

УДК 159.99

DOI <https://doi.org/10.30970/PS.2021.9.24>

## РОЗРАХУНОК ОБСЯГУ ВИБІРКИ ЯК НАРІЖНИЙ КАМІНЬ ПЛАНУВАННЯ НАУКОВОГО ДОСЛІДЖЕННЯ

**Валерій Олефір**

*Харківський національний університет імені В. Н. Каразіна,  
майдан Свободи, 4, м. Харків, Україна, 61022  
e-mail: vaolefir@gmail.com*

**Валерій Боснюк**

*Національний університет цивільного захисту України,  
вул. Чернишевська, 94, м. Харків, Україна, 61023  
e-mail: bosnyk@nuczu.edu.ua*

Оскільки результати наукових досліджень екстраполюються на генеральну сукупність, то під час планування роботи одним з головних питань є формування репрезентативної, оптимальної за обсягом вибірки. Мала вибірка провокує ненадійні результати та зменшує шанс виявлення очікуваної величини ефекту, водночас занадто велика вимагає залучення додаткових ресурсів та сприяє виявленню статистично значущих результатів, практична цінність яких досить низька. Тобто обґрунтування розміру вибірки є критичним аспектом дизайну дослідження. Необхідно залучити таку кількість осіб, щоб бути впевненим, якщо практично важливий ефект існує, то він із високим ступенем імовірності буде виявлений статистичним аналізом. Такий підхід у межах методології доказової психології дозволяє науковцям приймати статистичні рішення у дослідженнях більш раціонально та обґрунтовано.

Відповідь на питання про необхідний обсяг вибірки залежить від трьох основних факторів, які повинні враховуватися під час планування дослідження: 1) величини ймовірності помилки першого роду, на основі якої приймається рішення про прийняття/відхилення нульової гіпотези; 2) статистичної потужності дослідження як ступеня впевненості в одержанні значущого результату, якщо він справді існує; 3) величини очікуваного розміру ефекту.

У статті доказується, що загальноприйнятою практикою науковців має стати проведення апріорного аналізу статистичної потужності для визначення необхідного розміру вибірки в дослідженні. Для реалізації цієї необхідності пропонується скористатися безплатним програмним забезпеченням, таким як пакет G\*Power або програмно-статистичне середовище із відкритим кодом R. Наводяться деякі приклади застосування їх у практичних ситуаціях.

Як підсумок формулюється висновок, що під час планування психологічного дослідження, незважаючи на методологічні, практичні та етичні причини, науковці регулярно нехтують таким важливим аспектом методології доказової психології, як розрахунок необхідного обсягу вибірки. У разі відсутності такого розрахунку результати дослідження слід інтерпретувати з обережністю.

*Ключові слова:* обсяг вибірки, статистична потужність, розмір ефекту, помилка першого роду, помилка другого роду, доказова психологія.

**Постановка проблеми.** Доказова цінність дослідження і його внесок у кумулятивний науковий прогрес залежить від обґрунтованого дизайну та чіткої статистичної звітності. Для отримання науково обґрунтованих висновків на підставі результатів проведеного дослідження необхідно враховувати невизначеність багатьох характеристик.

Одним із недооцінених факторів коректної інтерпретації результатів аналізу даних є забезпечення дослідження оптимальною за обсягом вибіркою, під якою розуміється необ-

хідна й достатня кількість досліджуваних, за якої ймовірність прийняття правильного рішення є максимальною. Вибірки не повинні бути занадто великими або малими, оскільки вони можуть знизити рівень довіри до отриманих висновків та подальшого використання на практиці. Надмірно мала кількість досліджуваних зменшує шанс виявлення очікуваної величини ефекту та не дозволяє екстраполювати результати на широку генеральну сукупність, тоді як занадто велика вибірка сприятиме виявленню статистично значущих відмінностей, практична цінність яких досить низька. Тобто обґрунтування розміру вибірки є критичним аспектом дизайну дослідження. Проте на практиці у вітчизняних психологічних дослідженнях здебільшого його ігнорують, що також проявляється у відсутності звітності про попередній аналіз статистичної потужності та дискусії про величину можливого розміру ефекту.

Оцінка розміру вибірки на основі заданої потужності та спостережуваного ефекту – не формальна процедура, а наріжний камінь планування успішного дослідження. Вона необхідна для впевненості, що виявлений ефект не є результатом випадковості.

**Аналіз останніх публікацій та досліджень.** Натепер соціальні науки перебувають у кризі відтворюваності результатів. Вона розпочалася з медицини [7], але більше всього проявилась у психології. Криза остаточно стала очевидною, коли у 2015 році 220 учених перевірили 100 досліджень, опублікованих у провідних психологічних журналах, у 97 з яких автори заявляли про статистично значущі результати ( $p < 0,05$ ). Під час відтворення досліджень подібних результатів удалося отримати лише в 36 роботах. У середньому після реплікацій розмір ефекту був удвічі меншим порівняно з оригінальними результатами [11]. Головний редактор авторитетного журналу “The Lancet” Річард Хортон констатував: «Принаймні половина наукової літератури не відповідає дійсності. У результаті досліджень із малими обсягами вибірки, мізерними розмірами ефектів, некоректним аналізом та нав'язливою ідеєю слідувати модним тенденціям сумнівної цінності наука здійснила поворот у сторону невігластва» [6, с. 1380].

В основі систематичних помилок публікацій у психологічній науці лежать прийняті неправильні рішення в оцінці помилок першого (хибнопозитивні результати) та другого (хибнонегативні результати) роду. Для їх мінімізації найбільш прийнятним інструментом є ретельне планування дослідження із суворим дотриманням затвердженого плану та грамотне використання методів статистичного аналізу.

З метою зменшення кількості помилок першого типу та подолання кризи відтворюваності результатів досліджень 72 науковці у спільній статті-маніфесті запропонували змінити у наукових роботах поріг рівня статистичної значущості з 0,05 до 0,005. Перехід до  $\alpha = 0,005$ , залишаючи водночас 80% статистичної потужності, вимагатиме збільшення обсягу вибірки приблизно на 70%. На думку авторів, незважаючи на необхідність залучення більшої кількості досліджуваних, у майбутньому такий підхід дозволить зекономити значні ресурси через зменшення досліджень на хибних засадах, підвищення ефективності значно перевищить втрати [2].

В іншій статті-препринті, яку підписали 87 авторів, стверджується, що це нерозумно, і замість впровадження  $\alpha = 0,005$  ми повинні обґрунтувати експериментальний план та всі аналітичні рішення, у тому числі порогове значення « $p$ », статистичну потужність і обсяг вибірки, які застосовуються для визначення реального розміру ефекту [9].

Деякі редактори наукових журналів у своїх збірниках праць взагалі заборонили публікації результатів використання процедури перевірки значущості нульової гіпотези. У редакційній статті “Basic and Applied Social Psychology” вказується, що значення « $p$ » стало опорою для вчених, які працюють зі слабкими даними. Вони пишуть, що рівень  $p < 0,05$  дуже легко обійти та іноді є виправданням для досліджень низької якості. На заміну редактори вимагали чіткого опису статистику із зазначенням виявлених розмірів

ефекту. Було звернено увагу, що психологи здебільшого формують недостатні за обсягом вибірки. Збільшення кількості досліджуваних дозволить зробити описову статистику більш стійкою до випадкових результатів [13].

Дискусія щодо порогових значень «*p*» та їхнього використання, наявність систематичних помилок у публікаціях та криза відтворюваності результатів – усе це актуалізує питання доказової цінності наукових досліджень.

Проблема полягає в готовності багатьма дослідниками формулювати впевнені узагальнюючі висновки з недбало описаних даних, отриманих у погано спланованих дослідженнях [10]. Кожен дослідник у своїй галузі зобов'язаний дотримуватися вимог до організації дослідження та аналізу даних. Для покращення сучасної дослідницької практики та уникнення помилок першого та другого роду важливо ретельно планувати своє дослідження з обґрунтуванням обсягу вибірки досліджуваних.

**Мета дослідження** – систематизація проблеми визначення оптимально необхідно обсягу вибірки під час планування психологічних досліджень.

**Виклад основного матеріалу.** Для більшості психологічних досліджень визначення необхідного обсягу вибірки – досить проблематичне питання, адекватне обґрунтування якого відображене далеко не в кожному з них. Здебільшого дослідники орієнтуються на умовні рекомендації або на свої можливості зібрати первинні дані. Таке неправильне та халатне ставлення до вирішення цього питання може нівелювати результати будь-якого дослідження, піддати сумнівам коректність інтерпретації числових результатів статистичних методів. Насправді, кожен науковець на етапі планування дослідження повинен поставити собі питання такого типу: «Я зможу вивчити *X* осіб. Чи є сенс проводити таке дослідження?» або «Наскільки велику вибірку мені необхідно сформулювати, щоби виявити очікувану величину ефекту?» На ці сформульовані питання можна відповісти за допомогою аналізу статистичної потужності – набору методів для планування дослідження, які дозволяють оптимізувати використання наявних ресурсів.

Не всі розуміють, що питання розміру вибірки – справа не однієї формули, а досить складна тема, яка потребує аналізу популяції, на яку будуть поширюватися результати, детального усвідомлення дослідницьких гіпотез та завдань, розуміння дизайну дослідження, типів вимірювальних шкал, у яких представлені дані, направленість гіпотез, відношення між вибірками, чутливість статистичних критеріїв, визначення найменшої величини практично цінного розміру ефекту, передбачення вибування осіб у тривалих проспективних дослідженнях тощо.

Формування малої вибірки в дослідженні збільшує ймовірність прийняття помилкової гіпотези, результат часто буває недооціненим. А це означає, що люди були марно піддані опитуванню та/або експериментальному впливові. Крім того, фінансові та часові ресурси також були витрачені даремно, оскільки в кінцевому рахунку дослідження не покращило практику психологічної допомоги.

З іншого боку, не завжди коректна поширена думка, що великі вибірки ідеально підходять для досліджень та статистичного аналізу. Формування значної вибірки досліджуваних пов'язане з різними труднощами. Не етично залучати до експерименту більше осіб, ніж це необхідно, тому що вони піддаються впливові, результат якого зарання невідомий. Використання великої кількості досліджуваних також вимагає додаткових фінансових, організаційних та людських затрат, ніж це необхідно для вирішення поставлених завдань у дослідженні. Крім того, підвищується ймовірність виявлення статистично значущих результатів, які мають низьку практичну цінність.

Для вирішення цієї дилеми є компромісне і єдино правильне рішення – ретельне планування психологічного дослідження з розрахунком необхідного обсягу вибірки. Це є

обов'язковою умовою якісного дослідження в межах методології доказової психології та гарантує, якщо практично важливий ефект існує, то він із високим ступенем ймовірності буде виявлений статистичним аналізом. Такий підхід дозволяє приймати рішення більш раціонально та обгрунтовано. Він також корисний для критичної оцінки нової інформації, отриманої під час читання наукової літератури.

Відповідь на питання про необхідний обсяг вибірки залежить від трьох основних факторів [3], які повинен враховувати науковець під час планування дослідження:

1) величини ймовірності помилки першого роду, на основі якої приймається рішення про прийняття/відхилення нульової гіпотези;

2) статистичної потужності дослідження як ступеня впевненості в одержанні значущого результату, якщо він справді існує;

3) величини очікуваного розміру ефекту.

Максимально прийнятна ймовірність помилки першого роду позначається  $\alpha$ , чим нижчий її рівень, тим більший обсяг вибірки необхідно залучити в дослідженні. Цей рівень визначає сам експериментатор до початку збору даних як статистичну межу готовності помилитися під час прийняття рішення. Це ймовірність хибнопозитивного результату, тобто відхилення нульової гіпотези, коли в генеральній сукупності вона істинна. Якщо  $p$ -значення менше  $\alpha$ -рівня, то нульова гіпотеза вважається відхиленою як малоймовірна для отриманих результатів. У психології найчастіше ймовірність помилки першого роду встановлюється на рівні 0,05, хоча це лише данина традиції, кожен фахівець може вибрати свій рівень значущості залежно від предмета дослідження. Заповітне значення 0,05 свідчить, що отриманий результат може виявитися випадковим у 1 випадку з 20. А це не так і мало, тобто вплив ймовірності на результат дослідження часто буває недооцінений, особливо в дослідженнях із залученням багатьох діагностичних параметрів.

Необхідно зазначити, що низька величина помилки першого роду не завжди корисна, позаяк збільшується ймовірність помилки другого роду ( $\beta$ ) – ймовірність хибно-негативного результату, тобто прийняття рішення не відхиляти нульову гіпотезу, хоча в дійсності вона помилкова. У дослідженнях розмір цієї помилки не зазначається відкрито, але про неї слід обов'язково пам'ятати, формулюючи висновок, особливо працюючи з вибірками малого розміру. На практиці рекомендується, щоби цей параметр не перевищував значення в 0,2.

Помилка другого роду зворотно пов'язана зі статистичною потужністю дослідження, розраховується як  $1-\beta$ . Статистична потужність – здатність виявити статистично значущий ефект, якщо він справді існує. Іншими словами, ймовірність відхилення нульової гіпотези, коли вона помилкова. Зазвичай дослідниками вона визначається в діапазоні 80–90% (0,8–0,9). Проте якщо вивчаються рідкісні явища та дуже складно сформувати необхідний обсяг вибірки, то потужність можна знизити до 70%, щоби мати можливість все-таки провести дослідження. Чим вища статистична потужність, тим більшу кількість осіб необхідно набрати в кожну групу, причому це збільшення не пропорційне, а експоненційне, тобто кожне наступне зростання потужності в 1% вимагає все більшу кількість досліджуваних.

Перед початком дослідження критично важливо знати статистичну потужність. Є сенс розпочинати наукову роботу тільки у ситуації з високою ймовірністю виявлення значущого ефекту. Безвідповідально витрачати ресурси за інших умов, тому що ми ризикуємо проводити зарання приречені на невдачу дослідження. Наприклад, якщо у разі потужності у 40% не виявлено статистично значущих відмінностей, то це не означає, що ми підтвердили нульову гіпотезу, насправді ми отримали невизначений результат. Такий результат може бути пояснений як недостатньою кількістю досліджуваних, так і справді

відсутньою різницею. У наведеному прикладі ймовірність не помітити ефект становить 60%. Ми не виявили ефект, але не можемо стверджувати, що його немає. Отже, перш ніж формулювати висновок про відсутність відмінностей, дослідник повинен з'ясувати чи достатньою була потужність для її виявлення.

Необхідно дуже уважно оцінити баланс між помилками першого та другого роду, приймаючи рішення про рівень довіри до отриманих результатів та їхнє подальше використання на практиці (табл. 1). Зазвичай для більшості досліджень оптимальним поєднанням вважається 5%  $\alpha$ -рівень та 80% статистичної потужності. Проте, як уже зазначалося вище, баланс між ризиком прийняття неправильного рішення і дійсністю проведення дослідження визначається фахівцем індивідуально в кожному випадку, тільки йому відомо, наскільки важливі ті чи інші висновки та які можливості формування вибірки необхідного обсягу.

Таблиця 1

## Статистичні рішення для оцінки гіпотез

Рішення дослідника	У дійсності	
	Істинна гіпотеза $H_0$	Істинна гіпотеза $H_1$
Гіпотеза $H_0$ відхиляється	Хибне рішення. Помилка 1-го роду. Ймовірність = $\alpha$ .	Правильне рішення. Статистична потужність. Ймовірність = $1-\beta$ .
Гіпотеза $H_0$ приймається	Правильне рішення. Довірча ймовірність. Ймовірність = $1-\alpha$ .	Хибне рішення. Помилка 2-го роду. Ймовірність = $\beta$ .

Ще одним параметром дослідження, який впливає на визначення обсягу вибірки, є розмір ефекту.

Розмір ефекту – кількісне відображення величини (ступеня прояву) явища, що являє науковий інтерес [8, с. 140]. Позначається аббревіатурою ES (effect size). Розмір ефекту відображає силу взаємозв'язку між змінними, величину різниці ознаки між групами, відповідність теоретичної моделі емпіричним даним та дозволяє оцінити корисність, важливість, цінність, осмисленість такого відношення. Він за своєю суттю близький до поняття практичної значущості дослідження – головного орієнтира наукової роботи.

Розміри ефекту дозволяють дослідникам вийти за межі дихотомічного мислення, що нав'язує процедура перевірки значущості нульової гіпотези: «Є відмінність (взаємозв'язок) чи немає?» до більш важливого аспекту: «Яка величина цих відмінностей (сила взаємозв'язку)?», відійти від акцентування на ідентифікації статистичної значущості та перейти до більш цінного і простішого для розуміння кількісного опису величини розміру ефекту, більш наукового підходу до накопичення знань.

Як показник розміру ефекту можуть виступати нестандартизовані величини (різниця між двома середніми арифметичними значеннями, відсотками, нестандартизовані регресійні коефіцієнти тощо) та стандартизовані індекси, які розраховуються для представлення ефекту різних величин в єдиному масштабі. Нестандартизовані показники розміру ефекту корисні, коли досліджувані змінні мають безпосередньо зрозуміле значення (наприклад, швидкість прийняття рішення у стресових умовах). Стандартизовані індекси необхідно застосовувати, коли вимірювання не має внутрішнього значення, наприклад, числа в шкалі Лайкерта; коли в дослідженнях використовувалися різні шкали, тому пряме порівняння результатів неможливе; або коли розмір ефекту розглядається у контексті мінливості в популяції.

Розмір ефекту – це параметр, який найскладніше обґрунтувати під час розрахунку обсягу вибірки, тому що його визначення часто і є метою самого дослідження. Є декілька підходів для вирішення цієї проблеми.

Перший із них передбачає літературний аналіз раніше опублікованих даних із теми дослідження. Проте на практиці реалізація такого підходу ускладнюється малою кількістю метааналізів, які узагальнюють розміри ефектів у межах конкретної проблеми, а самостійний аналіз проблематичний через відсутність у більшості опублікованих робіт детального й чіткого опису процедурних особливостей дослідження та його результатів у термінах статистичних ефектів.

Інший підхід полягає у визначенні найменшого практично важливого розміру ефекту, який ми хочемо виявити. Дослідники повинні бути зацікавлені у встановленні межі, нижче якої виявлений ефект не буде мати цінності, навіть якщо він статистично значущий.

Також можна орієнтуватися на узагальнені значення для різних градацій ефекту, запропоновані Дж. Коеном [4], залежно від використаних статистичних критеріїв (табл. 2).

Таблиця 2

**Популярні індекси кількісної оцінки розміру ефекту та їх рекомендована інтерпретація**

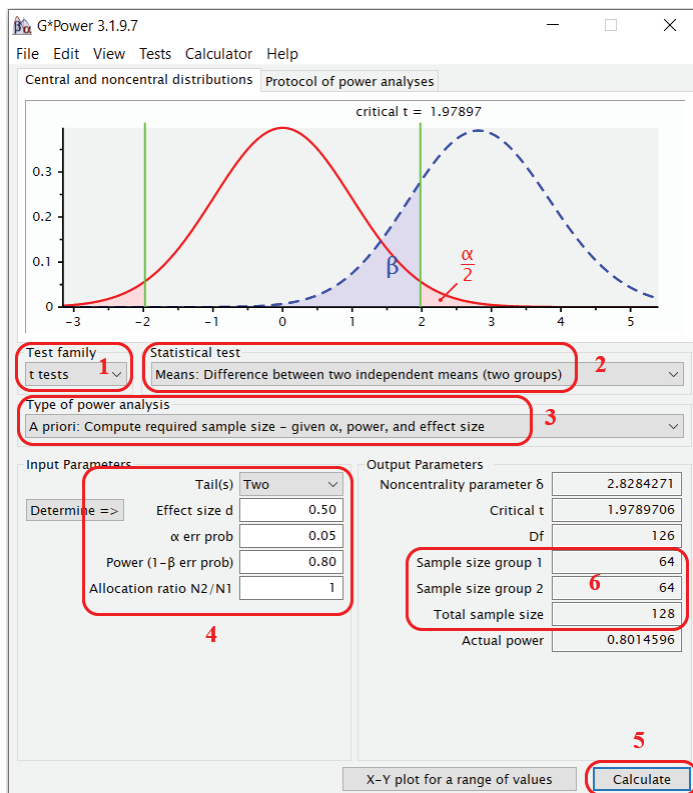
Статистичний критерій	Індекс ES	Розмір ефекту			
		несуттєвий	малий	середній	великий
Тест Стьюдента	<i>d</i>	0,00–0,19	0,20–0,49	0,50–0,79	0,80 >
Манна-Уїтні, Вілкоксона	<i>r</i>	0,00–0,09	0,10–0,29	0,30–0,49	0,50–1,00
Дисперсійний аналіз	$\eta^2$	0,00–0,009	0,01–0,05	0,06–0,13	0,14–1,00
Регресійний аналіз	<i>f</i> <sup>2</sup>	0,00–0,01	0,02–0,14	0,15–0,34	0,35–1,00
Кореляційний аналіз	<i>r</i>	0,00–0,09	0,10–0,29	0,30–0,49	0,50–1,00
$\chi^2$ -Пірсона	<i>w</i>	0,00–0,09	0,10–0,29	0,30–0,49	0,50–1,00
Тест пропорцій	<i>h</i>	0,00–0,19	0,20–0,49	0,50–0,79	0,80–1,00

Варто пам'ятати, що вказані еталонні значення, хоча й популярні серед психологів, проте не універсальні, для кожного окремого дослідження повинні застосовуватися індивідуальний підхід. Орієнтація в психології на узагальнені величини розміру ефекту ускладнюється значною різномірністю наукових напрямів. Показано, що є значна відмінність у медіанах розподілів розмірів ефекту в різних галузях психології [12].

Альтернативним підходом є зміна досліджуваних параметрів та спостереження за тим, як це відобразиться на обсязі вибірки та статистичній потужності [1]. Як варіант, можна провести самостійне невелике пілотне дослідження, щоби зрозуміти варіативність даних.

Узагальнюючи обговорення, можна стверджувати, що загальноприйнятою практикою має стати проведення апріорного аналізу статистичної потужності для визначення необхідного розміру вибірки в запланованому дослідженні. Для реалізації цієї необхідності можна скористатися безплатним програмним забезпеченням: пакетом G\*Power (<http://www.gpower.hhu.de/>) або програмно-статистичним середовищем із відкритим кодом R (<https://www.r-project.org/>).

Для прикладу за допомогою G\*Power визначимо необхідний обсяг вибірки у поширеній на практиці ситуації порівняння середніх значень статистичним критерієм Стьюдента у двох незалежних групах (див. рис. 1).



**Рис. 1.** Алгоритм розрахунку обсягу вибірки у статистичному пакеті G\*Power для порівняння середніх значень у двох незалежних групах

1. Вибираємо групу тестів: t tests.
2. Вказуємо необхідну процедуру: Means: Difference between two independent means (two groups).
3. Вибираємо опцію апіорного розрахунку необхідного обсягу вибірки – sample size.
4. Для більшості емпіричних ситуацій застосовується двобічна (ненаправлена) гіпотеза, тому в графі Tail(s) вибираємо Two; зазначаємо стандартизований мінімальний розмір ефекту, який для нас має цінність (Effect size  $d$ ), наприклад, 0,50 (інтерпретацію див. у табл. 2); максимальний прийнятний рівень ймовірності помилки першого роду визначимо на рівні 0,05 ( $\alpha$  err prob); статистичну потужність (Power) зазначимо 0,80. Якщо обсяг досліджуваних у двох групах планується однаковим, то залишаємо співвідношення рівне 1 (Allocation ratio  $N2/N1 = 1$ ).
5. Натискаємо на кнопку розрахунку (Calculate) та отримуємо результат.
6. Для нашого прикладу необхідним обсягом вибірки слід вважати по 64 досліджуваних у кожній групі, загальна кількість осіб – 128 (Total sample size).

Ми розглянули лише один простий типовий приклад до планування обсягу вибірки в дослідженні. Тим не менше принцип організації даних і здійснення розрахунків у пакеті G\*Power аналогічний і для інших завдань: встановлення кореляційних зв'язків, порівняння пропорцій, знаходження відмінностей у ситуації наявності більше двох вибірок,

аналіз регресії тощо. Для більш детальної інформації необхідно звернутися до Мануалу G\*Power – посібника користувача [5].

Також для здійснення розрахунків пропонується використовувати пакет `rwr` мови програмування R, який дозволяє виконати аналіз потужності за алгоритмами, описаними Дж. Коеном [4]. Деякі з найважливіших функцій представлені в табл. 3.

Таблиця 3

**Функції із пакета `rwr` для мови програмування R**

Функція	Розрахунок потужності для:
<code>rwr.2p.test()</code>	двох пропорцій (з однаковими обсягами вибірок)
<code>rwr.2p2n.test()</code>	двох пропорцій (з різними обсягами вибірок)
<code>rwr.anova.test()</code>	однофакторного ANOVA
<code>rwr.chisq.test()</code>	тесту $\chi^2$ -Пірсона
<code>rwr.f2.test()</code>	загальної лінійної регресійної моделі
<code>rwr.r.test()</code>	кореляції
<code>rwr.t.test()</code>	t-тесту (одновибіркового, для незалежних вибірок, для залежних вибірок)
<code>rwr.t2n.test()</code>	t-тесту (дві вибірки різного обсягу)

Для кожної з цих функцій необхідно вказати три з чотирьох величин (обсяг вибірки, розмір ефекту, рівень статистичної значущості, потужність) і розраховується остання.

Розглянемо для прикладу функцію `rwr.r.test()`, яка призначена для аналізу потужності коефіцієнтів кореляції. Для неї аргументами є: `n` – обсяг вибірки; `r` – коефіцієнт кореляції, який виступає показником розміру ефекту; `sig.level` – рівень статистичної значущості; `power` – потужність критерію; `alternative` – тип гіпотези: “two.sided” (ненаправлена), “greater”, “less” (направлена).

Нехай між змінними сила зв’язку в генеральній сукупності описується як середня ( $r = 0,3$ ), при цьому дослідник хоче забезпечити можливість виявлення цієї залежності з 80% ймовірністю, статистична значущість встановлена ним на рівні 0,05, а гіпотеза є ненаправленою:

```
> library(rwr)
> rwr.r.test(r=0.3, power=0.8, sig.level=0.05, alternative="two.sided")
Результат:
n = 84.74891
r = 0.3
sig.level = 0.05
power = 0.8
alternative = two.sided
```

Таким чином, щоб з високою ймовірністю виявляти середній за величиною зв’язок, необхідно залучити не менше 85 досліджуваних. Якщо у цьому прикладі підвищити рівень статистичної значущості до 0,01, то рекомендований обсяг вибірки зросте до 125 осіб. Якщо й при цьому зв’язок між змінними передбачається слабким ( $r = 0,1$ ), то необхідно за таких умов уже 1163 досліджуваних. Використовуючи такі обчислення на етапі планування наукової роботи, вчений може отримати відповідь на питання необхідного залучення обсягу вибірки для виявлення заданої величини розміру ефекту з високим ступенем впевненості за результат. Правильне і зворотне, цей аналіз дозволяє оцінити ймовірність виявлення ефекту заданої величини та ступеня впевненості у разі конкретного обсягу вибірки. Якщо ймовірність досить низька, то розумно змінити екс-



перимент або відмовитися від нього. Наприклад, якщо у вашому розпорядженні є вибірка в кількості 30 осіб і ви намагаєтесь виявити зв'язок середньої сили у разі встановленого рівня значущості 0,01, то якою буде статистична потужність? Виявляється, вона в такому разі становить усього 0,156, тобто існує дуже низька ймовірність виявити таку закономірність. Тому за відсутності статистично значущих ефектів в уже проведених дослідженнях досить корисно оцінити досягнуту апостеріорну потужність: якщо вона була низькою, скажемо, менше 0,8, то це означає, що наявного обсягу вибірки було просто не досить для обґрунтування висновку.

**Висновки.** Під час планування психологічного дослідження, незважаючи на методологічні, практичні та етичні причини, науковці регулярно нехтують таким важливим аспектом методології доказової психології, як розрахунок необхідного обсягу вибірки. Насправді, від того, наскільки адекватним буде це число, залежить рівень довіри до отриманих результатів. Оцінка обсягу вибірки необхідна для того, щоб упевнитися, якщо практично важливий ефект існує, то він із високим ступенем ймовірності буде виявлений статистичним аналізом. За відсутності такого розрахунку результати дослідження слід інтерпретувати з обережністю.

### Список використаної літератури

1. Кабаков Р.И. Р в действии. Анализ и визуализация данных в программе R. Москва : ДМК Пресс, 2014. 588 с.
2. Benjamin D.J., Berger J.O., Johannesson M., Nosek B.A., Wagenmakers E.J., Berk, R., ... Johnson, V.E. Redefine statistical significance. *Nature Human Behaviour*. 2018. Vol. 2. Issue 1. P. 6–10. URL: <https://doi.org/10.1038/s41562-017-0189-z>.
3. Bonett D.G. *Sample Size Planning for Behavioral Science Research*. Santa Cruz : University of California, 2016. 91 p. URL: <http://people.ucsc.edu/~dgbonett/sample.html>.
4. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edn. New York : Lawrence Erlbaum Associates, 1988. 567 p. <https://doi.org/10.4324/9780203771587>.
5. G\*Power 3.1 manual: March 1, 2017. URL: [https://www.psychologie.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche\\_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf](https://www.psychologie.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf).
6. Horton R. Offline: What is medicine's 5 sigma? *The Lancet*. 2015. Vol. 385. Issue 9976. P. 1380. URL: [https://doi.org/10.1016/S0140-6736\(15\)60696-1](https://doi.org/10.1016/S0140-6736(15)60696-1).
7. Ioannidis J.P. Why Most Published Research Findings Are False. *PLoS Med*. 2005. Vol. 2. Issue 8. e124. URL: <https://doi.org/10.1371/journal.pmed.0020124>.
8. Kelley K., Preacher K.J. On effect size. *Psychological Methods*. 2012. Vol. 17. Issue 2. P. 137–152. URL: <https://doi.org/10.1037/a0028086>.
9. Lakens D., Adolff F.G., Albers C.J., Anvari F., Apps M.A.J., Argamon S. E., ... Zwaan R.A. Justify your alpha. *Nature Human Behaviour*. 2018. Vol. 2. Issue 3. P. 168–171. URL: <https://doi.org/10.1038/s41562-018-0311-x>.
10. Matthews R., Wasserstein R.L., Spiegelhalter D. The ASA's p-value statement, one year on. *Significance*. 2017. Vol. 14. Issue 3. P. 38–41. URL: <https://doi.org/10.1111/j.1740-9713.2017.01021.x>.
11. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*. 2015. Vol. 349. Issue 6251. aac4716. URL: <https://doi.org/10.1126/science.aac4716>.
12. Schäfer T., Schwarz M.A. The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. *Frontiers in psychology*. 2019. Vol. 10. Article 813. URL: <https://doi.org/10.3389/fpsyg.2019.00813>.
13. Trafimow D., Marks M. Editorial. *Basic and Applied Social Psychology*. 2015. Vol. 37. P. 1–2. URL: <https://doi.org/10.1080/01973533.2015.1012991>.

14. Valentine J.C., Aloe A.M., Lau T.S. Life After NHST: How to Describe Your Data Without “p-ing” Everywhere. *Basic and Applied Social Psychology*. 2015. Vol. 37. Issue 5. P. 260–273. DOI: 10.1080/01973533.2015.1060240.

### **CALCULATION OF THE SAMPLE SIZE AS THE CORNERSTONE OF PLANNING SCIENTIFIC RESEARCH**

**Valerii Olefir**

*V. N. Karazin Kharkiv National University,  
4, Svobody square, Kharkiv, Ukraine, 61022  
e-mail: vaolefir@gmail.com*

**Valerii Bosniuk**

*National University of Civil Defence of Ukraine,  
94, Chernyshevska str., Kharkiv, Ukraine, 61023  
e-mail: bosnyk@nuczu.edu.ua*

Due to the fact that the scientific research results are extrapolated to the general population one of the main issues when planning work is the formation of a representative optimal sample size. A small sample provokes unreliable results and reduces the chance of detecting the expected effect value but at the same time it is too large because a small sample requires additional resources and helps to identify statistically significant results the practical value of which is quite low. So, justifying the sample size is a critical aspect of the design research. It is necessary in research to attract such a large number of people to be sure that if a practically important effect exists, it is highly likely to be detected by statistical analysis. This approach within the framework of the methodology of evidence-based psychology allows scientists to make statistical decisions in research more rationally and reasonably.

The answer to the question about the required sample size depends on three main factors that should be taken into account when planning the study: 1) the probability error of the first kind on the basis of which the decision is made to accept/reject the null hypothesis; 2) the statistical power of the study as a degree of confidence in obtaining a significant result, if it really exists; 3) the quantity of the expected effect size.

The article proves that it is a common practice of scientists to conduct a priori analysis of statistical power to determine the required sample size in the study. To implement this need it is suggested to use free software data analysis, such as the G\*Power package or the open source R statistical software environment. Some examples of their application in practical situations are given.

Eventually, the result of the analysis showed that during the planning of a psychological study, despite methodological, practical and ethical reasons, scientists regularly neglect such an important aspect of the methodology of evidence-based psychology as calculating the required sample size. In the absence of this calculation, the results of the study should be interpreted with caution.

*Key words:* sample size, statistical power, effect size, error of the first kind, error of the second kind, evidence-based psychology.