



Економетрика

ЛЕКЦІЯ 5. АСПЕКТИ МНОЖИННОЇ РЕГРЕСІЇ

Д.Е.Н., ПРОФЕСОР СТАВИЦЬКИЙ А.В.

Порівняння факторів за ступенем їх впливу

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_{k-1} x_{k-1t} + \varepsilon_t, t = \overline{1, n}$$

Регресійні коефіцієнти не можна використовувати для порівняння дії різних факторів.

Використовують два методи:

- порівняння коефіцієнтів в регресії відносно стандартизованих змінних;
- порівняння коефіцієнтів еластичності.

Регресія відносно стандартизованих змінних

$$\bar{y} = \frac{\sum_{t=1}^n y_t}{n}$$

– середнє значення залежної змінної,

$$\bar{x}_j = \frac{\sum_{t=1}^n x_{jt}}{n}, j = \overline{1, k-1}$$

– середнє значення j-ї незалежної змінної,

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^n (y_t - \bar{y})^2}{n-1}}$$

– середньоквадратичне відхилення залежної змінної,

$$\sigma_{x_j} = \sqrt{\frac{\sum_{i=1}^n (x_{jt} - \bar{x}_j)^2}{n-1}}, j = \overline{1, k-1}$$

– середньоквадратичне відхилення j-ї незалежної змінної,

$$y_t^* = \frac{y_t - \bar{y}}{\sigma_y}, t = \overline{1, n}$$

– значення стандартизованої залежної змінної в t-му спостереженні

$$x_{jt}^* = \frac{x_{jt} - \bar{x}_j}{\sigma_{x_j}}, t = \overline{1, n}, j = \overline{1, k-1}$$

– значення стандартизованої j-ї незалежної змінної в t-му спостереженні.

Модель зі стандартизованими змінними

- Середнє значення всіх стандартизованих дорівнює нулю – регресія не містить константи:

$$y_t^* = \beta_1^* x_{1,t}^* + \dots + \beta_{k-1}^* x_{k-1,t}^* + \varepsilon_t, t = \overline{1, n}$$

- Оскільки середньоквадратичні відхилення мають ті самі розмірності, що і змінні, стандартизовані змінні є безрозмірними величинами, то коефіцієнти регресії можна інтерпретувати як міру впливу незалежних змінних на залежну змінну

$$\hat{\beta}_j^* = \frac{\hat{\beta}_j \sigma_{x_j}}{\sigma_y}, j = \overline{1, k-1}$$

Коефіцієнти еластичності

- Для довільної залежності виду $y_t = f(x_{1,t}, x_{2,t}, \dots, x_{k-1,t})$

коефіцієнт еластичності змінної x_j відносно y визначається як:

$$e_j = \frac{\partial(\ln f(x_1, x_2, \dots, x_{k-1}))}{\partial(\ln x_j)} = \frac{\partial f}{\partial x_j} \frac{x_j}{f(x_1, x_2, \dots, x_{k-1})}$$

- Для множинної регресії:

$$e_j = \hat{\beta}_j \frac{x_j}{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_{k-1} x_{k-1}}$$

або

$$e_j = \hat{\beta}_j \frac{\bar{x}_j}{\bar{y}}$$

Приклад

Equation: EQ01 Workfile: CHICKEN::Chicken\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: Y
Method: Least Squares
Date: 10/09/13 Time: 18:07
Sample: 1 33
Included observations: 33

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	31.40959	1.376227	22.82297	0.0000
YD	0.001839	0.000405	4.538711	0.0001
PB	0.247457	0.070428	3.513599	0.0015
PC	-0.819809	0.089305	-9.179902	0.0000

R-squared	0.963632	Mean dependent var	35.87879
Adjusted R-squared	0.959870	S.D. dependent var	9.927763
S.E. of regression	1.988782	Akaike info criterion	4.326134
Sum squared resid	114.7024	Schwarz criterion	4.507529
Log likelihood	-67.38122	Hannan-Quinn criter.	4.387168
F-statistic	256.1347	Durbin-Watson stat	0.753680
Prob(F-statistic)	0.000000		

Equation: EQ01 Workfile: CHICKEN::Chicken\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Scaled Coefficients
Date: 10/09/13 Time: 18:06
Sample: 1 33
Included observations: 33

Variable	Coefficient	Standardized Coefficient	Elasticity at Means
C	31.40959	NA	0.875436
YD	0.001839	0.490206	0.199177
PB	0.247457	0.369997	0.210589
PC	-0.819809	-0.355427	-0.285202

Фіктивні змінні

- Фіктивні, або бінарні змінні можуть приймати лише два значення: 0 та 1. Якщо певна якісна ознака спостерігається у деякий період часу, то значення фіктивної змінної дорівнює 1, у протилежному випадку – 0.
- Використовуються для моделювання якісних ознак

Наявність дискримінації в оплаті праці

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_{k-1} x_{k-1t} + \beta_k q_t + \varepsilon_t, t = \overline{1, n}$$

- де y_t – рівень заробітної плати t -го робітника,
- x_1, x_2, \dots, x_{k-1} – фактори, що впливають на рівень заробітної плати, зокрема: кількість місяців навчання, кількість років роботи на фірмі, кількість іноземних мов, які використовуються робітником тощо;

$$q_t = \begin{cases} 1, & \text{якщо } t\text{-ий працівник є чоловіком,} \\ 0, & \text{якщо } t\text{-ий працівник є жінкою.} \end{cases}$$

- Після оцінки подібної моделі для формулювання кінцевого висновку необхідно перевірити гіпотезу:

$$H_0: \beta_k = 0$$

Вартість житла

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_{k-1} x_{k-1t} + \beta_k q_{1t} + \beta_{k+1} q_{2t} + \varepsilon_t, t = \overline{1, n}$$

- де y_t – ціна квадратного метра житла;
- x_1, x_2, \dots, x_{k-1} – фактори, що впливають на ціну квартири, зокрема: її метраж, віддаленість від центру у метрах, віддаленість від найближчої станції метро у метрах, кількість кімнат, балконів тощо. Крім цього розглянемо, на якому поверсі розташована квартира;

$$q_{1t} = \begin{cases} 1, & \text{якщо квартира розташована на першому поверсі,} \\ 0, & \text{якщо квартира розташована не на першому поверсі.} \end{cases}$$

$$q_{2t} = \begin{cases} 1, & \text{якщо квартира розташована на останньому поверсі,} \\ 0, & \text{якщо квартира розташована не на останньому поверсі.} \end{cases}$$

$$H_0 : \begin{cases} \beta_k = 0, \\ \beta_{k+1} = 0. \end{cases}$$

Моделювання сезонності

$$y_t = \beta_0 + \beta_1 q_1 + \beta_2 q_2 + \beta_3 q_3 + \varepsilon_t$$

$$y_t = \beta_0 + \beta_1 q_1 + \beta_2 q_2 + \beta_3 q_3 + \beta_4 t + \varepsilon_t$$

Фіктивні змінні для квартальних даних

$$q_1 = (1, 0, 0, 0, 1, 0, 0, 0, 1, 0, \dots)'$$

$$q_2 = (0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, \dots)'$$

$$q_3 = (0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, \dots)'$$

1 квартал

$$y_t = \beta_0 + \beta_1 + \varepsilon_t$$

2 квартал

$$y_t = \beta_0 + \beta_2 + \varepsilon_t$$

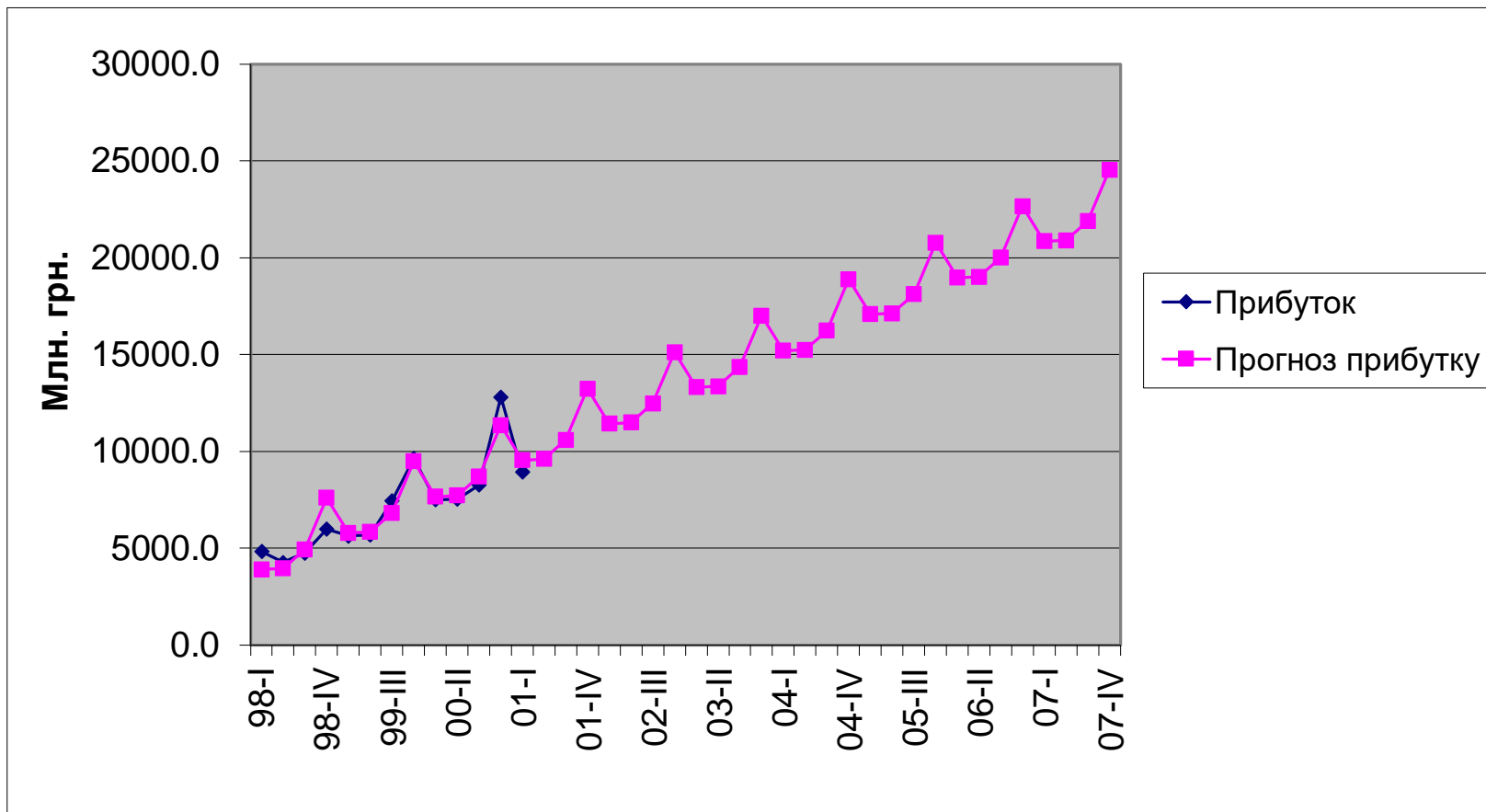
3 квартал

$$y_t = \beta_0 + \beta_3 + \varepsilon_t$$

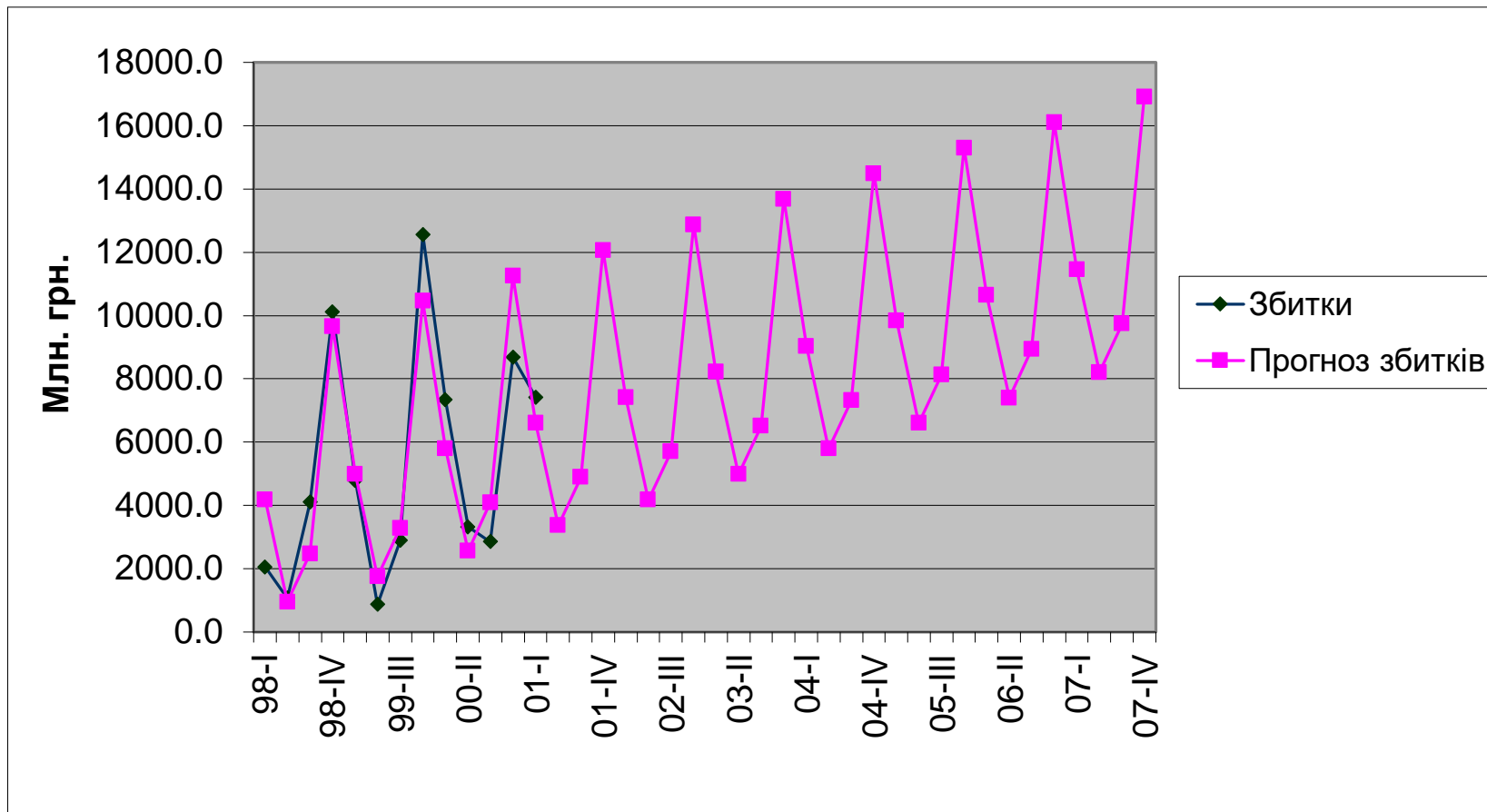
4 квартал

$$y_t = \beta_0 + \varepsilon_t$$

Приклад – I



Приклад – 2



Виділення сезонних коливань

Equation Estimation

Specification Options

Equation specification

Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like $Y=c(1)+c(2)*X$.

y c @seas(1) @seas(2) @seas(3)

Estimation settings

Method: LS - Least Squares (NLS and ARMA)

Sample: 1947q1 2005q4

OK Скасувати

Equation Estimation

Specification Options

Equation specification

Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like $Y=c(1)+c(2)*X$.

y c @seas(1) @seas(2) @seas(3) @trend

Estimation settings

Method: LS - Least Squares (NLS and ARMA)

Sample: 1947q1 2005q4

OK Скасувати

Виділення сезонних коливань

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Equation: UNTITLED Workfile: MACROMOD::Macromod\									
Dependent Variable: Y Method: Least Squares Date: 10/09/13 Time: 18:56 Sample (adjusted): 1947Q1 1999Q4 Included observations: 212 after adjustments									
Variable	Coefficient	Std. Error	t-Statistic	Prob.					
C	4059.243	270.9226	14.98303	0.0000					
@SEAS(1)	-100.1399	383.1424	-0.261365	0.7941					
@SEAS(2)	-64.37526	383.1424	-0.168019	0.8667					
@SEAS(3)	-30.88335	383.1424	-0.080605	0.9358					
R-squared	0.000365	Mean dependent var	4010.393						
Adjusted R-squared	-0.014052	S.D. dependent var	1958.633						
S.E. of regression	1972.346	Akaike info criterion	18.03052						
Sum squared resid	8.09E+08	Schwarz criterion	18.09385						
Log likelihood	-1907.235	Hannan-Quinn criter.	18.05612						
F-statistic	0.025345	Durbin-Watson stat	0.001709						
Prob(F-statistic)	0.994532								

View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Equation: UNTITLED Workfile: MACROMOD::Macromod\									
Dependent Variable: Y Method: Least Squares Date: 10/09/13 Time: 18:56 Sample (adjusted): 1947Q1 1999Q4 Included observations: 212 after adjustments									
Variable	Coefficient	Std. Error	t-Statistic	Prob.					
C	700.3215	65.86862	10.63210	0.0000					
@SEAS(1)	-5.964506	70.13732	-0.085040	0.9323					
@SEAS(2)	-1.591685	70.13147	-0.022696	0.9819					
@SEAS(3)	0.508438	70.12796	0.007250	0.9942					
@TREND()	31.39179	0.405201	77.47205	0.0000					
R-squared	0.966673	Mean dependent var	4010.393						
Adjusted R-squared	0.966029	S.D. dependent var	1958.633						
S.E. of regression	360.9997	Akaike info criterion	14.63893						
Sum squared resid	26976408	Schwarz criterion	14.71810						
Log likelihood	-1546.727	Hannan-Quinn criter.	14.67093						
F-statistic	1501.047	Durbin-Watson stat	0.014419						
Prob(F-statistic)	0.000000								

Застосування фіктивних змінних

- Виділення сезонних коливань
- Нівелювання кризових явищ
- Моделювання якісних показників
- Нівелювання пропущених значень

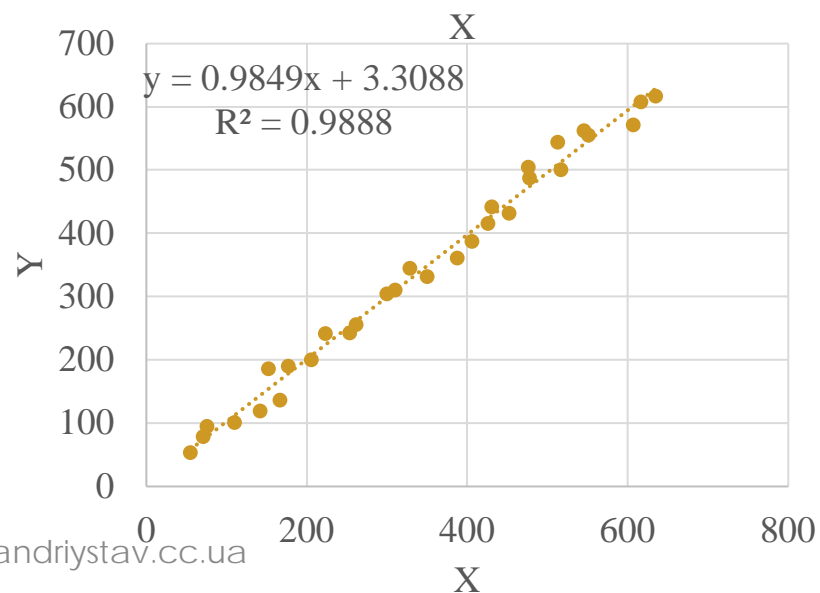
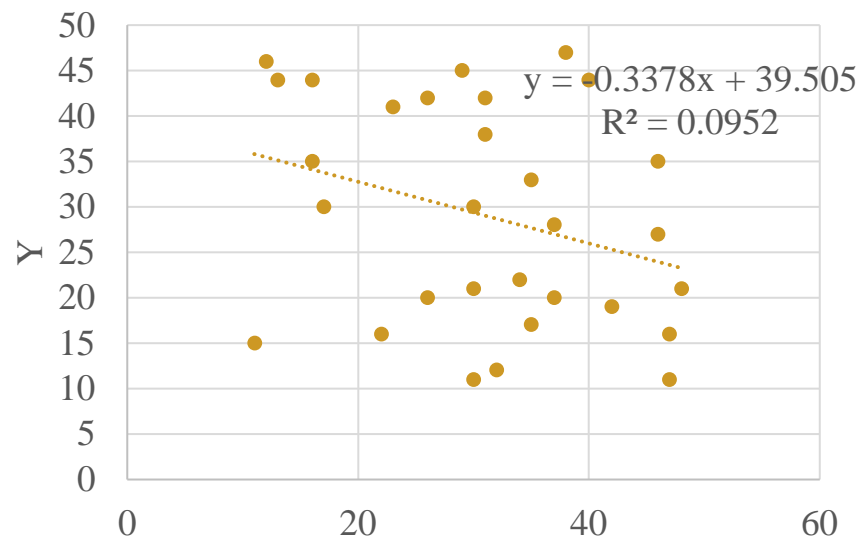
- Для відтворення в моделі впливу якісного фактора, який може приймати m рівнів, до моделі потрібно включити $m-1$ фіктивну змінну

Стаціонарний процес

- Випадковий процес $X(t)$ називається стаціонарним, якщо всі його імовірнісні характеристики не змінюються з часом t .

Чому стаціонарність важлива?

№	X	Y	T	X*	Y*
1	35	33	20	55	53
2	31	38	40	71	78
3	16	35	60	76	95
4	30	21	80	110	101
5	42	19	100	142	119
6	47	16	120	167	136
7	12	46	140	152	186
8	17	30	160	177	190
9	26	20	180	206	200
10	23	41	200	223	241
11	34	22	220	254	242
12	22	16	240	262	256
13	40	44	260	300	304
14	30	30	280	310	310
15	29	45	300	329	345
16	30	11	320	350	331
...					
30	35	17	600	635	617



Тест на стаціонарність

- У загальному випадку модель зводимо до виду

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$$

і перевіряємо гіпотезу

$$H^0 : \gamma = 0$$

$$H^1 : \gamma < 0$$

Приклад – I

Series: Y ...

View Proc Object Properties Prin

Last updated: 11/22/...

1	21.7
2	22.1
3	21.9
4	22.8
5	21.4
6	24.4
7	25.5
8	28.1
9	28.9
10	28.1
11	30.2
12	30.0
13	30.8
14	31.2
15	33.3
16	35.6
17	36.5
18	36.7
19	38.4
20	40.5
21	40.3
22	41.8
23	40.4
24	40.7
25	40.1
26	42.7
27	44.1
28	46.7
29	50.6
30	50.1
31	51.6
32	53.0
33	53.8

Series: Y ...

View Proc Object Properties Prin

- SpreadSheet
- Graph...
- Descriptive Statistics & Tests
- One-Way Tabulation...
- Correlogram...
- Long-run Variance...
- Unit Root Test...**
- Breakpoint Unit Root Test...
- Variance Ratio Test...
- BDS Independence Test...
- Forecast Evaluation...
- Label

Unit Root Test

Test type: Augmented Dickey-Fuller

Test for unit root in:

- Level
- 1st difference
- 2nd difference

Include in test equation:

- Intercept
- Trend and intercept
- None

Lag length:

- Automatic selection: Schwarz Info Criterion
- User specified: 1

Maximum lags: 8

OK Cancel

Series: Y Workfile: CHICKEN::Chicken\

View Proc Object Properties Print Name Freeze Sample Genr Sheet Graph Stat

Augmented Dickey-Fuller Unit Root Test on Y

Null Hypothesis: Y has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=8)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.508713	0.9844
Test critical values:		
1% level	-3.653730	
5% level	-2.957110	
10% level	-2.617434	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(Y)
Method: Least Squares
Date: 09/28/20 Time: 15:52
Sample (adjusted): 2 33
Included observations: 32 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	0.012625	0.024818	0.508713	0.6147
C	0.557211	0.907015	0.614335	0.5436

R-squared	0.008553	Mean dependent var	1.003125
Adjusted R-squared	-0.024496	S.D. dependent var	1.302723
S.E. of regression	1.318582	Akaike info criterion	3.451452
Sum squared resid	52.15974	Schwarz criterion	3.543061
Log likelihood	-53.22323	Hannan-Quinn criter.	3.481818
F-statistic	0.258789	Durbin-Watson stat	2.317387
Prob(F-statistic)	0.614674		

Приклад – 2

Series: Y Workfile: CHICKEN::Chicken\

View Proc Object Properties Print Name Freeze Sample Genr Sheet Graph Stat

Augmented Dickey-Fuller Unit Root Test on Y

Null Hypothesis: Y has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=8)

Unit Root Test

Test type: Augmented Dickey-Fuller

Test for unit root in: 1st difference

Lag length: Automatic selection: Schwarz Info Criterion, Maximum lags: 8

Include in test equation: Intercept

OK Cancel

R-squared	0.008553	Mean dependent var	1.003125
Adjusted R-squared	-0.024496	S.D. dependent var	1.302723
S.E. of regression	1.318582	Akaike info criterion	3.451452
Sum squared resid	52.15974	Schwarz criterion	3.543061
Log likelihood	-53.22323	Hannan-Quinn criter.	3.481818
F-statistic	0.258789	Durbin-Watson stat	2.317387
Prob(F-statistic)	0.614674		

Series: Y Workfile: CHICKEN::Chicken\

View Proc Object Properties Print Name Freeze Sample Genr Sheet Graph Stat

Augmented Dickey-Fuller Unit Root Test on D(Y)

Null Hypothesis: D(Y) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=8)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.210302	0.0000
Test critical values:		
1% level	-3.661661	
5% level	-2.960411	
10% level	-2.619160	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(Y,2)
Method: Least Squares
Date: 09/28/20 Time: 15:54
Sample (adjusted): 3 33
Included observations: 31 after adjustments

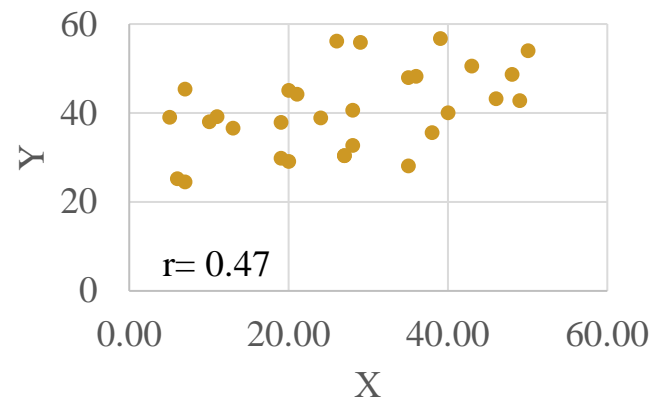
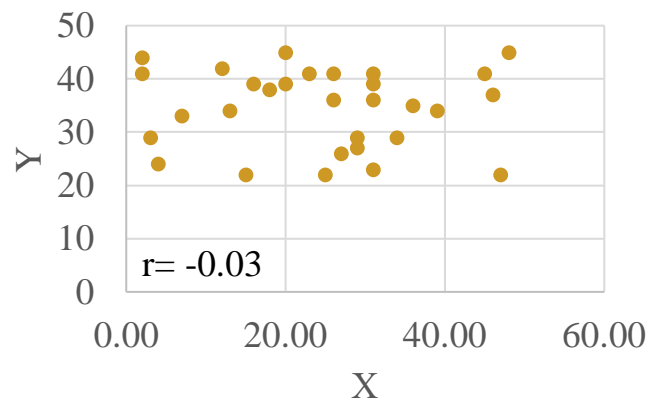
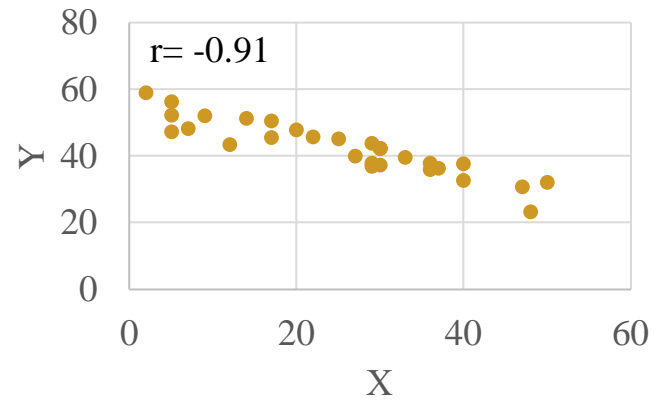
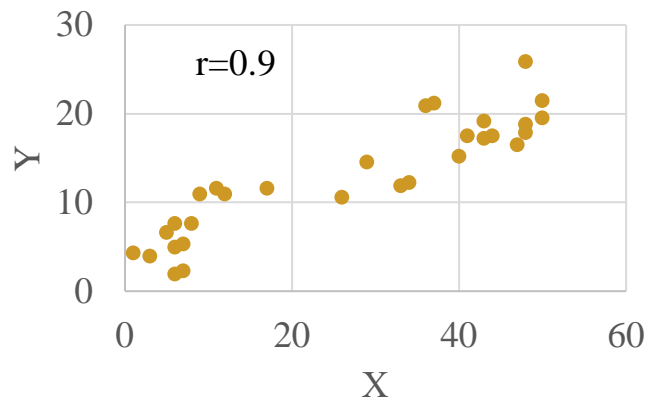
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(Y(-1))	-1.138428	0.183313	-6.210302	0.0000
C	1.162349	0.302058	3.848093	0.0006

R-squared	0.570802	Mean dependent var	0.012903
Adjusted R-squared	0.556002	S.D. dependent var	1.994616
S.E. of regression	1.329076	Akaike info criterion	3.469187
Sum squared resid	51.22688	Schwarz criterion	3.561702
Log likelihood	-51.77239	Hannan-Quinn criter.	3.499344
F-statistic	38.56785	Durbin-Watson stat	1.957202
Prob(F-statistic)	0.000001		

Вибір моделей на основі коефіцієнта детермінації

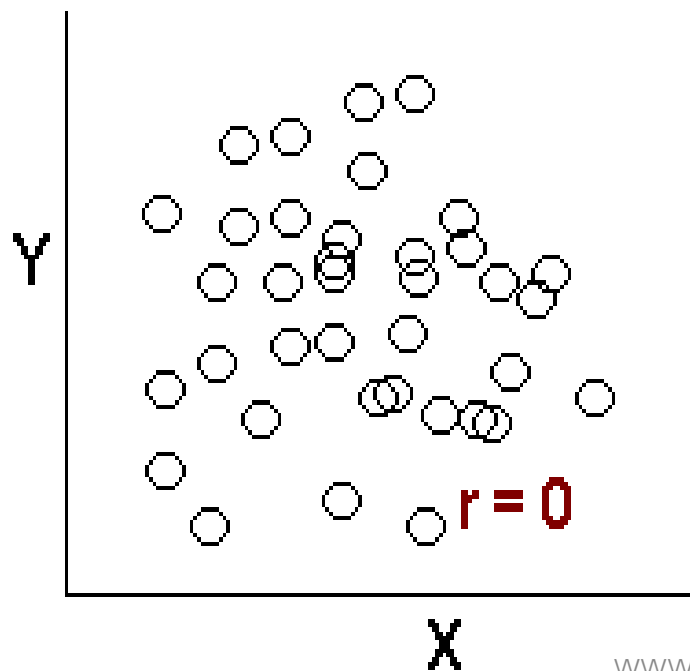
- В основному, порівнюємо скоригований коефіцієнт детермінації.
- Він пропорційний коефіцієнту кореляції.

Приклад



Але!

- Створюється враження, що близький до нуля коефіцієнт кореляції говорить про те, що зв'язки між змінними немає або майже немає.

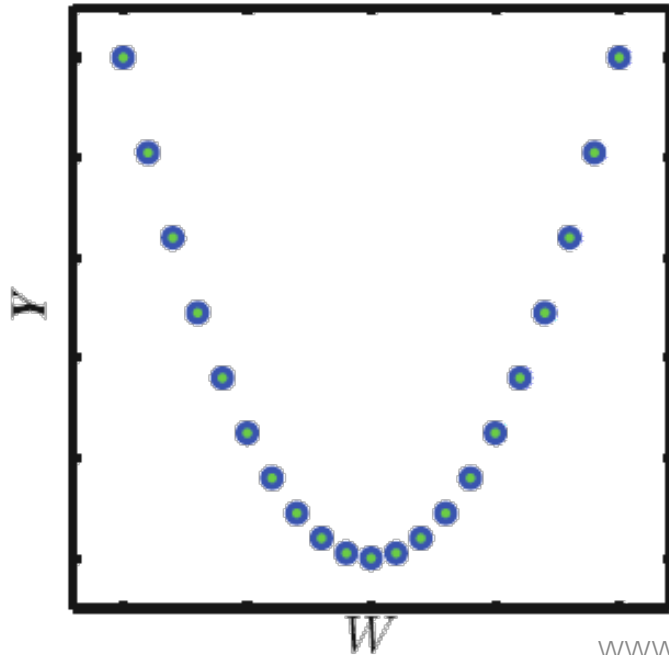


Тут її немає.

Але це не завжди так,
є винятки!

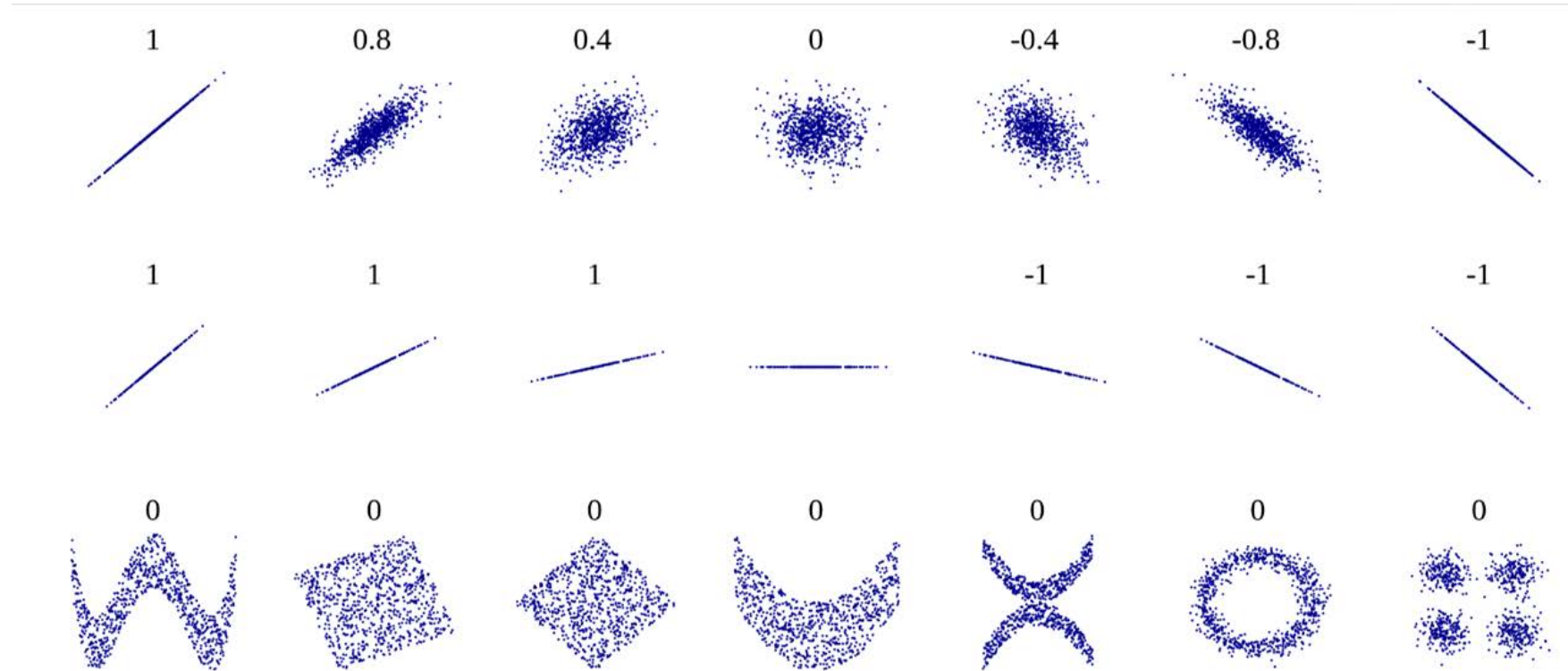
Недолік коефіцієнта кореляції

- Коефіцієнт кореляції оцінює тільки лінійний зв'язок змінних!
- Він не показує наявності нелінійного зв'язку!



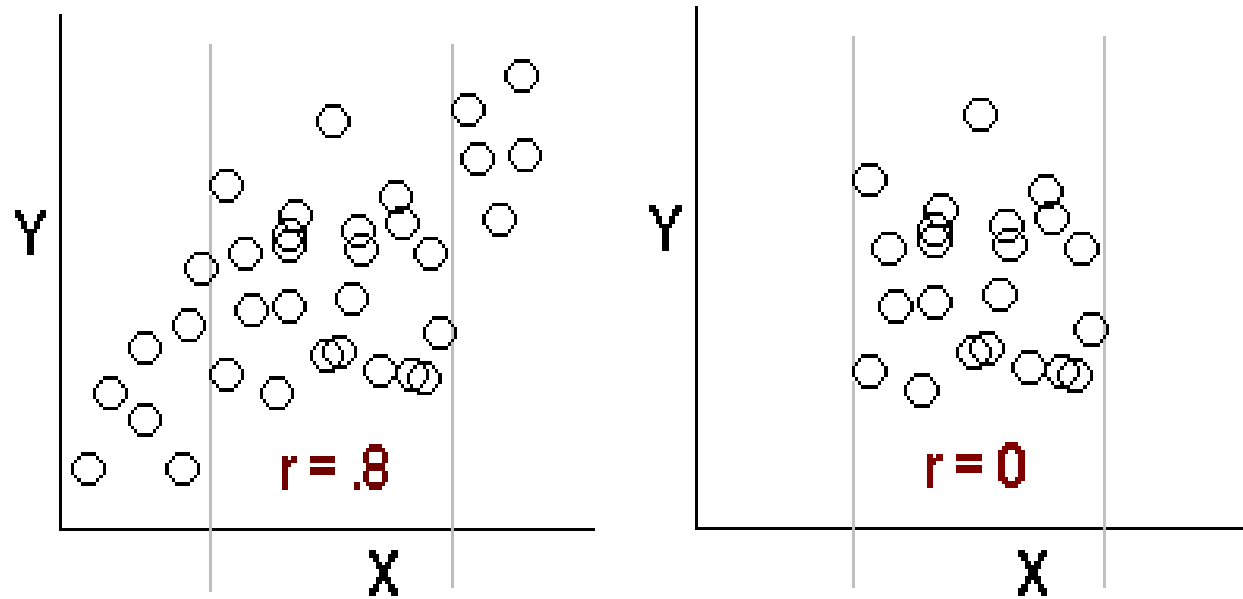
Тут зв'язок змінних
дуже сильний, але
 $r = 0.00$

Приклад

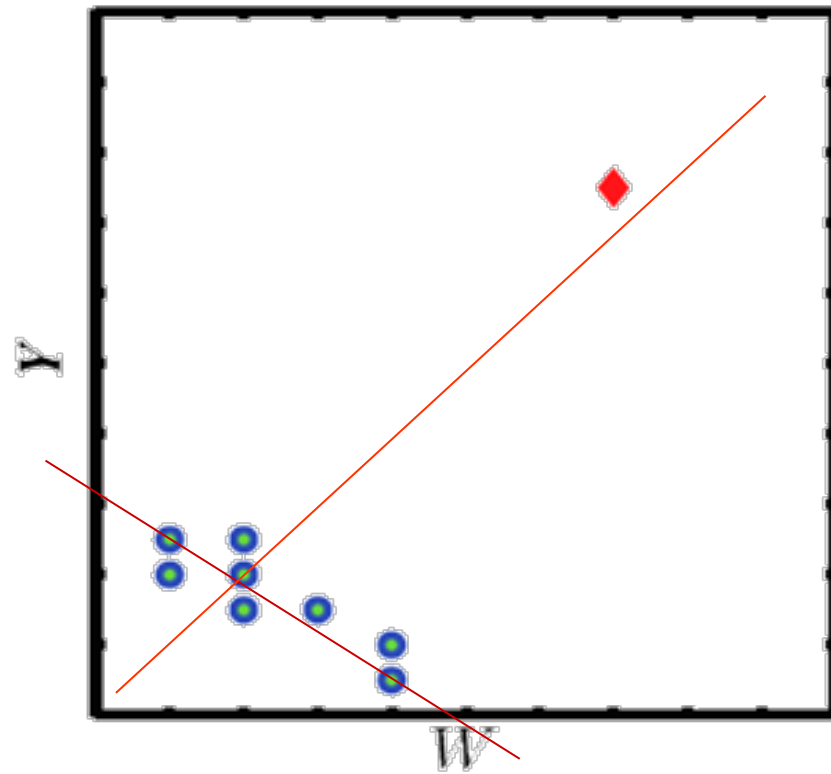


Пам'ятаємо!

- Необхідно, щоб в змінних була значна мінливість! Якщо сформулювати вибірку з однотипних осіб, сподіватися виявити там кореляцію не варто.

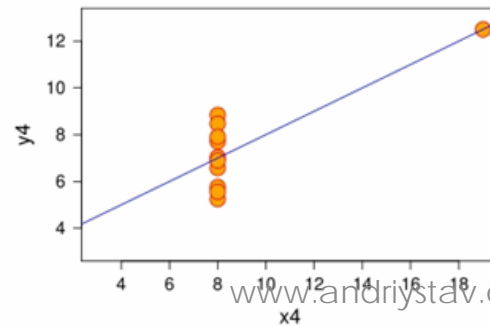
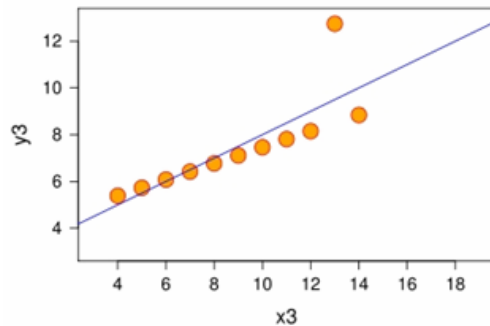
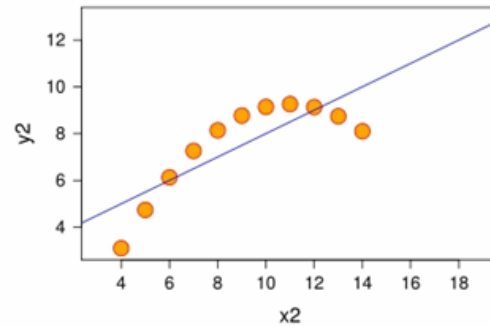
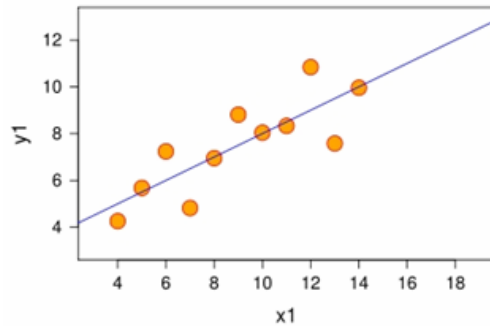


Викид




Рішення про залежність

- Коефіцієнт кореляції - параметр вибірки. Чи можна на його основі судити про популяції?
- Просто дивлячись на коефіцієнт - Ні.



Кореляція між
змінними всюди
 $r = 0.816$



Питання?