

5. Stratification to control prognostic variables

§5.1. Stratification and randomization schemes

When there are patient prognostic factors which are likely to affect the end-point measurements, these factors must be controlled. Stratification is a common method to use.

In stratification design, patients are stratified according to prognostic factors (usually more than one) into strata. Patients within strata are randomly assigned to receive different treatments under study.

The assignment is designed so that the treatments at each level of each prognostic factor are as balanced as possible.

Stratification is similar to randomized blocks design in that patients are classified according to

prognostic factors. But it is less strict than randomized blocks design. It does not require the strata to have the same size as the number of treatments as required by randomized block design.

- **Random permuted blocks within strata**

If number of strata is small and number of patients is big, the random permuted blocks within strata is an appropriate assignment scheme.

In this scheme, there are as many randomization lists as the number of strata. Each stratum has a separate randomization list. A random permuted blocks scheme is used to generate the randomization list.

This scheme achieves balance of treatments at each stratum, which automatically ensures the balance at each level of the prognostic factors.

Example 1: Randomization lists generated by random permuted blocks within strata for a trial in primary breast cancer (A =L-Pam, B=placebo).

Age	< 50	≥ 50	< 50	≥ 50
NPAN	1-3	1-3	≥ 4	≥ 4
	B	B	A	B
	A	B	A	A
	B	A	B	A
	A	A	B	B
	A	A	B	A
	B	A	A	B
	A	B	B	B
	B	B	A	A
	A	B	B	B
	B	A	A	B
	B	A	B	A
	A	B	A	A
	\vdots	\vdots	\vdots	\vdots

NPAN: No. of positive axillary nodes.

• Minimization method

If number of strata is large and number of patients is relatively small, as is the case in small scale trials, some strata might end up with less patients than the number of treatments.

Example 2: Distribution of patients across strata in an advanced breast cancer trial:

Performance	Age	DFI	Domi. metastatic lesion		
			Visceral	Osseous	Soft tissue
Ambu.	< 50	< 2	13	6	8
		≥ 2	1	0	1
	≥ 50	< 2	9	4	7
		≥ 2	5	0	7
Non-ambu.	< 50	< 2	3	0	2
		≥ 2	1	0	0
	≥ 50	< 2	7	4	0
		≥ 2	1	1	0

Two treatments are to be compared in this trial. But in many strata, there are only 1 patient.

The minimization method tries to balance the treatments at each level of prognostic factors.

In the implementation of the method, a card for each prognostic factor is used to record the number of patients assigned to each treatment at each level of the factor. There should be as many cards as the number of prognostic factors.

A patient is assigned to one of the treatment only at the time when he or she enters the trial.

Once a patient enters the trial, the number of patients on all cards at the levels corresponding to the new patient are added up for each treatment. The new patient is assigned to the treatment with the smallest summed

numbers. If there are several treatments having the smallest summed number, the new patient is assigned to one of these treatments at random.

Example 2 (cont.) Treatment assignments.

Factor	Level	No. on each treatment		Next patient
		A	B	
Performance status	Ambu.	30	31	←
	Non-ambu	10	9	
Age	< 50	18	17	←
	≥ 50	22	23	
Disease free interval	< 2	31	32	←
	≥ 2	9	8	
Dominant metastatic lesion	Visceral	19	21	←
	Osseous	8	7	
	Soft tissue	13	12	

The arrows in the table indicate the levels of the new patient at the four factors. The numbers of patients summed over these levels are:

$$\text{For A: } 30 + 18 + 9 + 19 = 76,$$

$$\text{For B: } 31 + 17 + 8 + 21 = 77.$$

The new patient is to be assigned to treatment A.

§5.2. Comparison of two treatments with stratification

The layout of data from a stratified comparison of two treatments is as follows:

Stratum	Treatment 1			Treatment 2		
	n	Mean	sd	n	Mean	sd
1	n_{11}	\bar{X}_{11}	s_{11}	n_{12}	\bar{X}_{12}	s_{12}
⋮						
a	n_{a1}	\bar{X}_{a1}	s_{a1}	n_{a2}	\bar{X}_{a2}	s_{a2}
⋮						
A	n_{A1}	\bar{X}_{A1}	s_{A1}	n_{A2}	\bar{X}_{A2}	s_{A2}

- **Analysis of treatment by stratum interaction**

The analysis of interaction is the first step.
Define

$$d_a = \bar{X}_{a1} - \bar{X}_{a2}, \quad w_a = \frac{n_{a1}n_{a2}}{n_{a1} + n_{a2}}.$$

Let

$$\bar{d} = \frac{\sum_{i=1}^A w_a d_a}{\sum_{i=1}^A w_a}.$$

The treatment by stratum interaction is measured by

$$\text{ISS} = \sum_{a=1}^A w_a (d_a - \bar{d})^2.$$

Let

$$s^2 = \frac{\sum_{a=1}^A \sum_{i=1}^2 (n_{ai} - 1) s_{ai}^2}{n_{..} - 2A},$$

where $n_{..} = \sum_{a=1}^A \sum_{i=1}^2 n_{ai}$.

The significance of the interaction is tested by the statistic given below:

$$F = \frac{\text{ISS}/(A - 1)}{s^2} \sim F_{A-1, n_{..} - 2A}.$$

Remark:

“The investigate is obliged to examine the data for evidence of certain interaction. For example, if the average difference between the treatments is in the direction favoring Treatment 1 but if there are one or more strata in which the difference is in the opposite direction, the investigator is obliged to try to identify those kinds of patients for whom Treatment 1 may be harmful. If, as another example, the average difference is close to zero but if difference is large and favors one of the treatments in one or more strata, the investigator is obliged to identify those kinds of patients who might benefit from that treatment.”

• Analysis of treatment mean difference

The significance of treatment difference is tested

by the statistic

$$F = \frac{\bar{d}^2 \sum_{a=1}^A w_a}{s^2} \sim F_{1, n.. - 2A}.$$

Remark:

1. In the absence of interaction, the treatment difference is the same across all strata, \bar{d} measures this difference.
2. In general, \bar{d} provides an approximated measurement of the mean difference between the two treatment over the population of the patients.

The estimate of the mean responses expected to Treatment 1 and 2 in the entire population of patients is

$$\bar{D} = \frac{\sum_{a=1}^A n_{a\cdot} (\bar{X}_{a1} - \bar{X}_{a2})}{n..},$$

where $n_{a\cdot} = n_{a1} + n_{a2}$.

When the treatments are balanced within strata, $w_a \approx n_a./4$, hence $\bar{d} \approx \bar{D}$.

3. This mean difference is scientifically meaningful, for example, in the development of a new drug, of interest is the mean difference of the drug from a standard drug over the population of patients.
 4. When the patients are over-stratified, where minimization method is used for patient assignment, the above method of analysis cannot be applied. The data is then analyzed by multiple regression method.
- **Example 3:** In a study comparing two methods for treating children with moderate neurological problems, the response is a measure of neurological functioning obtained eight weeks after the start of the treatment, the children were stratified according to their sex (it is known

males have poorer functioning than females) and their parent's social class (children from poorer family were expected to function worse). The summary data is in the following table:

	Social class	Sex	Treatment 1			Treatment 2		
			n	\bar{X}	s	n	\bar{X}	s
1	L	F	41	1.38	0.22	40	1.36	0.28
2	L	M	41	1.26	0.25	38	1.28	0.19
3	M	F	33	1.51	0.31	35	1.41	0.27
4	M	M	45	1.46	0.28	46	1.39	0.33
5	H	F	18	1.61	0.34	20	1.51	0.41
6	H	M	23	1.59	0.46	23	1.44	0.30

From the table, it is computed that

$$\bar{d} = 0.059, \quad s^2 = 0.0883, \quad \text{ISS} = 0.2963.$$

$$\sum w_a = 100.6747, \quad A = 6, \quad n_{..} = 403.$$

The F -statistic for the interaction is computed as

$$F = \frac{\text{ISS}/(A - 1)}{s^2} = \frac{0.2963/5}{0.0883} = 0.67,$$

which is non-significant compared with $F_{5,391,0.05}$.

The F -statistic for the mean difference is computed as

$$F = \frac{\bar{d}^2 \sum_{a=1}^A w_a}{s^2} = \frac{0.059^2 \times 100.6747}{0.0883} = 3.97,$$

which is significant compared with $F_{1,391,0.05}$.

§5.3. Comparison of more than two treatments

The layout of data:

Stratum	Treatment 1			Treatment 2			...	Treatment g		
	n	Mean	sd	n	Mean	sd	...	n	Mean	sd
1	n_{11}	\bar{X}_{11}	s_{11}	n_{12}	\bar{X}_{12}	s_{12}	...	n_{1g}	\bar{X}_{1g}	s_{1g}
⋮										
a	n_{a1}	\bar{X}_{a1}	s_{a1}	n_{a2}	\bar{X}_{a2}	s_{a2}	...	n_{ag}	\bar{X}_{ag}	s_{ag}
⋮										
A	n_{A1}	\bar{X}_{A1}	s_{A1}	n_{A2}	\bar{X}_{A2}	s_{A2}	...	n_{Ag}	\bar{X}_{Ag}	s_{Ag}

- **Measure of total treatment effect across strata**

Let

$$\mathbf{c}_j = (c_1^{(j)}, c_2^{(j)}, \dots, c_g^{(j)}), \quad j = 1, \dots, g-1,$$

be $g - 1$ linearly independent contrast vectors.

Let

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_{g-1} \end{bmatrix}.$$

Let $\mathbf{x}_a = (\bar{X}_{a1}, \bar{X}_{a2}, \dots, \bar{X}_{ag})'$ and

$$\mathbf{d}_a = \mathbf{C}\mathbf{x}_a, \quad a = 1, \dots, A.$$

The vector \mathbf{d}_a consists of $g - 1$ contrasts within stratum a . Let

$$\Lambda_a = \frac{\text{Var}(\mathbf{x}_a)}{s^2} = \text{Diag}\left(\frac{1}{n_{a1}}, \dots, \frac{1}{n_{ag}}\right).$$

Then

$$\begin{aligned}\text{Var}(\mathbf{d}_a) &= \mathbf{C}\text{Var}(\mathbf{x}_a)\mathbf{C}' = s^2\mathbf{C}\mathbf{\Lambda}\mathbf{C}' \\ &= s^2\mathbf{V}_a, \text{ say.}\end{aligned}$$

Let

$$\mathbf{W}_a = \mathbf{V}_a^{-1}.$$

Then the treatment effect across all the strata is measured by the following “weighted average” of the effects within strata:

$$\bar{\mathbf{d}} = \left[\sum_{a=1}^A \mathbf{W}_a \right]^{-1} \sum_{a=1}^A \mathbf{W}_a \mathbf{d}_a.$$

Let

$$\mathbf{S} = s^2 \left[\sum_{a=1}^A \mathbf{W}_a \right]^{-1}.$$

It can be shown that

$$\text{Var}(\bar{\mathbf{d}}) = \mathbf{S}.$$

The variation (standardized) due to the weighted average effects of the treatments is then given by

$$\bar{\mathbf{d}}' S^{-1} \bar{\mathbf{d}} \equiv \text{TSS}/s^2,$$

where TSS stands for treatment sum of squares, or un-standardized variation due to the weighted average effects of the treatments. The significance of the weighted average effects is tested by

$$F = \frac{\text{TSS}/(g - 1)}{s^2} \sim F_{g-1, n.. - gA}.$$

- **Measure of interaction effect**

The raw measure of interaction is

$$\mathbf{d}_a - \bar{\mathbf{d}}, \quad a = 1, \dots, A.$$

The variation due to interaction effect is measured by

$$\text{ISS} = \sum_{a=1}^A (\mathbf{d}_a - \bar{\mathbf{d}})' \mathbf{W}_a (\mathbf{d}_a - \bar{\mathbf{d}}).$$

The significance of interaction is tested by

$$F = \frac{\text{ISS}/(g-1)(A-1)}{s^2} \\ \sim F_{(g-1)(A-1), n..-gA}.$$

- **Invariance of TSS and ISS**

The TSS and ISS are invariant with respect to the contrast matrix \mathbf{C} ; that is, for any other contrast matrix \mathbf{C}^* , the TSS^* and ISS^* derived from \mathbf{C}^* are the same as TSS and ISS.

- **Multiple comparison**

Multiple comparison among the treatment effects can be carried out similarly to that in

parallel groups designs.

For any particular contrast C of the treatment means, it can be expressed as $C = \mathbf{c}'\bar{\mathbf{d}}$ for some vector \mathbf{c} .

The test statistic for this contrast is given by

$$L = \frac{C}{\sqrt{\mathbf{c}'S\mathbf{c}}} \sim t_{(g-1)(A-1)}.$$

L is then compared with the critical value of either Scheffe, Tukey, Dunnett or Bonferroni criterion according to the nature of the comparison.

- **Analysis using R**

The analysis can be realized by using the R function `lm`. It is illustrated by an example.

Example: Changes in systolic blood pressure associated with four treatments in a strat-

ified study. The following table gives summary data in each stratum.

Stratum		Treatment			
		1	2	3	4
1	n	6	5	3	5
	Mean	29.33	28.00	16.33	13.60
	sd	13.02	10.98	14.19	10.55
2	n	4	4	5	6
	Mean	28.25	33.50	4.40	12.83
	sd	5.85	2.08	6.91	10.34
3	n	5	6	4	5
	Mean	20.40	18.17	8.50	14.20
	sd	13.37	12.53	9.00	8.93

It is computed from the table that $s^2 = 110.4564$.

The following codes are used to generate the data to be used by `lm`.

```
x=c(29.33333,28.25,20.4,  
    28,33.5,18.166666667,  
    16.333333333,4.4,8.5,  
    13.6,12.833333333,14.2)  
n=c(6,4,5,5,4,6,3,5,4,5,6,5)  
S=factor(rep(c(1,2,3),4))  
T=factor(c(rep(1,3),rep(2,3),rep(3,3),rep(4,3)))
```

The data generated is of the following form:

```
      x S T n  
[1,] 29.33333 1 1 6  
[2,] 28.25000 2 1 4  
[3,] 20.40000 3 1 5  
[4,] 28.00000 1 2 5  
[5,] 33.50000 2 2 4  
[6,] 18.16667 3 2 6  
[7,] 16.33333 1 3 3  
[8,]  4.40000 2 3 5  
[9,]  8.50000 3 3 4  
[10,] 13.60000 1 4 5  
[11,] 12.83333 2 4 6  
[12,] 14.20000 3 4 5
```

The following code specifies the type of contrast:

```
options(contrasts=c("contr.treatment","contr.poly"))
```

With this option, the contrasts of the treatment effects are:

$$\mu_j - \mu_1, \quad j = 2, 3, 4.$$

The following code fit a linear model to the data and produces an anova table:

```
L1=lm(x~S*T, x=T, weights=n)  
anova(L1)
```

Note that the order of S and T appearing in the formula argument of `lm` is important. The SS's are computed in such a sequential way that the SS for the first factor is computed as if it is the only factor in the model and the SS for the second factor is computed within each level of the first factor in the manner discussed above.

The anova table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
S	2	488.64	244.32		
T	3	3063.43	1021.14		
S:T	6	707.27	117.88		
Residuals	0	0.00			

This anova table does not have a residual sum of squares since the model fitted using the summary data is saturated.

To get the F -ratios, the mean squares should be divided by s^2 .

Had the model been fitted to the raw data, s^2 would be given by the mean residual sum of squares.

To obtain the variance-covariance matrix of $\bar{\mathbf{d}}$, an additive model should be fitted. The matrix can be obtained from the variance matrix of the fitted coefficients which can be extracted from the fitted object.

```
L2=lm(x~S+T, x=T, weights=n)
summary(L2)
anova(L2)
V=vcov(L2)
```

summary extracts the following information:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	28.7818	3.3506	8.590	0.000137	***
S2	-2.1391	3.5568	-0.601	0.569572	
S3	-6.4341	3.4874	-1.845	0.114589	
T2	-0.1044	3.9713	-0.026	0.979881	
T3	-16.9958	4.2387	-4.010	0.007041	**
T4	-12.4690	3.9181	-3.182	0.019017	*

anova extracts the anova table associated with the additive model:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
S	2	488.64	244.32	2.0727	0.20685
T	3	3063.43	1021.14	8.6627	0.01338 *
Residuals	6	707.27	117.88		

Note: the SS of the interaction becomes residual SS in this model.

`vcov` extracts the variance-covariance matrix of the fitted coefficients:

	Int	S2	S3	T2	T3	T4
Int	11.226471	-5.4680063	-5.7294657	-7.4765494	-7.038313	-7.3855103
S2	-5.468006	12.6511326	6.2831128	-0.4188742	-1.897670	-1.2396412
S3	-5.729466	6.2831128	12.1619069	-0.8107938	-0.942467	-0.4272975
T2	-7.476549	-0.4188742	-0.8107938	15.7710804	7.921345	7.8870002
T3	-7.038313	-1.8976699	-0.9424669	7.9213449	17.966306	8.0444599
T4	-7.385510	-1.2396412	-0.4272975	7.8870002	8.044460	15.3512628

The following code extracts the variance-covariance of the estimated parameters associated with treatments (i.e., T2,T3,T4):

```
VT=V[4:6,4:6]
      T2      T3      T4
T2 15.771080  7.921345  7.88700
T3  7.921345 17.966306  8.04446
T4  7.887000  8.044460 15.35126
```

Note: This matrix is obtained by replacing s^2 with the mean residual sum of squares of the additive model. To get $\text{Var}(\bar{\mathbf{d}})$, this matrix must be divided by this mean residual sum of square and then multiplied by s^2 .

```

VT/117.88*110.4564
          T2          T3          T4
T2 14.777882  7.422491  7.390309
T3  7.422491 16.834862  7.537853
T4  7.390309  7.537853 14.384503

```

Note: this matrix differs from what is in the text book by Fleiss since the following different contrasts are used there: $\mu_j - \mu_4, j = 1, 2, 3$.

It is interesting to notice a relationship between Wald statistics and sum of squares. In the additive model, the Wald statistic for the coefficients of the treatments and strata equal their respective sum of squares divided by the mean residual's sum of squares of the model. It is also the case for interaction sum of squares in the interaction model fitted as L1. This is illustrated by TSS below:

```

b=L2$coef
bt=b[4:6]
t(bt)%*%solve(Vt)%*%bt*117.88

```

The last line of the code produces the number 3063.492 which differs from the TSS in the anova table only by rounding errors.

• Analysis using linear models

The stratified design model with treatment by stratum interaction can be described by a linear model:

$$\begin{aligned}
 X = & \mu_0 + \sum_{a=2}^A \alpha_a s_a + \sum_{k=2}^g \beta_k t_k \\
 & + \sum_{a=2}^A \sum_{k=2}^g \gamma_{ak} s_a t_k + \epsilon,
 \end{aligned}$$

where

$$\begin{aligned}
 s_a &= \begin{cases} 1, & \text{if in stratum } a, \\ 0, & \text{otherwise.} \end{cases} \\
 t_k &= \begin{cases} 1, & \text{if treatment } k, \\ 0, & \text{otherwise.} \end{cases}
 \end{aligned}$$

The parameters are interpreted as follows:

μ_0 : mean response in stratum 1 with treatment 1;

$\mu_0 + \beta_j$: mean response in stratum 1 with treatment j ;

$\mu_0 + \alpha_i + \beta_j + \gamma_{ij}$: mean response in stratum i with treatment j ;

β_j : difference of mean responses between treatment j and treatment 1 in strata 1;

$\beta_j - \beta_l$: difference of mean responses between treatment j and treatment l in strata 1;

$(\beta_j + \gamma_{ij}) - (\beta_l + \gamma_{il})$: difference of mean responses between treatment j and treatment l in strata i .

Some inference issues

1. Test statistic for the significance of interaction can be obtained either from the anova

table of the interaction model or from the Wald statistic on the interaction parameters. The F -statistic equal the Wald statistic divided by $(g - 1)(A - 1)$.

In the example,

$$F_{\text{ISS}} = \frac{707.2663/6}{110.4564} = 1.07 < F_{6,46,0.05}.$$

The interaction is not significant.

2. If interaction effect is significant, it is desirable to test whether certain contrasts are significant within stratum and how they differ across strata. Such contrasts can be formed from the estimates of the treatment differences explained above.

For instance, if difference between treatment j and l is of interest, the test statistic is

given by

$$\frac{(\hat{\beta}_j + \hat{\gamma}_{ij}) - (\hat{\beta}_l + \hat{\gamma}_{il})}{\sqrt{\text{Var}((\hat{\beta}_j + \hat{\gamma}_{ij}) - (\hat{\beta}_l + \hat{\gamma}_{il}))}},$$

where the variance can be obtained directly from the variance-covariance matrix of the fitted coefficients.

3. If interaction is non-significant, the F-statistic for the significance of treatment effects can still be obtained from the anova table of the interaction model.

In the example,

$$F_{\text{TSS}} = \frac{3063.4317/3}{110.4564} = 9.24 > F_{3,46,0.001}.$$

Statistically significant differences at level 0.001 exists among the four treatment means.

4. When the F-test for treatment effect is significant, a search for the sources of signif-

ificance is in order. But to make multiple comparison, an additive model should be fitted and inference be based on the fitted additive model. The interpretation of the model parameters remains the same when the terms associated with interactions are dropped. For example, the difference of mean responses between treatment j and l becomes $\beta_j - \beta_l$.

To test the significance of a contrast, say, $\beta_j - \beta_l$, the test statistic is given by

$$\frac{(\hat{\beta}_j - \hat{\beta}_l)}{\sqrt{\text{Var}(\hat{\beta}_j - \hat{\beta}_l)}},$$

where the variance can be obtained from the variance-covariance matrix of the additive model and then divided by the mean residual sum of squares of the additive model and multiplied by s^2 .

In the example, consider the difference between treatment 2 and 3, the estimate of the difference is

$$\begin{aligned}\hat{\beta}_2 - \hat{\beta}_3 &= -0.1043934 - (-16.99579) \\ &= 16.8914.\end{aligned}$$

The variance of the difference obtained from the model variance matrix is 17.8947. The modified variance is

$$17.8947 \times 110.4564 / 117.88 = 16.7678.$$

The test statistic is then

$$L = 16.8914 / \sqrt{16.7678} = 4.125,$$

which is significant.