

ДЕРЖАВНИЙ КОМІТЕТ СТАТИСТИКИ УКРАЇНИ

ЗАТВЕРДЖЕНО  
Наказ Держкомстату  
20.12.2011 р., № 356

**МЕТОДИКА**  
РОЗРАХУНКУ ХАРАКТЕРИСТИК НАДІЙНОСТІ ОЦІНЮВАННЯ  
ПОКАЗНИКІВ ЗА РЕЗУЛЬТАТАМИ ПОТОЧНИХ ОБСТЕЖЕНЬ  
ПІДПРИЄМСТВ ІЗ ПИТАНЬ СТАТИСТИКИ ПРАЦІ

Київ-2011

## АНОТАЦІЯ

Методика визначає основний зміст характеристик надійності оцінювання показників за результатами поточних обстежень підприємств із питань статистики праці, порядок їх розрахунку з урахуванням дизайну вибірки та принципи використання. Розглянуто інформаційне забезпечення розрахунків характеристик надійності та наведені конкретні приклади.

Ця методика орієнтована на здійснення розрахунків в автоматизованому режимі на основі стандартних процедур, реалізованих у пакеті статистичних програм *SPSS* і спеціалізованому модулі *SPSS Complex Sample*.

Методику призначено для використання працівниками органів державної статистики, насамперед на державному рівні. Вона є корисною також для працівників територіальних управлінь статистики, фахівців наукових організацій та навчальних закладів, що проводять дослідження в галузі методології вибірових обстежень.

Методику розроблено фахівцями Інституту демографії та соціальних досліджень НАН України В.Г.Саріогло, Г.І. Терещенко за безпосередньої участі директора департаменту статистики праці Держкомстату І.В. Сенік.

## ЗМІСТ

	с.
Передмова	4
Визначення основних термінів. Перелік скорочень і символів	5
1. Загальні положення	7
2. Визначення характеристик надійності обстеження підприємств із питань статистики праці	10
3. Інформаційне забезпечення	13
3.1 Вхідна інформація	13
3.2 Вихідна інформація	13
Перелік джерел	17
Додаток	18

## ПЕРЕДМОВА

Інформація щодо якості і, зокрема, надійності оцінювання показників за результатами вибіркового обстеження має вирішальне значення для забезпечення оптимальності дизайну обстеження, статистичної ефективності основних етапів обробки даних та, що найбільш важливо, для правильної інтерпретації їх результатів. Оцінки показників, надійність яких є недостатньою, не можуть бути використані для аналізу процесів і явищ і прийняття управлінських рішень.

Можливість визначення надійності оцінювання показників для генеральної сукупності за результатами окремого вибіркового обстеження є найважливішим фактором, який зумовлює ефективність вибіркового методу спостережень при зборі офіційної статистичної інформації. Як відомо, зазначена можливість забезпечується ймовірнісним характером, тобто репрезентативністю вибірки, та належною реалізацією процедур обробки даних і оцінювання показників.

У методиці розглянуто питання визначення характеристик надійності оцінювання показників за результатами поточних обстежень підприємств із питань статистики праці.

## **ВИЗНАЧЕННЯ ОСНОВНИХ ТЕРМІНІВ. ПЕРЕЛІК СКОРОЧЕНЬ І СИМВОЛІВ**

*Гранична похибка* – гранично припустима розбіжність між оцінкою показника за результатами вибіркового обстеження та дійсним значенням показника (невідомим) для генеральної сукупності, визначена на основі стандартної похибки вибірки та встановленого рівня довірчої ймовірності.

*Дизайн-ефект ( $deff(\hat{\theta})$ )* – міра відносної статистичної ефективності певного дизайну вибірки порівняно з простою випадковою вибіркою (відношення дисперсії вибірових оцінок показника для реального дизайну вибірки, до дисперсії оцінки цього показника, отриманої за допомогою простого випадкового відбору).

*Дисперсія вибірових оцінок* – середній квадрат відхилень оцінок показника за всіма можливими вибірками цього дизайну від середнього значення оцінки за всіма вибірками.

*Довірчі межі (межі довірчого інтервалу)* – крайні значення оцінки показника, між якими міститься дійсне значення показника.

*Коефіцієнт варіації (відносна стандартна похибка)* – відношення середньоквадратичного відхилення вибірових оцінок показника за всіма можливими вибірками певного дизайну до середнього значення оцінки показника за всіма вибірками. (було 2р.)

*Надійність* – ступінь близькості розрахованих оцінок показників до їх дійсних (точних) значень. (було методика відбору)

*Оцінка показника* – значення показника для генеральної сукупності, отримане за результатами вибіркового спостереження.

*Стандартна похибка* – середньоквадратичне відхилення вибірових оцінок показника за всіма можливими вибірками цього дизайну від дійсного значення показника для генеральної сукупності.

*Статистична вага* – число елементів сукупності з однаковими варіантами у вигляді абсолютних або відносних величин, які визначають значимість того чи іншого значення ознаки (варіанти) в певній статистичній сукупності і які використовуються для обчислення узагальнюючих показників.

### **Перелік скорочень**

ЄДРПОУ- Єдиний державний реєстр підприємств та організацій України  
КВЕД - Класифікація видів економічної діяльності  
ОПСП - Обстеження підприємств із питань статистики праці

**Символи:**

- $n$  – обсяг вибірки;
- $\hat{\theta}$  – оцінка показника  $\theta$  за результатами вибіркового обстеження;
- $\theta_L, \theta_R$  – нижня та верхня межі довірчого інтервалу;
- $CV(\hat{\theta})$  – коефіцієнт варіації вибірових оцінок показника  $\theta$ ;
- $d_{eff}(\hat{\theta})$  – дизайн-ефект при оцінюванні показника  $\theta$ ;
- $ME(\hat{\theta})$  – гранична похибка оцінювання показника  $\theta$ ;
- $SE(\hat{\theta})$  – стандартна похибка оцінювання показника  $\theta$ ;
- $V(\hat{\theta})$  – дисперсія вибірових оцінок показника  $\theta$ .

## 1. ЗАГАЛЬНІ ПОЛОЖЕННЯ

Поточні обстеження підприємств із питань статистики праці організовуються на основі комбінованого підходу. Тобто звітні одиниці з кількістю працівників 50 і більше осіб обстежуються суцільно, а з кількістю від 10 до 49 осіб включно – з використанням вибіркового методу [2].

Звітні одиниці, які обстежуються на суцільній основі, надають відомості протягом усього терміну проведення ОПСП. Сукупність звітних одиниць, що обстежується на основі вибіркового методу, формується з використанням стратифікованого випадкового відбору. При формуванні вибірки із її основи з використанням процедури систематичного відбору відбирається встановлена кількість звітних одиниць, які представляють усі регіони України та види економічної діяльності й обстежуються протягом двох років. Обсяг вибірки для обстеження становить близько 10% генеральної сукупності. За розробленою схемою ротації щороку оновлюється 50% вибіркової сукупності звітних одиниць [2].

За часом реєстрації даних ОПСП складається зі щомісячних, щоквартальних і річних спостережень. Щомісячні спостереження відрізняються від щоквартальних і річних вужчим колом питань, що вивчаються, та коротшими термінами обробки й оприлюднення інформації. Окремі блоки питань вивчаються один або два рази на рік.

За результатами ОПСП розраховуються оцінки показників ринку праці, які характеризують генеральну сукупність. Оцінювання показників передбачає розрахунок системи статистичних ваг обстежених звітних одиниць і власне розрахунок показників на основі стандартних оціночних функцій, мікроданих обстеження та зазначеної системи статистичних ваг.

Із статистичної точки зору, оцінка певного показника за результатами вибіркового обстеження є надійною, якщо вибіркові оцінки, побудовані на базі всіх можливих вибірок однакового дизайну й обсягу, сконцентровані біля дійсного значення показника [3]. Висока концентрація, тобто надійність, бажана тому, що в окремому процесі побудови вибірки та розрахунку оцінки, який майже завжди застосовується на практиці, існує лише деяка ймовірність наблизитися до дійсного значення (див. рис. 1).

Слід зазначити, що в теоретичній статистиці розвинуті підходи, які забезпечують можливість оцінки надійності без необхідності повторення обстежень для широкого спектру дизайнів вибірок. Ці підходи вимагають насамперед імовірнісного характеру вибірки: кожна одиниця генеральної сукупності повинна мати відому (не нульову) імовірність потрапити до вибірки. Для найпростіших дизайнів вибірки характеристики надійності оцінок показників за класичними оціночними функціями можуть бути визначені на основі аналітичних залежностей [3].

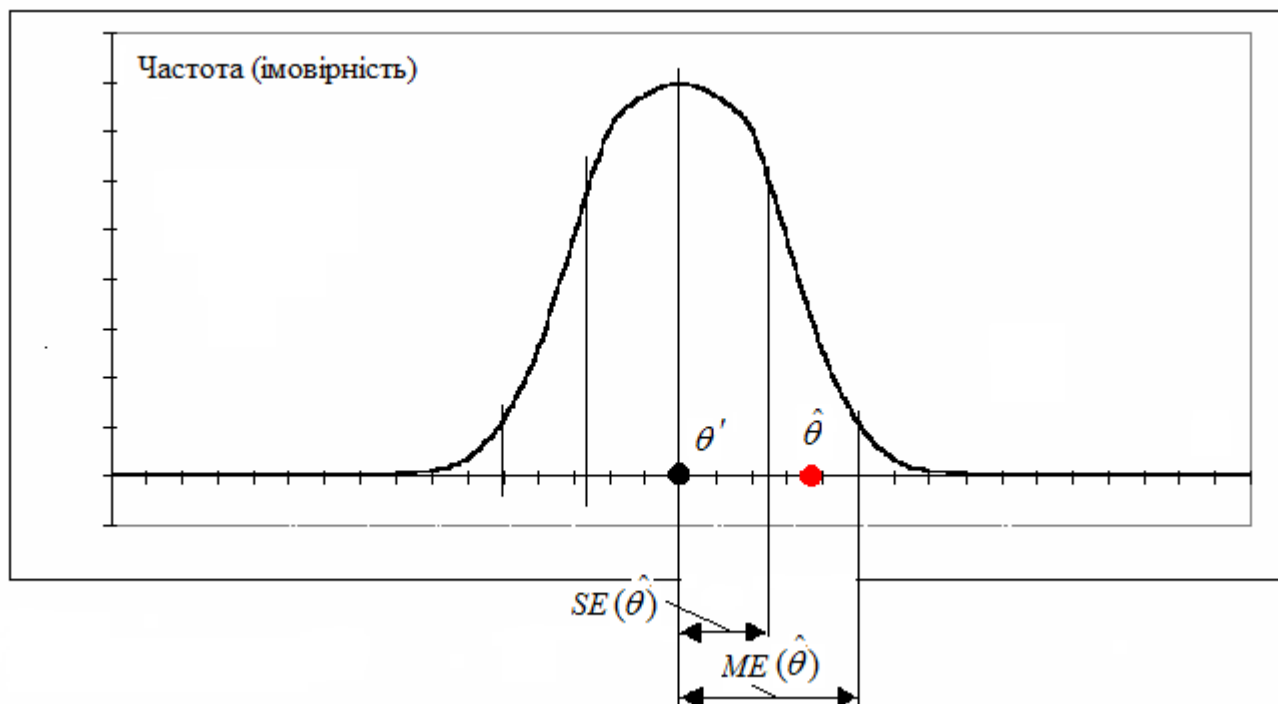


Рис. 1. Розподіл вибірових оцінок показника  $\theta$  за всіма можливими вибірками певного дизайну ( $\theta'$  - дійсне значення показника;  $\hat{\theta}$  - оцінка показника за певною вибіркою)

Близькість оцінки показника до реального значення показника, а саме її надійність, має два аспекти:

- оцінки мають малу дисперсію (дисперсію вибірових оцінок);
- середнє значення оцінок показника близьке до дійсного значення (оцінки показників є незміщеними або зміщення є незначним).

Дисперсія вибірових оцінок визначається за формулою [3, 4]:

$$V(\hat{\theta}) = \frac{1}{H} \sum_{i=1}^H (\hat{\theta}_i - \bar{\hat{\theta}})^2, \quad (1)$$

де  $\hat{\theta}_i$  – оцінка показника по вибірці  $i$ ,  $i = 1, 2, \dots, H$ ;  $\bar{\hat{\theta}}$  – середнє значення вибірових оцінок показника,  $\bar{\hat{\theta}} = \frac{1}{H} \sum_{i=1}^H \hat{\theta}_i$ ;  $H$  – кількість вибірок.

Похибка вибірки розраховується як стандартна похибка  $SE(\hat{\theta})$  на підставі величини оцінки дисперсії  $V(\hat{\theta})$  значень показника за формулою:

$$SE(\hat{\theta}) = \sqrt{V(\hat{\theta})}. \quad (2)$$



Для вибірок зі складним дизайном величина  $V(\hat{\theta})$  може бути визначена на основі класичної формули для дисперсії ознаки по одиницях вибірки при простому випадковому відборі  $\sigma_s^2$  [3]:

$$V(\hat{\theta}) = \text{deff}(\hat{\theta}) \cdot \frac{\sigma_s^2}{n}, \quad (3)$$

де  $\sigma_s^2$  – дисперсія ознаки, за якою визначається показник, по одиницях вибірки для простого випадкового відбору.

Коефіцієнт варіації  $CV(\hat{\theta})$  розраховується за формулою:

$$CV(\hat{\theta}) = \frac{SE(\hat{\theta})}{\hat{\theta}} \cdot 100 \%. \quad (4)$$

Таким чином, як видно із формули (4), коефіцієнт варіації є відносною стандартною похибкою. Відносні стандартні похибки в аналізі надійності результатів обстежень доцільно використовувати, оскільки вони не залежать від рівня ознаки.

Величина коефіцієнта варіації часто використовується як показник придатності даних для аналізу [4]. Так, якщо  $CV \leq 5\%$ , то оцінка вважається точною, якщо  $5\% < CV \leq 10\%$  – оцінка є придатною для кількісного аналізу, але її точність недостатньо висока, якщо  $10\% \leq CV \leq 25\%$  – оцінка придатна лише для якісного аналізу і її слід використовувати обережно (іноді публікують дані, для яких  $CV$  досягає 30% і навіть 40%).

Гранична похибка вибірки  $ME(\hat{\theta})$  розраховується на основі стандартної похибки вибірки за формулою [4]:

$$ME(\hat{\theta}) = t \cdot SE(\hat{\theta}) \quad (5)$$

де  $t$ -довірче число (квантиль нормального розподілу або квантиль розподілу ймовірностей) визначає співвідношення граничної та стандартної похибки при цій імовірності  $p$  ( $p$  - імовірність того, що похибка вибірки для оцінки показника не перевищить величину  $ME(\hat{\theta})$ ).

Гранична похибка вибірки використовується для побудови довірчих меж інтервальних оцінок (меж довірчих інтервалів). Наприклад, для побудови нижньої  $\hat{\theta}_L$  та верхньої  $\hat{\theta}_R$  довірчих меж інтервальної оцінки сумарного значення  $\hat{\theta}$  використовуються формули:

$$\hat{\theta}_L = \hat{\theta} - ME(\hat{\theta}); \quad \hat{\theta}_R = \hat{\theta} + ME(\hat{\theta}). \quad (6)$$

Зміщення оцінки показника  $B(\hat{\theta}; \theta)$  визначається як різниця між середнім значенням вибірових оцінок показника  $\bar{\theta}$  та дійсним значенням показника  $\theta$  за формулою:

$$B(\hat{\theta}; \theta) = \bar{\hat{\theta}} - \theta. \quad (7)$$

За наявності зміщення для визначення статистичної надійності оцінки показника використовується середньоквадратична похибка  $MSE(\hat{\theta}; \theta)$ , що визначається за формулою:

$$MSE(\hat{\theta}; \theta) = \text{var}(\hat{\theta}) + B^2(\hat{\theta}; \theta). \quad (8)$$

Таким чином, для визначення надійності оцінок показників у загальному випадку необхідно розраховувати дисперсію вибірових оцінок (на основі дисперсії визначаються стандартна похибка, гранична похибка, коефіцієнт варіації) та зміщення. Доцільно зауважити, що і дисперсія вибірових оцінок, і зміщення визначаються дизайном вибірки, оціночними функціями, які використовуються для оцінювання показників, та деякими особливостями проведення обстеження й обробки даних. Дизайн вибірки ОПСП, порядок збору, обробки даних і оцінювання показників розроблені так, що можливе зміщення оцінок показників за результатами цього обстеження є незначним і його можна не враховувати.

## 2. ВИЗНАЧЕННЯ ХАРАКТЕРИСТИК НАДІЙНОСТІ ОБСТЕЖЕННЯ ПІДПРИЄМСТВ ІЗ ПИТАНЬ СТАТИСТИКИ ПРАЦІ

В ОПСП надійність оцінок показників розраховується щомісячно та щоквартально за такими кластерами:

- Україна в цілому;
- Україна за видами економічної діяльності (секція та підсекція за КВЕД);
- регіони в цілому;
- регіони за видами економічної діяльності (секція та підсекція за КВЕД);
- адміністративно-територіальні одиниці (міста республіканського та обласного значення, райони) в цілому.

Основними показниками, для яких контролюється надійність за зазначеними кластерами, є "Середньооблікова кількість штатних працівників" та "Середня заробітна плата штатного працівника".

У відповідності з дизайном обстеження дисперсія вибірових оцінок показника  $V(\bar{\theta})$  по Україні, регіону або виду економічної діяльності для ймовірнісної вибірки визначається за формулою для стратифікованого відбору [3, 5]:

$$V(\hat{\theta}) = \sum_{s=1}^S W_s^2 \cdot V_s(\hat{\theta}_s), \quad (9)$$

де  $W_s$  – вага страти  $s$ :  $W_s = N_s / N$  ( $N_s$  – загальна кількість звітних одиниць у страті  $s$ ;  $N$  – загальна кількість звітних одиниць у сукупності, яка складається з  $S$  страт і по якій оцінюється показник);

$V_s(\hat{\theta}_s)$  – дисперсія вибірових оцінок показника  $\theta$  по страті  $s$ .

При розробці дизайну вибірки ОПСП реалізовано два варіанти стратифікації для оцінювання показників по кластерах. Перший варіант стратифікації забезпечує оцінювання показників для таких кластерів: Україна в цілому, Україна за видами економічної діяльності (секція та підсекція за КВЕД), регіони в цілому, регіони за видами економічної діяльності (секція та підсекція за КВЕД). Другий варіант стратифікації забезпечує оцінювання показників для кластера адміністративно-територіальні одиниці (міста республіканського та обласного значення, райони) в цілому.

За першим варіантом стратифікації оцінювання показників здійснюється на основі таких страт:

– страта, яка складається зі звітних одиниць, що обстежуються на суцільній основі (звітні одиниці з кількістю працівників 50 і більше та звітні одиниці секції В “Рибне господарство” та підсекції DF “Виробництво коксу, продуктів нафтоперероблення та ядерних матеріалів”);

– 559 страт, що складаються зі звітних одиниць із кількістю працівників 10 – 49 осіб і обстежуються на основі вибірового методу.

Стратифікація здійснюється на етапі побудови вибірки за ознаками: основний вид економічної діяльності за КВЕД - секції та підсекції; територіальна належність за місцем здійснення діяльності відповідно до адміністративно-територіального устрою України. Крім того, страти із загальною кількістю одиниць менше ніж 14 об’єднуються в окрему загальну страту (сукупність “малих” страт). У межах сукупності “малих” страт стратифікація здійснюється лише за ознакою основний вид економічної діяльності за КВЕД: секції та підсекції.

За другим варіантом стратифікації оцінювання показників здійснюється на основі таких страт:

– страта, яка складається зі звітних одиниць, що обстежуються на суцільній основі;

– 682 страти, що складаються зі звітних одиниць із кількістю працівників 10 – 49 осіб і обстежуються на основі вибірового методу.

Стратифікація здійснюється на етапі побудови вибірки на основі територіальної належності звітних одиниць до адміністративно-територіальних одиниць.

Таким чином, в ОПСП у межах окремих страт звітні одиниці обстежуються на суцільній основі або на основі вибірового методу, при цьому вибірова сукупність формується за процедурою простого випадкового відбору [2]. Для страт, які обстежуються на суцільній основі  $V_s(\hat{\theta}_s) = 0$ . Для страт, які обстежуються на основі вибірового методу,  $V_s(\hat{\theta}_s)$  визначається за такими формулами:

а) при оцінюванні середніх значень

$$V_s(\hat{\theta}_s) = (1 - f_s) \cdot \frac{\sigma_s^2}{n_s}, \quad (10)$$

де  $f_s$  - частка відбору в страті  $s$  ( $f_s = n_s / N_s$ );  $\sigma_s^2$  - дисперсія ознаки, по якій оцінюється середнє значення  $\hat{\theta}_s$ , по одиницях вибірки у страті  $s$  ( $\sigma_s^2 = \frac{\sum_{i=1}^{n_s} (\theta_i - \hat{\theta}_s)^2}{n_s - 1}$ );  $n_s$  - обсяг вибірки звітних одиниць по страті  $s$ .

На практиці, якщо  $f_s \leq 0,05$ , приймають  $f_s = 0$ .

б) при оцінюванні сумарних значень (зокрема показників «Кількість штатних працівників», «Фонд оплати праці штатних працівників»)

$$V_s(\hat{\theta}_s) = N_s^2 \cdot (1 - f_s) \cdot \frac{\sigma_s^2}{n_s}. \quad (11)$$

в) при оцінюванні показників за відношенням (зокрема, показника «Середня заробітна плата штатних працівників»)

$$V_s(\hat{r}_s) = \frac{1 - f_s}{\hat{\theta}_{2s}^2 n_s} (\sigma_s^2(\hat{\theta}_{1s}) + \hat{r}_s^2 \cdot \sigma_s^2(\hat{\theta}_{2s}) - 2\hat{r}_s \cdot \rho \cdot \sigma_s(\hat{\theta}_{1s}) \cdot \sigma_s(\hat{\theta}_{2s})), \quad (12)$$

де  $\hat{r}_s = \frac{\hat{\theta}_{1s}}{\hat{\theta}_{2s}}$ ;  $\rho$  - коефіцієнт кореляції розподілів  $\theta_1$  та  $\theta_2$  по одиницях сукупності.

Приклади розрахунків характеристик надійності показників ОПСП наведені в додатку А.

### 3. ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ

#### 3.1 Вхідна інформація

Вхідною інформацією для розрахунку характеристик надійності оцінювання показників за результатами ОПСП є файл первинних даних обстеження у форматі статистичного пакету SPSS, що містить ознаки по звітних одиницях, по яких будуються оцінки, та додаткові змінні, необхідні для розрахунків.

Додаткові змінні:

- статистична вага звітних одиниць для України ( $w\_q\_ua$ );
- статистична вага звітних одиниць для адміністративно-територіальних одиниць ( $w\_q\_rn$ );
- види економічної діяльності (секція та підсекція за КВЕД) ( $kved$ );
- страти для України ( $strt\_ua$ );
- обсяг генеральної сукупності за стратами для України ( $gen\_suk\_ua$ );
- код територій (регіонів) за КОАТУУ ( $rg$ );
- страти для адміністративно-територіальних одиниць ( $strt\_rn$ );
- обсяг генеральної сукупності за стратами для адміністративно-територіальних одиниць ( $gen\_suk\_rn$ );
- код територій (адміністративно-територіальних одиниць) за КОАТУУ ( $rn$ ).

Змінні  $strt\_ua$ ,  $strt\_rn$  для кожної звітної одиниці побудовані при формуванні основи вибірки. Статистичні ваги звітних одиниць  $w\_q\_ua$  та  $w\_q\_rn$  є кінцевими вагами звітних одиниць, тобто вони враховують рівень участі в обстеженні та різкі зміни характеристик звітних одиниць. Змінні  $gen\_suk\_ua$  та  $gen\_suk\_rn$  розраховуються як сума статистичних ваг по кожній звітній одиниці в межах страт.

### 3.2 Вихідна інформація

Розрахунок похибок оцінювання показників за результатами ОПСП здійснюється з використанням стандартного програмного пакета SPSS із модулем SPSS Complex Samples із використанням спеціально сформованих додаткових змінних (див. пп. 3.1). Зокрема, за допомогою цього модуля розраховуються оцінки показників, стандартні похибки та коефіцієнти варіації.

Далі за допомогою пакета EXCEL на основі стандартних похибок розраховуються значення граничних похибок для оцінок показників. Гранична похибка вибірки для оцінок показників за результатами ОПСП розраховується за 95% рівнем довірчої імовірності за формулою (5):

$$ME(\hat{\theta}) = 1,96 \cdot SE(\hat{\theta}) \quad (13)$$

Довірчі межі інтервальної оцінки визначаються за формулою (6).

Вхідною інформацією за результатами розрахунків похибок оцінювання є таблиця у форматі редактора виводу SPSS. У графі «ESTIMATE» таблиці виводиться оцінка показника, у графах «STDERROR» та «CV» – стандартна похибка та коефіцієнт варіації. У табл. 1 наведено приклад вихідної інформації

при оцінюванні показника «Середньооблікова кількість штатних працівників» за регіонами України

Таблиця 1

**Приклад вихідної таблиці у форматі редактора виводу SPSS**

<b>kobl_2</b>	<b>ESTIMATE</b>	<b>STDERROR</b>	<b>CV</b>
1	389895	1982,464	,005
5	313369	1849,430	,006
7	195772	1207,861	,006
12	924856	2852,144	,003
14	1145543	3033,757	,003
18	248334	1434,312	,006
21	195475	1047,063	,005
23	455037	1819,675	,004
26	220642	1511,906	,007
32	383591	2374,101	,006
35	203627	1529,226	,008
44	536087	2016,663	,004
46	542353	2496,555	,005
48	236933	1228,639	,005
51	517428	2750,294	,005
53	364231	1412,280	,004
56	211464	1319,960	,006
59	248616	1228,945	,005
61	179615	1178,246	,007
63	654308	2956,622	,005
65	195126	1165,358	,006
68	241092	1653,058	,007
71	265270	1651,984	,006
73	139696	1045,854	,007
74	230393	1796,754	,008
80	1250665	5124,163	,004
85	83681	576,837	,007

Приклад вихідної таблиці у форматі EXCEL, призначеної для аналізу результатів розрахунків похибок вибірки та їх публікації, наведено в табл. 2.

Таблиця 2

**Приклад вихідної таблиці у форматі EXCEL**

Регіони	Середньооблікова кількість штатних працівників,	Стандартна похибка вибірки ( <i>SE</i> ),	Гранична похибка вибірки	Коефіцієнт варіації ( <i>CV</i> ), у %
---------	---	---	--------------------------	--

	тис. осіб	тис.осіб	(ME), тис.осіб	
АР Крим	389,9	2,0	3,9	0,51
Вінницька	313,4	1,8	3,6	0,59
Волинська	195,8	1,2	2,4	0,62
Дніпропетровська	924,9	2,9	5,6	0,31
Донецька	1145,5	3,0	5,9	0,26
Житомирська	248,3	1,4	2,8	0,58
Закарпатська	195,5	1,0	2,1	0,54
Запорізька	455,0	1,8	3,6	0,40
Івано-Франківська	220,6	1,5	3,0	0,69
Київська	383,6	2,4	4,7	0,62
Кіровоградська	203,6	1,5	3,0	0,75
Луганська	536,1	2,0	4,0	0,38
Львівська	542,4	2,5	4,9	0,46
Миколаївська	236,9	1,2	2,4	0,52
Одеська	517,4	2,8	5,4	0,53
Полтавська	364,2	1,4	2,8	0,39
Рівненська	211,5	1,3	2,6	0,62
Сумська	248,6	1,2	2,4	0,49
Тернопільська	179,6	1,2	2,3	0,66
Харківська	654,3	3,0	5,8	0,45
Херсонська	195,1	1,2	2,3	0,60
Хмельницька	241,1	1,7	3,2	0,69
Черкаська	265,3	1,7	3,2	0,62
Чернівецька	139,7	1,0	2,1	0,75
Чернігівська	230,4	1,8	3,5	0,78
м. Київ	1250,7	5,1	10,0	0,41
м. Севастополь	83,7	0,6	1,1	0,69

Розрахунки характеристик надійності оцінювання показників за результатами ОПСП здійснюються щомісячно, щоквартально та щорічно.

Щомісячно розраховуються характеристики надійності оцінювання показників «Середньооблікова кількість штатних працівників» та «Середня заробітна плата штатних працівників» у таких розрізах:

- Україна в цілому;
- Україна за видами економічної діяльності (секція та підсекція за КВЕД);
- регіони в цілому.

Щоквартально розраховуються характеристики надійності оцінювання показників «Середньооблікова кількість штатних працівників» і «Середня заробітна плата штатних працівників» у таких розрізах:

- Україна в цілому;
- Україна за видами економічної діяльності (секція та підсекція за КВЕД);
- регіони в цілому;
- регіони за видами економічної діяльності (секція та підсекція за КВЕД);

- адміністративно-територіальні одиниці (міста республіканського та обласного значення, райони) в цілому.

Щорічно розраховуються характеристики надійності оцінювання показників «Середньооблікова кількість штатних працівників», «Середня заробітна плата штатних працівників», «Кількість працівників у віці 15-34 років», «Кількість працюючих пенсіонерів», «Кількість працівників, навчених новим професіям», «Кількість працівників, які підвищили кваліфікацію», «Кількість укладених та зареєстрованих колективних договорів», «Кількість працівників, охоплених колективними договорами» у таких розрізах:

- Україна в цілому;
- Україна за видами економічної діяльності (секція та підсекція за КВЕД);
- регіони в цілому.



## ПЕРЕЛІК ДЖЕРЕЛ

1. Вибіркове спостереження: Термінологічний словник. – К.: НТК статистичних досліджень, 2004.
2. Методика формування вибірки для проведення поточних обстежень підприємств з питань статистики праці / Затверджено наказом Держкомстату № 202 від 31.05.2010 р. – 34с.;
3. Kish L. Survey sampling. – Wiley Classics Library Edition Published 1995. – 643 р.
4. Методологічні положення з підготовки стандартних звітів щодо якості результатів державних вибірових обстежень населення (домогосподарств). – К: Держкомстат, 2008. – 92 с.
5. Кокрен У. Методы выборочного исследования. – М.: Статистика, 1976. – 440 с.
6. Методика оцінювання показників за результатами поточних обстежень підприємств із питань статистики праці ; затверджена Наказом Держкомстату від 16.03.2011 р. № 64 [Електронний ресурс]. – Режим доступу :  
[http://www.ukrstat.gov.ua/metod\\_polog/metod\\_doc/2011/64/metod.htm](http://www.ukrstat.gov.ua/metod_polog/metod_doc/2011/64/metod.htm)

## Додаток А

### Приклад розрахунку оцінок надійності показника «Середньооблікова кількість штатних працівників» для регіону АР Крим, вид діяльності «Діяльність готелів та ресторанів» (секція Н)

Цей кластер складається із двох страт:

- страта  $s_1$  містить звітні одиниці, що обстежуються на суцільній основі. У цій страті налічується 84 звітні одиниці, що взяли участь у обстеженні протягом II кварталу 2011 року;
- страта  $s_{2201}$  містить звітні одиниці з кількістю працівників 10-49 осіб і обстежується на основі вибіркового методу. У цій страті обстежуються у II кварталі 2011 року 24 звітні одиниці ( $n_{s_{2201}} = 24$ ). Ці 24 звітні одиниці репрезентують 183 звітні одиниці генеральної сукупності ( $N_{s_{2201}} = 183$ ).

1) Для страти  $s_1$  дисперсія оцінки показника «Середньооблікова кількість штатних працівників»  $V_{s_1}(\hat{\theta}_{s_1}) = 0$ .

2) Розрахунок для страти  $s_{2201}$  (дані для розрахунку наведені у табл. А.1):

– частка відбору в страті  $s_{2201}$  :

$$f_{s_{2201}} = n_{s_{2201}} / N_{s_{2201}} = 24 / 183 \approx 0,131;$$

– середнє значення показника по страті  $s_{2201}$  :

$$\hat{\theta}_{s_{2201}} = \frac{\sum_{i=1}^{n_{s_{2201}}} w_i \theta_i}{\sum_{i=1}^{n_{s_{2201}}} w_i} \approx 20,46;$$

– сума квадратів відхилення від середнього значення показника по страті  $s_{2201}$

$$S_{s_{2201}} = \sum_{i=1}^{n_{s_{2201}}} (\theta_i - \hat{\theta}_{s_{2201}})^2 = \sum_{i=1}^{24} (\theta_i - 20,46)^2 \approx 3209,96;$$

– оцінка дисперсії показника по одиницях сукупності розраховується за формулою по страті  $s_{2201}$  :

$$\sigma_{s_{2201}}^2 = \frac{S_{s_{2201}}}{n_{s_{2201}} - 1} = \frac{\sum_{i=1}^{n_{s_{2201}}} (\theta_i - \hat{\theta}_{s_{2201}})^2}{n_{s_{2201}} - 1} = \frac{3209,96}{24 - 1} \approx 139,56;$$

Таким чином, для страти  $s_{2201}$  дисперсія вибірових оцінок показника «Середньооблікова кількість штатних працівників»  $V(\hat{\theta})$  розраховується за формулою (10):

$$V_{s_{2201}}(\hat{\theta}_{s_{2201}}) = N_{s_{2201}}^2 \cdot (1 - f_{s_{2201}}) \cdot \frac{\sigma_{s_{2201}}^2}{n_{s_{2201}}} = 183^2 \cdot (1 - 0,13) \cdot \frac{139,56}{24} \approx 169227,79.$$

Таблиця А.1

Дані для розрахунку характеристик надійності показника «Середньооблікова кількість штатних працівників» по страті  $s_{2201}$

№ звітної одиниці	Значення показника «Середньооблікова кількість штатних працівників», $\theta_i$	Статистичні ваги, $w_i$	$(\theta_i - \hat{\theta}_{s_{2201}})^2$
1	34	7,63	183,38
2	37	7,63	273,63
3	8	7,63	155,21
4	6	7,63	209,04
5	22	7,63	2,38
6	12	7,63	71,54
7	20	7,63	0,21
8	12	7,63	71,54
9	37	7,63	273,63
10	9	7,63	131,29
11	29	7,63	72,96
12	2	7,63	340,71
13	6	7,63	209,04
14	17	7,63	11,96
15	45	7,63	602,29
16	40	7,63	381,88
17	20	7,63	0,21
18	29	7,63	72,96
19	13	7,63	55,63
20	23	7,63	6,46
21	18	7,63	6,04
22	15	7,63	29,79
23	23	7,63	6,46
24	14	7,63	41,71

3) Розрахунок для кластера в цілому:

Дисперсія показника «Середньооблікова кількість штатних працівників» по всьому кластеру розраховується за формулою:

$$V(\hat{\theta}) = \sum_{s=1}^S V_s(\hat{\theta}_s) = V_{s_1}(\hat{\theta}_{s_1}) + V_{s_{2201}}(\hat{\theta}_{s_{2201}}) = 0 + 169422,53 = 169227,79;$$

Стандартна похибка вибірки розраховується за формулою (2):

$$SE(\hat{\theta}) = \sqrt{V(\hat{\theta})} \approx 411;$$

Коефіцієнт варіації розраховується за формулою (4) (оцінка сумарного (абсолютного) значення показника «Середньооблікова кількість штатних працівників» по всьому кластеру  $\hat{\theta} = 12316$  осіб):

$$CV(\hat{\theta}) = \frac{SE(\hat{\theta})}{\hat{\theta}} \cdot 100\% = \frac{411}{12316} \approx 3,34\%.$$

Гранична похибка вибірки  $ME(\hat{\theta})$  розраховується на основі стандартної похибки вибірки за формулою (5) ( $t = 1,96$ ):

$$ME(\hat{\theta}) = 1,96 \cdot SE(\hat{\theta}) = 1,96 \cdot 411,37 = 806.$$

Нижня  $\hat{\theta}_L$  та верхня  $\hat{\theta}_R$  довірчі межі інтервальної оцінки показника «Середньооблікова кількість штатних працівників» розраховуються за формулами (6):

$$\hat{\theta}_L = \hat{\theta} - ME(\hat{\theta}) = 12316 - 806 = 11510;$$

$$\hat{\theta}_R = \hat{\theta} + ME(\hat{\theta}) = 12316 + 806 = 13122.$$

### **Приклад розрахунку оцінок надійності показника «Середньомісячна заробітна плата штатних працівників» для регіону АР Крим, вид діяльності «Діяльність готелів та ресторанів» (секція Н)**

1) Для страти  $s_1$  дисперсія оцінки показника «Середньомісячна заробітна плата штатних працівників»  $V_{s_1}(\hat{r}_{s_1}) = 0$ .

2) Розрахунок для страти  $s_{2201}$  (дані для розрахунку наведено в табл. А.2).

– частка відбору у страті  $s_{2201}$  :

$$f_{s_{2201}} = n_{s_{2201}} / N_{s_{2201}} = 24 / 183 \approx 0,131;$$

– середнє значення показника «Середньомісячний фонд оплати праці штатних працівників» по страті  $s_{2201}$  :

$$\hat{\theta}_{1s_{2201}} = \frac{\sum_{i=1}^{n_{s_{2201}}} w_i \theta_{1i}}{\sum_{i=1}^{n_{s_{2201}}} w_i} \approx 25998,61;$$

– середнє значення показника «Середньооблікова кількість штатних працівників» по страті  $s_{2201}$  :

$$\hat{\theta}_{2s_{2201}} = \frac{\sum_{i=1}^{n_{s_{2201}}} w_i \theta_{2i}}{\sum_{i=1}^{n_{s_{2201}}} w_i} \approx 20,46;$$

– значення показника «Середньомісячна заробітна плата штатних працівників» по страті  $s_{2201}$  :

$$\hat{r}_s = \frac{\hat{\theta}_{1s}}{\hat{\theta}_{2s}} \approx 1270,80;$$

**Дані для розрахунку характеристик надійності показника «Середньомісячна заробітна плата штатних працівників» по страті  $s_{2201}$**

№ звітної одиниці	Статистичні ваги, $w_i$	Значення показника «Середньо-облікова кількість штатних працівників», $\theta_{2_i}$	$(\theta_{2_i} - \hat{\theta}_{2_{s_{2201}}})^2$	Значення показника «Середньомісячний фонд оплати праці штатних працівників», $\theta_{1_i}$	$(\theta_{1_i} - \hat{\theta}_{1_{s_{2201}}})^2$
1	7,63	34	183,38	15516,67	109871083,63
2	7,63	37	273,63	62583,33	1338441676,50
3	7,63	8	155,21	8800	295792214,60
4	7,63	6	209,04	6283,33	388692298,34
5	7,63	22	2,38	37000	121030563,60
6	7,63	12	71,54	12950	170266244,68
7	7,63	20	0,21	19950	36585693,01
8	7,63	12	71,54	20433,33	30972350,75
9	7,63	37	273,63	55466,67	868366511,05
10	7,63	9	131,29	8433,33	308539090,75
11	7,63	29	72,96	41050	226544315,85
12	7,63	2	340,71	1616,67	594479038,80
13	7,63	6	209,04	4516,67	461473781,97
14	7,63	17	11,96	15083,33	119143355,67
15	7,63	45	602,29	47483,33	461593157,67
16	7,63	40	381,88	70050	1940524887,51
17	7,63	20	0,21	14250	138029856,51
18	7,63	29	72,96	47050	443160985,85
19	7,63	13	55,63	19150	46903470,35
20	7,63	23	6,46	33550	57023478,35
21	7,63	18	6,04	19516,67	42015556,97
22	7,63	15	29,79	17733,33	68314867,25
23	7,63	23	6,46	30750	22575699,01
24	7,63	14	41,71	14750	126531245,68

– значення коефіцієнта кореляції  $\rho$  між показниками «Середньомісячний фонд оплати праці штатних працівників» і «Середньооблікова кількість штатних працівників» по страті  $s_{2201}$  дорівнює 0,874349.

– сума квадратів відхилення від середнього значення показника «Середньомісячний фонд оплати праці штатних працівників» по страті  $s_{2201}$

$$S_{s_{2201}} = \sum_{i=1}^{n_{s_{2201}}} (\theta_{1_i} - \hat{\theta}_{1_{s_{2201}}})^2 = \sum_{i=1}^{24} (\theta_{1_i} - 25998,61)^2 \approx 8416871424,35;$$

– сума квадратів відхилення від середнього значення показника «Середньооблікова кількість штатних працівників» по страті  $s_{2201}$

$$S_{s_{2201}} = \sum_{i=1}^{n_{s_{2201}}} (\theta_{2i} - \hat{\theta}_{2s_{2201}})^2 = \sum_{i=1}^{24} (\theta_{2i} - 20,46)^2 \approx 3209,96;$$

– оцінка дисперсії показника «Середньомісячний фонд оплати праці штатних працівників» розраховується за формулою по страті  $s_{2201}$ :

$$\sigma_{s_{2201}}^2 (\hat{\theta}_{1s}) = \frac{\sum_{i=1}^{n_{s_{2201}}} (\theta_{1i} - \hat{\theta}_{1s_{2201}})^2}{n_{s_{2201}} - 1} = \frac{8416871424,35}{24 - 1} \approx 365950931,49;$$

– оцінка дисперсії показника «Середньооблікова кількість штатних працівників» розраховується за формулою по страті  $s_{2201}$ :

$$\sigma_{s_{2201}}^2 (\hat{\theta}_{2s}) = \frac{\sum_{i=1}^{n_{s_{2201}}} (\theta_{2i} - \hat{\theta}_{2s_{2201}})^2}{n_{s_{2201}} - 1} = \frac{3209,96}{24 - 1} \approx 139,56.$$

Таким чином, для страти  $s_{2201}$  дисперсія вибірових оцінок показника «Середньомісячна заробітна плата штатних працівників»  $V(\hat{r})$  розраховується за формулою (12):

$$V_{s_{2201}}(\hat{r}_{s_{2201}}) = \frac{1 - f_s}{\hat{\theta}_{2s_{2201}}^2 n_{s_{2201}}} (\sigma_{s_{2201}}^2 (\hat{\theta}_{1s_{2201}}) + \hat{r}_{s_{2201}}^2 \cdot \sigma_{s_{2201}}^2 (\hat{\theta}_{2s_{2201}}) - 2\hat{r}_{s_{2201}} \cdot \rho \cdot \sigma_{s_{2201}} (\hat{\theta}_{1s_{2201}}) \cdot \sigma_{s_{2201}} (\hat{\theta}_{2s_{2201}})) \approx 7709,94.$$

### 3) Розрахунок для кластера в цілому:

Дисперсія показника «Середньомісячна заробітна плата штатних працівників» по всьому кластеру дорівнює:

$$V(\hat{r}) \approx 3476707,52.$$

Стандартна похибка вибірки розраховується за формулою (2):

$$SE(\hat{r}) = \sqrt{V(\hat{r})} \approx 30,23.$$

Коефіцієнт варіації розраховується за формулою (4) (оцінка показника «Середньомісячна заробітна плата штатних працівників» по всьому кластеру  $\hat{r} = 1864,59$  грн.):

$$CV(\hat{r}) = \frac{SE(\hat{r})}{\hat{r}} \cdot 100\% = \frac{30,23}{1864,59} \approx 1,6\%.$$

Гранична похибка вибірки  $ME(\hat{r})$  розраховується на основі стандартної похибки вибірки за формулою (5) ( $t = 1,96$ ):

$$ME(\hat{r}) = 1,96 \cdot SE(\hat{r}) = 1,96 \cdot 30,23 = 59,24.$$

Нижня  $\hat{r}_L$  та верхня  $\hat{r}_R$  довірчі межі інтервальної оцінки показника «Середньомісячна заробітна плата штатних працівників» розраховуються за формулами (6):

$$\hat{r}_L = \hat{r} - ME(\hat{r}) = 1864,59 - 59,24 = 1805,35;$$

$$\hat{r}_R = \hat{r} + ME(\hat{r}) = 1864,59 + 59,24 = 1864,59 .$$