TABLE OF CONTENTS

Chapter 1: Poisson Distribution

Consider these random variables

- the number of emergency calls received by an ambulance control in an hour,
- the number of vehicles approaching a motorway toll bridge in a five-minute interval,
- the number of flaws in a metre length of material,
- the number of white corpuscles on a slide.

Assuming that each occurs randomly, they are all examples of variables that can be modelled using a Poisson distribution.

CONDITIONS FOR A POISSON MODEL

- Events occur singly and at random in a given interval of time or space,
- λ, the mean number of occurrences in the given interval, is known and is finite.

The variable X is the number of occurrences in the given interval.

If the above conditions are satisfied, X is said to follow a Poisson distribution, written

$X \sim \text{Po}(\lambda)$, where

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, 3, \dots \text{ to infinity}$$

c

Example 1

A student finds that the average number of amoebas in 10 ml of pond water from a particular pond is four. Assuming that the number of amoebas follows a Poisson distribution,

find the probability that in a 10 ml sample

(a) there are exactly five amoebas,
(b) there are no amoebas,
(c) there are fewer than three amoebas.

These two results are useful in general

If $X \sim \text{Po}(\lambda)$,

then $P(X = 0) = e^{-\lambda}$ and $P(X = 1) = \lambda e^{-\lambda}$

UNIT INTERVAL

Care must be taken to specify the interval being considered.
In Example 5.18 the mean number of amoebas in 10 ml of pond water from a particular pond is four so the number in 10 ml is distributed Po(4).
Now suppose you want to find a probability relating to the number of amoebas in 5 ml of water from the same pond. The mean number of amoebas in 5 ml is two, so the number in 5 ml is distributed Po(2).
Similarly, the number of amoebas in 1 ml of pond water is distributed Po(0.4).

Example 2

On average the school photocopier breaks down eight times during the school week (Monday to Friday). Assuming that the number of breakdowns can be modelled by a Poisson distribution, find the probability that it breaks down

(a) five times in a given week,
(b) once on Monday,
(c) eight times in a fortnight.

MEAN AND VARIANCE OF THE POISSON DISTRIBUTION

The mean number of occurrences in the interval, λ, is all that is needed to define the distribution completely; λ is the only parameter of the distribution.

In a Poisson distribution, it is obvious that the mean, $E(X) = \lambda$, but it is also the case that $\mathrm{Var}(X) = \lambda$. The following should be learnt:

If $X \sim \mathrm{Po}(\lambda)$
$E(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$

Example 3

X follows a Poisson distribution with standard deviation 1.5. Find $P(X \geqslant 3)$.

MODE OF THE POISSON DISTRIBUTION

The mode is the value of X that is most likely to occur, i.e. the one with the greatest probability.

From the diagrams, you can see that

when $\lambda = 1$, there are two modes, 0 and 1,
when $\lambda = 2$, there are two modes, 1 and 2,
when $\lambda = 3$, there are two modes, 2 and 3.

In general, if λ is an integer, there are two modes, $\lambda - 1$ and λ.

For example, if $X \sim \mathrm{Po}(8)$, the modes are 7 and 8.

Notice also that

when $\lambda = 1.6$, the mode is 1,
when $\lambda = 2.2$, the mode is 2,
when $\lambda = 3.8$, the mode is 3.

In general, if λ is not an integer, the mode is the integer below λ.

For example, if $X \sim \mathrm{Po}(4.9)$, the mode is 4.

Example 4

People arrive randomly and independently at the elevator in a block of flats at an average rate of 4 people every 5 minutes.

(i) Find the probability that exactly two people arrive in a 1-minute period.

(ii) Find the probability that nobody arrives in a 15-second period.

(iii) The probability that at least one person arrives in the next t minutes is 0.9. Find the value of t.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q6 June 2008]

Example 5

A shopkeeper sells electric fans. The demand for fans follows a Poisson distribution with mean 3.2 per week.

(i) Find the probability that the demand is exactly 2 fans in any one week.

(ii) The shopkeeper has 4 fans in his shop at the beginning of a week. Find the probability that this will not be enough to satisfy the demand for fans in that week.

(iii) Given instead that he has n fans in his shop at the beginning of a week, find, by trial and error, the least value of n for which the probability of his not being able to satisfy the demand for fans in that week is less than 0.05.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q6 November 2005]

Example 6

People arrive randomly and independently at a supermarket checkout at an average rate of 2 people every 3 minutes.

(i) Find the probability that exactly 4 people arrive in a 5-minute period.

At another checkout in the same supermarket, people arrive randomly and independently at an average rate of 1 person each minute.

(ii) Find the probability that a total of fewer than 3 people arrive at the two checkouts in a 3-minute period.

[Cambridge International AS and A Level Mathematics 9709, Paper 71 Q2 November 2010]

USING THE POISSON DISTRIBUTION AS AN APPROXIMATION TO THE BINOMIAL DISTRIBUTION

When n is large ($n > 50$) and p is small ($p < 0.1$), the binomial distribution

$X \sim B(n, p)$

can be approximated using a Poisson distribution with the same mean, i.e. $X \sim \text{Po}(np)$. The approximation gets better as n gets larger and p gets smaller.

Example 7

Eggs are packed into boxes of 500. On average 0.7% of the eggs are found to be broken when the eggs are unpacked. Find, correct to two significant figures, the probability that in a box of 500 eggs,

(a) exactly three are broken,
(b) at least two are broken.

Example 8

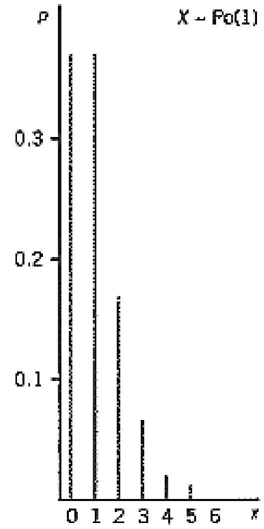
A Christmas draw aims to sell 5000 tickets, 50 of which will win a prize.

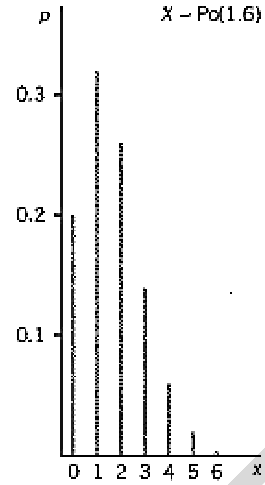
(a) A syndicate buys 200 tickets. Let X represent the number of these tickets that win a prize.
  (i) Justify the use of the Poisson approximation for the distribution of X.
  (ii) Calculate $P(X \leqslant 3)$.

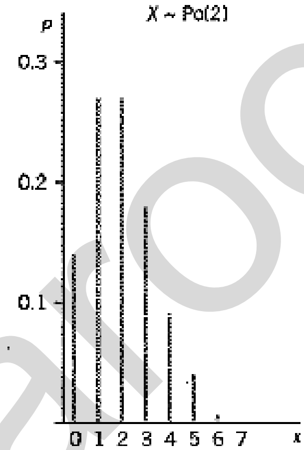
(b) Calculate how many tickets should be bought in order for there to be a 90% probability of winning at least one prize. (C)

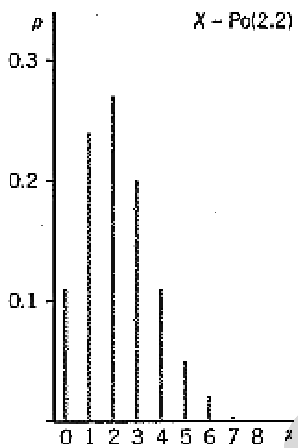
DIAGRAMMATIC REPRESENTATION OF THE POISSON DISTRIBUTION

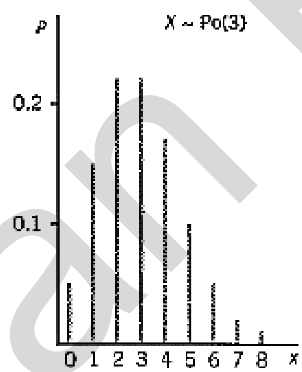
Notice that for small values of λ, the distribution is very skew, but it becomes more symmetrical as λ increases.

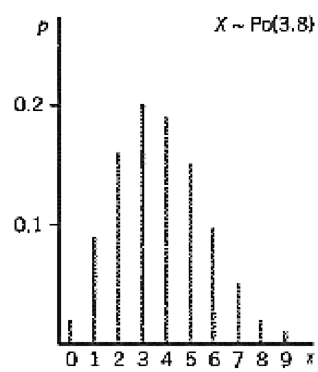


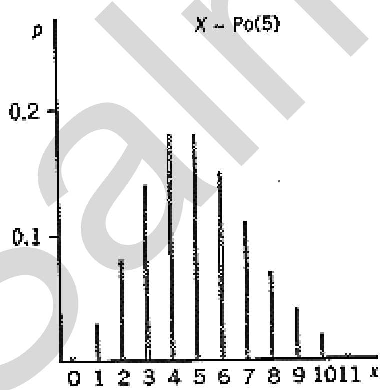


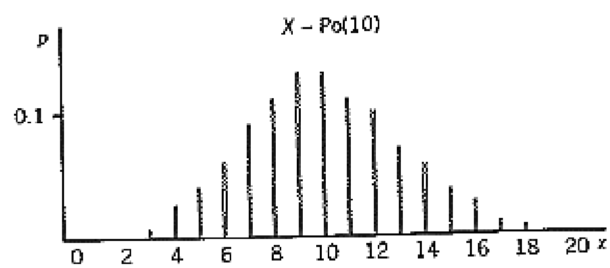












THE NORMAL APPROXIMATION TO THE POISSON DISTRIBUTION

If X follows a Poisson distribution with parameter λ, i.e. $X \sim \text{Po}(\lambda)$,

then $E(X) = \lambda$ and $\text{Var}(X) = \lambda$

When λ is large (say $\lambda > 15$), the normal distribution can be used as an approximation, where

$X \sim N(\lambda, \lambda)$.

As with the normal approximation to the binomial distribution, a continuity correction is needed, since you are using a continuous distribution as an approximation to a discrete one.

Example 9

A radioactive disintegration gives counts that follow a Poisson distribution with a mean count of 25 per second.

Find the probability that in a one-second interval the count is between 23 and 27 inclusive.

Example 10

At a petrol station cars arrive independently and at random times at constant average rates of 8 cars per hour travelling east and 5 cars per hour travelling west.

(i) Find the probability that, in a quarter-hour period

(a) one or more cars travelling east and one or more cars travelling west will arrive,

(b) a total of 2 or more cars will arrive.

(ii) Find the approximate probability that, in a 12-hour period, a total of more than 175 cars will arrive.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q6 June 2005]

Example 11

A dressmaker makes dresses for Easifit Fashions. Each dress requires 2.5 m^2 of material. Faults occur randomly in the material at an average rate of 4.8 per 20 m^2.

(i) Find the probability that a randomly chosen dress contains at least 2 faults.

Each dress has a belt attached to it to make an outfit. Independently of faults in the material, the probability that a belt is faulty is 0.03. Find the probability that, in an outfit,

(ii) neither the dress nor its belt is faulty,

(iii) the dress has at least one fault and its belt is faulty.

The dressmaker attaches 300 randomly chosen belts to 300 randomly chosen dresses. An outfit in which the dress has at least one fault and its belt is faulty is rejected.

(iv) Use a suitable approximation to find the probability that fewer than 3 outfits are rejected.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q6 June 2006]

Example 12

It is proposed to model the number of people per hour calling a car breakdown service between the times 0900 and 2100 by a Poisson distribution.

(i) Explain why a Poisson distribution may be appropriate for this situation.

People call the car breakdown service at an average rate of 20 per hour, and a Poisson distribution may be assumed to be a suitable model.

(ii) Find the probability that exactly 8 people call in any half hour.

(iii) By using a suitable approximation, find the probability that exactly 250 people call in the 12 hours between 0900 and 2100.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q5 June 2007]

Example 13

Major avalanches can be regarded as randomly occurring events. They occur at a uniform average rate of 8 per year.

(i) Find the probability that more than 3 major avalanches occur in a 3-month period.

(ii) Find the probability that any two separate 4-month periods have a total of 7 major avalanches.

(iii) Find the probability that a total of fewer than 137 major avalanches occur in a 20-year period.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q3 June 2009]

THE SUM OF INDEPENDENT POISSON VARIABLES

For independent variables, X and Y, if $X \sim \text{Po}(m)$ and $Y \sim \text{Po}(n)$,

then $X + Y \sim \text{Po}(m + n)$

Example 14

Two identical racing cars are being tested on a circuit. For each car, the number of mechanical breakdowns can be modelled by a Poisson distribution with a mean of one breakdown in 100 laps. If a car breaks down it is attended and continues on the circuit. The first car is tested for 20 laps and the second car for 40 laps.

Find the probability that the service team is called out to attend to breakdowns

(a) once,
(b) more than twice.

SELECTED PAST PAPER QUESTIONS

Question 1
N2011/72/Q6

Customers arrive at an enquiry desk at a constant average rate of 1 every 5 minutes.

(i) State one condition for the number of customers arriving in a given period to be modelled by a Poisson distribution. [1]

Assume now that a Poisson distribution is a suitable model.

(ii) Find the probability that exactly 5 customers will arrive during a randomly chosen 30-minute period. [2]

(iii) Find the probability that fewer than 3 customers will arrive during a randomly chosen 12-minute period. [3]

(iv) Find an estimate of the probability that fewer than 30 customers will arrive during a randomly chosen 2-hour period. [4]

Homework: The Poisson distribution – Variant 72

1 Computer breakdowns occur randomly on average once every 48 hours of use.

(i) Calculate the probability that there will be fewer than 4 breakdowns in 60 hours of use. [3]

(ii) Find the probability that the number of breakdowns in one year (8760 hours) of use is more than 200. [4]

(iii) Independently of the computer breaking down, the computer operator receives phone calls randomly on average twice in every 24-hour period. Find the probability that the total number of phone calls and computer breakdowns in a 60-hour period is exactly 4. [3]

Answers: **(i)** 2.5, 1.25; **(ii)** 5, 5. **J03/Q6**

2 At a certain airfield planes land at random times at a constant average rate of one every 10 minutes.

(i) Find the probability that exactly 5 planes will land in a period of one hour. [2]

(ii) Find the probability that at least 2 planes will land in a period of 16 minutes. [3]

(iii) Given that 5 planes landed in an hour, calculate the conditional probability that 1 plane landed in the first half hour and 4 in the second half hour. [3]

Answers: **(i)** 0.161; **(ii)** 0.475; **(iii)** 0.156. **J04/Q6**

3 At a petrol station cars arrive independently and at random times at constant average rates of 8 cars per hour travelling east and 5 cars per hour travelling west.

(i) Find the probability that, in a quarter-hour period,

(a) one or more cars travelling east and one or more cars travelling west will arrive, [4]

(b) a total of 2 or more cars will arrive. [2]

(ii) Find the approximate probability that, in a 12-hour period, a total of more than 175 cars will arrive. [3]

Answers: **(i)(a)** 0.617, **(b)** 0.835; **(ii)** 0.0593. **J05/Q6**

4 A dressmaker makes dresses for Easifit Fashions. Each dress requires 2.5 m^2 of material. Faults occur randomly in the material at an average rate of 4.8 per 20 m^2.

(i) Find the probability that a randomly chosen dress contains at least 2 faults. [3]

Each dress has a belt attached to it to make an outfit. Independently of faults in the material, the probability that a belt is faulty is 0.03. Find the probability that, in an outfit,

(ii) neither the dress nor its belt is faulty, [2]

(iii) the dress has at least one fault and its belt is faulty. [2]

The dressmaker attaches 300 randomly chosen belts to 300 randomly chosen dresses. An outfit in which the dress has at least one fault and its belt is faulty is rejected.

(iv) Use a suitable approximation to find the probability that fewer than 3 outfits are rejected. [3]

Answers: **(i)** 0.122; **(ii)** 0.532; **(iii)** 0.0135; **(iv)** 0.229. **J06/Q6**

5 It is proposed to model the number of people per hour calling a car breakdown service between the times 09 00 and 21 00 by a Poisson distribution.

(i) Explain why a Poisson distribution may be appropriate for this situation. [2]

People call the car breakdown service at an average rate of 20 per hour, and a Poisson distribution may be assumed to be a suitable model.

(ii) Find the probability that exactly 8 people call in any half hour. [2]

(iii) By using a suitable approximation, find the probability that exactly 250 people call in the 12 hours between 09 00 and 21 00. [4]

Answers: (i) People call randomly, independently, at an average uniform rate; (ii) 0.113; (iii) 0.0211 . **J07/Q5**

6 People arrive randomly and independently at the elevator in a block of flats at an average rate of 4 people every 5 minutes.

(i) Find the probability that exactly two people arrive in a 1-minute period. [2]

(ii) Find the probability that nobody arrives in a 15-second period. [2]

(iii) The probability that at least one person arrives in the next t minutes is 0.9. Find the value of t. [4]

Answers: (i) 0.144; (ii) 0.819; (iii) 2.88. **J08/Q6**

7 Major avalanches can be regarded as randomly occurring events. They occur at a uniform average rate of 8 per year.

(i) Find the probability that more than 3 major avalanches occur in a 3-month period. [3]

(ii) Find the probability that any two separate 4-month periods have a total of 7 major avalanches. [3]

(iii) Find the probability that a total of fewer than 137 major avalanches occur in a 20-year period. [4]

Answers: (i) 0.143
(ii) 0.118
(iii) 0.0316 **J09/71/Q3**

8 At a certain point on a road, cars pass by at random times at a constant average rate of 3 cars every 2 minutes.

(i) Find the probability that at most 2 cars pass by in a period of 3 minutes. [3]

(ii) Find the probability that the total number of cars that pass by in two separate periods of 1 minute and 4 minutes is 6. [3]

(iii) Find the probability that more than 100 cars pass by in a period of 1 hour. [4]

Answers: **(i)** 0.174
(ii) 0.137
(iii) 0.134

J09/72/Q3

---

9 In restaurant A an average of 2.2% of tablecloths are stained and, independently, in restaurant B an average of 5.8% of tablecloths are stained.

(i) Random samples of 55 tablecloths are taken from each restaurant. Use a suitable Poisson approximation to find the probability that a total of more than 2 tablecloths are stained. [4]

(ii) Random samples of n tablecloths are taken from each restaurant. The probability that at least one tablecloth is stained is greater than 0.99. Find the least possible value of n. [4]

Answers: **(i)** 0.815; **(ii)** 58.

J10/72/Q6

---

10 1.5% of the population of the UK can be classified as 'very tall'.

(i) The random variable X denotes the number of people in a sample of n people who are classified as very tall. Given that $E(X) = 2.55$, find n. [2]

(ii) By using the Poisson distribution as an approximation to a binomial distribution, calculate an approximate value for the probability that a sample of size 210 will contain fewer than 3 people who are classified as very tall. [3]

\- *Answers*: **(i)** $n = 170$; **(ii)** 0.390 .

N02/Q2

---

11 X and Y are independent random variables each having a Poisson distribution. X has mean 2.5 and Y has mean 3.1.

(i) Find $P(X + Y > 3)$. [4]

(ii) A random sample of 80 values of X is taken. Find the probability that the sample mean is less than 2.4. [4]

Answers: **(i)** 0.809; **(ii)** 0.286 .

N02/Q5

---

12 A certain machine makes matches. One match in 10 000 on average is defective. Using a suitable approximation, calculate the probability that a random sample of 45 000 matches will include 2, 3 or 4 defective matches. [5]

Answer: 0.471.

N03/Q2

13 The number of emergency telephone calls to the electricity board office in a certain area in t minutes is known to follow a Poisson distribution with mean $\frac{1}{80}t$.

(i) Find the probability that there will be at least 3 emergency telephone calls to the office in any 20-minute period. [4]

(ii) The probability that no emergency telephone call is made to the office in a period of k minutes is 0.9. Find k. [4]

Answers: **(i)** 0.00216; **(ii)** 8.43. **N03/Q4**

14 The number of radioactive particles emitted per second by a certain metal is random and has mean 1.7. The radioactive metal is placed next to an object which independently emits particles at random such that the mean number of particles emitted per second is 0.6. Find the probability that the total number of particles emitted in the next 3 seconds is 6, 7 or 8. [4]

Answer: 0.428. **N04/Q1**

15 Of people who wear contact lenses, 1 in 1500 on average have laser treatment for short sight.

(i) Use a suitable approximation to find the probability that, of a random sample of 2700 contact lens wearers, more than 2 people have laser treatment. [4]

(ii) In a random sample of n contact lens wearers the probability that no one has laser treatment is less than 0.01. Find the least possible value of n. [3]

Answers: **(i)** 0.269; **(ii)** 6908 or 6906. **N04/Q5**

16 A shopkeeper sells electric fans. The demand for fans follows a Poisson distribution with mean 3.2 per week.

(i) Find the probability that the demand is exactly 2 fans in any one week. [2]

(ii) The shopkeeper has 4 fans in his shop at the beginning of a week. Find the probability that this will not be enough to satisfy the demand for fans in that week. [4]

(iii) Given instead that he has n fans in his shop at the beginning of a week, find, by trial and error, the least value of n for which the probability of his not being able to satisfy the demand for fans in that week is less than 0.05. [4]

Answers: **(i)** 0.209; **(ii)** 0.219; **(iii)** $n = 6$. **N05/Q6**

17 In summer, wasps' nests occur randomly in the south of England at an average rate of 3 nests for every 500 houses.

(i) Find the probability that two villages in the south of England, with 600 houses and 700 houses, have a total of exactly 3 wasps' nests. [3]

(ii) Use a suitable approximation to estimate the probability of there being fewer than 369 wasps' nests in a town with 64 000 houses. [4]

Answers: **(i)** 0.0324; **(ii)** 0.215. **N06/Q4**

19 The random variable X denotes the number of worms on a one metre length of a country path after heavy rain. It is given that X has a Poisson distribution.

(i) For one particular path, the probability that $X = 2$ is three times the probability that $X = 4$. Find the probability that there are more than 3 worms on a 3.5 metre length of this path. [5]

(ii) For another path the mean of X is 1.3.

(a) On this path the probability that there is at least 1 worm on a length of k metres is 0.96. Find k. [4]

(b) Find the probability that there are more than 1250 worms on a one kilometre length of this path. [3]

Answers: **(i)** 0.918; **(ii)(a)** 2.48, **(b)** 0.915. **N07/Q6**

20 In their football matches, Rovers score goals independently and at random times. Their average rate of scoring is 2.3 goals per match.

(i) State the expected number of goals that Rovers will score in the first half of a match. [1]

(ii) Find the probability that Rovers will not score any goals in the first half of a match but will score one or more goals in the second half of the match. [2]

(iii) Football matches last for 90 minutes. In a particular match, Rovers score one goal in the first 30 minutes. Find the probability that they will score at least one further goal in the remaining 60 minutes. [3]

Independently of the number of goals scored by Rovers, the number of goals scored per football match by United has a Poisson distribution with mean 1.8.

(iv) Find the probability that a total of at least 3 goals will be scored in a particular match when Rovers play United. [3]

Answers: **(i)** 1.15; **(ii)** 0.216; **(iii)** 0.784; **(iv)** 0.776. **N08/Q6**

21 2% of biscuits on a production line are broken. Broken biscuits occur randomly. 180 biscuits are checked to see whether they are broken. Use a suitable approximation to find the probability that fewer than 4 are broken. [3]

Answer: 0.515. **N09/71/Q1**

22 A computer user finds that unwanted emails arrive randomly at a uniform average rate of 1.27 per hour.

(i) Find the probability that more than 1 unwanted email arrives in a period of 5 hours. [2]

(ii) Find the probability that more than 850 unwanted emails arrive in a period of 700 hours. [4]

Answers: **(i)** 0.987; **(ii)** 0.902. **N09/72/Q2**

23 An airline knows that some people who have bought tickets may not arrive for the flight. The airline therefore sells more tickets than the number of seats that are available. For one flight there are 210 seats available and 213 people have bought tickets. The probability of any person who has bought a ticket not arriving for the flight is $\frac{1}{50}$.

(i) By considering the number of people who do **not** arrive for the flight, use a suitable approximation to calculate the probability that more people will arrive than there are seats available. [4]

Independently, on another flight for which 135 people have bought tickets, the probability of any person not arriving is $\frac{1}{75}$.

(ii) Calculate the probability that, for both these flights, the total number of people who do not arrive is 5. [3]

Answers: (i) 0.202; (ii) 0.159. **N09/72/Q3**

24 People arrive randomly and independently at a supermarket checkout at an average rate of 2 people every 3 minutes.

(i) Find the probability that exactly 4 people arrive in a 5-minute period. [2]

At another checkout in the same supermarket, people arrive randomly and independently at an average rate of 1 person each minute.

(ii) Find the probability that a total of fewer than 3 people arrive at the two checkouts in a 3-minute period. [3]

Answers: (i) 0.184; (ii) 0.125. **N10/72/Q2**

25 A book contains 40 000 words. For each word, the probability that it is printed wrongly is 0.0001 and these errors occur independently. The number of words printed wrongly in the book is represented by the random variable X.

(i) State the exact distribution of X, including the values of any parameters. [1]

(ii) State an approximate distribution for X, including the values of any parameters, and explain why this approximate distribution is appropriate. [3]

(iii) Use this approximate distribution to find the probability that there are more than 3 words printed wrongly in the book. [3]

Answers: (i) B(40 000, 0.0001); (ii) Po(4); (iii) 0.567. **N10/72/Q3**

26 The number of goals scored per match by Everly Rovers is represented by the random variable X which has mean 1.8.

(i) State two conditions for X to be modelled by a Poisson distribution. [2]

Assume now that $X \sim \text{Po}(1.8)$.

(ii) Find $P(2 < X < 6)$. [2]

(iii) The manager promises the team a bonus if they score at least 1 goal in each of the next 10 matches. Find the probability that they win the bonus. [3]

Answers: **(i)** Two of: constant average rate of goals scored, goals occur at random, goals independent; **(ii)** 0.259; **(iii)** 0.164. **J11/72/Q3**

---

27 Bacteria of a certain type are randomly distributed in the water in two ponds, A and B. The average numbers of bacteria per cm^3 in A and B are 0.32 and 0.45 respectively.

(i) Samples of $8\,\text{cm}^3$ of water from A and $12\,\text{cm}^3$ of water from B are taken at random. Find the probability that the total number of bacteria in these samples is at least 3. [3]

(ii) Find the probability that in a random sample of $155\,\text{cm}^3$ of water from A, the number of bacteria is less than 35. [5]

Answers: **(i)** 0.986 **(ii)** 0.016(0) **J12/72/Q4**

---

28 It is known that 1.2% of rods made by a certain machine are bent. The random variable X denotes the number of bent rods in a random sample of 400 rods.

(i) State the distribution of X. [2]

(ii) State, with a reason, a suitable approximate distribution for X. [2]

(iii) Use your approximate distribution to find the probability that the sample will include more than 2 bent rods. [2]

Answer: Binomial; n=400 p=0.012 **J13/72/Q1**
Poisson n large and mean 4.8<5
0.857

---

29 **(i)** The random variable W has the distribution Po(1.5). Find the probability that the sum of 3 independent values of W is greater than 2. [3]

(ii) The random variable X has the distribution $\text{Po}(\lambda)$. Given that $P(X = 0) = 0.523$, find the value of λ correct to 3 significant figures. [2]

(iii) The random variable Y has the distribution $\text{Po}(\mu)$, where $\mu \neq 0$. Given that

$$P(Y = 3) = 24 \times P(Y = 1),$$

find μ. [3]

Answers: 0.826 J14/72/Q4
λ=0.648
μ=12

30 In a golf tournament, the number of times in a day that a 'hole-in-one' is scored is denoted by the variable X, which has a Poisson distribution with mean 0.15. Mr Crump offers to pay $200 each time that a hole-in-one is scored during 5 days of play. Find the expectation and variance of the amount that Mr Crump pays. [5]

Answer: 150 J15/72/Q3
30,000

31 In a certain lottery, 10 500 tickets have been sold altogether and each ticket has a probability of 0.0002 of winning a prize. The random variable X denotes the number of prize-winning tickets that have been sold.

(i) State, with a justification, an approximating distribution for X. [3]

(ii) Use your approximating distribution to find $P(X < 4)$. [3]

(iii) Use your approximating distribution to find the conditional probability that $X < 4$, given that $X \geqslant 1$. [4]

Answer: **(i)** Poisson n>50 np<5 J15/72/Q7
(ii) 0.839
(iii) 0.816

32 Customers arrive at an enquiry desk at a constant average rate of 1 every 5 minutes.

(i) State one condition for the number of customers arriving in a given period to be modelled by a Poisson distribution. [1]

Assume now that a Poisson distribution is a suitable model.

(ii) Find the probability that exactly 5 customers will arrive during a randomly chosen 30-minute period. [2]

(iii) Find the probability that fewer than 3 customers will arrive during a randomly chosen 12-minute period. [3]

(iv) Find an estimate of the probability that fewer than 30 customers will arrive during a randomly chosen 2-hour period. [4]

Answer: **(i)** Customers arrive independently; **(ii)** 0.161; **(iii)** 0.570; **(iv)** 0.869 N11/72/Q6

33 Each computer made in a factory contains 1000 components. On average, 1 in 30 000 of these components is defective. Use a suitable approximate distribution to find the probability that a randomly chosen computer contains at least 1 faulty component. [4]

Answer: 0.0328 N13/72/Q1

34 The number of radioactive particles emitted per 150-minute period by some material has a Poisson distribution with mean 0.7.

(i) Find the probability that at most 2 particles will be emitted during a randomly chosen 10-hour period. [3]

(ii) Find, in minutes, the longest time period for which the probability that no particles are emitted is at least 0.99. [5]

Answer: 0-469 **N13/72/Q4**
Max period is 2.15 mins

---

35 The probability that a randomly chosen plant of a certain kind has a particular defect is 0.01. A random sample of 150 plants is taken.

(i) Use an appropriate approximating distribution to find the probability that at least 1 plant has the defect. Justify your approximating distribution. [4]

The probability that a randomly chosen plant of another kind has the defect is 0.02. A random sample of 100 of these plants is taken.

(ii) Use an appropriate approximating distribution to find the probability that the total number of plants with the defect in the two samples together is more than 3 and less than 7. [3]

Answer: $n > 50$ and $np < 5$, 0.777, 0.398 **N14/72/Q2**

---

36 Failures of two computers occur at random and independently. On average the first computer fails 1.2 times per year and the second computer fails 2.3 times per year. Find the probability that the total number of failures by the two computers in a 6-month period is more than 1 and less than 4. [4]

Answer: 0.421 **N15/72/Q1**

---

37 On average, 1 in 2500 adults has a certain medical condition.

(i) Use a suitable approximation to find the probability that, in a random sample of 4000 people, more than 3 have this condition. [3]

(ii) In a random sample of n people, where n is large, the probability that none has the condition is less than 0.05. Find the smallest possible value of n. [4]

Answer: 0.0788 **N15/72/Q5**
7490

Homework: The Poisson distribution – Variant 71 & 73

1 A clinic deals only with flu vaccinations. The number of patients arriving every 15 minutes is modelled by the random variable X with distribution Po(4.2).

(i) State two assumptions required for the Poisson model to be valid. [2]

(ii) Find the probability that

(a) at least 1 patient will arrive in a 15-minute period, [2]

(b) fewer than 4 patients will arrive in a 10-minute period. [3]

(iii) The clinic is open for 20 hours each week. At the beginning of one week the clinic has enough vaccine for 370 patients. Use a suitable approximation to find the probability that this will not be enough vaccine for that week. [4]

Answers: (i) Patients arrive at a constant mean rate, patients arrive at random; (ii)(a) 0.985; (b) 0.692; (iii) 0.0300. **J10/73/Q7**

---

2 On average, 2 people in every 10 000 in the UK have a particular gene. A random sample of 6000 people in the UK is chosen. The random variable X denotes the number of people in the sample who have the gene. Use an approximating distribution to calculate the probability that there will be more than 2 people in the sample who have the gene. [4]

Answer: 0.121. **J11/71/Q1**

---

3 A hotel kitchen has two dish-washing machines. The numbers of breakdowns per year by the two machines have independent Poisson distributions with means 0.7 and 1.0. Find the probability that the total number of breakdowns by the two machines during the next two years will be less than 3. [4]

Answer: 0.340. **J11/73/Q1**

---

4 On average, 1 in 2500 people have a particular gene.

(i) Use a suitable approximation to find the probability that, in a random sample of 10 000 people, more than 3 people have this gene. [4]

(ii) The probability that, in a random sample of n people, none of them has the gene is less than 0.01. Find the smallest possible value of n. [3]

Answers: (i) 0.567; (ii) 11513 (Poisson) or 11511 (binomial). **J11/73/Q4**

---

5 A random variable X has the distribution Po(3.2).

(i) A random value of X is found.

(a) Find $P(X \geqslant 3)$. [2]

(b) Find the probability that $X = 3$ given that $X \geqslant 3$. [3]

Answers: (i)(a) 0.62(0) (b) 0.359 (ii)(a) (Approx) Normal with mean 3.2 and variance 3.2/120 (b) 0.730 **J12/71/Q5**

6 At work Jerry receives emails randomly at a constant average rate of 15 emails per hour.

(i) Find the probability that Jerry receives more than 2 emails during a 20-minute period at work. [3]

(ii) Jerry's working day is 8 hours long. Find the probability that Jerry receives fewer than 110 emails per day on each of 2 working days. [4]

(iii) At work Jerry also receives texts randomly and independently at a constant average rate of 1 text every 10 minutes. Find the probability that the total number of emails and texts that Jerry receives during a 5-minute period at work is more than 2 and less than 6. [4]

Answer: (i) 0.875 (ii) 0.0285 (iii) 0.247 **J12/71/Q7**

7 The number of lions seen per day during a standard safari has the distribution Po(0.8). The number of lions seen per day during an off-road safari has the distribution Po(2.7). The two distributions are independent.

(i) Susan goes on a standard safari for one day. Find the probability that she sees at least 2 lions. [2]

(ii) Deena goes on a standard safari for 3 days and then on an off-road safari for 2 days. Find the probability that she sees a total of fewer than 5 lions. [3]

(iii) Khaled goes on a standard safari for n days, where n is an integer. He wants to ensure that his chance of not seeing any lions is less than 10%. Find the smallest possible value of n. [3]

Answers: (i) 0.191 (ii) 0.112 (iii) n=3 **J12/73/Q4**

8 The probability that a new car of a certain type has faulty brakes is 0.008. A random sample of 520 new cars of this type is chosen, and the number, X, having faulty brakes is noted.

(i) Describe fully the distribution of X and describe also a suitable approximating distribution. Justify this approximating distribution. [4]

(ii) Use your approximating distribution to find

(a) $P(X > 3)$, [2]

(b) the smallest value of n such that $P(X = n) > P(X = n + 1)$. [3]

Answer: X~B(520,0.008) →approx Po(4.16) because n(=500) large and np(=4.16)<5 **J13/71/Q5**
0.597
Smallest n is 4

9 The independent random variables X and Y have the distributions Po(2) and Po(3) respectively.

(i) Given that $X + Y = 5$, find the probability that $X = 1$ and $Y = 4$. [4]

(ii) Given that $P(X = r) = \frac{2}{3}P(X = 0)$, show that $3 \times 2^{r-1} = r!$ and verify that $r = 4$ satisfies this equation. [2]

Answers: (i) 0.259. (ii) $3 \times 2^{r-1} = 3!$, Valid for r=4 **J13//73/Q4**

10 Calls arrive at a helpdesk randomly and at a constant average rate of 1.4 calls per hour. Calculate the probability that there will be

(i) more than 3 calls in $2\frac{1}{2}$ hours, [3]

(ii) fewer than 1000 calls in four weeks (672 hours). [4]

- *Answers:* (i) 0.463 (ii) 0.972 **J13/73/Q6**

11 The proportion of people who have a particular gene, on average, is 1 in 1000. A random sample of 3500 people in a certain country is chosen and the number of people, X, having the gene is found.

(i) State the distribution of X and state also an appropriate approximating distribution. Give the values of any parameters in each case. Justify your choice of the approximating distribution. [3]

(ii) Use the approximating distribution to find $P(X \leqslant 3)$. [2]

Answers: 0.826
λ=0.648
μ=12

J14/71/Q4

12 **(i)** The following tables show the probability distributions for the random variables V and W.

v	-1	0	1	>1
$P(V = v)$	0.368	0.368	0.184	0.080

w	0	0.5	1	>1
$P(W = w)$	0.368	0.368	0.184	0.080

For each of the variables V and W state how you can tell from its probability distribution that it does NOT have a Poisson distribution. [2]

(ii) The random variable X has the distribution $Po(\lambda)$. It is given that

$$P(X = 0) = p \quad \text{and} \quad P(X = 1) = 2.5p,$$

where p is a constant.

(a) Show that $\lambda = 2.5$. [1]

(b) Find $P(X \geqslant 3)$. [2]

(iii) The random variable Y has the distribution $Po(\mu)$, where $\mu > 30$. Using a suitable approximating distribution, it is found that $P(Y > 40) = 0.5793$ correct to 4 decimal places. Find μ. [5]

J14/71/Q8

13 On average 1 in 25 000 people have a rare blood condition. Use a suitable approximating distribution to find the probability that fewer than 2 people in a random sample of 100 000 have the condition. [3]

Answer: 0.0916 **J14/73/Q1**

14 A Lost Property office is open 7 days a week. It may be assumed that items are handed in to the office randomly, singly and independently.

(i) State another condition for the number of items handed in to have a Poisson distribution. [1]

It is now given that the number of items handed in per week has the distribution Po(4.0).

(ii) Find the probability that exactly 2 items are handed in on a particular day. [2]

(iii) Find the probability that at least 4 items are handed in during a 10-day period. [3]

(iv) Find the probability that, during a certain week, 5 items are handed in altogether, but no items are handed in on the first day of the week. [3]

Answers: **(i)** Constant mean rate **(ii)** 0.0921/2 **(iii)** 0.821 **(iv)** 0.0723 **J14/73/Q7**

15 A publishing firm has found that errors in the first draft of a new book occur at random and that, on average, there is 1 error in every 3 pages of a first draft. Find the probability that in a particular first draft there are

(i) exactly 2 errors in 10 pages, [2]

(ii) at least 3 errors in 6 pages, [3]

(iii) fewer than 50 errors in 200 pages. [4]

Answers: **(i)** 0.198 **(ii)** 0.323 **(iii)** 0.0178 **J15/71/Q6**

16 People arrive at a checkout in a store at random, and at a constant mean rate of 0.7 per minute. Find the probability that

(i) exactly 3 people arrive at the checkout during a 5-minute period, [2]

(ii) at least 30 people arrive at the checkout during a 1-hour period. [4]

People arrive independently at another checkout in the store at random, and at a constant mean rate of 0.5 per minute.

(iii) Find the probability that a total of more than 3 people arrive at this pair of checkouts during a 2-minute period. [4]

Answers: **(i)** 0.216 **(ii)** 0.973 **(iii)** 0.221 **J15/73/Q6**

17 People arrive randomly and independently at a supermarket checkout at an average rate of 2 people every 3 minutes.

(i) Find the probability that exactly 4 people arrive in a 5-minute period. [2]

At another checkout in the same supermarket, people arrive randomly and independently at an average rate of 1 person each minute.

(ii) Find the probability that a total of fewer than 3 people arrive at the two checkouts in a 3-minute period. [3]

Answers: **(i)** 0.184; **(ii)** 0.125. **N10/71/Q2**

19 A book contains 40 000 words. For each word, the probability that it is printed wrongly is 0.0001 and these errors occur independently. The number of words printed wrongly in the book is represented by the random variable X.

(i) State the exact distribution of X, including the values of any parameters. [1]

(ii) State an approximate distribution for X, including the values of any parameters, and explain why this approximate distribution is appropriate. [3]

(iii) Use this approximate distribution to find the probability that there are more than 3 words printed wrongly in the book. [3]

Answers: **(i)** B(40 000, 0.0001); **(ii)** Po(4); **(iii)** 0.567. **N10/71/Q3**

20 A random variable has the distribution Po(31). Name an appropriate approximating distribution and state the mean and standard deviation of this approximating distribution. [3]

Answers: normal; 31, √31 (5.57). **N10/73/Q1**

21 Customers arrive at an enquiry desk at a constant average rate of 1 every 5 minutes.

(i) State one condition for the number of customers arriving in a given period to be modelled by a Poisson distribution. [1]

Assume now that a Poisson distribution is a suitable model.

(ii) Find the probability that exactly 5 customers will arrive during a randomly chosen 30-minute period. [2]

(iii) Find the probability that fewer than 3 customers will arrive during a randomly chosen 12-minute period. [3]

(iv) Find an estimate of the probability that fewer than 30 customers will arrive during a randomly chosen 2-hour period. [4]

Answer: **(i)** Customers arrive independently; **(ii)** 0.161; **(iii)** 0.570; **(iv)** 0.869 **N11/71/Q6**

22 The numbers of men and women who visit a clinic each hour are independent Poisson variables with means 2.4 and 2.8 respectively.

(i) Find the probability that, in a half-hour period,

(a) 2 or more men and 1 or more women will visit the clinic, [4]

(b) a total of 3 or more people will visit the clinic. [3]

(ii) Find the probability that, in a 10-hour period, a total of more than 60 people will visit the clinic. [4]

Answer: **(i)** 0.254; **(ii)** 0.482; **(iii)** 0.119 **N11/73/Q7**

23 A random variable X has the distribution Po(1.6).

(i) The random variable R is the sum of three independent values of X. Find $P(R < 4)$. [3]

(ii) The random variable S is the sum of n independent values of X. It is given that

$$P(S = 4) = \frac{16}{3} \times P(S = 2).$$

Find n. [4]

(iii) The random variable T is the sum of 40 independent values of X. Find $P(T > 75)$. [4]

Answers: (i) 0.294; (ii) 5; (iii) 0.0753. **N12/71/Q7**

24 The number of workers, X, absent from a factory on a particular day has the distribution B(80, 0.01).

(i) Explain why it is appropriate to use a Poisson distribution as an approximating distribution for X. [2]

(ii) Use the Poisson distribution to find the probability that the number of workers absent during 12 randomly chosen days is more than 2 and less than 6. [3]

Answers: (i) $n > 50$, $np < 5$; (ii) 0.0800; (iii) There is evidence that the mean number of absent workers has decreased. **N12/73/Q7**

25 Each computer made in a factory contains 1000 components. On average, 1 in 30 000 of these components is defective. Use a suitable approximate distribution to find the probability that a randomly chosen computer contains at least 1 faulty component. [4]

Answer: 0.0328 **N13/71/Q1**

26 The number of radioactive particles emitted per 150-minute period by some material has a Poisson distribution with mean 0.7.

(i) Find the probability that at most 2 particles will be emitted during a randomly chosen 10-hour period. [3]

(ii) Find, in minutes, the longest time period for which the probability that no particles are emitted is at least 0.99. [5]

Answer: 0-469
Max period is 2.15 mins **N13/71/Q4**

27 Goals scored by Femchester United occur at random with a constant average of 1.2 goals per match. Goals scored against Femchester United occur independently and at random with a constant average of 0.9 goals per match.

(i) Find the probability that in a randomly chosen match involving Femchester,

(a) a total of 3 goals are scored, [2]

(b) a total of 3 goals are scored and Femchester wins. [3]

The manager promises the Femchester players a bonus if they score at least 35 goals in the next 25 matches.

(ii) Find the probability that the players receive the bonus. [4]

Answer: (i)(a) 0.189, (b)0.115, (ii) 0.206 **N13/73/Q4**

---

28 The probability that a randomly chosen plant of a certain kind has a particular defect is 0.01. A random sample of 150 plants is taken.

(i) Use an appropriate approximating distribution to find the probability that at least 1 plant has the defect. Justify your approximating distribution. [4]

The probability that a randomly chosen plant of another kind has the defect is 0.02. A random sample of 100 of these plants is taken.

(ii) Use an appropriate approximating distribution to find the probability that the total number of plants with the defect in the two samples together is more than 3 and less than 7. [3]

Answer: $n > 50$ and $np < 5$, 0.777, 0.398 **N14/71/Q2**

---

29 The number of calls received at a small call centre has a Poisson distribution with mean 2.4 calls per 5-minute period. Find the probability of

(i) exactly 4 calls in an 8-minute period, [2]

(ii) at least 3 calls in a 3-minute period. [3]

The number of calls received at a large call centre has a Poisson distribution with mean 41 calls per 5-minute period.

(iii) Use an approximating distribution to find the probability that the number of calls received in a 5-minute period is between 41 and 59 inclusive. [5]

Answers: (i) 0.195 , (ii) 0.176, (iii) 0.529 **N14/73/Q6**

---

30 Failures of two computers occur at random and independently. On average the first computer fails 1.2 times per year and the second computer fails 2.3 times per year. Find the probability that the total number of failures by the two computers in a 6-month period is more than 1 and less than 4. [4]

Answer: 0.421 **N15/71/Q1**

---

31 On average, 1 in 2500 adults has a certain medical condition.

(i) Use a suitable approximation to find the probability that, in a random sample of 4000 people, more than 3 have this condition. [3]

(ii) In a random sample of n people, where n is large, the probability that none has the condition is less than 0.05. Find the smallest possible value of n. [4]

Answer: 0.0788
7490 **N15/71/Q5**

32 The number of calls received per 5-minute period at a large call centre has a Poisson distribution with mean λ, where $\lambda > 30$. If more than 55 calls are received in a 5-minute period, the call centre is overloaded. It has been found that the probability of being overloaded during a randomly chosen 5-minute period is 0.01. Use the normal approximation to the Poisson distribution to obtain a quadratic equation in $\sqrt{\lambda}$ and hence find the value of λ. [5]

Answer: 40.7 **N15/73/Q2**

33 **(a)** A large number of spoons and forks made in a factory are inspected. It is found that 1% of the spoons and 1.5% of the forks are defective. A random sample of 140 items, consisting of 80 spoons and 60 forks, is chosen. Use the Poisson approximation to the binomial distribution to find the probability that the sample contains

(i) at least 1 defective spoon and at least 1 defective fork, [3]

(ii) fewer than 3 defective items. [3]

(b) The random variable X has the distribution $\text{Po}(\lambda)$. It is given that

$$P(X = 1) = p \quad \text{and} \quad P(X = 2) = 1.5p,$$

where p is a non-zero constant. Find the value of λ and hence find the value of p. [4]

Answer: 0.327 **J16/71/Q7**
0.757
3 0.149

34 1% of adults in a certain country own a yellow car.

(i) Use a suitable approximating distribution to find the probability that a random sample of 240 adults includes more than 2 who own a yellow car. [4]

(ii) Justify your approximation. [2]

J16/73/

35 On average, 1 clover plant in 10 000 has four leaves instead of three.

(i) Use an approximating distribution to calculate the probability that, in a random sample of 2000 clover plants, more than 2 will have four leaves. [3]

(ii) Justify your approximating distribution. [2]

Answers: (i) 0.00115 (ii) *n* large, *np*<5 **J17/71/Q1**

36 **(i)** A random variable X has the distribution $\text{Po}(42)$.

(a) Use an appropriate approximating distribution to find $P(X \geqslant 40)$. [4]

(b) Justify your use of the approximating distribution. [1]

(ii) A random variable Y has the distribution B(60, 0.02).

(a) Use an appropriate approximating distribution to find $P(Y > 2)$. [3]

(b) Justify your use of the approximating distribution. [1]

J17/73/15

37 The number of e-readers sold in a 10-day period in a shop is modelled by the distribution Po(5.1). Use an approximating distribution to find the probability that fewer than 140 e-readers are sold in a 300-day period. [4]

Answer: 0.138 **J18/71/Q3**

38 A random variable X has the distribution B(75, 0.03).

(i) Use the Poisson approximation to the binomial distribution to calculate $P(X < 3)$. [3]

(ii) Justify the use of the Poisson approximation. [1]

Answers: (i) 0.609 (ii) np <5, n large **J18/73/Q1**

39 The numbers, M and F, of male and female students who leave a particular school each year to study engineering have means 3.1 and 0.8 respectively.

(i) State, in context, one condition required for M to have a Poisson distribution. [1]

Assume that M and F can be modelled by independent Poisson distributions.

(ii) Find the probability that the total number of students who leave to study engineering in a particular year is more than 3. [3]

(iii) Given that the total number of students who leave to study engineering in a particular year is more than 3, find the probability that no female students leave to study engineering in that year. [3]

Answers: (i) Males leaving each year to study engineering have a constant mean (ii) 0.547 (iii) 0.308 **J18/73/Q4**

40 Particles are emitted randomly from a radioactive substance at a constant average rate of 3.6 per minute. Find the probability that

(i) more than 3 particles are emitted during a 20-second period, [3]

(ii) more than 240 particles are emitted during a 1-hour period. [4]

Answer: 0.0338
0.0478 **N16/71/Q3**

41 The random variable X has the distribution Po(3.5). Find $P(X < 3)$. [3]

Answer: 0.321 **N16/73/Q1**

42 A random variable, X, has the distribution Po(31). Use the normal approximation to the Poisson distribution to find $P(X > 40)$. [3]

Answer: 0.0441 or 0.0440 **N17/71/Q1**

43 An airline has found that, on average, 1 in 100 passengers do not arrive for each flight, and that this occurs randomly. For one particular flight the airline always sells 403 seats. The plane only has room for 400 passengers, so the flight is overbooked if the number of passengers who do not arrive is less than 3. Use a suitable approximation to find the probability that the flight is overbooked. [4]

Answer: 0.234 **N17/71/Q2**

44 Small drops of two liquids, A and B, are randomly and independently distributed in the air. The average numbers of drops of A and B per cubic centimetre of air are 0.25 and 0.36 respectively.

(i) A sample of $10\,\text{cm}^3$ of air is taken at random. Find the probability that the total number of drops of A and B in this sample is at least 4. [3]

5

(ii) A sample of $100\,\text{cm}^3$ of air is taken at random. Use an approximating distribution to find the probability that the total number of drops of A and B in this sample is less than 60. [5]

Answers: (i) 0.857 (ii) 0.424 **N18/71/Q4**

Chapter 2: Linear Combination of Variables

THE EXPECTATION AND VARIANCE OF A LINEAR FUNCTION OF A RANDOM VARIABLE

If $Y = X + b$, where b is a constant, then

$$E(Y) = E(X+b) = E(X) + b \qquad (2.1)$$

and

$$\text{Var}(Y) = \text{Var}(X+b) = \text{Var}(X). \qquad (2.2)$$

If $W = aX$, where a is a constant, then

$$E(W) = E(aX) = aE(X) \qquad (2.3)$$

and

$$\text{Var}(W) = \text{Var}(aX) = a^2\text{Var}(X). \qquad (2.4)$$

For any random variable X,

$$E(aX+b) = aE(X) + b, \qquad (2.5)$$

$$\text{Var}(aX+b) = a^2\text{Var}(X), \qquad (2.6)$$

where a and b are constants.

The temperature in degrees Fahrenheit on a remote island is a random variable with mean 59 and variance 27. Find the mean and variance of the temperature in degrees Centigrade, given that to convert degrees Fahrenheit to degrees Centigrade you subtract 32 and then multiply by $\frac{5}{9}$.

THE SUM OF INDEPENDENT NORMAL VARIABLES

If X and Y are any two random variables, discrete or continuous, and a and b are any two constants,

Sums

$E(X+Y)=E(X)+E(Y$ ... ①

$E(aX+bY)=aE(X)+bE(Y)$ ... ③

Also, if X and Y are independent, then

$\text{Var}(X+Y)=\text{Var}(X)+\text{Var}(Y)$ ... ⑤

$\text{Var}(aX+bY)=a^2\,\text{Var}(X)+b^2\,\text{Var}(Y)$ ... ⑦

Example 1

A coffee machine is installed in a students' common room. It dispenses white coffee by first releasing a quantity of black coffee, normally distributed with mean 122.5 ml and standard deviation 7.5 ml, and then adding a quantity of milk, normally distributed with mean 30 ml and standard deviation 5 ml.

Each cup is marked to a level of 137.5 ml and if this level is not attained the customer receives the drink free of charge.

What percentage of cups of white coffee will be given free of charge?

Example 2

Four runners, Andy, Bob, Chris and Dai, train to take part in a 1600 m relay race in which Andy is to run 100 m, Bob 200 m, Chris 500 m and Dai 800 m.
During training their individual times, recorded in seconds, follow normal distributions.
With obvious notation, these are:

$$A \sim N(10.8, 0.2^2),\ B \sim N(23.7, 0.3^2),\ C \sim N(62.8, 0.9^2) \text{ and } D \sim N(121.2, 2.1^2).$$

Find the probability that they run the relay race in less than 3 minutes 35 seconds.

MORE THAN ONE OBSERVATION OF A RANDOM VARIABLE

Consider now the special case when $X_1, X_2, \ldots, X_n$ are n independent observations from the *same* normal distribution

so $X_1 \sim N(\mu, \sigma^2)$, $\quad X_2 \sim N(\mu, \sigma^2)$, $\quad \ldots,\quad X_n \sim N(\mu, \sigma^2)$

then
$$\begin{aligned} E(X_1 + X_2 + \cdots + X_n) &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= \mu + \mu + \cdots + \mu \\ &= n\mu \end{aligned}$$

$$\begin{aligned} \text{Var}(X_1 + X_2 + \cdots + X_n) &= \text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_n) \\ &= \sigma^2 + \sigma^2 + \cdots + \sigma^2 \\ &= n\sigma^2 \end{aligned}$$

So $\quad X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)$

Example 3

Masses of a particular article are normally distributed with mean 20 g and standard deviation 2 g. A random sample of 12 such articles is chosen. Find the probability that the total mass is greater than 230 g.

Example 4

The maximum load a lift can carry is 450 kg. The weights of men are normally distributed with mean 60 kg and standard deviation 10 kg. The weights of women are normally distributed with mean 55 kg and standard deviation 5 kg. Find the probability that the lift will be overloaded by five men and two women, if their weights are independent. (*L*)

THE DIFFERENCE OF INDEPENDENT NORMAL VARIABLES

For two independent variables X and Y, where $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$

$E(X - Y) = E(X) - E(Y) = \mu_1 - \mu_2$ Result 2, page 403

$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) = \sigma_1^2 + \sigma_2^2$ Result 6, page 403

$X - Y$ is ***normally distributed***, so

$$X - Y \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

↑ (Remember the + sign here.)

Example 5

A machine produces rubber balls whose diameters are normally distributed with mean 5.50 cm and standard deviation 0.08 cm.
The balls are packed in cylindrical tubes whose internal diameters are normally distributed with mean 5.70 cm and standard deviation 0.12 cm.
If a ball, selected at random, is placed in a tube, selected at random, what is the distribution of the clearance? (The clearance is the internal diameter of the tube minus the diameter of the ball.)
What is the probability that the clearance is between 0.05 cm and 0.25 cm?

Example 6

A certain liquid drug is marketed in bottles containing a nominal 20 ml of drug. Tests on a large number of bottles indicate that the volume of liquid in each bottle is distributed normally with mean 20.42 ml and standard deviation 0.429 ml.
If the capacity of the bottles is normally distributed with mean 21.77 ml and standard deviation 0.210 ml, estimate what percentage of bottles will overflow during filling.

Example 7

In a cafeteria, baked beans are served either in ordinary portions or in children's portions. The quantity given for an ordinary portion is a normal variable with mean 90 g and standard deviation 3 g and the quantity given for a child's portion is a normal variable with mean 43 g and standard deviation 2 g.
What is the probability that Tom, who has two children's portions, is given more than his father, who has an ordinary portion?

MULTIPLES OF INDEPENDENT NORMAL VARIABLES

Remember that, for any constant a,

$E(aX) = aE(X)$ (page 246) and $\text{Var}(aX) = a^2\,\text{Var}(X)$ (page 250)

If X is a *normal* variable such that $X \sim N(\mu, \sigma^2)$

then $E(aX) = aE(X) = a\mu$
$\text{Var}(aX) = a^2\,\text{Var}(X) = a^2\sigma^2$

It can be shown that aX is also normally distributed

so $aX \sim N(a\mu, a^2\sigma^2)$

Now consider two independent normal variables X and Y where $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$

For any constants a, b, using the results on page 403

$E(aX + bY) = aE(X) + bE(Y) = a\mu_1 + b\mu_2$ ← Result ⑦
$E(aX - bY) = aE(X) - bE(Y) = a\mu_1 - b\mu_2$ ← Result ⑧

$\text{Var}(aX + bY) = a^2\,\text{Var}(X) + b^2\,\text{Var}(Y) = a^2\sigma_1^2 + b^2\sigma_2^2$ ← Result ⑦
$\text{Var}(aX - bY) = a^2\,\text{Var}(X) + b^2\,\text{Var}(Y) = a^2\sigma_1^2 + b^2\sigma_2^2$ ← Result ⑧

(Remember the + sign here.)

$aX + bY$ and $aX - bY$ are also normally distributed, so

$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$
$aX - bY \sim N(a\mu_1 - b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

Example 8

X and Y are independent random variables and $X \sim N(100, 8)$, $Y \sim N(55, 10)$. Find the probability that an observation from the population of X is more than twice the value of an observation from the population of Y.

Great care must be taken in distinguishing between a sum of random variables and a multiple of a random variable.
For example, if X is the weight of a small loaf, then the sum $X_1 + X_2 + X_3$ is the total weight of three loaves.

If $X \sim N(\mu, \sigma^2)$ then $X_1 + X_2 + X_3 \sim N(3\mu, 3\sigma^2)$.

But if there is a large economy-size loaf which is *three times the weight of a small loaf*, then the weight of an economy loaf is $3X$ (a multiple)

and $3X \sim N(3\mu, 9\sigma^2)$.

Example 9

A soft drinks manufacturer sells bottles of drinks in two sizes. The amount in each bottle, in

	Mean (ml)	Variance (ml^2)
Small	252	4
Large	1012	25

millilitres, is normally distributed as shown in the table:

(a) A bottle of each size is selected at random. Find the probability that the large bottle contains less than four times the amount in the small bottle.

(b) One large and four small bottles are selected at random. Find the probability that the amount in the large bottle is less than the total amount in the four small bottles.

Example 10

A mathematics module is assessed by an examination and by coursework. The examination makes up 75% of the total assessment and the coursework makes up 25%. Examination marks, X, are distributed with mean 53.2 and standard deviation 9.3. Coursework marks, Y, are distributed with mean 78.0 and standard deviation 5.1. Examination marks and coursework marks are independent. Find the mean and standard deviation of the combined mark $0.75X + 0.25Y$.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q2 June 2006]

Example 11

The cost of electricity for a month in a certain town under scheme A consists of a fixed charge of 600 cents together with a charge of 5.52 cents per unit of electricity used. Stella uses scheme A. The number of units she uses in a month is normally distributed with mean 500 and variance 50.41.

(i) Find the mean and variance of the total cost of Stella's electricity in a randomly chosen month.

Under scheme B there is no fixed charge and the cost in cents for a month is normally distributed with mean 6600 and variance 421. Derek uses scheme B.

(ii) Find the probability that, in a randomly chosen month, Derek spends more than twice as much as Stella spends.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q4 November 2007]

Example 12

The random variable X has the distribution $N(3.2, 1.2^2)$. The sum of 60 independent observations of X is denoted by S. Find $P(S > 200)$.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q2 June 2007]

Example 13

Weights of garden tables are normally distributed with mean 36 kg and standard deviation 1.6 kg. Weights of garden chairs are normally distributed with mean 7.3 kg and standard deviation 0.4 kg. Find the probability that the total weight of 2 randomly chosen tables is more than the total weight of 10 randomly chosen chairs.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q3 November 2008]

Example 14

A journey in a certain car consists of two stages with a stop for filling up with fuel after the first stage. The length of time, T minutes, taken for each stage has a normal distribution with mean 74 and standard deviation 7.3. The length of time, F minutes, it takes to fill up with fuel has a normal distribution with mean 5 and standard deviation 1.7. The length of time it takes to pay for the fuel is exactly 4 minutes. The variables T and F are independent and the times for the two stages are independent of each other.

(i) Find the probability that the total time for the journey is less than 154 minutes.

(ii) A second car has a fuel tank with exactly twice the capacity of the first car. Find the mean and variance of this car's fuel fill-up time.

(iii) This second car's time for each stage of the journey follows a normal distribution with mean 69 minutes and standard deviation 5.2 minutes. The length of time it takes to pay for the fuel for this car is also exactly 4 minutes. Find the probability that the total time for the journey taken by the first car is more than the total time taken by the second car.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q7 November 2005]

SELECTED PAST PAPER QUESTION

Question 1
N13/72/Q7

Kieran and Andreas are long-jumpers. They model the lengths, in metres, that they jump by the independent random variables $K \sim N(5.64, 0.0576)$ and $A \sim N(4.97, 0.0441)$ respectively. They each make a jump and measure the length. Find the probability that

(i) the sum of the lengths of their jumps is less than 11 m, [4]

(ii) Kieran jumps more than 1.2 times as far as Andreas. [6]

Question 2
N11/73/Q1

Test scores, X, have mean 54 and variance 144. The scores are scaled using the formula $Y = a + bX$, where a and b are constants and $b > 0$. The scaled scores, Y, have mean 50 and variance 100. Find the values of a and b. [4]

Question 3
N11/72/Q1

The random variable X has the distribution $Po(1.3)$. The random variable Y is defined by $Y = 2X$.

(i) Find the mean and variance of Y. [3]

(ii) Give a reason why the variable Y does not have a Poisson distribution. [1]

Homework: Linear Combination of variables– Variant 72

1 A fair coin is tossed 5 times and the number of heads is recorded.

(i) The random variable X is the number of heads. State the mean and variance of X. [2]

(ii) The number of heads is doubled and denoted by the random variable Y. State the mean and variance of Y. [2]

Answers: **(i)** 2.5, 1.25; **(ii)** 5, 5. **J03/Q1**

2 Machine A fills bags of fertiliser so that their weights follow a normal distribution with mean 20.05 kg and standard deviation 0.15 kg. Machine B fills bags of fertiliser so that their weights follow a normal distribution with mean 20.05 kg and standard deviation 0.27 kg.

(i) Find the probability that the total weight of a random sample of 20 bags filled by machine A is at least 2 kg more than the total weight of a random sample of 20 bags filled by machine B. [6]

Answers: **(i)** 0.0738; **(ii)** 216. **J03/Q7**

3 In athletics matches the triple jump event consists of a hop, followed by a step, followed by a jump. The lengths covered by Albert in each part are independent normal variables with means 3.5 m, 2.9 m, 3.1 m and standard deviations 0.3 m, 0.25 m, 0.35 m respectively. The length of the triple jump is the sum of the three parts.

(i) Find the mean and standard deviation of the length of Albert's triple jumps. [3]

(ii) Find the probability that the mean of Albert's next four triple jumps is greater than 9 m. [3]

Answers: **(i)** Mean = 9.5, Standard deviation = 0.524; **(ii)** 0.972. **J04/Q2**

4 The independent random variables X and Y are such that X has mean 8 and variance 4.8 and Y has a Poisson distribution with mean 6. Find

(i) $\mathrm{E}(2X - 3Y)$, [2]

(ii) $\mathrm{Var}(2X - 3Y)$. [4]

Answers: **(i)** –2; **(ii)** 73.2. **J04/Q3**

5 Exam marks, X, have mean 70 and standard deviation 8.7. The marks need to be scaled using the formula $Y = aX + b$ so that the scaled marks, Y, have mean 55 and standard deviation 6.96. Find the values of a and b. [4]

Answers: $a = 0.8$, $b = -1$. **J05/Q1**

6 A mathematics module is assessed by an examination and by coursework. The examination makes up 75% of the total assessment and the coursework makes up 25%. Examination marks, X, are distributed with mean 53.2 and standard deviation 9.3. Coursework marks, Y, are distributed with mean 78.0 and standard deviation 5.1. Examination marks and coursework marks are independent. Find the mean and standard deviation of the combined mark $0.75X + 0.25Y$. [4]

Answers: 59.4, 7.09. **J06/Q2**

7 A certain make of washing machine has a wash-time with mean 56.9 minutes and standard deviation 4.8 minutes. A certain make of tumble dryer has a drying-time with mean 61.1 minutes and standard deviation 6.3 minutes. Both times are normally distributed and are independent of each other. Find the probability that a randomly chosen wash-time differs by more than 3 minutes from a randomly chosen drying-time. [6]

Answer: 0.742. **J06/Q4**

8 The lengths of red pencils are normally distributed with mean 6.5 cm and standard deviation 0.23 cm.

(i) Two red pencils are chosen at random. Find the probability that their total length is greater than 12.5 cm. [3]

The lengths of black pencils are normally distributed with mean 11.3 cm and standard deviation 0.46 cm.

(ii) Find the probability that the total length of 3 red pencils is more than 6.7 cm greater than the length of 1 black pencil. [4]

Answers: **(i)** 0.938; **(ii)** 0.993. **J08/Q3**

9 When Sunil travels from his home in England to visit his relatives in India, his journey is in four stages. The times, in hours, for the stages have independent normal distributions as follows.

Bus from home to the airport:	N(3.75, 1.45)
Waiting in the airport:	N(3.1, 0.785)
Flight from England to India:	N(11, 1.3)
Car in India to relatives:	N(3.2, 0.81)

(i) Find the probability that the flight time is shorter than the total time for the other three stages. [6]

(ii) Find the probability that, for 6 journeys to India, the mean time waiting in the airport is less than 4 hours. [3]

Answer: (i) 0.324
(ii) 0.994 **J09/71/Q6**

10 The weekly distance in kilometres driven by Mr Parry has a normal distribution with mean 512 and standard deviation 62. Independently, the weekly distance in kilometres driven by Mrs Parry has a normal distribution with mean 89 and standard deviation 7.4.

(i) Find the probability that, in a randomly chosen week, Mr Parry drives more than 5 times as far as Mrs Parry. [5]

(ii) Find the mean and standard deviation of the total of the weekly distances in miles driven by Mr Parry and Mrs Parry. Use the approximation 8 kilometres = 5 miles. [3]

Answers: **(i)** 0.823; **(ii)** 376, 39.0. **J10/72/Q4**

11 Bottles of wine are stacked in racks of 12. The weights of these bottles are normally distributed with mean 1.3 kg and standard deviation 0.06 kg. The weights of the empty racks are normally distributed with mean 2 kg and standard deviation 0.3 kg.

(i) Find the probability that the total weight of a full rack of 12 bottles of wine is between 17 kg and 18 kg. [5]

(ii) Two bottles of wine are chosen at random. Find the probability that they differ in weight by more than 0.05 kg. [5]

Answers: (i) 0.813; (ii) 0.556 **N02/Q7**

12 Tien throws a ball. The distance it travels can be modelled by a normal distribution with mean 20 m and variance 9 m^2. His younger sister Su Chen also throws a ball and the distance her ball travels can be modelled by a normal distribution with mean 14 m and variance 12 m^2. Su Chen is allowed to add 5 metres on to her distance and call it her 'upgraded distance'. Find the probability that Tien's distance is larger than Su Chen's upgraded distance. [5]

Answer: 0.586. **N03/Q3**

13 The weights of men follow a normal distribution with mean 71 kg and standard deviation 7 kg. The weights of women follow a normal distribution with mean 57 kg and standard deviation 5 kg. The total weight of 5 men and 2 women chosen randomly is denoted by X kg.

(i) Show that $E(X) = 469$ and $Var(X) = 295$. [2]

(ii) The total weight of 4 men and 3 women chosen randomly is denoted by Y kg. Find the mean and standard deviation of $X - Y$ and hence find $P(X - Y > 22)$. [5]

Answers: (ii) 14, 23.8, 0.369. **N04/Q4**

14 A journey in a certain car consists of two stages with a stop for filling up with fuel after the first stage. The length of time, T minutes, taken for each stage has a normal distribution with mean 74 and standard deviation 7.3. The length of time, F minutes, it takes to fill up with fuel has a normal distribution with mean 5 and standard deviation 1.7. The length of time it takes to pay for the fuel is exactly 4 minutes. The variables T and F are independent and the times for the two stages are independent of each other.

(i) Find the probability that the total time for the journey is less than 154 minutes. [5]

(ii) A second car has a fuel tank with exactly twice the capacity of the first car. Find the mean and variance of this car's fuel fill-up time. [2]

(iii) This second car's time for each stage of the journey follows a normal distribution with mean 69 minutes and standard deviation 5.2 minutes. The length of time it takes to pay for the fuel for this car is also exactly 4 minutes. Find the probability that the total time for the journey taken by the first car is more than the total time taken by the second car. [5]

Answers: (i) 0.387; (ii) mean = 10, variance = 11.56; (iii) 0.647. **N05/Q7**

15 Climbing ropes produced by a manufacturer have breaking strengths which are normally distributed with mean 160 kg and standard deviation 11.3 kg. A group of climbers have weights which are normally distributed with mean 66.3 kg and standard deviation 7.1 kg.

(i) Find the probability that a rope chosen randomly will break under the combined weight of 2 climbers chosen randomly. [5]

Each climber carries, in a rucksack, equipment amounting to half his own weight.

(ii) Find the mean and variance of the combined weight of a climber and his rucksack. [3]

(iii) Find the probability that the combined weight of a climber and his rucksack is greater than 87 kg. [2]

Answers: **(i)** Ploughing has increased the number of metal pieces found; **(ii)** No significant increase at the 5% level; **(iii)** 0.395. **N06/Q6**

16 The cost of electricity for a month in a certain town under scheme A consists of a fixed charge of 600 cents together with a charge of 5.52 cents per unit of electricity used. Stella uses scheme A. The number of units she uses in a month is normally distributed with mean 500 and variance 50.41.

(i) Find the mean and variance of the total cost of Stella's electricity in a randomly chosen month. [5]

Under scheme B there is no fixed charge and the cost in cents for a month is normally distributed with mean 6600 and variance 421. Derek uses scheme B.

(ii) Find the probability that, in a randomly chosen month, Derek spends more than twice as much as Stella spends. [5]

Answers: **(i)** 3360, 1540; **(ii)** 0.0693. **N07/Q4**

17 Weights of garden tables are normally distributed with mean 36 kg and standard deviation 1.6 kg. Weights of garden chairs are normally distributed with mean 7.3 kg and standard deviation 0.4 kg. Find the probability that the total weight of 2 randomly chosen tables is more than the total weight of 10 randomly chosen chairs. [5]

Answer: 0.350. **N08/Q3**

18 The weights of pebbles on a beach are normally distributed with mean 48.5 grams and standard deviation 12.4 grams.

(i) Find the probability that the mean weight of a random sample of 5 pebbles is greater than 51 grams. [3]

(ii) The probability that the mean weight of a random sample of n pebbles is less than 51.6 grams is 0.9332. Find the value of n. [4]

Answers: **(i)** 0.326; **(ii)** 36. **N09/71/Q3**

19 The volume of liquid in cans of cola is normally distributed with mean 330 millilitres and standard deviation 5.2 millilitres. The volume of liquid in bottles of tonic water is normally distributed with mean 500 millilitres and standard deviation 7.1 millilitres.

(i) Find the probability that 3 randomly chosen cans of cola contain less liquid than 2 randomly chosen bottles of tonic water. [5]

(ii) A new drink is made by mixing the contents of 2 cans of cola with half a bottle of tonic water. Find the probability that the volume of the new drink is more than 900 millilitres. [4]

Answers: **(i)** 0.771; **(ii)** 0.890. **N09/71/Q7**

20 **(a)** Random variables Y and X are related by $Y = a + bX$, where a and b are constants and $b > 0$. The standard deviation of Y is twice the standard deviation of X. The mean of Y is 7.92 and is 0.8 more than the mean of X. Find the values of a and b. [3]

(b) Random variables R and S are such that $R \sim N(\mu, 2^2)$ and $S \sim N(2\mu, 3^2)$. It is given that $P(R + S > 1) = 0.9$.

(i) Find μ. [4]

(ii) Hence find $P(S > R)$. [3]

Answers: **(a)** −6.32, 2; **(b)(i)** 1.87, **(ii)** 0.699. **N09/72/Q7**

21 The marks of candidates in Mathematics and English in 2009 were represented by the independent random variables X and Y with distributions $N(28, 5.6^2)$ and $N(52, 12.4^2)$ respectively. Each candidate's marks were combined to give a final mark F, where $F = X + \frac{1}{2}Y$.

(i) Find $E(F)$ and $Var(F)$. [3]

(ii) The final marks of a random sample of 10 candidates from Grinford in 2009 had a mean of 49. Test at the 5% significance level whether this result suggests that the mean final mark of all candidates from Grinford in 2009 was lower than elsewhere. [5]

Answers: **(i)** 54, 69.8. **N10/72/Q5**

22 The weights of bags of fuel have mean 3.2 kg and standard deviation 0.04 kg. The total weight of a random sample of three bags is denoted by T kg. Find the mean and standard deviation of T. [4]

Answers: 9.6 kg; 0.0693 kg. **J11/72/Q1**

23 Fiona and Jhoti each take one shower per day. The times, in minutes, taken by Fiona and Jhoti to take a shower are represented by the independent variables $F \sim N(12.2, 2.8^2)$ and $J \sim N(11.8, 2.6^2)$ respectively. Find the probability that, on a randomly chosen day,

(i) the total time taken to shower by Fiona and Jhoti is less than 30 minutes, [4]

(ii) Fiona takes at least twice as long as Jhoti to take a shower. [4]

Answers: **(i)** 0.942 **(ii)** 0.0268 **J12/72/Q5**

24 Packets of cereal are packed in boxes, each containing 6 packets. The masses of the packets are normally distributed with mean 510 g and standard deviation 12 g. The masses of the empty boxes are normally distributed with mean 70 g and standard deviation 4 g.

(i) Find the probability that the total mass of a full box containing 6 packets is between 3050 g and 3150 g. [5]

(ii) A packet and an empty box are chosen at random. Find the probability that the mass of the packet is at least 8 times the mass of the empty box. [5]

Answer: 0.746
0.0717 **J13/72/Q5**

25 Each day Samuel travels from A to B and from B to C. He then returns directly from C to A. The times, in minutes, for these three journeys have the independent distributions $N(20, 2^2)$, $N(18, 1.5^2)$ and $N(30, 1.8^2)$, respectively. Find the probability that, on a randomly chosen day, the total time for his two journeys from A to B and B to C is less than the time for his return journey from C to A. [5]

Answer: 0.0047 **J14/72/Q2**

26 The independent random variables X and Y have standard deviations 3 and 6 respectively. Calculate the standard deviation of $4X - 5Y$. [3]

Answer: 32.3 **J15/72/Q1**

27 The random variable X has the distribution $Po(1.3)$. The random variable Y is defined by $Y = 2X$.

(i) Find the mean and variance of Y. [3]

(ii) Give a reason why the variable Y does not have a Poisson distribution. [1]

Answer: **(i)** Mean=2.6, Variance=5.2; **(ii)** Variance≠Mean **N11/72/Q1**

28 Three coats of paint are sprayed onto a surface. The thicknesses, in millimetres, of the three coats have independent distributions $N(0.13, 0.02^2)$, $N(0.14, 0.03^2)$ and $N(0.10, 0.01^2)$. Find the probability that, at a randomly chosen place on the surface, the total thickness of the three coats of paint is less than 0.30 millimetres. [5]

Answer: 0.0306 or 0.0307 **N11/72/Q3**

29 The cost of hiring a bicycle consists of a fixed charge of 500 cents together with a charge of 3 cents per minute. The number of minutes for which people hire a bicycle has mean 142 and standard deviation 35.

(i) Find the mean and standard deviation of the amount people pay when hiring a bicycle. [3]

(ii) 6 people hire bicycles independently. Find the mean and standard deviation of the total amount paid by all 6 people. [3]

Answers: **(i)** 926, 105; **(ii)** 5556, 257. **N12/72/Q3**

30 A random variable X has the distribution Po(1.6).

(i) The random variable R is the sum of three independent values of X. Find $P(R < 4)$. [3]

(ii) The random variable S is the sum of n independent values of X. It is given that

$$P(S = 4) = \frac{16}{3} \times P(S = 2).$$

Find n. [4]

(iii) The random variable T is the sum of 40 independent values of X. Find $P(T > 75)$. [4]

Answers: **(i)** 0.294; **(ii)** 5; **(iii)** 0.0753. **N12/72/Q7**

31 Kieran and Andreas are long-jumpers. They model the lengths, in metres, that they jump by the independent random variables $K \sim N(5.64, 0.0576)$ and $A \sim N(4.97, 0.0441)$ respectively. They each make a jump and measure the length. Find the probability that

(i) the sum of the lengths of their jumps is less than 11 m, [4]

(ii) Kieran jumps more than 1.2 times as far as Andreas. [6]

Answer: 0.889
0.176 **N13/72/Q7**

32 The masses, in grams, of potatoes of types A and B have the distributions $N(175, 60^2)$ and $N(105, 28^2)$ respectively. Find the probability that a randomly chosen potato of type A has a mass that is at least twice the mass of a randomly chosen potato of type B. [5]

Answer: 0.335 **N14/72/Q1**

33 The weights, in kilograms, of men and women have the distributions $N(78, 7^2)$ and $N(66, 5^2)$ respectively.

(i) The maximum load that a certain cable car can carry safely is 1200 kg. If 9 randomly chosen men and 7 randomly chosen women enter the cable car, find the probability that the cable car can operate safely. [5]

(ii) Find the probability that a randomly chosen woman weighs more than a randomly chosen man. [4]

Answer: 0.927
0.0815 **N15/72/Q6**

Homework: Linear Combination of variables– Variant 71 & 73

1 Yu Ming travels to work and returns home once each day. The times, in minutes, that he takes to travel to work and to return home are represented by the independent random variables W and H with distributions $N(22.4, 4.8^2)$ and $N(20.3, 5.2^2)$ respectively.

(i) Find the probability that Yu Ming's total travelling time during a 5-day period is greater than 180 minutes. [4]

(ii) Find the probability that, on a particular day, Yu Ming takes longer to return home than he takes to travel to work. [5]

Answers: **(i)** 0.983; **(ii)** 0.383. **J10/73/Q6**

2 Cans of drink are packed in boxes, each containing 4 cans. The weights of these cans are normally distributed with mean 510 g and standard deviation 14 g. The weights of the boxes, when empty, are independently normally distributed with mean 200 g and standard deviation 8 g.

(i) Find the probability that the total weight of a full box of cans is between 2200 g and 2300 g. [6]

(ii) Two cans of drink are chosen at random. Find the probability that they differ in weight by more than 20 g. [5]

Answers: (i) 0.896; (ii) 0.312. **J11/71/Q5**

3 Each drink from a coffee machine contains X cm^3 of coffee and Y cm^3 of milk, where X and Y are independent variables with $X \sim N(184, 15^2)$ and $Y \sim N(50, 8^2)$. If the total volume of the drink is less than 200 cm^3 the customer receives the drink without charge.

(i) Find the percentage of drinks which customers receive without charge. [4]

(ii) Find the probability that, in a randomly chosen drink, the volume of coffee is more than 4 times the volume of milk. [5]

Answers: (i) 2.28%; (ii) 0.325. **J11/73/Q5**

4 An examination consists of a written paper and a practical test. The written paper marks (M) have mean 54.8 and standard deviation 16.0. The practical test marks (P) are independent of the written paper marks and have mean 82.4 and standard deviation 4.8. The final mark is found by adding 75% of M to 25% of P. Find the mean and standard deviation of the final marks for the examination. [3]

Answers: Mean = 61.7, Standard Deviation = 12.1 **J12/71/Q2**

5 The independent random variables X and Y have the distributions $N(6.5, 14)$ and $N(7.4, 15)$ respectively. Find $P(3X - Y < 20)$. [5]

Answer: 0.747 **J12/73/Q2**

6 Weights of cups have a normal distribution with mean 91 g and standard deviation 3.2 g. Weights of saucers have an independent normal distribution with mean 72 g and standard deviation 2.6 g. Cups and saucers are chosen at random to be packed in boxes, with 6 cups and 6 saucers in each box. Given that each empty box weighs 550 g, find the probability that the total weight of a box containing 6 cups and 6 saucers exceeds 1550 g. [5]

Answer: 0.0147 **J13/71/Q3**

7 The mean and variance of the random variable X are 5.8 and 3.1 respectively. The random variable S is the sum of three independent values of X. The independent random variable T is defined by $T = 3X + 2$.

(i) Find the variance of S. [1]

(ii) Find the variance of T. [1]

(iii) Find the mean and variance of $S - T$. [3]

Answers: (i) 9.3 (ii) 27.9 (iii) mean = -2; variance =37.2 **J13/73/Q1**

8 The masses, in grams, of apples of a certain type are normally distributed with mean 60.4 and standard deviation 8.2. The apples are packed in bags, with each bag containing 8 randomly chosen apples. The bags are checked by Quality Control and any bag containing apples with a total mass of less than 436 g is rejected. Find the proportion of bags that are rejected. [4]

Answers: 250.75 38.5 **J14/71/Q1**

9 In an examination, the marks in the theory paper and the marks in the practical paper are denoted by the random variables X and Y respectively, where $X \sim N(57, 13)$ and $Y \sim N(28, 5)$. You may assume that each candidate's marks in the two papers are independent. The final score of each candidate is found by calculating $X + 2.5Y$. A candidate is chosen at random. Without using a continuity correction, find the probability that this candidate

(i) has a final score that is greater than 140, [5]

(ii) obtains at least 20 more marks in the theory paper than in the practical paper. [5]

Answers: (i) 0.0254 (ii) 0.983 **J14/73/Q8**

10 The daily times, in minutes, that Yu Ming takes showering, getting dressed and having breakfast are independent and have the distributions $N(9, 2.2^2)$, $N(8, 1.3^2)$ and $N(17, 2.6^2)$ respectively. The total daily time that Yu Ming takes for all three activities is denoted by T minutes.

(i) Find the mean and variance of T. [2]

(ii) Yu Ming notes the value of T on each day in a random sample of 70 days and calculates the sample mean. Find the probability that the sample mean is between 33 and 35. [4]

Answers: (i) 34, 13.29 (ii) 0.978 **J15/71/Q3**

11 The independent variables X and Y are such that $X \sim B(10, 0.8)$ and $Y \sim Po(3)$. Find

(i) $E(7X + 5Y - 2)$, [2]

(ii) $Var(4X - 3Y + 3)$, [4]

(iii) $P(2X - Y = 18)$. [4]

Answers: (i) 69 (ii) 52.6 (iii) 0.374 or 0.375 **J15/71/Q7**

12 The masses, in milligrams, of three minerals found in 1 tonne of a certain kind of rock are modelled by three independent random variables P, Q and R, where $P \sim N(46, 19^2)$, $Q \sim N(53, 23^2)$ and $R \sim N(25, 10^2)$. The total value of the minerals found in 1 tonne of rock is modelled by the random variable V, where $V = P + Q + 2R$. Use the model to find the probability of finding minerals with a value of at least 93 in a randomly chosen tonne of rock. [7]

Answer. 0.941. **N10/73/Q4**

13 The random variable X has the distribution Po(1.3). The random variable Y is defined by $Y = 2X$.

(i) Find the mean and variance of Y. [3]

(ii) Give a reason why the variable Y does not have a Poisson distribution. [1]

Answer. **(i)** Mean=2.6, Variance=5.2; **(ii)** Variance≠Mean **N11/71/Q1**

14 Three coats of paint are sprayed onto a surface. The thicknesses, in millimetres, of the three coats have independent distributions $N(0.13, 0.02^2)$, $N(0.14, 0.03^2)$ and $N(0.10, 0.01^2)$. Find the probability that, at a randomly chosen place on the surface, the total thickness of the three coats of paint is less than 0.30 millimetres. [5]

Answer. 0.0306 or 0.0307 **N11/71/Q3**

15 Test scores, X, have mean 54 and variance 144. The scores are scaled using the formula $Y = a + bX$, where a and b are constants and $b > 0$. The scaled scores, Y, have mean 50 and variance 100. Find the values of a and b. [4]

Answers. $b = 5/6$, $a = 5$ **N11/73/Q1**

16 Ranjit goes to mathematics lectures and physics lectures. The length, in minutes, of a mathematics lecture is modelled by the variable X with distribution $N(36, 3.5^2)$. The length, in minutes, of a physics lecture is modelled by the independent variable Y with distribution $N(55, 5.2^2)$.

(i) Find the probability that the total length of two mathematics lectures and one physics lecture is less than 140 minutes. [4]

(ii) Ranjit calculates how long he will need to spend revising the content of each lecture as follows. Each minute of a mathematics lecture requires 1 minute of revision and each minute of a physics lecture requires $1\frac{1}{2}$ minutes of revision. Find the probability that the total revision time required for one mathematics lecture and one physics lecture is more than 100 minutes. [4]

Answer. **(i)** 0.965; **(ii)** 0.985 **N11/73/Q6**

17 The cost of hiring a bicycle consists of a fixed charge of 500 cents together with a charge of 3 cents per minute. The number of minutes for which people hire a bicycle has mean 142 and standard deviation 35.

(i) Find the mean and standard deviation of the amount people pay when hiring a bicycle. [3]

(ii) 6 people hire bicycles independently. Find the mean and standard deviation of the total amount paid by all 6 people. [3]

Answers. **(i)** 926, 105; **(ii)** 5556, 257. **N12/72/Q3**

18 The lengths of logs are normally distributed with mean 3.5 m and standard deviation 0.12 m. Describe fully the distribution of the total length of 8 randomly chosen logs. [3]

Answers: Normal distribution, Mean 28, Variance 0.115. **N12/73/Q1**

19 The masses of a certain variety of potato are normally distributed with mean 180 g and variance 1550 g^2. Two potatoes of this variety are chosen at random. Find the probability that the mass of one of these potatoes is at least twice the mass of the other. [7]

Answer: 0.041 (2 significant figures sufficient here). **N12/73/Q4**

20 Kieran and Andreas are long-jumpers. They model the lengths, in metres, that they jump by the independent random variables $K \sim N(5.64, 0.0576)$ and $A \sim N(4.97, 0.0441)$ respectively. They each make a jump and measure the length. Find the probability that

(i) the sum of the lengths of their jumps is less than 11 m, [4]

(ii) Kieran jumps more than 1.2 times as far as Andreas. [6]

Answer: 0.889
0.176 **N13/71/Q7**

21 The lifetimes, in hours, of Longlive light bulbs and Enerlow light bulbs have the independent distributions $N(1020, 45^2)$ and $N(2800, 52^2)$ respectively.

(i) Find the probability that the total of the lifetimes of 5 randomly chosen Longlive bulbs is less than 5200 hours. [4]

(ii) Find the probability that the lifetime of a randomly chosen Enerlow bulb is at least 3 times that of a randomly chosen Longlive bulb. [6]

Answer: (i) 0.840, (ii) 0.0361 or 0.0362 **N13/73/Q6**

22 The masses, in grams, of potatoes of types A and B have the distributions $N(175, 60^2)$ and $N(105, 28^2)$ respectively. Find the probability that a randomly chosen potato of type A has a mass that is at least twice the mass of a randomly chosen potato of type B. [5]

Answer: 0.335 **N14/71/Q1**

23 The masses, in grams, of tomatoes of type A and type B have the distributions $N(125, 30^2)$ and $N(130, 32^2)$ respectively.

(i) Find the probability that the total mass of 4 randomly chosen tomatoes of type A and 6 randomly chosen tomatoes of type B is less than 1.5 kg. [5]

(ii) Find the probability that a randomly chosen tomato of type A has a mass that is at least 90% of the mass of a randomly chosen tomato of type B. [5]

Answers: (i) 0.987 , (ii) 0.576 **N14/73/Q4**

24 The weights, in kilograms, of men and women have the distributions $N(78, 7^2)$ and $N(66, 5^2)$ respectively.

(i) The maximum load that a certain cable car can carry safely is 1200 kg. If 9 randomly chosen men and 7 randomly chosen women enter the cable car, find the probability that the cable car can operate safely. [5]

(ii) Find the probability that a randomly chosen woman weighs more than a randomly chosen man. [4]

Answer: 0.927
0.0815

N15/71/Q6

---

25 The thickness of books in a large library is normally distributed with mean 2.4 cm and standard deviation 0.3 cm.

(i) Find the probability that the total thickness of 6 randomly chosen books is more than 16 cm. [4]

(ii) Find the probability that the thickness of a book chosen at random is less than 1.1 times the thickness of a second book chosen at random. [5]

Answer: 0.0147
0.705

J16/71/Q5

---

26 Bags of sugar are packed in boxes, each box containing 20 bags. The masses of the boxes, when empty, are normally distributed with mean 0.4 kg and standard deviation 0.01 kg. The masses of the bags are normally distributed with mean 1.02 kg and standard deviation 0.03 kg.

(i) Find the probability that the total mass of a full box of 20 bags is less than 20.6 kg. [5]

(ii) Two full boxes are chosen at random. Find the probability that they differ in mass by less than 0.02 kg. [5]

Answers: (i) 0.0684 (ii) 0.0836

J16/73/Q7

---

27 Large packets of sugar are packed in cartons, each containing 12 packets. The weights of these packets are normally distributed with mean 505 g and standard deviation 3.2 g. The weights of the cartons, when empty, are independently normally distributed with mean 150 g and standard deviation 7 g.

(i) Find the probability that the total weight of a full carton is less than 6200 g. [5]

Small packets of sugar are packed in boxes. The total weight of a full box has a normal distribution with mean 3130 g and standard deviation 12.1 g.

(ii) Find the probability that the weight of a randomly chosen full carton is less than double the weight of a randomly chosen full box. [5]

Answers: (i) 0.223 (ii) 0.965

J17/71/Q5

28 The mass, in tonnes, of iron ore produced per day at a mine is normally distributed with mean 7.0 and standard deviation 0.46. Find the probability that the total amount of iron ore produced in 10 randomly chosen days is more than 71 tonnes. [5]

Answer: 0.246 **J17/73/Q3**

29 The volume, in millilitres, of a small cup of coffee has the distribution N(103.4, 10.2). The volume of a large cup of coffee is 1.5 times the volume of a small cup of coffee.

(i) Find the mean and standard deviation of the volume of a large cup of coffee. [3]

(ii) Find the probability that the total volume of a randomly chosen small cup of coffee and a randomly chosen large cup of coffee is greater than 250 ml. [4]

Answers: (i) 155.1 4.79 (ii) 0.930 **J18/71/Q4**

30 The times, in minutes, taken to complete the two parts of a task are normally distributed with means 4.5 and 2.3 respectively and standard deviations 1.1 and 0.7 respectively.

(i) Find the probability that the total time taken for the task is less than 8.5 minutes. [4]

(ii) Find the probability that the time taken for the first part of the task is more than twice the time taken for the second part. [5]

Answers: (i) 0.904 (ii) 0.478 **J18/73/Q6**

31 Each week a farmer sells X litres of milk and Y kg of cheese, where X and Y have the independent distributions $N(1520, 53^2)$ and $N(175, 12^2)$ respectively.

(i) Find the mean and standard deviation of the total amount of milk that the farmer sells in 4 randomly chosen weeks. [2]

During a year when milk prices are low, the farmer makes a loss of 2 cents per litre on milk and makes a profit of 21 cents per kg on cheese, so the farmer's overall weekly profit is $(21Y - 2X)$ cents.

(ii) Find the probability that, in a randomly chosen week, the farmer's overall profit is positive. [5]

Answer: 6080 106
0.990 **N16/71/Q4**

32 Men arrive at a clinic independently and at random, at a constant mean rate of 0.2 per minute. Women arrive at the same clinic independently and at random, at a constant mean rate of 0.3 per minute.

(i) Find the probability that at least 2 men and at least 3 women arrive at the clinic during a 5-minute period. [4]

(ii) Find the probability that fewer than 36 people arrive at the clinic during a 1-hour period. [5]

Answer: 0.0505

Answer: 0.842

N16/73/Q7

---

33 The numbers of barrels of oil, in millions, extracted per day in two oil fields A and B are modelled by the independent random variables X and Y respectively, where $X \sim N(3.2, 0.4^2)$ and $Y \sim N(4.3, 0.6^2)$. The income generated by the oil from the two fields is $90 per barrel for A and $95 per barrel for B.

(i) Find the mean and variance of the daily income, in millions of dollars, generated by field A. [3]

(ii) Find the probability that the total income produced by the two fields in a day is at least $670 million. [5]

Answers: (i) mean = 288 variance = 1296, (ii) 0.653

N17/71/Q6

---

34 The times, in months, taken by a builder to build two types of house, P and Q, are represented by the independent variables $T_1 \sim N(2.2, 0.4^2)$ and $T_2 \sim N(2.8, 0.5^2)$ respectively.

(i) Find the probability that the total time taken to build one house of each type is less than 6 months. [4]

(ii) Find the probability that the time taken to build a type Q house is more than 1.2 times the time taken to build a type P house. [5]

Answers: (i) 0.941 (ii) 0.591

N18/71/Q5

Chapter 3: Continuous Random Variables

The following are examples of **continuous random variables:**

- the mass, in grams, of a bag of sugar packaged by a particular machine
- the time taken, in minutes, to perform a task,
- the height, in centimetres, of a five-year-old girl,
- the lifetime, in hours, of a 100-watt light bulb.

PROBABILITY DENSITY FUNCTION

A continuous random variable X is given by its **probability density function (p.d.f.)**, which is specified for the range of values for which x is valid. The function can be illustrated by a curve, $y = f(x)$. Note that this function cannot be negative throughout the specified range.

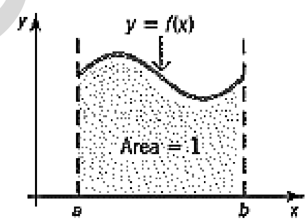
Probabilities are given by the area under the curve. It is sometimes possible to find an area by geometry, for example by using formulae for the area of a triangle or a trapezium. Often, however, areas need to be calculated using integration.

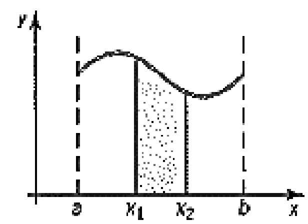
In general, for a continuous random variable X, with p.d.f. $f(x)$ valid over the range $a \leqslant x \leqslant b$

(a) $\int_a^b f(x)dx = 1$

(b) for $a \leqslant x_1 \leqslant x_2 \leqslant b$

$$P(x_1 \leqslant X \leqslant x_2) = \int_{x_1}^{x_2} f(x)dx$$





Remember that in an experimental approach, the area under the histogram represents frequency. In a theoretical approach, the area under the curve $y = f(x)$ represents probability.

Example 1

X is the delay, in hours, of a flight from Chicago, where

$$f(x) = 0.2 - 0.02x, \qquad 0 \leqslant x \leqslant 10$$

Find

(a) the probability that the delay will be less than four hours,
(b) the probability that the delay will be between two and six hours.

Example 2

X is the continuous variable, the mass, in kilograms, of a substance produced per minute in an industrial process, where

$$f(x) = \begin{cases} \frac{1}{36}x(6-x) & (0 \leqslant x \leqslant 6) \\ 0 & \text{otherwise} \end{cases}$$

Find the probability that the mass is more than 5 kg.

Example 3

A continuous random variable has p.d.f. $f(x) = kx^2$ for $0 \leqslant x \leqslant 4$.

(a) Find the value of the constant k.
(b) Find $P(1 \leqslant X \leqslant 3)$.

Example 4

The continuous random variable X has p.d.f. $f(x)$ where

$$f(x) = \begin{cases} k(x+2)^2 & -2 \leqslant x < 0 \\ 4k & 0 \leqslant x \leqslant 1\frac{1}{3} \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of the constant k.
(b) Sketch $y = f(x)$.
(c) Find $P(-1 \leqslant X \leqslant 1)$.
(d) Find $P(X > 1)$.

EXPECTATION OF X, E(X)

For a continuous random variable with p.d.f. $f(x)$,

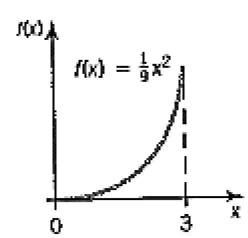
$$E(X) = \int_{\text{all } x} x f(x)\, dx$$

$E(X)$ is referred to as the mean or expectation of X and is often denoted by μ.

Example 5

The sketch shows the p.d.f. of X where $f(x) = \frac{1}{9}x^2$, $0 \leqslant x \leqslant 3$.

(a) Find μ, the mean of X.
(b) Find $P(X < \mu)$.



ution 6.5

(a) $\mu = E(X)$

If f(x) has a line of symmetry in the specified range, then E(X) can be found directly as in the following example.

Example 6

A continuous random variable X has p.d.f. $f(x)$ where

$$f(x) = \begin{cases} 0.25x & 0 \leqslant x < 2 \\ 1 - 0.25x & 2 \leqslant x \leqslant 4 \\ 0 & \text{otherwise} \end{cases}$$

Sketch $y = f(x)$ and find $E(X)$.

Example 7

A teacher of young children is thinking of asking her class to guess her height in metres. The teacher considers that the height guessed by a randomly selected child can be modelled by the random variable H with probability density function

$$f(h) = \begin{cases} \frac{3}{16}(4h - h^2) & 0 \leqslant h \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

Using this model,

(a) find $P(H < 1)$,
(b) show that $E(H) = 1.25$.

A friend of the teacher suggests that the random variable X with probability density function

$$g(x) = \begin{cases} kx^3 & 0 \leqslant x \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant, might be a more suitable model.

(c) Show that $k = \frac{1}{4}$.
(d) Find $P(X < 1)$.
(e) Find $E(X)$.
(f) Using your calculations in (a), (b), (d) and (e), state, giving reasons, which of the random variables H or X is likely to be the more appropriate model in this instance. (L)

VARIANCE OF X, VAR(X)

For a random variable X,

$$\text{Var}(X) = E(X-\mu)^2 \qquad \text{where } \mu = E(X).$$

As in the discrete case (see page 249) the formula can be written:

$$\begin{aligned}\text{Var}(X) &= E(X^2) - E^2(X)\\ &= E(X^2) - \mu^2\end{aligned}$$

If X is a continuous random variable with p.d.f. $f(x)$, then

$$\text{Var}(X) = \int_{\text{all } x} x^2 f(x)dx - \mu^2$$

where $\quad \mu = E(X) = \int_{\text{all } x} xf(x)dx$

The standard deviation of X is often written as σ, so $\sigma = \sqrt{\text{Var}(X)}$.

As in the case of the discrete random variable (see page 250), the following results also hold when X is continuous; where a and b are constants

1. $\text{Var}(a) = 0$
2. $\text{Var}(aX) = a^2\,\text{Var}(X)$
3. $\text{Var}(aX + b) = a^2\,\text{Var}(X)$

Example 8

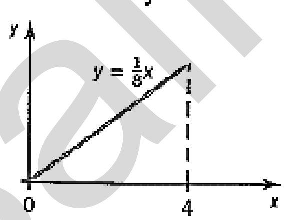
As an experiment a temporary roundabout is installed at the crossroads. The time, X minutes, which vehicles have to wait before entering the roundabout has probability density function

$$f(x) = \begin{cases} 0.8 - 0.32x & 0 \leqslant x \leqslant 2.5 \\ 0 & \text{otherwise} \end{cases}$$

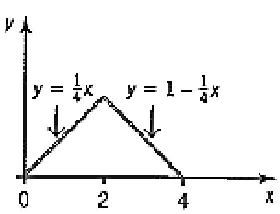
Find the mean and the standard deviation of X. *(AEB)*

THE MODE

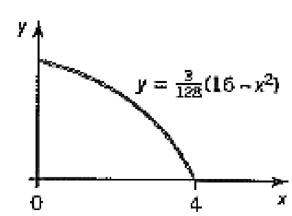
The mode is the value of X for which $f(x)$ is greatest in the given range of X.
To locate the mode it is a good idea to draw a sketch. Sometimes the mode can be deduced immediately.



Mode is 4



Mode is 2



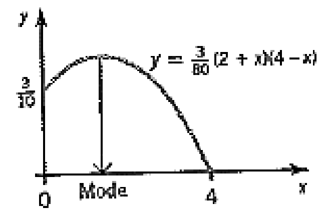
Mode is 0

For some probability density functions you will need to determine the maximum point on the curve $y = f(x)$ using the fact that, at a maximum point, $f'(x) = 0$, where $f'(x) = \frac{d}{dx} f(x)$.

Note that a maximum point is confirmed if $f''(x) < 0$, where $f''(x) = \frac{d}{dx} f'(x)$.

Example 9

X has p.d.f. defined by $f(x) = \frac{3}{80}(2+x)(4-x)$, for $0 \leqslant x \leqslant 4$ and is illustrated in the diagram. Find the mode.



Example 10

A random variable X has a probability density function

$f(x) = Ax(6-x)^2 \qquad 0 \leqslant x \leqslant 6$

$\quad = 0 \qquad$ elsewhere.

(a) Find the value of the constant A.

(b) Calculate

(i) the mean, (ii) the mode,

(iii) the variance, (iv) the standard deviation of X. *(AEB)*

Example 11

The time taken to perform a particular task, t hours, has the probability density function

$$f(t) = \begin{cases} 10ct^2 & 0 \leqslant t < 0.6 \\ 9c(1-t) & 0.6 \leqslant t \leqslant 1.0 \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant.

(a) Find the value of c and sketch the graph of this distribution.

(b) Write down the most likely time.

(c) Find the expected time.

(d) Determine the probability that the time will be

(i) more than 48 minutes,

(ii) between 24 and 48 minutes.

FINDING THE MEDIAN, QUARTILES AND OTHER PERCENTILES

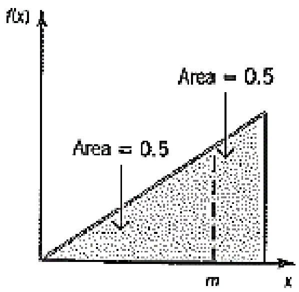
The median is the value 50% of the way through the distribution. It splits the area under the curve $y = f(x)$ into two halves. If m is the median, then for $f(x)$ defined for $a \leqslant x \leqslant b$,

$$\int_a^m f(x)\,dx = 0.5$$

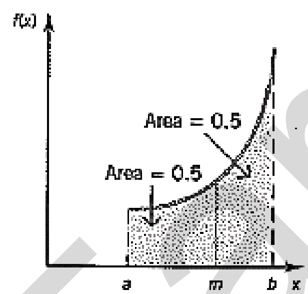
i.e. $\quad F(m) = 0.5$

For example:

(a)



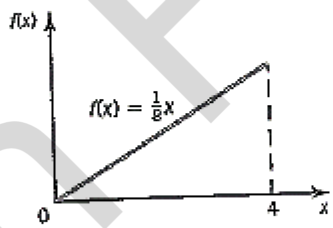
(b)



Example 12

X is a continuous random variable with p.d.f. as shown.

$f(x) = \frac{1}{8}x, \quad 0 \leqslant x \leqslant 4$



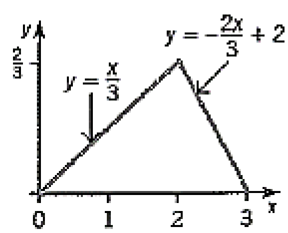
Find

(a) the cumulative distribution function $F(x)$ and sketch $y = F(x)$,
(b) $P(0.3 \leqslant X \leqslant 1.8)$,
(c) the median, m,
(d) the interquartile range.

Example 13

X is a continuous random variable with p.d.f. $f(x)$ where

$$f(x) = \begin{cases} \dfrac{x}{3} & 0 \leqslant x \leqslant 2 \\ -\dfrac{2x}{3} + 2 & 2 \leqslant x \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$



(a) Find the cumulative distribution function $F(x)$ and sketch it.
(b) Find $P(1 \leqslant X \leqslant 2.5)$.
(c) Find the median, m.

SELECTED PAST PAPER QUESTIONS

Question 1
N13/72/Q5

The volume, in cm^3, of liquid left in a glass by people when they have finished drinking all they want is modelled by the random variable X with probability density function given by

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

where k is a constant.

(i) Show that $k = \frac{3}{8}$. [2]

(ii) 20% of people leave at least $d\,\text{cm}^3$ of liquid in a glass. Find d. [3]

(iii) Find $\mathrm{E}(X)$. [3]

Question 2
J14/72/Q6

The time, T hours, spent by people on a visit to a museum has probability density function

$$f(t) = \begin{cases} kt(16-t^2) & 0 \leqslant t \leqslant 4, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{1}{64}$. [3]

(ii) Calculate the probability that two randomly chosen people each spend less than 1 hour on a visit to the museum. [4]

(iii) Find the mean time spent on a visit to the museum. [3]

Question 3
N11/72/Q7

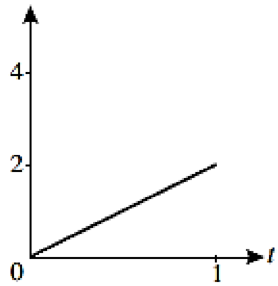


Fig. 1

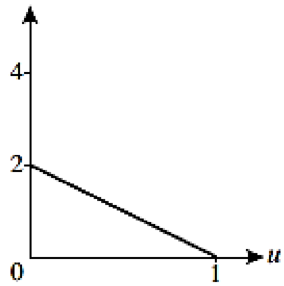


Fig. 2

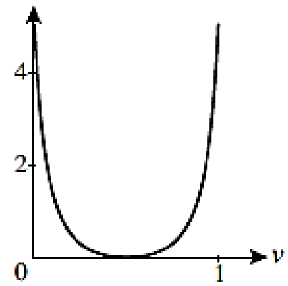


Fig. 3

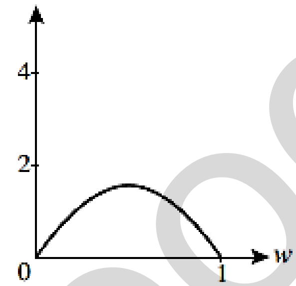


Fig. 4

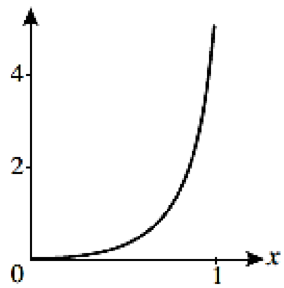


Fig. 5

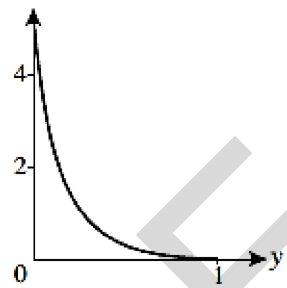


Fig. 6

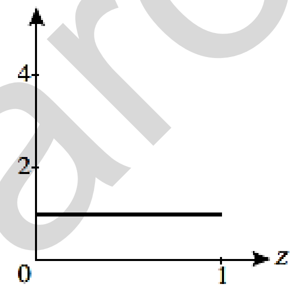


Fig. 7

Each of the random variables T, U, V, W, X, Y and Z takes values between 0 and 1 only. Their probability density functions are shown in Figs 1 to 7 respectively.

(i) (a) Which of these variables has the largest median? [1]

(b) Which of these variables has the largest standard deviation? Explain your answer. [2]

(ii) Use Fig. 2 to find $P(U < 0.5)$. [2]

(iii) The probability density function of X is given by

$$f(x) = \begin{cases} ax^n & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

where a and n are positive constants.

(a) Show that $a = n + 1$. [3]

(b) Given that $E(X) = \frac{5}{6}$, find a and n. [4]

Homework: Continuous Random Variables– Variant 62

1 A random variable X has probability density function given by

$$f(x) = \begin{cases} 1 - \frac{1}{2}x & 0 \leqslant x \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find $P(X > 1.5)$. [2]

(ii) Find the mean of X. [2]

(iii) Find the median of X. [3]

Answer: **(i)** 0.0625; **(ii)** $\frac{2}{3}$; **(iii)** 0.586 . **J03/Q4**

2 The queuing time, T minutes, for a person queuing at a supermarket checkout has probability density function given by

$$f(t) = \begin{cases} ct(25 - t^2) & 0 \leqslant t \leqslant 5, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant.

(i) Show that the value of c is $\frac{4}{625}$. [3]

(ii) Find the probability that a person will have to queue for between 2 and 4 minutes. [3]

(iii) Find the mean queuing time. [4]

Answers: **(ii)** 0.576; **(iii)** $\frac{8}{3}$. **J04/Q7**

3 The random variable X denotes the number of hours of cloud cover per day at a weather forecasting centre. The probability density function of X is given by

$$f(x) = \begin{cases} \dfrac{(x-18)^2}{k} & 0 \leqslant x \leqslant 24, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = 2016$. [3]

(ii) On how many days in a year of 365 days can the centre expect to have less than 2 hours of cloud cover? [3]

(iii) Find the mean number of hours of cloud cover per day. [4]

Answers: **(ii)** 104 or 105; **(iii)** 5.14. **J05/Q7**

4 The random variable X has probability density function given by

$$f(x) = \begin{cases} 4x^k & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a positive constant.

(i) Show that $k = 3$. [2]

(ii) Show that the mean of X is 0.8 and find the variance of X. [4]

(iii) Find the upper quartile of X. [2]

(iv) Find the interquartile range of X. [2]

Answers: **(ii)** 0.0267; **(iii)** 0.931; **(iv)** 0.223. **J06/Q5**

5 The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{3}{4}(x^2 - 1) & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Sketch the probability density function of X. [2]

(ii) Show that the mean, μ, of X is 1.6875. [3]

(iii) Show that the standard deviation, σ, of X is 0.2288, correct to 4 decimal places. [3]

(iv) Find $P(1 \leqslant X \leqslant \mu + \sigma)$. [3]

Answer: **(iv)** 0.822 . **J07/Q7**

6 If Usha is stung by a bee she always develops an allergic reaction. The time taken in minutes for Usha to develop the reaction can be modelled using the probability density function given by

$$f(t) = \begin{cases} \dfrac{k}{t+1} & 0 \leqslant t \leqslant 4, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \dfrac{1}{\ln 5}$. [4]

(ii) Find the probability that it takes more than 3 minutes for Usha to develop a reaction. [3]

(iii) Find the median time for Usha to develop a reaction. [3]

Answers: **(ii)** 0.139; **(iii)** 1.24 minutes. **J08/Q7**

7 The time in minutes taken by candidates to answer a question in an examination has probability density function given by

$$f(t) = \begin{cases} k(6t - t^2) & 3 \leqslant t \leqslant 6, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{1}{18}$. [3]

(ii) Find the mean time. [3]

(iii) Find the probability that a candidate, chosen at random, takes longer than 5 minutes to answer the question. [2]

(iv) Is the upper quartile of the times greater than 5 minutes, equal to 5 minutes or less than 5 minutes? Give a reason for your answer. [2]

Answers: **(ii)** 33/8
(iii) 4/27
(iv) UQ is less than 5

J09/71/Q5

---

8 The height in metres reached by a sunflower can be modelled by the probability density function given by

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

where k is a constant.

(i) Show that $k = \frac{3}{4}$. [3]

(ii) Find the mean height reached by a sunflower. [3]

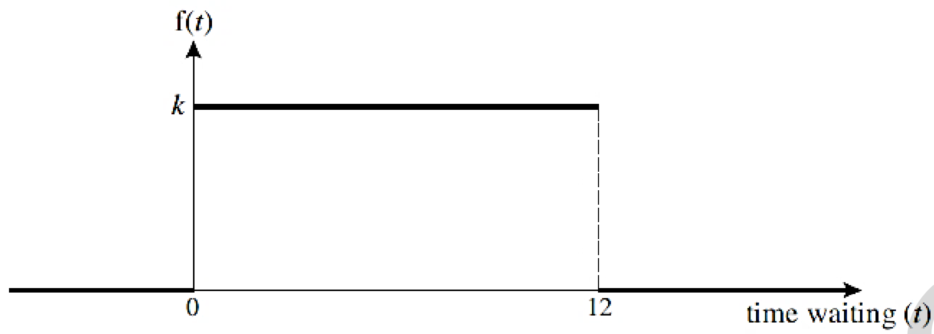
(iii) Find the probability that a randomly chosen sunflower reaches a height of more than 1.3 metres. [2]

(iv) Is the median height greater than 1.3 metres, less than 1.3 metres or equal to 1.3 metres? Justify your answer. [2]

Answers: **(ii)** 1.2
(iii) 0.437
(iv) Median is less than 1.3

J09/72/Q5

9



Fred arrives at random times on a station platform. The times in minutes he has to wait for the next train are modelled by the continuous random variable for which the probability density function f is shown above.

(i) State the value of k. [1]

(ii) Explain briefly what this graph tells you about the arrival times of trains. [1]

Answers: **(i)** $\frac{1}{12}$; **(ii)** Trains arrive every 12 minutes. **J10/72/Q1**

10 The random variable T denotes the time in seconds for which a firework burns before exploding. The probability density function of T is given by

$$f(t) = \begin{cases} ke^{0.2t} & 0 \leqslant t \leqslant 5, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \dfrac{1}{5(e-1)}$. [3]

(ii) Sketch the probability density function. [2]

(iii) 80% of fireworks burn for longer than a certain time before they explode. Find this time. [3]

Answer: **(iii)** 1.48 seconds. **J10/72/Q5**

11 The average speed of a bus, $x\,\text{km h}^{-1}$, on a certain journey is a continuous random variable X with probability density function given by

$$f(x) = \begin{cases} \dfrac{k}{x^2} & 20 \leqslant x \leqslant 28, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that $k = 70$. [3]

(ii) Find $E(X)$. [3]

(iii) Find $P(X < E(X))$. [2]

(iv) Hence determine whether the mean is greater or less than the median. [2]

Answers: **(ii)** 23.6; **(iii)** 0.528; **(iv)** Mean is greater. **N02/Q6**

12 The lifetime, x years, of the power light on a freezer, which is left on continuously, can be modelled by the continuous random variable with density function given by

$$f(x) = \begin{cases} ke^{-3x} & x > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = 3$. [2]

(ii) Find the lower quartile. [3]

(iii) Find the mean lifetime. [6]

Answers: **(ii)** 0.0959; **(iii)** $\frac{1}{3}$. **N03/Q7**

13 A continuous random variable X has probability density function given by

$$f(x) = \begin{cases} 3(1-x)^2 & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find

(i) $P(X > 0.5)$, [3]

(ii) the mean and variance of X. [6]

Answers: **(i)** 0.125; **(ii)** 0.25, 0.0375. **N04/Q6**

14 A continuous random variable X has probability density function given by

$$f(x) = \begin{cases} a + \frac{1}{3}x & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where a is a constant.

(i) Show that the value of a is $\frac{1}{2}$. [3]

(ii) Find $P(X > 1.8)$. [2]

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

Answers: **(ii)** 0.227; **(iii)** 1.53. **N05/Q5**

15 At a town centre car park the length of stay in hours is denoted by the random variable X, which has probability density function given by

$$f(x) = \begin{cases} kx^{-\frac{3}{2}} & 1 \leqslant x \leqslant 9, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Interpret the inequalities $1 \leqslant x \leqslant 9$ in the definition of $f(x)$ in the context of the question. [1]

(ii) Show that $k = \frac{3}{4}$. [2]

(iii) Calculate the mean length of stay. [3]

The charge for a length of stay of x hours is $\left(1 - e^{-x}\right)$ dollars.

(iv) Find the length of stay for the charge to be at least 0.75 dollars [3]

(v) Find the probability of the charge being at least 0.75 dollars. [2]

Answers: (i) All cars stayed between 1 and 9 hours; (iii) 3 hours; (iv) Greater than 1.39 hours; (v) 0.774. **N06/Q7**

---

16 The length, X cm, of a piece of wooden planking is a random variable with probability density function given by

$$f(x) = \begin{cases} \dfrac{1}{b} & 0 \leqslant x \leqslant b, \\ 0 & \text{otherwise,} \end{cases}$$

where b is a positive constant.

(i) Find the mean and variance of X in terms of b. [3]

The lengths of a random sample of 100 pieces were measured and it was found that $\Sigma x = 950$.

(ii) Show that the value of b estimated from this information is 19. [2]

Using this value of b,

(iii) find the probability that the length of a randomly chosen piece is greater than 11 cm, [1]

(iv) find the probability that the mean length of a random sample of 336 pieces is less than 9 cm. [4]

N07/Q5

Answers: (i) $\frac{b}{2}$, $\frac{b^2}{12}$; (iii) $\frac{8}{19}$; (iv) 0.0474.

17 The time in hours taken for clothes to dry can be modelled by the continuous random variable with probability density function given by

$$f(t) = \begin{cases} k\sqrt{t} & 1 \leqslant t \leqslant 4, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{3}{14}$. [3]

(ii) Find the mean time taken for clothes to dry. [4]

(iii) Find the median time taken for clothes to dry. [3]

(iv) Find the probability that the time taken for clothes to dry is between the mean time and the median time. [2]

Answers: **(ii)** 2.66 hours; **(iii)** 2.73 hours; **(iv)** 0.0243. **N08/Q7**

18 The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{3}x(k - x) & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

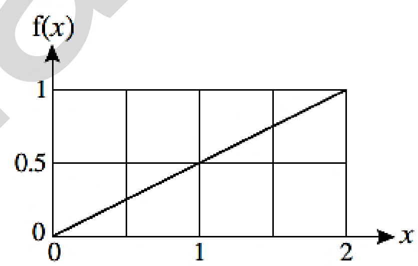
(i) Show that the value of k is $\frac{32}{9}$. [3]

(ii) Find $E(X)$. [2]

(iii) Is the median less than or greater than 1.5? Justify your answer numerically. [3]

Answers: **(ii)** 1.52; **(iii)** Median > 1.5. **N09/72/Q6**

19



The diagram shows the graph of the probability density function, f, of a random variable X which takes values between 0 and 2 only.

(i) Find $P(1 < X < 1.5)$. [2]

(ii) Find the median of X. [3]

(iii) Find $E(X)$. [2]

Answers: (i) $\frac{5}{16}$; (ii) $\sqrt{2}$; (iii) $\frac{4}{3}$. **N10/72/Q4**

---

20 A random variable X has probability density function given by

$$f(x) = \begin{cases} k(1-x) & -1 \leqslant x \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{1}{2}$. [2]

(ii) Find $P(X > \frac{1}{2})$. [1]

(iii) Find the mean of X. [3]

(iv) Find a such that $P(X < a) = \frac{1}{4}$. [3]

Answers: (ii) 0.0625; (iii) –0.333; (iv) –0.732. **J11/72/Q7**

---

21 At a certain shop the weekly demand, in kilograms, for flour is modelled by the random variable X with probability density function given by

$$f(x) = \begin{cases} kx^{-\frac{1}{2}} & 4 \leqslant x \leqslant 25, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{1}{6}$. [2]

(ii) Calculate the mean weekly demand for flour at the shop. [3]

(iii) At the beginning of one week, the shop has 20 kg of flour in stock. Find the probability that this will not be enough to meet the demand for that week. [2]

(iv) Give a reason why the model may not be realistic. [1]

Answers: (ii) 13 (iii) 0.176 (iv) Weekly demand may be >25 (or <4) **J12/72/Q6**

---

22 A random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{2}{3}x & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find $E(X)$. [3]

(ii) Find $P(X < E(X))$. [2]

(iii) Hence explain whether the mean of X is less than, equal to or greater than the median of X. [2]

Answer: 14/9
0.473
Mean<median

J13/72/Q2

---

23 The time, T hours, spent by people on a visit to a museum has probability density function

$$f(t) = \begin{cases} kt(16 - t^2) & 0 \leqslant t \leqslant 4, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{1}{64}$. [3]

(ii) Calculate the probability that two randomly chosen people each spend less than 1 hour on a visit to the museum. [4]

(iii) Find the mean time spent on a visit to the museum. [3]

Answers: 0.0147
2.13

J14/72/Q6

---

24 The waiting time, T minutes, for patients at a doctor's surgery has probability density function given by

$$f(t) = \begin{cases} k(225 - t^2) & 0 \leqslant t \leqslant 15, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{1}{2250}$. [3]

(ii) Find the probability that a patient has to wait for more than 10 minutes. [3]

(iii) Find the mean waiting time. [4]

Answer: **(i)** $\frac{4}{27}$

(ii) $\frac{45}{8}$

J15/72/Q6

25

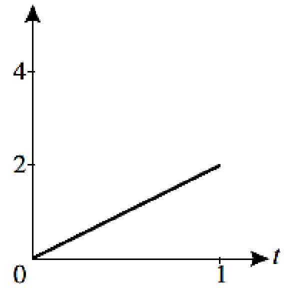


Fig. 1

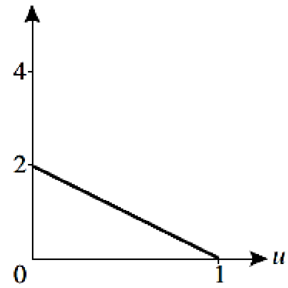


Fig. 2

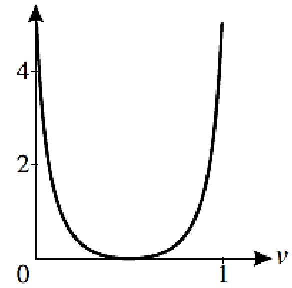


Fig. 3

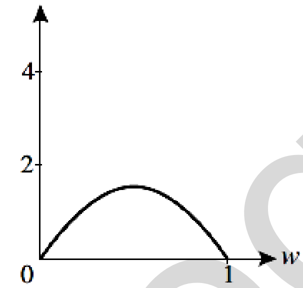


Fig. 4

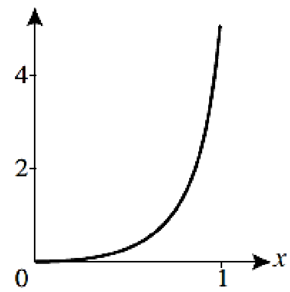


Fig. 5

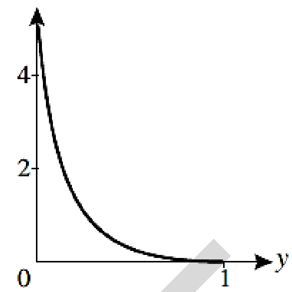


Fig. 6

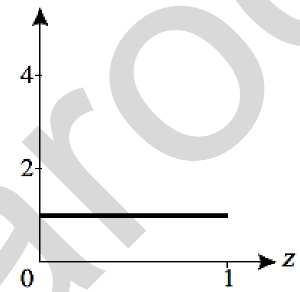


Fig. 7

Each of the random variables T, U, V, W, X, Y and Z takes values between 0 and 1 only. Their probability density functions are shown in Figs 1 to 7 respectively.

(i) **(a)** Which of these variables has the largest median? [1]

(b) Which of these variables has the largest standard deviation? Explain your answer. [2]

(ii) Use Fig. 2 to find $\mathrm{P}(U < 0.5)$. [2]

(iii) The probability density function of X is given by

$$\mathrm{f}(x) = \begin{cases} ax^n & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

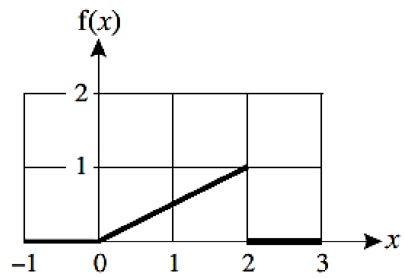
where a and n are positive constants.

(a) Show that $a = n + 1$. [3]

(b) Given that $\mathrm{E}(X) = \frac{5}{6}$, find a and n. [4]

Answer: (i) (a) X or 5; (b) V or 3, Higher and lower values more likely; (ii) 0.75; (iii) (b) a=5, n=4 **N11/72/Q7**

26



The diagram shows the graph of the probability density function, f, of a random variable X. Find the median of X. [3]

Answer: m = √2. **N12/72/Q1**

27 A random variable X has probability density function given by

$$f(x) = \begin{cases} \dfrac{k}{x-1} & 3 \leqslant x \leqslant 5, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \dfrac{1}{\ln 2}$. [4]

(ii) Find a such that $P(X < a) = 0.75$. [4]

Answer: (ii) 4.36. **N12/72/Q5**

28 The volume, in cm^3, of liquid left in a glass by people when they have finished drinking all they want is modelled by the random variable X with probability density function given by

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

where k is a constant.

(i) Show that $k = \frac{3}{8}$. [2]

(ii) 20% of people leave at least d cm^3 of liquid in a glass. Find d. [3]

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

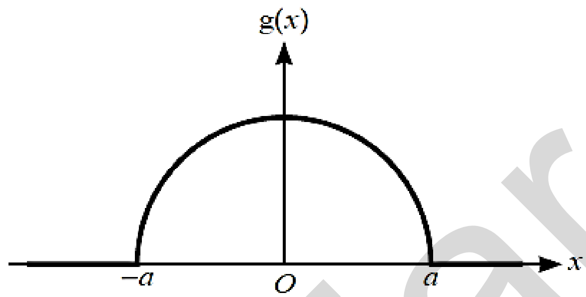
Answer: 0.830
1/2 **N13/72/Q5**

29 **(a)** The time for which Lucy has to wait at a certain traffic light each day is T minutes, where T has probability density function given by

$$f(t) = \begin{cases} \frac{3}{2}t - \frac{3}{4}t^2 & 0 \leqslant t \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

Find the probability that, on a randomly chosen day, Lucy has to wait for less than half a minute at the traffic light. [3]

(b)



The diagram shows the graph of the probability density function, g, of a random variable X. The graph of g is a semicircle with centre (0, 0) and radius a. Elsewhere $g(x) = 0$.

(i) Find the value of a. [2]

(ii) State the value of $E(X)$. [1]

(iii) Given that $P(X < -c) = 0.2$, find $P(X < c)$. [2]

Answer: 0.156, 0.798, 0, 0.8 **N14/72/Q3**

30 A random variable X has probability density function given by

$$f(x) = \begin{cases} k(3 - x) & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{2}{3}$. [3]

(ii) Find the median of X. [4]

Answer: 1.42 **N15/72/Q4**

Homework: Continuous Random Variables– Variant 71 & 73

1 The time, in minutes, taken by volunteers to complete a task is modelled by the random variable X with probability density function given by

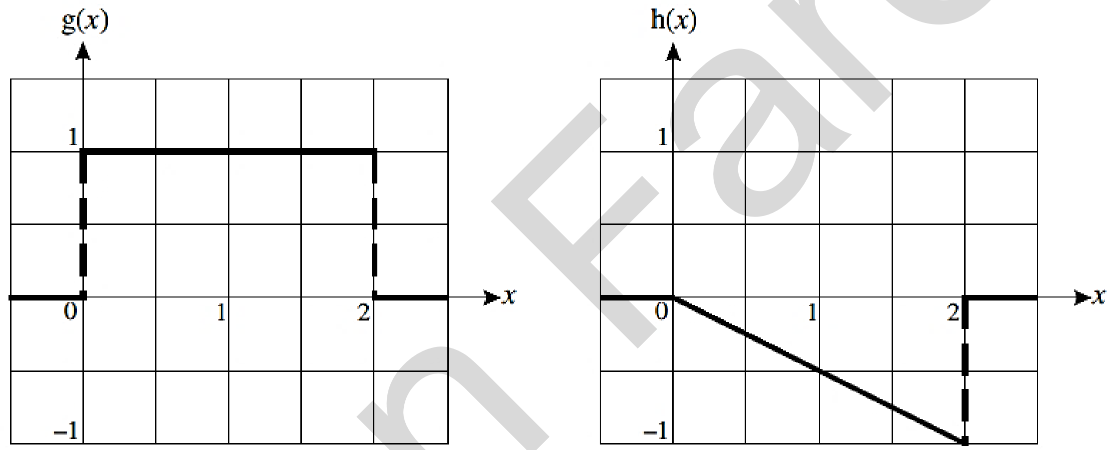
$$f(x) = \begin{cases} \dfrac{k}{x^4} & x \geqslant 1, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that $k = 3$. [2]

(ii) Find $E(X)$ and $Var(X)$. [6]

Answers: **(ii)** 1.5, 0.75. **J10/73/Q5**

2 **(a)**



The diagrams show the graphs of two functions, g and h. For each of the functions g and h, give a reason why it cannot be a probability density function. [2]

(b) The distance, in kilometres, travelled in a given time by a cyclist is represented by the continuous random variable X with probability density function given by

$$f(x) = \begin{cases} \dfrac{30}{x^2} & 10 \leqslant x \leqslant 15, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that $E(X) = 30 \ln 1.5$. [3]

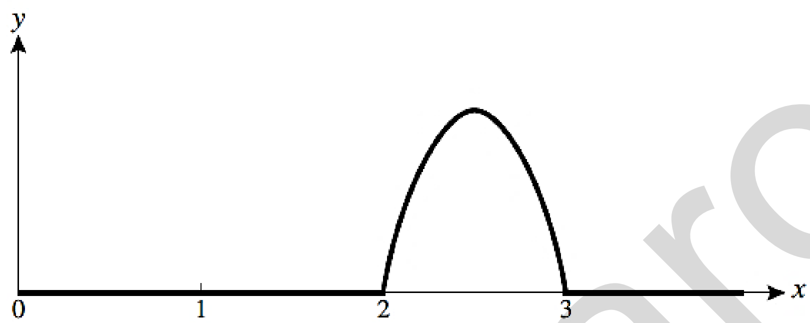
(ii) Find the median of X. Find also the probability that X lies between the median and the mean. [5]

Answers: **(a)** Area under graph is greater than 1; probability density function cannot take negative values; **(b)(ii)** 12, 0.0337 **J11/71/Q4**

3 The distance travelled, in kilometres, by a Grippo brake pad before it needs to be replaced is modelled by 10 000X, where X is a random variable having the probability density function

$$f(x) = \begin{cases} -k(x^2 - 5x + 6) & 2 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

The graph of $y = f(x)$ is shown in the diagram.



(i) Show that $k = 6$. [2]

(ii) State the value of E(X) and find Var(X). [4]

(iii) Sami fits four new Grippo brake pads on his car. Find the probability that at least one of these brake pads will need to be replaced after travelling less than 22 000 km. [3]

Answers: (ii) 2.5, 0.05; (iii) 0.355 or 0.356. **J11/73/Q6**

---

4 The random variable X has probability density function given by

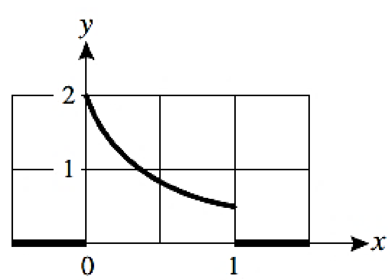
$$f(x) = \begin{cases} \dfrac{k}{(x+1)^2} & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = 2$. [2]

(ii) Find a such that $P(X < a) = \frac{1}{5}$. [3]

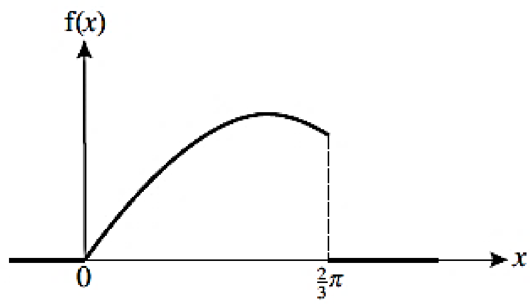
(iii)



The diagram shows the graph of $y = f(x)$. The median of X is denoted by m. Use the diagram to explain whether $m < 0.5$, $m = 0.5$ or $m > 0.5$. [2]

Answers: (ii) $a = 1/9$ (iii) $m<0.5$ **J12/71/Q4**

5



A random variable X has probability density function given by

$$f(x) = \begin{cases} k\sin x & 0 \leqslant x \leqslant \frac{2}{3}\pi, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant, as shown in the diagram.

(i) Show that $k = \frac{2}{3}$. [2]

(ii) Show that the median of X is 1.32, correct to 3 significant figures. [4]

(iii) Find $E(X)$. [4]

Answers: (i) $k = 2/3$ (AG) (ii) Median $= 1.32$ (3sf) (AG) (iii) 1.28 or equivalent surd form **J12/73/Q7**

6 The time in minutes taken by people to read a certain booklet is modelled by the random variable T with probability density function given by

$$f(t) = \begin{cases} \dfrac{1}{2\sqrt{t}} & 4 \leqslant t \leqslant 9, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the time within which 90% of people finish reading the booklet. [3]

(ii) Find $E(T)$ and $Var(T)$. [6]

Answer: 8.41 **J13/71/Q6**
19/3 2.09

7 A random variable X has probability density function given by

$$f(x) = \begin{cases} \dfrac{k}{x^3} & x \geqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = 2$. [2]

(ii) Find $P(1 \leqslant X \leqslant 2)$. [2]

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

Answers: (i) k=2 (ii) 0.75 (iii) 2 **J13/73/Q5**

---

8 A random variable X has probability density function given by

$$f(x) = \begin{cases} \dfrac{k}{x} & 1 \leqslant x \leqslant a, \\ 0 & \text{otherwise,} \end{cases}$$

where k and a are positive constants.

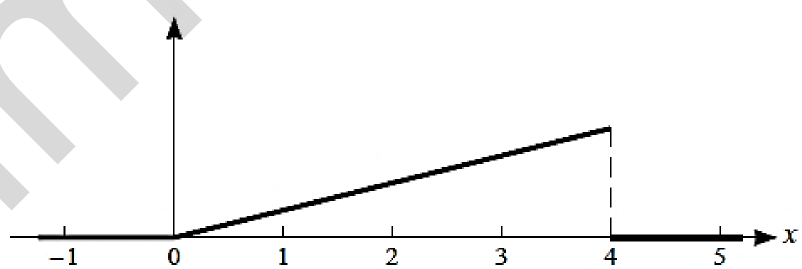
(i) Show that $k = \dfrac{1}{\ln a}$. [3]

(ii) Find $E(X)$ in terms of a. [3]

(iii) Find the median of X in terms of a. [4]

Answers: 2nd is more representative of *all* appointments **J14/71/Q7**
0.01 Concluding that times spent are too long when they are not
No reason to believe that appointments are too long
Normal population

9



A random variable X takes values between 0 and 4 only and has probability density function as shown in the diagram. Calculate the median of X. [3]

Answer: √8 or 2.83 or equivalent. **J14/73/Q2**

10 The lifetime, X years, of a certain type of battery has probability density function given by

$$f(x) = \begin{cases} \dfrac{k}{x^2} & 1 \leqslant x \leqslant a, \\ 0 & \text{otherwise,} \end{cases}$$

where k and a are positive constants.

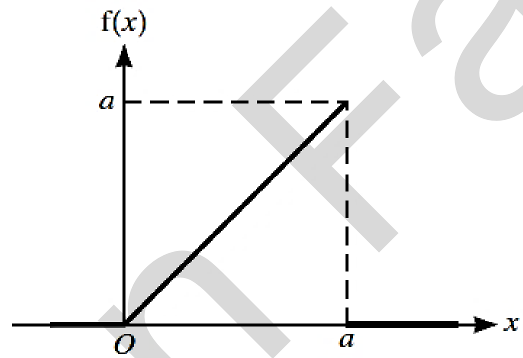
(i) State what the value of a represents in this context. [1]

(ii) Show that $k = \dfrac{a}{a-1}$. [3]

(iii) Experience has shown that the longest that any battery of this type lasts is 2.5 years. Find the mean lifetime of batteries of this type. [3]

Answers: **(i)** The longest lifetime of a battery **(ii)** $k = \frac{a}{a-1}$ AG **(iii)** 5/3 ln2.5 or 1.53 **J14/73/Q5**

11



The random variable X has probability density function, f, as shown in the diagram, where a is a constant. Find the value of a and hence show that $E(X) = 0.943$ correct to 3 significant figures. [5]

J15/71/Q1

12 The probability density function of the random variable X is given by

$$f(x) = \begin{cases} \frac{3}{4}x(c-x) & 0 \leqslant x \leqslant c, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant.

(i) Show that $c = 2$. [3]

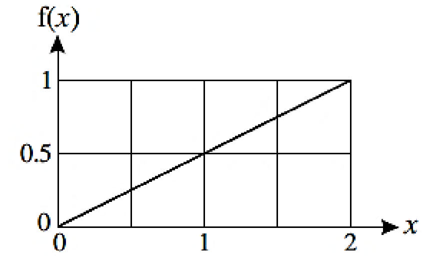
(ii) Sketch the graph of $y = f(x)$ and state the median of X. [3]

(iii) Find $P(X < 1.5)$. [4]

(iv) Hence write down the value of $P(0.5 < X < 1)$. [1]

Answers: **(ii)** Inverted parabola between (0,0) and (2,0) and 0 otherwise, median =1 **(iii)** 0.844 **(iv)** 0.344 **J15/73/Q7**

13



The diagram shows the graph of the probability density function, f, of a random variable X which takes values between 0 and 2 only.

(i) Find $P(1 < X < 1.5)$. [2]

(ii) Find the median of X. [3]

(iii) Find $E(X)$. [2]

Answers: **(i)** $\frac{5}{16}$; **(ii)** $\sqrt{2}$; **(iii)** $\frac{4}{3}$. **N10/71/Q4**

---

14 A continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{6}x & 2 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find $E(X)$. [3]

(ii) Find the median of X. [3]

(iii) Two independent values of X are chosen at random. Find the probability that both these values are greater than 3. [3]

Answers: **(i)** $\frac{28}{9}$ (3.11); **(ii)** $\sqrt{10}$ (3.16); **(iii)** $\frac{49}{144}$ (0.340). **N10/73/Q5**

15

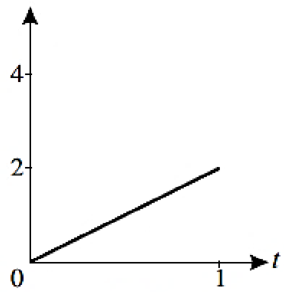


Fig. 1

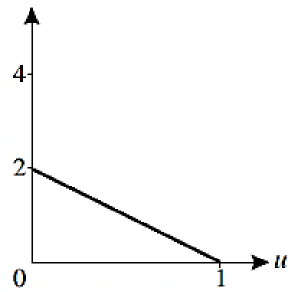


Fig. 2

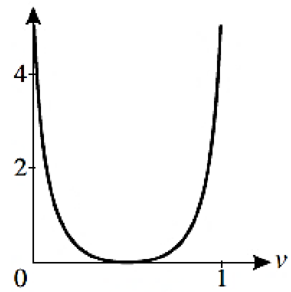


Fig. 3

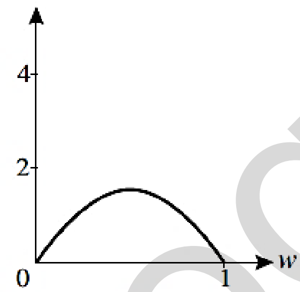


Fig. 4

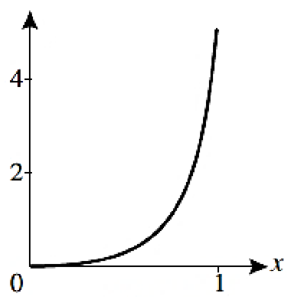


Fig. 5

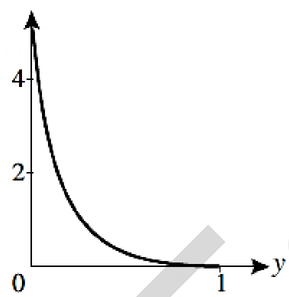


Fig. 6

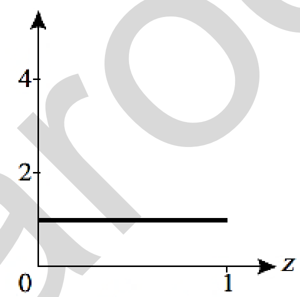


Fig. 7

Each of the random variables T, U, V, W, X, Y and Z takes values between 0 and 1 only. Their probability density functions are shown in Figs 1 to 7 respectively.

(i) **(a)** Which of these variables has the largest median? [1]

(b) Which of these variables has the largest standard deviation? Explain your answer. [2]

(ii) Use Fig. 2 to find $P(U < 0.5)$. [2]

(iii) The probability density function of X is given by

$$f(x) = \begin{cases} ax^n & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

where a and n are positive constants.

(a) Show that $a = n + 1$. [3]

(b) Given that $E(X) = \frac{5}{6}$, find a and n. [4]

Answer: (i) (a) X or 5; (b) V or 3, Higher and lower values more likely; (ii) 0.75; (iii) (b) a=5, n=4 **N11/71/Q7**

16 The random variable X has probability density function given by

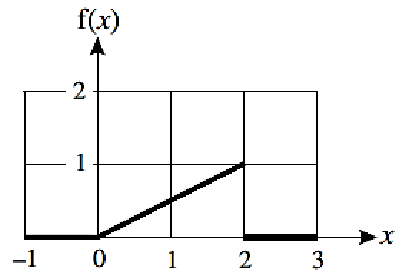
$$f(x) = \begin{cases} ke^{-x} & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that $k = \dfrac{e}{e - 1}$. [3]

(ii) Find $E(X)$ in terms of e. [4]

Answer: (ii) (e-2)/(e-1) **N11/73/Q4**

17



The diagram shows the graph of the probability density function, f, of a random variable X. Find the median of X. [3]

Answer: $m = \sqrt{2}$. **N12/71/Q1**

18 A random variable X has probability density function given by

$$f(x) = \begin{cases} \dfrac{k}{x-1} & 3 \leqslant x \leqslant 5, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \dfrac{1}{\ln 2}$. [4]

(ii) Find a such that $P(X < a) = 0.75$. [4]

Answer: **(ii)** 4.36. **N12/71/Q5**

19 Darts are thrown at random at a circular board. The darts hit the board at distances X centimetres from the centre, where X is a random variable with probability density function given by

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

where a is a positive constant.

(i) Verify that f is a probability density function whatever the value of a. [3]

It is now given that $E(X) = 8$.

(ii) Find the value of a. [3]

(iii) Find the probability that a dart lands more than 6 cm from the centre of the board. [3]

N12/73/Q6

Answers: **(ii)** $a = 12$; **(iii)** $\frac{3}{4}$.

20 The volume, in cm^3, of liquid left in a glass by people when they have finished drinking all they want is modelled by the random variable X with probability density function given by

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

where k is a constant.

(i) Show that $k = \frac{3}{8}$. [2]

(ii) 20% of people leave at least d cm^3 of liquid in a glass. Find d. [3]

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

Answer. 0.830
1/2

N13/71/Q5

21 The waiting time, T weeks, for a particular operation at a hospital has probability density function given by

$$f(t) = \begin{cases} \frac{1}{2500}(100t - t^3) & 0 \leqslant t \leqslant 10, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Given that $E(T) = \frac{16}{3}$, find $\text{Var}(T)$. [3]

(ii) 10% of patients have to wait more than n weeks for their operation. Find the value of n, giving your answer correct to the nearest integer. [5]

Answers: (i) 44/9 or 4.89, (ii) 8

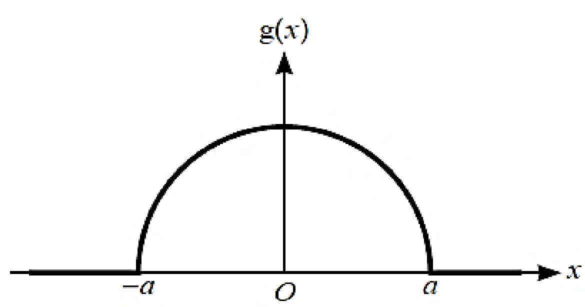
N13/73/Q3

22 **(a)** The time for which Lucy has to wait at a certain traffic light each day is T minutes, where T has probability density function given by

$$f(t) = \begin{cases} \frac{3}{2}t - \frac{3}{4}t^2 & 0 \leqslant t \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

Find the probability that, on a randomly chosen day, Lucy has to wait for less than half a minute at the traffic light. [3]

(b)



The diagram shows the graph of the probability density function, g, of a random variable X. The graph of g is a semicircle with centre (0, 0) and radius a. Elsewhere $g(x) = 0$.

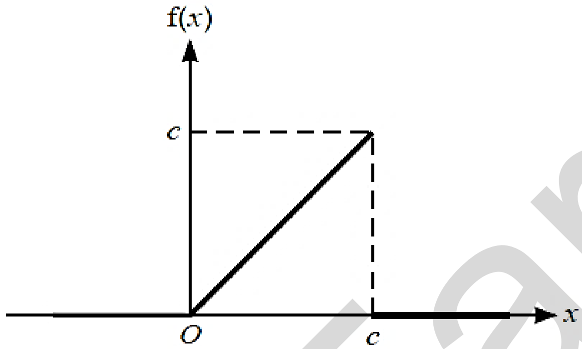
(i) Find the value of a. [2]

(ii) State the value of $E(X)$. [1]

(iii) Given that $P(X < -c) = 0.2$, find $P(X < c)$. [2]

Answer: 0.156, 0.798, 0, 0.8 **N14/71/Q3**

23



The diagram shows the graph of the probability density function, f, of a random variable X.

(i) Find the value of the constant c. [2]

(ii) Find the value of a such that $P(a < X < 1) = 0.1$. [4]

(iii) Find $E(X)$. [2]

Answers: (i) $c=\sqrt{2}$ or $c=1.41$, (ii) 0.894, (iii) 0.943 or $\frac{2\sqrt{2}}{3}$ **N14/73/Q2**

24 A random variable X has probability density function given by

$$f(x) = \begin{cases} k(3-x) & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{2}{3}$. [3]

(ii) Find the median of X. [4]

Answer: 1.42 **N15/71/Q4**

25 A random variable X has probability density function given by

$$f(x) = \begin{cases} k(4-x^2) & -2 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{3}{32}$. [3]

(ii) Sketch the graph of $y = f(x)$ and hence write down the value of $E(X)$. [2]

(iii) Find $P(X < 1)$. [3]

Answers: (i) 3/32 , (ii) inverted parabola between -2 and 2 only E(X) = 0, (iii) 27/32 or 0.844 **N15/73/Q4**

26 In each turn of a game, a coin is pushed and slides across a table. The distance, X metres, travelled by the coin has probability density function given by

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

where k is a constant.

(i) State the greatest possible distance travelled by the coin in one turn. [1]

(ii) Show that $k = \frac{3}{4}$. [3]

(iii) Find the mean distance travelled by the coin in one turn. [3]

(iv) Out of 400 turns, find the expected number of turns in which the distance travelled by the coin is less than 1 metre. [3]

Answer: 2m **J16/71/Q6**
1.2m.
125

27 The time, T minutes, taken by people to complete a test has probability density function given by

$$f(t) = \begin{cases} k(10t - t^2) & 5 \leqslant t \leqslant 10, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{3}{250}$. [3]

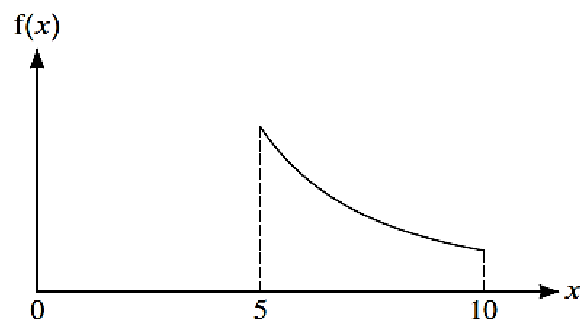
(ii) Find $E(T)$. [3]

(iii) Find the probability that a randomly chosen value of T lies between $E(T)$ and the median of T. [3]

(iv) State the greatest possible length of time taken to complete the test. [1]

Answers: (i) k =3/250 (ii) 6.875 (iii) 0.0361 (iv) 10 (minutes) **J16/73/Q5**

28



The time, X minutes, taken by a large number of runners to complete a certain race has probability density function f given by

$$f(x) = \begin{cases} \dfrac{k}{x^2} & 5 \leqslant x \leqslant 10, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant, as shown in the diagram.

(i) Without calculation, explain how you can tell that there were more runners whose times were below 7.5 minutes than above 7.5 minutes. [1]

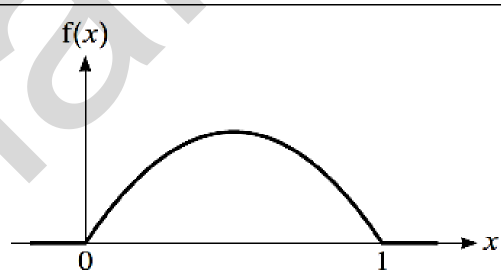
(ii) Show that $k = 10$. [3]

(iii) Find $\mathrm{E}(X)$. [3]

(iv) Find $\mathrm{Var}(X)$. [2]

Answers: (i) Greater area where $x < 7.5$ than $x > 7.5$ (iii) 6.93 (iv) 1.95 **J17/71/Q4**

29



The diagram shows the graph of the probability density function, f, of a continuous random variable X, where f is defined by

$$f(x) = \begin{cases} k(x - x^2) & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that the value of the constant k is 6. [3]

(ii) State the value of $E(X)$ and find $\text{Var}(X)$. [4]

(iii) Find $P(0.4 < X < 2)$. [3]

Answers: **(i)** $k = 6$ **(ii)** $E[X] = 0.5$, $\text{Var}[X] = 0.05$ **(iii)** 0.648 **J17/73/Q6**

---

30 The time, in minutes, taken by people to complete a test is modelled by the continuous random variable X with probability density function given by

$$f(x) = \begin{cases} \dfrac{k}{x^2} & 5 \leqslant x \leqslant 10, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = 10$. [3]

(ii) Show that $E(X) = 10\ln 2$. [2]

(iii) Find $P(X > 9)$. [3]

(iv) Given that $P(X < a) = 0.6$, find a. [3]

Answers: **(iii)** 0.111 **(iv)** 7.14 **J18/71/Q6**

---

31 A random variable X has probability density function defined by

$$f(x) = \begin{cases} k\left(\dfrac{1}{x^2} + \dfrac{1}{x^3}\right) & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \frac{8}{7}$. [3]

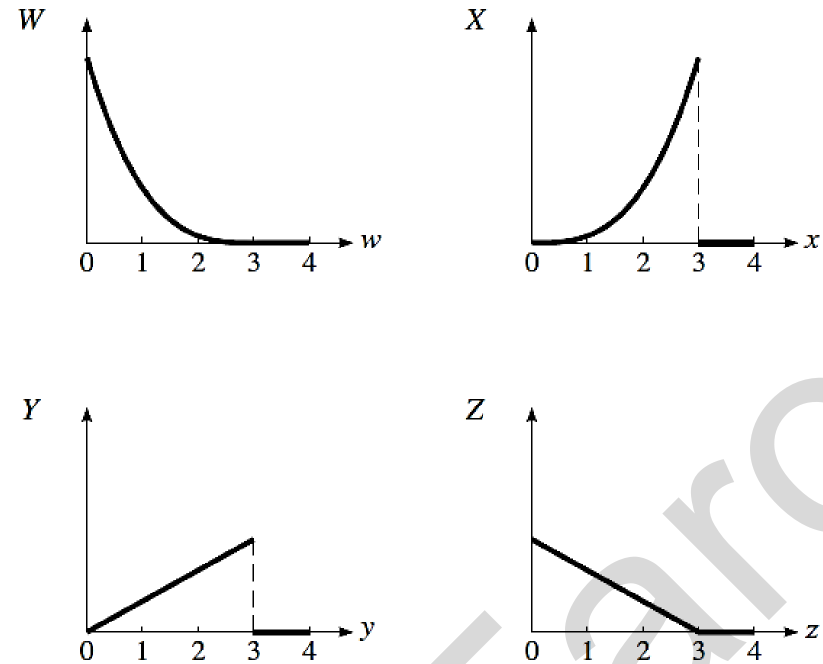
(ii) Find $E(X)$. [3]

(iii) Three values of X are chosen at random. Find the probability that one of these values is less than 1.5 and the other two are greater than 1.5. [5]

J18/73/Q7

Answers: **(i)** $k=\dfrac{8}{7}$ **(ii)** 1.36 **(iii)** 0.191

32



The diagrams show the probability density functions of four random variables W, X, Y and Z. Each of the four variables takes values between 0 and 3 only, and their medians are m_W, m_X, m_Y and m_Z respectively.

(i) List m_W, m_X, m_Y and m_Z in order of size, starting with the largest. [2]

(ii) The probability density function of X is given by

$$f(x) = \begin{cases} \frac{4}{81}x^3 & 0 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that $E(X) = \frac{12}{5}$. [3]

(b) Calculate $P(X > E(X))$. [3]

(c) Write down the value of $P(X < 2E(X))$. [1]

N16/71/Q

33 A continuous random variable, X, has probability density function given by

$$f(x) = \begin{cases} \frac{1}{4}(x+1) & 0 \leqslant x \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find $E(X)$. [3]

(ii) Find the median of X. [3]

Answers: (i) 7/6 or 1.17, (ii) 1.24 or $\sqrt{5}-1$ **N17/71/Q5**

34 The random variable X has probability density function given by

$$f(x) = \begin{cases} kx^{-1} & 2 \leqslant x \leqslant 6, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \dfrac{1}{\ln 3}$. [2]

(ii) Show that $E(X) = 3.64$, correct to 3 significant figures. [3]

9

(iii) Given that the median of X is m, find $P(m < X < E(X))$. [4]

Answers: (i) $k = \dfrac{1}{\ln 3}$ (ii) 3.64 (iii) 0.045 (2 sig. figs.)

N18/71/Q6

Chapter 4: Sampling and Estimation

SAMPLING

Population

In a statistical enquiry you often need information about a particular group. This group is known as the **population** or the **target population**, and it could be small, large or even infinite. Note that the word 'population' does not necessarily mean 'people'.
Here are some examples of populations:

- pupils in a class,
- people in England in full time employment,
- hospitals in Wales,
- cans of soft drink produced in a factory,
- ferns in a wood,
- rational numbers between 0 and 10.

SURVEYS

Information is collected by means of a survey. There are two types:

(a) a census,
(b) a sample survey.

Census

In a census every member of the population is surveyed.

When the population is small, this could be a straightforward exercise. For example, it would be easy to find out how each pupil in a class travelled to school on a particular morning. When populations are large, taking a census can be very time consuming and difficult to do with accuracy. Each year the government carries out a census in schools on the third Thursday in January. This requests the number of boys and girls in each age group on the roll of every school in the country. Its accuracy, though, relates only to that day. Even more difficult to carry out accurately is the population census taken every ten years. This attempts to provide details of different age groups for every area in Britain. When populations are very large, or infinite, it is not possible to survey every member.

On occasions it would not be sensible to survey every member. For example, if you performed a census to establish the length of life of a particular brand of light bulb, you would test each bulb until it failed and so you would destroy the population!

Sample Survey

When a survey covers less than 100% of the population, it is known as a **sample survey**. In many circumstances, taking a sample is preferable to carrying out a census. Sample data can be obtained relatively cheaply and quickly and, if the sample is representative of the population, a sample survey can give an accurate indication of the population characteristic being studied.

The size of the sample does not depend on the size of the population. It often depends on the time and money available to collect information. Note that large samples are more likely to give more reliable information than small ones. The next time that you read the results of a public opinion poll in the newspaper, look at the size of the sample – it is usually over 1000.

SAMPLE DESIGN

Once the purpose of a survey has been stated precisely, the **target population** must be defined, for example

- all the primary schools in England,
- all the oak trees in Hampshire,
- all the people admitted to the General Hospital in January suffering from a heart attack.

The **sampling units** must be defined clearly. These are the people or items to be sampled, for example

- the primary school,
- the oak tree,
- the person suffering from a heart attack.

Once the sampling units within a population are individually named or numbered to form a list, then this list of sampling units is called a **sampling frame**. It could take various forms (e.g. a list, a map, a set of maps), and should be as accurate as possible.

Ideally the sampling frame should be the same as the target population. For example, if the target population is all the first year students in a college, then the sampling frame and the target population should be the same, provided that the register is up-to-date and accurate. A sampling frame for people in Britain eligible to vote, however, is more difficult to form. The electoral register attempts to list all those who are eligible to vote throughout all the areas in the country, but it is never completely accurate, since many changes occur during the time that the information is being processed. Some people do not return the forms, people move in and out of the area, people die etc.

In some instances it is not possible to enumerate all the population, for example, the fish in a lake.

Example 1

(a) Explain briefly what you understand by
  (i) a population,
  (ii) a sampling frame.

(b) A market research organisation wants to take a sample of
  (i) owners of diesel motor cars in the UK,
  (ii) persons living in Oxford who suffered from injuries to the back during July 1996.

Suggest a suitable sampling frame in each case. (*L*)

BIAS

The purpose of sampling is to gain information about the whole population by selecting a sample from that population. You want the sample to be representative of the population so you must give every member of the population an equal chance of being included in the sample. This should eliminate any bias in the selection of the sample.

Sources of bias include

(a) the lack of a good sampling frame:
  – using the telephone directory misses all those who do not have a telephone or whose number is ex-directory,
  – using the electoral register in a city area misses the more mobile section of the population.

(b) the wrong choice of sampling unit:
  – choosing an individual rather than a particular group such as 'household'.

(c) non-response by some of the chosen units:
  – it might be difficult to locate the particular unit,
  – the cooperation of the respondent might not have been obtained,
  – the enquiry might not have been understood, for example, a questionnaire might have been badly designed. Questionnaires should be clear, specific, unambiguous and easily understood. Questions should be worded neutrally, especially in opinion surveys, to avoid bias caused by pointing towards a particular response.

(d) bias introduced by the person conducting the survey:
  – the interviewer might not question someone who appears uncooperative,
  – the style of questioning may influence the response.

It should be noted that a sample can only be representative of the population from which it is selected. If you select a sample of teachers from one school, the sample is representative of the teachers in that school, not of all teachers in all schools.

SAMPLING METHODS

Once a sampling frame has been established, you can choose a method of sampling. These fall into two categories:

- random sampling e.g. simple, systematic, stratified;
- non-random sampling e.g. quota, cluster

Simple Random Sampling

Suppose a population consists of N sampling units and you require a sample of n of these units. A sample of size n is called a **simple random sample** if all possible samples of size n are equally likely to be selected. Some form of random processes must be used to make the selection.

If the unit selected at each draw is **replaced** into the population before the next draw, then it can appear more than once in the sample. This is known as **sampling with replacement.**

If the unit selected at each draw is **not replaced** into the population before the next draw, this is known as **sampling without replacement.**

The second method of sampling without replacement is known as **simple random sampling.**

Two methods of simple random sampling are commonly used

- drawing lots,
- random number sampling.

For each, make a list of all N members of the population and give each member a different number.

Drawing lots

For each member, place a coloured ball into a container and then draw n balls out of the container at random and without replacement. If you wanted a sample of size 20, you would draw out 20 balls. This is suitable for a small population. Note, however, that the sample must be large enough to provide sufficiently accurate information about the population.

The sample should be selected at random. Any hint of possible bias should be avoided.

If the population is large then the method of drawing lots, sometimes described as 'drawing out of a hat' is not practical. You could instead make the choice by referring to random number tables. For your reference, a set is printed on page 653.

Using Random Number Tables

Random number tables consist of lists of digits 0, 1, 2, 3, ..., 9, such that each digit has an equal chance of occurring, so for example, the probability that a 3 occurs is 0.1. In random number tables the digits may appear singly or be grouped in some way. This is solely for convenience of printing.

Example 2

Here is an extract from a set of random number tables

6 8 7 2 5 3 8 1 5 9
2 5 3 4 7 0 5 4 9 5
3 2 6 8 7 4 4 7 0 5

Use it to select a random sample of

(a) eight people from a group of 100 people,
(b) eight people from a group of 60.

Solution Example 2

(a) To select a group of eight people from a target population of 100 people, allocate a two-digit number to each person, for example allocate 01 to the first on the list, 02 to the second, ... up to 98, 99, 00, calling the hundredth person 00 for convenience.

Using the list, starting at the beginning of the first row and reading along the rows, you would select people corresponding to the following numbers:

68 72 53 81 59 25 34 70

Alternatively, you could decide to read the digits backwards, from bottom right, in which case your sample would consist of people corresponding to the numbers

50 74 47 86 23 59 45 07

(b) To select a group of eight from a target population of 60 people, allocate each person a number from 01 to 60.

Using the tables, disregard any two-digit number outside the range.

Example 3

Use the following extract from random number tables to select a random sample of 12 numbers, each to two decimal places, from the continuous range $0 \leqslant x < 10$.

52	74	54	80	68	72	51	96	08	00
02	52	09	93	60	43	57	42	13	44

Solution 9.3

Since the sample values are required to two decimal place accuracy, consider groups of three digits, inserting the decimal point between the first and second digit.

In this case your sample would consist of the values

5.27, 4.54, 8.06, 8.72, 5.19, 6.08, 0.00, 2.52, 0.99, 3.60, 4.35, 7.42

Example 4

Here is a set of random numbers

848051 386103 153842 242330 580007 479971

Use it to select a random sample of four numbers, each to three decimal places, from the continuous range $0 \leqslant x < 5$.

Solution 9.4

Consider groups of four digits, inserting the decimal point between the first and second digit. Disregard any values that are out of range. This gives

~~8.480~~ ~~5.138~~ ~~6.103~~ 1.538 4.224 2.330 ~~5.800~~ 0.747

So the numbers chosen are 1.538, 4.224, 2.330, 0.747.

Systematic Sampling

Random sampling from a *very* large population is very cumbersome.
An alternative procedure is to list the population in some order, for example alphabetically or in order of completion on a production line, and then choose every kth member from the list after obtaining a random starting point. If you choose every tenth member from the list, for example every tenth vehicle passing a checkpoint, you would form a 10% sample. If you choose every twentieth item, for example every twentieth card in an index file, you would form a 5% sample.

Example 5

Describe how to choose a systematic sample of eight members from a list of 300.

Solution 9.5

Since you are going to choose every kth member, you need to find a suitable value for k. To do this, choose a convenient value close to $\frac{N}{n}$.

In this case, $\frac{N}{n} = \frac{300}{8} = 37.5$, so $k = 40$ will do.

Now choose a random starting point, for example if [Ran#] on your calculator gives 0.870 take the first member of the sample as 87 and then add 40 each time . The other members are 127, 167, 207, 247, 287, 27 and 67. Note that when you reach the end of the list, go back to the beginning.

So the sample consists of 27, 67, 87, 127, 167 207, 247, 287.

The advantages of systematic sampling are that it is quick to carry out and it is easy to check for errors. For large scale sampling, systematic selection is usually used in preference to taking simple random samples.

The disadvantage of this system is that there may be a periodic cycle within the frame itself. For example a machine may operate in such a manner that every tenth item is faulty. Systematic sampling of every fifth item, starting at 5, would result in half the items in the sample being faulty, whereas starting at 2 would produce a sample with no faulty items. Of course, if the periodic cycle is recognised then different samples could be taken by varying the starting points and the length of the interval between the chosen items.

Stratified Sampling

Stratified sampling is used when the population is split into distinguishable layers or strata that are quite different from each other and which together cover the whole population, for example

- age groups,
- occupational groups,
- topographical regions.

Separate random samples are then taken from each stratum and put together to form the sample from the population.

It is usual to represent the population proportionately in the strata, as in the following example.

Example 6

Competent Carriers employs 320 drivers, 80 administrative staff and 40 mechanics. A committee to represent all the employees is to be formed. The committee is to have 11 members and the selection is to be made so that there is as close a representation as possible without bias towards any individuals or groups. Explain how this could be done.

Solution 9.6

If you were to take a simple random sample of all 440 employees this would mean that every employee would have an equal chance of being selected. There is a high probability that the committee would consist of 11 drivers and therefore would not be representative of all employees.

A stratified random sample would provide a more accurate representation of the population and could be formed as follows:

Taking into account that drivers make up $\frac{320}{440}$ of the work force,

$$\text{number of drivers} = \frac{320}{440} \times 11 = 8$$

Similarly

$$\text{number of administrative staff} = \frac{80}{440} \times 11 = 2$$

$$\text{number of mechanics} = \frac{40}{440} \times 11 = 1$$

The required representation on the committee is eight drivers, two from the administrative staff and one mechanic. The people to be included can then be selected from each stratum by using simple random sampling or systematic sampling.

Non-Random Sampling

(a) Cluster sampling

Sometimes there is a natural sub-grouping of the population and these subgroups are called **clusters**. For example, in a population consisting of all children in the country attending state primary schools, the local education authorities form natural clusters. When a sample survey is carried out on a population that can be broken into clusters it is often more convenient to first choose a random sample of clusters and then to sample within each cluster chosen.

Unlike *stratified sampling* where the strata are as *different* from each other as possible, each *cluster* should be as *similar* to other clusters as possible.

One advantage of cluster sampling is that there is no need to have a complete sampling frame of the whole population. For the primary school children, you would need only a list of the pupils in the chosen local authority. Another advantage is that it is usually far less costly than random sampling. Consider the fees and travelling expenses paid to interviewers. Far less travelling and time is involved in an interviewer visiting individuals in a cluster than visiting individuals in the whole population.

The disadvantage of cluster sampling is that it is non-random. Suppose that a town has 7500 primary school children in 250 classes, each with an average class size of 30. If you want to select a sample of 90 children then you could use simple random sampling. It would however be quicker to use the classes as clusters and to take a sample consisting of three classes. This would give a sample of 90 children. The problem is that within each class there will be a certain amount of similarity between the children in say age, ability, home background. In selecting one whole class or cluster you are in fact selecting 30 similar children instead of 30 randomly chosen children from throughout the town. Therefore three clusters will not give as precise a picture of the whole population as 90 children chosen at random from throughout the town.

(b) Quota sampling

Quota sampling is widely used in market research where the population is divided into groups in terms of age, sex, income level and so on. Then the interviewer is told how many people to interview within each specified group, but is given no specific instructions about how to locate them and fulfil the quota. This is the method generally used in street interview surveys commonly carried out in shopping centres. It is quick to use, complications are kept to a minimum and, unlike random sampling, any member of the sample may be replaced by another member with the same characteristics.

If no sampling frame exists, then quota sampling may be the only practical method of obtaining a sample. The disadvantage of quota sampling, however, is that it is non-random. There is a possibility of bias in the selection process if, for example, the interviewer selects those easiest to question or those who look cooperative. The location of such surveys in shopping centres excludes a substantial part of the population in that area. It is difficult to find out about those who refuse to cooperate and they are simply replaced. One of the reasons put forward to explain the inaccuracy of the opinion polls before the British general election in 1992 was the high refusal rates of Conservative voters to take part in surveys.

SAMPLE STATISTICS

When you are trying to find out information about a population it seems sensible to take random samples and then consider the values obtained from them. It is therefore useful to know how these sample values are distributed.

THE DISTRIBUTION OF THE SAMPLE MEAN

Imagine carrying out the following procedure:

- Take a random sample of n independent observations from a population. Note that from a finite population, sampling should be with replacement to ensure that the observations are independent.
- Calculate the mean of these n sample values. This is known as the sample mean.
- Now repeat the procedure until you have taken all possible samples of size n, calculating the sample mean of each one.
- Form a distribution of all the sample means.

The distribution that would be formed is called the **sampling distribution of means.**

$$E(\bar{X}) = \mu \quad \text{and} \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

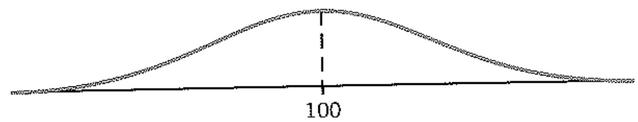
The standard deviation of the sampling distribution is $\sqrt{\frac{\sigma^2}{n}}$, usually written $\frac{\sigma}{\sqrt{n}}$. This is known as the **standard error of the mean.**

The mean of the sampling distribution is the same as the mean of the population.
The standard deviation of the sampling distribution is much smaller than that of the population since σ^2 has been divided by n. This implies that the sample means are much more clustered around μ than the population values are. In fact, the larger the sample size, the more clustered they are.

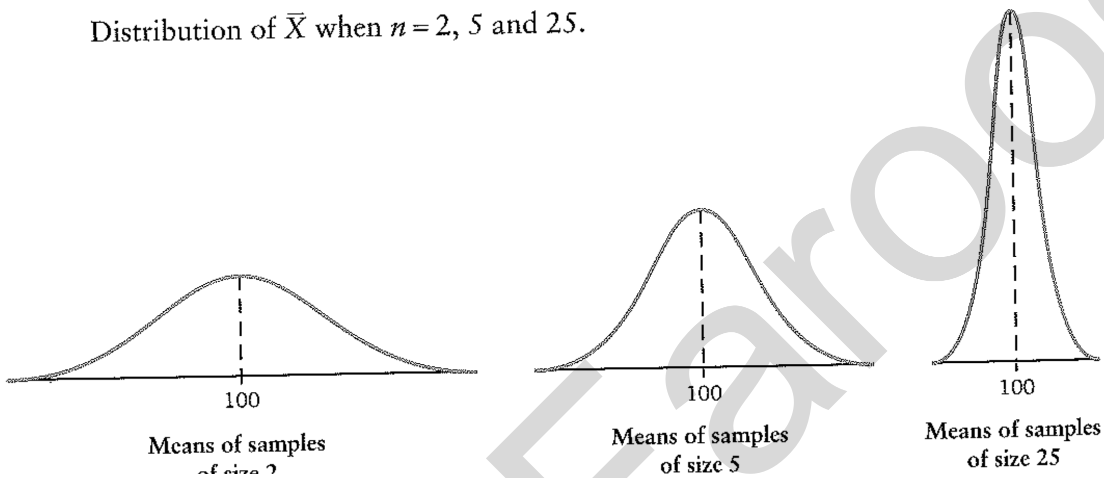
The following diagrams help to illustrate the shape of the sampling distribution of means resulting from different sized samples from given populations.

THE DISTRIBUTION OF $\bar{X}$ WHEN THE POPULATION OF X IS NORMAL

Distriubtion of X when $X \sim N(100, 64)$



Distribution of $\bar{X}$ when $n = 2$, 5 and 25.



Example 7

At a college the masses of the male students can be modelled by a normal distribution with mean mass 70 kg and standard deviation 5 kg.
Four male students are chosen at random. Find the probability that their mean mass is less than 65 kg.

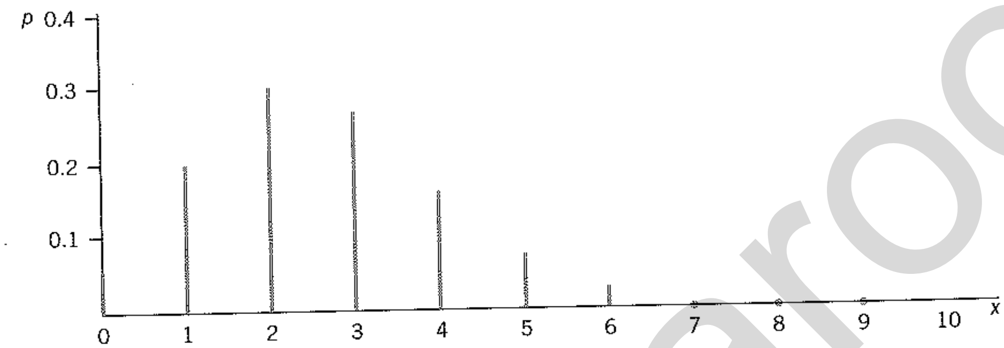
Example 8

The distribution of the random variable X is $N(25, 340)$. The mean of a random sample of size n drawn from this distribution is $\bar{X}$. Find the value of n, correct to two significant figures, given that $P(\bar{X} > 28)$ is approximately 0.005. *(C)*

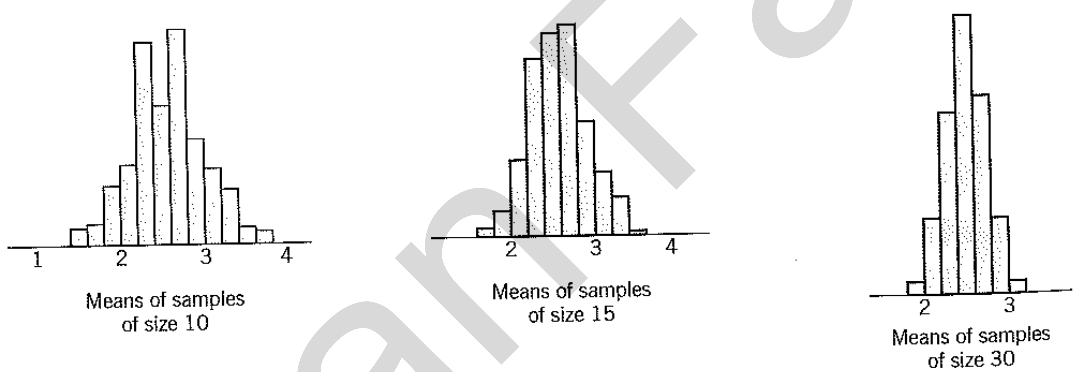
THE DISTRIBUTION OF $\bar{X}$ WHEN THE POPULATION OF X IS NOT NORMALLY DISTRIBUTED

The following diagrams illustrate the distribution of $\overline{X}$ for samples of different sizes taken from a population X:

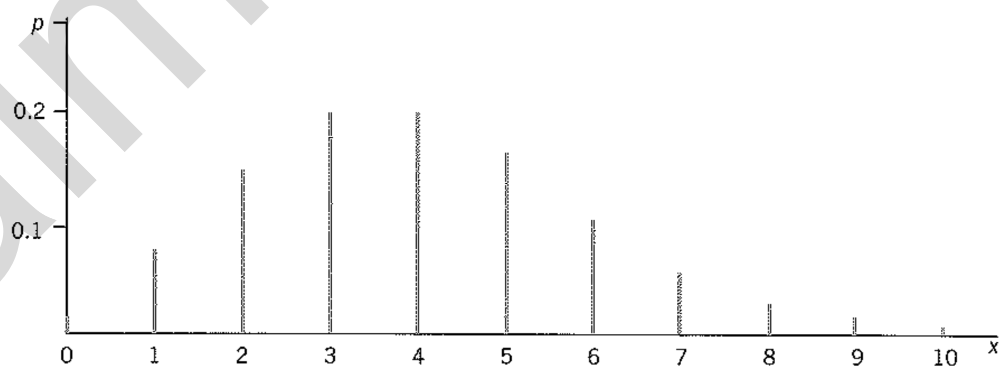
(i) Distribution of X when $X \sim B\,(10, 0.25)$



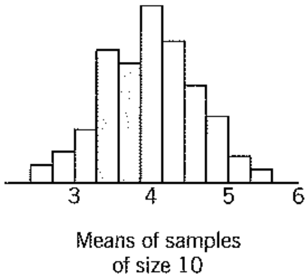
Distribution of $\overline{X}$ for samples of size 10, 15 and 30

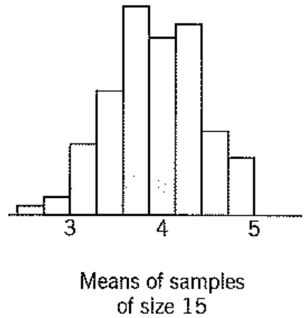


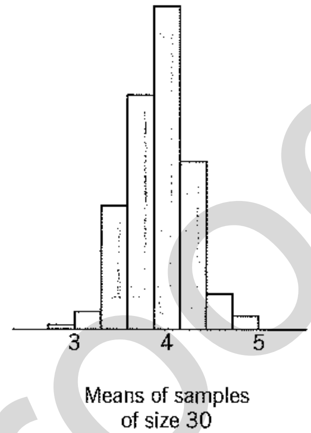
(ii) Distribution of X when $X \sim \text{Po}\,(4)$



Distribution of $\overline{X}$ for samples of size 10, 15 and 30







CENTRAL LIMIT THEOREM

From the diagrams you can see that when samples are taken from a population that is **not normally distributed**, the sampling distribution takes on the characteristic normal shape as the sample size increases. **For large *n*** the distribution of the sample mean is **approximately normal.**

This result is known as the **central limit theorem.** It is somewhat surprising, since it holds when the population of X is discrete (as in the binomial and Poisson distributions) and when X is continuous (as in the uniform distribution).

For samples taken from a non-normal population with mean μ and variance σ^2, by the central limit theorem, $\overline{X}$ is approximately normal

and $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

provided that the sample size, n, is large ($n \geqslant 30$ say).

Example 9

Thirty random observations are taken from each of the following distributions and the sample mean calculated. Find, in each case, the probability that the sample mean exceeds 5.

(a) X is the number of telephone calls made in an evening to a counselling service, where $X \sim Po(4.5)$.

(b) X is the number of heads obtained when an unbiased coin is tossed nine times.

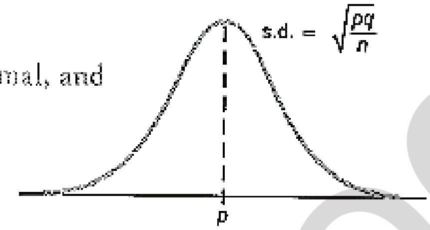
THE DISTRIBUTION OF THE SAMPLE PROPORTION

The distribution of P_s has mean p and variance $\frac{pq}{n}$.

When n is large, the distribution of P_s is approximately normal, and

$$P_s \sim N\left(p, \frac{pq}{n}\right)$$

The larger the sample size, n, the better the approximation.



The distribution of P_s is known as the **sampling distribution of proportions**. The standard deviation of this distribution is $\sqrt{\frac{pq}{n}}$ and it is known as the **standard error of proportion**.

NOTE: When considering the normal approximation to the binomial distribution, a continuity correction of $\pm\frac{1}{2}$ is needed (see page 383).

Since $P_s = \frac{1}{n} \times X$, use a continuity correction $\frac{1}{n} \times \left(\pm \frac{1}{2}\right)$ i.e. $\pm \frac{1}{2n}$.

Example 10

It is known that 3% of frozen pies delivered to a canteen are broken. What is the probability that, on a morning when 500 pies are delivered, 5% or more are broken?

UNBIASED ESTIMATES OF POPULATION PARAMETERS

In order to define a binomial distribution you need to know n and p; to define a Poisson distribution you need to know λ and to define a normal distribution you need to know μ and σ^2. These are known as the **population parameters** of the distributions.

Suppose that you do not know the value of a particular parameter of a distribution, for example the mean or the variance or the proportion of successes. It seems sensible that you would take a random sample from the distribution and use it in some way to make an estimate of the value of your unknown parameter.

This estimate is **unbiased** if the average (or expectation) of a large number of values taken in the same way is the true value of the parameter. There may be several ways of obtaining an unbiased estimate but the best (most efficient) estimate is the one with the smallest variance.

POINT ESTIMATES

If the random sample taken is of size n,

- the best unbiased estimate of p, the proportion of successes in the population, is $\hat{p}$ where
  $\hat{p} = p_s$ — p_s is the proportion of successes in the sample
- the best unbiased estimate of μ, the population mean, is $\hat{\mu}$ where
  $\hat{\mu} = \bar{x} = \dfrac{\sum x}{n}$ — $\bar{x}$ is the mean of the sample
- the best unbiased estimate of σ^2, the population variance, is $\hat{\sigma}^2$ where
  $\hat{\sigma}^2 = \dfrac{n}{n-1} \times s^2$ — s^2 is the variance of the sample

Example 11

A railway enthusiast simulates train journeys and records the number of minutes, x, to the nearest minute, trains are late according to the schedule being used. A random sample of 50 journeys gave the following times.

17	5	3	10	4	3	10	5	2	14
3	14	5	5	21	9	22	36	14	34
22	4	23	6	8	15	41	23	13	7
6	13	33	8	5	34	26	17	8	43
24	14	23	4	19	5	23	13	12	10

Given that $\sum x = 738$ and $\sum x^2 = 16\,526$, calculate to two decimal places, unbiased estimates of the mean and the variance of the population from which this sample was drawn. (*L*)

Example 12

For the data given in Example 9.17, estimate the proportion of trains that are more than 25 minutes late.

INTERVAL ESTIMATES

(a) Confidence interval for μ, the population mean

- **of a normal population,**
- **with known variance σ^2**
- **using any size sample, n large or small**

If $\bar{x}$ is the mean of a random sample of any size n taken from a normal population with known variance σ^2,
then a 95% confidence interval for μ is given by

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

Example 13

The mass of vitamin E in a capsule manufactured by a certain drug company is normally distributed with standard deviation 0.042 mg. A random sample of five capsules was analysed and the mean mass of vitamin E was found to be 5.12 mg. Calculate a symmetric 95% confidence interval for the population mean mass of vitamin E per capsule. Give the values of the end-points of the interval correct to three significant figures. (C)

Summary

A symmetric 90% confidence interval for μ is

$$\left(\bar{x} - 1.645\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 1.645\frac{\sigma}{\sqrt{n}}\right)$$

A symmetric 95% confidence interval for μ is

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

A symmetric 99% confidence interval for μ is

$$\left(\bar{x} - 2.576\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 2.576\frac{\sigma}{\sqrt{n}}\right)$$

(b) Confidence interval for μ, the population mean

- of a non-normal population,
- with a known variance σ^2
- using a large sample, $n \geqslant 30$ say

In this case, since the sample size is large, the **central limit theorem** can be used.

$\overline{X}$ is approximately normal and $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ (see page 442).

Example 14

The heights of men in a particular district are distributed with mean μ cm and the standard deviation σ cm.
On the basis of the results obtained from a random sample of 100 men from the district, the 95% confidence interval for μ was calculated and found to be (177.22 cm, 179.18 cm).

Calculate
(a) the value of the sample mean,
(b) the value of σ,
(c) a symmetric 90% confidence interval for μ.

Example 15

A plant produces steel sheets whose weights are known to be normally distributed with a standard deviation of 2.4 kg. A random sample of 36 sheets had a mean weight of 31.4 kg. Find 99% confidence limits for the population mean. *(L)*

Example 16

The result X of a stress test is known to be a normally distributed random variable with mean μ and standard deviation 1.3. It is required to have a 95% symmetrical confidence interval for μ with total width less than 2. Find the least number of tests that should be carried out to achieve this. *(L)*

(c) Confidence interval for μ, the population mean

- of a normal or non-normal population,
- with unknown variance σ^2
- using a large sample, n

Example 17

The fuel consumption of a new model of car is being tested. In one trial, 50 cars chosen at random, were driven under identical conditions and the distances, x km, covered on 1 litre of petrol were recorded. The results gave the following totals:

$\Sigma x = 525, \quad \Sigma x^2 = 5625.$

Calculate a 95% confidence interval for the mean petrol consumption, in kilometres per litre, of cars of this type.

Example 18

The height, x cm, of each man in a random sample of 200 men living in the UK was measured. The following results were obtained:

$\Sigma x = 35\,050, \Sigma x^2 = 6\,163\,109.$

(a) Calculate unbiased estimates of the mean and variance of the heights of men living in the UK.

(b) Determine an approximate 90% confidence interval for the mean height of men living in the UK. Name the theorem that you have assumed. *(NEAB)*

CONFIDENCE INTERVALS FOR THE POPULATION PROPORTION

You are then able to find *approximate* confidence intervals for p as follows:

	Confidence limits	Confidence interval	Width
90%	$p_s \pm 1.645\sqrt{\frac{p_s q_s}{n}}$	$\left(p_s - 1.645\sqrt{\frac{p_s q_s}{n}}, p_s + 1.645\sqrt{\frac{p_s q_s}{n}}\right)$	$2 \times 1.645\sqrt{\frac{p_s q_s}{n}}$
95%	$p_s \pm 1.96\sqrt{\frac{p_s q_s}{n}}$	$\left(p_s - 1.96\sqrt{\frac{p_s q_s}{n}}, p_s + 1.96\sqrt{\frac{p_s q_s}{n}}\right)$	$2 \times 1.96\sqrt{\frac{p_s q_s}{n}}$
99%	$p_s \pm 2.576\sqrt{\frac{p_s q_s}{n}}$	$\left(p_s - 2.576\sqrt{\frac{p_s q_s}{n}}, p_s + 1.96\sqrt{\frac{p_s q_s}{n}}\right)$	$2 \times 2.576\sqrt{\frac{p_s q_s}{n}}$

Remember that the sample size, n, should be large ($n \geqslant 30$ say), since the normal approximation to the binomial distribution is used in obtaining the distribution of sample proportions. Also, since a continuous distribution has been used as an approximation for a discrete distribution, continuity corrections should be used. These are usually omitted, however, when calculating confidence intervals.

Example 19

A manufacturer wants to assess the proportion of defective items in a large batch produced by a particular machine. He tests a random sample of 300 items and finds that 45 items are defective.

(a) Calculate an approximate 95% confidence interval for the proportion of defective items in the batch.

(b) If 200 such tests are performed and a 95% confidence interval calculated for each, how many would you expect to include the proportion of defective items in the batch?

Example 20

In a random sample of 400 carpet shops, it was discovered that 136 of them sold carpets at below the list prices recommended by the manufacturer.

(a) Estimate the percentage of all carpet shops selling below list price.

(b) Calculate an approximate 90% confidence interval for the proportion of shops that sell below list price and explain briefly what this means.

(c) What size sample would have to be taken in order to estimate the percentage to within ±2%, with 90% confidence?

Example 21

The weights in grams of oranges grown in a certain area are normally distributed with mean μ and standard deviation σ. A random sample of 50 of these oranges was taken, and a 97% confidence interval for μ based on this sample was (222.1, 232.1).

(i) Calculate unbiased estimates of μ and σ^2.

(ii) Estimate the sample size that would be required in order for a 97% confidence interval for μ to have width 8.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q2 June 2009]

Example 22

(i) Give a reason why, in carrying out out a statistical investigation, a sample rather than a complete population may be used.

(ii) Rose wishes to investigate whether men in her town have a different life-span from the national average of 71.2 years. She looks at government records for her town and takes a random sample of the ages of 110 men who have died recently. Their mean age in years was 69.3 and the unbiased estimate of the population variance was 65.61.

Example 23

(a) Calculate a 90% confidence interval for the population mean and explain what you understand by this confidence interval.

(b) State with a reason what conclusion about the life-span of men in her town Rose could draw from this confidence interval.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q4 November 2005]

Example 24

Diameters of golf balls are known to be normally distributed with mean μ cm and standard deviation σ cm. A random sample of 130 golf balls was taken and the diameters, x cm, were measured. The results are summarised by $\Sigma x = 555.1$ and $\Sigma x^2 = 2371.30$.

(i) Calculate unbiased estimates of μ and σ^2.

(ii) Calculate a 97% confidence interval for μ.

(iii) 300 random samples of 130 golf balls are taken and a 97% confidence interval is calculated for each sample. How many of these intervals would you expect **not** to contain μ?

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q4 November 2008]

Example 24

A random sample of n people were questioned about their internet use. 87 of them had a high-speed internet connection. A confidence interval for the population proportion having a high-speed internet connection is $0.1129 < p < 0.1771$.

(i) Write down the mid-point of this confidence interval and hence find the value of n.

(ii) This interval is an $\alpha\%$ confidence interval. Find α.

[Cambridge International AS and A Level Mathematics 9709, Paper 71 Q2 June 2010]

Example 25

(i) Explain what is meant by the term 'random sample'.

In a random sample of 350 food shops it was found that 130 of them had Special Offers.

(ii) Calculate an approximate 95% confidence interval for the proportion of all food shops with Special Offers.

(iii) Estimate the size of a random sample required for an approximate 95% confidence interval for this proportion to have a width of 0.04.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q3 November 2007]

Example 26

A survey of a random sample of n people found that 61 of them read *The Reporter* newspaper. A symmetric confidence interval for the true population proportion, p, who read *The Reporter* is $0.1993 < p < 0.2887$.

(i) Find the mid-point of this confidence interval and use this to find the value of n.

(ii) Find the confidence level of this confidence interval.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q3 June 2005]

SELECTED PAST PAPER QUESTIONS

Question 1
N11/73/Q3

Jack has to choose a random sample of 8 people from the 750 members of a sports club.

(i) Explain fully how he can use random numbers to choose the sample. [3]

Jack asks each person in the sample how much they spent last week in the club café. The results, in dollars, were as follows.

15 25 30 8 12 18 27 25

(ii) Find unbiased estimates of the population mean and variance. [3]

(iii) Explain briefly what is meant by 'population' in this question. [1]

Question 2
J14/72/Q7

A researcher is investigating the actual lengths of time that patients spend with the doctor at their appointments. He plans to choose a sample of 12 appointments on a particular day.

(i) Which of the following methods is preferable, and why?

- Choose the first 12 appointments of the day.
- Choose 12 appointments evenly spaced throughout the day. [2]

Question 3
N13/72/Q3

Following a change in flight schedules, an airline pilot wished to test whether the mean distance that he flies in a week has changed. He noted the distances, x km, that he flew in 50 randomly chosen weeks and summarised the results as follows.

$$n = 50 \qquad \Sigma x = 143\,300 \qquad \Sigma x^2 = 410\,900\,000$$

(i) Calculate unbiased estimates of the population mean and variance. [3]

Question 4
N11/72/Q4

The volumes of juice in bottles of Apricola are normally distributed. In a random sample of 8 bottles, the volumes of juice, in millilitres, were found to be as follows.

332 334 330 328 331 332 329 333

(i) Find unbiased estimates of the population mean and variance. [3]

A random sample of 50 bottles of Apricola gave unbiased estimates of 331 millilitres and 4.20 millilitres^2 for the population mean and variance respectively.

(ii) Use this sample of size 50 to calculate a 98% confidence interval for the population mean. [3]

(iii) The manufacturer claims that the mean volume of juice in all bottles is 333 millilitres. State, with a reason, whether your answer to part (ii) supports this claim. [1]

Question 5
N14/72/Q5

Mahmoud throws a coin 400 times and finds that it shows heads 184 times. The probability that the coin shows heads on any throw is denoted by p.

(i) Calculate an approximate 95% confidence interval for p. [4]

(ii) Mahmoud claims that the coin is not fair. Use your answer to part (i) to comment on this claim. [1]

(iii) Mahmoud's result of 184 heads in 400 throws gives an α% confidence interval for p with width 0.1. Calculate the value of α. [4]

Homework: Sampling & Estimation– Variant 62

1 A consumer group, interested in the mean fat content of a particular type of sausage, takes a random sample of 20 sausages and sends them away to be analysed. The percentage of fat in each sausage is as follows.

26 27 28 28 28 29 29 30 30 31 32 32 32 33 33 34 34 34 35 35

Assume that the percentage of fat is normally distributed with mean μ, and that the standard deviation is known to be 3.

(i) Calculate a 98% confidence interval for the population mean percentage of fat. [4]

(ii) The manufacturer claims that the mean percentage of fat in sausages of this type is 30. Use your answer to part (i) to determine whether the consumer group should accept this claim. [2]

Answers: **(i)** (29.4, 32.6), 30% is inside the interval; **(ii)** Accept claim (at 2% level). **J03/Q3**

2 Machine A fills bags of fertiliser so that their weights follow a normal distribution with mean 20.05 kg and standard deviation 0.15 kg. Machine B fills bags of fertiliser so that their weights follow a normal distribution with mean 20.05 kg and standard deviation 0.27 kg.

(i) Find the probability that the total weight of a random sample of 20 bags filled by machine A is at least 2 kg more than the total weight of a random sample of 20 bags filled by machine B. [6]

(ii) A random sample of n bags filled by machine A is taken. The probability that the sample mean weight of the bags is greater than 20.07 kg is denoted by p. Find the value of n, given that $p = 0.0250$ correct to 4 decimal places. [4]

Answers: **(i)** 0.0738; **(ii)** 216. **J03/Q7**

3 Packets of cat food are filled by a machine.

(i) In a random sample of 10 packets, the weights, in grams, of the packets were as follows.

374.6 377.4 376.1 379.2 371.2 375.0 372.4 378.6 377.1 371.5

Find unbiased estimates of the population mean and variance. [3]

(ii) In a random sample of 200 packets, 38 were found to be underweight. Calculate a 96% confidence interval for the population proportion of underweight packets. [4]

Answers: **(i)** –2; **(ii)** 73.2. **J04/Q3**

4 Jenny has to do a statistics project at school on how much pocket money, in dollars, is received by students in her year group. She plans to take a sample of 7 students from her year group, which contains 122 students.

(i) Give a suitable method of taking this sample. [1]

Her sample gives the following results.

13.40 10.60 26.50 20.00 14.50 15.00 16.50

(ii) Find unbiased estimates of the population mean and variance. [3]

(iii) Is the estimated population variance more than, less than or the same as the sample variance? [1]

(iv) Describe what you understand by 'population' in this question. [1]

Answers: **(i)** Put names in a hat and draw out; **(ii)** 16.6, 27.1; **(iii)** More; **(iv)** Pocket money of all pupils in Jenny's year at school. **J05/Q2**

---

5 A survey of a random sample of n people found that 61 of them read *The Reporter* newspaper. A symmetric confidence interval for the true population proportion, p, who read *The Reporter* is $0.1993 < p < 0.2887$.

(i) Find the mid-point of this confidence interval and use this to find the value of n. [3]

(ii) Find the confidence level of this confidence interval. [4]

Answers: **(i)** 0.244, 250; **(ii)** 90%. **J05/Q3**

---

6 A clock contains 4 new batteries each of which gives a voltage which is normally distributed with mean 1.54 volts and standard deviation 0.05 volts. The voltages of the batteries are independent. The clock will only work if the total voltage is greater than 5.95 volts.

(i) Find the probability that the clock will work. [4]

(ii) Find the probability that the average total voltage of the batteries of 20 clocks chosen at random exceeds 6.2 volts. [3]

Answers: **(i)** 0.982; **(ii)** 0.0367. **J05/Q5**

---

7 Packets of fish food have weights that are distributed with standard deviation 2.3 g. A random sample of 200 packets is taken. The mean weight of this sample is found to be 99.2 g. Calculate a 99% confidence interval for the population mean weight. [3]

Answer: (98.8, 99.6). **J06/Q1**

---

8 Random samples of size 120 are taken from the distribution B(15, 0.4).

(i) Describe fully the distribution of the sample mean. [3]

(ii) Find the probability that the mean of a random sample of size 120 is greater than 6.1. [3]

Answers: **(i)** Normal with mean 6, variance 0.03; **(ii)** 0.282. **J06/Q3**

9 The random variable X has the distribution B(10, 0.15). Find the probability that the mean of a random sample of 50 observations of X is greater than 1.4. [5]

Answer: 0.713 or 0.714 with continuity correction, 0.734 without continuity correction. **J07/Q1**

10 The random variable X has the distribution $N(3.2, 1.2^2)$. The sum of 60 independent observations of X is denoted by S. Find $P(S > 200)$. [5]

Answer: 0.195 . **J07/Q2**

11 The daily takings, \$$x$, for a shop were noted on 30 randomly chosen days. The takings are summarised by $\Sigma x = 31\,500$, $\Sigma x^2 = 33\,141\,816$.

(i) Calculate unbiased estimates of the population mean and variance of the shop's daily takings. [3]

(ii) Calculate a 98% confidence interval for the mean daily takings. [3]

The mean daily takings for a random sample of n days is found.

(iii) Estimate the value of n for which it is approximately 95% certain that the sample mean does not differ from the population mean by more than \$6. [3]

Answers: (i) 1050, 2304; (ii) (1030, 1070); (iii) 246. **J07/Q6**

12 A magazine conducted a survey about the sleeping time of adults. A random sample of 12 adults was chosen from the adults travelling to work on a train.

(i) Give a reason why this is an unsatisfactory sample for the purposes of the survey. [1]

(ii) State a population for which this sample would be satisfactory. [1]

A satisfactory sample of 12 adults gave numbers of hours of sleep as shown below.

4.6 6.8 5.2 6.2 5.7 7.1 6.3 5.6 7.0 5.8 6.5 7.2

(iii) Calculate unbiased estimates of the mean and variance of the sleeping times of adults. [3]

Answers: (i) Commuters are not representative of the whole population; (ii) People who travel to work on this train; (iii) 6.17, 0.657. **J08/Q1**

13 The lengths of time people take to complete a certain type of puzzle are normally distributed with mean 48.8 minutes and standard deviation 15.6 minutes. The random variable X represents the time taken in minutes by a randomly chosen person to solve this type of puzzle. The times taken by random samples of 5 people are noted. The mean time $\overline{X}$ is calculated for each sample.

(i) State the distribution of $\overline{X}$, giving the values of any parameters. [2]

(ii) Find $P(\overline{X} < 50)$. [3]

Answers: (i) $N(48.8, \frac{15.6^2}{5})$; (ii) 0.568. **J08/Q2**

14 The weights in grams of oranges grown in a certain area are normally distributed with mean μ and standard deviation σ. A random sample of 50 of these oranges was taken, and a 97% confidence interval for μ based on this sample was (222.1, 232.1).

(i) Calculate unbiased estimates of μ and σ^2. [4]

(ii) Estimate the sample size that would be required in order for a 97% confidence interval for μ to have width 8. [3]

Answer: (i) 227.1 265
(ii) 78

J09/71/Q2

15 The weights in grams of oranges grown in a certain area are normally distributed with mean μ and standard deviation σ. A random sample of 50 of these oranges was taken, and a 97% confidence interval for μ based on this sample was (222.1, 232.1).

(i) Calculate unbiased estimates of μ and σ^2. [4]

(ii) Estimate the sample size that would be required in order for a 97% confidence interval for μ to have width 8. [3]

Answers: (i) 227.1 265
(ii) 78

J09/72/Q2

16 A random sample of n people were questioned about their internet use. 87 of them had a high-speed internet connection. A confidence interval for the population proportion having a high-speed internet connection is $0.1129 < p < 0.1771$.

(i) Write down the mid-point of this confidence interval and hence find the value of n. [3]

(ii) This interval is an α% confidence interval. Find α. [4]

Answers: (i) 0.145, $n = 600$; (ii) 97.4.

J10/72/Q2

17 The time taken, T minutes, for a special anti-rust paint to dry was measured for a random sample of 120 painted pieces of metal. The sample mean was 51.2 minutes and an unbiased estimate of the population variance was 37.4 minutes^2. Determine a 99% confidence interval for the mean drying time. [3]

Answer: $49.8 < \mu < 52.6$.

N02/Q1

18 The result of a memory test is known to be normally distributed with mean μ and standard deviation 1.9. It is required to have a 95% confidence interval for μ with a total width of less than 2.0. Find the least possible number of tests needed to achieve this. [4]

Answer: $n = 14$.

N03/Q1

19

Over a long period of time it is found that the amount of sunshine on any day in a particular town in Spain has mean 6.7 hours and standard deviation 3.1 hours.

(i) Find the probability that the mean amount of sunshine over a random sample of 300 days is between 6.5 and 6.8 hours. [4]

(ii) Give a reason why it is not necessary to assume that the daily amount of sunshine is normally distributed in order to carry out the calculation in part **(i)**. [1]

Answers: **(i)** 0.580; **(ii)** 300 is large, so CLT can be applied. **N04/Q2**

20 A random sample of 150 students attending a college is taken, and their travel times, t minutes, are measured. The data are summarised by $\Sigma t = 4080$ and $\Sigma t^2 = 159\,252$.

(i) Calculate unbiased estimates of the population mean and variance. [3]

(ii) Calculate a 94% confidence interval for the population mean travel time. [4]

Answers: **(i)** 27.2, 324; **(ii)** (24.4, 30.0). **N04/Q3**

21 The number of words on a page of a book can be modelled by a normal distribution with mean 403 and standard deviation 26.8. Find the probability that the average number of words per page in a random sample of 6 pages is less than 410. [4]

Answer: 0.739. **N05/Q1**

22 **(i)** Give a reason why, in carrying out a statistical investigation, a sample rather than a complete population may be used. [1]

(ii) Rose wishes to investigate whether men in her town have a different life-span from the national average of 71.2 years. She looks at government records for her town and takes a random sample of the ages of 110 men who have died recently. Their mean age in years was 69.3 and the unbiased estimate of the population variance was 65.61.

(a) Calculate a 90% confidence interval for the population mean and explain what you understand by this confidence interval. [4]

(b) State with a reason what conclusion about the life-span of men in her town Rose could draw from this confidence interval. [2]

Answers: **(i)** for example: cheaper, less time consuming, not all destructive; **(ii)(a)**(68.0, 70.6), we are 90% confident that the true mean lies between 68.0 and 70.6, **(b)** 71.2 not in confidence interval, significant difference in life span from national average. **N05/Q4**

23 **(i)** Write down the mean and variance of the distribution of the means of random samples of size n taken from a very large population having mean μ and variance σ^2. [2]

(ii) What, if anything, can you say about the distribution of sample means

(a) if n is large, [1]

(b) if n is small? [1]

Answers: **(i)** μ, $\frac{\sigma^2}{n}$; **(ii)(a)** normal, **(b)** unknown, or normal if the population is normal. **N06/Q2**

24 A survey was conducted to find the proportion of people owning DVD players. It was found that 203 out of a random sample of 278 people owned a DVD player.

(i) Calculate a 97% confidence interval for the true proportion of people who own a DVD player. [4]

A second survey to find the proportion of people owning DVD players was conducted at 10 o'clock on a Thursday morning in a shopping centre.

(ii) Give one reason why this is not a satisfactory sample. [1]

Answers: (i) (0.672, 0.788); **(ii)** mainly unemployed, retired, or mothers with children i.e. not representative of the whole population. **N06/Q3**

25 **(i)** Explain what is meant by the term 'random sample'. [1]

In a random sample of 350 food shops it was found that 130 of them had Special Offers.

(ii) Calculate an approximate 95% confidence interval for the proportion of all food shops with Special Offers. [4]

(iii) Estimate the size of a random sample required for an approximate 95% confidence interval for this proportion to have a width of 0.04. [3]

Answers: **(ii)** (0.321, 0.422); **(iii)** 2241 **N07/Q3**

26 Alan wishes to choose one child at random from the eleven children in his music class. The children are numbered 2, 3, 4, and so on, up to 12. Alan then throws two fair dice, each numbered from 1 to 6, and chooses the child whose number is the sum of the scores on the two dice.

(i) Explain why this is an unsatisfactory method of choosing a child. [2]

(ii) Describe briefly a satisfactory method of choosing a child. [2]

N08/Q1

27 Diameters of golf balls are known to be normally distributed with mean μ cm and standard deviation σ cm. A random sample of 130 golf balls was taken and the diameters, x cm, were measured. The results are summarised by $\Sigma x = 555.1$ and $\Sigma x^2 = 2371.30$.

(i) Calculate unbiased estimates of μ and σ^2. [3]

(ii) Calculate a 97% confidence interval for μ. [3]

(iii) 300 random samples of 130 balls are taken and a 97% confidence interval is calculated for each sample. How many of these intervals would you expect **not** to contain μ? [1]

Answers: (i) 4.27, 0.00793; (ii) (4.25, 4.29); (iii) 9. **N08/Q4**

28 The lengths of sewing needles in travel sewing kits are distributed normally with mean μ mm and standard deviation 1.5 mm. A random sample of n needles is taken. Find the smallest value of n such that the width of a 95% confidence interval for the population mean is at most 1 mm. [4]

Answer: 35. **N09/71/Q2**

29 The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} k\cos x & 0 \leqslant x \leqslant \frac{1}{4}\pi, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(i) Show that $k = \sqrt{2}$. [2]

(ii) Find $P(X > 0.4)$. [2]

(iii) Find the upper quartile of X. [3]

(iv) Find the probability that exactly 3 out of 5 random observations of X have values greater than the upper quartile. [2]

Answers: **(ii)** 0.449; **(iii)** 0.559; **(iv)** 0.0879. **N09/71/Q5**

---

30 There are 18 people in Millie's class. To choose a person at random she numbers the people in the class from 1 to 18 and presses the random number button on her calculator to obtain a 3-digit decimal. Millie then multiplies the first digit in this decimal by two and chooses the person corresponding to this new number. Decimals in which the first digit is zero are ignored.

(i) Give a reason why this is not a satisfactory method of choosing a person. [1]

Millie obtained a random sample of 5 people of her own age by a satisfactory sampling method and found that their heights in metres were 1.66, 1.68, 1.54, 1.65 and 1.57. Heights are known to be normally distributed with variance $0.0052\,\text{m}^2$.

(ii) Find a 98% confidence interval for the mean height of people of Millie's age. [3]

Answer: **(ii)** (1.54, 1.70). **N09/72/Q1**

---

31 In a survey of 1000 randomly chosen adults, 605 said that they used email. Calculate a 90% confidence interval for the proportion of adults in the whole population who use email. [3]

Answer: [0.580, 0.630]. **N10/72/Q1**

---

32 **(a)** Give a reason why sampling would be required in order to reach a conclusion about

(i) the mean height of adult males in England, [1]

(ii) the mean weight that can be supported by a single cable of a certain type without the cable breaking. [1]

(b) The weights, in kg, of sacks of potatoes are represented by the random variable X with mean μ and standard deviation σ. The weights of a random sample of 500 sacks of potatoes are found and the results are summarised below.

$$n = 500, \qquad \Sigma x = 9850, \qquad \Sigma x^2 = 194\,125.$$

(i) Calculate unbiased estimates of μ and σ^2. [3]

(ii) A further random sample of 60 sacks of potatoes is taken. Using your values from part **(b)** **(i)**, find the probability that the mean weight of this sample exceeds 19.73 kg. [4]

(iii) Explain whether it was necessary to use the Central Limit Theorem in your calculation in part **(b)** **(ii)**. [2]

Answers: (a)(i) Population too large, (ii) Testing involves destruction; (b)(i) 19.7, 0.160, (ii) 0.281, (iii) Yes; *X* is not necessarily normal, and the sample is large. **N10/72/Q7**

33 X is a random variable having the distribution $B\left(12, \frac{1}{4}\right)$. A random sample of 60 values of X is taken. Find the probability that the sample mean is less than 2.8. [5]

Answer: 0.151. **J11/72/Q2**

34 A doctor wishes to investigate the mean fat content in low-fat burgers. He takes a random sample of 15 burgers and sends them to a laboratory where the mass, in grams, of fat in each burger is determined. The results are as follows.

9 7 8 9 6 11 7 9 8 9 8 10 7 9 9

Assume that the mass, in grams, of fat in low-fat burgers is normally distributed with mean μ and that the population standard deviation is 1.3.

(i) Calculate a 99% confidence interval for μ. [4]

(ii) Explain whether it was necessary to use the Central Limit theorem in the calculation in part (i). [2]

(iii) The manufacturer claims that the mean mass of fat in burgers of this type is 8 g. Use your answer to part (i) to comment on this claim. [2]

Answers: (i) [7.54, 9.26]; (ii) No, because the population is normally distributed, so sample means are normally distributed; (iii) Claim is justified. **J11/72/Q4**

35 A population has mean 7 and standard deviation 3. A random sample of size n is chosen from this population.

(i) Write down the mean and standard deviation of the distribution of the sample mean. [2]

(ii) Under what circumstances does the sample mean have

(a) a normal distribution, [1]

(b) an approximately normal distribution? [1]

Answers: (i) Mean = 7, Standard Deviation = 3/√n (ii)(a) Population is Normal (b) Large sample **J12/72/Q2**

36 In a sample of 50 students at Batlin college, 18 support the football club Real Madrid.

(i) Calculate an approximate 98% confidence interval for the proportion of students at Batlin college who support Real Madrid. [4]

(ii) Give one condition for this to be a reliable result. [1]

Answers: (i) 0.202 to 0.518 (ii) Sample random **J12/72/Q3**

37 The masses, in grams, of a certain type of plum are normally distributed with mean μ and variance σ^2. The masses, m grams, of a random sample of 150 plums of this type were found and the results are summarised by $\Sigma m = 9750$ and $\Sigma m^2 = 647\,500$.

(i) Calculate unbiased estimates of μ and σ^2. [3]

(ii) Calculate a 98% confidence interval for μ. [3]

Two more random samples of plums of this type are taken and a 98% confidence interval for μ is calculated from each sample.

(iii) Find the probability that neither of these two intervals contains μ. [2]

Answer: 65 92.3
63.2 to 66.8
0.0004

J13/72/Q4

---

38 The weights, in grams, of a random sample of 8 packets of cereal are as follows.

250 248 255 244 259 250 242 258

Calculate unbiased estimates of the population mean and variance. [3]

Answers: 250.75 38.5

J14/72/Q1

---

39 Mahmoud throws a coin 400 times and finds that it shows heads 184 times. The probability that the coin shows heads on any throw is denoted by p.

(i) Calculate an approximate 95% confidence interval for p. [4]

(ii) Mahmoud claims that the coin is not fair. Use your answer to part **(i)** to comment on this claim. [1]

(iii) Mahmoud's result of 184 heads in 400 throws gives an α% confidence interval for p with width 0.1. Calculate the value of α. [4]

Answers: 0.411 to 0.509
Claim not supported
95.5%

J14/72/Q5

---

40 The volumes, v millilitres, of juice in a random sample of 50 bottles of Cooljoos are measured and summarised as follows.

$$n = 50 \qquad \Sigma v = 14\,800 \qquad \Sigma v^2 = 4\,390\,000$$

(i) Find unbiased estimates of the population mean and variance. [3]

(ii) An α% confidence interval for the population mean, based on this sample, is found to have a width of 5.45 millilitres. Find α. [4]

Four random samples of size 10 are taken and a 96% confidence interval for the population mean is found from each sample.

(iii) Find the probability that these 4 confidence intervals all include the true value of the population mean. [2]

Answer: (i) 296 188
(ii) 84%
(iii) 0.849

J15/72/Q5

---

41 The volumes of juice in bottles of Apricola are normally distributed. In a random sample of 8 bottles, the volumes of juice, in millilitres, were found to be as follows.

332 334 330 328 331 332 329 333

(i) Find unbiased estimates of the population mean and variance. [3]

A random sample of 50 bottles of Apricola gave unbiased estimates of 331 millilitres and 4.20 millilitres^2 for the population mean and variance respectively.

(ii) Use this sample of size 50 to calculate a 98% confidence interval for the population mean. [3]

(iii) The manufacturer claims that the mean volume of juice in all bottles is 333 millilitres. State, with a reason, whether your answer to part **(ii)** supports this claim. [1]

Answer: (i) 331, 4.125; (ii) 330 to 332; (iii) No because 333 is not within the CI **N11/72/Q4**

---

42 In order to obtain a random sample of people who live in her town, Jane chooses people at random from the telephone directory for her town.

(i) Give a reason why Jane's method will not give a random sample of people who live in the town. [1]

Jane now uses a valid method to choose a random sample of 200 people from her town and finds that 38 live in apartments.

(ii) Calculate an approximate 99% confidence interval for the proportion of all people in Jane's town who live in apartments. [4]

(iii) Jane uses the same sample to give a confidence interval of width 0.1 for this proportion. This interval is an x% confidence interval. Find the value of x. [4]

Answers: (i) Excludes people without phones; (ii) 0.119 to 0.261; (iii) 93. **N12/72/Q6**

---

43 Heights of a certain species of animal are known to be normally distributed with standard deviation 0.17 m. A conservationist wishes to obtain a 99% confidence interval for the population mean, with total width less than 0.2 m. Find the smallest sample size required. [4]

Answer: Smallest n is 20 **N13/72/Q2**

44 In a survey a random sample of 150 households in Nantville were asked to fill in a questionnaire about household budgeting.

(i) The results showed that 33 households owned more than one car. Find an approximate 99% confidence interval for the proportion of all households in Nantville with more than one car. [4]

(ii) The results also included the weekly expenditure on food, x dollars, of the households. These were summarised as follows.

$$n = 150 \qquad \Sigma x = 19\,035 \qquad \Sigma x^2 = 4\,054\,716$$

Find unbiased estimates of the mean and variance of the weekly expenditure on food of all households in Nantville. [3]

(iii) The government has a list of all the households in Nantville numbered from 1 to 9526. Describe briefly how to use random numbers to select a sample of 150 households from this list. [3]

N14/72/Q3

---

45 The mean and standard deviation of the time spent by people in a certain library are 29 minutes and 6 minutes respectively.

(i) Find the probability that the mean time spent in the library by a random sample of 120 people is more than 30 minutes. [4]

(ii) Explain whether it was necessary to assume that the time spent by people in the library is normally distributed in the solution to part **(i)**. [2]

Answer: 0.034 **N15/72/Q2**

No; n is large; The CLT applies (sample mean approx. normally distributed)

---

46 Jagdeesh measured the lengths, x minutes, of 60 randomly chosen lectures. His results are summarised below.

$$n = 60 \qquad \Sigma x = 3420 \qquad \Sigma x^2 = 195\,200$$

(i) Calculate unbiased estimates of the population mean and variance. [3]

(ii) Calculate a 98% confidence interval for the population mean. [3]

Answer: 57, 4.41 **N15/72/Q3**

56.4 to 57.6

Homework: Sampling & Estimation– Variant 71 & 73

1 The weight, in grams, of a certain type of apple is modelled by the random variable X with mean 62 and standard deviation 8.2. A random sample of 50 apples is selected, and the mean weight in grams, $\bar{X}$, is found.

(i) Describe fully the distribution of $\bar{X}$. [3]

(ii) Find $P(\bar{X} > 64)$. [3]

Answers: **(i)** $N(62, \frac{8.2^2}{50})$; **(ii)** 0.0423. **J10/73/Q3**

2 **(a)** The time taken by a worker to complete a task was recorded for a random sample of 50 workers. The sample mean was 41.2 minutes and an unbiased estimate of the population variance was 32.6 minutes^2. Find a 95% confidence interval for the mean time taken to complete the task. [3]

(b) The probability that an α% confidence interval includes only values that are lower than the population mean is $\frac{1}{16}$. Find the value of α. [2]

Answers: **(a)** [39.6, 42.8]; **(b)** 87.5. **J11/71/Q2**

3 In a random sample of 70 bars of Luxcleanse soap, 18 were found to be undersized.

(i) Calculate an approximate 90% confidence interval for the proportion of all bars of Luxcleanse soap that are undersized. [4]

(ii) Give a reason why your interval is only approximate. [1]

Answers: **(i)** [0.171, 0.343]; **(ii)** Either: the variance (or standard deviation) is estimated; or that a normal approximation to a binomial distribution was being used. **J11/73/Q2**

4 The weights, in grams, of packets of sugar are distributed with mean μ and standard deviation 23. A random sample of 150 packets is taken. The mean weight of this sample is found to be 494 g. Calculate a 98% confidence interval for μ. [3]

Answer: 490 to 498 **J12/71/Q1**

5 A random variable X has the distribution Po(3.2).

(i) A random value of X is found.

(a) Find $P(X \geqslant 3)$. [2]

(b) Find the probability that $X = 3$ given that $X \geqslant 3$. [3]

(ii) Random samples of 120 values of X are taken.

(a) Describe fully the distribution of the sample mean. [2]

(b) Find the probability that the mean of a random sample of size 120 is less than 3.3. [3]

Answers: (i)(a) 0.62(0) (b) 0.359 (ii)(a) (Approx) Normal with mean 3.2 and variance 3.2/120 (b) 0.730 **J12/71/Q5**

6 Leaves from a certain type of tree have lengths that are distributed with standard deviation 3.2 cm. A random sample of 250 of these leaves is taken and the mean length of this sample is found to be 12.5 cm.

(i) Calculate a 99% confidence interval for the population mean length. [3]

(ii) Write down the probability that the whole of a 99% confidence interval will lie below the population mean. [1]

Answers: (i) 12.0, 13.0 (ii) 0.005 or 0.5% **J12/73/Q1**

7 Marie wants to choose one student at random from Anthea, Bill and Charlie. She throws two fair coins. If both coins show tails she will choose Anthea. If both coins show heads she will choose Bill. If the coins show one of each she will choose Charlie.

(i) Explain why this is not a fair method for choosing the student. [2]

(ii) Describe how Marie could use the two coins to give a fair method for choosing the student. [2]

Answer: One of each is more likely; P(one of each)=0.5, P(HH) = P(TT) = 0.25 **J13/71/Q1**
Choose Charlie only if H then T; throw again if T then H

8 The lengths, x m, of a random sample of 200 balls of string are found and the results are summarised by $\Sigma x = 2005$ and $\Sigma x^2 = 20\,175$.

(i) Calculate unbiased estimates of the population mean and variance of the lengths. [3]

(ii) Use the values from part **(i)** to estimate the probability that the mean length of a random sample of 50 balls of string is less than 10 m. [3]

(iii) Explain whether or not it was necessary to use the Central Limit theorem in your calculation in part **(ii)**. [2]

Answer: 10.025 0.376 **J13/71/Q4**
0.387
Yes; distribution of X unknown

9 Each of a random sample of 15 students was asked how long they spent revising for an exam. The results, in minutes, were as follows.

50 70 80 60 65 110 10 70 75 60 65 45 50 70 50

Assume that the times for all students are normally distributed with mean μ minutes and standard deviation 12 minutes.

(i) Calculate a 92% confidence interval for μ. [4]

(ii) Explain what is meant by a 92% confidence interval for μ. [1]

(iii) Explain what is meant by saying that a sample is 'random'. [1]

Answers: (i) (56.6, 67.4) (ii) 92% of intervals capture μ (iii) Each possible sample is equally likely. **J13/73/Q3**

10 A die is biased. The mean and variance of a random sample of 70 scores on this die are found to be 3.61 and 2.70 respectively. Calculate a 95% confidence interval for the population mean score. [5]

Answer: 0.0047 **J14/71/Q2**

11 The score on one throw of a 4-sided die is denoted by the random variable X with probability distribution as shown in the table.

x	0	1	2	3
$P(X = x)$	0.25	0.25	0.25	0.25

(i) Show that $\text{Var}(X) = 1.25$. [1]

The die is thrown 300 times. The score on each throw is noted and the mean, $\bar{X}$, of the 300 scores is found.

(ii) Use a normal distribution to find $P(\bar{X} < 1.4)$. [3]

(iii) Justify the use of the normal distribution in part **(ii)**. [1]

Answers: 0.411 to 0.509
Claim not supported
95.5% **J14/71/Q5**

12 A die is thrown 100 times and shows an odd number on 56 throws. Calculate an approximate 97% confidence interval for the probability that the die shows an odd number on one throw. [4]

Answer: 0.452 to 0.668 **J14/73/Q3**

13 The masses, m grams, of a random sample of 80 strawberries of a certain type were measured and summarised as follows.

$$n = 80 \qquad \Sigma m = 4200 \qquad \Sigma m^2 = 229\,000$$

(i) Find unbiased estimates of the population mean and variance. [3]

(ii) Calculate a 98% confidence interval for the population mean. [3]

50 random samples of size 80 were taken and a 98% confidence interval for the population mean, μ, was found from each sample.

(iii) Find the number of these 50 confidence intervals that would be expected to include the true value of μ. [1]

Answers: (i) 52.5, 108 (ii) 49.8 to 55.2 (iii) 49 **J15/71/Q5**

14 Jyothi wishes to choose a representative sample of 5 students from the 82 members of her school year.

(i) She considers going into the canteen and choosing a table with five students from her year sitting at it, and using these five people as her sample. Give two reasons why this method is unsatisfactory. [2]

(ii) Jyothi decides to use another method. She numbers all the students in her year from 1 to 82. Then she uses her calculator and generates the following random numbers.

231492 762305 346280

From these numbers, she obtains the student numbers 23, 14, 76, 5, 34 and 62. Explain how Jyothi obtained these student numbers from the list of random numbers. [3]

Answers: **(i)** Not all candidates in canteen, not all sat on tables of 5 **(ii)** Split into two digit blocks; ignore if outside range; ignore repeats. **J15/73/Q1**

---

15 A die is biased so that the probability that it shows a six on any throw is p.

(i) In an experiment, the die shows a six on 22 out of 100 throws. Find an approximate 97% confidence interval for p. [4]

(ii) The experiment is repeated and another 97% confidence interval is found. Find the probability that exactly one of the two confidence intervals includes the true value of p. [2]

Answers: **(i)** 0.13 to 0.31 **(ii)** 0.0582 **J15/73/Q3**

---

16 The marks, x, of a random sample of 50 students in a test were summarised as follows.

$$n = 50 \qquad \Sigma x = 1508 \qquad \Sigma x^2 = 51\,825$$

(i) Calculate unbiased estimates of the population mean and variance. [3]

(ii) Each student's mark is scaled using the formula $y = 1.5x + 10$. Find estimates of the population mean and variance of the scaled marks, y. [3]

Answers: **(i)** 30.16, 129 **(ii)** 55.24, 291 **J15/73/Q4**

---

17 In a survey of 1000 randomly chosen adults, 605 said that they used email. Calculate a 90% confidence interval for the proportion of adults in the whole population who use email. [3]

Answer: [0.580, 0.630]. **N10/71/Q1**

---

18 **(a)** Give a reason why sampling would be required in order to reach a conclusion about

(i) the mean height of adult males in England, [1]

(ii) the mean weight that can be supported by a single cable of a certain type without the cable breaking. [1]

(b) The weights, in kg, of sacks of potatoes are represented by the random variable X with mean μ and standard deviation σ. The weights of a random sample of 500 sacks of potatoes are found and the results are summarised below.

$$n = 500, \qquad \Sigma x = 9850, \qquad \Sigma x^2 = 194\,125.$$

(i) Calculate unbiased estimates of μ and σ^2. [3]

(ii) A further random sample of 60 sacks of potatoes is taken. Using your values from part **(b) (i)**, find the probability that the mean weight of this sample exceeds 19.73 kg. [4]

(iii) Explain whether it was necessary to use the Central Limit Theorem in your calculation in part **(b) (ii)**. [2]

Answers: (a)(i) Population too large, (ii) Testing involves destruction; (b)(i) 19.7, 0.160, (ii) 0.281, (iii) Yes; *X* is not necessarily normal, and the sample is large. **N10/71/Q7**

---

19 The editor of a magazine wishes to obtain the views of a random sample of readers about the future of the magazine.

(i) A sub-editor proposes that they include in one issue of the magazine a questionnaire for readers to complete and return. Give two reasons why the readers who return the questionnaire would not form a random sample. [2]

The editor decides to use a table of random numbers to select a random sample of 50 readers from the 7302 regular readers. These regular readers are numbered from 1 to 7302. The first few random numbers which the editor obtains from the table are as follows.

49757 80239 52038 60882

(ii) Use these random numbers to select the first three members in the sample. [2]

Answer: (ii) 4975, 3952, (0)386 (for example). **N10/73/Q2**

---

20 The masses of sweets produced by a machine are normally distributed with mean μ grams and standard deviation 1.0 grams. A random sample of 65 sweets produced by the machine has a mean mass of 29.6 grams.

(i) Find a 99% confidence interval for μ. [3]

The manufacturer claims that the machine produces sweets with a mean mass of 30 grams.

(ii) Use the confidence interval found in part **(i)** to draw a conclusion about this claim. [2]

(iii) Another random sample of 65 sweets produced by the machine is taken. This sample gives a 99% confidence interval that leads to a different conclusion from that found in part **(ii)**. Assuming that the value of μ has not changed, explain how this can be possible. [1]

Answers: (i) (29.3, 29.9); (ii) The claim is not supported. **N10/73/Q3**

---

21 The volumes of juice in bottles of Apricola are normally distributed. In a random sample of 8 bottles, the volumes of juice, in millilitres, were found to be as follows.

332 334 330 328 331 332 329 333

(i) Find unbiased estimates of the population mean and variance. [3]

A random sample of 50 bottles of Apricola gave unbiased estimates of 331 millilitres and 4.20 millilitres^2 for the population mean and variance respectively.

(ii) Use this sample of size 50 to calculate a 98% confidence interval for the population mean. [3]

(iii) The manufacturer claims that the mean volume of juice in all bottles is 333 millilitres. State, with a reason, whether your answer to part **(ii)** supports this claim. [1]

Answer: **(i)** 331, 4.125; **(ii)** 330 to 332; **(iii)** No because 333 is not within the CI **N11/71/Q4**

22 35% of a random sample of n students walk to college. This result is used to construct an approximate 98% confidence interval for the population proportion of students who walk to college. Given that the width of this confidence interval is 0.157, correct to 3 significant figures, find n. [5]

Answer: $n = 200$ **N11/73/Q2**

23 Jack has to choose a random sample of 8 people from the 750 members of a sports club.

(i) Explain fully how he can use random numbers to choose the sample. [3]

Jack asks each person in the sample how much they spent last week in the club café. The results, in dollars, were as follows.

15 25 30 8 12 18 27 25

(ii) Find unbiased estimates of the population mean and variance. [3]

(iii) Explain briefly what is meant by 'population' in this question. [1]

Answer: **(ii)** 20, 62.3; **(iii)** Amounts spent last week in the cafe by all club members. **N11/73/Q3**

24 In order to obtain a random sample of people who live in her town, Jane chooses people at random from the telephone directory for her town.

(i) Give a reason why Jane's method will not give a random sample of people who live in the town. [1]

Jane now uses a valid method to choose a random sample of 200 people from her town and finds that 38 live in apartments.

(ii) Calculate an approximate 99% confidence interval for the proportion of all people in Jane's town who live in apartments. [4]

(iii) Jane uses the same sample to give a confidence interval of width 0.1 for this proportion. This interval is an x% confidence interval. Find the value of x. [4]

Answers: **(i)** Excludes people without phones; **(ii)** 0.119 to 0.261; **(iii)** 93. **N12/71/Q6**

25 **(i)** A random variable X has mean μ and variance σ^2. The mean of a random sample of n values of X is denoted by $\overline{X}$. Give expressions for $E(\overline{X})$ and $Var(\overline{X})$. [2]

(ii) The heights, in centimetres, of adult males in Brancot are normally distributed with mean 177.8 and standard deviation 6.1. Find the probability that the mean height of a random sample of 12 adult males from Brancot is less than 176 cm. [3]

(iii) State, with a reason, whether it was necessary to use the Central Limit Theorem in the calculation in part **(ii)**. [1]

Answers: **(i)** μ, σ^2/n; **(ii)** 0.153; **(iii)** Not necessary population heights given as normally distributed. **N12/73/Q2**

26 It is claimed that, on average, people following the Losefast diet will lose more than 2 kg per month. The weight losses, x kilograms per month, of a random sample of 200 people following the Losefast diet were recorded and summarised as follows.

$$n = 200 \qquad \Sigma x = 460 \qquad \Sigma x^2 = 1636$$

(i) Calculate unbiased estimates of the population mean and variance. [3]

(ii) Test the claim at the 1% significance level. [5]

Answers: **(i)** 2.3, 2.90; **(ii)** There is evidence that the mean weight loss is more than 2 kg. **N12/73/Q5**

27 Heights of a certain species of animal are known to be normally distributed with standard deviation 0.17 m. A conservationist wishes to obtain a 99% confidence interval for the population mean, with total width less than 0.2 m. Find the smallest sample size required. [4]

Answer: Smallest n is 20 **N13/71/Q2**

28 A random sample of 80 values of a variable X is taken and these values are summarised below.

$$n = 80 \qquad \Sigma x = 150.2 \qquad \Sigma x^2 = 820.24$$

Calculate unbiased estimates of the population mean and variance of X and hence find a 95% confidence interval for the population mean of X. [6]

Answers: 1.8775 or 1.88 6.81 (1.31, 2.45) or 1.31 to 2.45 **N13/73/Q1**

29 In a survey a random sample of 150 households in Nantville were asked to fill in a questionnaire about household budgeting.

(i) The results showed that 33 households owned more than one car. Find an approximate 99% confidence interval for the proportion of all households in Nantville with more than one car. [4]

(ii) The results also included the weekly expenditure on food, x dollars, of the households. These were summarised as follows.

$$n = 150 \qquad \Sigma x = 19\,035 \qquad \Sigma x^2 = 4\,054\,716$$

Find unbiased estimates of the mean and variance of the weekly expenditure on food of all households in Nantville. [3]

(iii) The government has a list of all the households in Nantville numbered from 1 to 9526. Describe briefly how to use random numbers to select a sample of 150 households from this list. [3]

Answer: 0.133 to 0.307; 126.9, 11001.17; 4-digit numbers, ignore numbers > 9526, ignore repeats **N14/71/Q4**

---

30 The times, in minutes, taken by people to complete a walk are normally distributed with mean μ. The times, t minutes, for a random sample of 80 people were summarised as follows.

$$\Sigma t = 7220 \qquad \Sigma t^2 = 656\,060$$

(i) Calculate a 97% confidence interval for μ. [6]

(ii) Explain whether it was necessary to use the Central Limit theorem in part **(i)**. [2]

Answers: (i) the interval 88.4 to 92.1, (ii) the population was given as normally distributed, so it was necessary to use the Central Limit Theorem. **N14/73/Q3**

---

31 The mean and standard deviation of the time spent by people in a certain library are 29 minutes and 6 minutes respectively.

(i) Find the probability that the mean time spent in the library by a random sample of 120 people is more than 30 minutes. [4]

(ii) Explain whether it was necessary to assume that the time spent by people in the library is normally distributed in the solution to part **(i)**. [2]

Answer: 0.034 **N15/71/Q2**
No; n is large; The CLT applies (sample mean approx. normally distributed)

---

32 Jagdeesh measured the lengths, x minutes, of 60 randomly chosen lectures. His results are summarised below.

$$n = 60 \qquad \Sigma x = 3420 \qquad \Sigma x^2 = 195\,200$$

(i) Calculate unbiased estimates of the population mean and variance. [3]

(ii) Calculate a 98% confidence interval for the population mean. [3]

Answer: 57, 4.41 **N15/71/Q3**
56.4 to 57.6

---

33 It is known that the number, N, of words contained in the leading article each day in a certain newspaper can be modelled by a normal distribution with mean 352 and variance 29. A researcher takes a random sample of 10 leading articles and finds the sample mean, $\overline{N}$, of N.

(i) State the distribution of $\overline{N}$, giving the values of any parameters. [2]

(ii) Find $P(\overline{N} > 354)$. [3]

Answers: (i) normal 352 2.9 , (ii) 0.120 **N15/73/Q1**

---

34 From a random sample of 65 people in a certain town, the proportion who own a bicycle was noted. From this result an α% confidence interval for the proportion, p, of all people in the town who own a bicycle was calculated to be $0.284 < p < 0.516$.

(i) Find the proportion of people in the sample who own a bicycle. [1]

(ii) Calculate the value of α correct to 2 significant figures. [4]

Answers: (i) 0.4, (ii) 94 **N15/73/Q3**

---

35 The diameter, in cm, of pistons made in a certain factory is denoted by X, where X is normally distributed with mean μ and variance σ^2. The diameters of a random sample of 100 pistons were measured, with the following results.

$$n = 100 \qquad \Sigma x = 208.7 \qquad \Sigma x^2 = 435.57$$

(i) Calculate unbiased estimates of μ and σ^2. [3]

The pistons are designed to fit into cylinders. The internal diameter, in cm, of the cylinders is denoted by Y, where Y has an independent normal distribution with mean 2.12 and variance 0.000 144. A piston will not fit into a cylinder if $Y - X < 0.01$.

(ii) Using your answers to part **(i)**, find the probability that a randomly chosen piston will not fit into a randomly chosen cylinder. [6]

Answers: (i) 2.087 (2.09 accepted here) 0.000132 , (ii) 0.0832 **N15/73/Q7**

---

36 A researcher is investigating the lengths, in kilometres, of the journeys to work of the employees at a certain firm. She takes a random sample of 10 employees.

(i) State what is meant by 'random' in this context. [1]

The results of her sample are as follows.

1.5 2.0 3.6 5.9 4.8 8.7 3.5 2.9 4.1 3.0

(ii) Find unbiased estimates of the population mean and variance. [3]

(iii) State what is meant by 'population' in this context. [1]

Answer: Each employee has an equal chance of being chosen **J16/71/Q2**

4 4.36

Distances travelled by all employees at the firm

---

37 Based on a random sample of 700 people living in a certain area, a confidence interval for the proportion, p, of all people living in that area who had travelled abroad was found to be $0.5672 < p < 0.6528$.

(i) Find the proportion of people in the sample who had travelled abroad. [1]

(ii) Find the confidence level of this confidence interval. Give your answer correct to the nearest integer. [4]

Answer: 0.61 **J16/71/Q3**

98%

---

38 The time taken for a particular type of paint to dry was measured for a sample of 150 randomly chosen points on a wall. The sample mean was 192.4 minutes and an unbiased estimate of the population variance was 43.6 minutes^2. Find a 98% confidence interval for the mean drying time. [3]

Answer: 191 to 194 **J16/73/Q1**

39 X and Y are independent random variables with distributions Po(1.6) and Po(2.3) respectively.

(i) Find $P(X + Y = 4)$. [3]

A random sample of 75 values of X is taken.

(ii) State the approximate distribution of the sample mean, $\overline{X}$, including the values of the parameters. [2]

(iii) Hence find the probability that the sample mean is more than 1.7. [3]

(iv) Explain whether the Central Limit theorem was needed to answer part **(ii)**. [1]

Answers: **(i)** 0.195 **(ii)** N(1.6, 1.6/75) **(iii)** 0.247 **(iv)** Required as X not normally distributed. **J16/73/Q6**

40 **(a)** The waiting time at a certain bus stop has variance 2.6 minutes^2. For a random sample of 75 people, the mean waiting time was 7.1 minutes. Calculate a 92% confidence interval for the population mean waiting time. [3]

(b) A researcher used 3 random samples to calculate 3 independent 92% confidence intervals. Find the probability that all 3 of these confidence intervals contain only values that are greater than the actual population mean. [2]

(c) Another researcher surveyed the first 75 people who waited at a bus stop on a Monday morning. Give a reason why this sample is unsuitable for use in finding a confidence interval for the mean waiting time. [1]

Answers: **(a)** 6.77 to 7.43 **(b)** 0.000064 **(c)** e.g. Particular day or time of day was used **J17/71/Q3**

41 A residents' association has 654 members, numbered from 1 to 654. The secretary wishes to send a questionnaire to a random sample of members. In order to choose the members for the sample she uses a table of random numbers. The first line in the table is as follows.

1096 4357 3765 0431 0928 9264

The numbers of the first two members in the sample are 109 and 643. Find the numbers of the next three members in the sample. [3]

Answer: 573, (0)43 289 **J17/73/Q1**

42 In a random sample of 200 shareholders of a company, 103 said that they wanted a change in the management.

(i) Find an approximate 92% confidence interval for the proportion, p, of all shareholders who want a change in the management. [3]

(ii) State the probability that a 92% confidence interval does not contain p. [1]

Answers: (i) 0.453 to 0.577 (ii) 8% **J17/73/Q2**

43 A random sample of 75 values of a variable X gave the following results.

$$n = 75 \qquad \Sigma x = 153.2 \qquad \Sigma x^2 = 340.24$$

Find unbiased estimates for the population mean and variance of X. [3]

Answer: 2.04 0.369 **J18/71/Q1**

44 A six-sided die is suspected of bias. The die is thrown 100 times and it is found that the score is 2 on 20 throws. It is given that the probability of obtaining a score of 2 on any throw is p.

(i) Find an approximate 94% confidence interval for p. [3]

(ii) Use your answer to part **(i)** to comment on whether the die may be biased. [1]

Answers: **(i)** 0.125 to 0.275 **(ii)** No evidence of bias as $\frac{1}{6}$ is within this range **J18/71/Q2**

45 Amy has to choose a random sample from the 265 students in her year at college. She numbers the students from 1 to 265 and then uses random numbers generated by her calculator. The first two random numbers produced by her calculator are 0.213 165 448 and 0.073 165 196.

(i) Use these figures to find the numbers of the first four students in her sample. [2]

There were 25 students in Amy's sample. She asked each of them how much money, \$$x$, they earned in a week, on average. Her results are summarised below.

$$n = 25 \qquad \Sigma x = 510 \qquad \Sigma x^2 = 13\,225$$

(ii) Find unbiased estimates of the population mean and variance. [3]

(iii) Explain briefly what is meant by 'population' in this question. [1]

Answers: **(i)** 213,165, (0)73, 196 **(ii)** 20.4,118 **(iii)** The weekly earnings of all students in Amy's year **J18/73/Q2**

46 A researcher wishes to estimate the proportion, p, of houses in London Road that have only one occupant. He takes a random sample of 64 houses in London Road and finds that 8 houses in the sample have only one occupant. Using this sample, he calculates that an approximate α% confidence interval for p has width 0.130. Find α correct to the nearest integer. [5]

Answer: 88% **J18/73/Q3**

47 The weights, in kilograms, of a random sample of eight 16-year old males are given below.

58.9 63.5 62.7 59.4 66.9 68.0 60.4 68.2

Find unbiased estimates of the population mean and variance of the weights of all 16-year old males. [3]

Answer: 63.5 14.6 **N16/71/Q1**

48 **(a)** The masses, in grams, of certain tomatoes are normally distributed with standard deviation 9 grams. A random sample of 100 tomatoes has a sample mean of 63 grams. Find a 90% confidence interval for the population mean mass of these tomatoes. [3]

(b) The masses, in grams, of certain potatoes are normally distributed with known population standard deviation but unknown population mean. A random sample of potatoes is taken in order to find a confidence interval for the population mean. Using a sample of size 50, a 95% confidence interval is found to have width 8 grams.

(i) Using another sample of size 50, an $\alpha\%$ confidence interval has width 4 grams. Find α. [3]

(ii) Find the sample size n, such that a 95% confidence interval has width 4 grams. [2]

Answer: 61.5 to 64.5
67.3
200

N16/71/Q5

---

49 Dominic wishes to choose a random sample of five students from the 150 students in his year. He numbers the students from 1 to 150. Then he uses his calculator to generate five random numbers between 0 and 1. He multiplies each random number by 150 and rounds up to the next whole number to give a student number.

(i) Dominic's first random number is 0.392. Find the student number that is produced by this random number. [1]

(ii) Dominic's second student number is 104. Find a possible random number that would produce this student number. [1]

(iii) Explain briefly why five random numbers may not be enough to produce a sample of five student numbers. [1]

Answer: 59 *Answer:* most likely answer 0.693 *Answer:* possible repeats

N16/73/Q2

---

50 A men's triathlon consists of three parts: swimming, cycling and running. Competitors' times, in minutes, for the three parts can be modelled by three independent normal variables with means 34.0, 87.1 and 56.9, and standard deviations 3.2, 4.1 and 3.8, respectively. For each competitor, the total of his three times is called the race time. Find the probability that the mean race time of a random sample of 15 competitors is less than 175 minutes. [5]

Answer: 0.0356

N16/73/Q3

---

51 A variable X takes values 1, 2, 3, 4, 5, and these values are generated at random by a machine. Each value is supposed to be equally likely, but it is suspected that the machine is not working properly. A random sample of 100 values of X, generated by the machine, gives the following results.

$$n = 100 \qquad \Sigma x = 340 \qquad \Sigma x^2 = 1356$$

(i) Find a 95% confidence interval for the population mean of the values generated by the machine. [6]

(ii) Use your answer to part **(i)** to comment on whether the machine may be working properly. [2]

Answer: 3.12 to 3.68 or (3.12 , 3.68)

N16/73/Q6

Answer: the machine was probably not working properly

52 After an election 153 adults, from a random sample of 200 adults, said that they had voted. Using this information, an $\alpha\%$ confidence interval for the proportion of all adults who voted in the election was found to be 0.695 to 0.835, both correct to 3 significant figures. Find the value of α, correct to the nearest integer. [4]

Answer: 98 **N17/71/Q3**

53 The lengths, in millimetres, of rods produced by a machine are normally distributed with mean μ and standard deviation 0.9. A random sample of 75 rods produced by the machine has mean length 300.1 mm.

(i) Find a 99% confidence interval for μ, giving your answer correct to 2 decimal places. [3]

The manufacturer claims that the machine produces rods with mean length 300 mm.

(ii) Use the confidence interval found in part **(i)** to comment on this claim. [1]

Answers: (i) 299.83 to 300.37 or (299.83, 300.37), **(ii)** The confidence interval includes 300 so the claim is supported **N17/71/Q4**

54 The standard deviation of the heights of adult males is 7.2 cm. The mean height of a sample of 200 adult males is found to be 176 cm.

(i) Calculate a 97.5% confidence interval for the mean height of adult males. [3]

(ii) State a necessary condition for the calculation in part **(i)** to be valid. [1]

Answers: (i) 175 to 177 (ii) random sample **N18/71/Q1**

55 A population has mean 12 and standard deviation 2.5. A large random sample of size n is chosen from this population and the sample mean is denoted by $\overline{X}$. Given that $P(\overline{X} < 12.2) = 0.975$, correct to 3 significant figures, find the value of n. [4]

Answer: 600 **N18/71/Q3**

Chapter 5: Hypothesis testing using a Normal Distribution

In the production of ice packs for use in cool boxes, a machine fills packs with liquid and the packs are then frozen. Since space is needed in the packs for the liquid to expand, it is important that they are not over-filled. The volume of liquid in the packs follows a normal distribution with mean 524 ml and standard deviation 3 ml.

The machine breaks down and is repaired. In the next batch of production, there is a suspicion that the mean volume of liquid dispensed by the machine into the packs has increased and is greater than 524 ml. In order to investigate this, the supervisor takes a random sample of 50 packs and finds that the mean volume of liquid in these is 524.9 ml. Does this provide evidence that the machine is over-dispensing?

The mean volume of the sample, 524.9 ml, is higher than the established mean of 524 ml. But is it high enough to say that the mean volume of *all* the packs filled by the machine has increased? Perhaps the mean is still 524 ml and this higher value has occurred just because of sampling variation. A **hypothesis** (or **significance**) **test** will enable a decision to be made that is backed by statistical theory, not just based on a suspicion.

Let X be the volume, in millitres, of liquid dispensed into a pack after the machine has been repaired and let the mean of X be μ, where μ is unknown. Assuming that the standard deviation remains unchanged, $X \sim N(\mu, \sigma^2)$ with $\sigma = 3$.

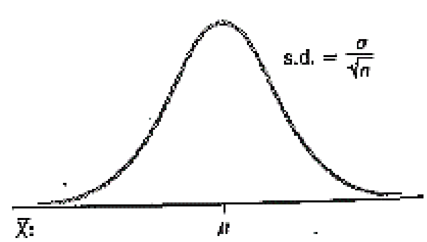
The hypothesis is made that μ is 524 ml, i.e. the mean has remained the same as it was prior to the repair. This is known as the **null hypothesis**, H_0 and is written

$H_0: \mu = 524$

Since it is suspected that the mean volume has *increased*, the **alternative hypothesis**, H_1, is that the mean is *greater than* 524 ml. This is written

$H_1: \mu > 524$

To carry out the test, the focus moves from X, the volume of liquid in a pack, to the distribution of $\overline{X}$, the *mean volume of a sample of 50 packs*. In this test, $\overline{X}$ is known as the **test statistic** and its distribution is needed. The distribution of $\overline{X}$ is known as the **sampling distribution of means**.



In Chapter 9 you saw that if $X \sim N(\mu, \sigma^2)$,

then, for samples of size n, $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

The hypothesis test starts by assuming that the value stated in the null hypothesis is true, so $\mu = 524$.

Since $\sigma = 3$ and $n = 50$,

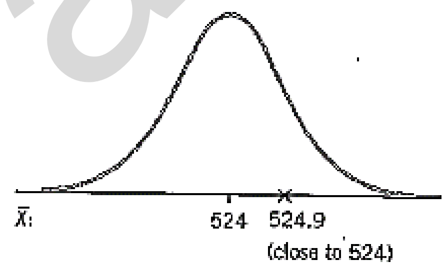
$$\overline{X} \sim N\left(524, \frac{3^2}{50}\right), \quad \text{i.e. } \overline{X} \sim N(524, 0.18).$$

The sampling distribution of means, therefore, follows a normal distribution with mean 524 ml and variance 0.18 ml^2. The standard deviation is $\sqrt{0.18}$ ml.

NOTE: This is sometimes left in the form $\sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{50}}$.

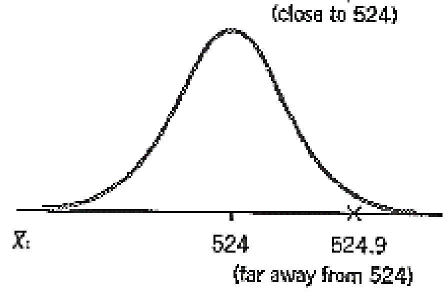
The result of the test depends on the whereabouts in the sampling distribution of the test value of 524.9 ml, the mean volume of the sample of 50 packs taken by the supervisor. She would need to find out whether 524.9 is close to 524 or far away from 524.

If it is *close to* 524 then it is likely to have come from a distribution with mean 524 ml and there would not be enough evidence to say that the mean volume has increased.



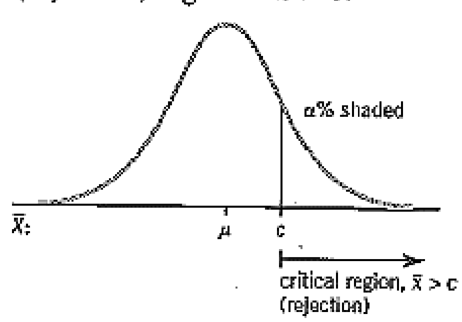
If it is *far away from* 524, i.e. in the right-hand (upper tail) of the distribution, then it is unlikely to have come from a distribution with a mean of 524 ml. The mean is likely to be higher than 524 ml.

Note that the *upper tail* is being used because the supervisor suspects that there is an *increase* in μ. This type of test is called a **one-tailed (upper tail) test.**



A decision needs to be taken about the cut off point, c, known as the **critical value**, which indicates the boundary of the region where values of $\bar{x}$ would be considered to be *too far away* from 524 ml and therefore would be *unlikely to occur*. The region is known as the **critical region or rejection region.**

The critical value and region are fixed using probabilities linked to the significance level of the test. In general, for an upper tail test at the $\alpha\%$ level, the critical value c is fixed so that $P(\overline{X} > c) = \alpha\%$ and the critical (rejection) region is $\bar{x} > c$.



The hypothesis test involves finding whether or not the sample value, $\bar{x}$, lies in the critical region of the sampling distribution of means, $\bar{X}$.

In this example, if $\bar{x}$ lies in the critical region, then a decision is taken that it is too far away from 524 ml to have come from a distribution with this mean. In statistical language, you would reject the null hypothesis, H_0 (that the mean is 524 ml), in favour of the alternative hypothesis, H_1 (that the mean is greater than 524 ml).

If $\bar{x}$ does not lie in the critical region, there is not have enough evidence to reject H_0, so H_0 is accepted. In this example, $\bar{x} < c$ is known as the acceptance region.

For a significance level of $\alpha\%$, if the sample mean lies in the critical (or rejection) region, then the result is said to be **significant at the $\alpha\%$ level.**

Note that if a result is significant at, say, the 1% level, then it is automatically significant at any level greater than 1%, for example 5% or 10%.

Say that the supervisor chooses a significance level of 5%. She will then reject H_0 if the test value (i.e. the mean volume of the sample of 50 cans) lies in the upper tail 5% of the distribution of sample means.

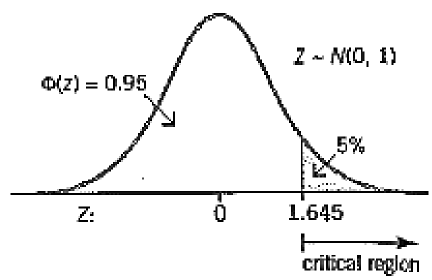
Since this distribution is normal, instead of finding c, the critical $\bar{x}$ value, it is possible to work in standardised values and find the z-value that gives 5% in the upper tail. Using standard normal tables (page 649),

if $\quad P(Z > z) = 0.05$

then $\quad P(Z < z) = 1 - 0.05 = 0.95$

i.e. $\quad \Phi(z) = 0.95$

$$z = \Phi^{-1}(0.95) = 1.645$$



So z-values that are greater than 1.645 lie in the upper tail 5% of the distribution.

This enables a statement to be made, known as the **rejection criterion**, which tells you when to reject the null hypothesis:

Reject H_0 if $z > 1.645$, where z is the standardised value of the mean of the sample of 50 packs,

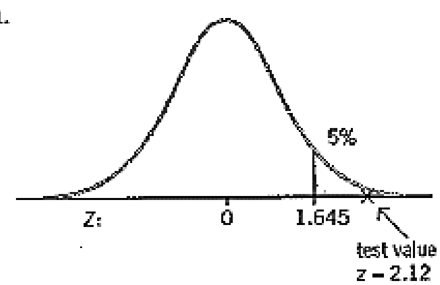
i.e. $\quad z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{\bar{x} - 524}{3/\sqrt{50}}$

Note that to avoid being influenced by sample readings, it is important that the rejection criterion is decided upon *before* any sample values are taken.

When the sample was taken, it was found that $\bar{x} = 524.9$,

so $\quad z = \dfrac{524.9 - 524}{3/\sqrt{50}}$

$= 2.12$ (2 d.p.)



The result of the test is now stated in statistical terms and then related to the context of the test, as follows:

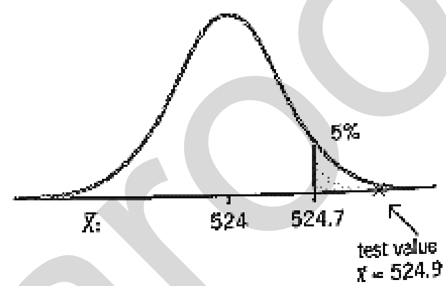
Since $z > 1.645$, H_0 is rejected in favour of H_1. The supervisor would conclude that the mean volume of liquid being dispensed by the machine is not 524 ml, but has increased, she would be wise therefore to stop production so that the setting on the machine could be adjusted.

Note that the critical $\bar{x}$-value, c, can be found by de-standardising the critical z-value of 1.645, where

$$\frac{c-524}{3/\sqrt{50}} = 1.645$$

$$c = 524 + 1.645 \times \frac{3}{\sqrt{50}}$$

$$= 524.7$$



Since the test value of 524.9 is greater than 524.7, it lies in the critical region, confirming the result obtained above.

If you want even more information, you can find out *exactly* where the sample mean lies in the distribution of $\bar{X}$. Note that this is the method used in Chapter 10 for discrete variables.

$$\bar{X} \sim N\left(524, \frac{3^2}{50}\right)$$

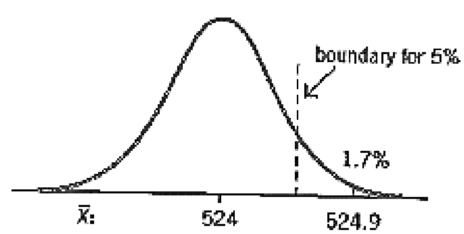
so $$P(\bar{X} > 524.9) = P\left(Z > \frac{524.9-524}{3/\sqrt{50}}\right)$$

$$= P(Z > 2.1213\ldots)$$

$$= 1 - 0.9831$$

$$= 0.0169$$

$$\approx 1.7\%$$



This probability is less than 5%, implying that the boundary for the critical region must lie to the left of the sample value of 524.9 and confirming that 524.9 lies in the critical region. This method also tells you that the test value of 524.9 will lie in the critical region for any level of significance above 1.7%.

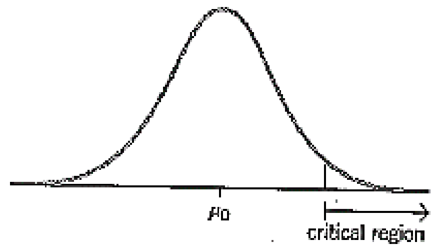
This probability method can be used, if preferred, in the hypothesis test to find whether the sample value lies in the rejection region. In this example, for a 5% level of significance, the rejection criterion would be to reject H_0 if $P(\bar{X} > \bar{x}) < 0.05$, where $\bar{x}$ is the sample mean.

ONE TAILED AND TWO TAILED TESTS

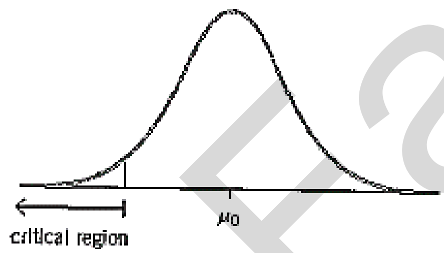
Say that the null hypothesis is $\mu = \mu_0$.

In a one-tailed test, the alternative hypothesis H_1 looks for an increase or a decrease in μ:

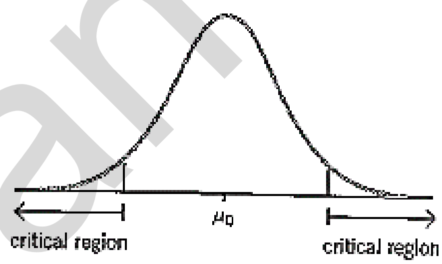
- for an increase, H_1 is $\mu > \mu_0$ and the critical region is in the upper tail,



- for a decrease, H_1 is $\mu < \mu_0$ and the critical region is in the lower tail,



In a two-tailed test, the alternative hypothesis H_1 looks for a change in μ without specifying whether it is an increase or a decrease and H_1 is $\mu \neq \mu_0$. The critical region is in two parts:

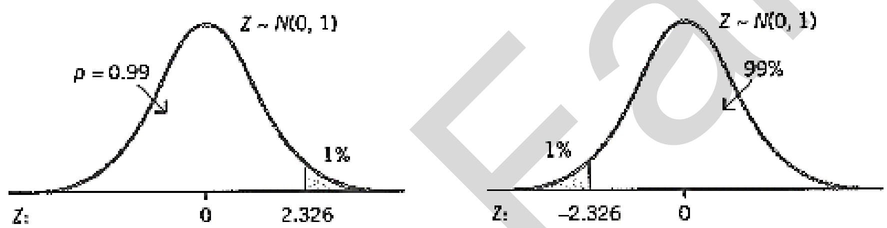


CRITICAL Z-VALUES

Critical values depend on the significance level and also whether the test is one- or two-tailed. The method of working in standardised values is widely used for tests involving the normal distribution because the critical z-values can be found easily from standard normal tables, as described on page 529. Sometimes the most commonly needed values are summarised in a critical value table. One such table is shown below and it is also printed in the appendix at the bottom of page 649. It gives the z-values for various values of p, where $p = P(Z < z) = \Phi(z)$.

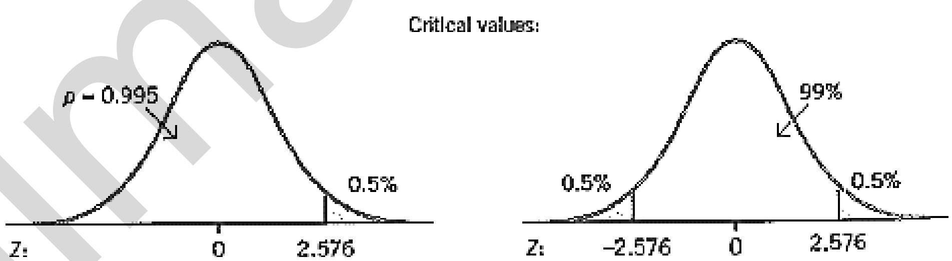
p	0.75	0.90	0.95	0.975	0.99	0.995	0.9975	0.999	0.9995
z	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

For example, *for a one-tailed test, at 1% level:* you want to find z such that $\Phi(z) = 0.99$, so look up $p = 0.99$.
From the table, $z = 2.326$. Therefore the upper tail critical value is 2.326. By symmetry, the lower tail critical value is -2.326.



For a two-tailed test, at the 1% level: the 1% in the tails is split evenly between the upper and lower tails with 0.5% in each. There are two critical values.

To find the upper tail value, you need to find z such that $\Phi(z) = 0.995$, so look up $p = 0.995$.
From the table $z = 2.576$.
So the upper tail critical value is 2.576 and the lower tail value (by symmetry) is -2.576.



SUMMARY OF CRITICAL VALUES AND REJECTION CRITERIA

The summary below shows the critical z-values and rejection criteria for the most commonly used levels of significance: 10%, 5% and 1%.

	One-tailed test (lower tail) $H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	One-tailed test (upper tail) $H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	Two-tailed test $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
10% significance level	Reject H_0 if $z < -1.282$	Reject H_0 if $z > 1.282$	Reject H_0 if $z < -1.645$ or $z > 1.645$ (written $\lvert z \rvert > 1.645$)
5% significance level	Reject H_0 if $z < -1.645$	Reject H_0 if $z > 1.645$	Reject H_0 if $z < -1.96$ or $z > 1.96$ (written $\lvert z \rvert > 1.96$)
1% significance level	Reject H_0 if $z < -2.326$	Reject H_0 if $z > 2.326$	Reject H_0 if $z < -2.576$ or $z > 2.576$ (written $\lvert z \rvert > 2.576$)

STAGES IN THE HYPOTHESIS TEST

When carrying out a hypothesis test, it is useful to work through the following stages. This is essentially the same procedure as in the tests for parameters of discrete distributions described in Chapter 10.

1. State the variable being considered.
2. State the null hypothesis H_0 and the alternative hypothesis H_1.

   Remember that if you are looking for
   a decrease, then $H_1: \ldots < \ldots$ (one-tailed test, lower tail)
   an increase, then $H_1: \ldots > \ldots$ (one-tailed test, upper tail)
   a change, then $H_1: \ldots \neq \ldots$ (two-tailed test, upper and lower tails)
3. Consider the distribution of the test statistic, assuming that the null hypothesis is true. If you are testing a sample mean, then the test statistic is $\overline{X}$, and the sampling distribution of means is considered.
4. State the type of test (i.e. whether it is one-tailed or two-tailed) and decide the significance level of the test.
5. Decide on your rejection criterion, remembering that you will reject H_0 if the test value lies in the critical (or rejection) region fixed by the significance level.

Now consider the value of the test statistic

6. Perform any calculations necessary to find out whether the test value is in the critical region.

7. Make your conclusion in statistical terms:

- If the test value is in the critical region, reject H_0 in favour of H_1.
- If the test value is not in the critical region, do not reject H_0.

Then relate your conclusion to the situation being tested.

There are several hypothesis tests involving continuous distributions and some of these are illustrated in the following text.

HYPOTHESIS TEST 1: TESTING μ (THE MEAN OF A POPULATION)

Consider a population X with unknown mean μ and variance σ^2.
A value for μ, call it μ_0, is specified in the hypotheses, for example

$H_0 : \mu = \mu_0$
$H_1 : \mu < \mu_0$ (or $\mu > \mu_0$ or $\mu \neq \mu_0$)

To test the hypotheses, take a sample of size n from the population and calculate the sample mean, $\bar{x}$. The test statistic is $\bar{X}$, and the sampling distribution of means is considered.

There are now several cases that may occur, depending on whether the population is normal or not, whether the sample size is large or small and whether the population variance is known or not.

Test 1a: Testing μ when the population X is normal and the variance σ^2 is known (any size sample)

Since the population is normally distributed, $X \sim N(\mu, \sigma^2)$. The sampling distribution of means, $\bar{X}$ is also normal for *all* sample sizes, with mean μ_0 (as specified in the null hypothesis).

When testing the mean of a normal population X with known variance σ^2 for samples of size n, the test statistic is

$$\bar{X}, \text{ where } \bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right).$$

In standardised form, the test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \text{ where } Z \sim N(0, 1).$$

Example 1

Each year trainees throughout the country sit a test. Over a period of time it has been established that the marks can be modelled by a normal distribution with mean 70 and standard deviation 6.

This year it was thought that trainees from a particular county did not perform as well as expected. The marks of a random sample of 25 trainees from the county were scrutinised and it was found that their mean mark was 67.3.

Does this provide evidence, at the 5% significance level, that trainees from this county did not perform well as expected?

Example 2

A sample of size 16 is taken from the distribution of $X \sim N(\mu, 3^2)$ and a hypothesis test is carried out at the 1% level of significance. On the basis of the value of the sample mean $\bar{x}$, the null hypothesis $\mu = 100$ is rejected in favour of the alternative hypothesis $\mu > 100$.

What can be said about the value of $\bar{x}$?

Test 1b: Testing μ when the population X is not normal, the variance σ^2 is known and the sample size n is large

Since the population is not normal, you cannot say that the distribution of $\overline{X}$ is normal for all sample sizes. If the sample size n is large, however, you can apply the central limit theorem (see page 442). This states that for large samples taken from a non-normal population, the sampling distribution of means $\overline{X}$ is *approximately* normal, whatever the distribution of the parent population.

When testing the mean of a non-normal population X with known variance σ^2, provided that the sample size n is large,

the test statistic is $\overline{X}$, where $\overline{X}$ is approximately normal, $\overline{X} \sim N\left(\mu_0, \frac{\sigma}{n}\right)$.

In standardised form,

the test statistic is $Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ where $Z \sim N(0, 1)$.

Example 3

The management of a large hospital states that the mean age of its patients is 45 years. Records of a random sample of 100 patients give a mean age of 48.4 years. Using a population standard deviation of 18 years, test at the 5% significance level whether there is evidence that the management's statement is incorrect. State clearly your null and alternative hypotheses. (C)

Test 1c: Testing the mean μ of a population X where the variance σ^2 is unknown and the sample size n is large

The variance of the population, σ^2, is unknown, but, as you saw in on page 447, an unbiased estimate, $\hat{\sigma}^2$ can be used instead,

where $\hat{\sigma}^2 = \frac{n}{n-1} \times s^2$ (s^2 is the sample variance)

Alternative formats:

$$\hat{\sigma}^2 = \frac{1}{n-1}\Sigma(x-\bar{x})^2 \quad \text{or} \quad \hat{\sigma}^2 = \frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)$$

Ideally the population distribution should be normal, but if it is not, then the central limit theorem can be applied, since the sample size is large.

When testing the mean of a population X with unknown variance σ^2, provided the sample size n is large,

the test statistic is $\overline{X}$ where $\overline{X} \sim N\left(\mu_0, \frac{\hat{\sigma}^2}{n}\right)$.

In standardised form,

the test statistic $Z = \frac{\overline{X} - \mu_0}{\hat{\sigma}/\sqrt{n}}$, where $Z \sim N(0, 1)$.

Example 4

The packaging on an electric light bulb states that the average length of life of bulbs is 1000 hours. A consumer association thinks that this is an overestimate and tests a random sample of 64 bulbs, recording the life x hours, of each bulb.

The results are summarised as follows:

$\Sigma x = 63\,910.4, \qquad \Sigma x^2 = 63\,824\,061.$

(a) Calculate the sample mean, $\bar{x}$.
(b) Calculate an unbiased estimate for the standard deviation of the length of life of all light bulbs of this type.
(c) Is there evidence, at the 10% significance level, that the statement on the packaging is overestimating the length of life of this type of light bulb?

TYPE I AND TYPE II ERRORS

When you make your decision about whether or not to reject H_0 there are two types of error that could be made. These were described in Chapter 10 (page 493) and are called Type I and Type II errors:

- a Type I error is made when you wrongly reject a true hypothesis,
- a Type II error is made when you wrongly accept a false hypothesis.

These can be summarised in a table:

		Test decision	
		Accept H_0	Reject H_0
Actual situation	H_0 is true	✓(correct)	Type I error
	H_0 is false	Type II error	✓ (correct)

Making an incorrect decision can be costly in various ways. For example, suppose that a fire alarm was tested to see whether it was still functioning correctly after a power cut. You might take as the null and alternative hypotheses

H_0: the alarm is functioning correctly,
H_1: the alarm is not functioning correctly.

A Type II error in this situation would mean that you assumed that the alarm was functioning correctly when in fact it was not. This could result in injury, loss of life or damage to property. A Type I error would mean that you thought the alarm was not working correctly when in fact it was. This could mean expenditure on unnecessary repairs or replacement.

When the distribution of the test statistic is continuous, P(Type I error) is equal to the significance level of the test.

The choice of significance level for a hypothesis test is thus related to the value of P(Type I error) which you are prepared to accept. The choice of a significance level should depend in the first instance on how serious the consequences of a Type I error are. The more serious the consequences, the lower the value of the significance level which should be used. For example, if the consequences of a Type I error were not serious, you might use a significance level of 10%; if the consequences were very serious you might use a significance level of 0.1%.

A Type II error involves accepting a false null hypothesis, which means that you fail to detect a difference in μ. You would expect the probability of this happening to depend on how much μ has changed: if there is a small difference in μ it could easily go undetected but if there is a big difference in μ then you would expect to detect it. This is why the alternative hypothesis has to be defined more exactly before P(Type II error) can be calculated.

Example 5

A machine fills 'one litre' water bottles. When the machine is working correctly the contents of the bottles are normally distributed with mean 1.002 litres and standard deviation 0.002 litres. The performance of the machine is tested at regular intervals by taking a sample of 9 bottles and calculating their mean content. If this mean content falls below a certain value, it is assumed that the machine is not performing correctly and it is stopped.

(a) Set up null and alternative hypotheses for a test of whether the machine is working correctly.
(b) For a test at the 5% significance level, find the rejection region taking the sample mean as the test statistic.
(c) Give the value for the probability of a Type I error.
(d) Find P(Type II error) if the mean content of the bottles has fallen to the nominal value of 1.000 litre.
(e) Find the range of values of μ for which the probability of making a Type II error is less than 0.001.

Example 6

Boxes of dried haricot beans have contents whose masses are normally distributed with mean μ and standard deviation 15 grams. A test of the null hypothesis $\mu = 375$ against the alternative hypothesis $\mu \neq 375$ is carried out at the 5% significance level using a random sample of 16 boxes.

(a) For what values of the sample mean is the alternative hypothesis accepted?
(b) Given that the actual value of μ is 380, find the probability of making a Type II error. (OCR, adapted)

Example 7

A study of a large sample of books by a particular author shows that the number of words per sentence can be modelled by a normal distribution with mean 21.2 and standard deviation 7.3. A researcher claims to have discovered a previously unknown book by this author. The mean length of 90 sentences chosen at random in this book is found to be 19.4 words.

(i) Assuming the population standard deviation of sentence lengths in this book is also 7.3, test at the 5% level of significance whether the mean sentence length is the same as the author's. State your null and alternative hypotheses.

(ii) State in words relating to the context of the test what is meant by a Type I error and state the probability of a Type I error in the test in part **(i)**.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q4 June 2005]

Example 8

The number of cars caught speeding on a certain length of motorway is 7.2 per day, on average. Speed cameras are introduced and the results shown in the following table are those from a random selection of 40 days after this.

Number of cars caught speeding	4	5	6	7	8	9	10
Number of days	5	7	8	10	5	2	3

(i) Calculate unbiased estimates of the population mean and variance of the number of cars per day caught speeding after the speed cameras were introduced.

(ii) Taking the null hypothesis H_0 to be $\mu = 7.2$, test at the 5% level whether there is evidence that the introduction of speed cameras has resulted in a reduction in the number of cars caught speeding.

(iii) State what is meant by Type I error in words relating to the context of the test in part **(ii)**. Without further calculation, illustrate on a suitable diagram the region representing the probability of this Type I error.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q7 June 2006]

Example 9

A machine has produced nails over a long period of time, where the length in millimetres was distributed as N(22.0, 0.19). It is believed that recently the mean length has changed. To test this belief a random sample of 8 nails is taken and the mean length is found to be 21.7 mm. Carry out a hypothesis test at the 5% significance level to test whether the population mean has changed, assuming that the variance remains the same.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q3 June 2007]

Example 10

In summer the growth rate of grass in a lawn has a normal distribution with mean 3.2 cm per week and standard deviation 1.4 cm per week. A new type of grass is introduced which the manufacturer claims has a slower growth rate. A hypothesis test of this claim at the 5% significance level was carried out using a random sample of 10 lawns that had the new grass. It may be assumed that the growth rate of the new grass has a normal distribution with standard deviation 1.4 cm per week.

(i) Find the rejection region for the test.

(ii) The probability of making a Type II error when the actual value of the mean growth rate of the new grass is m cm per week is less than 0.5. Use your answer to part **(i)** to write down an inequality for m.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q2 November 2007]

Homework: Hypothesis testing using Normal Distribution– Variant 62

1 Over a long period of time it is found that the time spent at cash withdrawal points follows a normal distribution with mean 2.1 minutes and standard deviation 0.9 minutes. A new system is tried out, to speed up the procedure. The null hypothesis is that the mean time spent is the same under the new system as previously. It is decided to reject the null hypothesis and accept that the new system is quicker if the mean withdrawal time from a random sample of 20 cash withdrawals is less than 1.7 minutes. Assume that, for the new system, the standard deviation is still 0.9 minutes, and the time spent still follows a normal distribution.

(i) Calculate the probability of a Type I error. [4]

(ii) If the mean withdrawal time under the new system is actually 1.5 minutes, calculate the probability of a Type II error. [4]

Answer: (i) 0.0234; (ii) 0.160 . **J03/Q5**

2 Each multiple choice question in a test has 4 suggested answers, exactly one of which is correct. Rehka knows nothing about the subject of the test, but claims that she has a special method for answering the questions that is better than just guessing. There are 60 questions in the test, and Rehka gets 22 correct.

(i) State null and alternative hypotheses for a test of Rehka's claim. [1]

(ii) Using a normal approximation, test at the 5% significance level whether Rehka's claim is justified. [4]

Answers: (i) H_0: $\mu = 15$ or $p = 0.25$, H_1: $\mu > 15$ or $p > 0.25$; (ii) Claim justified. **J04/Q1**

3 The lectures in a mathematics department are scheduled to last 54 minutes, and the times of individual lectures may be assumed to have a normal distribution with mean μ minutes and standard deviation 3.1 minutes. One of the students commented that, on average, the lectures seemed too short. To investigate this, the times for a random sample of 10 lectures were used to test the null hypothesis $\mu = 54$ against the alternative hypothesis $\mu < 54$ at the 10% significance level.

(i) Show that the null hypothesis is rejected in favour of the alternative hypothesis if $\bar{x} < 52.74$, where $\bar{x}$ minutes is the sample mean. [4]

(ii) Find the probability of a Type II error given that the actual mean length of lectures is 51.5 minutes. [4]

Answer: (ii) 0.103. **J04/Q5**

4 A study of a large sample of books by a particular author shows that the number of words per sentence can be modelled by a normal distribution with mean 21.2 and standard deviation 7.3. A researcher claims to have discovered a previously unknown book by this author. The mean length of 90 sentences chosen at random in this book is found to be 19.4 words.

(i) Assuming the population standard deviation of sentence lengths in this book is also 7.3, test at the 5% level of significance whether the mean sentence length is the same as the author's. State your null and alternative hypotheses. [5]

(ii) State in words relating to the context of the test what is meant by a Type I error and state the probability of a Type I error in the test in part **(i)**. [2]

Answers: **(i)** H_0: $\mu = 21.2$, H_1: $\mu \neq 21.2$, Significant evidence to say not the same sentence length (o author); **(ii)** Say it is not the same sentence length (or author) when it is, 5%. **J05/Q4**

5 The number of cars caught speeding on a certain length of motorway is 7.2 per day, on average. Speed cameras are introduced and the results shown in the following table are those from a random selection of 40 days after this.

Number of cars caught speeding	4	5	6	7	8	9	10
Number of days	5	7	8	10	5	2	3

(i) Calculate unbiased estimates of the population mean and variance of the number of cars per day caught speeding after the speed cameras were introduced. [3]

(ii) Taking the null hypothesis H_0 to be $\mu = 7.2$, test at the 5% level whether there is evidence that the introduction of speed cameras has resulted in a reduction in the number of cars caught speeding. [5]

(iii) State what is meant by a Type I error in words relating to the context of the test in part **(ii)**. Without further calculation, illustrate on a suitable diagram the region representing the probability of this Type I error. [3]

Answers: **(i)** 6.53, 2.87; **(iii)** Say there is a reduction in the number of cars caught speeding when there not. **J06/Q7**

6 A machine has produced nails over a long period of time, where the length in millimetres was distributed as N(22.0, 0.19). It is believed that recently the mean length has changed. To test this belief a random sample of 8 nails is taken and the mean length is found to be 21.7 mm. Carry out a hypothesis test at the 5% significance level to test whether the population mean has changed, assuming that the variance remains the same. [5]

Answer: Not enough evidence to say that the mean has changed. **J07/Q3**

7 People who diet can expect to lose an average of 3 kg in a month. In a book, the authors claim that people who follow a new diet will lose an average of more than 3 kg in a month. The weight losses of the 180 people in a random sample who had followed the new diet for a month were noted. The mean was 3.3 kg and the standard deviation was 2.8 kg.

(i) Test the authors' claim at the 5% significance level, stating your null and alternative hypotheses. [5]

(ii) State what is meant by a Type II error in words relating to the context of the test in part **(i)**. [2]

Answers: **(i)** $H_0 : \mu = 3$, $H_1 : \mu > 3$, Not enough evidence to support the claim; **(ii)** Say no extra weig loss when there is. **J08/Q4**

8 In Europe the diameters of women's rings have mean 18.5 mm. Researchers claim that women in Jakarta have smaller fingers than women in Europe. The researchers took a random sample of 20 women in Jakarta and measured the diameters of their rings. The mean diameter was found to be 18.1 mm. Assuming that the diameters of women's rings in Jakarta have a normal distribution with standard deviation 1.1 mm, carry out a hypothesis test at the $2\frac{1}{2}\%$ level to determine whether the researchers' claim is justified. [5]

Answer: Not enough evidence to support the claim that fingers are smaller. **J09/71/Q1**

9 The mean car engine size in Europe is known to be 1.746 litres. In Mauritius, a random sample of 230 cars was found to have a mean engine size of 1.765 litres. Assuming that the standard deviation of car engine sizes in Mauritius is 0.149 litres, test at the 10% significance level whether there is a difference between the mean engine sizes of cars in Europe and those in Mauritius. [5]

Answer: Evidence of a difference. **J09/72/Q1**

10 Metal bolts are produced in large numbers and have lengths which are normally distributed with mean 2.62 cm and standard deviation 0.30 cm.

(i) Find the probability that a random sample of 45 bolts will have a mean length of more than 2.55 cm. [3]

(ii) The machine making these bolts is given an annual service. This may change the mean length of bolts produced but does not change the standard deviation. To test whether the mean has changed, a random sample of 30 bolts is taken and their lengths noted. The sample mean length is m cm. Find the set of values of m which result in rejection at the 10% significance level of the hypothesis that no change in the mean length has occurred. [4]

Answers: **(i)** 0.941; **(ii)** $m < 2.53$, $m > 2.71$. **J10/72/Q3**

11 From previous years' observations, the lengths of salmon in a river were found to be normally distributed with mean 65 cm. A researcher suspects that pollution in water is restricting growth. To test this theory, she measures the length x cm of a random sample of n salmon and calculates that $\bar{x} = 64.3$ and $s = 4.9$, where s^2 is the unbiased estimate of the population variance. She then carries out an appropriate hypothesis test.

(i) Her test statistic z has a value of -1.807 correct to 3 decimal places. Calculate the value of n. [3]

(ii) Using this test statistic, carry out the hypothesis test at the 5% level of significance and state what her conclusion should be. [4]

Answers: **(i)** $n = 160$; **(ii)** Significant growth decrease. **N02/Q3**

12 The distance driven in a week by a long-distance lorry driver is a normally distributed random variable with mean 1850 km and standard deviation 117 km.

(i) Find the probability that in a random sample of 26 weeks his average distance driven per week is more than 1800 km. [3]

(ii) New driving regulations are introduced and in a random sample of 26 weeks after their introduction the lorry driver drives a total of 47 658 km. Assuming the standard deviation remains unchanged, test at the 10% level whether his mean weekly driving distance has changed. [5]

Answers: **(i)** 0.985; **(ii)** No significant change. **N03/Q5**

13 Flies stick to wet paint at random points. The average number of flies is 2 per square metre. A wall with area 22 m^2 is painted with a new type of paint which the manufacturer claims is fly-repellent. It is found that 27 flies stick to this wall. Use a suitable approximation to test the manufacturer's claim at the 1% significance level. Take the null hypothesis to be $\mu = 44$, where μ is the population mean. [5]

Answer: Claim justified. **N05/Q3**

14 The time taken for Samuel to drive home from work is distributed with mean 46 minutes. Samuel discovers a different route and decides to test at the 5% level whether the mean time has changed. He tries this route on a large number of different days chosen randomly and calculates the mean time.

(i) State the null and alternative hypotheses for this test. [1]

(ii) Samuel calculates the value of his test statistic z to be -1.729. What conclusion can he draw? [2]

Answers: **(i)** H_0: $\mu = 46$, H_1: $\mu \neq 46$; **(ii)** No significant difference in times. **N06/Q1**

15 In summer the growth rate of grass in a lawn has a normal distribution with mean 3.2 cm per week and standard deviation 1.4 cm per week. A new type of grass is introduced which the manufacturer claims has a slower growth rate. A hypothesis test of this claim at the 5% significance level was carried out using a random sample of 10 lawns that had the new grass. It may be assumed that the growth rate of the new grass has a normal distribution with standard deviation 1.4 cm per week.

(i) Find the rejection region for the test. [4]

(ii) The probability of making a Type II error when the actual value of the mean growth rate of the new grass is m cm per week is less than 0.5. Use your answer to part **(i)** to write down an inequality for m. [1]

Answers: (i) $x < 2.47$; (ii) $m < 2.47$. **N07/Q2**

---

16 The times taken for the pupils in Ming's year group to do their English homework have a normal distribution with standard deviation 15.7 minutes. A teacher estimates that the mean time is 42 minutes. The times taken by a random sample of 3 students from the year group were 27, 35 and 43 minutes. Carry out a hypothesis test at the 10% significance level to determine whether the teacher's estimate for the mean should be accepted, stating the null and alternative hypotheses. [5]

-

---

17 Photographers often need to take many photographs of families until they find a photograph which everyone in the family likes. The number of photographs taken until obtaining one which everybody likes has mean 15.2. A new photographer claims that she can obtain a photograph which everybody likes with fewer photographs taken. To test at the 10% level of significance whether this claim is justified, the numbers of photographs, x, taken by the new photographer with a random sample of 60 families are recorded. The results are summarised by $\Sigma x = 890$ and $\Sigma x^2 = 13\,780$.

(i) Calculate unbiased estimates of the population mean and variance of the number of photographs taken by the new photographer. [3]

(ii) State null and alternative hypotheses for the test, and state also the probability that the test results in a Type I error. Say what a Type I error means in the context of the question. [3]

(iii) Carry out the test. [4]

Answers: (i) 14.8, 9.80; (ii) $H_0: \mu = 15.2$, $H_1: \mu < 15.2$, 0.1. **N09/71/Q6**

---

18 The masses of packets of cornflakes are normally distributed with standard deviation 11 g. A random sample of 20 packets was weighed and found to have a mean mass of 746 g.

(i) Test at the 4% significance level whether there is enough evidence to conclude that the population mean mass is less than 750 g. [4]

(ii) Given that the population mean mass actually is 750 g, find the smallest possible sample size, n, for which it is at least 97% certain that the mean mass of the sample exceeds 745 g. [4]

Answer: (ii) 18. **N09/72/Q5**

19 The weights, X kilograms, of bags of carrots are normally distributed. The mean of X is μ. An inspector wishes to test whether $\mu = 2.0$. He weighs a random sample of 200 bags and his results are summarised as follows.

$$\Sigma x = 430 \qquad \Sigma x^2 = 1290$$

(i) Carry out the test, at the 10% significance level. [6]

(ii) You may now assume that the population variance of X is 1.85. The inspector weighs another random sample of 200 bags and carries out the same test at the 10% significance level.

(a) State the meaning of a Type II error in this context. [1]

(b) Given that $\mu = 2.12$, show that the probability of a Type II error is 0.652, correct to 3 significant figures. [7]

Answers: **(i)** No evidence that $\mu \neq 2.0$ **(ii)** Concluding that $\mu = 2.0$ although not true **J12/72/Q7**

---

20 The heights of a certain variety of plant have been found to be normally distributed with mean 75.2 cm and standard deviation 5.7 cm. A biologist suspects that pollution in a certain region is causing the plants to be shorter than usual. He takes a random sample of n plants of this variety from this region and finds that their mean height is 73.1 cm. He then carries out an appropriate hypothesis test.

(i) He finds that the value of the test statistic z is -1.563, correct to 3 decimal places. Calculate the value of n. State an assumption necessary for your calculation. [4]

(ii) Use this value of the test statistic to carry out the hypothesis test at the 6% significance level. [3]

Answer: n=18 assume s.d. for the region is 5.7
Evidence that plants are shorter **J13/72/Q3**

---

21 The number of calls per day to an enquiry desk has a Poisson distribution. In the past the mean has been 5. In order to test whether the mean has changed, the number of calls on a random sample of 10 days was recorded. The total number of calls was found to be 61. Use an approximate distribution to test at the 10% significance level whether the mean has changed. [5]

Answer: No evidence that the mean has changed **J14/72/Q3**

---

22 A researcher is investigating the actual lengths of time that patients spend with the doctor at their appointments. He plans to choose a sample of 12 appointments on a particular day.

(i) Which of the following methods is preferable, and why?

- Choose the first 12 appointments of the day.
- Choose 12 appointments evenly spaced throughout the day. [2]

Appointments are scheduled to last 10 minutes. The actual lengths of time, in minutes, that patients spend with the doctor may be assumed to have a normal distribution with mean μ and standard deviation 3.4. The researcher suspects that the actual time spent is more than 10 minutes on average. To test this suspicion, he recorded the actual times spent for a random sample of 12 appointments and carried out a hypothesis test at the 1% significance level.

(ii) State the probability of making a Type I error and explain what is meant by a Type I error in this context. [2]

(iii) Given that the total length of time spent for the 12 appointments was 147 minutes, carry out the test. [5]

(iv) Give a reason why the Central Limit theorem was not needed in part **(iii)**. [1]

J14/72/Q7

Answers: 2^{nd} is more representative of *all* appointments
0.01 Concluding that times spent are too long when they are not
No reason to believe that appointments are too long
Normal population

---

23 In the past, the flight time, in hours, for a particular flight has had mean 6.20 and standard deviation 0.80. Some new regulations are introduced. In order to test whether these new regulations have had any effect upon flight times, the mean flight time for a random sample of 40 of these flights is found.

(i) State what is meant by a Type I error in this context. [2]

(ii) The mean time for the sample of 40 flights is found to be 5.98 hours. Assuming that the standard deviation of flight times is still 0.80 hours, test at the 5% significance level whether the population mean flight time has changed. [4]

(iii) State, with a reason, which of the errors, Type I or Type II, might have been made in your answer to part **(ii)**. [2]

J15/72/Q4

Answers: (i) Conclude that flight times affected, when in fact they have not been.
(ii) No evidence that flight times affected
(iii) H_0 was not rejected. Type II.

---

24 An engineering test consists of 100 multiple-choice questions. Each question has 5 suggested answers, only one of which is correct. Ashok knows nothing about engineering, but he claims that his general knowledge enables him to get more questions correct than just by guessing. Ashok actually gets 27 answers correct. Use a suitable approximating distribution to test at the 5% significance level whether his claim is justified. [5]

J15/72/Q2

---

25 The management of a factory thinks that the mean time required to complete a particular task is 22 minutes. The times, in minutes, taken by employees to complete this task have a normal distribution with mean μ and standard deviation 3.5. An employee claims that 22 minutes is not long enough for the task. In order to investigate this claim, the times for a random sample of 12 employees are used to test the null hypothesis $\mu = 22$ against the alternative hypothesis $\mu > 22$ at the 5% significance level.

(i) Show that the null hypothesis is rejected in favour of the alternative hypothesis if $\bar{x} > 23.7$ (correct to 3 significant figures), where $\bar{x}$ is the sample mean. [3]

(ii) Find the probability of a Type II error given that the actual mean time is 25.8 minutes. [4]

Answer: (ii) 0.0172

N11/72/Q5

26 The heights of a certain type of plant have a normal distribution. When the plants are grown without fertilizer, the population mean and standard deviation are 24.0 cm and 4.8 cm respectively. A gardener wishes to test, at the 2% significance level, whether Hiergro fertilizer will increase the mean height. He treats 150 randomly chosen plants with Hiergro and finds that their mean height is 25.0 cm. Assuming that the standard deviation of the heights of plants treated with Hiergro is still 4.8 cm, carry out the test. [5]

Answer: There is evidence that Hiergo has increased the heights. **N12/72/Q2**

27 Following a change in flight schedules, an airline pilot wished to test whether the mean distance that he flies in a week has changed. He noted the distances, x km, that he flew in 50 randomly chosen weeks and summarised the results as follows.

$$n = 50 \qquad \Sigma x = 143\,300 \qquad \Sigma x^2 = 410\,900\,000$$

(i) Calculate unbiased estimates of the population mean and variance. [3]

(ii) In the past, the mean distance that he flew in a week was 2850 km. Test, at the 5% significance level, whether the mean distance has changed. [5]

Answer: 2866 4126.53 **N13/72/Q3**
No evidence that the mean distance changed.

28 The number of hours that Mrs Hughes spends on her business in a week is normally distributed with mean μ and standard deviation 4.8. In the past the value of μ has been 49.5.

(i) Assuming that μ is still equal to 49.5, find the probability that in a random sample of 40 weeks the mean time spent on her business in a week is more than 50.3 hours. [4]

Following a change in her arrangements, Mrs Hughes wishes to test whether μ has decreased. She chooses a random sample of 40 weeks and notes that the total number of hours she spent on her business during these weeks is 1920.

(ii) **(a)** Explain why a one-tail test is appropriate. [1]

(b) Carry out the test at the 6% significance level. [4]

(c) Explain whether it was necessary to use the Central Limit theorem in part **(ii)(b)**. [1]

Answer: 0.146; looking for a decrease; there is evidence that the mean time has decreased; no, th **N14/72/Q5**
population is normally distributed

Homework: Hypothesis testing using Normal Distribution– Variant 71 & 73

1 Dipak carries out a test, at the 10% significance level, using a normal distribution. The null hypothesis is $\mu = 35$ and the alternative hypothesis is $\mu \neq 35$.

(i) Is this a one-tail or a two-tail test? State briefly how you can tell. [1]

Dipak finds that the value of the test statistic is $z = -1.750$.

(ii) Explain what conclusion he should draw. [2]

(iii) This result is significant at the $\alpha\%$ level. Find the smallest possible value of α, correct to the nearest whole number. [2]

Answers: **(i)** two-tailed test; **(iii)** 8. **J10/73/Q2**

---

2 Past experience has shown that the heights of a certain variety of rose bush have been normally distributed with mean 85.0 cm. A new fertiliser is used and it is hoped that this will increase the heights. In order to test whether this is the case, a botanist records the heights, x cm, of a large random sample of n rose bushes and calculates that $\bar{x} = 85.7$ and $s = 4.8$, where $\bar{x}$ is the sample mean and s^2 is an unbiased estimate of the population variance. The botanist then carries out an appropriate hypothesis test.

(i) The test statistic, z, has a value of 1.786 correct to 3 decimal places. Calculate the value of n. [3]

(ii) Using this value of the test statistic, carry out the test at the 5% significance level. [3]

Answers: **(i)** 150; **(ii)** Evidence that the mean has increased. **J11/71/Q3**

---

3 Previous records have shown that the number of cars entering Bampor on any day has mean 352 and variance 121.

(i) Find the probability that the mean number of cars entering Bampor during a random sample of 200 days is more than 354. [4]

(ii) State, with a reason, whether it was necessary to assume that the number of cars entering Bampor on any day has a normal distribution in order to find the probability in part **(i)**. [2]

(iii) It is thought that the population mean may recently have changed. The number of cars entering Bampor during the day was recorded for each of a random sample of 50 days and the sample mean was found to be 356. Assuming that the variance is unchanged, test at the 5% significance level whether the population mean is still 352. [4]

Answers: **(i)** 0.0051; **(ii)** No, since n is large and CLT applies, or sample mean approximately normal distributed; **(iii)** Evidence that the population mean has changed. **J11/73/Q7**

4 A survey taken last year showed that the mean number of computers per household in Branley was 1.66. This year a random sample of 50 households in Branley answered a questionnaire with the following results.

Number of computers	0	1	2	3	4	> 4
Number of households	5	12	18	10	5	0

(i) Calculate unbiased estimates for the population mean and variance of the number of computers per household in Branley this year. [3]

(ii) Test at the 5% significance level whether the mean number of computers per household has changed since last year. [5]

(iii) Explain whether it is possible that a Type I error may have been made in the test in part **(ii)**. [1]

(iv) State what is meant by a Type II error in the context of the test in part **(ii)**, and give the set of values of the test statistic that could lead to a Type II error being made. [2]

Answer. **(i)** Mean = 1.96 Variance = 1.26(37) **(ii)** No evidence that the mean has changed **(iii)** No because H_0 was not rejected **(iv)** State mean has not changed when it has $-1.96<\text{test statistic}<1.96$ **J12/71/Q6**

5 Last year Samir found that the time for his journey to work had mean 45.7 minutes and standard deviation 3.2 minutes. Samir wishes to test whether his journey times have increased this year. He notes the times, in minutes, for a random sample of 8 journeys this year with the following results.

46.2 41.7 49.2 47.1 47.2 48.4 53.7 45.5

It may be assumed that the population of this year's journey times is normally distributed with standard deviation 3.2 minutes.

(i) State, with a reason, whether Samir should use a one-tail or a two-tail test. [2]

(ii) Show that there is no evidence at the 5% significance level that Samir's mean journey time has increased. [5]

(iii) State, with a reason, which one of the errors, Type I or Type II, might have been made in carrying out the test in part **(ii)**. [2]

Answers. **(i)** One-tailed as testing for an increase **(ii)** Accept H_0 based on a value of z in the range 1.481 1.503 **(iii)** Type II as H_0 accepted **J12/73/Q6**

6 The times taken by students to complete a task are normally distributed with standard deviation 2.4 minutes. A lecturer claims that the mean time is 17.0 minutes. The times taken by a random sample of 5 students were 17.8, 22.4, 16.3, 23.1 and 11.4 minutes. Carry out a hypothesis test at the 5% significance level to determine whether the lecturer's claim should be accepted. [5]

Answer. Claim cannot be accepted **J13/71/Q2**

7 In the past the weekly profit at a store had mean \$34 600 and standard deviation \$4500. Following a change of ownership, the mean weekly profit for 90 randomly chosen weeks was \$35 400.

(i) Stating a necessary assumption, test at the 5% significance level whether the mean weekly profit has increased. [6]

(ii) State, with a reason, whether it was necessary to use the Central Limit theorem in part **(i)**. [2]

The mean weekly profit for another random sample of 90 weeks is found and the same test is carried out at the 5% significance level.

(iii) State the probability of a Type I error. [1]

(iv) Given that the population mean weekly profit is now \$36 500, calculate the probability of a Type II error. [5]

Answers: (i) Evidence of an increase in profit (ii) As the distribution of weekly mean profit was unknov the CLT was required. (iii) 5% (iv) 0.0091 **J13/73/Q7**

8 The lengths, in centimetres, of rods produced in a factory have mean μ and standard deviation 0.2. The value of μ is supposed to be 250, but a manager claims that one machine is producing rods that are too long on average. A random sample of 40 rods from this machine is taken and the sample mean length is found to be 250.06 cm. Test at the 5% significance level whether the manager's claim is justified. [5]

Answer: No evidence that the mean has changed **J14/71/Q3**

9 The weights, X kilograms, of rabbits in a certain area have population mean μ kg. A random sample of 100 rabbits from this area was taken and the weights are summarised by

$$\Sigma x = 165, \qquad \Sigma x^2 = 276.25.$$

Test at the 5% significance level the null hypothesis $H_0 : \mu = 1.6$ against the alternative hypothesis $H_1 : \mu \neq 1.6$. [6]

Answer: Reject H_0. Evidence that $\mu \neq 1.6$ **J14/73/Q4**

10 A machine is designed to generate random digits between 1 and 5 inclusive. Each digit is supposed to appear with the same probability as the others, but Max claims that the digit 5 is appearing less often than it should. In order to test this claim the manufacturer uses the machine to generate 25 digits and finds that exactly 1 of these digits is a 5.

(i) Carry out a test of Max's claim at the 2.5% significance level. [5]

(ii) Max carried out a similar hypothesis test by generating 1000 digits between 1 and 5 inclusive. The digit 5 appeared 180 times. Without carrying out the test, state the distribution that Max should use, including the values of any parameters. [2]

(iii) State what is meant by a Type II error in this context. [1]

Answers: (i) Accept H_0. No evidence of fewer 5s (ii) N(200,180) (iii) Conclude that the machine produces an equal number of 5s when it actually produces fewer **J14/73/Q6**

11 Sami claims that he can read minds. He asks each of 50 people to choose one of the 5 letters A, B, C, D or E. He then tells each person which letter he believes they have chosen. He gets 13 correct. Sami says "This shows that I can read minds, because 13 is more than I would have got right if I were just guessing."

(i) State null and alternative hypotheses for a test of Sami's claim. [1]

(ii) Test at the 10% significance level whether Sami's claim is justified. [5]

Answers: **(i)** H_0 p=0.2, H_1 p>0.2 **(ii)** Not significant, insufficient evidence that Sami can read min **J15/71/Q2**

12 In the past, the time taken by vehicles to drive along a particular stretch of road has had mean 12.4 minutes and standard deviation 2.1 minutes. Some new signs are installed and it is expected that the mean time will increase. In order to test whether this is the case, the mean time for a random sample of 50 vehicles is found. You may assume that the standard deviation is unchanged.

(i) The mean time for the sample of 50 vehicles is found to be 12.9 minutes. Test at the 2.5% significance level whether the population mean time has increased. [4]

(ii) State what is meant by a Type II error in this context. [2]

(iii) State what extra piece of information would be needed in order to find the probability of a Type II error. [1]

Answers: **(i)** Not significant, no evidence of an increase in mean. **(ii)** Concluding mean time is still 12. **J15/71/Q4**
when it has increased. **(iii)** The value of the new mean.

13 The mean breaking strength of cables made at a certain factory is supposed to be 5 tonnes. The quality control department wishes to test whether the mean breaking strength of cables made by a particular machine is actually less than it should be. They take a random sample of 60 cables. For each cable they find the breaking strength by gradually increasing the tension in the cable and noting the tension when the cable breaks.

(i) Give a reason why it is necessary to take a sample rather then testing all the cables produced by the machine. [1]

(ii) The mean breaking strength of the 60 cables in the sample is found to be 4.95 tonnes. Given that the population standard deviation of breaking strengths is 0.15 tonnes, test at the 1% significance level whether the population mean breaking strength is less than it should be. [4]

(iii) Explain whether it was necessary to use the Central Limit theorem in the solution to part **(ii)**. [2]

Answers: **(i)** Testing destructive or too time consuming **(ii)** Test significant so evidence of drop in mea **J15/73/Q5**
(iii) Population distribution unknown so yes.

14 The marks of candidates in Mathematics and English in 2009 were represented by the independent random variables X and Y with distributions $N(28, 5.6^2)$ and $N(52, 12.4^2)$ respectively. Each candidate's marks were combined to give a final mark F, where $F = X + \frac{1}{2}Y$.

(i) Find $E(F)$ and $Var(F)$. [3]

(ii) The final marks of a random sample of 10 candidates from Grinford in 2009 had a mean of 49. Test at the 5% significance level whether this result suggests that the mean final mark of all candidates from Grinford in 2009 was lower than elsewhere. [5]

Answers: **(i)** 54, 69.8. **N10/71/Q5**

15 A clinic monitors the amount, X milligrams per litre, of a certain chemical in the blood stream of patients. For patients who are taking drug A, it has been found that the mean value of X is 0.336. A random sample of 100 patients taking a new drug, B, was selected and the values of X were found. The results are summarised below.

$$n = 100, \qquad \Sigma x = 43.5, \qquad \Sigma x^2 = 31.56.$$

(i) Test at the 1% significance level whether the mean amount of the chemical in the blood stream of patients taking drug B is different from that of patients taking drug A. [8]

(ii) For the test to be valid, is it necessary to assume a normal distribution for the amount of chemical in the blood stream of patients taking drug B? Justify your answer. [2]

Answer: **(ii)** No, n is large and Central Limit Theorem applies. **N10/73/Q6**

16 An engineering test consists of 100 multiple-choice questions. Each question has 5 suggested answers, only one of which is correct. Ashok knows nothing about engineering, but he claims that his general knowledge enables him to get more questions correct than just by guessing. Ashok actually gets 27 answers correct. Use a suitable approximating distribution to test at the 5% significance level whether his claim is justified. [5]

- *Answer*: **(i)** Mean=2.6, Variance=5.2; **(ii)** Variance≠Mean **N11/71/Q2**

17 The management of a factory thinks that the mean time required to complete a particular task is 22 minutes. The times, in minutes, taken by employees to complete this task have a normal distribution with mean μ and standard deviation 3.5. An employee claims that 22 minutes is not long enough for the task. In order to investigate this claim, the times for a random sample of 12 employees are used to test the null hypothesis $\mu = 22$ against the alternative hypothesis $\mu > 22$ at the 5% significance level.

(i) Show that the null hypothesis is rejected in favour of the alternative hypothesis if $\bar{x} > 23.7$ (correct to 3 significant figures), where $\bar{x}$ is the sample mean. [3]

(ii) Find the probability of a Type II error given that the actual mean time is 25.8 minutes. [4]

Answer: **(ii)** 0.0172 **N11/71/Q5**

18 Records show that the distance driven by a bus driver in a week is normally distributed with mean 1150 km and standard deviation 105 km. New driving regulations are introduced and in the next 20 weeks he drives a total of 21 800 km.

(i) Stating any assumption(s), test, at the 1% significance level, whether his mean weekly driving distance has decreased. [6]

(ii) A similar test at the 1% significance level was carried out using the data from another 20 weeks. State the probability of a Type I error and describe what is meant by a Type I error in this context. [2]

Answers: **(i)** There is evidence that the mean distance has decreased; **(ii)** 0.01, Concluding that the me: **N11/73/Q5**
driving distance has decreased when in fact it has not.

---

19 The heights of a certain type of plant have a normal distribution. When the plants are grown without fertilizer, the population mean and standard deviation are 24.0 cm and 4.8 cm respectively. A gardener wishes to test, at the 2% significance level, whether Hiergro fertilizer will increase the mean height. He treats 150 randomly chosen plants with Hiergro and finds that their mean height is 25.0 cm. Assuming that the standard deviation of the heights of plants treated with Hiergro is still 4.8 cm, carry out the test. [5]

Answer: There is evidence that Hiergo has increased the heights. **N12/71/Q2**

---

20 Following a change in flight schedules, an airline pilot wished to test whether the mean distance that he flies in a week has changed. He noted the distances, x km, that he flew in 50 randomly chosen weeks and summarised the results as follows.

$$n = 50 \qquad \Sigma x = 143\,300 \qquad \Sigma x^2 = 410\,900\,000$$

(i) Calculate unbiased estimates of the population mean and variance. [3]

(ii) In the past, the mean distance that he flew in a week was 2850 km. Test, at the 5% significance level, whether the mean distance has changed. [5]

Answer: 2866 4126.53 **N13/71/Q3**
No evidence that the mean distance changed.

---

21 A traffic officer notes the speeds of vehicles as they pass a certain point. In the past the mean of these speeds has been 62.3 km h^{-1} and the standard deviation has been 10.4 km h^{-1}. A speed limit is introduced, and following this, the mean of the speeds of 75 randomly chosen vehicles passing the point is found to be 59.9 km h^{-1}.

(i) Making an assumption that should be stated, test at the 2% significance level whether the mean speed has decreased since the introduction of the speed limit. [6]

(ii) Explain whether it was necessary to use the Central Limit theorem in part **(i)**. [2]

Answers: (i) Assume that the standard deviation is unchanged. No evidence that the mean speed has decreased. (Ii) yes. Population distribution was unknown. **N13/73/Q2**

22 The number of hours that Mrs Hughes spends on her business in a week is normally distributed with mean μ and standard deviation 4.8. In the past the value of μ has been 49.5.

(i) Assuming that μ is still equal to 49.5, find the probability that in a random sample of 40 weeks the mean time spent on her business in a week is more than 50.3 hours. [4]

Following a change in her arrangements, Mrs Hughes wishes to test whether μ has decreased. She chooses a random sample of 40 weeks and notes that the total number of hours she spent on her business during these weeks is 1920.

(ii) (a) Explain why a one-tail test is appropriate. [1]

(b) Carry out the test at the 6% significance level. [4]

(c) Explain whether it was necessary to use the Central Limit theorem in part **(ii)(b)**. [1]

Answer: 0.146; looking for a decrease; there is evidence that the mean time has decreased; no, the population is normally distributed **N14/71/Q5**

---

23 A researcher wishes to investigate whether the mean height of a certain type of plant in one region is different from the mean height of this type of plant everywhere else. He takes a large random sample of plants from the region and finds the sample mean. He calculates the value of the test statistic, z, and finds that $z = 1.91$.

(i) Explain briefly why the researcher should use a two-tail test. [1]

(ii) Carry out the test at the 4% significance level. [3]

Answer: There was no evidence that the mean height was different. **N14/73/Q1**

---

24 Parcels arriving at a certain office have weights W kg, where the random variable W has mean μ and standard deviation 0.2. The value of μ used to be 2.60, but there is a suspicion that this may no longer be true. In order to test at the 5% significance level whether the value of μ has increased, a random sample of 75 parcels is chosen. You may assume that the standard deviation of W is unchanged.

(i) The mean weight of the 75 parcels is found to be 2.64 kg. Carry out the test. [4]

(ii) Later another test of the same hypotheses at the 5% significance level, with another random sample of 75 parcels, is carried out. Given that the value of μ is now 2.68, calculate the probability of a Type II error. [5]

Answers: (i) there was evidence that μ had increased, (ii) 0.0345 or 0.0344 **N15/73/Q6**

---

25 In the past, the time spent by customers in a certain shop had mean 12.5 minutes and standard deviation 4.2 minutes. Following a change of layout in the shop, the mean time spent in the shop by a random sample of 50 customers is found to be 13.5 minutes.

(i) Assuming that the standard deviation remains at 4.2 minutes, test at the 5% significance level whether the mean time spent by customers in the shop has changed. [5]

(ii) Another random sample of 50 customers is chosen and a similar test at the 5% significance level is carried out. State the probability of a Type I error. [1]

26 In the past, the mean annual crop yield from a particular field has been 8.2 tonnes. During the last 16 years, a new fertiliser has been used on the field. The mean yield for these 16 years is 8.7 tonnes. Assume that yields are normally distributed with standard deviation 1.2 tonnes. Carry out a test at the 5% significance level of whether the mean yield has increased. [5]

Answer: Reject H_0, there is evidence that the mean yield has increased J16/73/Q2

27 Past experience has shown that the heights of a certain variety of plant have mean 64.0 cm and standard deviation 3.8 cm. During a particularly hot summer, it was expected that the heights of plants of this variety would be less than usual. In order to test whether this was the case, a botanist recorded the heights of a random sample of 100 plants and found that the value of the sample mean was 63.3 cm. Stating a necessary assumption, carry out the test at the 2.5% significance level. [6]

Answers: Assume standard deviation unchanged, No evidence that heights are shorter J17/71/Q2

28 The mass, in kilograms, of rocks in a certain area has mean 14.2 and standard deviation 3.1.

(i) Find the probability that the mean mass of a random sample of 50 of these rocks is less than 14.0 kg. [3]

(ii) Explain whether it was necessary to assume that the population of the masses of these rocks is normally distributed. [1]

(iii) A geologist suspects that rocks in another area have a mean mass which is less than 14.2 kg. A random sample of 100 rocks in this area has sample mean 13.5 kg. Assuming that the standard deviation for rocks in this area is also 3.1 kg, test at the 2% significance level whether the geologist is correct. [5]

Answers: (i) 0.324 (ii) No because n is large J18/71/Q5
(iii) There is evidence that the mean mass in this area is less than 14,2

29 The time taken for a particular train journey is normally distributed. In the past, the time had mean 2.4 hours and standard deviation 0.3 hours. A new timetable is introduced and on 30 randomly chosen occasions the time for this journey is measured. The mean time for these 30 occasions is found to be 2.3 hours.

(i) Stating any assumption(s), test, at the 5% significance level, whether the mean time for this journey has changed. [6]

(ii) A similar test at the 5% significance level was carried out using the times from another randomly chosen 30 occasions.

(a) State the probability of a Type I error. [1]

(b) State what is meant by a Type II error in this context. [1]

Answers: (i) Standard deviation unchanged. No evidence of a change in mean (ii) 5%. J18/73/Q5
(iii) To conclude that the mean was unchanged when in reality there is a change

30 In the past the time, in minutes, taken for a particular rail journey has been found to have mean 20.5 and standard deviation 1.2. Some new railway signals are installed. In order to test whether the mean time has decreased, a random sample of 100 times for this journey are noted. The sample mean is found to be 20.3 minutes. You should assume that the standard deviation is unchanged.

(i) Carry out a significance test, at the 4% level, of whether the population mean time has decreased. [5]

Later another significance test of the same hypotheses, using another random sample of size 100, is carried out at the 4% level.

(ii) Given that the population mean is now 20.1, find the probability of a Type II error. [5]

(iii) State what is meant by a Type II error in this context. [1]

Answer: No evidence that the population mean time has decreased **N16/71/Q7**
0.0567
Concluding that the mean time has not decreased when in fact it has

---

31 The manufacturer of a tablet computer claims that the mean battery life is 11 hours. A consumer organisation wished to test whether the mean is actually greater than 11 hours. They invited a random sample of members to report the battery life of their tablets. They then calculated the sample mean. Unfortunately a fire destroyed the records of this test except for the following partial document.

Test of the mean batte the table	
Sample size, n	
Sample mean (hours)	11.8
Is the result significant at the 5% level?	Yes
Is the result significant at the 2.5% level?	No

Given that the population of battery lives is normally distributed with standard deviation 1.6 hours, find the set of possible values of the sample size, n. [5]

Answer: the set of possible values of n was 11, 12, 13, 14, 15 **N16/73/Q4**

32 In order to test the effect of a drug, a researcher monitors the concentration, X, of a certain protein in the blood stream of patients. For patients who are not taking the drug the mean value of X is 0.185. A random sample of 150 patients taking the drug was selected and the values of X were found. The results are summarised below.

$$n = 150 \qquad \Sigma x = 27.0 \qquad \Sigma x^2 = 5.01$$

The researcher wishes to test at the 1% significance level whether the mean concentration of the protein in the blood stream of patients taking the drug is less than 0.185.

(i) Carry out the test. [7]

(ii) Given that, in fact, the mean concentration for patients taking the drug is 0.175, find the probability of a Type II error occurring in the test. [5]

Answers: (i) no evidence that the mean concentration with the drug is less than without the drug **(ii)** 0.0625 also answers in the range 0.0610 to 0.0628 accepted **N17/71/Q8**

33 A mill owner claims that the mean mass of sacks of flour produced at his mill is 51 kg. A quality control officer suspects that the mean mass is actually less than 51 kg. In order to test the owner's claim she finds the mass, x kg, of each of a random sample of 150 sacks and her results are summarised as follows.

$$n = 150 \qquad \Sigma x = 7480 \qquad \Sigma x^2 = 380\,000$$

(i) Carry out the test at the 2.5% significance level. [7]

You may now assume that the population standard deviation of the masses of sacks of flour is 6.856 kg. The quality control officer weighs another random sample of 150 sacks and carries out another test at the 2.5% significance level.

(ii) Given that the population mean mass is 49 kg, find the probability of a Type II error. [5]

Answers: (i) there was evidence that the population mean mass of sacks was less than 51 kg (ii) 0.0534 **N18/71/Q7**

Chapter 6: Hypothesis Testing Using a Binomial Distribution

HYPOTHESIS TEST FOR A BINOMIAL PROPORTION P

You may have seen advertisements for dairy spreads which claim that the spread cannot be distinguished from butter. How could you set about testing this claim? One way would be to take pairs of biscuits and put butter on one biscuit in each pair and the dairy spread on the other. The pairs of biscuits would be given to a number of tasters who would be asked to identify the biscuit with butter on it. Half the tasters would be given the buttered biscuit first, and the other half the buttered biscuit second.

Suppose you decided to use 10 tasters. How would you set about drawing a conclusion from your results? The method of hypothesis testing described in the last chapter can be adapted to this situation. First it is necessary to formulate a null hypothesis and an alternative hypothesis. It is usual to start from a position of doubt: you assume that the tasters cannot identify the butter and that they are guessing. In this situation the probability that a taster chosen at random will get the correct result is $\frac{1}{2}$. This can be expressed by the null hypothesis $H_0: p = \frac{1}{2}$. If some of the tasters can actually identify the butter then $p > \frac{1}{2}$. This can be expressed as an alternative hypothesis, $H_1: p > \frac{1}{2}$.

If H_0 is true, the number, X, of tasters who identify the buttered biscuit correctly is a random variable with distribution $B(10, \frac{1}{2})$. Fig. 7.1 shows this distribution. High values of X would suggest that H_0 should be rejected in favour of H_1. The most straightforward method of carrying out a hypothesis test for a discrete variable is to use the approach of Section 6.6 and calculate the probability that X takes the observed or a more extreme value assuming that H_0 is true and compare this probability with the specified significance level. Suppose that 9 out of the 10 people had identified the butter and you chose a significance level of 5%.

$$\begin{aligned} P(X \geqslant 9) &= P(X = 9) + P(X = 10) \\ &= \binom{10}{9}\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right) + \binom{10}{10}\left(\frac{1}{2}\right)^{10} \\ &= 0.009\,76\ldots + 0.000\,976\ldots \\ &= 0.010\,74\ldots \\ &= 1.07\%, \text{ correct to 3 significant figures.} \end{aligned}$$

This probability is less than 5% so the result is significant at the 5% level. H_0 is rejected and there is evidence, at the 5% significance level, that the proportion of people who can distinguish the butter is greater than $\frac{1}{2}$.

> To carry out a hypothesis test on a discrete variable, calculate the probability of the observed or a more extreme value and compare this probability with the significance level. For a one-tail test, reject the null hypothesis if this probability is less than the significance level; for a two-tail test reject the null hypothesis if this probability is less than half of the significance level.

Example 1

A national opinion poll claims that 40% of the electorate would vote for party R if there were an election tomorrow. A student at a large college suspects that the proportion of young people who would vote for them is lower. She asks 16 fellow students, chosen at random from the college roll, which party they would vote for. Three choose party R. Show, at the 10% significance level, that this indicates that the reported figure is too high for the young people at the student's college.

Example 2

A drugs company produced a new pain-relieving drug for migraine sufferers and its advertisements stated that the drug had a 90% success rate. A doctor doubted whether the drug would be as successful as the company claimed. She prescribed the drug for 15 of her patients. After six months, 11 of these patients said that their migraine symptoms had been relieved by the drug.

(a) Test the drug company's claim, at the 5% level of significance.
(b) Should the doctor continue to prescribe the drug?

TESTING A POPULATION PROPORTION FOR LARGE SAMPLES

When the sample is large, you can calculate probabilities by using the fact that the binomial distribution can be approximated by the normal distribution. The following examples illustrate the method.

Example 3

In a multiple choice paper a candidate has to select one of four possible answers to each question. On a paper with 100 questions a student gets 34 correct answers. Test, at the 5% significance level, the null hypothesis that the student is guessing the answers.

Example 4

If births are equally likely on any day of the week then the proportion of babies born at the week-end should be $\frac{2}{7}$. Out of a random sample of 490 children it was found that 132 were born at the week-end. Does this provide evidence, at the 5% significance level, that the proportion of babies born at the week-end differs from $\frac{2}{7}$?

TYPE I AND TYPE II ERRORS FOR TESTS INVOLVING THE BINOMIAL DISTRIBUTION

In Section 7.1 you met the idea that, for a discrete distribution, it is not usually possible to find a rejection region which corresponds exactly to the specified (nominal) significance level. You may find it helpful to look back at Section 7.1 before going on to the following example.

> For a hypothesis test involving a discrete variable the rejection region is defined so that
>
> $$\text{P}(\text{test statistic falls in rejection region} \mid H_0 \text{ true}) \leqslant \text{nominal significance of the test.}$$
>
> Actual significance level of the test $= \text{P}(\text{test statistic falls in rejection region} \mid H_0 \text{ true})$ and this is also the probability of a Type I error.

Example 5

An investigator suspects that operatives using a spring balance are reluctant to give 0 as the last value of a recorded weight, for example, 4.10 or 0.30. In order to test her theory she takes a random sample of 40 recorded weights and counts the number, X, which end in 0.

(a) State suitable hypotheses, involving a probability, for a hypothesis test which could indicate whether the operatives avoid ending a recorded weight with 0.

(b) Show that, for a test at the 10% significance level, the null hypothesis will be rejected if $X = 1$ but not if $X = 2$.

(c) State the rejection region for the test in terms of X.

(d) Calculate the value of $\text{P}(\text{Type I error})$.

(e) The nominal significance level of this test is 10%. What is the actual significance level of the test?

Example 6

A supplier of orchid seeds claims that their germination rate is 0.95. A purchaser of the seeds suspects that the germination rate is lower than this. In order to test this claim the purchaser plants 20 seeds in similar conditions, counts the number, X, which germinate. He rejects the claim if $X \leqslant 17$.

(a) Formulate suitable null and alternative hypotheses to test the seed supplier's claim.

(b) What is the probability of a Type I error using this test?

(c) Calculate $\text{P}(\text{Type II error})$ if the probability that a seed germinates is in fact 0.80.

Example 7

A manufacturer claims that the probability that an electric fuse is faulty is no more than 0.03. A purchaser tests this claim by testing a box of 500 fuses. A significance test is carried out at the 5% level using X, the number of faulty fuses in a box of 500, as the test statistic.

(a) For what values of X would you conclude that the probability that a fuse is faulty is greater than 0.03?
(b) Estimate P(Type I error) for this test.
(c) For this test estimate P(Type II error) if the probability that a fuse is faulty is, in fact, 0.06.

Example 8

Isaac claims that 30% of cars in his town are red. His friend Hardip thinks that the proportion is less than 30%. The boys decided to test Isaac's claim at the 5% significance level and found that 2 cars out of the random sample of 18 were red. Carry out the hypothesis test and state your conclusion.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q1 November 2007]

Example 9

At the 2009 election, $\frac{1}{3}$ of the voters in Chington voted for the Citizens Party. One year later, a researcher questioned 20 randomly selected voters in Chington. Exactly 3 of these 20 voters said that if there were an election next week they would vote for the Citizens Party. Test at the 2.5% significance level whether there is evidence of a decrease in support for the Citizens Party in Chington, since the 2009 election.

[Cambridge International AS and A Level Mathematics 9709, Paper 73 Q1 June 2010]

Example 10

At a certain airport 20% of people take longer than an hour to check in. A new computer system is installed, and it is claimed that this will reduce the time to check in. It is decided to accept the claim if, from a random sample of 22 people, the number taking longer than an hour to check in is either 0 or 1.

(i) Calculate the significance level of the test.
(ii) State the probability that a Type I error occurs.
(iii) Calculate the probability that a Type II error occurs if the probability that a person takes longer than an hour to check in is now 0.09.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q4 June 2007]

Example 11

A manufacturer claims that 20% of sugar-coated chocolate beans are red. George suspects that this percentage is actually less than 20% and so he takes a random sample of 15 chocolate beans and performs a hypothesis test with the null hypothesis $p = 0.2$ against the alternative hypothesis $p < 0.2$. He decides to reject the null hypothesis in favour of the alternative hypothesis if there are 0 or 1 red beans in the sample.

(i) With reference to this situation, explain what is meant by a Type I error.

(ii) Find the probability of a Type I error in George's test.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q2 November 2005]

Example 12

In a certain city it is necessary to pass a driving test in order to be allowed to drive a car. The probability of passing the driving test at the first attempt is 0.36 on average. A particular driving instructor claims that the probability of his pupils passing at the first attempt is higher than 0.36. A random sample of 8 of his pupils showed that 7 passed at the first attempt.

(i) Carry out an appropriate hypothesis test to test the driving instructor's claim, using a significance level of 5%.

(ii) In fact, most of this random sample happened to be careful and sensible drivers. State which type of error in the hypothesis test (Type I or Type II) could have been made in these circumstances and find the probability of this type of error when a sample of size 8 is used for the test.

[Cambridge International AS and A Level Mathematics 9709, Paper 7 Q4 June 2009]

Example 13

It is claimed that a certain 6-sided die is biased so that it is more likely to show a six than if it was fair. In order to test this claim at the 10% significance level, the die is thrown 10 times and the number of sixes is noted.

(i) Given that the die shows a six on 3 of the 10 throws, carry out the test.

On another occasion the same test is carried out again.

(ii) Find the probability of a Type I error.

(iii) Explain what is meant by a Type II error in this context.

[Cambridge International AS and A Level Mathematics 9709, Paper 71 Q6 November 2010]

TESTING THE POPULATION MEAN FOR A POISSON DISTRIBUTION

Example 14

In the past an office photocopier has failed, on average, three times every two weeks. A new, more expensive, photocopier is on trial which the manufacturers claim is more reliable. In the first four weeks of use this new photocopier fails once. Assuming that the failures of the photocopier occur independently and at random, test, at the 5% significance level, whether there is evidence that the new photocopier is more reliable than the old one.

Example 15

The number of failures in the rest of the first year is 57. Using the results for the whole year, test, at the 5% significance level, whether the new photocopier is more reliable than the old one.

TYPE I AND TYPE II ERRORS FOR TESTS INVOLVING THE POISSON DISTRIBUTION

Example 16

In an intensive survey of dune land it was found that the average number of plants of a particular species was 10 per m^2. After a very dry season it is suspected that these plants have tended to die off. In order to test this hypothesis a randomly chosen area of $1\ m^2$ is selected and the number of these plants, X, growing in it is counted. If X is greater than 4, it is assumed that the weather has no effect. You may assume that the distribution of X can be modelled by a Poisson distribution.

(a) State suitable null and alternative hypotheses. (b) What is P(Type I error)?

(c) Calculate P(Type II error) if the mean number of plants per square metre has changed to (i) 8 (ii) 4.

Do you think that the assumption of a Poisson model is likely to be justified in this situation?

Example 17

The average number of flaws per 100 metre length of yarn produced by a machine has been found to be 7. After the machine has been serviced, the number, X, of flaws in the first 300 metres of yarn produced by the machine is 27.

(a) Carry out a two-tail hypothesis test at the 5% level to test whether the average number of flaws produced by the machine has changed.

(b) For what values of X would the null hypothesis be rejected?

(c) Estimate the actual significance level of the test.

(d) Estimate P(Type II error) if the average has changed to 10.

Homework: Hypothesis Testing Using Poisson and Binomial– Variant 62

1 Before attending a basketball course, a player found that 60% of his shots made a score. After attending the course the player claimed he had improved. In his next game he tried 12 shots and scored in 10 of them. Assuming shots to be independent, test this claim at the 10% significance level. [5]

Answer: Accept claim at 10% level. **J03/Q2**

2 At a certain airport 20% of people take longer than an hour to check in. A new computer system is installed, and it is claimed that this will reduce the time to check in. It is decided to accept the claim if, from a random sample of 22 people, the number taking longer than an hour to check in is either 0 or 1.

(i) Calculate the significance level of the test. [3]

(ii) State the probability that a Type I error occurs. [1]

(iii) Calculate the probability that a Type II error occurs if the probability that a person takes longer than an hour to check in is now 0.09. [3]

Answers: (i) 0.0480; (ii) 0.0480; (iii) 0.601 . **J07/Q4**

3 When a guitar is played regularly, a string breaks on average once every 15 months. Broken strings occur at random times and independently of each other.

(i) Show that the mean number of broken strings in a 5-year period is 4. [1]

A guitar is fitted with a new type of string which, it is claimed, breaks less frequently. The number of broken strings of the new type was noted after a period of 5 years.

(ii) The mean number of broken strings of the new type in a 5-year period is denoted by λ. Find the rejection region for a test at the 10% significance level when the null hypothesis $\lambda = 4$ is tested against the alternative hypothesis $\lambda < 4$. [4]

(iii) Hence calculate the probability of making a Type I error. [1]

The number of broken guitar strings of the new type, in a 5-year period, was in fact 1.

(iv) State, with a reason, whether there is evidence at the 10% significance level that guitar strings of the new type break less frequently. [2]

Answers: (ii) 0 or 1; (iii) 0.0916; (iv) 1 is in the rejection region, there is evidence that the new guitar string lasts longer. **J08/Q5**

4 In a certain city it is necessary to pass a driving test in order to be allowed to drive a car. The probability of passing the driving test at the first attempt is 0.36 on average. A particular driving instructor claims that the probability of his pupils passing at the first attempt is higher than 0.36. A random sample of 8 of his pupils showed that 7 passed at the first attempt.

(i) Carry out an appropriate hypothesis test to test the driving instructor's claim, using a significance level of 5%. [5]

(ii) In fact, most of this random sample happened to be careful and sensible drivers. State which type of error in the hypothesis test (Type I or Type II) could have been made in these circumstances and find the probability of this type of error when a sample of size 8 is used for the test. [4]

Answers: **(i)** Accept the driving instructor's claim **J09/71/Q4**
(ii) 0.0293

5 In a certain city it is necessary to pass a driving test in order to be allowed to drive a car. The probability of passing the driving test at the first attempt is 0.36 on average. A particular driving instructor claims that the probability of his pupils passing at the first attempt is higher than 0.36. A random sample of 8 of his pupils showed that 7 passed at the first attempt.

(i) Carry out an appropriate hypothesis test to test the driving instructor's claim, using a significance level of 5%. [5]

(ii) In fact, most of this random sample happened to be careful and sensible drivers. State which type of error in the hypothesis test (Type I or Type II) could have been made in these circumstances and find the probability of this type of error when a sample of size 8 is used for the test. [4]

Answers: **(i)** Accept the driving instructor's claim **J09/72/Q4**
(ii) 0.0293

6 A hospital patient's white blood cell count has a Poisson distribution. Before undergoing treatment the patient had a mean white blood cell count of 5.2. After the treatment a random measurement of the patient's white blood cell count is made, and is used to test at the 10% significance level whether the mean white blood cell count has decreased.

(i) State what is meant by a Type I error in the context of the question, and find the probability that the test results in a Type I error. [4]

(ii) Given that the measured value of the white blood cell count after the treatment is 2, carry out the test. [3]

(iii) Find the probability of a Type II error if the mean white blood cell count after the treatment is actually 4.1. [3]

Answers: **(i)** The number of white blood cells has decreased when it has not, 0.0342; **(iii)** 0.915. **J10/72/Q7**

7 The number of accidents per month at a certain road junction has a Poisson distribution with mean 4.8. A new road sign is introduced warning drivers of the danger ahead, and in a subsequent month 2 accidents occurred.

(i) A hypothesis test at the 10% level is used to determine whether there were fewer accidents after the new road sign was introduced. Find the critical region for this test and carry out the test. [5]

(ii) Find the probability of a Type I error. [2]

Answers: **(i)** $X = 0$ or 1, Not enough evidence to say road sign has decreased accidents; **(ii)** 0.0477. **N02/Q4**

---

8 **(i)** Explain what is meant by

(a) a Type I error, [1]

(b) a Type II error. [1]

(ii) Roger thinks that a box contains 6 screws and 94 nails. Felix thinks that the box contains 30 screws and 70 nails. In order to test these assumptions they decide to take 5 items at random from the box and inspect them, replacing each item after it has been inspected, and accept Roger's hypothesis (the null hypothesis) if all 5 items are nails.

(a) Calculate the probability of a Type I error. [4]

(b) If Felix's hypothesis (the alternative hypothesis) is true, calculate the probability of a Type II error. [3]

Answers: **(i)(a)** Rejecting H_0 when it is true, **(b)** Accepting H_0 when it is false; **(ii)(a)** 0.266, **(b)** 0.168. **N03/Q6**

---

9 In a research laboratory where plants are studied, the probability of a certain type of plant surviving was 0.35. The laboratory manager changed the growing conditions and wished to test whether the probability of a plant surviving had increased.

(i) The plants were grown in rows, and when the manager requested a random sample of 8 plants to be taken, the technician took all 8 plants from the front row. Explain what was wrong with the technician's sample. [1]

(ii) A suitable sample of 8 plants was taken and 4 of these 8 plants survived. State whether the manager's test is one-tailed or two-tailed and also state the null and alternative hypotheses. Using a 5% significance level, find the critical region and carry out the test. [7]

(iii) State the meaning of a Type II error in the context of the test in part **(ii)**. [1]

(iv) Find the probability of a Type II error for the test in part **(ii)** if the probability of a plant surviving is now 0.4. [2]

Answers: **(i)** Not random, could be more light, etc.; **(ii)** One-tailed test, H_0: $p = 0.35$, H_1: $p > 0.35$, Critical region is 6,7,8 survive. No significant improvement in survival rate; **(iii)** Saying no improvement when there is; **(iv)** 0.950. **N04/Q7**

10 A manufacturer claims that 20% of sugar-coated chocolate beans are red. George suspects that this percentage is actually less than 20% and so he takes a random sample of 15 chocolate beans and performs a hypothesis test with the null hypothesis $p = 0.2$ against the alternative hypothesis $p < 0.2$. He decides to reject the null hypothesis in favour of the alternative hypothesis if there are 0 or 1 red beans in the sample.

(i) With reference to this situation, explain what is meant by a Type I error. [1]

(ii) Find the probability of a Type I error in George's test. [3]

Answers: (i) George says there are fewer than 20% red chocolate beans when there are 20%; (ii) 0.167. **N05/Q2**

11 Pieces of metal discovered by people using metal detectors are found randomly in fields in a certain area at an average rate of 0.8 pieces per hectare. People using metal detectors in this area have a theory that ploughing the fields increases the average number of pieces of metal found per hectare. After ploughing, they tested this theory and found that a randomly chosen field of area 3 hectares yielded 5 pieces of metal.

(i) Carry out the test at the 10% level of significance. [6]

(ii) What would your conclusion have been if you had tested at the 5% level of significance? [1]

Jack decides that he will reject the null hypothesis that the average number is 0.8 pieces per hectare if he finds 4 or more pieces of metal in another ploughed field of area 3 hectares.

(iii) If the true mean after ploughing is 1.4 pieces per hectare, calculate the probability that Jack makes a Type II error. [3]

Answers: (i) Ploughing has increased the number of metal pieces found; (ii) No significant increase at the 5% level; (iii) 0.395. **N06/Q6**

12 Isaac claims that 30% of cars in his town are red. His friend Hardip thinks that the proportion is less than 30%. The boys decided to test Isaac's claim at the 5% significance level and found that 2 cars out of a random sample of 18 were red. Carry out the hypothesis test and state your conclusion. [5]

Answer: Accept Isaac's claim. **N07/Q1**

13 Every month Susan enters a particular lottery. The lottery company states that the probability, p, of winning a prize is 0.0017 each month. Susan thinks that the probability of winning is higher than this, and carries out a test based on her 12 lottery results in a one-year period. She accepts the null hypothesis $p = 0.0017$ if she has no wins in the year and accepts the alternative hypothesis $p > 0.0017$ if she wins a prize in at least one of the 12 months.

(i) Find the probability of the test resulting in a Type I error. [2]

(ii) If in fact the probability of winning a prize each month is 0.0024, find the probability of the test resulting in a Type II error. [3]

(iii) Use a suitable approximation, with $p = 0.0024$, to find the probability that in a period of 10 years Susan wins a prize exactly twice. [3]

Answers: (i) 0.0202; (ii) 0.972; (iii) 0.0311. **N08/Q5**

14 The number of severe floods per year in a certain country over the last 100 years has followed a Poisson distribution with mean 1.8. Scientists suspect that global warming has now increased the mean. A hypothesis test, at the 5% significance level, is to be carried out to test this suspicion. The number of severe floods, X, that occur next year will be used for the test.

(i) Show that the rejection region for the test is $X > 4$. [5]

(ii) Find the probability of making a Type II error if the mean number of severe floods is now actually 2.3. [3]

Answer: **(ii)** 0.916. **N09/71/Q4**

15 It is not known whether a certain coin is fair or biased. In order to perform a hypothesis test, Raj tosses the coin 10 times and counts the number of heads obtained. The probability of obtaining a head on any throw is denoted by p.

(i) The null hypothesis is $p = 0.5$. Find the acceptance region for the test, given that the probability of a Type I error is to be at most 0.1. [4]

(ii) Calculate the probability of a Type II error in this test if the actual value of p is 0.7. [3]

Answers: **(i)** (2, 3, 4, 5, 6, 7, 8); **(ii)** 0.851. **N09/72/Q4**

16 It is claimed that a certain 6-sided die is biased so that it is more likely to show a six than if it was fair. In order to test this claim at the 10% significance level, the die is thrown 10 times and the number of sixes is noted.

(i) Given that the die shows a six on 3 of the 10 throws, carry out the test. [5]

On another occasion the same test is carried out again.

(ii) Find the probability of a Type I error. [3]

(iii) Explain what is meant by a Type II error in this context. [1]

Answers: **(ii)** 0.0697; **(iii)** Concluding that the die is fair when it is biased. **N10/72/Q6**

17 The number of adult customers arriving in a shop during a 5-minute period is modelled by a random variable with distribution Po(6). The number of child customers arriving in the same shop during a 10-minute period is modelled by an independent random variable with distribution Po(4.5).

(i) Find the probability that during a randomly chosen 2-minute period, the total number of adult and child customers who arrive in the shop is less than 3. [3]

(ii) During a sale, the manager claims that more adult customers are arriving than usual. In a randomly selected 30-minute period during the sale, 49 adult customers arrive. Test the manager's claim at the 2.5% significance level. [6]

Answers: **(i)** 0.359; **(ii)** There is evidence to support the claim. **J11/72/Q5**

18 Jeevan thinks that a six-sided die is biased in favour of six. In order to test this, Jeevan throws the die 10 times. If the die shows a six on at least 4 throws out of 10, she will conclude that she is correct.

(i) State appropriate null and alternative hypotheses. [1]

(ii) Calculate the probability of a Type I error. [3]

(iii) Explain what is meant by a Type II error in this situation. [1]

(iv) If the die is actually biased so that the probability of throwing a six is $\frac{1}{2}$, calculate the probability of a Type II error. [3]

Answers: **(i)** $H_0 : P(6) = \frac{1}{6}$, $H_1 : P(6) > \frac{1}{6}$; **(ii)** 0.0697; **(iii)** Die is biased but die shows a six on less than 4 throws; **(iv)** 0.172. **J11/72/Q6**

---

19 The number of new enquiries per day at an office has a Poisson distribution. In the past the mean has been 3. Following a change of staff, the manager wishes to test, at the 5% significance level, whether the mean has increased.

(i) State the null and alternative hypotheses for this test. [1]

The manager notes the number, N, of new enquiries during a certain 6-day period. She finds that $N = 25$ and then, assuming that the null hypothesis is true, she calculates that $P(N \geqslant 25) = 0.0683$.

(ii) What conclusion should she draw? [2]

Answers: **(i)** H_0 : Population mean=3 H_1 : Population mean>3 **J12/72/Q1**
(ii) No evidence that pop. mean has increased.

---

20 The number of cases of asthma per month at a clinic has a Poisson distribution. In the past the mean has been 5.3 cases per month. A new treatment is introduced. In order to test at the 5% significance level whether the mean has decreased, the number of cases in a randomly chosen month is noted.

(i) Find the critical region for the test and, given that the number of cases is 2, carry out the test. [5]

(ii) Explain the meaning of a Type I error in this context and state the probability of a Type I error. [2]

(iii) At another clinic the mean number of cases of asthma per month has the independent distribution Po(13.1). Assuming that the mean for the first clinic is still 5.3, use a suitable approximating distribution to estimate the probability that the total number of cases in the two clinics in a particular month is more than 20. [5]

Answer: CR is 0 or 1 cases. No evidence that the mean has decreased **J13/72/Q6**
Concluding that the mean has decreased when, in fact, it has not. 0.0314
0.312

21 Cloth made at a certain factory has been found to have an average of 0.1 faults per square metre. Suki claims that the cloth made by her machine contains, on average, more than 0.1 faults per square metre. In a random sample of 5 m^2 of cloth from Suki's machine, it was found that there were 2 faults. Assuming that the number of faults per square metre has a Poisson distribution,

(i) state null and alternative hypotheses for a test of Suki's claim, [1]

(ii) test at the 10% significance level whether Suki's claim is justified. [4]

Answer: (i) H_0: λ=0.5 **J15/72/Q2**
H_1: λ>0.5
(ii) There is evidence to support the claim

---

22 A cereal manufacturer claims that 25% of cereal packets contain a free gift. Lola suspects that the true proportion is less than 25%. In order to test the manufacturer's claim at the 5% significance level, she checks a random sample of 20 packets.

(i) Find the critical region for the test. [5]

(ii) Hence find the probability of a Type I error. [1]

Lola finds that 2 packets in her sample contain a free gift.

(iii) State, with a reason, the conclusion she should draw. [2]

Answers: (i) Critical region 0 or 1 packets contain gift; (ii) 0.0243; (iii) No evidence to reject claim. **N12/72/Q4**

---

23 At the last election, 70% of people in Apoli supported the president. Luigi believes that the same proportion support the president now. Maria believes that the proportion who support the president now is 35%. In order to test who is right, they agree on a hypothesis test, taking Luigi's belief as the null hypothesis. They will ask 6 people from Apoli, chosen at random, and if more than 3 support the president they will accept Luigi's belief.

(i) Calculate the probability of a Type I error. [3]

(ii) If Maria's belief is true, calculate the probability of a Type II error. [3]

(iii) In fact 2 of the 6 people say that they support the president. State which error, Type I or Type II, might be made. Explain your answer. [2]

Answer: 0.256 **N13/72/Q6**
0.117
Type I; they will reject Luigi's belief, although it may be true.

24 The number of accidents on a certain road has a Poisson distribution with mean 3.1 per 12-week period.

(i) Find the probability that there will be exactly 4 accidents during an 18-week period. [3]

Following the building of a new junction on this road, an officer wishes to determine whether the number of accidents per week has decreased. He chooses 15 weeks at random and notes the number of accidents. If there are fewer than 3 accidents altogether he will conclude that the number of accidents per week has decreased. He assumes that a Poisson distribution still applies.

(ii) Find the probability of a Type I error. [3]

(iii) Given that the mean number of accidents per week is now 0.1, find the probability of a Type II error. [3]

(iv) Given that there were 2 accidents during the 15 weeks, explain why it is impossible for the officer to make a Type II error. [1]

Answer: 0.186, 0.257, 0.191, He will reject H_0 **N14/72/Q6**

---

25 At a certain hospital it was found that the probability that a patient did not arrive for an appointment was 0.2. The hospital carries out some publicity in the hope that this probability will be reduced. They wish to test whether the publicity has worked.

(i) It is suggested that the first 30 appointments on a Monday should be used for the test. Give a reason why this is not an appropriate sample. [1]

A suitable sample of 30 appointments is selected and the number of patients that do not arrive is noted. This figure is used to carry out a test at the 5% significance level.

(ii) Explain why the test is one-tail and state suitable null and alternative hypotheses. [2]

(iii) State what is meant by a Type I error in this context. [1]

(iv) Use the binomial distribution to find the critical region, and find the probability of a Type I error. [5]

(v) In fact 3 patients out of the 30 do not arrive. State the conclusion of the test, explaining your answer. [2]

Answer: Probability could be different later in the day, or on a different day **N15/72/Q7**
Looking for a decrease H_0 p=0.2 H_1 p <0.2
Concluding that the probability has decreased when it has not
CR $X \le 2$ 0.0442
No evidence that p has decreased

Homework: Hypothesis Testing Using Poisson and Binomial– Variant 71 & 73

1 At the 2009 election, $\frac{1}{3}$ of the voters in Chington voted for the Citizens Party. One year later, a researcher questioned 20 randomly selected voters in Chington. Exactly 3 of these 20 voters said that if there were an election next week they would vote for the Citizens Party. Test at the 2.5% significance level whether there is evidence of a decrease in support for the Citizens Party in Chington, since the 2009 election. [5]

Answer: No evidence of a reduction in support for the Citizens Party. **J10/73/Q1**

2 At a power plant, the number of breakdowns per year has a Poisson distribution. In the past the mean number of breakdowns per year has been 4.8. Following some repairs, the management carry out a hypothesis test at the 5% significance level to determine whether this mean has decreased. If there is at most 1 breakdown in the following year, they will conclude that the mean has decreased.

(i) State what is meant by a Type I error in this context. [1]

(ii) Find the probability of a Type I error. [2]

(iii) Find the probability of a Type II error if the mean is now 0.9 breakdowns per year. [3]

Answers: (i) Conclude there is a reduced number of breakdowns when no reduction has occurred; (ii) 0.0477; (iii) 0.228. **J10/73/Q4**

3 The number of injuries per month at a certain factory has a Poisson distribution. In the past the mean was 2.1 injuries per month. New safety procedures are put in place and the management wishes to use the next 3 months to test, at the 2% significance level, whether there are now fewer injuries than before, on average.

(i) Find the critical region for the test. [5]

(ii) Find the probability of a Type I error. [1]

(iii) During the next 3 months there are a total of 3 injuries. Carry out the test. [3]

(iv) Assuming that the mean remains 2.1, calculate an estimate of the probability that there will be fewer than 20 injuries during the next 12 months. [5]

Answers: (i) $X \le 1$; (ii) 0.0134; (iii) No evidence that mean number of injuries has decreased; (iv) 0.128. **J11/71/Q6**

4 At an election in 2010, 15% of voters in Bratfield voted for the Renewal Party. One year later, a researcher asked 30 randomly selected voters in Bratfield whether they would vote for the Renewal Party if there were an election next week. 2 of these 30 voters said that they would.

(i) Use a binomial distribution to test, at the 4% significance level, the null hypothesis that there has been no change in the support for the Renewal Party in Bratfield against the alternative hypothesis that there has been a decrease in support since the 2010 election. [4]

(ii) **(a)** Explain why the conclusion in part **(i)** cannot involve a Type I error. [1]

(b) State the circumstances in which the conclusion in part **(i)** would involve a Type II error. [1]

Answers: (i) Accept H_0; (ii)(a) H_0 has been accepted; (ii) (b) If $p < 0.15$. **J11/73/Q3**

---

5 **(i)** Deng wishes to test whether a certain coin is biased so that it is more likely to show Heads than Tails. He throws it 12 times. If it shows Heads more than 9 times, he will conclude that the coin is biased. Calculate the significance level of the test. [3]

(ii) Deng throws another coin 100 times in order to test, at the 5% significance level, whether it is biased towards Heads. Find the rejection region for this test. [5]

Answers: (i) 1.93% or 1.9% (2% allowed with correct working) (ii) Region is ≥59 **J12/73/Q5**

---

6 Leila suspects that a particular six-sided die is biased so that the probability, p, that it will show a six is greater than $\frac{1}{6}$. She tests the die by throwing it 5 times. If it shows a six on 3 or more throws she will conclude that it is biased.

(i) State what is meant by a Type I error in this situation and calculate the probability of a Type I error. [3]

(ii) Assuming that the value of p is actually $\frac{2}{3}$, calculate the probability of a Type II error. [3]

Leila now throws the die 80 times and it shows a six on 50 throws.

(iii) Calculate an approximate 96% confidence interval for p. [4]

Answer: 0.0355 **J13/71/Q7**
0.210
0.514 to 0.736

---

7 A hockey player found that she scored a goal on 82% of her penalty shots. After attending a coaching course, she scored a goal on 19 out of 20 penalty shots. Making an assumption that should be stated, test at the 10% significance level whether she has improved. [5]

Answer: Insufficient evidence to show that the player had improved. **J13/73/Q2**

---

8 Stephan is an athlete who competes in the high jump. In the past, Stephan has succeeded in 90% of jumps at a certain height. He suspects that his standard has recently fallen and he decides to carry out a hypothesis test to find out whether he is right. If he succeeds in fewer than 17 of his next 20 jumps at this height, he will conclude that his standard has fallen.

(i) Find the probability of a Type I error. [4]

(ii) In fact Stephan succeeds in 18 of his next 20 jumps. Which of the errors, Type I or Type II, is possible? Explain your answer. [2]

Answers: 0.0147 **J14/71/Q6**
2.13

9 Marie claims that she can predict the winning horse at the local races. There are 8 horses in each race. Nadine thinks that Marie is just guessing, so she proposes a test. She asks Marie to predict the winners of the next 10 races and, if she is correct in 3 or more races, Nadine will accept Marie's claim.

(i) State suitable null and alternative hypotheses. [1]

(ii) Calculate the probability of a Type I error. [3]

(iii) State the significance level of the test. [1]

Answers: **(i)** $H_0\ p=\frac{1}{8}$ $H_1\ p>\frac{1}{8}$ **(ii)** 0.119 or 0.120 **(iii)** 12% **J15/73/Q2**

10 It is claimed that a certain 6-sided die is biased so that it is more likely to show a six than if it was fair. In order to test this claim at the 10% significance level, the die is thrown 10 times and the number of sixes is noted.

(i) Given that the die shows a six on 3 of the 10 throws, carry out the test. [5]

On another occasion the same test is carried out again.

(ii) Find the probability of a Type I error. [3]

(iii) Explain what is meant by a Type II error in this context. [1]

Answers: **(ii)** 0.0697; **(iii)** Concluding that the die is fair when it is biased. **N10/71/Q6**

11 In the past, the number of house sales completed per week by a building company has been modelled by a random variable which has the distribution Po(0.8). Following a publicity campaign, the builders hope that the mean number of sales per week will increase. In order to test at the 5% significance level whether this is the case, the total number of sales during the first 3 weeks after the campaign is noted. It is assumed that a Poisson model is still appropriate.

(i) Given that the total number of sales during the 3 weeks is 5, carry out the test. [6]

(ii) During the following 3 weeks the same test is carried out again, using the same significance level. Find the probability of a Type I error. [3]

(iii) Explain what is meant by a Type I error in this context. [1]

(iv) State what further information would be required in order to find the probability of a Type II error. [1]

Answers: **(ii)** 0.0357; **(iii)** Concluding that the mean sales have increased when this is not true; **(iv)** The new value of the mean. **N10/73/Q7**

12 A cereal manufacturer claims that 25% of cereal packets contain a free gift. Lola suspects that the true proportion is less than 25%. In order to test the manufacturer's claim at the 5% significance level, she checks a random sample of 20 packets.

(i) Find the critical region for the test. [5]

(ii) Hence find the probability of a Type I error. [1]

Lola finds that 2 packets in her sample contain a free gift.

(iii) State, with a reason, the conclusion she should draw. [2]

Answers: **(i)** Critical region 0 or 1 packets contain gift; **(ii)** 0.0243; **(iii)** No evidence to reject claim. **N12/71/Q4**

13 Joshi suspects that a certain die is biased so that the probability of showing a six is less than $\frac{1}{6}$. He plans to throw the die 25 times and if it shows a six on fewer than 2 throws, he will conclude that the die is biased in this way.

(i) Find the probability of a Type I error and state the significance level of the test. [3]

Joshi now decides to throw the die 100 times. It shows a six on 9 of these throws.

(ii) Calculate an approximate 95% confidence interval for the probability of showing a six on one throw of this die. [4]

Answers: **(i)** 6.29%; **(ii)** (0.0339, 0.146) also acceptable here is (0.034, 0.146). **N12/73/Q3**

14 The number of workers, X, absent from a factory on a particular day has the distribution B(80, 0.01).

(i) Explain why it is appropriate to use a Poisson distribution as an approximating distribution for X. [2]

(ii) Use the Poisson distribution to find the probability that the number of workers absent during 12 randomly chosen days is more than 2 and less than 6. [3]

Following a change in working conditions, the management wishes to test whether the mean number of workers absent per day has decreased.

(iii) During 10 randomly chosen days, there were a total of 2 workers absent. Use the Poisson distribution to carry out the test at the 2% significance level. [5]

Answers: (i) $n > 50$, $np < 5$; (ii) 0.0800; **(iii)** There is evidence that the mean number of absent workers has decreased. **N12/73/Q7**

15 At the last election, 70% of people in Apoli supported the president. Luigi believes that the same proportion support the president now. Maria believes that the proportion who support the president now is 35%. In order to test who is right, they agree on a hypothesis test, taking Luigi's belief as the null hypothesis. They will ask 6 people from Apoli, chosen at random, and if more than 3 support the president they will accept Luigi's belief.

(i) Calculate the probability of a Type I error. [3]

(ii) If Maria's belief is true, calculate the probability of a Type II error. [3]

(iii) In fact 2 of the 6 people say that they support the president. State which error, Type I or Type II, might be made. Explain your answer. [2]

Answer: 0.256 **N13/71/Q6**
0.117
Type I; they will reject Luigi's belief, although it may be true.

---

16 A fair six-sided die has faces numbered 1, 2, 3, 4, 5, 6. The score on one throw is denoted by X.

(i) Write down the value of $\mathrm{E}(X)$ and show that $\mathrm{Var}(X) = \frac{35}{12}$. [2]

Fayez has a six-sided die with faces numbered 1, 2, 3, 4, 5, 6. He suspects that it is biased so that when it is thrown it is more likely to show a low number than a high number. In order to test his suspicion, he plans to throw the die 50 times. If the mean score is less than 3 he will conclude that the die is biased.

(ii) Find the probability of a Type I error. [5]

(iii) With reference to this context, describe circumstances in which Fayez would make a Type II error. [2]

Answers: (i) 3.5 or 35/12, (ii) 0.0192, (iii) die is biased. mean of 50 throws ≥ 3 **N13/73/Q5**

---

17 The number of accidents on a certain road has a Poisson distribution with mean 3.1 per 12-week period.

(i) Find the probability that there will be exactly 4 accidents during an 18-week period. [3]

Following the building of a new junction on this road, an officer wishes to determine whether the number of accidents per week has decreased. He chooses 15 weeks at random and notes the number of accidents. If there are fewer than 3 accidents altogether he will conclude that the number of accidents per week has decreased. He assumes that a Poisson distribution still applies.

(ii) Find the probability of a Type I error. [3]

(iii) Given that the mean number of accidents per week is now 0.1, find the probability of a Type II error. [3]

(iv) Given that there were 2 accidents during the 15 weeks, explain why it is impossible for the officer to make a Type II error. [1]

Answer: 0.186, 0.257, 0.191, He will reject H_0 **N14/71/Q6**

---

18 It is known that when seeds of a certain type are planted, on average 10% of the resulting plants reach a height of 1 metre. A gardener wishes to investigate whether a new fertiliser will increase this proportion. He plants a random sample of 18 seeds of this type, using the fertiliser, and notes how many of the resulting plants reach a height of 1 metre.

(i) In fact 4 of the 18 plants reach a height of 1 metre. Carry out a hypothesis test at the 8% significance level. [5]

(ii) Explain which of the errors, Type I or Type II, might have been made in part **(i)**. [2]

Later, the gardener plants another random sample of 18 seeds of this type, using the fertiliser, and again carries out a hypothesis test at the 8% significance level.

(iii) Find the probability of a Type I error. [3]

Answers: (i) no evidence that the proportion of plants reaching 1m has increased. (ii) H_0 was not rejected so a Type II error might have been made. (iii) 0.0282 **N14/73/Q5**

19 At a certain hospital it was found that the probability that a patient did not arrive for an appointment was 0.2. The hospital carries out some publicity in the hope that this probability will be reduced. They wish to test whether the publicity has worked.

(i) It is suggested that the first 30 appointments on a Monday should be used for the test. Give a reason why this is not an appropriate sample. [1]

A suitable sample of 30 appointments is selected and the number of patients that do not arrive is noted. This figure is used to carry out a test at the 5% significance level.

(ii) Explain why the test is one-tail and state suitable null and alternative hypotheses. [2]

(iii) State what is meant by a Type I error in this context. [1]

(iv) Use the binomial distribution to find the critical region, and find the probability of a Type I error. [5]

(v) In fact 3 patients out of the 30 do not arrive. State the conclusion of the test, explaining your answer. [2]

Answer: Probability could be different later in the day, or on a different day **N15/71/Q7**
Looking for a decrease H_0 p=0.2 H_1 p <0.2
Concluding that the probability has decreased when it has not
CR $X \leq 2$ 0.0442
No evidence that p has decreased

20 **(a)** Narika has a die which is known to be biased so that the probability of throwing a 6 on any throw is $\frac{1}{100}$. She uses an approximating distribution to calculate the probability of obtaining no 6s in 450 throws. Find the percentage error in using the approximating distribution for this calculation. [4]

(b) Johan claims that a certain six-sided die is biased so that it shows a 6 less often than it would if the die were fair. In order to test this claim, the die is thrown 25 times and it shows a 6 on only 2 throws. Test at the 10% significance level whether Johan's claim is justified. [5]

Answers: (a) 2.29% , (ii) no reason to believe that the die was biased. **N15/73/Q5**

21 A six-sided die shows a six on 25 throws out of 200 throws. Test at the 10% significance level the null hypothesis: P(throwing a six) $= \frac{1}{6}$, against the alternative hypothesis: P(throwing a six) $< \frac{1}{6}$. [5]

Answer: There is some evidence that $p < 1/6$ **J16/71/Q1**

22 The number of sightings of a golden eagle at a certain location has a Poisson distribution with mean 2.5 per week. Drilling for oil is started nearby. A naturalist wishes to test at the 5% significance level whether there are fewer sightings since the drilling began. He notes that during the following 3 weeks there are 2 sightings.

(i) Find the critical region for the test and carry out the test. [5]

(ii) State the probability of a Type I error. [1]

(iii) State why the naturalist could not have made a Type II error. [1]

Answers: **(i)** $X < 2$, reject H_0. Evidence of fewer sightings. **(ii)** 0.0203 **(iii)** As H_0 rejected **J16/73/Q4**

---

23 The number of sports injuries per month at a certain college has a Poisson distribution. In the past the mean has been 1.1 injuries per month. The principal recently introduced new safety guidelines and she decides to test, at the 2% significance level, whether the mean number of sports injuries has been reduced. She notes the number of sports injuries during a 6-month period.

(i) Find the critical region for the test and state the probability of a Type I error. [6]

(ii) State what is meant by a Type I error in this context. [1]

(iii) During the 6-month period there are a total of 2 sports injuries. Carry out the test. [3]

(iv) Assuming that the mean remains 1.1, calculate the probability that there will be fewer than 30 sports injuries during a 36-month period. [4]

Answers: **(i)** $X \leqslant 1$ = 0.0103 **(ii)** Wrongly conclude that the mean number of sports injuries has decreased **(iii)** No evidence that the mean number of sports injuries has decreased **(iv)** 0.0543 **J17/71/Q6**

---

24 In the past the number of accidents per month on a certain road was modelled by a random variable with distribution Po(0.47). After the introduction of speed restrictions, the government wished to test, at the 5% significance level, whether the mean number of accidents had decreased. They noted the number of accidents during the next 12 months. It is assumed that accidents occur randomly and that a Poisson model is still appropriate.

(i) Given that the total number of accidents during the 12 months was 2, carry out the test. [6]

(ii) Explain what is meant by a Type II error in this context. [1]

It is given that the mean number of accidents per month is now in fact 0.05.

(iii) Using another random sample of 12 months the same test is carried out again, with the same significance level. Find the probability of a Type II error. [4]

Answers: **(i)** Accept H_0 no evidence that the number of accidents had fallen **J17/73/Q7**
(ii) Conclude that the number of accidents was unchanged, when it had decreased **(iii)** 0.122

25 The number of absences by girls from a certain class on any day is modelled by a random variable with distribution Po(0.2). The number of absences by boys from the same class on any day is modelled by an independent random variable with distribution Po(0.3).

(i) Find the probability that, during a randomly chosen 2-day period, the total number of absences is less than 3. [3]

(ii) Find the probability that, during a randomly chosen 5-day period, the number of absences by boys is more than 3. [2]

(iii) The teacher claims that, during the football season, there are more absences by boys than usual. In order to test this claim at the 5% significance level, he notes the number of absences by boys during a randomly chosen 5-day period during the football season.

(a) State what is meant by a Type I error in this context. [1]

(b) State appropriate null and alternative hypotheses and find the probability of a Type I error. [3]

(c) In fact there were 4 absences by boys during this period. Test the teacher's claim at the 5% significance level. [3]

Answers: (i) 0.920 (ii) 0.0656 **J18/71/Q7**
(iii)(a) Incorrectly concluding that there were more absences than usual when there were not
(b) $H_0: \lambda = 1.5$ $H_1: \lambda > 1.5$ 0.0186 (c) No evidence of more than usual male absences

---

26 A die has six faces numbered 1, 2, 3, 4, 5, 6. Manjit suspects that the die is biased so that it shows a six on fewer throws than it would if it were fair. In order to test her suspicion, she throws the die a certain number of times and counts the number of sixes.

(i) State suitable null and alternative hypotheses for Manjit's test. [1]

(ii) There are no sixes in the first 15 throws. Show that this result is not significant at the 5% level. [2]

(iii) Find the smallest value of n such that, if there are no sixes in the first n throws, this result is significant at the 5% level. [2]

Answer: H_0: P(6) = 1/6 H_1: P(6) < 1/6 **N16/71/Q2**
0.065 > 0.05
17

---

27 It is claimed that 30% of packets of Froogum contain a free gift. Andre thinks that the actual proportion is less than 30% and he decides to carry out a hypothesis test at the 5% significance level. He buys 20 packets of Froogum and notes the number of free gifts he obtains.

(i) State null and alternative hypotheses for the test. [1]

(ii) Use a binomial distribution to find the probability of a Type I error. [5]

Andre finds that 3 of the 20 packets contain free gifts.

(iii) Carry out the test. [2]

Answer: $H_0 : p = 0.3$ $H_1 : p < 0.3$ **N16/73/Q5**

Answer: 0.0355

Answer: no evidence to reject the claim that 30% of packets contain a free gift

28 In the past the number of cars sold per day at a showroom has been modelled by a random variable with distribution Po(0.7). Following an advertising campaign, it is hoped that the mean number of sales per day will increase. In order to test at the 10% significance level whether this is the case, the total number of sales during the first 5 days after the campaign is noted. You should assume that a Poisson model is still appropriate.

(i) Given that the total number of cars sold during the 5 days is 5, carry out the test. [6]

9

The number of cars sold per day at another showroom has the independent distribution Po(0.6). Assume that the distribution for the first showroom is still Po(0.7).

(ii) Find the probability that the total number of cars sold in the two showrooms during 3 days is exactly 2. [3]

Answers: **(i)** no evidence to suggest that the sales per day have increased, **(ii)** 0.154 **N17/71/Q7**

29 A headteacher models the number of children who bring a 'healthy' packed lunch to school on any day by the distribution $B(150, p)$. In the past, she has found that $p = \frac{1}{3}$. Following the opening of a fast food outlet near the school, she wishes to test, at the 1% significance level, whether the value of p has decreased.

(i) State the null and alternative hypotheses for this test. [1]

On a randomly chosen day she notes the number, N, of children who bring a 'healthy' packed lunch to school. She finds that $N = 36$ and then, assuming that the null hypothesis is true, she calculates that $P(N \leqslant 36) = 0.0084$.

(ii) State, with a reason, the conclusion that the headteacher should draw from the test. [2]

(iii) According to the model, what is the largest number of children who might bring a packed lunch to school? [1]

Answers: (i) H_0: p □ $\frac{1}{3}$ H_1: p < $\frac{1}{3}$ (ii) 0.0084 < 0.01 There is evidence that p has decreased **N18/71/Q2**

(iii) 150

6 Probability & Statistics 2 (for Paper 6)

Knowledge of the content of Paper 5: Probability & Statistics 1 is assumed, and candidates may be required to demonstrate such knowledge in answering questions. Knowledge of calculus within the content for Paper 3: Pure Mathematics 3 will also be assumed.

6.1 The Poisson distribution

Candidates should be able to:	Notes and examples
• use formulae to calculate probabilities for the distribution $\text{Po}(\lambda)$	
• use the fact that if $X \sim \text{Po}(\lambda)$ then the mean and variance of X are each equal to λ	Proofs are not required.
• understand the relevance of the Poisson distribution to the distribution of random events, and use the Poisson distribution as a model	
• use the Poisson distribution as an approximation to the binomial distribution where appropriate	The conditions that n is large and p is small should be known; $n > 50$ and $np < 5$, approximately.
• use the normal distribution, with continuity correction, as an approximation to the Poisson distribution where appropriate	The condition that λ is large should be known; $\lambda > 15$, approximately.

6.2 Linear combinations of random variables

Candidates should be able to:	Notes and examples
• use, when solving problems, the results that – $\text{E}(aX + b) = a\text{E}(X) + b$ and $\text{Var}(aX + b) = a^2\text{Var}(X)$ – $\text{E}(aX + bY) = a\text{E}(X) + b\text{E}(Y)$ – $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$ for independent X and Y – if X has a normal distribution then so does $aX + b$ – if X and Y have independent normal distributions then $aX + bY$ has a normal distribution – if X and Y have independent Poisson distributions then $X + Y$ has a Poisson distribution.	Proofs of these results are not required.

6 Probability & Statistics 2

6.3 Continuous random variables

Candidates should be able to:	Notes and examples
• understand the concept of a continuous random variable, and recall and use properties of a probability density function	For density functions defined over a single interval only; the domain may be infinite, e.g. $\frac{3}{x^4}$ for $x \geqslant 1$.
• use a probability density function to solve problems involving probabilities, and to calculate the mean and variance of a distribution.	Including location of the median or other percentiles of a distribution by direct consideration of an area using the density function. Explicit knowledge of the cumulative distribution function is not included.

6.4 Sampling and estimation

Candidates should be able to:	Notes and examples
• understand the distinction between a sample and a population, and appreciate the necessity for randomness in choosing samples	
• explain in simple terms why a given sampling method may be unsatisfactory	Including an elementary understanding of the use of random numbers in producing random samples. Knowledge of particular sampling methods, such as quota or stratified sampling, is not required.
• recognise that a sample mean can be regarded as a random variable, and use the facts that $\mathrm{E}(\overline{X}) = \mu$ and that $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$	
• use the fact that $(\overline{X})$ has a normal distribution if X has a normal distribution	
• use the Central Limit Theorem where appropriate	Only an informal understanding of the Central Limit Theorem (CLT) is required; for large sample sizes, the distribution of a sample mean is approximately normal.
• calculate unbiased estimates of the population mean and variance from a sample, using either raw or summarised data	Only a simple understanding of the term ‘unbiased’ is required, e.g. that although individual estimates will vary the process gives an accurate result ‘on average’.
• determine and interpret a confidence interval for a population mean in cases where the population is normally distributed with known variance or where a large sample is used	
• determine, from a large sample, an approximate confidence interval for a population proportion.	

6 Probability & Statistics 2

6.5 Hypothesis tests

Candidates should be able to:

- understand the nature of a hypothesis test, the difference between one-tailed and two-tailed tests, and the terms null hypothesis, alternative hypothesis, significance level, rejection region (or critical region), acceptance region and test statistic
- formulate hypotheses and carry out a hypothesis test in the context of a single observation from a population which has a binomial or Poisson distribution, using
  - direct evaluation of probabilities
  - a normal approximation to the binomial or the Poisson distribution, where appropriate
- formulate hypotheses and carry out a hypothesis test concerning the population mean in cases where the population is normally distributed with known variance or where a large sample is used
- understand the terms Type I error and Type II error in relation to hypothesis tests
- calculate the probabilities of making Type I and Type II errors in specific situations involving tests based on a normal distribution or direct evaluation of binomial or Poisson probabilities.

Notes and examples

Outcomes of hypothesis tests are expected to be interpreted in terms of the contexts in which questions are set.

Command words

The table below includes command words used in the assessment for this syllabus. The use of the command word will relate to the subject context.

Command word	What it means
Calculate	work out from given facts, figures or information
Describe	state the points of a topic / give characteristics and main features
Determine	establish with certainty
Evaluate	judge or calculate the quality, importance, amount, or value of something
Explain	set out purposes or reasons / make the relationships between things evident / provide why and/or how and support with relevant evidence
Identify	name/select/recognise
Justify	support a case with evidence/argument
Show (that)	provide structured evidence that leads to a given result
Sketch	make a simple freehand drawing showing the key features
State	express in clear terms
Verify	confirm a given statement/result is true

5 List of formulae and statistical tables (MF19)

PURE MATHEMATICS

Mensuration

Volume of sphere = $\frac{4}{3}\pi r^3$

Surface area of sphere = $4\pi r^2$

Volume of cone or pyramid = $\frac{1}{3} \times$ base area $\times$ height

Area of curved surface of cone = $\pi r \times$ slant height

Arc length of circle $= r\theta$ (θ in radians)

Area of sector of circle $= \frac{1}{2}r^2\theta$ (θ in radians)

Algebra

For the quadratic equation $ax^2 + bx + c = 0$:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

For an arithmetic series:

$$u_n = a + (n-1)d, \qquad S_n = \tfrac{1}{2}n(a+l) = \tfrac{1}{2}n\{2a + (n-1)d\}$$

For a geometric series:

$$u_n = ar^{n-1}, \qquad S_n = \frac{a(1-r^n)}{1-r} \quad (r \neq 1), \qquad S_\infty = \frac{a}{1-r} \quad (|r| < 1)$$

Binomial series:

$$(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \binom{n}{3}a^{n-3}b^3 + \ldots + b^n, \text{ where } n \text{ is a positive integer}$$

$$\text{and } \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \ldots, \text{ where } n \text{ is rational and } |x| < 1$$

Trigonometry

$$\tan\theta \equiv \frac{\sin\theta}{\cos\theta}$$

$$\cos^2\theta + \sin^2\theta \equiv 1, \qquad 1 + \tan^2\theta \equiv \sec^2\theta, \qquad \cot^2\theta + 1 \equiv \operatorname{cosec}^2\theta$$

$$\sin(A \pm B) \equiv \sin A\cos B \pm \cos A\sin B$$

$$\cos(A \pm B) \equiv \cos A\cos B \mp \sin A\sin B$$

$$\tan(A \pm B) \equiv \frac{\tan A \pm \tan B}{1 \mp \tan A\tan B}$$

$$\sin 2A \equiv 2\sin A\cos A$$

$$\cos 2A \equiv \cos^2 A - \sin^2 A \equiv 2\cos^2 A - 1 \equiv 1 - 2\sin^2 A$$

$$\tan 2A \equiv \frac{2\tan A}{1 - \tan^2 A}$$

Principal values:

$$-\tfrac{1}{2}\pi \leqslant \sin^{-1} x \leqslant \tfrac{1}{2}\pi, \qquad 0 \leqslant \cos^{-1} x \leqslant \pi, \qquad -\tfrac{1}{2}\pi < \tan^{-1} x < \tfrac{1}{2}\pi$$

Differentiation

$\mathbf{f}(x)$	$\mathbf{f}'(x)$
x^n	nx^{n-1}
$\ln x$	$\frac{1}{x}$
e^x	e^x
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\tan x$	$\sec^2 x$
$\sec x$	$\sec x\tan x$
$\operatorname{cosec} x$	$-\operatorname{cosec} x\cot x$
$\cot x$	$-\operatorname{cosec}^2 x$
$\tan^{-1} x$	$\frac{1}{1+x^2}$
uv	$v\frac{du}{dx} + u\frac{dv}{dx}$
$\frac{u}{v}$	$\frac{v\frac{du}{dx} - u\frac{dv}{dx}}{v^2}$

If $x = f(t)$ and $y = g(t)$ then $\frac{dy}{dx} = \frac{dy}{dt} \div \frac{dx}{dt}$

Integration

(Arbitrary constants are omitted; a denotes a positive constant.)

$\mathbf{f}(x)$	$\int \mathbf{f}(x)\,\mathbf{d}x$	
x^n	$\frac{x^{n+1}}{n+1}$	$(n \neq -1)$
$\frac{1}{x}$	$\ln\lvert x\rvert$	
e^x	e^x	
$\sin x$	$-\cos x$	
$\cos x$	$\sin x$	
$\sec^2 x$	$\tan x$	
$\frac{1}{x^2+a^2}$	$\frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right)$	
$\frac{1}{x^2-a^2}$	$\frac{1}{2a}\ln\left\lvert\frac{x-a}{x+a}\right\rvert$	$(x > a)$
$\frac{1}{a^2-x^2}$	$\frac{1}{2a}\ln\left\lvert\frac{a+x}{a-x}\right\rvert$	$(\lvert x\rvert < a)$

$$\int u\frac{\mathrm{d}v}{\mathrm{d}x}\,\mathrm{d}x = uv - \int v\frac{\mathrm{d}u}{\mathrm{d}x}\,\mathrm{d}x$$

$$\int \frac{\mathrm{f}'(x)}{\mathrm{f}(x)}\,\mathrm{d}x = \ln\lvert \mathrm{f}(x)\rvert$$

Vectors

If $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$ then

$$\mathbf{a}.\mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3 = \lvert\mathbf{a}\rvert\lvert\mathbf{b}\rvert\cos\theta$$

PROBABILITY & STATISTICS

Summary statistics

For ungrouped data:

$$\bar{x} = \frac{\Sigma x}{n}, \qquad \text{standard deviation} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

For grouped data:

$$\bar{x} = \frac{\Sigma xf}{\Sigma f}, \qquad \text{standard deviation} = \sqrt{\frac{\Sigma(x-\bar{x})^2 f}{\Sigma f}} = \sqrt{\frac{\Sigma x^2 f}{\Sigma f} - \bar{x}^2}$$

Discrete random variables

$$\mathrm{E}(X) = \Sigma xp, \qquad \mathrm{Var}(X) = \Sigma x^2 p - \{\mathrm{E}(X)\}^2$$

For the binomial distribution $\mathrm{B}(n, p)$:

$$p_r = \binom{n}{r} p^r (1-p)^{n-r}, \qquad \mu = np, \qquad \sigma^2 = np(1-p)$$

For the geometric distribution Geo(p):

$$p_r = p(1-p)^{r-1}, \qquad \mu = \frac{1}{p}$$

For the Poisson distribution $\mathrm{Po}(\lambda)$

$$p_r = \mathrm{e}^{-\lambda} \frac{\lambda^r}{r!}, \qquad \mu = \lambda, \qquad \sigma^2 = \lambda$$

Continuous random variables

$$\mathrm{E}(X) = \int x\,\mathrm{f}(x)\,\mathrm{d}x, \qquad \mathrm{Var}(X) = \int x^2\,\mathrm{f}(x)\,\mathrm{d}x - \{\mathrm{E}(X)\}^2$$

Sampling and testing

Unbiased estimators:

$$\bar{x} = \frac{\Sigma x}{n}, \qquad s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{1}{n-1}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)$$

Central Limit Theorem:

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Approximate distribution of sample proportion:

$$\mathrm{N}\left(p, \frac{p(1-p)}{n}\right)$$

FURTHER PROBABILITY & STATISTICS

Sampling and testing

Two-sample estimate of a common variance:

$$s^2 = \frac{\Sigma(x_1 - \bar{x}_1)^2 + \Sigma(x_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

Probability generating functions

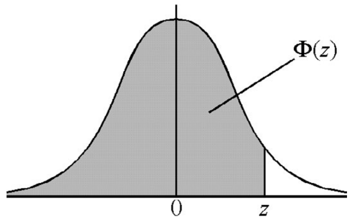
$\mathrm{G}_X(t) = \mathrm{E}(t^X)$, $\quad \mathrm{E}(X) = \mathrm{G}'_X(1)$, $\quad \mathrm{Var}(X) = \mathrm{G}''_X(1) + \mathrm{G}'_X(1) - \{\mathrm{G}'_X(1)\}^2$

THE NORMAL DISTRIBUTION FUNCTION

If Z has a normal distribution with mean 0 and variance 1, then, for each value of z, the table gives the value of $\Phi(z)$, where

$$\Phi(z) = \mathrm{P}(Z \leqslant z).$$

For negative values of z, use $\Phi(-z) = 1 - \Phi(z)$.



z	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
															ADD				
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	4	8	12	16	20	24	28	32	36
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	4	8	12	16	20	24	28	32	36
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	4	8	12	15	19	23	27	31	35
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	7	11	15	19	22	26	30	34
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	4	7	11	14	18	22	25	29	32
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	3	7	10	14	17	20	24	27	31
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	3	7	10	13	16	19	23	26	29
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	3	6	9	12	15	18	21	24	27
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	3	5	8	11	14	16	19	22	25
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	3	5	8	10	13	15	18	20	23
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	2	5	7	9	12	14	16	19	21
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	2	4	6	8	10	12	14	16	18
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	2	4	6	7	9	11	13	15	17
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	2	3	5	6	8	10	11	13	14
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1	3	4	6	7	8	10	11	13
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1	2	4	5	6	7	8	10	11
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1	2	3	4	5	6	7	8	9
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1	2	3	4	4	5	6	7	8
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1	1	2	3	4	4	5	6	6
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	1	1	2	2	3	4	4	5	5
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0	1	1	2	2	3	3	4	4
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0	1	1	2	2	2	3	3	4
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0	1	1	1	2	2	2	3	3
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0	1	1	1	1	2	2	2	2
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0	0	1	1	1	1	1	2	2
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0	0	0	1	1	1	1	1	1
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0	0	0	0	1	1	1	1	1
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0	0	0	0	0	1	1	1	1
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0	0	0	0	0	0	0	1	1
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0	0	0	0	0	0	0	0	0

Critical values for the normal distribution

If Z has a normal distribution with mean 0 and variance 1, then, for each value of p, the table gives the value of z such that

$$\mathrm{P}(Z \leqslant z) = p.$$

p	0.75	0.90	0.95	0.975	0.99	0.995	0.9975	0.999	0.9995
z	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291