MATH 523: EXAMPLE - POISSON DATA

An investigation of the incidence of the occurrences of four types of tumor at three different body locations involved recording counts from 400 randomly sampled cancer registry records.

- Tumor type: `type`
  1 Hutchinsons melanotic freckle
  2 Superficial spreading melanoma
  3 Nodular
  4 Indeterminate
- Location: `site`
  1 Head and Neck
  2 Trunk
  3 Extremities

		Site			
		1	2	3	Total
Tumor	1	22	2	10	34
	2	16	54	115	185
	3	19	33	73	125
	4	11	17	28	56
	Total	68	106	226	400

A Poisson GLM can be used to analyze these data: the response data are the counts of numbers of tumors observed in each of the $4 \times 3 = 12$ cross-categorized groups defined by the levels of the two factors.

Using a canonical log link function, four models can be fitted for the linear predictor,

- Model M_0: Null model
- Model M_1: Tumor type only main effect model
- Model M_2: Site only main effect model
- Model M_3: Tumor type plus Site additive main effects model

where all models are nested inside M_3, where for $j = 1, \ldots, 4, k = 1, \ldots, 3$

$$\log \mu_{jk} = \log \mu + \alpha_j + \beta_k$$

with the usual identifiability constraints. These models are written

$$\texttt{1}, \texttt{tumor}, \texttt{site}, \texttt{tumor+site}$$

respectively. Fits of the four models are summarized in the following table: p is the total number of parameters in the fitted model, Dev. is residual deviance, and rdf is the residual degrees of freedom.

Model	p	Dev.	rdf
M_0	1	295.200	11
M_1	4	150.100	8
M_2	3	196.900	9
M_3	6	51.795	6