The Question Bank of

Further Probability and Statistics

for CAIE 9231 paper 4.

v1.0

Edited by Thoridal

Instructions for Use

- This question bank is organized by chapter for systematic revision.
- This question bank is complied based on the 26-27 CAIE Further Probability and Statistics syllabus, which is included as appendix.
- Each question includes its source for reference.
- Mark schemes are provided in the separate answer booklet.
- The formula sheet (MF19) is included as appendix.
- Use this resource for targeted practice and exam preparation.

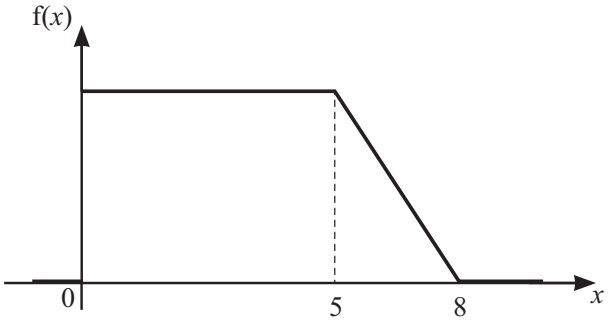
Edited by Y.Liu

Contents

Chapter 1

Continuous random variables

1. [9231/s25/41/q2]



As shown in the diagram, the continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} a & 0 \leqslant x \leqslant 5, \\ b - cx & 5 \leqslant x \leqslant 8, \\ 0 & \text{otherwise}, \end{cases}$$

where a, b and c are constants.

(a) Show that $a = \frac{2}{13}$ and find the values of b and c. [3]

(b) Find the mean of X. [3]

(c) Find the median of X. [2]

The random variable Y is defined by $Y = X^2$.

(d) Find the cumulative distribution function for Y. [4]

2. [9231/s25/43/q3]

A continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} kx & 0 \leqslant x < 1, \\ k(8-x) & 1 \leqslant x \leqslant 8, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(a) Show that $k = \frac{1}{25}$. [2]

(b) Find the median value of X. [3]

The random variable Y is defined by $Y = \sqrt[3]{X}$.

(c) Find the probability density function of Y. [5]

3. [9231/s25/44/q4]

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} kx & 0 \leqslant x < 1, \\ kx^2 & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that $k = \frac{6}{17}$. [2]

(b) Find the cumulative distribution function of X. [3]

(c) Find the median value of X. [2]

(d) Find $\mathrm{E}\left(\frac{1}{X}\right)$. [2]

4. [9231/w25/41/q5]

A continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{16}\sqrt{x} & 0 \leqslant x < 4, \\ \frac{1}{k\sqrt{x}} & 4 \leqslant x \leqslant 9, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(a) Show that $k = 3$. [2]

(b) Find the median value of X. [3]

The random variable Y is defined by $Y = \sqrt{X}$.

(c) Find the probability density function of Y. [5]

5. [9231/w25/42/q4]

A continuous random variable X has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 1, \\ \frac{1}{5}x + a & 1 \leqslant x < 4, \\ \frac{1}{50}x^2 + b & 4 \leqslant x \leqslant 6, \\ 1 & x > 6, \end{cases}$$

where a and b are constants.

(a) Find the value of a and the value of b. [2]

(b) Find the probability density function of X. [2]

(c) Given that $\mathrm{E}(X) = \frac{529}{150}$, find $\mathrm{Var}(X)$. [3]

(d) Find the 10th and 90th percentiles of X. [3]

6. [9231/w25/44/q2]

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{4}{9}(x+1) & 0 \leqslant x \leqslant 1, \\ (x-2)^2 & 1 < x \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the cumulative distribution function of X. [3]

(b) Find the exact value of the median of X. [2]

The random variable Y is defined by $Y = \sqrt{X}$.

(c) Find the cumulative distribution function of Y. [3]

(d) Determine whether the median of Y is greater than, or less than, the median of X. [2]

7. [9231/s24/41/q7]

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{x}{4}(4-x^2) & 0 \leqslant x \leqslant 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find $\mathrm{Var}\left(\sqrt{X}\right)$. [4]

The continuous random variable Y is defined by $Y = X^2$.

(b) Find the probability density function of Y. [4]

(c) Find the exact value of the median of Y. [2]

8. [9231/s24/43/q5]

The continuous random variable X has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 2, \\ \dfrac{(x-2)^2}{12} & 2 \leqslant x < 4, \\ 1 - \dfrac{(8-x)^2}{24} & 4 \leqslant x \leqslant 8, \\ 1 & x > 8. \end{cases}$$

(a) Sketch the graph of the probability density function of X. [3]

(b) Find $\mathrm{E}(X)$. [3]

(c) Find the exact value of the interquartile range of X. [4]

9. [9231/w24/41/q4]

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} kx^3 & 0 \leqslant x < 1, \\ k(5-x) & 1 \leqslant x \leqslant 5, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(a) Sketch the graph of f. [1]

(b) Show that $k = \frac{4}{33}$. [2]

(c) Find the cumulative distribution function of X. [3]

(d) Find the median value of X. [4]

10. [9231/w24/42/q4]

The random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{21}(x-1)^2 & 2 \leqslant x \leqslant 5, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the cumulative distribution function of X. [3]

The random variable Y is defined by $Y = (X-1)^4$.

(b) Find the probability density function of Y. [3]

(c) Find the median value of Y. [2]

(d) Find $E(Y)$. [2]

11. [9231/s23/41/q6]

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{3}{28}\left(\mathrm{e}^{\frac{1}{2}x} + 4\mathrm{e}^{-\frac{1}{2}x}\right) & 0 \leqslant x \leqslant 2\ln 3, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the cumulative distribution function of X. [3]

The random variable Y is defined by $Y = \mathrm{e}^{\frac{1}{2}(X)}$.

(b) Find the probability density function of Y. [3]

(c) Find the 30th percentile of Y. [3]

(d) Find $\mathrm{E}(Y^4)$. [2]

12. [9231/s23/43/q1]

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{6}\left(x^{-\frac{1}{3}} - x^{-\frac{2}{3}}\right) & 1 \leqslant x \leqslant 27, \\ 0 & \text{otherwise}. \end{cases}$$

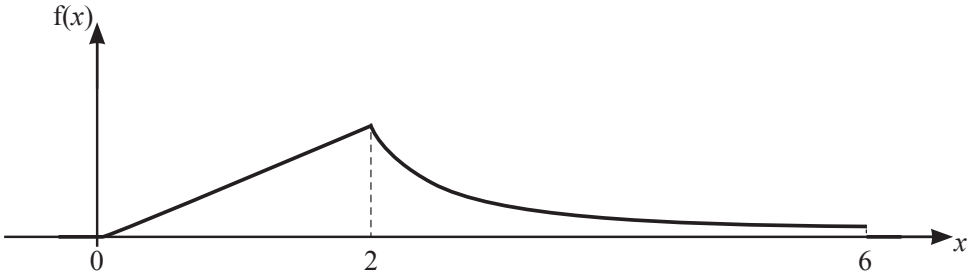
(a) Find the cumulative distribution function of X. [3]

The random variable Y is defined by $Y = X^{\frac{1}{3}}$.

(b) Find the probability density function of Y. [3]

(c) Find the exact value of the median of Y. [2]

13. [9231/w23/41/q4]



As shown in the diagram, the continuous random variable X has probability density function f given by

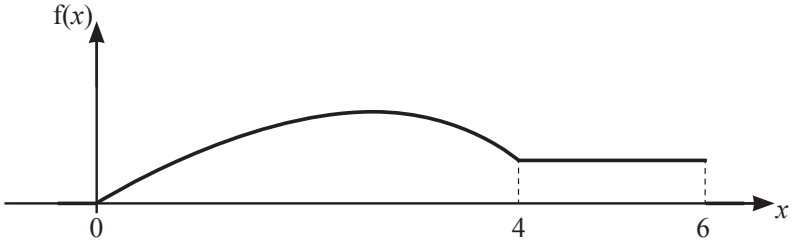
$$\mathrm{f}(x) = \begin{cases} mx & 0 \leqslant x \leqslant 2, \\ \dfrac{k}{x^2} + c & 2 \leqslant x \leqslant 6, \\ 0 & \text{otherwise,} \end{cases}$$

where m, k and c are constants.

(a) Given that $\mathrm{P}(X \leqslant 2) = \frac{1}{3}$, show that $m = \frac{1}{6}$ and find the values of k and c. [4]

(b) Find the exact numerical value of the interquartile range of X. [5]

14. [9231/w23/42/q4]



The diagram shows the continuous random variable X with probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{128}(4ax - bx^3) & 0 \leqslant x \leqslant 4, \\ c & 4 \leqslant x \leqslant 6, \\ 0 & \text{otherwise,} \end{cases}$$

where a, b and c are constants.

The upper quartile of X is equal to 4.

(a) Show that $c = \frac{1}{8}$ and find the values of a and b. [4]

(b) Find the exact value of the median of X. [3]

(c) Find $\mathrm{E}(\sqrt{X})$, giving your answer correct to 2 decimal places. [3]

15. [9231/s22/41/q3]

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} kx(4-x) & 0 \leqslant x < 2, \\ k(6-x) & 2 \leqslant x \leqslant 6, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(a) Show that $k = \frac{3}{40}$. [1]

(b) Given that $\mathrm{E}(X) = 2.5$, find $\mathrm{Var}(X)$. [3]

(c) Find the median value of X. [4]

16. [9231/s22/43/q4]

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{3}{8}\left(1 + \frac{1}{x^2}\right) & 1 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find $E\left(\sqrt{X}\right)$. [3]

The random variable Y is given by $Y = X^2$.

(b) Find the probability density function of Y. [4]

(c) Find the 40th percentile of Y. [3]

17. [9231/w22/41/q5]

The continuous random variable X has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0, \\ 1 - \frac{1}{144}(12-x)^2 & 0 \leqslant x \leqslant 12, \\ 1 & x > 12. \end{cases}$$

(a) Find the upper quartile of X. [2]

(b) Find $\mathrm{Var}(X^2)$. [5]

The random variable Y is given by $Y = \sqrt{X}$.

(c) Find the probability density function of Y. [3]

18. [9231/w22/42/q4]

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} k & 0 \leqslant x < 1, \\ kx & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(a) Show that $k = \frac{2}{5}$. [1]

(b) Find the interquartile range of X. [5]

(c) Find $\mathrm{Var}(X)$. [4]

19. [9231/s21/41/q3]

The continuous random variable X has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{81}x^2 & 0 \leqslant x \leqslant 9, \\ 1 & x > 9. \end{cases}$$

(a) Find $\mathrm{E}\left(\sqrt{X}\right)$. [3]

(b) Find $\mathrm{Var}\left(\sqrt{X}\right)$. [2]

(c) The random variable Y is given by $Y^3 = X$. Find the probability density function of Y. [3]

20. [9231/s21/43/q6]

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{8} & 0 \leqslant x < 1, \\ \frac{1}{28}(8-x) & 1 \leqslant x \leqslant 8, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the cumulative distribution function of X. [3]

(b) Find the value of the constant a such that $\mathrm{P}(X \leqslant a) = \frac{5}{7}$. [3]

The random variable Y is given by $Y = \sqrt[3]{X}$.

(c) Find the probability density function of Y. [5]

21. [9231/w21/41/q2]

The continuous random variable X has cumulative distribution function F given by

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

(a) Find the probability density function of X. [2]

(b) Find $\mathrm{P}\left(-\frac{1}{2} \leqslant X \leqslant \frac{1}{2}\right)$. [2]

(c) Find $\mathrm{E}(X^2)$. [2]

(d) Find $\mathrm{Var}(X^2)$. [2]

22. [9231/w21/42/q3]

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} a + \frac{1}{5}x & 0 \leqslant x < 1, \\ 2a - \frac{1}{5}x & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where a is a constant.

(a) Find the value of a. [3]

(b) Find $\mathrm{E}(X^2)$. [2]

(c) Find the cumulative distribution function of X. [3]

23. [9231/s20/41/q3]

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{3}{16}\left(2-\sqrt{x}\right) & 0 \leqslant x < 1, \\ \dfrac{3}{16\sqrt{x}} & 1 \leqslant x \leqslant 9, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find E(X). [3]

The random variable Y is such that $Y = \sqrt{X}$.

(b) Find the probability density function of Y. [5]

24. [9231/s20/43/q3]

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{5}x & 0 \leqslant x < 2, \\ \frac{2}{15}(5-x) & 2 \leqslant x \leqslant 5, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the cumulative distribution function of X. [3]

(b) Find the median value of X. [2]

(c) Find $\mathrm{E}(X^2)$. [2]

(d) Find $\mathrm{P}(1 \leqslant X \leqslant 3)$. [2]

25. [9231/w20/41/q6]

The continuous random variable X has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{60}(16x - x^2) & 0 \leqslant x \leqslant 6, \\ 1 & x > 6. \end{cases}$$

(a) Find the interquartile range of X. [4]

(b) Find $\mathrm{E}(X^3)$. [4]

The random variable Y is such that $Y = \sqrt{X}$.

(c) Find the probability density function of Y. [3]

26. [9231/w20/42/q4]

The continuous random variable X has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 2, \\ \frac{1}{60}x^2 - \frac{1}{15} & 2 \leqslant x \leqslant 8, \\ 1 & x > 8. \end{cases}$$

(a) Find $\mathrm{P}(3 \leqslant X \leqslant 6)$. [1]

(b) Find $\mathrm{E}\left(\sqrt{X}\right)$. [3]

(c) Find $\mathrm{Var}\left(\sqrt{X}\right)$. [2]

(d) The random variable Y is defined by $Y = X^3$. Find the probability density function of Y. [3]

27. [9231/s19/21/q7]

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \dfrac{3}{4x^2} + \dfrac{1}{4} & 1 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of X. [3]

(ii) Find the exact value of the interquartile range of X. [5]

28. [9231/w19/21/q7]

The time, T days, before an electrical component develops a fault has distribution function F given by

$$\mathrm{F}(t) = \begin{cases} 1 - \mathrm{e}^{-at} & t \geqslant 0, \\ 0 & \text{otherwise,} \end{cases}$$

where a is a positive constant. The mean value of T is 200.

(i) Write down the value of a. [1]

(ii) Find the probability that an electrical component of this type develops a fault in less than 150 days. [2]

A piece of equipment contains n of these components, which develop faults independently of each other. The probability that, after 150 days, at least one of the n components has not developed a fault is greater than 0.99.

(iii) Find the smallest possible value of n. [4]

29. [9231/w19/21/q10]

The random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{30}\left(\frac{8}{x^2} + 3x^2 - 14\right) & 2 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of X. [3]

The random variable Y is defined by $Y = X^2$.

(ii) Find the probability density function of Y. [4]

(iii) Find the value of y such that $P(Y < y) = 0.8$. [3]

30. [9231/s18/21/q6]

The continuous random variable X has distribution function given by

$$\mathrm{F}(x) = \begin{cases} 1 - \mathrm{e}^{-0.4x} & x \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find $\mathrm{P}(X > 2)$. [2]

(ii) Find the interquartile range of X. [4]

31. [9231/s18/23/q9]

The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{20}\left(3 - \frac{1}{\sqrt{x}}\right) & 1 \leqslant x \leqslant 9, \\ 0 & \text{otherwise.} \end{cases}$$

The random variable Y is defined by $Y = \sqrt{X}$.

(i) Show that the probability density function of Y is given by

$$g(y) = \begin{cases} \frac{1}{10}(3y - 1) & 1 \leqslant y \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$ [7]

(ii) Find the mean value of Y. [2]

32. [9231/w18/21/q6]

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{80}\left(3\sqrt{x} - \frac{8}{\sqrt{x}}\right) & 4 \leqslant x \leqslant 16, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of X. [3]

The random variable Y is defined by $Y = \sqrt{X}$.

(ii) Find the probability density function of Y. [3]

33. [9231/w18/22/q7]

The continuous random variable X has distribution function given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{90}(x^2 + x^4) & 0 \leqslant x \leqslant 3, \\ 1 & x > 3. \end{cases}$$

The random variable Y is defined by $Y = X^2$.

(i) Find the probability density function of Y. [4]

(ii) Find the mean value of Y. [2]

34. [9231/s17/21/q8]

The continuous random variable X has probability density function f given by

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

(i) Find the distribution function of X. [3]

The random variable Y is defined by $Y = (X-1)^3$.

(ii) Find the probability density function of Y. [4]

(iii) Find the median value of Y. [3]

35. [9231/s17/23/q9]

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} 0 & x < 0, \\ ae^{-x\ln 2} & x \geqslant 0, \end{cases}$$

where a is a positive constant.

(i) Find the value of a. [2]

(ii) State the value of $E(X)$. [1]

(iii) Find the interquartile range of X. [4]

The variable Y is related to X by $Y = 2^X$.

(iv) Find the probability density function of Y. [5]

36. [9231/w17/21/q7]

The random variable X has probability density function f given by

$$f(x) = \begin{cases} 0.2e^{-0.2x} & x \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of X. [2]

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

(iii) Find the median of X. [3]

37. [9231/s16/21/q8]

The random variable X has probability density function f given by

$$f(x) = \begin{cases} 2e^{-2x} & x \geqslant 0, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of X. [2]

(ii) Find the median value of X. [3]

The random variable Y is defined by $Y = e^X$.

(iii) Find the probability density function of Y. [4]

38. [9231/w16/21/q7]

The random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{6}x & 2 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of X. [3]

The random variable Y is defined by $Y = X^3$. Find

(ii) the probability density function of Y, [2]

(iii) the value of k for which $\mathrm{P}(Y \geqslant k) = \frac{7}{12}$. [3]

39. [9231/s15/23/q9]

The continuous random variable $\boldsymbol{X}$ has probability density function given by

$$f(\boldsymbol{x}) = \begin{cases} 0 & \boldsymbol{x} < 2, \\ \boldsymbol{a}e^{-(\boldsymbol{x}-2)} & \boldsymbol{x} \geqslant 2, \end{cases}$$

where $\boldsymbol{a}$ is a constant. Show that $\boldsymbol{a} = 1$. [1]

Find the distribution function of $\boldsymbol{X}$ and hence find the median value of $\boldsymbol{X}$. [5]

The random variable $\boldsymbol{Y}$ is defined by $\boldsymbol{Y} = e^{\boldsymbol{X}}$. Find

(i) the probability density function of $\boldsymbol{Y}$, [4]

(ii) $P(\boldsymbol{Y} > 10)$. [2]

40. [9231/w15/21/q7]

The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{21}x^2 & 1 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

The random variable Y is defined by $Y = X^2$. Show that Y has probability density function given by

$$g(y) = \begin{cases} \frac{1}{42}y^{\frac{1}{2}} & 1 \leqslant y \leqslant 16, \\ 0 & \text{otherwise.} \end{cases}$$ [5]

Find

(i) the median value of Y, [2]

(ii) the expected value of Y. [2]

Chapter 2

Inference using normal and t-distributions

1. [9231/s25/41/q1]

The manager of a hardware store is interested in whether there is a difference in the amount spent per customer on weekdays ($\$x$) compared to weekends ($\$y$). Random samples of 120 customers on weekdays and 80 customers on weekends are taken and the amount spent by each customer is recorded. The results are summarised as follows.

$$\sum x = 10\,470 \qquad \sum (x-\overline{x})^2 = 12\,283 \qquad \sum y = 6560 \qquad \sum (y-\overline{y})^2 = 13\,520$$

Test at the 1% significance level whether there is a difference in the mean amount spent per customer on weekdays compared to weekends. You should not assume that the population variances of the amounts spent on weekdays and weekends are equal. [7]

2. [9231/s25/41/q4]

A researcher is interested in whether there is a difference between two schools in students' aptitude for English. She randomly chooses 10 students from school X and 8 students of a similar age from school Y to take a written English test. The scores for the students from school X (x) and school Y (y) are summarised as follows.

$$\Sigma x = 612 \qquad \Sigma x^2 = 40\,104 \qquad \Sigma y = 444 \qquad \Sigma y^2 = 27\,460$$

You should assume that the two distributions are normal and have the same population variance.

(a) Find a 95% confidence interval for the difference in the mean scores for students from school X and students from school Y in the written English test. [6]

(b) Use the confidence interval you found in part **(a)** to explain why there is insufficient evidence at a 5% significance level to suggest that the English scores of students from school X and students from school Y are different. [1]

3. [9231/s25/43/q2]

A farmer is investigating whether using a new fertiliser will increase the yield of tomato plants. The farmer selects 40 tomato plants at random and gives them the new fertiliser. The crop mass, x kg, of each of these 40 plants is recorded. The farmer selects a further 60 tomato plants at random and gives them a standard fertiliser. The crop mass, y kg, of each of these 60 plants is recorded. The results are summarised as follows.

$$\Sigma x = 168 \quad \Sigma x^2 = 720 \quad \Sigma y = 228 \quad \Sigma y^2 = 900$$

Find a 90% confidence interval for the difference in mean crop mass associated with each type of fertiliser. [7]

4. [9231/s25/43/q5]

A doctor is investigating the concentration of blood glucose in patients at risk of developing type 2 diabetes, where blood glucose is measured in appropriate units. The doctor claims that a particular intervention reduces the concentration by more than k units on average. A group of 8 at risk patients is selected at random and each patient follows the intervention for six months. The blood glucose concentrations before and after the intervention are given in the following table.

Patient	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Before	183	165	172	165	143	176	161	153
After	164	148	164	149	134	153	155	148

(a) Use a t-test at the 5% significance level to find the range of values of k for which the result of the test is to reject the null hypothesis. [7]

(b) State an assumption necessary for the test in part **(a)** to be valid. [1]

5. [9231/s25/44/q1]

A random sample of 12 observations of a normal random variable is taken. The results give unbiased estimates for the population mean and variance as 10.24 and 0.52 respectively.

Test, at the 10% significance level, the null hypothesis that the population mean is 10.6 against the alternative hypothesis that the population mean is less than 10.6. [4]

6. [9231/s25/44/q6]

Lina and Mona are two statisticians who also write songs. The 'time' of a song is the number of minutes for which it lasts. For a random sample of 10 of her songs, Lina calculates a 95% confidence interval for the population mean time, μ minutes. This confidence interval is $2.95 \leqslant \mu \leqslant 3.13$. The times, x minutes, of Lina's songs are normally distributed.

(a) Find the values of $\sum x$ and $\sum x^2$ for the 10 songs in Lina's sample. [5]

Mona's songs have times, y minutes, that are normally distributed. The times for a random sample of 8 of Mona's songs are summarised as follows.

$$\sum y = 24.8 \qquad \sum y^2 = 76.98$$

Mona claims that the population mean time of her songs is greater than the population mean time of Lina's songs.

(b) Assuming that the two distributions have the same population variance, test at the 5% significance level whether there is evidence to support Mona's claim. [8]

7. [9231/w25/41/q1]

A group of 10 school children are asked to estimate the size of an angle $\theta°$ in a given acute angled triangle. These estimates, in degrees, are as follows.

84 85 77 85 84 87 86 88 83 85

(a) Stating any assumptions you make, calculate a 95% confidence interval for θ. [5]

(b) Give a reason why the assumptions made in part **(a)** may not be appropriate in this case. [1]

8. [9231/w25/41/q3]

A random sample of 10 newborn baby boys is taken and their masses in kg are recorded. From this sample, the population standard deviation of all newborn baby boys is estimated as 0.6 kg. A random sample of 5 newborn baby girls is taken and their masses in kg are recorded as follows.

3.9 3.1 2.9 3.1 3.6

It is assumed that the masses of newborn baby boys and girls have the same population standard deviation, σ kg.

By pooling the two samples, calculate an estimate of σ. [4]

9. [9231/w25/41/q6]

Nine athletes in a club have a new coach. The coach adopts a new training programme which he believes will reduce the race times of these athletes. Each athlete completes a 1500 m time trial before and after completing the new training programme. Their times, in seconds (s), are recorded.

Athlete	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Time before training (s)	250	251	252	267	276	291	310	320	335
Time after training (s)	245	251	253	261	275	293	302	313	320

(a) Carry out a paired *t*-test at the 5% significance level to test the coach's belief. [7]

Further research suggests that the effects of the training programme tend to reduce the times of the slower athletes by more than those of the faster athletes.

(b) Suggest a reason why the paired *t*-test used in part **(a)** may not have been an appropriate test in this case. [1]

(c) Suggest a suitable alternative test that could have been used instead of a paired *t*-test. [1]

10. [9231/w25/42/q2]

A factory produces packets of biscuits. The total mass of biscuits in a packet has a normal distribution with mean μ. A random sample of 12 packets is taken and the mass of the contents of each packet, x g, is recorded. The results are summarised as follows.

$$\sum x = 2390 \qquad \sum x^2 = 476117$$

(a) Find a 99% confidence interval for μ. [4]

A test of the null hypothesis $\mu = k$ is carried out on this sample using a 5% significance level. The test does not support the alternative hypothesis $\mu < k$.

(b) Find the greatest possible value of k. [3]

11. [9231/w25/42/q5]

An engineer is comparing the tensile strengths of steel rods made from two machines, A and B. The engineer randomly selects 8 rods from machine A and 6 rods from machine B. The tensile strengths, in appropriate units, are given in the following table.

Machine A	402	403	415	412	409	407	406	410
Machine B	401	398	395	397	410	405		

You should assume that the two distributions are normal and have the same population variance.

Use a t-test at the 5% significance level to test whether there is any difference in the mean tensile strengths of steel rods from the two machines. [9]

12. [9231/w25/44/q1]

Local residents are concerned about the speed of cars travelling through their village. They record the speed, in km h^{-1}, of a random sample of 13 cars travelling through their village. The recorded speeds are as follows.

40 53 59 42 43 48 62 67 46 82 66 45 70

(a) Construct a 90% confidence interval for the population mean speed of cars passing through the village. [5]

(b) State an assumption that is necessary for the confidence interval found in **1(a)** to be valid. [1]

13. [9231/w25/44/q3]

Two different types of juice extractor, machine X and machine Y, are being compared. The manufacturer claims that machine Y extracts more juice per orange on average than machine X.

A random sample of 20 oranges is selected and each carefully measured. Oranges of a similar size and mass are then paired and numbered from 1 to 10. One orange from each pair is randomly allocated to machine X with the other allocated to machine Y. The amount of juice, in ml, extracted from each orange is recorded in the following table.

	pair									
	1	2	3	4	5	6	7	8	9	10
machine X	65	73	58	61	72	79	64	65	69	71
machine Y	68	72	64	63	75	82	63	63	72	74

(a) Use a t-test to test the manufacturer's claim at the 1% significance level. [7]

(b) State an assumption required for the test in **3(a)** to be valid. [1]

The manufacturer notices that the amount of juice extracted from the oranges in pair 4 were recorded incorrectly. In fact, the amount of juice extracted from machine X was 60 ml and from machine Y was 62 ml.

(c) Explain, with justification, whether the conclusion of the test in **3(a)** remains the same. [2]

14. [9231/s24/41/q1]

The times taken by members of a large cycling club to complete a cross-country circuit have a normal distribution with mean μ minutes. The times taken, x minutes, are recorded for a random sample of 14 members of the club. The results are summarised as follows, where $\overline{x}$ is the sample mean.

$$\overline{x} = 42.8 \qquad \sum(x-\overline{x})^2 = 941.5$$

Find a 95% confidence interval for μ. [4]

15. [9231/s24/41/q6]

Jade is a swimming instructor at a sports college. She claims that, as a result of an intensive training course, the mean time taken by students to swim 50 metres has reduced by more than 1 second. She chooses a random sample of 10 students. The times taken, in seconds, before and after the training course are recorded in the table.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Time before course	54.2	47.4	52.1	59.0	55.3	51.0	48.9	52.2	58.4	51.4
Time after course	50.1	46.3	52.5	58.8	51.4	48.4	49.5	48.7	58.3	51.4

(a) Test, at the 10% significance level, whether Jade's claim is justified. [7]

(b) State an assumption that is necessary for this test to be valid. [1]

16. [9231/s24/43/q2]

A rowing club has a large number of members. A random sample of 12 of these members is taken and the pulse rate, x beats per minute (bpm), of each is measured after a 30-minute training session. A 98% confidence interval for the population mean pulse rate, μ bpm, is calculated from the sample as $64.22 < \mu < 68.66$.

(a) Find the values of $\sum x$ and $\sum x^2$. [6]

(b) State an assumption that is necessary for the confidence interval to be valid. [1]

17. [9231/s24/43/q6]

Seva is investigating the lengths of the tails of adult wallabies in two regions of Australia, X and Y. He chooses a random sample of 50 adult wallabies from region X and records the lengths, x cm, of their tails. He also chooses a random sample of 40 adult wallabies from region Y and records the lengths, y cm, of their tails. His results are summarised as follows.

$$\Sigma x = 1080 \qquad \Sigma x^2 = 23\,480 \qquad \Sigma y = 940 \qquad \Sigma y^2 = 22\,220$$

It cannot be assumed that the population variances of the two distributions are the same.

(a) Find a 90% confidence interval for the difference between the population mean lengths of the tails of adult wallabies in regions X and Y. [6]

The population mean lengths of the tails of adult wallabies in regions X and Y are μ_X cm and μ_Y cm respectively.

(b) Test, at the 10% significance level, the null hypothesis $\mu_Y - \mu_X = 1.1$ against the alternative hypothesis $\mu_Y - \mu_X > 1.1$. State your conclusion in the context of the question. [4]

18. [9231/w24/41/q1]

Ellie is investigating the heights of two types of beech tree, A and B, in a certain region. She has chosen a random sample of 60 beech trees of type A in the region, recorded their heights, x m, and calculated unbiased estimates for the population mean and population variance as 35.6 m and 4.95 m^2 respectively.

Ellie also chooses a random sample of 50 beech trees of type B in the region and records their heights, y m. Her results are summarised as follows.

$$\Sigma y = 1654 \qquad \Sigma y^2 = 54\,850$$

Find a 95% confidence interval for the difference between the population mean heights of type A and type B beech trees in the region. [6]

19. [9231/w24/41/q6]

Ansal is investigating the wingspans of Monarch butterflies in two different regions, X and Y. He takes a random sample of 8 Monarch butterflies from region X and records their wingspans, x cm. His results are as follows.

8.2 7.0 7.3 8.8 7.8 8.5 9.2 7.4

Ansal also takes a random sample of 9 Monarch butterflies from region Y and records their wingspans, y cm. His results are summarised as follows.

$$\Sigma y = 71.10 \qquad \Sigma y^2 = 567.13$$

Ansal suspects that the mean wingspan of Monarch butterflies from region X is greater than the mean wingspan of Monarch butterflies from region Y. It is known that the wingspans of Monarch butterflies in regions X and Y are normally distributed with equal population variances.

Test, at the 10% significance level, whether Ansal's suspicion is supported by the data. [8]

20. [9231/w24/42/q1]

A scientist is investigating the lengths of the leaves of a certain type of plant. The scientist assumes that the lengths of the leaves of this type of plant are normally distributed. He measures the lengths, x cm, of the leaves of a random sample of 8 plants of this type. His results are as follows.

3.5 4.2 3.8 5.2 2.9 3.7 4.1 3.2

Find a 90% confidence interval for the population mean length of leaves of this type of plant. [4]

21. [9231/w24/42/q5]

Dev owns a small company which produces bottles of juice. He uses two machines, X and Y, to fill empty bottles with juice. Dev is investigating the volumes of juice in the bottles. He chooses a random sample of 35 bottles filled by machine X and a random sample of 60 bottles filled by machine Y. The volumes of juice, x and y respectively, measured in suitable units, are summarised by

$$\Sigma x = 30.8, \qquad \Sigma x^2 = 29.0, \qquad \Sigma y = 62.4, \qquad \Sigma y^2 = 76.8.$$

Dev claims that the mean volume of juice in bottles filled by machine Y is greater than the mean volume of juice in bottles filled by machine X. A test at the α% significance level suggests that there is sufficient evidence to support Dev's claim.

Find the set of possible values of α. [9]

22. [9231/s23/41/q1]

The lengths of the leaves of a particular type of tree are normally distributed with mean μcm. The lengths, $x$cm, of a random sample of 12 leaves of this type are recorded. The results are summarised as follows.

$$\sum x = 91.2 \qquad \sum x^2 = 695.8$$

Find a 95% confidence interval for μ. [4]

23. [9231/s23/41/q2]

The children at two large schools, P and Q, are all given the same puzzle to solve. A random sample of size 10 is taken from the children at school P. Their individual times to complete the puzzle give a sample mean of 9.12 minutes and an unbiased variance estimate of 2.16 minutes^2. A random sample of size 12 is taken from the children at school Q. Their individual times, x minutes, to complete the puzzle are summarised by

$$\Sigma x = 99.6 \qquad \Sigma(x-\overline{x})^2 = 21.5,$$

where $\overline{x}$ is the sample mean. Times to complete the puzzle are assumed to be normally distributed with the same population variance.

Test at the 5% significance level whether the population mean time taken to complete the puzzle by children at school P is greater than the population mean time taken to complete the puzzle by children at school Q. [8]

24. [9231/s23/43/q2]

Shane is studying the lengths of the tails of male red kangaroos. He takes a random sample of 14 male red kangaroos and measures the length of the tail, x m, for each kangaroo. He then calculates a 90% confidence interval for the population mean tail length, μ m, of male red kangaroos. He assumes that the tail lengths are normally distributed and finds that $1.11 \leqslant \mu \leqslant 1.14$.

Find the values of Σx and Σx^2 for this sample. [6]

25. [9231/s23/43/q4]

An inspector is checking the lengths of metal rods produced by two machines, X and Y. These rods should be of the same length, but the inspector suspects that those made by machine X are shorter, on average, than those made by machine Y. The inspector chooses a random sample of 80 rods made by machine X and a random sample of 60 rods made by machine Y. The lengths of these rods are x cm and y cm respectively. Her results are summarised as follows.

$$\Sigma x = 164.0 \quad \Sigma x^2 = 338.1 \quad \Sigma y = 124.8 \quad \Sigma y^2 = 261.1$$

(a) Test at the 10% significance level whether the data supports the inspector's suspicion. [8]

(b) Give a reason why it is not necessary to make any assumption about the distributions of the lengths of the rods. [1]

26. [9231/w23/41/q1]

Maya is an athlete who competes in 1500-metre races. Last summer her practice run times had mean 4.22 minutes. Over the winter she has done some intense training to try to improve her times. A random sample of 10 of her practice run times, x minutes, this summer are summarised as follows.

$$\sum x = 42.05 \qquad \sum x^2 = 176.83$$

Maya's new practice run times are normally distributed. She believes that on average her times have improved as a result of her training.

Test, at the 5% significance level, whether Maya's belief is supported by the data. [6]

27. [9231/w23/41/q3]

Scientists are studying the effects of exercise on LDL blood cholesterol levels. Over a three-month period, a large group of people exercised for 20 minutes each day. For a randomly chosen sample of 10 of these people, the LDL blood cholesterol levels were measured at the beginning and the end of the three-month period. The results, measured in suitable units, are as follows.

	Person	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Cholesterol level	Beginning	72	84	120	90	102	135	64	75	80	88
	End	64	76	105	92	105	115	67	75	75	84

(a) Test, at the 2.5% significance level, whether there is evidence that the population mean LDL blood cholesterol level has reduced by more than 2 units after the three-month period. [7]

(b) State any assumption that you have made in part **(a)**. [1]

28. [9231/w23/42/q1]

A factory produces small bottles of natural spring water. Two different machines, X and Y, are used to fill empty bottles with the water. A quality control engineer checks the volumes of water in the bottles filled by each of the machines. He chooses a random sample of 60 bottles filled by machine X and a random sample of 75 bottles filled by machine Y. The volumes of water, x and y respectively, in millilitres, are summarised as follows.

$$\Sigma x = 6345 \quad \Sigma(x-\overline{x})^2 = 243.8 \quad \Sigma y = 7614 \quad \Sigma(y-\overline{y})^2 = 384.9$$

$\overline{x}$ and $\overline{y}$ are the sample means of the volume of water in the bottles filled by machines X and Y respectively.

Find a 95% confidence interval for the difference between the mean volume of water in bottles filled by machine X and the mean volume of water in bottles filled by machine Y. [6]

29. [9231/s22/41/q1]

A manager is investigating the times taken by employees to complete a particular task as a result of the introduction of new technology. He claims that the mean time taken to complete the task is reduced by more than 0.4 minutes. He chooses a random sample of 10 employees. The times taken, in minutes, before and after the introduction of the new technology are recorded in the table.

Employee	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Time before new technology	10.2	9.8	12.4	11.6	10.8	11.2	14.6	10.6	12.3	11.0
Time after new technology	9.6	8.5	12.4	10.9	10.2	10.6	12.8	10.8	12.5	10.6

(a) Test at the 10% significance level whether the manager's claim is justified. [7]

(b) State an assumption that is necessary for this test to be valid. [1]

30. [9231/s22/41/q5]

Raman is researching the heights of male giraffes in a particular region. Raman assumes that the heights of male giraffes in this region are normally distributed. He takes a random sample of 8 male giraffes from the region and measures the height, in metres, of each giraffe. These heights are as follows.

5.2 5.8 4.9 6.1 5.5 5.9 5.4 5.6

(a) Find a 90% confidence interval for the population mean height of male giraffes in this region. [5]

Raman claims that the population mean height of male giraffes in the region is less than 5.9 metres.

(b) Test at the 2.5% significance level whether this sample provides sufficient evidence to support Raman's claim. [4]

31. [9231/s22/43/q1]

The times taken by members of a large quiz club to complete a challenge have a normal distribution with mean μ minutes. The times, x minutes, are recorded for a random sample of 8 members of the club. The results are summarised as follows, where $\overline{x}$ is the sample mean.

$$\overline{x} = 33.8 \qquad \sum(x-\overline{x})^2 = 94.5$$

Find a 95% confidence interval for μ. [4]

32. [9231/s22/43/q6]

A company has two machines, A and B, which independently fill small bottles with a liquid. The volumes of liquid per bottle, in suitable units, filled by machines A and B are denoted by x and y respectively. A scientist at the company takes a random sample of 40 bottles filled by machine A and a random sample of 50 bottles filled by machine B. The results are summarised as follows.

$$\Sigma x = 1120 \qquad \Sigma x^2 = 31400 \qquad \Sigma y = 1370 \qquad \Sigma y^2 = 37600$$

The population means of the volumes of liquid in the bottles filled by machines A and B are denoted by μ_A and μ_B.

(a) Test at the 2% significance level whether there is any difference between μ_A and μ_B. [8]

(b) Find the set of values of α for which there would be evidence at the α% significance level that $\mu_A - \mu_B$ is greater than 0.25. [4]

33. [9231/w22/41/q1]

Jasmine is researching the heights of pine trees in forests in two regions A and B. She chooses a random sample of 50 pine trees in region A and records their heights, x m. She also chooses a random sample of 60 pine trees in region B and records their heights, y m. Her results are summarised as follows.

$$\sum x = 1625 \qquad \sum x^2 = 53200 \qquad \sum y = 1854 \qquad \sum y^2 = 57900$$

Find a 95% confidence interval for the difference between the population mean heights of pine trees in regions A and B. [7]

34. [9231/w22/41/q6]

A company manufactures copper pipes. The pipes are produced by two different machines, A and B. An inspector claims that the mean diameter of the pipes produced by machine A is greater than the mean diameter of the pipes produced by machine B. He takes a random sample of 12 pipes produced by machine A and measures their diameters, x cm. His results are summarised as follows.

$$\sum x = 6.24 \qquad \sum x^2 = 3.26$$

He also takes a random sample of 10 pipes produced by machine B and measures their diameters in cm. His results are as follows.

0.48 0.53 0.47 0.54 0.54 0.55 0.46 0.55 0.50 0.48

The diameters of the pipes produced by each machine are assumed to be normally distributed with equal population variances.

Test at the 2.5% significance level whether the data supports the inspector's claim. [9]

35. [9231/w22/42/q1]

A basketball club has a large number of players. The heights, x m, of a random sample of 10 of these players are measured. A 90% confidence interval for the population mean height, μ m, of players in this club is calculated. It is assumed that heights are normally distributed. The confidence interval is $1.78 \leqslant \mu \leqslant 2.02$.

Find the values of $\sum x$ and $\sum x^2$ for this sample. [6]

36. [9231/w22/42/q3]

A scientist is investigating the masses of birds of a certain species in country *X* and country *Y*. She takes a random sample of 50 birds of this species from country *X* and a random sample of 80 birds of this species from country *Y*. She records their masses in kg, *x* and *y*, respectively. Her results are summarised as follows.

$$\Sigma x = 75.5 \qquad \Sigma x^2 = 115.2 \qquad \Sigma y = 116.8 \qquad \Sigma y^2 = 172.6$$

The population mean masses of these birds in countries *X* and *Y* are μ_x kg and μ_y kg respectively.

Test, at the 5% significance level, the null hypothesis $\mu_x = \mu_y$ against the alternative hypothesis $\mu_x > \mu_y$. State your conclusion in the context of the question. [8]

37. [9231/s21/41/q1]

A random sample of 7 observations of a variable X are as follows.

8.26 7.78 7.92 8.04 8.27 7.95 8.34

The population mean of X is μ.

(a) Test, at the 10% significance level, the null hypothesis $\mu = 8.22$ against the alternative hypothesis $\mu < 8.22$. [6]

(b) State an assumption necessary for the test in part **(a)** to be valid. [1]

38. [9231/s21/41/q4]

A scientist is investigating the lengths of the leaves of birch trees in different regions. He takes a random sample of 50 leaves from birch trees in region A and a random sample of 60 leaves from birch trees in region B. He records their lengths in cm, x and y, respectively. His results are summarised as follows.

$$\sum x = 282 \qquad \sum x^2 = 1596 \qquad \sum y = 328 \qquad \sum y^2 = 1808$$

The population mean lengths of leaves from birch trees in regions A and B are μ_A cm and μ_B cm respectively.

Carry out a test at the 5% significance level to test the null hypothesis $\mu_A = \mu_B$ against the alternative hypothesis $\mu_A \neq \mu_B$. [8]

39. [9231/s21/43/q1]

Farmer A grows apples of a certain variety. Each tree produces 14.8 kg of apples, on average, per year. Farmer B grows apples of the same variety and claims that his apple trees produce a higher mass of apples per year than Farmer A's trees. The masses of apples from Farmer B's trees may be assumed to be normally distributed.

A random sample of 10 trees from Farmer B is chosen. The masses, x kg, of apples produced in a year are summarised as follows.

$$\sum x = 152.0 \qquad \sum x^2 = 2313.0$$

Test, at the 5% significance level, whether Farmer B's claim is justified. [6]

40. [9231/s21/43/q3]

The heights, x m, of a random sample of 50 adult males from country A were recorded. The heights, y m, of a random sample of 40 adult males from country B were also recorded. The results are summarised as follows.

$$\sum x = 89.0 \qquad \sum x^2 = 159.4 \qquad \sum y = 67.2 \qquad \sum y^2 = 113.1$$

Find a 95% confidence interval for the difference between the mean heights of adult males from country A and adult males from country B. [8]

41. [9231/w21/41/q1]

The times taken for students at a college to run 200 m have a normal distribution with mean μ s. The times, x s, are recorded for a random sample of 10 students from the college. The results are summarised as follows, where $\overline{x}$ is the sample mean.

$$\overline{x} = 25.6 \qquad \Sigma(x-\overline{x})^2 = 78.5$$

(a) Find a 90% confidence interval for μ. [4]

A test of the null hypothesis $\mu = k$ is carried out on this sample, using a 10% significance level. The test does not support the alternative hypothesis $\mu < k$.

(b) Find the greatest possible value of k. [3]

42. [9231/w21/41/q4]

Manet has developed a new training course to help athletes improve their time taken to run 800 m. Manet claims that his course will decrease an athlete's time by more than 2 s on average. For a random sample of 10 athletes the times taken, in seconds, before and after the course are given in the following table.

Athlete	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Before	150	146	131	135	126	142	130	129	137	134
After	145	138	129	135	122	135	132	128	127	137

Use a *t*-test, at the 5% significance level, to test whether Manet's claim is justified, stating any assumption that you make. [8]

43. [9231/w21/42/q1]

The number, x, of pine trees was counted in each of 40 randomly chosen regions of equal size in country A. The number, y, of pine trees was counted in each of 60 randomly chosen regions of the same equal size in country B. The results are summarised as follows.

$$\sum x = 752 \qquad \sum x^2 = 14\,320 \qquad \sum y = 1548 \qquad \sum y^2 = 40\,200$$

Find a 95% confidence interval for the difference between the mean number of pine trees in regions of this size in countries A and B. [7]

44. [9231/w21/42/q6]

A scientist is investigating the masses of a particular type of fish found in lakes A and B. He chooses a random sample of 10 fish of this type from lake A and records their masses, x kg, as follows.

2.1	1.8	0.9	3.0	2.4	2.6	1.8	2.2	1.9	2.5

The scientist also chooses a random sample of 12 fish of this type from lake B, but he only has a summary of their masses, y kg, as follows.

$$\Sigma y = 24.48 \qquad \Sigma y^2 = 53.75$$

Test at the 10% significance level whether the mean mass of fish of this type in lake A is greater than the mean mass of fish of this type in lake B. You should state any assumptions that you need to make for the test to be valid. [10]

45. [9231/s20/41/q4]

A company has two different machines, X and Y, each of which fills empty cups with coffee. The manager is investigating the volumes of coffee, x and y, measured in appropriate units, in the cups filled by machines X and Y respectively. She chooses a random sample of 50 cups filled by machine X and a random sample of 40 cups filled by machine Y. The volumes are summarised as follows.

$$\Sigma x = 15.2 \quad \Sigma x^2 = 5.1 \quad \Sigma y = 13.4 \quad \Sigma y^2 = 4.8$$

The manager claims that there is no difference between the mean volume of coffee in cups filled by machine X and the mean volume of coffee in cups filled by machine Y.

Test the manager's claim at the 10% significance level. [9]

46. [9231/s20/41/q5]

A large number of children are competing in a throwing competition. The distances, in metres, thrown by a random sample of 8 children are as follows.

19.8 22.1 24.4 21.5 20.8 26.3 23.7 25.0

(a) Assuming that distances are normally distributed, test, at the 5% significance level, whether the population mean distance thrown is more than 22.0 metres. [7]

(b) Find a 95% confidence interval for the population mean distance thrown. [3]

47. [9231/s20/43/q2]

A random sample of 40 observations of a random variable X and a random sample of 50 observations of a random variable Y are taken. The resulting values for the sample means, $\overline{x}$ and $\overline{y}$, and the unbiased estimates, s_x^2 and s_y^2, for the population variances are as follows.

$$\overline{x} = 24.4 \qquad \overline{y} = 17.2 \qquad s_x^2 = 10.2 \qquad s_y^2 = 11.1$$

Find a 90% confidence interval for the difference between the population means of X and Y. [5]

48. [9231/s20/43/q5]

Students at two colleges, A and B, are competing in a computer games challenge.

(a) The time taken for a randomly chosen student from college A to complete the challenge has a normal distribution with mean μ minutes. The times taken, x minutes, are recorded for a random sample of 10 students chosen from college A. The results are summarised as follows.

$$\Sigma x = 828 \qquad \Sigma x^2 = 68622$$

A test is carried out on the data at the 5% significance level and the result supports the claim that $\mu > k$.

Find the greatest possible value of k. [4]

(b) A random sample of 8 students is chosen from college B. Their times to complete the same challenge give a sample mean of 79.8 minutes and an unbiased variance estimate of 9.966 minutes^2.

Use a 2-sample test at the 5% significance level to test whether the mean time for students at college B to complete the challenge is the same as the mean time for students at college A to complete the challenge. You should assume that the two distributions are normal and have the same population variance. [7]

49. [9231/w20/41/q1]

Kayla is investigating the lengths of the leaves of a certain type of tree found in two forests X and Y. She chooses a random sample of 40 leaves of this type from forest X and records their lengths, x cm. She also records the lengths, y cm, for a random sample of 60 leaves of this type from forest Y. Her results are summarised as follows.

$$\sum x = 242.0 \qquad \sum x^2 = 1587.0 \qquad \sum y = 373.2 \qquad \sum y^2 = 2532.6$$

Find a 90% confidence interval for the difference between the population mean lengths of leaves in forests X and Y. [7]

50. [9231/w20/41/q4]

Members of the Sprints athletics club have been taking part in an intense training scheme, aimed at reducing their times taken to run 400 m. For a random sample of 9 athletes from the club, the times taken, in seconds, before and after the training scheme are given in the following table.

Athlete	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Time before	48.8	48.2	50.3	49.6	49.4	48.9	47.6	50.3	48.4
Time after	47.9	47.8	49.6	49.1	49.6	48.9	47.7	49.1	48.1

The organiser of the training scheme claims that on average an athlete's time will be reduced by at least 0.3 seconds.

Test at the 10% significance level whether the organiser's claim is justified, stating any assumption that you make. [8]

51. [9231/w20/42/q1]

The heights of the members of a large sports club are normally distributed. A random sample of 11 members of the club is chosen and their heights, x cm, are measured. The results are summarised as follows, where $\bar{x}$ denotes the sample mean of x.

$$\bar{x} = 176.2 \qquad \sum(x-\bar{x})^2 = 313.1$$

Test, at the 5% significance level, the null hypothesis that the population mean height for members of this club is equal to 172.5 cm against the alternative hypothesis that the mean differs from 172.5 cm. [5]

52. [9231/w20/42/q6]

Nassa is researching the lengths of a particular type of snake in two countries, A and B.

(a) He takes a random sample of 10 snakes of this type from country A and measures the length, x m, of each snake. He then calculates a 90% confidence interval for the population mean length, μ m, for snakes of this type, assuming that snake lengths have a normal distribution. This confidence interval is $3.36 \leqslant \mu \leqslant 4.22$.

Find the sample mean and an unbiased estimate for the population variance. [4]

(b) Nassa also measures the lengths, y m, of a random sample of 8 snakes of this type taken from country B. His results are summarised as follows.

$$\Sigma y = 27.86 \qquad \Sigma y^2 = 98.02$$

Nassa claims that the mean length of snakes of this type in country B is less than the mean length of snakes of this type in country A. Nassa assumes that his sample from country B also comes from a normal distribution, with the same variance as the distribution from country A.

Test at the 10% significance level whether there is evidence to support Nassa's claim. [8]

53. [9231/s19/21/q8]

A large number of runners are attending a summer training camp. A random sample of 6 runners is chosen and their times to run 1500 m at the beginning of the camp and at the end of the camp are recorded. Their times, in minutes, are shown in the following table.

Runner	*A*	*B*	*C*	*D*	*E*	*F*
Time at beginning of camp	3.82	3.62	3.55	3.71	3.75	3.92
Time at end of camp	3.72	3.55	3.52	3.68	3.54	3.73

The organiser of the training camp claims that a runner's time will improve by more than 0.05 minutes between the beginning and end of the camp. Assuming that differences in time over the two runs are normally distributed, test at the 10% significance level whether the organiser's claim is justified. [8]

54. [9231/s19/21/q11o]

A farmer grows two different types of cherries, Type A and Type B. He assumes that the masses of each type are normally distributed. He chooses a random sample of 8 cherries of Type A. He finds that the sample mean mass is 15.1 g and that a 95% confidence interval for the population mean mass, μ g, is $13.5 \leqslant \mu \leqslant 16.7$.

(i) Find an unbiased estimate for the population variance of the masses of cherries of Type A. [3]

The farmer now chooses a random sample of 6 cherries of Type B and records their masses as follows.

12.2 13.3 16.4 14.0 13.9 15.4

(ii) Test at the 5% significance level whether the mean mass of cherries of Type B is less than the mean mass of cherries of Type A. You should assume that the population variances for the two types of cherry are equal. [9]

55. [9231/s19/23/q9]

A farmer grows large amounts of a certain crop. On average, the yield per plant has been 0.75 kg. The farmer has improved the soil in which the crop grows, and she claims that the yield per plant has increased. A random sample of 10 plants grown in the improved soil is chosen. The yields, x kg, are summarised as follows.

$$\Sigma x = 7.85 \qquad \Sigma x^2 = 6.19$$

(i) Test at the 5% significance level whether the farmer's claim is justified, assuming a normal distribution. [7]

(ii) Find a 95% confidence interval for the population mean yield for plants grown in the new soil. [3]

56. [9231/s19/23/q11o]

A company produces packets of sweets. Two different machines, A and B, are used to fill the packets. The manager decides to assess the performance of the two machines. He selects a random sample of 50 packets filled by machine A and a random sample of 60 packets filled by machine B. The masses of sweets, x kg, in packets filled by machine A and the masses of sweets, y kg, in packets filled by machine B are summarised as follows.

$$\Sigma x = 22.4 \qquad \Sigma x^2 = 10.1 \qquad \Sigma y = 28.8 \qquad \Sigma y^2 = 16.3$$

A test at the $\alpha\%$ significance level provides evidence that the mean mass of sweets in packets filled by machine A is less than the mean mass of sweets in packets filled by machine B. Find the set of possible values of α. [12]

57. [9231/w19/21/q6]

A random sample of 9 members is taken from the large number of members of a sports club, and their heights are measured. The heights of all the members of the club are assumed to be normally distributed. A 95% confidence interval for the population mean height, μ metres, is calculated from the data as $1.65 \leqslant \mu \leqslant 1.85$.

(i) Find an unbiased estimate for the population variance. [3]

(ii) Denoting the height of a member of the club by x metres, find Σx^2 for this sample of 9 members. [4]

58. [9231/w19/21/q8]

A random sample of 8 elephants from region A is taken and their weights, x tonnes, are recorded. (1 tonne = 1000 kg.) The results are summarised as follows.

$$\Sigma x = 32.4 \qquad \Sigma x^2 = 131.82$$

A random sample of 10 elephants from region B is taken. Their weights give a sample mean of 3.78 tonnes and an unbiased variance estimate of 0.1555 tonnes^2. The distributions of the weights of elephants in regions A and B are both assumed to be normal with the same population variance. Test at the 10% significance level whether the mean weight of elephants in region A is the same as the mean weight of elephants in region B. [9]

59. [9231/s18/21/q7]

A large number of athletes are taking part in a competition. The masses, in kg, of a random sample of 7 athletes are as follows.

98.1 105.0 92.2 89.8 99.9 95.4 101.2

Assuming that masses are normally distributed, test, at the 10% significance level, whether the mean mass of athletes in this competition is equal to 94 kg. [7]

60. [9231/s18/21/q10]

The times taken to run 400 metres by students at two large colleges P and Q are being compared. There is no evidence that the population variances are equal. The time taken by a student at college P and the time taken by a student at college Q are denoted by x seconds and y seconds respectively. A random sample of 50 students from college P and a random sample of 60 students from college Q give the following summarised data.

$$\Sigma x = 2620 \qquad \Sigma x^2 = 138\,200 \qquad \Sigma y = 3060 \qquad \Sigma y^2 = 157\,000$$

(i) Using a 10% significance level, test whether, on average, students from college P take longer to run 400 metres than students from college Q. [9]

(ii) Find a 90% confidence interval for the difference in the mean times taken to run 400 metres by students from colleges P and Q. [3]

61. [9231/s18/23/q10]

During the summer months, all members of a large swimming club take part in intensive training. The times taken to swim 50 metres at the beginning of the summer and at the end of the summer are recorded for each member of the club. The time taken, in seconds, at the beginning of the summer is denoted by x and the time taken at the end of the summer is denoted by y. For a random sample of 9 members the results are shown in the following table.

Member	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
x	38.5	40.2	32.3	35.1	36.2	41.4	32.0	38.2	38.2
y	37.4	38.1	31.6	34.7	34.2	38.6	31.8	36.3	36.8

The swimming coach believes that, on average, the time taken by a swimmer to swim 50 metres will decrease by more than one second as a result of the intensive training.

(i) Stating suitable hypotheses and assuming a normal distribution, test the coach's belief at the 10% significance level. [8]

(ii) Find a 95% confidence interval for the population mean time taken to swim 50 metres after the intensive training, assuming a normal distribution. [4]

62. [9231/w18/21/q8]

The weekly salaries of employees at two large electronics companies, A and B, are being compared. The weekly salaries of an employee from company A and an employee from company B are denoted by $\$x$ and $\$y$ respectively. A random sample of 50 employees from company A and a random sample of 40 employees from company B give the following summarised data.

$$\Sigma x = 5120 \qquad \Sigma x^2 = 531\,000 \qquad \Sigma y = 3760 \qquad \Sigma y^2 = 375\,135$$

(i) The population mean salaries of employees from companies A and B are denoted by $\$\mu_A$ and $\$\mu_B$ respectively. Using a 5% significance level, test the null hypothesis $\mu_A = \mu_B$ against the alternative hypothesis $\mu_A \neq \mu_B$. [8]

(ii) State, with a reason, whether any assumptions about the distributions of employees' salaries are needed for the test in part **(i)**. [1]

63. [9231/w18/21/q9]

There are a large number of students at a particular college. The heights, in metres, of a random sample of 8 students are as follows.

1.75 1.72 1.62 1.70 1.82 1.75 1.68 1.84

You may assume that heights of students are normally distributed.

(i) Test, at the 5% significance level, whether the population mean height of students at this college is greater than 1.70 metres. [7]

(ii) Find a 95% confidence interval for the population mean height of students at this college. [3]

64. [9231/w18/22/q6]

The heights, in metres, of a random sample of 8 trees of a particular type are as follows.

14.2 11.3 10.8 8.4 12.8 11.5 12.1 9.2

Assuming that heights of trees of this type are normally distributed, calculate a 95% confidence interval for the mean height of trees of this type. [6]

65. [9231/w18/22/q11o]

In a particular country, large numbers of ducks live on lakes A and B. The mass, in kg, of a duck on lake A is denoted by x and the mass, in kg, of a duck on lake B is denoted by y. A random sample of 8 ducks is taken from lake A and a random sample of 10 ducks is taken from lake B. Their masses are summarised as follows.

$$\Sigma x = 10.56 \qquad \Sigma x^2 = 14.1775 \qquad \Sigma y = 12.39 \qquad \Sigma y^2 = 15.894$$

A scientist claims that ducks on lake A are heavier on average than ducks on lake B.

(i) Test, at the 10% significance level, whether the scientist's claim is justified. You should assume that both distributions are normal and that their variances are equal. [9]

A second random sample of 8 ducks is taken from lake A and their masses are summarised as

$$\Sigma x = 10.24 \qquad \text{and} \qquad \Sigma(x - \bar{x})^2 = 0.294,$$

where $\bar{x}$ is the sample mean. The scientist now claims that the population mean mass of ducks on lake A is greater than p kg. A test of this claim is carried out at the 10% significance level, using only this second sample from lake A. This test supports the scientist's claim.

(ii) Find the greatest possible value of p. [5]

66. [9231/s17/21/q7]

A farmer grows a particular type of fruit tree. On average, the mass of fruit produced per tree has been 6.2 kg. He has developed a new kind of soil and claims that the mean mass of fruit produced per tree when growing in this new soil has increased. A random sample of 10 trees grown in the new soil is chosen. The masses, x kg, of fruit produced are summarised as follows.

$$\Sigma x = 72.0 \qquad \Sigma x^2 = 542.0$$

Test at the 5% significance level whether the farmer's claim is justified, assuming a normal distribution. [7]

67. [9231/s17/21/q9]

Two fish farmers X and Y produce a particular type of fish. Farmer X chooses a random sample of 8 of his fish and records the masses, x kg, as follows.

1.2 1.4 0.8 2.1 1.8 2.6 1.5 2.0

Farmer Y chooses a random sample of 10 of his fish and summarises the masses, y kg, as follows.

$$\Sigma y = 20.2 \qquad \Sigma y^2 = 44.6$$

You should assume that both distributions are normal with equal variances. Test at the 10% significance level whether the mean mass of fish produced by farmer X differs from the mean mass of fish produced by farmer Y. [10]

68. [9231/s17/23/q6]

The independent variables X and Y have distributions with the same variance σ^2. Random samples of N observations of X and $2N$ observations of Y are taken, and the results are summarised by

$$\Sigma x = 4, \quad \Sigma x^2 = 10, \quad \Sigma y = 8, \quad \Sigma y^2 = 102.$$

These data give a pooled estimate of 10 for σ^2. Find N. [5]

69. [9231/s17/23/q8]

The number, x, of beech trees was counted in each of 50 randomly chosen regions of equal size in beech forests in country A. The number, y, of beech trees was counted in each of 40 randomly chosen regions of the same equal size in beech forests in country B. The results are summarised as follows.

$$\Sigma x = 1416 \qquad \Sigma x^2 = 41\,100 \qquad \Sigma y = 888 \qquad \Sigma y^2 = 20\,140$$

Find a 95% confidence interval for the difference between the mean number of beech trees in regions of this size in country A and in country B. [9]

70. [9231/s17/23/q11o]

The times taken to run 200 metres at the beginning of the year and at the end of the year are recorded for each member of a large athletics club. The time taken, in seconds, at the beginning of the year is denoted by x and the time taken, in seconds, at the end of the year is denoted by y. For a random sample of 8 members, the results are shown in the following table.

Member	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
x	24.2	23.8	22.8	25.1	24.5	24.0	23.8	22.8
y	23.9	23.6	22.8	24.5	24.2	23.5	23.6	22.7

$[\Sigma x = 191, \quad \Sigma x^2 = 4564.46, \quad \Sigma y = 188.8, \quad \Sigma y^2 = 4458.4, \quad \Sigma xy = 4510.99.]$

(i) Find, showing all necessary working, the equation of the regression line of y on x. [4]

The athletics coach believes that, on average, the time taken by an athlete to run 200 metres decreases between the beginning and the end of the year by more than 0.2 seconds.

(ii) Stating suitable hypotheses and assuming a normal distribution, test the coach's belief at the 10% significance level. [8]

71. [9231/w17/21/q10]

A factory produces bottles of an energy juice. Two different machines are used to fill empty bottles with the juice. The manager chooses a random sample of 50 bottles filled by machine X and a random sample of 60 bottles filled by machine Y. The volumes of juice, x and y respectively, measured in appropriate units, are summarised by

$$\Sigma x = 45.5, \quad \Sigma(x - \bar{x})^2 = 19.56, \quad \Sigma y = 72.3, \quad \Sigma(y - \bar{y})^2 = 30.25,$$

where $\bar{x}$ and $\bar{y}$ are the sample means of the volume of juice in the bottles filled by X and Y respectively.

(i) Find a 90% confidence interval for the difference between the mean volume of juice in bottles filled by machine X and the mean volume of juice in bottles filled by machine Y. [7]

A test at the α% significance level does not provide evidence that there is any difference in the means of the volume of juice in bottles filled by machine X and the volume of juice in bottles filled by machine Y.

(ii) Find the set of possible values of α. [6]

72. [9231/w17/21/q11o]

A large number of people attended a course to improve the speed of their logical thinking. The times taken to complete a particular type of logic puzzle at the beginning of the course and at the end of the course are recorded for each person. The time taken, in minutes, at the beginning of the course is denoted by x and the time taken, in minutes, at the end of the course is denoted by y. For a random sample of 9 people, the results are summarised as follows.

$$\Sigma x = 45.3 \quad \Sigma x^2 = 245.59 \quad \Sigma y = 40.5 \quad \Sigma y^2 = 195.11 \quad \Sigma xy = 218.72$$

Ken attended the course, but his time to complete the puzzle at the beginning of the course was not recorded. His time to complete the puzzle at the end of the course was 4.2 minutes.

(i) By finding, showing all necessary working, the equation of a suitable regression line, find an estimate for the time that Ken would have taken to complete the puzzle at the beginning of the course. [5]

The values of $x - y$ for the sample of 9 people are as follows.

$$0.2 \quad 0.8 \quad 0.5 \quad 1.0 \quad 0.2 \quad 0.6 \quad 0.2 \quad 0.5 \quad 0.8$$

The organiser of the course believes that, on average, the time taken to complete the puzzle decreases between the beginning and the end of the course by more than 0.3 minutes.

(ii) Stating suitable hypotheses and assuming a normal distribution, test the organiser's belief at the $2\frac{1}{2}\%$ significance level. [9]

73. [9231/s16/21/q7]

A random sample of 9 observations of a normal variable X is taken. The results are summarised as follows.

$$\Sigma x = 24.6 \qquad \Sigma x^2 = 68.5$$

Test, at the 5% significance level, whether the population mean is greater than 2.5. [8]

74. [9231/s16/21/q11o]

Petra is studying a particular species of bird. She takes a random sample of 12 birds from nature reserve A and measures the wing span, x cm, for each bird. She then calculates a 95% confidence interval for the population mean wing span, μ cm, for birds of this species, assuming that wing spans are normally distributed. Later, she is not able to find the summary of the results for the sample, but she knows that the 95% confidence interval is $25.17 \leqslant \mu \leqslant 26.83$. Find the values of Σx and Σx^2 for this sample. [6]

Petra also measures the wing spans of a random sample of 7 birds from nature reserve B. Their wing spans, y cm, are as follows.

23.2 22.4 27.6 25.3 28.4 26.5 23.6

She believes that the mean wing span of birds found in nature reserve A is greater than the mean wing span of birds found in nature reserve B. Assuming that this second sample also comes from a normal distribution, with variance the same as the first distribution, test, at the 10% significance level, whether there is evidence to support Petra's belief. [8]

75. [9231/s16/23/q8]

The coach of a national athletics team carried out an investigation into the effect of high altitude training on the times of 400-metre runners. The times, in seconds, were recorded before and after a six-week period of high altitude training, for a random sample of 8 athletes. The results are given in the following table.

Athlete	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Before	52.3	56.2	54.3	49.2	48.4	50.1	46.8	51.1
After	51.6	53.2	54.9	49.0	48.1	49.2	46.8	49.6

Stating any assumption that you make, test, at the 2.5% significance level, whether there is an improvement in athletes' times after high altitude training. [8]

76. [9231/s16/23/q11o]

The annual salaries of workers at two factories, A and B, are to be compared. The salaries, in tens of thousands of dollars, at A and B, are denoted by x and y respectively. For a random sample of 40 workers in factory A and a random sample of 50 workers in factory B the results are as follows.

$$\Sigma x = 256.0 \qquad \Sigma x^2 = 1910.8 \qquad \Sigma y = 382.9 \qquad \Sigma y^2 = 3148.8$$

The population mean salaries for A and B are denoted by μ_A and μ_B respectively. The population variances for salaries at A and B cannot be assumed to be equal. Test, at the 1% significance level, whether μ_B is greater than μ_A. [10]

The width of an α% confidence interval for $\mu_B - \mu_A$ is found to be 1.82. Find the value of α. [4]

77. [9231/w16/21/q6]

A random sample of 8 observations of a normal random variable X has mean $\bar{x}$, where

$$\bar{x} = 6.246 \quad \text{and} \quad \Sigma(x - \bar{x})^2 = 0.784.$$

Test, at the 5% significance level, whether the population mean of X is less than 6.44. [7]

78. [9231/w16/21/q8]

The amounts spent on the weekly food shopping by families in the big city P and the small town Q are to be compared. The amounts spent, in dollars, in P and Q are denoted by x and y respectively. For a random sample of 60 families in P and a random sample of 50 families in Q, the amounts are summarised as follows.

$$\Sigma x = 9600 \qquad \Sigma x^2 = 1\,560\,000 \qquad \Sigma y = 7200 \qquad \Sigma y^2 = 1\,052\,500$$

Assuming a common population variance, find

(i) a pooled estimate for the population variance, [4]

(ii) a 95% confidence interval for the difference in the population means in P and Q. [5]

79. [9231/s15/21/q6]

The independent random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ have distributions with the same variance σ^2. Random samples of $\boldsymbol{N}$ observations of $\boldsymbol{X}$ and 10 observations of $\boldsymbol{Y}$ are taken, and the results are summarised by

$$\Sigma \boldsymbol{x} = 5, \qquad \Sigma \boldsymbol{x}^2 = 11, \qquad \Sigma \boldsymbol{y} = 10, \qquad \Sigma \boldsymbol{y}^2 = 160.$$

These data give a pooled estimate of 12 for σ^2. Find $\boldsymbol{N}$. [4]

80. [9231/s15/21/q7]

A random sample of 8 sunflower plants is taken from the large number grown by a gardener, and the heights of the plants are measured. A 95% confidence interval for the population mean, μ metres, is calculated from the sample data as $1.17 < \mu < 2.03$. Given that the height of a sunflower plant is denoted by $\boldsymbol{x}$ metres, find the values of $\Sigma \boldsymbol{x}$ and $\Sigma \boldsymbol{x}^2$ for this sample of 8 plants. [7]

81. [9231/s15/21/q10]

Young children at a primary school are learning to throw a ball as far as they can. The distance thrown at the beginning of the school year and the distance thrown at the end of the same school year are recorded for each child. The distance thrown, in metres, at the beginning of the year is denoted by x; the distance thrown, in metres, at the end of the year is denoted by y. For a random sample of 10 children, the results are shown in the following table.

Child	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
x	5.2	4.1	3.7	5.4	7.6	6.1	3.2	4.0	3.5	8.0
y	6.2	4.8	5.0	5.6	7.7	7.0	4.0	4.5	3.6	8.5

[$\Sigma x = 50.8$, $\Sigma x^2 = 284.16$, $\Sigma y = 56.9$, $\Sigma y^2 = 347.59$, $\Sigma xy = 313.28$.]

A particular child threw the ball a distance of 7.0 metres at the beginning of the year, but he could not throw at the end of the year because he had broken his arm. By finding the equation of an appropriate regression line, estimate the distance this child would have thrown at the end of the year. [5]

The teacher suspects that, on average, the distance thrown by a child increases between the two throws by more than 0.4 metres. Stating suitable hypotheses and assuming a normal distribution, test the teacher's suspicion at the 5% significance level. [8]

82. [9231/s15/23/q8]

A large number of long jumpers are competing in a national long jump competition. The distances, in metres, jumped by a random sample of 7 competitors are as follows.

6.25 7.01 5.74 6.89 7.24 5.64 6.52

Assuming that distances are normally distributed, test, at the 5% significance level, whether the mean distance jumped by long jumpers in this competition is greater than 6.2 metres. [7]

The distances jumped by another random sample of 8 long jumpers in this competition are recorded. Using the data from this sample of 8 long jumpers, a 95% confidence interval for the population mean, μ metres, is calculated as $5.89 < \mu < 6.75$. Find the unbiased estimates for the population mean and population variance used in this calculation. [5]

83. [9231/s15/23/q10o]

The times taken, in hours, by cyclists from two different clubs, $\boldsymbol{A}$ and $\boldsymbol{B}$, to complete a 50 km time trial are being compared. The times taken by a cyclist from club $\boldsymbol{A}$ and by a cyclist from club $\boldsymbol{B}$ are denoted by $\boldsymbol{t_A}$ and $\boldsymbol{t_B}$ respectively. A random sample of 50 cyclists from $\boldsymbol{A}$ and a random sample of 60 cyclists from $\boldsymbol{B}$ give the following summarised data.

$$\Sigma \boldsymbol{t_A} = 102.0 \qquad \Sigma \boldsymbol{t_A^2} = 215.18 \qquad \Sigma \boldsymbol{t_B} = 129.0 \qquad \Sigma \boldsymbol{t_B^2} = 282.3$$

Using a 5% significance level, test whether, on average, cyclists from club $\boldsymbol{A}$ take less time to complete the time trial than cyclists from club $\boldsymbol{B}$. [10]

A test at the α% significance level shows that there is evidence that the population mean time for cyclists from club $\boldsymbol{B}$ exceeds the population mean time for cyclists from club $\boldsymbol{A}$ by more than 0.05 hours. Find the set of possible values of α. [4]

84. [9231/w15/21/q10o]

A farmer A grows two types of potato plants, Royal and Majestic. A random sample of 10 Royal plants is taken and the potatoes from each plant are weighed. The total mass of potatoes on a plant is x kg. The data are summarised as follows.

$$\Sigma x = 42.0 \qquad \Sigma x^2 = 180.0$$

A random sample of 12 Majestic plants is taken. The total mass of potatoes on a plant is y kg. The data are summarised as follows.

$$\Sigma y = 57.6 \qquad \Sigma y^2 = 281.5$$

Test, at the 5% significance level, whether the population mean mass of potatoes from Royal plants is the same as the population mean mass of potatoes from Majestic plants. You may assume that both distributions are normal and you should state any additional assumption that you make. [9]

A neighbouring farmer B grows Crown potato plants. His plants produce 3.8 kg of potatoes per plant, on average. Farmer A claims that her Royal plants produce a higher mean mass of potatoes than Farmer B's Crown plants. Test, at the 5% significance level, whether Farmer A's claim is justified. [5]

Chapter 3

x2-tests

1. [9231/s25/41/q3]

Eggs in a supermarket are sold in boxes of six. A supermarket manager wishes to model the number of broken eggs in the boxes sold in the store. A random sample of 2000 boxes is taken and the number of broken eggs recorded. The observed frequencies are shown in the table below.

Number of broken eggs	0	1	2	3	4	5	6
Observed frequency	1844	143	11	0	1	0	1

(a) Use the data to estimate the probability that an egg is broken. Give your answer correct to 4 significant figures. [1]

It is decided to carry out a goodness of fit test at the 0.5% significance level to determine whether a binomial distribution fits the data.

The observed frequencies and the expected frequencies are given in the following table.

Number of broken eggs	0	1	2	3	4	5	6
Observed frequency	1844	143	11	0	1	0	1
Expected frequency	1831.3	a	6.016	0.119	0.001	0.000	0.000

(b) Show that $a = 162.6$ correct to 1 decimal place. [1]

(c) Carry out a goodness of fit test at the 0.5% level of significance to test whether a binomial distribution is a satisfactory model for the data. [5]

(d) Give a reason why a binomial distribution may not be a suitable model in this situation. [1]

2. [9231/s25/43/q1]

A person's eye colour may be categorised as "brown", "blue" or "other". A scientist claims that these eye colours are uniformly distributed and hence are equally likely to occur in the population. A survey of 120 people from this population found that 38 people had brown eyes, 52 people had blue eyes and 30 people had eyes which were neither brown nor blue.

Use the data to carry out a goodness of fit test at the 5% significance level to test the scientist's claim. [6]

3. [9231/s25/44/q3]

A shop selling electrical goods has a team of three salespeople: Avril, Ben and Charlie. The manager wishes to investigate whether the salespeople are equally successful at selling particular types of items. The following table gives a record of a random sample of 250 sales of laptops, cameras and televisions, with the number sold by each of the three salespeople.

	Type of item			
	Laptop	Camera	Television	Total
Avril	31	40	24	95
Ben	23	45	29	97
Charlie	21	25	12	58
Total	75	110	65	250

Test, at the 10% significance level, whether there is independence between the type of item sold and the salesperson. [7]

4. [9231/w25/41/q2]

The manager of a car park claims that the number of cars entering the car park follows a Poisson distribution with mean 2.8. The numbers of cars entering the car park are recorded on a working day during successive 5-minute periods. The following table contains the observed frequencies, together with most of the expected frequencies and their contributions to the χ^2-test statistic.

Number of cars	0	1	2	3	4	5	$\geqslant 6$
Observed frequency	2	15	31	29	13	3	7
Expected frequency	6.081	17.03	23.84	p	15.57	8.721	6.511
χ^2-test statistic	2.739	0.241	2.152	q	0.425	3.753	0.037

(a) Find the value of p and the value of q. [2]

(b) Carry out a goodness of fit test at the 5% significance level to investigate the manager's claim. [4]

5. [9231/w25/42/q3]

A traffic expert claims that the number of breakdowns occurring each day on a busy section of a motorway follows a Poisson distribution with mean 0.7. The number of breakdowns each day over a 200-day period was recorded. The following table contains the observed frequencies together with some of the expected frequencies using the expert's distribution.

Number of breakdowns per day	0	1	2	3	4	$\geqslant 5$
Observed frequency	88	73	26	7	3	3
Expected frequency	99.317	m	24.333	5.678	0.994	n

(a) Find the value of m and the value of n, correct to 3 decimal places. [2]

.....

.....

.....

(b) Carry out a goodness of fit test at the 5% significance level to investigate the expert's claim. [6]

6. [9231/w25/44/q5]

A driving instructor believes that the performance (pass or fail) of students when taking a driving test is associated with their age. The following table summarises the number of students who pass and who fail, and the ages in years of the students taking the test, over a period of three years.

	age of student			
	under 20	20–40	over 40	total
pass	34	41	6	81
fail	16	39	9	64
total	50	80	15	145

Test, at the 10% significance level, whether performance is independent of age of student. [7]

7. [9231/s24/41/q5]

Two companies, P and Q, produce a certain type of paint brush. An independent examiner rates the quality of the brushes produced as poor, satisfactory or good. He takes a random sample of brushes from each company. The examiner's ratings are summarised in the table.

Company	Poor	Satisfactory	Good
P	18	43	64
Q	22	22	31

(a) Test, at the 5% significance level, whether quality of brushes is independent of company. [7]

(b) Compare the quality of the brushes produced by the two companies. [1]

8. [9231/s24/43/q3]

There are three bus companies in a city. The council is investigating whether the buses reliably arrive at their destination on time. The results from random samples of buses from each company are summarised in the following table.

		Bus company			
		A	*B*	*C*	Total
Arrival	Early	22	22	10	54
	On time	30	52	42	124
	Late	28	26	18	72
	Total	80	100	70	250

Test, at the 5% significance level, whether the reliability of buses is independent of bus company. [7]

9. [9231/w24/41/q3]

A statistician believes that the number of telephone calls received by an advice centre in a 10-minute interval can be modelled by the Poisson distribution Po(1.9). The number of calls received in a randomly chosen 10-minute interval was recorded on each of 100 days. The results are summarised in the table, together with some of the expected frequencies corresponding to the distribution Po(1.9).

Number of calls	0	1	2	3	4	5	6 or more
Observed frequency	10	18	35	21	11	4	1
Expected frequency	14.957	28.418	26.997				1.322

(a) Complete the table. [2]

(b) Carry out a goodness of fit test, at the 10% significance level, to determine whether the statistician's belief is reasonable. [6]

10. [9231/w24/42/q3]

Rosie sows 5 seeds in each of 150 plant pots. The number of seeds that germinate is recorded for each pot. The results are summarised in the following table.

Number of seeds that germinate	0	1	2	3	4	5
Number of pots	12	40	43	35	16	4

Rosie suggests that the number of seeds that germinate follows the binomial distribution B(5, p).

(a) Use Rosie's results to show that $p = 0.42$. [1]

(b) Carry out a goodness of fit test, at the 10% significance level, to test whether the distribution B(5, 0.42) is a good fit for the data. [9]

11. [9231/s23/41/q3]

A random sample of 50 values of the continuous random variable X was taken. These values are summarised in the following table.

Interval	$1 \leqslant x < 1.5$	$1.5 \leqslant x < 2$	$2 \leqslant x < 2.5$	$2.5 \leqslant x < 3$	$3 \leqslant x < 3.5$	$3.5 \leqslant x \leqslant 4$
Observed frequency	3	3	8	11	13	12

It is required to test the goodness of fit of the distribution with probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{24}\left(\frac{4}{x^2} + x^2\right) & 1 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

The expected frequencies, correct to 4 decimal places, are given in the following table.

Interval	$1 \leqslant x < 1.5$	$1.5 \leqslant x < 2$	$2 \leqslant x < 2.5$	$2.5 \leqslant x < 3$	$3 \leqslant x < 3.5$	$3.5 \leqslant x \leqslant 4$
Expected frequency	4.4271	a	6.1285	8.4549	b	14.9678

(a) Show that $a = 4.6007$ and find the value of b. [3]

(b) Carry out a goodness of fit test, at the 10% significance level, to test whether f is a satisfactory model for the data. [6]

12. [9231/s23/43/q6]

A scientist is investigating whether the ability to remember depends on age. A random sample of 150 students in different age groups is chosen. Each student is shown a set of 20 objects for thirty seconds and then asked to list as many as they can remember. The students are graded *A* or *B* according to how many objects they remembered correctly: grade *A* for 16 or more correct and grade *B* for fewer than 16 correct. The results are shown in the table.

	Age of students		
	11−12 years	13−14 years	15−16 years
Grade *A*	25	16	19
Grade *B*	28	45	17

(a) Carry out a χ^2-test at the 2.5% significance level to test whether grade is independent of age of student. [7]

The scientist decides instead to use three grades: grade *A* for 16 or more correct, grade *B* for 10 to 15 correct and grade *C* for fewer than 10 correct. The results are shown in the following table.

	Age of students		
	11−12 years	13−14 years	15−16 years
Grade *A*	25	16	19
Grade *B*	12	27	11
Grade *C*	16	18	6

With this second set of data, the test statistic is calculated as 10.91.

(b) Complete the χ^2-test at the 2.5% significance level for this second set of data. [2]

(c) State, with a reason, whether you would prefer to use the result from part **(a)** or part **(b)** to investigate whether the ability to remember depends on age. [1]

13. [9231/w23/41/q2]

A town council has published its plans for redeveloping the town centre and residents are being asked whether they approve or disapprove. A random sample of 250 responses has been selected from residents in the four main streets in the town: North, East, South and West Streets. The results are shown in the table.

	North Street	East Street	South Street	West Street
Approve	33	54	42	26
Disapprove	19	39	28	9

Test, at the 5% significance level, whether the opinions of the residents are independent of the streets on which they live. [7]

14. [9231/w23/42/q2]

The number of breakdowns on a particular section of road is recorded each day over a period of 90 days. It is suggested that the number of breakdowns follows a Poisson distribution with mean 3.5. The data is summarised in the table, together with some of the expected frequencies resulting from the suggested Poisson distribution.

Number of breakdowns per day	0	1	2	3	4	5	6	7	8 or more
Observed frequency	0	5	13	17	21	16	9	5	4
Expected frequency	2.718	9.512	16.646		16.993	11.895		3.469	2.407

(a) Complete the table. [2]

(b) Carry out a goodness of fit test, at the 10% significance level, to determine whether or not Po(3.5) is a good fit to the data. [6]

15. [9231/s22/41/q4]

A scientist is investigating the numbers of a particular type of butterfly in a certain region. He claims that the numbers of these butterflies found per square metre can be modelled by a Poisson distribution with mean 2.5. He takes a random sample of 120 areas, each of one square metre, and counts the number of these butterflies in each of these areas. The following table shows the observed frequencies together with some of the expected frequencies using the scientist's Poisson distribution.

Number per square metre	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency	12	20	36	32	13	6	1	0
Expected frequency	9.85	24.63	30.78	25.65	p	8.02	3.34	q

(a) Find the values of p and q, correct to 2 decimal places. [2]

(b) Carry out a goodness of fit test, at the 10% significance level, to test the scientist's claim. [6]

16. [9231/s22/43/q2]

A scientist is investigating the size of shells at various beach locations. She selects four beach locations and takes a random sample of shells from each of these beaches. She classifies each shell as large or small. Her results are summarised in the following table.

		Beach location				
		A	*B*	*C*	*D*	Total
Size of shell	Large	68	69	96	81	314
	Small	28	55	64	39	186
	Total	96	124	160	120	500

Test, at the 10% significance level, whether the size of shell is independent of the beach location. [7]

17. [9231/w22/41/q2]

An organisation runs courses to train students to become engineers. These students are taught in groups of 8. The director of the organisation claims that on average 60% of the students in a group achieve a pass. A random sample of 150 groups of 8 students is chosen. The following table shows the observed frequencies together with some of the expected frequencies using the appropriate binomial distribution.

Number of passes per group	0	1	2	3	4	5	6	7	8
Observed frequency	0	0	8	24	45	36	26	10	1
Expected frequency	p	1.180	6.193	18.579	34.836	q	r	13.437	2.519

(a) Find the values of p, q and r giving your answers correct to 3 decimal places. [2]

(b) Carry out a goodness of fit test, at the 10% significance level, to test whether there is evidence to reject the director's claim. [6]

18. [9231/w22/42/q2]

In the colleges in three regions of a particular country, students are given individual targets to achieve. Their performance is measured against their individual target and graded as 'above target', 'on target' or 'below target'. For a random sample of students from each of the three regions, the observed frequencies are summarised in the following table.

		Region			
		A	*B*	*C*	Total
Performance	Above target	62	41	44	147
	On target	102	94	95	291
	Below target	56	45	61	162
	Total	220	180	200	600

Test, at the 10% significance level, whether performance is independent of region. [7]

19. [9231/s21/41/q2]

A driving school employs four instructors to prepare people for their driving test. The allocation of people to instructors is random. For each of the instructors, the following table gives the number of people who passed and the number who failed their driving test last year.

	Instructor *A*	Instructor *B*	Instructor *C*	Instructor *D*	Total
Pass	72	42	52	68	234
Fail	33	34	41	58	166
Total	105	76	93	126	400

Test at the 10% significance level whether success in the driving test is independent of the instructor. [7]

20. [9231/s21/43/q5]

Chai packs china mugs into cardboard boxes. Chai's manager suspects that breakages occur at random times and that the number of breakages may follow a Poisson distribution. He takes a small sample of observations and finds that the number of breakages in a one-hour period has a mean of 2.4 and a standard deviation of 1.5.

(a) Explain how this information tends to support the manager's suspicion. [2]

The manager now takes a larger sample and claims that the numbers of breakages in a one-hour period follow a Poisson distribution. The numbers of breakages in a random sample of 180 one-hour periods are summarised in the following table.

Number of breakages	0	1	2	3	4	5	6	7 or more
Frequency	21	33	46	31	23	16	10	0

The mean number of breakages calculated from this sample is 2.5.

(b) Use the data from this larger sample to carry out a goodness of fit test, at the 10% significance level, to test the claim. [8]

21. [9231/w21/41/q3]

A supermarket sells pears in packs of 8. Some of the pears in a pack may not be ripe, and the supermarket manager claims that the number of unripe pears in a pack can be modelled by the distribution B(8, 0.15).

A random sample of 150 packs was selected and the number of unripe pears in each pack was recorded. The following table shows the observed frequencies together with some of the expected frequencies using the manager's binomial distribution.

Number of unripe pears per pack	0	1	2	3	4	5	$\geqslant 6$
Observed frequency	35	48	43	15	6	3	0
Expected frequency	40.874	p	35.641	12.579	2.775	0.392	q

(a) Find the values of p and q. [2]

(b) Carry out a goodness of fit test, at the 5% significance level, to test whether the manager's claim is justified. [6]

22. [9231/w21/42/q2]

It is claimed that the heights of a particular age group of boys follow a normal distribution with mean 125 cm and standard deviation 12 cm. Observations for a randomly chosen group of 60 boys in this age group are summarised in the following table. The table also gives the expected frequencies, correct to 2 decimal places, based on the normal distribution with mean 125 cm and standard deviation 12 cm.

Height, x cm	$x < 100$	$100 \leqslant x < 110$	$110 \leqslant x < 120$	$120 \leqslant x < 130$	$130 \leqslant x < 140$	$x \geqslant 140$
Observed frequency	0	3	15	23	11	8
Expected frequency	1.12	5.22	13.97	19.38	13.97	6.34

(a) Show how the expected frequency for $130 \leqslant x < 140$ is obtained. [2]

(b) Carry out a goodness of fit test, at the 5% significance level, to determine whether the claim is supported by the data. [6]

23. [9231/s20/41/q1]

Two randomly selected groups of students, with similar ranges of abilities, take the same examination in different rooms. One group of 140 students takes the examination with background music playing. The other group of 210 students takes the examination in silence. Each student is awarded a grade for their performance in the examination and the numbers from each group gaining each grade are shown in the following table.

	Grade awarded		
	A	B	C
Background music	49	51	40
Silence	93	68	49

Test at the 10% significance level whether grades awarded are independent of whether background music is playing during the examination. [6]

24. [9231/s20/43/q1]

Young children are learning to read using two different reading schemes, *A* and *B*. The standards achieved are measured against the national average standard achieved and classified as above average, average or below average. For two randomly chosen groups of young children, the numbers in each category are shown in the table.

	Standard achieved		
	Above average	Average	Below average
Scheme *A*	31	35	22
Scheme *B*	19	50	43

Test at the 5% significance level whether standard achieved is independent of the reading scheme used. [6]

25. [9231/w20/41/q3]

Apples are sold in bags of 5. Based on her previous experience, Freya claims that the probability of any apple weighing more than 100 grams is 0.35, independently of other apples in the bag.

The apples in a random sample of 150 bags are checked and the number, x, in each bag weighing more than 100 grams is recorded. The results are shown in the following table.

x	0	1	2	3	4	5
Frequency	12	39	46	37	12	4

Carry out a goodness of fit test at the 5% significance level and hence comment on Freya's claim. [7]

26. [9231/w20/42/q3]

A random sample of 200 observations of the continuous random variable X was taken and the values are summarised in the following table.

Interval	$0 \leqslant x < 0.5$	$0.5 \leqslant x < 1$	$1 \leqslant x < 1.5$	$1.5 \leqslant x < 2$	$2 \leqslant x < 2.5$	$2.5 \leqslant x < 3$
Observed frequency	5	23	40	41	46	45

It is required to test the goodness of fit of the distribution with probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{9}x(4-x) & 0 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

Most of the relevant expected frequencies, correct to 2 decimal places, are given in the following table.

Interval	$0 \leqslant x < 0.5$	$0.5 \leqslant x < 1$	$1 \leqslant x < 1.5$	$1.5 \leqslant x < 2$	$2 \leqslant x < 2.5$	$2.5 \leqslant x < 3$
Expected frequency	p	q	37.96	43.52	43.52	37.96

(a) Show that $p = 10.19$ and find the value of q. [3]

(b) Carry out a goodness of fit test, at the 5% significance level, to test whether f is a satisfactory model for the data. [4]

27. [9231/s19/21/q9]

A random sample of 50 observations of the continuous random variable X was taken and the values are summarised in the following table.

Interval	$0 \leqslant x < 0.8$	$0.8 \leqslant x < 1.6$	$1.6 \leqslant x < 2.4$	$2.4 \leqslant x < 3.2$	$3.2 \leqslant x < 4$
Observed frequency	18	16	8	6	2

It is required to test the goodness of fit of the distribution with probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{3}{16}(4-x)^{\frac{1}{2}} & 0 \leqslant x < 4, \\ 0 & \text{otherwise.} \end{cases}$$

The relevant expected frequencies, correct to 2 decimal places, are given in the following table.

Interval	$0 \leqslant x < 0.8$	$0.8 \leqslant x < 1.6$	$1.6 \leqslant x < 2.4$	$2.4 \leqslant x < 3.2$	$3.2 \leqslant x < 4$
Expected frequency	14.22	12.54	10.59	8.18	4.47

(i) Show how the expected frequency for $1.6 \leqslant x < 2.4$ is obtained. [3]

(ii) Carry out a goodness of fit test at the 5% significance level. [7]

28. [9231/s19/23/q8]

Two salesmen, A and B, work at a company that arranges different types of holidays: self-catering, hotel and cruise. The table shows, for a random sample of 150 holidays, the number of each type arranged by each salesman.

		Type of holiday		
		Self-catering	Hotel	Cruise
Salesman	A	25	38	21
	B	28	21	17

Test at the 10% significance level whether the type of holiday arranged is independent of the salesman. [8]

29. [9231/w19/21/q11o]

The number of puncture repairs carried out each week by a small repair shop is recorded over a period of 40 weeks. The results are shown in the following table.

Number of repairs in a week	0	1	2	3	4	5	$\geqslant 6$
Number of weeks	6	15	9	6	3	1	0

(i) Calculate the mean and variance for the number of repairs in a week and comment on the possible suitability of a Poisson distribution to model the data. [3]

Records over a longer period of time indicate that the mean number of repairs in a week is 1.6. The following table shows some of the expected frequencies, correct to 3 decimal places, for a period of 40 weeks using a Poisson distribution with mean 1.6.

Number of repairs in a week	0	1	2	3	4	5	$\geqslant 6$
Expected frequency	8.076	12.921	10.337	5.513	2.205	a	b

(ii) Show that $a = 0.706$ and find the value of the constant b. [3]

(iii) Carry out a goodness of fit test of a Poisson distribution with mean 1.6, using a 10% significance level. [8]

30. [9231/s18/21/q8]

A manufacturer produces three types of car: hatchbacks, saloons and estates. Each type of car is available in one of three colours: silver, blue and red. The manufacturer wants to know whether the popularity of the colour of the car is related to the type of car. A random sample of 300 cars chosen by customers gives the information summarised in the following table.

		Colour of car		
		Silver	Blue	Red
Type of car	Hatchback	53	36	41
	Saloon	29	40	31
	Estate	28	24	18

Test at the 10% significance level whether the colour of car chosen by customers is independent of the type of car. [8]

31. [9231/s18/23/q11o]

A scientist carries out an experiment to investigate the quantity X, which takes the values 0, 1, 2, 3, 4, 5 or 6. He believes that the values taken by X follow a binomial distribution. He conducts 250 trials. His results are summarised in the following table.

x	0	1	2	3	4	5	6
Observed frequency	22	83	72	53	17	3	0

(i) Show that unbiased estimates of the mean and variance for these results are 1.876 and 1.266 respectively, correct to 3 decimal places. By evaluating the mean and variance of the distribution B(6, 0.313), explain why X could have this distribution. [4]

The expected frequencies corresponding to the distribution B(6, 0.313) are shown in the following table.

x	0	1	2	3	4	5	6
Observed frequency	22	83	72	53	17	3	0
Expected frequency	26.3	71.9	81.8	49.7	17.0	3.1	0.2

(ii) Show how the expected frequency for $x = 4$ is calculated. [2]

(iii) Test at the 5% significance level whether the scientist's belief is correct. [8]

32. [9231/w18/21/q11o]

A machine is used to produce metal rods. When the machine is working efficiently, the lengths, x cm, of the rods have a normal distribution with mean 150 cm and standard deviation 1.2 cm. The machine is checked regularly by taking random samples of 200 rods. The latest results are shown in the following table.

Interval	$146 \leqslant x < 147$	$147 \leqslant x < 148$	$148 \leqslant x < 149$	$149 \leqslant x < 150$
Observed frequency	1	2	23	52

$150 \leqslant x < 151$	$151 \leqslant x < 152$	$152 \leqslant x < 153$	$153 \leqslant x < 154$
69	36	15	2

As a first check, the sample is used to calculate an estimate for the mean.

(i) Show that an estimate for the mean from this sample is close to 150 cm. [2]

As a second check, the results are tested for goodness of fit of the normal distribution with mean 150 cm and standard deviation 1.2 cm. The relevant expected frequencies, found using the normal distribution function given in the List of Formulae (MF10), are shown in the following table.

Interval	$x < 147$	$147 \leqslant x < 148$	$148 \leqslant x < 149$	$149 \leqslant x < 150$
Observed frequency	1	2	23	52
Expected frequency	1.24	8.32	30.94	59.50

$150 \leqslant x < 151$	$151 \leqslant x < 152$	$152 \leqslant x < 153$	$153 \leqslant x$
69	36	15	2
59.50	30.94	8.32	1.24

(ii) Show how the expected frequency for $151 \leqslant x < 152$ is obtained. [3]

(iii) Test, at the 5% significance level, the goodness of fit of the normal distribution to the results. [7]

33. [9231/w18/22/q10]

The number of accidents, x, that occur each day on a motorway are recorded over a period of 40 days. The results are shown in the following table.

Number of accidents	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency	3	5	8	10	5	7	2	0

(i) Show that the mean number of accidents each day is 2.95 and calculate the variance for this sample. Explain why these values suggest that a Poisson distribution might fit the data. [3]

A Poisson distribution with mean 2.95, as found from the data, is used to calculate the expected frequencies, correct to 2 decimal places. The results are shown in the following table.

Number of accidents	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency	3	5	8	10	5	7	2	0
Expected frequency	2.09	6.18	9.11	8.96	6.61	3.90	1.92	1.23

(ii) Show how the expected frequency of 6.61 for $x = 4$ is obtained. [2]

(iii) Test at the 5% significance level the goodness of fit of this Poisson distribution to the data. [7]

34. [9231/s17/21/q11o]

A shop is supplied with large quantities of plant pots in packs of six. These pots can be damaged easily if they are not packed carefully. The manager of the shop is a statistician and he believes that the number of damaged pots in a pack of six has a binomial distribution. He chooses a random sample of 250 packs and records the numbers of damaged pots per pack. His results are shown in the following table.

Number of damaged pots per pack (x)	0	1	2	3	4	5	6
Frequency	48	69	78	32	22	1	0

(i) Show that the mean number of damaged pots per pack in this sample is 1.656. [1]

The following table shows some of the expected frequencies, correct to 2 decimal places, using an appropriate binomial distribution.

Number of damaged pots per pack (x)	0	1	2	3	4	5	6
Expected frequency	36.01	82.36	a	39.89	b	1.74	0.11

(ii) Find the values of a and b, correct to 2 decimal places [5]

(iii) Use a goodness-of-fit test at the 1% significance level to determine whether the manager's belief is justified. [8]

35. [9231/s17/23/q10]

Roberto owns a small hotel and offers accommodation to guests. Over a period of 100 nights, the numbers of rooms, x, that are occupied each night at Roberto's hotel and the corresponding frequencies are shown in the following table.

Number of rooms occupied (x)	0	1	2	3	4	5	6	$\geqslant 7$
Number of nights	4	9	18	26	20	16	7	0

(i) Show that the mean number of rooms that are occupied each night is 3.25. [1]

The following table shows most of the corresponding expected frequencies, correct to 2 decimal places, using a Poisson distribution with mean 3.25.

Number of rooms occupied (x)	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency	4	9	18	26	20	16	7	0
Expected frequency	3.88	12.60	20.48	22.18	18.02	11.72		

(ii) Show how the expected value of 22.18, for $x = 3$, is obtained and find the expected values for $x = 6$ and for $x \geqslant 7$. [4]

(iii) Use a goodness-of-fit test at the 5% significance level to determine whether the Poisson distribution is a suitable model for the number of rooms occupied each night at Roberto's hotel. [7]

36. [9231/w17/21/q8]

Members of a Statistics club are voting to elect a new president of the club. Members must choose to vote either by post or by text or by email. The method of voting chosen by a random sample of 60 male members and 40 female members is given in the following table.

		Method of voting		
		Post	Text	Email
Gender	Male	10	12	38
	Female	5	21	14

Test, at the 1% significance level, whether there is an association between method of voting and gender. [8]

37. [9231/s16/21/q9]

Applicants for a national teacher training course are required to pass a mathematics test. Each year, the applicants are tested in groups of 6 and the number of successful applicants in each group is recorded. The overall proportion of successful applicants has remained constant over the years and is equal to 60% of the applicants. The results from 150 randomly chosen groups are shown in the following table.

Number of successful applicants	0	1	2	3	4	5	6
Number of groups	1	3	25	51	38	30	2

Test, at the 5% significance level, the goodness of fit of the distribution B(6, 0.6) for the number of successful applicants in a group. [10]

38. [9231/s16/23/q9]

A company produces four different flavours of ice cream: mint, strawberry, chocolate and fudge. Customers were asked which of the four flavours they preferred. The responses from a random sample of 80 male customers and 70 female customers are given in the following table.

	Mint	Strawberry	Chocolate	Fudge
Male	25	12	30	13
Female	16	22	25	7

Test, at the 10% significance level, whether there is a difference between ice cream flavour preferences of male and female customers. [9]

39. [9231/w16/21/q9]

The number of visitors arriving at an art exhibition is recorded for each 10-minute period of time during the ten hours that it is open on a particular day. The results are as follows.

Number of visitors in a 10-minute period	0	1	2	3	4	5	6	7	8	⩾9
Number of 10-minute periods	2	2	12	8	11	13	4	7	1	0

(i) Calculate the mean and variance for this sample and explain whether your answers support a suggestion that a Poisson distribution might be a suitable model for the number of visitors in a 10-minute period. [3]

(ii) Use an appropriate Poisson distribution to find the two expected frequencies missing from the following table. [2]

Number of visitors in a 10-minute period	0	1	2	3	4	5	6	7	8	⩾9
Expected number of 10-minute periods	1.10		8.79		11.72	9.38	6.25	3.57	1.79	1.28

(iii) Test, at the 10% significance level, the goodness of fit of this Poisson distribution to the data. [8]

40. [9231/s15/21/q11o]

Each of 200 identically biased dice is thrown repeatedly until an even number is obtained. The number of throws, $\boldsymbol{x}$, needed is recorded and the results are summarised in the following table.

$\boldsymbol{x}$	1	2	3	4	5	6	$\geqslant 7$
Frequency	126	43	22	3	5	1	0

State a type of distribution that could be used to fit the data given in the table above. [1]

Fit a distribution of this type in which the probability of throwing an even number for each die is 0.6 and carry out a goodness of fit test at the 5% significance level. [8]

For each of these dice, it is known that the probability of obtaining a 6 when it is thrown is 0.25. Ten of these dice are each thrown 5 times. Find the probability that at least one 6 is obtained on exactly 4 of the 10 dice. [5]

41. [9231/s15/23/q6]

The reliability of the broadband connection received from two suppliers, ***A*** and ***B***, is classified as good, fair or poor by a random sample of householders. The information collected is summarised in the following table.

		Reliability		
		Good	Fair	Poor
Supplier	***A***	65	63	33
	B	51	44	44

Test, at the 5% significance level, whether reliability is independent of supplier. [8]

42. [9231/w15/21/q8]

The number of goals scored by a certain football team was recorded for each of 100 matches, and the results are summarised in the following table.

Number of goals	0	1	2	3	4	5	6 or more
Frequency	12	16	31	25	13	3	0

Fit a Poisson distribution to the data, and test its goodness of fit at the 5% significance level. [10]

Chapter 4

Non-parametric tests

1. [9231/s25/41/q5]

Two jigsaw puzzles have the same number of pieces with identical shapes but have different pictures printed on them. One puzzle has a seaside picture and the other has a cartoon picture. A researcher believes that children will complete the cartoon puzzle more quickly. To test this belief, 10 children are randomly selected. The time taken in seconds for each child to complete each puzzle is recorded below.

Child	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Seaside	182	130	193	181	192	204	184	192	180	189
Cartoon	161	111	195	159	202	200	168	165	145	160

(a) Carry out a Wilcoxon matched-pairs signed-rank test at the 5% significance level to test the researcher's belief. [6]

(b) Show that using a paired-sample sign test at the 5% significance level would result in the opposite conclusion to that found in part **(a)**. [3]

It was later discovered that the experiment had been conducted such that each child completed the seaside puzzle first followed by the cartoon puzzle.

(c) Comment on the validity of using this experiment to test the researcher's belief. [1]

2. [9231/s25/43/q4]

A researcher claims that older people take longer to react to a sudden loud noise than younger people. To investigate this, the researcher randomly selects 6 people over 50 years old and 8 people under 25 years old and records their reaction times, in milliseconds, to a sudden loud noise. The reaction times are as follows.

Over 50	198	212	217	229	235	242		
Under 25	178	181	183	192	203	209	223	231

Carry out a Wilcoxon rank-sum test at the 5% significance level to test the researcher's claim. [8]

3. [9231/s25/44/q2]

The level of sound produced by a particular type of machine was measured for a random sample of 11 such machines. The results, in suitable units, are shown below.

Machine	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*
Sound level	7.66	8.48	8.21	7.98	8.01	7.77	8.25	8.11	8.03	8.16	7.92

(a) Use a Wilcoxon signed-rank test to test whether the average sound level produced by this type of machine is more than 8.00. Use a 5% significance level. [6]

(b) Give a reason why a Wilcoxon signed-rank test may be more appropriate than a *t*-test in this case. [1]

4. [9231/w25/41/q4]

A researcher believes that the median m of a population has changed from its known previous value m_0. The researcher collects a random sample of size 28. She ranks the data and calculates a test statistic T using the Wilcoxon signed-rank test. The conclusion of the test carried out at a 1% significance level is that there is not sufficient evidence to support her belief.

Using a normal approximation, find the least possible value of T. [5]

5. [9231/w25/42/q1]

A large company claims that the median salary of its employees is \$32 500. The salaries (\$) of 15 randomly selected employees are listed below.

18 750	30 500	125 000	42 500	25 000
26 000	52 500	23 000	27 500	19 500
25 500	33 000	30 000	21 500	29 000

(a) Explain why a Wilcoxon signed-rank test may not be appropriate to test the company's claim in this case. [1]

.....

.....

.....

(b) Carry out a sign test at the 10% significance level to investigate the company's claim. [5]

6. [9231/w25/44/q6]

Students of the same age from two schools, school A and school B, take a large number of quizzes throughout the year and are each awarded a mark out of 1000. The marks of 123 students in school A and 147 students in school B are ranked from lowest (rank 1) to highest (rank 270). The sum of the ranks of the students from school A is 15 355.

(a) Carry out a Wilcoxon rank-sum test at the 5% significance level to investigate whether there is a difference in average marks between the students in school A and school B. [6]

(b) State an assumption that is required for the Wilcoxon rank-sum test to be valid. [1]

7. [9231/s24/41/q2]

A large number of students are taking a Physics course. They are assessed by a practical examination and a written examination. The marks out of 100 obtained by a random sample of 15 students in each of the examinations are as follows.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*
Practical examination	66	63	24	52	59	76	88	51	48	36	91	72	68	67	60
Written examination	63	57	39	50	47	71	87	65	56	39	78	70	61	62	70

Use a sign test, at the 10% significance level, to test whether, on average, the practical examination marks are higher than the written examination marks. [5]

8. [9231/s24/41/q3]

A factory produces metal discs. The manager claims that the diameters of these discs have a median of 22.0 mm. The diameters, in mm, of a random sample of 12 discs produced by this factory are as follows.

22.4 20.9 22.8 21.5 23.2 22.9 23.9 21.7 19.8 23.6 22.6 23.0

(a) Carry out a Wilcoxon signed-rank test, at the 10% significance level, to test whether there is any evidence against the manager's claim. [7]

(b) State an assumption that is necessary for this test to be valid. [1]

9. [9231/s24/43/q1]

A college uses two assessments, X and Y, when interviewing applicants for research posts at the college. These assessments have been used for a large number of applicants this year.

The scores for a random sample of 9 applicants who took assessment X are as follows.

21.4 24.6 25.3 22.7 20.8 21.5 22.9 21.3 22.3

The scores for a random sample of 10 applicants who took assessment Y are as follows.

20.9 23.5 24.8 21.9 23.4 24.0 23.8 24.1 25.1 25.8

The interviewer believes that the population median score from assessment X is lower than the population median score from assessment Y.

Carry out a Wilcoxon rank-sum test, at the 1% significance level, to test whether the interviewer's belief is supported by the data. [7]

10. [9231/w24/41/q2]

A school with a large number of students is updating its logo. Each student has designed a new logo and two teachers have each awarded a mark out of 50 for each logo. The marks awarded to a random sample of 12 students are shown in the following table.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Teacher 1	36	38	40	36	22	34	45	44	48	35	28	30
Teacher 2	38	42	32	41	32	41	42	50	36	44	42	41

One of the students claims that Teacher 2 is awarding higher marks than Teacher 1.

(a) Carry out a Wilcoxon matched-pairs signed-rank test, at the 5% significance level, to test whether the data supports the claim. [7]

It was later discovered that Teacher 1 had entered her mark for student *C* incorrectly. Her intended mark was 24 not 40. This was corrected.

(b) Determine whether this correction affects the conclusion of the test carried out in part **(a)**. [2]

11. [9231/w24/42/q6]

A sports college keeps records of the times taken by students to run one lap of a running track. The population median time taken is 51.0 seconds. After a month of intensive training, a random sample of 22 new students run one lap of the track, giving times, in seconds, as follows.

51.3	52.0	53.4	49.2	49.3	51.1	52.2	47.2
53.0	48.5	49.4	50.3	50.8	51.6	49.1	52.3
51.8	52.4	47.9	48.9	50.6	51.9		

It is claimed that the intensive training has led to a decrease in the median time taken to run one lap of the track.

Carry out a Wilcoxon signed-rank test, at the 5% significance level, to test whether there is sufficient evidence to support the claim. [9]

12. [9231/s23/41/q4]

A random sample of 13 technology companies is chosen and the numbers of employees in 2018 and in 2022 are recorded.

Company	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*
Number in 2018	104	19	126	234	970	514	35	149	429	12	86	304	1104
Number in 2022	106	24	127	228	1012	525	32	156	449	24	78	294	1154

A researcher claims that there has been an increase in the median number of employees at technology companies between 2018 and 2022.

(a) Carry out a Wilcoxon matched-pairs signed-rank test, at the 5% significance level, to test whether the data supports this claim. [7]

The researcher notices that the figures for company *G* have been recorded incorrectly. In fact, the number of employees in 2018 was 32 and the number of employees in 2022 was 35.

(b) Explain, with numerical justification, whether or not the conclusion of the test in part **(a)** remains the same. [2]

13. [9231/s23/43/q3]

A large number of students took two test papers in mathematics. The teacher believes that the marks obtained in Paper 1 will be higher than the marks obtained in Paper 2. She chooses a random sample of 9 students and compares their marks. The marks are shown in the table.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Paper 1	46	73	55	64	86	42	66	68	60
Paper 2	41	66	61	63	90	40	58	42	70

(a) Carry out a Wilcoxon matched-pairs signed-rank test, at the 5% significance level, to test whether the data supports the teacher's belief. [7]

(b) State an assumption that you have made in carrying out the test in part **(a)**. [1]

14. [9231/w23/41/q6]

A school is conducting an experiment to see whether the distance that children can throw a ball increases in hot weather. On a cold day, all the children at the school were asked to throw a ball as far as possible. The distances thrown were measured and recorded. The median distance thrown by a random sample of 25 of the children was 22.0 m. The children were asked to throw the ball again on a hot day. The distances thrown by the same 25 children were measured and recorded and these distances, in m, are shown below.

21.2	23.5	22.9	18.6	19.4
22.1	26.5	20.2	25.7	20.6
22.3	17.4	22.2	27.0	23.9
28.2	22.6	27.2	23.0	23.7
19.8	22.7	23.3	21.5	24.3

The teacher claims that on average the distances thrown will be further when it is hot.

Carry out a Wilcoxon signed-rank test, at the 5% significance level, to test whether the data supports the teacher's claim. [10]

15. [9231/w23/42/q5]

A company is deciding which of two machines, X and Y, can make a certain type of electrical component more quickly. The times taken, in minutes, to make one component of this type are recorded for a random sample of 8 components made by machine X and a random sample of 9 components made by machine Y. These times are as follows.

Machine X	4.0	4.6	4.7	4.8	5.0	5.2	5.6	5.8	
Machine Y	4.5	4.9	5.1	5.3	5.4	5.7	5.9	6.3	6.4

The manager claims that on average the time taken by machine X to make one component is less than that taken by machine Y.

(a) Carry out a Wilcoxon rank-sum test at the 5% significance level to test whether the manager's claim is supported by the data. [6]

(b) Assuming that the times taken to produce the components by the two machines are normally distributed with equal variances, carry out a t-test at the 5% significance level to test whether the manager's claim is supported by the data. [9]

(c) In general, would you expect the conclusions from the tests in parts **(a)** and **(b)** to be the same? Give a reason for your answer. [1]

16. [9231/s22/41/q6]

A teacher at a large college gave a mathematical puzzle to all the students. The median time taken by a random sample of 24 students to complete the puzzle was 18.0 minutes. The students were then given practice in solving puzzles. Two weeks later, the students were given another mathematical puzzle of the same type as the first. The times, in minutes, taken by the random sample of 24 students to complete this puzzle are as follows.

18.2	17.5	16.4	15.1	20.5	26.5	19.2	23.2
17.9	18.8	25.8	19.9	17.7	16.2	17.3	16.6
17.1	20.1	20.3	12.6	16.0	21.4	22.7	18.4

The teacher claims that the practice has not made any difference to the average time taken to complete a puzzle of this type.

Carry out a Wilcoxon signed-rank test, at the 10% significance level, to test whether there is sufficient evidence to reject the teacher's claim. [10]

17. [9231/s22/43/q5]

A manager claims that the lengths of the rubber tubes that his company produces have a median of 5.50 cm. The lengths, in cm, of a random sample of 11 tubes produced by this company are as follows.

5.56 5.45 5.47 5.58 5.54 5.52 5.60 5.35 5.59 5.51 5.62

It is required to test at the 10% significance level the null hypothesis that the population median length is 5.50 cm against the alternative hypothesis that the population median length is not equal to 5.50 cm.

Show that both a sign test and a Wilcoxon signed-rank test give the same conclusion and state this conclusion. [9]

18. [9231/w22/41/q3]

A large college is holding a piano competition. Each student has played a particular piece of music and two judges have each awarded a mark out of 80. The marks awarded to a random sample of 14 students are shown in the following table.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*
Judge 1	79	54	63	74	69	52	50	57	55	42	63	55	56	48
Judge 2	75	62	60	73	76	41	31	51	45	55	49	50	65	36

(a) One of the students claims that on average Judge 1 is awarding higher marks than Judge 2. Carry out a Wilcoxon matched-pairs signed-rank test at the 5% significance level to test whether the data supports the student's claim. [7]

(b) Give a reason why it is preferable to use a Wilcoxon matched-pairs signed-rank test in this situation rather than a paired sample *t*-test. [1]

19. [9231/w22/42/q6]

The manager of a technology company *A* claims that his employees earn more per year than the employees at technology company *B*. The amounts earned per year, in hundreds of dollars, by a random sample of 12 employees from company *A* and an independent random sample of 12 employees from company *B* are shown below.

Company *A*	461	482	374	512	415	452	502	427	398	545	612	359
Company *B*	454	506	491	384	361	443	401	472	414	342	355	437

(a) Carry out a Wilcoxon rank-sum test at the 5% significance level to test whether the manager's claim is supported by the data. [9]

(b) Explain whether a paired sample *t*-test would be appropriate to test the manager's claim if earnings are normally distributed. [1]

20. [9231/s21/41/q5]

Georgio has designed two new uniforms X and Y for the employees of an airline company. A random sample of 11 employees are each asked to assess each of the two uniforms for practicality and appearance, and to give a total score out of 100. The scores are given in the table.

Employee	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*
Uniform X	82	74	42	59	60	73	94	98	62	36	50
Uniform Y	78	75	63	56	67	82	99	90	72	48	61

(a) Give a reason why a Wilcoxon signed-rank test may be more appropriate than a t-test for investigating whether there is any evidence of a preference for one of the uniforms. [1]

(b) Carry out a Wilcoxon matched-pairs signed-rank test at the 10% significance level. [7]

21. [9231/s21/43/q2]

A company is developing a new flavour of chocolate by varying the quantities of the ingredients. A random selection of 9 flavours of chocolate are judged by two tasters who each give marks out of 100 to each flavour of chocolate.

Chocolate	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Taster 1	72	86	75	92	98	79	87	60	62
Taster 2	84	72	74	95	85	87	82	75	68

Carry out a Wilcoxon matched-pairs signed-rank test at the 10% significance level to investigate whether, on average, there is a difference between marks awarded by the two tasters. [7]

22. [9231/w21/41/q6]

The blood cholesterol levels, measured in suitable units, of a random sample of 11 women and a random sample of 12 men are shown below.

Women	51	55	242	167	152	256	75	137	98	238	235	
Men	311	262	170	302	175	320	220	260	72	351	86	333

Carry out a Wilcoxon rank-sum test, at the 5% significance level, to test whether, on average, there is a difference in cholesterol levels between women and men. [9]

23. [9231/w21/42/q4]

Applicants for a particular college take a written test when they attend for interview. There are two different written tests, A and B, and each applicant takes one or the other. The interviewer wants to determine whether the medians of the distribution of marks obtained in the two tests are equal. The marks obtained by a random sample of 8 applicants who took test A and a random sample of 8 applicants who took test B are as follows.

Test A	46	32	29	12	33	18	25	40
Test B	36	28	49	37	48	35	41	31

(a) Carry out a Wilcoxon rank-sum test at the 5% significance level to determine whether there is a difference in the population median marks obtained in the two tests. [6]

The interviewer considers using the given information to carry out a paired sample t-test to determine whether there is a difference in the population means for the two tests.

(b) Give two reasons why it is not appropriate to use this test. [2]

24. [9231/s20/41/q2]

The times, in milliseconds, taken by a computer to perform a certain task were recorded on 10 randomly chosen occasions. The times were as follows.

6.44 6.16 5.62 5.82 6.51 6.62 6.19 6.42 6.34 6.28

It is claimed that the median time to complete the task is 6.4 milliseconds.

(a) Carry out a Wilcoxon signed-rank test at the 5% significance level to test this claim. [6]

(b) State an underlying assumption that is made when using a Wilcoxon signed-rank test. [1]

25. [9231/s20/43/q6]

A biologist is studying the effect of nutrients on the heights to which plants grow. A random sample of 24 similar young plants is divided into two equal groups A and B. The plants in group A are fed with nutrients and water and the plants in group B are given only water. After four weeks, the height, in cm, of each plant is measured and the results are as follows.

Group A	12.3	11.8	12.1	13.2	11.1	10.6	13.8	12.0	12.2	12.4	13.5	13.9
Group B	11.7	10.8	10.9	11.3	11.2	12.6	11.0	10.5	11.9	12.5	10.7	11.6

The biologist decides to carry out a test at the 5% significance level to test whether the nutrients have resulted in an increase in growth.

(a) She carries out a Wilcoxon rank-sum test. Give a reason why this is an appropriate choice of test. [1]

(b) Carry out the Wilcoxon rank-sum test for these results. [10]

26. [9231/w20/41/q2]

Metal rods produced by a certain factory are claimed to have a median breaking strength of 200 tonnes. For a random sample of 9 rods, the breaking strengths, measured in tonnes, were as follows.

210 186 188 208 184 191 215 198 196

A scientist believes that the median breaking strength of metal rods produced by this factory is less than 200 tonnes.

(a) Use a Wilcoxon signed-rank test, at the 5% significance level, to test whether there is evidence to support the scientist's belief. [6]

(b) Give a reason why a Wilcoxon signed-rank test is preferable to a sign test, when both are valid. [1]

27. [9231/w20/42/q2]

A large school is holding an essay competition and each student has submitted an essay. To ensure fairness, each essay is given a mark out of 100 by two different judges. The marks awarded to the essays submitted by a random sample of 12 students are shown in the following table.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Judge 1	62	74	52	48	68	55	56	64	37	70	81	59
Judge 2	65	70	47	49	76	74	67	54	50	77	72	75

(a) One of the students claims that Judge 2 is awarding higher marks than Judge 1.

Carry out a Wilcoxon matched-pairs signed-rank test at the 5% significance level to test whether the data supports the student's claim. [7]

It is discovered later that the marks awarded to student *A* have been entered incorrectly. In fact, Judge 1 awarded 65 marks and Judge 2 awarded 62 marks.

(b) By considering how this change affects the test statistic, explain why the conclusion of the test carried out in part **(a)** remains the same. [2]

Chapter 5

Probability generating functions

1. [9231/s25/41/q6]

Y is a discrete random variable which takes the values 0, 2, 4, … . The probability generating function of Y is given by

$$G_Y(t) = \frac{k}{1-at^2}.$$

(a) Find k in terms of a. [1]

(b) Show that $P(Y > 2) = a^2$. [3]

It is now given that $a = 0.2$.

(c) Find the value of $E(Y)$. [2]

2. [9231/s25/43/q6]

A bag contains 7 red balls and 3 blue balls. Kieran selects 2 balls at random, without replacement. The number of red balls selected by Kieran is denoted by X, and the number of different colours present in Kieran's selection is denoted by Y.

(a) Find the probability generating functions, $G_X(t)$ of X and $G_Y(t)$ of Y. [4]

The random variable Z is the sum of the number of red balls and the number of different colours present in Kieran's selection. Kieran claims that the probability generating function of Z is equal to $G_X(t) \times G_Y(t)$.

(b) Explain why Kieran is incorrect. [1]

(c) Find the probability generating function of Z, expressing your answer as a polynomial in t. [4]

(d) Use the probability generating function of Z to find $E(Z)$. [2]

3. [9231/s25/44/q5]

Eric has three identical coins, each of which is biased so that the probability of obtaining a head when it is thrown is $\frac{1}{3}$. The random variable X is the number of heads obtained when Eric throws the three coins at the same time.

(a) Find the probability generating function $G_X(t)$ of X. [2]

Eric also has two fair 6-sided dice with faces numbered 1 to 6. The random variable Y is the number of sixes obtained when Eric throws the two dice at the same time. It is given that the probability generating function of Y is $\frac{25}{36} + \frac{10}{36}t + \frac{1}{36}t^2$.

Eric throws the three coins and the two dice. The random variable Z is the sum of the number of heads obtained and the number of sixes obtained.

(b) Find the probability generating function $G_Z(t)$ of Z, expressing your answer as a polynomial in t. [3]

(c) Use $G_Z(t)$ to find $E(Z)$ and $\mathrm{Var}(Z)$. [5]

4. [9231/w25/41/q7]

A discrete random variable X takes values $r = 0, 1, 2$ with probabilities $\mathrm{P}(X = r)$ as given in the following table.

r	0	1	2
$\mathrm{P}(X = r)$	a	$2a$	b

(a) Write down the probability generating function of X, and use it to find an expression for $\mathrm{E}(X)$ in terms of a and b. [2]

.....

.....

.....

.....

.....

.....

.....

(b) Show that $\mathrm{Var}(X) = 2b + 2(a + b)(1 - 2a - 2b)$. [3

.....

The random variable Y is defined by $Y = X_1 + X_2 + X_3 + \cdots + X_{10}$ where $X_1, X_2, X_3, \ldots, X_{10}$ are ten independent observations of X.

(c) Using the probability generating function of Y, and your answer to part **(a)**, show that $\mathrm{E}(Y) = 10\mathrm{E}(X)$. [3]

(d) For the case $b = 0$, define fully the distribution of Y. [2]

5. [9231/w25/42/q6]

The discrete random variable X has probability generating function $G_X(t)$ given by

$$G_X(t) = \frac{t}{(3-2t)^2}.$$

(a) Find $E(X)$ and $\text{Var}(X)$. [5]

The discrete random variable Y has probability generating function $G_Y(t)$ given by

$$G_Y(t) = \frac{t^2}{(3-2t)^2}.$$

The random variable Z is the sum of the random variables X and Y.

(b) Assuming X and Y are independent, find $P(Z > 4)$. [5]

6. [9231/w25/44/q4]

The random variable X takes values 1 and 2 with probabilities $\frac{2}{5}$ and $\frac{3}{5}$ respectively.

(a) Write down the probability generating function $G_X(t)$ of X. [1]

The random variable Y is the sum of four independent observations of X.

(b) Find the probability generating function $G_Y(t)$ of Y. Give your answer in the form $G_Y(t) = at^m(b+ct)^n$, where a, b, c, m and n are constants to be determined. [2]

(c) Use $G_Y(t)$ to find $P(Y = 6)$. [2]

(d) Find $\text{Var}(Y)$. [5]

7. [9231/s24/41/q4]

The random variable Y is the sum of two independent observations of the random variable X. The probability generating function $G_Y(t)$ of Y is given by

$$G_Y(t) = \frac{t^2}{(4-3t)^4}.$$

(a) Find $E(Y)$. [3]

(b) Write down an expression for the probability generating function of X. [1]

(c) Find $P(X = 4)$. [3]

8. [9231/s24/43/q4]

The random variable X has probability generating function $G_X(t)$ given by

$$G_X(t) = ct(1+t)^5,$$

where c is a constant.

(a) Find the value of c. [1]

(b) Find the value of $E(X)$. [2]

The random variable Y is the sum of two independent values of X.

(c) Write down the probability generating function of Y and hence find $\text{Var}(Y)$. [4]

(d) Find $P(Y = 5)$. [2]

9. [9231/w24/41/q5]

Nikita has three coins. One coin is fair, one coin is biased so that the probability of obtaining a head is $\frac{1}{3}$ and the third coin is biased so that the probability of obtaining a head is $\frac{1}{5}$. The random variable X is the number of heads that Nikita obtains when he throws all three coins at the same time.

(a) Find the probability generating function of X. [3]

Rajesh has two fair six-sided dice with faces labelled 1, 2, 3, 4, 5, 6. The random variable Y is the number of 4s that Rajesh obtains when he throws the two dice.

The random variable Z is the sum of the number of heads obtained by Nikita and the number of 4s obtained by Rajesh.

(b) Find the probability generating function of Z, expressing your answer as a polynomial. [4]

(c) Use your answer to part **(b)** to find E(Z). [2]

10. [9231/w24/42/q2]

The random variable X has probability generating function $G_X(t)$ given by

$$G_X(t) = \tfrac{1}{5} + pt + qt^2,$$

where p and q are constants.

(a) Given that $E(X) = 1.1$, find the numerical value of $\text{Var}(X)$. [4]

The random variable Y has probability generating function $G_Y(t)$ given by

$$G_Y(t) = \tfrac{2}{3}t\left(1 + \tfrac{1}{2}t^2\right).$$

The random variable Z is the sum of independent observations of X and Y.

(b) Find the probability generating function of Z. [2]

(c) Find $P(Z > 2)$. [1]

(d) State the most probable value of Z. [1]

11. [9231/s23/41/q5]

Harry has three coins.

- One coin is biased so that, when it is thrown, the probability of obtaining a head is $\frac{1}{3}$.
- The second coin is biased so that, when it is thrown, the probability of obtaining a head is $\frac{1}{4}$.
- The third coin is biased so that, when it is thrown, the probability of obtaining a head is $\frac{1}{5}$.

The random variable X is the number of heads that Harry obtains when he throws all three coins together.

(a) Find the probability generating function of X. [3]

Isaac has two fair coins. The random variable Y is the number of heads that Isaac obtains when he throws both of his coins together. The random variable Z is the total number of heads obtained when Harry throws his three coins and Isaac throws his two coins.

(b) Find the probability generating function of Z, expressing your answer as a polynomial in t. [4]

(c) Use the probability generating function of Z to find $\mathrm{E}(Z)$. [2]

12. [9231/s23/43/q5]

The random variable X has probability generating function $G_X(t)$ given by

$$G_X(t) = k(1+3t+4t^2),$$

where k is a constant.

(a) Show that $E(X) = \frac{11}{8}$. [3]

The random variable Y has probability generating function $G_Y(t)$ given by

$$G_Y(t) = \tfrac{1}{3}t^2(1+2t).$$

The random variables X and Y are independent and $Z = X+Y$.

(b) Find the probability generating function of Z, expressing your answer as a polynomial in t. [2]

(c) Use your answer to part **(b)** to find the value of $\mathrm{Var}(Z)$. [3]

(d) Write down the most probable value of Z. [1]

13. [9231/w23/41/q5]

The random variable X has the geometric distribution Geo(p).

(a) Show that the probability generating function of X is $\frac{pt}{1-qt}$, where $q = 1-p$. [3]

(b) Use the probability generating function of X to show that $\mathrm{Var}(X) = \frac{q}{p^2}$. [5]

Kenny throws an ordinary fair 6-sided dice repeatedly. The random variable X is the number of throws that Kenny takes in order to obtain a 6. The random variable Z denotes the sum of two independent values of X.

(c) Find the probability generating function of Z. [2]

14. [9231/w23/42/q3]

Toby has a bag which contains 6 red marbles and 3 green marbles. He randomly chooses 3 marbles from the bag, without replacement. The random variable X is the number of red marbles that Toby obtains.

(a) Find the probability generating function of X. [3]

Ling also has a bag which contains 6 red marbles and 3 green marbles. He randomly chooses 2 marbles from his bag, without replacement. The random variable Y is the number of red marbles that Ling obtains. It is given that the probability generating function of Y is $\frac{1}{12}(1+6t+5t^2)$.

The random variable Z is the total number of red marbles that Toby and Ling obtain.

(b) Find the probability generating function of Z, expressing your answer as a polynomial in t. [3]

(c) Use the probability generating function of Z to find Var(Z). [4]

15. [9231/s22/41/q2]

The probability generating function, $G_Y(t)$, of the random variable Y is given by

$$G_Y(t) = 0.04 + 0.2t + 0.37t^2 + 0.3t^3 + 0.09t^4.$$

(a) Find Var(Y). [4]

The random variable Y is the sum of two independent observations of the random variable X.

(b) Find the probability generating function of X, giving your answer as a polynomial in t. [3]

16. [9231/s22/43/q3]

George throws two coins, A and B, at the same time. Coin A is biased so that the probability of obtaining a head is a. Coin B is biased so that the probability of obtaining a head is b, where $b < a$. The probability generating function of X, the number of heads obtained by George, is $G_X(t)$. The coefficients of t and t^2 in $G_X(t)$ are $\frac{5}{12}$ and $\frac{1}{12}$ respectively.

(a) Find the value of a. [2]

The random variable Y is the sum of two independent observations of X.

(b) Find the probability generating function of Y, giving your answer as a polynomial in t. [3]

(c) Find Var(Y). [3]

17. [9231/w22/41/q4]

Jason has three biased coins. For each coin the probability of obtaining a head when it is thrown is $\frac{2}{3}$. Jason throws all three coins. The number of heads obtained is denoted by X.

(a) Find the probability generating function $G_X(t)$ of X. [3]

Jason also has two unbiased coins. He throws all five coins. The number of heads obtained from the two unbiased coins is denoted by Y. It is given that $G_Y(t) = \frac{1}{4} + \frac{1}{2}t + \frac{1}{4}t^2$. The random variable Z is the total number of heads obtained when Jason throws all five coins.

(b) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(c) Find $E(Z)$. [2]

18. [9231/w22/42/q5]

A 6-sided dice, A, with faces numbered 1, 2, 3, 4, 5, 6 is biased so that the probability of throwing a 6 is $\frac{1}{4}$. The random variable X is the number of 6s obtained when dice A is thrown twice.

(a) Find the probability generating function of X. [2]

A second dice, B, with faces numbered 1, 2, 3, 4, 5, 6 is unbiased. The random variable Y is the number of 6s obtained when dice B is thrown twice.

The random variable Z is the total number of 6s obtained when both dice are thrown twice.

(b) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(c) Find $\mathrm{Var}(Z)$. [3]

(d) Use the probability generating function of Z to find the most probable value of Z. [1]

19. [9231/s21/41/q6]

Tanji has a bag containing 4 red balls and 2 blue balls. He selects 3 balls at random from the bag, without replacement. The number of red balls selected by Tanji is denoted by X.

(a) Find the probability generating function $G_X(t)$ of X. [2]

Tanji also has two coins, each biased so that the probability of obtaining a head when it is thrown is $\frac{1}{4}$. He throws the two coins at the same time. The number of heads obtained is denoted by Y.

(b) Find the probability generating function $G_Y(t)$ of Y. [2]

The random variable Z is the sum of the number of red balls selected by Tanji and the number of heads obtained.

(c) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(d) Use the probability generating function of Z to find $E(Z)$ and $\text{Var}(Z)$. [5]

20. [9231/s21/43/q4]

X is a discrete random variable which takes the values 0, 2, 4, … . The probability generating function of X is given by

$$G_X(t) = \frac{1}{3-2t^2}.$$

(a) Find $E(X)$ and $\text{Var}(X)$. [5]

(b) Find $P(X = 4)$. [3]

21. [9231/w21/41/q5]

Nine balls labelled 1, 2, 3, 4, 5, 6, 7, 8, 9 are placed in a bag. Kai selects three balls at random from the bag, without replacement. The random variable X is the number of balls selected by Kai that are labelled with a multiple of 3.

(a) Find the probability generating function $G_X(t)$ of X. [3]

The balls are replaced in the bag.

Jacob now selects two balls at random from the bag, without replacement. The random variable Y is the number of balls selected by Jacob that are labelled with an even number.

(b) Find the probability generating function $G_Y(t)$ of Y. [2]

The random variable Z is the sum of the number of balls that are labelled with a multiple of 3 selected by Kai and the number of balls that are labelled with an even number selected by Jacob.

(c) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(d) Use the probability generating function of Z to find $E(Z)$. [2]

22. [9231/w21/42/q5]

The random variable X is such that $\mathrm{P}(X = r) = kr^2$ for $r = 1, 2, 3, 4$, where k is a constant.

(a) Find the value of k. [1]

(b) Find the probability generating function $\mathrm{G}_X(t)$ of X. [2]

The random variable Y has probability generating function $\mathrm{G}_Y(t) = \frac{1}{4} + \frac{1}{2}t + \frac{1}{4}t^2$.

The random variable Z is the sum of X and Y.

(c) Assuming that X and Y are independent, find the probability generating function $\mathrm{G}_Z(t)$ of Z as a polynomial in t. [3]

(d) Given that $\mathrm{E}(Z) = \frac{13}{3}$, use $\mathrm{G}_Z(t)$ to find $\mathrm{Var}(Z)$. [3]

23. [9231/s20/41/q6]

A bag contains 4 red balls and 6 blue balls. Rassa selects two balls at random, without replacement, from the bag. The number of red balls selected by Rassa is denoted by X.

(a) Find the probability generating function, $G_X(t)$, of X. [2]

Rassa also tosses two coins. One coin is biased so that the probability of a head is $\frac{2}{3}$. The other coin is biased so that the probability of a head is p. The probability generating function of Y, the number of heads obtained by Rassa, is $G_Y(t)$. The coefficient of t in $G_Y(t)$ is $\frac{7}{12}$.

(b) Find $G_Y(t)$. [3]

The random variable Z is the sum of the number of red balls selected and the number of heads obtained by Rassa.

(c) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(d) Use the probability generating function of Z to find $E(Z)$. [2]

24. [9231/s20/43/q4]

The discrete random variable X has probability generating function $G_X(t)$ given by

$$G_X(t) = 0.2t + 0.5t^2 + 0.3t^3.$$

The random variable Y is the sum of two independent observations of X.

(a) Find the probability generating function of Y, giving your answer as an expanded polynomial in t. [3]

(b) Use the probability generating function of Y to find $E(Y)$ and $Var(Y)$. [5]

25. [9231/w20/41/q5]

Keira has two unbiased coins. She tosses both coins. The number of heads obtained by Keira is denoted by X.

(a) Find the probability generating function $G_X(t)$ of X. [1]

Hassan has three coins, two of which are biased so that the probability of obtaining a head when the coin is tossed is $\frac{1}{3}$. The corresponding probability for the third coin is $\frac{1}{4}$. The number of heads obtained by Hassan when he tosses these three coins is denoted by Y.

(b) Find the probability generating function $G_Y(t)$ of Y. [3]

The random variable Z is the total number of heads obtained by Keira and Hassan.

(c) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(d) Use the probability generating function of Z to find $E(Z)$. [2]

(e) Use the probability generating function of Z to find the most probable value of Z. [1]

26. [9231/w20/42/q5]

The random variable X has the binomial distribution B(n, p).

(a) Write down an expression for $\mathrm{P}(X = r)$ and hence show that the probability generating function of X is $(q+pt)^n$, where $q = 1-p$. [3]

(b) Use the probability generating function of X to prove that $\mathrm{E}(X) = np$ and $\mathrm{Var}(X) = np(1-p)$. [5]

Formula Sheet MF19

Cambridge Assessment
International Education

List MF19

List of formulae and statistical tables

Cambridge International AS & A Level
Mathematics (9709) and Further Mathematics (9231)

For use from 2020 in all papers for the above syllabuses.

CST319

Edited by Thoridal

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:

$$u_n = ar^{n-1}, \qquad S_n = \frac{a(1-r^n)}{1-r} \quad (r \neq 1), \qquad S_\infty = \frac{a}{1-r} \quad \left(|r| < 1\right)$$

Binomial series:

$$(a+b)^n = a^n + \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 + \binom{n}{3} a^{n-3}b^3 + \ldots + b^n$$, where n is a positive integer

and $\binom{n}{r} = \frac{n!}{r!(n-r)!}$

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \ldots$$, where n 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	$v\frac{\mathrm{d}u}{\mathrm{d}x} + u\frac{\mathrm{d}v}{\mathrm{d}x}$
$\frac{u}{v}$	$\frac{v\frac{\mathrm{d}u}{\mathrm{d}x} - u\frac{\mathrm{d}v}{\mathrm{d}x}}{v^2}$

If $x = \mathrm{f}(t)$ and $y = \mathrm{g}(t)$ then $\dfrac{\mathrm{d}y}{\mathrm{d}x} = \dfrac{\mathrm{d}y}{\mathrm{d}t} \div \dfrac{\mathrm{d}x}{\mathrm{d}t}$

Integration

(Arbitrary constants are omitted; a denotes a positive constant.)

$\mathbf{f}(x)$	$\int \mathbf{f}(x)\,\mathrm{d}x$	
x^n	$\dfrac{x^{n+1}}{n+1}$	$(n \neq -1)$
$\dfrac{1}{x}$	$\ln\lvert x\rvert$	
e^x	e^x	
$\sin x$	$-\cos x$	
$\cos x$	$\sin x$	
$\sec^2 x$	$\tan x$	
$\dfrac{1}{x^2+a^2}$	$\dfrac{1}{a}\tan^{-1}\left(\dfrac{x}{a}\right)$	
$\dfrac{1}{x^2-a^2}$	$\dfrac{1}{2a}\ln\left\lvert\dfrac{x-a}{x+a}\right\rvert$	$(x > a)$
$\dfrac{1}{a^2-x^2}$	$\dfrac{1}{2a}\ln\left\lvert\dfrac{a+x}{a-x}\right\rvert$	$\left(\lvert x\rvert < a\right)$

$$\int u\frac{\mathrm{d}v}{\mathrm{d}x}\mathrm{d}x = uv - \int v\frac{\mathrm{d}u}{\mathrm{d}x}\mathrm{d}x$$

$$\int \frac{\mathrm{f}'(x)}{\mathrm{f}(x)}\mathrm{d}x = \ln\lvert \mathrm{f}(x)\rvert$$

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

$$\mathbf{a}.\mathbf{b} = a_1b_1 + a_2b_2 + a_3b_3 = \lvert\,\mathbf{a}\,\rvert\,\lvert\,\mathbf{b}\,\rvert\cos\theta$$

FURTHER PURE MATHEMATICS

Algebra

Summations:

$$\sum_{r=1}^{n} r = \tfrac{1}{2}n(n+1), \qquad \sum_{r=1}^{n} r^2 = \tfrac{1}{6}n(n+1)(2n+1), \qquad \sum_{r=1}^{n} r^3 = \tfrac{1}{4}n^2(n+1)^2$$

Maclaurin's series:

$$\mathrm{f}(x) = \mathrm{f}(0) + x\,\mathrm{f}'(0) + \frac{x^2}{2!}\mathrm{f}''(0) + \ldots + \frac{x^r}{r!}\mathrm{f}^{(r)}(0) + \ldots$$

$$\mathrm{e}^x = \exp(x) = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^r}{r!} + \ldots \qquad \text{(all } x\text{)}$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots + (-1)^{r+1}\frac{x^r}{r} + \ldots \qquad (-1 < x \leqslant 1)$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots + (-1)^r \frac{x^{2r+1}}{(2r+1)!} + \ldots \qquad \text{(all } x\text{)}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots + (-1)^r \frac{x^{2r}}{(2r)!} + \ldots \qquad \text{(all } x\text{)}$$

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \ldots + (-1)^r \frac{x^{2r+1}}{2r+1} + \ldots \qquad (-1 \leqslant x \leqslant 1)$$

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots + \frac{x^{2r+1}}{(2r+1)!} + \ldots \qquad \text{(all } x\text{)}$$

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots + \frac{x^{2r}}{(2r)!} + \ldots \qquad \text{(all } x\text{)}$$

$$\tanh^{-1} x = x + \frac{x^3}{3} + \frac{x^5}{5} + \ldots + \frac{x^{2r+1}}{2r+1} + \ldots \qquad (-1 < x < 1)$$

Trigonometry

If $t = \tan\tfrac{1}{2}x$ then:

$$\sin x = \frac{2t}{1+t^2} \qquad \text{and} \qquad \cos x = \frac{1-t^2}{1+t^2}$$

Hyperbolic functions

$$\cosh^2 x - \sinh^2 x \equiv 1, \qquad \sinh 2x \equiv 2\sinh x \cosh x, \qquad \cosh 2x \equiv \cosh^2 x + \sinh^2 x$$

$$\sinh^{-1} x = \ln\left(x + \sqrt{x^2+1}\right)$$

$$\cosh^{-1} x = \ln\left(x + \sqrt{x^2-1}\right) \qquad (x \geqslant 1)$$

$$\tanh^{-1} x = \tfrac{1}{2}\ln\left(\frac{1+x}{1-x}\right) \qquad (|x| < 1)$$

Differentiation

$\mathbf{f}(x)$	$\mathbf{f}'(x)$
$\sin^{-1} x$	$\dfrac{1}{\sqrt{1-x^2}}$
$\cos^{-1} x$	$-\dfrac{1}{\sqrt{1-x^2}}$
$\sinh x$	$\cosh x$
$\cosh x$	$\sinh x$
$\tanh x$	$\operatorname{sech}^2 x$
$\sinh^{-1} x$	$\dfrac{1}{\sqrt{1+x^2}}$
$\cosh^{-1} x$	$\dfrac{1}{\sqrt{x^2-1}}$
$\tanh^{-1} x$	$\dfrac{1}{1-x^2}$

Integration

(Arbitrary constants are omitted; a denotes a positive constant.)

$\mathbf{f}(x)$	$\int \mathbf{f}(x)\,\mathbf{d}x$	
$\sec x$	$\ln\lvert\sec x+\tan x\rvert = \ln\lvert\tan(\frac{1}{2}x+\frac{1}{4}\pi)\rvert$	$\left(\lvert x\rvert<\frac{1}{2}\pi\right)$
$\operatorname{cosec} x$	$-\ln\lvert\operatorname{cosec} x+\cot x\rvert = \ln\lvert\tan(\frac{1}{2}x)\rvert$	$(0<x<\pi)$
$\sinh x$	$\cosh x$	
$\cosh x$	$\sinh x$	
$\operatorname{sech}^2 x$	$\tanh x$	
$\dfrac{1}{\sqrt{a^2-x^2}}$	$\sin^{-1}\left(\dfrac{x}{a}\right)$	$\left(\lvert x\rvert<a\right)$
$\dfrac{1}{\sqrt{x^2-a^2}}$	$\cosh^{-1}\left(\dfrac{x}{a}\right)$	$(x>a)$
$\dfrac{1}{\sqrt{a^2+x^2}}$	$\sinh^{-1}\left(\dfrac{x}{a}\right)$	

MECHANICS

Uniformly accelerated motion

$$v = u + at\,, \qquad s = \tfrac{1}{2}(u+v)t\,, \qquad s = ut + \tfrac{1}{2}at^2\,, \qquad v^2 = u^2 + 2as$$

FURTHER MECHANICS

Motion of a projectile

Equation of trajectory is:

$$y = x\tan\theta - \frac{gx^2}{2V^2\cos^2\theta}$$

Elastic strings and springs

$$T = \frac{\lambda x}{l}, \qquad E = \frac{\lambda x^2}{2l}$$

Motion in a circle

For uniform circular motion, the acceleration is directed towards the centre and has magnitude

$$\omega^2 r \qquad \text{or} \qquad \frac{v^2}{r}$$

Centres of mass of uniform bodies

Triangular lamina: $\frac{2}{3}$ along median from vertex

Solid hemisphere of radius r: $\frac{3}{8}r$ from centre

Hemispherical shell of radius r: $\frac{1}{2}r$ from centre

Circular arc of radius r and angle 2α: $\dfrac{r\sin\alpha}{\alpha}$ from centre

Circular sector of radius r and angle 2α: $\dfrac{2r\sin\alpha}{3\alpha}$ from centre

Solid cone or pyramid of height h: $\frac{3}{4}h$ from vertex

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:

$$s^2 = \frac{\Sigma(x_1 - \overline{x}_1)^2 + \Sigma(x_2 - \overline{x}_2)^2}{n_1 + n_2 - 2}$$

Probability generating functions

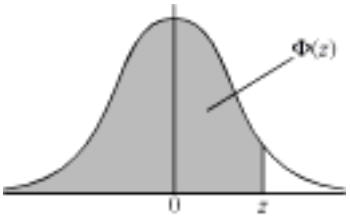
$$\mathrm{G}_X(t) = \mathrm{E}(t^X), \qquad \mathrm{E}(X) = \mathrm{G}'_X(1), \qquad \mathrm{Var}(X) = \mathrm{G}''_X(1) + \mathrm{G}'_X(1) - \{\mathrm{G}'_X(1)\}^2$$

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

$$\Phi(z) = P(Z \leqslant z).$$

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

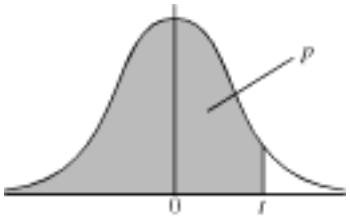
$$P(Z \leqslant z) = p.$$

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

CRITICAL VALUES FOR THE *t*-DISTRIBUTION

If T has a t-distribution with ν degrees of freedom, then, for each pair of values of p and ν, the table gives the value of t such that:

$P(T \leqslant t) = p$.



p	0.75	0.90	0.95	0.975	0.99	0.995	0.9975	0.999	0.9995
$\nu = 1$	1.000	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.894	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.689
28	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.660
30	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

CRITICAL VALUES FOR THE χ^2-DISTRIBUTION

If X has a χ^2-distribution with ν degrees of freedom then, for each pair of values of p and ν, the table gives the value of x such that

$$P(X \leqslant x) = p.$$

p	0.01	0.025	0.05	0.9	0.95	0.975	0.99	0.995	0.999
$\nu = 1$	$0.0^3 1571$	$0.0^3 9821$	$0.0^2 3932$	2.706	3.841	5.024	6.635	7.879	10.83
2	0.02010	0.05064	0.1026	4.605	5.991	7.378	9.210	10.60	13.82
3	0.1148	0.2158	0.3518	6.251	7.815	9.348	11.34	12.84	16.27
4	0.2971	0.4844	0.7107	7.779	9.488	11.14	13.28	14.86	18.47
5	0.5543	0.8312	1.145	9.236	11.07	12.83	15.09	16.75	20.51
6	0.8721	1.237	1.635	10.64	12.59	14.45	16.81	18.55	22.46
7	1.239	1.690	2.167	12.02	14.07	16.01	18.48	20.28	24.32
8	1.647	2.180	2.733	13.36	15.51	17.53	20.09	21.95	26.12
9	2.088	2.700	3.325	14.68	16.92	19.02	21.67	23.59	27.88
10	2.558	3.247	3.940	15.99	18.31	20.48	23.21	25.19	29.59
11	3.053	3.816	4.575	17.28	19.68	21.92	24.73	26.76	31.26
12	3.571	4.404	5.226	18.55	21.03	23.34	26.22	28.30	32.91
13	4.107	5.009	5.892	19.81	22.36	24.74	27.69	29.82	34.53
14	4.660	5.629	6.571	21.06	23.68	26.12	29.14	31.32	36.12
15	5.229	6.262	7.261	22.31	25.00	27.49	30.58	32.80	37.70
16	5.812	6.908	7.962	23.54	26.30	28.85	32.00	34.27	39.25
17	6.408	7.564	8.672	24.77	27.59	30.19	33.41	35.72	40.79
18	7.015	8.231	9.390	25.99	28.87	31.53	34.81	37.16	42.31
19	7.633	8.907	10.12	27.20	30.14	32.85	36.19	38.58	43.82
20	8.260	9.591	10.85	28.41	31.41	34.17	37.57	40.00	45.31
21	8.897	10.28	11.59	29.62	32.67	35.48	38.93	41.40	46.80
22	9.542	10.98	12.34	30.81	33.92	36.78	40.29	42.80	48.27
23	10.20	11.69	13.09	32.01	35.17	38.08	41.64	44.18	49.73
24	10.86	12.40	13.85	33.20	36.42	39.36	42.98	45.56	51.18
25	11.52	13.12	14.61	34.38	37.65	40.65	44.31	46.93	52.62
30	14.95	16.79	18.49	40.26	43.77	46.98	50.89	53.67	59.70
40	22.16	24.43	26.51	51.81	55.76	59.34	63.69	66.77	73.40
50	29.71	32.36	34.76	63.17	67.50	71.42	76.15	79.49	86.66
60	37.48	40.48	43.19	74.40	79.08	83.30	88.38	91.95	99.61
70	45.44	48.76	51.74	85.53	90.53	95.02	100.4	104.2	112.3
80	53.54	57.15	60.39	96.58	101.9	106.6	112.3	116.3	124.8
90	61.75	65.65	69.13	107.6	113.1	118.1	124.1	128.3	137.2
100	70.06	74.22	77.93	118.5	124.3	129.6	135.8	140.2	149.4

WILCOXON SIGNED-RANK TEST

The sample has size n.

P is the sum of the ranks corresponding to the positive differences.

Q is the sum of the ranks corresponding to the negative differences.

T is the smaller of P and Q.

For each value of n the table gives the **largest** value of T which will lead to rejection of the null hypothesis at the level of significance indicated.

Critical values of *T*

	Level of significance			
One-tailed	0.05	0.025	0.01	0.005
Two-tailed	0.1	0.05	0.02	0.01
$n = 6$	2	0		
7	3	2	0	
8	5	3	1	0
9	8	5	3	1
10	10	8	5	3
11	13	10	7	5
12	17	13	9	7
13	21	17	12	9
14	25	21	15	12
15	30	25	19	15
16	35	29	23	19
17	41	34	27	23
18	47	40	32	27
19	53	46	37	32
20	60	52	43	37

For larger values of n, each of P and Q can be approximated by the normal distribution with mean $\frac{1}{4}n(n+1)$ and variance $\frac{1}{24}n(n+1)(2n+1)$.

WILCOXON RANK-SUM TEST

The two samples have sizes m and n, where $m \leqslant n$.

R_m is the sum of the ranks of the items in the sample of size m.

W is the smaller of R_m and $m(n + m + 1) - R_m$.

For each pair of values of m and n, the table gives the **largest** value of W which will lead to rejection of the null hypothesis at the level of significance indicated.

Critical values of W

	Level of significance											
One-tailed	0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01
Two-tailed	0.1	0.05	0.02	0.1	0.05	0.02	0.1	0.05	0.02	0.1	0.05	0.02
n	$m = 3$			$m = 4$			$m = 5$			$m = 6$		
3	6	–	–									
4	6	–	–	11	10	–						
5	7	6	–	12	11	10	19	17	16			
6	8	7	–	13	12	11	20	18	17	28	26	24
7	8	7	6	14	13	11	21	20	18	29	27	25
8	9	8	6	15	14	12	23	21	19	31	29	27
9	10	8	7	16	14	13	24	22	20	33	31	28
10	10	9	7	17	15	13	26	23	21	35	32	29

	Level of significance											
One-tailed	0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01	0.05	0.025	0.01
Two-tailed	0.1	0.05	0.02	0.1	0.05	0.02	0.1	0.05	0.02	0.1	0.05	0.02
n	$m = 7$			$m = 8$			$m = 9$			$m = 10$		
7	39	36	34									
8	41	38	35	51	49	45						
9	43	40	37	54	51	47	66	62	59			
10	45	42	39	56	53	49	69	65	61	82	78	74

For larger values of m and n, the normal distribution with mean $\frac{1}{2}m(m+n+1)$ and variance $\frac{1}{12}mn(m+n+1)$ should be used as an approximation to the distribution of R_m.

BLANK PAGE

BLANK PAGE

Syllabus 26-27 Further Probability and Statistics

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 $\mathrm{Po}(\lambda)$	
• use the fact that if $X \sim \mathrm{Po}(\lambda)$ 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 – $\mathrm{E}(aX + b) = aE(X) + b$ and $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$ – $\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$ – $\mathrm{Var}(\mathrm{a}X + \mathrm{b}Y) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y)$ 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.

4 Further Probability & Statistics (for Paper 4)

Knowledge of Cambridge International AS & A Level Mathematics (9709) Papers 5 and 6: Probability & Statistics subject content is assumed for this component.

Please see the support document *Guide to prior learning for Paper 4 Further Probability & Statistics* on the Cambridge website for recommended prior knowledge for this paper.

4.1 Continuous random variables

Candidates should be able to:	Notes and examples
• use a probability density function which may be defined piecewise	
• use the general result $\mathrm{E}(\mathrm{g}(X)) = \int \mathrm{f}(x)\mathrm{g}(x)\mathrm{d}x$ where $\mathrm{f}(x)$ is the probability density function of the continuous random variable X and $\mathrm{g}(X)$ is a function of X	
• understand and use the relationship between the probability density function (PDF) and the cumulative distribution function (CDF), and use either to evaluate probabilities or percentiles	
• use cumulative distribution functions (CDFs) of related variables in simple cases.	e.g. given the CDF of a variable X, find the CDF of a related variable Y, and hence its PDF, e.g. where $Y = X^3$.

4.2 Inference using normal and *t*-distributions

Candidates should be able to:	Notes and examples
• formulate hypotheses and apply a hypothesis test concerning the population mean using a small sample drawn from a normal population of unknown variance, using a *t*-test	
• calculate a pooled estimate of a population variance from two samples	Calculations based on either raw or summarised data may be required.
• formulate hypotheses concerning the difference of population means, and apply, as appropriate – a 2-sample *t*-test – a paired sample *t*-test – a test using a normal distribution	The ability to select the test appropriate to the circumstances of a problem is expected.
• determine a confidence interval for a population mean, based on a small sample from a normal population with unknown variance, using a *t*-distribution	
• determine a confidence interval for a difference of population means, using a *t*-distribution or a normal distribution, as appropriate.	

4 Further Probability & Statistics

4.3 χ^2-tests

Candidates should be able to:	Notes and examples
• fit a theoretical distribution, as prescribed by a given hypothesis, to given data	Questions will not involve lengthy calculations.
• use a χ^2-test, with the appropriate number of degrees of freedom, to carry out the corresponding goodness of fit analysis	Classes should be combined so that each expected frequency is at least 5.
• use a χ^2-test, with the appropriate number of degrees of freedom, for independence in a contingency table.	Yates' correction is not required. Where appropriate, either rows or columns should be combined so that the expected frequency in each cell is at least 5.

4.4 Non-parametric tests

Candidates should be able to:	Notes and examples
• understand the idea of a non-parametric test and appreciate situations in which such a test might be useful	e.g. when sampling from a population which cannot be assumed to be normally distributed.
• understand the basis of the sign test, the Wilcoxon signed-rank test and the Wilcoxon rank-sum test	Including knowledge that Wilcoxon tests are valid only for symmetrical distributions.
• use a single-sample sign test and a single-sample Wilcoxon signed-rank test to test a hypothesis concerning a population median	Including the use of normal approximations where appropriate. Questions will not involve tied ranks or observations equal to the population median value being tested.
• use a paired-sample sign test, a Wilcoxon matched-pairs signed-rank test and a Wilcoxon rank-sum test, as appropriate, to test for identity of populations.	Including the use of normal approximations where appropriate. Questions will not involve tied ranks or zero-difference pairs.

4.5 Probability generating functions

Candidates should be able to:	Notes and examples
• understand the concept of a probability generating function (PGF) and construct and use the PGF for given distributions	Including the discrete uniform, binomial, geometric and Poisson distributions.
• use formulae for the mean and variance of a discrete random variable in terms of its PGF, and use these formulae to calculate the mean and variance of a given probability distribution	
• use the result that the PGF of the sum of independent random variables is the product of the PGFs of those random variables.	