**1.** A school has 15 classes and a sixth form. In each class there are 30 students. In the sixth form there are 150 students. There are equal numbers of boys and girls in each class. There are equal numbers of boys and girls in the sixth form. The head teacher wishes to obtain the opinions of the students about school uniforms.

Explain how the head teacher would take a stratified sample of size 40.

**(7)**

---

**2.** A workshop makes two types of electrical resistor.

The resistance, $X$ ohms, of resistors of Type A is such that $X \sim N(20, 4)$.

The resistance, $Y$ ohms, of resistors of Type B is such that $Y \sim N(10, 0.84)$.

When a resistor of each type is connected into a circuit, the resistance $R$ ohms of the circuit is given by $R = X + Y$ where $X$ and $Y$ are independent.

Find

(*a*) E($R$),

**(1)**

(*b*) Var($R$),

**(2)**

(*c*) P($28.9 < R < 32.64$)

**(6)**

---

**3.** The drying times of paint can be assumed to be normally distributed. A paint manufacturer paints 10 test areas with a new paint. The following drying times, to the nearest minute, were recorded.

82, 98, 140, 110, 90, 125, 150, 130, 70, 110.

(*a*) Calculate unbiased estimates for the mean and the variance of the population of drying times of this paint.

**(5)**

Given that the population standard deviation is 25,

(*b*) find a 95% confidence interval for the mean drying time of this paint.

**(5)**

Fifteen similar sets of tests are done and the 95% confidence interval is determined for each set.

(*c*) Estimate the expected number of these 15 intervals that will enclose the true value of the population mean $\mu$.

**(2)**

---

**4.** People over the age of 65 are offered an annual flu injection. A health official took a random sample from a list of patients who were over 65. She recorded their gender and whether or not the offer of an annual flu injection was accepted or rejected. The results are summarised below.

| Gender | Accepted | Rejected |
|---|---|---|
| Male | 170 | 110 |
| Female | 280 | 140 |

Using a 5% significance level, test whether or not there is an association between gender and acceptance or rejection of an annual flu injection. State your hypotheses clearly.

**(9)**

---

**5.** Upon entering a school, a random sample of eight girls and an independent random sample of eighty boys were given the same examination in mathematics. The girls and boys were then taught in separate classes. After one year, they were all given another common examination in mathematics.

The means and standard deviations of the boys' and the girls' marks are shown in the table.

| Examination marks | | | | |
|---|---|---|---|---|
| | Upon entry | | After 1 year | |
| | Mean | Standard deviation | Mean | Standard deviation |
| Boys | 50 | 12 | 59 | 6 |
| Girls | 53 | 12 | 62 | 6 |

You may assume that the test results are normally distributed.

(*a*) Test, at the 5% level of significance, whether or not the difference between the means of the boys' and girls' results was significant when they entered school.

**(7)**

(*b*) Test, at the 5% level of significance, whether or not the mean mark of the boys is significantly less than the mean mark of the girls in the 'After 1 year' examination.

**(5)**

(*c*) Interpret the results found in part (*a*) and part (*b*).

**(1)**

---

**6.** An area of grass was sampled by placing a 1 m × 1 m square randomly in 100 places. The numbers of daisies in each of the squares were counted. It was decided that the resulting data could be modelled by a Poisson distribution with mean 2. The expected frequencies were calculated using the model.

The following table shows the observed and expected frequencies.

| Number of daisies | Observed frequency | Expected frequency |
|---|---|---|
| 0 | 8 | 13.53 |
| 1 | 32 | 27.07 |
| 2 | 27 | $r$ |
| 3 | 18 | $s$ |
| 4 | 10 | 9.02 |
| 5 | 3 | 3.61 |
| 6 | 1 | 1.20 |
| 7 | 0 | 0.34 |
| $\geq 8$ | 1 | $t$ |

(*a*) Find values for $r$, $s$ and $t$.

**(4)**

(*b*) Using a 5% significance level, test whether or not this Poisson model is suitable. State your hypotheses clearly.

**(7)**

An alternative test might have been to estimate the population mean by using the data given.

(*c*) Explain how this would have affected the test.

**(2)**

**7.** The numbers of deaths from pneumoconiosis and lung cancer in a developing country are given in the table.

| Age group (years) | 20–29 | 30–39 | 40–49 | 50–59 | 60–69 | 70 and over |
|---|---|---|---|---|---|---|
| Deaths from pneumoconiosis (1000s) | 12.5 | 5.9 | 18.5 | 19.4 | 31.2 | 31.0 |
| Deaths from lung cancer (1000s) | 3.7 | 9.0 | 10.2 | 19.0 | 13.0 | 18.0 |

The correlation between the number of deaths in the different age groups for each disease is to be investigated.

(*a*) Give **one** reason why Spearman’s rank correlation coefficient should be used.
**(1)**

(*b*) Calculate Spearman’s rank correlation coefficient for these data.
**(6)**

(*c*) Use a suitable test, at the 5% significance level, to interpret your result. State your hypotheses clearly.
**(5)**

---

**TOTAL FOR PAPER: 75 MARKS**

**END**

**mock papers 2**

**1.** Some biologists were studying a large group of wading birds. A random sample of 36 were measured and the wing length, $x$ mm of each wading bird was recorded. The results are summarised as follows

$$\sum x = 6046 \qquad \sum x^2 = 1\,016\,338$$

(a) Calculate unbiased estimates of the mean and the variance of the wing lengths of these birds.

**(3)**

Given that the standard deviation of the wing lengths of this particular type of bird is actually 5.1 mm,

(b) find a 99% confidence interval for the mean wing length of the birds from this group.

**(5)**

**Question 1 continued**

Q1

**(Total 8 marks)**

Turn over

**2.** Students in a mixed sixth form college are classified as taking courses in either Arts, Science or Humanities. A random sample of students from the college gave the following results

| | | Course | | |
|---|---|---|---|---|
| | | Arts | Science | Humanities |
| Gender | Boy | 30 | 50 | 35 |
| | Girl | 40 | 20 | 42 |

Showing your working clearly, test, at the 1% level of significance, whether or not there is an association between gender and the type of course taken. State your hypotheses clearly.

**(11)**

**Question 2 continued**

Turn over

**Question 2 continued**

**Question 2 continued**

**(Total 11 marks)**

Q2

Turn over

**3.** The product moment correlation coefficient is denoted by $r$ and Spearman's rank correlation coefficient is denoted by $r_s$.

(a) Sketch separate scatter diagrams, with five points on each diagram, to show

(i) $r = 1$,

(ii) $r_s = -1$ but $r > -1$.

**(3)**

Two judges rank seven collie dogs in a competition. The collie dogs are labelled $A$ to $G$ and the rankings are as follows

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Judge 1 | *A* | *C* | *D* | *B* | *E* | *F* | *G* |
| Judge 2 | *A* | *B* | *D* | *C* | *E* | *G* | *F* |

(b) (i) Calculate Spearman's rank correlation coefficient for these data.

**(6)**

(ii) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the judges are generally in agreement.

**(5)**

**Question 3 continued**

Turn over

**Question 3 continued**

**Question 3 continued**

Q3

**(Total 14 marks)**

Turn over

**4.** The weights of adult men are normally distributed with a mean of 84 kg and a standard deviation of 11 kg.

(a) Find the probability that the total weight of 4 randomly chosen adult men is less than 350 kg.

**(5)**

The weights of adult women are normally distributed with a mean of 62 kg and a standard deviation of 10 kg.

(b) Find the probability that the weight of a randomly chosen adult man is less than one and a half times the weight of a randomly chosen adult woman.

**(6)**

**Question 4 continued**

Turn over

**Question 4 continued**

**Question 4 continued**

**(Total 11 marks)**

Q4

**5.** A researcher is hired by a cleaning company to survey the opinions of employees on a proposed pension scheme. The company employs 55 managers and 495 cleaners.

To collect data the researcher decides to give a questionnaire to the first 50 cleaners to leave at the end of the day.

(a) Give 2 reasons why this method is likely to produce biased results.
**(2)**

(b) Explain briefly how the researcher could select a sample of 50 employees using

(i) a systematic sample,

(ii) a stratified sample.
**(6)**

Using the random number tables in the formulae book, and starting with the top left hand corner (8) and working across, 50 random numbers between 1 and 550 inclusive were selected. The first two suitable numbers are 384 and 100.

(c) Find the next two suitable numbers.
**(2)**

**Question 5 continued**

**(Total 10 marks)**

Q5

Turn over

**6.** Ten cuttings were taken from each of 100 randomly selected garden plants. The numbers of cuttings that did not grow were recorded.

The results are as follows

| No. of cuttings which did not grow | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8, 9 or 10 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency | 11 | 21 | 30 | 20 | 12 | 3 | 2 | 1 | 0 |

(a) Show that the probability of a randomly selected cutting, from this sample, not growing is 0.223

**(2)**

A gardener believes that a binomial distribution might provide a good model for the number of cuttings, out of 10, that do not grow.

He uses a binomial distribution, with the probability 0.2 of a cutting not growing. The calculated expected frequencies are as follows

| No. of cuttings which did not grow | 0 | 1 | 2 | 3 | 4 | 5 or more |
|---|---|---|---|---|---|---|
| Expected frequency | $r$ | 26.84 | $s$ | 20.13 | 8.81 | $t$ |

(b) Find the values of $r$, $s$ and $t$.

**(4)**

(c) State clearly the hypotheses required to test whether or not this binomial distribution is a suitable model for these data.

**(2)**

The test statistic for the test is 4.17 and the number of degrees of freedom used is 4.

(d) Explain fully why there are 4 degrees of freedom.

**(2)**

(e) Stating clearly the critical value used, carry out the test using a 5% level of significance.

**(3)**

**Question 6 continued**

Turn over

**Question 6 continued**

**Question 6 continued**

**(Total 13 marks)**

Q6

**7.** A sociologist is studying how much junk food teenagers eat. A random sample of 100 female teenagers and an independent random sample of 200 male teenagers were asked to estimate what their weekly expenditure on junk food was. The results are summarised below.

| | $n$ | mean | s.d. |
|---|---|---|---|
| Female teenagers | 100 | £5.48 | £3.62 |
| Male teenagers | 200 | £6.86 | £4.51 |

(a) Using a 5% significance level, test whether or not there is a difference in the mean amounts spent on junk food by male teenagers and female teenagers. State your hypotheses clearly.

**(7)**

(b) Explain briefly the importance of the central limit theorem in this problem.

**(1)**

**Question 7 continued**

Turn over

**Question 7 continued**

**Q7**

**(Total 8 marks)**

**TOTAL FOR PAPER: 75 MARKS**

**END**

**1** The lengths of rivets produced by a certain factory are checked each day by measuring a random sample of 100 rivets. For a particular day's sample the lengths, $x$ mm, are summarised by $\Sigma x = 761.2$ and $\Sigma x^2 = 6115.04$. The mean and standard deviation of the lengths of all rivets produced that day are denoted by $\mu$ mm and $\sigma$ mm respectively.

**(i)** Find an unbiased estimate of $\sigma^2$. [2]

**(ii)** Calculate a 95% confidence interval for $\mu$. [3]

**(iii)** Explain what distributional assumptions (if any) are required for the validity of your calculated confidence interval. [1]

**2** The saturated fat content of a particular brand of olive spread is monitored regularly in order to maintain a mean percentage content of 12.6. This is carried out by measuring the saturated fat content in random samples of 10 cartons. For a particular sample, the sample mean and an unbiased estimate of the population variance are calculated. The unbiased estimate of the population variance is 0.1195. It may be assumed that percentage fat content has a normal distribution. Find the critical region for a test at the 10% significance level of whether the population mean percentage fat content exceeds 12.6. [5]

**3** A large sample of people were surveyed and classified by 4 levels of income and by which of 3 newspapers they read. The results were arranged in a contingency table consisting of 4 columns and 3 rows. In a $\chi^2$ test of independence between income and choice of newspaper, it was found necessary to combine two of the columns. The value of the test statistic was 12.32.

**(i)** State a suitable null hypothesis for the test. [1]

**(ii)** Determine the largest significance level, obtained from tables or calculator, for which independence would be accepted. [3]

**4** The continuous random variable $X$ has probability density function given by

$$f(x) = \begin{cases} 0 & x < 0, \\ \frac{4}{3}x^3 & 0 \leqslant x \leqslant 1, \\ \frac{4}{3x^3} & x > 1. \end{cases}$$

**(i)** Find $P(X < 2)$. [3]

**(ii)** Show that the median of $X$ exceeds 1. [3]

**(iii)** Find $E(X)$. [3]

**(iv)** Show that $\mathrm{Var}(X)$ is not finite. [3]

**5** The proportion of syringes of brand $A$ that are faulty is 2.2%. The corresponding proportion for brand $B$ is 2.5%. Random samples of 75 brand $A$ and 90 brand $B$ syringes are taken and the total number of faulty syringes is denoted by $X$.

**(i)** Show that the distribution of $X$ can be approximated by a Poisson distribution, and state its mean. [5]

**(ii)** Find $P(X > 5)$. [2]

**6** The proportion of teapots with faulty spouts produced in a factory is denoted by $p$. In a random sample of 50 teapots, the number with faulty spouts was found to be 6.

**(i)** Find a 98% confidence interval for $p$. [4]

**(ii)** Find an estimate of the sample size for which the sample proportion would differ from $p$ by less than 0.05 with 98% confidence. [5]

**7** A psychologist believed that teenage boys worry more than teenage girls and he devised a questionnaire to examine his belief. He gave the questionnaire to a random sample of 24 girls and a random sample of 18 boys. The scores, $x_G$ and $x_B$ for the girls and boys, are summarised by $\Sigma x_G = 1526.8$ and $\Sigma x_B = 1238.4$. Unbiased estimates of the respective population variances, obtained from the samples, are $s_G^2 = 86.79$ and $s_B^2 = 93.01$. Larger scores indicate greater levels of worry.

**(i)** State two assumptions required for the validity of a $t$-test to examine the psychologist's belief. [2]

**(ii)** Comment on one of these assumptions in the light of the data. [1]

**(iii)** Carry out the test at the 5% significance level. [9]

**[Question 8 is printed overleaf.]**

**[Turn over**

**8** The numbers of goals scored by my local football team in 80 matches are summarised in the following table.

| Number of goals | 0 | 1 | 2 | 3 | 4 | 5 | $\geqslant 6$ |
|---|---|---|---|---|---|---|---|
| Number of matches | 11 | 15 | 33 | 16 | 2 | 3 | 0 |

**(i)** Show that the mean of the distribution is 1.9, and find the variance of the distribution. [3]

**(ii)** Without carrying out a test, explain whether the values of the mean and variance indicate that a Poisson distribution could be a suitable model for the number of goals scored in a match. [2]

The table below gives the expected frequencies, correct to 2 decimal places, for a $\chi^2$ goodness of fit test of a Poisson distribution.

| Number of goals | 0 | 1 | 2 | 3 | 4 | $\geqslant 5$ |
|---|---|---|---|---|---|---|
| Expected frequency | 11.97 | 22.73 | 21.60 | 13.68 | 6.50 | 3.52 |

**(iii)** Show how the value 13.68 for 3 goals is obtained. [2]

**(iv)** Stating a required assumption regarding the data, carry out the test at the 5% significance level. Does the outcome of the test confirm your answer to part **(ii)**? [8]

**(v)** Without further calculation, state two ways in which the test would be different if it were a goodness of fit test of the distribution Po(2), also at the 5% significance level. [2]

**1** For the variables $A$ and $B$, it is given that $\text{Var}(A) = 9$, $\text{Var}(B) = 6$ and $\text{Var}(2A - 3B) = 18$.

**(i)** Find $\text{Cov}(A, B)$. **[3]**

**(ii)** State with a reason whether $A$ and $B$ are independent. **[1]**

**2** The probability generating function of the discrete random variable $X$ is $\dfrac{e^{4t^2}}{e^4}$. Find

**(i)** $E(X)$, **[3]**

**(ii)** $P(X = 2)$. **[3]**

**3** $X_1$ and $X_2$ are continuous random variables. Random samples of 5 observations of $X_1$ and 6 observations of $X_2$ are taken. No two observations are equal. The 11 observations are ranked, lowest first, and the sum of the ranks of the observations of $X_1$ is denoted by $R$.

**(i)** Assuming that all rankings are equally likely, show that $P(R \leqslant 17) = \frac{2}{231}$. **[5]**

The marks of 5 randomly chosen students from School $A$ and 6 randomly chosen students from School $B$, who took the same examination, achieving different marks, were ranked. The rankings are shown in the table.

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| School | *A* | *A* | *A* | *B* | *A* | *A* | *B* | *B* | *B* | *B* | *B* |

**(ii)** For a Wilcoxon rank-sum test, obtain the exact smallest significance level for which there is evidence of a difference in performance at the two schools. **[2]**

**4** The moment generating function of a continuous random variable $Y$, which has a $\chi^2$ distribution with $n$ degrees of freedom, is $(1 - 2t)^{-\frac{1}{2}n}$, where $0 \leqslant t < \frac{1}{2}$.

**(i)** Find $E(Y)$ and $\text{Var}(Y)$. **[5]**

For the case $n = 1$, the sum of 60 independent observations of $Y$ is denoted by $S$.

**(ii)** Write down the moment generating function of $S$ and hence identify the distribution of $S$. **[2]**

**(iii)** Use a normal approximation to estimate $P(S \geqslant 70)$. **[3]**

**5** In order to test whether the median salary of employees in a certain industry who had worked for three years was £19 500, the salaries $x$, in thousands of pounds, of 50 randomly chosen employees were obtained.

**(i)** The values $|x - 19.5|$ were calculated and ranked. No two values of $x$ were identical and none was equal to 19.5. The sum of the ranks corresponding to positive values of $(x - 19.5)$ was 867. Stating a required assumption, carry out a suitable test at the 5% significance level. **[10]**

**(ii)** If the assumption you stated in part **(i)** does not hold, what test could have been used? **[1]**

**6** Nuts and raisins occur in randomly chosen squares of a particular brand of chocolate. The numbers of nuts and raisins are denoted by $N$ and $R$ respectively and the joint probability distribution of $N$ and $R$ is given by

$$f(n, r) = \begin{cases} c(n+2r) & n = 0, 1, 2 \text{ and } r = 0, 1, 2, \\ 0 & \text{otherwise,} \end{cases}$$

where $c$ is a constant.

**(i)** Find the value of $c$. **[3]**

**(ii)** Find the probability that there is exactly one nut in a randomly chosen square. **[2]**

**(iii)** Find the probability that the total number of nuts and raisins in a randomly chosen square is more than 2. **[2]**

**(iv)** For squares in which there are 2 raisins, find the mean number of nuts. **[4]**

**(v)** Determine whether $N$ and $R$ are independent. **[2]**

**7** The continuous random variable $X$ has probability density function given by

$$f(x) = \begin{cases} \dfrac{x}{2\theta^2} & 0 \leqslant x \leqslant 2\theta, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta$ is an unknown positive constant.

**(i)** Find $E(X^n)$, where $n \neq -2$, and hence write down the value of $E(X)$. **[3]**

**(ii)** Find

**(a)** $\text{Var}(X)$,

**(b)** $\text{Var}(X^2)$.

**[4]**

**(iii)** Find $E(X_1 + X_2 + X_3)$ and $E(X_1^2 + X_2^2 + X_3^2)$, where $X_1$, $X_2$ and $X_3$ are independent observations of $X$. Hence construct unbiased estimators, $T_1$ and $T_2$, of $\theta$ and $\text{Var}(X)$ respectively, which are based on $X_1$, $X_2$ and $X_3$. **[6]**

**(iv)** Find $\text{Var}(T_2)$. **[2]**

**8** For the events $L$ and $M$, $P(L \mid M) = 0.2$, $P(M \mid L) = 0.4$ and $P(M) = 0.6$.

**(i)** Find $P(L)$ and $P(L' \cup M')$. **[3]**

**(ii)** Given that, for the event $N$, $P\big(N \mid (L \cap M)\big) = 0.3$, find $P(L' \cup M' \cup N')$. **[3]**

1 A random variable has a normal distribution with unknown mean $\mu$ and known standard deviation 0.19. In order to estimate $\mu$ a random sample of five observations of the random variable was taken. The values were as follows.

5.44 4.93 5.12 5.36 5.40

Using these five values, calculate,

**(i)** an estimate of $\mu$, **[1]**

**(ii)** a 95% confidence interval for $\mu$. **[4]**

2 In a Year 8 internal examination in a large school the Geography marks, $G$, and Mathematics marks, $M$, had means and standard deviations as follows.

| | Mean | Standard deviation |
|---|---|---|
| $G$ | 36.42 | 6.87 |
| $M$ | 42.65 | 10.25 |

Assuming that $G$ and $M$ have independent normal distributions, find the probability that a randomly chosen Geography candidate scores at least 10 marks more than a randomly chosen Mathematics candidate. Do not use a continuity correction. **[5]**

3 The continuous random variable $T$ has probability density function given by

$$f(t) = \begin{cases} 0 & t < 0, \\ \dfrac{a}{e} & 0 \leqslant t < 2, \\ ae^{-\frac{1}{2}t} & t \geqslant 2, \end{cases}$$

where $a$ is a positive constant.

**(i)** Show that $a = \frac{1}{4}e$. **[3]**

**(ii)** Find the upper quartile of $T$. **[4]**

4 A study in 1981 investigated the effect of water fluoridation on children's dental health. In a town with fluoridation, 61 out of a random sample of 107 children showed signs of increased tooth decay after six months. In a town without fluoridation the corresponding number was 106 out of a random sample of 143 children. The population proportions of children with increased tooth decay are denoted by $p_1$ and $p_2$ for the towns with fluoridation and without fluoridation respectively. A test is carried out of the null hypothesis $p_1 = p_2$ against the alternative hypothesis $p_1 < p_2$. Find the smallest significance level at which the null hypothesis is rejected. **[7]**

**5** An experiment with hybrid corn resulted in yellow kernels and purple kernels. Of a random sample of 90 kernels, 18 were yellow and 72 were purple.

**(i)** Calculate an approximate 90% confidence interval for the proportion of yellow kernels produced in all such experiments. **[4]**

**(ii)** Deduce an approximate 90% confidence interval for the proportion of purple kernels produced in all such experiments. **[1]**

**(iii)** Explain what is meant by a 90% confidence interval for a population proportion. **[2]**

**(iv)** Mendel's theory of inheritance predicts that 25% of all such kernels will be yellow. State, giving a reason, whether or not your calculations support the theory. **[2]**

**6** The continuous random variable $X$ has (cumulative) distribution function given by

$$F(x) = \begin{cases} 0 & x < \frac{1}{2}, \\ \dfrac{2x-1}{x+1} & \frac{1}{2} \leqslant x \leqslant 2, \\ 1 & x > 2. \end{cases}$$

**(i)** Given that $Y = \dfrac{1}{X}$, find the (cumulative) distribution function of $Y$, and deduce that $Y$ and $X$ have identical distributions. **[6]**

**(ii)** Find $E(X + 1)$ and deduce the value of $E\left(\dfrac{1}{X}\right)$. **[6]**

**7** **(i)** When should Yates' correction be applied when carrying out a $\chi^2$ test? **[1]**

Two vaccines against typhoid fever, $A$ and $B$, were tested on a total of 700 people in Nepal during a particular year. The vaccines were allocated randomly and whether or not typhoid had developed was noted during the following year. The results are shown in the table.

| | Vaccines | |
|---|---|---|
| | *A* | *B* |
| Developed typhoid | 19 | 4 |
| Did not develop typhoid | 310 | 367 |

**(ii)** Carry out a suitable $\chi^2$ test at the 1% significance level to determine whether the outcome depends on the vaccine used. Comment on the result. **[10]**

**[Question 8 is printed overleaf.]**

**Turn over**

**8** **(i)** State circumstances under which it would be necessary to calculate a pooled estimate of variance when carrying out a two-sample hypothesis test. **[1]**

**(ii)** An investigation into whether passive smoking affects lung capacity considered a random sample of 20 children whose parents did not smoke and a random sample of 22 children whose parents did smoke. None of the children themselves smoked. The lung capacity, in litres, of each child was measured and the results are summarised as follows.

For the children whose parents did not smoke: $n_1 = 20$, $\Sigma x_1 = 42.4$ and $\Sigma x_1^2 = 90.43$.

For the children whose parents did smoke: $n_2 = 22$, $\Sigma x_2 = 42.5$ and $\Sigma x_2^2 = 82.93$.

The means of the two populations are denoted by $\mu_1$ and $\mu_2$ respectively.

**(a)** State conditions for which a *t*-test would be appropriate for testing whether $\mu_1$ exceeds $\mu_2$. **[1]**

**(b)** Assuming the conditions are valid, carry out the test at the 1% significance level and comment on the result. **[11]**

**(c)** Calculate a 99% confidence interval for $\mu_1 - \mu_2$. **[3]**

**1** At a particular hospital, admissions of patients as a result of visits to the Accident and Emergency Department occur randomly at a uniform average rate of 0.75 per day. Independently, admissions that result from G.P. referrals occur randomly at a uniform average rate of 6.4 per *week*. The total number of admissions from these two causes over a randomly chosen period of four weeks is denoted by $T$. State the distribution of $T$ and obtain its expectation and variance. **[4]**

**2** The continuous random variable $U$ has (cumulative) distribution function given by

$$F(u) = \begin{cases} \frac{1}{5}e^{u} & u < 0, \\ 1 - \frac{4}{5}e^{-\frac{1}{4}u} & u \geqslant 0. \end{cases}$$

**(i)** Find the upper quartile of $U$. **[3]**

**(ii)** Find the probability density function of $U$. **[2]**

**3** In a random sample of credit card holders, it was found that 28% of them used their card for internet purchases.

**(i)** Given that the sample size is 1200, find a 98% confidence interval for the percentage of all credit card holders who use their card for internet purchases. **[4]**

**(ii)** Estimate the smallest sample size for which a 98% confidence interval would have a width of at most 5%, and state why the value found is only an estimate. **[4]**

**4** The weekly sales of petrol, $X$ thousand litres, at a garage may be modelled by a continuous random variable with probability density function given by

$$f(x) = \begin{cases} c & 25 \leqslant x \leqslant 45, \\ 0 & \text{otherwise}, \end{cases}$$

where $c$ is a constant. The weekly profit, in £, is given by $(400\sqrt{X} - 240)$.

**(i)** Obtain the value of $c$. **[1]**

**(ii)** Find the expected weekly profit. **[3]**

**(iii)** Find the probability that the weekly profit exceeds £2000. **[3]**

**5** The concentration level of mercury in a large lake is known to have a normal distribution with standard deviation 0.24 in suitable units. At the beginning of June 2008, the mercury level was measured at five randomly chosen places on the lake, and the sample mean is denoted by $\bar{x}_1$. Towards the end of June 2008 there was a spillage in the lake which may have caused the mercury level to rise. Because of this the level was then measured at six randomly chosen points of the lake, and the mean of this sample is denoted by $\bar{x}_2$.

**(i)** State hypotheses for a test based on the two samples for whether, on average, the level of mercury had increased. Define any parameters that you use. **[2]**

**(ii)** Find the set of values of $\bar{x}_2 - \bar{x}_1$ for which there would be evidence at the 5% significance level that, on average, the level of mercury had increased. **[4]**

**(iii)** Given that the average level had actually increased by 0.3 units, find the probability of making a Type II error in your test, and comment on its value. **[4]**

**6** A mathematics examination is taken by 29 boys and 26 girls. Experience has shown that the probability that any boy forgets to bring a calculator to the examination is 0.3, and that any girl forgets is 0.2. Whether or not any student forgets to bring a calculator is independent of all other students. The numbers of boys and girls who forget to bring a calculator are denoted by $B$ and $G$ respectively, and $F = B + G$.

**(i)** Find $\mathrm{E}(F)$ and $\mathrm{Var}(F)$. **[5]**

**(ii)** Using suitable approximations to the distributions of $B$ and $G$, which should be justified, find the smallest number of spare calculators that should be available in order to be at least 99% certain that all 55 students will have a calculator. **[8]**

**7** A tutor gives a randomly selected group of 8 students an English Literature test, and after a term's further teaching, she gives the group a similar test. The marks for the two tests are given in the table.

| Student | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* |
|---|---|---|---|---|---|---|---|---|
| First test | 38 | 27 | 55 | 43 | 32 | 24 | 51 | 46 |
| Second test | 37 | 26 | 57 | 43 | 30 | 26 | 54 | 48 |

**(i)** Stating a necessary condition, show by carrying out a suitable $t$-test, at the 1% significance level, that the marks do not give evidence of an improvement. **[8]**

**(ii)** The tutor later found that she had marked the second test too severely, and she decided to add a constant amount $k$ to each mark. Find the least integer value of $k$ for which the increased marks would give evidence of improvement at the 1% significance level. **[3]**

**[Question 8 is printed overleaf.]**

**Turn over**

**8** A soft drinks factory produces lemonade which is sold in packs of 6 bottles. As part of the factory's quality control, random samples of 75 packs are examined at regular intervals. The number of underfilled bottles in a pack of 6 bottles is denoted by the random variable $X$. The results of one quality control check are shown in the following table.

| Number of underfilled bottles | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Number of packs | 44 | 20 | 8 | 3 |

A researcher assumes that $X \sim B(3, p)$.

**(i)** By finding the sample mean, show that an estimate of $p$ is 0.2. **[3]**

**(ii)** Show that, at the 5% significance level, there is evidence that this binomial distribution does not fit the data. **[10]**

**(iii)** Another researcher suggests that the goodness of fit test should be for $B(6, p)$. She finds that the corresponding value of $\chi^2$ is 2.74, correct to 3 significant figures. Given that the number of degrees of freedom is the same as in part **(ii)**, state the conclusion of the test at the same significance level. **[1]**