

The Panel Employment, Wage and Working Hours for 2022 and 2023
Data Description, Sampling design and survey weights

Version 1.0
December, 2025
FOR PUBLIC RELEASE

Summary

This report describes the description for the data files with its unique identifiers, in addition to, the sampling design and weighting methodology used for the Employment, Wage and Working Hours panel for 2022 and 2023. The application of these weights in all analyses is essential to ensure that the sample accurately represents the characteristics of the target population within the country. Although weighting helps correct for observed imbalances, it cannot fully eliminate potential biases stemming from non-response on unobserved factors.

Data files

All data files are available as SPSS & STATA. dta files.

The Panel Employment, Wage and Working Hours for 2022 and 2023 data consist of 2 files as follows:

1) Establishment file

- This file includes information on the establishment's characteristics, such as the total number of employees and geographic details.
- Additionally, this file specifies the reason indicating whether the establishment interviewed in 2022 was successfully revisited in 2023 or not.
- **ser sector:** unique establishment identifier variables. These variables are repeated in all files if merge is required, where "ser" is the establishment number within the "sector" variable (private and public sectors).
- The data are constructed in **wide format for years 2022 & 2023**, directly compiled from the different cross-sectional tables.
- Each observation represents a single establishment, with variables reported across specific reference period.
- It must be noted that merging with the other files are not always one to one merge.

2) Left, hired, and expected occupation file

- This file captures worker flows (left and hired) as well as number of expected occupations reported in 2022 only, 2023 only, or in both years.
- **ser sector job_4:** unique occupation identifier variables within each establishment. The job_4 represents the 4th level of occupation code for the specified job.
- The data are constructed in **long format**, allowing multiple occupation records per establishment.

Sample

The sampling frame is constructed using explicit stratification by governorate and establishment size class. Stratification by economic activity is applied implicitly by sorting the list of establishments within each explicit stratum according to their 4-digit activity code, in order to avoid excessively large number of strata.

All public sector enterprises are included in the sample (100 percent). For enterprises operating multiple establishments ("branches"), up to five branches in addition to the headquarters (HQ) are selected. This approach ensures that the sampling unit reflects public sector establishments rather than enterprises, which may consist of several establishments carrying out different economic activities.

For the private sector, the sampling design is based on stratification by governorate and five establishment size classes defined by number of employees. Sampling rates vary by size class, with smaller establishments selected at lower rates and larger establishments at higher rates.

Establishments that provided completed responses in 2022 were followed up in 2023 and are included in the panel Employment, Wage and Working Hours Data for 2022 & 2023.

Weights

To reduce bias in a number of observable dimensions, inverse probability weighting was applied. The weights were constructed separately for public and private establishments.

The original establishment weight variable (w_i) was retained and adjusted to account for attrition, defined as whether an establishment was successfully revisited in 2023. A probit model incorporating survey weight for 2022 was used to estimate the probability that an establishment with given characteristics would be interviewed in the 2023 survey. The covariates included in the prediction model were:

- the economic activity of the establishment,
- the size of establishment,
- if the establishment is a headquarter or branch,
- the region in which the establishment is located.

Establishments that were permanently closed in 2023, demolished, or otherwise fell outside the sampling frame were excluded from the analysis and were not considered in the model.

The predicted probability from that model, p_i , was used to generate establishment weights for the Employment, Wage and Working Hours panel for 2022 and 2023:

$$w'_i = w_i * (1 / p_i)$$

General Notes:

- Further details on the weighting methodology applied in 2022 are available in the documentation attached to the ERF Data Portal.
- The worker-level file for the Employment, Wage and Working Hours survey does not constitute a panel for 2022 and 2023, as workers are newly sampled each year. Consequently, identical worker serial numbers appearing across years do not necessarily refer to the same individual.

- **Left, hired, and expected occupation file panel data** were constructed after aggregating occupations to the fourth-digit level separately for each year and then merging the datasets.