

INCGROSS variable in Serbian LFS

Milijana Smiljkovic

16th workshop on Labour Force Survey methodology

Lisbon, 25 May 2023

INGROSS variable in Serbian LFS



Content:

- Data collection
- Imputation methods in case of item non-response
- Net-to-gross conversion
- Preliminary results - testing the method for detection of outliers

What is collected in the LFS questionnaire?

INCOME IS COLLECTED IN NET AMOUNT.

Respondents found it easier to provide precise answers when reporting net values rather than gross values.

Concept and definition used do not differ from the description in the regulation.

First question- clear amount

How much was your NET income in last month?

Includes: actual net monthly payments, regular overtime, shift work, seniority bonuses, tips and commission and compensation for meals in cash, payments made on a higher than monthly periodicity (e.g. yearly or quarterly payments such as 13th month or holiday pay) proportionally included.

If answered

END

If not

Second question – income bands

Can you still indicate in which range is your net income in last month?

END

less than 30 000 RSD

30 001–45 000

45 001–60 000

60 001–80 000

80 001–100 000

100 001–150 000

150 001–200 000

200 001–300 000

300 001 RSD and more

What is collected in the LFS questionnaire?



Response rate 2021:

- ✓ direct question on exact amount of salary (INCSUM) = 26.6%,
- ✓ for the second question, income bands, (INCDECIL) = 27.3%.

Overall response rate to these (two) questions is very low, **below 60%**.



imputation from the administrative sources is necessary.

According to the new Regulation imputation is needed when non-response is more than 5%.

Imputation methods in case of item non-response



- A. Imputation of gross monthly pay is conducted using the Earnings Register, which is established based on income data from the Tax Authority (TA).

Approximately 54% of missing earnings data in the LFS, which corresponds to around 25% of all employees, are imputed using gross earnings data obtained from the Tax Authority.

- B. If it not feasible to impute the value from the Tax records, such as in cases of informal employment, an average value of earnings is imputed based on the occupation and activity at the two-digit level. This is done by considering data from both the tax records and the LFS, following specified criteria.

This method allows the imputation of 44% of missing data for the variable INCGROSS, which represents approximately 20% of all employees who need to respond to the salary question.

Net-to-gross conversion



The national taxation system allows us to convert net to gross value quite easy, because the corresponding taxes and contributions can be calculated based on the type of employment contract.

In the Serbian LFS starting from 2021, a new national variable called TEMPUG was introduced to assist in the conversion of net to gross earnings.

By combining the responses from four questions: STAPRO, TEMP, PREDRAD (also national variable), and TEMPUG, it is possible to ***distinguish between 4 types of (formal) employment contracts*** that are taxed differently according to our taxation legislation:

1. Employees having employment contract with definitive/indefinite duration (3 sub-groups with different taxation depending on the amount of income)
2. Temporary and occasional jobs
3. Service contract
4. Compensation for a person undergoing vocational training and professional specialization
and
5. Informal employees (net earnings = gross earnings)

Preliminary results - testing the method for detection of outliers



For outlier detection for INCGROSS is used the formula for SPSS extreme outlier, also given and recommended to be tested on INCGROSS data in the document *LFS_CONVAl*.

Outlier detection for INCGROSS:

Q1 and Q3 are the 1st and 3rd quartiles

Lower boundary: $\text{EXP} [\log Q1 - 3 * (\log Q3 - \log Q1)]$

Upper boundary: $\text{EXP} [\log Q3 + 3 * (\log Q3 - \log Q1)]$

Data on INCGROSS variable shown in this presentation are related to the 2021 and still be consider preliminary because are not yet published as well as because it not yet decided how outliers will be treated, and whether they should be treated at all?

After the implementation of this formula for detection of outliers the following results are obtained:

- ✓ The lower boundary obtained from the formula resulted in a negative value, which is not meaningful since income cannot be negative.
- ✓ On the other hand, the upper boundary amounted to 210,824 RSD, which corresponds to the 98th percentile of the National Monthly Survey of Earnings based on administrative (Tax) data.

Preliminary results - testing the method for detection of outliers

Table 1. Average earnings in national currency from Earnings statistics and LFS, 2021.

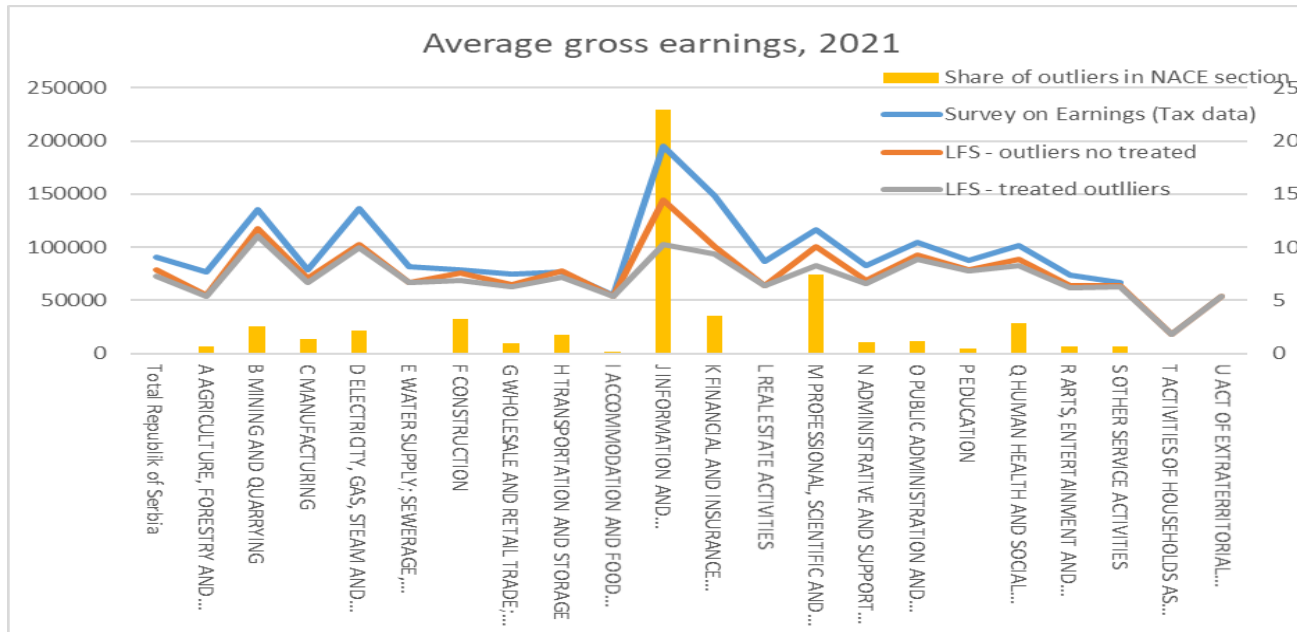
Average gross salary (Survey on Earnings)	90 784
INCGROSS (LFS - no treated outliers)	78 441
INCGROSS (LFS - treated outliers)	72 572

The preliminary results on INCGROSS variable from LFS in comparison with Earnings Survey show that data on gross monthly pay from the main job from LFS (both, treated and no treated outliers) is lower than the data from administrative data on Earnings statistics. The results are expected having in mind difference in the coverage (LFS cover informal employment) and the tendency of respondents to underestimate their earnings when reporting them in surveys.

Comparing the average earnings from the Earnings statistics and the LFS, it was found that the average gross salary from the Earnings statistics is 15% higher than the LFS data without treating outliers. When the outliers were treated, the difference increased to 25%.

Preliminary results - testing the method for detection of outliers

Graph 1. Comparison of average gross earnings from Survey on Earnings, LFS – with treated outliers and LFS – without treated outliers and share of detected outliers, by section of activity (NACE), 2021.



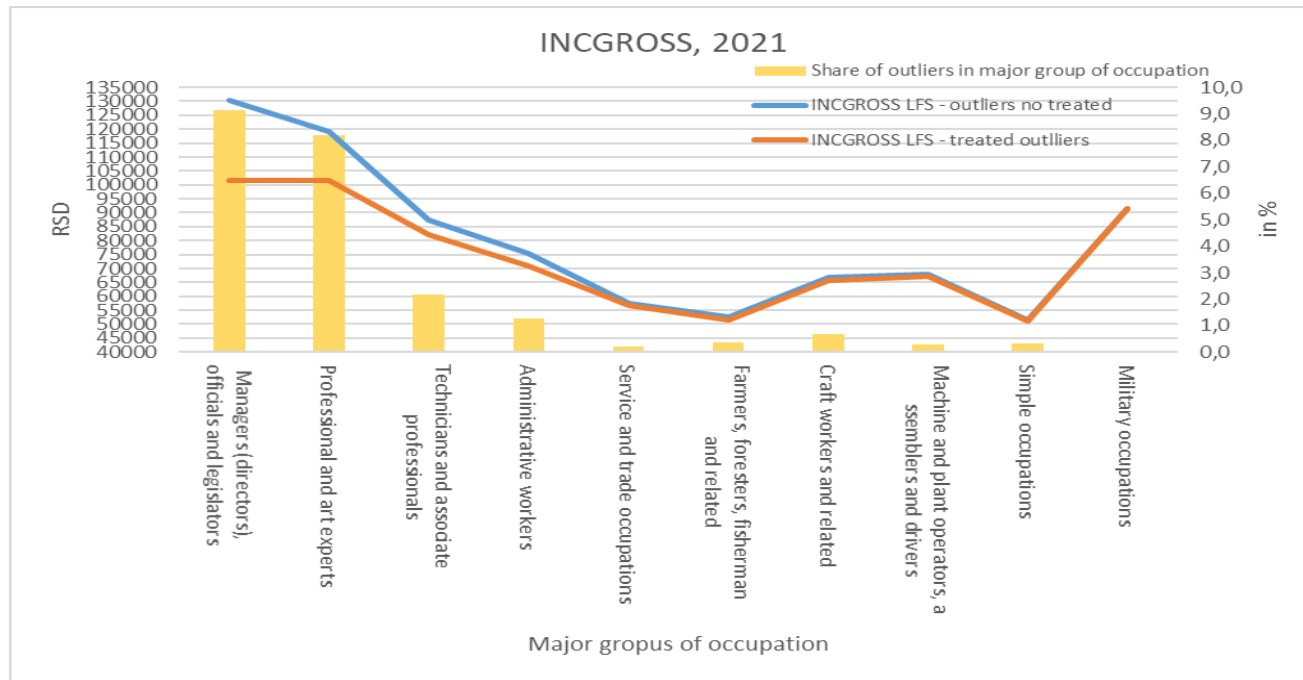
- The share of outliers in sector J amounts 22.9%, which means that almost every fourth employee in this sector should be excluded because they receive wages that are considered as outlier.
- In sector, M 7.4% of employees should be excluded in the calculation of INCGROSS variable if we apply mentioned criteria for detection of outliers.

Before and after the detection and removal of outliers:

- In three NACE sectors (E, T and U), there was no difference in the INCGROSS variable.
- In majority of sectors (A, B, C, D, G, F, H, I, N, O, P, Q, R, S) the difference between the INCGROSS variable with treated and without treated outliers ranges from 1 to 11%.
- The greatest difference in INCGROSS variable was noted in the sector J Information and communication, and amounted 40%, the second sector with significant difference was M Professional, scientific and technical activities, 22%.

Preliminary results - testing the method for detection of outliers

Graph 2. INCGROSS variable (with treated outliers and without treated outliers) and the share of outliers, by major groups of occupation, for 2021



- The greatest share of employees detected as outliers by groups of occupation is in first two groups, 9.1% and 8.9, respectively.
- Total number of outliers is 50,000, which is 2.3% of total employees. Of that total number of outliers (50,000) 32 000 i.e. 64% outliers belongs to the group of occupation Professional and art experts which is the main reason for doubting the justification of the used criterion for the detection of outliers.

The major share of outliers are in the first two groups of occupation which naturally have the largest earnings.

- The greatest difference in INCGROSS variable before and after the detection and removal of outliers is in the first group (Managers/directors, officials and legislators) and amounts 28%, while in the second group (Professional and art experts) amounts 17%.
- In next two groups (Technicians and associate professionals and Administrative workers) the difference is 6%,
- In other groups the difference is almost negligible, 1-2%.
- In Military occupation there is no difference.

Preliminary results - testing the method for detection of outliers



Table 2. Number of employees where INCGROSS > detected outlier (210 824 RSD)

	Count num. of employees	Share (%)
INCDECIL - interval response	5206	10.4
INCSUM - direct response	5354	10.7
Imputed net income for item non-response from the average (NACE and occupation at 2- digit level)	10198	20.4
Imputed gross income for item non-response from administrative sources (INCGROSS_F 25)	29340	58.6
Total	50098	100.0

- The majority of outliers come from the data that are imputed from administrative sources where the outliers are already treated in some way.
- In case of interval response the outlier are treated by defining the lower and upper limits of the bands.
- The other type of imputation is also done by using the averages.
- Only 10.7% of outliers come from direct answer.

Instead of conclusion questions??? 😊

Overall, the preliminary results highlight the challenges and dilemmas associated with outlier detection. It is important to consider the limitations of the methods used and further examine the appropriateness of outlier detection criteria.

Are you going to treat outliers and how?

Do the Eurostat going to apply the same rule for everyone?