Linear Algebra Primer

Matrices / Vectors

A matrix is an $n \times p$ object. Matrices are often denoted by a capital letter (or Greek symbol). A few common matrices will be

Vectors are essentially one-dimension vectors and will be denoted with an underline. We will assume vectors are $q \times 1$ dimension unless noted with a transpose.

The transpose operator will be denoted by $\underline{y}^T = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}$ or $\underline{y}^{'}$, both of which would result in a $1 \times n$ vector.

Matrix Multiplication

The most important component in matrix multiplication is tracking dimensions.

Consider a simple case with

$$\underline{\hat{y}} = X \times \underline{\hat{\beta}},$$

In R, we use `%*%` for matrix multiplication.

```
X <- matrix(c(1,2, 1 ,-1), nrow = 2, ncol = 2, byrow = T); X
```

```
##      [,1] [,2]
## [1,]    1    2
## [2,]    1   -1
```

```
#X <- matrix(c(1, 1, 2 ,-1), nrow = 2, ncol = 2); X
beta_hat <- matrix(c(3,2),nrow =2, ncol = 1); beta_hat
```

```
##      [,1]
## [1,]    3
## [2,]    2
```

```
y_hat <- X %*% beta_hat; y_hat
```

```
##      [,1]
## [1,]    7
## [2,]    1
```

Motivating Dataset: Washington (DC) housing dataset

Hopefully the connections to statistics are clear, using X and β, but let's consider a motivating dataset.

This dataset contains housing information from Washington, D.C. It was used for a STAT532 exam, so apologize in advance for any scar tissue.

```
DC <- read_csv('https://math.montana.edu/ahoegh/teaching/stat532/data/DC.csv')
```

```
## Parsed with column specification:
## cols(
##   BATHRM = col_double(),
##   HF_BATHRM = col_double(),
##   AC = col_character(),
##   BEDRM = col_double(),
##   STORIES = col_double(),
##   PRICE = col_double(),
##   CNDTN = col_character(),
##   LANDAREA = col_double(),
##   FULLADDRESS = col_character(),
##   ASSESSMENT_NBHD = col_character(),
##   WARD = col_character(),
##   QUADRANT = col_character()
## )
```

```
DC %>% group_by(WARD) %>%
  summarize(`Average Price (millions of dollars)` = mean(PRICE)/1000000, .groups = 'drop') %>%
  kable(digits = 3)
```

WARD	Average Price (millions of dollars)	.groups
Ward 1	0.879	drop
Ward 2	1.919	drop
Ward 3	1.294	drop
Ward 4	0.693	drop
Ward 5	0.592	drop
Ward 6	0.856	drop
Ward 7	0.321	drop
Ward 8	0.306	drop

```
DC %>% group_by(BEDRM) %>%
  summarize(`Average Price (millions of dollars)` = mean(PRICE)/1000000, .groups = 'drop') %>%
  kable(digits = 3)
```

BEDRM	Average Price (millions of dollars)	.groups
0	0.195	drop
1	0.442	drop
2	0.479	drop
3	0.620	drop
4	0.832	drop
5	1.355	drop
6	1.849	drop
7	1.666	drop
9	7.365	drop

Regression Model

There are many factors in this dataset that can are useful to predict housing prices.

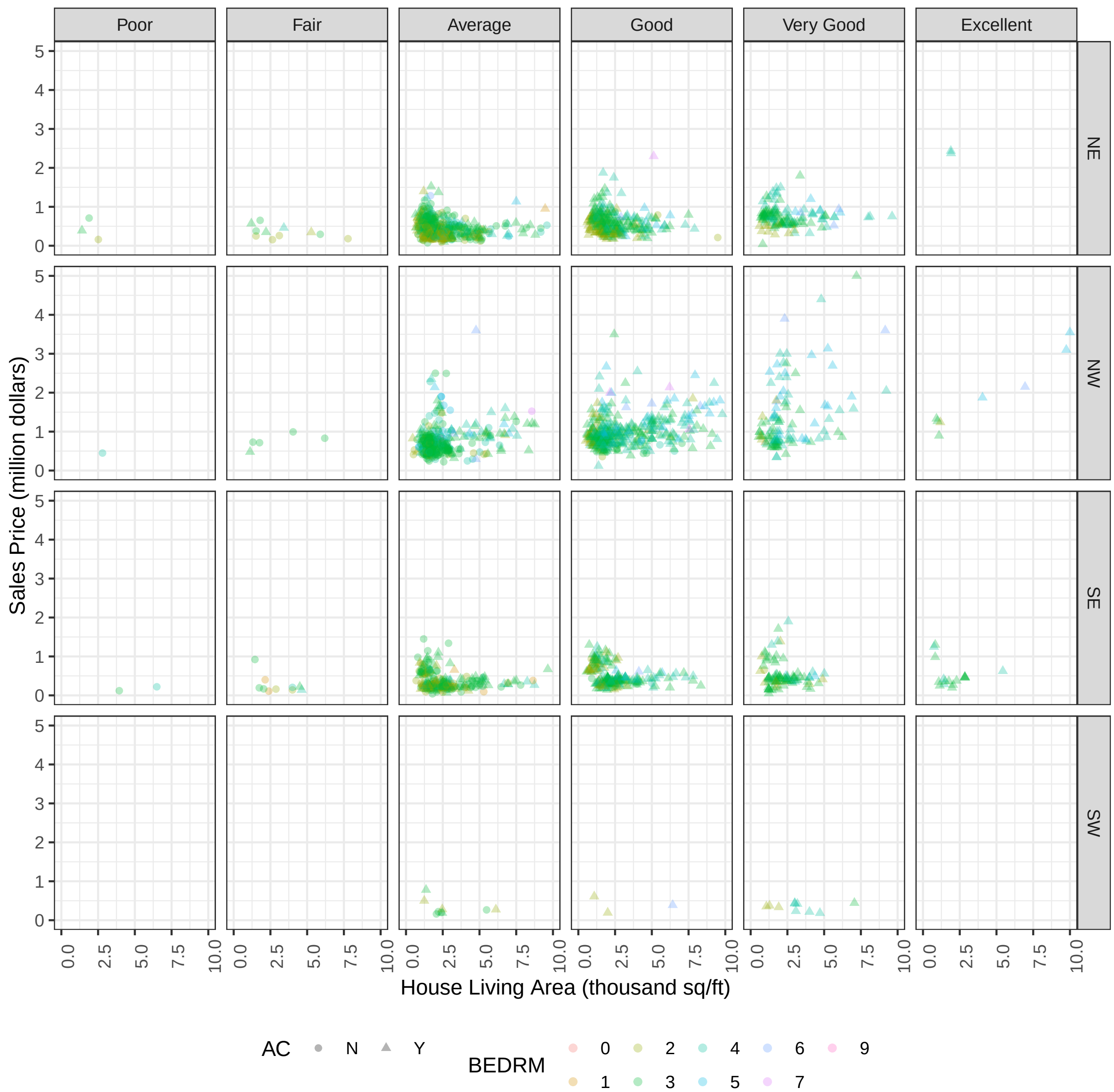


Figure 1: Washington DC Housing prices. Note the exploratory figure removes (~40) properties costing more than 5 million dollars or larger than 10,000 square feet

Diagonal Matrices

The matrix we previously specified, is referred to as a diagonal matrix.

Correlation Matrices

It turns out that I is the special case of what is referred to as a correlation matrix.

A correlation matrix is:

Similarly Σ is often referred to as a variance - covariance matrix (or just a covariance matrix). A covariance matrix:

Multivariate Normal Distribution

Formally, our matrix notation has used a multivariate normal distribution.

$$\underline{y} = X\underline{\beta} + \underline{\epsilon}, \tag{1}$$

where $\underline{\epsilon} \sim N(\underline{0}, \Sigma)$, which also implies $\underline{y} \sim N(X\underline{\beta}, \Sigma)$.

Partitioned Matrices

Conditional Multivariate Normal

Here is where the magic happens with correlated data. Let $\underline{y_1}|\underline{Y_2} = \underline{y_2}$ be a conditional distribution for $\underline{y_1}$ given that $\underline{y_2}$ is known. Then

$$\underline{y_1}|\underline{y_2} \sim N\left(X_1\beta + \Sigma_{12}\Sigma_{22}^{-1}\left(\underline{y_2} - X_2\beta\right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

Now let's consider a few special cases (in the context of the DC housing dataset.)

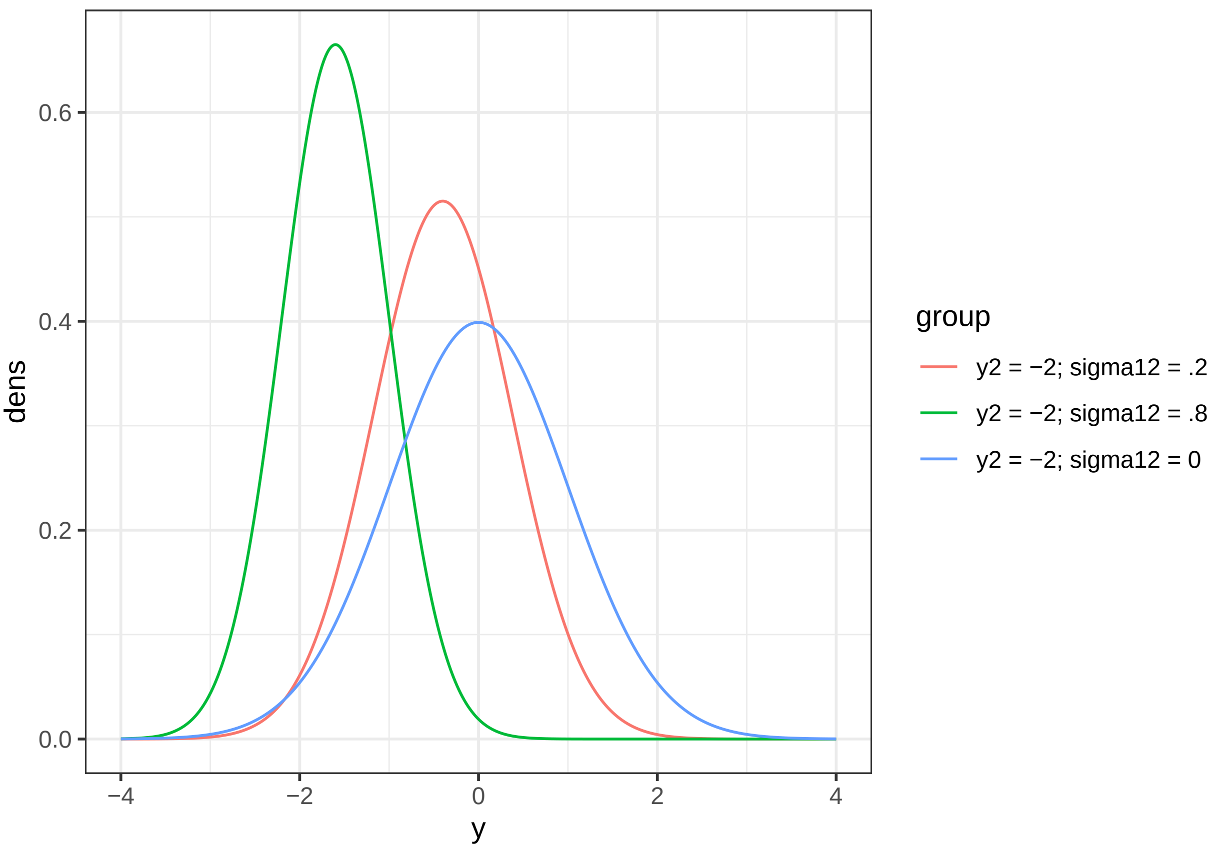
1. Let $\Sigma = \sigma^2 I$,

2. Otherwise, let $\Sigma = \sigma^2 H$ and we'll assume Σ_{12} has some non-zero elements.

First a quick interlude about matrix inversion. The inverse of a symmetric matrix is defined such that $E \times E^{-1} = I$. We can calculate the inverse of a matrix for a 1×1 matrix, perhaps as 2×2, matrix and maybe even a 3×3 matrix. However, beyond that it is quite challenging and time consuming. Furthermore, it is also (relatively) time intensive for your computer.

3. Let $n_1 = 1$ and $n_2 = 1$, then

Now consider an illustration for a couple simple scenarios. Let $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 1$. Now assume $y_2 = -2$ and we compare the conditional distribution for a few values of σ_{12}.



One last note, the marginal distributions for any partition $\underline{y_1}$ are quite simple.

$$\underline{y_1} \sim N\left(X_1\beta, \Sigma_{11}\right)$$

or just

$$y_1 \sim N\left(X_1\beta, \sigma_1^2\right)$$

if y_1 is scalar.