INSTITUTE FOR RESEARCH ON POPULATION AND SOCIAL POLICIES
INSTITUT DE RECHERCHE SUR LA POPULATION ET LES POLITIQUES SOCIALS

edited by - *édité par*
**Rossella Palomba - Adele Menniti**

# Minerva's Daughters
# *Filles de Minerve*

2003

# Appendix[22]

## 1. Introduction

This appendix describes how gender inequalities in the scientific field can be measured by means of composite indices (most of which have been used in this report), based on readily available data.

Capturing such a complex reality in a single, simple index is not easy. Therefore a set of different indicators and statistical methods will be described and the implementation of each will be discussed.

To begin with we can distinguish two different types of gender segregation:

- horizontal segregation which measures the concentration of women and men in some disciplinary sectors;
- vertical segregation, which concerns the positions of women and men in the hierarchies of science.

For each type of segregation one or more indicators will be described.

The reader must bear in mind that a single measurement of segregation on its own is of little interest; it is only through comparisons that we can give any meaningful interpretation to particular values. Such comparisons may involve trends over time or may be across different structures, for example comparing levels across countries, or across different research bodies or disciplinary sectors in the same country. In what follows we will generally define the various indices with respect to different disciplinary sectors, indicated by the suffix $i$.

Moreover we must stress that the computation of specific indicators should follow a preliminary descriptive analysis of the situation through appropriate tables and plots differentiated by gender, such as the distribution over time of male and female researchers at the different levels, the composition of the scientific boards, the share of female researchers by field of science, and so on.

After this preliminary analysis some further investigations should be made.

To summarise, the following flow-chart can be useful (Scheme A).

[22] The Appendix was written by Anna Gigli, IRPPS.

es in the scientific field
:s (most of which have
able data.
imple index is not easy.
istical methods will be
be discussed.
nt types of gender seg-

:oncentration of women

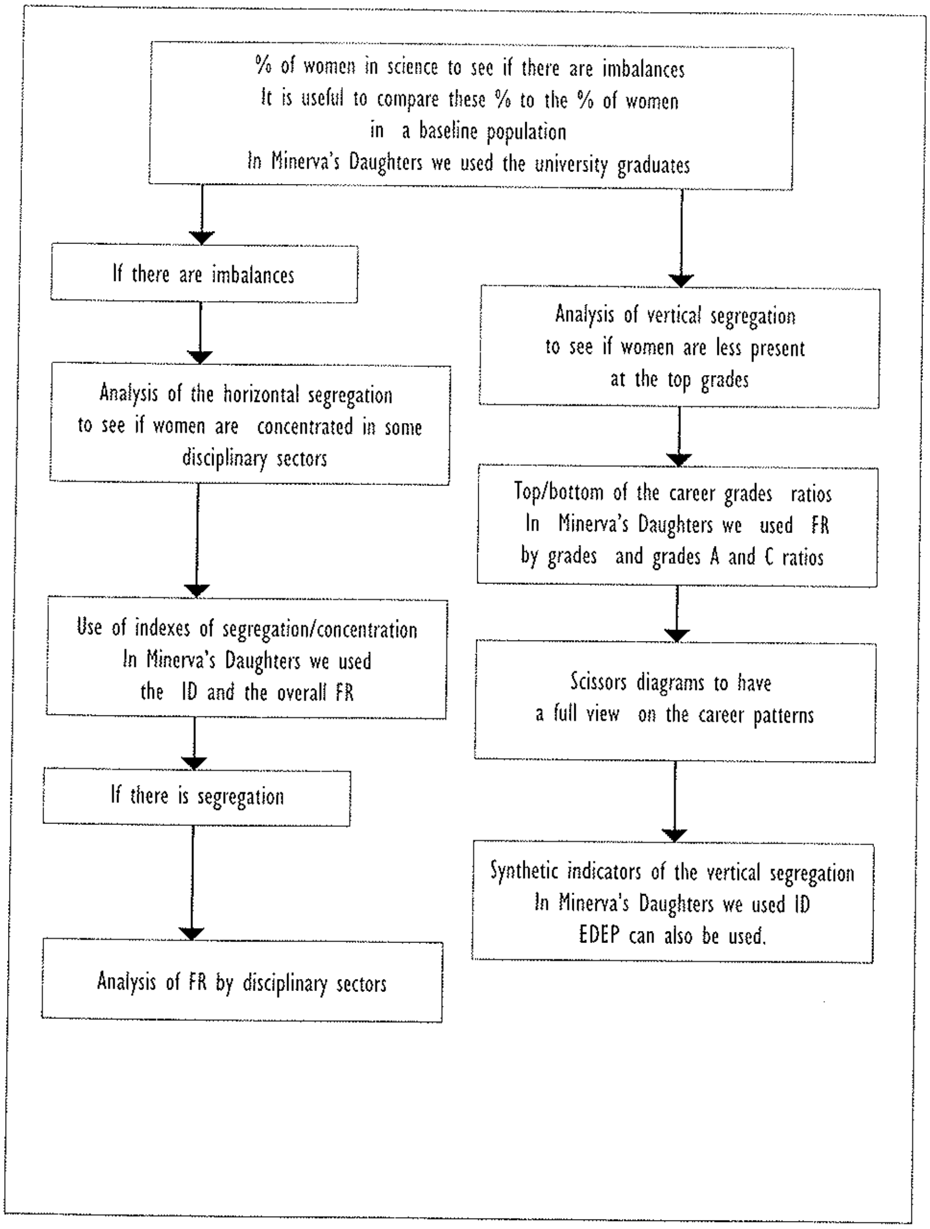
sitions of women and

cators will be described.
ieasurement of segrega-
irough comparisons that
particular values. Such
may be across different
oss countries, or across
s in the same country. In
ous indices with respect
he suffix *i*.
on of specific indicators
ialysis of the situation
iated by gender, such as
researchers at the differ-
iards, the share of female

investigations should be

be useful (Scheme A).






## 2. Horizontal segregation

### 2.1 *The Feminisation Ratio*

The Feminisation Ratio is the number of women present in a certain sector for every 100 men present in the same sector; in our case, if $F_i$ is the number of female researchers working in disciplinary sector *i*, and $M_i$ is the number of male researchers working in the same sector, the Feminisation Ratio is the ratio between $F_i$ and $M_i$ multiplied by 100:

$$FR_i = 100*(F_i/M_i)$$

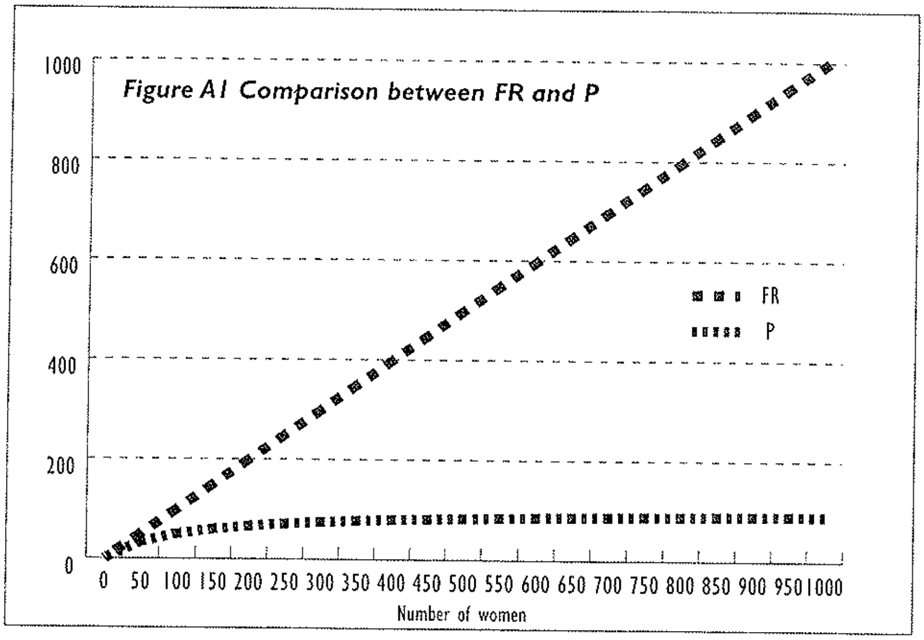
Note that the Feminisation Rate is different from the usual percentage (P) in which the baseline is the overall number of researchers (males and females): in disciplinary sector $i$, the percentage of female researchers $P_i$ is given by

$$P_i = 100* F_i/(F_i+M_i)$$

The indices FR and P are linked via the following mathematical relationship:

$$P = 100*FR/(100+FR)$$

which can be better interpreted by looking at a plot of FR and P computed for various values of female researchers per 100 male researchers (Figure A1).




When there are no women researchers in a certain sector (and therefore there is a total segregation in that sector) both $FR_i$ and $P_i$ take the value zero; if there is no segregation $FR_i$ takes the value 100 (there are 100 female researchers per 100 male researchers in sector $i$) and $P_i$ takes the value 50; when the number of female researchers grows and overtakes the number of male researchers, $FR_i$ takes values greater than 100 (with no upper limit), while $P_i$ takes values greater than 50, but cannot exceed 100. For this reason FR is better at evaluating even small i)

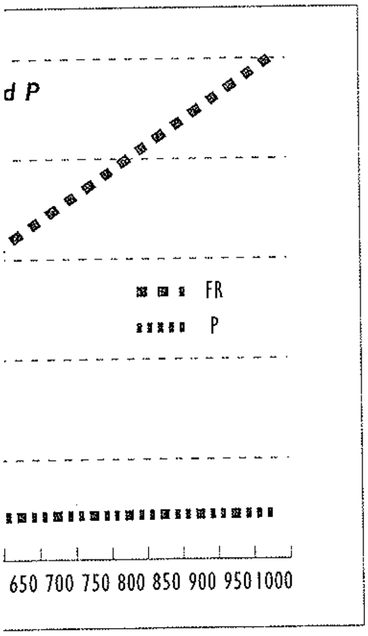
ıt from the usual percentage
mber of researchers (males
the percentage of female

$M_i$)

ollowing mathematical rela-

FR)

g at a plot of FR and P com-
ers per 100 male researchers




ı a certain sector (and there-
tor) both $FR_i$ and $P_i$ take the
akes the value 100 (there are
earchers in sector *i*) and $P_i$
male researchers grows and
$FR_i$ takes values greater than
values greater than 50, but
etter at evaluating even small



differences, while P tends to be less sensitive, hence FR is more suitable for studies on gender differences.

## 2.2 *The Dissimilarity Index*

An overall measure of horizontal gender segregation across different disciplinary sectors is given by the Dissimilarity Index, which provides an indication of the proportion of male and female researchers who should change their occupation in order for the proportion of females to be identical in all disciplinary sectors. Mathematically, if *i* denotes the *i-th* disciplinary sectors, it is expressed as:

$$DI= 1/2\ \Sigma_i\ |F_i/F - M_i/M|,$$

where F and M represent the total number of female and male researchers respectively (and | | indicates the positive value of the difference enclosed). Its minimum is 0, and it is achieved when there is an equal distribution of women and men across the disciplinary sectors; its maximum is 1, when in each sector only one of the two sexes is present.

The Dissimilarity Index is the most widely used measure for the analysis of occupational segregation by sex. It must be interpreted alongside the feminisation ratio, since the DI tells us whether there is an equal distribution between men and women and the FR indicates which is the dominant sex.

## 2.3 *The Segregation Curve*

To construct a Segregation (or Lorenz) Curve for gender segregation in disciplinary sectors, it is first necessary to order the sectors according to their feminisation ratios: $FR_{(1)}<FR_{(2)}<...<FR_{(n)}$; then for each ordered sector the cumulative proportions of women ($FP_{(i)}$) and men ($MP_{(i)}$) are computed:

$FP_{(1)}=F_{(1)}/F,$
$FP_{(2)}= [F_{(1)}+F_{(2)}]/F,$

.

.

.

$FP_{(n-1)}= [F_{(1)}+F_{(2)}+...+F_{(n-1)}]/F,$
$FP_{(n)}=1$
(and similarly for men);

the Segregation Curve is the plot of the cumulative proportions of women against the cumulative proportions of men. The diagonal of this graph represents zero segregation, and any deviation from the diagonal indicates the presence (and the extent) of gender segregation. There is an interesting relationship between the Segregation Curve and the Dissimilarity Index: DI represents the maximum distance between the Segregation Curve and the diagonal.

## 3. Vertical segregation

The measurement of vertical segregation investigates women's share of top positions in scientific careers, which indicates the level of female participation at the highest decision-making levels. A comparison between the percentage of women on the highest rung of the career ladder and those at the bottom is already a useful indicator of the real presence of segregation, particularly if this percentage gap is significantly different from that which applies to men.

### 3.1 *The Equally Distributed Equivalent Percentage in Science*

The Equally Distributed Equivalent Percentage (EDEP) is a derivation from the UN Gender Empowerment Measure (GEM) (UNDP, 1995), which is a general gender-equity-sensitive indicator. The methodology is used to compare disparities between the two sexes when the mean levels of achievement are different. Hence the EDEP can be used to compare situations in different countries, regardless of any differences in educational systems and career pathways. It is computed as follows:

1. Choose a number of variables representing the relative empowerment in the corresponding reference population; for example we can take the total number of female (FS) and male (MS) students as the reference population and compute the percentage of women and men in at the top of the career ladder (FT and MT), the percentage of women and men holding scientific responsibilities (FR and MR), and the percentage of women and men on scientific boards (FB and MB).

2. Choose the corresponding population weights; if FS and MS are the total number of female and male students respectively, then we can define the weights as the rate of female and male students:
   $FW = FS/(FS+MS)$ - for women
   $MW = MS/(FS+MS)$ - for men

3. For each variable defined in 1, compute the corresponding harmonic mean, which is the reciprocal of the weighted mean; for example:

the harmonic mean of the top career achievements is: $T=1/[FW/FT+MW/MT]$,
the harmonic mean of the scientific responsibility achievements is:

R=1/[FW/FR+MW/MR],
the harmonic mean of the scientific board achievements is:
B=1/[FW/FB+MW/MB]

4. The EDEP is the average of the indices computed in 3; in our example the EDEP will be:

$$EDEP = (T+R+B)/3.$$

In the most equitable case (for example, when both sexes share the same amount of top jobs and the number of overall male and female students is the same) the weights FW and MW are both equal to 1/2, while FT and MT are equal to 50%; then the index T is equal to 50%. In the least equitable case (i.e. when top jobs are allocated to one sex only) the index T is equal to zero. The overall EDEP is an average of several indices behaving in the same way as T, therefore it will vary between 0 and 50%.

More details regarding gender-equity-sensitive indicators can be found in the Human Development Report (UNDP, 1995; technical notes 1 and 2).

### 3.2. *Survival Analysis*

Another way to evaluate possible career differences between males and females is by looking at their promotion times: in the absence of any discrimination, the time of promotion up the career ladder should, on average, be the same for males and females.

Through a set of statistical methods mainly used in the medical and biological fields, known as Survival Analysis, it is possible to estimate the probability of being promoted at time $t$.

We will introduce here a few definitions which will be useful in the explanation of the method:

- let T represent the time of promotion;
- the function $S_{(t)}$=Prob(T>t) is the probability that the promotion time is greater than $t$ (that is the probability of not being promoted within time $t$) and is called the survival function;
- the hazard function $h_{(t)}$ is defined as the instant risk of being promoted at time $t$ for those who have not been promoted up to time $t$;
- the Cox regression model is a semiparametric model, also called the proportional hazard regression model, since it assumes that the promotion rates of any two individuals are proportional.

The Cox model is widely used in the analysis of medical survival data for identifying differences in survival due to treatment and prognostic factors in clinical trials.

In our case the factors will be a number of covariates, such as sex, age at promotion, geographic area, subject group of degree, number of publications, etc.

In this report, (cfr section 4.5) two phases are developed: first, the probability of remaining in the same position after 1,2,..,$t$ years is computed for males and females without any other covariate and then compared in the same graph; in the second phase this probability is related to some other factors, such as age at promotion, geographic area, subject group of degree, number of publications, etc.; again a graphical comparison through time is implemented separately for males and females.

Some statistical packages (such as SPSS) allow the user to implement the Cox model quite easily.

## 4. Another statistical tool: the Decision Tree

A classification system is a collection of decision rules that predict or classify future observations. It allows us to examine the data and discover important groupings of cases, to find key variables that identify group membership and to formulate rules for making predictions about the group membership of new cases.

The implementation of a classification system involves several steps. Once the decision rules are set by the user (for example, scientific production could be chosen as the dependent variable) one or more methods are applied to the data in order to split the data set into mutually exclusive subsets according to a set of predictive variables (for example, demographic variables, such as age and sex, or variables describing professional level, etc.).

At each successive step the decision rules are used to partition or segment the data into subgroups. The same procedure is then performed on each of the resulting subgroups.

The final outcome is a Decision Tree, a chart that graphically illustrates the results of the implementation of the decision rules, which is made up of:

- one root node that contains all the observations in the sample;
- a set of leaves containing mutually exclusive subsets of the data.

Examining the whole data set with respect to a chosen criterion requires a highly automated procedure which cannot be properly employed without a computer. Programs to implement Decision Tree algorithms are now available in most statistical packages.