ЄВРОПЕЙСЬКА КОМІСІЯ

ЄВРОСТАТ

Директорат B: Якість, методологія та інформаційні системи

**Підрозділ B-1: Якість; Класифікації**

Люксембург

ESTAT/B1/AB D(2011)

**РЕКОМЕНДАЦІЇ ЄСС З ІМЛЕМЕНТАЦІЇ**

**ІНДИКАТОРІВ ЯКОСТІ ТА ЕФЕКТИВНОСТІ 2010**



Ці індикатори були в останнє переглянуті Експертною групою Євростату з питань індикаторів якості у 2010 році

***Для отримання додаткової інформації стосовно цих індикаторів, будь ласка, звертайтеся до Підрозділу B-1:*** [***estat-qualit***](mailto:estat-quality@ec.europa.eu)[***y@ec.europa.eu***](mailto:y@ec.europa.eu)

Європейська Комісія, 2920 Люксембург, ЛЮКСЕМБУРГ - тел. +352 43011

Офіс: Бех - тел. пряма лінія +352 4301-35564 - факс +352 4301-33899

[http://epp.eurostat.ec.europa.eu](http://epp.eurostat.ec.europa.eu/) [Antonio.Baigorri@ec.europa.eu](mailto:Antonio.Baigorri@ec.europa.eu)

**РЕКОМЕНДАЦІЇ ЄСС З ІМЛЕМЕНТАЦІЇ**

**ІНДИКАТОРІВ ЯКОСТІ ТА ЕФЕКТИВНОСТІ 2010**

|  |  |
| --- | --- |
| **Назва:** | **R1. Повнота даних - рівень** |
| Визначення: | Співвідношення кількості наданих елементів даних (суб’єкти, які повинні бути визначені адміністратором домену Євростату) до кількості елементів даних, які потрібні Євростату або інших відповідних даних. Співвідношення вираховується для обраного набору даних та за певний період часу. |
| Застосовність: | Рівень доступних даних застосовується:   * до всіх статистичних процесів (в тому числі, використання адміністративних джерел); * до всіх користувачів та виробників (різний фокус та формули)   Вираховується лише Євростатом, але рекомендується також для включення в національні звіти з якості. |
| Формула розрахунку: | **Для конкретної ключової змінної величини:**  **Для виробників:**    *Drqd* у знаменнику це набір елементів з необхідними даними (тобто, за винятком обмежень/конфіденційності) та # *AD rqd* у чисельнику – це відповідний субнабір доступних/наданих елементів даних. Примітка # *D* означає кількість елементів у наборі D (кількісний числівник).    **Для користувачів**    *Drel*  у знаменнику – це набір елементів з відповідними даними (повне охоплення, тобто, за винятком тільки тих суб’єктів, до яких дані не будуть відноситися, як, наприклад, рибацький флот в Угорщині), та *AD rel* у числівнику – це відповідний субнабір доступних/наданих елементів даних. Примітка # *D* означає кількість елементів у наборі D (кількісний числівник).  Головна відмінність між двома формулами полягає у виборі наборів даних для знаменника.  Стосовно першої формули, для **виробників**, цей набір містить елементи даних, за винятком обмежень/конфіденційності, оскільки виробники зацікавлені в оцінці рівня відповідності вимогам.  З іншого боку, для **користувачів**, формула надає коефіцієнт співвідношення наданих елементів даних до тих, які є теоретично відповідними, що означає, що елементи відсутні через обмеження/ конфіденційність або з будь-якої іншої причини відсутності даних містяться тут, не включаючи лише ті елементи, для яких дані не будуть відповідними, наприклад, рибацький флот в Угорщині. |

|  |  |
| --- | --- |
| Цільове значення: | Цільове значення для цього індикатора – 1, тобто доступні 100% необхідних або відповідних елементів даних. |
| Рівні та принципи агрегації: | Розрахунки проводяться, для змістовного вибору адміністратором домену, на рівні тематичного домену. Агрегації рекомендуються на рівні ЄС для індикатора, орієнтованого на користувача.  Кількість наданих елементів з даними та кількість елементів з необхідними/відповідними даними з’єднуються окремо, після чого вираховується співвідношення. |
| Інтерпретація: | Індикатор вказує, в якій мірі статистичні дані є доступними порівняно з тим, як вони мають бути доступними.  **Для виробників:**  Може бути використаний для оцінки рівня відповідності певною державою-членом для заданого набору даних та періоду часу, що визначаються адміністратором домену.  **Для користувачів:**  На рівні ЄС, він може бути використаний для того, щоб   * визначити, чи відсутні важливі змінні для певної окремої держави-члена або навпаки * надати користувачам загальне вимірювання (сукупне по країнам та/або ключовим змінним) доступності статистики. |
| Особливі вказівки: | Цей індикатор має супроводжуватися інформацією про те, якої змінної не вистачає, та причин неповноти, а також, у відповідних випадках, інформацією про вплив відсутніх даних на сукупність ЄС та на плани для покращення повноти в майбутньому.  Розрахунок буде потребувати втручання адміністратора домену Євростату на початковому етапі (для визначення ключових змінних та періоду, який контролюється). Пізніше, індикатори повинні розраховуватись автоматично.  Обидві формули мають розраховуватись по кожній ключовій змінній, однак можливо розрахувати сукупність для всіх змінних.  **Для виробників:**  Цей індикатор є частиною моніторингу відповідності Євростату, таким чином, для виробників він має розраховуватися по державам-членам.  **Для користувачів:**  Якщо у звіт не внесені певні відповідні змінні, статистика є неповною. Це може статися, якщо дані не були зібрані, або зібрані дані низької якості, або конфіденційні. Для користувачів сукупність всіх ключових змінних по країнам може бути достатньою. |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). * ISO/IEC FDIS 11179-1 "Інформаційна технологія - Реєстри метаданих - Частина 1: Структура"б Березень 2004 (згідно з проектом загального словника метаданих SDMX, лютий 2008 р.). |

|  |  |
| --- | --- |
| **Назва:** | **А1. Помилка вибірки - індикатори** |
| Визначення: | Помилка вибірки може бути виражена:   1. у відносному вираженні, в такому випадку використовується відносна стандартна помилка або, синонімічно, коефіцієнт варіації (КВ). (Стандартна помилка оцінки  дорівнює кореню квадратному її відхилення ).) Очікувана відносна стандартна помилка (очікуваний КВ) це очікувана стандартна помилка оцінки поділена на очікуване значення параметру, див. формулу розрахунку нижче. 2. з точки зору інтервалів довіри, тобто інтервал, який з заданим рівнем довіри включає справжнє значення параметру . Ширина інтервалу залежить від стандартної помилки.   Статистична оцінка має враховувати план вибірки і надалі інтегрувати вплив на точність коригувань щодо неотримання даних, виправлень щодо неправильних класифікацій, використання допоміжної інформації через методи калібрування тощо. |
| Застосовність: | Індикатори помилок вибірки застосовуються до:   * статистичних процесів, які засновані на ймовірнісній вибірці або інших процедурах вибірки, що дозволяють обчислення такої інформації. * користувачів та виробників. |
| Формула розрахунку: | **Коефіцієнт варіації:**    Примітка: Нижній індекс "е" означає оцінку.  **Інтервал довіри, симетричний:**    Довжина інтервалу, яка складає 2·d, залежить від рівня довіри (наприклад, 95%), припущень, які зумовлюють розподіл оцінки параметру, та помилки вибірки. У багатьох випадках d має форму, представлену нижче, у той час як t залежить від розподілу та рівня довіри.    Щодо підсумкових результатів, засобів та співвідношень, то формули для агрегації коефіцієнтів варіації на рівні ЄС можна знайти у третьому посиланні, наданому нижче.  Формули розрахунку залежать від плану вибірки, оцінки та |

|  |  |
| --- | --- |
|  | методу, обраного для оцінки величини відхилення *V*() . |
| Цільове значення: | Чим менше КВ, стандартна помилка та ширина інтервалу довіри, тим точніша оцінка. Норми досліджень можуть включати специфікації стосовно порогів точності на різних рівнях сукупності. |
| Рівні та принципи агрегації: | Розрахунки проводяться для всіх статистичних даних, які основані на дослідженнях ймовірнісної вибірки або їхніх еквівалентів. Агрегації можливі на рівні держав-членів та ЄС, в залежності від статистичних оцінок та ступеня гармонізації.  Принцип розрахунку коефіцієнту варіації сукупності залежить від методу агрегації статистичної оцінки, що належить до тієї змінної. |
| Інтерпретація: | КВ - це відносна (безвимірна) міра точності статистичної оцінки, часто виражається у відсотках. Точніше кажучи, вона має властивість виключати одиниці вимірювання з мір точності, та однією з її ролей є уможливити порівняння між точністю статистичних оцінок різних індикаторів. Однак, ця властивість не має ніякого значення, яке додається у випадках з пропорціями (які за визначенням є безвимірними індикаторами). |
| Особливі вказівки: | Існує кілька мір точності, які можуть бути використані для оцінки випадкової варіації статистичної оцінки внаслідок вибірки, такі як коефіцієнт варіації, стандартні помилки та інтервал довіри.  Коефіцієнт варіації підходить для кількісних змінних з великими позитивними значеннями. Він не є стабільним для відсотків або змін та не використовується для статистичної оцінки даних з негативними значеннями, де вони можуть бути замінені абсолютними мірами точності (стандартні помилки та інтервали довіри).  Інтервал довіри, зазвичай, являє собою міру точності, якій надають перевагу користувачі даних. Цей найбільш чіткий спосіб для розуміння та інтерпретації мінливості вибірки.  Надання інтервалів довіри є добровільним.  КВ має перевагу бути безвимірним. Іноді перевага надається стандартній помилці або інтервалу довіри, як вже зазначалося. |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). * Методи оцінки розбіжностей у Європейському союзі, Монографії офіційної статистики, вид. 2002. |

|  |  |
| --- | --- |
| **Назва:** | **А2. Рівень надохоплення** |
| Визначення: | Рівень надохоплення це пропорція одиниць, доступних через структуру, які не належать до цільової сукупності (поза межами діапазону).  *Цільова сукупність* цесукупність, для якої робляться висновки. *Рамка* (або рамки) – це механізм, який дозволяє доступ до одиниць сукупності. *Рамкова сукупність* – це набір одиниць сукупності, до яких є доступ через рамку. Поняття рамки традиційно використовується для досліджень вибірки, але так само застосовується до деяких інших статистичних процесів, наприклад, переписів, процесів, які використовують адміністративні джерела, та процесів, що залучають декілька джерел даних. Дефіцит охоплення може бути спричинений затримками у звітуванні (типово для статистичного бізнесу) та помилками в ідентифікації, класифікації, кодуванні одиниць тощо. В такому випадку також використовуються адміністративні дані.  Рівень може бути розрахований як незначущий або як значущий по відношенню до загального рівня (рамка/сукупність, а не вибірка). Копії включені до надохоплення. Одиниці з невідомою прийнятністю обумовлюють невід’ємні труднощі, див. нижче. |
| Застосовність:: | Рівень надохоплення застосовується до:   * всіх статистичних процесів (в тому числі, використання адміністративних джерел); * користувачів та виробників.   Якщо в дослідженні присутні більше одного типу одиниць, рівень може бути розрахований для кожного типу. Якщо існує більше однієї рамки або якщо рівні надохоплення дуже відрізняються між субсукупностями, рівні необхідно відокремити. |
| Формула розрахунку: | Рівень надохоплення має три основні версії, написані в одній і тій самій формулі, як значущий рівень надохоплення *OCrw*    О набір одиниць поза межами діапазону (надохоплення, підтверджені та ті, що не належать цільовій сукупності)  E набір одиниць у межах діапазону (підтверджені одиниці, та такі, що належать цільовій сукупності; прийнятні одиниці)  Q набір одиниць невідомої прийнятності.  *wj* значущість одиниці *j*, описана нижче.  α Оцінена пропорція випадків невідомої прийнятності, які насправді є прийнятними. Має дорівнювати 1, якщо немає переконливих доказів для інших припущень на рівні країн. |

|  |  |
| --- | --- |
|  | Три основні випадки:  Незначущий рівень: *wj* = 1  Рівень значущості за планом: *w j* *d j* де по суті *d j* 1*j*, що означає, що значущість за планом є зворотною до ймовірності вибору.  Рівень значущості за розміром: *w j* *d j x j* де *x j* це значення змінної X.  Змінна Х, яка обирається суб’єктивно, показує розмір або важливість одиниць. Значення має бути відомим для всіх одиниць. Х – це допоміжна інформація, часто доступна в структурі. Прикладами є оборот для підприємств та сукупність для муніципалітетів.  Для рівня надохоплення найчастіше використовуються незначущі та значущі за планом альтернативи, див. Інтерпретацію нижче.  Рівень значущості за планом в основному використовується у вибіркових дослідженнях, але також може бути застосований до, наприклад, процесів індексації цін або процесів з багатьма джерелами даних. Значущість *d j* – це “збільшуючий” фактор, коли одиниця *j* представляє більше, ніж вона є. В іншому випадку *d j* дорівнює одиниці. Таким чином, маючи справу з адміністративними джерелами, незначущі та значущі за розміром версії рівня стають досить цікавими. |
| Цільове значення: | Цільове значення цього індикатора максимально наближене до 0. |
| Рівні та принципи агрегації: | * МS (держави-члени): індикатор, який розраховується для рамкових сукупностей, де він є значущим, наприклад, для промисловості. Тоді окремі рамкові сукупності розглядаються як одна рамкова сукупність. * ЄС: індикатор може бути поєднаний між країнами тільки там, де процеси статистичного виробництва повністю гармонізовані. Окремі рамкові сукупності розглядаються як одна рамкова сукупність для статистичних процесів, що застосовуються. У випадках, коли процеси виробництва різняться в країнах, нижчі та вищі рівні надохоплення можуть бути показані з метою демонстрації діапазону. |
| Інтерпретація: | *Надохоплення*: існують одиниці, доступні через рамку, які не належать до цільової сукупності (наприклад, померлі люди, все ще внесені у списки Перепису населення або вже непрацюючі підприємства у списках РП).  Відсоток індикатора залежить від статистичного процесу та шляхів ідентифікації надохоплення. Якщо адміністративні дані також використовуються для визначення цільової сукупності, цей індикатор зазвичай має невелике додаткове значення, окрім можливих повторень, якщо такі будуть знайдені. Він може надати загальне уявлення про якість реєстру/рамки та рівень змін сукупності.  Рівень незначущого надохоплення надає кількість одиниць, які були визначені як такі, що не належать до цільових, пропорційно до |

|  |  |
| --- | --- |
|  | загальної кількості досліджуваних одиниць. Кількість відноситься до вибірки, перепису або реєстру сукупності, що вивчається.  Рівень значущого за планом надохоплення – це оцінка рамкової сукупності у порівнянні з цільовою сукупністю, що основана на доступній інформації, зазвичай на вибірці.  Рівень значущого за розміром надохоплення виражає рівень відносно змінної обраного розміру, наприклад, оборот у статистиці підприємств. (Цей випадок менш цікавий для надохоплення ніж для неотримання даних). |
| Особливі вказівки: | - |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). |

|  |  |
| --- | --- |
| **Назва:** | **А3. Рівень неотримання даних одиниці** |
| Визначення: | Співвідношення кількості одиниць, що не мають інформації або містять інформацію, яку не можна використовувати (неотримання даних тощо) до загальної кількості одиниць у межах дослідження (прийнятних). Співвідношення може бути значущим або незначущим. |
| Застосовність: | Рівень неотримання даних одиниці застосовується до:   * всіх статистичних процесів (в тому числі, до прямого збору даних та адміністративних даних; термінологія різниться між статистичними процесами, але основний принцип однаковий; у деяких випадках може бути важко розрізнити між неотриманням даних одиниці та недоохопленням, особливо це стосується джерел адміністративних даних (як відомо, в останньому випадку одиниці існують, але дані відсутні, наприклад, через пізнє надання звіту або таку низьку якість, що інформація стає непридатною – в останньому випадку одиниці невідомі у побудові структури; * користувачів та виробників. |
| Формула розрахунку: | Рівень неотримання даних має три основні версії, записані в одній і тій самій формулі як значущий рівень неотримання даних одиниці *NRrw*    R набір відповідних прийнятних одиниць  NR набір невідповідних прийнятних одиниць  Q набір обраних одиниць з невідомою прийнятністю (непідтверджені обрані одиниці)  *wj* вага одиниці *j*, описана нижче.  α Оцінена пропорція випадків невідомої прийнятності, які насправді є прийнятними. Має бути визначена на рівні 1, якщо немає переконливих доказів на рівні країни для іншого припущення.  Три основні випадки:  Незначущий рівень: *w j* 1  Рівень значущості за планом: *w j* *d j* де по суті *d j* 1*j*, що означає, що значущість за планом є зворотною до ймовірності вибору.  Рівень значущості за розміром: *w j* *d j x j* де *x j* – це значення змінної X.  Змінна Х, яка обирається суб’єктивно, показує розмір або важливість одиниць. Значення має бути відомим для всіх одиниць. Х – це допоміжна інформація, часто доступна в структурі. Прикладами є оборот для підприємств та сукупність для муніципалітетів.  Для рівня неотримання даних на одиницю зазвичай використовуються всі три альтернативи, див. Інтерпретацію нижче. |

|  |  |
| --- | --- |
|  | Рівень значущості за планом в основному використовується у вибіркових дослідженнях, але також може бути застосований до, наприклад, процесів індексації цін або процесів з багатьма джерелами даних. Значущість *d j* це “збільшуючий” фактор, коли одиниця *j* представляє більше, ніж вона є. В іншому випадку *d j* дорівнює одиниці. Таким чином, маючи справу з адміністративними джерелами, незначущі та значущі за розміром версії рівня стають досить цікавими. |
| Цільове значення: | Цільове значення для цього індикатора максимально наближене до 0. |
| Рівні та принципи агрегації: | * МS: індикатор має бути розрахований на рівні статистичних процесів * ЄС: замість агрегації цього індикатора по країнам або розрахунку, середні, мінімальні або максимальні рівні неотримання даних одиниці можуть бути показані Євростатом для певної змінної на рівні статистичного процесу. |
| Інтерпретація: | Неотримання даних одиниці відбувається, коли не записуються дані стосовно прийнятної одиниці (або даних так мало або вони такої низької якості, що вони видаляються).  Незначущий рівень неотримання даних показує результат збору даних у вибірці (включаючи одиниці), а не непряму міру потенційної похибки, яка пов’язана з неотриманням даних. Якщо α=1, то передбачається, що всі одиниці з невідомою прийнятністю є прийнятними, тому вона надає консервативну оцінку А4 з урахуванням інших виборів α .  Рівень значущого за планом неотримання даних одиниці показує, наскільки добре спрацював збір даних з урахуванням сукупності, що вивчається.  Рівень значущого за розміром неотримання даних одиниці буде являти собою непрямий індикатор потенційної похибки, що спричинена неотриманням даних перед будь-якими налаштуваннями калібрування.  Необхідно зазначити, що похибка може бути незначною, навіть якщо рівень неотримання даних високий, в залежності від характеру неотримання даних та можливостей успішного урегулювання при неотриманні даних. |
| Особливі вказівки: | Неотримання даних є джерелом помилок в дослідженнях статистики в основному з двох причин:   * воно зменшує кількість отриманих даних і, таким чином, точність оцінок (це може бути особливо актуальним при використанні вибірки); * може призвести до похибок. Розмір похибки залежить від рівня неотримання даних, а також від розбіжностей між респондентами та особами, що не дали відповіді, з урахуванням змінної, що вивчається; крім того, від ефективності додаткової інформації. |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). * Статистичні стандарти якості Бюро перепису населення, перевидано в 2010 р. * Trépanier, Julien, та Kovar . "Звітування щодо рівня отримання даних, коли об’єднуються дослідження та адміністративні дані". *Протоколи Федерального комітету на науковій конференції з статистичної методології, 2005* |

|  |  |
| --- | --- |
| **Назва:** | **А4. Рівень неотримання даних елементу** |
| Визначення: | Рівень неотримання даних елементу для певної змінної визначається як (значуще) співвідношення між одиницями у межах дослідження, що не мають відповіді та одиницями у межах дослідження, які мають відповіді на конкретний елемент. |
| Застосовність: | Рівень неотримання даних елементу застосовується до:   * всіх статистичних процесів (в тому числі до прямого збору даних та адміністративних даних; термінологія різниться між статистичними процесами, але основний принцип однаковий; * користувачів та виробників (для ключових змінних або змінних з дуже високим рівнем неотримання даних елементу). |
| Формула розрахунку: | Рівень неотримання даних елементу має три основні версії, записані в одній і тій самій формулі як значущий рівень неотримання даних елементу *NRY rw* , який розраховується наступним чином:    *RY* набір прийнятних одиниць, що відповідають елементу *Y* (якщо необхідно)  *NRY* набір прийнятних одиниць, що не відповідають елементу *Y,* хоча цей елемент є необхідним. – Знаменник відноситься до набору одиниць, для яких елемент *Y* є необхідним. (Інші одиниці не отримують цей елемент, тому що їхні відповіді на попередні елементи дали їм можливість пропустити цей елемент, вони були "відфільтровані".)  *wj* значущість одиниці *j*, описана нижче  Три основні випадки:  Незначущий рівень: *w j* 1  Рівень значущості за планом: *w j* *d j* де по суті *d j* 1*j*, що означає, що значущість за планом є зворотною до ймовірності вибору.  Рівень значущості за розміром: *w j* *d j x j* де *x j* – це значення змінної X.  Змінна Х, яка обирається суб’єктивно, вказує на розмір або важливість одиниць. Значення має бути відомим для усіх одиниць. Х – це допоміжна інформація, часто доступна в структурі. Прикладами є оборот для підприємств та сукупність для муніципалітетів.  Для рівня неотримання даних елементу зазвичай використовуються всі три альтернативи, див. Інтерпретацію нижче.  Рівень значущості за планом в основному використовується у вибіркових дослідженнях, але також може бути застосований до, наприклад, процесів індексації цін або процесів з багатьма джерелами даних. Значущість *d j* це “збільшуючий” фактор, коли одиниця *j* представляє більше, ніж вона є. В іншому випадку *d j* дорівнює одиниці. Таким чином, маючи справу з адміністративними джерелами, незначущі та значущі за розміром версії рівня стають досить цікавими. |

|  |  |
| --- | --- |
| Цільове значення: | Цільове значення для цього індикатора максимально наближене до 0. |
| Рівні та принципи агрегації: | * МS: індикатор має бути розрахований на рівні статистичних процесів для ключових змінних та змінних з низькими рівнями. * ЄС: замість агрегації цього індикатора по країнам або розрахунку, середні, мінімальні або максимальні рівні неотримання даних елементу можуть бути показані Євростатом для певної змінної на рівні статистичного процесу. |
| Інтерпретація: | Високий рівень неотримання даних елементу вказує на труднощі надання інформації, наприклад, делікатне питання або нечітке формулювання для соціальної статистики або інформація, яка є недоступною у бухгалтерській системі для статистику підприємств.  Цей індикатор являє собою індикативний критерій можливої похибки, що викликана неотриманням даних елементу. Незважаючи на низький рівень неотримання даних елементу, похибка може бути низькою, в залежності від причин, зразків відповіді та додаткової інформації для регулювання/імпутації. |
| Особливі вказівки: | Незначущий рівень неотримання даних елементу може бути розрахований перед редагуванням та підстановкою даних для того, щоб виміряти вплив неотримання даних елементу для ключових змінних. |
| Посилання | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). * Статистичні стандарти якості Бюро перепису населення, перевидане 2010. * Trépanier, Julien, та Kovar . "Звітування щодо рівня отримання даних, коли об’єднуються дослідження та адміністративні дані". *Протоколи Федерального комітету на науковій конференції з статистичної методології, 2005* |

|  |  |
| --- | --- |
| **Назва:** | **А5. Рівень імпутації** |
| Визначення: | Імпутація – це процес, який використовується для призначення замінних значень для відсутніх, невірних або суперечливих даних, які не вдалося відредагувати. Це виключає продовження роботи з респондентами та ручну перевірку та виправлення (якщо доцільно). Таким чином, імпутація, як визначене вище, відбувається після збору даних, незалежно від того, з якого джерела або сполучення джерел були отримані дані, в тому числі, адміністративні дані.  Після імпутації файл даних зазвичай має містити лише правдоподібні та внутрішньо узгоджені записи даних.  На цей індикатор впливають як неотримання даних елементу, так і процес редагування. Він вимірює відносну кількість імпутованих значень та відносний вплив на остаточні оцінки з процедур імпутації.  Незначущий рівень імпутації для змінної – це співвідношення кількості імпутованих значень до загальної кількості значень, запитуваних для змінної.  Значущий рівень ампутації відображає відносний внесок імпутованих значень у статистику; зазвичай підсумковий для кількісної змінної. Для якісної змінної, відносний внесок базується на кількості одиниць з імпутованим значенням для якісного елемента. |
| Застосовність: | Рівень імпутації застосовується до:   * всіх статистичних процесів (з мікроданими; отже, наприклад, прямий збір даних та адміністративні дані); * користувачів та виробників |
| Формула розрахунку: | Рівень імпутації має три основні версії, що записані в одній і тій самій формулі як значущий рівень імпутації *IRY rw* для змінної Y    *IY* набір одиниць, для яких імпутується змінна *Y*  *KY* набір одиниць, для яких зберігається змінна *Y*  *wj* значущість одиниці *j*, описана нижче  Два набори *IY* та *KY* разом (об’єднання) – це теж саме, що і *RY* та  *NRY* разом.  У випадку з якісною змінною, значення *y j* 1, якщо одиниця *j* вказує на дану характеристику, в іншому випадку – 0. |

|  |  |
| --- | --- |
|  | Коли підраховується імпутація, треба взяти до уваги наступні зміни:   1. імпутація (непустого) значення для відсутнього елемента 2. імпутація (непустого) значення для виправлення знайденого невірного (непустого) значення 3. імпутація пустого значення для виправлення знайденої невірної (непустої) відповіді.   Три основні випадки для рівня імпутації:  Незначущий рівень: *w j* 1  Рівень значущості за планом: *w j* *d j* де по суті *d j* 1*j*, що означає, що значущість за планом є зворотною до ймовірності вибору.  Рівень значущості за розміром: *w j* *d j x j* де *x j* – це значення змінної X.  Змінна Х, яка обирається суб’єктивно, вказує на розмір або важливість одиниць. Значення має бути відомим для усіх одиниць. Х – це допоміжна інформація, часто доступна в структурі. Прикладами є оборот для підприємств та сукупність для муніципалітетів.  У розрахунках остаточних оцінок значущість за планом може бути змінена з метою виправлення для неотримання даних, недоохоплення тощо. Ця значущість за планом має бути використана, якщо рівні мають бути застосовані до остаточних оцінок, наприклад, рівень значущості за розміром.  Для рівня імпутації зазвичай використовуються всі три альтернативи, див. Інтерпретацію нижче.  Рівень значущості за планом в основному використовується у вибіркових дослідженнях, але також може бути застосований до, наприклад, процесів індексації цін або процесів з багатьма джерелами даних. Значущість *d j* це “збільшуючий” фактор, коли одиниця *j* представляє більше, ніж вона є. В іншому випадку *d j* дорівнює одиниці. Таким чином, маючи справу з адміністративними джерелами, незначущі та значущі за розміром версії рівня стають досить цікавими. |
| Цільове значення: | Бажаним є значення,яке дорівнює 0 або наближене до 0; імпутація вказує відсутні та недійсні значення. |
| Рівні та принципи агрегації: | * МS: Розрахунок проводиться для ключових змінних на рівні статистичного процесу. * ЄС: Агрегація може бути проведена на рівні ЄС на основі гармонізованих процесів статистичного виробництва по державам-членам, сприймаючи це як єдиний статистичний процес. Крім того, Євростат може доповідати про нижчий та вищий рівні імпутації для певної змінної на рівні статистичного процесу. |

|  |  |
| --- | --- |
| Інтерпретація: | Незначущий рівень показує, для конкретної змінної, пропорцію одиниць, для яких значення було введене через те, що початкове значення було відсутнє, неправдоподібне або суперечливе у порівнянні з кількістю одиниць зі значенням для цієї змінної. Одиниці з імпутацією пустого значення для виправлення невідповідної недійсної (непустої) відповіді (тип ііі) мають бути включені і в чисельник, і в знаменник.  Значущий рівень показує, для конкретної змінної, відносний внесок імпутованих значень в оцінку цього елемента/змінної. Очевидно, що цей значущий індикатор має значення, коли метою дослідження є оцінка загальної кількості або середнього значення змінної. Коли метою дослідження є оцінка складних індексів, значущий індикатор не має значення. |
| Особливі вказівки: | - |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). * Статистика Канади. Рекомендації з якості, 5-е видання - жовтень 2009 |

|  |  |
| --- | --- |
| **Назва:** | **А6. Спільні одиниці - пропорція** |
| Визначення: | Пропорція одиниць, охоплених як дослідженням так і адміністративними джерелами, по відношенню до загальної кількості одиниць у дослідженні. |
| Застосовність: | Пропорція застосовується до:   * змішаних статистичних процесів, де деякі змінні виходять з даних дослідження, а інші - з адміністративних джерел; * виробників та користувачів |
| Формула розрахунку: | *Ad* Кількість спільних одиниць серед даних дослідження та адміністративних джерел  Кількість унікальних одиниць у досліджуваних даних |
| Цільове значення: | - |
| Рівні та принципи агрегації: | - |
| Інтерпретація: | Індикатор використовується, коли адміністративні дані поєднуються з даними дослідження таким чином, що дані на рівні одиниць отримуються як з дослідження, так і з одного або більше адміністративних джерел (деякі змінні отримуються з досліджень, а інші змінні з адміністративних даних).  Індикатор надає уявлення про повноту/охоплення джерел –в якій мірі одиниці існують в адміністративних даних та в даних дослідження.  Цей індикатор не застосовується, якщо адміністративні дані використовуються лише для отримання оцінок без поєднання з даними дослідження. |
| Особливі вказівки: | Спільні одиниці відносяться до тих одиниць, які включені у дані, які виникають з адміністративних джерел та даних досліджень.  Для цілі цього індикатора, "унікальні одиниці в даних дослідження" у знаменнику означають, що якщо одиниця існує у більш ніж одному джерелі, вона має бути порахована лише один раз.  Якщо дослідження проводиться не для всіх одиниць в адміністративному джерелі (наприклад, проведення дослідження лише для великих підприємств), цей індикатор має бути розрахований для відповідного субнабору.  Помилки поєднання мають бути визначені та виправлені до розрахунку цього індикатора.  Якщо існує кілька спільних одиниць в результаті плану статистичного результату (наприклад, поєднання дослідження та адміністративних даних), це має бути пояснено. |
| Посилання: | ЄСС Результативне використання адміністративних та бухгалтерських даних у комерційній статистиці, WP6 Індикатори якості при використанні адміністративних даних у статистичних операціях, листопад 2010. |

|  |  |
| --- | --- |
| **Назва:** | **А7. Перегляд даних - середній обсяг** |
| Визначення: | Середня кількість переглядів ключового індикатора за період часу. "Перегляд" визначається як різниця між більш пізньою та більш ранньою оцінками ключового елементу.  Кількість випусків (*K*) ключового елементу (кількість разів його публікацій) є фіксованою та визначається у політиці переглядів. Зазвичай, перегляди включають часові ряди: під час публікації оцінки ключового індикатору, що відноситься до часу *t*, загальною практикою є випуск переглянутої версії індикатора, що відноситься до набору попередніх періодів.  У наступній таблиці ця ситуація проілюстрована для аналізу переглядів, де в політиці передбачено К переглядів та *n* звітнихперіодів, включених в аналіз.  Звітні періоди  Випуски 1 … *t* … *n*  1-й випуск *X* … *X* … *X*  11 1*t* 1*n*  … … … … … …  *k-й* випуск *X k*1 … *X kt* … *X kn*  … … … … … …  *K* та кінцевий випуск *X k*1 … *X kt* … *X kn*  Різні індикатори можуть бути отримані різними способами виведення середнього значення переглядів для часових рядів (перегляди можуть бути усереднені у абсолютних значеннях або ні, індикатор може бути абсолютним або відносним). |
| Застосовність: | Середній обсяг переглядів застосовується до:   * статистичних процесів, де початкова та наступна (переглянута) оцінки публікуються згідно з політикою переглядів (квартальні національні звіти, короткострокова статистика); * користувачів та виробників |
| Формула розрахунку: | З посиланням на двомірну ситуацію, описану у визначенні, існує кілька стратегій для розрахунку індикаторів: з або без знака, абсолютні або відносні значення, для конкретних пар переглядів за період часу або в межах послідовності переглядів тощо. Основною пропозицією є розглянути середнє значення для певного етапу перегляду в межах набору *n* звітних періодів.  **MAR (середній абсолютний перегляд):** |

|  |  |
| --- | --- |
|  | де:  *X* “пізніша” оцінка, *L*th випуск елемента на початку  *Lt*  відліку *t*;  *X* “раніша” оцінка, *P*th випуск елемента на початку  *Pt*  відліку *t*;  *n* = кількість оцінок (звітних періодів) у часовій послідовності враховується. *n* 20 рекомендується для квартальних оцінок, тоді як *n* 30 рекомендується для місячних оцінок. Цей індикатор не рекомендується для річних оцінок.  **MAR** надає уявлення про середній обсяг заданого етапу перегляду.  Цей індикатор може альтернативно бути виражений у відносних величинах:  **RMAR: (відносний середній абсолютний перегляд):**    Крім того – на рівні Євростату – і там, де знак викликає інтерес, існує середній перегляд від Випуску *P* до Випуску *L* за *n* звітних періодів:  **MR (середній перегляд):**    Різні комбінації *P* та *L* можуть бути розглянуті. Наприклад, ОЕСР пропонує порівнювати наступні випуски:  **Місячні дані** **Квартальні дані**  ***Випуск L*** ***Випуск P*** ***Випуск L*** ***Випуск P***  Через 2 місяці Перший Через 5 місяців Перший  Через 3 місяці Перший Через 1 рік Через 5 місяців Через 3 місяці Через 2 місяці Через 1 рік Перший  Через 1 рік Перший Через 2 роки Перший  Через 2 роки Перший Останній доступний Перший  Останній доступний Перший Через 2 роки Через 1 рік Через 2 роки Через 1 рік |
| Цільове значення: | - |
| Рівні та принципи агрегації: | * МS: індикатор має бути розрахований на рівні статистичних процесів * ЄС: індикатор розраховується по переглядам, зробленим для сукупності/індикатору ЄС. |

|  |  |
| --- | --- |
| Інтерпретація: | **MAR** надає уявлення про середній обсяг заданого етапу переглядудля ключового елементу за період часу**.**  Індикатор **RMAR** нормалізує міру MAR, використовуючи остаточні оцінки. Він полегшує міжнародні зіставлення та порівняння за періоди часу. При оцінці темпів зростання ця міра виправляє MAR щодо розміру зростання, і, таким чином, бере до уваги той факт, що перегляди ймовірно будуть більшими у періоди високого зростання, ніж у періоди повільного зростання.  Індикатори MAR і RMAR надають інформацію стосовно стабільності оцінок. Вони не надають інформацію щодо напрямків переглядів, оскільки розглядаються абсолютні значення переглядів. Така інформація надається за допомогою **MR**. Позитивний знак означає перегляд в бік підвищення (недооцінку), а негативний знак означає переоцінку у першому випадку. Іноді MR відносять до "середнього відхилення", але ненульовий MR не є достатнім для встановлення того, чи обсяг переглядів систематично відхиляється у певному напрямку. Для встановлення наявності відхилення, необхідно оцінити, чи є MR статистично відмінним від нуля (при відсутності змін у визначеннях, методологіях тощо). |
| Особливі вказівки: | MAR або RMAR мають бути представлені згідно з цим індикатором. Крім того, MR може також бути розрахований на рівні ЄС. |
| Посилання: | * ОЕСР: <http://stats.oecd.org/mei/default.asp?rev=1> |

|  |  |
| --- | --- |
| **Назва:** | **Т1. Проміжок часу - перші результати** |
| Визначення: | *Загальне визначення:*  Своєчасність статистичних результатів – проміжок часу між завершенням події або явища, які вони описують, та їх доступністю.  *Спеціальне визначення:*  Кількість днів (або тижнів або місяців) з останнього дня звітного періоду до дня публікації перших результатів. |
| Застосовність: | Цей показник застосовується до:   * всіх статистичних процесів з **випусками попередніх даних**; * користувачів та виробників.   T1 **не** застосовується до статистичних процесів з лише одним, безпосередньо остаточним набором результатів/статистичних даних – в такому випадку використовується лише T2. |
| Формула розрахунку: | *T*1 *d frst* *drefp*  *dfrst* … Дата випуску перших результатів;  *drefp*… Останній день (дата) звітного періоду статистики  *Одиниці вимірювання*: формат даних (календарні дні; якщо кількість днів велика, можлива конвертація у тижні або місяці)  Замість періоду, може використовуватися момент часу. |
| Цільове значення: | Цільові значення зазвичай визначаються законодавством або джентльменською угодою. Однак, менші значення означають більший високу своєчасність. |
| Рівні та принципи агрегації: | Розрахунок проводяться, для змістовного вибору, на рівні тематичного домену. Він може відноситися до поточного етапу виробництва або бути середнім за період часу. Агрегації можливі на рівні ЄС та на доменному рівні (наприклад, соціальна статистика, статистика підприємств). |
| Інтерпретація: | Цей індикатор представляє у кількісній формі проміжок часу між датою випуску перших результатів та датою звітного періоду для даних.  Порівняння можливо проводити між статистичними процесами з такою ж періодичністю. |
| Особливі вказівки: | Причини можливого тривалого періоду виробництва мають бути пояснені, та зусилля щодо поліпшення ситуації мають бути описані.  Для річної статистики або у випадках, коли своєчасність вимірюється роками, а не днями, достатньо речення, яке встановлює своєчасність. |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). |

|  |  |
| --- | --- |
| **Назва:** | **Т2. Проміжок часу - остаточні результати** |
| Визначення: | *Загальне визначення:*  Своєчасність статистичних результатів – проміжок часу між завершенням події або явища, яку вони описують, та їх доступністю.  *Спеціальне визначення:*  Кількість днів (або тижнів або місяців) з останнього дня звітного періоду до дня публікації перших та остаточних результатів. |
| Застосовність: | Цей показник застосовується до:   * всіх статистичних процесів * користувачів та виробників. |
| Формула розрахунку: | *T*1 *d frst* *drefp*  *dfrst* … Дата випуску остаточних результатів;  *drefp*… Останній день (дата) звітного періоду статистики  *Одиниці вимірювання*: формат даних (календарні дні; якщо кількість днів велика, можлива конвертація у тижні або місяці)  Замість періоду, може використовуватися момент часу. |
| Цільове значення: | Цільові значення зазвичай визначаються законодавством або джентльменською угодою. Однак, менші значення означають більший високу своєчасність. |
| Рівні та принципи агрегації: | Розрахунок проводяться, для змістовного вибору, на рівні тематичного домену. Він може відноситися до поточного етапу виробництва або бути середнім за період часу. Агрегації можливі на рівні ЄС та на доменному рівні (наприклад, соціальна статистика, статистика підприємств). |
| Інтерпретація: | Цей індикатор представляє у кількісній формі проміжок часу між датою випуску кінцевих результатів та датою завершення звітного періоду.  Порівняння можливо проводити між статистичними процесами з такою ж періодичністю. |
| Особливі вказівки: | Причини можливого тривалого періоду виробництва мають бути пояснені, та зусилля щодо поліпшення ситуації мають бути описані.  Для подальшого визначення тематичним доменом, з урахуванням політики переглядів, вони можуть бути розглянуті як "остаточні результати".  Для річної статистики або у випадках, коли звітний період вимірюється роками, а не днями, достатньо речення, яке встановлює своєчасність. |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). |

|  |  |
| --- | --- |
| **Назва:** | **Т3. Пунктуальність - доставка та публікація** |
| Визначення: | Пунктуальність – це проміжок часу між датою доставки/випуску даних та цільовою датою доставки/випуску, як узгоджено для доставки, або заявлено у офіційному календарі випусків, встановленого Регламентами або попередньо узгодженого між партнерами. |
| Застосовність: | Пунктуальність публікацій застосовується до:   * всіх статистичних процесів з встановленою/попередньо визначеною датою випуску, * всіх користувачів та виробників (різні аспекти та формули)   Підраховується лише Євростатом, але рекомендується до включення у національні звіти з якості. |
| Формула розрахунку: | **Для виробників:**  **Пунктуальність доставки даних Р3**  *P*3 *dact* *dsch*  dact .. Фактична дата ефективного надання статистики dsch…Запланована дата ефективного надання статистики  *Одиниці вимірювання*: формат даних (календарні дні)  **Для користувачів:**  **Рівень пунктуальності публікації даних P3R**  Відповідний для груп статистичних даних/результатів  **P3R**  – це рівень наборів даних, які відповідають даті в календарі випуску у групі наборів даних.    mpc… Кількість статистичних даних/результатів, які були опубліковані в день, заявлений у календарі або були випущені раніше (пунктуально)  mup… Кількість статистичних даних/результатів, які не відповідають даті, заявленій у календарі (не пунктуально) |
| Цільове значення: | Цільове значення для Р3 дорівнює 0, це означає, що не було затримки під час доставки/передачі даних.  Для P3R цільове значення дорівнює 1, це означає, що 100% елементів були опубліковані в день, попередньо визначений у календарі. |
| Рівні та принципи агрегації: | Існує два аспекти:   * Доставка національних даних до Євростату (орієнтовано на виробника), * Публікація/випуск Євростатом (орієнтовано на користувача), |

|  |  |
| --- | --- |
|  | Розрахунок проводиться на рівні статистичного процесу. Агрегації виконуються на рівні ЄС по країнам та доменам. |
| Інтерпретація: | Індикатор **Пунктуальність доставки даних** представляє у кількісній формі відмінність (проміжок часу) між фактичною та цільовою датою.  Це слід тлумачити відповідно до періодичності статистичного процесу.  Індикатор **Рівень пунктуальності** випуску (P3R) оцінює пунктуальність випуску групи конкретних наборів даних. |
| Особливі вказівки: | **Для виробників:**  З метою моніторингу відповідності цілям адміністратори доменів Євростату мають контролювати цей індикатор для окремих країн. Ця інформація може бути заповнена заздалегідь Євростатом, оскільки відомо, коли дані отримуються від держав-членів. У такому випадку необхідно застосувати формулу Р3.  Цей індикатор може бути представлений у форматі таблиці для різних держав-членів. Причини пізньої або не пунктуальної доставки мають бути зазначені разом з зазначенням їхнього впливу на статистичний продукт, мається на увазі, що через пізню доставку даних процедури забезпечення якості для всього продукту/серії можуть бути невиконані.  **Для користувачів:**  Достатньо сформувати цей індикатор як сумарне значення на рівні ESTAT. У такому випадку необхідно застосувати формулу P3R.  Слід надати користувачам деякі пояснення стосовно непунктуальних публікацій. |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). |

|  |  |
| --- | --- |
| **Назва:** | **АС1. Метадані - звернення** [**1**](#_bookmark0) |
| Визначення: | Кількість звернень до метаданих (ESMS) у межах статистичного домену за певний період часу.  Під "кількістю звернень" мається на увазі кількість разів перегляду файлу метаданих.  Деяка інформація доступна через щомісячний звіт з моніторингу у Електронному розповсюдженні Євростату та у файлах excel з докладними цифрами. |
| Застосовність | Цей показник застосовується до:   * всіх статистичних процесів * виробників (адміністраторів домену Євростату) та користувачів.   Підраховується лише Євростатом. |
| Формула розрахунку: | AC1 = # *ESMS*  де # *ESMS* позначає абсолютну кількість елементів у наборі ESMS (це так звана потужність набору). У такому випадку набір ESMS представляє файли ESMS, до яких зверталися щодо конкретного тематичного домену за певний період часу.  Примітка: внутрішні перегляди сторінки виключаються. |
| Цільове значення: | Не існує негайної інтерпретації низьких та високих значень цього індикатора, також не існує специфічної цілі. |
| Рівні та принципи агрегації: | Розрахунок проводиться на рівні статистичного процесу. Агрегація можлива на наступних рівнях:   * Конкретні файли ESMS доменів. * Річна агрегація.   Принцип полягає у підрахунку кількості звернень до файлів ESMS тематичними доменами. |
| Інтерпретація: | Індикатор сприяє оцінці потреб користувача у метаданих (рівень інтересу), для оцінки відповідності тематичних доменів.  Співвідношення може бути розраховане з метою надання уявлення про пропорції звернень до файлів ESMS, про які йде мова, у порівнянні з загальною кількістю звернень для усіх доменів. |
| Особливі вказівки: | Інформативний та простий спосіб репрезентації результату цього індикатора полягає у внесенні цифр за період часу в графік. Зокрема, це має бути графік, де горизонтальна вісь (х) представлятиме місяці, а вертикальна вісь (у) представлятиме кількість файлів ESMS, до яких були звернення. Стане можливим проводити моніторинг інтересу користувачів до кожного файлу ESMS на конкретному рівні домену.  Графік кількості звернень до таблиць даних (індикатор АС2) та файлів метаданих (ESMS) з відповідністю, з правильним налаштуванням, буде цікаво відобразити, з часом. |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). |



1 Індикатор має збиратися спільно з Одиницею D4 - Розповсюдження.

|  |  |
| --- | --- |
| **Назва:** | **АС2. Таблиці даних – звернення** [**2**](#_bookmark1) |
| Визначення: | Кількість звернень до таблиць даних у межах статистичного домену за певний період часу.  Під "кількістю звернень" мається на увазі кількість переглядів таблиць даних, де багаторазові перегляди під час однієї сесії враховуються лише один раз. Деяка інформація доступна через щомісячний звіт з моніторингу у Електронному розповсюдженні Євростату та у файлах excel з докладними цифрами. |
| Застосовність: | Кількість звернень до таблиць даних застосовуються до:   * всіх статистичних процесів з використанням он-лайн таблиць даних для розповсюдження статистики; * виробників (адміністраторів домену Євростату) та користувачів   Підраховується лише Євростатом, але рекомендується до включення у національні звіти з якості. |
| Формула розрахунку: | AC2 = # *CONS*  де # *CONS* позначає абсолютну кількість елементів у наборі CONS (це так звана потужність набору). У цьому випадку CONS представляє звернення до таблиць даних для конкретного тематичного домену.  Частота збору кількісних даних для цього індикатора повинна бути щомісячною.  Примітка: внутрішні перегляди сторінки виключаються. |
| Цільове значення: | Не існує негайної інтерпретації низьких та високих значень цього індикатора, також не існує специфічної цілі. |
| Рівні та принципи агрегації: | Розрахунок проводиться на рівні статистичного процесу. Агрегація можлива на наступному рівні:   * Спеціальні таблиці даних доменів. * Щорічна агрегація.   Принцип полягає у підрахунку кількості звернень до таблиць даних тематичними доменами. |
| Інтерпретація: | Цей індикатор має бути ретельно проаналізований та об’єднаний з іншою інформацією, яка доповнить аналіз.  Індикатор сприяє оцінці потреб користувача у метаданих (рівень інтересу), для оцінки відповідності тематичних доменів.  Співвідношення може бути розраховане з метою надання уявлення про пропорції звернень до файлів ESMS, про які йде мова, у порівнянні з загальною кількістю звернень для всіх доменів. |
| Особливі вказівки: | Інформативний та простий спосіб репрезентації результату цього індикатора полягає у внесенні цифр за період часу в графік. Зокрема, це має бути графік, де горизонтальна вісь (х) представлятиме місяці, а вертикальна вісь (у) представлятиме кількість наборів даних, до яких були звернення. Стане можливим проводити моніторинг інтересу користувачів до |



2 Індикатор має збиратися спільно з Одиницею D4 - Розповсюдження.

|  |  |
| --- | --- |
|  | кожного набору даних на рівні конкретного домену. Графік кількості звернень до таблиць даних та файлів ESMS (АС1), з правильним налаштування, буде цікаво відобразити. |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). |

|  |  |
| --- | --- |
| **Назва:** | **АС3. Повнота метаданих - рівень**[**3**](#_bookmark2) |
| Визначення: | Співвідношення кількості наданих елементів метаданих до загальної кількості застосовних елементів метаданих. |
| Застосовність: | Рівень повноти метаданих застосовується до:   * всіх статистичних процесів; * виробників (адміністраторів домену Євростату) та користувачів;   Підраховується лише Євростатом, але рекомендується до включення у національні звіти з якості. |
| Формула розрахунку: | *L* у знаменнику – це набір застосовних елементів метаданих, які розглядаються, а *M L* у числівнику – це субнабір *L* доступних елементів метаданих. Позначка # *L* означає кількість елементів у наборі *L* (кількісний числівник). Літера С в лівій частині формули означає як країни ЄС так і країни ЄАВТ.  Набір *L* отримується шляхом розрахунку для групи елементів метаданих, як пояснюється нижче на географічну одиницю (MS або ЄС+ЄАВТ), статистичний домен тощо.  Існує три групи метаданих, як описано нижче, разом з категоризацією, використовуючи поточні поняття EURO-SDMX  (у наступну розбивку включені лише головні поняття).   1. Метадані щодо статистичних результатів; поняття 3, 4, 5, 8.1, 9, 10; 2. Метадані щодо статистичних процесів;   поняття 11, 20.1, 20.2, 20.3, 20.4, 20.5, 20.6;   1. Метадані щодо якості: поняття 12-19   Розрахунки проводяться окремо для кожної з трьох груп і для кожної комбінації (група метаданих, рівень ЄС тощо). |
| Цільове значення: | Цільове значення дорівнює 1, мається на увазі, що 100% метаданих доступні з того, що вимагається/застосовується для статистичного процесу, або сумарна кількість, що досліджується. |
| Рівні та принципи агрегації: | Розрахунок проводиться на рівні файлів ESMS. Агрегації можливі на рівні МS, ЄС та на доменному рівні (наприклад, соціальна статистика, статистика підприємств).  Принцип полягає у розрахунку індикаторів як незначущий рівень на рівні МS та ЄС для статистичного домену (соціальна статистика, статистика підприємств тощо). |



3 Звірити з B6, чи можливо провести розрахунки централізовано

|  |  |
| --- | --- |
| Інтерпретація: | Кожний індикатор показує, в якій мірі метадані певного типу доступні у порівнянні з тим, як вони мають бути доступним.  Цей індикатор має бути ретельно проаналізований, оскільки цей рівень відображає лише існуючу кількість метаданих для певного статистичного процесу, а не якість тієї інформації. |
| Особливі вказівки: | Всю інформацію повинна бути вилучена з файлів ESMS.  У випадку, якщо ESMS порожній, для різних категорій, зазначених раніше, відсутня необхідність проводити розрахунок, але має бути замінений описовий текст.  Стосовно Євростату, існує можливість мати прямий доступ до тих файлів через веб-сайт Євростату, тоді як для МS буде можливим отримати доступ до файлів ESMS у найближчому майбутньому, через національну програму обслуговування реєстру та запитів.  Слід враховувати, що насправді означає доступність метаданих. |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). * Структура метаданих EURO-SDMX, редакція від березня 2009. |

|  |  |
| --- | --- |
| **Назва:** | **СС1. Довжина порівнюваних часових рядів** |
| Визначення: | Кількість звітних періодів у часовому ряді з моменту останньої перерви.  *Коментар*  Перерви у статистичному часовому ряді можуть виникати, коли відбувається зміна у визначенні параметру для оцінки (наприклад, змінна або сукупність) або методології, що використовується для проведення оцінки. Іноді перерву можна попередити, наприклад, шляхом з’єднання. |
| Застосовність: | Довжина порівнюваних часових рядів застосовується до:   * всіх статистичних процесів, що створюють часові ряди; * користувачів та виробників   Підраховується лише Євростатом, але рекомендується до включення у національні звіти з якості. |
| Формула розрахунку: | Звітні періоди нумеруються.  *CC*1 *Jlast* *J first* 1  *Jlast* … номер останнього звітного періоду з розповсюдженою статистикою.  *Jfirst* … номер першого звітного періоду з порівнянною статистикою. |
| Цільове значення: | Довгострокові періоди можуть здаватися бажаними, але це може бути мотивованим для внесення змін, наприклад, оскільки реальність мотивує появу нових понять, або ж для того, щоб досягти узгодженості з іншою статистикою. |
| Рівні та принципи агрегації: | Розрахунок проводиться на рівні статистичного процесу. Агрегація можлива на рівні МS, ЄС та на доменному рівні (наприклад, соціальна статистика, статистика підприємств).  Індикатор для рівня ЄС або доменного рівня має бути розрахований Євростатом, з урахуванням часового ряду сукупності ЄС. |
| Інтерпретація: | Якщо не було перерв, індикатор дорівнює кількості моментів часу у часовому ряді. |
| Особливі вказівки: | Довжина ряду з порівнюваною статистикою виражається як кількість часових періодів (моментів) у цьому ряді. Вона відраховується з першого періоду часу зі статистикою після перерви. Результат не залежить від довжини звітного періоду.  Застосовується тільки для статистичних даних, розповсюджених у послідовності регулярних часових періодів (моментів).  Якщо існує більше одного ряду для одного статистичного процесу, адміністратор домену має обрати прийнятні для розрахунку ряди. |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). |

|  |  |
| --- | --- |
| **Назва:** | **СС2. Асиметрія для дзеркальної статистики потоків - коефіцієнт** |
| Визначення: | *Загальне визначення:*  Розбіжності між даними, що стосуються потоків, наприклад, для пар країн.  *Спеціальне визначення (надається кілька версій) Двостороння дзеркальна статистика:*  Різниця або абсолютна різниця вхідних та вихідних потоків між парами країн, розділена на середнє з цих двох значень.  *Коментар*  Вхідні та вихідні потоки мають розглядатися як будь-які потоки, характерні для кожного тематичного домену (кількість проданих товарів, кількість людей, які відвідали країну з туристичними цілями тощо). |
| Застосовність: | Асиметрії дзеркальної статистики потоків застосовуються до:   * доменів, у яких доступна дзеркальна статистика (потоки, що стосуються торгівлі, міграції, статистики туризму, ЄБРР, платіжного балансу тощо). * користувачів (підлягає обговоренню) та виробників Підраховується Євростатом (попередньо заповнюється у звіті з якості) |
| Формула розрахунку: | **Двостороння дзеркальна статистика:**  Для кожної пари країн, припустимо: А - Країна А  В - Країна В    Спільна міра може бути отримана з двох відмінностей по відношенню до середнього потоку (кілька можливостей, одна надається нижче):    OFAB - вихідний потік, що йде з країни А до країни В  m IFAB – дзеркальний вхідний потік  IFBA - дзеркальний вхідний потік у країну В з країни А m OFAB - дзеркальний вихідний потік  **Багатостороння дзеркальна статистика:**  OFAiOj - вихідний потік, що йде з країни Ai у будь-яку іншу країну Oi  mIFAiOj – дзеркальний вхідний потік |

|  |  |
| --- | --- |
|  | Ai – країна Ai  Oj – інша країна Oj  K – кількість країн, з якими країна Ai може мати контакти  C – група країн ЄС + ЄАВТ |
| Цільове значення: | Значення цього індикатора має бути максимально наближеним до нуля, оскільки - принаймні в теорії - значення вхідного та вихідного потоків між парами країн мають співпадати. |
| Рівні та принципи агрегації: | * MS: Розрахунок проводиться для ключових змінних/субрядів, які обираються адміністратором домену Євростату. * ЄС: Агрегація можлива на рівні ЄС (див. формули багатосторонньої дзеркальної статистики). З іншого боку, коли наприклад доступна не вся інформація, нижнє та верхнє значення двосторонньої дзеркальної статистики можуть бути повідомлені для відображення діапазону. |
| Інтерпретація: | У доменах, де доступна дзеркальна статистика, можливо оцінити географічну порівнянність, що вимірює невідповідності між вхідними та вихідними потоками для пар країн.  Дзеркальні дані можуть допомогти перевірити узгодженість подання даних, даних, процесу звітування та використаних визначень. Зрештою, вони можуть допомогти оцінити відсутні дані. Для користувачів індикатори асиметрії надають певне уявлення стосовно загальної надійності даних.  Існує ідеальна симетрія (вихідні потоки дорівнюють дзеркальним вхідним потокам), коли коефіцієнт дорівнює нулю. Чим більше коефіцієнт відхиляється від нуля, тим більш важливою стає асиметрія між вихідними потоками та дзеркальними вхідними потоками. |
| Особливі вказівки: | Індикатори СC2AB та CC2BA можуть бути негативними та позитивними. Індикатор СC2AB завжди позитивний.  Вихідні потоки від держави-члена А до держави-члена В, як повідомляє А мають практично дорівнювати вхідним потокам у В, що йдуть від А, як повідомляє В. Оскільки деякі домени використовують різні принципи оцінювання, вхідні потоки можуть незначною мірою відрізнятися від вихідних потоків. Таким чином, порівняння, які стосуються дзеркальної статистики, мають проводитися обережно і враховувати існування цих розбіжностей.  Коефіцієнт асиметрії СC2AB є корисним, оскільки його можна буде контролювати з плином часу.  Індикатори СC2AB та CC2BA можуть бути як позитивними, так і негативними, і можуть використовуватися для оцінки того, наскільки країна глобально декларує верхній або нижній потоки, у порівнянні з дзеркальними потоками, задекларованими країнами-партнерами.  Індикатори СC2AB та CC2BA мають бути представлені у таблиці (наприклад, статистика зовнішньої торгівлі). |
| Посилання: | * ЄСС Довідник для звітів з якості - вид. 2009 (Євростат). * ЄСС Стандарт для звітів з якості - вид. 2009 (Євростат). * Міжнародна торгівля у статистиці надання послуг - Моніторинг прогресу з імплементації Посібника та оцінювання якості даних – ОЕСР Експертна нарада Євростату з питань торгівлі послугами, 2005 |