

OLLSCOIL NA hÉIREANN, GAILLIMH  
NATIONAL UNIVERSITY OF IRELAND, GALWAY

SUMMER EXAMINATIONS 1999

SECOND UNIVERSITY EXAMINATION

MA257 — STATISTICS

Dr. M.G. Kenward  
Professor I.G. O'Muircheartaigh  
Dr. A. McDermott

**Important notes:** Answer seven questions. You must answer *either* three questions from Part I and four questions from Part II or else four questions from Part I and three questions from Part II. All questions carry the same number of marks, 20. The total number of marks you obtain (maximum 140) will later be converted into a mark out of 100.

This paper is available ONLY to 2<sup>nd</sup> ENGINEERING candidates who are repeating the examination from a previous year.

**Part I**

1. (a) A list contains six numbers 8, 7, 14, 12, x and y. If the median of the six numbers is 10, state *any pair* of (x,y)-values which would be compatible with this result.
- (b) The frequency distribution given below summarises the sale prices (in thousand £) of 300 houses:

Sale price [£000]	Frequency
35 - 45	33
45 - 55	133
55 - 65	71
65 - 75	44
75 - 85	17
85 - 95	2

- i. Calculate the mean, the standard deviation, and the lower and upper quartiles for these data.
- ii. Given that the two highest data values are 85.2 and 86.3, and that the two lowest data values are 35.6 and 37.9, construct a boxplot to represent the data.

2. (a) Three students (Mr. A, Ms. B and Mr. C) apply for summer employment. Let A, B and C, respectively, be the events that Mr. A, Ms. B and Mr. C are offered jobs. Suppose it is known that:

$$\begin{aligned} P(A) &= .78, & P(B) &= .78 & P(C) &= .70 \\ P(A \text{ and } B) &= .58, & P(A \text{ and } C) &= .58, & P(B \text{ and } C) &= .58 \\ P(A \text{ and } B \text{ and } C) &= .48 \end{aligned}$$

- i. Are any two of the events A, B, and C mutually exclusive? Justify your answer.
  - ii. Are the events B and C independent? Justify your answer.
  - iii. What is the probability that only Mr. C will be offered a job?
- (b) A student representative committee of 3 persons is to be selected from a group of 6 men and 4 women. If the members of the committee are selected at random from the group, what are the probabilities
- i. that 3 women are selected?
  - ii. that 3 men are selected?
  - iii. that the committee is made up of a mixture of men and women?
- (c) In an industrial plant, machines A, B and C account, respectively, for 50 percent, 30 percent and 20 percent of the plant's output. Suppose that 2 percent of the units manufactured by machine A are defective, and the corresponding percentages for machines B and C are, respectively, 3 percent and 4 percent. If a unit is selected at random from a large stockpile,
- i. What is the probability that it will be defective?
  - ii. Given that the unit which is selected is defective, what is the probability that it was manufactured by machine A?

3. Experience has shown that the weekly advertising revenue of a certain local radio station follows a Normal distribution with mean £5,500 and standard deviation £500.

- (a) What is the probability that advertising revenue in a randomly selected week will exceed £4,500?
- (b) If the radio station is put up for sale, and the vendor wishes to state with probability 0.99 that weekly advertising revenue will be greater than £k, what is the value of k for which this claim will be true? [Note: you are requested to find k such that  $P(X > k) = 0.99$ , where X is the advertising revenue in a random week].
- (c) If  $\bar{X}$  is the mean advertising revenue for a random sample of 9 weeks, what is the probability that  $\bar{X}$  will lie between £5200 and £5800?
- (d) If the probability that the mean of a random sample of size n of the above weekly advertising revenues will lie between £5070 and £5930 is 0.99, what is the sample size n?

4. (a) Write down a *numerical expression* for each of the following:

- i. the number of ways the letters of the word **RANDOM** can be arranged so that the first letter is "R" *and* the two vowels are together (i.e. adjacent to each other);
- ii. the number of ways the letters of the word **RANDOM** can be arranged so that the first letter is not "R" *and* the two vowels are together (i.e. adjacent to each other);
- iii. the number of ways the letters of the word **REGISTRAR** can be arranged.

- (b) According to the Department of Health, 30% of all women will, by the age of 85, have suffered a hip fracture because of osteoporosis. Find the probability that in a random sample of five women aged 85, at least two of the women will have suffered a hip fracture due to osteoporosis [i.e. find  $P(X \geq 2)$ , where  $X$  = the number of the five women who have had a hip fracture due to osteoporosis].
- (c) If a random sample of 100 such women (i.e. aged 85) is selected, calculate, using the Normal approximation to the Binomial, the probability that more than 20 of the women will have suffered a hip fracture due to osteoporosis [i.e. approximate  $P(X > 20)$ , where  $X$  = the number of the 100 women who have had a hip fracture due to osteoporosis].

5. It was desired to estimate the population mean urinary concentration of lead in the blood of a certain group in a population. A random sample of 121 individuals was selected; the sample mean was 2.64 mmol/hr, and the sample standard deviation was 0.88 mmol/hr.

- (a) Construct a 98% confidence interval for the population mean urinary concentration of this group.
- (b) If all the information and numbers given in this question were the same, except that the sample size was 400 instead of 121, the width of the corresponding 98% confidence interval would be either
  - A: four times
  - B: one half      or
  - C: eleven twentiethsof the width of the interval constructed in part (a). State which one of A, B and C is correct.
- (c) What is the probability that the mean urinary concentration of a random sample of size 100 from the above population will be within  $\pm 0.1$  mmol/hr of the true population mean?
- (d) How large a random sample should be selected if it is desired that the mean urinary concentration of a random sample selected from this population will lie within  $\pm 0.01$  mmol/hr of the true population mean with probability 0.99?

## Part II

6. (a) If, for a given statistical hypothesis test of a null hypothesis  $H_0$  against an alternative hypothesis  $H_1$ , the probability of a type I error is 0.05, indicate for each of the following statements whether the statement is true or false (no explanation is required).

Statement A: The probability that  $H_0$  is true is 0.95. [one mark]

Statement B: The probability of rejecting  $H_0$  if it is true is 0.05 [one mark]

Statement C: Rejection of  $H_0$  implies that the probability that  $H_0$  is false is 0.95.

[one mark]

- (b) Assume that the heights (in inches) of a certain population follow a Normal distribution with unknown mean  $\mu$ , and standard deviation  $\sigma = 3$ . In order to test the null hypothesis  $H_0: \mu = 69$  against the alternative hypothesis  $H_1: \mu > 69$ , it is decided to take a random sample of  $n = 400$  individuals, and to reject  $H_0$  if  $\bar{x} > 69.25$ , where  $\bar{x}$  is the mean height for the sample.

i. For this decision rule:

A. Calculate  $\alpha = \text{Prob} [\text{Type I error}]$  [six marks]

B. Calculate  $\beta = \text{Prob} [\text{Type II error}]$ , if, in fact,  $\mu = 69.5$ . [six marks]

ii. Using the same decision rule (i.e. reject  $H_0$  if  $\bar{x} > 69.25$ ), for what value of  $\mu$  in  $H_1$  [i.e.  $> 69$ ] is the probability of a type II error equal to 0.01? [five marks]

7. An opinion poll commissioned in Northern Ireland revealed that 684 individuals out of a total random sample of 978 who were questioned supported the current peace process.

- (a) Construct an approximate 99% confidence interval for the actual percentage of the population who support the peace process. [six marks]
- (b) If it is desired to estimate this actual percentage to within 5% with probability 0.99, estimate the minimum sample size which is necessary to ensure this level of precision; you may take into account the information given in the sample in answering this part - i.e. 684 out of 978 indicated their support for the process. [seven marks]
- (c) If no knowledge at all about the actual percentage who support the peace process were available, what sample size would ensure that the same precision as that sought in part (b) of this question would be attained -- i.e. what sample size would ensure that the sample percentage would be within 5% of the actual population percentage with probability 0.99? [seven marks]

8. (a) State, without elaboration, whether the following statement is true or false:  
 "A necessary pre-requisite in order to be able to perform a valid two-sample t-test is that the two sample variances be equal".
- (b) A cannery uses two machines to produce cans of tomatoes. The plant supervisor selects and weighs 10 cans from each machine with the following results:

	Machine	
	A	B
Mean Weight (g)	225.6	218.1
Standard Deviation (g)	5.2	6.1

- (i) It is claimed that the population variability of mean weight is different for the two machines. Is there evidence in the data to support this claim?
- (ii) Show that the pooled sample standard deviation is 5.67.
- (iii) It is further claimed that the population mean weights is greater for Machine A than Machine B. Is there evidence in the data to support this claim?

9. (a) Consider the following list of hypothesis tests:  
 "Z-test for a single population mean"  
 "t-test for a single population mean"  
 "t-test for the difference between two population means when the two samples taken are independent and random"  
 "F-test for two population variances when the two samples taken are independent and random"  
 " $\chi^2$  goodness-of-fit test."

Now answer (i) and (ii) below separately:

- (i) Write down one test from the above list that could be used to determine whether or not the population of weights from which a random sample of 100 weights was selected has a population mean weight of 45 kg.
- (ii) Write down one test from the above list that could be used to test whether or not the marks of Engineering students taking a certain statistics course are more variable than those of Science students taking the same course, when independent random samples, each of size 25, of Engineering and Science students are taken, and the two populations of Engineering and Science marks in the course are assumed to follow Normal distributions.
- (b) A psychologist believes that the 4 types of personality A,B,C and D occur in the population in the ratio 6:3:3:1. Selecting 296 people at random he classifies the people as follows

	Personality Types			
	A	B	C	D
Frequencies	132	69	70	25

10. (a) State, without elaboration, whether the following statement is true or false: "A necessary pre-requisite in order to be able to perform a valid two-sample t-test is that the two sample standard deviations be equal".

[three marks for correct choice, - (minus) two marks for incorrect choice]

- (b) To compare the heights at ear emergence of 2 varieties of perennial ryegrass, 10 plants from each variety were grown in individual pots and their heights measured at ear emergence with the following results:

	Varieties	
	A	B
Mean plant height (cm)	79.11	75.67
Standard deviation	3.27	3.71

- i. It is claimed that the population variability of height at ear emergence is different for the two varieties. Is there evidence in the data to support this claim? [six marks]
- ii. Show that the pooled variance is 12.23. [three marks]
- iii. It is further claimed that the population mean height of ear emergence is greater for Variety A than for Variety B. Is there evidence in the data to support this claim? [eight marks]

11.

In a study of the relationship between cooking time,  $X$ , (in minutes) and a meat tenderness index  $Y$  (measured on a scale from 0 to 1000; the higher the value of  $Y$ , the more tender the meat) the following set of data was obtained:

Cooking time ( $x$ )	10	9	11	8	10	12	7
Tenderness ( $y$ )	500	550	425	550	450	350	625

[Note:  $\sum x_i = 67$ ;  $\sum y_i = 3,450$ ;  $\sum x_i^2 = 659$ ;  $\sum y_i^2 = 1,751,250$ ;  $\sum x_i y_i = 32,100$ ]

Assume that a population regression model of the form  $\mu_{Y|x} = \alpha + \beta x$ , together with the usual assumptions, relates these two variables.

- (a) Plot a scatter diagram to represent these data. [one mark]
- (b) Compute the least squares regression line of  $y$  on  $x$  to fit these data. [eight marks]
- (c) Based on your answer in (b), write down a point estimate of  $\mu_{Y|x=8.5}$ , [the population mean tenderness index, when the cooking time is 8.5 minutes], and calculate a 95% confidence interval for  $\mu_{Y|x=8.5}$  [five marks]
- (d) Calculate the sample correlation coefficient of  $x$  and  $y$ , and interpret its meaning. [three marks]
- (e) If the tenderness index for the sixth data point above were 550 instead of 350, would you expect the correlation coefficient for the revised data to be closer to -1.00 than that obtained in (d) above. A simple yes or no will suffice as an answer. [three marks]