**Пояснення для користувачів до файлу мікроданих "Анонімні мікродані щодо заробітної плати працівників за статтю, віком, освітою та професійними групами"**

**за 2016 рік**

За результатами державного статистичного спостереження "Рівень заробітної плати працівників за статтю, віком, освітою та професійними групами" підготовлено анонімні мікродані за основними показниками щодо заробітної плати та соціально-економічних ознак працівників за 2016 рік.

Детальну інформацію про організацію, методологію проведення спостереження, поширення його результатів на генеральну сукупність тощо наведено в Методологічних положеннях з організації державного статистичного спостереження "Рівень заробітної плати працівників за статтю, віком, освітою та професійними групами", які розміщено на вебсайті Держстату (доступно за посиланням: http://ukrstat.gov.ua/norm\_doc/2020/269/269.pdf).

Спостереження охоплює юридичні особи та відокремлені підрозділи юридичних осіб із кількістю найманих працівників 10 і більше осіб. Дані за 2016 рік не враховують тимчасово окуповану територію Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях.

Файл представлено у форматі "xlsx" та "sav" (програмне забезпечення SPSS). До нього включено змінні по працівниках щодо:

* статі;
* рівня освіти;
* професійної групи (на рівні 3-х знаків Класифікатора професій);
* вікової групи;
* групи за стажем роботи;
* регіону, в якому зайнятий працівник;
* виду економічної діяльності, в якому зайнятий працівник (на рівні секцій та розділів Класифікації видів економічної діяльності ДК 009:2010);
* розміру підприємства, на якому зайнятий працівник, за кількістю найманих працівників;
* ваг, які використовуються для поширення результатів обстеження на генеральну сукупність;
* кількості відпрацьованого робочого часу в жовтні, годин;
* кількості відпрацьованого робочого часу за рік, годин;
* нарахованої заробітної плати в жовтні, грн;
* середньомісячної заробітної плати в році, грн.

Файл є набором анонімних знеособлених мікроданих, до яких були застосовані методи захисту знеособлених мікроданих (глобального перекодування, агрегування та маскування) з метою мінімізації ризику опосередкованого встановлення конкретної статистичної одиниці.

Підготовка файлу проводилась із застосуванням до окремих змінних методів маскування, зокрема некоригувальних методів захисту знеособлених мікроданих, які не змінюють початкових даних. Так у файлі зменшено рівень деталізації шляхом глобального перекодування й агрегування в інтервальні ряди по змінних "Вікова\_група", "Група\_стаж", "Розмір\_підприємства". Наприклад, первинні дані містили інформацію щодо віку кожного працівника. Для зменшення ризику розкриття конфіденційної інформації у файлі мікроданих сформовано сім вікових груп інтервального розподілу: до 25 років, 25-34 років, 35-44 років, 45-54 років, 55-59 років, 60-64 років, 65 і більше років.

До змінних "Зарплата\_жовтень" та "Зарплата\_рік" були застосовані коригувальні методи маскування, а саме локального закриття (процедура верхнього перекодування top-coding − заміна найбільших значень на середнє значення цієї змінної по зазначеній групі одиниць). Наприклад, по змінній "Зарплата\_жовтень" було визначено вісім працівників, у яких значення заробітної плати значно перевищувало середнє значення по цій змінній по всій сукупності працівників (аутлаєри). Щоб захистити цих працівників від можливого розкриття конфіденційної інформації про них, такі значення-аутлаєри були замінені на середнє значення по цих восьми працівниках. При такому підході слід ураховувати, що використання процедури "top-coding" може вплинути на основні показники диференціації по заробітній платі.

Звертаємо увагу, що при аналізі змінних "Зарплата\_жовтень" та "Зарплата\_рік" слід брати до уваги значення змінних "Відпрац\_час\_жовтень" та "Відпрац\_час\_рік".

Також до файлу була застосована процедура маскування шляхом локального закриття окремих значень для забезпечення граничного рівня k-анонімності з параметром 2. Мета цієї процедури − досягнення встановленого граничного рівня k-анонімності при одночасній мінімізації кількості випадків необхідного приховування значення змінної. k-анонімність – це показник ризику, який ґрунтується на тому принципі, що кількість осіб, які мають однаковий ключ (набір змінних, які в поєднанні з іншими можуть призвести до ідентифікації осіб), становить не менш ніж k.

У результаті зазначеної процедури по змінній "Код\_професії" було приховано 50 записів.

Після застосування процедур анонімізації значення деяких змінних були змінені. У зв’язку з цим відповідні похідні змінні, які розраховуються на основі агрегування змінених первинних змінних, також приймають інші значення порівняно з показниками, наведеними в офіційних публікаціях. До таких показників належать: середня заробітна плата у жовтні, середньомісячна заробітна плата в році, а також показники диференціації заробітної плати.

Застосовані методи забезпечують високий рівень захисту від можливого розкриття конфіденційних даних із одночасним збереженням оцінок основних показників спостереження на рівні, прийнятному для аналізу. Відхилення середніх показників, розрахованих за даними файлу анонімних мікроданих, від відповідних оприлюднених рівнів не перевищує 3%.