Stat 206: Linear Models

Lecture 7

October 19, 2015

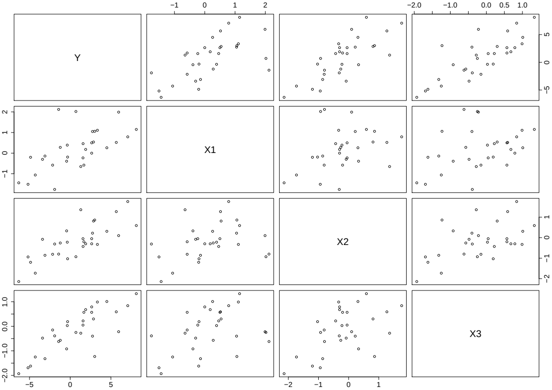
Multiple Regression: Example

$n = 30$ cases, response variable Y, three predictor variables X_1, X_2, X_3.

case	Y	X1	X2	X3
1	3.01	1.06	0.86	-1.23
2	-3.40	-0.30	-0.08	-0.48
3	2.74	1.05	0.22	-0.40
...	...	...	...	...
30	-1.42	2.12	-0.8	-0.62

Scatter Plot Matrix

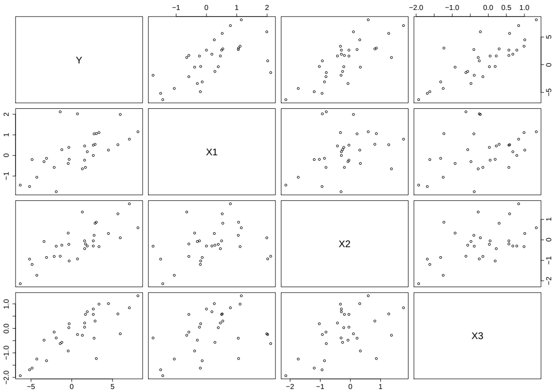
Figure: Scatter plots between response and predictors and among predictors



All variables appear to be correlated. No obvious nonlinearity.

Example: Scatter Plot Matrix

Figure: Scatter plots between response and predictors and among predictors



All variables appear to be positively correlated. No obvious nonlinearity.

Example: Model 1

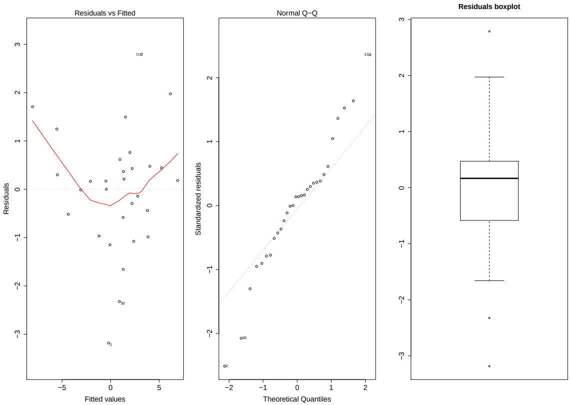
First-order model (only additive effects, a.k.a. *main effects*):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i, \quad i = 1, \cdots, 30.$$

R summary output:

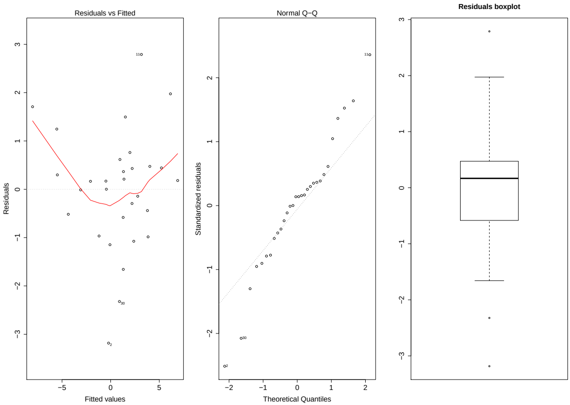
```
Call:
lm(formula = Y ~ X1 + X2 + X3, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1834 -0.5663  0.1673  0.4658  2.7901

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.2010     0.2541   4.727 6.91e-05 ***
X1            1.1107     0.2672   4.156 0.000311 ***
X2            1.7978     0.3287   5.469 9.78e-06 ***
X3            1.9596     0.3362   5.829 3.83e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.299 on 26 degrees of freedom
Multiple R-squared: 0.8883,     Adjusted R-squared: 0.8754
F-statistic: 68.93 on 3 and 26 DF,  p-value: 1.667e-12
```

Figure: Model 1: Residual Plots



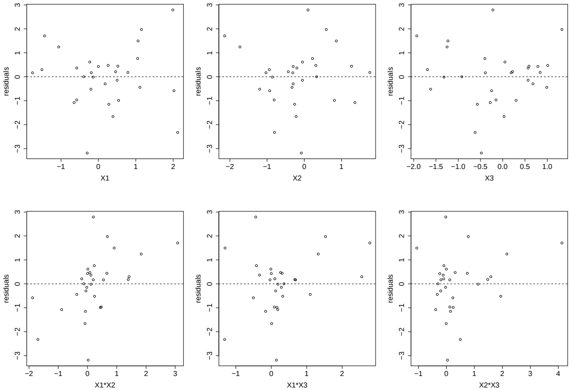
Residuals vs. fitted values plot shows . Residuals Q-Q plot shows . Residuals boxplot shows that most of residuals are in between 3, –3.

Figure: Model 1: Residual Plots



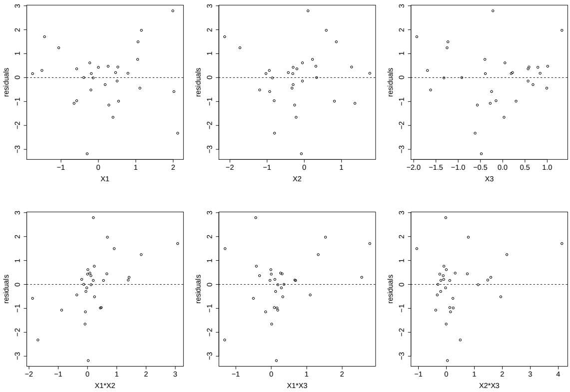
Residuals vs. fitted values plot shows nonlinearity. Residuals Q-Q plot shows heavy-tail. Residuals boxplot shows that most of residuals are in between 3, –3.

Figure: Model 1: Residuals vs. interaction term $X_1, X_2, X_3, X_1X_2, X_1X_3, X_2X_3$ Plots



Residuals vs. the interaction term shows a clear pattern. This term should in the model.

Figure: Model 1: Residuals vs. interaction term $X_1, X_2, X_3, X_1X_2, X_1X_3, X_2X_3$ Plots



Residuals vs. the interaction term X_1X_2 shows a clear linear pattern. This term should be included in the model.

Example: Model 2

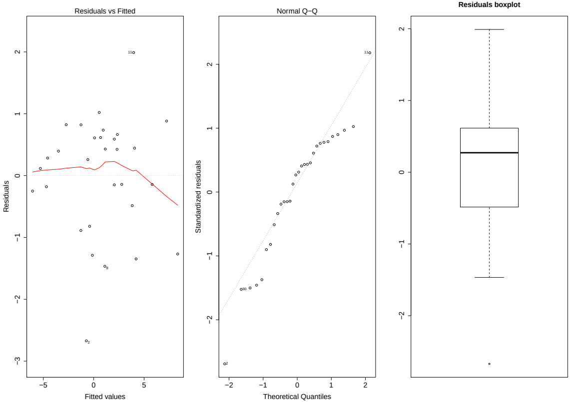
Nonadditive model with interaction between X_1 and X_2:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1} X_{i2} + \epsilon_i, \quad i = 1, \cdots, 30.$$

($p = 5$)

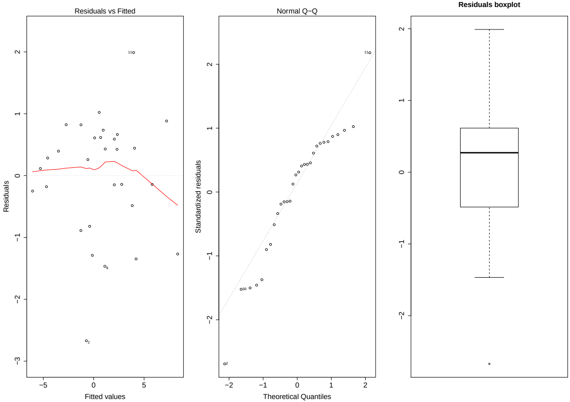
```
Call:
lm(formula = Y ~ X1 + X2 + X3 + X1:X2, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.6715 -0.4267  0.2715  0.6138  1.9901 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.8832     0.2153   4.103  0.00038 ***
X1            1.5946     0.2421   6.587 6.69e-07 ***
X2            1.7091     0.2605   6.560 7.16e-07 ***
X3            2.1266     0.2687   7.916 2.85e-08 ***
X1:X2         1.0076     0.2467   4.084  0.00040 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.026 on 25 degrees of freedom
Multiple R-squared: 0.933,     Adjusted R-squared: 0.9223 
F-statistic: 87.04 on 4 and 25 DF,  p-value: 2.681e-14
```

Figure: Model 2: Residual Plots



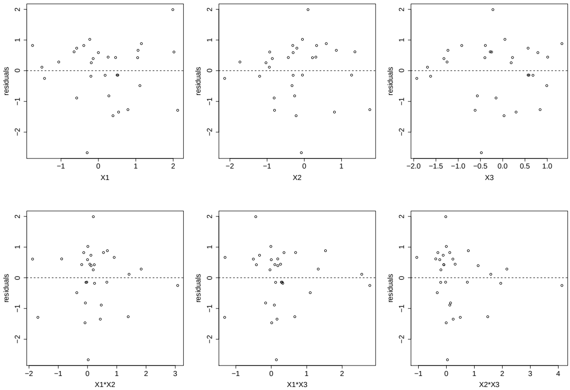
Residuals vs. fitted values plot shows
. Residuals Q-Q plot shows .
Residuals boxplot shows that most of residuals are in between
2, –2.

Figure: Model 2: Residual Plots



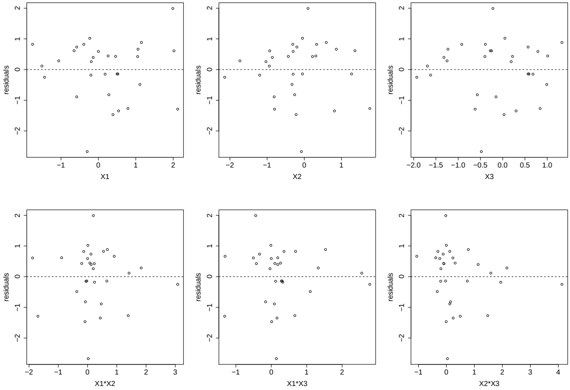
Residuals vs. fitted values plot shows no obvious nonlinearity. Residuals Q-Q plot shows no severe deviation from Normality. Residuals boxplot shows that most of residuals are in between 2, −2.

Figure: Model 2: Residuals vs. Each of $X_1, X_2, X_3, X_1X_2, X_1X_3, X_2X_3$ Plots



of these plots shows an obvious pattern. Model 2 seems to be .

Figure: Model 2: Residuals vs. Each of $X_1, X_2, X_3, X_1X_2, X_1X_3, X_2X_3$ Plots



None of these plots shows an obvious pattern. Model 2 seems to be adequate.

Example: Model 3

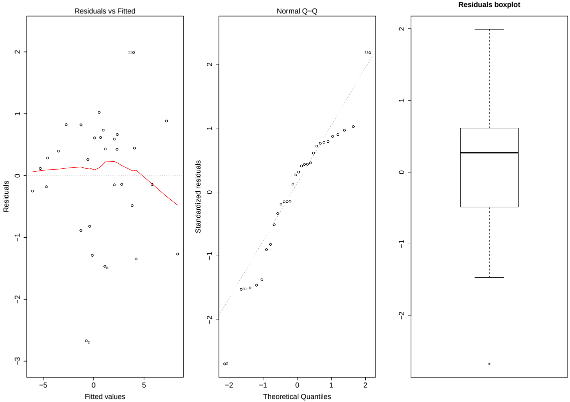
Nonadditive model with all three two-way interaction terms:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1} X_{i2} + \beta_5 X_{i1} X_{i3} + \beta_6 X_{i2} X_{i3} + \epsilon_i, \quad i = 1, \cdots, 30.$$

$(p = 7)$

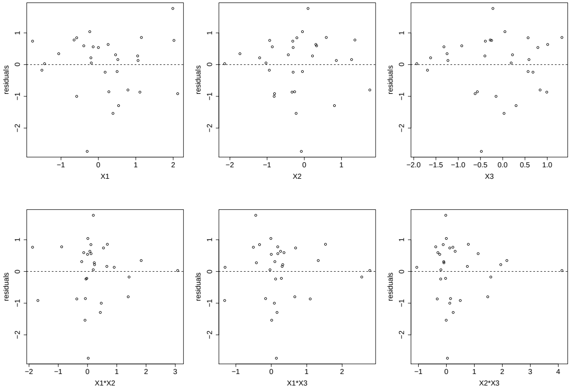
```
Call:
lm(formula = Y ~ X1 + X2 + X3 + X1:X2 + X1:X3 + X2:X3, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.7354 -0.6588  0.1868  0.6246  1.7705

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.8927     0.2278   3.920 0.000687 ***
X1            1.7179     0.2819   6.095 3.24e-06 ***
X2            1.5828     0.2925   5.411 1.69e-05 ***
X3            1.9982     0.3041   6.571 1.05e-06 ***
X1:X2         1.1925     0.3368   3.541 0.001744 **
X1:X3         0.2227     0.4009   0.555 0.583989
X2:X3        -0.4403     0.3675  -1.198 0.243074
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.038 on 23 degrees of freedom
Multiple R-squared: 0.937,      Adjusted R-squared: 0.9205
F-statistic: 56.99 on 6 and 23 DF,  p-value: 1.172e-12
```

Figure: Model 3: Residual Plots



Residuals vs. fitted values plot shows no obvious nonlinearity. Residuals Q-Q plot shows no severe deviation from Normality. Residuals boxplot shows that most of residuals are in between 2, −2.

Figure: Model 3: Residuals vs. Each of $X_1, X_2, X_3, X_1X_2, X_1X_3, X_2X_3$ Plots



None of these plots shows an obvious pattern. Model 3 seems to be adequate, but there is no obvious improvement over Model 2.

Analysis of Variance

Decomposition of total sum of squares:

- **Total sum of squares**:

$$SSTO = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \qquad , \quad d.f.(SSTO) =$$

- **Error sum of squares**:

$$SSE = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2 = \qquad , \quad d.f.(SSE) =$$

- **Regression sum of squares**:

$$SSR = \sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2 = \qquad , \quad d.f.(SSR) =$$

Analysis of Variance

$$\textbf{SSTO} = \textbf{SSE} + \textbf{SSR}, \quad \textbf{d.f.}(\textbf{SSTO}) = \textbf{d.f.}(\textbf{SSE}) + \textbf{d.f.}(\textbf{SSR}).$$

- **Total sum of squares**:

$$SSTO = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \mathbf{Y}'\left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right)\mathbf{Y}, \;\; d.f.(SSTO) = rank(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n) = n - 1.$$

- **Error sum of squares**:

$$SSE = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 = \mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}, \;\; d.f.(SSE) = rank(\mathbf{I}_n - \mathbf{H}) = n - p.$$

- **Regression sum of squares**:

$$SSR = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 = \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)\mathbf{Y}, \;\; d.f.(SSR) = rank(\mathbf{H} - \frac{1}{n}\mathbf{J}_n) = p - 1.$$

Sampling distributions of sums of squares (SS) under the Normal error model:

- *SSE* and *SSR* are .
  Notes: use the facts that $\mathbf{e}$ *are independent with* $\hat{\mathbf{Y}}$ *and* $\overline{Y}$*.*
  Why?
- $SSE \sim \sigma^2\chi^2_{(n-p)}$. *What is* $E(SSE)$*?*
- If $\beta_1 = \cdots = \beta_{p-1} = 0$, then $SSR \sim \sigma^2\chi^2_{(p-1)}$. *What is* $E(SSR)$ *in such a case?*

Sampling distributions of sums of squares (SS) under the Normal error model:

- SSE and SSR are independent.
  Notes: use the facts that $\mathbf{e}$ *are independent with* $\hat{\mathbf{Y}}$ *and* $\overline{Y}$.
  Why?
- $SSE \sim \sigma^2\chi^2_{(n-p)}$. *What is* $E(SSE)$*?*
- If $\beta_1 = \cdots = \beta_{p-1} = 0$, then $SSR \sim \sigma^2\chi^2_{(p-1)}$. *What is* $E(SSR)$ *in such a case?*

Mean squares (MS): **MS** $=$ **SS**/**d.f.**(**SS**).

- MSE (mean squared error):

$$MSE = \frac{SSE}{n-p}, \quad E(MSE) = \sigma^2.$$

MSE is an estimator of the error variance σ^2**.**

- MSR:

$$MSR = \frac{SSR}{p-1}.$$

$$E(MSR) = \begin{cases} \sigma^2 & \text{if} \quad \beta_1 = \cdots = \beta_{p-1} = 0 \\ & \text{if} \quad \text{otherwise} \end{cases}$$

- $MSTO = \frac{SSTO}{n-1}$.

Mean squares (MS): $\mathbf{MS} = \mathbf{SS}/\mathbf{d.f.}(\mathbf{SS})$.

- MSE (mean squared error):

$$MSE = \frac{SSE}{n-p}, \quad E(MSE) = \sigma^2.$$

MSE is an unbiased estimator of the error variance σ^2.

- MSR:

$$MSR = \frac{SSR}{p-1}.$$

$$E(MSR) = \begin{cases} \sigma^2 & \text{if} \quad \beta_1 = \cdots = \beta_{p-1} = 0 \\ > \sigma^2 & \text{if} \quad \textit{otherwise} \end{cases}$$

- $MSTO = \frac{SSTO}{n-1}$.

F Test of Regression Relation

Under the Normal error model:

- Test **whether there is a** **between the response variable** Y **and the set of** X **variables**:

- F ratio and its null distribution:

$$F^* = \qquad\qquad , \quad F^* \sim_{H_0} \qquad\qquad ,$$

  where $F_{p-1,n-p}$ denotes the F distribution with $(p-1, n-p)$ degrees of freedom.

- Decision rule at level α: reject H_0 if $F^* >$.

F Test of Regression Relation

Under the Normal error model

- Test **whether there is a regression relation between the response variable Y and the set of X variables**:

$$H_0 : \beta_1 = \cdots = \beta_{p-1} = 0 \ \ vs.$$

$$H_a : \text{not all } \beta_k\text{s equal zero.}$$

- F ratio and its null distribution:

$$F^* = \frac{MSR}{MSE}, \quad F^* \sim_{H_0} F_{p-1,n-p},$$

  where $F_{p-1,n-p}$ denotes the F distribution with $(p-1, n-p)$ degrees of freedom.

- Decision rule at level α: reject H_0 if $F^* > F(1-\alpha; p-1, n-p)$.

ANOVA Table

Source of Variation	SS	d.f.	MS	F^*
Regression	$SSR = \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J}_n)\mathbf{Y}$	$p-1$	$MSR = \frac{SSR}{p-1}$	$F^* = \frac{MSR}{MSE}$
Error	$SSE = \mathbf{Y}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}$	$n-p$	$MSE = \frac{SSE}{n-p}$	
Total	$SSTO = \mathbf{Y}'\left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right)\mathbf{Y}$	$n-1$		

Example Model 2: $n = 30, p = 5$.

Source of Variation	SS	d.f.	MS	F^*
Regression	$SSR = 366.4846$	4	$MSR = 91.62116$	$F^* = 87.03703$
Error	$SSE = 26.31672$	25	$MSE = 1.052669$	
Total	$SSTO = 392.8013$	29		

Pvalue $= P(F_{4,25} > 87.037) \approx 0$, so there is a significant regression relation between Y and X_1, X_2, X_3, X_1X_2.

Coefficient of Multiple Determination

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

- R^2 is the of the total variation in Y by using the X variables to explain Y.
- $0 \leq R^2 \leq 1$.
  When $R^2 = 0$? When $R^2 = 1$?
- **Adding more X variables to the model will always** R^2 because:
  (i) $SSTO$.
  (ii) SSE .

Coefficient of Multiple Determination

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

- R^2 is the proportional reduction of the total variation in Y by using the X variables to explain Y.
- $0 \leq R^2 \leq 1$.
  When $R^2 = 0$? When $R^2 = 1$?
- **Adding more X variables to the model will always increase** R^2 because:
  (i) $SSTO$ remains the same.
  (ii) SSE becomes smaller.

Coefficient of multiple correlation:

$$R := \sqrt{R^2}.$$

- When there is only one X variable, R equals to the absolute value of the (sample) correlation coefficient r between X and Y.
- In general, R is the maximum absolute (sample) correlation coefficient between Y and linear combinations of $X_1, \cdots, X_{p-1}$:

$$R = \max_{c_1, \cdots, c_{p-1}} |Corr(Y, c_1 X_1 + \cdots + c_{p-1} X_{p-1})|.$$

Since adding more X variables can only R^2, does this mean we should use as many X variables as possible?

- With more X variables, the model fits the observed data due to SSE.
- However, a model with many X variables that are unrelated to the response variable and/or are highly correlated with each other tends to
  - the observed data and often do a job for prediction due to sampling variability.
  - make interpretation .
  - make prediction more .

Since adding more X variables can only increase R^2, does this mean we should use as many X variables as possible?

- With more X variables, the model fits the observed data better due to smaller SSE.
- However, a model with many X variables that are unrelated to the response variable and/or are highly correlated with each other tends to
  - **overfit** the observed data and often do a poor job for prediction due to increased sampling variability.
  - make interpretation difficult.
  - make prediction more costly.

Adjusted Coefficient of Multiple Determination

Adjust for of X variables in the model:

- R_a^2 R^2.
- R_a^2 **may become when adding more** X **variables into the model** because:
  - the in SSE may be more than offset by the

    in SSE.

Adjusted Coefficient of Multiple Determination

Adjust for the number of X variables in the model:

$$R_a^2 = 1 - \frac{MSE}{MSTO} = 1 - \frac{n-1}{n-p}\frac{SSE}{SSTO}.$$

- $R_a^2 \leq R^2$.
- R_a^2 **may become smaller when adding more X variables into the model** because:
  - the decrease in SSE may be more than offset by the loss of degrees of freedom in SSE.

Example

- Model 1: $Y \sim X_1, X_2, X_3$

$$R^2 = 0.8883, \quad R_a^2 = 0.8754$$

- Model 2 : $Y \sim X_1, X_2, X_3, X_1X_2$

$$R^2 = 0.933, \quad R_a^2 = 0.9223.$$

- Model 3: $Y \sim X_1, X_2, X_3, X_1X_2, X_1X_3, X_2X_3$.

$$R^2 = 0.937, \quad R_a^2 = 0.9205.$$

(i) For each model, $R^2 > R_a^2$; (ii) Adding more X variable(s) increases R^2. The increase of R^2 is much more from Model 1 to Model 2 than from Model 2 to Model 3; (iii) Model 3 has a smaller R_a^2 than Model 2.

Inferences about Regression Coefficients

LS estimators:

$$\underset{p\times 1}{\hat{\boldsymbol{\beta}}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{p-1} \end{bmatrix} =$$

$$\underset{p\times 1}{\mathbf{E}\{\hat{\boldsymbol{\beta}}\}} = \qquad\qquad , \quad \underset{p\times p}{\boldsymbol{\sigma}^{\mathbf{2}}\{\hat{\boldsymbol{\beta}}\}} = \qquad\qquad .$$

The standard error of $\hat{\beta}_k$, $s(\hat{\beta}_k)$, is the

of $MSE(\mathbf{X}'\mathbf{X})^{-1}$.

Inferences about Regression Coefficients

LS estimators:

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

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

The standard error of $\hat{\beta}_k$, $s(\hat{\beta}_k)$, is the positive square-root of the $(k+1)th$ diagonal element of $MSE(\mathbf{X}'\mathbf{X})^{-1}$.

- Studentized quantity:

$$\frac{\hat{\beta}_k - \beta_k}{s\{\hat{\beta}_k\}} \sim \qquad\qquad\qquad .$$

- $(1-\alpha)$-Confidence interval for β_k:

- Two-sided T-Test: $H_0 : \beta_k = \beta_k^0 \;\; vs. \;\; H_a : \beta_k \neq \beta_k^0$.
- T statistic:

$$T^* = \qquad\qquad\qquad .$$

At level α, the decision rule is to reject H_0 if and only if $|T^*|$.

- Studentized quantity:

$$\frac{\hat{\beta}_k - \beta_k}{s\{\hat{\beta}_k\}} \sim t_{(n-p)}.$$

- $(1-\alpha)$-Confidence interval for β_k:

$$\hat{\beta}_k \pm t(1-\alpha/2; (n-p))s\{\hat{\beta}_k\}.$$

- Two-sided T-Test: $H_0 : \beta_k = \beta_k^0 \;\; vs. \;\; H_a : \beta_k \neq \beta_k^0$.
- T statistic:

$$T^* = \frac{\hat{\beta}_k - \beta_k^0}{s\{\hat{\beta}_k\}} \underset{H_0}{\sim} t_{(n-p)}.$$

At level α, the decision rule is to reject H_0 if and only if $|T^*| > t(1-\alpha/2; (n-p))$.

Example: Model 2

Nonadditive model with interaction between X_1 and X_2:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1} X_{i2} + \epsilon_i, \quad i = 1, \cdots, 30.$$

$(p = 5)$

```
Call:
lm(formula = Y ~ X1 + X2 + X3 + X1:X2, data = data)

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.8832     0.2153    4.103   0.00038 ***
X1            1.5946     0.2421    6.587 6.69e-07 ***
X2            1.7091     0.2605    6.560 7.16e-07 ***
X3            2.1266     0.2687    7.916 2.85e-08 ***
X1:X2         1.0076     0.2467    4.084  0.00040 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.026 on 25 degrees of freedom
Multiple R-squared: 0.933,      Adjusted R-squared: 0.9223
F-statistic: 87.04 on 4 and 25 DF,  p-value: 2.681e-14
```

◂ Model 3

Test whether there is an interaction between X_1 and X_2. Use $\alpha = 0.01$.

- H_0 : , *vs.*, H_a : .
- $T^* =$
- $n = 30, p = 5,$.
- Since , the null hypothesis and conclude that there is interaction effect between X_1 and X_2.
- Alternatively, pvalue= , so H_0.

Notes: pvalue for the two-sided alternative is in the R output.

Test whether there is an interaction between X_1 and X_2. Use $\alpha = 0.01$.

- $H_0 : \beta_4 = 0, \quad vs., \quad H_a : \beta_4 \neq 0.$
- $T^* = \frac{1.0076-0}{0.2467} = 4.084.$
- $n = 30, p = 5, t(0.995; 25) = 2.787.$
- Since $|4.084| > 2.787$, reject the null hypothesis and conclude that there is a significant interaction effect between X_1 and X_2.
- Alternatively, pvalue=$P(|t_{(25)}| > |4.084|) = 0.00040 < 0.01$, so reject H_0.

Notes: pvalue for the two-sided alternative is in the R output.

Estimation of the Mean Response

- For a given set of values of the X variables:

$$\underset{p\times 1}{\mathbf{X}_h} = \begin{bmatrix} 1 \\ X_{h1} \\ \vdots \\ X_{h,p-1} \end{bmatrix}$$

- Corresponding mean response:

$$E(Y_h) =$$

Estimation of the Mean Response

- For a given set of values of the X variables:

$$\underset{p\times 1}{\mathbf{X}_h} = \begin{bmatrix} 1 \\ X_{h1} \\ \vdots \\ X_{h,p-1} \end{bmatrix}$$

- Corresponding mean response:

$$E(Y_h) = \mathbf{X}_h'\boldsymbol{\beta} = \beta_0 + \beta_1 X_{h1} + \cdots + \beta_{p-1} X_{h,p-1}.$$

- $\widehat{Y}_h :=$ is an estimator of $E(Y_h)$:

$$E(\widehat{Y}_h) = \qquad\qquad .$$

$$\sigma^2(\widehat{Y}_h) = \qquad\qquad .$$

- Standard error of $\widehat{Y}_h$:

$$s(\widehat{Y}_h) = \qquad\qquad .$$

- $(1 - \alpha)$-confidence interval for $E(Y_h)$:

- $\widehat{Y}_h := \mathbf{X}_h'\hat{\beta}$ is an unbiased estimator of $E(Y_h)$:

$$E(\widehat{Y}_h) = E(\mathbf{X}_h'\hat{\beta}) = \mathbf{X}_h'\mathbf{E}\{\hat{\beta}\} = \mathbf{X}_h'\beta = E(Y_h).$$

$$\sigma^2(\widehat{Y}_h) = \mathbf{X}_h'\boldsymbol{\sigma}^{\mathbf{2}}\{\hat{\beta}\}\mathbf{X}_h = \sigma^2\left(\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h\right).$$

- Standard error of $\widehat{Y}_h$:

$$s(\widehat{Y}_h) = \sqrt{MSE\left(\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h\right).}$$

- $(1-\alpha)$-confidence interval for $E(Y_h)$:

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

Prediction of a New Observation

- $Y_{h(new)} = \mathbf{X}_h'\boldsymbol{\beta} + \epsilon_h$: with the observations Y_is.
- Predicted value: $\widehat{Y}_h :=$.

$\sigma^2(pred) :=$.

- Standard error for prediction:

$s(pred) =$.

- $(1 - \alpha)$-prediction interval for $Y_{h(new)}$:

Prediction of a New Observation

- $Y_{h(new)} = \mathbf{X}_h' \boldsymbol{\beta} + \epsilon_h$: independent with the observations Y_is.
- Predicted value: $\widehat{Y}_h := \mathbf{X}_h' \hat{\boldsymbol{\beta}}$

$$\sigma^2(pred) := Var(\widehat{Y}_h - Y_{h(new)}) = \sigma^2(\widehat{Y}_h) + \sigma^2(Y_{h(new)}) = \sigma^2 \mathbf{X}_h' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_h + \sigma^2.$$

- Standard error for prediction:

$$s(pred) = \sqrt{MSE \left[1 + \mathbf{X}_h' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_h \right]}.$$

- $(1 - \alpha)$-prediction interval for $Y_{h(new)}$:

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

Example

Estimate the mean response when $X_1 = 0.8, X_2 = 0.5, X_3 = -1$ under Model 2.

- $\mathbf{X}_h' =$.
- $n = 30, p = 5$:

$$\widehat{Y}_h := \mathbf{X}_h'\hat{\beta} = 1.290,$$

$$\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h = 0.170, \quad MSE = 1.053,$$

$$s(\widehat{Y}_h) = \qquad .$$

- A 99%-confidence interval for $E(Y_h)$: $t(0.995; 25) = 2.787$

$$1.290 \pm 2.787 \times 0.423 = [0.111, 2.469].$$

Example

Estimate the mean response when $X_1 = 0.8, X_2 = 0.5, X_3 = -1$ under Model 2.

- $\mathbf{X}_h' = \begin{bmatrix} 1 & 0.8 & 0.5 & -1 & 0.4 \end{bmatrix}$
- $n = 30, p = 5$:

$$\widehat{Y}_h := \mathbf{X}_h'\hat{\beta} = 1.290,$$

$$\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h = 0.170, \quad MSE = 1.053,$$

$$s(\widehat{Y}_h) = \sqrt{1.053 \times 0.170} = 0.423.$$

- A 99%-confidence interval for $E(Y_h)$: $t(0.995; 25) = 2.787$

$$1.290 \pm 2.787 \times 0.423 = [0.111, 2.469].$$

Predict a new observation when $X_1 = 0.8, X_2 = 0.5, X_3 = -1$ under Model 2.

- Standard error for prediction:

$$s(pred) = \qquad\qquad\qquad\qquad\qquad\qquad .$$

- A 99%-prediction interval for Y_{hnew}:

$$1.290 \pm 2.787 \times 1.1098 = [-1.803, 4.383].$$

- R codes.

```
> newX=data.frame(X1=0.8, X2=0.5, X3=-1)
> predict.lm(fit2, newX, interval="confidence",
+ level=0.99, se.fit=TRUE)

> predict.lm(fit2, newX, interval="prediction",
+ level=0.99, se.fit=TRUE)
```

Predict a new observation when $X_1 = 0.8, X_2 = 0.5, X_3 = -1$ under Model 2.

- Standard error for prediction:

$$s(pred) = \sqrt{1.053 \times (1 + 0.170)} = 1.1098.$$

- A 99%-prediction interval for Y_{hnew}:

$$1.290 \pm 2.787 \times 1.1098 = [-1.803, 4.383].$$

- R codes.

```
> newX=data.frame(X1=0.8, X2=0.5, X3=-1)
> predict.lm(fit2, newX, interval="confidence",
+ level=0.99, se.fit=TRUE)

> predict.lm(fit2, newX, interval="prediction",
+ level=0.99, se.fit=TRUE)
```

Example: Model 3

Nonadditive model with all three second-order interaction terms:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1} X_{i2} + \beta_5 X_{i1} X_{i3} + \beta_6 X_{i2} X_{i3} + \epsilon_i.$$

```
Call:
lm(formula = Y ~ X1 + X2 + X3 + X1:X2 + X1:X3 + X2:X3, data = data)

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.8927     0.2278   3.920 0.000687 ***
X1            1.7179     0.2819   6.095 3.24e-06 ***
X2            1.5828     0.2925   5.411 1.69e-05 ***
X3            1.9982     0.3041   6.571 1.05e-06 ***
X1:X2         1.1925     0.3368   3.541 0.001744 **
X1:X3         0.2227     0.4009   0.555 0.583989
X2:X3        -0.4403     0.3675  -1.198 0.243074
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 1.038 on 23 degrees of freedom
Multiple R-squared: 0.937,      Adjusted R-squared: 0.9205
F-statistic: 56.99 on 6 and 23 DF,  p-value: 1.172e-12
```

Model 2

Compare Model 2 and Model 3.

- SSE, R^2, R_a^2 are .
- In Model 3, the interaction terms X_1X_3 and X_2X_3 are .
- SEs are in Model 3, i.e., sampling variability, due to .
- of d.f. in Model 3 due to X variables $\Longrightarrow$ multipliers (critical values) are .
- Consequently, confidence intervals are under Model 3, i.e., precise.

Compare Model 2 and Model 3.

- SSE, R^2, R_a^2 are similar.
- In Model 3, the interaction terms X_1X_3 and X_2X_3 are not significant.
- SEs are larger in Model 3, i.e., increased sampling variability, due to multicollinearity.
- Loss of d.f. in Model 3 due to more X variables $\Longrightarrow$ multipliers (critical values) are bigger.
- Consequently, confidence intervals are wider under Model 3, i.e., less precise.