Course Materials May Not Be Distributed or Posted Electronically

These course materials are the sole property of Dr. Todd M. Gross. They are strictly for use by students enrolled in a course taught by Dr. Gross. They may not be altered, excerpted, distributed, or posted to any website or other document-sharing service without express permission.

PSTAT 126
Regression Analysis

Dr. Todd Gross
Department of Statistics and Applied Probability
UCSB

Lecture 7

Multiple Linear Regression

Lecture Outline

- Data Project – Pick your partner
- Multiple Linear Regression Model
  - Readings: Section 3.2 (ISL)
- Example with R Commands
- Inference in Multiple Regression
  - Readings: Section 3.2 (ISL)
- Least Squares Estimation

Introduction to Multiple Linear Regression

Multiple Linear Regression

- Use <u>multiple</u> predictor variables to predict a <u>single</u> outcome variable
- Multiple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

The model has the following features:

- 1 outcome variable (Y)
- *p-1* predictor variables (X_1 through X_{p-1})
- *p* regression coefficients (β_0 through β_{p-1})

Note: Some notation systems (ISL) refer to p predictors and p+1 coefficients

Multiple Regression Examples

- Predict length of marriage based on age, hours of interaction, income
- Predict Response to Medical Treatment based on demographic and baseline characteristics (BMI, age, sex, race, baseline severity)
- Predict college GPA based on High School GPA, SAT scores, ACT scores, number of AP courses taken
- What predicts your midterm score?

Meaning of the Regression Parameters

- Intercept – the predicted value of Y <u>when all</u> X = 0
- Slope of β_1 – the predicted increase in Y per unit increase in X_1 <u>when all other X are held constant</u>
- Model is linear; Y' is a linear function of X
- Model is additive; each X makes an independent contribution to Y'

Multiple Linear Model Example

- What variables predict your midterm score?
  - Hours of Study
  - Prior Stats Courses
  - Units Taken in Current Quarter
  - Hours Worked at Outside Job

- What is the expected influence of each predictor?

Multiple Linear Model Example

- What variables predict your midterm score?
- What is the expected influence of each predictor?
  - Hours of Study: increases score
  - Prior Stats Courses: increases score
  - Units Taken: decreases score
  - Hours Worked at Outside Job: decreases score

- Possible Resulting Linear Equation:

$$\text{Midterm} = 35 + 5(\text{studyhrs}) + 2(\text{courses}) - 1.5(\text{units}) - 1(\text{workhrs})$$

  - Start with a score of 35
  - Add 5 pts for each hour of study
  - Add 2 pts for each course taken
  - Subtract pts 1.5 for each unit taken
  - Subtract 1 pts for each hour of outside work

The Regression Plane (2 predictors)

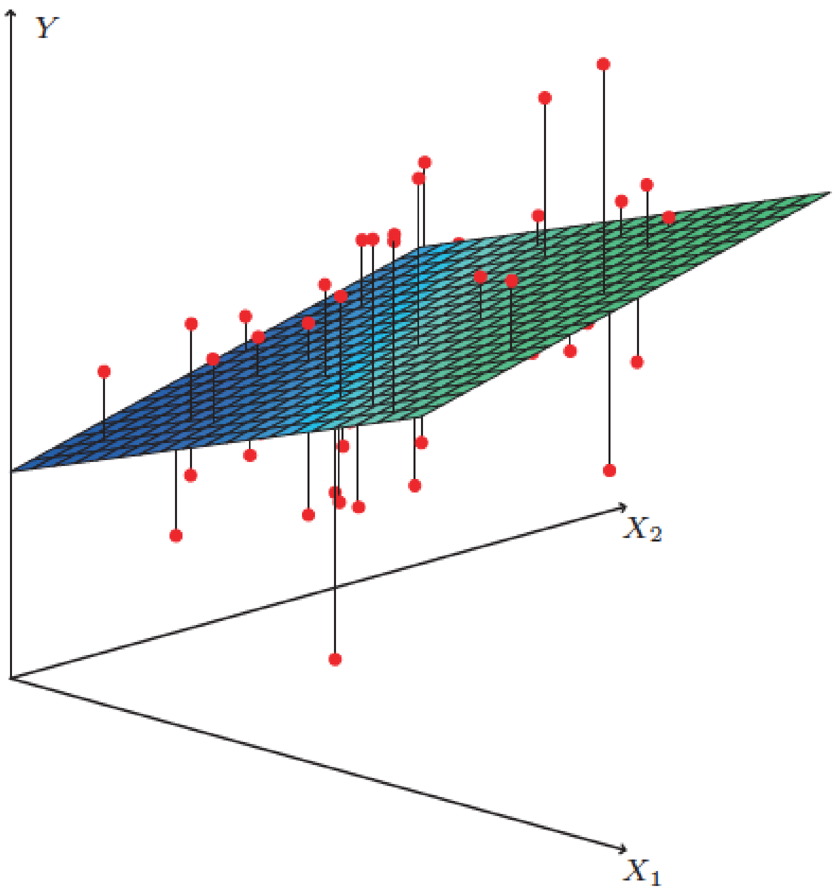


FIGURE 3.4. *In a three-dimensional setting, with two predictors and one response, the least squares regression line becomes a plane. The plane is chosen to minimize the sum of the squared vertical distances between each observation (shown in red) and the plane.*

Meaning of "Linear" in GLRM

"Linear" in a general linear regression model means the model depends linearly on the unknown parameters rather than on independent variables. The general linear regression model is quite flexible, it can be used to model nonlinear effects and/or interactions of independent variables.

Are the following models general linear regression models?

$$\boxed{\begin{aligned} Y &= \beta_1 + \beta_2 X_1 + \beta_3 X_1^2 + \beta_4 X_1 X_2 + \epsilon \\ Y &= \beta_1 + \beta_2 X_1 + \beta_3 \log X_2 + \epsilon \end{aligned}}$$

$$Y = \beta_1 + \beta_2 X^{\beta_3} + \epsilon$$

Similarities to Simple Regression

- There is only one outcome variable (Y)
- Use a regression equation to predict a value for Y
  - But we now insert a value for <u>each</u> X variable
- Residuals: $\varepsilon_i = (Y_i - Y'_i)$
- Sum of Squares Error: $SSE = \Sigma(Y_i - Y'_i)^2$

- Coefficient of Determination:

  $R^2 = SSR/SSTO = 1 - SSE/SSTO$

Multiple Linear Regression Example Using R

Example: Midterm Score

- In this example, we will predict your midterm score using two predictors, hours of study and units taken
  - Y = midterm score
  - X1 = hours of study
  - X2 = units taken
- Our dataset consists of three variables for each student

Student	Hours of Study	Units Taken	Midterm Score
1	5	16	41
2	8	13	77
3	15	10	93
4	16	10	70
5	9	11	54
(additional students)			

R Commands – Multiple Linear Model

```
>hours=c(5,8,15,16,9,7,9,6,0,15,9,3,2,2,16,14,14,14,9,8,17,
13,3,7,4)
>units=c(16,13,10,10,11,18,10,16,16,9,9,9,12,18,17,15,11,11
,10,14,13,13,12,18,19)
>midterm=c(41,77,93,70,54,40,59,34,10,75,63,22,33,9,66,88,9
2,95,50,46,98,88,32,58,48)
>fitmid=lm(midterm~hours+course+units+workhrs)
>summary(fitmid)
```

R Output – Multiple Linear Model

```
Call:
lm(formula = midterm ~ hours + units)

Residuals:
    Min      1Q  Median      3Q     Max 
-22.273  -8.656   1.270   8.189  23.845 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  20.8947    13.3094   1.570    0.131    
hours         4.5486     0.5150   8.833  1.1e-08 ***
units        -0.3176     0.8117  -0.391    0.699    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.16 on 22 degrees of freedom
Multiple R-squared:  0.807,       Adjusted R-squared:  0.7895 
F-statistic:    46 on 2 and 22 DF,  p-value: 1.383e-08
```

Strength of Fit

- `Coefficient of determination`, or R^2:

$$R^2 = \frac{\text{SSR}}{\text{SSTO}} = 1 - \frac{\text{SSE}}{\text{SSTO}}$$

- Adding more independent variables always increases R^2
- A large value of R^2 does not necessarily imply the fitted model is better, or useful

Adjusted R^2

- If we add an additional predictor, R^2 will always get larger, because additional predictors can only add to SSR, they can't reduce it
- Adjusted R^2 calculates R^2 by normalizing for the degrees of freedom, using Mean Square (MS or variance) instead of Sum of Squares (SS):

  Adj R^2 = 1 - MSE/MSTO
- Multiple R^2 is the descriptive measure of the percent of variance explained, and should be reported.
  - Adjusted R^2 is more useful for comparing different models (more on this in later lectures)

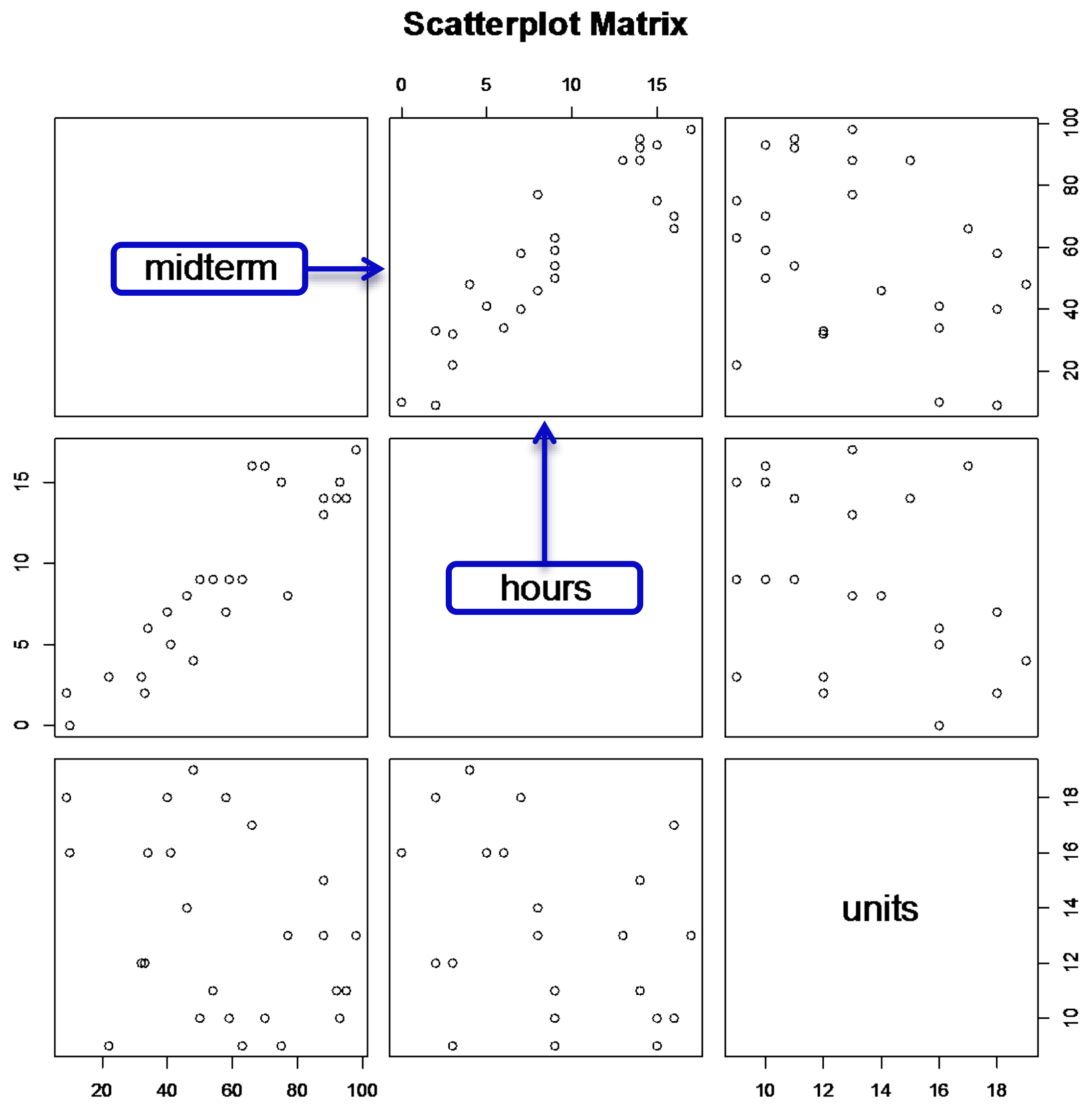
R Output – Multiple Linear Model

```
Call:
lm(formula = midterm ~ hours + units)

Residuals:
    Min      1Q  Median      3Q     Max
-22.273  -8.656   1.270   8.189  23.845

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.8947    13.3094   1.570    0.131
hours         4.5486     0.5150   8.833  1.1e-08 ***
units        -0.3176     0.8117  -0.391    0.699
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.16 on 22 degrees of freedom
Multiple R-squared:  0.807,       Adjusted R-squared:  0.7895
F-statistic:    46 on 2 and 22 DF,  p-value: 1.383e-08
```

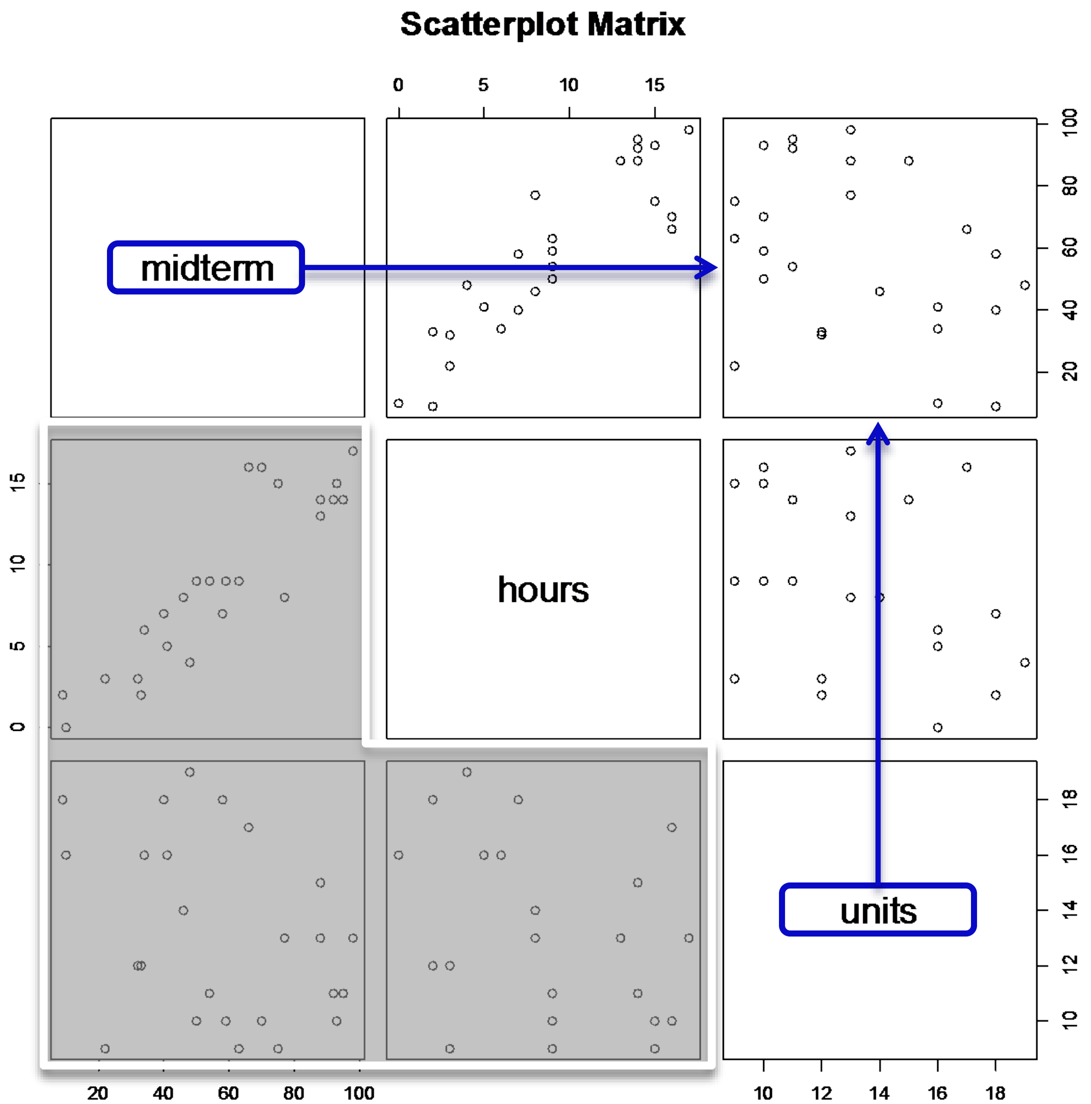
Useful Plot Commands

```
> pairs(midterm~hours+units,main="Scatterplot Matrix")
> plot(fitted(fitmid),residuals(fitmid),xlab="Predicted
Score",ylab="Residual",main="Residuals Plot")
> abline(h=0)
> qqnorm(residuals(fitmid))
> qqline(residuals(fitmid))
> hist(residuals(fitmid))
```

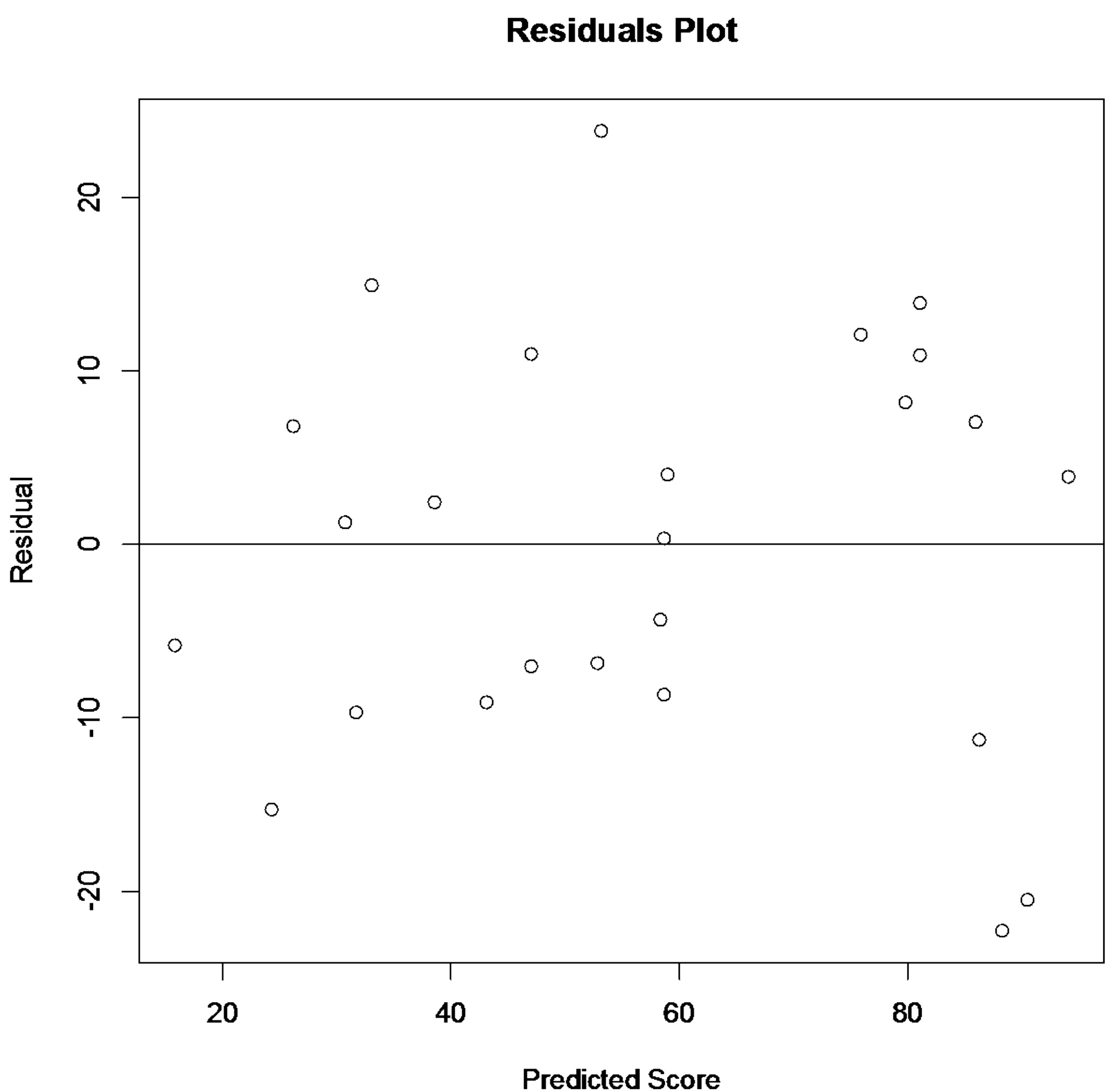
Scatterplot Matrix



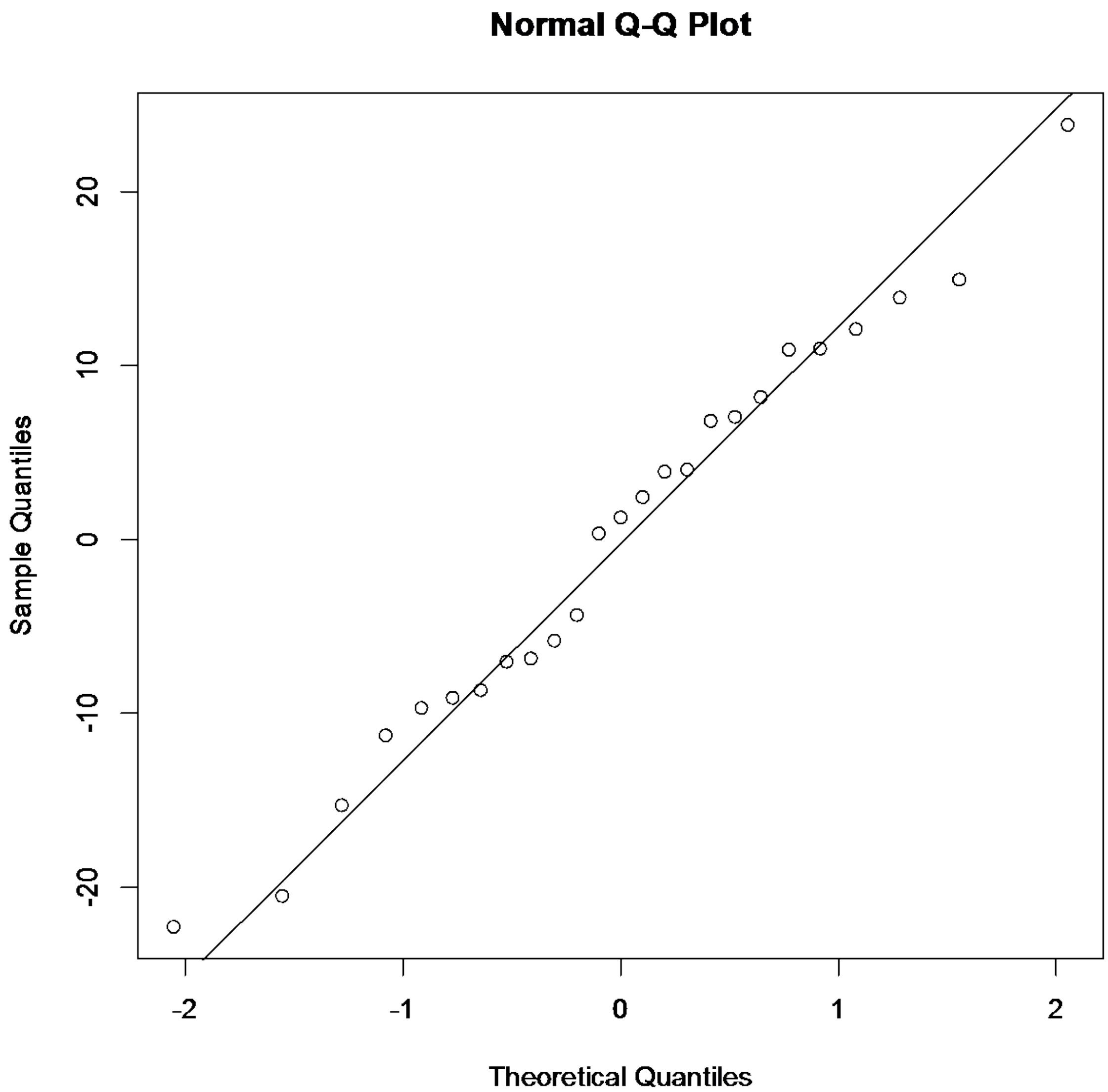
Scatterplot Matrix



Residuals Plot



QQ Normal Plot



Least Squares Estimation in Multiple Linear Regression

Data Structure of Observations

Suppose we have n observations on the dependent variable Y and $p-1$ independent variables $X_1, \cdots, X_{p-1}$:

$$
\begin{array}{cccccc}
\text{Observation} & Y & X_1 & X_2 & \cdots & X_{p-1} \\
\hline
1 & Y_1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\
2 & Y_2 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
n & Y_n & X_{n1} & X_{n2} & \cdots & X_{n,p-1}
\end{array}
\tag{1}
$$

Variables $X_1, \cdots, X_{p-1}$ do not need to represent different independent variables. Some variables may represent different transformations of an independent variable or dummy variables for a qualitative variable. Details later.

General Linear Regression Model

A general linear regression model assumes that

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

where ϵ_i's are random errors

- Similar to the simple linear regression model, we only need the assumption of $\mathrm{E}(\epsilon_i) = 0$, $\mathrm{Var}(\epsilon_i) = \sigma^2$ and $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$ for the LS estimation
- For inference, we need normalityassumption $\epsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$
- $\mathrm{E}(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1}$ and $\mathrm{Var}(Y_i) = \sigma^2$

Matrix Representation

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n\times p}{\mathbf{X}} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}$$

(6.18c) **(6.18d)**

$$\underset{p\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad \underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Then the general linear regression model can be represented in the matrix form

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon$$

Least Squares Estimation

- We estimate parameters β by minimizing the `least squares (LS)`

$$Q = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_{p-1} X_{i,p-1})^2 = ||\boldsymbol{Y} - \boldsymbol{X}\beta||^2$$

- Taking derivative of Q with respect to β and setting it equal to zero, we have the `normal equation`:

$$\boldsymbol{X}^T \boldsymbol{X} \beta = \boldsymbol{X}^T \boldsymbol{Y}$$

- Assume that X is full column rank. Then the LS estimates of β is

$$\boldsymbol{b} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}$$

Properties of LS Estimates

1. LS estimates $\boldsymbol{b}$ are linear functions of $Y_1, \cdots, Y_n$
2. $\boldsymbol{b}$ are *unbiased*

$$\mathrm{E}(\boldsymbol{b}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathrm{E}(\boldsymbol{Y}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\beta = \beta$$

3.

$$\mathrm{Var}(\boldsymbol{b}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathrm{Var}(\boldsymbol{Y})\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$$

4. Gauss-Markov theorem (BLUE): For any p-dimensional vector $\boldsymbol{c}$, $\boldsymbol{c}^T\boldsymbol{b}$ is the unique *best linear unbiased estimator* (BLUE) of $\boldsymbol{c}^T\beta$ in the sense that it has the minimum variance in the class of linear unbiased estimators for $\boldsymbol{c}^T\beta$
5. When $\epsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$, the LS estimates $\boldsymbol{b}$ are also MLEs

Some Special Cases of X Variables

- **First-order model** When all $X_1, \cdots, X_{p-1}$ represent independent variables
- **Polynomial regression** We often use polynomials such as quadratic and cubic to model nonlinear effect. For example, we may consider the following polynomial regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$

If we define $X_{i1} = X_i$ and $X_{i2} = X_i^2$, then the polynomial regression model is a special case of the general linear regression model
- **Transformed variables** The polynomial is one possible transformation. Other transformations may also be used for the dependent and/or independent variables. For example,

$$Y_i' = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$$

where $Y_i' = \log Y_i$, and $X_{i2} = \exp(X_{i1})$

Some Special Cases (con't)

- **Interaction model** Often the effect of one predictor depends on the level of another predictor. This is called `interaction`. One simple way to model interaction is to include a multiplicative term. For example,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i$$

  This is again a special case of the general linear regression model with $X_{i3} = X_{i1} X_{i2}$

- **Combination model** Combine several elements discussed above

Qualitative Independent Variables

- Qualitative variables are also known as categorical variables. Examples of qualitative variables include sex (Male and Female), hair color (Black, Blond, Brown, Gray, Red)
- For a qualitative independent variable with k levels (categories), we create $k - 1$ indicator (dummy) variables. For example, for variable sex, we may use an indicator variable

$$X = \begin{cases} 1 & \text{if female} \\ 0 & \text{if male} \end{cases}$$

For hair color, we need 4 indicator (dummy) variables:

$$X_1 = \begin{cases} 1 & \text{if Blond} \\ 0 & \text{otehrwise} \end{cases} \qquad X_2 = \begin{cases} 1 & \text{if Brown} \\ 0 & \text{otehrwise} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{if Gray} \\ 0 & \text{otehrwise} \end{cases} \qquad X_4 = \begin{cases} 1 & \text{if Red} \\ 0 & \text{otehrwise} \end{cases}$$

- Interpretation of parameters

Inference in Multiple Linear Regression

Partitioning the Total Sum of Squares

Define

$$J = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

Then

$$\begin{aligned} \mathbf{Y}^T J \mathbf{Y} &= \left(\sum_{i=1}^{n} Y_i, \cdots, \sum_{i=1}^{n} Y_i\right) \mathbf{Y} = \sum_{i=1}^{n} Y_i (1, \cdots, 1) \mathbf{Y} \\ &= \left(\sum_{i=1}^{n} Y_i\right)^2 = n^2 \bar{Y}^2 \end{aligned}$$

Partitioning the Total Sum of Squares

$$
\begin{aligned}
SSTO &= \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \\
&= \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 \\
&= \boldsymbol{Y}^T \boldsymbol{Y} - \frac{1}{n} \boldsymbol{Y}^T J \boldsymbol{Y} \\
&= \boldsymbol{Y}^T (I - \frac{J}{n}) \boldsymbol{Y} \\
&= \boldsymbol{Y}^T [(I - H) + (H - \frac{J}{n})] \boldsymbol{Y} \\
&= \boldsymbol{Y}^T (I - H) \boldsymbol{Y} + \boldsymbol{Y}^T (H - \frac{J}{n}) \boldsymbol{Y}
\end{aligned}
$$

Partitioning the Total Sum of Squares

- $SSE = \sum_{i=1}^{n} e_i^2 = \mathbf{e}^T\mathbf{e} = \mathbf{Y}^T(I-H)(I-H)\mathbf{Y} = \mathbf{Y}^T(I-H)\mathbf{Y}$
-

$$\begin{aligned} SSR &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^{n}\hat{Y}_i^2 - 2\bar{Y}\sum_{i=1}^{n}\hat{Y}_i + n\bar{Y}^2 \\ &= \sum_{i=1}^{n}\hat{Y}_i^2 - n\bar{Y}^2 \\ &= \hat{\mathbf{Y}}^T\hat{\mathbf{Y}} - n\bar{Y}^2 = (H\mathbf{Y})^T(H\mathbf{Y}) - n\bar{Y}^2 \\ &= \mathbf{Y}^T H\mathbf{Y} - \frac{1}{n}\mathbf{Y}^T J\mathbf{Y} = \mathbf{Y}^T(H - \frac{J}{n})\mathbf{Y} \end{aligned}$$

  We used the fact that $\mathbf{X}^T\mathbf{e} = 0$, then $\mathbf{X}^T\mathbf{Y} = \mathbf{X}^T\hat{\mathbf{Y}}$. Since elements in the first row of $\mathbf{X}^T$ are all 1, we have $\sum_{i=1}^{n}\hat{Y}_i = \sum_{i=1}^{n}Y_i = n\bar{Y}$
- $SSTO = SSR + SSE$
- Decomposition of df: $n-1 = (p-1) + (n-p)$

ANOVA Table for General Linear Regression Model: Matrix Algebra Formulas

Source of Variation	SS	df	MS
Regression	$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$	$p-1$	$MSR = \frac{SSR}{p-1}$
Error	$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$	$n-p$	$MSE = \frac{SSE}{n-p}$
Total	$SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y}$	$n-1$	

ANOVA Table (Simplified Form)

Source	SS	df	MS	F
Model	SSR	$p - 1$	MSR $=$ SSR$/$ (p-1)	MSR/MSE
Error	SSE	$n - p$	MSE $=$ SSE$/($ n-p $)$	
Total	SSTO	$n - 1$		

F-test for a Regression Relationship

We want to test if there is a regression relation between Y and independent variables $X_1, \cdots, X_{p-1}$:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0 \quad H_1: \text{ not all } \beta_k \text{ equal zero}$$

We use the F statistic

$$\boxed{F^* = \frac{SSR/\ p\text{-}1}{SSE/(\ n\text{-}p\ )} = \frac{MSR}{MSE}}$$

What is the distribution of F^* under H_0?

We are testing the null hypothesis that ALL of the slopes are zero (i.e., that none of the slopes are significant)

Distribution of F*

- Under H_0, $Y_1, \cdots, Y_n \overset{iid}{\sim} N(\beta_0, \sigma^2) \Rightarrow Y_i - \beta_0 \overset{iid}{\sim} N(0, \sigma^2)$
- $$\begin{aligned} SSTO &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}[(Y_i - \beta_0) - (\bar{Y} - \beta_0)]^2 \\ &= \sum_{i=1}^{n}(Y_i - \beta_0)^2 - n(\bar{Y} - \beta_0)^2 \end{aligned}$$
- We have $\sum_{i=1}^{n}(Y_i - \beta_0)^2 = n(\bar{Y} - \beta_0)^2 + SSR + SSE$
- Corresponding df: $n = 1 + (p-1) + (n-p)$
- Applying Cochran's theorem, we have $SSR/\sigma^2 \sim \chi^2_{p-1}$ and $SSE/\sigma^2 \sim \chi^2_{n-p}$ and are mutually independent
- Thus under H_0, $F^* \sim F_{p-1,n-p}$
- Reject H_0 if $F^* > F_{1-\alpha;p-1,n-p}$

R Output – Multiple Linear Model

```
Call:
lm(formula = midterm ~ hours + units)

Residuals:
    Min      1Q  Median      3Q     Max
-22.273  -8.656   1.270   8.189  23.845

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.8947    13.3094   1.570    0.131
hours         4.5486     0.5150   8.833  1.1e-08 ***
units        -0.3176     0.8117  -0.391    0.699
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '

Residual standard error: 12.16 on 22 degrees of freedom
Multiple R-squared:  0.807,       Adjusted R-squared:  0.7895
F-statistic:    46 on 2 and 22 DF,  p-value: 1.383e-08
```

Test of Overall Regression Model (H_0: All Slopes=0)

Testing Individual Slopes

- If the ANOVA is significant, then we know that not all slopes are zero.
  - At least one slope is non-zero
- We can test each slope individually, but this requires *p-1* hypothesis tests
- We now have a problem with multiplicity; multiple tests inflates the Type I error
- We usually will adjust the alpha level for the multiple tests.

Inference on parameters

- Assume that $\epsilon_i \overset{iid}{\sim} \text{N}(0, \sigma^2)$
- $\boldsymbol{b} \sim \text{N}(\beta, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$
- $MSE(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ provides an estimate of $\text{Var}(\boldsymbol{b})$
- Denote the diagonal elements of $MSE(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ as $s^2(b_k)$. They provide estimates of $\text{Var}(b_k)$
- $\frac{b_k - \beta_k}{s(b_k)} \sim t(n-p)$ for $k = 0, \cdots, p-1$

Inference on parameters

- A $100(1-\alpha)\%$ confidence interval for β_k is

$$b_k \pm t(1-\alpha/2; n-p)s(b_k)$$

- For the hypothesis

$$H_0: \ \beta_k = 0 \quad H_1: \ \beta_k \neq 0$$

  Use test statistic

$$t^* = \frac{b_k}{s(b_k)}$$

  Reject H_0 if $|t^*| > t(1-\alpha/2; n-p)$

R Output – Multiple Linear Model

```
Call:
lm(formula = midterm ~ hours + units)

Residuals:
    Min      1Q  Median      3Q     Max
-22.273  -8.656   1.270   8.189  23.845

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.8947    13.3094   1.570    0.131
hours         4.5486     0.5150   8.833  1.1e-08 ***
units        -0.3176     0.8117  -0.391    0.699
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.16 on 22 degrees of freedom
Multiple R-squared:  0.807,       Adjusted R-squared:  0.7895
F-statistic:    46 on 2 and 22 DF,  p-value: 1.383e-08
```

Tests of Individual Regression Coefficients

Confidence Intervals for Regression Coefficients

```
> confint(fitmid,level=.95)
                 2.5 %    97.5 %
(Intercept) -6.707222 48.496655
hours        3.480608  5.616574
units       -2.000922  1.365765
```

- For example:

"We are 95% confident that the population slope for hours of study is between 3.5 and 5.6"

Predicted Response

- What score would we predict for an <u>individual</u> student who studied 12 hours and is taking 16 units?

- What is the <u>mean</u> score we would predict for <u>all</u> students who studied 12 hours and are taking 16 units?

Estimation of mean response

We want to estimate the mean response $\mathrm{E}(Y_h)$ when $X_1 = X_{h,1}, \cdots, X_{p-1} = X_{h,p-1}$. Let

$$\boldsymbol{X}_h = \begin{pmatrix} 1 \\ X_{h,1} \\ \vdots \\ X_{h,p-1} \end{pmatrix}$$

An obvious estimate of $\mathrm{E}(Y_h)$ is

$$\hat{Y}_h = b_0 + b_1 X_{h,1} + \cdots + b_{p-1} X_{h,p-1} = \boldsymbol{X}_h^T \boldsymbol{b}$$

Estimation of mean response

- $$\begin{aligned} E(\hat{Y}_h) &= E(\boldsymbol{X}_h^T \boldsymbol{b}) = \boldsymbol{X}_h^T E(\boldsymbol{b}) = \boldsymbol{X}_h^T \beta = E(Y_h) \\ \mathrm{Var}(\hat{Y}_h) &= \boldsymbol{X}_h^T \mathrm{Var}(\boldsymbol{b}) \boldsymbol{X}_h = \sigma^2 \boldsymbol{X}_h^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}_h \end{aligned}$$
- $s^2(\hat{Y}_h) = MSE\,\boldsymbol{X}_h^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}_h$ provides an estimate of $\mathrm{Var}(\hat{Y}_h)$
- $\frac{\hat{Y}_h - E(Y_h)}{s(\hat{Y}_h)} \sim t(n-p)$
- A $100(1-\alpha)\%$ confidence interval for $E(Y_h)$ is

$$\hat{Y}_h \pm t(1-\alpha/2; n-p) s(\hat{Y}_h)$$

Predicted Response

- What score would we predict for an individual student who studied 12 hours and is taking 16 units?

```
> predict(fitmid,data.frame(hours=12,units=16),interval="prediction")
       fit      lwr      upr
1 70.39655 43.86154 96.93157
```

Prediction of new observation

We now consider prediction of a new observation Y_h when $X_1 = X_{h,1}, \cdots, X_{p-1} = X_{h,p-1}$.

- An estimate of Y_h is

$$\hat{Y}_h = b_0 + b_1 X_{h,1} + \cdots + b_{p-1} X_{h,p-1} = \boldsymbol{X}_h^T \boldsymbol{b}$$

- $\text{Var}(\hat{Y}_h) = \sigma^2 (1 + \boldsymbol{X}_h^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}_h)$
- $s^2(pred) = MSE(1 + \boldsymbol{X}_h^T (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}_h)$ provides an estimate of $\text{Var}(\hat{Y}_h)$
- $\frac{\hat{Y}_h - \text{E}(Y_h)}{s(pred)} \sim t(n-p)$
- A $100(1-\alpha)\%$ confidence interval for Y_h is

$$\hat{Y}_h \pm t(1-\alpha/2; n-p) s(pred)$$

Predicted Response

- What is the <u>mean</u> score we would predict for <u>all</u> students who studied 12 hours and are taking 16 units?

```
> predict(fitmid,data.frame(hours=12,units=16),interval="confidence")
       fit     lwr      upr
1 70.39655 62.1181 78.67501
```

Predicted Response

- What score would we predict for an individual student who studied 12 hours and is taking 16 units?

```
> predict(fitmid,data.frame(hours=12,units=16),interval="prediction")
       fit      lwr      upr
1 70.39655 43.86154 96.93157
```

- What is the mean score we would predict for all students who studied 12 hours and are taking 16 units?

```
> predict(fitmid,data.frame(hours=12,units=16),interval="confidence")
       fit     lwr      upr
1 70.39655 62.1181 78.67501
```

Diagnostics and Remedial Measures

- Similar to simple regression
- Inspect residuals (Y – Y')
  - Create residuals plot (against fitted Y' values, as there are now multiple predictors)
  - Create QQ normal plot
  - Create histogram of residuals
- Violations are diagnosed in the same way as simple regression
- Possible remedies are the same