



Monitoring gender remuneration inequalities in academia using biplots

IS Walters*

NJ le Roux†

Received: 14 December 2007; Revised: 8 April 2008; Accepted: 6 May 2008

Abstract

Gender remuneration inequalities at universities have been studied in various parts of the world. In South Africa, the responsibility largely rests with individual higher education institutions to establish levels of pay for male and female academic staff members. The multidimensional character of the gender wage gap includes gender differentials in research output, age, academic rank and qualifications. The aim in this paper is to demonstrate the use of modern biplot methodology for describing and monitoring changes in the gender remuneration gap over time. A biplot is considered as a multivariate extension of an ordinary scatterplot. Our case study includes the permanent fulltime academic staff at Stellenbosch University for the period 2002 to 2005. We constructed canonical variate analysis (CVA) biplots with 90% alpha bags for the five-dimensional data collected for males and females in 2002 and 2005 aggregated over faculties as well as for each faculty separately. The biplots illustrate, for our case study, that rank, age, research output and qualifications are related to remuneration. The CVA biplots show narrowing, widening and constant gender remuneration gaps in different faculties.

Key words: alpha-bag, academia, biplot, canonical variate analysis, gender inequality, multivariate, wage gap.

1 Introduction

In the 1960s, it became necessary to pass several laws and regulations addressing gender discrimination in salaries and fringe benefits in college and university faculties in the USA (see [7]). Since the passage of this legislation several studies have been conducted, focusing on what is called by Toutkoushian [39] the “total wage gap” in academia or the “unexplained wage gap”, *i.e.* that proportion of the total wage gap, that is not accounted for after controlling for mean differences in characteristics such as age, qualification and

*Department of Statistics and Actuarial Science, University of Stellenbosch, Private Bag X1, Matieland, 7602, South Africa.

†Corresponding author: Department of Statistics and Actuarial Science, University of Stellenbosch, Private Bag X1, Matieland, 7602, South Africa, email: njlr@sun.ac.za

seniority. Blinder [10] used regression models to quantify the proportion of the white-black earnings differential for white men as well as the male-female earnings differential that is due to “objective characteristics like education and work experience” and the proportion that is unexplained by the “objective” characteristics. The latter proportion is believed to be suggestive of discrimination. Oaxaca [34] proposed several multiple regression models for estimating a discrimination coefficient as a measure of discrimination. These regression models contained a wide variety of explanatory variables for controlling male-female differences. Analyses of the residuals and the regression coefficients obtained via these models were then proposed for arriving at an estimate of discrimination.

Using the Oaxaca model, Ashraf [3] studied the influence of race and gender on salaries earned in academia in the USA over the period 1969–1989. He found the racial earnings gap to have narrowed considerably more by the end of that period than the gender earnings differential. Furthermore, the gender gap was much higher among professors than among associate and assistant professors. Ashraf’s results showed that although the “discriminatory” component of the gender earnings differential decreased between 1969 and 1984, it rose again thereafter. It is argued by Benjamin [9] that the increase in the gender earnings differential in academia should partly be attributed to an increase in the relative proportion of women entering the academic profession. However, he also pointed out that substantial differences in salary, rank and tenure remained between males and females in academia. Furthermore, males and females were responding differently to a general decline in academic opportunities in the late nineties. Toutkoushian [40] confirmed the findings of Benjamin by demonstrating the increased participation (both in absolute numbers and relative proportion) of women in academia, but he went on to say that “once women have entered academia, there is less evidence that men and women do the same things and are rewarded in the same ways conditional on their qualifications”.

Barbezat [4] too was able to show that the gender wage gap in academia had narrowed in the USA from 1968 to 1977. A similar conclusion was reached by Ransom and Megdal [35] who analyzed USA survey data, information from institutions and published research. The latter authors reported that the gender salary gap in the academic labour market had closed considerably between 1965 and 1985 but they also found the salaries of women still to fall short of those of the men. Toutkoushian [39] applied various versions of the Oaxaca [34] model, including those of Barbezat [5], to data originating from two surveys conducted by the National Center for Education Statistics (NCES, [31]). He found that the unexplained gender differences in salary still had not decreased from previous years and were essentially similar to the findings of Barbezat [5]. In particular, Toutkoushian [39] was able to show that these salary differences were significant across institution type and discipline. However, an important finding was that the unexplained gender remuneration gap in academia was radically different between the younger and older academics.

Similar differentials than those reported for the USA existed also in other countries: McNabb and Wass [29], for example, write about the “male-female salary differentials in British universities”; Ward [44] investigates “the gender salary gap in British academia”; while Warman *et al.* [45] report on the “evolution of male-female wages differentials in Canadian universities”.

Of the several studies concerning the role of gender in academic reward reported since

the 1970s, some focused on individual universities (*e.g.* Ferber and Green [14]), while others took a broader view from a national perspective (Ward [44]). Certain authors, like Barbezat [4], were able to show that the gender wage gap in academia had narrowed in the United States, possibly due to affirmative action policies. On the other hand, there is some evidence that females received higher salaries and raises than their human capital, academic discipline, rank and research productivity would warrant (*e.g.* Lindley *et al.* [26]). Although these studies show contradictory results, Johnson and Stafford [24] by employing a human capital model and using national data for PhDs in academic employment in the USA, concluded “that it is likely that slightly over half of the academic salary differential by sex (in 1970) would have existed even in the absence of direct labor market discrimination because of historical differences between the sexes in patterns of work attachment” (see also Johnson and Stafford [25]). In reaction to the Johnson and Stafford [24] paper, Farber [13] responded by conducting a longitudinal study, while Strober and Quester [38] expressed skepticism about the basic finding of Johnson and Stafford as a “substitute for the discrimination explanation”. In their reply to these two papers, Johnson and Stafford [25] pointed out that despite some reservations about the models used by Farber, the latter’s longitudinal results provided evidence supporting their conclusions as to the “effect of differential labor force participation”. Referring to the Strober and Quester criticism, they denied maintaining direct discrimination in the labour market to be unimportant. Bellas [8] found from the literature another form of discrimination that might be responsible for depressing women’s wages *viz.* “the devaluation of work that is performed primarily by women”. Therefore, she examined the issue of comparable worth across faculties in the USA by considering the following three explanations for variation in salaries earned in academia: cultural devaluation of work that women do, labour-market conditions within academia and characteristics of individual faculties. She concluded that all three had a role in explaining the variation in salaries earned by males and females in academia.

Ward [44] pointed out that her data set, covering five of Scotland’s eight old established universities, revealed the females to be younger than their male counterparts and to be outnumbered by them in the more senior positions like senior lecturers and professors. The reasons for the persistence of the wage gap were explored further by Barbezat and Hughes [6]. They found the persistence of the gender wage gap to be due to differences in the salary hierarchy in different institutions (research versus liberal arts institutions).

Neumark [33] differentiated between two types of discrimination *viz.* nepotism toward men or discrimination against women. Nepotism results in women receiving the competitive wages, but men being overpaid; discrimination against women occurs where men receive the competitive wages, while women are underpaid. Thus, Neumark proposed a general discrimination model in which employers are allowed to have different preferences. The Neumark model was utilised by Appleton *et al.* [2] in one of the few studies devoted to the gender wage gap in African countries. These authors documented a narrower wage gap in the three African countries considered (Ethiopia, Uganda and the Côte d’Ivoire) than it might otherwise be, due to women being overrepresented in the better-paid public sector. They concluded their paper with the interesting remark that the above conclusion “is evident in the simple descriptive statistics but would have been obscured had we used the conventional decomposition techniques”, with “conventional decompositional techniques”,

referring to the Blinder [10] and Oaxaca [34] variants. Weichselbaumer and Winter-Ebmer [46] performed a meta-analysis of the international wage gap. They provide a quantitative review of the vast amount of empirical literature on gender wage differentials between men and women of equal productivity, worldwide, from the 1960s to the 1990s. Of the 1535 estimates they used in their meta-analysis only three percent were from Africa, with only two studies from South Africa included. It is not clear in which occupational groups these two studies were conducted. Comparing the reported wage gap and the reported wage residual for the different countries, part of the total wage gap may be attributed to differences in human capital (Côte d'Ivoire, Tanzania, Korea, Kenya, Cyprus, Japan, Indonesia and Nicaragua). In contrast, considering their human capital, women are more discriminated against than suggested by the total wage gap in Singapore, Guinea, Costa Rica, Sudan, Trinidad and Tobago, Philippines and South Africa. Looking at trends for country groups Weichselbaumer and Winter-Ebmer report that the ratio of what women should have earned in the absence of discrimination and their actual earnings to have decreased by 0.3% annually in the USA, 0.8% in Canada, Australia and New Zealand, 0.2% (not statistically significant) in Europe and 1.9% in post-communist countries. There was no discernable trend for Africa and Latin America. For Asia, the trend was even positive, increasing by 0.4% per year.

Gender wage differentials in the “new” South Africa were investigated by Hinks [22] who reported the focus of the changing nature of discrimination in South Africa to be on race with little concern shown for gender discrimination. Referring to the October Household Survey of 1995, Hinks reported female managers, on average, to earn 43% of what their male counterparts earn, while constituting only 18% of the labour market. Female professionals, on average, earn 71% of what their male counterparts earn, and constitute 43% of the labour market. In the general South African labour market, white and Asian females suffer greater gender discrimination than their black and coloured counterparts, possibly due to low wages in the black and coloured population. The magnitude of gender differentials in labour market outcomes in South Africa has been documented in very few studies thus far. Grün [21] studied direct and indirect gender discrimination in the South African labour market using statistical methods based upon modifications of the original Blinder [10] and Oaxaca [34] procedures. The October Household Survey data 1995, 1997 and 1999 were used to compare wages earned by South African men and women. The unexplained component of wages differentials was shown to have reached alarming dimensions. Moreover, gender discrimination was even increasing between 1995 and 1997. This was the case for both black and white workers. The Grün study also reveals black women to be particularly discriminated against at the hiring stage, whereas for white women the extent of direct wage discrimination (*i.e.* the unexplained component) has increased from 1995 to 1997, despite a reduction in the overall wage gap for that period.

In South Africa, prior to 2004, four formulae were used as a basis for the funding of universities: the so-called Holloway formula introduced in 1953 (documented in the Report of the Holloway Commission [23]); the Van Wyk de Vries formula introduced in 1977 (documented in the Report of the Van Wyk de Vries Commission [42]); the South African Post Secondary Education Information System (SAPSE) subsidy formula implemented in 1984 and revised for implementation in 1993 (documented in the Venter Report [43]). A detailed summary with a discussion of these four formulae and the interrelationships

among them is provided by Steyn and Vermeulen [37]. The SAPSE-subsidy formula introduced in 1984 in South Africa does not differentiate in the remuneration of male and female academic staff members at South African higher education institutions. Indeed, this was also the case under the earlier Holloway and Van Wyk de Vries formulae of 1953 and 1977, respectively. All the above formulae used salary scales as input parameters in the formula allocations to the various higher education institutions so that responsibility largely rested with the institutions to establish levels of pay for male and female academic staff members. After 1994, South African higher education institutions were faced with the reality of transforming to an organisational culture appropriate for the “new” South Africa. In a review of the implications for academic staff of institutional transformation at South African universities, Fourie [15] reported that female staff members are still mostly found at the lower ranks. Referring to the National Commission on Higher Education (NCHE) report of 1996 (see [32]), Mabokela [28] stated that 68% of the total research and teaching staff at South African universities in 1993 were male compared with 32% female. Moreover, the majority of female academics were employed as junior lecturers or lecturers. At the University of Cape Town, between 1983 and 1995 the proportion of female faculty increased slightly from 17.96% to 23.90%. At Stellenbosch University, for the same time, female faculty increased from 15.60% to 25.09%, but still disproportionately occupied the lower academic ranks as lecturer, junior lecturer and below, relative to their male counterparts. For Stellenbosch University in particular, Mabokela [28] reported that females comprised 15.6% of the academic staff in 1983 and 33.49% in 1995. In 1995, Stellenbosch had 2.66% female professors, 7.5% female associate professors, 16.73% female senior lecturers, 45.92% female lecturers, 89.46% junior lecturers and 100% below junior lecturers. However, Stellenbosch University has developed an employment equity policy to address these differences. The employment equity policy of Stellenbosch University [41] states that the University is committed to equity in the working environment, to respect diversity, to ensure fair labour practices and to eliminate unfair discrimination.

In this paper, we focus on the gender remuneration gap for academic staff at Stellenbosch University for the years 2002 to 2005. Based on the literature, we investigate the possibility that the gender wage gap may be explained by a gender research output gap, a gender age gap or a gender academic rank gap. Furthermore, it needs to be taken into account that the University must provide competitive remuneration for staff with qualifications that are in demand in the broader labour market, as is the case with staff in the Faculties of Law, Economic and Management Sciences and Engineering, and often for National Research Foundation evaluated researchers. Since these factors are not unrelated it is necessary to perform a multivariate analysis of the relevant data, and since in general “there are many patterns and relationships that are easier to discern in graphical displays than by any other data analysis method” [12], this analysis should preferably be accompanied by suitable graphics. In this regard we also refer to Gray [20] who examined the courts’ treatment of the statistical issues involved in studying possible gender discrimination in academia in the USA in her paper “Can statistics tell us what we do not want to hear?” We agree with her declaration in a rejoinder to a discussion of her paper: “Part of the basis of the expert’s judgment as to what feels right must be how the model can be explained to the finders of fact — judge or jury. Certainly the use of graphs and charts provides major assistance.” Although multivariate statistical techniques like multiple regression are used in the above

studies, a shortcoming is that the graphics employed are in the form of univariate plots like bar charts and histograms or bivariate plots like scatter plots or x - y line graphs. Since several variables are studied, *e.g.* remuneration, age, rank, research output and qualification, graphs that are able to incorporate the multidimensional character of the data should be considered. In this paper, it is illustrated how biplots may be utilised to meet this goal.

Therefore, the primary aim of this paper is to show by means of a case study how biplots could be used for describing and monitoring changes over time in the gender remuneration gap. In the biplot, the multidimensional character of the data is taken into account so that patterns and relationships might become visible that would not have been possible if only univariate (histograms, bar charts) or even bivariate (scatter plot) graphics were employed.

In the next section, we give a description of the data set constituting our case study. This is followed by a brief introduction to biplot methodology. In the main section (§4), we present the case study data in the form of biplots and discuss how to interpret them. Finally, we summarise our conclusions.

2 Case study data set

Our case study data set comprises the permanent full-time academic (C1) staff at Stellenbosch University over the period 2002 to 2005. The following measurements were provided by the Human Resources Division and the Research Development Division at Stellenbosch University:

- a coded identification number (complete anonymity was maintained and the true identity of all C1 staff members remained unknown to the researchers)
- gender
- faculty
- age (in years)
- the total annual cost to company, before deductions, for December 2002, 2003, 2004 and 2005 (Remun; in units of R10 000)
- research output (Resrch; measured on a continuous scale described below)
- academic position or rank (Rank; described below)
- academic qualification (Aqual; measured on a continuous scale described below).

Two data sets were prepared: one for 2002 and one for 2005.

Three remuneration adjustments that occurred in January 2003, January 2004 and January 2005 were included in the figures representing the total cost to company. Inflation was not taken into account, because it affected all C1 staff equally during the study period.

The research output for 2005 reflects the average research output of C1 staff members from 2002 to 2005, while the research output for 2002 contains only 2002 output. An average for 2005 was calculated by dividing the total research output for each C1 staff member by the number of years employed by the University from 2002 to 2005. This must be taken into account when the remuneration data sets for 2002 and 2005 are compared.

Research output is defined by the Department of Education as textual output resulting from an original, systematic investigation undertaken in order to gain new knowledge and understanding. Peer evaluation of the research output is a fundamental prerequisite for recognition. For the purpose of state subsidy, recognised research output in terms of this policy comprises journals, books and proceedings. Research output units are allocated according to the specifications in the Higher Education Act 101, 1997 [30] as follows:

- *Journals*: A research article published in an approved journal will be subsidised as a single unit. If two authors have contributed to the article, each will receive 0.5 units.
- *Books*: A book may be subsidised to a maximum of five units or portion thereof, based on the number of pages being claimed relative to the total number of pages in the book.
- *Proceedings*: Proceedings published as part of a peer-reviewed non-periodical research output from conferences, symposia or other meetings will be allocated a maximum of half a unit (0.5).

The academic positions or ranks were quantified according to the following scheme: Junior Lecturer: 1; Lecturer: 2; Senior Lecturer: 3; Associate Professor: 4 and Professor: 5.

Finally, academic qualifications were quantified as follows: Undergraduate Diploma or Certificate: 1; General academic first Bachelor's Degree: 2; Professional first Bachelor's Degree: 3; Post-graduate Diploma or Certificate: 4; Post-graduate Bachelor's Degree: 5; Honours Degree: 6; Master's Degree: 7; Second Master's Degree: 7.5; Doctoral Degree: 8; Doctoral Degree with second Master's Degree: 8.5; Second and third Doctoral Degrees: 9.

The number of permanent full-time C1 staff members in the various faculties according to gender for 2002, 2003, 2004 and 2005 is provided in Table 1. Note that the Faculty of Theology (and to a lesser extent the Faculty of Law) had very few permanent full-time C1 staff members in 2002, 2003, 2004 and 2005. In addition, the Faculty of Engineering employed very few female C1 staff during the period of study. Therefore, individual biplots for these faculties are not shown in this paper.

3 Biplot as a multidimensional extension of a scatter plot

Biplots were introduced by Gabriel [16] as a graphical display of the rows and columns of a data matrix \mathbf{X} : $n \times p$ in a single graph consisting of $n+p$ points/vectors: n row markers and p column markers. Any value x_{ij} is estimated by the scalar product of row marker i and column marker j . Gabriel [17] also treated biplots relevant to canonical analysis. Gabriel's work received a significant development in subsequent work, mainly by Gower, as described in Gower and Hand [19]. In this original approach, the biplot is seen as a multivariate analogue of the well-known scatter plot for displaying two variables graphically. In a scatter plot, the two variables are represented as the two perpendicular axes while the sample observations are displayed as points on the diagram. Gower and Hand showed how to generalise this simple graph to provide for samples having measurements for more than two variables (see the Appendix for more details). In our case study, we have five variables: remuneration, age, rank, qualifications and research output. These five variables

Faculty	2002		2003		2004		2005	
	M	F	M	F	M	F	M	F
ArtsSocSc	91	59	91	58	89 (75)	64 (50)	83 (68)	63 (44)
Science	103	40	103	40	100 (87)	43 (34)	90 (73)	38 (26)
Education	20	21	20	21	18 (16)	21 (16)	17 (15)	19 (13)
Agriscience	46	13	46	13	46 (38)	18 (12)	42 (32)	21 (11)
Law	22	7	22	7	22 (19)	9 (7)	18 (15)	10 (6)
Theology	11	1	10	1	11 (6)	1 (1)	13 (6)	—
EconManSc	81	39	81	39	85 (68)	50 (32)	81 (61)	46 (26)
Engineering	60	3	60	3	62 (51)	5 (2)	59 (44)	3 (1)
HealthSc	49	62	50	53	49 (42)	53 (46)	53 (42)	64 (42)
Total	483	245	483	235	482(402)	264(200)	456(356)	264(169)

Table 1: The number of full-time C1 staff members in the different Faculties according to Gender for 2002, 2003, 2004 and 2005. Individuals comprising the 2003 data set are almost identical to the individuals in the 2002 baseline data set. The numbers in brackets for the 2004 and 2005 data sets show the number of individuals who were also in the baseline data set of 2002.

are then represented as five biplot axes — each labelled with the corresponding variable’s name. Since we cannot have five perpendicular axes in just two dimensions, these axes will be non-perpendicular, but they can be calibrated in the original scales of measurement. Thus, biplot axes are used similarly to conventional scatter plot axes: a line is dropped perpendicular from any point in the display to any labelled axis and the value for that variable is determined from the calibrations appearing on the axis. By considering a biplot as a multivariate extension of an ordinary scatter plot, non-statisticians should comfortably understand the basics of biplot methodology, enabling them to interpret the biplot display with relative ease.

Since we would like to display the differences between the two gender groups as well as the changes that took place between 2002 and 2005, we would like the biplot to show optimally the five-dimensional group means of the four groups: 2002 females, 2002 males, 2005 females and 2005 males. In order to achieve this primary objective, canonical variate analysis (CVA) biplots will be constructed. The CVA follows naturally from a one-way multivariate analysis of variance (MANOVA) in which one attempts to reduce the MANOVA to a one-way univariate analysis of variance (ANOVA) by replacing the matrix \mathbf{X} by the n -vector $\mathbf{X}\mathbf{v}$, where \mathbf{v} : $p \times 1$ is the vector of coefficients of a linear combination applied to all the rows of \mathbf{X} (see the Appendix). Solving the problem of optimally choosing the linear combination \mathbf{v} , so as to maximise the value of the F -statistic, will yield a matrix \mathbf{V} : $p \times p$ defining the transformation $\mathbf{Y} = \mathbf{X}\mathbf{V}$ from the original data to the so-called canonical variables matrix \mathbf{Y} : $n \times p$. The canonical means, *i.e.* the group means of \mathbf{Y} possess a very important characteristic: only their first k elements differ, where k is defined as $\min\{\text{number of groups} - 1; \text{number of variables}\}$. Since in our case study, we are dealing with five variables and four groups, our five-dimensional means are effectively transformed to canonical means in which only the first three elements differ. Thus, we expect only a small loss of information when the three-dimensional data are represented in two dimensions. In passing, we remind the reader that the term *biplot* refers to the simultaneous display of both the rows and columns of a data matrix and not to the dimension of the

display space.

Recent literature on wage differentials in general, emphasises the need to evaluate the wage gap at different points of the wage distribution rather than at a mean or median value alone. Thus, for example, Mata and Machado [27] use quantile regression for decomposing changes in wage distributions. CVA biplots can also incorporate more than a mere display of means and axes for the variables: when a CVA biplot of the canonical means has been constructed, all the original samples, with different plotting symbols to distinguish between the groups, can be interpolated onto the biplot. The CVA biplot may then be overlaid with contours demarcating the inner $100\alpha\%$ of the observations. These contours are called α -bags and they are based upon the idea of a depth median and depth contours as introduced by Rousseeuw *et al.* [36]. Their construction is explained in Gardner [18] as well as in Aldrich *et al.* [1]. By equipping a CVA biplot with separate α -bags for the respective groups, detailed descriptions of the multidimensional separation and overlap among the different groups of observations are obtained. Since the proportion of bivariate points enclosed by α -bags can be specified, these bags provide not only a visual measure of the overlap between groups, but also a quantitative measure of such overlap: if two groups overlap for a small value of α they are more similar than two groups that overlap only for a larger value of α . Therefore, multidimensional change over time can be monitored by using CVA biplots together with α -bags.

4 CVA biplots of case study data

We constructed biplots that satisfy all requirements to be geometrically accurate: all distances (and directions) in these biplots can be measured with an ordinary ruler for inspection purposes. The software (see [11]) we used for constructing the biplots given in this paper, enables functionality such as zooming in or out and interactively turning axes on or off. In practice, differences between males and females can thus easily be determined with the aid of the calibrated axes and the zoom function.

In our case study the data set is the population under consideration and not just a sample from a population. Therefore describing the population is of primary importance and we are not concerned with testing of hypotheses regarding population parameters since these can be calculated directly. The practitioner should decide what magnitude of difference is considered to be practically significant.

As described in the previous section, we use α -bags superimposed on the biplots in order to describe the multidimensional character of the population under consideration. The α -bags are not confidence regions, but are descriptive measures which are not based upon any distributional assumptions. They should be seen as bivariate extensions of the univariate boxplot where the “box” is allowed not to contain only the central 50% of the values but any specified proportion of the data nearest to the (depth) median. The α -bags provide information concerning location, skewness, variability, outlying points, and the nature of overlap and separation among the four groups. Even if the means are similar, the shapes of the α -bags may be quite different pointing to, for example, differences in variation or the role of outlying values.

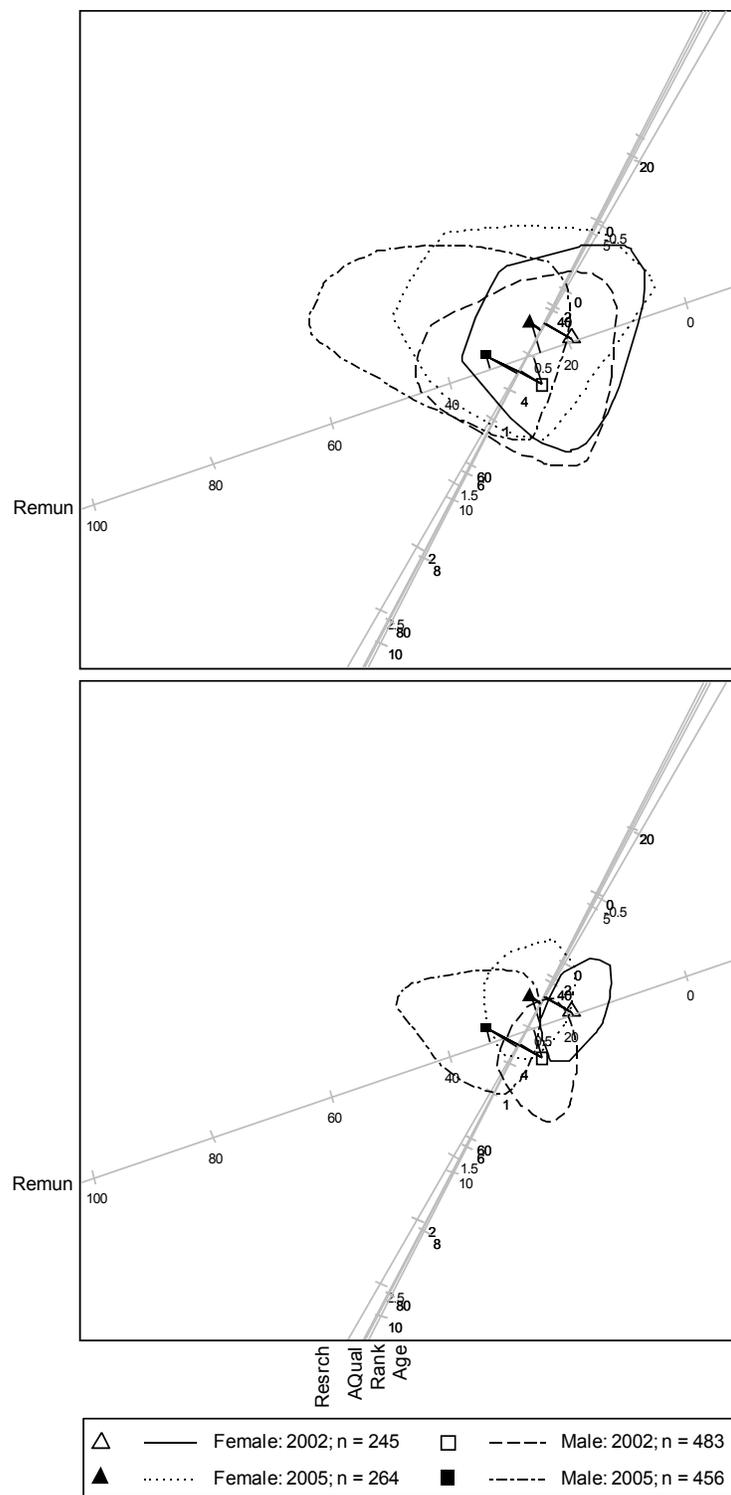


Figure 1: CVA biplot separating the data set of all C1 staff according to the 2002 Female, 2005 Female, 2002 Male and 2005 Male subgroups. Plotting of data points is suppressed. The upper panel shows the biplot overlaid with 90% bags; the lower panel with 50% bags.

In Figure 1 a CVA biplot of the complete data set is given. This biplot optimally separates the five-dimensional group means for the four groups: 2002 males, 2005 males, 2002 females and 2005 females. Although all sample points (1 448 altogether) were interpolated onto the biplot, their actual plotting has been suppressed to avoid a graph so cluttered with plotting symbols that it becomes useless. Instead, we equipped the biplot with 90% bags (the top panel) and 50% bags (the bottom panel) to show the overlap or separation between the different groups of sample points. In this paper, we follow the convention of labelling our biplot axes at the endpoints where the respective calibrations take on their largest values.

The calibrations on the biplot axis representing remuneration, increase with the calibrations on the axes representing research output, academic qualifications, rank and age, which are bundled together. The following can readily be seen in the biplot:

- the means of the females have not changed much between 2002 and 2005 with respect to research output, academic qualifications, rank and age;
- likewise, the means of the males have not changed much between 2002 and 2005 with respect to research output, academic qualifications, rank and age;
- but, for these variables the means of the females are smaller than those of the males;
- from the remuneration axis, the increase in remuneration within gender groups over time is obvious, as well as gender differentials;
- the gender remuneration differentials are closely related to the gender differentials with respect to research output, academic qualifications, rank and age;
- despite considerable overlap between the gender groups in 2002, as well as in 2005, the entire male configuration in both instances is shifted towards the higher end of the remuneration axis when compared with the corresponding configuration for the females.

Figure 2 contains the same biplot as in Figure 1, but on a larger scale to illustrate how the biplot axes may be used to read off values for all variables. In this case study, the values determined from the biplot axes match very closely the actual calculated values (see Table 2).

We highlight the following in Figure 2:

- the size of the gender remuneration gap is discernible from the remuneration axis — the mean of the males for 2002 projects to a higher value than the mean of the females even in 2005;
- the difference in lengths between the long grey and the long black arrows accurately shows a nominal increase in the gender remuneration gap (with respect to the mean) from 2002 to 2005, although a simple calculation reveals the percentage increase in the means to be 27.8% for the females and only 26.7% for the males;
- the difference in the lengths of the short black and the short grey arrows accurately shows that the females have increased their (mean) research output by more than the males between 2002 and 2005;
- although the mean research output of the males in 2005 is still higher (approximately twice) than that of the females, the latter have succeeded in reducing this gap. Moreover, while there is little change in the projections of the mean for the males onto the research, rank, age and qualification axes between 2002 and 2005, the

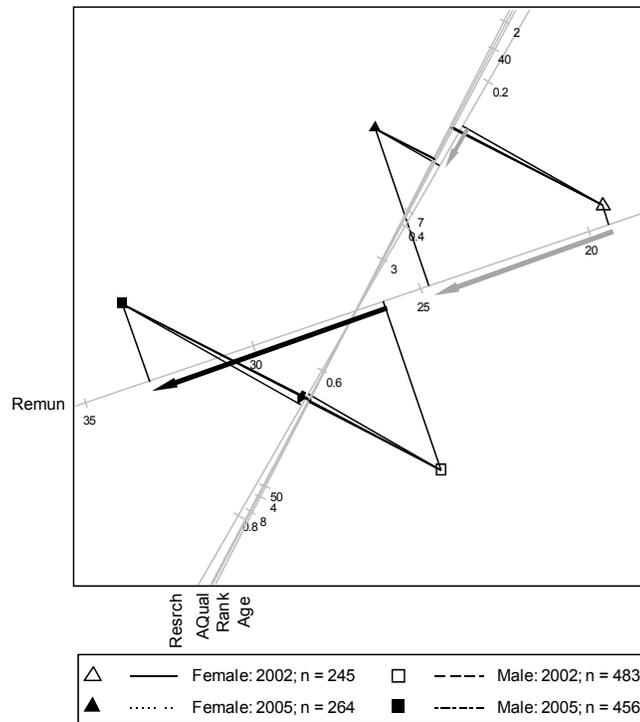


Figure 2: Large-scale CVA biplot showing the means of the 2002 Female, 2005 Female, 2002 Male and 2005 Male subgroups of the complete C1 data set. Also shown are perpendicular lines from each group mean to the five axes showing the biplot predictions of the different variables. The long grey and the long black arrows indicate the changes from 2002 to 2005 in the remuneration for the female and male staff members respectively. Similarly, the short grey and the short black arrows indicate the corresponding changes in the research output.

corresponding changes in the case of the females are not only larger, but constitute shifts toward the mean values of their male counterparts.

Is the state of affairs for all C1 staff suggested by Figures 1 and 2 similar to that of the individual faculties referred to in Table 1? In Figures 3 to 9 we give separate CVA biplots for seven of the individual faculties — the Faculties of Theology and Engineering are excluded due to very limited numbers of female C1 staff (see Table 1).

A comparison of Figure 3 with Figures 1 and 2 reveals very little difference between C1 staff in the university as a whole and those only in the Faculty of Arts and Social Sciences: the remuneration gender gap is similar, as well as the association between remuneration on the one hand and age, research, rank and academic qualification on the other hand.

While the biplot in Figure 4 for the Faculty of Economic and Management Sciences bears a resemblance to those in Figures 1, 2 and 3, it is characterised by a wide gender remuneration gap associated with wide gender age and research gaps. These gaps remain almost constant when 2005 is compared to 2002. Moreover, it seems that all gender differentials of 2002 are still very much the same in 2005 although the mean research output of the females shows some improvement from a very low value in 2002.

	Predicted mean values				Actual mean values			
	F2002	F2005	M2002	M2005	F2002	F2005	M2002	M2005
Remun	19.36	24.75	26.13	33.11	19.31	24.80	26.14	33.09
Resrch	0.26	0.32	0.63	0.65	0.26	0.32	0.63	0.65
Rank	2.45	2.58	3.60	3.59	2.44	2.59	3.60	3.59
Age	41.79	42.53	48.05	48.03	41.98	42.34	48.00	48.11
Qualif	6.72	6.83	7.63	7.63	6.72	6.83	7.63	7.63

Table 2: Predictions obtained from CVA biplot versus true means.

The CVA biplot for the Faculty of Science shown in Figure 5 demonstrates in general similar tendencies to that of the university as a whole. There is a widening in the gender remuneration gap between 2002 and 2005; the gender research differential has remained the same, but the gender age gap has become less pronounced due to an increase in the female mean age together with a decrease in that of the males.

The biplot in Figure 6 shows a widening of the gender remuneration gap from 2002 to 2005 in the Faculty of Agrisciences. According to Figure 6, there is also a gender age gap in this faculty with the mean age of the males approximately 10 years higher than that of the female C1 staff. It is to be noted that the mean research output of the males was already in 2002 more than twice that of the females and this gap has become even wider in 2005.

Apart from the axis representing research output, the CVA biplot in Figure 7 for the Faculty of Law shows the same general tendencies which are visible in Figures 1 to 6. From the axis representing research output, the following can be seen: both the male and female staff groups have relatively high research output, but there is a decline in this output from 2002 to 2005. Moreover, this decline is more pronounced in the case of the males — to such an extent that in 2005 the gender research gap has been reversed.

The CVA biplot for the Health Sciences in Figure 8 appears somewhat different from those considered previously. There seems to be pronounced gender differentials for all the variables considered. In addition, the group configurations have remained approximately constant from 2002 to 2005.

The 90% bags in the upper panel of Figure 9 differ from those in the previous figures: firstly, there is very little overlap between the bags for 2002 and those for 2005; secondly, the overlap in the 90% bags for the males and the females is more pronounced in 2005 than in 2002. It can thus be concluded that in the Faculty of Education, not only has the gender remuneration gap become smaller in 2005 (see also the Remuneration axis in the bottom panel of Figure 9), but the characteristics of the female C1 staff in general have become more similar to those of their male counterparts. In addition, the following is clear from the bottom panel of Figure 9: mean ages of females and males are approximately 50 or more; mean research output of the males at 1.65 in 2002 is relatively high and declines to approximately 1.4 in 2005, while the corresponding value for females is about 0.2 in 2002 and increases to about 0.25 in 2005.

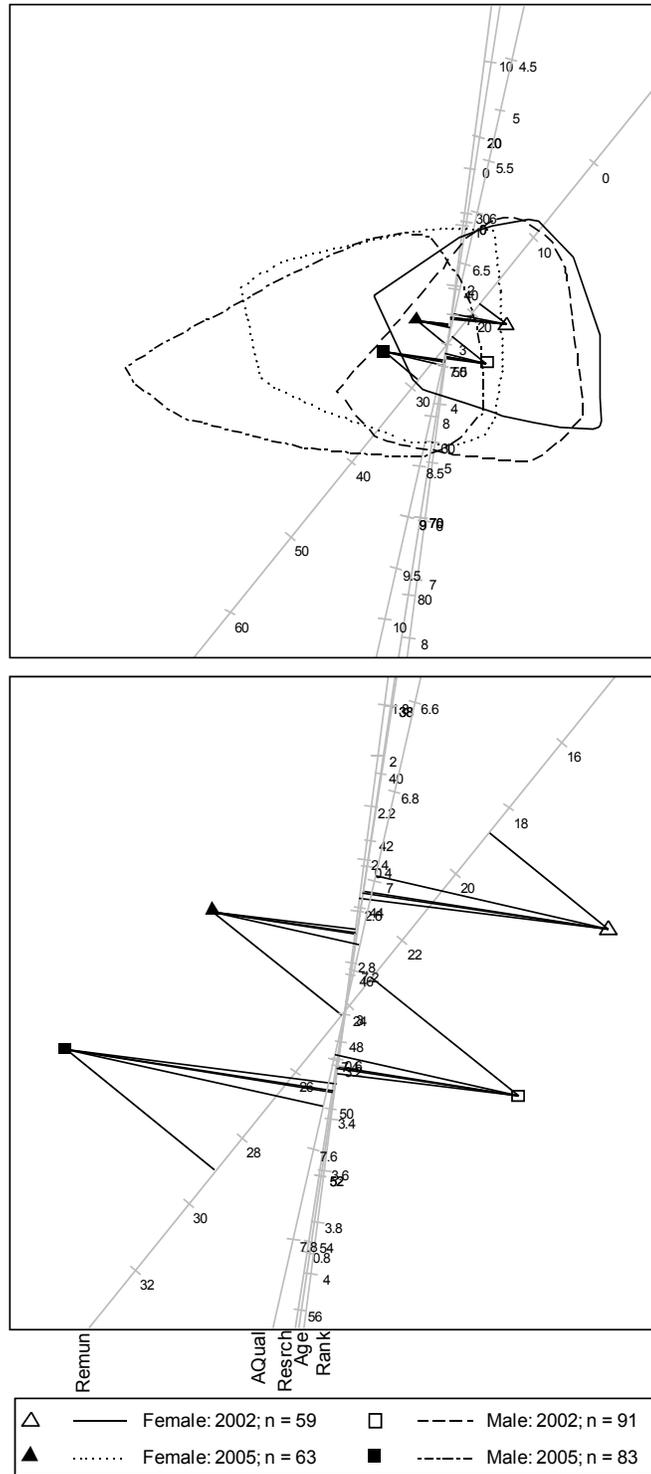


Figure 3: CVA biplot separating C1 staff in the Faculty of Arts according to the 2002 Female, 2005 Female, 2002 Male and 2005 Male subgroups. Plotting of data points is suppressed. The upper panel shows the biplot overlaid with 90% bags. The bottom panel shows the group means on a large scale together with perpendicular lines from each group mean to the five axes for determining the biplot predictions of the different variables.

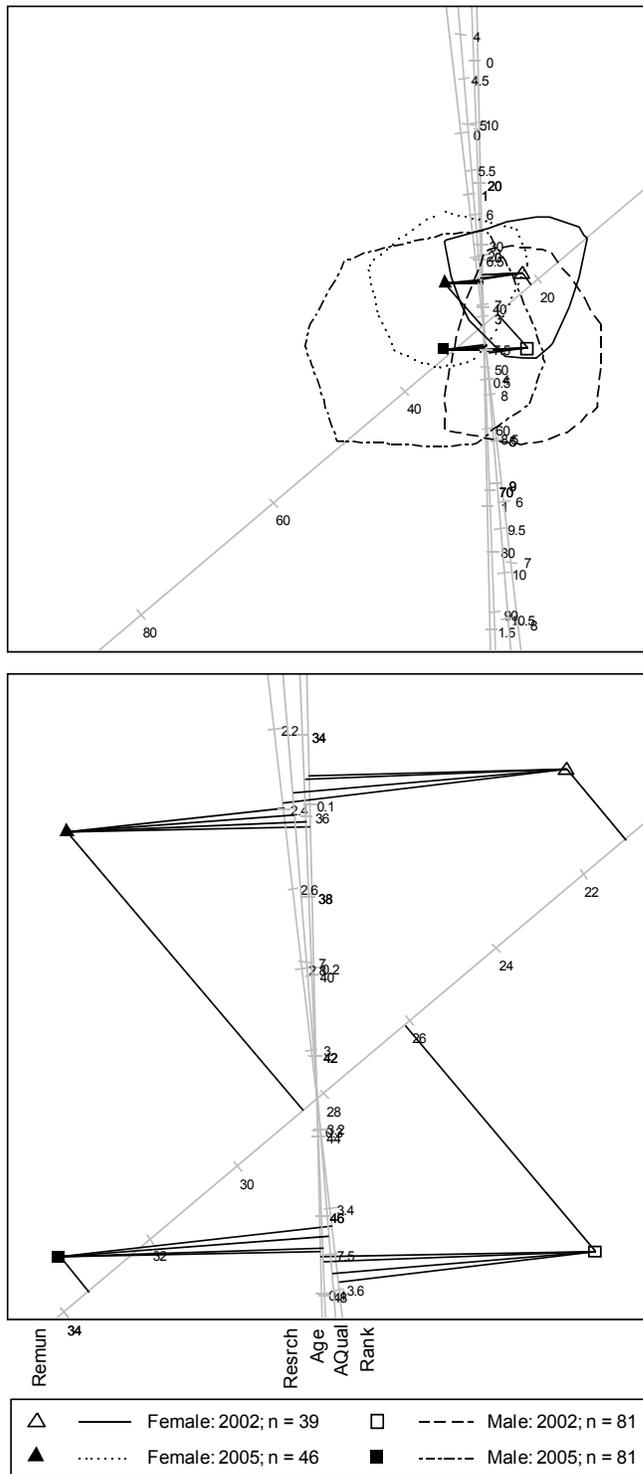


Figure 4: CVA biplot separating C1 staff in the Faculty of Economic and Management Science according to the 2002 Female, 2005 Female, 2002 Male and 2005 Male subgroups. Plotting of data points is suppressed. The upper panel shows the biplot overlaid with 90% bags. The bottom panel shows the group means on a large scale together with perpendicular lines from each group mean to the five axes for determining the biplot predictions of the different variables.

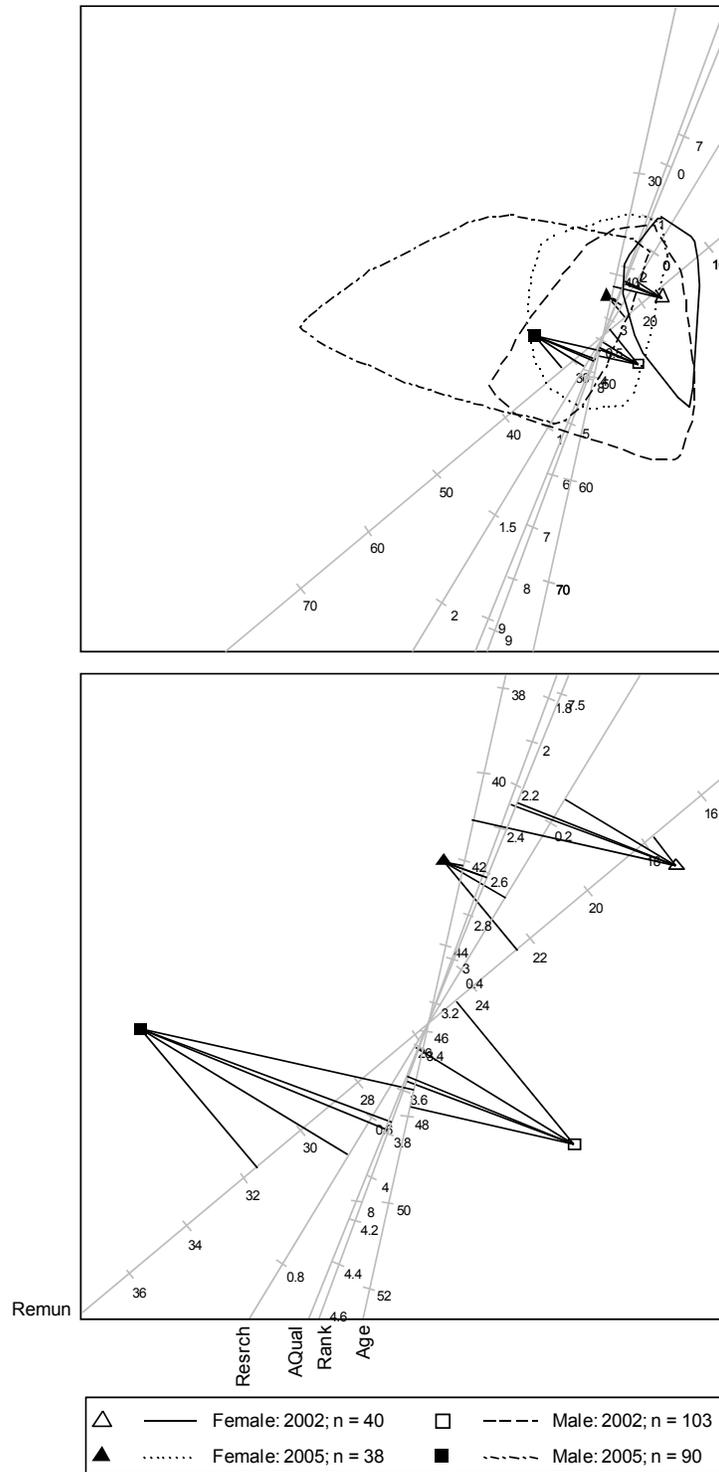


Figure 5: CVA biplot separating C1 staff in the Faculty of Science according to the 2002 Female, 2005 Female, 2002 Male and 2005 Male subgroups. Plotting of data points is suppressed. The upper panel shows the biplot overlaid with 90% bags. The bottom panel shows the group means on a large scale together with perpendicular lines from each group mean to the five axes for determining the biplot predictions of the different variables.

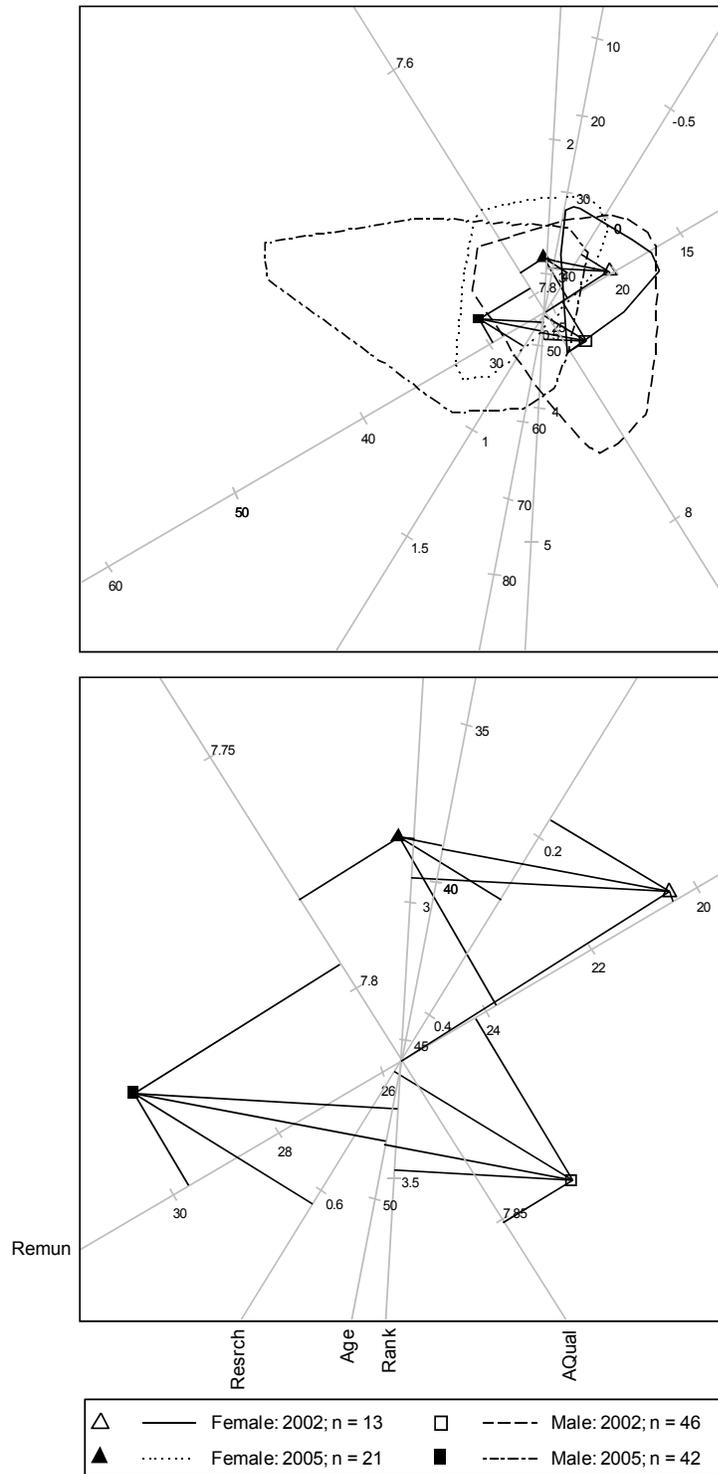


Figure 6: CVA biplot separating C1 staff in the Faculty of Agrisciences according to the 2002 Female, 2005 Female, 2002 Male and 2005 Male subgroups. Plotting of data points is suppressed. The upper panel shows the biplot overlaid with 90% bags. The bottom panel shows the group means on a large scale together with perpendicular lines from each group mean to the five axes for determining the biplot predictions of the different variables.

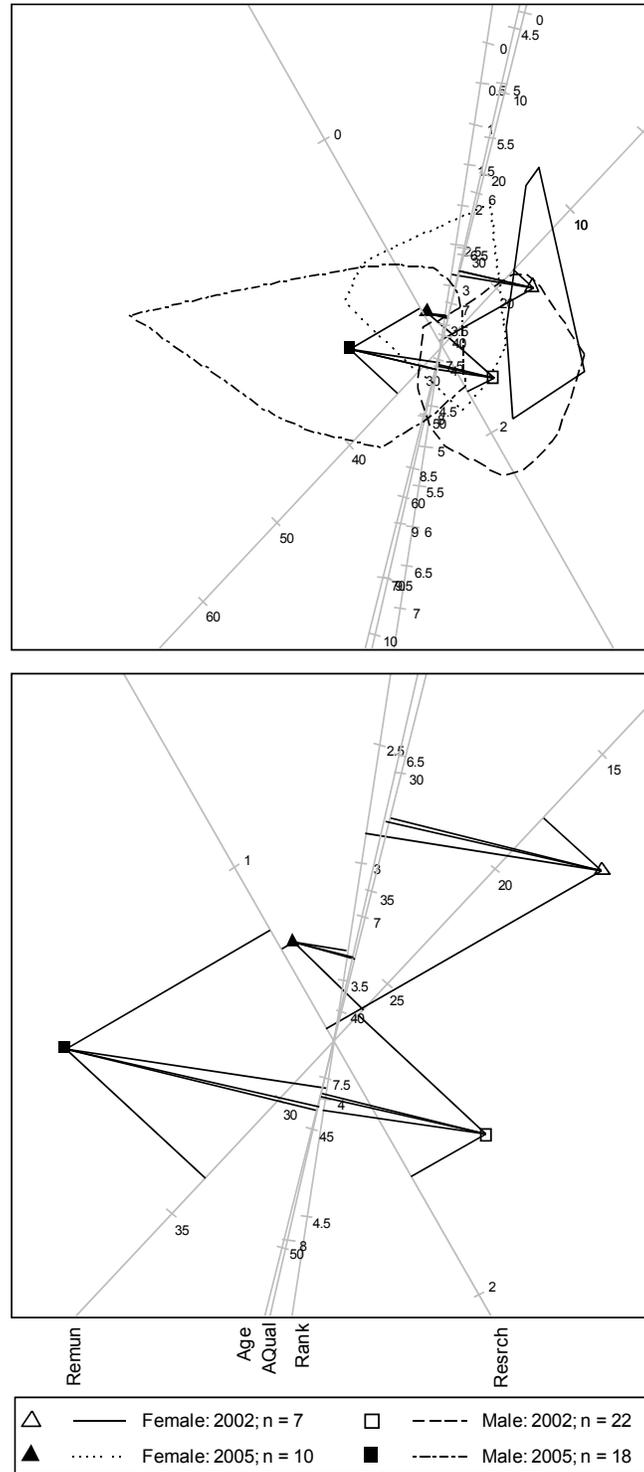


Figure 7: CVA biplot separating C1 staff in the Faculty of Law according to the 2002 Female, 2005 Female, 2002 Male and 2005 Male subgroups. Plotting of data points is suppressed. The upper panel shows the biplot overlaid with 90% bags. The bottom panel shows the group means on a large scale together with perpendicular lines from each group mean to the five axes for determining the biplot predictions of the different variables.

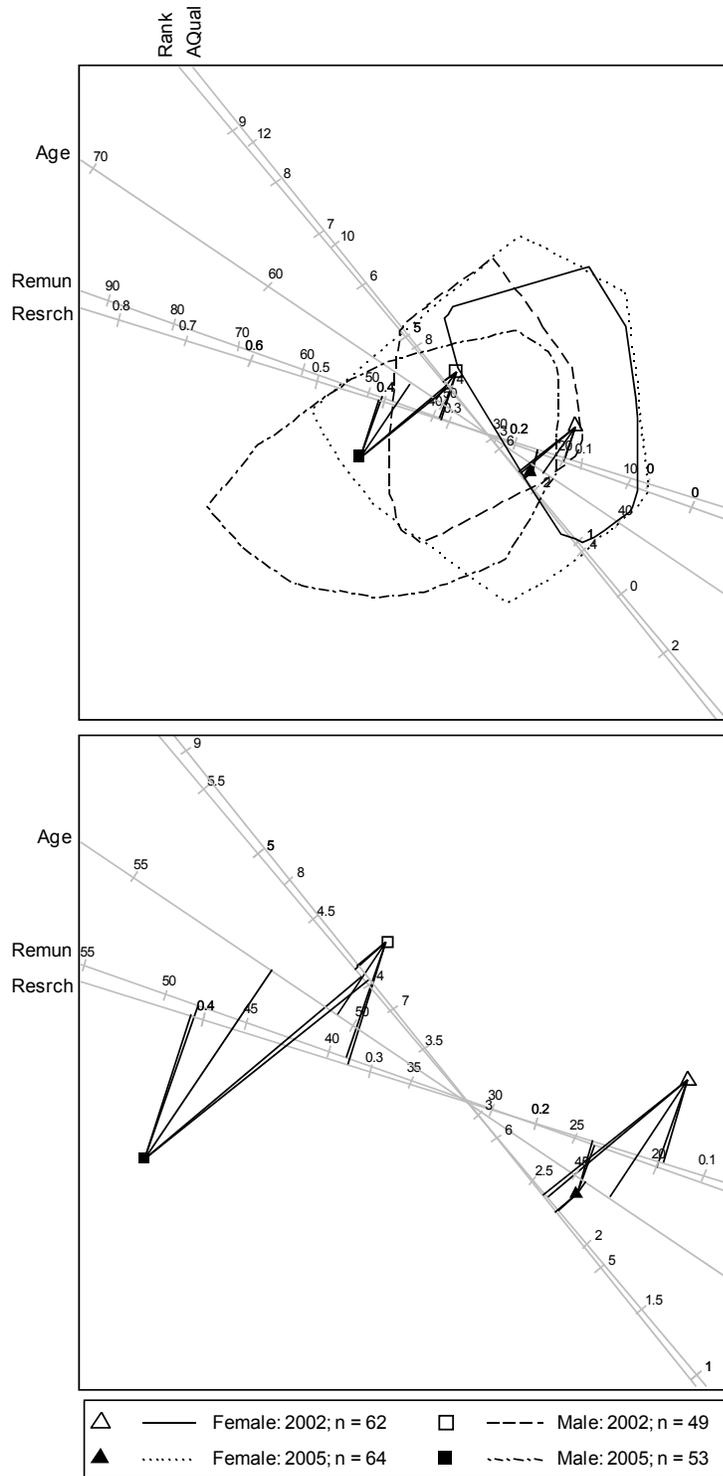


Figure 8: CVA biplot separating C1 staff in the Faculty of Health Sciences according to the 2002 Female, 2005 Female, 2002 Male and 2005 Male subgroups. Plotting of data points is suppressed. The upper panel shows the biplot overlaid with 90% bags. The bottom panel shows the group means on a large scale together with perpendicular lines from each group mean to the five axes for determining the biplot predictions of the different variables.

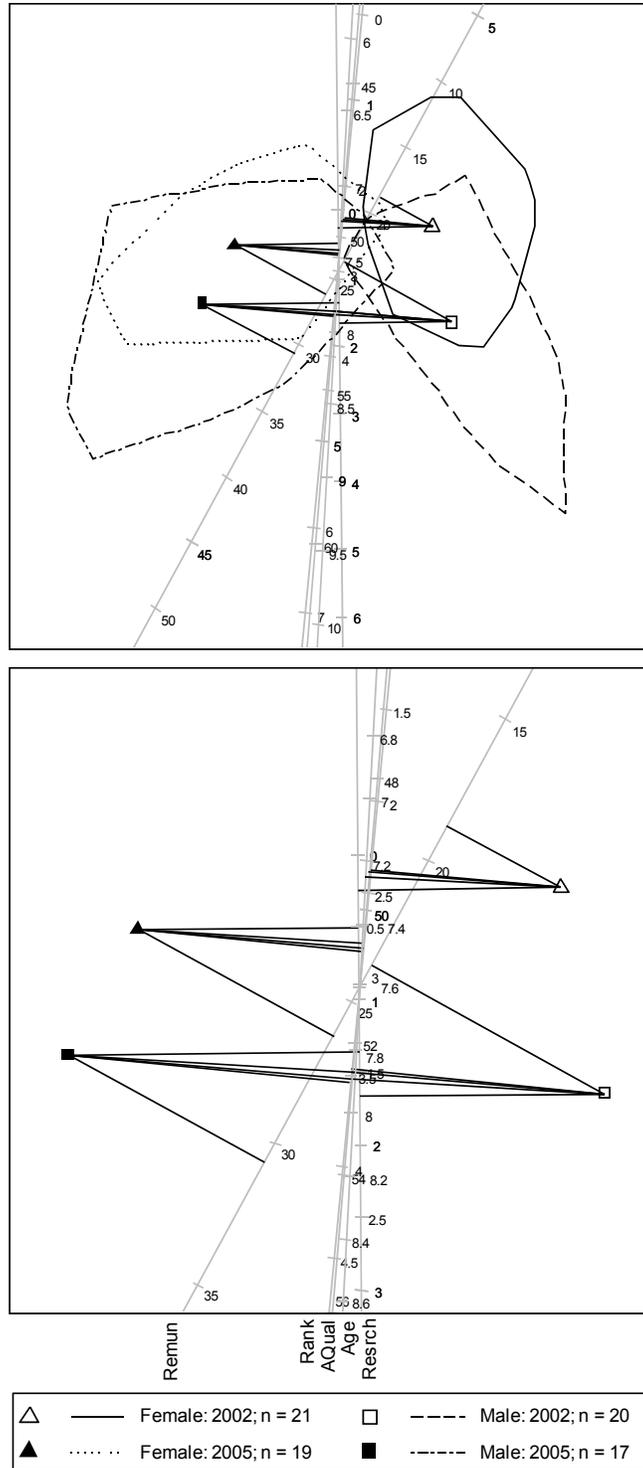


Figure 9: CVA biplot separating C1 staff in the Faculty of Education according to the 2002 Female, 2005 Female, 2002 Male and 2005 Male subgroups. Plotting of data points is suppressed. The upper panel shows the biplot overlaid with 90% bags. The bottom panel shows the group means on a large scale together with perpendicular lines from each group mean to the five axes for determining the biplot predictions of the different variables.

5 Conclusion

From the literature investigating the gender remuneration differential in academia in several countries, it may be concluded that this gender gap is closely associated with corresponding differentials in other areas of academia like research output, age, position held and academic qualifications. It follows that the gender remuneration differential should not be addressed without a serious consideration of all the associated gender differentials simultaneously. Instruments are therefore required to enable the relevant authorities to deal with these related issues in a multidimensional way. In this paper, we have illustrated how CVA biplots may be used in this respect. By presenting their data graphically in the form of a biplot, university management may obtain insight into the multidimensional nature of the gender differentials present in their institution. Such insight is a prerequisite for arriving at a lasting and fair solution to the problem. Moreover, a biplot equipped with α -bags provides a means for monitoring group separation or overlap over time. Therefore, the effect of steps taken to address the problem of gender differentials may be ascertained and monitored graphically. Furthermore, as the main features of a biplot are easy to understand, it provides management with an instrument to communicate policies and the results thereof to staff representatives. As our case study has shown, biplots may be constructed both for an institution as a whole and for its individual faculties. The unique circumstances of different faculties within the same institution may therefore be scrutinised.

Our case study demonstrates several aspects of gender differentials, and in particular, the gender remuneration gap, is highlighted by biplot representations. As with the other studies cited we have shown that one particular differential cannot be viewed in isolation. From our biplot displays, it can readily be ascertained that rank, age and research output are strongly related to remuneration of C1 staff. Therefore, it may be argued that in our case study, differentials in the remuneration of male and female C1 staff members are mainly due to relatively fewer female members within the senior ranks, in older categories, in higher research output categories and in higher qualifications categories. In some faculties (Law and Education) the canonical means in the CVA biplots provide evidence of a (slight) narrowing of the gap between male and female C1 staff members with respect to all the variables in the remuneration data set from 2002 to 2005. In other faculties (Agrisciences, Science and Health Sciences) there is a slight widening of the gender remuneration gap. The Arts and Social Sciences faculty and also Economic and Management Sciences show a constant remuneration gap between males and females that is associated with differences in age, research output, qualifications and rank. Moreover, the biplots suggest these gaps to remain almost constant over the period 2002–2005.

Finally, we remark that once a biplot is constructed it is easy to display interactively the axes one at a time or to vary the size of the α -bags. In doing so, detailed information of values of a particular variable, as well as of group separation and overlap, becomes available for management to aid and monitor their decisions.

Acknowledgements

We thank the referees for their invaluable contributions that enabled us to clarify and

improve the paper. Any remaining obscurities are ours alone. We would also like to thank the Human Resources Division and the Research Development Division at Stellenbosch University for providing us with the data used in our case study.

References

- [1] ALDRICH C, GARDNER S & LE ROUX NJ, 2004, *Monitoring of metallurgical process plants by using biplots*, American Institute for Chemical Engineering Journal, **50(9)**, pp. 2167–2186.
- [2] APPLETON S, HODDINOTT J & KRISHNAN P, 1999, *The gender wage gap in three African countries*, Economic Development and Cultural Change, **47**, pp. 289–312.
- [3] ASHRAF J, 1996, *The influence of gender on faculty salaries in the United States, 1969–89*, Applied Economics, **28**, pp. 857–864.
- [4] BARBEZAT DA, 1987, *Salary differentials by sex in the academic labor market*, The Journal of Human Resources, **22(3)**, pp. 422–428.
- [5] BARBEZAT DA, 1991, *Updating estimates of male-female salary differentials in the academic labor market*, Economic Letters, **36**, pp. 191–195.
- [6] BARBEZAT DA & HUGHES JW, 2005, *Salary structure effects and the gender pay gap in academia*, Research in Higher Education, **46**, pp. 621–640.
- [7] BAYER AE & ASTIN HE, 1975, *Sex differentials in the academic reward system*, Science, **188**, pp. 796–802.
- [8] BELLAS ML, 1994, *Comparable worth in academia: The effects on faculty salaries of the sex composition and labor-market conditions of academic disciplines*, American Sociological Review, **59**, pp. 807–821.
- [9] BENJAMIN E, 1999, *Disparities in the salaries and appointments of academic women and men*, Academe, **85**, pp. 60–62.
- [10] BLINDER AS, 1973, *Wage discrimination: reduced form and structural estimates*, The Journal of Human Resources, **8**, pp. 436–455.
- [11] DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE, 2008, *R library for constructing biplots*, 2008, Unpublished R code, Stellenbosch University, Stellenbosch.
- [12] EVERITT BS, 1994, *Exploring multivariate data graphically: A brief review with examples*, Journal of Applied Statistics, **21**, pp. 63–93.
- [13] FARBER S, 1977, *The earnings and promotion of women faculty: Comment*, The American Economic Review, **67(2)**, pp. 199–206.
- [14] FERBER MA & GREEN CA, 1982, *Traditional or reverse sex discrimination? A case study of a large public university*, Industrial and Labour Relations Review, **35(4)**, pp. 550–564.
- [15] FOURIE M, 1999, *Institutional transformation at South African universities: Implications for academic staff*, Higher Education, **38**, pp. 275–290.
- [16] GABRIEL KR, 1971, *The biplot graphical display of matrices with application to principal component analysis*, Biometrika, **58**, pp. 453–467.
- [17] GABRIEL KR, 1972, *Analysis of meteorological data by means of canonical decomposition and biplots*, Journal of Applied Meteorology, **11**, pp. 1071–1077.
- [18] GARDNER S, 2001, *Extensions of biplot methodology to discriminant analysis with applications of non-parametric principal components*, PhD Dissertation, University of Stellenbosch, Stellenbosch.
- [19] GOWER JC & HAND DJ, 1996, *Biplots*, Chapman & Hall, London.
- [20] GRAY MW, 1993, *Can statistics tell us what we do not want to hear? The case of complex salary structures*, Statistical Science, **8**, pp. 144–179.
- [21] GRÜN C, 2004, *Direct and indirect gender discrimination in the South African labour market*, International Journal of Manpower, **25**, pp. 321–342.
- [22] HINKS T, 2002, *Gender wage differentials and discrimination in the new South Africa*, Applied Economics, **34**, pp. 2043–2052.
- [23] HOLLOWAY COMMISSION OF ENQUIRY INTO UNIVERSITY FINANCES AND SALARIES, 1951, [Government Report], Government Publications, Pretoria.

- [24] JOHNSON GE & STAFFORD FP, 1974, *The earnings and promotion of women faculty*, The American Economic Review, **64**(6), pp. 888–903.
- [25] JOHNSON GE & STAFFORD FP, 1977, *Earnings and promotion of women faculty: Reply*, The American Economic Review, **67**(2), pp. 214–217.
- [26] LINDLEY JT, FISH M & JACKSON J, 1992, *Gender differences in salaries: An application to academe*, Southern Economic Journal, **59**, pp. 241–259.
- [27] MATA J & MACHADO JAF, 2005, *Counterfactual decomposition of changes in wage distributions using quantile regression*, Journal of Applied Econometrics, **20**(4), pp. 445–465.
- [28] MABOKELA RO, 2000, *'We Cannot Find Qualified Blacks': Faculty diversification programmes at South African universities*, Comparative Education, **36**, pp. 95–112.
- [29] MCNABB R & WASS V, 1997, *Male-female salary differentials in British universities*, Oxford Economic Papers, New Series, **49**(3), pp. 328–343.
- [30] MINISTRY OF EDUCATION, 2003, *Higher Education Act 101, 1997: Policy and procedures for measurement of research output of public higher education institutions*, Government Gazette, **460**(25583), Pretoria.
- [31] NATIONAL CENTER FOR EDUCATION STATISTICS (NCES), 1994, *Faculty and instructional staff: Who are they and what do they do?*, In 1993 National Study of Postsecondary Faculty, NCES 94-346, U.S. Department of Education, Washington (DC).
- [32] NATIONAL COMMISSION ON HIGHER EDUCATION (NCHE), 1996, [Government Report of 1996], Government Publications, Pretoria.
- [33] NEUMARK D, 1988, *Employers' discriminatory behaviour and the estimation of wage discrimination*, Journal of Human Resources, **23**, pp. 279–295.
- [34] OAXACA R, 1973, *Male-female wage differentials in urban labor markets*, International Economic Review, **14**, pp. 693–709.
- [35] RANSOM MR & MEGDAL SB, 1993, *Sex differences in the academic labor market in the affirmative action era*, Economics of Education Review, **12**, pp. 21–43.
- [36] ROUSSEEUW PJ, RUTS I & TUKEY JW, 1999, *The bagplot: A bivariate boxplot*, The American Statistician, **53**, pp. 382–387.
- [37] STEYN AGW & VERMEULEN PJ, 1997, *Perspektiewe op die finansiering van Suid-Afrikaanse universiteite*, Tydskrif vir Geesteswetenskappe, **37**, pp. 248–263.
- [38] STROBER MH & QUESTER AO, 1977, *The earnings and promotion of women faculty: Comment*, The American Economic Review, **67**(2), pp. 207–213.
- [39] TOUTKOUSHIAN RK, 1998, *Racial and marital status differences in faculty pay*, Journal of Higher Education, **69**, pp. 513–541.
- [40] TOUTKOUSHIAN RK, 1999, *The status of academic women in the 1990s: No longer outsiders, but not yet equals*, The Quarterly Review of Economics and Finance, **39**, pp. 679–698.
- [41] STELLENBOSCH UNIVERSITY, *Stellenbosch University's Employment Equity Policy*, [Online], [Cited May 5th, 2008], Available from <http://www.sun.ac.za/diensbillikheid/eng/>
- [42] VAN WYK DE VRIES COMMISSION OF ENQUIRY INTO UNIVERSITIES, 1974, [Government Report], Government Publications, Pretoria.
- [43] VENTER RH, 1985, *An investigation of government financing of universities*, SAPSE-110, [Government Report], Government Publications, Pretoria.
- [44] WARD M, 2001, *The gender salary gap in British academia*, Applied Economics, **33**, pp. 1669–1681.
- [45] WARMAN C, WOOLLEY F & WORSWICK C, 2006, *The evolution of male-female wages differentials in Canadian universities: 1970–2001*, Queen's Economics Department Working Paper No. 1099, Department of Economics, Queen's University.
- [46] WEICHELBAUMER D & WINTER-EBMER R, 2005, *A meta-analysis of the international gender wage gap*, Journal of Economic Surveys, **19**, pp. 479–511.

Appendix

1 One-way multivariate analysis of variance (MANOVA)

Consider a data matrix \mathbf{X} : $n \times p$ representing n p -variate independent homoscedastic samples, partitioned into g groups originating from distributions with possibly different means

$$\mathbf{X} : n \times p = [\mathbf{X}'_1, \dots, \mathbf{X}'_g]' \quad \text{where } \mathbf{X}_i : n_i \times p; \quad i = 1, 2, \dots, g, \quad \sum_{i=1}^g n_i = n.$$

The one-way MANOVA tests the hypothesis that all the group means are equal, and is based on the calculation of two statistics, both $p \times p$ matrices $\mathbf{B} : p \times p \equiv$ the “between groups” matrix and $\mathbf{W} : p \times p \equiv$ the “within groups” matrix.

In general, \mathbf{W} is nonsingular, whereas \mathbf{B} is of rank k , where $k = \min\{g - 1, p\}$. The forms of both \mathbf{W} and \mathbf{B} become much simpler if the data are centred, that is replaced by their deviations from the averages of the respective variables. This centring does not change the MANOVA and does only non-essentially modify the associated graphical displays. Therefore, we assume that \mathbf{X} was centred that way from the original data so that $\mathbf{1}'\mathbf{X} = \mathbf{0}'$.

2 Canonical variables

The obvious idea is to reduce a first stage MANOVA to a univariate analysis of variance (ANOVA), by replacing the matrix \mathbf{X} by a vector $\mathbf{X}\mathbf{v}$, containing the linear combinations of each sample’s values, with the same coefficients given by the vector \mathbf{v} . In this one-way ANOVA, the between- and within sums of squares are $\mathbf{v}'\mathbf{B}\mathbf{v}$ and $\mathbf{v}'\mathbf{W}\mathbf{v}$ respectively, and the F -test statistic is proportional to the ratio $\mathbf{v}'\mathbf{B}\mathbf{v}/\mathbf{v}'\mathbf{W}\mathbf{v}$. In the second stage, one will look for a vector \mathbf{v} maximizing this ratio; \mathbf{v} will be a solution of the two-sided eigenvalue problem $\mathbf{B}\mathbf{v} = \lambda\mathbf{W}\mathbf{v}$. This equation has p solutions, one optimal and $p - 1$ suboptimal, given as columns of a $p \times p$ matrix \mathbf{V} satisfying $\mathbf{B}\mathbf{V} = \mathbf{W}\mathbf{V}$, where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$; $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ and $\mathbf{V}'\mathbf{W}\mathbf{V} = \mathbf{I}$; $\mathbf{V}'\mathbf{B}\mathbf{V} = \mathbf{\Lambda}$.

The canonical variables are obtained by the transformation $\mathbf{y}' = \mathbf{x}'\mathbf{V}$, where \mathbf{x}' is any vector belonging to the row space of \mathbf{X} . The data matrix \mathbf{X} itself is transformed into the canonical variables values matrix $\mathbf{Y} = \mathbf{X}\mathbf{V}$. The canonical variables are sample “ \mathbf{W} uncorrelated” and have decreasing sample variances. For centred data \mathbf{X} , the g group averages of \mathbf{Y} have their last $p - k$ coordinates equal to zero. That means that by taking only the first k canonical variables, one obtains a perfect representation of the group means, although this does not hold for the individual samples.

A canonical variables plot is constructed by taking the first two canonical variables, to provide the coordinates for representing the n samples of \mathbf{X} as points in two dimensions. In this plot, one expects the group means to be represented by their canonical counterparts, better than the individual samples.

3 Canonical biplots

A canonical biplot consists of a canonical variables plot equipped with p linear axes, each one associated with one of the p variables. Each of these biplot axes is determined by a vector, which also induces a graduation on it.

There are two types of canonical biplots, each one characterized by its system of p linear axes, its aim and its corresponding geometry:

- The *interpolation biplot*, which has the aim of placing on the plot the image (y_1, y_2) of any new point \mathbf{x}' which belongs to the row space of \mathbf{X} .
- The *prediction biplot*, which has the aim of estimating the point \mathbf{x}' (*i.e.* the set of variable values) having as image a given point (y_1, y_2) in the plot.

After having placed the images of the group centroids on the canonical variables plot, we concentrate in this paper on the prediction biplot. In this plot, the p variable axes are given by the column vectors of the matrix obtained by taking the first two rows of the inverse \mathbf{V}^{-1} of \mathbf{V} . The points of concern will be the canonical centroids. For each one of them, the coordinates of the original centroid will be predicted by simply projecting in turn the canonical centroid onto each biplot axis representing one of the original variables and reading off the coordinate's estimated value.

The geometrical configuration obtained will yield a deep insight into the structure of the data at hand. At the same time, the “scatter plot”-like character of the prediction may be assimilated by the end user more readily than the scalar products of the classical Gabriel biplot.

