Regression Analysis

MIT 18.472

Dr. Kempthorne

Spring 2015

Outline

Multiple Linear Regression: Setup

Data Set

- n cases $i = 1, 2, \dots, n$
- 1 Response (dependent) variable

$$y_i,\ i = 1, 2, \dots, n$$

- p Explanatory (independent) variables

$$\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})^T,\ i = 1, 2, \dots, n$$

Goal of Regression Analysis:

- Extract/exploit relationship between y_i and $\mathbf{x}_i$.

Examples

- Prediction
- Causal Inference
- Approximation
- Functional Relationships

General Linear Model: For each case i, the conditional distribution $[y_i \mid x_i]$ is given by

$$y_i = \hat{y}_i + \epsilon_i$$

where

- $\hat{y}_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_{i,p} x_{i,p}$
- $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T$ are p regression parameters (constant over all cases)
- ϵ_i Residual (error) variable (varies over all cases)

Extensive breadth of possible models

- Polynomial approximation ($x_{i,j} = (x_i)^j$, explanatory variables are different powers of the same variable $x = x_i$)
- Fourier Series: ($x_{i,j} = sin(jx_i)$ or $cos(jx_i)$, explanatory variables are different sin/cos terms of a Fourier series expansion)
- Time series regressions: time indexed by i, and explanatory variables include lagged response values.

Note: *Linearity* of $\hat{y}_i$ (in regression parameters) maintained with non-linear x.

Steps for Fitting a Model

(1) Propose a model in terms of
- Response variable Y (specify the scale)
- Explanatory variables $X_1, X_2, \ldots X_p$ (include different functions of explanatory variables if appropriate)
- Assumptions about the distribution of ϵ over the cases

(2) Specify/define a criterion for judging different estimators.

(3) Characterize the best estimator and apply it to the given data.

(4) Check the assumptions in (1).

(5) If necessary modify model and/or assumptions and go to (1).

Specifying Assumptions in (1) for Residual Distribution

- Gauss-Markov: zero mean, constant variance, uncorrelated
- Normal-linear models: ϵ_i are i.i.d. $N(0, \sigma^2)$ r.v.s
- Generalized Gauss-Markov: zero mean, and general covariance matrix (possibly correlated,possibly heteroscedastic)
- Non-normal/non-Gaussian distributions (e.g., Laplace, Pareto, Contaminated normal: some fraction $(1-\delta)$ of the ϵ_i are i.i.d. $N(0, \sigma^2)$ r.v.s the remaining fraction (δ) follows some contamination distribution).

Specifying Estimator Criterion in (2)

- Least Squares
- Maximum Likelihood
- Robust (Contamination-resistant)
- Bayes (assume β_j are r.v.'s with known *prior* distribution)
- Accommodating incomplete/missing data

Case Analyses for (4) Checking Assumptions

- Residual analysis
  - Model errors ϵ_i are unobservable
  - Model residuals for fitted regression parameters $\tilde{\beta}_j$ are:
    $$e_i = y_i - [\tilde{\beta}_1 x_{i,1} + \tilde{\beta}_2 x_{i,2} + \cdots + \tilde{\beta}_p x_{i,p}]$$
- Influence diagnostics (identify cases which are highly 'influential'?)
- Outlier detection

Outline

Ordinary Least Squares Estimates

Least Squares Criterion: For $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$, define

$$\begin{aligned} Q(\boldsymbol{\beta}) &= \textstyle\sum_{i=1}^N [y_i - \hat{y}_i]^2 \\ &= \textstyle\sum_{i=1}^N [y_i - (\beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_{i,p} x_{i,p})]^2 \end{aligned}$$

Ordinary Least-Squares (OLS) estimate $\hat{\boldsymbol{\beta}}$: minimizes $Q(\boldsymbol{\beta})$.

Matrix Notation

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{p,n} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

Solving for OLS Estimate $\hat{\beta}$

$$\hat{\mathbf{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \mathbf{X}\beta \text{ and}$$

$$\begin{aligned} Q(\beta) &= \textstyle\sum_{i=1}^n (y_i - \hat{y}_i)^2 = (\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}}) \\ &= (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \end{aligned}$$

OLS $\hat{\beta}$ solves $\frac{\partial Q(\beta)}{\partial \beta_j} = 0, \quad j = 1, 2, \ldots, p$

$$\begin{aligned} \frac{\partial Q(\beta)}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \left(\textstyle\sum_{i=1}^n [y_i - (x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots x_{i,p}\beta_p)]^2 \right) \\ &= \textstyle\sum_{i=1}^n 2(-x_{i,j})[y_i - (x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots x_{i,p}\beta_p)] \\ &= -2(\mathbf{X}_{[j]})^T (\mathbf{y} - \mathbf{X}\beta) \quad \text{where } \mathbf{X}_{[j]} \text{ is the } j\text{th column of } \mathbf{X} \end{aligned}$$

Solving for OLS Estimate $\hat{\beta}$

$$\frac{\partial Q}{\partial \beta} = \begin{bmatrix} \frac{\partial Q}{\partial \beta_1} \\ \frac{\partial Q}{\partial \beta_2} \\ \vdots \\ \frac{\partial Q}{\partial \beta_p} \end{bmatrix} = -2 \begin{bmatrix} \mathbf{X}_{[1]}^T(\mathbf{y} - \mathbf{X}\beta) \\ \mathbf{X}_{[2]}^T(\mathbf{y} - \mathbf{X}\beta) \\ \vdots \\ \mathbf{X}_{[p]}^T(\mathbf{y} - \mathbf{X}\beta) \end{bmatrix} = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta)$$

So the OLS Estimate $\hat{\beta}$ solves the **"Normal Equations"**

$$\begin{aligned} \mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta) &= \mathbf{0} \\ \iff \quad \mathbf{X}^T\mathbf{X}\hat{\beta} &= \mathbf{X}^T\mathbf{y} \\ \implies \quad \hat{\beta} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \end{aligned}$$

N.B. For $\hat{\beta}$ to exist (uniquely)

$(\mathbf{X}^T\mathbf{X})$ must be invertible

$\iff$ $\mathbf{X}$ must have Full Column Rank

(Ordinary) Least Squares Fit

OLS Estimate:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Fitted Values:

$$\hat{\mathbf{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \begin{pmatrix} x_{1,1}\hat{\beta}_1 + \cdots + x_{1,p}\hat{\beta}_p \\ x_{2,1}\hat{\beta}_1 + \cdots + x_{2,p}\hat{\beta}_p \\ \vdots \\ x_{n,1}\hat{\beta}_1 + \cdots + x_{n,p}\hat{\beta}_p \end{pmatrix}$$

$$= \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{H}\mathbf{y}$$

Where $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is the $n \times n$ "Hat Matrix"

(Ordinary) Least Squares Fit

The Hat Matrix $\mathbf{H}$ projects R^n onto the column-space of $\mathbf{X}$

Residuals: $\hat{\epsilon}_i = y_i - \hat{y}_i$, $i = 1, 2, \ldots, n$

$$\hat{\boldsymbol{\epsilon}} = \begin{pmatrix} \hat{\epsilon}_1 \\ \hat{\epsilon}_2 \\ \vdots \\ \hat{\epsilon}_n \end{pmatrix} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$$

Normal Equations: $\mathbf{X}^T(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{X}^T\hat{\epsilon} = \mathbf{0}_p = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$

N.B. The Least-Squares Residuals vector $\hat{\epsilon}$ is orthogonal to the column space of $\mathbf{X}$

Outline

Vector-Valued Random Variables

Random Vector and Mean Vector

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad E[\mathbf{Y}] = \boldsymbol{\mu}_Y = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

where

- $Y_1, Y_2, \ldots, Y_n$ have joint pdf $f(y_1, y_2, \ldots, , y_n)$
- $E(Y_i) = \mu_i$, $i = 1, 2, \ldots, n$

Covariance Matrix

- $Var(Y_i) = \sigma_{ii}$, $i = 1, \ldots, n$
- $Cov(Y_i, Y_j) = \sigma_{ij}$, $i, j = 1, \ldots, n$

$\Sigma = ||\sigma_{i,j}||$: $(n \times n)$ matrix with (i, j) element σ_{ij}

Vector-Valued Random Variables

Covariance Matrix

$$Cov(\mathbf{Y}) = \mathbf{\Sigma} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,p} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n,1} & \sigma_{n,2} & \cdots & \sigma_{p,n} \end{bmatrix}$$

Theorem. Suppose

- $\mathbf{Y}$ is a random n-vector with $E(\mathbf{Y}) = \mu_{\mathbf{Y}}$ and $Cov(\mathbf{Y}) = \mathbf{\Sigma}_{YY}$
- $\mathbf{A}$ is a fixed $(m \times n)$ matrix
- $\mathbf{c}$ is a fixed $(m \times 1)$ vector.

Then for the random m-vector: $\mathbf{Z} = \mathbf{c} + \mathbf{AY}$

- $E(\mathbf{Z}) = \mathbf{c} + \mathbf{A}E(\mathbf{Y}) = \mathbf{c} + \mathbf{A}\mu_{\mathbf{Y}}$
- $Cov(\mathbf{Z}) = \mathbf{\Sigma}_{ZZ} = \mathbf{A\Sigma}_{YY}\mathbf{A}^T$

Vector-Valued Random Variables

Random m-vector: $\mathbf{Z} = \mathbf{c} + \mathbf{A}\mathbf{Y}$

Example 1

- Y_i i.i.d. with mean μ and variance σ^2.
- $\mathbf{c} = 0$ and $\mathbf{A} = [1, 1, \ldots, 1]^T$. $(m = 1)$

Example 2

- Y_i i.i.d. with mean μ and variance σ^2.
- $\mathbf{c} = 0$ and $\mathbf{A} = [1/n, 1/n, \ldots, 1/n]^T$. $(m = 1)$

Example 3

- Y_i i.i.d. with mean μ and variance σ^2.
- $\mathbf{c} = 0$ and $\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \cdots 0 \\ 1 & 1 & 0 & 0 \cdots 0 \\ 1 & 1 & 1 & 0 \cdots 0 \end{bmatrix}$

Vector-Valued Random Variables

Quadratic Form

- $\mathbf{A}$ an $(n \times n)$ symmetric matrix
- $\mathbf{x}$ an n-vector (an $n \times 1$ matrix)

$$\begin{aligned} QF(\mathbf{x}, \mathbf{A}) &= \mathbf{x}^T \mathbf{A} \mathbf{x} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i A_{ij} x_j \end{aligned}$$

Theorem. Let $\mathbf{X}$ be a random n-vector with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. For fixed $n \times n$ matrix $\mathbf{A}$

$$E[\mathbf{X}^T \mathbf{A} \mathbf{X}] = trace(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}$$

(trace of a square matrix is sum of diagonal terms).

Example: If $\boldsymbol{\Sigma} = \sigma^2 \mathrm{I}$, then $E[\sum_{i=1}^{n}(X_i - \overline{X})^2] = (n-1)\sigma^2$

- $\mathbf{A} = \mathrm{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$
- $\mathbf{X}^T \mathbf{A} \mathbf{X} = \sum_{i=1}^{n}(X_i - \overline{X})^2$

Vector-Valued Random Variables

Theorem. Let

- $\mathbf{X}$ be a random n-vector with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.
- For fixed $p \times n$ matrix $\mathbf{A}$
  $$\mathbf{Y} = \mathbf{A}\mathbf{X}$$
- For fixed $m \times n$ matrix $\mathbf{B}$
  $$\mathbf{Z} = \mathbf{B}\mathbf{X}$$

Then the cross-covariance matrix of $\mathbf{Y}$ and $\mathbf{Z}$ is

$$\boldsymbol{\Sigma}_{YZ} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^T$$

Example: If $\mathbf{X}$ is a random n-vector with

mean $\boldsymbol{\mu} = \mu\mathbf{1}$ and covariance $\boldsymbol{\Sigma} = \sigma^2\mathrm{I}$.

- $\mathbf{A} = \mathrm{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$
- $\mathbf{B} = \frac{1}{n}\mathbf{1}$

Solve for $\mathbf{Y}$, $\mathbf{Z}$ and $Cov(\mathbf{Y}, \mathbf{Z})$

Outline

Least Squares Estimate

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{AY}$$

- Mean:

$$\begin{aligned} E(\hat{\beta}) &= E(\mathbf{AY}) \\ &= \mathbf{A}E(\mathbf{Y}) = \mathbf{AX}\beta \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\beta \\ &= \beta \end{aligned}$$

- Covariance:

$$\begin{aligned} Cov(\hat{\beta}) &= \mathbf{A}Cov(\mathbf{Y})\mathbf{A}^T \\ &= \mathbf{A}(\sigma^2\mathrm{I})\mathbf{A}^T = \sigma^2\mathbf{AA}^T \\ &= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \end{aligned}$$

Outline

Normal Linear Regression Models

Distribution Theory

$$\begin{aligned} Y_i &= x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots x_{i,p}\beta_p + \epsilon_i \\ &= \mu_i + \epsilon_i \end{aligned}$$

Assume $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_n\}$ are i.i.d $N(0, \sigma^2)$.

$$\Longrightarrow [Y_i \mid x_{i,1}, x_{i,2}, \ldots, x_{i,p}, \boldsymbol{\beta}, \sigma^2] \sim N(\mu_i, \sigma^2),$$

independent over $i = 1, 2, \ldots n$.

Conditioning on X, $\boldsymbol{\beta}$, and σ^2

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \text{ where } \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \sim N_n(\mathbf{O}_n, \sigma^2\mathbf{I}_n)$$

Distribution Theory

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = E(\mathbf{Y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = \mathbf{X}\boldsymbol{\beta}$$

$$\mathbf{\Sigma} = Cov(\mathbf{Y} \mid \mathbf{X}, \beta, \sigma^2) = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_n$$

That is, $\mathbf{\Sigma}_{i,j} = Cov(Y_i, Y_j \mid \mathbf{X}, \beta, \sigma^2) = \sigma^2 \times \delta_{i,j}$.

Apply Moment-Generating Functions (MGFs) to derive

- Joint distribution of $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$
- Joint distribution of $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)^T$.

MGF of Y

For the n-variate r.v. $\mathbf{Y}$, and constant $n-$vector $\mathbf{t} = (t_1, \ldots, t_n)^T$,

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= E(e^{\mathbf{t}^T\mathbf{Y}}) = E(e^{t_1Y_1 + t_2Y_2 + \cdots t_nY_n}) \\ &= E(e^{t_1Y_1}) \cdot E(e^{t_2Y_2}) \cdots E(e^{t_nY_n}) \\ &= M_{Y_1}(t_1) \cdot M_{Y_2}(t_2) \cdots M_{Y_n}(t_n) \\ &= \prod_{i=1}^n e^{t_i\mu_i + \frac{1}{2}t_i^2\sigma^2} \\ &= e^{\sum_{i=1}^n t_i\mu_i + \frac{1}{2}\sum_{i,k=1}^n t_i \mathbf{\Sigma}_{i,k} t_k} = e^{\mathbf{t}^T\mathbf{u} + \frac{1}{2}\mathbf{t}^T\mathbf{\Sigma}\mathbf{t}} \end{aligned}$$

$\Longrightarrow \mathbf{Y} \sim N_n(\boldsymbol{\mu}, \mathbf{\Sigma})$

Multivariate Normal with mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$

MGF of $\hat{\beta}$

For the p-variate r.v. $\hat{\beta}$, and constant $p-$vector $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)^T$,

$$M_{\hat{\beta}}(\boldsymbol{\tau}) = E(e^{\boldsymbol{\tau}^T \hat{\beta}}) = E(e^{\tau_1 \hat{\beta}_1 + \tau_2 \hat{\beta}_2 + \cdots \tau_p \beta_p})$$

Defining $\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ we can express

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

and

$$\begin{aligned} M_{\hat{\beta}}(\boldsymbol{\tau}) &= E(e^{\boldsymbol{\tau}^T \hat{\beta}}) \\ &= E(e^{\boldsymbol{\tau}^T \mathbf{AY}}) \\ &= E(e^{\mathbf{t}^T \mathbf{Y}}), \quad \text{with } \mathbf{t} = \mathbf{A}^T \boldsymbol{\tau} \\ &= M_{\mathbf{Y}}(\mathbf{t}) \\ &= e^{\mathbf{t}^T \mathbf{u} + \frac{1}{2}\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}} \end{aligned}$$

MGF of $\hat{\beta}$

For

$$\begin{aligned} M_{\hat{\beta}}(\boldsymbol{\tau}) &= E(e^{\boldsymbol{\tau}^T\hat{\beta}}) \\ &= e^{\mathbf{t}^T\mathbf{u}+\frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t}} \end{aligned}$$

Plug in:

$$\begin{aligned} \mathbf{t} &= \mathbf{A}^T\boldsymbol{\tau} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\boldsymbol{\tau} \\ \boldsymbol{\mu} &= \mathbf{X}\beta \\ \boldsymbol{\Sigma} &= \sigma^2\mathbf{I}_n \end{aligned}$$

Gives:

$$\begin{aligned} \mathbf{t}^T\boldsymbol{\mu} &= \boldsymbol{\tau}^T\beta \\ \mathbf{t}^T\boldsymbol{\Sigma}\mathbf{t} &= \boldsymbol{\tau}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T[\sigma^2\mathbf{I}_n]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\boldsymbol{\tau} \\ &= \boldsymbol{\tau}^T[\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}]\boldsymbol{\tau} \end{aligned}$$

So the MGF of $\hat{\beta}$ is

$$\begin{aligned} M_{\hat{\beta}}(\boldsymbol{\tau}) &= e^{\boldsymbol{\tau}^T\beta+\frac{1}{2}\boldsymbol{\tau}^T[\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}]\boldsymbol{\tau}} \\ \iff \quad \hat{\beta} &\sim N_p(\beta, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}) \end{aligned}$$

Marginal Distributions of Least Squares Estimates

Because

$$\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1})$$

the marginal distribution of each $\hat{\beta}_j$ is:

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 C_{j,j})$$

where $C_{j.j} = j$th diagonal element of $(\mathbf{X}^T\mathbf{X})^{-1}$