# Chapter 3

# Multiple Linear Regression

# Chapter 3h

**Multicollinearity (pages 3-6)**

**Why do regression coefficients have the wrong sign? (pages 4-14)**

***X* matrix with orthogonal columns (pages 15-24)**

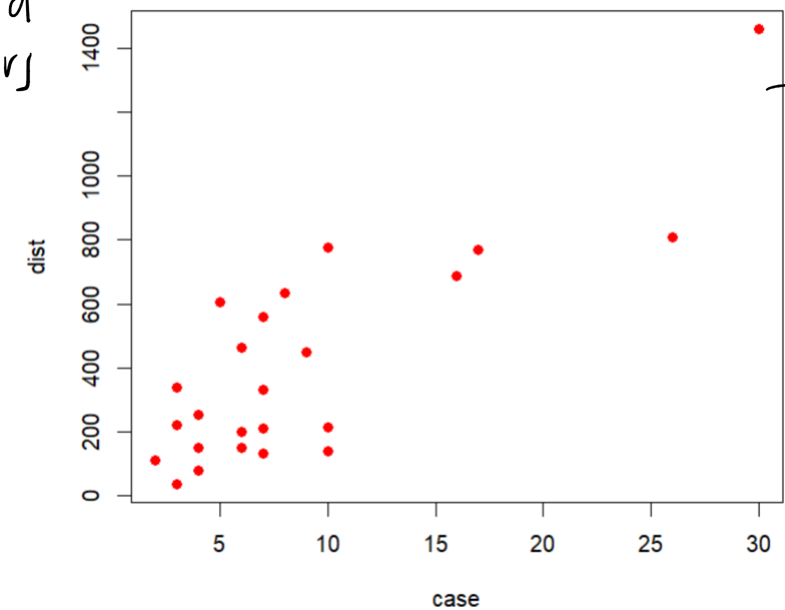
# Multicollinearity: near dependence among the regressor variables

## Example - Delivery time data

correlation matrix

$$\mathbf{W'W} = \begin{array}{cc} & \rho \\ \left[\begin{array}{cc} 1.00000 & 0.824215 \\ 0.824215 & 1.00000 \end{array}\right] \\ \rho & \end{array} \quad \text{and} \quad (\mathbf{W'W})^{-1} = \begin{bmatrix} 3.11841 & -2.57023 \\ -2.57023 & 3.11841 \end{bmatrix}$$

X'X based on standard regressors (unit length scale)




ρ = 0.82 (the correlation is quite high)

$$\frac{\mathrm{Var}\left(\hat{b}_1\right)}{\sigma^2} = \frac{\mathrm{Var}\left(\hat{b}_2\right)}{\sigma^2} = 3.11841 > 1$$

if there is dependence between variables

# Multicollinearity: ~~near dependence~~ independence among the regressor variables

## Example - Hypothetical data

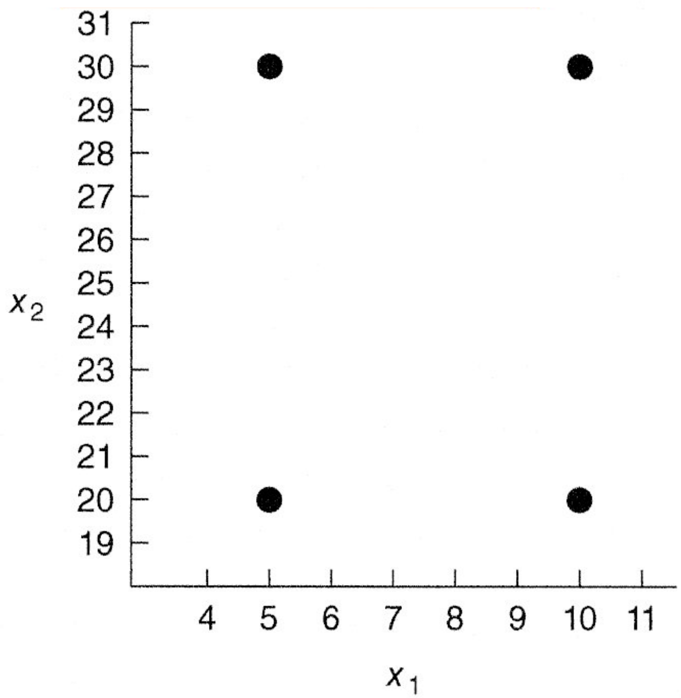
0 is the correlation (no correlation between $x_1$ and $x_2$)

identity matrix

$$\mathbf{W}'\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad (\mathbf{W}'\mathbf{W})^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

VIF (variance inflation factor)

$var(\beta) = \sigma^2 (X'X)^{-1}$



| $x_1$ | $x_2$ |
|---|---|
| 5 | 20 |
| 10 | 20 |
| 5 | 30 |
| 10 | 30 |
| 5 | 20 |
| 10 | 20 |
| 5 | 30 |
| 10 | 30 |

↑ variance, lesser the accuracy

$$\frac{\mathrm{Var}(\hat{b}_1)}{\sigma^2} = \frac{\mathrm{Var}(\hat{b}_2)}{\sigma^2} = 1$$

Variance Inflation Factors

1. The main diagonal elements of $(\boldsymbol{X'X})^{-1}$ are called the variance inflation factors (VIFs).

   $\boldsymbol{C} = (\boldsymbol{X'X})^{-1}$.

   Variance inflation factor, $\text{VIF}_j = C_{jj}, \quad j = 1, 2, ..., k$.

2. Suppose all the regressor variables are unit length scaled.

   (a) The $\boldsymbol{X}'\boldsymbol{X}$ matrix is in correlation form.

   (b) $\text{VIF}_j, C_{jj} = \dfrac{1}{1 - R_j^2}, \quad j = 1, 2, ..., k,$

   large if there is a dependency between the regressors since $R_j^2 \approx 1$

   where $R_j^2$ is the coefficient of determination obtained when $x_j$ is regressed on the remaining $k-1$ regressor variables.

   (c) If $x_j$ is nearly dependent on some subset of the remaining regressor variables, $R_j^2$ will be near one and hence $\text{VIF}_j$ will be much greater than one.

   indep, no multi-collinearity

   If $x_j$ is $\boxed{\text{orthogonal}}$ to the remaining regressor variables, $R_j^2$ will be near zero and hence $\text{VIF}_j$ will be near one.

   measure of dependency

   (d) The $\text{VIF}_j$ measures how much the variance of the regression coefficient $\hat{\beta}_j$ is affected by the relationship of $x_j$ with the other regressor variables.

   high correlation between regressors → strong dependence

   (e) The $\text{VIF}_j$ can be used to detect multicollinearity. In general, a VIF that is at least 2.5 provides some evidence of multicollinearity. Some researchers recommended 5 as the cutoff, so there is some subjectivity here.

   → BAD as it inflates variance of estimated regression

   use 2.5 in exams

   (f) The function `vif(fitted.model)` will return the VIFs for unit length scaled regressor variables whether the regressor variables are unit length scaled or not. Note that if a regressor variable $x$ is unit length scaled, then $x$ remains unchanged if it is unit length scaled again.

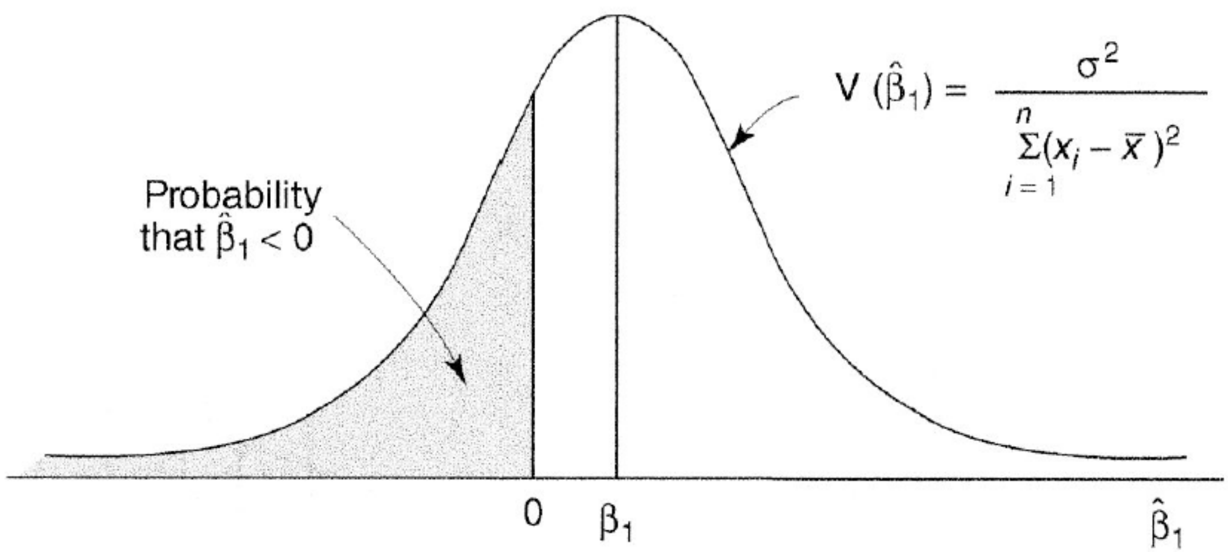
## Why do regression coefficients have the wrong sign?

1. The range of some of the regressors is too small.
2. Important regressors have not been included in the model.
3. Multicollinearity is present.
4. Computational errors have been made.

measure of spread of x value

# 1. The range of some of the regressors is too small.

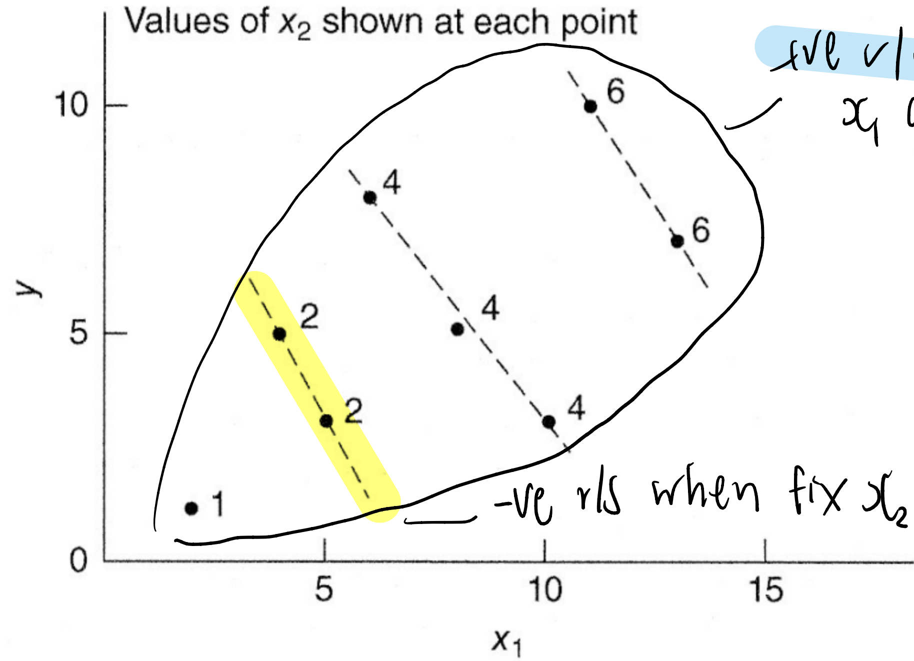
The variance of the regression coefficient $\hat{\beta}_1$ is $\mathrm{Var}(\hat{\beta}_1) = \sigma^2 / S_{xx} = \sigma^2 / \sum_{i=1}^{n} (x_i - \bar{x})$. Note that the variance of $\hat{\beta}_1$ is inversely proportional to the "spread" of the regressor. Therefore, if the levels of $x$ are all close together, the variance of $\hat{\beta}_1$ will be relatively large.




Sampling distribution of $\hat{\beta}_1$.

**2.** Important regressors have not been included in the model.

| $x_1$ | $x_2$ | $y$ |
|---|---|---|
| 2 | 1 | 1 |
| 4 | 2 | 5 |
| 5 | 2 | 3 |
| 6 | 4 | 8 |
| 8 | 4 | 5 |
| 10 | 4 | 3 |
| 11 | 6 | 10 |
| 13 | 6 | 7 |




+ve r/s between $x_1$ and $y$

-ve r/s when fix $x_2$ (just look at line)

**Figure 3.15** Plot of $y$ versus $x_1$.

$$\hat{y} = 1.835 + 0.463x_1$$

$SS_R(\beta_1 \mid \beta_2, \beta_0)$

$$\hat{y} = 1.036 - 1.222x_1 + 3.649x_2$$

when 2 regressor variables highly correlated

↓

variance ↑

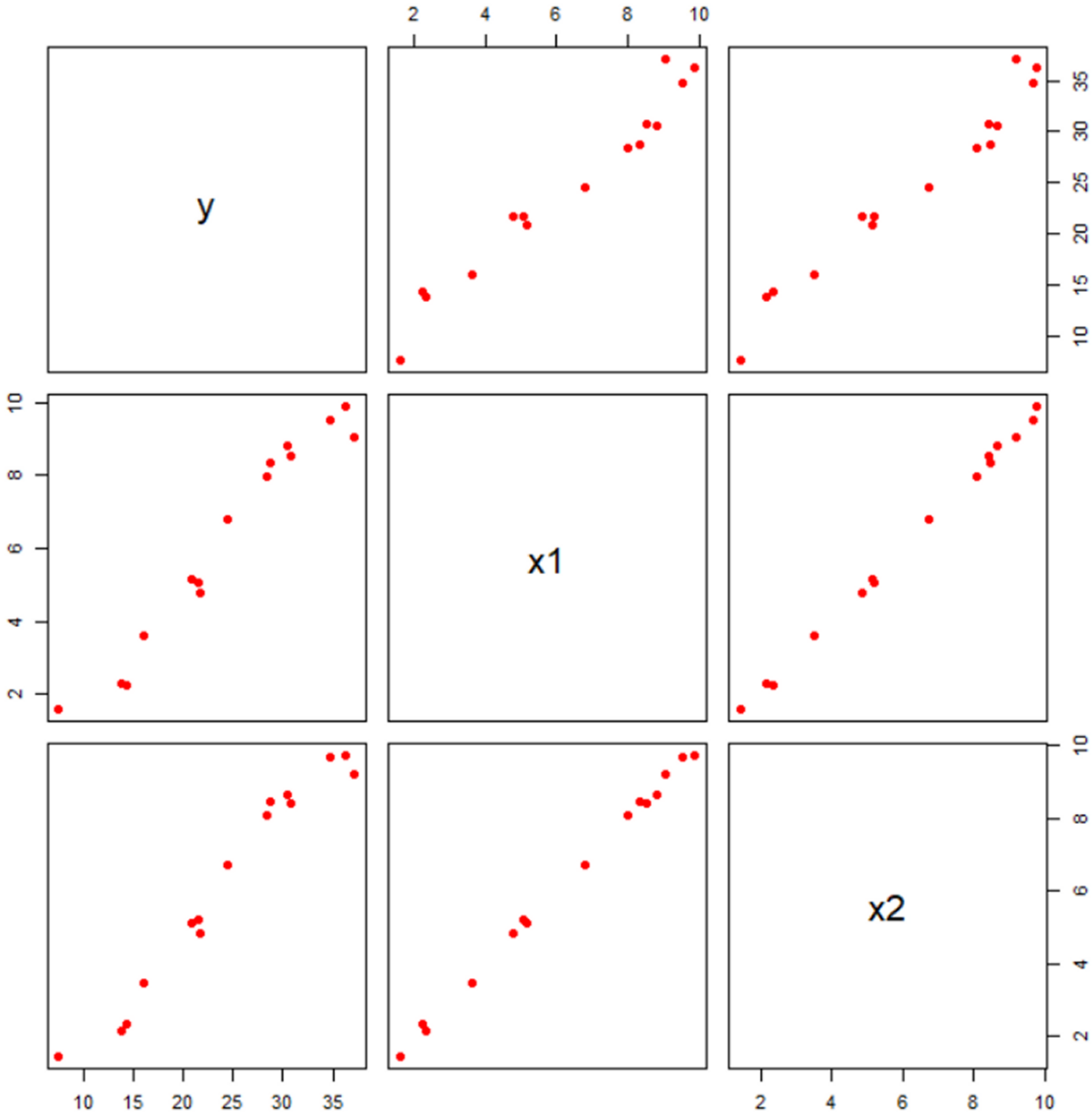
$\hat{\beta}_1 = -1.222$ in the multiple regression model is a partial regression coefficient; it measures the effect of $x_1$ **given** that $x_2$ is also in the model.

**3.** Multicollinearity is present.

## Example

```
> newdata <- data.frame(y,x1,x2)
> newdata
           y       x1       x2
1  28.381434 7.977843 8.079269
2  16.013798 3.612941 3.492279
3  36.304329 9.878469 9.745497
4  37.164591 9.053100 9.205440
5  30.506969 8.786434 8.665476
6   7.559907 1.595936 1.428194
7  21.682200 4.777778 4.857980
8  30.731600 8.538796 8.405191
9  13.795757 2.311567 2.159199
10 34.690268 9.523631 9.672278
11 14.276074 2.254322 2.350160
12 20.839655 5.153525 5.127385
13 24.490453 6.781726 6.726533
14 21.608255 5.066166 5.211069
15 28.792134 8.344372 8.479187
```

$x_1$, $x_2$ highly correlated



y
x1
x2

```
> summary(lm(y~x1))

Call:
lm(formula = y ~ x1)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.7557 -1.4760 -0.2993  1.0468  4.1619 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   5.4603     1.2037   4.536 0.000559 ***
x1            3.0423     0.1761  17.274 2.39e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.896 on 13 degrees of freedom
Multiple R-squared:  0.9583,	Adjusted R-squared:  0.955 
F-statistic: 298.4 on 1 and 13 DF,  p-value: 2.392e-10

> summary(lm(y~x2))

Call:
lm(formula = y ~ x2)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.4115 -1.3444 -0.1594  1.3384  3.7720 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   5.6476     1.1143   5.068 0.000215 ***
x2            3.0140     0.1628  18.515    1e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.774 on 13 degrees of freedom
Multiple R-squared:  0.9635,	Adjusted R-squared:  0.9607 
F-statistic: 342.8 on 1 and 13 DF,  p-value: 1.003e-10
```

```
> summary(lm(y~x1+x2))

Call:
lm(formula = y ~ x1 + x2)

Residuals:
   Min     1Q Median     3Q    Max 
-2.743 -1.031 -0.235  1.089  3.416 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)    5.902      1.179   5.005 0.000307 ***
x1            -3.030      3.931  -0.771 0.455705    
x2             6.005      3.883   1.546 0.147995    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 1.802 on 12 degrees of freedom
Multiple R-squared:  0.9652,	Adjusted R-squared:  0.9594 
F-statistic: 166.4 on 2 and 12 DF,  p-value: 1.78e-09
```

## $\boldsymbol{X}$ matrix with orthogonal columns

1. Definition (Independent vectors)

   Let $\boldsymbol{u} = (u_1, u_2, ..., u_k)'$ and $\boldsymbol{v} = (v_1, v_2, ..., v_k)'$ be two nonzero vectors, $c_1$ and $c_2$ are two scalar constants. The two vectors are said to be independent if $c_1\boldsymbol{u} + c_2\boldsymbol{v} = \boldsymbol{0}$, then $c_1 = c_2 = 0$. no $c_1$/$c_2$ can fulfil this eqn if indep

   Note: This means $\boldsymbol{u}$ cannot be written as a function of $\boldsymbol{v}$.

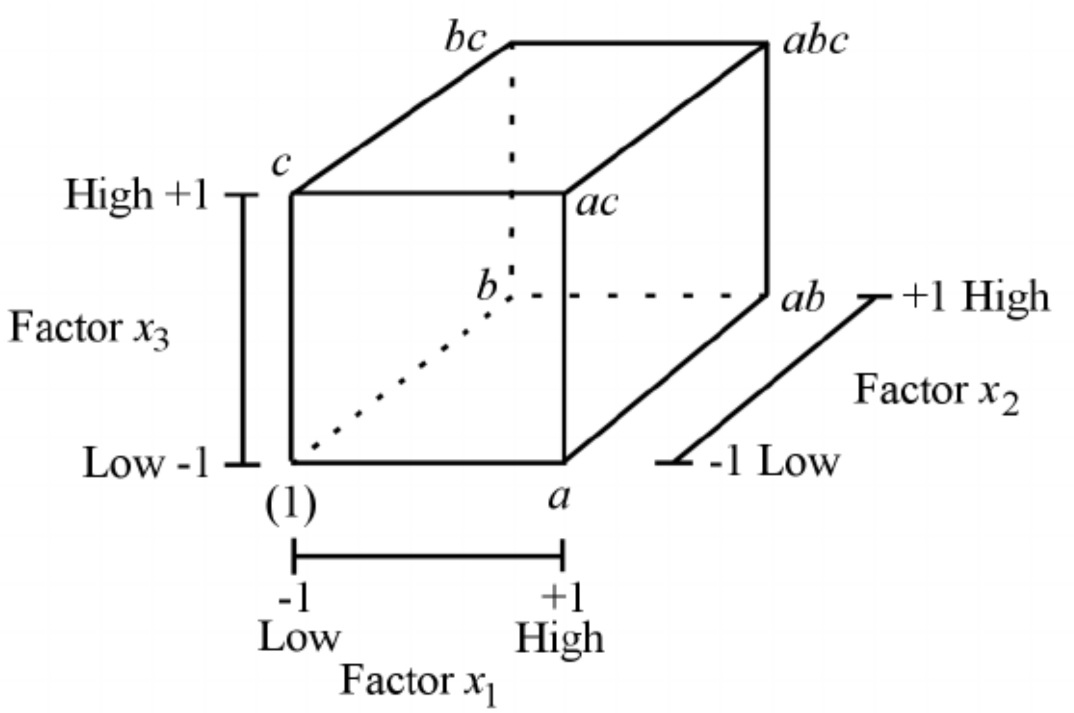
   Note: Two or more vectors are said to be linearly independent if none of them can be written as a linear combination of the others.

2. Definition (Orthogonal vectors)

   Let $\boldsymbol{u} = (u_1, u_2, ..., u_k)'$ and $\boldsymbol{v} = (v_1, v_2, ..., v_k)'$ be two nonzero vectors. The two vectors are said to be orthogonal if the dot product $\boldsymbol{u}'\boldsymbol{v} = 0$.

indep $\not\Rightarrow$ orthogonal but orthogonal $\Rightarrow$ indep

4. Cookie Example (Matrix $\boldsymbol{X}$ with orthogonal columns)




Consider baking cookies. In order to make good cookies, we need to know how to set the following variables:

(1) Temperature of the oven ($x_1$)

(2) Time for baking the cookies ($x_2$)

(3) Thickness of a cookie ($x_3$)

The response $y$ is a number between 0 and 10 that measures how good a cookie is.

Suppose we set the regressor variables as follows:

$$\text{Temperature} = \begin{cases} 190\ ^oC, & x_1 = -1, \\ 200\ ^oC, & x_1 = +1. \end{cases}$$

2 settings

$$\text{Time} = \begin{cases} 12\ min, & x_2 = -1, \\ 15\ min, & x_2 = +1. \end{cases}$$

2 settings

$$\text{Thickness} = \begin{cases} 0.5\ cm, & x_3 = -1, \\ 1.0\ cm, & x_3 = +1. \end{cases}$$

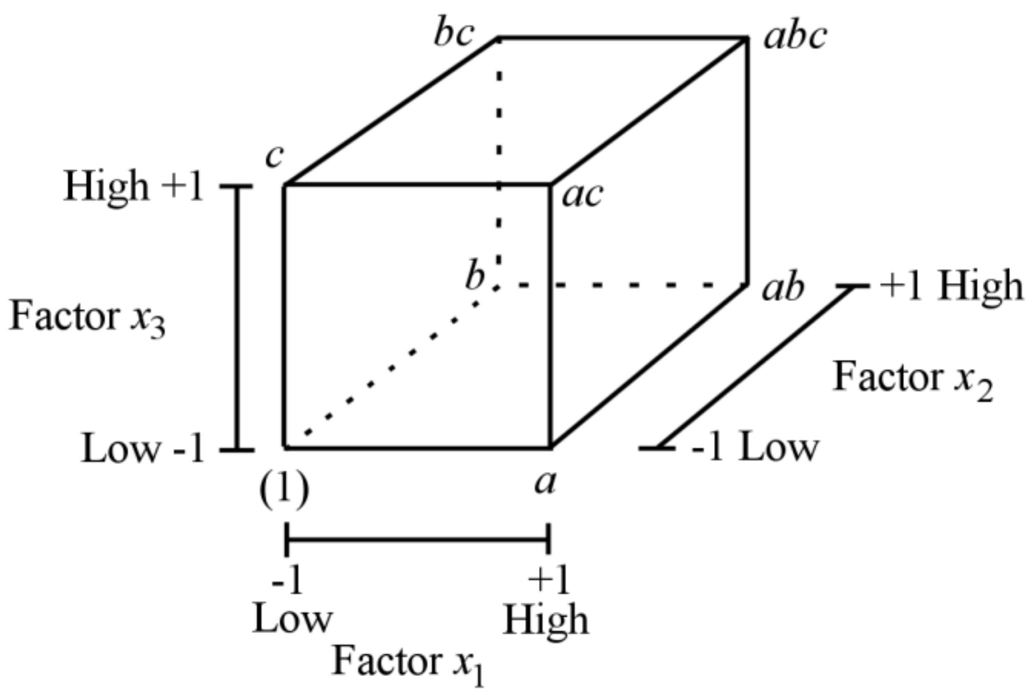
2 settings

$2^3 = 8$ combi

add up columns get 0

| Treatment | $I$ | $x_1$ | $x_2$ | $x_3$ | $x_1x_2$ | $x_1x_3$ | $x_2x_3$ | $x_1x_2x_3$ |
|---|---|---|---|---|---|---|---|---|
| (1) | $+1$ | $-1$ | $-1$ | $-1$ | $+1$ | $+1$ | $+1$ | $-1$ |
| $a$ | $+1$ | $+1$ | $-1$ | $-1$ | $-1$ | $-1$ | $+1$ | $+1$ |
| $b$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $+1$ | $-1$ | $+1$ |
| $c$ | $+1$ | $-1$ | $-1$ | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ |
| $ab$ | $+1$ | $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $-1$ |
| $ac$ | $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ |
| $bc$ | $+1$ | $-1$ | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ | $-1$ |
| $abc$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ | $+1$ |

all columns are orthogonal



Multiple linear regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 + \beta_7 x_1 x_2 x_3 + \epsilon$$

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

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ +1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \end{bmatrix} \qquad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \end{bmatrix}$$

Note: This $\boldsymbol{X}$ matrix is for one run of all the treatments. Multiple runs can also be conducted. It can be verified easily that all the columns are orthogonal.

5. Matrix $\boldsymbol{X}$ with orthogonal columns

(1) Multiple regression models

Full model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} = \boldsymbol{X_1\beta_1} + \boldsymbol{X_2\beta_2} + \boldsymbol{\epsilon}$$

Reduced models:

$$\boldsymbol{y} = \boldsymbol{X_1\beta_1} + \boldsymbol{\epsilon}$$
$$\boldsymbol{y} = \boldsymbol{X_2\beta_2} + \boldsymbol{\epsilon}$$

(2) $SS_R(\boldsymbol{\beta})$, $SS_R(\boldsymbol{\beta_1})$ and $SS_R(\boldsymbol{\beta_2})$

Full model:

$$\boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$$

0 for orthogonal regressors

$$\begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{y} \\ \boldsymbol{X}_2'\boldsymbol{y} \end{bmatrix}$$

If the columns of $\boldsymbol{X}_1$ are orthogonal to $\boldsymbol{X}_2$, then $\boldsymbol{X}_1'\boldsymbol{X}_2 = \boldsymbol{0}$ and $\boldsymbol{X}_2'\boldsymbol{X}_1 = \boldsymbol{0}$, we have

$$\boldsymbol{X}_1'\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1 = \boldsymbol{X}_1'\boldsymbol{y}$$
$$\boldsymbol{X}_2'\boldsymbol{X}_2\hat{\boldsymbol{\beta}}_2 = \boldsymbol{X}_2'\boldsymbol{y}$$

$$\hat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}$$
$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{y}$$

$$\begin{aligned} SS_R(\boldsymbol{\beta}) &= \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} \\ &= [\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2'] \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{y} \\ \boldsymbol{X}_2'\boldsymbol{y} \end{bmatrix} \\ &= \hat{\boldsymbol{\beta}}_1'\boldsymbol{X}_1'\boldsymbol{y} + \hat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{y} \end{aligned}$$

(2) $SS_R(\boldsymbol{\beta})$, $SS_R(\boldsymbol{\beta_1})$ and $SS_R(\boldsymbol{\beta_2})$

Full model:

$$SS_R(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}}_{\mathbf{1}}'\boldsymbol{X}_{\mathbf{1}}'\boldsymbol{y} + \hat{\boldsymbol{\beta}}_{\mathbf{2}}'\boldsymbol{X}_{\mathbf{2}}'\boldsymbol{y}$$

Reduced models:

$$SS_R(\boldsymbol{\beta_1}) = \hat{\boldsymbol{\beta}}_{\mathbf{1}}'\boldsymbol{X}_{\mathbf{1}}'\boldsymbol{y}$$
$$SS_R(\boldsymbol{\beta_2}) = \hat{\boldsymbol{\beta}}_{\mathbf{2}}'\boldsymbol{X}_{\mathbf{2}}'\boldsymbol{y}$$

(3) Relationship among $SS_R(\boldsymbol{\beta})$, $SS_R(\boldsymbol{\beta_1})$ and $SS_R(\boldsymbol{\beta_2})$

$$SS_R(\boldsymbol{\beta}) = SS_R(\boldsymbol{\beta_1}) + SS_R(\boldsymbol{\beta_2})$$
$$SS_R(\boldsymbol{\beta_1}|\boldsymbol{\beta_2}) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta_2}) = \overbrace{SS_R(\boldsymbol{\beta_1})}^{\text{don't depend on } \beta_2 \text{ coz orthogonal}}$$
$$SS_R(\boldsymbol{\beta_2}|\boldsymbol{\beta_1}) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta_1}) = SS_R(\boldsymbol{\beta_2})$$

6. Cookie Example (Matrix $\boldsymbol{X}$ with orthogonal columns)

Multiple linear regression model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} = \boldsymbol{X_1\beta_1} + \boldsymbol{X_2\beta_2} + \boldsymbol{\epsilon}$$

$$\boldsymbol{X} = \begin{bmatrix} +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & -1 & +1 & -1 & -1 & -1 \\ +1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 \\ +1 & -1 & +1 & +1 & -1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{bmatrix}$$

$$\boldsymbol{X_1} = \begin{bmatrix} +1 & -1 & -1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & -1 \\ +1 & +1 & -1 & +1 \\ +1 & -1 & +1 & +1 \\ +1 & +1 & +1 & +1 \end{bmatrix} \quad \boldsymbol{X_2} = \begin{bmatrix} +1 & +1 & +1 & -1 \\ -1 & -1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 \\ +1 & -1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \end{bmatrix} \quad \boldsymbol{\beta_1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \quad \boldsymbol{\beta_2} = \begin{bmatrix} \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \end{bmatrix}$$

$$SS_R(\boldsymbol{\beta_1}|\boldsymbol{\beta_2}) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta_2}) = SS_R(\boldsymbol{\beta_1})$$
$$SS_R(\boldsymbol{\beta_2}|\boldsymbol{\beta_1}) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta_1}) = SS_R(\boldsymbol{\beta_2})$$

7. Matrices with orthogonal columns are important in designs of experiments and more can be learnt in ST3232 Design and Analysis of Experiments

# The End