

METODY EKSPLOKACJI DANYCH
Laboratorium. Analiza regresji - regresja liniowa
Przekształcanie funkcji nieliniowych w równoważne liniowe
Budowa modelu na podstawie danych (ogólnie dostępnych)

ZADANIE

Mamy zgromadzone dane dot. dobrze znanego przedsiębiorstwa internetowego Meta w tym odnoszące się do portalu „Facebook”, które zostały zawarte w poniższych tabelach. Dane te dotyczą liczby użytkowników (klientów) w rozliczeniu na kwartały w poszczególnych latach. Pozostałe dane to przychody i zysk liczone w milionach dolarów amerykańskich. Ostatnia wielkość to zatrudnienie.

Tabela 1. Liczba użytkowników portalu społecznościowego „Facebook”

Liczba użytkowników		Liczba użytkowników		Liczba użytkowników	
Kwartał	w mln	Kwartał	w mln	Kwartał	w mln
Q3 '08	100	Q1 '09	197	Q1 '14	1276
		Q2 '09	242	Q2 '14	1317
		Q3 '09	305	Q3 '14	1350
		Q4 '09	360	Q4 '14	1393
		Q1 '10	431	Q1 '15	1441
		Q2 '10	482	Q2 '15	1490
		Q3 '10	550	Q3 '15	1545
		Q4 '10	608	Q4 '15	1591
		Q1 '11	680	Q1 '16	1654
		Q2 '11	739	Q2 '16	1712
		Q3 '11	800	Q3 '16	1788
		Q4 '11	845	Q4 '16	1860
		Q1 '12	901	Q1 '17	1936
		Q2 '12	955	Q2 '17	2006
		Q3 '12	1007	Q3 '17	2072
		Q4 '12	1056	Q4 '17	2129
		Q1 '13	1110		
		Q2 '13	1155		
		Q3 '13	1189		
		Q4 '13	1228		

Tabela 2. Przychody, zysk i zatrudnienie przedsiębiorstwa „Facebook”

Rok	Przychód w mln \$	Zysk w mln \$	Zatrudnienie
2007	153	-138	450
2008	272	-56	850
2009	777	229	1218
2010	1974	606	2127
2011	3711	1000	3200
2012	5089	53	4619
2013	7872	1500	6337
2014	12466	2940	9199
2015	17928	3688	12691
2016	27638	10217	17048
2017	40653	15934	25105

W analizie proszę:

- 1) Proszę przeanalizować dane i zastanowić się nad tym „Co chcemy zbadać i dlaczego?”, tzn. na jakie pytania chcemy sobie odpowiedzieć.
- 2) W powyższym kontekście, na gruncie regresji, należy zaproponować **model** lub **modele badające wybrane zależności** i wyliczyć m. in. ich parametry strukturalne, odchylenia standardowe, miary dopasowania, czy też przebadać hipotezy, itp.
- 3) Analizę proszę na początek przeprowadzić na podstawie danych, które zostały przedstawione w powyższych tabelach.
- 4) Proponowane modele należy uzasadnić oraz wskazać ewentualny sposób ich praktycznego użycia w tym np. do predykcji.
- 5) Na podstawie opracowanych modeli i przeprowadzonych obliczeń **sformułować własne wnioski**.
- 6) Następnie proszę odnaleźć w ogólnie dostępnych źródłach dane:
 - a) za lata 2018-2020 i sprawdzić zachowanie się (wykorzystanie) modeli np. w procesie predykcji za lata 2018-2020.
 - b) za lata 2021-2023 i sprawdzić zachowanie się (wykorzystanie) modeli np. w procesie predykcji za lata 2021-2023.
- 7) Wyniki swoich analiz zawierających: postawione pytania „badawcze”, hipotezy, wzory, udokumentowane wyniki obliczeń, wnioski, itp. proszę zawrzeć w postaci **sprawozdania**. Obliczenia można przeprowadzić w dowolnie wybranym narzędziu, np. które do tej pory wykorzystywaliśmy na zajęciach. Do sprawozdania proszę dodać jako załączniki wszystkie pliki z obliczeniami.

MODELE NIELINIOWE - LINEARYZOWANE
PRZYKŁADOWE METODY LINEARYZACJI
WYBRANYCH FUNKCJI NIELINIOWYCH

Poniżej przedstawiono przykłady linearyzacji nieliniowych modeli.

Przykład 1.

Niech model będzie **funkcją wykładniczą** z jedną zmienną objaśniającą:

$$\widehat{W} = b \cdot a^X$$

Powyższy model sprowadzamy do modelu liniowego logarytmując¹ obustronnie otrzymując:

$$\log \widehat{W} = \log b + X \cdot \log a$$

Przyjmując, że $\hat{Y} = \log \widehat{W}$, $a_0 = \log b$, $a_1 = \log a$ otrzymujemy następujący równoważny liniowy model:

$$\hat{Y} = a_1 \cdot X + a_0$$

Przykład 2.

Rozpatrzmy model oparty o **funkcję wykładniczą** z dwoma zmiennymi objaśniającymi:

$$\widehat{W} = b \cdot a^{X_1} \cdot c^{X_2}$$

Powyższy model sprowadzamy do modelu liniowego logarytmując obustronnie otrzymując:

$$\log \widehat{W} = \log b + X_1 \cdot \log a + X_2 \cdot \log c$$

Przyjmując, że $\hat{Y} = \log \widehat{W}$, $a_0 = \log b$, $a_1 = \log a$, $a_2 = \log c$ otrzymujemy następujący równoważny liniowy model:

$$\hat{Y} = a_1 \cdot X_1 + a_2 \cdot X_2 + a_0$$

Przykład 3.

Rozpatrzmy **model potęgowy** z dwoma zmiennymi objaśniającymi:

$$\widehat{W} = b \cdot X^{a_1} \cdot Z^{a_2}$$

Powyższy model sprowadzamy do modelu liniowego logarytmując obustronnie otrzymując:

$$\log \widehat{W} = \log b + a_1 \cdot \log X + a_2 \cdot \log Z$$

Przyjmując, że $\hat{Y} = \log \widehat{W}$, $a_0 = \log b$, $X_1 = \log X$, $X_2 = \log Z$ otrzymujemy następujący równoważny liniowy model:

$$\hat{Y} = a_1 \cdot X_1 + a_2 \cdot X_2 + a_0$$

¹ $\log(x) \equiv \log x$; podstawę logarytmu wybieramy wg. uznania, np. za podstawę logarytmu możemy przyjąć liczbę e

Przykład 4.

Rozpatrzmy model **wykładniczo-potęgowy** postaci:

$$\widehat{W} = b \cdot a^{X_1} \cdot Z^{a_2}$$

Powyższy model sprowadzamy do modelu liniowego logarytmując obustronnie otrzymując:

$$\log \widehat{W} = \log b + X_1 \cdot \log a + a_2 \cdot \log Z$$

Przyjmując, że $\hat{Y} = \log \widehat{W}$, $a_0 = \log b$, $a_1 = \log a$, $X_2 = \log Z$ otrzymujemy następujący równoważny liniowy model:

$$\hat{Y} = a_1 \cdot X_1 + a_2 \cdot X_2 + a_0$$

Przykład 5.

Niech modelem będzie **wielomianem** stopnia trzeciego (model z jedną zmienną objaśniającą):

$$\hat{Y} = a_0 + a_1 \cdot X + a_2 \cdot X^2 + a_3 \cdot X^3$$

Powyższy model sprowadzamy do modelu liniowego podstawiając

$$X = X_1, X^2 = X_2, X^3 = X_3$$

W konsekwencji wykonanych podstawień otrzymujemy równoważny liniowy model:

$$\hat{Y} = a_1 \cdot X_1 + a_2 \cdot X_2 + a_3 \cdot X_3 + a_0$$

Przykład 6.

Niech modelem (z jedną zmienną objaśniającą) będzie **funkcja wymierna ułamkowa**², której wykresem jest hiperbola:

$$\hat{Y} = a_1 \cdot \frac{1}{X} + a_0$$

Powyższy model sprowadzamy do modelu liniowego podstawiając

$$\frac{1}{X} = X_1$$

W konsekwencji wykonanych podstawień otrzymujemy równoważny liniowy model:

$$\hat{Y} = a_1 \cdot X_1 + a_0$$

² proporcjonalność odwrotna