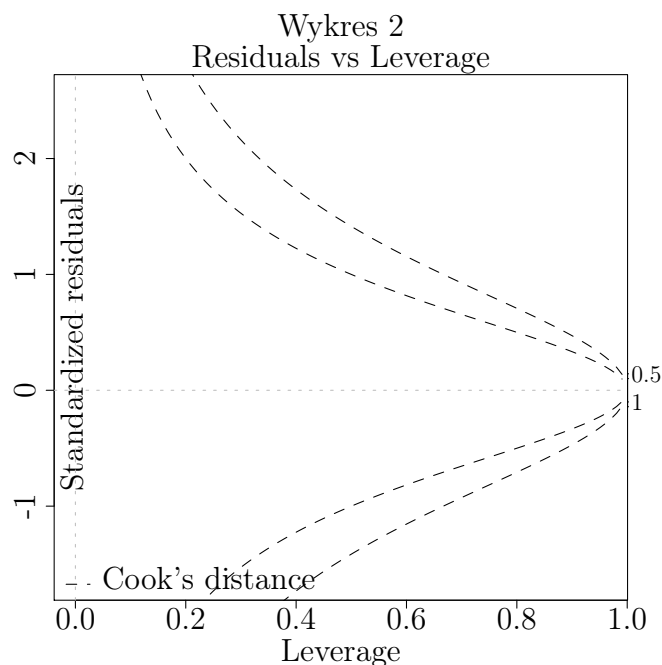
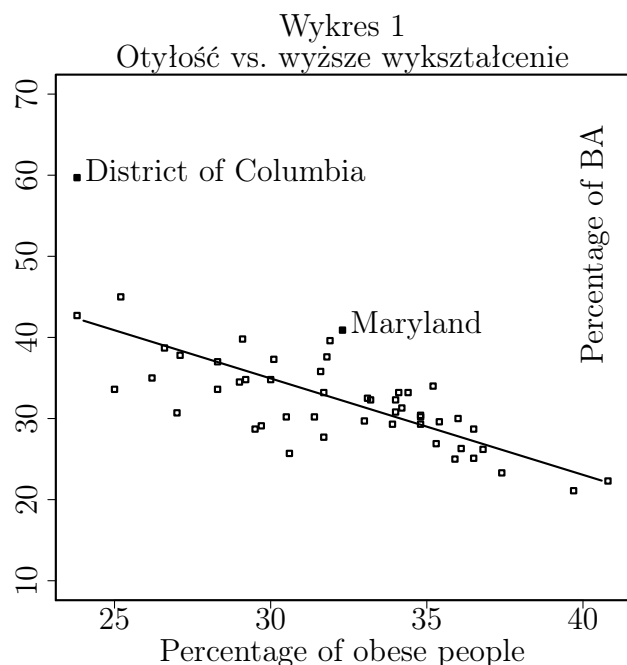


Egzamin z Ekonometrii dla IiE i MSEMat - 03.03.2020

Zadanie 1 (40%)

Alberto Cairo w swojej książce¹ analizował zależność między odsetkiem otyłych dorosłych (*Obese*) w zależności od odsetka osób z tytułem licencjata (*Bachelor's degree*) lub wyższym (zmienna *Bachelor*) pomiędzy wszystkimi stanami USA. Dane² zostały przedstawione na wykresie 1.



Przyjmij 5% poziom istotności. Na wydrukach standardowo używana jest kropka jako separator dziesiętny.

Call:

```
lm(formula = Bachelor_Obesity$Obesity ~ Bachelor_Obesity$Bachelor)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.5467	-2.0582	0.8916	1.9116	4.1721

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	46.65541	2.00875	23.226	< 2e-16 ***
Bachelor_Obesity\$Bachelor	-0.44967	0.06057	-7.424	1.65e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.754 on 48 degrees of freedom

Multiple R-squared: 0.5345, Adjusted R-squared: 0.5248

F-statistic: 55.11 on 1 and 48 DF, p-value: 1.653e-09

¹Alberto Cairo, *The Functional Art. An Introduction to information graphics and visualization*, New Riders, 2013.

²Dane są publicznie dostępne. Informacje o odsetku osób otyłych można znaleźć tu <https://www.cdc.gov/nccdphp/dnpao/data-trends-maps/index.html>, a o odsetku wyższego wykształcenia - tu <https://www.census.gov/acs/www/data/data-tables-and-tools/ranking-tables/>

```
> resettest(model, power=2:3, type="fitted")
RESET test
data:  model
RESET = 2.2838, df1 = 2, df2 = 46, p-value = 0.1133
```

```
> bptest(model)
studentized Breusch-Pagan test
data:  model
BP = 0.53541, df = 1, p-value = 0.4643
```

```
> jarque.bera.test(model$residuals)
Jarque Bera Test
data:  model$residuals
X-squared = 3.3014, df = 2, p-value = 0.1919
```

- a) (5%) Jaka jest zależność między odsetkiem osób otyłych a odsetkiem osób z wyższym wykształceniem? Uzasadnij.
- b) (5%) Czy Dystrykt Kolumbii jest którymkolwiek z: vertical outlier, bad leverage point, good leverage point? Umieść tę obserwację na wykresie 2. **Uzasadnij.**
- c) (5%) Czy można pominąć Dystrykt Kolumbii w analizie? Uzasadnij.
- d) (5%) Czy forma funkcyjna modelu jest prawidłowa? Uzasadnij.
- e) (5%) Czy występuje heteroskedastyczność reszt? Uzasadnij.
- f) (5%) Czy reszty mają rozkład normalny? Uzasadnij.
- g) (5%) Jak weryfikacja założeń tego modelu (forma funkcyjna, heteroskedastyczność, rozkład normalny) przekłada się na własności modelu?
- h) (5%) Jakie dodatkowe zmienne mogłyby być użyte do modelowania odsetka otyłych osób? Wymień dwie i uzasadnij.

Zadanie 2 (30%)

Dla danych o 880 domach z miejscowości Stockton w Kalifornii w Stanach Zjednoczonych oszacowano model dla ceny domu *price*, w którym zmiennymi objaśniającymi były *age* - wiek w latach, *beds* - liczba sypialni, *sqft* - powierzchnia mieszkania w stopach kwadratowych, *stories* jest zmienną zerojedynkową przyjmującą wartość 1 dla domów mających piętro i 0 dla jednopiętrowych i *baths* - liczba łazienek. W nawiasach podane są odchylenia standardowe oszacowań parametrów.

	model1	model2	model3	model4	model5
	b/se	b/se	b/se	b/se	b/se
beds	-17474.398???	-17616.319***	-18871.321***	29274.041***	
	(1983.61)	(1962.00)	(1899.98)	(2437.87)	
baths	-1332.919???				
	(2688.31)				
age	-299.074???	-287.782***	-296.001***	-309.598***	-183.874***
	(53.95)	(48.88)	(48.91)	(81.65)	(50.17)
stories	-6708.066???	-7340.456*			
	(3246.13)	(2983.77)			
sqft	98.139???	97.368***	95.186***		79.912***
	(2.99)	(2.56)	(2.40)		(1.95)
Constant	21485.424???	20293.376***	26474.579***	27601.888**	-11489.888**
	(6258.53)	(5775.83)	(5215.68)	(8707.74)	(3741.08)
RSS	7.0655	7.0675	7.1164	19.859	7.9178
F	431.770	540.117	714.021	102.872	919.338
R2	0.712	0.712	0.710	0.190	0.677
R2_skorygowane	0.710	0.710	0.709	0.188	0.676
N	880.000	880.000	880.000	880.000	880.000

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Przyjmij 5% poziom istotności. Na wydrukach standardowo używana jest kropka jako separator dziesiętny.

- (10%) Oceń, które zmienne w modelu 1 są statystycznie istotne na poziomie istotności 5%.
- (6%) Który model należy wybrać na podstawie odpowiedniej miary dopasowania?
- (7%) Dla modelu (1) zweryfikować hipotezę, że liczba łazienek, powierzchnia oraz to czy dom jest piętrowy czy parterowy łącznie wpływają na cenę domu.
- (7%) Postanowiono zweryfikować dla modelu (1) hipotezę, że kolor oczu sprzedającego (6 poziomów) ma wpływ na cenę domu. W związku z tym oszacowano model ze zmiennymi zerojedynkowymi dla koloru oczu i otrzymano wartość sumy kwadratów reszt równą 7.0245.

$F(d_1, d_2)$	$d_2 = 869$	$d_2 = 872$	$d_2 = 873$	$d_2 = 874$	$d_2 = 875$	$d_2 = 876$	$d_2 = 877$	$d_2 = 878$	$d_2 = 879$	$d_2 = 880$
1	3.852182	3.852145	3.852133	3.852120	3.852108	3.852096	3.852084	3.852072	3.852060	3.852048
2	3.006083	3.006048	3.006036	3.006024	3.006012	3.006000	3.005989	3.005977	3.005965	3.005954
3	2.615148	2.615112	2.615101	2.615089	2.615077	2.615066	2.615054	2.615042	2.615031	2.615019
4	2.382176	2.382140	2.382129	2.382117	2.382105	2.382094	2.382082	2.382070	2.382059	2.382047
5	2.224405	2.224370	2.224358	2.224346	2.224334	2.224323	2.224311	2.224299	2.224288	2.224276
6	2.108996	2.108960	2.108948	2.108937	2.108925	2.108913	2.108901	2.108889	2.108878	2.108866

Zadanie 3 (30%)

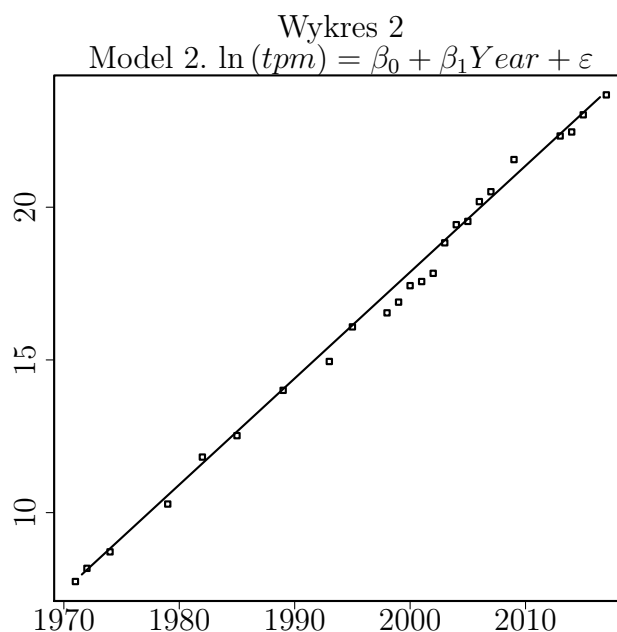
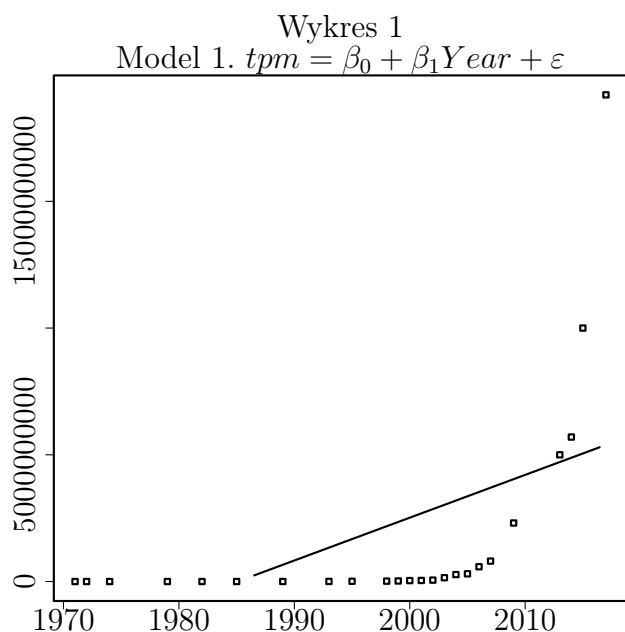
Prawo Moore'a zgodnie z definicją z Wikipedii³ jest to "prawo empiryczne, wynikające z obserwacji, że ekonomicznie optymalna liczba tranzystorów w układzie scalonym zwiększa się w kolejnych latach zgodnie z trendem wykładniczym (podwaja się w niemal równych odcinkach czasu). Autorstwo tego prawa przypisuje się Gordonowi Moore'owi, jednemu z założycieli firmy Intel, który w 1965 r. zaobserwował podwajanie się liczby tranzystorów co ok. 18 miesięcy. Liczba ta była następnie korygowana i obecnie przyjmuje się, że liczba tranzystorów w mikroprocesorach od wielu lat podwaja się co ok. 24 miesiące. Na zasadzie analogii, prawo Moore'a stosuje się też do wielu innych parametrów sprzętu komputerowego, np. pojemności dysków twardych czy wielkości pamięci operacyjnej."

Państwa wykładowca postanowił sprawdzić zasadność tego prawa i wykorzystał dane dostępne na stronie Our World in Data⁴. Następnie oszacował dwa modele regresji dla zmiennej kodującej liczbę tranzystorów w układzie scalonym na upływającym czasie w latach (zmienna *Year*).

$$tpm = -351182073587 + 176765146Year + e \quad (\text{Wykres 1})$$

$$\ln(tpm) = -678.682 + 0.348Year + e \quad (\text{Wykres 2})$$

Tablica w stylu publikacyjnym na następnej stronie przedstawia oszacowania obydwu modeli.



Przyjmij 5% poziom istotności. Na wydrukach standardowo używana jest kropka jako separator dziesiętny, a przecinek jest separatorem 1000.

- (10%) Zinterpretuj oszacowania parametrów obydwu modeli.
- (10%) Który z obydwu modeli lepiej wyjaśnia zmienną *tpm*? Uzasadnij.
- (10%) Czy Prawo Moore'a znajduje potwierdzenie w wynikach przedstawionych modeli? Uzasadnij.

³https://pl.wikipedia.org/wiki/Prawo_Moore%E2%80%99a

⁴<https://ourworldindata.org/grapher/transistors-per-microprocessor>

=====		
Dependent variable:		
	tpm	I(log(tpm))
	(1)	(2)

Year	176,765,146.000*** (56,790,574.000)	0.348*** (0.005)
Constant	-351,182,073,587.000*** (113,425,216,105.000)	-678.682*** (9.908)

Observations	24	24
R2	0.306	0.996
Adjusted R2	0.274	0.995
Residual Std. Error (df = 22)	3,768,754,581.000	0.329
F Statistic (df = 1; 22)	9.688***	4,926.809***
=====		
Note:	*p<0.1; **p<0.05; ***p<0.01	