Possible points: 100

1. IGLM book (ed. 2), Exercise 7.1 [20 pts]

The number of deaths from leukemia and other cancers among survivors of the Hiroshima atom bomb are shown in Table 7.11 of IGLM (edition 2), classified by the radiation dose received. The data refer to deaths during the period 1950 – 1959 among survivors who were aged 25 to 64 years in 1950 (from data set 13 of Cox and Snell, 1981, attributed to Otake, 1979).

Obtain a suitable model to describe the dose-response relationship between radiation and the proportional mortality rates for leukemia.

Table 1: Table 7.11 from the IGLM book (ed. 2): Deaths from leukemia and other cancers classified by radiation dose received from the Hiroshima atomic bomb.

	Radiation dose (rads)					
Deaths	0	1-9	10-49	50-99	100-199	200+
Leukemia	13	5	5	3	4	18
Other cancers	378	200	151	47	31	33
Total cancers	391	205	156	50	35	51

2. IGLM book (ed. 2), Exercise 9.1 (the last two parts only) [20 pts]

Let $Y_1, ..., Y_n$ be independent random variables with $Y_i \sim Poisson(\mu_i)$ and $\log \mu_i = \beta_1 + \sum_{j=2}^{J} x_{ij}\beta_j, \quad i = 1, ..., n.$

(a) Show that for maximum likelihood estimates, $\hat{\mu}_i$, it holds that $\sum \hat{\mu}_i = \sum y_i$.

(b) Deduce that the expression for the deviance, $D = 2(l(b_{max}; y) - l(b; y))$ can be written as:

$$D = 2\sum[o_i \log(o_i/e_i) - (o_i - e_i)]$$

and can be further simplified to:

$$D = 2\sum[o_i \log(o_i/e_i)]$$

in this case. (Recall that β_{max} denotes the parameter vector for the "saturated model" and that b_{max} denotes the maximum likelihood estimator of β_{max}.)

3. IGLM book (ed. 2), Exercise 9.2 [30 pts]

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 CLAIM S data set in Aitkin et al. (1989) obtained from a paper by Baxter, Coutts and Ross (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. (1989) 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 2: Table 9.13 from the IGLM book (ed. 2): Car insurance claims: based on the CLAIMS data set reported by Aitkin et al. (1989).

		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

4. Rats growth data [30 pts]

These data are weights (Gelfand et al (1990)) for 30 baby rats, measured every week for 5 weeks. Let's denote by Y_{jk} the weight of the j-th rat at age x_{jk} (in days) (where $j = 1, ..., 30$ and $k = 1, ..., 5$). The data are shown in Table 3.

(a) Conduct an exploratory analysis of these data. Provide plots of rat-specific growth trajectories (piecewise linear connection between observations for each rat, in a single figure). Provide summaries of individual rat linear models.

(b) Fit the "pooled" linear model

$$E(Y_{jk}) = \mu_j = \alpha + \beta x_{jk}, \quad \text{where } Y \sim N(\mu_j, \sigma^2)$$

assuming the random variables Y_{jk} are independent (i.e., ignoring the repeated measures on the same rat, i.e., the within-rat correlation). Provide summary of this pooled model, and check its model diagnostics. Discuss the adequacy of the linear model assumptions.

(c) Perform the simplified analysis: take only one single summary measure per rat (for example, select a single observation ($weight, age$) per rat at random) Fit a linear model to this "reduced" dataset. Comment on the diagnostics for this model.

(d) Compare the "pooled" estimates of the intercept α and slope β and their standard errors, with the results you obtain using the simplified analysis above.

(e) Fit a random effects model with random slope and intercept: $E(Y_{jk}) = \mu_j = \alpha_j + \beta_j x_{jk}$. Interpret the findings and discuss the validity of the model.

(f) Compare the results you obtain from each approach. Which method(s) do you think are most appropriate? Why?

Table 3: Rat growth: data from Gelfand et al. (1990).

	Age				
Rat	8	15	22	29	36
1	151	199	246	283	320
2	145	199	249	293	354
3	147	214	263	312	328
4	155	200	237	272	297
5	135	188	230	280	323
6	159	210	252	298	331
7	141	189	231	275	305
8	159	201	248	297	338
9	177	236	285	350	376
10	134	182	220	260	296
11	160	208	261	313	352
12	143	188	220	273	314
13	154	200	244	289	325
14	171	221	270	326	358
15	163	216	242	281	312
16	160	207	248	288	324
17	142	187	234	280	316
18	156	203	243	283	317
19	157	212	259	307	336
20	152	203	246	286	321
21	154	205	253	298	334
22	139	190	225	267	302
23	146	191	229	272	302
24	157	211	250	285	323
25	132	185	237	286	331
26	160	207	257	303	345
27	169	216	261	295	333
28	157	205	248	289	316
29	137	180	219	258	291
30	153	200	244	286	324

```
rats <- list(N = 30, T = 5,
    Y = structure(c(151, 145, 147, 155, 135, 159, 141, 159, 177, 134,
    160, 143, 154, 171, 163, 160, 142, 156, 157, 152, 154, 139, 146,
    157, 132, 160, 169, 157, 137, 153, 199, 199, 214, 200, 188, 210,
    189, 201, 236, 182, 208, 188, 200, 221, 216, 207, 187, 203, 212,
    203, 205, 190, 191, 211, 185, 207, 216, 205, 180, 200, 246, 249,
    263, 237, 230, 252, 231, 248, 285, 220, 261, 220, 244, 270, 242,
    248, 234, 243, 259, 246, 253, 225, 229, 250, 237, 257, 261, 248,
    219, 244, 283, 293, 312, 272, 280, 298, 275, 297, 350, 260, 313,
    273, 289, 326, 281, 288, 280, 283, 307, 286, 298, 267, 272, 285,
    286, 303, 295, 289, 258, 286, 320, 354, 328, 297, 323, 331, 305,
    338, 376, 296, 352, 314, 325, 358, 312, 324, 316, 317, 336, 321,
    334, 302, 302, 323, 331, 345, 333, 316, 291, 324), .Dim = c(30, 5)),
    x = c(8.0, 15.0, 22.0, 29.0, 36.0))
}
```