Exercise 3.2

Identify the experimental unit, the replication for each treatment and whether pseudo-replication is present in the following experiments.

a)   A pot experiment with 12 circular pots in a 2 × 6 array, in a uniform environment. Each pot contains four plants at the three-leaf stage, and each of four treatments (labelled A to D) were applied at random to one plant per pot as shown in Figure 3.8.

b)   A field experiment with 12 homogeneous rectangular plots in a 3 × 4 grid. Two treatments (labelled A and B) were applied at random to six plots each (Figure 3.9). At harvest, twenty-five plants are to be sampled per plot, and the plants from each plot will be processed as a single batch for measurement.

c)   The field experiment described in part (b) (Figure 3.9) with the height of 25 individual plants per plot measured and recorded *in situ* at four-weekly intervals from tillering until harvest.

Solution 3.2

a) As the treatments are applied directly to individual plants, the experimental unit is a plant, the replication is n = 12 (i.e. there are 12 replicates of each treatment and N = 48 units in total), and no pseudo-replication is present. This experiment is a RCBD with pots as blocks.

There is an interesting contrast here with the first example in the bulleted list of Section 3.1.1 (and Figure 3.1). Whilst the physical structure (plants within pots) is the same in both cases, the fact that treatments are applied at different levels of the structure (here to plants, there to pots) means that the experimental design structure and subsequent analysis differs.

b) As treatments are applied to whole plots, the experimental unit is a plot, the replication is n = 6, and pseudo-replication (the 25 plants from each plot) is present in the plot samples. However, as the sample of plants from each plot produces a single observation (as they are processed in bulk) this will not be apparent in the analysis which uses observations for the N = 12 plots.

c) Despite the changes to the sampling scheme, the underlying experiment here is exactly the same as that in (b), i.e. the experimental unit is still a plot. Pseudo-replication (the 25 plants from each plot) is present and each pseudo-replicate will give rise to several, say t, four-weekly repeated measurements. By the end of the experiment there will therefore be N = $300 \times t$ observations, but the true replication is still n = 6 for each treatment.

Exercise 3.4

Four novel herbicides (labelled A–D) are to be compared with a commercial product (labelled P) and a hand-weeded control (labelled H) in a field trial, giving six treatments in total. The field available can accommodate 24 plots in an array of four columns running north to south, by six rows running west to east (Figure 3.11). The field has a known pH gradient running west to east, i.e. along rows, which may affect crop growth. Produce a RCBD which accounts for this gradient with a randomized allocation of treatments to plots using (a) a pack of playing cards, and (b) a standard six-sided die.

Solution 3.4

To take the pH gradient into account we would construct a RCBD by grouping the plots into blocks running as columns on the plan with

Block 1:	Plots 1, 5, 9, 13, 17, 21
Block 2:	Plots 2, 6, 10, 14, 18, 22
Block 3:	Plots 3, 7, 11, 15, 19, 23
Block 4:	Plots 4, 8, 12, 16, 20, 24

Each plot in any given block will therefore have a similar pH value. Plots in different blocks will have different pH levels, with the maximum difference occurring between the most easterly and westerly blocks. This arrangement is illustrated in Figure S3.2.1. The six treatments should be allocated at random to plots within blocks using a different randomization for each block. Each treatment will therefore be tested within each pH condition. We assume other environmental factors are similar across the whole area.

If we define factors Block, with four levels 1–4 numbering blocks from west to east, and Plot, with six levels 1–6 numbering plots within blocks from north to south, then the structural component can be written as Blocks/Plots. This nested structure acknowledges that plots are nested within blocks, with no relationship among plots with the same level of factor Plot across blocks.

a) A valid randomization can be done using six playing cards. For example, allocate the 1–6 of Clubs to treatments A–D, P and H, respectively, then shuffle the six cards. Take the first card from the top of the pile and allocate the treatment it represents to the first unit in block 1 (plot 1). Use the next card to allocate a treatment to the second unit in block 1 (plot 5) and continue until all units in block 1 have a treatment allocated. Reshuffle the cards and repeat the process three more times to get a different randomization for blocks 2, 3 and 4. Our randomization is shown in Figure S3.2.1.

Figure S3.2.1. Randomization of six treatments to plots within blocks in a RCBD herbicide field trial. Treatments labelled as A–D (novel herbicides), P (commercial product) and H (hand-weeded control). Alternate blocks shaded in grey.

Note that the presence of treatment D twice in the last row here (plots 23 and 24) should not give cause for concern if our blocking is correct (i.e. the major trend in the field runs from west to east). This allocation would only be a problem if there was also a north to south trend that might affect the observations and, in that case, a design with a crossed blocking structure (Sections 3.2 and 3.3.3; also see Chapter 9) should be used.

Note that many other valid schemes are also possible, such as using cards 1–6 of a different suit for each block, allocating the cards to treatments in a different order, or allocating cards 5–10 to the treatments, etc., etc.

b) For this design, a valid randomization can also be done using a die. We first allocate treatments A–D, P and H to the numbers 1–6. We start with the randomization for the first block: roll the die to allocate a treatment to the first unit of block 1 (plot 1). Then roll the die to allocate a treatment to the second unit of block 1 (plot 5), but ignore the result and roll again if the treatment already allocated to plot 1 occurs. For the third to fifth unit, we again roll the die, but ignore the result and reroll if any treatment that has already occurred in the block is repeated. The treatment for the sixth unit is determined by elimination.

To determine whether this randomization scheme is valid, we consider the probabilities at each stage. We want each treatment to be equally likely to be applied to each unit within each block, each with probability of 1/6. In our process, assuming the die is fair, all treatments are allowed in the allocation of a treatment to the first unit of the block, all with equal probability (1/6). For the second unit of the block, we ignore the treatment chosen for the first unit but the other five treatments are all allowed, all with equal probability (1/5). The probability of the second unit receiving treatment A, say, is equal to the probability that the first unit does not receive treatment A multiplied by the probability that the second unit does receive treatment A, i.e. $\frac{5}{6} \times \frac{1}{5} = \frac{1}{6}$.

And so on. In fact the probability that any plot within a block receives treatment A (or any other specific treatment) is equal to 1/6.

This randomization process is valid because we work separately within each block and only have one replicate of each treatment within each block. If there were two replicates of any treatment within a block, a more complex scheme would be required.

In addition we provide details of how to generate this design using statistical software below.

GenStat 3.4

```
" Create RCBD design with 4 blocks and 6 plots per block,
  with treatments labelled A-D, P, H"
FACTOR [LEVELS=6; LABELS=!T(A,B,C,D,P,H)] IDENTIFIER=Treatment
AGHIERARCHICAL [PRINT=design; ANALYSE=no; SEED=26289] \
   BLOCKFACTORS=Block,Plot; TREATMENTFACTORS=*,Treatment; LEVELS=4,6
```

R 3.4

Note: this solution uses the agricolae package

```
library(agricolae)
trt <- c("A", "B", "C","D","H","P")
nblock <- 4
design <- design.rcbd(trt, r=nblock, seed=26289)
design$book
```

SAS 3.4

```
* Create RCBD design with 4 blocks and 6 plots per block with treatments labelled A-D, P, H;
proc plan seed=26289;
  factors Block=4 ordered Treatment=6 random;
  output out=RCBD
       Treatment cvals=('A' 'B' 'C' 'D' 'P' 'H') ordered;
run;

* Printing final scheme;
proc print data=RCBD;run;
```

Exercise 3.6

The effect of temperature on transmission of a virus by five aphid species is to be investigated. Three small growth chambers are available and three temperatures will be tested. The temperature for each chamber can be set and then applies to the whole chamber, and each chamber can hold five plants in individual pots. One aphid will be placed onto each plant using a clip cage. Forty-five plants and 15 aphids of each species are available. Assuming that chambers (and positions within chambers) can be considered homogeneous, suggest a design to test the effects of temperature and aphid species. What are the experimental units for each factor? Produce a randomized design for this experiment and write down the explanatory and structural components for the design. If you suspected that there were systematic differences between chambers, how would you modify your design? Write down the structural component for this new design.

Solution 3.6

Suggested design
With three chambers, we can run a maximum of 15 plants at one time (five per chamber). We are told that the chambers behave similarly (can be considered homogeneous) so we can consider the set of three chambers as a block and set one chamber at each temperature (allocated at random). Within each chamber there are five positions that can be considered homogeneous and we will also make assumptions of homogeneity within the sets of plants and aphids. We allocate plants at random to positions within chambers. We randomly allocate each of the five aphid species to one plant within each chamber and place one aphid of the allocated type onto each plant. This gives one replicate of each species × temperature combination and forms one run of the basic experiment. Replication will be achieved by running the experiment several times. In one run we use 15 plants and three aphids of each species and so with the resources available we will be able to repeat the experiment three times, using 45 plants and nine aphids of each species in total. Six aphids of each species remain unused as the number of plants is the limiting resource here.

Identify experimental units
This experiment has a hierarchical structure: temperatures are applied to chambers, and species are applied to plants within chambers; this is a split-plot design. Temperatures will be randomized to chambers, independently within each run. Aphid species will be randomized to plants within each chamber, with a different randomization in each chamber in each run. The experimental units for the temperature factor are chambers, and the experimental units for the aphid species factor are plants within chambers.

Explanatory and structural components of model
We have 15 experimental treatments with a 3 × 5 factorial structure, as each aphid species will be tested with each temperature within each run. There are three replicates of each treatment combination. A crossed structure is appropriate for the explanatory component, as the main effects associated with both factors are meaningful and the interaction between them is of interest. The hierarchical structural component has plants nested within chambers nested within runs. The explanatory and structural components for the design can therefore be written in symbolic notation as

Explanatory component:	Temperature * Species
	= Temperature + Species + Temperature.Species
Structural component:	Run / Chamber / Plant
	= Run + Run.Chamber + Run.Chamber.Plant

where Temperature and Species are factors with three and five levels labelling the temperature and species treatments, respectively, Run is a factor with three levels labelling the three runs of the experiment, Chamber is a factor with three levels labelling the chambers within runs, and Plant is a factor with five levels labelling the plants within each chamber.

Generating a randomized plan for the design
Randomization of the treatments can be achieved in many different ways. For example, we might write the numbers 1–3 on slips of paper to represent the three temperatures, then select the slips one at a time from a bag (without replacement) to allocate temperatures to chambers in the first run. The process would then be repeated to allocate temperatures to chambers in the second and third runs. A similar process with slips numbered 1–5 representing the different species can be used to allocate species to plants within each chamber; this process would need to be done nine times, once for each chamber in each run.

If a computer package is used then the allocation of temperatures to chambers, the allocation of aphid species to plants within chambers and the number of runs must all be specified. This differs between packages; some sample code is given below. Our randomization is shown in Figure S3.6.1 (with temperatures labelled T1–T3 and species as S1–S5).

```
T3  T3  T3  T3  T3 │ T2  T2  T2  T2  T2 │ T1  T1  T1  T1  T1
S2  S5  S1  S4  S3 │ S2  S4  S1  S3  S5 │ S4  S1  S2  S5  S3

T1  T1  T1  T1  T1 │ T2  T2  T2  T2  T2 │ T3  T3  T3  T3  T3
S4  S2  S3  S5  S1 │ S3  S1  S2  S4  S5 │ S2  S5  S1  S3  S4

T2  T2  T2  T2  T2 │ T3  T3  T3  T3  T3 │ T1  T1  T1  T1  T1
S2  S1  S4  S5  S3 │ S4  S2  S5  S3  S1 │ S5  S3  S1  S4  S2
```

Figure S3.6.1. Split-plot layout for growth chamber experiment. Runs correspond to rows of this plan and are separated with solid lines, chambers are separated by dashed lines, and plants within chambers are separated with dotted lines.

Systematic differences between chambers
Systematic differences between chambers might be suspected for various reasons. For example, a pilot study or observations from previous experiments might suggest that one chamber consistently ran a little below temperature and one a little above, the chambers might be from different manufacturers or of different ages, or the locations of the chambers might have slightly different background environments. In this case, chambers cannot be considered as nested within runs, as there is an association between results from the same cabinet across runs (due to the systematic differences) — the cabinets are no longer homogeneous (exchangeable).

In this case, a Latin square design might be used to allocate temperatures to chambers using a crossed structure such that each chamber is set to each temperature once over the three runs and each temperature still appears once in each run. The process for allocation of species to plants within chambers stays the same. The structural component now contains both crossing and nesting. The symbolic notation for this design is

Explanatory component:	Temperature * Species
Structural component:	(Run*Chamber) / Plant
	= Run + Chamber + Run.Chamber + Run.Chamber.Plant

where all factors are as defined earlier. Our randomization for this design is shown in Figure S3.6.2.

```
T2  T2  T2  T2  T2 │ T3  T3  T3  T3  T3 │ T1  T1  T1  T1  T1
S4  S3  S5  S2  S1 │ S5  S2  S4  S3  S1 │ S4  S3  S5  S1  S2

T1  T1  T1  T1  T1 │ T2  T2  T2  T2  T2 │ T3  T3  T3  T3  T3
S5  S3  S4  S1  S2 │ S4  S1  S5  S3  S2 │ S4  S2  S3  S1  S5

T3  T3  T3  T3  T3 │ T1  T1  T1  T1  T1 │ T2  T2  T2  T2  T2
S4  S1  S5  S2  S3 │ S3  S5  S1  S4  S2 │ S5  S3  S1  S2  S4
```

Figure S3.6.2. Crossed and nested layout for growth chamber experiment. Runs separated with solid lines, chambers separated by dashed lines, and plants within chambers separated with dotted lines.

Note the difference between the two plans. In the design with crossing and nesting (Figure S.3.6.2), each temperature appears once in each run and in each chamber, as desired. In the purely nested design (Figure S.3.6.1), each temperature appears once in each run, but there is no constraint in the allocation to chambers so, for example, chamber 2 is set twice to temperature T2 but never to temperature T1.

GenStat 3.6

```
" Create split-plot design with 3 blocks, 3 main plots per block and 5 subplots per main plot
"
AGHIERARCHICAL [PRINT=design; ANALYSE=no; SEED=1096] \
   BLOCKFACTORS=Run,Chamber,Plant; TREATMENTFACTORS=*,Temperature,Species; LEVELS=3,3,5

" Print factors "
PRINT STRUCTURE=Run,Chamber,Plant,Temperature,Species

" Plot experimental plan "
VARIATE IDENTIFIER=X,Y; VALUES=!((1...15)3),!(15(1...3))
DDESIGN [Y=Y; X=X; SIZE=0.5] FACTOR=Run,Temperature,Species; PEN=0,1,2; \
   PENGRID=1...3; LABELS=*,!t(T1,T2,T3),!t(S1,S2,S3,S4,S5)

" Introduce Latin square structure on chambers "

" Duplicate structural factors already set up "
DUPLICATE OLDSTRUCTURE=Run,Chamber,Plant; NEWSTRUCTURE=Run2,Chamber2,Plant2

" Get starting systematic allocations of temperatures to chambers (Latin square) and species
to plants (within chambers) "
FACTOR [LEVELS=3; VALUES=5(1,2,3,2,3,1,3,1,2)] IDENTIFIER=Temperature2
DUPLICATE OLDSTRUCTURE=Plant2; NEWSTRUCTURE=Species2

" Get randomized allocation of temps to chambers and species to plants "
RANDOMIZE [BLOCKSTRUCTURE=(Run2*Chamber2)/Plant2] Temperature2,Species2

" Print factors "
PRINT STRUCTURE=Run2,Chamber2,Plant2,Temperature2,Species2

" Plot experimental plan "
DDESIGN [Y=Y; X=X; SIZE=0.5] FACTOR=Run2,Temperature2,Species2; PEN=0,1,2; \
   PENGRID=1...3; LABELS=*,!t(T1,T2,T3),!t(S1,S2,S3,S4,S5)
```

R 3.6

Note: this solution uses the agricolae package

```
library(agricolae)

# Standard split-plot design
Temperature <- c("T1","T2","T3")
Species<-c("S1","S2","S3","S4","S5")
design <-design.split(Temperature, Species, r=3, seed=1097)
design$book

# Generating the Latin square variation is not straightforward within R
```

SAS 3.6

```
* Create split-plot design with 3 blocks, 3 main plots per block and 5 subplots per main plot;
proc plan seed=1096;
  factors Run=3 ordered
          Chamber=3 ordered
                     Temperature=3 random
                     Species=5 random;
  output out=SPLITPLOT
    Temperature cvals=('T1' 'T2' 'T3') ordered
    Species cvals=('S1' 'S2' 'S3' 'S4' 'S5') ordered;
run;

* Printing final scheme;
proc print data=SPLITPLOT;run;
```

Exercise 3.8

A glasshouse experiment to compare the effect of two nutrition regimes on the growth of three wheat varieties was set up as a RCBD with 12 blocks of six pots each, as shown in Figure 3.12. The treatments comprise the six combinations of nutrition regime (labelled N1, N2) and variety (labelled V1–V3). The blocks accommodate an expected temperature gradient running from the door to the far end of the glasshouse. Several characteristics of each plant, including height and number of leaves, are to be recorded every week. Suggest acceptable protocols for recording data if
   a)   you are the only person available to take the measurements;
   b)   there are two people available to take the measurements.
Which protocols would be unacceptable and why?

Solution 3.8

a) Given that it will take some time to make the measurements on each plant, we might expect that time-of-measurement is a possible source of heterogeneity and so should be accounted for by blocking. One easy way to implement this is to use the blocking already imposed. So you might start by measuring all the plants in the block nearest the door (pots numbered 12, 24, 36, 48, 60 and 72) and work your way, block by block, along the greenhouse to the far end. If you need to take a break then you can schedule it between two blocks. This scheme confounds the temperature effect and any time-of-measurement effect — so these two sources of heterogeneity cannot be separated — but this is not important unless you are interested in both effects. What is important is that circumstances in a given block are homogeneous, e.g. all the plants in a block are in the same environmental conditions and measured at a similar time.

b) If there are two people available then sensible protocols will involve allocating whole blocks to each person. For example, person A might assess the block nearest the door first (pots numbered 12, 24, 36, 48, 60 and 72) and then continue block by block towards the centre of the glasshouse. Person B might start by assessing the block nearest the far end (pots numbered 1, 13, 25, 37, 49 and 61), and then also work block by block towards the centre of the glasshouse. In practice, this protocol might lead to the two people assessing a different number of blocks (e.g. if one works faster than the other, or if one has to leave, etc.). With this protocol, person A will be associated with blocks at cooler temperatures and person B with blocks at higher temperatures but, if there is no requirement to separate recorder and temperature effects separately, this will be an acceptable protocol. Alternatively, the two people might be randomly allocated to blocks or might assess alternate blocks. It is always sensible to record which person assessed which block and in which order (also the order of measurements of pots within blocks); this information might be useful, for example, to produce index plots as part of the subsequent analysis (Section 5.2.2).

Protocols that are not advisable include

- Measure all plants of one treatment first and then all of another, etc. In this protocol, any time-of-measurement effect, e.g. due to physiological effects or you getting tired as the day goes on, will be confounded with the treatment effect you are trying to test.
- Work along rows, e.g. down one edge of the glasshouse first (pots numbered 1–12) and then come back along the next row, etc. In this protocol you cut across blocks, and it is difficult to efficiently account for any time-of-measurement effect in the statistical analysis (although it is now separated from temperature effects). If separation of temperature and time-of-measurement effects is required, these two structural sources of heterogeneity should be incorporated using a crossed blocking structure.