# 189.523B Assignment #4 (handed out 2001.03.19)
# Due Monday 2 April 2001 in class

*Notes*:

- This Assignment comprises 5 questions on 3 pages and 1 attachment.
- The data for questions 3, 4 and 5 are available in R format with extension `.dput` or in text format with extension `.dat` with the variable names the first row.
- Data files in R format can be read in using the `dget` command and `attach`'ed as in Assignment 1.

**1.** Data were collected following an outbreak of food poisoning after a party thrown by an insurance company. Out of the 320 guests, 304 responded to the questionnaire. Suspicion as to the cause of the food poisoning was eventually narrowed to the crabmeat and the potato salad, yielding the following table:

| | Food eaten | | | |
|---|---|---|---|---|
| | Crabmeat | | No crabmeat | |
| Consumer's | Potato salad | | Potato salad | |
| illness | Yes | No | Yes | No |
| Ill | 120 | 4 | 22 | 0 |
| Not ill | 80 | 31 | 24 | 23 |

*Source:* Korff, Taback & Beard (1952)

Is the food poisoning outbreak attributable to the crabmeat, the potato salad, or a combination of both?

**2.** [See the "Sex bias in admission" attached sheet.]
Analyze these data, assuming that the proportion of students admitted has a binomial distribution with $n$ =number of applicants.

**(a)** Test for a sex effect, ignoring the fact that the admission rates are different for different regions.

**(b)** **(i)** Test now for a sex effect, allowing for an effect from the type of major. Are your conclusions similar to those in (a)?

**(ii)** Explain how the model that you just fitted support the comment in the attachment that

> The first two majors were easy to get into.(...) The other four majors were much harder to get into. (...)

**(c)** Is there in fact an effect from the type of major?

**(d)** Does the sex bias (if it exists) depend on the major, or is it the same for all majors?

**3.** The table below gives the number of births, the number of congenital malformations and the amount of the herbicide 245-T sprayed (in kg/km$^2$/year) for three New Zealand South Island regions (Cities, West and East) during six time intervals. (The data is available under *nz245t.**.)

| | Cities | | | West | | |
|---|---|---|---|---|---|---|
| Period | Births | Malformations | 245-T | Births | Malformations | 245-T |
| 1959-65 | 6741 | 43 | 0.0 | 4533 | 70 | 0.0 |
| 1972 | 1034 | 16 | 0.2857 | 537 | 8 | 0.0857 |
| 1973 | 1028 | 15 | 1.0416 | 525 | 6 | 0.4475 |
| 1974 | 978 | 11 | 1.6757 | 552 | 3 | 0.6254 |
| 1975 | 1005 | 21 | 1.3370 | 457 | 7 | 0.5887 |
| 1976 | 1066 | 25 | 0.9835 | 497 | 5 | 0.8211 |

| | East | | |
|---|---|---|---|
| Period | Births | Malformations | 245-T |
| 1959-65 | 3726 | 35 | 0.0 |
| 1972 | 411 | 6 | 0.8006 |
| 1973 | 421 | 7 | 1.1820 |
| 1974 | 402 | 8 | 2.3422 |
| 1975 | 345 | 13 | 2.0095 |
| 1976 | 356 | 13 | 2.3290 |

The herbicide was introduced after 1965, so no herbicide was used in the first period. We wish to study the effect of herbicide on malformation proportion.

**(a)** Plot malformation proportion against herbicide usage. What is the apparent conclusion?

**(b)** Test the hypothesis that herbicide usage has an effect on the proportion of malformations, taking into account:

  **(i)** a constant term only;

  **(ii)** an effect due to time period;

  **(iii)** the possibility that regions may have different malformation proportions.

  What is your overall conclusion?

**(c)** Explain why it is not possible to test for the effect of herbicide in the presence of a region and period interaction.

**(d)** Is the herbicide effect the same for all periods? For all regions? Test these hypotheses.

**4.** Feigl & Zelen (1965) consider the survival of patients with acute myelogenous leukemia. The patients were divided into two groups, AG+ and AG-, according to the presence or absence respectively of a morphological character in the white blood cell (covariate $AG$). The white blood cell count (covariate $WBC$) was also recorded. Note that for leukemia, a higher blood cell count is usually associated with a poorer prognosis (i.e. a smaller probability of surviving a given length of time).

The data are available in file *wbc.**. AG is coded as a factor in *wbc.dput* with levels "pos" and "neg" for AG+ and AG- respectively. In *wbc.dat*, AG is in the first column, with values of 1 and 2 corresponding to AG+ and AG- respectively. Assume that the survival time (variable *surv*), in weeks, has a Gamma distribution, with a log link between expected survival and the linear models below.

**(a)** Carry out suitable tests to see if the mean survival time depends on:

**(i)** the log of the white blood cell count;

**(ii)** whether the patient is AG+ or AG-.

**(b)** Is the effect of log($WBC$) the same for AG+ and AG- patients?

**(c)** Suppose that the true distribution of the survival times is Exponential. How would you adjust the estimates and standard errors fitted from the larger model in part **(b)** to reflect this fact?

**(d)** Test to see if the assumption that the data have an Exponential distribution is satisfied.

**5.** Consider the motorettes example in the hand-out (the data are available in file *motors.**.).

**(a)** Fit a log-linear model, assuming that the time to failure has an Exponential distribution.

**(b)** Predict the mean time to failure for a motorette run at 130°C, which is the design temperature of interest in the experiment.

**(c)** What is the probability of a motor running at 130°C failing in the first year?

*End of Assignment 4.*