STATISTICAL MODELLING

X. Sample size and power

In the context of designed experiments determining the sample size amounts to finding the number of pure replicates of a treatment that we will denote by *r*.

X.A How it is done

Consider the following example.

Example X.1 Penicillin yield

In example IV.1 the effects of four treatments (A, B, C and D) on the yield of penicillin were investigated using an RCBD. It was concluded that the following treatment means were not significant:

Treatment			
A	B	C	D
84	85	89	86

However, perhaps there are differences between the penicillin treatments — it is just that 5 replicates was not enough given the size of the true differences between the treatment means. How many replicates should we take? ■

The general problem is to determine the number of pure replicates of a treatment that we will denote by *r*. To do this you have to specify:

1. the significance level, α;
2. the power (or probability of detecting a difference when there is actually a difference) desired, $1 - \beta$;
3. the number of treatments to be investigated
4. the minimum size of the difference to be detected between a pair of treatment means, as measured by Δ;
5. the uncontrolled variation, σ_U^2, expected.

In working out values for these often the results from a previous experiment you or another researcher has run will be useful.

A convenient method of computing the required number of pure replicates, *r*, is provided by the function `no.reps` — it uses the standard R function `optimize` to minimize the difference between the required power and the actual power of test numbers of replicates. The function `no.reps`, and the associated functions `power.exp` and `power.diff`, are available in the *dae* library. The usage and arguments for this function are as follows:

```
no.reps(multiple=1, df.num=1,
                  df.denom=expression((df.num+1)*(r-1)),
                  delta=1, sigma=1, alpha=0.05, power =0.8,
                  tol = 0.025, print=FALSE)
```

`multiple`: the multiplier, *m*, which when multiplied by the number of pure replicates of a treatment, *r*, gives the number of observations (rm) used in computing means for some, not necessarily proper, subset of the treatment factors; *m* is the replication arising from other treatment factors. However, for single treatment factor experiments the subset can only be the treatment factor and $m = 1$

`df.num`: the degrees of freedom of the numerator of the F for testing the term involving the treatment factor subset;

`df.denom`: an expression for the degrees of freedom of the denominator of the F for testing the term involving the treatment factor subset — it must involve *r*, the number of pure replicates, can involve other arguments to `no.reps` such as `multiple` and `df.num`, and must be enclosed in an `expression` function so that it is not evaluated when `no.reps` is called but will be evaluated as different values of *r* are tried during execution of `no.reps`;

`delta`: the true difference between a pair of means for some, not necessarily proper, subset of the treatment factors;

`sigma`: population standard deviation;

`alpha`: the significance level to be used;

`power`: the minimum power to be achieved;

`tol`: the maximum difference tolerated between the `power` required and the power computed in determining the number of replicates;

`print`: TRUE or FALSE to have or not have a table of power calculation details printed out.

Note that the values given for the arguments in the above expression for `power.exp` are the default values assigned to the arguments if they are not set in a call to the function.

Example X.1 Penicillin yield (continued)

We now determine the number of replicates required to achieve a power of 0.80 in detecting $\Delta = 5$ with $\alpha = 0.05$. In the analysis of variance for this experiment, the Residual MSq was 18.83 so we will take $\sigma^2_U \approx 20$. The output from the use of `no.reps` for the present example is as follows:

```
> no.reps(multiple=1, df.num=3, df.denom=expression(df.num*(r-1)),
+ delta=5, sigma=sqrt(20), power=0.8)

$nreps
[1] 19

$power
[1] 0.8055926
```

That is the required number of replicates is 19 and that this will achieve a power of 0.8055. ■

However, fundamental to these calculations is the concept of power and so we examine this concept in more detail.

X.B Power

a) Type I and II errors

It is important to keep in mind that your conclusion about the null hypothesis is not 100% certain to be correct. There is always the chance that you are wrong, although in some cases the chance is so small that you will be virtually 100% certain. Remember, however, that very unusual events do happen; for example, Baum and Scheuer (1976) report a case of the same person winning the same lottery twice, the chance of doing so being 500 million to 1.

The possible outcomes — the conclusion (or verdict) reached as a result of a hypothesis test — are illustrated in the following table:

Relationship between reality and verdict for H_0

	H_0 reality	
H_0 verdict	true	false
not rejected	correct (innocent cleared)	type II error (guilty cleared)
rejected	type I error (innocent convicted)	correct (guilty convicted)

Thus there are two types of errors that can be made in performing a hypothesis test.

Definition X.1: A **type I error** is made when the null hypothesis is true and it is rejected. The probability of a type I error is designated α and is:

$$P\left\{H_0 \text{ rejected} \mid \mathrm{H}_0 \text{ true}\right\} = \alpha .$$ ■

Definition X.2: A **type II error** is made when the null hypothesis is false and it is not rejected. The probability of a type II error is designated β and is:

$$P\left\{H_0 \text{ not rejected} \mid \mathrm{H}_0 \text{ false}\right\} = \beta$$ ■

Using the analogy that deciding between the null and alternative hypotheses in a hypothesis test is the same as deciding the verdict in a trial, a type I error is the same as convicting an innocent person while a type II error is the same as failing to convict a guilty person.

In a hypothesis test we set the probability of a type I error at α, which is called the significance level, α; that is we set the level of risk we are prepared to take in making a type I error. But what about the probability of a Type II error, β? Rather than consider β, we often consider $1-\beta$, which is called the power of the test.

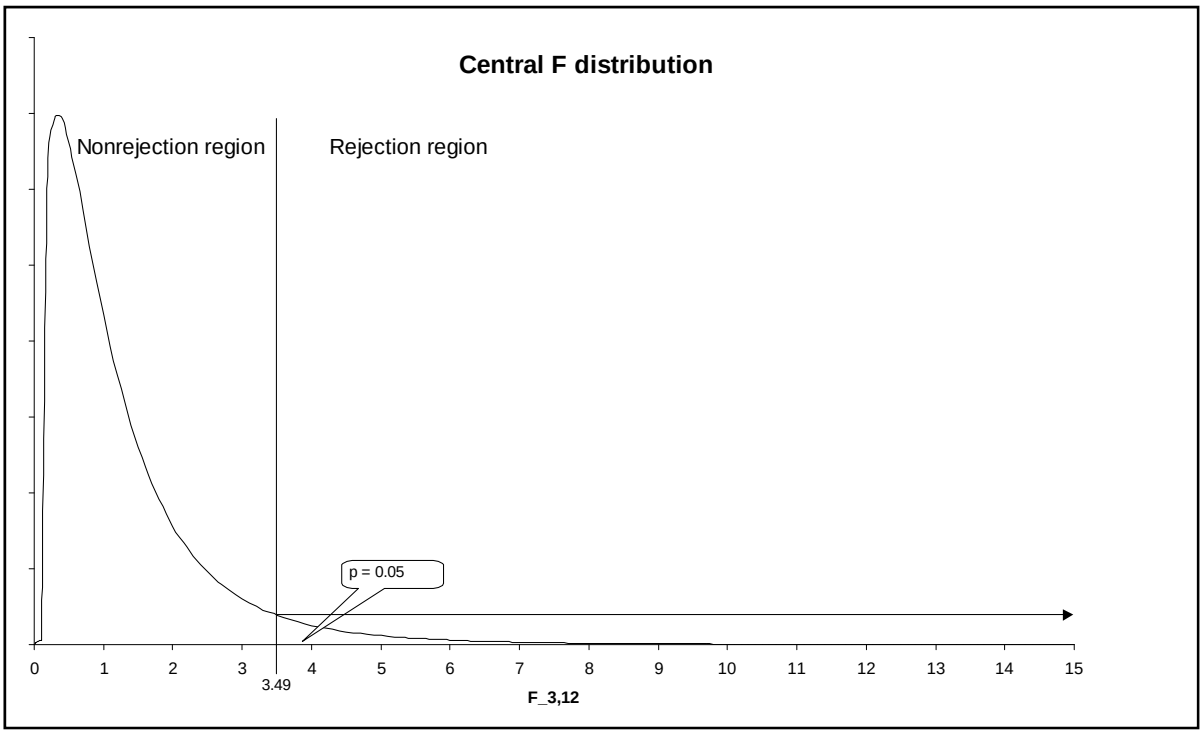
b) Power of a hypothesis test about expectation model terms

Definition X.3: The **power** of a hypothesis test is the probability of rejecting the null hypothesis when it is false is:

$$1-\beta = P\{H_0 \text{ rejected} \mid H_0 \text{ false}\}$$

■

Now to compute this probability means that we need to know the condition under which the null hypothesis is rejected. In general the null hypothesis is rejected when the computed value of the probability of the F statistics from the analysis of variance is less than α. This will occur whenever the observed value of the F is greater than the α% value from the Snedecor's F distribution as illustrated in the following diagram for $F_{0.05,3,12}$. In this case any observed value of the test statistic greater than 3.49 would result in a p-value of less than 0.05 and so be rejected at the 5% level.



The next quantity that needs to be specified is how false is the null hypothesis; that is, we know that there is a difference between the population treatment means. But, how big a difference? For the single, treatment-factor experiments a measure, relative to the magnitude of the uncontrolled variation, is given by

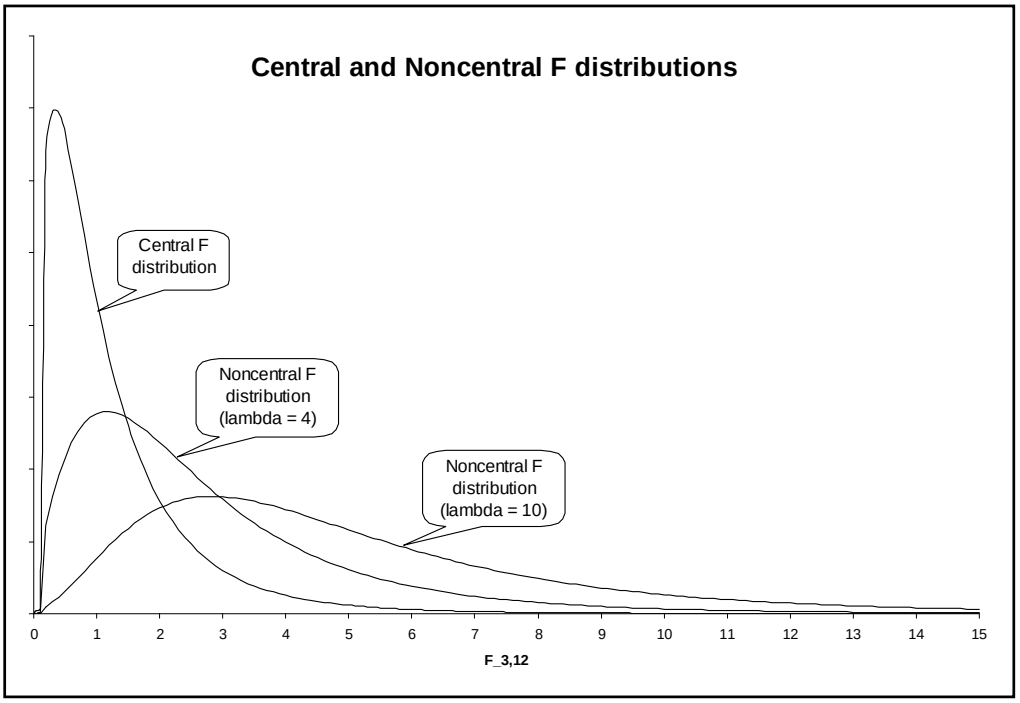
$$\lambda = \frac{r\sum_{j=1}^{t}\left(\tau_j - \bar{\tau}_.\right)^2}{\sigma_U^2} = \frac{(t-1)q_T(\psi)}{\sigma_U^2}$$

provided the number of replicates for all treatments is *r*. However, to use this formula requires that we know the values of the τs which is unlikely as estimating these is the purpose of the experiment. If they were known, the experiment would be irrelevant. A way to overcome this is to specify Δ, the difference such that if any two population means differ by this amount, the null hypothesis should be rejected. Then it can be shown that a general formula for the minimum value of λ is

$$\lambda = \frac{rm\Delta^2}{2\sigma_U^2}$$

where *r* is the number of pure replicates of each treatment and *m* is the multiplier of *r* that gives the number of observations $(=rm)$ used in computing one of the means being compared. For experiments involving a single treatment factor $m = 1$. For a single, treatment-factor RCBD $r = b$ and for a single, treatment-factor LS $r = t$.

We can now be more specific about computing the power: $1-\beta = P\{H_0 \text{ rejected} \mid H_0 \text{ false}\}$. We need the probability of getting a value of F greater than that of the α% value from Snedecor's F distribution when there is a difference between the treatments of the order specified by Δ (or λ). Clearly, Snedecor's distribution cannot be used to compute this probability as it is the distribution that applies when the null hypothesis is true: there is no difference between the τs in the population and $\Delta = 0$. We need a distribution of F for when the null hypothesis is false. This is provided by the noncentral F distribution, a modification of the (central) Snedecor's F distribution to incorporate a noncentrality parameter λ. The shape of the noncentral F distribution depends on ν_1, ν_2 and λ as illustrated in the following diagram— for the central F distribution $\lambda = 0$.



Distribution for:

$\lambda = 0$ is distribution of $F_{3,12}$ when H_0 true
$\lambda = 4$ is distribution of $F_{3,12}$ when H_0 is not true
$\lambda = 10$ is distribution of $F_{3,12}$ when H_0 is even less true

Specifically, to compute the power of an analysis of variance test for a fixed factor:

$$\begin{aligned} 1-\beta &= 1-P\left\{H_0 \text{ not rejected} \mid \mathrm{H}_0 \text{ false}\right\} \\ &= P\left\{H_0 \text{ rejected} \mid \mathrm{H}_0 \text{ false}\right\} \\ &= P\left\{F_{\nu_1,\nu_2,\lambda} \geq F_{\alpha,\nu_1,\nu_2}\right\} \end{aligned}$$

where F_{α,ν_1,ν_2} is the F value from the central F distribution such that $P\left\{F_{\nu_1,\nu_2} \geq F_{\alpha,\nu_1,\nu_2}\right\} = \alpha$.

The R function `pf` computes probabilities for the noncentral F distribution, for which the arguments are q (= F), df1, df2 and ncp (= λ). In addition, a function `power.exp` had been made available in the *dae* library for computing the power in an experiment. The usage and arguments for this function are:

```
power.exp(rm=5,df.num=1, df.denom=10, delta=1, sigma=1,
   alpha=0.05, print=FALSE)
```

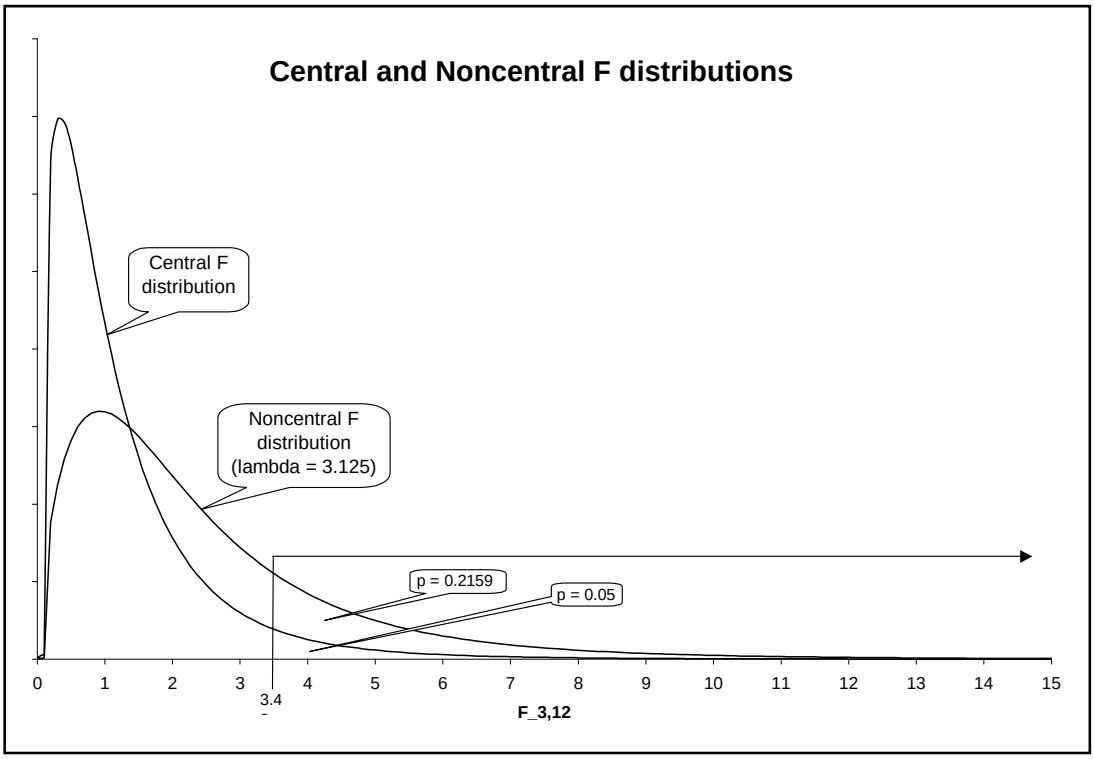
`rm`: the number of observations used in computing a mean.
`df.num`: the degrees of freedom of the numerator of the F for testing the term involving the means;
`df.denom`: the degrees of freedom of the denominator of the F for testing the term involving the means;
`delta`: the true difference between a pair of means;
`sigma`: population standard deviation;
`alpha`: the significance level to be used
`print`: TRUE or FALSE to have or not have a table of power calculation details printed out.

Example X.1 Penicillin yield (continued)

Suppose that it was expected that the minimum difference between a pair of treatment means is 5 and that $\alpha = 0.05$. We continue to take $\sigma_U^2 \approx 20$. Also $r = 5$ and $m = 1$. The output from the `power.exp` call to compute the power is given below. Note that alpha is not set in this call and so the default value of 0.05 will be used. Also, the expressions `3 * (rm - 1) and sqrt(20)` will be evaluated prior to the call to the function. To get the correct value of `rm` used in evaluating the expression `3 * (rm - 1)`, `rm` needs to be set prior to calling `power.exp`.

```
> rm <- 5
> power.exp(rm=rm, df.num=3, df.denom=3*(rm-1), delta=5, sigma=sqrt(20),
print=TRUE)
  rm df.num df.denom alpha delta    sigma lambda      powr
1  5      3       12  0.05     5 4.472136  3.125 0.2159032
[1] 0.2159032
```

The power for detecting a minimum difference of 5 in a pair of treatment means when $\sigma_U = 4.47$ is 0.2159. Note that $\lambda = 3.125$. The situation is illustrated in the following diagram. If there was no difference between the treatments, the probability of rejecting the null hypothesis is determined from the central F distribution to be 0.05. However, if the minimum size of the treatment difference is as measured by $\lambda = 3.125$, the probability of rejecting the null hypothesis is determined from the noncentral F distribution to be 0.2159.



The power of the hypothesis test being 0.22 is not good because we do not have a high chance of correctly rejecting the null hypothesis. How can we improve things? ■

For a fixed number of treatments, the power will increase as the noncentrality parameter and the residual degrees of freedom increase. Examination of the formula for the noncentrality parameter $\lambda = \left(rm\Delta^2\right)\Big/2\sigma_U^2$ leads us to conclude that, for a fixed set of treatments, the noncentrality parameter will increase if

1. the number of replicates, *r*; increases;
2. the size of the differences between the treatments, as measured by Δ, is increased;
3. the uncontrolled variation, σ_U^2, is decreased.

So we can get better power by taking more observations, by increasing the differences between the treatments or by decreasing the uncontrolled variation. The latter can be done either by improving the protocols used in conducting the experiment or by employing a design that reduces the uncontrolled variation affecting treatment comparisons. There may be little avenue for improving the protocols — hopefully, the experimenter is already being as careful as possible. It might be possible to improve the design — for example, if a CRD is being used, it might be possible to change to an RCBD which has the effect of removing block difference from σ_U^2 and so reducing it. However, most often it is the number of replicates that will have to be increased.

X.C Computing the required sample size for the CRD and RCBD with a single treatment factor

As already noted, in the context of designed experiments, determining the sample size amounts to finding the number of pure replicates of a treatment that we will denote by *r*. It is achieved by computing the power for different values of *r*. If the computed power for a particular *r* is a) too high, decrease *r*, or b) too low, increase *r*, continuing in both cases until you have identified the smallest *r* that has at least the required power.

Given the discussion on the power of a hypothesis test, as previously outlined, in order to compute the required number of pure replicates you will have to specify:

1. the significance level, α;
2. the power desired, $1 - \beta$;
3. the number of treatments to be investigated
4. the minimum size of the difference to be detected between a pair of treatment means, as measured by Δ;
5. the uncontrolled variation, σ_U^2, expected.

In working out values for these often the results from a previous experiment you or another researcher has run will be useful.

Example X.1 Penicillin yield (continued)

We now determine the number of replicates required to achieve a power of 0.80 in detecting $\Delta = 5$ with $\alpha = 0.05$. We continue to take $\sigma_U^2 \approx 20$. Because 5 replicates achieved a power of only 0.22 and is clearly much smaller than required, we take $r = 15$ as our first guess and find that this would provide a power of 0.6860. Next we increase *r* to 20 and this indicates a power of 0.8289 will be achieved. As this is in excess of the required power, *r* is reduced to 19 and the power reduces to 0.8056. As this is still in excess of the required power, *r* is reduced to 18 and the power reduces to 0.7798. Clearly, 19 replicates is required to achieve a power of at least 0.80 — fewer will have less power. The output from this is as follows:

```
> rm <- 15
> power.exp(rm=rm, df.num=3, df.denom=3*(rm-1), delta=5, sigma=sqrt(20))
[1] 0.6860223
> rm <- 20
> power.exp(rm=rm, df.num=3, df.denom=3*(rm-1), delta=5, sigma=sqrt(20))
[1] 0.828869
> rm <- 19
> power.exp(rm=rm, df.num=3, df.denom=3*(rm-1), delta=5, sigma=sqrt(20))
[1] 0.8055926
> rm <- 18
> power.exp(rm=rm, df.num=3, df.denom=3*(rm-1), delta=5, sigma=sqrt(20))
[1] 0.779796
```

A more convenient method of computing the required number of pure replicates, *r*, is provided by the previously described R function `no.reps` — it uses the standard R function `optimize` to minimize the difference between the required power and the actual power of test numbers of replicates. The output from the use of `no.reps`, with `print =TRUE` to show the way it works, for the present example is as follows:

```
> no.reps(multiple=1, df.num=3, df.denom=expression(df.num*(r-1)), delta=5,
+ sigma=sqrt(20), power=0.8, print=TRUE)
        rm df.num df.denom alpha delta    sigma   lambda     powr
1 20.33437      3  58.0031  0.05     5 4.472136 12.70898 0.836116
        rm df.num df.denom alpha delta    sigma   lambda      powr
1 31.66563      3  91.9969  0.05     5 4.472136 19.79102 0.9676862
        rm df.num df.denom alpha delta    sigma  lambda      powr
1 13.33126      3 36.99379  0.05     5 4.472136 8.33204 0.6226844
        rm df.num df.denom alpha delta    sigma   lambda      powr
1 22.64998      3 64.94994  0.05     5 4.472136 14.15624 0.8795446
        rm df.num df.denom alpha delta    sigma   lambda      powr
1 17.65942      3 49.97826  0.05     5 4.472136 11.03714 0.7704095
        rm df.num df.denom alpha delta    sigma   lambda      powr
1 18.62379      3 52.87137  0.05     5 4.472136 11.63987 0.7961913
        rm df.num df.denom alpha delta    sigma   lambda      powr
1 18.92537      3 53.77612  0.05     5 4.472136 11.82836 0.8037564
        rm df.num df.denom alpha delta    sigma  lambda      powr
1 18.77872      3 53.33616  0.05     5 4.472136 11.7367 0.8001066
        rm df.num df.denom alpha delta    sigma   lambda      powr
1 18.74539      3 53.23616  0.05     5 4.472136 11.71587 0.7992694
        rm df.num df.denom alpha delta    sigma   lambda     powr
1 18.81205      3 53.43616  0.05     5 4.472136 11.75753 0.800941
        rm df.num df.denom alpha delta    sigma  lambda      powr
1 18.77872      3 53.33616  0.05     5 4.472136 11.7367 0.8001066
$nreps
[1] 19

$power
[1] 0.8055926
```

It confirms that the required number of replicates is 19 and that this will achieve a power of 0.8055. ■

X.D Sample size for the Latin square design

The formulae for computing the power and sample size for an RCBD also apply to the Latin square except that the number of replicates is *t*, rather than *b* as in the case of the RCBD. However, there is the restriction that the number of rows, columns and treatments must be equal. So you cannot increase the treatment replication without changing the number of treatments. Consequently, the main interest will be to determine if the proposed design will have the desired power, rather than

determining how many replicates must be taken to achieve the desired power. This can be done using the function `power.exp`.

X.E Sample size for factorial experiments

Sample size for a two-way factorial experiment can be computed using the R function `no.reps` in a similar manner to the single, treatment-factor experiments described in the previous sections. The first step is to determine which effects you wish to specify to be determined with a nominated power: A main, B main and/or interaction effects. Then you will need to determine the power for each set of effects. As shown in the following table, the cells in this table that vary for the different types of effects are: DF numerator; the multiplier, *m*, which when multiplied by the number of pure replicates of a treatment, *r*, gives the number of observations (rm) used in computing a mean for the source of interest; and delta (Δ).

Source	DF numerator	*m*	Δ
A	$a-1$	*b*	$\alpha_i - \alpha_k$
B	$b-1$	*a*	$\beta_j - \beta_\ell$
A#B	$(a-1)(b-1)$	*1*	$\left\{\left[(\alpha\beta)_{ij} - (\alpha\beta)_{kj}\right] - \left[\left(\overline{\alpha\beta}\right)_{i.} - \left(\overline{\alpha\beta}\right)_{k.}\right]\right\}$ or $\left\{\left[(\alpha\beta)_{ij} - (\alpha\beta)_{i\ell}\right] - \left[\left(\overline{\alpha\beta}\right)_{.j} - \left(\overline{\alpha\beta}\right)_{.\ell}\right]\right\}$

where $i \neq k$ and $j \neq \ell$.

Note that for A and B, Δ is the difference between a pair of A or B means. For the A#B interaction, Δ is the difference between the simple effect for one of the factors and the corresponding main effect for that factor. From the table above, $\left\{\left[(\alpha\beta)_{ij} - (\alpha\beta)_{kj}\right] - \left[\left(\overline{\alpha\beta}\right)_{i.} - \left(\overline{\alpha\beta}\right)_{k.}\right]\right\}$ represents the difference between an A simple effect and the A main effect. In particular $(\alpha\beta)_{ij} - (\alpha\beta)_{kj}$ is the simple effect between the *i*th and *k*th level of A at the *j*th level of B and $\left(\overline{\alpha\beta}\right)_{i.} - \left(\overline{\alpha\beta}\right)_{k.}$ is the main effect between the *i*th and *k*th level of A. We note that, if there was no interaction between A and B, then all the simple effects between a pair of levels of A would be equal to the main effect for that pair of levels. Consequently Δ measures how much interaction there is in the sense of how much the simple effects might differ from the main effects.

Example X.4 Animal survival experiment

In this example, the Poison and Treatment did not interact in their effect on the death rate. However, consider the following table of means that combines the Poison and Treatment main effect tables of means with the interaction table of means.

	Treat				
Poison	1	2	3	4	Mean
1	2.487	1.163	1.863	1.690	1.801
2	3.268	1.393	2.714	1.702	2.269
3	4.803	3.029	4.265	3.092	3.797
Mean	3.519	1.862	2.947	2.161	

Consider the main effect for Treatments 2 and 3: it is 2.947 – 1.862 = 1.085. So if there is no interaction, then we would expect the differences between Treatments 2 and 3 for each Poison to be about 1.085. They are not exactly all equal to this value. But, how different from 1.085 would be an important difference? This is what Δ is. ■

The DF denominator depends on the design employed in randomizing the treatments as given in the following table:

Design	DF denominator
CRD	$ab(r-1)$
RCBD	$(ab-1)(r-1)$
LS	$(r-1)(r-2)$

Example X.2 Fertilizing oranges

Suppose that for the experiment to investigate 3 levels of N and 2 levels of P, outlined in example VII.1, you wish to determine the number of blocks to use in an RCBD. You would like to be able to detect with 80% power, a difference of between a simple effect and a main effect of 10. You believe that the standard deviation will be about 7.5 and you will use a significance level of 5%. How many replicates are required to achieve the desired power?

So we require the number of blocks in an RCBD to detect a difference associated with the interaction. The R output for obtaining this is as follows:

```
> no.reps(multiple=1, df.num=2, df.denom=expression(5*(r-1)), delta=10, sigma=7.5,
+ power=0.80)
$nreps
[1] 12

$power
[1] 0.819291
```

So 12 blocks needed to detect, with 80% power, a difference of 10 between a simple effect and a main effect. ■

X.F Sample size for the standard split-plot experiment

Determining the sample size for the standard split-plot experiment is complicated because they involve two sources of uncontrolled variation: main-plot variation $\left(\sigma_{\mathrm{M}}^2\right)$ and subplot variation $\left(\sigma_{\mathrm{S}}^2\right)$. We require guesstimates of the magnitude of these two variances. Then sample sizes required to detect the various effects with specified power are computed as described for ordinary factorial experiments in section X.E except that the values of the variance $\left(\sigma^2\right)$ are varied as outlined in the following table.

Effects	σ^2	Δ
A main effects (e.g. Variety effects)	σ_{M}^2	$\alpha_i - \alpha_k$
B main effects (e.g. Fertilizer effects)	σ_{S}^2	$\beta_j - \beta_\ell$
Interaction effects		
– at the same level of A	σ_{S}^2	$\left\{\left[(\alpha\beta)_{ij} - (\alpha\beta)_{i\ell}\right] - \left[\left(\overline{\alpha\beta}\right)_{.j} - \left(\overline{\alpha\beta}\right)_{.\ell}\right]\right\}$
– not at the same level of A	$(b-1)\sigma_{\mathrm{S}}^2 + \sigma_{\mathrm{M}}^2$	$\left\{\left[(\alpha\beta)_{ij} - (\alpha\beta)_{kj}\right] - \left[\left(\overline{\alpha\beta}\right)_{i.} - \left(\overline{\alpha\beta}\right)_{k.}\right]\right\}$

X.G Summary

In this chapter we have:

- Pointed out that the conclusions from experiments are subject to the possibility of types I and II errors;
  1. type I errors are controlled by setting the significance level, α, for the experiment;
  2. the control of type II errors involves ensuring that the experiment has sufficient power and this depends on the number of treatments, the number of pure replicates and the magnitude of the uncontrolled variation;
- One way of ensuring sufficient power is to ensure the sample size, the number pure replicates, is large enough to achieve that power;
- Computing the sample size using R is summarized in Appendix B, *Randomized layouts and sample size computations in R*;
- In power and sample size calculations, in addition to specifying delta, sigma, power and alpha, one has to supply a number of quantities that vary with the design of the experiment. The following table summarizes these for the common designs, giving the degrees of freedom of the denominator as a function of *r*. Note that rm is the number of replicates in means being compared.

Design	m	rm	*df.num* (ν_1)	*df.denom* (ν_2)
CRD	1	r	$t-1$	$t(r-1)$
RCBD	1	b	$t-1$	$(t-1)(b-1)$
LS	1	$r(=t)$	$r-1$	$(r-1)(r-2)$
Factorial				
A	b	br	$a-1$	CRD $ab(r-1)$,
B	a	ar	$b-1$	RCBD $(ab-1)(r-1)$
A#B	1	r	$(a-1)(b-1)$	or LS $(r-1)(r-2)$
Standard split-plot				
A	b	br	$a-1$	$(a-1)(r-1)$
B	a	ar	$b-1$	$a(b-1)(r-1)$
A#B	1	r	$(a-1)(b-1)$	$a(b-1)(r-1)^{\dagger}$

†only approximate for effects not at the same level of A

X.H Exercises

X.1 A chemist is planning to run an experiment to test the effect of four chemical agents on the strength of a particular type of cloth. Because there might be variability from one bolt of cloth to another, the chemist decides to use a randomized complete block design, with bolts of cloth used as blocks.

She wants to be at least 90% sure that she can detect a difference of 4 in the mean tensile strength and knows from previous experiments that a standard deviation of about 1.5 can be expected. Also, she is prepared to take a 5% risk of making a type I error. How many bolts of cloth should she observe?

X.2 The effect of five different ingredients on the reaction time of a chemical process is to be studied. Each batch of material to be used in the process is only large enough to permit five runs to be made using it and at least five runs can be made in one day. In order to control for batch and day differences, it is decided to use a randomized complete block design in which each batch is used for a day.

The experimenter is intending to have 6 blocks, is willing to run a 5% chance of making a type I error and would like to have a 95% chance of detecting any difference of 6 minutes or more in the reaction time between ingredients. A variance of 3 minutes for the variation in runs on the same day is expected in the experiment. Will the experiment have the desired power?

X.3 An experimenter wants to investigate the effects of four different rations on the apparent consumption of total carbohydrates (as a percentage) by calves. He has available four calves of around 280 kg. He plans to use a Latin square for two reasons. Firstly, so that each calf receives the four rations, one in each of four periods. Secondly, so that differences, such as climatic differences, between the periods are eliminated from treatment differences. The experimenter is willing to run a 5% chance of making a type I error and would like to have a 95% chance of detecting any difference of 7.5% or more in the apparent consumption between rations. A variance of 10% for the animal-period combinations is expected in the experiment. Will the Latin square have the desired power?

X.4 In exercise X.3 you considered the power of an experiment to investigate the effects of four different rations on the apparent consumption of total carbohydrates (as a percentage) by calves. The design used a 4×4 Latin square. The values used in obtaining the power were that the experimenter is willing to run a 5% chance of making a type I error and would like to have a 95% chance of detecting any difference of 7.5% or more in the apparent consumption between rations. A variance of 10% for the animal-period combinations is expected in the experiment. For this design it was computed that the power would be just over 0.5 — not nearly enough!

Suppose the experimenter could obtain another 4 animals and so conduct the experiment with 8 animals over the 4 periods. Would this experiment have at least the level of power desired?

X.5 Suppose that an experiment is to be conducted to investigate the effects of four temperatures and three pressures on the yield of a chemical process. It is planned to use a completely randomized design to assign the treatments and it is believed that the population standard deviation is about 0.1.

a) It is desired to be able to detect a difference of at least 0.25 in the overall differences between a pair of Temperatures. How many replicates should be observed if the power is to be 0.90 and the level of significance 0.05?
b) What power would be achieved in detecting a difference between a simple and the main Temperature effects of at least 0.25 with the number of replicates that you have computed in a)? How many replicates would be required to detect this latter difference with power 0.90 and significance level 0.05?
c) Use R to obtain a randomized layout for this experiment with the number of replicates you finally computed in b). Use a seed of 312 in generating the design.