

The mathematical analysis of crossover designs

Sadegh Rezaei, B.Sc (Hons) (Ahwaz University), M.Sc (Tarbiat
Modarres University)

Thesis submitted for the degree of

Doctor of Philosophy

in

Statistics

at

The University of Adelaide

(Faculty of Mathematical and Computer Sciences)

Department of Statistics



November 18, 1997

Contents

Contents	ii
List of Tables	viii
List of Figures	xii
Summary	xiii
Signed Statement	xv
Acknowledgements	xvi
1 Introduction and literature review	1
1.1 Introduction	1
1.2 Definitions	3
1.3 Fields of application	3
1.4 Background and problems	4
1.5 Review of previous work on two-treatment designs	5
1.5.1 Optimality for crossover designs	9
1.6 Structure of the thesis	9
2 Models for crossover designs	12
2.1 Introduction	12
2.2 The model and information matrices	13

2.2.1	Mixed linear model	13
2.2.2	Information matrices	15
2.3	Estimating the parameters of interest	17
2.4	Building the model and information matrices using averages	18
2.4.1	Linear model for averages	20
2.4.2	Variance matrix of the group $\times$ period means	21
2.4.3	Idempotent matrices for strata	22
2.5	Analysis of variance	22
2.6	Analysis based on means	23
2.6.1	The vector of means	23
2.6.2	Combining information about treatment parameters	26
2.6.3	Summary of Chi's paper	27
2.7	Equal group sequence analysis	28
2.8	Conclusion	30
3	Analysis of two-treatment two-period crossover designs	31
3.1	Introduction	31
3.2	Two-treatment, two-period crossover design	32
3.2.1	Analysis of design when λ is present	33
3.2.2	Treatment information	35
3.2.3	Analysis of the design when λ is not present	37
3.2.4	Analysis of design when subject effect is fixed	38
3.2.5	A comparison of the three cases of 2×2 crossover design	39
3.3	Baseline measurements in the 2×2 crossover design	40
3.3.1	One baseline measurement in the design	41
3.3.2	Two baseline measurements in the design	45
3.3.3	Conclusion and discussion	48

3.4	Balaam's design for two treatments	49
3.4.1	Balaam's design when subject effects are random and unequal group sizes	50
3.5	Two-treatment, three-period crossover design	54
4	A Bayesian analysis for the general crossover design	57
4.1	Introduction	57
4.2	The statistical linear model	59
4.3	Likelihood and parameter estimation	60
4.4	Bayesian analysis	63
4.4.1	Σ_0 is a nonsingular matrix	63
4.4.2	Σ_0 is a singular covariance matrix	64
4.5	Choice of prior distribution	67
4.6	Two-treatment, two-period design	69
4.6.1	Posterior estimates	69
4.7	Two-treatment, two-period with one baseline measurement	71
4.8	Two-treatment, two-period with two baseline measurements	73
4.9	Bayesian analysis of Balaam's design	74
4.10	Three-period designs with two groups	76
5	Analysis of a two-period crossover design for the comparison of two active treatments and placebo	78
5.1	Introduction	78
5.2	The linear model	79
5.3	Treatment information	81
5.4	Combination of the estimates	83
5.5	Conclusions	84

6 Cohort designs for two-treatment crossover trials with one baseline measurement	87
6.1 Introduction	87
6.2 Age-period-cohort design	88
6.3 Building the model for standard and cohort designs	91
6.4 Standard design	95
6.4.1 Information matrices and parameter estimates	95
6.4.2 Analysis of variance for standard design	96
6.5 Cohort design	97
6.6 Treatment information	102
6.6.1 Information matrices on $\beta_1 = \tau_0 + \lambda_0$ and $\beta_2 = \tau_0 - \lambda_0$	104
6.6.2 Where is the treatment information?	105
6.7 Combining information	107
6.7.1 Estimation of (τ_0, λ_0) in cohort design only	107
6.7.2 Estimation of (τ_D, λ_D) in both designs	109
6.7.3 Estimates of treatment effect minus baseline	110
6.8 Limiting estimates in terms of ρ	111
6.8.1 The estimates and covariance matrix when $\rho \rightarrow \infty$	111
6.8.2 Estimates and their covariance when $\rho \rightarrow 0$	113
6.9 Conclusion	114
7 Block structure of cohort designs	116
7.1 Introduction	116
7.2 Cohort design in general	117
7.3 Analysis of variance	119
7.4 Fitting periods and cohorts	120
7.5 James and Wilkinson theorem	122

7.5.1	Canonical variables	123
7.6	Expected mean squares of each block structure	125
7.7	Description of a function <code>cc</code> in S-PLUS to get the result of James and Wilkinson's theorem	129
7.7.1	Projector matrices	132
7.8	To get a pattern on ρ and the Cohort (elim. Periods) contrasts	133
7.9	Conclusions	134
8	Treatment structure of cohort designs	135
8.1	Introduction	135
8.2	Treatment structure for cohort and standard designs in general	136
8.3	Projector matrices for cohort and standard designs in general	137
8.3.1	Estimates of parameters and covariance matrix of estimates	139
8.4	Treatment information for two-treatment crossover design with two base- line measurements and a corresponding cohort design	140
8.4.1	Treatment information when the first-order carryover is present	142
8.4.2	Treatment information when carryover effect is not present	144
8.5	Treatment information for two-treatment, extra-period crossover design and its corresponding cohort design with one baseline measurement	145
8.5.1	Treatment information when first-order carryover effect is present	147
8.5.2	Treatment information when the first-order carryover is not present	151
8.6	Treatment information for three-treatment and three-period crossover design	152
8.6.1	Treatment information when carryover effect is present	153
8.6.2	Treatment information when carryover effect is not present	155
8.7	Treatment information for two-treatment with one baseline measurement with more than two cohorts	156
8.7.1	Treatment information when $c = 3$ and first-order carryover effect is present	157

8.8	Conclusion	159
Appendices		161
A	Some useful concepts from linear algebra	161
A.1	The Kronecker product	161
A.2	Idempotent matrices	162
A.3	Positive definite quadratic forms and matrices	162
A.4	Contrast and orthogonal contrasts	162
A.5	ANOVA sum of squares as quadratic forms	162
A.6	Summing vectors, and E -matrices	164
B	Bayes' theorem	165
B.1	Normal prior for multinormal sample	165
C	The S-PLUS program for cohort designs	167
C.1	S-PLUS functions	167
C.2	Use of S-PLUS functions for Chapter 6	170
Bibliography		175

List of Tables

2.1	Layout of the design	13
2.2	Group $\times$ period means	19
2.3	ANOVA table for standard design, no treatment terms.	22
2.4	ANOVA table for standard design, using means, no treatment terms. . .	26
2.5	ANOVA table for standard design, equal size groups, no treatment terms.	29
3.1	Notation and layout for the simple crossover design	32
3.2	Expectation of responses in 2×2 crossover design when λ is in the model	33
3.3	Weighted orthogonal matrices for four strata in the design	35
3.4	Estimates of parameters and their variances.	39
3.5	Two-treatment design with one baseline measurement	41
3.6	Expectation of responses in 2×2 crossover design when λ is in the model	42
3.7	Weighted orthogonal matrices for four strata in the design	43
3.8	Two-treatment crossover design with two baseline measurements	45
3.9	Weighted orthogonal matrices for four strata in the design	47
3.10	The layout of Balaam's design with the number of subjects in each group.	50
3.11	Expected value of mean in Balaam's design	50
3.12	Weighted orthogonal matrices for four strata in Balaam's design	51
3.13	Two-treatment, three-period crossover design	54
4.1	Layout of general crossover design	58

4.2	Idempotent matrices of each stratum	59
4.3	Table of idempotent matrices with their ranks.	61
4.4	Expected values of responses in two-treatment, two-period design.	67
4.5	Expected values of responses when treatment B is a standard treatment.	67
5.1	Layout of design with two active treatments and placebo	79
5.2	Expected values of responses in comparison of two active treatments and placebo.	80
5.3	Idempotent matrices for four strata in the design including placebo.	81
6.1	Two-treatment three-period crossover design with baseline measurements	91
6.2	Cohort design in two-treatment three-period crossover design with one baseline measurement	92
6.3	Two-treatment three-period crossover design with one baseline measurement repeated as 4 groups of n	92
6.4	Treatment effects for the 12 means of observations for two-treatment three-period crossover design	93
6.5	Analysis of variance for standard design	97
6.6	Projector matrices for cohort design for two orders of fitting	101
6.7	Matrices introduced in the table of projector matrices for cohort design	101
6.8	Information matrix in each stratum for both designs	103
6.9	Information matrix about β_1 and β_2 in each stratum for both designs	105
6.10	Percentage of information in each stratum about $\tau_0 + \lambda_0$ and $\tau_0 - \lambda_0$ for the cohort design	105
6.11	Those strata which contribute to estimate some parameters of interest	107
6.12	Contrasts for estimating parameters in cohort design	108
6.13	Contrasts for estimating parameters in standard design	108
7.1	ANOVA table for standard design, no treatment terms.	119

7.2	ANOVA table for general cohort design, no treatments term	120
7.3	Helmert contrasts for Periods	130
7.4	Helmert contrasts for Cohorts	131
7.5	Contrasts for three cohorts and three periods	132
7.6	Canonical correlation coefficients and orthogonal contrasts for two cohorts and various periods	133
7.7	Canonical correlation coefficients and orthogonal contrasts for three co- horts and various periods	134
8.1	Projector matrices for general standard design	138
8.2	Projector matrices for general cohort design	138
8.3	Two-treatment crossover design with two baseline measurements	140
8.4	Cohort design in two-treatment, two-period crossover design with two baseline measurements.	141
8.5	Expected values of two-treatment, two-period crossover with two baseline measurements when we consider the cohort design or double standard design.141	141
8.6	Information matrices for standard and cohort designs in two-treatment design with two baseline measurements, when first-order carryover effect is present	143
8.7	Orthogonal contrast on Cohort (elim. Periods)	144
8.8	Information matrices for two-treatment, two-period standard and cohort designs when first-order carryover effect is not present.	145
8.9	Two-treatment, extra-period crossover design with one baseline measurement146	146
8.10	Cohort design for two-treatment, extra-period crossover design with one baseline measurement.	146
8.11	Expected values for design of two-treatment, extra-period crossover trial. 147	147
8.12	Information matrices for standard and cohort two-treatment, extra-period designs with one baseline measurement.	149

8.13	Treatment information for both standard and cohort two-treatment, extra-period design with one baseline when the carryover effect is not present. .	151
8.14	Three-treatment, three-period crossover design	152
8.15	Cohort design in three-treatment and three-period crossover design . . .	153
8.16	Expected values for three-treatment, three-period crossover and its corresponding cohort design	153
8.17	Information matrices for standard and cohort designs for three-treatment, three-period.	154
8.18	Treatment information for both standard and the cohort designs in three-treatment, three-period when the carryover effect is not present.	155
8.19	Two-treatment crossover design with one baseline measurement when $c = 3$	156
8.20	The 18 means for two-treatment with one baseline measurement when $c = 3$	156
8.21	Information matrices for standard and cohort designs for two-treatment with a baseline, $c = 3$	158
8.22	Information on τ_0 and λ_0 in three cases of design	159

List of Figures

3.1	N times variance of (i) direct treatment effect and of (ii) carry-over effect estimator for 2×2 design in terms of $\rho = 0, 1, 2, 5, 10, 20$ and $q = n_1/(n_1 + n_2)$.	40
4.1	Conditional distribution of carryover effect given direct effect	68
4.2	Relation between posterior estimates when $k = 0.2$	73
5.1	Standard deviation of $\hat{\tau}_D$ as ρ changes	86
6.1	Diagram of cohort design and its main concepts in crossover trial	89
6.2	The relationship of projector matrices in cohort design	99
7.1	Full cohort design length $cpgn$	118
7.2	Full cohort design, length cp	121
7.3	Cohort design for three periods in each cohort	130

Summary

The mathematical analysis of crossover designs

Sadegh Rezaei

Department of Statistics

The common theme of this thesis is the theory and application of crossover designs. In this thesis both classical and Bayesian approaches are considered. The thesis covers three broad areas.

In the classical approach, the methodology of repeated measurements is used to develop structure and models to describe the properties of crossover designs with different numbers of periods and unequal numbers of subjects in each group. The common approach in this part is to look at the information available about treatment parameters and to see in which strata this treatment information resides. Estimates are obtained by combining the available treatment information from different strata, in particular, from the between-subject and within-subject strata. The estimates are equivalent to the generalised least squares estimates.

To obtain the Bayesian analysis of crossover designs, prior distributions are chosen for carryover and direct treatment effects to reflect the current state of information about these parameters and the relationship between them. The assumption made in this thesis is that since the effect of a treatment dies away over time, we might expect that the carryover effect in the next period is some proportion, less than 1, of the original treatment effect. It is assumed that the a priori information is summarised by the fact that we expect λ to be a proportion k of the treatment effect τ , but with uncertainty described by a variance σ_0^2 , where σ_0^2 is known and k is some small positive value less than one. In addition it is supposed that the τ has an un-informative uniform prior

distribution.

Traditional crossover designs, even if baseline measurements are included, still do not allow estimation of the difference between the average treatment effect and baseline, unless one is prepared to make strong assumptions about the period effects, in terms of their size or likely variance. In the third part of the thesis, an approach is developed in which some subjects enter the trial with a delayed start. If the situation justifies the twin assumptions that period effects are related to calendar time (e.g. time of year), and that there are no effects due to the length of the time in the trial, referred to here as ‘age’, then such designs allow the estimation of the average effect of treatment versus baseline. The designs have similarities to age-period-cohort designs, and they are referred to here as ‘cohort designs’. They represent an interesting example of designs with non-orthogonal block structure.

Signed Statement

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

SIGNED: .

DATE: ..2/12/1997.....

Acknowledgements

I would like to thank my supervisor, Prof. Richard Jarrett, for his support, encouragement and enthusiasm throughout the development of this work. I also thank him for his accessibility at all hours in considering aspects of the work, and the many hours spent in discussion and proof-reading.

Many thanks go to Dr Bill Venables due to his assistance relating to computer programs in particular, S-PLUS and MAPLE enquiries.

I also would like to acknowledge the financial assistance of the Ministry of Culture and Higher Education (MCHE) of Islamic Republic of Iran during the period 26/4/92 to 26/1/96 in the form of a University of Shahid Chamran Scholarship.



Chapter 1

Introduction and literature review

1.1 Introduction

Clinical trials attempting to compare the efficacy of two treatments often use the two-treatment two-period crossover design, in which each subject is randomly assigned to receive both treatments in one of two sequences. This can lead to estimates of the difference in treatment efficacy that have variances which are less than half of those provided by the standard parallel design, in which each subject receives, by random allocation, one treatment only. This reduction of variance results from the elimination of between-subject variability. Another reason for choosing the crossover design is the reduced cost of performance of the experiment.

Designs in which experimental units receive more than one treatment application in the course of the experiment are called **crossover designs**. Such designs are known by different names in the literature: **changeover designs**, or **repeated measurement designs**, in some cases. In this design, the experiment is split up into different periods. Each subject receives one treatment in each period. Usually each subject is observed for the same number of periods. Designs are composed of several treatment sequences, each subject being allocated to a sequence at the start of the experiment. It is often assumed that equal numbers of subjects are allocated to each sequence.

The response variable is often supposed to be continuous. The analysis of crossover designs when the response is binary has been discussed in the literature by, for example, Gart (1969), Prescott (1981), Farewell (1985), Jones & Kenward (1987), Kenward &

Jones (1987b), but little has been written concerning the design of such experiments. Gart (1969) has given an exact test for comparing matched proportions in the analysis of binary responses in the crossover designs. Layard & Arvesen (1978) discuss the analysis of Poisson data in crossover designs, although this seems to be restricted for all practical purposes to the two-treatment, two-period crossover trial.

As Hedayat & Afsarinejad (1978) pointed out, the need for this design can be justified in several ways.

1. Due to budget limitation, the experimenter has to use each subject for several tests.
2. In some experiments, the subjects are human beings or animals and often the nature of the experiment is such that it calls for special training over a long period of time. Therefore, due to time limitation, one is forced to use each subject for several tests.
3. One of the objectives of the experiment is to find out the effect of different treatments on the same subjects in drug, nutrition or learning experiments.
4. Sometimes the subjects are rare, therefore the subjects have to be used repeatedly.

The crossover design is a special case of a randomised control trial and has some appeal to statisticians, medical and psychological researchers. The crossover design allows each subject to serve as their own control and this, in theory, should reduce the background level of variation affecting treatment comparisons. Direct evidence moreover, can be obtained about individual subject preference since each subject receives two or more different treatments.

In the simple case known as the two-period two-treatment crossover design or the 2×2 design, each subject receives A or B in the first period and the alternative in the succeeding period. The order in which A and B are given to each subject is randomised. Ideally, half of the subjects receive the sequence AB and half of the subjects receive the sequence BA . This is so that any trend from first-period to second-period can be eliminated in the estimate of differences in response. In any particular case, the numbers in the groups may not be identically equal for a variety of reasons, including drop-out and the nature of the randomisation which might necessarily be done sequentially as subjects enter the trial over time.

1.2 Definitions

The area of crossover design like any other area of statistics contains certain terms which are not found or used elsewhere. Terms like “**direct effect**” and “**residual effect**” or “**carryover effect**”, **washout period** and **baseline measurement** are the most commonly used ones. These terms are defined below.

1. The effect that a treatment has during the period in which it is applied is referred to the **direct treatment effect**.
2. The effect of a treatment that persists into the next treatment period is referred to as the **carryover effect**.
3. Sometimes, steps are taken by the experimenter to prevent or make less severe the occurrence of carryover effects by use of a waiting period, commonly called a **washout period**, between applications of treatments.
4. In some situations, the experimenter takes a measurement from the subject before a treatment is given to the subject. These measurements are known as **baseline measurements**.

Many authors have discussed the design and analysis of the two-period and two-treatment crossover design with and without baseline measurements or using washout period or using extra treatment periods or extra groups in comparing two treatments and have addressed the issue of carryover effects. The general conclusion of this work is that the presence of carryover effect invalidates the use of this crossover design, and that, unless carryover effects are negligible, a parallel design should be employed, or, if a crossover design has been used, that the analysis should be based only on first period data.

1.3 Fields of application

Crossover designs have had application over many years in a broad spectrum of research areas, including agriculture experiments, Cochran (1939), animal husbandry, Cochran et al. (1941), bioassay procedures, Finney (1978)) food science, market research, medicine, pharmacology, psychology. However, among applications in occupational psychology,

Parkes (1982) gives an interesting example where the crossover design occurs naturally. Various examples of the use of these designs in industry can be found throughout the literature, for example Raghavarao (1989). An important area where crossover designs are often used is in clinical trials and the pharmaceutical industry. One particular design, the two-treatment, two-period design has been extensively used and widely studied in the literature. The book by Jones & Kenward (1989) lists more than one hundred papers that have been written on this subject. But there are still several challenging and practically useful unsolved problems awaiting solution.

1.4 Background and problems

Despite the advantages mentioned above, the design has fallen somewhat into disrepute because of the possibility of a carryover effect, or a period-by-treatment interaction. A term for carryover treatment effect is introduced into the model to allow for the absence or inadequacy of washout periods. If the effect of treatment does persist into the period following the period of administration, then a carryover term is included in the model and the estimates of the direct treatment effect will be based on all the data of design. For a more detailed discussion of the issues involved, see Abeyasekera & Curnow (1984). Unfortunately, if such a carryover effect is present, naive estimation of the direct treatment effect will be potentially completely misleading and this design cannot give an unbiased estimate of direct treatment effect. The US Food and Drug Administration suggested that this design should not be used, unless unequivocal external evidence of the absence of carryover was available. Brown (1980) took a similar unfavourable view of this particular crossover design. There are many papers that react against this feeling about crossover in general, see Healy, in discussion of Lewis (1983), Barker et al. (1982), Patel (1983) and Willan (1988). Moreover, although it is theoretically possible to test for a carryover effect, in most small experiments the test is not at all powerful. In addition as Senn (1988) has argued “the significance tests for a carryover effect are a form of self-delusion” and he agreed that the justification for a crossover design must depend on medical opinion. In this regard he recommended using a washout period to achieve the aim.

The main purpose of a crossover design has been to devise designs that allow the

treatment effect to be estimated within-subject. If the subject effects are assumed to be fixed, then any estimator of direct treatment effect must be within-subject. Some authors assume that the subject effects are random; this assumption leads to a between-subject estimate for direct treatments as well, and this can be combined with the within-subject estimates, as described by Chi (1991).

1.5 Review of previous work on two-treatment designs

The analysis of crossover designs has been done in the literature using various methods. In parametric methods, a linear model is set up with all the parameters of interest and normal least squares techniques are used to obtain estimates of these parameters or parameters of interest, thus allowing hypothesis tests to be performed. Despite many variations of approach there is one point on which authors are in substantial agreement, namely the desirability of a preliminary check for the presence of carryover effects in the design. The development of the methods starts from the most elementary technique, given by Grizzle (1965), reviewed recently by Senn (1991), and continues through the various statistical methods proposed by Balaam (1968), who used four groups to compare just two treatments. Chassan (1964), Ebbutt (1984), Federer & Atkinson (1964), Fletcher (1987), Freeman (1989), Jones & Kenward (1989) consider this design with various features of analysis to get unbiased estimates for direct treatments. The fully Bayesian methods with informative prior for the parameter of interest were initiated by Grieve (1985), Grieve (1986).

The analysis of this design was given by Grizzle (1965) (with a subsequent correction in Grizzle (1974)), who focused on the simple two-period, two-treatment crossover design under the model in which subject effects are random. In this paper, Grizzle (1965) proposed a mixed model for univariate analysis of crossover design. In his model, the hypothesis of equal carryover (H_0) effects for two treatments is tested from between-subject variability and since it is regarded as a preliminary test, a relatively high level of significance is used. If H_0 is not rejected the data from both periods are used for testing the hypothesis of equal treatment effects. Otherwise, the use of the data from the first period alone is justified for treatment comparison, resulting in a loss of information.

Based on the restrictions of Grizzle's mixed model, several researchers, e.g. Wallenstein & Fisher (1977), have been led to respond to the use of crossover designs. They generally pointed out the main disadvantages of this design, that is, (i) the loss of information when the carryover is present and (ii) the low power of the preliminary test for carryover, as it is based on between-subject variability, and (iii) the increased chance for bias in the test for equal direct treatment effects derived from the data of both periods.

Zimmermann & Rahlfs (1978) proposed a bivariate normal model and analysed crossover design using a multivariate analysis of variance approach to the repeated measurements design. Their approach leads to tests identical to those in Grizzle's mixed model approach. They also proposed a method for testing equal carryover effect and direct effect simultaneously.

Koch (1972) described non-parametric methods of analysis for the case of two-treatment, two-period crossover designs. He proposed a number of non-parametric procedures for performing various hypothesis tests in connection with the Grizzle (1965) model of crossover trials. One of the various tests which he used was a rank test for direct treatment effect in the presence of carryover effects. Koch's procedure consists of ranking the period differences for all subjects in the design and then using the Wilcoxon test for differences between the two sequence groups. Cornell (1980) extended Koch's result but only to a small extent.

Brown (1980) compares the crossover and parallel group or completely randomised one-period designs in terms of the number of replicates required to achieve a given power and concluded that the crossover design can yield great savings in cost if the assumption of no carryover effect is valid, but the design should not be used if this assumption is in doubt. He also showed that the lack of power of the pretest leads with a relatively high probability to the non-rejection of the hypothesis of no carryover effects when they exist and the following analysis of the treatment effect in crossover designs is based on a test biased by carryover effect.

Hills & Armitage (1979) describe particularly clearly the usual method of analysis for two period crossover design with one measurement in each period per subject for an ordinal response. For more than two periods it would appear that these analyses are of little use.

There are some papers in the study of crossover design in which baseline measurements

are considered. Hills & Armitage (1979), Armitage & Hills (1982), Kershner & Federer (1981) and Federer & Atkinson (1964) all applied the baseline measurements and have limited their discussion to the two-treatment case only. Patel (1983) proposed to consider baseline measurements which might be obtained prior to each period in a two period crossover design. He showed that these measurements can be used in a preliminary test to determine the validity of a test for treatment comparison and also for testing the hypothesis of equal treatment effects.

Wallenstein (1979) showed that, if baseline observations are available before each period, the two-period two-treatment crossover design may be used for valid estimation of direct effect even in the presence of carryover effects, under certain constraints about period effects.

Kershner & Federer (1981) restricted their attention to only two treatments and compared the variance of contrasts for many higher-order crossover designs in presence of a mixed effect due to treatment sequences in the model.

We can find an excellent introduction to crossover designs in the book by Jones & Kenward (1989). Kenward & Jones (1987a), when they introduced the treatment-by-period-interaction, said that it is desirable that a check be made for it in the statistical analysis.

Senn (1991) however completely rejected the significance test for carryover and he said that despite many variations of approach for analysing the crossover design, there is one common opinion upon which most authors are in substantial agreement, namely the desirability of a preliminary check for the presence of carryover effects. He in his paper has argued that the significance test for carryover effect "is a form of self-delusion and the justification for a carryover design must be dependent on medical opinion as to whether the wash-out period can be regarded as achieving its aim."

Chi (1991) has shown that we can recover information on direct and carryover treatment effects from a between-subject analysis as is done in an incomplete block design with subject as blocks. He proved that by combining the within-subject and between-subject estimates, we can obtain the generalised least square estimate (GLSE), such that the GLSE estimation is a weighted combination of the within-subject and between-subject estimates, where the weight depends on the ratio of the between-subject variance component σ_s^2 to the within-subject variance component σ^2 . Laird et al. (1992) also consider

the combination of information from between and within subjects and apply their results to a number of standard designs, such as those of Balaam (1968) and Koch et al. (1989).

Freeman (1989) considered the usual analysis of two-period, two-treatment crossover design, that is, a two stage procedure in which (i) the presence of carryover is first tested and then (ii), according as the preliminary test is or is not significant at some pre-specified level of probability, we either use just data from the first period or use all the data assuming there is no differential carryover effect. Because the preliminary test for carryover is highly correlated with the analysis of data from the first period only, he showed that actual significance levels are higher than the nominal level even when there is no differential carryover. In order to make inference about the difference in treatment effects he compared three options:

- Procedure PAR, which uses the simple parallel group design, ignoring the second period altogether.
- Procedure CROS, which uses the differences between first and second periods, but assumes that no carryover effect is present.
- Procedure TS, that is, two-stage procedure for 2×2 crossover design proposed by Grizzle (1965).

He examined the two-stage procedure in terms of mean square error of point estimate, confidence intervals and actual significance level of hypothesis tests for the differences between the effects of the two treatments. He showed that even when no carryover effect exists, the actual significance level of the two-stage procedure of Grizzle (1965) is substantially larger than the nominal level. The reason is that the pre-test against carryover effect is highly correlated with the first period comparison test. Thus the conditional first period test is biased.

One of the main advantages of the crossover design compared with a simple completely randomised design is that, for a given precision, the study requires fewer subjects. In order to keep this essential advantage, Lasserre (1991) determined the optimal design for crossover design with fixed size of population. He paid special attention to the design with various numbers of periods for comparing two treatments. Willan & Pater (1986) and Willan (1988) develop approaches based on the assumption regarding the relationship between carryover and direct treatment effects.

Some authors in terms of solving the problems have chosen the Bayesian approach.

Selwyn et al. (1981) have presented a method of analysis for direct treatment effects in 2×2 crossover design with an equal number of subjects in each sequence and an un-informative prior on the variance component. The Bayesian approach leads to a magnitude and precision of the experimental estimate, rather than the classical approach implied by a preliminary test. Grieve (1985) has given a Bayesian analysis based on the Bayes factor against unequal carryover effects which provide a mixture of the two models corresponding to “absence of the carryover effect” and “presence of the carryover effect”, in which the weights are both a function of the data and of the likelihood of a carryover effect. Grieve (1986) in a recent paper extended the Bayesian approach for two-period crossover when there are two baseline measurements in the design. In his paper he developed the study of a two-period crossover for displaying the dependence of posterior inferences concerning the treatment effect on unavoidable prior beliefs about the correct model.

1.5.1 Optimality for crossover designs

With the restrictions on the number of periods, p , and the number of treatments, t , in a crossover design, optimality is defined as finding a design that gives a minimum variance treatment estimator.

Just for two treatments, Laska et al. (1983) pointed out that for any number of periods exceeding two, we can find optimal designs. They have shown that for even numbers of periods, balanced uniform designs are optimal and for odd numbers of periods, an extra period design can be used. These results were obtained as special case by Laska & Meisner (1985). Kershner & Federer (1981) calculated the variance of estimators for direct treatment, carryover and total treatment effects for a number of two treatment designs, using from two to four periods. The various optimality results pointed out are: the most efficient three-period design is ABB , BAA and the most efficient four-period design is $AABB$, $BBAA$, $ABBA$, $BAAB$.

1.6 Structure of the thesis

The first main purpose of this thesis is focused on the analysis of two-treatment crossover and using the cohort design in this trial with one and two baseline measurements.

The second main purpose of this thesis is concerned with developing a Bayesian approach to analysis of two-treatment crossover designs by using an informative Normal density for the parameters of interest.

Chapter 2 develops a general analysis of crossover designs, with the aim of showing where the treatment information lies in different strata. The focus is on combining all information in all strata to obtain appropriate estimates and showing that the estimate is equivalent to the generalised least squares estimate. In Chapter 3 we apply the results from Chapter 2 to the two-treatment, two-period crossover design without and with one and two baseline measurements, Balaam's design and the two-treatment, three-period crossover design.

Bayesian analysis of general crossover design with an informative prior density with full rank and non-full rank variance-covariance matrix for the prior density is dealt with in Chapter 4 and application of the results are shown in this chapter. For choosing the prior distribution for the parameters of interest, we assume that since the effect of a treatment dies away over time, we might expect that the carryover effect in the next period is some proportion of the original treatment effect. It is assumed that the a priori information is summarised by the fact that we expect λ to be a proportion k of the treatment effect τ , where k is a small positive value less than one. In addition it is supposed that the τ has an un-informative uniform prior distribution. Analysis of two-treatment, two-period design with a placebo is shown in Chapter 5.

In the classical analysis of crossover designs, even if baseline measurements are included, we could not estimate the difference between the average treatment effect and baseline, unless we make strong assumptions about the period effects, in terms of their size or likely variance. In Chapter 6 an approach is developed for situations in which some subjects enter the trial with a delayed start. If the situation justifies the twin assumptions that period effects are related to calendar time (e.g. time of year), and that there are no effects due to the length of the time in the trial, referred to here as 'age', then such designs, which are referred to here as cohort designs, enable us to estimate the average effect of treatment against baseline. We compare the results of the cohort design with the standard design. This alternative design is like an age-period-cohort design applied to the crossover trial with baseline in order to recover information about average effects of treatment and carryover. By making some orthogonal contrasts for two

important strata we get the information on the parameters of interest in each stratum. This treatment information can then be combined to get estimates of the parameters of interest for the two designs.

In Chapter 7, the block structure of the experiment is presented for the general class of cohort designs. We use the work of James & Wilkinson (1971), who provide a way of looking at the array of means and identifying the contrasts for cohorts after eliminating the effects of periods, i.e. Cohorts(el. Periods), and their expected mean squares. For this purpose we need to split the Cohort(el. Periods) stratum into different 1 degree of freedom contrasts each with a different expected mean square and then to identify the projector matrix for projecting onto the vector space spanned by the columns of Cohort(el. Periods). This is done in detail for cohort designs with two and three cohorts. For these special cases we obtain the analysis of variance tables and by using a function written in S-PLUS (Venables & Ripley (1996)) we show their projector matrices and the expected mean square for each contrast in the various strata.

In Chapter 8 we apply treatments to the non-orthogonal block designs of Chapter 7 and use the results of James and Wilkinson's theorem to estimate the treatment effects in a cohort design and then compare it to a corresponding standard design. We want generally to see where the treatment information goes and what treatment information is available by splitting the observations into the several strata and using the idempotent matrices. To get the estimates of parameters we combine the estimates of parameters from those strata in which there is some information about parameters.

Chapter 2

Models for crossover designs

2.1 Introduction

As we mentioned in Chapter 1 the crossover design can be considered as a repeated measurements design, since each subject is used on more than one occasion.

The common way of analysing crossover design is to consider subject effects as fixed effects. As Chi (1991) pointed out, Milliken & Johnson (1984) considered the subjects as blocks in an incomplete block design and have given the within-subject analysis with analysis of the averages across periods for all subjects. Gough (1989) has used the REML approach to recover between-subject information. Chi (1991) also recovered between-subject information and has shown that the combination of between-subject and within-subject information is equivalent to a generalised least squares analysis of crossover design.

In this chapter the methodology of repeated measurements and combining information from different strata is used to obtain some information about parameters of interest in the general crossover design with p periods and N subjects. The data can be presented in a table as shown in Table 2.1. Our aim is to look at the information available about treatment parameters and to see in which strata this treatment information resides.

We will get the estimates by combining information from the different strata and will show that the estimate is equivalent to the generalised least squares estimate.

2.2 The model and information matrices

Suppose we arrange the data in a vector, y , reading in order across the rows, so that $y = (y_{11}, \dots, y_{1p}, \dots, y_{ij}, \dots, y_{Np})'$ denotes the vector of all observations on the N subjects and p periods, such that y_{ij} is an observation on the i th subject and in the j th period, as shown in Table 2.1. Treatments, yet to be defined, will then be applied to each subject in each period. Some treatments may in fact be null treatments, corresponding to baseline measurements.

		Period					
		1	2	...	j	...	p
Subject	1						
	2						
	$\vdots$						
	i						
	$\vdots$						
	N						

Table 2.1: Layout of the design

2.2.1 Mixed linear model

The mixed linear model for this general case of the crossover design can be given as

$$y = \mu \mathbf{1} + P\pi + B\beta + T\theta + \epsilon, \quad (2.2.1)$$

where

- y is the $Np \times 1$ response vector from all subjects in the design,
- μ is the grand mean,
- π is the vector of period effects,
- β is the vector of subject effects, which can be considered as random effects, normally distributed and independent with mean zero and variance σ_s^2 ,
- θ is a vector of fixed effects including direct treatment and carryover effects and in some designs may include group and second carryover effects,
- $\mathbf{1}$ is an $Np \times 1$ vector with all elements 1,
- P is the period design matrix of dimension $Np \times p$,
- B is the subject design matrix of dimension $Np \times N$,
- T is the treatment design matrix of dimension $Np \times t$, where t is the number of treatment parameters in the design,
- ϵ is an $Np \times 1$ error vector whose elements are assumed normally distributed and independent with mean zero and variance σ^2 .

Hence the total number of observations is Np .

For P and B we can put

$$P = \mathbf{1}_N \otimes I_p \tag{2.2.2}$$

$$B = I_N \otimes \mathbf{1}_p$$

where $\otimes$ notes the Kronecker product for which the definition and properties are given in Appendix A, I is the identity matrix, and the subscripts denote the length or size of the matrices, as appropriate.

For example, the Kronecker product of $\mathbf{1}_N$ and I_p is a matrix of dimension $Np \times p$

given by

$$\mathbf{1}_N \otimes I_p = \begin{bmatrix} I_p \\ \vdots \\ I_p \end{bmatrix}_{Np \times p},$$

and the Kronecker product of I_N and $\mathbf{1}_p$ is a matrix of dimension $Np \times N$ given by

$$I_N \otimes \mathbf{1}_p = \begin{bmatrix} \mathbf{1}_p & 0 & \cdots & 0 \\ 0 & \mathbf{1}_p & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \mathbf{1}_p \end{bmatrix}_{Np \times N},$$

where each $\mathbf{0}$ is a column of length p .

In this formulation of the model

$$\begin{aligned} \text{Var}(y) &= \sigma^2 I_{Np} + \sigma_s^2 BB' \\ &= \sigma^2 I_{Np} + p\sigma_s^2 (I_N \otimes J_p) \\ &= \sigma^2 (I_N \otimes I_p) + p\sigma_s^2 (I_N \otimes J_p), \\ &= \sigma^2 (I_N \otimes K_p) + (\sigma^2 + p\sigma_s^2) (I_N \otimes J_p), \end{aligned} \tag{2.2.3}$$

where J_m will be used for the $m \times m$ idempotent matrix of rank 1, with all elements equal to $1/m$. The matrix $K_m = (I - J_m)$ will be an $m \times m$ idempotent matrix of rank $(m - 1)$. Thus, applied to any vector of the appropriate length, J replaces each element by the mean, and K replaces each element by itself minus the mean.

2.2.2 Information matrices

If there were no terms in the model other than θ , the estimate of θ would be $(T'T)^{-1}T'y$ and the variance matrix for the estimates would be $\sigma^2(T'T)^{-1}$. The Fisher information, given by the expected value of minus the second derivative of the log likelihood, is the inverse $(T'T)/\sigma^2$. As a matter of convention, we shall regard the information about θ as being given by the matrix $T'T$. As further terms are included in the model, in particular those whose design matrices are not orthogonal to the columns of the T , the information available for estimating θ is partially lost in estimating the additional terms. For example, the columns of T may not be orthogonal to the vector $\mathbf{1}$, in which case some of the treatment information in $T'T$ will be confounded with the grand mean. The

information so confounded is $T'J_{Np}T$. Then the available treatment information in the space orthogonal to the grand mean is

$$\mathcal{I}_\theta = T'KT.$$

The information about the parameters of interest θ can be thought of as residing in 4 different strata, corresponding to the following orthogonal idempotent matrices:

- $Q_0 = J_N \otimes J_p$ for grand mean stratum,
- $Q_1 = J_N \otimes K_p$ for period stratum,
- $Q_2 = K_N \otimes J_p$ for between subject stratum,
- $Q_3 = K_N \otimes K_p$ for subject $\times$ period stratum,

where we have defined idempotent matrices and their properties in Appendix A.

Information about θ is distributed into these 4 strata as

$$T'Q_iT \quad (i = 0, 1, 2, 3).$$

Since Q_i are orthogonal idempotent matrices, we can split the data into four separate components $Q_i y$ and show the properties of each and the relation between them. For the grand mean stratum, this component can be written as

$$Q_0 y = \mu \mathbf{1} + (\mathbf{1}_N \otimes J_p)\pi + (J_N \otimes \mathbf{1}_p)\beta + J_{Np}T\theta + J_{Np}\epsilon.$$

As noted before, the information about θ here is $T'JT$ and we can see that the single value obtained for the grand mean estimates

$$\mu + (\mathbf{1}'\pi)/p + (\mathbf{1}'T\theta)/(Np),$$

with a variance $(\sigma^2 + p\sigma_s^2)/(Np)$. Then $T'(J_N \otimes J_p)T$ is the information about θ contained in the grand mean. This is completely confounded with μ and hence in the absence of any knowledge about μ , there is no recoverable information about θ , and for this reason, the only treatment information available to us is $T'KT$.

In the period stratum, the component is

$$Q_1 y = (\mathbf{1}_N \otimes K_p)\pi + Q_1 T\theta + Q_1 \epsilon,$$

so that the treatment information is $T'Q_1T$, and $\text{Var}(Q_1 y) = \sigma^2 Q_1$. We note, however, that this treatment information is compromised by the presence of the period effects. The information in Periods is only available if we assume that period differences are not

present or that they have some known variance. There are unlikely ever to be enough periods to estimate the variance of an assumed random effect for periods.

Similarly, in the subject stratum we have,

$$Q_2y = (K_N \otimes \mathbf{1}_p)\beta + Q_2T\theta + Q_2\epsilon,$$

so that the treatment information is $T'Q_2T$, and $\text{Var}(Q_2y) = (\sigma^2 + p\sigma_s^2)Q_2$.

Finally for subject $\times$ period stratum we can write

$$Q_3y = Q_3T\theta + Q_3\epsilon,$$

so that the treatment information is $T'Q_3T$, and $\text{Var}(Q_3y) = \sigma^2$. We note that, since $Q_iQ_j = 0$ ($i \neq j$), the components in different strata are uncorrelated. Further,

$$T'KT = \sum_{i=1}^3 T'Q_iT,$$

so that the total treatment information is split between these strata, each with its own precision and

$$\text{Var}(y) = V = \sum_{i=0}^3 \delta_i Q_i,$$

so that the weighting for the orthogonal idempotent matrices corresponding to each stratum is given by

$$\begin{aligned} \delta_0 &= \delta_2 = \sigma^2 + p\sigma_s^2, \\ \delta_1 &= \delta_3 = \sigma^2. \end{aligned} \tag{2.2.4}$$

2.3 Estimating the parameters of interest

In the general case we can rewrite the mixed linear model in (2.4.1) in the following form

$$y = T\theta + \xi \tag{2.3.1}$$

where all other terms are absorbed into the covariance matrix of ξ , so that $\xi \sim N(0, V)$ such that V is an $N \times N$ positive definite variance matrix. In doing this, we assume for the moment that all other terms, even the period effects, can be represented by random effects. Then

$$\hat{\theta} = (T'V^{-1}T)^{-1}T'V^{-1}y.$$

If we can write

$$V = \sum_i \delta_i Q_i,$$

where Q_i is the orthogonal idempotent matrix of i th stratum, then we can write

$$V^{-1} = \sum_i \delta_i^{-1} Q_i,$$

and the information about θ can be thought of as coming from the different strata, as $T'Q_iT$, and providing separate estimates

$$\hat{\theta}_i = (T'Q_iT)^- T'Q_iy.$$

where $\{-\}$ notes the generalised inverse of a matrix. To get the overall estimate of θ , we can combine all information in the strata with appropriate weights ($1/\delta_i$), and hence we can obtain the following estimate for parameters of interest

$$\begin{aligned} \hat{\theta} &= \left(\sum_i \delta_i^{-1} T'Q_iT \right)^- \sum_i \delta_i^{-1} T'Q_iy \\ &= \left(\sum_i \delta_i^{-1} T'Q_iT \right)^- \left(\sum_i \delta_i^{-1} T'Q_iT \hat{\theta}_i \right). \end{aligned} \tag{2.3.2}$$

The variance of this combined estimate is

$$Var(\hat{\theta}) = \left(\sum_i \delta_i^{-1} T'Q_iT \right)^-.$$

Care needs to be taken in deciding which parts of the information are included. For example, we decided that the treatment information $T'Q_0T$ is not useful. Similarly, if we believe that there are nonzero period effects, then the information $T'Q_1T$ will also contain period effects and hence will not provide reliable unbiased estimates of θ . Hence, we would generally use only $i = 2, 3$ in recovering the information. This is equivalent to allowing $\delta_0, \delta_1 \rightarrow \infty$, or regarding μ and π as random effects whose variances approach zero.

2.4 Building the model and information matrices using averages

In later chapters we consider designs in which subjects are allocated at random into one of g groups, where all subjects in the same group get the same treatment regime. Suppose now that the N subjects are divided into g groups, with n_i ($i = 1, 2, \dots, g$) in each group,

such that $N = \sum_{i=1}^g n_i$. The data now needs to be indexed with three subscripts as y_{ijk} , where i refers to the group, j refers to the subject within the group, and k refers to the period. If we write the data values as a single vector in standard order, reading across the rows of the table of data, we have, analogously to equation (2.2.1),

$$y = \mu \mathbf{1} + P\pi + B\beta + T\theta + \epsilon, \quad (2.4.1)$$

where y is now a vector of length Np . The matrix T can be conveniently partitioned into a separate matrix for each of the g groups of subjects as

$$T = \begin{bmatrix} \mathbf{1}_{n_1} \otimes T_1 \\ \mathbf{1}_{n_2} \otimes T_2 \\ \vdots \\ \mathbf{1}_{n_g} \otimes T_g \end{bmatrix},$$

where T_i is a $p \times t$ matrix defining the treatment assignment for each subject in group i , for the p periods and the t treatment parameters in θ . We can then consider a reduced table which contains means for each group and each period. In Table 2.2 we show the group $\times$ period means.

		Period				
		1	2	...	p	Average on Periods
Group	1	$\bar{y}_{11.}$	$\bar{y}_{12.}$	...	$\bar{y}_{1p.}$	$\bar{y}_{1..}$
	2	$\bar{y}_{21.}$	$\bar{y}_{22.}$	...	$\bar{y}_{2p.}$	$\bar{y}_{2..}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	g	$\bar{y}_{g1.}$	$\bar{y}_{g2.}$	...	$\bar{y}_{gp.}$	$\bar{y}_{g..}$
Average on Group		$\bar{y}_{.1.}$	$\bar{y}_{.2.}$	...	$\bar{y}_{.p.}$	$\bar{y}_{...}$

Table 2.2: Group $\times$ period means

The period means here are $\bar{y}_{.j.} = \sum n_i \bar{y}_{ij.} / N$, and the grand mean is $\bar{y}_{...} = \sum n_i \bar{y}_{i..} / N$.

To identify where the treatment information resides, we give the following G matrix as the idempotent matrix which produces a vector Gy , where each value y_{ijk} in the original data vector is replaced by its group $\times$ period mean $\bar{y}_{i.k}$:

$$G = \begin{bmatrix} J_{n_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & J_{n_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \dots & \dots & J_{n_g} \end{bmatrix} \otimes I_p = W \otimes I_p. \quad (2.4.2)$$

Thus, Gy has length Np and contains the means repeated so that in the i th group and j th period, each mean occurs n_i times. We note that W is an $N \times N$ idempotent matrix.

2.4.1 Linear model for averages

If we apply the G matrix in the model in (2.4.1), then we have

$$Gy = G\mathbf{1}\mu + GP\pi + GB\beta + GT\theta + G\epsilon,$$

where the matrix coefficient of μ is

$$G\mathbf{1}_{Np} = (W \otimes I_p)(\mathbf{1}_{Np}) = (W \otimes I_p)(\mathbf{1}_N \otimes \mathbf{1}_p) = (W\mathbf{1}_N) \otimes (\mathbf{1}_p) = (\mathbf{1}_N) \otimes (\mathbf{1}_p) = \mathbf{1}_{Np},$$

because W is a diagonal matrix with the idempotent matrices as elements on diagonal.

Similarly, for the coefficient of π , we can write

$$GP = (W \otimes I_p)(\mathbf{1}_N \otimes I_p) = (W\mathbf{1}_N) \otimes I_p = \mathbf{1}_N \otimes I_p.$$

Furthermore,

$$GB = (W \otimes I_p)(I_N \otimes \mathbf{1}_p) = (W \otimes \mathbf{1}_p) \quad (2.4.3)$$

Let $\beta = (\beta'_1, \dots, \beta'_g)'$ be the vector of subject effects, where β_i is a vector of length n_i ($i = 1, \dots, g$). Then for the subject effect we can define

$$\beta_i^* = \frac{1}{n_i} \mathbf{1}'_{n_i} \beta_i,$$

where β_i^* is a scalar, the average of the elements of β_i . It follows that

$$\begin{aligned} GB\beta &= (W \otimes \mathbf{1}_p)\beta \\ &= \begin{bmatrix} J_{n_1} \otimes \mathbf{1}_p & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & J_{n_2} \otimes \mathbf{1}_p & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & J_{n_g} \otimes \mathbf{1}_p \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_g \end{bmatrix} \\ &= \begin{bmatrix} (J_{n_1}\beta_1) \otimes \mathbf{1}_p \\ (J_{n_2}\beta_2) \otimes \mathbf{1}_p \\ \vdots \\ (J_{n_g}\beta_g) \otimes \mathbf{1}_p \end{bmatrix} = \begin{bmatrix} \beta_1^* \mathbf{1}_{n_1} \otimes \mathbf{1}_p \\ \beta_2^* \mathbf{1}_{n_2} \otimes \mathbf{1}_p \\ \vdots \\ \beta_g^* \mathbf{1}_{n_g} \otimes \mathbf{1}_p \end{bmatrix} \\ &= \left\{ \begin{bmatrix} \frac{1}{n_1} \mathbf{1}_{n_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{1}{n_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \frac{1}{n_g} \mathbf{1}_{n_g} \end{bmatrix} \otimes \mathbf{1}_p \right\} \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_g^* \end{bmatrix} \end{aligned}$$

$$= B^* \beta^*,$$

say, where $B^* = \text{diag}(\frac{1}{n_1} \mathbf{1}_{n_1}, \dots, \frac{1}{n_g} \mathbf{1}_{n_g}) \otimes \mathbf{1}_p$. For the treatment coefficient matrix T we have

$$GT = \begin{bmatrix} J_{n_1} \otimes I_p & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & J_{n_2} \otimes I_p & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & J_{n_g} \otimes I_p \end{bmatrix} \begin{bmatrix} \mathbf{1}_{n_1} \otimes T_1 \\ \mathbf{1}_{n_2} \otimes T_2 \\ \vdots \\ \mathbf{1}_{n_g} \otimes T_g \end{bmatrix} \quad (2.4.4)$$

$$(2.4.5)$$

$$= T \quad (2.4.6)$$

Then the mixed linear model for this purpose can be given by

$$Gy = \mu \mathbf{1}_{Np} + (\mathbf{1}_N \otimes I_p) \pi + B^* \beta^* + T\theta + \epsilon^*. \quad (2.4.7)$$

where $\epsilon^* = G\epsilon$.

2.4.2 Variance matrix of the group $\times$ period means

The variance matrix of $B^* \beta^*$ can be given, using 2.4.3, by

$$\begin{aligned} \text{Var}(B^* \beta^*) &= \sigma_s^2 G B B' G' \\ &= \sigma_s^2 (W \otimes \mathbf{1}_p)(W \otimes \mathbf{1}_p') \\ &= p \sigma_s^2 (W \otimes J_p) \end{aligned}$$

and the variance matrix for ϵ^* can be expressed as

$$\text{Var}(\epsilon^*) = \sigma^2 G G' = \sigma^2 G = \sigma^2 (W \otimes I_p) \quad (2.4.8)$$

Then the variance-covariance of Gy is

$$\begin{aligned} \text{Var}(Gy) &= \sigma^2 (W \otimes I_p) + p \sigma_s^2 (W \otimes J_p) \\ &= \sigma^2 (W \otimes K_p) + (\sigma^2 + p \sigma_s^2) (W \otimes J_p). \end{aligned} \quad (2.4.9)$$

Since Gy provides the vector of group $\times$ period means, the matrix $(I - G)$ is the idempotent matrix which projects onto the space orthogonal to this. From (2.4.4), we note that $(I - G)T = 0$ and hence that this space has no treatment information, so all treatment information is in Gy .

2.4.3 Idempotent matrices for strata

The information about parameters of interest θ can now be thought of as residing in 6 different strata, corresponding to the following orthogonal idempotent matrices:

- $Q_0 = J_N \otimes J_p$ for Grand mean stratum,
- $Q_1 = J_N \otimes K_p$ for Periods stratum,
- $Q_2 = (W - J_N) \otimes J_p$ for Groups stratum,
- $Q_3 = (W - J_N) \otimes K_p$ for Group $\times$ Period stratum,
- $Q_4 = (I - W) \otimes J_p$ for Subject within Groups stratum,
- $Q_5 = (I - W) \otimes K_p$ for Period $\times$ (Subject within Groups) stratum.

Now for the analysis of the model we will consider these 6 idempotent matrices, noting that in the last two strata, that is, in Q_4 and Q_5 , there is no treatment information.

2.5 Analysis of variance

In the previous section some general structure was provided for the general crossover design. Now in this section we will provide the analysis of variance table for this design, which simply identifies the components in the block structure and ignores, for the moment, the treatment structure.

Source	Idempotent matrix	d.f	EMS
GM	$J_N \otimes J_p$	1	
Period	$J_N \otimes K_p$	$p - 1$	σ^2
Group	$(W - J_N) \otimes J_p$	$g - 1$	$\sigma^2 + p\sigma_s^2$
Group $\times$ Period	$(W - J_N) \otimes K_p$	$(g - 1)(p - 1)$	σ^2
Subject within Group	$(I - W) \otimes J_p$	$N - g$	$\sigma^2 + p\sigma_s^2$
Period $\times$ (subjects within Group)	$(I - W) \otimes K_p$	$(N - g)(p - 1)$	σ^2
Total		Np	

Table 2.3: ANOVA table for standard design, no treatment terms.

The data vector y of length Np , written in standard order, can be partitioned into various components, corresponding to the terms that can be identified in the linear model in (2.4.1), and the expected mean squares determined. The general ANOVA for this

design is shown in Table 2.3. We note again that the last two strata contain no treatment information because $T'\{(I - W) \otimes A\}T = 0$, where $A = J$ or $A = K$.

2.6 Analysis based on means

It will be convenient to work with the table of group $\times$ period means, shown in Table 2.2. An analysis of this table will give the first four rows of the Table 2.3, but we need to see how the sums of squares should be determined from the vector y^* of means.

2.6.1 The vector of means

To arrange the data in a vector say, y^* in terms of averages, the vector y should be pre-multiplied by a matrix, G^* , of dimension $gp \times Np$. This matrix is similar to G except that it does not have repeated rows. Thus the G^* matrix can be given as

$$G^* = \begin{bmatrix} \frac{1}{n_1} \mathbf{1}'_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{1}{n_2} \mathbf{1}'_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \cdots & \frac{1}{n_g} \mathbf{1}'_{n_g} \end{bmatrix} \otimes I_p = (W^* \otimes I_p), \quad (2.6.1)$$

so that the G^* matrix produces the means in a vector y^* of length gp and contains just the means so that, in the i th group and j th period, each mean occurs once. We want to reproduce the analysis of variance using y^* , but with just grand mean, period, group and group $\times$ period strata. Now in terms of G^* we can write

$$y^* = G^* y.$$

In multiplying the model (2.4.1) by G^* , we have

$$G^* \mathbf{1}_{Np} = (W^* \otimes I_p)(\mathbf{1}_N \otimes \mathbf{1}_p) = \mathbf{1}_g \otimes \mathbf{1}_p = \mathbf{1}_{gp},$$

$$G^* P = (W^* \otimes I_p)(\mathbf{1}_N \otimes I_p) = (\mathbf{1}_g \otimes I_p),$$

and

$$G^* B = (W^* \otimes I_p)(I_N \otimes \mathbf{1}_p) = (W^* \otimes \mathbf{1}_p).$$

For $T^* = G^*T$, we can write

$$\begin{aligned}
G^*T &= \left\{ \begin{bmatrix} \frac{1}{n_1} \mathbf{1}' & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{1}{n_1} \mathbf{1}' & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \frac{1}{n_g} \mathbf{1}' \end{bmatrix} \otimes I_p \right\} \begin{bmatrix} \mathbf{1} \otimes T_1 \\ \mathbf{1} \otimes T_2 \\ \vdots \\ \mathbf{1} \otimes T_g \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{n_1} \mathbf{1}' \otimes I_p & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{1}{n_1} \mathbf{1}' \otimes I_p & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \frac{1}{n_g} \mathbf{1}' \otimes I_p \end{bmatrix} \begin{bmatrix} \mathbf{1} \otimes T_1 \\ \mathbf{1} \otimes T_2 \\ \vdots \\ \mathbf{1} \otimes T_g \end{bmatrix} \\
&= \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_g \end{bmatrix} = T^*.
\end{aligned}$$

Then the mixed linear model for the analysis with means can be expressed as

$$y^* = \mu \mathbf{1}_{gp} + (\mathbf{1}_g \otimes I_p) \pi + (W^* \otimes \mathbf{1}_p) \beta + T^* \theta + G^* \epsilon, \quad (2.6.2)$$

such that

$$\begin{aligned}
\text{Var}(y^*) &= \text{Var}(G^* \epsilon) + \text{Var}(G^* B \beta) \\
&= \sigma^2 (W^* \otimes I_p) (W^{*'} \otimes I_p) + \sigma_s^2 (W^* \otimes \mathbf{1}_p) (W^{*'} \otimes \mathbf{1}_p') \\
&= \sigma^2 (W^* W^{*'} \otimes I_p) + p \sigma_s^2 (W^* W^{*'} \otimes J_p) \\
&= \sigma^2 (\aleph^{-1} \otimes I_p) + p \sigma_s^2 (\aleph^{-1} \otimes J_p) \\
&= \sigma^2 (\aleph^{-1} \otimes K_p) + (\sigma^2 + p \sigma_s^2) (\aleph^{-1} \otimes J_p)
\end{aligned}$$

where

$$\aleph = (W^* W^{*'})^{-1} = \begin{bmatrix} n_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & n_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & n_g \end{bmatrix}.$$

We now take each term in Table 2.3 and write it in terms of y^* rather than y . All Q'_i s can be written in the form $W \otimes A$, $J \otimes A$ or $I \otimes A$, where $A = J_p$ or $A = K_p$. Thus, the

sum of squares of $W \otimes A$ is

$$SS_{WA} = y^{*'} W \otimes A y^*$$

and, since for W we can write

$$\begin{aligned} W &= \begin{bmatrix} \frac{1}{n_1} \mathbf{1}_{n_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{1}{n_2} \mathbf{1}_{n_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \frac{1}{n_g} \mathbf{1}_{n_g} \end{bmatrix} \begin{bmatrix} n_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & n_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & n_g \end{bmatrix} \begin{bmatrix} \frac{1}{n_1} \mathbf{1}'_{n_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \frac{1}{n_2} \mathbf{1}'_{n_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \frac{1}{n_g} \mathbf{1}'_{n_g} \end{bmatrix} \\ &= W^{*'} \begin{bmatrix} n_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & n_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & n_g \end{bmatrix} W^* = W^{*'} \aleph W^*, \end{aligned}$$

we can write

$$\begin{aligned} SS_{WA} &= y'(W^{*'} \otimes I)(\aleph \otimes A)(W^* \otimes I)y \\ &= y^{*'}(\aleph \otimes A)y^*. \end{aligned}$$

Thus, for $A = J_p$ and $A = K_p$, the sum of squares are

$$\begin{aligned} SS_{WJ} &= y^{*'}(\aleph \otimes J)y^*, \\ SS_{WK} &= y^{*'}(\aleph \otimes K)y^*. \end{aligned}$$

Similarly, the sum of squares for the grand mean and periods can be given as

$$SS_{JA} = y'(J \otimes A)y,$$

where A is replaced by J_p and K_p , respectively.

To get the above sum of squares in terms of y^* , the vector of means, we note that $W^* \aleph \mathbf{1}_g = \mathbf{1}_N$, so that

$$\begin{aligned} J_N &= \frac{1}{N} \mathbf{1}_N \mathbf{1}'_N \\ &= \frac{1}{N} W^{*'} \aleph \mathbf{1}_g \mathbf{1}'_g \aleph W^* \end{aligned}$$

and we can write

$$\begin{aligned} SS_{JA} &= \frac{1}{N} y'(W^{*'} \otimes I)(\aleph \mathbf{1}_N \mathbf{1}'_N \aleph \otimes A)(W^* \otimes I)y \\ &= \frac{1}{N} y^{*'}(\aleph \mathbf{1} \mathbf{1}' \aleph \otimes A)y^*. \end{aligned}$$

If we put $J^* = \frac{1}{N}(\mathbb{N}\mathbb{1}\mathbb{1}'\mathbb{N})$, then the sums of squares can be presented as $y^{*'}Q_i y^*$, where

- $Q_0^* = J_g^* \otimes J_p$ for the Grand Mean stratum,
- $Q_1^* = J_g^* \otimes K_p$ for the Period stratum,
- $Q_2^* = (\mathbb{N} - J_g^*) \otimes J_p$ for the Group stratum,
- $Q_3^* = (\mathbb{N} - J_g^*) \otimes K_p$ for the Group $\times$ Period stratum.

These matrices Q_i^* are no longer orthogonal idempotent matrices because the elements of y^* have different accuracy. They are however orthogonal with respect to the weight matrix $\mathbb{N}^{-1} \otimes I$, such that

$$Q_i^*(\mathbb{N}^{-1} \otimes I)Q_j^* = \begin{cases} 0 & i \neq j \\ Q_i^* & i = j \end{cases}$$

and

$$\sum_i Q_i^* = (\mathbb{N}^{-1} \otimes I).$$

The analysis of variance table corresponding to the first four lines in Table 2.3 can be given in Table 2.4.

Source	Sum of Squares	d.f	EMS
GM	$y^{*'}(J_g^* \otimes J_p)y^*$	1	$\sigma^2 + p\sigma_s^2$
Period	$y^{*'}(J_g^* \otimes K_p)y^*$	$p - 1$	σ^2
Group	$y^{*'}[(\mathbb{N} - J_g^*) \otimes J_p]y^*$	$g - 1$	$\sigma^2 + p\sigma_s^2$
Group $\times$ Period	$y^{*'}[(\mathbb{N} - J_g^*) \otimes K_p]y^*$	$(g - 1)(p - 1)$	σ^2
Total		gp	

Table 2.4: ANOVA table for standard design, using means, no treatment terms.

2.6.2 Combining information about treatment parameters

The total information about treatment parameters can be given as

$$T'T = \sum_i n_i T_i' T_i = T^{*'}(\mathbb{N} \otimes I)T^*,$$

and the treatment information in each stratum can be obtained by applying the methods of Section 2.6.1 using the fact $T^* = G^*T$. We then obtain

- $T^{*'}(J_g^* \otimes J_p)T^*$ in Grand Mean stratum,
- $T^{*'}(J_g^* \otimes K_p)T^*$ in Period stratum,

- $T^{*'}[(\mathbb{I} - J_g^*) \otimes J_p]T^*$ in Group stratum,
- $T^{*'}[(\mathbb{I} - J_g^*) \otimes K_p]T^*$ in Group $\times$ Period stratum.

As mentioned before there is no treatment information in the last two strata in Table 2.3. Now the estimation of treatment parameters in Grand mean, Period, Group and Group $\times$ Period strata respectively can be given as

$$\hat{\theta}_i = (T^{*'}Q_i^*T^*)^{-1}T^{*'}Q_i^*y^*, \quad (2.6.3)$$

and the overall estimate of θ can be obtained by combining the information from those strata which have useful treatment information, then

$$\hat{\theta} = (\sum_i \delta_i^{-1} T^{*'}Q_i^*T^*)^{-1} (\sum_i \delta_i^{-1} T^{*'}Q_i^*T^*\hat{\theta}_i), \quad (2.6.4)$$

and the covariance matrix of this estimate can be given by

$$Var(\hat{\theta}) = (\sum_i \delta_i^{-1} T^{*'}Q_i^*T^*)^{-1} \quad (2.6.5)$$

If we regard μ and π as fixed effects then the information in the Grand Mean and the Period strata give one observation or mean for each of the unknown parameters in μ and π . Hence, unless we know something about the values or distributions of μ and π , the strata represented by the Grand Mean and the Periods provide no accessible information about the treatment parameters θ . We therefore only recover information from Q_2^* and Q_3^* with weights δ_2 and δ_3 .

2.6.3 Summary of Chi's paper

Chi (1991) in his paper formalises the between- and within-subject analysis for a general crossover design and shows that the combined estimates are equivalent to the generalised least squares estimates. He defines the following mixed linear model for a general crossover design

$$y = T\theta + B\beta + e,$$

where θ is the vector of effects of parameters of interest includes the period effect π , and β is the vector of subject effects. He obtains the following within-subject estimate of θ , considering β in the mixed linear model as fixed effects:

$$\hat{\theta}_3 = \{T'[I - B(B'B)^{-1}B']T\}^{-1}T'[I - B(B'B)^{-1}B']y.$$

For the estimate of variance σ^2 he obtains

$$\hat{\sigma}^2 = y'[I - B(B'B)^{-1}B'](y - T\hat{\theta}_3)/[(p-1)(N-1) - 2(L-1)],$$

with $\hat{\beta} = (B'B)^{-1}B'(y - T\hat{\theta}_3)$ where N is the total number of subjects in the design and L is the number of direct treatments. The covariance matrix of $\hat{\theta}_3$ is given as

$$\hat{\Sigma}_3 = \hat{\sigma}^2 \{T'[I - B(B'B)^{-1}B']T\}^{-1}.$$

For the between-subject analysis, he assumes that the subject effects, β , formed a random vector distributed as $N(0, \sigma_s^2 I)$. Then the covariance structure for Y is:

$$\Sigma = \text{Var}(y) = \sigma_s^2 BB' + \sigma^2 I,$$

and he obtains the between subject estimate of θ as

$$\hat{\theta}_2 = \{T'B(B'B)^{-1}B'T\}^{-1} \{T'B(B'B)^{-1}B'y\}.$$

Chi then defines $\sigma_b^2 = \sigma_s^2 + \sigma^2/p$, and obtains the following estimate

$$\hat{\sigma}_b^2 = (Z'Z - Z'Q\hat{\theta}_2)/(N - L^*),$$

where $Z = B(B'B)^{-1}B'y$, $Q = B(B'B)^{-1}B'T$, and L^* is the rank of Q . The covariance matrix of $\hat{\theta}_2$ is

$$\hat{\Sigma}_2 = p\hat{\sigma}_b^2 [T'B(B'B)^{-1}B'T]^{-1}.$$

Finally he combines the two estimators of θ and got the following generalised least squares estimators for θ

$$\begin{aligned} \hat{\theta} &= (\hat{\Sigma}_2^{-1} + \hat{\Sigma}_3^{-1})^{-1}(\hat{\Sigma}_2^{-1}\hat{\theta}_2 + \hat{\Sigma}_3^{-1}\hat{\theta}_3) \\ &= (T'\hat{\Sigma}^{-1}T)^{-1}T'\hat{\Sigma}^{-1}y, \end{aligned}$$

where, $\hat{\Sigma} = \hat{\sigma}_b^2 BB' + \hat{\sigma}^2 I$ and $\hat{\sigma}_b^2 = \hat{\sigma}_b^2 - \hat{\sigma}^2/p$. He also shows that the combination of the two estimates of θ is equivalent to generalised least squares.

2.7 Equal group sequence analysis

If we assume that $n_1 = n_2 = \dots = N/g = n$, that is, if we assume that the number of subjects is equally allocated to each group, then the analysis of crossover designs simplifies. The linear mixed model for the means can be expressed as

$$y^* = \mu \mathbf{1}_{gp} + (\mathbf{1}_g \otimes I_p)\pi + (I_g \otimes \mathbf{1}_p)\beta^* + T^*\theta + (I_g \otimes I_p)\epsilon^*, \quad (2.7.1)$$

where $\epsilon^* = G^*\epsilon$ has elements which are $N(0, \sigma^2/n)$, and β^* has elements which are $N(0, \sigma_s^2/n)$. Then we have

$$\begin{aligned} \text{Var}(y^*) &= \text{Var}[(I_g \otimes I_p)\epsilon^*] + \text{Var}[(I_g \otimes \mathbf{1}_p)\beta^*] \\ &= \frac{\sigma^2}{n}(I_g \otimes I_p) + \frac{p\sigma_s^2}{n}(I_g \otimes J_p) \\ &= \frac{\sigma^2}{n}(I_g \otimes K_p) + \frac{\sigma^2 + p\sigma_s^2}{n}(I_g \otimes J_p), \end{aligned}$$

The Fisher information available in the space orthogonal to the grand mean is

$$n \sum_i T_i' K T_i$$

and the data vector distributes into the following strata

- $Q_0^* = nJ_g \otimes J_p$ for Grand Mean stratum,
- $Q_1^* = nJ_g \otimes K_p$ for Period stratum,
- $Q_2^* = nK_g \otimes J_p$ for Group stratum,
- $Q_3^* = nK_g \otimes K_p$ for Group $\times$ Period stratum,

which are orthogonal with respect to the weight matrix $(\frac{1}{n}I)$ and applied to the vector of means y^* , and

- $Q_4 = K_n \otimes I_g \otimes J_p$ for Subject within Group stratum,
- $Q_5 = K_n \otimes I_g \otimes K_p$ for Period $\times$ (Subject within Group) stratum,

which are applied to the data vector y . Then the ANOVA table for this case of design is shown in Table 2.5.

Source	Sum of Squares	d.f	EMS
GM	$ny^{*'}(J_g \otimes J_p)y^*$	1	
Period	$ny^{*'}(J_g \otimes K_p)y^*$	$p - 1$	σ^2
Group	$ny^{*'}(K_g \otimes J_p)y^*$	$g - 1$	$\sigma^2 + p\sigma_s^2$
Group $\times$ Period	$ny^{*'}(K_g \otimes K_p)y^*$	$(g - 1)(p - 1)$	σ^2
Subject within Group	$y'(K_n \otimes I_g \otimes J_p)y$	$g(n - 1)$	$\sigma^2 + p\sigma_s^2$
Period $\times$ (subjects within Group)	$y'(K_n \otimes I_g \otimes K_p)y$	$g(n - 1)(p - 1)$	σ^2
Total		ngp	

Table 2.5: ANOVA table for standard design, equal size groups, no treatment terms.

The treatment information in the i th stratum is the $T^{*'}Q_i^*T^*$, and the estimates and their covariance matrices are as given in Section 2.6.2.

2.8 Conclusion

This chapter provides a framework for the analysis of design in which multiple measurements are made on each subject and the variance structure is modelled by random subject effects and random measurement errors at each period. Treatment regimes are applied to groups of subjects, possibly with unequal numbers in each group. In the next chapter, we will apply these methods to some particular designs.

Chapter 3

Analysis of two-treatment two-period crossover designs

3.1 Introduction

This chapter is concerned with two-treatment, two-period crossover design, in which each of several groups of subjects receive a different sequence of treatments. In clinical trials to compare the effects of several treatments, large between- subject variability often reduces efficiency. To avoid that problem, resorting to crossover designs has become popular since they use each experimental subject repeatedly, and tests of treatment and carryover effects can often be performed using within subject variation leading to more powerful analysis.

In this chapter, we will analyse various aspects of two-treatment, two-period crossover designs, and have a look at some modifications such as Balaam designs and the use of baseline measurements, which give us new information about parameters in the model and resolve the difficulty which we face in the simple 2×2 crossover design.

The simplest design is the one known as the 2×2 design. In this design, subjects are divided into two groups at random. One group receives treatment A followed by treatment B , and the other group receives the treatments in the reverse order. We will analyse this design in Section 3.2 just to see the application of Chapter 2 to a well-known situation which is analysed by Jones & Kenward (1989). In Section 3.3 we will consider this design with one and two baseline measurements. The Balaam design will

be considered in Section 3.4 and in Section 3.5 we will give a modification which shows the optimum design to solve some problems in the 2×2 crossover design.

3.2 Two-treatment, two-period crossover design

Suppose we have the 2×2 crossover design with two sequences of treatments AB and BA . One measurement or observation is obtained per subject per period in the standard crossover design, although this measurement might itself be an average of several measurements of responses taken during the period. Table 3.1 shows the layout for the data.

Group	Subject	Period 1	Period 2
1	β_{11} $\vdots$ β_{1n_1}	Treatment	
		A	B
		y_{111} $\vdots$ y_{11n_1}	y_{121} $\vdots$ y_{12n_1}
		Treatment	
2	β_{21} $\vdots$ β_{2n_2}	B	A
		y_{211} $\vdots$ y_{21n_2}	y_{221} $\vdots$ y_{22n_2}
		Treatment	
		A	B

Table 3.1: Notation and layout for the simple crossover design

The subjects are divided into two groups of sizes n_1 and n_2 such that the n_1 subjects in group 1 receive the treatments in the order AB and the n_2 subjects in group 2 receive the treatments in the order BA . Following the method in Chapter 2, we assume that $\mathbf{Y}_{ijk}$ is a random variable with the observed value y_{ijk} which follows the mixed linear model given in (2.2.1). Thus, for the two-treatment, two-period crossover design the model terms for k th subject in group 1 would be:

$$y_{11k} = \mu + \beta_{1k} + \pi_1 + \tau_1 + e_{11k}, \quad (3.2.1)$$

$$y_{12k} = \mu + \beta_{1k} + \pi_2 + \tau_2 + \lambda_1 + e_{12k}.$$

and for group 2 they would be:

$$\begin{aligned} y_{21k} &= \mu + \beta_{2k} + \pi_1 + \tau_2 + e_{21k}, \\ y_{22k} &= \mu + \beta_{2k} + \pi_2 + \tau_1 + \lambda_2 + e_{22k}. \end{aligned} \quad (3.2.2)$$

We assume that the subject effects, β_{ik} are random variables from a Normal distribution with mean zero and variance σ_s^2 and the errors, e_{ijk} are i.i.d random variable from $N(0, \sigma^2)$. We also assume that the subject effects and errors are independent. In other words, we consider σ^2 as the variance within subjects and σ_s^2 as the variance between subjects. We define $N = n_1 + n_2$ as the total of subjects in the design. In order to make the parameter values unique, we introduce the following constraints on treatment parameters:

$$\begin{aligned} \tau_1 &= -\tau_2 = \tau, \\ \lambda_1 &= -\lambda_2 = \lambda. \end{aligned} \quad (3.2.3)$$

This section is divided into three parts. The 2×2 design is considered when λ is present and when it is not present with the subject effect as a random effect. We then discuss the case with λ in the model and the subject effect as a fixed effect.

3.2.1 Analysis of design when λ is present

With the constraints in (3.2.3), the expectations of responses in this case can be written in Table 3.2.

		Period	
		1	2
Group	1	$\mu + \pi_1 + \tau$	$\mu + \pi_2 - \tau + \lambda$
	2	$\mu + \pi_1 - \tau$	$\mu + \pi_2 + \tau - \lambda$

Table 3.2: Expectation of responses in 2×2 crossover design when λ is in the model

As in Section 2.6, we can examine the treatment information and obtain treatment estimates by looking at the vector of means y^* . This can be written in matrix notation

as:

$$\begin{bmatrix} \bar{y}_{11.} \\ \bar{y}_{12.} \\ \bar{y}_{21.} \\ \bar{y}_{22.} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ -1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \tau \\ \lambda \end{bmatrix} + \begin{bmatrix} \bar{\epsilon}_{11.} \\ \bar{\epsilon}_{12.} \\ \bar{\epsilon}_{21.} \\ \bar{\epsilon}_{22.} \end{bmatrix}, \quad (3.2.4)$$

where the dot notation in the formulas denotes averages and each mean is formed by averaging over the subjects in that group and that period. With the matrix notation of Chapter 2, we can write out the above mixed linear model as

$$y^* = \mu \mathbf{1}_4 + (\mathbf{1}_2 \otimes I_2) \pi + B^* \beta^* + T^* \theta + \epsilon^*, \quad (3.2.5)$$

where y^* and θ are known as data vector and treatment parameters respectively and $B^* = W^* \otimes \mathbf{1}_p$, where

$$W^* = \begin{bmatrix} \frac{1}{n_1} \mathbf{1}'_{n_1} & 0 \\ 0 & \frac{1}{n_2} \mathbf{1}'_{n_2} \end{bmatrix}.$$

The covariance matrix of Y^* can be obtained by

$$Var(y^*) = \sigma^2 (\aleph^{-1} \otimes K_2) + (\sigma^2 + 2\sigma_s^2) (\aleph^{-1} \otimes J_2) \quad (3.2.6)$$

where

$$\aleph = (W^* W^{*'})^{-1} = \begin{bmatrix} n_1 & 0 \\ 0 & n_2 \end{bmatrix}. \quad (3.2.7)$$

By referring to Chapter 2 and incorporating random subject effects, we can get the information matrices about θ in the four strata which we introduced in Chapter 2. Since we reduced the responses to means of observations over the subjects in each groups and period, we can put

$$J_2^* = \frac{1}{N} (\aleph \mathbf{1} \mathbf{1}' \aleph) = \frac{1}{N} \begin{bmatrix} n_1^2 & n_1 n_2 \\ n_1 n_2 & n_2^2 \end{bmatrix}. \quad (3.2.8)$$

Table 3.3 then shows the relevant projector matrices, which are orthogonal with respect to the weight matrix $\aleph^{-1} \otimes I_2$, for the four strata identified in Table 2.3. Each stratum corresponds to just one degree of freedom.

			Matrix
Grand Mean	Q_0^*	$J_2^* \otimes J_2$	$\frac{1}{2N} \begin{bmatrix} n_1^2 & n_1^2 & n_1 n_2 & n_1 n_2 \\ n_1^2 & n_1^2 & n_1 n_2 & n_1 n_2 \\ n_1 n_2 & n_1 n_2 & n_2^2 & n_2^2 \\ n_1 n_2 & n_1 n_2 & n_2^2 & n_2^2 \end{bmatrix}$
Period	Q_1^*	$J_2^* \otimes K_2$	$\frac{1}{2N} \begin{bmatrix} n_1^2 & -n_1^2 & n_1 n_2 & -n_1 n_2 \\ -n_1^2 & n_1^2 & -n_1 n_2 & n_1 n_2 \\ n_1 n_2 & -n_1 n_2 & n_2^2 & -n_2^2 \\ -n_1 n_2 & n_1 n_2 & -n_2^2 & n_2^2 \end{bmatrix}$
Group	Q_2^*	$(\mathbb{N} - J_2^*) \otimes J_2$	$\frac{n_1 n_2}{2N} \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$
Group $\times$ Period	Q_3^*	$(\mathbb{N} - J_2^*) \otimes K_2$	$\frac{n_1 n_2}{2N} \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$

Table 3.3: Weighted orthogonal matrices for four strata in the design

3.2.2 Treatment information

Now, to get the information about treatment parameters we consider the treatment information excluding Grand Mean stratum. This is given by

$$T^{*'}(\mathbb{N} \otimes I - Q_0^*)T^* = \frac{N}{2} \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} + \frac{2n_1 n_2}{N} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

This separates into a component in each of the three strata, given by

$$\begin{aligned} T^{*'}Q_1^*T^* &= \frac{l^2}{2N} \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} \\ T^{*'}Q_2^*T^* &= \frac{4n_1 n_2}{2N} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ T^{*'}Q_3^*T^* &= \frac{4n_1 n_2}{2N} \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}, \end{aligned} \tag{3.2.9}$$

where $N = n_1 + n_2$, and $l = n_1 - n_2$. We note that the rank of each of these matrices is 1. In particular, $T^{*'}Q_i^*T^*(i = 1, 3)$ contains information only about $(2\tau - \lambda)$ and no information about $(\tau + 2\lambda)$, while $T^{*'}Q_2^*T^*$ contains information only about λ . If $n_1 = n_2$, there is no information about treatment in $T^{*'}Q_1^*T^*$.

We will ignore, as discussed earlier, the information in Q_1^* , the Period stratum. The estimates of θ in the other strata can then be obtained as

$$\begin{aligned}\hat{\theta}_2 &= (T^{*'}Q_2^*T^*)^{-1}T^{*'}Q_2^*y^* = \frac{1}{2} \begin{bmatrix} 0 \\ (\bar{y}_{11.} + \bar{y}_{12.}) - (\bar{y}_{21.} + \bar{y}_{22.}) \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{y}_{1..} - \bar{y}_{2..} \end{bmatrix} \\ \hat{\theta}_3 &= (T^{*'}Q_3^*T^*)^{-1}T^{*'}Q_3^*y^* = \frac{1}{10} \begin{bmatrix} 2[(\bar{y}_{11.} - \bar{y}_{12.}) - (\bar{y}_{21.} - \bar{y}_{22.})] \\ -(\bar{y}_{11.} - \bar{y}_{12.}) + (\bar{y}_{21.} - \bar{y}_{22.}) \end{bmatrix} = \frac{1}{10}(d_{1.} - d_{2.}) \begin{bmatrix} 2 \\ -1 \end{bmatrix}\end{aligned}$$

where $d_{1.} = \bar{y}_{11.} - \bar{y}_{12.}$ and $d_{2.} = \bar{y}_{21.} - \bar{y}_{22.}$.

Thus, $\hat{\theta}_2$ provides an estimate of λ only, while $\hat{\theta}_3$ provides an estimate of $(2\tau - \lambda)$, since

$$\begin{bmatrix} 2 & -1 \end{bmatrix} \hat{\theta}_3 = \frac{1}{2}(d_{1.} - d_{2.}),$$

and this can be shown to have expected value $(2\tau - \lambda)$. On the other hand, $\begin{bmatrix} 1 & 2 \end{bmatrix} \hat{\theta}_3 = 0$, so that no information is available for the parameter $(\tau + 2\lambda)$ in this stratum.

In the case $n_1 \neq n_2$, there is information about $(2\tau - \lambda)$ in the Period stratum, but this is not recoverable unless we are willing to assume that either there is no period effect or that the period effect is random with a known variance. It can be seen from 3.2.9 that the information about $(2\tau - \lambda)$ is split between the Period stratum and the Group $\times$ Period stratum in the ratio $l^2 : (4n_1n_2)$, and hence that the proportion of the information lost into Periods is $l^2/N^2 = (1 - 2q)^2$, where $q = \frac{n_1}{N}$. Clearly, there is no information loss to Periods when $n_1 = n_2$.

The variance matrices for parameter estimates are

$$\begin{aligned}Var(\hat{\theta}_2) &= (\sigma^2 + 2\sigma_s^2)(T^{*'}Q_2^*T^*)^{-1} = \frac{N(\sigma^2 + 2\sigma_s^2)}{2n_1n_2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ Var(\hat{\theta}_3) &= \sigma^2(T^{*'}Q_3^*T^*)^{-1} = \frac{N\sigma^2}{50n_1n_2} \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}.\end{aligned}$$

The estimates can then be combined using (2.6.4) to give

$$\begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} = (\delta_2^{-1}T^{*'}Q_2^*T^* + \delta_3^{-1}T^{*'}Q_3^*T^*)^{-1}(\delta_2^{-1}T^{*'}Q_2^*T^*\hat{\theta}_2 + \delta_3^{-1}T^{*'}Q_3^*T^*\hat{\theta}_3)$$

(3.2.10)

$$= \frac{1}{2} \begin{bmatrix} \bar{y}_{11.} - \bar{y}_{21.} \\ 2(\bar{y}_{1..} - \bar{y}_{2..}) \end{bmatrix}, \quad (3.2.11)$$

and, if $\rho = \sigma_s^2/\sigma^2$, then the covariance matrix for the combined estimates is

$$\begin{aligned} \text{Var}\left\{ \begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} \right\} &= \left(\frac{1}{\sigma^2 + 2\sigma_s^2} T^{*'} Q_2^* T^* + \frac{1}{\sigma^2} T^{*'} Q_3^* T^* \right)^{-1} \\ &= \frac{N\sigma^2}{4n_1n_2} \begin{bmatrix} 1 + \rho & 1 + 2\rho \\ 1 + 2\rho & 2(1 + 2\rho) \end{bmatrix}. \end{aligned}$$

We see that an unbiased estimator of τ can be obtained as

$$\hat{\tau} = (\bar{y}_{11.} - \bar{y}_{21.})/2,$$

that is, $\hat{\tau}$ is based only on the first period data when λ is present in the model. In other words, if λ is in the model then the estimate of τ is based on between-subject information and is the estimator we would have obtained if the trial has been designed as a parallel-group study in only one period.

To test the null hypothesis that $\tau=0$ when λ is in the model, Jones & Kenward (1989), p. 28, recommend using a two-sample t-test, estimating $\sigma^2 + \sigma_s^2$ using only the first period data. This is different from the estimated variances which would be obtained by Chi (1991) who proposes estimating both σ^2 and σ_s^2 , separately, from the full data set, by equating the mean squares for Q_4 and Q_5 to their expectation in Section 2.6.

3.2.3 Analysis of the design when λ is not present

A further stage of course is that if we do have not a carry-over effect in our model then to get an unbiased estimator of τ , we can write

$$T^* = \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix}' \quad (3.2.12)$$

Then the treatment information in the three strata of interest can be given by

$$\begin{aligned} T^{*'} Q_1^* T^* &= \frac{2l^2}{N} \\ T^{*'} Q_2^* T^* &= 0 \\ T^{*'} Q_3^* T^* &= \frac{8n_1n_2}{N}, \end{aligned}$$

and since we have decided not to use the information in the Period stratum the estimate of τ just comes from the Group $\times$ Period stratum, namely

$$\hat{\tau} = (\bar{y}_{11.} - \bar{y}_{12.} - \bar{y}_{21.} + \bar{y}_{22.})/4$$

Now we have an unbiased estimator of τ based on within-subjects information, and

$$Var(\hat{\tau}) = \frac{N}{8n_1n_2}\sigma^2.$$

We note again that if $n_1 \neq n_2$ we lose a proportion $(1 - 2q)^2$ of the treatment information into the Period stratum. As noted by other authors, the assumption that $\lambda = 0$ leads to higher precision for the treatment estimate. The efficiency of estimating τ with λ in the model compared to estimating τ without λ in the model is the inverse ratio of the two variances, namely

$$\frac{N\sigma^2}{8n_1n_2} \times \frac{4n_1n_2}{N\sigma^2(1 + \rho)},$$

which is 50% when $\rho = 0$, that is there are no between-subject differences, and less the 50% for $\rho > 0$.

3.2.4 Analysis of design when subject effect is fixed

Our model has subject effects random. This enables us to recover information from differences between subjects. If the subject effects are fixed, the only information available is in the differences $d_i = \bar{y}_{i1.} - \bar{y}_{i2.}$ ($i = 1, 2$) between the two measurements on each subject. This implies that only the information from Q_1^* and Q_3^* is available. In order to derive a set of estimates for the parameters of interest, we write the period differences as.

$$\begin{aligned} d_1 &= \bar{y}_{11.} - \bar{y}_{12.} = 2\pi + 2\tau - \lambda + \bar{\xi}_1, \\ d_2 &= \bar{y}_{21.} - \bar{y}_{22.} = 2\pi - 2\tau + \lambda + \bar{\xi}_2, \end{aligned} \tag{3.2.13}$$

where $\pi = \pi_1 = -\pi_2$. By matrix notation

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} + \begin{bmatrix} 2 & -1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} \tau \\ \lambda \end{bmatrix} + \begin{bmatrix} \bar{\xi}_1 \\ \bar{\xi}_2 \end{bmatrix}, \tag{3.2.14}$$

where $Var(\bar{\xi}_i) = 2\sigma^2/n_i$, for $i = 1, 2$. It is clear that

$$\hat{\pi} = (d_1 + d_2)/4,$$

with variance

$$Var(\hat{\pi}) = \frac{\sigma^2}{8Nq(1-q)}.$$

However we can not distinguish between τ and λ since $(d_1 - d_2)$ estimates $2(2\tau - \lambda)$. If $\lambda = 0$, $(d_2 - d_1)/4$ estimates τ with a variance,

$$Var(\hat{\tau}) = \frac{N\sigma^2}{8Nq(1-q)},$$

but as discussed earlier, if $\lambda \neq 0$, the only estimate available for τ is based on between subject differences. If the subjects are fixed effects, we cannot obtain an estimate of τ at all.

3.2.5 A comparison of the three cases of 2×2 crossover design

In Table 3.4 we give the results of the analysis of the crossover design and have a comparison in terms of presence and absence of carryover effect in the model. Let us define $q = \frac{n_1}{n_1+n_2}$ and $\rho = \frac{\sigma_s^2}{\sigma^2}$.

	Case		
	1	2	3
Estimator	$\rho \neq 0$ and $q \neq 0.5$ and λ is present	$\rho \neq 0$ and $q \neq 0.5$ and λ is not present	$\rho = 0$ and $q \neq 0.5$ and λ is not present
$\hat{\tau}$	$\frac{1}{2}(\bar{y}_{11.} - \bar{y}_{21.})$	$\frac{1}{4}[\bar{y}_{11.} - \bar{y}_{12.} - \bar{y}_{21.} + \bar{y}_{22.}]$	$\frac{1}{4}[\bar{y}_{11.} - \bar{y}_{12.} - \bar{y}_{21.} + \bar{y}_{22.}]$
$\hat{\lambda}$	$\bar{y}_{1..} - \bar{y}_{2..}$	None	None
$Var(\hat{\tau})$	$\frac{\sigma^2(1+\rho)}{4Nq(1-q)}$	$\frac{\sigma^2}{8Nq(1-q)}$	$\frac{\sigma^2}{8Nq(1-q)}$
$Var(\hat{\lambda})$	$\frac{\sigma^2(1+2\rho)}{2Nq(1-q)}$	None	None

Table 3.4: Estimates of parameters and their variances.

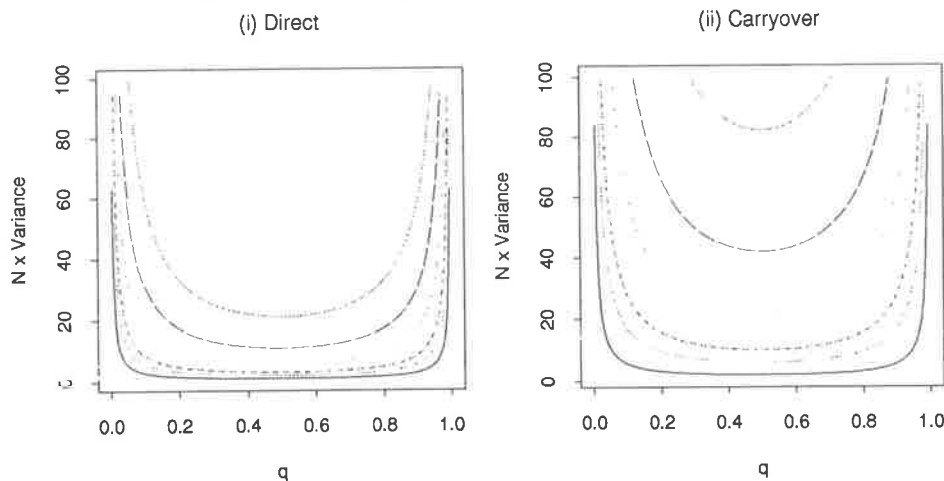
We note that the total information about (τ, λ) is

$$N \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix},$$

when $n_1 = n_2$, and hence the best we could hope to do for $Var(\hat{\tau})$ would be $\sigma^2/2N$ in the case where λ not in the model. This is achieved if $q = \frac{1}{2}$ above, but not otherwise. If λ is in the model, the smallest variance achievable for $\hat{\tau}$ is $\sigma^2(1+\rho)/N$ which is larger by a

factor $2(1+\rho)$. Thus, even if $\rho = 0$, only 50% of the information about τ is available when λ is in the model. To provide an illustration of the minimum variances of each estimator of the parameters, we show in Figure 3.1 the graphs of variance of direct treatment and of carryover effect, in terms of $\rho = 0, 1, 2, 5, 10, 20$ and q . In these graphs, we set $\sigma^2 = 1$. From these graphs we can see a clear minimum of variance of both parameters of interest when $\rho = 0$ and $q = 0.5$. Furthermore, the minimum of these variances occurs at $q = 0.5$ for all ρ .

Figure 3.1: N times variance of (i) direct treatment effect and of (ii) carry-over effect estimator for 2×2 design in terms of $\rho = 0, 1, 2, 5, 10, 20$ and $q = n_1/(n_1 + n_2)$.



3.3 Baseline measurements in the 2×2 crossover design

Up to now we have restricted our discussion to the case of the simplest of two-treatment designs, that is those with only two periods, just touching on the effect of adding baseline measurements. The difficulty with these designs is, as we have seen, the problem of confounding carryover or treatment $\times$ period interaction effects with the treatment effect. We now turn to consideration of more complex crossover designs. We shall, however, continue to restrict our discussion to studies involving only two treatments.

As we showed in the previous sections, in the analysis of 2×2 crossover design the main difficulty is that unbiased estimates for direct treatment effects are based on first period data only. In this section, we shall analyse this design by including baseline

measurements which some authors such as Freeman (1989) have recommended. We show that they can not solve our main difficulty, but they provide a way to increase the power of tests against the presence of carryover effect in the model.

Using the baseline in crossover trials has two potential advantages, in that we can overcome two basic difficulties which we face in the analysis of two-treatment, two-period crossover design without baseline measurements. The first difficulty is the low power of the test against of carryover effect. Kenward & Jones (1987a) have shown that one way to increase the power of this test is to use baseline measurements. The second difficulty, as Freeman (1989) has suggested, is the over-parameterisation of the design without baseline measurements. In addition, if there is baseline measurement in the model we can estimate the average effect of either treatment or carryover, provided we can make an assumption about the absence of period effects.

In this work, we consider baseline measurements in two sections. In the next section we will take a single baseline measurement before the subjects are given the first treatment. Then we shall analyse the consequences of the use of baseline measurements before the subject is given the first treatment as well as the second treatment.

3.3.1 One baseline measurement in the design

In this design there are three periods, in which the first period provides the baseline measurements. The layout of this design can be shown as in Table 3.5.

	Period		
group	1	2	3
	—	A	B
	—	B	A

Table 3.5: Two-treatment design with one baseline measurement

Here the hyphen sign (—) means that there is no treatment implemented in that period. The model that we apply to the above situation is the same as used in Section 3.2. Now, by employing the usual constraints on the parameters of interest, the expected values of observations are given in Table 3.6.

		Period		
Group	1	1	2	3
		$\mu + \pi_1$	$\mu + \pi_2 + \tau$	$\mu + \pi_3 - \tau + \lambda$
	2	$\mu + \pi_1$	$\mu + \pi_2 - \tau$	$\mu + \pi_3 + \tau - \lambda$

Table 3.6: Expectation of responses in 2×2 crossover design when λ is in the model

and the matrix form for the six means is

$$y^* = \begin{bmatrix} \bar{y}_{11.} \\ \bar{y}_{12.} \\ \bar{y}_{13.} \\ \bar{y}_{21.} \\ \bar{y}_{22.} \\ \bar{y}_{23.} \end{bmatrix} = \mu \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ -1 & 1 \\ 0 & 0 \\ -1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \tau \\ \lambda \end{bmatrix} + \begin{bmatrix} \bar{\epsilon}_{11.} \\ \bar{\epsilon}_{12.} \\ \bar{\epsilon}_{13.} \\ \bar{\epsilon}_{21.} \\ \bar{\epsilon}_{22.} \\ \bar{\epsilon}_{23.} \end{bmatrix}, \quad (3.3.1)$$

or

$$y^* = \mu 1_6 + (1_2 \otimes I_3)\pi + (I_2 \otimes 1_3)\beta^* + T^*\theta + \epsilon^*. \quad (3.3.2)$$

The variance-covariance of y^* can be shown as

$$\text{Var}(y^*) = \sigma^2(\aleph^{-1} \otimes K_3) + (\sigma^2 + 3\sigma_s^2)(\aleph^{-1} \otimes J_3).$$

By using J_2^* defined in Section 3.2.1, we can show that the weighted orthogonal idempotents for this case are as given in Table 3.7.

		Matrix
Q_0^*	$J_2^* \otimes J_3$	$\frac{1}{3N} \begin{bmatrix} n_1^2 \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} & n_1 n_2 \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \\ n_1 n_2 \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} & n_2^2 \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{bmatrix}$
Q_1^*	$J_2^* \otimes K_3$	$\frac{1}{3N} \begin{bmatrix} n_1^2 \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} & n_1 n_2 \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \\ n_1 n_2 \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} & n_2^2 \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \end{bmatrix}$
Q_2^*	$(\aleph - J_2^*) \otimes J_3$	$\frac{n_1 n_2}{3N} \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix}$
Q_3^*	$(\aleph - J_2^*) \otimes K_3$	$\frac{n_1 n_2}{3N} \begin{bmatrix} 2 & -1 & -1 & -2 & 1 & 1 \\ -1 & 2 & -1 & 1 & -2 & 1 \\ -1 & -1 & 2 & 1 & -2 & 1 \\ -2 & 1 & 1 & 2 & -1 & -1 \\ 1 & -2 & 1 & -1 & 2 & -1 \\ 1 & 1 & -2 & -1 & -1 & 2 \end{bmatrix}$

Table 3.7: Weighted orthogonal matrices for four strata in the design

With this design, the total information available is

$$T^{*'}(\aleph \otimes I - Q_0^*)T^* = \frac{N}{3} \begin{bmatrix} 6 & -3 \\ -3 & 2 \end{bmatrix} + \frac{4n_1 n_2}{3N} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

The treatment information in Period, Group and Group $\times$ Period strata respectively is given by

$$\begin{aligned} T^{*'}Q_1^*T^* &= \frac{l^2}{3N} \begin{bmatrix} 6 & -3 \\ -3 & 2 \end{bmatrix} \\ T^{*'}Q_2^*T^* &= \frac{4n_1n_2}{3N} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ T^{*'}Q_3^*T^* &= \frac{4n_1n_2}{3N} \begin{bmatrix} 6 & -3 \\ -3 & 2 \end{bmatrix}. \end{aligned}$$

The estimates in Group and Group $\times$ Period strata can be given by

$$\begin{aligned} \hat{\theta}_2 &= \frac{3}{2} \begin{bmatrix} 0 \\ \bar{y}_{1..} - \bar{y}_{2..} \end{bmatrix} \\ \hat{\theta}_3 &= \frac{1}{2} \begin{bmatrix} -\bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{21.} - \bar{y}_{22.} \\ -2\bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.} + 2\bar{y}_{21.} - \bar{y}_{22.} - \bar{y}_{23.} \end{bmatrix}, \end{aligned}$$

with covariance matrices

$$\begin{aligned} Var(\hat{\theta}_2) &= (\sigma^2 + 3\sigma_s^2)(T^{*'}Q_2^*T^*)^{-1} = \frac{3N(\sigma^2 + 3\sigma_s^2)}{4n_1n_2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ Var(\hat{\theta}_3) &= \sigma^2(T^{*'}Q_3^*T^*)^{-1} = \frac{N\sigma^2}{4n_1n_2} \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix}. \end{aligned}$$

These estimates can then be combined using (2.6.4) to give

$$\begin{aligned} \begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} &= (\delta_2^{-1}T^{*'}Q_2^*T^* + \delta_3^{-1}T^{*'}Q_3^*T^*)^{-1}(\delta_2^{-1}T^{*'}Q_2^*T^*\hat{\theta}_2 + \delta_3^{-1}T^{*'}Q_3^*T^*\hat{\theta}_3) \\ &= \frac{1}{2(\sigma^2 + \sigma_s^2)} \begin{bmatrix} (\sigma^2 + \sigma_s^2)(\bar{y}_{12.} - \bar{y}_{22.}) + \sigma_s^2(\bar{y}_{21.} - \bar{y}_{11.}) \\ (\sigma^2 + \sigma_s^2)(\bar{y}_{12.} + \bar{y}_{13.} - \bar{y}_{22.} - \bar{y}_{23.}) + 2\sigma_s^2(\bar{y}_{21.} - \bar{y}_{11.}) \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} \bar{y}_{12.} - \bar{y}_{22.} + \frac{\rho}{1+\rho}(\bar{y}_{21.} - \bar{y}_{11.}) \\ \bar{y}_{12.} + \bar{y}_{13.} - \bar{y}_{22.} - \bar{y}_{23.} + \frac{2\rho}{1+\rho}(\bar{y}_{21.} - \bar{y}_{11.}) \end{bmatrix} \end{aligned}$$

and the covariance matrix for the combined estimates is

$$\begin{aligned} Var\left\{ \begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} \right\} &= \left(\frac{1}{\sigma^2 + 3\sigma_s^2}T^{*'}Q_2^*T^* + \frac{1}{\sigma^2}T^{*'}Q_3^*T^* \right)^{-1} \\ &= \frac{N\sigma^2}{4n_1n_2(1+\rho)} \begin{bmatrix} 1+2\rho & 1+3\rho \\ 1+3\rho & 2(1+3\rho) \end{bmatrix}. \end{aligned}$$

In the case $\rho \rightarrow \infty$, the subject effects are regarded as providing no information and the estimates reduce to the within-subject $\hat{\theta}_3$. In case $\rho \rightarrow 0$, there no subject effects, and $\hat{\theta}_2, \hat{\theta}_3$ are combined with $\delta_2 = \delta_3$. As ρ decreases, the extra information about λ from the subjects stratum reduces the correlation between the estimates of τ and λ . However, this correlation still remains, being 0.71 even when $\rho = 0$. If λ is not in model, the information about τ in $T^*Q_3^*T^*$ is $8n_1n_2/N$, and the estimate is

$$(\bar{y}_{12.} - \bar{y}_{13.} - \bar{y}_{22.} + \bar{y}_{23.})/2,$$

with a variance of $N\sigma^2/(8n_1n_2)$. The strong covariance between $\hat{\tau}$ and $\hat{\lambda}$ implies that the variance of the within-subject estimate of τ when λ is in the model is increased to $N\sigma^2/(2n_1n_2)$, implying an efficiency of only 25%. Even this is an improvement over the case without baseline when there is no within-subject estimate available. Incorporation of between-subject information reduces the variance of the estimate of τ to

$$\frac{N\sigma^2(1+2\rho)}{4n_1n_2(1+\rho)},$$

implying that the efficiency of the combined estimate of τ when λ is in the model is

$$\frac{N\sigma^2}{8n_1n_2} \times \frac{4n_1n_2(1+\rho)}{N\sigma^2(1+2\rho)} = \frac{(1+\rho)}{2(1+2\rho)},$$

relative to the estimate of τ when λ is not in the model. This varies from 25% (as $\rho = \infty$) to 50% as ($\rho = 0$).

3.3.2 Two baseline measurements in the design

We shall now examine the effect of considering two baseline measurements to the 2×2 crossover design, one before the first active treatment period and another before the second treatment period. This case can be shown in Table 3.8 as follows.

Group	Periods			
	1	2	3	4
	—	A	—	B
	—	B	—	A

Table 3.8: Two-treatment crossover design with two baseline measurements

Employing the usual constraints on parameters, we have

$$\begin{bmatrix} \bar{y}_{11.} \\ \bar{y}_{12.} \\ \bar{y}_{13.} \\ \bar{y}_{14.} \\ \bar{y}_{21.} \\ \bar{y}_{22.} \\ \bar{y}_{23.} \\ \bar{y}_{24.} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ -1 & 1 \\ 0 & 0 \\ -1 & 0 \\ 0 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \tau \\ \lambda \end{bmatrix} + \begin{bmatrix} \bar{\epsilon}_{11.} \\ \bar{\epsilon}_{12.} \\ \bar{\epsilon}_{13.} \\ \bar{\epsilon}_{14.} \\ \bar{\epsilon}_{21.} \\ \bar{\epsilon}_{22.} \\ \bar{\epsilon}_{23.} \\ \bar{\epsilon}_{24.} \end{bmatrix},$$

or

$$y^* = \mu 1_8 + (1_2 \otimes I_4)\pi + (I_2 \otimes 1_4)\beta^* + T^*\theta + \epsilon^* \quad (3.3.3)$$

and we have variance-covariance matrix, V as given below

$$Var(y^*) = \sigma^2(\aleph^{-1} \otimes K_4) + (\sigma^2 + 4\sigma_s^2)(\aleph^{-1} \otimes J_4)$$

where $\aleph$ is $diag(n_1, n_2)$, as before.

The treatment information for this design also comes from those strata that we considered before, and is given by

$$\begin{aligned} T^{*'}Q_1^*T^* &= \frac{l^2}{4N} \begin{bmatrix} 8 & -4 \\ -4 & 3 \end{bmatrix} \\ T^{*'}Q_2^*T^* &= \frac{n_1n_2}{N} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ T^{*'}Q_3^*T^* &= \frac{n_1n_2}{N} \begin{bmatrix} 8 & -4 \\ -4 & 3 \end{bmatrix}, \end{aligned}$$

where Q_i^* are shown in Table 3.9.

Q_0^*	$J_2^* \otimes J_4$	$\frac{1}{4N}$	$\begin{bmatrix} n_1^2 \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} & n_1 n_2 \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \\ n_1 n_2 \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} & n_2^2 \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \end{bmatrix}$
Q_1^*	$J_2^* \otimes K_4$	$\frac{1}{4N}$	$\begin{bmatrix} n_1^2 \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} & n_1 n_2 \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} \\ n_1 n_2 \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} & n_2^2 \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} \end{bmatrix}$
Q_2^*	$(\aleph - J_2^*) \otimes J_4$	$\frac{n_1 n_2}{4N}$	$\begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{bmatrix}$
Q_3^*	$(\aleph - J_2^*) \otimes K_4$	$\frac{n_1 n_2}{4N}$	$\begin{bmatrix} 3 & -1 & -1 & -1 & -3 & 1 & 1 & 1 \\ -1 & 3 & -1 & -1 & 1 & -3 & 1 & 1 \\ -1 & -1 & 3 & 1 & 1 & 1 & -3 & 1 \\ -1 & -1 & -1 & 3 & 1 & 1 & 1 & -3 \\ -3 & 1 & 1 & 1 & 3 & -1 & -1 & -1 \\ 1 & -3 & 1 & 1 & -1 & 3 & -1 & -1 \\ 1 & 1 & -3 & 1 & -1 & -1 & 3 & -1 \\ 1 & 1 & 1 & -3 & -1 & -1 & -1 & 3 \end{bmatrix}$

Table 3.9: Weighted orthogonal matrices for four strata in the design

The estimates in the two strata, namely Group and Group $\times$ Period, can be given by

$$\hat{\theta}_2 = \frac{1}{2} \begin{bmatrix} 0 \\ \bar{y}_{1..} - \bar{y}_{2..} \end{bmatrix}$$

$$\hat{\theta}_3 = \frac{1}{4} \begin{bmatrix} 2(\bar{y}_{12.} - \bar{y}_{22.}) - (\bar{y}_{11.} + \bar{y}_{13.} - \bar{y}_{21.} - \bar{y}_{23.}) \\ 2[(\bar{y}_{12.} + \bar{y}_{14.} - \bar{y}_{11.} - \bar{y}_{13.}) + (\bar{y}_{21.} + \bar{y}_{23.} - \bar{y}_{22.} - \bar{y}_{24.})] \end{bmatrix},$$

with covariance matrices

$$Var(\hat{\theta}_2) = (\sigma^2 + 4\sigma_s^2)(T^{*'}Q_2^*T^*)^{-} = \frac{N(\sigma^2 + 4\sigma_s^2)}{n_1n_2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Var(\hat{\theta}_3) = \sigma^2(T^{*'}Q_3^*T^*)^{-} = \frac{N\sigma^2}{8n_1n_2} \begin{bmatrix} 3 & 4 \\ 4 & 8 \end{bmatrix}.$$

These estimates can then be combined using (2.6.4) to give

$$\begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} = (\delta_2^{-1}T^{*'}Q_2^*T^* + \delta_3^{-1}T^{*'}Q_3^*T^*)^{-}(\delta_2^{-1}T^{*'}Q_2^*T^*\hat{\theta}_2 + \delta_3^{-1}T^{*'}Q_3^*T^*\hat{\theta}_3)$$

$$= \frac{1}{2(\sigma^2 + 2\sigma_s^2)} \begin{bmatrix} (\sigma^2 + 2\sigma_s^2)(\bar{y}_{12.} - \bar{y}_{22.}) + \sigma_s^2(\bar{y}_{21.} - \bar{y}_{11.} - \bar{y}_{13.} + \bar{y}_{23.}) \\ (\sigma^2 + 2\sigma_s^2)(\bar{y}_{12.} + \bar{y}_{14.} - \bar{y}_{22.} - \bar{y}_{24.}) - 2\sigma_s^2(\bar{y}_{11.} + \bar{y}_{13.} - \bar{y}_{21.} - \bar{y}_{23.}) \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} \bar{y}_{12.} - \bar{y}_{22.} + \frac{\rho}{1+2\rho}(\bar{y}_{21.} - \bar{y}_{11.} - \bar{y}_{13.} + \bar{y}_{23.}) \\ (\bar{y}_{12.} + \bar{y}_{14.} - \bar{y}_{22.} - \bar{y}_{24.}) - \frac{2\rho}{1+2\rho}(\bar{y}_{11.} + \bar{y}_{13.} - \bar{y}_{21.} - \bar{y}_{23.}) \end{bmatrix},$$

and the covariance matrix for the combined estimates is

$$Var\left\{ \begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} \right\} = \left(\frac{1}{\sigma^2 + 4\sigma_s^2} T^{*'}Q_2^*T^* + \frac{1}{\sigma^2} T^{*'}Q_3^*T^* \right)^{-1}$$

$$= \frac{N\sigma^2}{4n_1n_2(1+2\rho)} \begin{bmatrix} 1+3\rho & 1+4\rho \\ 1+4\rho & 2(1+4\rho) \end{bmatrix}$$

3.3.3 Conclusion and discussion

As we can see in the analysis of the two-treatment, two-period case, we cannot estimate τ and λ within-subject; the only combined estimates require between-subject information. The total information in Group and Group $\times$ Period strata for the two-treatment, two-period design shown in Section 3.2 and both designs including baselines in Section 3.3 is

$$T^{*'}Q_2^*T^* + T^{*'}Q_3^*T^* = \frac{4n_1n_2}{N} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}. \quad (3.3.4)$$

The only difference between the designs is the way in which the information about λ is lost into the Subject stratum. In the two-treatment, two-period design, 50% of the information about λ is lost into the Subject stratum, leaving a singular information matrix. With one baseline, a third is lost into the Subject stratum, while with two baselines, only 25% is lost. Hence the baseline measurements allow us to keep more information about λ in the Subject $\times$ Period stratum and hence improve the estimate of τ . Thus, as move through these models, the variance of $\hat{\tau}$ changes from

$$\frac{\sigma^2(1+\rho)}{4Nq(1-q)} \text{ to } \frac{\sigma^2(1+2\rho)}{4Nq(1-q)(1+\rho)} \text{ to } \frac{\sigma^2(1+3\rho)}{4Nq(1-q)(1+2\rho)}.$$

If $\rho = 0$, there is no change, since the total information remains the same in all these three cases. However, if $\rho > 0$, $Var(\hat{\tau})$ is reduced as we include more baseline measurements.

3.4 Balaam's design for two treatments

As Jones & Kenward (1989) on page 140 have pointed out, the main disadvantages of the standard AB, BA design without baselines are that

1. the test for a carry-over effect or direct-by-period interaction lacks power because it is based on between-subject comparisons, and
2. the carryover effect, the treatment $\times$ period interaction and the group difference are completely aliased with one another in the sense that only two of the three can be estimated.

If designs with more sequences or more periods for two treatments are used, however, we are able to obtain within-subject estimators of the carryover effects. The problem of choosing the best estimates for direct treatment and carryover effects has been considered by a number of researchers, and the designs which they have chosen provide minimum variance unbiased estimators of the effects of interest.

One of the more efficient designs for comparing two treatments in the two period crossover design is the design of Balaam (1968) which we consider in this section.

We will analyse the Balaam design when the subject effects are random variables with mean zero and variance σ_s^2 and we assume that the within subject errors are independent random variables with mean zero and variance σ^2 . In addition we can assume that these

variables have a normal distribution. In Table 3.10 we show the layout of Balaam's design for two treatments and the number of subjects in each sequence group.

group	subjects	treatments
1	n_1	AB
2	n_1	BA
3	n_2	AA
4	n_2	BB

Table 3.10: The layout of Balaam's design with the number of subjects in each group.

3.4.1 Balaam's design when subject effects are random and unequal group sizes

We begin our analysis for Balaam's design by referring to the eight expected values in Table 3.11.

		Period	
		1	2
Group	1	$\mu + \pi_1 + \tau$	$\mu + \pi_2 - \tau + \lambda$
	2	$\mu + \pi_1 - \tau$	$\mu + \pi_2 + \tau - \lambda$
	3	$\mu + \pi_1 + \tau$	$\mu + \pi_2 + \tau + \lambda$
	4	$\mu + \pi_1 - \tau$	$\mu + \pi_2 - \tau - \lambda$

Table 3.11: Expected value of mean in Balaam's design

It will be noted that the all previous notation can be used here. By referring to Table 3.11 the mixed linear model for this case can be written as

$$y^* = \mu 1_8 + (1_4 \otimes I_2)\pi + (I_4 \otimes 1_2)\beta^* + T^*\theta + \epsilon^* \quad (3.4.1)$$

In this design the T^* matrix can be set as

$$T^* = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ -1 & 0 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ -1 & 0 \\ -1 & -1 \end{bmatrix}, \quad (3.4.2)$$

and by defining a weight matrix as

$$\aleph = \begin{bmatrix} n_1 & 0 & 0 & 0 \\ 0 & n_1 & 0 & 0 \\ 0 & 0 & n_2 & 0 \\ 0 & 0 & 0 & n_2 \end{bmatrix},$$

then we have

$$J_4^* = \frac{1}{N}(\aleph 1_4 1_4' \aleph) = \frac{1}{N} \begin{bmatrix} n_1^2 & n_1^2 & n_1 n_2 & n_1 n_2 \\ n_1^2 & n_1^2 & n_1 n_2 & n_1 n_2 \\ n_1 n_2 & n_1 n_2 & n_2^2 & n_2^2 \\ n_1 n_2 & n_1 n_2 & n_2^2 & n_2^2 \end{bmatrix} = \frac{2}{N} \begin{bmatrix} n_1^2 & n_1 n_2 \\ n_1 n_2 & n_2^2 \end{bmatrix} \otimes J_2,$$

where $N = 2n_1 + 2n_2$. Now we can get the following weighted orthogonal idempotent for the Balaam's design in Table 3.12.

		Matrix
Q_0^*	$J_4^* \otimes J_2$	$\frac{2}{N} \begin{bmatrix} n_1^2 & n_1 n_2 \\ n_1 n_2 & n_2^2 \end{bmatrix} \otimes J_2 \otimes J_2$
Q_1^*	$J_4^* \otimes K_2$	$\frac{2}{N} \begin{bmatrix} n_1^2 & n_1 n_2 \\ n_1 n_2 & n_2^2 \end{bmatrix} \otimes J_2 \otimes K_2$
Q_2^*	$(\aleph - J_4^*) \otimes J_2$	$\left\{ \begin{bmatrix} n_1 & 0 \\ 0 & n_2 \end{bmatrix} \otimes K_2 + \frac{4n_1 n_2}{N} K_2 \otimes J_2 \right\} \otimes J_2$
Q_3^*	$(\aleph - J_4^*) \otimes K_2$	$\left\{ \begin{bmatrix} n_1 & 0 \\ 0 & n_2 \end{bmatrix} \otimes K_2 + \frac{4n_1 n_2}{N} K_2 \otimes J_2 \right\} \otimes K_2$

Table 3.12: Weighted orthogonal matrices for four strata in Balaam's design

The treatment information in the strata of interest can be given as

$$\begin{aligned} T^{*'}Q_1^*T^* &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\ T^{*'}Q_2^*T^* &= \begin{bmatrix} 4n_2 & 2n_2 \\ 2n_2 & n_1 + n_2 \end{bmatrix} \\ T^{*'}Q_3^*T^* &= \begin{bmatrix} 4n_1 & -2n_1 \\ -2n_1 & n_1 + n_2 \end{bmatrix}, \end{aligned}$$

As we can see, the rank of $T^{*'}Q_3^*T^*$ is two, which is sufficient to estimate the two parameters of interest, so we have

$$\begin{aligned} \hat{\theta}_2 &= \frac{1}{4} \begin{bmatrix} \bar{y}_{21.} + \bar{y}_{22.} - \bar{y}_{11.} - \bar{y}_{12.} + \bar{y}_{31.} + \bar{y}_{32.} - \bar{y}_{41.} - \bar{y}_{42.} \\ 2(\bar{y}_{11.} + \bar{y}_{12.} - \bar{y}_{21.} - \bar{y}_{22.}) \end{bmatrix} \\ \hat{\theta}_3 &= \frac{1}{4} \begin{bmatrix} \bar{y}_{11.} - \bar{y}_{12.} - \bar{y}_{21.} + \bar{y}_{22.} - \bar{y}_{31.} + \bar{y}_{32.} + \bar{y}_{41.} - \bar{y}_{42.} \\ 2(\bar{y}_{32.} - \bar{y}_{31.} - \bar{y}_{42.} + \bar{y}_{41.}) \end{bmatrix} \end{aligned}$$

and covariance matrices of estimates are

$$\begin{aligned} Var(\hat{\theta}_2) &= (\sigma^2 + 2\sigma_s^2)(T^{*'}Q_2^*T^*)^{-} = \frac{\sigma^2 + 2\sigma_s^2}{8} \begin{bmatrix} \frac{N}{n_1 n_2} & -\frac{4}{n_1} \\ -\frac{4}{n_1} & \frac{8}{n_1} \end{bmatrix} \\ Var(\hat{\theta}_3) &= \sigma^2(T^{*'}Q_3^*T^*)^{-} = \frac{\sigma^2}{8} \begin{bmatrix} \frac{N}{n_1 n_2} & \frac{4}{n_2} \\ \frac{4}{n_2} & \frac{8}{n_2} \end{bmatrix}. \end{aligned}$$

Then the combined estimate for the parameters can be obtained as

$$\begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} = \left(\frac{1}{\sigma^2 + 2\sigma_s^2} T^{*'}Q_2^*T^* + \frac{1}{\sigma^2} T^{*'}Q_3^*T^* \right)^{-} \left(\frac{1}{\sigma^2 + 2\sigma_s^2} T^{*'}Q_2^*T^* \hat{\theta}_2 + \frac{1}{\sigma^2} T^{*'}Q_3^*T^* \hat{\theta}_3 \right)$$

If $N = 2(n_1 + n_2)$ = total number of subjects in the design, and $q = \frac{n_1}{n_1 + n_2}$ as before, then we have

$$\begin{aligned} \hat{\tau} &= \{q[2(1-q)\rho(1+\rho) + (1+2\rho)](\bar{y}_{11.} - \bar{y}_{21.}) \\ &\quad - (1-q)[2q\rho(1+\rho) - (1+2\rho)](\bar{y}_{31.} - \bar{y}_{41.}) \\ &\quad - 2q(1-q)(1+\rho)^2(\bar{y}_{12.} - \bar{y}_{22.} - \bar{y}_{32.} + \bar{y}_{42.})\} \\ &\quad / 2[4q(1-q)(1+\rho)^2 + (1+2\rho)] \end{aligned}$$

$$\begin{aligned}
\hat{\lambda} = & \{q[(1-q) + (-2+3q)(1+2\rho)](\bar{y}_{11.} - \bar{y}_{21.}) \\
& + q[(1-q) + (2-q)(1+2\rho)](\bar{y}_{12.} - \bar{y}_{22.}) \\
& + (1-q)(1+2\rho)[2q(1-\rho) - 1](\bar{y}_{31.} - \bar{y}_{41.}) \\
& + (1-q)(1+2\rho)[2q(1+\rho) + 1](\bar{y}_{32.} - \bar{y}_{42.})\} \\
& / 2 [4q(1-q)(1+\rho)^2 + (1+2\rho)],
\end{aligned}$$

and the covariance matrix which also depends on q and ρ can be obtained as

$$Var\left(\begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix}\right) = \left(\frac{1}{\sigma^2 + 2\sigma_s^2} T^{*'} Q_2^* T^* + \frac{1}{\sigma^2} T^{*'} Q_3^* T^*\right)^{-}$$

$$Var(\hat{\tau}) = \frac{\sigma^2(1+\rho)(1+2\rho)}{N[4q(1-q)(1+\rho)^2 + (1+2\rho)]}$$

$$Var(\hat{\lambda}) = \frac{2\sigma^2[1+2q\rho](1+2\rho)}{N[4q(1-q)(1+\rho)^2 + (1+2\rho)]}$$

$$Cov(\hat{\tau}, \hat{\lambda}) = \frac{\sigma^2(1+2\rho)[(-1+2q(1+\rho))]}{N[4q(1-q)(1+\rho)^2 + (1+2\rho)]}$$

As we can see in Balaam's design when the subject effect is random, estimates of parameters of interest and variances of estimates are completely dependent on q and ρ . In term of q and ρ there are several cases to consider.

1. When $q = 1$, Balaam's design is just the same as the two-treatment, two-period crossover design with $n_1 = n_2$. In fact when $q = 1$ we get the estimates of the parameters same as given in (3.2.10) in Section 3.2.1.
2. If we let $\rho \rightarrow \infty$, then Balaam's design gives just the within-subject estimates, $\hat{\theta}_3$.
3. If λ were not in the model, the information about τ is

$$4n_1 + 4n_2,$$

and hence the best possible variance for $\hat{\tau}$ could be $\sigma^2/2N$. In fact the estimate based on within-subject differences would have a variance $\sigma^2/4n_1 = \sigma^2/(2Nq)$, so that this suggests we should use q close to 1 if we believe that no carryover exists.

4. When λ in the model, the within-subject estimate of τ has a variance $N\sigma^2/(8n_1n_2) = \sigma^2/[(2N(q(1-q))]$ which recovers at most 25% of the information about τ , and then

only if $q = 1/2$, i.e. $n_1 = n_2$. The use of between-subject information reduces this variance to

$$\frac{\sigma^2(1 - \rho)(1 + 2\rho)}{N[4q(1 - q)(1 + \rho) + (1 + 2\rho)]},$$

which gives significant gains in efficiency, particularly when q is near $1/2$ and ρ is reasonably small, say less than 5. If $\rho = 0$ and $q = 1/2$, this achieves the minimum variance.

3.5 Two-treatment, three-period crossover design

As Mathews (1988) pointed out, although the use of more periods causes some problems in the experiment, there are several reasons why one may wish to extend the number of periods in a crossover design:

1. to allow the carryover effect to be estimated (within-subject estimate),
2. to make greater use of the experiment material,
3. to make it easier to compare more treatments in the one design.

Kershner & Federer (1981) calculate the variance of estimators for direct, carryover and total treatment effects for a number of two-treatment designs, using from two to four periods. The various optimality results were:

- the most efficient three-period design is ABB , BAA , and
- the most efficient four-period design is $AABB$, $BBAA$, $ABBA$, $BAAB$.

Table 3.13 shows the layout of the three-period design.

	Periods		
	1	2	3
Group	A	B	B
	B	A	A

Table 3.13: Two-treatment, three-period crossover design

The model is the same as that shown in the previous sections, but there are three active treatment periods. Thus, by using the usual constraints on the parameters, then the

linear model in terms of matrix form can be shown as below.

$$\begin{bmatrix} \bar{y}_{11.} \\ \bar{y}_{12.} \\ \bar{y}_{13.} \\ \bar{y}_{21.} \\ \bar{y}_{22.} \\ \bar{y}_{23.} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ -1 & -1 \\ -1 & 0 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \tau \\ \lambda \end{bmatrix} + \bar{\epsilon}, \quad (3.5.1)$$

or

$$y^* = \mu 1_6 + (1_2 \otimes I_3)P + (I_2 \otimes 1_3)B^* + T^*\theta + \epsilon^*. \quad (3.5.2)$$

If there are n_i subjects in the i th group ($i = 1, 2$), the variance-covariance of the design can be written as following.

$$Var(y^*) = \sigma^2(K^{-1} \otimes K_3) + (\sigma^2 + 3\sigma_s^2)(K^{-1} \otimes J_3).$$

The weighted orthogonal idempotent matrices are the same as those given earlier in Table 3.7.

Then the treatment information for this design can be written as

$$\begin{aligned} T^{*'}Q_1^*T^* &= \frac{2l^2}{3N} \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}, \\ T^{*'}Q_2^*T^* &= \frac{4n_1n_2}{3N} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \\ T^{*'}Q_3^*T^* &= \frac{8n_1n_2}{3N} \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}, \end{aligned}$$

and the estimates in the two strata are

$$\begin{aligned} \hat{\theta}_2 &= \frac{1}{2} \begin{bmatrix} [(\bar{y}_{21.} + \bar{y}_{22.} + \bar{y}_{23.}) - (\bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.})] \\ 0 \end{bmatrix} \\ \hat{\theta}_3 &= \frac{1}{8} \begin{bmatrix} [(2\bar{y}_{11.} - \bar{y}_{12.} - \bar{y}_{13.}) - (2\bar{y}_{21.} - \bar{y}_{22.} - \bar{y}_{23.})] \\ 2(\bar{y}_{12.} - \bar{y}_{13.} - \bar{y}_{22.} + \bar{y}_{23.}) \end{bmatrix} \end{aligned}$$

and the variance of estimates are

$$Var(\hat{\theta}_2) = (\sigma^2 + 3\sigma_s^2)(T^{*'}Q_2^*T^*)^{-} = \frac{3N(\sigma^2 + 3\sigma_s^2)}{4n_1n_2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$Var(\hat{\theta}_3) = \sigma^2(T^{*'}Q_3^*T^*)^{-1} = \frac{N\sigma^2}{32n_1n_2} \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}.$$

Combining estimates from the two strata we have the following estimates for parameters of interest by

$$\begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} = \left(\frac{1}{\sigma^2 + 3\sigma_s^2} T^{*'}Q_2^*T^* + \frac{1}{\sigma^2} T^{*'}Q_3^*T^* \right)^{-1} \left(\frac{1}{\sigma^2 + 3\sigma_s^2} T^{*'}Q_2^*T^*\hat{\theta}_2 + \frac{1}{\sigma^2} T^{*'}Q_3^*T^*\hat{\theta}_3 \right)$$

Then the estimates of two parameters can be obtained as

$$\hat{\tau} = \frac{1}{2(3 + 8\rho)} [(1 + 4\rho)(\bar{y}_{11.} - \bar{y}_{21.}) + (1 + 2\rho)(-\bar{y}_{12.} - \bar{y}_{13.} + \bar{y}_{22.} + \bar{y}_{23.})]$$

$$\hat{\lambda} = \frac{1}{4} [\bar{y}_{12.} - \bar{y}_{13.} - \bar{y}_{22.} + \bar{y}_{23.}],$$

and the covariance matrix for estimates is

$$Var\left(\begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix}\right) = \frac{N\sigma^2}{8n_1n_2} \begin{bmatrix} \frac{2(1+3\rho)}{3+8\rho} & 0 \\ 0 & 1 \end{bmatrix}.$$

As we can see the estimate of direct treatment is obtained from all data but it depends on variances and the variance of estimate not only is less than the corresponding variances in the standard design, that is, AB vs BA but also the correlation between them is zero, that is, $\hat{\tau}$ and $\hat{\lambda}$ is independent. In addition as we can see in this design the information in Subjects stratum is now about τ rather than λ . With this design the proportion of information on τ and λ lost into the Periods stratum is

$$\frac{2l^2}{8n_1n_2 + 2l^2} = \frac{2(n_1 - n_2)^2}{2(n_1 + n_2)^2} = (1 - 2q)^2.$$

The proportion of treatment information lost into the Subjects stratum is $\frac{4}{36} = \frac{1}{9}$, so there is not much treatment information to recover. This can be demonstrated by an equivalent formula which shows how little the estimate of τ is changed:

$$\hat{\tau} = \hat{\theta}_{3,1} - \frac{1}{8(3 + 8\rho)} \{(2\bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.}) - (2\bar{y}_{21.} + \bar{y}_{22.} + \bar{y}_{23.})\}. \quad (3.5.3)$$

and the variance of $\hat{\tau}$ is smaller than that of the within-subject estimate by a factor

$$8(1 + 3\rho)/\{3(3 + 8\rho)\},$$

which varies from $8/9$, when ρ is small and we recover all the between-subject variation, to 1, when ρ is large and there is no useful information to be recovered from between subjects.

Chapter 4

A Bayesian analysis for the general crossover design

4.1 Introduction

By sharing information among components of a statistical assessment and possibly including external objective evidence and personal opinion, Bayesian methods have the potential to produce more efficient and informative statistical analyses than those based on traditional approaches. In the Bayesian approach, the likelihood function is multiplied by another function due to prior belief in the values of parameters under consideration. This product gives us a new function which we refer to as the posterior distribution. Then in a Bayesian analysis we should consider the following concepts

1. an observed random variable or vector y ,
2. a parameter vector θ ,
3. the conditional density $f(y|\theta)$ of y given θ , and
4. the marginal density $p(\theta)$ which is called the prior distribution of θ .

In the two treatment, two-period crossover design, there are a series of papers which have used the Bayesian approach. Racine et al. (1986) and Grieve (1985), for example, presented a Bayesian analysis of the two-treatment, two-period design and Grieve (1986) used Bayesian analysis for this design when there are two baseline measurements in the

model. Recently Grieve (1994) considered missing values in the Bayesian approach to this design.

This chapter presents the Bayesian analysis of the crossover design. Emphasis is on analysis of the normal linear model together with normality of prior distribution for θ .

Most previous studies use un-informative prior distributions. It is argued here that we would generally have some expectation about the size of carryover as a proportion of the direct effect. If so this should be reflected in the prior distribution.

In Section 4.6 we analyse the Bayesian approach of two-treatment, two-period crossover design and in Section 4.7 we consider this design with one baseline measurement. Bayesian analysis of two-treatment with two baseline measurements is discussed in Section 4.8. Bayesian analysis of Balaam's design is considered in Section 4.9. In Section 4.10 we discuss Bayesian analysis of the two-treatment, three-period crossover design.

Suppose in a crossover design that treatments are to be compared over p periods and that there are gn subjects randomised to g equal-size groups, such that in each group a sequence of treatments is applied, and measurements made at intervals of time. The treatments may include absence of treatment and measurements may be taken in these periods. To see this more clearly, consider the following layout for general crossover designs shown in Table 4.1.

		Period			
		1	2	...	p
Group	1				
	2			T_{ij}	
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	g				

Table 4.1: Layout of general crossover design

The treatment T_{ij} is applied to all subjects in group i and period j ; where T_{ij} may be a null treatment or a combination of direct and carryover effects.

4.2 The statistical linear model

We assume that the model of responses follows the mixed linear model in (2.2.1) in Chapter 2, but with observations denoted by y_{ijk} , corresponding to the i th group, the j period, and the k th subject within the i th group ($k = 1, \dots, n$). Note that the order of these subscripts differs from that in Chapter 3, in order to facilitate the Kronecker product notation required in this chapter.

The model (2.3.1) can then be written

$$y = T\theta + \xi \quad (4.2.1)$$

where $y = (y'_{11}, \dots, y'_{1p}, \dots, y'_{g1}, \dots, y'_{gp})'$ is an $gpn \times 1$ vector of observation, such that $y_{ij} = (y_{ij1}, y_{ij2}, \dots, y_{ijn})'$ for $(i = 1, 2, \dots, g; j = 1, 2, \dots, p)$, T is a $gpn \times t$ matrix of known coefficients, θ is a $t \times 1$ vector of parameters and ξ is a $gpn \times 1$ vector of random effects. Again, for the moment, we subsume all other effects into the covariance matrix.

We also assume that the elements of ξ are jointly normally distributed. The model says simply that the conditional distribution of y given parameters θ and the variances of the random effects is the multivariate normal distribution $N(T\theta, Q)$. From Section 2.2.2, the matrices which appear in the covariance matrix Q are mutually orthogonal idempotent matrices so we can write Q as follows

$$Q = \sum_i \delta_i Q_i,$$

where Q_i are mutually orthogonal idempotent matrices of rank r_i , as given in Table 4.2. We note that in this formulation, the dimension of the first matrix in the Kronecker product is g , the second is p and the third is n . Where it is obvious from the context, we drop the subscripts from I, J and K .

i	Stratum	δ_i	r_i	Q_i
0	Grand Mean	$\sigma^2 + p\sigma_s^2$	1	$J \otimes J \otimes J$
1	Period	σ^2	$p - 1$	$J \otimes K \otimes J$
2	Subject	$\sigma^2 + p\sigma_s^2$	$ng - 1$	$(I \otimes J \otimes I) - (J \otimes J \otimes J)$
3	Period $\times$ Subject	σ^2	$(p - 1)(ng - 1)$	$(I \otimes K \otimes I) - (J \otimes K \otimes J)$

Table 4.2: Idempotent matrices of each stratum

As before, an analysis of this model to determine the estimates of θ in the various strata, would not include $i = 0$ since there is no useful information about treatment parameters in the Grand Mean stratum. Throughout this chapter, therefore, we will omit the term for δ_0 . We may, however, include $i = 1$ if there were information available, in the form of σ_π^2 , which would allow recovery of the treatment information in the Periods stratum. Generally, however, we would set $\sigma_\pi^2 = \infty$, so that $\delta_1 = 0$ and the Periods stratum would not contribute to the estimation of θ .

The fact that all n observations in a group have the same treatment regime implies that

$$E(y) = \eta \otimes 1 = T^* \theta \otimes 1 \quad (4.2.2)$$

and

$$Var(y) = Q,$$

where η is $\{\eta_{ij}\}$ in standard order and $T = T^* \otimes 1_n$. In order to work on means let us define the average of responses by

$$\bar{y} = (\bar{y}_{11}, \bar{y}_{12}, \dots, \bar{y}_{1p}, \dots, \bar{y}_{gp})' = \frac{1}{n}(I \otimes I \otimes 1)'y \quad (4.2.3)$$

4.3 Likelihood and parameter estimation

The likelihood function becomes

$$\begin{aligned} L &= p(y|\theta, Q) = |Q|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y - T\theta)'Q^{-1}(y - T\theta)\right\} \\ &= \prod_{i=1}^3 \delta_i^{-\frac{1}{2}r_i} \exp\left\{-\frac{1}{2} \sum_{i=1}^3 \left\{\frac{1}{\delta_i}(y - T\theta)'Q_i(y - T\theta)\right\}\right\} \end{aligned} \quad (4.3.1)$$

since

$$\begin{aligned} |Q| &= \det(Q) = \prod_{i=0}^3 \delta_i^{r_i}, \\ Q^{-1} &= \sum_{i=1}^3 \frac{1}{\delta_i} Q_i. \end{aligned}$$

To simplify the exponent term in the likelihood function we define

$$y = Gy + (I - G)y;$$

Maximising L can be achieved by maximising the logarithm of L which shall be denoted by, l , namely

$$l = -\frac{1}{2} \sum_{i=1}^3 r_i \log \delta_i - \frac{1}{2} \sum_{i=1}^3 \left\{ \frac{1}{\delta_i} (\bar{y} - T^* \theta)' Q_i^* (\bar{y} - T^* \theta) \right\} - \frac{S_2}{2(\sigma^2 + p\sigma_s^2)} - \frac{S_3}{2\sigma^2}, \quad (4.3.4)$$

and the maximum likelihood estimate of θ is obtained by differentiating l and equating it to zero

$$\frac{\partial l}{\partial \theta} = - \sum_{i=1}^3 \frac{1}{\delta_i} T^{*'} Q_i^* (\bar{y} - T^* \theta) = 0. \quad (4.3.5)$$

If we define $Q^* = \sum \frac{1}{\delta_i} Q_i^*$, then the solution is obtained as

$$\hat{\theta} = (T^{*'} Q^* T^*)^{-1} T^{*'} Q^* \bar{y}, \quad (4.3.6)$$

and the distribution of $\hat{\theta}$ given θ is

$$\hat{\theta} | \theta \sim N\{\theta, (T^{*'} Q^* T^*)^{-1}\},$$

provided σ^2 and $\sigma^2 + p\sigma_s^2$ are known. Using the marginal likelihood for S_2 and S_3 , we obtain

$$\begin{aligned} \hat{\sigma}^2 + p\hat{\sigma}_s^2 &= \left[\frac{S_2}{g(n-1)} \right], \\ \hat{\sigma}^2 &= \left[\frac{S_3}{g(p-1)(n-1)} \right], \end{aligned}$$

providing unbiased estimates of σ^2 and $\sigma^2 + p\sigma_s^2$ respectively. S_2 has $g(n-1)$ and S_3 has $g(p-1)(n-1)$ degrees of freedom. If in fact the rank of T^* is less than gp , there will be some degrees of freedom available which may provide further estimates of σ^2 and/or $\sigma^2 + p\sigma_s^2$. In general, they will have relatively few degrees of freedom and hence we choose, for simplicity, to ignore this information.

If σ^2 and $\sigma^2 + p\sigma_s^2$ are unknown and have to be replaced by estimates, clearly the multivariate Normal distribution now becomes similar to a multivariate t -distribution. In literature, in this regard, little has been done in general. However, in the two-treatment two-period case the combined estimate of λ has the distribution

$$\hat{\lambda} = \bar{y}_{1..} - \bar{y}_{2..} \sim N \left(\lambda, \frac{m\sigma^2(1+2\rho)}{2} \right),$$

where $m = N/(n_1 n_2)$, and an estimate of the variance is provided by

$$S_2 \sim \sigma^2(1+2\rho)\chi_{2(n-1)}^2.$$

Hence

$$\frac{(\hat{\lambda} - \lambda)}{\sqrt{mS_2/4\{(n-1)\}}} \sim t_{2(n-1)}.$$

However,

$$\hat{\tau} = \frac{1}{2}(\bar{y}_{11.} - \bar{y}_{21.}) \sim N\left(\tau, \frac{m\sigma^2(1+\rho)}{4}\right),$$

and we do not have a χ^2 -distribution to estimate $\sigma^2(1+\rho)$. The estimate of this relies on a combination of S_2 and S_3 . The result of this, as Grieve (1985) shows in Bayesian context, is that the joint distribution of $(\hat{\tau}, \hat{\lambda})$ is a combination of two independent multivariate t -distributions, which provides a Behrens-Fisher distribution. Patil (1965) shows that this is well approximated by a multivariate t -distribution.

4.4 Bayesian analysis

Suppose that the prior of θ is $\theta \sim N(\theta_0, \Sigma_0)$, where Σ_0 may be singular. Now we discuss the Bayesian approach for our model in two sections, one when Σ_0 is a nonsingular covariance matrix and another section when it is a singular matrix.

4.4.1 Σ_0 is a nonsingular matrix

In this case of a design with g groups and p periods, when Σ_0 is full rank, by following Lindley & Smith (1972) and using Bayes theorem, the posterior distribution for θ is given by

$$\theta|\hat{\theta} \sim N(\theta_p, \Sigma_p) \quad (4.4.1)$$

where

$$Var(\theta|\hat{\theta}) = \Sigma_p = \left(\sum_i \frac{1}{\delta_i} T^{*'} Q_i^* T^* + \Sigma_0^{-1}\right)^{-1}$$

and

$$E(\theta|\hat{\theta}) = \theta_p = \left(\sum_i \frac{1}{\delta_i} T^{*'} Q_i^* T^* + \Sigma_0^{-1}\right)^{-1} \left(\sum_i \frac{1}{\delta_i} T^{*'} Q_i^* \bar{y} + \Sigma_0^{-1} \theta_0\right). \quad (4.4.2)$$

The posterior mean is a vector-weighted average of the prior mean θ_0 and the maximum likelihood estimate, $\hat{\theta}$. The weights are respectively the prior information matrix Σ_0^{-1} and the inverse of the covariance matrix of the maximum likelihood $\hat{\theta}$. The posterior

covariance matrix is the inverse of the sum of these weights, and the posterior covariance matrix is smaller than either the prior covariance Σ_0 or

$$(\sum_i \frac{1}{\delta_i} T^{*'} Q_i^* T^*)^{-1}$$

in the sense that

$$(\sum_i \frac{1}{\delta_i} T^{*'} Q_i^* T^*)^{-1} - (\sum_i \frac{1}{\delta_i} T^{*'} Q_i^* T^* + \Sigma_0^{-1})^{-1},$$

is non-negative definite. It follows that

$$Var(d'\theta|\hat{\theta}) \leq Var(d'\hat{\theta}), \quad (4.4.3)$$

for all vectors d .

In general, suppose we are interested in some contrast $d'\theta$. For example, if $\theta = [\mu, \tau_1, \tau_2, \lambda_1, \lambda_2]'$ and we are interested in the posterior distribution for $(\tau_1 - \tau_2)$, then

$$Var(\tau_1 - \tau_2|\hat{\theta}) = d'(T^{*'} Q^* T^* + \Sigma_0^{-1})^{-1} d,$$

where $d = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \end{bmatrix}'$. Then the posterior distribution of $(\tau_1 - \tau_2)$ would be

$$(\tau_1 - \tau_2)|\hat{\theta} \sim N(d'\theta_p, d'\Sigma_p d). \quad (4.4.4)$$

4.4.2 Σ_0 is a singular covariance matrix

If the covariance matrix Σ_0 is not full rank, this implies that certain directions in the l -dimensional parameter space have no prior information. This can be expressed by regarding the information matrix Σ_0^- as having zero eigenvalues in these directions. Thus we write

$$\Sigma_0 = \sum_{i=1}^k \gamma_i m_i m_i' = M \Gamma M', \quad (4.4.5)$$

where $k = rank(\Sigma_0)$, $m_i' m_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta. Then the information matrix or g-inverse with minimum rank is

$$\Sigma_0^- = \sum_{i=1}^k \frac{1}{\gamma_i} m_i m_i',$$

and the posterior distribution for θ is Normal with

$$Var(\theta|\hat{\theta}) = \Sigma_p = (\sum_i \frac{1}{\delta_i} T^{*'} Q_i^* T^* + \Sigma_0^-)^{-1}$$

and

$$E(\theta|\hat{\theta}) = \theta_p = (\sum_i \frac{1}{\delta_i} T^{*'} Q_i^* T^* + \Sigma_0^-)^{-1} (\sum_i \frac{1}{\delta_i} T^{*'} Q_i^* \bar{y} + \Sigma_0^- \theta_0). \quad (4.4.6)$$

On all occasions we will use the minimum rank symmetric g -inverse of Σ_0 .

If σ^2 and $\sigma^2 + p\sigma_s^2$ are unknown and have to be replaced by estimates, then $\hat{\theta}|\theta$ is a mixture of two non-central multivariate t -distributions. There is little on this in the literature. However, Grieve (1985) considers the two-treatment, two-period crossover design with N subjects and, possibly, unequal group sizes of n_1 and n_2 , and from the classical analysis approach of this design, he summarises the following independent distributions for all parameters in the model.

$$\begin{aligned} (\hat{\mu}, \hat{\lambda})' &\sim N\{(\mu, \lambda), \Sigma_2\}; \\ (\hat{\pi}, \hat{\tau})' &\sim N\{(\pi, \tau - \lambda/2), \Sigma_3\}; \\ S_2 &\sim (\sigma^2 + 2\sigma_s^2) \chi_{N-2}^2; \\ S_3 &\sim \sigma^2 \chi_{N-2}^2 \end{aligned}$$

where

$$\begin{aligned} \Sigma_2 &= (\sigma^2 + 2\sigma_s^2) \begin{bmatrix} m/8 & l/4 \\ l/4 & m/2 \end{bmatrix}, \\ \Sigma_3 &= \sigma^2 \begin{bmatrix} m/8 & l/8 \\ l/8 & m/4 \end{bmatrix}, \end{aligned}$$

where $N = n_1 + n_2$, $m = N/(n_1 n_2)$ and $l = (n_1 - n_2)/(n_1 n_2)$. By applying the following un-informative prior for all parameters in the model

$$p(\mu, \pi, \tau, \lambda, \sigma^2, \sigma_s^2) \propto \frac{1}{\sigma^2(\sigma^2 + 2\sigma_s^2)},$$

Grieve (1985) shows approximately that the marginal distribution of $\phi_1 = \tau - \lambda/2$ and $\phi_2 = \lambda/2$ can be obtained as

$$\begin{aligned} p(\phi_1) &\propto \left\{ S_3 + \frac{8(\phi_1 - \hat{\tau})^2}{m} \right\}^{-\frac{N-1}{2}} \\ p(\phi_2) &\propto \left\{ S_2 + \frac{8(\phi_2 - \hat{\lambda}/2)^2}{m} \right\}^{-\frac{N-1}{2}}. \end{aligned}$$

By referring to the result of Patil (1965) who showed that, if ϕ_1 and ϕ_2 are two parameters which have independent shifted and scaled t distributions, then the sum of two

parameters has Behrens-Fisher distribution, Grieve (1985) then shows that $\tau = \phi_1 + \phi_2$ is approximated by

$$p(\tau) \propto \left\{ 1 + \frac{8[\tau - (\hat{\tau} + \hat{\lambda}/2)]^2}{mh} \right\}^{-\frac{f+1}{2}}.$$

where

$$\begin{aligned} f &= \frac{(S_3 + S_2)^2(N-6)}{S_3^2 + S_2^2} + 4, \\ h &= \frac{(f-2)(S_3 + S_2)}{N-4}. \end{aligned}$$

Law (1987) in his thesis develops a Bayesian analysis of higher-order crossover designs. For a subset of the parameters, θ , of length r , he use a prior distribution which is multivariate Normal distribution, with mean θ_0 and covariance matrix $\sigma^2 \Sigma_0$. He also uses a Gamma Distribution for the prior of σ^2 . Thus

$$\begin{aligned} p(\sigma^2) &\propto \left[\frac{1}{\sigma} \right]^{2(\alpha+1)} \exp \left[-\frac{\beta}{\sigma^2} \right] \\ p(\theta|\sigma^2) &\propto \left[\frac{1}{\sigma} \right]^r \exp \left[-\frac{(\theta - \theta_0)' \Sigma_0^{-1} (\theta - \theta_0)}{2\sigma^2} \right], \end{aligned}$$

where α , β , θ_0 and Σ_0 are assumed known. Then the prior density of θ can be derived as

$$p(\theta) \propto \left[1 + \frac{1}{2\alpha} (\theta - \theta_0)' \frac{2\alpha \Sigma_0^{-1}}{2\beta} (\theta - \theta_0) \right]^{-\frac{(r+2\alpha)}{2}} \sim t_r \left[\theta_0, \frac{\beta}{\alpha} \Sigma_0, 2\alpha \right], \quad (4.4.7)$$

a multivariate t -distribution, where r is dimension of θ . He then uses the within-subject estimate of θ as $\hat{\theta}|\theta \sim N \{ \theta, \sigma^2(T'T)^{-1} \}$ and shows that the marginal posterior density for θ is also the multivariate t -distribution

$$\theta|y \sim t_r [\theta_p, \Sigma_p, \nu + r + 2\alpha],$$

where

$$\begin{aligned} \theta_p &= (T'T + \Sigma_0^{-1})^{-1} (T'T\hat{\theta} + \Sigma_0^{-1}\theta_0) \\ \Sigma_p &= a(T'T + \Sigma_0^{-1})^{-1}, \end{aligned}$$

and

$$a = \frac{1}{(\nu + r + 2\alpha)} \left[2\beta + \theta_0' \Sigma_0^{-1} \theta_0 + y'y - (T'T\hat{\theta} + \Sigma_0^{-1}\theta_0)' (T'T + \Sigma_0^{-1})^{-1} (T'T\hat{\theta} + \Sigma_0^{-1}\theta_0) \right],$$

and ν is the degrees of freedom of S_3 . The only reason this method works is that Law (1987) makes the somewhat unusual assumption that the prior for θ involves the unknown σ^2 . Without this, the multivariate t -distribution is not obtained. Similarly, in our case, to include the between-subject estimates as well, the complication is that there is a factor $\sigma^2(1 + p\rho)$ and, unless ρ is known, the terms do not combine conveniently.

4.5 Choice of prior distribution

We have four strata in the crossover design which may contain treatment information, with weights $\delta_i (i = 0, \dots, 3)$. As before we are generally not interested in using $i = 0, 1$, although we might use $i = 1$ if we had a prior distribution for the period effects π . In addition in practice the $\delta_2 = \sigma^2 + p\sigma_s^2$ and $\delta_3 = \sigma^2$ are not known and we should replace these terms by their estimates or use prior distributions and integrate them out if necessary.

In the following sections we will show the Bayesian analysis in certain crossover designs. But first, we need to consider the type of prior distributions which might be used for (τ, λ) .

As we showed in the previous chapter, the expected values of responses in the two-treatment, two-period crossover design when the period effect is random can be represented in Table 4.4.

$\mu + \tau$	$\mu - \tau + \lambda$
$\mu - \tau$	$\mu + \tau - \lambda$

Table 4.4: Expected values of responses in two-treatment, two-period design.

In our models, the difference between the two treatments A and B is equal to 2τ , and the difference in their carryover effects is equal to 2λ . Suppose that treatment B is a standard on-going treatment, like a baseline, for which the true value is μ_o and for which no carryover effect is expected. If we now look at the way treatment A departs from this baseline, we can model the four means by the revised Table 4.5.

$\mu_0 + 2\tau$	$\mu_0 + 2\lambda + \pi$
μ_0	$\mu_0 + 2\tau + \pi$

Table 4.5: Expected values of responses when treatment B is a standard treatment.

This table provides a model in which the active drug A has a treatment effect relative to baseline and in which we might expect any carryover effect to be a percentage of that treatment effect. Such a relationship of treatment and carryover effect can be shown in

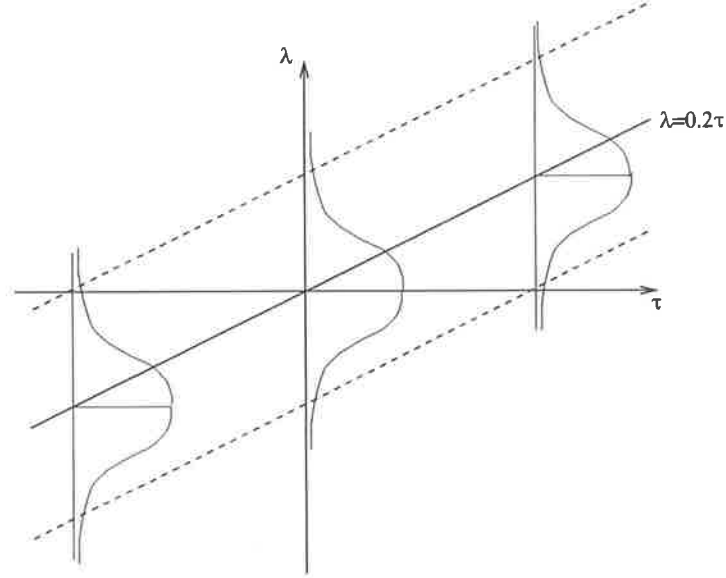
Figure 4.1, where our prior distribution for the carryover effect is that it is a proportion k of the treatment effect, as expressed by

$$p(\lambda|\tau) \sim N(k\tau, \sigma_0^2), \quad (4.5.1)$$

where k is a small value, probably in the range $(0 - 0.5)$ and σ_0^2 expresses our degree of prior certainty about that relationship. We note that the case $k = 0$ corresponds to the situation considered by Selwyn et al. (1981). If we assume an un-informative prior on τ , the joint distribution of λ and τ is

$$p(\tau, \lambda) = p(\tau)p(\lambda|\tau) \propto \exp\{-(\lambda - k\tau)^2/2\sigma_0^2\}. \quad (4.5.2)$$

Figure 4.1: Conditional distribution of carryover effect given direct effect



Now we can write

$$(\lambda - k\tau)^2/2\sigma_0^2 = \frac{1}{2\sigma_0^2} \begin{bmatrix} \tau & \lambda \end{bmatrix} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} \begin{bmatrix} \tau \\ \lambda \end{bmatrix}, \quad (4.5.3)$$

so that the prior density can be represented by a singular bivariate Normal with information matrix

$$\Sigma_0^- = \frac{1}{\sigma_0^2} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix}.$$

Following (4.4.5), this can be represented as a singular covariance matrix

$$\Sigma_0 = \frac{\sigma_0^2}{(1 + k^2)^2} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix}.$$

4.6 Two-treatment, two-period design

In Chapter 3, we obtained between-subject and within-subject estimates for θ in the Group and the Period $\times$ Group strata as

$$\begin{aligned}\hat{\theta}_2 &= \begin{bmatrix} 0 \\ \bar{y}_{1..} - \bar{y}_{2..} \end{bmatrix} \sim N \left\{ \begin{bmatrix} 0 \\ \lambda \end{bmatrix}, \frac{N(\sigma^2 + 2\sigma_s^2)}{2n_1n_2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}, \\ \hat{\theta}_3 &= \frac{1}{10}(d_{1.} - d_{2.}) \begin{bmatrix} 2 \\ -1 \end{bmatrix} \sim N \left\{ (2\tau - \lambda) \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \frac{N\sigma^2}{50n_1n_2} \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} \right\}.\end{aligned}$$

We noted that, in combining these estimates, we needed to weight $\hat{\theta}_2$ and $\hat{\theta}_3$ by the inverse of those covariance matrices, using a minimum rank g-inverse in cases where the covariance matrix is singular.

When this is extended to include a prior distribution for θ , the same principle applies, so that if

$$\theta \sim N(\theta_0, \Sigma_0), \quad (4.6.1)$$

where Σ_0 may be singular, the posterior distribution for θ , given (σ^2, ρ) , is given by

$$(\theta | \hat{\theta}, \sigma^2, \rho) \sim N(\theta_p, \Sigma_p), \quad (4.6.2)$$

where

$$\Sigma_p = (T^{*'}Q^*T^* + \Sigma_0^-)^{-1}$$

and

$$\begin{aligned}\theta_p &= \Sigma_p(T^{*'}Q^*T^*\hat{\theta} + \Sigma_0^-\theta_0) \\ &= \Sigma_p\{(T^{*'}Q^*T^* + \Sigma_0^-)\hat{\theta} + \Sigma_0^-(\theta_0 - \hat{\theta})\} \\ &= \hat{\theta} - \Sigma_p\Sigma_0^-(\hat{\theta} - \theta_0),\end{aligned} \quad (4.6.3)$$

so the mean of the posterior is shifted from $\hat{\theta}$ towards θ_0 depending on the distance between $\hat{\theta}$ and θ_0 and the strength of the prior information.

4.6.1 Posterior estimates

Combining the two estimates appropriately with the prior for (τ, λ) we obtain

$$\Sigma_p = \left\{ \frac{2n_1n_2}{N\sigma^2} \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} + \frac{2n_1n_2}{N(\sigma^2 + 2\sigma_s^2)} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{\sigma_0^2} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} \right\}^{-1}$$

$$= \frac{N\sigma^2(1+2\rho)}{2n_1n_2} \left\{ a \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} + b \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}^{-1}$$

where $b = (1 + 2\rho)$, and $a = N\sigma^2(1 + 2\rho)/(2n_1n_2\sigma_0^2)$. Then, after simplification,

$$\Sigma_p = \frac{N\sigma^2(1+2\rho)}{2n_1n_2\Delta} \left\{ a \begin{bmatrix} 1 & k \\ k & k^2 \end{bmatrix} + b \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\}, \quad (4.6.4)$$

where

$$\Delta = a[b(2-k)^2 + k^2] + 4b.$$

If we assume that $\theta_0 = 0$, the posterior distribution for θ is $N(\theta_p, \Sigma_p)$ where Σ_p is given above and (4.6.3) implies that

$$\begin{aligned} \theta_p &= \hat{\theta} - \frac{a}{\Delta} \left\{ a \begin{bmatrix} 1 & k \\ k & k^2 \end{bmatrix} + b \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\} \begin{bmatrix} k \\ -1 \end{bmatrix} \begin{bmatrix} k & -1 \end{bmatrix} \hat{\theta} \\ &= \hat{\theta} + \frac{a}{\Delta} \begin{bmatrix} k - b(2-k) \\ -2b(2-k) \end{bmatrix} (\hat{\lambda} - k\hat{\tau}). \end{aligned}$$

If $\hat{\lambda} = k\hat{\tau}$ then the estimate is unchanged, as would be expected since the data is consistent with the prior distribution.

In other cases, however, the maximum likelihood estimate $\hat{\theta}$ is shifted towards the line $\lambda = k\tau$.

$$\begin{aligned} \theta_p &= \frac{1}{\Delta} \begin{bmatrix} [\Delta - ak\{k - b(2-k)\}]\hat{\tau} + a\{k - b(2-k)\}\hat{\lambda} \\ 2abk(2-k)\hat{\tau} + \{\Delta - 2ab(2-k)\}\hat{\lambda} \end{bmatrix} \\ &= \frac{4b}{\Delta} \begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} + \frac{a}{\Delta} \begin{bmatrix} 1 \\ k \end{bmatrix} \{k\hat{\lambda} + b(2-k)(2\hat{\tau} - \hat{\lambda})\} \end{aligned} \quad (4.6.5)$$

$$(4.6.6)$$

Now, k will typically be small, certainly less than about 0.5, and b is at least one. Thus, the last term is dominated by the fact that most of our information in the data is about $(2\hat{\tau} - \hat{\lambda})$ and there is relatively little information about λ .

The two particular cases of interest here are (i) when $a = 0$ in which case there is no prior information and the estimate reverts to $\hat{\theta}$, and (ii) when $a \rightarrow \infty$, for then we know that $\lambda = k\tau$ and the estimate has to lie along that line. The above formula then

simplifies to

$$\theta_p = \frac{k\hat{\lambda} - (1+2\rho)(2-k)(\hat{\lambda} - 2\hat{\tau})}{k^2 + (1+2\rho)(2-k)^2} \begin{bmatrix} 1 \\ k \end{bmatrix}, \quad (4.6.7)$$

so that the posterior estimate for λ is k times the posterior estimate for τ . As σ_0^2 varies between these two extremes, the posterior estimate (4.6.5) moves along a straight line joining $\hat{\theta}$ and the point (4.6.7) on the line $\lambda = k\tau$.

4.7 Two-treatment, two-period with one baseline measurement

This section presents the Bayesian analysis of the 2×2 crossover design with one baseline measurement taken before a treatment is given to the subject. We consider that the experiment yields complete data on n_1 subjects in the first group and n_2 subjects in the second group.

From Chapter 3 for this design we obtained the following between-subject and within-subject estimates from Group and Group $\times$ Period strata respectively

$$\begin{aligned} \hat{\theta}_2 &= \begin{bmatrix} 0 \\ \bar{y}_{1..} - \bar{y}_{2..} \end{bmatrix} \sim N \left\{ \begin{bmatrix} 0 \\ \lambda \end{bmatrix}, \frac{3N(\sigma^2 + 3\sigma_s^2)}{4n_1n_2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}, \\ \hat{\theta}_3 &= \frac{1}{2} \begin{bmatrix} \bar{y}_{12.} - \bar{y}_{11.} + \bar{y}_{21.} - \bar{y}_{22.} \\ \bar{y}_{12.} + \bar{y}_{13.} - 2\bar{y}_{11.} - \bar{y}_{22.} - \bar{y}_{23.} + 2\bar{y}_{21.} \end{bmatrix} \\ &\sim N \left\{ \begin{bmatrix} \tau \\ \lambda \end{bmatrix}, \frac{N\sigma^2}{4n_1n_2} \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix} \right\} \end{aligned}$$

If we apply the joint prior density in (4.5.2), then these estimates can be combined with the prior for (τ, λ) to obtain

$$\begin{aligned} \Sigma_p &= \left\{ \frac{4n_1n_2}{3N\sigma^2} \begin{bmatrix} 6 & -3 \\ -3 & 2 \end{bmatrix} + \frac{4n_1n_2}{3N(\sigma^2 + 3\sigma_s^2)} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{\sigma_0^2} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} \right\}^{-1} \\ &= \frac{3N(1+3\rho)\sigma^2}{4n_1n_2} \left\{ a \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} + b \begin{bmatrix} 6 & -3 \\ -3 & 2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}^{-1} \end{aligned}$$

where $b = 1 + 3\rho$ and $a = 3N\sigma^2(1 + 3\rho)/(4n_1n_2\sigma_0^2)$. Then

$$\Sigma_p = \frac{3N(1 + 3\rho)\sigma^2}{4n_1n_2\Delta} \left\{ a \begin{bmatrix} 1 & k \\ k & k^2 \end{bmatrix} + b \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\},$$

where $\Delta = 2ab(k^2 - 3k + 3) + 3b(b + 2) + ak^2$. Using (4.6.3), we obtain the posterior mean as given by

$$\begin{aligned} \theta_p &= \hat{\theta} - \Sigma_p \frac{1}{\sigma_0^2} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} \hat{\theta} \\ &= \hat{\theta} - \frac{a}{\Delta} \left\{ a \begin{bmatrix} 1 & k \\ k & k^2 \end{bmatrix} + b \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \begin{bmatrix} k \\ -1 \end{bmatrix} \begin{bmatrix} k & -1 \end{bmatrix} \hat{\theta} \\ &= \hat{\theta} + \frac{a}{\Delta} \left\{ b \begin{bmatrix} 2k - 3 \\ 3k - 6 \end{bmatrix} + \begin{bmatrix} k \\ 0 \end{bmatrix} \right\} (\hat{\lambda} - k\hat{\tau}) \\ &= \hat{\theta} + \frac{a}{\Delta} \begin{bmatrix} k + (2k - 3)(1 + 3\rho) \\ (3k - 6)(1 + 3\rho) \end{bmatrix} (\hat{\lambda} - k\hat{\tau}) \end{aligned} \quad (4.7.1)$$

As we can see, if $\hat{\lambda} = k\hat{\tau}$, the estimate of θ is unchanged. Otherwise the mean of the posterior for θ is given above. We can re-express this in terms of a weighted mean between $\hat{\theta}$ on the one hand, and a vector which lies on the line $\lambda = k\tau$, as follows:

$$\begin{aligned} \theta_p &= \frac{1}{\Delta} \begin{bmatrix} [\Delta - ak\{k - b(3 - 2k)\}]\hat{\tau} + a\{k - b(3 - 2k)\}\hat{\lambda} \\ 3abk(2 - k)\hat{\tau} + \{\Delta - 3ab(2 - k)\}\hat{\lambda} \end{bmatrix} \\ &= \frac{3b(b + 2)}{\Delta} \begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} + \frac{a}{2\Delta} \begin{bmatrix} 1 \\ k \end{bmatrix} \{3b(2 - k)(2\hat{\tau} - \hat{\lambda}) + k(b + 2)\hat{\lambda}\}, \end{aligned} \quad (4.7.2)$$

which again highlights, since k is typically very much less than 1, that there is strong information about $(2\tau - \lambda)$ but rather weaker information about λ .

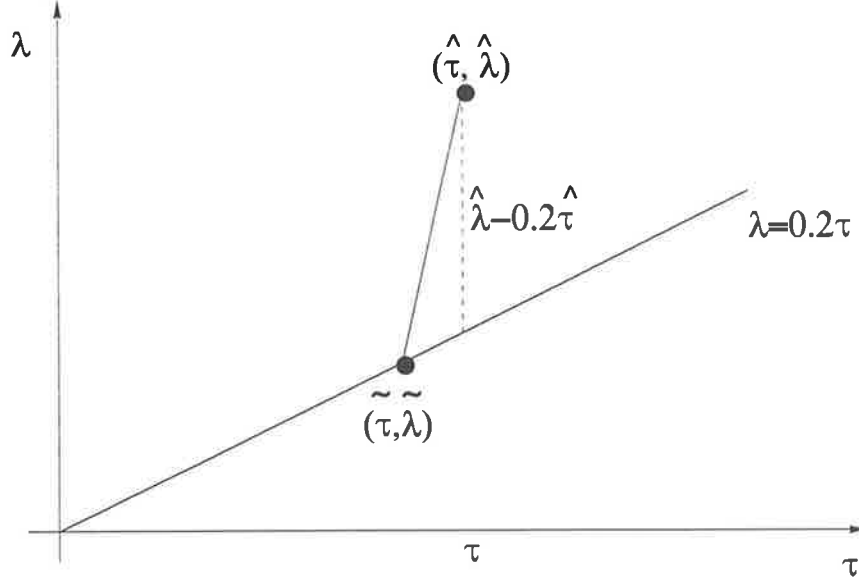
As an example of how the parameter estimates change with a and b , consider the case $k = 0.2$. Then, ignoring the relatively small ak^2 term in Δ , (4.7.2) reduces to

$$\begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} - \frac{a(\hat{\lambda} - 0.2\hat{\tau})}{4.88a + 3(b + 2)} \begin{bmatrix} 2.6 \\ 5.4 \end{bmatrix}$$

Figure 4.2 shows what happens according to the strength of the prior information.

If $\sigma_0^2 = \infty$, we have $a = 0$ and we get the maximum likelihood estimates $(\hat{\tau}, \hat{\lambda})$, an arbitrary point on the plane. The vertical distance from this point to the line $\lambda = 0.2\tau$ is then $(\hat{\lambda} - 0.2\hat{\tau})$. If this is positive, as in the figure, the posterior mean θ_p will lie along

Figure 4.2: Relation between posterior estimates when $k = 0.2$



the steep angled line, and will be somewhere between $\hat{\theta}$ and the point denoted by $(\hat{\tau}, \hat{\lambda})$ in Figure 4.2. This point is the one obtained as $a \rightarrow \infty$.

4.8 Two-treatment, two-period with two baseline measurements

This section presents the Bayesian analysis of the 2×2 crossover design with two baseline measurements which have been taken before each treatment is given to the subject. We consider that the experiment yields complete data on n_1 subjects in the first group and n_2 subjects in the second group.

From Chapter 3 for this design we obtained the following estimates

$$\begin{aligned} \hat{\theta}_2 &= \frac{1}{2} \begin{bmatrix} 0 \\ \bar{y}_{1..} - \bar{y}_{2..} \end{bmatrix} \sim N \left\{ \begin{bmatrix} 0 \\ \lambda \end{bmatrix}, \frac{N(\sigma^2 + 4\sigma_s^2)}{n_1 n_2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \\ \hat{\theta}_3 &= \frac{1}{2} \begin{bmatrix} 2(\bar{y}_{12.} - \bar{y}_{22.}) - (\bar{y}_{11.} + \bar{y}_{13.} - \bar{y}_{21.} - \bar{y}_{23.}) \\ 2(\bar{y}_{12.} + \bar{y}_{14.} - \bar{y}_{11.} - \bar{y}_{13.}) + (\bar{y}_{21.} + \bar{y}_{23.} - \bar{y}_{22.} - \bar{y}_{24.}) \end{bmatrix} \\ &\sim N \left\{ \begin{bmatrix} \tau \\ \lambda \end{bmatrix}, \frac{N\sigma^2}{8n_1 n_2} \begin{bmatrix} 3 & 4 \\ 4 & 8 \end{bmatrix} \right\} \end{aligned}$$

If we apply the joint prior density in (4.5.2), then the combination of these densities using Bayes theorem can be given by combining the two estimates appropriately with a prior

for (τ, λ) to obtain

$$\begin{aligned}\Sigma_p &= \left\{ \frac{n_1 n_2}{N \sigma^2} \begin{bmatrix} 8 & -4 \\ -4 & 3 \end{bmatrix} + \frac{n_1 n_2}{N(\sigma^2 + 4\sigma_s^2)} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{\sigma_0^2} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} \right\}^{-1} \\ &= \frac{N(1 + 4\rho)\sigma^2}{n_1 n_2} \left\{ a \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} + b \begin{bmatrix} 8 & -4 \\ -4 & 3 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}^{-1}\end{aligned}$$

where $b = 1 + 4\rho$ and $a = N\sigma^2(1 + 4\rho)/(n_1 n_2 \sigma_0^2)$. Then

$$\Sigma_p = \frac{N(1 + 4\rho)\sigma^2}{n_1 n_2 \Delta} \left\{ a \begin{bmatrix} 1 & k \\ k & k^2 \end{bmatrix} + b \begin{bmatrix} 3 & 4 \\ 4 & 8 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\},$$

where $\Delta = ab(3k^2 - 8k + 8) + 8b(b + 1) + ak^2$. The posterior distribution is $N(\theta_p, \Sigma_p)$, where

$$\begin{aligned}\theta_p &= \hat{\theta} - \left\{ \Sigma_p^{-1} \frac{1}{\sigma_0^2} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} \right\} \hat{\theta} \\ &= \hat{\theta} - \frac{a}{\Delta} \left\{ a \begin{bmatrix} 1 & k \\ k & k^2 \end{bmatrix} + b \begin{bmatrix} 3 & 4 \\ 4 & 8 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\} \begin{bmatrix} k \\ -1 \end{bmatrix} \begin{bmatrix} k & -1 \end{bmatrix} \hat{\theta} \\ &= \hat{\theta} + \frac{a}{\Delta} \left\{ b \begin{bmatrix} 3k - 4 \\ 4k - 8 \end{bmatrix} + \begin{bmatrix} k \\ 0 \end{bmatrix} \right\} (\hat{\lambda} - k\hat{\tau}) \\ &= \hat{\theta} + \frac{a}{\Delta} \begin{bmatrix} k + (3k - 4)(1 + 4\rho) \\ (4k - 8)(1 + 4\rho) \end{bmatrix} (\hat{\lambda} - k\hat{\tau})\end{aligned}$$

We note that if $\hat{\lambda} = k\hat{\tau}$, the estimate of θ is unchanged. However, in general, for different values of a , the posterior mean for θ will lie along a line between $\hat{\theta}$ and a point on the line $\lambda = k\tau$:

$$\begin{aligned}\theta_p &= \frac{1}{\Delta} \begin{bmatrix} [\Delta - ak\{k - b(4 - 3k)\}]\hat{\tau} + a\{k - b(4 - 3k)\}\hat{\lambda} \\ 4abk(2 - k)\hat{\tau} + \{\Delta - 4ab(2 - k)\}\hat{\lambda} \end{bmatrix} \\ &= \frac{8b(b + 1)}{\Delta} \begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} + \frac{a}{\Delta} \begin{bmatrix} 1 \\ k \end{bmatrix} \{2b(2 - k)(2\hat{\tau} - \hat{\lambda}) + k(b + 1)\hat{\lambda}\}, \quad (4.8.1)\end{aligned}$$

4.9 Bayesian analysis of Balaam's design

We have seen in Chapter 3 that Balaam's design is an optimum design to compare two treatments. In this section we consider the Bayesian analysis of Balaam's design when we will restrict ourselves to n subjects in each sequence group.

From Chapter 3 for this design we obtained the estimates as

$$\begin{aligned}\hat{\theta}_2 &= \frac{1}{4} \left[\frac{\bar{y}_{21.} + \bar{y}_{22.} - \bar{y}_{11.} - \bar{y}_{12.} + \bar{y}_{31.} + \bar{y}_{32.} - \bar{y}_{41.} - \bar{y}_{42.})}{2(\bar{y}_{11.} + \bar{y}_{12.} - \bar{y}_{21.} - \bar{y}_{22.})} \right] \\ &\sim N \left\{ \begin{bmatrix} \tau \\ \lambda \end{bmatrix}, \frac{\sigma^2 + 2\sigma_s^2}{2n} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \right\}, \\ \hat{\theta}_3 &= \frac{1}{4} \left[\frac{\bar{y}_{11.} - \bar{y}_{12.} - \bar{y}_{21.} + \bar{y}_{22.} - \bar{y}_{31.} + \bar{y}_{32.} + \bar{y}_{41.} - \bar{y}_{42.})}{2(\bar{y}_{32.} - \bar{y}_{31.} - \bar{y}_{42.} + \bar{y}_{41.})} \right] \\ &\sim N \left\{ \begin{bmatrix} \tau \\ \lambda \end{bmatrix}, \frac{\sigma^2}{2n} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \right\}\end{aligned}$$

If we apply the joint prior density in (4.5.2), then the combinations of these densities using Bayes theorem can be given by combining the two estimates appropriately with the prior for (τ, λ) to obtain

$$\begin{aligned}\Sigma_p &= \left\{ \frac{2n}{\sigma^2} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} + \frac{2n}{(\sigma^2 + 2\sigma_s^2)} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + \frac{1}{\sigma_0^2} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} \right\}^{-1} \\ &= \frac{\sigma^2(1+2\rho)}{2n} \left\{ a \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} + b \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \right\}^{-1},\end{aligned}$$

where $b = 1 + 2\rho$ and $a = \sigma^2(1 + 2\rho)/(2n\sigma_0^2)$. Then

$$\Sigma_p = \frac{\sigma^2(1+2\rho)}{2n\Delta} \left\{ a \begin{bmatrix} 1 & -k \\ -k & k^2 \end{bmatrix} + b \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \right\},$$

where $\Delta = ab[(k-1)^2 + 1] + a[(k+1)^2 + 1] + (2b^2 + 6b + 1)$.

The mode of the posterior distribution when $\theta_0 = 0$ is

$$\begin{aligned}\theta_p &= \hat{\theta} - \Sigma_p \frac{1}{\sigma_0^2} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} \hat{\theta} \\ &= \hat{\theta} - \frac{a}{\Delta} \left\{ a \begin{bmatrix} 1 & k \\ k & k^2 \end{bmatrix} + b \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \right\} \begin{bmatrix} k \\ -1 \end{bmatrix} \begin{bmatrix} k & -1 \end{bmatrix} \hat{\theta} \\ &= \hat{\theta} + \frac{a}{\Delta} \left\{ b \begin{bmatrix} k-1 \\ k-2 \end{bmatrix} + \begin{bmatrix} 1+k \\ -(2+k) \end{bmatrix} \right\} (\hat{\lambda} - k\hat{\tau}) \\ &= \hat{\theta} + \frac{a}{\Delta} \begin{bmatrix} (1+k) + (k-1)(1+2\rho) \\ -(2+k) + (k-2)(1+2\rho) \end{bmatrix} (\hat{\lambda} - k\hat{\tau})\end{aligned}$$

Now if $\hat{\lambda} = k\hat{\tau}$, the estimate of θ is unchanged. But again we can show that, as a changes, the posterior mean moves along a straight line joining $(\hat{\tau}, \hat{\lambda})$ and a point $(\tilde{\tau}, \tilde{\lambda})$ on the line $\lambda = k\tau$.

$$\begin{aligned}\theta_p &= \frac{1}{\Delta} \left[[\Delta - ak\{1 + k - b(1 - k)\}]\hat{\tau} + a\{1 + k - b(1 - k)\}\hat{\lambda} \right] \\ &= \frac{2b^2 + 6b + 1}{\Delta} \begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} + \frac{1}{\Delta} \begin{bmatrix} 1 \\ k \end{bmatrix} [\{2 + k + b(2 - k)\}\hat{\tau} + a\{1 + k - b(1 - k)\}\hat{\lambda}],\end{aligned}$$

4.10 Three-period designs with two groups

From Chapter 3 for this design we obtained the estimates as

$$\begin{aligned}\hat{\theta}_2 &= \frac{1}{2} \begin{bmatrix} [(\bar{y}_{21.} + \bar{y}_{22.} + \bar{y}_{23.}) - (\bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.})] \\ 0 \end{bmatrix} \\ &\sim N \left\{ \begin{bmatrix} \tau \\ \lambda \end{bmatrix}, \frac{3N(\sigma^2 + 3\sigma_s^2)}{32n_1n_2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\} \\ \hat{\theta}_3 &= \frac{1}{8} \begin{bmatrix} 2(\bar{y}_{11.} - \bar{y}_{12.} - (\bar{y}_{13.}) - (2\bar{y}_{21.} - \bar{y}_{22.} - \bar{y}_{23.})) \\ 2(\bar{y}_{12.} - \bar{y}_{13.} - \bar{y}_{22.} + \bar{y}_{23.}) \end{bmatrix} \\ &\sim N \left\{ \begin{bmatrix} \tau \\ \tau + \lambda/2 \end{bmatrix}, \frac{N\sigma^2}{8n_1n_2} \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} \right\}\end{aligned}$$

If we apply the joint prior density in (4.5.2), then the combination of these densities using Bayes theorem can be given by combining the two estimates appropriately with the prior for (τ, λ) to obtain

$$\begin{aligned}\Sigma_p &= \left\{ \frac{8n_1n_2}{3N\sigma^2} \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} + \frac{4n_1n_2}{3N(\sigma^2 + 3\sigma_s^2)} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{\sigma_0^2} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} \right\}^{-1} \\ &= \frac{3N(1 + 3\rho)\sigma^2}{4n_1n_2} \left\{ a \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} + b \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\}^{-1}\end{aligned}$$

where $b = 2(1 + 3\rho)$ and $a = N\sigma^2(1 + 3\rho)/(4n_1n_2\sigma_0^2)$. Then

$$\Sigma_p = \frac{3N(1 + 3\rho)\sigma^2}{4n_1n_2\Delta} \left\{ a \begin{bmatrix} 1 & k \\ k & k^2 \end{bmatrix} + b \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\},$$

where $\Delta = ab(3k^2 + 4) + 12b^2 + 3b + a$. The posterior distribution is then $N(\theta_p, \Sigma_p)$, where

$$\begin{aligned}\theta_p &= \hat{\theta} - \Sigma_p \frac{1}{\sigma_0^2} \begin{bmatrix} k^2 & -k \\ -k & 1 \end{bmatrix} \hat{\theta} \\ &= \hat{\theta} - \frac{a}{\Delta} \left\{ a \begin{bmatrix} 1 & k \\ k & k^2 \end{bmatrix} + b \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \begin{bmatrix} k \\ -1 \end{bmatrix} \begin{bmatrix} k & -1 \end{bmatrix} \hat{\theta} \\ &= \hat{\theta} + \frac{a}{\Delta} \begin{bmatrix} 3k(1 + 3\rho) \\ -4(1 + 3\rho) - 1 \end{bmatrix} (\hat{\lambda} - k\hat{\tau})\end{aligned}$$

Now if $\hat{\lambda} = k\hat{\tau}$, the estimate of θ is unchanged. But again we can show that, as a changes, the posterior mean moves along a straight line joining $(\hat{\tau}, \hat{\lambda})$ and a point $(\tilde{\tau}, \tilde{\lambda})$ on the line $\lambda = k\tau$.

$$\begin{aligned}\theta_p &= \frac{1}{\Delta} \begin{bmatrix} (\Delta - 3abk^2)\hat{\tau} + 3abk\hat{\lambda} \\ ak(4b + 1)\hat{\tau} + \{\Delta - a(4b + 1)\}\hat{\lambda} \end{bmatrix} \\ &= \frac{3b(4b + 1)}{\Delta} \begin{bmatrix} \hat{\tau} \\ \hat{\lambda} \end{bmatrix} + \frac{1}{\Delta} \begin{bmatrix} 1 \\ k \end{bmatrix} \{a(4b + 1)k\tau + 3abk\hat{\lambda}\}, \quad (4.10.1)\end{aligned}$$

In general, in all these examples, the case $\sigma_0^2 = 0$ will produce an estimate satisfying $\lambda = k\tau$ by moving along a straight line from $(\hat{\tau}, \hat{\lambda})$ to the line $\lambda = k\tau$. If $\sigma_0^2 \neq 0$, then the posterior mean is only part of the way along that straight line, as shown in Figure 4.2.

Chapter 5

Analysis of a two-period crossover design for the comparison of two active treatments and placebo

5.1 Introduction

This chapter considers the analysis to the design in Koch et al. (1989) involving three treatments in two periods. Those authors were considering a situation involving a chronic health disorder, in which two treatments were active agents (labelled A and B), with the third treatment being a placebo, labelled P . In their paper, the authors supposed that there were m subjects in each of the groups 1 and 2, which received the active treatments in the order AB and BA respectively, for every one subject in each of the groups 3,4,5, and 6 which received the treatments in the order AP , PA , BP and PB respectively. In this chapter, we will suppose that $m = 1$.

The analysis given by Koch et al. (1989) considers several models which have either no carryover effects, equal carryover effects for the two active treatments, or distinct carryover effects for the two treatments. We note particularly that they only consider between-subject estimates and, as we show in this chapter, there is considerable advantage to be gained by recovering the between-subject information.

This design is considered further by Laird et al. (1992) who consider two-period crossover designs in general and who argue for the combination of the two components of

the variation. While they consider the variances of the estimates in the different strata, they do not look closely at the covariance matrices of the estimates in the different strata, nor do they present either the combined estimates or the covariance matrix of the combined estimates.

5.2 The linear model

Following the method in Chapter 2, we assume that $\mathbf{Y}_{ijk}$ is a random variable with the observed value y_{ijk} which follows the mixed linear model given there. The layout is shown in Table 5.1. We shall suppose that the groups each have n subjects, thus there $N = 6n$ subjects.

Group	Period	
	1	2
1	<i>A</i>	<i>B</i>
2	<i>B</i>	<i>A</i>
3	<i>A</i>	<i>P</i>
4	<i>P</i>	<i>A</i>
5	<i>B</i>	<i>P</i>
6	<i>P</i>	<i>B</i>

Table 5.1: Layout of design with two active treatments and placebo

Following the model in Section 2.3 of Koch et al. (1989), we shall use a model in which

- there is no residual effect for the placebo,
- the direct effects of the placebo, *A* and *B* are respectively, 0, τ_1 and τ_2 ,
- the residual effect of placebo, *A* and *B* are respectively, 0, λ_1 and λ_2 .

We shall see shortly that the information in the design separates conveniently into a component corresponding to the average treatment effect for *A* and *B* and another corresponding to the difference between the effects of *A* and *B*. Accordingly, we define:

- $\tau_0 = (\tau_1 + \tau_2)/2$ and $\lambda_0 = (\lambda_1 + \lambda_2)/2$ as the average effects of treatment and carryover respectively, and
- $\tau_D = (\tau_1 - \tau_2)/2$ and $\lambda_D = (\lambda_1 - \lambda_2)/2$ as the half difference between direct treatment and carryover effects of the two treatments respectively.

From the twelve cells which are the combination of six groups and two periods, the expected values corresponding to the cells in Table 5.1 are as shown in Table 5.2.

	Period	
	1	2
Group	$\mu + \pi_1 + \tau_0 + \tau_D$	$\mu + \pi_2 + \tau_0 + \lambda_0 - \tau_D + \lambda_D$
	$\mu + \pi_1 + \tau_0 - \tau_D$	$\mu + \pi_2 + \tau_0 + \lambda_0 + \tau_D - \lambda_D$
	$\mu + \pi_1 + \tau_0 + \tau_D$	$\mu + \pi_2 + \lambda_0 + \lambda_D$
	$\mu + \pi_1$	$\mu + \pi_2 + \tau_0 + \tau_D$
	$\mu + \pi_1 + \tau_0 - \tau_D$	$\mu + \pi_2 + \lambda_0 - \lambda_D$
	$\mu + \pi_1$	$\mu + \pi_2 + \tau_0 - \tau_D$

Table 5.2: Expected values of responses in comparison of two active treatments and placebo.

These equations lead to the following form for the Group $\times$ Period means

$$\begin{bmatrix} \bar{y}_{11.} \\ \bar{y}_{12.} \\ \bar{y}_{21.} \\ \bar{y}_{22.} \\ \bar{y}_{31.} \\ \bar{y}_{32.} \\ \bar{y}_{41.} \\ \bar{y}_{42.} \\ \bar{y}_{51.} \\ \bar{y}_{52.} \\ \bar{y}_{61.} \\ \bar{y}_{62.} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \beta^* + \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & -1 & 1 \\ 1 & 0 & -1 & 0 \\ 1 & 1 & 1 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} \tau_0 \\ \lambda_0 \\ \tau_D \\ \lambda_D \end{bmatrix} + \epsilon^*, \quad (5.2.1)$$

or

$$\bar{y} = \mu 1_{12} + (1_6 \otimes I_2)\pi + (I_6 \otimes 1_2)\beta^* + T^*\theta + \epsilon^*, \quad (5.2.2)$$

where $\theta = \begin{bmatrix} \tau_0 & \lambda_0 & \tau_D & \lambda_D \end{bmatrix}'$. The covariance matrix of $\bar{y}$ can be obtained by

$$Var(\bar{y}) = \frac{\sigma^2}{n}(J_6 \otimes K_2) + \frac{\sigma^2 + 2\sigma_s^2}{n}(J_6 \otimes J_2) \quad (5.2.3)$$

We want to examine the treatment information and obtain treatment estimates by looking at the vector of means $\bar{y}$. By referring to Chapter 2 and incorporating random subject effects, we can get the information matrices about θ in the four strata which we introduced in Chapter 2. Table 5.3 shows the relevant projector matrices for the four strata Grand Mean, Periods, Groups and Group $\times$ Period.

Stratum	Matrix	
Grand Mean	Q_0^*	$J_6 \otimes J_2$
Period	Q_1^*	$J_6 \otimes K_2$
Group	Q_2^*	$K_6 \otimes J_2$
Group $\times$ Period	Q_3^*	$K_6 \otimes K_2$

Table 5.3: Idempotent matrices for four strata in the design including placebo.

5.3 Treatment information

To get the information about treatment parameters as before we consider all strata except the Grand Mean stratum. The treatment information on the four parameters is then identified as:

$$T^{*'}Q_1^*T^* = \frac{n}{3} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$T^{*'}Q_2^*T^* = \frac{n}{3} \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 6 & 3 \\ 0 & 0 & 3 & 6 \end{bmatrix},$$

$$T^{*'}Q_3^*T^* = \frac{n}{3} \begin{bmatrix} 6 & -3 & 0 & 0 \\ -3 & 2 & 0 & 0 \\ 0 & 0 & 18 & -9 \\ 0 & 0 & -9 & 6 \end{bmatrix}.$$

The Period stratum has some information about the parameter λ_0 , corresponding to half of the available information, but, since it is estimated by the difference between the two period means, and we suspect that there may be period differences anyway, we do not attempt to recover this information, concentrating instead on the information in the last two strata. We see that 75% of the information about the treatment parameters is in the Group $\times$ Period stratum, and 25% in the Groups stratum, while the information available, i.e. excluding that lost to the Period stratum, about the carryover parameters is split equally between these two strata. We note that Koch et al. (1989) only consider the within-subject estimates available from the Group $\times$ Period stratum.

Now the between-subject and within-subject estimates and their covariance matrices of parameters of interest respectively are given by

$$\hat{\theta}_2 = (T^{*'}Q_2^*T^*)^{-1}T^{*'}Q_2^*\bar{y}$$

$$= \begin{bmatrix} \bar{y}_{1..} + \bar{y}_{2..} - \bar{y}_{3..} - \bar{y}_{5..} \\ \bar{y}_{3..} - \bar{y}_{4..} + \bar{y}_{5..} - \bar{y}_{6..} \\ (-\bar{y}_{1..} + \bar{y}_{2..} + \bar{y}_{3..} + 2\bar{y}_{4..} - \bar{y}_{5..} - 2\bar{y}_{6..})/3 \\ (2\bar{y}_{1..} - 2\bar{y}_{2..} + \bar{y}_{3..} - \bar{y}_{4..} - \bar{y}_{5..} + \bar{y}_{6..})/3 \end{bmatrix}$$

$$\hat{\theta}_3 = (T^{*'}Q_3^*T^*)^{-1}T^{*'}Q_3^*\bar{y}$$

$$= \begin{bmatrix} (-\bar{y}_{11.} + \bar{y}_{12.} - \bar{y}_{21.} + \bar{y}_{22.} + \bar{y}_{31.} - \bar{y}_{32.} + \bar{y}_{51.} - \bar{y}_{52.})/2 \\ [2(-\bar{y}_{11.} + \bar{y}_{12.} - \bar{y}_{21.} + \bar{y}_{22.}) + \bar{y}_{31.} - \bar{y}_{32.} + \bar{y}_{41.} - \bar{y}_{42.} + \bar{y}_{51.} - \bar{y}_{52.} + \bar{y}_{61.} - \bar{y}_{62.})] / 4 \\ [(\bar{y}_{11.} - \bar{y}_{12.} - \bar{y}_{21.} + \bar{y}_{22.} - \bar{y}_{31.} + \bar{y}_{32.} + \bar{y}_{51.} - \bar{y}_{52.}) - 2(\bar{y}_{41.} - \bar{y}_{42.} - \bar{y}_{61.} + \bar{y}_{62.})] / 4 \\ (-\bar{y}_{31.} + \bar{y}_{32.} - \bar{y}_{41.} + \bar{y}_{42.} + \bar{y}_{51.} - \bar{y}_{52.} + \bar{y}_{61.} - \bar{y}_{62.})/2 \end{bmatrix}$$

$$Var(\hat{\theta}_2) = \frac{\sigma^2(1+2\rho)}{3n} \begin{bmatrix} 6 & -3 & 0 & 0 \\ -3 & 2 & 0 & 0 \\ 0 & 0 & 18 & -9 \\ 0 & 0 & -9 & 6 \end{bmatrix}$$

$$Var(\hat{\theta}_3) = \frac{\sigma^2}{3n} \begin{bmatrix} 6 & 9 & 0 & 0 \\ 9 & 18 & 0 & 0 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 3 & 6 \end{bmatrix}$$

The variance matrices for the parameter estimates reveal some important properties. The within-subject estimates given by $\hat{\theta}_3$ show very high positive correlation between the estimates of τ_0 and λ_0 , and also between the estimates of τ_D and λ_D . We note that the corresponding correlations in the between-subject estimates are positive, suggesting that the combination of estimates across the two strata may be advantageous not just for bringing together the information in the two strata, but to some extent cancelling out the correlations between the estimates.

5.4 Combination of the estimates

The between and within subject estimates may then be combined by the appropriate weighting as

$$\hat{\theta} = \{T^{*'}Q_2^*T^* + (1 + 2\rho)T^{*'}Q_3^*T^*\}^{-1}\{T^{*'}Q_2^*T^*\hat{\theta}_2 + (1 + 2\rho)T^{*'}Q_3^*T^*\hat{\theta}_3\}$$

The estimates of the four parameters may then be given, using the same layout as Table 5.1, by:

$$\hat{\tau}_0 = \frac{1}{2(7 + 14\rho + 3\rho^2)} \begin{array}{|c|c|} \hline 2 + \rho - 3\rho^2 & 3(1 + \rho)^2 \\ \hline 2 + \rho - 3\rho^2 & 3(1 + \rho)^2 \\ \hline 2 + 7\rho + 3\rho^2 & -3(1 + \rho)^2 \\ \hline -4(1 + 2\rho) & 0 \\ \hline 2 + 7\rho + 3\rho^2 & -3(1 + \rho)^2 \\ \hline -4(1 + 2\rho) & 0 \\ \hline \end{array}$$

$$\begin{aligned}
\hat{\lambda}_0 &= \frac{1}{2(7 + 14\rho + 3\rho^2)} \begin{array}{|c|c|} \hline 1 - \rho - 6\rho^2 & 5 + 13\rho + 6\rho^2 \\ \hline 1 - \rho - 6\rho^2 & 5 + 13\rho + 6\rho^2 \\ \hline 1 + 2\rho + 3\rho^2 & 2 + \rho - 3\rho^2 \\ \hline -(2 + \rho - 3\rho^2) & -(7 + 14\rho + 3\rho^2) \\ \hline 1 + 2\rho + 3\rho^2 & 2 + \rho - 3\rho^2 \\ \hline -(2 + \rho - 3\rho^2) & -(7 + 14\rho + 3\rho^2) \\ \hline \end{array} \\
\hat{\tau}_D &= \frac{1}{2(7 + 14\rho + 3\rho^2)} \begin{array}{|c|c|} \hline 2 + 5\rho + \rho^2 & -(1 + \rho)^2 \\ \hline -(2 + 5\rho + \rho^2) & (1 + \rho)^2 \\ \hline 2 + 3\rho - \rho^2 & -(1 + \rho)^2 \\ \hline -2(1 + \rho) & -2(1 + \rho)^2 \\ \hline -(2 + 3\rho - \rho^2) & (1 + \rho)^2 \\ \hline 2(1 + \rho) & (1 + \rho)^2 \\ \hline \end{array} \\
\hat{\lambda}_D &= \frac{1}{2(7 + 14\rho + 3\rho^2)} \begin{array}{|c|c|} \hline 1 + \rho & 3 + 5\rho \\ \hline -(1 + \rho) & -(3 + 5\rho) \\ \hline 1 - 3\rho^2 & 4 + 9\rho + 3\rho^2 \\ \hline -\rho(1 + 3\rho) & 1 + 4\rho + 3\rho^2 \\ \hline -(1 - 3\rho^2) & -(4 + 9\rho + 3\rho^2) \\ \hline \rho(1 + 3\rho) & -(1 + 4\rho + 3\rho^2) \\ \hline \end{array}
\end{aligned}$$

The variance matrix for these combined estimates $\hat{\theta}$ is

$$\begin{aligned}
Var(\hat{\theta}) &= \left(\frac{1}{\sigma^2 + 2\sigma_s^2} T^{*'} Q_2^* T^* + \frac{1}{\sigma^2} T^{*'} Q_3^* T^* \right)^{-1} \\
&= \frac{\sigma^2(1 + 2\rho)}{2n(7 + 14\rho + 3\rho^2)} \begin{bmatrix} 6(1 + \rho) & 3(1 + 3\rho) & 0 & 0 \\ 3(1 + 3\rho) & 6(2 + 3\rho) & 0 & 0 \\ 0 & 0 & 2(1 + \rho) & 1 + 3\rho \\ 0 & 0 & 1 + 3\rho & 2(2 + 3\rho) \end{bmatrix},
\end{aligned}$$

5.5 Conclusions

The estimates obtained in this way vary according to the value of $\rho = \sigma_s^2/\sigma^2$, which needs to be estimated from the data. Even with only 5 subjects per group, the within and between subject variances are estimated on 24 degrees of freedom and will give a stability to the estimate of ρ which will provide estimates largely unaffected by the errors

in that estimate. The estimates of $\sigma^2 + 2\sigma_S^2$ and σ^2 are the mean squares W_2 and W_1 , the Subjects within Groups and the Period $\times$ Subjects within Groups mean squares, respectively. The estimate of ρ is then given by equating the observed variance ratio to its expectation, so that

$$\hat{\rho} = (\frac{W_2}{W_1} - E_F)/(2E_F),$$

where $E_F = 24/22$ is the expectation of an F -distribution with (24,24) degrees of freedom. The standard deviation of $\hat{\rho}$ then follows from the variance of the same F -distribution and is given by

$$SD(\hat{\rho}) = \sqrt{\frac{46}{2 \times 24 \times 20}} = 0.22$$

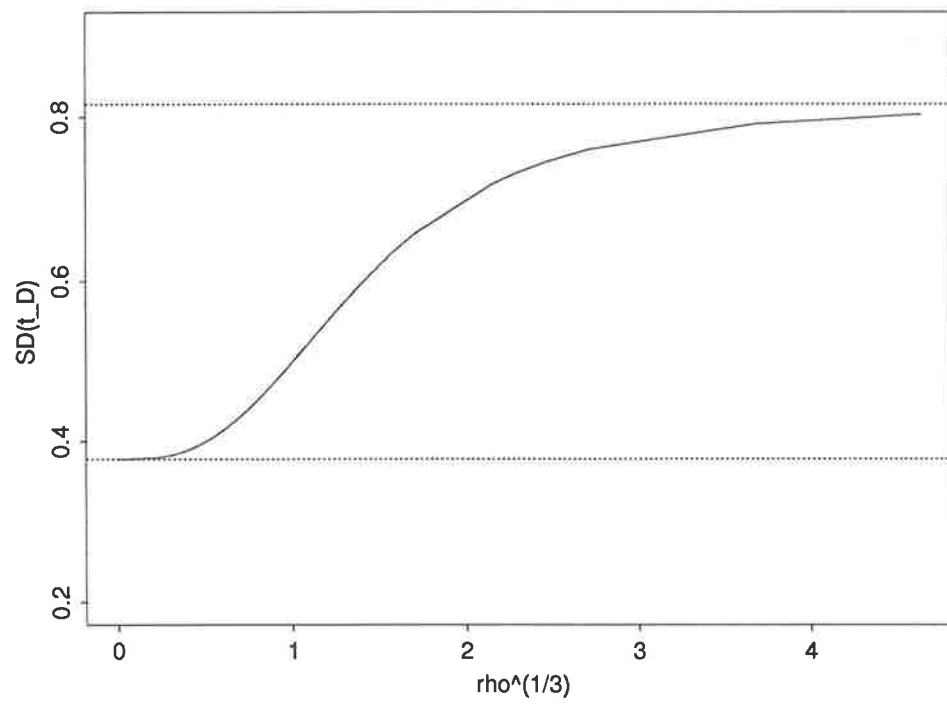
(Johnson & Kotz, 1970, Chapter 26). Except when the estimate of ρ is close to zero, changes in ρ of the order of one standard deviation will have a small impact on both the form and the standard deviations of the estimates.

As ρ changes, the estimates move between the estimates at $\rho = 0$ which ignore the grouping and treat all observations as independent, and the estimates at $\rho = \infty$ which regards the subject differences as so large that only the within-subject estimates are useful. In the latter case, it can be seen that the estimates revert to $\hat{\theta}_3$; while in the former case they are not simply the average of $\hat{\theta}_2$ and $\hat{\theta}_3$ due to the fact that the variance matrices are not proportional. For example, the estimates of τ_0 and λ_0 in the case $\rho = 0$ may be written as

$$\begin{bmatrix} \hat{\tau}_0 \\ \hat{\lambda}_0 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 5\hat{\tau}_{02} + 9\hat{\tau}_{03} + 4(\hat{\lambda}_{02} - \hat{\lambda}_{03}) \\ 9\hat{\lambda}_{02} + 5\hat{\lambda}_{03} + 6(\hat{\tau}_{02} - \hat{\tau}_{03}) \end{bmatrix}.$$

Figure 5.1 shows a plot of the standard deviation of $\hat{\tau}_D$ as ρ changes. It ignores the factor σ^2/n introduced by having n subjects in each group. The horizontal axis here is $\rho^{1/3}$ in order to show the change more clearly. The upper limit of $\sqrt{2/3} = 0.816$ is the standard deviation achieved using the within-subject estimates, which are the estimates used as $\rho \rightarrow \infty$, while the lower limit of $\sqrt{1/7} = 0.378$ is achieved if there is no between-subject variance and $\rho = 0$.

Figure 5.1: Standard deviation of $\hat{\tau}_D$ as ρ changes



Chapter 6

Cohort designs for two-treatment crossover trials with one baseline measurement

6.1 Introduction

In previous chapters, we showed that the average effect of treatment against baseline was not estimable unless strong assumptions were made about the absence of period effects, or the existence of a random effect for periods with known variance. In this chapter we apply an alternative design which under some conditions enables us to estimate the average effect of treatment against baseline. This alternative design is like an age-period-cohort design, in that it supposes that periods correspond to the times of the year and that some groups of subjects, referred to here as cohorts, enter the trial with a delayed start. In particular, we suppose that each successive cohort is delayed by exactly one period. Provided we can assume that period effects are related to the time of year rather than the length of time in the trial, referred to here as the ‘age’, this design will allow the estimation of the average effect of treatment against baseline. Section 6.2 summarises Age-Period-Cohort designs as they currently exist in the literature and illustrates the association between these and our proposed ‘cohort designs’. In Section 6.3 we build the linear model for both the standard design used earlier and the new ‘cohort design’. The standard design with one baseline measurement taken before the first period treatment is analysed in Section 6.4 and the corresponding cohort design with just two cohorts is

considered in Section 6.5. In Section 6.6 we discuss the treatment information for both designs and in Section 6.7 we combine the treatment information in each stratum to get estimates of the parameters of interest for the two designs. The limiting cases of $\rho = \frac{\sigma_3^2}{\sigma^2} \rightarrow \infty$ and zero are given in Section 6.8 and finally in Section 6.9 we make some general conclusions about using cohort designs rather than standard designs.

6.2 Age-period-cohort design

The aim of this section is to give a general overview of age-period-cohort designs. Up to now there are a lot of publications in which they attempted to show all features of this design and to explain all sorts of things on the basis of three variables, namely: age, period and cohort which occupy a central position in cohort design. Based on our literature review on some articles in this regard we can claim that the precise meaning of the variables age, period and cohort, and thus also what is meant by age, period and cohort effects, varies considerably from one study to another.

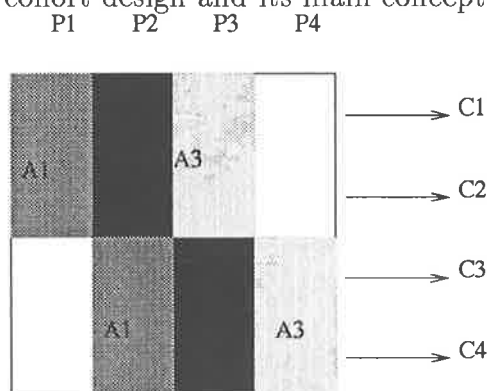
Kupper et al. (1983) discuss the specific problem of age-period-cohort analysis within the general framework and give the following definition for age-period-cohort design: “age-period-cohort analysis concerns methods for statistically analysing of such data gathered on human population followed over time, the purpose of such analysis being to quantify accurately patterns in the separate effects of age, period of time (e.g., as measured by the year of occurrence), and cohort membership (e.g., as measured by year of birth).” Hagenaaars (1978) in his paper gives the following definitions for these three variables. He says that “in general it may be said that age is regarded as an indicator for all the possible changes which are related to becoming older. Age can for example be taken as an indicator for phenomena which are mainly of a biological nature, for example becoming sexually mature, becoming tired more easily, getting ill more often.” For definition of period he says that “Period refers to all events which have taken place at or between the moments of observation and which have influenced the phenomenon being studies. Here too it will often be difficult to indicate exactly which event is responsible for changes observed.” Finally he defines cohort as “A collection of people born within the same period or a collection of people who have experienced a fundamental event during the same period.”

Therefore in our study we would like to draw attention to the suitability of the cohort design for the analysis of crossover trials with the following definitions for the age, period and cohort variables:

- (i) Age can be defined as the length of time in trial and geometrically can be represented as each diagonal in Figure 6.1,
- (ii) Period in this situation is the same as defined generally for each crossover trial in which experimenters take observations separately on that duration of time and is represented by the columns in Figure 6.1. Period may be of short duration (e.g. hours) or long duration (e.g. months).
- (iii) Cohort is a group of subjects which start together in the trial at the same time is represented by the rows in Figure 6.1.

To clearly show this we represent it schematically in Figure 6.1 emphasising the definitions of age(A), period(P) and cohort(C) for our so-called cohort design.

Figure 6.1: Diagram of cohort design and its main concepts in crossover trial



The main and interaction effects of age, period, and cohort are not all estimable. This difficulty, usually referred to as the identification problem, is caused by the fact that a certain fixed, direct relationship exists between age, cohort and period. We can express algebraically that age (A), is the difference between the moment(P) of observation and the start time (C), that is $A = P - C$. In this framework, the rows, C are orthogonal to the diagonals, A . Hence, the P can be thought of as part of the $A \times C$ interaction. Thus, fitting the linear effects for C , the linear trend for A and the linear trend for P provide 3 contrasts, only two of which are estimable.

This identification problem can perhaps be shown even more clearly by means of a more formal demonstration. To show this we refer to the general development model of

Schaie (1965). He gives an additive model in which each observation is seen as a weighted sum of the linear effects of the three independent variables

$$Y = b_a A + b_c C + b_p P.$$

Since $A = P - C$ it is also true that

$$Y = b_a(P - C) + b_c C + b_p P = (b_a + b_p)P + (b_c - b_a)C,$$

this can be reformulated as

$$Y = q_p P + q_c C.$$

In other words the additive model with three linear terms can be reduced to a model with two linear terms which predicts exactly the same Y value as the three variable model. Only two independent effect parameters are found to exist, that is, q_p and q_c , from which it is no longer possible to deduce the three original coefficients b_a, b_c, b_p .

In our situation, we will be assuming that ‘age’ effects are not present, but that ‘period’ effects may be. This will only be appropriate in particular cases, such as asthma trials where a subject’s condition may be related quite strongly to the time of year in which a treatment is applied.

The potential advantage of the baseline measurement in crossover trials is that it enables estimation of the absolute effect of direct treatment by comparison with the baseline. This advantage is lost if there are period effects present. The use of the cohort design as described here is that the absolute effect of treatment may again be estimable, even when period effects are present, provided age effects, as described, are not present. For this to be successful, we require the following :

- (i) Subjects enter according to a protocol so that no essential differences exist between those subjects entering as the first or second cohort.
- (ii) Length of time or experience of being in the trial does not influence the results. This might not be satisfied, for example, if blood pressure is related to anxiety and the first cohort is less anxious in period two than the second cohort. Different blood pressure between first and second cohorts at period two may then be partly due to anxiety level, and hence related to what we have called here ‘age’ or time in trial.

The potential advantages of this design in crossover design are:

- it is useful when we suspect period effects (e.g. time of year), and

- we can estimate absolute effects of treatment and carryover.

6.3 Building the model for standard and cohort designs

To show the efficiency of the cohort design we compare it to the standard design of crossover trial with baseline shown in Table 6.1. Here we use two groups of subjects, where each *row* represents a group of subjects, and each *column* represents one period. In this situation, periods refer to the time a subject has been in the trial and to calendar time.

		Period		
		1	2	3
Group	1	—	A	B
	2	—	B	A

Table 6.1: Two-treatment three-period crossover design with baseline measurements

Suppose, however, that subjects enter the trial at different times, as the trial progresses. It may be that we anticipate an effect related to the time of year in which the measurement is made, but that we do not expect an effect related to the length of time a subject has been in the trial.

To illustrate the idea, suppose that a study is undertaken as above, but with half the subjects having a delayed entry at period 2 as shown in Table 6.2. If there are $4n$ subjects, this design would appear as shown in Table 6.2, with $2n$ subjects in each cohort, and the subjects in each cohort being randomly allocated n to each of the two treatment regimes. We note, for example, that a direct between-group comparison of treatment with baseline is available in period 2 under certain conditions which we will explore in the next section.

		Period			
		1	2	3	4
Group	1	—	A	B	
	2	—	B	A	
	3		—	A	B
	4		—	B	A

Table 6.2: Cohort design in two-treatment three-period crossover design with one baseline measurement

In order to compare this with the standard design, it is convenient to present the standard design in a similar way, with 4 groups of n subjects, as shown in Table 6.3.

		Period		
		1	2	3
Group	1	—	A	B
	2	—	B	A
	3	—	A	B
	4	—	B	A

Table 6.3: Two-treatment three-period crossover design with one baseline measurement repeated as 4 groups of n .

Our purpose in this chapter is to compare the cohort design in Table 6.2 with the corresponding standard design in Table 6.3. We assume that subject effects are random; and that we have fixed effects for the period and mean μ . The treatment parameters are defined as:

- τ_0 , the average effect of treatment; i.e. $\frac{A+B}{2}$ minus baseline;
- λ_0 , the average effect of carryover; i.e. $\frac{A+B}{2}$, for carryover effects, minus baseline;
- $2\tau_D$, the difference between A and B ; i.e. τ_D for A and $-\tau_D$ for B ;
- $2\lambda_D$, the difference in carryover between A and B ; i.e. λ_D for carryover of A and $-\lambda_D$ for carryover of B .

To consider what treatment information is available, we can write out as in Table 6.4 the 12 means in terms of the treatment effects in the following table which could be applied to either the standard or the cohort design.

μ	$\mu + \tau_0 + \tau_D$	$\mu + \tau_0 + \lambda_0 - \tau_D + \lambda_D$
μ	$\mu + \tau_0 - \tau_D$	$\mu + \tau_0 + \lambda_0 + \tau_D - \lambda_D$
μ	$\mu + \tau_0 + \tau_D$	$\mu + \tau_0 + \lambda_0 - \tau_D + \lambda_D$
μ	$\mu + \tau_0 - \tau_D$	$\mu + \tau_0 + \lambda_0 + \tau_D - \lambda_D$

Table 6.4: Treatment effects for the 12 means of observations for two-treatment three-period crossover design

If we write $\bar{y} = (\bar{y}_{11}, \dots, \bar{y}_{43})'$, corresponding to the 12 means written in order with periods changing quickest, then groups, and cohorts changing slowest, then we have

$$\bar{y} = T_0 \theta_0 + \epsilon,$$

where

$$T_0 = 1_2 \otimes \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & -1 & 0 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 \end{bmatrix} \quad (6.3.1)$$

and $\theta_0 = [\mu, \tau_0, \lambda_0, \tau_D, \lambda_D]'$, and ϵ represents the random components, to be defined. If $\epsilon \sim N(0, Q/n)$, then the overall information matrix is

$$n T_0' Q^{-1} T_0.$$

In many cases, Q can be expressed as $\sum_{i=0}^k \delta_i Q_i$, where the Q_i are orthogonal idempotent matrices, such that $\sum_{i=0}^k Q_i = I$. Then the total information matrix will be

$$n \sum_{i=0}^k \delta_i^{-1} T_0' Q_i T_0, \quad (6.3.2)$$

where the δ_i are estimated from data. For some values of i , for example for the grand mean $Q_0 = J$, the term may be omitted if we feel that it contributes no useful information

on the parameters of interest. Since $\sum Q_i = I$, it is useful to consider

$$n \sum_{i=0}^3 T_0' Q_i T_0 = n T_0' T_0 = 2n \begin{bmatrix} 6 & 4 & 2 & 0 & 0 \\ 4 & 4 & 2 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 4 & -2 \\ 0 & 0 & 0 & -2 & 2 \end{bmatrix},$$

as the overall information matrix which will be partitioned into components and then recombined with different weights as in (6.3.2) in the final analysis.

Information about the parameters of interest is only contained in the space orthogonal to the grand mean. Thus, while $T_0' T_0$ gives the overall information matrix, the term $T_0' J T_0$, corresponding to the grand mean, represents information which is not available to us since it just represents an average level of response. The information available about the parameters $\theta = (\tau_0, \lambda_0, \tau_D, \lambda_D)'$ of interest is obtained by using the component of T_0 which is orthogonal to the mean vector $\mu 1$. In particular, this reduces our matrix T_0 to the matrix

$$K T_0 = \frac{1}{6} 1_2 \otimes \begin{bmatrix} 0 & -4 & -2 & 0 & 0 \\ 0 & -4 & -2 & 0 & 0 \\ 0 & 2 & -2 & 6 & 0 \\ 0 & 2 & -2 & -6 & 0 \\ 0 & 2 & 4 & -6 & 6 \\ 0 & 2 & 4 & 6 & -6 \end{bmatrix},$$

from which we can discard the first column of zeros to obtain

$$T = \frac{1}{6} 1_2 \otimes \begin{bmatrix} -4 & -2 & 0 & 0 \\ -4 & -2 & 0 & 0 \\ 2 & -2 & 6 & 0 \\ 2 & -2 & -6 & 0 \\ 2 & 4 & -6 & 6 \\ 2 & 4 & 6 & -6 \end{bmatrix}. \quad (6.3.3)$$

Hence the available information about $\theta = (\tau_0, \lambda_0, \tau_D, \lambda_D)'$ is contained in the information matrix

$$\mathcal{I} = nT'T = \frac{n}{3} \begin{bmatrix} 8 & 4 & 0 & 0 \\ 4 & 8 & 0 & 0 \\ 0 & 0 & 24 & -12 \\ 0 & 0 & -12 & 12 \end{bmatrix}. \quad (6.3.4)$$

This is the same for both standard and cohort designs. We want to look at $T'Q_iT$, the information contained in the i th stratum. This will be different for the standard and cohort design.

However, we want to see where the treatment information goes, and how much is lost when we take out periods. We could do that by putting a term for periods in the model and comparing joint information matrices for θ with and without periods.

An alternative is to consider the projection matrix which takes out period effects and see which part of $\mathcal{I}$ ends up in that subspace. Thus, we are seeing Q as not just a variance matrix, but rather as a way of splitting up $\bar{y} \in R^{12}$ into separate components based on the strata in the experiment, to see how much treatment information is lost into each subspace. Later on we can make the choice as to which components we recombine to get treatment estimates.

6.4 Standard design

We first look at describing the treatments effects for the 12 means contained in both standard and cohort design defined in Section 6.3.

6.4.1 Information matrices and parameter estimates

In the crossover design with baseline measurements described in Section 6.3 we assume that we have means with expected values as shown in Table 6.4.

$$(I - J)\bar{y} = T\theta + \epsilon, \quad (6.4.1)$$

where $\epsilon \sim N(0, Q/n)$ and

$$Q = \sum_{i=1}^3 \delta_i Q_i \quad (6.4.2)$$

where $Q_1 = J \otimes K \otimes J$ represents the projector matrix for Periods, $Q_2 = (I \otimes J \otimes I - J \otimes J \otimes J)$ represents the projector matrix for Groups, and $Q_3 = (I \otimes K \otimes I - J \otimes K \otimes J)$ represents the projector matrix for Periods $\times$ Groups, and the matrices in the Kronecker products are of dimension 2, 3 and 2, corresponding to Cohorts, Periods, and Groups within Cohorts, respectively. The coefficients δ_i in (6.4.2) are $\delta_2 = \sigma^2 + 3\sigma_s^2$, $\delta_1 = \delta_3 = \sigma^2$, although δ_1 would also include a component relating to the variation between Periods. Since it is unlikely that we would regard the periods as a random rather than a fixed effect, we would not use δ_1 in our recovery of treatment information.

We can think of the information on θ as coming from the three different strata. Thus, since $\sum_{i=1}^3 Q_i = (I - J)$, and since $T'JT = 0$,

$$T'T = \sum_{i=1}^3 T'Q_iT.$$

Hence, the information matrix for the i th stratum is $T'Q_iT$. These will be given in the first column of Table 6.8.

The estimate of θ from the i th stratum is

$$\hat{\theta}_i = (T'Q_iT)^{-1}T'Q_i\bar{y}.$$

and these can be combined using

$$\hat{\theta} = (\sum_i \delta_i^{-1} T'Q_iT)^{-1} (\sum_i \delta_i^{-1} T'Q_iT \hat{\theta}_i), \quad (6.4.3)$$

provided suitable estimates of δ_i are available. The cases $i = 0, 1$ correspond to the grand mean and periods, respectively, and we will not attempt to recover information from these strata. Hence the summation in (6.4.3) will be for $i = 2, 3$ only.

6.4.2 Analysis of variance for standard design

For the standard design from equations (6.4.1) and (6.4.2), we can write out the table of the analysis of variance for the standard design. We can write

$$nE(\bar{y}\bar{y}') = \{Q + nT\theta\theta'T'\},$$

where we have ignored terms involving the periods, from which we can get the expected sum of squares for each stratum as

$$\begin{aligned} nE\{\bar{y}'Q_i\bar{y}\} &= \text{trace}\{Q_iQ + n\theta'T'Q_iT\theta\}, \\ &= \delta_i \text{trace}(Q_i) + n\theta'T'Q_iT\theta \end{aligned}$$

Since $\text{trace}(Q_i) = \text{rank}(Q_i)$, then the expected mean squares are

$$\text{EMS}_i = \delta_i + n\theta'T'Q_iT\theta/\text{rank}(Q_i),$$

as shown in Table 6.5.

Source	DF	SS	EMS
Grand Mean	1	$n\bar{y}'(J \otimes J \otimes J)\bar{y}$	
Periods	2	$n\bar{y}'Q_1\bar{y}$	
Groups	3	$n\bar{y}'Q_2\bar{y}$	$\sigma^2 + 3\sigma_s^2 + n\theta'T'Q_2T\theta/3$
Period $\times$ Group	6	$n\bar{y}'Q_3\bar{y}$	$\sigma^2 + n\theta'T'Q_3T\theta/6$
Subject within Groups	$4(n-1)$		$\sigma^2 + 3\sigma_s^2$
(Subject within Group) $\times$ Period	$8(n-1)$		σ^2
Total	$12n$		

Table 6.5: Analysis of variance for standard design

6.5 Cohort design

In the cohort design we have the same model (6.4.1) for the 12 means, and the matrix Q still has the same covariance structure in terms of the between and within subject effects. However, the projector matrix for Periods is now changed, because the 12 means are now assigned to five different periods, rather than the original three. Furthermore, Groups and Periods are not orthogonal. This implies that we cannot just use the Groups stratum as it stands because it has some period effects in it. We can explore the different strata by considering:

- Periods (ignoring Groups), with 3 degrees of freedom,
- Groups(ignoring Periods), with 3 degrees of freedom, and
- the Group $\times$ Period interaction, with 5 degrees of freedom.

The first two of these can be examined by considering the following contrasts. Contrasts on Groups can be expressed as the row contrasts g_1 , g_2 and g_3 :

-1	-1	-1	
-1	-1	-1	
	1	1	1
	1	1	1

g_1

-1	-1	-1	
1	1	1	
	-1	-1	-1
	1	1	1

g_2

-1	-1	-1	
1	1	1	
	1	1	1
	-1	-1	-1

g_3

from which a projector matrix for Groups (ignoring Periods) can be formed as

$$Q_G = \sum_{i=1}^3 g_i g_i' / (g_i' g_i). \quad (6.5.1)$$

Similarly, contrasts for Periods are the column contrasts p_1 , p_2 and p_3 :

0	-1	1	
0	-1	1	
	-1	1	0
	-1	1	0

p_1

-2	1	1	
-2	1	1	
	1	1	-2
	1	1	-2

p_2

-1	0	0	
-1	0	0	
	0	0	1
	0	0	1

p_3

and the projector matrix for Periods (ignoring Groups) can be given as

$$Q_P = \sum_{i=1}^3 p_i p_i' / (p_i' p_i). \quad (6.5.2)$$

Due to the fact that groups and periods here are not orthogonal, $Q_G Q_P \neq 0$. We can identify the relationships between the six contrasts by the following matrix of cross-products. If we let $W = (g_1, g_2, g_3, p_1, p_2, p_3)$, then

$$W'W = \begin{bmatrix} 12 & 0 & 0 & 0 & 0 & 4 \\ 0 & 12 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 24 & 0 \\ 4 & 0 & 0 & 0 & 0 & 4 \end{bmatrix} \quad (6.5.3)$$

which shows the lack of orthogonality between groups and periods because $g_1' p_3 \neq 0$. This selection of contrasts is such that:

- (g_2, g_3) capture differences between groups within cohorts and are orthogonal to periods,
- (p_1, p_2) capture differences between periods which are orthogonal to groups, and

- g_1 , the difference between cohorts, is not orthogonal to p_3 , the contrast between the first and last period.

The implication is that 2 degrees of freedom for groups (ignoring periods), corresponding to g_2 and g_3 , are the same whether we fit groups before periods or periods before groups. Similarly, 2 degrees of freedom for periods (ignoring groups), corresponding to p_1 and p_2 , are the same whether we fit periods first or second. The only difference between the two analyses arises from the order of fitting g_1 and p_3 .

If we fit groups and periods together, the projector matrix is

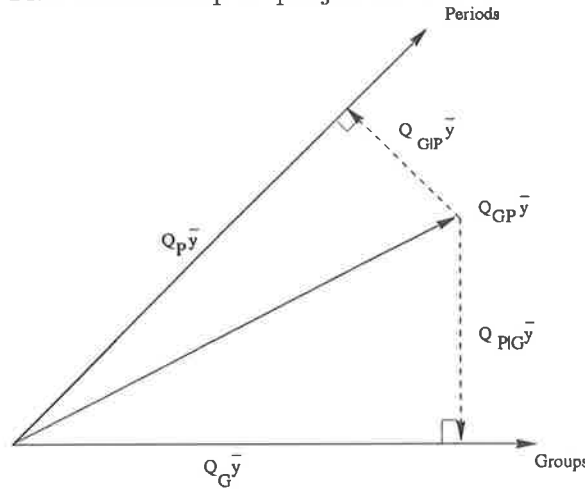
$$Q_{GP} = W(W'W)^{-1}W', \quad (6.5.4)$$

and, since all these columns are contrasts and hence $1'W = 0$, the projector matrix for the interaction stratum is

$$Q_{Int} = K - Q_{GP}.$$

Figure 6.2 shows the relationship between the vectors produced using these projector matrices. Thus, if we consider a vector $Q_{GP}\bar{y}$, as the projection of the vector $\bar{y}$ onto the space of Groups and Periods, in a case when the Groups and Periods are not orthogonal to one another, then the vector can be decomposed in two distinct ways. The first of these has a component in the Groups space, $Q_G\bar{y}$, and a remainder, $Q_{P|G}\bar{y}$, in the Periods (eliminating Groups) space. The second has a component in the Periods space, $Q_P\bar{y}$, and a remainder, $Q_{G|P}\bar{y}$, in the Groups (eliminating Periods) space.

Figure 6.2: The relationship of projector matrices in cohort design



We can obtain Q_{GP} directly, or we can obtain it by orthogonalizing the columns of W . This is conveniently done in one of two ways:

(i) we can rewrite W , equivalently, as

$$W^* = \begin{pmatrix} g_1 & g_2 & g_3 & p_1 & p_2 & (3p_3 - g_1) \end{pmatrix}$$

in which all columns are orthogonal then

$$Q_{GP} = Q_G + Q_{P|G},$$

where

$$Q_{P|G} = \frac{1}{8}p_1p_1' + \frac{1}{24}p_2p_2' + \frac{1}{24}(3p_3 - g_1)(3p_3 - g_1)'$$

represents the projector matrix for Periods eliminating Groups.

(ii) We can rewrite W , equivalently, as

$$W^+ = \begin{pmatrix} (g_1 - p_3) & g_2 & g_3 & p_1 & p_2 & p_3 \end{pmatrix}$$

in which, again, all columns are orthogonal, so that

$$Q_{GP} = Q_P + Q_{G|P}$$

and

$$Q_{G|P} = \frac{1}{8}(g_1 - p_3)(g_1 - p_3)' + \frac{1}{12}(g_2g_2' + g_3g_3')$$

represent the projector matrix for Groups eliminating Periods. From (6.5.3) we can derive Q_{GP} directly. We need the inverse consisting of the first and last row and column submatrix in (6.5.3) that is, the inverse of $\begin{bmatrix} 12 & 4 \\ 4 & 4 \end{bmatrix}$, namely $\frac{1}{8} \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix}$, so that

$$\begin{aligned} Q_{GP} &= \frac{1}{12}(g_2g_2' + g_3g_3') + \frac{1}{8}p_1p_1' + \frac{1}{24}p_2p_2' + \frac{1}{8} \begin{bmatrix} g_1 & p_3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} g_1' \\ p_3' \end{bmatrix} \\ &= \frac{1}{12}(g_2g_2' + g_3g_3') + \frac{1}{8}p_1p_1' + \frac{1}{24}p_2p_2' + \frac{1}{8}(g_1g_1' - g_1p_3' - p_3g_1' + 3p_3p_3') \end{aligned}$$

For the cohort design we can show the projector matrices in two different orders that is, in one case we fit periods first and in the second case we fit groups first. Table 6.6 shows these projector matrices for the cohort design. The matrices $A_1, A_2, B_1, B_2, C_1, C_2, D_1, D_2$ are shown in Table 6.7.

Projector Matrix	Cohort Design (Periods first)	Cohort Design (Groups first)
Periods	$Q_P = \frac{1}{6} \begin{pmatrix} A_1 & A_2 \\ A_2' & A_1^\dagger \end{pmatrix} \otimes J_2$	$Q_{P G} = \frac{1}{12} \begin{pmatrix} B_1 & B_2 \\ B_2' & B_1^\dagger \end{pmatrix}$
Groups	$Q_{G P} = \frac{1}{24} \begin{pmatrix} C_1 & C_2 \\ C_2' & C_1^\dagger \end{pmatrix}$	$Q_G = K_4 \otimes J_3$
Interaction	$Q_{Int} = \frac{1}{24} \begin{pmatrix} D_1 & D_2 \\ D_2' & D_1^\dagger \end{pmatrix}$	

($A_1^\dagger$ represents A_1 transposed about the reverse diagonal; that is, about the diagonal running from bottom left to top right of the matrix.)

Table 6.6: Projector matrices for cohort design for two orders of fitting

$A_1 = \begin{pmatrix} 5 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$	$A_2 = \begin{pmatrix} -1 & -1 & -1 \\ 2 & -1 & -1 \\ -1 & 2 & -1 \end{pmatrix}$
$B_1 = \begin{pmatrix} 8 & -4 & -4 \\ -4 & 5 & -1 \\ -4 & -1 & 5 \end{pmatrix}$	$B_2 = \begin{pmatrix} 0 & 0 & 0 \\ 3 & -3 & 0 \\ -3 & 3 & 0 \end{pmatrix}$
$C_1 = \begin{pmatrix} 4 & -4 & 4 & -4 & 4 & -4 \\ -4 & 4 & -4 & 4 & -4 & 4 \\ 4 & -4 & 7 & -1 & 7 & -1 \\ -4 & 4 & -1 & 7 & -1 & 7 \\ 4 & -4 & 7 & -1 & 7 & -1 \\ -4 & 4 & -1 & 7 & -1 & 7 \end{pmatrix}$	$C_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -3 & -3 & -3 & -3 & 0 & 0 \\ -3 & -3 & -3 & -3 & 0 & 0 \\ -3 & -3 & -3 & -3 & 0 & 0 \\ -3 & -3 & -3 & -3 & 0 & 0 \end{pmatrix}$
$D_1 = \begin{pmatrix} 8 & -8 & -4 & 4 & -4 & 4 \\ -8 & 8 & 4 & -4 & 4 & -4 \\ -4 & 4 & 11 & -5 & -7 & 1 \\ 4 & -4 & -5 & 11 & 1 & -7 \\ -4 & 4 & -7 & 1 & 11 & -5 \\ 4 & -4 & 1 & -7 & -5 & 11 \end{pmatrix}$	$D_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -3 & -3 & 3 & 3 & 0 & 0 \\ -3 & -3 & 3 & 3 & 0 & 0 \\ 3 & 3 & -3 & -3 & 0 & 0 \\ 3 & 3 & -3 & -3 & 0 & 0 \end{pmatrix}$

Table 6.7: Matrices introduced in the table of projector matrices for cohort design

Thus for the cohort design we have four strata i.e Groups with 2 degrees of freedom, Periods with 2 degrees of freedom, the rest of the Group+Period stratum with 2 degrees of freedom and the Group $\times$ Period interaction with 5 degrees of freedom. The partially confounded 2 degrees of freedom in Group+Period can be presented as the two contrasts

Gps(ig. Per)				Pers(elim. Gps)			
-1	-1	-1		-2	1	1	
-1	-1	-1		-2	1	1	
	1	1	1		-1	-1	2
	1	1	1		-1	-1	2
g_1				$3p_3 - g_3$			

or as the two contrasts:

Per(ig. Gps)				Gps(elim. Per)			
-1	0	0		0	-1	-1	
-1	0	0		0	-1	-1	
	0	0	1		1	1	0
	0	0	1		1	1	0
p_3				$g_1 - p_3$			

Period effects may be large and hence we would prefer to remove them first and base our treatment estimates solely on the information in the Group (elim. Periods) and interaction strata for which the projector matrices are $Q_{G|P}$ and Q_{Int} .

6.6 Treatment information

The matrix in (6.4.2) gives the total treatment information available after removal of the grand mean for both the standard and cohort designs. We now consider where that information resides in the different strata. For the standard design, this involves calculating $T'Q_iT$ ($i = 1, 2, 3$), while for the cohort design, we need $T'Q_PT$, $T'Q_{G|P}T$, $T'Q_{Int}T$. In each case, our intention is to discard the information contained in Periods (ig. Groups) on the grounds that there are few periods and that the effects of periods cannot reasonably be separated from the treatment effects contained therein. Table 6.8 shows where the treatment information occurs in the different strata.

Stratum	Standard Design	Cohort Design (Periods first)	Cohort Design (Groups first)
Period	$\frac{n}{6} \begin{pmatrix} 16 & 8 & 0 & 0 \\ 8 & 16 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 10 & 8 & 0 & 0 \\ 8 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 13 & 11 & 0 & 0 \\ 11 & 13 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
Group	$\frac{n}{6} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 3 & 3 & 0 & 0 \\ 3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}$
Int	$\frac{n}{6} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 48 & -24 \\ 0 & 0 & -24 & 16 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 3 & -3 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 0 & 0 & 48 & -24 \\ 0 & 0 & -24 & 16 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 3 & -3 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 0 & 0 & 48 & -24 \\ 0 & 0 & -24 & 16 \end{pmatrix}$
Total	$\frac{n}{6} \begin{pmatrix} 16 & 8 & 0 & 0 \\ 8 & 16 & 0 & 0 \\ 0 & 0 & 48 & -24 \\ 0 & 0 & -24 & 24 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 16 & 8 & 0 & 0 \\ 8 & 16 & 0 & 0 \\ 0 & 0 & 48 & -24 \\ 0 & 0 & -24 & 24 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 16 & 8 & 0 & 0 \\ 8 & 16 & 0 & 0 \\ 0 & 0 & 48 & -24 \\ 0 & 0 & -24 & 24 \end{pmatrix}$

Table 6.8: Information matrix in each stratum for both designs

Table 6.8 shows that the information about τ_D and λ_D is distributed between the strata in the same way for both standard and cohort designs. However, for τ_0 and λ_0 , the only information in the standard design is contained in the Period stratum. Thus information is unavailable unless we make strong assumptions about the period effects. For the cohort design, and in particular when we take out Period (ig. Groups) first, some of the information about (τ_0, λ_0) is available in both the Groups (elim. Periods) and Group $\times$ Periods interaction strata. The matrices in the middle column of Table 6.8 shows that the information matrices for (τ_0, λ_0) in Groups (elim. Periods) and Group $\times$ Periods are each of rank 1, the first corresponding to the sum and the second to the difference of the two parameters. This is formalised in the next section.

6.6.1 Information matrices on $\beta_1 = \tau_0 + \lambda_0$ and $\beta_2 = \tau_0 - \lambda_0$

By referring to the cohort design information matrices in Table 6.8, we can see that the information matrices of Groups (elim. Periods) and Interaction strata about τ_0 and λ_0 parameters are of rank 1 and have information on $(\tau_0 + \lambda_0)$ and $(\tau_0 - \lambda_0)$ respectively. Hence to get information matrices for our parameters of interest that is, τ_0 and λ_0 , we consider two new parameters, that is β_1 and β_2 , where

$$\tau_0 = (\beta_1 + \beta_2)/2$$

$$\lambda_0 = (\beta_1 - \beta_2)/2.$$

The information matrix for the parameters $\beta = \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix}'$ can be obtained in terms of the information matrix for the parameters $\theta_0 = \begin{bmatrix} \tau_0 & \lambda_0 \end{bmatrix}'$ by using the formula

$$\mathcal{I}_\beta = G\mathcal{I}_{\theta_0}G',$$

where

$$G = \frac{\partial \theta_0}{\partial \beta'} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}.$$

Then the information matrix about β_1 and β_2 for the Period (ig. Groups) is

$$GHG' = \frac{n}{24} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 10 & 8 \\ 8 & 10 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{n}{6} \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix},$$

where H is the top 2×2 of matrix in Table 6.8 corresponding to (τ_0, λ_0) from Period (ig. Groups) stratum. If this is done for the other information matrices in the table, we are led to the information matrices for β given in Table 6.9.

Stratum	Standard Design	Cohort Design (first Periods)	Cohort Design (first Groups)
Period	$\frac{n}{6} \begin{pmatrix} 12 & 0 \\ 0 & 4 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 12 & 0 \\ 0 & 1 \end{pmatrix}$
Group	0	$\frac{n}{6} \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}$	0
Interaction	0	$\frac{n}{6} \begin{pmatrix} 0 & 0 \\ 0 & 3 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 0 & 0 \\ 0 & 3 \end{pmatrix}$
Total	$\frac{n}{6} \begin{pmatrix} 12 & 0 \\ 0 & 4 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 12 & 0 \\ 0 & 4 \end{pmatrix}$	$\frac{n}{6} \begin{pmatrix} 12 & 0 \\ 0 & 4 \end{pmatrix}$

Table 6.9: Information matrix about β_1 and β_2 in each stratum for both designs

Since all matrices are diagonal, the estimates of β_1 and β_2 from each stratum will be independent. As we can see, information about $\beta_1 = (\tau_0 + \lambda_0)$ and $\beta_2 = (\tau_0 - \lambda_0)$ in the cohort design is present in each stratum. In the standard design, for each of β_1, β_2 all the information is in the Period stratum. The following Table 6.10 shows the percentage of information about β_1 and β_2 in each stratum for the cohort design, when we fit periods first.

Parameter	$\tau_0 + \lambda_0$	$\tau_0 - \lambda_0$
Periods(ig. Groups)	75	25
Groups (elim. Periods)	25	0
Interaction	0	75

Table 6.10: Percentage of information in each stratum about $\tau_0 + \lambda_0$ and $\tau_0 - \lambda_0$ for the cohort design

6.6.2 Where is the treatment information?

To get the estimate of θ , we must combine information from those strata which give us some information about θ and in addition we should determine the δ_i corresponding to each stratum which contributes to the information. We partition θ into two components that is, (τ_0, λ_0) and (τ_D, λ_D) , and we are looking for where the information for these

two parts of the parameters of interest lies. By referring to Section 6.5 we can say that Groups (elim. Periods) stratum can be decomposed into two substrata as Groups within Cohorts with the projector

$$(\frac{1}{g'_2 g_2} g_2 g'_2 + \frac{1}{g'_3 g_3} g_3 g'_3),$$

with 2 degrees of freedom and $\delta_2 = \sigma^2 + 3\sigma_s^2$ and the one degree of freedom for Cohorts (elim. Periods) with the projector $\frac{1}{d'd} dd'$, where $d = (g_1 - p_3)$. We need to establish the appropriate variance to use for the contrast d . The projector matrix corresponding to this contrast is $d(d'd)^{-1}d'$ so that the sum of squares for this stratum is

$$SS_{CeP} = \bar{y}' d(d'd)^{-1} d' \bar{y} = \frac{1}{d'd} (d' \bar{y})^2 = \frac{1}{8} (d' \bar{y})^2,$$

and the expected sum of squares ignoring the treatment effects

$$E(SS_{CeP}) = \frac{1}{8} \text{trace} E(d' \bar{y} \bar{y}' d).$$

Now we know that

$$\begin{aligned} \text{Var}(y) &= \sigma^2(I \otimes K) + (\sigma^2 + 3\sigma_s^2)(K \otimes J) \\ &= \sigma^2 I + 3\sigma_s^2(I \otimes J) - \sigma^2 J \\ &= \sigma^2(I - J) + 3\sigma_s^2(I \otimes J), \end{aligned} \tag{6.6.1}$$

and we can show that $d'(I - J)d = \delta = 8$ and $d'(I \otimes J)d = 16/3$. It follows that, ignoring treatments for the moment,

$$E(SS_{CeP}) = \sigma^2 + 2\sigma_s^2.$$

The contrasts g_2 and g_3 form the Group within Cohort stratum with two degrees of freedom and they provide the sum of squares

$$SS_{GwC} = \frac{1}{12} \text{trace} \{g'_2 \bar{y} \bar{y}' g_2 + g'_3 \bar{y} \bar{y}' g_3\},$$

for which the expected mean square can be expressed, using (6.6.2), as

$$E(MS_{GwC}) = \frac{1}{24} \text{trace} \left\{ \sum_{i=2}^3 [\sigma^2 g'_i (I - J) g_i + 3\sigma_s^2 g'_i (I \otimes J) g_i] \right\}.$$

Now we have

$$g'_i (I - J) g_i = 12, g'_i (I \otimes J) g_i = 12$$

for $i = 2, 3$, so that, ignoring the treatment effects again, we obtain

$$E(MS_{GwC}) = \sigma^2 + 3\sigma_s^2.$$

So to get θ , we have the following Table 6.11.

Stratum	Parameter estimable	degree of freedom	MSE
Cohort (elim. Periods)	(τ_0, λ_0)	1	$\sigma^2 + 2\sigma_s^2$
Group within Cohort	(τ_D, λ_D)	2	$\sigma^2 + 3\sigma_s^2$
Interaction	$(\tau_0, \lambda_0, \tau_D, \lambda_D)$	5	σ^2

Table 6.11: Those strata which contribute to estimate some parameters of interest

6.7 Combining information

The parameter estimates in (6.4.3) can now be obtained by combining the information from these three strata. The pair $(\hat{\tau}_0, \hat{\lambda}_0)$ and the pair $(\hat{\tau}_D, \hat{\lambda}_D)$ are uncorrelated with each other in all strata, and hence we can look at their estimates separately. Information about (τ_0, λ_0) is available in two strata, with weights σ^2 and $(\sigma^2 + 2\sigma_s^2)$, while information about (τ_D, λ_D) is available with weights σ^2 and $(\sigma^2 + 3\sigma_s^2)$. Estimates of σ^2 and σ_s^2 are obtained from the appropriate residual lines in the analysis of variance in Table 6.5, these being calculated in exactly the same way for both the standard and cohort design.

6.7.1 Estimation of (τ_0, λ_0) in cohort design only

For the average effect of treatments, that is (τ_0, λ_0) , we can obtain estimates from each of the Group (elim. Periods) and the interaction strata, as shown in the middle columns of Table 6.12, and then combine them with appropriate weights using:

$$\begin{aligned}
\begin{bmatrix} \hat{\tau}_0 \\ \hat{\lambda}_0 \end{bmatrix} &= 6 \left\{ \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} + (1 + 2\rho) \begin{bmatrix} 3 & -3 \\ -3 & 3 \end{bmatrix} \right\}^{-1} T_1' \{Q_{G|P} + (1 + 2\rho)Q_{Int}\} \bar{y} \\
&= \frac{1}{(1 + 2\rho)} \begin{bmatrix} 1 + \rho & \rho \\ \rho & 1 + \rho \end{bmatrix} T_1' (Q_{G|P} + (1 + 2\rho)Q_{Int}) \bar{y},
\end{aligned}$$

where T_1 is a submatrix of matrix T given by the first two columns in (6.3.3), corresponding to the pair of parameters (τ_0, λ_0) . This leads to the contrasts represented in the last column of Table 6.12. The covariance matrix of these combined estimates is

$$\Sigma_1 = \frac{\sigma^2}{n} \begin{bmatrix} 1 + \rho & \rho \\ \rho & 1 + \rho \end{bmatrix} \quad (6.7.1)$$

		Cohort design																																																		
Estimator	Periods (ig. Groups)	Groups (elim. Periods)	Interaction —	Combined —																																																
$\hat{\tau}_0$	$\frac{1}{6}$ <table><tr><td>-4</td><td>1</td><td>2</td></tr><tr><td>-4</td><td>1</td><td>2</td></tr><tr><td>1</td><td>2</td><td>-2</td></tr><tr><td>1</td><td>2</td><td>-2</td></tr></table>	-4	1	2	-4	1	2	1	2	-2	1	2	-2	$\frac{1}{36}$ <table><tr><td>0</td><td>9</td><td>9</td></tr><tr><td>0</td><td>9</td><td>9</td></tr><tr><td>-9</td><td>-9</td><td>0</td></tr><tr><td>-9</td><td>-9</td><td>0</td></tr></table>	0	9	9	0	9	9	-9	-9	0	-9	-9	0	$\frac{1}{36}$ <table><tr><td>0</td><td>9</td><td>-9</td></tr><tr><td>0</td><td>9</td><td>-9</td></tr><tr><td>-9</td><td>9</td><td>0</td></tr><tr><td>-9</td><td>9</td><td>0</td></tr></table>	0	9	-9	0	9	-9	-9	9	0	-9	9	0	$\frac{1}{2}$ <table><tr><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td></tr><tr><td>-1</td><td>0</td><td>0</td></tr><tr><td>-1</td><td>0</td><td>0</td></tr></table>	0	1	0	0	1	0	-1	0	0	-1	0	0
-4	1	2																																																		
-4	1	2																																																		
1	2	-2																																																		
1	2	-2																																																		
0	9	9																																																		
0	9	9																																																		
-9	-9	0																																																		
-9	-9	0																																																		
0	9	-9																																																		
0	9	-9																																																		
-9	9	0																																																		
-9	9	0																																																		
0	1	0																																																		
0	1	0																																																		
-1	0	0																																																		
-1	0	0																																																		
$\hat{\lambda}_0$	$\frac{1}{6}$ <table><tr><td>2</td><td>-2</td><td>-1</td></tr><tr><td>2</td><td>-2</td><td>-1</td></tr><tr><td>-2</td><td>-1</td><td>4</td></tr><tr><td>-2</td><td>-1</td><td>4</td></tr></table>	2	-2	-1	2	-2	-1	-2	-1	4	-2	-1	4	$\frac{1}{36}$ <table><tr><td>0</td><td>9</td><td>9</td></tr><tr><td>0</td><td>9</td><td>9</td></tr><tr><td>-9</td><td>-9</td><td>0</td></tr><tr><td>-9</td><td>-9</td><td>0</td></tr></table>	0	9	9	0	9	9	-9	-9	0	-9	-9	0	$\frac{1}{36}$ <table><tr><td>0</td><td>-9</td><td>9</td></tr><tr><td>0</td><td>-9</td><td>9</td></tr><tr><td>9</td><td>-9</td><td>0</td></tr><tr><td>9</td><td>-9</td><td>0</td></tr></table>	0	-9	9	0	-9	9	9	-9	0	9	-9	0	$\frac{1}{2}$ <table><tr><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>-1</td><td>0</td></tr><tr><td>0</td><td>-1</td><td>0</td></tr></table>	0	0	1	0	0	1	0	-1	0	0	-1	0
2	-2	-1																																																		
2	-2	-1																																																		
-2	-1	4																																																		
-2	-1	4																																																		
0	9	9																																																		
0	9	9																																																		
-9	-9	0																																																		
-9	-9	0																																																		
0	-9	9																																																		
0	-9	9																																																		
9	-9	0																																																		
9	-9	0																																																		
0	0	1																																																		
0	0	1																																																		
0	-1	0																																																		
0	-1	0																																																		

Table 6.12: Contrasts for estimating parameters in cohort design

	Standard Design															
Estimator	Period	Group	Interaction	Combined												
$\hat{\tau}_0$	<table><tr><td>-1</td><td>1</td><td>0</td></tr><tr><td>-1</td><td>1</td><td>0</td></tr><tr><td>-1</td><td>1</td><td>0</td></tr><tr><td>-1</td><td>1</td><td>0</td></tr></table>	-1	1	0	-1	1	0	-1	1	0	-1	1	0	None	None	None
-1	1	0														
-1	1	0														
-1	1	0														
-1	1	0														
$\hat{\lambda}_0$	<table><tr><td>0</td><td>-1</td><td>1</td></tr><tr><td>0</td><td>-1</td><td>1</td></tr><tr><td>0</td><td>-1</td><td>1</td></tr><tr><td>0</td><td>-1</td><td>1</td></tr></table>	0	-1	1	0	-1	1	0	-1	1	0	-1	1	None	None	None
0	-1	1														
0	-1	1														
0	-1	1														
0	-1	1														

Table 6.13: Contrasts for estimating parameters in standard design

Now we can have a look in each stratum at the contrasts for the two designs and compare them. First, for the standard design, we have the results shown in Table 6.13. However, these results are not useful, because all information about the two parameters are in the Period stratum and in the presence of strong or unknown period effects, we cannot estimate the average effects of treatments. In fact this information in the Period stratum on τ_0 and λ_0 in Table 6.13 is not available when we suspect period effects. The cohort design offers a way to get more useful information for average effects of treatments.

6.7.2 Estimation of (τ_D, λ_D) in both designs

Table 6.8 shows that the information about (τ_D, λ_D) is in Groups (elim. Periods) and the Interaction strata, for both standard and cohort designs. If we put $\rho = \sigma_s^2/\sigma^2$, then combined estimates can then be found as

$$\begin{aligned} \begin{bmatrix} \hat{\tau}_D \\ \hat{\lambda}_D \end{bmatrix} &= 6 \left\{ \begin{bmatrix} 0 & 0 \\ 0 & 8 \end{bmatrix} + (1 + 3\rho) \begin{bmatrix} 48 & -24 \\ -24 & 16 \end{bmatrix} \right\}^{-1} T_2' \{Q_{G|P} + (1 + 3\rho)Q_{Int}\} \bar{y} \\ &= \frac{1}{4(1 + \rho)} \begin{bmatrix} (1 + 2\rho)/(1 + 3\rho) & 1 \\ 1 & 2 \end{bmatrix} T_2' \{Q_{G|P} + (1 + 3\rho)Q_{Int}\} \bar{y}, \end{aligned}$$

where T_2 is a submatrix of matrix T given by the last two columns of (6.3.3) corresponding to the pair of parameters (τ_D, λ_D) . This leads to two contrasts which should be applied to the vector $\bar{y}$, and these may be represented by:

$$\begin{aligned} \hat{\tau}_D &= \frac{1}{4(1 + \rho)} \begin{bmatrix} -\rho & (1 + \rho) & 0 \\ \rho & -(1 + \rho) & 0 \\ -\rho & (1 + \rho) & 0 \\ \rho & -(1 + \rho) & 0 \end{bmatrix} \\ \hat{\lambda}_D &= \frac{1}{4(1 + \rho)} \begin{bmatrix} -2\rho & (1 + \rho) & (1 + \rho) \\ 2\rho & -(1 + \rho) & -(1 + \rho) \\ -2\rho & (1 + \rho) & (1 + \rho) \\ 2\rho & -(1 + \rho) & -(1 + \rho) \end{bmatrix} \end{aligned}$$

The covariance matrix of these estimates is

$$\Sigma_2 = \frac{\sigma^2}{4n(1+\rho)} \begin{bmatrix} 1+2\rho & 1+3\rho \\ 1+3\rho & 2(1+3\rho) \end{bmatrix}. \quad (6.7.2)$$

Except for the change in the definition of n , these estimates and their covariance matrix agree with the results of Section 3.3.1.

6.7.3 Estimates of treatment effect minus baseline

We might want to look at estimates of $\tau_A - \mu$ and $\tau_B - \mu$ where μ is the baseline measurement. The estimates of (τ_0, λ_0) and (τ_D, λ_D) are uncorrelated, and hence the covariance matrices of $\hat{\theta}$ can be found in (6.7.1) and (6.7.2). As we defined our parameters in Section 6.3 we have

$$\begin{aligned} \tau_A - \mu &= \tau_0 + \tau_D \\ \tau_B - \mu &= \tau_0 - \tau_D \\ \lambda_A - \mu &= \lambda_0 + \lambda_D \\ \lambda_B - \mu &= \lambda_0 - \lambda_D \end{aligned} \quad (6.7.3)$$

If we put $\hat{\eta}' = \begin{pmatrix} \hat{\tau}_A - \hat{\mu} & \hat{\tau}_B - \hat{\mu} & \hat{\lambda}_A - \hat{\mu} & \hat{\lambda}_B - \hat{\mu} \end{pmatrix}$, then to get the estimates and covariance matrix of new vector $\hat{\eta}$ from (6.7.3) we can write

$$\hat{\eta} = \begin{pmatrix} \hat{\tau}_A - \hat{\mu} \\ \hat{\tau}_B - \hat{\mu} \\ \hat{\lambda}_A - \hat{\mu} \\ \hat{\lambda}_B - \hat{\mu} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{pmatrix} \hat{\tau}_0 \\ \hat{\lambda}_0 \\ \hat{\tau}_D \\ \hat{\lambda}_D \end{pmatrix} = F\hat{\theta},$$

for which the estimates are:

$$\hat{\tau}_A - \hat{\mu} = \frac{1}{4(1+\rho)} \begin{array}{|c|c|c|c|} \hline -\rho & 3(1+\rho) & 0 & \\ \hline \rho & 1+\rho & 0 & \\ \hline & -(2+3\rho) & 1+\rho & 0 \\ \hline & -(2+\rho) & -(1+\rho) & 0 \\ \hline \end{array}$$

$$\begin{aligned}
\hat{\tau}_B - \hat{\mu} &= \frac{1}{4(1+\rho)} \begin{array}{|c|c|c|c|} \hline \rho & 1+\rho & 0 & \\ \hline -\rho & 3(1+\rho) & 0 & \\ \hline & -(2+\rho) & -(1+\rho) & 0 \\ \hline & -(2+3\rho) & 1+\rho & 0 \\ \hline \end{array} \\
\hat{\lambda}_A - \hat{\mu} &= \frac{1}{4(1+\rho)} \begin{array}{|c|c|c|c|} \hline -2\rho & 1+\rho & 3(1+\rho) & \\ \hline 2\rho & -(1+\rho) & 1+\rho & \\ \hline & -2\rho & -(1+\rho) & 1+\rho \\ \hline & 2\rho & -3(1+\rho) & -(1+\rho) \\ \hline \end{array} \\
\hat{\lambda}_B - \hat{\mu} &= \frac{1}{4(1+\rho)} \begin{array}{|c|c|c|c|} \hline 2\rho & -(1+\rho) & 1+\rho & \\ \hline -2\rho & 1+\rho & 3(1+\rho) & \\ \hline & 2\rho & -3(1+\rho) & -(1+\rho) \\ \hline & -2\rho & -(1+\rho) & 1+\rho \\ \hline \end{array}
\end{aligned}$$

These four contrasts should be applied to the vector of means $\bar{y}$ to estimate the effects. The covariance matrix of $\hat{\eta}$ can be obtained by using the fact that

$$\text{Var}(\hat{\eta}) = \Sigma_{\hat{\eta}} = F\Sigma_{\hat{\theta}}F',$$

which gives

$$\Sigma_{\hat{\eta}} = \frac{\sigma^2}{4n(1+\rho)} \begin{bmatrix} 5+10\rho+4\rho^2 & 3+6\rho+4\rho^2 & 1+7\rho+4\rho^2 & -1+\rho+4\rho^2 \\ 3+6\rho+4\rho^2 & 5+10\rho+4\rho^2 & -1+\rho+4\rho^2 & 1+7\rho+4\rho^2 \\ 1+7\rho+4\rho^2 & -1+\rho+4\rho^2 & 6+14\rho+4\rho^2 & 2+2\rho+4\rho^2 \\ -1+\rho+4\rho^2 & 1+7\rho+4\rho^2 & 2+2\rho+4\rho^2 & 6+14\rho+4\rho^2 \end{bmatrix}$$

6.8 Limiting estimates in terms of ρ

In this section we consider the estimates of the parameter of interest when ρ tends to zero and infinity.

6.8.1 The estimates and covariance matrix when $\rho \rightarrow \infty$

If $\rho \rightarrow \infty$, that is when $\sigma_s^2 \rightarrow \infty$, the subjects are considered as fixed effects and we obtain the following estimates of τ_D and λ_D :

$$\hat{\tau}_D = \frac{1}{4} \begin{array}{|c|c|c|c|} \hline -1 & 1 & 0 & \\ \hline 1 & -1 & 0 & \\ \hline -1 & 1 & 0 & \\ \hline 1 & -1 & 0 & \\ \hline \end{array}$$

$$\hat{\lambda}_D = \frac{1}{4} \begin{array}{|c|c|c|c|} \hline -2 & 1 & 1 & \\ \hline 2 & -1 & -1 & \\ \hline -2 & 1 & 1 & \\ \hline 2 & -1 & -1 & \\ \hline \end{array}$$

These estimates are orthogonal to both Groups and Periods, and are those estimates obtained solely from the Interaction stratum in Table 6.8.

Because the estimates of τ_0 and λ_0 are independent of ρ , then those estimates do not change when ρ changes. The covariance matrix of estimators however, tends to infinity when $\rho \rightarrow \infty$ because the contrasts involve comparisons between as well as within subjects. For $\hat{\eta}$, the cohort design gives the estimates:

$$\hat{\tau}_A - \hat{\mu} = \frac{1}{4} \begin{array}{|c|c|c|c|} \hline -1 & 3 & 0 & \\ \hline 1 & 1 & 0 & \\ \hline & -3 & 1 & 0 \\ \hline & -1 & -1 & 0 \\ \hline \end{array}$$

$$\hat{\tau}_B - \hat{\mu} = \frac{1}{4} \begin{array}{|c|c|c|c|} \hline 1 & 1 & 0 & \\ \hline -1 & 3 & 0 & \\ \hline & -1 & -1 & 0 \\ \hline & -3 & 1 & 0 \\ \hline \end{array}$$

$$\hat{\lambda}_A - \hat{\mu} = \frac{1}{2} \begin{array}{|c|c|c|c|} \hline -2 & 1 & 3 & \\ \hline 2 & 1 & 1 & \\ \hline & -2 & -1 & 1 \\ \hline & 2 & -3 & -1 \\ \hline \end{array}$$

$$\hat{\lambda}_B - \hat{\mu} = \frac{1}{2} \begin{array}{|c|c|c|c|} \hline 2 & -1 & 1 & \\ \hline -2 & 1 & 3 & \\ \hline & 2 & -3 & -1 \\ \hline & -2 & -1 & 1 \\ \hline \end{array}$$

All elements of the covariance matrix for $\hat{\eta}$ in this case become large as $\rho \rightarrow \infty$ because the estimates are not orthogonal to subjects.

6.8.2 Estimates and their covariance when $\rho \rightarrow 0$

In this subsection we consider the estimates and their covariance when $\rho \rightarrow 0$, or in other words when there is no difference between the subjects. Since the estimate of the pair (τ_0, λ_0) is independent of ρ then it remains the same, but for the pair (τ_D, λ_D) we have

$$\hat{\tau}_D = \frac{1}{4} \begin{array}{|c|c|c|} \hline 0 & 1 & 0 \\ \hline 0 & -1 & 0 \\ \hline 0 & 1 & 0 \\ \hline 0 & -1 & 0 \\ \hline \end{array}$$

$$\hat{\lambda}_D = \frac{1}{4} \begin{array}{|c|c|c|} \hline 0 & 1 & 1 \\ \hline 0 & -1 & -1 \\ \hline 0 & 1 & 1 \\ \hline 0 & -1 & -1 \\ \hline \end{array}$$

The covariance matrix of $\hat{\theta}$ is

$$\lim_{\rho \rightarrow 0} \Sigma_{\hat{\theta}} = \frac{\sigma^2}{4n} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

The result in this case for $\hat{\eta}$ is given below and shows that the estimates are orthogonal to Periods, but not to subjects:

$$\begin{aligned}
\hat{\tau}_A - \hat{\mu} &= \frac{1}{4} \begin{bmatrix} 0 & 3 & 0 & \\ 0 & 1 & 0 & \\ & -2 & 1 & 0 \\ & -2 & -1 & 0 \end{bmatrix} \\
\hat{\tau}_B - \hat{\mu} &= \frac{1}{4} \begin{bmatrix} 0 & 1 & 0 & \\ 0 & 3 & 0 & \\ & -2 & -1 & 0 \\ & -2 & 1 & 0 \end{bmatrix} \\
\hat{\lambda}_A - \hat{\mu} &= \frac{1}{2} \begin{bmatrix} 0 & 1 & 3 & \\ 0 & -1 & 1 & \\ & 0 & -1 & 1 \\ & 0 & -3 & -1 \end{bmatrix} \\
\hat{\lambda}_B - \hat{\mu} &= \frac{1}{2} \begin{bmatrix} 0 & -1 & 1 & \\ 0 & 1 & 3 & \\ & 0 & -3 & -1 \\ & 0 & -1 & 1 \end{bmatrix}
\end{aligned}$$

The covariance matrix for $\hat{\eta}$ in this case is

$$\lim_{\rho \rightarrow 0} \Sigma_{\hat{\eta}} = \frac{\sigma^2}{4n} \begin{bmatrix} 5 & 3 & 1 & -1 \\ 3 & 5 & -1 & 1 \\ 1 & -1 & 6 & 2 \\ -1 & 1 & 2 & 6 \end{bmatrix}$$

6.9 Conclusion

By using the cohort design in crossover design, we can recover some information about (τ_0, λ_0) and obtain estimates for them within and between subjects. This method opens the way for a more detailed examination of the topic.

It has been shown that the block effects corresponding to Cohorts and Periods are not orthogonal to one another. This will also be true in more general cases. The resolution of this in general requires the application of the results of James & Wilkinson (1971) which

allows the appropriate partitioning of these spaces. Although the original work of James & Wilkinson (1971) is concerned with Blocks and Treatments, we apply it to a situation with two nonorthogonal blocking factors.

Chapter 7

Block structure of cohort designs

7.1 Introduction

In the previous Chapter, we considered a particular example of a cohort design in which a second cohort was delayed by one period. It was shown that treatment information was available in the Cohort (elim. Periods) and Interaction strata.

We will now compare a variety of such designs with the corresponding standard design in which subsequent cohorts do not have delayed entry.

In cohort designs two basic types of structures are the block structure of the experiment and the treatment structure. The purpose of this chapter is to present an analysis of the block structure of the general class of cohort designs. We will consider the general case where each of c cohorts are observed for p periods, but with each successive cohort entering the study after a delay of one period. Thus, the experiment will occur over a total of $(p + c - 1)$ periods. Within each cohort we will assume there are g groups of n subjects, where each subject within a group receives the same treatment regime. We will show the ANOVA tables for the standard and the cohort designs with a data vector of length $cp gn$ in terms of the block structure, where g is the number of groups in each cohort and n is the number of subjects in each group. We show that the differences between them are only due to differing idempotent matrices for cohort, periods and the interaction.

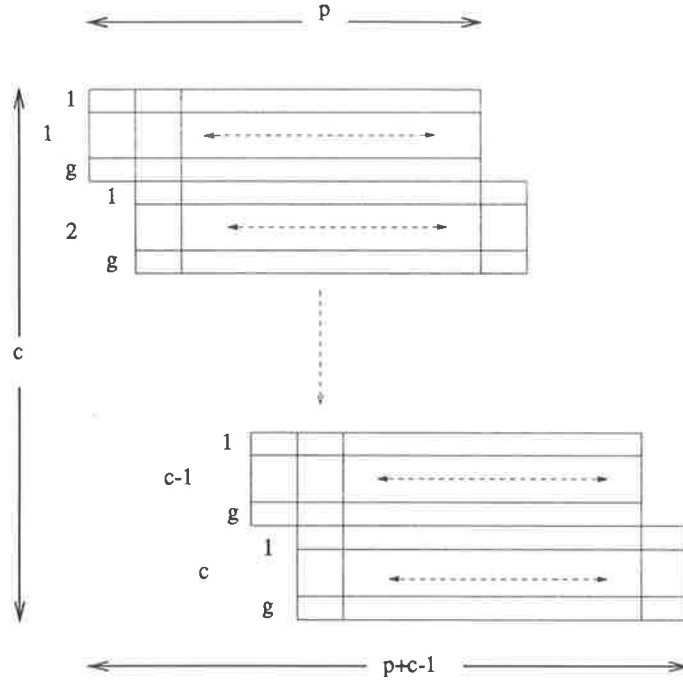
We use the work of James & Wilkinson (1971), because James and Wilkinson's geometrical formulation provides a way of looking at the $c \times (p + c - 1)$ array of means

and identifying the contrasts for Cohort(elim. Periods) and their expected mean squares. For this purpose we need to split the contrasts for Cohort(elim. Periods) into different 1 degree of freedom contrasts each with a possibly different expected mean square and then to identify the projector matrix for projecting onto the vector space spanned by the columns of Cohort(elim. Periods). We then consider several examples in detail. In a special study we have got the orthogonal contrasts for the three strata for two and three cohort designs with up to 7 periods in each cohort. For these special cases we obtain the analysis of variance table in Section 7.3 and by using a function which we have written in S-PLUS (Venables & Ripley (1996)) we show their projector matrices and the expected mean square for each stratum.

7.2 Cohort design in general

The design and analysis of cohort designs for the two-treatment three-period design with one baseline measurement is carried out in the previous chapter. In that design, it was found that there was nonorthogonality between periods and cohorts, and this led to using the Cohort (elim. Periods) sum of squares to provide some of the treatment information. In this section, we shall describe the block structure of a general cohort design and our response will be supposed a linear function of parameters with a covariance matrix whose spectral form is governed by the blocking arrangement. The model has c cohorts, p periods in each cohort where each successive cohort starts one period later than the previous one, such that the design has a total of $p + c - 1$ periods. Within each cohort we have g groups and within each group there are n subjects. These groups will eventually each be assigned a different treatment regime, but the analysis in this Section will rely on the fact that the ng subjects in a cohort are randomly allocated to these groups. The design is shown in the following layout in Figure 7.1.

Figure 7.1: Full cohort design length $cpgn$



In this and subsequent tables, we suppose that the data are written in standard order, running over the last subscript first, where the data values are given as y_{ijkl} , for the i th cohort ($i = 1, 2, \dots, c$), j th period ($j = 1, 2, \dots, p$), k th group in i th cohort ($k = 1, 2, \dots, g$), and l th subject in k th group ($l = 1, 2, \dots, n$). Data from the experiment will be denoted by y , an $cpgn \times 1$ column vector, with mean $E(y)$ and variance-covariance matrix Q

We consider a model which can be represented as

$$y_{ijkl} = \mu + \gamma_i + \pi_j + \beta_{ikl} + \tau_{(ijkl)} + \epsilon_{ijkl} \quad (7.2.1)$$

where μ , γ_i and π_j are the grand mean, i th cohort, j th period effect respectively, and β_{ikl} is subject effect from the l th subject in k th group in i th cohort and $\tau_{(ijkl)}$ at this stage is used to describe the set of treatment effects applied to the l th subject in k th group in j th period in i th cohort. In this model, the β_{ikl} are independent random variables which are normally distributed with mean 0 and variance σ_s^2 and are independent of the ϵ_{ijkl} which are independent random normal variables with mean 0 and variance σ^2 .

7.3 Analysis of variance

In the previous section some general structure was laid down for the general cohort design. Now in this section we will provide the analysis of variance tables for both standard and cohort designs which simply identifies the components in the block structure and ignores, for the moment, the treatment structure. This same model could also be used for the standard design in which there were no delays from one cohort to the next. The data vector y of length $cpgn$, written in standard order, can be partitioned into various components, corresponding to the terms that can be identified in the linear model in (7.2.1), and the expected mean squares determined. The general ANOVA for the standard design is shown in Table 7.1.

Source	Idempotent matrix	Degrees of Freedom	EMS
GM	$J_c \otimes J_p \otimes J_g \otimes J_n$	1	
Period	$J_c \otimes K_p \otimes J_g \otimes J_n$	$p - 1$	
Cohort	$K_c \otimes J_p \otimes J_g \otimes J_n$	$c - 1$	$\sigma^2 + p\sigma_s^2$
Group within Cohort	$I_c \otimes J_p \otimes K_g \otimes J_n$	$c(g - 1)$	$\sigma^2 + p\sigma_s^2$
Subject within Group	$I_c \otimes J_p \otimes I_g \otimes K_n$	$cg(n - 1)$	$\sigma^2 + p\sigma_s^2$
Subjects		$cgn - 1$	
Period $\times$ Cohort	$K_c \otimes K_p \otimes J_g \otimes J_n$	$(c - 1)(p - 1)$	σ^2
Period $\times$ (Group within Cohort)	$I_c \otimes K_p \otimes K_g \otimes J_n$	$c(p - 1)(g - 1)$	σ^2
Periods $\times$ (Subject within Group)	$I_c \otimes K_p \otimes I_g \otimes K_n$	$c(p - 1)g(n - 1)$	σ^2
Periods $\times$ Subjects		$(cgn - 1)(p - 1)$	
Total		$cpgn$	

Table 7.1: ANOVA table for standard design, no treatment terms.

If we apply the model to the cohort design shown in Figure 7.1, the matrix for periods becomes more complex, since there are now $(p + c - 1)$ periods. As noted in the previous chapter, periods and cohorts are no longer orthogonal to one another. However, groups within cohorts and subjects within groups are still orthogonal to periods and hence we can write Table 7.2 for the cohort design.

Source	Idempotent matrix	Degrees of Freedom	EMS
GM	$J_c \otimes J_p \otimes J_g \otimes J_n$	1	
Periods (ig. Cohort)	$S_1 \otimes J_g \otimes J_n$	$p + c - 2$	
Cohorts (elim. periods)	$S_2 \otimes J_g \otimes J_n$	$c - 1$	
Group within Cohort	$I_c \otimes J_p \otimes K_g \otimes J_n$	$c(g - 1)$	$\sigma^2 + p\sigma_s^2$
Subject within Group	$I_c \otimes J_p \otimes I_g \otimes K_n$	$cg(n - 1)$	$\sigma^2 + p\sigma_s^2$
Subjects		$cgn - 1$	
Period $\times$ Cohort	$S_3 \otimes J_g \otimes J_n$	$(c - 1)(p - 2)$	σ^2
Period $\times$ (Group within Cohort)	$I_c \otimes K_p \otimes K_g \otimes J_n$	$c(p - 1)(g - 1)$	σ^2
Periods $\times$ (Subject within Group)	$I_c \otimes K_p \otimes I_g \otimes K_n$	$c(p - 1)g(n - 1)$	σ^2
Periods $\times$ Subjects		$(cgn - 1)(p - 1) - (c - 1)$	
Total		$cpgn$	

Table 7.2: ANOVA table for general cohort design, no treatments term

In Table 7.2 we consider the projector matrix which takes out Period effects from the model and then consider a stratum named Cohort (elim. Periods) which is orthogonal to Periods (ig. Cohort) stratum. As we can see, many lines in Tables 7.1 and 7.2 are the same. The differences occur only in Periods (ig. Cohort), Cohort (elim. Periods) and Periods $\times$ Cohort interaction. For these three lines, the idempotent matrices always end in $J_g \otimes J_n$, indicating that we can average over all subjects within a cohort. Thus, to get a block structure for the general cohort design we need only consider the vector of (Cohort $\times$ Period) means of length cp . In Table 7.2, S_1 , S_2 and S_3 are projector matrices for Period (ig. Cohort), Cohort (elim. Periods) and Period $\times$ Cohort interaction applied in each case to the vector of Cohort $\times$ Period means of length cp . These three projector matrices add together to give idempotent matrix $(I - J)$ of dimension $(cp - 1)$. We will show how these can be presented in the next section.

7.4 Fitting periods and cohorts

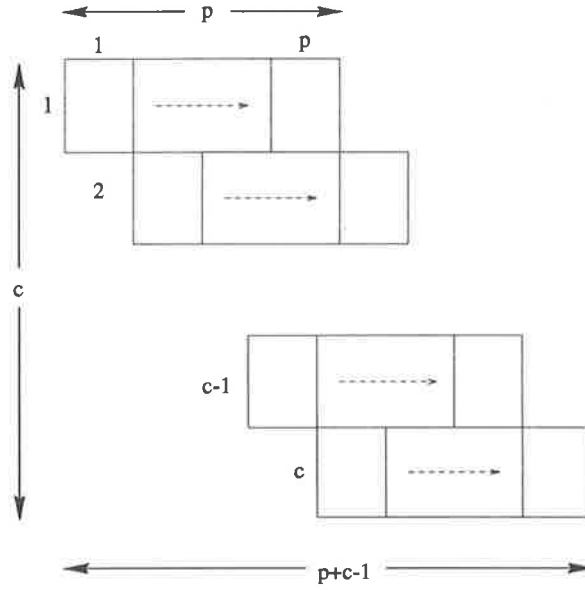
We now consider just the table of Cohort $\times$ Period means with length cp , as shown in Figure 7.2. Considering this in standard order across the rows, we get the model

$$y = \mu 1 + P\pi + C\gamma + T\tau + \epsilon, \quad (7.4.1)$$

where $T\tau$ considers whatever treatments are applied and ϵ indicates the random effects due to subject and error. In this case, each mean will now have a variance of $(\sigma^2 + \sigma_s^2)/gn$, with observations in the same cohort having covariance σ_s^2/gn . We will ignore the factor gn in this section, for convenience, and regard each value as a single observation rather than as a mean of gn values. The variance-covariance matrix of ϵ can then be written

$$Var(\epsilon) = \sigma^2(I_c \otimes K_p) + (\sigma^2 + p\sigma_s^2)(I_c \otimes J_p).$$

Figure 7.2: Full cohort design, length cp



For the Cohort $\times$ Period table of means we can identify the components in the block structure. This will provide five strata whose orthogonal projector matrices can be shown in the following;

- $J_c \otimes J_p$ for grand mean;
- $S_0 = K_c \otimes J_p$ for Cohorts ignoring periods;
- $S_1 = P(P'P)^{-1}P' - J$ for Periods ignoring Cohorts;
- $S_2 = X(X'X)^{-1}X' - P(P'P)^{-1}P'$ for Cohort eliminating periods;
- $S_3 = I - X(X'X)^{-1}X'$ for Interaction stratum;

where $X = \begin{bmatrix} P & : & C \end{bmatrix}$. Note that $C = I_c \otimes 1_p$, and $(X'X)$ is singular with rank of $(p+2c-3)$. In this formulation, the S_i determine the block structure of the cohort design

experiment and with respect to the usual Euclidean norm divide $\mathbf{R}^{cp}$ into subspaces called strata. We can think of the columns of C and P as generating subspaces of $\mathbf{R}^{cp}$ with dimension $c - 1$ and $p + c - 2$ respectively, and corresponding idempotent matrices $S_0 = K_c \otimes J_p$ and S_1 , after removing the grand mean. These subspaces are not orthogonal to each other. To explore the relationship between these subspaces, we turn to the results of James & Wilkinson (1971).

7.5 James and Wilkinson theorem

In the following section we summarise a geometrical approach to the analysis of nonorthogonal designs generated by projection matrices which was initially worked out by James & Wilkinson (1971). Such a geometrical approach has quite general application to problems in experimental design. In our cohort design, due to the nonorthogonality of two strata in the design, we need to use the result of the above mentioned paper.

James and Wilkinson show that, if the projector matrices onto the spaces $\mathcal{U}$ and $\mathcal{V}$ with dimension (q, m) , such that $q \geq m$, are U and V then there exist spaces

$$\begin{aligned} \mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_k, \mathcal{U}_{k+1}, \\ \mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_k, \end{aligned}$$

and the projector matrices

$$\begin{aligned} U_1, U_2, \dots, U_k, U_{k+1} \\ V_1, V_2, \dots, V_k, \end{aligned}$$

such that $\text{rank}(U_i) = \text{rank}(V_i) = r_i$, $\sum_{i=1}^k r_i = m$ and $\sum_{i=1}^{k+1} r_i = q$, such that,

$$U_i V_j = 0 \quad (i \neq j), \tag{7.5.1}$$

$$U_i U_j = 0 \quad (i \neq j),$$

$$V_i V_j = 0 \quad (i \neq j),$$

$$U_i V_i U_i = \rho_i U_i, \tag{7.5.2}$$

$$V_i U_i V_i = \rho_i V_i \quad (i, j = 1, 2, \dots, k),$$

where $1 \geq \rho_1 > \dots > \rho_k \geq 0$. Then ρ_i are the distinct eigenvalues of VUV and UVU , and if α_i is such that $\cos^2 \alpha_i = \rho_i$, then α_i is the angle between the projection $U_i y$ of any

vector into $\mathcal{U}_i$ and the projection $V_i y$ of that same vector into $\mathcal{V}_i$. Now in the next section we show how we can get these ρ_i 's as the canonical correlation coefficients between the columns of C and the columns of P .

7.5.1 Canonical variables

In the analysis of cohort designs we use the concept of the canonical correlation of two canonical variables. Krzanowski & Marriott (1994) develop this idea as follows. Suppose P and C are $N \times q$ and $N \times m$ respectively, and suppose the variance-covariance V of the rows of $X = [P:C]$ is partitioned as

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{bmatrix} \quad (7.5.3)$$

so that V_{11} and V_{22} are the variance-covariance matrices of the columns of P and C respectively, while V_{12} contains the covariances between the columns in P and those in C . In particular, we can show that $V_{11} = \frac{1}{N-1} P'(I - J)P$, $V_{12} = \frac{1}{N-1} P'(I - J)C$, and $V_{22} = \frac{1}{N-1} C'(I - J)C$, where $(I - J)$ in each case is $N \times N$. Now consider the correlation between any linear combination $P\mathbf{a}$ of the rows of P and any linear combination $C\mathbf{b}$ of the rows of C , namely

$$\rho = \frac{\mathbf{a}'V_{12}\mathbf{b}}{\sqrt{\mathbf{a}'V_{11}\mathbf{a} \cdot \mathbf{b}'V_{22}\mathbf{b}}}. \quad (7.5.4)$$

Next, choose $\mathbf{a}$ and $\mathbf{b}$ to maximise this correlation. This is equivalent to maximising

$$F = \mathbf{a}'V_{12}\mathbf{b} - \frac{1}{2}\lambda(\mathbf{a}'V_{11}\mathbf{a} - 1) - \frac{1}{2}\mu(\mathbf{b}'V_{22}\mathbf{b} - 1)$$

where λ and μ are Lagrange multipliers, since we can choose $\mathbf{a}$ and $\mathbf{b}$ to make $\mathbf{a}'V_{11}\mathbf{a}$ and $\mathbf{b}'V_{22}\mathbf{b} = 1$. Differentiating with respect to $\mathbf{a}$ and $\mathbf{b}$ gives

$$\begin{aligned} \frac{\partial F}{\partial \mathbf{a}} &= V_{12}\mathbf{b} - \lambda V_{11}\mathbf{a} = 0, \\ \frac{\partial F}{\partial \mathbf{b}} &= V_{21}\mathbf{a} - \mu V_{22}\mathbf{b} = 0. \end{aligned} \quad (7.5.5)$$

Substituting for $\mathbf{b}$ gives

$$V_{12}V_{22}^{-1}V_{21}\mathbf{a} - \lambda\mu V_{11}\mathbf{a} = 0. \quad (7.5.6)$$

This equation has non-trivial roots only if

$$|V_{12}V_{22}^{-1}V_{21} - \rho V_{11}| = 0 \quad (7.5.7)$$

where $\rho = \lambda\mu$. Let the values of ρ , the roots of the determinantal equation (7.5.7), be denoted $\rho_1, \dots, \rho_q$, where some of the roots may be repeated, and let the vectors corresponding to these be $\mathbf{a}_1, \dots, \mathbf{a}_q$, since we can scale these vectors so that $\mathbf{a}_i' V_{11} \mathbf{a}_i = 1$.

It follows from (7.5.6) that

$$\mathbf{b}_i = (1/\mu) V_{22}^{-1} V_{21} \mathbf{a}_i,$$

for $i = 1, \dots, q$, and that, by the properties of eigenvectors of symmetric matrices,

$$\mathbf{a}_i' V_{11} \mathbf{a}_j = 0,$$

for all $i \neq j$. Again, from (7.5.6),

$$\mathbf{a}_i' V_{12} \mathbf{b}_j = \lambda \mathbf{a}_i' V_{11} \mathbf{a}_j = 0,$$

and, similarly, $\mathbf{b}_i' V_{22} \mathbf{b}_j = 0$, for $i \neq j$.

The vectors $P\mathbf{a}_1$ and $C\mathbf{b}_1$, corresponding to the largest eigenvalue ρ_1 are called the first canonical variates and, from the definition (7.5.4), ρ_1 represents the correlation between these two vectors. Subsequent roots will correspond to other pairs of vectors $\mathbf{a}$ and $\mathbf{b}$ which have the property that they maximise the correlation subject to $P\mathbf{a}$ and $C\mathbf{b}$ being orthogonal to each of the earlier canonical vectors as demonstrated above.

Suppose now that we choose P to have $(p+c-2)$ linearly independent columns and C have $(c-1)$ linearly independent columns, where in each case the columns are contrasts in the sense that $(I-J)P = 0, (I-J)C = 0$. Then the matrices V_{11} and V_{22} will be positive definite and the roots of $|V_{12}V_{22}^{-1}V_{21} - \rho V_{11}| = 0$ are non-negative. If we assume, without loss of generality, that $q \geq m$, there are exactly $m = (c-1)$ non-zero roots, which may well have repeated roots among them. The remaining $q - m = p - 1$ roots will be zero and correspond to a set of vectors $\mathbf{a}_i, i = c, \dots, p+c-2$ which span a part of the Cohorts space which is orthogonal to Periods.

The vectors $P\mathbf{a}_i$ and $C\mathbf{b}_i$ now each form a set of orthonormal vectors in the space of contrasts, and they provide a partition into orthogonal one degree of freedom contrasts of the Periods(ig. Cohorts) and Cohorts(ig. Periods) strata, respectively, fulfilling the conditions of the James and Wilkinson model. The only difference is that we have not collapsed repeated eigenvalues into subspaces of higher dimensionality. The projector matrices for these one dimensional spaces are then $P\mathbf{a}_i\mathbf{a}_i'P'$ and $C\mathbf{b}_i\mathbf{b}_i'C'$, respectively.

To obtain the vectors which span the Cohorts(elim. Periods) space, we need to look, for each i , at the linear combination of $P\mathbf{a}_i$ and $C\mathbf{b}_i$ which is orthogonal to Periods, i.e. to $P\mathbf{a}_i$. This can easily be shown to be the vector

$$\frac{1}{\sqrt{1 - \rho_i^2}}(C\mathbf{b}_i - \rho_i P\mathbf{a}_i),$$

which has been normalised to length one. This will be established more formally in the next section.

7.6 Expected mean squares of each block structure

We choose to use the Cohort (elim. Periods) to capture treatment information, so we need the expected mean square of Cohort (elim. Periods) stratum and we need this for each degree of freedom contrast. In this situation we refer to equation (7.4.1), and we know that the projector matrices, that is $S_i (i = 0, 1, 2, 3)$, will change as p and c change. We also know that

$$Var(y) = \sigma^2 I_{cp} + p\sigma_s^2(I_c \otimes J_p) \quad (7.6.1)$$

We have now a situation analogous to that of James and Wilkinson. In the space R^p , we have two subspaces $\mathcal{U}$, for Periods(ig. Cohorts), and $\mathcal{V}$, for Cohorts(ig. Periods), such that these two subspaces are not orthogonal to one another. As we know the respective projector matrices for these subspaces are S_1 and $S_0 = K_c \otimes J_p$, of rank $(p + c - 2)$ and $(c - 1)$ respectively such that $p + c - 2 > c - 1$ for $p > 1$. Based on the result of James and Wilkinson's paper it is possible to find projector matrices $U_1, U_2, \dots, U_{k+1}$ in Periods(ig. Cohorts) and $V_1, V_2, \dots, V_k$ in Cohorts(ig. Periods) satisfying the results in (7.5.2) and with ranks $r_1, r_2, \dots, r_k, r_{k+1}$, such that, in our case $\sum_{i=1}^k r_k = (c - 1)$, $r_{k+1} = p - 1$. We also can get the subspaces and the quantities ρ_i by orming the canonical correlations between sets of columns spanning the two subspaces. Let P be any set of linearly independent columns spanning $\mathcal{U}$ and C be any set of linearly independent columns spanning $\mathcal{V}$. The canonical correlations ρ_i can be obtained by finding $\mathbf{a}_i$ and $\mathbf{b}_i$ such that

$$\begin{aligned} \rho_1 &= \max_{\mathbf{a}_1, \mathbf{b}_1} Corr\{P\mathbf{a}_1, C\mathbf{b}_1\}, \\ \rho_2 &= \max_{\mathbf{a}_2, \mathbf{b}_2} Corr\{P\mathbf{a}_2, C\mathbf{b}_2\}, \\ &\vdots \end{aligned}$$

$$\rho_{c-1} = \max_{\mathbf{a}_{c-1}, \mathbf{b}_{c-1}} \text{Corr}\{P\mathbf{a}_{c-1}, C\mathbf{b}_{c-1}\},$$

subject to the orthogonality conditions $\mathbf{a}_i' \mathbf{1} = \mathbf{b}_i' \mathbf{1} = 0$, $\mathbf{a}_i' P' P' \mathbf{a}_j = \delta_{ij}$ and $\mathbf{b}_i' C' C \mathbf{b}_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta. There is also an additional set of vectors $\mathbf{a}_c, \dots, \mathbf{a}_{c+p-2}$, satisfying the same conditions.

If P and C are the Helmert contrasts on the Periods (ig. Cohorts) and Cohorts (ig. Periods) respectively in the general cohort design, and if we make each such column orthogonal to the vector of ones, we can obtain the canonical variables for Cohorts (ig. Period) and Periods (ig. Cohorts). We can then write

$$X_p = P \times [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{p+c-2}] \quad (7.6.2)$$

$$= [p_1, p_2, \dots, p_{p+c-2}],$$

$$X_c = C \times [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{c-1}]$$

$$= [c_1, c_2, \dots, c_{c-1}], \quad (7.6.3)$$

$$\rho = [\rho_1, \rho_2, \dots, \rho_{c-1}],$$

where, from the earlier results, $c_i' c_j = p_i' p_j = \delta_{ij}$, $c_i' p_j = 0$ for $i \neq j$, and $c_i' p_i = \rho_i$. Note that $\rho_1, \rho_2, \dots, \rho_{c-1}$ may include repeated values.

To obtain the projector matrix onto Cohort (elim. Periods), we can use the combined (Cohort, Periods) space and the projector matrix S_1 for Period (ig. Cohort). Now the projector matrix onto the (Cohort, Period) space, S_{12} , is given by

$$S_{12} = \begin{bmatrix} X_p & X_c \end{bmatrix} \begin{bmatrix} X_p' X_p & X_p' X_c \\ X_c' X_p & X_c' X_c \end{bmatrix}^{-1} \begin{bmatrix} X_p' \\ X_c' \end{bmatrix}.$$

Since $X_p' X_p = I_{p+c-2}$ and $X_c' X_c = I_{c-1}$ and if we put $X_c' X_p = [\text{diag}(\rho_1, \dots, \rho_{c-1}), \mathbf{0}_{(c-1) \times (p-1)}] = D$, say, then the projector matrix for Cohort (elim. Periods) is

$$S_2 = S_{12} - S_1 = \begin{bmatrix} X_p & X_c \end{bmatrix} \begin{bmatrix} I_{p+c-2} & D' \\ D & I_{c-1} \end{bmatrix}^{-1} \begin{bmatrix} X_p' \\ X_c' \end{bmatrix} - X_p X_p',$$

Now, using Rao (1973, p.29), we have

$$\begin{bmatrix} I & D' \\ D & I \end{bmatrix}^{-1} = \begin{bmatrix} I + D' E^{-1} D & -D' E^{-1} \\ -E^{-1} D & E^{-1} \end{bmatrix},$$

where

$$E = (I - DD')_{(c-1) \times (c-1)} = \text{diag} \left(1 - \rho_1^2, 1 - \rho_2^2, \dots, 1 - \rho_{c-1}^2 \right).$$

Then, it follows that

$$\begin{aligned} S_2 &= \begin{bmatrix} X_p & X_c \end{bmatrix} \begin{bmatrix} I + D'E^{-1}D & -D'E^{-1} \\ -E^{-1}D & E^{-1} \end{bmatrix} \begin{bmatrix} X'_p \\ X'_c \end{bmatrix} - X_p X'_p \\ &= \begin{bmatrix} X_p & X_c \end{bmatrix} \begin{bmatrix} D'E^{-1}D & -D'E^{-1} \\ -E^{-1}D & E^{-1} \end{bmatrix} \begin{bmatrix} X'_p \\ X'_c \end{bmatrix} \\ &= \begin{bmatrix} X_p & X_c \end{bmatrix} \begin{bmatrix} -D' \\ I \end{bmatrix} E^{-1} \begin{bmatrix} -D & I \end{bmatrix} \begin{bmatrix} X'_p \\ X'_c \end{bmatrix} \\ &= [X_c - X_p D'] E^{-1} [X'_c - D X'_p], \end{aligned}$$

where

$$X_p D' = \begin{bmatrix} p_1, & \dots, & p_{p+c-2} \end{bmatrix} \begin{bmatrix} \rho_1 & 0 & \dots & 0 \\ 0 & \rho_2 & \dots & 0 \\ 0 & \dots & \dots & \rho_{c-1} \\ 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & 0 \end{bmatrix} = \begin{bmatrix} \rho_1 p_1, & \dots, & \rho_{c-1} p_{c-1} \end{bmatrix}.$$

Thus

$$S_2 = \sum_{i=1}^{c-1} \frac{1}{1 - \rho_i^2} (c_i - \rho_i p_i)(c_i - \rho_i p_i)'.$$

We note that $(c_i - \rho_i p_i)' p_j = 0$, for all j , so that this space is indeed orthogonal to Periods. Then Cohort (elim. periods) has $(c - 1)$ degrees of freedom and they can be given individually by the rank 1 projector matrices

$$C_i = \frac{1}{1 - \rho_i^2} (c_i - \rho_i p_i)(c_i - \rho_i p_i)'.$$

The sum of squares for the i th of these contrasts is just $y' C_i y$. The expected value of this will typically include treatment components, so we obtain

$$\begin{aligned} E(y' C_i y) &= \text{tr} \{ C_i (\sigma^2 I_{cp} + p \sigma_s^2 I_c \otimes J_p) \} + \tau' T' C_i T \tau \\ &= \sigma^2 \text{tr}(C_i) + p \sigma_s^2 \text{tr} \{ C_i (I_c \otimes J_p) \} + \tau' T' C_i T \tau \\ &= \sigma^2 + p \sigma_s^2 \text{tr} \{ C_i (I_c \otimes J_p) \} + \tau' T' C_i T \tau. \end{aligned}$$

To further simplify this, we note that the orthogonal projector matrix for Cohort (ig. Periods) is

$$X_c(X'_c X_c)^{-1} X'_c = \sum_{i=1}^{c-1} c_i c'_i$$

and we also know that this orthogonal projector matrix can be given as

$$\{(I - J)_c \otimes J_p\} = I_c \otimes J_p - J_c \otimes J_p.$$

Therefore, it follows that

$$\sum_{i=1}^{c-1} c_i c'_i = I_c \otimes J_p - J_c \otimes J_p,$$

or

$$I_c \otimes J_p = \sum_{i=1}^{c-1} c_i c'_i + J_c \otimes J_p.$$

Since $C_i 1 = 0$,

$$tr\{C_i(I_c \otimes J_p)\} = tr\left[\frac{1}{1 - \rho_i^2}(c_i - \rho_i p_i)(c_i - \rho_i p_i)' \sum_{j=1}^{c-1} c_j c'_j\right].$$

Now on the basis that

$$c'_i c_j = \delta_{ij}, \quad p'_i c_j = \rho_i \delta_{ij},$$

where δ_{ij} is Kronecker delta given by

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases},$$

we have

$$tr\{C_i(I_c \otimes J_p)\} = \frac{1}{1 - \rho_i^2} \{(c_i - \rho_i p_i)' c_i\}^2 = (1 - \rho_i^2),$$

so that

$$E(y' C_i y) = \sigma^2 + p(1 - \rho_i^2) \sigma_s^2 + \tau' T' C_i T \tau.$$

Thus the expected mean square for each contrast may be different, according to the degree of correlation between that contrast and the appropriate contrast in the period stratum. Overall, the expected sum of squares in the Cohort (elim. Periods) stratum is

$$\begin{aligned} E(SS_{CeP}) &= \sum_{i=1}^{c-1} [\sigma^2 + p(1 - \rho_i^2) \sigma_s^2] + \sum_{i=1}^{c-1} \tau' T' C_i T \tau \\ &= (c - 1) \sigma^2 + p \sigma_s^2 \sum_{i=1}^{c-1} (1 - \rho_i^2) + \sum_{i=1}^{c-1} \tau' T' C_i T \tau \\ &= (c - 1) (\sigma^2 + p \sigma_s^2) - p \sigma_s^2 \sum_{i=1}^{c-1} \rho_i^2 + \sum_{i=1}^{c-1} \tau' T' C_i T \tau, \end{aligned}$$

For the expected sum of squares for the Period (ig. Cohort) stratum, we know that the projector matrix is $S_1 = \sum_{i=1}^{p+c-2} p_i p'_i$, so we can write

$$\begin{aligned}
E(SS_{PiC}) &= E[\text{tr}(S_1 y y')] \\
&= \text{tr}[S_1 \{\sigma^2 I_{cp} + p\sigma_s^2 (I_c \otimes J_p)\}] + \tau' T' S_1 T \tau + \pi' P' S_1 P \pi \\
&= \sigma^2 \text{tr}(S_1) + p\sigma_s^2 \text{tr}[S_1 (I_c \otimes J_p)] + \tau' T' S_1 T \tau + \pi' P' S_1 P \pi \\
&= (p + c - 2)\sigma^2 + p\sigma_s^2 \text{tr}[S_1 (I_c \otimes J_p)] + \tau' T' S_1 T \tau + \pi' P' S_1 P \pi,
\end{aligned}$$

since trace and rank of any idempotent matrix is same. It follows that the expected mean square has a term σ^2 but now has a component involving σ_s^2 , indicating that some of the cohort differences have ‘leaked’ into the period stratum.

Finally for Cohort (ig. Periods) we know that the projector matrix is the same as projector matrix for standard design that is, we have $S_0 = K_c \otimes J_p$. Then we can write

$$\begin{aligned}
E(SS_{CiP}) &= E(y' S_0 y) = E[\text{tr}(S_0 y y')] \\
&= \text{tr}[S_0 \{\sigma^2 I_c + p\sigma_s^2 (I_c \otimes J_p)\}] + \tau' T' S_0 T \tau + \pi' P' S_0 P \pi \\
&= \sigma^2 \text{tr}(S_0) + p\sigma_s^2 \text{tr}[S_0 (I_c \otimes J_p)] + \tau' T' S_0 T \tau + \pi' P' S_0 P \pi \\
&= (c - 1)\sigma^2 + p\sigma_s^2 \text{tr}[(K_c \otimes J_p)(I_c \otimes J_p)] + \tau' T' S_0 T \tau + \pi' P' S_0 P \pi \\
&= (c - 1)\sigma^2 + p\sigma_s^2 \text{tr}[(K_c \otimes J_p)] + \tau' T' S_0 T \tau + \pi' P' S_0 P \pi \\
&= (c - 1)(\sigma^2 + p\sigma_s^2) + \tau' T' S_0 T \tau + \pi' P' S_0 P \pi,
\end{aligned}$$

Since there are $(c - 1)$ degrees of freedom, the mean square for Cohort (ig. Periods) has expected value $\sigma^2 + p\sigma_s^2$, plus whatever treatment and period effects are present. Any contrast of length 1 in the Cohort (ig. Periods) stratum has variance $\sigma^2 + p\sigma_s^2$, but cannot generally be used for estimation since it contains the (unknown) period effects.

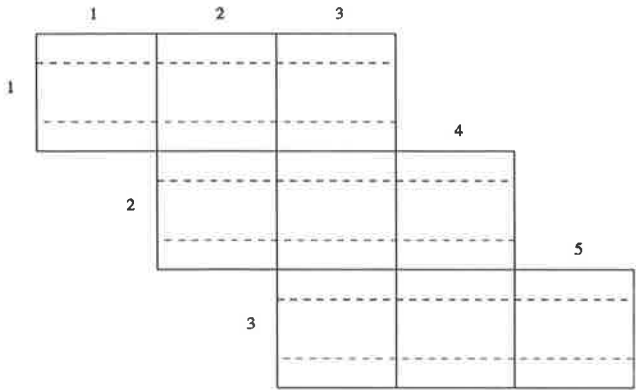
7.7 Description of a function `cc` in S-PLUS to get the result of James and Wilkinson’s theorem

To see the result of applying James and Wilkinson’s theorem, we use a function `cc` written in the S-PLUS program and listed in Appendix B. The function requires as input the value of p , the number of measurements on each subject, and c , the number of cohorts. Columns P and C are formed giving the levels of periods and cohorts corresponding to

the vector of cp means y written in standard order. From these, design matrices X_p and X_c are created. In the second stage, for any cohort design, the Helmert contrasts for $p + c - 2$ and $c - 1$ degrees of freedom for both the Period(ig. Cohorts) and Cohorts(ig. Periods) strata respectively are created. Initially, due to the unequal replication of the periods, the Helmert contrasts are not orthogonal to the grand mean. The use of the function `cc` produces columns which are true contrasts. In the third stage, the function `cc` gives us the two sets of linearly independent columns labelled P_c and C_c spanning the spaces due to periods and cohorts respectively and satisfying the James and Wilkinson theorem.

To show these stages we refer to the special case of a cohort design with $c = 3, p = 3$, which is shown in Figure 7.3 and for which observations are made over a total of 5 periods.

Figure 7.3: Cohort design for three periods in each cohort



As expected in this case, we have four Helmert contrasts on the period stratum, say $P_1, \dots, P_4$, and two Helmert contrasts on the cohort stratum, say C_1 and C_2 . These are shown in Tables 7.3 and 7.4 respectively.

P_1	-1	1	0		
		1	0	0	
			0	0	0

P_2	-1	-1	2		
		-1	2	0	
			2	0	0
P_4	-1	-1	-1		
		-1	-1	-1	
			-1	-1	4

P_3	-1	-1	-1		
		-1	-1	3	
			-1	3	0

Table 7.3: Helmert contrasts for Periods

C_1	-1	-1	-1		
		1	1	1	
			0	0	0

C_2	-1	-1	-1		
		-1	-1	-1	
			2	2	2

Table 7.4: Helmert contrasts for Cohorts

The Helmert contrasts given here for Periods are not, as they stand, orthogonal to the grand mean and hence an additional step is required to make this so. Now in the next stage the function `cc` gives us orthogonal contrasts for Cohort (ig. Periods) and Periods (ig. Cohort) and since these contrasts are not orthogonal to each other we use the result of James and Wilkinson's paper. Function `cc` considers C as any linear independent columns spanning the space due to Cohort (ig. Periods) stratum by vectors of C_1 , C_2 and P also as any linear independent columns spanning the space due to Periods (ig. Cohort) stratum by vectors of $P_1, \dots, P_4$. Now, we consider a new stratum named Cohort (elim. Periods) which is orthogonal to the Periods (ig. Cohort) stratum. To reach this aim function `cc` gives us the canonical correlation coefficients and a set of vectors **a** and **b**. For example, the vectors of **a** and **b** and the corresponding canonical correlation coefficients, ρ , associated with them are in this case:

```

$a:           $b:           $rho:
      [,1]  [,2]           [,1]  [,2]  [,3]  [,4]  [1] 0.707 0.408
[1,] 0.204 -0.354  [1,] 0.144 -0.433 -0.401  0.077
[2,] 0.204  0.118  [2,] 0.144 -0.048  0.045  0.231
                        [3,] 0.144 -0.096  0.089 -0.077
                        [4,] 0.144  0.115 -0.107 -0.015

```

When these columns are applied to the matrices X_c and X_p , respectively, we obtain the contrasts (c_i, p_i) as in James and Wilkinson theorem. These span the Cohort (ig. Periods) and Period (ig. Cohorts) strata respectively. For the Cohorts (elim. Periods) stratum, it is necessary to calculate $(c_i - \rho_i p_i)$. All these contrasts are shown in Table 7.5, although they are not shown standardised to length 1.

c	p	$1 - \rho_i^2$	Cohort(ig. Periods)	Periods(ig. Cohort)	Cohorts (elim. Periods)
3	3	1/2	$\begin{pmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}$	$\begin{pmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{pmatrix}$
		5/6	$\begin{pmatrix} 1 & 1 & 1 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 2 & -1 & 0 \\ -1 & 0 & -1 \\ 0 & -1 & 2 \end{pmatrix}$	$\begin{pmatrix} 0 & 3 & 2 \\ -3 & -4 & -3 \\ 2 & 3 & 0 \end{pmatrix}$
				$\begin{pmatrix} 4 & -5 & 1 \\ -5 & 1 & 4 \\ 1 & 4 & -5 \end{pmatrix}$	
				$\begin{pmatrix} -2 & -1 & 3 \\ -1 & 3 & -2 \\ 3 & -2 & -1 \end{pmatrix}$	

Table 7.5: Contrasts for three cohorts and three periods

For convenience, we write the Cohort $\times$ Periods array as a rectangle in which cohorts are represented by rows and successive periods are represented by successive reverse diagonals of the array; that is, the sets of value running along diagonals from top right to bottom left. It can then be seen that the period contrasts are in fact contrasts between periods, while the Cohort (elim. Periods) contrasts are orthogonal to periods.

7.7.1 Projector matrices

For this design we have

- $S_0 = K_3 \otimes J_3$ for Cohort (ig. Periods)
- $S_1 = \sum_{i=1}^4 p_i p_i'$ for Periods (ig. Cohort)
- $S_2 = \sum_{i=1}^2 \frac{1}{1-\rho_i^2} (c_i - \rho_i)(c_i - \rho_i p_i)'$ for Cohort (elim. Periods)
- $S_3 = (I_3 \otimes I_3) - (J_3 \otimes J_3) - (S_1 + S_2)$ for Period $\times$ Cohort interaction,

where it is assumed that both c_i and p_i are standardized to length 1.

7.8 To get a pattern on ρ and the Cohort (elim. Periods) contrasts

To get a pattern on ρ and the contrasts for the Cohort (elim. Periods) stratum we have written the results in Table 7.6 for $c = 2$ and Table 7.7 for $c = 3$. For $c = 2$, there is one contrast for Cohorts (elim. Periods), and it is obtained by deleting the values in the first and last periods and taking the average difference between the other means. This has $(1 - \rho^2) = (p - 1)/p$, so that the expected mean square for this contrast is $\sigma^2 + (p - 1)\sigma_s^2$, plus whatever treatment terms it estimates.

c	p	$1 - \rho_1^2$	Cohort(elim. Periods)									
2	2	0.500		0	-1							
					1	0						
2	3	0.667		0	-1	-1						
					1	1	0					
2	4	0.750		0	-1	-1	-1					
					1	1	1	0				
2	5	0.800		0	-1	-1	-1	-1				
					1	1	1	1	0			
2	6	0.833		0	-1	-1	-1	-1	-1			
					1	1	1	1	1	0		
2	7	0.857		0	-1	-1	-1	-1	-1	-1		
					1	1	1	1	1	1	0	
2	8	0.875		0	-1	-1	-1	-1	-1	-1	-1	
					1	1	1	1	1	1	1	0

Table 7.6: Canonical correlation coefficients and orthogonal contrasts for two cohorts and various periods

When $c = 3$, there are two such contrasts, with $(1 - \rho_i^2)$ equal to $1 - 3/(2p)$ and $1 - 1/(2p)$. The two contrasts have a consistent pattern, representing, respectively, a linear and quadratic contrast of the cohorts through the middle $p - 2$ periods, with slight adjustments at the ends, and having expected mean squares of $\sigma^2 + (p - \frac{3}{2})\sigma_s^2$ and $\sigma^2 + (p - \frac{1}{2})\sigma_s^2$, respectively.

c	p	$1 - \rho_1^2$	$1 - \rho_2^2$	Cohort(elim. Periods)						Cohort(elim. Periods)						
3	2	0.250	0.750	0	-1					0	1					
					1	-1				-1	-1					
						1	0				1	0				
3	3	0.500	0.833	0	-1	-2				0	3	2				
					1	0	-1			-3	-4	-3				
						2	1	0			2	3	0			
3	4	0.625	0.875	0	-1	-2	-2			0	3	2	2			
					1	0	0	-1			-3	-4	-4	-3		
						2	2	1	0			2	2	3	0	
3	5	0.700	0.900	0	-1	-2	-2	-2		0	3	2	2	2		
					1	0	0	0	-1		-3	-4	-4	-4	-3	
						2	2	2	1	0		2	2	2	3	0
3	6	0.750	0.917	0	-1	-2	-2	-2	-2	0	3	2	2	2	2	
					1	0	0	0	0	-1		-3	-4	-4	-4	-3
						2	2	2	2	1	0		2	2	2	2
3	7	0.786	0.929	0	-1	-2	-2	-2	-2	0	3	2	2	2	2	
					1	0	0	0	0	-1		-3	-4	-4	-4	-4
						2	2	2	2	1	0		2	2	2	2

Table 7.7: Canonical correlation coefficients and orthogonal contrasts for three cohorts and various periods

7.9 Conclusions

The above analysis provides a set of contrasts among the Cohorts, orthogonal to Periods, which will enable estimation of treatment effects. Each contrast will have a different expected mean square, so that the treatment information from each normalised contrast will need to be combined with a different weight. Furthermore treatment information will be available in the Cohort $\times$ Period interaction, and each normalised contrast in that space will have variance σ^2 . In the next chapter, we consider some particular designs and demonstrate how much information is available and how it can be recovered.

Chapter 8

Treatment structure of cohort designs

8.1 Introduction

Having examined the block structure of cohort designs in the previous chapter, we now turn our attention to treatment structure. We consider the specification of linear treatment models when treatments are applied to the experimental subjects, and we get formulae for the analysis of cohort designs which have a linear treatment model and a non-orthogonal block structure. To extend our approach to analyse crossover trials efficiently as we did in Chapter 6, we apply the cohort design methodology to a number of different crossover trials. In Section 8.2, we describe how to obtain treatment estimates from cohort designs, and in Section 8.3 we extend the results of Chapter 7 to the case where there are g groups of subjects in each of the cohorts, the idea being that each group within a cohort is given a different treatment regime. Subsequent sections then consider a number of possible designs and compare the results with those obtained earlier in this thesis for the corresponding standard design.

8.2 Treatment structure for cohort and standard designs in general

To show the efficiency of the cohort design and the results of James and Wilkinson's theorem in a nonorthogonal block structure, we estimate the treatment effects in a cohort design and then compare it to a corresponding standard design. To refer to previous chapter we now consider the general cohort and standard designs with length $cpgn$, for both designs, where there are c cohorts, g groups within each cohort, and each of the n subjects within each (cohort, group) combination is observed for p periods. Since the n subjects within a group all receive the same treatment regime, we can consider just the vector of length cpg which consists of the group means at each time. In terms of seeing where treatment information resides, we can without further loss of generality consider the case where $n = 1$.

We assume that the design has s treatment parameters and the linear model for treatment and grand mean for both designs is expressed in Chapter 6 in equation (6.3.1). Suppose $\theta_0 = \begin{bmatrix} \mu & \theta' \end{bmatrix}'$ represents the parameters other than the group and period parameters, and suppose that the vector of treatment effects θ has s elements. Let T_0 represent the corresponding design matrix and suppose that its $(s + 1)$ columns are linearly independent.

The vector θ might include τ_0 , τ_D , λ_0 , and λ_D , which we introduced in Chapter 6. However, we want to consider at this stage a general set of treatments which might even include factorial contrasts, for example. To consider what treatment information is available, we take the matrix T_0 for treatments in a model with length cpg . Then the information matrix overall is $T_0' T_0$ when the grand mean is considered. Some of the columns of T_0 corresponding to particular parameters may be orthogonal to the $\mathbf{1}$ vector. Since the estimate of the grand mean μ does not provide any useful information about treatments, the only usable information is contained in the bottom right-hand $s \times s$ submatrix of $T_0'(I - J)T_0$. If we let $T_0 = \begin{bmatrix} \mathbf{1} & T_0^* \end{bmatrix}$, and define

$$T = (I - J)T_0^*,$$

then the required information matrix for θ is $T' T$.

If we split the observation vector into several strata, using orthogonal idempotent

matrices $Q_0, Q_1, \dots, Q_k$ where $\sum_{i=0}^k Q_i = I$, then we can write

$$y = (Q_0 + Q_1 + \dots + Q_k)y,$$

where $Q_0y, Q_1y, \dots, Q_ky$ are components of the variation among different strata, and the treatment information for the i th stratum is $T'Q_iT$. In the case where $Var(\epsilon) = \sigma^2 I$,

$$\hat{\theta} = (T'T)^{-1}T'y,$$

but if each stratum has variance matrix $\delta_i Q_i$ then the combined estimate is

$$\hat{\theta} = \left\{ \sum_i \delta_i^{-1} T' Q_i T \right\}^{-1} \sum_i \delta_i^{-1} T' Q_i y$$

and

$$Var(\hat{\theta}) = \left\{ \sum_i \delta_i^{-1} T' Q_i T \right\}^{-1},$$

where, as in previous chapters, the summation is made over those values of i for which we believe we have useful information. As before, we would exclude components such as the grand mean ($i = 0$), and strata involving period effects since they can not generally be well estimated or a variance component determined. In the case of a cohort design, we would generally not try to use information from Period(ig Cohort) stratum, but the information from the Cohorts(elim. Period) stratum would be used.

8.3 Projector matrices for cohort and standard designs in general

We know that the projector matrices for the standard design, in the case $n = 1$, are as shown in Table 8.1.

And for the cohort design we have the projector matrices shown in Table 8.2 where S_i are $cp \times cp$ matrices defined in the previous chapter.

Stratum	Projector matrices for Standard Design
Periods	$J_c \otimes K_p \otimes J_g$
Cohorts	$K_c \otimes J_p \otimes J_g$
Groups within Cohort	$I_c \otimes J_p \otimes K_g$
Period $\times$ Cohort	$K_c \otimes K_p \otimes J_g$
Periods $\times$ (Groups within Cohort)	$I_c \otimes K_p \otimes K_g$

Table 8.1: Projector matrices for general standard design

Stratum	Projector matrices for Cohort Design
Periods (ig. Cohorts)	$S_1 \otimes J_g$
Cohorts (elim. Periods)	$S_2 \otimes J_g$
Groups within Cohort	$I_c \otimes J_p \otimes K_g$
Periods $\times$ Cohort	$S_3 \otimes J_g$
Periods $\times$ (Groups within Cohorts)	$I_c \otimes K_p \otimes K_g$

Table 8.2: Projector matrices for general cohort design

We note that in Cohort (elim. Periods) in the general case the information is contained in $(c - 1)$ separate 1-dimensional subspaces and is given by

$$T'(C_i \otimes J_g)T \quad (i = 1, \dots, c - 1)$$

where C_i can be interpreted as $d_i d'_i$ such that d_i is a contrast of length 1. Then the projector matrix for d_i is

$$d_i(d'_i d_i)^{-1} d'_i = d_i d'_i,$$

where

$$d_i = \frac{1}{\sqrt{1 - \rho_i^2}}(c_i - \rho_i p_i),$$

and c_i, ρ_i and p_i are as defined in Chapter 7.

$$\begin{aligned}
T'(C_i \otimes J_g)T &= T'(d_i d'_i \otimes J_g)T \\
&= \frac{1}{g} T'(d_i d'_i \otimes \mathbf{1} \mathbf{1}')T \\
&= \frac{1}{g} T'(d_i \otimes \mathbf{1})(d_i \otimes \mathbf{1})'T \\
&= \frac{1}{g} w_i w'_i.
\end{aligned} \tag{8.3.1}$$

It follows that the information in the i th 1-dimensional subspace of Cohort (elim. Periods) is given by $\frac{1}{g}w_i w'_i$, where $w_i = T'(d_i \otimes \mathbf{1})$. Correspondingly, the projector matrix for Periods (ig. Cohorts) stratum is $\sum_{i=1}^{p+c-2} p_i p'_i$, and since the length of p_i 's are 1 then the information matrix for this stratum can be written as

$$\begin{aligned} T'\{S_1 \otimes J_g\}T &= T'\left\{\sum_{i=1}^{p+c-2} p_i p'_i \otimes J_g\right\}T \\ &= \frac{1}{g}T'\left\{\sum_{i=1}^{p+c-2} (p_i \otimes \mathbf{1})(p_i \otimes \mathbf{1})'\right\}T \\ &= \frac{1}{g}\sum_{i=1}^{p+c-2} q_i q'_i, \end{aligned} \quad (8.3.2)$$

where $q_i = T'(p_i \otimes \mathbf{1})$.

If we have $n > 1$ subjects in each group, then the ANOVA's will have two more lines, corresponding to Subjects within Groups and Periods $\times$ (Subjects within Groups) strata. As before, these will contain no treatment information, and will provide estimates of σ^2 and $\sigma^2 + p\sigma_s^2$ respectively.

8.3.1 Estimates of parameters and covariance matrix of estimates

To get the estimates of parameters we combine the estimates of parameters from those strata in which there are some information about parameters. Thus we can write

$$\begin{aligned} \hat{\theta} &= \left\{ \sum_{j=1}^{c-1} \left[\frac{1}{\eta_j} T'(C_j \otimes J_g)T \right] + \sum_{i=3}^5 \left[\frac{1}{\delta_i} T'(Q_i \otimes J_g)T \right] \right\}^{-1} \\ &\quad \left\{ \sum_{j=1}^{c-1} \left[\frac{1}{\eta_j} T'(C_j \otimes J_g) \right] + \sum_{i=3}^5 \left[\frac{1}{\delta_i} T'(Q_i \otimes J_g) \right] \right\} y, \end{aligned} \quad (8.3.3)$$

and covariance matrix of $\hat{\theta}$ is

$$Var(\hat{\theta}) = \left\{ \sum_{j=1}^{c-1} \left[\frac{1}{\eta_j} T'(C_j \otimes J_g)T \right] + \sum_{i=3}^5 \left[\frac{1}{\delta_i} T'(Q_i \otimes J_g)T \right] \right\}^{-1} \quad (8.3.4)$$

where $\eta_j = \sigma^2 + p\sigma_s^2(1 - \rho_j^2)$ ($j = 1, 2, \dots, c-1$) and $\delta_3 = \delta_5 = \sigma^2$ and $\delta_4 = \sigma^2 + p\sigma_s^2$.

In the rest of this chapter we examine the result of James and Wilkinson theorem in some special cases of standard and cohort designs.

8.4 Treatment information for two-treatment crossover design with two baseline measurements and a corresponding cohort design

We now examine the treatment information for the two-treatment, crossover with two baseline measurements, one before the first treatment period and one after this period. As Kenward & Jones (1987b) concluded, one reason to use two baseline measurements in the 2×2 crossover designs is to increase the power of test for the presence of carryover effects by splitting the sum of squares between-subjects into various components which can be used to analyse these designs. Freeman (1989) gives us another reason to take the baseline measurements in the 2×2 crossover design. He and others have suggested that use of baseline measurements, leads to a unique resolution of the estimation difficulties which we face in the simple crossover designs. In this section, we consider the consequence of the use of baseline measurements and in the next subsection we look at what happens if we withdraw the carryover effect in this design. These extra periods give the design which is shown in Table 8.3.

	Period			
	1	2	3	4
Group 1	—	A	—	B
2	—	B	—	A

Table 8.3: Two-treatment crossover design with two baseline measurements

Now in this study we consider a cohort design with $c = 2$, $p = 4$ and $g = 2$ but $n = 1$, then the layout of design is shown in Table 8.4 with a delayed entry at period 2 for the second cohort. Subjects within a cohort would be randomly allocated in equal-sized groups to the treatment regimes.

		Period				
		1	2	3	4	5
Group	1	—	A	—	B	
	2	—	B	—	A	
	3		—	A	—	B
	4		—	B	—	A

Table 8.4: Cohort design in two-treatment, two-period crossover design with two baseline measurements.

By referring to Table 8.4, we first write out the 16 expected values in the following table which could be applied to either the standard or the cohort design. Using the notation of Chapter 5, the expected values in each cell are as shown in Table 8.5.

μ	$\mu + \tau_0 + \tau_D$	$\mu + \lambda_0 + \lambda_D$	$\mu + \tau_0 - \tau_D$
μ	$\mu + \tau_0 - \tau_D$	$\mu + \lambda_0 - \lambda_D$	$\mu + \tau_0 + \tau_D$
μ	$\mu + \tau_0 + \tau_D$	$\mu + \lambda_0 + \lambda_D$	$\mu + \tau_0 - \tau_D$
μ	$\mu + \tau_0 - \tau_D$	$\mu + \lambda_0 - \lambda_D$	$\mu + \tau_0 + \tau_D$

Table 8.5: Expected values of two-treatment, two-period crossover with two baseline measurements when we consider the cohort design or double standard design.

Our purpose in the two following subsections is to get the treatment information for this case. We now first consider the model which includes the first-order carryover effect and in the next subsection we analysis the model without the first-order carryover effect.

8.4.1 Treatment information when the first-order carryover is present

If we let y be a vector of observations corresponding to the 16 expected values written in order across the rows, then by referring to the linear model expressed in 8.5 we have

$$T_0 = 1_2 \otimes \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & -1 \\ 1 & 1 & 0 & -1 & 0 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix},$$

where 1_2 arises from the fact that there are two cohorts, both using the same treatment structure and $\theta_0 = \left(\mu, \tau_0, \lambda_0, \tau_D, \lambda_D \right)'$. The matrix T , obtained by eliminating the effect of the $\mathbf{1}$ vector from the columns of T_0 is then given by

$$T = \frac{1}{8} 1_2 \otimes \begin{bmatrix} -4 & -2 & 0 & 0 \\ -4 & -2 & 0 & 0 \\ 4 & -2 & 8 & 0 \\ 4 & -2 & -8 & 0 \\ -4 & 6 & 0 & 8 \\ -4 & 6 & 0 & -8 \\ 4 & -2 & -8 & 0 \\ 4 & -2 & 8 & 0 \end{bmatrix}. \quad (8.4.1)$$

so that the overall treatment information is

$$T'T = \begin{bmatrix} 4 & -2 & 0 & 0 \\ -2 & 3 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix},$$

Table 8.6 shows the information matrices for the different strata in each design obtained by calculating $T'Q_iT$ for the appropriate idempotent matrices Q_i . In the case of the standard design, we use the Kronecker product matrices shown in Table 8.1. For the

cohort design, we refer to Table 7.6 in the previous chapter, which shows that when $c = 2$ and $p = 4$, we have one degree of freedom for Cohort (elim. Periods) with $\rho_1 = 0.50$, and d_1 given by the contrast in Table 8.7.

Stratum	Standard Design			Cohort Design			
	σ^2	σ_s^2	Information Matrix	ρ	σ^2	σ_s^2	Information Matrix
Periods(ig. Cohorts)			$\begin{bmatrix} 4 & -2 & 0 & 0 \\ -2 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$				$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Cohorts(elim. Periods)	1	4	None	0.50	1	3	$\frac{1}{3} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Group within Cohorts	1	4	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$		1	4	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Periods $\times$ Cohorts	1	0	None		1	0	$\frac{2}{3} \begin{bmatrix} 4 & -3 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Periods $\times$ (Grp.w.Coh.)	1	0	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$		1	0	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$
Total			$\begin{bmatrix} 4 & -2 & 0 & 0 \\ -2 & 3 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$				$\begin{bmatrix} 4 & -2 & 0 & 0 \\ -2 & 3 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$

Table 8.6: Information matrices for standard and cohort designs in two-treatment design with two baseline measurements, when first-order carryover effect is present

0	-1	-1	-1	
0	-1	-1	-1	
	1	1	1	0
	1	1	1	0

Table 8.7: Orthogonal contrast on Cohort (elim. Periods)

In comparing the results in Table 8.6, in terms of treatment information in each design and in each stratum, the following points can be noted:

- The treatment information in Group within Cohorts and Periods $\times$ (Grp.w.Coh.) strata for both designs is the same, because the projector matrix in each stratum is the same for both designs.
- In the standard design, all information about (τ_0, λ_0) is in the Periods stratum, so that the possible presence of period effects would imply that we cannot estimate (τ_0, λ_0) with this design.
- In the cohort design, most of the information about (τ_0, λ_0) is available in Period $\times$ Cohort stratum with a variance of σ^2 and some information (11%) about τ_0 is in the Cohorts (elim. Periods) stratum with a variance of $\sigma^2 + 3\sigma_s^2$.

8.4.2 Treatment information when carryover effect is not present

Now we consider this crossover design when there is no carryover effect in the model. We refer to Table 8.5 and do not consider the two parameter λ_0 and λ_D . Then the total treatment information for direct treatment only can be written as

$$T'T = \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix}$$

The treatment information in five different strata for direct treatment effect for standard and the cohort designs is given in Table 8.8.

Stratum	Standard Design			Cohort Design			
	σ^2	σ_s^2	Information Matrix	ρ	σ^2	σ_s^2	Information Matrix
Periods(ig. Cohorts)			$\begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}$				$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
Cohorts(elim. Periods)	1	4	None	0.50	1	3	$\frac{1}{3} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
Group within Cohorts	1	4	None		1	4	None
Periods $\times$ Cohorts	1	0	None		1	0	$\frac{1}{3} \begin{bmatrix} 8 & 0 \\ 0 & 0 \end{bmatrix}$
Periods $\times$ (Grp.w. Coh.)	1	0	$\begin{bmatrix} 0 & 0 \\ 0 & 8 \end{bmatrix}$		1	0	$\begin{bmatrix} 0 & 0 \\ 0 & 8 \end{bmatrix}$
Total			$\begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix}$				$\begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix}$

Table 8.8: Information matrices for two-treatment, two-period standard and cohort designs when first-order carryover effect is not present.

Like the situation in the previous subsection with this same design but with first-order carryover effect, cohort design is more efficient than the standard design. The parameter τ_0 is estimable in all strata except the Groups within Cohort stratum and for τ_D , the difference between direct treatments, Periods $\times$ (Groups within Cohorts) in both designs has all the information.

8.5 Treatment information for two-treatment, extra-period crossover design and its corresponding cohort design with one baseline measurement

In this section we consider the extra-period crossover design with two treatments and one extra period for baseline measurement. This design with three treatment periods with two groups is known as a dual balanced design, because this design is made up of equally replicated pairs of dual groups. The reason to choose this design is that the additional treatment period decreases many of the problems associated with analysis of

two-period crossover design. For example, as Ebbutt (1984) has shown, an appropriate extra-period design allows us to use all the data to estimate and test direct treatment effects when first order carryover effects are present. Kershner & Federer (1981) discuss the estimation of contrasts for direct and carryover treatment effects for a wide variety of two-treatment crossover designs with two, three and four periods. They also compare the variances of these contrasts for many alternative designs and evaluate the impact of including baseline measurements as suggested by Wallenstein (1979). Hafner & Koch (1988) give some alternatives to the analysis of extra-period crossover designs with two treatments. In their paper for these designs, they have emphasised the formulation of appropriate within-subject linear functions to analyse the design with parametric and non-parametric methods. The reason to choose this linear function in non-parametric analysis is that one only needs the independence and common distribution assumptions on the resulting within-subject linear functions. In the parametric approach one also requires normality assumptions on the resulting within-subject linear function for small sample cases. Now in this section we attempt to compare the treatment information for both the standard and the cohort designs for this crossover trial with one baseline measurement. This design with one baseline measurement is given in Table 8.9.

		Period			
		1	2	3	4
Group	1	—	A	B	B
	2	—	B	A	A

Table 8.9: Two-treatment, extra-period crossover design with one baseline measurement

The Cohort design corresponding to this standard design is shown in Table 8.10.

		Period				
		1	2	3	4	5
Group	1	—	A	B	B	
	2	—	B	A	A	
	3		—	A	B	B
	4		—	B	A	A

Table 8.10: Cohort design for two-treatment, extra-period crossover design with one baseline measurement.

The expected values of the design for this model are given in Table 8.11

μ	$\mu + \tau_0 + \tau_D$	$\mu + \tau_0 + \lambda_0 - \tau_D + \lambda_D$	$\mu + \tau_0 + \lambda_0 - \tau_D - \lambda_D$
μ	$\mu + \tau_0 - \tau_D$	$\mu + \tau_0 + \lambda_0 + \tau_D - \lambda_D$	$\mu + \tau_0 + \lambda_0 + \tau_D + \lambda_D$
μ	$\mu + \tau_0 + \tau_D$	$\mu + \tau_0 + \lambda_0 - \tau_D + \lambda_D$	$\mu + \tau_0 + \lambda_0 - \tau_D - \lambda_D$
μ	$\mu + \tau_0 - \tau_D$	$\mu + \tau_0 + \lambda_0 + \tau_D - \lambda_D$	$\mu + \tau_0 + \lambda_0 + \tau_D + \lambda_D$

Table 8.11: Expected values for design of two-treatment, extra-period crossover trial.

Now, as in the previous section, we deal with this design in two subsections in terms of presence and absence of first-order carryover effect in the model.

8.5.1 Treatment information when first-order carryover effect is present

The total treatment information for both the standard and the cohort designs after taking out the vector 1 due to the grand mean can be obtained as

$$T'T = \begin{bmatrix} 3 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$$

The output from SPLUS and Maple software provides the information matrices for various strata for both designs as given in Table 8.12. Treatment information about the differences τ_D and λ_D remain the same for the two designs, with a small amount of information in the Groups within Cohorts stratum. The situation, however, is different from the treatment information which looks at how the average of the two treatments differs from baseline. In the standard design, this is totally lost to the Period stratum. For the cohort design, not much is recovered, with only a third of the information about τ_0 and a quarter of the information about λ_0 being recovered, the rest being lost into the Period stratum.

The Cohort(elim. Periods) stratum has the information about (τ_0, λ_0) , and this is contained in the one degree of freedom contrast

0	-1	-1	-1	
0	-1	-1	-1	
	1	1	1	0
	1	1	1	0

which in this case has expectation

$$2(2\tau_0 + \lambda_0) - 2(3\tau_0 + 2\lambda_0) = -2(\tau_0 + \lambda_0),$$

and variance $12(\sigma^2 + 3\sigma_s^2)$. The Period $\times$ Cohort stratum has just 2 degrees of freedom.

These may be represented by

0	-1	2	-1	
0	-1	2	-1	
	1	-2	1	0
	1	-2	1	0

,

0	-1	0	1	
0	-1	0	1	
	1	0	-1	0
	1	0	-1	0

which have expectations of $2(2\lambda_0 - \tau_0)$ and $-2\tau_0$, and variances of $24\sigma^2$ and $8\sigma^2$, respectively. It follows that the information matrix for these two contrasts is given by

$$\frac{1}{6} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \begin{bmatrix} -1 & 2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix},$$

as given in Table 8.12.

Stratum	Standard Design			Cohort Design			
	σ^2	σ_s^2	Information Matrix	ρ	σ^2	σ_s^2	Information Matrix
Periods(ig. Cohorts)			$\begin{bmatrix} 3 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$				$\begin{bmatrix} 2 & 2 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Cohorts(elim. Periods)	1	4	None	0.50	1	3	$\frac{1}{3}$ $\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Group within Cohorts	1	4	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$		1	4	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Periods $\times$ Cohorts	1	0	None		1	0	$\frac{1}{3}$ $\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Periods $\times$ (Grp.w. Coh.)	1	0	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 11 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$		1	0	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 11 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$
Total			$\begin{bmatrix} 3 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$				$\begin{bmatrix} 3 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$

Table 8.12: Information matrices for standard and cohort two-treatment, extra-period designs with one baseline measurement.

This information can be combined to form treatment estimates by weighting these contrasts by their information matrices, i.e.

$$\begin{bmatrix} \hat{\tau}_0 \\ \hat{\lambda}_0 \end{bmatrix} = \left\{ \frac{1}{3} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \frac{1+3\rho}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \right\}^{-1}$$

$$\times \left\{ \frac{1}{3} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \hat{\tau}_{01} \\ \hat{\lambda}_{01} \end{bmatrix} + \frac{1+3\rho}{6} \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} \hat{\tau}_{02} \\ \hat{\lambda}_{02} \end{bmatrix} + \frac{1+3\rho}{2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\tau}_{03} \\ \hat{\lambda}_{03} \end{bmatrix} \right\},$$

in which $\hat{\tau}_{01}$, $\hat{\tau}_{02}$, $\hat{\tau}_{03}$ represent estimates obtained from the three contrasts, one from Cohorts(elim. Periods) and the other two from Periods $\times$ Cohorts. Now we can write

$$\begin{aligned} \begin{bmatrix} \hat{\tau}_0 \\ \hat{\lambda}_0 \end{bmatrix} &= \begin{bmatrix} 3+6\rho & -3\rho \\ -3\rho & 3+6\rho \end{bmatrix}^{-1} \\ &\times \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} (\hat{\tau}_{01} + \hat{\lambda}_{01}) + \frac{1+3\rho}{2} \begin{bmatrix} 1 \\ -2 \end{bmatrix} (\hat{\tau}_{02} - 2\hat{\lambda}_{02}) + \frac{3(1+3\rho)}{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \hat{\tau}_{03} \right\} \\ &= \frac{1}{3(1+\rho)(1+3\rho)} \begin{bmatrix} 1+\rho & \rho \\ \rho & 1+2\rho \end{bmatrix} \\ &\times \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} (\hat{\tau}_{01} + \hat{\lambda}_{01}) + \frac{1+3\rho}{2} \begin{bmatrix} 1 \\ -2 \end{bmatrix} (\hat{\tau}_{02} - 2\hat{\lambda}_{02}) + \frac{3(1+3\rho)}{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \hat{\tau}_{03} \right\} \\ &= \frac{1}{3(1+\rho)} \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} (\hat{\tau}_{01} + \hat{\lambda}_{01}) + \frac{1}{2} \begin{bmatrix} 1 \\ -2-3\rho \end{bmatrix} (\hat{\tau}_{02} - 2\hat{\lambda}_{02}) + \frac{3}{2} \begin{bmatrix} 1+2\rho \\ \rho \end{bmatrix} \hat{\tau}_{03} \right\}. \end{aligned}$$

Then the estimate of τ_0 can be obtained as the following linear combination of the 16 treatment means:

$$\hat{\tau}_0 = \frac{1}{2(1+\rho)} \begin{array}{|c|c|c|c|c|} \hline 0 & 1+\rho & 0 & -\rho & \\ \hline 0 & 1+\rho & 0 & -\rho & \\ \hline & -(1+\rho) & 0 & \rho & 0 \\ \hline & -(1+\rho) & 0 & \rho & 0 \\ \hline \end{array}$$

and the estimate of λ_0 is similarly given by

$$\hat{\lambda}_0 = \frac{1}{2(1+\rho)} \begin{array}{|c|c|c|c|c|} \hline 0 & 0 & (1+\rho) & -\rho & \\ \hline 0 & 0 & (1+\rho) & -\rho & \\ \hline & 0 & -(1+\rho) & \rho & 0 \\ \hline & 0 & -(1+\rho) & \rho & 0 \\ \hline \end{array}$$

The variance of these estimates is given by

$$Var \begin{bmatrix} \hat{\tau}_0 \\ \hat{\lambda}_0 \end{bmatrix} = \frac{\sigma^2}{(1+\rho)} \begin{bmatrix} 1+2\rho & \rho \\ \rho & 1+2\rho \end{bmatrix}.$$

We note again that this refers to the case $n = 1$ subject in each group in each cohort, and an increase in the value of n would simply reduce this variance by a factor of n .

8.5.2 Treatment information when the first-order carryover is not present

We here again consider the design in the previous subsection, but we discard the first-order carryover effect from the model. Just for the direct treatment effect we have the total treatment information

$$T'T = \frac{1}{2} \begin{bmatrix} 3 & 0 \\ 0 & 12 \end{bmatrix},$$

and subsequently the treatment information separately for various strata is obtained for both standard and the cohort designs as in Table 8.13.

Stratum	Standard Design			Cohort Design			
	σ^2	σ_s^2	Information Matrix	ρ	σ^2	σ_s^2	Information Matrix
Periods(ig. Cohorts)			$\begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$				$\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$
Cohorts(elim. Periods)	1	4	None	0.50	1	3	$\frac{1}{3} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
Group within Cohorts	1	4	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$		1	4	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$
Periods $\times$ Cohorts	1	0	None		1	0	$\frac{1}{3} \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$
Periods $\times$ (Grp.w. Coh.)	1	0	$\begin{bmatrix} 0 & 0 \\ 0 & 11 \end{bmatrix}$		1	0	$\begin{bmatrix} 0 & 0 \\ 0 & 11 \end{bmatrix}$
Total			$\begin{bmatrix} 3 & 0 \\ 0 & 12 \end{bmatrix}$				$\begin{bmatrix} 3 & 0 \\ 0 & 12 \end{bmatrix}$

Table 8.13: Treatment information for both standard and cohort two-treatment, extra-period design with one baseline when the carryover effect is not present.

8.6 Treatment information for three-treatment and three-period crossover design

In designs with more than two treatments, unlike the two treatment design, the main focus in planning is that we must decide which contrasts of treatments are to be estimated with the highest precision. This idea may lead us to want a design such that all pairwise differences between the treatments are estimated with the same precision. A design that possesses this property is known as variance balanced. Thus, our design comparing three treatments A , B and C , will be variance balanced if it satisfies the following property.

$$Var(\hat{\tau}_A - \hat{\tau}_B) = Var(\hat{\tau}_A - \hat{\tau}_C) = Var(\hat{\tau}_B - \hat{\tau}_C) = \nu\sigma^2,$$

where τ_A, τ_B , and τ_C are the direct treatment effects of treatment A , B and C respectively and ν is a constant for all treatments. A layout which achieves this is shown in Table 8.14.

		Period		
		1	2	3
Group	1	A	B	C
	2	B	C	A

Table 8.14: Three-treatment, three-period crossover design

A balanced incomplete block (BIB) design is one in which each treatment is replicated the same number of times and each pair of treatments occur together in the same block κ times, where κ is a constant integer. Then as can be seen the design in Table 8.14 is a balanced incomplete block design with periods as blocks with $t = 3$ treatments and $k = 2$ units in each of the three blocks, and the efficiency of this (BIB) design is

$$E = \frac{t(k-1)}{k(t-1)} = \frac{3 \times 1}{2 \times 2} = \frac{3}{4}.$$

This implies that 25% of the information on treatment differences is lost into the Period stratum. The above design is thus balanced if periods are regarded as blocks. We note that treatments are orthogonal to groups. The presence of carryover effects here may, however, destroy this balanced property if they are included in the model. The cohort design corresponding to this design is shown in Table 8.15.

		Period			
		1	2	3	4
Group	1	A	B	C	
	2	B	C	A	
	3		A	B	C
	4		B	C	A

Table 8.15: Cohort design in three-treatment and three-period crossover design

If the usual constraints

$$\tau_A + \tau_B + \tau_C = 0, \quad \lambda_A + \lambda_B + \lambda_C = 0,$$

are applied to the direct treatment and first-order carryover effects, we can put τ_1 and τ_2 for direct treatment effects of A and B respectively, λ_1 and λ_2 for first-order carryover effects for A and B respectively, and have $-(\tau_1 + \tau_2)$ and $-(\lambda_1 + \lambda_2)$ for direct treatment and carryover effects respectively for treatment C . Then the expected values of the cohort design of this model is given in Table 8.16.

$\mu + \tau_1$	$\mu + \tau_2 + \lambda_1$	$\mu - \tau_1 - \tau_2 + \lambda_2$
$\mu + \tau_2$	$\mu - \tau_1 - \tau_2 + \lambda_2$	$\mu + \tau_1 - \lambda_1 - \lambda_2$
$\mu + \tau_1$	$\mu + \tau_2 + \lambda_1$	$\mu - \tau_1 - \tau_2 + \lambda_2$
$\mu + \tau_2$	$\mu - \tau_1 - \tau_2 + \lambda_2$	$\mu + \tau_1 - \lambda_1 - \lambda_2$

Table 8.16: Expected values for three-treatment, three-period crossover and its corresponding cohort design

8.6.1 Treatment information when carryover effect is present

By following the expected values in Table 8.16 the design matrices T_0 , and T are written as

$$T_0 = 1_2 \otimes \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & -1 & -1 & 0 & 1 \\ 1 & -1 & -1 & 0 & 1 \\ 1 & 1 & 0 & -1 & -1 \end{bmatrix}, \quad T = \frac{1}{6} 1_2 \otimes \begin{bmatrix} 6 & 0 & 0 & -1 \\ 0 & 6 & 0 & -1 \\ 0 & 6 & 6 & -1 \\ -6 & -6 & 0 & 5 \\ -6 & -6 & 0 & 5 \\ 6 & 0 & -6 & -7 \end{bmatrix}$$

where for T matrix we have taken out the vector 1 due to the grand mean from matrix T_0 . Then the total information on $(\tau_1, \tau_2, \lambda_1, \lambda_2)$ is given by:

$$T'T = \frac{1}{6} \begin{bmatrix} 24 & 12 & -6 & -18 \\ 12 & 24 & 6 & -12 \\ -6 & 6 & 12 & 6 \\ -18 & -12 & 6 & 17 \end{bmatrix},$$

and treatment information for these parameters is given in Table 8.17.

Stratum	Standard Design				Cohort Design				
	σ^2	σ_s^2	Information Matrix		ρ	σ^2	σ_s^2	Information Matrix	
Periods(ig. Cohorts)			$\frac{1}{3}$	$\begin{bmatrix} 6 & 3 & -3 & -3 \\ 3 & 6 & 3 & 0 \\ -3 & 3 & 6 & 3 \\ -3 & 0 & 3 & 2 \end{bmatrix}$				$\frac{1}{12}$	$\begin{bmatrix} 9 & 9 & 0 & -3 \\ 9 & 18 & 9 & 0 \\ 0 & 9 & 9 & 3 \\ -3 & 0 & 3 & 2 \end{bmatrix}$
Cohorts(elim. Periods)	1	3		None	0.577	1	2	$\frac{1}{8}$	$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 4 & 2 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Group within Cohorts	1	3	$\frac{1}{3}$	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 2 \\ 0 & 0 & 2 & 1 \end{bmatrix}$		1	3	$\frac{1}{3}$	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 2 \\ 0 & 0 & 2 & 1 \end{bmatrix}$
Periods× Cohorts	1	0		None		1	0	$\frac{1}{8}$	$\begin{bmatrix} 9 & 0 & -9 & -6 \\ 0 & 0 & 0 & 0 \\ -9 & 0 & 9 & 6 \\ -6 & 0 & 6 & 4 \end{bmatrix}$
Periods ×(Grp.w. Coh.)	1	0	$\frac{1}{3}$	$\begin{bmatrix} 18 & 9 & -3 & -15 \\ 9 & 18 & 3 & -12 \\ -3 & 3 & 2 & 1 \\ -15 & -12 & 1 & 14 \end{bmatrix}$		1	0	$\frac{1}{3}$	$\begin{bmatrix} 18 & 9 & -3 & -15 \\ 9 & 18 & 3 & -12 \\ -3 & 3 & 2 & 1 \\ -15 & -12 & 1 & 14 \end{bmatrix}$
Total			$\frac{1}{3}$	$\begin{bmatrix} 24 & 12 & -6 & -18 \\ 12 & 24 & 6 & -12 \\ -6 & 6 & 12 & 6 \\ -18 & -12 & 6 & 17 \end{bmatrix}$				$\frac{1}{3}$	$\begin{bmatrix} 24 & 12 & -6 & -18 \\ 12 & 24 & 6 & -12 \\ -6 & 6 & 12 & 6 \\ -18 & -12 & 6 & 17 \end{bmatrix}$

Table 8.17: Information matrices for standard and cohort designs for three-treatment, three-period.

By referring to Table 8.17 for τ_1 and τ_2 , 75% of the information is in the Periods×(Group within Cohort) stratum. However, half of what is left for τ_1 is recovered in the Periods×Cohorts stratum with a variance σ^2 and a small amount more in the Cohorts(elim. Periods) stratum. The Periods×Cohort stratum actually provides information on only one treatment

combination, namely $3\tau_1 - 3\lambda_1 - 2\lambda_2$, while Cohorts(elim. Periods) has information only on the linear combination $\tau_1 + 2\lambda_2 + \lambda_1$.

8.6.2 Treatment information when carryover effect is not present

The total information for just direct treatment effect after taking out the grand mean from the matrix design of the model is

$$T'T = \begin{bmatrix} 8 & 4 \\ 4 & 8 \end{bmatrix},$$

and treatment information for both standard and cohort design is given in Table 8.18. It shows that, of the 25% of the information lost to Periods in the standard design, about half of this is recovered for τ_1 from the Periods $\times$ Cohorts stratum, and a quarter of it for τ_2 is recovered from the Cohorts(elim. Periods) stratum. It may be possible to do better than this by astute choice of design or by extending the design to three cohorts.

Stratum	Standard Design			Cohort Design			
	σ^2	σ_s^2	Information Matrix	ρ	σ^2	σ_s^2	Information Matrix
Periods(ig. Cohorts)			$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$			$\frac{3}{4}$	$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$
Cohorts(elim. Periods)	1	3	None	0.577	1	2	$\frac{1}{8} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$
Group within Cohorts	1	3	None		1	3	None
Periods $\times$ Cohorts	1	0	None		1	0	$\frac{1}{8} \begin{bmatrix} 9 & 0 \\ 0 & 0 \end{bmatrix}$
Periods $\times$ (Grp.w. Coh.)	1	0	$3 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$		1	0	$3 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$
Total			$4 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$				$4 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

Table 8.18: Treatment information for both standard and the cohort designs in three-treatment, three-period when the carryover effect is not present.

8.7 Treatment information for two-treatment with one baseline measurement with more than two cohorts

In Chapter 6 we analysed this design with two cohorts, now we see what happens when we analyse with more than two cohorts. The layout for this case is shown in Table 8.19.

		Period				
		1	2	3	4	5
Group	1	—	A	B		
	2	—	B	A		
	3		—	A	B	
	4		—	B	A	
	5			—	A	B
	6			—	B	A

Table 8.19: Two-treatment crossover design with one baseline measurement when $c = 3$

The means corresponding to these 18 groups can be written as in Table 8.20

μ	$\mu + \tau_0 + \tau_D$	$\mu + \tau_0 + \lambda_0 - \tau_D + \lambda_D$
μ	$\mu + \tau_0 - \tau_D$	$\mu + \tau_0 + \lambda_0 + \tau_D - \lambda_D$
μ	$\mu + \tau_0 + \tau_D$	$\mu + \tau_0 + \lambda_0 - \tau_D + \lambda_D$
μ	$\mu + \tau_0 - \tau_D$	$\mu + \tau_0 + \lambda_0 + \tau_D - \lambda_D$
μ	$\mu + \tau_0 + \tau_D$	$\mu + \tau_0 + \lambda_0 - \tau_D + \lambda_D$
μ	$\mu + \tau_0 - \tau_D$	$\mu + \tau_0 + \lambda_0 + \tau_D - \lambda_D$

Table 8.20: The 18 means for two-treatment with one baseline measurement when $c = 3$

As for the previous designs, we consider this case in two stages, with and without first-order carryover effect.

8.7.1 Treatment information when $c = 3$ and first-order carry-over effect is present

By referring to Chapter 7 and considering the means repeated three times due to $c = 3$, and also the different ordering to put the design matrix, then the design matrix T_0 can be presented as below

$$T_0 = 1_3 \otimes \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & -1 & 0 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 \end{bmatrix} \quad T = \frac{1}{6} 1_3 \otimes \begin{bmatrix} -4 & -2 & 0 & 0 \\ -4 & -2 & 0 & 0 \\ 2 & -2 & 6 & 0 \\ 2 & -2 & -6 & 0 \\ 2 & 4 & -6 & 6 \\ 2 & 4 & 6 & -6 \end{bmatrix}$$

and the total information for this case is

$$T'T = \begin{bmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 12 & -6 \\ 0 & 0 & -6 & 6 \end{bmatrix}.$$

This information is now partitioned among the different strata as shown in Table 8.21. We see that, in the cohort design, about 60% of the information about τ_0 and λ_0 is recovered in the various strata, with some appearing in the Cohorts (elim. Periods) and some in the Cohort $\times$ Period interaction.

Stratum	Standard Design			Cohort Design			
	σ^2	σ_s^2	Information Matrix	ρ	σ^2	σ_s^2	Information Matrix
Periods(ig. Cohorts)			$\begin{bmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$			$\frac{1}{3}$	$\begin{bmatrix} 5 & 4 & 0 & 0 \\ 4 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Cohorts(elim. Periods)	1	3	None	0.707	1	1.5	$\frac{1}{2}$ $\begin{bmatrix} 3 & 3 & 0 & 0 \\ 3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
	1	3	None	0.408	1	2.5	$\frac{1}{30}$ $\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Group within Cohorts	1	3	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$		1	3	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$
Periods $\times$ Cohorts	1	0	None		1	0	$\frac{4}{5}$ $\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$
Periods $\times$ (Grp.w. Coh.)	1	0	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & -6 \\ 0 & 0 & -6 & 4 \end{bmatrix}$		1	0	$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & -6 \\ 0 & 0 & -6 & 4 \end{bmatrix}$
Total			$\begin{bmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 12 & -6 \\ 0 & 0 & -6 & 6 \end{bmatrix}$				$\begin{bmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 12 & -6 \\ 0 & 0 & -6 & 6 \end{bmatrix}$

Table 8.21: Information matrices for standard and cohort designs for two-treatment with a baseline, $c = 3$

8.8 Conclusion

As we showed, classical analysis of crossover design even if baseline measurements are included and in the presence of strong or unknown period effects is unable to recover information about the difference between the average treatment effect and baseline. Using the cohort design in crossover design, we can recover some information about (τ_0, λ_0) and obtain estimates for them from within and between subjects. This method opens the way for a more detailed examination of the topic and in this chapter we have examined various crossover designs.

In comparing the cohort and standard designs and to show the abilities of cohort designs to recover some information about the parameters of interest, consider the information matrices in Table 8.22 in three cohort designs in the analysis of two-treatment, two-period crossover design with 1, 2 and 3 cohorts, or equivalently, $c = 1$, $c = 2$ and $c = 3$ respectively. We note that $c = 1$ is equivalent to the standard design.

	Information matrix		
Strata	$c = 1$	$c = 2$	$c = 3$
Periods(ig. Cohorts)	$\frac{1}{6} \begin{bmatrix} 8 & 4 \\ 4 & 8 \end{bmatrix}$	$\frac{1}{6} \begin{bmatrix} 10 & 8 \\ 8 & 10 \end{bmatrix}$	$\frac{1}{6} \begin{bmatrix} 10 & 8 \\ 8 & 10 \end{bmatrix}$
Cohort(elim. Periods)		$\frac{1}{6} \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix}$	$\frac{1}{6} \begin{bmatrix} 9 & 9 \\ 9 & 9 \end{bmatrix}$ $\frac{1}{6} \begin{bmatrix} 0.2 & -0.2 \\ -0.2 & 0.2 \end{bmatrix}$
Periods $\times$ Cohorts		$\frac{1}{6} \begin{bmatrix} 3 & -3 \\ -3 & 3 \end{bmatrix}$	$\frac{1}{6} \begin{bmatrix} 4.8 & -4.8 \\ -4.8 & 4.8 \end{bmatrix}$
Total	$\frac{1}{6} \begin{bmatrix} 8 & 4 \\ 4 & 8 \end{bmatrix}$	$\frac{1}{6} \begin{bmatrix} 16 & 8 \\ 8 & 16 \end{bmatrix}$	$\frac{1}{6} \begin{bmatrix} 24 & 12 \\ 12 & 24 \end{bmatrix}$

Table 8.22: Information on τ_0 and λ_0 in three cases of design

As can be seen, the total information will increase when we increase the number of cohorts in the model. In addition in the standard design ($c = 1$), all information about the parameters of interest is in the Periods(ig. Cohorts) stratum, although in the cohort designs ($c = 2, 3$) the information is distributed in all strata, and the proportion of the

information recoverable increases, being close to 40% when $c = 2$ and nearly 60% when $c = 3$.

APPENDICES

A Some useful concepts from linear algebra

In this appendix, we give some basic matrix results which are need at various stages in this thesis.

A.1 The Kronecker product

We begin with a brief review of the definition of a right Kronecker product and some of its properties. Thus, if A and B denote matrices of dimension $p \times q$ and $m \times n$, then the right Kronecker product of A and B is defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1q}B \\ \vdots & & \vdots \\ a_{p1}B & \dots & a_{pq}B \end{bmatrix}$$

Observe that $A \otimes B$ is a matrix of dimension $pm \times qn$ made up of pq submatrices with each submatrix equal to the scalar multiplication of an element of A with the matrix B . Some useful properties of the above Kronecker product include

- (1) $(A \otimes B)' = A' \otimes B'$,
- (2) $(A \otimes B)(C \otimes D) = AC \otimes BD$;
- (3) $rank(A \otimes B) = rank(A)rank(B)$
- (4) $(A + B) \otimes (C + D) = (A \otimes C) + (A \otimes D) + (B \otimes C) + (B \otimes D)$;
- (5) $(AB \otimes CD) = (A \otimes C)(B \otimes D)$;
- (6) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
- (7) $(A \otimes B)^- = A^- \otimes B^-$
- (8) $|A \otimes B| = |A|^p |B|^m$,

where we assume that the matrices involved in regular matrix multiplication are of appropriate sizes and in property numbers 6-8 the two matrices, A and B , are square matrices. These properties are established , for example, in (Rao (1973), p. 29)

A.2 Idempotent matrices

A matrix A of order $m \times m$, is said to be idempotent if

$$AA = A.$$

We note that if A is idempotent, each of its eigenvalues is order 0 or 1. If A is idempotent, then the matrix $I - A$ is also idempotent, since we have

$$(I - A)(I - A) = I - A - A + AA = I - A,$$

since $AA = A$. The rank of an idempotent matrix is equal to its trace.

A.3 Positive definite quadratic forms and matrices

Q is said to be a quadratic form in the n variables $(x_1, \dots, x_n)$ if

$$Q = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = x'Ax$$

We refer to $A = (a_{ij})$ as the matrix of quadratic form and we will assume that A is symmetric. A matrix A is said to be positive definite if the quadratic form $Q = x'Ax > 0$ for all $x \neq 0$. A matrix A is said to be positive semidefinite if $Q = x'Ax \geq 0$ for all $x \neq 0$.

A.4 Contrast and orthogonal contrasts

As Jones & Kenward (1989) pointed out a contrast on a vector $\theta = (\tau_1, \tau_2, \dots, \tau_t)$ is a linear combination $c_1\tau_1 + c_2\tau_2 + \dots + c_t\tau_t$, in which $c_1 + c_2 + \dots + c_t = 0$. For example with a vector with two dimensions there is only one such contrast and that is $\tau_1 - \tau_2$. The coefficients of this simple contrast are $c_1 = 1$ and $c_2 = -1$. With more than two dimensions, however, a number of different sets of contrasts can be written. One contrast $c_1\tau_1 + c_2\tau_2 + \dots + c_t\tau_t$ is orthogonal to another $d_1\tau_1 + d_2\tau_2 + \dots + d_t\tau_t$ if and only if $c_1d_1 + c_2d_2 + \dots + c_td_t = 0$. In general we can always write down a set which includes up to $t - 1$ orthogonal contrasts on a vector with dimension of t .

A.5 ANOVA sum of squares as quadratic forms

In the following discussion we denote the $k \times k$ identity matrix by I_k and the $k \times k$ of ones by E_k . In terms of matrices like I_k and E_k , a neat expression may be given of the

various quadratic forms in terms of matrices J_k and K_k denoted by

$$J_k = k^{-1} E_k \quad \text{and} \quad K_k = I_k - J_k,$$

where J and K are matrices of rank 1 and $(k - 1)$, respectively.

Suppose X is an $r \times c$ matrix with X_{ij} as (i, j) th element and Y is a $rc \times 1$ vector obtained by stacking the rows of X starting with first row, i.e, $Y = (X'_1, \dots, X'_r)'$ where $X_i = (X_{i1}, \dots, X_{ic})'$ for $i = 1, \dots, r$. To avoid confusion, the reader should note that subscripts in this note play a dual role, namely, to indicate the dimension of a matrix as a means for referencing an element or a row of a matrix. The first basic sum of squares as a matrix quadratic form is

$$\sum_{i=1}^r \sum_{j=1}^c X_{ij}^2 = Y' [I_r \otimes I_c] Y. \quad (\text{A.1})$$

This follows from the definition of the inner product of Y with itself and the fact that $[I_r \otimes I_c] = I_{rc}$. Next, we have

$$\left[\sum_{i=1}^r \sum_{j=1}^c X_{ij} \right]^2 = Y' [E_r \otimes E_c] Y, \quad (\text{A.2})$$

because the left side of (A.2) is equal to

$$[1'_{rc} Y]^2 = Y' 1_{rc} 1'_{rc} Y = Y' [E_{rc}] Y = Y' [E_r \otimes E_c] Y,$$

where 1_{rc} is the $rc \times 1$ vector of ones. The third basic sum of squares as a quadratic form is given by

$$\sum_{i=1}^r \left[\sum_{j=1}^c X_{ij} \right]^2 = Y' [I_r \otimes E_c] Y, \quad (\text{A.3})$$

since the left side of (A.3) may be written as

$$\sum_{i=1}^r [1'_r X_i]^2 = \sum_{i=1}^r X'_i E_c X_i = Y' [I_r \otimes E_c] Y,$$

The final basic sum of squares as a matrix quadratic form is

$$\sum_{j=1}^c \left[\sum_{i=1}^r X_{ij} \right]^2 = Y' [E_r \otimes I_c] Y, \quad (\text{A.4})$$

To see this, we expand the left side (A.4) to obtain

$$\sum_{ij1}^c \sum_{i=1}^r X_{ij}^2 + \sum_{j=1}^c \sum_{i=1}^r \sum_{k \neq i} X_{ij} X_{kj},$$

In terms of vectors representing the i th and k th rows of matrix X , the preceding expression may be rewritten as

$$\sum_{j=1}^c \sum_{i=1}^r X_{ij}^2 + \sum_{j=1}^c \sum_{i=1}^r \sum_{k \neq i} X_{ij} X_{kj} = \begin{bmatrix} X'_1 & \dots & X'_r \end{bmatrix} \begin{bmatrix} I_c & \dots & I_c \\ \vdots & & \vdots \\ I_c & \dots & I_c \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_r \end{bmatrix}$$

A.6 Summing vectors, and E -matrices

Vectors having every element equal to unity are called summing vectors and are denoted by $\mathbf{1}$, using a subscript to present order when necessary. They are called summing vectors because, with $\mathbf{x}' = [x_1, x_2, \dots, x_n]$, for example $\mathbf{1}'_n \mathbf{x} = \sum_{i=1}^n x_i$. In particular, the inner product of $\mathbf{1}_n$ with itself is n : $\mathbf{1}'_n \mathbf{1}_n = n$. A product of a summing vector with a matrix yields a vector of either column totals or row totals, of the matrix involved: for $\mathbf{B}$ having elements b_{ij} , the product $\mathbf{1}'\mathbf{B}$ is a row vector of column totals $b_{.j}$, and $\mathbf{B}\mathbf{1}$ is a column vector of row totals $b_{i.}$. Outer products of summing vectors with each other are matrices having every element unity. They are denoted by $\mathbf{E}$. For example,

$$\mathbf{1}_n \mathbf{1}'_n = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} = \mathbf{E}_n.$$

For $\mathbf{E}$ square and of size n , $\mathbf{E}^2 = n\mathbf{E}$, and $\mathbf{E}\mathbf{1}_n = n\mathbf{1}_n$ and $tr(\mathbf{E}_n) = n$. Two useful variants of $\mathbf{E}_n$ are

1. $\mathbf{J}_n = \frac{1}{n}\mathbf{E}_n$
2. $\mathbf{K}_n = \mathbf{I}_n - \mathbf{J}_n$,

which are idempotent matrices of rank 1 and $(n - 1)$, respectively.

B Bayes' theorem

Suppose that $y = (y_1, y_2, \dots, y_n)'$ is a vector of n observations whose probability distribution $p(y|\theta)$ depends on the values of k parameters $\theta = (\theta_1, \dots, \theta_k)'$. Suppose also that θ itself has a probability distribution $p(\theta)$. Then,

$$p(y|\theta)p(\theta) = p(y, \theta) = p(\theta|y)p(y) \quad (\text{B.1})$$

Given the observed data, y , the conditional distribution of θ is

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \quad (\text{B.2})$$

where $p(y) = E_\theta[p(y|\theta)] = c^{-1} = \begin{cases} \int p(y|\theta)p(\theta)d\theta & \text{if } \theta \text{ is continuous} \\ \sum p(y|\theta)p(\theta) & \text{if } \theta \text{ is discrete} \end{cases}$,

where the sum or the integral is taken over the admissible range of θ , and where $E_\theta[f]$ is the mathematical expectation of f with respect to the distribution $p(\theta)$. Thus we may write (B.1) alternatively as

$$p(\theta|y) = cp(y|\theta)p(\theta)$$

or

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

where the proportionality depends only on y , not θ .

B.1 Normal prior for multinormal sample

Suppose that $(x_1, x_2, \dots, x_n)$ are independently and identically distributed as $N(\theta, \Sigma)$.

The likelihood function is

$$\begin{aligned} f(x_1, x_2, \dots, x_n|\theta, \Sigma) &\propto \prod_{i=1}^n |\Sigma|^{-1/2} \exp\{-(x_i - \theta)' \Sigma^{-1} (x_i - \theta)/2\} \\ &= |\Sigma|^{-n/2} \exp\{-\sum_{i=1}^n (x_i - \theta)' \Sigma^{-1} (x_i - \theta)/2\} \end{aligned} \quad (\text{B.3})$$

Now expanding the quadratic forms we have

$$\begin{aligned} \sum_{i=1}^n (x_i - \theta)' \Sigma^{-1} (x_i - \theta) &= n\theta' \Sigma^{-1} \theta - n\bar{x}' \Sigma^{-1} \theta - n\theta' \Sigma^{-1} \bar{x} + \sum_{i=1}^n x_i' \Sigma^{-1} x_i \\ &= n(\theta - \bar{x})' \Sigma^{-1} (\theta - \bar{x}) + r, \end{aligned}$$

where $\bar{x} = n^{-1} \sum_i x_i$ is the sample mean vector, and

$$\begin{aligned} r &= \sum_{i=1}^n x_i' \Sigma^{-1} x_i - n \bar{x}' \Sigma^{-1} \bar{x} \\ &= \sum_{i=1}^n (x_i - \bar{x})' \Sigma^{-1} (x_i - \bar{x}) \\ &= \sum_{i=1}^n \text{trace} \Sigma^{-1} (x_i - \bar{x})(x_i - \bar{x})' \\ &= n \text{trace} \Sigma^{-1} S \end{aligned}$$

and $S = n^{-1} \sum_i^n (x_i - \bar{x})(x_i - \bar{x})'$ is the sample covariance matrix. Therefore

$$f(x_1, x_2, \dots, x_n | \theta, \Sigma) \propto |\Sigma|^{-n/2} \exp\{-n(\theta - \bar{x})' \Sigma^{-1} (\theta - \bar{x})/2 - n \text{trace}(\Sigma^{-1} S/2)\} \quad (\text{B.4})$$

Now if Σ is known, two terms can be dropped from (B.4) leaving

$$f(x_1, x_2, \dots, x_n | \theta, \Sigma) \propto \exp\{-n(\theta - \bar{x})' \Sigma^{-1} (\theta - \bar{x})/2\}. \quad (\text{B.5})$$

If we let the prior distribution of θ be $N(\theta_0, \Sigma_0)$ we find

$$\begin{aligned} f(\theta | x_1, x_2, \dots, x_n) &\propto f(\theta) f(x_1, x_2, \dots, x_n | \theta) \\ &\propto \exp\{-(\theta - \theta_0)' \Sigma_0^{-1} (\theta - \theta_0)/2 - n(\theta - \bar{x})' \Sigma^{-1} (\theta - \bar{x})/2\}. \end{aligned}$$

Now complete the square by

$$\begin{aligned} (\theta - \theta_0)' \Sigma_0^{-1} (\theta - \theta_0) + n(\theta - \bar{x})' \Sigma^{-1} (\theta - \bar{x}) &= \theta' (\Sigma_0^{-1} + n \Sigma^{-1}) \theta - (\theta_0' \Sigma_0^{-1} + n \bar{x}' \Sigma^{-1}) \theta \\ &\quad - \theta' (\Sigma_0^{-1} \theta_0 + n \Sigma^{-1} \bar{x}) + (\theta_0' \Sigma_0^{-1} \theta_0 + n \bar{x}' \Sigma^{-1} \bar{x}) \\ &= (\theta - \theta_p)' (\Sigma_p)^{-1} (\theta - \theta_p) + R, \end{aligned}$$

where

$$\begin{aligned} \Sigma_p &= (\Sigma_0^{-1} + n \Sigma^{-1})^{-1} \\ \theta_p &= \Sigma_p (\Sigma_0^{-1} \theta_0 + n \Sigma^{-1} \bar{x}), \end{aligned}$$

and R is a constant. The posterior distribution of θ is therefore $N(\theta_p, \Sigma_p)$. The posterior mean is a matrix-weighted average of the prior mean θ_0 and the sample mean $\bar{x}$. The weights are respectively the prior information matrix Σ_0^{-1} and the analogous data information $n \Sigma^{-1}$ (being the inverse of the covariance matrix of $\bar{x}$). The posterior information matrix is the sum of Σ_0^{-1} and $n \Sigma^{-1}$, and the posterior covariance matrix is smaller than either the prior covariance matrix Σ_0 or the data analogue $n^{-1} \Sigma$, in the sense that both $\Sigma_0 - \Sigma_p$ and $n^{-1} \Sigma - \Sigma_p$ are positive definite. Then the distribution of θ is $N(\theta_p, \Sigma_p)$.

C The S-PLUS program for cohort designs

Chapters 6, 7 and 8 require the calculation of the canonical correlations and the canonical variables for the columns of two design matrices P and C representing Periods (ig. Cohorts) and Cohorts(ig. Periods), respectively. The following functions in S-PLUS require that the user specify only the number of cohorts `nc`, the number of periods `np`, and the number of groups within each cohort `ng`. It returns the canonical correlations in the vector 'rho' and the contrasts required for the various subspaces defined in the text.

C.1 S-PLUS functions

Firstly, we need a function which will form the Kronecker products. This is provided in the S-PLUS library `design`:

```
kron<-function(...)
{
  l <- list(...)
  a <- as.matrix(l[[1]])
  for(b in l[-1]) {
    b <- as.matrix(b)
    a <- matrix(aperm(a%o%b, c(3, 1, 4, 2)), nrow=nrow(a)*
               nrow(b), ncol = ncol(a)*ncol(b))
  }
  a
}
```

The function `cc` then forms the required projector matrices, as defined in Table 7.2, where $V_4 = I_c \otimes J_p \otimes K_g$ and $V_5 = I_c \otimes K_p \otimes K_g$:

```
cc<-function(nc, np, ng)
{
  # Set up Period and Cohort factors
  P <- factor(outer(1:np, (1:nc) - 1, "+"))
  C <- factor(rep(1:nc, rep(np, nc)))
}
```

```

    design <- data.frame(C, P)
# Set up Helmert contrasts in P and C
    Xp <- model.matrix( ~ P, design)[, -1]
    Xc <- model.matrix( ~ C, design)[, -1]
# Make orthogonal to grand mean
    Xp <- scale(Xp, scale = F)
    Xc <- scale(Xc, scale = F)
# Form canonical correlations
    ccor <- cancel(Xc, Xp)
# Form contrasts for PiC and CiP
    Pc <- Xp %*% ccor$ycoef
    Cc <- Xc %*% ccor$xcoef
# Form contrasts for CeP
    rho <- ccor$cor
    m <- 1:length(rho)
    if(length(rho) > 1)
        {Cepc <- (Cc[,m]-Pc[,m]%*%diag(rho))%*%diag(1/sqrt(1-rho^2))}
    if(length(rho) == 1)
        {Cepc <- (Cc[ m]-Pc[,m]*rho)/sqrt(1-rho^2)}
# Find projection matrices
S1<-Pc%*%t(Pc)
S2<-Cepc%*%t(Cepc)
    one<-rep(1,np*nc)
S3<-diag(1,np*nc)-one %*%t(one)/(np*nc)
S3<-S3-S1-S2
# Now include the ng groups within each cohort
    Jg <- matrix(1/ng, ng, ng)
    Ig <- diag(1, ng, ng)
    Kg <- Ig - Jg
    V1 <- kron(S1, Jg)
    V2 <- kron(S2, Jg)
    V3 <- kron(S3, Jg)
    Ic <- diag(1, nc, nc)

```

```

      Jp <- matrix(1/np, np, np)
      Kp <- diag(1, np, np) - Jp
# Groups within Cohort
      S4 <- kron(Ic, Jp)
      V4 <- kron(S4, Kg)
# Periods x (Groups within Cohort)
      S5 <- kron(Ic, Kp)
      V5 <- kron(S5, Kg)
# Periods(elim. Groups)
      S6 <- S1+S2-Cc%*%t(Cc)
      V6 <- kron(S6, Jg)
      list(Xc = round(Xc, 6), Xp = round(Xp, 6), rho = round(rho, 6),
           Cc = round(Cc, 6), Pc = round(Pc, 6), Cepc = round(Cepc, 6),
           V1 = round(V1, 6), V2 = round(V2, 6), V3 = round(V3, 6),
           V4 = round(V4, 6), V5 = round(V5, 6), V6 = round(V6, 6))
}

```

The above function provides the projector matrices, but to get the information matrices, we need a matrix T whose columns describe the treatment parameters and are orthogonal to the vector of ones:

```

infn<-function(tm,V1,V2,V3,V4,V5,V6)
{ tm<-scale(tm,scale=F)
  onec<-matrix(rep(1,nc),nrow=nc)
  t<-kron(onec,tm)
# Now the treatment information matrices are obtained using the
# projector matrices obtained from cc()
  I1<-t(t) %*% V1 %*% t
  I2<-t(t) %*% V2 %*% t
  I3<-t(t) %*% V3 %*% t
  I4<-t(t) %*% V4 %*% t
  I5<-t(t) %*% V5 %*% t
  I6<-t(t) %*% V6 %*% t
  list(I1 = round(I1, 6), I2 = round(I2, 6), I3 = round(I3, 6),

```

```

I4 = round(I4, 6), I5 = round(I5, 6), I6 = round(I6, 6))
}

```

C.2 Use of S-PLUS functions for Chapter 6

We now illustrate the use of these functions in the analysis of the design given in Chapter 6. Firstly, we find the projector matrices using `cc`:

```

cc232<-cc(2,3,2)
# Projector matrices
6*S1
  1  2  3  4  5  6
1  5 -1 -1 -1 -1 -1
2 -1  2 -1  2 -1 -1
3 -1 -1  2 -1  2 -1
4 -1  2 -1  2 -1 -1
5 -1 -1  2 -1  2 -1
6 -1 -1 -1 -1 -1  5
SPLUS > round(24*(V2+V4))
  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]   4  -4   4  -4   4  -4   0   0   0   0   0   0
[2,]  -4   4  -4   4  -4   4   0   0   0   0   0   0
[3,]   4  -4   7  -1   7  -1  -3  -3  -3  -3   0   0
[4,]  -4   4  -1   7  -1   7  -3  -3  -3  -3   0   0
[5,]   4  -4   7  -1   7  -1  -3  -3  -3  -3   0   0
[6,]  -4   4  -1   7  -1   7  -3  -3  -3  -3   0   0
[7,]   0   0  -3  -3  -3  -3   7  -1   7  -1   4  -4
[8,]   0   0  -3  -3  -3  -3  -1   7  -1   7  -4   4
[9,]   0   0  -3  -3  -3  -3   7  -1   7  -1   4  -4
[10,]  0   0  -3  -3  -3  -3  -1   7  -1   7  -4   4
[11,]  0   0   0   0   0   0   4  -4   4  -4   4  -4
[12,]  0   0   0   0   0   0  -4   4  -4   4  -4   4
SPLUS > round(24*(V3+V5))
  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]

```

[1,]	4	-4	-2	2	-2	2	0	0	0	0	0	0
[2,]	-4	4	2	-2	2	-2	0	0	0	0	0	0
[3,]	-2	2	5	-3	-4	0	-1	-1	2	2	0	0
[4,]	2	-2	-3	5	0	-4	-1	-1	2	2	0	0
[5,]	-2	2	-4	0	6	-3	2	2	-1	-1	0	0
[6,]	2	-2	0	-4	-3	6	2	2	-1	-1	0	0
[7,]	0	0	-1	-1	2	2	5	-3	-4	0	-2	2
[8,]	0	0	-1	-1	2	2	-3	5	0	-4	2	-2
[9,]	0	0	2	2	-1	-1	-4	0	6	-3	-2	2
[10,]	0	0	2	2	-1	-1	0	-4	-3	6	2	-2
[11,]	0	0	0	0	0	0	-2	2	-2	2	4	-4
[12,]	0	0	0	0	0	0	2	-2	2	-2	-4	4

SPLUS > round(24*V6)

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
[1,]	8	8	-4	-4	-4	-4	0	0	0	0	0	0
[2,]	8	8	-4	-4	-4	-4	0	0	0	0	0	0
[3,]	-4	-4	5	5	-1	-1	3	3	-3	-3	0	0
[4,]	-4	-4	5	5	-1	-1	3	3	-3	-3	0	0
[5,]	-4	-4	-1	-1	5	5	-3	-3	3	3	0	0
[6,]	-4	-4	-1	-1	5	5	-3	-3	3	3	0	0
[7,]	0	0	3	3	-3	-3	5	5	-1	-1	-4	-4
[8,]	0	0	3	3	-3	-3	5	5	-1	-1	-4	-4
[9,]	0	0	-3	-3	3	3	-1	-1	5	5	-4	-4
[10,]	0	0	-3	-3	3	3	-1	-1	5	5	-4	-4
[11,]	0	0	0	0	0	0	-4	-4	-4	-4	8	8
[12,]	0	0	0	0	0	0	-4	-4	-4	-4	8	8

Then, we need to calculate the information matrices given in Table 6.8, using the treatment design matrix of given by the last four columns of the matrix in (6.3.1).

#This is the T as given in Chapter 6: Coh x Per x Grp

tm<-matrix(c(0,0,1,1,1,1, 0,0,0,0,1,1,

0,0,1,-1,-1,1, 0,0,0,0,1,-1),nrow=6)

inf<-infn(tm,cc232\$V1,cc232\$V2,cc232\$V3,cc232\$V4,cc232\$V5,cc232\$V6)

Information matrices

SPLUS > 6*I1

	[,1]	[,2]	[,3]	[,4]
[1,]	10	8	0	0
[2,]	8	10	0	0
[3,]	0	0	0	0
[4,]	0	0	0	0

SPLUS > 6*I2

	[,1]	[,2]	[,3]	[,4]
[1,]	3	3	0	0
[2,]	3	3	0	0
[3,]	0	0	0	0
[4,]	0	0	0	0

SPLUS > 6*I3

	[,1]	[,2]	[,3]	[,4]
[1,]	3	-3	0	0
[2,]	-3	3	0	0
[3,]	0	0	0	0
[4,]	0	0	0	0

SPLUS > 6*I4

	[,1]	[,2]	[,3]	[,4]
[1,]	0	0	0	0
[2,]	0	0	0	0
[3,]	0	0	0	0
[4,]	0	0	0	8

SPLUS > 6*I5

	[,1]	[,2]	[,3]	[,4]
[1,]	0	0	0	0
[2,]	0	0	0	0
[3,]	0	0	48	-24
[4,]	0	0	-24	16

SPLUS > 6*I6

	[,1]	[,2]	[,3]	[,4]
--	------	------	------	------

```

[1,]  13  11   0   0
[2,]  11  13   0   0
[3,]   0   0   0   0
[4,]   0   0   0   0

```

Finally, for the standard design, we can obtain the information matrices using the same design matrix and applying the idempotent matrices given after (6.4.2):

```

# Set up basic matrices
I2<-matrix(c(1,0,0,1),nrow=2)
J2<-matrix(c(1,1,1,1),nrow=2)/2
K2<-I2-J2
I3<-matrix(c(1,0,0,0,1,0,0,0,1),nrow=3)
J3<-matrix(rep(1,9),nrow=3)/3
K3<-I3-J3

# Set up idempotents
Q1<-kron(J2,K3,J2)
Q2<-kron(I2,J3,I2)-kron(J2,J3,J2)
Q3<-kron(I2,K3,I2)-kron(J2,K3,J2)

# Fix up design matrix
tm<-scale(tm,scale=F)
onec<-matrix(rep(1,nc),nrow=nc)

# Calculate information matrices
t<-kron(onec,tm)
Is1<-t(t) %*% Q1 %*% t
Is2<-t(t) %*% Q2 %*% t
Is3<-t(t) %*% Q3 %*% t
SPLUS > round(6*Is1)
[1,]  16   8   0   0
[2,]   8  16   0   0
[3,]   0   0   0   0
[4,]   0   0   0   0
SPLUS > round(6*Is2)
[,1] [,2] [,3] [,4]

```

[1,]	0	0	0	0
[2,]	0	0	0	0
[3,]	0	0	0	0
[4,]	0	0	0	8

SPLUS > round(6*Is3)

	[,1]	[,2]	[,3]	[,4]
[1,]	0	0	0	0
[2,]	0	0	0	0
[3,]	0	0	48	-24
[4,]	0	0	-24	16

Bibliography

- ABEYASEKERA, S. & CURNOW, R. N. (1984). The desirability of adjusting for residual effect in a crossover designs. *Biometrics* **40**, 1071–1078.
- ARMITAGE, A. P. & HILLS, M. (1982). The two-period crossover trials. *Statistician* **31**, 118–131.
- BALAAM, L. N. (1968). A two-period design with t^2 experimental units. *Biometrics* **24**, 61–73.
- BARKER, N., HEWS, R. J., & POLONIECKI, J. (1982). The two-period crossover trials. *Bull. in Appl. Statist.* **9**, 67–116.
- BROWN, B. W. (1980). The crossover experiment for clinical trials. *Biometrics* **36**, 69–79.
- CHASSAN, J. B. (1964). On the analysis of simple cross-over with unequal numbers of replicates. *Biometrics* **20**, 206–208.
- CHI, E. M. (1991). Recovery of inter-block information in cross-over trial. *Statistics in Medicine* **10**, 1115–1122.
- COCHRAN, W. G. (1939). Long-term agricultural experiments (with discussion). *Journal of the Royal Statistical Society, B*, **6**, 104–48.
- COCHRAN, W. G., AUTREY, K. M., & CANNON, C. Y. (1941). A double change-over design for cattle feeding experiments. *Journal of Dairy Science*. **24**, 937–51.
- CORNELL, R. G. (1980). Evaluation of bioavailability data using non-parametric statistics, in Drug Absorption and Disposition. *Statistical Considerations* (ed. K. S. Albert). *American Pharmaceutical Association* pages pp. 51–7.
- EBBUTT, A. F. (1984). Three-period crossover designs for two treatments. *Biometrics* **40**, 219–224.
- FAREWELL, V. T. (1985). Some remarks on the analysis of cross-over trials with a binary response. *Applied Statistics* **34**, 121–128.
- FEDERER, W. T. & ATKINSON, G. F. (1964). Tied-double-change-over designs. *Bio-*

- metrics* **20**, 168–181.
- FINNEY, D. J. (1978). *Statistical Method in Biological Assay*. Griffin, London.
- FLETCHER, D. J. (1987). A new class of change-over designs for factorial experiments. *Biometrika* **74**, 649–654.
- FREEMAN, P. R. (1989). The performance of the two-stage analysis of two-treatment, two-period crossover trials. *Statistics in Medicine* **8**, 1421–39.
- GART, J. J. (1969). An exact test for comparing matched proportions in crossover designs. *Biometrika* **56**, 75–80.
- GOUGH, K. (1989). Letter to the editor. *Statistics in Medicine* **8**, 891–892.
- GRIEVE, A. P. (1985). A Bayesian analysis of the two-period crossover designs for clinical trials. *Biometrics* **41**, 979–990.
- GRIEVE, A. P. (1986). Corrigenda to Grieve(1985). *Biometrics* **42**, 459.
- GRIEVE, A. P. (1994). Extending a Bayesian analysis of the two-period crossover to allow for baseline measurements. *Statistics in Medicine* **13**, 905–929.
- GRIZZLE, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics* **21**, 467–80.
- GRIZZLE, J. E. (1974). Corrigenda to Grizzle (1965). *Biometrics* **30**, 727.
- HAFNER, K. B. & KOCH, G. G. (1988). Some analysis strategies for three-period changeover designs with two treatments. *Statistics in Medicine* **7**, 471–481.
- HAGENAARS, J. A. (1978). Age, cohort and period: A general model for the analysis of social change. *The Netherlands' Journal of Sociology* **14**, 59–91.
- HEDAYAT, A. & AFSARINEJAD, K. (1978). Repeated measurement designs, II. *Annals of Statistics* **6**, 619–28.
- HILLS, M. & ARMITAGE, P. (1979). The two-period cross-over clinical trial. *British Journal of Clinical Pharmacology* **8**, 7–20.
- JAMES, A. T. & WILKINSON, G. N. (1971). Factorization of the residual operator and canonical decomposition of nonorthogonal factors in the analysis of variance. *Biometrika* **58**, 279–294.
- JOHNSON, N. L. & KOTZ, S. (1970). *Continuous Univariate Distributions - 2*. Houghton Mifflin, London.
- JONES, B. & KENWARD, M. G. (1987). Modelling binary data from a three-period crossover trial. *Statistics in Medicine* **6**, 555–564.
- JONES, B. & KENWARD, M. G. (1989). *Design and Analysis of Cross-over Trials*.

- Chapman and Hall, London, New York.
- KENWARD, M. G. & JONES, B. (1987a). The analysis of data from 2×2 cross-over trials with baselines measurements. *Statistics in Medicine* **10**, 911–26.
- KENWARD, M. G. & JONES, B. (1987b). A log-linear model for binary cross-over data. *Applied Statistics*. **36**, 192–204.
- KERSHNER, R. P. & FEDERER, W. T. (1981). Two-treatment crossover designs for estimating a variety of effects. *Journal of the American Statistical Association* **76**, 612–18.
- KOCH, G. G. (1972). The use of non-parametric methods in the statistical analysis of the two-period change-over design. *Biometrics* **28**, 901–914.
- KOCH, G. G., AMARA, I. A., BROWN, B. W., COLTON, T., & GILLINGS, D. B. (1989). A two-period crossover design for the comparing of two active treatments and placebo. *Statistics in Medicine* **8**, 487–504.
- KRZANOWSKI, W. J. & MARRIOTT, F. H. C. (1994). *Multivariate Analysis. Part 1: Distribution, ordering and inference*. Edward Arnold, New York - Toronto.
- KUPPER, L. L., JANIS, J. M., SALAMA, C. N., GREENBERG, Y. G., & GREENBERG, B. G. (1983). Age-period-cohort analysis: An illustration of the problem in assessing in one observation per cell data. *Communication in Statistics -Theory and Method* **12**, 2779–2807.
- LAIRD, N. M., SKINNER, J., & KENWARD, M. G. (1992). An analysis of two-period crossover designs with carryover effects. *Statistics in Medicine* **11**, 1967–1979.
- LASKA, E. & MEISNER, M. (1985). A variational approach to optimal two-treatment crossover designs: application to carryover effect models. *Journal of the American Statistical Association* **80**, 704–710.
- LASKA, E., MEISNER, M., & KUSHNER, H. B. (1983). Optimal crossover designs in the presence of carryover effect. *Biometrics* **39**, 1087–91.
- LASSERVE, B. W. (1991). Determination of optimal designs using linear models crossover trials. *Statistics In Medicine* **36**, 909–924.
- LAW, M. G. (1987). A Bayesian analysis of higher-order cross-over designs. *Unpublished MSc dissertation University of Kent at Canterbury*, 18–25.
- LAYARD, M. W. & ARVESEN, J. N. (1978). Analysis of Poisson data in crossover experimental designs. *Biometrics* **34**, 421–428.
- LEWIS, J. A. (1983). Clinical trials:statistical developments of practical benefit to the

- pharmaceutical industry (with discussion). *Journal of the Royal Statistical Society. A* **146**, 362–393.
- LINDLEY, D. V. & SMITH, A. F. M. (1972). Bayes Estimates for the Linear Model (with discussion). *Journal of the Royal Statistical Society B* **34**, 1–41.
- MATHEWS, J. N. S. (1988). Recent Developments in Crossover Designs. *International Statistical Review* **56**, 117–127.
- MILLIKEN, G. A. & JOHNSON, D. E. (1984). *Analysis of Messy Data, Volume 1: Designed Experiments*. Van Nostrand Reinhold, New York.
- PARKES, K. R. (1982). Occupational stress among student nurses: a natural experiment. *Journal of Applied Psychology* **67**, 784–796.
- PATEL, H. I. (1983). Use of baseline measurements in the two-period cross-over design. *Communication in Statistics - Theory and Methods* **12**, 269–2712.
- PATIL, V. H. (1965). Approximation to the Behrens-Fisher distributions. *Biometrika* **52**, 267–271.
- PRESCOTT, R. J. (1981). The comparison of success rates in cross-over trials in the presence of an order effect. *Applied Statistics* **30**, 9–15.
- RACINE, A., GRIEVE, A. P., FLUHLER, H., & SMITH, A. F. M. (1986). Bayesian methods in practice; experiences in the pharmaceutical industry. *Applied Statistics* **35**, 93–150.
- RAGHAVARAO, D. (1989). *Crossover designs in industry, in Design and Analysis of Experiments with Applications to Engineering and Physical Sciences* (ed. Ghosh), Marcel Dekker, New York.
- RAO, C. R. (1973). *Linear Statistical Inference and its Application*. 2nd edition. Wiley, New York.
- SCHAE, K. K. (1965). A general model for the study of development problem. *Psychological Bulletin* **64**, 92–107.
- SELWYN, M. R., DEMPSTER, A. R., & HALL, N. R. (1981). A Bayesian approach to bioequivalence for the 2×2 change-over design. *Biometrics* **37**, 11–21.
- SENN, S. J. (1988). Letter to the editor: Cross-over trials, carry-over effects and the art of self-delusion. *Statistics in Medicine* **7**.
- SENN, S. J. (1991). Problem with the two stage analysis of crossover trials. *British Journal of Clinical Pharmacology* **32**, 133.
- VENABLES, W. N. & REZAEI, S. (1996). Splus Program for James and Wilkinson's

Theorem.

- WALLENSTEIN, S. (1979). Inclusion of baseline values in the analysis of crossover designs (abstract). *Biometrics* **35**, 894.
- WALLENSTEIN, S. & FISHER, A. C. (1977). Analysis of the two-period repeated measurements crossover design with application to clinical trials. *Biometrics* **33**, 261–269.
- WILLAN, A. R. (1988). Using the maximum test statistic in the two-period crossover clinical trial. *Biometrics* **44**, 211–18.
- WILLAN, A. R. & PATER, J. L. (1986). Using baseline measurements in the two-period crossover clinical trial. *British Journal of Clinical Pharmacology* **7**, 282–289.
- ZIMMERMANN, H. & RAHLFS, W. (1978). Model building and testing for the change-over design. *Biometrical Journal* **22**, 197–220.