

# Geometric Data Analysis (GDA)

– an alternative approach to the analyses of gender differences

BY CLAUD D. HANSEN

---

## ABSTRACT

The aim of this paper is threefold: First, the criticism of quantitative methods raised by feminist and gender researchers is reiterated and illustrated using gender differences in job attribute preferences as an example. Second, the paper compares this ‘standard quantitative methods’ approach to Geometric Data Analysis (GDA), an approach that e.g. makes use of principal components analysis. I argue that GDA breaks with many of the problematic features of traditional statistics by being multi-dimensional (as opposed to one-dimensional), having a statistical model formulated at the individual level (as opposed to treating individuals as mere ‘residuals’) and visualising the results (as opposed to just presenting the results exclusively in numbers). Third, the empirical analyses from the first part of the paper are then used as an example and analysed again, thereby introducing the basic concepts and principles which comprise GDA. Data used in the paper stem from the study *Youth on the margin* where a sample of young men and women from the North Denmark Region were asked to fill out a battery of job attribute preferences among other things. This is an important topic because such preferences are widely thought to be closely related to the continuing segregation of the Danish labour market.

## KEYWORDS

*gender similarities, critique of quantitative methods, principal components analysis, survey.*

Claus D. Hansen is Associate Professor of Sociology at the Department of Sociology and Social Work, Aalborg University, Denmark. His research focuses on masculinity/gender studies, work environment and occupational accidents. His recent papers have appeared in NORA, BMC Psychology and Safety Science on topics such as ‘masculinity ideals’, ‘bullying’ and ‘masculinity and safety violations’.

Traditionally, quantitative methods have been frowned upon in feminist research (see Mies 1983; Risman 1993) and by extension in gender research in general. These methods have been viewed as anti-feminist and acting in the service of patriarchy by supporting the masculine bias inherent in much social science research, namely because it was conducted almost exclusively on male subjects and only concerned itself with topics defined by males (see Oakley 1998 for a discussion). In the words of Barbara DuBois, women were not only ‘unknown but virtually unknowable’ (DuBois 1983:107) due to the androcentrism of the social sciences.

While the feminist critique of quantitative methods is unique in certain respects (for example, the critique of masculine bias), many similarities with the criticism of these methods raised in sociology more generally remain. Herbert Blumer’s critique of ‘variable sociology’ (Blumer 1956), Pierre Bourdieu’s critique of ‘structural causality’ (Bourdieu 1984) and Andrew Abbott’s critique of ‘general linear reality’ (Abbott 1988) all point to problematic features of quantitative research with which the above mentioned feminist scholars would concur. Most notable among these criticisms is the lack of attention to context but also the use of unrealistic and untestable statistical assumptions (for a similar argument see Freedman 1991).

For many years, feminist scholars doing quantitative analyses have argued that there is not necessarily a discrepancy between doing feminist research and using quantitative methods (see Risman 1993). Gradually, a more pragmatic approach towards the use of quantitative methodologies in gender research has emerged, encouraged not only by scholars such as Risman but also by the increasing interest in ‘feminist scholarship’ by sociologists focused on the importance of gender for various research topics (Scott

2010). Instead of seeing certain methodologies as inherently problematic, the focus is redirected to the use these methods are put to. Kelly and colleagues argue that the disparagement of quantitative methods for being distanced, objective and keeping to the researcher’s agenda is similar to criticism originally raised regarding the qualitative research interview; “This suggests that what makes research feminist is less the method used, and more how it is used and what it is used for” (Kelly, Regan & Burton 2003: 150). Thus, if the method is used in pursuit of human emancipation it does not matter which method is employed.

It is, however, not only the perspective in gender research that has changed and become more inclusive of quantitative methods. Proponents of the latter have also changed. As Scott notes, ‘quantitative researchers are not naïve positivists’ (Scott 2010: 223). In fact, positivism is no longer – and has not been for a very long time – a viable ‘theory of science’ (Philips & Burbules 2000). The dominant *post-positivist* position, namely critical realism, emphasises the way in which knowledge is socially constructed and for that reason depends heavily upon well-constructed and reflective research designs. In other words, quantitative methodologies are not inherently more objective than qualitative methods. However, neither are they inherently more problematic because of their aim to generalise and to examine patterns or the correlation between two phenomena.

From the perspective of intersectionality theorists, one of the most problematic features of quantitative methods has been the uncritical use of taken-for-granted social categories. For example, this is notable in what Connell (2005) somewhat derogatorily calls ‘sex differences’ research. This type of research takes the categories of men and women and compares them on different parameters without at the same time

examining and acknowledging the possible heterogeneity inherent in these social categories, i.e. that women of colour with a working-class background are substantially different from white middle-class women; there is a lack of attention to context. In this paper, I want to examine an alternative quantitative approach that may be more compatible with a theoretical perspective that acknowledges some of the criticism of the use of social categories in social research. But before I introduce the approach dubbed ‘Geometric Data Analysis’ (GDA) by Le Roux and Rouanet (2004), I want to re-examine some of the problematic features of a ‘standard regression analysis’ approach often employed in quantitative sociology and in sex differences research, using segregated labour markets as an empirical example. Prior to the actual statistical analyses, it is worth framing the example of segregated labour markets by reviewing selected portions of the literature on the subject.

#### SEGREGATED LABOUR MARKETS AND JOB ATTRIBUTE PREFERENCES

Men and women often work in different occupations. This is particularly true of the Danish labour market which continues to be clearly segregated despite a trend of desegregating in recent years (EGGE 2009). In 2010, approximately a quarter of the workforce worked in occupations with an equal distribution of gender (i.e. no more than 60% of one gender) (Burchell, Hardy, Rubery & Smith 2014). For both men and women, the great majority work in occupations that are dominated by their own gender. However, men are slightly more likely to work in female-dominated occupations than the other way around: 21% of the employed males work in female-dominated occupations while the opposite is only the case for 11% of the employed females. It appears that male access to female-dominated occupations is easier than female ac-

cess to male-dominated occupations. Another way of measuring the gender segregation of the labour market is to calculate the share of employees needed to change jobs to achieve complete equality – in Denmark, this figure is approximately 25%, indicating that either half of the employed men or employed women would need to change jobs to achieve a non-segregated labour market.

Segregation in the labour market is interesting for several reasons: Gender segregation is widely thought to be one of the primary mechanisms through which gender inequality continues, for example in relation to the pay gap. This production of inequality by segregation takes place in many ways. First, because women still do a larger share of domestic work, segregation may be a consequence of ‘a *scarcity of jobs* where it is possible to combine work and family responsibilities or where those following non-linear careers may be recruited’ (Burchell et al. 2014: 29). Second, segregation may lead to *undervaluing work carried out by women* because it is easier to defend large wage differences across occupations and explain it as a consequence of ‘neutral’ market mechanisms. Third, segregation may also be linked with gendered expectations and stereotypes of what are considered appropriate occupational and – by extension – educational choices for men and women respectively. Segregation reinforces social norms, influencing the choices made by those in the labour market.

One should carefully note that segregation is not always a bad thing – nor has it negative consequences for women alone. Segregation may very well make it more difficult for men to take on a more active role in family responsibilities such as taking a bigger share of parental leave (Bekkengen 2006). In addition, this segregation also often results in a concentration of dangerous jobs in the hands of men who have a much higher risk of fatal occupational accidents than women (Jensen et al. 2014). Segrega-

FIGURE 1: GENDER DIFFERENCES IN JOB ATTRIBUTE PREFERENCES PERCENT

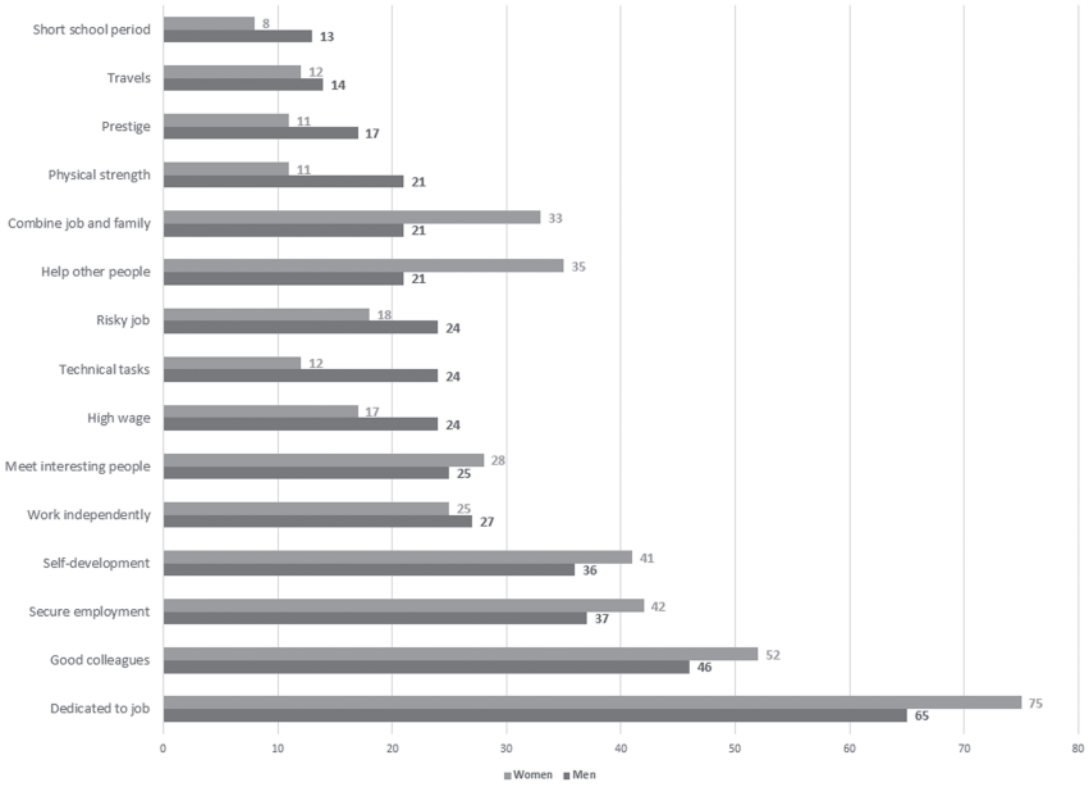


TABLE 1: JOB ATTRIBUTE PREFERENCES BY GENDER. PERCENT INDICATING JOB ATTRIBUTE TO BE ‘VERY IMPORTANT’.

	Men	Women	P-value	Cohen’s <i>d</i>
Help other people	21	35	<0.001	-0.33
Technical tasks	24	12	<0.001	0.31
Physical strength	21	11	<0.001	0.27
Combine job and family	21	33	<0.001	-0.28
Dedicated to job	65	75	<0.001	-0.21
High wage	24	17	<0.001	0.18
Prestige	17	11	<0.001	0.17
Short school period	13	8	<0.001	0.17
Risky job	24	18	<0.001	0.16
Good colleagues	46	52	0.004	-0.12
Secure employment	37	42	0.015	-0.10
Self-development	36	41	0.012	-0.10
Meet interesting people	25	28	0.219	-0.05
Travels	14	12	0.287	0.04
Work independently	27	25	0.476	0.03

tion may in some cases prove positive because it may make it easier for women to become part of the labour market.

One of the ‘root causes’ of gender segregation of the labour market identified is differences in educational aspirations, educational choices and job attribute preferences between genders. Various explanations for this exist, some emphasising the fundamental differences in men’s and women’s brains that result in different preferences for jobs, while other explanations highlight the influence of gendered norms and stereotypes in the choices young boys and girls are exposed to at an early age, resulting in clear differences in young men’s and women’s ideas about what constitutes a good job (for a discussion of this see Lippa 2005, 2010). If young men and women have different preferences for the types of job they strive for and women more often prefer jobs that are not so highly valued this helps to explain (and partly legitimise) the gender wage gap.

#### GENDER DIFFERENCES IN JOB ATTRIBUTE PREFERENCES. THE ‘STANDARD’ APPROACH

Having established that gendered preferences may be one prominent root cause of segregated labour markets, I am now in a position to carry out the statistical analyses of gender differences in job attribute preferences. The aim is first and foremost to show the problematic aspects of what could be called a ‘standard’ approach to this analysis.

The paper examines gender differences<sup>1</sup> in job attribute preferences by looking at a sample of young men and women from the North Denmark Region. The data used for the analyses are taken from *Youth at the margin*, a survey conducted in 2013 with a sample of 2,437 young men and women aged 16-25 enrolled at a secondary educational institution in the North of Denmark. Four classes at each educational institution

were selected randomly, except at those institutions where the pupils did not attend school continually, e.g. those attending vocational education training programmes. Fieldworkers administered the survey and were present in the classroom while the young people completed it online. In most cases, school staff was present as well. Overall, the response rate was almost 76%.

Part of the questionnaire asked the respondents to indicate how important 15 different job attributes were when considering the type of job, they would prefer. In figure 1 below, the share indicating the different job attributes to be ‘very important’ is presented by gender. In other words, figure 1 and table 1 constitute examples of the ‘standard’ sex differences analysis widely criticised by feminist scholars (see Eagly & Riger 2014).

As can be seen from table 1, for all but three of the job attributes a statistically significant difference<sup>2</sup> between men’s and women’s preferences exists: The young men prefer jobs with high wages and prestige that involves using their physical strength, whereas the young women prefer jobs where they can help other people, have good colleagues and combine job and family more easily. This is all very predictable and confirms the stereotypical view of gender differences in this area. In themselves, these results are not problematic; it all depends on the way they are interpreted. If the results are presented in a one-sided manner, we risk reinforcing stereotypical beliefs about men’s and women’s preferences and by extension legitimising the inequalities resulting from these differences by arguing that they just reflect that women have different preferences than men.

However, one thing is the fact that we have small p-values and by extension ‘statistically significant gender differences’. If we do not tread lightly, we risk committing the ‘practical significance fallacy’ (Kirk 1996) by relying solely on the p-value while ignoring the effect size of the difference and

its practical implications. In other words, by focusing on the p-value alone, we ignore discussing whether the actual differences in preferences between men and women are so large that they have practical implications (for a discussion of this see Kline 2013). The size of the p-value depends upon three things: the sample size (if we sample more people, the p-value will decrease even if the difference in means remains the same), the variability of the parameter estimated (if there are big differences within the groups compared this will result in larger p-values) and the effect size (if the ‘true’ difference between men and women in the population studied is bigger, the p-value will be smaller).

If we incorporate Cohen’s  $d$  as a measure of the effect size into the interpretation of gender differences in job attribute preferences, we reach a somewhat different conclusion; instead of focusing on the gender differences which exist in all but three of the job attributes measured, we find that in reality there are only small differences for five of the attributes and the rest are negligible. Cohen’s  $d$  is a standardised way of measuring the size of differences so researchers can compare these across various studies; it was first suggested by Jacob Cohen (1988). As a rule of thumb, Cohen suggests that a value of  $d = 0.2$  indicates a small difference in means, a  $d$ -value of 0.5 indicates a medium-sized difference and a  $d$ -value of 0.8 indicates a large difference. Table 1 is arranged by the absolute value of Cohen’s  $d$ , and you can see that the largest difference in means between the young men and women is -0.33, which is, at most, classified as a small difference.

In order to fully appreciate what this means, we can try to translate Cohen’s  $d$  into something more easily understood. For a  $d$ -value of 0.3, we would expect that a randomly chosen young woman would have a 58% chance of preferring jobs where she could ‘help other people’ if we compared her to a randomly chosen young

man. In the case of complete similarity of preferences between men and women, the probability would have been 50%, i.e. the difference (8 percentage points) is not very large. Another way of stating this would be that 62% of the women would have a higher likelihood of selecting ‘helping other people’ as a very important job attribute. If there had been no difference, only 50% of the women would have had a higher likelihood of indicating a preference for this job attribute.

Bear in mind that I am commenting on the *largest* difference in job attribute preferences here. For most job attributes, the difference between men and women is even smaller. This latter point is clearly linked to the criticism raised of much sex differences research that states that the exclusive interest in establishing whether there is a statistically significant difference *between* men and women in the mean of some measure overshadows the fact that the difference *within* the group is often bigger (Connell 2005: 21-22). While there are some sex differences that are not only statistically significant but also of substantial interest (see the discussion in Lippa 2005), the general picture is that men and women are much more alike than different and that sex difference research therefore has a much too narrow focus (Hyde 2005).

Overall, the ‘standard approach’ to examining gender differences is likely to lead to an exaggeration of the actual differences between men and women because the focus is primarily on the test of significance instead of the evaluation of the size of the difference and its ‘practical significance’. This approach thus risks essentialising differences between men and women even when there are only very minor ones (Prentice & Miller 2006). It should be noted that the use of these statistical techniques in themselves does not necessarily result in these consequences. As is the case with methods in general, it is all about the way they are employed in specific analyses.



However, as can be seen from the discussion above, it is not wrong to state that gender differences have very often been examined in this way without further contextual analysis.

#### GEOMETRIC DATA ANALYSIS – A SHORT INTRODUCTION TO MULTI-DIMENSIONAL ANALYSES

Having described some problematic features of the ‘standard approach’ to examining gender differences, I can now turn to the second aim of this paper: to introduce GDA and make a comparison to the approach described above.

The term geometric data analysis is the term suggested by Patrick Suppes to cover the statistical approach originally developed by J.P. Benzécri from the 1960s onwards (Suppes 2004). Correspondence Analysis (CA) is probably the most famous of the techniques connected to this label, but it covers Multiple Correspondence Analysis (MCA) and Principal Components Analysis (PCA) as well. The central difference between the three approaches has primarily to do with the types of data tables analysed. When analysing questionnaire data such as the data used in this paper, MCA (understood as Individuals x Variables table) is the most relevant. However, depending on the types of variables included in the questionnaire, PCA may also be used.<sup>3</sup>

Central to GDA is a unique data analysis philosophy where description is taken as the most important task of statistics: “Description comes first. Statistics is not probability!” (Le Roux & Rouanet 2004: 6). This is important in this context because one of the criticisms raised of traditional uses of statistics in gender research is the lack of attention to context. In GDA, context is important because we are primarily interested in letting “the data speak for themselves!” (Le Roux & Rouanet 2004: 10). By approaching the data in an exploratory manner, we can more easily see

things that do not conform to our (androcentric?) preconceptions. We are also less prone to commit the ‘significance fallacy’ described above by being able to distinguish between statistically significant results and results that are of substantial sociological significance.

In GDA, the focus is first on describing the associations established in the analysis and to ascertain, which of the associations are large enough to warrant a more thorough examination. Through the use of concentration ellipsoids and the geometrical and visual output connected to GDA, the approach also differs from mainstream statistical methods in being able to highlight and take into account the heterogeneity of the social categories being analysed. This is intimately linked to the fact that the statistical model in GDA is ‘formulated at the individual level’, and for this reason, Le Roux & Rouanet (2004: 15) discuss how GDA brings about ‘the rehabilitation of individuals’ in contrast to traditional statistical approaches. As we shall see in the final section of this paper, this means that we are able to visualise the individual differences between men and women in a much more detailed way than the traditional approach allowed for.

Having analysed the preferences by examining the overall gender differences, the next step in the ‘standard quantitative approach’ would often be to choose one or two preferences and analyse them in more depth using regression analysis techniques. This would then enable us to determine, which factors were responsible for ‘explaining’ the gender difference detected in the cross tables presented above. We could, for example, examine to what extent the difference in preference for jobs where ‘solving technical tasks’ is involved could be explained by differences in ideologies of masculinity or femininity. We might conclude that young women prefer this type of job to a lesser degree because stereotypical beliefs about what is feminine are not easily

compatible with a job where you primarily ‘solve technical tasks’. While this approach has many merits if one is interested in specific research questions, it may not be the best choice if one’s interest is broader analysis and if you seek to identify overall patterns in the data. GDA serves as an alternative in this respect because the complete set of job attribute preferences would be included in a *joint* analysis. The aim of this analysis would be to construct a ‘space of job attribute preferences’.

In the social sciences, GDA has been compared to “the construction of social spaces” (Le Roux & Rouanet 2004: 15). For instance, in Pierre Bourdieu’s study *The Distinction* (1984), he constructs a ‘space of lifestyles’, and the analysis of this social space shows that the basic divisions in this are related to differences in economic, cultural and social capital. A precondition for constructing social spaces is *homogeneity* between the variables entered into the analysis as well as *exhaustivity*. Homogeneity requires that the variables analysed are comparable, for example, by having the same number of response categories. As Le Roux and Rouanet note (2004: 14), *exhaustivity* is a radically different ideal from the ideal strived for in the ‘standard approach’ advanced before, namely *parsimony*. Parsimony entails making the model and analyses as simple as possible by including as few variables as possible and by focusing exclusively on how *one* independent variable is associated with *one* dependent variable. By taking a more holistic approach

and examining the potential multi-dimensional nature of the phenomenon under investigation as well as including the individuals much more directly in the interpretation of the analysis, GDA breaks with some of the problematic features of the ‘standard approach’ criticised above. Having said that, the best way to illustrate the advantages of GDA is to conduct a small case study using the same data on job attribute preferences we have already examined.

### CONSTRUCTING A SPACE OF JOB ATTRIBUTE PREFERENCES USING GDA

As mentioned above, the precondition for using GDA is homogeneity and exhaustivity in the data used. The variables are clearly homogenous as they all have the same 4 point Likert scale response categories (‘very important’, ‘important’, ‘less important’, ‘not at all important’). One can debate whether the preferences are exhaustive of overall job preferences; however, the 15 job attributes featured in the analysis cover approximately half of those analysed in a review of sex differences in job attribute preferences (Konrad, Ritchie, Lieb & Corrigan 2000). In this case, the 15 variables are first standardised using ‘personal equation rescaling’ (Murtagh 2005: 98-99). After this, the variables are *doubled* to facilitate the interpretation of the principal components analysis (Greenacre 2007: 184). The first step of GDA is to determine how many principal axes are needed to adequately grasp the social space being con-

TABLE 2: PRINCIPAL COMPONENTS ANALYSIS OF JOB ATTRIBUTE PREFERENCES. EIGENVALUES AND VARIANCE RATES.

Principal variable	Eigenvalue	Variance rates	Cumulative variance rates
1	2.29	16.36	16.36
2	1.56	11.17	27.54
3	1.48	10.58	38.12
4	1.21	8.61	46.72



structed. One important rule of thumb is that all principal axes with an eigenvalue  $> 1$  are retained for the analysis because these principal axes explain more variance in the space of job attribute preferences than the remaining variables do. If we focus only on the principal axes with an eigenvalue  $> 1$ , we can derive the following table of the eigenvalues and rates of explained variance.

As can be seen from table 2, we need at least four principal axes to account for the diversity and variation in job attribute preferences. I write at least four dimensions because another criterion used to determine the number of axes retained in the analysis is the cumulative rate of explained variance that preferably needs to be at least 75%. In this case, the cumulative rate of variance is 47%, indicating that the first 4 principal axes account for only half of the variance in the space of job attribute preferences. This is somewhat small and points to a very complex geometric space with multiple dimensions that are only partly summed up by the analysis carried out. One could argue that the relatively small level of explained variance indicates that a joint analysis of the job attributes is not ideal because the preferences are, to a great extent, independent of each other.

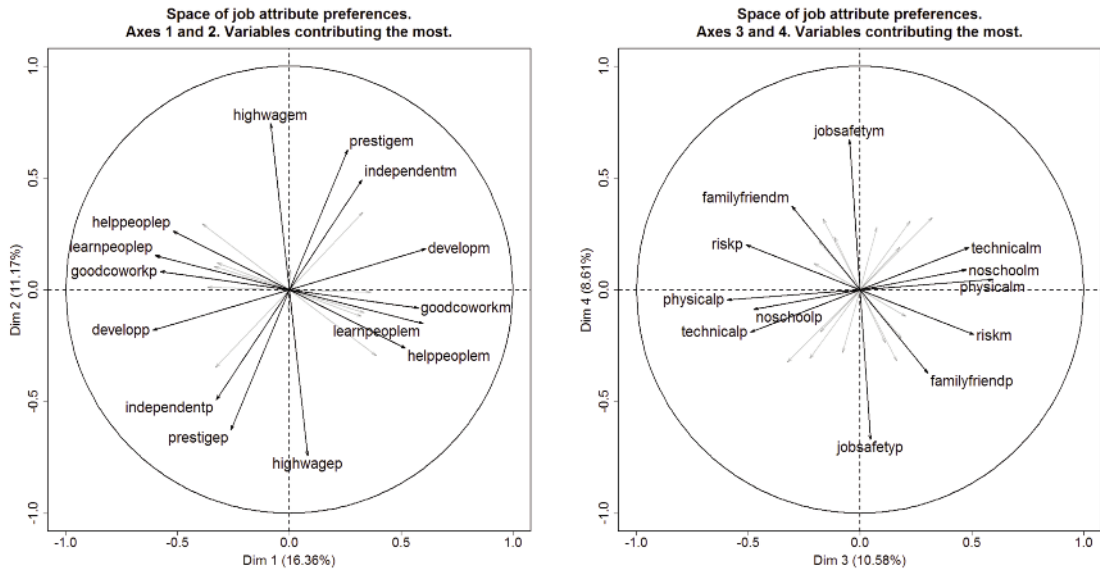
The next step in the analysis is to determine how to interpret the principal axes. In the words of Benzécri:

Interpreting an axis amounts to finding out what is similar, on the one hand, between all the elements figuring on the right of the origin and, on the other hand between all that is written on the left; and expressing with conciseness and precision, the contrast (or opposition) between the two extremes. (Benzécri 1992, 405).

This is done by selecting the variables with the highest contributions for each of the principal axes retained. The contributions are a measure of how important each job attribute preference is for the principal axes

derived from the analysis. The average contribution, in this case, is 3.33<sup>4</sup> because there are 15 variables which are double-coded. This means we must focus on variables that have the greatest contribution to the axis because they are most important for the interpretation. In the case of axis 1, there are four variables clearly important for the interpretation, here listed in order of importance: 'Self-development' (8.1), 'Meet interesting people' (7.8), 'Good colleagues' (7.3) and 'Help other people' (5.9). Apart from the absolute contribution, we also need to focus on the signs of the variables. In this case, all of the above have the same positive sign which means that individuals who prefer jobs with these attributes are situated left of the origin with negative coordinates on the first axis, whereas those who do not prefer these job attributes have positive coordinates on this axis. This procedure is repeated for each of the four principal axes retained for the analyses, and when this has been done, the analysis is presented visually as has been done in figures 2a and 2b. These two figures show the result of the principal components analysis. The arrows indicate the different job attribute preferences and how they are located in the space of job attribute preferences. There are two arrows for each job attribute because of the doubling procedure applied; for example, the label named highwagep (p for plus) situated at the lower end of figure 2a indicates that individuals with a preference for jobs that lead to high wages are most likely situated in the lower part of the figure. In contrast, highwagem (m for minus) indicates that individuals situated in this part of the cloud have a lower preference for this type of job relative to the other preferences they have rated. The overall interpretation of figure 2a is that it distinguishes between individuals motivated by *extrinsic factors*, most notably wages and prestige, along the second principal axis (the vertical axis in the figure). Along the horizontal axis, the ana-

FIGURE 2A+B:



lysis distinguishes between those motivated by *social factors* such as being able to ‘meet interesting people’, ‘help others’ and have ‘good co-workers’ and those who are not motivated by these.

If we turn to figure 2b, this shows the third and fourth principal axes that emphasise other job attributes. Along the third axis, the analysis distinguishes between those preferring a job that lets you use your ‘physical strength’, ‘solve technical tasks’, ‘involves risk and thrill’ and where you ‘do not have to go to school for many years’. Along the fourth axis, the most important distinction is between those preferring jobs that are ‘stable and secure’ vs. those who do not have that preference. Being able to combine the job with children and family is also situated along this fourth principal axis leading to an interpretation of the axis as one that emphasises ‘family friendliness’ in terms of both flexibility as well as security.

As can be seen from this very brief analysis of the space of job attributes, it would have been immensely difficult to do justice

to the complexity of this phenomenon by choosing one or two job attribute preferences and focusing on them in a more advanced analysis than the one presented above. The job attributes are associated with each other, and this overarching analysis is better able to acknowledge this complexity (although it, of course, still simplifies things, which can be clearly seen because the analysis only accounts for 47% of the variance in the original point cloud and the quality of the representation ( $\cos^2$ ) for most of the variables that contribute the most is below 50%).

EXPLORING GENDER DIFFERENCES IN JOB ATTRIBUTE PREFERENCES IN THE SPACE OF INDIVIDUALS

Having constructed the space of job attribute preferences, we are now able to return to the question of gender differences in job attributes. However, instead of comparing the job attributes individually as we did before, we will now examine where the

TABLE 3: EUCLIDEAN DISTANCES IN THE SPACE OF JOB ATTRIBUTE PREFERENCES BY GENDER.

Principal axis	Distance (in standard deviations)	Mean coordinate of women	Mean coordinate of men
1. Social aspects of job	0.44	-0.221	0.219
2. Extrinsic factors, e.g. wage	0.18	0.089	-0.088
3. Technical and physical aspects of job	1.01	0.505	-0.501
4. Job security and family	0.19	-0.097	0.096

individual men and women are situated in this multi-dimensional space of job attribute preferences. In GDA, the analysis consists of both a space of variables (already commented on above) as well as a *space of individuals* that is complementary to it. Even though gender was not part of the active variables constituting the space of job attribute preferences, we can examine to what extent men and women are situated differently in the space by utilising gender and other structuring factors as *supplementary variables* that are inserted into the analysis. In table 3, I have listed the mean distance between men and women for the first four principal axes we analysed above. As a rule of thumb, Le Roux and Rouanet (2004: 234) suggest using a distance of more than 1 standard deviation between the mean points of two categories in the space of individuals as a large difference, whereas differences below 0.5 are regarded as 'small'. As can be seen from table 3, in the multi-dimensional analysis, we are once again reminded that men and women are very similar.

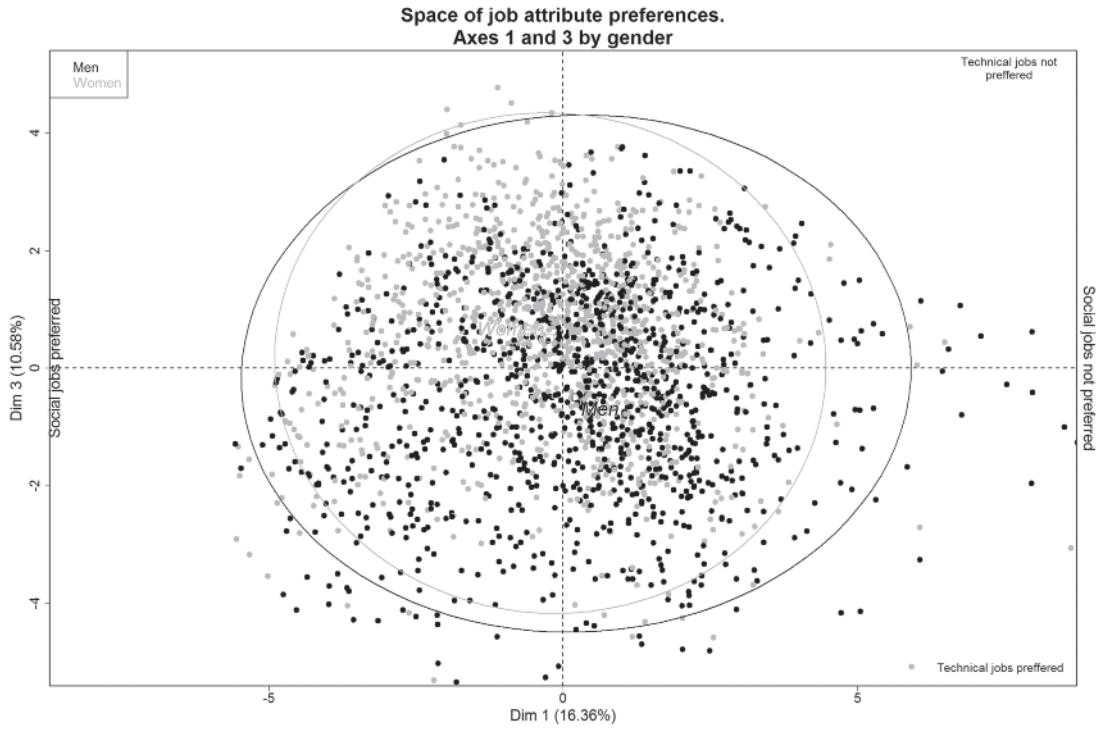
There is a tendency for the women to be slightly more oriented towards social aspects of future jobs, but the difference between men and women is just below the threshold of 0.5. In addition, there seems to be a relatively large difference when it

comes to preferring jobs where you need to solve technical tasks, use your physical strength and where you do not need to go to school for too many years. For the third principal axis, the mean difference between men and women is large and suggests that men more often than women prefer these types of physical and technical jobs.

Another way of illustrating this gender similarity is to visualise the differences between men and women by plotting the space of individuals for the two principal axes with the largest mean differences between men and women. In figure 3, axis 1 and 3 are plotted against each other to yield a social space that distinguishes between those individuals preferring jobs with clear social dimensions or not along the horizontal axis and on the vertical axis between those preferring jobs with a technical element versus those who do not prefer this. As can be seen from the figure, the mean point of women is situated in the first quadrant of the space, i.e. the average woman in the sample prefers jobs that are more social than not, as well as being more non-technical than not. In contrast, the average man prefers a job that does not involve social elements but technical skills (relative to his other preferences).

However, if we disregard the mean point – reminding ourselves that the mean

FIGURE 3:



(as was the case in the first part of the analysis) can cover quite different situations – and instead focus on the actual distribution of men and women in the figure, we are reminded to state the conclusion of this part of the analysis more modestly. In the figure, the grey dots are the individual women and the black dots are the individual men. There is a tendency for the women to be skewed in the direction of the first quadrant and for more men to have negative coordinates along the third principal axis. However, for the great majority of men and women, there is a very large overlap between their positions in the space of job attribute preferences even for the two axes where we would expect the largest differences. The overlap is also very clearly illustrated by the concentration ellipsoids drawn up in figure 3. The size of the ellip-

soids also indicates that the group of men is somewhat more heterogeneous than the group of women.

The use of supplementary variables can be extended to more variables, and if we recall some of the important points raised by McCall (2005) in her discussion of different approaches to intersectional analyses, one could argue that GDA offers the potential to carry out *intracategorical* analyses quite easily. The intracategorical approach, McCall (2005: 1774) writes, acknowledges the stable and even durable relationships that social categories represent at any given point in time, though it also maintains a critical stance towards categories. It is especially the latter point that GDA is particularly apt at illustrating as we saw in figure 3: although the analysis of the job attribute preferences shows that on average more

TABLE 4: EUCLIDEAN DISTANCES IN THE SPACE OF JOB ATTRIBUTE PREFERENCES BY THE INTERSECTION OF GENDER AND EDUCATION

Intersection of gender and education	Coordinates on principal axes (mean points for category)			
	1. Social aspects of job	2. Extrinsic factors, e.g. wage	3. Technical and physical aspects of job	4. Job security and family
Female – Vocational training (EUD)	-0.944	0.567	-0.610	0.320
Male – Vocational training (EUD)	-0.317	0.246	-1.343	0.157
Female – Gymnasium (STX)	-0.142	0.077	0.817	0.008
Male – Gymnasium (STX)	0.405	0.005	-0.176	-0.390
Female – Commercial (HHX)	-0.019	-0.454	0.501	0.295
Male – Commercial (HHX)	0.401	-0.923	0.020	0.023
Female – Technical (HTX)	0.393	-0.292	0.199	-0.053
Male - Technical (HTX)	0.436	0.001	-0.495	0.115

men than women prefer, for example, technical jobs, the concentration ellipsoids acknowledge that there is a great heterogeneity within both genders.

Let me illustrate this point with one more example from the analysis of job attribute preferences where I insert the intersection of gender and education in the space of individuals by creating a variable where I have combined the categories from the two variables. This is admittedly still a very simple analysis from an intersectionality perspective but nevertheless serves to illustrate my main argument here. Table 4 shows the mean coordinates of the different combinations of gender and education. We can interpret the distances in the same manner as in table 3 above, designating distances of more than 1 standard deviation as large and distances below 0.5 as small.

From this, we can see that the social aspects of a job are particularly important to women taking vocational training which is not so surprising as this will include women training to become a home carer or taking a basic course in pedagogy. Men and women taking vocational training are somewhat

different as illustrated by the distance of 0.627 between the categories whereas men and women attending a technical gymnasium are very similar. Here, the distance is only 0.043, i.e. virtually none. However, if we compare the men and women attending a technical gymnasium, they are somewhat different when it comes to the importance of preferences for technical tasks and physical work summarised along the third principal axis.

Space does not permit a detailed substantive analysis of the intersection between gender and education in the space of job attribute preferences. However, the above table illustrates that it is quite possible to carry out intracategorical intersectional analyses as part of the GDA framework.

## CONCLUSION

In this paper, I have tried to argue for the use of GDA as a better statistical method for the analysis of gender differences and similarities when compared to more standard quantitative approaches often used in ‘sex differences’ research. GDA has, in par-

ticular, three merits that are worth emphasising: 1) The switch from a one-dimensional to a multi-dimensional analysis of the research object. This is an improvement over standard quantitative approaches which often oversimplify the phenomena investigated in order to employ regression techniques in the analysis. Because the dimensions analysed in GDA are not specified in advance, this also means a more inclusive approach that has the potential to be more sensitive to complex phenomena. One should note, however, that more standard regression-like types of analyses could also be utilised in a more exploratory manner. This is, however, not very often the case. 2) The statistical model used in GDA is formulated on an individual level, which means that the actual persons analysed are not merely ‘residuals’, as is the case in standard regression techniques, but instead ‘carry all the information’. This should also encourage a more modest approach in the use of quantitative methods that acknowledge the importance of the individuals that form the basis of the analysis. 3) Finally, an important part of GDA is the visualisation of the results of the analyses. Not only are the associations between the variables that go into the analysis visualised in order to interpret the meaning of the principal axes, but the exploration of the space of individuals makes it even clearer when we should be careful not to overstate the differences found in our statistical analyses – not least because of the “significance fallacy: confusing significant deviations with important ones and non-significant deviations with negligible ones” (Le Roux & Rouanet 2004: 64). By using GDA, we are less likely to commit that mistake, leading one to conclude that perhaps quantitative methods should be given a bigger role in gender research in the future than they do today.

## NOTES

1. I use the term gender differences even if I only analyse differences between the two demographic groups, men and women. This is to indicate that I assume that most of the differences are of a social nature and therefore not related to biological differences between men and women. Measuring gender with only these two categories is neither ideal nor sufficient if you are trying to break away from the problematic aspects of taken-for-granted social categories (for a recent discussion see Westbrook & Saperstein 2015).
2. It is beyond the scope of this paper to introduce basic statistical concepts such as the significance test and the p-value. But when I argue that a common way of interpreting table 1 is to focus on whether the p-value for the comparison between the mean between men and women is below 0.5, this is in line with how widely used textbooks on statistical methods in the social sciences teach these things (see Agresti & Finlay 2014; Gliner, Leech & Morgan 2002).
3. A ‘Individuals x Variables’ table is a standard data matrix in which each row represents an individual who answered a questionnaire and each column represents the variables in the questionnaire. The table can be coded in different ways, depending on the type of GDA procedure used: for categorical variables one would normally use MCA, and for numerical variables we would instead use PCA (for a discussion of this see Le Roux & Rouanet 2004).
4. The average contribution is calculated by dividing 100 by the number of variables used in the analysis. In this case, we have 15 job attribute preferences, but as we have doubled them (i.e. coded them both negatively and positively), we have 30 variables entered into the analysis, which gives  $100/30 = 3.33$  for the average contribution.

## REFERENCES

- Abbott, Andrew (1988): Transcending General Linear Reality, in *Sociological Theory* 1988/6.
- Agresti, Alan & Finlay, Barbara (2014): *Statistical Methods for the Social Sciences*. Pearson, Harlow.
- Bekkengen, Lisbeth (2006): Men’s parental leave: A manifestation of gender equality or child-orientation, in: Lena Gonäs & Jan Ch. Karlsson (eds.): *Segregating. Divisions of Work in Post-Industrial Welfare States*. Ashgate, Aldershot.
- Benzécri, J.P. (1992): *Correspondence Analysis Handbook*. Marcel Dekker, New York.



- Blumer, Herbert (1956): Sociological Analysis and the 'Variable', in: *American Sociological Review* 1956/6.
- Bourdieu, Pierre (1984): *Distinction. A Social Critique of the Judgement of Taste*. Routledge, London.
- Burchell, Brendan; Hardy, Vincent; Rubery, Jill & Smith, Mark (2014): *A New Method to Understand Gender Segregation in European Labour Markets*. European Commission, Luxembourg.
- Cohen, Jacob (1988): *Statistical Power Analysis for the Behavioral Sciences. Second Edition*. Erlbaum, Hillsdale.
- Connell, R.W. (2005): *Masculinities. Second Edition*. Polity, Cambridge.
- DuBois, Barbara (1983): Passionate Scholarship: Notes on Values, Knowing and Method in Feminist Social Sciences, in: Gloria Bowles & Renate Duelli Klein (eds.): *Theories of Women's Studies*. Routledge and Kegan Paul, London.
- Eagly, Alice & Riger, Stephanie (2014): Feminism and Psychology. Critiques of Methods and Epistemology, in: *American Psychologist* 2014/7.
- EGGE – European Commission's Expert Group on Gender and Employment (2009): *Gender segregation in the labour market. Root causes, implications and policy responses in the EU*. European Commission, Luxembourg.
- Freedman, (1991): Statistical Models and Shoe Leather, in: *Sociological Methodology* 1991/21.
- Gliner, Jeffrey; Leech, Nancy & Morgan, George (2002): Problems With Null Hypothesis Significance Testing (NHST): What Do the Textbooks Say?, in: *The Journal of Experimental Education* 2002/1.
- Greenacre, Michael (2007): *Correspondence Analysis in Practice. Second Edition*. Chapman & Hall/CRC, Boca Raton.
- Hyde, Janet (2005): The Gender Similarities Hypothesis, in: *American Psychologist* 2005/6.
- Jensen, Sune Q.; Kyed, Morten; Christensen, Ann-Dorte; Bloksgaard, Lotte; Hansen, Claus D. & Nielsen, Kent (2014): A gender perspective on work-related accidents, in: *Safety Science* 2014/64.
- Kirk, Roger (1996): Practical Significance: A Concept Whose Time Has Come, in: *Educational and Psychological Measurement* 1996/5.
- Kline, Rex (2013): *Beyond Significance Testing: Statistics Reform in the Behavioral Sciences. Second Edition*. American Psychological Association, Washington.
- Konrad, Alison; Ritchie, J. Edgar; Lieb, Pamela & Corrigan, Elizabeth (2000): Sex Differences and Similarities in Job Attribute Preferences: A Meta-Analysis, in: *Psychological Bulletin* 2000/4.
- Le Roux, Brigitte & Rouanet, Henry (2004): *Geometric Data Analysis. From Correspondence Analysis to Structured Data Analysis*. Kluwer Academic Publishers, Dordrecht.
- Lippa, Richard (2005): *Gender, Nature and Nurture. Second Edition*. Psychology Press, New York.
- Lippa, Richard (2010): Sex Differences in Personality Traits and Gender-Related Occupational Preferences across 53 Nations: Testing Evolutionary and Social-Environmental Theories, in: *Archives of Sexual Behavior* 2010/3.
- McCall, Leslie (2005): The Complexity of Intersectionality, in: *Signs* 2005/3.
- Mies, Maria (1983): Towards a methodology for feminist research, in: Gloria Bowles & Renate Duelli Klein (eds.): *Theories of Women's Studies*. Routledge and Kegan Paul, London.
- Murtagh, Fionn (2005): *Correspondence Analysis and Data Coding with Java and R*. Chapman & Hall/CRC, Boca Raton.
- Philips, D.C. & Burbules, Nicholas (2000): *Post-positivism and Educational Research*. Rowman & Littlefield, Lanham.
- Prentice, Deborah & Miller, Dale (2006): Essentializing Differences Between Women and Men, in: *Psychological Science* 2006/2.
- Reinharz, Shulamit (1993): Neglected Voices and Excessive Demands in Feminist Research, in: *Qualitative Sociology* 1993/1.
- Risman, Barbara (1993): Methodological Implications of Feminist Scholarship, in: *The American Sociologist* 1993/3-4.
- Scott, Jacqueline (2010): Quantitative methods and gender inequalities, in: *International Journal of Social Research Methodology* 2010/3.
- Suppes, Patrick (2004): Foreword, in: Brigitte Le Roux & Henry Rouanet: *Geometric Data Analysis. From Correspondence Analysis to Structured Data Analysis*. Kluwer Academic Publishers, Dordrecht.
- Westbrook, Laurel & Saperstein, Aliya (2015): New Categories Are Not Enough, in: *Gender & Society* 2015/4.
- Westmarland, Nicole (2001): The Quantitative/Qualitative Debate and Feminist Research: A Subjective View of Objectivity, in: *Forum: Qualitative Social Research* 2001/1.