

OLLSCOIL NA hÉIREANN, GAILLIMH
NATIONAL UNIVERSITY OF IRELAND, GALWAY

SEMESTER II EXAMINATIONS 2002-2003

MA 307 – BIOSTATISTICS

Dr. D. Harrington.
Dr. J. Newell.

Time allowed: *Two hours.*

Answer ALL QUESTIONS in Section A- each of which is worth 5 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. Correlation and linear regression are two techniques that are used in exploring the relationship between two continuous variables. Explain the distinction between these two techniques and illustrate your answer with suitable examples.

2. A multiple linear regression model with two explanatory variables, incorporating an interaction term, takes the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \text{error}$$

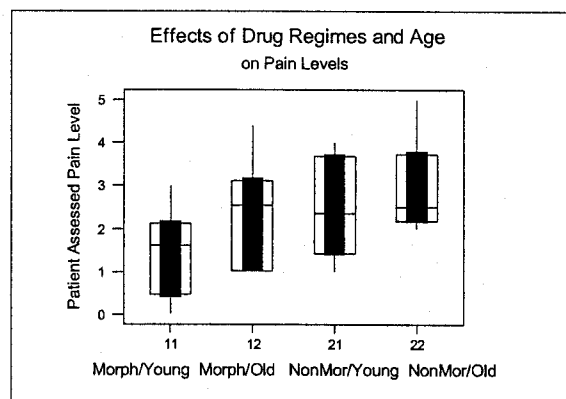
- i) Provide a clear interpretation of the parameters β_0 , β_1 , β_2 and β_3 in this model
 - ii) *State* the assumptions underlying this model.
3. Describe briefly (with illustrations) the type of context in which you would use each of the following methods of analysis:
 - i) One Way Analysis of Variance
 - ii) One Way Analysis of Covariance.

4. In a study of pain-relieving drug regimes on patients who have just undergone bypass heart surgery, samples of 20 young (<50 years) and 20 older patients were randomised to one of two post-operative drug regimes - one based on Morphine, the other not based on Morphine.

Each patient had their pain levels assessed 12 hours after the (successful) end of their operation. The resulting data are presented below in the form of Boxplots and two-way Anova table and follow-up multiple comparisons of all 4 drug/age combinations.

What, on the basis of these, should you conclude about such average pain levels with respect to the two drug regimes and to age?

Is there any alternative form of analysis that might account for the effect of Age in a better fashion than separating by whether or not one is less than 50 years old?



Analysis of Variance for Pain Score

Source	DF	SS	MS	F	P
Drug	1	6.889	6.889	5.86	0.021
Age	1	5.041	5.041	4.29	0.046
Drug*Age	1	0.625	0.625	0.53	0.471
Error	36	42.340	1.176		
Total	39	54.895			

Set of Bonferroni Multiple Comparisons for One-Way ANOVA

Group (minus)	Group	Interval Estimate
11	12	(-2.314 , 0.394)
11	21	(-2.434 , 0.274)
11	22	(-2.894 , -0.186)
12	21	(-1.474 , 1.234)
12	22	(-1.934 , 0.774)
21	22	(-1.814 , 0.894)

where Group 11 is Morphine Treated Younger patients,
 Group 12 is Morphine Treated Older patients,
 Group 21 is Non-Morphine Treated Younger patients etc.

5. In a clinical trial to compare two treatments for patients with lymphoma (a type of cancer), 273 patients were randomly allocated to receive treatment 'BP' or treatment 'CP'. The outcome for each patient was assessed by a clinician as a Complete Response (CR), Partial Response (PR), No Change (NC) or Worse (W). The following annotated output was obtained from Minitab.

Table 1 Expected counts are printed below observed counts.

		Outcome				Total
		CR	PR	NC	W	
Treatment	BP	26 (28.8)	51 (55.6)	21 (16.2)	40 (37.4)	138
	CP	31 (28.2)	59 (54.4)	11 (15.8)	34 (36.6)	135
		57	110	32	74	273

$$X^2 = 4.6, df = 3, p = 0.20$$

Table 2 Corresponding *row* percentages are:

19%	37%	15%	29%
23%	44%	8%	25%

State the null and alternative hypotheses being tested, and interpret these results.

Section B.

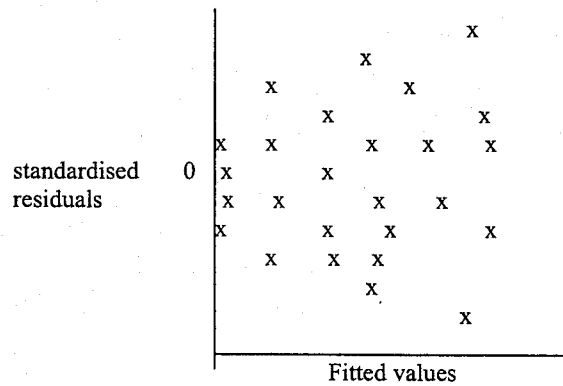
1. A population correlation of 0.8 for two random variables X and Y means that
 - A) there is no relationship between X and Y;
 - or B) there is a strong positive relationship between X and Y;
 - or C) there is a strong negative relationship between Y and X.

2. The height (cm) and vital capacity (a measurement of total lung volume, in litres) were measured for a random sample of men. The aim was to predict vital capacity from height. A linear regression was found to be appropriate and provided the following two pieces of information:
 - a 95% confidence interval for the average vital capacity of men of height 175cm.
 - a 95% prediction interval for the vital capacity of a man of height 175cm measured in the future.Compared to the confidence interval, the prediction interval will be
 - A) wider
 - or B) of equal width
 - or C) narrower

3. In multiple regression, when using 'best subsets' regression, the final choice of the best model to use should be made from among the models with
 - A) a high value of R^2_{adj} and a high value of C_p ;
 - or B) a high value of R^2_{adj} and a low value of C_p ;
 - or C) a high value of R^2 and a high value of s .

4. In a two-way Analysis of Variance the interaction between two factors measures
 - A) the difference between the two factors;
 - or B) the additive effect of the two factors;
 - or C) the extra 'combined' effect of the two factors?

5. In linear regression, residual plots are used to check the assumptions underlying the model. The following plot was obtained following a linear regression analysis of Y on X:



The most likely interpretation of this is:

- A) the relationship between Y and X is not linear;
- or B) the variability of Y is not constant for different values of X;
- or C) the assumptions underlying the linear regression are reasonable.