Linear Regression

Professor Jie Peng, PhD

Department of Statistics

University of California, Davis

Multiple Regression Model: Overview

Motivation

Often a number of variables affect the response variable in important and distinctive ways such that any single one wouldn't have provided an adequate description. E.g.,

- 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.

Multiple Regression Model

if $p-1=1$
↳ simple regression

$$Y_i = \overbrace{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1}}^{\text{fixed}} + \overset{\text{random}}{\epsilon_i}, \quad i = 1, \cdots n$$

- Y_i : value of the response variable in the *ith* case
- $X_{i1}, \cdots, X_{i,p-1}$: values of the X variables in the *ith* case
- $\beta_0, \beta_1, \cdots, \beta_{p-1}$: regression coefficients
- ϵ_i: random errors

$$E(\epsilon_i) = 0, \quad Var(\epsilon_i) = \sigma^2, \quad Cov(\epsilon_i, \epsilon_j) = 0 \text{ for } i \neq j$$

- *Response function (surface)/ mean response*:

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

First-Order (Additive) Models

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

- 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).**

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**.

- How education level affects income may depend on gender.
- Interactions are often represented by cross product terms among predictors.

Non-additive Model with Two Predictors

interaction

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

income gender years of edu.

- This model is in the form of the multiple regression 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 $\beta_0, \beta_1, \beta_2$, β_3 but is not linear in the original predictors X_1, X_2. $\because X_1 X_2$

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
...	...	...

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$

Design matrix of a first-order model: $\underset{n\times 3}{\mathbf{X}} = \begin{array}{c} \begin{array}{ccc} & x_1 & x_2 \end{array} \\ \begin{bmatrix} 1 & 4.0 & 2.0 \\ 1 & 4.0 & 4.0 \\ 1 & 4.0 & 2.0 \\ 1 & 4.0 & 4.0 \\ \vdots & \vdots & \vdots \end{bmatrix} \end{array}$

$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i} \cdot$ — — —

Design matrix of a non-additive model: $\underset{n\times 4}{\mathbf{X}} = \begin{bmatrix} 1 & 4.0 & 2.0 & 8.0 \\ 1 & 4.0 & 4.0 & 16.0 \\ 1 & 4.0 & 2.0 & 8.0 \\ 1 & 4.0 & 4.0 & 16.0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$

Polynomial Regression Models

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

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

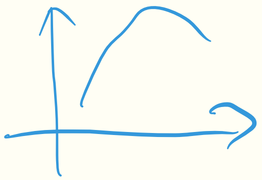
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i, \quad i = 1, \cdots, n.$$

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

$p = 3$ $\quad$ p: # of regression

Example

	y	x
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:

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

Models with Transformed Variables

- 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 multiple regression model by defining $Y_i^* := \log Y_i$.

What Makes a 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 **original predictors**.

- In contrasts, **nonlinear regression models** are nonlinear in the parameters.
- The model below can not be expressed in the form of a linear regression model through transformations or introducing new X variables:

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

Multiple Regression: Example

Data

30 cases, one response variable and three predictor variables:

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

- First examine each variable marginally: Variable type, summary statistics, histogram, boxplot, pie chart, missing values? outliers? etc.
- Then explore their relationships through pairwise scatter plots.

Scatter Plot Matrix

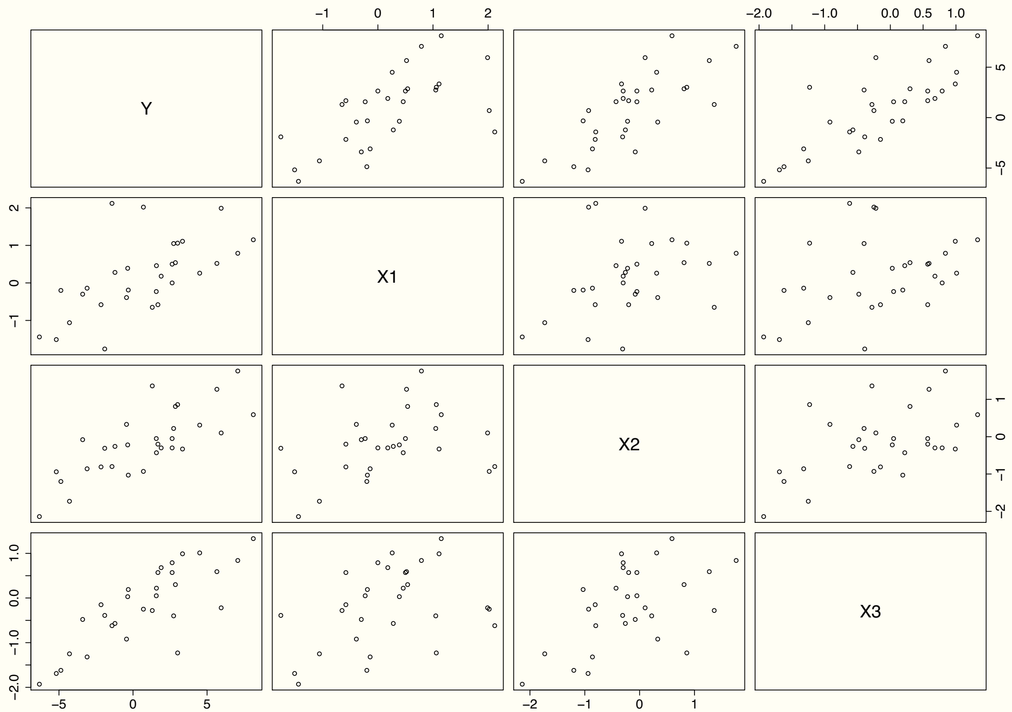


Figure: Pairwise scatter plots

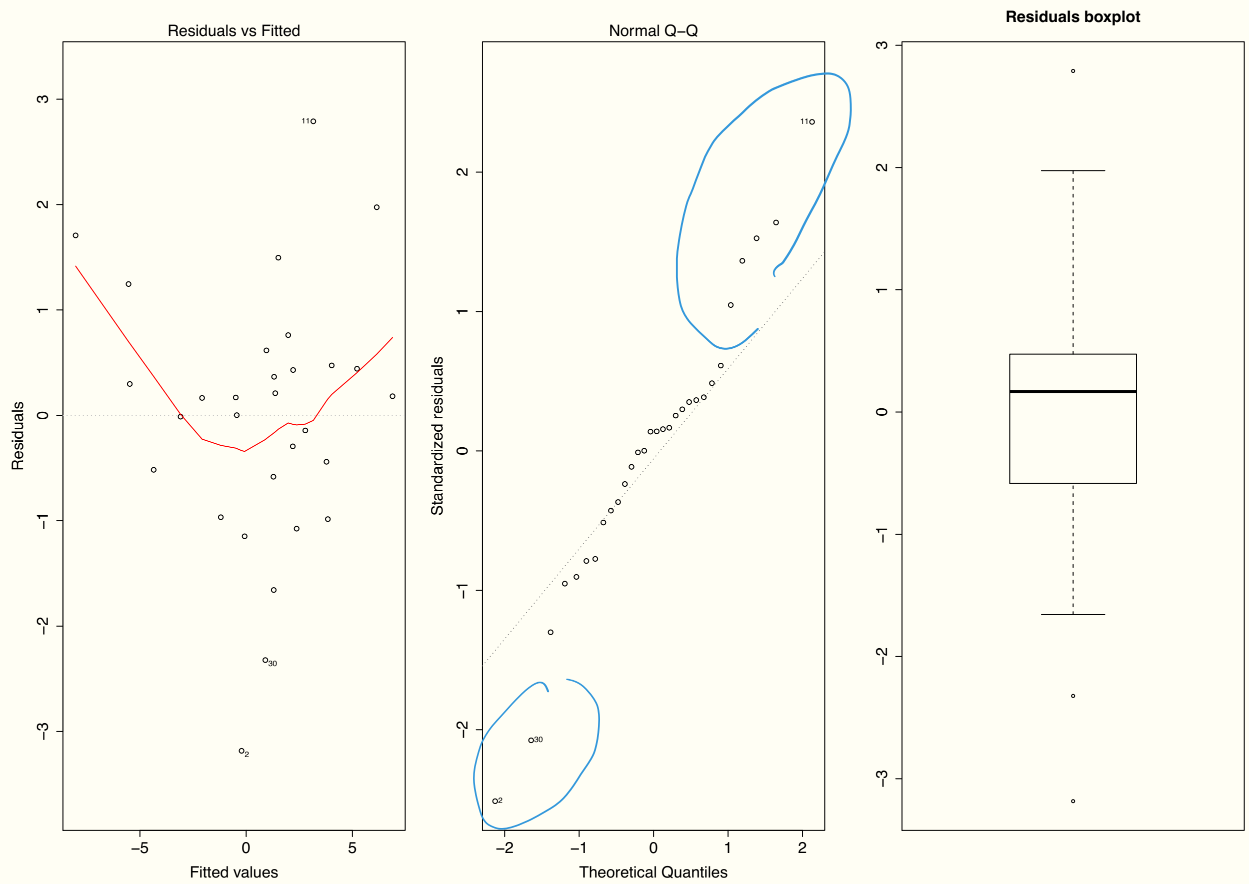
All variables appear to be positively related. No obvious nonlinearity.

Model 1: First-order Model

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

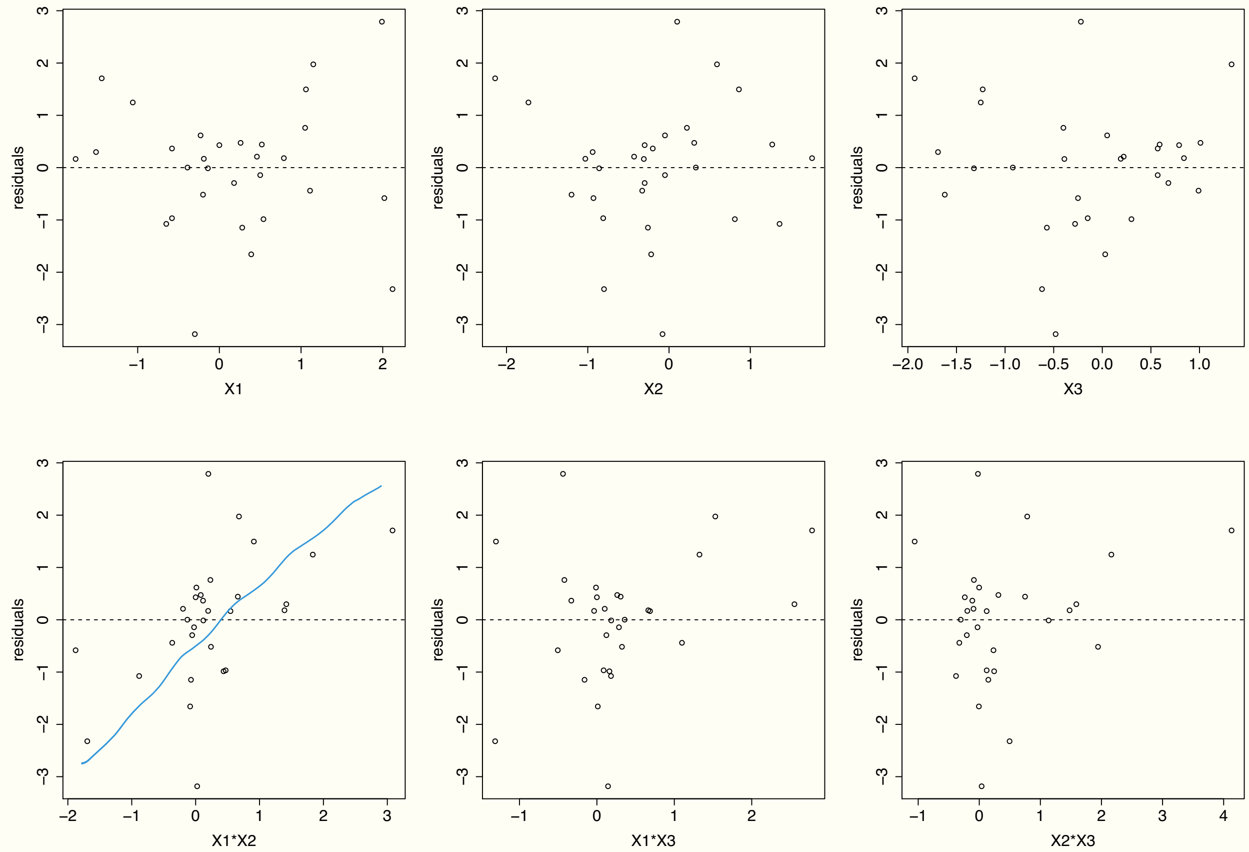
```
Call:
lm(formula = Y ~ X1 + X2 + X3, data = data)

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 ***
---

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



Residual vs. fitted value plot shows nonlinearity. Residual Q-Q plot shows heavy-tail. Residual boxplot shows range from −3 to 3.

Figure: Model 1: residual vs. interaction terms



Others: pattern unclear

Residual vs. $X_1 X_2$ shows a clear linear pattern → this term should be included in the model.

we should not see pattern in resid. if ✓ (should include)

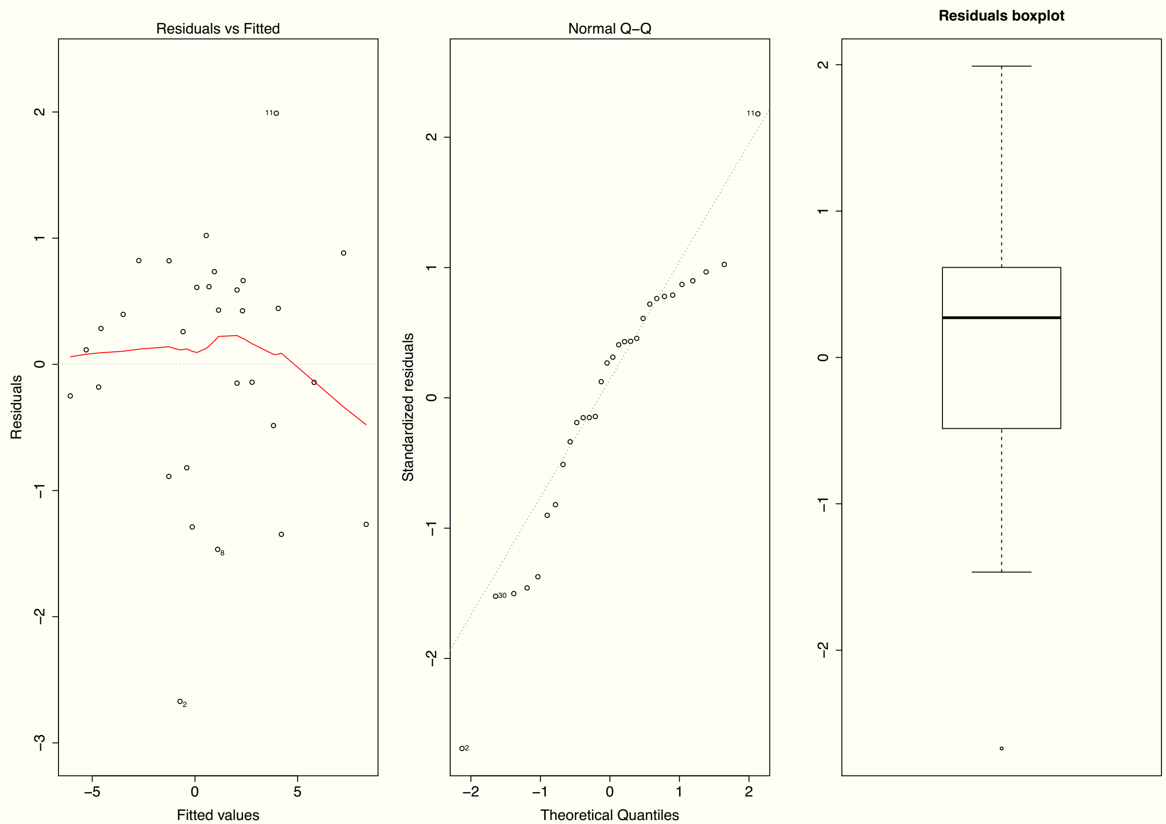
Model 2: Nonadditive Model with Interaction $X_1 X_2$

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

interaction

```
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 ***
---

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

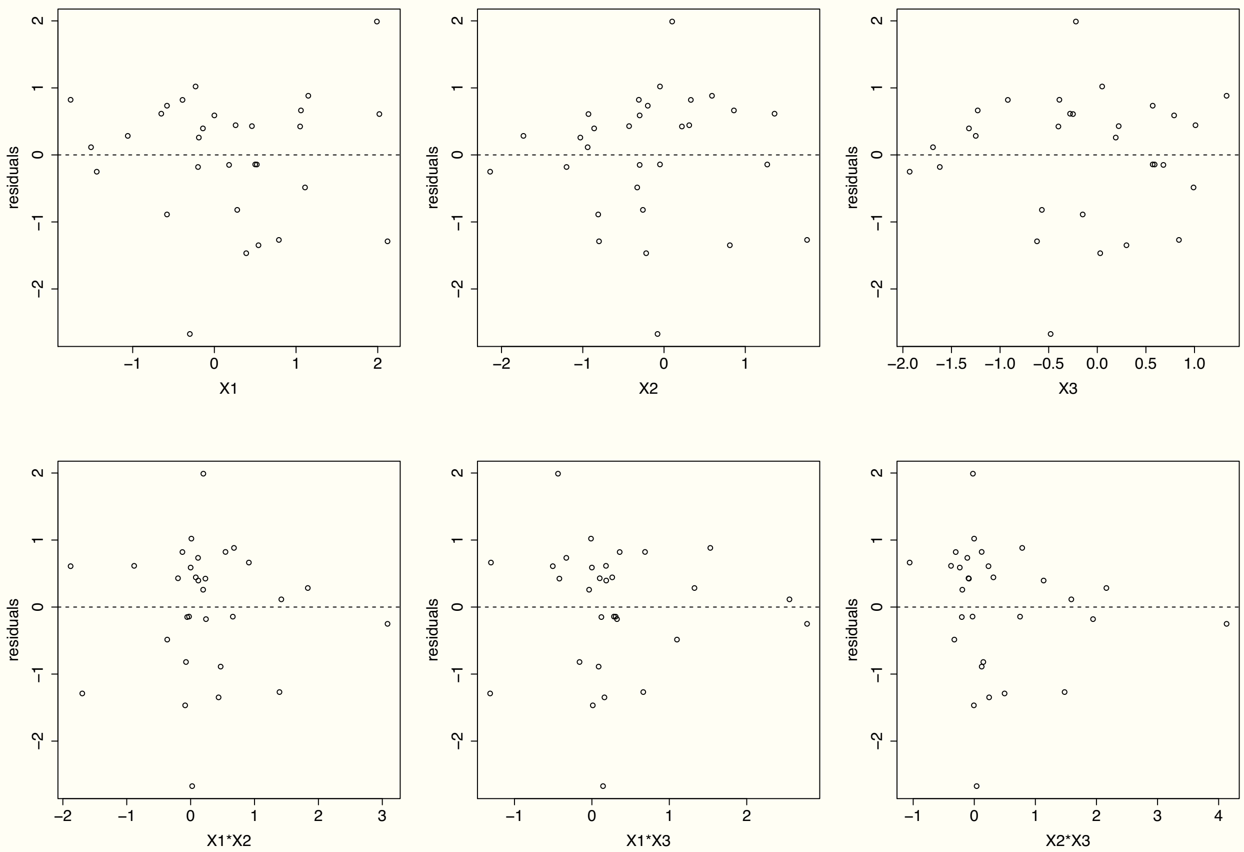


Residual vs. fitted value plot shows no obvious nonlinearity.

Residual Q-Q plot shows no severe deviation from Normality.

Residual boxplot shows range from −2 to 2. improved model

Figure: Model 2: residual vs. interaction terms



None of these plots shows an obvious pattern → Model 2 appears adequate.

3-way: $X_1 X_2 X_3$

Model 3: Nonadditive Model with All Two-way Interactions

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

lose statistical efficiency

```
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
---
```

Model 2: df = 25

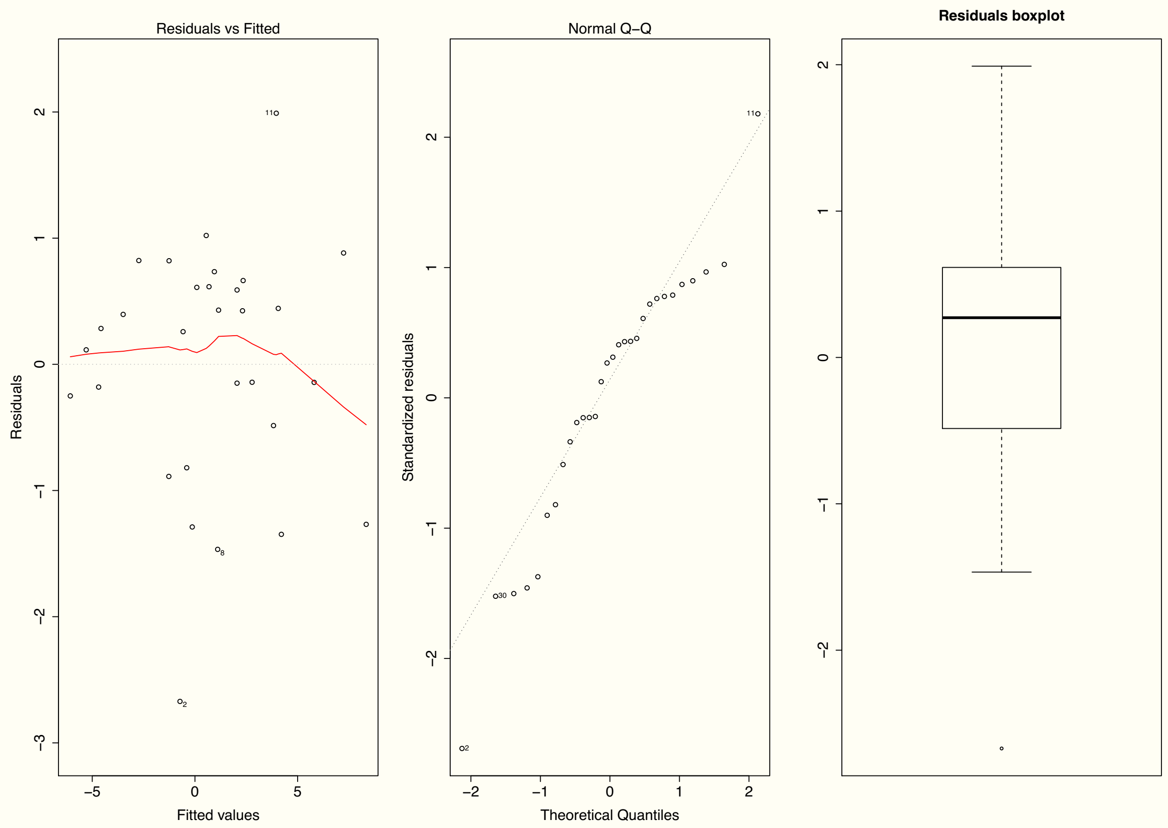
lose df!

```
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
```

we will inflate SE.

Model 2 F larger, p smaller

Figure: Model 3: residual plots

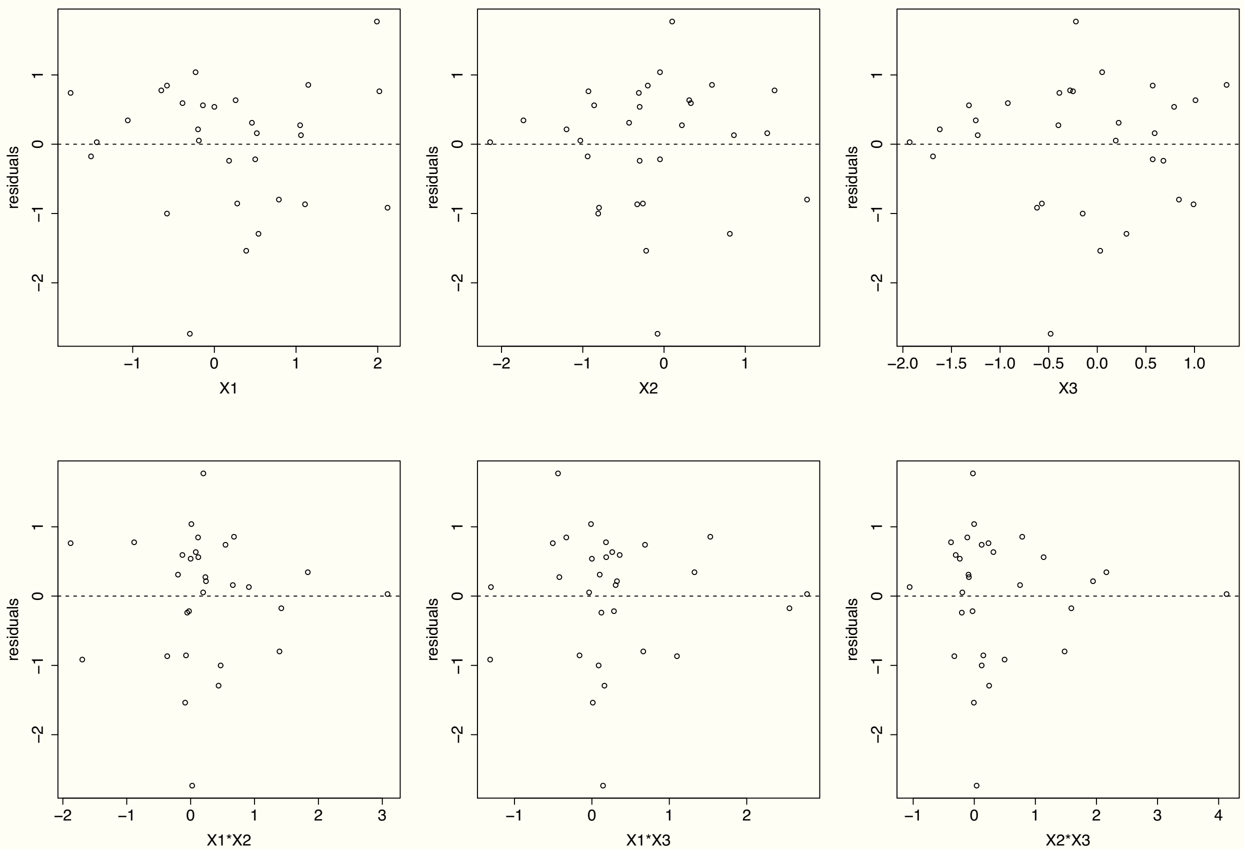


Residual vs. fitted value plot shows no obvious nonlinearity.

Residual Q-Q plot shows no severe deviation from Normality.

Residual boxplot shows range from -2 to 2.

Figure: Model 3: residual vs. interaction terms



None of these plots shows an obvious pattern → Model 3 appears adequate, but there is also no obvious improvement over Model 2.

Multiple Regression:

Matrix Form

Model Equations

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

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

Each row of **X** corresponds to a case and each column of X corresponds to an X variable.

Model Assumptions

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

- In terms of the observations:

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

- Under the Normal error model, $\boldsymbol{\epsilon}$ and $\mathbf{Y}$ are vectors of independent normal random variables.

Least Squares Estimators

SSE (new) = min Q — — — ??

10:46

- Least squares criterion:

boundary df.

$$Q(\mathbf{b}) = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{i1} - \cdots - b_{p-1} X_{i,p-1})^2$$

$$= (\underset{n\times 1}{\mathbf{Y}} - \underset{n\times p}{\mathbf{X}} \overset{p\times 1}{\vec{b}})' (\mathbf{Y} - \mathbf{X}b), \quad \underset{p\times 1}{\mathbf{b}} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix}.$$

with respect to $\vec{b}$

- 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:

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

$\underset{n\times p}{X}$: rank $(X) \le \min(n, p)$

[given $X^T X$ is invertible]

from now on

i.e. rank $(X^T) \le p$

$\Rightarrow$ $\boxed{p \le n}$

- $\hat{\boldsymbol{\beta}}$ is an unbiased estimator for $\boldsymbol{\beta}$: $E(y) = X\beta$

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

- Variance-covariance matrix of $\hat{\boldsymbol{\beta}}$: $\sigma^2\{y\} = \sigma^2 \cdot I_n$

$= \underbrace{(X^T X)^{-1} \cdot X^T}\, \sigma^2 \cdot I_n \cdot X \cdot (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}$

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

Fitted Values and Residuals

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

$$\mathbf{E}\{\widehat{\mathbf{Y}}\} = \overset{\substack{\text{mean representation} \\ \downarrow}}{\mathbf{X}\beta} = \mathbf{E}\{\mathbf{Y}\}, \quad \boldsymbol{\sigma}^2\{\widehat{\mathbf{Y}}\} = \sigma^2\mathbf{H}.$$

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

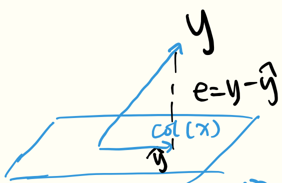
- Linear transformations of the observations vector **Y**
- Under the Normal error model, they are normally distributed

Hat Matrix

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

$\mathrm{rank}(H) = \mathrm{Tr}(H)$
$= \mathrm{Tr}\left(X(X^TX)^{-1}X^T\right)$
$= \mathrm{Tr}\left((X^TX)^{-1}X^TX\right) = \mathrm{Tr}(I_p)$
$= p$
$[\mathrm{rank}(X)]$

- **H** and $\mathbf{I}_n - \mathbf{H}$ are projection matrices: symmetric and idempotent.



p dimensional linear subset of $\mathbb{R}$

- $rank(\mathbf{H}) = p, \quad rank(\mathbf{I}_n - \mathbf{H}) = n - p.$
- **H** is the projection matrix to $\mathrm{col}(\mathrm{X})$:
  - Fitted value vector $\widehat{\mathbf{Y}} = \mathbf{HY}$ is the projection of the observations vector **Y** to $\mathrm{col}(\mathrm{X})$.
  - Residual vector $\mathbf{e} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y}$ is orthogonal to $\mathrm{col}(\mathrm{X})$.

Multiple Regression:

ANOVA

only df diff.

Decomposition of Total Variation

$$SSTO = SSE + SSR$$

- **Total sum of squares**:

Jn : n×n of 1

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

- **Error sum of squares**:

$= \Sigma e_i^2$

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

- **Regression sum of squares**:

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

$= Tr(H - \frac{1}{n} J_n)$

$= Tr(H) - Tr(\frac{1}{n} J_n) = p - 1$

Sampling Distributions of Sums of Squares

under H_0:

$$\frac{SSE}{\sigma^2} \sim \chi^2_{(n-p)}$$

indep.

$$\frac{SSR}{\sigma^2} \sim \chi^2_{(p-1)}$$

Under the Normal error model:

rescaled χ^2 distr.

- $SSE \sim \sigma^2 \chi^2_{(n-p)}$
- SSE and SSR are independent.
- If $\underbrace{\beta_1 = \cdots = \beta_{p-1} = 0}$, then $SSR \sim \sigma^2 \chi^2_{(p-1)}$.

i.e. there's no regression relation between y and $(x_1, x_2 \cdots x_{p-1})$

$$\frac{\frac{SSR}{\sigma^2 (p-1)}}{\frac{SSE}{\sigma^2 (n-p)}} = \frac{MSR}{MSE} \sim F_{p-1, n-p}$$

Mean Squares

$\xrightarrow{} \text{rank}(X)$

$E(SSE) = (n-p)\sigma^2$ use Tr

- MSE: an unbiased estimator of the error variance σ^2

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

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

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

F Test for Regression Relation

Test whether the response variable and the set of X variables are related:

no regression relation

- $H_0 : \beta_1 = \cdots = \beta_{p-1} = 0$ vs. H_a: not all β_ks equal to zero
- F ratio and its null distribution:

null (reference) 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	$SSTO = \mathbf{Y}'\left(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\right)\mathbf{Y}$	$n-1$		

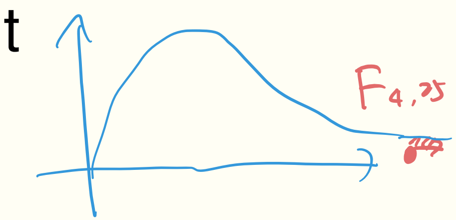
Example: Model 2

$n=30$

$p=5$

Source of Variation	SS	d.f.	MS	F^*
Regression	$SSR = 366.4846$	4	$MSR = 91.62116$	$F^* = 87.03703$ large!
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 $\underbrace{X_1, X_2, X_3, X_1X_2}_{p-1=4 \Rightarrow p=5}$.



Multiple Regression: Coefficient of Multiple Determination

Coefficient of Multiple Determination

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

- R^2 is the proportion of total variation in Y that may be explained by the X variables . $\rightarrow SSR$

$$\underset{\geq 0}{SSTO} = \underset{\geq 0}{SSE} + \underset{\geq 0}{SSR}$$

- $0 \leq R^2 \leq 1$
- Adding more X variables to the model will never decrease R^2:
  (i) $SSTO$ remains the same. $SSTO = \sum (y_i - \bar{y})^2$

  (ii) SSE will not increase $\leftrightarrow$ SSR will not decrease.

Use As Many X Variables As Possible?

but less likely to generalize to other data

- With more X variables, the model does fit the observed data better, indicated by smaller SSE. goodness-of-fit (one data only)
- 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.
  - makes interpretation more difficult.
  - makes model maintenance more costly.

Adjusted Coefficient of Multiple Determination

Adjust for the number of X variables in the model:

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

MSE

↳ replace SS by MS

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

- $R_a^2 \leq R^2$
- It's possible for R_a^2 to decrease when adding more X variables into the model:
  - 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 : $Y \sim X_1, X_2, X_3, X_1X_2$ Improve ↑↑

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

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

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

Multiple Regression: Inference of Regression Coefficients

LS Estimator: Standard Error

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

MSE

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

$s(\hat{\beta}_k)$ – the standard error of $\hat{\beta}_k$ – is the positive square-root of the $(k+1)th$ diagonal element of $MSE(\mathbf{X}'\mathbf{X})^{-1}$. $(k=0,1,2,\ldots p-1)$

Under Normal error model:

- $(1-\alpha)100\%$-confidence interval of β_k:

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

- T statistic:

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

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

  At level α, the decision rule is to reject H_0 if and only if

  $|T^*| > t(1-\alpha/2; (n-p))$.

model 2

Example: Nonadditive Model with Interaction $X_1 X_2$

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

```
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 ***
---
```

$\hat{\beta}_4$ $SE(\hat{\beta}_4)$ $T = \frac{\hat{\beta}_4 - 0}{SE(\hat{\beta}_4)}$ 2-sided p-value

$\sqrt{MSE}$

```
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
```

R^2 ? R_a^2

Test whether there is an interaction between X_1 and X_2 at significance level 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.

Multiple Regression: Estimation of Mean Response

Mean Response

For a given set of X values:

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

the corresponding mean response is:

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

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

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

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

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

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

- $(1-\alpha)100\%$-confidence interval of $E(Y_h)$:

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

Example: Nonadditive Model with Interaction $X_1 X_2$

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

```
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 ***
---

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
```

Estimate the mean response when $X_1 = 0.8, X_2 = 0.5, X_3 = -1$:

- $\mathbf{X}_h' = \begin{array}{c} \begin{array}{ccccc} 1 & X_1 & X_2 & X_3 & X_1 \cdot X_2 \end{array} \\ \begin{bmatrix} 1 & 0.8 & 0.5 & -1 & 0.8 \times 0.5 \end{bmatrix}_{1\times 5} \end{array}$ $\quad (X^T X)^{-1} \quad (\ \ )\, X_h$
- Estimator $\widehat{Y}_h = \mathbf{X}_h' \hat{\beta} = 1.290$:

$$\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h = 0.170, \quad \sqrt{MSE} = 1.026$$

$$s(\widehat{Y}_h) = 1.026 \times \sqrt{0.170} = 0.423$$

- $n = 30, p = 5$: $t(0.995; 25) = 2.787$
- A 99%-confidence interval of $E(Y_h)$:

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

Multiple Regression:

Prediction

Prediction of a New Observation

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

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

- Standard error of prediction:

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

- $(1-\alpha)100\%$-prediction interval of $Y_{h(new)}$:

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

$n=28$
$\Downarrow$

Example: Nonadditive Model with Interaction $X_1 X_2$

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

```
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 ***
---

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
```

Predict a new observation when $X_1 = 0.8, X_2 = 0.5, X_3 = -1$:

- Predicted value $\widehat{Y}_h = \mathbf{X}_h' \hat{\beta} = 1.290$:

$$\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h = 0.170, \quad \sqrt{MSE} = 1.026$$

$$s(pred) = 1.026 \times \sqrt{1 + 0.170} = 1.1098$$

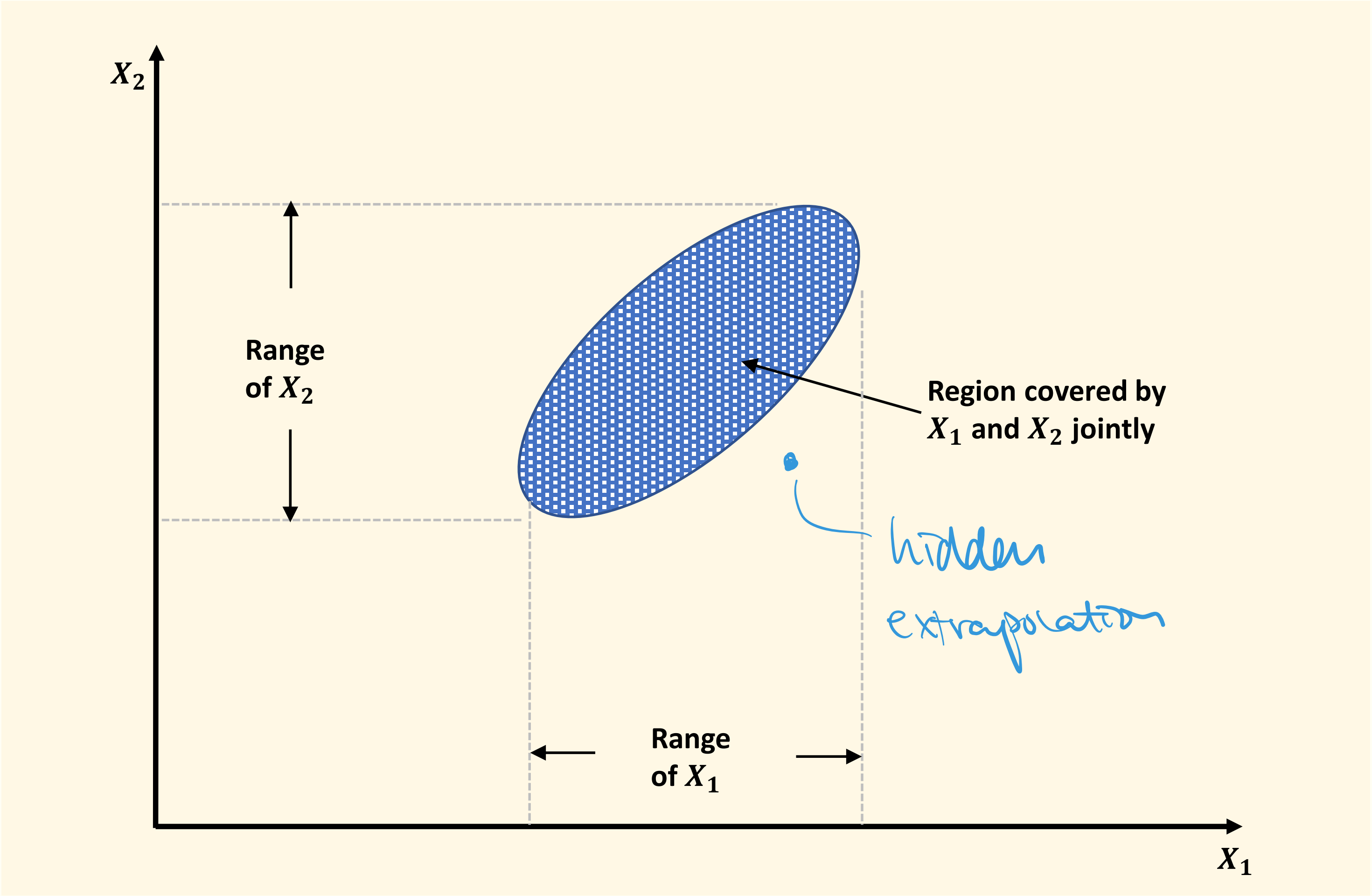
- A 99%-prediction interval of Y_{hnew}:

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

Hidden Extrapolations

- Extrapolation occurs when predicting the response variable for values of the X variable(s) lying outside the range of the observed data.
- With more than one X variables, the levels of all X variables jointly define the region of the observations.

With two X variables, we can look at their scatter plot to determine the region of observations.



Multiple Regression: Extra Sum of Squares

Notation

eg. $I = \{1, 2\}$ $I = \{2, 3, 5\}$

- $\mathcal{I}$: an index set

$X_I = \{X_1, X_2\}$ $X_I = \{X_2, X_3, X_5\}$

- $X_{\mathcal{I}} := \{X_i : i \in \mathcal{I}\}$
- Example: $\mathcal{I} = \{2, 3\}$, $X_{\mathcal{I}} = \{X_2, X_3\}$

$SSE(X_I)$: SSE when regressing
y to x_1 & x_2

- $SSE(X_{\mathcal{I}})$ and $SSR(X_{\mathcal{I}})$ denote the error sum of squares and regression sum of squares, respectively, under the regression model with $X_{\mathcal{I}} := \{X_i : i \in \mathcal{I}\}$ being the set of X variables.

SSE & SSR when regressing y to $X_I = X\{\ \in I\}$

— — — ?

Extra Sum of Squares

$$SSR(X_{\mathcal{J}}|X_{\mathcal{I}}) := SSE(X_{\mathcal{I}}) - SSE(X_{\mathcal{I}}, X_{\mathcal{J}}),$$

additional x (to original)

$= SSR(X_I, X_J) - SSR(X_I)$

where $\mathcal{I}$ and $\mathcal{J}$ are two **non-overlapping** index sets.

decrease in SSR

- It is the reduction in error sum of squares by adding $X_{\mathcal{J}}$ to the model where $X_{\mathcal{I}}$ is the set of X variables.
- degrees of freedom: the number of additional X variables being added: $d.f.(SSR(X_{\mathcal{J}}|X_{\mathcal{I}})) = |\mathcal{J}|$

  # of indices in J
- Mean squares:

$$MSR(X_{\mathcal{J}}|X_{\mathcal{I}}) := \frac{SSR(X_{\mathcal{J}}|X_{\mathcal{I}})}{d.f.(SSR(X_{\mathcal{J}}|X_{\mathcal{I}}))}$$

Properties

- $SSR(X_{\mathcal{J}}|X_{\mathcal{I}}) \geq 0$
- In general, $SSR(X_{\mathcal{J}}|X_{\mathcal{I}}) \neq SSR(X_{\mathcal{I}}|X_{\mathcal{J}})$
- $SSR(X_{\mathcal{J}}|X_{\mathcal{I}}) = SSR(X_{\mathcal{I}}, X_{\mathcal{J}}) - SSR(X_{\mathcal{I}})$, so it is also the marginal increase of the regression sum of squares by adding $X_{\mathcal{J}}$ to the model.

SSTO = SSE + SSR

fixed, not depend on model

Multiple Regression: ESS

Examples

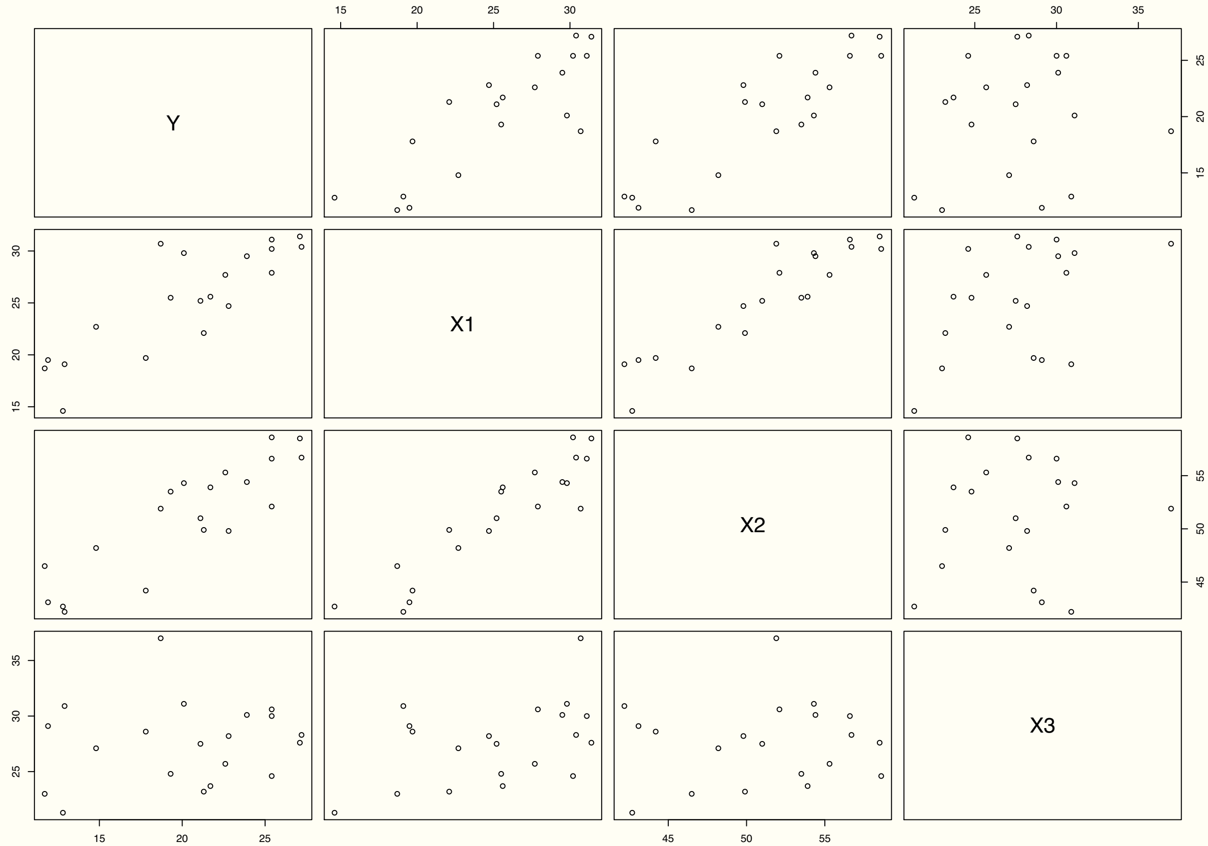
Body Fat

A researcher measured the amount of body fat (Y) of 20 healthy females 25 to 34 years old, together with three (potential) predictor variables, triceps skinfolds thickness (X_1), thigh circumference (X_2), and midarm circumference (X_3). The amount of body fat was obtained by a cumbersome and expensive procedure requiring immersion of the person in water. Thus it would be helpful if a regression model with some or all of these predictors could provide reliable estimates of body fat as these predictors are easy to measure.

A snapshot of the data.

case	X1	X2	X3	Y
Triceps	Thigh	MidArm	BodyFat	
1	19.5	43.1	29.1	11.9
2	24.7	49.8	28.2	22.8
3	30.7	51.9	37.0	18.7
4	29.8	54.3	31.1	20.1
5	19.1	42.2	30.9	12.9
6	25.6	53.9	23.7	21.7
...	...	...	...	...

Figure: Scatter plot matrix



No obvious nonlinearity

Correlation matrix

	X1	X2	X3	Y
X1	1.0000000	0.9238425	0.4577772	0.8432654
X2	0.9238425	1.0000000	0.0846675	0.8780896
X3	0.4577772	0.0846675	1.0000000	0.1424440
Y	0.8432654	0.8780896	0.1424440	1.0000000

X_1 and X_2 are strongly correlated, X_1 and X_3 are moderately correlated, X_2 and X_3 are weakly correlated. Moreover, X_1, X_2 are strongly correlated with Y and X_3 is weakly correlated with Y.

- Model 1: regression of Y on X_1

$$Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- Model 2: regression of Y on X_2

$$Y_i = \beta_0 + \beta_2 X_{i2} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- Model 3: regression of Y on X_1 and X_2

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- Model 4: regression of Y on X_1, X_2 and X_3

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i, \quad i = 1, \cdots, 20.$$

Boy Fat: Model 1

```
Call:
lm(formula = Y ~ X1, data = fat)

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.4961     3.3192  -0.451    0.658
X1            0.8572     0.1288   6.656 3.02e-06 ***
---
Residual standard error: 2.82 on 18 degrees of freedom
Multiple R-squared: 0.7111,     Adjusted R-squared: 0.695
F-statistic:  44.3 on 1 and 18 DF,  p-value: 3.024e-06

Analysis of Variance Table
Response: Y
Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 352.27  352.27  44.305 3.024e-06 ***
Residuals 18 143.12    7.95
```

Boy Fat: Model 2

```
Call:
lm(formula = Y ~ X2, data = fat)

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -23.6345     5.6574  -4.178 0.000566 ***
X2            0.8565     0.1100   7.786  3.6e-07 ***
---
Residual standard error: 2.51 on 18 degrees of freedom
Multiple R-squared: 0.771,      Adjusted R-squared: 0.7583
F-statistic: 60.62 on 1 and 18 DF,  p-value: 3.6e-07

Analysis of Variance Table
Response: Y
Df Sum Sq Mean Sq F value  Pr(>F)
X2          1 381.97  381.97  60.617 3.6e-07 ***
Residuals 18 113.42    6.30
```

Boy Fat: Model 3

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

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.1742     8.3606  -2.293   0.0348 *
X1            0.2224     0.3034   0.733   0.4737
X2            0.6594     0.2912   2.265   0.0369 *
---
Residual standard error: 2.543 on 17 degrees of freedom
Multiple R-squared: 0.7781,     Adjusted R-squared: 0.7519
F-statistic:  29.8 on 2 and 17 DF,  p-value: 2.774e-06

Analysis of Variance Table
Response: Y
Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 352.27  352.27 54.4661 1.075e-06 ***
X2         1  33.17   33.17  5.1284    0.0369 *
Residuals 17 109.95    6.47
```

Boy Fat: Model 4

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

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.085     99.782   1.173    0.258
X1             4.334      3.016   1.437    0.170
X2            -2.857      2.582  -1.106    0.285
X3            -2.186      1.595  -1.370    0.190
---
Residual standard error: 2.48 on 16 degrees of freedom
Multiple R-squared: 0.8014,     Adjusted R-squared: 0.7641
F-statistic: 21.52 on 3 and 16 DF,  p-value: 7.343e-06

Analysis of Variance Table
Response: Y
Df Sum Sq Mean Sq F value    Pr(>F)
X1          1 352.27  352.27 57.2768 1.131e-06 ***
X2          1  33.17   33.17  5.3931   0.03373 *
X3          1  11.55   11.55  1.8773   0.18956
Residuals 16  98.40    6.15
```

Body Fat: ESS

- From Model 1, $SSE(X_1) = 143.12$ and from Model 3, $SSE(X_1, X_2) = 109.95$:

$$SSR(X_2|X_1) = SSE(X_1) - SSE(X_1, X_2) = 143.12 - 109.95 = 33.17$$

- From Model 2, $SSE(X_2) = 113.42$:

$$SSR(X_1|X_2) = SSE(X_2) - SSE(X_1, X_2) = 113.42 - 109.95 = 3.47$$

- The reduction of SSE by adding X_2 to the model with X_1 is much more than the reduction of SSE by adding X_1 to the model with X_2.

- From Model 4, $SSE(X_1, X_2, X_3) = 98.40$:

$$\begin{aligned} SSR(X_3|X_1, X_2) &= SSE(X_1, X_2) - SSE(X_1, X_2, X_3) \\ &= 109.95 - 98.40 = 11.55 \end{aligned}$$

- Moreover,

$$SSR(X_2, X_3|X_1) = SSE(X_1) - SSE(X_1, X_2, X_3) = 143.12 - 98.40 = 44.72,$$

$$SSR(X_1, X_3|X_2) = SSE(X_2) - SSE(X_1, X_2, X_3) = 113.42 - 98.40 = 15.02.$$

- These two extra sums of squares have degrees of freedom 2:

$MSR(X_2, X_3|X_1) = 44.72/2 = 22.36,$

$MSR(X_1, X_3|X_2) = 15.02/2 = 7.51$

Multiple Regression: Decomposition of SSR

Decomposition of SSR into ESS

For a model with multiple X variables, the regression sum of squares (SSR) can be expressed as the sum of several extra sums of squares.

- $SSR(X_1, X_2) = SSR(X_1) + SSR(X_2|X_1)$: $SSR(X_1)$ measures the contribution by having X_1 alone in the model, whereas $SSR(X_2|X_1)$ measures the additional contribution when X_2 is added, given that X_1 is already in the model.
- However, such decomposition is usually not unique: $SSR(X_1, X_2) = SSR(X_2) + SSR(X_1|X_2)$.

- More X variables, more decompositions.
- For example, with three X variables:

$$
\begin{aligned}
SSR(X_1, X_2, X_3) &= SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1, X_2) \\
SSR(X_1, X_2, X_3) &= SSR(X_2) + SSR(X_1|X_2) + SSR(X_3|X_1, X_2) \\
SSR(X_1, X_2, X_3) &= SSR(X_1) + SSR(X_2, X_3|X_1), \quad \cdots, \cdots
\end{aligned}
$$

Body Fat

- From Model 1, $SSR(X_1) = 352.27$; Also $SSR(X_2|X_1) = 33.17$ and $SSR(X_3|X_1, X_2) = 11.55$. So

$$SSR(X_1, X_2, X_3) = 352.27 + 33.17 + 11.55 = 396.99.$$

- From Model 2, $SSR(X_2) = 381.97$; Also $SSR(X_1|X_2) = 3.47$. So

$$SSR(X_1, X_2, X_3) = 381.97 + 3.47 + 11.55 = 396.99.$$

R output: anova()

```
lm(formula = Y ~ X1 + X2 + X3, data = fat)
> anova(fit4)
Analysis of Variance Table
Df Sum Sq Mean Sq F value    Pr(>F)
X1          1 352.27  352.27 57.2768 1.131e-06 ***
X2          1  33.17   33.17  5.3931   0.03373 *
X3          1  11.55   11.55  1.8773   0.18956
Residuals 16  98.40    6.15
```

Decomposition of *SSR* into single d.f. ESS, by order of the *X* variables entering the model:

Source of Variation	SS	d.f.	MS
Regression	396.99	3	132.33
X_1	352.27	1	352.27
$X_2 \vert X_1$	33.17	1	33.17
$X_3 \vert X_1, X_2$	11.55	1	11.55
Error	98.40	16	6.15
Total	495.39	19	

- $SSR(X_2, X_3|X_1) = SSR(X_2|X_1) + SSR(X_3|X_1, X_2) = 33.17 + 11.55 = 44.72.$
- How to get $SSR(X_2|X_1, X_3)$ from the R output? Enter the X variables in a different order, i.e., X_1, X_3, X_2:

```
lm(formula = Y ~ X1 + X3 + X2, data = fat)
> anova(fit4.alt2)
Analysis of Variance Table
Df Sum Sq Mean Sq F value    Pr(>F)
X1         1 352.27  352.27 57.2768 1.131e-06 ***
X3         1  37.19   37.19  6.0461   0.02571 *
X2         1   7.53    7.53  1.2242   0.28489
Residuals 16  98.40    6.15
```

- $SSR(X_2|X_1, X_3) = 7.53$

Multiple Regression:

General Linear Tests

General Linear Tests

$\mathcal{I}$ and $\mathcal{J}$ are two non-overlapping index sets:

- **Full model**: with both $X_{\mathcal{I}}$ and $X_{\mathcal{J}}$
- **Reduced model**: with only $X_{\mathcal{I}}$
- Test whether $X_{\mathcal{J}}$ may be dropped out of the full model:

$$H_0 : \beta_j = 0, \quad \text{for } \textbf{all } j \in \mathcal{J} \quad vs. \quad H_a : \text{not all } \beta_j : j \in \mathcal{J} \text{ is zero}$$

- H_0 corresponds to the reduced model with only $X_{\mathcal{I}}$.

F Test

Compare *SSE* under the full model with *SSE* under the reduced model by an F ratio:

$$F^* = \frac{\frac{SSE(R)-SSE(F)}{df_R-df_F}}{\frac{SSE(F)}{df_F}} = \frac{MSR(X_{\mathcal{J}}|X_{\mathcal{I}})}{MSE(F)}$$

- Under H_0 (i.e., the reduced model):

$$F^* \sim_{H_0} F_{df_R-df_F, df_F}$$

- Reject H_0 at level α iff the observed

  $F^* > F(1-\alpha; df_R - df_F, df_F)$.

Multiple Regression:

General Linear Tests

Examples

F-test for Regression Relation

- Full model with $X_1, \cdots, X_{p-1}$:

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

- Reduced model with no X variable:

$$Y_i = \beta_0 + \epsilon_i, \quad i = 1, \cdots, n, \quad SSE(R) = SSTO, \; df_R = n - 1$$

- $SSE(R) - SSE(F) = SSTO - SSE(F) = SSR(F)$, and $df_R - df_F = (n-1) - (n-p) = p - 1 = d.f.(SSR(F))$
- $F^* = \frac{SSR(F)/(p-1)}{SSE(F)/(n-p)} = \frac{MSR(F)}{MSE(F)}$

Test whether a Single $\beta_k = 0$

Body Fat: for the model with all three predictors, test whether the midarm circumference (X_3) can be dropped.

- Full model: $SSE(F) = 98.40$ with d.f. 16:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- Reduced model: $SSE(R) = 109.95$ with d.f. 17:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- $F^* = \frac{11.55/1}{98.40/16} = 1.88$; Pvalue=$P(F_{1,16} > 1.88) = 0.189$, so X_3 can be dropped.

Equivalence between F-test and T-test

- $H_0 : \beta_k = 0 \;\; vs. \;\; H_a : \beta_k \neq 0$
- T-test:

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

where $\hat{\beta}_k$ is the LS estimator of β_k and $s\{\hat{\beta}_k\}$ is its standard error. At level α, reject H_0 when $|T^*| > t(1-\alpha/2; n-p)$.

- $F^* = (T^*)^2$ and $F(1-\alpha; 1, n-p) = (t(1-\alpha/2; n-p))^2 \rightarrow$ F-test and two-sided T-test are equivalent.

For one-sided alternatives, we still need the T-tests.

Test whether Several $\beta_k = 0$

Body Fat: Test whether both X_2 and X_3 can be dropped from the model with all three predictors:

- Full model: $SSE(F) = 98.40$ with d.f. 16:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- Reduced model: $SSE(R) = 143.12$ with d.f. 18:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i, \quad i = 1, \cdots, 20.$$

- $F^* = \frac{44.72/2}{98.40/16} = 3.635$; Pvalue$= P(F_{2,16} > 3.635) = 0.0499$

Test Equality of Several β_ks

- Full model: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i$
- For $q \leq p-1$: $H_0 : \beta_1 = \cdots = \beta_q$ vs. $H_a : \beta_1, \cdots, \beta_q$ are not all equal
- Reduced model: $Y_i = \beta_0 + \beta_c (X_{i1} + \cdots + X_{iq}) + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i$
- β_c denotes the common value of $\beta_1, \cdots, \beta_q$ under H_0, and $X_1 + \cdots + X_q$ is the corresponding (new) X variable. $SSE(R)$ has d.f. $n - (p - q + 1)$.
- $F^* = \frac{(SSE(R) - SSE(F))/(q-1)}{SSE(F)/(n-p)} \underset{H_0}{\sim} F_{q-1,n-p}$