Problems

1. If $\Pr(A) = 0.1$, $\Pr(B) = 0.2$, $\Pr(C) = 0.3$, find $\Pr(A \cup B \cup C)$

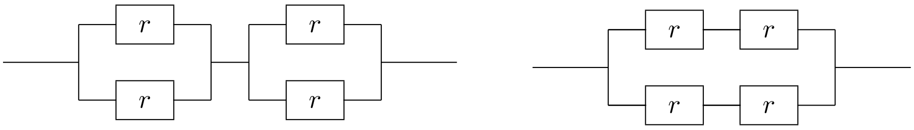
i. if A, B and C are mutually exclusive;

ii. if A, B and C are independent.

2. If $\Pr(A) = 0.43$, $\Pr(B) = 0.37$, $\Pr(A \cap B) = 0.21$, find $\Pr(A')$, $\Pr(A \cap B')$, $\Pr(A' \cap B)$, $\Pr(A' \cup B)$, $\Pr(A \mid B)$, $\Pr(A \mid B')$.

3. A particular fault in an item can only be conclusively determined by destructive testing. However, a non-destructive test is proposed which is found to be such that it gives a positive result (i.e. indicates a fault) for 99% of items which have the fault, and gives a negative result (i.e. indicates no fault) for 99% of items which do not have the fault. If this test is applied to items coming off the production line which produces 3% defective items, find the probability that an item is defective if the test indicates it is defective.

4. Calculate the reliabilities (i.e. the probabilities of operating successfully) of the two systems indicated in the diagram below, in which the components operate independently, each having reliability r (i.e. Pr(failure) = $1-r$). Which system is more reliable?



5. A supplier sends boxes of items to a factory: 90% of the boxes contain 1% defective, 9% contain 10% defective, and 1% contain 100% defective (e.g. wrong size). What percentage of items supplied are defective?

Two items are chosen from a randomly selected box. What is the probability that both are defective? Given that both are defective, what is the probability that the box is 100% defective?

6. If A and B are independent events with probabilities 0.3 and 0.4, find $\Pr(A \cap B)$ and $\Pr(A \cup B)$.

Let X denote the number of A and B that occur, so that if both A and B occur then $X = 2$ and if neither occurs then $X = 0$. Find $\Pr(X = 0)$, $\Pr(X = 1)$ and $\Pr(X = 2)$.

7. A testing device which tests items coming off a production line signals a fault with probability 0.99 when there is a fault present, but it also signals a fault with probability 0.10 when there is no fault present. If about 5% of the items are faulty, what is the probability that when the testing device indicates a fault the item is actually faulty.

8. Items are tested to failure and then examined to determine whether components p or q had failed. It is found that in 27% of a large number of failed items component p had failed, in 18% q had failed, and in 8% both p and q had failed.

Taking these values as good approximations for the probabilities, find $\Pr(P \mid Q)$ and $\Pr(Q \mid P)$ where P denotes the event that item p failed and Q the event that item q failed.

Are P and Q independent? negatively related? positively related?

9. Calls arriving at an exchange are independently switched to one of five lines with equal probability. Consider what happens to the next five calls that arrive at the exchange; find

(a) the probability that at least one is switched to line 5;

(b) the probability that at least two are switched to line 5;

(c) the probability that all are switched to line 5.

10. (a) A uniform six-sided die is thrown six times. What is the probability of obtaining at least one six?

(b) The chance that a worker in a particular factory has an accident in a given week is 1 in 100. If there are 100 workers at the factory, find the probability that an accident occurs in the week. Assume that the workers act independently.

(c) Consider n independent trials each having probability $\frac{1}{n}$ of success. Show that the probability of at least one success in n trials is approximately $1 - e^{-1}$ if n is large.

11. Two people each toss three fair coins. Find the probability that they obtain the same number of heads.

12. Let A, B and C be independent events with probabilities 0.3, 0.4 and 0.5 respectively. Find the probability that exactly one of A, B and C occurs, and hence evaluate the probability that exactly k of A, B and C occurs, for $k = 0, 1, 2, 3$.

13. *N-year design magnitudes* A system (such as a dam) is said to be designed for the N-year flood (or other extreme event) if it has a capacity which will be exceeded only by a flood greater than the N-year flood. The magnitude of the N-year flood is that which is exceeded with probability $\frac{1}{N}$ in any given year. Assume that successive annual floods are independent.

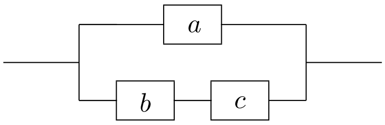
(a) What is the probability that one or more floods will exceed the 50-year flood in 50 years?

(b) What is the probability that exactly one flood in excess of the 50-year flood will occur in a 50-year period?

(c) If an agency designs each of 20 independent systems — i.e. systems at widely scattered locations — for its 500-year flood, what is the distribution of the number of systems which will fail at least once within the first 50 years after their construction. Assume that $(1 - x)^n \approx 1 - xn$, if $xn \ll 1$.

(d) In 1978, the 50-year flood was estimated to be a particular size. In the next ten years, two floods were observed in excess of that size. If the original estimate was correct, what is the probability of at least two such floods in ten years? (Such a rare event may be so unlikely that the engineer prefers to believe (i.e. act as if) the original estimate was wrong.

14. In the system indicated below, each of the components (a, b and c) operates independently with reliability 0.9.



Let A denote the event that a operates, and let W denote the event that the system operates. Find:

i. the reliability of the system, $\Pr(W)$;

ii. $\Pr(A \,|\, W)$;

iii. $\Pr(A' \,|\, W')$

15. A machine performs repeated operations independently. At each operation the machine operates successfully with probability 0.99. If it fails it is checked and serviced — but in fact this has no effect on the probability of successful operation. Company policy is that after ten failures the machine is scrapped. Let N denote the total number of successful operations performed by this machine in its lifetime. Evaluate, using a formula, the mean and standard deviation of N and hence give an approximate 95% interval within which N will lie.

16. Evaluate:

(a) $\Pr(5 \leqslant X \leqslant 10)$, where $X \overset{\mathrm{d}}{=} \mathrm{G}(0.4)$;

(b) $\Pr(5 \leqslant Y \leqslant 10)$, where $Y \overset{\mathrm{d}}{=} \mathrm{Bi}(20, 0.4)$.

17. A company obtains components from three suppliers U, V and W: 50% from U, 30% from V and 20% from W. Of the components from U, 1% are defective on average, 2% of those from V and 3% of those from W. Find the overall proportion of defective components.

A batch of ten components are tested and it is found that 2 are defective. What is the probability that this batch came from supplier W?

18. A production process when in control produces 4% defective items. If a week's production is 1000 items, specify the mean and variance of X, the number of defective items produced in a week and hence give an approximate 95% interval within which X will lie.

19. (a) Packages, each containing thirty manufactured articles, are subjected to the following sampling plan. Five articles, selected at random from the package are tested and the package is accepted if none of them is defective. Find the probability that a package which actually contains three defective items (and twenty-seven non-defective items) is accepted using this procedure.

(b) Packages, each containing a very large number of items, are subjected to the same sampling plan. Find probability that a package which actually contains 10% defective items is accepted.

20. Consider the following sampling plan. Test a random sample of $n = 50$ items chosen from a lot containing a large number of items, and accept the lot if the number of defective items found, $X \leqslant 2$.

Construct the OC-curve for this test, i.e., sketch the graph of P_A against p, where P_A denotes the probability of acceptance of the batch and p the proportion of defective items in the lot.

21. (a) A random sample of $n = 100$ items is selected from a batch of $N = 400$ which contains $R = 50$ defective items. If X denotes the number of defective items in the sample, find the mean and variance of X. Specify an approximate 95% probability interval for X.

(b) A fair coin is tossed four hundred times. Specify an approximate 95% probability interval for the number of heads obtained.

22. It is claimed that, on average, 98 percent of the components shipped out by a particular company are in good working order. If this claim is correct, find the probability that among twenty components that are shipped out there are $0, 1, 2, \ldots$ defectives.

23. A machine normally makes items of which 5 percent are defective. The machine is checked every hour, by drawing a random sample of ten items from the hour's production. If the sample contains no defectives the machine is allowed to run for another hour, otherwise it is checked. What is the probability that under this procedure the machine is left alone when it is producing 10 percent defective items?

24. A quality control engineer wants to check whether (in accordance with specifications) 95% of the items shipped by the company are in good working condition. To check this, 15 items are randomly selected from each lot of 200 ready to be shipped. The lot is passed if all the sampled items are in good working condition; otherwise each of the components in the lot is checked.

(a) What is the probability that a lot is totally inspected even though only 5% of the items are defective?

(b) What is the probability that a lot is passed without further inspection even though 10% of the items are defective?

(c) Over a large number of such lots, what is the average number of items inspected if 5% of the items are defective? if 10% of the items are defective?

25. Consider a multiple choice test of twenty questions each with five alternatives as a sampling plan. A lot is a student. The sampled items are the student's responses to the test questions. A defective item is an incorrect response. The lot is accepted (the student passes the test) if the number of incorrect responses is suitably small: $X \leqslant c$. Find c so that if the student is guessing (so that $p = 0.8$) then the probability of passing is not more than 0.05.

Interpret the results in terms of the scores on the test. What mark is required to pass? Find the probability of passing the test if $p = 0.5$.

26. Consider the following sampling plan. Test a random sample of $n = 100$ items chosen from a lot containing a large number of items, and accept the lot if the number of defective items found, $X \leqslant 3$.

The supplier of the items is concerned that if the quality level is 99% then the batch should be accepted. What can you tell the supplier about the probability of acceptance?

The consumer wants to reject batches for which the percentage of defectives exceeds 6%. What reassurances can you give the consumer?

27. A machine producing large numbers of items is checked every hour, by drawing a random sample of twenty items from the hour's production. If the sample contains at most one defective, the machine is allowed to run for another hour, otherwise it is checked and re-set. What is the probability that under this procedure the machine stopped and checked when it is producing 5 percent defective items?

28. Suppose the following double sampling plan is used: a sample of five items is selected and if no defective items are obtained the lot is accepted; if two or more defective items are obtained the lot is rejected; while if one defective item is obtained a further sample of five items is selected from the lot. The lot is then accepted if there are no defectives among the second sample; and it is rejected otherwise.

i. Assuming the lot size is large, construct the OC-curve for this plan.

ii. What advantages might a double sampling plan have over a single sampling plan?

29. A particular fault in an item can only be conclusively determined by destructive testing. However, a non-destructive test is proposed which is found to be such that it gives a positive result (i.e. indicates a fault) for 90% of items which have the fault, and gives a negative result (i.e. indicates no fault) for 95% of items which do not have the fault.

i. If this test is applied to items coming off the production line which produces 4% defective items, find the probability that an item is defective if this test indicates it is defective.

ii. Suppose a random sample of ten items actually contains two defectives. The non-destructive test is applied to the ten items. What is the probability that the test indicates at least two defectives in the sample?

30. A particular type of machine component is such that its performance level after k years of use is well described by a Markov chain on the states 1 = good, 2 = fair and 3 = unsatisfactory. The transition probability matrix is given by:

$$P = \begin{bmatrix} 0.95 & 0.05 & 0 \\ 0 & 0.9 & 0.1 \\ 0 & 0 & 1 \end{bmatrix}.$$

(a) Of a large number of such components, which are all good initially, what proportion can be expected to be unsatisfactory after two years? after four years? after ten years? (Use a computer.)

(b) If a machine component is initially classified as good, what is the distribution of N_G, the number of years for which it is classified as good? [Note that $N_G \geqslant 1$.] Specify the mean and variance of N_G.

31. The status of components off an assembly line follows a Markov chain model: if a component is defective then the probability that the next component that comes off the line is also defective is 0.5; whereas if a component is non-defective then the probability that the next component is also non-defective is 0.99. Consider this process as a Markov chain with states 1 = defective and 2 = non-defective.

i. Specify the transition probability matrix.

ii. Find the overall proportion of defective components produced.

iii. Find the mean length of a run of non-defective components.

32. Consider a communications system which transmits only the digits 0 and 1. The signal must pass through a number of stages. At each stage, 0 is correctly transmitted with probability 0.99, and 1 is correctly transmitted with probability 0.96. At each stage either a 0 or a 1 is transmitted.

Consider a system consisting of four such stages. Find the probabilities of correct transmission of the signals 0 and 1: let X_0 denote the signal fed into the system, let X_n denote the signal transmitted after the nth stage; model the system by a Markov chain and hence evaluate $\Pr(X_4 = 0 \mid X_0 = 0)$ and $\Pr(X_4 = 1 \mid X_0 = 1)$.

33. The weather in Parktown is either 1 = fair or 2 = foul. It is well described by a Markov chain with transition probabilities given by:

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}.$$

If Thursday is fine, what is the probability that (i) Saturday is foul? (ii) Sunday is foul? (iii) both Saturday and Sunday are foul?

What is the long-run proportion of fair days in Parktown?

34. An engineering student enrolled in the kth year of a four year course has probability 0.9 of passing the year, 0.05 of failing and having to repeat and 0.05 of having to quit the course. Consider this as a Markov chain with states 0 = quit, 1 = first year, 2 = second year, 3 = third year, 4 = fourth year and 5 = graduated.

Specify the transition probability matrix for this Markov chain and find the probability that the student eventually graduates.

35. Consider a communications system which transmits only the digits 0 and 1. The signal must pass through three stages. At each stage, 0 is transmitted correctly with probability 0.996, and 1 is transmitted correctly with probability 0.992.

(a) Evaluate $\Pr(X_3 = 0 \,|\, X_0 = 0)$ and $\Pr(X_3 = 1 \,|\, X_0 = 1)$.

(b) Given that the input is 80% 0s, i.e., $\Pr(X_0 = 0) = 0.8$,

i. find $\Pr(X_3 = 0)$ – i.e., find the proportion of output that is zero;

ii. find $\Pr(X_0 = 1 \,|\, X_3 = 1)$, i.e., find the probability that the signal was 1 given that the output was 1.

36. Consider a Markov chain on the states $\{1, 2, 3\}$ with transition probability matrix

$$P = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.2 \\ 0 & 0 & 1 \end{bmatrix}.$$

(a) Draw a state transition diagram for this Markov chain.

(b) Given that $X_0 = 1$, what is the probability that $X_2 = 3$? what is the probability that $X_n = 3$?

(c) In the long-run, what is the probability that the process is in states 1, 2 and 3?

37. An accident insurance company has found that about 0.1% of the population has a particular kind of accident each year. This year the company has insured $10\,000$ persons against this type of accident. What is the probability that the company will have to pay out for more than fifteen such accidents?

38. Material produced by a particular process contains flaws occurring at random at the average rate of 0.12 flaws per square metre of material. The material is sold in sheets of standard size 4 m × 1 m. Such a sheet is regarded as unsatisfactory if it contains two or more flaws. Find the expected proportion of unsatisfactory sheets.

39. In a container there are 10^{23} atoms of a particular element each having probability 10^{-23} of disintegrating in the next minute independently of all other atoms.

(a) Find the probability that at least one atom disintegrates in the next minute.

(b) Find the probability that at least two atoms disintegrate in the next minute.

40. (a) If $X \overset{\mathrm{d}}{=} \mathrm{Pn}(3)$, find $\Pr(X \geqslant 6)$.

(b) If $Y \overset{\mathrm{d}}{=} \mathrm{Pn}(30)$, specify an approximate 95% probability interval for Y.

(c) Consider a Poisson process with rate 3. Let U denote the number of "events" in the interval $(0, 1)$, and let V denote the number of "events" in the interval $(1, 2)$. Find $\Pr(U = 0)$, $\Pr(V = 3)$, $\Pr(U + V = 3)$.

41. Points are randomly distributed in a plane at the rate of 1 point per cm^2.

(a) Find the probability that there are at least four points in a region of area 4 cm^2.

(b) Find the probability that there is at least one point in each of four regions each of which has area 1 cm^2.

42. Consider writing onto a computer disk and then sending it through a certifier that counts the number of missing pulses. Suppose this number X has a Poisson distribution with parameter $\lambda = 0.2$.

(a) What is the probability that a disk has exactly one missing pulse?

(b) What is the probability that a disk has exactly two missing pulses?

(c) If two disks are independently selected, what is the probability that neither contains a missing pulse?

43. Use the Poisson approximation to construct an approximate OC-curve for sampling 50 items from a large batch and accepting the batch if at most one defective is found.

44. If $X \overset{\mathrm{d}}{=} \mathrm{Hg}(n = 100, R = 2000, N = 200000)$ find approximately $\Pr(X \geqslant 2)$.

45. The random variable Z has cdf given by

$$F_Z(z) = e^{-e^{-z}}.$$

Find c_q, the q-quantile of Z. Hence find the median and the quartiles of Z: i.e. $c_{0.25}$, $c_{0.5}$ and $c_{0.75}$. Is the distribution of Z positively skew? negatively skew? symmetrical?

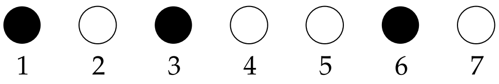
46. Suppose that X has pdf $f(x) = 20x^3(1-x)\ \ (0 < x < 1)$.

(a) Find the mode, $\textsc{m}$;

(b) find the mean μ;

(c) find the cdf, F, and by evaluating $F(\textsc{m})$ and $F(\mu)$, show that the median, m, lies between the mode and the mean: $\mu < m < \textsc{m}$.

47. Suppose that three black balls and four white balls are randomly arranged in a line. A rank is assigned to each ball by numbering the balls from left to right after arrangement:



Let X denote the sum of the ranks of the black balls, so that $X = 10$ in the arrangement shown above. Find the pmf of X, and hence find the mean of X.

48. The random variable X has pmf given by:

x	0	1	2	3	4	5	6	7	8	9	10
$p(x)$	0.021	0.085	0.163	0.206	0.195	0.149	0.094	0.051	0.024	0.009	0.003

(a) Find the mean, median and mode of X;

(b) find the quartiles of X;

(c) find the variance of X and hence evaluate $\Pr(\mu - 2\sigma < X < \mu + 2\sigma)$.

49. Find the mean and variance of the following random variables:

(a) X, where X has pmf

x	0	1	2	3
$p_X(x)$	0.3	0.4	0.2	0.1

(b) Y, where Y has pdf $f_Y(y) = 6y(1-y)$, $(0 < y < 1)$.

50. Points are randomly spread in a plane so that points occur at a mean rate of $\frac{1}{2\pi}$ per square metre. The nearest neighbour distance, T, between points is given by:

$$f(t) = te^{-\frac{1}{2}t^2} \quad (t > 0)$$

(a) Find the mode of T, i.e. the value of t for which the pdf is a maximum.

(b) Find the median of T.

(c) Find $\Pr(T > 1)$.

(d) Sketch the graph of f.

51. The random variable T has cumulative distribution function (cdf) given by

$$F(t) = 1 - (t+1)e^{-t} \quad (t > 0).$$

(a) Find $\Pr(T > 1)$.

(b) Find the median T, approximately.

(c) Find the pdf of T.

(d) Find the mode of T, i.e., the value of t for which the pdf is a maximum.

(e) Find the mean and variance of T.
Hint: You may use the fact that $\int_0^\infty e^{-z}z^n\,dz = n!$.

(f) Evaluate $\Pr(\mu - 2\sigma < T < \mu + 2\sigma)$.

52. Consider a random experiment in which two uniform six-sided dice are tossed and the results observed. Let U denote the sum of the numbers obtained, and let V denote the larger of the numbers obtained. Find the pmf of U, and the pmf of V.

53. Sketch pdfs for which the following relations hold:

(a) mode = median = mean;

(b) mode > median > mean;

(c) mean > standard deviation

54. The proportion of alcohol in a randomly selected sample of a particular substance is a random variable having pdf given by:

$$f_Z(z) = 30z^4(1-z) \quad (0 < z < 1).$$

Find the mean proportion of alcohol found in samples of this substance.

55. The random variable, X has pdf given by

$$f(x) = Kx(1-x)^2 \quad (0 < x < 1).$$

(a) Show that $K = 12$; and evaluate the mean and standard deviation of X.
Use the result: $\displaystyle\int_0^1 x^a(1-x)^b dx = \frac{a!\,b!}{(a+b+1)!}$.

(b) Verify that the cdf is given by $F(x) = 6x^2 - 8x^3 + 3x^4 \quad (0 < x < 1)$;
and hence evaluate $\Pr(\mu - 2\sigma < X < \mu + 2\sigma)$.

56. If U has pdf

$$f_U(u) = \frac{2u}{(1+u^2)^2} \quad (u > 0),$$

find c_q, the q-quantile of U. Hence find the median and the quartiles of U: $c_{0.25}$, $c_{0.5}$ and $c_{0.75}$.

57. A country petrol station is supplied once a week. Its weekly volume of sales, Z, in thousands of litres, has pdf:

$$f_Z(z) = 4(1-z)^3 \quad (0 < z < 1).$$

Show that the cdf of Z is given by

$$F_Z(z) = 1 - (1-z)^4 \quad (0 < z < 1),$$

and hence find:

i. the probability that more than 100 litres are sold in one week;

ii. the capacity of the tank so that the probability that the week's supply is exhausted is 0.01.

58. (a) If U denotes the number of tails before the 100th head in a sequence of tosses of a fair coin, find the mean and standard deviation of U and hence specify an approximate 95% probability interval for U.

(b) If V denotes the time until the 100th 'event' in a Poisson process with rate 2, find the mean and standard deviation of V and hence specify an approximate 95% probability interval for V.

59. A store has found from experience that the number of customers wishing to buy a particular type of item in a week, X, has a Poisson distribution with mean 4. At the start of the week, the store has five of these items in stock and no more are available until the next week.

(a) i. Find the pmf of the number of items sold for the week.
ii. Find the mean number of items sold.

(b) How do these answers change if there are six items in stock?

60. If $X \overset{\mathrm{d}}{=} \mathrm{R}(30, 90)$, i.e. X is a continuous random variable which is uniformly distributed on the interval $30 < x < 90$, find

i. $\Pr(X > 50)$;
ii. $\Pr(X > 50 \,|\, X > 40)$;
iii. c such that $\Pr(X > c) = 0.2$.

61. (a) If $Z \overset{\mathrm{d}}{=} \mathrm{N}(0, 1)$, find:
i. $\Pr(Z < 0.6)$;
ii. $\Pr(-1.4 < Z < 0.6)$;
iii. $\Pr(Z < 0.6 \,|\, Z > -1.4)$.

(b) If $Y \overset{\mathrm{d}}{=} \mathrm{N}(60, 225)$, find:
i. $\Pr(Y > 50)$;
ii. $\Pr(Y > 50 \,|\, Y > 40)$;
iii. c such that $\Pr(Y > c) = 0.2$.

62. Radiation counts can be modelled by a Poisson process. For a particular count the rate is supposed to be 125 per minute.

(a) Find the probability that the number of counts in five minutes is less than 600.

(b) Find the probability that the number of counts in five minutes is more than 700.

(c) What would you think if you got a count of 723 in five minutes?

63. The electrical resistance of a coil is subject to an upper specification of 25 ohms and a lower specification of 24 ohms. Examination of a large number of coils indicates that the manufacturer is producing coils such that the resistances are normally distributed with mean 24.62 ohms and standard deviation 0.22 ohms. What proportion of the coils would you expect to find outside each specification?

Four of these coils are used in series in a production component. Assuming that the coils are selected randomly, what is the probability that the sum of their resistances is more than 100 ohms?

The final unit contains two of these components and it is important that the total resistance in the separate components should agree closely. Within what limits would you expect 95% of the differences between the resistances to lie?

64. (a) If X is an integer valued random variable which is approximately normally distributed with mean 67 and standard deviation 12, find an approximate value for the probability that $X \geq 50$.

(b) Failure stresses of many standard pieces of Norwegian pine have been found to be approximately normally distributed with a mean of 25.44 N/m^2 and a standard deviation of 4.65 N/m^2. The "statistical minimum failure stress" is defined as the value such that 99% of failure stress test results may be expected to exceed it. What is its value in this case? e

65. Bolts are manufactured to a nominal diameter of 10 mm, but the process actually produces a normal distribution of diameters with mean 10 mm and standard deviation 0.11 mm. A bolt is rejected if its diameter lies outside the range 9.72 mm to 10.28 mm.

(a) Find the proportion of bolts that are rejected.

(b) Find the probability of rejecting none of a random sample of 100 bolts.

66. If $Y \overset{\mathrm{d}}{=} \mathrm{N}(\mu = 60, \sigma^2 = 100)$, find:

i. $\Pr(Y > 65)$,

ii. $\Pr(Y > 75)$,

iii. $\Pr(Y > 75 \,|\, Y > 65)$.

67. Suppose that the failure time of a particular component is well approximated by a normal distribution: $T \overset{\mathrm{d}}{=} \mathrm{N}(25, 4)$. Evaluate the hazard function of T, i.e. $h_T(t) = f_T(t)/[1 - F_T(t)]$ for $t = 20, 22, 24, 26, 28, 30, 32$.

Note: $f_T(t) = \frac{1}{2\sqrt{2\pi}}\, e^{-\frac{1}{8}(t-25)^2}$ *and* $1 - F_T(t) = \Pr(T > t)$.

Because the upper tail of the distribution of T is ________ (shorter/longer) than exponential, the hazard function is ________ (increasing/decreasing).

68. Consider a failure time, T, which has hazard function

$$h(t) = \frac{0.025}{\sqrt{t}} + 0.004\, t \quad (0 < t < 60).$$

(a) Plot the graph of h.

(b) In general, if T has hazard function h, then the cdf of T is given by :

$$F(t) = 1 - e^{-\int_0^t h(s)ds}.$$

Use this relation to evaluate and hence plot the cdf and the pdf of T.

(c) Find approximately

i. $\Pr(T > 20)$;

ii. the median of T;

iii. the mean of T.

69. After a year's use, the probability that components of a particular type are still functioning is 0.6. Of 100 such items, what is the probability that at least 50 are still functioning after a year?

70. Let Z denote the number of failures before the occurrence of the 60th success in a sequence of independent trials each having probability of success 0.6, so that $Z = T_1 + T_2 + \cdots + T_{60}$ where the T_is are independent $\mathrm{G}(0.6)$ random variables. Specify the mean and variance of Z and find approximately $\Pr(25 \leqslant Z \leqslant 50)$.

Show that $Z \overset{\mathrm{d}}{=} \mathrm{Nb}(r = 60, p = 0.6)$ and hence use MATLAB to find the probability exactly.

71. Thirty fair dice are rolled. Find the probability that the total score (i.e. the sum of the scores on the thirty dice) is 100 or more.

72. Suppose that 250 gram bars of butter are cut from larger slabs by a machine. We assume that the larger slabs are uniform in density. If the length of the bar is exactly 10 cm then the bar will weigh 250 grams. Suppose that the actual length, X cm, is such that X is equally likely to take any value in the interval 9.90 to 10.30. Assuming that the lengths of bars cut by the machine are independent, find:

i. the probability that four such bars will all weight at least 250 grams;
ii. the probability that the lightest bar weighs at least 250 grams;
iii. the probability that the heaviest bar weighs at least 250 grams;
iv. the probability that the heaviest bar weighs at least 252 grams.

73. If $X \overset{\mathrm{d}}{=} \mathrm{Hg}(n = 100, R = 20\,000, N = 200\,000)$, find approximately $\Pr(X \geqslant 15)$. How might a random variable with this probability distribution occur?

74. (a) A random digit is equally likely to take any of the values $0, 1, 2, \ldots, 9$. If R denotes a random digit, find the mean and variance of R.

(b) Suppose that 100 independent random digits $(R_1, R_2, \ldots, R_{100})$ are added, specify an interval within which the total $T = R_1 + R_2 + \cdots + R_{100}$ will lie with a probability of approximately 0.95.

75. For a particular type of observation, rounding errors can be assumed to be uniformly distributed in the interval $(-0.5, 0.5)$. Assuming independence, find the probability that the sum of 100 such observations is in error by more than 5.

76. In the AFL competition: if the Victorian teams and the interstate teams were equal, their rankings should be a random selection from $\{1, 2, \ldots, 16\}$. Under the equality hypothesis, statistical theory says that the sum of the ranks of the interstate teams is such that:

$$R_1 \overset{\mathrm{d}}{\approx} \mathrm{N}(\tfrac{1}{2}n_1(n_1{+}n_2{+}1), \tfrac{1}{12}n_1 n_2(n_1{+}n_2{+}1)),$$

where n_1 denotes the number of interstate teams (6), and n_2 denotes the number of Victorian teams (10). At the end of the 2006 season, it was observed that $r_1 = 35$.

Find $\Pr(R_1 \leqslant 35)$ under the equality hypothesis. Comment.

77. X and Y are independent random variables and $X \overset{\mathrm{d}}{=} \mathrm{N}(60, 20^2)$, $Y \overset{\mathrm{d}}{=} \mathrm{N}(70, 10^2)$.

(a) Sketch the two pdfs on the same graph.

(b) Specify the distribution of $X - Y$ and hence find $\Pr(X > Y)$.

78. Suppose that X_1, X_2 and X_3 are independent random variables such that $X_1 \overset{\mathrm{d}}{=} \mathrm{Bi}(25, 0.4)$, $X_2 \overset{\mathrm{d}}{=} \mathrm{N}(10, 9)$ and $X_3 \overset{\mathrm{d}}{=} \mathrm{R}(7, 13)$. Find the mean and variance of $U = X_1 + X_2 + X_3$, and hence find an approximate 95% probability interval for U.

79. Consider a manufacturing process producing items which include a metal frame which must be encased in a plastic casing.

(a) Each metal frame requires twenty welds. If each weld is satisfactory with probability 0.99 independently of the others, find the probability that all welds on a frame are satisfactory.

(b) Bubbles in the plastic casing occur at the rate of 1 per 100 cm^3. The casing is rejected if it contains any bubbles. If the casing contains 20 cm^3 of plastic, what proportion of casings are rejected?

(c) The casing needs to fit the component snugly. The critical inner dimension of the casing (C) has mean 15.1 mm and standard deviation 0.03 mm; the corresponding outer measurement for the frame (F) has mean 14.9 mm and standard deviation 0.04 mm. Find $\Pr(0.1 < C - F < 0.3)$, assuming the dimensions are approximately normally distributed.

80. In mass production of rectangular plates, the length is X cm and the width is Y cm, where $X \overset{\mathrm{d}}{=} \mathrm{N}(60, 0.04)$, $Y \overset{\mathrm{d}}{=} \mathrm{N}(40, 0.03)$ and X and Y are independent. The plate perimeters are to be covered with a strip of expensive material which is cut to a length Z cm where $Z \overset{\mathrm{d}}{=} \mathrm{N}(\zeta, 0.01)$ and is independent of X and Y. Find ζ so that the probability that the strip is too short for a randomly selected plate is 0.001.

81. If $X \overset{\mathrm{d}}{=} \exp(0.1)$, then the probability density function of X is given by

$$f_X(x) = 0.1e^{-0.1x} \quad (x > 0).$$

(a) Find the cumulative distribution function of X and hence obtain the median and the quartiles of X.

(b) Let $Y = \ln X$.

i. Find $\Pr(Y < 0)$.

ii. Specify the median and the quartiles of Y.

iii. Draw a rough sketch of the probability density function of Y.

82. If $Y \overset{\mathrm{d}}{=} \ell\mathrm{N}(2, 1)$ (a lognormal distribution), then $\ln Y \overset{\mathrm{d}}{=} \mathrm{N}(2, 1)$. Find:

(a) $\Pr(Y > 10)$;

(b) the median and the quartiles of Y;

(c) approximate values for the mean and variance of Y.

83. A manufacturing process requires three stages: the total time taken (in hours) for the process, $T = T_1 + T_2 + T_3$, where T_i denotes the time taken for the ith stage. It can be assumed that T_1, T_2 and T_3 are independent. It is known that:

$$\mathrm{E}(T_1) = 40, \mathrm{E}(T_2) = 30, \mathrm{E}(T_3) = 20; \quad \mathrm{sd}(T_1) = 3, \mathrm{sd}(T_2) = 2, \mathrm{sd}(T_3) = 5.$$

There is a deadline of 100 hours. Give an approximate probability that the deadline is met, i.e. find $\Pr(T \leqslant 100)$. Assume that the times are approximately normally distributed.

Which stage is most influential in determining whether the deadline is met?

84. (a) If $K = 40V^{-0.4}$, and V is subject to error such that $\mathrm{E}(V) = 10.52$ and $\mathrm{sd}(V) = 1.93$, find approximately the mean and standard deviation of K.

(b) If X has mean 500 and standard deviation 50, find approximate values for the mean and standard deviation of $Y = \ln X$.

(c) Suppose that U has mean 10 and standard deviation 2; and $Z = \sqrt{10/U}$. Find approximations for the mean and standard deviation of Z.

85. The variable X has mean 1000 and standard deviation 40. It is multiplied by a "doubling factor", D, which is supposed to be 2, but in fact $D \overset{\mathrm{d}}{=} \mathrm{R}(1.8, 2.2)$. Assuming that X and D are independent, find the mean and standard deviation of $Y = XD$.
Hint: use $\mathrm{var}(Y) = \mathrm{E}(Y^2) - \mathrm{E}(Y)^2$.

86. (a) Generate fifty numbers which are uniformly distributed on $(0, 100)$, i.e. fifty observations on $X \overset{\mathrm{d}}{=} \mathrm{R}(0, 100)$.

This can be done in `EXCEL` *using* `= 100*RAND()` *in cells* `A1:A50`; *in* MATLAB *using* `>> x=unifrnd(0,100,1,50)`; *in* MINITAB *using* `random 50 x; uniform 0 100`; *or by using tables of random numbers.*

(b) Calculate the mean, standard deviation and quartiles for these data.

In `EXCEL` *these can be obtained using* `= AVERAGE(A1:A50)`, `= STDEV(A1:A50)`, `= QUARTILE(A1:A50,1)`, `= MEDIAN(A1:A50)` *and* `= QUARTILE(A1:A50,3)`.
In MATLAB, *use* `>>mean(x)` `>>std(x)` `>>prctile (x,25)` `>>median(x)` *and* `>>prctile(x,75)`. *In* MINITAB, *they are generated by* `describe x`.

What values should these statistics be close to?

(c) Construct a boxplot for these data.

87. (a) Calculate $Y = -50\ln(X/100)$ for the fifty observations obtained in the previous question.

This can be done in EXCEL using `= -50*LN(A1/100)` in cell `B1` and filling down; in MATLAB: `>> y = -50*log(x/100);` in MINITAB: `let y = -50*log(x/100);` or just take logs for each of the fifty observations.

(b) Calculate the mean, standard deviation and quartiles for these data.

(c) Draw a dotplot or a histogram for these data.

(d) Draw a boxplot for these data.

(e) Describe (in a sentence) the distribution of Y. Which of the distributions you have met might fit it?

88. For the following sample:

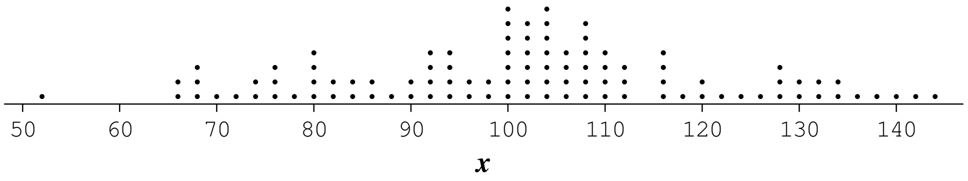
47.2	29.7	33.0	41.2	49.5	50.1	56.9	40.8	51.8	44.4
49.7	48.6	59.4	46.3	35.1	41.4	25.9	40.0	47.4	

(a) find the sample median and the sample interquartile range;

(b) draw the boxplot for the sample;

(c) calculate the sample mean and the sample standard deviation.

89. A random sample of $n = 100$ observations on $X \overset{\mathrm{d}}{=} \mathrm{N}(100, 20^2)$ is generated, and a dotplot produced, along with some standard descriptive statistics.



```
Variable     N     Mean  StDev  Minimum      Q1  Median       Q3  Maximum
x          100   101.58  19.60    52.29   87.86  102.45   112.59   143.10
```

Specify values for the sample mean, sample median, sample standard deviation and sample interquartile range for these data. What values should these be close to?

90. The following data sets were obtained supposedly as random samples from a normally distributed population:

```
sample 1:  34.7   57.6   49.3   36.6   54.5   96.1   56.7   50.9   34.5   52.1
           73.4   54.4   48.2   45.4   42.1   52.2   41.1   45.1   49.8   44.8

sample 2:  67.0   54.6   47.8   45.9   52.1   42.5   44.1   42.1   44.6   49.9
           34.7   63.6   49.3   36.6   54.5   69.1   61.7   50.9   34.5   52.1
```

Construct a normal QQ-plot for these two data sets, and decide that one is probably not from a normal population, giving an explanation.

For the other (acceptably normal) sample, use the QQ-plot to estimate the population mean and standard deviation.

91. (a) Generate a sample of $n = 10$ observations on $Y \overset{\mathrm{d}}{=} \mathrm{N}(\mu = 50, \sigma^2 = 10^2)$.

(b) Construct a QQ-plot for your data, i.e. plot the points $\{(\Phi^{-1}(\frac{k}{n+1}),\ y_{(k)}),\ k{=}1, 2, \ldots, 10\}$, where Φ denotes the standard normal cdf.

(c) Fit a straight line to the QQ-plot, and hence estimate μ and σ.

(d) Use MATLAB to generate a Normal probability plot for your data.

(e) Now suppose that all observations greater than 60 are censored.
For example: in an accelerated failure test, items are stressed for an hour. If the item fails in that time, the failure time (Y) is observed; otherwise we only observe $Y > 60$.
Thus an observation like 61.3 would be replaced by 60+.
Construct a QQ-plot for these censored data and use this to estimate μ and σ.

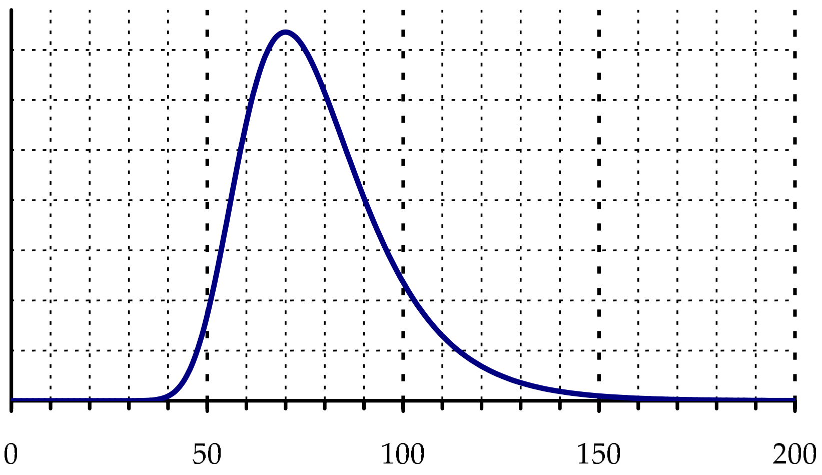
92. The following data were obtained as the number of items in batches of ten which had a particular characteristic.

x	0	1	2	3	4	5	6	7	8	9	10
freq(x)	25	30	31	23	14	10	8	6	3	0	0

(a) Verify that $\bar{x} = 2.540$ and $s = 2.075$.

(b) A binomial distribution would be appropriate for such data if the items were independent and each was equally likely to have the characteristic. Explain why these data are apparently incompatible with the assumption of sampling from a binomial distribution.

93. The diagram below is a sketch of the graph of the probability density function of X:



(a) Give approximate values for $\mathrm{E}(X)$ and $\mathrm{sd}(X)$, explaining the reasons for your choices.

(b) Draw a dotplot for a random sample of $n = 10$ observations on X.

(c) If a random sample of $n = 100$ observations is obtained on X, specify an approximate 95% probability interval for the sample mean.

(d) Draw a boxplot for a random sample of $n = 1000$ observations on X.

94. For the following sample:

x	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
freq(x)	1	3	1	3	4	7	7	7	6	4	2	1	1	2	1

(a) draw a graph of the sample pmf;

(b) calculate the sample mean and the sample standard deviation.

(c) If the data are from a Poisson distribution, what is your best guess at the value of λ? If that were the correct value of λ, what would the sample standard deviation be close to?

(d) Is it plausible that these data were obtained by random sampling from a Poisson distribution? Explain.

95. The numbers given below represent a random sample on an exponentially distributed random variable with mean λ (rounded to the nearest integer):

24	1	6	18	8	44	1	17	6	26	18	8
22	26	16	6	35	1	11	8	4	80	15	33
13	4	31	2	0	5	22	24	39	23	24	1
3	3	59	15	7	59	14	4	46	32	7	61
4	32	15	4	11	1	12	0	11	0	4	11
17	27	29	11	48	74	0	35	40	4	26	20
7	10	1	22	15	2	13	4	8	4	1	24

(a) Find the sample median.

(b) The sample mean is an estimate of λ. Calculate it.

(c) Show that for an exponential distribution, the population median, $m = \lambda \ln 2$. By assuming that the sample median relates to λ in a similar way, obtain an alternative estimate of λ based on the sample median.

96. The frequency distribution below is a summary of the results of measuring the diameters of 200 rivets (in mm)

range	freq	range	freq	range	freq
13.10 – 13.14	2	13.30 – 13.34	27	13.50 – 13.54	25
13.15 – 13.19	1	13.35 – 13.39	30	13.55 – 13.59	17
13.20 – 13.24	8	13.40 – 13.44	37	13.60 – 13.64	7
13.25 – 13.29	17	13.45 – 13.49	27	13.65 – 13.69	2

(a) Draw a histogram for these data.

(b) Draw the graph of the sample cdf and hence obtain approximately the sample median and the sample quartiles.

(c) Draw a boxplot for these data.

(d) Calculate estimates of the mean and the standard deviation of the population of rivet diameters.

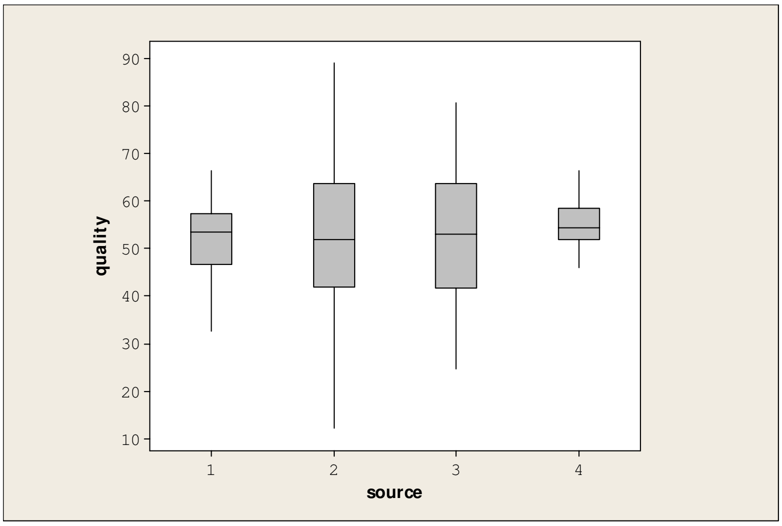
97. (a) Using MATLAB, or otherwise, generate 100 observations from an Extreme Value distribution: $X \overset{\text{d}}{=} \text{EV}(\theta = 100, \phi = 10)$, for which $\mu = 105.77$ and $\sigma = 12.83$.
In EXCEL, put `= 100-10*LN(-LN(RAND()))` in cell `A1` and then fill down;
in MATLAB: `>>r=rand(1,100); >>x=100-10*log(-log(r));` or use the Extreme Value generator `>>x=-evrnd(-100,10,1,100)` *(but be careful with the negative signs)*.

(b) Calculate the mean and standard deviation for these data.
In EXCEL: `= AVERAGE(A1:A100)` and `= STDEV(A1:A100);`
in MATLAB: `>> mean(x); >> std(x);`
What values should these statistics be close to?

(c) Plot a graph of the sample cdf $\tilde{F}$, i.e. plot the points $(x_{(k)}, \frac{k}{n+1})$.
In EXCEL, order the data in `A1:A100` (from smallest to largest) and putting $\frac{1}{101}$, $\frac{2}{101}, \ldots, \frac{100}{101}$ in `B1:B100`; and plot with `A1:A100` on the horizontal axis.
In MATLAB: `>> k=1:100; >> fk=k/101; >> xo=sort(x); >> plot(xo,fk);`
Indicate the sample median and quartiles on your graph.

(d) A check of whether the distribution is actually an Extreme Value distribution is provided by a QQ-plot: plotting $y = x_{(k)}$ vs $x = -\ln(-\ln(\frac{k}{n+1}))$. Generate such a plot.
In EXCEL, enter `= -LN(-LN(B1))` in `C1`, and `= A1` in `D1` and fill down; then plot with `C1:C100` on the horizontal-axis.
In MATLAB, use `>>eq=-log(-log(fk)); >>qqplot(eq,xo)`
If $X \overset{\text{d}}{=} \text{EV}(\theta, \phi)$ then this plot should be close to a straight line with intercept θ and slope ϕ. So this provides a check on the validity of the Extreme Value model.
The intercept and the slope of the fitted line can be used to obtain estimates of θ and ϕ. Obtain the intercept and slope of the fitted straight line.
In EXCEL, use `=INTERCEPT(D1:D100, C1:C100)` and `=SLOPE(D1:D100, C1:C100)`.
In MATLAB, use Tools>Basic Fitting in the Figure plot.
What values should these statistics (intercept and slope) be close to?

98. The quality of output from four sources is indicated by the box-plots below, in which the better the quality product the greater the value on the quality axis. Each box-plot is based on a sample of twenty items from each source.



Which source is preferred? Explain your choice.

99. A population has mean $\mu = 50$ and standard deviation $\sigma = 10$.

(a) If a random sample of 10 is obtained, find approximately, $\Pr(49 < \bar{X} < 51)$.

(b) If a random sample of 100 is obtained, find approximately, $\Pr(49 < \bar{X} < 51)$.

100. (a) The following is a random sample from a population which can be assumed to have a Poisson distribution:

$$12, 20, 8, 15, 12, 10, 16, 15, 23, 14$$

Give an estimate of λ. Specify the standard error of your estimate. What (roughly) are the likely values of λ?

(b) The sampling is continued until $n = 100$ observations have been obtained from this population, and for this sample $\bar{x} = 15.0$. Give an estimate of λ. Specify the standard error of your estimate. What (roughly) are the likely values of λ?

101. A random sample of 25 components is taken from a large lot of components, and these components are tested to failure. The sample mean of the component failure times is found to be 1410 hours. Assume that the failure time distribution is normal with standard deviation 200 hours. Find a 95% confidence interval for the mean failure time.

102. Of a random sample of sixty items from a large population, twenty-four had a particular characteristic.

(a) Use the chart in the Statistical Tables (Table 2) to find a 95% confidence interval for the population proportion.

(b) Use the formula $\hat{p} \pm 2\text{se}(\hat{p})$ to find an approximate 95% confidence interval for the population proportion.

103. Of a random sample of $n = 40$ items, it is found that $x = 8$ had a particular characteristic. Use the chart in the Statistical Tables (Table 2) to find a 95% confidence interval for the population proportion.

Repeat the process to complete the following table:

n	x	$\hat{p}$	95% CI: (a, b)
60	12		
100	20		
200	40		
40	32		
60	48		
100	80		
200	160		

104. A random sample of fifty observations is obtained on $X \overset{\mathrm{d}}{=} \mathrm{Pn}(\lambda = 50)$. Specify the mean and variance of the sample mean, $\bar{X}$. What are the likely values of the sample mean?

105. (a) A population of items contains 8% defective items. A random sample of 400 items is selected from this population, and X denotes the number of defective items in the sample.

i. Find an approximate 95% range for X.

ii. Find an approximate 95% range for the sample proportion defective.

(b) A random sample of 400 items is checked and 36 defectives are found. Find approximate 95% confidence interval for the population proportion defective.

106. A random sample of $n = 400$ observations on Y yields 68 zeros, so that $\hat{p}(0) = 0.17$. Find a 95% confidence interval for $p(0) = \Pr(Y = 0)$.

107. Consider the discrete random variable, X, with pmf such that $\Pr(X = 0) = 0.5$, $\Pr(X = 1) = 0.3$ and $\Pr(X = 2) = 0.2$. A sample of 10 independent observations on X were simulated, with results given below.

```
0    1    0    0    0    0    0    2    1    0
```

Calculate the sample mean, $\bar{x}$, for the ten observations. How does it compare to μ?

An array of independent observations on X is put into a 100×10 array; by generating another 99 rows generated like the one above:

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1,10} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2,10} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{100,1} & x_{100,2} & x_{100,3} & \cdots & x_{100,10} \end{bmatrix} \begin{matrix} \longrightarrow \\ \longrightarrow \\ \\ \longrightarrow \end{matrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_{100} \end{bmatrix}$$

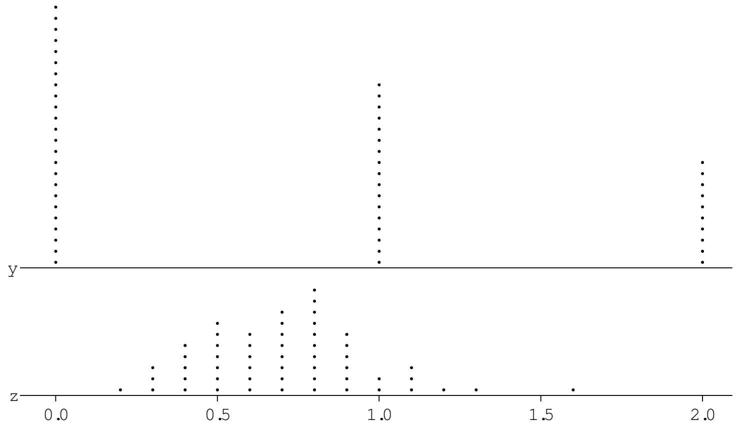
The average of each row is calculated as indicated so that $\bar{x}_i$ denotes the average of a sample of ten independent observations on X.

Let y denote the first column of the data matrix above; and let z denote the column of averages. Descriptive statistics for these data sets (i.e. y and z) are given below.

```
        N     Mean    StDev  Minimum       Q1   Median       Q3  Maximum
y     100   0.7200   0.7665   0.0000   0.0000   1.0000   1.0000   2.0000
z     100   0.7080   0.2477   0.2000   0.5000   0.7000   0.8750   1.6000
```

What are the theoretical value of $\mathrm{E}(X)$? $\mathrm{sd}(X)$? $\mathrm{E}(\bar{X})$? $\mathrm{sd}(\bar{X})$? Are these values reflected by these data?

Comment on the dotplots corresponding to y and z given below.



108. A random sample of 400 observations is obtained on $X \overset{\mathrm{d}}{=} \exp(\alpha = 0.1)$. Specify the mean and variance of the sample mean, $\bar{X}$. What are the likely values of the sample mean?

109. A random sample of $n = 25$ observations is obtained on $X \overset{\mathrm{d}}{=} \mathrm{N}(\mu{=}50, \sigma^2{=}100)$.

(a) Find a 0.95 probability interval for $\bar{X}$.

(b) Find a 0.95 probability interval for S.

110. The following sample is obtained:

$$21, 28, 19, 29, 20, 14, 21, 27, 31, 28, 37, 28, 42, 14, 35, 38, 19, 37$$

(a) Give an estimate of the population mean.

(b) Specify the standard error of your estimate.

(c) Hence obtain an approximate 95% confidence interval for the population mean.

111. The following sample is thought to be from a population which has a Poisson distribution:

```
14    21    15    25    17    26    22    16    19    19
21    18    31    14    21    24    16    15    20    26
22    22    21    14    16    21    21    19
```

```
       N     Mean  StDev  Minimum       Q1  Median       Q3  Maximum
x     28   19.857  4.187   14.000   16.000  20.500   22.000   31.000
```

(a) How can you check that a Poisson distribution is a reasonable assumption?

(b) Assuming the population distribution is Poisson, find a 95% confidence interval for the population mean.

(c) What is your opinion of the proposition that $\mu = 18$?

112. A random sample of $n = 100$ observations is obtained on $X \overset{\mathrm{d}}{=} \mathrm{N}(\mu, 25)$. The sample gives $\bar{x} = 24.67$.

(a) Find a 95% confidence interval for μ.

(b) Find a 95% prediction interval for X.

113. The following is a random sample from a population with mean μ and standard deviation σ.

```
48.9   40.4   43.8   43.8   19.3   40.8   32.6   57.2   42.6
39.9   50.7   25.9   45.0   40.5   35.3   55.2   36.6   30.4
52.3   27.1
```

(a) Is it plausible that this could have come from a normally distributed population?

(b) Give estimates of μ and σ.

(c) Give a 95% confidence interval for μ.

114. From the daily inspections of items from the production line over the past month, of 420 items inspected, 28 were found to be defective. Find an approximate 95% upper confidence limit for the proportion of defectives in the sampled population of items. Is it plausible that the population proportion is 10%? Explain.

115. A random sample of $n = 25$ observations is obtained on $X \overset{\mathrm{d}}{=} \mathrm{N}(\mu, \sigma^2)$. The sample gives: $\bar{x} = 35.71$ and $s = 5.69$.

(a) Find a 95% confidence interval for μ.

(b) Find a 95% confidence interval for σ.

(c) Find a 95% prediction interval for X.

116. The thickness (in mm) of each of a random sample of sixty tiles was measured. For this sample, the sample mean is 5.72 and the sample standard deviation 0.16.

(a) Find a 95% confidence interval for the mean tile thickness, and hence test the hypothesis $\mu = 5.75$.

(b) Find a 95% confidence interval for the standard deviation of the tile thickness, and hence test the hypothesis $\sigma = 0.10$.

(c) Find a 95% prediction interval for the tile thicknesses.

117. A random sample of 30 observations on $X \overset{\mathrm{d}}{=} \mathrm{Pn}(\lambda)$ yields $\bar{x} = 72$. Find an approximate 95% confidence interval for λ.

A further sample of 20 observations yields $\bar{x} = 74$. Find an approximate 95% confidence interval for λ based on all 50 observations.

118. A standard determination of the sulphur content in oil expressed as a percentage of the weight, gave the following results:
1.12, 1.11, 1.08, 1.09, 1.06, 1.08, 1.13, 1.14, 1.11, 1.14, 1.12, 1.15, 1.11, 1.12, 1.09, 1.08, 1.11.
Assuming normality of the observations, find a 95% confidence interval for the true sulphur content.

119. In the production of polyol, it is reacted with isocyanate in a foam moulding process. Variations in the moisture content of polyol causes problems in controlling the reaction with isocyanate. Production has set a target moisture content of 2.125%. The following data represent 18 moisture measurements over a nine-week period:

2.29	2.22	1.94	1.90	2.15	2.02	2.15	2.09	2.18
2.00	2.06	2.02	2.15	2.17	2.17	1.90	1.72	1.98

Construct a 95% confidence interval for the mean moisture content. What conclusions do you draw about the foam moulding process?

120. Consider the following random sample on X:

```
0   1   3   0   1   0   2   0   0   1
0   0   2   2   1   1   0   1   0   1
1   1   4   0   0   0   0   1   0   3

x       0   1   2   3   4
freq   14  10   3   2   1
```

(a) Making no assumptions about the distribution of X, find a 95% confidence interval for $\Pr(X = 0)$.

(b) If it is assumed that $X \overset{\mathrm{d}}{=} \mathrm{Pn}(\lambda)$, use the following information to find a 95% confidence interval for λ and hence for $e^{-\lambda} = \Pr(X = 0)$.

```
 N      Mean     Median      StDev
30     0.867      1.000      1.074
```

121. The likelihood function for data set D is given by

$$L(\theta) = \mathrm{K}\,\theta^8 e^{-7\theta} e^{-\frac{1}{2}\theta^2} \quad (\theta > 0).$$

i. Find the maximum likelihood estimate $\hat{\theta}$ and its standard error.

ii. Specify an approximate 95% confidence interval for θ.

iii. Sketch, roughly, the graph of the relative log-likelihood function.

122. The number of successful repetitions of a severe strain test completed by a particular type of item has probability distribution given by:

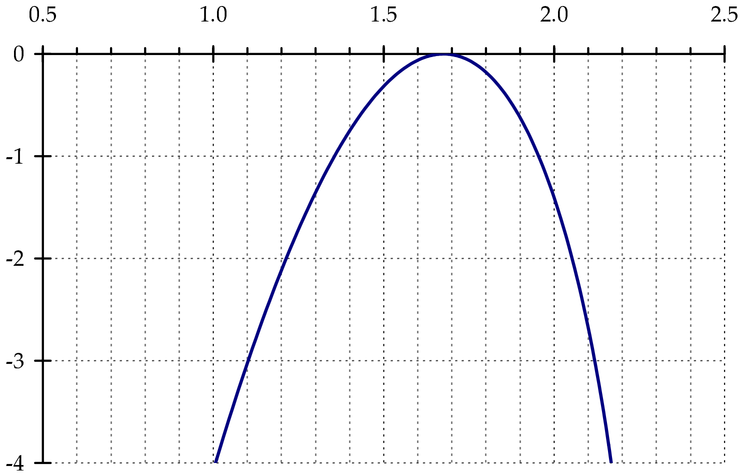
	$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$	$x > 4$
probability	$1 - \theta$	$\theta(1 - \theta)$	$\theta^2(1 - \theta)$	$\theta^3(1 - \theta)$	$\theta^4(1 - \theta)$	θ^5

Observations on 100 such items yielded the following data:

	$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$	$x > 4$
frequency	19	25	21	15	8	12

Find the maximum likelihood estimate of θ and its standard error.

123. For a particular sample on Y, the distribution of which is dependent on the parameter θ, the likelihood function has been computed, and a graph of the log likelihood is shown below.



Specify the maximum likelihood estimate $\hat{\theta}$ and an approximate 95% confidence interval for θ.

124. The likelihood function for an observed set of data is given by:

$$L(\theta) = e^{-10\theta}\theta^6(1-\theta)^3 \quad (0 < \theta < 1).$$

i. Find the maximum likelihood estimate of θ.
ii. Obtain an approximate value for the standard error of your estimate.

125. The Rayleigh distribution, $\mathrm{Ra}(\theta)$ has cdf and pdf given by

$$F(x;\theta) = 1 - \exp\big(-\frac{x^2}{2\theta^2}\big) \ (x > 0); \quad f(x;\theta) = \frac{x}{\theta^2}\exp\big(-\frac{x^2}{2\theta^2}\big) \ (x > 0).$$

The following data are observations of a random variable X for a sample of five individuals, where it is assumed that $X \overset{\mathrm{d}}{=} \mathrm{Ra}(\theta)$.

$$2, 3^*, 4, 5^*, 6; \qquad (\sum x = 20, \sum x^2 = 90).$$

In this sample the observations marked with an asterisk have been censored: that is, the observation 5^* indicates that it is only known that x is greater than 5 for that individual.

Specify the likelihood function for this data set.
Hence find the maximum likelihood estimate of θ and its standard error.

126* A random sample is obtained on $X \overset{\mathrm{d}}{=} \mathrm{G}(\theta)$, so that:

$$\Pr(X = k) = \theta(1-\theta)^k \quad (k = 0, 1, 2, \ldots).$$

(a) Let $U = \hat{p}(0)$. Find the mean and variance of U.

(b) Let $V = 1/(1+\bar{X})$. Find approximate expressions for the mean and variance of V. *Hint: Assume that* $\mathrm{E}(X) = \mu = \frac{1-\theta}{\theta}$ *and* $\mathrm{var}(\bar{X}) = \frac{\sigma^2}{n} = \frac{1-\theta}{n\theta^2}$ *and use the approximation formulae.*

(c) Which is the better estimator of θ?

(d) Use the following sample on X to compute estimates of θ based on U and V and also their standard errors:

$$4, 2, 1, 0, 0, 1, 0, 7, 2, 1, 0, 1, 5, 3, 2, 1, 0, 1, 4, 1, 2, 0, 1, 2, 1, 0, 0, 8, 3.$$

127. KILLEM flyspray in a given concentration is believed to kill on the average 60% of flies. To test this hypothesis, twenty flies are exposed to the flyspray in the given concentration, and if less than 8 are killed it will be assumed that the kill rate is less than 0.60.

i. State H_0 and H_1.
ii. What is the size of the test?
iii. What is the power of the test for a kill rate of 0.30?

128. The pH of water coming out of a filtration plant is supposed to be 7.0. Twelve water samples independently selected from this plant give the following results:

6.8 6.7 7.1 6.9 6.8 7.0 6.7 6.8 6.6 6.9 6.6 6.9

Is there reason to doubt that the plant's specification is being maintained? Give a detailed statistical argument.

129. A random sample of 10 observations is taken on $X \overset{\mathrm{d}}{=} \mathrm{N}(\mu, 2^2)$ to test the null hypothesis H_0: $\mu = 10$ against the alternative H_1: $\mu \neq 10$, using a significance level of 0.05. It is observed that $\bar{x} = 11.31$

(a) Specify the critical region, and use it to determine whether H_0 should be rejected.

(b) Find a 95% confidence interval for μ, and use it to determine whether H_0 should be rejected.

(c) Find the P-value for the test, and use it to determine whether H_0 should be rejected.

Must all these methods produce the same decision?

130. A critical quality measure in the production of tiles is variability of thickness. A process in current use produces tiles with σ = 0.2 mm. Tests on a proposed new process gave s = 0.16 mm from 25 observations.

Assuming a normal distribution for tile thickness, find (approximately) the P-value for a test of whether the new process has smaller variability. Is it reasonable to conclude that the new process has smaller variability? Give reasons for your answer.

131. Tests on paper towels produced the following results for the dry breaking strength of SupaStrong paper towels (in kgf):

9.1	9.5	9.2	8.4	9.7	9.6	9.9	9.8	8.7	9.2
9.3	10.0	9.7	9.3	9.0	9.6	8.5	8.6	9.3	9.5

You are required to report on the mean and the standard deviation of the breaking strength of SupaStrong paper towels. In particular, management is concerned that the mean should be no less than 9.3 and that the standard deviation should be no greater than 0.4. Is there any significant evidence in the data to suggest that these conditions are not met?

Report your findings.

132. A random sample of 10 observations is to be taken on $X \overset{\mathrm{d}}{=} \mathrm{N}(\mu, 2^2)$ to test H_0: $\mu = 10$ versus H_1: $\mu = 12$. Find the critical region for a test of size 0.05, and find the power of the test.

133. A random sample of 9 observations on $X \overset{\mathrm{d}}{=} \mathrm{N}(\mu, 5^2)$ gave $\bar{x} = 22.6$.

(a) Find the critical region of size 0.05 for testing H_0: $\mu = 20$ versus H_1: $\mu > 20$. Is H_0 accepted or rejected?

(b) Find the power of the test when $\mu = 23$.

134. The Rockwell hardness index for steel is determined by pressing a diamond point into the steel and measuring the depth of penetration. For 20 specimens of a particular type of steel, the Rockwell hardness index averaged 61.2 with a standard deviation of 4.5. The manufacturer claims that this steel has an average hardness index of at least 64. Do the data presented here contain significant evidence against the manufacturer's claim?

135. It is specified that the stress resistance of a particular plastic should have mean 30 psi. The results from 12 specimens of this plastic show a mean of 27.4 psi and a standard deviation of 1.1 psi. Is there sufficient evidence to question the specification? What assumptions are you making?

136. A random observation sample of 100 observations on $X \overset{\mathrm{d}}{=} \mathrm{N}(\mu, \sigma^2)$ gives $\bar{x} = -0.25$ and $s^2 = 1.47$.

(a) Test the hypothesis $\mu = 0$.

(b) Test the hypothesis $\sigma^2 = 1$.

Specify the P-value in each case.

137. Tests on plastic strips produced the following results for the strength of the strips:

90	94	92	85	97	96	99	98	87	92
93	96	97	93	89	96	85	86	93	92

You are required to report on the mean and the standard deviation of the strength of the plastic strips. In particular, management is concerned that the mean should be no less than 90 and that the standard deviation should be no greater than 5. Is there any significant evidence in the data to suggest that these conditions are not met?

138. The break angle in a torsion test of wire used in wrapping concrete pipe is specified to be 40 deg. Tests of wire wrapping a particular malfunctioning pipe produced the following results:

41.28	32.73	39.17	31.67	40.74	26.98	33.92	35.60
33.27	38.24	39.76	43.94	43.19	31.96	43.48	

Is there significant evidence in these data to suggest that the wire used on the malfunctioning pipe does not meet the specification? Give a detailed explanation of your conclusions suitable for management.

139. If X has an exponential distribution with mean μ then X has pdf

$$f(x;\mu) = (1/\mu)e^{-x/\mu} \ \ (x > 0).$$

i. Verify that $\Pr(X > x) = e^{-x/\mu} \ \ (x > 0)$.

The following observations, obtained on time to failure of a particular item, are supposed to be exponential with mean μ:

9.6	9.7	10.1	22.7	24.9	36.7	54.3	58.1	59.4	63.5
81.3	89.9	95.9	100*	100*	100*	100*	100*	100*	100*

The asterisks indicate that the last seven observations are actually censored, i.e. in these two cases the item has not failed by time 100, so in each of these cases, all that we know is that $(x > 100)$.

ii. Show that the log-likelihood function for these data is given by

$$\ln L = -13\ln\mu - \frac{1316.1}{\mu}.$$

iii. Hence find the maximum likelihood estimate of μ and a 95% confidence interval for μ.

iv. Use the relative log-likelihood function to test the hypothesis H_0: $\mu = 100$ against a two-sided alternative.

140. A Poisson process with rate α is observed over a period of time and 200 times between successive events recorded. These are as follows:

$$1.0, 0.4, 1.2, 0.5, 0.2, 2.7, 1.7, 3.9, 0.1, 0.3, 0.9, \ldots, 0.7$$

For these data, $\bar{x} = 0.632$ and $s^2 = 0.419$. Use these results to test the hypothesis that $\alpha = 2$ against the alternative that $\alpha < 2$. Specify the P-value.

141. A chemical company manufactures a transparent plastic in 1 m × 2 m sheets. Sheets quite often contain 'flaws', which means that part of the sheet has to be cut off and recycled. Changes have recently been made in the hope of improving matters. Prior to the changes, the rate of flaws was 15%. In the week after the changes, it was found that 7 out of 134 were flawed after the changes. Has there been a significant improvement? Give a statistical explanation.

142. Construct sequential tests for H_0: $p = 0.1$ vs H_1: $p = 0.2$

i. with size 0.05 and power 0.95;

ii. with size 0.05 and power 0.90;

iii. with size 0.01 and power 0.90.

143. Each hour, a random sample of twenty items is obtained from a production line. These items are tested and the number of defective items is recorded. Let x_i denote the number of defective items in the ith sample. Suppose that the proportion of defective items produced is p.

Show that a sequential test of H_0: p=0.05 vs H_1: p=0.1 with size 0.05 and power 0.95 (approximately) is defined as follows: if $\sum_{i=1}^{k} x_i < -4.01 + 1.45k$ then accept H_0; if $\sum_{i=1}^{k} x_i > 4.01 + 1.45k$, then accept H_1; otherwise continue.

144. Consider a sequence of independent observations on a $\mathrm{Pn}(\lambda)$ random variable. We wish to test H_0: $\lambda = \lambda_0$ against H_1: $\lambda = \lambda_1$, where $\lambda_1 > \lambda_0$. Show that

$$U_k = \ln L_1^{(k)} - \ln L_0^{(k)} = y_k \ln \frac{\lambda_1}{\lambda_0} - k(\lambda_1 - \lambda_0), \text{ where } y_k = \textstyle\sum_{i=1}^k x_i.$$

Deduce that a sequential test for $\lambda = 1$ against $\lambda = 2$ of size 0.05 and power 0.95 (approximately) has continuation region given by $-4.33 + 1.44k < y_k < 4.33 + 1.44k$, i.e. $-4.33 < y_k^* < 4.33$, where $y_k^* = \sum_{i=1}^k (x_i - 1.44)$.

Generate a sequence of observations on a $\mathrm{Pn}(1)$ distribution until a conclusion is reached.

145. It can be assumed that $X \overset{\mathrm{d}}{=} \mathrm{N}(\theta, 10^2)$.

It is required to test H_0: $\theta = 50$ vs H_1: $\theta > 50$, so that the size of the test is 0.05 and the power of the test, when $\theta = 55$, is 0.95.

i. Show that this is achieved by using a sample of $n = 44$ observations and rejecting H_0 when $\bar{x} > 52.5$.

ii. Construct a sequential test of H_0: $\theta = 50$ vs H_1: $\theta = 55$ with $\alpha = \beta = 0.05$.
[Hint: use $y_k^* = \sum_{i=1}^k (x_i - 52.5)$.*]*

iii. Generate observations on $X \overset{\mathrm{d}}{=} \mathrm{N}(55, 10^2)$ (so that H_1 is true), as follows:
use `>> x=normrnd(55,10,1,200);` to generate observations on x;
and `>> y=cumsum(x-52.5);` to obtain the cumulative sums.
Plot the cumulative sum and your test limits.
Report your decision, and the number of trials required to reach the decision.
How does this compare with the fixed sample test?

146. A high-voltage power supply should have a nominal output voltage of 350 V. A sample of four units are selected each day and tested for process-control purposes. The data shown below give values of x_i = (observed voltage on unit i – 350)×10.

sample	x_1	x_2	x_3	x_4	sample	x_1	x_2	x_3	x_4
1	6	9	10	15	11	8	12	14	16
2	10	4	6	11	12	6	13	9	11
3	7	8	10	5	13	16	9	13	15
4	8	9	6	13	14	7	13	10	12
5	9	10	7	13	15	11	7	10	16
6	12	11	10	10	16	15	10	11	14
7	16	10	8	9	17	9	8	12	10
8	7	5	10	4	18	15	7	10	11
9	9	7	8	12	19	8	6	9	12
10	15	16	10	13	20	14	15	12	16

Set up $\bar{x}$ and R charts for this process. Is the process in statistical control?

147. Suppose that a process is such that if it is in control then the quality variable is such that $Q \overset{\mathrm{d}}{=} N(\mu = 10,\ \sigma^2 = 1)$. Find the probability that twenty successive points will all be within the control limits if:

i. the process is in control;

ii. the process is out of control: $Q \overset{\mathrm{d}}{=} \mathrm{N}(\mu = 11, \sigma^2 = 1)$;

iii. the process is out of control: $Q \overset{\mathrm{d}}{=} \mathrm{N}(\mu = 12, \sigma^2 = 1)$;

iv. the process is out of control: $Q \overset{\mathrm{d}}{=} \mathrm{N}(\mu = 11, \sigma^2 = 4)$.

How does this relate to hypothesis testing?

148. Each day for 23 days, a random sample of five items from the day's production is selected and carefully measured: the average of the five measurements ($\bar{x}$) and the range of the five measurements (R) are calculated and recorded each day. At the end of the 23 days, the average of the daily averages, $\bar{\bar{x}} = 31.46$; and the average of the daily ranges, $\bar{R} = 7.13$. Assume that the measurements are approximately normal with mean μ and variance σ^2.

(a) Explain why $\bar{R} \approx 2.326\sigma$.

(b) Determine control limits for an $\bar{x}$-chart.

(c) Determine control limits for an R-chart.

149. Each day for twenty days, eight items are randomly selected from the day's production at a factory. These items are tested with the results given in the table below. For each day the sample mean, range and standard deviation are given; the average values of these statistics over the twenty days are also given.

	1	2	3	4	5	6	7	8	x-bar	range	stdev
1	53	53	55	54	54	55	52	54	53.8	3	1.04
2	55	54	54	56	56	56	52	52	54.4	4	1.69
3	55	55	56	55	54	54	58	58	55.6	4	1.60
4	53	52	56	57	53	56	54	55	54.5	5	1.77
5	58	55	53	56	56	59	57	52	55.8	7	2.38
6	54	52	52	56	55	54	53	54	53.8	4	1.39
7	55	56	55	54	58	58	56	56	56.0	4	1.41
8	52	58	54	54	59	53	57	57	55.5	7	2.56
9	55	56	58	56	51	54	59	55	55.5	8	2.45
10	52	58	59	54	52	55	55	56	55.1	7	2.53
11	57	55	55	54	58	57	54	52	55.3	6	1.98
12	54	56	54	59	55	59	58	54	56.1	5	2.23
13	57	50	54	54	55	52	57	51	53.8	7	2.60
14	52	58	60	54	52	55	57	54	55.3	8	2.87
15	53	54	51	52	55	58	53	57	54.1	7	2.42
16	52	58	55	54	59	58	51	56	55.4	8	2.92
17	58	56	57	55	54	56	53	56	55.6	5	1.60
18	58	52	55	55	54	56	53	52	54.4	6	2.07
19	54	56	54	56	55	55	53	56	54.9	3	1.13
20	56	56	57	56	55	55	56	55	55.8	2	0.71
								[average]	55.02	5.5	1.97

Use this information to determine control limits for an $\bar{x}$-chart and for an R-chart.

Plot an $\bar{x}$-chart and an R-chart for the following ten days data, given below, and comment on the result.

	1	2	3	4	5	6	7	8	x-bar	range	stdev
21	53	55	56	53	58	57	52	56	55.0	6	2.14
22	54	50	57	53	56	58	52	54	54.3	8	2.66
23	55	55	54	53	59	56	55	52	54.9	7	2.10
24	58	53	54	57	56	55	55	53	55.1	5	1.81
25	53	51	54	54	52	57	53	53	53.4	6	1.77
26	62	58	57	61	55	60	56	54	57.9	8	2.90
27	53	50	59	62	56	55	54	54	55.4	12	3.70
28	58	56	54	53	54	55	55	61	55.8	8	2.60
29	57	55	54	56	59	56	56	56	56.1	5	1.46
30	56	55	54	57	55	57	60	53	55.9	7	2.17

150. A manufacturer takes daily samples of 100 items, carefully inspects each item and records the number of items each day that are imperfect with the following results:

$$3, 2, 8, 2, 6, 1, 3, 2, 2, 7, 4, 5, 2, 6, 2, 3, 1, 11, 4, 4, 3, 2, 5, 4, 8.$$

For these data, there are 25 observations and the sum of the observations is 100.

Construct an appropriate control chart for these data and check for any out of control points.

151. A sample of 200 ROM computer chips was selected on each of 30 consecutive working days and the number of nonconforming chips on each day as follows:

10, 18, 24, 17, 37, 19, 7, 25, 11, 24, 29, 15, 16, 21, 18,
17, 15, 22, 12, 20, 17, 18, 12, 24, 30, 16, 11, 20, 14, 28.

Construct a p-chart and examine it for any out of control points.

152. A furniture manufacturer carefully inspects each item and keeps track of the number of minor blemishes per item with the following results:

3, 2, 8, 1, 6, 1, 3, 2, 2, 7, 4, 5, 2, 6, 2, 3, 8.

Construct an appropriate control chart for these data and check for any out of control points.

153. The table below is obtained from data on moisture content for specimens of a particular type of fabric. Each sample consisted of five fabric specimens.

sample	1	2	3	4	5	6	7	8	9	10
mean	12.72	12.80	13.04	13.14	12.52	13.20	12.78	13.18	13.28	12.74
range	1.2	0.9	1.7	1.3	1.3	1.5	2.2	1.3	2.4	0.9
sample	11	12	13	14	15	16	17	18	19	20
mean	12.78	12.92	12.94	13.18	12.88	12.92	13.02	13.04	13.14	12.84
range	0.9	1.2	1.4	1.2	1.2	1.3	1.8	0.5	1.6	1.5

(a) Determine control limits for a chart based on $\mu = 13.0$ and $\sigma = 0.6$, and comment on its appearance.

(b) Use the sample data to construct $\bar{x}$ and R charts. Does the process appear to be in control?

154. Water samples from eight sites on a river before and two years after an antipollution program was started, gave the following results. The numbers represent scores for a combined pollution measure, higher scores indicating greater pollution:

site	1	2	3	4	5	6	7	8
"before"	88	69	98	81	96	74	65	72
"after"	87	70	91	76	97	69	64	68

Test whether the antipollution program has been effective in reducing pollution.

155. Random samples were selected from two populations with the following results:

	A	$\bar{A}$	
sample 1	63	37	$(n_1 = 100)$
sample 2	48	52	$(n_2 = 100)$

Test the hypothesis that the proportion of individuals with characteristic A is the same in both populations, against the alternative that the proportion is higher in population 1.

156. Test the hypothesis that the populations from which the following random samples were independently obtained are identically distributed:

sample 1:	27, 31, 42, 37, 40, 43, 33, 32, 28, 29, 30, 32, 36, 41, 40, 39, 36, 32, 33, 33.
sample 2:	35, 32, 39, 44, 43, 40, 29, 38, 35, 37

It may be assumed that the populations are normally distributed.

157. The following are results of independent random samples on $X_1 \overset{\mathrm{d}}{=} \mathrm{N}(\mu_1, \sigma^2)$ and $X_2 \overset{\mathrm{d}}{=} \mathrm{N}(\mu_2, \sigma^2)$ (so that $\mathrm{var}(X_1) = \mathrm{var}(X_2)$).

$$n_1 = 30, \quad \bar{x}_1 = 14.37, \quad s_1^2 = 8.39;$$
$$n_2 = 20, \quad \bar{x}_2 = 12.95, \quad s_2^2 = 11.04.$$

Test the hypothesis that $\mu_1 = \mu_2$ against a two-sided alternative.

158. We wish to investigate the possible improvement in precision for a new method of gas analysis. Let σ_1^2 denote the new error variance and σ_2^2 denote the old error variance. Trial runs with instruments gave the following results:

$$s_1^2 = 0.251 \quad (n_1 = 13) \quad \text{and} \quad s_2^2 = 0.897 \quad (n_2 = 25).$$

Test the hypothesis that $\sigma_1^2 = \sigma_2^2$ against the alternative that $\sigma_1^2 < \sigma_2^2$.

159. Population A is normally distributed with mean $\mu_A = 28$ and standard deviation $\sigma_A = 7$; population B is normally distributed with mean $\mu_B = 26$ and standard deviation $\sigma_B = 5$. A random sample of $n_A = 20$ is obtained from population A and a random sample of $n_B = 10$ is obtained from population B.

(a) Find $\Pr(|\bar{X}_A - \bar{X}_B| > 1)$, where $\bar{X}_A$, $\bar{X}_B$ denote the respective sample means.

(b) Find, approximately, $\Pr(S_B < S_A)$, where S_A, S_B denote the respective sample standard deviations.

160. The following samples are from normally distributed populations A and B, with equal variances:

A:	4.37, 3.42, 3.76, 4.12, 2.87
B:	4.57, 5.13, 5.07, 4.88, 6.00, 5.50

(a) Test the hypothesis that the populations have equal means using a t-test.

(b) Give a confidence interval for the difference between the means.

161. An engineer is required to compare the time required for two different work sequences in a garment factory used to measure the shear strength of polyester fibres. To collect some data, 12 workers are randomly divided into two groups. The first group measures the shear strength using work sequence A and the second group using work sequence B. The data recorded were the twelve completion times.

work sequence A:	235, 214, 197, 206, 214, 220
work sequence B:	247, 223, 215, 219, 207, 236

Assuming equal variances, test whether the difference in average completion times is significantly different from zero.

162. Of two hundred individuals trying to stop smoking, one hundred are given treatment A and one hundred treatment B. Of those given treatment A, 23 stopped, while of those given treatment B, only 11 stopped. Find a 95% confidence interval for the difference in probabilities of stopping.

163. Two items are randomly selected from each of ten batches. One is untreated. The other is treated with a hardening agent. Both items are tested to failure and the result recorded. Higher scores indicate a stronger, longer lasting item.

batch	1	2	3	4	5	6	7	8	9	10
untreated	123	117	107	121	104	117	133	112	115	126
treated	134	122	107	128	102	120	142	118	114	131

Test whether the treatment produces a significantly stronger product. Give a 95% confidence interval for the mean increase.

164. Yields from a particular process using raw materials from two different sources were as given below:

(A) 75.1 80.3 85.7 77.2 83.9 85.1 86.7 80.5 75.2 70.0
(B) 80.4 83.1 85.6 90.7 86.6 95.9 82.8 80.9 78.7 89.4

Test the hypothesis that the source material has no effect on yield.

165. The use of carpeting in hospitals, while having obvious aesthetic merits, raises a question as to whether carpeted floors are sanitary. One way to get an answer to this question is to compare bacterial levels in carpeted and uncarpeted rooms. Airborne bacteria can be counted by passing room air at a known rate over a growth medium, incubating that medium, and then counting the number of bacterial colonies that form. In one such study this procedure was repeated in sixteen patient rooms, eight carpeted and eight uncarpeted with the following results:

carpeted rooms	bacteria/m^3	uncarpeted rooms	bacteria/m^3
# 212	0.34	# 210	0.35
# 216	0.23	# 214	0.24
# 220	0.20	# 215	0.11
# 223	0.37	# 217	0.21
# 225	0.31	# 221	0.34
# 226	0.29	# 222	0.32
# 227	0.42	# 224	0.29
# 228	0.40	# 229	0.39

Assuming equal variances, test whether carpeting has any effect on the level of airborne bacteria in patient rooms.

166. An experiment is performed to measure the effect of different diets on the liver content of vitamin A in rats. A group of 20 fairly similar rats is divided at random into two groups, with ten rats in each group. The first group receives a normal diet, but the second group receives a diet deficient in vitamin E. At a later stage, the vitamin A in the liver of each rat is measured.
For the first group the measurements have mean 3375 and standard deviation 626; for the second group, the measurements have mean 2570 and standard deviation 538.

(a) Has the vitamin E deficient diet caused a significant reduction in the vitamin A content of the rat livers?

(b) Specify an approximate 95% confidence interval for the mean reduction in vitamin A content.

167. A randomly selected group of 180 engineering students are given a spelling test: 70/80 females pass the test whereas 80/100 males pass. Find a 95% confidence interval for the difference $p_f - p_m$, where p_f and p_m denote the probability of passing this spelling test for female engineering students and male engineering students respectively.

168. The following table gives the number of children with dental cavities in surveys conducted before and after fluoridation of a city's water supply:

	with cavities	*without cavities*
before fluoridation	143	57
after fluoridation	62	88

Set up an appropriate hypothesis and test it.

169. The data below give the yield of a chemical process with each of three catalysts:

C_1	91	93	90	94
C_2	98	96	95	94
C_3	93	92	90	87

(a) Test for differences in yields with the three catalysts using an analysis of variance.

(b) Find 95% confidence intervals for the mean yield with each catalyst.

(c) Find 95% confidence intervals for the differences in mean yields between each pair of catalysts.

170. Determinations of the strength of a fibre after using three treatments were as follows:

		mean	sd
(1) control	$67, 72, 61, 76$	69.0	6.48
(2) treatment A	$74, 83, 81, 90, 84$	82.4	5.77
(3) treatment B	$86, 92, 89$	89.0	3.00

(a) Complete the following analysis of variance table derived from the above data:

	df	SS	MS	F
between methods		753.7		
within methods				
total				

[Note: the within methods sum of squares, $W = \sum(n_i - 1)s_i^2$; and the total sum of squares $T = \sum(y - \bar{y})^2 = (N-1)s_y^2$.]

(b) Test the hypothesis H_0: $\mu_1 = \mu_2 = \mu_3$ giving an approximate P-value.

(c) Find a 95% confidence interval for the mean strength with no treatment, i.e., μ_1.

(d) Find a 95% confidence interval for the effect of treatment B relative to the control treatment, i.e., $\mu_3 - \mu_1$.

(e) Test the hypothesis that $\mu_2 = \mu_3$ against a two-sided alternative.

171. Random samples were obtained on each of five normal populations having equal variances, σ^2. The samples contained $n_1 = 5, n_2 = 3, n_3 = 7, n_4 = 2$ and $n_5 = 3$ observations respectively.

(a) Complete the following analysis of variance table derived from these samples:

	df	SS	MS	F
between samples		13.76		
within samples				
total	19	28.16		

(b) Show that the hypothesis that the populations have equal means is rejected using a test of size 0.05.

(c) Give an estimate of and a 95% confidence interval for σ^2.

172. The data below give the running costs for three types of truck:

A	108	107	104	105		
B	101	98	94	100	101	94
C	98	96	94			

(a) Test for differences in running costs of the three types of truck using an analysis of variance.

(b) Find 95% confidence intervals for the mean running cost of each type of truck.

(c) Find 95% confidence intervals for the differences between the mean running costs for each pair of truck types.

173. The following observations are standardised yields from 18 experimental units divided into three blocks each containing six homogeneous plots. Three treatments were applied twice in each block as indicated.

	Treatment 1	Treatment 2	Treatment 3	
Block 1	0, 1	1, –1	2, 0	3
Block 2	–1, –2	2, 1	1, 2	3
Block 3	0, 2	1, 2	4, 3	12
	0	6	12	18

For these data $\Sigma\Sigma\Sigma y^2 = 56$.

Assuming an additive model with independent normally distributed errors having equal variances,

(a) test the significance of the treatment effects;

(b) find an estimate of the error variance;

(c) find 95% confidence intervals for the treatment differences.

174. In the production of a particular material two factors are of interest: the catalyst used in the experiment (three catalysts: C_1, C_2 and C_3) and the washing time (W) of the product following the cooling process (W = 15 minutes and W = 20 minutes). Three runs were made at each combination of the factors. The coded yields are as follows:

	C_1	C_2	C_3
$W = 15$	10.7 10.8 11.3	10.3 10.2 10.5	11.2 11.6 12.0
$W = 20$	10.9 12.1 11.5	10.5 11.1 10.3	12.2 11.7 11.0

From these data the following, incomplete, analysis of variance table was obtained.

	df	SS	MS	F
washing times		0.405		
catalysts			1.982	
interaction		0.213		
error				
total		6.949		

Assuming independent normal errors with constant variance,

(a) carry out a test to show that a strictly additive model is reasonable;

(b) assuming a strictly additive model, test for differences between catalysts and between washing times;

(c) assuming a strictly additive model, find 95% confidence intervals for the difference in mean yield for the two washing times, and for differences between the mean yields for pairs of catalysts.

175. The data below were obtained in a study to compare the moisture absorption of five different concrete aggregates. For each aggregate, six samples were exposed to moisture for 48 hours, and the amount of moisture absorbed was determined as a percentage, by weight, of aggregate.

aggregate	absorbed moisture					
1	551	457	450	731	499	632
2	595	580	508	583	633	517
3	639	615	511	573	648	677
4	417	449	517	438	415	555
5	563	631	522	613	656	679

(a) Complete the following analysis of variance table derived from the above data:

	df	SS	MS	F
between aggregates		85356		
within aggregates				
total		209377		

(b) Find a 95% confidence intervals for the mean absorption of aggregate 4.

(c) Find a 95% confidence intervals for the differences in mean absorption between aggregates 1 and 4.

176. An investigation was conducted to determine the source of reduction in yield of a certain chemical product. It was known that the loss in yield occurred in the mother liquid, that is, the material removed at the filtration stage. It was felt that different blends of the original material may result in different yield reductions at the mother liquid stage. The following are results of the percent reduction for three batches at each of four preselected blends:

blend			
1	2	3	4
25.6	25.2	20.8	31.6
24.3	28.6	26.7	29.8
27.9	24.7	22.2	34.3

(a) Perform an analysis of variance on these data, and state your conclusion.

(b) Use 95% confidence intervals to determine which blends differ significantly from each other.

177. Four laboratories are used to perform chemical analyses. Samples of the same material are sent to the laboratories for analysis as part of a study to determine whether or not they give, on average, the same results. The results are as follows:

Laboratory			
A	B	C	D
58.7	62.7	55.9	60.7
61.4	64.5	56.1	60.3
60.9	63.1	57.3	60.9
59.1	59.2	55.2	61.4
58.2	60.3	58.1	62.3

Complete the following analysis of variance table and state your conclusions.

	df	SS	MS	F
between laboratories			28.64	
within laboratories				
total		120.31		

178. The data below were obtained in a study to calibrate four thermometers. The study consisted of using each of the thermometers to measure the melting temperature of each of four chemical cells. The nature of the cells, and the study, were such that only one thermometer could be used for each of the (16) trials. Temperature was measured in degrees Centigrade, but only the third and fourth decimal places are given, as the readings agreed up to the last two places.

	Thermometer			
Chemical Cell	1	2	3	4
1	36	38	36	30
2	17	18	26	17
3	30	39	41	44
4	30	45	38	33

From these data the following, incomplete, analysis of variance table was obtained.

	df	SS	MS	F
between cells			302.92	
between thermometers		136.25		
error				
total		1239.75		

Assuming an additive model with independent normally distributed errors having equal variances,

(a) find an estimate of and a 95% confidence interval for the error variance;

(b) test for differences between the thermometers;

(c) find 95% confidence intervals for differences between pairs of thermometers.

179. When metal pipe is buried in soil it is desirable to apply a coating to retard corrosion. Four coatings are under consideration for use with pipe that will ultimately be buried in three types of soil. An experiment to investigate the effects of these coatings and soils was carried out by first selecting 12 pipe segments and applying each coating to three segments. The segments were then buried in soil for a specified period in such a way that each soil type received one piece with each coating. The resulting data (depth of corrosion) is given below.

	soil type		
coating	1	2	3
1	68	53	54
2	53	51	48
3	44	42	47
4	51	43	52

Assuming that there is no interaction between coating type and soil type, test for possible effects of soil type and coating and find 95% confidence intervals for differences between the means of the levels of any significant effects.

180. Certain voltage regulators are required to operate within the range 15.8–16.4 volts. One regulator was set by five operatives, each using a different setting machine, and was then tested at each of four testing stations. The results are given below.

Setting	Testing station			
station	1	2	3	4
1	16.5	16.5	16.6	16.6
2	15.5	15.5	15.3	15.6
3	15.9	16.0	16.0	16.5
4	16.0	16.0	15.9	16.3
5	16.2	16.1	15.8	16.0

Complete the following analysis of variance table to test for biases between testing stations and between setting stations.

	df	SS	MS	F
between settings			0.583	
between testings		0.204		
error				
total		2.788		

181. The following data were obtained in a study involving two factors, A and B.

	B		
A	1	2	3
1	17.5 18.5 22.1	14.8 13.6 12.2	15.9 14.8 13.6
2	13.3 14.6 16.2	17.2 15.5 14.2	16.1 14.7 13.4

Assuming independent normal errors with constant variance, carry out a complete analysis of these data, including interpretation of the interaction, if significant, and state your conclusions.

182. (a) Twenty-four plots are available in six blocks of four. It is required to run an experiment to compare four treatments. Explain how the treatments should be assigned to the plots.

(b) The experiment described in (a) is carried out, with results indicated in the table below. The observations are standardised yields.

	trt 1	trt 2	trt 3	trt 4	total
block 1	0	1	2	5	8
block 2	–1	1	1	3	4
block 3	0	2	4	6	12
block 4	3	4	7	10	24
block 5	2	1	5	8	16
block 6	2	3	5	10	20
total	6	12	24	42	84

```
Analysis of Variance for y
source          df       ss       ms       f
blocks                   70
treatments
error
total                   210
```

For these data, show that the sum of squares due to treatments is equal to 126, and complete the above analysis of variance table.

Assuming an additive model with independent normally distributed errors having equal variances

i. test the significance of the treatment effects;

ii. find an estimate of the error variance;

iii. find a 95% confidence interval for the mean yield with treatment 4;

iv. find a 95% confidence interval for the difference in mean yield with treatment 4 and treatment 1.

183. An evaluation of diffusion bonding of zircaloy components is performed. The main objective is to determine which of three elements — nickel, iron or copper — is the best bonding agent. A series of zircaloy components is bonded, using each of the possible bonding agents. Since there is a great deal of variation among components from different ingots, a randomised block design is used, blocking on the ingots. A pair of components from each ingot is bonded together, using each of the three agents, and the pressure (in units of 1000 pounds per square inch) required to separate the bonded components is measured. The following data are obtained:

ingot	nickel	iron	copper
1	67.0	71.9	72.2
2	67.5	68.8	66.4
3	76.0	82.6	74.5
4	72.7	78.1	67.3
5	73.1	74.2	73.2
6	65.8	70.8	68.7
7	75.6	84.9	69.0

Is there evidence of a difference in pressure required to separate the components among the three bonding agents?

184. A randomised block design was conducted to compare the mean responses for three treatments (A, B and C) in four blocks. The data are shown below:

	block 1	block 2	block 3	block 4
treatment A	3	6	1	2
treatment B	5	7	4	6
treatment C	2	3	2	2

(a) Compute the appropriate sums of squares and complete the following analysis of variance table:

source	df	SS	MS	F
treatments	..	..	..	..
blocks	..	..	..	..
error	..	..	..	
total	..	..		

(b) Do the data provide sufficient evidence to indicate a difference among the treatment effects?

(c) Do the data provide sufficient evidence to indicate that blocking was effective in reducing experimental error?

185. (a) i. Write a few lines on the importance of randomisation in the design of experiments.

ii. Define 'block' and explain its importance in the context of the design of experiments.

(b) A randomised block experiment to compare four treatments is to be conducted in eight blocks with eight plots per block. Each treatment is to be replicated twice in each block. Explain in detail how you would decide which treatment would be assigned to which plot.

(c) The experiment described in (b) is carried out, resulting in the analysis of variance given below. For the analysis, it can be assumed that the observations are independent and normally distributed with equal variances.

	df	SS	MS
blocks	..	...	..
treatments	..	30	..
error	..	...	1
total	..	223	

i. Complete the ANOVA table and hence test the significance of the treatment effects.

ii. Specify the standard error of the estimate of $\alpha = \frac{1}{2}(\tau_1 + \tau_2) - \frac{1}{2}(\tau_3 + \tau_4)$.
If $\hat{\alpha} = -0.75$, test the hypothesis that $\alpha = 0$ against a two-sided alternative.

186. A randomised block experiment to compare four treatments (T_1, T_2, T_3, T_4) is carried out in six blocks ($B_1, B_2, \ldots, B_6$) of four plots/block with the following results:

	B_1	B_2	B_3	B_4	B_5	B_6	av
T_1	39	51	59	25	27	33	39
T_2	47	55	61	33	33	41	45
T_3	..	..	..	..	..	..	..
T_4	..	..	..	..	..	..	..

(in which the dots indicate observations, the values of which are omitted).

The following sums of squares are given:

source	df	SS	MS
T		96	
B		3200	
error		60	
total		3356	

(a) Complete the above analysis of variance table and hence test the hypothesis that there is no difference between the treatment effects.

(b) Give an estimate and 95% confidence interval for

i. the error variance;

ii. the difference in effect of treatment 2 and treatment 1.

187. A randomised block experiment to compare three treatments was conducted in eight blocks with six plots per block. Each treatment was replicated twice in each block. It can be assumed that the observations are independent and normally distributed with equal variances.

i. Complete the following ANOVA table and hence test the significance of the treatment effects.

	df	SS	MS
blocks	...	154	...
treatments	...	...	...
error	...	...	2
total	...	254	

ii. Find a 95% confidence interval for the error variance.

iii. It is required to estimate the difference in effects of treatment 3 and treatment 1. This is done by comparing the average yield with treatment 1 and the average yield with treatment 3. The resulting estimate is 2.3. Find its standard error and hence obtain a 95% confidence interval for this difference.

188. A consumer protection organization wants to compare the annual power consumption of five different brands of dehumidifiers. Because power consumption depends on the prevailing humidity level, each brand was tested at four different humidity levels, ranging from moderate to heavy humidity. For each brand, a sample of four humidifiers was randomly assigned, one each, to the four humidity levels. The resulting annual power consumption (in kWh) is given in the following table:

	H_1	H_2	H_3	H_4
B_1	685	792	838	875
B_2	722	806	893	953
B_3	733	802	880	941
B_4	811	888	952	1005
B_5	828	920	978	1023

(a) Can you conclude that there is a difference between the power consumptions of the five brands?

(b) Can you conclude that there are differences in power consumption between the levels of the blocking factor "humidity"? Does this results support the experimenters' use of humidity as a blocking factor?

189. A randomised block experiment for comparing three different methods (A, B and C) of curing concrete is carried out. Different batches of concrete are used as the blocks in the experiment. For convenience, the data are repeated here:

batch	Method A	Method B	Method C
1	30.7	33.7	30.5
2	29.1	30.6	32.6
3	30.0	32.2	30.5
4	31.9	34.6	33.5
5	30.5	33.0	32.4
6	26.9	29.3	27.8
7	28.2	28.4	30.7
8	32.4	32.4	33.6
9	26.6	29.5	29.2
10	28.6	29.4	33.2

i. Can you conclude that there is a difference in mean concrete strength between the three curing methods?

ii. Can you conclude that there are differences between the batch means?

iii. Suppose that you ignore the fact that the batches are blocks in this experiment and that you simply run a one-factor ANOVA test, treating the three columns of data as three random samples. What conclusion do you reach regarding the differences between the three curing methods?

190. A consumer protection organization carried out a study to compare the electricity usage for four different types of residential air-conditioning systems. Each system was installed in given homes and the monthly electricity usage (in kilowatt-hours) was measured for a particular summer month. Because of the many differences that can exist between residences (e.g. floor space, type of insulation, type of roof, etc.), five different groups of homes were identified for study. From each group of homes of a similar type, four homes were randomly selected to receive one of the four air-conditioning systems. The resulting data is given in the table:

	H_1	H_2	H_3	H_4	H_5
AC_1	116	118	97	101	115
AC_2	171	131	105	107	129
AC_3	138	131	115	93	110
AC_4	144	141	115	93	99

i. Construct an ANOVA table for this experiment.

ii. Can you conclude that there is a difference between the monthly mean kilowatt-hours of electricity used by the four types of air-conditioners?

191. A Latin square experiment for four treatments (A, B, C, D) with row and column blocking produces the following data:

	col=1	col=2	col=3	col=4
row=1	(A) 9.1	(C) 22.0	(D) 19.0	(B) 20.9
row=2	(D) 14.7	(B) 21.8	(C) 14.7	(A) 7.0
row=3	(C) 14.3	(A) 16.6	(B) 21.9	(D) 17.2
row=4	(B) 13.9	(D) 17.1	(A) 5.3	(C) 8.7

Using these data the following output was obtained:

```
Analysis of Variance for y
source      df    SS     MS       F
rows         *    --    37.069  45.62
columns      *    --    33.758  41.55
treatments   *    --    73.827  90.86
error        *    --     0.813
total        *    --    S=0.90
```

i. Specify the missing degrees of freedom in the above table.

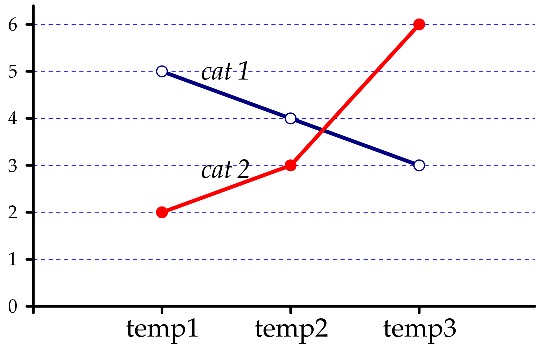
ii. Test the significance of the treatment effects.

iii. Specify the standard error of $\hat{\tau}_C - \hat{\tau}_D$, i.e. the estimate of the difference between the effects of treatment C and treatment D.

192. A fixed effects model is used to analyse two factors, each of which has five levels. Three replicated measurements are available for each combination of factor levels. Complete the following ANOVA table for this experiment, and interpret your results.

```
source          df   SS      MS     F
Factor A             20
Factor B                           8.1
AB interaction
Error                         2
Total               200
```

193. The yield of a chemical process is observed at three temperature levels (1=low, 2=moderate and 3=high) for each of two catalysts. Four replicate observations are obtained for each catalyst-temperature combination and the averages of these sets of four observations are plotted in the following diagram.



For these data, the error mean square, $s^2 = 2$.

i. Indicate the number of degrees of freedom for each of the components in the standard analysis of variance for this situation. Indicate also whether the corresponding F-tests are likely to be significant for these data, giving reasons for your answers.

ii. Find a 95% confidence interval for the expected difference in yield with catalyst 2 at high temperature and catalyst 1 at low temperature, i.e. $\mu_{23} - \mu_{11}$.

194. The yield percentage (y) of a certain precipitate depends on the concentration of the reactant (C) and the rate of addition of diammonium hydrogen phosphate. Experiments were run at three different concentration levels (C_1, C_2 and C_3) and three addition rates (R_1, R_2 and R_3). The yield percentages, with two observations per treatment, were as follows:

	C_1	C_2	C_3
R_1	90.1, 90.3	92.4, 91.8	96.4, 96.8
R_2	91.2, 92.3	84.3, 93.9	98.2, 97.6
R_3	92.4, 92.5	96.1, 95.8	99.0, 98.9

The following split up of the sum of squares is obtained for these data:

	df	SS	MS	F
addition rate		23.974		
concentration level		122.368		
interaction		1.702		
error		1.200		
total		149.244		

(a) Complete the analysis of variance table and determine the significant effects.

(b) Give a brief interpretation of the results.

195. The following data was obtained in an experiment to investigate whether the yield from a certain chemical process depends on either the chemical formulation of the input materials or the mixer speed, or on both factors:

	60	70	80
F_1	189.7 188.6 190.1	185.1 179.4 177.3	189.0 193.0 191.0
F_2	165.1 165.9 167.6	161.7 159.8 161.6	163.3 166.6 170.3

A statistical software package gave the following sums of squares:
SS(formulation) = 2253.44, SS(speed) = 230.81,
SS(interaction) = 18.58, and SS(error) = 71.87.

i. Does there appear to be interaction between the two factors?

ii. Does the yield appear to depend on either the formulation or the speed?

196. In an automated chemical coating process, the speed with which objects on a conveyer belt are passed through a chemical spray (belt speed, S), the amount of chemical sprayed (spray volume, V), and the brand of chemical used (brand, B) are factors that may affect the uniformity of the coating applied. A replicated 2^3 experiment was conducted in an effort to increase the coating uniformity. In the following table, higher values of the response variable are associated with higher surface uniformity:

run	V	S	B	r_1	r_2
1	−	−	−	40	36
2	+	−	−	25	28
3	−	+	−	30	32
4	+	+	−	50	48
5	−	−	+	45	43
6	+	−	+	25	30
7	−	+	+	30	29
8	+	+	+	52	49

i. Create the ANOVA table for this experiment. Which factors appear to have an effect on surface uniformity?

ii. Draw the main effects or interaction effects plot for the factors identified in i.

iii. Which settings (high or low) of the factors identified in part i. lead to maximizing surface uniformity?

197. Coal-fired power plants used in the electrical industry have gained increased public attention because of the environmental problems associated with solid wastes generated by large-scale combustion. A study was conducted to analyse the influence of three factors, binder type (A), amount of water (B) and land disposal scenario (C), that affect certain leaching characteristics of solid wastes from combustion. Each factor was studied at two levels. An unreplicated 2^3 experiment was run, and a response value EC50 (the Effective Concentration, in mg/L, that decreases 50% of the light in a luminescence bioassay) was measured for each combination of factor levels. The experimental data is given in the following table:

	A	B	C	EC50
1	0	0	0	23100
2	1	0	0	28500
3	0	1	0	71400
4	1	1	0	76000
5	0	0	1	37000
6	1	0	1	33200
7	0	1	1	17000
8	1	1	1	16500

i. Calculate the 'complete' ANOVA for this experiment.

ii. Create a half-normal plot of the effects. Which effects appear to be important?

iii. Which settings (high or low) of the factors in part ii. lead to minimizing EC50?

iv. Which settings (high or low) of the factors in part ii. lead to maximizing EC50?

198. Impurities in the form of iron oxides lower the economic value and usefulness of industrial minerals, such as kaolins, to ceramic and paper-processing industries. A 2^4 experiment was conducted to assess the effects of four factors on the percentage of iron removed from kaolin samples. The factors and their levels are given by:

factor	description	units	(0)	(1)
A	H_2SO_4	M	0.10	0.25
B	Thiourea	g/l	0.0	5.0
C	Temperature	degC	70	90
D	Time	min	30	150

The data from an unreplicated 2^4 experiment are listed in the table below:

run	%	run	%
(1)	7	d	28
a	11	ad	51
b	7	bd	33
ab	12	abd	57
c	21	cd	70
ac	41	acd	95
bc	27	bcd	77
abc	48	abcd	99

i. Calculate an ANOVA including all main effects and two-factor interaction effects for this experiment.

ii. Create a half-normal plot of the effects. Which effects appear to be important?

iii. Which settings (high or low) of the factors in part ii. lead to maximizing the percentage of iron extracted?

iv. Write a model for predicting iron extraction percentage from the factors identified in part ii.

199. In an effort to reduce the variation in copper plating thickness on printed circuit boards, a fractional factorial design was used to study the effect of three factors, anode height (up or down), circuit board orientation (in or out) and anode placement (spread or tight) on plating thickness. The following factor combinations were run:

Anode height	Board orientation	Anode placement	Thickness variation
0	0	0	11.63
0	1	1	3.57
1	0	1	5.57
1	1	0	7.36

(a) Specify k and p for this 2^{k-p} design.

(b) Calculate estimates of the main effects.

(c) Obtain the ANOVA for the additive model. What conclusions do you reach?

(d) If the objective of the study is to minimize the variation in plating thickness, what setting of each factor do you recommend?

200. A half-replicate of a 2^5 experiment is used to study the effects of heating time (A), quenching time (B), drawing time (C), position of heating coils (D) and measurement position (E) on the hardness of steel castings. The following data were obtained:

a	70.4	acd	66.6
b	72.1	ace	67.5
c	70.4	ade	64.0
d	67.4	bcd	66.8
e	68.0	bce	70.3
abc	73.8	bde	67.9
abd	67.0	cde	65.9
abe	67.8	abcde	68.0

Assuming that second and higher order interactions are negligible, conduct tests for the presence of main effects.

201. A nitride etch process was studied on a single-wafer plasma etcher. The response of interest was the etch rate. These were the experimental factors: A-Gap, the spacing between the anode and the cathode B-Pressure in the reactor chamber C-Flow rate of the reactant gas (C_2F_6) D-Power applied to the cathode For the purposes of this discussion, factors A (gap) and B (pressure) are the control factors. Factors C (flow rate) and D (power) are the noise factors. The factor levels used were as listed in the following table:

	A	B	C	D
level	Gap	Pressure	C_2F_6 Flow	Power
Low (0)	0.8	450	125	275
High (1)	1.2	550	200	325

The table below summarizes the design and observed results. Perform an appropriate analysis of this experiment.

	A	B	C	D	y
run	gap	pressure	C_2F_6 flow	power	etch rate
1	0	0	0	0	550
2	1	0	0	0	669
3	0	1	0	0	604
4	1	1	0	0	650
5	0	0	1	0	633
6	1	0	1	0	642
7	0	1	1	0	601
8	1	1	1	0	635
9	0	0	0	1	1037
10	1	0	0	1	749
11	0	1	0	1	1052
12	1	1	0	1	868
13	0	0	1	1	1075
14	1	0	1	1	860
15	0	1	1	1	1063
16	1	1	1	1	729

202. A fractional factorial experiment for five factors was run with results as indicated below.

```
     a    b    c    d    e       y
1    0    0    0    0    0    10.5
2    0    0    0    1    1     7.0
3    0    1    1    0    0    13.7
4    0    1    1    1    1     9.7
5    1    0    1    0    1    13.4
6    1    0    1    1    0    13.3
7    1    1    0    0    1     9.4
8    1    1    0    1    0    11.0
```

```
Estimated Effects and Coefficients
Term       Effect      Coef  Std.Coef t-value      P
Constant             11.000    0.2215   49.66  0.000
a           1.550     0.775    0.2215    3.50  0.073
b          -0.100    -0.050    0.2215   -0.23  0.842
c           3.050     1.525    0.2215    6.88  0.020
d          -1.500    -0.750    0.2215   -3.39  0.077
e          -2.250    -1.125    0.2215   -5.08  0.037

Analysis of Variance
Source           DF        SS       MS      F      P
Main Effects      5   38.0550   7.6110  19.39  0.050
Residual Error    2    0.7850   0.3925
Total             7   38.8400
```

(a) What fraction of the possible factor combinations have been used?

(b) Verify that the effect of A is the difference between the average yield with A at level 1 and the the average yield with A at level 0.

(c) Give an estimate of the error variance.

(d) Which factors have significant effects?

(e) What are your conclusions and recommendations?

203. A fractional factorial design was used to determine which of five possible factors influenced the determination of carbon in cast iron. The factors and their levels are:

A — Testing time	Morning	Afternoon
B — KOH conc	38	42
C — Heating time	45	75
D — Oxygen flow	slow	fast
E — Muffle temp	950	1100

The experimental results are given in the following table:

x_1	x_2	x_3	x_4	x_5	y
0	0	0	0	0	3.130
0	0	1	1	0	3.065
0	1	0	0	1	3.105
0	1	1	1	1	2.940
1	0	0	1	1	2.940
1	0	1	0	1	3.110
1	1	0	1	0	3.145
1	1	1	0	0	3.240

Identify this design and analyse the experimental results.

204. An experiment was carried out to investigate the effect of tool speed (A), depth (B) and feed rate (C) on the life of a cutting tool. The experimental results are as follows:

Speed (rpm)	Depth (in.)	Feed (ips)	Tool Life (min)
1000	0.02	2	200
2000	0.02	2	225
1000	0.08	2	28
2000	0.08	2	35
1000	0.02	10	150
2000	0.02	10	154
1000	0.08	10	17
2000	0.08	10	21

i. Estimate all of the main effects and interactions.

ii. Plot the estimated effects on a half-normal probability plot.

iii. Interpret your results.

205. An experiment to investigate the surface charge on a silicon wafer is carried out as follows. The factors thought to influence induced surface charge are cleaning method (spin rinse dry or SRD and spin dry or SD) and the position on the wafer (left, right) where the charge was measured. The surface charge ($\times 10^{11}$ $\mathrm{q/cm^3}$) data are:

	L	R
SD	1.66	1.84
	1.90	1.84
	1.92	1.62
SRD	–4.21	–7.58
	–1.35	–2.20
	–2.08	–5.36

Analyse these experimental results.

206. An engineer is interested in the effect of cutting speed (A), metal hardness (B), and cutting angle (C) on the life of a cutting tool. Two levels of each factor are chosen, and two replicates of a 2^3 factorial design are run. The tool life data (in hours) are shown in the following table.

	r1	*r2*
(1)	221	311
a	325	435
b	354	348
ab	552	472
c	440	453
ac	406	377
bc	605	500
abc	392	419

i. Analyse the data from this experiment.

ii. Specify an appropriate model that explains tool life in terms of the variables used in the experiment.

207. The data shown here represent a single replicate of a 2^5 study design that is used in an experiment to study the compressive strength of concrete. The factors are mix (A), time (B), laboratory (C), temperature (D), and drying time (E).

(1) = 700	e = 800
a = 900	ae = 1200
b = 3400	be = 3500
ab = 5500	abe = 6200
c = 600	ce = 600
ac = 1000	ace = 1200
bc = 3000	bce = 3000
abc = 5300	abce = 5500
d = 1000	de = 1900
ad = 1100	ade = 1500
bd = 3000	bde = 4000
abd = 6100	abde = 6500
cd = 800	cde = 1500
acd = 1100	acde = 2000
bcd = 3300	bcde = 3400
abcd = 6000	abcde = 6800

i. Estimate the factor effects

ii. Which effects appear important? Use a half-normal probability plot.

iii. Determine an appropriate model and analyse the residuals from this experiment. Comment on the adequacy of the model.

iv. If it is desirable to maximize the strength, in which direction would you adjust the process variables?

208. The following data were obtained by measuring the tensile strength (y) and hardness (x) of each of fifteen specimens of cold-drawn copper.

specimen	x	y	specimen	x	y
1	104.2	38.9	9	104.0	37.6
2	106.1	40.4	10	101.5	33.2
3	105.6	39.9	11	101.9	33.9
4	106.3	40.8	12	100.6	29.9
5	101.7	33.7	13	104.9	39.5
6	104.4	39.5	14	106.2	40.6
7	102.0	33.0	15	103.1	35.1
8	103.8	37.0			

For these data:

$$\begin{array}{ccc} n = 15 & \sum xy = 57464.62 & \sum(x-\bar{x})(y-\bar{y}) = 89.0267 \\ \sum x = 1556.3 & \sum x^2 = 161520.87 & \sum(x-\bar{x})^2 = 49.5573 \\ \sum y = 553.0 & \sum y^2 = 20555.80 & \sum(y-\bar{y})^2 = 168.5333 \end{array}$$

(a) Plot a scatter diagram.

(b) Fit a straight line to the regression of y on x using the method of least squares.

(c) Obtain a 95% confidence interval for the mean tensile strength if the hardness $x = 102$.

(d) Obtain a 95% prediction interval for the tensile strength if the hardness $x = 102$.

(e) Compute the sample correlation coefficient, r.

(f) Carry out a test of H_0: $\rho = 0$ against H_0: $\rho \neq 0$, assuming the population is bivariate normal.

209. An investigation into the relationship between the amount of nickel and the volume percent austenite in various steels yielded the results below.

amount of nickel, x	percent austenite, y
0.608	2.11
0.634	1.95
0.651	2.27
0.658	1.95
0.675	2.05
0.677	2.09
0.702	2.54
0.710	2.51
0.730	2.33
0.750	2.26
0.772	2.47
0.802	2.80
0.819	2.95

For these data, the following values have been calculated:

$$\begin{array}{ccc} n = 13 & \sum xy = 21.6083 & \sum(x-\bar{x})(y-\bar{y}) = 0.207328 \\ \sum x = 9.188 & \sum x^2 = 6.544592 & \sum(x-\bar{x})^2 = 0.050796 \\ \sum y = 30.28 & \sum y^2 = 71.7122 & \sum(y-\bar{y})^2 = 1.183092 \end{array}$$

(a) Plot a scatter diagram to verify that it is reasonable to assume that the regression of Y on x is linear.

(b) Fit a straight line regression by the method of least squares.

(c) Construct the analysis of variance table, and indicate how to obtain from it:

i. an estimate of the error variance;

ii. a test of the hypothesis that the slope of the regression line is zero.

(d) Find a 95% confidence interval for the slope of the regression line.

210. The table below gives the approximate leaf area (y) of a particular type of tree at age x years.

x	8	11	17	20	23	26
y	14.8 9.0	17.3	20.8 23.7	24.4 28.9	29.3	35.0 33.4

Assuming that the observations are normally distributed with $\mathrm{E}(Y \mid x) = \alpha + \beta x$ and $\mathrm{var}(Y \mid x) = \sigma^2$:

(a) Calculate estimates of α, β and σ^2 using the method of least squares.

(b) Plot the observations and your fitted line.

(c) Find $\mathrm{se}(\hat{\alpha})$ and $\mathrm{se}(\hat{\beta})$.

(d) Test the hypothesis that $\beta = 1.25$.

(e) Test the hypothesis that $\alpha = 0$.

(f) Find a 95% confidence interval for the expectation of Y when $x = 20$.

(g) Find a 95% prediction interval for Y when $x = 20$.

211. The life-lengths (Y) of particular manufactured articles when a quantity (x) of an expensive material is used in the manufacture is approximately normally distributed with $\mathrm{E}(Y \mid x) = \alpha + \beta x$ and $\mathrm{var}(Y \mid x) = \sigma^2$. The observations below are independent values of Y obtained with selected values of x.

x	1	2	3	4	5	6
y	108	111	117	124	126	139

(a) Estimate α and β and give standard errors of your estimates.

(b) Give a 95% prediction interval for the life-length of an article with $x = 5$

212. The following data are observations on Y at specified values of x:

x	0	1	2	3	4
y	14	25	27	32	34
	15	23	29	33	35

The following MINITAB output was obtained using these data:

```
The regression equation is y = 12.1 + 4.85 x
Predictor       Coef       Stdev    t-ratio        P
Constant      12.150       1.651       7.36    0.000
x             4.8500      0.4978       9.74    0.000
s = 2.226       R-sq = 92.2%     R-sq(adj) = 91.3%

Analysis of Variance
source      df       SS        MS         F        P
Regression   1   470.45    470.45     94.92    0.000
Error        8    39.65      4.96
Total        9   510.10

Analysis of variance
source      df         SS        MS         F        P
Between      4     504.60    126.15    114.68    0.000
Within       5       5.50      1.10
Total        9     510.10
                           INDIVIDUAL 95 PCT CI'S FOR MEAN
 LEVEL  N    MEAN  STDEV  ---+---------+---------+---------+---
 1      2  14.500  0.707   (--*-)
 2      2  24.000  1.414                (-*--)
 3      2  28.000  1.414                     (--*--)
 4      2  32.500  0.707                            (-*--)
 5      2  34.500  0.707                               (-*--)
                          ---+---------+---------+---------+---
POOLED STDEV = 1.049       14.0      21.0      28.0      35.0
```

It can be assumed that the observations are independent and normally distributed with equal variances.

(a) What is the correlation between x and y?

(b) Find a 95% confidence interval for $\mathrm{E}(Y \mid x = 3)$:

i. assuming a straight line regression model;

ii. making no assumptions about the regression;

(c) Give an assessment of the goodness of fit of the straight-line regression.

213. The table below gives the weights y kg of a particular type of animal at age x years.

x	y	
0	0.90	1.25
1	2.79	3.00
2	3.28	3.76
3	4.19	3.67
4	3.80	4.56

(a) Assuming that $\mathrm{E}(Y \mid x) = \alpha + \beta e^{-x}$ and $\mathrm{var}(Y \mid x) = \sigma^2$, obtain estimates of α, β and σ^2 using the method of least squares.

(b) Plot the observations and your fitted curve.

214. The following data are observations on a variable y at specified values of a variable x. Assume that the regression of y on x is linear and that the observations are independent and normally distributed with equal variances.

x	11	12	13	14	15	16	17	18	19
y	68.8	61.5	63.3	85.8	78.1	71.5	91.5	82.5	93.2
	55.4	56.0	78.4	68.4	89.1	91.8	92.4	94.2	107.7
	65.6	62.5	74.0	70.4	71.8	71.6	90.1	104.9	94.2
	56.6	66.9	76.3	86.5	72.4	87.2	85.0	94.3	99.4

For these data it is found that: $\bar{x} = 15.0$, $\bar{y} = 79.43$; $\hat{\alpha} = 9.150$, $\mathrm{se}(\hat{\alpha}) = 7.04$; and $\hat{\beta} = 4.685$, $\mathrm{se}(\hat{\beta}) = 0.463$

Further, the analysis of variance yields:

```
              df      SS       MS
Regression     1   5267.8   5267.8
Error         34   1745.9     51.4
Total         35   7013.7
```

(a) i. Test the hypothesis that $\beta = 0$.
ii. Specify a 95% confidence interval for β.

(b) i. What is the standard error of $\bar{y}$?
ii. What is the standard error of $\hat{\mu} = \bar{y} + 2\hat{\beta}$?
iii. Find a 95% confidence interval for $\mathrm{E}(Y \mid x = 17)$.

(c) What is the correlation between x and y?

215. Consider the following bivariate data set:

x	41	52	36	47	53	36	27	58	39	50
y	87	63	73	50	55	76	80	42	70	49

Assume that these data were obtained from a bivariate normal population with correlation ρ, such that

$$\mathrm{E}(Y \mid x) = \alpha + \beta x \quad \mathrm{var}(Y \mid x) = \sigma^2$$

(a) i. Obtain an estimate of β and a 95% confidence interval for β.
ii. Find a 95% confidence interval for $\mu = \alpha + 50\beta$.
iii. Find a 95% prediction interval for an observation on Y with $X = 50$.

(b) i. Find the sample correlation coefficient, r.
ii. Find a 95% confidence interval for ρ.
iii. Test the hypothesis that X and Y are independent.

216. A random sample of fifty observations on (X, Y) produced a sample correlation coefficient, $r = -0.42$. Is this significant evidence that X and Y are not independent? Explain.

217. An investigation of a die-casting process resulted in the data below for x_1 = furnace temperature, x_2 = die close time, and y = temperature difference on the die surface yielded the following data.

x_1	1250	1300	1350	1250	1300	1250	1300	1350	1350
x_2	6	7	6	7	6	8	8	7	8
y	80	95	101	85	92	87	96	106	108

MINITAB output from fitting the multiple linear regression model

$$\eta = \mathrm{E}(y \,|\, x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

is given below:

```
The regression equation is  y = - 200 + 0.210 x1 + 3.00 x2
Predictor      Coef   SE Coef       T      P
Constant    -199.56     11.64  -17.14  0.000
x1          0.2100   0.00864   24.30  0.000
x2          3.0000    0.4321    6.94  0.000
S = 1.05848   R-Sq = 99.1%   R-Sq(adj) = 98.8%

Analysis of Variance
Source          DF      SS      MS       F      P
Regression       2  715.50  357.75  319.31  0.000
Residual Error   6    6.72    1.12
Total            8  722.22
```

(a) The model utility test is a test of $\beta_1 = \beta_2 = 0$, i.e. whether the regression is significantly different from zero. Use the regression MS to test this hypothesis.

(b) Calculate and interpret the 95% confidence interval for β_2.

(c) When x_1=1300 and x_2=7, the standard error of $\hat{\eta}$ is 0.353. Calculate a 95% confidence interval for η when furnace temperature is 1300 and die close time is 7.

(d) Calculate a 95% prediction interval for the temperature difference resulting from a single experimental run with a furnace temperature of 1300 and a die close time of 7.

218. The data (not given here) come from an experiment to investigate how the resistance of rubber to abrasion is affected by the hardness of the rubber and its tensile strength. Each of 30 samples of rubber was tested for hardness (x_1), for tensile strength (x_2) and was then subjected to steady abrasion for a fixed time. The weight loss due to abrasion (y) was measured in grams per hour.

The following output gives the fit of $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$ and the correlations between y, x_1 and x_2:

```
The regression equation is y = 885 - 6.57 x1 - 1.37 x2
Predictor      Coef  SE Coef       T      P
x0           885.16    61.75   14.33  0.000
x1          -6.5708   0.5832  -11.27  0.000
x2          -1.3743   0.1943   -7.07  0.000
S = 36.4893   R-Sq = 84.0%

          y      x1      x2
y            -0.738  -0.298
x1   -0.738          -0.299
x2   -0.298  -0.299
```

(a) Give a 95% confidence interval for β_2.

(b) What is R^2 when only x_1 is fitted?

(c) Does x_2 add anything significant to the model? Explain.

219. An experiment to investigate the relationship between the heat evolved during the hardening of cement and the composition of the cement gave rise to the following:

x_1	x_2	x_3	y
7	26	60	78
11	52	20	104
11	55	22	109
3	71	6	102
2	31	44	74
3	54	22	93
21	47	26	115
1	40	34	83
11	66	12	113
10	68	14	109

where x_1, x_2, x_3 denote percentages by weight of three components, and y denotes the heat evolved in calories per gram of cement.

(a) Assuming the model:
$$\mathrm{E}(Y \mid x_1, x_2, x_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3, \quad \mathrm{var}(Y \mid x_1, x_2, x_3) = \sigma^2$$
use MATLAB to estimate β_0, β_1, β_2, β_3, σ^2.

(b) The above data gave the following results when models involving all subsets of the predictor variables $\{x_1, x_2, x_3\}$ were fitted:

```
     x x x
     1 2 3    R-Sq      S
     X        55.0   10.69
       X      60.2   10.06
         X    61.4    9.90
     X X      97.8    2.50
     X   X    97.4    2.76
       X X    62.0   10.51
     X X X    98.4    2.32

       y       x1       x2       x3
              0.742    0.776   -0.784
x1   0.742             0.177   -0.196
x2   0.776    0.177            -0.965
x3  -0.784   -0.196   -0.965
```

i. Use MATLAB to check the result for the model with x_1 and x_2 as predictors.

ii. Which of x_1, x_2 and x_3 is most highly correlated with y?

iii. Which model is best? Explain your reasoning.

220. Observations on undrained shear strength of sandy soil (y, in kPa), depth (x_1, in m) and water content (x_2, in %) yielded the following data:

x_1	8.9	36.6	36.8	6.1	6.9	6.9	7.3	8.4	6.5	8.0	4.5	9.9	2.9	2.0
x_2	31.5	27.0	25.9	39.1	39.2	38.3	33.9	33.8	27.9	33.1	26.3	37.8	34.6	36.4
y	14.7	48.0	25.6	10.0	16.0	16.8	20.7	38.3	16.9	27.0	16.0	24.9	7.3	12.8

Fitting the model with predictors x_1 and x_2 only gives residual sum-of-squares = 894.95, whereas fitting the complete second-order model, with terms for x_1, x_2, x_1^2, x_2^2 and $x_1 x_2$ gives residual sum-of-squares = 390.64.

Carry out a test of the hypothesis to decide whether at least one of the second-order terms provides useful information about shear strength.

221. The data below come from an experiment to investigate how the resistance of rubber to abrasion is affected by the hardness of the rubber and its tensile strength. Each of 30 samples of rubber was tested for hardness (x_1), for tensile strength (x_2) and was then subjected to steady abrasion for a fixed time. The weight loss due to abrasion (y) was measured in grams per hour.

y	x_1	x_2	y	x_1	x_2
372	45	162	206	55	233
175	61	232	154	66	231
136	71	231	112	71	237
55	81	224	45	86	219
221	53	203	166	60	189
164	64	210	113	68	210
82	79	196	32	81	180
228	56	200	196	68	173
128	75	188	97	83	161
64	88	119	249	59	161
219	71	151	186	80	165
155	82	151	114	89	128
341	51	161	340	59	146
283	65	148	267	74	144
215	81	134	148	86	127

Investigate these data.

222. (a) We wish to maximise yield (y) which depends on two factors x_1 and x_2. The yield (y) at each of the four points (x_1, x_2) = $(50, 40)$, $(50, 41)$, $(52, 40)$, $(52, 41)$ is obtained. A regression model is fitted which results in the following regression equation:

$$\hat{\mu}(x_1, x_2) = 76.3 + 1.50x_1 + 3.00x_2.$$

On an (x_1, x_2) diagram, indicate the four points at which yields were obtained, and the direction of search for maximum yield.

(b) After a series of such experiments, the neighbourhood of maximum yield has been reached. An experiment is performed with yields at a number of points, expressed in terms of (z_1, z_2), where $z_1 = (x_1 - 60)/3$ and $z_2 = (x_2 - 55)/2$. A quadratic model is fitted, which results in the following regression equation:

$$\hat{\mu}(z_1, z_2) = 281.4 + 0.2z_1 - 0.1z_2 - 0.5z_1^2 - 1.0z_2^2 - 1.0z_1z_2.$$

i. Explain how this model would be fitted.

ii. Specify the values of x_1 and x_2 for maximum yield.

223. In an experiment to maximise yield (y) depending on three factors x_1, x_2 and x_3, the neighbourhood of the maximum yield has been reached. It is thought to be around (c_1, c_2, c_3). All factors appear influential. How many points need to be used to estimate the point of maximum yield? Suggest a possible arrangement of points around the point (c_1, c_2, c_3).