Chapter 3 Log-linear Models

3.1 Introduction:

Motivating example:

Ship type	Year of construction	Period of operation	Aggregate months service	Number of damage incidents
A	1960 – 64	1960 – 74	127	0
A	1960 – 64	1975 – 79	63	0
A	1965 – 69	1960 – 74	1095	3
A	1965 – 69	1975 – 79	1095	4
A	1970 – 74	1960 – 74	1512	6
A	1970 – 74	1975 – 79	3353	18
A	1975 – 79	1960 – 74	0	0^*
A	1975 – 79	1975 – 79	2244	11
⋮	⋮	⋮	⋮	⋮
E	1960 – 64	1960 – 74	45	0
E	1960 – 64	1975 – 79	0	0^{**}
E	1965 – 69	1960 – 74	789	7
E	1965 – 69	1975 – 79	437	7
E	1970 – 74	1960 – 74	1157	5
E	1970 – 74	1975 – 79	2161	12
E	1975 – 79	1960 – 74	0	0^*
E	1975 – 79	1975 – 79	542	1

Response: the number of damage incidents,

$$y_i, i = 1, \cdots, 40.$$

Covariates:

- **Ship type:** $A - E$.
- **Year of construction:** $1960 - 64, 1965 - 69, 1970 - 74, 1975 - 79$.
- **Period of operation:** $1960 - 74, 1975 - 79$.

In addition, it is reasonable to suppose that the number of damage incidents is directly proportional to the other variable, the aggregate month service or total period of risk.

Objective: We are concerned with the effects of the above factors (covariates) on the risk of damage. That is, the relationship between the number of damage incidents and these factors. A natural model is as follows:

$$\begin{aligned} log(\text{expected number of damage incidents}) \\ = \beta_0 + \log(\text{aggregate months service}) \\ + (\text{effect due to ship type}) \\ + (\text{effect due to year of construction}) \\ + (\text{effect due to service period}). \end{aligned}$$

Note:

It is quite reasonable to assume the number of damage incident is Poisson distributed. Thus, the above model is associated with canonical link for Poisson data.

In this chapter, we are concerned mainly with counted data not in the form of proportions. Typical examples involve counts of events in a Poisson or Poisson-like process, where the upper limit to the number is infinite or effectively so. However, departures from the idealized Poisson model are to be expected, for example, over-dispersion. Therefore, we avoid the assumption of Poisson variation and assume only that

$$Var(Y_i) = \sigma^2 E(Y_i).$$

The dependence of $\mu_i = E(Y_i)$ on the covariate x_i is assumed to be

$$g(\mu_i) = log(\mu_i) = \eta_i = x_i\beta.$$

The term log-linear models are referred to the above log-linear relationship.