Uva Wellassa University, Sri Lanka
End Semester Examination – December 2009
BIO 441-1 Data Handling and Statistics

Time: Three (03) Hours

Answer All Questions
Graph paper and Official tables are provided
Minitab Statistical Package will be provided on personal computers

01. Wild bears were anesthetized, and their bodies were measured and weighed. One goal of the study was to make a table (or perhaps a set of tables) for hunters, so they could estimate the weight of a bear based on other measurements. This data set includes only the first measurement for each bear taken in September through November. He needs to find out the **factors that could affect the bear's weight.**

Information was collected from 50 bears and the data are stored in a Minitab work sheet named "***Bears.mtw***" given on your computer. Descriptions of the data columns are as follows.

Column	Name	Description
C1	ID	Identification number
C2	Sex	M or F
C3	Age	Bear's age, in months. Note, wild bears are always born in January, so an expert can estimate the bear's age without directly asking it how old it is.
C4	Head.L	Length of the head, in inches
C5	Head.W	Width of the head, in inches
C6	Neck.G	Girth (distance around) the neck, in inches
C7	Chest.G	Girth (distance around) the chest, in inches
C8	Weight	Weight of the bear, in pounds

I. Comment on the correlation between the weight of the bear and any other quantitative variables that you find to affect the weight of the bear. Compare the strength of relationships.
II. Formulate the best multiple regression model to represent the above data. State any assumptions regarding the model.
III. Construct the ANOVA table, test for significance of the model and state your conclusions.
IV. Interpret the estimated value of the parameters of the model.
V. Explain how you would check whether the assumptions made in part (II) are reasonable for the data.

(25 marks)

02. The hormone Leptin is thought to be important in energy expenditure in humans. In a study to investigate the effect of diet on leptin levels, the leptin levels (in ng/ml) of 19 healthy volunteers were measured before and after a 72-hour controlled diet.

The data are stored in a Minitab work sheet named "***Leptin.mtw***" given on your computer.

I. Use an appropriate parametric test to determine whether there is a difference (at the 5% significance level) between the mean leptin levels before and after the diet.

II. Suggest non-parametric tests, either of which might be used here if the distribution of within-pair differences could not reasonably be assumed to be Normal.

III. Calculate the sample mean and sample median of the differences. How do these results affect your choice of test?

(15 marks)

03. A health psychologist is studying, fear of falling in old people. He is examining the relationships between three psychometric scales using a random sample of 64 people. Scale A measures fear of falling, where a high score corresponds to high fear. Scale B measures confidence doing tasks where there is a risk of falling; here a high score corresponds to high confidence. Scale C is a measure of general anxiety, where a high score corresponds to high anxiety.

Pearson correlations from his sample are as follows.

	Scale A	Scale B
Scale B	-0.771	
Scale C	0.637	-0.328

Coefficients from multiple regressions, modeling Scale B as the dependent variable, are as follows.

Dependent variable: Scale B

	Coefficient	Standard error
Constant:	126.28	2.976
Scale A:	-1.486	0.165
Scale C:	1.439	0.566

Given the negative correlation between Scales B and C, the psychologist expected the coefficient of Scale C to be negative. Provide a possible explanation for the positive coefficient.

(05 marks)

04. In Sri Lanka, all patients surviving a stroke are supposed to have their cholesterol levels measured soon after their stroke and regularly thereafter. A sample of medical records of men and women who had suffered a stroke was examined to determine whether there was a difference between the sexes in the proportions of stroke survivors who had a recently recorded cholesterol measurement. The following data were obtained.

	Cholesterol Level Recorded	
Sex	*No*	*Yes*
Female	109	22
Male	97	77

I. Perform a suitable test of the null hypothesis that there is no association between an individual's sex and the chance of he or she having a recently recorded cholesterol measurement.

II. Compute and interpret an approximate 95% confidence interval for the difference between the proportions of females and males having a recently recorded cholesterol measurement. (Estimated variance of difference between the proportions of females and males is 0.002485)

(15 marks)

05. An experiment was conducted to investigate the toxic effects of four different chemical compounds *A*, *B*, *C* and *D* on the skin. Four adjacent regions, each a square of side 3 cm, were marked on the left forearm of each of six subjects, and each chemical was applied to each subject, choosing the sites of application of the chemicals at random for each subject. After three hours, the skin was examined and scored from 0 to 10 depending on the degree of irritation, with 0 representing no irritation and 10 severe irritations. The data are given in the following table.

Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
D 5	A 7	B 2	B 4	A 3	D 6
B 3	C 4	A 1	A 6	B 1	B 7
A 3	D 7	D 3	C 6	D 2	C 3
C 2	B 6	C 1	D 7	C 2	A 5

I. Carry out a suitable analysis of these data.

II. Explain clearly your conclusions and comment on the differences, if any, between the toxic effects of the compounds. Discuss briefly whether all the necessary assumptions can safely be made.

III. Summarize your conclusions in non-technical language which the experimenter would understand.

(20 marks)

06. A conservationist wishes to assess the effectiveness of four management treatments for promoting meadow brome *Bromus commatatus* in a hay meadow nature reserve. The treatments are as follows.

A: Hay cut and harvested (removed from site)
B: Hay flail cut and left on ground in windrows
C: Hay scythed and left *in situ*
D: Hay not cut

The meadow slopes in one direction towards a river. On another side, at right angles, it is bounded by a motorway. The experiment was arranged as a Latin square design with the columns representing distance from the river and the rows representing distance from the motorway.

I. Explain why this arrangement would be better than complete randomisation of treatments to plots. Describe in detail how you would choose at random a 4×4 Latin square design for the layout of this experiment.

II. The treatment allocation and the data obtained during a fixed period after the intended harvest date are shown in the following table. The data were transformed into units such that high values show a more effective treatment.

		(Coded) distance from river			
		1	2	3	4
(Coded) distance from motorway	1	D 34.45	A 64.16	B 53.13	C 49.02
	2	A 66.42	D 31.31	C 49.60	B 51.94
	3	C 52.54	B 54.94	A 62.73	D 25.84
	4	B 58.05	C 53.73	D 29.33	A 56.79

a. Analyze the data to determine the significance of the effects due to distance from motorway, distance from river and hay management treatment, and briefly report your results. (Made the comparison between treatments)

b. State the assumptions needed for the validity of the analysis of variance in (a). Based on these graphs (Normal probability plot of the residuals and a plot of residuals against fitted values), explain whether you consider any of the assumptions for the analysis of variance to be invalid.

(20 marks)

Dataset for Question No 01 : "*Bears.mtw*"

ID	Sex	Age	Head.L	Head.W	Neck.G	Chest.G	Weight
41	F	23	12.5	5	20.5	38	142
48	M	81	15.5	8	31	54	416
69	M	25	16	8	32	52	432
83	M	117	15.5	7.5	32	54.5	476
238	M	70	15	6.5	28	45	334
274	F	57	13.5	7	20	38	204
518	M	45	13.5	7	24	39	204
520	F	9	9	4.5	12	19	26
522	M	21	13	6	19	30	120
525	M	177	16	9.5	30	48	436
527	F	57	12.5	5	19	32	125
529	F	81	13	5	20	33	132
531	M	21	13	5	17	28	90
533	M	9	10	4	13	23	40
535	M	45	16	6	24	42	220
538	M	9	10	4	13.5	23	46
541	M	33	13.5	6	22	34	154
543	F	57	13	5.5	17.5	31	116
545	F	45	13	6.5	21	34.5	182
547	M	21	14.5	5.5	20	34	150
549	M	10	9.5	4.5	16	26	65
551	F	82	13.5	6.5	28	48	356
553	F	70	14.5	6.5	26	48	316
555	M	10	11	5	17	29	94
557	M	10	11.5	5	17	29.5	86
560	M	34	13	7	21	35	150
562	M	34	16.5	6.5	27	44.5	270
564	M	34	14	5.5	24	39	202
568	F	58	13.5	6.5	21.5	40	202
570	M	58	15.5	7	28	50	365
572	M	11	11.5	6	16.5	31	79
576	M	23	12	6.5	19	38	148
579	M	70	15.5	7	28	55	446
581	F	11	9	5	15	27	62
583	F	83	14.5	7	23	44	236
585	M	35	13.5	8.5	23	44	212
598	M	35	14.5	6.5	26	41	278
626	M	16	13.5	6.5	22	36	190
630	M	24	13	6	19.5	31	140
644	M	70	13	5.5	19.5	35	162
648	M	19	13	5.5	19.5	32.5	126

5

659	M	75	17	8.5	30.5	48.5	368
661	M	36	12	5.5	18	32	116
663	F	87	14	7	21	37	160
665	M	41	13	6.5	20.5	36.5	154
670	M	71	16	7.5	28	45	316
673	F	36	13.5	5.5	19.5	35	158
675	F	23	12.5	5.5	19	32	120
679	M	81	15.5	7.5	25.5	43	324
681	M	56	14.5	7	22	38	196

Dataset for Question No 02 : "*Leptin.mtw*"

Leptin level before diet (ng/ml)	Leptin level after diet (ng/ml)
11.58	12.54
13.47	12.96
13.62	9.84
8.94	9.84
10.80	12.45
17.91	12.33
11.22	11.88
16.62	18.24
9.12	14.67
11.31	14.01
12.12	15.45
15.12	15.60
9.45	11.67
11.34	13.08
12.12	13.80
15.78	16.56
9.78	9.54
12.60	13.83
13.23	14.61