# PAD 07

**Praca Domowa 7 – Testy statystyczne i modelowanie statystyczne**

1. Przed rozpoczęciem zainstaluj biblioteki scipy i statmodels poleceniem:

*pip install NAZWA*  lub *conda install NAZWA*

**Zadanie 1**

Przeprowadź test T studenta na danych w pliku wyniki.csv

Hipoteza zerowa brzmi: nie ma istotnej różnicy w Ocenie końcowej pomiędzy uczniami i uczennicami.

Hipoteza alternatywna brzmi: istnieje istotna różnica (bez wskazania kierunku) w Ocenie końcowej pomiędzy uczniami i uczennicami.

Przyjmij significance level (alpha) na poziomie 0.05. Czy uda się odrzucić hipotezę zerową?

Pomocne wideo: <https://www.khanacademy.org/math/ap-statistics/xfb5d8e68:inference-quantitative-means/two-sample-t-test-means/v/two-sample-t-test-for-difference-of-means> - przeprowadzanie testu dla dwóch zmiennych.

**Zadanie 2**

Korzystając z danych w pliku ZyskiFirmyX.csv stwórz model regresji liniowej i biblioteki statmodels:

* Wyświetl wykres pokazujący zyski w czasie.
* Czy da się zaobserwować trend? Jaki?
* Wyświetl podsumowanie modelu.

**Zadanie 3**

Korzystając z danych w pliku spendingscores.csv i biblioteki statmodels:

1. stwórz model regresji liniowej, gdzie zmienną zależną jest Spending score, a zmiennymi niezależnymi pozostałe zmienne.
2. Z modelu wyświetl:
   1. p values
   2. standard err
   3. coefficients
3. Sprawdź czy istnieje korelacja pomiędzy zmiennymi. Pokaż to na wykresie, np. z biblioteki plotly express.
4. Korzystając z eliminacji wstecznej usuń najmniej istotną zmienną i jeszcze raz wyświetl model.

**UWAGA: zwróć uwagę na zmienne będące kategoriami**!