PAPER CODE NO.
**MATH 161**

THE UNIVERSITY OF LIVERPOOL

**JANUARY 2006 EXAMINATIONS**

Bachelor of Science: Year 1
Master of Mathematics: Year 1

**INTRODUCTION TO STATISTICS**

TIME ALLOWED: Two Hours and a Half

---

INSTRUCTIONS TO CANDIDATES

Answer **all** questions in Section A.

The marks for the **best three** answers in Section B will be used in the assessment.

Marks for parts of questions may be subject to small adjustments.

Normal and Chi-squared tables are provided at the end of this paper.

**NOTES, DECEMBER 2006**

1. This version of the paper shows the marks as used when marking the scripts, i.e. *after* the small adjustments mentioned in the instructions section above.

2. In question 8, the diagram had labels added by hand, not easily reproducible electronically. Instead, an explanation has been added immediately before the diagram, intended to make clear where the labels were.

3. The Normal and Chi-squared tables provided as part of the paper are not attached here.

THE UNIVERSITY
*of* LIVERPOOL

## SECTION A

1. The median of a set of data is 18 and the quartiles are 15 and 22. The three highest values in the dataset are 30, 34 and 35. The three lowest values are 6, 8 and 8. Represent the data by a boxplot, explaining the procedure you follow. Comment on the plot. **(8 marks)**

2. Suppose that A, B and C are three events occurring with non-zero probability in a sample space $\Omega$.
   (i) (a) If A,B and C are mutually exclusive, express Pr $(A \cup B \cup C)$ in terms of Pr (A), Pr (B) and Pr (C). **(1 mark)**
   (b) If they are also exhaustive, i.e. Pr $(A \cup B \cup C) = \Omega$, express Pr (C) in terms of Pr (A) and Pr (B). **(1 mark)**
   (ii) If, instead, A and B are <u>not</u> mutually exclusive, but are independent events, express Pr $(A \cup B)$ in terms of Pr (A) and Pr (B). **(2 marks)**
   (iii) (a) If, instead, A and B are neither mutually exclusive nor independent, explain what is meant by Pr $(B|A)$ and express it in terms of Pr $(A \cap B)$ and Pr (A). **(2 marks)**
   (b) Finally, express Pr $(B|A)$ in terms of Pr $(A|B)$, Pr (A) and Pr (B). **(2 marks)**

3. A woman who is making random telephone calls on behalf of a charity estimates that, if the call is answered by a woman, there is a constant 30% chance of making a sale, but that if the call is answered by a man, there is only a (constant) 20% chance of making a sale. Her next ten calls are answered by three men and seven women, who behave mutually independently.
   (i) What is the probability that she will make exactly one sale to the three men answering the phone calls? **(2 marks)**
   (ii) What is the probability that she will make no more than two sales to the seven women answering the phone calls? **(3 marks)**
   (iii) What is the probability that she will make no more than one sale in the ten calls? **(3 marks)**

THE UNIVERSITY *of* LIVERPOOL

4. A study of 60 patients who received a particular drug as a treatment for a bone disease revealed that 34 had suffered side effects. It was suspected that these might be related to a blood problem, so the patients were classified as shown:

| | Side effects | no side effects |
|---|---|---|
| Blood problem | 28 | 8 |
| No blood problem | 6 | 18 |

Use a $\chi^2$ test to assess whether or not the blood problem and side effects are associated. **(8 marks)**

5. The random variable X has an exponential distribution with parameter $\lambda$, i.e.its p.d.f. is given by $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ and $f(x) = 0$ otherwise.
(i) Show that the distribution function of X is given by $1 - e^{-\lambda x}$ for $x \geq 0$. **(2 marks)**
(ii) Suppose that $\lambda = 3$. Use the result of part (i) to find the probability that $0.25 \leq X \leq 0.5$. **(2 marks)**
(iii) Prove that, for all $s, t > 0$,
$\Pr(X > s+t \mid X > s) = \Pr(X > t)$. **(4 marks)**

THE UNIVERSITY of LIVERPOOL

SECTION B

6. If X has a binomial, B (n ; p), distribution,
(i) show that

$$\Pr(X = r+1) = \frac{(n-r)}{(r+1)}\frac{p}{(1-p)}\Pr(X = r)$$

for r=0,1,2,…,n-1; **(2 marks)**

(ii) find Pr (X = 26) in terms of Pr (X = 25) for the case where n=55 and p=0.52; **(2 marks)**

(iii) for what value of r will Pr (X=r) take its highest value in the case given in (ii)? **(4 marks)**

(iv) Explain how, and for what values of n and p, B (n ; p) may be approximated by

(a) a Poisson P ($\lambda$) distribution **(2 marks)**

(b) a Normal N ($\mu$, $\sigma^2$) distribution **(3 marks)**

(v) For which, if either, of these approximations would use of a continuity correction be worth considering, and why? **(2 marks)**
(vi) Find Pr (X ≥ 29) for B (55; 0.52) by using a suitable approximating distribution. **(5 marks)**

THE UNIVERSITY
*of* LIVERPOOL

7. The number of errors found by a spell-checker in randomly-selected pages from an electronic library (which may be assumed to have a large enough stock that its number of pages may be treated as infinite) is believed to have a Poisson distribution with a mean of 2.
(i) Verify that Pr (X = 0) = 0.135 to 3 d.p. **(1 mark)**
(ii) Use the Poisson recurrence relationship

$$\Pr(X = x+1) = \frac{\lambda}{(x+1)} \Pr(X=x) \text{ for } x = 0, 1, \ldots$$

to find Pr (X=x+1) for each integer in the range $0 \le x \le 5$. **(6 marks)**
(iii) What is Pr (X>6)? **(1 mark)**
(iv) A sample of 100 pages produces the following frequency table:

| Errors per page | Frequency |
|---|---|
| 0 | 11 |
| 1 | 32 |
| 2 | 26 |
| 3 | 14 |
| 4 | 12 |
| 5 | 4 |
| 6 | 1 |

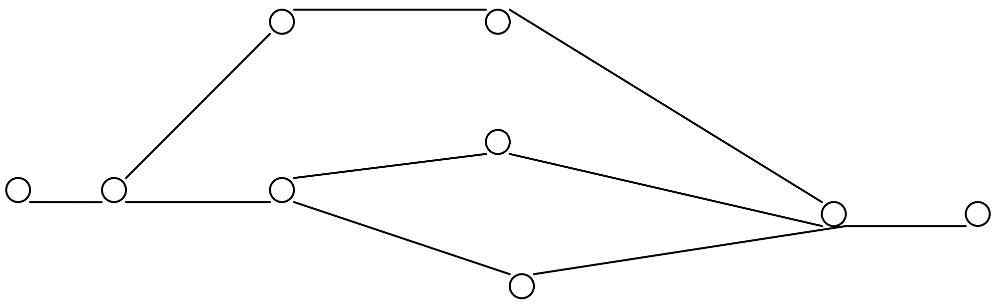
(a) Represent this table by a suitable plot, justifying your choice. **(4 marks)**

(b) Use a $\chi^2$ goodness-of-fit test to assess the fit of P(2) to the data. Comment on your results. **(8 marks)**

**In the following question, the labels on the nodes have the pattern:**

**B D**
**E**
**START AC G FINISH**
**F**

8. Current flows through a network as shown below:



The probabilities that the various gates (shown as nodes in the diagram) allow current to pass through are given by :
Pr (A) = Pr (G) = 0.9, Pr ( B) = Pr (D) = 0.8, Pr (C) = 0.7 and
Pr (E) = Pr (F) = 0.6.
It may also be assumed that the gates operate independently. What is the probability that the current will flow from START to FINISH?

**(20 marks)**

THE UNIVERSITY of LIVERPOOL

9. A biathlon event consists of a one-mile swim followed by a 10 mile run. John has discovered that his times in minutes for the swim follow a N (25 , 4 ) distribution, whilst those for the run follow an independent N (75 , 9 ) distribution. On a particular day, John starts his swim at 1.00 p.m.

(i) What is the probability that John starts his run later than 1.30 p.m.? **(3 marks)**

(ii) What is the probability that his run takes between 70 and 80 minutes? **(4 marks)**

(iii) What is the probability that John beats a club record by completing the whole event in under 95 minutes? **(6 marks)**

(iv) On another occasion, Michael enters the same event. His times in minutes are independently distributed as N (27, 8) for the swim and N (72, 15) for the run. What is the probability that he will beat John? **(7 marks)**

10. A new airline is considering what maximum leg room to allow between seats. Leg room needed (in cm.) is believed to vary as N($\mu$, 36) where $\mu$ is determined by the airline's assumed mix of customers. A random sample of 9 customers gives the following results in terms of leg room in cm. considered adequate:
70, 67, 76, 77, 71, 73, 65, 68, 72.

(i) State the sampling distribution of the sample mean, $\overline{X}$ , and the mean and variance of this distribution. **(4 marks)**

(ii) Construct a two-sided 95% confidence interval for $\mu$. **(6 marks)**

(iii) Interpret your results for the manager of the airline. **(3 marks)**

(iv) (a) The manager wishes to test at the 5% significance level a null hypothesis that $\mu = 75$ against an alternative that $\mu \neq 75$. Use the results above to carry out this test, explaining your argument fully. **(3 marks)**

(b) Would your conclusions change if the manager wished to use a 1% significance level instead? Explain your answer. **(4 marks)**