THE UNIVERSITY OF SYDNEY

FACULTIES OF ARTS, ECONOMICS, EDUCATION AND SCIENCE

MATH1905

STATISTICS (ADVANCED)

November 2005

TIME ALLOWED: One and a half hours

LECTURER: M Raimondo

This Examination has 3 Printed Components.

(1) AN EXTENDED ANSWER QUESTION PAPER (THIS BOOKLET, GREEN 8017A): 3 PAGES NUMBERED 1 TO 3; 3 QUESTIONS NUMBERED 1 TO 3.

(2) A MULTIPLE CHOICE QUESTION PAPER (YELLOW 8017B): 5 PAGES NUMBERED 1 TO 5; 15 QUESTIONS NUMBERED 1 TO 15.

(3) A MULTIPLE CHOICE ANSWER SHEET (WHITE 8017C): 1 PAGE.

Components 2 and 3 MUST NOT be removed from the examination room.

This Examination has 2 Sections: **Extended Answer** and **Multiple Choice**.

The **Extended Answer Section** *is worth 75% of the total marks for the paper: all questions may be attempted; questions are of equal value; working must be shown.*

The **Multiple Choice Section** *is worth 25% of the total marks for the paper: all questions may be attempted; questions are of equal value; answers must be coded onto the* **Multiple Choice Answer Sheet**.

Calculators will be supplied; no other electronic calculators are permitted.
Notes for use in this examination are printed after the Multiple Choice questions.

1. (a) The data below shows the number of fatal road accidents in NSW (y) over the period 1961-71 (x).

Year (x)	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
Accidents (y)	850	798	818	903	1026	1042	1022	1069	1070	1135	1096

(i) Sketch the scatter plot of the data with Year on the horizontal axis. Comment on the relationship between Accidents and Years.

(ii) Given that $\sum_{i=1}^{11} x_i = 21626$, $\sum_{i=1}^{11} y_i = 10829$, $\sum_{i=1}^{11} x_i^2 = 42516826$, $\sum_{i=1}^{11} y_i^2 = 10803863$, $\sum_{i=1}^{11} x_i y_i = 21293476$, find the Least Square (LS) regression line used to predict Accidents from Years 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 Accidents) below

```
> summary(y)
  Min. 1st Qu. Median   3rd Qu.    Max.
 798.0   876.5 1026.0   1070.0  1135.0
```

to sketch a boxplot of Accidents.

(b) For paired observations $(x_1, y_1), ..., (x_n, y_n)$ we denote r the sample correlation coefficient. Show that $|r| \leq 1$. <u>Hint</u>: $\sum_{i=1}^{n} \left((x_i - \bar{x})/\sqrt{S_{xx}} + (y_i - \bar{y})/\sqrt{S_{yy}}\right)^2 \geq 0$.

2. (a) The following table gives the observed frequencies (O_i) of genotypes A, B, and C of 100 plants from a genetic cross:

Genotype	A	B	C	Total
O_i	18	55	27	100

Compute the expected frequencies, E_i's under the hypothesis that A, B, and C are in the ratio of 1:2:1.

(b) Using the same data as in part (a) test whether the model 1:2:1 fits the data well. (Use the appropriate R output on page 3 to answer this)

(c) A 12 sided die in thrown 10,000 times. An observer has recorded that the number 7 appeared 901 times. Does this suggest that the die is loaded to favour 7? (Justify your answer)

(d) Let A and B be two events. Use the axioms of probability to show that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. <u>Hint</u>: $A = (A \cap B) \cup (A \cap B^c)$.

3. The following data are measurements of weight gain (in gm) after 10 Males rats and 10 Females rats were given the same Diet over the same period of time. The 10 Males rats and 10 Females rats were chosen independently.

Male (x)	2.6	4.8	12.5	8.7	9.7	8.2	9.4	8.7	9.2	10.0
Female (y)	8.1	7.6	10.5	8.9	11.2	6.9	11.7	12.6	10.3	7.1

(a) Given that $\sum_{i=1}^{10} x_i = 83.8, \sum_{i=1}^{10} y_i = 94.9 \sum_{i=1}^{10} x_i^2 = 771.76, \sum_{i=1}^{10} y_i^2 = 938.03$, is there evidence of a difference in weight gains between Male rats and Female rats?
(Use the appropriate R output below to answer this)

(b) The function

$$\begin{aligned} f(x) &= Cxe^{-2x}, \qquad x > 0, \\ &= 0, \qquad \text{otherwise} \end{aligned}$$

is a probability density function.

(i) Find C and calculate $E(X)$.

(ii) Let $X \sim N(1, \frac{1}{2})$. Use Chebyshev's inequality to bound $P(X > 2)$. Is Chebyshev's bound sharp in this case? (Justify your answer)

(c) Let X be a continuous random variable with probability density function $g(x)$. We suppose that $EX = 0$ and that $EX^2 = \sigma^2 < \infty$. Prove that X satisfies Chebyshev's inequality. Hint: $S = \{x : |x| > c\,\sigma\}$.

R output

```
> p=c(.25,.1,.05,.025,.01,.005,.001)
> q=1-p
[1] 0.750 0.900 0.950 0.975 0.990 0.995 0.999
> qt(q,8)
[1] 0.706 1.397 1.860 2.306 2.896 3.355 4.501
> qt(q,9)
[1] 0.703 1.383 1.833 2.262 2.821 3.250 4.297
>qt(q,10)
[1] 0.700 1.372 1.812 2.228 2.764 3.169 4.144
> qt(q,18)
[1] 0.688 1.330 1.734 2.101 2.552 2.878 3.610
> qt(q,19)
[1] 0.688 1.328 1.729 2.093 2.539 2.861 3.579
> qt(q,20)
[1] 0.687 1.325 1.725 2.086 2.528 2.845 3.552

>x=c(0.2,0.8,1.5,1.8,2.4,3.1,32.3)
> pchisq(x,1)
[1] 0.35 0.63 0.78 0.82 0.88 0.92 1.00
> pchisq(x,2)
[1] 0.10 0.33 0.53 0.59 0.70 0.79 1.00
> pchisq(x,3)
[1] 0.02 0.15 0.32 0.39 0.51 0.62 1.00
> pchisq(x,4)
[1] 0.00 0.06 0.17 0.23 0.34 0.46 1.00
```