

imię, nazwisko, nr indeksu: \_\_\_\_\_

**Ekonometria — egzamin**  
**02/02/2011**

1. Egzamin trwa 90 minut.
2. Rozwiązywanie zadań należy rozpocząć po ogłoszeniu początku egzaminu a skończyć wraz z ogłoszeniem końca egzaminu. Złamanie tej zasady skutkuje usunięciem z sali i unieważnieniem pracy.
3. W trakcie egzaminu wolno używać jedynie długopisu o innym kolorze atramentu niż czerwony oraz kalkulatora.
4. Przed przystąpieniem do pisania egzaminu należy podpisać **wszystkie kartki** pracy (na dole w przewidzianym miejscu). Złożenie podpisu pod regulaminem oznacza jego akceptację. Do egzaminu mogą przystąpić osoby, które akceptują regulamin.
5. W razie braku podpisu lub numeru zadania na kartce, kartka nie zostanie oceniona. Nie będą też oceniane rozwiązania wpisane na kartkach innych, niż te rozdawane przez prowadzących.
6. Rozwiązanie każdego zadania należy zapisać na kartce z tymże zadaniem, ewentualnie na czystych kartkach znajdujących się na końcu arkusza egzaminacyjnego lub na dodatkowych kartkach uzyskanych od prowadzących egzamin.
7. Na jednej kartce może znajdować się rozwiązanie tylko jednego zadania. Oceniane jest rozwiązanie tylko tego zadania, którego numer widnieje na kartce.
8. Egzamin składa się z czterech pytań teoretycznych i 3 zadań.
9. Posiadanie przy sobie wszelkich materiałów drukowanych (w tym książek) oraz innych np. wykonanych własnoręcznie materiałów zostanie uznane za ściąganie.
10. Rozmowy z innymi zdającymi będą traktowane identycznie jak ściąganie.
11. Każda zauważona próba ściągania skutkuje usunięciem z egzaminu.
12. Wszystkie pytania należy kierować bezpośrednio do osób pilnujących.
13. Warunkiem uzyskania oceny pozytywnej jest zdobycie conajmniej 50 % punktów z części teoretycznej egzaminu oraz min. 40 % punktów z części zadaniowej.

Warszawa, 02/02/2011, \_\_\_\_\_

*podpis*

Powodzenia :-)

**Pytanie 1.**

Wyprowadź wzór na estymator MNK w Klasycznym Modelu Regresji Liniowej z wieloma zmiennymi objaśniającymi.

**Pytanie 2.**

Wyprowadź postać macierzy wariancji-kowariancji estymatora wektora parametrów i podaj interpretację jej elementów.

**Pytanie 3.**

Pokaż, że w modelu oszacowanym na podstawie losowej próby estymator MNK jest nieobciążony. Zapisz niezbędne do przeprowadzenia dowodu założenia.

**Pytanie 4.**

Opisz korzyści i niebezpieczeństwa związane z nakładaniem ograniczeń na parametry modelu.

1	2	3	4	$\Sigma$

ZADANIE 1 Oszacowano parametry modelu

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

Dla wektora  $\mathbf{y}' = [1, 2, 2, y_4, y_5]$  uzyskano oszacowania  $\hat{\mathbf{y}}' = [2, 2, 1, 1, 2]$ .

1. Oblicz wartość  $y_4$  oraz  $y_5$
2. Oblicz wartość estymatora wariancji błędu losowego
3. Oblicz wartość współczynnika determinacji liniowej
4. Zweryfikuj hipotezę  $\beta_1 = \beta_2 = 0$ , wiedząc, że  $F(1, 2) = 199,5$ ;  $F(1, 3) = 215,71$ ;  $F(2, 1) = 18,51$ ;  $F(2, 2) = 19$ ;  $F(2, 3) = 19,16$ ;  $F(3, 1) = 10,13$ ;  $F(3, 2) = 9,55$ ;  $F(3, 3) = 9,28$ .

1	2	3	4	5	$\Sigma$

ZADANIE 2 Na podstawie danych pochodzących z badania budżetów gospodarstw domowych zbudowano model uzależniający wydatki na żywność od dochodu w gospodarstwie domowym (model A), oraz wydatki na żywność na osobę od dochodu w gospodarstwie domowym na osobę (model B)

Number of obs = 37358; R-squared = 0.2512

Model A	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dochg	.3600137	.0032163	111.93	0.000	.3537097	.3663178
_cons	1383.3	13.0976	105.61	0.000	1357.629	1408.972

Number of obs = 37358; R-squared = 0.1994

Model B	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dochg_os	.3163207	.0032796	96.45	0.000	.3098926	.3227487
_cons	627.1354	5.653069	110.94	0.000	616.0552	638.2156

Bez zmiany całkowitej sumy kwadratów modelu zaproponuj sposób modyfikacji modelu A lub B (wskaz modyfikowany model) w celu zbadania czy wydatki na żywność zależą od

1. liczby osób w gospodarstwie domowym (lo)
2. liczby dzieci w gospodarstwie domowym (ld)
3. liczby i płci dzieci w gospodarstwie domowym

Zapisz postać testowanej hipotezy.

Do modelu wprowadzono dodatkową zmienną zero-jedynkową informującą o lokalizacji gospodarstwa domowego w mieście lub na wsi, Wiedząc, że po wprowadzeniu tej zmiennej współczynnik determinacji linowej modelu zwiększył się o 4 punkty procentowe stwierdź czy (wartość krytyczna z statystyki testowej wynosi  $F^* = 3$ )

4. lokalizacja gospodarstwa domowego wpływa na poziom konsumpcji
5. czy dla odpowiedzi na pytanie w pkt 4 ma znaczenie który model wykorzystujesz?

1	2	3	4	5	6	$\Sigma$

ZADANIE 3

Na podstawie danych pochodzących badania aktywności ekonomicznej ludności w trzecim kwartale 2008 roku starano się zbadać hipotezę o niedostosowaniu wykształcenia uzyskanego przed 1995 rokiem do potrzeb rynku pracy (zmienna *po1995* przyjmuje wartość 1 dla wykształcenia uzyskanego w 1995 i później). W tym celu oszacowano na próbie mężczyzn równanie logarytmu miesięcznej płacy jako funkcję doświadczenia zawodowego *dosw* i jego kwadratu *dosw2*.

Source	SS	df	MS	Number of obs =	187
Model	5.29190715	3	1.76396905	F( 3, 483) =	8.43
Residual	101.041131	183	.209194889	Prob > F =	0.0000
Total	106.333039	186	.21879226	R-squared =	0.0498
				Adj R-squared =	0.0439
				Root MSE =	.45738

lzarobki	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
po_1995	-.064972	.0613751	-1.06	0.290	-.1855671 .0556231
dosw	.0233748	.0073829	3.17	0.002	.0088682 .0378814
dosw2	-.00063	.000149	-4.23	0.000	-.0009229 -.0003372
_cons	7.241466	.0931081	77.77	0.000	7.058519 7.424413

Breusch-Pagan	chi2(3)	=	4.50	Prob > chi2 =	0.2123
Breusch-Pagan	chi2(1)	=	1.15	Prob > chi2 =	0.2842
Ramsey RESET	F(3, 180)	=	1.04	Prob > F =	0.3741

Przyjmując poziom istotności  $\alpha = 0,05$  odpowiedz na poniższe pytania:

1. Oceń istotność i łączną istotność oszacowanych parametrów.
2. Oblicz efekt krańcowy dla doświadczenia zawodowego jeżeli przeciętne doświadczenie zawodowe wynosi 25,9 roku.
3. Czy weryfikowana w badaniu hipoteza jest prawdziwa?
4. Zbadaj czy założenia KMRL są spełnione.

Dokonano ponownego oszacowania parametrów modelu na podstawie próby liczącej 5000 obserwacji

Source	SS	df	MS	Number of obs =	4105
Model	44.3273847	3	14.7757949	F( 3, 4101) =	68.65
Residual	882.717531	4101	.21524446	Prob > F =	0.0000
Total	927.044916	4104	.225888137	R-squared =	0.0478
				Adj R-squared =	0.0471
				Root MSE =	.46394

lzarobki	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
po_1995	.0718284	.0215547	3.33	0.001	.0295696 .1140873
dosw	.0337558	.0026494	12.74	0.000	.0285616 .0389501
dosw2	-.000761	.0000538	-14.14	0.000	-.0008665 -.0006555
_cons	7.046439	.0335917	209.77	0.000	6.980581 7.112297

Breusch-Pagan	chi2(3)	=	49.40	Prob > chi2 =	0.0000
Breusch-Pagan	chi2(1)	=	18.58	Prob > chi2 =	0.0000
Ramsey RESET	F(3, 4098)	=	0.13	Prob > F =	0.9433

5. Skomentuj ponownie obliczone wartości statystyk testów diagnostycznych.
6. Jeżeli model nie spełnia założeń KMRL opisz jakie to ma konsekwencje dla poprawności wnioskowania statystycznego, oraz jakie są metody radzenia sobie z tym problemem.

Każdą odpowiedź uzasadnij wynikiem odpowiednich testów diagnostycznych zapisując wartość statystyki testowej lub jej *p-value*, oraz interpretację wyniku.

## Rozwiązania zadań

### Zadanie 1

1. Z własności hiperpłaszczyzny regresji wiemy, że  $\bar{y} = \bar{\hat{y}}$ , oraz że  $\hat{y}'e = 0$ . Zatem

$$\begin{cases} 5 + y_4 + y_5 = 8 \\ e_4 + 2e_5 = 0 \end{cases}$$

więc  $y_4 = 0, y_5 = 3$ .

2.  $e = [-1, 0, 1, -1, 1]$ . Zatem  $RSS=4$ , a  $S^2 = \frac{1}{5-3}4 = 2$

3.  $R^2 = \frac{14}{18}$

4. Ponieważ współczynnik determinacji modelu ze stałą wynosi 0 to

$$F = (N - k)/g \frac{R^2 - R_R^2}{R^2} = (5 - 3)/2 = 1 < F^* = 19$$

Zatem brak jest podstaw do odrzucenia hipotezy  $\beta_1 = \beta_2 = 0$

### Zadanie 2

- Model A: dodanie zmiennej  $lo$ :  $H_0 : \beta_{lo} = 0$ . Nie jest możliwe testowanie hipotezy w ramach modelu B.
- Model A: dodanie zmiennej  $ld$ :  $H_0 : \beta_{ld} = 0$ . Model B: dołączenie zmiennej  $ld/lo$ :  $H_0 : \beta_{ld/lo} = 0$
- Model A: dołączenie zmiennych np.  $L_{dziewczynek}, L_{chlopcow}$ :  $H_0 : \beta_{L_{dziewczynek}} = \beta_{L_{chlopcow}}$ .
- 

$$F = \frac{RSS_R - RSS_U}{(1 - RSS_U)/(N - k)} = (N - k) \frac{R^2 - R_R^2}{R^2}$$

Dla modelu A:

$$F_A \approx \frac{0.04}{0.76} * 37000 \approx 1950 > F^*$$

Dla modelu B:

$$F_B \approx \frac{0.04}{0.72} * 37000 \approx 2055 > F^*$$

5. Dla wyników testu nie ma znaczenia który model zostanie wykorzystany (obliczenia w pkt 4)

### Zadanie 3

- Oceny parametrów są łącznie statystycznie istotne o czym świadczy  $p$ -value statystyki  $F = 0.00 < \alpha$ . Pojedynczo istotne są oceny parametrów dla zmiennych  $dosw$  ( $p$ -value statystyki  $t^* = 0.002 < \alpha$ ),  $dosw2$  ( $p$ -value statystyki  $t^* = 0.000 < \alpha$ ), oraz stała ( $p$ -value statystyki  $t^* = 0.000 < \alpha$ ). Nie jest istotna statystycznie ocena parametru przy zmiennej  $po1995$  ( $p$ -value statystyki  $t^* = 0.29 > \alpha$ )
- Ponieważ wpływ doświadczenia zawodowego na wysokość zarobków mirza dwie zmienne i jest on nieliniowy, aby uzyskać interpretację ilościową należy obliczyć efekt krańcowy

$$\frac{\partial E(lzarobki | dosw, dosw2)}{\partial dosw} = \beta_{dosw} + 2\beta_{dosw2}dosw = 0.0234 - 0.0012 \times 25.9 = -0.0077$$

3. Nie ma znaczenia dla wysokości uzyskiwanych zarobków czy wykształcenie uzyskano przed czy po 1995 roku, gdyż ocena parametru okazała się być nieistotna statystycznie. P-value dla zmiennej po1995  $t^* = 0.29 > \alpha$ .
4.
  - Na podstawie wyniku testu RESET ( $0.37 > \alpha$ ) brak jest podstaw do odrzucenia hipotezy o poprawności formy funkcyjnej.
  - Na podstawie wyniku testów Breuscha-Pagana  $p$ -value ( $0.21 > \alpha$ ), oraz ( $0.28 > \alpha$ ), zatem brak jest podstaw do odrzucenia hipotezy o homoscedastyczności składnika losowego
5. W dużej próbie odrzucamy założenie o homoscedastyczności składnika losowego. Może wynikać to ze sposobu obliczenia wartości statystyki testowej, która jest czuła na niewielkie odchylenia od prawidłowych wartości.
6. Model nie spełnia założenia o stałości wariancji. Rozwiązaniem jest wykorzystanie tzw. odpornego estymatora, np. estymatora White'a albo wykorzystanie SUMNK.