ST104

UNIVERSITY OF WARWICK

December 1998

STATISTICAL LABORATORY I: Practical Examination

Open book exam: all course notes and books are allowed.

ALL questions should be attempted.

Your answer to each question should contain a final section headed: "Constructive Criticism".

Time allowed: 2 hours

Read carefully the instructions on the answer book and make sure that the particulars required are entered on each answer book.

---

1. Between 16 December 1902 (when an earthquake killed 4,500 in Turkestan) and 4 March 1977 (when an earthquake killed 2,000 in Romania), there were another 61 major earthquakes. Here a 'major earthquake' is one in which over 1,000 were killed, or whose magnitude was at least 7.5 on the Richter scale. The 62 'waiting times' between consecutive earthquakes are shown in the following table, reading row-by-row:

Time intervals between major earthquakes (days)									
840	157	145	44	33	121	150	280	434	736
584	887	263	1901	695	294	562	721	76	710
46	402	194	759	319	460	40	1336	335	1354
454	36	667	40	556	99	304	375	567	139
780	203	436	30	384	129	9	209	599	83
832	328	246	1617	638	937	735	38	365	92
82	220								

   (a) Produce a stem-and-leaf plot of the data and comment on its shape.

   (b) Assuming that the waiting times are independent and identically distributed (with an unknown underlying distribution F), estimate the mean, median and mode of this distribution F.

   (c) What type(s) of distribution would you consider for modelling F? Briefly justify your choice(s).

   (d) *Without carrying out any further large calculations*, outline how you would estimate the unknown parameter(s) of F. How you would then check whether the numbers of observed waiting times below 50 days, or above 5 years, are reasonable given your model?

Continued...

2. The amount of nitrogen-bound bovine serum albumin produced by three groups of mice was measured. The groups were: normal mice treated with a placebo (i.e. an inert substance), alloxan-diabetic mice treated with a placebo, and alloxan-diabetic mice treated with insulin. The resulting data, as reported in 'Elements of Statistics' by F. Daly, D. J. Hand, M. C. Jones, A. D. Lunn and K. J. McConway, Addison-Wesley 1995, were:

Normal + placebo	Alloxan-diabetic + placebo	Alloxan-diabetic + insulin
156	391	82
282	46	100
197	469	98
297	86	150
116	174	243
127	133	68
119	13	228
29	499	131
253	168	73
122	62	18
349	127	20
110	276	100
143	176	72
64	146	133
26	108	465
86	276	40
122	50	46
455	73	34
655		44
14		

(a) Produce appropriate graphical display(s) and numerical summaries of these data, and comment on what can be learnt from these.

(b) Carry out a two-sample t-test to check whether insulin treatment can be assumed to have the same effect as a placebo on alloxan-diabetic mice.

(c) Include your 'constructive criticism'

Continued...

3. A survey of differences in mortality between different occupations for working males in England and Wales produced the following table of values for the standardized smoking ratio and the standardized mortality ratio (SMR) for lung cancer:

	Occupational group	Smoking ratio (x)	Lung cancer SMR (y)
1	Farmers, foresters, fishermen	77	84
2	Miners and quarrymen	137	116
3	Gas, coke and chemical makers	117	123
4	Glass and ceramics makers	94	128
5	Furnace, forge, foundry, rolling mill workers	116	155
6	Electrical and electronic workers	102	101
7	Engineering and allied trades not included elsewhere	111	118
8	Woodworkers	93	113
9	Leather workers	88	104
10	Textile workers	102	88
11	Clothing workers	91	104
12	Food, drink and tobacco workers	104	129
13	Paper and printing workers	107	86
14	Makers of other products	112	96
15	Construction workers	113	144
16	Painters and decorators	110	139
17	Drivers of stationary engines, cranes, etc.	125	113
18	Labourers not included elsewhere	133	146
19	Transport and communication workers	115	128
20	Warehousemen, storekeepers, bakers, bottlers	105	115
21	Clerical workers	87	79
22	Sales workers	91	85
23	Service, sport and recreation workers	100	120
24	Administrators and managers	76	60
25	Professional, technical workers, artists	66	51

In this table, the occupational groups are chosen so that every working male appears in one and only one group, and x and y are adjusted to allow for the different age patterns across the occupational groups. Thus, if all ages of workers in a given group have the same risk of lung cancer as in the general population, then $y = 100$. Similarly, if each age group smokes on average 10% more cigarettes per individual than the same age group in the general population, then $x = 110$.

(a) Produce an appropriate scatterplot of the data, and add any enhancements that you think help in interpreting the plot (for example, you might decide to plot the line $y = x$).

(b) Identify any occupational groups that are extreme in any way, and comment on the apparent relationship between smoking and lung cancer.

(c) Your 'constructive criticism' should include, amongst other things, a discussion of other explanatory variables that may be useful in predicting y.

END.