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 2007)

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. The data below refer to the number of cars sold each month by a motor trader over a 6 month period:

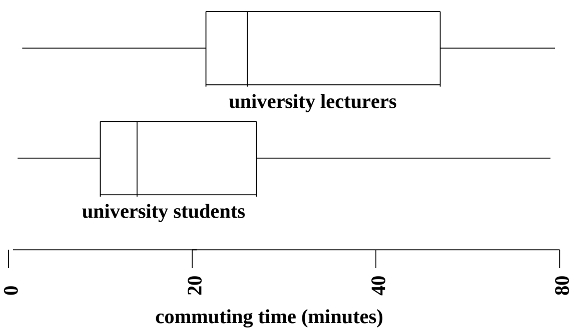
$$31, 39, 29, 22, 35, 42$$

Which of the following are true?

(i) The mean number of cars sold per month is 33.
(ii) The median number of cars sold per month is 33.
(iii) The range in the number of cars sold per month is 33.

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

A2. The two box plots shown below summarise the time taken for a sample of university students and university lecturers to commute from home to the university:

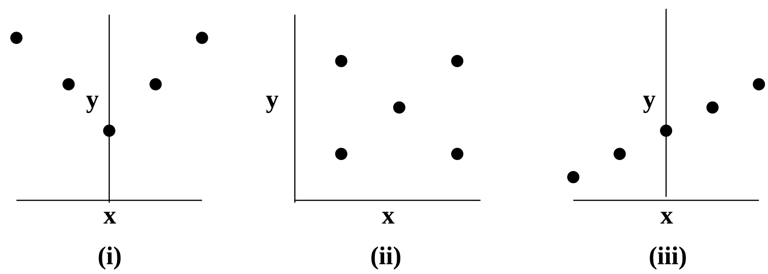


Which of the following are true?

(i) The commuting times for both students and lecturers show positive skewness.
(ii) The interquartile range (IQR) of commuting times is similar for students and lecturers.
(iii) The range of commuting times is similar for students and lecturers.

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

A3. Which of the following scatter diagrams show a relationship between the variables x and y with a correlation coefficient near 0?



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

A4. A group of 20 executives, earning between £25,000 and £75,000 per year, have recently bought new company cars. The variables "salary per year", x, and "spend on company car", y, (each in £1000) follow a regression line $y = 10.56 + 0.37x$. 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 spend of $y = £23,510$ on a company car for an executive earning $£35,000$.

(ii) The regression line can be used to predict a spend of $y = £84,560$ on a company car for an executive earning $£200,000$.

(iii) The regression line can be used to predict a spend of $y = £15,000$ on a company car for an executive earning $£12,000$.

A: (i) only **B**: all of these **C**: none of these

A5. If events A and B are statistically independent with $\mathrm{P}(A) = 0.4$ and $\mathrm{P}(B) = 0.25$, which of the following are true?

(i) $\mathrm{P}(A \cap B) = 0.1$.

(ii) $\mathrm{P}(A|B) = 0.4$.

(iii) $\mathrm{P}(B') = 0.75$.

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

A6. Five cars are to be parked in a line in a company car park. Three of the cars are black, B, and two are silver, S. How many different orderings of the car colours are possible (for example, one of the orderings is $SSSBB$)?

A:120 **B**: 10 **C**: 30

A7. Consider a discrete random variable, X, with probability function $\mathrm{f}(x) = \mathrm{P}(X = x)$. Which of the following are true?

(i) $0 \leq \mathrm{f}(x) \leq 1$ for all values of x.

(ii) $\sum_x \mathrm{f}(x) = 1$.

(iii) $\mathrm{E}(X) = \sum_x x^2 \mathrm{P}(X = x)$.

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

A8. The discrete random variable Y follows a Po(0.5) distribution. Which of the following are true?

(i) $\mathrm{P}(Y = 0) = 0.6065$.

(ii) $\mathrm{E}(Y) = 0.5$.

(iii) $\mathrm{P}(Y = 1) = 0.5\mathrm{P}(Y = 0)$.

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

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

(i) $\mathrm{P}(Z > -1.64) = \Phi(1.64) = 0.9495$.

(ii) $\mathrm{P}(Z < 1.64) = \Phi(1.64) = 0.9236$.

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

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

A10. Consider two independent random variables, $X_1 \sim \mathrm{N}(0, 10)$ and $X_2 \sim \mathrm{N}(0, 6)$. What is $\mathrm{P}(X_1 + X_2 < 1)$?

A:0.5987 **B**: 0.4013 **C**: 0.5239

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

A11. Explain what is meant by (i) Descriptive Statistics and (ii) Inferential Statistics.

A12. Explain what is meant by negatively skewed data.

A13. Draw a scatter diagram that shows two variables, x and y, with a correlation coefficient, r, close to -1.

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. Consider a sample space, S, consisting of $n(S)$ outcomes. Consider an event, A, consisting of $n(A)$ outcomes. How many outcomes are there for the complementary event A'? Illustrate your answer by drawing a Venn diagram.

A16. Explain what is meant by the Basic Principle of Counting.

A17. Explain what is meant by the variance of a random variable.

A18. State the three conditions that events must satisfy if they are described by a Poisson process.

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

A20. State the Central Limit Theorem for independent random variables $X_1, X_2, \ldots, X_n$ each having the same distribution with mean μ and variance σ^2. Briefly explain why this result is important.

SECTION B
Attempt TWO questions from Section B.

B1. (a) The following data refer to the number of spam emails I received each day over a period of two weeks:

Day of week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Week 1	11	21	18	19	10	1	3
Week 2	22	25	16	24	15	4	8

Combine the data from the two weeks and construct a stem-and-leaf diagram.

(b) Calculate the mean number of spam emails received at weekends (i.e. on Saturdays and Sundays), $\bar{x}$, and the variance of the number of spam emails received at weekends, s_x^2. Calculate the mean number of spam emails received on weekdays (i.e. on Mondays, Tuesdays, Wednesdays, Thursdays and Fridays), $\bar{y}$, and the variance of the number of spam emails received on weekdays, s_y^2.

(c) It may be shown that, for Week 1, $Q_1 = 6.5$, median$= 11.0$ and $Q_3 = 18.5$. For Week 2, $Q_1 = 11.5$, median$= 16.0$ and $Q_3 = 23.0$ (you do not need to check this). Use this information to construct box plots to compare the number of spam emails received during the two different weeks.

(d) Using your results from (c), compare the number of spam emails received during the two different weeks. Using your results from (b), compare the number of spam emails received on weekdays and weekends.

B2. (a) The manager of a small clothing shop records the number of minutes customers spend browsing in the shop. She groups the data and uses a coded value, z_i, for the groups. The data for a sample of 50 customers are summarised in the table below.

Browsing time (minutes)	Coded value, z_i	Frequency, f_i	$z_i f_i$	$z_i^2 f_i$
0–10	1	21	a_1	a_7
11–20	2	15	a_2	a_8
21–30	3	8	a_3	a_9
31–40	4	5	a_4	a_{10}
41–50	5	1	a_5	a_{11}
		$n = \sum_{i=1}^{5} f_i = 50$	$\sum_{i=1}^{5} z_i f_i = a_6$	$\sum_{i=1}^{5} z_i^2 f_i = a_{12}$

(i) Calculate the values that should be inserted in place of the constants $a_1, a_2, \ldots, a_{12}$.

(ii) Estimate the mean, $\bar{z}$, and the standard deviation s_z of the coded data. Interpret your results in terms of the number of minutes customers spend browsing in the shop.

(b) The manager suspects the amount a customer spends in the shop (in £) might be linearly related to the time they spend browsing in the shop. The following data are collected, where x denotes the time the customer browses in the shop (in minutes), and y denotes the amount the customer spends in the shop (in £).

x	5	3	10	15	20	35
y	0	5	10	11	30	50

$$\bar{x} = 14.67, \quad \bar{y} = 17.67, \quad S_{xx} = 693.33, \quad S_{yy} = 1773.33, \quad S_{xy} = 1075.33.$$

(i) Construct a scatter diagram to display the data.

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

(iii) Use the regression line to predict the amount, y, a shopper who browses for $x = 7$ minutes will spend in the shop.

(iv) Can the regression line be used to predict how long, x, a shopper who spends $y = £100$ in the shop will have spent browsing?

B3. (a) Consider a family with three children. Using B to denote a boy and G to denote a girl, writing the gender of the youngest child first and the oldest child last, the following eight families are possible:

$$BBB, BBG, BGB, BGG, GBB, GBG, GGB, GGG$$

A couple plans to have children until they have at least one boy and one girl, but they agree they will not have more than three children in total. Using B to denote a boy and G to denote a girl, write out the six possible families the couple could have.

(b) I define the events:

A: The couple has three children.
B: The couple has at least one girl.
C: The couple has exactly two boys.

Assume that each child is equally likely to be a boy or a girl, and multiple births are impossible.

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

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

(c) Let the random variable X denote the number of boys the couple has.

(i) Determine the probability distribution of X.

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

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

(d) All of the bedrooms in the couple's house are painted pink (!) The couple will require three tins of blue paint for each boy they have. Let T denote the total number of tins of blue paint they must buy. Calculate the expected number of tins the couple buys, $\mathrm{E}(T)$, and the variance of the number of tins the couple must buy, $\mathrm{Var}(T)$.

B4. (a) An airline has 5 "premium" seats for a flight to Venice. The airline estimates that 1 in 10 of its premium seats are left vacant. If X is the number of "premium" seats left vacant, what distribution does X have? Calculate the probability that at most 2 of the 5 "premium" seats on this flight are vacant. What assumptions are you making?

(b) There are 250 customers waiting to check in for the flight to Venice. The airline estimates that 1 in 50 of its customers will request a "premium" seat. Using a Poisson approximation, determine the probability that there will be enough "premium" seats on the flight to meet customer demand.

(c) The airline estimates that the time taken to check in a customer for the Venice flight has a mean of 0.5 minutes and a variance of 0.25 minutes^2. The 250 customers have all arrived and are waiting in a queue. The flight will leave in 2 hours time. Use the Central Limit Theorem to estimate the probability that all of the customers will have checked in when the flight departs.

FORMULAE SHEET

Representation and summary of data

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Probability

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

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 $\mathrm{Po}(\lambda = np)$.

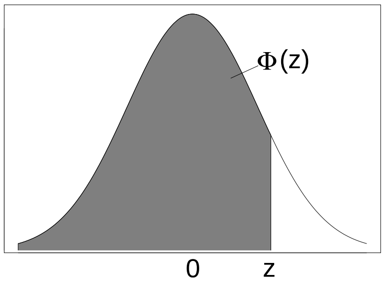
The normal distribution

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

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

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