Chapter 9

Poisson regression:

- Different exposure
- Explanatory variables: factors and covariates

Log-linear models

- Same exposure (e.g. not relevant)
- Explanatory variables: A few factors
  $\Rightarrow$ contingency tables.

Poisson regression

Response: $Y_i \sim Po(\mu_i)$, $\mu_i = n_i\theta_i$

Link: $\eta_i = \log(\mu_i) = \log(n_i) + \log(\theta_i)$

Linear component: $\eta_i = \log(n_i) + x_i^T\beta$

Skin cancer (Ch 9.3.1)

Tumor type and site for 400 patients.

	Head & neck	Trunk	Extremities	Total
Type 1	22	2	10	34
Type 2	16	54	155	185
Type 3	19	33	73	125
Type 4	11	17	28	56
Total	68	106	226	400

- Any association between tumor type and site?

Randomized control trail of influenza vaccine (Ch 9.3.2)

- Patients randomly chosen to a group; vaccine or placebo.
- Response: Levels { Small, Moderate, Large } of antibody found six weeks later.

	Small	Moderate	Large	Total
Placebo	25	8	5	38
Vaccine	6	18	11	35

- Does the response patter differ between the groups?

Case control study ulcer type and aspirin use (Ch 9.3.3.)

Case: Ulcer patients

Control: Without known ulcer, and similar to case group wrt age, sex, etc

	No aspirin	Aspirin	Total
Gastric ulcer			
Control	62	6	68
Cases	39	25	64
Duodenal ulcer			
Control	53	8	61
Cases	49	8	57

1. Gastric ulcer associated with aspirin use?
2. Duodenal ulcer associated with aspirin use?
3. Any association same for the two ulcer sites?

Probability models for contingency tables

$y = (y_1, \dots, y_N)$, frequency's in N cells.

Poisson

- No constraints on Ys
- $Y_i \sim Po(\mu_i)$, $\eta_i = \log(\mu_i)$

Multinomial, ex skin cancer

- Constraint: $\sum_{i=1}^{N} Y_i = n$
- $Y \sim m(n, \theta_1, \dots, \theta_N)$, $\eta_i = \log(n) + \log(\theta_i)$

Product multinomial, ex ulcer

- Constraint: $\sum_{j=1}^{J} \sum_{k=1}^{K} Y_{jkl} = n_{jk}$
- $Y_{jk} \sim m(n_{jk}, \theta_{jk1}, \dots, \theta_{jkL})$, $\eta_{jk} = \log(n_{jk}) + \log(\theta_{jk})$

Overdispersion

Response:

- $Y_i \sim Po(\mu_i)$, $\Rightarrow E(Y_i) = \mu_i$ and $Var(Y_i) = \mu_i$
- $Y_i \sim Bin(n_i, p_i) \Rightarrow E(Y_i) = n_i p_i$ and $Var(Y_i) = n_i p_i(1 - p_i)$

We fit $E(Y_i)$, which gives variance.

Overdispersion: $Var(Y_i) > E(Y_i)$

Possible reasons:

- Relevant explanatory variables omitted.
- Dependent observations (know why? use GLMMs, Chp 11)

Possible solution:

Include extra parameter, overdispersion parameter: $Var(Y_i) = \phi E(Y_i)$

- $Y_i \sim Bin()$ use *quasi binomial*
- $Y_i \sim Po()$ use *negative binomial*