PATTERN ANALYSIS OF GENOTYPE X ENVIRONMENT INTERACTIONS AND COMPARISONS WITH ALTERNATIVE ANALYSES

A thesis presented in partial fulfilment
of the requirements for the degree
of
Master of Science
in
Plant Science
at
Massey University

Carol Cullen

1981

Abstract

The occurrence of genotype-environment interactions is a problem affecting the interpretation of cultivar trials. Several analyses have been used to try to resolve the inconsistencies of cultivar performances which occur when these interactions are present.

An assessment of several techniques was carried out using three sets of data. Two sets of barley data came from one season's trials covering the barley growing areas of New Zealand. Ten wheat cultivars were tested in four locations in the lower North Island in two seasons.

The analyses which were examined were Analysis of variance, linear regression, Cluster Analysis and Principal Component Analysis. The parameters of Wricke, Hanson and Eberhart and Russell were also studied.

The Analysis of Variance revealed significant location, genotype and genotype-location interaction effects for the barley data. The wheat data had significant years, years x locations, genotype and genotype-year interactions effects.

There was a strong linear relationship between the genotype means and the environmental index for the Finlay-Wilkinson regression analysis. Following refinement of the error term $\beta_{i's}$ with significant differences from 1.0 could be seen for several Barley and Wheat genotypes. It was noted that a conflict existed between the aim of finding significant differences from 1.0 and the assumption of

independence of effects for the underlying model. It was suggested that an independent measure of environments be used.

The parameters of Wricke, Hanson and Eberhart & Russell were each related to different concepts of stability and the genotypes ranked accordingly. The three parameters gave reasonably consistent results for the rankings of the cultivars.

In the barley data the cultivars Goldmarker and Magnum had uniformly low rankings. The wheat cultivar Gamenya was generally found to rank highly. These were measures of variability over environments so a high ranking infers a low level of variability and vice versa.

A comparison of the different clustering strategies available was carried out and Wards Incremental Sums of Squares method was chosen as the major strategy. This was applied to each data set using both genotype-environment effects and means. A probabilistic cut off measure was used for truncation of the dendrogram.

The clusters formed could be related to the previous analyses and seemed to adequately summarise the different responses present.

A Principal Component Analysis was carried out and the number of components needed to account for 75% of the total variation were examined. For the barley data sets relatively large numbers of components were needed for this (five and six). This made interpretation and presentation of the genotypic performances difficult. For the wheat data two components

explained a satisfactory level of the total variation and the arrangement of the genotypes on these two axes agreed closely with the clustering results.

Varimax rotation did not aid greatly in the interpretation of the components.

It was felt that the roles of these different analyses were complementary in interpreting genotypic performances.

Acknowledgements

I would especially like to thank my supervisor, Dr Ian Gordon for his assistance and guidance throughout this study.

I would also like to thank:-

- The Department of Agriculture, University of Queensland, in particular D.E. Byth, for supplying the programme YLGR.

- C.B. Dyson, MAF, Ruakura for supplying the data from the Barley cultivar trials.

- J.M. McEwan and staff at CRD, DSIR, Palmerston North for providing the data from the Wheat Cultivar Trials.

- The staff at the Computer Centre, Massey University, for assistance and advice.

Finally I would like to thank my parents for financial and moral support, and my flatmates for tolerance.

CONTENTS

List of Tables

LIST OF FIGURES

Introduction

In order to assess the worth of a potential cultivar a Plant Breeder will carry out a series of trials which are designed to give an estimation of its relative performance. Ideally the trials should cover the range of environments, both sites and seasons, which will be experienced by the cultivar when grown commercially.

Decisions based on the results of these trial systems become confounded when genotype-environmental interactions are present. These interactions are seen as inconsistencies of relative genotypic performances over environments and mean that comparisons of genotype means are not a true indication of the situation.

The occurrence of genotype-environment interactions has been reported in many crops over a long period of time and has resulted in several different analyses being proposed to handle them (Hill, 1975). The adaptability of a genotype is a concept concerned with its responsiveness to environmental change. This is closely linked to genotype-environment interaction problems and some measure of adaptability has been the aim of some of these analyses.

The aim of a plant breeding program may be for specific adaptation to a particular environmental range or general adaptability to a wide range of environmental conditions. Some measure of adaptation will therefore be important.

Parameters designed for this purpose have involved analysis of variance techniques, partitioning into variance components and regression analysis. A more recent approach has been to describe the pattern of response of a genotype. This has involved the application of several multivariate analyses to trial data, using each test environment as a variable in the analysis.

This study has been designed to assess some of these methods in an attempt to identify the advantages and weaknesses of each. Of particular interest are the Pattern Analysis techniques since these are relatively untried in this field.

The data which has been used comes from regional trials for the evaluation of wheat and barley cultivars in New Zealand. Interpretation of these for the various stability and adaptability concepts has been carried out and related to the New Zealand situation.

Chapter 1 Literature Review

1.1 Introduction to the concept of genotype x environment interactions

In any plant breeding program assessments have to be made on genotypic performance. These assessments are made on phenotypic measurements reflecting both genetic and non-genetic components.

i.e. Phenotype = Genotype + environment.

Unfortunately these nongenetic effects are not independent of the genetic effects as is often seen in the changes of relative genotype rankings when measured in different environments. This genotype x environment interaction (G x E) reduces the correlation between genotype and phenotype and makes accurate assessment of genotype performance difficult (Moll & Stuber, 1974). Genotype x environment interactions therefore pose considerable problems for plant breeders and many attempts have been made to categorise and describe them.

Allard & Bradshaw (1964) note that where 10 genotypes are measured in 10 different environments, 10^{145} different interaction types may occur, making assessment very complex.

They also make the distinction between predictable environmental variation, such as seasonal or treatment differences, and non-predictable environmental variation, such as year to year fluctuations.

To counteract the predictable variation they suggest the breeding of cultivars specifically for regions or management practices. However, the non-predictable variations are harder to account for and they stress the need to test cultivars over seasons.

The selection of cultivars exhibiting stability across environmental variation would therefore be desirable. Such a cultivar can be said to be well buffered and this can be achieved in two ways (Allard & Bradshaw, 1964). Firstly the cultivar can be made up of a number of genotypes each adapted to a different range of environmental conditions. Secondly the individuals themselves may be well buffered, such as heterozygotes often are.

Recognition of the occurrence of genotype x environment interactions has been reported in the literature for most of this century (Hill, 1975).

1.2 Analysis of Variance

The technique of Analysis of Variance has been widely used to determine the presence of significant genotype x environment interactions. This analysis partitions the total phenotypic variance into components as defined by the model and its associated experimental design.

Early work was done by Federer & Sprague (1947) and Sprague & Federer (1951). They outlined the application of the Analysis of Variance to the trial situation where genotypes are tested over a series of environments.

Miller, Williams and Robinson (1959) applied it to their variety evaluation program for cotton which involved trials over seasons as well as locations.

The models used by Sprague & Federer and Miller et al. are set out as Models 1 and 2 in Table 1.1.

The two models differ in that the environmental component of model 1 is split into a year, a location and a year x location component in model 2. Where the data involved is measured over both years and locations model 2 is therefore applicable. Where only one factor of environment is involved i.e. either locations or years then model 1 is appropriate. Both are random-effect models, implying that both genotypes and environments describe a random sample from the populations of inference from which they are drawn.

The expected Mean Squares for both models are given in Table 1.2.

The assumptions underlying the Analysis of Variance as discussed by Cochran (1947) and Eisenhart (1947) may be summarised as follows:

1. That the random effects are independent.

2. Experimental errors are random, have a common variance and are normally distributed.

Cochran (1947) found that heterogeneity of error variances was an important feature of data since this may result in a loss of efficiency in significance testing and Mean Square estimation.

For the model 1 situation the significance of the genotype x environment interaction term is estimated from the F ratio of G x E Mean Square/error Mean Square using the degrees of freedom appropriate to each mean square.

The G x E variance component is estimated from the mean squares (G x E mean square - error mean square)/b

where b = the number of blocks.

From such an analysis therefore it is possible to get an estimate of the genotype x environment interaction component and its significance.

Table 1.1 Analysis of Variance Models

Model 1

$$\gamma_{ijk} = \mu + \alpha_i + \eta_\kappa + \alpha\eta_{i\kappa} + \beta_{(j)\kappa} + \varepsilon_{ij\kappa}.$$

$\gamma_{ij\kappa}$ = phenotypic value of the i^{th} genotype in the κ^{th} environment and j^{th} replicate.

μ = grand mean over genotypes, environments and blocks.

α_i = effect of the i^{th} genotype.

η_κ = effect of the κ^{th} environment.

$\alpha\eta_{i\kappa}$ = interaction effect for the i^{th} genotype and the κ^{th} environment.

$\beta_{(j)\kappa}$ = effect of the j^{th} block in the κ^{th} environment.

$\varepsilon_{ij\kappa}$ = residual deviation for the $ji\kappa^{th}$ observation.

Model 2

$$\gamma_{ij\kappa\ell} = \mu + \alpha_i + \lambda_\kappa + T_\ell + \beta_{(j)\kappa\ell} + \alpha\lambda_{i\kappa} + \alpha T_{i\ell} + \alpha\lambda T_{i\kappa\ell} + \lambda T_{\kappa\ell} + \varepsilon_{ij\kappa\ell}.$$

λ_κ = effect of the κ^{th} site.

T_ℓ = effect of the ℓ^{th} year.

$\beta_{(j)\kappa\ell}$ = effect of the j^{th} block in each $\kappa\ell^{th}$ combination.

$\alpha\lambda_{i\kappa}$ = interaction effect of i^{th} genotype and κ^{th} site.

$\alpha T_{i\ell}$ = interaction effect of i^{th} genotype and ℓ^{th} year.

$\alpha\lambda T_{i\kappa\ell}$ = second order interaction effect for i^{th} genotype κ^{th} site and ℓ^{th} year.

$\lambda T_{\kappa\ell}$ = interaction effect of κ^{th} site and ℓ^{th} year.

$\varepsilon_{ij\kappa\ell}$ = residual deviation

and other symbols previously defined above.

Table 1.2 Expectations Mean Squares

Model 1

Source of variation	Degrees of freedom	E(MS)
environment	$\varepsilon-1$	$\sigma^2 + g\sigma^2_{\beta(\varepsilon)} + b\sigma^2_{g\varepsilon} + gb\sigma^2_{\varepsilon}$
blocks	$\varepsilon(b-1)$	$\sigma^2 + g\sigma^2_{\beta(\varepsilon)}$
genotypes	$g-1$	$\sigma^2 + b\sigma^2_{g\varepsilon} + b\varepsilon\sigma^2_{g}$
genotype-environment interaction	$(g-1)(\varepsilon-1)$	$\sigma^2 + b\sigma^2_{g\varepsilon}$
error	$\varepsilon(b-1)(g-1)$	σ^2

From Gordon et al. (1972)

Model 2

Source of variation	Degrees of Freedom	E(MS)
sites	$s-1$	$\sigma^2 + b\sigma^2_{gsy} + by\sigma^2_{gs} + g\sigma^2_{(b)sy} + gb\sigma^2_{sy} + gby\sigma^2_{s}$
years	$y-1$	$\sigma^2 + b\sigma^2_{gsy} + bs\sigma^2_{gy} + g\sigma^2_{(b)sy} + gb\sigma^2_{sy} + gbs\sigma^2_{y}$
sites x years	$(s-1)(y-1)$	$\sigma^2 + b\sigma^2_{gsy} + g\sigma^2_{b(sy)} + gb\sigma^2_{sy}$
blocks	$sy(b-1)$	$\sigma^2 + g\sigma^2_{b(sy)}$
genotypes	$g-1$	$\sigma^2 + b\sigma^2_{gys} + bs\sigma^2_{gy} + by\sigma^2_{gs} + bsy\sigma^2_{g}$
genotypes x sites	$(g-1)(s-1)$	$\sigma^2 + b\sigma^2_{gys} + by\sigma^2_{gs}$
genotypes x years	$(g-1(y-1)$	$\sigma^2 + b\sigma^2_{gys} + bs\sigma^2_{gy}$
genotypes x sites x years	$(g-1)(y-1)(s-1)$	$\sigma^2 + b\sigma^2_{gys}$
error	$sy(b-1)(g-1)$	σ^2

From Le Clerg, et al. (1962)

ε = no. environments g = no. genotypes s = no. sites

y = no. years b = no. blocks.

1.3 Partitioning of the G x E Variance Component

A development from the Analysis of Variance technique has been to partition the G x E sums of squares into individual variety contributions.

Plaisted and Peterson (1959) and Plaisted (1960) achieved this by omitting varieties one at a time from the computation.

A similar parameter, the ecovalence, was defined by Wricke (1962) /as reported in Easton (1971)/ as

$$\varepsilon_i = \sum_{\kappa}(X_{i\cdot\kappa/b} - X_{i\cdot\cdot/\varepsilon b} - X_{\cdot\cdot\kappa}/gb + X_{\cdot\cdot\cdot/bg\varepsilon})^2$$

The dot notation implies summation over the omitted subscript

e.g. $X_{i.\kappa} = \sum_j X_{ij\kappa}$

Subtraction of the expectations of these means shows:-

$$\bar{X}_{i\cdot\kappa} = \mu + \alpha_i + \eta_\kappa + \alpha\eta_{i\kappa}$$

$$-\ \bar{X}_{i\cdot\cdot} = \mu + \alpha_i$$

$$-\ \bar{X}_{\cdot\cdot\kappa} = \mu + \eta_\kappa$$

$$+\ \bar{X}_{\cdot\cdot\cdot} = \mu$$

$$\alpha\eta_{i\kappa}$$

Hence ε_i is the sum of $\alpha\eta_{i\kappa}$'s over all environments. Easton (1971) examined ecovalences and found that on a regression line of varietal performance against an environment index the ecovalence measured the deviation from a line of slope 1.0 with its y-ordinate being equal to the genotype mean. He concluded that a genotype with a slope greater than 1.0 would have a high ecovalence and thus be classed as unstable although it may be a desirable genotype.

Shukla (1972) described a stability variance for each genotype defined as the variance over environments of $(\alpha_i + \varepsilon_{i\cdot\kappa})$ where $\varepsilon_{i\cdot\kappa}$ is the mean of $\varepsilon_{ij\kappa}$ over blocks. He found an unbiased estimate of this variance σ_i^2 and described an F test of σ_i^2/σ_o^2 where σ_o^2 = pooled error Mean Square to test the instability of a genotype.

The stability analysis of Tai (1971) also partitions the interaction term but he defines two parameters:-

$$\alpha_i = \frac{\text{MS env}}{\text{MS env-MS error}} \times (\beta-1) \quad , \quad \lambda i = \frac{\sigma^2_{\delta i} + \sigma^2_{\varepsilon}}{\sigma^2_{\varepsilon}}$$

Where β = regression coefficient of genotypic means regressed on an environmental index.

$\sigma^2_{\delta i}$ = variance component for deviations from regression.

σ^2_{ε} = error variance

i.e. α_i = measures the linear response of the i^{th} genotype to the environment.

λ_i measures the deviation from the linear response in terms of the magnitude of the error variance.

A genotype with $\alpha = o$ $\lambda = 1$ is defined as having average stability (Tai, 1979).

Hanson's measure of stability incorporates the contribution of the i^{th} genotype to the G x E sums of squares with its response to environmental change (Hanson, 1970).

His particular concept of stability was defined with reference to a space of ℓ dimensions (ℓ = no. environments), with a stable genotype located at the origin.

The position of the i^{th} genotype in this space can be determined by the deviations of the genotype performance from the expected. The expected performance was defined as

$$Z_{i\kappa} = \bar{X}_{...} + \bar{X}_{i..} + \alpha\bar{X}_{..\kappa}$$

$$\eta_{i\kappa} = (X_{i\kappa} - Z_{i\kappa})$$

where $\alpha = 1 + \min \beta_i$

The stability parameter is $D_i^2 = \sum_{\kappa} \eta_{i\kappa}^2$

For comparative genotypic stability, a distance measure between the i^{th} and i'^{th} genotypes is

$$D_{(ii')}^2 = \sum_{\kappa} (\eta_{i\cdot\kappa} - \eta_{i'\cdot\kappa})^2$$

1.4 Joint Regression Analysis

(i) Basic Models

A widely used technique for the identification of adaptable varieties is that known as Joint Regression Analysis.

This type of analysis was originally suggested by Yates and Cochran (1938) for the identification of genotypes which were superior over a range of environments.

Apparently unaware of this early suggestion Finlay and Wilkinson (1963) outlined a similar technique but emphasized its potential for identifying different adaptation responses in genotypes. Their basic approach is outlined as Model 3 in Table 1.3.

The environmental index upon which the variety yields are regressed was obtained by taking the mean of all varieties at each of the environments ($\bar{x}_{..k}$). They suggested that β_i was a useful stability parameter indicating a variety's response to the range of environments tested. They also examined the way in which the regression coefficient and variety mean yield varied together and constructed the diagram presented in figure 1.1.

Finlay and Wilkinson nominated an 'ideal' variety as one having general adaptability, i.e. maximum yield potential in favourable environments but with maximum stability i.e. $\beta_i = 0$, $\bar{x}_{i..}$ maximum.

Table 1.3 Regression Models

Model 3

$$\alpha\eta_{i\kappa} = \beta_i I_\kappa + \delta_{i\kappa}.$$

Rewritten in terms of Model 1, Table 1.1,

$$\gamma_{ij\kappa} = \mu + \alpha_i + (1 + \beta_i) I_\kappa + \alpha_{i\kappa} + \varepsilon_{ij\kappa}.$$

$\gamma_{ij\kappa}$ = phenotypic value of the i^{th} genotype in the κ^{th} environment and j^{th} replicate.

$\alpha\eta_{i\kappa}$ = interaction effect for the i^{th} genotype and the κ^{th} environment.

μ = grand mean over genotypes, environments and replicates.

α_i = effect of the i^{th} genotype.

β_i = regression coefficient measuring response of i^{th} genotype to all environments.

I_κ = Environmental Index.

$\delta_{i\kappa}$ = deviation from regression of i^{th} genotype of κ^{th} environment.

From Perkins and Jinks (1967).

Model 4

$$\gamma_{i\kappa} = \mu_i + \beta_i I_\kappa + \delta i\kappa$$

$\gamma_{i\kappa}$ = phenotypic value of the i^{th} genotype in the κ^{th} environment.

μ_i = mean of the i^{th} genotype over all environments.

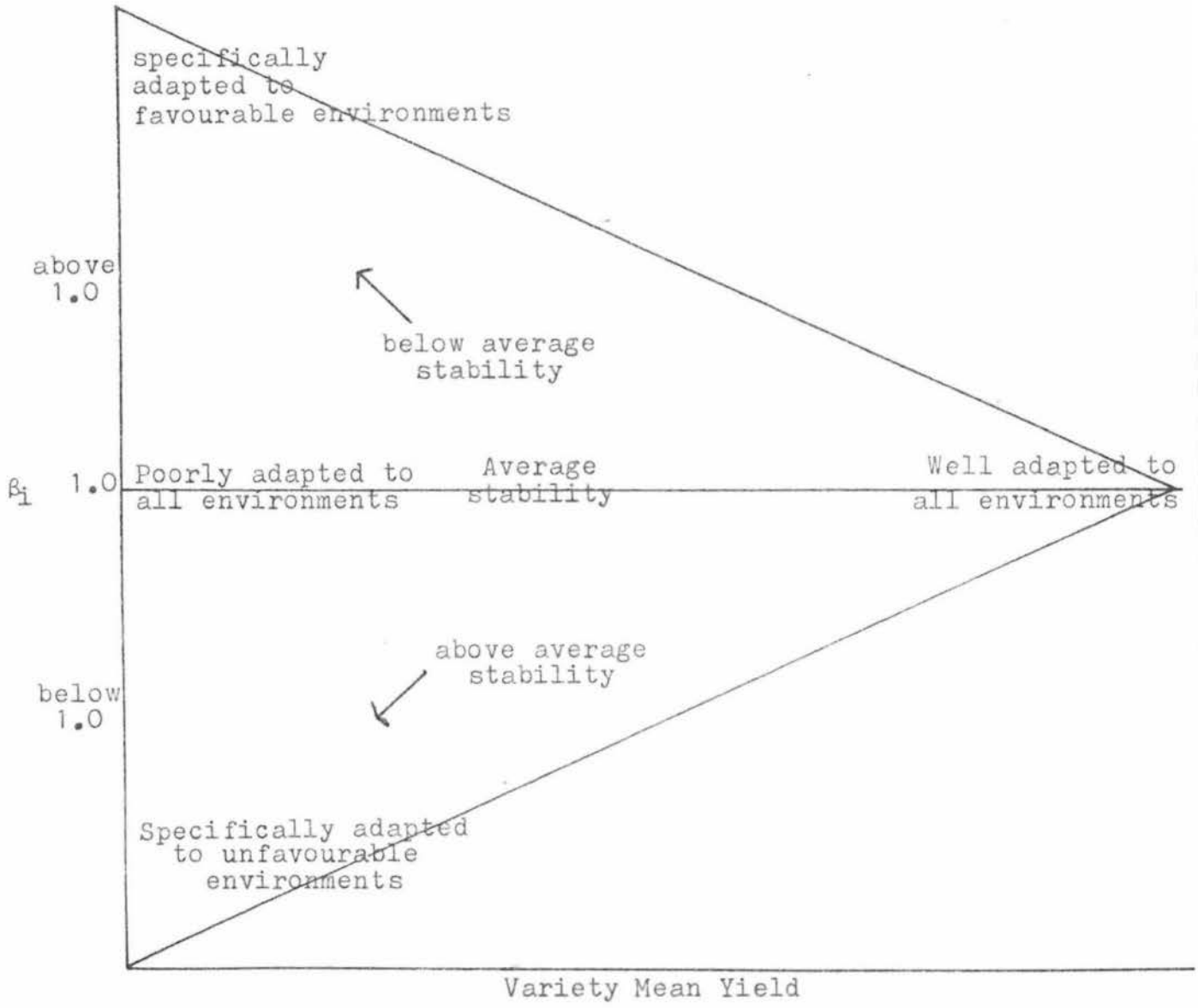
β_i as before.

I_κ

$\delta_{i\kappa}$

Figure 1.1

A generalised interpretation of regression coefficients plotted against variety mean yields



From Finlay and Wilkinson (1963)

Eberhart and Russell (1966) also outlined a regression approach given as model 4 of Table 1.3. Their partitioning is into:

(i) a linear component between genotypes with 1 d.f.

(ii) a linear component of the G x E interaction with (g-1) d.f.

(iii) deviations from regression with (ε -2) d.f.

They suggested that in addition to the regression coefficient a plant breeder may be interested in the degree to which a cultivar deviates from regression on an environmental index. Since large deviations from regression indicate unpredictable behaviour a parameter measuring this can be used to indicate stability of a cultivar.

Their parameter is given as such:-

$$S^2_{di} = \left[\sum_{\kappa} \delta_{i\kappa}{}^2/(n-2)\right] - S\varepsilon^2/r$$

Where $S\varepsilon^2/r$ = an estimate of the pooled error of $\bar{\chi}_{i\cdot\kappa}$'s

n = number of environments

r = number of blocks.

This can be related to the division of total variance about regression into pure error and lack of fit which is standard procedure with regression problems (Draper and Smith, 1966).

Eberhart and Russell considered that an 'ideal' cultivar would be one with a regression coefficient of 1 and S^2_{di} of 0.

Another group of workers to have suggested regression analysis used biometrical-genetical parameters instead of

straight biometrical ones. Bucio Alanis (1966) and Bucio Alanis and Hill (1966) began by describing 2 inbred lines differing at 1 locus.

$$P_1 = \mu - [d] + \varepsilon - \gamma d$$

$$P_2 = \mu + [d] + \varepsilon + \gamma d$$

i.e. the phenotypic effect of each line has a genetic effect $[d]$, an environmental effect ε and a G x E effect γ. $[d]$ is a constant and is the average deviation of the 2 lines from the midparent μ. They found that a linear relationship existed between ε and γ so that $y = [d] + \beta\varepsilon$.
This was expressed in more general terms by Perkins and Jinks (1967) in a model equivalent to model 3 at Table 1.3.

Various tests of significance are available to ascertain the effectiveness of the regression analysis. These are generally standard regression techniques such as the F test for Regression Mean Square/Deviation Mean Square and the calculation of R^2 (Draper and Smith, 1966).

The significance of β_i can also be tested as the difference between β_i and one.

It may also be useful to calculate the percentage of the G x E sum of squares accounted for by the regression sum of squares.

(ii) Modifications of Regression Analysis

Baker (1969) extended the basic regression approach of Finlay and Wilkinson in two ways. Firstly he used two different models, the additional one being for sites-years data. Secondly by carrying out the appropriate subtraction of means he regressed the $\alpha\eta_{i\kappa}$ effects on to η_{κ} effects, the $\alpha\lambda_{i\kappa}$ effects on to λ_{κ} effects and the $\alpha\lambda T_{i\kappa\ell}$ effects on to the $\lambda T_{\kappa\ell}$ effects. The possibility of regressing $\alpha T_{i\ell}$ effects on to T_{ℓ} effects was noted but not carried out because the variance component for $\alpha T_{i\ell}$ was not significant. The regression coefficients for the effects regression centred around zero as opposed to those of Finlay and Wilkinson regression which centred around one. In several cases he found significant effects regression coefficients indicating non independence of the different effects in the model. This indicates violation of one of the assumptions underlying the analysis of variance.

Simmonds (1979) suggests that for comparisons of new cultivars with standard ones the regression of the different genotype means over environments ($\bar{x}_{i\cdot\cdot's}$) on to each other may be informative. His application of this to potato trials Simmonds (1980) found that these regression coefficients had little predictive value in practice.

Eagles, Hinz and Frey (1977) examined the heterogeneity of regression to see if genetic variation for the character in question was present. Where they found significant heterogeneity they tested for convergence of the regression

lines and calculated the point of convergence to try and identify the type of interaction present. They concluded that the successful application of Finlay and Wilkinson regression depends on:-

(a) significant heterogeneity but small convergence variation;

(b) significant heterogeneity and significant convergence variation with the point of convergence lying within normal environmental range.

It was suggested by Freeman and Crisp (1979) that when two characters both show g x e interaction regressing one on to the other may help to interpret the relationship between them and help find the character of primary importance to g x e interaction. Verma et al. (1978) proposed that an ideal genotype would have a relatively high yield and low sensitivity in less favourable environments and high sensitivity in favourable environments. By dividing the environmental index into two subsets of less than favourable and more than favourable and comparing the two regression coefficients to see if they differed, they hoped to identify such genotypes. In a study of 10 inbred Nicotiana rustica lines they found one by this method which complied with their ideal genotype concept.

Similar work was done by Jinks and Pooni (1979) who investigated the possibility of two intersecting straight lines being more successful than a linear or quadratic model. They found that in cases where linearity failed this was an improvement.

Work was done by Suzuki and Abe (1975) and Nor and Cady (1979) in developing a regression model where the environmental index used was constructed through the use of Principal Component Analysis on a number of environmental variables. Both groups found this technique to be successful.

Such an environmental index has the advantages of being an independent measure incorporating as many environmental factors as possible.

It was postulated by Wright (1971) that if regression on to α_i was successful as well as the normal regression on to η_κ then there may be a regression on to both jointly. This can be detected by fitting a single parameter (κ) and this is outlined as model 5 of Table 1.4. He later relates this model to selection under various situations (Wright, 1976). He found that when using such a model selection for general adaptation will be an efficient strategy, as opposed to selecting in particular environments.

Two papers by Lin and Thompson (1975) and Lin, Binn and Thompson (1977) outline a method of grouping together genotypes with similar intercepts and regression coefficients from the standard Finlay & Wilkinson regression.

They calculate a statistic, d, which is a variance ratio for testing the hypothesis of a common regression line against the alternative of r independent regression lines.

$$d\ (\nu_1\nu_2\ldots\nu_r) = \frac{S.Sqs\ (\nu_1\nu_2\ldots\nu_r) - \sum_{i=1}^{r} SS(\nu_i)}{2(r-1)s^2},$$

where $^{2} = \sum_{i=1}^{m}$ S Sqs. $(\nu_i)/[(m-1)(n-2)]$,

S Sqs. $(\nu_1\ \nu_2 \cdots \nu_r)$ = Sums of Squares of deviation from regression for genotypes ν_1, $\nu_2 \cdots \nu_r$,

m = no. of genotypes,

and n = no. of environments.

They relate d to the Group Average clustering method of Sokal and Michener (1958) and extend this to encompass an F-test based stopping measure. For an analysis on carp data the clustering produced groups which were very consistent with their member's crossing backgrounds.

Hardwick and Wood (1972) outlined a method of finding a linear function of a set of environmental variables which could then be used as an independent environmental index for regression. Wood (1976) compared this method to that of a Principal Component analysis on environmental variables and found that they produced similar results.

Table 1.4 Regression Model

Model 5.

$$\hat{Y}_{i\kappa} = \mu + \alpha_i + \eta_\kappa + \kappa\alpha_i\eta_\kappa + s_{i\kappa}$$

μ = grand mean.

α_i = effect of the i^{th} genotype.

η_κ = effect of the κ^{th} environment.

κ = regression coefficient of interaction effects on to genotypes & environments jointly.

$s_{i\kappa}$ = the component of $Y_{i\kappa}$ independent of regression on to $\alpha_i\eta_\kappa$. (residual about regression).

From Wright, A.J. (1971).

(iii) Criticisms of Regression Analysis

Freeman and Perkins (1971) criticized the standard regression techniques because they use the environmental means as the environmental index. Because this index is not wholly independent of the y variable they pointed out that the regression analysis is invalidated since the sums of squares for the joint regression (df = 1) is not independent of the total sums of squares between environments (df = e -1). To overcome this problem they proposed an alternative approach where the index is obtained either by replication of the genotypes being investigated or the use of control genotypes (such as inbred parents).

Fripp and Caten (1971) used the standard technique as well as regressing on to values from control genotypes and found that a similar result was reached with both indices. Studies done by Williams (1974), Tan et al. (1979), Perkins and Jinks (1973), Jinks and Connolly (1973) and Snoad and Arthur (1976) all used methods of measuring an independent index of the environment and all found little difference between conclusions based on regressing on independent or non-independent environmental values.

Knight (1970) also examined the Finlay and Wilkinson regression technique by applying it to data where a precisely measured environmental factor was varied. He made several valuable points as a result of this.

1. A genotype which differed strongly from the others tested would show a markedly deviant regression since the others would largely determine the environmental means.

2. Different limiting factors result in equally low mean yields. Genotypes are unlikely to rank similarly for these factors but this is not revealed by Finlay and Wilkinson regression.

3. Combining data from different lengths of growth or different growth phases can be misleading.

4. Interpretation may be affected by the scale used and different scales may be appropriate for different genotypes.

Hardwick and Wood (1972) showed that a bias arises in the estimation of the $\beta_{i's}$ because it is assumed that the environmental mean is measured without error. In large experiments this will be small so it is not normally a problem.

Recent criticism of the regression analysis has been made either because of a lack of **linearity** in the data or because it has failed to account for a high percentage of the G x E sums of squares.

Blyth et al. (1976) found that overall linear regression accounted for only 9% of the G x E interaction in their wheat data and Baker (1969) had a very similar result.

Funnah & Mak (1980), Francis and Kannenberg (1978) and Easton and Clements (1973) had relatively large numbers of genotypes with non-significant regression coefficients.

To improve the linearity of their data Finlay & Wilkinson (1963) and Jowett (1972) used a log transformation of their raw data.

1.5 Pattern Analysis

(i) Introduction

The term Pattern Analysis has been put forward by W.T. Williams and co-workers as a name to encompass the hypothesis generating methods used to simplify complex data sets (Williams, 1975).

They make the distinction between these methods and those of classical statistics which are concerned largely with hypothesis testing.

The Pattern Analysis techniques have only a weak hypothesis on which they are based which is "There is some simpler pattern present in this data". (Williams, 1975). The classical null hypothesis that all elements are the same is not applicable.

Significance tests are seldom available in Pattern Analysis. Those that have been suggested e.g. for divisions in hierarchical classification are often sensitive to data anormality or have other problems. There are two distinct types of analyses involved:-

1. Ordination

The model underlying these methods is that the multiplicity of attributes measured were generated by a small number of underlying factors usually independent of each other.

2. Classification

The model for classification is that the population is best regarded as a set of mutually distinct smaller populations

rather than one single one.

Williams and Gillard (1971) point out that Pattern Analysis techniques can handle mixtures of different types of attributes and that they are most powerful when the number of attributes is large.

(ii) Classification

Classification has been described as a technique for allocating entities to initially undefined classes so that individuals in a class are in some sense close to one another (Cormack, 1971). Cluster Analysis is one method by which this is done. Several authors have attempted to classify clustering techniques Everitt (1974), Cormack (1971). That of Williams (1971) is presented as follows.

1. A classification may be
   (a) exclusive - an element occurs only in one class, i.e. no membership overlapping.

or (b) non-exclusive - an element may occur simultaneously in more than 1 class, i.e. it may have the necessary attributes to qualify for more than one class.

2. Exclusive classifications may be
   (a) intrinsic where all attributes are regarded as equivalent

or (b) extrinsic where the grouping must reflect discontinuity in an external attribute while still being based on internal attributes.

3. Exclusive intrinsic classifications may be:
   (a) a hierarchical strategy optimizing a route or phylogeny from individuals to the complete population = agglomerative, or from the population to the individuals = divisive. The fissions or fusions are always dichotomous.

(b) non-hierarchical, where no route is defined between groups and individuals. The strategy works on optimizing the homogeneity of the groups.

4. The non-hierarchical strategies are of 2 types:

(a) the groups are obtained serially so that successive groups are defined and removed from the population one at a time.

(b) the groups are obtained by simultaneous optimization by an iterative process.

5. Hierarchical strategies may be further subdivided depending on whether they are monothetic, i.e. based on one attribute or polythetic, based on many attrubutes.

Since an agglomerative method begins with individuals a monothetic agglomerative method is impossible.

There are problems associated with all of the three final types of classification. The divisive polythetic systems are relatively undeveloped; the divisive monothetic methods may produce misclassifications since grouping is done on the strength of 1 attribute.

The agglomerative, polythetic methods are highly developed but involve large amounts of computation time and may produce misclassifications since they start at the inter-individual level where error is greatest (Williams, 1971).

There are two considerations to be made in cluster analysis methods. Firstly there has to be a measure made of inter-individual likeness, and, secondly, this measure has to be

implemented by a strategy to extract groups or clusters. Cluster Analysis results are often presented in branching two dimensional diagrams called dendrograms (Clifford & Stephenson, 1975).

The Measurement of Similarity or Distance

A distinction may be made between a dissimilarity measure which has zero for identity, and a similarity measure which has its maximum positive value for identity. Most modern measures are constrained within 0 and 1. There are many possible measures of "likeness" but Williams (1971) identifies 3 main classes.

1. those based on the first Minkowski metric

$$\sum_{\kappa=1}^{\rho} \omega_\kappa | X_{i\kappa} - X_{j\kappa} |$$

When $\omega_\kappa = 1$ it is "city block" or "Manhatten" metric,

when $\omega_\kappa = \left| \frac{1}{X_{i\kappa} + X_{j\kappa}} \right|$ it is "Canberra" metric.

(Clifford & Stephenson, 1975), (Lance & Williams, 1967).

2. those based on the second Minkowski metric

$$\sum_{\kappa=1}^{\rho} \omega_\kappa ((X_{i\kappa} - X_{j\kappa})^2)^{\frac{1}{2}}$$

= Euclidean distance.

3. Information Statistics.

The basic measure of diversity by information statistic is the Shannon index.

$$H = N \log N - \sum_{\kappa=1}^{\rho} N_\kappa \log N_\kappa$$

N = no. individuals. ρ = no attributes.

(Clifford & Stephenson, 1975).

In addition to these the correlation coefficient has been used in some fields. This is a shape measure as opposed to a size measure as it measures identity when attrubutes occur in equal proportions.

A numerical function of pairs of points is said to be metric, if:

1. $d(x,y) = d(y,x)$;

2. $d(x,z) \leq d(x,y) + d(y,z)$;

3. When $d(x,y) = o$, $x = y$;

and 4. $d(x,x) = o$

This means that metric measures are geometrically attractive and perceptible. They are also advantageous in hierarchical systems where the measure should decrease as the hierarchy descends. This last is true of information statistics as well (Williams & Dale, 1965).

Williams (1971) suggests that for highly skewed binary data an information statistic would be best as the Euclidean measure is sensitive to outliers in data. If the data is defined by a small number of continuous variables, then Euclidean distance is preferable. Data with a few extreme outliers that should not dominate and is non-negative would suit the Canberra metric. With data of no striking features the different measures often produce similar solutions.

Sorting Strategies

Sorting strategies are methods of forming groups or clusters based on the similarity measures. The most well developed and commonly used strategies are the Agglomerative polythetic hierarchical ones. These are subdivided as follows.

1. Based on successive information gain. These are advantageous when the data is binary and disordered multistate with possible missing values. When the data is not of this form their weaknesses in matching zeros (absences) and in extra computational time negates this (Clifford & Stephenson, 1975)..

The basis of these strategies is the minimal information gain ΔI at each fusion where $\Delta I = IC - (IA + IB)$

ΔI = increase in diversity of attributes.

IC = diversity of the combined attributes.

IA, IB = diversity at the individual attributes before fusion.

2. The remaining strategies can be used with any indices of dissimilarity. They have different properties, affecting their suitability to different situations, which are outlined in the following.

(a) The properties of space-conserving, space-dilating and space-contracting refer to the degree that a group will appear to move nearer or away from the remaining elements.

Problems may occur with space-dilating strategies where "non-conformist groups of peripheral elements" may be formed.

Also with space-contracting strategies chaining may result (Lance & Williams, 1967).

(b) Strategies are combinatorial where the original data is not stored after the first measures are calculated. A non-combinatorial strategy, however, needs the original data retained for further calculations.

(c) A monotonic strategy is one which will not cause reversals in successive fusions in a dendrogram. With a non-montonic strategy the similarity measure may be less than that of the entities before they merged (Clifford & Stephenson, 1975).

(d) A compatible strategy has the same measures throughout the analysis. Where this is not true interpretation may be difficult.

There are several major sorting strategies which are widely used to form clusters on the basis of the distance measures. Their properties are documented by Anderberg (1973), Cormack (1971), Clifford & Stephenson (1975) and Lance & Williams (1967).

1. Single Linkage on Nearest Neighbour.

This is the oldest of the conventional strategies and the simplest. The distance between two groups is defined as the distance between their closest elements. It is combinatorial, compatible and space-contracting. Because of its well known tendency to chain it is often criticized and avoided. However, Jardin and Sibson (1968) support it on mathematical grounds.

2. Complete Linkage or Furthest Neighbour.

This operates in the opposite fasion to single linkage in that fusions are based on the minimal distance between the most remote entities in two groups. It is combinatorial, compatible and space-dilating. Applications of this method are apparently few but it has been found to be useful in ecological studies (Clifford & Stephenson, 1975).

3. Centroid.

In this method a group is considered defined in Euclidean space and is replaced by the coordinates of its centroid. The groups are fused on minimal distance between these centroids. The strategy is compatible, space conserving and combinatorial with metric measures. It may lack monotinicity and as a result reversals are often a problem.

4. Median.

This was suggested by Gower (1966) with a view to preventing large groups from dominating classifications. This is done by regarding the groups of unit size with the new centroid lying at the midpoint of the two fused ones. It is combinatorial, compatible except with correlation coefficients and space conserving. It may lack monotinicity and produce reversals. Teow (1978) found that it was so space-conserving that it produced chaining.

5. Group Average.

In this strategy fusion of an entity to a cluster occurs between those with the shortest mean distance between the

entity and all members of the cluster. With two clusters the mean distance between the entities of the two are calculated. The strategy is combinatorial, compatible and mildly space-conserving. Lance & Williams consider it a promising strategy.

6. Average Linkage Within the New Group.

This is similar to Group Average but in addition the sums of within group pairwise similarities are considered (Anderberg, 1973). Details of its properties are not known but it is thought to produce results like complete linkage (Anderberg, 1973).

7. Incremental Sums of Squares or Words Method.

A pair of elements whose distance measure is minimum are fused and thereafter entities are fused such that the sum of squares of distances within a cluster increases by the smallest amount. The total sum of squares is constant so as the sum of squares of distances within a cluster increases minimally, the sum of squares of distances between clusters increases maximally. Its properties are only known when squared Euclidean distance is used and is space-dilating and monotonic (Burr, 1970).

8. Flexible Method.

Lance and Williams (1967) generalised the preceding strategies into a single system by the relation

$$d_{hk} = \alpha_i d_{hi} + \alpha_j d_{hj} + \beta d_{ij} + \gamma |d_{hi} - d_{hj}|$$

where d = intergroup measure.

i, j, h = groups

α_i, α_j, β and γ are parameters determining the nature of

the sorting strategy. By varying these parameters a different strategy can be implemented and they showed that a system must be monotonic if $(\alpha_i + \alpha_j + \beta) \leqslant 1$. Their flexible strategy has the constraints of

$$\alpha_i + \alpha_j + \beta = 1 \qquad \alpha_i = \alpha_j \qquad \gamma = 0 \text{ and } \beta < 1.$$

They term β the cluster intensity coefficient as this can be varied according to whether a space-contracting, conserving or dilating strategy is required. Conventionally $\beta = -0.25$ for preliminary investigation which means it is mildly space-dilating (Clifford & Stephenson, 1975). A positive value for β means that the strategy will be space-contracting.

Which Strategy to Use?

Lance & Williams (1967) believe that any space-contracting strategy is obsolete since this means that classificatory boundaries are weakened. They include in this the single linkage method.

They suggest that the choice between space-conserving and space-dilating methods be made depending on the use to which the classification is put.

Williams (1971) suggests that a weakly clustering system may be best for situations where no a priori knowledge of discontinuities in the population is available.

A further restriction which can be placed on available strategies is the occurrence of reversals which are conceptually illogical and should be avoided (Clifford & Stephenson, 1975).

How Many Clusters?

In any agglomerative hierarchical clustering system the number of clusters present during the course of one exercise ranges from n, the number of individuals to one. Generally what is wanted is a compromise between these two extremes so some way of determining a truncation point is required. The normal practice appears to be to make a subjective chocie based on previous knowledge of the structure of the data set (Anderberg, 1973). Other workers admit to making arbitrary cut-off points, e.g. Mungomery et al. (1974). Teow (1978) implemented a probabilistic cut-off measure, based on the F ratio of amongst cluster sums of squares/within cluster sums of squares.

Application of Cluster Analysis to G x E Interaction

Although not working specifically to examine varietal adaptation the work of Abou-El-Fittouh et al. (1969) was the first to apply Cluster Analysis to this field. They clustered together environments by minimising the within cluster genotype-environment interactions and then identified the best zonation pattern for cotton variety trials in the U.S.A.

Mungomery et al. (1974) considered a genotypes performance in each environment as its coordinates in a multi-dimensional space. By applying clustering techniques they identified groups of cultivars which had similar responses over environments. They mapped the means of each cluster against an environmental index to further examine environmental responses.

A study done by Byth et al.(1976) clustered both genotypes and environments and they found that the genotype grouping reflected differences in genetical and selectional origins.

Clustering of environments was carried out by Shorter et al.(1977) to determine a basis for soybean test sites.

All of these previous studies used the Group Average strategy.

(iii) Ordination

The term ordination was first proposed by Goodall (1954) when he noted that analyses such as Factor Analysis do not result in classification in the ordinary sense but in the arrangement of data in a multi-dimensional series.

Ordination is thus a broad term encompassing several different multivariate techniques whose primary aim is to detect the underlying pattern in the data.

The term Factor Analysis is a broad term encompassing several procedures which analyse the intercorrelations within a set of variables (Cooley and Lohnes, 1971).

The intercorrelations may be calculated between each pair of variables (R type analysis) or between each pair of individuals (Q type analysis).

The factors may be extracted via a closed model where the factors are said to be defined because they are exact transformations of the original data or an open model where the factors are said to be inferred. In this case each variable is said to be influenced by commonly shared determinants (communalities) and unique determinants, the latter ones being ignored in the analysis as they do not contribute to relationships between variables (Cattell, 1965).

Where the factors are defined the general term for the analysis is Principal Component Analysis. However, if the correlation matrix is Q type the analysis may be termed

Principal Coordinate Analysis. In addition the linear function may be standardised in which case it may be called Principal Factor Analysis (Cooley and Lohnes, 1971). Where the factors are inferred the communalities are inserted in the leading diagonal of the correlation matrix, the extracted eigenvalues are standardised and the analysis is known as Factor Analysis.

It has been pointed out that Factor Analysis involves a greater element of subjectivity than Principal Component Analysis, and in addition involves greater computation time (Williams, 1975), (Rummel, 1970).

Computation of Principal Component Analysis

Principal Component Analysis (PCA) describes multivariate data by considering selected linear combinations of the variables $X_1 \ldots X_p$ rather than the variables themselves (Bofinger, 1975), (Cooley and Lohnes, 1971).

To look for the projection of the distribution which gives maximum variance the transformation must be found

$$z = \nu X$$

such that the variance of z is maximised subject to

$$\nu' \nu = 1$$

If the variance-covariance matrix of the standardised variables $X_1 \ldots X\rho$ is Σ then it can be shown that the variance of z will be $\nu' \Sigma \nu$ and so the maximisation is $\nu' \Sigma \nu$ subject to $\nu' \nu = 1$.

Using a Lagrange multiplier λ the determinental equation

$$/(\Sigma - \lambda I)\nu / = o \text{ is solved}$$

where I is the identity matrix.

The λ's which are the solutions of the equation are the eigenvalues of Σ and the solution vectors $\nu_1, \nu_2 \ldots \nu\rho$ are the mutually orthogonal eigenvectors.

The largest eigenvalue λ_1 associated with the eigenvector ν_1 gives the length and direction of the axis of maximum variation.

The λ_2 and ν_2 give the axis of largest variation in the set of all axes which are orthogonal to the first.

A factor structure matrix contains the correlation of each variable with the new components (or factors in the case of standardisation).

A factor score matrix contains each individual's score for the new components or factors.

Rotation to a Terminal Solution

A typical result of Principal Component Analysis is the production of one general factor and one or more bipolar factors (Cooley & Lohnes, 1971). Such situations are generally easily interpreted in terms of the original variables.

Where interpretation is not straightforward, rotation of the original reference axes may be carried out to give a "simpler" structure.

Rotation sets out to achieve the situation where

1. Many points lie near the final axes.
2. A large number of the points will lie near the origin.
3. Only a small number will remain removed from any axis.

In terms of the factor structure matrix, the rows and/or columns are simplified by making as many values in each row and/or column as close to zero as possible.

The different types of rotation are:-

1. Quartimax which emphasizes simplification of the rows so that a variable may load high on one factor but almot zero on all others. This allows interpretation of each variable in terms of its contribution to the factors.

2. Varimax which emphasizes simplification of the columns of the factor structure matrix so that a factor may load high or low on the variables. Each factor may then be interpreted more readily in terms of the variables contributing to it.

3. Equimax where both columns and rows are simplified.

4. Oblique rotation where the requirement of orthogonality is relaxed so that the axes are allowed to rotate freely to best summarise any clustering of variables.

How Many Principal Components?

A Principal Component Analysis may produce as many components as there are variables. Generally the first few components explaining the larger proportions of the total variance will be important and the later ones can be ignored.

Morrison (1967) suggests that if a suitably large proportion (e.g. 75%) of the variance is not explained by the first four or five components it is fruitless to persist in extracting vectors as the later components are likely to be very difficult to interpret. Approximate tests of significance are available for seeing how many Principal components are meaningful.

Bartlett (1950) outlined a chi-square test which examines the hypothesis that all subsequent roots are zero. This is sensitive to multivariate anormality and singularity of the dispersion (correlation) matrix.

An approximate F-test suggested by Mandel (1971) was used by Freeman and Dowker (1973) and Perkins (1971). Perkins found that seven components were significant but subsequently ignored this as the first two accounted for 84% of the total variation.

It has been found by Williams (1976) that the tests are of little use in practice because of the problems mentioned above and that consideration of the components which are interpretable and account for large portions of the variance is often the case.

Application of Principal Component Analysis to G x E Interactions

The application of PCA to 2 sets of 8 Nicotiana rustica lines measured over 10 and 9 seasons produced components which were readily interpretable (Perkins, 1971). The first component was related to general responses to environmental differences and the second was related to the difference at a single locus in the inbred lines.

Freeman and Dowker (1973) applied PCA to data from a series of carrot trials. They analysed the within-environment variation, the within-genotype variation and the residual variation in separate analyses. For the first two components in each analysis an interpretation in terms of environmental differences or genotypic differences could be made.

PCA was used primarily to confirm the results from clustering methods in the work by Mungomery et al. (1974). Their plots of the genotype positions in the first two axes (corresponding to the first two components) indicated two broad subpopulations within the genotypes tested and these could be related to different origins.

Work done by Shorter et al. (1977) used PCA to estimate the contributions of each of the environments to the first component using the factor structure matrix. An environment making a relatively large contribution was considered to be most useful as a test site for genotypes.

Okuno et al. (1971) [as reported in JIPB synthesis] have evidently used PCA to aid in the elucidation of adaptation of genotypes over environments as well.

1.6 Crop Evaluation in New Zealand

Wheat and barley are the two most important cereals grown in New Zealand in terms of both acreage grown and monetary value (Clardige, 1972).

Breeding programs are carried out for both wheat and barley and these produce new lines to be tested each year.

Douglas et al. (1977) examined the basis on which cultivars were assessed in New Zealand and made several recommendations. The acceptance or rejection of a new cultivar depends on its performance relative to the standard at the time. This is estimated by the degree by which the cultivar exceeds the standard, 5% greater than being the level generally set for acceptance. Depending on the amount of between-trial variability the number of trials needed to show this will vary.

They carried out an analysis of variance on some wheat data and noted the significance of the genotype x location interaction component. As their data was not suitable for analysis over seasons there was no indication of the significance of genotype x year interactions. However, they recommended testing genotypes over several years as they noted that such an interaction component was often important. On the basis of the significant genotype x location component they suggested the use of test site zones within which locational differences would be non-significant.

Work published by Mitchel and Cross (1977) quotes the results of an analysis of variance on four wheat genotypes over four sites and two years in the southern half of the North Island. For yield data the genotype x year component

was highly significant but the genotype x location component was non-significant. They also calculated stability parameters by the method of Shukla (1972) and found that the cultivar Karamu was "rather unstable".

Chapter 2 Materials and Methods

2.1 Objectives

The primary objective of this study was to make an assessment of the major analyses available to examine genotype-environment interactions.

Of particular interest was the different Pattern Analysis techniques since these have only recently been applied to this problem and they showed potential to overcome some of the problems which the more traditional methods have.

As a result of applying these various analyses to the data sets it was hoped that useful information on the performances of the trial cultivars would be produced.

In order to effectively assess the different methods of G x E analysis a set of data was required which had a good sample of genotypes measured in a range of environments.

In order to obtain data with a significant G x E interaction component the wider the range of genotypes and environments sampled, the better.

Such data sets where the sample of genotypes and trial methods are consistent are not common. However, national cultivar evaluation trials do provide suitable data and were therefore ideal for the purposes of this study.

2.2 Barley Data

The Ministry of Agriculture and Fisheries (MAF) ran a series of Barley cultivar evaluation trials in the 1978/1979 season. Thirteen cultivars were tested over seventeen locations from the barley growing areas of New Zealand (Dyson pers. comm). Leeston had two trials, one by MAF and the other by DSIR.

A second set of Barley data was obtained by including a 14^{th} cultivar (Hassan) which is an important standard in parts of the South Island (J.M. McEwan, pers. comm). This necessitated decreasing the number of sites to eleven.

Both of these sets of data were carried through the full analyses.

The MAF trials had variable plot sizes, ranging from 18 x 1.05 m to 40 x 1.37 m. The difference in plot size was not adjusted for as it was decided that they were all sufficiently large to avoid bias. There were three replicates per trial.

The parental pedigrees (Table 2.1) indicate that the gene base involved is probably a narrow one as all were either bred in European countries or have European parents (J.M. McEwan, pers. comm).

Table 2.1 Barley Cultivars

Cultivar	Parental Lines	Origin
Georgie	Vada x Zephyr	U.K.
Koreke	Ruby x Carlsberg x Lenta	Lincoln, N.Z.
Jupiter	Betina x Midas	U.K.
Neptune	Betina x Midas	U.K.
Goldmarker	TCE 131 x Midas	U.K.
000-72	Being tested from Europe Pedigrees not available.	
6487		
124-72		
Ark Royal	Proctor x HP5466	U.K.
Magnum	Magnif 104 x Universe	U.K.
Kaniere	(Carlsberg x Volla) x (Gazelle x kenia)	Lincoln, N.Z.
Zephyr	Heine 2149 x Carlsberg	Holland
Mata	(Britta x Carlsberg) x (Research x Ingrid)	Lincoln, N.Z.
Hassan	Delta x /Ägio x (Kenia x Arabische)/7	Holland

Table 2.2 Barley Trial Sites and Soil Types

Site	Area	Soil
Gore	Inland Southland	Waikoikoi silt loam
Pleasant Point	South Canterbury	Templeton silt loam
Fairview	Canterbury	Timaru silt loam
Outram	South Otago	Wehenga silt loam
Takapau	Hawkes Bay	Takapau silt loam
Bulls	Manawatu	Parewanui silt loam
Marton	Manawatu	Marton silt loam
Winfield	Canterbury	Waberton clay loam
Leeston	Canterbury	Waberton clay loam
Milton	South Otago	Tokomairiro silt loam
Incholme	North Otago	Claremont silt loam
Mitcham	Inland Canterbury	Hatfield silt loam
Methven	Inland Canterbury	Lynhurst silt loam
Kirwee	Central Canterbury	Lismore silt loam
Lincoln	Central Canterbury	Templeton silt loam
Dunsandel	Central Canterbury	Templeton silt loam

The genotype sample would have been even smaller if data was considered over several seasons trials. For this reason the data used for this study was restricted to one season. The character used for this study was grain yield, adjusted to 15% moisture expressed on a kg/ha basis. The trial sites and their soil types are listed in Table 2.2.

2.3 Wheat Data

Because of the narrow gene base in the Barley data and the restriction to one season it was decided to obtain another data set where these limitations were not present. For this purpose data from trials carried out by Crop Research Division, DSIR (Palmerston North), was gathered together.

This set consisted of two seasons trials from 1976/1977 and 1977/1978 over four sites in the lower North Island.

Ten genotypes were included in this study representing New Zealand standard cultivars, newly released cultivars and elite lines under trial (see Table 2.3).

The character used in this study was grain yield, adjusted to 15% moisture and expressed on a kg/ha basis.

The trials were set out in a randomised complete block design with three replicates per trial and a plot size of 50 m^2 (Mitchel and Cross, 1979b).

Table 2.4 lists the trial sites and their soil types.

Table 2.3 Wheat Cultivars

Cultivar	Parental Lines
61,01 = Tiritea Bulk	Raven x 66 RN430
300-301	II-60-155 x 3 Karamu/9
Oroua Bulk	Skemer x (Pitic 62 x II-53-526) x 2 sonora 64
293-201	Triple Dirk x 2 Karamu/9
Pataka	Complex cross from Minnesota
Gamenya	Gabo x [(5 Gabo x Mentana) x (2 Gaba x Kenya 117A)]
Rongotea Bulk	Raven x (Tezanos Pintos Precoz x 3 Anahvac)
61,06	Raven x 66 RN430
Karamu/9	Lerma Rojo - (Norin 10 - Brevor) x 3 Andes Enaro
301-302	II-60-155 x 3 Karamu/9

Table 2.4 Wheat Trial Sites

Site	Soil Type	Area
Halcombe	Halcombe silt loam	Rangitikei
Aorangi	Manawatu fine-sandy loam	Manawatu
Kairanga	Te Arakura sandy loam	Manawatu
Tiko Kino	Takapau silt loam	Hawkes Bay

2.4 Data Analysis

The initial step for the G x E interaction analysis was to carry out a Pooled Analysis of variance. This gave an indication of the size and significance of the different variance components, in particular that for G x E interaction.

From this analysis the G x E sums of squares were partitioned into ecovalences.

A regression analysis was carried out on both the G x E effects and the G x E means. This provided two parameters, the $\beta_{i's}$ and the variance about regression $\sigma_{Y \cdot X}^2$.

Both classification and ordination analyses were carried out. Cluster Analysis using Wards Incremental sums of squares strategy was done on both the G x E means and the G x E effects.

A Principal component analysis was also carried out.

(i) Programming

It was necessary to develop a computer program for the analysis of variance, regression analysis and ecovalence calculations since there was no appropriate program available.

Based primarily on a program YLGR from the University of Queensland, PANSY has six subprograms as well as the central program and calls on several functions to do additional calculations.

PANSY can be used for both sites-years or environments models and has the facility to handle multiple observations per plot and takes into account missing data.

Table 2.5

Outline of Program Pansy

PANSY
- File equations
- Total Array space set

↓ ↑

PRELIM
- Titles and data parameters read in
- Data read in and transformed if required - from "YLGRD". → ← **TRANS** - data transformed
  - square root x
  - log 10 x
  - Arcsin x
  - 1/1 + x
  - Arcsin 1/100x
- Titles and parameters printed out
- Work arrays dimensioned.

↓ ↑

ANAL
- Work arrays are initialised
- Data placed into appropriate arrays for analysis. → ← **INDEX** - indices determined for data placement in arrays.
- unadjusted sums of squares calculated.
- corrected sums of squares calculated.
- degrees of freedom calculated.
- homogeneity of error variances tested for. → ← **HOMER** - homogeneity of error variances tested by chi-square test.
- mean squares calculated.
- degrees of freedom for F tests calculated. → ← **DFCF** - degrees of freedom for complex F tests
- F values calculated.
- significance of F tests. → ← **PRBF** - significance of F values.
- Anova table printed out with coefficient of variation.
- All totals converted to means and print out.
- G x E means (G x SY) put on to disc file "Group".
- Finlay & Wilkinson regression and/or effects regression calculated.
- ecovalences and ratio to least ecovalence calculated.
- F = Reg. S.Sqs/Dev MS calculated.
- Significance of F tests. → ← **PRBF** - significance of F values.
- Regression parameters printed out
- Hanson's stability parameter calculated and printed out.

An outline of the different subprograms involved in PANSY is given in Table 2.5. Listings of these subprograms are presented in Appendix 3.

(ii) Analysis of Variance

A pooled analysis of variance was calculated for each of the sets of data. For the two barley sets the model used was the environments model (Model 1 of Table 1.1). For the wheat data a sites-years model was used (Model 2 of Table 1.1).

The Expected Mean Squares were those outlined in Table 1.2. By equating observed Mean Squares to expected mean squares, variance components and their standard errors were calculated.

The degrees of freedom for the complex F-tests were calculated using the formula of Satterthwaite (1946).

(iii) Ecovalences

Partitioning of the G x E sums of squares into the portions due to each genotype was carried out according to the method of Wricke (1962).

$$\varepsilon_i = \sum_{\kappa} (X_{i\cdot\kappa}/b - X_{i\cdot\cdot}/\varepsilon b - X_{\cdot\cdot\kappa}/gb + X\cdots/bg\varepsilon)^2$$

where b = no blocks g = no genotypes ε = no environments.

(iv) Hansons Parameter

Program Pansy also has the facility for calculating Hanson's parameter D_i^2 (Hanson, 1970).

$D_i^2 = \sum_{\kappa} Z_{i\kappa}^2$ where $Z_{i\kappa}$ represents the deviations of the i^{th} genotype performance from the expected in the κ^{th} environment.

The expected performance was estimated as

$$Z_{i\kappa} = \bar{X}\cdots + \bar{X}_i\cdot\cdot + \alpha X\cdot\cdot\kappa,$$

where $\alpha = 1 + \min \beta$.

The minimum β used for the analysis was calculated by the Finlay and Wilkinson regression analysis.

(v) Regression Analysis

Two types of Regression Analysis were carried out.

(i) Genotype yields in each environment were regressed against environment means (Finlay & Wilkinson, 1963).
i.e. $\bar{X}_{i\cdot\kappa}$ against $\bar{X}\cdot\cdot_{\kappa}$ for barley
$\bar{X}_{i\cdot\kappa\ell}$ against $\bar{X}_{\cdot\cdot\kappa\ell}$ for wheat.

(ii) Genotype effects for each environment were regressed against environmental effects (Baker, 1969).
i.e. $\alpha\eta_{i\kappa}$ against η_{κ} for barley.
$\alpha\lambda T_{i\kappa\ell}$ against $\lambda T_{\kappa\ell}$ for wheat.

Significance Testing for Regression

(i) The significance of β_1 from 1 was tested by the following:

$$\tau = \frac{(\hat{\beta}_1 - 1)}{\hat{\sigma}_{\beta 1}}$$ (Draper & Smith, 1966).

Where $\hat{\sigma}_{\beta 1} = \sqrt{\frac{\hat{\sigma}\cdot y\cdot x^2}{\Sigma (X_i - \bar{X})^2}}$,

with degrees of freedom $(\varepsilon - 2)$ and ε = no. environments.

(ii) The significance of regression was calculated by

$$F = \frac{(\Sigma\hat{Y}-\bar{Y})/1}{\Sigma(Y-\hat{Y})/(N-2)} = \frac{\text{Sum of Squares for regression}}{\text{Sum of Squares for deviations}/(\varepsilon-2)}$$

with degrees of freedom of 1 and $(\varepsilon-2)$.

This may also be interpreted as testing significance for $\hat{\beta}_1 = 0$.

(iii) The R^2 measures the proportion of the total variation about the mean $\bar{Y}$ explained by the regression (Draper & Smith, 1966).

This is calculated by

$$R^2 = \frac{\text{Sum of Squares of regression}}{\text{Total variance in } Y}$$

(Draper & Smith, 1966).

From the regression analyses two parameters are estimated which are of use to the analysis of G x E interaction.

The $\beta_{i's}$ indicate the functional relationship between genotype performance and environment.

The variance about regression $\hat{\sigma}_{Y\cdot X}$ indicates the degree by which a genotypes performance varies from this predicted linear trend.

(vi) Cluster Analysis

The data matrices on which this analysis was performed were of two sorts:

1. The yield of the individual genotypes in each environment i.e. $\bar{\chi}_{i\cdot\kappa's}$ for barley
   $\bar{\chi}_{i\cdot\kappa\ell's}$ for wheat.

2. The effects for each genotype-environment combination i.e. $\alpha\eta_{i\kappa's}$ for barley
   $\alpha\lambda T_{i\kappa\ell's}$ for wheat.

The program SIMMAT developed by Teow (1978) from Anderberg (1973) was used to calculate Squared Euclidean Distance (D^2) measures between each pair of genotypes.

$$D_{ij}^2 = \left[\sum_{\kappa=1}^{n} \omega_\kappa \, (X_{i\kappa} - X_{j\kappa})^2 \right]$$

Where i and j are two genotypes measured on n environments
ω_κ provides the facility for standardising the variables but this was not done since there were no strong outliers evident in the data and the variables were not measured on different scales (Everitt, 1974) (Anderberg, 1973).

CLUSAN, a clustering program developed from Anderberg (1973) provided seven different sorting strategies.

1. Centroid
2. Complete Linkage.
3. Group Average.
4. Average within New Group.
5. Single linkage.
6. Median.
7. Wards Incremental Sums of Squares.

In addition the Flexible method of Lance and Williams (1966) was added to CLUSAN (Cameron, 1980 unpublished).

In order to choose the most appropriate strategy for this study it was decided to analyse one set of data with all eight. Once the decision was made the other sets of data were analysed with that one strategy.

Anderberg (1973) suggested that it would be informative to examine the growth of Within-Cluster Sums of Squares/Total Sums of Squares as clustering proceeds. It is then possible to identify dominant variables, (i.e. those where among-cluster sums of squares is high in the early stages) and dormant variables (i.e. those where among-cluster sums of squares is low until later in the clustering process).

The program SEFWIG, Teow (1978) developed from Anderberg (1973) was used for this purpose.

Also produced by this program is the F ratio of AMS/WMS which, when tested for probability, can be used to indicate a cut off point for the dendrogram.

Teow (1978) took the point of minimum probability as the truncation point and found this to be successful. This was also the criterion used for this study.

A study was done of cluster structure and membership by the program POSTCA from Anderberg (1973).

The subroutine DIFFS Gordon (unpublished) tested the means of the clusters by the least significant difference test (Steel & Torrie, 1980).

(vii) Principal Component Analysis

A Principal Component Analysis was carried out using the Factor subprogram of SPSS (Nie et al, 1975).

The analysis was an R type one where the correlations were calculated between each pair of variables, the variables being environments for this study.

The data matrix consisted of the individual genotypes yield in each environment i.e. $\bar{X}_{i \cdot k}$'s for barley

$\bar{X}_{i \cdot k\ell}$'s for wheat.

The correlation matrix of the standardised variables Σ was calculated and the determinental equation

$/ (\Sigma - \lambda I)\mathbf{v} / = 0$ was solved.

Where λ is a Lagrange multiplier (the eigenvalues in this case)

I is the Identity matrix

and $\mathbf{v}$ the eigenvector giving the length and direction of the axis of maximum variation.

Rotation of the axes may be carried out to aid interpretation of the components. For this study both rotated and unrotated Factor Structure Matrices were produced.

The rotation method used was varimax rotation which simplifies the columns of the factor structure matrix so that a factor loads either high or low on the original variables.

Following the Principal Component Analysis, Bartlett's chi-square test was applied to the results to determine the number of significant components (Bartlett, 1950).

To test the significance of the remaining eigenvalues

$$\chi^2 = -\{n - 1/6(2\rho + 5) - 2/3\ \kappa\} \log_\varepsilon R_{\rho-\kappa}$$

n = no variables ρ = no. eigenvalues κ = no. eigenvalues removed.

$$R_{\rho-\kappa} = \text{determinant of correlation matrix}/\lambda_1 \ldots \lambda_\kappa \left[\frac{\rho-\lambda_1 \ldots \lambda_k}{\rho-\kappa}\right]^{\rho-\kappa}$$

Interpretation of the 'meaning' of each component was done from the Factor structure matrix which contains the correlation of each variable with each of the new components.

Ordination of the original variables was done by considering the factor scores of each individual on the principal components.

(viii) Ratio to a Standard Cultivar

The standard method of assessing cultivars in N.Z. is to compare each cultivar's performance to that of the standard (Douglas et al, 1977).

In order to compare this with the preceding methods the procedure was carried out with all three sets of data.

For barley the ratio $\dfrac{\bar{X}_{i \cdot \kappa}}{\bar{X}_{s \cdot \kappa}}$ was calculated

s being the standard cultivar zephyr.

For wheat the ratio $\dfrac{\bar{X}_{i \cdot \kappa \ell}}{\bar{X}_{s \cdot \kappa \ell}}$ was calculated

s being the standard cultivar karamu.

The number of times that a cultivar scored 5% higher or lower than the standard was noted, this being the normal "significance" level used (Douglas et al, 1977).

The average ratio over all environments was also calculated.

i.e. For barley $\frac{\bar{x}_{i..}}{\bar{x}_{s..}}$ for wheat $\frac{\bar{x}_{i...}}{\bar{x}_{s...}}$

The standard cultivars used for this study are those used by MAF and DSIR for their own analyses (Malcolm, 1978), (Mitchel and Cross, 1979b).

Chapter 3. Results and Associated Discussion

3.1 Analysis of Variance

(i) Barley: 13 genotypes

The test for homogeneity of error variances found a chi-square value of 127.875 with an associated probability of 0.000. This indicates that the error variances were heterogeneous. For the significance testing of genotypes and genotype x environment interaction the pooled error mean square is still the most appropriate error term to use. However for the testing of environments bias may occur when using the pooled error term (Cochran, 1947).

The number of environments involved in this pooled analysis was seventeen.

The coefficient of variation was low indicating a high level of precision in the pooled experiment (Table 3.11).

There was a high level of significance for the mean squares for all of the effects. The locations component was the largest (Table 3.11). This would indicate that bias from heterogeneity of error variances would be trivial. The significant differences found between locations gives an indication of the wide range of sites involved.

The high significance of the G x E interaction component indicates the need for further analyses to help interpret these trials results.

Table 3.11 Pooled Analysis of Variance: 13 Barley Genotypes

Source	df	Variance Component	s.e. Var. Comp.	F-Test df		F	Significance
Locations	16	2,335,984.55	795,741.55	16	75	43.23	****
Blocks within locations	34	97,176.66	25,804.73	34	408	8.94	****
Genotypes	12	33,643.47	17,606.29	12	192	3.34	****
Genotype x locations	192	191,858.25	25,141.72	192	408	4.62	****
Error	408	159,164.46	11,116.53				

**** Significant at 0.1% level.

Coefficient of Variation = 8.4%.

(ii) Barley: 14 genotypes

Heterogeneity of the error variances was also present in this data set with a chi-square value of 100.784 and a probability of 0.000.

The coefficient of variation was again low (Table 3.12). This set of trials differed from the previous one by the omission of six Canterbury sites but the locational differences were still very evident in the Pooled Analysis of Variance (Table 3.12).

The standard cultivar Hassan was added to the genotypes but the genotypic differences were less significant for these trials. The G x E interaction effect was again highly significant.

Table 3.12 Pooled Analysis of Variance: 14 Barley Genotypes

Source	df	Variance Component	s.e. Var. Comp.	F-Test	df	F	Significance
Locations	10	2,261,548.02	299,209.28	10	50	38.57	****
Blocks within locations	22	104,682.22	33,906.25	22	286	9.23	****
Genotypes	13	25,470.10	19,369.19	13	130	1.95	*
Genotype x locations	130	235,458.84	10,852.90	130	286	4.97	****
Error	286	178,011.50	14,834.29				

* = significant at 5% level.

**** = significant at 0.1% level.

Coefficient of Variation = 8.0%.

(iii) Wheat 10 genotypes

The wheat genotypes were tested over four sites in two seasons. The error variances were again heterogeneous with a chi-square value of 35.645 and an associated probability of 0.000. (Table 3.13).

The mean square for sites was non-significant which may be a reflection of the relatively small area involved in the trials, i.e. the lower North Island.

The year component was significant at the 5% level and the highly significant sites x years component indicates the variation in the effect of years at the different sites.

The differences between genotypes was significant at the 5% level.

There are three genotype x environment interaction components for this model. The genotype-year component was highly significant indicating that the genotypes reacted differently to the seasonal changes. There was no significant genotype x site and genotype x site x year interactions.

Table 3.13 Pooled Analysis of Variance: 10 Wheat Genotypes

Source	df	Variance Component	s.e. Var. Comp.	F-Test	df	F	Significance
Sites	3	422,226.42	528,730.14	3	3	2.22	NS
Years	1	2,335,363.02	2,051,582.78	1	3	14.05	*
Years x sites	3	613,535.21	430.025.28	3	19	9.63	***
Blocks within sites	16	179,785.12	64,650.13	16	144	13.75	****
Genotypes	9	94,087.18	61,032.34	10	14	2.66	*
Genotypes x locations	27	16,054.02	15,265.43	27	27	1.50	NS
Genotypes x years	9	73,223.18	19,170.37	9	27	5.54	****
Genotypes x sites x years	27	17,563.20	17,827.11	27	144	1.37	NS
Error	144	141,020.40	16,505.20				

NS = not significant

* = 5% level.

*** = 1% level.

**** = 0.1% level.

3.2 Stability Parameters

Three different parameters were calculated which had been described as measuring stability of a genotype by their various authors.

The ecovalence of Wricke assigns relative stability to a genotype with a low proportion of the G-E sums of squares. Hanson's parameter involves the definition of a stability space with a stable genotype lying close to the origin. Eberhart and Russell related the $\hat{\sigma}^2_{\gamma \cdot x}$ to stability saying that the more deviation from regression shown by a genotype the less prdictable and therefore the less stable its performance was.

These three parameters are presented in two ways:

(a) the actual values are listed along with the cultivars mean yield over all environments;

(b) the cultivars ranking for each parameter with the highest ranking being assigned to the most stable cultivar according to that parameters definition of stability. stability.

(i) Barley: 13 genotypes

For this set of data the extreme rankings for the different parameters agree very well, i.e.Goldmarker and Magnum share the bottom rankings and Koreke and Kaniere have uniformly high rankings (Table 3.21).

An examination of the cultivar mean yields reveals that Goldmarker was highest for this but its evident lack of

Table 3.21 Stability Parameters: 13 Barley Genotypes

GENOTYPE	MEAN YIELD	S.E. MEAN YIELD	ECOVALENCE(1)	HANSON'S(2)	s^2d_i (3)
Georgie	5006.31	1669.67	2071223.35	2058.7	62865.52
Koreke	4464.67	1405.39	2026870.60	1199.8	62802.93
Jupiter	4600.29	1606.05	2969776.26	2061.1	141786.14
Neptune	4551.82	1632.49	3021360.92	2140.3	141074.84
Goldmarker	5127.55	1813.94	6131501.17	3032.3	296900.12
000-72	4553.39	1560.93	4624411.44	2255.3	254384.81
6487	5002.43	1590.35	3172824.93	2059.1	157310.35
124-72	4593.67	1562.69	4172773.75	2178.0	224811.54
Ark Royal	4722.08	1423.47	4645534.18	1925.7	225900.47
Magnum	4947.00	1711.44	5774420.91	2771.4	317109.54
Kaniere	4903.27	1454.78	2235259.87	1444.5	87314.38
Zephyr	4565.78	1506.32	2129046.11	1587.6	87543.14
Mata	4794.16	1388.60	4048304.98	1702.1	177748.98

Rankings for Stability Parameters and Mean Yield

	(1)	(2)	(3)	Mean Yield
Georgie	2	6	2	2
Koreke	1	1	1	13
Jupiter	5	8	6	8
Neptune	6	9	5	12
Goldmarker	13	13	12	1
000-72	10	11	11	11
6487	7	7	7	3
124-72	9	10	9	9
Ark Royal	11	5	10	7
Magnum	12	12	13	4
Kaniere	4	2	3	5
Zephyr	3	3	4	10
Mata	8	4	8	6

stability should be taken into account as well.

The standard variety Zephyr scored relatively highly for stability but was amongst the lowest for mean yield. Kaniere the cultivar noted for its wide adaptability (Coles, 1980) also scored well for stability and its mean yield was amongst the highest.

The cultivar Georgie did well for both stability and mean yield. Koreke, the most stable cultivar, had the lowest mean yield.

(ii) Barley: 14 genotypes

Goldmarker and Magnum again had uniformly low stability rankings and the additional cultivar Hassan took the place of Koreke as the most stable genotype overall (Table 3.22).

Hassan has been found to be a promising cultivar in North Island trials and is used as a standard in some South Island sites (Malcolm, 1978).

Georgie, Hassan and Kaniere appear to be the best cultivars in terms of high mean yield and relative stability.

Table 3.22 Stability Parameters: 14 Barley Genotypes

GENOTYPE	MEAN YIELD	S.E. MEAN YIELD	ECOVALENCE(1)	HANSONS(2)	s^2d_i (3)
Georgie	5534.24	1681.05	189914.90	1864.5	108459.57
Koreke	4863.03	1403.14	1625406.10	1170.9	109418.03
Jupiter	5104.24	1593.51	2464979.98	1815.9	208251.98
Neptune	5040.00	1680.79	2815072.05	2050.9	222449.33
Goldmarker	5633.03	1818.04	4686879.12	2600.1	374908.92
000-72	5087.88	1513.95	3777612.45	1950.4	356768.29
6487	5549.70	1555.39	2700952.13	1795.1	240119.9
124-72	5128.48	1417.23	1432607.83	1142.0	92268.38
Ark Royal	5176.67	1340.04	3802405.38	1652.7	300564.2
Magnum	5378.18	1679.61	4717507.98	2391.6	453016.26
Kaniere	5410.91	1333.44	2159232.36	1174.1	140595.49
Hassan	5303.03	1544.04	663471.18	1230.9	7555.92
Mata	5249.39	1311.27	3777607.98	1595.0	282937.08
Zephyr	5014.24	1515.58	1799830.78	1494.2	140541.87

Rankings for Stability Parameters and Mean Yield

	(1)	(2)	(3)	Mean Yield
Georgie	4	10	3	3
Koreke	3	2	4	14
Jupiter	6	9	7	10
Neptune	9	12	8	12
Goldmarker	12	14	13	1
000-72	10	11	12	11
6487	7	8	9	2
124-72	2	1	2	9
Ark Royal	11	7	11	8
Magnum	13	13	14	5
Kaniere	5	3	6	4
Hassan	1	4	1	6
Mata	14	6	10	7
Zephyr	8	5	5	13

(iii) Wheat: 10 genotypes

The cultivar Karamu has been noted for its high yielding ability (Mitchel & Cross, 1977). This is borne out by its high mean yield for these trials. Its stability ranking however (Table 3.23) is uniformly low for all of the parameters, so its performance is markedly inconsistent over environments.

There was less correspondence between the parameters rankings for each genotype in this data set.

Gamenya has the lowest mean yield and is very stable in its performance according to Hanson's parameter and s^2d_i. Its ecovalence however was the highest of all the genotypes.

The line 61,01 (now a cultivar known as Tiritea) had high stability rankings and a relatively high mean yield.

Table 3.23 Stability Parameters for Wheat Genotypes

GENOTYPE	MEAN YIELD	S.E. MEAN YIELD	ECOVALENCE(1)	HANSONS(2)	$s^2_{d_i}$ (3)
61,01	5284.83	1456.84	151415.60	1182.9	21830.43
300-301	5731.33	1748.03	862315.28	2033.9	22586.86
Oroua	5056.13	1345.67	536832.86	959.9	23931.32
293-201	5174.08	1206.83	1189701.21	726.5	54860.65
Pataka	5164.83	1478.72	672191.60	1384.9	64970.72
Gamenya	4573.29	1071.89	1451267.68	0.0	4539.93
Rongotea	5294.46	1645.73	771732.20	1800.0	37059.72
61,06	4809.42	1415.58	341791.79	1117.2	7538.55
Karamu/9	5831.58	1646.62	1058375.60	1856.8	90216.26
301-302	5367.50	1607.29	535237.78	1665.6	12304.52

Rankings for Stability Parameters and Mean Yield

	(1)	(2)	(3)	Mean Yield
61,01	1	5	2	5
300,301	7	10	1	2
Oroua	4	3	6	8
293-201	9	2	8	6
Pataka	5	6	9	7
Gamenya	10	1	3	10
Rongotea	6	8	7	4
61,06	2	4	4	9
Karamu	8	9	10	1
301-302	3	7	5	3

3.3 Finlay and Wilkinson Regression

(i) Barley: 13 genotypes

Table 3.31 contains the results for this regression analysis. Initially the significance of the β_i's from 1.0 was carried out using the $\hat{\sigma}^2_{\gamma \cdot x}$ in the calculation. When none of the β_i's were found to be significant in this respect it was decided to refine the error term by subtracting an estimated pure error term from $\hat{\sigma}^2_{\gamma \cdot x}$. Draper and Smith (1966) describe this procedure and call the remaining error term the "lack of fit". Using this lack of fit term the significance tests from 1.0 were recalculated and these results are reported in Table 3.31. It was found that three genotypes were significant at the 5% level, Georgie, Koreke and Mata. A genotype such as Goldmarker had a relatively high β_i but its "lack of fit" was very large making the test non-significant. The t-tests between individual β's found no significant differences between any combination of genotypes (Steel & Torrie, 1980), indicating again that no differences existed between these genotypes in their response to environmental change.

The strength of the linear relationship between the $\bar{X}_{i \cdot \kappa}$'s and the $\bar{X}_{\cdot \cdot \kappa}$'s is illustrated by the high R^2 values and the significance of the Regression Mean Squares.

The cultivars are ranked according to their β_i value using Finlay and Wilkinson's definition of a stable cultivar having a slope of 0 as a basis. Goldmarker, Georgie and Magnum would be judged the least stable by this definition. The alternative definition of stability as given by Eberhart and Russell (1966)

Table 3.31 Barley Regression Parameters: 13 Genotypes

Genotypes	β_i	Signif. of β_i from 1	R^2	$\hat{\sigma}^2_{Y \cdot X}$	Signif. of Regression	Rank of β_i
Georgie	1.093[a]	*	0.963	115920.34	*****	12
Koreke	0.913[a]	*	0.948	115857.75	*****	2
Jupiter	1.035[a]	NS	0.933	194840.96	*****	9
Neptune	1.053[a]	NS	0.935	194129.68	*****	10
Goldmarker	1.152[a]	NS	0.906	349954.94	*****	13
000-72	0.982[a]	NS	0.888	307439.63	*****	5
6487	1.021[a]	NS	0.926	210365.17	*****	8
124-72	0.988[a]	NS	0.899	277866.36	*****	7
Ark Royal	0.890[a]	NS	0.878	278955.29	*****	1
Magnum	1.076[a]	NS	0.888	370164.36	*****	11
Kaniere	0.937[a]	NS	0.918	140369.20	*****	3
Zephyr	0.985[a]	NS	0.948	140597.96	*****	6
Mata	0.977[a]	*	0.945	230803.80	*****	4

NS = not significant

* = significant at the 5% level

***** = significant at the 0.01% level.

Letter subscripts indicate significance groupings of β_i's as calculated by individual t-tests (Steel and Torrie, 1980).

where $\beta_i = 1$ would indicate that Goldmarker, Ark Royal and Koreke were the least stable genotypes.

It would appear from the significance tests from 1.0 and the individual t-tests that little "real" difference occurred between the genotypes in respect to their response to environmental change. In terms of figure 1.1 genotypes Koreke and Mata are specifically adapted to less favourable environments and Georgie to favourable environments.

(ii) Barley: 14 genotypes

The results for this analysis are presented in Table 3.32. The significance from 1.0 for the β's was again recalculated using the lack of fit error. It was then found that Georgie and Hassan had β's significantly different from 1.0.

No significant differences were found between the β's when all combinations of the genotypes were tested.

A strong linear relationship was again found as illustrated by the R^2 values and the high significance of the Regression mean squares.

Using the Finlay and Wilkinson criterion the ranking of the cultivars indicates that again Goldmarker, Georgie and Magnum would be judged least stable. (Table 3.32).

Using the Eberhart and Russell definition Goldmarker, Georgie, Mata and Ark Royal would be seen to be the least stable genotypes.

Again there did not appear to be any significant differences between genotypes in their response to environmental change.

Table 3.32 Barley Regression Parameters: 14 Genotypes

Genotypes		Signif. of β from 1	R^2	$\hat{\sigma}^2_{\gamma \cdot x}$	Signif. of Regression	Rank of β_i
Georgie	1.129[a]	(NS)	0.951	167796.83	*****	13
Koreke	0.932[a]	NS	0.930	168755.19	*****	4
Jupiter	1.049[a]	NS	0.914	267589.14	*****	9
Neptune	1.109[a]	NS	0.918	281786.49	*****	12
Goldmarker	1.183[a]	NS	0.893	434246.08	*****	14
000-72	0.962[a]	NS	0.851	416105.45	*****	6
6487	1.016[a]	NS	0.898	299457.06	*****	8
124-72	0.946[a]	NS	0.938	151605.54	*****	5
Ark Royal	0.844[a]	NS	0.836	359901.36	*****	2
Magnum	1.067[a]	NS	0.851	512353.42	*****	11
Kaniere	0.875[a]	NS	0.908	199932.65	*****	3
Hassan	1.051[a]	*	0.977	66893.08	*****	10
Mata	0.827[a]	NS	0.837	342274.24	*****	1
Zephyr	1.006[a]	NS	0.929	199879.03	*****	7

NS = not significant
(NS) = significant at the 10% level
* = significant at the 5% level
***** = significant at the 0.01% level.

Letter subscripts indicate significance groupings of β_i's by individual t-tests (Steel and Torrie, 1980).

(iii) Wheat genotypes

These results are presented in Table 3.33.

The significance testing of β_i's from 1.0 was done using the lack of fit error term. This showed that five of the cultivars had slopes significantly different from 1.0. Three of these cultivars (300-301, Rongotea, and 301-302) were steeper than average which in terms of the Finlay and Wilkinson diagram (figure 1.1) would mean that they were of below average stability and better adapted to favourable environments. Two cultivars (293-201 and Gamenya) were flatter than average, indicating that they were of above average stability and specifically adapted to unfavourable environments.

The R^2 values and the significance of the Regression Mean Squares indicate a strong linear relationship between the variables.

The individual t-tests between the β_i's showed no significant differences were present between the genotypes responses to environmental change.

Table 3.33 Wheat Regression Parameters

Genotypes	β_i	Signif. of β_i from 1	R^2	$\hat{\sigma}^2_{\gamma \cdot x}$	Signif. of Regression	Rank of β_i
61,01	1.0046^a	NS	0.991	25176.37	*****	6
300-301	1.2072^a	***	0.994	24419.94	*****	10
Oroua	0.9183^a	NS	0.971	70938.12	*****	3
293-201	0.8137^a	*	0.947	101867.45	*****	2
Pataka	1.0044^a	NS	0.962	111977.52	*****	5
Gamenya	0.7321^a	*****	0.972	42466.87	*****	1
Rongotea	1.1266^a	*	0.977	84066.52	*****	9
61,06	0.9705^a	NS	0.979	54545.35	*****	4
Karamu	1.1187^a	NS	0.962	137223.06	*****	8
301-302	1.1037^a	***	0.983	59311.32	*****	7

NS = not significant.
* = significant at 5% level.
*** = significant at 1% level.
***** = significant at 0.1% level.

Letter subscripts indicate significance groupings of β_i's as calculated by individual t-tests (Steel and Torrie, 1980).

3.4 Effects Regression

Tables 3.41, 3.42 and 3.43 contain the regression parameters produced from regressing $\alpha\eta_{i\kappa}$ on η_{κ} for barley,

$\alpha\lambda T_{i\kappa\ell}$ on $\lambda T_{\kappa\ell}$ for wheat.

In contrast to the Finlay and Wilkinson regression β_i, which has an average value of 1.0, the β_i value for the effects regression will have an average value 0.0. This is because the effects of μ, α_i and η_{κ} have been removed from the regression sum of squares leaving only those for $\alpha\eta_{i\kappa}$ and η_{κ} which sum to zero by definition. This can be seen in the results in the above tables.

The R^2 values and the general lack of significance for the Regression Mean Square (an indirect test for $H_0:\beta = 0$) reflect the lack of linear relationship between the variables. As has been already stated one of the assumptions underlying an analysis model is that the effects are independent of one another. Therefore the lack of linearity seen here is to be expected.

However the wheat data suggests that for two cultivars 300-301 and Gamenya there was a strong correlation between the $\alpha\lambda T_{i\kappa\ell}$'s and the $\lambda T_{\kappa\ell}$'s. Since this evident lack of independence of effects may have caused bias in the Analysis of Variance estimates the regression coefficients have been used to get an estimate of the covariance between the effects and the Analysis of Variance estimates adjusted accordingly. This has been presented in Tables 3.44 and 3.45. A comparison of this with the unadjusted Anova table (Table 3.13) shows that no difference in the significance tests resulted from this so the effect of the covariance was minimal. This result suggests that even though one of the basic assumptions of Anova was violated the analysis was not biased by this and can therefore be seen to be quite robust.

Table 3.41

Regression Parameters for Effects 13 Barley Genotypes

Genotype	β_i's	R^2	Signif. of Regression
Georgie	0.09	0.160	NS
Koreke	-0.09	0.143	NS
Jupiter	0.04	0.016	NS
Neptune	0.05	0.036	NS
Goldmarker	0.15	0.144	NS
000-72	-0.02	0.002	NS
6487	0.02	0.005	NS
124-72	-0.01	0.001	NS
Ark Royal	-0.11	0.099	NS
Magnum	0.08	0.041	NS
Kaniere	-0.06	0.058	NS
Zephyr	-0.02	0.009	NS
Mata	-0.12	0.145	NS

NS = not significant.

Table 3.42

Regression Parameters for Effects 14 Barley Genotypes

	β_i's	R^2	Signif. of Regression
Georgie	0.13	0.205	NS
Koreke	-0.07	0.066	NS
Jupiter	0.05	0.023	NS
Neptune	0.11	0.099	NS
Goldmarker	0.18	0.166	NS
000-72	-0.04	0.008	NS
6487	0.02	0.002	NS
124-72	-0.05	0.048	NS
Ark Royal	-0.16	0.148	NS
Magnum	0.07	0.023	NS
Kaniere	-0.12	0.167	NS
Hassan	0.05	0.093	NS
Mata	-0.17	0.184	NS
Zephyr	0.01	0.001	NS

NS = not significant.

Table 3.43

Regression Parameters for Effects Wheat

Genotype	$\beta_{i's}$	R^2	Signif. of Regression
61,01	0.00	0.002	NS
300-301	0.21	0.830	***
Oroua	-0.08	0.207	NS
293-201	-0.19	0.486	(NS)
Pataka	0.00	0.001	NS
Gamenya	-0.27	0.824	***
Rongotea	0.13	0.346	NS
61,06	-0.03	0.042	NS
Karamu/9	0.12	0.222	NS
301-302	0.10	0.335	NS

NS = not significant
(NS) = significant at 10% level
*** = significant at 1% level.

Table 3.44

Adjustment for Covariance Amongst $\alpha\lambda T$ and λT Effects

Source	E(MS)
sites	$\sigma^2+b\sigma^2_{gs\gamma}+b\gamma\sigma^2_{gs}+g\sigma^2_{b(s\gamma)}+gb\sigma^2_{s\gamma}+gb\gamma\sigma^2_{s}+(2)b\ \text{cov.}_{g\gamma s,\gamma s}$
years	$\sigma^2+b\sigma^2_{gs\gamma}+bs\sigma^2_{g\gamma}+g\sigma^2_{b(s\gamma)}+gb\sigma^2_{s\gamma}+gbs\sigma^2_{\gamma}+(2)b\ \text{cov.}_{g\gamma s,\gamma s}$
sites x years	$\sigma^2+b\sigma^2_{gs\gamma}+g\sigma^2_{b(s\gamma)}+gb\sigma^2_{s\gamma}+(2)b\ \text{cov.}_{g\gamma s,\gamma s}$
blocks $(s\gamma)$	$\sigma^2+g\sigma^2 b(s\gamma)$
genotypes	$\sigma^2+b\sigma^2_{g\gamma s}+bs\sigma^2_{g\gamma}+b\gamma\sigma^2_{gs}+bs\gamma\sigma^2_{g}+b\ \text{cov.}_{g\gamma s,\gamma s}$
genotypes x sites	$\sigma^2+b\sigma^2_{g\gamma s}+b\gamma\sigma^2_{gs}+b\ \text{cov.}_{g\gamma s,\gamma s}$
genotypes x years	$\sigma^2+b\sigma^2_{g\gamma s}+bs\sigma^2_{g\gamma}+b\ \text{cov.}_{g\gamma s,\gamma s}$
genotypes x sites x years	$\sigma^2+b\sigma^2_{g\gamma s}+b\ \text{cov.}_{g\gamma s,\gamma s}$
error	σ^2

Genotype	Covariance
61,01	0
300-301	500195
Oroua	-190550
293-201	-452557
Pataka	0
Gamenya	-643107
Rongotea	722503
61,06	71456
Karamu	285825
301-302	238188
Total	389040

Table 3.45 Adjusted Pooled Analysis of Variance

Source	Adjusted S.S.	Adjusted M.S.	Re-estimated F	Signif.
sites	136704500.20	45568166.67	2.47	NS
years	300741778.55	300741778.55	14.22	*
sites x years	60414772.17	20138257.39	9.57	***
blocks (sy)				
genotypes	30452202.48	3383578.05	2.73	*
genotypes x sites	7441881.44	275625.24	1.54	NS
genotypes x years	9262453.50	1029161.50	5.74	*****
genotypes x sites x years	4841129.61	179301.10	1.27	NS

NS = not significant
* = significant at 5% level
*** = significant at 1% level
***** = significant at 0.1% level.

3.5 Cluster Analysis

(i) Comparison of Clustering Strategies

The eight available clustering strategies were applied to the G x E means data for the 13 barley genotypes. This allowed an assessment to be made of the strategies so that the most suitable one could be chosen.

The single Linkage Strategy exhibited a mild chaining effect (figure 3.51). This is due to its space-contracting tendencies which have been noted by previous authors (Williams, 1971). There were single entity clusters present until late stages of clustering.

Both the Centroid (figure 3.52) and Median (figure 3.53) methods had the problem of reversals occurring in the clustering process. For the Centroid method this occurred at two stages, for the Median method at one stage. This problem had been noted before by Lance and Williams (1967), Teow (1978), Anderberg (1973).

A mild chaining effect could be seen with the Average Within Groups strategy (figure 3.55).

Program SEFWIG was run for the Group Average (figure 3.54), Average Within, Complete Linkage (figure 3.56), Wards (figure 3.57) and Flexible (figure 3.58) strategies. This determined a truncation point for each dendrogram based on the minimum F value for Amongst Cluster Sum of Squares/Within Cluster Sum of Squares ratio (Teow, 1978).

Figure 3.51 Dendrogram for Single Linkage Clustering of 13 Barley Genotype Means

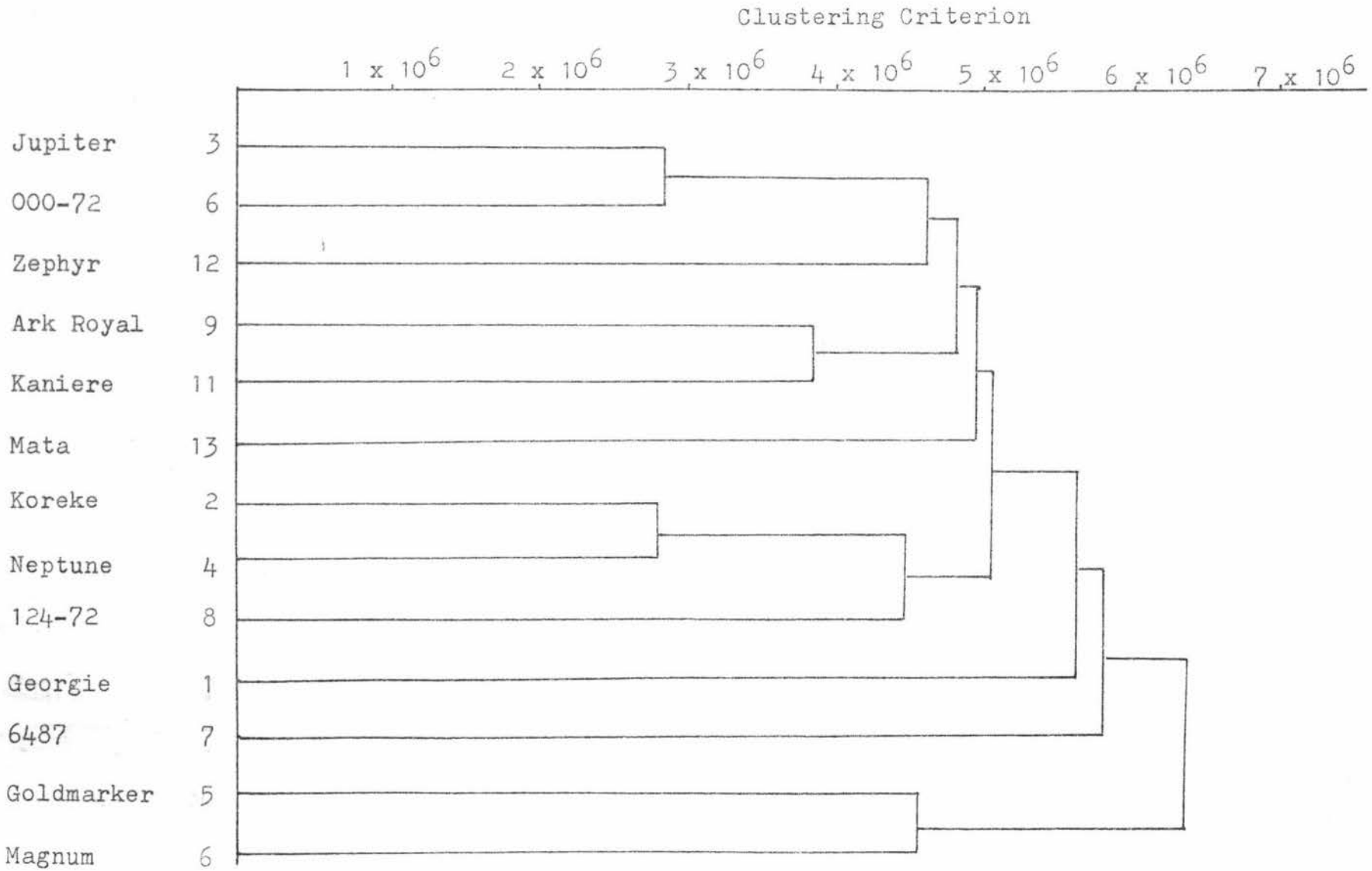


Figure 3.52 Dendrogram for Centroid Clustering of 13 Barley Genotype Means

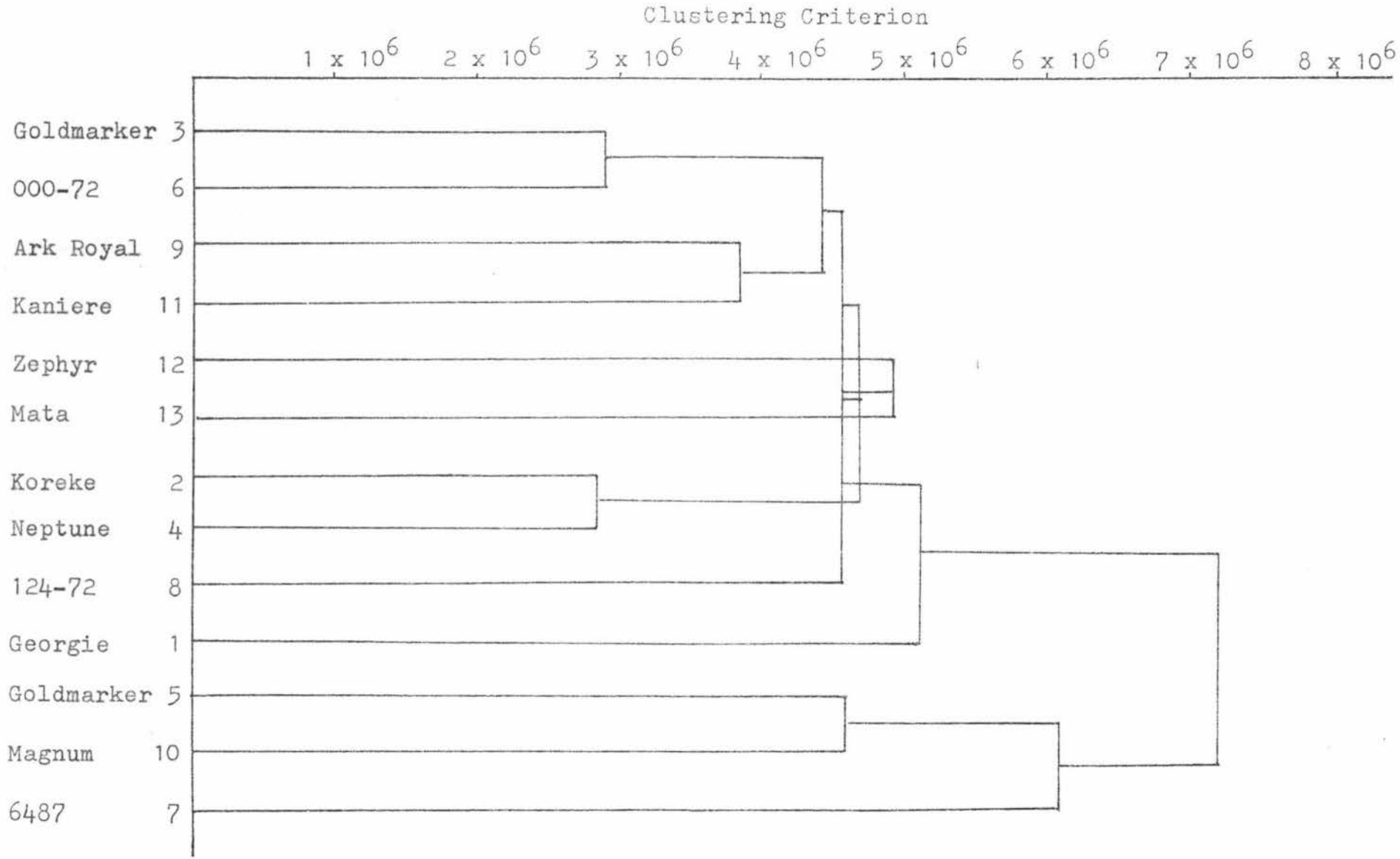


Figure 3.53 Dendrogram for Median Clustering of 13 Barley Genotype Means

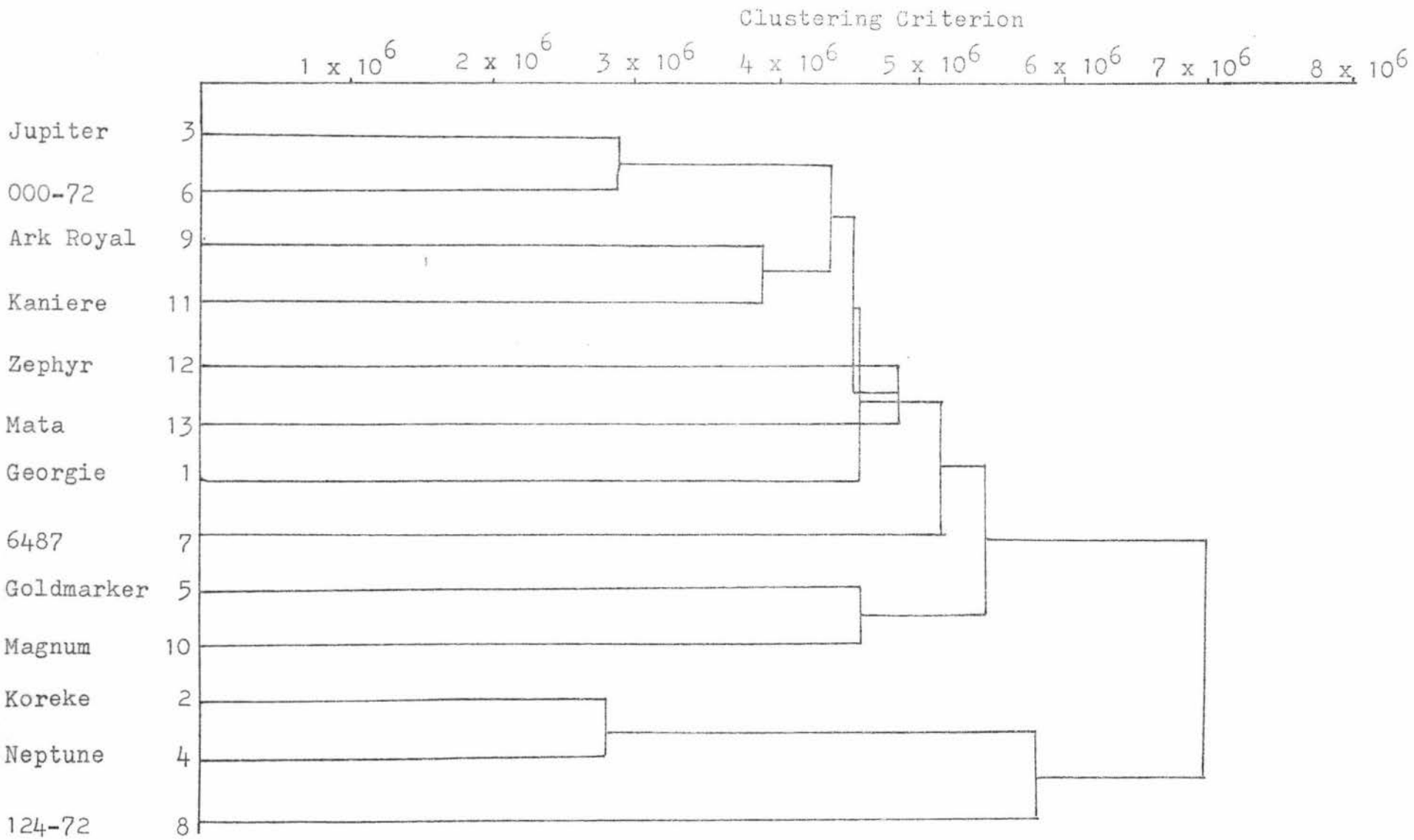


Figure 3.54 Dendrogram for Group Average Clustering of 13 Barley Genotype Means

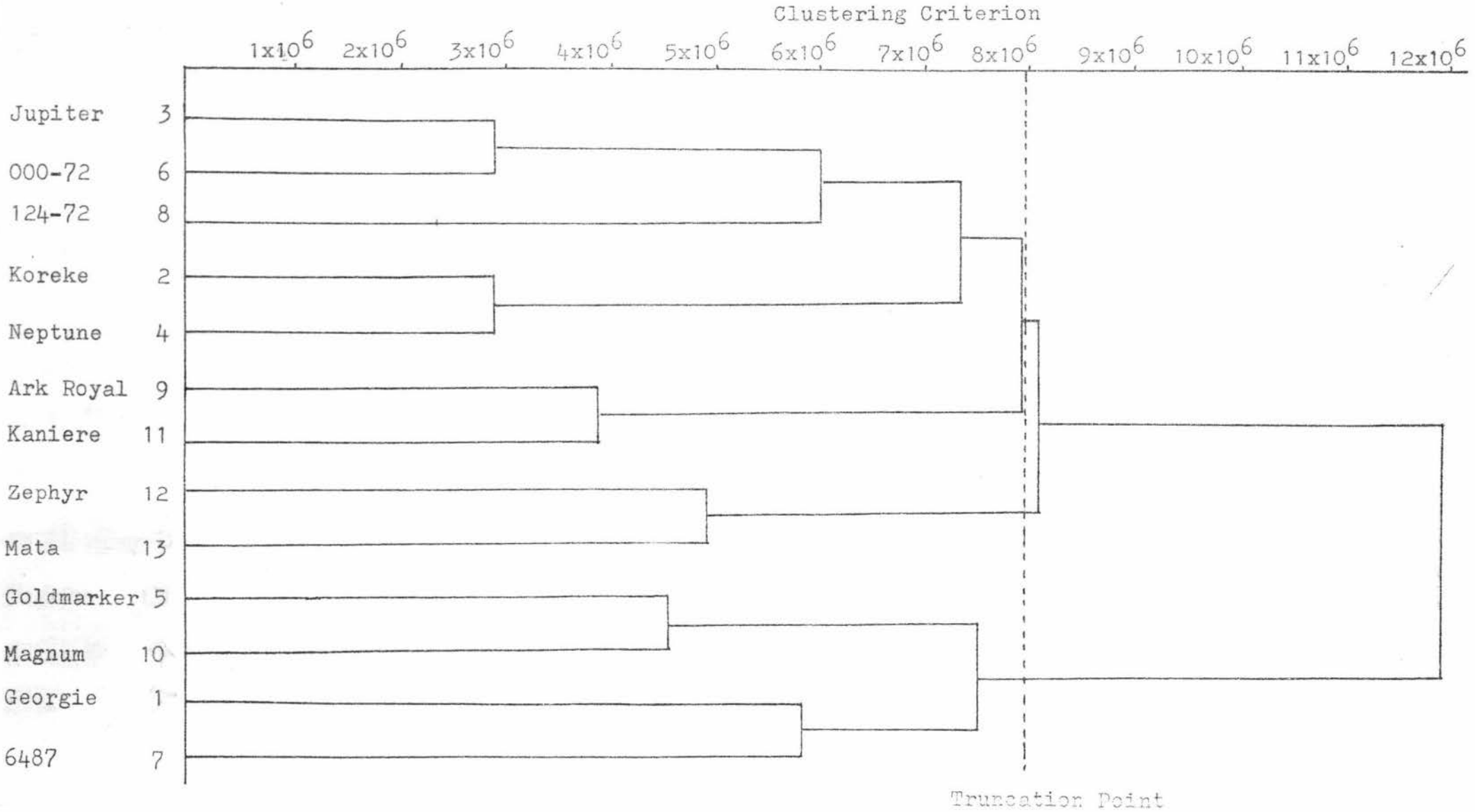


Figure 3.55 Dendrogram for Average Within Group Clustering of 13 Barley Genotype Means

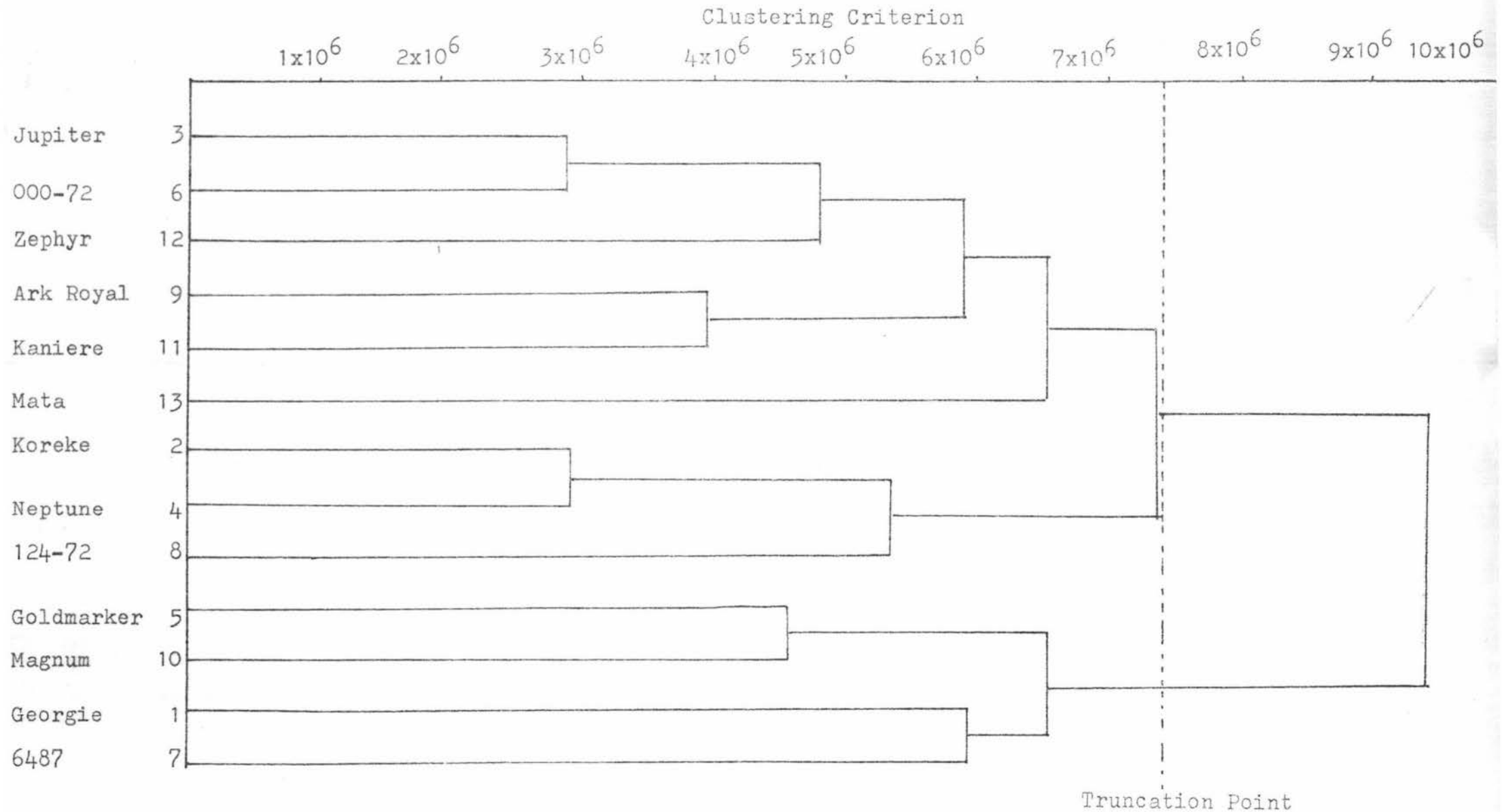


Figure 3.56 Dendrogram for Complete Linkage Clustering of 13 Barley Genotype Means

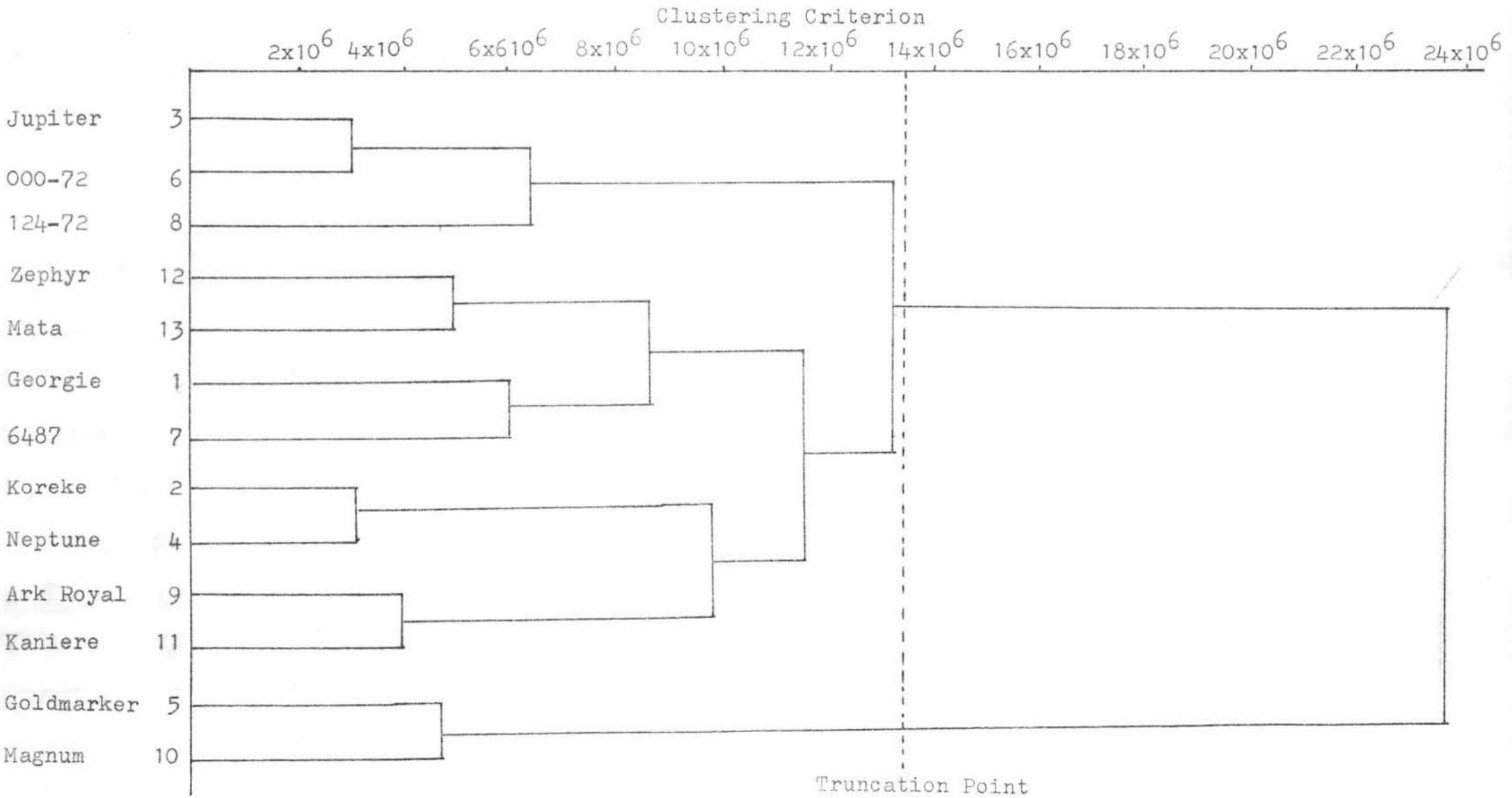


Figure 3.57 Dendrogram for Wards Clustering of 13 Barley Genotype Means

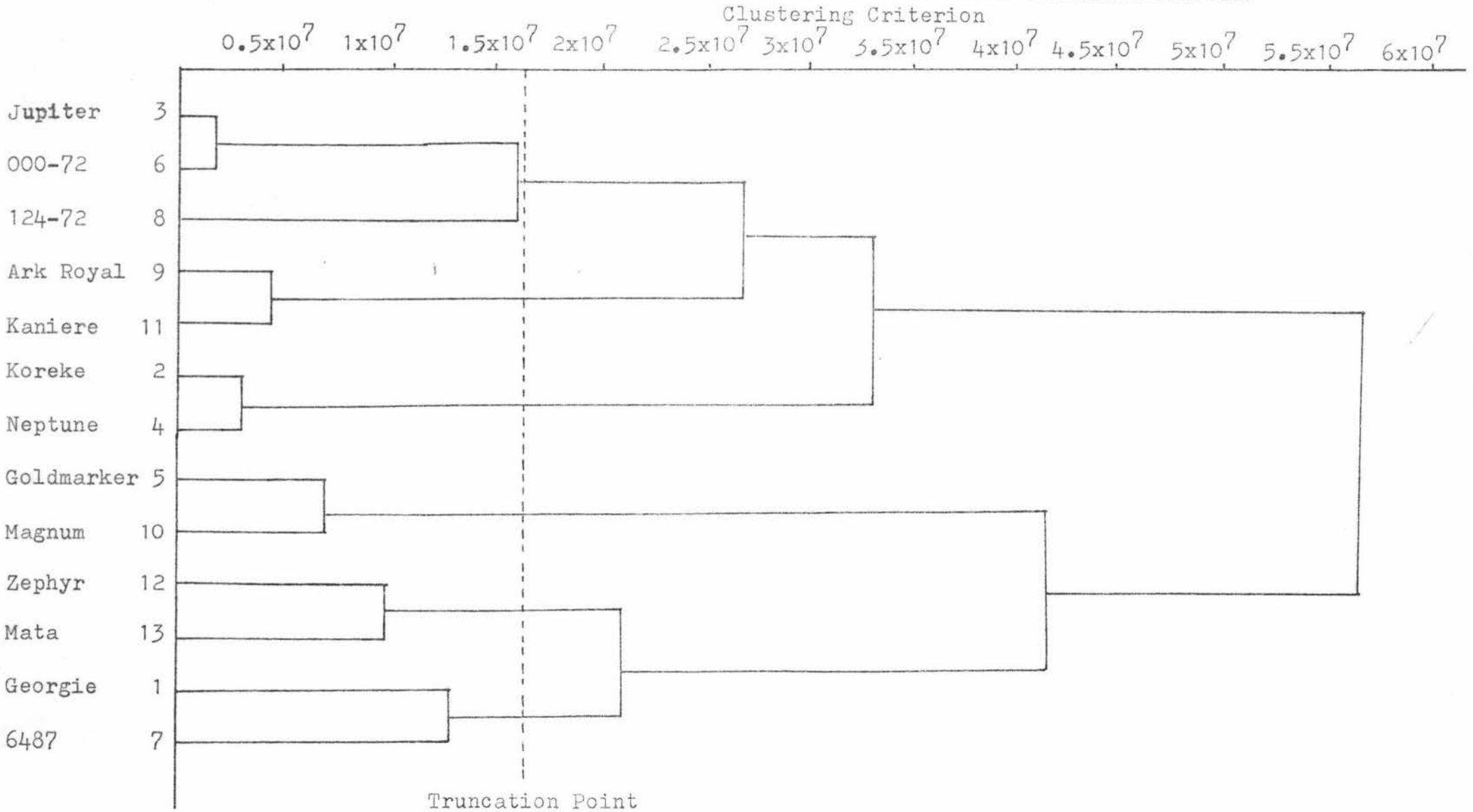
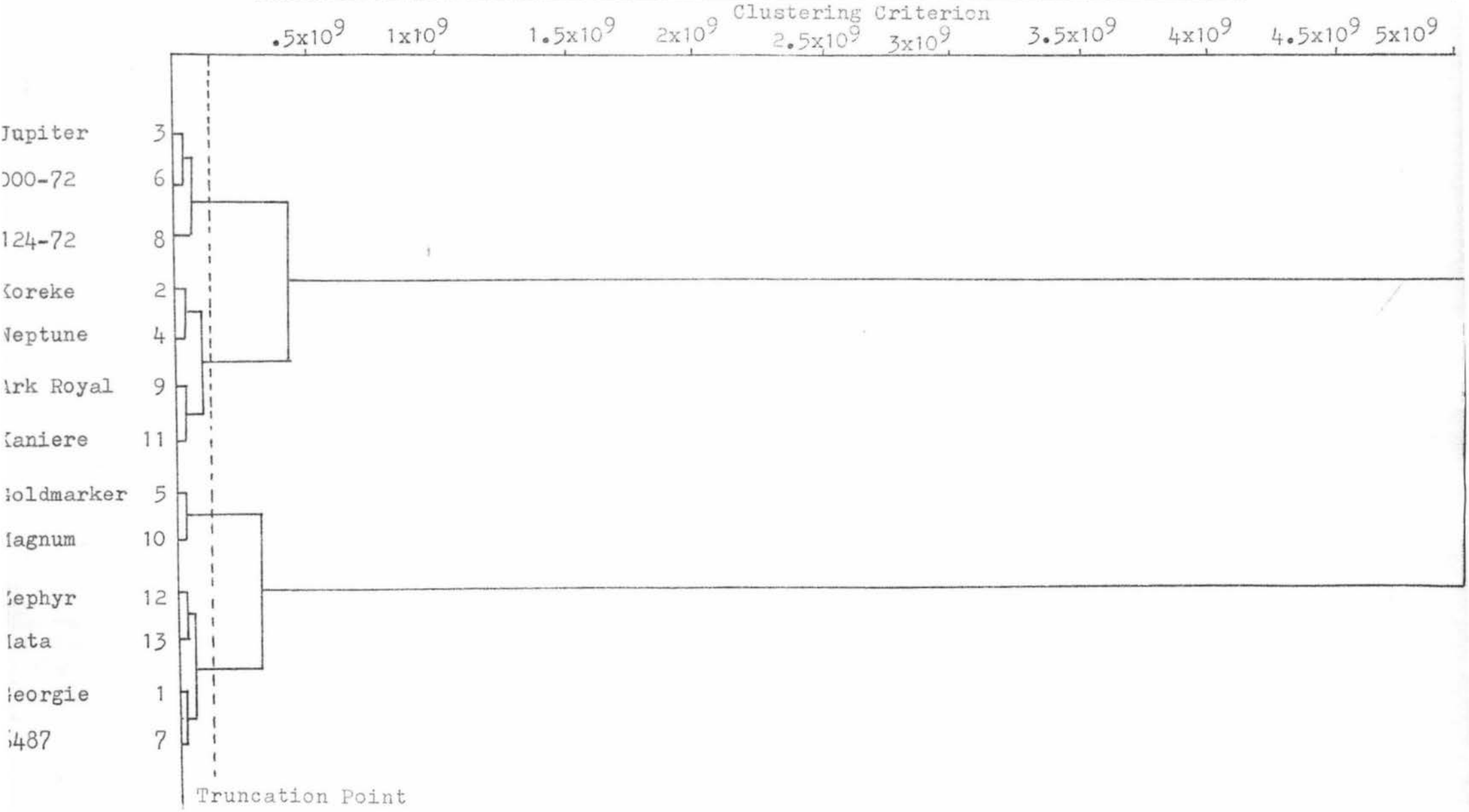


Figure 3.58 Dendrogram for Flexible Clustering of 13 Barley Genotype Means



The truncation points have been marked on each dendrogram. Of these five methods Wards and Flexible strategies produced the most even sized clusters and in addition the cluster memberships produced by these two strategies were quite similar.

Because of the chaining and reversal problems of the first three methods, and the desirability of having even sized clusters, Wards and Flexible strategies would appear to be the most suitable. Previous reports of the successful application of Wards Incremental Sum of Squares strategy (Teow, 1978), (Cameron, 1979) and the conceptually attractive basis of this method meant that it was chosen as the strategy to be used for the remaining data sets.

(ii) Barley 13 genotypes

1. Clustering on means.

The results of the Cluster Analysis of Genotype-Environment means by Wards method are presented in Tables 3.51 and 3.52 and figure 3.57.

Using the method described by Teow (1978) the truncation point of the dendrogram was determined at the 6 cluster level and the membership and means for each attribute (environment) for these clusters is reported in Table 3.51.

An examination of WSS/TSS ratio (Table 3.52) gives an indication of the proportion of the total sums of squares of each attribute not explained by cluster differences. The truncation point is marked in the table and at this point dormant and dominant characters (environments) can be identified. At this point, Winfield had only 15.8% of its Total Sums of Squares (TSS) not accounted for by clustering. This implies that it was a dominant attribute in the cluster formation. Similarly Outram, Bulls and Methven were relatively dominant characters.

Dormant environments were Pleasant Point, Leeston MAF, Mitcham, and Leeston CRD, i.e. a small proportion of their TSS was accounted for by clustering.

The significance testing of the means of the clusters Table 3.51 indicated that Cluster 4, whose members are Goldmarker and Magnum, had a significantly higher mean than the other clusters for 6 out of the 17 sites. These 6 sites were all relatively dominant characters. These two cultivars had been previously noted as having a high mean yield and low stability rankings - section 3.2.

Barley 15 g. means

Table 3.51 Cluster Means for Each Environment

Letters indicate significance grouping of Means, by LSD, between clusters.

Environments	1.	2.	3.	4.	5.	6.
1. Gore	7892.2^{a}	7671.7^{a}	7990.0^{a}	8788.3^{a}	7663.3^{a}	8840.0^{a}
2. Pleasant Pt	5103.3^{a}	5160.0^{a}	5091.7^{a}	4921.7^{a}	4791.7^{a}	4816.7^{ab}
3. Fairview	5133.3^{b}	5901.7^{ab}	5576.7^{ab}	6233.3^{a}	5691.6^{ab}	5620.0^{ab}
4. Outram	6253.3^{ab}	6590.0^{a}	6336.7^{a}	6570.0^{a}	4905.0^{b}	6260.0^{ab}
5. Takapau	5175.6^{a}	4795.0^{a}	4688.3^{a}	4440.0^{a}	5060.0^{a}	5348.3^{a}
6. Bulls	5385.6^{ab}	4973.3^{b}	5718.3^{ab}	6925.0^{a}	5861.7^{ab}	6280.0^{ab}
7. Marton	4990.0^{ab}	4980.0^{ab}	3935.0^{b}	5425.0^{a}	5375.0^{a}	5545.0^{a}
8. Winfield	4806.7^{b}	5208.3^{ab}	5328.3^{ab}	6745.0^{a}	5733.3^{ab}	6230.0^{ab}
9. Leeston MAF	2355.5^{a}	2576.7^{a}	2308.3^{a}	2406.7^{a}	2471.7^{a}	2353.3^{a}
10. Milton	6324.4^{a}	6515.0^{a}	4895.0^{a}	4963.3^{a}	6083.3^{a}	6115.0^{a}
11. Incholme	2755.6^{ab}	3865.0^{a}	2598.3^{b}	3143.3^{ab}	2813.3^{ab}	3553.3^{ab}
12. Mitcham	2495.3^{a}	2551.2^{a}	2355.5^{a}	2523.3^{a}	2570.7^{a}	2624.7^{a}
13. Methven	2746.7^{b}	4258.8^{ab}	3653.2^{ab}	4808.7^{a}	3825.5^{ab}	4157.8^{ab}
14. Kirwee	3133.9^{a}	3161.0^{a}	3123.7^{a}	3323.2^{a}	3328.2^{a}	3443.0^{a}
15. Lincoln	5632.8^{ab}	5718.7^{ab}	5280.0^{b}	6462.3^{a}	5718.7^{ab}	5555.3^{ab}
16. Leeston CRD	2922.3^{a}	2866.3^{a}	2880.3^{a}	2777.7^{a}	2915.5^{a}	3032.2^{a}
17. Dunsandel	4759.1^{a}	5009.2^{a}	4880.8^{a}	5176.8^{a}	4751.0^{a}	5299.7^{a}
	Jupiter	Ark Royal	Koreke	Gold-marker	Zephyr	Georgie
MEMBERSHIP	000-72 124-72	Kaniere	Neptune	Magnum	Mata	6487

Table 3.52

Barley 13 g Means Wards

Proportion of Variance Not Explained by Clustering (WSS/TSS)

No. of Clusters	Clusters fusing		Criterion	Environments 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
12	3	6	$.1447 \times 10^7$	.045	.075	.231	.004	.006	.005	.009	.010	.015	.001	.002	.350	.001	.034	.018	.453	.015
11	2	4	$.2895 \times 10^7$	.112	.154	.251	.033	.088	.023	.017	.012	.131	.012	.003	.529	.031	.063	.035	.465	.060
10	9	11	$.4861 \times 10^7$	.117	.245	.253	.115	.309	.034	.023	.033	.219	.021	.004	.542	.058	.319	.138	.494	.088
9	5	10	$.7208 \times 10^7$	.136	.263	.253	.117	.429	.068	.023	.097	.403	.073	.092	.682	.078	.342	.209	.734	.122
8	12	13	$.9695 \times 10^7$	.155	.749	.287	.149	.441	.102	.068	.130	.566	.073	.137	.683	.126	.348	.214	.776	.134
7	1	7	$.1267 \times 10^8$	.224	.809	.297	.180	.444	.151	.265	.158	.762	.129	.214	.725	.128	.352	.240	.836	.169
6	3	8	$.1625 \times 10^8$	.230	.848	.300	.193	.446	.207	.294	.158	.780	.356	.267	.790	.216	.430	.241	.836	.364
5	1	12	$.2104 \times 10^8$	.593	.849	.303	.567	.485	.236	.300	.194	.811	.356	.421	.797	.230	.470	.254	.868	.740
4	3	9	$.2690 \times 10^8$	.608	.851	.608	.594	.565	.269	.300	.221	.941	.361	.836	.806	.854	.473	.259	.877	.808
3	2	3	$.3343 \times 10^8$	.621	.852	.619	.595	.639	.326	.619	.249	1.000	.762	.981	.899	.601	.475	.362	.878	.808
2	1	5	$.4172 \times 10^8$	.722	.862	.811	.860	.999	.484	.620	.361	1.000	.975	.982	.918	.716	.491	.802	.998	.846
1	1	2	$.5682 \times 10^8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

2. Clustering on Effects

The results for the Cluster Analysis by Wards Method on the G x E effects are reported in Tables 3.53, 3.54 and figure 3.59. Five clusters were indicated.

The dominant environments involved in the formation of these clusters can be identified as Milton, Bulls and Outram.

The environments with the least proportion of their TSS involved in clustering were Pleasant Point, Gore and Incholme.

There was a degree of correspondence between the clustering of means and effects. In both cases Goldmarker and Magnum were grouped together by themselves as were Koreke and Neptune. Jupiter and 000-72, and Ark Royal and Kaniere remained together in both analyses.

Bulls and Outram were dominant environments in both analyses.

The cluster containing Goldmarker and Magnum, cluster 5, showed significantly different means in five out of the seventeen environments. This could reflect the relatively high level of instability noted earlier.

Table 3.53 Cluster Means for Each Environment

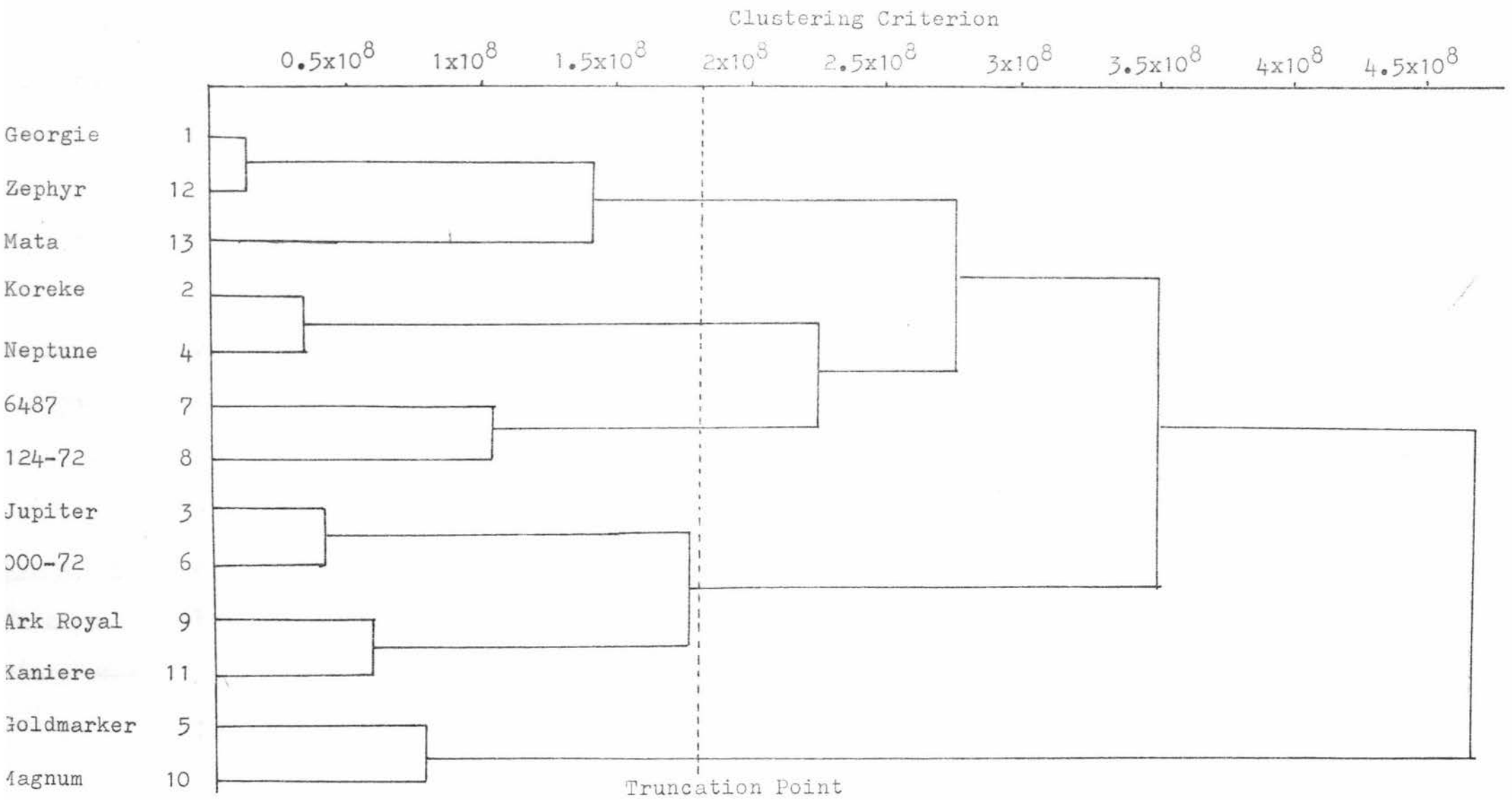
Environments	Clusters 1	2	3	4	5
1. Gore	22.5[a]	116.3[a]	83.2[a]	-309.4[a]	385.6[a]
2. Pleasant Pt	-146.0[a]	349.5[a]	-80.3[a]	149.6[a]	-349.5[a]
3. Fairview	22.3[a]	175.0[a]	-334.8[a]	-88.2[a]	302.7[a]
4. Outram	-928.2[b]	424.5[a]	91.4[ab]	373.8[a]	128.8[ab]
5. Takapau	167.7[ab]	-1.3[ab]	338.9[a]	94.7[ab]	-778.6[b]
6. Bulls	18.8[ab]	145.4[ab]	403.9[ab]	-700.3[b]	823.0[a]
7. Marton	128.8[ab]	-854.6[b]	692.3[a]	-68.6[ab]	106.4[ab]
8. Winfield	363.2[a]	-31.3[a]	-284.4[a]	-542.7[a]	856.4[a]
9. Leeston MAF	62.1[a]	148.7[a]	-236.1[a]	138.1[a]	-282.0[a]
10. Milton	364.7[ab]	-712.0[ab]	-466.8[ab]	902.3[a]	-1172.7[b]
11. Incholme	-189.1[a]	-246.9[a]	381.6[a]	189.9[a]	230.9[a]
12. Mitcham	7.3[a]	85.4[a]	112.4[a]	33.5[a]	-275.8[a]
13. Methven	110.8[ab]	82.2[ab]	-784.6[b]	-86.2[ab]	708.7[a]
14. Kirwee	82.7[a]	128.7[a]	81.1[a]	-66.5[a]	-200.8[a]
15. Lincoln	-34.4[a]	-192.5[a]	-224.4[a]	3.9[a]	460.8[a]
16. Leeston CRD	55.7[ab]	224.7[a]	-28.4[ab]	63.6[ab]	-407.0[b]
17. Dunsandel	-108.7[a]	158.1[a]	255.0[a]	-87.5[a]	-74.9[a]
Membership	Georgie Zephyr Mata	Koreke Neptune	6487 124-72	Jupiter 000-72 Ark Royal Kaniere	Goldmarker Magnum

Letters indicate significant groupings by LSD between environments.

Table 3.54 Proportion of Variance Not Explained by Clustering (WSS/TSS)

No. of clusters	I	J	Criterion	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
12	1	12	$.1196 \times 10^7$	.172	.043	.015	.014	.002	.001	.060	.000	.017	.000	.009	.178	.015	.212	.049	.006	.004
11	2	4	$.2579 \times 10^7$	.253	.081	.030	.035	.089	.018	.082	.000	.134	.006	.009	.356	.048	.284	.092	.023	.147
10	3	6	$.4007 \times 10^7$	.314	.143	.372	.041	.098	.023	.098	.021	.137	.008	.015	.548	.049	.300	.111	.232	.160
9	9	11	$.5700 \times 10^7$	.314	.255	.375	.096	.217	.027	.099	.035	.144	.027	.027	.553	.114	.374	.191	.232	.335
8	5	10	$.7771 \times 10^7$	.322	.292	.388	.108	.271	.053	.104	.099	.179	.058	.214	.573	.128	.379	.377	.420	.339
7	7	8	$.1083 \times 10^8$	.323	.549	.391	.109	.282	.070	.146	.153	.388	.058	.246	.657	.370	.440	.553	.496	.352
6	1	13	$.14078 \times 10^8$	.635	.737	.425	.241	.283	.104	.258	.278	.439	.063	.260	.657	.394	.459	.557	.497	.372
5	3	9	$.1759 \times 10^8$	.700	.739	.650	.241	.412	.142	.260	.284	.442	.104	.696	.679	.561	.519	.569	.569	.416
4	2	7	$.2223 \times 10^8$	.700	.813	.832	.265	.456	.158	.965	.299	.647	.111	.852	.680	.703	.526	.569	.629	.438
3	1	2	$.2775 \times 10^8$	.705	.867	.844	.780	.456	.184	.988	.403	.674	.289	.896	.703	.773	.528	.606	.632	.824
2	1	3	$.3539 \times 10^8$	.854	.885	.849	.992	.461	.627	.992	.613	.740	.629	.950	.706	.775	.722	.641	.632	.970
1	1	5	$.4702 \times 10^8$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Figure 3.59 Dendrogram for Wards Clustering of 13 Barley Genotype Effects



(iii) Barley 14 genotypes

1. Clustering on means.

The results for clustering of means by Wards Method are presented in Tables 3.55 and 3.56, and figure 3.510.

The probabilistic cut off method indicated that four clusters be considered.

From Table 3.56 it can be seen that the dominant attributes involved in clustering were Bulls and Milton with Winfield also having a relatively low percentage of its TSS not accounted for by clustering.

The dormant characters could be identified as Takapau, Pleasant Point and Leeston MAF.

The cluster membership reflected some similarities with the 13 genotype analysis, with such pairs of genotypes as Goldmarker and Magnum, Koreke and Neptune, Jupiter and 000-72 being grouped together still.

The cluster of which Goldmarker and Magnum were members had significantly high means for five out of the eleven environments.

The dormant attributes exhibited no significant differences for cluster means.

Table 3.55 Cluster Means for Each Environment

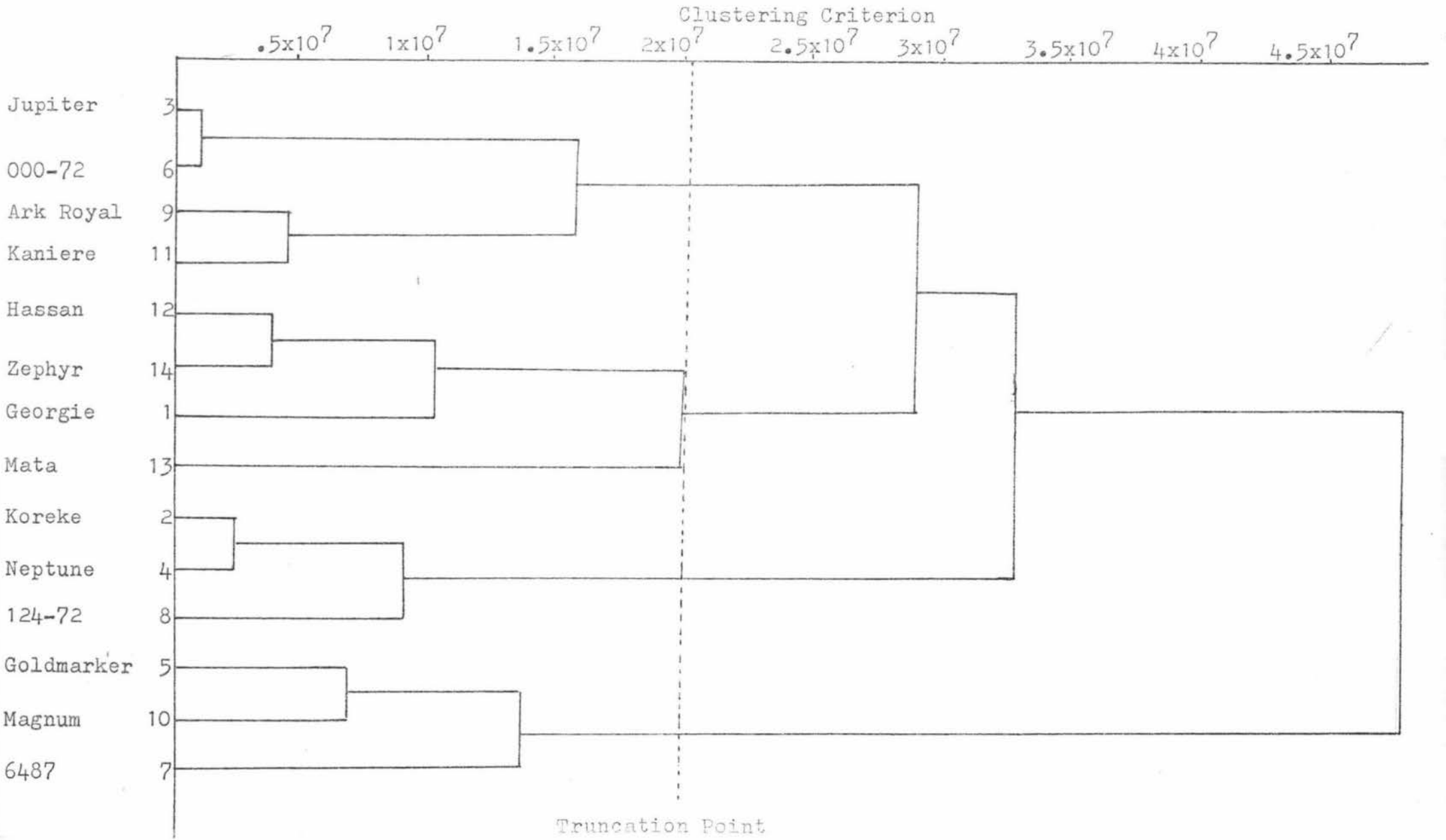
Environments	Clusters 1	2	3	4
1. Gore	7750.8[b]	8123.3[a]	7998.9[ab]	8684.5[a]
2. Pleasant Pt	5078.3[a]	4907.5[a]	5166.7[a]	4810.0[a]
3. Fairview	5500.0[a]	5703.3[a]	5452.2[a]	5992.2[a]
4. Outram	6472.5[a]	5502.5[b]	6241.1[ab]	6558.9[a]
5. Takapau	4970.8[a]	5081.7[a]	4870.0[a]	4761.1[a]
6. Bulls	5059.2[b]	5867.5[a]	5767.8[ab]	6838.9[a]
7. Marton	4907.5[ab]	5290.0[a]	4390.0[b]	5697.8[a]
8. Winfield	5003.3[b]	5945.0[ab]	5156.7[ab]	6468.9[a]
9. Leeston MAF	2484.2[a]	2435.0[a]	2300.0[a]	2318.9[a]
10. Milton	6695.8[a]	6290.0[ab]	5003.3[b]	5188.9[ab]
11. Incholme	3221.7[a]	2881.7[a]	2768.9[a]	3403.3[a]
Cluster Membership	Jupiter 000-72 Ark Royal Kaniere	Hassan Zephyr Georgie Mata	Koreke Neptune 124-72	Goldmarker Magnum 6487

Letters indicate means grouping by LSDs between environments.

Table 3.56 Proportion of Variance Not Explained by Clustering (WSS/TSS)

No. of Clusters	I	J	Criterion	1	2	3	4	5	6	7	8	9	10	11
13	3	6	$.10435 \times 10^7$	.045	.075	.231	.004	.006	.005	.008	.010	.014	.001	.002
12	2	4	$.21060 \times 10^7$	.111	.154	.251	.033	.088	.023	.017	.012	.124	.011	.003
11	12	14	$.3575 \times 10^7$	.113	.365	.260	.141	.088	.032	.043	.018	.126	.017	.008
10	9	11	$.47580 \times 10^7$	.118	.457	.262	.224	.309	.042	.048	.039	.208	.026	.008
9	5	10	$.66045 \times 10^7$	.137	.475	.262	.226	.429	.076	.048	.103	.381	.077	.093
8	2	8	$.86633 \times 10^7$	.137	.494	.302	.237	.520	.079	.287	.128	.382	.085	.141
7	1	12	$.10863 \times 10^8$	.428	.581	.307	.248	.561	.083	.314	.166	.513	.095	.197
6	5	7	$.13676 \times 10^8$	.445	.623	.457	.248	.847	.090	.400	.252	.610	.131	.307
5	3	9	$.19567 \times 10^8$	.451	.639	.735	.259	.904	.095	.404	.255	.682	.147	.756
4	1	13	$.19567 \times 10^9$	.598	.857	.755	.468	.910	.117	.448	.313	.827	.154	.773
3	1	3	$.25830 \times 10^8$	.670	.891	.790	.850	.921	.329	.505	.570	.837	.193	.835
2	1	2	$.32736 \times 10^8$	.672	.929	.811	.879	.946	.362	.715	.601	.953	.773	.883
1	1	5	$.45797 \times 10^8$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Figure 3.510 Dendrogram for Wards Clustering of 14 Barley Genotype Means



2. Clustering on effects

These results are found in Tables 3.57 and 3.58, and figure 3.511. A relatively large number of clusters were indicated, i.e. eight, and of these three had single genotype memberships.

Again previously associated pairs of genotypes were grouped together i.e. Jupiter and 000-72, Koreke and Neptune, Hassan and Zephyr. Goldmarker and Magnum, the cultivars which had previously been strongly associated were both in single genotype clusters.

Dominant characters involved in cluster formation were Bulls, Milton and Incholme. There were no really dormant characters because of the relatively early stage of truncation, however Pleasant Point had 52% of its TSS not accounted for by clustering at this stage.

Table 3.57 Cluster Means

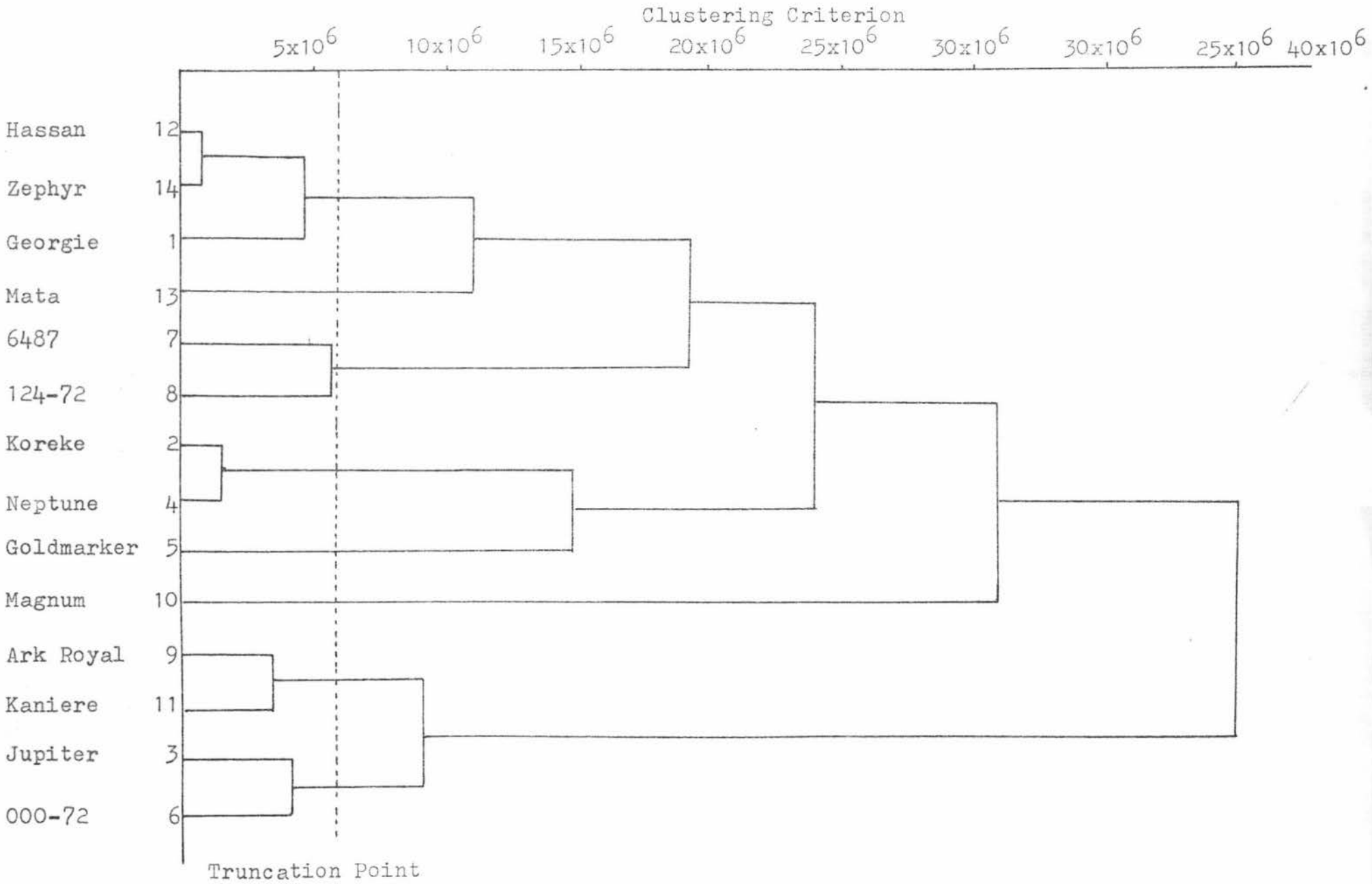
Environments	Cluster 1	2	3	4	5	6	7	8
1. Gore	193.7^{a}	-638.5^{a}	45.2^{a}	176.1^{a}	484.6^{a}	356.1^{a}	-484.5^{a}	-128.5^{a}
2. Pleasant Pt	-297.8^{a}	451.1^{a}	-130.3^{a}	397.3^{a}	-579.2^{a}	-74.4^{a}	123.3^{a}	157.7^{a}
3. Fairview	-48.0^{a}	235.3^{a}	-387.7^{a}	219.9^{a}	188.4^{a}	456.6^{a}	202.6^{a}	-403.0^{a}
4. Outram	-405.6^{ab}	-1538.9^{b}	38.0^{ab}	468.9^{a}	-55.9^{ab}	352.3^{a}	379.9^{a}	342.7^{a}
5. Takapau	78.3^{ab}	239.4^{a}	291.4^{a}	48.9^{ab}	-520.9^{ab}	-986.1^{b}	-186.7^{ab}	362.7^{a}
6. Bulls	-97.9^{ab}	362.0^{a}	352.3^{a}	191.6^{a}	1041.7^{a}	-1869.4^{b}	-895.7^{ab}	-526.3^{ab}
7. Marton	40.1^{ab}	630.1^{a}	605.4^{a}	-843.7^{b}	-36.9^{ab}	221.2^{ab}	-141.0^{ab}	88.2^{ab}
8. Winfield	472.7^{ab}	-223.9^{ab}	-345.3^{ab}	5.6^{ab}	1209.1^{a}	527.3^{ab}	-461.7^{b}	-663.9^{b}
9. Leeston MAF	-72.1^{ab}	266.8^{a}	-86.4^{ab}	209.7^{ab}	-170.2^{ab}	-322.0^{b}	135.7^{ab}	148.4^{ab}
10. Milton	428.8^{ab}	187.7^{ab}	-555.3^{ab}	-702.7^{ab}	-855.9^{ab}	-1521.1^{b}	474.2^{ab}	1134.4^{a}
11. Incholme	-292.2^{ab}	28.9^{ab}	359.2^{ab}	-171.6^{ab}	-704.7^{b}	343.5^{ab}	752.8^{a}	-336.1^{ab}
Cluster Membership	Hassan Zephyr Georgie	Mata	6487 124-72	Koreke Neptune	Goldmarker	Magnum	Ark Royal Kaniere	Jupiter 000-72

Letters indicate means grouping by LSDs between environments.

Table 3.58 Proportion of Variance Not Explained by Clustering (WSS/TSS)

No. of Clusters	Clusters Fusing I	J	Criterion	1	2	3	4	5	6	7	8	9	10	11
13	12	14	$.7730 \times 10^6$	.007	.063	.002	.061	.020	.000	.008	.038	.094	.000	.002
12	2	4	$.1663 \times 10^7$	.068	.086	.006	.075	.140	.007	.040	.038	.302	.003	.004
11	9	11	$.2677 \times 10^7$	.069	.210	.012	.124	.250	.008	.040	.048	.305	.014	.020
10	3	6	$.3719 \times 10^8$	.138	.264	.319	.128	.257	.012	.054	.065	.312	.015	.024
9	1	12	$.4884 \times 10^7$	.379	.266	.342	.130	.257	.015	.188	.074	.317	.015	.033
8	7	8	$.6123 \times 10^7$	.380	.522	.346	.130	.269	.025	.227	.125	.346	.015	.061
7	3	9	$.8723 \times 10^7$	.434	.522	.562	.130	.390	.045	.228	.134	.347	.067	.498
6	1	13	$.1170 \times 10^8$	.656	.684	.597	.342	.398	.067	.304	.214	.490	.072	.526
5	2	5	$.1498 \times 10^8$	.683	.930	.598	.383	.484	.137	.430	.424	.650	.074	.596
4	1	7	$.1848 \times 10^8$	.685	.930	.730	.538	.500	.158	.497	.545	.671	.210	.757
3	1	2	$.2384 \times 10^8$	.749	.957	.853	.779	.581	.193	.969	.591	.707	.368	.836
2	1	10	$.3138 \times 10^8$	.775	.957	.967	.839	.983	.772	.979	.613	.876	.552	.910
1	1	3	$.4035 \times 10^8$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Figure 3.511 Dendrogram for Wards Clustering of 14 Barley Genotype Effects



(iv) Wheat

1. Clustering on means

Tables 3.59 and 3.510 and figure 3.512 give the results for this analysis.

The truncation point was determined to be at the three cluster level.

Examination of the cluster means shows that the cluster 1, with 6 genotypes in it, ranked consistently in the middle for each environment.

The cluster containing Gamenya and 61,06 had significantly lower means for all of the environments while the cluster with 300-301 and Karamu/9 had significantly higher means.

The mean yields listed in Table 3.23 also indicated this trend so the clustering seems to account for differences in yield quite closely.

There was one outstanding dormant environment for this analysis. Tikokino 1978 had 81.5% of its TSS not accounted for by clustering.

Kairanga 1978 was the most dominant character but Tikokino 1977, Halcombe 1977 and Halcombe 1978 were also important.

Table 3.59

Cluster Means for Each Environment

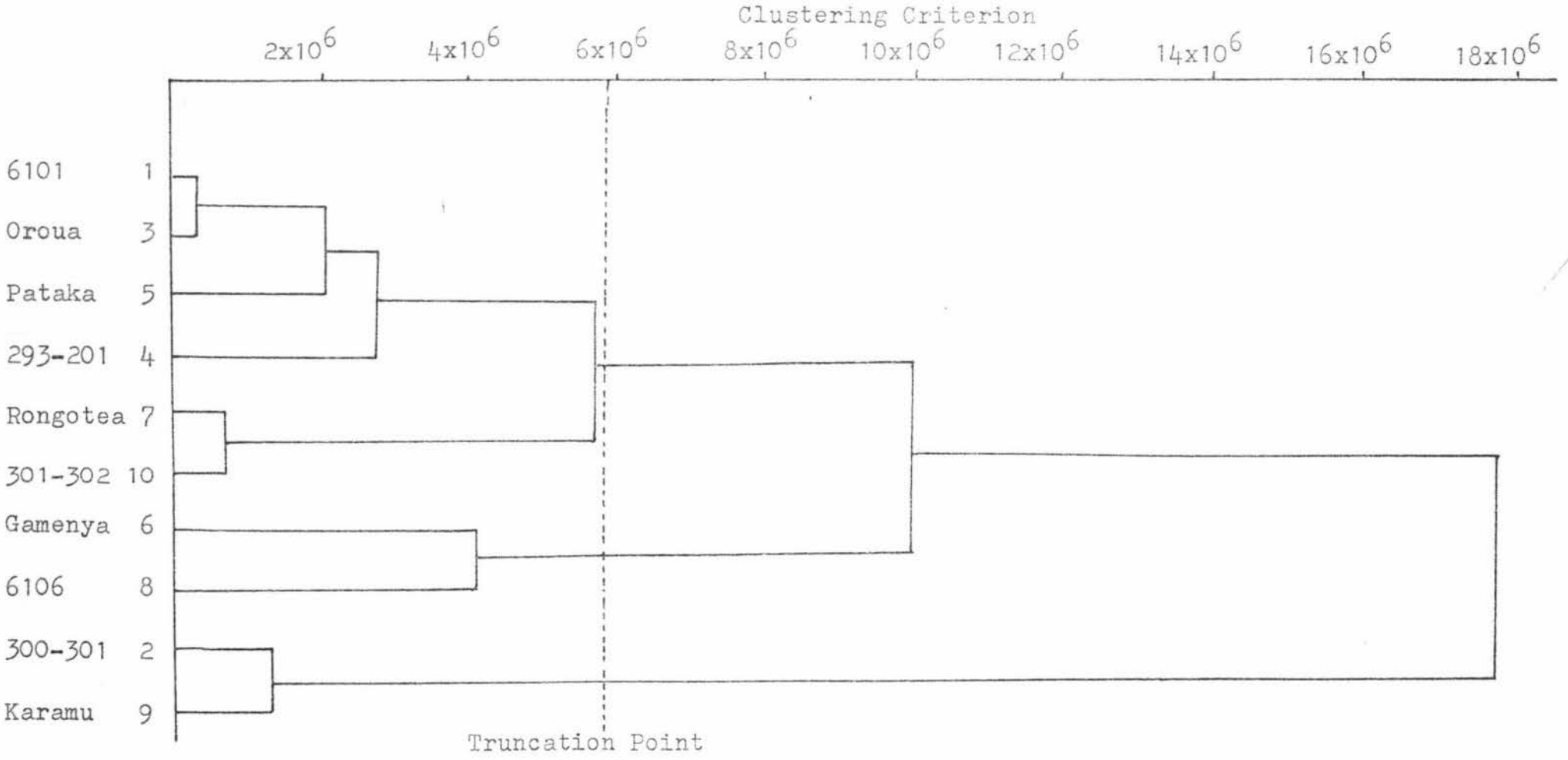
Environment	Cluster 1	Cluster 2	Cluster 3	Wheat Means
Halcombe '77	6002.1^{ab}	5344.5^{b}	6694.2^{a}	
Aorangi '77	7543.8^{ab}	6453.3^{b}	8202.2^{a}	
Kairanga '77	5816.6^{ab}	5334.5^{b}	6803.2^{a}	
Tikokino '77	5896.7^{ab}	5451.7^{b}	6931.2^{a}	
Halcombe '78	5293.9^{ab}	4923.0^{b}	5947.2^{a}	
Aorangi '78	5066.7^{ab}	4433.7^{b}	5155.7^{a}	
Kairanga '78	3628.1^{ab}	3171.7^{b}	3799.5^{a}	
Tikokino '78	2541.2^{ab}	2413.5^{b}	2718.7^{a}	
Cluster Membership	6101 Bulk Oroua Pataka 293-201 Rongotea 301-302	Gamenya 6106	300-301 Karamu/9	

Table 3.510 Proportion of Variance not Explained by Clustering (WSS/TSS)

Wheat Means Wards Method

No. of Clusters	Clusters Fusing I	J	Criterion	Hal '77	Aor '77	Kai '77	Tik '77	Hal '78	Aor '78	Kai '78	Tik '78
9	1	3	$.38095 \times 10^6$	0.019	0.007	0.017	0.039	0.060	0.005	0.027	0.005
8	7	10	$.79120 \times 10^6$	0.076	0.017	0.019	0.055	0.109	0.065	0.030	0.041
7	2	9	$.13937 \times 10^7$	0.103	0.018	0.033	0.129	0.121	0.082	0.030	0.449
6	1	5	$.20594 \times 10^7$	0.119	0.033	0.121	0.170	0.174	0.083	0.058	0.462
5	1	4	$.28881 \times 10^7$	0.127	0.129	0.155	0.174	0.260	0.098	0.076	0.567
4	6	8	$.40472 \times 10^7$	0.267	0.146	0.296	0.174	0.265	0.132	0.167	0.786
3	1	7	$.58267 \times 10^7$	0.276	0.343	0.369	0.220	0.287	0.421	0.167	0.815
2	1	6	$.10044 \times 10^8$	0.534	0.713	0.464	0.317	0.418	0.919	0.755	0.863
1	1	2	$.17851 \times 10^8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Figure 3.512 Dendrogram for Wards Clustering of Wheat Genotype Means



2. Clustering on effects

Tables 3.511 and 3.512 and figure 3.513 give the results for this analysis.

The cut off point was at the three cluster level again. Halcombe 1977 was the most noticeably dormant environment with Tikokino 1977 and Halcombe 1978 also being relatively dormant. Aorangi 1977, Aorangi 1978 and Tikokino 1978 were dominant environments.

Not a great deal of correspondence occurred with the cluster analysis on means either in respect to cluster membership or importance of environments.

However cluster 1 was again the middle ranking cluster for most environments.

Table 3.511

Cluster Means for Each Environment

Environment	Cluster 1	Cluster 2	Cluster 3
Halcombe '77	20.1[a]	-152.9[a]	56.3[a]
Aorangi '77	-49.8[ab]	-639.6[b]	369.6[a]
Kairanga '77	20.1[ab]	-467.6[b]	213.7[a]
Tikokino '77	-183.4[a]	4.8[a]	181.0[a]
Halcombe '78	-101.5[a]	214.7[a]	-5.9[a]
Aorangi ''78	235.2[a]	218.7[a]	-344.5[a]
Kairanga '78	34.0[ab]	305.0[a]	-186.5[b]
Tikokino '78	25.2[ab]	517.1[a]	-283.7[b]
Cluster Membership	61,01 bulk 6106 Oroua Pataka	Gamenya	300-301 301-302 Rongotea Karamu/9

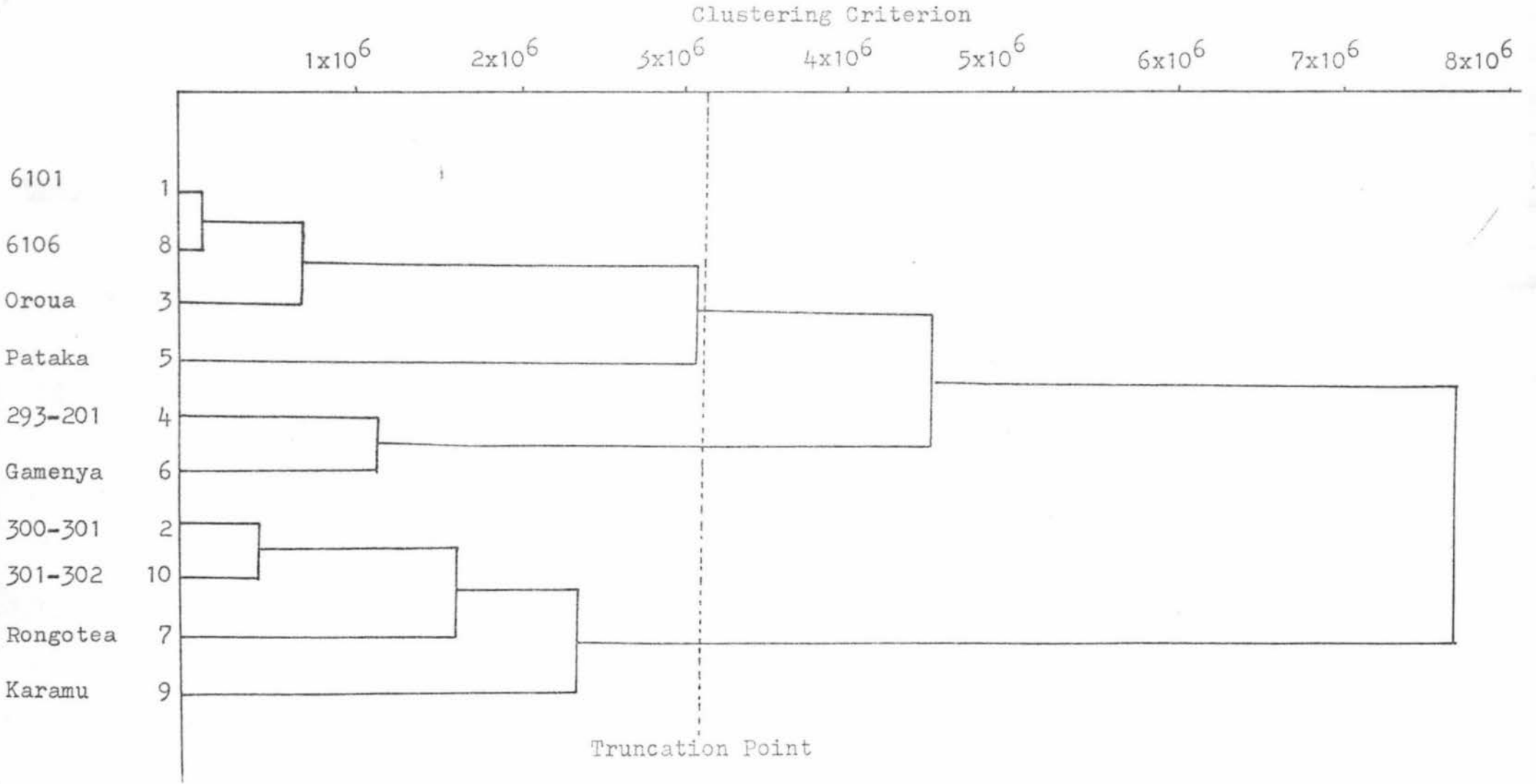
Letters indicate significance groupings of cluster means by LSD.

Table 3.512 Proportion of Variance Not Explained by Clustering (WSS/TSS)

Wheat Effects

No. of Clusters	Clusters fusing I	J	Criterion	Hal '77	Aor '77	Kai '77	Tik '77	Hal '78	Aor '78	Kai '78	Tik '78
9	1	8	$.1593 \times 10^6$	0.058	0.033	0.008	0.003	0.000	0.003	0.061	0.018
8	2	10	$.4225 \times 10^6$	0.081	0.034	0.027	0.005	0.010	0.058	0.091	0.136
7	1	3	$.7205 \times 10^6$	0.135	0.042	0.049	0.081	0.081	0.142	0.135	0.156
6	4	6	$.1151 \times 10^7$	0.421	0.051	0.051	0.099	0.092	0.234	0.181	0.248
5	2	7	$.1672 \times 10^7$	0.787	0.113	0.053	0.140	0.321	0.277	0.241	0.260
4	2	9	$.2306 \times 10^7$	0.849	0.169	0.113	0.523	0.354	0.307	0.327	0.292
3	1	5	$.3133 \times 10^7$	0.885	0.251	0.476	0.656	0.639	0.317	0.331	0.297
2	1	4	$.4555 \times 10^7$	0.960	0.504	0.744	0.718	0.999	0.317	0.530	0.561
1	1	2	$.7571 \times 10^7$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3.513 Dendrogram for Wards Clustering of Wheat Genotypes Effects



3.6 Principal Component Analysis

(i) Barley: 13 genotypes

For the unrotated analysis, Table 3.60 presents the eigenvalues and cumulative percentage for each component, Table 3.61 presents the Factor Structure matrix and Table 3.62 the factor scores.

For the rotated analysis Table 3.63 presents the factor structure matrix and Table 3.64 the factor scores.

It was found that for this set of data the correlation matrix was singular and therefore could not be inverted.

The inversion of the correlation matrix is not essential for the subsequent calculations for Principal Component Analysis (Cooley and Lohnes, 1971). However the determinant of the correlation matrix is needed for the chi-square test for the significance of the components. This may be calculated by multiplying together all of the eigenvalues (Bartlett, 1950), but in this case 5 eigenvalues equalled 0.0000 making the determinant also 0 and the test inapplicable.

It was thought that the singularity was caused by the number of attributes (17) exceeding the number of observations per attribute (13). This was examined by removing the five environments (attributes) which had the lowest communalities and therefore contributed least to the components. With these five removed the number of attributes was less than the number of observations and it was found that the correlation matrix was no longer singular. It was found that at least five

Table 3.60 Principal Component Analysis Barley 13 Genotypes

Component	Eigenvalue	% of variation	Cumulative %
1	5.47746	32.2	32.2
2	2.76262	16.3	48.5
3	1.88862	11.1	59.6
4	1.72453	10.1	69.7
5	1.29954	7.6	77.4
6	1.08207	6.4	83.7
7	0.88126	5.2	88.9
8	0.76377	4.5	93.4
9	0.53041	3.1	96.5
10	0.31928	1.9	98.4
11	0.21067	1.2	99.6
12	0.05977	0.4	100.0
	rest 0.00000		

Table 3.61 Factor Structure Matrix

	Component 1	Component 2	Component 3	Component 4	Component 5
Gore	0.71494	-0.18976	-0.27467	0.12660	0.09229
Pleasant Pt	-0.32423	-0.38657	0.56280	0.22512	0.13345
Fairview	0.74822	-0.33767	0.30038	-0.39204	-0.03591
Outram	0.13536	-0.39525	0.03965	0.31272	-0.41060
Takapau	-0.09247	0.90532	-0.09011	0.17839	0.07216
Bulls	0.81869	-0.12068	-0.30673	0.15294	0.16096
Marton	0.54354	0.48324	0.22778	0.36572	0.03046
Winfield	0.83019	-0.09434	-0.18026	-0.03129	0.38016
Leeston MAF	0.15376	0.38060	0.52982	-0.53458	0.24426
Milton	-0.37791	0.46337	0.43421	-0.20229	0.08022
Incholme	0.37344	0.03899	0.67944	0.46245	-0.31796
Mitcham	0.51852	0.65723	0.02277	-0.27431	-0.41903
Methven	0.65364	-0.36682	0.40622	-0.20898	0.16352
Kirwee	0.76125	0.51871	-0.20328	0.10431	0.13724
Lincoln	0.61658	-0.12706	0.16093	-0.08484	0.27918
Leeston CRD	-0.31434	0.12256	0.20155	0.65830	0.57261
Dunsandel	0.74771	0.05035	0.11743	0.32546	-0.36413

Table 3.62 Factor Scores: 13 Barley Genotypes

	Component 1	Component 2	Component 3	Component 4	Component 5
Georgie	0.643788	0.525747	0.255464	0.405274	1.306019
Koreke	-0.584618	-0.087699	-0.869497	-0.397332	-0.358118
Jupiter	-0.622711	0.508553	-0.411969	-1.515196	-1.067782
Neptune	-0.747270	-2.065158	-0.868615	-0.015358	-0.067044
Goldmarker	1.929766	-0.318062	-0.411482	-1.176750	0.099638
000-72	-1.867604	0.116778	0.060806	1.102198	1.237173
6487	1.078185	1.015331	-0.409525	1.835393	-1.290971
124-72	-0.423015	0.628068	-0.563072	1.092318	-0.868777
Ark Royal	-0.569793	-0.835443	2.027480	-0.413692	-1.369084
Magnum	1.160939	-1.783501	0.133369	0.849704	0.669084
Kaniere	0.194435	0.669151	1.381211	-0.206185	-0.351985
Zephyr	-0.184124	0.772478	-1.486467	-0.963743	0.534855
Mata	-0.007978	0.853757	1.162299	-0.596632	1.525727

Table 3.63 Varimax Rotated Factor Structure Matrix

13 Barley Genotypes

	Component 1	Component 2	Component 3	Component 4	Component 5
Gore	0.62775	0.40028	0.01579	-0.22545	0.02363
Pleasant Pt	-0.04117	-0.75215	0.15725	0.06269	0.28761
Fairview	0.79482	-0.15829	0.21630	0.01899	-0.47471
Outram	0.03828	-0.09155	0.10629	-0.08027	-0.02968
Takapau	-0.37580	0.69719	0.19479	0.38504	0.28708
Bulls	0.65488	0.34721	0.21506	-0.52145	0.03044
Marton	0.20588	0.53806	0.73981	-0.02855	0.15507
Winfield	0.84874	0.40726	0.00858	-0.13114	0.06990
Leeston MAF	0.30884	0.05792	0.05751	0.80069	-0.13495
Milton	-0.26984	0.03849	-0.03124	0.82449	0.13393
Incholme	0.12984	-0.22559	0.88169	0.11520	0.05788
Mitcham	0.03737	0.58503	0.48778	0.20346	-0.55146
Methven	0.81876	-0.21072	0.18047	0.13947	-0.15556
Kirwee	0.42994	0.76688	0.36163	-0.04330	0.02817
Lincoln	0.68473	0.03433	0.17763	-0.02547	-0.01396
Leeston CRD	-0.12288	-0.10685	0.10106	0.06745	0.93556
Dunsandel	0.34200	0.24052	0.69256	-0.23163	-0.18253

Table 3.64 Factor Scores for Rotated Factors

	Component 1	Component 2	Component 3	Component 4	Component 5
Georgie	1.012474	0.899925	-0.156539	1.028158	1.333970
Koreke	-0.664143	0.107303	-0.691659	-0.522401	-0.506762
Jupiter	-0.846950	0.437449	-0.802041	0.800837	-1.516007
Neptune	-0.016298	-1.129772	-1.460422	-1.161005	-0.063896
Goldmarker	1.876050	0.839167	-0.252261	0.007871	-1.121978
000-72	-1.060925	-0.369844	-0.883719	0.549848	2.151928
6487	-0.550130	1.330916	1.945440	-0.963014	0.170071
124-72	-1.347374	0.304707	0.886346	-1.104624	0.067645
Ark Royal	-0.432926	-2.044375	0.964484	0.825215	-1.058696
Magnum	1.586808	-1.037191	0.379795	-1.506408	0.522476
Kaniere	0.107703	0.040287	0.730957	1.682712	-0.087567
Zephyr	-0.206109	1.192690	-1.190201	-0.256290	-0.352299
Mata	0.541820	-0.571262	0.529821	0.619111	0.461116

environments had to be removed to solve the problem, thereby supporting the above explanation. The results of the analysis with twelve environments are presented in Appendix 1. A comparison of the factor scores for both the full and restricted analyses reveals that the differences between them are minimal.

Because the chi-square test was not applicable here the number of components to be studied was decided on the basis of 75% of the total variation explained. The first five components satisfied this criterion (Table 3.60).

The factor structure matrices for both the unrotated (Table 3.61) and the varimax rotated (Table 3.63) analyses are given for these five components.

Component 1 accounted for 32.2% of the total variation and was by definition the most important component.

For the unrotated analysis Winfield and Bulls were the environments with the highest positive correlation with Component 1 although Kirwee, Fairview and Dunsandel were also highly correlated. Three sites: Pleasant Point, Milton and Leeston CRD had negative correlations with the component although these were not large.

For the varimax rotated analysis, Fairview and Winfield had the highest positive correlation with component, with Lincoln, Bulls and Gore also being important.

Takapau, Milton and Leeston CRD had small negative correlations. It has been often found that the first Principal component represents a general factor in the data (Cooley & Lohnes, 1971). This does not appear to be the case here as the

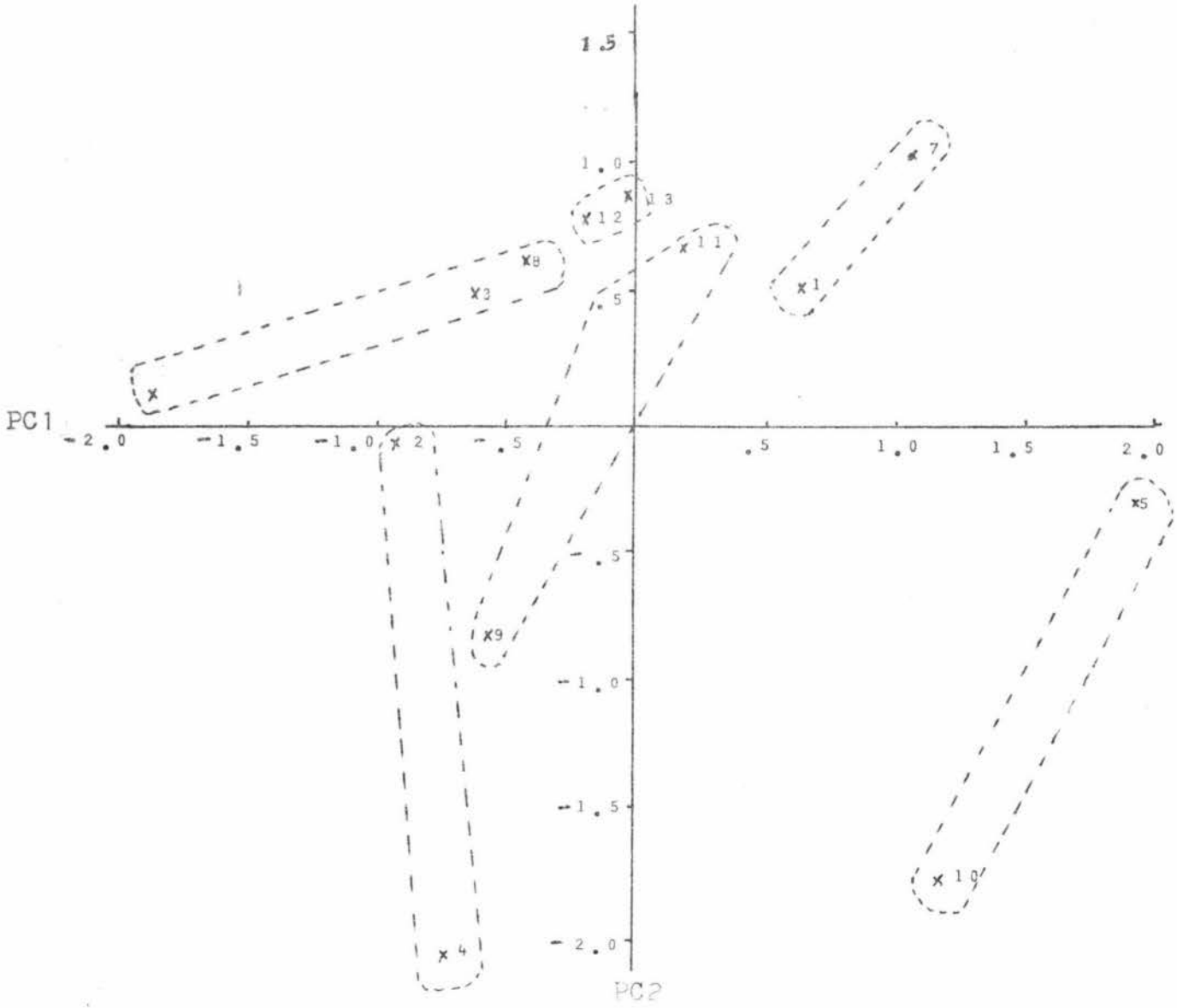
environments vary considerably in their correlations with the component. The varimax rotation indicated that Fairview and Winfield were especially important for this component. Both are Canterbury sites, but other factors which might single them out from the other environments were not obvious.

Component 2 accounted for 16.3% of the total variation taking its cumulative total with Component 1 to just under 50%. For the unrotated analysis Takapau had a very high positive correlation with this component and there were again some low negatively correlated sites i.e. Pleasant Point, Fairview, Outram and Methven.

The varimax rotated factor structure however suggests some degree of bipolarity of the environments with Kirwee and Takapau's high positive correlations being opposed by the high negative correlation of Pleasant Point. Again any interpretation of this feature was not obvious from the information available on these environments.

The genotypic scores for the components are represented in Tables 3.62 and 3.64, and graphs of the scores for Component 1 against Component 2 are presented in figures 3.61 and 3.62. The clusters formed by Wards Method on means are indicated on the unrotated scores. It is apparent from this that there is little correspondence between the grouping of the genotypes by cluster analysis and their arrangement on these two axes. Magnum and Goldmarker were grouped together by cluster analysis and here they are separated into a quadrant by themselves so there is some agreement there. Since the first two components only account for

Figure 3.61 Genotype Scores for components 1 and 2 for 13 Barley Genotypes - unrotated

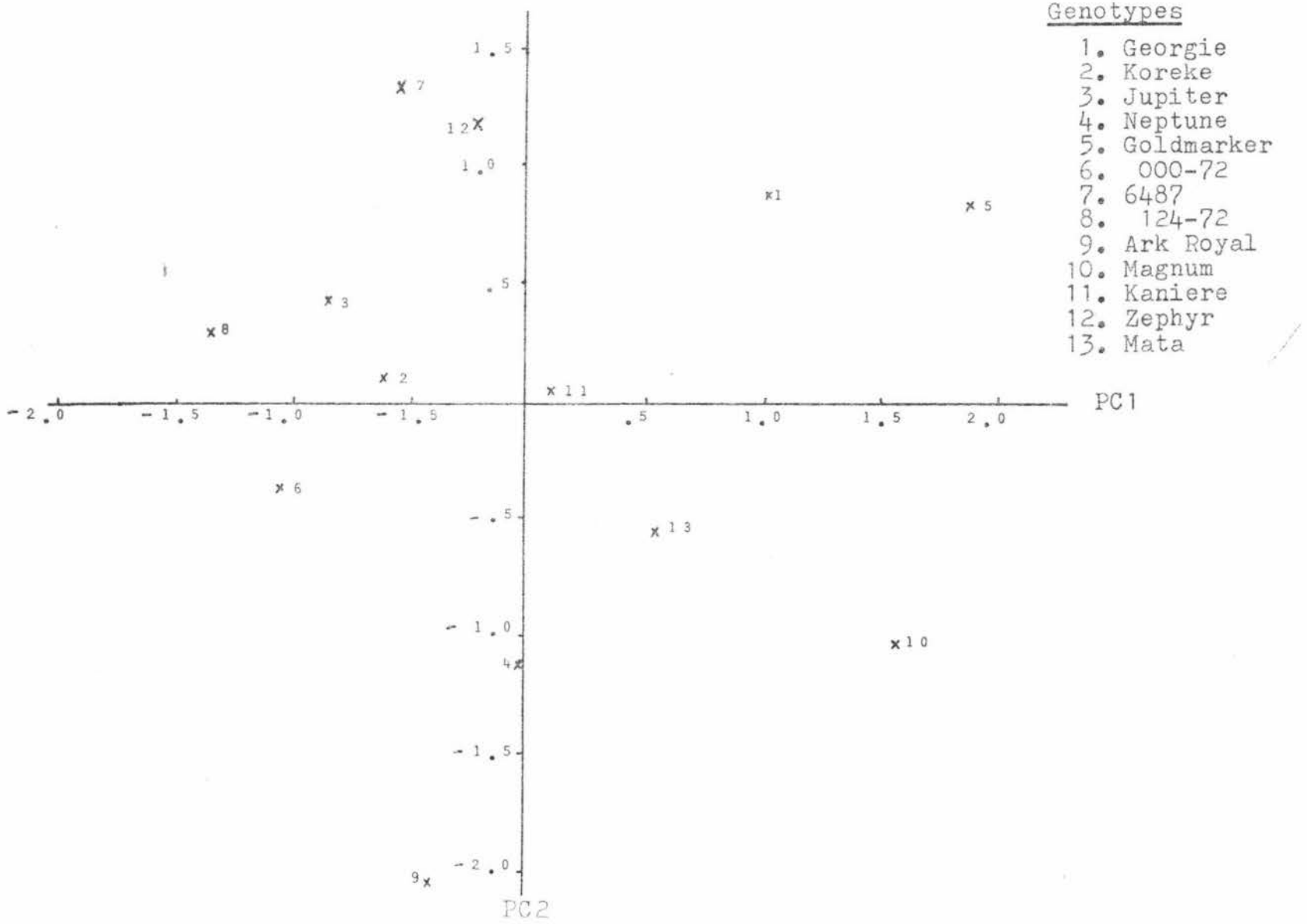


Genotypes

1. Georgie
2. Koreke
3. Jupiter
4. Neptune
5. Goldmarker
6. 000-72
7. 6487
8. 124-72
9. Ark Royal
10. Magnum
11. Kaniere
12. Zephyr
13 Mata

--- lines indicate clusters from Wards on means.

Figure 3.62 Genotype Scores for Components 1 and 2 for 13 Barley Genotypes - Varimax Rotated



~50% of the total variation this lack of correspondence with cluster analysis is not really surprising.

The arrangement of the genotypes on these two axes seems to suggest the separateness of OOO-72, Neptune, Magnum and Goldmarker from the remaining genotypes and each other. On the varimax rotated graph Ark Royal is also apart from the others.

Similarly little correspondence could be seen between the dominant/dormant environments in Cluster Analysis and those showing strong correlations in Principal Component Analysis. Winfield was a dominant environment for the clustering on means and had a high positive correlation with Component 1 for both the rotated and unrotated PCA's. Other similarities were not evident however.

(ii) Barley 14 genotypes

The unrotated results are presented in Tables 3.65, 3.66 and 3.67. Figure 3.63 is a graph of the unrotated factor scores for the first two components.

The varimax rotated results are presented in Tables 3.68 and 3.69. Figure 3.64 is a graph of the rotated genotype scores.

The determinant of the correlation matrix was 0.000513. Bartlett's Chi-Square test indicated that there were six significant components. To be consistent with the previous data set the first four components are considered here together accounting for 73.3% of the total variation.

The first component accounted for 30.5% of the total variation and there were four environments which had high positive correlations with it (i.e. Gore, Fairview, Bulls and Winfield) in the unrotated analysis. Milton had a strong negative correlation and Pleasant Point and Takapau had small negative ones. For the varimax rotated analysis Gore, Bulls and Winfield again were highly correlated with the component with Pleasant Point and Milton having notable negative correlations.

Component 2 accounted for 17.2% of the variation taking the cumulative total to 47.8%. Takapau and Marton had high positive correlations. Pleasant Point and Outram had negative ones. The varimax rotated analysis showed Takapau being strongly positively correlated with this component and Pleasant Point and Fairview being negatively correlated with it.

Table 3.65 Principal Component Analysis Barley 14 Genotypes

Component	Eigenvalue	% of Variation	Cumulative %	X^2
1	3.35665 *	30.5	30.5	66.9
2	1.89680 *	17.2	47.8	52.0
3	1.54609 *	14.1	61.8	45.5
4	1.25871 *	11.4	73.3	38.1
5	1.06847 *	9.7	83.0	29.7
6	0.72390 *	6.6	89.6	21.4
7	0.58676	5.3	94.9	12.2
8	0.27938	2.5	97.4	
9	0.19739	1.8	99.2	
10	0.05668	0.5	99.7	
11	0.02916	0.3	100.0	

* denotes significance with Bartlett's Chi-Square Test.

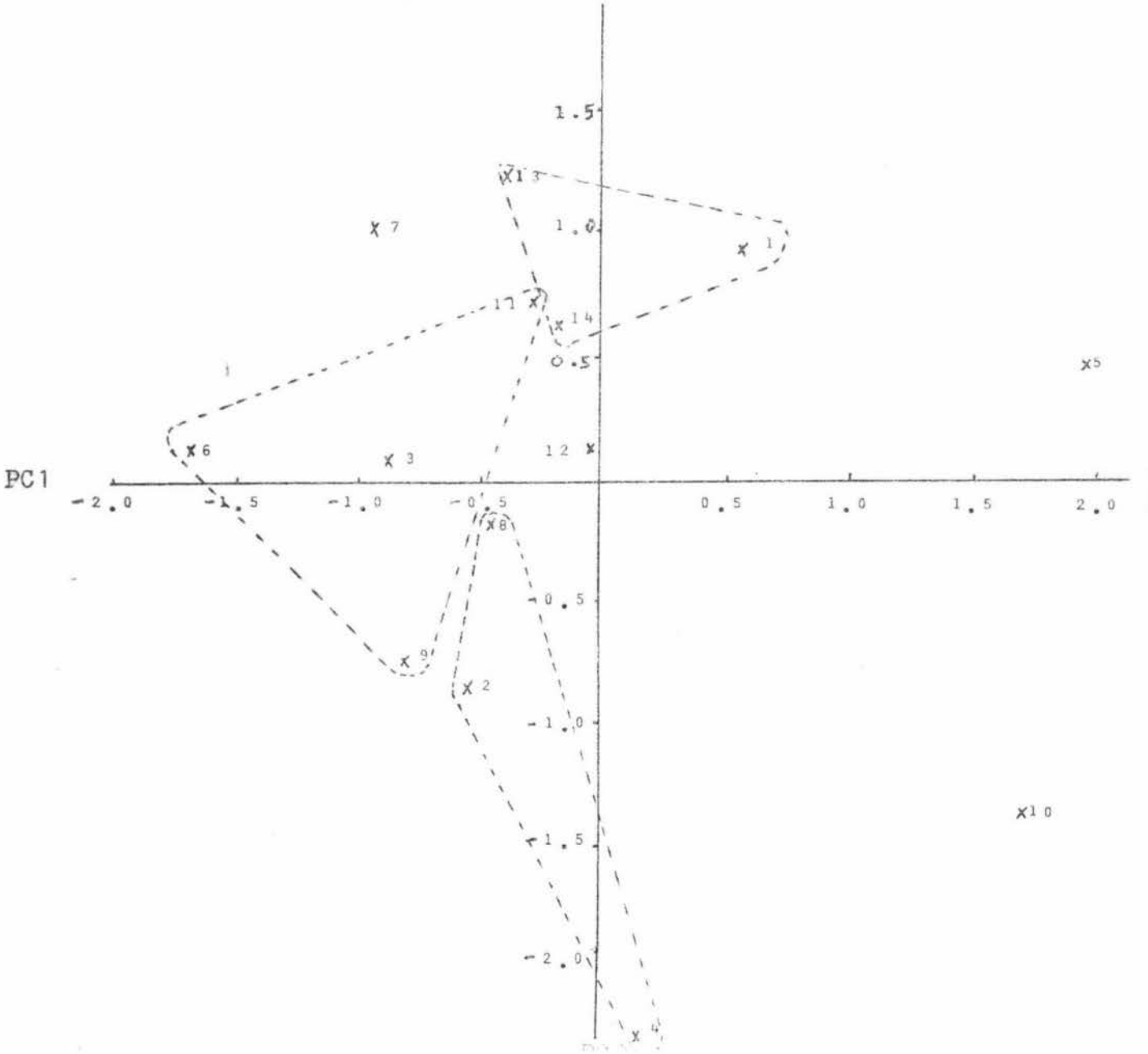
Table 3.66 Factor Structure Matrix

	Component 1	Component 2	Component 3	Component 4
Gore	0.79550	0.02201	-0.14592	0.01340
Pleasant Pt	-0.26114	-0.34677	0.57764	-0.09310
Fairview	0.71002	-0.06489	0.43425	-0.41267
Outram	0.22848	-0.37262	0.21159	0.46417
Takapau	-0.29884	0.77882	-0.22831	0.17962
Bulls	0.88424	0.10927	-0.17024	0.03515
Marton	0.36044	0.65650	0.20284	0.43485
Winfield	0.86011	0.23865	-0.15487	-0.17972
Leeston	-0.08467	0.48966	0.39165	-0.65749
Milton	-0.54441	0.51140	0.18244	-0.03408
Incholme	0.23422	0.15899	0.79049	0.41924

Table 3.67 Factor Scores: 14 Barley Genotypes

	Component 1	Component 2	Component 3	Component 4
Georgie	0.570079	0.925757	-0.000572	-0.744147
Koreke	-0.566884	-0.847484	-0.933814	-0.802940
Jupiter	-0.871437	0.091062	-0.486134	-0.379196
Neptune	0.121206	-2.273767	-0.446846	-0.274623
Goldmarker	1.963166	0.492373	-0.312684	-1.071806
000-72	-1.678791	0.122963	-0.525612	0.947717
6487	0.935070	1.043631	-0.138201	2.431432
124-72	-0.484280	-0.175060	-0.138885	0.954181
Ark Royal	-0.827948	-0.741471	2.094530	-0.137911
Magnum	1.690205	-1.361501	0.633924	0.493915
Kaniere	-0.286608	0.725642	1.565036	0.037133
Hassan	-0.012163	0.142858	-0.313257	0.426284
Mata	-0.379709	1.242250	0.813829	-1.439883
Zephyr	-0.171906	0.612745	-1.811315	-0.440155

Figure 3.63 Scores for components 1 and 2 for 14 genotypes - unrotated



Genotypes

1. Georgie
2. Koreke
3. Jupiter
4. Neptune
5. Goldmarker
6. 000-72
7. 6487
8. 124-72
9. Ark Royal
10. Magnum
11. Kaniere
12. Hassan
13. Mata
14. Zephyr

---- indicates clusters found by Wards Method.

Table 3.68 Varimax Rotated Factor Structure Matrix

14 Barley Genotypes

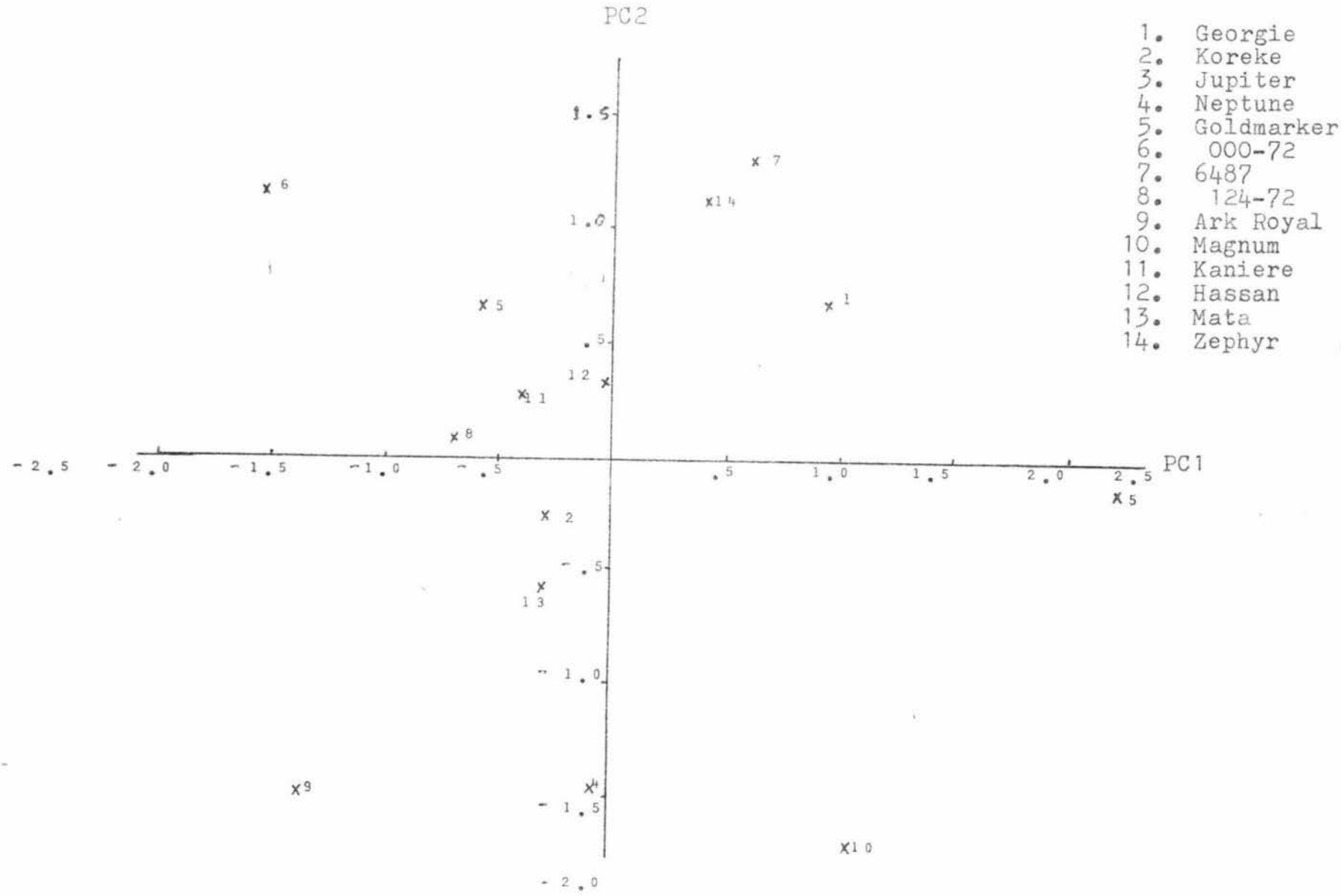
	Component 1	Component 2	Component 3	Component 4
Gore	0.81541	0.02162	-0.01235	-0.02112
Pleasant Pt	-0.44950	-0.58592	0.19894	0.11159
Fairview	0.62048	-0.56094	0.19661	0.35831
Outram	0.08435	-0.07516	0.06331	-0.13076
Takapau	-0.11373	0.83670	0.15880	0.20959
Bulls	0.85347	-0.10094	0.23551	-0.26283
Marton	0.28182	0.31361	0.82699	-0.08688
Winfield	0.92449	0.03266	0.06056	-0.09458
Leeston	0.04691	0.02623	0.01676	0.93145
Milton	-0.43194	0.47487	0.09057	0.52059
Incholme	-0.02157	-0.20640	0.84516	0.13976

Table 3.69 Factor Scores on Rotated Factors

14 Barley Genotypes

	Component 1	Component 2	Component 3	Component 4
Georgie	0.942317	0.664378	-0.276265	1.320919
Koreke	-0.300606	-0.262565	-1.494304	-0.224980
Jupiter	-0.518751	0.671034	-0.972111	0.541252
Neptune	-0.081922	-1.440038	-1.415479	-0.960013
Goldmarker	2.238439	-0.109651	-0.296106	0.758946
000-72	-1.546525	1.167700	-0.305664	-0.438847
6487	0.620268	1.309000	1.907535	-1.404334
124-72	-0.708368	0.090158	0.519042	-1.068428
Ark Royal	-1.392979	-1.490585	0.821893	0.791760
Magnum	1.050445	-1.700311	0.681341	-1.075933
Kaniere	-0.377304	0.270119	0.756176	1.691872
Hassan	-0.017819	0.310717	0.174603	-0.480456
Mata	-0.318418	-0.591704	0.925851	1.008843
Zephyr	0.411224	1.111748	-1.026512	-0.460601

Figure 3.64 Genotype Scores for components 1 and 2 for 14 Barley Genotypes - varimax Rotated



Interpretation of either the unrotated or rotated-components in terms of site information again proved difficult with no obvious trend being evident.

There was some correspondence between this data set and that for 13 genotypes, e.g. the importance of Takapau for the second component.

The graph of the unrotated genotype scores did not correspond very well with the clusters produced by cluster analysis. Neptune, Goldmarker and Magnum again appear to be apart from the other genotypes when mapped on the first two components.

For the varimax rotated components 000-72 also appears to be distinctly apart. This agrees well with the results found for the previous data set.

(iii) Wheat

The results of this Principal Component Analysis are presented in Tables 3.610, 3.611 and 3.612.- unrotated and Tables 3.613, 3.614 - varimax rotated.

The determinant of the correlation matrix was 0.0000149 Bartlett's Chi-Square test indicated that there were four significant components (Table 3.610).

The first two components together accounted for 77% of the total variation and since this met with the previously stated criterion these two alone are considered further. The first component accounted for 62.2% of the total variation.

The factor structure matrix Table 3.611 shows that there was a positive correlation between all the environments and the first component. With the exceptions of Tikokino 1978 and Aorangi 1978 the correlations were all very high.

After the varimax rotation the importance of Tikokino 1978 and Aorangi 1978 were lessened further and Halcombe 1977, Halcombe 1978 and Tikokino 1977 have the highest correlations with this component.

The second component accounted for 14.9% of the total variation. The unrotated Factor Structure Matrix, (Table 3.611), showed Tikokino 1978 having a strong positive correlation with this component and three environments, Halcombe 1977, Aorangi 1977 and Kairanga 1977 having moderately negative correlations. Following varimax rotation there were no negative correlations and Aorangi 1978 had a notably high positive correlation.

Table 3.610 Wheat: Principal Component Analysis

Component	Eigenvalue	% of Variation	Cumulative %	χ^2
1	4.97521 *	62.2	62.2	40.81
2	1.18966 *	14.9	77.1	31.51
3	0.91941 *	11.5	88.6	21.18
4	0.40300 *	5.0	93.6	12.64
5	0.27659	3.5	97.0	5.39
6	0.20211	2.5	99.6	
7	0.02995	0.4	99.9	
8	0.00407	0.1	100.0	

* = significance according to Bartlett's Chi-Square Test.

Table 3.611 Factor Structure Matrix

	Component 1	Component 2
Halcombe '77	0.90259	-0.22979
Aorangi '77	0.82313	-0.33771
Kairanga '77	0.83696	-0.36672
Tikokino '77	0.89701	0.12425
Halcombe '78	0.89523	-0.01121
Aorangi '78	0.55949	0.15407
Kairanga '78	0.83945	0.30824
Tikokino '78	0.39841	0.86834

Table 3.612 Factor Scores

	Component 1	Component 2
61,01	0.199834	-0.260904
300-301	1.341273	-1.021764
Oroua	-0.457483	0.123452
293-201	0.044245	1.501390
Pataka	-0.229436	0.297590
Gamenya	-1.659441	1.107329
Rongotea	0.085331	-0.841602
61,06	-1.244292	-1.528020
Karamu/9	1.570204	1.041631
301-302	0.349765	-0.419102

Figure 3.65 Genotypic Scores for Components 1 and 2 for Wheat Genotypes - unrotated

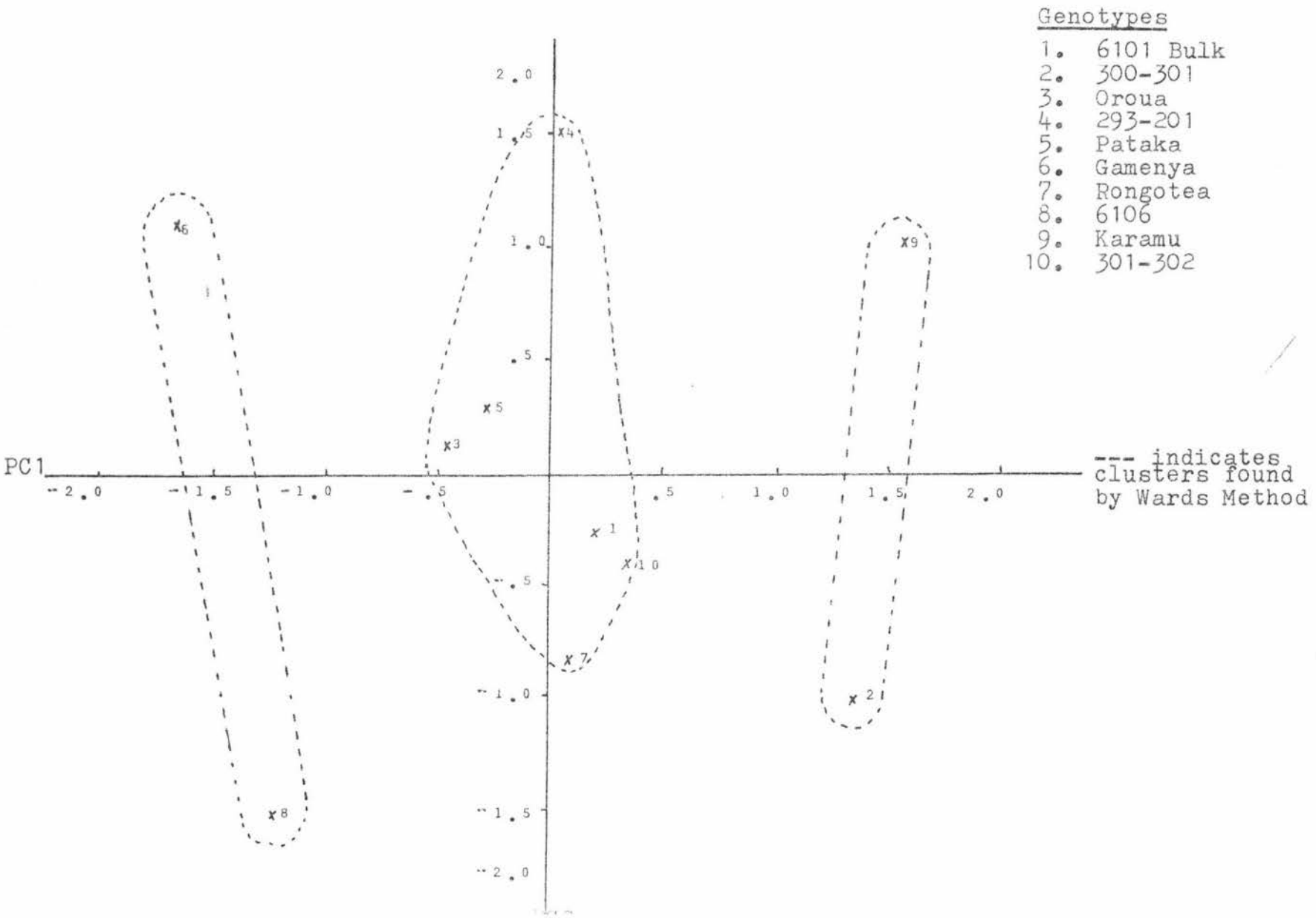


Table 3.613 Wheat: Principal Component Analysis

Varimax Rotated Factor Matrix

	Component 1	Component 2
Halcombe '77	0.81674	0.35898
Aorangi '77	0.41575	0.21264
Kairanga '77	0.73592	0.09073
Tikokino '77	0.67643	0.04671
Halcombe '78	0.88814	0.24063
Aorangi '78	0.18035	0.95645
Kairanga '78	0.36144	0.60034
Tikokino '78	0.09252	0.09376

Table 3.614 Factor Scores on Rotated Factors: Wheat

	Component 1	Component 2
61,01	0.071915	0.655779
300-301	1.694268	0.662596
Oroua	-0.963256	0.960352
293,201	0.240765	1.282884
Pataka	-1.104026	0.691903
Gamenya	-1.245826	-1.471524
Rongotea	-0.806364	-0.027419
61,06	0.395104	-1.262456
Karamu/9	1.066558	-0.421258
301-302	0.650862	-1.070857

Figure 3.66 Genotypic Scores for Components 1 and 2 for Wheat Genotypes - Varimax Rotated

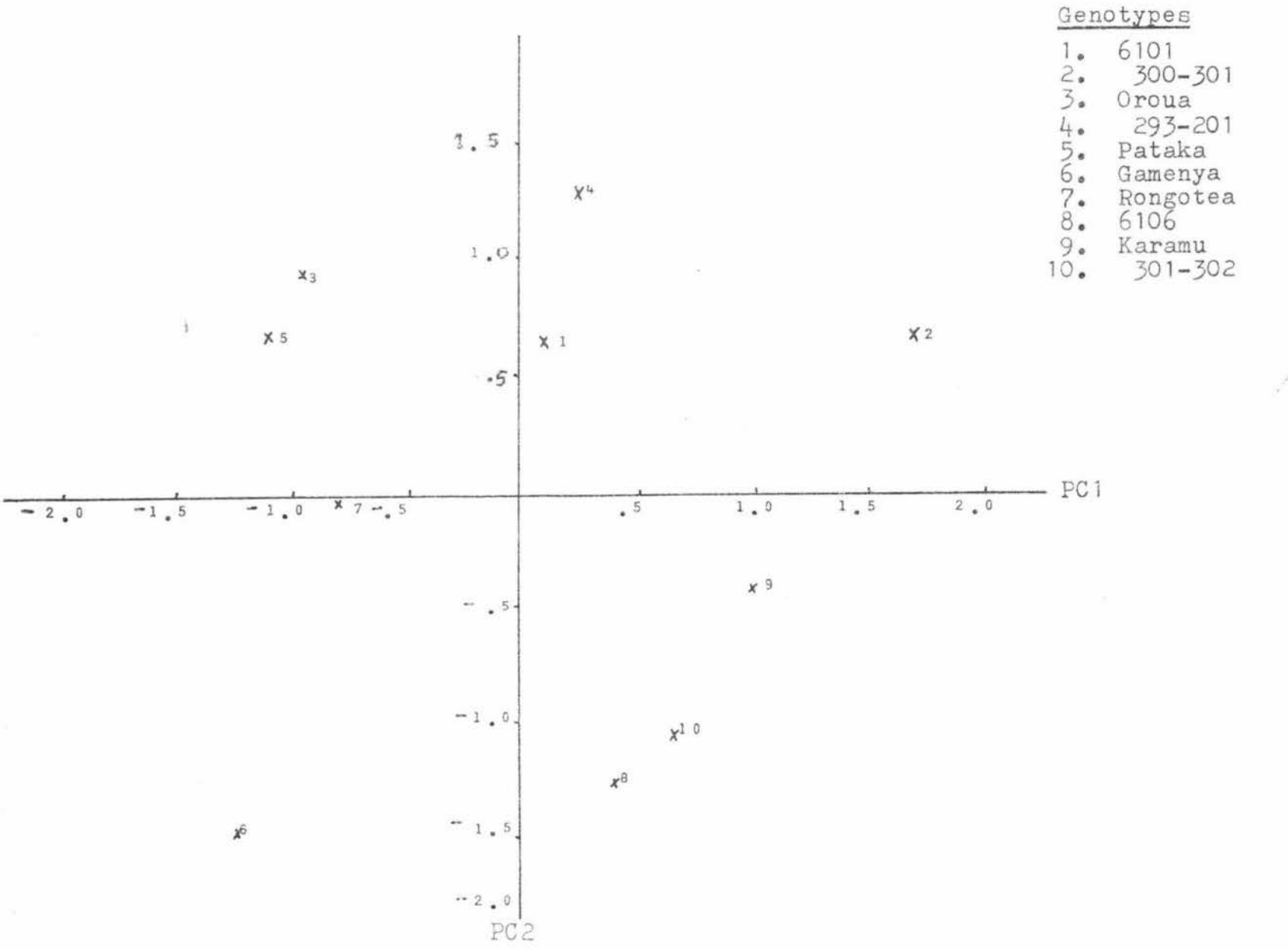


Figure 3.65 is a graph of the genotypic scores for the unrotated components 1 and 2. The clusters formed by cluster analysis are marked on the graph and for this set of data there is a marked agreement between the genotypes positions and the cluster memberships. The separation of the clusters is along the axis of the first component with the high yielding cultivars scoring positively and the low yielding cultivars negatively. With the exception of cultivar 293-201 none of the middle cluster score very highly on either component.

Figure 3.66 is the graph of the genotypic scores on the varimax rotated components. The cultivar Gamenya had notably high negative scores for both components 1 and 2 and is separate from the other cultivars. Cultivar 300-301 is also rather removed from the others.

It is of interest to note that Tikokino 1978 and Aorangi 1978 were the least important environments for the clustering analysis with the remaining environments all being dominant. This corresponds closely with the correlations of the environments with component 1, both unrotated and rotated.

Based on the unrotated analysis the first component may be interpreted as representing a general environmental trend with no particular sites or years distinction involved.

The positioning of the genotypes along this axis seems to be closely related to their mean yields over all environments.

3.7 Ratios to Standard Cultivars

(i) Barley: 13 genotypes

The standard cultivar for barley in New Zealand is Zephyr (Malcolm, 1978). The results of the ratios for the other cultivars with Zephyr are presented in Table 3.71. The number of times (out of a possible 17) that each cultivar was significantly higher or lower than Zephyr is also given. All of the cultivars bettered Zephyr in some environments but Georgie, Goldmarker, 6481 and Magnum did so most notably. In the cluster analysis on means (section 3.5) Goldmarker and Magnum formed a cluster which had generally significantly higher means for each environment. Georgie and 6487 also formed a cluster by themselves, in that analysis so there is some degree of agreement between the two analyses. The highest average ratio to Zephyr $\left(\frac{\bar{x}i..}{\bar{x}s..}\right)$ was produced by Goldmarker, the lowest by Koreke.

(ii) Barley: 14 genotypes

Again Zephyr was the standard cultivar used for this analysis. The results are presented in Table 3.72.

Georgie and Goldmarker were significantly higher than Zephyr in 9 out of 11 environments. Neptune and 000-72 were significantly lower than Zephyr in relatively large numbers of environments. Goldmarker again had the highest average ratio to Zephyr.

No correspondence could be seen in these results and groupings formed by the other analyses.

Table 3.71 Ratio to Standard Cultivar (Zephyr) 13 Genotypes

	Average ratio to Zephyr*	No. times signif. higher than Zephyr (5%)	No. times signif. lower than Zephyr (5%)	Rank
Georgie	1.096	12	0	2
Koreke	0.978	4	4	11
Jupiter	1.008	5	5	8
Neptune	0.997	5	8	13
Goldmarker	1.123	13	2	1
000-72	0.997	5	6	12
6487	1.096	10	2	3
124-72	1.006	7	4	9
Ark Royal	1.034	8	4	7
Magnum	1.083	11	3	4
Kaniere	1.074	9	2	5
Mata	1.050	7	2	6
Zephyr	1.000			10

*for 17 trials.

Table 3.72 Ratio to Standard Cultivar (Zephyr) 14 Barley Genotypes

	Average ratio to Zephyr*	No. times signif. higher than Zephyr (5%)	No. times signif. lower than Zephyr (5%)	Rank
Georgie	1.104	9	0	3
Koreke	0.970	3	3	14
Jupiter	1.018	4	3	10
Neptune	1.005	4	5	12
Goldmarker	1.123	9	1	1
000-72	1.015	4	3	11
6487	1.107	7	2	2
124-72	1.023	6	3	9
Ark Royal	1.032	7	2	8
Magnum	1.073	7	2	8
Kaniere	1.079	6	2	4
Hassan	1.058	6	1	6
Mata	1.047	7	3	7
Zephyr	1.000			13

*for 11 trials.

(iii) Wheat

Karamu was the standard cultivar used for this analysis (Mitchel and Cross, 1979b). The results are presented in Table 3.73. Only the cultivars 300-301 and 293-201 achieved yields significantly higher than Karamu and that was only in one environment each.

This illustrates the superior yielding ability of Karamu. The lowest average ratio to Karamu was that of Gamenya.

Table 3.73 Ratio to Standard Cultivar (Karamu) Wheat

	Average ratio to Karamu*	No. times signif. higher Karamu (5%)	No. times signif. lower Karamu (5%)	Rank
61,01	0.906	0	6	5
300-301	0.983	1	2	2
Oroua	0.867	0	7	8
293-201	0.887	1	6	6
Pataka	0.886	0	7	7
Gamenya	0.784	0	8	10
Rongotea	0.908	0	5	4
61,06	0.825	0	8	9
301-302	0.920	0	5	3
Karamu	1.000			1

* 8 trials.

Chapter 4 General Discussion

4.1 Analysis of Variance

The application of Analysis of Variance to trial data is a necessary initial step to further detailed studies. The importance of this analysis has been illustrated in this study. The analysis provided an indication of the presence and magnitude of any interactions present and the necessity of further analyses can be based on this.

There were two indications that not all the assumptions underlying Analysis of Variance were satisfied. Firstly all three pooled Anova's had heterogeneity of error variances which means that when testing for significant differences between environments the use of the pooled error term may cause bias. For the other significance tests the pooled error term is still a valid basis (Cochran, 1947). Secondly, the regression of effects for the wheat data indicated that for two genotypes there was a linear relationship between the $\alpha\lambda T$'s and the λT's suggesting non-independence of effects.

The analysis of the wheat data showed that the genotype x years component was larger than the genotype x locations component. This is commonly found (Simmonds, 1979) and emphasizes the necessity for testing cultivars over seasons.

4.2 Stability Parameters

The parameters of Wricke, Hanson and Eberhart and Russell have been termed stability parameters in this study because their primary purpose is to give some estimation of the degree to which each genotype interacts with the environments it is grown in.

The three parameters studied here represent only a sample of those that have been proposed.

Of these Ecovalences and the Variance about Regression (S^2d_i) are perhaps the most commonly used while Hansons is relatively untried.

All three are based on different concepts of stability. Hanson's defines a stability space where a genotype's stability is measured by its deviation from the expected performance for each environment. Ecovalences are a partitioning of the G-E Sums of Squares and S^2d_i measures the deviations from a linear trend over environments.

Despite this there was a certain degree of consistency found in the genotypic rankings for these parameters, indicating that they were all successfully detecting consistency of yield in a cultivar.

Ecovalences and S^2d_i are both off-shoots from other analyses, Anova and regression respectively and as such do not require large amounts of computation time and extra expertise. In comparison Hanson's parameter requires a more complex analysis and the concept on which it is based may be less easily understood.

As a single statistic upon which genotypic assessment can be made they are valuable especially to the practical plant breeder. They have the advantage of ease of interpretability and relative ease of computation.

The major criticism of them is that they do not describe the pattern of response of the cultivar but only the degree of variability over environments. Some indication of the form of the response may give additional information to the plant breeder (Byth, 1977).

4.3 Ratio to a Standard Cultivar

This method of analysis also produces a single statistic for each genotype. The average ratio to a standard cultivar $\frac{\bar{x}_{i..}}{\bar{x}_{s..}}$ involves a comparison of the α_i effects.

While this gives an indication of relative genotype average performance it ignores any degree of variability over environments that may be present and is therefore of limited value. The genotypic rankings for this ratio follow closely those for the mean yields, as would be expected.

The recording of the number of times a cultivar's performance exceeds a standard's performance may be of value to give a general idea of performance but the performance of a standard must influence these results as is illustrated by using Karamu for wheat where it is rarely exceeded and Zephyr for barley where it often is.

4.4 Regression Analysis - Finlay and Wilkinson

The linear Regression Analysis, as Finlay and Wilkinson described it, gives an indication of the form of a genotype's response, and in addition some idea of stability as judged by the regression slope β_i. If the Eberhart and Russell parameter S^2d_i is considered as part of the regression analysis then a further indication of stability is provided. In this sense stability refers to the degree of variation about the linear trend. This analysis has proven to be very useful and has been widely applied to many different crops.

It has been noted previously that the use of the $\bar{x}_{..k}$'s as an environmental index causes statistical problems in the regression analysis (Freeman and Perkins, 1971). Several studies have been done comparing analyses using independent environmental measures with those using the $\bar{x}_{..k}$'s. The general conclusion to be drawn from these is that little difference is to be found between the two. However consideration of the basic premise of Finlay and Wilkinson regression leads to the conclusion that there is a conflict between what the analysis sets out to show and an assumption underlying the experimental model. In order to successfully compare genotypic performances, significant differences from a slope of 1.0 are tested for. If they occur, however, it is implied that covariance of effects exists for the model underlying the analysis. The independence of effects is a basic assumption for a linear model.

The regression approach may still be a useful one, however, as long as an alternative environmental index is used. One promising possibility for this is to create such an index by estimating a Principal Component Analysis on a set of environmental variables e.g. climatic measurements, soil. This has been successfully done by Suzuki and Abe (1975) and Nor and Cady (1979).

Other criticisms have centred on the lack of linearity of the data or the failure to explain an adequate percentage of the G x E sums of squares. This study found a strong linear relationship between the variables so on this basis the analysis was successfully applied to these sets of data.

One factor in the wide application of regression analysis to G x E studies is the relatively easy computation involved and the ready availability of computer programs to do such an analysis. Statistical tests are associated with the analysis which allow objective decisions to be made concerning linearity, deviations from 1.0, etc.

The conceptual basis on which it stands of finding the linear relationship between a genotypes performance and an environmental index is one which is readily interpreted. It would be statistically more acceptable however if such an index was an independent one and not the $\bar{x}_{..k}$'s as suggested by Finlay and Wilkinson.

4.5 Regression Analysis - Effects

The regression of effects from a linear model would theoretically show that no linear relationship existed between them.

This was largely borne out by this study where only in two exceptional cases was any trend shown to exist.

Since such a negative conclusion is of no value to the interpretation of genotype x environment interactions this analysis is of no practical use for this purpose. It may be used to check for independence of effects if this is desired.

4.6 Pattern Analysis

The advantages of using Cluster Analysis and Principal Component Analysis to interpret Genotype x environment interactions have been listed by Eisemann *et al.* (1977a) as follows:-

1. Comparisons of response patterns are simplified through data reduction.

2. No *a priori* form of the response pattern is assumed, as is the case with linear regression.

3. Summarisation of the degrees of similarity among genotypes in their response patterns, is produced.

4. The contrasting of different response patterns permits identification of specific response differences.

5. Environments giving rise to specific contrasts in response patterns can be identified.

6. Differences in response patterns can be related to genetic backgrounds and environmental differences.

7. In general enables the formulation of hypotheses about genotypic adaptation, G x E interactions and differences in environmental factors.

These authors have stressed the importance they place on point 2 in several papers (Eisemann et al, 1977b; Byth et al, 1976; Mungomery et al, 1974). They have evidently encountered data where the linear relationship of the data was not strong although they do not publish details of this. However as they state themselves a basic assumption underlying cluster analysis is that there are discontinuities in the data. The successful application of this analysis must therefore depend on this and if there are no such divisions present cluster analysis may misrepresent any trends that are there. The variance basis of Wards Incremental Sum of Squares method is the best strategy from this point of view as it is not concerned with distances between groups or individuals but with variance partitioning.

Cluster Analysis produces groups of similarly responding genotypes over the environments tested and shows a phyletic relationship amongst them. Principal Component Analysis also can be used to produce such groupings but there is no way to assess the phyletic relationship between the groups as there is with cluster analysis. Principal Component Analysis does not assume discontinuities in the data as Cluster Analysis does and considers the total dispersion i.e. variances and covariances, in G x E performance.

Both these analyses bring about a reduction in the number of genotypic comparisons to be made which may be valuable for large data sets but the important aspect of such groups is that they may suggest new hypotheses about genotypic adaptation and environmental influences.

To be successful in this respect the clusters have to be further assessed in light of background knowledge such as environmental variables, previous stability rankings (as in this study), regression results and genotypic backgrounds. This need for further analyses to aid interpretation can be seen as a disadvantage of Pattern Analysis techniques. The simplicity of interpreting the analysis in terms of genotype adaptability and stability is not possible as it was with the earlier analyses. The possibility of using rotation to aid interpretation of Principal Components is an advantage of this analysis over Clustering.

An additional disadvantage of Cluster Analysis is that several subjective decisions are needed in its application. Choice of distance measures, strategies and truncation points are made on the basis of previous experience and knowledge about the data set. The solution of a Principal Component Analysis is unique and no decisions have to be made on a purely subjective basis.

Both Cluster Analysis and Principal Component Analysis have been successfully applied by other workers. This study however encountered some problems with Principal Component Analysis, illustrating that its application may not always be straightforward and its interpretation not always obvious. Both analyses involve complex computer programs which are often not readily available to a plant breeder. Cluster Analysis is perhaps the more available

method but the degree of subjectivity involved means some experience with it is desirable.

4.7 Barley Data Sets

The Pooled Analysis of Variance indicated that highly significant genotype x location occurred for both sets of barley data. The trials sties for both sets covered a wide geographical range and included all the major Barley growing areas of New Zealand.

The presence of such interactions is therefore an important aspect to be considered for Barley Cultivar evaluations.

The presence of significant location components tends to suggest that growing trials over this range of environments is justified. Significant genotypic variation apparently existed within this sample of genotypes.

A comparison of three different stability parameters revealed that similar rankings were produced by each one. In particular genotypes showing a high level of variability over environments had very consistent rankings for these methods.

It could be seen that examination of cultivar mean yield over environments and mean ratios to a standard cultivar did not supply all the information about a genotype's performance.

A strong linear relationship existed between the genotype yields and the environmental index and none of the significance tests suggested that the regression technique was not applicable to this data. Following refinement of the $\hat{\sigma}_{\gamma \cdot \times}$ several genotypes were found to have slopes significantly different from 1.0. Georgie had a slope steeper than 1.0, Koreke and Mata had slopes

flatter than one. For the 14 genotypes data set Hassan had a slope greater than 1.0. No significant differences between the $\beta_{i's}$ were found.

The regression of the $\alpha\eta_{ik}$ effects on to the $\eta_{k's}$ was not successful for the Barley data as no linear relationship was present. This was as expected illustrating the lack of practical use for this type of regression.

The Cluster Analysis allowed examination of two different aspects of the data. Firstly the identification of groups of similarly responding genotypes with respect to G-E means or G-E effects provided a summary of the response patterns present in the data. Secondly the environments important in the formation of the groups could be identified. Clustering on the basis of G-E means ($\bar{x}_{i\cdot k}$) is an examination of the general pattern of response for each genotype. Clustering $\alpha\eta_{ik}$ effects is looking in particular at each genotype x environment combination relative to the whole set of G-E effects.

For both the 13 genotype and 14 genotype barley sets, some correspondence was seen between the means cluster analysis and the effects cluster analysis. There were several pairs of genotypes which remained clustered together and these could be related to the stability rankings and mean yields. This would seem to indicate that Cluster Analysis was successfully summarising the response patterns to the environments involved.

The Principal Component Analysis of the Barley data was not effective in interpretation of the response patterns for several reasons. Firstly the first few components did not account for

large amounts of the total variation and this meant consideration of relatively large numbers of components.

The interpretation of the components and genotypic scores in such cases would therefore be expected to be difficult.

In addition relating any of the components to specific combinations of external factors such as geographic position climatic data and soil types was not successful.

The fact that many small components had to be extracted may indicate that there was no simple influence underlying the pattern of the data (Williams, 1976). He discusses such situations and suggests two possible causes.

1. When the variances of the original measurements are approximately equal and the measurements are only very weakly correlated. In such a case no really efficient reduction of the space is possible.

2. The correlations between the original measurements may be strongly curvilinear in which case the Principal Component Analysis is "helpless" because it takes out the best straight lines.

An examination of the correlation matrices for the barley Principal Component Analyses (Appendix 2) shows uniformly low correlations between the variables so the first situation may be the most likely here. Varimax rotation did not appear to aid interpretation of the components.

Superimposing cluster boundaries on to the genotypic scores for components 1 and 2 did not show a high degree of association between Cluster Analysis and Principal Component Analysis.

However as the first two components explained less than half of the total variance, this is not unexpected. The value of considering only the first two component scores, must be lessened when so little of the total variance is accounted for.

Generally similar results for the two sets of barley data were produced for all the analyses. Hassan (the additional genotype for the 14 genotype set) proved to be a cultivar of note as it ranked highly on the stability parameters and had a regression slope steeper than 1.0. Its mean yield ranked in the middle of this sample of genotypes.

The correlation coefficient between mean yield and β_i was 0.46 (13 genotypes) and 0.52 (14 genotypes) indicating that independent manipulation of these may be possible.

4.8 Wheat Data

The pooled analysis of variance over sites and years indicated that there was a significant genotype x years interaction but none for genotype x locations. A significant years component emphasized the importance of testing over seasons whereas the non-significant locations component meant that perhaps the use of several trial sites in the lower North Island was not necessary to ascertain genotypic performance there. Similar results were found by Mitchel and Cross (1977) using similar locations.

There was genotypic variation present as indicated by the significant genotypic component.

The genotypic rankings for the different stability parameters were not as consistent as was found with the barley data.

However once again the unstable (Karamu) and the stable (Gamenya) could readily be identified as such.

A strong linear relationship was apparent in the Finlay and Wilkinson regression so there was no question of invalidity of this analysis from this point of view. Gamenya had a highly significant flat slope which by Finlay and Wilkinson's definition meant that it is a stable cultivar. This result corresponded well with the stability parameters rankings. Cultivars 300-301, 301-302 and Rongotea had significantly steep slopes.

Two genotypes showed a linear relationship when the effects regression was applied indicating non-independence of effects. The cluster analysis of means gave some interesting results. Three clusters were formed with one having consistently higher means and one having consistently lower. Apparently most of the sites x years combinations were important for the cluster formation. Clustering on effects showed some correspondence with that on means.

The Principal Component Analysis was found to be effective in determining the pattern for this data set. Because the first two components accounted for a large proportion of the total variance, interpretation of these done could identify the major sources of influence i.e. which sites x year combinations were important.

In addition the scores of the genotypes on these two components produced a pattern corresponding closely to that of clustering i.e. both analyses suggested that the wheat cultivars formed three distinct groups.

The first principal component was related to a general environmental trend as there were strong correlations with most of the environments and no distinct seasonal or sites effect. The varimax rotation did not aid in the interpretation of the component. The second component showed a polarity effect with the 1977 trials having small negative correlations. The rotated factor matrix did not support such interpretation however. Other studies have found that a similar interpretation could be given to the first and second components (Perkins, 1971; Cooley and Lohnes, 1971).

There was close correspondence between the dominant/dormant environments in Cluster Analysis on Means and the unrotated factor structure matrix.

The correlation between mean yield over environments and the linear regression coefficient was 0.806 suggesting that selection for either particular responses or yield may result in strong correlated responses in the other parameter (Brennan and Byth, 1979).

4.9 Barley Cultivar Assessments

The fourteen barley genotypes were analysed by a wide variety of techniques and several trends could be discerned from their results. Firstly the pooled Anova found that G - E interactions occurred for both sets of data which means that some consideration of this aspect should be included in cultivar assessment.

The present practice of comparing each cultivar to a standard is likely to be confused and complicated by the presence of such interactions.

The ratio of cultivar 1 to a standard s at any one test site κ can be expressed as the following.

$$\frac{\bar{x}_{1\cdot\kappa}}{\bar{x}_{s\cdot\kappa}} = \frac{\mu + \alpha_1 + \eta_\kappa + \alpha\eta_{1\kappa}}{\mu + \alpha_s + \eta_\kappa + \alpha\eta_{s\kappa}}$$

Such a comparison can be seen to involve the $\alpha\eta_{i\kappa}$ effects as well as the $\alpha_{i's}$. Since at each site the $\alpha\eta_{s\kappa}$ for the standard may be different and is unknown the comparison by this ratio is not merely a simple ratio of genotypic performances i.e. $\alpha_{i's}$.

Some assessment of a cultivar's stability or degree of variability, by whatever definition, may be valuable information for judging a cultivar's worth. By such analyses it could be inferred that cultivars Goldmarker and Magnum were relatively unstable and that their response patterns singled them out from the others in the analyses. Koreke, Kaniere and Hassan appeared to be relatively stable across environments.

The adaptation trends described by Finlay and Wilkinson regression suggests that Hassan and Georgie changed to a greater extent than average with changing environments and that Koreke and Mata changed to a lesser extent than average.

The standard cultivar Zephyr ranked poorly for ecovalences showing that it had a high proportion at G x E variation. This feature would affect the assessment of others using the ratio technique described above. A cultivar such as Hassan with a stable performance and an average yielding ability may be more suitable for this purpose.

4.10 Wheat Cultivar Assessments

When the comparison of new cultivars occurs over seasons as well as over locations further interaction effects are involved in means ratios, as follows:-

$$\frac{\bar{x}_{1\cdot k\ell}}{\bar{x}_{s\cdot k\ell}} + \frac{\mu + \alpha_1 + \lambda_k + T_\ell + \alpha\lambda_{1k} + \alpha T_{1\ell} + \lambda T_{k\ell} + \alpha\lambda T_{1k\ell}}{\mu + \alpha_s + \lambda_k + T_\ell + \alpha\lambda_{sk} + \alpha T_{s\ell} + \lambda T_{k\ell} + \alpha\lambda T_{sk\ell}}$$

For the wheat data tested here the $\alpha\lambda T_{ik\ell}$ and the $\alpha\lambda_{ik}$ components were not significant. However the $\alpha T_{i\ell}$ components were and would therefore be confounded with α_i in any ratio comparison with the standard.

The adaptation analyses applied in this study found that there were three distinct groups within this set of wheat cultivars:

1. Karamu/9 and 300-301 had high mean yields and ranked poorly for stability.

2. Gamenya and 61,06 had low mean yields and ranked highly for stability.

3. The remaining six cultivars were grouped together and were apparently not extreme for either property.

These trends in cultivar performance are similar to those suggested by Finlay and Wilkinson (1963) (refer to figure 1.1). Previous published work, Mitchel and Cross (1977) had also found that Karamu/9 was unstable for yield.

Conclusions

1. The pooled Analysis of Variance showed that significant genotype-location interactions were present in the barley data, and that significant genotype-year interactions were in the wheat data.

2. The three parameters measuring the degree of genotypic variability over environments agreed reasonably well for cultivar rankings. They were useful for the interpretation of Cluster analysis and Principal Component analysis.

3. Lack of linearity found by other workers for Finlay and Wilkinson regression was not evident here. Some distinction between cultivars for their regression slopes could be made. A conceptual conflict was seen to exist in the use of a non-independent environmental index. The use of an independent index perhaps produced by Principal Component Analysis of environmental variables was suggested.

4. A comparison of eight different clustering strategies produced results in agreement with other authors.

   Wards Incremental Sums of Squares method produced clusters which could be interpreted in the light of mean yields and stability measures. The probabilistic cut off measure appeared to produce reasonable numbers of clusters of even size.

5. The Principal Component Analysis was effectively applied to the wheat data where the first two components accounted for 77% of the total variation. For the barley data large

numbers of components were needed to account for a satisfactory amount of the total variation making interpretation difficult. Varimax rotation was not of great value in interpretation of the components.

6. There was generally close correspondence between the two sets of barley data. The fourteenth cultivar Hassan could be distinguished as having above average adaptation but also had little variability over environments as measured by the stability parameters. Other notable genotypes - Goldmarker and Magnum with relative instability were also singled out by the Pattern Analysis.

7. Both cluster analysis and principal component analysis suggested that the wheat cultivars fell into three groups. Those with high mean yields and relative instability, those with low mean yields and high stability and the remainder which were not extreme for either property.

8. It was felt that although all of the analyses studied had some problems they all produced information of value for genotype-environment interaction interpretation. Their roles were in general complementary to each other. The choice of analysis depends largely on facilities and expertise available and the purpose for which the results are needed.

Appendix 1

Principal Component Analysis Barley 13 Genotypes with 12 Environments

Component	Eigenvalue	% of Variation	Cumulative %
1	4.24919	35.4	35.4
2	2.14824	17.9	53.3
3	1.53529	12.8	66.1
4	1.43941	12.0	78.1
5	0.98739	8.2	86.3
6	0.79423	6.6	92.9
7	0.31644	2.6	95.6
8	0.24235	2.0	97.6
9	0.17353	1.4	99.1
10	0.08021	0.7	99.7
11	0.02818	0.2	100.0
12	0.00554	0.0	100.0

Factor Matrix

	Component 1	Component 2	Component 3	Component 4
Gore	0.76499	-0.29920	0.23686	-0.25561
Fairview	0.70506	-0.25146	-0.48635	0.17694
Outram	0.13911	-0.50758	0.26524	0.42853
Takapau	-0.00622	0.85945	0.38939	-0.06226
Bulls	0.80508	-0.17453	0.11464	-0.35873
Winfield	0.80883	-0.14699	0.12535	-0.34920
Leeston MAF	0.16803	0.57835	-0.38587	0.16188
Incholme	0.34178	-0.03100	0.25553	0.83219
Mitcham	0.57557	0.63068	-0.29396	0.19927
Kirwee	0.78421	0.44422	0.24298	-0.14842
Leeston CRD	-0.34565	0.12787	0.77058	0.02289
Dunsandel	0.79644	-0.01536	0.19595	0.35159

Factor Scores 13 Barley Genotypes 12 Environments

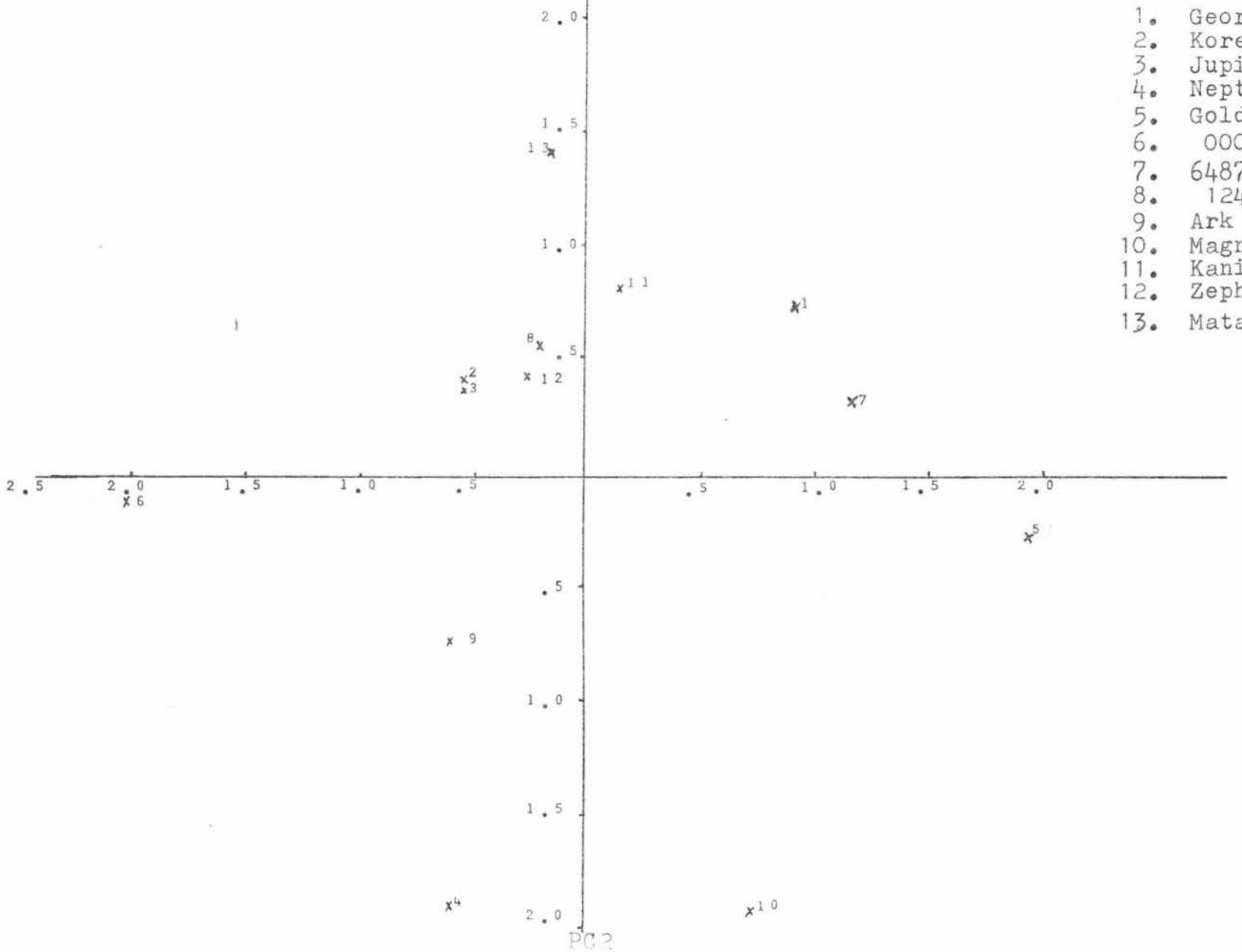
	Component 1	Component 2	Component 3	Component 4
Georgie	0.918832	0.547215	1.243080	-0.454754
Koreke	-0.505563	0.418056	-0.261291	-0.155138
Jupiter	-0.526688	0.394133	-1.43441	-0.131581
Neptune	-0.616365	-1.919559	-0.122959	-0.773292
Goldmarker	1.951928	-0.298751	-1.035963	-0.883138
000-72	-2.010095	-0.113032	1.49453	-0.529705
6487	1.163323	0.330468	1.577013	0.745033
124-72	-0.214715	0.558367	0.538316	0.219806
Ark Royal	-0.594847	-0.739343	-1.137822	2.019230
Magnum	0.721374	-1.914251	0.385409	0.157747
Kaniere	0.138296	0.810656	-0.013355	1.736576
Zephyr	-0.280031	0.494106	-0.495336	-1.459789
Mata	-0.145450	1.431935	-0.734105	-0.490994

Genotypic Scores for Components 1 and 2 for 13 Barley Genotypes: 12 Environments - Unrotated

Appendix 1

Genotypes

1. Georgie
2. Koreke
3. Jupiter
4. Neptune
5. Goldmarker
6. 000-72
7. 6487
8. 124-72
9. Ark Royal
10. Magnum
11. Kaniere
12. Zephyr
13. Mata



Appendix 2

Correlation Coefficients for 13 Barley Genotypes: 17 Environments

	Gore	Pleasant Point	Fairview	Outram	Takapau	Bulls	Marton	Winfield	Leeston MAF	Milton	Incholme
Gore	1.000										
Pleasant Point	-0.2029	1.0000									
Fairview	0.4502	-0.0361	1.0000								
Outram	0.3241	0.0063	0.0732	1.0000							
Takapau	-0.0886	-0.2456	-0.4859	-0.1807	1.0000						
Bulls	0.6744	-0.1541	0.4851	0.0199	-0.0881	1.0000					
Marton	0.1307	-0.1643	0.1427	-0.1454	0.3226	0.4799	1.0000				
Winfield	0.7431	-0.3794	0.5397	0.0561	-0.1111	0.7382	0.3109	1.0000			
Leeston MAF	-0.1049	0.1034	0.3184	-0.1884	0.3263	-0.1290	0.0393	0.1215	1.0000		
Milton	-0.2356	0.0355	-0.3099	-0.1489	0.3990	-0.5811	0.0692	-0.2819	0.3783	1.0000	
Incholme	0.0463	0.1896	0.3498	0.2947	-0.0520	0.0431	0.5340	0.0609	0.0409	0.0462	1.0000
Mithcham	0.1518	-0.4514	0.3093	-0.1909	0.4469	0.2530	0.4921	0.1620	0.3970	0.0581	0.2211
Methven	0.4187	0.0367	0.7763	0.1036	-0.4403	0.4578	0.2018	0.6640	0.2707	-0.0621	0.3762
Kirwee	0.5144	-0.5546	0.3519	-0.1080	0.4648	0.6121	0.5790	0.6580	0.2057	-0.2053	0.2241
Lincoln	0.3502	-0.1253	0.5861	0.1552	-0.2535	0.4383	0.4097	0.4902	0.2546	-0.2613	0.2058
Leeston CRD	-0.1666	0.3180	-0.4657	-0.0807	0.2840	-0.2263	0.0833	-0.1033	-0.0653	0.1219	0.1812
Dunsandel	0.5601	-0.0488	0.4186	0.2354	0.0196	0.5783	0.4615	0.4897	0.0115	-0.3479	0.5724

	Mitcham	Methven	Kirwee	Lincoln	Leeston CRD	Dunsandel
Gore						
Pleasant Point						
Fairview						
Outram						
Takapau						
Bulls						
Marton						
Winfield						
Leeston MAF						
Milton						
Incholme						
Mitcham	1.0000					
Methven	0.0389	1.0000				
Kirwee	0.6354	0.2009	1.0000			
Lincoln	0.1842	0.3463	0.3941	1.0000		
Leeston CRD	-0.4892	-0.2150	-0.0241	-0.0676	1.0000	
Dunsandel	0.5285	0.4123	0.5198	0.1885	-0.1748	1.0000

Appendix 2

Correlation Coefficients for 14 Barley Genotypes: 11 Environments

	Gore	Pleasant Point	Fairview	Outram	Takapau	Bulls	Marton	Winfield	Leeston	Milton	Incholme
Gore	1.0000										
Pleasant Point	-0.2028	1.0000									
Fairview	0.4262	-0.0358	1.0000								
Outram	0.3211	0.0064	0.0739	1.0000							
Takapau	-0.0871	-0.2457	-0.4863	-0.1810	1.0000						
Bulls	0.6713	-0.1540	0.4852	0.0202	-0.0883	1.0000					
Marton	0.1097	-0.1586	0.1462	-0.1362	0.3109	0.4706	1.0000				
Winfield	0.7349	-0.3783	0.5405	0.0575	-0.1119	0.7376	0.3159	1.0000			
Leeston MAF	-0.0826	0.0986	0.3009	-0.1887	0.3204	-0.1282	-0.0169	0.1032	1.0000		
Milton	-0.2451	0.0362	-0.2988	-0.1419	0.3894	-0.5689	0.1069	-0.2658	0.3164	1.0000	
Incholme	0.0601	0.1848	0.3374	0.2847	-0.0481	0.0399	0.4693	0.0485	0.0849	0.0101	1.0000

CODON/PANSY

```
C     THE MAIN PROGRAMME FOR PANSY                                                00100000  T  000:0000:5
C                                                                                 00200000  T  000:0000:5
C     ONLY THE DIMENSIONS OF THE MAIN PROGRAM NEED TO BE CHANGED IF               00300000  T  000:0000:5
C     THE DATA HAS DIMENSIONS GREATER THAN 4000 OR Z ARRAY REQUIRED IS            00400000  T  000:0000:5
C     BIGGER THAN 14500                                                           00500000  T  000:0000:5
FILE  19(TITLE="REGPAR",KIND=DISK,MAXRECSIZE=14,BLOCKSIZE=420)                    00600000  T
FILE  21(TITLE="YLOUT",KIND= PRINTER,MAXRECSIZE=22)                               00700000  T
FILE  22(TITLE="GROUP",KIND = DISK,MAXRECSIZE=14,BLOCKSIZE=420)                   00800000  T
FILE  20(TITLE="YLGRD",KIND = DISK,MAXRECSIZE=14,BLOCKSIZE=420)                   00900000  T
                                                                                          START OF SEGMENT 002
      DIMENSION Z(14500),DATA(4000)                                               01000000  T  002:0000:0
      LIMZ=14500                                                                  01100000  T  002:0000:0
      LIMD=4000                                                                   01200000  T  002:0001:0
      CALL PRELIM(Z,DATA,LIMZ,LIMD)                                               01300000  T  002:0002:0
      END                                                                         01400000  T  002:0005:0
                                                                                          SEGMENT 002 IS 0009 LONG

##################################################################################################################################

                                                                                          START OF SEGMENT 003
WARNING:THE SUBROUTINE "PRELIM" WAS NOT FOUND                                                   (150)

                                                                                          START OF SEGMENT 005
                                                                                    SEGMENT 005 IS 0011 LONG
                                                                                    SEGMENT 003 IS 0015 LONG

      NO ERRORS DETECTED.  NUMBER OF CARDS = 14.
      COMPILATION TIME = 9 SECONDS ELAPSED,  0.94 SECONDS PROCESSING.
      D2 STACK SIZE = 10 WORDS.  FILESIZE = 2564 WORDS.  ESTIMATED CORE STORAGE REQUIREMENT = 21159 WORDS.
      TOTAL PROGRAM CODE = 81 WORDS.  ARRAY STORAGE = 18500 WORDS.
      NUMBER OF PROGRAM SEGMENTS = 5.  NUMBER OF DISK SEGMENTS = 16.
      PROGRAM CODE FILE = (AG0938CULLEN)CODON/PANSY ON PACK.
      COMPILER COMPILED ON 10/31/78 (FORTRAN ON PACK).
```

```
LIST SYMBOL/PRELIM/TWO
DATE 03/02/81          TIME IS 15:27
SYSTEM/DUMPALL VERSION  3.0.140
LASTRECORD =     176
MAXRECSIZEIN =    14    BLOCKSIZEIN =  420
*** EBCDIC ***          UNITS=WORDS

      1    C      PROGRAM  PANSY                                                          00100000
      2    C      POOLED ANOVA FOR SITES/YEARS MODEL                                    00200000
      3    C                                                                            00300000
      4    C      ***************************************************************       00400000
      5    C                                                                            00500000
      6    C      ADAPTED BY CAROL CULLEN                                               00600000
      7    C                                                                            00700000
      8    C      FROM YLGR:-PROGRAMMERS P.J.TAYLOR,R.EISEMANN,R.SHORTER,I.DE LACY.     00800000
      9    C      UNIVERSITY OF QUEENSLAND                                              00900000
     10    C      ***************************************************************       01000000
     11    C                                                                            01100000
     12    C      PANSY CARRIES OUT A POOLED ANALYSIS OF VARIANCE AND A REGRESSION      01200000
     13    C      ANALYSIS.                                                             01300000
     14    C                                                                            01400000
     15    C      VARIOUS STABILITY PARAMETERS ARE PRODUCED AND A TABLE OF MEANS        01500000
     16    C      IS PRINTED OUT.                                                       01600000
     17    C                                                                            01700000
     18    C      OUTPUT FILES "YLOUT" IS FOR THE ANALYSIS; "GROUP" IS THE MATRIX       01800000
     19    C      OF GENOTYPE * ENVIRONMENT MEANS                                       01900000
     20    C                                                                            02000000
     21    C      DATA INPUT FILE IS "YLGRD"                                            02100000
     22    C                                                                            02200000
     23    C      ***************************************************************       02300000
     24           SUBROUTINE PRELIM (Z,DATA,LIMZ,LIMD)                                  02400000
     25    C                                                                            02500000
     26    C      COUNTERS ARE DETERMINED WHICH DIMENSION THE ARRAYS USED IN THE        02600000
     27    C      ANALYSIS                                                              02700000
     28           DIMENSION DATA(1),TEMP(16,10),L(4),Z(1),KG(5,220),KT(13),KS(5,220)02800000
     29          1,KY(5,15),FMT(13),XIN(90,20,16),N(20,90)                              02900000
     30    C                                                                            03000000
     31            KOUNT=0                                                            03100000
     32    C       DATA PARAMETERS: NV=NO. OF VARIABLES; NT=NY*NL*NG*NR; NY=NO. OF     03200000
     33    C       YEARS; NL=NO. OF LOCATIONS; NG=NO. OF GENOTYPES; NR=NO. OF REPS.    03300000
     34            READ (5,1) KT                                                      03400000
     35        1   FORMAT(13A6)                                                       03500000
     36            READ(5,2)NV,NT,NY,NL,NG,NR,NP                                      03600000
     37        2   FORMAT(7I6)                                                        03700000
     38            READ 3,((KG(O,I),O=1,5),I=1,NG)                                    03800000
     39        3   FORMAT(5A6)                                                        03900000
     40            READ 7,((KS(O,I ),O=1,5),I=1,NL)                                   04000000
     41            READ 7,((KY(O,I),O=1,5),I=1,NY)                                    04100000
     42        7   FORMAT(5A6)                                                        04200000
     43            WRITE (21,61)KT                                                    04300000
     44       61   FORMAT(20X,1H*,13A6,1H*)                                           04400000
     45            WRITE (21,62)NY,NL,NG,NR                                           04500000
     46       62   FORMAT(///10X,12HNO. OF YEARS,I5//10X,12HNO. OF SITES,I5//10X,10HN 04600000
     47           10. OF GENOTYPES,I5//10X,11HNO. OF REPS,I5/////)                    04700000
     48            GO TO 6                                                            04800000
     49        4   IF (NV-KOUNT)70,70,21                                              04900000
     50        6   N1 = NY+1                                                          05000000
```

```
      N2=N1+NL                                                          05100000
      N3=N2+NG                                                          05200000
      N4=N3+NR                                                          05300000
      N5=N4+NY*NL                                                       05400000
      N6=N5+NY*NG                                                       05500000
      N7=N6+NY*NR                                                       05600000
      N8=N7+NL*NG                                                       05700000
      N9=N8+NL*NR                                                       05800000
      N10=N9+NG*NR                                                      05900000
      N11=N10+NY*NL*NG                                                  06000000
      N12=N11+NY*NL*NR                                                  06100000
      N13=N12+NY*NG*NR                                                  06200000
      N14=N13+NL*NG*NR                                                  06300000
      N15=N14+NY*NL                                                     06400000
      N16=N15+NY*NL*NG                                                  06500000
      N17=N16+NG                                                        06600000
      N18=N17+NG                                                        06700000
      N19=N18+NG                                                        06800000
       N20=N19+NG                                                       06900000
      N21=N20+NG                                                        07000000
      N22=N21+NG                                                        07100000
      N23=N22+NG                                                        07200000
      N24=N23+NG                                                        07300000
      N25=N24+NR                                                        07400000
      N26=N25+NG                                                        07500000
      N27=N26+NG                                                        07600000
      N28=N27+NY                                                        07700000
      N29=N28+NL                                                        07800000
      N30=N29+NY*NL                                                     07900000
      N31=N30+NY*NG                                                     08000000
      N32=N31+NL*NG                                                     08100000
      N33=N32+NL*NY*NG                                                  08200000
      N34=N33+NG                                                        08300000
C                                                                       08400000
C     A CHECK IS MADE FOR SUFFICIENT ARRAY STORAGE SPACE                08500000
C                                                                       08600000
      LIMDAT=NR*NL*NY*NG                                                08700000
      IF(LIMD.LE.LIMDAT) GO TO 10                                       08800000
      IF(N33.LE.LIMZ) GO TO 20                                          08900000
      WRITE(21,5)N33                                                    09000000
    5 FORMAT (/"THE Z ARRAY IS TOO SMALL FOR THIS DATA.IT MUST BE       09100000
     1GREATER THAN OR EQUAL TO ",I5,/)                                  09200000
      CALL EXIT                                                         09300000
   10 WRITE(21,15)LIMD,LIMDAT                                           09400000
   15 FORMAT(/"THE DATA ARRAY STORAGE(",I5,")IS TOO SMALL               09500000
     1IT MUST BE GREATER THAN OR EQUAL TO(",I5,")"/)                    09600000
      CALL EXIT                                                         09700000
   20 CONTINUE                                                          09800000
      READ(5,26)FMT                                                     09900000
   26 FORMAT(13A6)                                                      10000000
C                                                                       10100000
C     NA=SUBSCRIPT OF VARIABLE TO BE ANALYSED.                          10200000
C     IT=TRANSFORMATION CODE.                                           10300000
C     NM=NAME OF VARIABLE TO BE ANALYSED.                               10400000
C                                                                       10500000
   21 READ(5,25)NA,IT,NM                                                10600000
      KOUNT = KOUNT +1                                                  10700000
   25 FORMAT (2I6,2A6)                                                  10800000
   30 LY=NT/NY                                                          10900000
      LL=LY/NL                                                          11000000
      LG=LL/NG                                                          11100000
```

```
      L(1)=NY                                                                   11200000
      L(2)=NL                                                                   11300000
      L(3)=NG                                                                   11400000
      L(4)=NR                                                                   11500000
C     INITIALISE Z ARRAY FOR USE IN ANAL                                        11600000
      DO 35 I=1,N34                                                             11700000
   35 Z(I)=0.0                                                                  11800000
      MT=NY*NL*NR                                                               11900000
C                                                                               12000000
C     IY=YEAR IDENTIFIER, IL=LOCATION IDENTIFIER, IR=REP.IDENTIFIER.            12100000
C     IG=GENOTYPE IDENTIFIER.                                                   12200000
C     READ IN DATA FOR EACH YEAR ,LOCATION &REP.                                12300000
C                                                                               12400000
      IF (NA-1)39,39,48                                                         12500000
   39 DO 38 J=1,MT                                                              12600000
   40 READ (20,41)IY,IL,IR                                                      12700000
   41 FORMAT(3I6)                                                               12800000
C     FIRST FIVE COLUMNS ARE FOR YRS. SITES REPS CODE.                          12900000
   45 DO 38 K=1,NG                                                              13000000
      DO 42 M = 1,NP                                                            13100000
   42 READ(20,FMT)IG,(TEMP(I,M),I=1,NV)                                         13200000
      DO 38 I = 1,NV                                                            13300000
      NLIST = 0                                                                 13400000
      SUM = 0                                                                   13500000
      DO 47 M = 1,NP                                                            13600000
      IF (TEMP(I,M))47,46,46                                                    13700000
   46 NLIST = NLIST + 1                                                         13800000
      SUM = SUM + TEMP(I,M)                                                     13900000
   47 CONTINUE                                                                  14000000
      N(K,J)=LY*(IY-1)+LL*(IL-1)+LG*(IG-1)+IR                                   14100000
   38 XIN(J,K,I)=SUM/NLIST                                                      14200000
   48 DO 55 J=1,MT                                                              14300000
0     DO 55 K=1,NG                                                              14400000
      ZZ=XIN(J,K,NA)                                                            14500000
C     TRANSFORM THE DATA, & DETERMINE ITS POSITION IN THE DATA ARRAY.           14600000
      IF (IT-1)50,50,49                                                         14700000
   49 CALL TRANS(ZZ,IT)                                                         14800000
   50 I=N(K,J)                                                                  14900000
   55 DATA(I)=ZZ                                                                15000000
      IF (IY*IL*IR .LT. MT )GO TO 40                                            15100000
      WRITE(21,72)                                                              15200000
   65 WRITE(21,66)NM                                                            15300000
   66 FORMAT(10X,9HCHARACTER,2A6///)                                            15400000
      WRITE(21,67)                                                              15500000
   67 FORMAT(//5X,15HGENOTYPES ARE :/)                                          15600000
      WRITE(21,68)(I,(KG(O,I),O=1,5),I=1,NG)                                    15700000
   68 FORMAT(6X,I5,2X,5A6 /)                                                    15800000
      WRITE(21,69)                                                              15900000
   69 FORMAT (//5X,11HSITES ARE :/)                                             16000000
      WRITE (21,71) (I,(KS(O,I),O=1,5),I=1,NL)                                  16100000
   71 FORMAT(6X,I5,2X,5A6/)                                                     16200000
      WRITE(21,73)                                                              16300000
   73 FORMAT(//5X,11HYEARS ARE :/)                                              16400000
      WRITE(21,71)(I,(KY(O,I),O=1,5),I=1,NY)                                    16500000
      WRITE(21,72)                                                              16600000
   72 FORMAT(1H1)                                                               16700000
      CALL ANAL(DATA,L,Z(1),Z(N1),Z(N2),Z(N3),Z(N4),Z(N5),Z(N6),Z(N7),          16800000
     1Z(N8),Z(N9),Z(N10),Z(N11),Z(N12),Z(N13),Z(N14),Z(N15),Z(N16),             16900000
     2Z(N17),Z(N18),Z(N19),Z(N20),Z(N21),Z(N22),Z(N23),Z(N24),NY,NL,NG,         17000000
     3NR,Z(N25),Z(N26),Z(N27),Z(N28),Z(N29),Z(N30),Z(N31),Z(N32),Z(N33),        17100000
     4Z(N34))                                                                   17200000
```

```
      GO TO 4                                                          17300000
70    WRITE(21,75)                                                     17400000
75    FORMAT(//1X,"END OF RUN")                                        17500000
      RETURN                                                           17600000
      END                                                              17700000
```

```
LIST SYMBOL/ANAL/TWO

DATE 03/02/81          TIME IS 15:28

SYSTEM/DUMPALL VERSION  3.0.140

LASTRECORD =     523
MAXRECSIZEIN =    14    BLOCKSIZEIN =  420
*** EBCDIC ***          UNITS=WORDS

      1          SUBROUTINE ANAL(DATA,L,A,B,C,D,AB,AC,AD,BC,BD,CD,ABC,ABD,ACD,      00100000
      2         1BCD,X,Y,BETA,REG,DEV,SSQ,CON,DEVM,F1,R,STAB,NY1,NL1,NG1,NR1,SST,   00200000
      3         2RATIO,FSIG,X2,X3,X4,Y2,Y3,Y4,YS)                                   00300000
      4    C                                                                        00400000
      5          DOUBLE PRECISION G,NIL,SSE,MSE,PMSE                              00500000
      6    C                                                                        00600000
      7          INTEGER DFE,PDFE,ID                                              00700000
      8          DIMENSION A(NY1),B(NL1),C(NG1),D(NR1),AB(NY1,NL1),AC(NY1,NG1),   00800000
      9         1AD(NY1,NR1),BC(NL1,NG1),BD(NL1,NR1),CD(NG1,NR1),ABC(NY1,NL1,NG1),00900000
     10         2ABD(NY1,NL1,NR1),ACD(NY1,NG1,NR1),BCD(NL1,NG1,NR1),L(4),S(20),   01000000
     11         3G(16),ID(16),F(16),P(16),E(15),DATA(1),ALP(16),X(NY1,NL1),       01100000
     12         4Y(NY1,NL1,NG1),BETA(NG1),REG(NG1),DEV(NG1),SSQ(NG1),CON(NG1),    01200000
     13         5DEVM(NG1),F1(NG1),R(NG1),STAB(NG1),FORM(14),FSIGNF(16),FDF1(10),  01300000
     14         6FDF2(16),SSG(5,25),SSB(5,25),M(5,25),U(5,25),CF(5,25),SST(NY1,NL1)01400000
     15         7,RATIO(NG1),FSIG(NG1),X2(NY1),X3(NL1),X4(NY1,NL1),Y2(NY1,NG1),Y3(  01500000
     16         8NL1,NG1),Y4(NY1,NL1,NG1),YS(NG1)                                   01600000
     17    C                                                                        01700000
     18          DATA ALP/'YEARS','LOCATS','Y*LOCS','ENVTOT','BLOCKS','GENOTS',     01800000
     19         1'GEN*Y','GEN*L','G*L*Y','GEN*E','R*GEN','R*G*Y','R*G*L','R*GLY',' 01900000
     20         2ERROR','TOTAL'/                                                    02000000
     21          COMMON/ERRM/SSE(5,25),MSE(5,25),DFE,JSIG                         02100000
     22    C                                                                        02200000
     23    C     THIS S/R PERFORMS THE CALCULATIONS FOR THE A. O. V. & REGRESSIONS  02300000
     24    C     IF DESIRED HANSONS STABILITY PARAMETERS ARE CALCULATED             02400000
     25    C                                                                        02500000
     26    C     FUNCTIONS DFCF ,PRBF, SIGNIF ARE NEEDED FOR SIGNIFICANCE TESTS     02600000
     27    C                                                                        02700000
     28    C     CODE:- A=YEARS,                                                    02800000
     29    C            B=LOCATIONS,                                              02900000
     30    C            C=GENOTYPES,                                              03000000
     31    C            D=REPLICATIONS.                                           03100000
     32    C                                                                        03200000
     33    C     L ARRAY CONTAINS DATA PARAMETERS                                   03300000
     34    C     DATA IS STORED IN NESTED FORM YEARS(LOCATIONS(REPS(GENOTYPES)))    03400000
     35    C                                                                        03500000
     36          NY=L(1)                                                          03600000
     37          NL=L(2)                                                          03700000
     38          NG=L(3)                                                          03800000
     39          NR=L(4)                                                          03900000
     40          NT=NY*NL*NG*NR                                                   04000000
     41    C     INITIALISE S ARRAY                                               04100000
     42          DO 10 I=1,20                                                     04200000
     43       10 S(I)=0.0                                                         04300000
     44          DO 11 I=1,NY                                                     04400000
     45          DO 11 J=1,NL                                                     04500000
     46          DO 11 K=1,NG                                                     04600000
     47          DO 11 M=1,NR                                                     04700000
     48          A(I)=0.0                                                         04800000
     49          B(J)=0.0                                                         04900000
     50          C(K)=0.0                                                         05000000
```

```
      D(M)=0.0                                                          05100000
      AB(I,J)=0.0                                                       05200000
      AC(I,K)=0.0                                                       05300000
      AD(I,M)=0.0                                                       05400000
      BC(J,K)=0.0                                                       05500000
      BD(J,M)=0.0                                                       05600000
      CD(K,M)=0.0                                                       05700000
      ABC(I,J,K)=0.0                                                    05800000
      ABD(I,J,M)=0.0                                                    05900000
      ACD(I,K,M)=0.0                                                    06000000
C                                                                       06100000
      BCD(J,K,M)=0.0                                                    06200000
   11 SST(I,J)=0.0                                                      06300000
C     FIND CLASS TOTALS                                                 06400000
      DO 15 I=1,NT                                                      06500000
      CALL INDEX(I,IY,IL,IG,IR,L)                                       06600000
      T=DATA(I)                                                         06700000
      A(IY)=A(IY)+T                                                     06800000
      B(IL)=B(IL)+T                                                     06900000
      C(IG)=C(IG)+T                                                     07000000
      D(IR)=D(IR)+T                                                     07100000
      SST(IY,IL)=SST(IY,IL) +T*T                                        07200000
      AB(IY,IL)=AB(IY,IL)+T                                             07300000
      AC(IY,IG)=AC(IY,IG)+T                                             07400000
      AD(IY,IR)=AD(IY,IR)+T                                             07500000
      BC(IL,IG)=BC(IL,IG)+T                                             07600000
      BD(IL,IR)=BD(IL,IR)+T                                             07700000
      CD(IG,IR)=CD(IG,IR)+T                                             07800000
      ABC(IY,IL,IG)=ABC(IY,IL,IG)+T                                     07900000
      ABD(IY,IL,IR)=ABD(IY,IL,IR)+T                                     08000000
      ACD(IY,IG,IR)=ACD(IY,IG,IR)+T                                     08100000
      BCD(IL,IG,IR)=BCD(IL,IG,IR)+T                                     08200000
      S(19)=S(19)+T                                                     08300000
   15 S(20)=S(20)+T*T                                                   08400000
C     E ARRAY CONTAINS DIVISORS FOR S.S. CALCULATIONS.                  08500000
      DO 20 I=1,4                                                       08600000
   20 E(I)=NT/L(I)                                                      08700000
      E(5)=E(1)/L(2)                                                    08800000
      E(6)=E(1)/L(3)                                                    08900000
      E(7)=E(1)/L(4)                                                    09000000
      E(8)=E(2)/L(3)                                                    09100000
      E(9)=E(2)/L(4)                                                    09200000
      E(10)=E(3)/L(4)                                                   09300000
      E(11)=E(5)/L(3)                                                   09400000
      E(12)=E(5)/L(4)                                                   09500000
      E(13)=E(6)/L(4)                                                   09600000
      E(14)=E(8)/L(4)                                                   09700000
      E(15)=NT                                                          09800000
C     COMPUTE ALL UNADJUSTED SUM OF SQUARES                             09900000
      CT=S(19)*S(19)/E(15)                                              10000000
      DO 40 I=1,NY                                                      10100000
      S(1)=S(1)+A(I)*A(I)/E(1)                                          10200000
      DO 31 J=1,NL                                                      10300000
      S(5)=S(5)+AB(I,J)*AB(I,J)/E(5)                                    10400000
      W(I,J)=0.0                                                        10500000
      U(I,J)=0.0                                                        10600000
      DO 25 K=1,NG                                                      10700000
C                                                                       10800000
      S(11)=S(11)+ABC(I,J,K)**2/E(11)                                   10900000
   25 W(I,J)=W(I,J)+ABC(I,J,K)**2/E(11)                                 11000000
C                                                                       11100000
```

```
      C                                                                 11200000
            DO 31 K=1,NR                                                11300000
      C                                                                 11400000
         30 S(12)=S(12)+ABD(I,J,K)*ABD(I,J,K)/E(12)                     11500000
         31 U(I,J)=U(I,J)+ABD(I,J,K)**2/E(12)                           11600000
            DO 35 J=1,NG                                                11700000
            S(6)=S(6)+AC(I,J)*AC(I,J)/E(6)                              11800000
            DO 35 K=1,NR                                                11900000
         35 S(13)=S(13)+ACD(I,J,K)*ACD(I,J,K)/E(13)                     12000000
            DO 40 J=1,NR                                                12100000
         40 S(7)=S(7)+AD(I,J)*AD(I,J)/E(7)                              12200000
            DO 50 I=1,NL                                                12300000
            S(2)=S(2)+B(I)*B(I)/E(2)                                    12400000
            DO 45 J=1,NG                                                12500000
            S(8)=S(8)+BC(I,J)*BC(I,J)/E(8)                              12600000
            DO 45 K=1,NR                                                12700000
         45 S(14)=S(14)+BCD(I,J,K)*BCD(I,J,K)/E(14)                     12800000
            DO 50 J=1,NR                                                12900000
         50 S(9)=S(9)+BD(I,J)*BD(I,J)/E(9)                              13000000
            DO 55 I=1,NG                                                13100000
            S(3)=S(3)+C(I)*C(I)/E(3)                                    13200000
            DO 55 J=1,NR                                                13300000
         55 S(10)=S(10)+CD(I,J)*CD(I,J)/E(10)                           13400000
            DO 60 I=1,NR                                                13500000
         60 S(4)=S(4)+D(I)*D(I)/E(4)                                    13600000
      C     COMPUTE ALL CORRECTED SUMS OF SQUARES                       13700000
      C     IN P ARRAY.                                                 13800000
            P(1)=S(1)-CT                                                13900000
            P(2)=S(2)-CT                                                14000000
            P(4)=S(5)-CT                                                14100000
            P(3)=P(4)-(P(1)+P(2))                                       14200000
            P(5)=S(12)-CT-P(4)                                          14300000
            P(6)=S(3)-CT                                                14400000
            P(7)=S(6)-CT-P(1)-P(6)                                      14500000
            P(8)=S(8)-CT-P(2)-P(6)                                      14600000
            P(9)=S(11)-CT-P(4)-P(7)-P(8)-P(6)                           14700000
            P(10)=P(7)+P(8)+P(9)                                        14800000
            P(11)=S(10)-S(4)-P(6)                                       14900000
            P(12)=S(13)-S(7)-S(10)+S(4)-P(7)                            15000000
            P(13)=S(14)-S(9)-P(6)-P(8)-P(11)                            15100000
            P(16)=S(20)-CT                                              15200000
            XX=0.0                                                      15300000
            DO 65 I=1,13                                                15400000
         65 XX=XX+P(I)                                                  15500000
            P(14)=P(16)-XX+P(10)+P(4)                                   15600000
            P(15)=P(11)+P(12)+P(13)+P(14)                               15700000
      C     COMPUTE THE DEGREES OF FREEDOM IN ID ARRAY.                 15800000
            ID(1)=NY-1                                                  15900000
            ID(2)=NL-1                                                  16000000
            ID(3)=ID(1)*ID(2)                                           16100000
            ID(4)=(NY*NL)-1                                             16200000
            ID(5)=(NY*NL)*(NR-1)                                        16300000
            ID(6)=NG-1                                                  16400000
            ID(7)=(NY-1)*(NG-1)                                         16500000
            ID(8)=(NL-1)*(NG-1)                                         16600000
            ID(9)=(NY-1)*(NL-1)*(NG-1)                                  16700000
            ID(10)=ID(7)+ID(8)+ID(9)                                    16800000
            ID(11)=(NG-1)*(NR-1)                                        16900000
            ID(12)=ID(11)*(NY-1)                                        17000000
            ID(13)=ID(11)*(NL-1)                                        17100000
            ID(14)=ID(13)*(NY-1)                                        17200000
```

```
      ID(15)=ID(11)+ID(12)+ID(13)+ID(14)                                17300000
      ID(16)=NY*NL*NG*NR-1                                              17400000
C     TEST FOR HOMOGENEITY OF ERRORS.                                   17500000
      JSIG=1                                                            17600000
      PDFE = ID(15)                                                     17700000
      DFE=ID(11)                                                        17800000
C                                                                       17900000
      DO 70 I=1,NY                                                      18000000
      DO 70 J=1,NL                                                      18100000
      SSB(I,J)=0.0                                                      18200000
      SSG(I,J)=0.0                                                      18300000
      SSE(I,J)=0.0                                                      18400000
      MSE(I,J)=0.0                                                      18500000
      CF(I,J)=0.0                                                       18600000
      CF(I,J)=AB(I,J)**2/(NG*NR)                                        18700000
      SSG(I,J)=W(I,J)-CF(I,J)                                           18800000
      SSB(I,J)=U(I,J)-CF(I,J)                                           18900000
      SSE(I,J)=SST(I,J)-CF(I,J)-SSG(I,J)-SSB(I,J)                       19000000
      MSE(I,J)=SSE(I,J)/DFE                                             19100000
C                                                                       19200000
C                                                                       19300000
   70 CONTINUE                                                          19400000
      CALL HOMER(NY,NL,PDFE,PMSE)                                       19500000
C     COMPUTE MEAN SQUARES IN G ARRAY.                                  19600000
      DO 75 I=1,16                                                      19700000
      IF (ID(I).EQ.0) ID(I)=1                                           19800000
   75 G(I)=P(I)/ID(I)                                                   19900000
C     COMPUTE COMPLEX F-TESTS IN F ARRAY.                               20000000
      IF(NY.EQ.1) GO TO 76                                              20100000
      F(1)=(G(1)+G(9))/(G(3)+G(7))                                      20200000
      F(2)=(G(2)+G(9))/(G(3)+G(8))                                      20300000
      F(3 )=(G(3)+G(15))/(G(5)+G(9))                                    20400000
      F(4)=G(4)/G (15)                                                  20500000
      F(5)=G(5)/G(15)                                                   20600000
      F(6)=(G(6)+G(9))/(G(7)+G(8))                                      20700000
      F(7)=G(7 )/G(9 )                                                  20800000
      F(8)=G(8 )/G(9)                                                   20900000
      F(9)=G(9)/G(15)                                                   21000000
      F(10)=G(10)/G(15)                                                 21100000
      GO TO 79                                                          21200000
C     ALTERNATIVE F TESTS IF YEARS = 1.                                 21300000
   76 F(1)=0.0                                                          21400000
      F(2)=(G(2)+G(15))/(G(5)+G(10))                                    21500000
      F(3)=0.0                                                          21600000
      F(4)=F(2)                                                         21700000
      F(5)=G(5)/G(15)                                                   21800000
      F(6)=G(6)/G(10)                                                   21900000
      F(7)=0.0                                                          22000000
      F(8)=G(8)/G(15)                                                   22100000
      F(9)=0.0                                                          22200000
      F(10)=G(10)/G(15)                                                 22300000
   79 CONTINUE                                                          22400000
      DO 80 I=11,16                                                     22500000
   80 F (I)=0.0                                                         22600000
C                                                                       22700000
C                                                                       22800000
C     FIND THE DEGREES OF FREEDOM.                                      22900000
      NIL =0.0                                                          23000000
      IF (NY.EQ.1) GO TO 81                                             23100000
      FDF1(1)=DFCF(G(1),G (9),NIL,NIL,ID(1 ),ID(9),0,0 )                23200000
      FDF2(1)=DFCF(G(3),G(7),NIL,NIL,ID(3),ID(7),0,0 )                  23300000
```

```
      FDF1 (2)=DFCF(G(2),G(9),NIL,NIL,ID(2),ID(9),0,0)                  23400000
      FDF2(2)=DFCF(G(3),G(8),NIL,NIL,ID(3),ID(8),0,0 )                  23500000
      FDF1(3)=DFCF(G(3),G(15),NIL,NIL,ID(3),ID (15),0,0)                23600000
      FDF2(3)=DFCF (G(5),G(9),NIL,NIL,ID(5),ID(9),0,0)                  23700000
      FDF1(4)=ID(4)                                                     23800000
      FDF2(4)=ID (15)                                                   23900000
      FDF1(5)=ID(5 )                                                    24000000
      FDF2(5)=ID(15)                                                    24100000
      FDF1(6)=DFCF(G(6),G(15),NIL,NIL,ID(6),ID(15),0,0)                 24200000
      FDF2(6)=DFCF(G(7),G(8),NIL,NIL,ID(7),ID (8),0,0)                  24300000
      FDF1(7)=ID(7)                                                     24400000
      FDF2(7)=ID(9)                                                     24500000
      FDF1(8)=ID(8)                                                     24600000
      FDF2(8)=ID(9)                                                     24700000
      FDF1(9)=ID(9)                                                     24800000
      FDF2(9)=ID(15)                                                    24900000
      FDF1(10)=ID(10)                                                   25000000
      FDF2(10)=ID(15)                                                   25100000
      GO TO 89                                                          25200000
C       ALTERNATIVE DEGREES OF FREEDOM IF YEARS = 1.                    25300000
   81 FDF1(1)=0                                                         25400000
      FDF2(1)=0                                                         25500000
      FDF1(2)=DFCF(G(2),G(15),NIL,NIL,ID(2),ID(15),0,0)                 25600000
      FDF2(2)=DFCF(G(5),G(10),NIL,NIL,ID(5),ID(10),0,0)                 25700000
      FDF1(3)=0                                                         25800000
      FDF2(3)=0                                                         25900000
      FDF1(4)=FDF1(2)                                                   26000000
      FDF2(4)=FDF2(2)                                                   26100000
      FDF1(5)=ID(5)                                                     26200000
      FDF2(5)=ID(15)                                                    26300000
      FDF1(6)=ID(6)                                                     26400000
      FDF2(6)=ID(10)                                                    26500000
      FDF1(7)=0                                                         26600000
      FDF2(7)=0                                                         26700000
      FDF1(8)=ID(8)                                                     26800000
      FDF2(8)=ID(15)                                                    26900000
      FDF1(9)=0                                                         27000000
      FDF2(9)=0                                                         27100000
      FDF1(10)=ID(10)                                                   27200000
      FDF2(10)=ID(15)                                                   27300000
   89 CONTINUE                                                          27400000
C     FIND SIGNIFICANCE OF F TEST.                                      27500000
      DO 90 I=1,10                                                      27600000
   90 FSIGNF(I)=PRBF(FDF1(I),FDF2(I),F(I))                              27700000
      DO 95 I=1,10                                                      27800000
   95 FSIGNF (I)=SIGNIF(FSIGNF(I))                                      27900000
      DO 100 I=11,16                                                    28000000
  100 FSIGNF(I)='    '                                                  28100000
      DO 105 I=11,16                                                    28200000
      FDF1(I)= 0                                                        28300000
  105 FDF2(I)= 0                                                        28400000
C     WRITE OUT THE ANOVA TABLE.                                        28500000
      WRITE(21,109)                                                     28600000
  109 FORMAT(1H1)                                                       28700000
      WRITE (21,110)                                                    28800000
  110 FORMAT (/24X,'POOLED ANALYSIS OF VARIANCE'//)                     28900000
      WRITE(21,115)                                                     29000000
  115 FORMAT(24X,6HSOURCE,4X,2HDF,3X,10HSUM OF SQS,2X,11HMEAN SQUARE,   29100000
     15X,1HF,5X,6HSIGNIF,2X,7HTEST DF/)                                 29200000
      DO 120 I=1,16                                                     29300000
  120 WRITE(21,125)ALP(I),ID(I),P(I),G(I),F(I),FSIGNF(I),FDF1(I),FDF2(I)29400000
```

```
      125 FORMAT(24X,A6,2X,I4,1X,F12.2,2X,F11.2,2X,F7.2,2X,A4,3X,I3,I4/)
          GM=S(19)/E(15)
          EMS =G(15)
          CV=SQRT(EMS)*100.0/GM
          WRITE(21,130)CV
      130 FORMAT(//20X,'COEFFICIENT OF VARIATION',3X,F10.1)
          WRITE(21,135)
      135 FORMAT(1H1)
          WRITE(21,140)
      140 FORMAT(///,' TABLE OF MEANS'//)
C         CONVERT ALL TOTALS TO MEANS.
          DO 160 I=1,NY
          A(I)=A(I)/E(1)
          DO 150 J=1,NL
          AB(I,J)=AB(I,J)/E(5)
          DO 145 K=1,NG
      145 ABC(I,J,K)=ABC(I,J,K)/E(11)
          DO 150 K=1,NR
      150 ABD(I,J,K)=ABD(I,J,K)/E(12)
          DO 155 J=1,NG
          AC(I,J)=AC(I,J)/E(6)
          DO 155 K=1,NR
      155 ACD(I,J,K)=ACD(I,J,K)/E(13)
          DO 160 J=1,NR
      160 AD(I,J)=AD(I,J)/E(7)
          DO 170 I=1,NL
          B(I)=B(I)/E(2)
          DO 165 J=1,NG
          BC(I,J)=BC(I,J)/E(8)
          DO 165 K=1,NR
      165 BCD(I,J,K)=BCD(I,J,K)/E(14)
          DO 170 J=1,NR
      170 BD(I,J)=BD(I,J)/E(9)
          DO 175 I=1,NG
          C(I)=C(I)/E(3)
          DO 175 J=1,NR
      175 CD(I,J)=CD(I,J)/E(10)
          DO 180 I=1,NR
      180 D(I)=D(I)/E(4)
      185 FORMAT (8(11F10.2/))
C         WRITE OUT THE TABLE OF MEANS
          WRITE(21,190)
      190 FORMAT(2X,10HGRAND MEAN/)
          WRITE(21,185)GM
          WRITE(21,195)
      195 FORMAT(//2X,5HYEARS,/)
          WRITE(21,185)(A(I),I=1,NY)
          WRITE(21,200)
      200 FORMAT(//2X,5HSITES,/)
          WRITE(21,185)(B(I),I=1,NL)
          WRITE(21,205)
      205 FORMAT(//2X,9HGENOTYPES,/)
          WRITE(21,185)(C(I),I=1,NG)
          WRITE(21,210)
      210 FORMAT(//2X,14HSITES BY YEARS,/)
          DO 215 I=1,NY
      215 WRITE(21,185)(AB(I,J),J=1,NL)
          WRITE(21,220)
      220 FORMAT(//2X,18HYEARS BY GENOTYPES/)
          DO 225 I=1,NY
      225 WRITE(21,185)(AC(I,J),J=1,NG)
```

```
      WRITE(21,230)                                                     35600000
  230 FORMAT(//2X,18HSITES BY GENOTYPES/)                               35700000
      DO 235 I=1,NL                                                     35800000
  235 WRITE(21,185)(BC(I,J),J=1,NG)                                     35900000
      WRITE(21,240)                                                     36000000
  240 FORMAT(//2X,27HYEARS BY SITES BY GENOTYPES/)                      36100000
      DO 245 I=1,NY                                                     36200000
      DO 245 J=1,NL                                                     36300000
  245 WRITE(21,185)(ABC(I,J,K),K=1,NG)                                  36400000
      READ(5,255) (FORM(I),I=1,14)                                      36500000
  255 FORMAT(14A6)                                                      36600000
C     WRITE OUT THE TABLE OF GENOTYPE * ENVIRONMENT MEANS.              36700000
      DO 260 K=1,NG                                                     36800000
  260 WRITE(22,FORM)((ABC(I,J,K),J=1,NL),I=1,NY)                        36900000
      WRITE(21,135)                                                     37000000
      SPE=P(15)/E(11)                                                   37100000
      SE=P(4)/E(11)                                                     37200000
      SG=P(6)/E(11)                                                     37300000
      SEG=P(10)/E(11)                                                   37400000
      WRITE(21,264)                                                     37500000
  264 FORMAT(1H1)                                                       37600000
      READ(5,263) REGOP                                                 37700000
  263 FORMAT(I1)                                                        37800000
      WRITE(21,265)                                                     37900000
  265 FORMAT(//5X,'ENV. INDICES'//5X,'(ENV.MN.-GRD.MN.)'//5X,2HYR,1X,   38000000
     13HLOC/)                                                           38100000
C     CALCULATE ENVIRONMENTAL INDICES.                                  38200000
      DO 275 I=1,NY                                                     38300000
      DO 275 J=1,NL                                                     38400000
      X(I,J)=AB(I,J)-GM                                                 38500000
  275 WRITE(21,270)I,J,X(I,J)                                           38600000
  270 FORMAT(6X,I1,2X,I2,1X,F10.2/)                                     38700000
      IF(REGOP.EQ.1) GO TO 278                                          38800000
C     ALTERNATIVE ENV. INDICES.                                         38900000
      DO 277 I=1,NY                                                     39000000
      X2(I)=A(I)-GM                                                     39100000
      DO 277 J=1,NL                                                     39200000
      X4(I,J)=AB(I,J)-A(I)-B(J)+GM                                      39300000
  277 X3(J)=B(J)-GM                                                     39400000
C     BEGIN THE REGRESSION ANALYSIS.                                    39500000
  278 WRITE(21,281)                                                     39600000
  281 FORMAT(//5X,'G*E EFFECTS'//)                                      39700000
      ECOMIN=10000000.0                                                 39800000
      DO 279 K=1,NG                                                     39900000
      SY=0.0                                                            40000000
      SXX=0.0                                                           40100000
      SXY=0.0                                                           40200000
      SYY=0.0                                                           40300000
      CON(K)=0.0                                                        40400000
      SSQ(K)=0.0                                                        40500000
      REG(K)=0.0                                                        40600000
      DEV(K)=0.0                                                        40700000
      BETA(K)=0.0                                                       40800000
      DO 280 I=1,NY                                                     40900000
      DO 280 J=1,NL                                                     41000000
      Y(I,J,K)=ABC(I,J,K)-C(K)                                          41100000
      YS(K)=YS(K)+Y(I,J,K)*Y(I,J,K)                                     41200000
      SY=SY+Y(I,J,K)                                                    41300000
      SXX=SXX+X(I,J)*X(I,J)                                             41400000
      SXY=SXY+X(I,J)*Y(I,J,K)                                           41500000
      Y(I,J,K)=ABC(I,J,K)-AB(I,J)-C(K)+GM                               41600000
```

```
      280 SYY=SYY+Y(I,J,K)*Y(I,J,K)                                                  41700000
C                                                                                    41800000
C                                                                                    41900000
C                                                                                    42000000
C                                                                                    42100000
          DO 271 I=1,NY                                                              42200000
      271 WRITE(21,283)(Y(I,J,K),J=1,NL)                                             42300000
      283 FORMAT(8(11F10.2/))                                                        42400000
C                                                                                    42500000
          BETA(K)=SXY/SXX                                                            42600000
          SSQ(K)=SYY                                                                 42700000
C                                                                                    42800000
C                                                                                    42900000
C                                                                                    43000000
          REG(K)=BETA(K)*SXY                                                         43100000
          DEV(K)=YS(K)-REG(K)                                                        43200000
          IF (REGOP.EQ.1) GO TO 279                                                  43300000
C         ALTERNATIVE REG. ANALYSIS.                                                 43400000
          DO 284 I=1,NY                                                              43500000
          Y2(I,K)=AC(I,K)-A(I)-C(K)+GM                                               43600000
          WRITE(19,282) X2(I),Y2(I,K)                                                43700000
      282 FORMAT(20(F10.2,2X,F10.2/))                                                43800000
          DO 284 J=1,NL                                                              43900000
          Y4(I,J,K)=ABC(I,J,K)-AB(I,J)-AC(I,K)-BC(J,K)+C(K)+A(I)+B(J)-GM             44000000
          WRITE(19,282) X4(I,J),Y4(I,J,K)                                            44100000
          Y3(J,K)=BC(J,K)-B(J)-C(K)+GM                                               44200000
      284 WRITE(19,282)X3(J),Y3(J,K)                                                 44300000
      279 CONTINUE                                                                   44400000
          DFDEV=E(10)-2                                                              44500000
          SUMSSQ=0.0                                                                 44600000
          SUMDEV=0.0                                                                 44700000
          SUMREG=0.0                                                                 44800000
          DO 290 K=1,NG                                                              44900000
          SUMSSQ=SUMSSQ+SSQ(K)                                                       45000000
          SUMDEV=SUMDEV+DEV(K)                                                       45100000
          SUMREG=SUMREG+REG(K)                                                       45200000
          DEVM(K)=ABS(DEV(K)/DFDEV)                                                  45300000
          F1(K)=REG(K)/DEVM(K)                                                       45400000
          FSIG(K)=PRBF(1,E(10)-2,F1(K))                                              45500000
          FSIG(K)=SIGNIF(FSIG(K))                                                    45600000
          R(K)=100.*REG(K)/SSQ(K)                                                    45700000
      290 STAB(K)=1./SSQ(K)                                                          45800000
          WRITE(21,295)SE                                                            45900000
          WRITE(21,300)SG                                                            46000000
      295 FORMAT(//20X,'ENVIR SSQ',3F13.2//)                                         46100000
      300 FORMAT(20X,'GENOT SSQ',3F13.2)                                             46200000
C         WRITE OUT THE REGRESSION STABILITY PARAMETERS FOR EACH GENOTYPE.           46300000
          WRITE(21,264)                                                              46400000
          WRITE(21,305)                                                              46500000
      305 FORMAT(///45X,'REGRESSION STABILITY PARAMETERS')                           46600000
          WRITE(21,310)                                                              46700000
      310 FORMAT(//25X,'G',5X,'REG SS',7X,'DEV SS',7X,'DEV MS',5X,'F',4X,'SI'46800000
     1GNIF',//)                                                                      46900000
          DO 315 K=1,NG                                                              47000000
      315 WRITE(21,320)K,REG(K),DEV(K),DEVM(K),F1(K),FSIG(K)                         47100000
      320 FORMAT(23X,I3,3(F11.2,2X),F6.2,3X,A4)                                      47200000
          CALL BART(NG,DEVM,DFDEV)                                                   47300000
          WRITE(21,325)                                                              47400000
      325 FORMAT(//24X,'NG',2X,'TOT REG SS',2X,'TOT DEV SS')                         47500000
          WRITE(21,341)NG,SUMREG,SUMDEV                                              47600000
      341 FORMAT(23X,I3,2(F11.2,2X)//)                                               47700000
```

```
      WRITE(21,326)                                                     47800000
  326 FORMAT(//25X,'G',1X,' G*E SSQS.',2X,'RATIO TO LEAST SSQS',3X,'REG47900000
     1 % G*E',6X,'(1/G*E)',6X,'BETA'//)                                 48000000
C                                                                       48100000
      DO 327 K=1,NG                                                     48200000
      RATIO(K)=SSQ(K)/ECOMIN                                            48300000
  327 WRITE(21,328)K,SSQ(K),RATIO(K),R(K),STAB(K),BETA(K)               48400000
  328 FORMAT(23X,I3,F11.2,5X,F7.2,11X,F11.2,2X,E11.4,2X,F11.2)          48500000
      PGEAR=(SUMREG/SUMSSQ)*100.0                                       48600000
      WRITE(21,329)                                                     48700000
  329 FORMAT(//24X,'NG',1X,'TOT G*E SS',12X,'% G*E SS @ BY REG')        48800000
      WRITE(21,342)NG,SUMSSQ,PGEAR                                      48900000
  342 FORMAT(/23X,I3,F11.2,12X,F13.2)                                   49000000
      READ(5,340) ANS                                                   49100000
  340 FORMAT(A1)                                                        49200000
      IF (ANS.EQ.'N')GO TO 385                                          49300000
      WRITE(21,345)                                                     49400000
  345 FORMAT(1H1,///,45X,'HANSON STABILITY PARAMETERS'//)               49500000
C     C ARRAY CONTAINS GENOTYPE MEANS.                                  49600000
C     ABC ARRAY HAS G*E MEANS.                                          49700000
      READ(5,350)ALPHA                                                  49800000
  350 FORMAT(F4.2)                                                      49900000
      WRITE(21,355)                                                     50000000
  355 FORMAT(40X,'GENOTYPE',16X,'RELATIVE STABILITY'/)                  50100000
      DO 380 KA=1,NG                                                    50200000
      LIM=NY*NL                                                         50300000
      QA=0.0                                                            50400000
      DO 360 I=1,NY                                                     50500000
      DO 360 J=1,NL                                                     50600000
  360 QA=QA+(ABC(I,J,KA)-C(KA)-X(I,J)*ALPHA)**2                         50700000
      IF (ID(15).EQ.0)GO TO 365                                         50800000
      T1=ALPHA**2-2.0*ALPHA                                             50900000
      T3=NG                                                             51000000
      XKI=(LIM-1)*(T3+T1)/T3                                            51100000
      T1=XKI*SPE/ID(15)                                                 51200000
      DDA=QA-T1                                                         51300000
      IF(DDA.LT.0.0)DDA=0.0                                             51400000
      DDA=SQRT(DDA)                                                     51500000
      GO TO 370                                                         51600000
  365 DDA=SQRT(QA)                                                      51700000
  370 WRITE(21,375)KA,DDA                                               51800000
  375 FORMAT(40X,I5,19X,F14.1)                                          51900000
  380 CONTINUE                                                          52000000
  385 CLOSE(22,DISP=CRUNCH)                                             52100000
      CLOSE(19,DISP=CRUNCH)                                             52200000
      RETURN                                                            52300000
      END                                                               52400000
```

CODON/INDEX

```
                                                                                                    START OF SEGMENT 002
      SUBROUTINE INDEX(I,IY,IL,IG,IR,L)                                                  00100000   C   002:0000:0
C                                                                                        00200000   C   002:0000:0
C     THIS S/R DETERMINES YEAR,LOCATION,GENOTYPE & REP. INDECES FROM                     00300000   C   002:0000:0
C     THEIR POSITION IN THE DATA ARRAY.                                                  00400000   C   002:0000:0
      DIMENSION L(4)                                                                     00500000   C   002:0000:0
      N=1                                                                                00600000   C   002:0000:0
      DO 10 K=1,4                                                                        00700000   C   002:0000:4
   10 N=N*L(K)                                                                           00800000   C   002:0002:0
      LY=N/L(1)                                                                          00900000   C   002:0006:5
      IY=(I-1)/LY+1                                                                      01000000   C   002:0008:4
      LL=LY/L(2)                                                                         01100000   C   002:000A:5
      IT=I-LY*(IY-1)                                                                     01200000   C   002:000D:0
      IL=(IT-1)/LL+1                                                                     01300000   C   002:000F:2
      IS=IT-LL*(IL-1)                                                                    01400000   C   002:0011:1
      LG=LL/L(3)                                                                         01500000   C   002:0013:3
      IG=(IS-1)/LG+1                                                                     01600000   C   002:0015:5
      IR=IS-LG*(IG-1)                                                                    01700000   C   002:0018:0
      RETURN                                                                             01800000   C   002:001A:2
      END                                                                                01900000   C   002:001A:5
                                                                                                    SEGMENT 002 IS 0028 LONG

#######################################################################################################################
                                                                                                    START OF SEGMENT 003
                                                                                                    SEGMENT 003 IS 0004 LONG

      NO ERRORS DETECTED.  NUMBER OF CARDS = 19.
      COMPILATION TIME = 5 SECONDS ELAPSED,  0.84 SECONDS PROCESSING.
      D2 STACK SIZE = 3 WORDS.  FILESIZE = 0 WORDS.  ESTIMATED CORE STORAGE REQUIREMENT = 83 WORDS.
      TOTAL PROGRAM CODE = 59 WORDS.  ARRAY STORAGE = 0 WORDS.
      NUMBER OF PROGRAM SEGMENTS = 3.  NUMBER OF DISK SEGMENTS = 14.
      PROGRAM CODE FILE = (AG0938CULLEN)CODON/INDEX ON PACK.
      COMPILER COMPILED ON 10/31/78 (FORTRAN ON PACK).
```

```
                                                                                    START OF SEGMENT 002
      SUBROUTINE HOMER (NY,NL,PDFE,PMSE)                                 00100000  T  002:0000:0
C                                                                        00200000  T  002:0000:0
C     COMPUTES HOMOGENEITY OF ERROR VARIANCES.                           00300000  T  002:0000:0
C                                                                        00400000  T  002:0000:0
      DOUBLE PRECISION ADDE,POOLE,PSSE,PMSE,SSE,MSE,CMSE,CPMSE           00500000  T  002:0000:0
      INTEGER PDFE,DFE                                                   00600000  T  002:0000:0
      DIMENSION CMSE(5,25)                                               00700000  T  002:0000:0
      COMMON/ERHM/SSE(5,25),MSE(5,25),DFE,JSIG                           00800000  T  002:0000:0
      PSSE=0.0                                                           00900000  T  002:0000:0
      ADDE=0.0                                                           01000000  T  002:0000:5
      POOLE=0.0                                                          01100000  T  002:0001:4
      DO 5 I=1,NY                                                        01200000  T  002:0002:3
      DO 5 J=1,NL                                                        01300000  T  002:0004:0
      CONST=0.0                                                          01400000  T  002:0005:0
    4 IF (MSE(I,J) .EQ. CONST) CONST = 0.0000001                         01500000  T  002:0005:4
      PSSE=PSSE+SSE(I,J)                                                 01600000  T  002:000A:3
      CMSE(I,J)=MSE(I,J)+CONST                                           01700000  T  002:000D:4
    5 ADDE = ADDE + DFE * DLOG(DABS(CMSE(I,J)))                          01800000  T  002:0012:5
      PMSE=PSSE/PDFE                                                     01900000  T  002:001C:2
      IF (JSIG.EQ.0) GO TO 40                                            02000000  T  002:001D:4
      CONST = 0.0000001                                                  02100000  T  002:001F:2
      CPMSE=PMSE+CONST                                                   02200000  T  002:0021:3
      POOLE=PDFE*DLOG(DABS(CPMSE))                                       02300000  T  002:0022:5
      CHIER=POOLE-ADDE                                                   02400000  T  002:0025:1
      CHICF=1+1/(3*(NL*NY)-1)*((NY*NL)/(DFE-1)/PDFE)                     02500000  T  002:0026:2
      CHIER=CHIER/CHICF                                                  02600000  T  002:002C:0
      PRBCER=PRBF((NY*NL)-1,1000.1,CHIER/(NL*NY-1))                      02700000  P  002:002D:2
      WRITE(21,10)                                                       02800000  T  002:0033:2
                                                                                      FIB IS 0006 LONG
   10 FORMAT(5X,21HHOMOGENEITY OF ERRORS///23X,2HYR,1X,3HLOC,2X,2HDF,8X, 02900000  T  002:0037:2
     12HSS,12X,2HMS/)                                                    03000000  T  002:0037:2
      DO 11 L=1,NY                                                       03100000  T  002:0037:2
         DO 11 M=1,NL                                                    03200000  T  002:0038:0
   11 WRITE(21,15) L,M,DFE,SSE(L,M),MSE(L,M)                             03300000  T  002:0039:0
   15 FORMAT(24X,I1,2X,I2,1X,I3,1X,F13.2,2X,F12.2 )                      03400000  T  002:004D:4
      WRITE(21,20)PDFE,PSSE,PMSE,CHIER,PRBCER                            03500000  T  002:004D:4
   20 FORMAT(1HO,23X,4HPOOL,2X,I3,1X,F13.4,2X,F12.4///24X,12HCHI-SQUARE   03600000  T  002:005A:2
     1=,F7.3,5X,5HPROB=,F5.3//)                                          03700000  T  002:005A:2
      IF(PRBCER.LE.0.05)GO TO 30                                         03800000  T  002:005A:2
      WRITE (21,25)                                                      03900000  T  002:005C:4
   25 FORMAT(24X,49HINDIVIDUAL ERROR VARIANCES FORM A HOMOGENEOUS SET/   04000000  T  002:0060:2
     124X,66HFOLLOWING POOLED ANALYSIS IS VALID WITH RESPECT TO THAT ASS 04100000  T  002:0060:2
     2UMPTION//////30X,60(1H*))                                          04200000  T  002:0060:2
      GO TO 40                                                           04300000  T  002:0060:2
   30 WRITE(21,35)                                                       04400000  T  002:0060:5
   35 FORMAT(24X,44HINDIVIDUAL ERROR VARIANCES ARE HETEROGENEOUS/24X,    04500000  T  002:0065:2
     137HFOLLOWING POOLED ANALYSIS IS INVALID /22X,65H**IT IS PRESENTED  04600000  T  002:0065:2
     2ONLY FOR COMPARISON WITH MORE VALID ANALYSIS **/////30X,60(1H*))    04700000  T  002:0065:2
   40 RETURN                                                             04800000  T  002:0065:2
      END                                                                04900000  T  002:0065:5
                                                                                    FORMAT SEGMENT IS 0086 LONG
                                                                                    SEGMENT 002 IS 0075 LONG
```

CODON/TRANS

```
                                                                                                    START OF SEGMENT 002
      SUBROUTINE TRANS(Z,IT)                                                                 00100000  C   002:0000:0
C                                                                                            00200000  C   002:0000:0
C     A SUBROUTINE FOR TRANSFORMING THE DATA:                                                00300000  C   002:0000:0
C                                                                                            00400000  C   002:0000:0
C     IT=1   Y=X                                                                             00500000  C   002:0000:0
C     IT=2   Y=SQRT(X+0.5)                                                                   00600000  C   002:0000:0
C     IT=3   Y=LOG10(X+1)                                                                    00700000  C   002:0000:0
C     IT=4   Y=ARCSIN(SQRT(X));0.LE.X.LE.1                                                   00800000  C   002:0000:0
C     IT=5   Y=1/(1+X)                                                                       00900000  C   002:0000:0
C     IT=6   Y=ARCSIN(SQRT(X/100));0.LE.X.LE.100                                             01000000  C   002:0000:0
C                                                                                            01100000  C   002:0000:0
      GO TO (10,11,12,13,14,16)IT                                                            01200000  C   002:0000:0
   10 RETURN                                                                                 01300000  C   002:0007:0
   11 IF(Z.LT.0)GO TO 15                                                                     01400000  C   002:0007:3
      Z=SQRT(Z+0.5)                                                                          01500000  C   002:0008:4
   15 RETURN                                                                                 01600000  C   002:000B:5
   12 IF(Z.LT.0)Z=0.0                                                                        01700000  C   002:000C:2
      Z=ALOG10(Z+1)                                                                          01800000  C   002:000E:1
      RETURN                                                                                 01900000  C   002:0010:0
   13 Z=ARSIN(SQRT(Z))                                                                       02000000  C   002:0010:3
      RETURN                                                                                 02100000  C   002:0012:4
   16 Z=ARSIN(SQRT(Z/100))                                                                   02200000  C   002:0013:1
      RETURN                                                                                 02300000  C   002:0015:5
   14 IF(Z.EQ.-1)Z=0                                                                         02400000  C   002:0016:2
      Z=1/(Z+1)                                                                              02500000  C   002:0018:2
      RETURN                                                                                 02600000  C   002:0019:5
      END                                                                                    02700000  C   002:001A:2
                                                                                                    SEGMENT 002 IS 0021 LONG

################################################################################################################################

                                                                                                    START OF SEGMENT 006
                                                                                                SEGMENT 006 IS 0004 LONG

      NO ERRORS DETECTED.  NUMBER OF CARDS = 27.
      COMPILATION TIME = 5 SECONDS ELAPSED, 0.85 SECONDS PROCESSING.
      D2 STACK SIZE = 3 WORDS.  FILESIZE = 0 WORDS.  ESTIMATED CORE STORAGE REQUIREMENT = 65 WORDS.
      TOTAL PROGRAM CODE = 56 WORDS.  ARRAY STORAGE = 0 WORDS.
      NUMBER OF PROGRAM SEGMENTS = 6.  NUMBER OF DISK SEGMENTS = 13.
      PROGRAM CODE FILE = (AG0938CULLEN)CODON/TRANS ON PACK.
      COMPILER COMPILED ON 10/31/78 (FORTRAN ON PACK).
```

CODON/BART

```
                                                                                                  START OF SEGMENT 002
      SUBROUTINE BART(N,AR,DI)                                                    00100000  T   002:0000:0
C                                                                                 00200000  T   002:0000:0
C     THIS S/R PERFORMS A BARTLETTS TEST ON THE INDIVIDUAL DEVN M.S.              00300000  T   002:0000:0
C     FROM REGRESSION.                                                            00400000  T   002:0000:0
C                                                                                 00500000  T   002:0000:0
      DIMENSION AR(1)                                                             00600000  T   002:0000:0
      SA=0.0                                                                      00700000  T   002:0000:0
      SB=0.0                                                                      00800000  T   002:0000:4
      DO 100 K=1,N                                                                00900000  T   002:0001:2
      X=AR(K)                                                                     01000000  T   002:0002:0
      SA=SA+X                                                                     01100000  T   002:0004:0
  100 SB=SB+ALOG10(X)                                                             01200000  T   002:0005:2
      SA=ALOG10(SA/N)                                                             01300000  T   002:0009:3
      XP=2.3026*DI*(N*SA-SB)                                                      01400000  T   002:000B:3
      XC=1.+(N+1.)/(3*N*DI)                                                       01500000  T   002:000F:3
      CR=XP/XC                                                                    01600000  T   002:0012:3
      PRB = PRBF(N-1,1000.1,CR/(N-1))                                             01700000  P   002:0013:5
      WRITE(21,1)CR,PRB                                                           01800000  T   002:0018:5
                                                                                                  FIB IS 0006 LONG
    1 FORMAT(/59X,'CHISQ',3X,F13.2/59X,'PROB',3X,F14.3//)                         01900000  T   002:0021:2
      RETURN                                                                      02000000  T   002:0021:2
      END                                                                         02100000  T   002:0021:5
                                                                                                  FORMAT SEGMENT IS 000F LONG
                                                                                                  SEGMENT 002 IS 002C LONG

################################################################################################################################

                                                                                                  START OF SEGMENT 005
                                                                                                      (150)
WARNING:THE FUNCTION "PRBF" WAS NOT FOUND

                                                                                                  START OF SEGMENT 008
                                                                                                  SEGMENT 008 IS 000C LONG
                                                                                                  SEGMENT 005 IS 000B LONG

      NO ERRORS DETECTED.  NUMBER OF CARDS = 21.
      COMPILATION TIME = 8 SECONDS ELAPSED,  1.15 SECONDS PROCESSING.
      D2 STACK SIZE = 6 WORDS.  FILESIZE = 78 WORDS.  ESTIMATED CORE STORAGE REQUIREMENT = 195 WORDS.
      TOTAL PROGRAM CODE = 94 WORDS.  ARRAY STORAGE = 0 WORDS.
      NUMBER OF PROGRAM SEGMENTS = 7.  NUMBER OF DISK SEGMENTS = 18.
      PROGRAM CODE FILE = (AG0938CULLEN)CODON/BART ON PACK.
      COMPILER COMPILED ON 10/31/78 (FORTRAN ON PACK).
```

[Abbreviations of Journal Titles According to The American Chemical Society (1974)]

List of References

Abou-El-Fittouh H.A., J.O. Rawlings and P.A. Miller (1969)
Classification of environments to control genotype by environment interactions with an application to cotton. Crop Sci. 9: 135-140.

Allard R.W. and A.D. Bradshaw (1964)
Implications of genotype-environmental interactions in applied Plant Breeding. Crop Sci. 4: 503-507.

Anderberg M.R. (1973)
Cluster Analysis for Applications. New York, Academic Press 359 pp.

Baker R.J. (1969)
Genotype-environment interactions in yield of wheat. Can. J. Plant Sci. 49: 743-751.

Bartlett, M.S. (1950)
Tests of significance in Factor Analysis. Br. J. Psychol. Statistical Section 3: 77-85.

Bofinger V.J. (1975)
An introduction to some multivariate techniques with applications in field experiments. p.73-83.
In Bofinger V.J. and J.L. Wheeler ed. Developments in Field Experiment Design and Analysis. UK Commonwealth Agricultural Bureaux, pp.196.

Brennan P.S. and D.E. Byth (1979)
Genotype-environmental interactions for wheat yields and selection for widely adapted wheat genotypes. Aust. J. Agric. Res. 30: 221-232.

Bucio Alanis L. (1966)
Environmental and genotype-environmental components of Variability I. Inbred lines. Heredity 21: 387-397.

Bucio Alanis L. and J. Hill (1966)
Environmental and genotype-environmental components of variability II. Heterozygotes. Heredity 21: 399-405.

Burr E.J. (1970)
Cluster sorting with mixed character types
II Fusion Strategies. Aust. Comput. J. 2(3): 98-103.

Byth D.E. (1977)
A conceptual basis of G x E interactions for plant improvement. In 3rd SABRAO proceedings No.1, 3(d)-1-3(d)-11.

Byth D.E., R.L. Eisemann and I.H. DeLacy (1976)
Two way pattern analysis of a large data set to evaluate genotypic adaptation Heredity 37(2): 215-230.

Cameron N.E. (1979)
A study of the acceptability of Holcus sp. to Perendale sheep. M. Ag. Sci Thesis, Massey University.

Cattell R.B. (1965)
Factor Analysis: An introduction to essentials. Parts I and II, Biometrics 21: 190-435.

Causton D.R. (1977)
A Biologists Mathematics. London, Arnold 326 pp.

Claridge J.H. (1973)
Arable Farm Crops of New Zealand
DSIR, A.H. and A.W. Reed, 345 pp.

Clifford H.T. and W. Stephenson (1975)
An Introduction to Numerical Classification. New York, Academic Press 229 pp.

Cochran W.G. (1947)
Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics 3: 22-38.

Coles G.D. (1980)
State involvement in barley breeding - A personal view. Crop Res. News, DSIR, N.Z. 23: 24-29.

Comstock R.E. and R.H. Moll (1963)
Genotype-environment interactions. In Statistical Genetics and Plant Breeding ed. Hanson W.D. and H.F. Robinson. National Academy of Science, publication No. 982, p.164.

Cooley W.W. and P.R. Lohnes (1971)
Multivariate Data Analysis, Wiley 364 pp.

Cormack R.M. (1971)
A review of classification. JR. Stats. Soc. A 134: 321-358.

Cross R.J. (1980)
Initiation of Spring Wheat Breeding at Lincoln.
Cereal News DSIR No.8, 19-20.

Douglas J.A., C.B. Dyson and J.E Waller (1977)
A basis for crop cultivar evaluation. Proc. Agron. Soc. N.Z. 7: 103-106.

Draper N. and H. Smith (1966)
Applied Regression Analysis. John Wiley and Sons, 407 pp.

Dyson C.B. (1979)
pers. comm.

Eagles H.A., P.N. Hinz and K.J. Frey (1977)
Selection of superior cultivars of oats by using regression coefficients. Crop Sci. 17: 101-105.

Easton H.S. (1971)
The interaction of wheat genotypes with a specific factor of the environment. M.Ag. Sci Thesis, Massey University.

Easton, H.S. and R.J. Clements (1973)
The interaction of wheat genotypes with a specific factor of the environment. J. Agric. Sci, Camb 80: 43-52.

Eberhart S.A. and W.A. Russell (1966)
Stability parameters for comparing varieties.
Crop Sci. 6: 36-40.

Eisemann R.L., D.E. Byth, I.H. DeLacy and P.J. Taylor (1977a)
A new approach to the analysis of genotypic adaptation and genotype-environment interactions. In 3rd SABRAO proceedings, No.1 3(d)-16-3(d)-22.

Eisemann R.L., D.E. Byth, I.H. DeLacy and P.J. Taylor (1977b)
A comparison of some methods of analysis of G x E interactions and adaptation responses in a large data set. In 3rd SABRAO proceedings No.1 3(d)-41-3(d)-47.

Eisenhart C. (1947)
The assumptions underlying the analysis of variance.
Biometrics 3: 1-21.

Everitt B. (1974)
Cluster Analysis. Heinemann, London 122 pp.

Finlay K.W. and G.N. Wilkinson (1963)
The analysis of adaptation in a plant breeding programme.
Aust. J. Agric. Res. 14: 742-754.

Federer W.T. and G.F. Sprague (1947)
A comparison of variance components in corn yield trials.
J. Am. Soc. Agron. 39: 453-463.

Francis T.R. and L.W. Kannenberg (1978)
Yield stability studies in short-season maize I.A. descriptive method for grouping genotypes. Can. J. Plt. Sci 58: 1029-1034.

Freeman G.H. (1973)
Statistical methods for the analysis of genotype-environment interactions. Heredity 31: 339-354.

Freeman G.H. and P. Crisp (1979)
The use of related variables in explaining genotype-environment interactions. Heredity 41: 1-11.

Freeman G.H. and B.D. Dowker (1973)
The analysis of variation between and within genotypes and environments. Heredity 30(2): 97-109.

Freeman G.H. and J.M. Perkins (1971)
Environmental and genotype-environmental components of variability. Heredity 27: 15-23.

Fripp Y.J. and C.E. Caten (1971)
Genotype-environmental interactions in Schizophyllum commune. I. Analysis and character. Heredity 27: 393-407.

Funnuh S.M. and C. Mak (1980)
Yield stability studies in soya beans (Glycine max). Exp. Agric. 16: 387-392.

Goodall D.W. (1954)
Objective methods for the classification of vegetation. III An essay in the use of factor analysis. Aust. J. Bot. 2: 304-324.

Gordon I.L., D.E. Byth and L.N. Balaam (1972)
Variance of heritability ratios estimated from phenotypic variance components. Biometrics 28: 401-415.

Gower J.C. (1966)
Some distance properties and latent root and vector methods used in multivariate analysis. Biometrika 53: 325-338.

Hanson, W.D. (1970)
Genotypic stability. Theor. Appl. Genet. 40: 226-231.

Hardwick R.C. and J.T. Wood (1972)
Regression methods for studying genotype-environment interactions. Heredity 28: 209-222.

Hill J. (1975)
Genotype-environment interactions - a challenge for plant breeding. J. Agric. Sci, Camb. 85: 477-493.

Jardine N. and R. Sibson (1968)
The construction of hierarchic and non-hierarchic classiciations. Comput. J. 11: 177-184.

Jinks J.L. and V. Connolly (1973)
Selection for specific and general response to environmental differences. Heredity 30: 33-40.

Jinks J.L. and H.S. Pooni (1979)
Non-linear genotype-environment interactions arising from response thresholds. 1. Parents, $F_{1's}$ and selections. Heredity 43(1): 57-70.

Jowett D. (1972)
Yield stability parameters for sorghum in East Africa. Crop Sci. 12: 314-317.

Kim J.O. (1975)
Factor Analysis, p.468-513, In SPSS Manual, 2nd edition. (Nie N.H. et al, eds), McGraw-Hill, USA.

Knight R. (1970)
The measurement and interpretation of genotype-environment interactions. Euphytica 19: 225-235.

Lance G.N. and W.T. Williams (1967)
A general theory of classificatory sorting strategies. 1. Hierarchical systems. Comput. J. 9: 373-380.

Langer I., K.J. Frey and T. Bailey (1979)
Associations among productivity production response and stability indexes in oat varieties. Euphytica 28: 17-24.

Le Clerg E.L., W.H. Leonard and A.G. Clark (1962)
Field plot technique. Burgess, Minneapolis, USA, pp.271.

Lin C.S.and B. Thompson (1975)
An empirical method of grouping genotypes based on a linear function of the genotype-environment interaction. Heredity 34(2): 255-263.

Lin C.S., M.R. Binns and B.K. Thompson (1977)
The use of regression methods to study genotype-environment interactions. Heredity 38(3): 309-319.

McEwan J.M. (1979)
pers comm.

Malcolm J.P. (1978)
Barley cultivars. A guide. Ag link Information Services MAF NZ.

Mandel J. (1971)
A new analysis of variance model for non-additive data. Technometrics 13: 1-18.

Miller, P.A., H.F. Robinson and O.A. Pope (1962)
Cotton variety testing: additional information on variety x environment interactions. Crop Sci. 2: 349-352.

Miller P.A., J.C. Williams and H.F. Robinson (1959)
Variety x Environment interactions in cotton variety tests and their implications on testing methods. Agron. J. 51: 132-134.

Mitchel W.J.P. and R.J. Cross (1977)
The performance of prospective wheat cultivars in the Southern half f the North Island. Proc. Agron. Soc. NZ 7: 55-56.

Mitchel W.J.P. and R.J. Cross (1979a)
Yield evaluation of barley under North Island conditions N.Z. J. Exp. Agric 7: 375-376.

Mitchel W.J.P. and R.J. Cross (1979b)
Wheat cultivar evaluation under North Island conditions. NZ J. Exp. Agric. 7: 377-378.

Moll R.H. and C.W. Stuber (1974)
Quantitative Genetics. Adv. Agron. 26: 277-313.

Morgenstern E.K. and A.H. Teich (1969)
Phenotypic Stability of height growth of Jack Pine Provenances. Can. J. Genet. Cytol. 11: 110-117.

Morrison D.F. (1967)
Multivariate Statistical Methods. McGraw-Hill, USA, 338 pp.

Mungomery V.E., R. Shorter and D.E. Byth (1974)
Genotype x Environment Interactions. I Pattern Analysis - application to soya bean populations. Aust. J. Agric Res. 25: 59-72.

Nie N.H., C. Hadlai Hull, J.G. Jenkins, K. Steinbrenner and D.H. Bent
1975 2nd ed. Statistical Package for the Social Sciences. McGraw-Hill, 675 pp.

Nor K.M. and F.B. Cady (1979)
Methodology for identifying wide adaptability in crops. Agron. J. 71: 556-559.

Okuno T., F. Kikuchi, K. Kunagai, C. Okuno, M. Shiyomi and H. Tabachi (1971)
Evaluation of varietal performance in several environments. Bull. Nat. Inst. Agric. Sci (Japan) Series A. No.18, 93-147. As reported in JIBP Synthesis vol.6 ed. T. Matsuo. Adaptability in Plants (1975).

Perkins, J.M. (1971)
The Principal Component Analysis of Genotype-Environment interactions and physical measures of the environment. Heredity 29: 51-70.

Perkins J.M. and J.L. Jinks (1967)
Environmental and Genotype-environmental components of variability. III Multiple lines and crosses. Heredity 23: 339-356.

Perkins J.M. and J.L. Jinks (1973)
The assessment and specificity of environmental and genotype-environmental components of variability. Heredity 30 (2): 111-126.

Plaisted R.L. (1960)
A shorter method for evaluating the ability of selections to yield consistently over locations. Am. Potato J. 37: 166-172.

Plaisted R.L. and L.C. Peterson (1959)
A technique for evaluating the ability of selections to yield consistently in different locations and seasons. Am. Potato J. 36: 381-385.

Rummel R.J. (1970)
Applied Factor Analysis. Evanston: Northwestern University Press, USA, 617 pp.

Satterthwaite F.E. (1946)
An approximate distribution of estimates of variance components. Biometrics 2: 110-114.

Shorter R., D.E. Byth and V.E. Mungomery (1977)
Genotype x Environment Interactions and Environmental Adaptation. II Assessment of environmental conditions. Aust. J. Agric Res. 28: 223-235.

Shukla G.K. (1972)
Some statistical aspects of partitioning genotype-environmental components of variability. Heredity 29: 237-245.

Simmonds N.W. (1979)
Principles of Crop Improvement. Longman, New York 408 pp.

Simmonds N.W. (1980)
Comparisons of the Yields of Four potato varieties in Trials and in Agriculture. Exp. Agric. 16: 393-398.

Snoad B. and A.E. Arthur (1976)
The use of regression techniques for predicting the response of peas to environment. Theor. Appl. Genet. 47: 9-19.

Sokal R.R. and C.D. Michener (1958)
A statistical model for evaluating systematic relationships. Univ. Kansas Sci Bull. 38, 1409-1438. As reported in Lin C.S. and B. Thompson (1975).

Sprague G.F. and W.T.Federer (1951)
A comparison of variance components in Corn Yield Trials: II Error, Year x Variety, Location x Variety, and Variety Components. Agron. J. 43: 535-541.

Steel R.G.D. and J.H. Torrie (1980)
Principles and Procedures of Statistics, 2nd Edition. McGraw-Hill, USA, 633 pp.

St-Pierre C.A., H.R. Klinck and F.M. Gauthier (1967)
Early generation selection under different environments as it influences adaptation of barley. Can. J. Plt. Sci 47: 507-517.

Suzuki S. and J. Abe (1975)
A study on the adaptability of Forage Crop Cultivars, p.109-121. in Matsuo T. ed. JIBP Synthesis Adaptability in Plants, vol.6.

Tai G.C.C. (1971)
Genotypic stability analysis and its application to potato Regional Trials. Crop Sci 11: 184-190.

Tai G.C.C. (1979)
Analysis of genotype-environment interactions of potato yield. Crop Sci 19: 434-438.

Tan W.K., G.Y. Tan and P.D. Walton (1979)
Regression Analysis of Genotype-Environment Interaction in Smooth Bromegrass. Crop Sci. 19: 393-396.

Taylor, Eisemann, Shorter and De Lacy (1978)
YLGR, University of Queensland.

Teow, S.H.(1978)
M.Ag Sci, Massey University.
Evaluation of Variability in a Fog Grass Gene Pool.

Verma M.M. and G.S. Chahal and B.R. Murty (1978)
Limitations of Conventional Regression Analysis. A Proposed Modification. Theor. Appl. Genets. 53: 89-91.

Williams H. (1974)
Genotype-environment interaction in strawberry cultivars. Hortic. Res. 14: 18-88.

Williams W.T. (1971)
Principles of Clustering. Annu. Rev. Ecol. Syst. 2: 303-326.

Williams W.T. (1975)
Pattern Analysis in Field Experiments, p.107-117. In Bofinger V.J. et al Developments in Field Experiment Design and Analysis. Commonwealth Agricultural Bureaux, UK.

Williams W.T. (1976) ed.
Pattern Analysis in Agricultural Science, CSIRO, Melbourne pp.331.

Williams W.T. and M.B. Dale (1965)
Fundamental Problems in numerical taxonomy. Adv. Bot. Res. 2: 35-68.

Williams W.T. and P. Gillard (1971)
Pattern Analysis of a Grazing Experiment. Aust. J. Agric Res. 22: 245-260.

Wood J.T. (1976)
The use of environmental variables in the interpretation of genotype-environment interaction. Heredity 37(1): 1-7.

Wricke G. (1962)
Ubereine methode zur erfassung der ökologischen streubreite in Feldversuchen. Z. Pfl Zücht. 47, 92-96. [As reported in Easton and Clements, 1973].

Wright A.J. (1971)
The analysis and prediction of some two factor interactions in grass breeding. J. Agric. Sci, Camb. 76: 301-306.

Wright A.J. (1976)
The significance for breeding of linear regression analysis of genotype-environmental interactions. Heredity 37(1) 83-93.

Yates F. and W.G. Cochran (1938)
The analysis of groups of experiments. J. Agric. Sci, Camb. 28: 556-579.