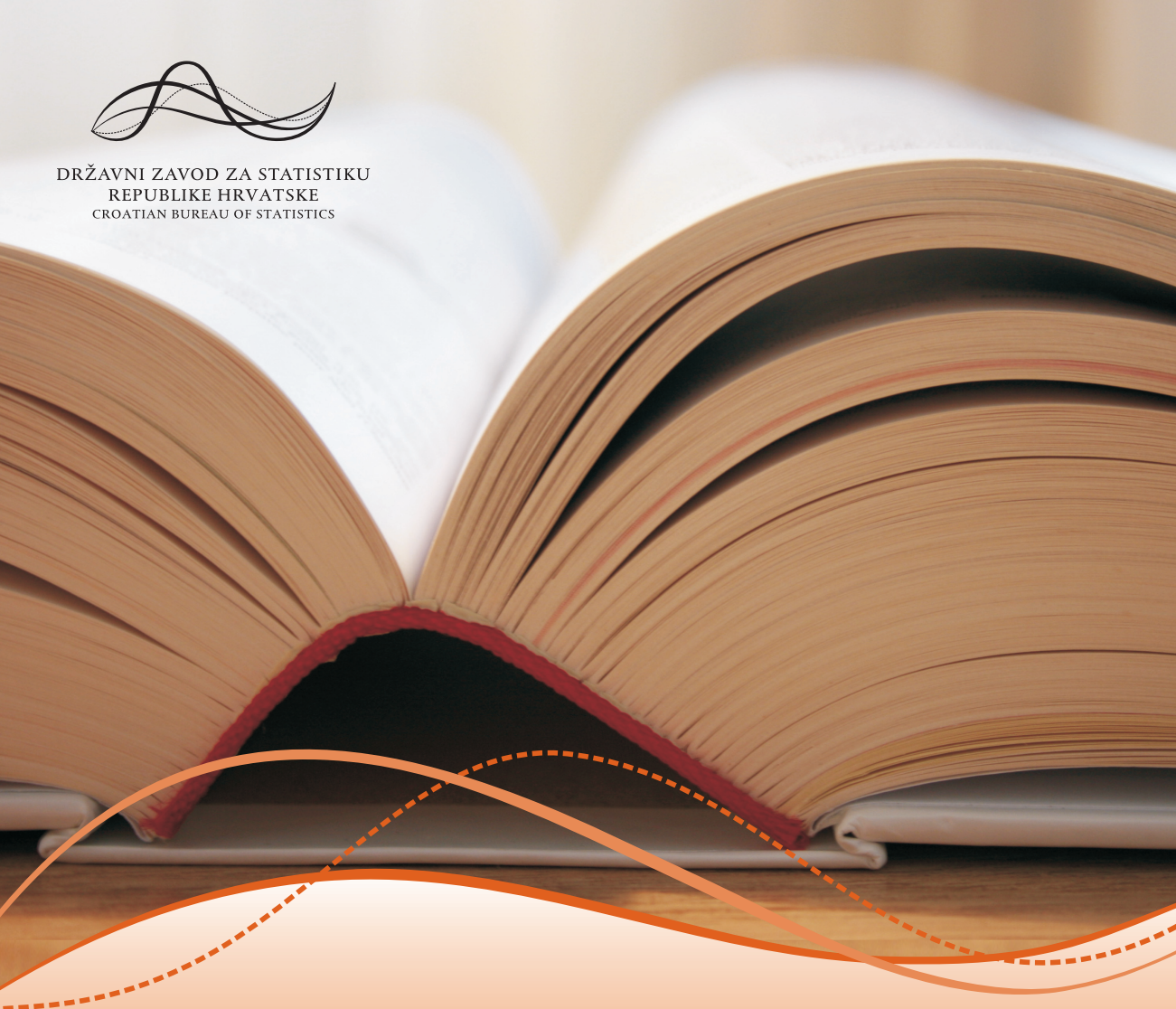




DRŽAVNI ZAVOD ZA STATISTIKU  
REPUBLIKE HRVATSKE  
CROATIAN BUREAU OF STATISTICS



## Priručnik za izračun pokazatelja kvalitete

### *Handbook for Calculation of Quality Indicators*

Metodološke upute 73  
*Methodological Guidelines 73*

Zagreb, 2015.



DRŽAVNI ZAVOD ZA STATISTIKU  
REPUBLIKE HRVATSKE  
CROATIAN BUREAU OF STATISTICS

## **Priručnik za izračun pokazatelja kvalitete**

### ***Handbook for Calculation of Quality Indicators***

Metodološke upute 73

*Methodological Guidelines 73*

Zagreb, 2015.

Objavljuje i tiska Državni zavod za statistiku Republike Hrvatske, Zagreb, Ilica 3, p. p. 80.  
*Published and printed by the Croatian Bureau of Statistics, Zagreb, Ilica 3, P. O. B. 80*

Telefon/ Phone: +385 (0) 1 4806-111  
Telefaks/ Fax: +385 (0) 1 4817-666

Elektronička pošta/ E-mail: ured@dzs.hr  
Internetske stranice/ Web site: <http://www.dzs.hr>

Odgovara ravnatelj Marko Krištof.  
*Person responsible: Marko Krištof, Director General*

Priradili: Članovi radne skupine Tehnička pomoć u poslovnim statistikama i izrada dokumentacije  
*Prepared by:* o kvaliteti podataka: Mario Gavrić i Dario Legčević

Redaktor: Žaklina Čizmović  
*Sub-editor:*

Urednica: Ljiljana Ostroški  
*Editor-in-Chief:*

Lektorica za hrvatski jezik: Anđa Matić  
*Language Editors for the Croatian Language:*

Prevoditelj i lektor za engleski jezik: Thomas Peter  
*Translator and Language Editor for the English Language:*

Tehnička urednica: Ankica Bajzek  
*Technical Editor:*

ISBN 978-953-273-053-1

**MOLIMO KORISNIKE DA PRI KORIŠTENJU PODATAKA NAVEDU IZVOR.**  
*USERS ARE KINDLY REQUESTED TO STATE THE SOURCE*

Elektroničko izdanje  
*Electronic edition*

Služba za komunikaciju s korisnicima  
*User Communication Service*

• Informacije i korisnički zahtjevi  
*Information and user requests*

Telefon/ Phone:  
+385 (0) 1 4806-138, 4806-154, 4811-212  
Elektronička pošta/ E-mail:  
[stat.info@dzs.hr](mailto:stat.info@dzs.hr)  
Telefaks/ Fax:  
+385 (0) 1 4806-148, 4806-199

• Pretplata publikacija  
*Subscription*

Telefon/ Phone:  
+385 (0) 1 4814-791  
Elektronička pošta/ E-mail:  
[prodaja@dzs.hr](mailto:prodaja@dzs.hr)  
Telefaks/ Fax:  
+385 (0) 1 4806-148, 4806-199

• Novinarski upiti  
*Press corner*

Telefon/ Phone:  
+385 (0) 1 4806-121, 4806-196  
Elektronička pošta/ E-mail:  
[press@dzs.hr](mailto:press@dzs.hr)  
Telefaks/ Fax:  
+385 (0) 1 4806-148, 4806-199

# SADRŽAJ

## CONTENTS

PREDGOVOR <i>PREFACE</i> .....	5
KRATICE <i>ABBREVIATIONS</i> .....	9
UVOD <i>INTRODUCTION</i> .....	11
1. KLJUČNI POKAZATELJI <i>KEY INDICATORS</i> .....	14
1.1. TOČNOST <i>ACCURACY</i> .....	14
1.1.1. Uzoračka greška (A1) <i>Sampling error (A1)</i> .....	15
1.1.2. Stopa nadobuhvata (A2) <i>Over-coverage rate (A2)</i> .....	20
1.1.3. Stopa neodgovora jedinica (A4) <i>Unit non-response rate (A4)</i> .....	23
1.1.4. Stopa neodgovora na određenu varijablu (A5) <i>Item non-response rate (A5)</i> .....	25
1.1.5. Stopa imputacije (A7) <i>Imputation rate (A7)</i> .....	31
1.1.6. Prosječna veličina revizije podataka (A6) <i>Data revision-average size (A6)</i> .....	33
1.2. PRAVODOBNOST I VREMENSKA ODREĐENOST <i>TIMELINESS AND PUNCTUALITY</i> .....	36
1.2.1. Pravodobnost prvih rezultata (TP1) <i>Timeliness of first results (TP1)</i> .....	37
1.2.2. Pravodobnost konačnih rezultata (TP2) <i>Timeliness of final results (TP2)</i> .....	39
1.2.3. Vremenska određenost (TP3) <i>Punctuality (TP3)</i> .....	41
1.3. USPOREDIVOST <i>COMPARABILITY</i> .....	42
1.3.1. Dužina usporedivih vremenskih serija (CC2) <i>Length of comparable time series (CC2)</i> .....	43

2. POMOĆNI POKAZATELJI <i>AUXILIARY INDICATORS</i> .....	45
2.1. RELEVANTNOST <i>RELEVANCE</i> .....	45
2.1.1. Stopa potpunosti podataka (R1) <i>Data completeness rate (R1)</i> .....	46
2.2. TOČNOST <i>ACCURACY</i> .....	47
2.2.1. Pristranost u procesu selekcije (A8) <i>Bias due to selection process (A8)</i> .....	47
2.2.2. Stopa uređivanja podataka (A9) <i>Editing rate (A9)</i> .....	49
2.2.3. Stopa jedinica koje nisu prošle LRK (A10) <i>Edit failure rate (A10)</i> .....	50
2.2.4. Stopa učinkovitosti LRK (A11) <i>Hit rate (A11)</i> .....	52
2.2.5. Stopa nepravilnog razvrstavanja (A12) <i>Misclassification rate (A12)</i> .....	53
2.3. SMISLENOST <i>COHERENCE</i> .....	56
2.3.1. Smislenost između različitih izvora podataka (CH1) <i>Coherence between different sources (CH1)</i> .....	56
LITERATURA <i>LITERATURE</i> .....	62

## PREDGOVOR

Važna obveza Državnog zavoda za statistiku (u nastavku teksta: DZS) kao glavnog nositelja, diseminatora i koordinatora sustava službene statistike Republike Hrvatske je osigurati visoku kvalitetu i pouzdanost statističkih rezultata. Službena statistika pruža, na nepristranoj osnovi, državi, gospodarstvu i javnosti podatke o gospodarskome, demografskome, socijalnome i ekološkom stanju, djelatnostima ili događajima koji se mogu mjeriti statističkim metodama te osigurava ispunjavanje međunarodnih obveza Republike Hrvatske koje se odnose na razmjenu statističkih podataka s drugim državama i međunarodnim organizacijama.

DZS shvaća politiku kvalitete kao korporativnu kulturu s pet temeljnih stupova kvalitete (glavni čimbenici), međusobno povezanih s modernim alatima za upravljanje. Ti su stupovi:

- *Nezavisnost nacionalne statistike.* Trenutačno uređenje DZS-a osigurava visoku razinu profesionalne neovisnosti u proizvodnji nacionalnih i regionalnih statistika. Ta pozicija bi se svakako trebala održati i ojačati jer se samo profesionalne i politički neovisne statistike smatraju pouzdanima i kao takve važne korisnicima.
- *Korisnici podataka i pružatelji podataka.* Postaje sve teže ostvariti ravnotežu između korisnika statističkih podataka i informacija te zahtjeva prema kućanstvima i poslovnim subjektima za pružanjem podataka potrebnih za proizvodnju tih informacija. Stoga je iznimno važno pratiti u kojoj su mjeri objavljene statistike ispunile očekivanja i potrebe korisnika, te u isto vrijeme pratiti opterećenost ispitanika s obzirom na njihovu obvezu davanja podataka za statističke svrhe. Smanjenje opterećenja ispitanika te osiguranje povjerljivosti i zaštite dostavljenih podataka (koji se moraju koristiti isključivo za statističke svrhe) temeljne su zadaće nacionalne statistike.

## PREFACE

*The important commitment of the Croatian Bureau of Statistics (hereinafter: CBS), as a central disseminator and coordinator of the official statistics of the Republic of Croatia, is to provide high quality and reliability of its statistical results. Produced on an impartial basis, the official statistics provide the country, the economy and the public with the data on economic, demographic, social and environmental state, activities or events that can be measured by statistical methods and fulfils the international obligations of the Republic of Croatia relating to the exchange of statistical data with other countries and international organisations.*

*CBS perceives its quality policy as part of a corporate culture with five basic quality pillars (main factors) interconnected with the modern management tools in use in the CBS. These pillars are:*

- *The independence of national statistics.* CBS currently provides a high level of professional independence in its production of national and regional statistics. This position should be certainly maintained and strengthened, as only a professional and politically independent statistics are seen as reliable and relevant to the customers.
- *Data users and data providers.* It becomes more difficult to achieve a balance between the users of statistical data and information and requests generally to households and businesses to provide the data necessary to produce this information. Therefore, it is extremely important to monitor the extent to which the published statistics meet the expectations and needs of users, at the same time and to monitor the burden on the respondents with respect to their obligation to provide data for statistical purposes. Reducing the respondent burden and assuring the confidentiality and protection of submitted data (which must be used only for statistical purposes) are fundamental tasks of national statistics.

- *Kvaliteta statističkih proizvoda i usluga.* Radi pružanja kvalitetnih statističkih proizvoda i usluga DZS bi trebao djelovati u skladu sa standardnim definicijama kvalitete i načelima Kodeksa prakse Europske statistike. Nadalje, potrebno je kontinuirano razvijati sustav za diseminaciju informacija o kvaliteti podataka svojim korisnicima te se upravo u sklopu nacionalne IPA-e postigao veliki napredak u proizvodnji izvještaja o kvaliteti provedenih istraživanja standardizaciji istih za sva istraživanja koja provodi DZS.
- *Orijentiranost na procese.* U procesima pripreme statističkih podataka važnu ulogu ima pružanje informacija o izvorima pojedinih statističkih istraživanja, primijenjenim metodologijama, procedurama i troškovima vezanima za statistička istraživanja. Dakle, bolja kvaliteta rezultata i bolja troškovna učinkovitost istraživanja može se ostvariti isključivo s transparentnim statističkim procesima te jasno dokumentiranim postupcima.
- *Razvoj ljudskih resursa.* Obuka zaposlenika DZS-a radi povećanja razine kvalitete statističkih proizvoda i usluga uključuje nekoliko aspekata: metodološko znanje, *know-how* i promicanje razmjene dobre prakse. Važno je da su zaposlenici u sustavu nacionalne statistike svjesni sadržaja Kodeksa prakse Europske statistike i da svakodnevno rade u skladu s njim.
- *The quality of statistical products and services.* In order to provide good quality in its statistical products and services, CBS should operate in accordance with the standard definitions and principles of quality in the European Statistics Code of Practice. Further, the system for disseminating information about the quality of statistics to their users should continue to be developed. As part of the national IPA project, much progress has been achieved within the component on the production of survey reports relating to their quality, and standardisation of these quality reports across the statistical survey work of the CBS.
- *Process orientation.* In the statistical data preparation processes, providing information on individual statistical sources, methodologies used, procedures and costs of statistical surveys plays an important role. With transparent statistical processes, and clearly documented procedures, higher quality of results and better cost-efficiency of surveys can be obtained.
- *Human resources development.* Training of CBS employees in order to increase the quality of statistical products and services include several aspects: methodological knowledge, know-how and exchange of good practices promotion. In the national statistical system the employee, awareness on European Statistics Code of Practice, content and everyday use of it is very important.

U skladu s Kodeksom prakse europske statistike, cilj DZS-a je jačanje povjerenja u neovisnost i cjelovitost statistika koje se proizvode i diseminiraju. DZS tako teži primjeni najboljih međunarodnih statističkih načela, metoda i prakse kako bi se osigurala proizvodnja i diseminacija reprezentativnih te međunarodno usporedivih statističkih podataka u skladu s preporukama iz Kodeksa. Stoga je od presudne važnosti praćenje kvalitete podataka te kontrola kvalitete svih statističkih procesa koje DZS koristi u proizvodnji svojih podataka.

*In accordance with the European Statistics Code of Practice, the goal of CBS is to strengthen the confidence in the independence and integrity, of the produced and disseminated statistics. So, CBS seek to implement the best international statistical principles, methods and practices to ensure the production and dissemination of representative and internationally comparable statistical data, required within these recommendations. Therefore it is of crucial importance to monitor the data quality and quality control, of all statistical processes that CBS is using in the production of its statistics.*

Nadalje, kvaliteta podataka je više-dimenzionalni koncept koji se ne odnosi samo na statističku preciznost podataka nego i na usporedivost, relevantnost, pravodobnost i vremensku određenost, dostupnost te jasnoću podataka. Budući da sve zemlje EU provode ankete i istraživanja u skladu s preporukama Eurostata, potrebno je u skladu s tim preporukama ispuniti izvještaje o kvaliteti podataka s detaljnom analizom pokazatelja kvalitete podataka. Stoga je pripremljen ovaj priručnik kako bi zaposlenici DZS-a, korisnici proizvoda DZS-a te istraživači mogli dobiti mnogo vrijednih informacija o kvaliteti podataka te izračunu pokazatelja kvalitete.

Kako bi se procijenila koja je razina kvalitete podataka postignuta istraživanjem<sup>1)</sup>, potrebno je detaljno proučiti podatke i primijenjene procedure, i to gledajući sa strane i inputa i outputa implementiranih statističkih procesa. To uključuje analizu metodologije i implementiranog procesa proizvodnje statistika, znači, kako su podaci prikupljeni, a zatim i kako su statistički tretirani, obrađeni i analizirani. Potonje, uključuju usporedbu rezultata ostvarenih prema odgovarajućim standardima s alternativnim izvorima podataka, prethodnim znanjem ili logičnim očekivanjima. Naravno, svi ti aspekti kvalitete od izrazite su važnosti.

U ovom priručniku pokazatelji kvalitete objašnjeni su empirijski i u dubinu. Prilikom rada oko ovih pokazatelja kvalitete najvažnije je bilo istražiti i upozoriti s kojim se sve ograničenjima tijekom analize kvalitete prikupljenih podataka te proizvedenih statistika istraživači susreću, posebno što se tiče ograničenja dostupnih podataka iz registara ili baza podataka. Razumijevanje kvalitete istraživanja i proizvoda je pitanje od velike važnosti za pravilnu upotrebu podataka te za razvoj i poboljšanje samih anketa i istraživanja. Iz tog razloga treba jasno istaknuti razne metode analize kvalitete podataka i pokazatelje.

*Furthermore, data quality is a multi-dimensional concept that applies not only to the statistical accuracy of the data, but also to the comparability, relevance, timeliness and punctuality, accessibility and clarity of the data. As all the countries of the EU are conducting statistical surveys in accordance with the Eurostat recommendations, the quality reports have to be completed with the detailed analysis of data quality indicators required within these recommendations. This handbook has been prepared in order CBS staff and users of CBS outputs and researchers can gain a lot of valuable information about the data quality and the calculation of the quality indicators.*

*To estimate the achieved level of data quality of a statistical survey<sup>1)</sup>, it is necessary to study in depth information about the data itself and the procedures applied, and by looking to both the input and output side of the implemented statistical processes. It includes the analysis of the methodology, and the implemented statistical production process. That means how the data were collected, and then how they are statistically treated, processed and analysed. Further, this includes a comparison of achieved results according to the relevant standards with alternative sources of information, prior knowledge and logical expectations. Of course, all these quality aspects are of high importance.*

*In this handbook the quality indicators are explained empirically and in depth. Through this work on quality indicators the most important thing was to investigate and warn researchers, about the survey data/output constraints during the analysis of data quality, particularly regarding the limitations of available data from registers or databases. Understanding survey/output quality is of crucial importance for the proper use of data and for development and improvement of statistical surveys. For this reason it should be clearly pointed out the various methods of data quality analysis and quality indicators.*



Provedene ankete i istraživanja su jedinstven i koristan izvor podataka. Ipak, postoje određeni problemi i nedostaci koje treba analizirati kako bismo unaprijedili kvalitetu podataka koji se objavljuju od tih istraživanja. Ti nedostaci i problemi se razlikuju s obzirom na to koliko ih lako možemo ispraviti. Isto tako, očito je da je za neke potrebno duže razdoblje i dalekosežne promjene. Te se promjene odnose i na pravnu osnovu svakoga pojedinog istraživanja kako bi se što preciznije proveli postupci poboljšanja kvalitete podataka, pa je tako dovoljno čak i umjereno ulaganje DZS-a kako bi se ta kvaliteta poboljšala. S tim umjerenim ulaganjem poboljšala bi se i ukupna vrijednost provedenih anketa.

*The statistical surveys are unique and useful source of information. There are certain problems and deficiencies that need to be analysed in order to improve the quality of disseminated data from these surveys. These deficiencies vary with regard to how easily they can be corrected. It is also obvious that some of statistical surveys need a longer period of time to be changed. These changes relate also to the legal basis of each statistical survey in order to carry out the actions needed for improving data quality. A moderate investment by CBS is sufficient for improvement of data quality. With this investment the overall value of conducted surveys will be improved in one or more aspects of each survey quality.*

Marko Krištof

ravnatelj Državnog zavoda za statistiku

Marko Krištof

Director General  
of Croatian Bureau of Statistics

1) Pod pojmom "istraživanje" se prema pojmovniku OECD-a podrazumijevaju sva statistička istraživanja: ankete, poslovna istraživanja, popis, istraživanja iz administrativnih izvora, itd.

1) According to the OECD glossary, the term "research" refers to all statistical surveys: surveys, business research, census, research from administrative data sources, etc.

## KRATICE

ADS	Anketa o dohotku stanovništva
APK	Anketa o potrošnji kućanstava
ARS	Anketa o radnoj snazi
DZS	Državni zavod za statistiku
ESS	Europski statistički sustav
EU	Europska unija
Eurostat	Statistički ured Europskih zajednica
LRK	Logičko-računska kontrola
NKD	Nacionalna klasifikacija djelatnosti

## ABBREVIATIONS

<i>CIS</i>	<i>Community Innovation Survey</i>
<i>ESS</i>	<i>European Statistical System</i>
<i>EU</i>	<i>European Union</i>
<i>Eurostat</i>	<i>Statistical Office of the European Communities</i>
<i>SILC</i>	<i>Statistics on Income and Living Conditions</i>
<i>HBS</i>	<i>Household Budget Survey</i>
<i>IPA</i>	<i>Instrument for Pre-Accession Assistance</i>
<i>ISCED</i>	<i>International Standard Classification of Education</i>
<i>LFS</i>	<i>Labour Force Survey</i>
<i>NKD</i>	<i>Nacionalna klasifikacija djelatnosti, National Classification of Activities</i>
<i>NSO</i>	<i>National Statistical Office</i>
<i>NSI</i>	<i>National Statistical Institute</i>
<i>OECD</i>	<i>Organisation for Economic Co-operation and Development</i>



## UVOD

Mjerenje i praćenje kvalitete statističkih procesa i proizvoda ima važnu ulogu u procesima kojima se bave svi statistički uredi Europske unije, pa tako i Državni zavod za statistiku. Stoga je potrebno definirati određene pokazatelje kvalitete koji će kvantitativno mjeriti kvalitetu objavljenih podataka. Pokazatelji kvalitete pružaju objektivnu procjenu kvalitete postignutih rezultata, služe za praćenje promjena u kvaliteti tijekom vremena te za usporedbu kvalitete podataka i procesa između različitih istraživačkih područja.

Ovaj priručnik donosi prvi put na jednome mjestu metodološke definicije različitih pokazatelja kvalitete i konkretne primjere s postupcima njihova izračuna i to upravo one koji se koriste za praćenje kvalitete procesa i proizvoda u DZS-u. Glavni cilj ovog priručnika je uspostaviti jasan i transparentan standard izračuna pokazatelja kvalitete te omogućiti uvid i informacije o njima svim zaposlenicima DZS-a (pogotovo stručnim odjelima koji provode različite ankete i istraživanja), ali i vanjskim korisnicima radi lakšeg shvaćanja i tumačenja objavljenih podataka istraživanja i rezultata.

Budući da je Eurostat uspostavio radnu skupinu upravo za praćenje i kontrolu kvalitete statističkih procesa i proizvoda, glavni zadatak te radne skupine svakako je razvoj standardiziranih metoda izvješćivanja o kvaliteti među zemljama članicama EU. Tako metodološki dokumenti, nastali kao glavni rezultat rada te radne skupine, definiraju standardne strukture izvješćivanja te točno definiraju i opisuju šest sastavnica kvalitete statističkih rezultata, a to su: relevantnost, točnost, pravodobnost i vremenska određenost, usporedivost, dostupnost i jasnoća te smislenost podataka. Kao sedma sastavnica kvalitete navode se još troškovi i obveze, koji nisu izričita sastavnica kvalitete, ali imaju znatan utjecaj na svih šest navedenih sastavnica.

## INTRODUCTION

*Measuring and monitoring the quality of statistical processes and products play an important role in the processes of all National Statistical institutes of the EU, including the Croatian Bureau of Statistics. It is necessary to define specific quality indicators for measuring the quality of the disseminated data in quantitative terms. The quality indicators provide an objective assessment of the quality of the achieved results, and are used for monitoring quality changes over time and to compare the quality of data and processes between different survey domains.*

*This handbook brings for the first time into one place, methodological definitions of various quality indicators, and concrete examples of their calculation methods, particularly for those that are used for monitoring the quality of processes and products in CBS. The main objective of this handbook is to establish clear and transparent standards for quality indicators and to provide insight and information for all CBS employees (especially and information involved with various statistical surveys), and for external users in order to facilitate their better comprehension and interpretation of the disseminated survey data and results.*

*Since Eurostat has established a working group dedicated to monitor and control the quality of statistical processes and products, a main task of this working group is certainly the development of standardised methods for quality reporting in the EU member states. The main working results of this group are the methodological documents, which define a standard report structures and describe the six components of quality of statistical results. These are: relevance, accuracy, timeliness and punctuality, comparability, accessibility and clarity as well as coherence. As a seventh quality component also listed are costs and liabilities, which are not explicitly a quality component, but have a significant impact on all six mentioned components.*

U DZS-u je ovaj posao o pokazateljima kvalitete objedinjen u izvještajima o kvaliteti statističkih istraživanja. Važan dio svakog izvješća o kvaliteti svakako su pokazatelji kvalitete, koji kroz numeričke vrijednosti pokazuju ostvarenu razinu kvalitete u kontekstu kvalitete svake pojedine istražene komponente. Glavni cilj izvještaja o kvaliteti je pružiti proizvođačima raznih statistika analitički uvid u cjelokupni statistički proces, a s druge strane korisnicima statističkih rezultata pružiti dodatne informacije za ispravno korištenje i tumačenje rezultata.

Budući da je DZS 2013. počeo s pripremom i skupljanjem dokumentacije o kontroli kvalitete svih statističkih procesa i proizvoda kojima se bavi, ovaj priručnik je važan prvi korak u upoznavanju svih DZS-ovih istraživača s procesom kontrole kvalitete. Radi boljeg tumačenja i primjene, kako istraživačima iz stručnih odjela DZS-a pa tako i vanjskim korisnicima podataka, pokazatelji su objašnjeni na jednostavan način te su na praktičnim primjerima protumačeni njihovi izračuni, što je izrazito bitno jer su oni glavna sastavnica svakoga standardiziranog izvješća o kvaliteti.

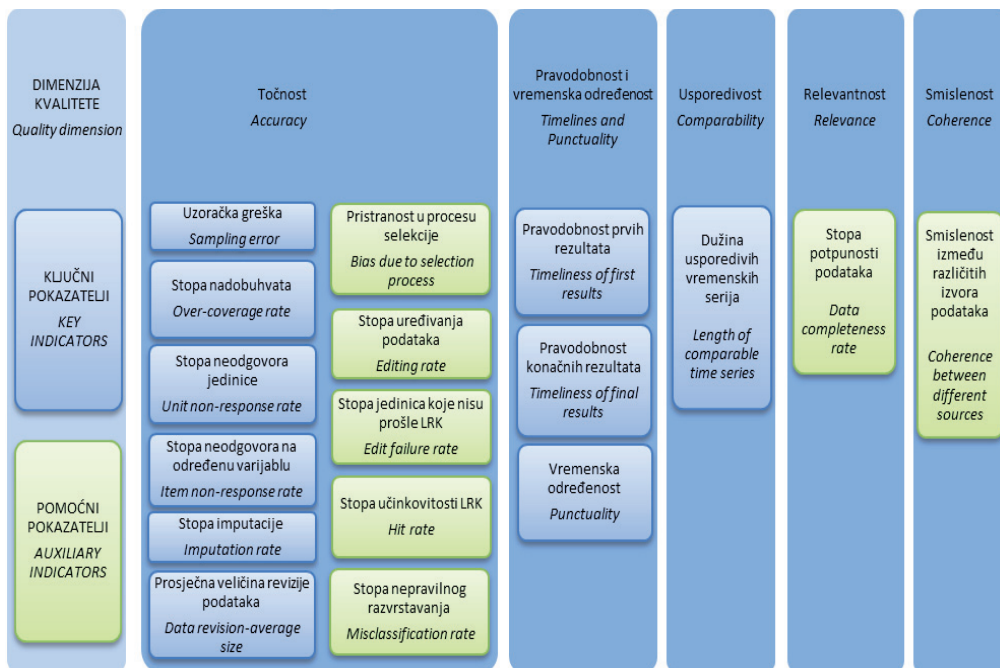
Priručnik je sastavljen tako da su pokazatelji kvalitete razvrstani u dvije kategorije, na ključne i pomoćne pokazatelje kvalitete. Ključni pokazatelji kvalitete osnovni su pokazatelji koje zahtijeva Eurostat u svojim istraživanjima (onima pokrivenima Regulativama EU), dok su pomoćni pokazatelji dodatna mjerila kvalitete koja se mogu izračunavati na dobrovoljnoj osnovi i općenito su bitni isključivo za statistike na nacionalnoj razini. U nastavku je grafički prikaz osnovne strukture prema kojoj je sastavljen priručnik za izračun pokazatelja kvalitete. Pokazatelji kvalitete razvrstani su na ključne i pomoćne, uzevši u obzir dimenzije kvalitete relevantnost, točnost, pravodobnost i vremensku određenost, usporedivost i smislenost.

*This work on quality indicators in the CBS is brought together in survey quality reports. The important part of any quality report is certainly its quality indicators, which show the numerical value of the achieved quality level in the context of each individual quality component examined in a survey. The main quality report goal is to give producers of different statistics the analytical insight into the overall statistical process, and for the users of the statistical results to provide them additional information for their use and interpretation of results in a correct way.*

*Since the CBS has started with the preparation and collection of documentation on the quality control of all statistical processes and products in 2013, this handbook is the important first step in making CBS's researchers familiar with the process of quality control. In order to be useful to professionals from CBS and to all external data users, the indicators are illustrated in a simple way and their calculations are explained in this handbook with the practical examples, which is extremely important because they are main component of every standardised quality report.*

*The handbook has been compiled in a way that the quality indicators are classified into two categories - key indicators and auxiliary indicators. The key quality indicators are the basic indicators required by Eurostat for their surveys, (i.e. those covered by EU Regulations) while auxiliary indicators represent additional quality measures that can be calculated on a voluntary basis and are generally important only for statistics at national level. A graph below shows the basic structure which is used as structure of the prepared Handbook for the calculation of quality indicators. Taking into account the quality dimensions relevance, accuracy, timeliness and punctuality, comparability and coherence the quality indicators are then divided into key and auxiliary indicators.*

**Slika 1. Pokazatelji kvalitete**  
*Picture 1 Quality indicators*



## 1. KLJUČNI POKAZATELJI

Kao što je spomenuto u uvodnom dijelu, ključni pokazatelji kvalitete odnose se na pokazatelje koje zahtijeva Eurostat u svojim istraživanjima pa svi nacionalni statistički uredi imaju obvezu njihova dostavljanja u Eurostat.

Ključni pokazatelji kvalitete, sa šiframa pokazatelja, strukturirani su na sljedeći način:

1. Dimenzija kvalitete: Točnost
  - a. Uzoračka greška – A1
  - b. Stopa nadobuhvata – A2
  - c. Stopa neodgovora jedinica – A4
  - d. Stopa neodgovora na određenu varijablu – A5
  - e. Stopa imputacije – A7
  - f. Prosječna veličina revizije podataka – A6
2. Dimenzija kvalitete: Pravodobnost i vremenska određenost
  - a. Pravodobnost prvih rezultata – TP1
  - b. Pravodobnost konačnih rezultata – TP2
  - c. Vremenska određenost – TP3
3. Dimenzija kvalitete: Usporedivost
  - a. Dužina usporedivih vremenskih serija – CC2

### 1.1. TOČNOST

U općem statističkom smislu dimenzija kvalitete točnost označava moguću razliku procijenjenih i stvarnih podataka populacije. Statistički podaci nisu jednaki stvarnim vrijednostima zbog varijabilnosti (kada se vrijednosti mijenjaju zbog slučajnih utjecaja, a odnose se na greške koje se pojavljuju u provedbi ankete) i pristranosti (kada se vrijednosti mijenjaju zbog sustavnih djelovanja, a odnose se na greške koje se pojavljuju u provedbi ankete).

## 1. KEY INDICATORS

*As mentioned in the introduction, the key quality indicators refer to the indicators required by Eurostat in their surveys and therefore all national statistical offices have an obligation to deliver them to Eurostat.*

*The key quality indicators, with the indicator codes, are structured as follows:*

1. *Quality dimension: Accuracy*
  - a. *Sampling error – A1*
  - b. *Over-coverage rate – A2*
  - c. *Unit non-response rate – A4*
  - d. *Item non-response rate – A5*
  
  - e. *Imputation rate – A7*
  - f. *Data revision-average size – A6*
2. *Quality dimension: Timeliness and punctuality*
  - a. *Timeliness of the first results – TP1*
  - b. *Timeliness of the final results – TP2*
  
  - c. *Punctuality – TP3*
3. *Quality dimension: comparability*
  - a. *Length of comparable time series – CC2*

### 1.1. ACCURACY

*In the general statistical sense the quality dimension accuracy denotes as much as it can the difference between estimated and the real population data. The statistical data are not equal to their real values because of variability (when values change due to random effects, which relate to the errors that occur during the implementation of the statistical survey), and bias (when values change due to the systematic effects, related to the errors that occur during the implementation of statistical survey).*

### 1.1.1. Uzoračka greška (A1)

#### Definicija

Zbog velikih troškova provedbe istraživanja na cijeloj populaciji parametri populacije se procjenjuju na temelju podataka prikupljenih na uzorku. Iz tog razloga, radi ocjene preciznosti procjena, nužno je izračunati uzoračku grešku procjene za istraživanja na uzorku. Uzoračka greška je rezultat provedbe istraživanja cijele populacije na uzorku, a ne na punom obuhvatu.

Uzoračka greška se može izraziti na sljedeće načine:

- u apsolutnom smislu – kao standardna pogreška,
- u relativnom smislu – kao koeficijent varijacije, ili
- u smislu pouzdanosti – kao interval pouzdanosti

Standardna pogreška je procjena apsolutne razlike procjene parametara populacije na temelju uzorka i stvarne vrijednosti tog parametra u populaciji. Statistika standardne pogreške predstavlja vrstu statističkih podataka koji se u mnogim inferencijalnim statistikama prikazuju kao izlazni podaci, no funkcioniraju kao deskriptivna statistika. Izraz "standardna pogreška" odnosi se na skupinu statističkih podataka koji pružaju informacije o raspršenju vrijednosti unutar nekog skupa. Kod primjene standardne pogreške pretpostavlja se da je korisnik upoznat s centralnim graničnim teoremom i pretpostavkama o skupu podataka s kojim istraživač radi.

Središnji granični teorem je temeljna pretpostavka cjelokupne parametarske inferencijalne statistike. Za njegovu je primjenu nužno da uzorak bude slučajan te da su zapažanja o svakom pojedinom ispitaniku neovisna o zapažanjima o bilo kojemu drugom ispitaniku. Teorem ustvrđuje da će srednje vrijednosti iz velikog broja izabranih slučajnih uzoraka biti normalno distribuirane bez obzira na oblik ishodišne populacije (1).

### 1.1.1. Sampling error (A1)

#### Definition:

*Due to the high costs of carrying out a statistical survey based on the entire population, the required population parameters are estimated on the data collected by sample. For this reason, in order to evaluate the precision of estimates, it is necessary to calculate the sampling error of estimates from the sample surveys. The sampling error is the result of the survey of the whole population being conducted on a sample and not on the full coverage.*

*The sampling error can be expressed as follows:*

- *in absolute terms - as the standard error,*
- *in relative terms - as the coefficient of variation, or*
- *in terms of reliability – as the interval of confidence*

*The standard error is an estimate of absolute difference between parameter estimate based on a sample and the real value for the parameter of population. The standard error statistics represent the type of statistical data that are in many inferential statistics presented as output data, but function as descriptive statistics. The term "standard error" refers to a set of statistics that provide information about the dispersion of values within the data set. When applying the standard error it is assumed that the user is familiar with the central limit theorem, and hypotheses about the data set with which the researcher works.*

*The central limit theorem is the fundamental hypothesis of the parametric inferential statistics. For its application it is necessary that the sample is random and that the observation on each respondent is independent of any other respondent. The theorem determines that the mean values of sampling distribution obtained from a large number of random samples will represent a normal distribution regardless the form of the initial population (1).*



Točnije, iako mali uzorci mogu rezultirati nenormalnim distribucijama, s povećanjem veličine uzoraka (tj. povećanjem  $n$ ) oblik distribucije srednjih vrijednosti uzoraka ubrzano će se približavati obliku normalne distribucije. Druga generalizacija na temelju centralnog graničnog teorema je da kako raste  $n$ , tako se smanjuje varijabilnost srednje vrijednosti uzoraka (2). To je važno jer pojam distribucije uzoraka tvori teorijsku osnovu koja istraživačima omogućuje da donose zaključke o populaciji iz uzoraka.

Istraživači obično uzimaju samo uzorak kao podskup populacije jer nije moguće mjeriti cijelu populaciju. Za to nemaju ni vremena ni novca. Iz istih razloga istraživači ne mogu izabrati mnogo uzoraka iz ciljane populacije. Za njih je stoga osnovno da mogu odrediti vjerojatnost prema kojoj će mjere njihova uzorka pouzdano predstavljati cjelokupnu populaciju, za koju će se zatim raditi procjene temeljene na uzorku. Određivanje reprezentativnosti određenog uzorka temelji se na teorijskoj raspodjeli uzoraka čije ponašanje opisuje centralni granični teorem. Statistika standardne pogreške predstavlja procjene intervala u kojem je moguće procijeniti parametre populacije, a time i stupanj preciznosti u kojem statistika uzorka predstavlja parametar populacije. Što je standardna pogreška manja, to je statistika uzorka bliža parametru populacije. Standardna pogreška statistike je stoga standardno odstupanje distribucije uzorka za tu statistiku (3).

S druge strane, koeficijent varijacije je procijenjena standardna pogreška procjenitelja podijeljena s procijenjenom vrijednošću parametra te je izražen u relativnom smislu. Budući da se izražava kao relativna mjera, mnogo je zahvalniji za ocjenu reprezentativnosti procjene od standardne pogreške.

Interval pouzdanosti je interval stvarnih vrijednosti parametara uz određenu razinu pouzdanosti.

*More precisely, although small samples can lead to abnormal distributions, with an increase of sample size (i.e., increasing  $n$ ) the form of sample distribution of mean values is rapidly approaching the form of a normal distribution. The second generalisation based on the central limit theorem is that as  $n$  increases, the variability of the sample means is reducing (2). This is important because the sampling distribution term forms the theoretical basis for mathematics, which allows researchers to draw conclusions about the population based on the sample.*

*The researchers usually take a sample as a subset of the population because it is not possible to measure the entire population. For doing that, they do not have either the time or the money. Therefore, the researchers cannot select a lot of samples from a target population. Fundamental for them is to determine the probability that measures of the sample represents accurately the entire population, for which they will do the estimates based on the sample. Determining the representativeness of the samples is based on the theoretical distribution of samples whose behaviour describes the central limit theorem. The standard error statistics represent the estimation of the interval in which it is possible to estimate the population parameters, and thereby the precision degree in which the sample statistic represents the population parameter. When the standard error is lower, the sample statistic is closer to the population parameter. The standard error statistics is therefore the standard deviation of the sample distribution for that statistic (3).*

*On the other hand, the coefficient of variation is the estimated standard error of the estimator divided by the estimated value of the parameter expressed in relative terms. Since it is expressed as a relative measure, it is much more acceptable for assessing the representativeness of estimates than standard error.*

*The confidence interval is the interval of the actual parameter values with a certain confidence level.*

### Formula za izračun

Dani su primjeri formula za jednostavan slučajni uzorak i procjenu prosječne vrijednosti populacije. Naime, postoje i druge složenije formule za izračun uzoračke greške za različite nacрте uzorka te za druge procjene (totala, proporcija, medijana itd.).

Nepristran procjenitelj parametra populacije aritmetičkom sredinom:

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$\hat{Y}$  – procijenjena vrijednost aritmetičke sredine populacije  
 $y_i$  – vrijednost varijable iz uzorka  
 $n$  – veličina uzorka

Varijanca uzorka:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{y} - y_i)^2$$

$s^2$  – varijanca  
 $\bar{y}$  – aritmetička sredina vrijednosti varijable  $y_i$   
 $y_i$  – vrijednost varijable iz uzorka  
 $n$  – veličina uzorka

Standardna pogreška procjene aritmetičke sredine:

$$se(\hat{Y}) = \frac{s}{\sqrt{n}} \quad \text{za } f < 0,05 \text{ ili}$$

$$se(\hat{Y}) = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad \text{za } f > 0,05$$

$s$  – standardna devijacija uzorka  
 $n$  – veličina uzorka  
 $N$  – veličina populacije

pri čemu je  $f$  frakcija izbora uzorka koja je jednaka  $\frac{n}{N}$ .

### Formula for calculation

These are the examples of formula for a simple random sample and for estimate of a population mean. There are other more complex formulas for calculating sampling errors for the various sample designs and for other estimation types (totals, proportions, median, etc.).

The unbiased estimator of the population parameter by mean:

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$\hat{Y}$  – estimated value of population mean  
 $y_i$  – sample variable value  
 $n$  – sample size

Sample variance:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{y} - y_i)^2$$

$s^2$  – variance  
 $\bar{y}$  – average value of variable  $y_i$   
 $y_i$  – sample estimator value  
 $n$  – sample size

The standard error of the estimated mean:

$$se(\hat{Y}) = \frac{s}{\sqrt{n}} \quad \text{for } f < 0,05 \text{ or}$$

$$se(\hat{Y}) = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad \text{for } f > 0,05$$

$s$  – sample standard deviation  
 $n$  – sample size  
 $N$  – population size

where  $f$  is the sampling fraction, which is equal to  $\frac{n}{N}$ .

Prva jednostavnija formula za standardnu pogrešku se primjenjuje u slučajevima kad je uzorak manji od 5% populacije (kao što je slučaj kod socijalnih anketa na kućanstvima i osobama), dok se kod uzoraka većih od 5% populacije (kao što je slučaj kod poslovnih istraživanja) standardna greška još dodatno množi s faktorom korekcije za konačne populacije  $\sqrt{\frac{N-n}{N-1}}$ .

Koeficijent varijacije:

$$cv(\hat{Y}) = \frac{se(\hat{Y})}{\hat{Y}}$$

$se(\hat{Y})$  – standardna pogreška procjene aritmetičke sredine  
 $\hat{Y}$  – procijenjena vrijednost aritmetičke sredine populacije

Interval pouzdanosti:

$$\left(\hat{Y} - 1,96 \cdot se(\hat{Y}) < \bar{Y} < \hat{Y} + 1,96 \cdot se(\hat{Y})\right) = 95\%$$

$$\left[\hat{Y} - 1,96 \cdot se(\hat{Y}), \hat{Y} + 1,96 \cdot se(\hat{Y})\right]$$

$\bar{Y}$  – stvarna vrijednost aritmetičke sredine populacije  
 $\hat{Y}$  – procijenjena vrijednost aritmetičke sredine populacije  
 $se(\hat{Y})$  – standardna pogreška procjene aritmetičke sredine

### Primjer

Prema Anketi o radnoj snazi, u četvrtom tromjesečju 2012. procijenjena je stopa nezaposlenosti od 17,98% dok je varijanca uzorka ( $s^2$ ) iznosila 26% (procjena varijance populacije na temelju uzorka). Veličina populacije je 100.000, a uzorka 1.000.

*The first simple formula for standard error is used in cases when the sample is less than 5% of the population (as it is the case in the social statistical surveys on households and persons), while on samples bigger than 5% of the population (as it is the case with business statistical surveys) the standard error is additionally multiplied by the factor of correction for finite population  $\sqrt{\frac{N-n}{N-1}}$ .*

*Coefficient of variation:*

$$cv(\hat{Y}) = \frac{se(\hat{Y})}{\hat{Y}}$$

*$se(\hat{Y})$  – standard error of estimate of the mean  
 $\hat{Y}$  – estimated value of the population mean*

*Confidence interval:*

$$\left(\hat{Y} - 1,96 \cdot se(\hat{Y}) < \bar{Y} < \hat{Y} + 1,96 \cdot se(\hat{Y})\right) = 95\%$$

$$\left[\hat{Y} - 1,96 \cdot se(\hat{Y}), \hat{Y} + 1,96 \cdot se(\hat{Y})\right]$$

*$\bar{Y}$  – real value of the population mean  
 $\hat{Y}$  – estimated value of the population parameter  
 $se(\hat{Y})$  – standard error of estimate of the mean*

### Example

*According to the Labour Force Survey, in the fourth quarter of 2012 the estimated unemployment rate amounts 17.98% while the sample variance ( $s^2$ ) was equal to 26% (population variance estimate on the sample). The size of the population is 100,000, and sample 1,000.*

Primjer prikazuje izračun koeficijenta varijacije i 95%-tnog intervala pouzdanosti procjene stope nezaposlenosti na jednostavnome slučajnom uzorku.

Postupak izračuna započinje utvrđivanjem frakcije izbora.

$$f = \frac{n}{N} = \frac{1000}{100000} = 0,01 = 1\%$$

Budući da je uzorak manji od 5% populacije standardna greška procjene aritmetičke sredine jednaka je:

$$se(\hat{Y}) = \frac{s}{\sqrt{n}} = \frac{\sqrt{26}}{\sqrt{1000}} = 0,16$$

$$CV(\hat{Y}) = \frac{se(\hat{Y})}{\hat{Y}} = \frac{0,16}{17,98} \cdot 100 = 0,89\%$$

Interval pouzdanosti je jednak:

$$P(\hat{Y} - 1,96 \cdot se(\hat{Y}) < Y < \hat{Y} + 1,96 \cdot se(\hat{Y})) = 95\%$$

$$P(17,98 - 1,96 \cdot 0,16 < Y < 17,98 + 1,96 \cdot 0,16) = 95\%$$

$$P(17,67 < Y < 18,29) = 95\%$$

$$[17,67, \quad 18,29]$$

Iz dobivenih rezultata možemo zaključiti da je prosječno odstupanje stope nezaposlenosti od njene prosječne vrijednosti 0,16 ili u relativnom iznosu 0,89%. Ujedno je radi poboljšanja preciznosti procjena izračunan i interval pouzdanosti stope nezaposlenosti kojim se uz 95% pouzdanosti procjenjuje da se stopa nezaposlenosti nalazi u intervalu između 17,67% i 18,29%.

*The following example shows the calculation of the coefficient of variation and the 95% confidence interval of the estimated unemployment rate on the simple random sample.*

*Calculation begins with calculating the sample fraction.*

$$f = \frac{n}{N} = \frac{1000}{100000} = 0,01 = 1\%$$

*Since the sample is less than 5% of the population, standard error of the estimated mean is equal to:*

$$se(\hat{Y}) = \frac{s}{\sqrt{n}} = \frac{\sqrt{26}}{\sqrt{1000}} = 0,16$$

$$CV(\hat{Y}) = \frac{se(\hat{Y})}{\hat{Y}} = \frac{0,16}{17,98} \cdot 100 = 0,89\%$$

*The confidence interval is equal to:*

$$P(\hat{Y} - 1,96 \cdot se(\hat{Y}) < Y < \hat{Y} + 1,96 \cdot se(\hat{Y})) = 95\%$$

$$P(17,98 - 1,96 \cdot 0,16 < Y < 17,98 + 1,96 \cdot 0,16) = 95\%$$

$$P(17,67 < Y < 18,29) = 95\%$$

$$[17,67, \quad 18,29]$$

*From the results obtained it can be concluded that the average deviation of unemployment rate from its average value equals to 0.16 or 0.89% in relative terms. At the same time in order to improve the precision of estimates the confidence interval was calculated, which shows with 95% reliability the unemployment rate between 17.67% and 18.29%.*

## 1.1.2. Stopa nadobuhvata (A2)

### Definicija

Stopa nadobuhvata predstavlja udio jedinica iz okvira uzorka koje ne pripadaju ciljanoj populaciji (npr. poduzeća koja više ne rade ili koja se bave drugom djelatnošću koja nije pokrivena istraživanjem, ili osobe iz kućanstava koje su se iselile u drugu zemlju ili su preminule). Ako se istraživanje provodi na uzorku, stopa nadobuhvata se procjenjuje na temelju podataka prikupljenih na uzorku.

Ukoliko se stopa nadobuhvata računa temeljem podataka iz uzorka mogu se računati neutežena i utežena stopa nadobuhvata.

Ako se istraživanje temelji na uzorku pa svaka jedinica iz uzorka predstavlja određen broj jedinica u ciljanoj populaciji, logično je računati uteženu stopu nadobuhvata. U slučaju da sve jedinice iz uzorka imaju jednaku težinu, težina se u postupku izračuna zanemaruje i tada govorimo o izračunu neutežene stope nadobuhvata.

### Formula za izračun

$$OCRr_w = \frac{\sum_o w_j + (1 - \alpha) \sum_Q w_j}{\sum_o w_j + \sum_E w_j + \sum_Q w_j}$$

*O* – skup jedinica izvan obuhvata (ne pripadaju ciljanoj populaciji)

*E* – skup jedinica unutar obuhvata (pripadaju ciljanoj populaciji ili valjane jedinice)

*Q* – skup jedinica nepoznate valjanosti

*w* – težina jedinice

$\alpha$  – procijenjeni udio jedinica nepoznate valjanosti koje su zapravo valjane (u praksi se uglavnom pretpostavlja da iznosi 1: brojnik u tom slučaju uvijek obuhvaća samo jedinice izvan obuhvata)

## 1.1.2. Over-coverage rate (A2)

### Definition

*Over-coverage rate is the proportion of units in the frame that do not belong to the target population (e.g. companies that are no longer working or doing other economic activity that is not covered by the statistical survey, or the persons from households that have emigrated to another country, or are deceased). If the statistical survey is carried out on a sample, the over-coverage rate is estimated on the data collected on sample.*

*If the over-coverage rate is calculated on the basis of sample data either weighted or un-weighted over-coverage rate could be calculated.*

*If the research is based on a sample, and each unit in the sample represents a certain number of units in the target population, it is logical to calculate weighted over-coverage rate. In case that all units in the sample have the same weight, the weights in the process of calculation are ignored and then we talk about calculating of un-weighted over-coverage rates.*

### Formula for calculation

$$OCRr_w = \frac{\sum_o w_j + (1 - \alpha) \sum_Q w_j}{\sum_o w_j + \sum_E w_j + \sum_Q w_j}$$

*O* – set of out-of-scope units (that do not belong to the target population)

*E* – set of in-scope units (that belong to the target population, or eligible units)

*Q* – set of units of unknown eligibility

*w* – unit weight

$\alpha$  – estimated proportion of unknown eligibility units that are actually eligible (in practice is usually equal to 1: in that case numerator consists of out-of-scope units only)

Ako se želimo izračunati usporedive vrijednosti stope nadobuhvata za poslovna istraživanja, moraju se definirati standardizirani statusi valjanosti poslovnih subjekata iz uzorka, te odrediti koji označavaju nadobuhvat.

### Primjer

U uzorak za četvrto tromjesečje 2012. Ankete o radnoj snazi izabrano je 5 616 adresa (stanova), od čega je 764 bilo praznih stanova. Nadalje, na 194 adrese stambena jedinica nije bila dostupna (zbog elementarnih nepogoda i sl.) ili je nije bilo moguće pronaći. Primjer prikazuje postupak izračuna neutežene stope nadobuhvata (tj. stope za adrese) pod pretpostavkom da sve jedinice imaju istu težinu.

$$OCRr = \frac{764 + 194}{5616} = 0,1706 = 17,06\%$$

Pokazatelj računamo stavljanjem u omjer jedinica koje nisu valjane (dakle praznih stanova) uključujući jedinice koje nisu bile dostupne, sa svim jedinicama koje su izabrane u uzorak. Zaključujemo kako 17,06% svih jedinica iz okvira uzorka ne pripada ciljanoj populaciji.

### Primjer

Primjer prikazuje izračun utežene stope nadobuhvata pod pretpostavkom da sve jedinice imaju svoju težinu, i to prema oznaci valjanosti jedinice, kako je navedeno u nastavku.

Statusi valjanosti jedinice:

- 1 – valjana jedinica koja pripada ciljanoj populaciji
- 2 – jedinica koja ne pripada ciljanoj populaciji (jedinice izvan obuhvata) – nevaljane jedinice
- 3 – jedinica nepoznate valjanosti – npr. stambena jedinica koja nije bila dostupna (zbog elementarnih nepogoda i sl.) ili stambena jedinica koju nije bilo moguće pronaći.

*If comparable values of over-coverage rate are to be calculated for business surveys there needs to be an agreement for standardised status codes that can be applied to sampled businesses and which status codes leads to over-coverage.*

### Example

*In the sample of the Labour Force Survey for the fourth quarter of 2012 5616 addresses (flats) were selected, of which 764 were empty flats. In addition, there were 194 addresses where either housing unit was not available (due to natural disasters, etc.) or it was not possible to find a housing unit. The following example shows the calculation of un-weighted coverage rate (i.e. the rate for address) assuming that all units have the same weight.*

$$OCRr = \frac{764 + 194}{5616} = 0,1706 = 17,06\%$$

*The indicator is calculated as a ratio of units that are not eligible (i.e. empty flats) including units that were not available, with all the units that were selected in the sample. It could be concluded that 17.06% of all sample frame units does not belong to the target population.*

### Example

*The example shows the calculation of weighted over-coverage rate assuming that all units have their weight, according to the label of eligibility, as specified below.*

Statuses of the unit eligibility:

- 1 – valid unit that belongs to the target population
- 2 – unit that does not belong to the target population (out-of-scope unit) - ineligible unit
- 3 – unit of unknown eligibility – e.g. housing unit which was not available (due to natural disasters, etc.) or housing unit which was not possible to find.

### T 1. Izračun stope nadobuhvata na hipotetskom primjeru

*Calculation of the over-coverage rate on the hypothetical example*

Oznaka valjanosti jedinice <i>Label of the unit eligibility</i>	Težina jedinice <i>Unit weight</i>
1	5,4
1	16,7
3	24,0
2	3,5
1	125,6
2	110,7
3	34,5
2	67,2
2	14,5
3	6,2
1	8,6
1	46,2

Izvor: hipotetski primjer

Source: *hypothetical example*

$$OCR_{r_w} = \frac{(3,5 + 110,7 + 67,2 + 14,5) + (24,0 + 34,5 + 6,2)}{463,1}$$

$$= 0,563 = 56,27\%$$

U brojniku se nalazi suma težina svih jedinica koje iz određenih razloga nisu valjane ili s njima nije bilo moguće ostvariti kontakt. Dakle, to su težine jedinica koje u primjeru imaju oznake 2 i 3. Nazivnik sadrži sume težina svih jedinica iz uzorka.

Zaključujemo kako 56,27% svih jedinica iz okvira uzorka ne pripada ciljanoj populaciji.

$$OCR_{r_w} = \frac{(3,5 + 110,7 + 67,2 + 14,5) + (24,0 + 34,5 + 6,2)}{463,1}$$

$$= 0,563 = 56,27\%$$

*The numerator consists of the sum of the weights of all units that for certain reasons are not valid, or it was not possible to make contact with them. Those units were marked in the example by 2 and 3. The denominator contains the sum of the weights for all units in the sample.*

*We conclude that 56.27% of all units from the sampling frame do not belong to the target population.*

### 1.1.3. Stopa neodgovora jedinica (A4)

#### Definicija

Stopa neodgovora jedinica je pokazatelj koji govori o tome koliko jedinica, poduzeća ili kućanstava, nije odgovorilo na upitnik u cijelosti, a definira se kao udio jedinica za koje nisu dostupni podaci u odnosu na ukupan broj valjanih jedinica (koje pripadaju ciljanoj populaciji).

Ovisno o tome računamo li pokazatelje za uzorak ili cijelu populaciju, može se izračunati utežena i neutežena stopa neodgovora jedinica.

#### Formula za izračun

Stopa neodgovora jedinica ima tri osnovne verzije koje su zapisane u istoj formuli. Moguće je računati neuteženu stopu, uteženu stopu s težinama izbora uzorka (*design weight*) ili uteženu stopu s korekcijom težina za pomoćnu varijablu.

$$NRr_w = 1 - \frac{\sum_R w_j}{\sum_R w_j + \sum_{NR} w_j + \alpha \sum_Q w_j}$$

$R$  – skup valjanih jedinica koje su odgovorile

$NR$  – skup valjanih jedinica koje nisu odgovorile

$Q$  – skup odabranih jedinica nepoznate valjanosti

$w_j$  – težina jedinice

$\alpha$  – procijenjeni udio jedinica nepoznate valjanosti koje su zapravo valjane (u praksi se uglavnom pretpostavlja da iznosi 1).

Kao i kod stope nadobuhvata, izračun stopa odgovora će također ovisiti o unaprijed definiranim standardiziranim statusima odgovora koje pripisujemo svakoj jedinici uzorka dijeleći ih tako na valjane, nevaljane jedinice te jedinice nepoznate valjanosti.

### 1.1.3. Unit non-response rate (A4)

#### Definition

The unit non-response rate is an indicator which shows how many units, enterprises or households did not respond to the survey questionnaire as a whole. It is defined as the proportion of units for which the data are not available compared to the total number of eligible units (units that belong to the target population).

Depending on the calculation of indicators for a sample or a target population, weighted and un-weighted unit non-response rate can be calculated.

#### Formula for calculation

The unit non-response rate has three basic versions, which are written in the same formula. It is possible to calculate the un-weighted unit non-response rate, weighted unit non-response rate by design weight or weighted unit non-response rate where weights are corrected using auxiliary variable.

$$NRr_w = 1 - \frac{\sum_R w_j}{\sum_R w_j + \sum_{NR} w_j + \alpha \sum_Q w_j}$$

$R$  – set of eligible units of respondents

$NR$  – set of eligible units of non-respondents

$Q$  – set of selected units of unknown eligibility

$w_j$  – unit weight

$\alpha$  – estimated proportion of unknown eligibility units that are actually eligible (in practice is usually equal to 1)

As with the over-coverage rate the calculation of response rates will depend on the response status codes allocated in each survey and their consistent allocation including eligible, ineligible and unknown eligibility units.



## Primjer

Slijedeći primjer prikazuje istraživanje sa šest poslovnih subjekata, na temelju čega treba izračunati neuteženu i uteženu stopu neodgovora jedinica. Status 1 predstavlja odgovor; status 2 neodgovor; status 3 jedinicu koja ne pripada ciljanoj populaciji (jedinica koje nije valjana). Podatak o zaposlenima je pomoćna varijabla koja je poznata iz Statističkog poslovnog registra.

## Example

The following example shows a survey of six business entities based on which should be calculated un-weighted and weighted unit non-response rate. Status 1 represents the response, the status 2 non-response and status 3 unit that does not belong to the target population (ineligible unit). The data on the employment is an auxiliary variable that is known from the Statistical Business Register.

### T 2. Izračun neodgovora jedinica na hipotetskom primjeru

*Calculation of unit non-response on hypothetical example*

ID <i>ID</i>	Status <i>Status</i>	Težina uzorka <i>Sampling weight</i>	Broj zaposlenih <i>Number of employees</i>
1	1	1,45	35
2	1	1,32	48
3	1	1,00	123
4	2	5,23	13
5	2	3,20	21
6	3	1,82	8

Izvor: hipotetski primjer

Source: hypothetical example

Neutežena stopa neodgovora jedinica:

$$NRr = 1 - \frac{1 + 1 + 1}{(1 + 1 + 1) + (1 + 1)} =$$

$$= 1 - 0,6 = 0,4 = 40\%$$

Neutežena stopa neodgovora u ovom primjeru se računa na način da se stave u omjer valjane jedinice koje nisu odgovorile na anketu sa svim valjanim jedinicama i oduzmu od jedan. Dobiveni rezultat od 40% tako pokazuje ukupan broj svih jedinica koje nisu odgovorile na istraživanje u cijelosti.

Un-weighted unit non-response rate:

$$NRr = 1 - \frac{1 + 1 + 1}{(1 + 1 + 1) + (1 + 1)} =$$

$$= 1 - 0,6 = 0,4 = 40\%$$

Un-weighted unit non-response rate in this example is calculated as a ratio of eligible units that did not respond to the questionnaire with all eligible units, subtracted from one. Thus the obtained result of 40% shows the total number of all units that did not respond to the questionnaire in the whole.

Utežena stopa neodgovora jedinica s težinama izbora uzorka:

$$NRr_w = 1 - \frac{1,45 + 1,32 + 1,00}{(1,45 + 1,32 + 1,00) + (5,23 + 3,20)}$$

$$= 1 - \frac{3,77}{12,2} = 0,691 = 69,1\%$$

Utežena stopa neodgovora jedinica s korekcijom težina za pomoćnu varijablu:

$$NRr_w = 1 - \frac{1,45 * 35 + 1,32 * 48 + 1 * 123}{(1,45 * 35 + 1,32 * 48 + 1 * 123) + (5,23 * 13 + 3,2 * 21)}$$

$$= 1 - \frac{237,11}{372,3} = 0,3631 = 36,31\%$$

Za razliku od neutežene stope neodgovora kod računanja utežene stope ne uzima se u obzir broj jedinica u pojedinom skupu odgovora/neodgovora već se zbrajaju njihove težine. Stoga je kod stope gdje se koriste težine izbora uzorka dovoljno od jedan oduzeti omjer sume težina valjanih jedinica neodgovora i težina svih valjanih jedinica. Ukoliko se želi dobiti kvalitetniji pokazatelj može se koristiti težine korigirane s pomoćnom varijablom, na način da se prije računanja stope svaka težina jedinice pomnoži sa pripadajućim iznosom pomoćne varijable.

#### 1.1.4. Stopa neodgovora na određenu varijablu (A5)

##### Definicija

Stopa neodgovora na određenu varijablu definira se kao uteženi ili neuteženi omjer između jedinica u obuhvatu koje nisu odgovorile na određenu varijablu. Općenito, to je ograničeno na sve valjane jedinice od interesa koje moraju odgovoriti na određenu varijablu.

Bitna razlika u odnosu na stopu neodgovora jedinica kod koje u izračun ulaze sve valjane jedinice, je u tome što se pri izračunu uzimaju u obzir samo valjane jedinice od

*Weighted unit non-response rate by design weight:*

$$NRr_w = 1 - \frac{1,45 + 1,32 + 1,00}{(1,45 + 1,32 + 1,00) + (5,23 + 3,20)}$$

$$= 1 - \frac{3,77}{12,2} = 0,691 = 69,1\%$$

*Weighted unit non-response rate where weights are corrected using auxiliary variable:*

$$NRr_w = 1 - \frac{1,45 * 35 + 1,32 * 48 + 1 * 123}{(1,45 * 35 + 1,32 * 48 + 1 * 123) + (5,23 * 13 + 3,2 * 21)}$$

$$= 1 - \frac{237,11}{372,3} = 0,3631 = 36,31\%$$

*Unlike the un-weighted unit non-response rate, for the calculation of weighted non-response rate we take in the account the sum of weights of units in a particular set of response/nonresponse and not just number of them. Therefore, by the rate for which the design weights are used, it is enough to subtract the ratio of sum of eligible non-respondent weights and the sum of weights for all eligible units from one. If you want to get a higher quality indicator, corrected weights, which we get by multiplying design weights by the corresponding amount of an auxiliary variable, can be used.*

#### 1.1.4. Item non-response rate (A5)

##### Definition

*Item non-response rate is defined as a weighted or un-weighted ratio of the number of in-scope units which did not respond for a particular variable. Generally this is restricted to all of units in-scope that are required to respond to a particular variable.*

*An important difference compared to the unit non-response rate in which the calculation includes all valid units, is that the calculation of this rate take into account only eligible*

kojih se zahtijeva odgovor na određeno pitanje. Prema tome, jedinice kod kojih dolazi do preskoka pitanja ne ulaze u izračun stope neodgovora na određenu varijablu koju su preskočili. Na primjer, računamo li stopu neodgovora za varijablu iznos naknade za nezaposlene, osobe koje su na prethodnom pitanju (primaju li takvu naknadu) odgovorile negativno, ne uzimamo u obzir pri izračunu.

### Formula za izračun

Stopa neodgovora na određenu varijablu ima tri osnovne verzije koje su zapisane u istoj formuli. Kao i kod stope neodgovora jedinica, moguće je računati neuteženu stopu, uteženu stopu s težinama izbora uzorka ili uteženu stopu s korigiranim težinama s vrijednostima pomoćne varijable.

$$NR_{Y\Gamma_w} = 1 - \frac{\sum_{R_Y} w_j}{\sum_{R_Y} w_j + \sum_{NR_Y} w_j}$$

$R_Y$  – skup valjanih jedinica koje su odgovorile na varijablu  $Y$

$NR_Y$  – skup valjanih jedinica koje nisu odgovorile na varijablu  $Y$  iako se odgovor na nju zahtijeva.

$w_j$  – težina jedinice

Tri osnovna načina izračuna stope jesu:

1. Neutežena stopa:  $w_j = 1$
2. Stopa utežena težinom nacrtu uzorka:  $w_j = d_j$ , pri čemu je  $d_j = \frac{1}{\pi_j}$  težina nacrtu uzorka recipročna vrijednost vjerojatnosti izbora
3. Stopa utežena značajnošću jedinica:  $w_j = d_j x_j$  pri čemu je  $x_j$  vrijednost pomoćne varijable  $X$ .

Varijable  $X$ , koja se odabire subjektivno, pokazuje veličinu ili značajnost jedinica i njena vrijednost bi trebala biti poznata za sve jedinice. Vrijednost varijable  $X$  je pomoćna informacija, često dostupna u okviru. Primjeri su promet za poslovne subjekte ili ukupni broj stanovnika u općini.

*units required to give an answer to a specific question. Accordingly, the units on which are allowed by the questionnaire to skip over questions are not included in the calculation of item non-response for these skipped questions. For example, if we count the item non-response rate for the amount of unemployment benefits, persons who responded negatively on the previous question (whether receiving such benefits), are not taken into account.*

### Formula for calculation

*Item non-response rate has three basic versions that are written in the same formula. As well as the unit non-response rate, it is possible to calculate the item non-response rate as an un-weighted rate, weighted rate by design weight, or and weighted rate where weights are corrected using an auxiliary variable.*

$$NR_{Y\Gamma_w} = 1 - \frac{\sum_{R_Y} w_j}{\sum_{R_Y} w_j + \sum_{NR_Y} w_j}$$

$R_Y$  – set of eligible units of respondents for variable  $Y$

$NR_Y$  – set of eligible units of non-respondents for variable  $Y$ , although the response is required

$w_j$  – unit weight

*Three basic ways for calculating the rate are:*

1. *Un-weighted rate:  $w_j = 1$*
2. *Rate weighted by the sample design weight:  $w_j = d_j$ , whereby  $d_j = \frac{1}{\pi_j}$  as a sample design weight is a reciprocal of the sample selection probability*
3. *Rate weighted by the unit significance:  $w_j = d_j x_j$ , whereby  $x_j$  is a value of the variable  $X$ .*

*The variable  $X$  which is selected subjectively represents the significance size of the unit. Its value should be known for all units. The value of  $X$  is the auxiliary information, often available in the sample frame. Examples are turnover for business entities or the total number of inhabitants in the municipality.*

## Primjer

Sljedeći primjer prikazuje istraživanje s osam poslovnih subjekata, na temelju čega treba izračunati neuteženu i uteženu stopu neodgovora na pojedinu stavku upitnika, na koju odgovaraju poduzeća s više od 10 zaposlenih. Status 1 predstavlja odgovor na varijablu; status 2 neodgovor na upitnik u cijelosti; status 3 jedinicu koja ne pripada ciljanoj populaciji (nevaljana jedinica); status 4 neodgovor na varijablu. Podatak o zaposlenima je pomoćna varijabla koja je poznata iz Statističkoga poslovnog registra prije istraživanja.

## Example

The following example represents a statistical survey with eight business entities, based on which should be calculated the un-weighted and weighted item non-response rate, on which are responding the business entities with 10 or more employees. Status 1 represents the response on item; status 2 non-response on the whole questionnaire; status 3 unit that does not belong to the target population (in-eligible unit); status 4 item non-response. The data on employees is the auxiliary variable known from the Statistical Business Register prior the survey.

### T 3. Izračun stope neodgovora na određenu varijablu na hipotetskom primjeru Calculation of item non-response rate on a hypothetical example

ID ID	Status za ključnu varijablu Status of key variable	Težina uzorka Sampling weight	Broj zaposlenih Number of employees
1	3	2,56	86
2	4	3,57	48
3	1	1,95	123
7	1	7,20	20
8	3	8,55	6
4	2	0,00	13
5	2	0,00	21
6	3	0,00	18

Izvor: hipotetski primjer

Source: hypothetical example

Neutežena stopa neodgovora na određenu varijablu:

$$NR_{Yr} = 1 - \frac{2}{3} = 1 - 0,667 = 0,3333 = 33,33\%$$

Neutežena stopa neodgovora je izračunana na način da se od jedan oduzeo omjer valjanih jedinica koje su odgovorile na varijablu (status 1) i valjanih jedinica od kojih se zahtijevalo da odgovore na varijablu (status 1 i status 4).

Un-weighted item non-response rate:

$$NR_{Yr} = 1 - \frac{2}{3} = 1 - 0,667 = 0,3333 = 33,33\%$$

Un-weighted item non-response rate is calculated in a way that are placed in the ratio all eligible respondents for the item (status 1) and all eligible units that are required to respond to a variable (status 1 and status 4), then subtracted from one.

Utežena stopa neodgovora na određenu varijablu s težinama izbora uzorka:

$$NR_Y r_w = 1 - \frac{1,95 + 7,20}{(1,95 + 7,20) + 3,57} =$$

$$= 1 - \frac{9,15}{12,72} = 1 - 0,7193 = 28,07\%$$

Utežena stopa neodgovora na određenu varijablu s korekcijom težina za pomoćnu varijablu:

$$NRr = 1 - \frac{1,95 * 123 + 7,20 * 20}{(1,95 * 123 + 7,20 * 20) + 3,57 * 48} =$$

$$= 1 - \frac{383,85}{555,21} = 1 - 0,6914 = 30,86\%$$

Za razliku od neutežene stope neodgovora na određenu varijablu kod računanja utežene stope ne uzima se u obzir broj jedinica u pojedinom skupu odgovora/neodgovora već se zbrajaju njihove težine. Dakle, ovdje je potrebno staviti u omjer sumu težina valjanih jedinica neodgovora na varijablu i svih valjanih jedinica od kojih se zahtijevao odgovor, umanjeno od jedan. Ako se koristi utežena stopa neodgovora na određenu varijablu s korekcijom težina, težina svake jedinice se množi sa korekcijskim faktorom (iznosom pomoćne varijable), a tek zatim se težine zbrajaju.

### Primjer

U idućem primjeru ukratko je analiziran neodgovor na pojedinu varijablu upitnika za ADS 2010. U tablici možemo vidjeti kratak pregled neodgovora na pojedinu varijablu upitnika nekih od osnovnih varijabli dohotka. Za tablicu su se, radi analize tog tipa neuzoračke greške, izračunali postoci kućanstava koja su primila određene iznose dohotka te postoci kućanstava s nedostajućim vrijednostima ili sa samo djelomičnim informacijama. Navedeni postoci računali su se na sljedeći način:

*Weighted item non-response rate by design weight:*

$$NR_Y r_w = 1 - \frac{1,95 + 7,20}{(1,95 + 7,20) + 3,57} =$$

$$= 1 - \frac{9,15}{12,72} = 1 - 0,7193 = 28,07\%$$

*Weighted item non-response rate where weights are corrected using auxiliary variable:*

$$NRr = 1 - \frac{1,95 * 123 + 7,20 * 20}{(1,95 * 123 + 7,20 * 20) + 3,57 * 48} =$$

$$= 1 - \frac{383,85}{555,21} = 1 - 0,6914 = 30,86\%$$

*Unlike the un-weighted item non-response rate, for the calculation of weighted item non-response rate we take in the account the sum of weights of units in a particular set of response/nonresponse and not just number of them. So, here it is necessary to subtract from one ratio of sum of all eligible non-respondents weights to a particular variable (item) and sum of all eligible units required to respond to this particular item. If you want to get a higher quality indicator, we can use corrected weights, which we get by multiplying design weights by the corresponding amount of an auxiliary variable, and then the weights are summed.*

### Example

*In the following example, the item non-response in the questionnaire is briefly analysed for SILC 2010. In the following table, a brief overview of the non-response to a particular item in the questionnaire regarding some basic income variables can be shown. In order to analyse this type of non-sampling error, percentages of households that have received specific income and percentages of households with missing data or with partial information, are calculated in the table. The percentages were calculated as follows:*

- postotak kućanstava koja su primila određene iznose: broj kućanstava (ili osoba) koja su primila barem neki iznos / ukupan broj kućanstava (osoba)
- postotak kućanstava s nedostajućim podacima ili samo s djelomičnim informacijama: broj kućanstava (ili osoba) koja tvrde da su primili neku vrijednost, ali nisu dali točan iznos ili su primili nešto, a dali su samo djelomične informacije o tome (ako iznosi nisu dani za sve komponente dohotka koje su primili) / ukupan broj kućanstava (osoba) koja su primili barem neki iznos

- *The percentage of households that received certain money amount: the number of households (or persons) which received at least some money amount / total number of households (persons)*
- *The percentage of households with missing data or only partial information: number of households (or persons) who claimed that have received money, but did not give the exact amount of money or are received something, and they gave only partial information (if the amounts of money are not available for all received income components) / total number of households (persons) who received some amount*

Varijable koje se odnose na različite komponente dohotka iz tablice sortirane su u padajućem redoslijedu prema postotku nedostajućih vrijednosti ili, drugim riječima, postotku neodgovora na pojedinu stavku upitnika.

*The variables that refer to the different income components are sorted in the table using dropdown order by percentage of missing values or, in other words, the percentage of non-response for specific item of the column of the questionnaire.*

**T 4. Neodgovor na određenu varijablu upitnika ADS 2010.**  
*Item non-response in the SILC 2010*

Vrsta dohotka <i>Income type</i>	Kućanstva koja su primila određene iznose <i>Households that received certain amount of money</i>		Kućanstva s potpunim informacijama o primljenim iznosima <i>Households with full information on received amount of money</i>		Kućanstva s nedostajućim vrijednostima ili s djelomičnim informacijama <i>Households with missing data or with the partial information</i>	
	N	%	N	%	N	%
Dohodak od iznajmljivanja imovine ili zemlje (neto) <i>Income from rental of property or land (net)</i>	157	4,24	68	43,31	89	56,69
Dohodak u naturi od nesamostalnog rada (neto) <i>Non-cash employee income (net)</i>	343	4,03	160	46,65	183	53,35
Neto novčani dohodak od samostalnog rada (neto) <i>Cash benefits or losses from self-employment (net)</i>	1 438	16,90	705	49,03	733	50,97
Ukupni raspoloživi dohodak kućanstva prije socijalnih transfera i mirovina <i>Total disposable household income before social transfers including old-age and survivor's benefits</i>	2 684	72,48	1 430	53,28	1 254	46,72
Ukupni raspoloživi dohodak kućanstava <i>Total disposable household income</i>	3 681	99,41	2 068	56,18	1 613	43,82
Ukupni bruto dohodak kućanstva <i>Total household gross income</i>	3 680	99,38	2 078	56,47	1 602	43,53
Neto novčani dohodak od nesamostalnog rada <i>Employee cash or near cash income</i>	2 877	33,80	2 038	70,84	839	29,16
Vrijednost dobara iz vlastite potrošnje <i>Value of goods produced for own consumption</i>	1 402	37,86	1 012	72,18	390	27,82
Starosna mirovina <i>Old-age benefits</i>	1 760	20,68	1 417	80,51	343	19,49
Obiteljska mirovina <i>Survivor's benefits</i>	630	7,40	545	86,51	85	13,49
Dohodak od porodičnih naknada, dječji doplataci, alimentacije, naknade za opremu novorođenčadi <i>Family/children related allowances</i>	492	13,29	436	88,62	56	11,38
Naknada za obrazovanje <i>Education-related allowances</i>	57	0,67	51	89,47	6	10,53
Naknada za nezaposlenost <i>Unemployment benefits</i>	241	2,83	223	92,53	18	7,47
Porez na luksuzne proizvode <i>Regular taxes on wealth</i>	1	0,03	1	100,00	0	0,00
Porez na dohodak i socijalni doprinosi <i>Tax on income and social contributions</i>	2 255	60,90	2 255	100,00	0	0,00
Socijalni doprinosi koje plaća poslodavac (iz i na plaću) <i>Employer's social insurance contribution</i>	2 877	33,80	2 877	100,00	0	0,00
Iznos uplata u privatni mirovinski fond <i>Contributions to individual private pension plans</i>	218	2,56	218	100,00	0	0,00
Bruto iznos mjesečnog dohotka za zaposlenike <i>Gross monthly earnings for employees</i>	2 861	33,62	2 861	100,00	0	0,00

Izvor: interni podaci DZS-a, Anкета o dohotku stanovništva 2010.  
 Source: internal CBS data, SILC 2010

### 1.1.5. Stopa imputacije (A7)

#### Definicija

Imputacija je proces koji se primjenjuje za zamjenu ne prikupljenih, nevažećih ili nedosljednih podataka koji se nisu uspjeli urediti. To isključuje praćenje ispitanika i ručni pregled, kao i ispravak podataka tj. određenog zapisa (ako je primjenjivo). Nakon imputacije svi podaci trebali bi biti uvjerljivi i interno konzistentni.

Razlikujemo dva načina izračuna stope imputacije, a time i dvije vrste stope. Neutežena stopa imputacije za varijable je omjer broja jedinica za koje je varijabla Y imputirana i ukupnog broja jedinica za koje je vrijednost varijable Y ostala nepromijenjena ili imputirana.

Utežena stopa pokazuje relativni doprinos statističkim podacima od imputiranih vrijednosti, obično za kvantitativne varijable. Za kvalitativne varijable, relativni doprinos temelji se na broju jedinica s imputiranim vrijednostima za kvalitativnu varijablu.

#### Formula za izračun

$$IR_{Yr_w} = \frac{\sum_{I_y} w_j y_j}{\sum_{I_y} w_j y_j + \sum_{K_y} w_j y_j}$$

$I_y$  – skup jedinica za koje je varijabla Y imputirana

$K_y$  – skup jedinica za koje je vrijednost varijable Y ostala nepromijenjena

$w_j$  – težina jedinice

#### Primjer

Na temelju podataka iz hipotetskog primjera izračunana je utežena i neutežena stopa imputacije.

### 1.1.5. Imputation rate (A7)

#### Definition

*Imputation is a process used for replacing missing, invalid or inconsistent data that could not be edited. It excludes monitoring of respondents and manual control, as well as correction on a record (if it is applicable). All the data after imputation should be credible and internally consistent.*

*Two ways for calculating the imputation rate and thereby two types of imputation rate are distinguished. The un-weighted imputation rate for variables is defined as ratio of units for which the variable Y is imputed and total number of units for which the value of the variable Y is remained unchanged or imputed.*

*The weighted imputation rate represents the relative contribution to the statistical data of the imputed values, usually for quantitative variables. For qualitative variables the relative contribution is based on the number of units with an imputed value on qualitative variable.*

#### Formula for calculation

$$IR_{Yr_w} = \frac{\sum_{I_y} w_j y_j}{\sum_{I_y} w_j y_j + \sum_{K_y} w_j y_j}$$

$I_y$  – set of units for which the variable Y is imputed

$K_y$  – set of units for which the value of the variable Y is remained unchanged

$w_j$  – unit weight

#### Example

*Based on the data from the hypothetical example, weighted and un-weighted imputation rates are calculated.*



## T5. Izračun stope imputacije na hipotetskom primjeru

*Calculation of the imputation rate on the hypothetical example*

ID <i>ID</i>	Vrijednost imputirana <i>Imputed value</i>	Vrijednost zadržana <i>Value retained</i>	Težina <i>Weight</i>
1	0	1	1,45
2	1	0	3,46
3	0	1	25,40
4	0	0	13,10
5	1	0	2,20
6	1	0	1,60
7	0	0	7,20

Izvor: hipotetski primjer

*Source: hypothetical example*

Neutežena stopa imputacije:

$$IR_{Yr} = \frac{1 + 1 + 1}{(1 + 1 + 1) + (1 + 1)} = \frac{3}{5} = 0,6 = 60\%$$

Stopa je izračunana kao omjer broja vrijednosti varijable koje su ispravljene ili imputirane i ukupnog broja vrijednosti varijable u upitniku. Iz izračuna zaključujemo kako je 60% originalnih vrijednosti zamijenjeno imputiranim vrijednostima.

Utežena stopa imputacije:

$$IR_{Yr_w} = \frac{3,46 + 2,20 + 1,60}{(1,45 + 25,40) + (3,46 + 2,20 + 1,60)}$$
$$= \frac{7,26}{34,11} = 0,2128 = 21,28\%$$

Radi dobivanja preciznijeg pokazatelja zamijene nedostajućih vrijednosti izračunana je i utežena stopa imputacije koja pokazuje kako je 21,28% originalnih vrijednosti zamijenjeno, odnosno toliko iznosi relativan doprinos statističkim podacima od procesa imputacije.

*Un-weighted imputation rate:*

$$IR_{Yr} = \frac{1 + 1 + 1}{(1 + 1 + 1) + (1 + 1)} = \frac{3}{5} = 0,6 = 60\%$$

*The rate is calculated as the ratio of the number of values of variable that have been corrected or imputed and the total number of values of variable in the questionnaire. It can be concluded that 60% of the original values of variable have been imputed.*

*Weighted imputation rate:*

$$IR_{Yr_w} = \frac{3,46 + 2,20 + 1,60}{(1,45 + 25,40) + (3,46 + 2,20 + 1,60)}$$
$$= \frac{7,26}{34,11} = 0,2128 = 21,28\%$$

*To obtain a more precise indicator of missing values replacement, weighted imputation rate was calculated. It shows that 21.28% of the original values were replaced, so is the relative contribution to the statistics of the imputation process.*

### 1.1.6. Prosječna veličina revizije podataka (A6)

#### Definicija

Prosječna veličina revizije podataka je prosjek revizija ključnih varijabli tijekom određenog razdoblja.

Podaci koji se objavljuju, a odnose se na određeno izvještajno razdoblje  $t$ , tijekom vremena se mogu mijenjati. Broj izvještajnih razdoblja za koje se podaci objavljuju označava se sa  $n$ , a broj objava tijekom vremena sa  $K$ . Revizija podataka nastaje između pojedinih objava vezano uz određeno izvještajno razdoblje te se prema tome definira kao razlika između kasnije i ranije procjene.

### 1.1.6. Data revision - average size (A6)

#### Definition

The data revision - average size is the average of revised key items during a specific period of time.

The publishing data which relate to the  $t$  reference period can be changed in time. The number of reference periods for which the data are published are marked with  $n$ , and the number of releases over time with  $K$ . The data revision occurs between specific releases related to the certain period and is therefore defined as the difference between the earlier and the later estimates.

Objava Release	Izveštajno razdoblje Reference period				
	1	...	$t$	...	$n$
1	$X_{11}$	...	$X_{1t}$	...	$X_{1n}$
...	...	...	...	...	...
$k$	$X_{k1}$	...	$X_{kt}$	...	$X_{kn}$
...	...	...	...	...	...
$K$	$X_{K1}$	...	$X_{Kt}$	...	$X_{KN}$

#### Formula za izračun

S obzirom na dvodimenzionalnu situaciju koja je opisana u definiciji, postoji nekoliko strategija za izračun pokazatelja: s predznakom, apsolutne ili relativne vrijednosti, za određene parove revidiranja kroz vrijeme ili preko niza revizija itd.

Glavni prijedlog je uzeti u obzir prosjek za određene objave za  $n$  izvještajnih razdoblja, odnosno MAR (prosječna apsolutna revizija).

#### Formula for calculation

Having in mind the two-dimensional situation described in the definition, there are several strategies for calculating the indicator: with sign, as an absolute or relative value, for specific pairs of revisions over time or over a sequence of revisions, etc.

The main proposal is to take into account the average of certain releases for  $n$  reference periods, i.e. MAR (Mean Absolute Revision).

MAR (prosječna apsolutna revizija):

$$MAR = \frac{1}{n} \sum_{t=1}^n |X_{L_t} - X_{P_t}|$$

$X_{L_t}$  – kasnija objava  
 $X_{P_t}$  – prijašnja objava  
 $n$  – broj revizija

Broj revizija ( $n$ ) se računa kao broj razlika u objavama u vremenskim serijama. U načelu, to je broj objava umanjen za 1, tj. ako imamo šest objava  $n = 6 - 1 = 5$ .

Reviziju podataka možemo računati i kao razliku prve i posljednje objave – AR (apsolutna revizija).

AR (apsolutna revizija):

$$AR = |X_{L_t} - X_{P_t}|$$

$X_{L_t}$  – kasnija objava  
 $X_{P_t}$  – prijašnja objava

Izračunavanje pokazatelja za tromjesečne procjene se ne preporučuje ako je  $n \geq 20$ , dok se za mjesečne procjene ne preporučuje ako je  $n \geq 30$ . Pokazatelj nije preporučljiv za godišnje procjene.

Prosječna relativna revizija (RMAR):

$$RMAR = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_{L_t} - X_{P_t}}{X_{P_t}} \right|$$

Prosječna apsolutna revizija uglavnom se primjenjuje kod indeksa, proporcija i ostalih relativnih podataka, dok se prosječna relativna revizija primjenjuje za podatke u apsolutnim iznosima (npr. broj zaposlenih).

*MAR (Mean Absolute Revision):*

$$MAR = \frac{1}{n} \sum_{t=1}^n |X_{L_t} - X_{P_t}|$$

$X_{L_t}$  – later release  
 $X_{P_t}$  – earlier release  
 $n$  – number of revisions

*Number of revisions ( $n$ ) is calculated as a number of differences between releases in time series. Usually it is a number of releases subtracted by one, i.e. for six releases it is equal  $n = 6 - 1 = 5$ .*

*The data revision can be calculated as the difference between previous and last release – AR (Absolute Revision).*

*AR (Absolute Revision):*

$$AR = |X_{L_t} - X_{P_t}|$$

$X_{L_t}$  – later release  
 $X_{P_t}$  – earlier release

*The calculation of indicator for quarterly estimates is not recommended if  $n \geq 20$ , while for the monthly estimates is not recommended for  $n \geq 30$ . The indicator is not recommended for annual estimates.*

*Relative Mean Absolute Revision (RMAR):*

$$RMAR = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_{L_t} - X_{P_t}}{X_{P_t}} \right|$$

*The Mean Absolute Revision is applicable for indices, proportions and other data in relative terms, while the Relative Mean Absolute Revision is applicable for the data in absolute terms (e.g. number of employees).*

## Primjer

Provodi se istraživanje XY u kojem se mjesečni indeks u Z izvještajnom razdoblju revidira u šest uzastopnih mjeseci te je izračunana revizija podataka. Napomena: treći stupac tablice pokazuje izračun razlika između objava. Budući da se radi o indeksima računa se prosječna apsolutna revizija (MAR).

## Example

A survey is carried out whereby the monthly index is revised during six months consecutively and the data revision is calculated. Remark: third column of the table represents the calculation of differences between releases. Since these are the indices, Mean Absolute Revision (MAR) is calculated.

### T6. Izračun prosječne veličine revizije podataka na hipotetskom primjeru

*Calculation of the data revision - average size on hypothetical example*

Objava <i>Release</i>	Mjesečni indeks <i>Monthly index</i>	Razlika <i>Difference</i>
1.	97,50	-
2.	98,10	0,60
3.	98,00	0,10
4.	98,05	0,05
5.	98,03	0,02
6.	97,70	0,33

Izvor: hipotetski primjer

Source: hypothetical example

$$MAR = \frac{0,60 + 0,10 + 0,05 + 0,02 + 0,33}{5} = \frac{1,10}{5} = 0,22$$

$$MAR = \frac{0,60 + 0,10 + 0,05 + 0,02 + 0,33}{5} = \frac{1,10}{5} = 0,22$$

Zaključujemo kako se mjesečni indeks u šest uzastopnih objava u prosjeku mijenjao za 0,22.

*It can be concluded that the monthly index in six consecutive releases has changed for 0.22 in average.*

## Primjer

U sljedećem primjeru prikazan je izračun prosječne relativne revizije podataka o ukupnom broju zaposlenih u Republici Hrvatskoj u 2012. prema podacima iz Mjesečnog istraživanja o zaposlenima i plaćama za pravne osobe (RAD-1). Budući da je pokazatelj izražen u apsolutnim vrijednostima računa se prosječna relativna revizija (RMAR). Pokazatelj je prvo izračunan za svaki pojedini mjesec (zadnji stupac u tablici), a zatim i za cijelu godinu kao prosjek svih prosječnih relativnih revizija po mjesecima.

## Example

*The following example shows the calculation of the Relative Mean Absolute Revision (RMAR) on the total number of employees in the Republic of Croatia in 2012, according to data from the Monthly survey on employment and wages in legal persons (RAD-1). Since the indicator is expressed in absolute values the calculated indicator is the average relative revision (RMAR). The indicator was first calculated for each month (the last column in the table), and then for the whole year as the average of all the average relative audits per month.*

## T 7. Izračun prosječne veličine revizije podataka za mjesečno istraživanje o zaposlenima i plaćama u pravnim osobama (RAD-1)

*Calculation of the data revision – average size for the monthly survey on the employment and earnings in legal entities (RAD-1)*

Mjesec <i>Month</i>	Privremeni rezultati <i>Preliminary results</i>	Konačni rezultati <i>Final results</i>	Prosječna relativna revizija <i>Relative Mean Absolute Revision</i>
I.	1 135 504	1 135 504	0,00%
II.	1 129 276	1 143 848	1,29%
III.	1 128 769	1 148 525	1,75%
IV.	1 135 156	1 155 127	1,76%
V.	1 143 678	1 163 888	1,77%
VI.	1 151 286	1 171 813	1,78%
VII.	1 152 918	1 173 578	1,79%
VIII.	1 148 210	1 168 788	1,79%
IX.	1 140 166	1 160 482	1,78%
X.	1 130 848	1 150 831	1,77%
XI.	1 120 899	1 140 537	1,75%
XII.	1 109 719	1 129 006	1,74%
2012.	-	-	1,58%

Izvor: interni podaci DZS-a, Mjesečno istraživanje o zaposlenima i plaćama za pravne osobe (RAD-1)  
*Source: internal CBS data, Monthly survey on the employment and earnings in legal entities (RAD-1)*

## 1.2. PRAVODOBNOST I VREMENSKA ODREĐENOST

Pravodobnost objavljivanja podataka odnosi se na duljinu vremena između refrentnog razdoblja na koje se podaci odnose i datuma objave podataka, dok se vremenska određenost odnosi na razdoblje između najavljenog datuma objave (primjerice u Kalendaru objavljivanja statističkih podataka) i stvarnog datuma objave podataka.

## 1.2. TIMELINESS AND PUNCTUALITY

*The timeliness of data publishing refers to the length of time between periods when the event of statistical phenomena and the date of publication of the data, while the punctuality refers to the time lag between the target date when data should have been delivered (e.g. date in the Calendar of Statistical Data Issues) and the actual date of the publication of data.*

### 1.2.1. Pravodobnost prvih rezultata (TP1)

#### Definicija

Pravodobnost prvih statističkih rezultata je dužina vremena između kraja razdoblja za koji su podaci izračunani i datuma prve objave tih podataka.

Izračunava se kao broj dana (ili tjedana ili mjeseci) od posljednjeg dana izvještajnog razdoblja do dana objavljivanja prvih rezultata.

Izražava se kao  $T + n$ , gdje je  $n$  broj dana ili mjeseci.

Za mjesečna, tromjesečna i polugodišnja istraživanja se izražava u ukupnom broju dana (ne samo radnim danima) koji su prošli, dok se za godišnja i višegodišnja istraživanja izražava u mjesecima.

#### Formula za izračun

$$T_1 = d_{frst} - d_{refp}$$

$d_{frst}$  – datum objavljivanja prvih rezultata  
 $d_{refp}$  – posljednji dan (datum) referentnog razdoblja statističkih podataka

#### Primjer

U tablici su ispisani datumi prvih objava mjesečnih podataka iz Trgovine na malo i izračun pravodobnosti prvih rezultata. Pokazatelj se izračunava tako da se za svaki mjesec izračuna razlika između datuma prve objave ( $d_{frst}$ ) i posljednjeg dana u referentnom mjesecu na koje se podaci odnose ( $d_{refp}$ ). Kako je riječ o mjesečnim objavama, izračun pokazatelja pravodobnosti iskazuje se u danima, u obliku  $T +$  broj dana.

### 1.2.1. Timeliness of first results (TP1)

#### Definition

The timeliness of the first statistical results is the length of time between the end of the period for which the statistical data are calculated and the date of first publication of this data.

It is calculated as a number of days (or weeks, or months) since the last day of the reference period until the date of dissemination the first results.

It is expressed as  $T + n$ , where  $n$  is a number of days/months.

It is generally expressed for monthly, quarterly and semi-annual statistical surveys in elapsed days (not just working days), while for annual and multi-annual statistical surveys it is expressed in months.

#### Formula for calculation

$$T_1 = d_{frst} - d_{refp}$$

$d_{frst}$  – date of dissemination of first results  
 $d_{refp}$  – last day (date) of the reference period of statistical data

#### Example

In the table below are listed the first release dates on monthly data from survey on the Retail trade and calculation of the timeliness of first results. The indicator is prepared in a way that for each month is calculated the difference between the date of the first release ( $d_{frst}$ ) and the last day of the reference month ( $d_{refp}$ ). As these are monthly releases, the indicator is expressed in days, in the form of  $T +$  number of days.

**T8. Izračun pravodobnosti prvih rezultata za mjesečno istraživanje Trgovine na malo**  
*Calculation of timeliness of first results for monthly survey on Retail trade*

Mjesec <i>Month</i>	Datum prve objave <i>Date of first release</i>	Pravodobnost prvih rezultata <i>Timeliness of first results</i>
I. 12.	6. 3. 2012.	T + 35
II. 12.	4. 4. 2012.	T + 35
III. 12.	4. 5. 2012.	T + 34
IV. 12.	5. 6. 2012.	T + 36
V. 12.	4. 7. 2012.	T + 34
VI. 12.	6. 8. 2012.	T + 37
VII. 12.	4. 9. 2012.	T + 35
VIII. 12.	4. 10. 2012.	T + 34
IX. 12.	6. 11. 2012.	T + 37
X. 12.	4. 12. 2012.	T + 34
XI. 12.	4. 1. 2013.	T + 35
XII. 12.	4. 2. 2013.	T + 35

Izvor: interni podaci DZS-a, Mjesečno istraživanje iz Trgovine na malo  
*Source: internal CBS data, Monthly survey on Retail trade*

**Primjer**

U tablici su ispisani datumi prvih objava Ankete o dohotku stanovništva i izračun pravodobnosti prvih rezultata. Datum prve objave je  $d_{frst}$ , a  $d_{refp}$  je posljednji dan referentnog razdoblja. Pokazatelj se izračunava kao razlika između navedena dva datuma (kao i u prethodnom primjeru). Budući da je riječ o godišnjim podacima, iskazuje se u mjesecima, u obliku  $T +$  broj mjeseci.

**Example**

*In the table below are listed the first release dates of the Statistics on Income and Living Conditions survey. The first release date is  $d_{frst}$ , and  $d_{refp}$  is the last date of the reference period. The indicator is calculated as a difference between these two dates (as it was done in the previous example). The annual data are presented in a form  $T +$  number of months.*

**T 9. Izračun pravodobnosti prvih rezultata za Anketu o dohotku stanovništva**  
*Calculation of timeliness of first results on the Statistics on Income and Living Conditions survey*

Godina <i>Year</i>	Datum prve objave <i>Date of first release</i>	Pravodobnost prvih rezultata <i>Timeliness of first results</i>
2010.	30. 11. 2011.	T + 11
2011.	26. 11. 2012.	T + 11

Izvor: interni podaci DZS-a, Anкета o dohotku stanovništva 2010  
*Source: internal CBS data, Statistics on Income and Living Conditions survey.*

## 1.2.2. Pravodobnost konačnih rezultata (TP2)

### Definicija

Pravodobnost konačnih statističkih rezultata je dužina vremena između kraja referentnog razdoblja na koji se pojava odnosi i konačne objave.

Izražava se kao broj dana ili mjeseci od posljednjeg dana izvještajnog razdoblja do dana objavljivanja konačnih rezultata (u obliku  $T + n$ , gdje je  $n$  broj dana ili mjeseci).

Isto kao i kod pravodobnosti prvih rezultata, izražava se u danima za mjesečna, tromjesečna i polugodišnja istraživanja, dok se za godišnja i višegodišnja istraživanja izražava u mjesecima.

### Formula za izračun

$$T_1 = d_{finl} - d_{refp}$$

$d_{finl}$  – datum objavljivanja konačnih rezultata

$d_{refp}$  – posljednji dan (datum) referentnog razdoblja statističkih podataka

### Primjer

U tablici su ispisani datumi objava priopćenja (konačni podaci) Trgovine na malo i izračun pravodobnosti konačnih rezultata. Pravodobnost konačnih rezultata izračunava se tako da se za svaki mjesec izračuna razlika između datuma konačne objave ( $d_{finl}$ ) i posljednjeg dana u referentnom mjesecu na koje se podaci odnose ( $d_{refp}$ ). Kako je riječ o mjesečnim objavama pokazatelj se izražava u danima, u obliku  $T +$  broj dana.

## 1.2.2. Timeliness of final results (TP2)

### Definition

The timeliness of final results is the length of time between the end of the reference period of statistical phenomena and the dissemination of final results.

It can be expressed as number of days or months since the last date of the reference period until the date of dissemination of final results (in the form  $T + n$ , where  $n$  is a number of days or months).

As with the timeliness of first results, it can be expressed as number of days for monthly, quarterly and semi-annual surveys, while for annual and multi-annual surveys it is expressed in months.

### Formula for calculation

$$T_1 = d_{finl} - d_{refp}$$

$d_{finl}$  – date of final results release

$d_{refp}$  – last day (date) of the reference period of statistical data

### Example

In the table are listed release dates (final data) of the Monthly Statistical Report of the Retail trade and the calculation of timeliness of final results. The timeliness of the final results is calculated for each month as a difference between the date of the final release ( $d_{fst}$ ), and the last date of the reference period ( $d_{refp}$ ). As here is spoken about monthly releases the indicator is calculated in days in a form  $T +$  number of days.



**T 10. Izračun pravodobnosti konačnih rezultata za mjesečno istraživanje  
Trgovine na malo**

*Calculation of timeliness of final results for monthly survey on Retail trade*

Mjesec Month	Datum objave Release date	Pravodobnost konačnih rezultata Timeliness of final results
I. 12.	15. 3. 2012.	T + 44
II. 12.	12. 4. 2012.	T + 43
III. 12.	10. 5. 2012.	T + 40
IV. 12.	13. 6. 2012.	T + 44
V. 12.	11. 7. 2012.	T + 41
VI. 12.	10. 8. 2012.	T + 41
VII. 12.	12. 9. 2012.	T + 43
VIII. 12.	12. 10. 2012.	T + 42
IX. 12.	13. 11. 2012.	T + 44
X. 12.	11. 12. 2012.	T + 41
XI. 12.	11. 1. 2013.	T + 42
XII. 12.	13. 2. 2013.	T + 44

Izvor: interni podaci DZS-a, Mjesečno istraživanje iz Trgovine na malo  
Source: internal CBS data, monthly survey on Retail trade

**Primjer**

U tablici su ispisani datumi konačnih objava Ankete o dohotku stanovništva i izračun pravodobnosti konačnih rezultata. Datum konačne objave je  $d_{finl}$ , a  $d_{refp}$  je posljednji dan referentnog razdoblja. Pokazatelj se izračunava kao razlika između navedena dva datuma (kao i u prethodnom primjeru). Budući da je riječ o godišnjim podacima, iskazuje se u mjesecima, u obliku  $T +$  broj mjeseci.

**Example**

*In the table below are listed the dates of the final releases of the survey on Statistics on Income and Living Conditions and the calculation of timeliness of final results. The release date of the final data is  $d_{finl}$ , and the  $d_{refp}$  is the last day of the reference period. The indicator is calculated as the difference between these two dates (as in the previous example). Since it is about annual data, it is expressed in months in the form of  $T +$  number of months.*

**T 11. Izračun pravodobnosti konačnih rezultata za Anketu o dohotku stanovništva**

*Calculation of timeliness of final results for Statistics on Income and Living Conditions*

Godina Year	Datum objave Dissemination date	Pravodobnost prvih rezultata Timeliness of first results
2010.	30. 11. 2011.	T + 11
2011.	1. 2. 2013.	T + 13

Izvor: interni podaci DZS-a, Anкета o dohotku stanovništva 2010.  
Source: internal CBS data, Statistics on Income and Living Conditions

### 1.2.3. Vremenska određenost (TP3)

#### Definicija

Vremenska određenost je razdoblje između isporuke ili objave podataka i ciljanog datuma za isporuku ili objavu podataka. Prema tome, pokazateljem se izračunava stopa skupa podataka koji su objavljeni točno na datum iz kalendara objava.

#### Formula za izračun

$$P_{3R} = \frac{m_{pc}}{m_{pc} + m_{up}}$$

$m_{pc}$  – broj statističkih rezultata koji su objavljeni na datum određen kalendarom objava ili prije (pravodobno)

$m_{up}$  – broj statističkih rezultata koji nisu objavljeni do datuma određenog kalendarom objava (nepravodobno)

#### Primjer

U sljedećoj tablici dan je primjer izračuna vremenske određenosti na hipotetskim podacima. Na temelju datuma iz Kalendara objavljivanja statističkih podataka i stvarnih datuma objave statističkih rezultata najprije treba izračunati broj statističkih rezultata koji su objavljeni pravodobno, odnosno nepravodobno. Nakon toga broj pravodobnih objava dijeli se s ukupnim brojem objava.

### 1.2.3. Punctuality (TP3)

#### Definition

*The punctuality is the time between the actual delivery or publication of the data and the target date when the data should have been delivered or published. Thus the indicator represents the rate at which CBS data is published in accordance with the Calendar of Statistical Data Issues.*

#### Formula for calculation

$$P_{3R} = \frac{m_{pc}}{m_{pc} + m_{up}}$$

$m_{pc}$  – number of statistical results that are disseminated on the date defined by the Calendar of Statistical Data Issues or earlier (on time)

$m_{up}$  – number of statistical results that are not disseminated until the date defined by the Calendar of Statistical Data Issues (not on time)

#### Example

*In the table below is presented an example of calculating punctuality on the hypothetical data. Based on the Calendar of Statistical Data Issues and the real dissemination dates of statistical results it is necessary to calculate the number of statistical results that are disseminated on time and not on time. After that the number of on time disseminations is divided by the total number of disseminations.*

**T 12. Izračun vremenske određenosti na hipotetskom primjeru**  
*Calculation of punctuality on a hypothetical example*

Tromjesečje <i>Quarter</i>	Datum određen Kalendarom objavljivanja statističkih podataka <i>Date defined by the Calendar of Statistical Data Issues</i>	Datum objave rezultata <i>Dissemination date of results</i>	Objava na datum određen Kalendarom ili prije <i>Published on the date defined by the Calendar or earlier</i>
1. 2010.	25. 04. 2010.	25. 04. 2010.	1
2. 2010.	25. 07. 2010.	23. 07. 2010.	1
3. 2010.	25. 10. 2010.	25. 10. 2010.	1
4. 2010.	25. 01. 2011.	28. 01. 2011.	0
1. 2011.	24. 04. 2011.	27. 04. 2011.	0
2. 2011.	24. 07. 2011.	24. 07. 2011.	1
3. 2011.	24. 10. 2011.	22. 10. 2011.	1
4. 2011.	24. 01. 2012.	26. 01. 2012.	0
1. 2012.	28. 04. 2012.	28. 04. 2012.	1
2. 2012.	28. 07. 2012.	26. 07. 2012.	1
3. 2012.	28. 10. 2012.	31. 10. 2012.	0
4. 2012.	28. 01. 2013.	29. 01. 2013.	0

Izvor: hipotetski primjer  
*Source: hypothetical example*

$$P_{3R} = \frac{m_{pc}}{m_{pc} + m_{up}} = \frac{7}{7 + 5} = 0,583 = 58,3\%$$

$$P_{3R} = \frac{m_{pc}}{m_{pc} + m_{up}} = \frac{7}{7 + 5} = 0,583 = 58,3\%$$

### 1.3. USPOREDIVOST

Usporedivost kao jedna od temeljnih dimenzija kvalitete odnosi se na potrebu da dobiveni podaci i informacije budu usporedivi u vremenu, između zemljopisnih područja te između različitih domena.

### 1.3. COMPARABILITY

*The comparability as one of the basic quality dimensions is referring to the need that received data and information should be comparable over time, between geographical areas, and between different domains.*

### 1.3.1. Dužina usporedivih vremenskih serija (CC2)

#### Definicija

Dužina usporedivih vremenskih serija je broj izvještajnih razdoblja u vremenskoj seriji od posljednjeg prekida. Prekidi u vremenskim serijama mogu nastati zbog promjena u definicijama parametara koji se procjenjuju (npr. varijable ili populacija) ili metodologija koje se primjenjuju za procjene.

#### Formula za izračun

$$CC_1 = J_{last} - J_{first} + 1$$

$J_{first}$  – redni broj prvoga izvještajnog razdoblja s usporedivim statističkim podacima

$J_{last}$  – redni broj posljednjega izvještajnog razdoblja s diseminiranim statističkim podacima

Ovaj pokazatelj nije dostatan za ocjenu usporedivosti kao dimenzije kvalitete budući da postoji i drugih razloga zašto podaci nisu usporedivi iako nema očiglednog prekida u vremenskoj seriji podataka.

#### Primjer

Idući primjer prikazuje tromjesečnu vremensku seriju prometa u trgovini na veliko (NKD 469). Kako je početkom 2009. promijenjen NKD te je uz to promijenjen i uzorak izvještajnih jedinica, nije bilo moguće uspostaviti odgovarajuću vezu sa starim podacima te je došlo do prekida u vremenskoj seriji.

Na podacima iz sljedeće tablice izračunana je dužina usporedivih vremenskih serija. Kako vremenska serija ima prekid u određenom dijelu unutar serije, a podaci su se i dalje nastavili objavljivati moguće je izračunati dvije vrijednosti pokazatelja. Budući je riječ o tromjesečnom istraživanju, mjerna jedinica duljine je jedno tromjesečje.

### 1.3.1. Length of comparable time series (CC2)

#### Definition

The length of comparable time series is a number of reference periods in the time series since the last break. The breaks in the time series can be result of changes in the estimation parameter definitions (e.g. variable or population) or changes in the methodology implemented for estimations.

#### Formula for calculation

$$CC_1 = J_{last} - J_{first} + 1$$

$J_{first}$  – ordinal number of the first reference period with a comparable statistical data

$J_{last}$  – ordinal number of the last reference period of disseminated statistical data

This is quite a poor indicator for estimating level of comparability as one of the main quality dimensions as there may be many reasons why data are not comparable even though there are no clear breaks in the series.

#### Example

The following example represents the quarterly time series of wholesale trade (NKD 469). As at the beginning of 2009 the National Classification of Activities (NKD) was changed and consequently the sample of reporting units had changed as well, therefore it was not possible to make a connection between new and old data due to the break in time series.

The length of comparable time series has been calculated for the data from the table below. As the time series has a time break and the data is still continued to publish after this break it is possible to calculate two values of indicator. As this is a case of quarterly statistical survey, the measurement unit is one quarter.

**T 13. Indeksi prometa u trgovini na veliko raznovrsnim proizvodima (NKD 469)**  
*Indices of turnover in non-specialised wholesale trade (NKD 469)*

Tromjesečje <i>Quarter</i>	Indeksi 2005. = 100 <i>Indices 2005 = 100</i>	Broj izvještajnog razdoblja <i>Number of reference period</i>
1. 2005.	83,8	1
2. 2005.	106,2	2
3. 2005.	107,0	3
4. 2005.	103,0	4
1. 2006.	86,2	5
2. 2006.	109,5	6
3. 2006.	109,0	7
4. 2006.	115,0	8
1. 2007.	93,3	9
2. 2007.	117,4	10
3. 2007.	117,6	11
4. 2007.	117,1	12
1. 2008.	95,3	13
2. 2008.	118,1	14
3. 2008.	115,9	15
4. 2008.	110,5	16
1. 2009.	PREKID / <i>BREAK</i>	PREKID / <i>BREAK</i>
2. 2009.	102,8	1
3. 2009.	94,6	2
4. 2009.	85,3	3
1. 2010.	69,3	4
2. 2010.	90,5	5
3. 2010.	86,4	6
4. 2010.	85,4	7
1. 2011.	72,3	8
2. 2011.	94,3	9
3. 2011.	89,5	10
4. 2011.	85,6	11
1. 2012.	67,0	12
2. 2012.	79,9	13
3. 2012.	80,5	14
4. 2012.	75,6	15

Izvor: interni podaci DZS-a, indeksi kretanja trgovine na veliko  
*Source: internal CBS data, indices on wholesale movement*

$$CC_1 = J_{last} - J_{first} + 1 = 16 - 1 + 1 = 16$$

$$CC_1 = J_{last} - J_{first} + 1 = 16 - 1 + 1 = 16$$

$$CC_2 = J_{last} - J_{first} + 1 = 15 - 1 + 1 = 15$$

$$CC_2 = J_{last} - J_{first} + 1 = 15 - 1 + 1 = 15$$

## 2. POMOĆNI POKAZATELJI

Pomoćni pokazatelji jesu pokazatelji za koje ne postoji obaveza izračunavanja i slanja podataka u Eurostat. Oni su dodatna mjerila, bitna za ocjenu kvalitete statističkih podataka isključivo na nacionalnoj razini, te se stoga izračunavaju na dobrovoljnoj osnovi.

Pomoćni pokazatelji kvalitete, s oznakama pokazatelja, strukturirani su na sljedeći način:

1. Dimenzija kvalitete: Relevantnost
  - a. Stopa potpunosti podataka – R1
2. Dimenzija kvalitete: Točnost
  - a. Pristranost u procesu selekcije – A8
  - b. Stopa uređivanja podataka – A9
  - c. Stopa jedinica koje nisu prošle logičko računsku kontrolu – A10
  - d. Stopa učinkovitosti LRK – A11
  - e. Stopa nepravilnog razvrstavanja – A12
3. Dimenzija kvalitete: Smislenost
  - a. Smislenost između različitih izvora podataka – CH1

### 2.1. RELEVANTNOST

Relevantnost je stupanj u kojem statistički podaci zadovoljavaju sadašnje i potencijalne potrebe korisnika. To znači jesu li sve statistike koje se proizvode doista potrebne i jesu li primijenjene mjere (definicije, klasifikacije, itd.) koje odražavaju potrebe korisnika.

## 2. AUXILIARY INDICATORS

*The auxiliary indicators are indicators for which there is no obligation for them to be calculated and sent to Eurostat. They are considered as additional measurements, important for estimation of quality of statistical data only on national level, and therefore are calculated on the voluntary basis.*

*The auxiliary quality indicators, with the indicator codes are structured as follows:*

1. *Quality dimension: Relevance*
  - a. *Data completeness rate – R1*
2. *Quality dimension: Accuracy*
  - a. *Bias due to selection process – A8*
  - b. *Editing rate – A9*
  - c. *Edit failure rate A10*
  - d. *Efficiency rate LCC – A11*
  - e. *Rate of misclassification – A12*
3. *Quality dimension: Coherence*
  - a. *Coherence between different sources – CH1*

### 2.1. RELEVANCE

*The relevance is the level in which the statistical data fulfil current known and unmet user needs. That means whether all produced statistical data are really needed and whether the measures reflecting user needs are implemented (definitions, classifications, etc.).*

### 2.1.1. Stopa potpunosti podataka (R1)

#### Definicija

Stopa potpunosti podataka jedan je od pomoćnih pokazatelja čiji izračun može biti zahtjevan od strane Eurostata, no uglavnom se računa za potrebe nacionalnih statistika.

To je udio broja prikupljenih podataka u ukupnom broju traženih podataka u određenom razdoblju.

#### Formula za izračun

$$R1_U = \frac{\#A_d^{rel}}{\#D^{rel}}$$

$D^{rel}$  – skup relevantnih podataka (sve statistike koje su obvezatne prema EU propisima)

$A_d^{rel}$  – odgovarajući podskup dostupnih podataka (koliko je podataka izračunano od traženih podataka)

# – broj elemenata u skupu/podskupu

#### Primjer

Prema priopćenju Ankete o inovacijama u hrvatskim poduzećima, DZS objavljuje sljedeća četiri pokazatelja: poduzeća prema inovativnosti, djelatnosti i veličini, poduzeća prema vrsti inovacija, djelatnosti i veličini, poduzeća prema važnosti tržišta, inovatori proizvoda i procesa prema vrsti inovacijske aktivnosti.

Eurostat objavljuje sljedećih šesnaest pokazatelja: osnovne informacije o poduzećima, poduzeća prema tipu inovacija, inovacije proizvoda i procesa, inovacijske aktivnosti i troškovi, javno financiranje inovacijskih aktivnosti, izvori informacija za inovacije proizvoda i procesa, vrste partnerske suradnje u inovacijama proizvoda i procesa, ciljevi inovacija proizvoda i

### 2.1.1. Data completeness rate

#### Definition

The data completeness rate is one of auxiliary indicators that can be required by Eurostat, but mainly it is calculated for purposes of each NSI.

It is the ratio of the number of data cells provided to the number of data cells required in particular period of time.

#### Formula for calculation

$$R1_U = \frac{\#A_d^{rel}}{\#D^{rel}}$$

$D^{rel}$  – set of relevant data (all statistical data that are obligatory according to the EU legislation)

$A_d^{rel}$  – appropriate subset of available data (how many data are calculated in comparison with the required data)

# – number of elements in the set/subset

#### Example

According to the First Release on Innovation activities in the Croatian enterprises, CBS disseminates the following four indicators: enterprises according to the innovation, economic activity and size of the enterprise, according to the market importance, innovators of the products and processes according to the type of innovation activity.

Eurostat disseminates sixteen following indicators: basic information on enterprises according to the type of innovation, innovation on products and processes, innovation activities and expenses, public financing of innovation activities, information sources for innovation of products and processes, types of partners cooperation in the innovation of products and processes,

procesa, otežavajuće inovacijske aktivnosti, organizacijske i marketinške inovacije, oblici uvođenja novih organizacijskih metoda, ciljevi organizacijskih inovacija, oblici uvođenja novih marketinških metoda, ciljevi marketinških inovacija, unutarne i vanjske vještine u poduzećima, metode stimulacije novih ideja i kreativnosti.

U skladu s navedenim pokazateljima stopa potpunosti podataka računa se tako da podijelimo ukupan broj pokazatelja koje objavljuje DZS s ukupnim brojem pokazatelja koji su obvezatni prema regulativama EU-a i koje objavljuje Eurostat.

$$R1_U = \frac{4}{16} = 0,25 = 25\%$$

## 2.2. TOČNOST

Kao što je već ranije spomenuto dimenzija kvalitete koja predstavlja točnost prikazuje razlike između procijenjenih i stvarnih podataka populacije. Kao pomoćni pokazatelj kojim se mjeri kvaliteta statističkih istraživanja može se izračunavati pristranost u procesu selekcije.

### 2.2.1. Pristranost u procesu selekcije (A8)

#### Definicija

Pristranost u procesu selekcije je pokazatelj koji se uglavnom povezuje s namjerno izabranim uzorkom u kojem se određen broj jedinica iz ciljane populacije namjerno ispušta u biranju uzorka (tzv. uzorak na ograničenoj populaciji kao što su npr. poduzeća s 10 i više zaposlenih).

Upravo zbog pristranosti važno je imati na umu kako uzoračka greška nije reprezentativan pokazatelj kvalitete procjena populacije na temelju namjernog uzorka.

*goals of products and process innovation, making difficult innovative activities, organisational and marketing innovation, types of implementing new organisational methods, goals of organisational innovations, types of introducing new marketing methods, goals of marketing innovation, internal and external skills in enterprises, methods of stimulating new ideas and creativity.*

*In accordance with the mentioned indicators the data completeness rate is calculated by dividing total number of CBS disseminated indicators with the total number of obligatory indicators regulated by EU and published by Eurostat.*

$$R1_U = \frac{4}{16} = 0,25 = 25\%$$

## 2.2. ACCURACY

*As mentioned earlier quality dimension of accuracy represents the difference between estimated and the actual population data. As an auxilliary indicator which measures the quality of statistical surveys, the bias due to the selection process can be calculated.*

### 2.2.1. Bias due to selection process (A8)

#### Definition

*Bias due to selection process is the indicator which usually relates to the purposive sample where one part of units in the target population was deliberately omitted in the process of sample selection (e.g. cut-off sample such as of business with 10 or more employees).*

*Because of bias, it is important to note that sampling error is not representative indicator of the quality of population estimates based on a purposive sample. Therefore, in such*



Zato se umjesto uzoračke greške u takvim slučajevima kao pokazatelj kvalitete primjenjuje pristranost u procesu selekcije.

S obzirom na to da je pristranost u procesu selekcije teško procijeniti, obično se računa korištenjem pomoćnih varijabli, primjerice prihoda poduzeća, a moguće je i upotrebom broja zaposlenih.

### Formula za izračun

$$\beta = \left| \frac{\hat{\theta} - \theta}{\hat{\theta}} \right|$$

- $\beta$  – pristranost u procesu selekcije
- $\theta$  – stvarna vrijednost parametra populacije
- $\hat{\theta}$  – procijenjena vrijednost parametra uzorka na ograničenoj populaciji

### Primjer

Prema Anketi o slobodnim radnim mjestima, 2012. raspoložemo podatkom o visini ukupnih prihoda poduzeća iz ciljane populacije u iznosu od  $6,607 * 10^{13}$  kuna. Nakon izdvajanja dijela populacije s manje od 20 zaposlenih, prihod poduzeća iznosio je  $5,337 * 10^{13}$  kuna. Ukupni prihod poduzeća iz cijele populacije stvarna je vrijednost parametra populacije, dok je ukupni prihod poduzeća s više od 20 zaposlenih procijenjena vrijednost parametra uzorka na ograničenoj populaciji. Na temelju navedenih iznosa moguće je izračunati pristranost u procesu selekcije.

$$\beta = \left| \frac{6,607 * 10^{13} - 5,337 * 10^{13}}{6,607 * 10^{13}} \right| = 0,1922$$

= 19,22%

*cases instead of the sampling error as a quality indicator the bias due to selection process is applied.*

*Due to the fact that the bias due to selection process is difficult to estimate, usually is calculated using auxiliary variables, i.e. revenue of the enterprise, or possibly an employment variable.*

### Formula for calculation

$$\beta = \left| \frac{\theta - \hat{\theta}}{\hat{\theta}} \right|$$

- $\beta$  – bias due to selection process
- $\theta$  – actual value of the population parameter
- $\hat{\theta}$  – estimated parameter value of cut-off sample

### Example

*According to the Job vacancy statistics survey for 2012, total revenues of enterprises in the target population amounted  $6,607 * 10^{13}$  HRK. After extracting one part of the population with less than 20 employees, total revenues of enterprises amounted  $5,337 * 10^{13}$  HRK. Total revenues of enterprises in the population are the actual value of the population parameter while the total revenue of enterprises with more than 20 employees is the estimated value of the cut-off sample parameter. Based on the above mentioned values it is possible to calculate the bias due to selection process.*

$$\beta = \left| \frac{6,607 * 10^{13} - 5,337 * 10^{13}}{6,607 * 10^{13}} \right| = 0,1922$$

= 19,22%

## 2.2.2. Stopa uređivanja podataka (A9)

### Definicija

Stopa uređivanja podataka je definirana za određene ključne varijable kao odnos između broja jedinica za koje su originalne vrijednosti ispravljene nakon provedene logičko-računske kontrole (LRK) u odnosu na ukupan broj jedinica.

Jednostavnije rečeno, to je omjer broja ispravljenih podataka (bilo ponovnim telefonskim pozivima bilo logičkim ispravicima koje rezultiraju iz LRK) i ukupnog broja podataka s kojima raspolažemo, tj. onih podataka koji su prošli provjere.

### Formula za izračun

Neutežena stopa uređivanja podataka:

$$T6 = \frac{\sum_{i=1}^p E_i}{p}$$

- $E_i$  – indikatorska varijabla promjene vrijednosti ključne varijable  $Y$
- $p$  – ukupan broj vrijednosti kojima raspolažemo za ključnu varijablu  $Y$

Indikatorska varijabla  $E_i$  prikazuje promjene vrijednosti ključne varijable  $Y$  i poprima vrijednosti 0 ili 1. Ako je podatak ispravljen iznosi 1, a ako zadržan kao ispravan iznosi 0.

Utežena stopa uređivanja podataka:

$$T6_w = \frac{\sum_{i=1}^p E_i * w_i * y_i}{\sum_{i=1}^p w_i * y_i}$$

- $E_i$  – indikatorska varijabla promjene vrijednosti ključne varijable  $Y$
- $p$  – ukupan broj vrijednosti kojima raspolažemo za ključnu varijablu  $Y$
- $w_i$  – konačna težina varijable  $Y$
- $y_i$  – vrijednost ključne varijable  $Y$

## 2.2.2. Editing rate (A9)

### Definition

*Editing rate is defined for certain key variables as the the ratio between the number of units for which the original values were corrected after editing in relation to the total number of units.*

*Simply put it is the ratio of the corrected data (either by repeated phone calls or by logical adjustments resulting from the editing and the total number of available data, i.e. those who have passed the control).*

### Formula for calculation

*Un-weighted editing rate:*

$$T6 = \frac{\sum_{i=1}^p E_i}{p}$$

- $E_i$  – indicator of value change for key variable  $Y$
- $p$  – total number of available values for key variable  $Y$

*$E_i$  as an indicator variable of key variable changes takes the value 0 or 1. If the data is corrected it equals to 1 and if the data is retained as correct it equals to 0.*

*Weighted editing rate:*

$$T6_w = \frac{\sum_{i=1}^p E_i * w_i * y_i}{\sum_{i=1}^p w_i * y_i}$$

- $E_i$  – indicator of value change for key variable  $Y$
- $p$  – total number of available values for key variable  $Y$
- $w_i$  – final weight of variable  $Y$
- $y_i$  – value of key variable  $Y$

## Primjer

Na hipotetskim podacima s brojem ispravljenih vrijednosti varijable *Y* izračunana je neutežena i utežena stopa uređivanja podataka. Radi jednostavnosti izračuna dana je suma uteženih vrijednosti varijable *Y*.

## Example

Based on the hypothetical data with the number of corrected values of variable *Y* is calculated un-weighted and weighted editing rate. For simplicity of calculation the weighted sum of variable *Y* has given.

### T 14. Izračun neutežene i utežene stope uređivanja podataka na hipotetskom primjeru Calculation of un-weighted and weighted editing rate on the hypothetical example

	Broj jedinica Number of units	Uteženi zbroj varijable <i>Y</i> Weighted sum of variable <i>Y</i>
Broj podataka za varijablu <i>Y</i> Number of data for variable <i>Y</i>	1 000	25 000
Ispravljeni podaci varijable <i>Y</i> Corrected data of the variable <i>Y</i>	380	7 000

Izvor: hipotetski primjer

Source: hypothetical example

Neutežena stopa uređivanja podataka:

$$T6 = \frac{380}{1000} = 0,38$$

Un-weighted editing rate:

$$T6 = \frac{380}{1000} = 0,38$$

Utežena stopa uređivanja podataka:

$$T6_w = \frac{7000}{25000} = 0,28$$

Weighted editing rate:

$$T6_w = \frac{7000}{25000} = 0,28$$

### 2.2.3. Stopa jedinica koje nisu prošle logičko računsku kontrolu (A10)

#### Definicija

Stopa jedinica koje nisu prošle logičko-računsku kontrolu (LRK) pokazuje broj jedinica koje imaju barem jednu LRK grešku. Drugim riječima, to je udio jedinica koje su pale barem na jednoj logičko-računskoj kontroli u ukupnom broju jedinica koje su dale odgovor na upitnik.

Razlika u odnosu na prethodnu stopu je u tome što se ova stopa računa na razini jedinice, dok se prethodna stopa računa na razini varijable. Dakle, kod prethodne stope, ako jedna jedinica ima grešku na 5 varijabli, stopa se računa za svaku od tih 5 varijabli, dok se kod stope jedinica koje nisu prošle LRK računa stopa samo za tu jedinicu, a ne za sve varijable koje su kod te jedinice pogrešne.

### 2.2.3 Edit failure rate (A10)

#### Definition

Edit failure rate shows the number of units that have at least one error in the editing process. In other words, this is the proportion of units that have failed on at least one algorithm in the editing process compared to the total number of respondents of questionnaire.

The difference compared to the previous rate is that this rate is calculated on the level of units, and previous rate is calculated on the level of each variable. That means in edit failure rate, if on one analysis unit are found five errors in variables, it is still counted as a single unit with an error, while for the editing rate all five entered data would be counted in the total number of errors for calculating this indicator.

Budući da kod pojedinih istraživanja (anketa) zbog iscrpnog broja varijabli koje prolaze LRK postoji i veliki broj kontrola koje se moraju proći, odabiru se najvažnije varijable od interesa te se samo za njih računa stopa jedinica koje nisu prošle logičko računsku kontrolu – broj jedinica koje nisu prošle kontrolu za pojedinu varijablu u odnosu na ukupan broj jedinica koje su dale odgovor za tu varijablu.

Ipak, preporučuje se da se taj pokazatelj računa zasebno i za različite vrste kontrola koje rezultiraju različitim greškama obrade, a ne samo kontrole općenito, kako bismo dobili informaciju na kojim kontrolama najčešće padaju jedinice.

### Formula za izračun

$$T7 = \frac{\sum_{i=1}^m U_i}{m}$$

$U_i$  – indikatorska varijabla jedinica s greškama na nekoj varijabli od interesa

$m$  – ukupan broj jedinica analize koje su odgovorile na upitnik

$U_i$  je indikatorska varijabla kojom se označavaju jedinice analize koje imaju neke greške na nekoj varijabli od interesa. Ako jedinica analize ima barem jednu grešku na nekoj od prikupljenih varijabli tada  $U_i$  ima vrijednost 1. Ako jedinica ima sve ispravne i zadržane podatke tada je vrijednost 0.

### Primjer

Idući primjer pokazuje izračun stope jedinica koje nisu prošle logičko-računsku kontrolu na primjeru Ankete o dohotku stanovništva (ADS) 2013. Na upitnik kućanstava za ADS 2013 odgovorilo je 3 765 kućanstava. Tijekom prve eliminacijske logičko-računske kontrole uočeno je da se na 324 kućanstva pojavila barem jedna greška na nekoj od varijabli od interesa, tj. pitanju iz upitnika.

*Because of the comprehensive number of variables and thus the number of editing algorithms in some surveys, the most important variables of interest are selected for which the edit failure rate is going to be calculated. In that sense edit failure rate represents the number of units that have not pass to the editing process for certain variable comparing to the total number of respondents for this variable.*

*In order to receive information about the quality of different control types in the editing process which results with failures, it is recommended to calculate this indicator also for each control type and not just for editing process in general.*

### Formula for calculation

$$T7 = \frac{\sum_{i=1}^m U_i}{m}$$

$U_i$  – indicator of editing failures for certain variable of interest

$m$  – total number of respondents to the questionnaire

*$U_i$  is an indicator that represents the number of units of analysis that have some errors on some variables of interest. It has a value of 1 if a unit of analysis has at least one mistake on one of the collected variables. If all the data are correct and their original values are retained then  $U_i$  is equal to 0.*

### Example

*The next example shows the calculation of weighted and un-weighted edit failure rate on the example of Survey on Income and Living Conditions 2013. For SILC 2013 questionnaire 3765 households responded. During the first editing process it was determined that at least one error was detected on 324 households on some variable of interest, i.e. question from the questionnaire.*

Budući da u ADS-u postoji iscrpan broj kontrola odabiru se najvažnije varijable od interesa. Za potrebe ovog primjera uzeti su podaci o ukupnome godišnjem iznosu primljenih naknada na osnovi nezaposlenosti. Logičko-računska kontrola za tu varijablu pokazala je da su od ukupno 324 jedinica koje su odgovorile na tu varijablu 32 pale na kontroli. Osim za navedenu varijablu od interesa, vrijednost pokazatelja izračunana i zasebno za različite vrste kontrola procesa uređivanja podataka.

Stopa jedinica koje nisu prošle LRK na razini kontrole:

$$T7 = \frac{324}{3765} = 0,086 = 8,6\%$$

Stopa jedinica koje nisu prošle LRK za varijablu od interesa:

$$T7 = \frac{32}{324} = 0.0984 = 9.84\%$$

#### 2.2.4. Stopa učinkovitosti logičko računске kontrole (A11)

##### Definicija

Stopa učinkovitosti logičko računске kontrole (LRK) je udio stvarnih grešaka u podacima u odnosu na ukupan broj signaliziranih grešaka proizašlih iz LRK.

Razlika od stope uređivanja podataka je što je kod stope uređivanja u nazivniku ukupan broj prikupljenih podataka, a kod stope ispravljenih grešaka je u nazivniku samo ukupan broj pogrešnih podataka kao rezultat LRK ili neke druge kontrole. Brojnik kod obje stope pokazuje pogrešne podatke koji su nakon LRK ispravljeni.

*As in SILC exists a comprehensive number of controls, the most important variables of interests are selected. For the example purposes the data on the total annual unemployment benefits are selected as a key variable of interest. The result of the editing process has shown that from 328 respondents on this variable 32 of them failed in the editing process. Except for the above mentioned variable of interest the indicator has been also calculated for different control types of editing process.*

*Edit failure rate for the editing process:*

$$T7 = \frac{324}{3765} = 0,086 = 8,6\%$$

*Edit failure rate for the variable of interest:*

$$T7 = \frac{32}{328} = 0.0984 = 9.84\%$$

#### 2.2.4. Hit rate (A11)

##### Definition

*The hit rate is the proportion of actual errors in the data compared to the total number of determined errors arising from the editing process.*

*The difference from the editing rate is that the editing rate in the denominator has the total number of collected data, and by the editing rate in the denominator is only the total number of incorrect data arising from the editing process or other control. In the numerator of both rates are false data which are corrected after the editing process.*

## Formula za izračun

$$T8 = \frac{\sum_{i=1}^n P_i}{n}$$

$P_i$  – indikatorska varijabla koja pokazuje da li je signalizirana greška na podacima stvarno greška

$n$  – ukupan broj podataka za koje je signalizirana greška

## Primjer

Logičko-računska kontrola za Istraživanje o inovacijskim aktivnostima poduzeća (CIS 2010.) pokazala je da je na odgovorenih 3 389 poduzeća (od uzorka veličine 4 500 poduzeća) ukupno 11 234 podataka bilo evidentirano kao potencijalno pogrešnima. Naknadnom kontrolom ustanovljeno je da je njih 5 051 stvarno pogrešno te ih je upravo toliko i ispravljeno. Stopa učinkovitosti LRK u tom je slučaju jednaka:

$$T8 = \frac{5051}{11234} = 0,4496 = 44,96\%$$

### 2.2.5. Stopa nepravilnog razvrstavanja (A12)

#### Definicija

Stopa nepravilnog razvrstavanja je broj jedinica za koje je informacija prikupljena istraživanjem pokazala da su jedinice bile pogrešno razvrstane u okviru.

Najjednostavnije rečeno, budući da se većina anketa u DZS-u provodi na stratificiranom uzorku, stopa nepravilnog razvrstavanja pokazuje udio jedinica koje su promijenile stratum iz kojeg su izabrane, ili su promijenile djelatnost (pogrešna klasifikacija prema NKD-u definirana u okviru), ili je na primjer u istraživanju iz poljoprivrede određena pogrešna veličina farme (ako su izabrane iz okvira kao male farme, a na terenu se pokazalo da su velike).

## Formula for calculation

$$T8 = \frac{\sum_{i=1}^n P_i}{n}$$

$P_i$  – indicator which shows whether the determined error in data is really an error

$n$  – total number of data with potential errors

## Example

The result of the editing process for the Community Innovation Survey (CIS 2010) showed that 3 389 enterprises that responded to the questionnaire (from sample size of 4 500 enterprises) in total 11 234 data were to be potentially wrong. After an additional control it was found that 5051 of them were really false and therefore were corrected. In this case the hit rate is equal to:

$$T8 = \frac{5051}{11234} = 0,4496 = 44,96\%$$

### 2.2.5. Misclassification rate (A12)

#### Definition

The misclassification rate is number of units for which the collected information showed that the units were classified wrongly in the sample frame.

Put simply, as the major statistical surveys are carried using a stratified sample, the misclassification rate shows the share of units that have changed the stratum from which they were selected or have changed the activity (wrong classification of economic activity in the sample frame by NKD classification) or for example in the statistics of agriculture when the farm size was not correctly defined (if in the sample frame were defined as the small farms and during the field work they were identified as large farms).

Uz stopu nadobuhvata, stopa nepravilnog razvrstavanja je jedan od najvažnijih pokazatelja grešaka okvira.

Može se računati za ukupan uzorak ili za pojedine podskupine populacije/uzorka.

### Formula za izračun

$$MC_r = \frac{m_{mc}}{m} * 100$$

$m_{mc}$  – ukupan broj pogrešno razvrstanih jedinica

$m$  – ukupan broj jedinica analize koje su odgovorile na upitnik

### Primjer

Idući primjer pokazuje izračun stope nepravilnog razvrstavanja za Istraživanje o inovacijskim aktivnostima poduzeća (CIS) 2010. za različite podskupine uzorka prema veličini poduzeća (mala, srednja i velika poduzeća prema broju zaposlenih) te za ukupni uzorak (4 500 jedinica). Dakle, ovaj pokazatelj mjeri postotak poduzeća koja su promijenila stratum u vremenu između posljednjeg ažuriranja okvira i vremena kada se obavljao terenski rad. Stopa je procijenjena na temelju obilježja anketiranih poduzeća.

Beside the over-coverage rate, it is one of the most important indicators of the sample frame errors.

It could be calculated for a total sample or for the specific subgroups of the population/sample.

### Formula for calculation

$$MC_r = \frac{m_{mc}}{m} * 100$$

$m_{mc}$  – total number of wrong classified units

$m$  – total number of units that responded on the questionnaire

### Example

The following example shows the calculation of misclassification rate for the Community Innovation Survey 2010 (CIS) for different sample subgroups according to the enterprise size (small, medium and large enterprises according to the number of employees) and for the total sample. Therefore this indicator measures the percentage of enterprises that have change the stratum during the time between the last update of sample frame and the time when the field work was carried out. The rate was estimated based on the characteristics of surveyed enterprises.

## T 15. Izračun stopa nepravilnog razvrstavanja prema veličini poduzeća

Calculation of the misclassification rate by enterprise size

	Mala poduzeća <i>Small enterprises</i>	Srednja poduzeća <i>Medium enterprises</i>	Velika poduzeća <i>Large enterprises</i>	Ukupno <i>Total</i>
Broj anketiranih poduzeća prema okviru <i>Number of surveyed enterprises according to the sample frame</i>	1 743	1 324	323	3 390
Broj anketiranih poduzeća koja su promijenila stratum <i>Number of surveyed enterprises that have changed stratum</i>	26	145	38	209

Izvor: interni podaci DZS-a, Istraživanje o inovacijskim aktivnostima poduzeća 2010.

Source: internal CBS data, the Community Innovation Survey

Prvi redak (broj anketiranih poduzeća prema okviru) odnosi se samo na jedinice koje su ispunile upitnik i pripadaju ciljanoj populaciji, što znači da se stopa ne računa nad svim jedinicama iz uzorka. Dakle, stopa se računa samo na jedinicama koje su odgovorile na upitnik (3 390).

*First row (number of surveyed enterprises according to the sample frame) refers only to units that have fulfilled questionnaire and are belonging to the target population. That means that the rate was not calculated on all sample units. Therefore the rate was calculated only on units that have responded (3 390).*

Stopa nepravilnog razvrstavanja za mala poduzeća:

$$MC_r = \frac{26}{1743} * 100 = 1,49\%$$

*Misclassification rate for small enterprises:*

$$MC_r = \frac{26}{1743} * 100 = 1,49\%$$

Stopa nepravilnog razvrstavanja za srednja poduzeća:

$$MC_r = \frac{145}{1324} * 100 = 10,95\%$$

*Misclassification rate for medium enterprises:*

$$MC_r = \frac{145}{1324} * 100 = 10,95\%$$

Stopa nepravilnog razvrstavanja za velika poduzeća:

$$MC_r = \frac{38}{323} * 100 = 11,76\%$$

*Misclassification rate for large enterprises:*

$$MC_r = \frac{38}{323} * 100 = 11,76\%$$

Stopa nepravilnog razvrstavanja za sva poduzeća:

$$MC_r = \frac{209}{3390} * 100 = 6,17\%$$

*Misclassification rate for all enterprises:*

$$MC_r = \frac{209}{3390} * 100 = 6,17\%$$

Kao što se može vidjeti, stope nepravilnog razvrstavanja dosta su visoke za pojedine podskupine uzorka. Dokle god su procijenjeni koeficijenti varijacije mali, te visoke stope nisu velik problem. One bi mogle postati ozbiljan problem samo u slučaju kada bi se ozbiljno narušila zastupljenost malih jedinica u uzorku.

*It can be stated that the misclassification rate are quite high for certain sample subgroups. As long as the estimated coefficients of variation are small, high misclassification rates are not a big problem. They could become a serious problem only in case the presence of small units in the sample becomes seriously undermined.*



## 2.3. SMISLENOST

Smislenost se odnosi na usklađenost podataka sa sličnim podacima iz drugih izvora. Problemi s usklađenošću podataka mogu se pojaviti kada podaci potječu iz potpuno različitih izvora ili iz različitih statističkih istraživanja, pri čemu primijenjeni pojmovi, klasifikacije i metodološki standardi nisu bili usklađeni, tj. smisleni.

### 2.3.1. Smislenost između različitih izvora podataka (CH1)

#### Definicija

Smislenost između različitih izvora podataka pokazuje razliku između statističkih rezultata koji su proizvedeni istraživanjima i sličnih statističkih rezultata iz drugih izvora (npr. kratkoročnih i strukturnih poslovnih pokazatelja, i/ili nacionalnih računa, i/ili administrativnih izvora).

Može se iskazati u apsolutnome ili relativnom obliku, ovisno o vrsti podataka. Ako je riječ o apsolutnim iznosima, smislenost se prikazuje u relativnom obliku, a ako je riječ o relativnim podacima (tj. indeksi, postoci), smislenost se izražava apsolutno.

#### Formula za izračun

Smislenost između različitih izvora podataka –apsolutna razlika

$$S_i = |X_{i_{ref}} - X_i|$$

$X_{i_{ref}}$  – referentni podatak dobiven anketom

$X_i$  – podatak iz drugog izvora

$i$  – broj izvora podataka

## 2.3. COHERENCE

*The coherence refers to the compliance of data with similar data from other sources. The problems with the data compliance can occur when the data come from the completely different source or from different statistical surveys, whereby applied concepts, classifications and methodological standards were not harmonised, that means coherent.*

### 2.3.1. Coherence between different sources (CH1)

#### Definition

*The coherence between different sources represents the difference between statistical results produced as a result of statistical surveys and similar statistical results from other sources (e.g. short term and structural business statistics indicators and/or national accounts, and/or administrative sources.)*

*It can be presented in absolute or relative terms, depending on the data type. If it is a case of absolute amounts, the coherence is represented relatively, but if it is case of relative data (i.e. indices, percentages) the coherence is expressed in absolute terms.*

#### Formula for calculation

*The coherence between different sources – absolute difference:*

$$S_i = |X_{i_{ref}} - X_i|$$

$X_{i_{ref}}$  – reference data as a result of statistical survey

$X_i$  – data from other source

$i$  – number of data sources

Smislenost između različitih izvora podataka  
– relativna razlika

$$S_i = \frac{|X_{i_{ref}} - X_i|}{X_i}$$

$X_{i_{ref}}$  – referentni podatak dobiven  
anketama

$X_i$  – podatak iz drugog izvora

$i$  – broj izvora podataka

### Primjer

Sljedeći primjer prikazuje izračun smislenosti između različitih izvora podataka. Referentni podaci dobiveni su iz Ankete o dohotku stanovništva (ADS) za 2010. i želimo ih usporediti s podacima iz njoj najsličnijih socijalnih anketa, a to su Anketa o potrošnji kućanstava (APK) i Anketa o radnoj snazi (ARS). Podaci na temelju kojih želimo izračunati smislenost - Najviši završeni ISCED stupanj obrazovanja i Status u zaposlenosti prikazane su u sljedeće dvije tablice.

*The coherence between different sources –  
relative difference:*

$$S_i = \frac{|X_{i_{ref}} - X_i|}{X_i}$$

$X_{i_{ref}}$  – *reference data as a result of  
statistical survey*

$X_i$  – *data from other source*

$i$  – *number of data sources*

### Example

*The following example represents the calculation of coherence between different sources. The reference data are received from the Survey on Income and Living Conditions (SILC) for 2010 and it should be compared with the data of similar social surveys, and these are the Household Budget Survey (HBS) and the Labour Force Survey (LFS). The data used for calculating the coherence are “The highest completed ISCED education level” and “Status in employment” presented in following two tables.*

**T 16. Usporedba varijable "Najviši završeni ISCED stupanj obrazovanja "s drugim izvorima u Republici Hrvatskoj 2010.**

*Comparison of variables „The highest completed ISCED education level“ with other data sources in the Republic of Croatia*

ISCED razina obrazovanja <i>ISCED education level</i>	ADS <i>SILC</i>	ARS <i>LFS</i>	APK <i>HBS</i>
Predškolsko obrazovanje (dječji vrtići, predškola) <i>Pre-primary education (kindergartens, preschool)</i>	-	1,76%	-
Osnovno obrazovanje (niži razredi) <i>Primary education (lower classes)</i>	6,19%	7,25%	8,85%
Osnovno obrazovanje (viši razredi) <i>Primary education (higher classes)</i>	23,49%	24,68%	22,63%
Srednje obrazovanje <i>Secondary education</i>	55,28%	51,47%	53,06%
Obrazovanje nakon srednjega koje nije ni više ni visoko <i>Post-secondary non-tertiary education</i>	0,77%	-	1,00%
Više i visoko obrazovanje, magisterij <i>Tertiary education</i>	12,21%	14,60%	12,35%
Doktorat <i>Doctorate</i>	0,30%	0,24%	0,29%
Bez podatka <i>Without data</i>	1,75%	-	1,83%
UKUPNO <i>TOTAL</i>	100,00%	100,00%	100,00%

Izvor: interni podaci Državnog zavoda za statistiku, Anketa o dohotku stanovništva 2010.

Source: internal data of Croatian Bureau of Statistics, Survey on Income and Living Conditions 2010

**T 17. Usporedba varijable "Status u zaposlenosti" u ADS-u s podacima iz drugih izvora u Republici Hrvatskoj**

*Comparison of variable »Status in employment« in SILC with the data from other data sources in the Republic of Croatia*

Status u zaposlenosti <i>Status in employment</i>	ADS <i>SILC</i>	ARS <i>LFS</i>	APK <i>HBS</i>
Zaposleni <i>Employed persons</i>	32,35%	32,10%	-
Samozaposleni <i>Self-employed persons</i>	4,27%	6,83%	43,93%
Nezaposleni <i>Unemployed persons</i>	14,25%	9,23%	9,25%
Učenci, studenti, stažisti <i>Pupils, students, trainees</i>	8,97%	8,29%	8,54%
Umirovljenici <i>Pensioners</i>	32,08%	35,01%	32,10%
Nesposobni za rad <i>Unable to work</i>	1,29%	1,77%	1,49%
Obavljaju kućanske poslove i vode brigu o kućanstvu <i>Perform household work and take care of the household</i>	5,45%	6,77%	4,69%
Ostali ekonomski neaktivni <i>Other economic groups of non-active persons</i>	1,33%	0,00%	0,00%

Izvor: interni podaci Državnog zavoda za statistiku, Anketa o dohotku stanovništva 2010.

Source: internal data of Croatian Bureau of Statistics, Survey on Income and Living Conditions 2010

Da bismo mogli usporediti navedene podatke, najprije je potrebno provesti dodatne analize, ponajprije u smislu veličine uzorka i procjene veličine populacije na temelju uzorka. U sljedećoj tablici uspoređujemo ciljane varijable dohotka te broj osoba koje su primile dohodak za svaku od komponenata dohotka iz ADS s tim istim varijablama iz drugih izvora, primjerice s varijablama iz APK.

*In order to compare the listed data, it is necessary to carry out additional analysis, primarily in sense of the sample size and estimated population size based on the sample. In the following table are the comparisons of target variables on income and number of persons that received some income component for each SILC income component with the same variables from other sources, e.g. from HBS.*

#### **T 18. Usporedba ciljanih varijabli dohotka iz ADS-a s vanjskim izvorima u Republici Hrvatskoj 2010.**

*Comparison of target income variables from SILC with external sources in the Republic of Croatia*

Komponenta dohotka <i>Income component</i>	ADS SILC		APK HBS	
	broj osoba – utežen <i>number of persons – weighted</i>	broj osoba – neutežen <i>number of persons – un-weighted</i>	broj osoba – utežen <i>number of persons – weighted</i>	broj osoba – neutežen <i>number of persons – un-weighted</i>
Dohodak zaposlenih <i>Income of employees</i>	1 293 819,60	2 877	1 243 636,60	2 786
Dohodak od samostalne djelatnosti <i>Income from self-employment</i>	609 345,40	1 438	435 247,60	1 014
Starosne mirovine <i>Regular pensions</i>	680 143,50	1 760	847 473,60	1 896
Obiteljske mirovine <i>Family pensions</i>	222 510,70	630	191 693,70	457
Naknade za tjelesna oštećenja <i>Compensation for physical disability</i>	343 207,70	866	231 720,10	526
Bez dohotka <i>Without income</i>	762 102,30	1 746	872 536,10	2 024

Izvor: interni podaci Državnog zavoda za statistiku, Anкета o dohotku stanovništva 2010.

*Source: internal data of Croatian Bureau of Statistics, Survey on Income and Living Conditions 2010*

Analizom podataka iz gornje tablice utvrđujemo sličnost referentnog izvora podataka (ADS) s drugim izvorom (APK) te na temelju toga možemo izračunati smislenost. Budući da je riječ o postotcima, smislenost se računa kao apsolutna razlika.

*Based on the taken analysis from the above table it can be stated the similarity of reference data source (SILC) with other source (HBS) and based on that the coherence can be calculated. As percentages were used, the coherence is calculated as an absolute difference.*

Smislenost podataka za niže razrede osnovne škole između ADS-a i APK ( $S_2$ ) računa se na sljedeći način:

$$S_2 = |X_{i_{ref}} - X_i| = |6,19\% - 8,85\%| = 2,66\%$$

*The coherence for lower classes of primary school between SILC and HBS ( $S_2$ ) is calculated as follows:*

$$S_2 = |X_{i_{ref}} - X_i| = |6,19\% - 8,85\%| = 2,66\%$$

#### T 19. Izračun smislenosti ADS-a s ARS-om i APK-om za varijablu "najviši završeni ISCED stupanj obrazovanja"

*Calculation of coherence of SILC in comparison with LFS and HBS for ISCED variables »level of education«*

ISCED razina obrazovanja <i>ISCED education level</i>	$S_1$	$S_2$
Predškolsko obrazovanje (dječji vrtići, predškola) <i>Pre-primary education (kindergartens, preschool)</i>	-	-
Osnovno obrazovanje (niži razredi) <i>Primary education (lower classes)</i>	1,06%	2,66%
Osnovno obrazovanje (viši razredi) <i>Primary education (higher classes)</i>	1,19%	0,86%
Srednje obrazovanje <i>Secondary education</i>	3,81%	2,22%
Obrazovanje nakon srednjega koje nije ni više ni visoko <i>Post-secondary non-tertiary education</i>	-	0,23%
Više i visoko obrazovanje, magisterij <i>Tertiary education</i>	2,39%	0,14%
Doktorat <i>Doctorate</i>	0,06%	0,01%
Bez podatka <i>Without data</i>	-	0,08%

Izvor: izračun autora

*Source: authors' calculation*

Na isti način možemo izračunati i smislenost za varijablu "status u zaposlenosti".

*In the same way can be calculated the coherence for the variable „Status in employment“*

**T 20. Izračun smislenosti ADS-a s ARS-om i APK-om za varijablu "status u zaposlenosti"**  
*Calculation of coherence of SILC in comparison with LFS and HBS for variable »status in employment«*

Status u zaposlenosti <i>Status in employment</i>	$S_1$	$S_2$
Zaposleni <i>Employed persons</i>	0,25%	-
Samozaposleni <i>Self-employed persons</i>	2,56%	39,66%
Nezaposleni <i>Unemployed persons</i>	5,02%	5,00%
Učenici, studenti, stažisti <i>Pupils, students, trainees</i>	0,68%	0,43%
Umirovljenici <i>Pensioners</i>	2,93%	0,02%
Nesposobni za rad <i>Unable to work</i>	0,48%	0,20%
Obavljaju kućanske poslove i vode brigu o kućanstvu <i>Perform household work and take care of the household</i>	1,32%	0,76%
Ostali ekonomski neaktivni <i>Other economic groups of non-active persons</i>	1,33%	1,33%

Izvor: izračun autora  
*Source: authors' calculation*

# LITERATURA

## LITERATURE

### Knjige

#### Books

- Atkinson, A. B. et al. (2002). *Social Indicators: The EU and social inclusion*, Oxford: Oxford University Press, USA, 1st edition
- Biemer, P. P. et al. (1991), *Measurement Errors in Surveys*, Wiley & Sons, Inc, USA
- Biemer, P. P. (2003), *Introduction to Survey Quality*, Wiley & Sons, Inc., Hoboken, New Jersey
- Biewen, M., Jenkins, S. P. (2006). Variance estimation for generalized entropy and Atkinson inequality indices: The complex survey data case, *Oxford Bulletin of Economics and Statistics*, 68(3)
- Cochran, W. G. (1977). *Sampling Techniques*, Wiley & Sons, Inc., New York
- Foreman, E. K. (1991), *Survey Sampling Principles*, Marcel Dekker, Inc., New York
- Groves, R. M. et al. (2002). *Survey Nonresponse*, John Wiley & Sons, Inc., New York
- Groves, R. M. et al. (1998). *Nonresponse in Household Interview Survey*, Wiley & Sons, Inc., Canada
- Hussmans, R. et al. (1990), *Surveys of the Economically Active Population, Employment, Unemployment and Underemployment*, International Labour Organisation Manual on Concepts and Methods, Geneva, 1990.
- Lyberg, L., et al. (1997). *Survey Measurement and Process Quality*, Wiley & Sons, Inc., New York

### Članci i izvještaji

#### Articles and reports

- Brackstone, G. (1999): *Managing Data Quality in a Statistical Agency*. U: *Survey Methodology* 25, str. 139-149
- Burg, T. (2004): *Quality Reports at Statistics Austria*. Paper presented at the European Conference on Quality and Methodology in Official Statistics (Q2004), Mainz, Germany, 24-26 May 2004.
- Elvers, E. (2004): *Quality and Metadata – Some Connections and Reflections*. Paper presented at the European Conference on Quality and Methodology in Official Statistics (Q2004), Mainz, Germany, 24-26 May 2004.
- Hahn, M. and Lindén, H. (2006): *The European Statistics Code of Practice for a High Quality*
- *European Statistical System*. Paper presented at the European Conference on Quality in Survey Statistics (Q2006), Cardiff, United Kingdom, 24-26 April 2006.
- Olson, K., West, B.T., *How Much of Interviewer Variance is Really Nonresponse Error Variance?*, *Oxford Journals, Social Sciences, Public Opinion Quarterly*, Volume 74, Issue 5, str.1004-1026
- Verma, V., et al. (1996). *An Analysis of Sampling Errors for the Demographic and Health Surveys*, *International Statistical Review*, 64(3), str. 265–294.

## Metodološki dokumenti na europskoj razini (priručnici)

### *Methodological documents on the european level (handbooks)*

- Standard Quality Indicators - Producer oriented: Working Group "Assessment of quality in statistics", Sixth meeting, Luxembourg 2-3 October 2003
- Methodological documents, Handbook "How to make quality report: Working Group "Assessment of quality in statistics", Sixth meeting, Luxembourg 2-3 October 2003
- Methodological documents, Standard Report: Working Group "Assessment of quality in statistics", Sixth meeting, Luxembourg 2-3 October 2003
- Eurostat (2007): "Handbook on Data Quality Assessment Methods and Tools (DatQAM)", Handbook written by Körner, T., Bergdahl, M., Elvers, E., Lohaus, P., Nimmergut, A., Sæbø, H.V., Szép, K., Timm, U., and Zilhão M.J.
- Eurostat (2003): "Handbook On improving quality by analysis of process variables", Handbook written by Aitken, A., Hörngren, J., Jones, N., Lewis, D. and Zilhão M.J.
- Eurostat (2009): "ESS Handbook for Quality Reports", Methodologies and Working Papers, Eurostat, Luxembourg
- Eurostat (2009): "ESS Standard for Quality Reports", Methodologies and Working Papers, Eurostat, Luxembourg
- Eurostat, (1997), Household Budget Surveys Methodology and Recommendations for Harmonisation 1997, theme 3E, Eurostat (1997) (odlomak 4.1.2)
- Statistics Canada (2003): Statistics Canada Quality Guidelines, 4th edition, October 2003.
- Statistics Finland (2007): Quality Guidelines for Official Statistics.
- ONS (2007), Guidelines for Measuring Statistical Quality, Version 3.1, London.

## Internetske stranice

### **Web sites**

- Državni zavod za statistiku. Dostupno na: <http://www.dzs.hr/> (21. 02. 2013.)
- <http://poq.oxfordjournals.org/content/74/5/1004.full> ,How Much of Interviewer Variance is Really Nonresponse Error Variance?, West, B., Olson, K., (23. 2. 2013.)
- [http://epp.eurostat.ec.europa.eu/portal/page/portal/income\\_social\\_inclusion\\_living\\_conditions/quality/national\\_quality\\_reports](http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/quality/national_quality_reports); Izvještaji o kvaliteti podataka za anketu EU-SILC zemalja članica EU; (28. 2. 2012.)
- Eurostat, 'ESS Handbook for Quality Reports' (2009.)  
[http://epp.eurostat.ec.europa.eu/portal/page/portal/ver1/quality/documents/EHQR\\_FINAL.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/ver1/quality/documents/EHQR_FINAL.pdf)
- Stranica Eurostata posvećena kvaliteti  
<http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/introduction>
- OECD-ova stranica s metodološkim dokumentima  
[http://www.oecd.org/findDocument/0,3354,en\\_2649\\_34257\\_1\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/findDocument/0,3354,en_2649_34257_1_1_1_1_1,00.html)
- Konferencija za kvalitetu Q2004 (Mainz)  
<http://q2004.destatis.de/>
- Konferencija za kvalitetu Q2006 (Cardiff)  
<http://www.statistics.gov.uk/events/q2006/default.asp>
- Konferencija za kvalitetu Q2008 (Rim)  
<http://q2008.istat.it/>



## **Statističke baze**

### ***Statistical data bases***

- Državni zavod za statistiku Republike Hrvatske (Anketa o potrošnji kućanstava, Anketa o dohotku stanovništva, Trgovina na malo –TRG-1, ...)
- Eurostat (različiti izvještaji o kvaliteti statističkih ureda EU-a te metodološki priručnici i radovi)
- CIRCA (Communication & Information Resource Centre Administrator)