Handout 11

Repeated Measures and Split-plot designs.

This handout lists a few more repeated measures designs which are a bit more complicated than the ones in Handout 6. In addition it also presents the basic ideas behind Split-Plot designs, which are quite useful in practice

Repeated measures designs: Two-factors with repeated measures on one factor (repeated measures on factor A, but not on B)

A national retail chain wanted to study the effects of two advertising campaigns (factor A, fixed) on the volume of sales of athletic shoes over time (factor B). Ten similar markets (subjects, S) were chosen at random to participate in this study. The two advertising campaigns (A_1 and A_2) were similar in all respects except that a different national sports personality was used in each. Sales data (Y) were collected for three weeks for three two-week periods (B_1: two weeks prior to campaign; B_2: two weeks during which campaign occurred; B_3: two-weeks after campaign was concluded). The experiment was conducted conducted during a six-week period when sales of athletic shoes are quite stable

Advertising	Test	Time period (B)		
Campaign (A)	Market (S)	$k=1$	$k=2$	$k=3$
	$i=1$	958	1047	933
	$i=2$	1005	1122	986
$j=1$	$i=3$	351	436	339
	$i=4$	549	632	512
	$i=5$	730	784	707
	$i=1$	780	897	718
	$i=2$	229	275	202
$j=2$	$i=3$	883	964	817
	$i=4$	624	695	599
	$i=5$	375	436	351

Note that test market (subject, random) is nested in advertising campaign (factor A). Factors A (advertising campaign), time period (factor B) are crossed. The model is

$$Y_{ijk} = \mu_{...} + \rho_{i(j)} + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk}, i = 1, \ldots, s, j = 1, \ldots, a, k = 1, \ldots, b,$$

where

(a) $\mu_{...}$ is the overall mean,

(b) α_j's are main effect of factor A, $\sum \alpha_j = 0$

(c) β_k's are main effects of factor B, $\sum \beta_k = 0$

(d) $(\alpha\beta)_{jk}$'s are the interactions (between A and B), $\sum_j (\alpha\beta)_{jk} = 0$ for each k, $\sum_k (\alpha\beta)_{jk} = 0$ for each j,

(e) $\rho_{i(j)}$'s are subject effect nested in A are iid $N(0, \sigma_\rho^2)$,

(f) ε_{ijk}'s are iid $N(0, \sigma^2)$,

(g) $\{\rho_{i(j)}\}$ and $\{\varepsilon_{ijk}\}$ are independent.

Here we have

$$
\begin{aligned}
E(Y_{ijk}) &= \mu_{\cdots} + \alpha_j + \beta_k + (\alpha\beta)_{jk}, \\
Var(Y_{ijk})) &= \sigma_\rho^2 + \sigma^2, \\
Cov(Y_{ijk}, Y_{ijk'}) &= \sigma_\rho^2, \quad k = k', \\
Cov(Y_{ijk}, Y_{i'j'k'}) &= 0, \ i \neq i' \text{ and/or } j \neq j'.
\end{aligned}
$$

Important: The sums of squares, df's, mean squares are calculated as if all the three factors are fixed (not random). However, the expected values of the sums of squares are different from a design in which the factor effects are fixed. The expected values reflect if there are some random factors..

Estimates are

$$
\begin{aligned}
&\hat{\mu}_{\cdots} = \bar{Y}_{\cdots}, \ \hat{\alpha}_j = \bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdots}, \ \hat{\beta}_k = \bar{Y}_{\cdot\cdot k} - \bar{Y}_{\cdots}, \\
&\widehat{(\alpha\beta)}_{jk} = \bar{Y}_{\cdot jk} - \bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdot\cdot k} + \bar{Y}_{\cdots}, \hat{\rho}_{i(j)} = \bar{Y}_{ij\cdot} - \bar{Y}_{\cdot j \cdot}
\end{aligned}
$$

Thus the fitted Y values and residuals are

$$
\begin{aligned}
\hat{Y}_{ijk} &= \hat{\mu}_{\cdots} + \hat{\rho}_{i(j)} + \hat{\alpha}_j + \hat{\beta}_k + \widehat{(\alpha\beta)}_{jk} = \bar{Y}_{\cdot jk} + \bar{Y}_{ij\cdot} - \bar{Y}_{\cdot j \cdot}. \\
e_{ijk} &= Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - (\bar{Y}_{\cdot jk} + \bar{Y}_{ij\cdot} - \bar{Y}_{\cdot j \cdot}).
\end{aligned}
$$

The expressions for SSTO, SSA, SSB, SSAB remain the same as before

$$
\begin{aligned}
SSTO &= \sum\sum\sum (Y_{ijk} - \bar{Y}_{\cdots})^2, \ df = sab - 1, \\
SSA &= \sum\sum\sum \hat{\alpha}_j^2, df = a - 1, \\
SSB &= \sum\sum\sum \hat{\beta}_k^2, \ df = b - 1, \\
SSAB &= \sum\sum\sum \widehat{(\alpha\beta)}_{jk}^{\,2}, df = (a-1)(b-1).
\end{aligned}
$$

The expressions for subjects (SSS(A)) and SSE are

$$
\begin{aligned}
SSS(A) &= \sum\sum\sum (\bar{Y}_{ij\cdot} - \bar{Y}_{\cdot j \cdot})^2, df = a(s-1), \\
SSE &= \sum\sum\sum e_{ijk}^2, df = a(s-1)(b-1).
\end{aligned}
$$

Note that SSE has been called SSB.S(A) in the textbook.

[**The df for SSE** equals total number of observations minus the number of parameters estimated (i.e., α's, β's, $(\alpha\beta)$'s and $\rho_{i(j)}$'s)].

The decomposition of SSTO also holds

$$SSTO = SSA + SSB + SSAB + SSS(A) + SSE.$$

Each mean square is defined in the usual manner, i.e., sum of square divided by its df. The following table given the expected values of the mean squares

Source	df	SS	E(MS)
A	a-1	SSA	$\sigma^2 + b\sigma_\rho^2 + bs\sum\alpha_j^2/(a-1)$
B	b-1	SSB	$\sigma^2 + bs\sum\beta_k^2/(b-1)$
AB	(a-1)(b-1)	SSAB	$\sigma^2 + s\sum\sum(\alpha\beta)_{jk}/[(a-1)(b-1)]$
Subject, nested in A	a(s-1)	SSS(A)	$\sigma^2 + b\sigma_\rho^2$
Error	a(s-1)(b-1)	SSE	σ^2
Total	sab-1	SSTO	

ANOVA table.(A, fixed; S(A) (i.e., subject nested in A); random; B fixed);$a = 2, b = 3, s = 5$.

Source	df	SS	MS	F	p-val
A	$a-1=1$	168151	168151	$MSA/MSS(A) = 0.73$	0.417
B	$b-1=2$	67073	33537	$MSB/MSE = 93.69$	0.000
A*B	$(a-1)(b-1)2$	391	196	$MSAB/MSE = 0.55$	0.589
S(A)	$a(s-1)=8$	1833681	229210	$MSS(A)/MSE = 640.31$	0.000
Error	$a(s-1)(b-1)=16$	5727	358		
Total	$sab-1=29$				

The main effect of A (advertising campaign) and interaction AB seem to be statistically not significant. However, main effect of B (time period) and test market (nested in A) seem to be present.

Since the interaction effect is not present we may wish to estimate contrasts in $\mu_{\cdot j \cdot}$'s and contrasts in $\mu_{\cdot\cdot k}$'s. The textbook gives some examples.

Repeated measures designs. Two-factors with repeated measures on both factors.

A clinician studied the effects of two drugs used either alone or together on the blood flow in human subjects. Twelve healthy middle-aged males participated in the study and they are viewed as a random sample of from a relevant population of middle-aged males. The four combinations of the treatments are:

A_1B_1: placebo (neither drug); A_1B_2: drug B alone; A_2B_1 : drug A alone; A_2B_2: both drugs A and B

[Thus factor A has two levels: placebo and drug A; factor B has two levels: placebo and drug B]

The 12 subjects received each of the four treatments in independently randomized orders. The response variable is the increase in drug flow from before to shortly after the administration of the treatment. The treatments were administered on successive days. This wash-out period prevented any carryover effect because the effect of each drug is short-lived. The experiment was conducted in a double-blind fashion so that neither the physician nor the subject knew which treatments administered when the change in blood flow was measured. The data are given below. Note that a negative entry denotes a decrease in blood flow.

Subject (i)	Treatment			
	A_1B_1	A_1B_2	A_2B_1	A_2B_2
1	2	10	9	25
2	-1	8	6	21
3	0	11	8	24
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
10	-2	10	10	28
11	2	8	10	25
12	-1	8	6	23

Note that subject effect is random, but factors A and B are fixed. The model here is

$$Y_{ijk} = \mu_{...} + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + (\rho\alpha)_{ij} + (\rho\beta)_{ik} + \varepsilon_{ijk}, i = 1, \ldots, s, j = 1, \ldots, a, k = 1, \ldots, b,$$

where

(a) $\mu_{...}$ is the overall mean,

(b) ρ_i's iid $N(0, \sigma_\rho^2)$,

(c) α_j's are main effect of factor A, $\sum \alpha_i = 0$,

(d) β_k's are main effects of factor B, $\sum \beta_k = 0$,

(e) $(\alpha\beta)_{jk}$'s are the interactions (between A and B), $\sum_j (\alpha\beta)_{jk} = 0$ for each k and $\sum_k (\alpha\beta)_{jk} = 0$ for each j,

(f) $(\rho\alpha)_{ij}$'s $N(0, ((a-1)/a)\sigma_{\rho\alpha}^2)$ subject to the constraint $\sum_j (\rho\alpha)_{ij} = 0$ for each i, and $Cov((\rho\alpha)_{ij}, (\rho\alpha)_{ij'}) = -\sigma_{\rho\alpha}^2/a, j \neq j'$,

(g) $(\rho\beta)_{ik}$'s $N(0, ((b-1)/b)\sigma_{\rho\beta}^2)$ subject to the constraint $\sum_k (\rho\beta)_{ik} = 0$ for each i, and $Cov((\rho\beta)_{ik}, (\rho\beta)_{ik'}) = -\sigma_{\rho\beta}^2/b, k = k'$,

(h) ε_{ijk}'s are iid $N(0, \sigma^2)$,

(i) $\{\rho_i\}$, $\{(\rho\alpha)_{ij}\}$, $\{(\rho\beta)_{ik}\}$ and $\{\varepsilon_{ijk}\}$ are independent.

For this model

$$E(Y_{ijk}) = \mu_{...} + \alpha_j + \beta_k + (\alpha\beta)_{jk},$$
$$Var(Y_{ijk}) = \sigma_\rho^2 + ((a-1)/a)\sigma_{\rho\alpha}^2 + ((b-1)/b)\sigma_{\rho\beta}^2 + \sigma^2.$$

There are two tables in the text both numbered 27.11 - the first gives you expected values of the mean squares and the second one gives you actual ANOVA table of the data.

Important: As in the last example, the sums of squares, df's, mean squares are calculated as if all the three factors were fixed (not random). However, the expected values of the sums of squares are different from a design in which the factor effects are fixed. The expected values reflect the fact that the subject effect is random. In this case it is like a three factor model without the three factor intercation and with one observation with n = 1.

Estimates are

$$\hat{\mu}_{...} = \bar{Y}_{...},\ \hat{\alpha}_j = \bar{Y}_{\cdot j\cdot} - \bar{Y}_{...},\ \hat{\beta}_k = \bar{Y}_{\cdot\cdot k} - \bar{Y}_{...}, \hat{\rho}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{...},$$
$$\widehat{(\alpha\beta)}_{jk} = \bar{Y}_{\cdot jk} - \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot k} + \bar{Y}_{...},$$
$$\widehat{(\rho\alpha)}_{ij} = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{...},$$
$$\widehat{(\alpha\beta)}_{ik} = \bar{Y}_{i\cdot k} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot k} + \bar{Y}_{...}$$

Thus the fitted Y values and residuals are

$$\hat{Y}_{ijk} = \hat{\mu}_{...} + \hat{\alpha}_j + \hat{\beta}_k + \widehat{(\alpha\beta)}_{jk} + \widehat{(\rho\alpha)}_{ij} + \widehat{(\alpha\beta)}_{ik}$$
$$= \bar{Y}_{ij\cdot} + \bar{Y}_{i\cdot k} + \bar{Y}_{\cdot jk} - 2\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot k} + .2\bar{Y}_{...}$$
$$e_{ijk} = Y_{ijk} - \hat{Y}_{ijk}.$$

The expressions for SSTO, SSA, SSB, SSAB remain the same as before

$$SSTO = \sum\sum\sum(Y_{ijk} - \bar{Y}_{...})^2,\ df = sab - 1$$
$$SSA = \sum\sum\sum \hat{\alpha}_j^2, df = a - 1$$
$$SSB =, \sum\sum\sum \hat{\beta}_k^2,\ df = b - 1$$
$$SSS = \sum\sum\sum \hat{\rho}_i^2,\ df = s - 1,$$
$$SSAB = \sum\sum\sum \widehat{(\alpha\beta)}_{jk}^2, df = (a-1)(b-1),$$
$$SSAS = \sum\sum\sum \widehat{(\rho\alpha)}_{ij}^2, df = (s-1)(a-1),$$
$$SSBS = \sum\sum\sum \widehat{(\rho\beta)}_{ik}^2, df = (s-1)(b-1),$$
$$SSE = \sum\sum\sum e_{ijk}^2, df = (s-1)(a-1)(b-1).$$

Note that SSE is denoted by SSABS in the textbook.

ANOVA table

Source	df	SS	MS	F	p-val
A	$a-1=1$	1587.00	1587.00	$MSA/MSAS = 775.87$	0.000
B	$b-1=1$	2028.00	2028.00	$MSB/MSAB = 524.89$	0.000
Subject	$s-1=1$	258.50	23.50	$MSS/MSE = 20.58$	0.000
AB	$(a-1)(b-1)=1$	147.00	147.00	$MSAB/MSE = 129.36$	0.000
Subject*A	$(s-1)(a-1)=11$	22.50	2.05	$MSAS/MSE = 1.80$	0.172
Subject*B	$(s-1)(b-1)=11$	42.50	3.86	$MSBS/MSE = 3.40$	0.027
Error	$(s-1)(a-1)(b-1)=11$	12.50			
Total	$sab-1=47$	4098.00			

Except for Subject*A interaction all the other effects seems to be prsent, though Subject*B interaction is less conclusive. One can now consult the text in order to estimate various parameters and their confidence intervals etc. These are discussed in the text.

More complicated repeated measures designs.

There can be repeated measures design in which there can be more than two factors, with repeated measure on some and not the others. Also, the unbalanced case is handled pretty much the same way as discussed as on other designs via a regression approach. There can be repeated measures with covariates as in analysis of covariance models.

Split Plot Designs

Only an outline is given here of the split plot design using the example given in the text. It is generally known that yield of a crop, say wheat, depends on fertilizer and the amount/manner in which water is applied. Suppose it is desired to estimate the main effects and interaction of irrigation method (factor A, fixed) and fertilizer (factor B, fixed). Suppose that factor A is at $a = 2$ levels and factor B is at $b = 3$ levels. In order to estimate the main effects and interactions we need at least two replications for each combination of the factors. Thus we need at least $2ab = 12$ observations on yield. Now modern irrigation method can only be applied to a field, a large piece of land, and not small plots of land. Thus it may be prohibitively expensive to carry out an experiment with 24 plots. In contrast to irrigation method, fertilizers can be applied to small plots of land. Split plot design addresses this issue where one cannot have 24 fields for irrigation.

Split-plot method proposes that we need only 4 fields of equal size (instead of 24). Two fields are randomly selected and irrigation method 1 is on these fields; and irrigation method 2 is applied to other 2 randomly selected 2 fields. Now each field is randomly split in three smaller equal sized plots and the three fertilizers are given randomly to these plots. Thus the number of smaller plots is now 12. In the jargon of split-plot design, the fields are call whole-plots and the smaller plots (which are obtained by splitting the fields) are called split plots. The following table points out how the observations of a split plot looks like.

	Irrigation method			
	$j=1$		$j=2$	
Whole plot	$i=1$	$i=2$	$i=1$	$i=2$
$k=1$	Y_{111}	Y_{211}	Y_{121}	Y_{221}
$k=2$	Y_{112}	Y_{212}	Y_{122}	Y_{222}
$k=3$	Y_{113}	Y_{213}	Y_{123}	Y_{223}

Since the whole plots tend to be large, its effect (i.e., the whole-plot effect) should be considered entered in the model. However, the whole-plots are nested in factor A (irrigation method). Thus if Y_{ijk} is the yield when irrigation (factor A, fixed, a levels) is at level j, whole plot (number of levels s) is i and fertilizer (factor B, fixed, b levels) is at level k, then we may write the model as

$$Y_{ijk} = \mu_{...} + \rho_{i(j)} + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk}, k = 1, \dots, b, j = 1, \dots, a, i = 1, \dots, s$$

where

(a) $\mu_{...}$ is the overall mean,

(b) α_j are main effect of factor A, $\sum \alpha_j = 0$

(c) β_k are main effects of factor B, $\sum \beta_k = 0$

(d) $(\alpha\beta)_{jk}$ are the interactions (between A and B), $\sum_j (\alpha\beta)_{jk} = 0$ for each k, $\sum_k (\alpha\beta)_{jk} = 0$ for each j,

(e) $\rho_{i(j)}$'s are whole-plot effects nested in A (if the whole-plots are selected at random from a population of plots the are $\rho_{i(j)}$'s iid $N(0, \sigma_\rho^2)$),

(f) ε_{ijk} are iid $N(0, \sigma^2)$,

[(g) $\{\rho_{i(j)}\}$ and $\{\varepsilon_{ijk}\}$ are independent if $\rho_{i(j)}$'s are random.]

Estimates are

$$\hat{\mu}_{...} = \bar{Y}_{...},\ \hat{\alpha}_j = \bar{Y}_{.j.} - \bar{Y}_{...},\ \hat{\beta}_k = \bar{Y}_{..k} - \bar{Y}_{...},$$
$$\widehat{(\alpha\beta)}_{jk} = \bar{Y}_{.jk} - \bar{Y}_{.j.} - \bar{Y}_{..k} + \bar{Y}_{...}, \hat{\rho}_{i(j)} = \bar{Y}_{ij.} - \bar{Y}_{.j.}$$

Thus the fitted Y values and residuals are

$$\hat{Y}_{ijk} = \hat{\mu}_{...} + \hat{\rho}_{i(j)} + \hat{\alpha}_j + \hat{\beta}_k + \widehat{(\alpha\beta)}_{jk} = \bar{Y}_{.jk} + \bar{Y}_{ij.} - \bar{Y}_{.j.}.$$
$$e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - (\bar{Y}_{.jk} + \bar{Y}_{ij.} - \bar{Y}_{.j.}).$$

So the sums of squares are (notation: SSW(A): sum of squares due to whole plots (nested in A))

$$
\begin{aligned}
SSA &= \sum\sum\sum \hat{\alpha}_j^2, df = a-1, \\
SSB &= \sum\sum\sum \hat{\beta}_k^2, df = b-1, \\
SSAB &= \sum\sum\sum \widehat{(\alpha\beta)_{jk}}^2, df = (a-1)(b-1), \\
SSW(A) &= \sum\sum\sum \hat{\rho}_{i(j)}, df = a(s-1), \\
SSE &= \sum\sum\sum e_{ijk}^2, df = a(s-1(b-1), \\
SSTO &= \sum\sum\sum (Y_{ijk} - \bar{Y}_{...})^2, df = sab-1.
\end{aligned}
$$

Important: In the text SSE is called $SSB.W(A)$.

The mean squares etc. are the same as any other design. In terms of hypotheses testing, how the F-statistics are formed depend on whether whole-plot effects are random or fixed.