Department of Statistics

COURSE STATS 330/762

Assignment 3, 2011

Instructions: Hand in your completed assignment to the Student Resource Centre by **4pm September 15th.**

The data set for this assignment is in the file **insurance.txt** which is available on the course web page.

These data derive from a civil rights legal case in the city of Chicago in the 1970's. A civil rights group alleged that certain insurance companies were denying Chicago residents insurance on the basis of their race. The complicating factor is that if applicants lived in disadvantaged neighbourhoods, they were more likely to be denied insurance. In this case, the residents were forced to accept a more modest insurance policy from the city, the so-called FAIR plan.

Data were collected at the zip-code level – the zip code is the US postal code which serves to define small areas. The variables below refer to those areas.

The variables in the data set are

race:	racial composition in percent minority
fire:	fires per 100 housing units
theft:	theft per 1000 population
age:	percent of housing units built before 1939
involact:	new FAIR plan policies and renewals per 100 housing units
income:	median family income in thousands of dollars
side:	North or South side of Chicago

Your task in this assignment is to write me a report of not more than four pages (excluding appendices). The report should attempt to provide an answer to the question "Is the proportion of new FAIR plan policies in the zip code district related to the racial composition"? To answer the question, you should fit a regression model to the data, taking whatever remedial steps seem necessary. When you are satisfied with your model, you should interpret the coefficients in the appropriate way.

Your report should consist of the following parts: (1) an executive summary, (2) a main part (introduction, main section, conclusions) and a technical appendix, containing the details of the statistical analysis, including the R code. Don't make

the first two parts too technical. You may find the presentation "Report writing presentation" on the course web site helpful.

The assignment is worth 40 marks, split up as follows:

Report: good structure 10 marks, clear arguments 10 marks
Analysis: Thorough analysis, fixing up any deficiencies in the model: 10 marks.
Correct conclusions drawn from the analysis: 10 marks.

The data are shown below: note that the 5-digit zip codes (e.g.60626) are row labels.

	race	fire	theft	age	involact	income	side
60626	10.0	6.2	29	60.4	0.0	11.744	n
60640	22.2	9.5	44	76.5	0.1	9.323	n
60613	19.6	10.5	36	73.5	1.2	9.948	n
60657	17.3	7.7	37	66.9	0.5	10.656	n
60614	24.5	8.6	53	81.4	0.7	9.730	n
60610	54.0	34.1	68	52.6	0.3	8.231	n
60611	4.9	11.0	75	42.6	0.0	21.480	n
60625	7.1	6.9	18	78.5	0.0	11.104	n
60618	5.3	7.3	31	90.1	0.4	10.694	n
60647	21.5	15.1	25	89.8	1.1	9.631	n
60622	43.1	29.1	34	82.7	1.9	7.995	n
60631	1.1	2.2	14	40.2	0.0	13.722	n
60646	1.0	5.7	11	27.9	0.0	16.250	n
60656	1.7	2.0	11	7.7	0.0	13.686	n
60630	1.6	2.5	22	63.8	0.0	12.405	n
60634	1.5	3.0	17	51.2	0.0	12.198	n
60641	1.8	5.4	27	85.1	0.0	11.600	n
60635	1.0	2.2	9	44.4	0.0	12.765	n
60639	2.5	7.2	29	84.2	0.2	11.084	n
60651	13.4	15.1	30	89.8	0.8	10.510	n
60644	59.8	16.5	40	72.7	0.8	9.784	n
60624	94.4	18.4	32	72.9	1.8	7.342	n
60612	86.2	36.2	41	63.1	1.8	6.565	n
60607	50.2	39.7	147	83.0	0.9	7.459	n
60623	74.2	18.5	22	78.3	1.9	8.014	s
60608	55.5	23.3	29	79.0	1.5	8.177	s
60616	62.3	12.2	46	48.0	0.6	8.212	s
60632	4.4	5.6	23	71.5	0.3	11.230	s
60609	46.2	21.8	4	73.1	1.3	8.330	s
60653	99.7	21.6	31	65.0	0.9	5.583	s
60615	73.5	9.0	39	75.4	0.4	8.564	s
60638	10.7	3.6	15	20.8	0.0	12.102	s
60629	1.5	5.0	32	61.8	0.0	11.876	s
60636	48.8	28.6	27	78.1	1.4	9.742	s
60621	98.9	17.4	32	68.6	2.2	7.520	s
60637	90.6	11.3	34	73.4	0.8	7.388	s
60652	1.4	3.4	17	2.0	0.0	13.842	s
60620	71.2	11.9	46	57.0	0.9	11.040	s

60619	94.1	10.5	42	55.9	0.9	10.332	s
60649	66.1	10.7	43	67.5	0.4	10.908	s
60617	36.4	10.8	34	58.0	0.9	11.156	s
60655	1.0	4.8	19	15.2	0.0	13.323	s
60643	42.5	10.4	25	40.8	0.5	12.960	s
60628	35.1	15.6	28	57.8	1.0	11.260	s
60627	47.4	7.0	3	11.4	0.2	10.080	s
60633	34.0	7.1	23	49.2	0.3	11.428	s
60645	3.1	4.9	27	46.6	0.0	13.731	n