Lista 1

1. Czterech pasażerów Janek, Franek, Karolina i Agnieszka postanowili polecieć z Warszawy do Dubaju tym samym samolotem, lecz kupili bilety w różnych odstępach czasowych:

Table 1: Cena biletu Warszawa–Dubaj (WAW-DXB) w zależności od momentu zakupu

Pasażer	**Dni przed wylotem (WAW-DXB)**	**Cena biletu (w tys. PLN)**
Karolina	60	1.4
Janek	30	1.7
Agnieszka	15	2.1
Franek	1	2.8

Znajdź model regresji liniowej przedstawiającej cenę biletu z Warszawy do Dubaju w zależności od dnia zakupu biletu przed odlotem. Na podstawie modelu oszacuj cenę lotu do Dubaju jeśli zdecyduję się na lot 45 dni przed odlotem.

2. Czterech studentówww (dwóch studentów i dwie studentki) – Gilbert, Hubert, Aisza i Jaśmina zanotowało, jaki procent swojego wolnego czasu poświęcają na naukę, oraz jakie uzyskują średnie oceny. Dane przedstawiono w tabeli. Model ma na celu zbadanie, czy większy nakład pracy w naukę przekłada się na wyższe wyniki.

Table 2: Średnie wyniki w nauce w zależności od czasu przeznaczonego na naukę

Student(ka)	**Czas wolny przeznaczony na naukę**	**Średnia ocen**
Gilbert	20	3.11
Hubert	45	4.70
Aisza	55	4.60
Jaśmina	18	3.75

Na podstawie powyższych danych należy: oszacować model regresji liniowej opisujący zależność średniej oceny od procentu wolnego czasu przeznaczanego na naukę, zapisać równanie regresji w postaci analitycznej i zinterpretować parametry modelu. O ile średnio wzrasta ocena przy wzroście czasu nauki o 10% czasu wolnego?

3. Na 4 terenach Polski przeprowadzono badania gleboznawcze mające na celu ocenę, czy zawartość próchnicy w glebie ma wpływ na plon pszenicy ozimej. Przyjmijmy, że na polach eksperymentalnych nawożenie było identyczne, a warunki pogodowe podobne.

Table 3: Plon pszenicy w zależności od zawartości próchnicy (pięć regionów Polski)

Region (okolice)	**Zawartość próchnicy X [%]**	**Plon pszenicy Y [t/ha]**
Zielona Góra	1.8	4.0
Zakopane	3.0	4.8
Zamość	3.5	5.4
Kołobrzeg	1.5	3.7
Włocławek	2.2	4.3

Na podstawie danych w tabelce oszacuj model regresji liniowej w glebie w zależności od zawartości próchnicy w glebie. Naszkicuj wykres badanej zależności.

4. Dla celów ilustracyjnych postanowiono sprawdzić, czy między wielkością plonów tulipanów w Holandii a średnim utargiem kin w Gnieźnie istnieje jakikolwiek związek statystyczny. Dane z lat 2015–2024, choć na pierwszy rzut oka niepowiązane, przedstawiono w Tabeli 4.

   Na podstawie danych z Tabeli 4 oszacuj model regresji liniowej oraz:

   (a) Zinterpretuj znak i wartość współczynnika kierunkowego β_1.

   (b) Wyjaśnij, dlaczego mimo bardzo wysokiego współczynnika determi-

Table 4: Plony tulipanów w Holandii a utarg kina w Gnieźnie (lata 2015-2024)

Rok	**Plony tulipanów X [mln szt.]**	**Utarg kina Y [mln PLN]**
2015	520	15.0
2016	545	14.4
2017	560	13.7
2018	585	13.2
2019	600	12.6
2020	625	11.7
2021	640	11.5
2022	665	10.6
2023	680	10.3
2024	705	9.5

nacji R^2 nie można uznać tej zależności za przyczynową.

Komentarz: Wstępna analiza wykazała bardzo silną a ujemną korelację pomiędzy obiema zmiennymi. Zjawisko to jest przykładem tzw. **regresji pozornej**, ponieważ oba procesy zależą od wspólnego czynnika pogodowego: duże nasłonecznienie sprzyja uprawie tulipanów, lecz jednocześnie zniechęca niektórych mieszkańców Gniezna do odwiedzin kina na rzecz plaży, jeziora, czy zawodów żużlowych).

5. W 10 miastach zebrano dane dotyczą ce:

   - liczby lekarzy przypadającej na 10 000 mieszkańców (X),
   - oraz rocznej liczby zgonów na 1 000 mieszkańców (Y).

   Dane przedstawiono w Tabeli 5.

   Narysuj wykres i opisz, dlaczego intuicyjny wniosek jest błędny.

   Komentarz: Wynik zasugeruje, że **im więcej lekarzy, tym więcej zgonów**. Wniosek wydaje się absurdalny, ponieważ zależność jest pozorna i wynika z **odwróconej przyczynowości**: to nie lekarze powodują zgony,

Table 5: Liczba lekarzy i liczba zgonów w wybranych miastach

Miasto	**Lekarze na 10 000 os.** (X)	**Zgony na 1 000 os.** (Y)
1	8	5.1
2	9	5.4
3	10	5.7
4	11	6.2
5	12	6.4
6	13	6.7
7	14	7.1
8	15	7.3
9	16	7.8
10	17	8.1

lecz **w większych i bardziej obciążonych chorobami regionach zatrudnia się więcej lekarzy**.

6. Celem zadania jest zbadanie, czy istnieje statystycznie istotny związek między liczebnością populacji żab a liczbą par lęgowych bociana białego w wybranych regionach pogranicza Wielkopolski, Dolnego Śląska i Ziemii Lubuskiej.

   Poniższa tabela zawiera dane dla roku 2024. Liczebność żab podano w tysiącach osobników (indeks populacyjny), a liczba par bocianów oznacza aktywne gniazda z lęgami.

Table 6: Dane dla roku 2024: populacja żab a liczba par bociana białego.

Lokalizacja	**Populacja żab [tys.]**	**Pary bocianów [liczba]**
Stawy Milickie & Dolina Baryczy	182	36
Nadleśnictwo Dębno Polskie (Rawicz)	128	41
Okolice Lginia i Brenna (Wschowa)	96	33
Okolice Osiecznej i Leszna	84	35

(a) Narysuj wykres punktowy zależności między liczebnością żab a liczbą par bocianów.

(b) Oszacuj prosty model regresji liniowej postaci:

$$\text{bociany} = \beta_0 + \beta_1 \cdot \text{żaby} + \varepsilon,$$

gdzie:

- zaby – liczebność populacji żab (w tys.),
- bociany – liczba par lęgowych bociana białego.

(c) Oblicz współczynniki $\hat{\beta}_0$, $\hat{\beta}_1$, i wartość R^2;

(d) Zinterpretuj, czy zależność między liczebnością żab a bocianów jest statystycznie istotna.

(e) Zastanów się, co mogłoby wyjaśniać ewentualny brak istotności statystycznej mimo logicznego powiązania ekologicznego.

Komentarz: Dla bociana żaba to to samo, czym dla człowieka są **orzeszki, paluszki albo kawałek czekolady** coś po chątnie sięgnie, ale bez czego można się doskonale obejść.