8029A SEMESTER 2 2006

THE UNIVERSITY OF SYDNEY

FACULTIES OF ARTS, ECONOMICS, EDUCATION,
ENGINEERING AND SCIENCE

MATH1905
STATISTICS (ADVANCED)

November 2006 LECTURERS: M Raimondo

TIME ALLOWED: **One and a half hours**

Name: .....

SID: ..... Seat Number: .....

This examination has two sections: Multiple Choice and Extended Answer.

The Multiple Choice Section is worth 25% of the total examination;
there are 15 questions; the questions are of equal value;
all questions may be attempted.

Answers to the Multiple Choice questions must be coded onto
the Multiple Choice Answer Sheet.

The Extended Answer Section is worth 75% of the total examination;
there are 3 questions; the questions are of equal value;
all questions may be attempted;
working must be shown.

THE QUESTION PAPER MUST NOT BE REMOVED FROM THE EXAMINATION ROOM.

Extended Answer Section

Answer these questions in the answer book(s) provided.
Ask for extra books if you need them.

1. (a) The data below shows measurements of ozone concentration (y) (ppb) and temperature (x) (degrees F), over 11 summer days.

Temperature (x)	93	87	84	80	78	75	73	81	76	77	71
Ozone concentration (y)	91	47	32	20	23	21	24	44	21	28	9

(i) Sketch the scatter plot of the data with Temperature on the horizontal axis. Comment on the relationship between Ozone concentration and Temperature.

(ii) For this data set, you are given the following R output:

```
> c(sum(x),sum(y),sum(x^2),sum(y^2),sum(x*y))
[1] 875   360      70019   16702     29916
```

find the Least Square (LS) regression line used to predict Ozone concentration from Temperature and add the line to your scatter plot. Compute the correlation coefficient and comment on the quality of the LS fit.

(iii) Use the R output (5 number summary of the residuals) below and the data

```
> summary(residuals)
     Min.   1st Qu.    Median    3rd Qu.      Max.
-14.41    -6.80     2.230     4.95       16.96
```

to sketch a boxplot of the residuals, comment on the plot. (Hint: the 2nd largest residual is $e_7 = y_7 - a - bx_7$)

(b) For paired observations $(x_1, y_1), ..., (x_n, y_n)$ we denote $\hat{y}_i = a + bx_i$ the ith predicted value using the regression line. Show that $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \frac{S_{xy}^2}{S_{xx}}$. Explain why this result is useful in regression analysis.

2. (a) A doctor has prescribed a diet for 10 of his patients. Unknown to him the treatment was successful in exactly 8 of the 10 cases. He asks his secretary to contact any 5 of the 10 patients as a follow-up. Assuming that the selection is made randomly, find the probability that exactly 4 of those contacted had a successful result with the diet.

(b) A population is modelled by the random variable X with density function $f(x)$.

(i) If $f(x) = kx(1-x), 0 < x < 1$, and $f(x) = 0$ otherwise. Find the value of k.

(ii) If $f(x) = 12x^2(1-x), 0 < x < 1$, and $f(x) = 0$ otherwise. Calculate $E(X)$ and $\text{var}(X)$ and $P(X \leq \frac{2}{3})$.

(iii) Four independent observations are taken from the same population as in (ii) What is the probability that exactly two are less than $\frac{2}{3}$?

(*iv*) If the random variable X has the same distribution as in part (ii), use Chebyshev's inequality to bound $P(|X - \frac{3}{5}| \geq 0.25)$. Compare this bound to the exact probability.

3. (a) Ten patients who want to loose weight are given Diet A and Diet B in two periods the order in which the diets were given being determined randomly. The following weight were recorded over the periods:

Diet A	68.9	75.5	81.0	67.6	75.8	71.2	69.1	69.1	70.6	69.3
Diet B	89.6	74.8	70.3	76.9	81.9	75.1	69.7	71.9	76.1	77.8

(*i*) Is there evidence of diet B being less effective than diet A? (Use a t-test to answer this (1)) For these data we have access to the following R-output:

```
> c(sum(diet.A),sum(diet.B),sum(diet.A^2),sum(diet.B^2),sum(diet.A*diet.B))
[1]   718.10   764.10      51729.77      58698.27      54817.48
```

(*ii*) Obtain a 99% confidence interval for the difference in weight under the two diets.

(*iii*) Use a sign test to answer the question in (i). (use the normal approximation with correction for continuity and the normal table)

(b) Let A and B be two independent events. Use the probability rules to show that $P(A^c \cap B^c) = P(A^c)P(B^c)$. Conclude.

(1) *Relevant R-output may be found at the end of the multiple choice section.*

End of Extended Answer Section