2007

An evaluation of multivariate statistical techniques for the analysis of yield from barley (Hordeum vulgare L.) breeding trials data

ABDULLAH, AHMED

http://hdl.handle.net/10026.1/1972

http://dx.doi.org/10.24382/3499
University of Plymouth

An evaluation of multivariate statistical techniques for the analysis of yield from barley *(Hordeum vulgare L.)* breeding trials data

by

AHMED ABDULLAH

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Biological Sciences
Faculty of Science

July 2007

Copyright Statement

Abstract

An evaluation of multivariate statistical techniques for the analysis of yield from barley *(Hordeum vulgare L.)* breeding trials data

by

Ahmed Abdullah

This project involved two locations (Breda and Tel Hadya) over two seasons (1993 and 1994).

Yield was found to have been affected by many factors including environment, genotype and morphological characters. A genotype-environment interaction (GEI) was also discovered.

To investigate the influence of morphological characters on yield parameters, multivariate statistical techniques (canonical analysis, factor analysis and multiple regression analysis (linear and exponential)) were used. Multivariate statistical techniques were applied to three hybrids (Hybrid 1, 2 and 3) in replicated field plots at two locations (Breda and Tel Hadya) in two seasons.

Canonical analysis and factor analysis revealed a significant relationship between yield parameters and morphological characters. However, this relationship was not significant for each hybrid because there were insufficient data for each hybrid.

Stepwise multiple regression analysis showed that plant height, vegetative duration and length of growing season were the significant factors influencing yield parameters,.while leafiness was not. The relationship can approximate nonlinear in that it gives more realistic predictions. Consequently, stepwise multiple exponential equation fitted the data better than stepwise multiple linear equation.

The relationship between yield parameters and morphological characters was affected by environment but not by genotype.

List of Contents

Contents Page no.

List of Tables

List of Figures

Dedication

I dedicate this thesis to my Father, Hatem; my mother, Sameha; My wife, Hanan my Children, Hazem and Shemaa.

Acknowledgements

Studying abroad was an exciting experience and definitely more challenging than studying in my own country. I can say that these four years were exciting and challenging. I had the privilege of living in the lovely city of Plymouth and met a lot of people without whom I would not be here, and completing this thesis would not have been possible. To all of you, close and far I am endlessly thankful. However, let me mention some people by name.

First and foremost I would like to express my gratitude to Dr. John Eddison and Prof. Mick Fuller, my supervisors, due to their support and encouragement throughout this research. It was only due to their valuable guidance, cheerful enthusiasm and ever-friendly nature that I was able to complete my research in a respectable manner.

My special thanks to Prof. Abdo Kaseem my supervisor in Syria for his concern and opinions during my study. Also I wish to thank Prof. Gassan Khayat, Prof. Mahmoud Yaseen, Prof. Falah Abo-Nokta and Prof. Fasal Hamed for their valuable advice and friendly help.

I am grateful beyond expression to the people I left when coming here "my parents" thank you for your continuous support, encouragement and faith in me. Especially when living abroad, it is important to know that you are always on my side. I think that now, at the end of my Ph.D., is a good moment to thank you as well for all the time and effort you put into my education.

I owe my loving thanks to my wife Hannan Najar and my children Hazem and Shemaa. They have lost a lot due to my research abroad. Without their encouragement and understanding it would have been impossible for me to finish this work. My thanks are due to my best friend Jack Shannon for supporting me.

Finally I wish to thank my other colleagues from the School of Biological Sciences, for a great working atmosphere and the good times.

Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

I declare that the work submitted in this thesis is the results of my own investigations except where reference is made to published literature and where assistance is acknowledged.

Candidate

Director of studies

Word account of main body thesis: 37669 words

Course attended (Modules)

Module code	Title	Result
EFP0113	Postgraduate Academic reading and Writing	A
EFP0115	Information Technology and Skills for Postgraduate study	A
EFP0211	Introduction to Professional English	A
EFP0213	Preparing for IELTS	A
EFP0214	Introduction to Postgraduate Research	A
IMS0111	British Culture and Society Life in Britain	A
IMS0112	Educational and Research Skills and Information Technology	A
IMS0217Y	English for Academic Purposes	A
STAT011	Business Data Analysis and Optimisation	A

Paper published

- **Abdullah, A., Fuller M.P.** 2006. Genotype environment interaction in breeding lines of Syrian Barley. *Bibliotheca Fragmenta Agronomica.* **11**, 17-18.
- **Abdullah, A., Kalandar, H.** 2001. Relation between the Properties and Ostensible Properties of barley. *Aleppo University within the series of Economic Sciences.* **27**

Conference poster presentation

- **Abdullah, A., Eddison, J.C., Fuller M.P. 2006.** *Genotype environment interaction in breeding lines of Syrian Barley(Hordeum vulgare L).* IX European Society for Agronomy Congress. " European education and research in agronomy", Warsaw, Poland, 4- 7 September 2006.
- **Abdullah, A., Eddison, J.C., Fuller M.P. 2006.** *Genotype environment interaction in breeding lines of Syrian Barley(Hordeum vulgare L).* VC's Research and Innovation Conference. Collaboration in research and innovation: opportunities and benefits. Wednesday 4th April 2007 Sherwell Centre, University of Plymouth.

Attendance

- Multilevel analysis (at the University of Bristol).
- SPSS (part one and part two).
- GTA course: Learning, Teaching and Assessment: (Theory and Practice).
- EndNote program, introduction and advanced.
- Introduction to Ph.D.
- The Transfer Process.
- An Introduction to Qualitative Research.
- Risk Management.
- Preparing to Teach.
- Research- Owning and Using.
- Microsoft PowerPoint.
- Essential tips for using Microsoft Word 2002.
- Project Management.
- Academic writing.

Chapter 1

Introduction and Literature Review

1.1 Introduction

Barley is one of the best cereals for cultivation in dry and semi-dry areas of the world. It is the world's fourth most important cereal in terms of production, (after wheat, rice, and maize) while it is the second most important (after wheat) in the Middle East and North Africa (Naesah, 1996). Barley is grown in significant quantity in many places around the world. The production of barley in the main producing countries and regions for 1995/1996 and 2005/2006 is summarized in Table 1.1.

Barley production in 1995/1996 was in fact higher than in 2005/2006 despite considerable research in between with the aim of increasing yield. Iran and Australia were not among the main producing countries in 1995/1996 although they were in 2005/2006. Production in most of the major barley-producing countries outside the European Union (EU) decreased, although within the EU production has increased, for example, barley production in Morocco fell by 2,618 Thousand Metric Tonnes between 1995/1996 and 2005/2006.

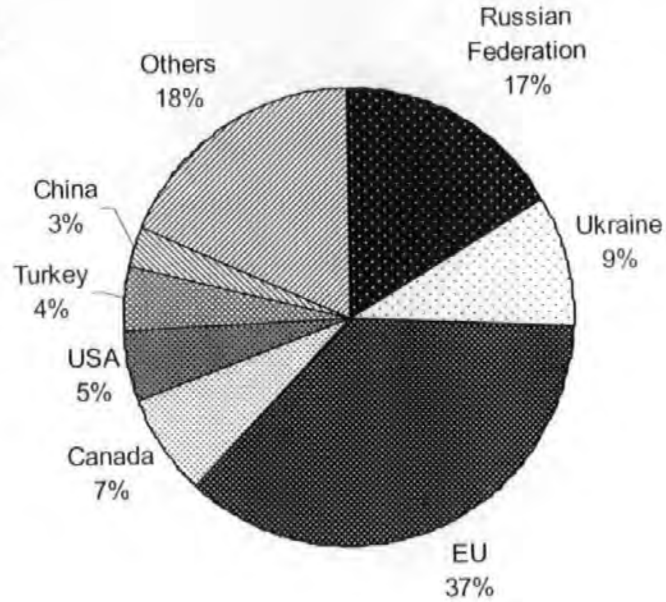
From the data in Table 1.1, the percentage of the world barley production in the main producing countries and regions were derived (Figure 1.1.a and Figure 1.1,b).

Table 1.1. Production of barley in the main producing countries for 1995/1996 and 2005/2006.

Country or region	1995/1996 Production ('000 metric tonnes)		Country or region	2005/2006 Production ('000 metric tonnes)	
Russian Federation	27054		Russian Federation	15800	
Ukraine	14509		Ukraine	9000	
Germany	10903		EU	52917	
France	7898		-	-	
Spain	7416		-	-	
U.K.	5950		-	-	
Denmark	3446		-	-	
Other EU	11806		Iran	2900	
Eastern Europe	11459		Australia	9869	
Among Which:	-		Others	4515	
Bulgaria		1143	Ethiopia		1785
Hungary		1558	Iraq		1250
Poland		2686	India		1080
Czech Rep.		2419	Algeria		400
Romania		2134	-	-	
Slovakia		874	-	-	
Canada	11690		Canada	12481	
USA	8161		USA	4613	
Turkey	7000		Turkey	7600	
Kazakhstan	6497		Kazakhstan	1500	
Belarus	3012		Belarus	1800	
China	4500		China	3400	
Morocco	3720		Morocco	1102	
Others	16465		Other	10643	
World Total	**161486**		**World Total**	**138140**	

Source : USDA Foreign Agricultural Service (Anon, 1999; Anon, 2007).

1.1.a
(1995/1996)



1.1.b
(2005/2006)

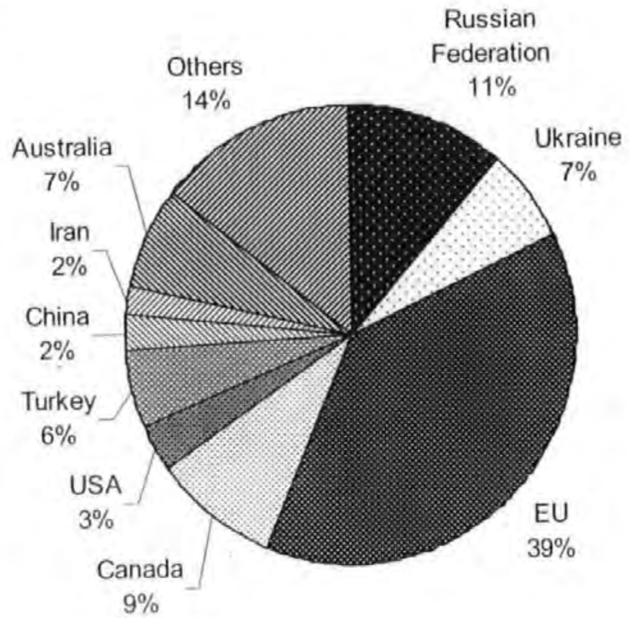


Figure 1.1.The percentage of the world production of barley in the main production countries and regions in a) 1995/1996 and b) 2005/2006 (derived from Table 1.1).

Over a ten year period these statistics demonstrate that the share of world production can very greatly between countries and regions. For example, in 1995/1996 the Russian Federation produced 17% of world production whilst in 2005/2006 this fell to only 11%, a reduction of over 10 Mt.

Australia is currently the largest exporter of barley and malt (Figure 1.2) while Saudi Arabia, Japan and China are the largest importers of the world's barley (Anon, 2006b).

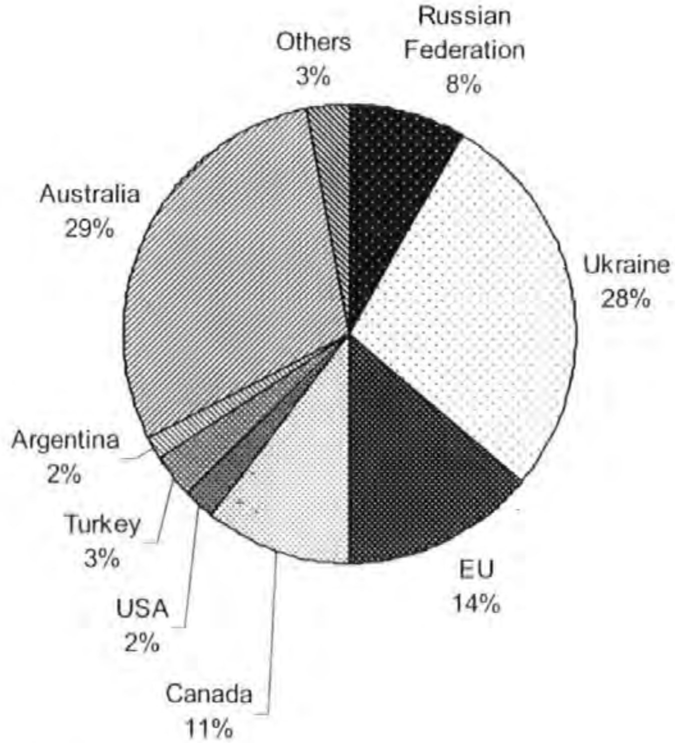


Figure 1.2. World barley exports 2005/2006 (USDA Foreign Agricultural Service (Anon, 2007)).

Syria is a Middle Eastern cereal-producing country with a national agriculture production similar to Iraq, Morocco and Bulgaria (Naesah, 1996). The area and production of barley in cultivation and the average yield between 1981 and 2005 in Syria is presented in Table 1.2 and shows that the production of barley was not stable during this period.

Syrian barley area increased slightly between 1981 and 1992 but then decreased until 2000. It has been nearly constant at 1,300 thousand hectares since 2001 but production levels have varied. It was estimated that the Syrian barley production in the crop year 2003 was approximately 1,100,000 t from 1.3 Mha harvest area (Anon, 2006b) while it was estimated in 2004 and 2005 at approximately 900,000 and 750,000 t, respectively, from the same production box (Table 1.2). The reduction in harvested area in 2004 and 2005 was attributed to severe drought in spring (Anon, 2006b) .

Barley Grain yield in Syria is both low by overall world standards, and varies tremendously from year to year with as much as 15-fold variation between consecutive years (1988 and 1989; Figure 1.3).

Table 1.2. The area and yield of barley cultivation in Syria 1981-2005

Date	Area in Syria ('000 ha)	Production ('000 metric tonnes)
1981	1347	1406
1982	1589	661
1983	1520	1043
1984	1289	304
1985	1386	740
1986	1548	1116
1987	1570	576
1988	1844	2836
1989	2892	271
1990	2729	846
1991	2233	917
1992	2267	1091
1993	2169	1553
1994	1894	1482
1995	1963	1705
1996	1550	1653
1997	1572	983
1998	1543	869
1999	1414	426
2000	250	130
2001	1400	1300
2002	1400	920
2003	1300	1100
2004	1300	900
2005	1300	750

Source : USDA Foreign Agricultural Service and (Anon 2002a; Anon 2006b).

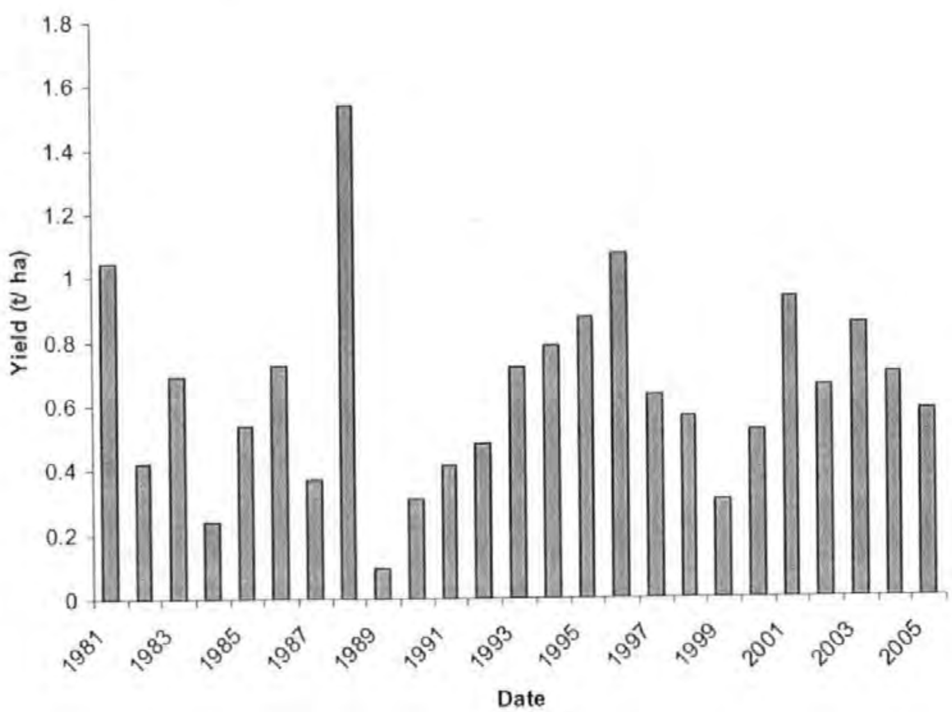


Figure 1.3. Grain yield of barley in Syria between 1981 and 2005 (derived from Table 1.1).

Barley in the Middle East is mainly used to feed ruminant animals (generally sheep), and also on a very limited scale for feeding egg-laying poultry. During the winter period there is no grass available for sheep fodder, therefore HOBOOB[1] regularly retains a stock of barley for this purpose. Syria normally requires about 1.3 - 1.4 million tonnes of barley per year for feed use and for seed production (Anon, 2002b), and it can be seen from Table 1.2 that Syria is frequently forced to import barley.

Barley in Syria in 2003 and 2004 was mainly imported from Iraq and Eastern Europe countries due to lower prices in the international market compared to the locally produced crop (Table 1.3). An import permit is required for such imports from the Ministry of Economy and Trade (MET), after receiving approval from the Ministry of Agriculture and Agrarian Reform (MAAR).

The domestic price for barley has been stable at 7,000 Syrian pounds per tonne (US $136). However, the international market in 2004 offered more favourable prices, and therefore HOBOOB purchased large quantities of barley from Iraq, Turkey and East Europe during that year. Despite domestic shortfalls, the Syrian public sector fulfilled pre-arranged contracts by exporting large quantities of barley to Jordan in 2003 and in the early part of 2004 (Table 1.4).

[1] HOBOOB is a Syrian government company responsible for the nationwide collection of barley and its export. HOBOOB literally translates as " crops"

Table 1.3. Import Trade matrix (Barley ; Syria).

2003 (1000 t)			2004 (1000 t)	2005 (1000 t)
Iraq	373	Iraq	281	200
Turkey	15	Ukraine	279	300
Germany	33	Bulgaria	53	-
Russia	11	Turkey	-	200
Other	168	Other	12	100
Total	**600**	**Total**	**625**	**800**

Source : USDA Foreign Agricultural Service (Anon, 2006b).

Table 1.4. Export Trade matrix (Barley ; Syria).

2003 (1000 t)		2004 (1000 t)		2005 (1000 t)	
Jordan	448				
Algeria	29				
Cyprus	34				
Saudi Arabia	20	Jordan	194	other	50
Lebanon	10				
Libya	5				
Other	54				
Total	**600**	**Total**	**194**	**Total**	**50**

Source : USDA Foreign Agricultural Service (Anon, 2006b).

A great deal of research has been carried out with the aim of increasing barley yield in Syria. The most important centres involved in barley breeding research are the Arab Centre for the Study of Arid and Dry Land (ACSAD) and the International Centre for Agricultural Research in Dry Areas (ICARDA). ACSAD is a Syrian Government funded centre, whilst ICARDA receives both Government funding and substantial international development funding. ICARDA was established in 1977 in recognition of the biodiversity of barley, wheat and lentil gemplan in both the cultivated and natural areas in northern Syria.

Yield in barley, and in cereals in general, is a function of the interaction of genetics and environment; and in Syria critical environmental factors, especially rainfall, can vary enormously between seasons. The influence of genetics was clearly demonstrated in the period 1988-1989, when a large-scale national change in the barley variety grown had disastrous results (Naesah, 1996), resulting in a reduction of barley production by *90%* (Table 1.2).

The morphology and physiology of cultivars is clearly a vital factor governing yield. Important morphological characteristics of barley include leafiness (number and size of leaves), length of growing season, plant height, and vegetative duration (Briggs, 1978). An understanding of both climate and morphology is of central importance for increasing yield, and this dissertation aims to further this understanding by means of novel applications of statistical analyses in studying the effects of seasons on barley cultivation.

Fraser and Eaton (1983) described many statistical analyses that have been used to study the relationship between yield and the variables which affect it, including correlation analysis and simple linear regression analysis. Fraser & Eaton (1983) have stressed that these kinds of methods are not very effective, especially (i) when the independent variables in regression equations have high correlations between them and (ii) when the variables are measured in different units (Naesah, 1996). Consequently, multi-colinearity has been found which leads to errors in the analytical results. To avoid this problem, To avoid this problem, statisticians generally advocate the following approaches (i) some of the independent variables which have a high correlation between them are removed; (ii) new statistical analyses are used, for example, multivariate analysis, canonical correlation analysis, or multiple regression analysis which are less affected by co-linearity. These statistical analyses are popular in many areas of biological science, but have not been used extensively in agriculture, especially in yield improvement studies. The reason for this has been cited as the difficulty and complexity involved in solving the equations when the equation has many variables. More recently, improvements in computer power have given statisticians the opportunity to investigate these methods more widely.

1.2 Regression Analysis

Regression analysis is defined as the relationship among variables (Chatterjee and Price, 1977). It is a method of organizing data and standing the relationship between two variables. Sometimes it is appropriate to show data as points on a graph, and then try to estimate a line of best fit through the data by the technique known as the sum of the least square.

Many people believe that linear regression and correlation coefficient were discovered and developed by Karl Pearson. However, they were the inventions of Sir Francis Galton (cousin of Charles Darwin) in the early 19th Century. He used correlation and regression with genetics and heredity.

"Galton's fascination with genetics and heredity provided the initial inspiration that led to regression and the PPMC[2]" (Jeffrey, 2001).

For many years, regression analysis has been developed and used in many types of science [3] for example:

- **Health:**

Multiple linear regression has been used in many areas of Health and Medicine. The multiple linear regression method was used to study haematological changes due to chronic exposure to natural gas leakage, in order to detect differences among exposed groups for haematological markers (Saadat and Bahaoddini, 2004).

Multiple linear regression was used in a study of methods preferred by surgeons and radiologists, respectively, in the treatment of severe limb ischemia. In this case, stepwise multiple linear regression was able to identify significant statistical differences which affected responses from the entire group and from surgeons and radiologists separately (Bradbury, *et al.*, 2004).

- **Food :**

Some statisticians have used multiple regression analysis in dietary research. Other statisticians used stepwise multiple linear regression to detect milk mixtures in Halloumi cheese (Recio, *et al.*, 2004).

[2] This statistic is the known as the Pearson Product Moment correlation.
[3] most statisticians speak of "multiple regression " rather than " multiple linear regression"

- **Business:**

Multiple regression analysis has been used in different types of business, mainly by companies which aim to increase their competitive advantage by improving operational performance (De Cerio, 2003). It also has been used to study insurance policies (Chenevert and Tremblay, 2002). There are many applications of multiple regression in business and most of these applications aim to increase the profits of companies.

- **Agriculture:**

Multiple linear regression has been employed in some types of agricultural research, for example, in studies of the influence of weather on agriculture and in detailed physiological studies for example, the relationship between multispectral band features and nitrate in potato leaves (Borhan, *et al.*, 2004). Stepwise multiple regression analysis has been used to study the influences on seed yield of some agronomic and seed characters of sunflower. It indicated that seed yield is strongly affected by the date of physiological maturity, plant height and oil content (Qaizar, *et al.*, 1991). Stepwise regression has also been used with wheat to study the intra- and inter-generation relationship among yield, its components and other related characteristics. Stepwise regression suggested that 1000-kernel weight, grain yield per plant and number of tillers per plant (in order) were the most significant factors in determining F_4 line yield (Lungu, *et al.*, 1990). Regression analysis was also used to study the genotype environment interaction for tea yields (Francis, *et al.*, 2002). Stepwise regression analysis was used to study spatial and temporal variability in nitrogen uptake by corn across a variable landscape. It showed the effect of tillage, legume, nitrogen fertilizer and organic carbon on cumulative nitrogen uptake at different growth stages (Dharmakeerthia, *et al.*, 2006).

- **Barley**

Many barley agronomic experiments are analysed by using multiple linear regression analysis, and most of them use this kind of analysis with the objective to increase yield. Much of this research has dealt with the effects of environment on barley yield. For example, using multiple linear regression analysis, research has revealed a strong relationship between climate and barley cultivation, (Sharratt, *et al.*, 2003). Multiple regression analysis has also been used to study the influence of rainfall and temperature on the feeding value of barley straw in semi-arid regions (Goodchild, 1997). It was also used to study yield response of barley to rainfall and temperature in northern Syria (Vanoosterom, *et al.*, 1993a); to study the relationship between zonal variation and all quantitative morpho-agronomic characters, except plant height (Kebebew, *et al.*, 2001); and to study increased dry area cropping intensity with no-till barley (Schillinger, *et al.*, 1999). In a study of the use of field spectroscopy for the ranking of cereal breeding plots during the early stages of crop growth, multiple linear regression was used to find the optimal relationship between spectral reflectance and biomass (Smith, *et al.*, 1993). Multiple regression analysis was also used to predict cadmium concentrations and pH in barley grain by testing soil properties (Adams, *et al.*, 2004).

In the realm of genetic research, procedures in multiple linear regression analysis have been used to explore the relationships between the phenotype and genotype of barley (Zhu, *et al.*, 1999). Stepwise multiple regression analysis also was used in a project to identify a region of the barley genome contributing to variation in height (Barua, *et al.*, 1993). Stepwise multiple regression was used to study the linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. It indicated that for complex traits with costly measurements, the association mapping approaches can be a viable alternative to classical quantitative trait locus approaches based on crosses between inbred lines (Arnold, *et al.*, 2004).

Multiple linear regression equations were used to improve the standard of prediction for protein in barley and malt (Fox, *et al.*, 2002), and in a study of barley samples to investigate the feasibility of producing nitrogen-corrected true metabolizable energy (tmen) content (Zhang, *et al.*, 1994).

Scientists have used multiple linear regression analysis and correlation coefficients to decide whether or not there are relationships between the variables in their research. Adams, *et al.* (2004) found that a linear equation was not useful in their research and concluded that there are no relationships between the variables. However, when these workers investigated alternative equations such as the **exponential equation** (Adams, *et al.*, 2004) relationships were established. The significance of the equation, coefficient of determination, standardised residuals and significance of independent variables are used to determine whether or not linear regression is better than exponential regression. Exponential equations have been used in many kinds of science. For example, it has been used in medicine to describe the cytotoxic activity of several unrelated drugs (Breier, *et al.*, 2000). In agricultural research on barley, they were used to describe the effect of time on radio caesium fixation (Absalom, *et al.*, 2001); and to study thin layer experiments for germinating malting barley varieties (Bala and Woods, 1992). They were also used to study the relationship between leaf appearance rate (LAR) and temperature (Xue, *et al.*, 2004); and to describe total gas production which was correlated with intake, digestible dry matter intake and growth rate (Blummel and Orskov, 1993).

1.3 Factor Analysis

Harman and other scientists believe that factor analysis was developed by Pearson in 1901. Nevertheless, "Spearman, who devoted the remaining forty years of his life to the development of factor analysis, is regarded as the father of the subject" (Harman, 1967). The first improvement to the methods was made by Hotelling (1933) (Harman, 1967). Before the invention of the digital computer, factor analysis was regarded as an unreliable method since there existed many variations, each offering different results. These discrepancies were due to a necessary dependence on simplifying assumptions and auxiliary conditions. In the present era, factor analysis has been transformed by the development of computer techniques and the computational speed that they often. The older methods, with their conflicting results, have been largely abandoned. This development has led to an enormous increase in the number of methodologies. For example, factor analysis has been used to determine characters for grain yield selection in chickpea (Toker and Cagirgan, 2004).

The modern definition of factor analysis is as follows: "factor analysis is a multivariate technique for reducing matrices of data to their lowest dimensionality by the use of orthogonal factor space and transformations that yield prediction and/or recognizable factors" (Utexas, 1995).

According to this definition, factor analysis is a model which attempts to explain the correlation between a large set of variables. The main applications of factor analytical techniques are :

(i) to reduce the number of variables

(ii) to detect structure in the relationship between variables, that is, to classify variables (Anon, 2004c).

Factor analysis has been used in several different areas in agricultural science. For example, it has been used to study the genetics of beans in order to increase yield and improve quality (Nasser, 2002). It was also used to determine all the interrelationship between the characters of durum wheat

(Wassouf, 1996). Factor analysis was applied to the data to study the effects of soil, climate and cultivation techniques on cotton yield in central Greece. It was applied in two different ways: (i) variables including yield (ii) variables excluding yield (Kalivas and Kollias, 2001). Factor analysis was used to analyse the relationship between yield components, morphological structures above the flag leaf node, and three developmental stages in spring wheat (*Triticum aestivum L.*) (Walton, 2004). It also used to study the yield in dry beans through the analysis of plant variables (Denis and Adams, 1978). Factor analysis was used to study the dependence relationships between yield components and morphological characters for different genotypes of grasspea (*Lathyrus sativus L.*) (Tadesse and Bekele, 2001). In a study of breadmaking quality data of the wheat cultivar 'Manitou' (*Triticum aestivum L.),* factor analysis was used for samples taken from different years to find whether simple factors could be isolated to explain better the interrelationship of the quality parameters measured (Jardine, *et al.*, 1971).

With regard to barley, factor analysis has been used with breeding programmes. For example, a large set of barley data from South Australia was used to reduce the number of factors and to study the genetics of barley (Smith, *et al.*, 2001). Factor analysis was also employed to reduce the number of morphological and phenological measurements of populations of wild barley before multiple regression was utilised (Volis, *et al.*, 2002). Bratos and Szanyi (1992) used factor analysis and stepwise regression analysis to study yield and chemical composition of winter barley in a nitrogen fertilization trial. In a study intended to find the best barley for malt, factor analysis was used to discover an ideotype of a barley with a low and stable grain protein content (GPC) suitable to be used for malt (Bertholdsson, 1999).

In another field of agricultural research, factor analysis was used to study the influence of temperature change on numbers of bacteria (Nasser, 2002); and in a study of the structure of two agriculture soils, factor analysis was used to show that soil microbial communities from various plants species may differ depending on the plant species cultivated in the field (Ibekwe and Kennedy, 1998).

Furthermore, it should be noted that factor analysis has been used in many other areas of science. For example, it has been used in education to select students for universities in accordance with their results in A Level GCE examinations (Nasser, 2002).

1.4 Canonical correlation analysis

The relationship between variables can be measured by several kinds of correlations. For example, *correlation coefficients* (r) measure the relationship between two variables; *multiple regression* allows the assessment of the relationship between a set of independent variables and a dependent variable; *multiple correspondence analysis* is useful for a set of categorical variables (e.g. sex, geographic location, ethnicity).

Canonical correlation is the correlation between a set (group) of independent data and dependent data, also as a set.

"The utilization of canonical correlation analysis can provide information concerning:

1. The nature of the links or patterns of interdependency that join the two sets.
2. The number of (statistically significant) links between the sets.
3. The extent to which the variance in one set is conditional upon, or redundant given the other set" (Levine 1977).

Canonical correlation analysis has only occasionally been used in experiments with barley. It was used to study the interrelationship between quantitative characters and resistance to *Rhynchosporium secalis* in barley (Zhang, *et al.*, 1991). Canonical correlation analysis has also been used to study the effects seasonal and locations differences on barley cultivation (Vadiveloo and Phang, 1996). It was also used to discover the relationship between two sets of characters in the barley cultivation world wide (Zhang, *et al.*, 1991).

Canonical correlation analysis has been used in various research on wheat. For example, it was used to establish a relationship between yield components and morphological characters of durum wheat in one step (Wassouf, 1996); and also to study the ecological factors and nutrient-content variables in wheat (Bartos, *et al.*, 1991). In animal feed science, canonical correlation analysis has been used to study the relationship between wheat meal variability and compacting behaviour of wheat meals (Nathierdufour, *et al.*, 1995). Also relying on canonical correlation analysis, scientists discovered a strong relationship between quality characteristics and physico-chemical, rheological and protein components of wheat cultivars (Butt, *et al.*, 2001).

Canonical correlation analysis has been used in a wide range of research projects. For example, in marine science, canonical correlation analysis was used to find the relationship between zooplankton and physical variables (where four zooplankton assemblages were identified in the South western Atlantic Ocean; Marrari, *et al.*, 2004). Canonical correlation analysis has also been used in sports science to study the effect of motivational climate on sportspersonship among young competing male and female football players (Miller, *et al.*, 2004). In production research, canonical correlation analysis has been used with industry to explore the relationship between competitive strategy and the perceived value of tactical and/ or strategic flexibility (Cannon and St. John, 2004). In climatology, the relationship between bi-monthly precipitation and sea-surface temperature (SST) has also been studied by canonical correlation analysis (Berri and Bertossa, 2004). It has been used in sensory studies, to examine the relationship between oral and non-oral evaluation of texture in acid milk gels (Pereira, *et al.*, 2004). In environmental science, canonical correlation has been employed to study the relationship among spectral and phytometric variables for 20 winter-wheat fields (Korobov and Railyan, 1993); and the relationship between environment and population variables (Walsh, *et al.*, 1999). Finally, in ecological research, canonical correlation analysis has been used to study the relationship between ecological parameters for bird strategies and habitat variables describing sample plots (Tworek, 2002).

1.5 Hypothesis Testing

In real life problems, hypothesis testing is one of the most important tools for statistical science.

"There are two types of statistical inferences: estimation of population parameters and hypothesis testing" (Swinscow and Campbell, 1996). In any hypothesis testing there are four components:

1. **Null Hypothesis:** defined as no difference or relationship between the procedures and denoted by H_0 *or* H_N.
2. **Alternative Hypothesis:** a hypothesis which states that there is a difference relationship between the procedures and denoted by H_1 or H_A (Wilson and Sankaran, 1997). For example, let μ_1 be the average of the first population and μ_2 the average of the second population.

There are a lot of cases for the null hypothesis and the alternative hypothesis. Some of these are given in Table 1.5:

Table 1.5. Various types of H_0 and H_1

Cases of H_0 and H_1		
Case	H_0	H_1
1	$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$
2	$\mu_1 > \mu_2$	$\mu_1 \leq \mu_2$
3	$\mu_1 < \mu_2$	$\mu_1 \geq \mu_2$

3. **Test Statistic:** A test statistic is a quantity calculated from a sample of data. Its value is used to decide whether or not the null hypothesis should be rejected in the hypothesis test (Valerie and McColl, 2004).

4. **Conclusion:** when statistical result is related to the biological description.

Hypothesis testing is used in most areas of science; and it is used in multiple regression analysis to test whether or not the equations are statistically significant. It is employed in many forms of technological research. For example, it was used by computer programmers to evaluate five industrial-sized C++ systems (Counsell, *et al.*, 2004). Hypothesis testing has been used to study probability models of image regions in Synthetic Aperture Radar (SAR; Beaulieu, 2004). It was also used in a study designed to improve the performance of automatic speech recognizers (ASRs) with regard to spontaneous and its filled pauses (“ah”,”em”,etc.), discriminant features for filled pause detection were selected by means of Bartlett hypothesis testing (Wu and Yan, 2004). In brain physiology research, hypothesis testing has been used to study the spatio-temporal pattern of magneto-encephalography signals (Lutkenhoner, 2003). Hypothesis testing has also been used in a study of wheat genomics, in the task of assigning gene function to EST (expressed sequence tagged) databases. Hypothesis testing is particularly useful in strategies based on high resolution EST mapping, candidate gene analysis, gene expression profiling and proteomics (Lagudah, *et al.*, 2001).

In business management, researchers in the U.S. airline industry have used hypothesis testing to study different models of oligopoly (i.e. domination of the market by a few large companies; Fischer and Kamerschen, 2003).

Finally, hypothesis testing was used to evaluate questionnaires given to urban and rural El Salvadorian adolescents to assess their knowledge of a fruit called ujushte (*Brosimum alicastrum* SW.(moraceae)) (Yates and Ramirez-Sosa, 2004).

1.6 Summary

This chapter has given:

- Some comparative information about barley cultivation (area and production) in the main producing countries, with particular regard to Syria.
- The history of four kinds of statistical analysis, viz:

- History of multiple regression analysis and its uses in various fields of research and especially with regard to barley yield.
- History of factor analysis and its uses in different types of scientific enquiry including some research into barley yield.
- Canonical correlation analysis, of which a definition was given. Some examples were given of various types of scientific investigations, using canonical correlation analysis including a reference to research on barley cultivation.
- Hypothesis testing, citing its four ingredients and some examples of the use of hypothesis testing in diverse areas of research.

Chapter 2

Theoretical Descriptions

2.1 Introduction

This section will provide the mathematical models for: Regression Analysis, Factor Analysis, Canonical Correlation, Hypothesis test and Wilcoxon Signed-Rank Test which will be used to analyse the data.

2.2 Regression analysis

2.2.1 Simple and Multiple Linear Regression

2.2.1.1 Introduction

The relationship between the variables is expressed in the form of an equation connecting the response or dependent variable y, and one or more independent variables $x_1, x_2, \ldots, x_n$ (Rawlings, *et al.*, 1998). It takes the form:

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + \ldots + a_n x_{ni} + \varepsilon_i$$

where $a_0, a_1, a_2 \ldots a_n$ are called the regression coefficients. They are determined from the data. An equation that has only one independent variable is called a simple regression equation. An equation containing more than one independent variable is called a multiple regression equation (Aiken and West, 1991).

2.2.2.2 Simple Linear Regression

Simple linear relationships are used in any scientific disciplines. Simple linear regression is given by the equation:

$$y = ax + b \qquad (2 - 1)$$

Calculation of Slope and Intercept:

The steepness of a line is indicated by a number known as 'gradient'. The kind of relationship between x and y (positive or negative) is determined by slope (Allison, 1999). The slope and intercept are given by:

$$Gradient = \hat{a} = SXY / SXX = r_{xy}\frac{S_y}{S_x} = r_{xy}\left(\frac{SYY}{SXX}\right)^{1/2} \qquad (2 - 2)$$

$$Intercept = \hat{b} = \bar{y} - a\bar{x} \qquad (2 - 3)$$

Where (SXX) the corrected sum of squares for the x_i's (x_i is the x value for i^{th} data point). It is given by the equation:

$$SXX = \sum(x_i - \bar{x})^2 = \sum x_i^2 - (\sum x_i)^2 / n = \sum x_i^2 - n(\bar{x})^2 \qquad (2 - 4)$$

and the corrected sum of cross-products (SXY), covariance, is given by the equation :

$$SXY = \sum(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - (\sum x_i)(\sum y_i) / n \qquad (2 - 5)$$

The equations which given by (2 - 2) and (2 - 3) is used to find a and b in (2 - 1).

r_{XY} is the correlation coefficient which is given by the equation:

$$r_{XY} = SXY / \sqrt{(SXX)(SYY)} = S_{xy} / S_x S_y \qquad (2 - 6)$$

S_X is the standard deviation of x_i (Aiken and West, 1991). It is given by the equation:

$$S_X = \sqrt{S_X^2} = \sqrt{SXX/(n-1)} \qquad (2 - 7)^1$$

SYY is the corrected sum of squares for the y_i's, sample variance, and S_y is the sample standard deviation. They are given by similar equations. They are:

$$SYY = \sum (y_i - \bar{y})^2 = \sum y_i^2 - (\sum y_i)^2 / n = \sum y_i^2 - n(\bar{y})^2 \qquad (2 - 8)$$

$$S_y^2 = SYY/(n-1)$$

$$S_y = \sqrt{S_y^2} = \sqrt{SYY/(n-1)} \qquad (2 - 9)$$

$\hat{y}_i$ is the y value for the i'th data point and is given by the quation:

$$\hat{y}_i = \hat{a}x_i + \hat{b}$$

The value of residuals is given by: $\hat{E}_i = y_i - \hat{y}_i$.

The residual sum of squares, RSS, is given by:

$$RSS = \sum \hat{E}_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum \left[y_i - (\hat{a}x_i + \hat{b})\right]^2 \qquad (2 - 10)$$

or

$$RSS = SYY - \frac{(SXY)^2}{SXX} = SYY - \hat{b}^2 SXX$$

In simple regression, the residual degrees of freedom *(d.f.)* are given by:

$$d.f. = n\text{-}2 \qquad (2 - 11)$$

(residual d.f. = number of cases – number of parameters in model). The degrees of freedom for model (2 - 10) are (*n-2*).The sum of squares due to

[1] A sample the division is on (*n-1*) but for the population the division is on *n*

regression, SS_{reg}, is defined as the difference between the sum of squares at the equation (2 - 4) and the equation (2 - 10). It is given in (2 - 12)

$$SS_{reg} = SYY - RSS$$
$$= SYY - \left(SYY - \frac{(SXY)^2}{SXX} \right)$$

$$SS_{reg} = \frac{(SXY)^2}{SXX} \qquad (2 - 12)$$

Residual Variance[2], σ^2, is obtained by dividing RSS by its degrees of freedom *(d.f.)*. So $\hat{\sigma}^2$ is given by :

$$\hat{\sigma}^2 = \frac{RSS}{n-2} \qquad (2 - 13)$$

The degrees of freedom for model (2 - 13) are given by :

$$d.f.=(n-1)-(n-2)=1$$

The summary of results is given in ANOVA (Table 2.1).

Table 2. 1. Analysis of variance for simple regression with *n* data points

Source	Degrees of Freedom(d.f.)	Sum of Squares(SS)	Mean Square(MS)	F
Regression on X	1	SS_{reg}	$SS_{reg}/1$ or MS_{reg}	
Residual for larger model	n-2	RSS	$\hat{\sigma}^2$ or MSE	MS_{reg}/MSE
Total corrected sum of squares	n-1	SYY		

The coefficient of determination, R^2, is given by :

$$R^2 = r_{xy}^2 = \frac{SS_{reg}}{SYY} = 1 - \frac{RSS}{SYY} \qquad (2 - 14)$$

[2] this is also commonly referred to as error variance.

R^2 describes how much of the variation in the dependent variable y is explained by the equation. If R^2 is near 1, then x explains a large part of the variation in y (Timm, 2002).

The correlation coefficient is given by:

$$R = \sqrt{R^2} = r_{xy} = \sqrt{\frac{SS_{reg}}{SYY}} = \sqrt{1 - \frac{RSS}{SYY}} \qquad (2 - 15)$$

It is necessary to know whether or not R is consistent with the null hypothesis.

The appropriate statistic for testing this hypothesis is:

$$t = \frac{|R|\sqrt{n-2}}{\sqrt{1-R^2}} \qquad (2 - 16)$$

The test is carried out by comparing the observed t value with a tabulated t value with appropriate degrees of freedom (Aiken and West, 1991).

2.2.2.3 Multiple Regression Model

The relationships between several independent or predictor variables and a dependent or criterion variable are called multiple regressions (Dillon and Goldstein, 1984).

The data consists of n observations on a dependent variable y and p independent variables $x_1, x_2, ..., x_p$. The observations are usually represented as follows (Aiken and West, 1991):

Observation number	y	x_1	x_2	...	x_p
1	y_1	x_{11}	x_{21}	...	x_{p1}
2	y_2	x_{12}	x_{22}	...	x_{p2}
.	.	.	.	...	.
.	.	.	.	...	.
n	y_n	x_{1n}	x_{2n}	...	x_{pn}

The multiple regression equation is formulated as a linear model:

$$y = a_0 + a_1x_1 + a_2x_2 + ... + a_px_p$$

$a_0, a_2, ..., a_p$ are called the regression coefficients. Before computers, it was very difficult to find them. There are p equations to find the regression coefficients. These equations are given by:

$$\sum y_i = na_0 + a_1\sum x_{1i} + a_2\sum x_{2i} + a_3\sum x_{3i} + ... + a_p\sum x_{pi}$$
$$\sum y_i x_{1i} = a_0\sum x_{1i} + a_1\sum x^2{}_{1i} + a_2\sum x_{2i}x_{1i} + ... + a_p\sum x_{pi}x_{1i}$$
$$\sum y_i x_{2i} = a_0\sum x_{2i} + a_1\sum x_{1i}x_{2i} + a_2\sum x^2{}_{2i} + ... + a_p\sum x_{pi}x_{2i}$$

....

....

$$\sum y_i x_{pi} = a_0\sum x_{pi} + a_1\sum x_{1i}x_{pi} + a_2\sum x_{2i}x_{pi} + ... + a_p\sum x^2{}_{pi}$$

Now computers have solved this problem with many statistical packages, for example, SPSS, SAS and MINITAB.

The standard error of a_i ($SE_{a_i}, i = 1,2,.....,n$) is given by the equation:

$$SE_{a_1} = \frac{S}{\sqrt{\sum(x_{1i} - \bar{x}_1)^2}} \qquad (2 - 17)$$

where:

$$S^2 = \frac{SSE}{n - p - 1}$$

and:

$$RSS = SSE = \sum \left(y_i - \left(\hat{a}_0 + \hat{a}_1 x_1 + \ldots + \hat{a}_p x_p\right)\right)^2 \qquad (2-18)$$

Each of the independent variables can be tested by the equation:

$$t = \frac{a_i}{SE_{ai}}, \quad where\ i = 1,2,\ldots,p \qquad (2-19)$$

After fitting the linear model, the coefficient of determination, R^2, must be examined. The coefficient of determination (R^2) takes a value between *0* and *1*. If the value is 1: there is no difference between the estimated *y-value* and the actual *y-value*. But if coefficient of determination is *0*, then the regression equation is not useful.

The coefficient of determination is given by:

$$R^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \qquad (2-20)$$

F-test: the F statistic is used to determine whether the relationship between *y* and $x_1, x_2, \ldots, x_p$ occurs by chance. It is given by the equation:

$$F = \frac{\left(\Sigma (y_i - \bar{y})^2 - \Sigma (y_i - \hat{y}_i)^2\right) / p}{\Sigma (y_i - \hat{y}_i)^2 / (n - p - 1)} \qquad (2-21)$$

If *F* statistic is greater than the *F* table with degrees freedom *p*, *n-p-1* then the relationship between *y* and the independent variables ($x_1, x_2, \ldots, x_p$) is significant (Kendall, 1975).

Finally, a summary to multiple regression is given by ANOVA Table (Table 2.2).

Table 2. 2. Analysis of variance for multiple regression with *p* parameters and *n* data points.

Source	Degrees of Freedom(d.f.)	Sum of Squares(SS)	Mean Square(MS)	F
Regression on X	p-1	SS_{reg}	SS_{reg} /p or MS_{reg}	
Residual for larger model	n-p	RSS	$\hat{\sigma}^2$ or MSE	MS_{reg} / MSE
Total corrected sum of squares	n-1	SYY		

2.1.2 Multiple exponential regression

Introduction:

Sometimes the relationship among the variables is nonlinear. An appropriate transformation of the variable can make the relationship among the transformed variables linear (Chatterjee and Price, 1991). Many nonlinear models can be transformed into linear. For example, the function which is given by the equation $y = ax^b$ can be changed to linear, taking the natural logarithms of both sides of that equation and putting $y' = \log y$ and $x' = \log x$; the equation is: $y' = \log a + bx'$ which describes the straight line graph of log(y) against log(x) with an intercept of log(a) and a gradient of b (Eddison, 2000). There are, however, many simple nonlinear models that cannot be linearized. Consider for example: $y = a + bc^x$

Before starting to find an equation, it should be considered whether the equation suits the data. If there is not a relationship between y-variable and x-variable then a suitable model should be found. Sometimes a graph is the best method of determining the model of an equation. For example, the equation: $y = \alpha x^{\beta}$ is given by the Figure 2.1 (a and b)

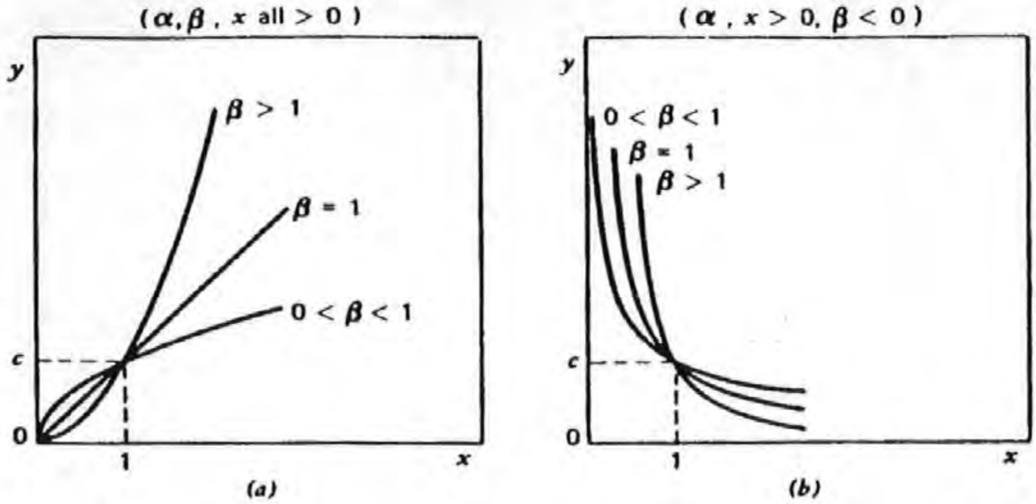


Figure 2.1. Graphs of linearizable functions $y = \alpha x^{\beta}$ (Chatterjee and Price, 1991).

Also the equation : $y = \dfrac{x}{\alpha x - \beta}$ is given by the Figure(2.2)

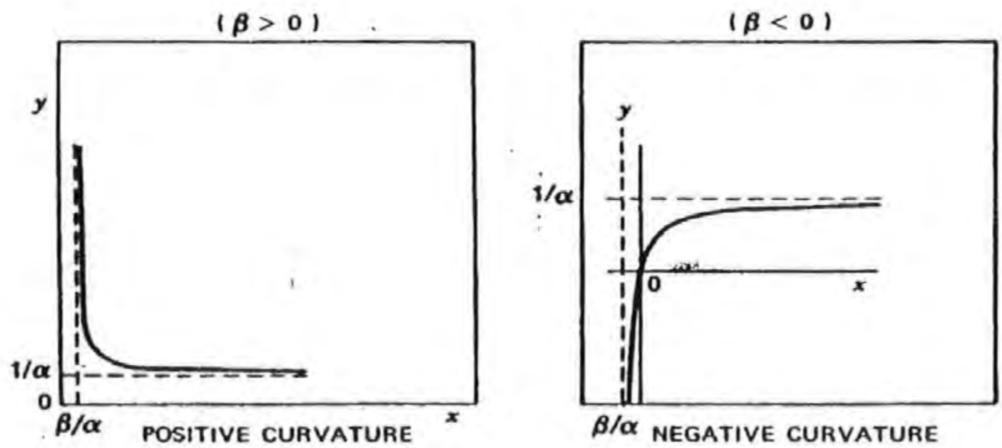


Figure 2.2. Graphs of linearizable functions $y = \dfrac{x}{\alpha x - \beta}$ (Chatterjee and Price, 1991).

Finally, the equation: $y = \frac{e^{a+bx}}{1+e^{a+bx}}$ is given by the Figure (2.3)

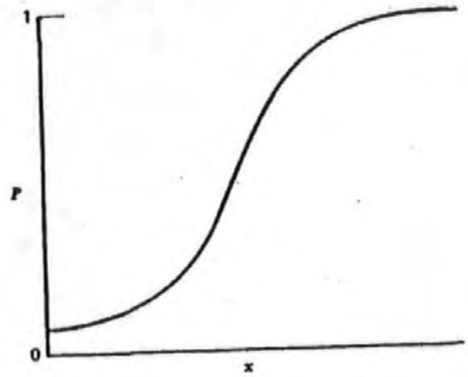


Figure 2.3. Graph of linearizable functions $y = \frac{e^{a+bx}}{1+e^{a+bx}}$ (Chatterjee and Price, 1991).

Exponential equation regression:

The relationship expressed in the form of an exponential equation is given by the form:

$$y = a_0 e^{a_1 x_1 + a_2 x_2 + \ \dots \ + a_n x_n} \qquad (2 - 22)$$

where y is a dependent variable, $x_1, x_2, \dots\dots, x_n$ are an independent variables and $a_0, a_1, a_2 \ \dots\ a_n$ are called the regression coefficients.

The equation (2-22) can be changed to linear by taking the natural logarithms of both sides of that equation and putting $y' = \log y$; the equation is:

$$y' = log\, a_0 + a_1 x_1 + a_2 x_2 + \ \dots \ + a_n x_n \qquad (2 - 23)$$

The equation (2 - 23) is a linear equation so the equation can be studied like a linear equation.

2.2 Factor analysis

2.1.1 Introduction

Factor analysis is one of the most important models in multivariate analysis. Statisticians define factor analysis as: a generic term for a family of statistical techniques concerned with reduction of a set of observable variables in terms of a small number of latent factors (Anon, 1995).

Before applying factor analysis in this research it is important to give some statistical background about factor analysis:

- **Factor loadings**, also called component loading, are the correlation coefficients between the variables and factors (Group, 2005).

- **Common variance** the variance shared with other variables in the factor analysis (Hair, *et al.* 2005).
- **Communality** h^2, is the squared multiple correlation for the variables as dependent using the factors as predictors (Group, 2005).
- **Correlation matrix** shows the inter-correlations among all variables (Hair, *et al.* 2005).
- **Eigenvalue**, also called characteristic or latent root, equals the sum of the column of squared loading for each factor (Group, 2005).
- **Varimax rotation** is an orthogonal rotation of the factor axes to maximize the variance of the squared loadings of a factor (column) on all the variables (rows) in a factor matrix, which make it as easy as possible to identify each variable with single factor (Group, 2005).
- **Error variance** Variance of a variable due to errors in data collection or measurement (Hair, *et al.* 1995).
- **Factor scores**, also called component scores, are the scores of each case (row) on each factor (column; (Group, 2006).
- **Trace** is the sum of variances for all factors, which is equal to the number of variables since the variance of a standardized variable is *1.00* (Group, 2005).
- **Unique variance** is the variability of a variable minus its communality (Group, 2006).

Let X (p,1) be a vector then the Mathematical Expectation for this vector is given by:

$$X=\begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_p \end{bmatrix} \quad \Rightarrow \quad E(X)=\begin{bmatrix} E(x_1) \\ E(x_2) \\ . \\ . \\ E(x_p) \end{bmatrix}=\begin{bmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ \mu_p \end{bmatrix} \qquad (2-24)$$

where $\mu_1,\mu_2,...,\mu_p$ are the means of $x_1,x_2,...,x_p$, respectively.

The variance and covariance to X vector is given by:

$$\Sigma = E(X-\mu)(X-\mu)' = E\left(\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ . \\ . \\ x_p-\mu_p \end{bmatrix} . \begin{bmatrix} x_1-\mu_1 & x_2-\mu 2 & . & . & x_p-\mu_p \end{bmatrix}\right)$$

$$=\begin{bmatrix} E(x_1-\mu_1)^2 & E(x_1-\mu_1)(x_2-\mu_2) & . \ . & E(x_1-\mu_1)(x_p-\mu_p) \\ E(x_2-\mu_2)(x_1-\mu_1) & E(x_2-\mu_2)^2 & . \ . & E(x_2-\mu_2)(x_p-\mu_p) \\ . & . & . \ . & . \\ . & . & . \ . & . \\ E(x_p-\mu_p)(x_1-\mu_1) & E(x_p-\mu_p)(x_2-\mu_2) & . \ . & E(x_p-\mu_p)^2 \end{bmatrix}$$

$$\Sigma=\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \sigma_{2p} \\ & & \\ & & \\ \sigma_{p1} & \sigma_{p2} & \sigma_{pp} \end{bmatrix} \qquad (2-25)$$

$$where\ \sigma_{ij}=E(x_i-\mu_1)(x_j-\mu_j);\ i,j=1,2,......,p$$

Σ[3] is called variance- covariance matrix but in this research it will be called covariance matrix. The matrix has p variances (σ_{ii}) and *p(p-1)/2* covariance ($\sigma_{ij}\ where\ i<j$)[4] (Hair, *et al.* 2005).

---

[3] Σ or σ^2 is called variance.

Correlation matrix is defined as:

$$\rho = \begin{bmatrix} 1 & \rho_{12} & . & . & \rho_{1p} \\ \rho_{21} & 1 & . & . & \rho \\ . & . & . & . & . \\ . & . & . & . & . \\ \rho_{p1} & \rho & . & . & 1 \end{bmatrix} \tag{2 - 26}$$

where:

- $\rho_{ij} = \rho_{ji}$
- $-1 \le \rho_{ij} \le 1$

The correlation coefficient is given by the equation:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} \ ; \ i, j = 1,2,\ldots\ldots, p \tag{2 - 27}$$

The standard variation matrix for X is given as:

$$V = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & . & . & 0 \\ 0 & \sqrt{\sigma_{22}} & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & \sqrt{\sigma_{pp}} \end{bmatrix} \tag{2 - 28}$$

According to (2-25), (2-26), (2-27), and (2-28) ρ *and* Σ are given by the equations:

$$\rho_{ij} = \mathbf{V}^{-1} \Sigma \mathbf{V}^{-1} \tag{2 - 29}$$

$$\Sigma = V \rho V \tag{2 - 30}$$

The linear combinations $\mathbf{Z} = \mathbf{CX}$ have:

$$\mu_Z = E(Z) = E(CX) = C\mu_X$$

$$\Sigma_Z = Cov(Z) = Cov(CX) = C \Sigma_X C' \tag{2 - 31}$$

[4] x_i , x_j are independent variables then $\operatorname{cov}(x_i , x_j) = 0$

Eigenvalue and Eigenvector: Let $\mathbf{S}(k \times k)$ be a matrix and $I(k \times k)$ be a identity matrix. Then the scalars $\lambda_1, \lambda_2, ..., \lambda_k$ satisfying the polynomial equation $|\mathbf{S} - \lambda I| = 0$ are called eigenvalues of matrix ***S***. If $\mathbf{E}(k \times 1)$ is a nonzero vector ($(\mathbf{E} \neq \mathbf{0})$ such that :

$$\mathbf{SE} = \lambda \mathbf{E} \tag{2 - 32}$$

then **E** is said to be an eigenvector (characteristic vector) of the matrix **S** associated with the eigenvalue λ (Johnson and Wichern, 1998).

2.2.2 The Factor Model

Figure 2.1 is a sample illustration of the factor analysis mode

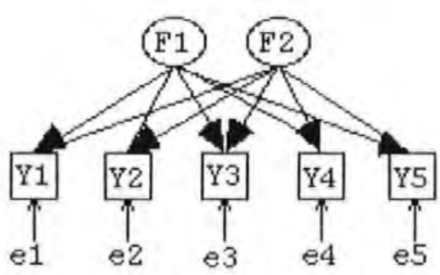


Figure 2.4. Factor analysis mode.

F_1 *and* F_2 are two common factors. y_i *where* $i = 1,...,5$ is a random vector. e_i *where* $i = 1,...,5$ are unique factors, which are assumed to be uncorrelated with each other.

The mathematical factor model is defined as:

$$x_1 - \mu_1 = l_{11}f_1 + l_{12}f_2 + ... + l_{1m}f_m + \varepsilon_1$$
$$x_2 - \mu_2 = l_{21}f_1 + l_{22}f_2 + ... + l_{2m}f_m + \varepsilon_2 \qquad (2 - 33)$$
$$...$$
$$x_p - \mu_p = l_{p1}f_1 + l_{p2}f_2 + ... + l_{pm}f_m + \varepsilon_p$$

or as matrix :

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon} \qquad (2 - 34)$$

where $X\,(p \times 1)$ is a random vector with mean μ and covariance matrix Σ. $F\,(m \times 1)$ are common (latent) factors and $m<p$. $L\,(p \times m)$ is called the factor loadings. $\varepsilon\,(p \times 1)$ is called specific or unique factors.

This definition also gives :

- $E(F) = 0 \quad and \quad E(\varepsilon) = 0$.
- $Cov(F) = I \quad and \quad Cov(\varepsilon) = \psi$. (2 - 35)
- $Cov(F, U) = 0 \quad and \quad Cov(\varepsilon_i, \varepsilon_j) = 0 \quad where\ i \neq j$

From this definition:

1. $\text{cov}(\varepsilon) = \psi = diag(\psi_{11}, ..., \psi_{pp})$.
2. $(X - \mu)(X - \mu)' = (LF + \varepsilon)(LF + \varepsilon)'$

$$(X - \mu)(X - \mu)' = (LF + \varepsilon)((LF)' + \varepsilon')$$
$$(X - \mu)(X - \mu)' = (LF)(LF)') + (\varepsilon(LF)') + LF\varepsilon + \varepsilon\varepsilon$$
$$E(X - \mu)(X - \mu)' = E((LF)(LF)') + (\varepsilon(LF)') + LF\varepsilon + \varepsilon\varepsilon)$$

From (2 - 25) and (2 - 35) we can write:

$$\Sigma = LL' + \psi \qquad (2 - 36)$$

From (2 - 36) covariance matrix can be written by factor loadings and covariance ε.

3. According to (2) the equation can be written by:

$$\sigma_{ii} = \text{var}(x_i) = l_{i1}^2 + l_{i2}^2 + \text{.........} + l_{im}^2 + \psi_{ii}$$

Let :

$$h_i^2 = l_{i1}^2 + l_{i2}^2 + \text{.........} + l_{im}^2 \qquad (2 - 37)$$

then :

$$\sigma_{ii} = \text{var}(x_i) = h_i^2 + \psi_{ii} \qquad where \quad i = 1,2,\ldots\ldots,p \qquad (2 - 38)$$

h_i^2 is called the communality and represents x_i.

ψ_{ii} is called the specific or unique variance and is due to the unique factor ε_i.

4. According to (2 - 34) :

$$(X - \mu) = (LF + \varepsilon)$$
$$(X - \mu)F' = (LF + \varepsilon)F'$$
$$(X - \mu)F' = LFF' + \varepsilon F'$$

but :

$$Cov(X,F) = E(X - \mu)(F - o)' \qquad from\ (2-35) \qquad E(F) = 0$$
$$Cov(X,F) = E(X - \mu)F'$$

then:

$$Cov(X,F) = LE(FF') + E(\varepsilon F') = L$$
$$Cov(X,F) = L \qquad (2 - 39)$$

Let T be $(m \times m)$ orthogonal matrix, then $TT' = T'T = I$. The equation (2 - 34) can be written

$$X - \mu = LF + \varepsilon = LTT'F + \varepsilon = L^*F^* + \varepsilon \qquad (2 - 40)$$

and

$$\Sigma = LL' + \psi = LTT'L' + \psi = (L^*)(L^*)' + \psi \qquad (2 - 41)$$

where

$$\mathbf{L}^* = \mathbf{LT} \qquad and \qquad \mathbf{F}^* = \mathbf{T}'\mathbf{F}$$

since

$$E(\mathbf{F}^*) = \mathbf{T}'E(\mathbf{F}) = 0$$
$$Cov(\mathbf{F}^*) = \mathbf{T}'Cov(\mathbf{F})\mathbf{T} = \mathbf{T}'\mathbf{T} = I$$

2.2.3 Methods of Estimation

The quantity s is the difference between the number of unique values in the data's $p \times p$ correlation matrix and the number of parameters in the factor model:

$$s = \frac{1}{2}(p-m)^2 - \frac{1}{2}(p+m) \quad ; \quad m \le p \qquad (2 - 42)$$

Two methods of estimating the parameters of the factor model when $s>0$ will be given in this research.

2.2.3.1 Principal Factor Analysis

Estimation of parameters of the *m-factor* model is given by principal factor analysis when $s > 0$ (Krzanowski and Marriott, 1994).

Two common estimates of the i^{th} communality (h_i^2) are:

i. The square of the multiple correlation coefficient of the i^{th} variable with all other variables.

ii. The largest correlation coefficient between the ith variable and one of the other variables.

$\hat{h}_i^2$ is higher when x_i is highly correlated with the other variables.

Variances for variables are estimated by $\hat{\sigma}_{ii} = s_{ii}$.

The reduced correlation matrix is given by the matrix $R-\psi$, where the 1 s on the diagonal here been replaced by the equation:

$$\hat{h}_i^2 = 1 - \hat{\psi}_{ii} \qquad (2 - 43)$$

According to spectral decomposition theorem:

$$R - \hat{\psi} = \sum_{i=1}^{p} \lambda_i e_{(i)} e'_{(i)} \qquad (2 - 44)$$

where $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_p$ are called eigenvalues of $R-\hat{\psi}$ and $\lambda_i \geq 0\,; i = 1,2,..,p$.

$e_{(1)},........, e_{(p)}$ are called eigenvectors of $R-\hat{\psi}$.

The *i*th column of L is estimated by the equation :

$$\hat{l}_{(i)} = \lambda_i^{1/2} e_{(i)} \quad where \;\; i = 1,2,......,m \qquad (2 - 45)$$

then :

$$\hat{L} = \gamma A^{1/2} \qquad (2 - 46)$$

where $\gamma = (e_{(1)}, e_{(2)}, ..., e_{(m)})$ *and* $A^{1/2} = diag(\lambda_1, ..., \lambda_m)$[5].

Estimates of the specific variances are given by the equation:

$$\hat{\psi}_{ii} = 1 - \sum_{i=1}^{p} \hat{l}_{ij}^2 \quad where \;\; i = 1,2,........, p \qquad (2 - 47)$$

Now if $\hat{\psi}_{ii} \geq 0$ then the principal factor solution is permissible.

$\hat{\psi}_{ii}$ is given in other equation :

$$\hat{\psi}_{ii} = \frac{1}{r^{ii}} = s^{ii} = 1 - \sum_{i=1}^{p} \hat{l}_{ij}^2 \qquad (2 - 48)$$

[5] The standardized variables each of whose estimated true variance is 1

where r^{ii} is the number i from the diagonal in the R^{-1} matrix, and s^{ii} is the number i from the diagonal in the S^{-1} matrix(S is a variance matrix).

Then from (2 - 43) the equation is:

$$\hat{h}_i^2 = 1 - \hat{\psi}_{ii} = \sum_{i=1}^{p} \hat{l}_{ij}^2 = 1 - \frac{1}{r^{ii}} = 1 - s^{ii} \qquad (2 - 49)$$

Parameters of the m-factor model are estimated by these equations.

2.2.3.2 Maximum Likelihood Factor Analysis

"It has been found to provide valid results with sample size as small as 50, but the recommended minimum sample size to ensure stable maximum likelihood estimation solution are 100 to 150" (Hair, *et al.* 2005). Also when factors F and the specific factors ε can be assumed to be normal distribution, then the maximum likelihood can be used to estimate the parameters (Child, 1990). The likelihood is given by the equation:

$$\begin{aligned} L(\mu,\Sigma) &= (2\pi)^{\frac{-np}{2}} |\Sigma|^{\frac{-n}{2}} e^{-\left(\frac{1}{2}\right)tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})' + n(\bar{x}-\mu)(\bar{x}-\mu)'\right)\right]} \\ L(\mu,\Sigma) &= (2\pi)^{-\frac{(n-1)p}{2}} |\Sigma|^{-\frac{(n-1)}{2}} e^{-\left(\frac{1}{2}\right)tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right)\right]} \times (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\left(\frac{n}{2}\right)(\bar{x}-\mu)\Sigma^{-1}(\bar{x}-\mu)} \end{aligned} \qquad (2 - 50)$$

This equation depends on $\mathbf{L}$ and ψ also $\Sigma = \mathbf{LL}' + \psi$, but this model is not well defined. The convenient uniqueness condition to define L well is given by:

$$\mathbf{L}'\psi^{-1}\mathbf{L} = \Delta \qquad (2 - 51)$$

where Δ is a diagonal matrix.

As result: let $x_1, x_2, x_3, \ldots\ldots\ldots, x_p$ are a random sample form $N_p(\mu,\Sigma)$, where

$\Sigma = \mathbf{LL}' + \psi$ is the covariance matrix for the m common factor. The maximum likelihood estimators $\hat{L}, \hat{\psi}, and\ \hat{\mu} = \overline{X}$ maximize (2 - 49) subject to $\hat{L}'\hat{\psi}^{-1}\hat{L}$ being diagonal. The maximum likelihood estimates of the communalities are:

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \dots\dots\dots\dots + \hat{l}_{im}^2 \qquad where\ i = 1,2,\dots\dots, p$$

so :

$$\begin{array}{l} \Pr oportion\ of\ total \\ sample\ \ \mathrm{var}iance\ due\ to\ jth\ factor = \dfrac{\hat{l}_{1j}^2 + \hat{l}_{2j}^2 + \dots\dots\dots\dots + \hat{l}_{pj}^2}{S_{11} + S_{22} + \dots\dots\dots\dots + S_{pp}} \end{array} \qquad (2 - 52)$$

A Large Sample Test for the Number of Common Factor:

When the sample is large, a normal distribution is assumed. Let m be the number of common the factor in this model. In this case, $\Sigma = LL' + \psi$.

Depending on hypothesis testing

$$H_0 : \underset{(p\times p)}{\Sigma} = \underset{(p\times m)}{L}\ \underset{(m\times p)}{L'} + \underset{(p\times p)}{\psi} \qquad (2 - 53)$$

$$H_1 : \Sigma\ any\ other\ positive\ definite\ matrix.$$

When Σ does-not have any special form, the maximum of likelihood function is given by:

$$|S_n|^{-n/2} e^{-np/2} \qquad (2 - 54)$$

where

$$S_n = \frac{n-1}{n} S$$

The maximum of the likelihood function is proportional to

$$|\hat{\Sigma}|^{-n/2}\exp\left(-\frac{1}{2}tr\left[\hat{\Sigma}^{-1}\left(\sum_{j=1}^{n}(X_j-\overline{X})(X_j-\overline{X})'\right)\right]\right)= \\ |\hat{L}\hat{L}'+\hat{\psi}|^{-n/2}\exp\left(-\frac{1}{2}ntr\left[(\hat{L}\hat{L}'+\psi)^{-1}S_n\right]\right) \quad (2-55)$$

where $\hat{\mu}=\overline{X}$ *and* $\hat{\Sigma}=\hat{L}\hat{L}'+\hat{\psi}$ and $\hat{L},\hat{\psi}$ are the maximum likelihood estimate of L and ψ, respectively.

The likelihood ratio statistic for testing H_0 with large sample is given:

$$-2\ln\Lambda=-2\ln\left[\frac{\max imized\ likelihood\ under\ H_0}{\max imized\ likelihood}\right] \\ =-2\ln\left(\frac{|\hat{\Sigma}|}{|S_n|}\right)^{-n/2}+n\left[tr\left(\hat{\Sigma}^{-1}S_n\right)-p\right] \quad (2-56)$$

but $tr\left(\hat{\Sigma}^{-1}S_n\right)-p=0$ and $\hat{\Sigma}=\hat{L}\hat{L}'+\hat{\psi}$ is the maximum likelihood estimate of $\Sigma=LL'+\psi$ so:

$$-2\ln\Lambda==n\ln\left(\frac{|\hat{\Sigma}|}{|S_n|}\right) \quad (2-57)$$

Degrees of freedom are given by:

$$d.f.=\frac{1}{2}\left[(p-m)^2-p-m\right] \quad (2-58)$$

Using Bartlett's correction, H_0 is rejected at the α level of significance if

$$(n-1-(2p+4m+5)/6)\ln\frac{|\hat{L}\hat{L}'+\hat{\psi}|}{|S_n|}>\chi^2_{\frac{1}{2}[(p-m)^2-p-m]}(\alpha) \quad (2-59)$$

To use this test, n and n-p must be large, and the degrees of freedom are positive that is

$$\frac{1}{2}\left[(p-m)^2-p-m\right]>0$$

then

$$\frac{1}{2}(2p+1-\sqrt{8p+1}) > m \qquad (2-60)$$

in order to apply (2-59).

2.2.4 Factor Rotation

Rotation is usually necessary to facilitate the interpretation of factors.

Rotation does not affect the sum of eigenvalues, but rotation will alter the eigenvalues (and percent of variance explained) of particular factors and will change the factor loadings (Group, 2005).

According to the equations:

$$\begin{aligned} \mathbf{L}^{*} &= \hat{\mathbf{L}}\mathbf{T}, \qquad \textit{where} \quad \mathbf{TT}' = \mathbf{T}'\mathbf{T} = I \\ \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\psi} &= \mathbf{L}'\mathbf{TT}'\hat{\mathbf{L}}' + \hat{\psi} = \hat{\mathbf{L}}^{*}\hat{\mathbf{L}}^{*\prime} + \hat{\psi} \end{aligned} \qquad (2-61)$$

then

$$\mathbf{S}_n - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\psi} = \mathbf{S}_n - \hat{\mathbf{L}}^{*}\hat{\mathbf{L}}^{*\prime} - \hat{\psi}$$

"Kaiser has suggested an analytical measure of simple structure known as varimax criterion (ϕ)" (Johnson and Wichern, 1998).The function ϕ is the sum of the variances of the squared loadings within each column of the loading matrix, where each row of loadings is normalized by its communality; that is

$$\phi = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{i=1}^{p}\tilde{l}_{ij}^{*4} - \left(\sum_{i=1}^{p}\tilde{l}_{ij}^{*2}\right)^{2} \Big/ p\right] \qquad (2-62)$$

where $\tilde{l}_{ij}^* = \hat{l}_{ij}^* / \hat{h}_i$. The varimax criterion ϕ is a function of **T**, and the iterative algorithm proposed by Kaiser finds the orthogonal matrix **T** which maximizes ϕ.

In the case where m=2, the calculations simplify. For then **T** is given by:

$$\begin{cases} T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} & \textit{Clockwise rotation} \\ T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} & \textit{Counterclockwise rotation} \end{cases} \qquad (2-63)$$

From (2-39) and (2-41) (clockwise rotation):

$$\underset{(p\times 2)}{\mathbf{L}^*} = \underset{(p\times 2)}{\hat{\mathbf{L}}} \underset{(2\times 2)}{\mathbf{T}}$$

then

$$l_{i1}^* = \{\hat{l}_{i1}\cos\theta - \hat{l}_{i2}\sin\theta\}, \qquad l_{i2}^* = \{\hat{l}_{i1}\sin\theta + \hat{l}_{i2}\cos\theta\} \qquad (2-64)$$

2.3 Canonical correlation analysis

Canonical correlation is "a multivariate statistical model that facilitates the study of interrelationship among sets of multiple dependent variables and multiple independent variables" (Hair, *et al.* 2005). Canonical correlation analysis (CCA) which was first described by Hotelling (1935)(Shafto, *et al.*, 2005), is used in many fields like chemistry, biology, meteorology, sociology, economics,....etc. (Shafto, *et al.*, 2005).

Before starting with a canonical correlation model some definitions may be useful:

1- **Canonical function** is the relationship between two separate linear composites (canonical variates). One linear composite is for the set of criterion variables, the other for the set of predictor variables. Canonical correlation gives the strength of the relationship between the two sets (Hair, *et al.* 2005).

2- **Canonical loadings** measure the simple linear correlation between independent variable and their respective canonical variates. Canonical loadings are sometimes called canonical structure correlations. They can be interpreted in the same way as factor loadings.

3- **Canonical roots (eigenvalues)** are squared canonical correlations. They are used to give an estimate of the degree of shared variance between the linear composites (canonical variate), optimally weighted, of criterion and predictor variables respectively (Hair, *et al.* 2005).

4- **Canonical variates** are also known as **linear composites**, linear compounds, linear combinations, or canonical variables. They represent the weighed sum of two variables. They can be defined for either criterion variables or for predictor variables (Tabachnick and Fidell, 2001).

5- **Canonical coefficients (canonical weight)** are used to measure the relative importance in a conical correlation of the contribution of individual variables.

6- **Criterion variables:** dependent variables.

7- **Predictor variables:** independent variables.

8- **Pooled canonical correlation** is "the sum of the squares of all the canonical correlation coefficients, representing all the orthogonal dimensions in the solution by which the two sets of variable are related. Pooled canonical correlation is used to assess the extent to which one set of variables can be predicted or explained by the other set"(Anon, 2004a).

An analysis of the relationship with canonical correlation is given by the Figure 2.5:

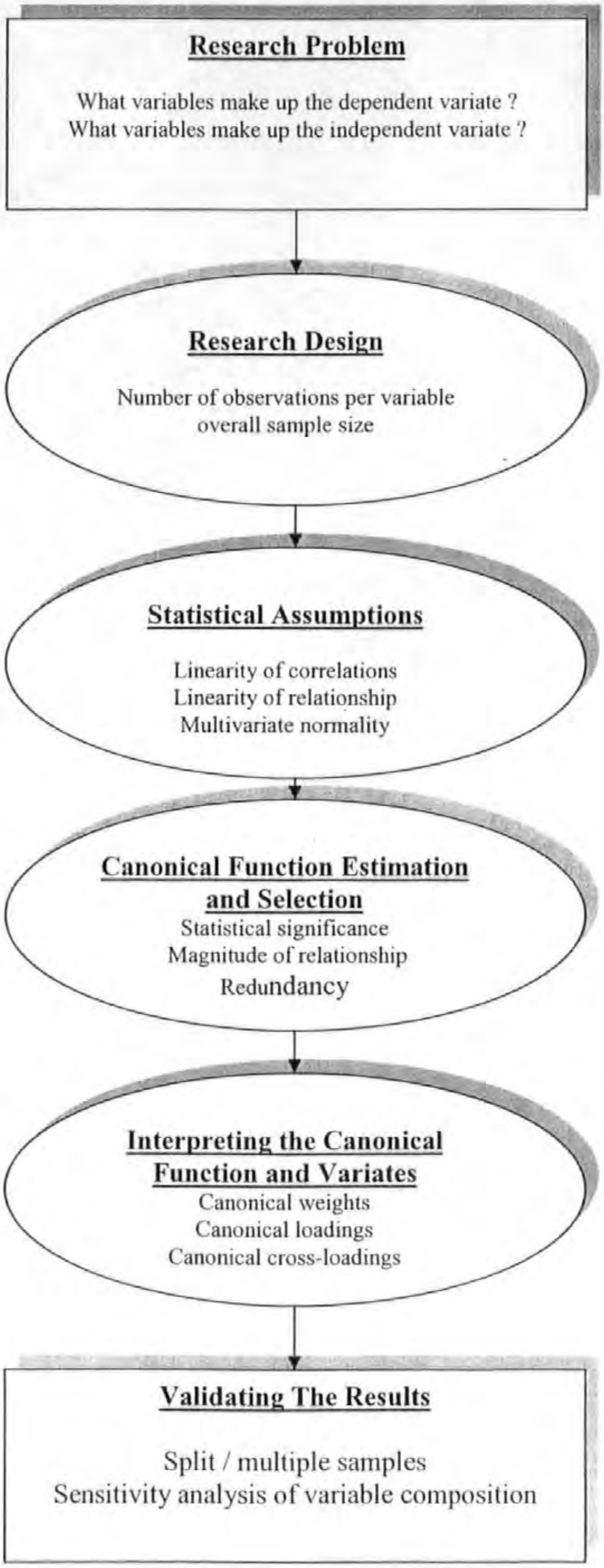


Figure 2.5. **Canonical Correlation Analysis (Hair, 2005).**

2.3.1 Mathematical canonical correlation model

Let $X^{(1)}$ is a random vector $(p\times 1)$ and $X^{(2)}$ is the second group $(q\times 1)$ and $p\le q$. For these vectors let:

$$\begin{aligned} E(X^{(1)}) &= \mu^{(1)} \quad , \quad \operatorname{cov}(X^{(1)})=\Sigma_{11} \\ E(X^{(2)}) &= \mu^{(2)} \quad , \quad \operatorname{cov}(X^{(2)})=\Sigma_{22} \\ \operatorname{cov}(X^{(1)},X^{(2)}) &= \Sigma_{12}=\Sigma_{21} \end{aligned} \tag{2 - 65}$$

then the random vector is given by:

$$X_{((p+q)\times 1)}=\begin{bmatrix} X^{(1)} \\ ---- \\ X^{(2)} \end{bmatrix}=\begin{bmatrix} X_1^{(1)} \\ .... \\ X_p^{(1)} \\ ---- \\ X_1^{(2)} \\ .... \\ X_q^{(2)} \end{bmatrix} \tag{2 - 66}$$

and:

$$\mu_{((p+q)\times 1)}=E(X)=\begin{bmatrix} E(X^{(1)}) \\ ---- \\ E(X^{(2)}) \end{bmatrix}=\begin{bmatrix} \mu^{(1)} \\ ---- \\ \mu^{(2)} \end{bmatrix} \tag{2 - 67}$$

The covariance of $X\,((p+q)\times 1)$ is given by the equation:

$$\Sigma_{((p+q)\times 1)}=E(X-\mu)(X-\mu)'$$

$$\Sigma_{((p+q)\times 1)}=\begin{bmatrix} E(X^{(1)}-\mu^{(1)})(X^{(1)}-\mu^{(1)})' & \vdots & E(X^{(1)}-\mu^{(1)})(X^{(2)}-\mu^{(2)})' \\ ---------- & --- & ---------- \\ E(X^{(2)}-\mu^{(2)})(X^{(1)}-\mu^{(1)})' & \vdots & E(X^{(2)}-\mu^{(2)})(X^{(2)}-\mu^{(2)})' \end{bmatrix}$$

$$\Sigma_{((p+q)\times 1)}=\begin{bmatrix} \underset{p\times p}{\Sigma_{11}} & \vdots & \underset{p\times q}{\Sigma_{12}} \\ --- & -- & --- \\ \underset{q\times p}{\Sigma_{21}} & \vdots & \underset{q\times q}{\Sigma_{22}} \end{bmatrix} \tag{2 - 68}$$

$X^{(1)},X^{(2)}$ (two sets of variables) are written as linear combinations as:

$$U = a'X^{(1)}$$
$$V = b'X^{(2)} \qquad (2-69)$$

According to (2-8), there are:

$$Var(U) = a'Cov(X^{(1)})a = a'\Sigma_{11}\, a$$
$$Var(V) = b'Cov(X^{(2)})b = b'\Sigma_{22}\, b \qquad (2-70)$$
$$Cov(U,V) = a'Cov(X^{(1)}, X^{(2)})b = a'\Sigma_{12}\, b$$

then the canonical correlation is given by the equation:

$$Corr(U,V) = \frac{a'\Sigma_{12}\, b}{\sqrt{a'\Sigma_{11}\, a}\sqrt{b'\Sigma_{22}\, b}} \qquad (2-71)$$

Theorem 4-1: Suppose $p \le q$ and let the random vectors $\underset{p\times 1}{X^{(1)}}$ and $\underset{q\times 1}{X^{(2)}}$ have $Cov(X^{(1)}) = \underset{(p\times p)}{\Sigma_{11}}$, $Cov(X^{(2)}) = \underset{(q\times q)}{\Sigma_{22}}$, $Cov(X^{(1)}, X^{(2)}) = \underset{(p\times q)}{\Sigma_{12}}$ where Σ has full rank. For coefficient vectors *a* and *b*, from the linear combinations $U = a'X^{(1)}$ *and* $V = b'X^{(2)}$. Then:

$$\max_{a,b} Corr(U,V) = \rho_1^*$$

attained by the linear combinations(first canonical variate pair)

$$U_1 = e_1'\Sigma_{11}^{-1/2} X^{(1)} \quad and \quad V_1 = f_1'\Sigma_{22}^{-1} X^{(2)}$$

Then:

$$a_1' = e_1'\Sigma_{11}^{-1/2} \quad and \quad b_1' = f_1'\Sigma_{22}^{-1/2}$$

The *k*th pair of canonical variates, *k=2,3,......,p*

$$U_k = e_k'\Sigma_{11}^{-1/2} X^{(1)} \quad and \quad V_k = f_k'\Sigma_{22}^{-1} X^{(2)}$$

maximizes

$$Corr(U_k, V_k) = \rho_k^*$$

among those linear combinations uncorrelated with the preceding *1,2,......,k-1* canonical variables.

Here $\rho_1^{*2} \geq \rho_2^{*2} \geq .... \geq \rho_p^{*2}$ are the eigenvalues of $\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$, and $e_1, e_2, ..., e_p$ are the associated $(p \times 1)$ eigenvectors. (The quantities $\rho_1^{*2} \geq \rho_2^{*2} \geq .... \geq \rho_p^{*2}$ are also the p largest eigenvalue of the matrix $\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2}$ with corresponding $(q \times 1)$ eigenvectors $f_1, f_2, ..., f_p$. Each f_i is proportional to $\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1}e_i.)$. The canonical variates have the properties:

$$\begin{aligned} Var(U_k) &= Var(V_k) = 1 \\ Cov(U_k, U_l) &= Corr(U_k, U_l) = 0 \quad k \neq l \\ Cov(V_k, V_l) &= Corr(V_k, V_l) = 0 \quad k \neq l \\ Cov(U_k, V_l) &= Corr(U_k, V_l) = 0 \quad k \neq l \end{aligned}$$

for $k, l = 1, 2, ..., p$ (Johnson and Wichern, 1998).

2.3.2 Tests of significance

The test for relationship between $X^{(1)}$ and $X^{(2)}$ was proposed by Bartlett and is given by the equation:

$$\chi^2 = -\{n - \frac{1}{2}(p + q + 3)\}\sum \log_e(1 - \lambda_i) \qquad (2 - 72)$$

where n is the number of cases for which data are available. r is the minimum among p and q . λ_i is the square of canonical correlation. The value of X^2 can be compared with the percentage point of chi-squared distribution with pq degrees of freedom. A non-significant result indicates that canonical correlation can be accounted for by sampling variation only. A significantly large value provides evidence that at a one of the r canonical correlations is significant. (Manly, 1994).

Note: canonical correlation analysis is used only when: ($n > p + q$) (Clark, 1977).

2.4 Hypothesis testing

Let X_1, X_2 two samples from the same population then the hypothesis testing for the difference between the means is given by the equation (Wilson and Sankaran, 1997):

$$T = \frac{\overline{X}_1 - \overline{X}_2}{S_{(\overline{X}_1, \overline{X}_2)}}$$

where :

- $\overline{X}_1$ is the average of the first sample.
- $\overline{X}_2$ is the average of the second sample.
- $S_{(\overline{X}_1, \overline{X}_2)}$ is the standard error of the difference which is given by the equation:

$$S_{(\overline{X}_1, \overline{X}_2)} = \sqrt{\frac{\sigma^2_{\overline{X}_1}}{n_1} + \frac{\sigma^2_{\overline{X}_2}}{n_2}}$$

where:

- $\sigma^2_{\overline{X}_1}$ is the variance of the first sample.

- $\sigma^2_{\bar{x}_2}$ is the variance of the second sample.
- n_1 is the number of variables in the first sample.
- n_2 is the number of variables in the second sample.

t statistic and t tabular will be compared in order to establish whether or not the weather and location affect barley breeding experiments.

2.5 Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank sum test is a non-parametric or distribution free test. It is used to test that the median of equal to some value (Shier, 2004). The Wilcoxon signed rank sum test is given as:

1. The null hypothesis: the difference between the members of each pair (x, y) has median value zero.
2. Calculate each paired different: $d_i = x_i - y_i$.
3. Rank the differences without regard to the sign of the difference.
4. Calculate all positive ranks (W^+) and all negative ranks (W^-) (where $W^+ + W^- = \frac{n(n+1)}{2}$, n is the number of observation).
5. Calculate the mean and variance which are given by the equations: $\mu = \frac{n(n+1)}{4}$, $\sigma^2_W = \frac{n(n+1)(2n+1)}{24}$ ($n>8$ or Normal distribution).

6. Calculate $Z = \frac{W - \mu}{\sigma_W}$, where W is the maximum between W^+ and W^-.

7. Find the probability of observing a value Z by using tables of critical values for Wilcoxon signed-rank sum test.

Chapter 3

Materials and Methodology

3.1 Description and Transferring the Data

The experiments from which the data sets were drawn for this research were conducted during two successive seasons 1992/1993 and 1993/1994 at two experimental stations Tel Hadya and Breda by the International Centre for Agricultural Research in the Dry Area (ICARDA).

There were three (F_5) hybrids, and each hybrid had one hundred families derived from crossing two six-rowed genotypes of barley MO.B1337/W1291//Zambaka. Their parents were planted at Tel Hadya during the first season. During the second season (F_6) hybrids and their parents (the original parents family numbers were 101 and 102) were planted at Tel Hadya and Breda and a number of parameters recorded (Table 3.1).

Table 3.1. The parameters (n=102 families) measured for each of 3 hybrids at 2 sites and 2 seasons (see Table a.2 in appendices).

Season	Tel Hadya	Breda
Season 1	Total plant yield (kg ha^{-1}) Grain yield (kg ha^{-1}) Straw yield (kg ha^{-1}) Harvest index Thousand grain weight (TGW) Ear number Protein content of grain Plant height (cm) Vegetative duration (day)	-
Season 2	Total plant yield (kg ha^{-1}) Grain yield (kg ha^{-1}) Straw yield (kg ha^{-1}) Harvest index Thousand grain weight (TGW) Ear number Protein content of grain Plant height (cm) Vegetative duration (day) Length of growing season (day) Leafiness	Total plant yield (kg ha^{-1}) Grain yield (kg ha^{-1}) Straw yield (kg ha^{-1}) Harvest index Thousand grain weight (TGW) Ear number Protein content of grain Plant height (cm) Vegetative duration (day) Length of growing season (day) Leafiness

In the first season, a Randomized Block Design with three replicates was used in Tel Hadya and Table 3.2 gives a summary of agriculture inputs during the first season.

In the second season, a Randomized Block Design with two replicates was used in Tel Hadya and Breda and Table 3.3 gives a summary of inputs during the second season.

Table 3.2. Agricultural input factors during the first season in Tel Hadya.

Factor	Tel Hadya
Date of Cultivation	29/11/1992
Date of Germination	26/12/1993
Date of Harvest	6/7/1993
Seed rate (kg ha^{-1})	100
Rate of Nitrogen (kg (N) ha^{-1})	40
Rate of Phosphate (kg (P_2O_5)ha^{-1})	40 Before Cultivation
Herbicide for Grass control	W46(3 litre ha^{-1})
Rainfall (mm/year)	290.1

Table 3.3. Agricultural input factors during the second season in Tel Hadya and Breda.

Factor	Tel Hadya	Breda
Date of Cultivation	12/12/1993	15/12/1993
Date of Germination	24/12/1993	5/1/1994
Date of Harvest	3/6/1994	9/6/1994
Seed rate (kg ha^{-1})	100	100
Rate of Nitrogen (kg (N) ha^{-1})	60	60
Herbicides	Deblozan (1.5 l ha^{-1})	-
Rainfall (mm/year)	273.3	291.2

3.2 Description of the data (attributes of barley)

The data are divided into two sets, morphological characters and yield parameters.

3.2.1 Morphological characters

a. **Plant height** was estimated by taking the mean of three random samples from each experimental plot.

b. **Leafiness** was recorded as a five point scores estimated by eye at ear emergence.

Score 1: very low leafiness.

Score 2: low.

Score 3: medium.

Score 4: high.

Score 5: very high.

c. **Vegetative duration** was the number of days from germination until 50% ear emergence.

d. **Length of growing season** was the number of days from germination to harvest.

3.2.2 Yield parameters

a. **Total plant yield**: a two metre plot length of barley was harvested from each experimental plot after removing the border rows. Total plant yield was recorded and expressed as kg ha^{-1}.

b. **Grain yield:** was measured after mechanically separating grain from straw expressed as kg ha^{-1}.

c. **Harvest index:** is defined as the ratio of grain yield to the total plant yield at harvest.

d. **Thousand grain weight (TGW)**: 200 seeds were separated from each grain sample weighed and the result multiplied by five.

e. **Protein content of grain**: the protein content of a 20 g seed sample was measured using a NIRSYSTEM-500.

f. **Ear number:** the number of ears was counted in a 50 cm row length and from this the number of ears per square metre was calculated by multiplying by 2 and dividing by the row width.

3.3 Aims and methodology

It was difficult for the breeders to decide which family produced the highest yield. Since when family yields were ranked there appeared to be a genotype-environment interaction for grain yield, this is illustrated for Hybrid 1 (H1) (Table 3.4). Total rank indicated that the genotype was not stable because the the total ranks was between *19* (family 28) and *276* (family 34)(Table 3.4), and for example family 39 was 2nd in Tel Hadya in season 1, but 41st in season 2 and 28th in Breda in season 2.

Table 3.4. Example of a data set - grain yield (kg ha^{-1}) for H1 in rank order and family number at 2 sites.

family	Tel Hadya season 1		Tel Hadya season 2		Breda season 2		Total of ranks
	Grain yield	Rank	Grain yield	Rank	Grain yield	Rank	
1	2378	99	4475	39	1578	42	180
2	3442	59	3093	96	1500	64	219
3	3622	49	3916	74	1858	15	138
4	3662	46	5750	6	1766	24	76
5	4080	27	5333	12	1775	22	61
6	3800	41	5168	16	1516	60	117
7	4298	19	4466	40	1841	18	77
8	4009	30	4433	45	1256	87	162
9	3333	66	4233	58	1783	20	144
10	2653	93	3350	91	1250	89	273
11	3978	33	4125	66	941	101	200
12	2604	95	4310	52	1541	53	200
13	2311	100	5341	11	1558	46	157
14	3676	44	3253	92	966	100	236
15	3168	76	4521	35	1478	65	176
16	3013	84	3191	94	1403	72	250
17	2622	94	3525	83	1786	19	196
18	2307	101	3658	80	1858	16	197
19	3391	64	4401	48	1473	67	179
20	5435	1	5775	5	1628	34	40
21	4635	10	4001	71	1358	78	159
22	3840	39	3833	77	1666	30	146
23	3600	53	2675	102	1450	69	224
24	3866	37	4453	42	1550	52	131
25	4986	5	4223	59	1890	11	75
26	5200	4	3846	76	1941	10	90
27	3444	58	4608	32	1511	63	153
28	4493	14	5791	4	2400	1	19
29	3302	67	3958	73	2111	6	146
30	4084	26	4875	25	1525	57	108
31	2675	90	4186	62	1953	9	161
32	4217	23	3460	87	1661	32	142
33	4186	25	4476	38	1345	79	142
34	3191	75	2883	99	933	102	276
35	3520	56	4508	36	1600	38	130
36	3235	72	3000	98	1775	23	193
37	2862	86	4758	27	1866	14	127
38	4622	11	3983	72	1370	76	159
39	5400	2	4458	41	1703	28	71
40	3257	70	4075	69	1125	96	235
41	3284	69	4958	22	1978	8	99
42	3111	81	3476	84	1711	27	192
43	3828	40	2875	100	1200	93	233
44	3617	50	4600	33	1575	43	126
45	4591	12	4291	54	1516	61	127
46	4724	9	5450	8	1345	80	97

Table 3.4. (contd.)

family	Tel Hadya season 1		Tel Hadya season 2		Breda season 2		Total of ranks
	Grain yield	Rank	Grain yield	Rank	Grain yield	Rank	
47	3462	57	3126	95	1378	75	227
48	4288	20	4210	60	1108	97	177
49	3164	77	4291	55	1150	94	226
50	3524	55	3660	79	1558	47	181
51	5360	3	3425	88	1145	95	186
52	2288	102	2800	101	1875	12	215
53	3920	36	4416	46	1253	88	170
54	2848	87	4750	28	1558	48	163
55	4346	17	4941	23	1625	35	75
56	3408	61	4453	43	1283	84	188
57	3666	45	4076	68	1616	36	149
58	3657	47	3358	90	1700	29	166
59	4533	13	5433	9	2028	7	29
60	4280	21	3566	82	1525	58	161
61	4302	18	4833	26	1528	56	100
62	4480	15	5350	10	1533	54	79
63	2475	98	4033	70	1316	82	250
64	3222	73	3633	81	1300	83	237
65	3855	38	4450	44	1586	41	123
66	3133	79	5800	3	1258	86	168
67	3604	52	4383	49	2183	4	105
68	3408	62	5216	15	1533	55	132
69	3084	82	5010	19	1783	21	122
70	4844	6	5141	17	1066	98	121
71	4804	7	4083	67	1053	99	173
72	3244	71	4916	24	2316	2	97
73	2791	88	5316	13	1391	74	175
74	3573	54	5266	14	1456	68	136
75	3933	35	4360	51	1558	49	135
76	3650	48	6193	1	1225	90	139
77	3768	42	4666	31	1333	81	154
78	2742	89	4150	63	1225	91	243
79	2666	92	4416	47	1600	39	178
80	3715	43	3066	97	1558	50	190
81	4764	8	3800	78	1558	51	137
82	3013	85	3400	89	1595	40	214
83	4053	28	4726	29	2166	5	62
84	4000	31	4383	50	1666	31	112
85	3608	51	4968	21	1370	77	149
86	3435	60	4260	56	1428	71	187
87	4191	24	4258	57	1475	66	147
88	3124	80	5126	18	1858	17	115
89	4377	16	3475	85	2236	3	104
90	4222	22	6125	2	1650	33	57
91	3297	68	4301	53	1875	13	134
92	2502	97	4691	30	1450	70	197
93	3022	83	3475	86	1266	85	254
94	3991	32	4500	37	1516	62	131

Table 3.4. (contd.)

family	Tel Hadya season 1		Tel Hadya season 2		Breda season 2		Total of ranks
	Grain yield	Rank	Grain yield	Rank	Grain yield	Rank	
95	3355	65	3875	75	1716	26	166
96	3400	63	4150	64	1561	45	172
97	2671	91	4188	61	1611	37	189
98	3204	74	4126	65	1211	92	231
99	4044	29	4593	34	1725	25	88
100	3951	34	5666	7	1570	44	85
101	3137	78	4991	20	1525	59	157
102	2503	96	3253	93	1400	73	262

The highest and the lowest ten families for grain yield were extracted from Table 3.4 and are presented in Table 3.5.

Table 3.5. The highest and the lowest ten families for each site/season (H1).

Rank	Tel Hadya season 1 family	Tel Hadya season 2 family	Breda season 2 family	Rank	Tel Hadya season 1 family	Tel Hadya season 2 family	Breda season 2 family
1	20	76	28	93	10	102	43
2	39	90	72	94	17	16	49
3	51	66	89	95	12	47	51
4	26	28	67	96	102	2	40
5	25	20	83	97	92	80	48
6	70	4	29	98	63	36	70
7	71	100	59	99	1	34	71
8	81	46	41	100	13	43	14
9	46	59	31	101	18	52	11
10	21	62	26	102	52	23	34

No families in the top ten highest grain yield rank appeared in top ten at any other location (Tel Hadya and Breda) and season (season one and season two; Table 3.5) at the same time with exception of family 26. However, three families were strongly affected by environment since they changed from the highest yield to the lowest. For example, family 51 produced the third highest grain yield in Tel Hadya, season one, while it produced the eighth lowest yield of all for Breda, season two. The genotype for Hybrid 2 (H2) and Hybrid 3

(H3) was also affected by environment (Table a.3; Table a.4 respectively Appendix). For example for example family 80th for H2 was 6th in Tel Hadya in season 1, but 23rd in season 2 and 101st in Breda in season 2.

The highest and the lowest ten families for grain yield H2 and H3 are derived from Table a.3 and Table a.4, presented respectively in Table 3.6 and 3.7. Only one family (11) for H2 and family 21 for H3 from the highest grain yield rank appeared in the other location (Tel Hadya and Breda; Table 3.6 and 3.7). this indicated examination of just one of the 9-11 parameters measured demonstrates how difficult it is to understand fully what the data shows. Genotype- environment interactions are clearly apparent for yield and are likely to exist for the other parameters. Whether such variations are in harmony between yield and other parameters can only be investigated by statistical methods (more information will be given in Chapter 4).

Table 3.6. The highest and the lowest ten families for each site/season (H2).

Rank	Tel Hadya season 1 family	Tel Hadya season 2 family	Breda season 2 family	Rank	Tel Hadya season 1 family	Tel Hadya season 2 family	Breda season 2 family
1	63	85	28	93	78	38	15
2	46	5	11	94	15	35	14
3	55	102	58	95	38	6	53
4	40	9	13	96	23	90	79
5	100	22	34	97	16	79	65
6	80	60	96	98	43	71	89
7	10	62	70	99	13	39	101
8	1	52	23	100	76	57	31
9	27	11	8	101	47	2	80
10	14	93	25	102	42	89	7

Table 3.7. The highest and the lowest ten families for each site/season (H3).

Rank	Tel Hadya season 1 family	Tel Hadya season 2 family	Breda season 2 family	Rank	Tel Hadya season 1 family	Tel Hadya season 2 family	Breda season 2 family
1	87	17	21	93	32	8	77
2	67	66	13	94	30	62	68
3	71	6	85	95	25	51	93
4	21	35	51	96	82	32	16
5	80	95	82	97	44	5	33
6	89	84	9	98	83	93	4
7	34	57	63	99	73	71	22
8	28	21	46	100	97	63	73
9	88	52	94	101	76	54	66
10	13	39	41	102	57	16	67

The ultimate goal of this research was:

- to study whether or not the genotypes were affected by environment and to determine which was the best hybrid in terms of productivity.
- to determine whether or not yield parameters were affected by morphological characters; also to determine whether or not the effectiveness was altered by environmental changes associated with locations (Tel Hadya and Breda) and seasons (1992/1993-1993/1994) with the aim of finding appropriate analytical tools to improve the implementation of barley breeding programmes and assist the interpretation of the data generated.
- to identify and measure any single character which gives the best prediction; and also to discover the best statistical methods for discovering this character.
- to investigate the form of the relationship between yield parameters and morphological characters.

Various kinds of statistical analysis have previously been used to study genotype-environment interactions. Some methods have been found to be useful, however the constraints of the experimental design of this investigation and the data preclude their use here. Specifically, (i) the data in Tel Hadya were collected in two seasons while the data in Breda were collected in the second season only. (ii) Length of growing season and leafiness were measured in the second season whereas they were not measured in the first season. (iii) The sample sizes for each hybrid were insufficient.

The key methods that have not been applied in this study include:

- AMMI: Additive main effects and multiplicative interaction (AMMI) is a type of multivariate method used to supply a biplot for discovering the main effects and interactions between genotype and environment (Kang, 2002).
- GLM: General linear model (GLM) is also a type of multivariate method used to find genotype-environment interactions (genotype × years, genotype × locations and genotype × years × locations) (Anon, 2005).
- Principal coordinate analysis, as detailed by Westcott (1987), analyses genotype means for each environment, highlighting performance features. His method, in certain circumstances, has advantages over methods based on regression, cluster and principal components analyses.
- SHUKLA's stability variance (Shukla, 1972): a univariate parametric ANOVA method that provides "an unbiased GEI variance attributed to each genotype." (Kang 2002).

Other methods could have been applied to study the relationship between yield parameters and morphological characters; for example, multilevel analysis (multilevel factor analysis and multilevel regression analysis) (Hox, 1995). Similarly, however, the experimental design and data quality of this investigation preclude its use.

New approaches to the statistical analysis of barley breeding experiments, specifically *Factor Analysis* and *Canonical Correlation*, will be used to study the relationship between yield parameters as a set and morphological characters, also as a set. *Stepwise multiple regression analysis* between morphologies and yield will be used to determine whether or not there is a relationship, and which is the best equation for describing barley breeding data (linear or exponential). Finally, the relationship between yield parameters and morphological characters for each hybrid and area will be studied by multiple regression analysis (linear and exponential).

Five stages will be studied in this project:

Stage 1: Hypothesis testing and Genotype-Environment interaction

The yield component is affected by many factors, for example, temperature location, and rainfall. This stage will study the hypothesis testing for different independent samples. This stage also studies whether or not there is genotype-environment interaction.

Stage 2: Canonical analysis

The research will aim to determine whether there is relationship between the yield parameters as a set and morphological characters also as a set, and will display alteration due to differences of season or location.

Canonical correlation is used to study the relationship between morphologies of barley as a set (vegetative duration x_1, plant height x_2, length of growing season x_3 and leafiness x_4) and yield parameters of barley, also as a set, (total yield y_1, grain yield y_2, straw yield y_3 and thousand grain weight (TGW) y_4). When a canonical correlation is found, the statistical significance in canonical correlation uses the Bartlett test (Levine, 1977). Then canonical analysis will demonstrate whether or not there are relationships between the morphologies and yield parameters.

Stage 3: Factor analysis

This investigation will use factor analysis to find the interrelationship among various morphological characters and yield parameters. Factor analysis will be used to reduce a large number of correlated variables into a smaller number of unrelated factors.

Stage 4: Multiple regression analysis

In this section, regression analysis (stepwise and multiple regression) will be used with the aims to determine whether there is a relationship between the morphological characters and the yield parameters, and whether or not this relationship will be altered by differences of season or location.

In this research, the effectiveness of *multiple exponential regression* in agricultural analyses will be compared with *multiple linear regression*.

Stage 5: Conclusion, Summary and Future plans

The project can be summarized by the Figure 3.1

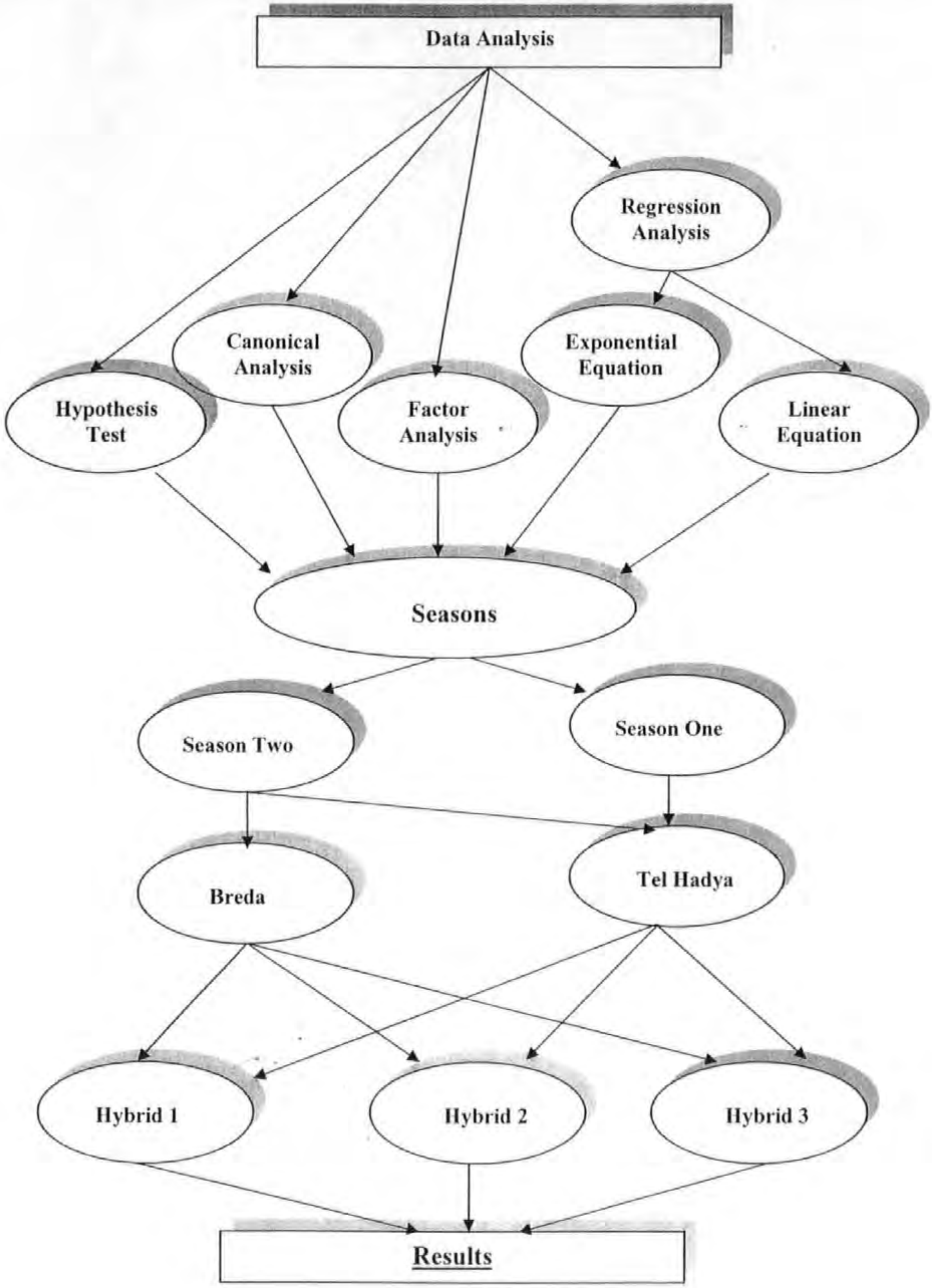


Figure 3.1. Project plan summary.

Chapter 4

Hypothesis Testing and Genotype-Environment Interaction

4.1 Introduction

Barley yield is affected by many factors but these can be grouped as either environmental or genotypical characters. The final yield of the crop is the result of the expression of the genotype resulting in morphological and physiological processes which then interact with the environment in order to create biomass which eventually ends up as grain and straw yield. The data available for analysis in this investigation were not collected from a set of designed physiological experiments but from breeding trials, and thus only a limited number of parameters have been recorded. Nevertheless, the aim of this chapter is to investigate these characters, their stability with respect to genotype and their stability with respect to environment in order to evaluate their usefulness and reliability for use in further analysis. The parameters and effects will be analysed by hypothesis test, employing different independent samples (yield parameters) in different cases.

Firstly, some definitions are necessary:

- Genotype: genotype refers to cultivar rather than to an individual's genetic make-up (Annicchiarico, 2002).
- Environment: Environment relates to a the set of climatic, soil, biotic (pests and diseases) and management conditions in an individual trial carried out at a given location in one year or over several years (Annicchiarico, 2002).

This chapter is divided into three sections:

A. The first will deal with the difference between yield parameters (total plant yield, grain yield and straw yield) in relation to levels of leafiness during the second season. The section will also suggest reasons for different levels of leafiness and will also examine whether or not the yield was affected by levels of leafiness.

B. The second section will compare the mean yield parameters and morphological characters in different seasons in the same area to examine the effects of the weather.

C. The third section will study the genotype environmental interaction. Firstly, two different seasons in the same area will be compared to assess further the effects of the environment on the results genotype× year. Secondly, the interaction between environment and genotype (genotype× location). Finally, the reasons why three different hybrids (Hybrid 1 (H1), Hybrid 2 (H2), and Hybrid 3 (H3)) gave different yield parameters will be investigated.

4.2 The influence of leafiness on yield

Leaves begin as a regular series of primordial, localized outgrowths on the sides of the apical dome of a vegetative shoot (Hay and Walker, 1989) and then undergo extensive cell division, differentiation and expansion resulting in leaf emergence. Barley leaves comprise a leaf blade and a leaf sheath and are connected to the crown (vegetative condition) or the stem (reproductive condition) at a node. Leaves are one of the most important factors influencing yield since they are the photosynthetic factory of the plant, fixing light into dry matter which ultimately determines yield. Many factors influence the development of crop leaves. The four main ones are: (i) temperature, which affects the rate of leaf production and the rate and duration of leaf expansion (Hay and Walker, 1989); (ii) nitrogen, which has an effect on leaf size and longevity, and photosynthetic ability (Hay and Walker, 1989); (iii) population density, which can limit leaf size through interplant competition; (iv) water supply, which affects leaf size and longevity as well as photosynthetic ability (Hay and Walker, 1989) . There are secondary controls on crop leaf area such as abiotic and biotic stresses (frost, high temperature, wind, diseases pests). A physiological study would normally measure crop leaf area and derive the leaf area index (LAI) but the score used here in the data, *leafiness*, is a crude breeders scale to reflect this characteristic.

First, this section will study whether or not leafiness is a stable character in the data or whether it varies from site to site. Second, the section will study the differences in yield in relation to level of leafiness to determine whether leafiness can be used to predict yield or not.

There are different methods of discovering whether or not genotype is affected by the environment; and in this section, the crossover interaction will be described in diagrammatic form and analysed using Spearman correlation coefficient.

Firstly, some definitions will be given:

- Genotype environment interaction (GEI): The usual definitions of genotype environmental interaction implies that interaction exists if differences between genotypes are not consistent from one environment to another (Baker, 1988). Unay, *et, al.* (2004) gave another definition, they stated: "GEI results from changes in relative rankings or magnitudes of differences among entries over environments. They also comment that these variations cause some difficulties in making the subsequent selection of genotypes by plant breeders and growers.
- Crossover: The crossover interaction diagram can be used to visualize when two genotypes change in rank order of performance when evaluated in different environments" (Baker, 1988) (Figure 4.1).
- Spearman's correlation coefficient is a non-parametric measure of how rank orders are related.

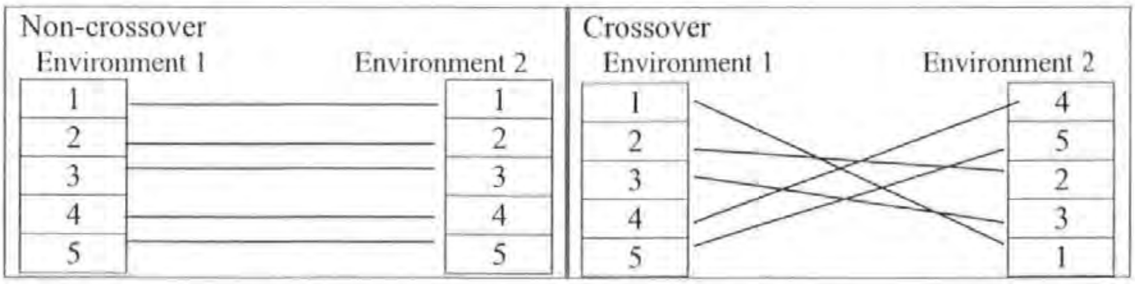


Figure 4.1. Example of significant crossover interaction of 5 genotypes ranked in order of performance in 2 environments.

Results

In order to compensate for differences in overall performance at a site leafiness interactions between different locations were investigated by Crossover after calculating the difference between the actual leafiness (x) and the mean leafiness ($\bar{x}$) for the site/year Δx ($\Delta x = x - \bar{x}$) of change by 0.5 of an integer; and placing these into 4 categories (Δx < *-0.5; -0.5-0; 0-+0.5; >+0.5*) (Figure 4.2).

The crossover interaction between Tel Hadya and Breda for H1 (Figure 4.2) showed a high degree of crossover. Of the total of 102 families in H1 there are only 25 families in the same category at each site (4 in the first group, 9 in the second group, 12 in the third group), while 77 families have crossover between categories. This indicates that the leafiness for H1 is strongly affected by the environment .

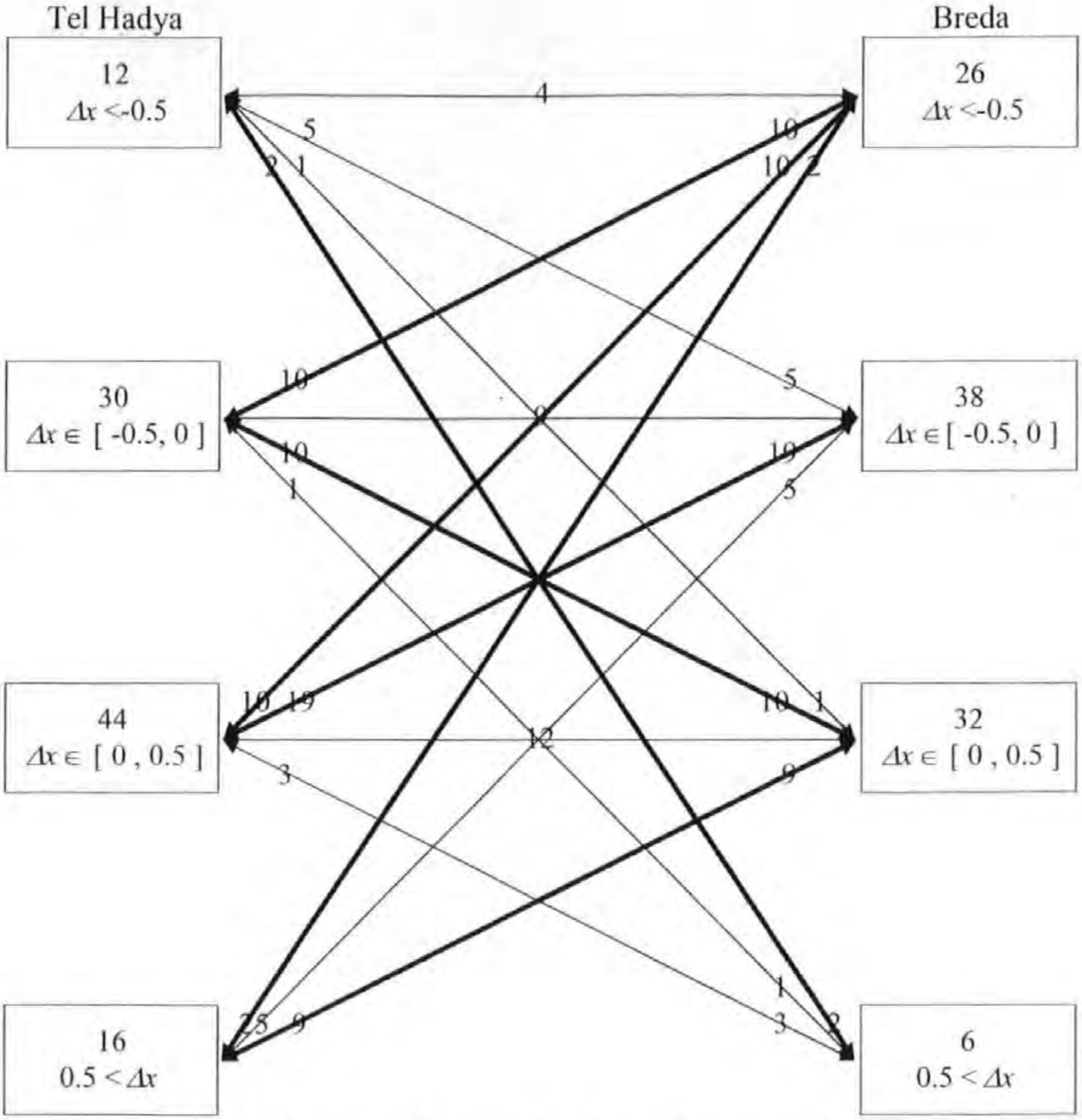


Figure 4.2. Leafiness crossover diagram for Hybrid one between Tel Hadya and Breda. (Numbers in boxes show the number of individual families in each category and numbers on crossover arrows indicate the number of families changing to another category at the second location the "weight of the line graphically indicates the numbers of families crossing over).

This result is supported by lack of significance in the Spearman correlation coefficient between the 2 sites for H1, ($r_s = 0.15, P=0131$).

However, since a non-significant result can never prove a null hypothesis, this statistical result cannot be used as objective evidence of GEI.

Similarly, H3 shows a high degree of crossover (Figure 4.3) with only 29 families in the same categories between sites (4 first group, 13 second group and 12 third group) and there were no families in the fourth group in Tel Hadya, while there are 22 families in the same group in Breda. This result is supported the lack of significance of the Spearman correlation coefficient, ($r_s = 0.174, P=0.08$).

In contrast, for H2 there are no large differences between the numbers of families in each category at the two sites and the number of crossovers is less (Figure 4.4) and the Spearman correlation is significant ($r_s = 0.196, P=0.048$) indicating a lower degree of interaction in this hybrid.

In conclusion, it appears that leafiness is not a stable characteristic for H1 and H3 with some genotypes performing well in Tel Hadya, but not in Breda. For example, family numbers (11,27) were in the first group (best performing) in Tel Hadya whilst they were in the fourth group (worst performing) in Breda. Other genotypes perform well in Breda, but not in Tel Hadya, for example: family numbers (36,89).

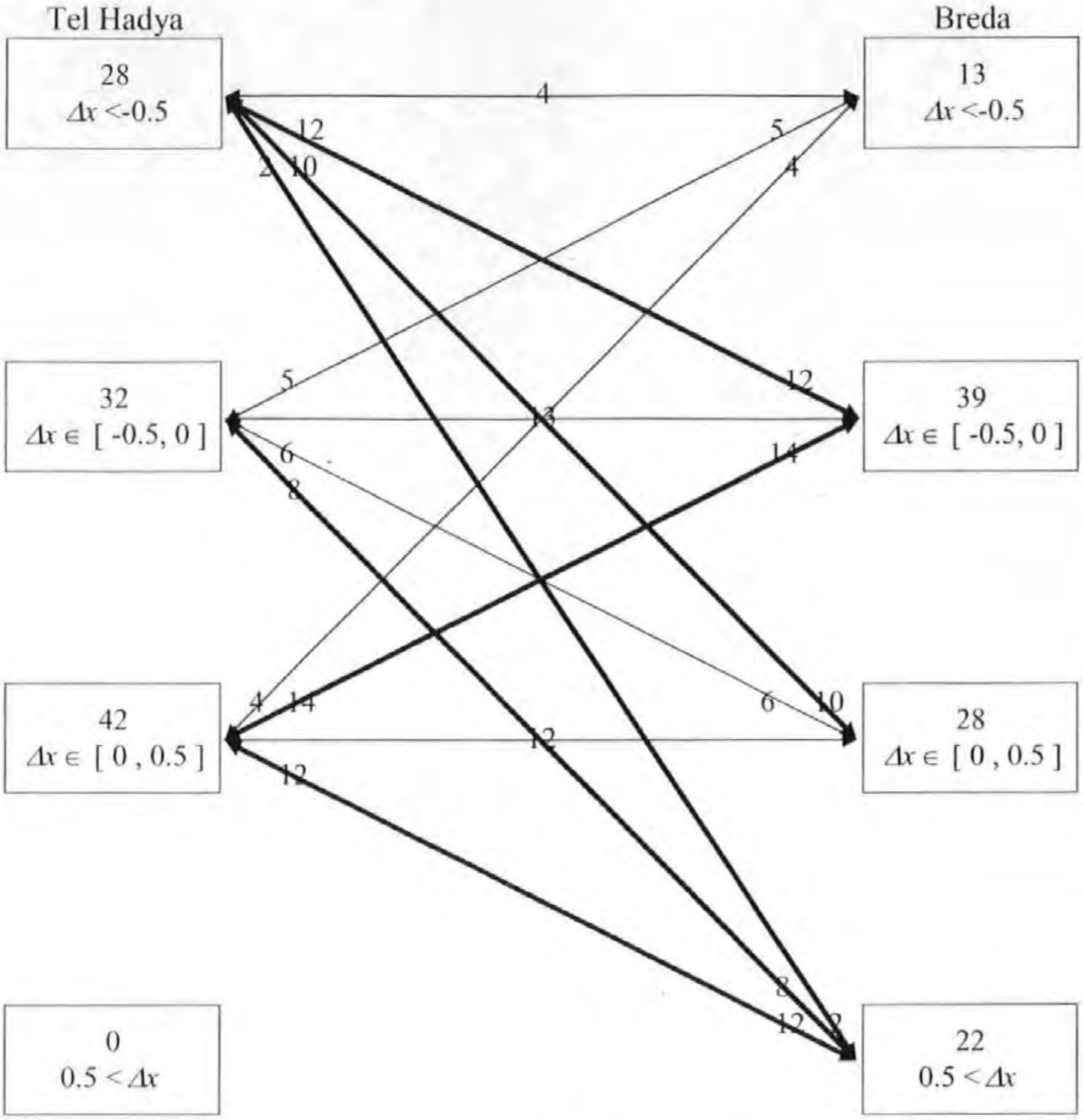


Figure 4.3. Leafiness crossover diagram for H3 between Tel Hadya and Breda.

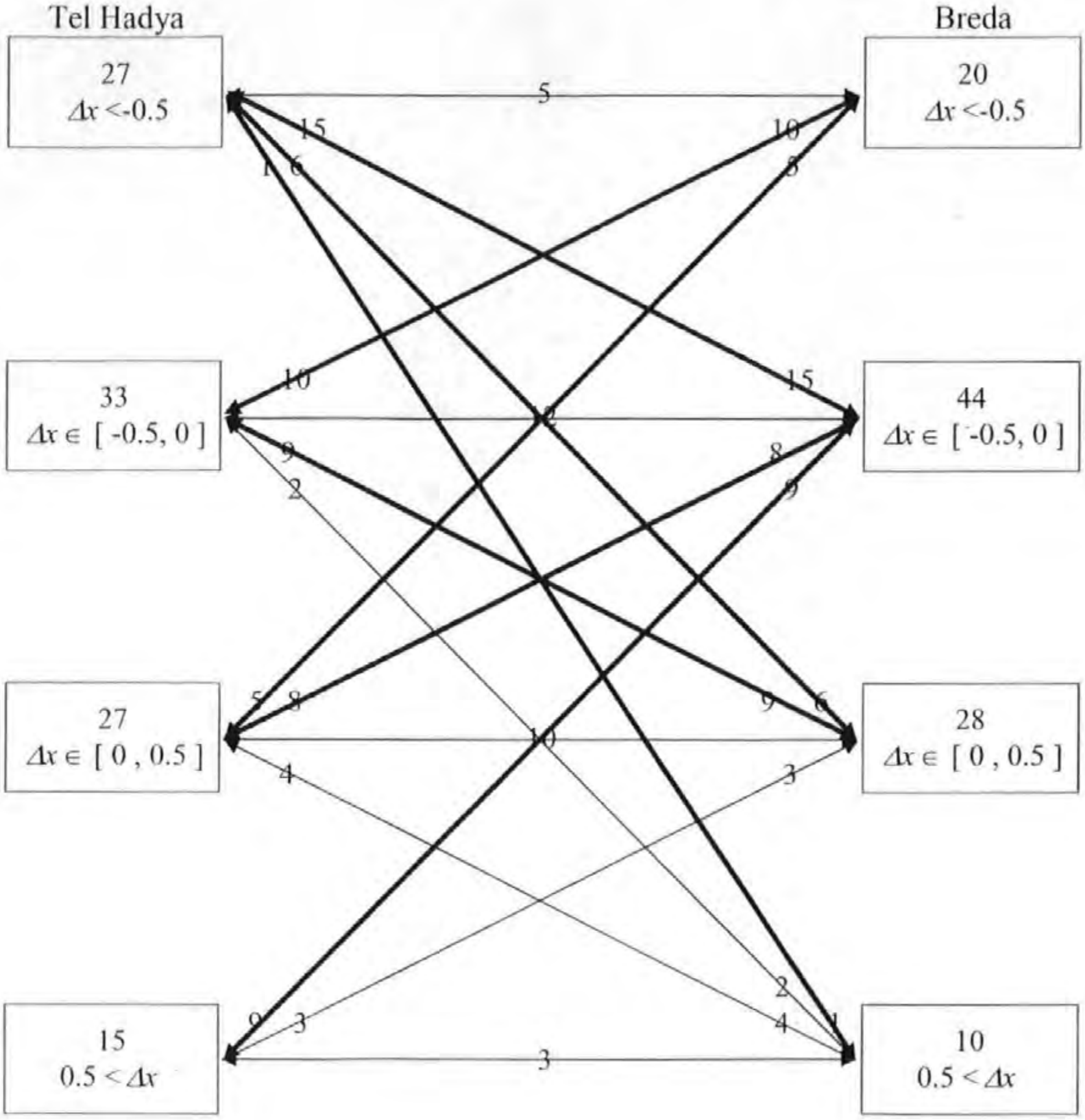


Figure 4.4. Leafiness crossover diagram for H2 between Tel Hadya and Breda.

There are several possible reasons for the high degree of interaction in leafiness scores (including high phenotypic plasticity), but since the characteristic is a score it is also possible that it has been inconsistently applied between sites.

Despite the levels of crossover discussed above, mean yield (grain and straw) appeared to have been affected by mean levels of leafiness since it was observed that for an individual hybrid, area and season gave different yields in accordance with different levels of leafiness (Figure 4.5 and Tables a.1.a to a.1.f in Appendix). For example, in Tel Hadya, second season, hybrid one, there were significant differences between the yields based on levels of leafiness with levels 4-4.5 giving the highest yields in two areas (Tel Hadya, Breda). There are however some inconsistencies, e.g. sometimes level (5) was no different from the other levels.

Figures 4.5.a, 4.5.b and 4.5.c show that the grain yield in Tel Hadya and Breda increased steadily according to leafiness, but there was a sudden decrease in the grain yield in Tel Hadya, H1, between the levels 4.5-5.

Straw yield in Breda and Tel Hadya rose gradually according to leafiness (Figures 4.5.d, 4.5.e, and 4.5.f).

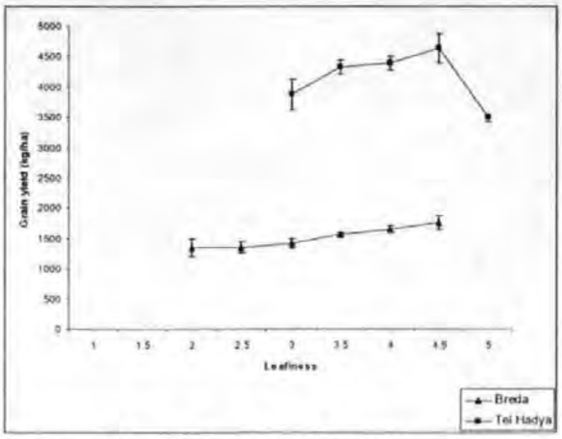


Figure 4.5.a. Grain yield (H1) .

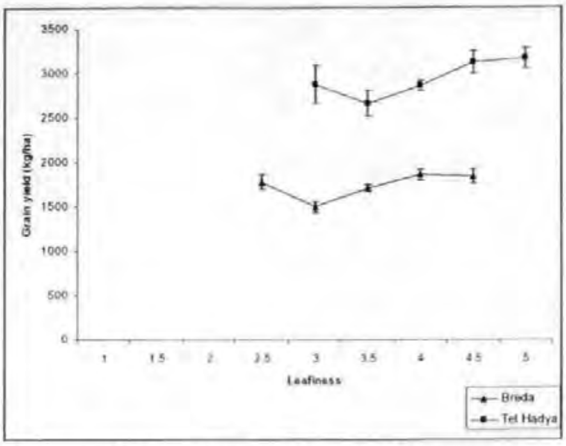


Figure 4.5.b. Grain yield (H2).

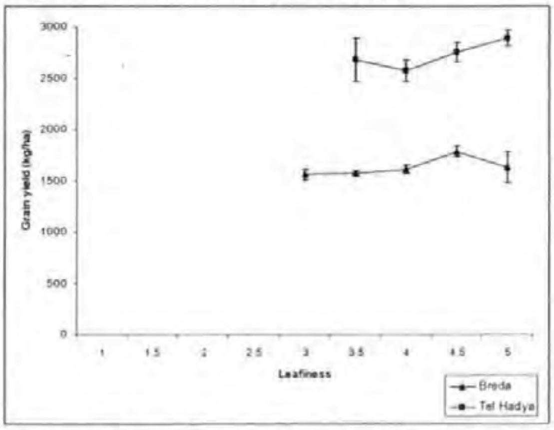


Figure 4.5.c. Grain yield (H3).

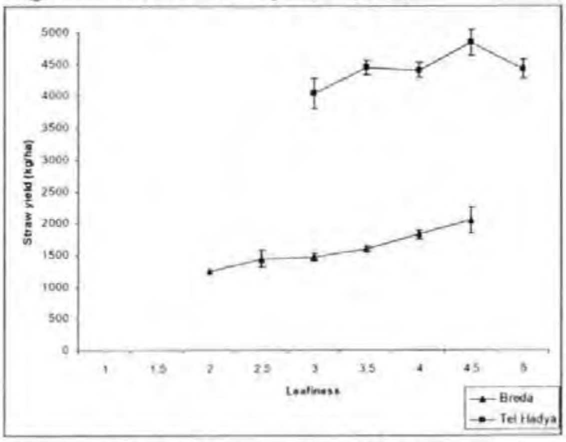


Figure 4.5.d. Straw yield (H1).

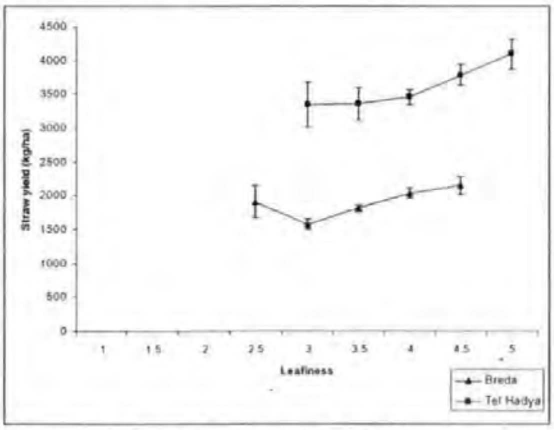


Figure 4.5.e. Straw yield (H2).

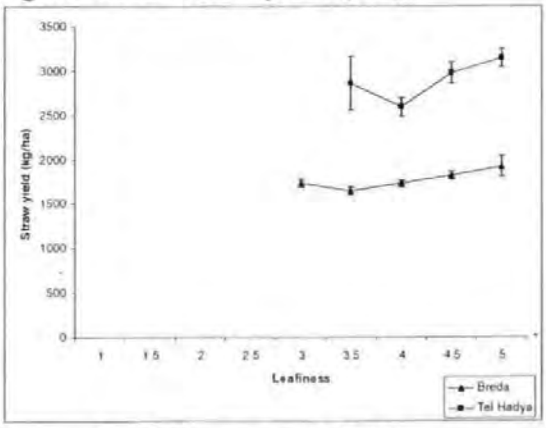


Figure 4.5.f. Straw yield (H3).

Figure 4.5. Grain and straw yield (mean and standard error) according to leafiness in Tel Hadya and Breda for hybrids 1, 2 and 3.

Statistical description:

Leafiness had a significant influence on yields across seasons at Tel Hadya and Breda. This was shown by: firstly, a significant difference (*P<0.05*) between the mean total plant yield depending on levels of leafiness in two areas (Tel Hadya, Breda) with all three hybrids; and secondly, grain yield in Breda was significantly affected by leafiness during the second season. As shown in Tables a.1.a to a.1.f, and according to levels of leafiness, there were significant differences between grain yields (H1 (*P=0.011*), H2 (*P=0.001*), and H3 (*P=0.007*). However, at Tel Hadya, leafiness had no real effect on grain yield except for H2 (*P<0.05*), since the mean grain yields for given levels of leafiness in H2 were significantly different during the second season. Mean straw yields were affected by levels of leafiness during the second season except for H1 in Tel Hadya where there were no significant differences between means straw yield according to leafiness scores. (*P=0.098*). A significant difference between mean straw yields is illustrated in Tables a.1.a to a.1.f (Tel Hadya: H2 (*P=0.038*), H3 (*P=0.012*); Breda: H1 (*P<0.001*), H2 (*P<0.001*), H3 (*P=0.036*)).

Using the Least Significant Difference (LSD) test it can be seen that for:

⇒ H1: the mean yields at leafiness levels 4.5 and 4 were the best in Tel Hadya and Breda. The rank order of leafiness levels according to total plant yield were: in Tel Hadya (4.5, 4, 3.5, 5, 3) and in Breda (4.5, 4, 3.5, 3, 2.5, 2).

⇒ H2: there are no large differences between H1 and H2. However, it can be seen that, at Tel Hadya, the best yield parameters (total plant yield, grain yield, and straw yield) were with leafiness level 5 and the rank order of leafiness levels was: in Tel Hadya (5,4.5,4,3,3.5) and in Breda (4.5,4,2.5,3.5,3).

⇒ H3: at Tel Hadya, leafiness level 4 was the lowest for production whilst level 5 was highest. The order at Tel Hadya was (5,4.5,3.5,4); while at Breda it was (4.5,4,3.5,3,5).

In summary, leafiness level 5 was the best in H2 and H3 in Tel Hadya, while it was the worst for H1 at Tel Hadya and H3 at Breda. This suggests that leafiness was not stable for hybrids 1 and 3.

In conclusion, yield parameters can be strongly affected by leafiness but leafiness can also be affected by location and H1 and H3 appeared not to be stable. As a consequence, the interaction between genotype and environment and also the relationship between yield parameters and morphological characters must be investigated more thoroughly.

4.3 The influence of weather on yield and morphological characters

Weather is one of the most important factors influencing barley yields. Table 4.1 indicated that there were significant differences (*P<0.05*) between the mean yield parameters during the two seasons (season 1, season 2) in Tel Hadya . There could be several reasons for these differences. For example:

1. Date of cultivation and germination: the date of cultivation for the first season was two weeks earlier than for the second season, however, the date of germination for the two seasons was the same (25th Dec). Thus the seeds took about one month to germinate in the first season but only 2 weeks in the second. Many factors affect germination but the most important of these are temperature and moisture availability (Briggs, 1978). During the period of germination in the second season average rainfall was 18 mm and the temperatures ranged from 3 ^{0}C to 10.5 ^{0}C, while in the first season, the rainfall was 50 mm and the temperatures ranged from 2 ^{0}C to 8.3 ^{0}C. The low temperatures in the second season explain the slow establishment rate and possibly increased plant losses during establishment, lower plant populations can reduce yield (Briggs, 1978).

Table 4.1. Differences between mean data in different seasons (season one –season two).

	Total yield (kg ha^{-1})	Grain yield (kg ha^{-1})	Straw yield (kg ha^{-1})	Harvest index	No.of ears m^{-2}	TGW (gram)	Protein %	vegetative duration (day)	Plant height (cm)	Length of growing season (day)	Leafiness
(a) Tel Hadya, H1, different seasons (first season $\bar{X}_1$, second season $\bar{X}_2$)											
$\bar{X}_1$	7857	3642	4215	0.46	229.5	45.8	9.63	81.54	58.45	-	-
$\bar{X}_2$	8760	4325	4435	0.49	265.8	45.2	8.53	100.23	64.32	134.2	3.82
T[1]	-5.19	-7.15	-2.31	-6.79	-8.40	1.87	16.27	-100.32	-9.66	-	-
P	<0.05	<0.05	<0.05	<0.05	<0.05	>0.05	<0.05	<0.05	<0.05	-	-
(b) Tel Hadya, H2 different seasons (first season and second season $\bar{X}_2$)											
$\bar{X}_1$	6726	2813	3913	0.42	254.2	49.5	10.24	80.97	61.39	-	-
$\bar{X}_2$	6551	2942	3609	0.45	265.7	46.0	9.29	99.98	67.05	134.2	4.1
T	0.99	-1.54	2.66	-5.05	-3.59	11.67	8.04	-55.89	-9.60	-	-
P	>0.05	>0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	<0.05	-	-
(c) Tel Hadya, H3 different seasons (first season $\bar{X}_1$, second season $\bar{X}_2$)											
$\bar{X}_1$	6162	2564	3598	0.41	249.3	46.2	10.84	84.38	58.85	-	-
$\bar{X}_2$	5717	2765	2953	0.49	265.1	45.5	9.41	101.24	60.86	132.89	4.54
T	2.45	-2.34	6.10	-13.59	-4.20	1.45	12.00	-79.25	3.07	-	-
P	<0.05	<0.05	<0.05	<0.05	<0.05	>0.05	<0.05	<0.05	<0.05	-	-

[1] See Chapter two (Hypothesis test, P 56)

2. Rainfall during the growth period: soil moisture availability can severely affect the yield of cereal crops. Irrigation of cereals is rarely undertaken in Tel Hadya and Breda and was not applied in any of the trials supplying data to the current investigation. Both the total rainfall during the growing season (December to July) and that falling during the grainfilling period (April to July) can influence yield. For Tel Hadya the total rainfall during the first season (290 mm) was marginally higher than the second season (273.3 mm) but most of this rainfall (240 mm) fell during the vegetative period (Figure 4.3) with little falling during grainfilling. There was however, significant rainfall in May during the second season and this could have the effect of improving yield parameters during this season. There was no significant difference between thousand grain weight (TGW) in the first season and second season.

   The means of rainfall in January, February and April in the second season were greater than in the first season (Figure 4.6).

3. Temperature: Mean maximum and minimum temperatures in the second season were consistently above those in first season (Figure 4.7) resulting in approximately 50 °Cd more accumulation of thermal time per month in the second season. Warmer temperatures accelerate crop development meaning that in the second season the crop would accelerate through its growth stages faster and mature earlier. Given that barley requires approximately 1200-1500 °Cd from sowing to maturity with 700 – 800°Cd from sowing to anthesis (Briggs, 1978), it can be estimated that

the second season crop matured approximately 2 weeks earlier than the first season crop and this is supported by the time to flowering data recorded in the experiments. Accelerated development and early maturity normally reduces yield potential in cereals since it limits total crop photosynthesis but under Tel Hadya and Breda conditions it can sometimes be an advantage because grainfilling is brought forward and completed before the full intensity of the summer drought takes effect. This was further accentuated by the high April/May temperatures being combined with low rainfall at the critical early grainfill stage of development . Figure 4.7 also shows that during early vegetative growth in January and February, low temperatures were experienced and in the first season mean monthly temperatures were around 0 ^{0}C indicating that incidences of frost would have occurred. Since these barley crosses were Spring types with low frost tolerance, these frosts may have caused some kill or damage.

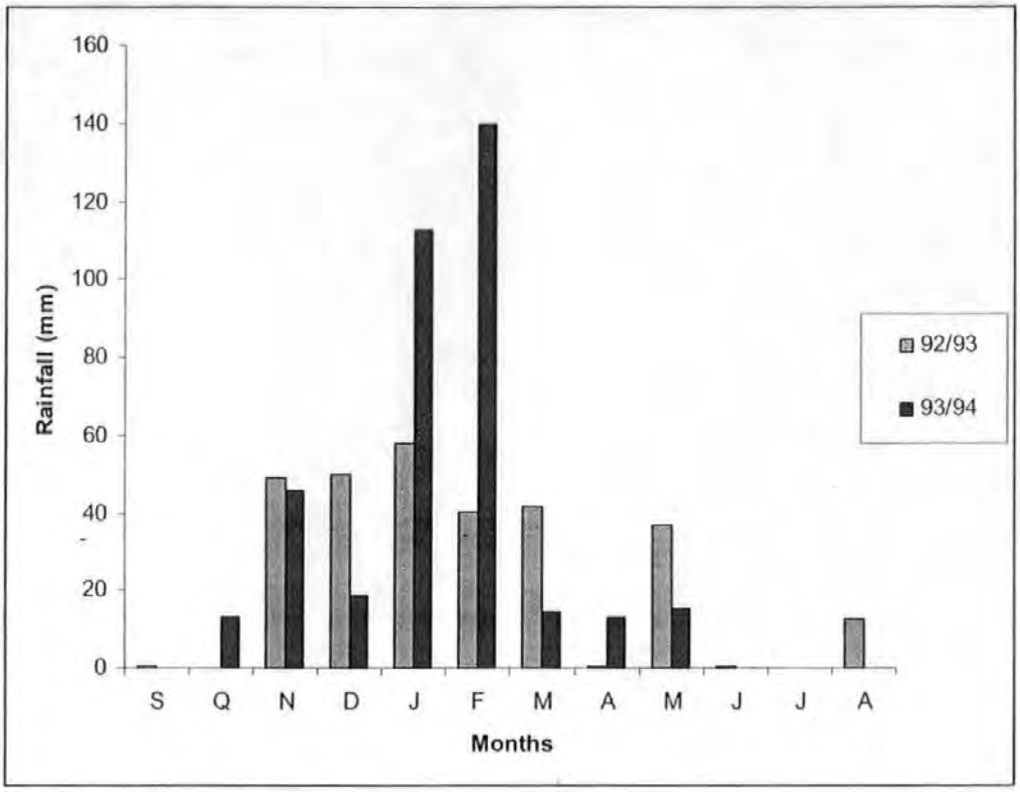


Figure 4.6. Means of rainfall during two season in Tel Hadya.

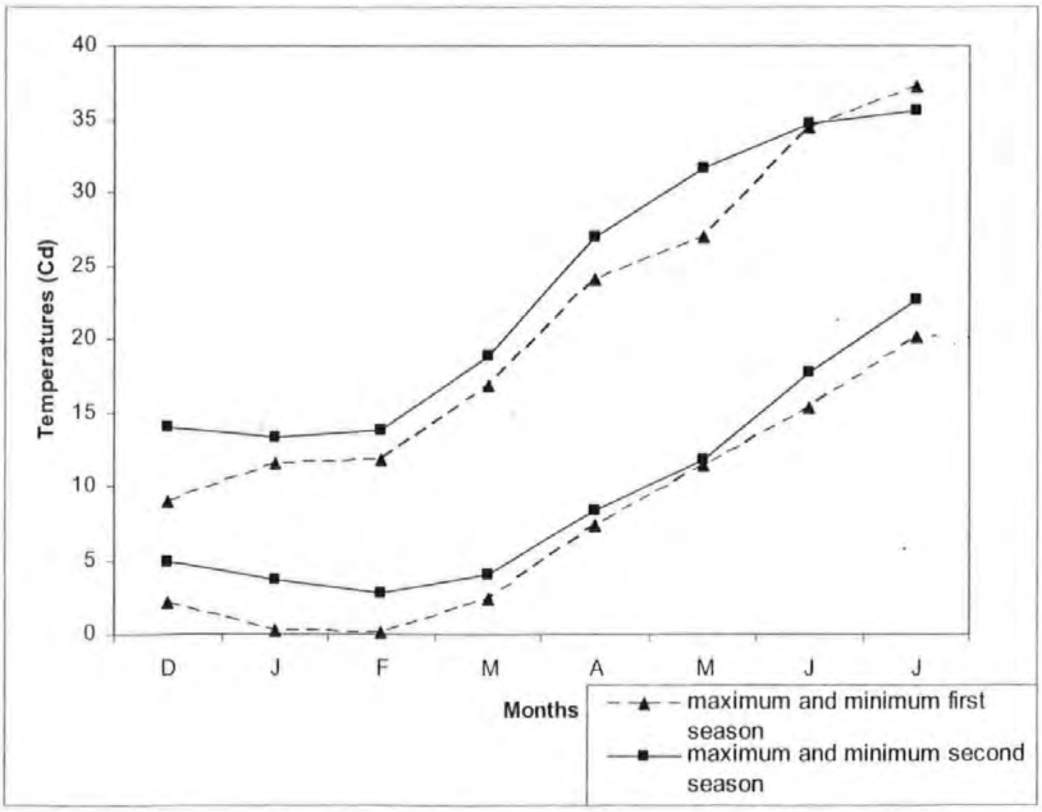


Figure 4.7. Mean monthly maximum and minimum temperatures (° Cd) during the growing two seasons in Tel Hadya.

4. Nitrogen nutrition. Nitrogen improves photosynthetic capacity by increasing levels of the CO_2 harvesting protein (Rubisco) and increased chlorophyll-binding protein leading to higher chlorophyll levels (greener leaves) (Soffe, 2003). This in turn leads to greater production of carbohydrate which leads to increased leaf size and thereby increased Leaf Area Index (LAI). Increased LAI in cereals in the absence of severe water restriction improves yield capacity (Dennis, *et al.*, 1980) but under water stress this effect is less important. Individual plant development (and thus grain yield) can be affected by the amount and timing of nitrogen application and this also affects the quality of grain harvest. Heavy nitrogen applications can also lead to taller weaker stems of crops and result in lodging.

Soils in the growing regions of Tel Hadya and Breda tend to be of relatively low fertility and calcareous in nature (Naesah, 1996), and therefore require the addition of nitrogen to grow arable crops. In the second season, 60 kg ha^{-1} of nitrogen was used, while in the first season only 40 kg ha^{-1} was used.

Whilst these interpretations of the effects of the agroclimate are based on very limited weather and soil data and are therefore speculative, they do indicate that the environments were different in the two seasons and could be responsible for the significant GE interactions reported herein.

Statistically the weather had a major impact on the expression of genotypes. Table 4.1 shows significant differences between the mean yield parameters (total plant yield, grain yield, straw yield, number of ears). In the Tel Hadya area all the hybrids (H1, H2, or H3) gave different mean yield parameters in the two seasons. With the exception of H2, there were no significant differences between the mean total plant yields or the mean grain yields $P=0.323$; $P=0.126$, respectively. These results are shown in Figure 4.8.a and Figure 4.8.b, respectively. H1 gave the best yield in the two seasons and H2 was consistently better than H3. Total plant yield and straw yield for H1 (Tel Hadya) in the second season were greater than in the first season, while they were greater in the first season for H2 and H3 (Figure 4.8.a, Figure 4.8.b, and Figure 4.8.c) indicating interaction of genotypes and years.

An unpaired t-test was performed, showing that the number of ears in the second season was significantly greater than in the first season for all three hybrids ($t=-8.40$; $P<0.001$; $t=-3.59$; $P=0.001$; $t=-4.20$; $P<0.001$ for hybrids 1, 2 and 3 respectively) as shown in Figure 4.8.d. In the first season there were significant differences between the numbers of ears; however, in the second season these differences diminished so that the number of ears for different hybrids were almost the same. There were no significant differences between mean thousand grain weight TGW for H1 and H3 ($P>0.05$), while TGW in H2 first season was significantly higher than in the second season. There were no significant differences between TGW for H1 and H3 ($P>0.05$) (Figure 4.8.f). The mean grain yields in the second season were significantly greater than in the first season, while in H2 there were significant variations in TGW between the two season. However, the grain yields were not significantly different.

The environment also had an effect on the morphological characters. There was a significant difference between the mean morphological characters (plant height and vegetative duration) in the two seasons (season 1 and season 2; Figure 4.8.g and Figure 4.8.h) . Also plant height and vegetative duration in the second season was greater than those in the first season.

Interaction diagram Figure 4.8 illustrates that there was a difference in response between hybrids i.e it can be seen that the lines are not parallel, so there is an interaction between the environment and genotype.

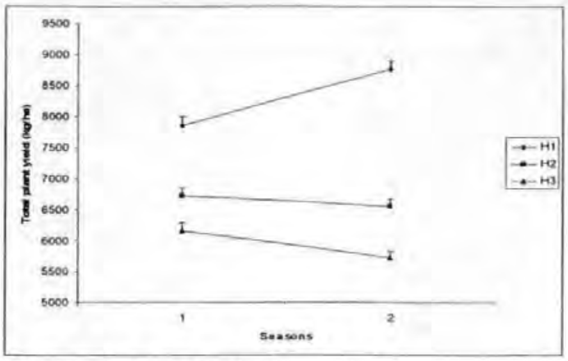


Figure 4.8.a.Total plant yield.

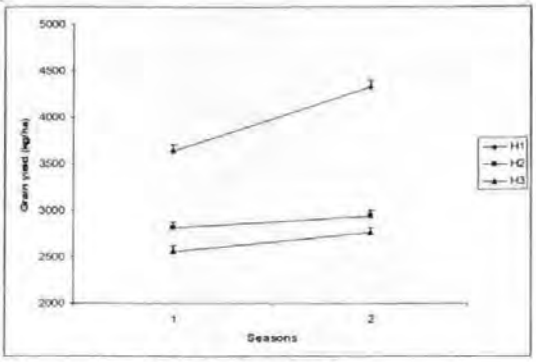


Figure 4.8.b. Grain yield.

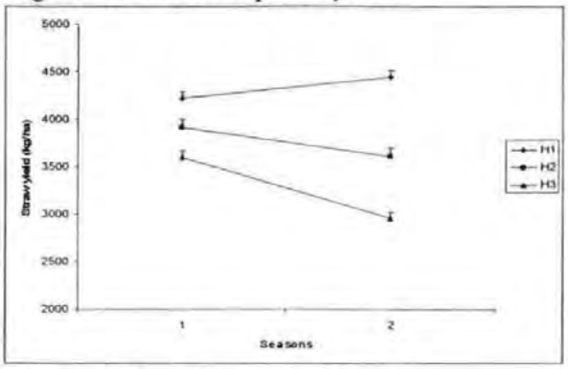


Figure 4.8.c. Straw yield .

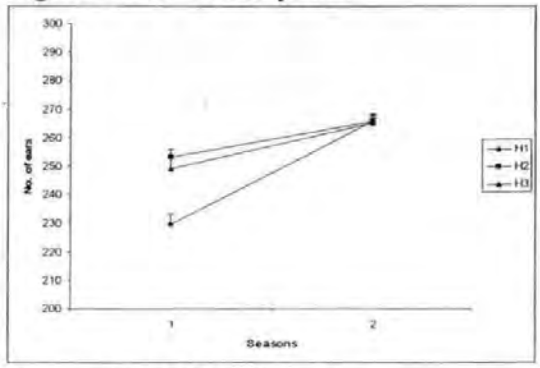


Figure 4.8.d. Number of ears.

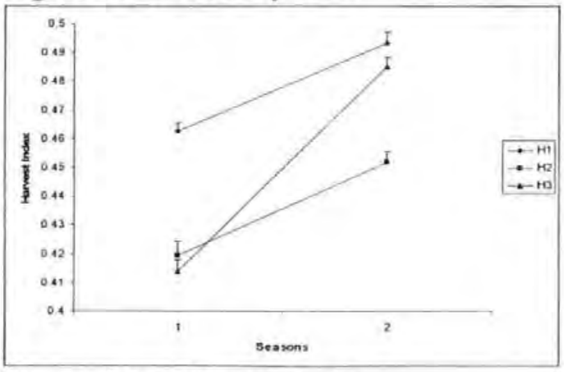


Figure 4.8.e. Harvest index.

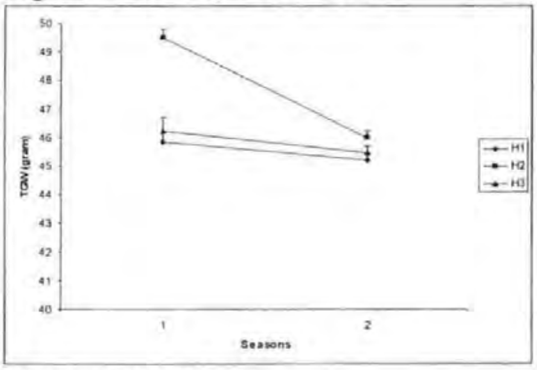


Figure 4.8.f. TGW.

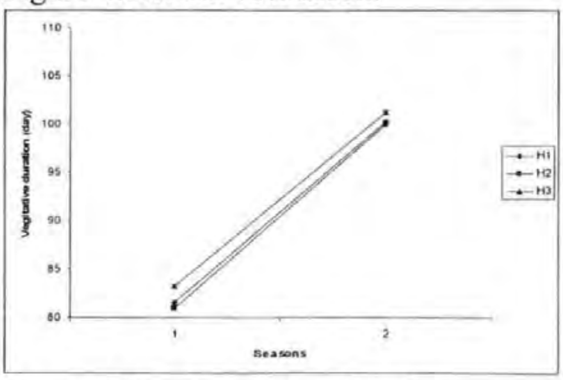


Figure 4.8.g. Vegetative duration.

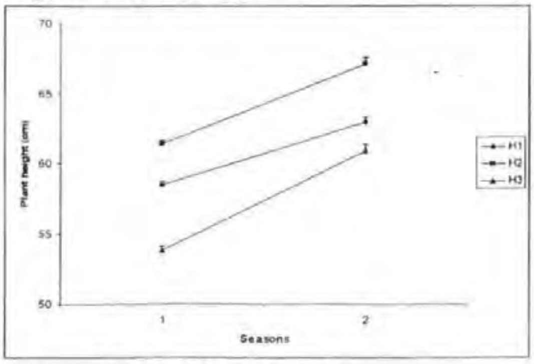


Figure 4.8.h. Plant height.

Figure 4.8. Interaction diagrams of the influence of weather on yield and morphological characters in Tel Hadya.

4.4 Genotype–Environment Interaction

Genotype-Environment Interactions (GEI) are extremely important in the development and evaluation of plant varieties because they reduce the genotypic stability values under diverse environments (Akqura, *et al.*, 2004).

4.4.1 Genotype-Environment Interaction (GEI) in Tel Hadya during different seasons (genotype × year)

genotypes in Tel Hadya was affected by the environment (weather). For example, the crossover interaction for total plant yield is high (Figure 4.A); therefore the total plant yield for H1 is affected by the weather. This result is also supported by Spearman correlation coefficient, ($r_S = 0.18$, $P=0.07$). However, since a non-significant result can never prove a null hypothesis, this statistical result cannot be used as objective evidence of GEI. For the same reason, the total plant yields for H2 is affected by environment since there is no significant Spearman correlation coefficients (H2: $r_S=0.120, P=0.231$).

There are no significant Spearman correlation coefficients between the grain yields for the three hybrids (Tel Hadya, first season and second season) since $P>0.05$. Spearman correlation coefficients were: H1: $r_S=0.17, P=0.084$, H2: $r_S=0.097, P=0.33$, and H3: $r_S=-0.170, P=0.08$.

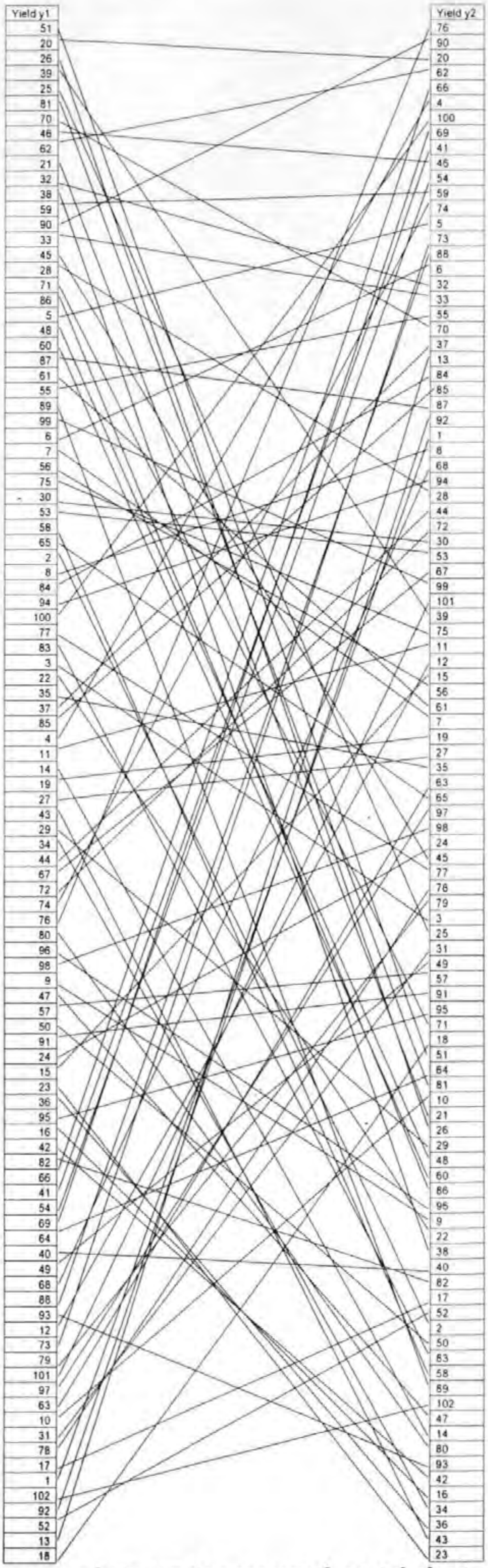


Figure 4.10. Crossover between two seasons for total plant yield (H1; Tel Hadya).

The straw yields in Tel Hadya for the three hybrids were also affected by environment (weather) because Spearman correlation coefficients are not significant for H1 and H2. However, for H3 it is significant, although it is less than *0.25* (H1: r_s*=0.106, P=0.289*, H2: r_s*=0.059, P=0.556*, and H3: r_s=-*0.224, P=0.024*).

H3 has a negative Spearman correlation coefficient and it is significant for total plant yield and straw yield ($P<0.05$), so the ranking order is reversed. For example, in the total plant yield, the fifth in rank in the first season became the ninetieth in the second season, and the seventeenth in the first season became ninety-third in rank in the second season.

The morphological characters are also affected by environment because there were no significant Spearman correlation coefficients for vegetative durations (H1: r_s*=-0.06, P=0.582*, H2: r_s*=-0.070 P=0.484*, and H3: r_s*=-0.087, P=0.38*), and also none between plant heights. Negative Spearman correlations can also be observed in relation to morphological characters. Therefore, it can be seen that weather has a significant effect on morphological characters.

In conclusion, the strong GxE interactions demonstrated the major effect that weather can have on crop yield in Tel Hadya. These findings suggest that factors can influence plant yield independently or concomitantly in the determination of either plant height or leafiness. In the physiology of barley, this points clearly to significant environmental influences during growth and grain filling.

4.4.2 Genotype-Environment Interaction (GEI) in second season in different areas

There were locational genotype environment interactions in the second season between Tel Hadya and Breda for yield. There are no significant Spearman correlations between yield parameters in different areas (Tel Hadya and Breda) except for H2. Total plant yields in H1 and H2 have been affected by location since there are different ranks between them. This result is also shown by the Spearman correlations. While there are no significant Spearman correlations between total plant yields for H1 and H3 (H1, $r_S=0.068$, $P=0.5$; H2, $r_S=0.223$, $P=0.024<0.05$; H3, $r_S=0.119$, $P=0.24$), genotype for total plant yield has been affected by location. The same result (including H2) also appears for grain yields (H1, $r_S=0.021$, $P=0.84$; H2, $r_S=0.021$, $P=0.84$; H3, $r_S=0.063$, $P=0.53$) and Straw yields (H1, $r_S=0.021$, $P=0.84$; H2, $r_S=0.021$, $P=0.84$; H3, $r_S=0.063$, $P=0.53$).

TGW however was not affected by location. Figure 4.10 shows that there was no significant effect for H1, since *53* families from Tel Hadya had the same rank in Breda. Figure 4.10 shows that each group has almost the same number of families in both Tel Hadya and Breda. For example, group ($\Delta r<-0.5$) has 43 families in Tel Hadya and 42 families in Breda. Group ($\Delta r \in [-0.5, 0]$) has 6 families in Tel Hadya while there are 7 families in Breda. The result, TGW has not been affected by location for H1, is also supported by Spearman's correlation (*0.41*) which is highly significant (*P<0.0001*).

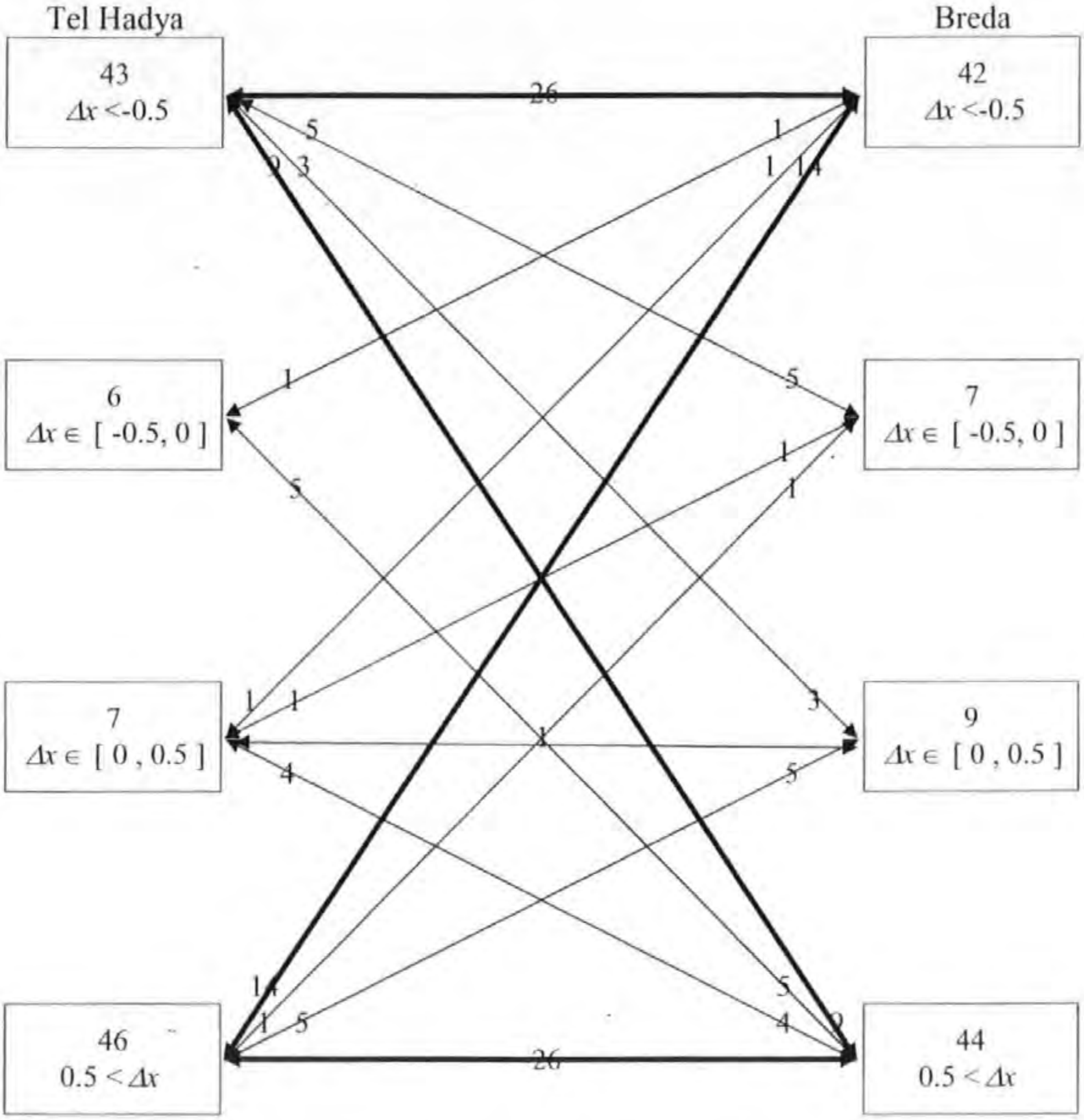


Figure 4.10. Tel Hadya (H1) and Breda (H1) TGW.

Figure 4.11 shows the same result for H2 with a slight difference, there are *46* families with the same rank but the Spearman's correlation (*0.264*) is smaller, but is still significant ($P<0.008$). Also there is no great difference between the number of families in each group.

The TGW for H3 is affected by location as shown by Figure 4.12 and Spearman's correlation is not significant ($P=0.478>0.05$). Figure 4.12 shows that there is a difference between the number of families in each group. In group three ($\Delta r \in [0\ ,\ 0.5\ ]$), Tel Hadya has 2 families while Breda has 15 and group four ($0.5 < \Delta r$), Tel Hadya has 45 families while Breda has 34 only. In conclusion, TGW is stable for H1 and H2, while it is not stable for H3.

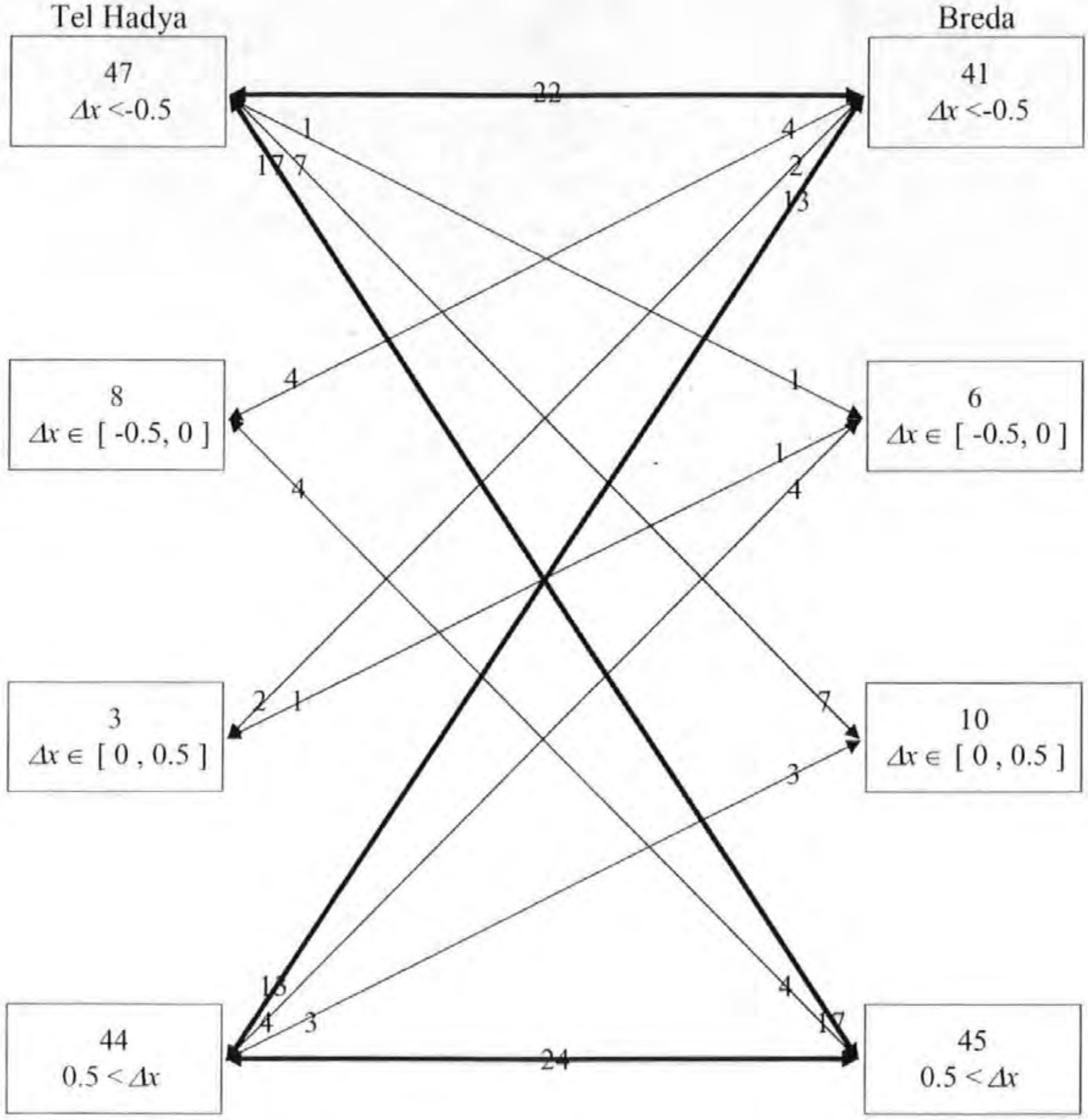


Figure 4.11. Tel Hadya (H2) and Breda (H2) TGW.

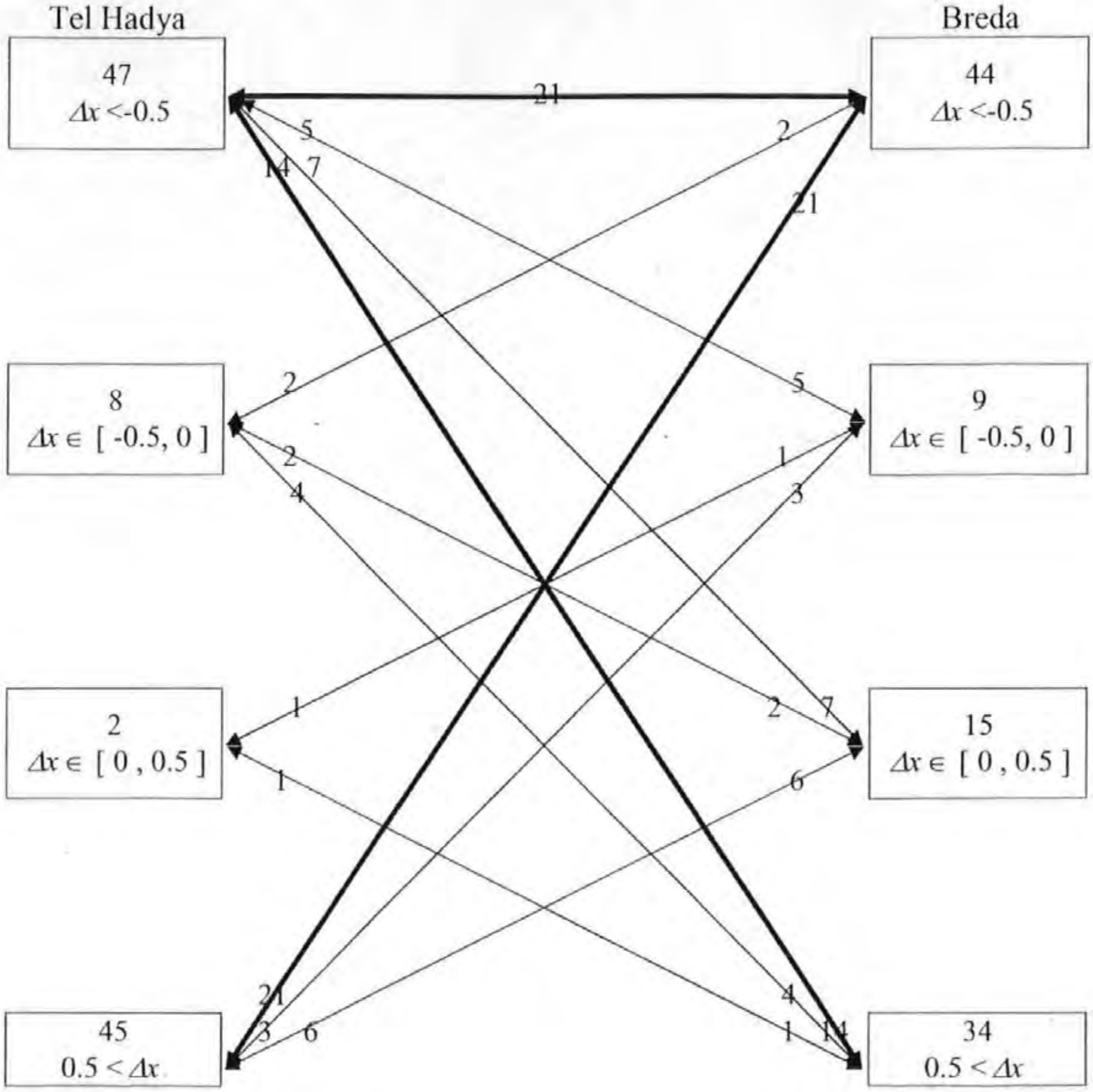


Figure 4.12. Tel Hadya (H3) and Breda (H3) TGW.

The number of ears was affected by location: Figures 4.13, 4.14 and 4.15 Figure 4.13 shows that some families give a high number of ears in Tel Hadya while in Breda they do not. For example, family number (*90*) gives *306* ears per square metre in Tel Hadya while in Breda it only gives *186* ears. On the other hand, the family number *55* has *313* ears per square metre in Breda while in Tel Hadya it only has *246*. Figure 4.14 and 4.15 show that there are no correlations between them. There are no significant Spearman's correlations for all three hybrids ($P>0.05$). In conclusion, the number of ears has been significantly affected by the environment.

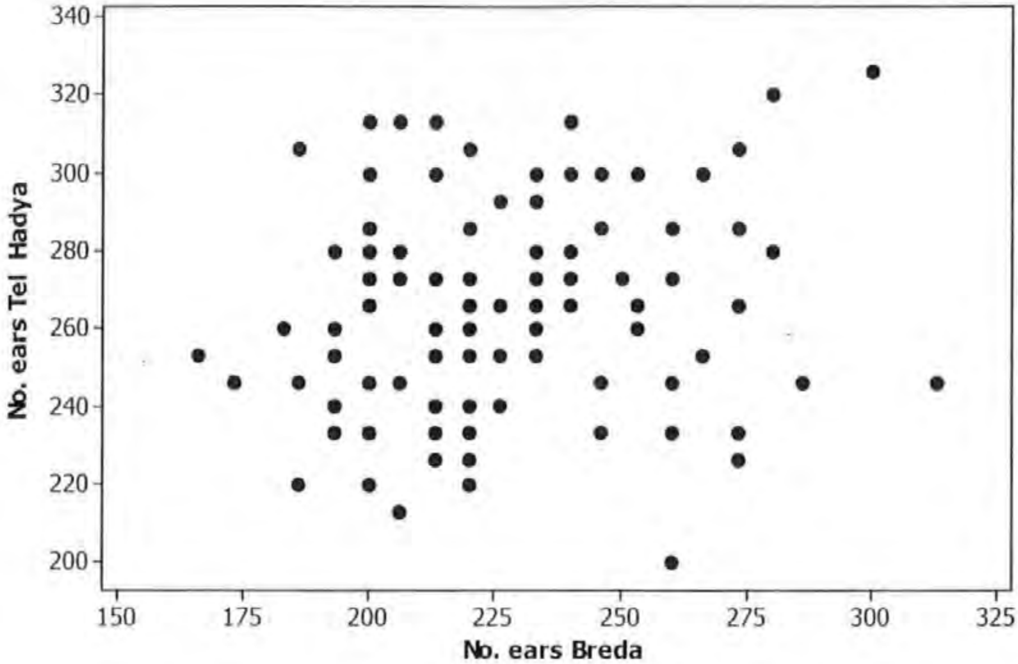


Figure 4.13. Scattergram of No. ears m^{-2} at Tel Hadya vs No. ears m^{-2} at Breda (H1).

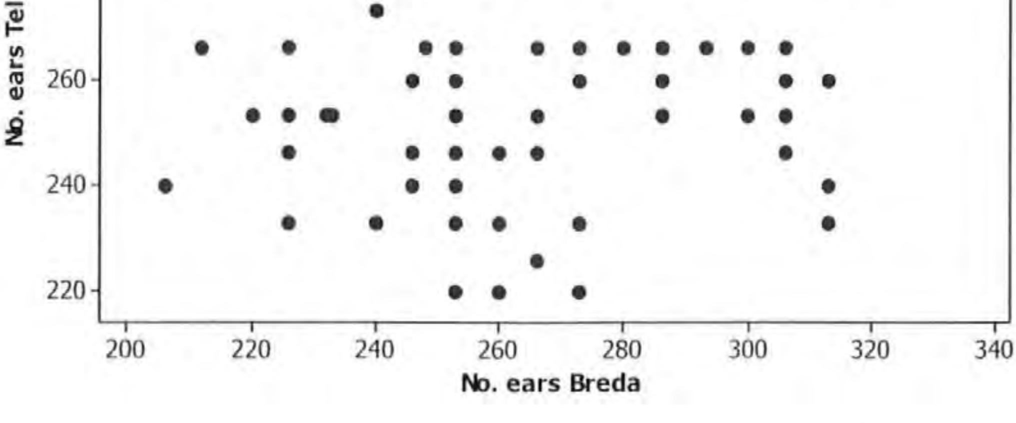


Figure 4.14. Scattergram of No. ears m^{-2} at Tel Hadya vs No. ears m^{-2} at Breda (H2).

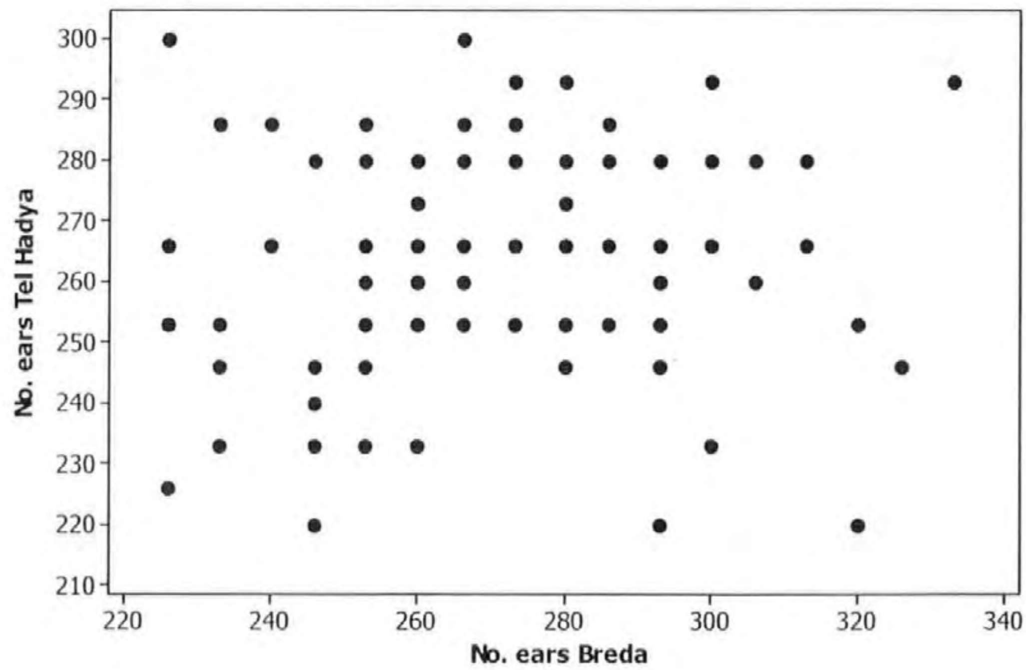


Figure 4.15. Scattergram of No. ears m^{-2} at Tel Hadya vs No. ears m^{-2} at Breda (H3).

Figure 4.16, 4.17 and 4.18 show that vegetative duration has not been affected by environment. According to Fig 4.16, there are *76* families with the same ranks. Also there are 59 families which have the same ranks for H2 (Fig 4.17) and 50 families with the same ranks for H3 (Fig 4.18).

Spearman correlations were significant for vegetative duration (H1: $r_s = 0.2$, $P = 0.048$, H2: $r_s = 0.83$, $P < 0.0001$, and H3: $r_s = 0.82$, $P < 0.0001$). Consequently, it can be concluded that vegetative duration has not been strongly affected by the environment.

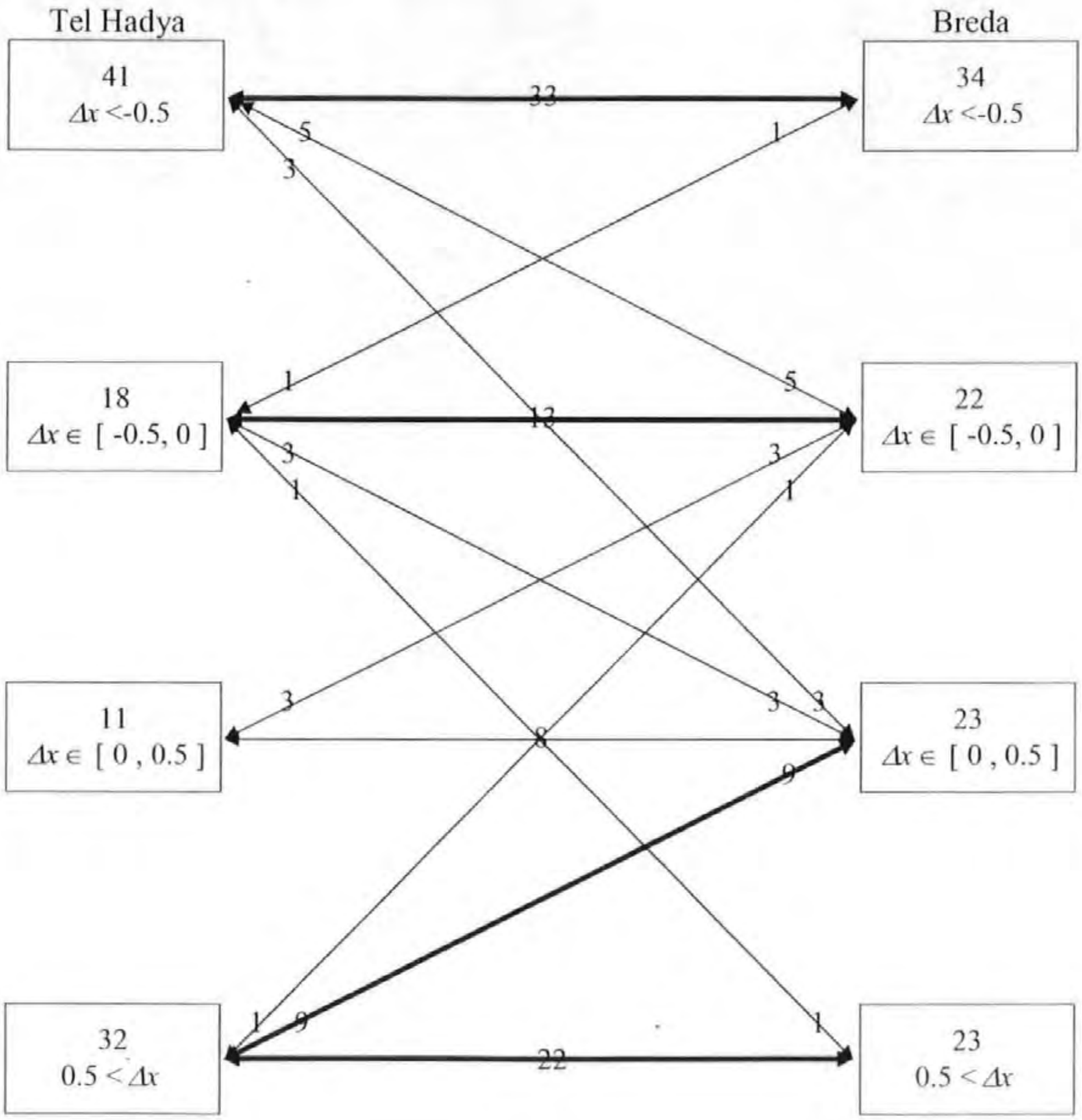


Figure 4.16. Tel Hadya (H1) and Breda (H1) Vegetative duration.

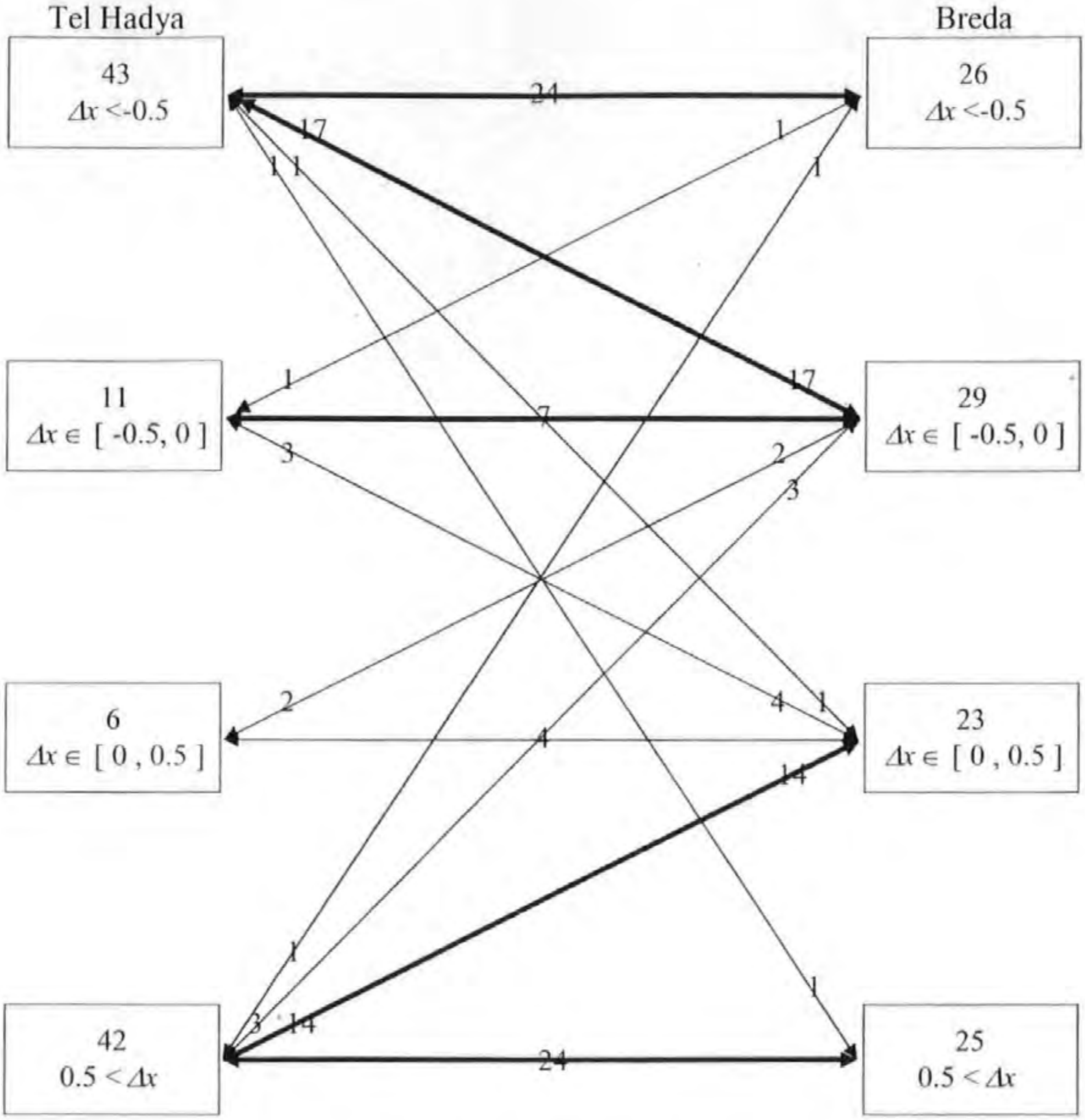


Figure 4.17. Tel Hadya (H2) and Breda (H2) Vegetative duration.

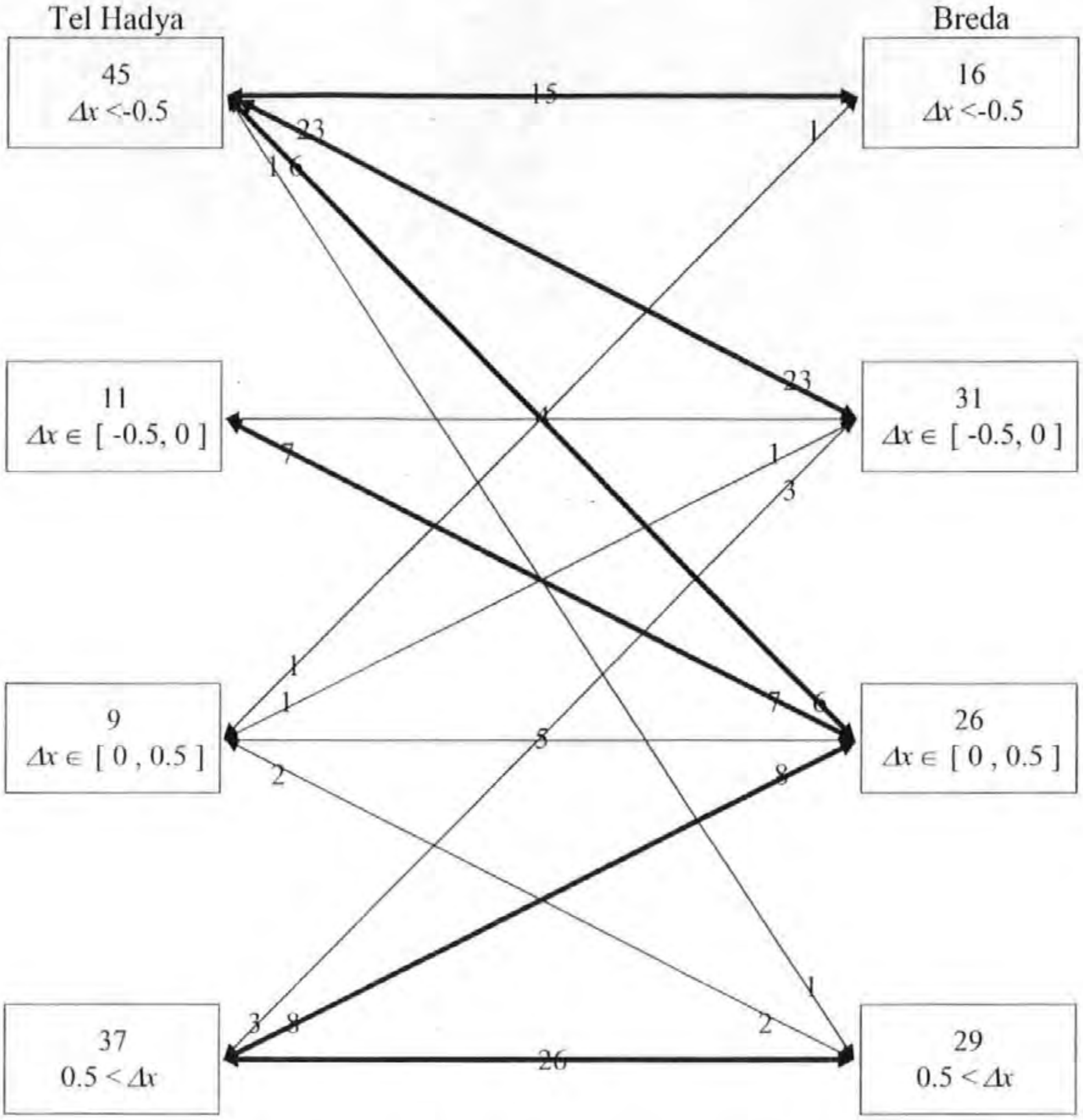


Figure 4.18. Tel Hadya (H3) and Breda (H3) Vegetative duration.

Similarly there are no significant Spearman correlation coefficients between plant heights in the different areas. In conclusion, there are genotype environmental interactions in the second season between Tel Hadya and Breda.

Table 4.2 shows the mean data in the different areas for the same hybrid (H1, H2, and H3) and season.

The environment had a significant impact on the yield parameters in two areas whenever the same hybrids were used in the same season. This conclusion is illustrated in Table 4.2. There is a significant difference between yields ($P<0.05$). Yields in Tel Hadya were much greater than in Breda for several reasons for example, average rainfall during the second season in Tel Hadya (291 mm) was higher than in Breda (273.3 mm). The temperature in Tel Hadya was greater than in Breda. Also the soil in Tel Hadya was different from that in Breda. Consequently, rainfall, temperature and soil in Tel Hadya resulted in greater growth and higher production. The environment was different and the production may be due to one or more factors or the interaction between them.

Table 4.2. Differences between mean data in different areas (Tel Hadya- Breda).

Season two, H1 different areas (Tel Hadya $\bar{x}_1$, Breda $\bar{x}_2$)											
	Total yield (kg ha^{-1})	Grain yield (kg ha^{-1})	Straw yield (kg ha^{-1})	Harvest index	No.of ears m^{-2}	TGW (gram)	Protein %	vegetative duration (day)	Plant height (cm)	Length of growing season (day)	Leafiness
$\bar{x}_1$	8760	4325	4435	0.49	265.8	45.2	8.53	100.23	64.32	134.2	3.82
$\bar{x}_2$	3213	1557	1656	0.49	226.5	40.0	11.44	90.2	45.67	131.1	3.53
T[2]	39.54	35.35	34.06	1.13	11.13	25.1	-53.78	103.52	32.74	9.11	4.11
P	<0.001	<0.001	<0.001	**>0.05**	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Season two, H2 different areas (Tel Hadya $\bar{x}_1$, Breda $\bar{x}_2$)											
$\bar{x}_1$	6551	2942	3609	0.45	265.7	46.0	9.29	99.98	67.05	134.2	4.1
$\bar{x}_2$	3592	1726	1866	0.48	265.3	42.2	12.1	89.78	50.48	129.9	3.63
T	22.81	21.21	20.73	-6.21	0.11	13.4	-39.56	110.25	31.11	17.69	7.14
P	<0.001	<0.001	<0.001	<0.001	**>0.05**	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Season two, H3 different areas (Tel Hadya $\bar{x}_1$, Breda $\bar{x}_2$)											
$\bar{x}_1$	5717	2765	2953	0.49	265.1	45.46	9.41	101.24	60.86	132.89	4.54
$\bar{x}_2$	3343	1622	1721	0.49	271.2	40.6	12.45	90.41	44.29	129.2	3.8
T	20.62	20.23	18.35	-0.09	-2.08	16.12	-34.43	100.71	-37.79	-18.64	12.05
P	<0.001	<0.001	<0.001	**>0.05**	<0.05	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

[2] See Chapter two (Hypothesis Test; P 56)

From Table 4.2, it can be seen that the number of ears in Breda was greater than the number of ears in Tel Hadya. However, there was a significant difference in (TGW) and yields, with those in Tel Hadya being greater than those in Breda.

Environment also had an impact on the morphological characters. There were significant differences between morphological characters in Tel Hadya and morphological characters in Breda (plant height, vegetative duration, length of growing season , leafiness). For example, plant height in Tel Hadya was higher than in Breda. Leafiness in Tel Hadya was (3-5), while in Breda it was (2-4.5).

4.4.3 The influence of genotype on yield and morphological characters

Table 4.3 shows the mean data for different hybrids (H1, H2, and H3) in same area (Tel Hadya, and Breda) and season (first season, second season).

Table 4.3. Differences between mean data for different hybrids (ANOVA).

	Total yield (kg ha^{-1})	Grain yield (kg ha^{-1})	Straw yield (kg ha^{-1})	Harvest index	No.of ears m^{-2}	TGW (gram)	Protein %	vegetative duration (day)	Plant height (cm)	Length of growing season (day)	Leafiness
Season one, Tel Hadya different hybrids (H1 $\bar{X}_1$, H2 $\bar{X}_2$ H3 $\bar{X}_3$)											
$\bar{X}_1$	7857	3642	4215	0.46	229.5	45.8	9.63	81.54	58.45	-	-
$\bar{X}_2$	6726	2813	3913	0.42	254.2	49.5	10.24	80.97	61.39	-	-
$\bar{X}_3$	6162	2564	3598	0.41	249.3	46.2	10.84	84.38	58.85	-	-
F	42.38	72.58	15.63	41.87	15.28	28.50	51.65	37.61	0.29	-	-
Season two, Tel Hadya different hybrids (H1 $\bar{X}_1$, H2 $\bar{X}_2$, H3 $\bar{X}_3$)											
$\bar{X}_1$	8760	4325	4435	0.49	265.8	45.2	8.53	100.23	64.32	134.2	3.82
$\bar{X}_2$	6551	2942	3609	0.45	265.7	49.5	9.29	99.98	67.05	134.2	4.1
$\bar{X}_3$	5717	2765	2953	0.49	265.1	46.2	9.41	101.24	60.86	132.89	4.54
F	161.1	187.8	101.4	35.37	0.02	40.41	66.63	17.49	40.11	6.96	52.95
Season two, Breda different hybrids (H1 $\bar{X}_1$, H2 $\bar{X}_2$, H3 $\bar{X}_3$)											
$\bar{X}_1$	3213	1557	1656	0.49	226.5	40.0	11.44	90.2	45.67	131.1	3.53
$\bar{X}_2$	3592	1726	1866	0.48	265.3	42.2	12.1	89.78	50.48	129.9	3.63
$\bar{X}_3$	3343	1622	1721	0.49	271.2	40.6	12.45	90.42	44.29	129.2	3.8
F	12.15	9.92	10.49	0.41	80.43	50.15	94.27	15.97	192.35	88.13	7.42

The genotype had an effect on barley yields and morphological characters (Table 4.3) and showed that in the first season there were strongly significant differences between hybrids in terms of yield parameters (total plant yields (*F=42.38; P<0.001*), grains yields (*F=72.6; P<0.001*), straw yields (*F=15.63; P<0.001*), number of ears (*F=15.28; P<0.001*)). However, it can be seen that the genotype had no real impact on plant height, since there were no significant differences between them (*F=0.29; P=0.747*).

In the second season in Tel Hadya, there were no significant differences between the number of ears (F=*0.02; P=0.976*), but there were strongly significant differences between yield parameters (total plant yield (*F=161.1; P<0.001*), grain yield (*F=187.8; P<0.001*); straw yield (*F=101.4; P<0.001*)). The reasons for these differences are: (i) the number of grains in the ears was different from one hybrid to another (grain yield is affected by number of grains per ear). (ii) Straw yield is affected by plant height and leafiness. There were significant differences between morphological characters (vegetative duration (*F=17.49; P<0.001*), plant height (*F=40.11; P<0.001*), length of growing season (*F=6.96; P=0.001*), and leafiness (*F=52.95; P<0.001*)).

In Breda (second season), there were significant differences in yield parameters and morphological characters.

4.5 Summary

This chapter studied some of the factors affecting yield and yield parameters (environment, morphological characters, and genotype) and the significant findings were:

1- The influence of leafiness on yield: yield parameters have been affected by leafiness and the LSD test was used to study significant differences. The leafiness was stable only with H2 while H1 and H3 were affected strongly by location.

2- The influence of weather on (genotype) yield and morphological characters was studied by a comparative study of the same area during different seasons : the yield parameters and morphological characters were significantly affected by weather and H1 gave the best production at all hybrids.

3- The influence of environment on yield parameters and morphological characters: there were significant differences between mean data in different areas in the same season. Yields in Tel Hadya for each hybrid were much better than Breda.

4- Genotype tended to rank differently in yield parameters and morphological characters at different locations.

5- Genotype tended to rank differently in yield parameters and morphological characters in different years.

6- Vegetative duration and TGW were not affected by location (significant Spearman correlation).

7- The influence of genotype on yield and morphological characters: ANOVA was used to show that the genotype had an effect on yield parameters and morphological characters and H1 gave the best yield.

It can be seen that many factors have affected yield parameters and morphological characters when studied individually. However, the relationships between yield parameters and morphological characters now needs to be studied to see whether or not these relationships have been changed by groups of factors associated with the weather or environment. Canonical correlation analysis will be used to study the relationship between yield parameters as a set and morphological character also as a set in the next chapter.

Chapter 5

A study of the relationship between yield parameters as a set and morphological characters as a set using canonical correlation

5.1 Introduction

Canonical correlation is a type of multivariate statistical analysis which is used to analyze the relationships between multiple independent and multiple dependent variables (Shafto, *et al.*, 2005). While canonical correlation is a form of correlation relating two sets, there may be more than one significant canonical correlation. The maximum number of canonical correlations between two sets of variables is the number of variables in the smaller set (Anon, 2004a). Most of the relationship between two sets is explained by the first canonical correlation (see Chapter 2).

In this chapter, canonical correlation will be used to establish whether or not there is a relationship between morphological characters of barley as a set and yield parameters (biomass and grain) of barley, also as a set. Canonical correlation analysis will also be used to determine which variable among morphological characters has the strongest relationship with yield parameters. Pearson correlation will only be used to study the relationship between two variables, while canonical correlation will be used to study the relationship between sets. Thus, canonical correlation will be used to study the relationship between morphological characters and yield parameters and partial correlation will be used to study the relationship between the variables (yield parameters and morphological characters) and canonical variables. When the canonical correlation is found, the statistical significance in canonical correlation will be studied by means of the Bartlett test.

This chapter has three sections. The first will study the relationship between yield parameters as a set and morphological characters also as a set in each of the two areas Tel Hadya and Breda, in the second season. The second section develops from the first section, in which this relationship will be studied for the whole data during the second season using two different type of equations for yield parameters (linear and logarithmic). Finally, the comparison between canonical correlations (the first canonical correlation between yield parameters and morphological characters, and the second canonical correlation between logarithmic yield parameters and morphological characters) will be studied to see which is the best and for what reasons.

5.2 The relationship between yield parameters as a set and morphological characters also as a set in Tel Hadya

The canonical correlation coefficients between morphological characters as a set (vegetative duration, plant height, length of growing season, leafiness) and yield parameters of barley, also as a set, (total plant yield, grain yield, straw yield and weight of 1000 barley seeds (TGW)) were (*0.3981, $P<0.001$*), (*0.2703, $P<0.001$*), (*0.2368, $P<0.002$*) and (*0.0226, $P=0.6958$*). The first three canonical correlation coefficients were significant ($P<0.05$), while the fourth canonical correlation coefficient was not significant ($P>0.05$). That set has a statistically significant correlation. Therefore, there was a relationship between morphological characters and yield parameters. The first set of equations will be given to study this relationship. The equations are:

$$U = -0.4977x_1 + 0.6784x_2 + 0.5262x_3 - 0.3019x_4 \quad \text{5-1}$$

$$V = -573.756y_1 + 300.807y_2 + 308.233y_3 + 0.1649y_4 \quad \text{5-2}$$

where:

U is a canonical variable for morphological characters.

V is a canonical variable for yield parameters.

x_1 is a scaled vegetative duration [1].

x_2 is a scaled plant height .

x_3 is a scaled length of growing season.

x_4 is a scaled leafiness.

And

y_1 is a scaled total plant yield.

y_2 is a scaled grain yield.

y_3 is a scaled straw yield.

y_4 is a scaled TGW.

Figure 5.1 shows a scatter plot of the first set of canonical variables.

According to the coefficients in the equations 5-1 and 5-2, U is primarily related to plant height. The order of contribution of morphological characters is: plant height, length of growing season, vegetative duration and leafiness. According to the second equation, the canonical variable V is affected by the total plant yield. The order is: total plant yield, straw yield, grain yield and

[1] Scaled $X = \dfrac{x - mean(x)}{sd(x)}$

TGW. Plant height and length of growing season have a positive influence on the first canonical loading, while vegetative duration and leafiness have a negative influence on the first canonical loading. The second canonical loading had been affected positively by all variables except total plant yield.

The simple correlation for both the morphological characters and yield parameters with canonical variables (Table 5.1) showed that there are differences between the values of simple correlations (morphological characters). The simple correlations range for morphological characters and *U* was from *0.279* (length of growing season and *U*, $P<0.05$) to *0.703* (plant height and *U*), since vegetative duration and leafiness have a negative simple correlations with *U*. Total plant yield, grain yield and straw yield are equally important to *U* and they have significant simple correlations with canonical variable *U* ($P<0.01$), while TGW has no significant simple correlation with *U*. Also, Table 5.1 shows that there were simple correlations between *V* and morphological characters.

Table 5.1. The simple correlation between original variables and canonical variables (Tel Hadya).

Morphological characters	Canonical variables		Yield parameters	Canonical variables	
	$\hat{U}$	$\hat{V}$		$\hat{U}$	$\hat{V}$
Vegetative duration	*-0.567***	*-0.226***	Total plant yield	*0.361***	*0.907***
Plant height	*0.703***	*0.280***	Grain yield	*0.296***	*0.743***
Length of growing season	*0.279***	*0.111*	Straw yield	*0.385***	*0.967***
Leafiness	*-0.320***	*-0.127**	TGW	*0.062*	*0.157***

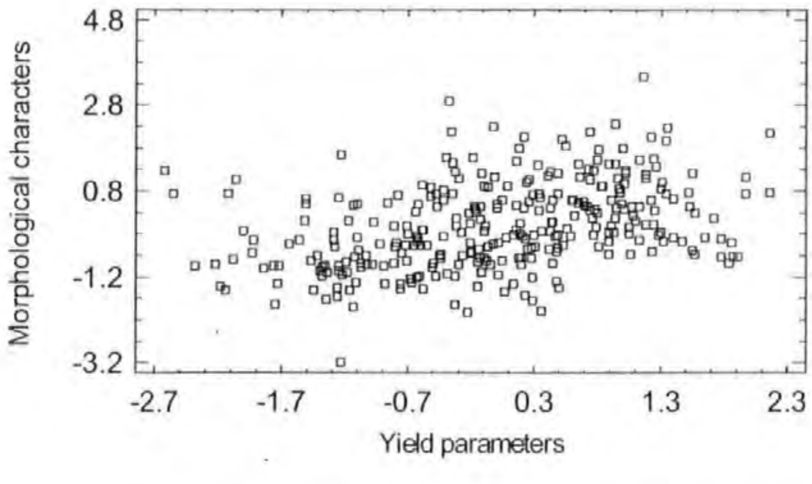


Figure 5.1. A scatter plot of the first canonical variables (Tel Hadya).

V has positive simple correlation with plant height (*0.280, P<0.01*), while it has a negative simple correlation with vegetative duration (*-0.226, P<0.01*). There was no significant relationship between length of growing season and canonical variable *V*, since there is a significant negative simple correlation between leafiness and *V* but this simple correlation was not strong (*-0.127, P<0.05*).The simple correlation between V and total plant yield was (*0.907, P<0.01*), *V* and grain yield was (*0.743, P<0.01*), while the strongest relationship appeared between straw yield and canonical variable *V (0.967, P<0.01)*. It can be seen that total plant yield, grain yield and straw yield were variables affected by morphological characters, while TGW was not affected by morphological characters. Consequently, TGW can be removed.

The proportion of the variance (each domain) is accounted for by canonical variates:

$$U^{*} = \frac{(-0.567)^2 + (0.703)^2 + (0.279)^2 + (-0.320)^2}{4} = 0.2490 \qquad 5-3$$

$$V^{*} = \frac{(0.907)^2 + (0.743)^2 + (0.967)^2 + (0.157)^2}{4} = 0.5836 \qquad 5-4$$

58% of the variance in the *Y* set is accounted for by *V*, while only *25%* of the *X* set is accounted for by *U*.

The simple correlation for cross-loading (*V* with morphological characters and *U* with yield parameters) indicates that length of growing season had no correlation with canonical variable V. Also TGW had no correlation with U. Approximately *nine per cent* of the variance in plant height is explained by

V, while less than *6%* of the variance for vegetative duration is explained by *V*. Also, *13%* of the variance for total plant yield and straw yield are explained by canonical variable *U*, while less than *10%* of variance for grain yield is explained.

In conclusion, there were significant canonical correlations between morphological characters as a set and yield parameters also as a set. The simple correlations between morphological characters and the canonical variable *V* indicated that plant height and vegetative duration were the most important variables of the morphological characters affecting yield parameters. There were no simple correlations between TGW and morphological characters. However, since there are insufficient data, the conclusion about the relationship remains equivocal.

5.3 The relationship between yield parameters as a set and morphological characters also as a set in Breda

The canonical correlation coefficients between morphological characters and yield parameters were not given because the variables (yield parameters) were linearly dependent[2]. Since a non-canonical correlation exists between morphological characters and yield parameters, it will be necessary to analyse the relationship between morphological characters and logarithmic yield parameters (logarithmic total plant yield, grain yield, logarithmic straw yield and TGW). The canonical correlation coefficients were *0.560, 0.183, 0.082 and 0.005*. Since the first canonical correlation was significant ($P<0.0001$),

---

[2] As out lined in Chapter 2.

there is relationship between the sets (morphological characters and logarithmic yield parameters). The sets of canonical variables are given by:

$$U = -0.0076x_1 + 0.7018x_2 - 0.1539x_3 + 0.6234x_4 \qquad 5\text{-}5$$

$$V = -1.8345y_1 + 1.2614y_2 + 1.7667y_3 + 0.1258y_4 \qquad 5\text{-}6$$

where: x_1 scaled vegetative duration, x_2 scaled plant height, x_3 scaled length of growing season, x_4 scaled leafiness, y_1 scaled logarithmic total plant yield, y_2 scaled grain yield, y_3 scaled straw yield and y_4 scaled TGW.

Figure 5.2 shows a scatter plot of the first set of canonical variables.

The morphological characters set shows that vegetative duration, and length of growing season have a negative influence on canonical variables, while plant height and leafiness have a positive influence.

The entire logarithmic yield parameters (except logarithmic total plant yield) have a positive effect on canonical variable *V*. Plant height and leafiness were the most important variables affecting *U* . The main contributors, in the morphological characters set, in order of importance, were: plant height, leafiness, length of growing season and vegetative duration. Vegetative duration and length of growing season did not have a significant effect on canonical variable *U*. Logarithmic total plant yield and logarithmic straw yield were the most important variable effecting *V*, while *V* was not affected by TGW.

The simple correlations between morphological characters and logarithmic yield parameters, each with canonical variables (Table 5.2), showed that the range for simple correlations for morphological characters and *U* differed from *0.252* (vegetative duration and *U*) to *0.756* (plant height and *U*). the results demonstrate that vegetative duration and length of growing season have negative simple correlations with *U* and *V*. Plant height has the strongest simple correlation with *U* (*0.756*; $P<0.01$) and with *V* (*0.423*; $P<0.01$). As shown in Table 5.2, the TGW did not have strong relationship with canonical variable *U* and *V*, while logarithmic total plant yield, logarithmic grain yield and logarithmic straw yield had a very strong relationship with *V* (*0.972*, *0.867*, *0.942*, respectively; $P<0.01$*in each case*). Plant height and leafiness had positive simple correlations with *U* and *V*. Also yield parameters had positive simple correlations with canonical variables *U* and *V*.

Table 5.2. Simple correlation between original variables and canonical variables in Breda.

Morphological characters	Canonical variables		Yield parameters	Canonical variables	
	$\hat{U}$	$\hat{V}$		$\hat{U}$	$\hat{V}$
Vegetative duration	*-0.252***	*-0.141**	Total plant yield	*0.544***	*0.972***
Plant height	*0.756***	*0.423***	Grain yield	*0.485***	*0.867***
Length of growing season	*-0.277***	*-0.155***	Straw yield	*0.527***	*0.942***
Leafiness	*0.682***	*0.381***	TGW	*0.113**	*0.201***

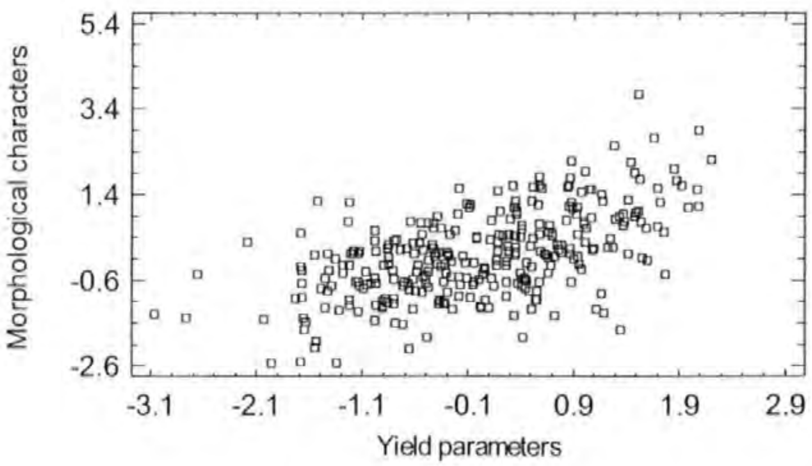


Figure 5.2. A scatter plot of the first canonical variables (Breda).

5.4 The relationship between yield parameters as a set and morphological characters also as a set in the second season

In this section, the canonical correlation coefficients between morphological characters as a set (vegetative duration, plant height, length of growing season, leafiness) and yield parameters of barley, also as a set, (total plant yield, grain yield, straw yield, thousand grain weight (TGW)) will be studied for the whole data in the second season . There is a strong significant canonical correlation between morphological characters and yield parameters (*0.87*, $P<0.001$). The eigenvalue, or canonical root, is the square of the canonical correlation; and this measures the proportion of variance of canonical variables. The canonical correlation showed that *76%* of yield parameters are explained by the correlation between the two canonical variables.

Table 5.3 shows the canonical correlations, canonical r-square, probability of exceeding the critical value (Chi-Square), percentage trace[3] and cumulative trace (%).

According to Table 5.3, there are two significant canonical correlations. However, the first canonical correlation gives about *92%* of trace.

There are two sets of canonical variables, but only the first set will be studied. The first set of canonical variables is given by the equations:

$$\begin{aligned} Yield &= 37.092 \times Total\ Plant\ yield - 17.994 \times Grain\ yield \\ &\quad - 19.230 \times Straw\ yield + 0.431 \times TGW \end{aligned} \qquad 5\text{-}7$$

[3] % Trace= $r_i^2 * 100 / \sum r_i^2$

$$morpho\log y = 0.398 \times vegetative\ duration + 0.550 \times Plant\ height + 0.152 \times Length\ of\ growing\ season - 0.011 \times Leafiness \quad 5\text{-}8$$

The first canonical variable (yield) for yield parameters is the weighted difference of total plant yield (*37.092*), grain yield (*-17.994*) and straw yield (*-19.230*).

The coefficient for TGW is close to *0*. Equations 5-7 indicated that total plant yield and straw yield are the most important variables in canonical variable (yield). There was *87%* of the variation in the total plant yield and straw yield explained by the first canonical variable (yield), *79%* for grain and 67% for TGW (Table 5.4). There were significant and positive correlation coefficients between the first canonical variable (yield) and yield parameters , *0.93* for total plant yield and straw yield with canonical variable yield, *0.89* for grain yield and *0.82* for TGW (Table 5.4).

The coefficient of grain yield and straw yield in the first canonical variable (yield) and the correlation have opposite signs. The correlation coefficients between yield variables and the second canonical variable (morphology) were positive and significant, *0.80* for total plant yield and straw yield with canonical variable morphology, *0.77* for grain yield and *0.71* for TGW (Table 5.4).

Table 5.3. The canonical correlations, canonical r-square, probability of exceeding the critical value (*Chi-Square*), percentage trace and cumulative trace(%).

I	Canonical correlations	Canonical r-square	P	Trace %	Cumulative Trace %
1	0.87	0.76	<0.001	92.36	92.36
2	0.22	0.05	<0.001	5.91	98.27
3	0.10	0.01	0.076	1.22	99.49
4	0.065	0.00	0.111	0.51	100.00

Table 5.4. The simple correlations between yield parameters, morphological characters and canonical variables.

Yield parameters	Canonical yield	Canonical morph_	Morphological characters	Canonical yield	Canonical morph_
Total yield	*0.93*	*0.80*	Vegetative duration	*0.81*	*0.94*
Grain yield	*0.89*	*0.77*	Plant height	*0.82*	*0.95*
Straw yield	*0.93*	*0.80*	Length of growing season	*0.63*	*0.72*
TGW	*0.82*	*0.71*	Leafiness	*0.35*	*0.41*

The first canonical variable (morphology) for morphological characters also shows that plant height is the most important variable which had an effect on canonical variable (morphology). The coefficients have different weights on the first canonical variable (morphology). Plant height is *0.55*, vegetative duration is *0.40* and length of growing season is *0.15*. It can be seen that canonical variable (morphology) has not been strongly affected by leafiness (*-0.01*) (Equation 5-8).

The simple correlation coefficients between the first canonical variable (morphology) and morphological characters were significant: *0.95* for plant height, *0.94* for vegetative duration, *0.72* for length of growing season and *0.41* for leafiness. The morphological characters' variables have a differing degree of variation, which is explained by the first canonical variable (morphology). Approximately *89%* of the variation in the plant height is explained by the first canonical variable (morphology), while only *17%* of the variation in leafiness is explained.

There were significant simple correlation coefficients between the first canonical variable (yield) and morphological characters' variables. The relationship between them is: weighted difference of plant height (*0.67*); vegetative duration (*0.66*); length of growing season (*0.39*); and finally leafiness (*0.13*), with plant height having greater significance. The results indicate that leafiness is a suppressor variable.

5.5 Conclusion and summary

In terms of biology, there was a very strong relationship between morphological characters and yield parameters. Examining the relationship between them it is clear that:

- With regard to plant height, tall barley has a greater yield than shorter barley. Hence, the correlation between yield parameters (total plant yield, grain yield, straw yield, TGW) and plant height were positive. For example, in Tel Hadya the plant height was greater than in Breda so the yield was bigger.
- Barley with longer vegetative duration and length of growing season tends to give better yields than barley with shorter vegetative duration and length of growing season, since these variables correlate positively with yield parameters.
- The relationship between yield parameters and leafiness does not appear to be very strong because the genotype was not stable. The result is supported by lack of significance in the Wilcoxon test[4], (H1: $P=0.261$; H2: $P=0.121$; H3:$P=0.078$). Table 5.5 shows that Δx ($\Delta x = x - \bar{x}$) for H1 in Tel Hadya which has 56 families' with higher ranks than Breda. On the other hand, there are 46 families in Breda which have higher ranks than Tel Hadya. H3 has same number of families 51.

[4] The Wilcoxon signed-ranks method tests the null hypothesis that the medians of two related populations are the same (see Chapter 2).

Table 5.5. Wilcoxon Signed Ranks Test Tel Hadya.

		N[5]	Sum of Ranks	Z	Sig, P
Δx Breda - Δx Tel Hadya H1	Negative Ranks[6]	56	2292.00	-1.124	0.261
	Positive Ranks[7]	46	2961.00		
	Ties[8]	0	0		
	Total	102			
Δx Breda - Δx Tel Hadya H2	Negative Ranks	65	3088.00	-1.549	0.121
	Positive Ranks	37	2165.00		
	Ties	0	0		
	Total	102			
Δx Breda - Δx Tel Hadya H3	Negative Ranks	51	3160.00	-1.792	0.073
	Positive Ranks	51	2093.00		
	Ties	0	0		
	Total	102			

[5] Number of families

[6] Δx Breda < Δx Tel Hadya

[7] Δx Breda > Δx Tel Hadya

[8] Δx Breda = Δx Tel Hadya

- TGW was not affected strongly by morphological characters.
- The results indicate that plant height is the most important variable among morphological characters influencing yield.

Summary:

There was a significant relationship between morphological characters and yield parameters. However, these relationships were least evident in each of the two areas (Tel Hadya and Breda) because the sample sizes were insufficient to provide definite conclusions.

Studying canonical correlation between logarithmic yield parameters and morphological characters in Breda gives better results than studying canonical correlation between yield parameters and morphological characters.

While there are significant relationships between morphological characters and yield parameters, Chapter Six is going to study the interrelationship for morphological characters and yield parameters by using factor analysis.

Chapter 6

Using factor analysis to study the interrelationship among yield parameters morphological characters

6.1 Introduction

Factor analysis is a mathematical model used to reduce the number of variables (Thompson, 2004), being used for many purposes, including the following:

- To reduce a large number of variables to a smaller number of factors for modeling purposes.
- To select a subset of variables from a larger set, according to correlation.
- To create a set of independent factors.
- To determining which sets of variables cluster together.

This chapter will use factor analysis to study the interrelationship among yield parameters and morphological characters (total plant yield, grain yield, straw yield, TGW, vegetative duration, plant height, length of season growing and leafiness) and to reduce them to a smaller number of key factors.

There are three sections in this chapter. The first will be a mathematical description of H1 in Tel Hadya. Mathematical methods will be used to discover how each factor is calculated and to compare the results of using principal components factor analysis and maximum likelihood factor analysis, respectively.

Second, factor analysis will be used to study the interrelationship of all data (H1, H2 and H3) in each area (Tel Hadya and Breda). Mathematical and biological descriptions will also be given in this section. Third, a conclusion and a summary will be given.

6.2 The interrelationship between morphological characters and yield parameters for H1 in Tel Hadya

According to the data for H1 in Tel Hadya, there are eight variables (total plant yield x_1, grain yield x_2, straw yield x_3, TGW x_4, vegetative duration x_5, plant height x_6, length of growing season x_7 and leafiness x_8). Firstly, A correlation matrix is used by factor analysis to determine which sets of variables cluster together (Anon, 2005). The matrix of mean ($\bar{X}$) and correlation matrix (R) for the variables are:

$$\bar{X}' = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \bar{x}_3 & \bar{x}_4 & \bar{x}_5 & \bar{x}_6 & \bar{x}_7 & \bar{x}_8 \\ 8759.7 & 4324.8 & 4434.8 & 45.16 & 100.23 & 62.92 & 133.8 & 3.82 \end{bmatrix}$$

$$R = \begin{bmatrix} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ x_2 & 0.897^{**} & & & & & & \\ x_3 & 0.884^{**} & 0.588^{**} & & & & & \\ x_4 & 0.058 & 0.057 & 0.046 & & & & \\ x_5 & -0.218^{*} & -0.161 & -0.230^{*} & 0.319^{**} & & & \\ x_6 & 0.323^{**} & 0.256^{*} & 0.322^{**} & 0.026 & -0.304^{**} & & \\ x_7 & -0.45 & 0.020 & -0.105 & 0.324^{**} & 0.265^{**} & -0.133 & \\ x_8 & 0.210^{*} & 0.162 & 0.214^{*} & 0.089 & 0.108 & 0.290^{**} & 0.049 \end{bmatrix}$$

** $P < 0.01$

* $P<0.05$.

Total plant yield and straw yield had a weak correlation with vegetative duration (*r=-0.218, -0.230, P<0.05, respectively*), plant height (*r=0.323, 0.322, P<0.01 respectively*) and leafiness (*r=0.210, 0.214, P<0.05 respectively*) while grain yield correlated weakly with plant height (*r=0.256,* *P<0.05*). Length of growing season did not correlate with yield parameters except TGW (*r=0.324, P<0.05*) and vegetative duration (*r=0.265, P<0.05*). Leafiness correlated significantly with total plant yield, straw yield and plant height.

6.2.1 Principal Components Factor Analysis

According to the correlation matrix, there are three groups of correlated variables. The first group consisted of total plant yield, grain yield and straw yield. The second group consisted of TGW, vegetative duration and length of growing season. Finally, the third group consisted plant height and leafiness.

The eigenvalues and eigenvectors are given by the equation (2 - 32; Chapter 2):

$$
\begin{array}{ll}
\hat{\lambda}_1 = 2.930 & \tilde{e}'_1 = [0.561 \quad 0.490 \quad 0.509 \quad 0.008 \quad -0.213 \quad 0.312 \quad -0.080 \quad 0.192] \\
\hat{\lambda}_2 = 1.642 & \tilde{e}'_2 = [0.097 \quad 0.128 \quad 0.042 \quad 0.583 \quad 0.522 \quad -0.107 \quad 0.540 \quad 0.238] \\
\hat{\lambda}_3 = 1.066 & \tilde{e}'_3 = [-0.227 \quad -0.272 \quad -0.128 \quad 0.041 \quad 0.009 \quad 0.578 \quad -0.136 \quad 0.710] \\
\hat{\lambda}_4 = 0.805 & \tilde{e}'_4 = [-0.103 \quad -0.085 \quad -0.099 \quad 0.407 \quad -0.510 \quad 0.478 \quad 0.379 \quad -0.417] \\
\hat{\lambda}_4 = 0.700 & \tilde{e}'_5 = [0.004 \quad -0.093 \quad 0.106 \quad 0.0580 \quad 0.243 \quad 0.100 \quad -0.709 \quad -0.269] \\
\hat{\lambda}_4 = 0.470 & \tilde{e}'_6 = [-0.026 \quad -0.236 \quad 0.203 \quad 0.358 \quad -0.573 \quad -0.556 \quad -0.059 \quad 0.366] \\
\hat{\lambda}_4 = 0.386 & \tilde{e}'_7 = [-0.015 \quad 0.627 \quad -0.691 \quad 0.169 \quad -0.184 \quad -0.100 \quad -0.183 \quad 0.151] \\
\hat{\lambda}_4 = 3.07E-06 & \tilde{e}'_8 = [0.783 \quad -0.452 \quad -0.427 \quad 0.000 \quad 0.000 \quad 0.000 \quad 0.000 \quad 0.000]
\end{array}
$$

There are several methods for determining the number of factors. Here two methods are used: Scree test and the eigenvalues greater than 1.0 rule (known as the Kaiser criterion; Krzanowski, 1990).

Scree test (developed by Cattell in 1966 (Thompson, 2004)): used to identify the optimum number of factors that can be extracted before the amount of unique variance begins to dominate the common variance structure (Hair, 2005). Figure 6.1 shows that starting with the first factor, the slopes steeply downward initially and then (after third factor) slowly becomes an approximately horizontal line. This plot suggests that there are three or four factors.

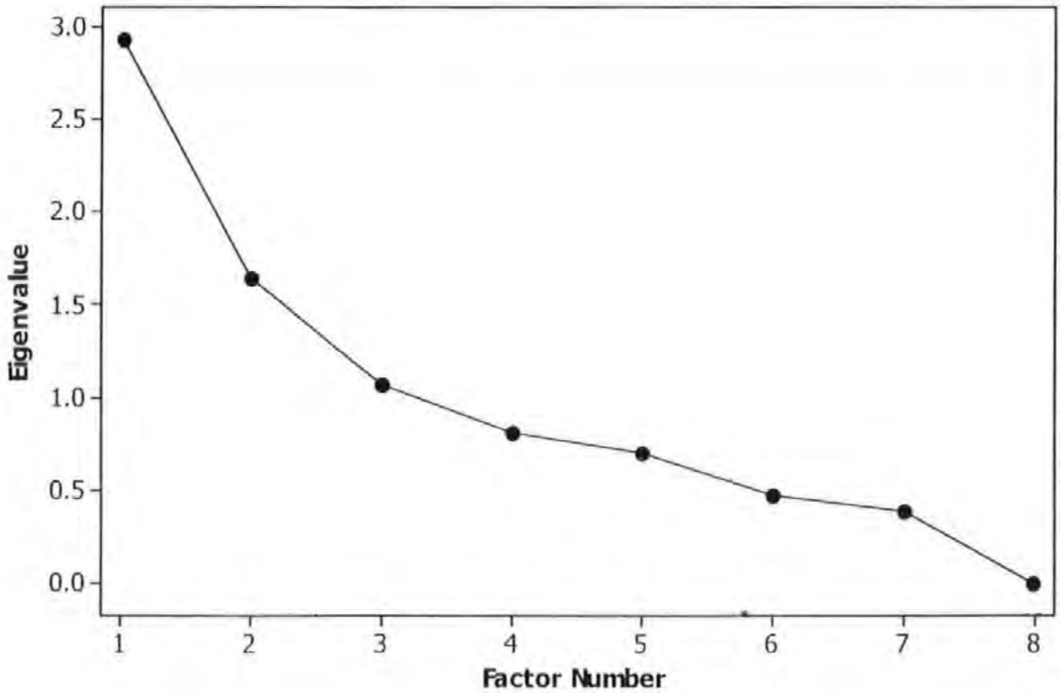


Figure 6.1. Scree plot for seven variables in Tel Hadya.

Eigenvalues greater than *1.0* rule:

The outstanding factors should have eigenvalues greater than *1.0*. This idea was suggested by Cuttman in 1954 (Thompson, 2004). The factors with eigenvalues greater than *1.00* are considered significant; all factors with eigenvalues less than *1.00* are considered insignificant and disregarded (Hair, 2005). Whilst this criterion is a useful rule of thumb, it is very important for the researcher to remember that all values contain some sampling error. Therefore, in the context of particular investigation, a researcher could ,for example, choose to retain a factor with an eigenvalue of slightly less than *1.0* (Thompson, 2004). There are three factors since their eigenvalues are greater than *1.00*.

According to Scree test rule and Eigenvalues greater than *1.0* rule, there are three critical factors.

Cumulative proportion: it is used by scientists to determine the total variance explained by the number of factors. One factor explained about *36.63%*[1] of the total variance; and two and three factors explain *57.16%*[2], *70.50%* respectively of the total variance (Table 6.1).

---

[1] $\frac{\hat{\lambda}_1}{p} = \frac{2.930}{8} = 0.3663$

[2] $\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{p} = \frac{2.930 + 1.642}{8} = 0.5716$

Table 6.1. Number of factors and Communalities.

Number of factors	Cumulatively
1	36.63
2	57.16
3	70.50
4	80.50
5	89.32
6	95.24
7	1
8	1

Factor loadings: the correlations between variables and factors are called factor loadings (Kline, 1994). According to Equation 2-45 $\left(\tilde{L} = \left[\sqrt{\hat{\lambda}_1}\hat{e}_1 \quad \sqrt{\hat{\lambda}_2}\hat{e}_2 \quad \sqrt{\hat{\lambda}_3}\hat{e}_3\right]\right)$, the estimation of factor loadings are:

$$\tilde{L} = \begin{bmatrix} 0.960 & 0.124 & -0.234 \\ 0.839 & 0.164 & -0.281 \\ 0.872 & 0.054 & -0.132 \\ 0.013 & 0.747 & 0.042 \\ -0.365 & 0.670 & 0.009 \\ 0.534 & -0.137 & 0.597 \\ -0.138 & 0.692 & -0.140 \\ 0.328 & 0.305 & 0.732 \end{bmatrix}$$

If the loading factor is greater than *0.6* (negative or positive) it is regarded as high; between *0.6* and *0.3* it is moderately high; and finally if it is less than *0.3* it is ignored (Kline, 1994). According to the $\tilde{L}$ matrix (factor loadings), first factor correlates with total plant yield (*0.960*), grain yield (*0.839*) and straw yield (*0.872*) so it is termed "yield". The second factor is termed "plant period" because it is highly correlated with TGW (*0.747*), vegetative duration

(*0.670*) and length of growing season (*0.692*). The third factor correlated with leafiness (*0.732*) and moderately highly correlated with plant height (*0.597*) is termed "morphology".

A matrix $\tilde{L}$ shows that the total plant yield, grain yield and straw yield have positive loadings on the first factor while they are ignored on the other factors (less than *0.3*). TGW and length of growing season are ignored on the first and third factors (less than *0.30*). Length of growing season, vegetative duration and plant height have positive loadings on the second factor (matrix $\tilde{L}$). Vegetative duration is ignored on the third factor while it is moderately high on the first factors (*-0.365*) (matrix $\tilde{L}$). Plant height has positive moderately high loadings on the first and the third factors since it is ignored on the second factor (matrix $\tilde{L}$). Leafiness has positive loadings on all three factors.

Communality and residuals: Thompson (2004) defines communality (h^2) as follows; the communality for a measured variable reflects how much of the variance of a given measured variable is useful in delineating the factor as a set. It is the sum of squared factor loading for that variable, and the value of communality is between 0.0 and 1.0. When the communality has value 1.0 that means all of the variance in the variables is explained by all of the factors.

According to Equation (2 - 37; Chapter 2) ($h_i^2 = l_{i1}^2 + l_{i2}^2 + l_{i3}^2$) where l_{ij} is the loading factor from matrix *L*, the estimation of communalities are:

$$h_1^2 = (0.985)^2 + (0.124)^2 + (-0.234)^2 = 0.991,\ h_2^2 = 0.810,\ h_3^2 = 0.780$$

$$h_4^2 = 0.560,\ \ h_5^2 = 0.582,\ \ h_6^2 = 0.660,\ \ h_7^2 = 0.518,\ \ h_8^2 = 0.737$$

where:

h_1^2 is a communality for total plant yield.

h_2^2 is a communality for grain yield.

h_3^2 is a communality for straw yield.

h_4^2 is a communality for TGW.

h_5^2 is a communality for vegetative duration.

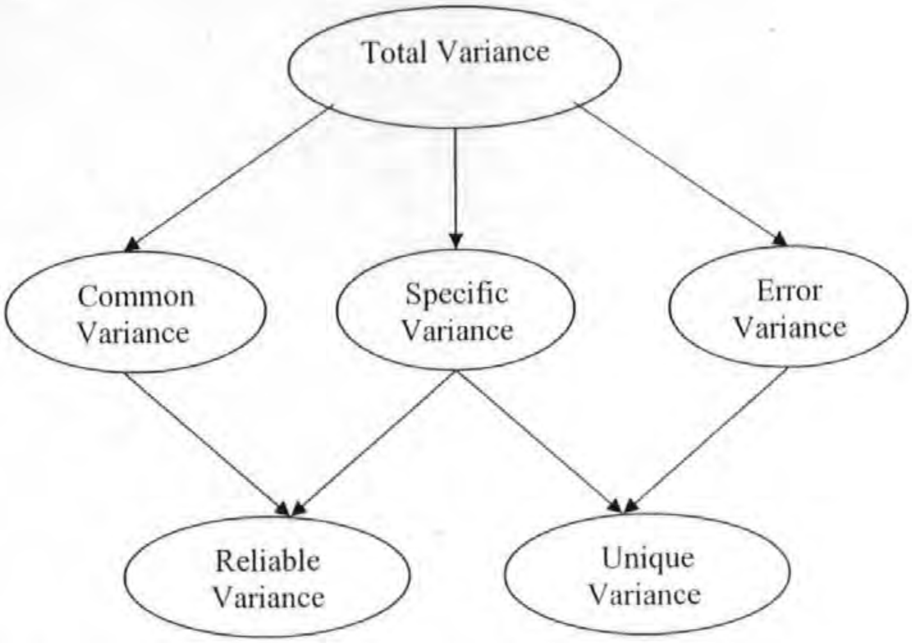
h_6^2 is a communality for plant height.

h_7^2 is a communality for length of growing season.

h_8^2 is a communality for leafiness.

99% of variance in total plant yield is explained by three factors, but most of this variance is explained by the first factor. Also *81%* of variance in grain yield and *78%* in straw yield are explained by three factors. *56%* of variance in TGW is explained by three factors, while vegetative duration *58%*, plant height *66%*, length of growing season *52%* and leafiness *74%*.

Unique variance ψ_i(i=1,2,..,7) is the variance which is not explained by factors and not associated with other factors. It is composed of specific and error variance. It is shown by the following diagram:



It is given by the equation:

$$\hat{\psi}_i = 1 - \sum_{j=1}^{2} l_{ij}^2$$

$$\left.\begin{array}{l}
\tilde{\psi}_1 = 1 - 0.991 = 0.009 \\
\tilde{\psi}_2 = 1 - 0.810 = 0.190 \\
\tilde{\psi}_3 = 1 - 0.780 = 0.220 \\
\tilde{\psi}_4 = 1 - 0.560 = 0.440 \\
\tilde{\psi}_5 = 1 - 0.582 = 0.418 \\
\tilde{\psi}_6 = 1 - 0.660 = 0.340 \\
\tilde{\psi}_7 = 1 - 0.518 = 0.482 \\
\tilde{\psi}_8 = 1 - 0.737 = 0.263
\end{array}\right\} \Rightarrow \tilde{\psi} = \begin{bmatrix}
0.009 & & & & & & & \\
& 0.190 & & & & & & \\
& & 0.220 & & & & & \\
& & & 0.440 & & & & \\
& & & & 0.418 & & & \\
& & & & & 0.340 & & \\
& & & & & & 0.482 & \\
& & & & & & & 0.263
\end{bmatrix}$$

where :

$\tilde{\psi}_1$ is the unique variance for total plant yield.

$\tilde{\psi}_2$ is the unique variance for grain yield.

$\tilde{\psi}_3$ is the unique variance for straw yield.

$\tilde{\psi}_4$ is the unique variance for TGW.

$\tilde{\psi}_5$ is the unique variance for vegetative duration.

$\tilde{\psi}_6$ is the unique variance for plant height.

$\tilde{\psi}_7$ is the unique variance for length of growing season.

$\tilde{\psi}_8$ is the unique variance for leafiness.

Residual matrix for variables (yield parameters and morphological characters) is the covariance matrix of the parts of the variables unexplained by the factors. Sometimes it can be used to extract factors until the residual matrix is very small (Kline, 1994).

The residual matrix corresponding to the three factors is:

$$R-\tilde{L}\tilde{L}'-\tilde{\psi}=\begin{bmatrix} 0 \\ 0.006 & 0 \\ 0.010 & -0.190 & 0 \\ -0.038 & -0.065 & 0.000 & 0 \\ 0.051 & 0.038 & 0.053 & -0.177 & 0 \\ -0.032 & -0.001 & -0.057 & 0.097 & -0.022 & 0 \\ -0.032 & -0.017 & -0.040 & -0.185 & -0.247 & 0.119 & 0 \\ 0.029 & 0.042 & 0.008 & -0.174 & 0.016 & -0.281 & -0.015 & 0 \end{bmatrix}$$

The residual matrix indicated that there was about 25% of covariance between vegetative duration and length of growing season and 28% of covariance between length of growing season and leafiness. However, the residual matrix was small.

The mathematical technique for factor analysis depending on principal components is summarized by Table 6.2 of the mathematical technique.

Rotation Factors:

Firstly the determination of minimum number of factors and the communalities of each factor should be studied. The next step in factor analysis is to deal with rotation factors. Rotation factors are a very important tool in interpreting factors, while keeping the number of factors and communalities of each variable fixed. The varimax rotation[3] has been used to rotate the axes to make all of the loadings close to zero or one (Group, 2005). According to the Equation (2 - 63) in Chapter 2, the factor transformation matrix is:

$$T = \begin{bmatrix} 0.917 & -0.174 & 0.359 \\ 0.141 & 0.983 & 0.118 \\ -0.374 & -0.058 & 0.926 \end{bmatrix}$$

According to the equation $\underset{(p\times 2)}{L^*} = \underset{(p\times 2)}{\hat{L}} \underset{(2\times 2)}{T}$ new factor loadings will be collected.

Rotated and unrotated factor loadings are shown by Table 6.3

[3] Varimax rotation is a type of rotation to obtain a simplified factor solution.

Table 6.2. Factor analysis by using principal components (Tel Hadya- H1)

Variables	Three-factors solution			Estimated communality	Specific variances
	F_1	F_2	F_2	$\hat{h}_i^2$	ψ_i^2
Total plant yield	*0.960*	*0.124*	*-0.234*	*0.991*	*0.009*
Grain yield	*0.839*	*0.164*	*-0.281*	*0.810*	*0.190*
Straw yield	*0.872*	*0.054*	*-0.132*	*0.780*	*0.220*
TGW	*0.013*	*0.747*	*0.042*	*0.560*	*0.440*
Vegetative duration	*-0.365*	*0.670*	*0.009*	*0.582*	*0.418*
Plant height	*0.534*	*-0.137*	*0.597*	*0.660*	*0.340*
Length of growing season	*-0.138*	*0.692*	*-0.140*	*0.518*	*0.482*
Leafiness	*0.328*	*0.305*	*0.732*	*0.737*	*0.263*
Eigenvalues	2.930	1.642	1.066	Total % of variance	
% of Variance	**36.63**	**20.47**	**13.40**	**70.50**	

Table 6.3. Rotated factor loadings

Variables	Rotated factor loadings		
	F_1	F_2	F_3
Total plant yield	***0.985***	*-0.032*	*0.143*
Grain yield	***0.898***	*0.031*	*0.060*
Straw yield	***0.856***	*-0.092*	*0.197*
TGW	*0.101*	***0.729***	*0.132*
Vegetative duration	*-0.244*	***0.721***	*-0.043*
Plant height	*0.247*	*-0.263*	***0.728***
Length of growing season	*0.024*	***0.712***	*-0.097*
Leafiness	*0.070*	*0.200*	***0.832***

Table 6.3 indicated that yield parameters (Total plant yield, grain yield and straw yield) after rotation are close to the first factor (omitted from the second and third factors). TGW, vegetative duration and length of growing season are close to the second factor (omitted on the first and third factors). Plant height and leafiness allocated on the third factor. There are three factors: the first factor is called "yield" (total plant yield, grain yield, straw yield), the second is called "plant period" (vegetative duration, length of growing season and TGW) and the third is called "morphology" (plant height and leafiness) (Figure 6.2) and Table 6.3.

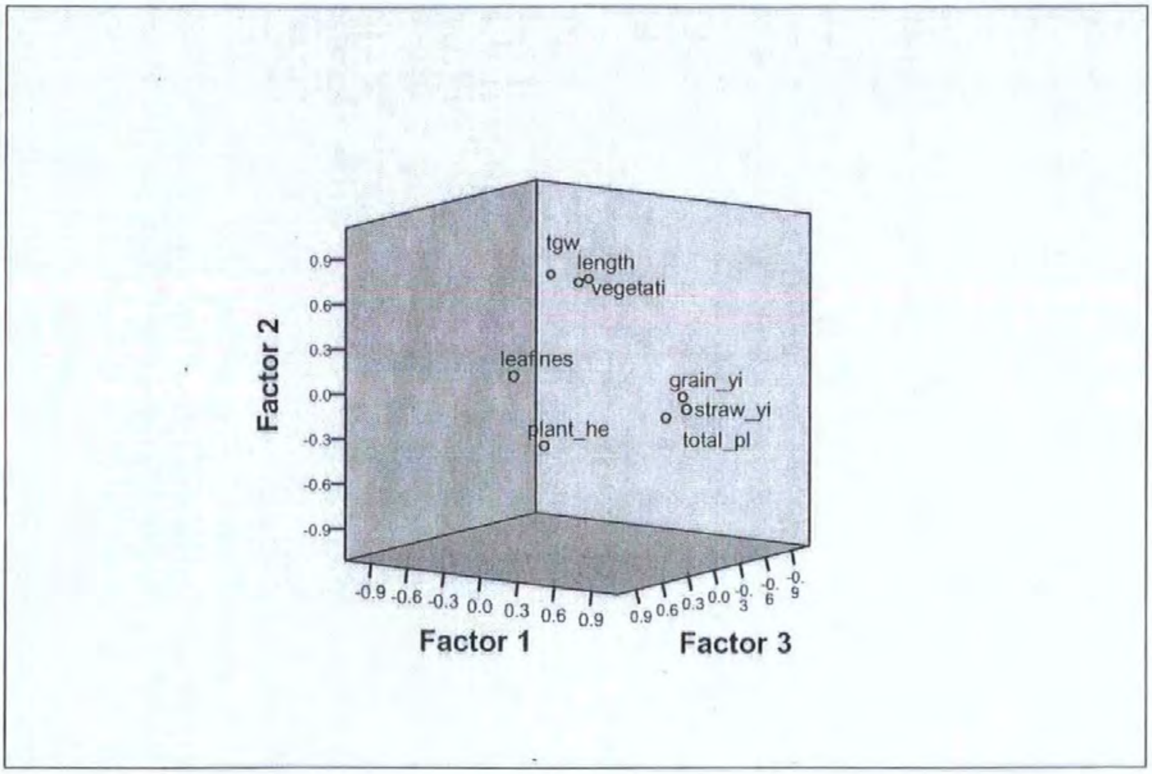


Figure 6.2. Factors rotation (Tel Hadya, H1).

Factor score coefficients: the estimation of the common factors, called factor scores (Kim and Mueller, 1978) (See Chapter 2). The estimation of values for unobserved random factor vectors F_j, $j = 1,2,3$ are indicated by Table 6.4.

Table 6.4. Factor scores for Tel Hadya , H1.

Variable	Factor 1 (Yield)	Factor 2 (Plant period)	Factor 3 (Morphology)
Total plant yield	*0.393*	*0.030*	*-0.077*
Grain yield	*0.375*	*0.063*	*-0.130*
Straw yield	*0.324*	*-0.013*	*-0.004*
TGW	*0.053*	*0.444*	*0.092*
Vegetative duration	*-0.060*	*0.422*	*0.012*
Plant height	*-0.054*	*-0.147*	*0.574*
Length of growing season	*0.065*	*0.430*	*-0.089*
Leafiness	*-0.128*	*0.123*	*0.698*

6.2.2 Maximum Likelihood Factor Analysis

The estimation of unrotated factor loadings, rotated factor loadings, communality, unique variance, eigenvalue, total sample variance explained by each factor and total variance are shown in Table 6.5 (for more information see Chapter 2).

Maximum likelihood factor analysis for H1 in Tel Hadya gives a result consistent with the Heywood Case[4]. Table 6.5 indicated that TGW, vegetative duration, plant height, length of growing season and leafiness have small loadings (less than 0.3) and these are omitted . Total plant yield, grain yield and straw yield have positive loadings on the first factor and negative on the second factor (before rotation). However straw yield changed to positive on the second factor (after rotation).

The estimation of communalities ($\hat{h}_i^2$) for grain yield which are calculated by factors are larger for the maximum likelihood than for that method principal components. Other variables (TGW, vegetative duration, length of growing season, leafiness) have much smaller communalities when calculated by maximum likelihood than when calculated by principal components. The total sample variances calculated by each factor are larger for principal components factor analysis than for maximum likelihood. Also the principal

[4] "Heywood Case occurs when the minimum of the discrepancy function is obtained with one or more negative values as estimates for the variance of the unique variables" **StatSoft, I.** 2006. *Text book.* [on-line] Available: www.statsoft.com/textbook/glosh.html [date accessed: 2005]

component estimates $\tilde{L}$ and $\tilde{\psi}$ are better than the maximum likelihood because the residual matrix is smaller.

Finally, there are different results between maximum likelihood estimate and that of principal components because TGW, vegetative duration, plant height, length of growing season and leafiness are omitted for the maximum likelihood while they explain by principal components.

Table 6.5. Factor loadings is estimated by Maximum likehood for Tel Hadya, H1.

Variables	Unrotated factor loadings		Rotated factor loadings		$\hat{h}_i^2$	ψ_i^2
	F_1	F_2	F_1	F_2		
Total plant yield	0.714	-0.534	0.891	-0.036	0.795	0.215
Grain yield	0.548	-0.824	0.918	-0369	0.978	0.022
Straw yield	0.729	-0.108	0.662	0.323	0.543	0.457
TGW	-0.040	-0.078	0.011	-0.088	0.008	0.992
Vegetative duration	-0.129	0.076	-0.149	-0.010	0.022	0.978
Plant height	0.248	-0.111	0.268	0.049	0.074	0.926
Length of growing season	-0.142	-0.123	-0.048	-0.181	0.035	0.965
Leafiness	0.266	-0.014	0.227	0.139	0.071	0.929
Eigenvalue	2.222	0.304	2.222	0.304	*Total variance*	
% Variance	27.8	3.8	27.8	3.8	31.6	

6.3 The interrelationship between morphological characters and yield parameters

Yield parameters except TGW are not distributed normally while ln (total plant yield), ln (grain yield) and ln (straw yield) have more normal distributions. For example, Figures 6.3, 6.4 and 6.5 show that ln(total plant yield), ln (grain yield) and ln(straw yield) for the second season had a normal distribution while total plant yield, grain yield and straw yield were not distributed normally. Consequently, factor analysis was used to study the interrelationship between the variables (ln(total plant yield) x_1, ln(grain yield) x_2, ln(straw yield) x_3, TGW x_4, vegetative duration x_5, plant height x_6, length of growing season x_7 and leafiness x_8).

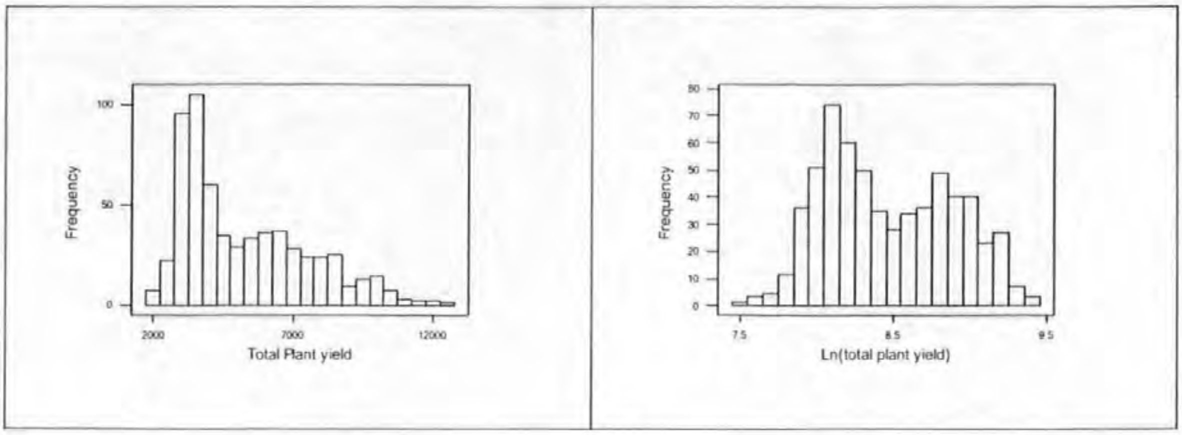


Figure 6.3. Distributions of total plant yield and ln(total plant yield).

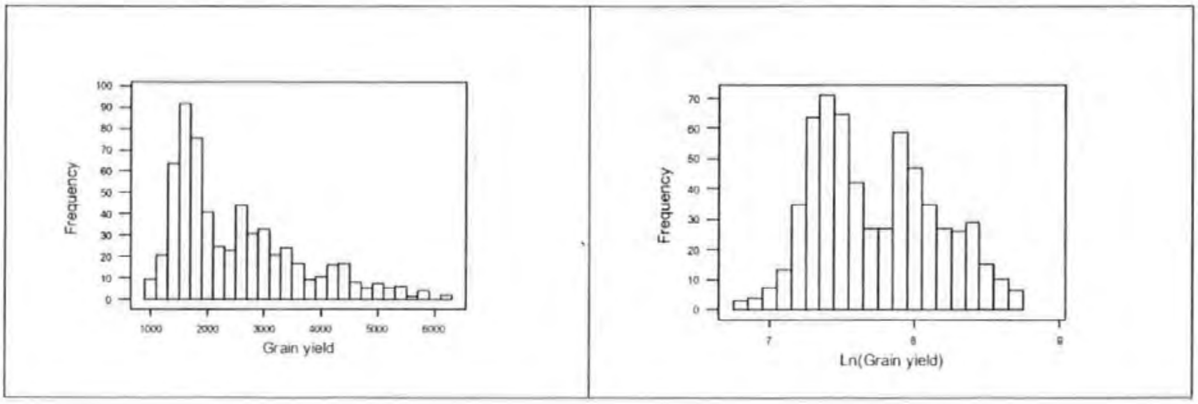


Figure 6.4. Distributions of grain yield and ln(grain yield).

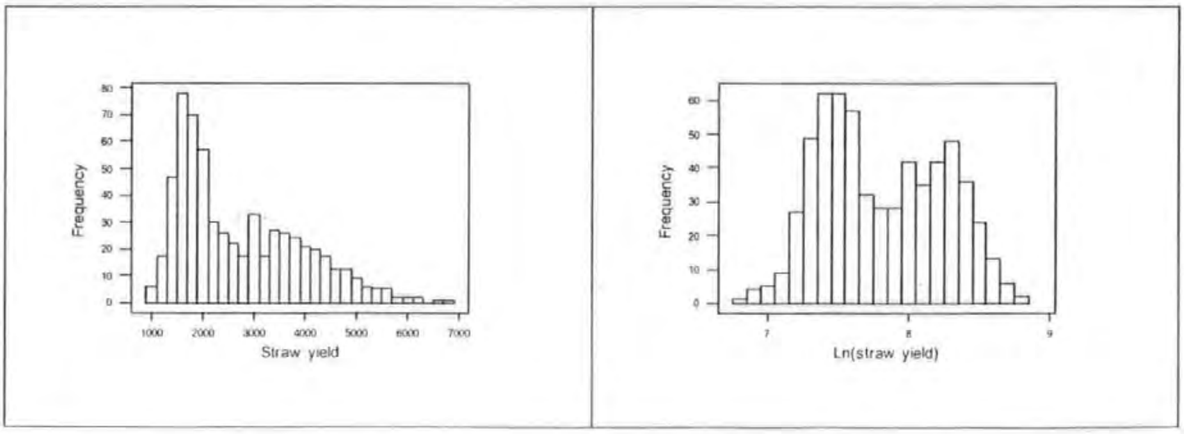


Figure 6.5. Distributions of straw yield and ln(straw yield).

The total variance explained by factors is indicated in Tables 6.6 and 6.7, the first 3 factors, which account for *71.8%* (H2, Tel Hadya), *79.1%* (H3, Tel Hadya), *72.7%* (H1, Breda), *74.8%* (H2, Breda), *69.8%* (H3, Breda), 72.2 % (Tel Hadya), *66.4%* (Breda) of the total variances, are important. However, the total variance explained by two factors in the second season (Tel Hadya and Breda) was *80.0%* of the total variance.

Factor 1 for Tel Hadya (H1, H2, H3 and Tel Hadya), which accounted for about *40%* of the variation, was strongly associated with total plant yield, grain yield and straw yield. However, in Breda (H1, H2, H3 and Breda overall) Factor 1, which accounted for about *38%* of the variation, was also strongly associated with total plant yield, grain yield and straw yield. It also was associated with leafiness. In season two, Factor 1, which accounted *63.7%* of the variation, was strongly influenced by total plant yield, grain yield, straw yield, TGW, vegetative duration, plant height and length of growing season (Table 6.7). All variables had positive loadings in factor 1. Factor 1 was termed "yield" factor since it consisted of total plant yield, grain yield and straw yield.

Factor 2, which accounted for about *20%* of the variation, was strongly correlated with:

- Thousand grain weight (TGW), vegetative duration and length of growing season for H2 (Tel Hadya) Table 6.6. H3 (Tel Hadya) and Tel

Hadya, it was strongly correlated with vegetative duration and length of growing season only. However it was correlated with TGW (*0.470* for H3 and -*0.506* for Tel Hadya) Tables 6.6 and 6.7.

- In Breda it was strongly correlated with vegetative duration, length of growing season and correlated with plant height for H1. The hybrids 2 and 3, factor 2 was strongly correlated with TGW and length of growing season. However, it was correlated with plant height (*-0.435* for H2 and -*0.539* for H3; Tables 6.6 and 4.7). Leafiness (*0.409*) was correlated with factor 2 for H2 (Breda). Table 6.7 indicated that factor 2 was strongly associated with vegetative duration and plant height for Breda.
- Factor 2 was strongly associated with leafiness for season two since there were only two factors (Table 6.7 season two).

Vegetative duration, length of growing season and leafiness had positive loadings with factor 2 except for vegetative duration in Breda, while plant height had a negative loadings.

Table 6.6. Factor loadings are estimated by principal components (H2, H3 in Tel Hadya and H1, H2 in Breda).

Variables	Factor loadings (Tel Hadya, H2)					Factor loadings (Tel Hadya, H3)				
	F_1	F_2	F_3	$\hat{h}_i^2$	ψ_i^2	F_1	F_2	F_3	$\hat{h}_i^2$	ψ_i^2
ln (Total plant yield)	***0.987***	*-0.024*	*0.137*	*0.994*	*0.006*	***0.989***	*-0.104*	*0.091*	*0.997*	*0.003*
ln (Grain yield)	***0.908***	*-0.048*	*0.121*	*0.841*	*0.139*	***0.938***	*-0.059*	*0.064*	*0.887*	*0.113*
ln (Straw yield)	***0.930***	*-0.009*	*0.151*	*0.887*	*0.113*	***0.937***	*-0.130*	*0.111*	*0.908*	*0.092*
TGW	*0.075*	***-0.506***	*0.191*	*0.299*	*0.701*	*-0.102*	*0.470*	***0.683***	*0.698*	*0.302*
Vegetative duration	*-0.085*	***-0.776***	*-0.107*	*0.621*	*0.379*	*-0.109*	***0.723***	*-0.165*	*0.562*	*0.438*
Plant height	*0.078*	*0.251*	***0.799***	*0.708*	*0.292*	*0.133*	*-0.329*	***0.688***	*0.600*	*0.400*
Length of growing season	*0.076*	***-0.823***	*-0.199*	*0.739*	*0.261*	*-0.062*	***0.824***	*-0.101*	*0.693*	*0.307*
Leafiness	*0.237*	*-0.162*	***0.759***	*0.658*	*0.342*	*0.203*	*-0.330*	***0.663***	*0.590*	*0.410*
Eigenvalue	*2.744*	*1.644*	*1.3578*	*Total Var %*		*3.217*	*1.670*	*1.113*	*Total Var %*	
% Variance	*34.3*	*20.5*	*17.0*	*71.8*		*40.2*	*20.9*	*18.0*	*79.1*	
Variables	Factor loadings (Breda, H1)					Factor loadings (Breda, H2)				
	F_1	F_2	F_3	$\hat{h}_i^2$	ψ_i^2	F_1	F_2	F_3	$\hat{h}_i^2$	ψ_i^2
ln (Total plant yield)	***0.965***	*-0.102*	*0.054*	*0.945*	*0.055*	***0.977***	*-0.113*	*-0.044*	*0.970*	*0.030*
ln (Grain yield)	***0.854***	*-0.077*	*0.007*	*0.735*	*0.295*	***0.904***	*-0.090*	*0.000*	*0.826*	*0.174*
ln (Straw yield)	***0.900***	*-0.104*	*0.085*	*0.827*	*0.173*	***0.921***	*-0.110*	*-0.078*	*0.867*	*0.133*
TGW	*-0.029*	*0.089*	***0.934***	*0.880*	*0.120*	*-0.019*	***0.743***	*0.063*	*0.556*	*0.444*
Vegetative duration	*-0.006*	***0.789***	*-0.080*	*0.629*	*0.371*	*0.026*	*0.064*	***0.914***	*0.841*	*0.159*
Plant height	*0.489*	*-0.418*	*0.351*	*0.537*	*0.463*	*0.362*	*-0.435*	***-0.538***	*0.610*	*0.390*
Length of growing season	*-0.023*	***0.751***	*0.421*	*0.742*	*0.258*	*-0.108*	***0.861***	*0.144*	*0.773*	*0.227*
Leafiness	***0.662***	*0.205*	*-0.193*	*0.518*	*0.482*	***0.559***	*0.409*	*-0.253*	*0.543*	*0.457*
Eigenvalue	*3.149*	*1.439*	*1.23*	*Total Var %*		*3.077*	*1.686*	*1.223*	*Total Var %*	
% Variance	*39.4*	*18.0*	*15.3*	*72.7*		*38.5*	*21.1*	*15.3*	*74.8*	

Table 6.7. Factor loadings are estimated by principal component (H3 in Breda, Tel Hadya, Breda and season 2).

Variables	Factor loadings (Breda, H3)					Factor loadings (Tel Hadya)				
	F_1	F_2	F_3	$\hat{h}_i^2$	ψ_i^2	F_1	F_2	F_3	$\hat{h}_i^2$	ψ_i^2
ln (Total plant yield)	**0.983**	-0.084	0.023	0.974	0.026	**0.989**	0.038	0.038	0.980	0.020
ln (Grain yield)	**0.904**	-0.036	0.026	0.819	0.181	**0.934**	-0.027	0.105	0.884	0.116
ln (Straw yield)	**0.881**	-0.115	0.017	0.790	0.210	**0.945**	0.096	-0.031	0.903	0.097
TGW	-0.075	**0.811**	0.055	0.666	0.334	0.047	**-0.506**	**-0.610**	0.630	0.370
Vegetative duration	-0.002	0.036	**0.962**	0.927	0.073	-0.269	**-0.716**	0.045	0.587	0.413
Plant height	0.408	**-0.539**	-0.304	0.549	0.451	0.244	0.446	**-0.630**	0.654	0.346
Length of growing season	-0.040	**0.810**	-0.046	0.660	0.340	0.178	**-0.784**	0.057	0.650	0.350
Leafiness	**0.396**	-0.125	-0.174	0.202	0.798	-0.223	0.055	**-0.656**	0.484	0.516
Eigenvalue	2.890	1.642	1.055	Total Var %		2.959	1.597	1.218	Total Var %	
% Variance	36.1	20.5	13.2	69.8		37.0	20.0	15.2	72.2	

Variables	Factor loadings (Breda)					Factor loadings (season two)			
	F_1	F_2	F_3	$\hat{h}_i^2$	ψ_i^2	F_1	F_2	$\hat{h}_i^2$	ψ_i^2
ln (Total plant yield)	**0.957**	0.204	-0.007	0.958	0.042	**0.919**	0.242	0.903	0.097
ln (Grain yield)	**0.876**	0.171	-0.030	0.798	0.202	**0.907**	0.213	0.869	0.131
ln (Straw yield)	**0.893**	0.204	0.010	0.838	0.162	**0.899**	0.262	0.877	0.123
TGW	0.068	0.223	**-0.766**	0.640	0.360	**0.731**	0.181	0.568	0.432
Vegetative duration	0.039	**-0.824**	-0.153	0.704	0.296	**0.874**	0.261	0.832	0.168
Plant height	0.313	**0.710**	-0.204	0.644	0.356	**0.833**	0.322	0.797	0.203
Length of growing season	-0.160	-0.302	**-0.640**	0.526	0.474	**0.771**	-0.059	0.597	0.403
Leafiness	**0.602**	-0.246	0.140	0.443	0.557	0.181	**0.963**	0.959	0.041
Eigenvalue	2.973	1.498	1.081	Total Var %		5.094	1.307	Total Var %	
% Variance	37.2	18.7	13.5	66.4		63.7	16.3	80.0	

Factor 2 was regarded as a "plant period" factor since it consisted of days to vegetative duration and length of growing season.

Plant height and leafiness in Tel Hadya (H2, H3 and Tel Hadya) were strongly associated with Factor 3, which accounted for about *17%* of the variation. Factor 3 was strongly correlated with TGW for H3 (Tel Hadya) and Tel Hadya Tables 6.6 and 6.7. In Breda, Factor 3, which accounted for about *15%* of the variation, was strongly associated with vegetative duration and plant height. All variables had positive loadings in factor 3 (H2 and H3 in Tel Hadya) while variables had negative loadings in factor 3 in Tel Hadya. Factor 3 was termed "morphology" factor since it consisted of plant height and leafiness.

The total variance for total plant yield, explained by factors, was indicated in Tables 6.6 and 6.7. The three factors accounted between *94.5%* (H1 in Breda) and *99.7%* (H3 in Tel Hadya), but most of the variance was explained by the first factor (yield). *90.3%* of the variance in total plant yield in the second season was explicable because there were only two factors (Table 6.7). Also more than *80%* of the variance for grain yield and straw yield in season two were explicable for the same reason (two factors). Most of the variance is explained by the first factor (yield).

The variance for TGW and leafiness, which are explained by three factors, were not stable. For example, in H2 (Tel Hadya), there was only *29.9%* of variance which was explained by three factors (Table 6.6), while for H1

(Breda) *88%* of variance for TGW was explained by three factors (Table 6.6). Also for H2 (Breda), the total variance for leafiness, explained by three factors, was *20.2%* (Table 6.7), while in season two there was *97.8%* of variance for leafiness was explained by three factors (Table 6.7).

The total variance explained for: vegetative duration was between *56.2%* H3 (Tel Hadya) and *92.7 %* H3 (Breda), length of growing season was between *52.6%* Breda and *80.6%* season two. Finally, the total variance for plant height, explained by three factors, was between *53.7%* H1 (Breda) and *79.9%* season two (Tables 6.6 and 6.7).

Generally, the relationship between yields (total plant yield, grain yield and straw yield) and morphological characters (vegetative duration, plant height, length of growing season and leafiness) was not significant because there are no significant correlations among them, except in the second season. TGW was related to vegetative duration and length of growing season (Tables 6.6 and 6.7)

In the second season, vegetative duration, plant height and length of growing season are accepted as the most important characters due to their close relationship with total plant yield, grain yield, straw yield and TGW (Table 6.7).

Residual: in Tel Hadya, there was about *25%* of covariance between vegetative duration and length of growing season (H1) , *25%* of covariance between TGW and vegetative duration (H2) and *28%* of covariance between plant height and leafiness (Tel Hadya, H1 and H2). These results are indicated by Tables 6.8. and 6.9. However, the residual matrix was small in magnitude. In Breda, there was about *24%* of covariance between vegetative duration and plant height (H1 and H2) while there was about *25%* of covariance between TGW and length of growing season (H3). The residual matrix for season two was very small in magnitude (Table 6.9).

Table 6.8. The residual matrix corresponding for factors (H2, H3 in Tel Hadya and H1, H2 in Breda).

Tel Hadya (H2)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0							
x_2	0.003	0						
x_3	0.009	−0.193	0					
x_4	−0.031	−0.058	−0.005	0				
x_5	0.042	0.023	0.063	−0.179	0			
x_6	−0.028	0.006	−0.054	0.094	−0.018	0		
x_7	−0.028	−0.014	−0.039	−0.181	−0.251	0.118	0	
x_8	0.027	0.041	0.004	−0.182	0.023	−0.282	−0.023	0

Tel Hadya (H3)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0							
x_2	−0.021	0						
x_3	0.021	−0.130	0					
x_4	−0.010	0.044	−0.044	0				
x_5	0.029	−0.003	0.045	−0.254	0			
x_6	0.020	0.010	0.024	−0.057	0.118	0		
x_7	−0.010	−0.024	0.000	−0.171	−0.153	0.063	0	
x_8	−0.019	−0.024	−0.013	−0.194	−0.061	−0.270	0.029	0

Breda (H1)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0							
x_2	0.055	0						
x_3	0.048	−0.117	0					
x_4	−0.008	0.012	−0.022	0				
x_5	−0.023	−0.007	−0.037	0.033	0			
x_6	−0.108	−0.122	−0.088	−0.095	0.247	0		
x_7	0.022	0.005	0.035	−0.131	−0.240	−0.054	0	
x_8	−0.137	−0.174	−0.089	0.097	−0.092	0.088	−0.046	0

Breda (H2)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0							
x_2	0.024	0						
x_3	0.032	−0.100	0					
x_4	0.012	0.083	−0.043	0				
x_5	−0.061	−0.068	−0.047	0.033	0			
x_6	−0.057	−0.034	−0.065	0.160	0.209	0		
x_7	0.023	−0.012	0.048	−0.199	−0.026	0.020	0	
x_8	−0.099	−0.133	−0.061	−0.0213	0.148	0.086	−0.074	0

Where : x_1 is a total plant yield, x_2 is a grain yield, x_3 is a straw yield, x_4 is a TGW, x_5 is a vegetative duration, x_6 plant height, x_7 length of growing season and x_8 leafiness.

Table 6.9. The residual matrix corresponding for factors (H3 Breda; Tel Hadya; Breda and season 2).

Breda (H3)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0							
x_2	0.006	0						
x_3	0.039	−0.155	0					
x_4	−0.013	−0.007	−0.017	0				
x_5	−0.041	−0.020	−0.052	0.008	0			
x_6	−0.057	−0.010	−0.090	0.013	0.133	0		
x_7	−0.032	−0.023	−0.033	−0.254	0.085	0.136	0	
x_8	−0.113	−0.077	−0.123	−0.011	0.135	−0.061	0.054	0

Tel Hadya

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0							
x_2	0.018	0						
x_3	0.019	−0.070	0					
x_4	−0.030	−0.015	−0.040	0				
x_5	0.054	0.041	0.064	−0.183	0			
x_6	−0.048	−0.061	−0.031	0.091	0.094	0		
x_7	−0.050	−0.068	−0.031	−0.149	−0.190	0.148	0	
x_8	0.076	0.092	0.055	−0.284	0.099	−0.239	0.001	0

Breda

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0							
x_2	0.033	0						
x_3	0.043	−0.109	0					
x_4	−0.042	−0.025	−0.045	0				
x_5	−0.002	0.008	−0.012	0.037	0			
x_6	−0.055	−0.061	−0.046	−0.156	0.187	0		
x_7	0.036	0.003	0.055	−0.364	−0.211	0.048	0	
x_8	−0.137	−0.151	−0.106	0.110	−0.167	0.077	0.012	0

Season two

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	0							
x_2	0.095	0						
x_3	0.096	0.060	0					
x_4	−0.115	−0.109	−0.115	0				
x_5	−0.070	−0.064	−0.072	0.003	0			
x_6	−0.031	−0.047	−0.017	−0.011	0.000	0		
x_7	−0.111	−0.106	−0.113	−0.004	0.048	−0.085	0	
x_8	−0.027	−0.022	−0.031	0.006	0.008	−0.048	0.123	0

Where : x_1 is a total plant yield, x_2 is a grain yield, x_3 is a straw yield, x_4 is a TGW, x_5 is a vegetative duration, x_6 plant height, x_7 length of growing season and x_8 leafiness.

6.4 Conclusions and summary

The conclusion will be given in two sections. The first section will summarize the results for each hybrid and each area. The second will give the conclusion for season two.

Three factors in each hybrid (H1, H2 and H3) and each area (Tel Hadya and Breda) accounted for about *73 %* of the total variability. The communalities were high for total plant yield, grain yield and straw yield (*>0.85*). Tables 6.6 and 6.7 indicated that there were three factors which explained eight variables (total plant yield, grain yield, straw yield, TGW, vegetative duration, plant height, length of growing season and leafiness). The most important factor in Tel Hadya (H1, H2, H3 and Tel Hadya) was the first factor (yield) since it consisted of total plant yield, grain yield and straw yield. However, In Breda (H1, H2, H3 and Breda), the variability was nearly the same (about *38%*); this factor also included the leafiness parameter. The second most important factor contained three variables (TGW, vegetative duration and length of growing season) in Tel Hadya (H1, H2, H3 and Tel Hadya) and accounted for *20.99%, 20.5%, 20.90%* and *20.0%*, respectively, of the total variability. There appeared to be no obvious factor equivalent to this one in the Breda data (H1, H2, H3 and Breda). However, this factor accounted for about *20%* of the total variability. The summary is given in Table 6.10 and 6.11.

Table 6.10. The summary for the total variability (Tel Hadya).

H1 (Tel Hadya)			H2 (Tel Hadya)		
Factor	% Total Variability	Variables	Factor	% Total Variability	Variables
1	33.1	Total plant yield, grain yield and straw yield	1	34.3	Total plant yield, grain yield and straw yield
2	21.0	TGW, vegetative duration and length of growing season.	2	20.5	TGW, vegetative duration and length of growing season
3	16.4	Plant height and leafiness.	3	17.0	Plant height and leafiness.
	70.5			71.8	
H3 (Tel Hadya)			**Tel Hadya**		
1	40.2	Total plant yield, grain yield and straw yield	1	37.0	Total plant yield, grain yield and straw yield
2	20.9	Vegetative duration and length of growing season	2	20.0	TGW*, vegetative duration and length of growing season
3	18.0	Plant height and leafiness.	3	15.2	TGW*, plant height and leafiness.
	79.1			72.2	

* Variables so marked have loadings greater than 0.50 in more than one factor.

Table 6.11. The summary for the total variability (Breda).

		H1 (Breda)			H2 (Breda)
1	39.4	Total plant yield, grain yield, straw yield and leafiness	1	38.5	Total plant yield, grain yield, straw yield and leafiness
2	18.0	Vegetative duration and length of growing season	2	21.1	TGW and length of growing season
3	15.3	TGW	3	15.3	Height plant and vegetative duration
	72.7			74.9	
		H3 (Breda)			Breda
1	36.1	Total plant yield, grain yield and straw yield	1	37.2	Total plant yield, grain yield, straw yield and leafiness
2	20.5	TGW, plant height and length of growing season	2	18.7	Vegetative duration and plant height
3	13.2	Vegetative duration	3	13.5	TGW and length of growing season
	69.8			66.4	

In season two, *80 %* of the total variation was accounted by common factors. The first two factors accounted for *63.7%* and *16.3%* of the variation after rotation, respectively (Table 6.7). The first factor consisted of all variables except leafiness. Ideally, the major variables in the first factor would show high positive loading values (it was between *0.731 for length of growing season* and *0.919 for total plant yield*).

Finally, Season Two reveals a relationship between yield parameters and morphological characters while this relationship for each hybrid and area did not appear for certain reasons, (for example, insufficient size of simple of data in each hybrid or area)

Chapter 7 will study the relationship between yield parameters and morphological characters.

Chapter 7

Using multiple regression analysis to study the relationship between yield parameters and morphological characters

7.1 Introduction

Crop yield is a product of the expression of the genotype within an environment ($G \times E$), that is to say, set of environment and morphological characters. Frequently, only simple relationships between these factors and yield are reported in the literature, but often these factors may interact. A stepwise multiple regression may provide a more applicable approach. viz:

1. *Stepwise. multiple regression.* Using this method, the research will aim to determine whether there is a relationship between the morphological characters and the yield parameters.
2. In this research, the effectiveness of *multiple exponential regression* in agricultural analyses will be compared with *multiple linear regression*, especially for studying the relationship between morphological characters and yield parameters in experiments involving barley breeding.

Multjple regression analysis is frequently used to study the effect of sets of independent variables X's (morphologies of barley (leafiness -number of leaves- length of growing season, plant height, vegetative duration) on one variable yield parameters of barley (total plant yield, grain yield and straw yield)) which are defined as a function. In general, stepwise programmes are designed to maximise the coefficient of determination with a minimum number of independent variables. However, this may not succeed very well in practice because it depends on many factors. For example, the significance of independent coefficients and the significance of equation (Cohen and Cohen, 1975). This research will use forward stepwise regression to study this effect in two ways: stepwise linear regression equations and stepwise exponential regression equations.

Exponential functions are frequently found in biological relationships[1], for example, in the study of the relationship between leaf appearance rate (LAR) and temperature (Xue, *et al.*, 2004) Also exponential equation has been chosen because it is very close to linear equation.

By means of forward stepwise regression the correlation coefficients will be found between the independent variables and the function. Then the independent variable which has the strongest correlation coefficients with y will be chosen to find the equations (linear and exponential). Subsequently, F-tests will be used to study the significance of these equations and t-test to study the significance of correlation coefficients and also regression coefficients (Allison, 1999). When the equations (linear and exponential) are used, the independent variable will be kept and a new independent variable (which has a strong correlation with y) will be added, and the same methods will be repeated until the multiple regression equations between the independent variables (morphologies) and the function (yield parameters) are

[1] Exponential relationships occur in situations where the absolute value of variable changes at a rate that is proportional to its instantaneous value (Burton, 1998).

optimised (depending on, determination coefficient, the significance of equation, the significance of the independent variables and the residuals). In this section, at every step, linear equation and exponential equation will be compared to determine which equation is better.

7.2 Relationship between yield parameters and morphological characters in season two:

Before starting to use stepwise multiple regression, the correlation coefficients between morphological characters and yield parameters were calculated (Table 7.1).

Table 7.1. Matrix of correlation coefficients

$R =$	Total plant yield	Grain yield	Straw yield
Vegetative duration	0.750**	0.726**	0.740**
Plant height	0.767**	0.720**	0.778**
Length of growing season	0.571**	0.564**	0.553**
Leafiness	0.307**	0.275**	0.324**

According to the correlation matrix (above), there are positive correlations between morphological characters and yield parameters.

7.2.1 The relationship between total plant yield and morphological characters

According to the R matrix:

$$\left|R_{(Total\ yield.Plant\ height)}\right| > \left|R_{(Total\ yield.Vegetative\ duration)}\right| > \left|R_{(Total\ yield.length\ of\ growing\ season)}\right| > \left|R_{(Total\ yield.Leafiness)}\right|$$ [2]

[2] $|R|$ is Abstract value of R.

So the relationship between morphological characters and yield parameters will be studied as: firstly, total plant yield as a function of plant height as an independent variable; secondly, total plant yield as a function of vegetative duration and plant height as independent variables; and thirdly, total plant yield as a function of length of growing season, vegetative duration and plant height as independent variables. Finally, total plant yield as a function of morphological characters as independent variables.

7.2.1.1 The relationship between the total plant yield and the plant height

The linear relationship between total plant yield (*y*) and plant height (*x*) is shown in Figure 7.1. There is a significant positive correlation between the total plant yield and plant height (Figure 7.1). The regression equation is:

Total plant yield = - 4509 + 176 plant height

Estimated gradient, also known as the regression coefficient, $\hat{b} = 176$, there is an increase of *176* (kg ha^{-1}) in the total plant yield for each increase of one centimetre.

The plant height was significant in this equation ($P<0.001$)[3]. The constant in this equation also is significant ($P<0.001$).

Coefficient of determination: (R^2)

The coefficient of determination tells us how much of the variation in the dependent variable y is explained by the equation. In other words, R^2 is greater than *0.85* means the data fit the equation well, and R^2 near zero means

[3] The estimated of t is given as follows:

$$t = |b \div SE_b| = |208.96 \div 46.18| = 4.52$$

The plant height is a significant variant affecting the total plant yield ($t=4.52$; $P<0.05$).

a poor fit. However, R^2 does not indicate the significance of the regression (Phillips, 2002).

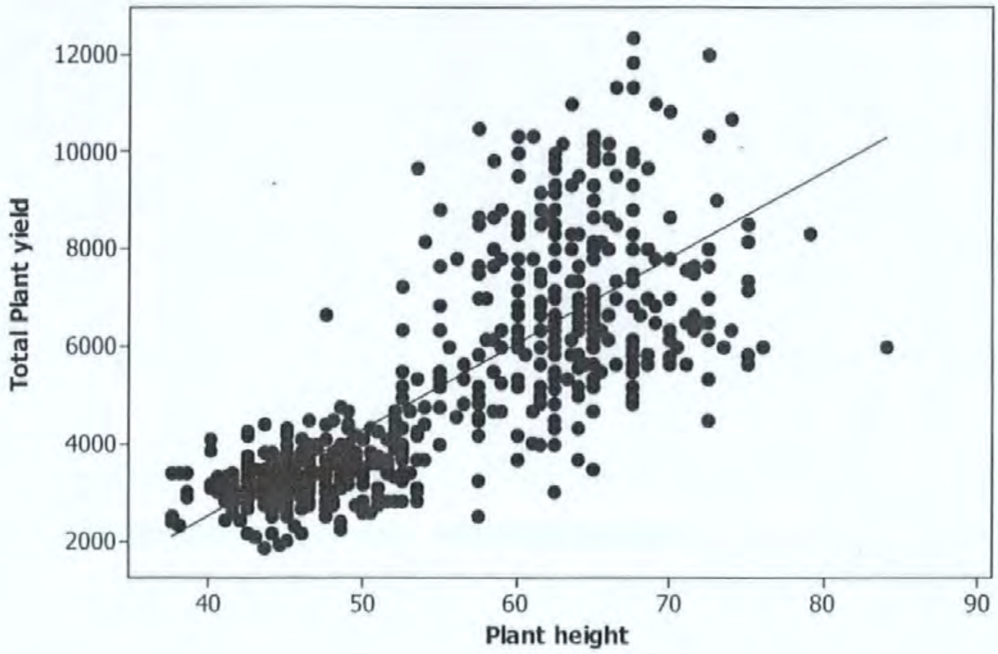


Figure 7.1. The linear relationship between the total plant yield and plant height[4].

[4] Figure 7.1 shows that the goodness of fit appears to be better at low plant heights than at high plant heights. In fact, lower plant heights are from Breda and highest group is from Tel Hadya). The relationship between the yield parameters and morphological characters in each area will be analysed later in this chapter.

The coefficient of determination was $R^2\% = 58.8\%$ $(P<0.001)$[5]. It indicated that *59%* of the variation in the total plant yield is explained by the equation.

F-test determines whether the equation is significant or not. It is given by the equation:

$$F = \frac{MS_{regression}}{MS_{error}} = \frac{1819230874}{2087268} = 871.58$$

The equation is significant ($F_{1,610} = 871.58; P < 0.001$).

Exponential Equation:

The exponential relationship between *X* and *Y* is shown in Figure 7.2

The exponential equation ($Y = ae^{bX}$) and the summary can be given as:

The regression equation is

$$Total\ Yield = 706.2717\, e^{0.0347\ Plant\ height}$$

P-value is less than *0.001*, so the equation is significant. The coefficient of determination in exponential equation (*68.8%, P<0.001*) was better than the coefficient of determination in linear equation.

The plant height and constant were significant in this equation (*P<0.001*).

[5] This determines whether the correlation coefficient is significant or not using the following equation. It is given by:

$$t = \frac{|R|\sqrt{N-M}}{\sqrt{1-R^2}} = \frac{0.767\sqrt{1612-2}}{\sqrt{1-0.588}} = 45.97$$

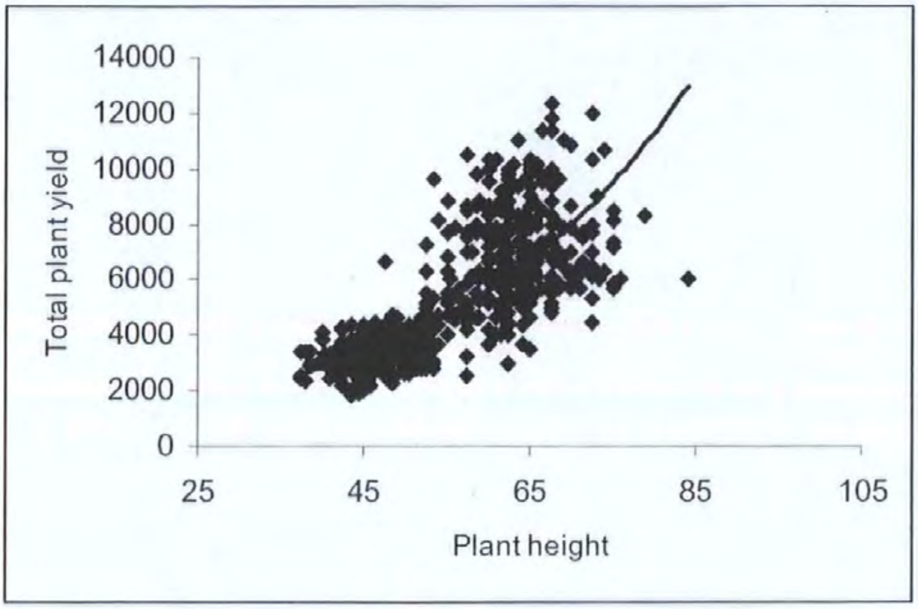


Figure 7.2. The exponential relationship between the total plant yield and plant height there are two groups (see footnote 2 page 6).

Unusual observations (Linear, Exponential):

Large residuals:

In the linear equation there are *42* points from *612* points that are more than two standardised residuals away from the expected value, but in the exponential equation, *28* points are more than two standardised residuals away from the expected value.

Influential points:

In the linear and exponential equations, there are two points that are identified as being particularly influential.

In conclusion:

- The total plant yield has been strongly affected by plant height.
- The exponential equation was better than linear equation because the coefficient of determination was higher and the residual error in the exponential equation was lower than the residual error in the linear equations.

7.2.1.2 The relationship between the total plant yield and morphological characters (season two)

The results for stepwise multiple (linear and exponential) regression analysis are summarized in Table 7.2.

Table 7.2 indicated that there was a significant relationship between total plant yield and morphological characters. There were *64%* and *72%* of the variation in the total plant yield is explained by the linear equation and exponential equation respectively.

In linear and exponential equations, the morphological characters variables except leafiness were highly positively correlated with total plant yield (Table 7.2). The coefficient of determination increased slightly (Table 7.2); since in linear equation, the coefficient of determination was *58.8%* (total plant yield and height plant) then increased to be *63.6%* (total plant yield and plant height and vegetative duration). When length of growing season and leafiness were taken into account in the linear regression, the coefficient of determination did not improve more than *1%*. The coefficient of determination using the exponential equation was about *65%* (total plant yield and plant height) increased until about *72%* (total plant yield with plant height, vegetative duration and length of growing season) since the variable leafiness did not improve the coefficient of determination (Table 7.2).

Table 7.2. Stepwise multiple regression analysis results between total plant yield and morphological characters using the linear and exponential equations (n=612).

Linear equations ($y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$)

Step	**1**	**2**	**3**	**4**
Constant	-4509***	-15649***	-26690***	-26021***
Plant height	176.4***	106.7***	109.4***	111.2***
Vegetative duration		157***	119***	125***
Length of growing season			111**	105**
Leafiness				-142
R^2%	58.83***	63.60***	64.27***	64.38***
R	0.77	0.80	0.80	0.80
P (equation significant)	<0.0001	<0.0001	<0.0001	<0.0001

Exponential equations ($y = a_0e^{a_1x_1+a_2x_2+a_3x_3+a_4x_4}$)

Step	**1**	**2**	**3**	**4**
Constant	706.27***	75.11***	2.99***	19.11***
Plant height	0.037***	0.0206***	0.0210***	0.0209***
Vegetative duration		0.0316***	0.0270***	0.0266***
Length of growing season			0.0133**	0.0136**
Leafiness				0.0084
R^2%	65.84***	71.34***	71.71***	71.71***
R	0.81	0.84	0.85	0.85
P (equation significant)	<0.0001	<0.0001	<0.0001	<0.0001

P<0.001 ***, P<0.01 **, P<0.05 *.

y **: total plant yield,** x_1 **: plant height,** x_2 **: vegetative duration,** x_3 **: length of growing season,** x_4 **: leafiness.**

In the linear equation there are forty two points that are more than two standardised residuals away from the expected value, but in the exponential equation, thirty two points are more than two standardised residuals away from the expected value.

In the linear and exponential equations, there are nine points that are identified as being particularly influential.

Tables 7.3 and 7.4 show the analysis of variance for the relationship between total plant yield and morphological characters (linear and exponential respectively). Tables 7.3 and 7.4 show that *F=0.72, P-Value=0.88* and *F=1.1, P-Value= 0.42* respectively; these determine whether the residuals follow a normal distribution (Pure error lack of fit test; Minitab Inc.). Test of hypotheses is given by:

H_0: the residuals follow the normal distribution.

H_1: the residuals do not follow the normal distribution.

P-Value for linear and exponential equations were greater than *0.05*, so the null hypothesis cannot be rejected. That is, the residuals follow a normal distribution.

Table 7.3. Analysis of Variance for the relationship between total plant yield and morphological characters (linear).

Source	*DF*	*SS*	*MS*	*F*	*P*
Regression	*4*	*1991009847*	*497752462*	*274.31*	*0.0001*
Residual Error	*607*	*1101454503*	*1814587*		
Lack of Fit	*587*	*1051554276*	*1791404*	*0.72*	*0.882*
Pure Error	*20*	*49900227*	*2495011*		
Total	*611*	*3092464350*			

Table 7.4. Analysis of Variance for the relationship between total plant yield and morphological characters (exponential).

Source	*DF*	*SS*	*MS*	*F*	*P*
Regression	*4*	*76.518*	*19.130*	*384.91*	*0.0001*
Residual Error	*607*	*30.167*	*0.050*		
Lack of Fit	*587*	*29.264*	*0.050*	*1.10*	*0.419*
Pure Error	*20*	*0.903*	*0.045*		
Total	*611*	*106.685*			

In conclusion:

- ✓ The total plant yield had a significant relationship with morphological characters.
- ✓ The increase in coefficient of determination from two to three variables was small and R^2 is close to the maximum of $R^2\% = 63.60$ for linear regression and $R^2\% = 71.34$ for exponential equation. Thus, the R^2 criterion leads to the choice of the two variable subsets containing plant height and vegetative duration.
- ✓ The total plant yield was not significantly affected by leafiness.
- ✓ The stepwise exponential equation was better than stepwise linear equation for studying the relationship between total plant yield and morphological characters because:
    - The coefficient of determination in exponential equations was higher than the coefficient of determination in linear equation in each step.
    - Residual errors in exponential regression are lower than residual errors in linear regressions.

7.2.2 The relationship between grain yield and morphological characters (season two)

Stepwise multiple regression was computed, using grain yield as dependent variable and morphological characters as independent. The variables (plant height and vegetative duration and length of growing season) were accepted by the stepwise multiple regression in the linear and exponential model with coefficients of determination (R^2) value *58.55* and *67.69* respectively (Table 7.5). Table 7.5 indicated that *58.8 %* and *67.71%* of the variation in the grain yield is explained by the linear equation and exponential equation respectively. Grain yield was affected by vegetative duration more than plant height because the simple correlation between vegetative duration and grain yield was higher than the simple correlation between grain yield and plant height. All morphological characters, except leafiness, had large positive indirect affects on grain yield (linear and exponential) (Table 7.5).

The coefficient of determination was close to maximum of $R^2\% = 57.73$ for linear regression and $R^2\% = 67.35$ for exponential equation with two variables. Thus, the R^2 criterion leads to the choice of the two variable subsets containing plant height and vegetative duration. The coefficient of determination for exponential regression was higher than the coefficient of determination for linear regression in each step (Table 7.5). For example, the coefficient of determination between grain yield and vegetative duration was *52.77%* for linear equation while it was *61.55%* for exponential equations.

Tables 7.6 and 7.7 show that the residual for linear and exponential equations follow a normal distribution (*P=0.892 and 0.410*, respectively).

There were forty points which were more than two standardised residuals away from the expected value for exponential equation, while in the linear equation there were thirty eight points.

In the linear and exponential equations, there are ten points that are identified as being particularly influential.

Table 7.5. Stepwise multiple regression analysis results between grain plant yield and morphological characters using the linear and exponential equations (n=612).

Linear equations ($y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$)

Step	1	2	3	4
Constant	-11752***	-8048***	-14113***	-13621***
Vegetative duration	149.4***	85.8***	64.7***	45.7***
Plant height		42.9***	44.4***	69.0***
Length of growing season			61.0***	56.4**
Leafiness				-104
R^2%	52.77***	57.73***	58.58***	58.8***
R	0.73	0.76	0.77	0.77
P (equation significant)	<0.0001	<0.0001	<0.0001	<0.0001

Exponential equations ($y = a_0e^{a_1x_1+a_2x_2+a_3x_3+a_4x_4}$)

Step	1	2	3	4
Constant	6.79***	30.81***	7.19***	19.11***
Vegetative duration	0.0610***	0.0350***	0.0300***	0.0301***
Plant height		0.0175***	0.0179***	0.0179***
Length of growing season			0.0146***	0.0144**
Leafiness				-0.0038
R^2	61.55***	67.35***	67.69***	67.71***
R	0.78	0.82	0.82	0.82
P (equation significant)	<0.0001	<0.0001	<0.0001	<0.0001

P<0.001 ***, P<0.01 **, P<0.05 *.

y : grain yield, x_1**: vegetative duration,** x_2**: plant height,** x_3**: length of growing season,** x_4**: leafiness.**

Table 7.6. Analysis of variance for the relationship between grain yield and morphological characters (linear).

Source	DF	SS	MS	F	P
Regression	*4*	*435239489*	*108809872*	*216.91*	*0.0001*
Residual Error	*607*	*304495600*	*501640*		
Lack of Fit	*587*	*290494113*	*494879*	*0.71*	*0.892*
Pure Error	*20*	*14001487*	*700074*		
Total	*611*	*739735089*			

Table 7.7. Analysis of variance for the relationship between grain yield and morphological characters (exponential).

Source	DF	SS	MS	F	P
Regression	*4*	*71.593*	*17.898*	*317.94*	*0.0001*
Residual Error	*607*	*34.170*	*0.056*		
Lack of Fit	*587*	*33.155*	*0.056*	*1.11*	*0.410*
Pure Error	*20*	*1.015*	*0.051*		
Total	*611*	*105.763*			

In conclusion:

- ✓ The grain yield had a significant relationship with morphological characters except leafiness.
- ✓ The increase in coefficient of determination from two to three variables was small. Thus, the R^2 criterion leads to the choice of the two variable subsets containing plant height and vegetative duration.
- ✓ The stepwise exponential equation was better than stepwise linear equation for studying the relationship between grain yield and morphological characters because the coefficient of determination in exponential equations was higher than the coefficient of determination in linear equation in each step.

7.2.3 The relationship between straw yield and morphological characters (season two)

Stepwise multiple regression was computed using straw yield as a dependent variable and morphological characters as independent variables. The best single variable was plant height which gave $R^2\% = 60.55\%$ and $R^2\% = 66.54\%$ for linear and exponential equations respectively. Furthermore, the relationship between straw yield and plant height was significant ($P<0.001$) (Table 7.8). The second best independent variable was vegetative duration which increased the coefficient of determination to *64.00%* for linear regression and *70.76%* for exponential regression. The relationship between straw yield and both plant height and vegetative duration was significant ($P<0.001$). Table 7.8 indicted that all the morphological characters variables were positive and significant except leafiness (*P>0.05*). The exponential equation was better than the linear equation to describe the relationship between straw yield and morphological characters because the coefficient of determination was higher in each step (Table 7.8). For example, the coefficient of determination for straw yield with plant height and vegetative duration was *64.00%* for linear equation while it was *70.76%* for exponential equations. The analysis of variance for the relationship between straw yield and morphological characters is given by Table 7.9 (linear) and Table 7.10 (exponential).

Tables 7.9 and 7.10 indicated that the residuals follow a normal distribution (*P>0.05*).

There were thirty eight points which were more than two standardised residuals away from the expected value for exponential equation, while in the linear equation there were thirty six points.

In the linear and exponential equations there were nine points that were identified as being particularly influential.

Table 7.8. Stepwise multiple regression analysis results between straw yield and morphological characters using the linear and exponential equations (n=612).

Linear equations ($y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$)

Step	1	2	3	4
Constant	-2545***	-7601***	-12578***	-12401***
Plant height	95.4***	63.8***	65.1***	65.5***
vegetative duration		71.3***	53.9***	55.5***
Length of growing season			50.0**	48.3**
Leafiness				-37.5
R^2%	60.55***	64.00***	64.48***	64.5***
R	0.78	0.80	0.80	0.80
P (equation significant)	<0.0001	<0.0001	<0.0001	<0.0001

Exponential equations ($y = a_0e^{a_1x_1+a_2x_2+a_3x_3+a_4x_4}$)

Step	1	2	3	4
Constant	339.00***	45.20***	14.21***	12.81***
Plant height	0.0360***	0.0234***	0.0237***	0.0235***
vegetative duration		0.0284***	0.0244***	0.0235***
Length of growing season			0.0116**	0.0125*
Leafiness				0.0210
R^2%	66.54***	70.76***	70.96***	71.00***
R	0.82	0.84	0.84	0.84
P (equation significant)	<0.0001	<0.0001	<0.0001	<0.0001

P<0.001 ***, P<0.01 **, P<0.05 *.

y : straw yield, x_1 : plant height, x_2 : vegetative duration, x_3 : length of growing season, x_4 : leafiness.

Table 7. 9. Analysis of Variance for the relationship between straw yield and morphological characters (linear).

Source	*DF*	*SS*	*MS*	*F*	*P*
Regression	*4*	*567531559*	*141882890*	*275.84*	*0.0001*
Residual Error	*607*	*312214922*	*514357*		
Lack of Fit	*587*	*299237477*	*509774*	*0.79*	*0.810*
Pure Error	*20*	*12977445*	*648872*		
Total	*611*	*879746481*			

Table 7.10. Analysis of Variance for the relationship between straw yield and morphological characters (exponential).

Source	*DF*	*SS*	*MS*	*F*	*P*
Regression	*4*	*81.124*	*20.281*	*372.01*	*0.0001*
Residual Error	*607*	*33.092*	*0.055*		
Lack of Fit	*587*	*32.074*	*0.055*	*1.07*	*0.454*
Pure Error	*20*	*1.019*	*0.051*		
Total	*611*	*114.216*			

In conclusion:

- ✓ The straw yield had a significant relationship with morphological characters except leafiness.
- ✓ The best one-variable (plant height) subset accounted for $R^2 = 60.55\%$ (linear equation) and $R^2 = 66.54\%$ (exponential equation) of the variation in straw yield, the best two-variable (plant height and vegetative duration) subset accounted for $R^2 = 64.00\%$ (linear equation) and $R^2 = 70.76\%$ (exponential equation) and the best three-variable (plant height, vegetative duration and length of growing season) subset accounted for $R^2 = 64.48\%$ (linear equation) and $R^2 = 70.96\%$ (exponential equation). The increase in coefficient of determination from two to three variables was small. Thus, straw yield was affected by plant height and vegetative duration more than other variables.
- ✓ The stepwise exponential equation was better than stepwise linear equation.

Conclusion and summary for stepwise multiple regression:

The yield parameters had a significant relationship with morphological characters. Yield parameters were not affected by leafiness ($P>0.05$) but plant height and vegetative duration had a strong influence on yield parameters. The coefficients of determination were very strong and significant. However, the increase in coefficients of determination from two to three independent variables was small. The stepwise exponential equations were better than stepwise linear equations for studying the relationship between yield parameters and morphological characters.

7.3 Using multiple regression analysis to study the relationship between yield parameters and morphological characters

This section reviews the multiple regression results for the linear and exponential model in one dependent variable (yield parameters) with four independent variables (morphological characters). The aims of this section are to see whether or not the relationships between the yield parameters and morphological character for each hybrid (H1, H2 and H3) in each area (Tel Hadya and Breda) were stable. Also this section aims to compare linear and exponential regression analysis for the relationship between yield parameters and morphological characters. The section is divided into three parts. Part one will study the relationship between total plant yield and morphological characters (plant height, vegetative duration, length of growing season and leafiness). The second part will study the relationship between grain yield and morphological characters. Finally, in part three, the relationship between straw yield and morphological characters will be studied.

7.3.1 Using multiple regression analysis to study the relationship between total plant yield and morphological characters

The results for multiple (linear and exponential) regression analysis for the relationship between total plant yield and morphological characters are summarized by Table 7.11.

Table 7.11. The relationship between total plant yield (y) and the variables: vegetative duration (x_1), plant height (x_2), length of growing season (x_3) and leafiness (x_4).

Area	Hybrid	Equation	$P_{constant}$	P_{a_1}	P_{a_2}	P_{a_3}	P_{a_4}	$P_{equation}$	$R^2\%$	P_{R^2}
Tel Hadya	1	$y = 14156 - 139x_1 + 80.6x_2 + 12.6x_3 + 471x_4$	0.18	0.09	**0.03**	0.81	0.11	**<0.001**	14.4	**<0.05**
		$y = 14472.4e^{-0.0152x_1 + 0.00869x_2 + 0.00177x_3 + 0.0574x_4}$	**<0.001**	0.11	**0.04**	0.77	0.09	**<0.001**	13.4	**<0.05**
	2	$y = -6605 - 32.9x_1 + 37.8x_2 + 87.2x_3 + 562x_4$	0.54	0.73	0.15	0.27	**0.02**	**0.02**	11.9	**<0.05**
		$y = 1422.3e^{-0.0115x_1 + 0.0054x_2 - 0.0144x_3 + 0.0921x_4}$	**<0.001**	0.44	0.18	0.24	**<0.001**	**<0.001**	12.8	**<0.05**
	3	$y = 13266 - 75.7x_1 + 25.3x_2 - 26.2x_3 + 454x_4$	0.17	0.30	0.33	0.69	0.09	**0.03**	10.5	**<0.05**
		$y = 22026e^{-0.0138x_1 + 0.0043x_2 - 0.0045x_3 + 0.0809x_4}$	**<0.001**	0.29	0.34	0.70	0.08	**0.03**	10.6	**<0.05**
Breda	1	$y = 3116 - 83.4x_1 + 68.6x_2 + 22.7x_3 + 430x_4$	0.63	0.17	**<0.001**	0.55	**<0.001**	**<0.001**	33.7	**<0.001**
		$y = 3428.9e^{-0.0275x_1 + 0.0204x_2 + 0.0074x_3 + 0.137x_4}$	**<0.001**	0.151	**<0.001**	0.53	**<0.001**	**<0.001**	33.5	**<0.001**
	2	$y = 15301 + 38.1x_1 + 51.7x_2 - 153x_3 + 576x_4$	0.27	0.58	**0.03**	0.10	**<0.001**	**<0.001**	29.1	**<0.001**
		$y = 80821e^{0.0067x_1 + 0.0161x_2 - 0.0391x_3 + 0.159x_4}$	**<0.01**	0.74	**0.03**	0.15	**<0.001**	**<0.001**	27.9	**<0.001**
	3	$y = 828 + 24.5x_1 + 56.6x_2 - 21.7x_3 + 174x_4$	0.92	0.64	**<0.001**	0.67	**0.02**	**<0.001**	19.1	**<0.001**
		$y = 1199.9e^{0.0078x_1 + 0.0174x_2 - 0.0049x_3 + 0.0498x_4}$	**<0.001**	0.62	**<0.001**	0.75	0.75	**0.03**	19.0	**<0.001**
Tel Hadya		$y = 7584 - 235x_1 + 55.8x_2 + 161x_3 - 496x_4$	0.352	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	12.8	**<0.05**
		$y = 8955.3e^{-0.0352x_1 + 0.0839x_2 + 0.0231x_3 - 0.0657x_4}$	**<0.001**	**<0.001**	**0.003**	**0.001**	**0.008**	**<0.001**	13.0	**<0.05**
Breda		$y = 5007 + 2.0x_1 + 54.1x_2 - 44.2x_3 + 392x_4$	0.228	0.95	**<0.001**	**0.045**	**<0.001**	**<0.001**	30.0	**<0.001**
		$y = 7555.3e^{-0.0009x_1 + 0.0153x_2 - 0.0145x_3 + 0.119x_4}$	**<0.001**	0.93	**<0.001**	**0.03**	**<0.001**	**<0.001**	29.7	**<0.001**

For each hybrid in Tel Hadya, there were significant equations (*P<0.05*). However, most of the coefficients did not have significant influence on total plant yield. Plant height was significant for H1 in Tel Hadya (linear and exponential) and leafiness was significant for H2 in Tel Hadya (linear and exponential) Table 7.11. There was a weak correlation between total plant yield and morphological characters. The coefficient of determinations (each hybrid in Tel Hadya) were significant but very weak ($10.5 \leq R^2 \leq 14.4$) (Table 7.11). The linear and exponential equations (hybrids in Tel Hadya) indicated a negative correlation between total plant yield and vegetative duration. The linear and exponential equations for each hybrid in Breda were significant. However, the coefficients for vegetative duration and length of growing season were not significant. Plant height and leafiness (hybrids in Breda) were positively correlated with total plant yield (Table 7.11). The coefficients of determination for each hybrid in Breda, between *19.00%* and *33.7%*, were not very strong (Table 7.11). The coefficient of determination for Tel Hadya and Breda was not significantly increased compared with the coefficient of determination for each hybrid in each area. However, the independent variables (morphological characters) had significant influence on total plant yield with the exception of vegetative duration in Breda (Table 7.11). According to analysis of variance for the relationship between total plant yield and morphological characters in Tel Hadya and Breda, the residuals do not differ significantly from normal distribution (*P>0.05*). The standardised residuals versus the fitted values for Tel Hadya and Breda are shown in Figure 7.3.

In Tel Hadya, Figures 7.3.a and 7.3.b indicated ten points for linear regression and eight points for exponential regression which were more than two standardised residuals away, and seven points which were influential (linear and exponential). In Breda, there were fifteen points for linear regression and exponential regression which were more than two standardised residuals away, and three points which were influential (linear and exponential) Figures 7.3.c and 7.3.d. Table 7.11 and Figure 7.3 indicated that the relationship between total plant yield and morphological characters in Breda was more significant than the relationship in Tel Hadya.

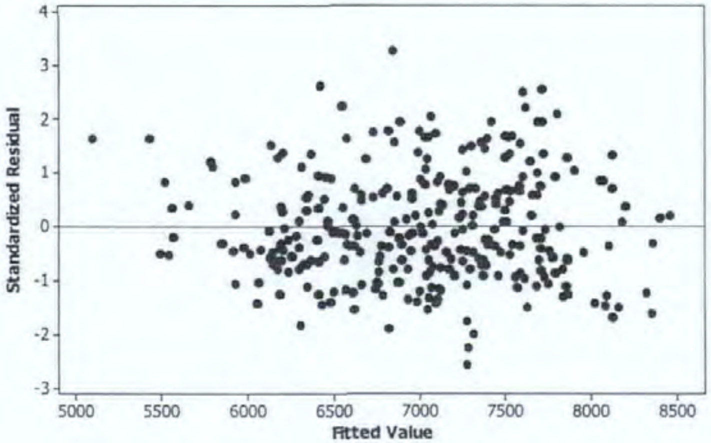


Figure 7.3.a. The residuals between the total plant yield and morphological characters in Tel Hadya (linear).

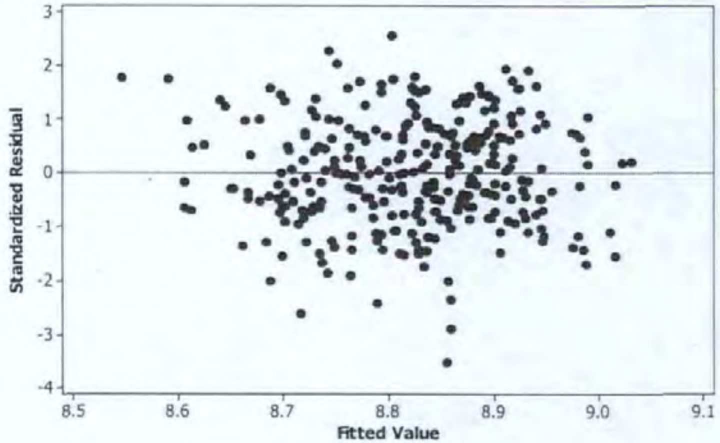


Figure 7.3.b. The residuals between the total plant yield and morphological characters in Tel Hadya (exponential).

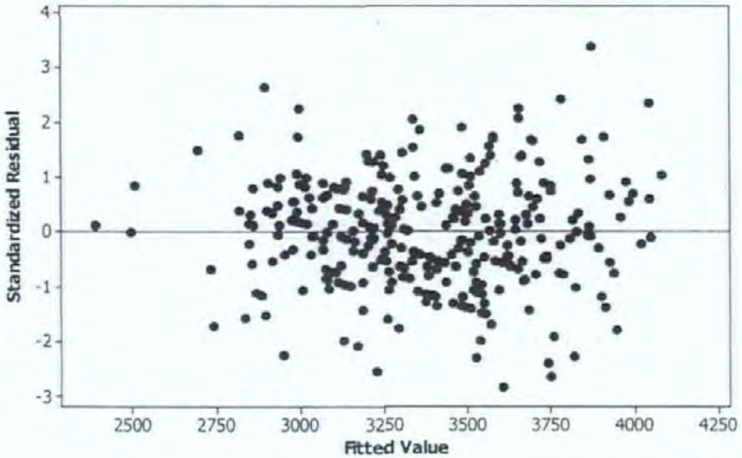


Figure 7.3.c. The residuals between the total plant yield and morphological characters in Breda (linear).

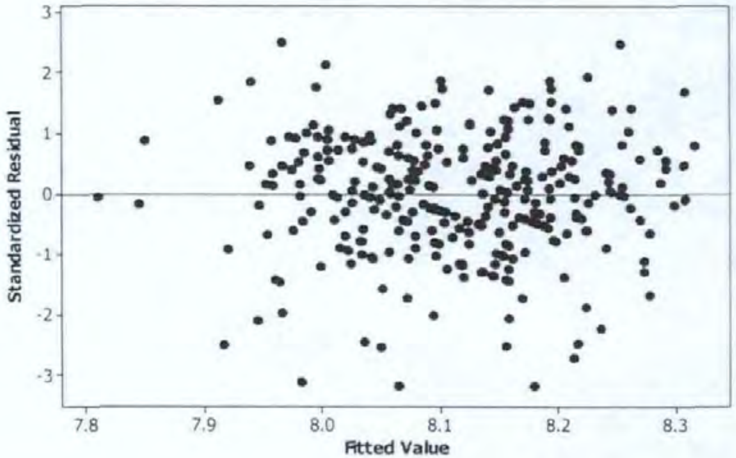


Figure 7.3.d. The residuals between the total plant yield and morphological characters in Breda (exponential).

Figure 7.3. The residuals between the total plant yield and morphological characters (linear and exponential).

In conclusion

- ✓ Total plant yield for each hybrid (Tel Hadya and Breda) was not affected by vegetative duration and length of growing season while total plant yield correlated with plant height and leafiness.
- ✓ The coefficients of determination for Tel Hadya were very weak but significant. Approximately *33%* , *28%* and *19%* of the variations in the total plant yield were explained by the linear and exponential equations for H1, H2 and H3 respectively in Breda. The coefficients of determination in Breda were better than Tel Hadya because the data for Breda were more stable than those for Tel Hadya.
- ✓ The relationship between total plant yield and morphological characters was not clear for several reasons. For example: there were insufficient observations for each hybrid and moreover, not all the independent variables were stable (Chapter 4).
- ✓ Table 7.11 indicated that environment influenced the relationship between total plant yield and morphological characters. For example, plant height was significant in the equations for hybrids 2 and 3 in Breda while it was not significant in Tel Hadya. Leafiness was significant for hybrids 1 and 3 in Breda while it was not significant in Tel Hadya. Also the coefficients of determination in Breda were higher than the coefficients of determination in Tel Hadya.
- ✓ The relationship between total plant yield and morphological characters was not greatly affected by genotype. For example, total

plant yield is affected by plant height and leafiness for all hybrids in Breda.

✓ The multiple exponential equation is better than the multiple linear equation for studying the relationship between total plant yield and morphological characters because:

- The constant was significant in the exponential equation, while the constant was not significant in the linear equation.
- Residual errors in exponential regression were lower than residual errors in linear regressions.

7.3.2 Using multiple regression analysis to study the relationship between grain yield and morphological characters

The results for linear and exponential regression for the relationship between grain yield and morphological characters are summarized by Table 7.12.

There were no significant relationships between grain yield and morphological characters for H1 and H3 in Tel Hadya (*P>0.05*) (Table 7.12) and there was a weak relationship between grain yield and morphological characters for H2 in Tel Hadya (*P<0.05*).The coefficient of determination for H2 was very weak but significant (*P<0.05*) and the coefficients were not significant, except for leafiness (*P<0.05*); (Table 7.12). For each hybrid in Breda, the results in Table 7.12 indicated that there were significant relationships between grain yield and morphological characters (*P<0.05*). The coefficient of determination (between *18.3%* and *20.4%*) was not strong, although it was significant (Table 7.12). Grain yield for each hybrid in Breda was positively affected by the coefficients plant height and leafiness (*P<0.05*) while vegetative duration and length of growing season were not significant enough to influence grain yield (linear and exponential); (Table 7.12). The coefficient of determination for the data in each area Tel Hadya and Breda did not increase when the coefficient of determination was studied by each hybrid. Approximately *10%* to *23%* of the variations in the grain yield were explained by the linear and exponential equations for Tel Hadya and Breda respectively. In Tel Hadya and Breda, there were significant relationships between grain yield and morphological characters (*P<0.05*) (Table 7.12). In Tel Hadya, the grain yield was significantly affected by vegetative duration, length of growing season and leafiness whereas plant height was not significant (*P>0.05*) (Table 7.12). In Breda, the grain yield was significantly affected by plant height, length of growing season and leafiness while vegetative duration did not have significant influence on grain yield (*P>0.05*) (Table 7.12). Leafiness and vegetative duration had a negative influence on grain yield in Tel Hadya while they had a positive influence in Breda. Grain yield was positively affected by length of growing season in Tel Hadya while it was negatively affected in Breda. The standardised residuals versus the fitted values for Tel Hadya and Breda are shown in Figure 7.4. In Tel Hadya Figures 7.4.a and 7.4.b indicate fourteen points for linear regression and nine points for exponential regression which were more than two standardised residuals away, while there were seven points (linear and exponential). In Breda, there were fifteen points for linear regression and twelve points for exponential regression that were more than two standardised residuals away. Also there were three points that were influential (linear and exponential) Figures 7.4.c and 7.4.d.

Table 7.12. The relationship between grain yield (y) and the variables: vegetative duration (x_1), plant height (x_2), length of growing season (x_3) and leafiness (x_4).

Area	Hybrid	Equation	$P_{constant}$	P_{a_1}	P_{a_2}	P_{a_3}	P_{a_4}	$P_{equation}$	R^2%	P_{R^2}
Tel Hadya	1	$y = 4214 - 63.1x_1 + 39.1x_2 + 24.1x_3 + 197x_4$	0.50	0.21	0.08	0.45	0.26	0.053	9.1	**<0.05**
		$y = 3677.54e^{-0.014x_1 + 0.0085x_2 + 0.0063x_3 - 0.0476x_4}$	**<0.001**	0.24	0.11	0.41	0.26	**0.08**	8.1	**<0.05**
	2	$y = -1956 - 21.7x_1 + 13.5x_2 + 39.6x_3 + 218x_4$	0.68	0.61	0.24	0.25	**0.03**	**0.04**	9.5	**<0.05**
		$y = 880.07e^{-0.0122x_1 + 0.00405x_2 + 0.0136x_3 + 0.0800x_4}$	**<0.001**	0.39	0.30	0.25	**0.02**	**0.03**	10.4	**<0.05**
	3	$y = 6456 - 28.2x_1 + 8.2x_2 - 16.0x_3 + 174x_4$	0.17	0.43	0.51	0.62	0.17	0.13	7.0	>0.05
		$y = 10198e^{-0.0110x_1 + 0.00286x_2 - 0.0052x_3 + 0.0651x_4}$	**<0.001**	0.40	0.53	0.66	0.16	0.13	7.0	>0.05
Breda	1	$y = 1700 - 22.9x_1 + 25.3x_2 + 1.4x_3 + 166x_4$	0.62	0.28	**0.01**	0.94	**0.002**	**<0.001**	20.3	**<0.05**
		$y = 1998.19e^{-0..0163x_1 + 0.0156x_2 + 0.0007x_3 - 0.112x_4}$	**0.001**	0.44	**0.014**	0.955	**0.001**	**<0.001**	20.4	**<0.05**
	2	$y = 6529 + 26.4x_1 + 24.7x_2 - 71.1x_3 + 225x_4$	0.34	0.44	**0.04**	0.119	**<0.001**	**<0.001**	23.1	**<0.05**
		$y = 29732e^{0.0122x_1 + 0.0150x_2 - 0.0396x_3 - 0.131x_4}$	**0.02**	0.57	**0.046**	0.165	**0.001**	**<0.001**	20.9	**<0.05**
	3	$y = -1992 + 17.2x_1 + 30.7x_2 + 2.9x_3 + 95.2x_4$	0.66	0.54	**0.001**	0.92	**0.02**	**0.001**	18.3	**<0.05**
		$y = 109.95e^{0.0129x_1 + 0.0196x_2 + 0.0035x_3 - 0.0549x_4}$	0.09	0.45	**0.001**	0.84	**0.03**	**0.001**	18.4	**<0.05**
Tel Hadya		$y = 3311 - 111x_1 + 14.1x_2 + 86.0x_3 - 278x_4$	0.45	**0.001**	0.181	**0.001**	**0.002**	**<0.001**	10.3	**<0.05**
		$y = 3533.34e^{-0.0335x_1 + 0.0043x_2 + 0.0248x_3 - 0.0733x_4}$	**<0.001**	**0.001**	0.16	**0.001**	**0.006**	**<0.001**	9.9	**<0.05**
Breda		$y = 2164 + 9.4x_1 + 23.8x_2 - 23.6x_3 + 162x_4$	0.31	0.60	**<0.001**	**0.037**	**<0.001**	**<0.001**	23.7	**<0.05**
		$y = 2952.29e^{0.0050x_1 + 0.0140x_2 - 0.0160x_3 + 0.120x_4}$	**<0.001**	0.66	**<0.001**	**0.026**	**<0.001**	**<0.001**	22.5	**<0.05**

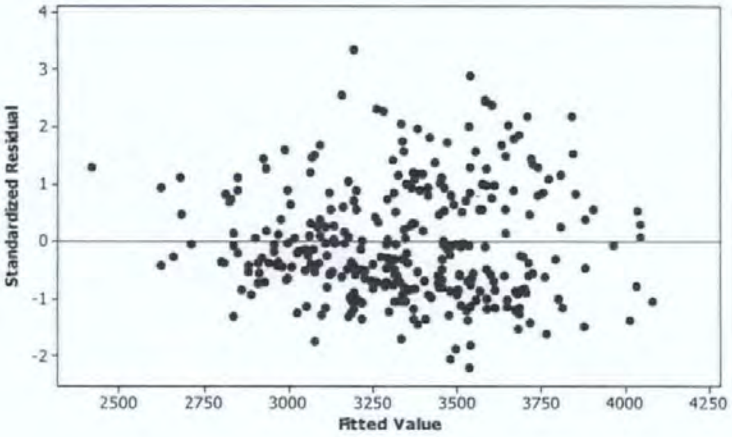


Figure 7.4.a. The residuals between the grain yield and morphological characters in Tel Hadya (linear).

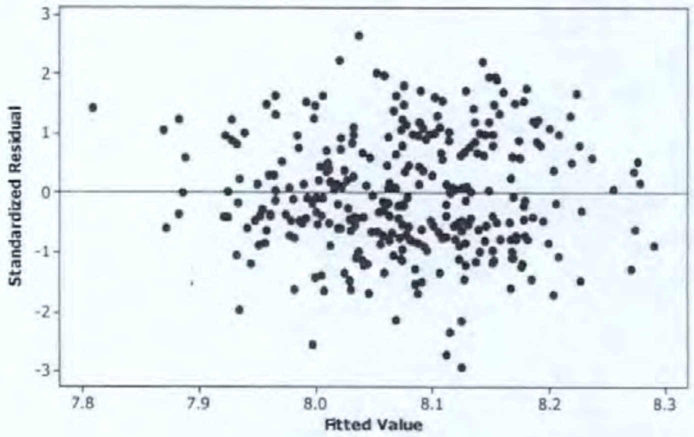


Figure 7.4.b. The residuals between the grain yield and morphological characters in Tel Hadya (exponential).

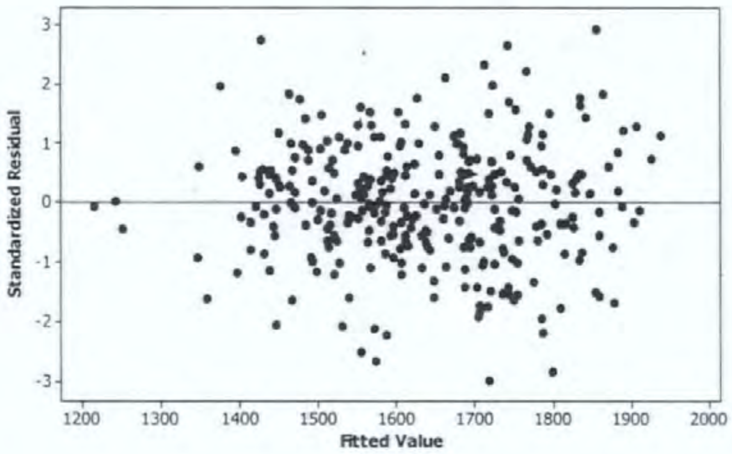


Figure 7.4.c. The residuals between the grain yield and morphological characters in Breda (linear).

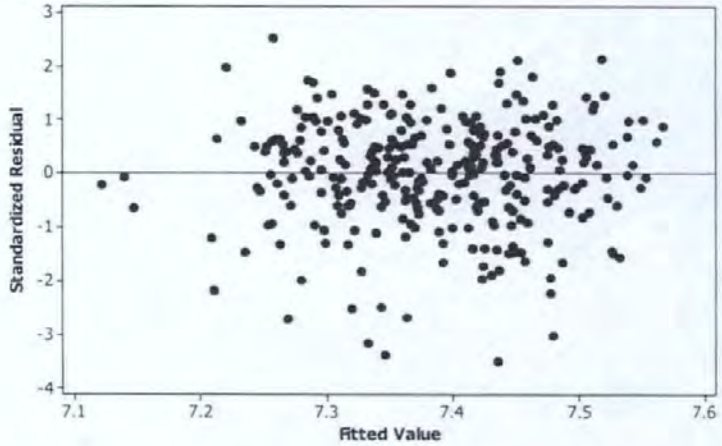


Figure 7.4.d. The residuals between the grain yield and morphological characters in Breda (exponential).

Figure 7.4. The residuals between the grain yield and morphological characters (linear and exponential).

In conclusion:

- ✓ Grain yield for each hybrid in Tel Hadya was not affected by morphological characters since grain yield for each hybrid in Breda was positively affected by plant height and leafiness.
- ✓ The coefficients of determination for Tel Hadya were very weak and significant for H1 and H2 since the coefficient of determination for H3 was not significant. The coefficients of determination in Breda were better than in Tel Hadya because the data in Breda was stable more than in Tel Hadya.
- ✓ While there were short observation data in each hybrid and some of the variables were not stable (Chapter 4), the relationship between grain yield and morphological characters did not appear clear.
- ✓ The relationship between grain yield and morphological characters was affected by environment. For example, plant height and leafiness were significant for each hybrid in Breda while they were not significant in Tel Hadya (Table 7.12). The genotype did not influence the relationship between grain yield and morphological characters because plant height and leafiness were significant in the equations for all hybrids in Breda (Table 7.12).
- ✓ The multiple exponential equation is better than the multiple linear equation because the constant was significant in the exponential equation, while the constant was not significant in the linear equation.

Residual errors in exponential regression were lower than residual error in linear regressions.

7.3.3 Using multiple regression analysis to study the relationship between straw yield and morphological characters

There was a relationship between straw yield and morphological character for each hybrid (H1, H2 and H3) and each area (Tel Hadya and Breda) ($P<0.05$) (Table 7.13). Each hybrid in Tel Hadya and Breda, Table 7.13 indicated that vegetative duration and length of growing season did not affect straw yield (vegetative duration and length of growing season were not significant ($P>0.05$)). Straw yield was affected by leafiness for each hybrid ($P<0.05$) except H1 and H3 in Tel Hadya and H3 in Breda ($P>0.05$). Plant height affected straw yield for H1 in Tel Hadya ($P<0.05$) since it affected straw yield for each hybrid in Breda ($P<0.05$). The variation in straw yield explained by the linear and exponential equations is indicated by Table 7.13. The variation for each hybrid in Tel Hadya explained by linear and exponential equations was between (10.7% and 15.0%). However, the coefficient of determination for each hybrid was significant ($P<0.05$). In each area (Tel Hadya and Breda), coefficients of determination were quite low (*14%* and *28%* respectively) and it is quite evident that the linear and exponential regression model is not suitable. However, the morphological variables were significant except for vegetative duration in Breda.

Figure 7.5 shows the standard residual between the straw yield and morphological characters in Tel Hadya and Breda (linear and exponential).

Table 7.13. The relationship between straw yield (y) and the variables: vegetative duration (x_1), plant height (x_2), length of growing season (x_3) and leafiness (x_4).

Area	Hybrid	Equation	$P_{conestan}$	P_{a_1}	P_{a_2}	P_{a_3}	P_{a_4}	$P_{equation}$	R^2%	P_{R^2}
Tel Hadya	1	$y = 9939 - 76.1 x_1 + 41.5 x_2 - 11.5 x_3 + 274 x_4$	0.09	0.10	**0.04**	0.69	0.09	**0.003**	15.0	**<0.05**
		$y = 12835.88e^{-0.0148 x_1 + 0.00938 x_2 - 0.00316 x_3 + 0.0630 x_4}$	**<0.001**	0.155	**0.046**	0.64	0.09	**0.005**	14.1	**<0.05**
	2	$y = -4649 - 11.2 x_1 + 24.3 x_2 + 47.6 x_3 + 344 x_4$	0.510	0.858	0.159	0.361	**0.020**	**0.026**	10.7	**<0.05**
		$y = 555.57e^{-0.0108 x_1 + 0.00695 x_2 + 0.0151 x_3 + 0.107 x_4}$	**0.001**	0.53	0.144	0.291	**0.009**	**0.009**	12.9	**<0.05**
	3	$y = 6810 - 47.5 x_1 + 17.1 x_2 - 10.3 x_3 + 280 x_4$	0.24	0.27	0.26	0.79	0.07	**0.018**	11.4	**<0.05**
		$y = 11271.13e^{-0.0167 x_1 + 0.00593 x_2 - 0.0035 x_3 + 0.0943 x_4}$	**<0.001**	0.24	0.24	0.79	0.07	**0.012**	12.2	**<0.05**
Breda	1	$y = 1417 - 60.5 x_1 + 43.3 x_2 + 21.2 x_3 + 264 x_4$	0.717	0.106	**<0.001**	0.357	**<0.001**	**<0.001**	34.9	**<0.05**
		$y = 1635.98e^{-0.0394 x_1 + 0.0242 x_2 + 0.0142 x_3 + 0.160 x_4}$	**0.002**	0.078	**<0.001**	0.301	**<0.001**	**<0.001**	34.3	**<0.05**
	2	$y = 8772 + 11.7 x_1 + 27.0 x_2 - 81.6 x_3 + 350 x_4$	0.31	0.78	0.069	0.150	**<0.001**	**<0.001**	26.9	**<0.05**
		$y = 40134.84e^{0.0013 x_1 + 0.0175 x_2 - 0.0366 x_3 + 0.185 x_4}$	**0.024**	0.95	**0.03**	0.23	**<0.001**	**<0.001**	27.8	**<0.05**
	3	$y = 2820 + 7.3 x_1 + 25.9 x_2 - 24.6 x_3 + 79.1 x_4$	0.56	0.81	**0.011**	0.40	0.07	**0.005**	14.1	**<0.05**
		$y = 2751.77e^{0.0035 x_1 + 0.0155 x_2 - 0.0127 x_3 + 0.0463 x_4}$	**0.006**	0.85	**0.01**	0.466	0.074	**0.005**	14.0	**<0.05**
Tel Hadya		$y = 4273 - 123 x_1 + 41.7 x_2 + 75.4 x_3 - 218 x_4$	0.32	**<0.001**	**<0.001**	**0.004**	**0.017**	**<0.001**	13.9	**<0.05**
		$y = 4582.5e^{-0.0360 x_1 + 0.0122 x_2 + 0.0212 x_3 - 0.0572 x_4}$	**<0.001**	**<0.001**	**<0.001**	**0.004**	**0.024**	**<0.001**	14.8	**<0.05**
Breda		$y = 2843 - 7.3 x_1 + 30.3 x_2 - 20.6 x_3 + 230 x_4$	0.26	0.73	**<0.001**	0.124	**<0.001**	**<0.001**	28.2	**<0.05**
		$y = 2271.13e^{-0.0070 x_1 + 0.0163 x_2 - 0.0135 x_3 + 0.135 x_4}$	**<0.001**	0.57	**<0.001**	0.08	**<0.001**	**<0.001**	28.0	**<0.05**

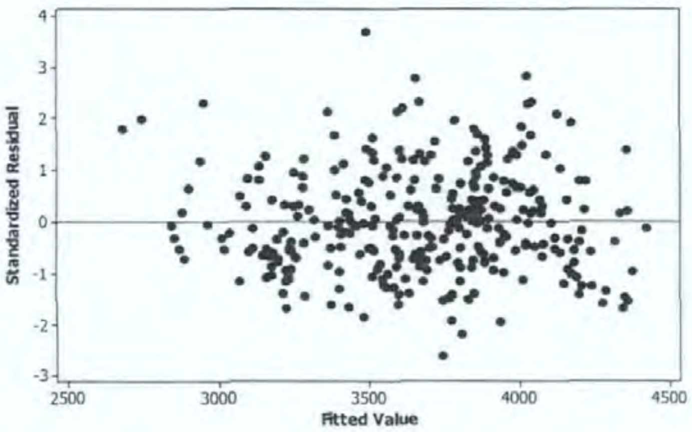


Figure 7.5.a. The residuals between the straw yield and morphological characters in Tel Hadya (linear).

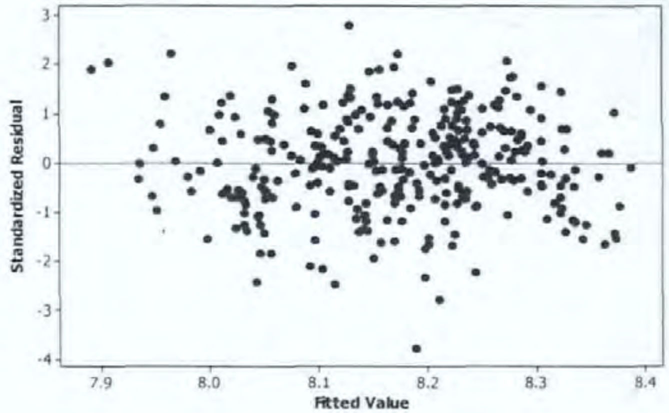


Figure 7.5.b. The residuals between the straw yield and morphological characters in Tel Hadya (exponential).

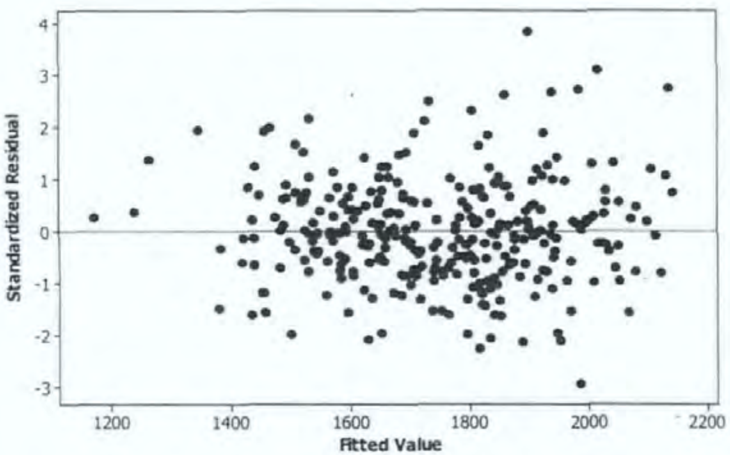


Figure 7.5.c. The residuals between the straw yield and morphological characters in Breda (linear).

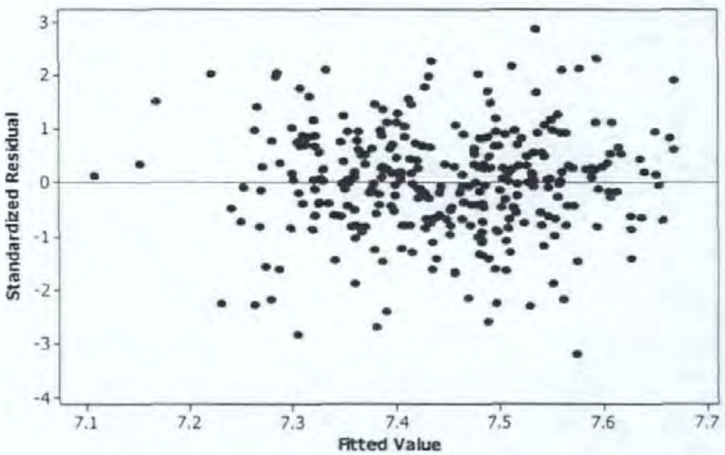


Figure 7.5.d. The residuals between the straw yield and morphological characters in Breda (exponential).

Figure 7.5. The residuals between straw yield and morphological characters (linear and exponential).

7.3.4 Conclusion and summary for each hybrid and area

- ✓ The statistical analyses presented here clearly indicate a common dilemma for the biological interpretation of the results. Thus, whilst significant relationships between variables have been identified, the percentage of variation accounted for by the relationship is very low, indicating that perhaps insufficient data or the existence of other, unmeasured variables are greatly influencing the significance of the relationships.
- ✓ The relationship between yield parameters and morphological characters did not appear at all clearly because there were insufficient observations in each hybrid. Also independent variables were not stable (Chapter 4).
- ✓ Yield parameters for each hybrid were affected by plant height and leafiness more than by vegetative duration and length of growing season.
- ✓ The coefficients of determination were quite low. However, most of the equations were significant.
- ✓ The relationship between yield parameters and morphological characters was affected by environment.
- ✓ The relationship between yield parameters and morphological characters was not affected by genotype except for the relationship between straw yield and morphological characters.

- ✓ The relationship between yield parameters and morphological characters in Breda was more significant than the relationship in Tel Hadya.
- ✓ Although there was no evidence that the linear and exponential regression was suitable, the multiple exponential equation is better than the multiple linear equation for studying the relationship between yield parameters and morphological characters.

Chapter 8

General Discussion and Conclusion

8.1 Influence on yield

It was clear in the current investigation that there were several factors affecting barley yield; these were grouped as either environmental characters or genotypical factors and these characters and factors were linked in different ways according to location. Shakhatreh et al. (2001) also found that there was a significant and positive relationship between grain yield, biological yield, straw yield, plant height and harvest index, but this was irrespective of location.

It is well known that drought can reduce grain size, the number of spikelets/ear, plant height, tiller number, tiller fertility, the number of grains/ear and the extent of tiller mortality (Briggs, 1978). In the present study the yield in Breda was lower than the yield in Tel Hadya due to various factors but principally because the weather in Breda was drier than the weather in Tel Hadya. This result agreed with the findings of Shakhatreh et al. (2001), who indicated a high negative correlation between a drought susceptibility index and grain yield at the driest site and a lower, and sometimes, positive correlation at the wettest site. This was demonstrated in Chapter 4, where the genotype was affected by environment with significant differences between the mean yield parameters (total plant yield, grain yield, straw yield, number of ears and TGW) and morphological characters (vegetative duration, plant height, length of growing season and leafiness) for each of the three hybrids studied over both of the seasons studied and for each hybrid in the different regions (Tel Hadya and Breda). Furthermore, the same area and season gave different yield parameters and differing importance to the morphological characters for the different hybrids, and this supports the work of Fox, *et al.* (2006).

There was a significant interaction for genotype × year (in the same area and hybrid but in different seasons) and according to Spearman correlation, there was a significant difference between each pair of hybrids in Tel Hadya in different seasons[1]. Also there was a significant interaction for genotype × location (within the same hybrid and season but in different locations) and a difference between each pair of hybrids in different areas (Breda and Tel Hadya) and similars result were obtained by Vanoosterom, *et al.* (1993b) and Dehghania, *et al.* (2006).

Leafiness was affected by environment, and in Tel Hadya the level of leafiness was greater than the level in Breda. Variations in leafiness between sites is also reported by other workers (Johnson and Whittington, 1977) indicating an apparent plasticity of this factor. Despite this variation yield parameters were affected by levels of leafiness and ANOVA (*LSD*) indicated that there was a significant difference between yield parameters according to levels of leafiness. It was evident that the highest level of leafiness (*5*) did not correspond with the highest yield with level 5 for H3 in Breda and H1 in Tel

[1] H1 season one with H1 season two
H2 season one with H2 season two
H3 season one with H3 season two

Hadya giving lower yields than other leafiness levels. Leafiness level 4.5 was shown to be the best level for leafiness in most of the different environments (locations) and this level always correlated with the best yield production. This is contrary to the situation in temperate rainfed agriculture in Western Europe, but is not surprising for Syrian situations since high leafiness would have higher soil water demands and if this cannot be met by sufficient irrigation then it will exacerbate terminal drought in cereal crops.

As a consequence of these analyses and findings, the stability with respect to genotype and the stability with respect to environment, evaluated in terms of usefulness and reliability, was used in subsequent analyses of the relationship between yield parameters and morphological characters. Furthermore this relationship was investigated to see whether or not these relationships were changed by groups of factors associated with the environment.

8.2 The relationship between morphological characters and yield parameters

A full understanding of the relationship between yield parameters and morphological characters could not be answered by using only one type of analysis. Different analyses brought different interpretations to the data sets under investigation.

8.2.1 Canonical correlation analysis and factor analysis

Canonical correlation was chosen because it is the method which is most suitable for studying the relationship between two sets of parameters (Zhang, *et al.*, 1991; Vadiveloo and Phang, 1996). The aim of using canonical correlation was to discover whether yield parameters as a set were affected by morphological characters also as a set.

Canonical correlation indicated a relationship between yield parameters and morphological characters. Yield parameters were affected positively by plant height, vegetative duration and length of growing season. In essence, tall barley with longer vegetative duration and longer length of growing season had greater yield than shorter barley with shorter vegetative duration and shorter length of growing season. Thousand grain weight (TGW) was not, however, significantly affected by morphological characters. Indeed, TGW is normally affected by environmental factors more than other factors as shown by Hacett, *et al.*, (2001).

Although leafiness influenced yield parameters overall, this was not apparent in detailed analyses because the genotypes appeared to be unstable (Wilcoxon test). In this investigation, the suggestion was that the influence between yield parameters and level of leafiness should be studied separately.

Canonical analysis indicated that the relationship between yield parameters and morphological characters may have been affected by environment since the canonical correlation in Breda ($r_{canonical\ correlation} = .0.560;\ P < 0.05$) was more significant than in Tel Hadya ($r_{canonical\ correlation} = .3981;\ P < 0.05$). Also the order of contribution of morphological characters in Breda (plant height, leafiness, length of growing season and vegetative duration) was different to that of Tel

Hadya (plant height, length of growing season, vegetative duration and leafiness). This difference is explored in more detail using factor analysis and multiple regression analysis (below).

Finally, canonical correlation suggested that there was a relationship between yield parameters and morphological characters, and among the morphological characters, plant height was the most important parameter positively affecting yield, a finding which contradicts the findings of ShouFu et al . (2007), whose research indicated that, "For plant height, neither tall nor short varieties are good. Plants of intermediate height will have enough growing quantity and will result in big ear, many grains and heavy grains"

Whilst factor analysis has been used elsewhere to study the interrelationship between barley parameters, for example, the interrelationship between morphology and phenology of wild barley (Volis, *et al.*, 2002), studies that bring all morphological characters and yield parameters together are not commonly found in the literature. The aim of using factor analysis was to discover the interrelationship between parameters (morphological characters and yield parameters).

Factor analysis indicated similar findings to canonical correlation for the relationship between yield parameters and morphological characters for all data. However, this relationship was not apparent for each hybrid or area because the sample sizes were insufficiently large. There were three factors for each hybrid and area which explained eight variables (total plant yield, grain yield, straw yield, TGW, vegetative duration, plant height, length of growing season and leafiness). The first factor "yield" was the most important factor since it consisted of total plant yield, grain yield and straw yield and showed a strong correlation between the three parameters. The second and third factors revealed a difference between Tel Hadya and Breda. The second factor in Tel Hadya strongly correlated with vegetative duration and length of growing season and was regarded as the "plant period". Factor 3 in Tel Hadya was termed a "morphology" factor since it consisted of plant height and leafiness. In Breda, the second and third factors were not stable because factors 2 and 3 correlated with different variables for each hybrid. For example, factor 2 for H1 in Breda was correlated with vegetative duration and length of growing season, in H2 it was correlated with TGW and length of growing season, whilst in H3 it was correlated with TGW, plant height and length of growing season and in Breda the second factor correlated with vegetative duration and plant height. Clearly, the relationship between the variables was affected by the environments leading to complex interactions.

Factor analysis for season two (all data) indicated that yield parameters were affected by all morphological characters except leafiness.

In conclusion, canonical analysis and factor analysis suggested that environment influenced the relationship between variables; and for season two the yield parameters had been affected by all morphological characters except leafiness.

8.2.2 Multiple regression analysis to study the relationship between yield parameters and morphological characters

Whilst simple regression analysis has been used to analyze specific characters in barley growth and development, for example, TGW with climate (Beavan, 1947); plant height with yield (Adams, *et al.*, 2004) and vegetative duration with grain yield (Sinebo, 2002). There is very little in the literature that brings all of these characters together. This findings reported in this report have shown that multiple regression analysis has the following advantages: (i) to determine statistically significant morphological character variables that explain the variation in each yield parameter; (ii) to assist the making of reliable inferences for the relationship between yield parameters and morphological characters.

Stepwise multiple regression (linear and exponential) for season two (all data) indicated that yield parameters had a significant relationship with all morphological characters except leafiness. Plant height and vegetative duration were the most important variables which positively influenced yield parameters. A similar result was obtained by Sinebo (2002), who also found that grain yield was correlated positively with vegetative duration and vegetative height. However, the coefficients of determination slightly increased when the length of growing season was added to the equation. This suggested that using a stepwise exponential equation was better than using a stepwise linear equation for studying the relationship between yield parameters and morphological characters because the coefficients of determination in exponential equations were consistently higher than the coefficients of determination in linear equations in each step. Residual errors in exponential regression were also shown to be lower than residual errors in linear regressions.

Because there were insufficient observations for each hybrid and area, and because independent variables (morphological characters) were not stable, the overall relationship was not at all clear. However, yield parameters were affected by plant height and leafiness more than vegetative duration and length of growing season. The relationship between yield parameters and morphological characters was not stable because the environment influenced the relationship. For example, the length of growing season in Tel Hadya correlated positively with grain yield while it correlated negatively in Breda. These findings concur with those of Shakhatreh *et al.*, (2001), at five locations in Jordan. At the wettest of these locations, there was a positive relationship between grain yield and length of grain filling period, while at the driest location the relationship was negative. Also total plant yield in Breda was affected by plant height for Hybrids 2 and 3 while it was non-significant in Tel Hadya and grain yield was affected by plant height and leafiness for each hybrid in Breda while they were not significant in Tel Hadya.

8.3 Conclusion

Yield parameters were affected by many factors including, environment, genotype and morphological characters. H1 produced the highest yield in Tel Hadya for two seasons while in Breda H2 produced the highest yield.

There was strong evidence of genotype-environment interactions. However, TGW and vegetative duration were not affected by location (Breda and Tel Hadya) so it would be difficult for the breeder to decide which family gave the highest output based on these parameters.

Yield parameters were affected by level of leafiness and the greatest yield was correlated with leafiness level 4.5.

Multivariate analysis techniques (canonical analysis, factor analysis and multiple regression analysis) indicated that yield parameters had been affected by all morphological characters except leafiness in season two. This technique was not suitable for studying the relationship between leafiness and yield parameters because leafiness in stepwise regression analysis was not a significant variable in the equations (linear or exponential). In factor analysis leafiness did not correlate with yield parameters and was categorized as a separate factor. The most important variables positively affecting yield parameters were plant height and vegetative duration while length of growing season was less important, whilst still significant in the equations (linear and exponential). Consequently, all morphological characters (plant height, vegetative duration, length of growing season, leafiness) affected yield but multivariate techniques were not suitable for studying the relationship between leafiness and yield parameters.

Exponential equations were better than linear equations for studying the relationship between yield parameters and morphological characters for season two for the following reasons: (i) the coefficient of determination for exponential equation was higher than the coefficient of determination for linear equation (Table 8.1), (ii) residual errors in exponential regression were lower than residual errors in linear regressions.

Table 8.1. The coefficients of determination for the relationship between yield parameters and morphological characters (linear and exponential).

R^2%	Linear	Exponential
Total plant yield	*64.38*	*71.71*
Grain yield	*58.8*	*67.71*
Straw yield	*64.5*	*71*

The relationship between yield parameters and morphological characters for each hybrid in Tel Hadya was not clearly apparent since the coefficient of determination was very weak and most of the independent variables were not significant. In Breda the relationship was more significant than in Tel Hadya for each hybrid because the coefficient of determination was higher and there were two independent significant variables (plant height and leafiness) in this relationship.

Multivariate analysis techniques (canonical analysis, factor analysis and multiple regression analysis) indicated that the relationship between yield and morphological characters was strongly affected by environment while this relationship was not affected by genotype.

8.4 Overview of the research

The outputs of this research were successful in confirming that yield parameters were affected by morphological characters. This relationship was studied with the aim of improving barley yield. It was not the aim to study the influence of environment on yield, so the techniques were used to see whether a relationship existed between yield parameters and morphological characters and to discover whether this relationship was linear or exponential. Also the study aimed to discover whether the relationship was affected by environment or genotype.

This thesis illustrates that the multivariate technique (canonical correlation, factor analysis, stepwise regression analysis and multiple regression analysis) can be useful to barley breeders in order to provide the following informative advice:

- It is advisable to evaluate breeding lines at several sites and be aware that G × E effects are likely.
- Breeding lines should be evaluated over several seasons.
- Care should be taken in assigning leafiness scores to trials at different sites and seasons.

- Leafiness should not be relied on as an indicator of yield potential.
- It should be recognised that vegetative duration and plant height are reasonably good indicators of yield potential.
- It should be recognised that G × E effects may affect individual morphological characters differently between families within a hybrid and variability can be high. This indicates that families even at the F5 generation in an inbreeding crop like barley may still be unstable.
- The stressful growing conditions likely to prevail in Syria, particularly drought stress, are likely to exacerbate G × E effects.
- The use of multiple exponential equations to study the relationship between yield parameters and morphological characters is a sound approach with the exception of leafiness.

8.5 Future work

- The number of location and season should be extended to provide further evidence on the genotype-environment interactions.
- To study the genotype-environment interaction the use of General Linear Model (GLM) would be useful. For example, genotype × location; genotype× year and genotype × location × year.
- With the extension of location and season new methods of analysis should be used. For example, Multilevel Factor Analysis, Multilevel Regression Analysis.

Appendices

Table a.1.a. LSD for differences between mean yields (H1; season two in Tel Hadya).

Dependent Variable	leafiness (a)	Frequency	Mean	leafiness (b)	Mean Difference (a-b)	Individual P value of difference	ANOVA P value
Total Plant yield	medium 3	12	7916.2	medium high 3.5	-855.70	.060	.046
				high 4	-872.04	.044	
				medium very high 4.5	-1571.62	.003	
				very high 5	-.33	1.000	
	medium high 3.5	30	8771.9	high 4	-16.34	.958	
				medium very high 4.5	-715.92	.096	
				very high 5	855.37	.375	
	high 4	44	8788.2	medium very high 4.5	-699.58	.086	
				very high 5	871.70	.362	
	medium very high 4.5	14	9487.8	very high 5	1571.29	.117	
Grain yield	medium 3	12	3874.6	medium high 3.5	-451.45	.086	.064
				high 4	-509.94	.043	
				medium very high 4.5	-764.49	.012	
				very high 5	379.08	.516	
	medium high 3.5	30	4326.0	high 4	-58.49	.747	
				medium very high 4.5	-313.04	.207	
				very high 5	830.53	.139	
	high 4	44	4384.5	medium very high 4.5	-254.55	.279	
				very high 5	889.02	.110	
	medium very high 4.5	14	4639.1	very high 5	1143.57	.050	
Straw yield	medium 3	12	4041.6	medium high 3.5	-404.25	.105	.093
				high 4	-362.10	.128	
				medium very high 4.5	-806.42	.006	
				very high 5	-379.42	.494	
	medium high 3.5	30	4445.8	high 4	42.15	.806	
				medium very high 4.5	-402.17	.089	
				very high 5	24.83	.963	
	high 4	44	4403.7	medium very high 4.5	-444.32	.048	
				very high 5	-17.32	.974	
	medium very high 4.5	14	4848.0	very high 5	427.00	.437	

Table a.1.b. LSD for differences between mean yields (H1; season two in Breda).

Dependent Variable	leafiness (a)	Frequency	Mean	leafiness (b)	Mean Difference (a-b)	Individual P value of difference	ANOVA P value
Total Plant yield	low 2	2	2600.0	medium medium 2.5	-197.29	.651	<.0001
				medium 3	-279.12	.493	
				medium high 3.5	-559.71	.159	
				high 4	-873.16	.030	
				medium very high 4.5	-1205.33	.008	
	medium medium 2.5	7	2797.9	medium 3	-81.83	.738	
				medium high 3.5	-362.42	.108	
				high 4	-675.87	.004	
				medium very high 4.5	-1008.05	.001	
	medium 3	17	2879.1	medium high 3.5	-280.59	.080	
				high 4	-594.04	.000	
				medium very high 4.5	-926.22	.001	
	medium high 3.5	38	3159.7	high 4	-313.45	.018	
				medium very high 4.5	-645.62	.008	
	high 4	32	3473.1	medium very high 4.5	-332.18	.172	
Grain yield	low 2	2	1350.0	medium medium 2.5	-6.14	.978	.011
				medium 3	-62.47	.765	
				medium high 3.5	-212.82	.295	
				high 4	-297.81	.145	
				medium very high 4.5	-403.67	.079	
	medium medium 2.5	7	1356.14	medium 3	-56.33	.653	
				medium high 3.5	-206.67	.074	
				high 4	-291.67	.014	
				medium very high 4.5	-397.52	.012	
	medium 3	17	1412.47	medium high 3.5	-150.35	.067	
				high 4	-235.34	.006	
				medium very high 4.5	-341.20	.011	
	medium high 3.5	38	1562.82	high 4	-85.00	.206	
				medium very high 4.5	-190.85	.122	
	high 4	32	1647.8	medium very high 4.5	-105.85	.395	
Straw yield	low 2	2	1250.0	medium medium 2.5	-191.14	.472	<.0001
				medium 3	-216.65	.383	
				medium high 3.5	-346.89	.151	
				high 4	-575.34	.019	
				medium very high 4.5	-801.67	.004	
	medium medium 2.5	7	1441.1	medium 3	-25.50	.864	
				medium high 3.5	-155.75	.255	
				high 4	-384.20	.006	
				medium very high 4.5	-610.52	.001	
	medium 3	17	1466.7	medium high 3.5	-130.25	.180	
				high 4	-358.70	<.0001	
				medium very high 4.5	-585.02	<.0001	
	medium high 3.5	38	1596.9	high 4	-228.45	.005	
				Medium very high 4.5	-454.77	.002	
	high 4	32	1825.3	Medium very high 4.5	-226.32	.127	

Table a.1.c. LSD for differences between mean yields (H2; seasons two in Tel Hadya).

Dependent Variable	leafiness (a)	Frequency	Mean	leafiness (b)	Mean Difference (a-b)	Individual P value of difference	ANOVA P value
Total Plant yield	medium 3	9	6220.7	medium high 3.5	204.28	.685	.018
				high 4	-97.12	.834	
				Medium very high 4.5	-689.52	.148	
				very high 5	-1034.67	.049	
	medium high 3.5	18	6016.4	high 4	-301.40	.405	
				Medium very high 4.5	-893.80	.019	
				very high 5	-1238.94	.005	
	high 4	33	6317.8	Medium very high 4.5	-592.40	.066	
				very high 5	-937.55	.016	
	medium very high 4.5	27	6910.2	very high 5	-345.15	.385	
Grain yield	medium 3	9	2878.9	medium high 3.5	218.11	.318	.025
				high 4	17.74	.930	
				Medium very high 4.5	-245.67	.234	
				very high 5	-287.11	.204	
	medium high 3.5	18	2660.8	high 4	-200.37	.202	
				Medium very high 4.5	-463.78	.005	
				very high 5	-505.22	.008	
	high 4	33	2861.2	Medium very high 4.5	-263.40	.060	
				very high 5	-304.85	.069	
	medium very high 4.5	27	3124.6	very high 5	-41.44	.810	
Straw yield	medium 3	9	3341.8	medium high 3.5	-13.83	.967	.038
				high 4	-114.86	.705	
				Medium very high 4.5	-443.85	.155	
				very high 5	-747.56	.030	
	medium high 3.5	18	3355.6	high 4	-101.03	.669	
				Medium very high 4.5	-430.02	.082	
				very high 5	-733.72	.011	
	high 4	33	3456.6	Medium very high 4.5	-328.99	.119	
				very high 5	-632.70	.013	
	medium very high 4.5	27	3785.6	very high 5	-303.70	.244	

Table a.1.d. LSD for differences between mean yields (H2; season two in Breda).

Dependent Variable	leafiness (a)	Frequency	Mean	leafiness (b)	Mean Difference (a-b)	Individual P value of difference	ANOVA P value
Total Plant yield	medium 2.5	2	3683.0	medium 3	613.06	.151	<.0001
				medium high 3.5	164.43	.690	
				high 4	-210.86	.614	
				Medium very high 4.5	-308.30	.486	
	medium 3	18	3069.9	medium high 3.5	-448.62	.006	
				high 4	-823.91	<.0001	
				Medium very high 4.5	-921.36	<.0001	
	medium high 3.5	44	3518.6	high 4	-375.29	.008	
				Medium very high 4.5	-472.73	.020	
	high 4	28	3893.9	Medium very high 4.5	-97.44	.643	
Grain yield	medium 2.5	2	1776.5	medium 3	281.72	.178	.001
				medium high 3.5	68.20	.736	
				high 4	-82.29	.687	
				Medium very high 4.5	-65.80	.761	
	medium 3	18	1494.7	medium high 3.5	-213.52	.007	
				high 4	-364.01	<.0001	
				Medium very high 4.5	-347.52	.002	
	medium high 3.5	44	1708.3	high 4	-150.49	.028	
				Medium very high 4.5	-134.00	.173	
	high 4	28	1858.8	Medium very high 4.5	16.49	.873	
Straw yield	medium 2.5	2	1906.5	medium 3	331.33	.201	<.0001
				medium high 3.5	96.23	.701	
				high 4	-128.57	.612	
				Medium very high 4.5	-242.50	.367	
	medium 3	18	1575.2	medium high 3.5	-235.11	.017	
				high 4	-459.90	<.0001	
				Medium very high 4.5	-573.83	<.0001	
	medium high 3.5	44	1810.3	high 4	-224.80	.008	
				Medium very high 4.5	-338.73	.006	
	high 4	28	2035.1	Medium very high 4.5	-113.93	.373	

Table a.1.e. LSD for differences between mean yields (H3;season two in Tel Hadya).

Dependent Variable	leafiness (a)	Frequency	Mean	leafiness (b)	Mean Difference (a-b)	Individual P value of difference	ANOVA P value
Total plant yield	medium high 3.5	5	5532.8	high 4	368.58	.487	.023
				Medium very high 4.5	-196.04	.704	
				very high 5	-500.60	.325	
	high 4	23	5164.2	Medium very high 4.5	-564.63	.056	
				very high 5	-869.19	.002	
	medium very high 4.5	32	5728.8	very high 5	-304.56	.228	
Grain yield	medium high 3.5	5	2673.0	high 4	102.00	.689	.125
				Medium very high 4.5	-81.09	.744	
				very high 5	-216.29	.377	
	high 4	23	2571.0	Medium very high 4.5	-183.09	.197	
				very high 5	-318.29	.019	
	medium very high 4.5	32	2754.1	very high 5	-135.19	.266	
Straw yield	medium high 3.5	5	2859,8	high 4	266.58	.392	.012
				Medium very high 4.5	-114.95	.705	
				very high 5	-284.32	.341	
	high 4	23	2593.2	Medium very high 4.5	-381.53	.029	
				very high 5	-550.90	.001	
	medium very high 4.5	32	2974.8	very high 5	-169.37	.254	

Table a.1.f. LSD for differences between mean yields (H3; season two in Breda).

Dependent Variable	leafiness (a)	Frequency	Mean	leafiness (b)	Mean Difference (a-b)	Individual P value of difference	ANOVA F value
Total Plant yield	medium 3	13	3294.5	medium high 3.5	-855.70	.060	.012
				high 4	-872.04	.044	
				Medium very high 4.5	- 1571.62	.003	
				very high 5	-.33	1.000	
	medium high 3.5	39	3218.7	high 4	-16.34	.958	
				Medium very high 4.5	-715.92	.096	
				very high 5	855.37	.375	
	high 4	28	3336.6	Medium very high 4.5	-699.58	.086	
				very high 5	8o71.70	.362	
	medium very high 4.5	19	3605.0	very high 5	1571.29	.117	
Grain yield	medium 3	13	1560.5	medium high 3.5	-451.45	.086	.007
				high 4	-509.94	.043	
				Medium very high 4.5	-764.49	.012	
				very high 5	379.08	.516	
	medium high 3.5	39	1573.4	high 4	-58.49	.747	
				Medium very high 4.5	-313.04	.207	
				very high 5	830.53	.139	
	high 4	28	1606.6	Medium very high 4.5	-254.55	.279	
				very high 5	889.02	.110	
	medium very high 4.5	19	1785.5	very high 5	1143.57	.050	
Straw yield	medium 3	13	1733.9	medium high 3.5	-404.25	.105	.036
				high 4	-362.10	.128	
				Medium very high 4.5	-806.42	.006	
				very high 5	-379.42	.494	
	medium high 3.5	39	1645.3	high 4	42.15	.806	
				Medium very high 4.5	-402.17	.089	
				very high 5	24.83	.963	
	high 4	28	1730.0	Medium very high 4.5	-444.32	.048	
				very high 5	-17.32	.974	
	medium very high 4.5	19	1819.5	very high 5	427.00	.437	

Table a.2. The data for H3, season 2 in Breda.

Number	Total Plant yield	Grain yield	Straw yield	Harvest index	No.of ears	TGW	protein %	vegetative duration	Plant height	Leangth of growing season	Leafiness
1	3083	1523	1560	0.49	280	42.0	12.5	89.5	43.5	129.0	4.0
2	3166	1661	1505	0.52	280	40.3	12.4	91.5	43.0	130.0	4.0
3	3500	1470	2030	0.42	233	38.2	12.2	91.0	44.0	128.5	3.5
4	2833	1325	1508	0.47	286	35.4	13.1	90.0	47.5	129.5	3.5
5	3000	1433	1567	0.48	233	39.3	12.2	89.5	40.5	128.5	3.5
6	3333	1750	1583	0.53	273	40.6	12.5	90.5	42.5	129.0	4.5
7	3500	1583	1917	0.45	273	46.8	11.1	90.0	42.5	131.0	3.5
8	4166	1866	2300	0.45	273	39.5	12.9	90.0	45.0	128.5	4.0
9	4000	1975	2025	0.49	266	39.0	12.3	89.0	45.0	128.5	4.0
10	2916	1500	1416	0.51	293	39.3	12.1	90.5	44.5	128.5	4.0
11	3000	1453	1547	0.48	273	39.8	12.0	91.5	43.0	129.0	3.5
12	3333	1658	1675	0.50	293	41.0	11.9	90.5	47.5	129.5	4.0
13	4250	2091	2159	0.49	333	41.8	12.5	90.5	42.5	128.5	4.5
14	2750	1490	1260	0.54	273	42.9	11.9	90.0	43.5	129.5	3.5
15	3000	1453	1547	0.48	300	41.0	11.5	90.5	46.0	129.0	4.0
16	3000	1345	1655	0.45	253	43.5	13.5	90.5	38.5	131.0	3.5
17	2833	1395	1438	0.49	253	44.8	11.9	90.5	41.0	130.0	3.5
18	3166	1628	1538	0.51	260	38.9	12.6	89.5	44.0	129.0	4.0
19	2916	1403	1513	0.48	286	37.8	12.0	90.0	44.0	129.5	3.5
20	3000	1450	1550	0.48	253	40.1	13.6	91.5	45.0	129.0	4.5
21	4416	2291	2125	0.52	313	39.4	12.6	90.0	43.5	129.5	4.5
22	3333	1311	2022	0.39	253	39.9	11.8	91.0	41.0	128.5	3.0
23	3416	1658	1758	0.49	266	46.0	12.3	91.0	41.5	129.5	3.5
24	3833	1908	1925	0.50	326	39.3	12.4	90.5	46.0	129.5	4.5
25	3333	1475	1858	0.44	313	40.3	13.0	90.0	43.5	129.0	4.5
26	3166	1466	1700	0.46	253	41.3	12.0	89.5	43.5	129.5	3.0

Table a.2. (contd.)

27	2916	1420	1496	0.49	280	38.5	13.5	91.0	38.5	129.5	4.5
28	3083	1533	1550	0.50	266	39.4	12.8	91.0	42.5	128.0	3.5
29	2750	1386	1364	0.50	226	41.5	12.1	90.5	41.5	129.0	3.5
30	3666	1775	1891	0.48	300	39.9	13.1	90.0	45.0	129.0	4.0
31	3333	1583	1750	0.47	253	40.0	11.0	91.0	40.5	127.0	3.5
32	3083	1486	1597	0.48	233	36.9	12.5	90.0	43.0	129.5	3.5
33	3166	1333	1833	0.42	266	40.0	12.9	89.5	45.0	129.0	3.0
34	3416	1746	1670	0.51	293	37.3	12.9	90.0	44.5	128.5	4.0
35	2916	1420	1496	0.49	280	39.1	12.0	90.0	42.0	129.0	3.0
36	2583	1375	1208	0.53	266	41.8	11.9	92.0	41.5	129.5	3.5
37	3083	1458	1625	0.47	253	41.4	12.2	92.0	41.0	129.5	4.0
38	3583	1786	1797	0.50	293	40.4	11.8	91.0	42.5	129.5	3.5
39	3416	1500	1916	0.44	226	43.8	12.3	91.0	38.0	129.5	3.0
40	3416	1650	1766	0.48	273	43.6	12.0	92.0	41.5	130.0	4.0
41	3750	1916	1834	0.51	293	39.9	13.0	89.5	42.5	129.5	4.5
42	3333	1653	1680	0.50	253	42.0	13.2	90.5	44.0	128.5	3.5
43	3500	1625	1875	0.46	293	37.3	13.4	89.5	49.5	128.0	3.5
44	3500	1633	1867	0.47	246	40.8	13.0	90.0	43.5	129.5	3.5
45	4083	1916	2167	0.47	260	41.0	11.0	89.5	40.0	129.5	3.5
46	4083	1933	2150	0.47	293	40.9	13.0	90.5	46.0	128.5	5.0
47	3166	1525	1641	0.48	253	43.4	13.0	91.0	41.5	129.0	4.0
48	2833	1500	1333	0.53	240	41.9	12.5	91.0	45.0	129.5	3.5
49	3333	1600	1733	0.48	300	45.0	13.5	90.5	46.0	128.5	3.0
50	3333	1486	1847	0.45	320	38.9	13.1	90.0	45.0	128.5	5.0
51	4000	2073	1927	0.52	293	41.0	12.4	91.5	45.0	128.5	4.5
52	3916	1908	2008	0.49	260	38.3	12.6	90.5	47.0	128.5	4.5
53	3916	1875	2041	0.48	280	40.8	13.4	90.0	48.5	128.5	3.5
54	3416	1666	1750	0.49	273	40.8	12.1	90.5	43.5	129.0	3.0
55	3416	1633	1783	0.48	240	40.9	12.4	90.5	37.5	130.0	4.5
56	3166	1478	1688	0.47	233	38.9	11.3	90.5	40.0	131.0	3.5

Table a.2. (contd.)

57	3083	1500	1583	0.49	266	37.6	12.4	90.0	43.0	129.0	4.0
58	3583	1716	1867	0.48	293	38.8	12.3	90.0	44.0	128.5	4.0
59	3000	1508	1492	0.50	273	39.5	12.9	90.5	44.0	129.0	3.5
60	3083	1486	1597	0.48	293	42.8	13.0	91.5	42.5	129.5	3.5
61	3000	1475	1525	0.49	246	40.9	12.5	92.5	41.0	130.0	3.5
62	3250	1561	1689	0.48	226	39.4	12.2	91.0	43.5	129.0	3.5
63	4000	1933	2067	0.48	320	37.7	12.6	91.0	48.5	130.0	4.5
64	3250	1570	1680	0.48	226	39.3	11.6	92.0	41.0	129.5	3.0
65	3250	1475	1775	0.45	293	40.7	12.8	90.5	46.0	129.5	5.0
66	2333	1220	1113	0.52	266	44.7	12.3	90.5	38.0	130.0	3.5
67	2450	1041	1409	0.42	226	40.5	12.2	89.0	41.0	128.5	4.0
68	2833	1366	1467	0.48	266	38.3	12.5	90.0	42.5	128.5	4.0
69	3833	1816	2017	0.47	246	38.1	11.5	90.0	46.0	129.5	4.0
70	3270	1458	1812	0.45	253	42.9	11.9	90.0	41.5	131.5	4.0
71	3883	1845	2038	0.48	260	38.7	12.7	92.5	40.0	129.5	4.0
72	3250	1736	1514	0.53	300	39.6	12.7	90.0	42.5	129.5	4.5
73	2666	1283	1383	0.48	226	42.8	12.5	90.0	42.5	129.5	3.5
74	3500	1773	1727	0.51	293	41.8	11.9	89.5	42.5	130.0	3.5
75	3083	1550	1533	0.50	273	43.0	12.4	89.0	40.0	130.0	3.5
76	3500	1666	1834	0.48	286	41.0	12.3	89.5	43.5	129.0	3.5
77	3000	1370	1630	0.46	246	42.4	12.6	90.0	41.5	129.0	4.0
78	3500	1700	1800	0.49	253	42.0	13.2	90.0	45.0	128.0	4.5
79	3666	1816	1850	0.50	286	35.1	13.0	90.0	47.5	128.5	4.0
80	3500	1645	1855	0.47	246	43.1	11.1	90.5	44.0	131.5	3.5
81	3333	1725	1608	0.52	266	41.1	12.0	90.0	42.5	130.0	4.0
82	4333	2066	2267	0.48	300	38.5	12.0	90.5	45.0	128.0	3.5
83	3166	1378	1788	0.44	273	41.0	14.0	90.0	41.0	129.0	4.0
84	3333	1620	1713	0.49	300	41.3	13.1	89.5	44.0	129.0	4.5
85	4000	2090	1910	0.52	306	40.4	12.8	91.0	45.0	129.0	4.5
86	3666	1708	1958	0.47	293	39.0	13.0	91.0	44.0	128.0	4.0

Table a.2. (contd.)

87	3538	1806	1732	0.51	286	42.3	12.9	90.0	42.5	129.5	3.5
88	3250	1595	1655	0.49	286	39.8	12.6	90.5	41.5	130.0	4.5
89	3333	1756	1577	0.53	260	40.3	12.8	90.5	45.0	128.5	3.0
90	2916	1408	1508	0.48	253	42.0	12.3	91.0	41.0	129.5	3.0
91	3500	1908	1592	0.55	273	39.8	12.3	89.5	44.0	129.5	3.0
92	3333	1653	1680	0.50	286	43.5	12.5	89.5	42.5	130.0	4.5
93	3000	1358	1642	0.45	286	40.3	12.5	91.0	42.5	130.0	4.0
94	3666	1928	1738	0.53	286	41.5	12.3	91.0	44.5	129.0	4.0
95	3666	1683	1983	0.46	266	42.9	12.4	90.5	46.5	130.0	4.5
96	2750	1428	1322	0.52	226	40.5	12.4	90.0	41.5	129.0	3.5
97	3250	1516	1734	0.47	260	38.9	12.2	90.0	42.5	128.5	3.0
98	3500	1875	1625	0.54	260	41.9	11.6	90.0	43.5	131.5	3.5
99	3416	1741	1675	0.51	280	41.1	12.0	90.5	38.5	129.0	3.5
100	3833	1833	2000	0.48	240	40.1	12.4	91.5	43.5	129.5	3.0
101	3416	1541	1875	0.45	306	38.1	13.2	90.0	45.0	128.5	4.0
102	3583	1895	1688	0.53	280	45.3	12.4	90.5	44.0	129.5	3.5

Table a.3. Grain yield for H2 in rank order and family number at 2 sites.

family	Tel Hadya season 1		Tel Hadya season 2		Breda season 2		Total of ranks
	Grain yield	Rank	Grain yield	Rank	Grain yield	Rank	
1	3866	8	2808	59	1540	78	145
2	2813	46	1650	101	1691	63	210
3	2635	61	2691	70	1916	26	157
4	2728	55	2650	75	1666	67	197
5	2906	39	4350	2	1845	36	77
6	2380	79	2333	95	1866	33	207
7	2795	47	2800	60	916	102	209
8	2533	66	2583	82	2166	9	157
9	2875	41	4141	4	1716	57	102
10	3937	7	3143	30	1358	90	127
11	2844	43	3708	9	2316	2	54
12	2733	53	2933	45	1791	44	142
13	1604	99	2635	78	2225	4	181
14	3786	10	2975	41	1308	94	145
15	1982	94	2760	63	1341	93	250
16	1939	97	2816	58	2053	13	168
17	3057	28	3058	33	1586	73	134
18	2351	82	3225	25	1575	74	181
19	2986	33	3158	29	2091	12	74
20	3084	26	3166	28	1808	42	96
21	3053	30	2950	42	1350	92	164
22	3315	19	4075	5	1725	52	76
23	1955	96	3426	17	2191	8	121
24	2053	92	2350	92	1441	86	270
25	2933	37	3466	15	2141	10	62
26	2964	36	2758	64	1675	65	165
27	3822	9	2893	51	1491	82	142
28	3195	22	3235	24	2575	1	47
29	2400	77	2733	68	1733	50	195
30	3209	21	3283	21	1600	72	114
31	2315	84	2401	91	1020	100	275
32	2408	76	2683	71	1375	89	236
33	3511	14	2591	81	1633	68	163
34	3124	25	3050	34	2216	5	64
35	3480	15	2333	94	1466	84	193
36	2880	40	2920	48	1700	59	147
37	2555	65	2866	54	1683	64	183
38	1982	95	2350	93	1725	53	241
39	2706	56	1943	99	1816	40	195
40	4160	4	2476	88	1825	37	129
41	2431	74	3476	14	1816	41	129
42	1577	102	2658	74	2111	11	187
43	1880	98	2701	69	1791	45	212
44	2257	87	2976	40	1753	49	176
45	3515	13	2900	49	1886	28	90
46	4791	2	3183	26	2025	14	42

Table a.3. (contd.)

family	Tel Hadya season 1		Tel Hadya season 2		Breda season 2		Total of ranks
	Grain yield	Rank	Grain yield	Rank	Grain yield	Rank	
47	1600	101	3116	31	1766	46	178
48	3164	24	2560	85	1825	38	147
49	2440	73	3500	13	1520	79	165
50	2613	62	2883	52	1895	27	141
51	2488	69	3050	35	1358	91	195
52	2377	80	3775	8	1695	61	149
53	2155	90	2800	61	1283	95	246
54	2991	32	3583	11	2025	15	58
55	4453	3	2816	56	1758	48	107
56	2200	89	3000	37	1958	17	143
57	2448	71	1826	100	1950	19	190
58	3582	11	2408	90	2275	3	104
59	2817	45	3441	16	1725	54	115
60	2773	48	4066	6	1461	85	139
61	2586	64	3500	12	1966	16	92
62	3177	23	3966	7	1945	20	50
63	4840	1	2533	87	1925	25	113
64	2271	86	2793	62	1766	47	195
65	3400	17	2641	77	1233	97	191
66	3555	12	2566	84	1550	75	171
67	2848	42	2666	72	1928	23	137
68	3035	31	2933	44	1516	80	155
69	3217	20	2983	38	1883	29	87
70	2666	59	2925	47	2200	7	113
71	2822	44	2025	98	1545	77	219
72	3071	27	2933	43	1861	34	104
73	2604	63	2933	46	1875	31	140
74	2520	67	3283	22	1825	39	128
75	2671	58	2900	50	1716	58	166
76	1604	100	2658	73	1508	81	254
77	2973	34	2741	67	1416	87	188
78	2044	93	2535	86	1608	70	249
79	2471	70	2100	97	1250	96	263
80	4040	6	3235	23	983	101	130
81	2706	57	2979	39	1695	62	158
82	2755	51	3385	20	1700	60	131
83	3057	29	2816	57	1875	32	118
84	2502	68	3400	19	1883	30	117
85	2097	91	5035	1	1725	55	147
86	2773	49	2616	80	1953	18	147
87	3368	18	2741	66	1800	43	127
88	2640	60	2866	53	1603	71	184
89	2346	83	1583	102	1161	98	283
90	2373	81	2296	96	1725	56	233
91	2968	35	2750	65	1475	83	183
92	2431	75	2850	55	1945	21	151
93	2444	72	3608	10	1928	24	106
94	2924	38	2616	79	1858	35	152

Table a.3. (contd.)

family	Tel Hadya season 1		Tel Hadya season 2		Breda season 2		Total of ranks
	Grain yield	Rank	Grain yield	Rank	Grain yield	Rank	
95	2773	50	3026	36	1550	76	162
96	2213	88	2643	76	2216	6	170
97	2395	78	3418	18	1633	69	165
98	2733	54	3183	27	1383	88	169
99	3440	16	3083	32	1728	51	99
100	4128	5	2566	83	1675	66	154
101	2315	85	2475	89	1111	99	273
102	2751	52	4208	3	1941	22	77

Table a.4. Grain yield for H3 in rank order and family number at 2 sites.

family	Tel Hadya season 1		Tel Hadya season 2		Breda season 2		Total of ranks
	Grain yield	Rank	Grain yield	Rank	Grain yield	Rank	
1	2235	73	3100	27	1523	61	161
2	3155	17	2600	62	1661	39	118
3	2631	45	2158	91	1470	76	212
4	2537	54	3358	16	1325	98	168
5	2711	37	2033	97	1433	83	217
6	3257	14	3793	3	1750	28	45
7	2053	80	3091	28	1583	53	161
8	1795	89	2091	93	1866	18	200
9	2382	65	2683	49	1975	6	120
10	3155	18	2425	75	1500	64	157
11	2311	68	3116	25	1453	80	173
12	3142	19	3450	13	1658	40	72
13	3302	10	2533	68	2091	2	80
14	3297	12	2408	78	1490	68	158
15	2133	77	2650	54	1453	81	212
16	2791	34	1533	102	1345	96	232
17	1866	87	4150	1	1395	89	177
18	2942	27	2616	58	1628	48	133
19	2368	66	2591	66	1403	88	220
20	2826	32	2600	64	1450	82	178
21	3702	4	3585	8	2291	1	13
22	2124	78	2510	70	1311	99	247
23	2388	64	2683	48	1658	41	153
24	2542	53	3075	29	1908	12	94
25	1671	95	2610	61	1475	73	229
26	2813	33	2216	89	1466	77	199
27	2671	44	2260	87	1420	85	216
28	3413	8	2925	36	1533	59	103
29	2462	58	2791	41	1386	90	189
30	1684	94	3166	24	1775	25	143
31	2608	47	2233	88	1583	54	189
32	1688	93	2060	96	1486	69	258
33	2951	26	2400	80	1333	97	203
34	3453	7	2616	57	1746	29	93
35	3284	13	3700	4	1420	86	103
36	2844	31	2785	42	1375	92	165
37	2697	41	2666	50	1458	78	169
38	2551	51	2416	77	1786	24	152
39	3137	20	3551	10	1500	65	95
40	2604	48	3033	31	1650	44	123
41	2471	57	2908	37	1916	10	104
42	1777	90	2926	35	1653	42	167
43	2991	24	2701	47	1625	49	120
44	1631	97	2125	92	1633	46	235
45	2875	29	3326	19	1916	11	59
46	2271	69	2908	38	1933	7	114

Table a.4. (contd.)

family	Tel Hadya season 1		Tel Hadya season 2		Breda season 2		Total of ranks
	Grain yield	rank	Grain yield	rank	Grain yield	rank	
47	1906	86	2486	72	1525	60	218
48	2177	74	2951	33	1500	66	173
49	3302	11	2500	71	1600	51	133
50	2062	79	2600	65	1486	70	214
51	2502	55	2075	95	2073	4	154
52	1697	92	3566	9	1908	13	114
53	2475	56	3350	17	1875	16	89
54	2697	42	1683	101	1666	37	180
55	2906	28	2433	74	1633	47	149
56	2164	75	2850	40	1478	72	187
57	1382	102	3650	7	1500	67	176
58	2848	30	2600	63	1716	33	126
59	3208	16	2326	81	1508	63	160
60	2417	61	3450	14	1486	71	146
61	2426	60	2450	73	1475	74	207
62	2440	59	2088	94	1561	56	209
63	2137	76	1841	100	1933	8	184
64	3102	21	2658	52	1570	55	128
65	2706	38	2758	45	1475	75	158
66	2248	72	3891	2	1220	101	175
67	4142	2	3383	15	1041	102	119
68	2564	50	2523	69	1366	94	213
69	2697	43	2775	44	1816	21	108
70	2257	70	2950	34	1458	79	183
71	4102	3	1950	99	1845	19	121
72	2751	36	2425	76	1736	31	143
73	1546	99	3108	26	1283	100	225
74	2702	40	2408	79	1773	26	145
75	2035	81	3316	21	1550	57	159
76	1466	101	3550	11	1666	38	150
77	1720	91	2726	46	1370	93	230
78	2400	63	3333	18	1700	35	116
79	3240	15	2283	84	1816	22	121
80	3511	5	3183	23	1645	45	73
81	2991	25	2316	83	1725	32	140
82	1644	96	2616	60	2066	5	161
83	1617	98	2583	67	1378	91	256
84	1968	83	3660	6	1620	50	139
85	2631	46	2616	59	2090	3	108
86	2355	67	2875	39	1708	34	140
87	4677	1	2616	56	1806	23	80
88	3355	9	2658	51	1595	52	112
89	3484	6	2260	86	1756	27	119
90	2706	39	2783	43	1408	87	169
91	1986	82	3186	22	1908	14	118
92	2253	71	3326	20	1653	43	134
93	2408	62	1958	98	1358	95	255
94	2586	49	2658	53	1928	9	111

Table a.4. (contd.)

family	Tel Hadya season 1		Tel Hadya season 2		Breda season 2		Total of ranks
	Grain yield	Rank	Grain yield	Rank	Grain yield	Rank	
95	1920	85	3668	5	1683	36	126
96	1844	88	2325	82	1428	84	254
97	1515	100	3050	30	1516	62	192
98	3093	22	2633	55	1875	17	94
99	1946	84	2970	32	1741	30	146
100	3088	23	3500	12	1833	20	55
101	2773	35	2266	85	1541	58	178
102	2551	52	2183	90	1895	15	157

References

References

Absalom, J.P., Young, S.D., Crout, N.M.J., Sanchez, A., Wright, S.M., Smolders, E., Nisbet, A.F. and Gillett, A.G. 2001. Predicting the transfer of radiocaesium from organic soils to plants using soil characteristics. *Journal of Environmental Radioactivity* 52: 31-43.

Adams, M.L., Zhao, F.J., McGrath, S.P., Nicholson, F.A. and Chambers, B.J. 2004. Predicting cadmium concentrations in wheat and barley grain using soil properties. *Journal of Environmental Quality* 33: 532-541.

Aiken, L.S. and West, S.G. 1991. *Multiple regression : testing and interpreting interactions.* Sage, xi,211pp.

Akqura, M., Kaya, Y. and Taner, S. 2004. Genotype-Environment Interaction and Phenotypic Stability Analysis for Grain Yield of Durum Wheat in the Central Anatolian Region. *Bahri Dağdaş International Agricultural Research Institute* 29: 369-375.

Allison, P.D. 1999. *Multiple regression : a primer.* Pine Forge Press, Thousand Oaks, Calif. ; London xviii, 202 pp.

Annicchiarico, P. 2002. *Genotype x Environment Interactions - Challenges and Opportunities for Plant Breeding and Cultivar Recommendations.* [on-line] Available: http://www.fao.org/documents/show_cdr.asp?url_file=/DOCREP/005/Y4391E/y4391e04.htm [date accessed: 22/3/205]

Anon 1995. *Factor Analysis Using SAS PROC FACTOR.* [on-line] Available: www.utexas.edu/cc/docs/stat53.html [date accessed: 5/8/2004]

Anon 1999. *Barley/ Malt/ Beer.* [on-line] Available: http://www.fao.org/tc/tci/sectors/beer.pdf [date accessed: 2/6/2004]

Anon 2002a. *Syria Grain and Feed Annual 2002.* [on-line] Available: [date accessed: 1/4/2007]

Anon 2002b. *Syria Grain and Feed Annual.* [on-line] Available: http://www.fas.usda.gov/gainfiles/200310/145986466.pdf [date accessed: 01/04/2007]

Anon 2004a. *Canonical Analysis.* [on-line] Available: www.worldhistory.com/wiki/c/canonical-correlation.htm [date accessed: 27/7/2004]

Anon 2004b. *Canonical Correlation.* [on-line] Available: http://www2.chass.ncsu.edu/garson/pa765/canonic.htm [date accessed: 6/9/2004]

Anon 2004c. *Principal Components and Factor Analysis.* [on-line] Available: http://www.statsoftinc.com/textbook/stfacan.html [date accessed: 28/6/2004]

Anon 2005. *General Linear Models (GLM).* [on-line] Available: http://www.statsoft.com/textbook/stglm.html#basic_ideas [date accessed: 10/7/2007]

Anon 2006a. *Barley: Genus Hordeum, family poaceae, a cereal grain*. [on-line] Available: http://www.grains.org/page.ww?section=Barley%2C+Corn+%26+Sorghum&name=Barley [date accessed: 5/04/2007]

Anon 2006b. *Syria Grain and Feed Report 2006-10-16*. [on-line] Available: http://damascus.usembassy.gov/uploads/images/L6Nef5epvB9DWVM4Z1wZRA/Syria_Grain_and_Feed_Report_2006.pdf [date accessed: 01/04/2007]

Anon 2007. *Barley/Malt/Beer*. [on-line] Available: www.fao.org [date accessed: 2/2/2007]

Arnold, T., Kraakman, W., Niks, R.E., M., P.M.M., Stam., P. and A., F. 2004. Linkage Disequilibrium Mapping of Yield and Yield Stability in Modern Spring Barley Cultivars *Genetics* 168: 435-446.

Baker, R.J. 1988. Tests for Crossover Genotype-Environmental Interactions. *Canadian Journal of Plant Science* 68: 405-410.

Bala, B.K. and Woods, J.L. 1992. Thin-Layer Drying Models for Malt. *Journal of Food Engineering* 16: 239-249.

Bartos, A., Fekete, A. and Sarvari, B. 1991. Study of Ecological Factors and Nutrient-Content Variables in Wheat Applying Canonical Correlation-Analysis. *Novenytermeles* 40: 111-123.

Bartos, A. and Szanyi, M. 1992. Study of Yield and Chemical-Composition of Winter Barley by Factor-Analysis in a Nitrogen-Fertilization Trial. *Novenytermeles* 41: 511-523.

Barua, U.M., Chalmers, K.J., Thomas, W.T.B., Hackett, C.A., Lea, V., Jack, P., Forster, B.P., Waugh, R. and Powell, W. 1993. Molecular Mapping of Genes Determining Height, Time to Heading, and Growth Habit in Barley (Hordeum-Vulgare). *Genome* 36: 1080-1087.

Beaulieu, J.M. 2004. Utilisation of contour criteria in micro-segmentation of SAR images. *International Journal of Remote Sensing* 25: 3497-3512.

Beavan, E.S. 1947. *Barley*. Duckworth, London.

Berri, G.J. and Bertossa, G.I. 2004. The influence of the tropical and subtropical Atlantic and Pacific Oceans on precipitation variability over southern central South America on seasonal time scales. *International Journal of Climatology* 24: 415-435.

Bertholdsson, N.O. 1999. Characterization of malting barley cultivars with more or less stable grain protein content under varying environmental conditions. *European Journal of Agronomy* 10: 1-8.

Blummel, M. and Orskov, E.R. 1993. Comparison of Invitro Gas-Production and Nylon Bag Degradability of Roughages in Predicting Feed-Intake in Cattle. *Animal Feed Science and Technology* 40: 109-119.

Borhan, M.S., Panigrahi, S., Lorenzen, J.H. and Gu, H. 2004. Multispectral and color imaging techniques for nitrate and chlorophyll determination of potato leaves in a controlled environment. *Transactions of the Asae* 47: 599-608.

Breier, A., Drobna, Z., Docolomansky, P. and Barancik, M. 2000. Cytotoxic activity of several unrelated drugs on L1210 mouse leukemic cell sublines with P-glycoprotein (PGP) mediated multidrug resistance (MDR) phenotype. A QSAR study. *Neoplasma* 47: 100-106.

Briggs, D.E. 1978. *Barley.* Chapman and Hall, London. xviii,612pp.

Burton, R.F. (Ed.) 1998. *Biology by numbers : an encouragement to quantitative thinking.* Cambridge University Press, Cambridge. xvi, 238pp

Butt, M.S., Anjum, F.M., van Zuilichem, D.J. and Shaheen, M. 2001. Development of predictive models for end-use quality of spring wheats through canonical analysis. *International Journal of Food Science and Technology* 36: 433-440.

Cannon, A.R. and St John, C.H. 2004. Competitive strategy and plant-level flexibility. *International Journal of Production Research* 42: 1987-2007.

Chatterjee, S. and Price, B. 1977. *Regression analysis by example.* Wiley, New York. xv,228ppp.

Chatterjee, S. and Price, B. 1991. *Regression analysis by example.* 2nd ed. Wiley, xvii, 278 pp.

Chenevert, D. and Tremblay, M. 2002. Role of external and internal strategies in the choice of compensation policies. *Relations Industrielles-Industrial Relations* 57: 331-353.

Clark, D. 1977. *Understanding canonical correlation analysis* Geo Abstracts Ltd, Norwich 36pp.

Cohen, J. and Cohen, P. 1975. *Applied multiple regression correlation analysis for the behavioral sciences* Lawrence Erlbaum Associates, Hillsdale xxi,490pp.

Counsell, S., Newson, P. and Mendes, E. 2004. Design level hypothesis testing through reverse engineering of object-oriented software. *International Journal of Software Engineering and Knowledge Engineering* 14: 207-220.

De Cerio, J.M.D. 2003. Quality management practices and operational performance: empirical evidence for Spanish industry. *International Journal of Production Research* 41: 2763-2786.

Dehghania, H., Ebadia, A. and Yousefib, A. 2006. Biplot Analysis of Genotype by Environment Interaction for Barley Yield in Iran *Agronomy Journal* 98: 388-393.

Denis, J.C. and Adams, M.W. 1978. A factor analysis of plant variables related to yield in dry beans. I. Morphological traits. *Crop Sci.* 18: 74-78.

Dennis, C., Hurd, R.G.R.G., 1932- and Biscoe, P.V. 1980. *Opportunities for increasing crop yields : taken from the proceedings of a meeting held at the University of Reading, 17-21 September 1979, to celebrate the 75th anniversary meeting of the Association of Applied Biology [i.e. Biologists].* Pitman, Boston. xi,410pp.

Dharmakeerthia, R.S., Kayb, B.D. and Beauchampb, E.G. 2006. Spatial Variability of In-Season Nitrogen Uptake by Corn Across a Variable Landscape as Affected by Management *Agronomy Journal* 98: 255-264.

Dillon, W.R. and Goldstein, M. 1984. *Multivariate analysis : methods and applications* Wiley, xii,587pp.

Eddison, J. 2000. *Quantitative investigations in the biosciences using Minitab.* Chapman & Hall, Boca Raton, Fla. XI, 458pp.

Fischer, T. and Kamerschen, D.R. 2003. Price-cost margins in the US airline industry using a conjectural variation approach. *Journal of Transport Economics and Policy* 37: 227-259.

Fox, G., Onley-Watson, K., Ferguson, R., Skerman, A., Poulsen, D., Inkerman, A. and Johnston, P. 2006. *Grain and Malt Quality Evaluation from a Structured Environmental Trial.* [on-line] Available: http\\.www.Grain and Malt Quality Evaluation from a Structured Environmental Trial.htm [date accessed: 22/12/2006]

Fox, G.P., Onley-Watson, K. and Osman, A. 2002. Multiple linear regression calibrations for barley and malt protein based on the spectra of hordein. *Journal of the Institute of Brewing* 108: 155-159.

Francis, W., Wilson, N.E., Joseph, O. and George, M. 2002. Genotype X environment interactions for tea yields. *Euphytica (Euphytica) ISSN 0014-2336* 127: 289-296

Fraser, J. and Eato, G.E. 1983. Application of Yield cmponent analysis to crop research *Filed Crop* 36: 787-797.

Goodchild, A.V. 1997. Effects of rainfall and temperature on the feeding value of barley straw in a semi-arid Mediterranean environment. *Journal of Agricultural Science* 129: 353-366.

Group, S. 2005. *Factor analysis.* [on-line] Available: http://www.mathworks.com/products/demo/statistics/factorandemo.html [date accessed: 15/12/2005]

Group, S. 2006. *Factor Analysis.* [on-line] Available: http://www2.chass.ncsu.edu/garson/pa765/factor.htm [date accessed: 31/10/2005]

Hacett, C.A., Meyer, R.C. and Thomas, W.T.B. 2001. Multi-trait QTL mapping in barley using multivariate regression. *Genet. Res., Camb.* 77: 95-106.

Hair, J.F., Black, W.C., Anderson, R.E. and Tatham, R.L. 1995. *Multivariate data analysis with readings.* 4th ed ed. Prentice Hall International, 745pp.

Hair, J.F., Black, W.C., Babin, B.G., Anderson, R.E. and Tatham, R.L. 2005. *Multivariate data analysis.* 6th ed ed. Prentice Hall International, Upper Saddle River, NJ xxiv, 899 pp.

Harman, H.H. 1967. *Modern factor analysis.* 2na ed. ed. University of Chicago Press,

Hay, R.K.M., 1946- and Walker, A.J., 1951- 1989. *An introduction to the physiology of crop yield.* Longman Scientific & Technical, 400pp.

Hox, J.J. 1995. *Applied Multilevel Analysis.* Cip-gegevens koninklijke bibliotheek, The Hague. 119pp.

Ibekwe, A.M. and Kennedy, A.C. 1998. Fatty acid methyl ester (FAME) profiles as a tool to investigate community structure of two agricultural sails. *Plant and Soil* 206: 151-161.

Jardine, R., Moss, H.J. and Mullaly, J.V. 1971. An application of factor analysis to some breadmaking. quality data. *Crop Sci.* 12: 44-46.

Jeffrey, M.S. 2001. Galton, Pearson, and the Peas: A Brief History of Linear Regression for Statistics Instructors.*Journal of Statistics Education* 9:

Johnson, G.F. and Whittington, W.J. 1977. Genotype-environment interaction effects in F1 barley hybrids *Euphytica* 26: 67-73.

Johnson, R.A. and Wichern, D.W. 1998. *Applied multivariate statistical analysis.* 4th ed ed. Prentice-Hall, 816pp.

Kalivas, D.P. and Kollias, V.J. 2001. Effects of soil, climate and cultivation techniques on cotton yield in Central Greece, using different statistical methods. *Agronomie* 21: 73-89.

Kang, M.S. 2002. *quantitative genetics, genomics and plant breeding.* CAB international New York. 400pp.

Kebebew, F., Tschaye, Y. and McNeilly, T. 2001. Morphological and farmers cognitive diversity of barley (Hordeum vulgare L. [Poaceae]) at Bale and North Shewa of Ethiopia. *Genetic Resources and Crop Evolution* 48: 467-481.

Kendall, M.G. 1975. *Multivariate analysis* Griffin, London [7],210pp.

Kim, J.-o. and Mueller, C.W. 1978. *Factor analysis : statistical methods and practical issues.* Sage Publications, Beverly Hills, Calif 88pp.

Kline, P. 1994. *An easy guide to factor analysis.* Routledge, viii,194pp.

Korobov, R.M. and Railyan, V.Y. 1993. Canonical Correlation Relationships among Spectral and Phytometric Variables for 20 Winter-Wheat Fields. *Remote Sensing of Environment* 43: 1-10.

Krzanowski, W.J. 1990. *Principles of multivariate analysis* Clarendon Press, Oxford.

Krzanowski, W.J. and Marriott, F.H.C. 1994. *Multivariate analysis* Edward Arnold, London. ix,280pp.

Lagudah, E.S., Dubcovsky, J. and Powell, W. 2001. Wheat genomics. *Plant Physiology and Biochemistry* 39: 335-344.

Levine, M.S. 1977. *Canonical analysis and factor comparison*. Sage Publications, Beverly Hills ; London. 62,[1]ppp.

Lungu, D.M., Kaltsikes, P.J. and Larter, E.N. 1990. Intra- and intergeneration relationships among yield, its components and other related characteristics in spring wheat. *Euphytica (Euphytica) ISSN 0014-2336* 45: 139-153.

Lutkenhoner, B. 2003. Magnetoencephalography and its Achilles' heel. *Journal of Physiology-Paris* 97: 641-658.

Manly, B.F.J. 1994. *Multivariate statistical methods : a primer*. 2nd ed ed. Chapman and Hall, xiii,215pp.

Marrari, M., Vinas, M.D., Martos, P. and Hernandez, D. 2004. Spatial patterns of mesozooplankton distribution in the Southwestern Atlantic Ocean (34 degrees-41 degrees S) during austral spring: relationship with the hydrographic conditions. *Ices Journal of Marine Science* 61: 667-679.

Miller, B.W., Roberts, G.C. and Ommundsen, Y. 2004. Effect of motivational climate on sportspersonship among competitive youth male and female football players. *Scandinavian Journal of Medicine & Science in Sports* 14: 193-202.

Naesah, G. 1996. *Inhalants of quality characters of three barley hybrids*. Aleppo University, Aleppo.

Nasser, M. 2002. *Applied Multivariate analysis and SPSS in Statistical Analysis*. Damascus Uni., Damascus. 145pp.

Nathierdufour, N., Angue, Y., Devaux, M.F., Bertrand, D. and Ledeschaultdemonredon, F. 1995. Influence of Wheat Meal Variability Upon Compacting Behavior During Pelleting. *Animal Feed Science and Technology* 51: 255-268.

Pereira, R.B., Singh, H., Jones, V.S. and Munro, P.A. 2004. Relationship between oral and nonoral evaluation of texture in acid milk gels. *Journal of Sensory Studies* 19: 67-82.

Phillips, W.J. 2002. *Simple Linear Regression*. [on-line] Available: http://www.engmath.dal.ca/courses/engm6671/notes/node1.html [date accessed: 21/2/2005]

Qaizar., A., Rana, M.A. and Siddiqui, S.U.H. 1991. Sunflower seed yield as influenced by some agronomic and seed characters. *Euphytica (Euphytica) ISSN 0014-2336* 56: 137-142.

Rawlings, J.O., Pantula, S.G. and Dickey, D.A. 1998. *Applied regression analysis : a research tool* 2nd ed ed. Springer, New York xviii, 657 pp.

Recio, I., Garcia-Risco, M.R., Amigo, L., Molina, E., Ramos, M. and Martin-Alvarez, P.J. 2004. Detection of milk mixtures in Halloumi cheese. *Journal of Dairy Science* 87: 1595-1600.

Saadat, M. and Bahaoddini, A. 2004. Hematological changes due to chronic exposure to natural gas leakage in polluted areas of Masjid-i-Sulaiman (Khozestan province, Iran). *Ecotoxicology and Environmental Safety* 58: 273-276.

Schillinger, W.F., Cook, R.J. and Papendick, R.I. 1999. Increased dryland cropping intensity with no-till barley. *Agronomy Journal* 91: 744-752.

Shafto, M.G., Degani, A. and Kirlik, A. 2005. *Canonical Correlation Analysis of Data on Human-Automation Interaction.* [on-line] Available: http://human-factors.arc.nasa.gov/IHpublications/shafto/canonical-corr/canonical-corr.html [date accessed: 31/10/2005]

Shakhatreh, Y., Kafawin, O., Ceccarelli, S. and Saoub, H. 2001. Selection of Barley lines for Drought Tolerance in Low-Rainfall Areas. *J.Agronomy& Crop Science* 1: 119-127.

Sharratt, B.S., Knight, C.W. and Wooding, F. 2003. Climatic impact on small grain production in the Subarctic region of the United States. *Arctic* 56: 219-226.

Shier, R. 2004. *The Wilcoxon signed rank sum test.* [on-line] Available: http://mlsc.lboro.ac.uk/documents/wsrt.pdf [date accessed: 15-1-2007]

ShouFu, X., FengJun, W. and RunSheng, J. 2007. *Correlation Analysis of Several Quantitative Characters of Barley.* [on-line] Available: http://wheat.pw.usda.gov/ggpages/bgn/27/xf1txt.html [date accessed: 27/4/2007]

Shukla, G.K. 1972. Some statistical aspects of partitoning genotype-enveronmental component of variability. *Heredity* 29: 237-245.

Sinebo, W. 2002. Yield Relationships of Barleys Grown in a Tropical Highland Environment *Crop Sci* 42: 428-437.

Smith, A., Cullis, B. and Thompson, R. 2001. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57: 1138-1147.

Smith, R.C.G., Wallace, J.F., Hick, P.T., Gilmour, R.F., Belford, R.K., Portmann, P.A., Regan, K.L. and Turner, N.C. 1993. Potential of Using Field Spectroscopy During Early Growth for Ranking Biomass in Cereal Breeding Trials. *Australian Journal of Agricultural Research* 44: 1713-1730.

Soffe, R.J. 2003. *Primrose McConnell's the agricultural notebook.* 20th ed. Blackwell Science, Oxford. xvi, 744pp.

StatSoft, I. 2006. *Text book.* [on-line] Available: www.statsoft.com/textbook/glosh.html [date accessed: 2005]

Swinscow, T.D.V. and Campbell, M.J. 1996. *Statistics at square one.* 9th ed / ed. BMJ, London. viii, 140 ppp.

Tabachnick, B.G. and Fidell, L.S. 2001. *Using multivariate statistics* 4th ed. Allyn and Bacon, Boston, MA

Tadesse, W. and Bekele, E. 2001. Factor analysis of components of yield in grasspea (Lathyrus sativus L.). *Lathyrus Lathyrism Newsletter 2 (2001)* 2: 91-93.

Thompson, B. 2004. *Exploratory and confirmatory factor analysis :understanding concepts and applications.* American Psychological Association, Washington, DC.

Timm, N.H. 2002. *Applied multivariate analysis* Springer,

Toker, C. and Cagirgan, M.I. 2004. The use of phenotypic correlation and factor analysis in determining characters for grain yield selection in chickpea(Cicer arietinum L.). *Hereditas* 140: 226-228.

Tworek, S. 2002. Different bird strategies and their responses to habitat changes in an agricultural landscape. *Ecological Research* 17: 339-359.

Unay, A., Konak, C. and Sezener, V. 2004. Stability analysis for malting barley in aegean region. *Cereal Research Communications* 32: 39-43.

Utexas 1995. *Factor Analysis Using SAS PROC FACTOR.* [on-line] Available: http://www.utexas.edu/cc/docs/stat53.html [date accessed: 28/6/2004]

Vadiveloo, J. and Phang, O.C. 1996. Differences in the nutritive value of two rice straw varieties as influenced by season and location. *Animal Feed Science and Technology* 61: 247-258.

Valerie, J.E. and McColl, J.H. 2004. *Hypothsies Testing.* [on-line] Available: http://www.cas.lancs.ac.uk/glossary_v1.1/hyptest.html [date accessed: 9/4/2005]

Vanoosterom, E.J., Ceccarelli, S. and Peacock, J.M. 1993a. Yield Response of Barley to Rainfall and Temperature in Mediterranean Environments. *Journal of Agricultural Science* 121: 307-313.

Vanoosterom, E.J., Kleijn, D., Ceccarelli, S. and Nachit, M.M. 1993b. Genotype-by-environment interactions of Barley in the Mediterranean region. *Crop Sci* 33: 669-674

Volis, S., Mendlinger, S. and Ward, D. 2002. Differentiation along a gradient of environmental productivity and predictability in populations of Hordeum spontaneum Koch: multilevel selection analysis. *Biological Journal of the Linnean Society* 75: 313-318.

Walsh, S.J., Evans, T.P., Welsh, W.F., Entwisle, B. and Rindfuss, R.R. 1999. Scale-dependent relationships between population and environment in northeastern Thailand. *Photogrammetric Engineering and Remote Sensing* 65: 97-105.

Walton, P.D. 2004. Factor analysis of yield in spring wheat (Triticum aestivum L.). *Crop Sci* 12: 731-733.

Wassouf, M.Z. 1996. Multivariate analysis of yield,yield components, morphological and quality characters in durum what. *Agri. Sciences Series* 25: 11-25.

Westcott, B. 1987. Amethod of assessing the yield stability of crop genotype *Journal of Agricaltural Scinece* 108: 267-274.

Wilson, G. and Sankaran, K. 1997. *Hypothesis Testing.* [on-line] Available: http://www.cee.vt.edu/program_areas/environmental/teach/smprimer/hypotest/ht.html [date accessed: 1/6/2004]

Wu, C.H. and Yan, G.L. 2004. Acoustic feature analysis and discriminative modeling of filled pauses for spontaneous speech recognition. *Journal of Vlsi Signal Processing Systems for Signal Image and Video Technology* 36: 91-104.

Xue, Q.W., Weiss, A. and Baenziger, P.S. 2004. Predicting leaf appearance in field-grown winter wheat: evaluating linear and non-linear models. *Ecological Modelling* 175: 261-270.

Yates, S. and Ramirez-Sosa, C.R. 2004. Ethnobotanical knowledge of Brosimum alicastrum (Moraceae) among urban and rural El Salvadorian adolescents. *Economic Botany* 58: 72-77.

Zhang, Q.F., Maroof, M.A.S. and Webster, R.K. 1991. Spatial and Temporal Patterns of Associations between Quantitative Characters and Resistance to Scald in Barley. *Hereditas* 115: 1-8.

Zhang, W.J., Campbell, L.D. and Stothers, S.C. 1994. An Investigation of the Feasibility of Predicting Nitrogen- Corrected True Metabolizable Energy (Tmen) Content in Barley from Chemical-Composition and Physical Characteristics. *Canadian Journal of Animal Science* 74: 355-360.

Zhu, H., Briceno, G., Dovel, R., Hayes, P.M., Liu, B.H., Liu, C.T. and Ullrich, S.E. 1999. Molecular breeding for grain yield in barley: an evaluation of QTL effects in a spring barley cross. *Theoretical and Applied Genetics* 98: 772-779.