Chapter 2 Binary Data

2.1 Introduction:

Motivating example:

$$Z = 1: recovered; \;\; Z = 0: not \;\; recovered$$
$$x_1 = 1: hospital \;\; A; \;\; x_1 = 2: hospital \;\; B$$
$$x_2 = 1: surgical \;\; procedure \;\; I; \;\; x_2 = 2: surgical \;\; procedure \;\; II$$

The data are:

Table (a)

Data subject	Covariate	Response
1	$(1,1)$	0
2	$(1,2)$	1
3	$(1,2)$	0
4	$(2,1)$	0
5	$(2,2)$	1
6	$(1,2)$	1
7	$(1,1)$	1

Let $Z_i, i = 1,2,\cdots,7,$ be the responses indicating whether the patients are recovered or not and let $x_i = (x_{i1}, x_{i2}), i = 1,2,\cdots,7,$ be the hospitals and surgical procedures for the patients. Suppose

$$P(Z_i = 1) = \pi(x_i) = \pi_i$$

and

$$P(Z_i = 0) = 1 - \pi(x_i) = 1 - \pi_i.$$

Objective:

We want to investigate the relationship between the response probability π_i and the explanatory variable x_i. That is, whether the recovery of the patient is correlated to the hospital he chose or the surgical procedure conducted.

The original ungrouped data can be organized to the grouped data in the following table:

Table (b)

Covariate	Class size	Response
$(1,1)$	2	1
$(1,2)$	3	2
$(2,1)$	1	0
$(2,2)$	1	1

The responses in table (b) are

$$\boldsymbol{Y_i, 0 \leq Y_i \leq m_i, i = 1, 2, 3, 4;}$$
$$\boldsymbol{m_1 = 2, m_2 = 3, m_3 = 1, m_4 = 1.}$$

The most commonly used link function in practice is

$$\boldsymbol{g(\pi) = log\left(\frac{\pi}{1-\pi}\right) = logit(\pi).}$$