

## Exercise 22: Poisson regression – cancer data

The dataset cancer gives lung cancer cases and person-years at risk for British physicians. The data are from a study of lung cancer incidence as a function of cigarette smoking, and we will use it to illustrate Poisson regression<sup>13</sup>. The data are available on the course web-page, and a summary is given in the table below.

Column	Contains
1	Age groups. Values are midpoints of the age groups 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, and 75-79
2	Number of cigarettes smoked per day. Each value is the average within a range; the ranges are 0, 1-9, 10-14, 15-19, 20-24, 25-34, and 35+.
3	Number of man-years at risk in that category.
4	Number of cases of lung cancer.

1. Perform poisson regression on the cancer data with number of cases being responsvariable and years of cigarette smoking and number of cigarettes smoked per day being covariates.
2. Are the two covariates significant important? Perform a formal test for each of the covariates.
3. Extend the model from 1) by including second order and interaction terms. Use step-wise regression procedures for chosing the “best” model.

<sup>13</sup>From E. L. Frome (1983), The analysis of rates using Poisson regression models, Biometrics Vol 39, 665–674.

Age	Cigarettes per day						
	Non-smokers	1-9	10-14	15-19	20-24	25-34	> 35
15-20	1 (10366)	0 (3121)	0 (3577)	0 (4317)	0 (5683)	0 (3042)	0 (670)
20-25	0 (8162)	0 (2397)	1 (3286)	0 (4214)	1 (6385)	1 (4050)	0 (1166)
25-30	0 (5969)	0 (2288)	1 (2546)	0 (3185)	1 (5483)	4 (4290)	0 (1482)
30-35	0 (4496)	0 (2015)	2 (2219)	4 (2560)	6 (4687)	9 (4268)	4 (1580)
35-40	0 (3152)	1 (1648)	0 (1826)	0 (1893)	5 (3646)	9 (3529)	6 (1336)
40-45	0 (2201)	2 (1310)	1 (1386)	2 (1334)	12 (2411)	11 (2424)	10 (924)
45-50	0 (1421)	0 (927)	2 (988)	2 (849)	9 (1567)	10 (1409)	7 (556)
50-55	0 (1121)	3 (710)	4 (684)	2 (470)	7 (857)	5 (663)	4 (255)
55-60	2 (826)	0 (606)	3 (449)	5 (280)	7 (416)	3 (284)	1 (104)

Table: Lung cancer deaths (with the number of person-years in parenthesis) broken down on age and the daily cigarette consumption.