ІНСТРУКЦІЯ
підготовки словника змінних
у CodeBook Wizard

1. Призначення

CodeBook – супровідний документ *табличного набору даних*, який пояснює зміст змінних, їхні типи, одиниці вимірювання та коди значень. Він зберігається окремим файлом разом із відповідними даними.

Для його підготовки використовується програма CodeBook Wizard.

Поточна версія працює з CSV та XLSX і не змінює вихідний файл даних. Для моделей, зображень та інших нетабличних матеріалів готується окремий опис структури й параметрів; автоматичне документування таких матеріалів у CodeBook Wizard не передбачено.

2. Підготовка файлу Excel

Дані в таблиці Excel мають бути підготовлені за правилами, які забезпечують їх коректний опис у CodeBook, експорт до програм статистичного аналізу та перетворення в інші формати для подальшого опрацювання.

Назва файлу:

– Назва файлу має бути короткою та зрозумілою.
– Бажано використовувати латиницю.
– Якщо назва складається з кількох слів, їх розділяють символом підкреслення _.
– Не слід використовувати пробіли та спеціальні символи.

Приклад: *knee_oa_data.xlsx.*

Структура таблиці:

– Дані розміщують на одному аркуші Excel.
– Аркуш має містити лише один набір даних у вигляді суцільної таблиці.
– Перший рядок містить унікальні назви змінних, а не їх розгорнуті описи або мітки.
– Кожен стовпець відповідає одній змінній.
– Кожен рядок відповідає одному спостереженню: пацієнту, зразку, вимірюванню, анкетуванню тощо.
– Таблиця не повинна містити додаткових заголовків, порожніх рядків або підсумків усередині набору даних.

Назви змінних

– Назви змінних бажано подавати латиницею, без пробілів і спеціальних символів (`#`, `/`, `()`, `!` тощо).
– Для розділення слів використовувати символ підкреслення _.
– Кожна назва має бути короткою, зрозумілою та унікальною.
– Назва змінної не повинна починатися з цифри.
– Слід уникати надто загальних назв, наприклад *Data1, Var2, Test3.*

Приклади правильних назв: *patient_id, sex, age, VAS_1, IL6_pre.*

Якість даних:

– Усі значення в межах одного стовпця повинні належати до одного типу: число, текст, дата або логічне значення.
– Категоріальні значення в таблиці, як правило, подають у вигляді числових або коротких текстових кодів. Кожна категорія повинна мати окремий унікальний код. Однакові коди необхідно використовувати послідовно в усьому стовпці. Розшифрування кодів і відповідних їм категорій наводять у CodeBook. Коди не повинні змінюватися в межах одного набору даних.

– Формат дат має бути однаковим у всьому стовпці.

Рекомендований формат *YYYY-MM-DD*.

– У таблиці не повинно бути об’єднаних клітинок.

– Формули перед передаванням файлу необхідно перетворити на значення за допомогою команди *«Вставити як значення»*.

– Відсутні значення позначають однаково в усьому наборі даних: `NA` або порожньою клітинкою.

– Для відсутніх значень не слід використовувати `0`, дефіс, крапку або довільний текст, якщо вони не є реальними значеннями змінної.

Однорідність стовпців:

– Таблиця не повинна містити стовпців із коментарями, виносками, поясненнями або службовою інформацією, які не належать до набору даних.

– Кожна змінна, що потребує окремого опису або аналізу, повинна мати власний стовпець.

– В одній клітинці не слід об’єднувати кілька значень, наприклад `120/80`, `5 кг; 10 см` або `так, після операції`. Такі дані потрібно розподіляти між окремими змінними.

– Одиниці вимірювання не включають до значень клітинок. Їх зазначають у CodeBook або, за потреби, у назві змінної.

Приклад: у стовпці `weight` записують `72.5`, а одиницю вимірювання `kg` зазначають у CodeBook.

Зберегти остаточний файл, для якого створюється CodeBook.

CSV має використовувати кому як роздільник.

3. Завантаження та вибір локальної папки

Програма **CodeBook_Wizard** працює у браузері та доступна за посиланням: https://archive.sytenko.org.ua/codebook_wizard/

– Відкрити додаток у браузері.

– Натиснути «Вибрати / оновити папку».

– Вибрати папку в якій розміщені документи.

– За потреби залишити увімкненим «Ураховувати вкладені папки».

– Перевірити кількість доступних файлів.

При виборі папки передаються лише назви файлів XLSX та CSV.

Після перезавантаження програми, додавання, перейменування або переміщення файлів папку потрібно вибрати повторно.

4. Опис файлу даних

У полі **«Load file / Завантажте файл»** натиснути Browse і вибрати підготовлений файл.

Після завантаження перевірити назву файла, кількість змінних і спостережень, а також перелік змінних. Якщо структура прочитана неправильно, файл слід виправити до початку опису.

У полі **«Dataset description / Короткий опис файлу даних»** зазначається, що містить таблиця, яких об’єктів вона стосується та які часові точки або умови представлені.

Приклад: Динамометрія м’язів стегна у пацієнтів до та через 1, 3 і 6 місяців після ендопротезування кульшового суглобу

5. Опис змінних

Для кожної змінної заповнюються поля:

Поле	Зміст
Label / Змінна (розшифровка)	Коротка зрозуміла назва показника.
Type / Тип змінної	Числова, номінальна, порядкова, текстова, дата або ідентифікатор.
Units / Одиниці виміру	Фактична одиниця вимірювання або подання. Якщо не застосовується, поле залишається порожнім.
Description / Опис змінної	Зміст показника, інтерпретація, особливості вимірювання чи обчислення; за потреби – допустимі значення та позначення пропусків.
Identifier / Ідентифікатор"	Унікальний код записи

Для номінальних і порядкових змінних відкривається **«Code List / Коди змінної»**. Навпроти кожного наявного значення вказується його значення словами, наприклад:

```
1 – класичний доступ
2 – експериментальний доступ
```

У полі Description / Опис змінної описується додаткові відомості про змінну, наявність пропусків, особливостей вимірювання, розрахунків, тощо.

Переходи між змінними виконуються стрілками ліворуч і праворуч. При переході поточні записи зберігаються в сеансі. Список змінних є інформаційним.

Індикатор Progress враховує заповнення Label, Type і Description.

6. Збереження CodeBook

Після опису змінних натискається **Export CodeBook**. Формується TXTфайл з назвою на зразок `muscle_force_codebook.txt`.

Файл містить:

- назву вихідного файла та загальний опис;
- кількість змінних і спостережень;
- назву, розшифровку, тип, одиниці та опис кожної змінної;
- пояснення кодів номінальних і порядкових змінних.

Експорт враховує останні введені зміни поточної змінної без додаткового переходу стрілкою. Незаповнені поля не блокують експорт, тому готовий TXT необхідно переглянути перед передачею матеріалів.

CodeBook за замовченням зберігається поруч із відповідними даними або за вибором користувача. Його назва зазначається в картці набору в README Wizard.

7. Чернетка та продовження роботи

При необхідності перервати роботу над описом даних, рекомендується зберегти чернетку для подальшої роботи.

Збереження: натиснути **«Зберегти чернетку JSON»** і зберегти файл на зразок `muscle_force_codebook_draft.json` у робочій папці.

Чернетка містить загальний опис, описи змінних, типи, одиниці, значення категорій і їхні мітки, поточну позицію та відомості для перевірки джерела.

Відновлення:

- Завантажити той самий незмінений CSV/XLSX.
- Вибрати відповідну JSON-чернетку.
- Натиснути **«Відновити чернетку»** та підтвердити заміну поточних описів.
- Перевірити відновлені поля й продовжити роботу.

Відповідність файла перевіряється за контрольною сумою та структурою. Навіть повторне збереження XLSX без видимих змін може змінити внутрішній вміст файла; для відновлення слід зберігати початковий файл незмінним.

> Автозбереження відсутнє. Перед закриттям додатка або завантаженням іншого набору потрібну роботу зберігають у JSON. TXT не використовується для відновлення форми.

8. Контроль результату

Перед передачею CodeBook перевіряються:

- відповідність конкретному файлу даних;
- наявність описів усіх змінних;
- правильність типів та одиниць;
- пояснення кодів категорій і пропущених значень;
- коректність українських символів у TXT;
- відповідність назви CodeBook посиланню в README.

Чернетка JSON є робочим матеріалом і не потребує включення до опублікованого набору, але його краще зберегти до затвердження матеріалів і наборів даних до публікації