ST553 HW9

Nick Sun

June 3, 2019

Problem 1

We are examining a crossover design here. We have 5 treatments and 5 subjects with 5 different treatment periods - this study has a Latin Squares design.

The model prescribed by this design will be

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}$$

where y_{ijk} is the response of the *ith* treatment with the *jth* subject during the *kth* treatment period, μ is the overall mean, α_i is the effect of the *ith* treatment level, β_j is the effect of the *jth* subject, γ_k is the effect of the *kth* treatment period, and ϵ_{ijk} is the random error.

Our constraints and our assumptions are that $\sum \alpha_i = \sum \beta_j = \sum \gamma_k = 0$ and $\epsilon_{ijk} \sim N(0, \sigma^2)$.

Our ANOVA table with the degrees of freedom will look like

Source	df
Treatments	g-1
Subject	g-1
Treatment Period	g-1
Error	(g-1)(g-2)
Total	$g^2 - 1$

In our case, we have g = 5, so the degrees of freedom for the sum of squares of treatments, subject, and treatment period are 4, the SSE df is 12, and the Total sum of squares df is 24.

Problem 2

For this problem, we are looking at a randomized complete block design comparing soybean varieties and weed treatments. There are 16 varieties and 3 weed treatments and 2 replicates per unique combination, 1 from two different cities in Minnesota.

We can treat the location of the replicates as a blocking factor. If we want to compare this RCBD design to a completely randomized design, we can calculate the relative efficiency of these designs.

The formula is given as

$$\hat{E}_{RCBD:CRD} = \frac{\hat{\sigma}^2_{CRD}}{\hat{\sigma}^2_{RCBD}} \left(\frac{(\nu_{RCBD} + 1)(\nu_{CRD} + 3)}{(\nu_{RCBD} + 3)(\nu_{CRD} + 1)} \right)$$

where σ^2_{CRD} is estimated using

$$\hat{\sigma}^2_{CRD} = \frac{(r-1)MS_{BLOCK} + ((g-1) + (g-1)(r-1))MSE}{(r-1) + (g-1) + (g-1)(r-1)}$$

Plugging in the numbers we get from this table from `PROC MIXED`:

Source	DF	Sum of Squares	Mean Square
treatment	2	85183540	42591770
variety	15	25587029	1705802
location*treatment	2	11746744	5873372
location	1	50634150	50634150
Residual	75	66737633	889835

and using r = 2 for the blocking factor and g = 48 for each unique combination of the treatments.

We calculate the relative efficiency as 1.435 (the RCBD is 1.4 times as efficient as CRD). Therefore, in order to get the same precision using a balanced CRD as in our RCBD with n = 96, we would need `96*1.4`, or around 135 experimental units

Problem 3

We are looking at a Latin Squares design for this problem. Our response is tissue growth in rabbits. There are 4 treatments in this experiments:

- Treatment A is the positive control
- Treatment B is the negative control
- Treatments C and D are experimental devices

There are 4 rabbit subjects and 4 locations on each rabbit. Our rows and columns in our Latin Square are rabbits and locations.

We are interested in checking the effectiveness of the new devices C and D. We can analyze this data by using `PROC MIXED` to specify the blocking variables and then using `Tukey-Kramer` to conduct pairwise tests on the individual treatments.

Differences of Least Squares Means

Effect	Tmt	_Tmt	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
Tmt	A	B	0.3300	0.05476	6	6.03	0.0009	Tukey-Kramer	0.0038	0.05	0.1960	0.4640	0.1404	0.5196
Tmt	A	C	-0.02000	0.05476	6	-0.37	0.7275	Tukey-Kramer	0.9818	0.05	-0.1540	0.1140	-0.2096	0.1696
Tmt	A	D	0.2875	0.05476	6	5.25	0.0019	Tukey-Kramer	0.0077	0.05	0.1535	0.4215	0.09793	0.4771
Tmt	B	C	-0.3500	0.05476	6	-6.39	0.0007	Tukey-Kramer	0.0028	0.05	-0.4840	-0.2160	-0.5396	-0.1604
Tmt	B	D	-0.04250	0.05476	6	-0.78	0.4672	Tukey-Kramer	0.8627	0.05	-0.1765	0.09150	-0.2321	0.1471
Tmt	C	D	0.3075	0.05476	6	5.62	0.0014	Tukey-Kramer	0.0055	0.05	0.1735	0.4415	0.1179	0.4971

Least Squares Means

Effect	Tmt	Estimate	Standard Error	DF	t Value	Pr > |t|
Tmt	A	0.6150	0.08344	6	7.37	0.0003
Tmt	B	0.2850	0.08344	6	3.42	0.0142
Tmt	C	0.6350	0.08344	6	7.61	0.0003
Tmt	D	0.3275	0.08344	6	3.93	0.0078

We can see that new device C has a significantly higher mean response than the negative control B, but we do not have evidence to make the same claim about C's performance compared to the positive control A. We also see that new device D has a significantly lower mean response than the positive control A and

no evidence to say that it performs differently from negative control B. Therefore, we can summarise this by saying that the new experimental devices are not practically better than the controls - we have evidence to say that the new device D is outperformed by control A. We also do not have evidence to say that new device C does better than control A.

We can also choose to look at the average of the controls and the new devices using a contrast estimate.

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
experimental devices vs controls	-0.06250	0.07745	6	-0.81	0.4505

However, our large p-value indicates that we don't have strong statistical evidence to say that the mean of the new experimental devices are different than the controls, which again corroborates our findings from Tukey-Kramer.

Appendix

Here is the code I used for Problem 3

```
proc mixed data=dat2 method=type3;
    class Loc Subject Tmt;
    model Y=Tmt;
    random Loc Subject;
    lsmeans Tmt;
    estimate 'experimental devices vs controls' Tmt 1 1 -1 -1;
run;
```