19. CONTROLLING ERROR

19.1 MISSING VALUES

We have noted the importance of balance in experimental design. Balance aids not only the ease of subsequent statistical analysis but also in the interpretation of results. Thus balance could be taken to be an experimental principal just as randomization or the running of factorial arrangements. Sometimes balance is impossible to attain, e.g. observations are lost. If the departure from balance is not serious, i.e. only a small proportion of observations are missing then it 'seems reasonable' to estimate the missing values and then proceed - with the balanced analysis.

Suppose then that the original design matrix is X with data y and the design matrix for the data $y^* = \binom{y}{y^{**}}$ where y^{**} are missing values is $X_* = \binom{X}{X_{**}}$.
We estimate y^{**} by minimizing the function

$$(y_* - X_* \beta)' (y_* - X_* \beta)$$

which is minimized by choosing β = b and $y^{**} = X_{**} b$.

Thus missing value estimates are the predicted values at the relevant input setting obtained from the least squares estimates in the original model.

Note that the least squares estimates are the same in both situations and also the SS for error remains constant.

The other SS are calculated using the missing values, y*, and they will naturally be somewhat inflated thus biasing our results in favour of significance. Note that in estimating y^{**} we have not increased our degrees of freedom - for example if we had an a x b factorial with 1 missing value then the ANOVA table takes the form

Source	DF
Mean	1
A	a-1
B	b-1
A x B	(a-1)(b-1)
Error	(nab-1) - (a-1)(b-1)-(b-1) - (a-1) – (**#missing**-1)
	nab-1

Note that the error line has lost one degree of freedom.

19.2 CONTROLLING ERROR

A basic method available to the investigator for controlling the error in his/her experimental results is varying the number of applications of each treatment (i.e. the individual sample sizes). Other techniques such as blocking, analysis of covariance, etc. may also be applicable. We will discuss this later. Determining the sample size for each treatment is thus an important component in the design of an experiment. We will ensure that the experiment is balanced so we need only select one sample size for all treatments.

19.3 MULTIPLE COMPARISONS

Often in an investigation we do not have a prescribed set of contrasts for each factor which we wish to make inferences about. For example in a 1 factor design with > 2 levels we may not have an obvious set of orthogonal contrasts to use - but want to compare the levels of the factor on a pairwise basis.

These contrasts will obviously not be mutually orthogonal and thus even in the balanced case the SS will not be statistically independent. Various methods of multiple comparisons have been designed to deal with such situations.

A problem which must be faced when carrying out multiple comparisons procedures is that even if no differences exist it can happen that significances will show with high probability. Suppose we are going to test m hypotheses and we have m mutually statistically independent test procedures, one for each hypothesis and suppose we reject the hypotheses if the OLS is less than or equal to ~ and accept it otherwise.

Then if the hypotheses are all true the probability of rejecting at least one of them is $1-(1-\alpha)^m$.

Even for small values of α this probability can be quite high for moderate values of m.

Example

m / α	**.05**	**.01**
5	.23	.05
10	.40	.10
15	.54	.14
20	.64	.18

Thus if the comparisons we wish to make are not specified before the data is actually data it is very dangerous to go on 'fishing expeditions' for differences. This applies whether or not the test procedures are statistically independent.

Thus when making tests of significance based upon the data we must select α conservatively or proceed in a different fashion.

19.4 FISHER'S LEAST SIGNIFICANT DIFFERENCE TEST

R.A. Fisher (1949) developed a procedure for making pairwise comparisons among a set of k population means called the least significant difference (LSD). It can be used for orthogonal preplanned comparisons or all pairwise comparisons.

For a specified value of α, the LSD for comparing μ_i to μ_j with equal n is

$$LSD = t_{\frac{\alpha}{2}} \frac{\sqrt{2s^2}}{n}$$

where n is the number of observations on each treatment and t is the critical t value with df corresponding to s^2 or MSE. All pairs of means are compared, if $\left|\bar{y}_i - \bar{y}_j\right| > LSD$ the corresponding population means are different.

For each pairwise comparison the probability of Type 1 error is fixed at a specified value of α

19.5 TUKEY'S RANGE TEST

Tukey (1953) proposed a procedure that makes use of the Studentized range distribution. One ranks the t sample means and the two population means, μ_i and μ_j are declared different if $\left|\bar{y}_i - \bar{y}_j\right| \geq w$.

$$w = q_\alpha(t,v)\frac{s}{\sqrt{n}}$$

where q_α (t,v) is the upper tail critical value of the Studentized Range for comparing t different populations. The probability of observing an experiment with one or more pairwise comparisons falsely declared significant is specified at α.

19.6 STUDENT NEWMAN-KEULS TEST (SNK)

Keuls (1952) makes use of the Studentized range as did Tukey above. The t sample means are ranked and are r steps apart, the μ_i and μ_j are declared different if $\left|\bar{y}_i - \bar{y}_j\right| \geq w_r$.

$$W_r = q_\alpha(r,v)\frac{s}{\sqrt{n}}$$

where q_α (r,v) is the critical value for the Studentized Range.

19.7 Example

Lewis (1943), The Effect of Noise and Vibration on Certain Psychomotor Response, *Research Report* Iowa, examined the performance of six groups (four are listed below), 5 people per group working under different experimental conditions. Electrocardiograms recorded heartheat and averages were calculated per quarter minute. The data are listed below.

	Groups 1	2	3	4
	22.07	25.62	21.21	22.84
	19.95	26.55	17.37	25.43
	17.84	23.77	22.65	22.09
	22.54	24.20	21.59	20.76
	19.17	27.76	22.05	22.29
T_i	101.57	127.90	104.87	113.41
$\overline{X}_i$	20.31	25.58	20.97	22.68

19.8 SYNTAX COMPUTER IMPLEMENTATION

```
data list/ G 1 HB 3-7
begin data
1 22.07
1 19.95
. .
. .
. .
4 20.29
4 22.29
```

Once the data has been entered, open a new syntax file and type in the following command.

```
UNIANOVA
  hb  BY g
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /POSTHOC = g ( BONFERRONI )
  /PRINT = DESCRIPTIVE
  /CRITERIA = ALPHA(.05)
  /DESIGN = g .
```

After the command is typed, highlight the entire text and press the black error located on the tool bar in order to obtain the output. Click here for an SPSS windows example.

The output is given below. Note the form of the command statement ONEWAY which conducts a 1-Factor ANOVA on the data. HB, the dependent variable is followed by the factor group G which has a minimum value of 1 and a maximum of 4. The RANGES statement indicates that Fishers Least Significant Difference test is to be performed. Other options such as SNK can also be specified.

The ANOVA table indicates there are significant differences in the means of the four groups, $F(3,16) = 7.94$, $p = .002$.

Tests of Between-Subjects Effects

Dependent Variable: HB

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	82.881[a]	3	27.627	7.944	.002
Intercept	10024.003	1	10024.003	2882.463	.000
G	82.881	3	27.627	7.944	.002
Error	55.641	16	3.478		
Total	10162.525	20			
Corrected Total	138.522	19			

a. R Squared = .598 (Adjusted R Squared = .523)

The output of the LSD comparison procedure is listed below. The means are reported below and a matrix of group comparisons is also reported. The matrix is read as follows. A star (*) indicated the two groups have means that differ significantly from one another. For example, the elements; group(1,2),(3,3),(4,2) have stars indicating there are significant differences between these means. Subsets of means that *do not* differ are also reported, for example, G1, G3 and G4 do not differ.

HB

Student-Newman-Keuls[a,b]

G	N	Subset 1	Subset 2
1.00	5	20.3140	
3.00	5	20.9740	
4.00	5	22.6820	
2.00	5		25.5800
Sig.		.142	1.000

Means for groups in homogeneous subsets are displayed.
Based on Type III Sum of Squares
The error term is Mean Square(Error) = 3.478.

a. Uses Harmonic Mean Sample Size = 5.000.

b. Alpha = .05.

19.9 Exercises

1. For problem 1, chapter 14.7, use a multiple comparison procedure to compare the means.

2. For problem 2, chapter 14.7, use a multiple comparison procedure to compare the means.

3. For problem 1, chapter 14.7, assume that the last two animals in the placebo group, values 80, 80 are missing data. Redo the analysis with this information. Report the ANOVA table and state your conclusions clearly.