ON THE ACCURACY OF CURRENT APPROXIMATE VARIANCES OF TREATMENT DIFFERENCES IN RANDOMISED BLOCK DESIGNS WITH MISSING OBSERVATIONS

By M. N. Das and G. A. Kulkarni

Indian Council of Agricultural Research, New Delhi

Yates (1933) suggested that in designs with several missing observations an approximate value of the variance of treatment differences can be obtained by estimating what he called an effective number of replicates for each of the treatments involved in a difference. The effective number of replicates of any of the two treatments is obtained by subtracting from the existing number of replications half the number of those replicates of the other treatment which are in the blocks in which the first treatment is not missing. Sometimes an approximate variance is also obtained by dividing the error variance by the existing number of replicate of each of the treatments. Later, Taylor (1948) suggested an improvement that the factor $\frac{1}{2}$ in Yates' method should be replaced by $1/(k-1)$ where k is number of treatments.

So far there was no way of judging the accuracy of these approximate variances. One of the authors, Das (1954), recently obtained expressions for the variance in all possible cases of three and four missing observations in Randomised Block Designs as also when the observations missing from different blocks form themselves into B.I.B. designs. An attempt has, thus, been made in the present note to examine the accuracy of the approximate variances in all the balanced as also some non-balanced cases.

If in a R.B.D. with 'k' treatments and 'r' blocks, 'p' ($p < k$) is the number of treatments each having 'q' ($q \leqslant r$) replicates missing; 's' is the number of blocks in each of which observations from 'n' plots are missing such that the number of times any pair of treatments is missing in the same block is constant being equal to, say λ, then the exact variance of the difference of any two treatments both of which are affected has been obtained by Das (1954), as:

$$\frac{2\sigma^2 (k-n)}{v - q + \lambda}, \quad \text{where} \quad v = (r-q)(k-n)$$

i.e.,

$$\frac{2\sigma^2}{(r-q)-\frac{q-\lambda}{k-n}}. \tag{1}$$

The approximate variance of the same difference according to the method suggested by Yates comes to be

$$\frac{2\sigma^2}{(r-q)-\frac{q-\lambda}{2}}. \tag{2}$$

This shows that when $k-n$, *i.e.*, the number of existing plots in the different blocks is 2, the variances become equal. In all other cases excepting when $q = \lambda$, *i.e.*, a set of treatments is missing all in each of a number of blocks and in none others are missing in any block, the approximate variance of Yates is always greater than the true variance. This implies that by applying the approximate variance there will be less number of significant cases, than what it should be for all such comparisons.

Again if we obtain the approximate variance by dividing by the existing number of replications only, the variance of the difference will in such cases be $2\sigma^2/(r-q)$. This like the other variance suggested by Yates becomes equal to the true variance when $q = \lambda$ but in all cases it is less than the true variance and this will result in making the 't' value greater. Hence there will be more number of significant cases than what it should be, so that by using such variances some of the non-significant differences will pass out as significant.

The exact variance of the difference between two treatments only one of which is affected has been obtained by Das (1954) as:

$$\mathrm{Var}\,(t_i - t_m) = \sigma^2 \left[\frac{1}{r} + \frac{1}{v + (p-1)(q-\lambda)} \times \left\{\frac{q}{r} + \frac{(k-n)\,\overline{v + (q-\lambda)}\,\overline{(p-2)}}{v - q + \lambda}\right\}\right] \tag{3}$$

where $v = (r-q)(k-n)$, and t_m stands for the affected treatment and t_i for unaffected.

The approximate variances of the same difference is equal to

$$\sigma^2 \left[\frac{1}{r-q} + \frac{1}{r-\frac{q}{2}}\right] \tag{4}$$

according to Yates' method and

$$\sigma^2\left[\frac{1}{r}+\frac{1}{r-q}\right] \tag{5}$$

according to the other.

It appears that direct comparison of these expressions is not possible excepting that exact variance (3) and Yates' variance (4) are both always greater than (5). Actually the variance at (5) is the limit of the variance at (3) when k is very large. Detailed examination for different situations has been made by obtaining the tables of coefficients of σ^2 in the variance of the differences between the treatments.

A scrutiny of the tables, some of which have been appended, shows that if the number of treatments is five or more in a R.B.D. with five or more blocks (which alone are the useful designs) the approximate variance of Yates for the differences $(t_i - t_m)$ is almost always greater than or equal to the exact variance, excepting where all the missing plots are in the same block. In the latter case, if the number of the affected treatments be $n+2$ or less, the approximate variance is less than the exact whatever the number of blocks. If there be only four treatments, it is found that the Yates' variance is greater than the exact one in most of the designs with six or more number of blocks, excepting the case where four plots are missing affecting the two treatments in each of the two blocks.

These tables thus indicate that there are cases under such types of comparison where application of Yates' method may pass out some of the comparisons as significant when they are really not so. The position of the other approximate variance remains the same as found earlier, *viz.*, it is always less than the exact variance whatever may be the type of comparison. The tables, however, indicate that if either or both of the number of treatments and replications be large the difference between the two variances becomes small.

Broadly the tables show that in all cases having error *d.f.* 12 or more the average of the two variances, *viz.*, that suggested by Yates and the other approximate formula is a very close approximate to the exact variance.

An examination of Taylor's improvement shows that its tendency is to make the variance smaller than the exact one resulting in more number of significant cases. In the different cases examined it has been found that the method suggested gives generally better approximate values than what can be obtained from Taylor's method and is less likely to give more number of significant cases in all cases of practical interest.

SUMMARY

Accuracy of the approximate variances of treatment differences in Randomised Block Design as obtained (i) from the effective number of replications (suggested by Yates) and (ii) from the existing number of replications has been examined by comparing them with the corresponding exact variances. It appears that the variance through Yates' method gives in most of the cases an overestimate of the true variance, while the other approximate variance is always an underestimate. The average of these two expressions gives a very close approximation to the exact variance and is an improvement over what was suggested by Taylor.

ACKNOWLEDGEMENTS

The authors feel grateful to Dr. V. G. Panse, Statistical Adviser, Indian Council of Agricultural Research, for drawing their attention to the problem.

REFERENCES

1. Yates, F. .. *Emp. J. Exp. Agri.*, 1933, **1**, 129–42.
2. Taylor, J. .. *Nature*, 1948, **162**, 262.
3. Das, M. N. .. *J. Ind. Soc. Agri. Stat.*, 1954, **6**, 58–76.

APPENDIX

Variance of $(t_i - t_m)$ where t_i stands for unaffected treatment and t_m for affected one in the different cases of balanced incompleteness. [The figures in the tables below are coefficients of σ^2 obtained from formulæ (3), (4) and (5) in the text.]

CASE I. *Three plots missing such that* $p = s = 3, \ n = q = 1, \ \lambda = 0$

r \ k	4	5	6	7	8	9	$\frac{1}{r-q}+\frac{1}{r-q/2}$	$\frac{1}{r}+\frac{1}{r-q}$	Average of (4) and (5)
	By formula (3)						by (4)	by (5)	
3 ..	0·900	0·881	0·870	0·864	0·859	0·856	0·900	0·833	0·866
4 ..	0·614	0·605	0·601	0·601	0·596	0·594	0·619	0·583	0·601
5 ..	0·467	0·463	0·460	0·458	0·457	0·456	0·472	0·450	0·461
6 ..	0·378	0·375	0·373	0·372	0·371	0·371	0·381	0·367	0·374

CASE II. *Three plots missing such that* $p = n = 1, \ s = q = 3, \ \lambda = 0$

r \ k	2	3	4	5	6	7	8	$\frac{1}{r-q}+\frac{1}{r-q/2}$	$\frac{1}{r}+\frac{1}{r-q}$	Average of (4) and (5)
	By formula (3)							by (4)	by (5)	
4 ..	2·000	1·625	1·500	1·437	1·400	1·375	1·357	1·400	1·250	1·325
5 ..	1·000	0·850	0·800	0·775	0·760	0·750	0·743	0·786	0·700	0·743
6 ..	0·667	0·583	0·556	0·542	0·533	0·528	0·524	0·556	0·500	0·528
7 ..	0·500	0·446	0·428	0·420	0·414	0·411	0·408	0·432	0·393	0·417

CASE III. *Three plots missing such that* $p = n = 3,\ s = q = 1,\ \lambda = 1$

$r \backslash k$	4	5	6	7	8	9	10	$\frac{1}{r-q}+\frac{1}{r-q/2}$	$\frac{1}{r}+\frac{1}{r-q}$	Average of (4) and (5)
	By formula (3)							by (4)	by (5)	
2 ..	2·000	1·750	1·667	1·625	1·600	1·583	1·571	1·667	1·500	1·583
3 ..	1·000	0·917	0·889	0·875	0·867	0·861	0·857	0·900	0·833	0·866
4 ..	0·667	0·625	0·611	0·604	0·600	0·597	0·595	0·619	0·583	0·601
5 ..	0·500	0·475	0·467	0·462	0·460	0·458	0·457	0·472	0·450	0·461
6 ..	0·400	0·383	0·378	0·375	0·373	0·372	0·371	0·381	0·367	0·374

CASE IV. *Four plots missing such that* $p = s = 4,\ q = n = 1,\ \lambda = 0$

$r \backslash k$	5	6	7	8	9	10	11	$\frac{1}{r-q}+\frac{1}{r-q/2}$	$\frac{1}{r}+\frac{1}{r-q}$	Average of (4) and (5)
	By formula (3)							by (4)	by (5)	
2 ..	1·714	1·656	1·622	1·600	1·584	1·573	1·564	1·667	1·500	1·583
3 ..	0·883	0·872	0·865	0·860	0·856	0·853	0·851	0·900	0·833	0·866
4 ..	0·606	0·601	0·598	0·596	0·594	0·593	0·592	0·619	0·583	0·601
5 ..	0·463	0·460	0·459	0·457	0·456	0·456	0·455	0·472	0·450	0·461
6 ..	0·375	0·373	0·372	0·371	0·371	0·370	0·370	0·381	0·367	0·374

CASE V. *Four plots missing such that* $p = q = 2,\ s = 4,\ n = 1,\ \lambda = 0$

$r \backslash k$	3	4	5	6	7	8	9	$\frac{1}{r-q}+\frac{1}{r-q/2}$	$\frac{1}{r}+\frac{1}{r-q}$	Average of (4) and (5)
	By formula (3)							by (4)	by (5)	
4 ..	0·958	0·844	0·808	0·792	0·782	0·776	0·772	0·833	0·750	0·791
5 ..	0·600	0·569	0·557	0·551	0·547	0·545	0·543	0·583	0·533	0·558
6 ..	0·450	0·436	0·430	0·427	0·425	0·424	0·423	0·450	0·417	0·433
7 ..	0·363	0·355	0·352	0·349	0·348	0·347	0·347	0·367	0·343	0·355
8 ..	0·305	0·300	0·298	0·296	0·295	0·295	0·294	0·310	0·292	0·301

CASE VI. *Four plots missing such that* $p = n = 1,\ s = q = 4,\ \lambda = 0$

r \ k		2	3	4	5	6	7	8	$\frac{1}{r-q}+\frac{1}{r-q/2}$	$\frac{1}{r}+\frac{1}{r-q}$	Average of (4) and (5)
		By formula (3)							by (4)	by (5)	
5	..	2·000	1·600	1·467	1·400	1·360	1·333	1·314	1·333	1·200	1·266
6	..	1·000	0·833	0·778	0·750	0·733	0·722	0·714	0·750	0·667	0·708
7	..	0·667	0·571	0·540	0·524	0·514	0·508	0·503	0·533	0·476	0·504
8	..	0·500	0·437	0·417	0·406	0·400	0·396	0·393	0·417	0·375	0·396
9	..	0·400	0·355	0·341	0·333	0·329	0·326	0·324	0·343	0·311	0·327

CASE VII. *Four plots missing such that* $p = q = s = n = \lambda = 2$

r \ k		3	4	5	6	7	8	9	$\frac{1}{r-q}+\frac{1}{r-q/2}$	$\frac{1}{r}+\frac{1}{r-q}$	Average of (4) and (5)
		By formula (3)							by (4)	by (5)	
3	..	2·000	1·667	1·555	1·500	1·467	1·444	1·428	1·500	1·333	1·416
4	..	1·000	0·875	0·833	0·812	0·800	0·792	0·786	0·833	0·750	0·791
5	..	0·667	0·600	0·578	0·567	0·560	0·555	0·552	0·583	0·533	0·558
6	..	0·500	0·458	0·444	0·437	0·433	0·430	0·429	0·450	0·417	0·433
7	..	0·400	0·371	0·362	0·357	0·354	0·352	0·351	0·367	0·343	0·355

CASE VIII. *Four plots missing such that* $p = n = 4,\ q = s = 1,\ \lambda = 1$

r \ k		5	6	7	8	9	10	11	$\frac{1}{r-q}+\frac{1}{r-q/2}$	$\frac{1}{r}+\frac{1}{r-q}$	Average of (4) and (5)
		By formula (3)							by (4)	by (5)	
2	..	2·000	1·750	1·667	1·625	1·600	1·583	1·467	1·667	1·500	1·583
3	..	1·000	0·917	0·889	0·875	0·867	0·861	0·857	0·900	0·833	0·866
4	..	0·667	0·625	0·611	0·604	0·600	0·597	0·595	0·619	0·583	0·601
5	..	0·500	0·475	0·467	0·462	0·460	0·458	0·457	0·472	0·450	0·461
6	..	0·400	0·383	0·378	0·375	0·373	0·372	0·371	0·381	0·367	0·374