MATH 244 Applied Statistics

Final Examination December 17, 2001

Time allowed : 3 hours
Answer all six questions.

1. (*22%*) The height of soap bubbles in the dishpan is of importance to soap manufacturers. An experiment was performed by varying the amount of soap and measuring the height of the bubbles in a standard dishpan after a given amount of stirring. The data are as follows:

Grams of Product (X)	4.0	4.5	5.0	5.5	6.0	6.5	7.0
Bubble Height in mm (Y)	33	43	46	51	53	61	63

You are asked to analyse the data based on the following regression model.

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad , \; i = 1,2,\ldots,7 \quad ; \quad \varepsilon_i \overset{iid}{\sim} N\left(0, \sigma^2\right)$$

(a) Fit the regression line of Y on X. (*5 marks*)

(b) Construct the ANOVA table. (*6 marks*)

(c) Calculate the coefficient of determination. (*2 marks*)

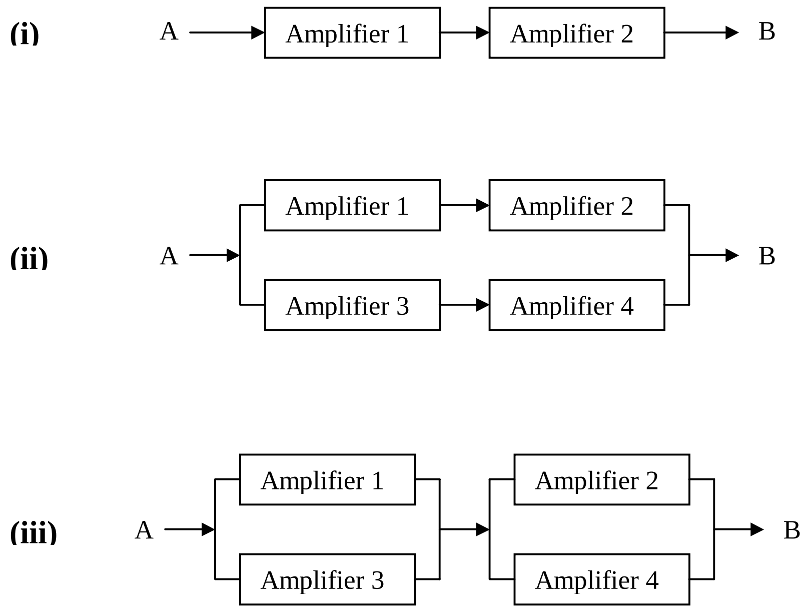
(d) Find the 95% prediction interval of the height of bubbles when 5 grams of soap product is used. (*4 marks*)

(e) Is the prediction of 45mm, based on the third pair of values, inferior to the prediction by the regression line obtained in (a)? (*2 marks*)

(f) An experimenter stated that the model $Y = \alpha + \beta X + \varepsilon$ was “a ridiculous model unless $\alpha = 0$, for anyone knows that if you don’t put any soap in the dishpan there will be no soap bubbles.” Thus he insists on using the model $Y = \beta X + \varepsilon$. Comment on the experimenter’s statement. (*3 marks*)

P. T. O.

2. (*18%*) An amplifier circuit must be designed to achieve a gain of 100. The “minimum” circuit, which will achieve this gain, is shown in figure (i) below. It consists of two small amplifiers; each small amplifier magnifies the volume of sound by a factor of 10. Two alternative circuits are shown in (ii) and (iii) which will achieve the desired gain. Each circuit can function as long as signals can pass from A to B. Suppose each individual small amplifier has a reliability (chance of functioning properly) of 0.9. Assume that they are functioning independently.



(a) Compute the reliability (chance of functioning properly) of each of the three circuits. Which circuit is the most reliable? (*8 marks*)

(b) Let X be the number of amplifiers functioning properly in circuit (iii). Write down the distribution, expected value and variance of X. (*6 marks*)

(c) If two of the amplifiers in circuit (iii) were broken down, what is the probability that this circuit is still functioning? (*4 marks*)

3. (*15%*) The following data are guesses of the outcome of Super Bowl 1994 by 16 faculty members and staff. Team = Buffalo (B) or Dallas (D); Points = total number of points to be scored in the game.

Staff	**Team**	**Points**	**Staff**	**Team**	**Points**
Alice	B	46	Jay	D	45
Barbara	D	56	Joan	B	49
Carole	B	49	Joe	D	68
Dan	D	66	Larry	B	60
Dave	D	61	Lynne	B	51
Emmy	D	40	Ned	D	45
Erl	B	33	Nick	B	45
Jack	B	57	Ralph	D	59

(a) What are the measurement scales of the variables "Team" and "Points"? (*4 marks*)

(b) Find the first quartile and 80^{th} percentile of the points in whole dataset. (*4 marks*)

(c) Test whether there is significance difference on their guesses on the two teams at $\alpha = 0.05$. (*4 marks*)

(d) Write down the assumption(s) and/or approximation(s) you had made in doing part (c). (*3 marks*)

4. (*10%*) From past experience, a manufacturer knows that the lifetime (in days) of the central processing unit of a certain type of microcomputer can be modelled by a gamma distribution with $\alpha = 3$ and $\lambda = 0.002$.

(a) What is the probability that a central processing unit will have a lifetime of at least 2660 days? (*3 marks*)

After the manufacturer tightened the quality control requirements, he claimed that the mean lifetime of the central processing units produced was increased. A random sample of 100 central processing units was tested. The sample mean and sample standard deviation are 1681 days and 759 days respectively.

(b) To establish the manufacturer's claim, what hypotheses should be test? (*3 marks*)

(c) Test the hypotheses in (b) at 5% significance level. What can be concluded? (*4 marks*)

P. T. O.

5. (*20%*) Four different television commercials are rated on a scale of 0 to 100. One commercial features a product package shot, another features a cast, a third features a celebrity, and the fourth commercial has a jingle. The advertising company testing these commercials wants to know if they are all equally liked, on average, by the population of viewers ($H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$), or whether some are liked better than others (H_1 : not H_0). The Minitab outputs for the data are given below.

One-Way Analysis of Variance

```
Analysis of Variance
Source      DF        SS        MS        F        p
Factor     (i)      (iv)     (vii)    68.07    0.000
Error     (ii)       (v)      93.7
Total    (iii)      (vi)

                                       Individual 95% CIs For Mean
                                       Based on Pooled StDev
 Level       N      Mean     StDev    --------+---------+---------+--------
Package     26    50.640     9.613             (--*-)
Cast        26    36.815     9.616    (--*-)
Cele        26    56.917     9.969                 (--*--)
Jingle      26    74.560     9.509                             (--*-)
                                      --------+---------+---------+--------
                                             45        60        75
Pooled StDev =   9.678
```

(a) Write down the appropriate model and assumptions. (*3 marks*)

(b) Fill in the intentionally omitted information (i) – (vii) in the above outputs. (*5 marks*)

(c) Is H_0 rejected at 1% significance level? (*3 marks*)

(d) What are the advantage(s) and disadvantage(s) of using the above single test instead of the two samples t-tests on all the possible pairs of samples? (*4 marks*)

(e) Find the 90% confidence interval for each of the pairwise differences among the four treatment effects. What can be concluded? (*5 marks*)

6. (*15%*) To determine the possible effect of a chemical treatment on the rate of seed germination, 100 chemically treated seeds and 150 untreated seeds are sown. The numbers of seeds that germinate are recorded in the following table.

	Germinated	Did not germinated
Treated	84	16
Untreated	132	18

Let p_T, p_U be the germination rates of treated and untreated seeds respectively.

(a) Suppose we want to establish the assertion that treated seeds will have a lower germination rate. What hypotheses should we test? Write down the hypotheses in terms of p_T and p_U. (*3 marks*)

(b) Test your hypotheses in (a) at 3% significance level. Find the *p*-value. (*5 marks*)

(c) Construct a 95% confidence interval for $p_T - p_U$. (*4 marks*)

(d) What can be concluded from the results in (b) and (c)? (*3 marks*)

< E N D >