Normal Equations and the LS Estimator

To find the minimizer of $Q(\mathbf{b})$, differentiate $Q(\mathbf{b})$ wrt $\mathbf{b}$:

$$\frac{\partial Q(\mathbf{b})}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\mathbf{b},$$

set the derivative to $\mathbf{0}$ and solve.

Notice Xc: any linear combination of columns of X.

Notice the solution must satisfy $\mathbf{X}' \cdot (\mathbf{Y} - \mathbf{X}\mathbf{b}) = \mathbf{0}$.

We want that for b^*, vector of resid $(Y - Xb^*)$ is orthogonal to X; i.e.,

The least squares *normal equations* are $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$.

To solve, premultiply both sides by $(\mathbf{X}'\mathbf{X})^{-1}$ (assume this exists):

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

$c^T X^T \cdot (Y - X b^*) = 0$

Since $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$ and $\mathbf{I}\mathbf{b} = \mathbf{b}$, we then find the solution

$$\underset{m\times 1}{\mathbf{b}^*} = \underset{m\times m}{(\mathbf{X}'\mathbf{X})^{-1}}\,\underset{m\times 1}{\mathbf{X}'\mathbf{Y}},$$

the LS estimator that minimizes $Q(\mathbf{b})$.

ISLR 3.6.3

```
> lm.fit=lm(medv~lstat+age,data=Boston)
> summary(lm.fit)

Call:
lm(formula = medv ~ lstat + age, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max
-15.98   -3.98   -1.28    1.97   23.16

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   33.2228     0.7308   45.46   <2e-16 ***
lstat         -1.0321     0.0482  -21.42   <2e-16 ***
age            0.0345     0.0122    2.83   0.0049 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 6.17 on 503 degrees of freedom
Multiple R-squared: 0.551,     Adjusted R-squared: 0.549
F-statistic:  309 on 2 and 503 DF,  p-value: <2e-16
```

$= \hat\beta_0, \hat\beta_1, \hat\beta_2$

lstat X_1

age X_2

est. SE of $\hat\beta_1$

value of TS for $H_0: \beta_1 = 0$

If H_0 is true, then $TS = \frac{\hat\beta_1 - 0}{\widehat{SE}(\hat\beta_1)}$ follows student's t with $df = 3$.

2-sided p-value $H_A: \beta_1 \neq 0$.

estimated value for σ. (I.e., $(RSE)^2 = SSE/(n-p)$).

(unadjusted) coeff. of determination R^2.

F-statistic: for $H_0: \beta_1 = 0$ and $\beta_2 = 0$.
H_A: at least one of these coef's $\neq 0$.

Recall credit card data balances example:

1 categorical covariate, ethnicity, has 3 levels,

3 levels: African American, Asian, Caucasian

Model 1: The mean is the same for every group, $E(Y_i) = \mu \ \forall i$.

Model 2: model with subgroup means:

$E(Y_i) = \mu_j$ if the ith client belongs to group j.

Representation 1:

$y_i = \mu_1 + \varepsilon_i$ if ith person is AA

$y_i = \mu_2 + \varepsilon_i$ if —"— is Asian

$y_i = \mu_3 + \varepsilon_i$ if —"— is Caucasian

Representation 2:

$y_i = \beta_0 + \varepsilon_i$.

$y_i = \beta_0 + \beta_1 + \varepsilon_i$

$y_i = \beta_0 + \beta_2 + \varepsilon_i$.

example: suppose we have 2 clients (consecutive) from each group.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} \overset{\text{equivalent representation}}{=} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_6 \end{bmatrix}$$