MATH036501

This question paper consists of 11 printed pages, each of which is identified by the reference **MATH0365**.

Graph paper is provided.
A formulae sheet is attached.
A normal table is attached.
Only approved basic scientific calculators may be used.

Examination for the Module MATH0365
(January 2005)

FOUNDATION PROBABILITY AND STATISTICS

Time allowed: **2 hours**

Attempt **ALL** questions in Section A and **TWO** questions from Section B.

Questions A1 to A10 require you to write down a single letter answer.
Questions A11 to A20 require you to write down a short explanation.

Sections A and B are each worth 50% of the examination marks.
Questions A11 to A20 are each worth 1.5 times the marks of questions A1 to A10.

SECTION A

Attempt all questions in Section A.

Questions A1 to A10 require you to write down a single letter answer.

A1. Consider the following stem-and-leaf plot which shows the scores of 20 adults taking a psychometric test:

8	67
9	045999
10	011111235678

Table 1: Key: 10|2 means 102

Which of the following are true?

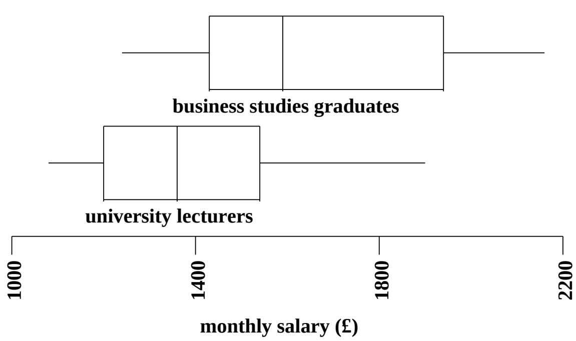
(i) The lowest score recorded was 86.

(ii) The highest score recorded was 103.

(iii) The coding 9|5 corresponds to a test score of 95.

A: all of these **B**: (i) (ii) **C**: (i) (iii)

A2. The two box plots shown below summarise the monthly salaries of 21 business studies graduates and 21 university lecturers:



Which of the following are true?

(i) The median salary of business studies graduates is greater than the median salary of university lecturers.

(ii) Business studies graduates' salaries are less variable than university lecturers' salaries.

(iii) Business studies graduates' salaries are skewed.

A: (i) (iii) **B**: all of these **C**: (i) (ii)

A3. The variables "temperature", x, and "ice cream sales", y, have a correlation coefficient $r = 0.89$. Based on this information alone, which of the following are true?

(i) There is strong positive correlation between temperature and ice cream sales.

(ii) The points (x_i, y_i) must approximate a straight line with positive slope.

(iii) An increase in temperature must be the cause of increased demand for ice cream.

A: (i) only **B**: all of these **C**: (i) (ii)

A4. A class of 25 students spent between 2 and 10 hours studying for an exam. The variables "hours studied", x, and "exam mark", y, follow a regression line $y = 22.71 + 3.98x$. Assuming the regression line is an appropriate summary of the data gathered, which of the following are true?

(i) The regression line can be used to predict a mark of $y = 42.61$ for a student that studied for $x = 5$ hours.

(ii) The regression line shows that a mark of $y = 100$ can be achieved if a student studies for $x = 19.41$ hours.

(iii) For the students in the class, each extra hour studied was associated with an average of 3.98 extra marks on the exam.

A: (i) only **B**: (i) (ii) **C**: (i) (iii)

A5. If events A and B are such that $\mathrm{P}(A) = 0.5$, $\mathrm{P}(B) = 0.75$ and $\mathrm{P}(A \cap B) = 0.25$. Which of the following are true?

(i) Events A and B are exhaustive.

(ii) Event B is more likely to occur than event A.

(iii) $\mathrm{P}(A \cup B) = 1$.

A: (i) (ii) **B**: all of these **C**: (i) (iii)

A6. Four professors, three lecturers and two students are queuing in a line at the coffee bar. How many different arrangements of professors, lecturers and students are possible?

A:362880 **B**: 1260 **C**: 2520

A7. The discrete random variable X follows a $\mathrm{B}(n, p)$ distribution. Which of the following are true?

(i) The random variable X is the number of events that occur within a section of space (or time).

(ii) Each of the n "trials" is either a success or a failure.

(iii) The "trials" are performed independently.

A: (ii) (iii) **B**: all of these **C**: (i) (iii)

CONTINUED...

A8. The discrete random variable Y follows a Po(5) distribution. Which of the following are true?

(i) $\mathrm{P}(Y = 1) = 0.0337$.

(ii) $\mathrm{E}(Y) = 5$.

(iii) $\mathrm{Var}(Y) = 5$.

A: (i) only **B**: (i) (iii) **C**: all of these

A9. For $Z \sim \mathrm{N}(0, 1)$ which of the following are true?

(i) $\mathrm{P}(a < Z < b) = \Phi(b) - \Phi(a)$.

(ii) $\mathrm{P}(Z < -z) = \mathrm{P}(Z < z)$.

(iii) $\mathrm{P}(Z < 0) = \Phi(0) = 0.5$.

A: all of these **B**: (i) (iii) **C**: (ii) (iii)

A10. The weight of a loaf of bread produced in a small bakery is modelled by a normal distribution with $\mu = 300$ grams and standard deviation $\sigma = 15$ grams. What is the probability that a loaf of bread will weigh more than 262.5 grams?

A:0.9938 **B**: 0.0062 **C**: 0.9878

Questions A11 to A20 require you to write down a short explanation.

A11. Explain what is meant by qualitative data, and give an example.

A12. Explain why the interquartile range (IQR) is preferred to the range as a measure of variability.

A13. Draw a scatter diagram that shows two variables, x and y, with strong negative correlation.

A14. Draw a diagram that shows the residuals $\{r_i = y_i - (a + bx_i)\}$ from the regression line $y = a + bx$. Give a condition that the residuals must satisfy when a and b are estimated.

A15. Explain what is meant by the sample space of a statistical experiment, and give an example.

A16. Explain what is meant if events A and B are statistically independent, and give a probability statement that they must satisfy.

A17. State the two conditions that the probability function of a discrete random variable must satisfy.

A18. Explain what is meant by a Poisson process, and state any two of the three conditions that a Poisson process must satisfy.

A19. Explain, with the aid of diagrams, how the parameters μ and σ affect the shape of the normal probability density function.

A20. Draw a sketch to show the relationship between $\mathrm{P}(Z < -z)$ and $\mathrm{P}(Z > z)$ where $Z \sim \mathrm{N}(0, 1)$.

SECTION B
Attempt TWO questions from Section B.

B1. (a) The following data refer to the time (in minutes) taken by 13 trains to travel from London Euston to Glasgow Central on the West Coast Mainline:

$$314, 329, 297, 324, 332, 316, 325, 314, 339, 396, 422, 396, 388$$

Calculate the median journey time, the lower quartile, Q_1, and the upper quartile, Q_3, for these data.

(b) The following data refer to the time (in minutes) taken by 13 trains to travel from London Kings Cross to Edinburgh Waverley on the East Coast Mainline:

$$259, 260, 260, 261, 261, 266, 269, 273, 279, 284, 285, 303, 303$$

It may be shown that the median journey time for these data is 269 minutes, and the quartiles are $Q_1 = 261$ and $Q_3 = 284$. Use this information, together with the calculations from part (a), to produce box plots comparing the journey times on the two different routes.

(c) For the data in parts (a) and (b), calculate the quartile coefficients of skewness, and identify any outliers.

(d) What can you say about journey times on the two different routes?

B2. (a) The table below summarises the weights (to the nearest gram) of a random sample of 60 pears from a small orchard.

Weight	Frequency f_i	Midpoint y_i	$z_i = (y_i - 133)/5$	$z_i f_i$	$z_i^2 f_i$
131-135	8	133	0	0	a_1
136-140	13	138	1	13	a_2
141-145	12	143	2	24	a_3
146-150	20	148	3	60	a_4
151-155	7	153	4	28	a_5
	$n = \sum_{i=1}^{5} f_i = 60$			$\sum_{i=1}^{5} z_i f_i = 125$	$\sum_{i=1}^{5} z_i^2 f_i = a_6$

(i) Calculate the values that should be inserted in place of the constants $a_1, a_2, a_3, a_4, a_5, a_6$.

(ii) Using the coding $z_i = (y_i - 133)/5$ estimate the mean weight of these pears, $\bar{y}$, and the variance in the weights of these pears, s_y^2.

QUESTION B2 CONTINUED...

(b) The orchard owner suspects that the weight of a pear might be linearly related to the age of the tree on which it grows. The following data are collected, where x denotes the age of a tree (in years), and y denotes the weight of a randomly selected pear from the tree (in grams).

x	2	3	5	7	7
y	135	138	145	153	149

$$\sum_{i=1}^{5} x_i = 24, \quad \sum_{i=1}^{5} x_i^2 = 136, \quad \sum_{i=1}^{5} y_i = 720, \quad \sum_{i=1}^{5} y_i^2 = 103904, \quad \sum_{i=1}^{5} x_i y_i = 3523.$$

(i) Construct a scatter diagram of the data.

(ii) Calculate the values of a and b in the regression line $y = a + bx$.

(iii) Add the regression line to the plot produced in part (i).

B3. I invite you to throw a green four-sided die and a red four-sided die. Both dice are fair. However, the green die has sides numbered $1, 2, 2, 4$, and the red die has sides numbered $1, 3, 4, 4$.

(a) I define the events:

A: The score on the green die is 2.
B: The total of the scores on the two dice is 6.
C: The total of the scores on the two dice is 8.

(i) Calculate $\mathrm{P}(A)$, $\mathrm{P}(B)$ and $\mathrm{P}(C)$.

(ii) Calculate $\mathrm{P}(A \cap B)$ and $\mathrm{P}(A|B)$. Are events A and B statistically independent?

(iii) Calculate $\mathrm{P}(A \cap C)$. Are events A and C mutually exclusive?

(b) Let the random variable X denote the total of the scores on the two dice.

(i) Determine the probability distribution of X.

(ii) Calculate $\mathrm{E}(X)$.

(iii) Calculate $\mathrm{E}(X^2)$ and $\mathrm{Var}(X)$.

(c) I will charge you £5 to throw the dice and pay you £X (i.e. the total of the scores on the two dice in pounds). Would you accept my invitation? Justify your answer.

B4. (a) A solicitor is currently handling 5 criminal cases. From past experience the probability that a criminal case will end in conviction is 0.63. Use a suitable model to calculate the probability that exactly 3 of the 5 cases will end in conviction. What assumptions are you making?

(b) The solicitor's firm usually deals with one large fraud case every 3 years. Use a suitable model to calculate the probability that the firm will deal with 2 or more large fraud cases over the next 6 years. What assumptions are you making?

(c) The notes for a particular criminal case are 200 pages long. From past experience, the solicitor believes that the probability a page will contain errors is 0.05. Using a suitable approximation, determine the probability that the notes contain less than 3 pages with errors.

(d) The length of time taken to prepare a set of case notes has mean $\mu = 5$ days and standard deviation $\sigma = 1.5$ days. Assuming that the time taken to prepare case notes can be modelled by a normal distribution, calculate the probability it will take between 3.5 days and 7.25 days to prepare a set of case notes.

FORMULAE SHEET

Representation and summary of data

$$\text{median} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd,} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)}}{2} & \text{if } n \text{ is even.} \end{cases}$$

$$\text{mean} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

$$\bar{x} \text{ (grouped data)} = \frac{\sum_{i=1}^{m} x_i f_i}{n}, \quad n = \sum_{i=1}^{m} f_i.$$

$$\text{lower quartile} = Q_1 = x_{\left(\frac{n+3}{4}\right)}.$$

$$\text{upper quartile} = Q_3 = x_{\left(\frac{3n+1}{4}\right)}.$$

$$\text{range} = x_{(n)} - x_{(1)}.$$

$$\text{IQR} = Q_3 - Q_1.$$

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

$$s^2 \text{ (grouped data)} = \frac{1}{n-1}\left[\sum_{i=1}^{m} x_i^2 f_i - \frac{\left(\sum_{i=1}^{m} x_i f_i\right)^2}{n}\right], \quad n = \sum_{i=1}^{m} f_i.$$

$$s = \sqrt{s^2}.$$

$$\text{if } y_i = \frac{x_i - a}{b} \qquad \bar{x} = a + b\bar{y}, \quad s_x^2 = b^2 s_y^2.$$

$$\text{quartile coefficient of skewness} = \frac{Q_3 - (2 \times \text{median}) + Q_1}{Q_3 - Q_1}.$$

$$\text{outliers are outside the limits} \quad \left[\frac{1}{2}\left(5Q_1 - 3Q_3\right), \frac{1}{2}\left(5Q_3 - 3Q_1\right)\right].$$

Correlation and regression

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}.$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}.$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}.$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}.$$

In the regression line $y = a + bx$ $\qquad b = \frac{S_{xy}}{S_{xx}}, \quad a = \bar{y} - b\bar{x}.$

CONTINUED...

Probability

$$\begin{aligned}
\mathrm{P}(A') &= 1 - \mathrm{P}(A). \\
\mathrm{P}(A \cup B) &= \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B). \\
\mathrm{P}(A|B) &= \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}. \\
n! &= n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1, \\
0! &= 1. \\
{}^nP_r &= \frac{n!}{(n-r)!}. \\
\binom{n}{r} &= \frac{n!}{(n-r)!r!}.
\end{aligned}$$

A box contains N balls. The balls are of k different types. There are N_1 balls of type 1, N_2 balls of type 2 etc. (with $\sum_{i=1}^{k} N_i = N$). The probability that the sample contains exactly n_1 balls of type 1, n_2 balls of type 2 etc. (with $\sum_{i=1}^{k} n_i = n$) is:

$$\frac{\binom{N_1}{n_1} \times \binom{N_2}{n_2} \times \cdots \times \binom{N_k}{n_k}}{\binom{N}{n}}.$$

Discrete random variables

$$\begin{aligned}
\mathrm{E}(X) &= \sum_x x\mathrm{P}(X = x). \\
\mathrm{E}(X^2) &= \sum_x x^2\mathrm{P}(X = x). \\
\mathrm{Var}(X) &= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2. \\
\mathrm{E}(aX + b) &= a\mathrm{E}(X) + b. \\
\mathrm{Var}(aX + b) &= a^2\mathrm{Var}(X).
\end{aligned}$$

discrete uniform probability function $\quad \mathrm{f}(x) = \frac{1}{n}, \quad (x = 1, \ldots, n).$

if X is discrete uniform then $\quad \mathrm{E}(X) = \frac{1}{2}(n+1), \quad \mathrm{Var}(X) = \frac{1}{12}(n^2 - 1).$

if $X \sim \mathrm{B}(n,p)$ $\quad \mathrm{f}(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad (x = 0, 1, 2, \ldots, n).$

if $X \sim \mathrm{B}(n,p)$ $\quad \mathrm{E}(X) = np, \quad \mathrm{Var}(X) = np(1-p).$

if $X \sim \mathrm{Po}(\lambda)$ $\quad \mathrm{f}(x) = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}, \quad (x = 0, 1, 2, \ldots).$

if $X \sim \mathrm{Po}(\lambda)$ $\quad \mathrm{E}(X) = \lambda, \quad \mathrm{Var}(X) = \lambda.$

if $X \sim \mathrm{B}(n,p)$ with $n > 50$ and $p < 0.1$ $\quad X$ is approximately $\mathrm{Po}(\lambda = np)$.

The normal distribution

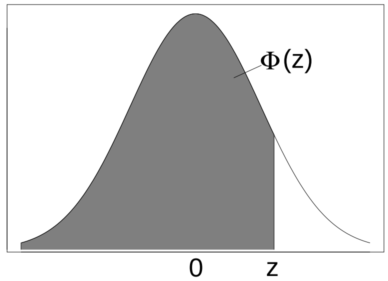
if $X \sim \mathrm{N}(\mu, \sigma^2)$ $\quad \mathrm{E}(X) = \mu, \quad \mathrm{Var}(X) = \sigma^2.$

if $X \sim \mathrm{N}(\mu, \sigma^2)$ $\quad Z = \frac{X - \mu}{\sigma} \sim \mathrm{N}(0, 1).$

CONTINUED...

TABLE OF $\Phi(z)$ FOR THE STANDARD NORMAL DISTRIBUTION

For $Z \sim \mathrm{N}(0,1)$, the table shows $\Phi(z) = \mathrm{P}(Z < z)$ where $z \geq 0$.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
+0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
+0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
+0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
+0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
+0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
+0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
+0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
+0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
+0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
+0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
+1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
+1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
+1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
+1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
+1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
+1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
+1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
+1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
+1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
+1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
+2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
+2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
+2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
+2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
+2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
+2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
+2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
+2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
+2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
+2.9	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
+3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
+3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
+3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
+3.3	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
+3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

END