

Zadania z kursu języka R

Wojciech Sobczuk

Wszystkie zadania będą bazowały na tych samych danych, z plików `dane_badane.csv` oraz `dane_kontrolne.csv`. Dane pochodzą z ankiet dotyczących reakcji dzieci na COVID-19. Przygotowaliśmy dwie osobne, choć podobne ankiety, dla rodziców dzieci z ASD (grupa badana) i dla dzieci neurotypowych (grupa kontrolna).

Zad. 1

Załaduj pliki `dane_badane.csv` oraz `dane_kontrolne.csv`. Ze zbioru danych wybierz zmienne opisujące wiek dziecka, płeć dziecka, typ szkoły do której uczęszcza dziecko oraz poziom lęku dziecka w poszczególnych okresach. Dokonaj pobieżnej analizy eksploracyjnej - upewnij się, że wszystkie zmienne mają odpowiedni typ (tzn. numeryczny, kategoriowy lub tekstowy), jeśli nie mają to je zmień (np. za pomocą funkcji `as.factor`). Na koniec, ogranicz zbiory danych względem wieku do przedziału $[4;19]$ tak, aby powstał nowy `data.frame` na którym będziesz pracował/a przez resztę zadań. Dodatkowy punkt za użycie biblioteki `dplyr`.

Uwaga Użyj funkcji `read.csv` z parametrem `encoding = "UTF-8"`, nazwa zmiennej dla płci do przekopowania - `X.U.FEFFF.Plec.dziecka`

Zad. 2

Narysuj histogram rozkładu wieku dziecka osobno dla grupy badanej oraz dla grupy kontrolnej. Oblicz i wyświetl podstawowe miary statystyczne dla wieku (średnia, odch. std., min, max). Sprawdź testem statystycznym czy średnie wieku dziecka są równe między grupami. (dodatkowy punkt za użycie biblioteki `ggplot2`)

Zad. 3

a Stwórz tabele zliczające liczbę chłopców i dziewczynek w obu grupach osobno, oblicz i wyświetl odsetek chłopców oraz sprawdź testem statystycznym (t-Studenta), czy te odsetki są równe między grupami. (pomiń sprawdzanie założeń testu statystycznego)

b Stwórz tabelę zliczającą liczbę dzieci uczęszczających do szkół publicznych i niepublicznych w obu grupach. Wylicz odsetek dzieci uczęszczających do szkół publicznych i sprawdź testem statystycznym (t-Studenta), czy te odsetki są równe między grupami. (pomiń sprawdzanie założeń testu statystycznego)

Zad. 4

a Stwórz `data.frame` (jeden dla obu grup), w którym będą informacje o poziomie lęku w każdym z okresów, grupie z której pochodzi dziecko. Na jego podstawie oblicz średni poziom lęku w każdym z okresów w obu grupach.

b Zwizualizuj średnie z poprzedniego podpunktu na wykresie słupkowym łączonym (tj. dla okresu I ma być słupek dla grupy badawczej i kontrolnej obok siebie, w różnych kolorach, powtórz dla każdego okresu).

c Sprawdź testem statystycznym (t-Studenta) czy te poziomy są sobie równe między grupami dla każdego z okresów osobno. (pomiń sprawdzanie założeń testu statystycznego)

Zad. 5

Do `data.frame` stworzonego w poprzednim zadaniu dodaj zmienne, które będą odpowiadały zmianie poziomu lęku między okresami 1-3, 2-3 i 1-2. Zmianę poziomu lęku nazwiemy dalej dynamikami lęku (tj. zmiana 1-3 będzie dynamiką 1-3) Wybierz jedną z dynamik i narysuj histogram rozkładu wartości dynamiki dla obu

grup na dwóch wykresach koło siebie (tzn. wykres dla grupy badanej po jednej stronie, wykres dla grupy kontrolnej po drugiej stronie).