E

TWC/16/4
ORIGINAL: English
DATE: May 6, 1998

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

# TECHNICAL WORKING PARTY ON AUTOMATION AND COMPUTER PROGRAMS

## Sixteenth Session
## Melle, Belgium, June 16 to 19, 1998

SOME REMARKS ON THE COMBINED OVER YEARS DISTINCTNESS CRITERION

*Document prepared by experts from Germany*

# Some remarks on the combined over years distinctness criterion

## Summary

The combined over years distinctness (COYD) criterion (TWC/14/7) is based on a mixed model, which implies that the variance-covariance structure has the so-called compound symmetry (CS) form. The LSD computed for COYD produces a valid test only when the CS assumption is met. A measure for the departure from the CS assumption is suggested and computed for a *Lolium perenne* data set of the Bundessortenamt. The results indicate appreciable departure from CS. It is shown, that under departure from CS, an increased sampling variation is to be expected among variety x year mean squares for different groups of years. This may partly explain observations reported in past TWC documents. It is concluded that the validity of the LSD procedure suggested for COYD may be hampered by departure from the CS assumption. It therefore appears necessary to check the validity of CS assumption over a number of crops and traits.

The COYD procedure requires balanced data sets. If the CS assumption is tenable for a three-year data set, balanced subsets of two or three years may be formed and analysed by COYD for simplicity. If the balanced subsets suffer from a lack of residual degrees of freedom (< 20), the unbalanced three-year data may be analysed by the method of fitting constants (FITCON). This seems preferable to long-term LSD (TWC/14/19), since it cannot be ruled out that time trends in variability and changes in the composition of varieties invalidate the CS assumption.

TWC is encouraged to search for powerful and simple alternatives to COYD for data that violate the CS assumption. Under departure from CS, the paired t-test may be used for assessing distinctness. Unfortunately, this test has low power due to the lack of degrees of freedom. To compensate for the loss of power, the significance level could be reduced from 1% to 5%. Appropriate choice of significance level requires further study and should involve power considerations.

## 1. Background

Document TWC/14/7 describes the Combined Over Years Distinctness (COYD) criterion for deciding whether two varieties are distinct. The criterion is based on a two-way analysis of variance (ANOVA) with factors variety and year. Two varieties are considered distinct, when their difference is statistically significant at the 1% level by an LSD (least significant difference) test. The error term for computing the LSD is the variety x year mean square (MS). When the analysis is based on variety x year means, as will be assumed subsequently, this is equivalent to the residual MS. The method requires a balanced data set, i.e. a complete two-way table of variety x year means. Many varieties are tested for only two years, while others are tested for three years. In order to compute the COYD criterion for varieties subjected to three years of testing, data from varieties tested for only two years may be discarded to obtain a complete two-way table. Similarly, a balanced two-way table for varieties under test for only two years can be obtained by dropping data on varieties tested for more than two years.

A problem with this approach is a loss of degrees of freedom (d.f.) for the variety x year mean square. The d.f. can become so small that the LSD is excessively large. The reason for this is that the variety x year interaction variance cannot be estimated with good accuracy. Therefore, TWC has agreed to use the long-term LSD, when there are less than 20 d.f. (TWC/14/19). The long-term LSD is based on the variety x year mean square from (possibly unbalanced) data accumulated over several years. A potential problem with long-term LSD is that the set of varieties which are to be compared is not identical to the set of varieties used for computing the LSD. This may yield an inappropriate LSD when the variances of varieties differ between both sets. Thus, it may be preferable to use unbalanced three-year data for computing the LSD. In this document we critically assess the merits and demerits of these different approaches.

## 2. Some problems

Long-term LSD has been the subject of several TWC documents, and some problems concerning the use of long-term LSD raised in these documents are listed below:
(a) Variation between varieties close together can be less than variation between varieties further apart (TWC/VIII/10).
(b) Variation can be related to level of expression of a character (TWC/VIII/10).
(c) LSD in groups of 2-3 years may vary markedly between groups of years (TWC/12/4) and it may be very different from long-term LSD (TWC/VIII/10).

These problems have led to uncertainty as to which method of computing LSD is preferable. We will give a recommendation at the end of this document. Before doing so, it will be argued in the subsequent sections that all of the possible approaches to computing LSD (two years, balanced; three years, balanced; long-term, unbalanced; three years, unbalanced) are based on the same underlying statistical model. It is pointed out that this model is quite restrictive and the evidence reported above (problems a-c) suggests that the model may not be appropriate in all circumstances, so that neither of the above-mentioned options for computing LSD is entirely satisfactory. To substantiate this view, we will now introduce a mixed modelling framework.

## 3. Mixed modelling

Computing the LSD using the variety x year interaction mean square as an error term implies a mixed model perspective, where years are considered as a random factor, while varieties are taken as fixed. COYD is based on the following model for variety means:

$$y_{ij} = \mu + u_j + v_i + e_{ij} \quad \text{, where} \qquad (1)$$

$y_{ij}$ = yield of variety $i$ ($i$ = 1, ..., $I$) in year $j$ ($j$ = 1,..., $J$)
$\mu$ = fixed overall mean
$u_j$ = random effect of year $j$, distributed with zero mean and variance $\sigma_u^2$
$v_i$ = fixed effect of variety $i$
$e_{ij}$ = random residual term, comprising variety x year interaction plus mean error distributed with zero mean and variance $\sigma_e^2$

The random effects $u_j$ and $e_{ij}$ are assumed to be independent. This is the underlying model, no matter how the LSD is computed (two years, balanced; three years, balanced; long-term, unbalanced; three years, unbalanced), as long as the interaction MS is used as an error term. The COYD procedure is designed to detect differences among variety main effects $v_i$, while variety x year interactions are regarded as “error“-effects not contributing to the distinctness of varieties. This latter assumption is worthy of some discussion (see a second paper to be presented by first author at this TWC meeting), but in this contribution we will not consider this issue. Here, it will be assumed that differences among variety main effects are, in fact, of primary interest. The main focus is on some statistical problems arising from the application of the mixed model (1) to the computation of an LSD for detecting differences among variety main effects.

When model (1) holds, it is best to use all available data in order to estimate the residual error variance $\sigma_e^2$ with good accuracy. The problems a-c mentioned in the foregoing section imply, however, that the simple mixed model (1) is not generally valid. In statistical terms, it implies that there may be some sort of variance heterogeneity between varieties and/or between years, and that the usual LSD is biased. Such deviations can be expected to be quite common, because, as will be argued subsequently, model (1) is extremely restrictive as regards the variance-covariance structure. Under model (1) the variance of an observation $y_{ij}$ is

$$\text{Var}(y_{ij}) = \sigma_u^2 + \sigma_e^2 \qquad (2)$$

The variance is independent of both years and varieties. This assumption is not in accordance with a large body of literature on yield stability, which suggests that the variance may depend on variety. Moreover, Slafer and Kernich (1996), in investigating long-term yield data of Australian yield trials with cereals (1900-1992), found year-dependent trends in variability. This may be a result of the change in the set of varieties under trial and variance differences among varieties (also see problem c). The shift in variance may also be caused by changes in management practices (fertilisation, pesticide use etc.). The occasionally observed dependence of the variance on the mean (problem b) is a special case of variance heterogeneity, which renders the simple model (1) inappropriate.

According to model (1) the covariance of two observations $y_{ij}$ and $y_{i'j}$ in the same year is

$$\text{Cov}(y_{ij}, y_{i'j}) = \sigma_u^2 \qquad (3)$$

which is also independent of both years and varieties. It is often found, however, that the covariance varies among pairs of varieties. For example, two varieties, which are genetically closely related and which therefore show a very similar response to changing environmental conditions, are expected to have a comparatively large covariance (see problem a). By contrast, two varieties with quite different response patterns will usually have a smaller covariance (Piepho, 1996).

The variance-covariance structure implied by model (1) is termed compound symmetry (CS) in the repeated measures literature (Winer et al., 1991; Hand and Crowder, 1996). Instead of working with the simple model (1), it is instructive to consider a more general model, which encompasses model (1) as a special case (Piepho, 1996). The model reads

$$y_{ij} = \mu_i + f_{ij} \tag{4}$$

where $\mu_i$ is the expected value of the $i$-th variety and $f_{ij}$ is a random deviation of $y_{ij}$ from the expected value $\mu_i$ in the $j$-th year. Under this general model, it will be assumed that

$$\mathrm{Var}(y_{ij}) = \mathrm{Var}(f_{ij}) = \sigma_{ii} \quad \text{and} \tag{5}$$

$$\mathrm{Cov}(y_{ij}, y_{i'j}) = \mathrm{Cov}(f_{ij}, f_{i'j}) = \sigma_{ii'} \tag{6}$$

Under this model, each variety has its own variance ($\sigma_{ii}$) and each pair of varieties has its own covariance ($\sigma_{ii'}$). This variance-covariance structure may be termed “unstructured“, because no specific model is imposed on the variances and covariances. The model is still restrictive because the variance-covariance structure is independent of years, but it is clearly much more flexible than model (1). Model (1) is obtained as a special case of model (4) by letting $\mu_i = \mu + v_i$ and $f_{ij} = u_j + e_{ij}$. Partitioning the random deviation as $f_{ij} = u_j + e_{ij}$ and assuming that random effects are independent with constant variance induces the CS structure. Scheffé (1959), who considers the general model (4) at some length, cautions the reader regarding the rather restrictive CS assumption. He says: „*We do not recommend* (Scheffé’s italics) that this assumption ordinarily be made in applications, where there usually exists no real symmetry corresponding to it.“ This caution probably also applies to DUS trails.

We may now look at the expectation of the variety x year mean square ($MS_{int}$) under the general model (4). For balanced data the expectation is known to be (Scheffé, 1959, § 8.1)

$$\mathrm{E}(MS_{int}) = (I-1)^{-1}\sum_i(\sigma_{ii} - \bar{\sigma}_{\bullet\bullet}) \quad \text{where} \tag{7}$$

$$\bar{\sigma}_{\bullet\bullet} = I^{-2}\sum_i\sum_{i'}\sigma_{ii'} \tag{8}$$

Equation (7) shows that the variety x year mean square will tend to increase as the variances $\sigma_{ii}$ of the varieties increase (problem a). It also shows that an increase in the covariances $\sigma_{ii'}$ will reduce the variety x year mean square. This may explain, why, e.g., analyses within maturity groups of perennial ryegrass can produce smaller variety x year variance components than analyses comprising several maturity groups (TWC/VIII/10; compare problem a). Within a maturity group, the average of all covariances are probably larger than in a bulk of several groups.

It is stressed that under the general model (4) the usual analysis-of-variance F-test is not valid, nor is the LSD procedure. This is so because the variance of a difference depends on the pair of varieties considered. A valid comparison of varieties under the general model is possible by conducting paired t-tests separately for each pair. Unfortunately, this approach aggravates the problem of lack of degrees-of-freedom (d.f.), since there are only $J-1$ d.f. for such a paired $t$-test. For example, at a nominal Type I error of 1%, the critical t-value for 1 d.f. (two years of paired data) is 63.657. This is 22.5 times the critical t-value for 20 d.f. (2.845). Thus, for two years of data, the critical difference for a paired t-test is expected to be 22.5 times as large as the COYD-LSD based on 20 residual d.f., assuming the CS assumption is valid. Consequently, the paired t-test is not a useful approach for two-year data. By comparison, the critical t-value for 2 d.f. (three years of paired data) is 9.925, i.e. 3,5 times the value for 20 d.f., which seems more acceptable, but still involves a considerable loss of power. To

compensate the loss of power, one might consider reducing the significance level from 1% to 5%. The critical t-value at the 5% level of significance for 2 d.f. is 4.303, which is only 1.5 times the 1% value for 20 d.f. A more thorough analysis for the choice of significance level should involve joint consideration of Type I error, power and minimum difference to be detected.
An alternative to the paired t-test is to find a subset of varieties, for which both the variances and the covariances are nearly homogeneous, i.e. for which the CS assumption is not too severely violated. Finding such subsets may be difficult in practice. The need to identify such subsets does not usually make this approach a very practical solution for routine use. Sometimes, certain covariates such as maturity group may be used to form appropriate groups.

The foregoing discussion has shown that if variety comparisons are to be based on LSD with $MS_{int}$ as an error term, it is necessary to check the validity of the implicit CS assumption. It would seem desirable to investigate larger data sets (many years) for various crops and traits to check in which cases the CS assumption is tenable. The next section proposes a measure that can be used for this purpose.

## 4. A measure of departure from compound symmetry

Box (1954) suggested a measure $\varepsilon$ of the departure from the CS assumption. More generally, $\varepsilon$ is a measure of departure from the so-called Huynh-Feldt (1976) condition, which in the present case implies that the variance of a difference of two varieties is the same for all pairs of varieties. $\varepsilon$ is given by

$$\varepsilon = \frac{I^2(\bar{\sigma}_{ii} - \bar{\sigma}_{\bullet\bullet})^2}{(I-1)\left\{\sum_i \sum_{i'} \sigma_{ii'}^2 - 2I\sum_i \bar{\sigma}_{i\bullet}^2 + I^2\bar{\sigma}_{\bullet\bullet}^2\right\}} \tag{9}$$

where $\bar{\sigma}_{i\bullet} = I^{-1}\sum_{i'}\sigma_{ii'}$. It can be shown that $(I-1)^{-1} \le \varepsilon \le 1$. Under CS we have $\varepsilon = 1$. The larger the deviation from CS, the smaller becomes $\varepsilon$. One may estimate $\varepsilon$ by simply plugging in the sample variances and covariances, as suggested by Geisser and Greenhouse (1958). It has been shown, however, that for $\varepsilon > 0.75$ and $J < 2I$ this estimate may be seriously biased. Huynh and Feldt (1976) therefore suggested a less biased estimator, which is preferable particularly when the true $\varepsilon$ is close to 1.

Table 1: Geisser-Greenhouse (1958) (G-G) estimate of $\varepsilon$ and Huynh-Feldt (1976) (H-F) estimate of $\varepsilon$ for *Lolium perenne* L. data set of Bundessortenamt.

| | | | Estimate of $\varepsilon$ | |
|---|---|---|---|---|
| Trait | Code | $I^{\S}$ | G-G | H-F |
| Flag leaf length | M11 | 112 | 0.0219 | 0.0380 |
| Flag leaf width | M12 | 112 | 0.0426 | 0.2199 |
| Length of longest stem | M21 | 111 | 0.0388 | 0.1461 |
| Length of inflorescence | M22 | 111 | 0.0350 | 0.1051 |
| Time of inflorescence emergence | M31 | 113 | 0.0222 | 0.0393 |
| Natural plant height at inflorescence emergence | M41 | 113 | 0.0347 | 0.1061 |

| | | | | |
|---|---|---|---|---|
| Plant growth habit at inflorescence emergence | M42 | 113 | 0.0379 | 0.1414 |
| Plant length in autumn of year of sowing | M51 | 115 | 0.0422 | 0.2341 |
| Plant growth habit in autumn of year of sowing | M52 | 115 | 0.0310 | 0.0807 |

([§]$I$ = number of varieties)

We estimated $\varepsilon$ for a *Lolium perenne* L. data set of the Bundessortenamt (see Table 1). A balanced subset was used with varieties that were tested in seven consecutive years. Emphasis is given here to the bias-corrected Huynh-Feldt estimate of $\varepsilon$. The Geisser-Greenhouse estimate is reported for comparison. The results indicate that there is notable departure from the Huynh-Feldt condition. Thus, use of the COYD-LSD procedure seems problematic for this data set.

## 5. Variance-mean dependence

A plot of variety variances $[s_i^2 = (J-1)^{-1}\sum_j (y_{ij} - \bar{y}_{i\bullet})^2]$ vs. variety means $(\bar{y}_{i\bullet})$ revealed a variance-mean dependence for traits M11, M31 and M51 (problem b) (see Figs. 1, 3 and 5). After a log-transform, this dependence apparently disappeared for traits M11 and M51 (see Figs. 2 and 4), but the estimates of $\varepsilon$ remained rather unaffected (M11: G-G-$\varepsilon$ = 0.0204; H-F-$\varepsilon$ = 0.0334; M51: G-G-$\varepsilon$ = 0.0446; H-F-$\varepsilon$ = 0.3203). Heading date (M31) is an exceptional case, because the variance decreases with the mean. We have not found a simple transformation that removes variance heterogeneity. This trait may be appropriately analysed by the modified joint regression analysis proposed in TWC/14/7.

## 6. Pairs of years

In the past it has been observed that LSD may vary notably among different groups of years, the variation being larger than one would usually expect by pure sampling variation (problem c). It will be shown in the next section that under the CS assumption, sampling variation is, in fact, expected to be relatively small, but that sampling variation of LSD (via $MS_{int}$) can be substantially increased, when the CS assumption is violated. Thus, a large variation among groups of years may be seen as evidence of departure from the CS assumption.

In this section we present some further empirical evidence of large variation among groups of years. For the *Lolium* data set (years 1990-1996), we estimated $MS_{int}$ for balanced subsets comprising only two years of data. The subsets contained only those varieties, for which the trait under consideration was observed in all seven years. Thus, variation among pairs of years cannot be caused by variation in the composition of varieties. The results are given in Table 2.

The variation of $MS_{int}$ among pairs of years is considerable. Note that two pairs of years are not independent when the same year is present in both pairs. Thus, the coefficient of variation (CV) of $MS_{int}$ reported in Table 2 ($CV^*$) is likely to be a biased estimate of the CV for independent pairs of years. For this reason, the distribution of $CV^*$ under the CS assumption was simulated. Data sets of the same dimensions as the balanced subsets of the *Lolium* data used in Table 2 (seven years, 111 to 115 varieties) were generated according to model (1). The effects $\mu$, $u_j$ and $v_i$ were set equal to zero. Note that $MS_{int}$ is free of these effects. The residual $e_{ij}$ was generated from a standard normal distribution. The CV of $MS_{int}$ is independent of the variance of the random normal deviates $e_{ij}$. 10,000 data sets were simulated. For each data set, $MS_{int}$ was computed for all pairs of years. Then the CV of all pairwise $MS_{int}$ values of a simulated data set was computed ($CV^{sim}$). The distribution of the 10,000 $CV^{sim}$ values constitutes the simulated distribution of $CV^*$ under CS. The 95% percentile and the mean (expected value) of this distribution are reported in Table 2. A $CV^*$ value markedly above the 95% percentile indicates departure from the assumption of CS (and normality). For all traits, $CV^*$ is clearly larger than the 95% percentile, so the *CS* assumption is doubtful. It is noted that there is a formal test for the CS assumption (see, e.g., Winer et al., 1991, and Hand and Crowder, 1996), but this is not applicable unless $J > I$, which is rarely the case in variety testing. Thus, one has to rely on unconventional procedures as those suggested here for checking the CS assumption.

Table 2: $MS_{int}$ for pairs of years using the *Lolium* data set and estimated coefficient of variation ($CV^*$) of $MS_{int}$ values across pairs. All 21 possible pairs out of seven years were formed. For a trait, the same subset of varieties was used in each pair.

| | Trait | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Year pair | M11 x $10^{-4}$ | M12 x $10^{-1}$ | M21 x $10^{-3}$ | M22 x $10^{-2}$ | M31 x $10^{-2}$ | M41 x $10^{-3}$ | M42 x $10^{-1}$ | M51 x $10^{-2}$ | M52 |
| 1-2 | 248 | 148 | 136 | 138 | 234 | 95 | 335 | 344 | 582 |
| 1-3 | 223 | 134 | 245 | 174 | 598 | 184 | 289 | 365 | 316 |
| 1-4 | 512 | 127 | 105 | 479 | 201 | 108 | 168 | 454 | 423 |
| 1-5 | 151 | 118 | 111 | 364 | 264 | 87 | 126 | 458 | 405 |
| 1-6 | 251 | 129 | 112 | 365 | 192 | 163 | 210 | 455 | 379 |
| 1-7 | 741 | 175 | 113 | 349 | 573 | 213 | 391 | 474 | 595 |
| 2-3 | 163 | 74 | 217 | 161 | 221 | 152 | 314 | 396 | 879 |
| 2-4 | 244 | 99 | 129 | 458 | 101 | 99 | 352 | 512 | 989 |
| 2-5 | 88 | 66 | 183 | 289 | 111 | 110 | 224 | 431 | 612 |
| 2-6 | 158 | 97 | 149 | 368 | 130 | 173 | 395 | 429 | 748 |
| 2-7 | 270 | 105 | 139 | 298 | 230 | 228 | 508 | 590 | 1274 |
| 3-4 | 430 | 120 | 261 | 517 | 221 | 139 | 420 | 661 | 256 |
| 3-5 | 155 | 87 | 131 | 392 | 164 | 193 | 228 | 541 | 468 |
| 3-6 | 249 | 116 | 142 | 428 | 377 | 318 | 320 | 447 | 320 |
| 3-7 | 479 | 135 | 127 | 406 | 125 | 409 | 604 | 563 | 384 |
| 4-5 | 205 | 63 | 149 | 340 | 59 | 106 | 139 | 644 | 503 |
| 4-6 | 103 | 56 | 144 | 332 | 138 | 170 | 211 | 517 | 311 |
| 4-7 | 171 | 117 | 115 | 310 | 220 | 227 | 387 | 631 | 367 |
| 5-6 | 86 | 63 | 107 | 219 | 171 | 132 | 141 | 627 | 321 |
| 5-7 | 337 | 66 | 83 | 190 | 192 | 226 | 336 | 954 | 610 |
| 6-7 | 272 | 98 | 95 | 240 | 372 | 370 | 460 | 416 | 480 |
| d.f. | 111 | 111 | 110 | 110 | 112 | 112 | 112 | 114 | 114 |
| $CV^*$ | 61.1 | 30.8 | 33.1 | 33.0 | 60.4 | 47.8 | 41.1 | 26.3 | 48.0 |
| 95% percentile of $CV^{sim}$ (#) | 16.4 | 16.4 | 16.4 | 16.4 | 16.3 | 16.3 | 16.1 | 16.1 | 16.1 |
| Mean of $CV^{sim}$(#) | 12.3 | 12.3 | 12.4 | 12.4 | 12.3 | 12.3 | 12.3 | 12.1 | 12.1 |

(#): Simulated under the assumptions of CS and normality ($n$ = 10,000). A value of $CV^*$ exceeding simulated 95% percentile of $CV^{sim}$ indicates departure from CS. For details see text.

## 7. The variance of $MS_{int}$

In follows from results in Box (1954) that the variance of $MS_{int}$ is

$$\mathrm{Var}(MS_{int}) = \frac{2\left[\mathrm{E}(MS_{int})\right]^2}{(I-1)(J-1)\varepsilon} \tag{10}$$

The coefficient of variation of $MS_{int}$ is

$$\mathrm{CV}(MS_{int}) = \sqrt{\frac{\mathrm{Var}(MS_{int})}{\left[\mathrm{E}(MS_{int})\right]^2}} = \sqrt{\frac{2}{(I-1)(J-1)\varepsilon}} \tag{11}$$

Under CS ($\varepsilon$ = 1) we can therefore expect a low coefficient of variation of $MS_{int}$, while for larger departures from CS ($\varepsilon$ < 1) the uncertainty of $MS_{int}$ can be substantial. For example, when $J$ = 2 and $I$ = 111, we have CV($MS_{int}$) = 0.13 = 13%, when the CS assumption is met. By contrast, when $\varepsilon$ = 0.1, one finds CV($MS_{int}$) = 0.42 = 42%. Thus, under departure from CS, a much larger variability of $MS_{int}$ among pairs of years can occur than when the CS assumption is met. For the *Lolium* data the H-F estimate of $\varepsilon$ is much smaller than 1 for all traits (Table 1). This may explain why the variation among $MS_{int}$ between different pairs of years (Table 2) is larger than expected under CS. It should be noted, however, that the CV of $MS_{int}$ is also influenced by non-normality. Specifically, the CV will increase with the kurtosis (standardised forth moment) (Searle et al., 1992, p. 412).

## 8. Three-year data versus long-term data

Table 3: $MS_{int}$ for triplets of consecutive years and for complete data set (seven years). Due to unbalancedness of the complete data set, the subset of varieties may differ among triplets.

| | Trait | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Years | M11 x $10^{-4}$ | M12 x $10^{-1}$ | M21 x $10^{-3}$ | M22 x $10^{-2}$ | M31 x $10^{-2}$ | M41 x $10^{-3}$ | M42 x $10^{-1}$ | M51 x $10^{-2}$ | M52 |
| 1-7 | 257 | 101 | 138 | 306 | 207 | 192 | 305 | 590 | 478 |
| 1-3 | 209 | 118 | 201 | 163 | 358 | 142 | 314 | 361 | 587 |
| 2-4 | 288 | 100 | 203 | 382 | 188 | 126 | 352 | 506 | 636 |
| 3-5 | 256 | 88 | 180 | 428 | 143 | 136 | 240 | 696 | 375 |
| 4-6 | 131 | 62 | 141 | 299 | 119 | 135 | 165 | 685 | 346 |
| 5-7 | 237 | 80 | 99 | 217 | 198 | 230 | 304 | 755 | 446 |
| Average of triplets | 224 | 89 | 165 | 298 | 201 | 154 | 275 | 601 | 478 |

Table 4: Huynh-Feldt estimates of $\varepsilon$ for triplets of consecutive years and for complete data set (seven years). Due to unbalancedness of the complete data set, the subset of varieties may differ among triplets. For each subset, only those varieties tested in all years in th subset, were included.

| | Trait | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Years | M11 | M12 | M21 | M22 | M31 | M41 | M42 | M51 | M52 |
| 1-7 | 0.038 | 0.220 | 0.146 | 0.105 | 0.039 | 0.106 | 0.141 | 0.234 | 0.081 |
| 1-3 | 0.308 | 0.115 | 0.160 | 1.000 | 0.027 | 0.127 | 1.000 | 1.000 | 0.049 |
| 2-4 | 0.047 | 0.232 | 0.111 | 0.049 | 0.092 | 0.304 | 0.732 | 0.194 | 0.034 |
| 3-5 | 0.030 | 0.119 | 0.115 | 0.539 | 0.031 | 0.138 | 0.033 | 0.306 | 0.109 |
| 4-6 | 0.032 | 1.000 | 0.592 | 0.215 | 0.046 | 0.108 | 0.196 | 0.119 | 0.132 |
| 5-7 | 0.021 | 0.077 | 0.248 | 1.000 | 0.067 | 0.040 | 0.034 | 0.025 | 0.062 |
| Average of triplets | 0.082 | 0.308 | 0.245 | 0.560 | 0.053 | 0.143 | 0.399 | 0.329 | 0.077 |

We investigated the *Lolium* data with respect to the question, whether three-year data are preferable to long-term data (seven years in this example). From the seven years, all five possible triples of three consecutive years were formed. For each triplet, both $MS_{int}$ and $\varepsilon$ were computed and averaged across triplets. These results were compared to $MS_{int}$ and $\varepsilon$ computed for the long-term data. Both $MS_{int}$ (Table 3) and estimates of $\varepsilon$ (Table 4) vary markedly among triplets. With many traits, the averages of $MS_{int}$ across triplets are somewhat smaller than $MS_{int}$ for the complete data. Similarly, the average of estimates of $\varepsilon$ across triplets are larger than for the complete data set in all traits except M52. This is weak indication that analysis based on three years of (unbalanced) data may be preferable to analysis based on long-term data.

## 9. Alternative methods of analysis for unbalanced data under compound symmetry

In the foregoing sections, it has been shown that the CS assumption implicit in the mixed model used for COYD may be violated, thus invalidating the LSD procedure. If it has been established that for a given data set (possibly a subset of a larger data set) the CS assumption is met, an analysis based on model (1) is appropriate. It then needs to be decided, how the data should be analysed. For example, the question arises whether a balanced subset should be formed or not. A balanced subset simplifies the analysis, but by forming a subset, valuable information on the variance components is discarded. In order to gain maximum accuracy of variance component estimates, it is desirable to use all available information. This will usually require the analysis of an unbalanced data set. Here, we wish to point out two options for unbalanced data.

One option is the method of fitting constants (FITCON; see TWC/12/4). In this analysis, years are formally treated as a fixed, despite the fact that the year factor is random. Variety means

adjusted for year effects are estimated by the method of least squares. An estimate of the variance component $\sigma_e^2$ is computed from the residual mean square after fitting main effects for varieties and years. This mean square is then used to compute standard errors of variety means and differences of variety means. Due to the unbalancedness of the data, no single LSD can be computed. This type of analysis is analogous to an analysis of an incomplete block design without recovery of inter-block information (Cochran and Cox, 1957), with years corresponding to incomplete blocks. Thus, FITCON does not allow recovery of "inter-year" information.
Alternatively, the year main effect may be regarded as random. This allows the recovery of inter-year information. The recovery of inter-year information is based on a weighted least squares analysis, where the weights depend on the variances of both year main effects $u_j$ and residuals $e_{ij}$. The variance components can be estimated by standard procedures, e.g. by the REML (Restricted Maximum Likelihood) method, which is readily available with common statistical packages (GENSTAT, SAS). When the variance components are known, the weighted least squares estimator (WLSE) is best linear unbiased, but in practice estimates of variance components have to be used. An analysis based on estimated weights is no longer optimal, and it is not clear whether such an analysis is superior to the FITCON procedure. The most critical issue seems to be the small d.f. for estimating the year variance component $\sigma_u^2$, which will certainly make an adjustment necessary to the usual Wald tests based on WLSE with estimated weights (Kenward and Roger, 1997).

Simulations and scrutiny of some data sets indicate that the results obtained by both of these methods differ only marginally. This observation can be explained by the fact that inter-year information is usually small, because the year variance component $\sigma_u^2$ tends to be relatively large. It is suggested that the FITCON procedure may be used in practice for its simplicity. It is stressed again, however, that this method also assumes CS.

## 10. Conclusion

- The COYD criterion of distinctness is based on a mixed model, which assumes that the compound symmetry (CS) assumption holds. This assumption may be violated in some cases.
- If the CS assumption is tenable, it would seem best to use as much data as possible, i.e. long-term data, to estimate variance components, provided the CS assumption can be extended to long-term data. This ensures maximum attainable precision for the variance estimates. Analysis may be done by FITCON. However, since it cannot be ruled out that long-term time trends of variability (which are difficult to detect) and the effect of a changing composition of the set of varieties under test invalidate the CS assumption for long-term data, it seems safer to generally use only three-year data. Inspection of several traits in a *Lolium* data set indicates that three-year data tend to have a somewhat smaller residual mean square ($MS_{int}$) and to better met the CS assumption required for FITCON analysis. When balanced two-year or three-year subsets can be formed with more than 20 residual d.f., such subsets may be used for computing LSDs. If balanced subsets have less than 20 residual d.f., an unbalanced three-year data set can be analysed using FITCON. It should be borne in mind that all of these procedures are based on the assumption of CS.
- The evidence of real data sets suggests that the CS assumption underlying standard mixed model analyses of variety x year data may at times be violated. Under marked departure

from CS, only the paired t-test can be recommended as a valid procedure. A problem with this approach is the extremely small number of d.f. (d.f. = $J - 1$, where $J$ is the number of years in which both varieties were tested together). Three years of paired data appears to be the minimum requirement for this test. To compensate for the loss of power, one might consider reducing the significance level from 1% to 5%. No satisfactory alternative to the paired t-test can be suggested at this stage, and it is unclear whether a simple and powerful alternative under departure from CS is forthcoming. It is hoped, however, that the theoretical results presented in this document may help someway towards finding a viable procedure.

## *References*

Box GEP 1954 Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification. Annals of Mathematical Statistics 25, 484-498.

Cochran WG, Cox GM 1957 Experimental designs. Wiley, New York.

Geisser S, Greenhouse SW 1958 An extension of Box’s results on the use of the F-distribution in multivariate analysis. Annals of Mathematical Statistics 29, 885-891.

Hand D, Crowder M 1996 Practical longitudinal data analysis. Chapman and Hall, London.

Huynh H, Feldt LS 1976 Estimation of the Box correction for degrees of freedom from sample data in randomized block and split plot designs. Journal of Educational Statistics 1, 69-82.

Kenward MG, Roger JH 1997 Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53, 983-997.

Piepho HP 1996 Comparing cultivar means in multilocation trials when the covariance structure is not circular. Heredity 76, 198-203.

SAS Institute Inc. 1997 SAS/STAT Software: Changes and enhancements through release 6.12. SAS Institute, Cary.

Scheffé H 1959 The analysis of variance. Wiley, New York.

Searle SR, Casella G, McCulloch CE 1992 Variance components. Wiley, New York.

Slafer GA, Kernich GC 1996 Have changes in yield (1900-1992) been accompanied by a decreased yield stability in Australian cereal production? Australian Journal of Agricultural Research 47, 323-334.

TWC/VIII/10 Estimation of COYD variance and long-term LSD (1990).

TWC/12/4 Evaluation of long term LSD’s (1994).

TWC/14/7 The combined over years distinction criterion (1996).

TWC/14/19 Report of Fourteenth Session, Hannover. June 4 to 6, 1996 (1996).

Winer BJ, Brown DR, Michels KM 1991 Statistical principles in experimental design. McGraw-Hill, New York.

H P Piepho
Institut für Nutzpflanzenkunde
Gh Kassel
Steinstrasse 19
37213 Witzenhausen
Germany

F Laidig
Bundessortenamt
Osterfelddamm 80
30627 Hannover
Germany

May 1998

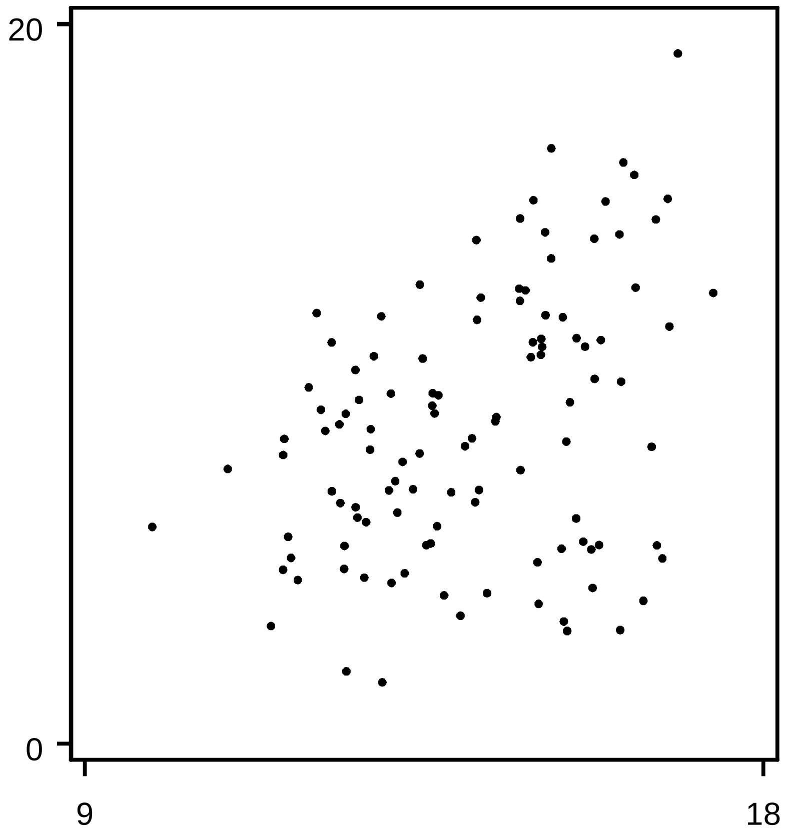


Fig. 1: Plot of variance (ordinate) vs. mean (abscissa) for trait M11. Original data (x $10^{-3}$).

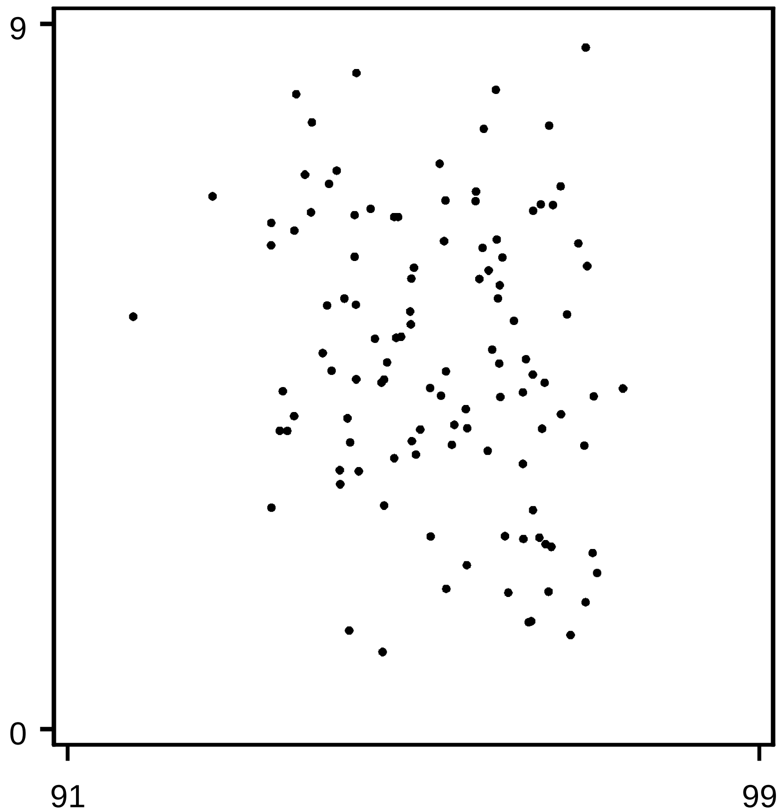


Fig. 2: Plot of variance vs. mean for trait M11. Log-transformed data (x $10^{-1}$).

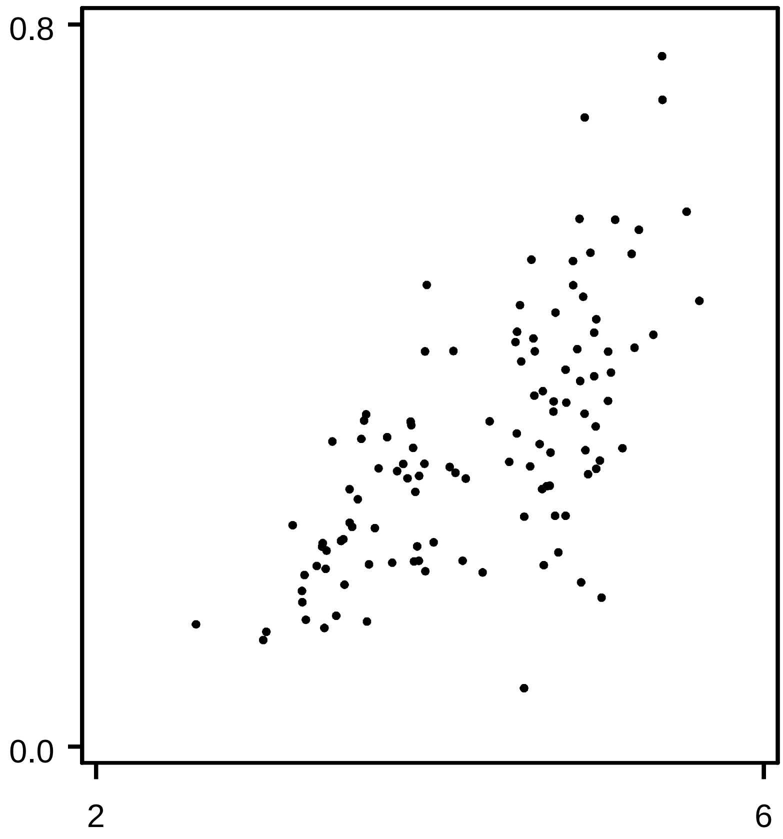


Fig. 3: Plot of variance vs. mean for trait M51. Original data (x $10^{-3}$).

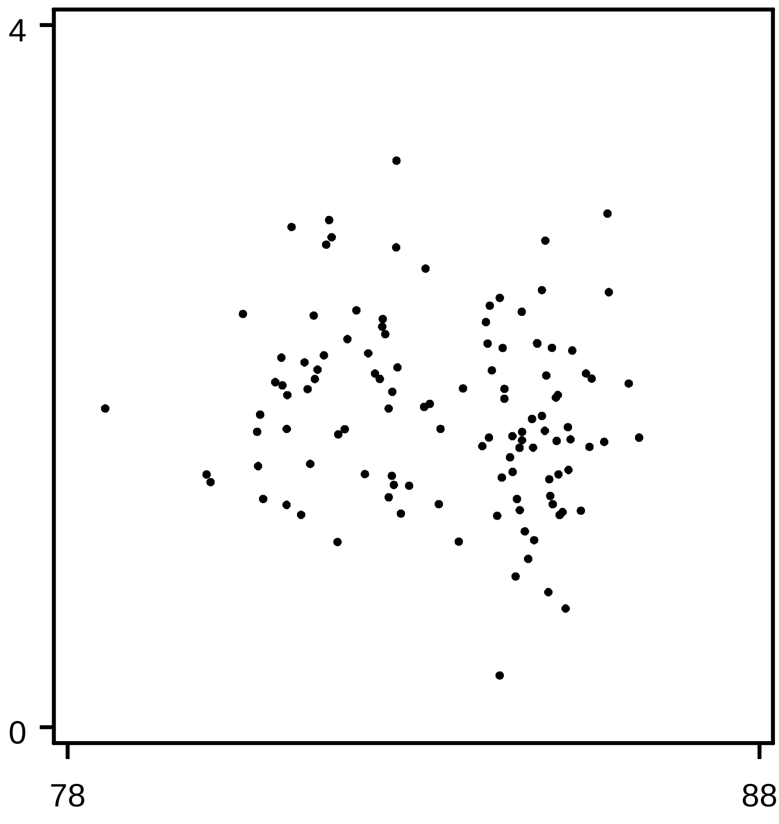


Fig. 4: Plot of variance vs. mean for trait M51. Log-transformed data (x $10^{-1}$).

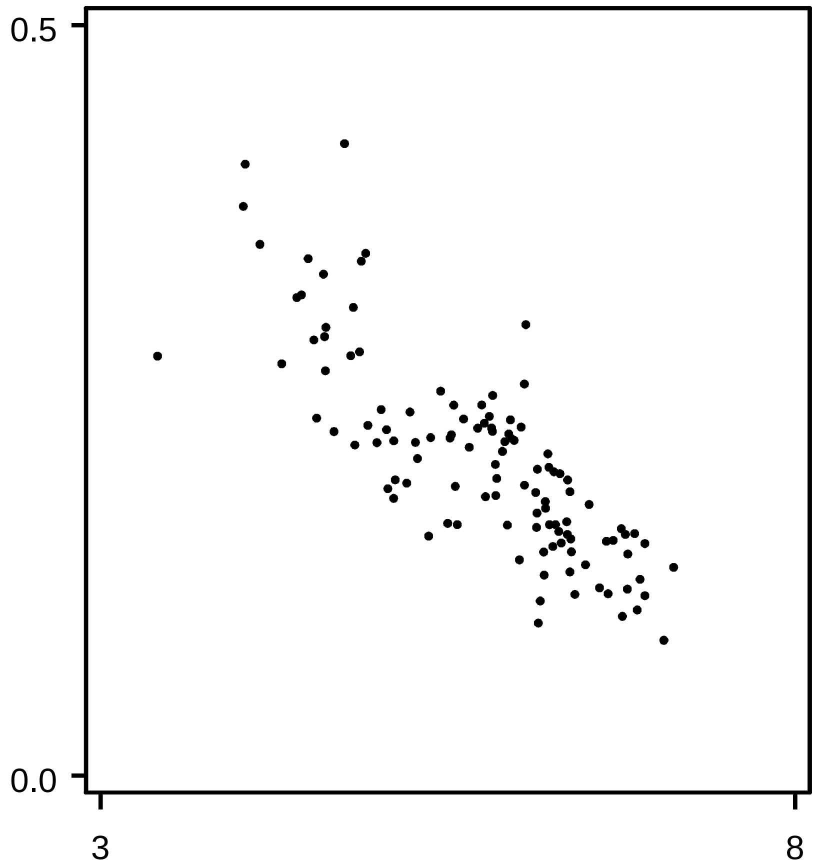


Fig. 5: Plot of variance vs. mean for trait M31. Original data (x $10^{-3}$).

[End of document]