Recap: (change the office hours)

\[ X^T X \beta = X^T y \] (N.E.s)

When \( X^T X \) is invertible we know \( \beta \) the solution to the N.E.s is
\[ \hat{\beta} = (X^T X)^{-1} X^T y \]

\( X^T X \) is not invertible, there are infinitely many solutions. We want to characterize them.

**Vector Space:** \( \mathbb{C} \subseteq \mathbb{R}^n \) is a set of vectors such that if \( x, y \in S \), then \( \alpha x + \beta y \in S \) and \( 0 \in S \).

**Subspace:** It is a subset of vector space and it is also a vector space.

\( \mathbb{R}^3 = \{ (x, y, z) \mid x, y, z \in \mathbb{R} \} \) is a vector space.

\( \mathcal{N} = \{ (x, y, 0) \mid x, y \in \mathbb{R} \} \) in a subspace of \( \mathbb{R}^3 \).

Any vector space is a subspace of itself, i.e., \( \mathbb{R}^3 \) in a subspace of \( \mathbb{R}^3 \).

If we take \( x_1, \ldots, x_k \in S \), define

\[ M = \{ y \mid y = \sum c_i x_i, c_i \text{ coefficients} \} \]

\( M \) is called the space spanned by \( x_1, \ldots, x_k \).
The column space of a matrix $A$, denoted by $C(A)$, in the vector space spanned by the columns of the matrix $A$.

\[
A = \begin{bmatrix}
    a_{11} & \cdots & a_{1n} \\
    \vdots & \ddots & \vdots \\
    a_{m1} & \cdots & a_{mn}
\end{bmatrix}
\]

\[
C(A) = \left\{ \mathbf{c} \mid \mathbf{c} = \sum_{j=1}^{n} x_j \begin{bmatrix}
    a_{1j} \\
    \vdots \\
    a_{mj}
\end{bmatrix} \right\}, \text{ for coefficients } x_1, \ldots, x_n
\]

$C(A)$ consists of all possible linear combinations of the columns of $A$.

\[
\sum_{j=1}^{n} x_j \begin{bmatrix}
    a_{1j} \\
    \vdots \\
    a_{mj}
\end{bmatrix} = \begin{bmatrix}
    a_{11} x_1 + \cdots + a_{1n} x_n \\
    \vdots \\
    a_{m1} x_1 + \cdots + a_{mn} x_n
\end{bmatrix}
\]

\[
= A \mathbf{x}
\]

\[
\mathbf{x} = \begin{pmatrix}
    x_1 \\
    \vdots \\
    x_n
\end{pmatrix}
\]

$C(A) = \left\{ \mathbf{c} \mid A \mathbf{x} = \mathbf{c} \text{, for some } \mathbf{x} \right\}$

\[
3x_1: \quad A = \begin{bmatrix}
    1 & 0 \\
    1 & 2 \\
    0 & 0
\end{bmatrix} \quad C(A) = \left\{ \begin{pmatrix}
    a \\
    b
\end{pmatrix} \mid a, b \in \mathbb{R} \right\}
\]

\[
\begin{pmatrix}
    a \\
    b
\end{pmatrix} = \begin{bmatrix}
    1 & 0 \\
    1 & 2
\end{bmatrix} x_1 + \begin{bmatrix}
    0 & 1
\end{bmatrix} x_2 \quad (\text{there should be some } x_1, x_2 \text{ for which this holds})
\]
\[ a = x_1 \quad \cdots \quad 0 \]
\[ b = x_1 + 2x_2 \quad \cdots \quad 2 \]
\[ \Rightarrow \quad x_1 = a, \quad x_2 = \frac{b-a}{2} \]

\[ \exists \ x_1, x_2 \text{ (given as above), such that} \]

\[ \begin{pmatrix} a \\ b \\ c \end{pmatrix} \text{ can be written as a linear combination of columns of } A. \]

\[ \text{Out: } \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} \in \text{C}(A). \]

Note the fact that \( X^p \in \text{C}(X) \)

\[
X^p = \begin{bmatrix}
\vdots & \vdots & \vdots \\
\tilde{x}_1 & \ldots & \tilde{x}_p \\
\vdots & \vdots & \vdots
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\vdots \\
\beta_p
\end{bmatrix}
= \begin{bmatrix}
\tilde{x}_1 \\
\vdots \\
\tilde{x}_p
\end{bmatrix} \beta_1 + \ldots + \begin{bmatrix}
\tilde{x}_1 \\
\vdots \\
\tilde{x}_p
\end{bmatrix} \beta_p
\]

When does an equation of the form \( A\vec{x} = \vec{c} \)

have a solution?

\( \vec{c} \in \text{C}(A) \)

If a system of equations has solution then that system is called consistent.

Recall the Normal equations

\[ X^T X \beta = X^T y. \]

Is it a consistent system of equations?

\( X^T y \in \text{C}(X^T) \) and \( X^T X \beta \in \text{C}(X^T X) = \text{C}(X^T) \)

It can be shown that \( \text{C}(X^T) = \text{C}(X^T X) \)

\( (3) \)
Linear dependence:

Let \( x_1, \ldots, x_n \) be vectors in \( S \). If there exists scalars \( \alpha_1, \ldots, \alpha_n \) not all zero so that

\[
\sum_{i=1}^{n} \alpha_i x_i = 0
\]

then \( x_1, \ldots, x_n \) are known as linearly dependent.

If \( \sum_{i=1}^{n} \alpha_i x_i = 0 \Rightarrow \alpha_1 = \cdots = \alpha_n = 0 \), then \( x_1, \ldots, x_n \) are called linearly independent.

**Ex:** Look at vectors \( \left( \begin{array}{c} 1 \\ 0 \\ 0 \end{array} \right), \left( \begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right), \left( \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right) \).

\[
\alpha_1 \left( \begin{array}{c} 1 \\ 0 \\ 0 \end{array} \right) + \alpha_2 \left( \begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right) + \alpha_3 \left( \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right) = \left( \begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right)
\]

\[\Rightarrow \alpha_1 = 0, \alpha_2 = 0, \alpha_3 = 0\]

So, these three vectors are linearly independent.

**Ex:** \( \left( \begin{array}{c} 1 \\ 0 \\ 0 \end{array} \right), \left( \begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right), \left( \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right) \). Are they linearly independent?

\[
\alpha_1 \left( \begin{array}{c} 1 \\ 0 \\ 0 \end{array} \right) + \alpha_2 \left( \begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right) + \alpha_3 \left( \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right) = \left( \begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right)
\]

\[\Rightarrow \alpha_1 = 0, \alpha_2 = 0, \alpha_3 - \alpha_2 = 0 \Rightarrow \alpha_3 = 0\]

**Ex:** \( x_1 = \left( \begin{array}{c} 1 \\ -1 \end{array} \right), \ x_2 = \left( \begin{array}{c} 1 \\ 2 \end{array} \right), \ x_3 = \left( \begin{array}{c} 2 \\ 1 \end{array} \right) \).

\[
\alpha_1 \left( \begin{array}{c} 1 \\ -1 \end{array} \right) + \alpha_2 \left( \begin{array}{c} 1 \\ 2 \end{array} \right) + \alpha_3 \left( \begin{array}{c} 2 \\ 1 \end{array} \right) = \left( \begin{array}{c} 0 \\ 0 \end{array} \right)
\]

\(\)
\[\begin{align*}
&= a_1 + a_2 + 2a_3 = 0 \quad \cdots \quad (1) \\
&= -a_1 + 2a_2 + a_3 = 0 \quad \cdots \quad (2)
\end{align*}\]

Consider \(a_1 = 1\), \(a_2 = 1\), \(a_3 = -1\). This is a solution to (1) & (2). Thus not all \(a_i\)'s are zero.

\[a_1 + a_2 = a_3 \implies a_1 + a_2 - a_3 = 0\]

Thus \(a_1, a_2, a_3\) are not linearly independent.

**Basis:** If \(M\) is a subspace of \(S\) and if \(\{x_1, \ldots, x_k\}\) is a linearly independent set of vectors which span \(M\) then \(\{x_1, \ldots, x_k\}\) is called a basis for \(M\).

**Ex:** \(\mathbb{R}^3\), \(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\), \(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\), \(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\) are linearly independent and, the space spanned by \(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\), \(\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\), \(\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\) in \(\mathbb{R}^3\). Take any \(\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3\). Clearly

\[
\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}
\]

\(\implies \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\), \(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\), \(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\) is a basis for \(\mathbb{R}^3\).

You can also check, \(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\), \(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\), \(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\) in another basis of \(\mathbb{R}^3\).
Def: For any \( m \times n \) matrix \( A \), \text{rank} of the matrix \( A \), denoted by \( \text{rank}(A) \), is the number of linearly independent rows or columns of \( A \).

In a \( m \times n \) matrix \( A \) if \( \text{rank}(A) = n \), then we say that \( A \) has \( \text{the full row rank} \).

\( \bigcirc \) If \( \text{rank}(A) = n \) then we call that \( A \) has the full column rank.

We will see that in a linear regression we care about a design matrix \( X \) having the full column rank. We will see that if \( X \) has the full column rank, then the solution to the Normal Equations is unique.

Def: Let \( A \) be an \( m \times n \) matrix.

\( A \) is nonsingular if there exists a matrix \( A^{-1} \) such that \( A A^{-1} = A^{-1} A = I \).

If no such matrix exists then we call \( A \) a singular matrix.

\( A^{-1} \) is called the inverse of \( A \).

If \( \text{rank}(A) = n \) then \( A \) is nonsingular. If \( A \) is singular if \( \text{rank}(A) < n \).

Ex: \( A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 5 & 5 \\ -2 & 1 & 3 \end{pmatrix} \) the third column is the second column - first column.

the \( \text{rank}(A) < 3 \).
\[ A = \begin{pmatrix} 1 & -2 \\ 2 & 3 \\ 4 & 5 \end{pmatrix} \Rightarrow \kappa(A) = 2 \Rightarrow A \text{ is invertible.} \]

Note the fact that \( \Box \)

Recall the linear regression model with \( p \) predictors and \( n \) observations:

\[ y = X \beta + \epsilon \]

Here \( X \) is an \( n \times p \) matrix and \( y \) is an \( n \times 1 \) vector.

\( X^T X \) is a \( p \times p \) matrix.

\( \kappa(X^T X) = \) number of linearly independent columns of \( X^T X \)

\( \kappa(X^T X) = \) number of linearly independent columns of \( X \)

(as \( \kappa(X^T X) = \kappa(X^T) \) (we discussed earlier).

\( \kappa(X^T X) = \kappa(X) \Rightarrow \kappa(X) = p \Rightarrow \kappa(X^T X) \leq p \Rightarrow X^T X \text{ is invertible.} \)

\( \hat{\beta} = (X^T X)^{-1} X^T y \)

If all columns of \( X \) are not linearly independent, then \( \kappa(X^T X) < p \) and \( X^T X \) is not invertible.

\( \square \)
Example: $y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \ j = 1, \ldots, O3, \ i = 0, 1, \ldots, 3$

\[
\begin{align*}
  y_{11} &= \mu + \alpha_1 + \epsilon_{11} \\
  y_{12} &= \mu + \alpha_1 + \epsilon_{12} \\
  y_{13} &= \mu + \alpha_1 + \epsilon_{13} \\
  y_{14} &= \mu + \alpha_2 + \epsilon_{21} \\
  y_{22} &= \mu + \alpha_2 + \epsilon_{22} \\
  y_{23} &= \mu + \alpha_2 + \epsilon_{23} \\
  y_{31} &= \mu + \alpha_3 + \epsilon_{31} \\
  y_{32} &= \mu + \alpha_3 + \epsilon_{32} \\
  y_{33} &= \mu + \alpha_3 + \epsilon_{33}
\end{align*}
\]

\[\Rightarrow \quad y = \begin{bmatrix}
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0
\end{bmatrix} \begin{bmatrix}
  \mu \\
  \alpha_1 \\
  \alpha_2 \\
  \alpha_3
\end{bmatrix} + \epsilon
\]

The first column is the sum of 2nd, 3rd, and 4th columns. Thus $X$ does not have the full column rank.

However, in order to estimate parameters, we often put a restriction in one way ANOVA.

$R$ puts $\alpha_1 + \alpha_2 + \alpha_3 = 0$  \[\alpha_3 = -\alpha_1 - \alpha_2\]

$X$ becomes

\[\begin{bmatrix}
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0
\end{bmatrix}
\]
\[ y = X^\top \beta + \varepsilon \]

\[ \hat{\beta} = \left( \begin{array}{c} \mu \\ \alpha_1 \\ \alpha_2 \end{array} \right) \quad \alpha_3 = -\alpha_1 - \alpha_2 \]

\[ X \mathbb{I} = \left[ \begin{array}{cccc} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & -1 \\ 1 & 0 & 1 & -1 \\ 1 & 0 & 1 & -1 \end{array} \right] \]

The standard linear regression equation is if \( X \) has full column rank, \( X^T X \) is invertible.

\[ \hat{\beta} = (X^T X)^{-1} X^T y \]
height <- c(169.6, 166.8, 157.1, 181.1, 158.4, 165.6, 166.7, 155.6, 168.1, 165.3)
weight <- c(71.2, 58.2, 56.6, 64.5, 53.4, 56.8, 49.2, 55.6, 77.8)

###command

lm(weight~height)

###output

\[
y = \text{weight} \\
x = \text{height} \\
y = \beta_0 + \beta_1 x + e.
\]

Call:

\[\text{lm(formula = weight} \sim \text{height)}\]

Coefficients: parameter estimates

<table>
<thead>
<tr>
<th></th>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-36.6097</td>
</tr>
<tr>
<td>weight</td>
<td>0.5808</td>
</tr>
</tbody>
</table>

###command

model.1 <- lm(weight~height)
anova(model.1)

###output

Analysis of Variance Table

Response: weight

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>height</td>
<td>1</td>
<td>164.96</td>
<td>164.961</td>
<td>2.3275</td>
<td>0.1656</td>
</tr>
</tbody>
</table>
Residuals 8.567.00 70.875

```r
# command
summary(model.1)
```

```r
# output

Call:
lm(formula = weight ~ height)

Residuals:
 Min  1Q Median  3Q  Max
-7.1687 -4.4384 -2.8973  0.5096 18.4055

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|---------|
| (Intercept) | -36.6097 | 63.0342 | -0.581 0.577 |
| height     | 0.5808    | 0.3807 | 1.526 0.166 |

Residual standard error: 8.419 on 8 degrees of freedom
Multiple R-squared: 0.2254, Adjusted R-squared: 0.1285
F-statistic: 2.327 on 1 and 8 DF, p-value: 0.1656

```r
# command
new = data.frame(height=c(170,180))
predict.lm(model.1,new,new=TRUE,interval="confidence",level=0.95)
```
## output

$f_{\text{fit}}$

<table>
<thead>
<tr>
<th>fit</th>
<th>lwr</th>
<th>upr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62.12420</td>
<td>54.79042</td>
</tr>
<tr>
<td>2</td>
<td>67.93208</td>
<td>53.74439</td>
</tr>
</tbody>
</table>

$\text{se.fit}$

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3.180296</td>
<td>6.152498</td>
</tr>
</tbody>
</table>

$\text{df}$

[1] 8

$\text{residual.scale}$

[1] 8.418732

### command

```r
plot(model.1$f_{\text{fit}}, weight-model.1$f_{\text{fit}}, xlab="fitted", ylab="residual",
     main="residual vs. fitted")
```

### command

```r
race <- c(1,0,0,1,0,0,0,0,0,0,0)
height <- c(169.6,166.8,157.1,181.1,158.4,165.6,166.7,155.6,168.1,165.3)
weight <- c(71.2,58.2,56,64.5,53,52.4,56.8,49.2,55.6,77.8)
```
model.2 <- lm(weight~height+race)
summary(model.2)

Call:
lm(formula = weight ~ height + race)

Residuals:
  Min  1Q Median   3Q  Max
-5.870 -4.712  -2.350  0.249  19.688

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.2790    91.2199    0.069    0.947
height      0.3136     0.5595     0.560    0.593
race        6.5867     9.7814     0.673    0.522

Residual standard error: 8.722 on 7 degrees of freedom
Multiple R-squared: 0.2725,  Adjusted R-squared: 0.06464
F-statistic: 1.311 on 2 and 7 DF,  p-value: 0.3284