda)
powerF
```

```
## [1] 0.07123778
```

The power is 0.07123778.

```
powerf <- c()
for (n in 1:200){
  m <- 4
  N < m*n
  lambda <- n*(1^2 + 5^2 +3^2)/20^2
  alpha <- 0.05
  powerf[n] <- 1 - pf(qf(1-alpha, m-1, N-m), m-1,N-m, ncp = lambda)
}
powerf
```

```
##   [1] 0.05410027 0.05827608 0.06252553 0.06684673 0.07123778 0.07569680
##   [7] 0.08022188 0.08481116 0.08946276 0.09417482 0.09894546 0.10377286
##  [13] 0.10865516 0.11359054 0.11857718 0.12361329 0.12869706 0.13382673
##  [19] 0.13900053 0.14421671 0.14947355 0.15476933 0.16010235 0.16547093
##  [25] 0.17087340 0.17630813 0.18177348 0.18726785 0.19278965 0.19833731
##  [31] 0.20390928 0.20950404 0.21512007 0.22075588 0.22641002 0.23208103
##  [37] 0.23776749 0.24346800 0.24918118 0.25490567 0.26064012 0.26638323
##  [43] 0.27213371 0.27789027 0.28365167 0.28941669 0.29518411 0.30095276
##  [49] 0.30672148 0.31248912 0.31825457 0.32401674 0.32977456 0.33552697
```

```
##  [55] 0.34127294 0.34701147 0.35274158 0.35846230 0.36417269 0.36987183
##  [61] 0.37555881 0.38123277 0.38689285 0.39253820 0.39816802 0.40378150
##  [67] 0.40937787 0.41495639 0.42051630 0.42605691 0.43157750 0.43707741
##  [73] 0.44255598 0.44801257 0.45344657 0.45885736 0.46424437 0.46960704
##  [79] 0.47494481 0.48025717 0.48554359 0.49080358 0.49603667 0.50124239
##  [85] 0.50642031 0.51156999 0.51669102 0.52178302 0.52684558 0.53187837
##  [91] 0.53688101 0.54185318 0.54679456 0.55170485 0.55658374 0.56143097
##  [97] 0.56624627 0.57102938 0.57578008 0.58049814 0.58518334 0.58983549
## [103] 0.59445440 0.59903990 0.60359183 0.60811002 0.61259435 0.61704468
## [109] 0.62146090 0.62584290 0.63019059 0.63450387 0.63878267 0.64302693
## [115] 0.64723658 0.65141159 0.65555190 0.65965751 0.66372837 0.66776449
## [121] 0.67176586 0.67573248 0.67966436 0.68356153 0.68742402 0.69125186
## [127] 0.69504509 0.69880376 0.70252793 0.70621767 0.70987304 0.71349411
## [133] 0.71708098 0.72063373 0.72415245 0.72763724 0.73108821 0.73450547
## [139] 0.73788913 0.74123932 0.74455615 0.74783977 0.75109029 0.75430787
## [145] 0.75749265 0.76064477 0.76376438 0.76685164 0.76990670 0.77292974
## [151] 0.77592091 0.77888038 0.78180832 0.78470492 0.78757035 0.79040478
## [157] 0.79320840 0.79598140 0.79872397 0.80143629 0.80411856 0.80677097
## [163] 0.80939372 0.81198701 0.81455104 0.81708601 0.81959212 0.82206959
## [169] 0.82451861 0.82693940 0.82933216 0.83169710 0.83403444 0.83634439
## [175] 0.83862716 0.84088296 0.84311202 0.84531454 0.84749075 0.84964085
## [181] 0.85176507 0.85386363 0.85593673 0.85798460 0.86000747 0.86200554
## [187] 0.86397903 0.86592816 0.86785316 0.86975424 0.87163162 0.87348551
## [193] 0.87531614 0.87712371 0.87890846 0.88067059 0.88241032 0.88412787
## [199] 0.88582346 0.88749729
```

```r
ind = match(1, powerf >= .8)
ind
```

```
## [1] 160
```

We need sample size of at least 160.

```r
powerf1 <- c()
for (sigma in 1:20){
  m <- 4
  n <- 10
  N < m*n
  lambda <- n*(1^2 + 5^2 +3^2)/sigma^2
  alpha <- 0.05
  powerf1[sigma] <- 1 - pf(qf(1-alpha, m-1, N-m), m-1,N-m, ncp = lambda)
}
powerf1
```

```
##  [1] 1.00000000 0.99999999 0.99831702 0.94793280 0.80143629 0.63067154
##  [7] 0.48887444 0.38406460 0.30892904 0.25490567 0.21544558 0.18604445
## [13] 0.16368815 0.14635755 0.13268293 0.12171906 0.11280150 0.10545445
## [19] 0.09933119 0.09417482
```

```r
ind = match(1, powerf1 <= .8) -1
ind
```

```
## [1] 5
```

Largest value σ can take is 5.

2

(a)

```
X <- matrix(c(1,1,1,1, 1,1,1,1, 1,1,0,0, 0,0,1,1, 0,0,1,1, 0,0,-1,-1,
              0,0,0,0, 1,1,-1,-1), byrow = F, ncol = 4)
X
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    1    0    0
## [2,]    1    1    0    0
## [3,]    1    0    1    0
## [4,]    1    0    1    0
## [5,]    1    0    0    1
## [6,]    1    0    0    1
## [7,]    1    1   -1   -1
## [8,]    1    1   -1   -1
```

We have obtained this matrix because of the constraint $\tau_1 - \tau_2 - \tau_3 = \tau_4$. The rows of the matrix are Y_{ij}, namely $Y_{11}, Y_{12}\dots$ And the columns are $\mu, \tau_1, \tau_2, \tau_3$.

(b)

```
XT <- t(X)
invxtx <- solve(XT %*% X)
H <- X %*% invxtx %*% XT
H
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,]  0.5  0.5  0.0  0.0  0.0  0.0  0.0  0.0
## [2,]  0.5  0.5  0.0  0.0  0.0  0.0  0.0  0.0
## [3,]  0.0  0.0  0.5  0.5  0.0  0.0  0.0  0.0
## [4,]  0.0  0.0  0.5  0.5  0.0  0.0  0.0  0.0
## [5,]  0.0  0.0  0.0  0.0  0.5  0.5  0.0  0.0
## [6,]  0.0  0.0  0.0  0.0  0.5  0.5  0.0  0.0
## [7,]  0.0  0.0  0.0  0.0  0.0  0.0  0.5  0.5
## [8,]  0.0  0.0  0.0  0.0  0.0  0.0  0.5  0.5
```

We have obtained the H matrix using $H = X(X^TX)^{-1}X^T$.

(c) Because of the constraints, we know $\tau_1 = \tau_2 = 2\tau_3 = -10\tau_4$. So we can write everything in terms of τ_3 as follows:

```
X_prime <- matrix(c(1,1,1,1, 1,1,1,1, 2,2,2,2, 1,1,-5,-5), byrow = F, ncol = 2)
X_prime
```

```
##      [,1] [,2]
## [1,]    1    2
## [2,]    1    2
## [3,]    1    2
## [4,]    1    2
## [5,]    1    1
## [6,]    1    1
## [7,]    1   -5
## [8,]    1   -5
```

3

(a)

```r
X_weight <- matrix(c(1,1,1, 0,0,0, 0,0,0, 0,0,0, 1,1,1, 0,0,0,
                     0,0,0, 0,0,0, 1,1,1), byrow = F, ncol = 3)
X_weight
```

```
##       [,1] [,2] [,3]
##  [1,]    1    0    0
##  [2,]    1    0    0
##  [3,]    1    0    0
##  [4,]    0    1    0
##  [5,]    0    1    0
##  [6,]    0    1    0
##  [7,]    0    0    1
##  [8,]    0    0    1
##  [9,]    0    0    1
```

(b)

```r
XXTinv <- solve(t(X_weight) %*% X_weight)
Y = matrix(c(78,86.8,103.8,83.7,89,99.2,83.8,81.5,86.2), ncol = 1)
beta <- XXTinv %*% t(X_weight) %*% Y
beta
```

```
##          [,1]
## [1,] 89.53333
## [2,] 90.63333
## [3,] 83.83333
```

Therefore, the estimate $\hat{\beta}_1$ for β_1 is the first entry in the above vector, i.e., 89.5333.

(c)

```r
mn_1 <- 3*2
sse <- sum((Y[1:3] - mean(c(78,86.8,103.8)))^2 + (Y[4:6] - mean(c(83.7,89,99.2)))^2
           +  (Y[7:9] - mean(c(83.8,81.5,86.2)))^2)/mn_1
se <- sqrt(sse) * sqrt(XXTinv[1,1])
se
```

```
## [1] 5.159673
```

The standard error of $\hat{\beta}_1$ is 5.159673.

(d) By theorem, we know that $\frac{\hat{\beta}_1-\beta_1}{SE(\hat{\beta}_1)}$ follows t_{n-p} distribution.

```r
n = 9
p = 3
t_c <- qt(0.95,6)
CI = c(beta[1,1] - t_c * se, beta[1,1] + t_c * se)
CI
```

```
## [1] 79.50716 99.55951
```

The 90% confidence interval for $\hat{\beta}_1$ is [79.50716, 99.55951].

(e) Yes they are independent. By our model, we know μ_i are fixed. Since ϵ_{ij} are independent, Y_{ij} is also independent because it is a linear function of ϵ_{ij}. For each $\hat{\mu_i}$, we know that $\hat{\mu_i} = X(X^TX)^{-1}Y$ and therefore is a linear function of the Y, which is independent. Therefore, $\hat{\mu}_{beef}$ and $\hat{\mu}_{pork}$ are independent.