GX2035

OLLSCOIL NA hÉIREANN, GAILLIMH
NATIONAL UNIVERSITY OF IRELAND, GALWAY

SEMESTER II EXAMINATIONS 2003-2004

MA 307 –BIOSTATISTICS
Dr. T.C. Bailey
Dr. J. Newell

Time allowed: ***Two* hours.**
*Answer **ALL QUESTIONS** in Section A- each of which is worth 10 **MARKS**. Answer **ALL QUESTIONS** in Section B – each question is worth 2 **MARKS** with a loss of 1 **MARK** for a wrong answer.*

Section A.

1. In a survey of the relationship between the smoking and drinking habits of students, a random sample of 504 students at an Irish University were classified as Smokers, Ex-smokers or Non-smokers and then as Non-drinkers, Occasional Drinkers (≤ 2 units per week) or Regular Drinkers (> 2 units per week). The following (annotated) output was obtained from Minitab:

Table 1 Expected counts are printed below observed counts

	Regular Drinkers	Occasional Drinkers	Non-Drinkers	Total
Smokers	46	5	0	51
	(37.3)	(7.8)	(5.9)	
Ex-Smokers	48	9	3	60
	(43.9)	(9.2)	(6.9)	
Non-Smokers	275	63	55	393
	(287.8)	(60.0)	(45.2)	
Total	369	77	58	504

$X^2 = 14.29$, $df = 4$, $p < 0.01$

Table 2 Rowpercents

	Regular Drinkers	Occasional Drinkers	Non-Drinkers
Smokers	90	10	0
Ex-Smokers	80	15	5
Non-Smokers	70	16	14

Is there any evidence of a significant association between the smoking and drinking habits of students? If so, briefly describe the form of the association.

2. A multiple linear regression model with two explanatory variables takes the form

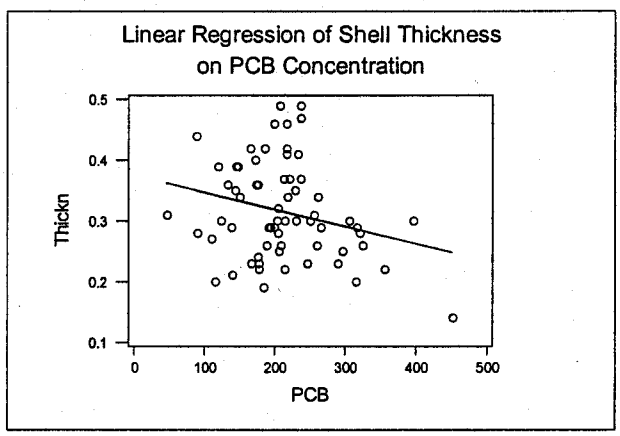
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \text{error}$$

i) Provide a clear interpretation of the parameters β_0, β_1 and β_2 in this model

ii) State the usual assumptions underlying this model.

3. In a study of the effects of industrial pollutants, a random sample of 65 Anacapa pelican eggs were selected for study. The thickness of each shell (mm) was measured along with the concentration of polychlorinated biphenyl (PCB, an industrial pollutant, in parts per million) in each shell.

A linear regression of shell thickness on PCB concentration in Minitab gave the following output:




```
MTB > Regress 'Thickn' 1 'PCB'.

The regression equation is
Thickn = 0.375 -0.000279 PCB

Predictor       Coef         Stdev     t-ratio         p
Constant     0.37494     0.02990       12.54     0.000
PCB       -0.0002790   0.0001345       -2.07     0.042

s = 0.07848      R-sq = 6.4%      R-sq(adj) = 4.9%
```

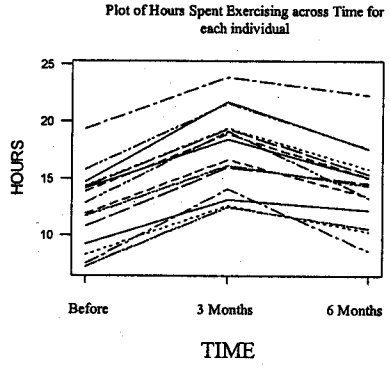
Summarise the dependence of shell thickness on PCB by providing a clear interpretation of the regression equation, the hypothesis tests, s and R^2.

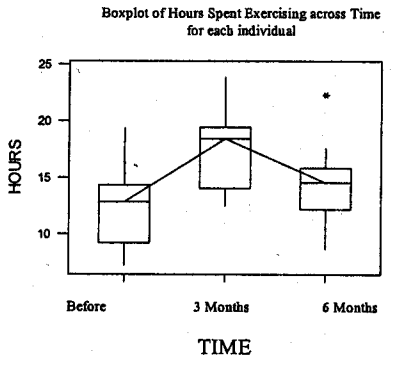
What additional information would you require to complete this analysis?

4. A local sports hall carried out a study to investigate if a new course on fitness encouraged individuals to spend more time exercising. A group of fifteen fitness conscious individuals took part in the study. Each individual had recorded the number of hours spent exercising in a typical week immediately before attending the fitness course as well as the typical number of hours spent exercising weekly 3 and 6 months after having attended the course.

What form of analysis would be appropriate here?

Appropriate plots of the data are presented below:-





What is your subjective impression from the above plots?

The results of a formal Repeated Measures Analysis of Variance and appropriate follow-up analysis are given below.

Analysis of Variance for HOURS

Source	DF	SS	MS	F	P
Time	2	203.561	101.781	155.18	0.000
Individual	14	474.190	33.871	51.64	0.000
Error	28	18.365	0.656		
Total	44	696.116			

Follow-up analysis

Set of simultaneous 95% Bonferroni Multiple Comparisons are

TimePt1	minus	TimePt2	Interval Estimate
Before		3 Months	(-5.790 , -4.547)
Before		6 Months	(-2.722 , -1.311)
3 Months		6 Months	(2.125 , 4.178)

What should be concluded from this analysis?

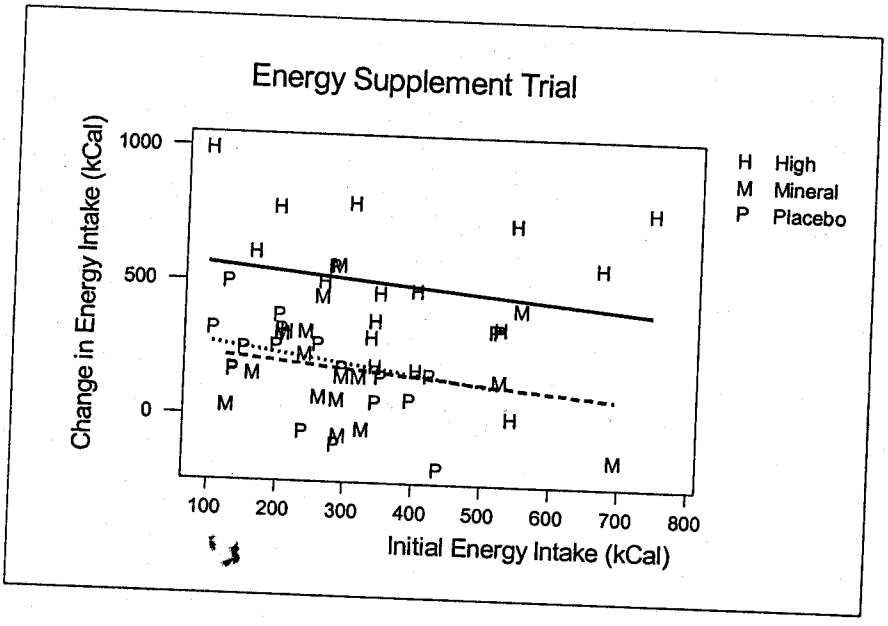
5. In a study of the relationship between nutritional supplements and Energy Intake of 1 to 2 year-old children in Third World rural communities, a sample of 53 such children were randomised to one of three supplements

(i.e. a High Energy (**H**) or a Mineral (**M**) or a Placebo (**P**) supplement)

The children had their typical daily Energy Intake (E.I.) assessed ***before*** starting a daily dose of their allocated supplement and then six months ***after.*** Plotted below are the ***changes*** (i.e. After E.I. - Before E.I.) against the **Initial** (i.e. Before) E.I. labelled by supplement and with **separate** linear regressions plotted for each supplement.

On the basis of this plot and the Minitab output from fitting separate and parallel regressions (i.e. in the usual form for an Analysis of Covariance) as well as the Multiple Comparisons of the Adjusted Mean changes in Energy Intake for the Supplements, report on which, if any, of the Supplements differ on average in terms of change in Energy Intake over a six-month period for such children.

Also discuss the usefulness of any of the regressions of Change in Energy Intake on Initial Energy Intake (parallel or otherwise).



question continued ...

Question 5. continued

Fit of Different Slopes Model

Source	DF	Seq SS	Adj SS	Adj MS	F	P
InitEI	1	1824	92893	92893	1.90	0.174
Supplmnt	2	1154373	192353	96177	1.97	0.151
Supplmnt*InitEI	2	3115	3115	1557	0.03	0.969
Error	47	2293855	2293855	48805		
Total	52	3453167				

Fit of Parallel Lines Model

Source	DF	Seq SS	Adj SS	Adj MS	F	P
InitEI	1	1824	92698	92698	1.98	0.166
Supplmnt	2	1154373	1154373	577186	12.31	0.000
Error	49	2296971	2296971	46877		
Total	52	3453167				

Multiple Comparisons of Supplements Adjusted for Initial E.I.

Supplement	Adjusted Mean (kCal)
H	510
M	172
P	194

---- these are evaluated at the covariate equal to 322

Set of Bonferroni Multiple Comparisons (simultaneous 95% confidence)

Group	minus	Group	Interval Estimate
H		M	(148 , 527)
H		P	(132 , 500)
M		P	(-200 , 160)

SECTION B MULTIPLE CHOICE SECTION

1. Which of the following would be the most appropriate for measuring the strength of relationship between height (cm) and weight (kg) based on a random sample of subjects selected from the target population:-

a) a chi square test of association;

or b) a correlation coefficient;

or c) a paired t-test?

2. A repeated measures ANOVA on a single repeated measures factor corresponding to two time periods is identical to

   a) a two sample t-test;

   or b) a paired t-test;

   or c) an analysis of covariance?

3. In a survey, data on hair and eye colour were recorded for a random sample drawn from the population. The aim was to determine if hair and eye colour are associated. The appropriate null hypothesis would be:-

   a) there is no association between hair and eye colour in the population;

   or b) there is a significant association between hair and eye colour in the population;

   or c) there is no association between hair and eye colour in this sample?

4. A study was carried out to assess the dependence of Forced Expiratory Volume; (FEV; litres) on Age (years), Height (cm), Weight (kg) of a sample of 196 men and women aged 45-64. Gender was coded in the variable 'GENDER' with codes 1: Male, 0: Female. Multiple regression analysis gave the following output:

   FEV = -1.61 - 0.0295 AGE + 0.0105 WEIGHT + 0.332 GENDER + 0.0268 HEIGHT

   The coefficient for 'GENDER' in this regression equation is interpreted as:

   a) On average, FEV is 0.332 litres higher for a male than for a female of the same age, height and weight;

   or b) On average, FEV increases by 0.332 litres for every one unit of increase in gender, if age, height and weight are held constant;

   or c) On average, FEV is 0.332 litres higher for a female than for a male of the same age, height and weight?

5. In a One-Way Analysis of Covariance the main reason for including a covariate in the model is

   a) to improve the normality assumption of the response;

   or b) to correct the response for a known effect of the covariate;

   or c) to correct the covariate for a known effect of the response?