Stat 206: Linear Models

Lecture 6

October 14, 2015

General Linear Regression Models

- Often a number of variables affect the response variable in important and distinctive ways such that any single variable wouldn't have provided an adequate description.
- Examples. The weight of a person may be affected by height, gender, age, diet, etc. The income of a person may be affected by age, gender, years of education, etc. The body fat of a person may be associated with age, gender, weight, height, etc.

General linear regression model: for $i = 1, \cdots n$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i. \qquad (1)$$

- Y_i : value of the response variable Y in the *ith* case.
- $X_{i1}, \cdots, X_{i,p-1}$: values of the variables $X_1, \cdots, X_{p-1}$ in the *ith* case.
- $\beta_0, \beta_1, \cdots, \beta_{p-1}$: regression coefficients.
  - p: the number of regression coefficients.
  - In simple regression $p =$.
- ϵ_i: error terms where
- Response function (surface)/ mean response:

General linear regression model: for $i = 1, \cdots n$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i. \tag{2}$$

- Y_i : value of the response variable Y in the *ith* case.
- $X_{i1}, \cdots, X_{i,p-1}$: values of the variables $X_1, \cdots, X_{p-1}$ in the *ith* case.
- $\beta_0, \beta_1, \cdots, \beta_{p-1}$: regression coefficients.
    - p: the number of regression coefficients.
    - In simple regression $p = 2$.
- ϵ_i: error terms where $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$, $Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$.
- Response function (surface)/ mean response:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{p-1} X_{p-1}. \tag{3}$$

First-Order Models

$X_1, \cdots, X_{p-1}$ represent $p-1$ predictor variables.

- Response function is a in $\mathbb{R}^p$.
- β_k indicates the change in with a unit increase in the predictor X_k, when all other predictors are held constant. This change is irrespective of the levels at which other predictors are held.
- **The effects of the predictor variables are**

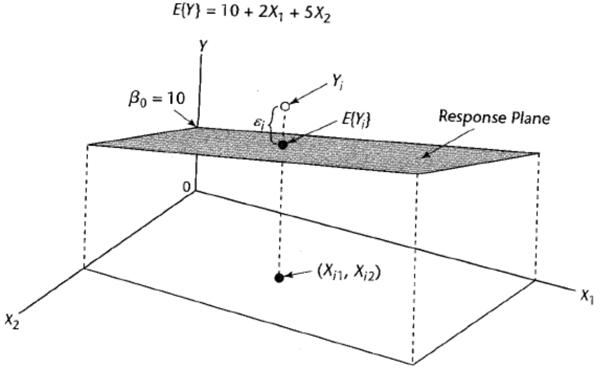
.

First-Order Models

$X_1, \cdots, X_{p-1}$ represent $p-1$ **distinct** predictor variables.

- Response function is a **hyperplane** in $\mathbb{R}^p$.
- β_k indicates the change in mean response $E(Y)$ with a unit increase in the predictor X_k, when all other predictors are held constant. This change is the same irrespective of the levels at which other predictors are held.
- **The effects of the predictor variables are additive without interactions.**

Figure: Response plane for a first-order model with two predictors.



Models with Interactions

Sometimes the effect of one predictor depends on
of the other predictor(s), i.e., the effects are .

- For example: How education level affects income may depend on gender.
- These models include the terms.
- Example. Non-additive model with two predictors:
  - This model is in the form of the general linear model with $p - 1 =$ by defining $X_{i3} :=$.
  - The mean response $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$ is in the parameters $\beta_0, \beta_1, \beta_2$, but is in the original predictors X_1, X_2.

Models with Interactions

Sometimes the effect of one predictor depends on the value(s) of the other predictor(s), i.e., the effects are **non-additive or interacting**.

- For example: How education level affects income may depend on gender.
- These models include the cross product terms.
- Example. Non-additive model with two predictors:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i, \quad i = 1, \cdots, n.$$

  - This model is in the form of the general linear model with $p - 1 = 3$ by defining $X_{i3} := X_{i1} X_{i2}$.
  - The mean response $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$ is linear in the parameters $\beta_0, \beta_1, \beta_2$, but is not linear in the original predictors X_1, X_2.

Example

Brand-liking (Y)	Moisture (X1)	Sweetness (X2)
64.0	4.0	2.0
73.0	4.0	4.0
61.0	4.0	2.0
76.0	4.0	4.0
...	...	...

Design matrix of the first-order model:

Design matrix of the non-additive model:

Example

Brand-liking (Y)	Moisture (X1)	Sweetness (X2)
64.0	4.0	2.0
73.0	4.0	4.0
61.0	4.0	2.0
76.0	4.0	4.0
...	...	...

Design matrix of a first-order model:

$$\mathbf{X} = \begin{bmatrix} 1 & 4.0 & 2.0 \\ 1 & 4.0 & 4.0 \\ 1 & 4.0 & 2.0 \\ 1 & 4.0 & 4.0 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

Design matrix of a non-additive model:

$$\mathbf{X} = \begin{bmatrix} 1 & 4.0 & 2.0 & 8.0 \\ 1 & 4.0 & 4.0 & 16.0 \\ 1 & 4.0 & 2.0 & 8.0 \\ 1 & 4.0 & 4.0 & 16.0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Polynomial Regression Models

These models contain terms of the predictor variable(s), making the response function .

- Example. 2nd-order polynomial regression model with one predictor:

  - By defining, , this model is in the form of the general linear model with $p - 1 =$.

Polynomial Regression Models

These models contain squared and/or higher-order terms of the predictor variable(s), making the response function curvilinear.

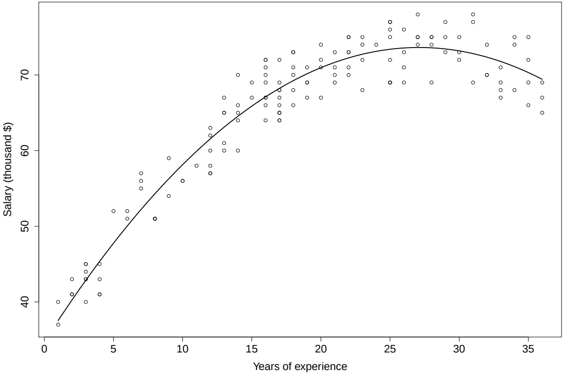
- 2nd-order polynomial regression model with one predictor:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i, \quad i = 1, \cdots, n.$$

  - By defining, $X_{i1} := X_i, X_{i2} := X_i^2$, this model is in the form of the general linear model with $p - 1 = 2$.

Example

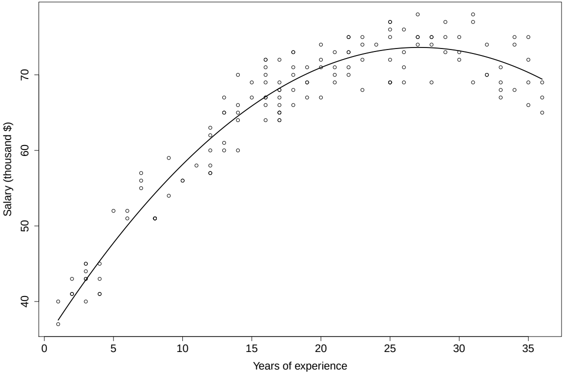
Figure: Scatter plot of salary against years of experience



The regression relation appears to be .

Example

Figure: Scatter plot of salary against years of experience



The regression relation appears to be quadratic.

Case	Salary	Experience
1	71	26
2	69	19
3	73	22
4	69	17
5	65	13
6	75	25
...	...	...

Design matrix of a 2nd-order polynomial regression model:

Case	Salary	Experience
1	71	26
2	69	19
3	73	22
4	69	17
5	65	13
6	75	25
...	...	...

Design matrix of a 2nd-order polynomial regression model:

$$\mathbf{X} = \begin{bmatrix} 1 & 26 & 26^2 \\ 1 & 19 & 19^2 \\ 1 & 22 & 22^2 \\ 1 & 17 & 17^2 \\ 1 & 13 & 13^2 \\ 1 & 25 & 25^2 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

Models with Transformed Variables

These models often have complex response functions/surfaces.

- Example. Model with logarithm-transformed response variable:

  - This model is in the form of the general linear model by defining .

Models with Transformed Variables

These models often have complex curvilinear response functions/surfaces.

- Example. Model with logarithm-transformed response variable:

$$\log Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i, \quad i = 1, \cdots n.$$

  - This model is in the form of the general linear model by defining $\tilde{Y}_i := \log Y_i$.

Key defining features of the general linear regression model:
The response function is in the regression coefficients: $\beta_0, \beta_1, \cdots, \beta_{p-1}$. However, the response function does not need to be linear in the predictors, i.e., the response surface could be .

- In contrasts, **nonlinear regression models** are in the parameters. For example:

$$Y_i = \beta_0 \exp(\beta_1 X_i) + \epsilon_i, \quad i = 1, \cdots n.$$

- The above model can not be expressed in the form of

  by taking transformations and/or introducing new X variables.

Key defining features of the general linear regression model: The response function is linear in the regression coefficients: $\beta_0, \beta_1, \cdots, \beta_{p-1}$. However, the response function does not need to be linear in the predictors, i.e., the response surface could be nonlinear.

- In contrasts, **nonlinear regression models** are nonlinear in the parameters. For example:

$$Y_i = \beta_0 \exp(\beta_1 X_i) + \epsilon_i, \quad i = 1, \cdots n.$$

- The above model can not be expressed in the form of model (2) by taking transformations and/or introducing new X variables.

General Linear Regression Model in Matrix Form

Model equations:

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\epsilon}},$$

where the design matrix $\mathbf{X}$ and the coefficients vector $\boldsymbol{\beta}$:

$$\underset{n\times p}{\mathbf{X}} := \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & X_{i1} & X_{i2} & \cdots & X_{i,p-1} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}, \quad \underset{p\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}.$$

Each row of $\mathbf{X}$ corresponds to and each column of X corresponds to .

General Linear Regression Model in Matrix Form

Model equations:

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\epsilon}},$$

where the design matrix $\mathbf{X}$ and the coefficients vector $\boldsymbol{\beta}$:

$$\underset{n\times p}{\mathbf{X}} := \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{i1} & X_{i2} & \cdots & X_{i,p-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}, \quad \underset{p\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}.$$

Each row of $\mathbf{X}$ corresponds to a case and each column of X corresponds to the n observations of an X variable.

Model assumptions:

- The response vector has:

- Under the Normal error model, **Y** is a vector of

.

Model assumptions:

$$\mathbf{E}\{\boldsymbol{\epsilon}\} = \mathbf{0}_n, \quad \boldsymbol{\sigma}^2\{\boldsymbol{\epsilon}\} = \sigma^2 \mathbf{I}_n.$$

- The response vector has:

$$\mathbf{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}, \quad \boldsymbol{\sigma}^2\{\mathbf{Y}\} = \sigma^2 \mathbf{I}_n.$$

- Under the Normal error model, **Y** is a vector of independent normal random variables.

Least Squares Estimators

- Least squares criterion:

$$
\begin{aligned}
Q(\mathbf{b}) &= \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{i1} - \cdots - b_{p-1} X_{i,p-1})^2 \\
&= (\mathbf{Y} - \mathbf{X}b)' (\mathbf{Y} - \mathbf{X}b), \quad \underset{p\times 1}{\mathbf{b}} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix}.
\end{aligned}
$$

- Differentiate $Q(\cdot)$ and set the gradient to zero $\Longrightarrow$ normal equation:

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}.$$

LS estimators are solutions of the normal equation:

- $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$:

- Variance-covariance matrix of $\hat{\boldsymbol{\beta}}$:

Notes: hereafter, assume $\mathbf{X}'\mathbf{X}$ *is of full rank* p*.*

LS estimators are solutions of the normal equation:

$$\underset{p\times 1}{\hat{\boldsymbol{\beta}}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{p-1} \end{bmatrix} = \underset{p\times p}{(\mathbf{X}'\mathbf{X})^{-1}} \underset{p\times n}{\mathbf{X}'} \underset{n\times 1}{\mathbf{Y}}. \tag{4}$$

- $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$:

$$\mathbf{E}\{\hat{\boldsymbol{\beta}}\} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}\{\mathbf{Y}\} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}.$$

- Variance-covariance matrix of $\hat{\boldsymbol{\beta}}$:

$$\boldsymbol{\sigma}^{\mathbf{2}}\{\boldsymbol{\beta}\} = \sigma^2 \underset{p\times p}{(\mathbf{X}'\mathbf{X})^{-1}}.$$

Notes: hereafter, assume $\mathbf{X}'\mathbf{X}$ *is of full rank p.*

Fitted Values and Residuals

- Both are of the observations vector **Y**.
- Under the Normal error model, both are .
- Expectations and variance-covariance matrices of the fitted values vector and residuals vector:

Fitted Values and Residuals

$$\underset{n\times 1}{\widehat{\mathbf{Y}}} := \begin{bmatrix} \widehat{Y}_1 \\ \widehat{Y}_2 \\ \vdots \\ \widehat{Y}_n \end{bmatrix} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}, \quad \underset{n\times 1}{\mathbf{e}} := \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \mathbf{Y} - \widehat{\mathbf{Y}} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y}.$$

- Both are linear transformations of the observations vector $\mathbf{Y}$.
- Under the Normal error model, both are normally distributed.
- Expectations and variance-covariance matrices of the fitted values vector and residuals vector:

$$\mathbf{E}\{\widehat{\mathbf{Y}}\} = \mathbf{X}\beta = \mathbf{E}\{\mathbf{Y}\}, \quad \boldsymbol{\sigma}^2\{\widehat{\mathbf{Y}}\} = \sigma^2\mathbf{H}.$$

$$\mathbf{E}\{\mathbf{e}\} = \mathbf{E}\{\mathbf{Y}\} - \mathbf{E}\{\widehat{\mathbf{Y}}\} = \mathbf{0}_n, \quad \boldsymbol{\sigma}^2\{\mathbf{e}\} = \sigma^2(\mathbf{I}_n - \mathbf{H}).$$

Hat Matrix

- $\mathbf{H}$ and $\mathbf{I}_n - \mathbf{H}$ are matrices: symmetric and idempotent.
- $rank(\mathbf{H}) =$, $rank(\mathbf{I}_n - \mathbf{H}) =$.
- $\mathbf{H}$ is the projection matrix to the column space $\langle X \rangle$ of the design matrix $\mathbf{X}$.
  - Fitted values vector $\widehat{\mathbf{Y}} = \mathbf{HY}$ is the projection of the observations vector $\mathbf{Y}$ to .
  - Residuals vector $\mathbf{e} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y}$ is to $\langle X \rangle$.

What are the covariances between $\mathbf{e}$ and $\hat{\mathbf{Y}}$, $\mathbf{e}$ and $\overline{Y}$? What's the implication under the Normal error model?

Hat Matrix

$$\underset{n\times n}{\mathbf{H}} := \underset{n\times p}{\mathbf{X}} \underset{p\times p}{(\mathbf{X}'\mathbf{X})^{-1}} \underset{p\times n}{\mathbf{X}'}.$$

- $\mathbf{H}$ and $\mathbf{I}_n - \mathbf{H}$ are projection matrices: symmetric and idempotent.
- $rank(\mathbf{H}) = p, \quad rank(\mathbf{I}_n - \mathbf{H}) = n - p.$
- $\mathbf{H}$ is the projection matrix to the column space $\langle X \rangle$ of the design matrix $\mathbf{X}$.
  - Fitted value vector $\widehat{\mathbf{Y}} = \mathbf{HY}$ is the projection of the response vector $\mathbf{Y}$ to $\langle X \rangle$.
  - Residual vector $\mathbf{e} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y}$ is orthogonal to $\langle X \rangle$.

What are the covariances between $\mathbf{e}$ and $\hat{\mathbf{Y}}$, $\mathbf{e}$ and $\overline{Y}$? What's the implication under the Normal error model?

Geometric Interpretation of Linear Regression

Figure: Orthogonal projection of response vector **Y** onto the linear subspace of $\mathbb{R}^n$ generated by the columns of the design matrix **X**.