

STATYSTYCZNE METODY WSPOMAGANIA DECYZJI

zestaw zadań nr 7

Cel: analiza regresji – regresja prosta i wieloraka

Zadanie 1

W zamieszczonej poniżej tabeli podano wysokość rocznego dochodu i wartość posiadanego domu dziewięciu rodzin wybranych w sposób losowy spośród mieszkańców pewnego okręgu:

Roczny dochód (\$ 1000)	36	64	49	21	28	47	58	19	32
Wartość domu (\$ 1000)	129	310	260	92	126	242	288	81	134

- Wyznaczyć prostą regresji wartości domu względem dochodu.
- Przeanalizować dopasowanie modelu.
- Oszacować wartość domu rodziny, której roczny dochód wynosi \$40000.
- Wyznaczyć 95% przedział ufności dla szacowanej wartości domu tej rodziny.

Zadanie 2

Wyznaczyć prostą regresji poziomu cholesterolu względem wieku dziesięciu losowo wziętych mężczyzn.

Zweryfikować dopasowanie modelu.

Wiek	58	69	43	39	63	52	47	31	74	36
Poziom cholesterolu	189	235	193	177	154	191	213	175	198	181

Zadanie 3

W poniższej tabeli podano liczbę ludności USA (w mln) w latach 1890-2007:

Rok	1890	1900	1910	1920	1930	1940	1950	1960	1970
Ludność	62.947	75.994	91.972	105.710	122.775	131.669	150.697	179.323	203.235

Rok	1980	1990	2000	2007	2008	2009
Ludność	226.542	248.718	281.422	301.140	305.529	309.237

- Przyjmując wykładniczy model wzrostu populacji, oszacować parametry tego modelu i zweryfikować jego dopasowanie.
- Oszacować przewidywaną wielkość populacji USA w 2015 i w 2020 roku.

Zadanie 4

Niech X oznacza przeciętną liczbę samochodów poruszających się autostradą w ciągu dnia, natomiast Y liczbę wypadków samochodowych, która ma miejsce w ciągu miesiąca na autostradzie. Na podstawie danych zamieszczonych w poniższej tabeli wyznaczyć następujący model regresji

$$\sqrt{Y} = a + b \cdot X,$$

opisujący zależność liczby wypadków od natężenia ruchu na autostradzie. Oszacować liczbę wypadków, jakiej można się spodziewać przy natężeniu ruchu odpowiadającemu 3500 samochodom poruszającym się autostradą w ciągu dnia.

X	2000	2300	2500	2600	2800	3000	3100	3400	3700	3800	4000	4600	4800
Y	15	27	20	21	31	26	22	23	32	39	27	43	53

Zadanie 5

Dokonano osiem niezależnych pomiarów wielkości drgań pionowych gruntu powstałych w wyniku trzęsienia ziemi w różnej odległości od epicentrum trzęsienia. Otrzymano następujące wyniki:

Odległość od epicentrum (km)	20	30	40	50	80	140	200	250
Wielkość drgań pionowych (cm)	4.8	3.2	2.5	2.5	1.5	1.8	1.2	0.8

- Wyznaczyć funkcję regresji wielkości drgań gruntu względem odległości od epicentrum.
- Zweryfikować dopasowanie modelu.
- Oszacować wielkość drgań w odległości 100 km od epicentrum.

Zadanie 6

Korzystając z danych zawartych w poniższej tabeli wyznaczyć funkcję regresji, opisującą zależność między liczbą cykli do zniszczenia pewnego detalu a wywieranym na ten detal naprężeniem. Oszacować liczbę cykli do zniszczenia detalu, pracującego pod naprężeniem 40 tys. psi.

Naprężenie (w tys. psi)	55	50.5	43.5	42.5	42	41	35.7	34.5	33	32
Liczba cykli do zniszczenia (w mln cykli)	0.223	0.925	6.75	18.1	29.1	50.5	126	215	445	420

Zadanie 7

W pewnej firmie postanowiono zbadać zależność między wielkością tygodniowej sprzedaży produktów chemicznych tej firmy, a wydatkami poniesionymi na reklamę radiowo-telewizyjną oraz wydatkami poniesionymi na pokazy w sklepach. Oto dane (w tys. \$) pochodzące z 10 tygodni:

Wartość tygodniowej sprzedaży	72	76	78	70	68	80	82	65	62	90
Wydatki na reklamę radiowo-telewizyjną	12	11	15	10	11	16	14	8	8	18
Wydatki na pokazy w sklepach	5	8	6	5	3	9	12	4	3	10

- Wyznaczyć liniową funkcję regresji opisującą badaną zależność.
- Zweryfikować dopasowanie modelu.
- Wykorzystać uzyskane równanie regresji do prognozy wielkości sprzedaży, gdy wydatki na reklamę radiowo-telewizyjną wyniosą 8000\$, natomiast wydatki na pokazy w sklepach 12000\$.

Zadanie 8

Pośrednik w handlu nieruchomościami jest zainteresowany oszacowaniem wpływu powierzchni budynku i jego odległości od centrum miasta na wartość budynku. Poniższa tabela zawiera informacje o dziewięciu losowo wybranych budynkach.

Wartość budynku (tys. \$)	345	320	452	422	328	375	660	466	290
Powierzchnia (m ²)	150	180	200	160	175	180	300	170	135
Odległość od centrum (km)	5.6	1.2	2.4	7.2	2.9	2.5	5.5	4.8	1.6

- Wyznaczyć liniową funkcję regresji opisującą zależność, którą interesuje się ów pośrednik.
- Zweryfikować dopasowanie modelu.
- Podać przewidywaną wartość domu o powierzchni 160 m², położonego w odległości 3 km od centrum miasta.

Zadanie 9

Onkolodzy postanowili przetestować na myszach skuteczność nowego leku antyrakowego. W tym celu 10 myszom mającym guz o wadze 4g podawano różne dawki owego leku. Po ustalonym czasie zmierzono stopień redukcji wielkości nowotworu. Wyniki badania zamieszcza poniższa tabela.

Wielkość dawki	1	2	3	4	5	6	7	8	9	10
Redukcja wagi nowotworu	0.50	0.90	1.20	1.35	1.50	1.60	1.53	1.38	1.21	0.65

Wyznaczyć kwadratowy model regresji opisujący zależność stopnia redukcji wagi guza od wielkości dawki nowego leku.

Zadanie 10

Dane zamieszczone w pliku **geodezja.csv** zawierają wyniki doświadczenia, którego celem było zbadania wpływu środowiska na dokładność pomiarów geodezyjnych. Kolejne kolumny odpowiadają następującym pomiarom:

- kat – kąt refrakcji
- temp – temperatura powietrza
- wilg – wilgotność względna powietrza (w %)
- cis – ciśnienie atmosferyczne (w mm Hg)
- odl – odległość pomiędzy stanowiskiem a celem (w m)
- wys – średnia wysokość celowania (w m).

Znaleźć najlepszy model (liniowy) opisujący zależność kąta refrakcji od cech geodezyjnych (odległość i wysokość celowania) i atmosferycznych (kwadrat temperatury, wilgotność, ciśnienie).

Zadanie 11

Na podstawie danych zawartych w pliku **cardata.csv** wyznaczyć model liniowy najlepiej opisujący zależność zużycia paliwa od przyspieszenia, mocy silnika, liczby cylindrów, wagi i roku produkcji samochodu.