Handout 10

Repeated measures design

This design is often employed in practice when the same subject (or object) is subjected to a number of treatments. It is a two-factor model in which one factor (subject) is usually treated as random whereas the treatment is a fixed factor. Another way is to treat it as a randomized block design where the block effects are random. The following example will clarify this.

Blood pressure: The relationship between the dose of a drug that increases blood pressure and the actual amount of increase in mean systolic blood pressure was investigated in a laboratory experiment. Twelve rabbits received in random order six different dose levels of the drug, with a suitable time interval between each drug administration. The increase in blood pressure was used as the response variable. The data on blood pressure increase follow.

Rabbit	Dose j						Rabbit	Dose j					
i	.1	.3	.5	1.0	1.5	3.0	i	.1	.3	.5	1.0	1.5	3.0
1	21	21	23	35	36	48	7	9	12	17	22	33	40
2	19	24	27	36	36	46	8	20	20	30	30	38	41
3	12	25	27	26	33	40	9	18	18	27	31	42	49
4	9	17	18	27	34	39	10	8	12	11	24	26	31
5	7	10	19	25	31	38	11	18	22	25	32	38	38
6	18	26	26	29	39	44	12	17	23	26	28	34	35

Let Y_{ij} be the blood pressure increase of the i^{th} rabbit on the j^{th} dose of the drug. Following the notations in the book we will assume that we have s subjects and there are r levels of the treatment (dose). The model is

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

where $\mu_{\cdot\cdot}$ is the overall mean, τ_j's are the treatment effects (fixed), ρ_i's are the subject effects (random), ε_{ij}'s are independent $N(0, \sigma^2)$. It is assumed that ρ_i's are independent $N(0, \sigma^2)$, and $\{\rho_i\}$ and $\{\varepsilon_{ij}\}$ are independent. The total number of observations is $n_T = sr$. Here are a few facts

$$E(Y_{ij}) = \mu_{\cdot\cdot} + \tau_j, \ \ Var(Y_{ij}) = \sigma_\rho^2 + \sigma^2,$$
$$Cov(Y_{ij}, Y_{ij'}) = \sigma_\rho^2, \ j \neq j', Cov(Y_{ij}, Y_{i'j'}) = 0, i \neq i'.$$

As in the case of single factor random effects model, the observations Y_{ij}'s are no longer independent. Note that the ratio $\sigma_\rho^2/(\sigma_\rho^2 + \sigma^2)$ is the proportion of variability in blood pressure increase that can be attributed to the variability among subjects.

Estimates of $\mu_{\cdot\cdot}$ and τ_j are

$$\hat{\mu}_{\cdot\cdot} = \bar{Y}_{\cdot\cdot}, \hat{\tau}_j = \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}.$$

Sums of squares and mean squares

The sums of squares and the mean squares are

$$\begin{aligned} SSTO &= \sum\sum(Y_{ij}-\bar{Y}_{..})^2, df = rs-1, \\ SSS &= r\sum(\bar{Y}_{i.}-\bar{Y}_{..})^2, df = s-1, MSS = \frac{SSS}{s-1}, \\ SSTR &= s\sum(\bar{Y}_{.j}-\bar{Y}_{..})^2, df = r-1, MSTR = \frac{SSTR}{r-1}, \\ SSE &= \sum\sum(Y_{ij}-\bar{Y}_{i.}-\bar{Y}_{.j}+\bar{Y}_{..})^2, df = (s-1)(r-1), MSE = \frac{SSE}{(s-1)(r-1)}. \end{aligned}$$

Note that SSS is stands for "sum of squares due to subject effects". The following hold

$$SSTO = SSS + SSTR + SSE,$$

$$df(SSTO) = df(SSS) + df(SSTR) + df(SSE).$$

The following important fact will guide us to hypothesis testing.

Fact i) $E(MSE) = \sigma^2$, ii) $E(MSS) = r\sigma_\rho^2 + \sigma^2$, iii) $E(MSTR) = \sigma^2 + \frac{s}{r-1}\sum\tau_j^2$.

Remark

a) MSE estimates σ^2,

b) Estimate of σ_ρ^2 is given by $s_\rho^2 = \frac{MSS-MSE}{r}$. If MSS is smaller than MSE, then s_ρ^2 is set to zero.

Hypothesis testing:

1. If we want to test $H_0 : \tau_j = 0$ for all j against H_1 : not all τ_j's are zero, then the F-statistic is $F^* = \frac{MSTR}{MSE}$. Degrees of freedom for the F-test are $(r-1, (s-1)(r-1))$.

2. If we want to test $H_0 : \sigma_\rho^2 = 0$ against $H_1 : \sigma_\rho^2 \neq 0$, then the F-statistic is $F^* = \frac{MSS}{MSE}$. The degrees of freedom for this F-test are $(s-1, (s-1)(r-1))$.

Confidence intervals

The following table will tell us how to estimate various parameters and construct confidence intervals.

Parameter	Estimate	$Var(estimate)$	$s^2(estimate)$
$\mu_{..}$	$\hat{\mu}_{..} = \bar{Y}_{..}$	$Var(\hat{\mu}_{..}) = \frac{r\sigma_\rho^2+\sigma^2}{n_T}$	$s^2(\hat{\mu}_{..}) = \frac{MSS}{n_T}$
$D = \tau_j - \tau_{j'}$	$\hat{D} = \bar{Y}_{.j} - \bar{Y}_{.j'}$	$Var(\hat{D}) = \frac{2}{s}\sigma^2$	$s^2(\hat{D}) = \frac{2}{s}MSE$
Contrast: $L = \sum d_j\tau_j$	$\hat{L} = \sum d_j\bar{Y}_{.j}$	$Var(\hat{L}) = \frac{\sum d_j^2}{s}\sigma^2$	$s^2(\hat{L}) = \frac{\sum d_j^2}{s}MSE$

Tukey's test for interactions

In order to check if the interaction effects can be ignored we may use Tukey's method as in chapters 20 and 21. We may also use the plot of Y_{ij}'s (as in chapters 20 and 21) to investigate the presence or absence of interaction effects.

Blood pressure data:

We have

$$s = 12, r = 6, n_T = 72, \bar{Y}_{\cdot\cdot} = 26.89,$$
$$\bar{Y}_{\cdot 1} = 14.67, \bar{Y}_{\cdot 2} = 19.17, \bar{Y}_{\cdot 3} = 23.00, \bar{Y}_{\cdot 4} = 28.75, \bar{Y}_{\cdot 5} = 35.00, \bar{Y}_{\cdot 6} = 40.75,$$
$$\hat{\tau}_1 = -12.22, \hat{\tau}_2 = -7.72, \hat{\tau}_3 = -3.89, \hat{\tau}_4 = 1.86, \hat{\tau}_5 = 8.11, \hat{\tau}_6 = 13.11.$$

We write down below the ANOVA table for the blood pressure data

Source	df	SS	MS	F	p-value
Subject	$s-1=11$	1197.44	108.86	12.81	0.000
Dose	$r-1=5$	5826.28	1165.26	137.09	0.000
Error	$(s-1)(r-1)=55$	467.39	8.498		
Total	$sr-1=71$	7491.11			

Clearly, both the subject and treatment effects seem to be present here.
Estimate of σ^2 is $MSE = 8.498$. Estimate of σ^2_ρ is $s^2_\rho = \frac{MSS-MSE}{r} = \frac{108.86-8.498}{6} = 16.73$.
An estimate of $\sigma^2_\rho/(\sigma^2_\rho+\sigma^2)$ is $\frac{s^2_\rho}{s^2_\rho+MSE} = \frac{16.73}{16.73+8.498} = .663$.
About 66.3% of the variability in blood pressure increase is due to variability among subjects.

Now suppose that we wish to find the dose which causes the largest increase in blood pressure. For this we will need to construct simultaneous confidence intervals for all pairwise differences of $\mu_{\cdot j} = \mu_{\cdot\cdot} + \tau_j$'s. Let us assume that we will use Tukey's method for this purpose. There are $\binom{6}{2} = 15$ intervals to be constructed. Fortunately, we need not construct all the confidence intervals in order to detect the dose which causes the largest increase in blood pressure. We need to compare only those doses for which $\bar{Y}_{\cdot j}$ are the two largest, but use the Tukey multiplier. Note that the $\bar{Y}_{\cdot 6}$ and $\bar{Y}_{\cdot 5}$ are the two largest. So we will construct a 95% confidence interval for $D = \mu_{\cdot 6} - \mu_{\cdot 5} = \tau_6 - \tau_5$ suing Tukey's multiplier.

Note that

$$\hat{D} = \bar{Y}_{\cdot 6} - \bar{Y}_{\cdot 5} = 40.75 - 5.75, s(\hat{D}) = \sqrt{\frac{2}{s}MSE} = \sqrt{\frac{2}{12}(8.498)} = 1.1901,$$
$$T = \frac{1}{\sqrt{2}}q(1-\alpha; r, (s-1)(r-1))$$
$$= \frac{1}{\sqrt{2}}q(.95; 6, 55) \approx \frac{1}{\sqrt{2}}(4.19) = 2.9628.$$

[Note that from the table of Studentized range distribution, we have $q(.95; 6, 40) = 4.23$ and $q(.95; 6, 60) = 4.16$. So $q(.95; 6, 55) \approx 4.19$ by interpolation.]

So using Tukey's method for simultaneous confidence intervals we get the following interval for $D = \tau_6 - \tau_5$

$$\hat{D} \pm Ts(\hat{D}), i.e., 5.75 \pm (2.9628)(1.1901), i.e., 5.75 \pm 3.53, i.e., (2.22, 9.28).$$

Since this interval does not include zero, we may conclude with 95% confidence that dose 6 causes the largest increase in blood pressure.