

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

- Wyznacz prostą regresji wartości domu względem dochodu.
 - Przeanalizuj dopasowanie modelu.
 - Oszacuj wartość domu rodziny, której roczny dochód wynosi \$40000.
 - Wyznacz 95% przedział ufności dla szacowanej wartości domu tej rodziny.
2. Wyznaczyć prostą regresji poziomu cholesterolu względem wieku dziesięciu losowo wziętych mężczyzn. Zweryfikuj 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

3. W poniższej tabeli podano liczbę ludności USA (w mln) w latach 1890-2020:

Rok	Populacja	Rok	Populacja	Rok	Populacja
1890	62.947	1940	131.669	1990	248.718
1900	75.994	1950	150.697	2000	281.422
1910	91.972	1960	179.323	2010	308.746
1920	105.710	1970	203.235	2020	323.996
1930	122.775	1980	226.542	2030	?

- Przyjmując wykładniczy model wzrostu populacji, oszacuj parametry modelu i zweryfikuj jego dopasowanie.
 - Oszacuj przewidywaną wielkość populacji USA w 2030 roku.
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 wyznacz następujący model regresji $\sqrt{Y} = a + b \cdot X$, opisujący zależność liczby wypadków od natężenia ruchu na autostradzie. Oszacuj 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

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

- Wyznacz funkcję regresji wielkości drgań gruntu względem odległości od epicentrum.
- Zweryfikuj dopasowanie modelu.
- Oszacuj wielkość drgań w odległości 100 km od epicentrum.

6. Korzystając z danych zawartych w poniższej tabeli wyznacz funkcję regresji, opisującą zależność między liczbą cykli do zniszczenia pewnego detalu a wywieranym na ten detal naprężeniem. Oszacuj 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

7. Pewna firma, chcąc sprawdzić jak liczba reklam zamieszczanych w gazetach w ciągu miesiąca wpływa na zainteresowanie klientów pewnym towarem, przeprowadziła badania otrzymując następujące wyniki:

liczba reklam	8	10	12	14	15	17	20	21	23	26
odsetek zainteresowanych klientów	0.4	0.42	0.43	0.44	0.46	0.45	0.48	0.49	0.5	0.51

- Wyznacz liniowy model regresji dla odsetka klientów zainteresowanych tym towarem w zależności od liczby reklam.
 - Zweryfikuj dopasowanie modelu.
 - Na podstawie dobranego modelu oszacuj odsetek klientów zainteresowanych tym towarem, jeżeli firma zamieści w ciągu tygodnia 19 reklam. Podać 95% przedział ufności dla tej prognozy.
8. Badano zależność jednostkowego kosztu produkcji pewnego wyrobu od wielkości jego rocznej produkcji. Otrzymano następujące dane:

wielkość produkcji (w tys. sztuk)	10	25	35	50	70	80	95	110
koszt jednostkowy (w zł)	25	21	20	17	14	9	4	2

- Wyznacz liniowy model regresji opisujący badaną zależność.
 - Zweryfikuj dopasowanie modelu.
 - Na podstawie dobranego modelu podaj prognozę jednostkowego kosztu produkcji tego wyrobu przy produkcji wynoszącej 40 tys. sztuk. Podać 95% przedział ufności dla tej prognozy.
9. 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

- Wyznacz liniową funkcję regresji opisującą zależność, którą interesuje się ów pośrednik.
 - Zweryfikuj dopasowanie modelu.
 - Podaj przewidywaną wartość domu o powierzchni 160 m², położonego w odległości 3 km od centrum miasta.
10. Badano zależność między liczbą wypalanych dziennie papierosów, a prawdopodobieństwem zachorowania na raka płuc w populacji 40-letnich palaczy, palących od 10 lat. Uzyskano następujące dane.

liczba papierosów	5	10	20	30	40	50	60
prawdopodobieństwo	0.061	0.113	0.192	0.259	0.339	0.401	0.461

- Wyznacz potęgowy model regresji opisujący badaną zależność.
- Zweryfikuj dopasowanie modelu.
- Oszacuj prawdopodobieństwo zachorowania na raka płuc przez palacza wypalającego 35 papierosów dziennie.

11. *Sprawdzano jak zależy efektywność nowego programu od wielkości danych wejściowych. Efektywność była mierzona liczbą wykonanych żądań na godzinę. Uruchamiając program przy różnej wielkości danych wejściowych otrzymano następujące wyniki

rozmiar danych (w GB)	6	7	7	8	10	10	15
liczba przetworzonych żądań	40	55	50	41	17	26	16

- a) Wyznacz liniowy model regresji opisujący badaną zależność.
b) Zweryfikuj dopasowanie modelu.
12. *W celu poprawy modelu efektywności programu (zadanie 3) dodatkowo wzięto pod uwagę liczbę tabel w jakich zamieszczono dane wejściowe oraz rodzaj systemu operacyjnego.

rozmiar danych (w GB)	6	7	7	8	10	10	15
liczba tabel	4	20	20	10	10	2	1
system operacyjny	A	A	A	A	B	B	B
liczba przetworzonych żądań	40	55	50	41	17	26	16

- a) Wyznacz liniowy model regresji opisujący badaną zależność.
b) Zweryfikuj dopasowanie modelu. Czy otrzymany model jest lepszy od modelu wyznaczonego w zadaniu 3?
13. *Przesłano 30 plików o średnim rozmiarze 126 KB i z odchyleniem standardowym 35 KB. Średni czas przesyłania pliku wyniósł 0,04 s z odchyleniem standardowym 0,01 s. Współczynnik korelacji między czasem przesyłania a wielkością pliku wyniósł 0,86. Na podstawie tych danych wyznacz model prostej regresji liniowej oraz współczynnik determinacji R^2 . Zakładając poprawność modelu, oszacuj ile czasu może zająć przesłanie pliku o wielkości 200 KB.
14. *Na podstawie danych zawartych w pliku *samochody.csv* wyznacz model liniowy najlepiej opisujący zależność zużycia paliwa od przyspieszenia, mocy silnika, liczby cylindrów, wagi i roku produkcji samochodu.
15. *Na podstawie danych zawartych w ramce danych *Carseats* z pakietu *ISLR* wyznacz model liniowy najlepiej opisujący zależność wielkości sprzedaży (Sales) od pozostałych zmiennych.