

О. В. Гончар,

к. е. н., с. н. с., в. о. завідувача кафедри теорії статистики, Національна академія статистики, обліку та аудиту, завідувач відділу, НТК статистичних досліджень

ПИТАННЯ КЛАСИФІКАЦІЇ ЗМІННИХ, ЩО ВИКОРИСТОВУЮТЬСЯ ПРИ ОЦІНЮВАННІ ЯКОСТІ СТАТИСТИЧНИХ ОБСТЕЖЕНЬ

У статті автором запропоновано класифікувати процесні змінні, що використовуються при оцінюванні якості процесів виробництва статистичної інформації за типами статистичних процесів. Розроблено класифікацію процесних змінних для статистичних обстежень.

In the paper author offers to classify of process variables used for quality estimation of statistical information production processes by statistical process type. The classification of process variables for statistical survey is developed.

Ключові слова: оцінювання якості, процесна змінна, статистичне обстеження, якість статистичної інформації.

ВСТУП

Виробники статистичної інформації намагаються довести користувачам, що вироблена ними інформація має високу якість, що забезпечується шляхом контролю якості на усіх етапах її виробництва. Користувачі статистичної інформації передусім зацікавлені у якості статистичної інформації та метаданих, які супроводжують цю інформацію, для правильної її інтерпретації. Процес виробництва інформації цікавить користувачів не сам по собі, а лише з погляду на його впливу на якість статистичної інформації. Для самих виробників статистичної інформації якість процесів виробництва має вирішальне значення, оскільки вони усвідомлюють її величезний вплив на якість статистичної інформації.

Досягти високої якості можна шляхом ідентифікації ключових процесних змінних, вимірювання їхніх величин, регулювання процесу, який базується на цих вимірюваннях, та перевірки зміни якості статистичної інформації. Процесні змінні використовуються як індикатори якості процесів, індикатори якості статистичної інформації або для обчислення індикаторів якості статистичної інформації.

Варто зазначити, що до 90-х років ХХ ст. національні статистичні служби країн не займалися систематичним збиранням та аналізом даних щодо процесів виробництва, передусім через відсутність методології вимірювання процесних змінних. Проте з часом розпочався активний розвиток цього напрямку досліджень. Питаннями забезпечення якості процесів виробництва статистичної інформації займалися такі зарубіжні вчені, як П. Бімер, А. Ліберг, Д. Моргенштейн, Д. Маркер та інші [1; 2]. В Україні окремі спроби здійснювались фахівцями Державної служби статистики, однак значних успіхів не досягли через відсутність підрозділу, відповідального за питання якості. Нерозробленими залишаються питання класифікації процесних змінних, чому й буде присвячена дана стаття.

ПОСТАНОВКА ЗАДАЧ

Метою статті є розроблення класифікації процесних змінних, що використовуються при оцінюванні якості процесів виробництва статистичної інформації, яка отримується зі статистичних обстежень.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

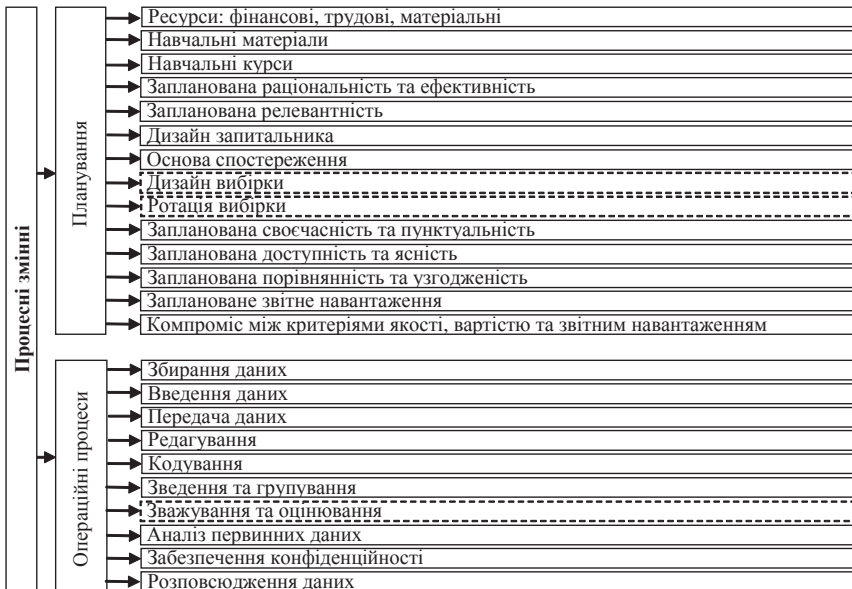
Як зазначалося в [3], статистична інформація виробляється у різних за сутністю статистичних процесах (суцільних та вибіркового обстеженнях, процесах, що викорис-

товують адміністративні дані, дані з багатьох джерел, обчислення економічних індексів, статистичні компіювання, статистичне моделювання та прогнозування). Залежно від типу статистичного процесу відрізняються і процеси виробництва інформації, які є підпроцесами статистичного процесу. Спільною ознакою всіх статистичних процесів є наявність у виробничому циклі статистичного процесу ланок планування статистичного процесу та перегляду методології і організації. Проте очевидно, що виконувани на цих ланках виробничі підпроцеси відрізнятимуться за типами статистичних процесів. Ланка архівування також є спільною для усіх статистичних процесів, але архівування має незвичний зв'язок із якістю статистичної інформації, і тому ми його не розглядаємо.

На проміжних ланках між плануванням та переглядом методології і організації виконуються операційні виробничі процеси. Якість статистичної інформації створюється в усіх ланках виробничого циклу статистичного процесу, хоча ступінь впливу окремих ланок на її якість є різним. Саме тому для покращення якості статистичної інформації необхідно регулярно вимірювати ключові процесні змінні (проводити моніторинг) та покращувати якість усіх виробничих процесів, задіяних у циклі виробництва статистичної інформації. З цієї ж причини процесні змінні повинні обчислюватися не лише при збиранні та обробленні статистичних даних, а й при плануванні процесів виробництва статистичної інформації.

Планування виробництва статистичної інформації базується на відповідних потребах і вимогах користувачів. На етапі планування передусім необхідно дізнатись про потреби та вимоги користувачів і відповідно до них визначити обсяг і зміст статистичної інформації та сформулювати завдання дослідження. Потреби та вимоги користувачів статистичної інформації визначаються шляхом проведення обстежень потреб та вимог користувачів, а також національними і міжнародними нормативами і стандартами. Далі необхідно визначити основу сукупності, використовувани концепції, статистичні показники, класифікації, деталізацію даних тощо.

Після цього визначається тип статистичного процесу, в якому вироблятиметься статистична інформація. Якщо це статистичне обстеження, то мають бути вирішені стратегічні питання щодо методів збирання та оброблення даних, зокрема методів редагування та імпуації. Для цього розробляється дизайн запитальника та проводиться його тестуван-



* курсивом позначені блоки процесних змінних, характерні лише для вибіркового обстеження.

Рис. 1. Процесні зміни у статистичних обстеженнях*

ня, визначаються кошти на проведення обстеження, проводиться навчання інтерв'юєрів, вирішуються організаційні питання. Розробляється також детальний план проведення кожного окремого етапу обстеження з визначенням усіх стадій, кількість яких залежить від специфіки обстеження.

Як вже зазначалось, усі процеси впливають на якість статистичної інформації, зокрема процеси планування, тому моніторинг має передбачати вимірювання процесних змінних усіх процесів планування.

У роботі [4] розглядаються такі основні блоки процесних змінних у плануванні статистичних обстежень: фінансові ресурси, дизайн вибірки, дизайн запитальника, навчальні тренінги, оновлення основи вибірки. На наш погляд, для ефективного планування якості статистичної інформації цього недостатньо, і тому ми розробили власну класифікацію процесних змінних у статистичних обстеженнях (рис. 1).

Якість статистичних даних прямо залежить від фінансових ресурсів, кваліфікованих кадрів та матеріально-технічного забезпечення, а тому процесні зміни, які характеризують ці аспекти, є водночас індикаторами якості процесів виробництва статистичної інформації.

Для виробництва статистичної інформації високої якості на етапі планування обстеження необхідно оцінити потребу в кадрах відповідної кваліфікації. У статистичних обстеженнях зазвичай беруть участь не лише статистики, а й інші фахівці, зокрема інтерв'юєри, науковці, управлінці, аналітики, програмісти, оператори ЕОМ, технічний персонал тощо. Тому необхідно визначити процесні зміни, які характеризують необхідність у кадрах конкретної кваліфікації, наявність відповідних кадрів та їхні кваліфікаційні характеристики, а також ступінь відповідності кваліфікаційних характеристик наявних кадрів необхідним кваліфікаційним вимогам.

Важливо також оцінювати процесні зміни, які характеризують матеріально-технічне забезпечення, зокрема обчислювальну техніку, засоби зчитування даних, транспортні засоби, статистичний інструментарій, друкарські засоби тощо.

Одним з основних індикаторів якості, який стосується фінансового забезпечення, є індикатор фінансування, який розраховується як відношення наявних фінансових ресурсів до необхідних для забезпечення високої якості. Можна розраховувати аналогічні індикатори і за окремими категоріями матеріальних і людських ресурсів.

Як відомо, вартість обстеження прямо залежить від кількості безпосередньо обстежуваних одиниць, а тому можна визначити ще один індикатор, який стосується фінансового забезпечення, — відношення максимального за наявних ресурсів обсягу вибірки до оптимального її обсягу. Для прикладу можна навести структурне спостереження підприємств у Греції. Відповідно до потреб користувачів у цьому обстеженні має бути забезпечена висока

точність на рівні чотирьох знаків Європейської класифікації видів економічної діяльності (NACE) для кожного регіону на рівні NUTS II (відносна вибіркова похибка не перевищує 5%). Було доведено, що для досягнення такої точності необхідно обстежити 8650 підприємств. Проте фактичне фінансування дозволяло обстежити лише 6000 підприємств. Враховуючи наявні кошти, вибірку було модифіковано таким чином, щоб забезпечити точність на рівні трьох знаків NACE для кожного регіону на рівні NUTS II (відносна вибіркова похибка не перевищує 8%). Таким чином, індикатор фінансування дорівнює 6000 / 8500, тобто 0,7 [4].

Планування процесів виробництва статистичної інформації передбачає також планування їхньої результативності та ефективності, які водночас є важливими критеріями якості процесів виробництва статистичної інформації.

Важливими елементами забезпечення якості статистичної інформації є посібники з якості та навчальні курси для фахівців. Основними процесними змінними, пов'язаними з навчальними тренінгами та посібниками з якості, є:

- кількість семінарів для навчання інтерв'юєрів;
- час, витрачений на семінари з метою навчання інтерв'юєрів;
- кількість навчальних посібників з якості.

При плануванні якості необхідно обчислити планові індикатори якості, зокрема за такими аспектами якості, як релевантність, своєчасність та пунктуальність, доступність та ясність, порівнянність та узгодженість, які водночас є процесними змінними. При оцінюванні якості виробленої статистичної інформації ці індикатори можна використати для оцінювання досягнення запланованої якості статистичної інформації. Зауважимо, що спланувати такий критерій якості, як точність, можна лише в частині вибіркової похибки. Оскільки вибіркової похибки залежать від дизайну вибірки, то процесні зміни, які їх характеризують, належать до блоку дизайну вибірки.

При плануванні релевантності важливе значення мають питання, пов'язані з використанням при розробленні статистичної інформації статистичних класифікацій та концепцій. Використовувані статистичні класифікації повинні бути релевантними до вимог ключових користувачів статистичної інформації та мати високу якість. Тому варто визначити процесні зміни якості класифікацій та їх відповідності вимогам користувачів.

Використовувані при виробництві статистичної інформації концепції мають наукову та статистичну основу і застосовуються: 1) для визначення сутності ознак спостереження; 2) для визначення цільової сукупності та статистичних одиниць.

У блоках запланованої релевантності і порівнянності та узгодженості варто обчислювати також процесні зміни, які характеризують відповідність класифікацій та концепцій міжнародним нормативам та стандартам, що свідчить про релевантність вимогам користувачів за межами України, так і про порівнянність даних з іншими країнами. Також можна обчислити процесні зміни відповідності методів, визначень, процедур загальноприйнятим у міжнародній практиці (якщо такі існують).

Якість збирання первинних статистичних даних прямо залежить від дизайну, використовуваного в обстеженні запитальника. Обчислення процесних змінних, які характеризують дизайн запитальника, варто виконувати за такими аспектами, як:

- легкість введення та кодування даних;
- орієнтовний час на заповнення та оброблення запитальника;
- складність заповнення запитальника (наявність даних первинного обліку тощо);
- вартість запитальника, введення та оброблення даних для одного запитальника;
- обсяг процедур редагування та імпутації даних;
- термінологічна узгодженість відповідей зі стандартними концепціями та визначеннями.

Основними процесними змінними, пов'язаними з дизайном запитальника, є:

- довжина запитальника — визначається як кількість запитань, поділена на кількість обстежуваних ознак (визначається зазвичай у відсотках);
- частка відкритих запитань, поділена на загальну кількість запитань;
- очікуваний середній час інтерв'ювання (час, витрачений на заповнення запитальника);
- запланований час інтерв'ювання в розрахунок на одного інтерв'юера.

До цього блоку варто включити також процесні змінні, отримані при тестуванні запитальника з метою оцінювання зрозумілості запитань респондентами.

Забезпечення релевантності та точності статистичних даних потребує використання коректної основи спостереження для: 1) розроблення дизайну спостереження; 2) збирання даних; 3) оцінювання статистичних показників у вибіркових обстеженнях.

Як у суцільних, так і у вибіркових обстеженнях важливе значення для виробництва статистичної інформації високої якості мають охоплення, повнота, свіжість, зміст та точність основи спостереження. При формуванні основи спостереження визначають процесні змінні, які характеризують відповідність основи спостереження цільовій сукупності. Оскільки основи спостережень зазвичай формуються із статистичних або адміністративних реєстрів, тому визначають процесні змінні якості реєстру (наприклад, індикатор, запропонований у роботі [5]). Для розроблення дизайну вибіркового обстеження основа вибірки повинна містити також додаткову інформацію, яка може використовуватись для стратифікації, виділення нетипових одиниць тощо. Тому визначають процесні змінні, що характеризують якість додаткової інформації, яка міститься в основі вибірки, а також якість контактної інформації, зокрема її свіжість.

Велике значення для отримання якісних даних у вибіркового обстеження мають дизайн та ротація вибірки. Розроблення дизайну вибірки передбачає визначення виду вибіркового обстеження, визначення кількості ступенів відбору, визначення критеріїв стратифікації/ кластеризації, визначення схеми та процедури відбору, методів оцінювання тощо. Ротація проводиться з метою оновлення вибіркової сукупності та зниження звітного навантаження на респондентів. На кожному етапі розроблення дизайну вибірки та її ротації визначаються власні процесні змінні.

Основними процесними змінними, пов'язаними з дизайном вибірки, є:

- кількість критеріїв стратифікації;
- розмір страт;
- обсяг вибірки у цілому та за стратами;
- частка відбору;
- кількість та частка нетипових одиниць, виявлених на етапі планування вибіркового обстеження (за даними аналогічного обстеження, обстеження попереднього року або пробного обстеження);
- середні значення та дисперсії ознак у цілому та за стратами (за даними аналогічного обстеження, обстеження попереднього року або пробного обстеження);
- коефіцієнти варіації ознак у цілому та за стратами (за даними аналогічного обстеження, обстеження попереднього року або пробного обстеження);
- кінцева кількість страт;
- ваги дизайну вибірки;
- дизайн-ефект;
- очікувана вибірка похибка у цілому та за стратами.

Процеси планування статистичного обстеження, особливо вибіркового, зазвичай виконуються у вигляді ітераційних процедур, а тому процесні змінні, які входять до одного блоку, можуть визначатися у різних процесах планування. Крім того, більшість процесних змінних визначається повторно на кожній ітерації до моменту досягнення компромісу між запланованими критеріями якості, вартістю та звітним навантаженням на респондентів.

Оскільки між різними критеріями якості статистичної інформації існує конфлікт, на етапі планування необхідно описати в термінах процесних змінних компроміс між ними, якого планують досягти органи державної статистики опираючись на виробництво статистичної інформації. Операційні процеси в статистичних обстеженнях включають процеси збирання, оброблення (введення, передача, редагування, кодування), зведення та групування / поширення (звжу-

вання та оцінювання), аналізу та розповсюдження, забезпечення конфіденційності статистичних даних.

У статистичній практиці використовуються такі форми збирання даних:

- інтерв'ю (опитування респондента інтерв'юером вічна-віч);
- поштове опитування (надсилання бланку запитальника та отримання заповненого запитальника від респондента поштою). Модифікацією поштового опитування є передача бланку запитальника та отримання заповненого запитальника від респондента факсом;
- телефонне опитування (опитування респондента інтерв'юером по телефону);
- заповнення запитальника респондентом в електронному вигляді та його повернення в статистичний орган за допомогою електронної пошти;
- он-лайн-опитування;
- подання заповненого запитальника респондентом безпосередньо в статистичний орган.

Зазначені форми поділяються на дві групи: 1) з участю інтерв'юера та 2) без безпосередньої участі інтерв'юера (самозаповнення). До першої групи належить інтерв'ю та телефонне опитування. Проміжним варіантом є подання заповненого запитальника респондентом безпосередньо в статистичний орган. У цьому разі інтерв'юером є фахівець, який приймає заповнені запитальники. Хоча він сам не проводить опитування, проте надає консультації із заповнення запитальника та перевіряє вже заповнені запитальники, вказуючи, за необхідності, на недоліки, які потрібно виправити.

Невідповіді в обстеженнях та похибки охоплення спричиняють зміщення, вплив якого є невідомим. Для елімінації зміщення, спричиненого похибками охоплення та невідповідей, застосовується зважування. Ваги є випадковими змінними, а тому вони приводять до зростання дисперсії оцінок показників у обстеженні. Індикаторами якості процесів збирання та зважування даних, які характеризують зміщення показників, спричинене невідповідями та неякісним охопленням цільової сукупності, можуть бути показники, які показують вплив зважування на дисперсію оцінки показника та дизайн-ефекти.

Основними процесними змінними, пов'язаними зі збиранням даних, можуть бути:

- час інтерв'ювання (час, витрачений на заповнення запитальника);
- кількість інтерв'юерів;
- час проведення обстеження;
- частка відповідей;
- частка повних відповідей;
- частка відмов;
- частка тимчасово відсутніх (у обстеженнях домогосподарств);
- частка часткових невідповідей;
- частки невідповідей за причинами;
- частка одиниць поза полем обстеження;
- частка невідповідей поза полем обстеження;
- частка одиниць, включених до обстеження кілька разів;
- частка одиниць з неправильно заповненими значеннями певної ознаки (наприклад, з позначкою "так" і "ні" одночасно, з кількісною величиною замість позначки, з позначками у числовому полі, з більш ніж однією позначкою для ознаки, яка допускає позначку лише для однієї відповіді на запитання тощо).

Основними процесними змінними, які можуть бути визначені при введенні даних, є:

- час, витрачений на введення даних;
- кількість операторів, які вводили дані;
- частка неправильно введених значень ознаки (значення ознаки неправильне, розташоване не на належному місці тощо);
- частка неправильно розпізнаних значень ознаки (при скануванні).

При передачі файлів даних можуть бути визначені такі основні процесні змінні:

- час передачі даних;
- кількість файлів даних за форматами;
- кількість записів та полів у файлах до і після передачі;
- зміна показників до і після передачі.

Основними процесними змінними, які можуть бути обчислені при редагуванні, є:

- час, витрачений на редагування даних;
- кількість фахівців, які проводили редагування даних;

- кількість виявлених при редагуванні похибок;
- питома вага похибок, виявлених при автоматизованому редагуванні;
- питома вага похибок, виявлених при ручному редагуванні;
- питома вага скоригованих на етапі редагування похибок;
- вплив редагування на оцінки показників;
- час, витрачений на імпутацію;
- кількість ознак, за якими проводилася імпутація;
- частка імпутованих значень для кожної ознаки;
- частка запитальників, у яких імпутовані значення усіх ознак;
- вплив імпутації на оцінки показників.

Основними процесними змінними, пов'язаними з кодуванням, є:

- час, витрачений на кодування даних;
- частка відкритих запитань, поділена на загальну кількість запитань;
- частка невірних кодувань;
- час, витрачений на виявлення та виправлення невірних кодів;
- частка скоригованих кодів у загальній кількості перевірених кодів.

При зведенні та групуванні можна визначити такі процесні змінні:

- час, витрачений на зведення та групування даних;
- кількість використаних стандартних статистичних класифікацій;
- кількість групувальних ознак;
- кількість утворених груп;
- межі утворених груп;
- ширина інтервалів групування;
- частотні характеристики груп;
- коефіцієнти варіації ознак за групами.

Для компенсації відсутніх відповідей у обстеженнях, виявлених на етапі оброблення даних нетипових одиниць, та забезпечення узгодженості з даними з інших джерел використовують зважування даних. Основними процесними змінними, пов'язаними зі зважуванням та оцінюванням, є:

- час, витрачений на автоматизоване зважування;
- час, витрачений на ручну перевірку ваг;
- час, витрачений на оцінювання та табуляцію;
- базові та підсумкові ваги;
- коефіцієнти перезважування;
- кількість виділених на етапі оцінювання нетипових одиниць;
- зміна оцінок показників через виділення нетипових одиниць;

- середньоквадратичні похибки;
- стандартні похибки;
- довірчі інтервали;
- зміщення, спричинене невідповідями.

Аналіз первинних результатів передбачає:

- макроредагування для виявлення похибок шляхом:
 - 1) просторових та часових порівнянь,
 - 2) дослідження узгодженості з показниками для тих самих соціально-економічних доменів;
 - 3) аналізу неочікувано високих вибірових похибок;
- подальші коригування виявлених похибок (за можливістю);
- табулювання.

Основними процесними змінними, пов'язаними з аналізом первинних даних, є:

- частка похибок, виявлених та виправлених на макрорівні, у загальній кількості записів;
- час, витрачений на макроредагування;
- відношення похибок, спричинених невірними вагами, до сумарної кількості записів.

Забезпечення конфіденційності має важливе значення для налагодження взаємовигідних відносин з респондентами. Загалом забезпечення конфіденційності передбачає:

- ідентифікацію вимог, потреб та пріоритетів користувачів;
- визначення інформації, оприлюднення якої потенційно може призвести до непрямой ідентифікації одиниць;
- застосування методів захисту конфіденційності;
- документування.

Основними процесними змінними, пов'язаними з забезпеченням конфіденційності, є:

- частка втрат інформації про основні ознаки внаслідок контролю непрямой ідентифікації;

- частка комірок, які містять заборонену для оприлюднення інформацію;
- частка втрат інформації про основні ознаки внаслідок контролю прямої ідентифікації;
- час, витрачений на захист конфіденційності табульованих даних;
- час, витрачений на захист конфіденційності мікроданих.

Розповсюдження даних/ інформації передбачає:

- розповсюдження стандартних агрегованих результатів у вигляді тексту, діаграм, таблиць, на паперових та електронних носіях та через веб-сайт;
- розповсюдження мікроданих у формі, що забезпечує конфіденційність;
- розповсюдження метаданих та звітів з якості.

Основними процесними змінними, пов'язаними з розповсюдженням інформації, є:

- змінні своєчасності та пунктуальності випусків даних;
- кількість засобів, використаних для розповсюдження та документування агрегованих даних;
- кількість засобів, використаних для доступу до інформації та документування мікроданих.

Зауважимо, що успішне виконання роботи на операційному процесі оцінюється ступенем залежності якості кінцевих результатів від якості кожного підпроцесу загального операційного процесу.

Процеси перегляду методології та організації мають дуже важливе значення для удосконалення процесів виробництва статистичної інформації та покращення її якості. Вони передбачають оцінювання відповідності кінцевих результатів очікуваним на стадії планування. Тобто вони вказують, чи проводились на етапі планування роботи з визначення якості кінцевих результатів, що підлягають публікації, і наскільки реальна якість кінцевих результатів відповідає вимогам до якості статистичної інформації. Аналіз у процесі перегляду дозволяє визначити відповідність вироблених кінцевих статистичних продуктів концепціям, визначенням та критеріям якості, прийнятим на стадії планування. Цей аналіз переважно базується на обчисленні індикаторів якості не тільки кінцевих статистичних продуктів, а й процесів їх виробництва.

ВИСНОВКИ

Запропонована в роботі класифікація може застосовуватися при оцінюванні якості процесів виробництва статистичної інформації в українських статистичних обстеженнях. Особливістю цієї класифікації є те, що вона враховує процеси циклу виробництва статистичної інформації, а тому її використання може значно полегшити впровадження системи оцінювання якості процесів виробництва статистичної інформації.

Очевидно, що розглянуті нами процесні змінні є лише незначною частиною процесних змінних, які можуть бути обчислені у процесі проведення статистичних обстежень. При виконанні інших статистичних процесів, таких як процеси, що використовують адміністративні дані, процеси, що використовують дані з багатьох джерел, обчислення економічних індексів, статистичні компіювання, статистичне моделювання та прогнозування, лише частина процесних змінних буде аналогічною до процесних змінних, обчислюваних у статистичних обстеженнях. Таким чином, перспективою виконаного дослідження є розроблення класифікації процесних змінних для цих статистичних процесів.

Література:

1. Biemer P. Continuous quality improvement for survey operations: some general principles and applications / P. Biemer, R. Caspar // Journal of Official Statistics. — 1994. — № 10. — P. 307—326.
2. Morganstein D. R. Continuous quality improvement in statistical organization / D. R. Morganstein, D. A. Marker // Survey measurement and process quality. — New York: Wiley, 1997. — P. 475—500.
3. Гончар О. В. Процеси виробництва та критерії якості статистичної інформації / О. В. Гончар // Економіст. — 2011. — № 12. — С. 33—37.
4. Quality improvement of the survey processes / I. Nikolaidis. — Piraeus: Eurostat, 2009. — 142 p. — (Methodological documents)
5. Васечко О. О. Новий аспект виміру якості статистичного реєстру підприємств // Статистика України. — 2004. — № 2. — С. 4—7.

Стаття надійшла до редакції 24.05.2012 р.