Revision exercises 1

R1.1 (a) Consider a system of components that works if at least k out of n independent components operate. Denote such a system by $[k/n]$. Consider systems made up of independent components each of which has reliability 0.9.
Find the reliability of a [3/4] system.

(b) Minor flaws occur randomly in a production process. Suppose that each item coming off the production line has an average of 0.4 flaws.
What is the expected proportion of items free of flaws?

(c) The hazard density function (hdf) of a continuous random variable T is given by

$$h(t) = \frac{f(t)}{1 - F(t)}$$

where f and F denote the pdf and cdf of T.
If T has pdf $f(t) = 1/(1+t)^2 \quad (t > 0)$, find the hdf of T.

(d) X is a random variable with mean 20 and standard deviation 2. Let $Y = \ln X$.
Find an approximate value for the standard deviation of Y.

(e) A random sample of 11 observations is obtained on $Y \stackrel{\mathrm{d}}{=} \mathrm{N}(67, 10^2)$.
Find a 95% probability interval for the sample standard deviation.

(f) Sixty independent trials, each with probability p of success, yielded 48 successes.
Find a 95% confidence interval for p.

(g) Of 100 independent 95% confidence intervals, let Z denote the number of these confidence intervals that contain the true parameter value.
Specify the distribution of Z.

(h) If $W = |Z|$, where $Z \stackrel{\mathrm{d}}{=} \mathrm{N}(0,1)$, show that $c_q(W) = \Phi^{-1}(\frac{1+q}{2})$, where Φ denotes the standard normal cdf. Hence find the median of W.

(i) The log-likelihood function for a particular data set is given by

$$\ln L = -50(\theta - 1)^2.$$

Find $\hat{\theta}$ and $\mathrm{se}(\hat{\theta})$.

(j) Each day for fifty days, a random sample of 16 items from the day's production is selected and measured: the average of the 16 measurements ($\bar{x}$) and the range of the 16 measurements (R) are calculated and recorded each day.
At the end of the fifty days, the average of the daily averages, $\mathrm{av}(\bar{x}) = 14.50$; and the average of the daily ranges, $\mathrm{av}(R) = 7.06$
Determine control limits for an $\bar{x}$-chart.

R1.2 (a) The events A and B are such that $\Pr(B) = 0.2$, $\Pr(A \mid B) = 0.4$ and $\Pr(A \mid B') = 0.1$.

i. Find $\Pr(A)$.

ii. Find $\Pr(B \mid A)$.

iii. Are A and B positively related? Explain.

(b) A production process consists of two stages. Items begin in stage 1. As a result of stage 1, three things can happen: the item is scrapped, with probability 0.1; or the item is re-worked, i.e. sent through stage 1 again, with probability 0.4; or the item moves along to stage 2. As a result of stage 2, three things can happen: the item is returned to stage 1 with probability 0.2, or it is sent through stage 2 again with probability 0.3, or it is deemed satisfactory and complete

Consider this process as a Markov chain with four states: 0 = scrapped, 1 = stage 1, 2 = stage 2, and 3 = complete.

i. Write down the transition probability matrix, P.

Powers of P were computed with the following results:

```
P^4  =  1.0000  0.0000  0.0000  0.0000
        0.1834  0.1166  0.1575  0.5425
        0.0434  0.0630  0.0851  0.8085
        0.0000  0.0000  0.0000  1.0000

P^40 =  1.0000  0.0000  0.0000  0.0000
        0.2188  0.0000  0.0000  0.7812
        0.0625  0.0000  0.0000  0.9375
        0.0000  0.0000  0.0000  1.0000
```

ii. Specify the probability that an item is still in the system after four cycles.

iii. Specify the proportion of items scrapped in the production process.

iv. If an item has reached stage 2, specify the value for its probability of successful completion.

R1.3 (a) Consider the discrete random variable X with pmf given by

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

Show that $\mathbb{E}(X) = \mathrm{var}(X) = 1$.

(b) Suppose that X_1 and X_2 are independent random variables each with the pmf given in (a). Find $\Pr(X_1 = X_2)$.

(c) Suppose that $X_1, X_2, \ldots, X_{100}$ are independent random variables each with the pmf given in (a). Let $S = X_1 + \cdots + X_{50}$ and $T = X_{51} + \cdots + X_{100}$.

i. Find the mean and standard deviation of $S - T$.

ii. Use the central limit theorem to obtain an approximate value for $\Pr(S = T)$.

R1.4 A random sample of 100 observations is obtained from a Normally distributed random variable $X \stackrel{d}{=} \mathrm{N}(60, 10^2)$.

(a) Specify approximate values you would expect to obtain for each of the following statistics for this sample:

i. the sample mean, $\bar{x}$;

ii. the sample standard deviation, s;

iii. the number of observations less than 50, $\mathrm{freq}(X < 50)$;

iv. the sample upper-quartile, $Q3$;

v. the sample maximum, $x_{(100)}$.

(b) i. Sketch a boxplot that would be not unreasonable for this sample.

ii. Indicate in a sketch, the likely form of a Normal QQ-plot for this sample, showing its important features.

iii. What would the Normal QQ-plot look like if observations less than 50 were censored?

R1.5 (a) A random sample of 400 observations is obtained on $T \stackrel{d}{=} \exp(0.01)$, for which

$$F(t) = 1 - e^{-0.01t} \ (t > 0); \ \text{ and } \ f(t) = 0.01e^{-0.01t} \ (t > 0).$$

i. Write down the mean and standard deviation of the sample mean, $\bar{T}$.
The summary notes may be used.

ii. Specify an approximate 95% probability interval for $\bar{T}$.
Give your answers to one decimal place.

(b) The following is a random sample from a Normal population:

3.0, 5.0, 6.0, 7.0, 9.0.

i. Verify that $\bar{x} = 6.0$ and $s^2 = 5.0$.

ii. Find a 95% confidence interval for μ.

iii. Find a 95% confidence interval for σ.

iv. Find a 95% prediction interval for X.
Give your answers to two significant figures.

R1.6 (a) A following sample is obtained on $X \stackrel{d}{=} \mathrm{N}(\mu, \sigma^2)$ to test the hypothesis μ=50 against $\mu \neq 50$ using a significance level of 0.05.

54.0, 45.0, 39.9, 41.5, 55.6, 48.8, 36.6, 49.0, 47.4, 45.6, 39.8, 51.3, 34.2, 32.8, 59.3, 36.0.

For this sample $n = 16$, $\bar{x} = 44.8$ and $s = 8.0$.

i. Show that $t = -2.60$, specify the appropriate critical value and hence show that H_0 is rejected.

ii. Specify the p-value for this test.

(b) Suppose the above sample represents results for the strength of the adhesion for glued tiles. You are required to report to management on the mean strength of the adhesion. In particular, management is concerned that the mean should be no less than 50. Test the hypothesis $\mu = 50$ vs $\mu < 50$.

Write a brief statement summarising your conclusions for management.

R1.7 (a) Independent random samples are obtained from Normally distributed populations, $X_1 \stackrel{d}{=} \mathrm{N}(\mu_1, \sigma^2)$ and $X_2 \stackrel{d}{=} \mathrm{N}(\mu_2, \sigma^2)$, with the following results:

$n_1 = 15$; $\bar{x}_1 = 25.0$, $s_1^2 = 60.0$;
$n_2 = 10$; $\bar{x}_2 = 31.5$, $s_2^2 = 44.7$.

It is assumed that the population variances are equal, and so the sample variances are pooled to give $s^2 = 54.0$.

i. Explain how this pooled variance is obtained.

ii. Find a 95% confidence interval for $\mu_1 - \mu_2$.

(b) Determinations of a strength measure after using three treatments (P, Q and R) were as follows:

		mean	var
[1] treatment P	5, 4, 10, 7	6.5	7.0
[2] treatment Q	12, 8, 9, 15	11.0	10.0
[3] treatment R	15, 9, 12, 14	12.5	7.0

For these data $\sum(y - \bar{y})^2 = 150$.

i. Show that $s^2 = 8$ and hence, or otherwise, derive the analysis of variance table for the above data.

ii. Test the hypothesis H_0: $\mu_P = \mu_Q = \mu_R$ giving an approximate P-value.

iii. Find a 95% confidence interval for the effect of treatment R relative to treatment Q, i.e., $\mu_R - \mu_Q$.

iv. State the conclusions you reach from your analysis.

R1.8 (a) Twenty-four experimental units are available in blocks of six. It is required to run an experiment to compare three treatments: a control C, treatment A and treatment B. Give an appropriate assignment of treatments to the experimental units. Explain your method.

(b) The experiment described in (a) is carried out, with results as indicated in the table below:

	C	A	B
B_1	9, 12	23, 19	12, 15
B_2	5, 8	16, 18	11, 14
B_3	10, 14	23, 20	15, 20
B_4	10, 12	23, 18	19, 14
(av)	10.0	20.0	15.0

This yielded the following incomplete analysis of variance table:

```
source      df   SS   MS
blocks      **  400  ***
treatments  **  ***  ***
error       **  ***    5
total       **  574
```

i. Complete this analysis of variance table.

ii. Assuming an additive model with independent normally distributed errors having equal variances, test the significance of the treatment effects.

iii. Give an estimate of $\mu_A - \mu_C$, and the standard error for your estimate.

R1.9 The table below gives the results of one replicate of a 2^4 experiment with factors A, B, C and D. Some relevant computer output is also given.

```
     y  A  B  C  D
 1   4  0  0  0  0
 2   8  0  0  0  1
 3   4  0  0  1  0
 4  16  0  0  1  1
 5   2  0  1  0  0
 6   8  0  1  0  1
 7   6  0  1  1  0
 8  14  0  1  1  1
 9   8  1  0  0  0
10  14  1  0  0  1
11  12  1  0  1  0
12  20  1  0  1  1
13  10  1  1  0  0
14  16  1  1  0  1
15  12  1  1  1  0
16  22  1  1  1  1
```

```
anova
source  df   SS
A        1  169
B        1    1
C        1   81
D        1  225
AB       1    4
AC       1    0
AD       1    0
BC       1    0
BD       1    0
CD       1   16
ABC      1    1
ABD      1    1
ACD      1    1
BCD      1    1
ABCD     1    4
Error    0
Total   15  504
```

```
         av
A0     7.75
A1    14.25

C0     8.75
C1    13.25

D0     7.25
D1    14.75

C0D0   6.00
C0D1  11.50
C1D0   8.50
C1D1  18.00
```

i. How might a half-normal plot be used to indicate which interactions might be non-zero? Indicate in a rough sketch its appearance in this case.

ii. Show that A, C, D and CD are significant.

iii. Show that the estimate of the error standard deviation, $s = \sqrt{\frac{13}{11}}$.

iv. Give an estimate of, and a 95% confidence interval for the effect of A. Use $s \approx 1.1$.

v. Sketch a cell-mean plot to indicate the interaction between C and D. Write a sentence describing the interaction.

R1.10 The table below gives the corresponding values of variables x and y.

x	5	6	7	8	10	11	12	13	14	14
y	20	22	26	24	34	30	38	36	32	38

For these data, the following statistics were calculated:

$n = 10$, $\bar{x} = 10$, $\bar{y} = 30$; $\sum(x - \bar{x})^2 = 100$, $\sum(x-\bar{x})(y-\bar{y}) = 180$, $\sum(y-\bar{y})^2 = 400$.

i. Assuming that $\mathbb{E}(Y \mid x) = \alpha + \beta x$ and $\text{var}(Y \mid x) = \sigma^2$, obtain estimates of α and β using the method of least squares.
ii. Show that $s^2 = 9.5$ and hence obtain $\text{se}(\hat{\beta})$.
iii. Plot the observations and your fitted line.
iv. Find the sample correlation, and give an approximate 95% confidence interval for the population correlation.

Revision exercises 2

R2.1 (a) If $\Pr(G) = 0.4$ and $\Pr(H) = 0.2$, find $\Pr(G \cup H)$ if

i. G and H are mutually exclusive;
ii. G and H are independent;
iii. $\Pr(G \mid H) = 0.8$.

(b) The events A, B and C are independent with probabilities 0.5, 0.6 and 0.8 respectively.

i. Find the probability that at least one of the events A, B and C occur.
ii. Find the probability that at least two of the events A, B and C occur.

R2.2 Consider the following sampling plan. Test a random sample of 100 items chosen from a lot containing a very large number of items, and accept the lot if at most five defective items are found.

Use an appropriate Poisson approximation to sketch the graph of the OC curve for this sampling plan, for $0 \leqslant p \leqslant 0.1$, where p denotes the lot defective proportion.

R2.3 A production process consists of three stages. Items begin in stage 1, and as a result of each cycle, three things can happen: the item is scrapped, with probability 0.05; or the item is reworked, i.e. sent through the same stage again, with probability 0.25; or the item moves along to the next stage. Consider this process as a Markov chain with five states: 0 = item scrapped, 1 = stage 1, 2 = stage 2, 3 = stage 3 and 4 = complete.

(a) Write down the transition probability matrix, P.

(b) Given that

```
P^5  =  1.0000         0         0         0         0
        0.1799    0.0010    0.0137    0.0766    0.7289
        0.1279         0    0.0010    0.0137    0.8575
        0.0666         0         0    0.0010    0.9324
             0         0         0         0    1.0000
```

i. Specify the probability that an item is still in the system after five cycles.
ii. Give an approximate value for the proportion of items scrapped in the production process.
iii. If an item has reached stage 2, give an approximate value for its probability of successful completion.

R2.4 (a) Minor flaws occur randomly in the production process. Suppose that each sprocket coming off the production line has an average of 1.2 minor flaws. What proportion of items are free of minor flaws?

(b) Sprockets are checked for major flaws before being used. 90% pass the test. The other 10% are scrapped. Of 18 sprockets tested, what is the probability that at most two are scrapped?

(c) The sprockets must fit in their place in the assembly, so their measurement needs to be within 0.20 mm of specifications. If the deviation of this measurement from specifications is roughly normal with mean –0.05 mm and standard deviation 0.10 mm, find approximately the proportion of sprockets rejected on this measurement check.

R2.5 A random sample of sixteen observations is obtained from a population having a standard normal distribution, i.e. $\mathrm{N}(0, 1)$. Find approximate values for:

i. the probability that the sample mean is greater than 0.5;

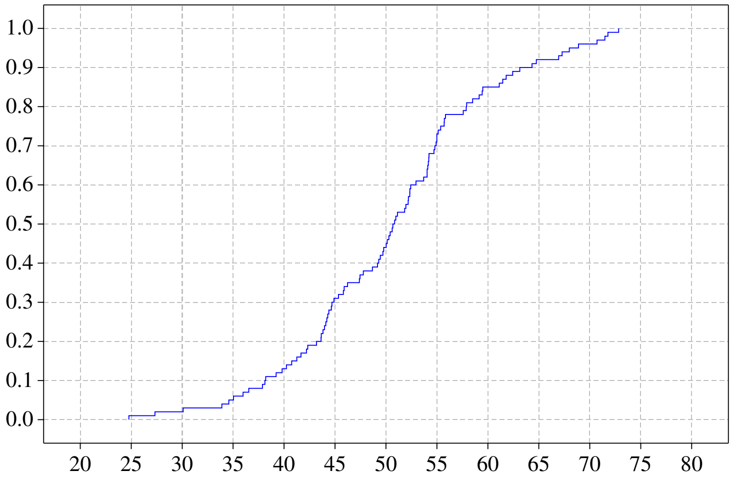
ii. the probability that the sample variance is greater than 2;

Consider the distribution of $X_{(16)}$, the maximum of a random sample of sixteen observations from a standard normal distribution.

iii. Use the Statistical Tables to specify the mean of $X_{(16)}$.

iv. Use the Statistical Tables to show that $\Pr(X_{(16)} \leqslant 1.2816) = 0.1853$.

R2.6 The diagram below gives the sample cdf for a random sample of 100 observations on X.



i. Find the sample median.

ii. Sketch a boxplot for this sample.

iii. Draw a rough graph of what you think the population pdf might look like.

iv. Give a rough estimate of the population standard deviation.

v. Explain briefly how a sample cdf relates to a probability plot.

R2.7 (a) The following is a random sample of ten observations on $Y \stackrel{\mathrm{d}}{=} \mathrm{Pn}(\lambda)$:
$\{$`7,12,9,2,4,7,8,11,4,6`$\}$.
For this sample $\sum y = 70$ *and* $\sum y^2 = 580$.
Give an estimate of λ and its standard error.

(b) Sixty independent trials, each with probability p of success, yielded 18 successes. Find a 95% confidence interval for p.

(c) A random sample of $n = 15$ observations on $X \stackrel{\mathrm{d}}{=} \mathrm{N}(\mu, \sigma^2)$ has sample mean $\bar{x} = 50.0$ and sample variance $s^2 = 60.0$.

i. Find a 90% confidence interval for μ.

ii. Find a 90% prediction interval for X.

Give your answers to one decimal place.

(d) Each day for twenty days, a random sample of ten items from the day's production is selected and carefully measured: the average of the ten measurements ($\bar{x}$) and the variance of the ten measurements (s^2) are calculated and recorded each day.

At the end of the twenty days, the average of the daily averages, $\mathrm{av}(\bar{x}) = 8.2$; and the average of the daily ranges, $\mathrm{av}(s^2) = 1.6$

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

R2.8 A random sample of n observations is obtained on X which has pdf and cdf given by

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

(a) Find the maximum likelihood estimate of θ and an expression for its standard error.

(b) i. Show that the q-quantile of X is such that $\theta \ln(1+c_q) = -\ln(1-q)$.

ii. Let $x_{(k)}$ denote the kth order statistic for this sample on X; and define $z_k = -\ln(1-\frac{k}{n+1})$. Explain how a plot of $\ln(1+x_{(k)})$ against z_k yields: (A) a check of the distributional assumption; and (B) an estimate of θ.

iii. Indicate in a *rough* sketch what such a plot might look like if the model is correct.

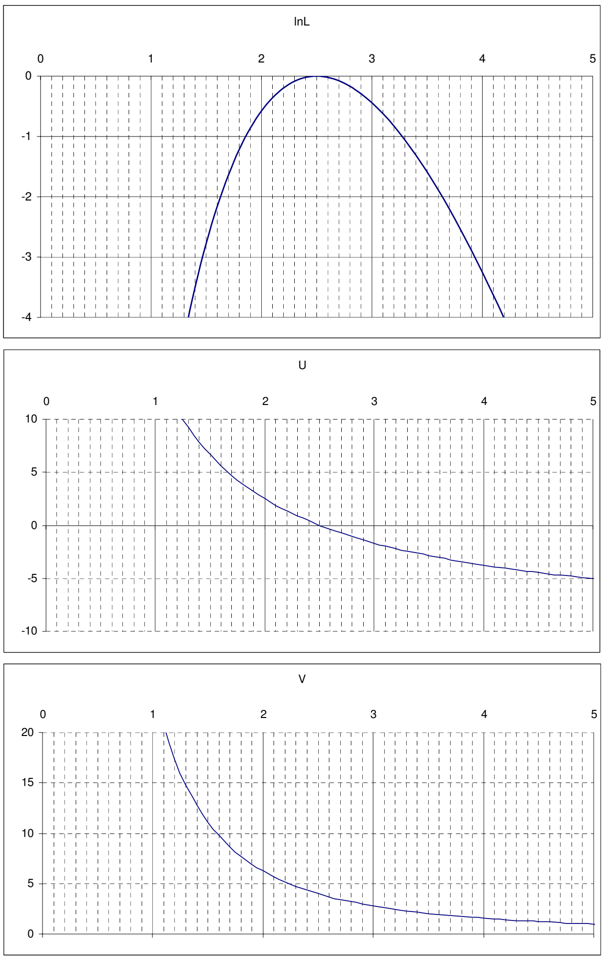
R2.9 *For a random sample of n observations taken from a normal population, the variance of the sample variance, S^2, is given by* $\text{var}(S^2) = \frac{2\sigma^4}{n-1}$.

Use this result to show that an approximate expression for the standard error of the sample standard deviation s is given by $\text{se}(s) \approx \frac{s}{\sqrt{2(n-1)}}$.

Hence find an approximate 95% confidence interval for σ based on a random sample of $n = 33$ observations from a normal population, for which $s = 16$.

The exact 95% confidence interval in this case is $(12.9 < \sigma < 21.2)$. Explain how this exact confidence interval is obtained.

R2.10 A random sample yields the following graphs of the log likelihood, $\ln L$; the score function, $U = \frac{\partial \ln L}{\partial \theta}$; and the observed information function, $V = -\frac{\partial^2 \ln L}{\partial \theta^2}$.



Use the above graphs to determine

i. the maximum likelihood estimate of θ;

ii. an approximate 95% confidence interval for θ;

iii. an approximate standard error for $\hat{\theta}$.

R2.11 Consider a sequence of independent trials with probability of success θ. We wish to test H_0: $\theta = 0.05$ vs H_1: $\theta = 0.1$.

(a) Define the size and the power of a statistical test.

(b) Given the following MATLAB output, which gives the cdf of a binomial distribution, specify a test, based on n=300 observations which has size $\leqslant 0.05$ and power $\geqslant 0.95$ to test H_0: $\theta = 0.05$ vs H_1: $\theta = 0.1$.

```
>> x = [19 20 21 22 23]
x = 19    20    21    22    23
>> binocdf(x,300,0.05)
ans = 0.8810    0.9224    0.9514    0.9708    0.9832
>> binocdf(x,300,0.1)
ans = 0.0171    0.0287    0.0458    0.0699    0.1024
```

(c) In the sequential likelihood ratio test to test H_0: $\theta = \theta_0$ vs H_1: $\theta = \theta_1$, we define $U_k = \ln\left(\frac{L_k(\theta_1)}{L_k(\theta_0)}\right)$, where $L_k(\theta)$ denotes the likelihood of the data set consisting of the first k observations. Then, we accept H_1 if $U_k \geqslant 3$, accept H_0 if $U_k \leqslant -3$, and continue sampling otherwise.

i. Give a brief explanation of why the cut-off values ± 3 correspond roughly to size = 0.05 and power = 0.95.

ii. In the case of testing H_0: $\theta = 0.05$ vs H_1: $\theta = 0.1$ for independent trials, indicate why $U_k \approx 0.75x_k - 0.05k$, where x_k denotes the number of successes in k trials.

iii. Verify that, if we obtained 20 successes in 200 trials, then this sequential test would mean that H_1 would be accepted.

iv. Comment on the advantages and disadvantages of a sequential test.

R2.12 PQR-Co is concerned with the quality of its major product, the deconvolving sprocket. In particular, the springiness of the sprocket must be at least 0.2. To determine the springiness conclusively, a destructive test needs to be used. But this is clearly pointless for products to be shipped. In the past it has tested a random sample of 1% of the products shipped. A new non-destructive testing procedure is proposed which is inexpensive and could be applied to all products shipped. However, it is felt that this new testing procedure over-estimates the springiness. To test this, a series of sprockets is tested using each of the methods: it is first tested by the new method (N) and then using the destructive test (D), after which no further testing is possible. These results are given below.

	N	D
1	0.280	0.228
2	0.228	0.227
3	0.271	0.242
4	0.217	0.197
5	0.225	0.209
6	0.247	0.227
7	0.209	0.190
8	0.226	0.210
9	0.235	0.215
10	0.241	0.239

Describe how you would test the null hypothesis that $\mu_N = \mu_D$ against the alternative $\mu_N > \mu_D$. Give a rough analysis, without actually performing the test.

R2.13 Independent samples are obtained from normally distributed populations, $X_1 \stackrel{d}{=} N(\mu_1, \sigma_1^2)$ and $X_2 \stackrel{d}{=} N(\mu_2, \sigma_2^2)$, with the following results:

$$n_1 = 8; \quad \bar{x}_1 = 80, \quad s_1^2 = 40;$$
$$n_2 = 8; \quad \bar{x}_2 = 50, \quad s_2^2 = 32.$$

i. Find a 95% confidence interval for σ_1/σ_2; and verify that the hypothesis $\sigma_1^2 = \sigma_2^2$ would be accepted.

ii. Specify the pooled variance estimate based on both samples.

iii. Using the pooled variance estimate, obtain a 95% confidence interval for $\mu_1 - \mu_2$.

R2.14 (a) Twenty experimental units are available in blocks of four. It is required to run an experiment to compare four treatments. Give an appropriate assignment of treatments to plots. Explain your method.

(b) The experiment described in (a) is carried out, with results as indicated in the table below:

	T_1	T_2	T_3	T_4	(sum)
B_1	16	25	9	30	80
B_2	9	17	14	20	60
B_3	4	21	16	19	60
B_4	17	28	24	31	100
B_5	14	29	27	30	100
(sum)	60	120	90	130	400

This yielded the following incomplete approximate analysis of variance table:

```
source       df     SS    MS
blocks       **    400   ***
treatments   **    ***   ***
error        **    ***    15
total        **   1180
```

i. Complete this analysis of variance table.

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

ii. test the significance of the treatment effects.

The experiment above is actually a five replicate 2^2 factorial experiment, with $T_1 = P_0Q_0$, $T_2 = P_1Q_0$, $T_3 = P_0Q_1$ and $T_4 = P_1Q_1$. This allows the split up of the treatment sum of squares into components due to P, Q and PQ: $SS_P = 500$, $SS_Q = 80$ and $SS_{PQ} = 20$.

iii. Show that the effect of P is highly significant.

iv. Give an estimate of the effect of P, and the standard error for your estimate.

R2.15 The table below gives the results of one replicate of a 2^4 experiment with factors P, Q, R and S. Some relevant computer output is also given.

```
      y    P  Q  R  S
 1  49.0   0  0  0  0
 2  56.2   0  0  0  1
 3  49.8   0  0  1  0
 4  49.0   0  0  1  1
 5  52.8   0  1  0  0
 6  62.2   0  1  0  1
 7  51.8   0  1  1  0
 8  57.2   0  1  1  1
 9  52.6   1  0  0  0
10  55.4   1  0  0  1
11  50.2   1  0  1  0
12  56.6   1  0  1  1
13  57.2   1  1  0  0
14  63.2   1  1  0  1
15  57.0   1  1  1  0
16  61.6   1  1  1  1
```

```
     av.y
P0  53.50
P1  56.73

Q0  52.35
Q1  57.88

R0  56.08
R1  54.15

S0  52.55
S1  57.68
```

```
anova-1
source df     SS
P       1   41.603
Q       1  122.103
R       1   14.823
S       1  105.063
PQ      1    1.102
PR      1    5.522
PS      1    0.122
QR      1    0.003
QS      1    6.003
RS      1    6.003
PQR     1    0.062
PQS     1    3.063
PRS     1   12.602
QRS     1    0.062
PQRS    1    5.062
Error   0
Total  15  323.198
```

```
anova-2
source  df     SS       MS      F      P
P        1   41.60    41.60  11.55  0.006
Q        1  122.10   122.10  33.91  0.000
R        1   14.82    14.82   4.12  0.067
S        1  105.06   105.06  29.18  0.000
Error   11   39.61     3.60
Total   15  323.20
```

i. Indicate the steps in the analysis of this experiment. Include answers to the following questions: How might a half-normal plot be used to indicate which interactions might be non-zero? How is the second analysis of variance obtained from the first? Which effects are significant? What is the estimate of the error variance?

ii. Give an estimate of, and a 95% confidence interval for the effect of P.

R2.16 (a) The table below gives the values of an independent variable x and a dependent variable y.

x	0	1	2	3	4
y	90	87	82	74	67

For these data, the following statistics were calculated:

$\bar{x} = 2$, $\bar{y} = 80$; $\sum(x - \bar{x})^2 = 10$, $\sum(x - \bar{x})(y - \bar{y}) = -59$, $\sum(y - \bar{y})^2 = 358$.

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

ii. Find s^2 and hence obtain $\mathrm{se}(\hat{\beta})$.

iii. Plot the observations and your fitted curve.

(b) A random sample of $n = 50$ observations are obtained on (X, Y). For this sample, it is found that $\bar{x} = \bar{y} = 50$, $s_x = s_y = 10$ and the sample correlation $r_{xy} = -0.5$.

i. Indicate, with a rough sketch, the general nature of the scatter plot for this sample.

ii. On your diagram, indicate the regression line for the regression of y on x.

iii. Give an approximate 95% confidence interval for the population correlation.

Revision exercises 3

R3.1 **(a)** The events A and B are such that $\Pr(A) = 0.5$, $\Pr(B) = 0.4$ and $\Pr(B \mid A) = 0.6$. Find $\Pr(B \mid \bar{A})$.

(b) Items from a production line are classified as having no fault, having a minor fault or having a major fault. On average, 90% have no fault, 8% have a minor fault and 2% have a major fault.

A 'fault detector' is such that the probability that it signals a fault is 0.01 for items with no fault; 0.6 for items with a minor fault; and 0.95 for items with a major fault.

i. For what proportion of items is a fault signalled?

ii. Of those items for which a fault is signalled, what proportion have a major fault?

R3.2 **(a)** A production process when in control produces 10% defective items. If a week's production is 400 items, find 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.

(b) Consider the following sampling plan. Test a random sample of 100 items chosen from a lot containing a large number of items, and accept the lot if the number of defective items found is at most two.

Construct the OC-curve for this sampling plan.

R3.3 A production process consists of three stages. As a result of each stage, three things can happen: the item is scrapped; the item is reworked, i.e. sent through the same stage again; or the item moves along to the next stage. The probabilities of these events at each stage are set out in the following table:

	scrapped	reworked	next stage
stage 1	0.1	0.2	0.7
stage 2	0.1	0.2	0.7
stage 3	0	0.6	0.4

Consider this process as a Markov chain with five states: 0 = scrapped, 1 = stage 1, 2 = stage 2, 3 = stage 3 and 4 = complete.

i. Write down the transition probability matrix, P.

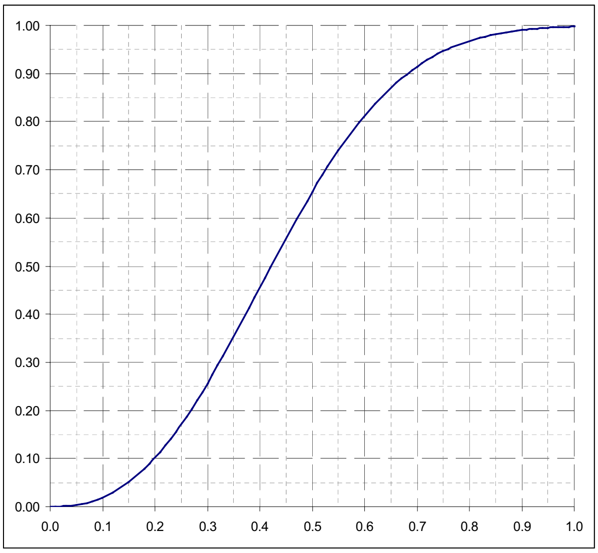
ii. If N denotes the number of cycles (i.e. stages and repeated stages) required to complete an item, specify approximately $\mathbb{E}(N)$.

iii. Given that

$$P^5 = \begin{bmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.234 & 0.000 & 0.006 & 0.227 & 0.533 \\ 0.125 & 0.000 & 0.000 & 0.136 & 0.739 \\ 0.000 & 0.000 & 0.000 & 0.078 & 0.922 \\ 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \end{bmatrix}$$

a. Specify the probability that an item is still in the production process after five cycles.
b. Specify the probability that an item is complete after five cycles.
c. Give an approximate value for the proportion of items scrapped in this production process.

R3.4 **(a)** The graph below represents the cdf of a random variable W



Use the graph to obtain approximate values for

i. $\Pr(0.2 < W < 0.5)$ and $\Pr(W > 0.5)$;
ii. the median and quartiles of W.

Draw a rough sketch of the graph of the pdf of W.

(b) Suppose that T is a continuous random variable with pdf given by

$$f(t) = 12t^2(1-t) \quad (0 < t < 1).$$

Find the mean and standard deviation of T.

R3.5 **(a)** Suppose that X and Y are independent random variables, which are such that $X \stackrel{d}{=} \mathrm{N}(10, 4^2)$ and $Y \stackrel{d}{=} \mathrm{N}(12, 3^2)$. Find

i. $\Pr(X > 12)$;
ii. $\Pr(X > Y)$.

(b) If $Z \stackrel{d}{=} \mathrm{Pn}(25)$, find $\Pr(Z \geqslant 33)$.

(c) If $\mathbb{E}(U) = 10$ and $\mathrm{sd}(U) = 5$, specify approximate values for the mean and standard deviation of $\ln U$.

R3.6 The following is a random sample of 10 observations on the integer-valued random variable Y

`6, 4, 9, 3, 2, 4, 11, 8, 6, 7.`

For this sample $\sum y = 60$ and $\sum y^2 = 432$.

(a) Find:

i. the second order statistic, $y_{(2)}$;
ii. the sample mean;
iii. the sample standard deviation;
iv. the sample median;
v. the sample quartiles.

(b) If $Y \stackrel{\text{d}}{=} \text{Pn}(\lambda)$, give an estimate of λ and its standard error.

R3.7 **(a)** A sequence of $n = 30$ independent trials yields $x = 9$ successes.
Use the tables to specify a 95% confidence interval for p, the probability of success.

(b) A random sample of $n = 15$ on $X \stackrel{\text{d}}{=} \text{N}(\mu, \sigma^2)$ yields sample mean $\bar{x} = 160$ and sample variance $s^2 = 60$.

i. Find a 95% confidence interval for μ.
ii. Find a 95% prediction interval for X.

(c) The log-likelihood for a data set is given by

$$\ln L = 200 \ln \theta - 25\theta^2.$$

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

R3.8 **(a)** A random sample of twenty-five observations is obtained on $X \stackrel{\text{d}}{=} \text{N}(\mu, 10^2)$. The sample mean for this sample is $\bar{x} = 46.4$. Test the null hypothesis H_0: $\mu = 50$ against the alternative H_1: $\mu < 50$. Specify the P-value.

(b) Each day for twenty days, a random sample of 25 items from the day's production is selected and carefully measured: the average of the 25 measurements ($\bar{x}$) and the range of the 25 measurements (R) are calculated and recorded each day.

At the end of the twenty days, the average of the daily averages, $\bar{\bar{x}} = 22.0$; and the average of the daily ranges, $\bar{R} = 7.86$.

It is assumed that the measurements are approximately normal with mean μ and variance σ^2.

i. Explain why $\bar{R} \approx 3.93\sigma$.
ii. Determine control limits for an $\bar{x}$-chart.

R3.9 Independent samples are obtained from normal populations $X_1 \stackrel{\text{d}}{=} \text{N}(\mu_1, \sigma_1^2)$ and $X_2 \stackrel{\text{d}}{=} \text{N}(\mu_2, \sigma_2^2)$, with the following results:

$$n_1 = 8; \quad \bar{x}_1 = 80, \quad s_1^2 = 40;$$
$$n_2 = 8; \quad \bar{x}_2 = 50, \quad s_2^2 = 32.$$

i. Find a 95% confidence interval for σ_1/σ_2; and verify that the hypothesis $\sigma_1^2 = \sigma_2^2$ would be accepted.

ii. Specify the pooled variance estimate based on both samples; and state its distribution under the assumption that $\sigma_1^2 = \sigma_2^2$.

iii. Using the pooled variance estimate, obtain a 95% confidence interval for $\mu_1 - \mu_2$, and specify the value of the t-statistic used to test $\mu_1 = \mu_2$.

R3.10 Random samples were obtained on each of four normal populations having equal variances, σ^2. The samples contained $n_1 = 4, n_2 = 3, n_3 = 5$ and $n_4 = 5$ observations respectively.

i. Complete the following analysis of variance table derived from the above samples:

	df	SS	MS	F
between samples				
within samples			25	
total		625		

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

iii. Give an estimate of σ^2, and a 95% confidence interval for σ^2.

iv. If $\bar{x}_1 = 31$ specify a 95% confidence interval for μ_1.

v. Specify the standard error of $\hat{\mu}_3 - \hat{\mu}_4$.

R3.11 **(a)** An experiment to compare the effect of three treatments is to be conducted using eighteen plots. Six of the plots are classed as "good quality", six as "moderate quality" and six as "poor quality". Describe in detail how you would decide which treatments would be allocated to which plots.

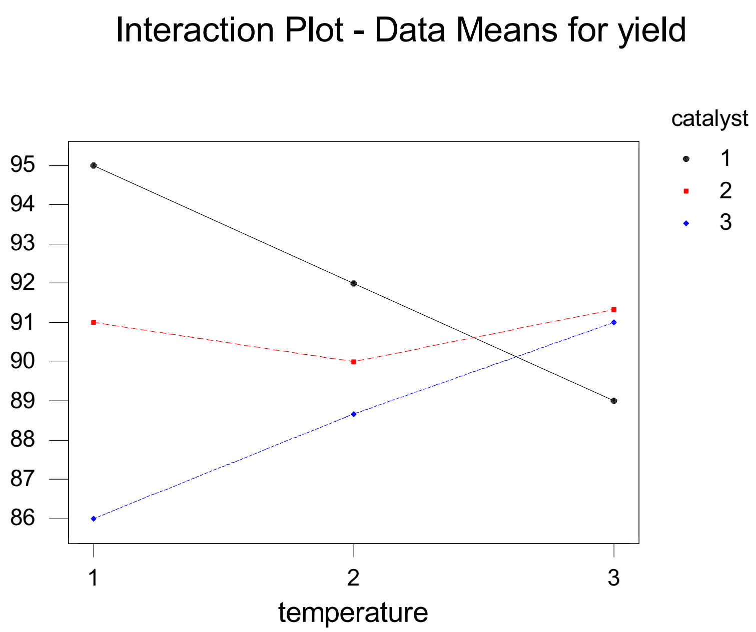
This experiment is carried out giving the following analysis of variance table:

	df	SS	MS
blocks	..	48	..
treatments	..	24	..
error	..	39	..
total	..	111	

i. Test the significance of the treatment effects.

ii. Specify the standard error of the estimate of the difference in effects of treatments 1 and 2.

(b) The yield of a chemical process is observed at three temperature levels (1=low, 2=medium and 3=high) for each of three 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, $s^2 = 1$. Indicate the number of degrees of freedom for each of the components in the standard analysis of variance for this situation; and indicate whether the corresponding F-tests are likely to be significant for these data, giving reasons for your answers.

R3.12 A study was conducted to investigate the relationship between strength (y) and density (x) of a particular material. Seventeen specimens were tested and for these data the following statistics were calculated:

$$\bar{x} = 30, \quad \bar{y} = 50;$$

$$\sum(x-\bar{x})(y-\bar{y}) = 200, \ \sum(x-\bar{x})^2 = 400, \ \sum(y-\bar{y})^2 = 400.$$

i. Evaluate the sample correlation coefficient r and give a *rough* 95% confidence interval for ρ.

ii. Fit the straight line regression model $y = \alpha + \beta x + e$; i.e. give estimates of α and β.

iii. Roughly sketch a possible scatter plot for these data.

iv. Complete the following analysis of variance table:

	df	SS	MS	F
regression	..	..	..	..
residual	..	..	20	
total	..	..		

Hence, or otherwise, test the hypothesis $\beta = 0$.