

Jak przygotować dane do analizy w Jamovi

Paweł Kleka

2026-03-15

Spis treści

1 Krok 1 — Import i sprawdzenie struktury danych	2
2 Krok 2 — Sprawdzenie zakresów wartości	3
2.1 Zmienne nominalne — duplikaty kategorii	3
2.2 Braki danych	4
3 Krok 3 — Przekodowanie itemów odwróconych	5
3.1 Co to jest item odwrócony?	5
4 Krok 4 — Obliczenie wyników skalowych	6
5 Krok 5 — Attention check: wykluczenie nieuważnych uczestników	7
5.1 Co to jest attention check?	7
6 Krok 6 — Wstępna eksploracja zmiennych skalowych	8
7 Krok 7 — Sprawdzenie normalności rozkładu	9
8 Podsumowanie: co zrobiliśmy	9

Zanim zaczniesz testować hipotezy, upewnij się, że dane są czyste i gotowe do analizy.

Zbieranie danych to dopiero połowa pracy. Pisałem już o [schludnych danych teoretycznie](#), natomiast ten wpis ma pokazać jak takie dane przygotować w Jamovi. Zanim uruchomisz pierwszy test statystyczny, musisz sprawdzić, czy dane nie zawierają błędów, przekodować itemy odwrócone, obliczyć wyniki skalowe i upewnić się, że osoby badane faktycznie uważnie wypełniały kwestionariusz. Ten wpis przeprowadzi Cię przez cały proces — od importu surowego pliku do momentu, w którym dane będą gotowe do analiz. Popracujemy na [przykładowym](#)

zbiornie danych o pracownikach ($N = 200$) z czterema skalami: dobrostanu, osamotnienia, wielowymiarowego lęku technologicznego i wsparcia organizacji.

1 Krok 1 — Import i sprawdzenie struktury danych

Otwórz plik ze swoim zbiorem danych:

☒ → **Otwórz** → **Ten komputer** → **wskaż plik**

Gdy plik się załaduje, pierwsze co robisz to **sprawdzasz typy zmiennych**. Przy nazwie każdej kolumny widać małą ikonę — to informacja o tym, jak Jamovi traktuje daną zmienną. Upewnij się, że typy odpowiadają treści zmiennych:

Zmienna	Powinien być typ	Dlaczego?
wiek	Ciągła	Wartość liczbowa, ma sens matematyczny
plec	Nominalna	Kategorie bez porządku
rodzaj_pracy	Nominalna	Kategorie: Zdalna, Stacjonarna, Hybrydowa
dobr_1 ... wsp_org_4	Ciągła	Itemy Likerta — chcemy liczyć średnie
ac_1, ac_2	Porządkowa	Będziemy porównywać z konkretną liczbą

Aby zmienić typ: kliknij dwukrotnie nagłówek kolumny i wybierz właściwy typ z listy rozwijanej.

Na koniec policz obserwacje — w prawym dolnym rogu ekranu Jamovi wyświetla liczbę wierszy. Sprawdź: czy zgadza się z oczekiwanym $N = 200$?

☒ **Sprawdź:** Czy masz dokładnie 200 wierszy? Jeśli mniej — może plik się nie załadował w całości. Jeśli więcej — może zawiera duplikaty lub wiersz z nagłówkami został wczytany jako dane.

2 Krok 2 — Sprawdzenie zakresów wartości

Każda zmienna ma swój dopuszczalny zakres wartości. Itemy na skalach Likerta 1–5 nie powinny zawierać wartości 0 czy 6. Wiek nie powinien wynosić 3 ani 200. Jak to sprawdzić?

Eksploracja → Statystyki opisowe

Wstaw zmienną, którą chcesz sprawdzić, i w statystykach zaznacz opcje: **Rozstęp**, **Min**, **Max**.

Przykład: sprawdzasz zmienną `lekt_1` (item skali lęku technologicznego, Likert 1–5). W wynikach widzisz:

Statystyka	Wartość
Min	0
Max	6
Rozstęp	6

Minimum powinno wynosić 1, a maximum 5. Wartości 0 i 6 to błędy — ktoś mógł źle zakodować odpowiedź albo wpisać coś przypadkiem. Takie wartości trzeba usunąć lub oznaczyć jako brakujące.

Jak naprawić? Kliknij dwukrotnie nagłówek zmiennej `lekt_1`, przejdź do sekcji **Braki danych** i wpisz warunki wykluczenia, np.:

```
== 0  
>= 6
```

Jamovi zamieni te wartości na braki danych (nie usunie całego wiersza — jedynie ta konkretna komórka zostanie potraktowana jako brak).

☒ **Wskazówka:** Zrób to dla wszystkich itemów Likertowych. W naszym zbiorze itemy tak jak `dobr_1`–`dobr_5` powinny mieć wartości wyłącznie z zakresu 1–5. Sprawdź też `wiek` — czy nie ma wartości ujemnych albo skrajnie wysokich?

2.1 Zmienne nominalne — duplikaty kategorii

Zmienne nominalne mogą wyglądać czysto, a mimo to zawierać „duplikaty” — tę samą kategorię zapisaną różnymi sposobami. Najczęstszym problemem jest wielkość liter lub zbędna spacja.

Eksploracja → Statystyki opisowe → wstaw rodzaj_pracy → zaznacz Tabele częstości

Możesz zobaczyć coś takiego:

Kategoria	Częstość
Stacjonarna	24
Zdalna	25
hybrydowa	88
stacjonarna	28
zdalna	35

Zamiast 3 kategorii masz 5. Jamovi traktuje „Stacjonarna” i „stacjonarna” jako dwie różne wartości — bo różnią się wielkością pierwszej litery.

Jak naprawić? Nie zmieniaj danych źródłowych. Zamiast tego utwórz nową, przekształconą zmienną. Kliknij nagłówek pustej kolumny obok `rodzaj_pracy` i wstaw nową zmienną jako zmienną wyliczoną (Computed variable). Użyj formuły:

```
IF($source == 'Stacjonarna', 'stacjonarna',  
  IF($source == 'Zdalna', 'zdalna',  
    $source))
```

Formuła działa tak: jeśli oryginalna wartość to „Stacjonarna” (z wielkiej), zamień na „stacjonarna” (z małej). Jeśli to „Zdalna” — zamień na „zdalna”. W każdym innym przypadku zostaw wartość bez zmian (bo „hybrydowa” już jest poprawnie).

Alternatywnie — kliknij w nagłówek następnej wolnej kolumny, wstaw pośrodku nową zmienną jako Przekształcenie (Transform) i użyj takiej samej logiki. Dzięki temu dane źródłowe pozostają nienaruszone i możesz w każdej chwili zmienić formułę. Dotyczy to obu sposobów i nie ma różnicy w ich działaniu. Po prostu ten drugi sposób zapamiętuje formułę jako przekształcenie, które można użyć dla wielu zmiennych, co jest przydatne np. przy rekodowaniu.

☒ **Sprawdź:** Po utworzeniu przekształconej zmiennej zrób ponownie Tabele częstości — czy teraz widzisz dokładnie 3 kategorie: stacjonarna, hybrydowa, zdalna?

2.2 Braki danych

Eksploracja → Statystyki opisowe → wstaw wszystkie zmienne → zaznacz: Braki danych

Jamovi wyświetli liczbę brakujących wartości (missing) dla każdej zmiennej. Zwróć uwagę na zmienne z wyjątkowo dużą liczbą braków — czy to przypadek, czy może osoby z określonej grupy (np. pracujący zdalnie) częściej pomijały te pytania?

Możesz to sprawdzić: dodaj zmienną grupującą (np. `rodzaj_pracy`) do opcji **Podziel według** i porównaj liczbę braków w każdej grupie.

☒ **Sprawdź:** Które zmienne mają najwięcej braków? Czy braki rozkładają się losowo, czy koncentrują się w konkretnej grupie?

3 Krok 3 — Przekodowanie itemów odwróconych

3.1 Co to jest item odwrócony?

To pytanie (item), które mierzy ten sam konstrukt co pozostałe itemy skali, ale jest sformułowane „na odwrót”. Na przykład w skali dobrostanu większość pytań brzmi w stylu „Czuję się dobrze”, a item odwrócony brzmi „Czuję się źle”. Osoba z wysokim dobrostanem odpowie *wysoko* na normalne itemy, ale *nisko* na ten odwrócony. Zanim policzysz wynik skali, musisz odwrócić punktację takich itemów, żeby wszystkie „mówiły w tym samym kierunku”.

W naszym zbiorze mamy na przykład następujące itemy odwrócone: `dobr_2`, `dobr_4` oraz `osam_3`, `osam_5`.

Jak odwrócić? Kliknij nagłówek kolumny (np. `dobr_2`) → wybierz **Przekształć**. Jamovi utworzy nową zmienną, nadaj jej nazwę (np. `dobr_2_r`). Jako formułę wpisz:

`6 - $source`

Dlaczego 6? Bo skala ma zakres 1–5, więc 6 – wartość odwraca punktację: 1→5, 2→4, 3→3, 4→2, 5→1. Ogólna zasada: **odwrócona = (minimum + maximum) – wartość**, a dla skali 1–5 to właśnie 6 – wartość.

Powtórz to samo dla pozostałych itemów odwróconych:

Tabela 4: W zbiorze danych jest więcej odwróconych pozycji.

Item oryginalny	Nowa zmienna	Formuła
<code>dobr_2</code>	<code>dobr_2_r</code>	<code>6 - \$source</code>
<code>dobr_4</code>	<code>dobr_4_r</code>	<code>6 - \$source</code>
<code>osam_3</code>	<code>osam_3_r</code>	<code>6 - \$source</code>
<code>osam_5</code>	<code>osam_5_r</code>	<code>6 - \$source</code>

¶ **Wskazówka:** Gdy raz zdefiniujesz transformację 6 - \$source, możesz jej użyć ponownie — kliknij nagłówek kolejnego itemu odwróconego → Przekształcić → wybierz istniejącą transformację z listy. Nie musisz pisać formuły od nowa.

¶ **Jak rozpoznać item odwrócony?** Jeśli masz klucz kodowania — sprawa jest prosta, bo jest tam wprost napisane. Ale jeśli pracujesz z własnymi itemami albo nie masz dokumentacji, sprawdź korelację danego itemu z pozostałymi itemami skali. Jeśli korelacja jest ujemna (podczas gdy inne itemy korelują ze sobą dodatnio), prawdopodobnie jest to item odwrócony.

4 Krok 4 — Obliczenie wyników skalowych

Teraz tworzymy zmienne reprezentujące wynik na całej skali — jako średnią lub sumę poszczególnych itemów. Kliknij ikonę **f(x)** u góry **pustej** kolumny, aby utworzyć **Wyliczoną zmienną** (Computed variable).

Nowa zmienna	Metoda	Formuła
DOBROSTAN	Średnia	MEAN(dobr_1, dobr_2_r, dobr_3, dobr_4_r, dobr_5, ignore_missing=1)
OSAMOTNIENIE	Suma	SUM(osam_1, osam_2, osam_3_r, osam_4, osam_5_r, ignore_missing=1)
WSPARCIE	Średnia	MAX(wsp_org_1, wsp_org_2, wsp_org_3, wsp_org_4)

¶ **Uwaga:** Zwróć uwagę, że w formule na DOBROSTAN używamy `dobr_2_r` i `dobr_4_r` — czyli nazw *przekodowanych* zmiennych z etapu 3a, a nie oryginalnych `dobr_2` i `dobr_4`. To samo dotyczy OSAMOTNIENIA: `osam_3_r` i `osam_5_r`, gdzie używamy sumy. Inna funkcja to `max`, która zwraca największy wynik w zbiorze. Inne przydatne funkcje są opisane w Jamovi pod przyciskiem **f(x)**.

Jeśli nie chcesz tworzyć osobnych zmiennych przekodowanych, możesz umieścić odwrócenie bezpośrednio w formule. Np. suma osamotnienia bez osobnych przekodowań wyglądałaby tak:

```
SUM(osam_1, osam_2, 6 - osam_3, osam_4, 6 - osam_5, ignore_missing=1)
```

Oba podejścia dają identyczny wynik — wybierz to, które jest dla Ciebie czytelniejsze.

Parametr `ignore_missing=1` oznacza, że jeśli jedna z odpowiedzi jest brakująca, Jamovi obliczy wynik na podstawie pozostałych. Bez tego parametru wystarczy jeden brak w dowolnym itemie, by cały wynik skali był pusty.

☐ **Sprawdź:** Uruchom Statystyki opisowe (Eksploracja → Statystyki opisowe) na nowych zmiennych skalowych. Czy średnie wyglądają sensownie? Dla skali Likert 1–5 liczonej jako średnia spodziewaj się wartości gdzieś w okolicy środka skali (tj. 2–4). Dla sumy 5 itemów Likerta 1–5 spodziewaj się zakresu 5–25, z typową średnią w pobliżu 15.

5 Krok 5 — Attention check: wykluczenie nieuważnych uczestników

5.1 Co to jest attention check?

To pytanie kontrolne ukryte w kwestionariuszu, które sprawdza, czy osoba badana faktycznie czyta treść pytań, a nie klika losowo. Typowy attention check brzmi np. „Proszę zaznaczyć odpowiedź jeden” — jeśli ktoś zaznaczy cokolwiek innego, prawdopodobnie nie czytał pytań. Stosujemy go, bo nieuważne odpowiedzi wprowadzają szum do danych i mogą zniekształcić wyniki analiz. Wykluczenie takich obserwacji zwiększa jakość zbioru.

W naszym zbiorze mamy dwa pytania kontrolne: - `ac_1` — poprawna odpowiedź: 1 - `ac_2` — poprawna odpowiedź: 7

Aby odfiltrować osoby, które nie przeszły attention check:

Dane → Filtry

Wpisz formułę:

```
IF(ac_1 == 1) and IF(ac_2 == 7)
```

Jamovi wyszarzy (ale nie usunie!) wiersze, które nie spełniają tego warunku. Wszystkie analizy będą teraz wykonywane tylko na przefiltrowanej próbie.

☐ **Sprawdź:** Ile osób zostało wykluczonych? Spójrz na liczbę aktywnych wierszy w prawym dolnym rogu. Jeśli np. z 200 pozostało 187, to 13 osób nie przeszło attention checku. Czy to dużo? Przyjmuje się, że do ok. 10–15% wykluczeń to norma. Jeśli odpada więcej — zastanów się, czy pytanie kontrolne nie było zbyt trudne albo niejednoznaczne.

☒ **Wskazówka:** Możesz w każdej chwili wyłączyć filtr (odznaczając go w menu Dane → Filtry), żeby wrócić do pełnej próby. Warto przeprowadzić analizy w obu wariantach i porównać — jeśli wyniki się znacząco różnią, to sygnał, że nieuważne odpowiedzi miały wpływ na dane.

6 Krok 6 — Wstępna eksploracja zmiennych skalowych

Dane są czyste, skale obliczone, nieuważni uczestnicy odfiltrowani — czas przyjrzeć się zmiennym, zanim przejdziemy do właściwych analiz.

Eksploracja → Statystyki opisowe

Wstaw zmienne skalowe (DOBROSTAN, OSAMOTNIENIE, LEKT, WSPARCIE) i zaznacz: **Histogramy** oraz **Wykresy skrzynkowe**.

Histogramy pokażą Ci kształt rozkładu każdej zmiennej, a wykresy skrzynkowe — medianę, rozrzut i ewentualne wartości odstające (outliers).

Oto przykładowe wyniki i ich interpretacja:

Zmienna	Śred- nia	SD	Ocena
DOBROSTAN	3.1	1.2	☒ OK — średnia blisko środka skali, rozsądny rozrzut (średnia, skala 1–5)
LEK_T	1.1	0.2	☒ Podejrzane — prawie wszyscy odpowiadali „1”. SD bliskie zeru oznacza niemal zerowe zróżnicowanie. Sprawdź, czy to nie efekt błędu kodowania, np. wartości spoza zakresu zamienione na braki mogły wyzerować zmienną, albo czy skala nie ma efektu podłogi
OSAMOTNIENIE	18.4	4.1	☒ OK — nieco powyżej środka (15), ale realistycznie (suma, skala 5–25)

☒ **Sprawdź:** Czy Min i Max Twoich zmiennych skalowych mieszczą się w oczekiwanym zakresie? Czy na histogramach nie ma dziwnych „kolców” przy skrajnych wartościach? Które zmienne mają wyraźną skośność (wartości odchylenia od zera powyżej ± 1)?

7 Krok 7 — Sprawdzenie normalności rozkładu

Dlaczego to ważne? Wiele testów statystycznych (np. test t, ANOVA, korelacja Pearsona) zakłada, że dane mają rozkład zbliżony do normalnego. Jeśli rozkład zmiennej jest silnie skośny lub ma ciężkie ogony, wyniki tych testów mogą być niewiarygodne. W takiej sytuacji lepiej sięgnąć po testy nieparametryczne (np. test Manna-Whitneya zamiast testu t, korelację Spearmana zamiast Pearsona), które nie wymagają założenia o normalności.

Eksploracja → Statystyki opisowe → wstaw zmienne skalowe

Zaznacz: **Test Shapiro-Wilka, Skośność, Kurtoza, K-K plot (Q-Q plot).**

Test Shapiro-Wilka sprawdza, czy rozkład zmiennej istotnie odbiega od rozkładu normalnego. Interpretacja: - $p > 0.05 \rightarrow$ brak podstaw do odrzucenia normalności - $p \leq 0.05 \rightarrow$ rozkład istotnie odbiega od normalnego

K-K plot to wykres, na którym punkty (Twoje dane) powinny układać się mniej więcej wzdłuż linii prostej. Im bardziej punkty odchylają się od tej linii, tym bardziej rozkład odbiega od normalnego. To dobre uzupełnienie testu Shapiro-Wilka, bo daje Ci wizualny obraz — widzisz *gdzie i jak* rozkład odbiega (np. ciężkie ogony, skośność).

⚠ **Uwaga:** Test Shapiro-Wilka jest bardzo czuły przy dużych próbach — przy $N = 200$ może pokazać istotny wynik nawet przy niewielkim odchyleniu od normalności. Dlatego nie polegaj wyłącznie na wartości p. Zawsze sprawdź też histogram, K-K plot oraz wartości skośności i kurtozy. Jeśli skośność mieści się w przedziale ± 1 , a K-K plot wygląda w miarę prosto, rozkład jest zwykle „wystarczająco normalny” dla testów parametrycznych.

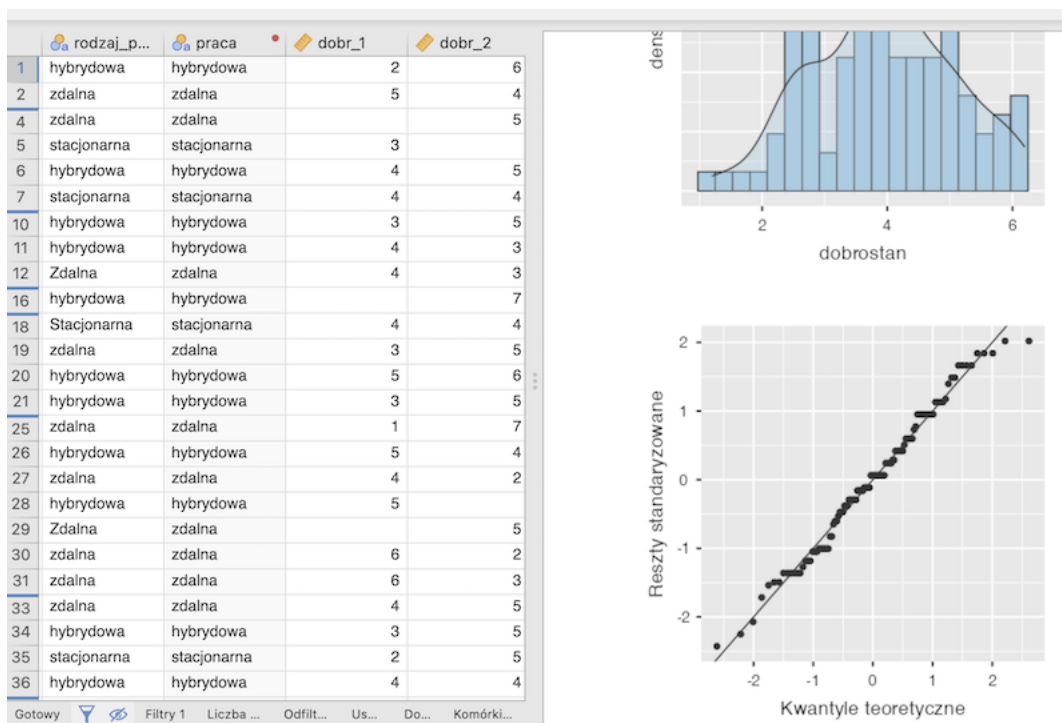
⚠ **Sprawdź:** Dla których zmiennych test Shapiro-Wilka daje $p > 0.05$? Jak wyglądają ich K-K ploty? Czy zmienne z niskim p (np. OSAMOTNIENIE) mają widoczną skośność na histogramie?

8 Podsumowanie: co zrobiliśmy

Krok	Czynność
1	Import danych, sprawdzenie typów zmiennych i liczby obserwacji
2	Weryfikacja zakresów wartości, naprawa błędów kodowania, braki danych

Krok	Czynność
3	Przekodowanie itemów odwróconych (6 - \$source)
4	Obliczenie wyników skalowych (średnie i sumy)
5	Filtrowanie nieuważnych uczestników (attention check)
6	Wstępna eksploracja rozkładów (histogramy, wykresy skrzynkowe)
7	Sprawdzenie normalności (Shapiro-Wilk, K-K plot)

Dopiero teraz dane są gotowe do testowania hipotez. Cały ten proces może wydawać się żmudny, ale to fundament wiarygodnych wyników — jeśli dane na wejściu są zaszumione, żaden test statystyczny tego nie naprawi.



Rysunek 1: Podgląd na dane i efekty pracy nad ich przygotowaniem

Na koniec — **zapisz plik w formacie .omv** (☒ → Zapisz jako). Format .omv to natywny format Jamovi, który zachowuje nie tylko dane, ale też wszystkie Twoje przekształcenia, filtry,

wyliczone zmienne i wyniki analiz. Dzięki temu gdy wrócisz do projektu za tydzień, wszystko będzie dokładnie tam, gdzie zostało. Dalszą analizę warto rozpocząć od utworzenia kopii tego pliku.