STATISTICS MASTER'S/PH.D. QUALIFYING EXAM: Take-home Portion
Handed out January 7, 2002. Due: January 11, 2002 at noon.

DIRECTIONS: Turn in your answers by noon on Friday, January 11, 2002, to the main office or to the proctor of the exam. Answer each question in two or less typed pages; you are encouraged to create a brief, well-organized appendix for each answer. You are to work on the exam independently.

1 Kangaroo skull data

The data to be analyzed here are selected skull measurements on 148 kangaroos of known sex and species. There are 11 columns of data, corresponding to the following features. Columns, from left to right:

1. sex (1=M, 2=F)
2. species (0=M. giganteus, 1=M. f. melanops, 2=M. f. fuliginosus)
3. post orbit width
4. rostal width
5. supra-occipital - paroccipital depth
6. crest width
7. incisive foramina length
8. mandible length
9. mandible width
10. mandible depth
11. ascending ramus height (cols 3-11 are in mm times 10)

A partial data set is given below; the full data set may be found at

`http://www.math.unm.edu/~hanson/kangaroo.txt`

Some of the observations in the data set are missing. These are represented by a period.

We are interested in you building a model, or models, that relates the mandible width (i.e. the response) to the mandible length and mandible depth. Furthermore, you need to quantify how this relationship depends, if at all, on both the species and the sex of the kangaroo.

You can use any tools to answer this question. However, there is a lot you can do with this problem using simple plots and simple analyses.

Write up a careful summary of your analysis, and provide a careful summarization of your findings. Partial data:

sex	sp	pow	rw	sopd	cw	ifl	ml	mw	md	arh
1	0	249	227	531	153	88	1086	131	179	591
1	0	233	248	632	141	100	1158	148	181	643
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	2	236	289	680	145	106	1334	153	211	739
2	2	253	291	699	188	103	1354	153	223	807

2 Fat in diets data

A researcher studied the effects of three experimental diets ($j = 1, 2, 3$) with varying fat contents on the total lipid (fat) level in blood plasma; total lipid level is a widely used predictor of coronary heart disease. Fifteen male subjects who were within 20% of their ideal body weight were grouped into five blocks ($i = 1, 2, 3, 4, 5$) according to age. Within each block, the three experimental diets were randomly assigned to the three subjects. Data on reduction in lipid level (in grams per liter) after the subjects were on the diet for a fixed period of time follow.

	Fat Content of Diet		
Block	$j = 1$	$j = 2$	$j = 3$
i	Extremely Low	Fairly Low	Moderately Low
1: Ages 15-24	0.73	0.67	0.15
2: Ages 25-34	0.86	0.75	0.21
3: Ages 35-44	0.94	0.81	0.26
4: Ages 45-54	1.40	1.32	0.75
5: Ages 55-64	1.62	1.41	0.78

You are to provide a complete analysis of this data set, keeping in mind the implied goal of the researcher and the experimental design used in collecting data. In your appendix, you should provide appropriate residual plots and diagnostics with interpretation to convince a reader that their is no gross deficiencies in the assumptions for the statistical model you select. You may want consider the use of Tukey's test for additivity, if appropriate, as well as determining and analyzing contrasts of interest.