Lecture 4

September 4, 2018

Multiple Regression

- There is a response variable y and p explanatory variables $x_1, \ldots, x_p$. The goal is understand the relationship between y and $x_1, \ldots, x_p$.

Multiple Regression

- There is a response variable y and p explanatory variables $x_1, \ldots, x_p$. The goal is understand the relationship between y and $x_1, \ldots, x_p$.
- There are n subjects and data is collected on the variables from these subjects.

Multiple Regression

- There is a response variable y and p explanatory variables $x_1, \ldots, x_p$. The goal is understand the relationship between y and $x_1, \ldots, x_p$.
- There are n subjects and data is collected on the variables from these subjects.
- Data on the response variable is $y_1, \ldots, y_n$.

Multiple Regression

- There is a response variable y and p explanatory variables $x_1, \ldots, x_p$. The goal is understand the relationship between y and $x_1, \ldots, x_p$.
- There are n subjects and data is collected on the variables from these subjects.
- Data on the response variable is $y_1, \ldots, y_n$.
- Data on the jth explanatory variable x_j is $x_{1j}, x_{2j}, \ldots, x_{nj}$.

Multiple Regression

- There is a response variable y and p explanatory variables $x_1, \ldots, x_p$. The goal is understand the relationship between y and $x_1, \ldots, x_p$.
- There are n subjects and data is collected on the variables from these subjects.
- Data on the response variable is $y_1, \ldots, y_n$.
- Data on the jth explanatory variable x_j is $x_{1j}, x_{2j}, \ldots, x_{nj}$.
- The data for the ith subject is $(y_i, x_{i1}, \ldots, x_{ip})$.

Multiple Regression

- There is a response variable y and p explanatory variables $x_1, \ldots, x_p$. The goal is understand the relationship between y and $x_1, \ldots, x_p$.
- There are n subjects and data is collected on the variables from these subjects.
- Data on the response variable is $y_1, \ldots, y_n$.
- Data on the jth explanatory variable x_j is $x_{1j}, x_{2j}, \ldots, x_{nj}$.
- The data for the ith subject is $(y_i, x_{i1}, \ldots, x_{ip})$.
- The linear regression model assumes that these are independent with

$$\mathbb{E}(y_i | x_{i1}, \ldots, x_{ip}) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}, \quad \text{for } i = 1, \ldots, n.$$

Multiple Regression

- Equivalently, we can write the model as

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + e_i, \quad \text{with} \quad \mathbb{E}(e_i | x_{i1}, \ldots, x_{ip}) = 0.$$

Estimation of $\beta_0, \beta_1, \ldots, \beta_p$

- As before, we use least squares: Estimate $\beta_0, \beta_1, \ldots, \beta_p$ by the minimizers of

$$s(b) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - \ldots b_p x_{ip})^2 \tag{1}$$

over $b_0, \ldots, b_p$.

Estimation of $\beta_0, \beta_1, \ldots, \beta_p$

- As before, we use least squares: Estimate $\beta_0, \beta_1, \ldots, \beta_p$ by the minimizers of

$$s(b) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - \ldots b_p x_{ip})^2 \tag{1}$$

  over $b_0, \ldots, b_p$.
- Like in the simple regression case, we can try to take partial derivatives with respect to $b_0, \ldots, b_p$, set them equal to zero and then solve the resulting equations.

Estimation of $\beta_0, \beta_1, \ldots, \beta_p$

- As before, we use least squares: Estimate $\beta_0, \beta_1, \ldots, \beta_p$ by the minimizers of

$$s(b) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - \ldots b_p x_{ip})^2 \tag{1}$$

  over $b_0, \ldots, b_p$.
- Like in the simple regression case, we can try to take partial derivatives with respect to $b_0, \ldots, b_p$, set them equal to zero and then solve the resulting equations.
- Doing this directly using the expression (1) is painful. Things are greatly simplified if one uses matrices and vectors.

Estimation of $\beta_0, \beta_1, \ldots, \beta_p$

- As before, we use least squares: Estimate $\beta_0, \beta_1, \ldots, \beta_p$ by the minimizers of

$$s(b) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - \ldots b_p x_{ip})^2 \tag{1}$$

  over $b_0, \ldots, b_p$.
- Like in the simple regression case, we can try to take partial derivatives with respect to $b_0, \ldots, b_p$, set them equal to zero and then solve the resulting equations.
- Doing this directly using the expression (1) is painful. Things are greatly simplified if one uses matrices and vectors.
- Data on the response variable, $y_1, \ldots, y_n$, are represented by the column vector $Y = (y_1, \ldots, y_n)^T$ (the T here stands for transpose).

- Data on the jth explanatory variable x_j is $x_{1j}, x_{2j}, \ldots, x_{nj}$.

- Data on the jth explanatory variable x_j is $x_{1j}, x_{2j}, \ldots, x_{nj}$.
- The data on all the explanatory variables is represented by the $n \times (p+1)$ matrix X whose first column consists of ones, second column has the values $x_{11}, x_{21}, \ldots, x_{n1}$ corresponding to the first explanatory variable, third column has values for the second explanatory variable and so on.

- Data on the jth explanatory variable x_j is $x_{1j}, x_{2j}, \ldots, x_{nj}$.
- The data on all the explanatory variables is represented by the $n \times (p+1)$ matrix X whose first column consists of ones, second column has the values $x_{11}, x_{21}, \ldots, x_{n1}$ corresponding to the first explanatory variable, third column has values for the second explanatory variable and so on.
- With this notation, the sum of squares in (1) can be rewritten as

$$s(b) = \|Y - Xb\|_2^2.$$

where the norm of the vector x is defined as $\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$. Note the equality $\|x\|_2^2 = x^T x$.

- Data on the jth explanatory variable x_j is $x_{1j}, x_{2j}, \ldots, x_{nj}$.
- The data on all the explanatory variables is represented by the $n \times (p+1)$ matrix X whose first column consists of ones, second column has the values $x_{11}, x_{21}, \ldots, x_{n1}$ corresponding to the first explanatory variable, third column has values for the second explanatory variable and so on.
- With this notation, the sum of squares in (1) can be rewritten as

$$s(b) \;=\; \|Y - Xb\|_2^2.$$

  where the norm of the vector x is defined as $\|x\|_2 \;=\; \sqrt{\sum_{i=1}^{n} x_i^2}$. Note the equality $\|x\|_2^2 = x^T x$.
- Using this, we can write

$$s(b) = (Y - Xb)^T (y - Xb) = Y^T Y - 2b^T Xy + b^T X^T Xb.$$

prove : https://math.stackexchange.com/questions/2753210/when-can-we-say-that-atb-bta

- Data on the jth explanatory variable x_j is $x_{1j}, x_{2j}, \ldots, x_{nj}$.
- The data on all the explanatory variables is represented by the $n \times (p+1)$ matrix X whose first column consists of ones, second column has the values $x_{11}, x_{21}, \ldots, x_{n1}$ corresponding to the first explanatory variable, third column has values for the second explanatory variable and so on.
- With this notation, the sum of squares in (1) can be rewritten as

$$s(b) = \|Y - Xb\|_2^2.$$

where the norm of the vector x is defined as $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$. Note the equality $\|x\|_2^2 = x^T x$.
- Using this, we can write

$$s(b) = (Y - Xb)^T (y - Xb) = Y^T Y - 2Y^T Xb + b^T X^T Xb.$$

- This can be minimized via calculus. Take partial derivatives with respect to i for $i = 0, 1, \ldots, p$ and equate them to 0. It is easy to check that

$$\nabla S(b) = 2X^T Xb - 2X^T y$$

$$\begin{aligned} S(b) &= \sum e_i^2 = e'e = (y - Xb)'(y - Xb) \\ &= y'y - y'Xb - b'X'y + b'X'Xb. \end{aligned} \tag{3.6}$$

Derivation of least squares estimator

T

The minimum of $S(b)$ is obtained by setting the derivatives of $S(b)$ equal to zero. Note that the function $S(b)$ has scalar values, whereas b is a column vector with k components. So we have k first order derivatives and we will follow the convention to arrange them in a column vector. The second and third terms of the last expression in (3.6) are equal (a 1×1 matrix is always symmetric) and may be replaced by $-2b'X'y$. This is a linear expression in the elements of b and so the vector of derivatives equals $-2X'y$. The last term of (3.6) is a quadratic form in the elements of b. The vector of first order derivatives of this term $b'X'Xb$ can be written as $2X'Xb$. The proof of this result is left as an exercise (see Exercise 3.1). To get the idea we consider the case $k = 2$ and we denote the elements of $X'X$ by c_{ij}, $i, j = 1, 2$, with $c_{12} = c_{21}$. Then $b'X'Xb = c_{11}b_1^2 + c_{22}b_2^2 + 2c_{12}b_1b_2$. The derivative with respect to b_1 is $2c_{11}b_1 + 2c_{12}b_2$ and the derivative with respect to b_2 is $2c_{12}b_1 + 2c_{22}b_2$. When we arrange these two partial derivatives in a 2×1 vector, this can be written as $2X'Xb$. See Appendix A (especially Examples A.10 and A.11 in Section A.7) for further computational details and illustrations.

- where

$$\nabla S(b) = \left(\frac{\partial S(b)}{\partial \beta_0}, \ldots, \frac{\partial S(b)}{\partial \beta_p}\right).$$

denotes the gradient of $S(b)$ with respect to $b = (b_1, \ldots, b_p)^T$.

- where
$$\nabla S(b) = \left(\frac{\partial S(b)}{\partial \beta_0}, \ldots, \frac{\partial S(b)}{\partial \beta_p}\right).$$
denotes the gradient of $S(b)$ with respect to $b = (b_1, \ldots, b_p)^T$.
- It follows therefore that the minimizer of $S(b)$ satisfies the equality
$$X^T X b = X^T y.$$

- where

$$\nabla S(b) = \left(\frac{\partial S(b)}{\partial \beta_0}, \ldots, \frac{\partial S(b)}{\partial \beta_p} \right).$$

  denotes the gradient of $S(b)$ with respect to $b = (b_1, \ldots, b_p)^T$.
- It follows therefore that the minimizer of $S(b)$ satisfies the equality

$$X^T X b = X^T y.$$

- This gives p linear equations for the $p + 1$ unknowns $b0, \ldots, b_p$.

- where

$$\nabla S(b) = \left(\frac{\partial S(b)}{\partial \beta_0}, \ldots, \frac{\partial S(b)}{\partial \beta_p}\right).$$

  denotes the gradient of$S(b)$ with respect to $b = (b_1, \ldots, b_p)^T$.
- It follows therefore that the minimizer of $S(b)$ satisfies the equality

$$X^T X b = X^T y.$$

- This gives p linear equations for the $p + 1$ unknowns $b0, \ldots, b_p$.
- This important set of equations are called normal equations.

- where
$$\nabla S(b) = \left(\frac{\partial S(b)}{\partial \beta_0}, \ldots, \frac{\partial S(b)}{\partial \beta_p}\right).$$
denotes the gradient of $S(b)$ with respect to $b = (b_1, \ldots, b_p)^T$.
- It follows therefore that the minimizer of $S(b)$ satisfies the equality
$$X^T X b = X^T y.$$
- This gives p linear equations for the $p + 1$ unknowns $b0, \ldots, b_p$.
- This important set of equations are called normal equations.
- Their solution, denoted by $\hat{\beta} = (\hat{\beta}_0, \ldots, \hat{\beta}_p)^T$ gives an estimate of β called the least squares estimate.

- When $X^T X$ is invertible, we have a unique solution to the normal equations that is given by $(X^T X)^{-1} X^T Y$. Thus our least squares estimate is

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

‘

- When $X^T X$ is invertible, we have a unique solution to the normal equations that is given by $(X^T X)^{-1} X^T Y$. Thus our least squares estimate is

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

- What happens when $X^T X$ is not invertible?

‘

- When X^TX is invertible, we have a unique solution to the normal equations that is given by $(X^TX)^{-1}X^TY$. Thus our least squares estimate is

$$\hat{\beta} = (X^TX)^{-1}X^TY.$$

- What happens when X^TX is not invertible? We shall deal with this later.

‘

- When $X^T X$ is invertible, we have a unique solution to the normal equations that is given by $(X^T X)^{-1} X^T Y$. Thus our least squares estimate is

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

- What happens when $X^T X$ is not invertible? We shall deal with this later.
- Our next goal is to understand the properties of the least squares estimator. Before that, let us take a detour and learn some formulae for dealing with mean and variance of random vectors.

‘

Basic Mean and Covariance Formulae for Random Vectors

- We next want to explore properties of $\hat{\beta} = (X^T X)^{-1} X^T Y$ as an estimator of β in the linear model. For this we need a few facts about means and covariances.

Basic Mean and Covariance Formulae for Random Vectors

- We next want to explore properties of $\hat{\beta} = (X^T X)^{-1} X^T Y$ as an estimator of β in the linear model. For this we need a few facts about means and covariances.
- Let $Z = (Z_1, \ldots, Z_k)^T$ be a random vector. Its expectation $\mathbb{E}Z$ is defined as a vector whose ith entry is the expectation of Z_i i.e., $\mathbb{E}Z = (\mathbb{E}Z_1, \mathbb{E}Z_2, \ldots, \mathbb{E}Z_k)^T$.

Basic Mean and Covariance Formulae for Random Vectors

- We next want to explore properties of $\hat{\beta} = (X^T X)^{-1} X^T Y$ as an estimator of β in the linear model. For this we need a few facts about means and covariances.
- Let $Z = (Z_1, \dots, Z_k)^T$ be a random vector. Its expectation $\mathbb{E}Z$ is defined as a vector whose ith entry is the expectation of Z_i i.e., $\mathbb{E}Z = (\mathbb{E}Z_1, \mathbb{E}Z_2, \dots, \mathbb{E}Z_k)^T$.
- The covariance matrix of Z, denoted by $Cov(Z)$, is a $k \times k$ matrix whose (i, j)th entry is the covariance between Z_i and Z_j.

Basic Mean and Covariance Formulae for Random Vectors

- We next want to explore properties of $\hat{\beta} = (X^T X)^{-1} X^T Y$ as an estimator of β in the linear model. For this we need a few facts about means and covariances.
- Let $Z = (Z_1, \ldots, Z_k)^T$ be a random vector. Its expectation $\mathbb{E}Z$ is defined as a vector whose ith entry is the expectation of Z_i i.e., $\mathbb{E}Z = (\mathbb{E}Z_1, \mathbb{E}Z_2, \ldots, \mathbb{E}Z_k)^T$.
- The covariance matrix of Z, denoted by $Cov(Z)$, is a $k \times k$ matrix whose (i, j)th entry is the covariance between Z_i and Z_j.
- If $W = (W1, \ldots, W_m)^T$ is another random vector, the covariance matrix between Z and W, denoted by $Cov(Z, W)$, is a $k \times m$ matrix whose (i, j)th entry is the covariance between Z_i and W_j. Note then that, $Cov(Z, Z) = Cov(Z)$.

Basic Mean and Covariance Formulae for Random Vectors

The following formulae are very important:

Basic Mean and Covariance Formulae for Random Vectors

The following formulae are very important:

- $\mathbb{E}(AZ + c) = A\mathbb{E}(Z) + c$ for any constant matrix A and any constant vector c.

Basic Mean and Covariance Formulae for Random Vectors

The following formulae are very important:

- $\mathbb{E}(AZ + c) = A\mathbb{E}(Z) + c$ for any constant matrix A and any constant vector c.
- $Cov(AZ + c) = ACov(Z)A^T$ for any constant matrix A and any constant vector c.

Basic Mean and Covariance Formulae for Random Vectors

The following formulae are very important:

- $\mathbb{E}(AZ + c) = A\mathbb{E}(Z) + c$ for any constant matrix A and any constant vector c.
- $Cov(AZ + c) = ACov(Z)A^T$ for any constant matrix A and any constant vector c.
- $Cov(AZ + c, BW + d) = ACov(Z, W)B^T$ for any pair of constant matrices A and B and any pair of constant vectors c and d.

Assumptions

- The linear model is

$$Y = X\beta + e.$$

Assumptions

- The linear model is

$$Y = X\beta + e.$$

- The zero conditional mean assumption on e can be written as $\mathbb{E}(e|X) = 0$. For variance calculations, we also make the assumption that $Cov(e|X) = \sigma^2 I_n$.

Assumptions

- The linear model is

$$Y = X\beta + e.$$

- The zero conditional mean assumption on e can be written as $\mathbb{E}(e|X) = 0$. For variance calculations, we also make the assumption that $Cov(e|X) = \sigma^2 I_n$.
- Because of the above formulae, we can write

$$\mathbb{E}(Y|X) = X\beta, \quad cov(Y|X) = \sigma^2 I_n.$$

Properties of the Least Squares Estimator

- Assume that $X^T X$ is invertible (equivalently, that X has rank p + 1) and consider the least squares estimator

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

Properties of the Least Squares Estimator

- Assume that $X^T X$ is invertible (equivalently, that X has rank p + 1) and consider the least squares estimator

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$

- What properties does $\hat{\beta}$ have as an estimator of β?

Linearity

- An estimator of β is said to be linear if it can be written as AY for some matrix A.

Linearity

- An estimator of β is said to be linear if it can be written as AY for some matrix A.
- Clearly $\hat{\beta} = (X^T X)^{-1} X^T Y$ is of this form and hence it is a linear estimator of β.

Unbiasedness

- An estimator for a parameter is said to be unbiased if its expectation equals the parameter (for all values of the parameter).

Unbiasedness

- An estimator for a parameter is said to be unbiased if its expectation equals the parameter (for all values of the parameter).
- The expectation of the least squares estimator is (using the formula for expectation: $\mathbb{E}AZ = A\mathbb{E}Z$)

.

Unbiasedness

- An estimator for a parameter is said to be unbiased if its expectation equals the parameter (for all values of the parameter).
- The expectation of the least squares estimator is (using the formula for expectation: $\mathbb{E}AZ = A\mathbb{E}Z$)

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}((X^TX)^{-1}X^TY)$$

.

Unbiasedness

- An estimator for a parameter is said to be unbiased if its expectation equals the parameter (for all values of the parameter).
- The expectation of the least squares estimator is (using the formula for expectation: $\mathbb{E}AZ = A\mathbb{E}Z$)

$$\begin{aligned}\mathbb{E}(\hat{\beta}) &= \mathbb{E}((X^TX)^{-1}X^TY) \\ &= (X^TX)^{-1}X^T\mathbb{E}(Y)\end{aligned}$$

.

Unbiasedness

- An estimator for a parameter is said to be unbiased if its expectation equals the parameter (for all values of the parameter).
- The expectation of the least squares estimator is (using the formula for expectation: $\mathbb{E}AZ = A\mathbb{E}Z$)

$$\begin{aligned}\mathbb{E}(\hat{\beta}) &= \mathbb{E}((X^TX)^{-1}X^TY) \\ &= (X^TX)^{-1}X^T\mathbb{E}(Y) \\ &= (X^TX)^{-1}X^TX\beta\end{aligned}$$

.

Unbiasedness

- An estimator for a parameter is said to be unbiased if its expectation equals the parameter (for all values of the parameter).
- The expectation of the least squares estimator is (using the formula for expectation: $\mathbb{E}AZ = A\mathbb{E}Z$)

$$\begin{aligned}
\mathbb{E}(\hat{\beta}) &= \mathbb{E}((X^TX)^{-1}X^TY) \\
&= (X^TX)^{-1}X^T\mathbb{E}(Y) \\
&= (X^TX)^{-1}X^TX\beta \\
&= \beta.
\end{aligned}$$

.

Unbiasedness

- An estimator for a parameter is said to be unbiased if its expectation equals the parameter (for all values of the parameter).
- The expectation of the least squares estimator is (using the formula for expectation: $\mathbb{E}AZ = A\mathbb{E}Z$)

$$\begin{aligned}
\mathbb{E}(\hat{\beta}) &= \mathbb{E}((X^TX)^{-1}X^TY) \\
&= (X^TX)^{-1}X^T\mathbb{E}(Y) \\
&= (X^TX)^{-1}X^TX\beta \\
&= \beta.
\end{aligned}$$

- In particular, this means that $\mathbb{E}\hat{\beta}_i = \beta_i$ for each i which implies that each $\hat{\beta}_i$ is an unbiased estimator of β_i.

Covariance Matrix

- The Covariance matrix of the estimator $\hat{\beta}$ can be easily calculated using the formula: $Cov(AZ) = ACov(Z)AT$:

Covariance Matrix

- The Covariance matrix of the estimator $\hat{\beta}$ can be easily calculated using the formula: $Cov(AZ) = ACov(Z)AT$:

$$Cov(\hat{\beta}) \quad = \quad Cov((X^T X)^{-1} X^T Y)$$

Covariance Matrix

- The Covariance matrix of the estimator $\hat{\beta}$ can be easily calculated using the formula: $Cov(AZ) = ACov(Z)AT$:

$$
\begin{aligned}
Cov(\hat{\beta}) &= Cov((X^T X)^{-1} X^T Y) \\
&= (X^T X)^{-1} X^T Cov(Y) X (X^T X)^{-1}
\end{aligned}
$$

Covariance Matrix

- The Covariance matrix of the estimator $\hat{\beta}$ can be easily calculated using the formula: $Cov(AZ) = ACov(Z)AT$:

$$\begin{aligned}
Cov(\hat{\beta}) &= Cov((X^TX)^{-1}X^TY) \\
&= (X^TX)^{-1}X^TCov(Y)X(X^TX)^{-1} \\
&= (X^TX)^{-1}X^T(\sigma^2 I_n)X(X^TX)^{-1}
\end{aligned}$$

Covariance Matrix

- The Covariance matrix of the estimator $\hat{\beta}$ can be easily calculated using the formula: $Cov(AZ) = ACov(Z)AT$:

$$\begin{aligned}
Cov(\hat{\beta}) &= Cov((X^TX)^{-1}X^TY) \\
&= (X^TX)^{-1}X^TCov(Y)X(X^TX)^{-1} \\
&= (X^TX)^{-1}X^T(\sigma^2 I_n)X(X^TX)^{-1} \\
&= \sigma^2(X^TX)^{-1}.
\end{aligned}$$

Covariance Matrix

- The Covariance matrix of the estimator $\hat{\beta}$ can be easily calculated using the formula: $Cov(AZ) = ACov(Z)AT$:

$$\begin{aligned} Cov(\hat{\beta}) &= Cov((X^TX)^{-1}X^TY) \\ &= (X^TX)^{-1}X^TCov(Y)X(X^TX)^{-1} \\ &= (X^TX)^{-1}X^T(\sigma^2 I_n)X(X^TX)^{-1} \\ &= \sigma^2(X^TX)^{-1}. \end{aligned}$$

- In particular, the variance of $\hat{\beta}_i$ equals σ^2 multiplied by the ith diagonal element of $(X^TX)^{-1}$.

Covariance Matrix

- The Covariance matrix of the estimator $\hat{\beta}$ can be easily calculated using the formula: $Cov(AZ) = ACov(Z)AT$:

$$\begin{aligned}
Cov(\hat{\beta}) &= Cov((X^TX)^{-1}X^TY) \\
&= (X^TX)^{-1}X^TCov(Y)X(X^TX)^{-1} \\
&= (X^TX)^{-1}X^T(\sigma^2 I_n)X(X^TX)^{-1} \\
&= \sigma^2(X^TX)^{-1}.
\end{aligned}$$

- In particular, the variance of $\hat{\beta}_i$ equals σ^2 multiplied by the ith diagonal element of $(X^TX)^{-1}$.
- Once we learn how to estimate σ^2, we can use this to obtain standard errors for $\hat{\beta}_i$.

Optimality - The Gauss-Markov Theorem

- The Gauss-Markov Theorem states that $\hat{\beta}$ is BLUE (Best Linear Unbiased Estimator).

Optimality - The Gauss-Markov Theorem

- The Gauss-Markov Theorem states that $\hat{\beta}$ is BLUE (Best Linear Unbiased Estimator).
- This means that $\hat{\beta}$ is the best estimator among all linear and unbiased estimators of β. Here, best is in terms of variance.

Optimality - The Gauss-Markov Theorem

- The Gauss-Markov Theorem states that $\hat{\beta}$ is BLUE (Best Linear Unbiased Estimator).
- This means that $\hat{\beta}$ is the best estimator among all linear and unbiased estimators of β. Here, best is in terms of variance.
- This implies that $\hat{\beta}_i$ has the smallest variance among all linear and unbiased estimators of i for every i.

Optimality - The Gauss-Markov Theorem

- The Gauss-Markov Theorem states that $\hat{\beta}$ is BLUE (Best Linear Unbiased Estimator).
- This means that $\hat{\beta}$ is the best estimator among all linear and unbiased estimators of β. Here, best is in terms of variance.
- This implies that $\hat{\beta}_i$ has the smallest variance among all linear and unbiased estimators of *i* for every *i*.
- The Gauss-Markov theorem is actually quite simple to prove. We shall prove it in the next class.