Handout 4

Randomized Block Design

This handout is on randomized block design. We will not learn any new statistical method here, but the concepts are very important. Let us refer to the data on "Auditor training". There are three training methods for the auditors and the response Y is a proficiency score after the training are completed. Ideally we would like to compare the three methods among those who are as similar as possible in their educational background. How do we achieve this? One way is compare the three different training methods among those whose time since graduation from college are about the same. Suppose then we have ten such groups (of three individuals each). Group 1 consists of those who graduated recently, group 2 people graduated between 1 and two years ago, and group 10 consists of those who graduated some time in the past (say, ten years or more). Time since graduation is called the block (or a blocking factor) and treatment is the training method. Methodologically speaking, we have a two factor model with one observation for each combinations of the factors. So the analysis is exactly the same as in chapter 20 of App. Lin. Stat. Model, if we identify block (time since graduation) as factor A and the training method (treatment) as factor B. However, the notations given in your text are somewhat different from chapter 20. The number of levels of the block is $n_b = 10$ and the number of levels for the treatment factor (training) is $r = 3$.

Let Y_{ij} be the score of the person in the i^{th} block assigned to training method j. The model is

$$Y_{ij} = \mu_{\cdot\cdot} + \rho_i + \tau_j + \varepsilon_{ij},\ j = 1, ..., r, i = 1, ..., n_b,$$

where ε_{ij}'s are independent $N(0, \sigma^2)$, $\sum \rho_i = 0$ and $\sum \tau_j = 0$. Here $\mu_{\cdot\cdot}$ is the overall mean, ρ_i is the block effect and τ_j is the treatment effect. Once again note that this is an additive model with one observation for each combination of the factors. In fact if we think of ρ_i as α_i and τ_j as β_j, then all the analysis including Tukey's test for interactions are exactly the same as given in chapter 20

Here the total number of observations is $n_T = n_b r$. Let us follow the usual notations

$$\bar{Y}_{\cdot\cdot} = \sum\sum Y_{ij}/(n_b r),\ \bar{Y}_{i\cdot} = \frac{1}{r}\sum_{j=1}^{r} Y_{ij},\ \bar{Y}_{\cdot j} = \frac{1}{n_b}\sum_{i=1}^{n_b} Y_{ij}.$$

Parameter estimates, the fitted values and the residuals are

$$\begin{aligned}
&\hat{\mu}_{\cdot\cdot} = \bar{Y}_{\cdot\cdot}, \hat{\rho}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot}, \hat{\tau}_j = \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot},\\
&\hat{Y}_{ij} = \hat{\mu}_{\cdot\cdot} + \hat{\rho}_i + \hat{\tau}_j = \bar{Y}_{i\cdot} + \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot},\ \ e_{ij} = Y_{ij} - \hat{Y}_{ij}.
\end{aligned}$$

The sums of squares and the mean squares are

$$\begin{aligned}
SSTO &= \sum\sum (Y_{ij} - \bar{Y}_{\cdot\cdot})^2,\ df = n_b r - 1,\\
SSBL &= \sum\sum \hat{\rho}_i^2 = r\sum \hat{\rho}_i^2,\ \ df = n_b - 1,\ \ MSBL = \frac{SSBL}{n_b - 1}\\
SSTR &= \sum\sum \hat{\tau}_j^2 = n_b \sum \hat{\tau}_j^2,\ \ df = r - 1,\ \ MSTR = \frac{SSTR}{r - 1},\\
SSE &= \sum\sum e_{ij}^2,\ \ \ df = (n_b - 1)(r - 1),\ \ MSE = \frac{SSE}{(n_b - 1)(r - 1)}.
\end{aligned}$$

As usual the following identities hold

$$SSTO = SSBL + SSTR + SSE,$$
$$df(SSTO) = df(SSBL) + df(SSTR) + df(SSE).$$

Important: In your text SSE is denoted by $SSBL.TR$ and MSE is denoted by $MSBL.TR$.

Fact: (a) $E(MSE) = \sigma^2$, (b) $E(MSBL) = \sigma^2 + \frac{r\sum \rho_i^2}{n_b - 1}$, (c) $E(MSTR) = \sigma^2 + \frac{n_b \sum \tau_j^2}{r-1}$.

Note that MSE estimates σ^2.

F-tests:

If we want to test $H_0 : \rho_i = 0$ for all i, against H_1 : not all ρ_i are zero, at level α, the F statistic is $F^* = \frac{MSBL}{MSE}$. The decision rule is: reject H_0 if $F^* > F(1-\alpha; n_b - 1, (n_b - 1)(r-1))$.
If we want to test $H_0 : \tau_j = 0$ for all j, against H_1 : not all τ_j are zero, at level α, the F statistic is $F^* = \frac{MSTR}{MSE}$. The decision rule is: reject H_0 if $F^* > F(1-\alpha; r-1, (n_b - 1)(r-1))$.
[Note that the analysis of treatment effects is usually of primary interest.]

Auditor training:

An accounting firm, prior to introducing in the firm widespread training in statistical sampling for auditing, tested three training methods: (1) study at home with programmed training materials, (2) training sessions at local offices conducted by local staff, and (3) training session in Chicago conducted by national staff. Thirty auditors were grouped into 10 blocks of three: according to time elapsed since college graduation, and the auditors in each block were randomly assigned to the three training methods. At the end of the training, each auditor was asked to analyze a complex case involving statistical applications; a proficiency measure based on this analysis was obtained for each auditor. The results were (block 1 consists of auditors graduated most recently, block 10 consist of those graduated most distantly):

Block	Training method (j)			Block	Training method (j)		
i	$j=1$	$j=2$	$j=3$	i	$j=1$	$j=2$	$j=3$
1	73	81	92	6	73	75	86
2	76	78	89	7	68	72	88
3	75	76	87	8	64	74	82
4	74	77	90	9	65	73	81
5	76	71	88	10	62	69	78

Here

$$\bar{Y}_{..} = 77.1, \bar{Y}_{.1} = 70.6, \bar{Y}_{.2} = 74.6, \bar{Y}_{.3} = 86.1,$$
$$\hat{\mu}_{..} = 77.1, \hat{\tau}_1 = -6.5, \hat{\tau}_2 = -2.5, \hat{\tau}_3 = 9.0.$$

ANOVA table

Source	df	SS	MS	F	p-value
Block	$n_b - 1 = 9$	433.37	48.15	7.72	0.000
Treatment	$r - 1 = 2$	1295.0	647.50	103.75	0.000
Error	$(n_b - 1)(r-1) = 18$	112.33	6.24		
Total	$n_b r - 1 = 29$	1840.70			

The training methods seem to be different. It also seems that block effect is present.

Let us now compare the three training methods by constructing simultaneous 95% confidence intervals for $D_1 = \tau_1 - \tau_2$, $D_2 = \tau_1 - \tau_3$ and $D_3 = \tau_2 - \tau_3$ using Tukey's method. Estimates are

$$\begin{aligned}
\hat{D}_1 &= \bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 2} = 70.6 - 74.6 = -4.0,\\
\hat{D}_2 &= \bar{Y}_{\cdot 1} - \bar{Y}_{\cdot 3} = 74.6 - 86.1 = -11.5,\\
\hat{D}_3 &= \bar{Y}_{\cdot 2} - \bar{Y}_{\cdot 3} = 70.6 - 86.1 = -15.5.
\end{aligned}$$

Estimate of $\sigma^2(\hat{D}_1)$ is $s^2(\hat{D}_1) = \frac{2}{n_b} MSE = \frac{2}{10}(6.2406) = 1.2481$. So $s(\hat{D}_1) = 1.1172$. Here $s^2(\hat{D}_1) = s^2(\hat{D}_2) = s^2(\hat{D}_3) = \frac{2}{n_b} MSE$. Hence we have $s(\hat{D}_2) = s(\hat{D}_3) = 1.1172$.

From the table of Studentized Range distribution we get $q(1-\alpha; r, (n_b - 1)(r-1)) = q(0.95; 3, 18) = 3.61$. So $T = \frac{1}{\sqrt{2}} q(.95; 3, 18) = 2.5525$. So simultaneous 95% confidence intervals are

$$\begin{aligned}
D_1 &: \ \hat{D}_1 \pm T s(\hat{D}_1), \text{ i.e., } -4.0 \pm (2.5525)(1.1172), \text{ i.e., } -4.0 \pm 2.85, \text{ i.e., } (-6.85, -1.15),\\
D_2 &: \ \hat{D}_2 \pm T s(\hat{D}_2), \text{ i.e., } -11.5 \pm (2.5525)(1.1172), \text{ i.e., } -11.5 \pm 2.85, \text{ i.e., } (-14.35, -8.65),\\
D_3 &: \ \hat{D}_3 \pm T s(\hat{D}_3), \text{ i.e., } -15.5 \pm (2.5525)(1.1172), \text{ i.e., } -15.5 \pm 2.85, \text{ i.e., } (-18.35, -12.65).
\end{aligned}$$

None of these intervals include zero. So we may conclude that the three training methods are different from each other.

Tukey's test for additivity:

Recall that Tukey's interaction model is

$$Y_{ij} = \mu_{\cdot\cdot} + \rho_i + \tau_j + D\rho_i\tau_j + \varepsilon_{ij}.$$

According to this model, when $D = 0$ we have no interaction. So we would like to test $H_0 : D = 0$ against $H_1 : D \neq 0$. Note that

$$\begin{aligned}
\sum\sum Y_{ij}\hat{\rho}_i\hat{\tau}_j &= -48.667,\\
\sum \hat{\rho}_i^2 &= \frac{SSBL}{r} = \frac{43.37}{3}, \sum \hat{\tau}_j^2 = \frac{SSTR}{n_b} = \frac{1295}{10},\\
\hat{D} &= \frac{\sum\sum Y_{ij}\hat{\rho}_i\hat{\tau}_j}{\sum \hat{\rho}_i^2 \sum \hat{\tau}_j^2} = -0.0260,\\
SSBLTR^* &= \sum\sum (\hat{D}\hat{\rho}_i\hat{\tau}_j)^2 = \hat{D}^2 \sum \hat{\rho}_i^2 \sum \hat{\tau}_j^2 = 1.2651,\\
SS\,\mathrm{Re}\,m^* &= SSTO - SSBL - SSTR - SSBLTR^*\\
&= 1840.70 - 433.37 - 1295.00 - 1.2651 = 111.0649,\\
F^* &= \frac{MSBLTR^*}{MS\,\mathrm{Re}\,m^*} = \frac{\{1.2651/1\}}{\{111.0649/(29 - 9 - 2 - 1)\}} = 0.1936.
\end{aligned}$$

Degrees of freedom for this F-test are $(1, 17)$. It can be checked that p-value here is larger than 0.50. Hence we cannot reject H_0. So we can conclude that use of an additive model is reasonable here.

What happens if we ignore the blocks?

When we ignore the blocks we get a one-factor study with training method as the factor. This is called a completely randomized study. If the block effects are ignored then the ANOVA table is

Source	df	SS	MS	F	p-value
Treatment	$r-1=2$	1295.00	647.50	32.04	0.000
Error	27	545.70	20.21		
Total	29	1840.70			

[Note that the residual sum of squares here is equal to 433.37+112.33=545.70]

The conclusion still is that the treatment is significant. However, the value of the F-statistic for treatment effects is substantially lower.

Efficiency of blocking is defined to be equal to

$$E = \frac{\text{variance of } \varepsilon\text{'s for completely randomized design}}{\text{variance of } \varepsilon\text{'s in randomized block design}}.$$

Larger the value of E, clearer is the need for using a randomized block design. An estimate of E is given by

$$\hat{E} = \frac{MSE \text{ for completely randomized design}}{\text{MSE for randomized block design}} = \frac{20.21}{6.24} = 3.29.$$

It can be shown that

$$\begin{aligned}\hat{E} &= \frac{\{(n_b-1)MSBL + (n_b-1)(r-1)MSE_{RBD}\}/\{(n_b-1)r\}}{MSE_{RBD}} \\ &= w\frac{MSBL}{MSE_{RBD}} + (1-w), \text{ where } w = \frac{n_b-1}{(n_b-1)r},\end{aligned}$$

and MSE_{RBD} is the MSE for the randomized block design. So $\hat{E}$ is a weighted average of $\frac{MSBL}{MSE_{RBD}}$ and 1. Hence $\hat{E} \geq 1$ when and only when $\frac{MSBL}{MSE_{RBD}} \geq 1$. Also note that if block effects exist, then $\frac{MSBL}{MSE_{RBD}}$ is much larger than 1 and hence $\hat{E}$ is substantially larger than 1 and vice-versa.

It is possible to obtain a better estimate of E than $\hat{E}$. Define

$$\begin{aligned}df_1 &= df(\text{MSE}) \text{ for randomized block design} = 18, \\ df_2 &= df(\text{MSE}) \text{ for completely randomized design} = 27.\end{aligned}$$

Then the improved estimate of E is given by

$$\hat{E}' = \frac{(df_2+1)(df_1+3)}{(df_2+3)(df_1+1)}\hat{E} = \frac{(27+1)(18+3)}{(27+3)(18+1)}(3.29) = 3.39.$$

In this case, the value of $\hat{E}'$ is not really all that different from that of $\hat{E}$.

Figure 1: Auditor Training Data