



TGP/8.5 Draft 3

ORIGINAL: English

DATE: May 6, 2004

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

Associated Document
to the
General Introduction to the Examination
of Distinctness, Uniformity and Stability and the
Development of Harmonized Descriptions of New Varieties of Plants (document TG/1/3)

DOCUMENT TGP/8
“USE OF STATISTICAL PROCEDURES IN
DISTINCTNESS, UNIFORMITY AND STABILITY TESTING”

Section TGP/8.5: Statistical Methods for DUS Examination

Document prepared by experts from the United Kingdom

to be considered by the

*Technical Working Party on Automation and Computer Programs (TWC),
at its twenty-second session to be held in Tsukuba, Japan, from June 14 to 17, 2004*

SECTION 8.5 STATISTICAL METHODS FOR DUS EXAMINATION	3
8.5.1 Analysis of variance	3
<u>Two-way anova</u>	3
<i>The data model</i>	3
<i>The two-way analysis of variance table</i>	5
<u>One-way anova</u>	6
<i>The data model</i>	6
<i>The one-way analysis of variance table</i>	8
<u>Assumptions about the data</u>	9
<u>The precision of means and the differences between means</u>	10
<u>Comparisons of means</u>	10
<u>Higher-order anova's</u>	11
<u>Unbalanced data and the method of Fitting Constants</u>	12
<u>The paired t-test</u>	13
<i>The paired t-test using a one-sample t-test of differences</i>	13
<i>The paired t-test using two way anova</i>	14
APPENDICES.....	16
<u>Example of two-way anova (same type as Example A)</u>	16
<u>Example of two-way anova (same type as Example B)</u>	17
<u>Example of one-way anova (same type as Example C)</u>	19
<u>Example of one-way anova (same type as Example D)</u>	20
<u>Example of a paired t-test (same type as Example E)</u>	22
<i>The paired t-test analysis using a one-sample t-test of differences</i>	22
<i>The paired t-test analysis using two way anova</i>	23

SECTION 8.5

STATISTICAL METHODS FOR DUS EXAMINATION

8.5.1 Analysis of variance

1. The analysis of variance (anova) of data from a designed experiment has two purposes. Firstly it subdivides the data's total variation into separate components with each component representing a different source of variation, so that the relative importance of the different sources can be assessed. Secondly it provides an estimate of the random variation in the data. This may be used as an estimate of precision when comparing means calculated from the data.

2. Anova can take many forms. Just two forms will be considered in detail here. These are the two forms which arise as part of the statistical techniques recommended by UPOV. At their simplest, both operate on an $n \times m$ table of data. They are the :-

- Two-way anova, e.g. as used in the analysis of variety-by-year means for a characteristic for v varieties grown in each of y years in the Combined Over Years Distinctness (COYD) criterion. The paired t-test is a special case of two-way anova.
- One-way anova, e.g. as used in the analysis of variety-by-year adjusted $\log(\text{SD}+1)$'s (a measure of uniformity) for a characteristic for v reference varieties grown in each of y years in the Combined Over Years Uniformity (COYU) criterion.

3. The particular form an anova takes depends on the origins of the data. This determines the model for the data, i.e. what factors are likely to cause the data to vary, which in turn determines what components the total variation is divided into, and hence the form of the anova.

Two-way anova*The data model*

4. In two-way anova, the $n \times m$ table of data corresponds to nm data values classified by two factors: Factor 1 with m levels and Factor 2 with n levels. Usually, only one of the factors will be of interest while the other will be present simply because it explains variability in the data. For example (Example A), each data value might be the mean over all plants in a plot for a characteristic from a trial with vb plots laid out in b blocks (Factor 1) of v varieties (Factor 2), the factor of interest.

Example A: data from a trial with v varieties and b blocks

	Block 1	Block 2	Block 3	...	Block b
Variety 1	-	-	-	...	-
Variety 2	-	-	-	...	-
Variety 3	-	-	-	...	-
...
Variety v	-	-	-	...	-

Diagram annotations:

- mean of all plants in the plot in block b with variety 2 (arrow pointing to the cell for Variety 1, Block b)
- vb data values (arrow pointing to the entire data table)
- mean of all plants in the plot in block 2 with variety 3 (arrow pointing to the cell for Variety 3, Block 2)

5. Alternatively, for the above COYD example (Example B), the data values might consist of the vy variety-by-year means for a characteristic for v varieties (Factor 2, the factor of interest) grown in each of y years (Factor 1).

Example B: data for COYD example with v varieties grown in y years

	Year 1	Year 2	Year 3	...	Year y
Variety 1	-	-	-	...	-
Variety 2	-	-	-	...	-
Variety 3	-	-	-	...	-
...
Variety v	-	-	-	...	-

variety-by-year mean for variety 2 in year 3

vy data values

variety-by-year mean for variety v in year 2

6. If x represents one of the nm data values in the $n \times m$ table of data, the model explaining the variation in the data is as follows :-

$$x = \text{Factor 1 effect} + \text{Factor 2 effect} + \left(\text{Factor 1} \times \text{Factor 2} \text{ interaction effect} + \text{random variation} \right)$$

Thus each of the nm data values is made up of a sum of effects. The “Factor 1 effect” and “Factor 2 effect” are due to the particular levels of Factors 1 and 2 influencing the data value. An interaction between two factors is when the effects of one factor differ, i.e. are inconsistent, from level to level of the other factor. So a Factor 1×Factor 2 interaction is when Factor 1 effects differ from level to level of Factor 2. In the above model the remainder, or residual, of the data value is the amount, additional to its Factor 1 effect and its Factor 2 effect, which appears to be due to its particular combination of Factor 1 and 2 levels. This amount might be partly due to a genuine interaction or it might be just due to random variation. As there is just a single data value for each combination of Factor 1 and 2 levels it is not possible to determine which.

7. For Example A, the model explaining the variation in the data is as follows :

$$x = \text{block effect} + \text{variety effect} + \left(\text{variety} \times \text{block} \text{ interaction effect} + \text{random variation} \right)$$

In this example each of the vb data values (one from each of the vb plots) is made up of the sum of a “block effect”, due to the block the plot is in, plus a “variety effect”, due to the variety sown on the plot, plus a remainder. This remainder, or residual represents the amount that is additional to the variety and block effects. As there is just a single data value for each variety in each block, it is not possible to separate interaction effects and random variation. However, as the variety effects are not expected to differ from block to block, or in other words any variety×block interaction is expected to be negligible, the residual is likely to be due to random variation.

8. For Example B (COYD), the model explaining the variation in the data is as follows :-

$$x = \text{year effect} + \text{variety effect} + \left(\begin{array}{l} \text{variety} \times \text{year} \\ \text{interaction effect} \end{array} + \text{random variation} \right)$$

Here each of the vy variety-by-year means is made up of a sum of effects. The “year effect” is an amount due to the year the variety-by-year mean was recorded in. The year effects might or might not be the same for all years. The “variety effect” is an amount due to the variety of the variety-by-year mean, and might or might not be the same for all varieties. The remainder, or residual of the variety-by-year mean represents the amount that is additional to its variety effect and its year effect, which appears to be due to that particular variety in that particular year. This amount may be partly due to a genuine variety×year interaction effect or it may be just due to random variation caused by the means having been calculated from different plants grown on different plots, and possibly due to measurement error. As there is just a single variety-by-year mean for each variety in each year it is not possible to distinguish between interaction effects and random variation.

The two-way analysis of variance table

9. Two-way anova produces a table as follows:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Factor 1	$m - 1$	-	-	-
Factor 2	$n - 1$	-	-	-
Residual	$(n - 1)(m - 1)$	-	-	-
Total	$nm - 1$	-	-	-

[As the anova computations are likely to be done by computer, details are not given here. The interested reader can find them in a good statistical book such as DAGNELIE (1998 and 1981), Kala (2002), Mead et al (1993), and Sokal and Rohlf (1995).]

10. For Example A, the two-way anova table is as follows:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Block	$b - 1$	-	-	-
Variety	$v - 1$	-	-	-
Residual	$(b - 1)(v - 1)$	-	-	-
Total	$vb - 1$	-	-	-

11. For Example B (COYD), the two-way anova table is as follows:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Year	$y - 1$	-	-	-
Variety	$v - 1$	-	-	-
Residual	$(y - 1)(v - 1)$	-	-	-
Total	$vy - 1$	-	-	-

12. The total variation in the data is measured by the Total Sum of Squares, which is the sum of the squared deviations of all the data from their mean, i.e. $\sum(x - \bar{x})^2$. It is subdivided into “sums of squares” representing the three component sources of variation included in the data model: variation due to Factor 1, variation due to Factor 2 and residual variation. These sums of squares are divided by their degrees of freedom (df) to give “mean squares”, which are directly compared in order to assess the relative magnitude of the different sources of variation. This is done in the final column where the F-ratio’s are the ratios of each of the Factor 1 and Factor 2 mean squares to the residual mean square. Providing the assumptions discussed below about the data are valid, these F-ratio’s can be used to perform “F-test’s” of the significance of the variation due to each of Factors 1 and 2, i.e. used to test whether Factors 1 and 2 have significant effects. This is done by comparing the F-ratio’s with critical values from F tables on the df of the numerator and the denominator mean squares: if the F-ratio is greater than or equal to the $100 \times p\%$ critical value, the factor is declared to have a significant effect at the $p\%$ level. Otherwise, it does not have a significant effect at the $p\%$ level.

13. The residual mean square is a variance. It estimates the combined variation due to any Factor 1×Factor 2 interaction and random variation. Hence, it is often referred to as the “Factor 1-by-Factor 2 mean square”, e.g. the “variety-by-block mean square” in Example A or the “variety-by-year mean square” in the COYD example (Example B).

14. Statistical theory shows that in two-way anova it is appropriate to use the residual mean square to estimate the variance or standard errors of means calculated from the data. This is the case whether the Factor 1×Factor 2 interaction is assumed to be negligible, such as in Example A, or not, such as in the COYD example (Example B).

15. Worked examples of two-way anova are given in Appendix A1. These are of the same types as Examples A and B above.

One-way anova

The data model

16. In one-way anova, the $n \times m$ table of data corresponds to data classified into m groups by a single factor of interest such that there are n independent replicates within each group. It is important to note that the replicates are different within each group, i.e. there is nothing in common between the i^{th} replicate in one group and the i^{th} replicate in another group. For example (Example C), each data value might be the plot yield from a trial with tr plots laid out with r replicate plots of each of t treatments (the factor of interest). Alternatively each data value might be the mean over all plants in a plot for a characteristic for the tr plots.

Example C: data from a trial with r replicate plots of t treatments

	Treat' 1	Treat' 2	Treat' 3	...	Treat' t
Replicate 1	-	-	-	...	-
Replicate 2	-	-	-	...	-
Replicate 3	-	-	-	...	-
...
Replicate r	-	-	-	...	-

plot yield in replicate 2 plot of treatment t

tr data values

plot yield in replicate r plot of treatment 2

17. Alternatively, as in the above COYU example (Example D), they might consist of variety-by-year adjusted $\log(\text{SD}+1)$'s for v reference varieties (replicates) grown in each of y years (the factor of interest) for a characteristic.

Example D: data for COYU example with v varieties grown in y years

	Year 1	Year 2	Year 3	...	Year y
Variety 1	-	-	-	...	-
Variety 2	-	-	-	...	-
Variety 3	-	-	-	...	-
...
Variety v	-	-	-	...	-

variety-by-year adjusted $\log(\text{SD}+1)$ for variety 2 in year 3

vy data values

variety-by-year adjusted $\log(\text{SD}+1)$ for variety v in year 2

18. It might surprise the reader to see the adjusted $\log(\text{SD}+1)$'s of the v varieties within a year regarded just as replicates, and not as a second factor like in two-way anova. Year is included as a factor in the anova because the overall levels of uniformity, as measured by the adjusted $\log(\text{SD}+1)$'s, can be expected to vary from year to year. However, by regarding the uniformity values of the v varieties within a year as replicates, this allows the variation between them to be used as an estimate of the random variation in uniformity between the reference varieties, which are all considered to be uniform. It will be seen in the following that this estimate of the random variation in uniformity among the (uniform) reference varieties is used to compare the uniformity of a candidate variety with the mean uniformity of the (uniform) reference varieties. This is done in order to see whether the uniformity of the candidate variety is extreme relative to the uniformity of the reference varieties.

19. If x represents one of the nm data values in the $n \times m$ table of data, the model explaining the variation in the data is as follows :-

$$x = \text{Factor effect} + \text{random variation}$$

Thus each of the nm data values is made up of the sum of a "Factor effect", which is due to the particular level of the factor influencing the data value, plus a residual amount which is random variation. This means that the variation between data values within a group is considered to be random variation.

20. For Example C, the model explaining the variation in the data is as follows :-

$$x = \text{treatment effect} + \text{random variation}$$

Here each of the tr plot yields is the sum of a “treatment effect”, due to the treatment the plot receives, plus a residual amount due to random variation.

21. For Example D (COYU), the model explaining the variation in the data is as follows :-

$$x = \text{year effect} + \text{random variation}$$

Here each of the vy variety-by-year adjusted $\log(\text{SD}+1)$'s, i.e. measures of uniformity, is the sum of a “year effect” plus a residual amount due to random variation. This is equivalent to recognising that uniformity is likely to vary from year to year, and is expected to vary at random from variety to variety within a year. Note: the absence of a variety effect in the model indicates that, within a year, apart from random variation, uniformity is expected to be the same for all reference varieties.

The one-way analysis of variance table

22. One-way anova produces a table as follows:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Factor	$m - 1$	-	-	-
Residual	$m(n - 1)$	-	-	-
Total	$nm - 1$	-		

[Again, details of the anova computations are not given here but may be found in DAGNELIE (1998 and 1981), Kala (2002), Mead et al (1993), and Sokal and Rohlf (1995).]

23. For Example C, the one-way anova table is as follows:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Treatment	$t - 1$	-	-	-
Residual	$t(r - 1)$	-	-	-
Total	$tr - 1$	-		

24. For Example D (COYU), the one-way anova table is as follows:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Year	$y - 1$	-	-	-
Residual	$y(v - 1)$	-	-	-
Total	$vy - 1$	-	-	-

25. The total variation in the data is subdivided into “sums of squares” representing the two component sources of variation in the data model, i.e. variation due to the factor of interest and residual or random variation. The sums of squares are divided by their degrees of freedom (df) to give directly comparable “mean squares” used to compare the two sources of variation. This is done in the final column where the F-ratio is the ratio of the factor mean square to the residual mean square. Providing the assumptions discussed below about the data are valid, comparison of the F-ratio with critical values from F tables on $m - 1$ and $m(n - 1)$ df provides an “F-test” of the significance of the variation due to the factor of interest, i.e. whether that factor has a significant effect. If the F-ratio is greater than or equal to the $100 \times p\%$ critical value, the factor of interest is declared to have a significant effect at the $p\%$ level. Otherwise, it does not have a significant effect at the $p\%$ level.

26. The residual mean square is a measure pooled over groups of the variation in the data from replicate to replicate within a group. Thus it is a variance and estimates the random variation in the $n \times m$ table of data that has been analysed. Consequently, it can be used to estimate the variance or standard errors of means calculated from the data.

27. Worked examples of one-way anova are given in Appendix A2. These are of the same types as Examples C and D above.

Assumptions about the data

28. Two assumptions are necessary for one-way and two-way anova. They are that:-

- (a) the variability of the data in the $n \times m$ table of data is the same for the different levels of the classifying factors.

Thus, for two-way anova it is assumed that the variability of the vb means is the same for all varieties and for all blocks in Example A, and that the variability of the vy variety-by-year means is the same for all varieties and for all years in Example B (COYD).

For one-way anova it is assumed that the variation in the plot yields between the replicates within a treatment is the same for all treatments in Example C. Likewise, in the one-way anova COYU example (Example D) it is assumed that the variation between the adjusted $\log(SD+1)$'s of the different varieties in a year is the same from year to year;

- (b) the model describes the data adequately in that the effects of the classifying

factors are additive. For example (Example A), it is assumed that the expected difference in the data values for two varieties is the same in one block as it is in any other block.

Failure of this assumption will lead to large residuals, as the residual is the part of a data value which is not explained by the additive factor effects. This in turn will lead to a large residual mean square, which will give large standard errors of means, and so large differences will be required between factor means in order for them to be declared significant.

An example of such a failure would be when a variety×year interaction occurs, i.e. when the variety effects are inconsistent over years in the two-way anova COYD example (Example B). Here only large differences between varieties would be declared significant.

29. For F-tests of ratios of anova mean squares and t-tests based on anova mean squares (or the equivalent use of LSD's based on anova mean squares) it is also necessary to assume that the data values are independent and that the random variation in the data has an approximately Normal distribution.

The precision of means and the differences between means

30. Let \bar{x}_1 and \bar{x}_2 be factor means of r_1 and r_2 data values from the $n \times m$ table of data that has been analysed by anova (or from equivalent data). The precision of a mean \bar{x}_i , where $i = 1$ or 2 , is measured by its standard error ($SE(\bar{x}_i)$), which is estimated by

$$SE(\bar{x}_i) = \sqrt{\frac{RMS}{r_i}}$$

Where RMS is the residual mean square from the anova of the $n \times m$ table of data. The precision of the difference in two means, $\bar{x}_1 - \bar{x}_2$, is measured by its standard error, ($SE(\bar{x}_1 - \bar{x}_2)$), which is estimated by

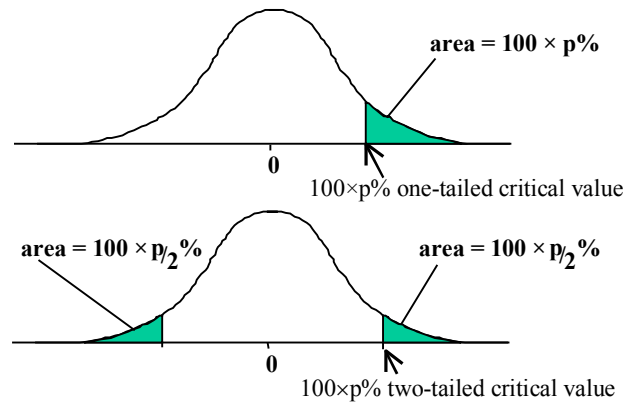
$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{RMS \left(\frac{1}{r_1} + \frac{1}{r_2} \right)}$$

Comparisons of means

31. The significance of the difference between \bar{x}_1 and \bar{x}_2 can be tested by either:-

- Comparing the two sample t-statistic $t = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}$ with critical values from

Student's t-tables on the df of the RMS . If it is known *a priori* which mean will be larger, the critical values should be the one-tailed values, and the test is called a "one-tailed test". If it is NOT known *a priori* which mean will be larger, the critical values should be the two-tailed values, and the test is called a "two-tailed test".



If the absolute value of t is greater than or equal to the $100 \times p\%$ critical value, the means are declared to be significantly different at the $p\%$ level. Otherwise, the means are declared to be not significantly different at the $p\%$ level.

- Comparing the absolute difference between the means, $|\bar{x}_1 - \bar{x}_2|$, with the $100 \times p\%$ least significant difference (LSD), i.e. comparing $|\bar{x}_1 - \bar{x}_2|$ with $t \times SE(\bar{x}_1 - \bar{x}_2)$

where t is the $100 \times p\%$ critical value from Student's t -tables on the df of the RMS . The critical value should be the one-tailed value if it is known *a priori* which mean will be larger, and two-tailed otherwise.

If the absolute difference between the means is greater than or equal to the $100 \times p\%$ LSD, the means are declared to be significantly different at the $p\%$ level. Otherwise, the means are declared to be not significantly different at the $p\%$ level.

Higher-order anova's

32. Anova has been introduced as the subdivision of the total variation among the data values in an $n \times m$ table of data such that it :-

- allows a comparison of the different sources of variation
- provides an estimate of the random variation affecting the nm data values

The $n \times m$ table of data may consist of means calculated from a higher order table of data such as an $l \times n \times m$ table of data or a $k \times l \times n \times m$ table of data. For example, in the above two-way anova COYD example (Example B) the $v \times y$ table of data are variety-by-year means for v varieties in each of y years and these can be viewed as having been calculated from an $l \times v \times y$ table of plot means from trials with l blocks and v varieties in each of y years. Alternatively they can be viewed as having been calculated from a $k \times l \times v \times y$ table of measurements on k plants in each plot of trials with l blocks and v varieties in each of y years.

33. If the $n \times m$ table of data takes the form of means calculated from a higher order table, the experimenter can analyse the data using a multi-way anova that is a logical extension of the two-way anova. In this the total variation is subdivided into components for each of the factors classifying the data table plus components for two-way, three-way and all higher order interactions between the factors. As with the two-way anova, the components of variation can

be compared using ratios of mean squares. Also the residual mean square is a variance which estimates the random variation at the level of the data values in the table of data that has been analysed.

34. Given data values in a more-than-two-way table of data, the experimenter has the choice of analysing it by multi-way anova or by calculating an $n \times m$ table of means and using two-way anova. If the data values in the $n \times m$ table are the means of r of the original data values, then the mean squares in the multi-way anova are r times the size of the equivalent mean squares in the two-way anova. Thus whichever approach is used, the relative sizes of the mean squares are the same and the variances estimated by the residual mean square in the two-way anova can be derived from the mean squares in the higher order anova. However, it is important for the purposes of COYD that the variety means are compared using variances or standard errors based on the variety-by-year mean square as an estimate of random variation, such as is provided by the residual mean square in two-way anova.

Unbalanced data and the method of Fitting Constants

35. An $n \times m$ table of data that has a data value present in each of the nm table cells is balanced. If data values are missing from one or more cells it is unbalanced or incomplete.

36. Although the data for one-way anova was introduced for simplicity as having equal (n) replication in each of the m groups, i.e. balanced, this is not a necessary requirement for one-way anova. If, instead of being balanced, the data consist of a total of w data values unequally replicated within the m groups, the computations are straight forward, and the anova table looks similar to that for one-way anova except that the df differ. The total df is $w-1$, the factor df is $m-1$, and the residual df is calculated by subtracting the factor df from the total df, i.e. $(w-1) - (m-1)$.

37. By contrast, the data for two-way anova must be balanced. If the data is unbalanced, i.e. some of the $n \times m$ table's cells have no data, two-way anova cannot be used. Instead, some other method of analysing two-way data such as the method of Fitting Constants or Fitcon (Yates (1933)) or restricted maximum likelihood (REML) (Patterson and Thompson (1971)) must be used. Unbalanced data would arise in the above two-way anova COYD example (Example B) if one or more of the v varieties either was not present or failed to grow in one or more of the y years. It also arises in the calculation of Long Term LSD for use in Long Term COYD. In this a table of variety-by-year means that extends over more years and varieties than are present in the test years is to be analysed. As not all varieties are present in all years, this table is unbalanced. Like two-way anova, Fitcon subdivides the total variation in the data into different components for the different sources of variation, and the residual mean square provides an estimate of the random error variation. The df are as for the two-way anova except that the total df is $w-1$ where w is the total number of data values, and the residual df is calculated by subtracting the df for each of Factor 1 and Factor 2 from the total df, i.e. $(w-1) - (m-1) - (n-1)$.

38. If the data is unbalanced in either one-way anova or two-way analysis such as Fitcon or REML, the standard errors needed for LSD's or t-tests for comparing factor means are more complicated because they differ depending on which factor mean is being compared with which. In one-way anova this simply affects the replication of the factor means (r_1 and r_2) and the standard errors are calculated as given above. In two-way analysis the standard errors are further complicated because they depend on the pattern of missing values for the two

factor means being compared. However they are easily calculated by a computer program.

The paired t-test

39. The paired t-test is a special case of two-way anova where the factor of interest has just two levels. Thus, it is used when there is an $n \times 2$ table of data corresponding to $2n$ data values classified by two factors: Factor 1 with n levels and Factor 2, the factor of interest with 2 levels.

40. Examples of the paired t-test include:-

- (a) Each data value might be the plot yield from a trial with $2b$ plots laid out in b blocks (Factor 1) of 2 treatments (Factor 2);
- (b) in DUS testing when a special test is set up to test distinctness between two varieties using additional characteristics. A number of plants of each variety are sown in a trial. This is repeated in time with s independent sowings. In this example (Example E) the data values are the $2s$ means for the additional characteristic, each based on a fixed number of established plants, from the s sowings (Factor 1) of the two varieties (Factor 2).

Example E: data from a special test with 2 varieties and s sowings

	Sowing 1	Sowing 2	Sowing 3	...	Sowing s
Variety 1	-	-	-	...	-
Variety 2	-	-	-	...	-

41. The analysis can be done in either of the two following ways:-

The paired t-test using a one-sample t-test of differences

42. The difference in the two data values (one data value for each of the two levels of Factor 2, the factor of interest) for each of the n levels of Factor 1 is obtained. This produces n differences denoted $d_i, i=1, \dots, n$. The mean difference and the variance of the differences are calculated and used to produce the one sample t-statistic as follows:-

$$t = \frac{\bar{d}}{SE(\bar{d})}$$

where $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i$ is the mean difference, and $SE(\bar{d}) = \sqrt{\frac{s_d^2}{n}}$ is the standard error of the mean

difference, and $s_d^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2$ is the variance of the differences, which is more easily

$$\text{computed as } s_d^2 = \frac{1}{n-1} \left[\sum_{i=1}^n d_i^2 - \frac{\left(\sum_{i=1}^n d_i \right)^2}{n} \right].$$

Providing the assumptions about the data discussed above are valid, this t-statistic can be compared with critical values from Student's t-tables on $n-1$ df. The critical values should be one-tailed values if the sign of the mean difference is known *a priori*, and two-tailed values otherwise. The test is of whether the mean difference is significantly different from zero, i.e. whether Factor 2 has a significant effect.

If the absolute value of t is greater than or equal to the $100 \times p\%$ critical value, the mean difference is declared to be significantly different from zero at the $p\%$ level. Otherwise, the mean difference is declared to be not significantly different from zero at the $p\%$ level.

The paired t-test using two way anova

43. A standard two-way anova of the $n \times 2$ table of data will provide an anova table with F-ratio's of each of the Factor 1 and Factor 2 mean squares to the residual mean square. Providing the assumptions about the data discussed above are valid, comparison of these F-ratio's with F tables on the df of the numerator and the denominator mean squares will provide "F-test's" to test whether Factors 1 and 2 have significant effects.

44. For Example E, the two-way anova table is as follows:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Sowing Variety	$s - 1$	-	-	-
Residual	1	-	-	-
Total	$s - 1$	-	-	-
	$2s - 1$	-		

45. The residual mean square from the two-way anova can be used to estimate the variance or standard errors of means calculated from the data. These can be used to calculate a two-sample t-statistic (or the equivalent LSD), which can be used to test the significance of the difference between the two factor means for Factor 2, i.e. test whether Factor 2 has a significant effect

46. Whichever method of analysis is used, the significance of the test of the effect of Factor 2 will be the same. In other words the significance of the t-test in the former method will be the same as the significance of the F-test of the Factor 2 effects and the significance of the t-test of the two Factor 2 means in the latter method.

47. A worked example of a paired t-test of the same type as Example E is given in Appendix A3.

REFERENCES

Dagnelie Pierre. (1981). *Principes d'expérimentation*

Dagnelie Pierre. (1998). *Statistique théorique et appliquée volume 2 inférence statistique à une et deux dimensions*. Bibliothèque des universités Statistique

Kala, R. (2002). *Statystyka dla przyrodników*, Agric. Univ. of Poznan

Mead, R., Curnow, R. N. and Hasted, R. M. (1993). *Statistical Methods in Agriculture and Experimental Biology*. Chapman & Hall, London.

Patterson, H. D. and Thompson, R. (1971). *Recovery of interblock information when block sizes are unequal*. Biometrika, 58, 545-554.

Sokal, R. R. and Rohlf, F. J. (1995). *Biometry*, W. H. Freeman Company

Yates, F. (1933). *The principles of orthogonality and confounding in replicated experiments*. Journal of Agricultural Science, Cambridge, 23, 108-45.

APPENDICES

APPENDIX A1

Example of two-way anova (same type as Example A)

1. A trial with four blocks of five varieties of Kale was sown with 15 plants per plot. The data below are the plot means over all the plants in a plot for the “petiole length in mm” characteristic.

Variety	Block			
	I	II	III	IV
J	361	375	361	399
K	388	383	376	401
L	356	386	365	382
M	383	373	385	405
N	386	385	389	413

2. Two-way anova of the data produces the following table:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Block	3	2116.00	705.33	9.95
Variety	4	1316.30	329.08	4.64
Residual	12	850.50	70.88	
Total	19	4282.80		

3. From F-tables the 5%, 1% and 0.1% critical values on 4 and 12 df are 3.259, 5.412 and 9.633 respectively. Comparison of the Variety F-ratio with these shows that there is a significant variety effect at the 5% level ($P < 0.05$). The residual mean square or variety-by-block mean square is an estimate of the random variability of the 20 values in the above data table. It may be used to estimate the variance or standard errors of means calculated from the data.

4. The variety means are the means of 4 data values and are as follows:-

Variety	Mean
J	374.0
K	387.0
L	372.2
M	386.5
N	393.2

5. Their standard error, $SE(\bar{x})$, is estimated by

$$SE(\bar{x}) = \sqrt{\frac{RMS}{r}} = \sqrt{\frac{70.88}{4}} = 4.21$$

6. The standard error of the difference in two means, $(SE(\bar{x}_1 - \bar{x}_2))$, is estimated by

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{RMS\left(\frac{1}{r_1} + \frac{1}{r_2}\right)} = \sqrt{70.88\left(\frac{1}{4} + \frac{1}{4}\right)} = 5.95$$

7. The significance of the difference between pairs of variety means can be tested by comparing the absolute difference between pairs of means with, for example, the 5% LSD, where

$$5\% \text{ LSD} = t \times SE(\bar{x}_1 - \bar{x}_2) = 2.179 \times 5.95 = 12.97$$

and t is the 5% two-tailed critical value from Student's t-tables on 12 df. Thus varieties J and K are significantly different at the 5% level, whereas varieties J and L and varieties M and N are not significantly different at the 5% level etc..

Example of two-way anova (same type as Example B)

8. This example illustrates the calculation of the COYD criterion. The data are the variety-by-year means for 11 varieties of italian ryegrass in three years for the "plant width in cm at ear emergence" characteristic.

Variety	Year 1	Year 2	Year 3
L	60.66	61.47	55.18
N	58.91	62.28	55.66
O	54.46	56.68	51.32
P	57.69	54.75	54.94
Q	56.57	57.62	51.46
R	51.33	53.40	49.18
S	58.59	59.08	51.67
T	63.47	58.94	54.97
V	66.14	65.49	60.15
W	62.63	63.90	58.84
AC	60.36	58.42	58.51

9. Two-way anova of the data produces the following table:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Year	2	148.821	74.4106	26.843
Variety	10	383.679	38.3679	13.841
Residual	20	55.443	2.7721	
Total	31	587.944		

10. From F-tables the 5%, 1% and 0.1% critical values on 10 and 20 df are 2.348, 3.368 and 5.075 respectively. Comparison of the Variety F-ratio with these shows that there is a very highly significant variety effect at the 0.1% level ($P < 0.001$). The residual mean square or variety-by-year mean square is an estimate of the random variability of the 33 values in the

above data table. It may be used to estimate the variance or standard errors of means calculated from the data.

11. The variety means are the means of 3 data values and are as follows:-

Variety	Mean
L	59.103
N	58.950
O	54.153
P	55.793
Q	55.217
R	51.303
S	56.447
T	59.127
V	63.927
W	61.790
AC	59.097

12. Their standard error, $SE(\bar{x})$, is estimated by

$$SE(\bar{x}) = \sqrt{\frac{RMS}{r}} = \sqrt{\frac{2.7721}{3}} = 0.9613$$

13. The standard error of the difference in two means, $(SE(\bar{x}_1 - \bar{x}_2))$, is estimated by

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{RMS \left(\frac{1}{r_1} + \frac{1}{r_2} \right)} = \sqrt{2.7721 \left(\frac{1}{3} + \frac{1}{3} \right)} = 1.3594$$

14. The significance of the difference between pairs of variety means can be tested by comparing the absolute difference between pairs of means with the 1% LSD where

$$1\% \text{ LSD} = t \times SE(\bar{x}_1 - \bar{x}_2) = 2.845 \times 1.3594 = 3.868$$

and t is the 1% two-tailed critical value from Student's t-tables on 20 df. Thus varieties L and N are not significantly different at the 1% level, whereas varieties L and O and varieties L and Q are significantly different at the 1% level etc..

15. For more information on the COYD criterion see TGP 9.7.

APPENDIX A2

Example of one-way anova (same type as Example C)

1. The data below are the plot yields in Kg from a glasshouse experiment consisting of 18 pots (plots) of potato plants to which 3 replicates of each of six fungicide treatments were allocated at random.

Replicate	Treat' 1	Treat' 2	Treat' 3	Treat' 4	Treat' 5	Treat' 6
1	1.07	0.23	1.07	0.66	1.07	0.91
2	0.74	0.54	0.63	0.85	1.31	0.94
3	0.89	0.57	1.08	0.78	1.50	0.66

2. One-way anova of the data produces the following table:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Treatments	5	1.1236	0.2247	6.48
Residual	12	0.4161	0.0347	
Total	17	1.5398		

3. From F-tables the 5%, 1% and 0.1% critical values on 5 and 12 df are 3.106, 5.064 and 8.892 respectively. Comparison of the Treatment F-ratio with these shows that there is a highly significant treatment effect at the 1% level ($P < 0.01$) on the plot yield. The residual mean square is an estimate of the random variability of the 18 values in the above data table. It may be used to estimate the variance or standard errors of means calculated from the data.

4. The treatment means are the means of 3 data values and are as follows:-

Treatment	Mean
1	0.900
2	0.447
3	0.927
4	0.763
5	1.293
6	0.837

5. Their standard error, $SE(\bar{x})$, is estimated by

$$SE(\bar{x}) = \sqrt{\frac{RMS}{r}} = \sqrt{\frac{0.0347}{3}} = 0.1075$$

6. The standard error of the difference in two means, $(SE(\bar{x}_1 - \bar{x}_2))$, is estimated by

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{RMS \left(\frac{1}{r_1} + \frac{1}{r_2} \right)} = \sqrt{0.0347 \left(\frac{1}{3} + \frac{1}{3} \right)} = 0.1521$$

7. The significance of the difference between pairs of treatment means can be tested by comparing the absolute difference between pairs of means with, for example, the 5% LSD, where

$$5\% \text{ LSD} = t \times \text{SE}(\bar{x}_1 - \bar{x}_2) = 2.179 \times 0.1521 = 0.3313$$

and t is the 5% two-tailed critical value from Student's t -tables on 12 df. Thus treatments 1 and 2 and treatments 1 and 5 are significantly different at the 5% level, whereas treatments 1 and 3 are not significantly different at the 5% level etc..

Example of one-way anova (same type as Example D)

8. This example illustrates a stage in the calculation of the COYU criterion. The data are the variety-by-year adjusted $\log(\text{SD}+1)$ of the "days to ear emergence" characteristic for 11 reference varieties of perennial ryegrass in three years. The data have been adjusted for any relationships between $\log(\text{SD}+1)$ and mean values for the characteristic. The data are as follows:-

Variety	Year 1	Year 2	Year 3
R1	2.36	2.13	2.30
R2	2.32	2.00	2.00
R3	2.42	2.10	1.95
R4	2.43	1.96	2.06
R5	2.52	2.14	1.96
R6	2.36	1.84	2.16
R7	2.43	2.19	1.80
R8	2.44	1.70	1.91
R9	2.52	2.16	2.24
R10	2.33	2.23	2.09
R11	2.28	1.78	1.96

9. One-way anova of the data produces the following table:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Year	2	1.011	0.5053	25.06
Residual	30	0.605	0.0202	
Total	32	1.616		

10. From F-tables the 5%, 1% and 0.1% critical values on 2 and 30 df are 3.316, 5.390 and 8.773 respectively. Comparison of the Year F-ratio with these shows that there is a very highly significant year effect at the 0.1% level ($P < 0.001$) on uniformity. However, this F-test is of minor importance in calculating the COYU criterion. Of real importance is the overall mean adjusted $\log(\text{SD}+1)$ for all the reference varieties, and the residual mean square. The residual mean square provides an estimate of the random variability of the 33 values in the above data table, i.e. the variation between reference varieties within years. It may be used to estimate the variance or standard errors of means calculated from the data. In particular, it allows the overall mean of the reference varieties to be compared with a candidate variety's mean adjusted $\log(\text{SD}+1)$.

11. The overall mean adjusted log(SD+1) is 2.154. It is the mean of the reference varieties' 33 data values. Its standard error, $SE(\bar{x})$, is estimated by

$$SE(\bar{x}) = \sqrt{\frac{RMS}{r}} = \sqrt{\frac{0.0202}{33}} = 0.0247$$

12. If \bar{x}_1 represents the overall mean adjusted log(SD+1) and \bar{x}_2 a candidate variety's mean adjusted log(SD+1) which is the mean of the equivalent of 3 data values, then the standard error of the difference in the two means, $(SE(\bar{x}_1 - \bar{x}_2))$, is estimated by

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{RMS\left(\frac{1}{r_1} + \frac{1}{r_2}\right)} = \sqrt{0.0202\left(\frac{1}{33} + \frac{1}{3}\right)} = 0.0857$$

13. The significance of the difference between \bar{x}_1 , the overall mean adjusted log(SD+1) and \bar{x}_2 , a candidate variety's mean adjusted log(SD+1) can be tested by comparing

$$|\bar{x}_1 - \bar{x}_2| = \bar{x}_2 - \bar{x}_1 \text{ with } t \times SE(\bar{x}_1 - \bar{x}_2) = 3.118 \times 0.0857 = 0.2672$$

Where t is the 0.2% one-tailed critical value (one-tailed because the candidate's mean is only rejected if it is larger than the overall mean) from Student's t-tables on 30 df. This is equivalent to comparing

$$\bar{x}_2 \text{ with } 0.2672 + \bar{x}_1 = 2.4212$$

14. Thus if a candidate variety has a mean adjusted log(SD+1) greater than or equal to 2.42*, this mean is significantly greater than the mean of the reference varieties at the 0.2% level, and hence the candidate variety is considered to be significantly non-uniform compared to the reference varieties. If the candidate variety's mean adjusted log(SD+1) is less than 2.42*, the mean will be considered to be not significantly different from the mean of the reference varieties at the 0.2% level, and hence the candidate variety not significantly different from the reference varieties in uniformity.

* The above is based on the conventional use of the two sample t-test. However, in order to give the benefit of the doubt to the breeder, in the practical application of COYU the accepted procedure is to reject a variety as non-uniform if its mean adjusted log(SD+1) is *greater than* the uniformity criterion, and accept it as uniform if its mean adjusted log(SD+1) is *less than or equal to* the uniformity criterion.

15. For more information on the COYU criterion see TGP/10.3.1.

APPENDIX A3

Example of a paired t-test (same type as Example E)

1. A special test was set up in order to test distinctness between two varieties of perennial ryegrass using the additional characteristic “Seedling width of vegetative leaf”. Thirty plants of each variety were sown at each independent sowing in a glasshouse, and when the plants were established, the “Seedling width of vegetative leaf” was observed in mm on each plant, and the mean for each variety calculated. There were 6 independent sowings and the data below are the 12 means: one for each variety for each sowing.

Variety	Sowing					
	I	II	III	IV	V	VI
J	6.9	7.9	5.2	5.8	6.5	4.8
N	5.9	6.7	3.8	5.3	3.6	3.6

2. The data will be analysed using each of the described methods.

The paired t-test analysis using a one-sample t-test of differences

3. Subtracting the variety N data value from the variety J data value for each sowing gives the following differences:-

Difference (d_i)	Sowing					
	I	II	III	IV	V	VI
	1.0	1.2	1.4	0.5	2.9	1.2

4. The mean difference \bar{d} , the variance of the differences s_d^2 , and the standard error of the mean difference $SE(\bar{d})$ are calculated from the differences as follows:-

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i = \frac{1}{6} \sum_{i=1}^6 d_i = \frac{1}{6} (1.0 + 1.2 + 1.4 + \dots + 1.2) = \frac{1}{6} \times 8.2 = 1.37$$

$$s_d^2 = \frac{1}{n-1} \left[\sum_{i=1}^n d_i^2 - \frac{\left(\sum_{i=1}^n d_i \right)^2}{n} \right] = \frac{1}{6-1} \left[\sum_{i=1}^6 d_i^2 - \frac{\left(\sum_{i=1}^6 d_i \right)^2}{6} \right]$$

$$= \frac{1}{5} \left[\sum_{i=1}^6 1.0^2 + 1.2^2 + 1.4^2 + \dots + 1.2^2 - \frac{(1.0 + 1.2 + 1.4 + \dots + 1.2)^2}{6} \right]$$

$$= \frac{1}{5} [14.5 - (8.2)^2/6] = 0.6587$$

$$SE(\bar{d}) = \sqrt{\frac{s_d^2}{n}} = \sqrt{\frac{0.6587}{6}} = 0.3313$$

5. These are used to calculate the one sample t-statistic as follows:-

$$t = \frac{\bar{d}}{SE(\bar{d})} = \frac{1.37}{0.3313} = 4.12$$

6. From Student's t-tables the 5%, 1% and 0.1% two-tailed critical values on 5 df are 2.571, 4.032 and 6.869 respectively. Comparison of the t-statistic with these shows that the mean difference is highly significantly different from zero at the 1% level ($P < 0.01$), i.e. Variety has a highly significant effect. From the sign of the mean difference it is apparent that variety J has a greater "Seedling width of vegetative leaf" than variety N.

The paired t-test analysis using two way anova

7. Two-way anova of the data produces the following table:-

Source of variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio
Sowing	5	13.8900	2.7780	8.44
Variety	1	5.6033	5.6033	17.01
Residual	5	1.6467	0.3293	
Total	11	21.1400		

8. From F-tables the 5%, 1% and 0.1% critical values on 1 and 5 df are 6.608, 16.26 and 47.18 respectively. Comparison of the Variety F-ratio with these shows that there is a highly significant Variety effect at the 1% level ($P < 0.01$). The residual mean square or variety-by-sowing mean square is an estimate of the random variability of the 12 values in the above data table. It may be used to estimate the variance or standard errors of means calculated from the data.

9. The variety means are the means of 6 data values and are as follows:-

Variety	Mean
J	6.18
N	4.82

10. Their standard error, $SE(\bar{x})$, is estimated by

$$SE(\bar{x}) = \sqrt{\frac{RMS}{r}} = \sqrt{\frac{0.3293}{6}} = 0.2343$$

11. The standard error of the difference in two means, $(SE(\bar{x}_1 - \bar{x}_2))$, is estimated by

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{RMS \left(\frac{1}{r_1} + \frac{1}{r_2} \right)} = \sqrt{0.3293 \left(\frac{1}{6} + \frac{1}{6} \right)} = 0.3313$$

12. The significance of the difference between the two variety means can be tested by calculating the two sample t-statistic:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)} = \frac{6.18 - 4.82}{0.3313} = 4.12$$

From Student's t-tables on 5 df, the 5%, 1% and 0.1% two-tailed critical values are 2.571, 4.032 and 6.869 respectively. Comparison of the t-statistic with these shows that the difference in the variety means is highly significant at the 1% level ($P < 0.01$), i.e. variety J has a greater 'Seedling width of vegetative leaf' than variety N.

[End of document]