

Тема. Надійність інструменту виміру.

План.

1. Надійність виміру.
2. Стійкість виміру.
3. Обґрунтованість виміру.

1. Надійність виміру.

Надійність виміру та надійність відбору – два аспекти надійності соціологічної інформації.

Надійність інструменту і, отже, надійність даних, одержуваних за допомогою цього інструменту, передбачає досягнення при шкалюванні таких вимог як:

1) обґрунтованість. Обґрунтованість шкали у тому, що з її допомогою вимірюють цілком певне властивість чи ознака, не змішуючи його з іншими. Обґрунтованість передбачає упевненість у цьому, що вимірюється цілком задане властивість об'єкта, а чи не якесь інше, нею схоже. Інакше кажучи, обґрунтованість – це відповідність інструментарію вимірюваному властивості об'єкта.

На підвищення обґрунтованості використовують різні способи. Насамперед, це логічні міркування на основі досвіду та здорового глузду.

Наприклад, необхідно виміряти середню інтенсивність перегляду телепередач за допомогою масового опитування. Можна запитати: "Скільки приблизно годин на день Ви проводите біля телевізора?". Пропонується 5 інтервалів: не більше 1 години, від 1 до 2 годин, від 2 до 3 годин, від 3 до 4 годин, понад 4 години. Виникають сумніви щодо обґрунтованості такого методу. Більш логічним буде виділення будніх, суботніх та недільних днів, а також формулювання питань про те, в які дні та скільки в середньому годин проводять респонденти у ці дні біля телевізора. Таким чином, ми зможемо вирахувати середньопереглядний час.

Використовується тест з "еталонної групи" - сильніший спосіб. Сенс його полягає у зіставленні даних, отриманих шляхом виміру за шкалою, з достовірними відомостями про об'єкт виміру.

Так, наприклад, вимір правосвідомості та дотримання прав можна провести шляхом порівняння еталона негативного порядку (правопорушників) та еталона законослухняних громадян.

Перевірка обґрунтованості шкали, вимірювання може бути здійснена за допомогою незалежного критерію, як якого виступає фіксована величина, наприклад, самообстеження, самофіксація витрат часу. З таким показником може бути порівняна величина середнього перегляду часу, отримана під час опитування респондентів.

При відборі шкал та його значень іноді використовується метод суддівства. Як суддів виступають типові представники основної маси опитуваних. Склад суддів має бути репрезентативним за основними соціальними параметрами (стаття, вік, рівень освіти, статус тощо).

2) правильність виміру.

Правильність виміру, тобто. відповідність результату виміру будь-якої величини її істинного значення. Правильність виміру – це здатність використовованого інструменту належним чином диференціювати (розрізнити) досліджувану сукупність. В іншому випадку може виявитися, що "не працює" якась частина шкали або бракує градацій, тобто. розкиду значень.

Якщо відповіді респондентів фіксуються лише в одному пункті шкали, це свідчить про її непридатність. Таке може бути в тому випадку, коли запропоновані значення не мають відношення до визначення досліджуваної властивості об'єкта, або вони важко сприймаються, або відчувається вплив на респондентів загальноприйнятої думки. Така шкала нерелевантна.

Подібна ситуація зустрічається при використанні шкал семантичного диференціалу Ч. Осгуда, тому що при вимірі установок необхідно використовувати антоніми типу "чоловічий - жіночий", "гарячий - холодний", які, на перший погляд, навряд чи можуть мати до такої установки якась відношення .

Нерідко під час опитувань ми бачимо, що використовується лише частина шкали або один із її полюсів. Таке зустрічається, коли оцінюється явно позитивне явище і в шкалах не використовуються негативні оцінки. Те саме може статися у разі, якщо зміст значень змінної нерівноцінно або шкала представлена надмірно дробовою. Респонденту вибрати з великої кількості запропонованих варіантів завжди важко, він звертає увагу на основний, базовий, найбільш зрозумілий варіант.

Таким чином, недостатня правильність вимірювання може виявлятися:

1) без розкиду в значеннях за шкалою, тобто. всі респонденти відзначили один варіант відповіді, що говорить про те, що шкала побудована неправильно;

2) коли респондентами використовується лише частина шкали, наприклад, негативні значення шкали уникають респондентів, що говорить про те, що ця частина шкали не працює;

3) як нерівномірного використання окремих пунктів шкал. Шкальне значення, виражене у балах, має істотно відрізнитися від сусідніх за частотою чи за кількістю респондентів. Якщо ця відмінність присутня, то це значення перевіряється статистично.

3) стійкість;

Стійкість – повторюваність одержуваних результатів при повторних вимірах, відсутність будь-яких суттєвих відхилень у яких при зміні процедури використання цієї шкали.

Стійкість може бути повною, абсолютною та відносною.

Повна стійкість характеризується часткою відповідей, що збігаються в ході повторних вимірювань.

Відносна стійкість показує здатність шкали відображати мінливість об'єкта, що досліджується, його вимірюваних властивостей.

Апробується якась шкала на певній групі респондентів, а через 2-3 тижні та ж група респондентів оцінюється за тією самою анкетною. Після проведення повторного опитування щодо кожного респондента проводиться порівняння результатів. Дивляться на повторюваність результатів, таким чином

враховується стійкість. Оцінка стійкості – це відсоток збігу відповіді питання двох вимірів.

Ця проблема має і методичний, і теоретичний аспект.

Вона пов'язана з правильністю переведення теоретичних понять в операційні, а потім в емпіричні індикатори. Підвищення стійкості виміру досягається за рахунок покращення чутливості шкали шляхом підбору оптимального числа її градацій. При мінімальному числі градацій шкала є менш чутливою, але надійнішою, зменшуються можливості помилки. При великій кількості значень шкала буде більш чутливою, але менш надійною, з більшим ризиком помилок. Рекомендується використовувати стільки градацій у шкалі, щоб помилки були зведені до мінімуму.

Перш ніж братися до вивчення таких компонентів надійності, як стійкість і обґрунтованість. Необхідно переконатися в правильності вибраного інструменту вимірювання (шкали або системи шкал). Можливо, що наступні етапи виявляться зайвими, якщо на початку з'ясується повна нездатність даного інструменту на необхідному рівні диференціювати сукупність, що вивчається, або може виявитися, що систематично не використовується якась частина шкали або її окрема градація. Насамперед потрібно ліквідувати або зменшити такого роду недоліки шкали і тільки потім використовувати її в дослідженні,

Відсутність розкиду, відповідей за значеннями шкали. Попадання відповідей в один пункт свідчить про повну непридатність вимірювального інструменту – шкали. Така ситуація може виникнути або через «нормативний» тиск у бік загальноприйнятої думки; або через те, що градації (значення) шкали крім ставлення до визначення даної властивості об'єктів, що розглядаються (нерелевантні).

Наприклад, якщо всі опитувані респонденти згодні із твердженням «добре, коли робота чи завдання вимагають універсальних знань», немає жодної відповіді «не згоден», залишається лише зафіксувати цей факт, проте подібна шкала не допоможе диференціювати сукупність, що вивчається, стосовно респондентів» до роботи .

Часто прикладом нерелевантності є багато вихідних шкал методики семантичного диференціала Осгуда. Так, зокрема, при вивченні установок інженера в роботі вимірювання респондентів за шкалами «чоловічий – жіночий», «гарячий – холодний» та ін. давали оцінку тільки в середині шкали, в нейтральній точці. На думку респондентів, немає, мають відношення до досліджуваних установок.

Використання частини шкали. Досить часто - виявляється, що практично працює лише якась частина шкали, якийсь один з його полюсів із прилеглою більш менш широкою зоною. Так, якщо респондентам для оцінки пропонується шкала, що має позитивний і негативний полюс, зокрема від +3 до -3, то при оцінюванні якоїсь свідомо позитивної ситуації респонденти не використовують негативні оцінки, а диференціюють свою думку лише за допомогою позитивних. Для того, щоб обчислити значення відносної помилки вимірювання, дослідник повинен знати точно, якою ж метрикою користується респондент - усіма сімома

градаціями шкали або тільки чотирма позитивними. Так, помилка виміру в 1 бал мало про що говорить, якщо ми не знаємо, якою є дійсна варіація думок.

Вважатимемо, що якщо ймовірність попадання в кінець шкали перевищує 5% при досить малому рівні значущості ($\alpha = 0,05$ або $\alpha = 0,01$), то попадання відповідей є випадковими і відповідна частина шкали «не працює».

Для питань, що мають якісні градації відповідей, можна застосовувати подібну вимогу щодо кожного пункту шкали: кожен із них має набирати не менше 5% відповідей, інакше вважаємо цей пункт шкали непрацюючим.

Вимогу 5%-го рівня наповнення у двох розглянутих завданнях не слід розглядати як суворо обов'язкове; Залежно від завдань дослідження може бути висунуті більші чи менші значення цих рівнів.

Нерівномірне використання окремих пунктів шкали. Трапляється, особливо при використанні впорядкованих шкал, градації яких супроводжуються словесними описами, що деяке значення змінної (ознака) систематично випадає з поля зору респондентів, хоча сусідні градації, що характеризують нижчий і вищий рівень вираженості ознаки, мають суттєве наповнення.

Наприклад, найчастіше десятибальну шкалу респонденти розцінюють як деяку п'ятибальну модифікацію, припускаючи, що «десять» відповідає «п'яти», «вісім» – «чотирьом», «п'ять» – «трьом» і т.д.

При цьому базові оцінки використовуються значно частіше за інші.

Визначення грубих помилок. У процесі виміру іноді виникають грубі помилки, причиною яких можуть бути неправильні записи вихідних даних, погані розрахунки, некваліфіковане використання вимірювальних засобів і т. п. Це проявляється в тому, що в рядах вимірів трапляються дані, що різко відрізняються від сукупності всіх інших значень. Щоб з'ясувати, чи потрібно ці значення визнати грубими помилками, встановлюють критичну межу так, щоб ймовірність перевищення її крайніми значеннями була досить малою та відповідала деякому рівню значущості. Це ґрунтується на тому, що поява у вибірці надмірно великих значень хоч і можлива як наслідок природної варіабельності значень, але мало ймовірна.

Якщо виявиться, що якісь крайні значення сукупності належать їй з дуже малою ймовірністю, то такі значення визнаються грубими помилками і виключаються з подальшого розгляду. Виявлення грубих помилок особливо важливо проводити для вибірок малих, обсягів: не виключеними з аналізу, вони суттєво спотворюють параметри вибірки.

2. Стійкість виміру.

Про високу надійність шкали можна говорити лише в тому випадку, якщо повторні вимірювання за допомогою тих самих об'єктів дають подібні результати стійкість перевіряється на одній і тій же вибірці досліджуваних об'єктів (респондентів). Порівняння ж середніх оцінок різних вибірок нічого не говорить про стійкість виміру як такого, а лише про репрезентативність вибірок та їх відповідності однієї, і тієї ж сукупності. Зазвичай стійкість перевіряй проведенням двох послідовних вимірів з певним часовим інтервалом - таким,

щоб цей проміжок не був занадто великий, щоб позначилася зміна самого об'єкта, але не надто травень, щоб респондент міг пам'яті «підтягувати» дані другого виміру до попереднього (тобто. його протяжність залежить від (об'єкт вивчення і коливається від двох до трьох тижнів).

1. Показником абсолютної стійкості шкали назвемо величину, що показує частку відповідей, що збігаються в послідовних пробах.

Тут придатні показники нестійкості, т. е. величини помилки, які враховують не просто факт розбіжності відповідей, а ступінь цього розбіжності. Помилки розраховуються принаймні на порядкові ознаки.

Підвищення стійкості виміру. Для вирішення цього завдання необхідно з'ясувати розрізняючі можливості пунктів: шкали, що передбачає чітку фіксацію респондентами окремих значень: кожна оцінка повинна бути суворо відокремлена від сусідньої. Насправді це означає, що у послідовних пробах респонденти практично повторюють свої оцінки. Отже, високої помітності поділів шкали повинна відповідати мала помилка.

Це завдання можна описати в термінах чутливості шкали, яка характеризується кількістю поділів, що припадають на одну і ту ж різницю в значеннях вимірюваної величини, тобто чим більше градацій, шкалою, тим/більше її чутливість. Однак чутливість не можна підвищувати простим збільшенням дрібності, бо висока чутливість при низькій стійкості є зайвою (наприклад, шкала 100 балів, а помилка вимірювання ± 10 балів).

Во і при малій кількості градацій, тобто при низькій чутливості, може бути низька стійкість, і тоді слід збільшити дрібність шкали. Так буває, коли респонденту нав'язують категоричні відповіді «так», «ні», а він віддав би перевагу менш жорстким оцінкам. І тому він вибирає в повторних випробуваннях іноді так, іноді ні для характеристики свого нейтрального становища.

У тому випадку, якщо виявлено змішання градацій, застосовують один із двох способів.

Перший спосіб, і підсумковому варіанті зменшують дрібність шкали (наприклад, зі шкали в 7 інтервалів переходять на шкалу в 3 інтервали).

Другий спосіб. Для пред'явлення респонденту зберігають колишню дрібність шкали і лише під час обробки укрупнюють відповідні її пункти (як це було показано вище).

Другий спосіб здається кращим, оскільки, як правило, велика дрібність шкал спонукає респондента і до більш активної реакції. При обробці даних інформацію слід перекодувати відповідно до проведеного аналізу розрізняючої здатності вихідної шкали.

Отже, запропоновані методи аналізу доцільні при відпрацюванні остаточного варіанта методики. Аналіз стійкості окремих питань шкали дозволяє; а) виявити погано сформульовані питання, їхнє неадекватне розуміння різними респондентами; б) уточнити інтерпретацію шкали запропонованої для оцінки того чи іншого явища, виявити більш оптимальний варіант дрібності значення шкали.

Вивчення стійкості остаточного варіанта методики дасть уявлення про надійність даних (пов'язану стійкістю), які будуть отримані в основному дослідженні.

3. Обґрунтованість виміру.

Перевірка обґрунтованості шкали здійснюється лише після того, як встановлено достатню правильність і стійкість вимірювання вихідних даних. Як уже зазначалося, перевірка обґрунтованості – досить складний процес і, як правило, не до кінця можна розв'язати. І тому недоцільно спочатку застосовувати трудомістку техніку виявлення обґрунтованості, а потім переконуватися в неприйнятності даних внаслідок їх низької стійкості.

Обґрунтованість даних виміру – це доказ відповідності між тим, що виміряно, і тим, що мало бути виміряне. Деякі дослідники вважають за краще виходити з так званої готівкової обґрунтованості, тобто обґрунтованості у поняттях використаної процедури.

Наприклад, вважають, що задоволеність роботою – це властивість, що міститься у /відповідях питанням: «Чи задоволені Ви роботою?» У серйозному соціологічному дослідженні, яке має на меті перевірку деяких теоретичних гіпотез, такий суто емпіричний підхід є неприйнятним.

Зупинимося на можливих формальних підходах до з'ясування обґрунтованості методики. Їх можна розділити втричі групи: 1) конструювання, типології відповідно до цілями дослідження з урахуванням кількох ознак; 2) використання паралельних даних; 3) суддівські процедури.

Перший варіант не можна вважати формальним методом - це лише деяка схематизація логічних міркувань, початок процедури обґрунтування, яка може бути на цьому і закінчена, а може бути підкріплена більш потужними засобами.

Другий варіант вимагає використання принаймні двох джерел для виявлення однієї й тієї ж властивості. Обґрунтованість визначається ступенем узгодженості відповідних даних.

В останньому випадку ми покладаємося на компетентність суддів, яким пропонується визначити, чи вимірюємо > ми потрібну Вам властивість чи щось інше.

Розглянемо запропоновані варіанти послідовно.

Конструювання, типологія. Один із способів – використання контрольних питань, які разом із основними дають більше наближення до змісту досліджуваного властивості, розкриваючи різні його сторони.

Наприклад, можна визначати задоволеність роботою лобовим питанням: «Чи влаштує Вас Ваша нинішня робота?» Комбінація його з двома іншими непрямыми: «Чи бажаєте Ви перейти на іншу роботу?» та «Припустимо, що Ви з якихось причин тимчасово не працюєте. Повернулися б Ви на своє колишнє місце роботи?» дозволяє зробити надійнішу диференціацію респондентів.

Паралельні дані. Нерідко доцільно розробити два рівноправні прийоми вимірювання заданої ознаки, що дозволяє встановити обґрунтованість методів один одного, тобто. підвищити загальну обґрунтованість шляхом зіставлення двох незалежних результатів.

Класифікуємо паралельні процедури в залежності від співвідношення методів та виконавців: а) кілька методів – один виконавець. б) один спосіб – кілька виконавців; в) кілька методів – кілька виконавців.

Декілька методів – один виконавець. Тут один і той же виконавець використовує два або більше різних методів для вимірювання однієї й тієї ж властивості.

Один метод – кілька виконавців. Якщо метод надійний, різні виконавці дадуть східну інформацію, але якщо їх результати погано узгоджуються, то або виміри ненадійні, або результати окремих виконавців не можна вважати рівноцінними. В останньому випадку треба встановити, чи не можна вважати якусь групу результатів заслуговує на більшу довіру. Вирішення цього завдання тим паче важливе, якщо передбачається, що однаково припустиме отримання інформації будь-яким із методів (наприклад, використання самооцінок проти оцінок). Аналіз паралельних даних за допомогою наведених нижче процедур дозволить встановити правильність такого припущення.

Для кількісних ознак під час вирішення питання узгодженості оцінок кількох виконавців пропонується виявити помилки відповідності однією з прийомів, розглянутих щодо стійкості.

Насамперед, оскільки ми маємо тут випадок прямих групових спостережень, найбільш адекватною оцінкою збігу даних є середня помилка.

Декілька методів і кілька виконавців. Одним із способів встановлення обґрунтованості вимірювання деякої якості в одного і того ж респондента (об'єкта) є фіксування даної якості різними виконавцями, що володіють різними методами. Як і попередніх випадках, тут не можна встановити якусь абсолютну обґрунтованість, тому розглядається лише обґрунтованість одного способу щодо іншого.

Метод суддівства при обґрунтуванні процедур виміру. Один із найпоширеніших підходів до встановлення обґрунтованості – це використання так званих суддів. Дослідники звертаються до певної групи людей з проханням виступити як судді або компетентні особи. Їм пропонують набір ознак, призначений для вимірювання явища, що вивчається, і просять оцінити правильність віднесення кожної з ознак до цього об'єкта. Спільна обробка думок суддів дозволить присвоїти ознакам ваги або, що те саме, шкільні оцінки у вимірі явища, що вивчається. Як набір ознак може виступити список окремих суджень, серія предметів, сукупність обстежуваних осіб і т.д.

Процедури суддівства різноманітні. Спосіб виявлення ставлення ознак до вимірюваної властивості визначає сутність методу. Це можуть бути методи парних порівнянь, ранжирування, послідовних інтервалів тощо. У кожному разі, обираючи ту чи іншу техніку суддівства, необхідно враховувати її специфічні можливості, що впливають на рівень обґрунтування суддівських оцінок.

Питання про те, кого слід вважати суддями, є досить дискусійним. Судді, які обираються як представники досліджуваної сукупності, так чи інакше повинні представляти її мікромодель: за оцінками суддів дослідник визначає, наскільки адекватно будуть витлумачені респондентами пункти опитувальної процедури чи інші звернені до респондента стимули.

Однак при відборі суддів виникає питання, який вплив власних установок суддів на їх оцінки, адже ці установки. Можуть істотно відрізнятись від установок обстежуваних щодо того самого об'єкта.

Зрозуміло, що в кожному конкретному випадку слід здійснювати контроль таких помилок стосовно даної вибірки суддів.

Так, використовуючи чоловіків і жінок як суддів для оцінки потенційних творчих можливостей різних занять на дозвіллі, виявили, що установки суддів-чоловіків істотно відрізняються від установок суддів-жінок. Більше того, їх установки залежать від того, чи захоплюється сам суддя, що судить даним видом.

Наприклад, жінки, які займаються рукоділлям, значно вищі оцінюють творчі можливості цього заняття, ніж ті, які не займаються.

У загальному вигляді рішення, проблеми полягає в тому, щоб:

а) уважно проаналізувати склад суддів з погляду адекватності їх життєвого досвіду та ознак соціального статусу відповідним показникам обстежуваної генеральної сукупності;

б) виявити ефект індивідуальних ухилень у оцінках суддів щодо загального розподілу оцінок. Нарешті, слід оцінити як якість, а й обсяг вибіркової сукупності суддів. Тут також немає однастайності між фахівцями. Рекомендується брати то 25-30 чоловік, то 200-300 і більше. Серйозних обґрунтувань в обох випадках не наводиться.

Численні експерименти з виявлення рівня надійності вихідної інформації дозволяють укласти, що в процесі відпрацювання інструментів вимірювання з боку їх надійності доцільна наступна послідовність основних етапів роботи:

1. Попередній контроль обґрунтованості методів вимірювання первинних даних на стадії проб методики. Тут перевіряється, наскільки інформація відповідає своєму призначенню по суті та які межі подальшої інтерпретації даних. Для цієї мети достатні невеликі вибірки в 10-20 спостережень з подальшим коригуванням структури методики.

2. Пілотаж методики і ретельна перевірка стійкості вихідних даних, особливо підсумкових показників, індексів, багатовимірних шкал тощо. п. На цьому етапі необхідна вибірка щонайменше 100 людина, що представляє мікромодель реальної сукупності обстежуваних з урахуванням представництва за суттєвими характеристиками об'єкта дослідження.

3. У період загального пілотажу здійснюються всі необхідні операції, що стосуються перевірки, рівня обґрунтованості. Результати аналізу даних генерального пілотажу призводять до вдосконалення методики, доопрацювання всіх її деталей й у результаті – отримання остаточного варіанта методики для основного дослідження.

4. На початку основного дослідження бажано провести перевірку використовуваного варіанта методики на стійкість для того, щоб розрахувати точні показники її стійкості. Досліджуване уточнення меж обґрунтованості проходить через аналіз самого дослідження.