Full marks can be obtained for correct answers to
five questions

Data provided: New Cambridge Elementary Statistical Tables by D.V. Lindley and W.F. Scott.

Some useful Formulae

1) For any two events A and B

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$,

$P(A \cap B) = P(A \mid B)\, P(B)$.

$P(A \cap \bar{B}) = P(A) - P(A \cap B)$

2) For three events A, B and C

$P\{A \cap (B \cup C)\} = P\{(A \cap B) \cup (A \cap C)\}$,

3) If X has a Binomial distribution with parameters n and p

$$P(X = x) = \binom{n}{x} p^{x} (1 - p)^{n-x} \qquad (x = 0, 1, \ldots, n)$$

and E(X) = np, V(X) = np(1-p).

Moreover, under suitable conditions,

$$P(a \le X \le b) = \Phi(\beta) - \Phi(\alpha),$$

where $\beta = \dfrac{(b + 0.5 - np)}{\{np(1 - p)\}^{‰}}$, $\alpha = \dfrac{a - 0.5 - np}{\{np(1 - p)\}^{‰}}$

and $\Phi(z)$ denotes the area to the left of z for a standard Normal distribution.

4) The t-statistic for testing the hypothesis of a zero mean is

$$t = \frac{\bar{x}}{\sqrt{s^2 / n}},$$

where$x$ and s^2 denote the sample mean and variance respectively, and n denotes the sample size.

5) The two-sample t statistic for testing the equality of two means is

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{1}{m} \right)}},$$

where

$$\hat{\sigma}^2 = (n+m-2)^{-1} \left\{ \sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{m} (y_i - \bar{y})^2 \right\}$$

denotes the `pooled' estimate of variance,x and , respectively, denote the means of the two samples, n and m their sizes and the x_i and y_i denote the individual observations in the two samples.

1a) Given that $\log_2 3 = 1.58496$ and $\log_2 5 = 2.32193$, find

i) $\log_2(15)$; ii) $\log_2(5/3)$; iii) $\log_3 5$; iv) $2^{3.90689}$

[Tables of logs and exponentials, or these functions on a calculator, are not to be used except for reassurance: credit will be given only for answers clearly obtained by other methods.][3 marks]

b) If $y = 3x^2 + x - 30$, find the value at which y has a local maximum or minimum and state which it is. [2 marks]

c) The following is a list of the weights in kilograms of 33 male players on the books of a professional football club at the start of a football season:

60.91	77.73	78.18	76.36	70.91	75.91
71.36	61.82	63.18	64.55	71.36	79.09
70.00	72.27	80.00	68.64	82.73	70.45
82.27	71.82	69.09	66.82	70.45	63.64
72.27	69.55	74.09	75.45	61.82	70.91
61.82	71.82	73.18			

Represent the data by a stem-and-leaf plot. [10 marks]

Comment on possible features of the data brought out by your plot. [2 marks]

Discuss, giving reasons, whether you would expect that valid inferences could be drawn from the above data about the weight distribution of

i) all male professional footballers playing for football clubs in UK. [2 marks]

ii) all males in the UK in the same age range as those playing for this particular club.

[1 mark]

2. Marks, out of hundred, obtained by two groups of students in two subjects, A and B, say, are listed below:

A:	35,	40,	41,	44,	48,	49,	51,	56,
	56,	56,	58,	59,	60,	60,	68,	68,
	69,	73,	73,	84				
B:	19,	21,	25,	26,	28,	34,	36,	47,
	47,	49,	59,	61,	64,	66,	68,	69,
	70,	71,	72,	73,	92,	94		

Obtain a comparative Box plot of the data. [15 marks]

Comment on your results. [5 marks]

3. A family with two pre-school children, called Sarah and Jane, will watch a particular afternoon TV programme only if the mother (M) does not object to watching the programme and either or both Sarah (S) and Jane (J) do not object as well. Let M denote the event that the mother does not object to watching the programme and define S and J in the same way. It is know that

$P(S) = 0.23$, $P(J) = 0.25$, $P(M) = 0.5$,

$P(S \cap J) = 0.06$, $P(S \cap M) = 0.125$, $P(J \cap M) = 0.12$,

$P(S \cap J \cap M) = 0.029$.

Find the probability that

a) Sarah or Jane or both do not object to watching the programme. [3 marks]

b) Jane does not object to watching the programme but Sarah does. [4 marks]

c) The family watches the TV programme. [5 marks]

d) Either Sarah or Jane or both do not object to watching the programme but the mother does. [8 marks]

4. In a series of 7 independent trials, the probability of a `success' is constant from trial to trial and equals p. Write down an expression for the probability that

i) exactly one success is observed,

ii) no more than one success is observed. [3 marks]

A scientist inoculates seven mice, one at a time, with a disease germ and after an appropriate time has elapsed, she examines each mouse individually and records whether or not it has contracted the disease. Suppose that the probability of contracting the disease from the inoculated germ is $^1/_6$ for each mouse and that the mice contract the disease independently of each other. What is the probability that

iii) no more than one mouse has contracted the disease,

iv) at least four mice have contracted the disease. [5 marks]

Suppose instead that the number, n say, of mice inoculated with the germ is increased to 150. Find the probability that

v) 15 or less mice contract the disease,

vi) 35 or more mice contract the disease

and deduce that the probability that the number of mice contracting the disease falls between 16 and 34, inclusive, is approximately 96%. [7 marks]

A scientist at a different laboratory continues to inoculate the mice with the disease germ until she finds two mice that have developed the disease. It may be assumed that at this laboratory also the probability that an inoculated mouse contracts the disease is $^1/_6$ for each mouse and that the mice contract the disease independently of each other. Find the probability that

vii) exactly two mice are required,

viii) exactly three mice are required,

ix) exactly eight mice are required.

[5 marks]

5. The reaction time, in seconds, of a laboratory animal to a stimulus is a continuous random variable, X, say, with probability density function

$$f(x) = \begin{cases} \left(\frac{3}{2}\right)\frac{1}{x^2} & 1 \le x \le 3, \\ 0 & \text{elsewhere.} \end{cases}$$

Verify that f(x) defines a valid probability density function. [3 marks]

Find the probability that the reaction time is

a) less than 1.5 seconds,

b) greater than 2.5 seconds. [5 marks]

Compute the expected reaction time, E(X). [4 marks]

If the animal takes more than 1.5 seconds to react, a light comes on and stays on until either one second has elapsed or until the animal reacts (whichever happens first). Let the random variable Y denote the length of time the light stays on. Deduce that Y is a mixture of continuous and discrete random variables defined by

$$Y = \begin{cases} 0 & \text{if } X < 1.5, \\ X\text{-}1.5 & \text{if } 1.5 \le X < 2.5, \\ 1 & \text{if } X \ge 2.5. \end{cases}$$

[3 marks]

Find

i) P(Y = 0), [1 mark]

ii) P(Y = 1), [1 mark]

iii) E(Y). [3 marks]

6. Human beta-endorphin (HBE) is a hormone secreted by the pituitary gland under condition of stress. A researcher conducted a study to investigate whether a program of regular exercise might affect the resting (unstressed) concentration of HBE in the blood. He measured blood HBE levels, in January and again in May, in ten participants in a physical fitness programme. The results were as shown in the following table (Lobstein, D.D., 1983):

PARTICIPANT	HBE LEVEL (pg/ml)		
	January	May	Difference
1	42	22	20
2	47	29	18
3	37	9	28
4	9	9	0
5	33	26	7
6	70	36	34
7	54	38	16
8	27	32	-5
9	41	33	8
10	18	14	4
Mean	37.8	24.8	13.0
SD	17.6	10.9	12.4

Construct a 95% confidence interval for the mean difference in HBE levels between January and May, and hence test the hypothesis that a program of regular exercise does ***not*** affect the resting concentration of HBE in the blood. [7 marks]

How many more participants would be needed to halve the length of this confidence interval, assuming that the sample standard deviation were to remain constant?

[4 marks]

Suppose that the researcher replicates the study with different but fewer subjects six months later with the following results:

Question 6 is continued overleaf

Q6 contd

PARTICIPANT	HBE LEVEL (pg/ml)		
	June	November	Difference
11	48	32	16
12	24	16	8
13	34	20	14
14	58	30	28
15	20	9	11
16	5	5	0
17	15	14	1
Mean	29.1	18.0	11.1
SD	18.7	10.1	9.6

Test whether the mean difference in HBE level for the second study is significantly different from that for the first study. [9 marks]

7. In a study of `Social Class' and `Season of Birth' the following data were considered:

Table: 10% sample of legitimate births, England and Wales, July 1963-June 1964, by Social class and three-monthly period of birth

Period of Birth	Social Class of F ather			Total
	I & II	III	IV & V	
Dec - Feb.	3266	10317	5063	18646
March - May	3743	11409	5411	20563
June - Aug	3443	10529	5182	19154
Sept - Nov	3266	9934	4902	18102
Total	13718	42189	20558	76465

These figures omit births to women who were unmarried or who were married to men who were in the armed forces, of `inadequately' described social class, students or unoccupied.

`Social Class' key:
- I = Professional etc. occupations
- II = Intermediate between I and III
- III = Skilled workers
- IV = Intermediate between III and V
- V = Unskilled workers

a) On ignoring the social class of father and considering only the `Total' column, test whether the number of births is uniformly distributed over the four periods, December-February, March-May, June-August and September-November, of three months each. [5 marks]

b) Test also whether there is evidence of an association between the Social Class of Father and the `Period of Birth'. [15 marks]