HW 2 - Predictive Modeling in Finance and Insurance

Dennis Goldenberg

2024-02-01

1. Nursing Home Utilization

```
# import packages
library(ggplot2)
library(magrittr)
```

```
# read in data
# WNH <- read.csv(file)
WNH <- read.csv('WiscNursingHome.csv', header = TRUE)
```

1a) Estimation of Coefficients

```
#Generate variables to analyze
WNH$LOGTPY <- log(WNH$TPY)
WNH$LOGNUMBED <- log(WNH$NUMBED)
```

Using the generated variables, I calculate $x^T x$, adding in a column for the intercept:

```
x <- cbind(1,WNH$LOGNUMBED)
xTx <- t(x) %*% x
xTx
```

```
##          [,1]      [,2]
## [1,]  717.000  3195.999
## [2,] 3195.999 14419.186
```

Then, I find $\left(x^T x\right)^{-1}$:

```
xTxInv <- xTx^{-1}
xTxInv
```

```
##               [,1]         [,2]
## [1,] 0.0013947001 3.128912e-04
## [2,] 0.0003128912 6.935204e-05
```

Finally, I find $x^T y$:

```
y <- WNH$LOGTPY
xTy <- t(x) %*% y
xTy
```

```
##           [,1]
## [1,]  3129.203
## [2,] 14124.171
```

Using the formula for linear regression that $\beta = (x^T x)^{-1} x^T y$:

```r
beta <- xTxInv %*% xTy
beta
```

```
##           [,1]
## [1,] 8.783629
## [2,] 1.958640
```

1b. The prediction Matrix