Questions 1-1 and 1-2 refer to the following information:
In a factory, machines U, V and W produce 60%, 30% and 10% of the total output respectively. Of these outputs, 2%, 3% and 4% are defective.

1-1 The percentage of defective items produced at the factory is

A. 9.0%
B. 3.0%
C. 2.8%
D. 2.5%
E. 2.2%

1-2 An item chosen at random from the combined output of a day's production is found to be defective. The probability that it was produced by machine W is equal to:

A. 0.10
B. 0.16
C. 0.20
D. 0.33
E. 0.40

1-3 The Markov chain with state space $\{0, 1, 2\}$ has transition probability matrix P, where

$$P = \begin{bmatrix} 0.70 & 0.20 & 0.10 \\ 0.20 & 0.60 & 0.20 \\ 0.20 & 0.30 & 0.50 \end{bmatrix} \quad P^2 = \begin{bmatrix} 0.55 & 0.29 & 0.16 \\ 0.30 & 0.46 & 0.24 \\ 0.30 & 0.37 & 0.33 \end{bmatrix} \quad P^4 = \begin{bmatrix} 0.44 & 0.35 & 0.21 \\ 0.37 & 0.39 & 0.24 \\ 0.37 & 0.38 & 0.25 \end{bmatrix}$$

$\Pr(X_4 = 2 \mid X_2 = 2)$ is equal to:

A. 0.16
B. 0.21
C. 0.25
D. 0.30
E. 0.33

1-4 X and Y are independent random variables with $\mathrm{E}(X) = 20$, $\mathrm{sd}(X) = 4$; and $\mathrm{E}(Y) = 30$, $\mathrm{sd}(Y) = 3$. If $Z = X + Y + 2$, then the standard deviation of Z is

A. 9
B. $\sqrt{29}$
C. 7
D. $\sqrt{27}$
E. 5

1-5 Suppose that T is normally distributed, i.e. $T \stackrel{\mathrm{d}}{=} \mathrm{N}(200, 20^2)$.
Because the upper tail of the distribution of T is $(\cdots\cdots)$ than exponential,
the hazard function of T is $(\cdots\cdots)$.
The missing words are

A. (longer), (increasing);
B. (longer), (decreasing);
C. (shorter), (increasing);
D. (shorter), (decreasing);
E. (about the same), (approximately constant).

1-6 If $Z \stackrel{\text{d}}{=} \text{N}(0, 1)$ then $\Pr(Z > 1 \mid Z > 2)$ is equal to

A. 1;

B. $\dfrac{0.8413}{0.9772}$;

C. $\dfrac{0.0228}{0.1587}$;

D. $\dfrac{0.0228}{0.8413}$;

E. $\frac{1}{2}$.

1-7 A probability interval, a confidence interval and a prediction interval are statements about:

A. a parameter, a statistic, and an observation, respectively;

B. a statistic, an observation, and a parameter, respectively;

C. a statistic, a parameter, and an observation, respectively;

D. a parameter, an observation, and a statistic, respectively;

E. an observation, a statistic, and a parameter, respectively.

1-8 A 95% confidence interval for a mean μ of a random variable with known variance, based on the normal distribution, is found to be $(15.1, 15.9)$. Without any further calculations, the P-value for test of H_0: $\mu{=}15$ versus H_1: $\mu{\neq}15$ is such that

A. $P > 0.05$;

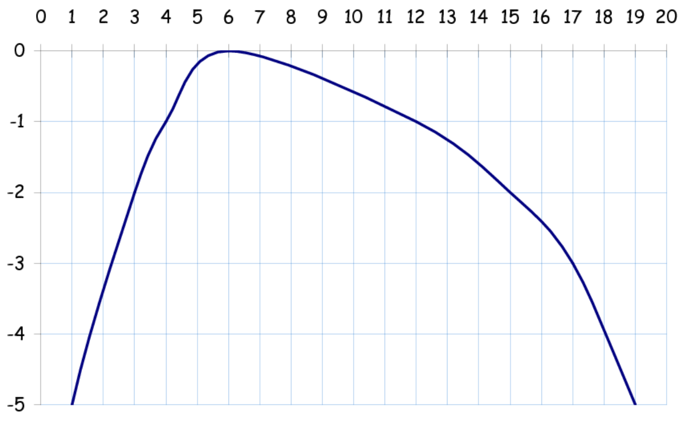
B. $0.01 \leq P < 0.05$;

C. $0.001 \leq P < 0.01$;

D. $P < 0.001$;

E. cannot be determined without knowing the sample size.

1-9 The graph of the relative likelihood function, $\text{RLL}(\theta)$ is shown in the diagram below.



An approximate 95% confidence interval for θ is given by:

A. $4 < \theta < 8$;

B. $4 < \theta < 12$;

C. $3 < \theta < 9$;

D. $3 < \theta < 15$;

E. $1 < \theta < 19$.

1-10 In a $4{\times}4$ Latin square experiment with one-replicate, the following sums of squares for the analysis of variance were obtained:

	SS
rows	300
columns	600
treatments	120
error	60
total	1110

From the analysis of variance the F-ratio for treatments is such that

A. $F = 4$ and so the treatments are not significant;

B. $F = 4$ and so the treatments are significant;

C. $F = 2$ and so the treatments are not significant;

D. $F = 2$ and so the treatments are significant;

E. $F = 1$ and so the treatments are not significant.

2. (a) If $\Pr(F) = 0.2$, $\Pr(G) = g$ and $\Pr(F \cup G) = 0.8$, find g:
   i. if F and G are mutually exclusive;
   ii. if F and G are independent.
   iii. Show that $\Pr(F \mid G) \leqslant 0.25$.

   (b) Items from a production line are checked for major flaws before being used. 90% pass the test. The other 10% are recycled. Of twenty items tested, find the probability that at most two are recycled.

   (c) In a production process, each item has to pass through three stages. At each stage, the probability of successful completion of the stage is 0.8. Of those that fail, half repeat the stage and half are returned to stage 1. All items failing stage 1 must repeat stage 1. Consider the four-state Markov chain describing an item's progress through the production process, with states 1 = stage 1, 2 = stage 2, 3 = stage 3 and 4 = complete.
   i. Write down the transition probability matrix for this Markov chain.
   ii. Suppose that each procedure at each stage takes one hour. Explain how you could obtain, using matrix multiplication, the probability that an item is completed within five hours of processing time.

3. (a) Let $Y = \sum_{i=1}^{160} X_i$, where the X_i are independent and identically distributed random variables, each having probability mass function: $p(0) = 0.2$, $p(1) = 0.6$, $p(2) = 0.2$. Find $\Pr(Y \leq 150)$.

   (b) Let T denote a positive continuous random variable with hazard density function $h(t) = 0.03\sqrt{t},\ t > 0$. Find $\Pr(T > 4)$.

   (c) If $\mathrm{E}(Z) = \theta$ and $\mathrm{var}(Z) = c\theta^2$, find approximate expressions for $\mathrm{E}(\sqrt{Z})$ and $\mathrm{var}(\sqrt{Z})$. For what values of c will the approximation be best?

   (d) Suppose that $X \stackrel{\mathrm{d}}{=} \mathrm{N}(46, 4^2)$.
   i. Find the probability that X exceeds the threshold, $t = 50$.
   ii. Suppose that the threshold is random: $T \stackrel{\mathrm{d}}{=} \mathrm{N}(50, 3^2)$, and that T and X are independent Find the probability that X exceeds the threshold T.

4. (a) Evaluate the sample mean and sample standard deviation for the following data set:

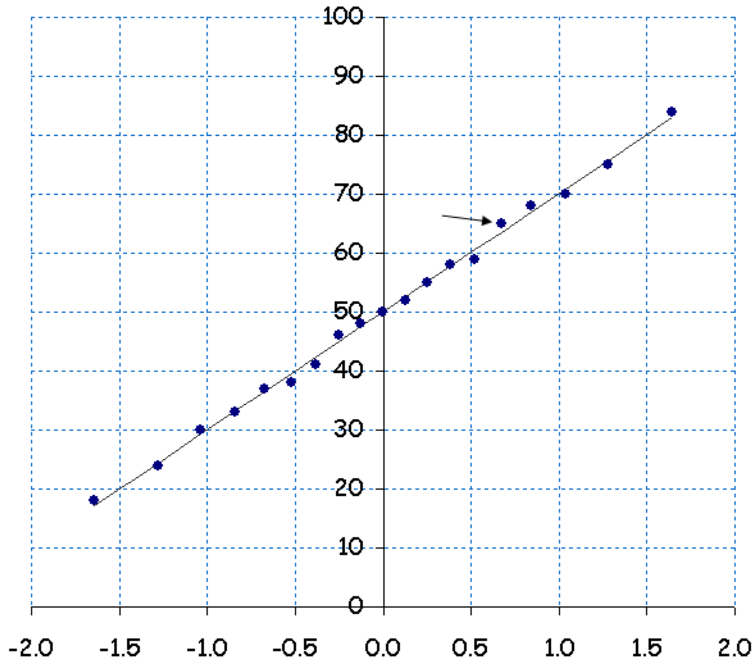
   23.2, 37.4, 29.9, 12.4, 17.0, 43.5, 31.5, 19.6, 22.2, 34.7.

   (b) A random sample of 49 observations is obtained from a normal population, $\mathrm{N}(320, 40^2)$. Specify values you might expect to observed for min, Q1, med, Q3 and max. Hence sketch a "typical" boxplot for such a sample.

   (c) The following is a sample of $n = 19$ observations on $X \stackrel{\mathrm{d}}{=} \mathrm{N}(\mu, \sigma^2)$.

```
84  37  33  24  58  75  55  46  65  59
18  30  48  38  70  68  41  52  50
```

   The graph below is the normal QQ plot for this sample.



   Specify the coordinates of the indicated point, explaining how they are obtained. Use the diagram to obtain estimates of μ and σ.

5. Twelve independent observations are obtained on $X \stackrel{\mathrm{d}}{=} \mathrm{N}(\mu, 5^2)$. To test H_0: μ=40 vs H_1: $\mu \neq 40$, the decision rule is to reject H_0 if $|\bar{X} - 40| > 3$.

   i. Find a 95% probability interval for $\bar{X}$ if $\mu = 40$.

   ii. Find the size of the specified test.

   iii. Find the P-value if $\bar{x} = 44.0$.

   iv. Find the power of the specified test if μ=45.

   v. Find a 95% confidence interval for μ if $\bar{x} = 44.0$.

   vi. Find a 95% prediction interval for X if $\bar{x} = 44.0$.

6. (a) A random sample of ten observations is obtained on an exponentially distributed random variable with unknown mean λ. The sample has mean $\bar{x} = 5.7$ and standard deviation $s = 7.2$. Give an estimate of λ and its standard error.

(b) The log-likelihood for a set of data is given by $\ln L(\theta) = -20\theta + 100 \ln \theta$. Find the maximum likelihood estimate of θ and its standard error.

(c) Each day for thirty days, a random sample of 12 items from the day's production is selected and carefully measured: the average of the 12 measurements ($\bar{x}$) and the range of the 12 measurements (R) are calculated and recorded each day.

At the end of the thirty days, the average of the daily averages, $\bar{\bar{x}} = 36.29$; and the average of the daily ranges, $\bar{R} = 7.32$.

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

i. Explain why $\bar{R} \approx 3.258\sigma$; and hence derive an estimate of σ.

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

iii. Instead of R, it is suggested that the difference between the second largest and the second smallest observation be used. If this difference is denoted by Q, find c such that $\bar{Q} \approx c\sigma$. What advantage/disadvantage might an estimator based on $\bar{Q}$ have over the estimator based on $\bar{R}$?

7. (a) Sketch a graph, roughly indicating location and spread, of each of the following:

i. the pdf of t_{10};
ii. the pdf of χ^2_{10};
iii. the pdf of $F_{10,10}$.

(b) A random sample of nine observations is obtained on $X_1 \stackrel{d}{=} N(\mu_1, \sigma_1^2)$, and it is found that $\bar{x}_1 = 38.9$, $s_1 = 7.41$.

An independent random sample of nine observations is obtained on $X_2 \stackrel{d}{=} N(\mu_2, \sigma_2^2)$, for which $\bar{x}_2 = 31.5$ and $s_2 = 5.67$.

i. Find a 95% confidence interval for σ_1/σ_2; and verify that there is no significant evidence against the null hypothesis H_0: $\sigma_1 = \sigma_2$.
ii. Specify the pooled standard deviation estimate based on both samples.
iii. Assuming $\sigma = \sigma_1 = \sigma_2$, obtain a 95% confidence interval for σ.
iv. Using the pooled variance estimate, obtain a 95% confidence interval for $\mu_1 - \mu_2$.

8. (a) Determinations of the strength of a fibre after using three treatments were as follows:

	number	mean	variance
control	6	76	20
treatment A	6	82	24
treatment B	6	84	31

i. Show that the error mean square is 25; and hence, or otherwise, complete the following analysis of variance table:

	df	SS	MS	F
treatments	**	******	****	***
error	**	******	****	
total	**	575.00		

ii. Test the hypothesis H_0: $\mu_1 = \mu_2 = \mu_3$ giving an approximate P-value. What do you conclude?

(b) A study was made to determine if humidity conditions have an effect on the force required to pull apart pieces of glued plastic. Two types of plastic were tested using three levels of humidity. The results are given in the table below. There are two observations on each factor combination; the number in brackets is the average of these two observations.

	humidity		
type	30%	60%	90%
A	41.2, 40.6 (40.9)	38.6 37.8 (38.2)	35.5, 33.3 (34.4)
B	39.0, 40.8 (39.9)	34.6, 37.4 (36.0)	23.2, 26.4 (24.8)

The following analysis of variance was obtained for these data:

	df	*SS*	*MS*	*F*
type	1	54.6	54.6	24.1
humidity	2	245.0	122.5	354.5
interaction	2	43.4	21.7	9.6
error	6	13.6	2.26	
total	11	356.6		

Give a brief analysis and interpretation of these results.

9. The following represent the results of a one replicate 2^3 experiment.

y	P	Q	R
17.2	0	0	0
12.3	0	0	1
16.6	0	1	0
12.8	0	1	1
20.4	1	0	0
15.8	1	0	1
17.3	1	1	0
14.8	1	1	1

	av.y
P0	14.7
P1	17.1
Q0	16.4
Q1	15.4
R0	17.9
R1	13.9

The following split-up of the sum of squares for these data was obtained using MATLAB, with X1=P, X2=Q and X3=R:

```
Source     Sum Sq.  d.f.   Mean Sq.    F    Prob>F
-------------------------------------------------------
X1          11.1     1      11.1      Inf    NaN
X2           2.2     1       2.2      Inf    NaN
X3          31.2     1      31.2      Inf    NaN
X1*X2        2       1       2        Inf    NaN
X1*X3        0.3     1       0.3      Inf    NaN
X2*X3        1.3     1       1.3      Inf    NaN
X1*X2*X3     0.1     1       0.1      Inf    NaN
Error       -0       0      -0
Total       48.2     7
```

(a) Why are the p-values undefined?

(b) Generate a half-normal plot for the root mean-squares and use this to identify the possibly important effects.
Note: $\Phi^{-1}(\frac{1+q}{2}) = 0.16, 0.32, 0.49, 0.67, 0.89, 1.15, 1.53$, *for* $q = \frac{1}{8}, \frac{2}{8}, \ldots, \frac{7}{8}$.

(c) Produce a revised analysis of variance and determine the significant effects.

(d) Obtain an estimate and a 95% confidence interval for the effect of R.

(e) Give an interpretation of your results.

10. (a) The following tables gives two observations on y for each of the specified values of x. The average value of y for each x-value and the overall average are also given.

x	0	1	2	3	4	
y	13	25	27	32	37	
	16	23	29	33	35	
$\bar{y}$	14.5	24.0	28.0	32.5	36.0	$\bar{\bar{y}}$ = 27.0

The following results were obtained using these data:

Regression Analysis

```
The regression equation is  y = 16.7 + 5.15 x

                est        se        t       P
intercept     16.70     1.155    14.46   0.000
slope          5.15     0.471    10.93   0.000
                 DF        SS       MS        F       P
Regression        1   530.45   530.45   119.37   0.000
Residual Error    8    35.55     4.44
Total             9   566.00
                                 S = 2.108  R-Sq = 93.7%
```

One-way ANOVA

```
Source          DF       SS       MS       F       P
Between groups   4   555.00   138.75   63.07   0.000
Within groups    5    11.00     2.20
Total            9   566.00
                                 S = 1.483  R-Sq = 98.06%
```

Find a 95% confidence interval for $\mathrm{E}(Y \mid x=2)$:

i. assuming a straight line regression model;
ii. making no assumptions about the regression.

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

(b) A random sample of $n=50$ observations is obtained on (X, Y), from which the sample correlation is found to be $r=-0.5$.

i. Indicate with a sketch the general nature of the scatter plot for these data.
ii. Give an approximate 95% confidence interval for the population correlation.