STAT 5385: Lab 5

Willliam Ofosu Agyapong

03/03/2022

```
toluca <- read.table("../Data Sets/Chapter  1 Data Sets/CH01TA01.txt")
colnames(toluca) <- c("lot_size","Work_hours")
kable(head(toluca, 10))
```

lot_size	Work_hours
80	399
30	121
50	221
90	376
70	361
60	224
120	546
80	352
100	353
50	157

```
senic <- read.table("../Data Sets/Appendix C Data Sets/APPENC01.txt")
colnames(senic) <- c("ID","LOS","Age","Infec","Cul","Xray","beds","Med","region","avg","nurses","fands")
kable(head(senic,10))
```

ID	LOS	Age	Infec	Cul	Xray	beds	Med	region	avg	nurses	fands
1	7.13	55.7	4.1	9.0	39.6	279	2	4	207	241	60
2	8.82	58.2	1.6	3.8	51.7	80	2	2	51	52	40
3	8.34	56.9	2.7	8.1	74.0	107	2	3	82	54	20
4	8.95	53.7	5.6	18.9	122.8	147	2	4	53	148	40
5	11.20	56.5	5.7	34.5	88.9	180	2	1	134	151	40
6	9.76	50.9	5.1	21.9	97.0	150	2	2	147	106	40
7	9.68	57.8	4.6	16.7	79.0	186	2	3	151	129	40
8	11.18	45.7	5.4	60.5	85.8	640	1	2	399	360	60
9	8.67	48.2	4.3	24.4	90.8	182	2	3	130	118	40
10	8.84	56.3	6.3	29.6	82.6	85	2	1	59	66	40

```
toluca %>% dfSummary() %>% view()
```

0.1 Regression matrix calculations

```
j <- rep(1,nrow(toluca)) # create a vector of ones for the intercept
X <- cbind(j,toluca$lot_size) # Create the design matrix
y <- toluca$Work_hours;y # Extract response variable
```

```
##  [1] 399 121 221 376 361 224 546 352 353 157 160 252 389 113 435 420 212 268 377
```

```
## [20] 421 273 468 244 342 323
```

```
kable(t(X))# making sure everything went right
```

j	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	80	30	50	90	70	60	120	80	100	50	40	70	90	20	110	100	30	50	90	110	30	90	40	80	

```
# matrix computations
(xpx <- t(X) %*%X)
```

```
##      j
## j   25   1750
##   1750 142300
```

```
(xpy <- t(X) %*%y)
```

```
##     [,1]
## j   7807
##   617180
```

```
(ypy <- t(y) %*%y)
```

```
##         [,1]
## [1,] 2745173
```

```
xpx^-1 # is this the correct inverse? This computes elementwise inverse; probably not what we want.
```

```
##           j
## j 0.0400000 5.714e-04
##   0.0005714 7.027e-06
```

```
solve(xpx) # computes the inverse of a matrix
```

```
##           j
## j  0.287475 -3.535e-03
##   -0.003535  5.051e-05
```

```
# Using the SENIC data
names(senic) # just to remind ourselves of the variable names
```

```
##  [1] "ID"     "LOS"    "Age"    "Infec"  "Cul"    "Xray"   "beds"   "Med"
##  [9] "region" "avg"    "nurses" "fands"
```

```
j <- rep(1, nrow(senic))
X2 <- as.matrix(cbind(j, senic[, c(4,6,12)]))
kable(head(X2, 10)) # looking at few rows to make sure everything went right
```

j	Infec	Xray	fands
1	4.1	39.6	60
1	1.6	51.7	40
1	2.7	74.0	20
1	5.6	122.8	40
1	5.7	88.9	40
1	5.1	97.0	40
1	4.6	79.0	40
1	5.4	85.8	60
1	4.3	90.8	40
1	6.3	82.6	40

```r
y2 <- senic$LOS

(xpx2 <- t(X2) %*%X2)
```

```
##            j   Infec   Xray  fands
## j      113.0   492.1   9224   4877
## Infec  492.1  2344.4  41488  22181
## Xray  9224.0 41487.8 794935 401791
## fands 4877.0 22180.6 401791 236367
```

```r
(xpy2 <- t(X2) %*%y2)
```

```
##        [,1]
## j      1090
## Infec  4901
## Xray  90582
## fands 48212
```

```r
(ypy2 <- t(y2) %*%y2)
```

```
##       [,1]
## [1,] 10928
```

```r
solve(xpx2)
```

```
##                j      Infec       Xray      fands
## j      0.221158 -0.0047174 -1.685e-03 -1.256e-03
## Infec -0.004717  0.0075030 -2.143e-04 -2.425e-04
## Xray  -0.001685 -0.0002143  3.023e-05  3.488e-06
## fands -0.001256 -0.0002425  3.488e-06  4.697e-05
```