Euphytica 1 (1952): 112–122

THE USE OF THE „STUDENTIZED RANGE" IN CONNECTION WITH AN ANALYSIS OF VARIANCE

M. KEULS
Institute of Horticultural Plant Breeding, Wageningen

Received 26. Mar. 1952

CONTENTS

Introduction 112
Numerical example 112
Analysis of variance 114
Range-test 116
Conclusion 121
Summary 121
Samenvatting 122
List of references 122

INTRODUCTION

In breeding agricultural and horticultural crops it is, in many cases, of much importance to compare the different selections obtained, e.g. in regard to their productive capacity. This is usually done in field trials involving these selections. The different plot yields will give us an impression of the productivity of the selections grown. In order to find out how far such impressions are reliable, the yield figures are mathematically worked out. As a rule a so-called analysis of variance (described hereafter) will be carried out. The following conclusions may be drawn from it: the selections will show or will not show differences which are statistically reliable, leaving unrevealed, however, *which* selections differ. In practice often a t-test is applied to find this out. This article means to show that a t-test is not allowed, at the same time suggesting a practical method to learn more about each pair of varieties.

NUMERICAL EXAMPLE

The calculation concerns a trial on white cabbage carried out in 1950. A trial field had been divided into 39 plots, grouped into 3 blocks of 13 plots each. In each block the 13 varieties to be investigated were planted out (randomized blocks design). During this trial all plots were treated in exactly the same way. The purpose was to learn which variety would give the highest gross yield per head of cabbage and which the lowest, in other words to find approximately the order of the varieties according to gross yield per cabbage.

The numerical example (see fig. 1) shows 39 gross weights which were determined for each plot, arranged according to variety 1, 2, ..., 13 and to block (A, B and C). Thus, variety 1 in the plot of block A yielded 2090 grammes per (head of) cabbage. For simpler computation purposes this number was first rounded off to 3 figures, thus giving 209, and then it was reduced by a „basic number" 120, so that in the end the calcultation is based on the number 89 („reduced yield"). Therefore in the numerical example the 39 numbers of the 13 varieties in the 3 blocks indicate the reduced yields. Behind column C figures the sum of the columns A, B and C. Thus, for variety 1 the

FIG. 1. NUMERICAL EXAMPLE

Basic number 120	Gross weight per cabbage				
Number of variety	A	B	C	sum	mean
1	89	46	33	168	176.0
2	0	- 5	-21	-26	111,3
3	-25	-16	-26	-67	97.7
4	22	7	- 3	26	128.7
5	8	6	-12	2	120.7
6	31	10	- 5	36	132.0
7	54	15	- 4	65	141.7
8	- 8	-20	-30	-58	100.7
9	31	7	- 5	33	131.0
10	26	14	-27	13	124.3
11	67	29	2	98	152.7
12	65	21	6	92	150.7
13	23	7	- 3	27	129.0
	383	121	-95	409	

The nrs 1, 2, 3..... 13 indicate a variety
The letters A, B and C indicate three blocks

Cause of variation	Degrees of freedom	Sum of squares	Variance	F
general mean	1	4289.26		
varieties	12	16713.74	1392.81	11.21++
blocks[1]	2	8814.97	4407.49	35.46++
error	24	2983.03	124 .29	
total	39	32801.00		

$m = 130.49$
$\sigma \simeq \sqrt{124.29} = 11.15$
$\sigma/m\ \% \simeq 8.54$

Number of variety	Rank	Mean	Rank numbers	n_1+1	$\Delta_{0.05}$
1	**1**	176.0	1 — 1	2	18.80
11	**2**	152.7	2 — 9	3	22.73
12	**3**	150.7	2 — 10	4	25.11
7	**4**	141.7	2 — 10	5	26.91
6	**5**	132.0	2 — 11	6	28.20
9	**6**	131.0	2 — 11	7	29.29
13	**7**	129.0	2 — 11	8	30.26
4	**8**	128.7	2 — 11	9	31.16
10	**9**	124.3	2 — 13	10	31.87
5	**10**	120.7	3 — 13	11	32.51
2	**11**	111.3	5 — 13	12	33.09
8	**12**	100.7	9 — 13	13	33.61
3	**13**	97.7	9 — 13		

[1] In reality the blocks were different *trial fields*. To simplify the discussion we have spoken throughout the whole article of *blocks*. This accounts for the large F = 35.46++. Questions arising from this are not the object of this article.

sum of the reduced yields of the 3 blocks is 168 and the mean reduced yield of each plot 168:3 = 56,0. So the mean yield is 56,0 + 120 = 176,0 or 1760 grammes. On the last line figures the total of the above 13 lines. Therefore the total yield of all the varieties in block A is 383 and that of all varieties in all blocks = 409. The reduced mean yields of the 13 plots of block A and of all 39 plots are 383:13 = 29,5 and 409:39 = 10,5 respectively.

Analysis of variance

It is known that the means of each variety or block will afford the best estimates of the productive capacity of the variety or block. We are especially interested in the variety means. Now the analysis of variance was applied to find out whether the 13 variety means would vary more than could be expected when the 13 variety numbers represented one variety only. It is based on the following principles:

The 3 means 56,0, 29,5 and 10,5, as discussed above, may be considered as the best estimates of respectively:

productive capacity of variety 1, that of block A and the „normal" productive capacity. The differences 56,0 – 10,5 = 45,5 and 29,5 – 10,5 = 19,0 are the best estimates of the extra capacity of variety 1 and of the extra capacity of block A, or briefly, of the contributions of variety 1 and block A. Thus we may explain the yield of plot A, as follows:

Yield of A_1 = normal yield + contribution of 1 + contribution of A + remainder

$$89 = 10{,}5 + 45{,}5 + 19{,}0 + 14{,}0$$
$$a = b + c + d + e$$

The number 14,0, the closing entry in this equation, represents the remainder, say „error", unexplained contribution, which we interpret more or less rightly as due to chance. From the fact that the contributions of both varieties and blocks have been estimated as accurately as possible it follows that the chance deviation of the yield has been accumulated in the „error" as much as possible. It is obvious that we are justified to write in this way the yield a of each plot as the total of 4 components, $b + c + d + e$. For instance the reader himself may ascertain that the last plot C 13 will result in –3 = 10,5–1,5–17,8 + 5,8.

We call Sa^2 the total of the squares of the 39 numbers a, Sb^2 that of the 39 numbers b. It should be noted that only the numbers a and e can all differ. The numbers b are all alike; the numbers c form 13 groups, each of 3 equal numbers. Calculating the 5 square totals gives:

$Sa^2 = 89^2 +$	$+ (-3)^2$	= 32801	=	„total"
$Sb^2 = 10{,}5^2 +$	$+ (10{,}5)^2$	= 4289,26	=	„total mean"
$Sc^2 = 45{,}5^2 +$	$+ (-1{,}5)^2$	= 16713,74		„varieties"
$Sd^2 = 19{,}0^2 +$	$+ (-17{,}8)^2$	= 8814,97	=	„blocks"
$Se^2 = 14{,}0^2 +$	$+ 5{,}8^2$	= 2983,03	=	„error"

In reality these sums of squares are calculated in a simpler way than indicated above, but this unnecessarily complicates the explanation (for numerical computation see the well-known textbooks e.g. of G. W. Snedecor).

A so-called „relation of orthogonality" between the square totals (a sort of Pythagoras' theorem) states that:

with: $a = b + c + d + e$
goes: $Sa^2 = Sb^2 + Sc^2 + Sd^2 + Se^2$
or in the example: $32801 = 4289 + 16714 + 8815 + 2983$.

This relation results from the form of the table of figures (being a complete rectangular table without blank places). The last three sums of squares are to some extent measures of the relative causes of variation „varieties", „blocks" and „error".

In order to ascertain the effective power of these causes, we make use of the following proposition:

> Suppose the yields of the 39 plots have chance deviations of the *same order of magnitude* σ, such deviations being *uncorrelated*, and suppose there is a relation of orthogonality between the sums of squares then these sums of squares, apart from systematic contributions (varieties, blocks, total-mean) are estimates of n σ^2, n representing the number of degrees of freedom.

The number of degrees of freedom is a constant for each sum of squares representing the smallest number of independent outcomes involved. Thus it is easy to see that, in our example, the number of degrees of freedom of Sa^2 is $n_a = 39$, for the 39 numbers *a* are mutually independent. If only 38 were given, it would be impossible to calculate the 39th from them. In the same way $n_b = 1$, for the 39 numbers *b* are all alike. Since the numbers *c* have been divided into 13 groups of 3 equal numbers each, only 13 of them at most need to be given, viz. 1 out of each group of 3. But the sum of these 13 numbers must equal 0, for they represent the differences between the 13 variety means and the total mean. Therefore only 12 numbers *c* need to be given to calculate all of them, thus $n_c = 12$, and $n_d = 2$. The 39 numbers *e*, arranged in 13 rows and 3 columns may all differ. It is, however, possible to prove that the total of the 3 numbers in each row and also the total of the 13 numbers in each column must equal 0. Should one row and one column not be given then it can be calculated from the other 24 numbers, so that $n_e = 24$. We may verify that all numbers of degrees of freedom have been rightly calculated, for in addition to the above equations of „Pythagoras' theorem" the following equation also holds good:

$n_a = n_b + n_c + n_d + n_e$, or in the example
$39 = 1 + 12 + 2 + 24$.

In this example we are only interested in sums of squares Sc^2 and Se^2, „varieties" and „error". Suppose that the conditions of the foregoing proposition are fulfilled, dividing the square totals Sc^2 and Se^2 by their respective numbers of degrees of freedom n_c and n_e, gives (apart from systematic contributions) two estimates s_c^2 and s_e^2:

$$s_c^2 = 16713{,}74 : 12 = 1392{,}81$$
$$s_e^2 = 2983{,}03 : 12 = 124{,}29$$

the so-called „mean squares" for „varieties" and „error".

Of both estimates of σ^2, s_c^2 is only a right error estimate if the 13 varieties have the same productive capacity. On the other hand, s_e^2 is always right, even in the case of unequal productive capacity of varieties or blocks. Therefore the latter will measure σ^2. (The accuracy per plot of the trial can be expressed by the magnitude of σ in relation to the mean yield m, being here 120 + 10,5 = 130,5. That accounts for the statement that $\sigma/m = 8,54\ \%$).

It is of importance that $s_c^2 = 11,21 \times s_e^2$. The difference $s_c^2 - s_e^2 = 1268,52$ may be considered as an estimate of the contributions to the variety mean square by „real" differences between the varieties.

The following considerations serve to ascertain how far we may put confidence in a conclusion that there are „real" differences:

> If we assume that chance deviations apart from being uncorrelated and of the same order of magnitude follow moreover the *normal law of error*, a proposition states that $F = s_c^2/s_e^2$ will follow a fixed (tabulated) probability distribution.

When has been verified, that F has so great a value, that chance outcomes as great or greater have a probability of 0,05 at most, we conclude that it is improbable that the variety means form a random sample from one and the same normal population. If we maintain the supposition that the error distribution of the plot yields, therefore, of the variety means are normal, uncorrelated and have equal variances, then we can only conclude that the variety means are samples from normal populations, which only differ with respect to their expected (= real) mean values or more practically that *the variety means show „real" differences*. In the case before us the F-value (11,21) is even greater than may be expected on a probability of 0,01. This we have indicated by ++. So far for the analysis of variance itself.

RANGE-TEST

The conclusion that the variety means form samples from populations not having the same expected mean values is rather vague. It is clear that the varieties may be arranged in some way or other into groups of different productive capacity. Now we should like to have an answer to the question; which are these groups? Or, to put the question more vaguely, *which varieties do not belong to one and the same group*? In deciding upon this question one should have to ascertain for each pair of varieties if there are any objections to their being placed into one and the same group.

Should the trial only provide two variety means, then the t-test would be valid. When $\sigma = 11,15$, estimated at 24 degrees of freedom, the difference D between the two variety means must be at least equal to

$$D_{0,05;\,24} = t_{0,05;\,24} \times \sigma \sqrt{\tfrac{2}{3}} = 2,064 \times \frac{11,15 \times 1,414}{1,73} = 18,80.$$

For instance, suppose the trial only comprises the varieties 1 and 3 with yields 176,0 and 97,7. Then we find the difference 176,0 − 97,7 = 78,2 to be much greater than 18,8, consequently these varieties belong to different populations. Briefly they really differ in productive capacity. In a trial with two varieties $t = D/\sigma \sqrt{2/3}$, is the root from the value found for F. The outcome of F-and t-test are not only of the same practical value but also identical. In this case the t-test renders the F-test superfluous.

In case the trial involves more than two varieties it is easy to see that any arbitrary use of the t-test may lead to absurd conclusions. Suppose the 13 varieties really belong to one population only and the value for F is found to be 1,0. The right conclusion can be immediately drawn from the F-test viz. that there is no indication for differences between the varieties. However, the 13 variety means will show a range of variation (= difference between the highest and lowest value = range) as is common for a random sample of 13 from a normal distribution. This range of variation is $\pm$ 3,34 times the standard deviation (see Table XX. Scores for ranked data: Statistical Tables of FISHER and YATES), being in our sample $3{,}34 \times 11{,}15/1{,}73 = 21{,}5$. This range is greater than the 18,80, which is needed according to the t-test. Therefore we are justified to expect in each trial with 13 varieties presenting no racial differences at all at least one difference of mean yields which should be significant according to the t-test. It is clear that the t-test should not be used in this manner. Then, is it impossible to use the t-test in a trial with 13 varieties? When it is desired to compare a pair of varieties or groups of varieties chosen *independent of a knowledge about the mean yields obtained in the experiment*, the t-test may be used. In fact, however, there are only two varieties or objects involved in such a trial. As soon as the trial contains more objects and more than one difference has to be tested, one is naturally testing sufficiently great differences, thus choosing the greatest. In applying the t-test in that case an „error of selection" would be made, for not an arbitrary deviation, but the greatest deviations from a normal population will be tested.

More or less intuitively I think, that I have succeeded in solving this problem by making use of a „range test", which is regularly applied by our Institute since early 1950. Results and practicability were satisfactory.

We use the table of the „studentized" range by E. S. PEARSON and H. O. HARTLEY (6). For our purpose we have added to this table as much graphic matter as could be safely done.

Along the vertical axis in the graph the range $R_{0,05}$ (see fig. 2) or $R_{0,01}$ (see fig. 3) has been plotted. The number of error degrees of freedom n_2 has been plotted along the horizontal axis. For each number of varieties $n_1 + 1$, a line has been traced representing the $R_{0,05}$ and $R_{0,01}$, respectively, at a certain value for n_2. In the example we worked with 24 error degrees of freedom. We only used the $R_{0,05}$. In the graph (fig. 2) the vertical line at $n_2 = 24$ intersects the curves of $n_1 + 1 = 13,12$ etc. down to 2, successively in the points $R_{0,05}$ = 5,22; 5,14; 5,05; 4,95; 4,83; 4,70; 4,54; 4,38; 4,18; 3,90; 3,53 and 2,92. Multiplying these numbers by the standard deviations of the variety means $\sigma/\sqrt{3} = 6{,}45$ gave the column $\triangle_{0,05}$ = 18,80; 22,73;..... 33,61 of fig. 1.

The graph is not very handy for frequent use. Therefore we have established an extensive table from the graph, with many values n_1 and n_2. In the numerical example we use the column $\triangle_{0,05}$ as follows: First the variety means are arranged from high to low, thus: variety 1 = 176,0, variety 11 = 152,7, etc...... variety 3 = 97,7. Next we ascertain whether the greatest difference between two varieties, i.e. 176,0 – 97,7 = 78,3, is greater than may be expected on a probability of only 0,05 for a random sample of thirteen from a normal distribution, of which the σ with 24 degrees of freedom has been estimated at $11{,}15/\sqrt{3}$. Stated briefly we ascertain if 78,3 is greater than 33,61, which indeed is the case.

We conclude now that the two varieties 1 and 3 belong to different populations, in

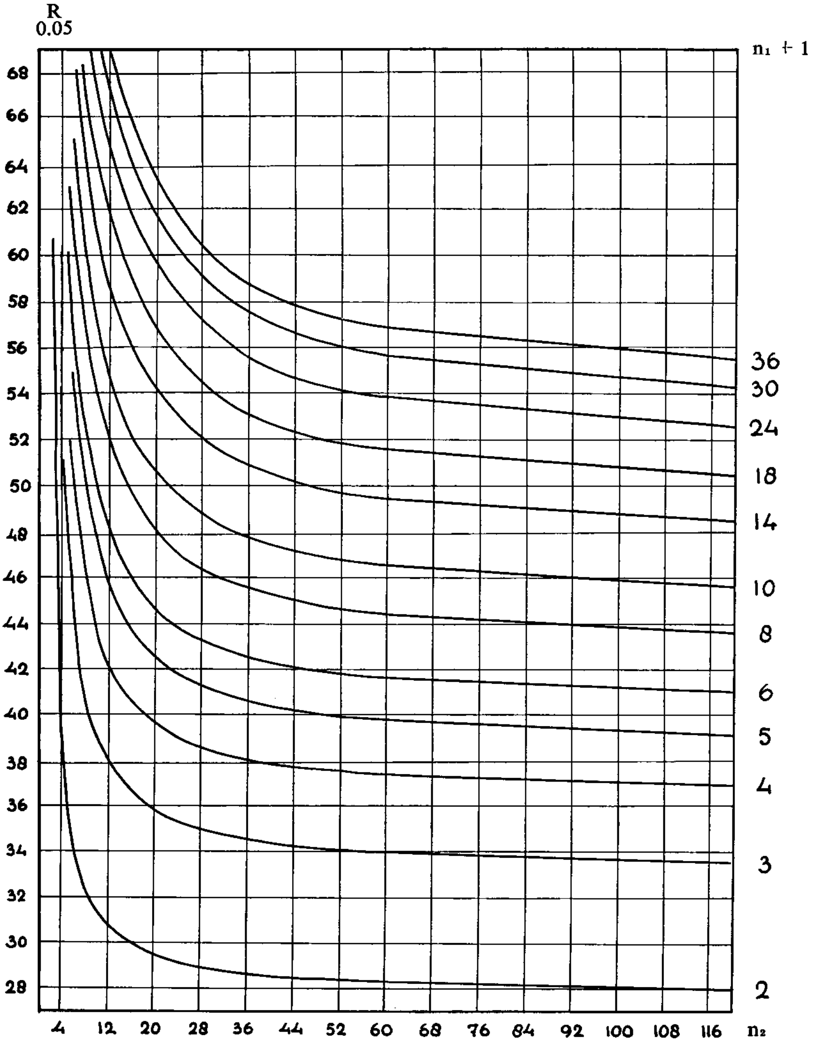

FIG. 2. GRAPH OF THE STUDENTIZED RANGE ACCORDING TO E. S. PEARSON AND H. O. HARTLEY. BIOMETRICA 33 (1943) : 89 (TABLE 2).

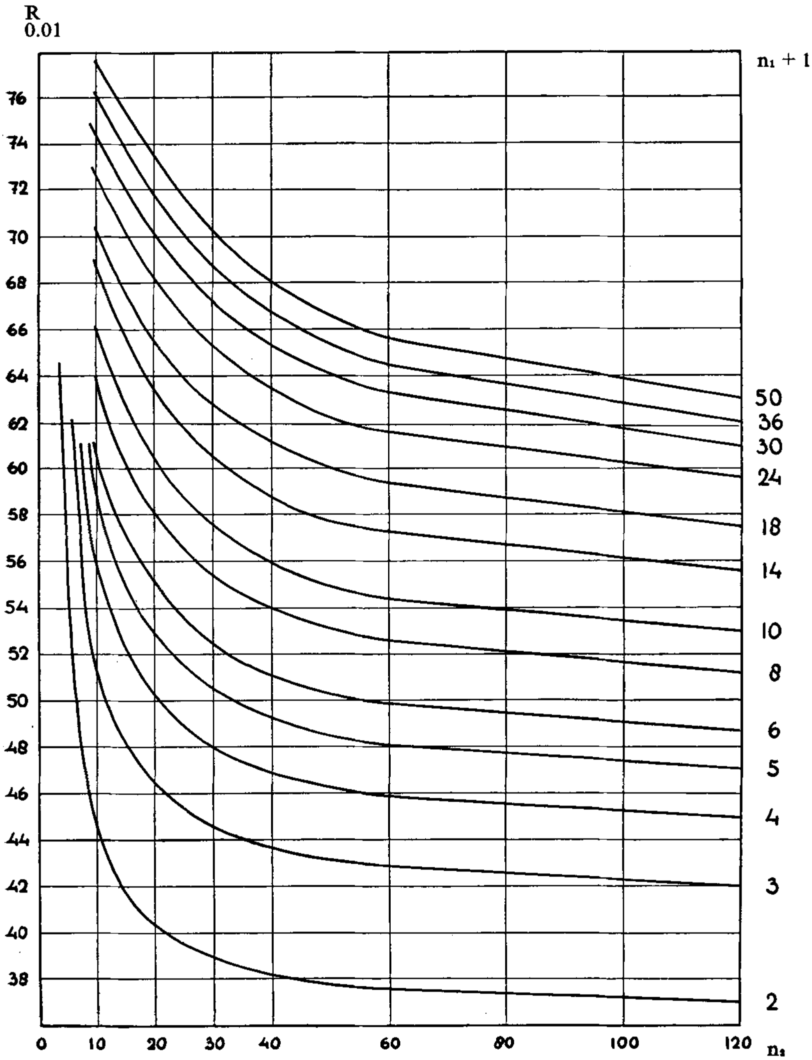

FIG. 3. GRAPH OF THE STUDENTIZED RANGE ACCORDING TO E. S. PEARSON AND H. O. HARTLEY. BIOMETRICA 33 (1943) : 89 (TABLE 2).

other words, that variety 1 „really" differs from variety 3. At this stage it is desirable to note that the „variety variance" 1393 is closely linked up to the range 78,3. From the fact that the $F = 11{,}21^{++}$ or $78{,}3 > 33{,}61$ we might draw exactly the same conclusion. Both criteria though not being identical, are nearly equivalent. Our experience is that, if $F > F_{0,05}$, also $D > \triangle_{0,05}$ holds good, so that we should also have been able to predict on account of the F-Test that the greatest difference i.e. that between the varieties 1 and 3 is „real". It is clear that the range test has the advantage of practically never contradicting the F-test (contrary to the t-test).

We cannot compare other pairs of varieties without obtaining small deviations from the 5 per cent basis, for we proceed as follows. After having concluded that variety 3 is distinct from variety 1, we ascertain whether 8 is distinct from variety 1. To be exact in this test we should not overlook the possibility that the first conclusion may not be right, even if the odds are less than 5 %, say 2,6 %. In that case the following should be taken into account. If our first conclusion is wrong, there are two possibilities.

1) There are no differences between the varieties 1, 2, 3,...., 13. Then there is only a risk of 2,6 % of making such wrong conclusion with respect to any couple of varieties.

2) There are real differences between some of the varieties but not between the couple of varieties with the largest range. The probability for such couple to show the largest difference will be certainly less than 2,6 %. So the probability of making wrongly the statement that the largest difference is real will be at most 2,6 %.

If our first conclusion is right one may ask: should the hypothesis that variety 1 and 8 are the extremes in a random test of 12 from one normal distribution be rejected? Criterion: is $176{,}0 - 100{,}7 = 75{,}3$ greater than 33,09?

If our first statement is wrong (the probability then of making the statement is at most 2,6 %) one may ask: should the hypothesis that variety 1 and 8 are the highest, respectively the lowest but one in a random test of 13 from one normal distribution be rejected? Criterion?

One should have to average both criteria in some way or other to obtain a criterion on a 5 basis for the highest and the lowest but one in the row of averages. This averaging, however, is impossible, as only one of the situations is the real one, briefly it is a problem difficult to solve. Therefore we have neglected the possible error in the first conclusion. Besides it is quite likely that the two criteria do not diverge very much. For both the requirements are less severe than for the greatest difference. Also the first criterion is more severe than the second from which the conclusion may be drawn that the difference 33,09 differs very little from the exact one and *on the safe side.*

Now variety 1 may be compared with the third and fourth lowest variety by removing each time a lowest variety from the random test. Thus variety 1 differs significantly from the 13th, the 12th, the 11th, and the 2nd variety. This result we indicate by giving to variety 1 the rank values 1 up to and including 1, in other words: variety 1 belongs to a population, to which at most the varieties 1 up to 1 inclusive, belong. Now this conclusion results from the foregoing twelve. There is only a small probability of the first conclusion being wrong, the chances that the next conclusions are wrong become greater and greater because accumulation of faulty conclusions may occur. The smaller the remaining group of varieties undergoing the range test the greater the chances of a less exact test. It may be noted that as soon as the group only consists of the varieties 1 and 11, viz. 176,0 and 152,7, the range test ($\triangle_{0,05} =$ 18,80 becomes identical with the t-test ($D_{0,05} = 18,80$).

Therefore the range test remains entirely on the safe side in regard to the t-test. We may also ascertain here how much the t-test will exaggerate the favorable results of the tests. For the greatest difference we find that $\triangle_{0,05} = 33,61$. $D_{0,05}$, on the contrary, $= 18,80$. Consequently the „accuracy" suggested by the t-test is $(33,61/18,80)^2$ or 3,2 times too big.

By eliminating each time the highest variety from the group of 13 varieties, a group of 5 varieties will remain: 10, 5, 2, 8, 3 with a range of variation 124,3 – 97,7 = 26,6, which is smaller than $\triangle_{0,05} = 26,91$. Consequently variety 3 does no longer differ significantly from the varieties 10, 5, 2 and 8 and has the values 9 up to and including 13. A variety 4, in the middle of the table, ranks eighth. Should 4 be placed into a group of the first 8 varieties, it becomes evident that 4 is different from the highest variety. If placed into a group of the six lower varieties, 4 differs from 12 up to 13 inclusive. Therefore variety 4 has the values 2 up to and including 11.

Conclusion

The idea to abolish the use of the t-test in connection with an analysis of variance, is no longer new. Some authors have never referred to the t-test in their publications on analyses of variance, others have mentioned the t-test in recent publications without accepting any responsibility as to the interpretation (Snedecor, 7, Cox and Cochran, 1).

D. Newman (4) is the first to mention the range test in connection with the analysis of variance in Biometrika 31, providing a table of ranges as well. Newman tries to group the objects. In 1943 Pearson and Hartley (6) published an emended and extended table of ranges. In Biometrics 1949 Tukey (8) provides some other criteria which can be used following up the analysis of variance. [1])

At first I myself, not knowing this literature, made a range by converting the F-test into a range test by means of table XX from Tables of Fisher and Yates:

$R_{0,05;\,n_1,\,n_2} = \sqrt{F}_{0,05;\,n_1,\,n_2} \times \hat{R}_{n_1,\,\infty}$

In a more mathematical way a fair approximation may be attained as investigated bij Patnaik (5) and Florin (2).

The problem forced itself on me after I had attended a lecture by Dr Drion for the „Studiekring voor Proeftechniek" (minutes of the 22nd meeting on 13 October 1948) on the 5 % and 1 % point for the greatest value in a series of F-values in the analysis of variance, a problem presenting itself in the 2-factor tests and bearing much resemblance to this question. I mention from this lecture the article by K. R. Nair (3).

Mathematical study about a good test remains to be done.

Summary

A numerical example is given of the analysis of variance applied on yields per cabbage.

After having concluded from a F-test, that the varieties show significant differences,

[1]) It may be remarked that the range test applied as indicated in our article gives more detailed conclusions than drawn by Newman in his two examples and the same results as attained by Tukey in the same two examples. Comparing our method with that of Tukey the former is more plausible.

a discussion is given of a new method to decide *which* varieties are different.

The t-test though in frequent use, gives wrong conclusions. The method indicated in this article diverges from those discussed by NEWMAN and TUKEY and is I suppose the more plausible.

SAMENVATTING

Het gebruik van de „studentized range" in verband met de variatieanalyse

Een rekenvoorbeeld wordt gegeven van de variatieanalyse toegepast op spitskool-opbrengstcijfers.

Na op grond van een F-test te hebben geconcludeerd dat de rassen duidelijke verschillen tonen, wordt een nieuwe methode besproken om uit te maken *welke* rassen verschillen. Hoewel de t-test hiertoe geregeld gebruikt wordt geeft deze verkeerde uitkomsten. De aangegeven methode wijkt af van die welke door NEWMAN en TUKEY worden besproken en lijkt mij de meest voor de hand liggende van de drie.

LIST OF REFERENCES

1. COX, G. and COCHRAN W. G., Experimental designs. John Wiley, New York, 1950.
2. FLORIN, H., Over het gebruik van de range tot schatten van de standaarddeviatie. Meded. v. d. Kon. Vlaamse Acad. van Wetensch. **12** (1951) nr 6.
3. NAIR, K. R., The studentized form of the extreme mean square test in the analysis of variance. Biometrika **35** (1948): 16.
4. NEWMAN, D., The distribution of range in samples from a normal population expressed in terms of an independent estimate of standard deviation. Biometrika **31** (1939): 20.
5. PATNAIK, B. P., The use of mean range as an estimator of variance in statistical tests. Biometrika **37** (1950): 78.
6. PEARSON, E. S. and HARTLEY, H. O., Tables of the probability integral of the studentized range. Biometrika **33** (1943): 89.
7. SNEDECOR, G. W., Statistical Methods (4th ed.), Iowa State College Press. 1946.
8. TUKEY, J. W., Comparing individual means in the analysis of variance. Biometrics **5** (1949): 99.