

Projekt RPiSM - wymagania

dr inż. W. Sałabun

1 Dane

1.1 Źródła:

- <https://stat.gov.pl/>
- <https://bdl.stat.gov.pl/>
- <https://dane.gov.pl/>
- <https://www.kaggle.com/>
- <https://archive.ics.uci.edu/ml/datasets.php>
- i inne.

1.2 Wymagania

- Wybieramy jedną zbiorowość statystyczną (np. osoby poniżej 20 roku życia).
- Jednostka statystyczna (np. osoba poniżej 20 roku życia) musi być opisana conajmniej za pomocą dwóch cech statystycznych (np. wysokość, waga, liczba dni absencji w pracy).
- Wielkość próby: $N \geq 2000$
- **Dane przedstawiamy jako szereg szczegółowy oraz odpowiedni szereg rozdzielczy**
- Obliczenia można wykonać:
 - na kartce - ręcznie ;)
 - arkusz kalkulacyjny
 - Python 3 (ale funkcje obliczające miary statystyczne mają być napisane przez studenta)

2 Realizacja projektu

Realizacja polega na dokonaniu wyliczeń i ich interpretacji dla wszystkich cech w wybranym problemie. W dokumencie zamieszczamy:

- Krótki opis wybranych danych oraz ich źródło (około pół strony). Definicje zmiennych x_i , y_i , itp.
- Miary statystyki opisowej:
 - średnia arytmetyczna
 - odchylenie standardowe
 - dominanta
 - mediana
 - kwartył dolny
 - kwartył górny
 - odchylenie ćwiartkowe

- klasyczne i pozycyjne *xtyp*
- klasyczne i pozycyjne wsp. zmienności
- asymetria A_1 , A_2 oraz A_3
- koelacje pearsona pomiędzy wszystkimi parami cech statystycznych
- estymacje przedziałowe (z parametrem $\alpha = 0.025, 0.05$ oraz 0.1) dla wszystkich cech dla:
 - średnia
 - odchylenie standardowe
 - korelacja (dla wszystkich par)
- testy statystyczne dla 1 zmiennej (z parametrem $\alpha = 0.025, 0.05$ oraz 0.1) dla wszystkich cech dla:
 - średniej (wartość testowana dobrana przez studenta)
 - odchylenie standardowe (wartość testowana dobrana przez studenta)
 - korelacja (dla wszystkich par oraz wartość testowana dobrana przez studenta)
- testy statystyczne dla par zmiennych (z parametrem $\alpha = 0.025, 0.05$ oraz 0.1) dla:
 - średniej (wartość testowana dobrana przez studenta)
 - odchylenie standardowe (wartość testowana dobrana przez studenta)
 - korelacja (dla wszystkich par oraz wartość testowana dobrana przez studenta)

3 Wyniki

- Jako wyniki wysyłamy 3 pliki (BEZ KOMPRESJI):
 - Raport w **formacie pdf** z krótkim opisem danych; wynikami i interpretacjami
 - Obliczenia (jako skan, plik arkusza kalkulacyjnego lub skrypt Python 3)
 - Dane w dowolnym formacie (1 plik!!!)
- **PROJEKTY KTÓRE NIE SPEŁNIAJĄ POWYŻSZEGO WYMAGANIA ZOSTANĄ ZWRÓCONE Z OCENĄ 2.0**
- Data wysłania to data wskazana na Moodle
- Praca oceniana w czasie do 7 dni roboczych
- waga w ocenie końcowej to 0,4
- projekty oddane po terminie będą sprawdzane z zastrzeżeniem, iż ocena będzie obniżona a czas oceniania może przekraczać 7 dni roboczych
- **PRACE PRZESYŁANE W INNY SPOSÓB NIŻ NA MOODLE NIE BĘDĄ OCENIANE**

W razie wątpliwości proszę pytać na zajęciach lub w szczególnych okolicznościach pisać na adres:

- wojciech.salabun@wsb.szczecin.pl