THE UNIVERSITY OF SYDNEY
FACULTIES OF ARTS, ECONOMICS, EDUCATION, ENGINEERING AND SCIENCE

MATH1905
STATISTICS

November 2008 LECTURER: N. Weber

TIME ALLOWED: **90 minutes**

Name: ....

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

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

The Multiple Choice Section is worth 35% of the total examination;
there are 20 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 65% of the total examination;
there are 4 questions; the questions are of equal value;
all questions may be attempted;
working must be shown.

Calculators will be supplied; no other calculators are permitted.
Statistical tables will be supplied.
Notes for use in this examination are printed after the last question in this booklet.

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.

Extended Answer Question Paper

1. Levine (1973) conducted a study on the effect of smoking on platelet aggregation. Blood samples were drawn from 10 patients before and after they smoked a cigarette and the extent of aggregation measured in each sample. The data are

Patient	1	2	3	4	5	6	7	8	9	10
Before (x_i):	25	25	27	44	30	67	53	53	52	60
After (y_i):	27	29	37	56	46	82	57	80	61	59
Differences (d_i):	2	4	10	12	16	15	4	27	9	-1

where $d_i = y_i - x_i$.

Summary statistics include:

Before:	sample mean $\bar{x} = 43.6$	$S_{xx} = 2216.4$
After:	sample mean $\bar{y} = 53.4$	$S_{yy} = 3250.4$
Differences:	$\Sigma_i d_i^2 = 1572$	

(a) (5 marks) Calculate the coefficient of correlation between the before and after scores. What does this tell you about the scatterplot of the before versus the after scores? Do NOT draw the scatterplot.

(b) (7 marks) Assuming that the differences follow a normal distribution analyse the data to test $H_0 : \mu_D = 0$ against $H_1 : \mu_D > 0$, where μ_D is the mean change in platelet aggregation level. Calculate the test statistic and give a clear expression for the P-value. What can you conclude?

2. (a) (3 marks) Use the data in Question 1 to find the P-value if the sign test statistic is used to test the hypotheses in Question 1 (b)

(b) (5 marks) A 1985 Gallop poll estimated the support among Americans for 'right to die' laws. For the survey 1528 adults were asked if they were in favour of *voluntary* withholding of life support systems from the terminally ill. The number in favour was 1238. Find a 99% confidence interval for the true percentage, p, of adult Americans who are in favour.

(c) (4 marks) What sample size would be required if a future survey wishes to estimate the proportion using a 95% confidence interval of width no more than 0.04? (Use the conservative confidence bound.)

3. (a) (7 marks) The number of radioactive counts in 100 one minute intervals for a particular material were

No. of counts (x):	0	1	2	3	4	5
Frequency:	15	28	23	16	12	6

Suppose that we can model the number of counts in one minute by a Poisson random variable X, where

$$P(X = j) = e^{-\lambda}\, \lambda^j / j!, \quad j = 0, 1, \ldots.$$

(i) Calculate the expected number of 0's in a sample of size 100 from a Poisson random variable with mean 2.

(ii) You are given that the above sample average is $\bar{x} = 2$. Complete the table of expected frequencies below and test the goodness of fit of the Poisson distribution as a model for the number of radioactive counts.

No. of counts (x):	0	1	2	3	4	5
Expected Frequency:	?	27.07	27.07	18.04	9.02	?

(b) (5 marks) The random variable X has probability generating function

$$P(s) = e^{3(s^2-1)}.$$

(i) Use this function to find $E(X)$.

(ii) Evaluate $P(X = 0)$.

(iii) Show that the probability that X takes an odd value is 0.

4. In each game of Sic Bo three regular, six-sided dice are thrown once.

(a) (1 mark) In a single game, what is the probability that all 3 dice show the number '4'?

(b) (3 marks) What is the probabillity that exactly two dice show '4'?

(c) (3 marks) A sequence of 90 independent games is played. How many games do you expect to have 3 different numbers showing on the dice?

(d) (5 marks) Find the exact probability that in 90 games there are at least 2 games with all three dice showing the same number. Find an approximation using the Poisson approximation to the binomial probabilities for this senario.

End of Extended Answer Section

FORMULA SHEET FOR MATH1905 STATISTICS

- **Calculation formulae:**
  - *For a sample* $x_1, x_2, \ldots, x_n$

Sample mean $\bar{x}$	$\frac{1}{n}\sum_{i=1}^{n} x_i$
Sample variance s^2	$\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right].$

  - *For paired observations* $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

S_{xy}	$\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$
S_{xx}	$\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$
S_{yy}	$\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$
r	$\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$

For the regression line $y = a + bx$:	
b	$\frac{S_{xy}}{S_{xx}}$
a	$\bar{y} - b\bar{x}$

- **Some probability results:**

For any two events A and B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ and $P(A \cap B) = P(A)P(B\|A)$
If A and B are mutually exclusive (m.e.)	$P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$
If A and B are independent	$P(A \cap B) = P(A)P(B)$

- If $X \sim \mathcal{B}(n, p)$, then :
  $P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$, $r = 0, \ldots, n$, $E(X) = np$ and $var(X) = np(1-p)$

- **Some test statistics** and sampling distributions under appropriate assumptions and hypotheses:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

$$\frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim t_{n_x+n_y-2}, \text{ where}$$

$$S_p^2 = [(n_x - 1)S_x^2 + (n_y - 1)S_y^2]/(n_x + n_y - 2)$$

$$\sum_i \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_\nu, \text{ for appropriate } \nu$$