9.2 The data in Table 9.13 are numbers of insurance policies, n, and numbers of claims, y, for cars in various insurance categories, CAR, tabulated by age of policy holder, AGE, and district where the policy holder lived ($DIST = 1$, for London and other major cities, and $DIST = 0$, otherwise). The table is derived from the $CLAIMS$ data set in Aitkin et al. (2005) obtained from a paper by Baxter et al. (1980).

(a) Calculate the rate of claims y/n for each category and plot the rates by AGE, CAR and $DIST$ to get an idea of the main effects of these factors.

(b) Use Poisson regression to estimate the main effects (each treated as categorical and modelled using indicator variables) and interaction terms.

(c) Based on the modelling in (b), Aitkin et al. (2005) determined that all the interactions were unimportant and decided that AGE and CAR could be treated as though they were continuous variables. Fit a model incorporating these features and compare it with the best model obtained in (b). What conclusions do you reach?

Table 9.13 *Car insurance claims: based on the CLAIMS data set reported by Aitkin et al. (2005).*

		$DIST = 0$		$DIST = 1$	
CAR	AGE	y	n	y	n
1	1	65	317	2	20
1	2	65	476	5	33
1	3	52	486	4	40
1	4	310	3259	36	316
2	1	98	486	7	31
2	2	159	1004	10	81
2	3	175	1355	22	122
2	4	877	7660	102	724
3	1	41	223	5	18
3	2	117	539	7	39
3	3	137	697	16	68
3	4	477	3442	63	344
4	1	11	40	0	3
4	2	35	148	6	16
4	3	39	214	8	25
4	4	167	1019	33	114