

Elementy niezbędne do złożenia danych w



Dokumenty

1. Podpisany przez właściciela praw autorskich do składanego zbioru **formularz zgody na jego rozpowszechnianie**. Może być w postaci skanu.
Formularz można pobrać stąd: [PADSFORM02.docx](#).
2. **Dokumentacja badania** wraz z opisem zbioru lub zbiorów danych.
Szablon dokumentu można pobrać stąd: [Opis badania.docx](#).
3. **Narzędzia badania** (kwestionariusz w przypadku sondaży, wywiadów itp.) lub opis narzędzia (w przypadku danych pozyskanych w inny sposób).

Dokumentacja badania

Dokumentacja badania (książka kodów, codebook) musi zawierać następujące informacje:

1. Dane kierownika projektu. Imię nazwisko, tytuł, afiliacja w czasie wykonywania badań oraz informacje na temat instytucji realizującej badanie.
2. Oficjalny tytuł badań.
3. Wzór cytowania danych.
4. Źródła finansowania, numer grantu oraz opis innych, związanych z badaniem, sposobów finansowania.
5. Skład osobowy zespołu przetwarzającego dane.
6. Opis projektu, w tym:
 - o opis celu badań,
 - o krótka historia projektu z opisem ważniejszych problemów i decyzji podejmowanych w trakcie jego realizacji.
7. Informacje na temat próby i procedury jej doboru. Sekcja ta powinna opisywać badaną populację i metodę użytą do wyłonienia z niej próby. Powinna tu także znaleźć się odpowiedź na pytanie, czy stosowanie – przy tak wylosowanej próbie – standardowych procedur wyliczania błędów estymatorów jest możliwe. Jeśli nie, należy zamieścić procedury szacowania błędów estymatorów dostosowane do danego modelu losowania próby. Jeśli zbiór danych zawiera wagę, musi być szczegółowo opisana procedura jej tworzenia i sposób użycia. Niezbędną informacją w tej sekcji jest dokładne określenie liczebności próby założonej (wylosowanej) oraz liczebności próby zrealizowanej. Informacje te są niezbędne do poprawnego stosowania procedur szacowania błędów estymatorów. Nie mniej ważne jest dokładne określenie okresu, w którym przeprowadzane było badanie terenowe.
8. Obszar geograficzny, na którym badanie zostało zrealizowane oraz termin jego realizacji.
9. Źródła danych, tj. informacje na temat technik użytych do zbierania danych.
10. Informacje na temat tego, co było jednostką badania.

11. Nieważone rozkłady liczebności dla wszystkich zmiennych ze wszystkich przynależących do deponowanego badania zbiorów danych. Rozkłady muszą zawierać niezbędne informacje:

- **Nazwa zbioru danych** wraz z czytelną dla człowieka i jednoznaczną **etykietą** (w przypadku badań zawierających więcej niż jeden zbiór danych).
- **Nazwa zmiennej** wraz z czytelną dla człowieka i jednoznaczną **etykietą**.
- Pełne zestawienie **kodów wartości** oraz kodów **specjalnych** dla każdej zmiennej niemierzonej na skali ciągłej (nominalnej, porządkowej itp.) wraz z ich czytelną dla człowieka i jednoznaczną **etykietą**.
- Jeśli zmienna mierzona jest na skali ciągłej, należy podać nazwę jednostki, w jakiej została wyrażona (np. rok w przypadku wieku) oraz czytelne dla człowieka i jednoznaczne **etykiety** kodów **specjalnych**.
- **Wszelkie kody specjalne** takie jak "nie dotyczy", "brak danych" itp. muszą posiadać czytelna dla człowieka i jednoznaczną **etykietę**.
W zbiorze danych **nie mogą występować systemowe braki danych** (puste komórki). Jeśli w zbiorze są braki danych, należy nadać im unikalny kod i dołączyć jego opis.
- Jeśli zmienna nie jest zmienną pierwotną, czyli powstała w wyniku obliczeń na zmiennych pierwotnych, lub jeśli zmienna została poddana jakiegokolwiek innej edycji, należy zamieścić dokładny opis użytych procedur.
- Struktura, układ oraz kolejność zmiennych w książce kodów i dokumentacji powinna odpowiadać kolejności zmiennych w zbiorze danych (kolejnych zbiorach danych). W przypadku jakichkolwiek rozbieżności należy załączyć opis niezgodności i podać jej przyczynę.

12. Nieważone statystyki opisowe dla wszystkich zmiennych ze wszystkich przynależących do deponowanego badania zbiorów danych. Przez statystyki opisowe należy tu rozumieć zestawienie liczebności, minimum, maksimum, średnią oraz odchylenie standardowe wyliczone dla każdej zmiennej występującej w zbiorze wraz z nazwą, pod jaką występuje i jej etykietą.

Narzędzia badania

Narzędzia badania (kwestionariusze, listy zapowiednie, metryki, karty pytań, karty gospodarstw domowych itp. w przypadku sondaży, wywiadów itp.) lub opis narzędzia (w przypadku danych pozyskanych w inny sposób) można dołączyć bezpośrednio do pliku dokumentacji badania lub w osobnych plikach.

Formaty zbiorów danych

Preferowane są formaty otwarte. W przypadku formatów zamkniętych preferowane są należące do najpopularniejszych pakietów statystycznych i arkuszy kalkulacyjnych.

1. Formaty otwarte:

1. Format **ASCII** (fixed width) wraz ze słownikiem definiującym (data definition statements), zawierającym informacje na temat szerokości kolumn dla każdej zmiennej, ich etykiety, etykiety wartości zmiennych i definicje kodów specjalnych. Przez format ASCII należy tu rozumieć dane w postaci tekstowej bez separacji kolumn.

2. Format tekstowe z separacją kolumn (**CSV, TSV**) wraz ze słownikiem definiującym (data definition statements), zawierającym etykiety każdej zmiennej, jej wartości oraz definicje kodów specjalnych.
 3. Format **R** wraz ze słownikiem definiującym (data definition statements), zawierającym etykiety każdej zmiennej, jej wartości oraz definicje kodów specjalnych.
2. Formaty zamknięte (własnościowe):
1. SPSS (SAV, POR),
 2. STATA (DTA),
 3. SAS (sas7bdat + sas7bcat, XPT wersja 5 i 8),
 4. STATISTICA (STA),
 5. Excell (XLS, XLSX) wraz ze słownikiem definiującym (data definition statements), zawierającym etykiety każdej zmiennej, jej wartości oraz definicje kodów specjalnych.