Пігіда Антон Олександрович. Назва дисертаційної роботи: "Оцінювання та врахування дизайн-ефекту у багатоступеневих стратифіковиних вибірках для загальнонаціональних соціальних досліджень в Україні"

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ІМЕНІ ТАРАСА ШЕВЧЕНКА

На правах рукопису

Пігіда Антон Олександрович

УДК 303.533

ОЦІНЮВАННЯ ТА ВРАХУВАННЯ ДИЗАЙН-ЕФЕКТУ У

БАГАТОСТУПЕНЕВИХ СТРАТИФІКОВАНИХ ВИБІРКАХ ДЛЯ

ЗАГАЛЬНОНАЦІОНАЛЬНИХ СОЦІАЛЬНИХ ДОСЛІДЖЕНЬ В УКРАЇНІ

Спеціальність 22.00.02 — методологія та методи соціологічних досліджень

Дисертація на здобуття наукового ступеня кандидата

соціологічних наук

Науковий керівник:

Сидоров Микола Володимир-Станіславович,

кандидат фізико-математичних наук, доцент

Київ — 2015

2

ЗМІСТ

ВСТУП 4

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ВИНИКНЕННЯ

ДИЗАЙН-ЕФЕКТУ ТА ЙОГО РІЗНОВИДИ.

12

1.1. Загальні положення формування вибіркових

сукупностей.

12

1.2. Дизайн-ефект при стратифікованому відборі. 32

1.3. Дизайн-ефект при кластерному відборі. 39

1.4. Дизайн-ефект, що виникає від нерівних шансів

респондентів потрапити до вибірки.

44

1.5. Зменшення похибок, які виникають в результаті

округлення вибірки.

48

Висновки до першого розділу 62

РОЗДІЛ 2. ПРАКТИЧНА РЕАЛІЗАЦІЯ ПРОЦЕСУ

ПОБУДОВИ ВИБІРКИ З УРАХУВАННЯМ ДИЗАЙНЕФЕКТУ.

65

2.1. Етап стратифікації вибіркової сукупності. 65

2.2. Етап кластерного відбору. 82

2.2.1. Усунення впливу нерівних шансів кластерів

потрапити до вибірки.

84

2.3. Етап відбору респондентів. 91

Висновки до другого розділу 95

РОЗДІЛ 3. МЕТОДИЧНІ РЕКОМЕНДАЦІЇ ЩОДО

ПОБУДОВИ ВИБІРОК З УРАХУВАННЯМ ДИЗАЙНЕФЕКТУ ДЛЯ СОЦІАЛЬНИХ ДОСЛІДЖЕНЬ В УКРАЇНІ.

97

3.1. Оптимальне розміщення при кластерному відборі. 97

3.2. Оптимальне розміщення при стратифікованому

відборі.

114

3

3.3. Алгоритм нерівноймовірнісного відбору кластерів. 129

Висновки до третього розділу 141

ВИСНОВКИ 144

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 148

4

ВСТУП

Актуальність теми. Майже будь-яке кількісне дослідження має

етап побудови вибірки. І від цього етапу залежить якість отриманих

результатів дослідження. Тому метою будь-якого дослідника є

побудова якомога досконалішої вибірки з точки зору досягнутої

точності та витрат ресурсів на реалізацію вибірки. Пошук можливостей

підвищити точність вибірки завжди буде актуальною темою. З

теоретичної точки зору базовим і найпростішим типом вибірки є проста

випадкова вибірка. У цій вибірці кожен елемент генеральної сукупності

має рівні шанси потрапити до вибіркової, а також ці шанси є

незалежними від інших факторів. Проте вона не завжди є

найпростішою у реалізації, оскільки потребує механізму, який

забезпечить рівноймовірнісний та незалежний відбір одиниць до

вибіркової сукупності. Найчастіше це вимагає переліку всіх одиниць

генеральної сукупності. На практиці соціальних досліджень такої

можливості зазвичай немає. Крім того, простота такої вибірки ще не

означає її оптимальність з точки зору ресурсів, які необхідно витратити

на реалізацію цієї вибірки. Швидше за все, проста випадкова вибірка

виявиться навіть найдорожчим варіантом з усіх можливих.

Тому у реальних дослідженнях дослідники вдаються переважно

до інших методів побудови вибіркової сукупності. Вони можуть досить

сильно відрізнятись за точністю та складністю реалізації, а отже

похибка вибірки для них обчислюють по-різному. Крім того, кожен

метод залежить від умов його застосування. Для того, щоб можна було

порівняти ефективність різних вибірок між собою, вводять міру

відносної (стосовно простої випадкової вибірки) статистичної

ефективності певного дизайну, яку називають дизайн-ефектом, тобто,

який вплив (ефект) мав поточний метод (дизайн) формування

вибіркової сукупності на репрезентативність вибірки, і порівнюють цей

5

вплив з простою випадковою вибіркою. Отже, дизайн-ефект виражає

міру відхилення наявної вибірки від стандартної простої випадкової

вибірки.

Концепція дизайн-ефекту розроблена для кількісної оцінки міри,

в якій помилка вибірки в дослідженні відрізняється від помилки

вибірки, яку можна очікувати при простій випадковій вибірці.

Спосіб відбору респондентів (наприклад, стратифікований,

кластерний, спрямований, квотний відбір) визначає те, як ці дані

аналізувати статистично. Більшість статистичних тестів розраховані на

те, що дані були отримані за допомогою простої випадкової вибірки.

Але коли дані насправді не були зібрані таким чином, статистичні

тести стають некоректними до невідомого рівня. Тому дизайн-ефект

розраховується як частка від ділення фактичного розміру вибірки на

ефективний розмір вибірки.

Ефективний розмір вибірки — це оцінка розміру вибірки

дослідження, проведеного за допомогою простої випадкової вибірки,

що вимагала б досягнути такої ж помилки вибірки, яка була обчислена

у дослідженні, що не передбачає простого випадкового відбору.

Дизайн-ефект — це вплив методу формування вибірки на

необхідний обсяг вибірки. Чим складніший метод формування, тим

сильніше порушується умова незалежності вибору і тим вищий дизайнефект.

Звісно метою побудови вибірки є отримання якомога меншої

похибки. А отже і дизайн-ефект вибірки має бути також якомога

нижчим, що залежить від безлічі не зовсім очевидних чинників.

Серед публікацій, автори яких розглядають проблеми якості

результатів вибіркових обстежень, насамперед слід відмітити роботи

M. Хансена і В. Гурвіца, Л. Кіша, У. Кокрена, К.-Е. Сарндала та Б.

Свенсона, Р. Джессена, Ф. Йейтса, Г. Колтона. Ці автори звернули

увагу на новітні аспекти розуміння проблем якості результатів

6

вибіркових обстежень та особливості впливу їх дизайну на окремі

характеристики якості.

Вибіркові методи досліджень розвивались і застосовувались на

практиці також відомими радянськими вченими В.Н. Криловим, А.Я.

Боярським, Б.В. Гнеденком, Н.К. Дружиніним, Ю.К.Бєляєвим та

іншими. Але їх дослідження були спрямовані насамперед на

забезпечення якості і надійності данних, а також окремих результатів

відбору та представництва в природничих і технічних процесах.

Фундаментальні дослідження з питань якості соціологічної інформації

були виконані В.Е.Шляпентохом, Б.З.Докторовим, Г.І. Саганенко. Вони

здійснили глибокий аналіз характеристик надійності та способів їх

вимірювання. Але результати цих досліджень стосувались насамперед

проблем первинного вимірювання ознак при проведенні соціологічних

обстежень.

В Україні вагомий внесок у вивчення проблеми якості даних

вибіркових обстежень зробили М.М.Чурілов, В.І. Паніотто, О.І.

Черняк, В.Г. Саріогло, А.М. Єріна, В.М.Пархоменко, О.М.Гладун та

інші. Результати їх робіт стосуються насамперед особливостей

організації обстежень та окремих аспектів обробки даних та

оцінювання показників.

Загалом математичний апарат у вирішенні цього питання вже

розроблений, але він не передбачає конкретних дій щодо побудови

вибірок у реальних умовах. Постає ситуація, що практика

соціологічних досліджень не застосовує всіх можливих досягнень

теорії. Отже, розробка методичних засобів щодо побудови

оптимальних вибіркових сукупностей залишається актуальною. Тому

необхідно, по-перше, показати які фактори мають значний вплив на

дизайн-ефект вибірки. А, по-друге, розробити комплекс методичних

рекомендації щодо побудови оптимальних вибірок з точки зору

похибки та можливостей реалізації цих рекомендацій. Звичайно, що

7

неможливо розглянути всі типи вибірок для всіх випадків, тому дане

дисертаційне дослідження обмежене лише типовими вибірками для

соціальних кількісних опитувань в Україні.

Наукова проблема даної роботи полягає у подоланні

невідповідності між потребою методологічного обгрунтування

емпіричних досліджень на основі багатоступеневих стратифікованих

вибірок та відсутністю методичних рекомендацій для побудови

репрезентативних вибіркових сукупностей з урахуванням дизайнефекту для всеукраїнських соціальних досліджень.

Метою дисертаційного дослідження є розробка комплексу

методичних рекомендацій щодо побудови якісних вибірок для

соціальних досліджень з урахуванням дизайн-ефекту. У процесі

реалізації цієї мети було виділено кілька підпорядкованих їй завдань:

1. проаналізувати загальні принципи та положення формування

вибіркових сукупностей;

2. виявити специфічність методів оцінювання дизайн-ефекту, що

виникає при побудові різних видів вибіркових сукупностей;

3. окреслити підходи до зменшення дизайн-ефекту на різних

етапах побудови вибірки;

4. обгрунтувати можливість розробки методичних рекомендацій

щодо побудови вибірок з урахуванням дизайн ефекту;

5. запропонувати оптимальне розміщення при кластерному

відборі;

6. запропонувати оптимальне розміщення при стратифікованому

відборі;

7. конкретизувати алгоритм невірноймовірнісного відбору

кластерів.

Об’єктом дослідження є дизайн-ефект у багатоступеневих

стратифікованих вибірках.

8

Предметом дослідження є методи оцінювання та врахування

дизайн-ефекту у багатоступеневих стратифікованих вибірках для

всеукраїнських соціальних досліджень.

Методи дослідження. У дисертаційному дослідженні були

використані методи кластерного, факторного, дисперсійного аналізу.

Розробка та доведення методик, запропонованих у роботі, здійснені на

основі використання математичного аналізу та математичної

статистики.

Комп’ютерне опрацювання масивів емпіричних даних

здійснювалось за допомогою прикладних програм OCA, SPSS, Excel, R

та розробленого в рамках реалізації дисертаційного дослідження

власного програмного забезпечення.

Емпіричною базою дисертаційної роботи є масиви даних

національних екзит-полів компанії TNS в Україні 2012-2014 років,

масиви даних European social survey хвиль 2008-2012 років, масиви

даних результатів українських президентських та парламентських

виборів 2012-2014 років.

Наукова новизна одержаних результатів.

На основі вирішення дослідницьких завдань були отримані

наступні наукові результати, що мають наукову новизну і виносяться

на захист:

1. Вперше запропоновано методику оптимального розміщення

при побудові кластерної вибірки з використанням функції витрат для

загальнонаціональних соціальних досліджень в Україні. Доведено, що

застосування даної методики дозволяє досліднику отримати

мінімальний дизайн-ефект від кластеризації елементів сукупності в

межах заданих ресурсів дослідження при використанні заданої функції

витрат.

9

2. Вперше запропоновано методику оптимального з точки зору

дизайн-ефекту розбиття генеральної сукупності на страти при

застосуванні стратифікованої вибірки. Дана методика дозволяє

згрупувати елементи генеральної сукупності у страти з точки зору

мінімізації дизайн-ефекту такого розбиття.

3. Вперше запропоновано методику округлення квотних таблиць,

що дозволяє зберегти всі задані розподіли. При застосуванні звичайних

методик округлення є загроза втрати розподілів за заданими

характеристиками при переході від дробових до цілих чисел. Дана

методика дозволяє зберегти ці розподіли.

4. Дістали подальшого розвитку методики корегування

ймовірності елементів генеральної сукупності потрапити до вибірки, з

метою мінімізації дизайн-ефекту дослідження. Порушення принципу

рівності шансів елементів генеральної сукупності потрапити до

вибіркової сукупності призводить до збільшення дизайн-ефекту

дослідження. Тому виникає потреба корегування цих ймовірностей.

5. Отримала подальшого розвитку систематизація різновидів

дизайн-ефектів, які виникають на різних етапах побудови вибірки для

загальнонаціональних соціальних досліджень в Україні.

Проаналізовано їх комплексний вплив на загальну репрезентативність

дослідження.

6. Доведено, що пропорційна стратифікація вибіркової сукупності

не може призводити до збільшення дизайн-ефекту дослідження. Будьяке розбиття сукупності на страти не може призвести до збільшення

похибки вибірки.

Практичне значення одержаних результатів. Розроблений

комплекс методичних рекомендацій щодо побудови вибірок для

соціальних досліджень з урахуванням дизайн-ефекту має значну

практичну цінність для соціологів, які займаються емпіричними

соціологічними дослідженнями, зокрема електоральними

10

дослідженнями, вивченням громадської думки щодо процесів

соціальної інтеграції та диференціації, соціальної мобільності,

соціальної стратифікації, а також соціального конфлікту, автономії та

контролю. Результати роботи можуть бути використані в практиці

досліджень в соціальному управлінні для оцінки ефективності

можливих управлінських впливів, у діагностиці напрямків

трансформаційних процесів і оцінці функціонування соціальних систем

різного рівня.

Запропонований комплекс методичних рекомендацій може бути

використаний у навчальному процесі при підготовці та викладанні

таких дисциплін як організація та методи вибіркових досліджень,

кількісні методи в соціологічних дослідженнях, аналіз соціальних

мереж, загальна теорія соціології тощо.

Застосування запропонованих у роботі методичних рекомендацій

у практиці соціологічних досліджень дозволяє знизити похибку

вибірки, зменшити її обсяг, підвищити рівень репрезентативності,

знизити витрати на дослідження, оптимальніше розподілити наявні

ресурси.

Особистий внесок здобувача. Дисертаційна робота є

завершеною працею, в якій викладено авторські підходи до розуміння

проблеми. Викладені в дисертації наукові положення та висновки, що

виносяться на захист, здобуті автором особисто, а саме розробка

комплексу методичних рекомендацій щодо побудови якісних вибірок

для соціальних досліджень з урахуванням дизайн-ефекту.

Апробація результатів дисертації. Основні результати

дисертаційного дослідження доповідались та обговорювались на таких

конференціях, семінарах та «круглих столах»: VІ Міжнародна

конференція студентів та молодих науковців “Соціологія і сучасні

соціальні трансформації” 22 листопада 2013 року; X Всеукраїнська

науково-практична конференція "Проблеми розвитку соціологічної

11

теорії: держава – людина – соціальна відповідальність" 24 травня 2013;

науково-методологічні семінари факультету соціології Київського

національного університету імені Тараса Шевченка.

Публікації.

Основні результати дисертаційного дослідження знайшли

відображення у 5 наукових працях, 4 з яких опубліковано у наукових

виданнях, включених ДАК України до переліку фахових, а 1

опубліковано у іноземному фаховому виданні.

Структура та обсяг дисертації.

Структура та обсяг дисертації обумовлені специфікою предмета

дослідження, логікою розкриття проблеми, а також метою і завданнями

дисертаційного дослідження. Дисертація складається із вступу, трьох

розділів, висновків та списку використаних джерел. Повний обсяг

дисертації складає 156 сторінок. Список використаних джерел на 9

сторінках містить 106 найменувань, з яких 40 іноземними мовами.

ВИСНОВКИ

ДануроботупідназвоюОцінюваннятаврахуваннядизайнефектуубагатоступеневихстратифікованихвибіркахдля

загальнонаціональнихсоціальнихдослідженьвУкраїнібуло

присвяченоявищудизайнефектущопритаманнемайжебудьякому

практичномудослідженню

Удисертаційнійроботіавторомрозробленометодикита

методичнірекомендаціїспрямованіназменшеннядизайнефектуу

загальнонаціональнихсоціальнихдослідженьвУкраїні

Застосуваннязапропонованихуроботіметодичнихрекомендацій

упрактицісоціологічнихдослідженьдозволяєзнизитипохибку

вибіркизменшитиїїобсягпідвищитирівеньрепрезентативності

знизитивитратинадослідженняоптимальнішерозподілитинаявні

ресурси

Длякожногоетапупобудовивибіркивизначаєтьсяякісаме

факторивпливаютьнадизайнефектвибіркитаможливостіконтролю

тавпливудослідникомнаціфакториДлястратифікованоївибірки

цимифакторамиєспіввідношенняпропорційміжстратамитасереднє

значенняознакиукожнійзістратТобтодизайнефектзалежитьвід

тогоякимчиномрозбитисукупністьнастратиУкластернійвибірці

дизайнефектзалежитьвідтогонаскількиподібнікластериувибірціта

якаїхкількістьвикористовуєтьсяувідборіЧимбільшнесхожі

кластериміжсобоютимбільшимбудедизайнефектудослідженні

АлезбільшеннякількостікластерівнаетапівідборузменшуєдизайнефектТакожнадизайнефектвпливаєймовірністьреспондентів

потрапитидовибіркиУвипадкуякщопорушуєтьсярівністьшансів

всіхреспондентівпотрапитидовибіркицепризводитьдозбільшення



дизайнефектуТакимчиномнеобхіднозаможливістюзабезпечувати

рівністьцихшансів

Однієюзголовнихумовувибірковомудослідженнієрівні

шансиреспондентівпотрапитидовибіркиУвипадкуякщо

респондентимаютьрізнішансипотрапитидовибіркицепризводитьдо

збільшеннядизайнефектуоскількивибіркастаєзміщеноюубіктих

респондентівякімаливищийшанспотрапитидовибірки

Кластернийвідбірдаєточнішуоцінкуніжпростийвипадковий

відбірякщосередньоквадратичневідхиленняміжелементами

всерединіодиницьбільшеніжсередньоквадратичневідхилення

середніхзначеньелементівміжодиницямиІнакшекажучикластери

маютьбутисхожіміжсобоюалеприцьомузалишатись

неодноріднимивсерединіМірасхожостікластерівміжсобою

називаєтьсякоефіцієнтомміжкластерноїкореляціїІвіднеїзалежить

величинадизайнефектуякийвиникаєприкластерномувідборі

Похибкакластерноївибіркизалежитьвідзагальногообсягу

вибіркикількостікластерівувибірцітамірисхожостікластерівміж

собоюкоефіцієнтміжкластерноїкореляціїПричомузаумови

обмеженостіресурсівнадослідженнязагальнийобсягвибіркита

кількістькластерівувибірцізалежатьвідобсягуцихресурсівта

функціївитрат

Длятогощобстворитикластернувибіркузнайнижчою

можливоюпохибкоюдослідникмаєрозрахуватиякимчиномслід

оптимальнорозміститивибіркузакластерамизалишаючисьвмежах

наявнихресурсівДляцьогойомунеобхіднодізнатисякоефіцієнт

міжкластерноїкореляціїдосліджуваноїознакиабоознакиякувін

можевикористатизамістьнеїоскількидлярозрахункукоефіцієнту

міжкластерноїкореляціїнеобхіднознатисереднєзначенняознакиу

кожномукластерітавціломуподосліджуванійсукупностіта

пов’язатийогозфункцієювитратякамаєпоказуватиякпов’язані



загальнівитратинадослідженнязобсягомвибіркитакількістю

кластерівунійТакимчиномрезультатомвиконанняцьогозавданняє

методиказастосуванняякоїдозволяєоптимальнимчиномвикористати

наявніресурсинареалізаціюкластерноївибіркиПобудованаза

допомогоюданоїметодикивибіркабудематинайменшийзможливих

заданихумовдизайнефект

Авторвиділяєщеоднеджерелопохибокпідчаспобудови

вибіркипов’язанезокругленнямчиселРозрахункипідчаспобудови

вибіркизазвичайведутьсязадопомогоюдробовихчисела

результуючавибіркамаєскладатисязцілихчиселІнаетапі

округленнячиселможутьвиникатипохибкиізсувивибіркиОсобливо

цязагрозаактуальнадляквотнихвибірокоскількиприокругленні

квотнихтаблицьєзагрозавтратитизаданірозподіли

Стратифікаціягенеральноїсукупностідозволяєспростити

відбірреспондентівізнизитипохибкувибіркиЦевідбуваєтьсяза

рахуноктогощореспондентивмежаходнієїстративиявляються

більшсхожимиміжсобоюніжміжстратамиОтжезавдання

дослідниканаетапістратифікаціїполягаєвтомущобзнайтитаке

розбиттянастратиприякомудизайнефектбудемінімальнимВролі

елементівмножиниякінеобхіднорозбитинапідмножинистрати

можутьвиступатиобластіокругинаселеніпунктитериторіальні

дільниціпоштовівідділеннятощо

Длятогощоброзрахуватиоптимальнерозбиттясукупностіна

стратидослідникунеобхіднознатидлякожногоелементуйоговагу

часткуісереднєзначеннядосліджуваноїознакиабоознаквцьому

елементіБулозапропонованоітеративнийалгоритмпошуку

оптимальногогрупуванняелементіввстратиНакожномукроці

алгоритмунеобхіднооб’єднуватитакідвіпідмножиниякідадуть

мінімальнийдизайнефектвпідсумку



Оскількивирішеннязавданняпошукуоптимальногорозбиття

сукупностінастратизточкизорудизайнефектудляоднієї

досліджуваноїознакиєнедужекориснимзпрактичноїточкизоруто

булозапропонованопокращеннящодозволяєврахуватибудьяку

кількістьважливихдлядослідниказміннихЗастосуванняданої

методикидозволяєзнизитипохибкувибіркиабозменшитиїїобсягщо

підвищуєоптимальністьвибіркитазнижуєвитратинадослідження

Такожбулодоведенощобудьякерозбиттясукупностінастрати

неможепризвестидозбільшенняпохибкивибіркиТобтонезалежно

відобраноговаріантустратифікаціїдизайнефекттакогорозбиттяне

можеперевищуватиодиницю

Одинзпринципівпобудовивибірковихсукупностейполягаєу

томущовсіреспондентимаютьматирівнішансипотрапитидо

вибіркиУвипадкупорушенняцьогопринципудизайнефектвибірки

зростаєУходівиконанняцьогозавданняаналізуютьсяетапинаяких

можепорушуватисьданийпринципДослідникмаєабокоректно

врахуватидизайнефектякийвиникаєвнаслідокпевноготипувідбору

абозмінитиметодикувідборуякадозволитьнівелюватизсувиу

порушеннірівнихшансівреспондентівпотрапитидовибіркиУ

випадкукластерноговідборуданийзсувможливоподолатиіавтор

пропонуєметодикуякуслідзастосовувати