



EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS 2024 ESTORIL - PORTUGAL





EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL



Moving from experimental to official statistics: increasing the scope of statistics on earnings based on new administrative data sources

Daniela Ramos, Célio Oliveira, Ricardo Cotrim
Statistics Portugal



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Overview

1. Introduction
2. From experimental to official statistics: gross monthly earnings per employee based on administrative data at enterprise level
3. One step further: experimental statistics on gross monthly earnings per employee based on administrative data at worker level
4. Conclusions and future developments



eurostat 

The conference is partly
financed by the European Union



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Overview

1. Introduction

2. From experimental to official statistics: gross monthly earnings per employee based on administrative data at enterprise level
3. One step further: experimental statistics on gross monthly earnings per employee based on administrative data at worker level
4. Conclusions and future developments



1. Introduction

- Society demands for more information
- Respondents complain they are overloaded and asked the same information repeatedly
 - ➔ **Administrative data analysis becomes a priority for the statistical systems**
- Replacing traditional 10-yearly Census with 1-year continuous updated Census based on administrative data



Development of a **Resident Population Database** integrated into a **National Data Infrastructure**



Extension of the domains covered + Frees up resources + Greater innovation



1. Introduction

- **Protocols** for statistical purposes with entities that hold **administrative data on gross earnings**
 - SS - Social Security
 - CGA - *Caixa Geral de Aposentações* (Portuguese civil servants' retirement and survivor pensions fund)
 - AT - Tax Authority
- Data at **two levels**:
 1. Enterprise
 2. Employee
- **Dedicated area** at Statistics Portugal website 
 - Projects still under development
 - But provide useful information for economic and social analysis



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Overview

1. Introduction

2. From experimental to official statistics: gross monthly earnings per employee based on administrative data at enterprise level

3. One step further: experimental statistics on gross monthly earnings per employee based on administrative data at worker level

4. Conclusions and future developments



eurostat 

The conference is partly
financed by the European Union



2. From experimental to official statistics: administrative data at enterprise level

- **2018:** Monthly Statement of Earnings from Social Security (DMR-SS)
- **2019:** Contributory relation from CGA

Release of new statistical product

Average Gross Earnings per
Employee (DMR-SS)

May
2019



Nov.
2019

**1st Integration of SS and
CGA data on earning**
(DMR-SS & CGA)

Official Statistics (DMR-SS & CGA)

Nov.
2021

- Quarterly release
- Monthly (3-mma) and yearly data
- Data since 2014
- Nominal and real values
- Several disaggregations



2. From experimental to official statistics: administrative data at enterprise level

CGA

- Monthly data at enterprise level, definitive once received

DMR-SS

- Monthly data at enterprise level, can be continuously updated
 - Since each month is received 4 times, it is assumed to be final by the 4th update
- Nonetheless, last 3 months are provisional in their first dissemination and are revised in the following release
- To minimize revision, these data are imputed for:
 1. Companies systematically delay sending data
 2. Companies regularly make substantial corrections to values reported in previous months
- The detection of these companies is ensured by a combination of two methods:
 1. Ad hoc criteria
 2. Supervised machine learning algorithm in the Support Vector Machine (SVM) version



2. From experimental to official statistics: administrative data at enterprise level

DMR-SS
CGA



Databases linked by using the employers' tax identification number (NIF) or the legal person identification number (NIPC)



The **final database** covers almost the **entire universe of employees**

Those with more than one job **are counted as many times** as the number of jobs they have

$$RBMMT_{Q,c} = \frac{\sum_i^k \text{Earning paid by the company } i_{Q,c}}{\sum_i^k \text{Number of the company's employees } i_Q}$$

where:

Q: moving quarter ended in month **M**, covering months $\{M-2, M-1, M\}$

c: earning component $\{Total, Regular, Base\}$

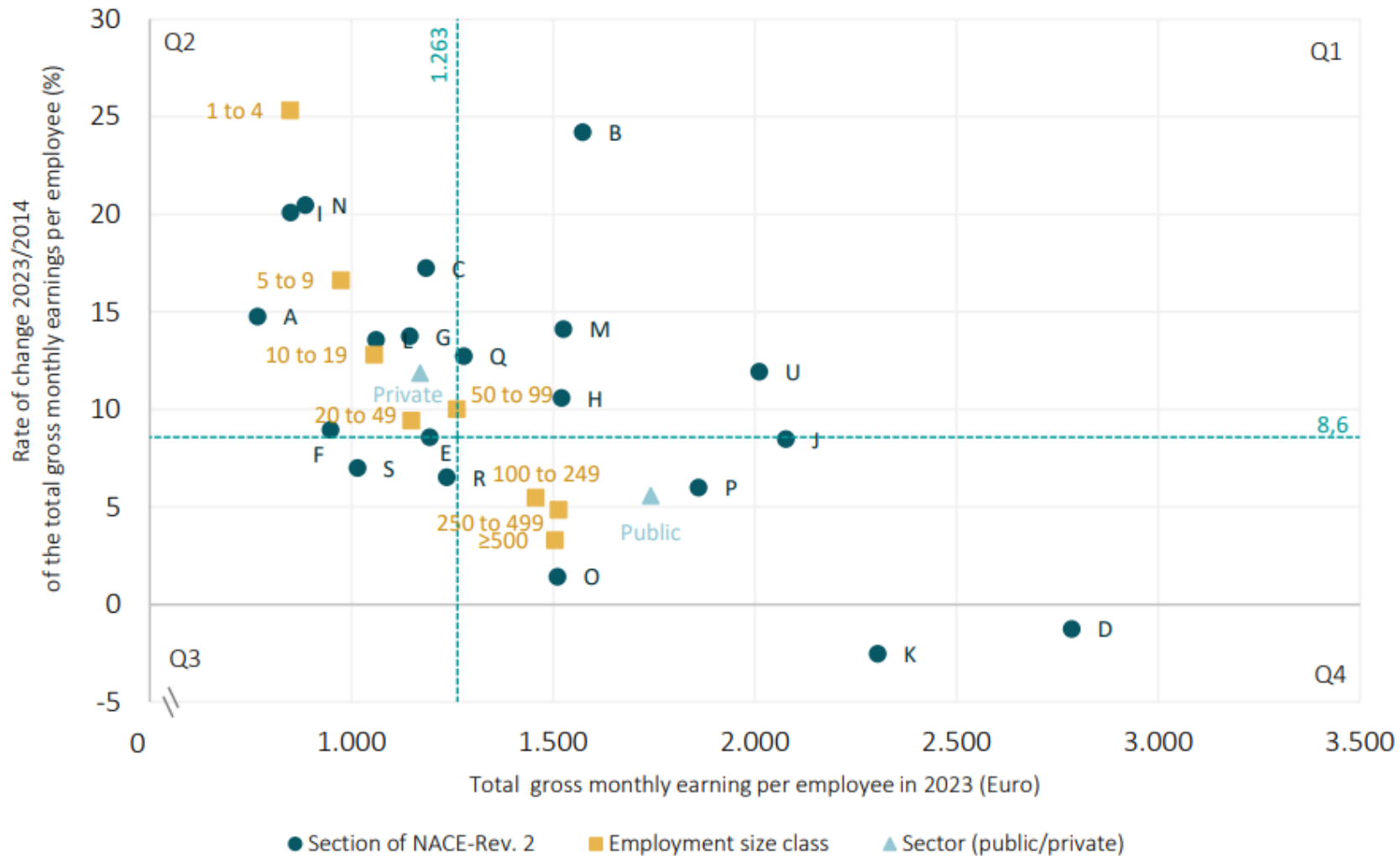
i: company number

k: total number of companies

$\sum_i^k \text{Earning paid by the company } i_{Q,c}$: volume of earnings paid in the period **Q** in the component **c**



2. From experimental to official statistics: administrative data at enterprise level



Example of annual analysis

Total gross monthly earnings per employee in 2023 and rate of change since 2014, in real terms, by economic activity (NACE-Rev. 2), employment size class and the institutional sector (public and private)



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Overview

1. Introduction
2. From experimental to official statistics: gross monthly earnings per employee based on administrative data at enterprise level
- 3. One step further: experimental statistics on gross monthly earnings per employee based on administrative data at worker level**
4. Conclusions and future developments



eurostat 

The conference is partly
financed by the European Union



3. One step further: ... administrative data at worker level

Companies monthly tax and social security obligation



	Social Security ^(a)		Tax Authority	
Desaggregation level	Enterprise		Employee	
Observation unit	Job ^(b)		Employee	Job ^(b)
<i>Annual average in 2021</i>	<i>4.2 millions</i>		<i>4.6 millions</i>	<i>5.5 millions</i>
Social protection schemes covered	Social Security ^(a)		All schemes of social protection ^(c)	
Type of income/earning	Only those subject to tax/contribution		All income or earnings	

(a) This includes data from CGA.

(b) Corresponds to the unique combination of employee and company.

(c) This includes the various pension funds, the Social Security and the CGA, for example.



3. One step further: ... administrative data at worker level

Integration of DMR-AT with National Data Infrastructure



Socioeconomic variables

Coverage rates (%) of the dimensions characterising individuals and enterprises

Portugal	2019	2020	2021
Total	100.0	100.0	100.0
Sex	95.8	96.5	94.7
Age group	95.8	96.5	94.7
Educational level	86.1	83.8	77.8
Region of residency (NUTS 2)	96.9	96.4	89.9
Economic activity (NACE)	99.7	99.9	99.9
Occupation (ISCO)	75.0	73.6	63.1



3. One step further: ... administrative data at worker level

In 2023, two Press Releases

- **Yearly data from 2019 to 2021**
 - Distribution moments of earnings (mean, median, deciles and percentiles)
 - Inequality indicators (Gini coefficient, S80/S20 and P50/P10)
 - Excel file with dynamic graphs and tables by sex, age group, level of education, NUTS 2 region of residency, economic activity and occupation
- **Year of 2021**
 - Analysis focused on comparing the distribution moments of earnings between institutional sector (public and private) by sex, age groups and level of education

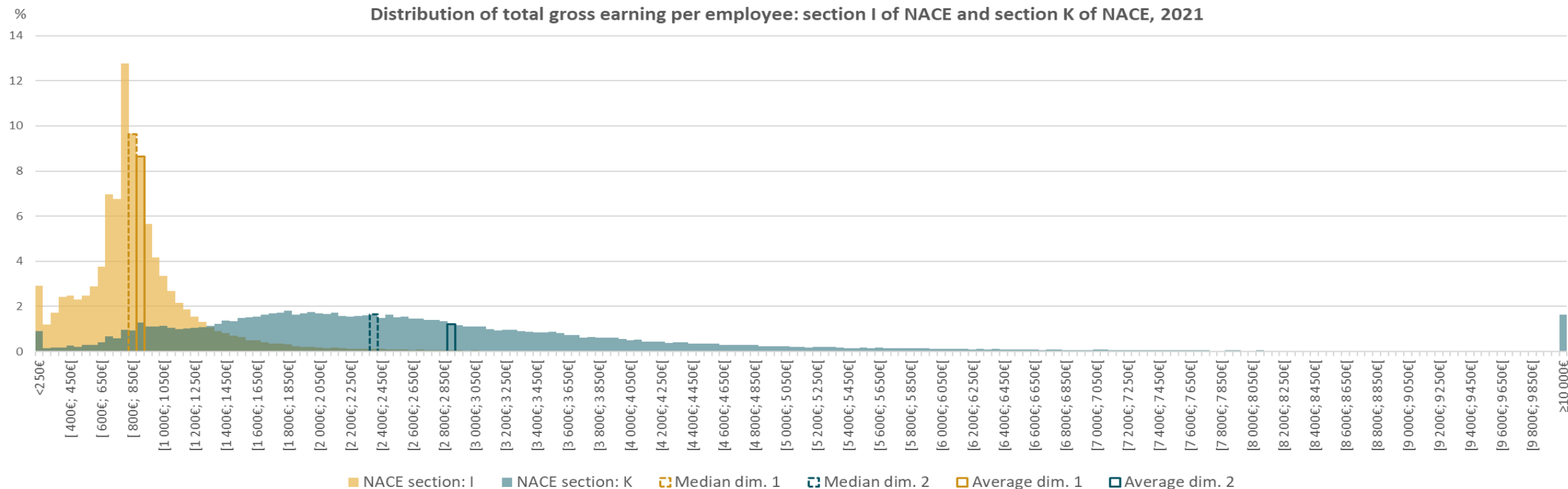


3. One step further: ... administrative data at worker level

Example of the dynamic Excel file from the 1st STATS lab Press Release

Year	Dimension 1	Designation
2021	Section NACE: I	Accommodation and food service activities
Year	Dimension 2	Designation
2021	Section NCAE: K	Financial and insurance activities

Distribution of total gross earning per employee: section I of NACE and section K of NACE, 2021

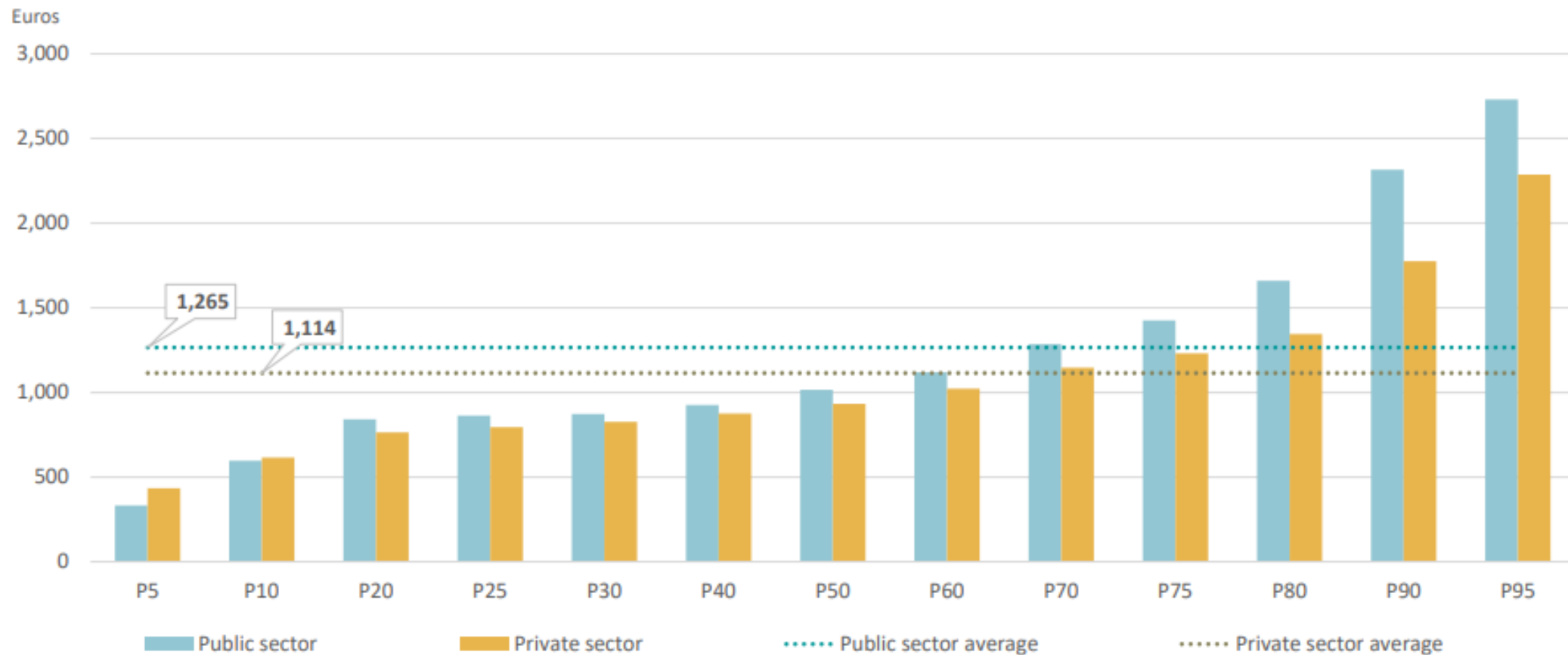




3. One step further: ... administrative data at worker level

Example of the 2nd STATS Press Release

Distribution of gross monthly per employee who has completed, at most, the lower secondary education by institutional sector





EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL

Overview

1. Introduction
2. From experimental to official statistics: gross monthly earnings per employee based on administrative data at enterprise level
3. One step further: experimental statistics on gross monthly earnings per employee based on administrative data at worker level

4. Conclusions and future developments



4. Conclusion and future developments

The **use of administrative data** makes it possible to:

- Increase the frequency of information
- Reduce the response burden, as well as costs
- Produce new statistical products

The **integration** of administrative data on earnings **with the National Data Infrastructure** still **requires further development** to move from experimental to official statistics regarding data at employee level.

However, **future projects** are:

- Take advantage of data from the DMR-SS and CGA at the employee level to explore sociodemographic characterization of workers
- Updating the procedure to deal with non-responses to further minimize revisions to DMR-SS data
- Regular analysis of earnings distribution and inequality indicators
- Study of movements into and out of the labour market and associated wage changes



EUROPEAN CONFERENCE ON
QUALITY IN OFFICIAL STATISTICS
2024 ESTORIL - PORTUGAL



Obrigada!

Session 9: Experimental Statistics
Lisboa, 5 July 2024

Daniela Ramos, daniela.ramos@ine.pt

Célio Oliveira, celio.oliveira@ine.pt

Ricardo Cotrim, ricardo.cotrim@ine.pt



EUROPEAN CONFERENCE ON QUALITY IN OFFICIAL STATISTICS 2024 ESTORIL - PORTUGAL