Економетрика

ЛЕКЦІЯ 2. ПРОСТА ЛІНІЙНА РЕГРЕСІЯ

Д.Е.Н., ПРОФЕСОР СТАВИЦЬКИЙ А.В.

Моделі простої регресії

Рівень доходу → Споживання

Ціна товару → Попит

Розмір основних фондів → Обсяг виробництва

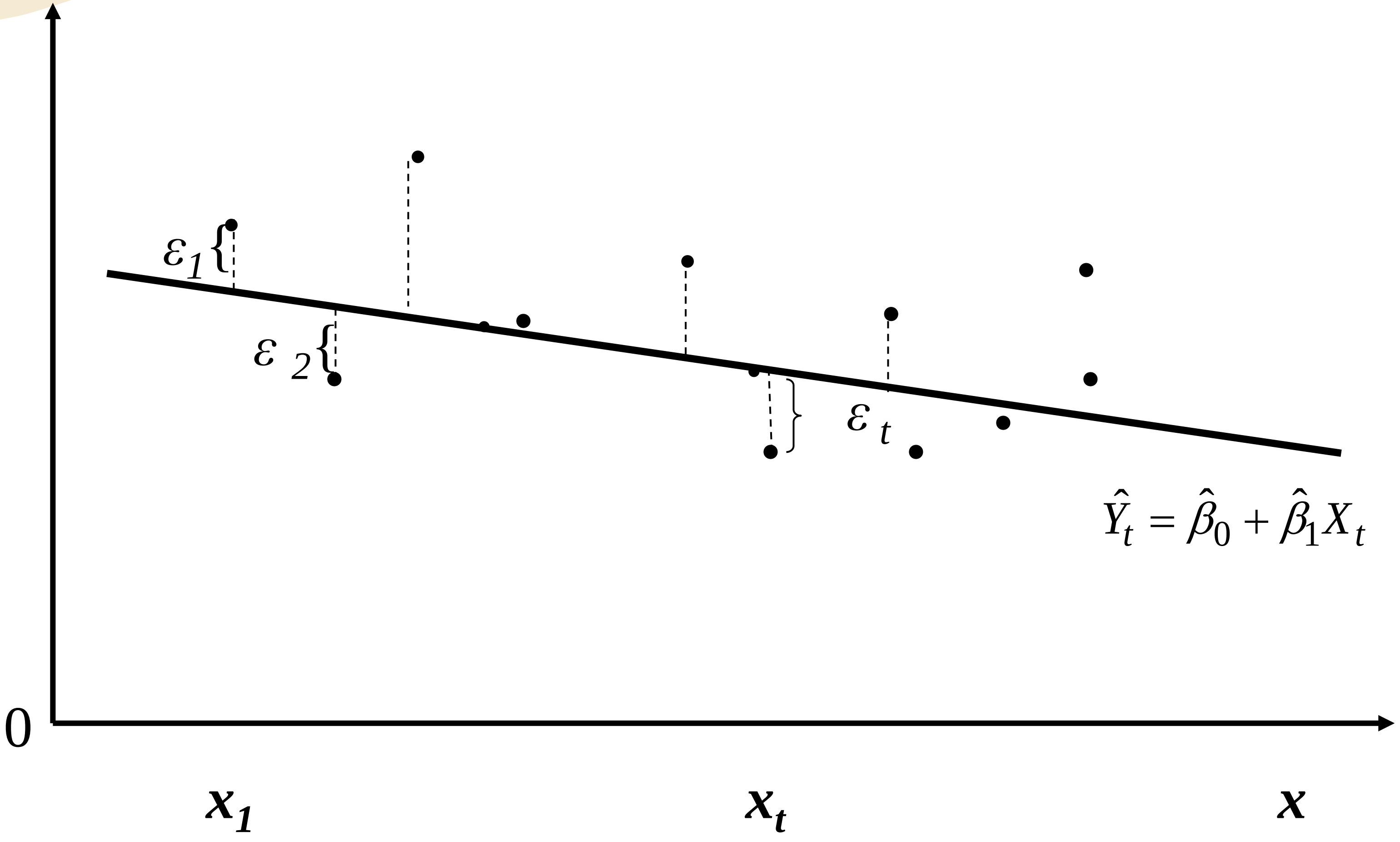
Опис моделі

- Нехай існує дві змінні X та Y, між якими необхідно встановити зв'язок.
- Дані підготовлені (прибрані інфляція, випадкові події тощо)
- Питання стаціонарності

Дані

x_1	x_2	x_3	$\cdots$	x_{n-1}	x_n
y_1	y_2	y_3	$\cdots$	y_{n-1}	y_n

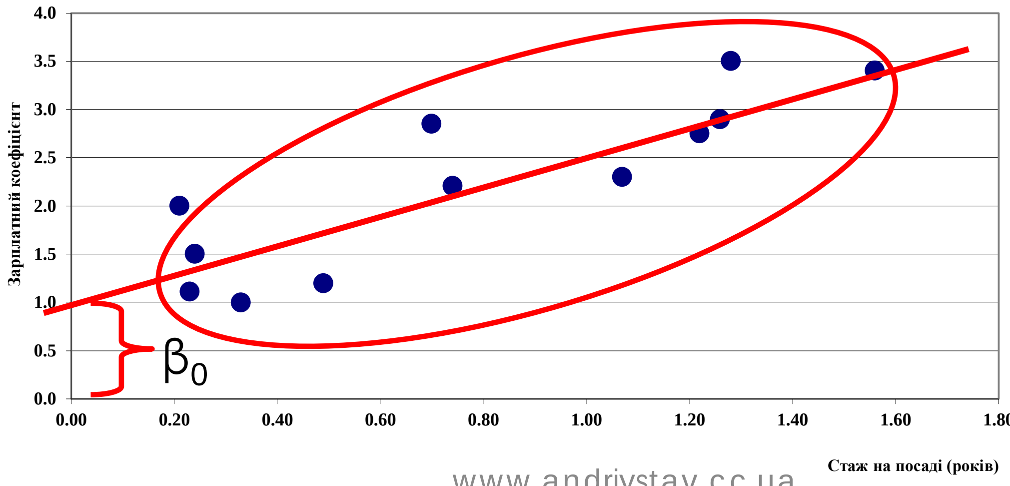
Помилки (залишки, збурення, шоки) моделі



Проста лінійна регресія

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

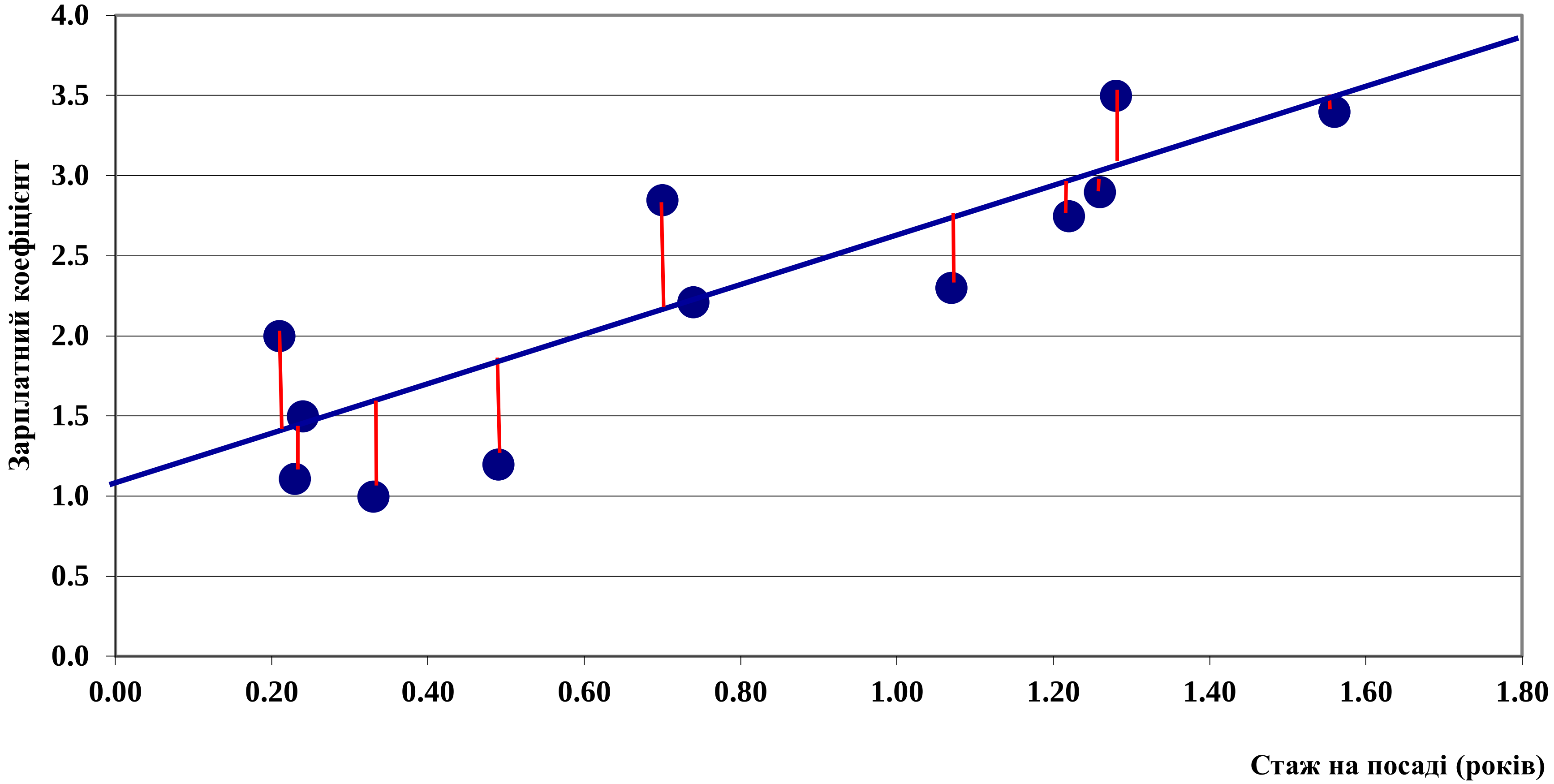
- Y – залежна змінна
- X – незалежна змінна
- β_0 і β_1 – коефіцієнти регресії
- β_1 – характеризує НАХИЛ прямої;
- β_0 – визначає точку перетину прямої з віссю OY.



«Найкраща» лінія регресії

$$\varepsilon_t = Y_t - \widehat{Y}_t$$

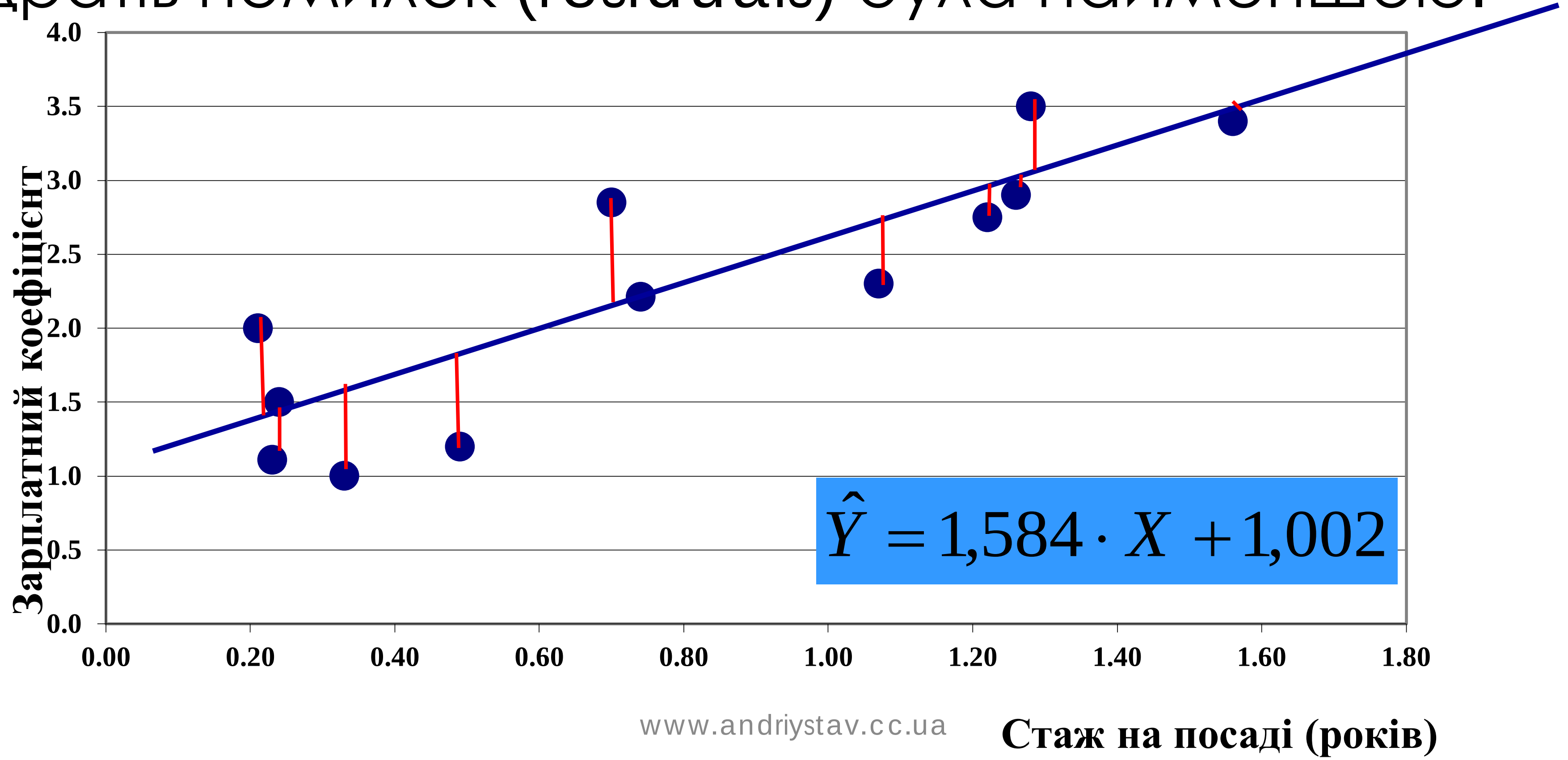
$$\sum \varepsilon_t = 0$$



Метод найменших квадратів

$$\sum \varepsilon_t^2 \rightarrow \min$$

- лінію регресії підбирають таку, щоб загальна сума квадратів помилок (residuals) була найменшою.



Припущення щодо збурень

1. Нульове середнє: $M\varepsilon_t = 0, t = \overline{1, n}$
2. Рівність дисперсій (гомоскедастичність): $D\varepsilon_t = M\varepsilon_t^2 = \sigma^2 = const, t = \overline{1, n}$
3. Незалежність збурень: $\mathrm{cov}(\varepsilon_t, \varepsilon_\tau) = M\varepsilon_t\varepsilon_\tau = 0, t \neq \tau$
4. Незалежність збурень та регресора: $\mathrm{cov}(\varepsilon_t, x_t) = 0, \forall t$
5. Нормальність збурень: $\varepsilon_t \sim N(0, \sigma^2)\ \forall t$

Знаходження оцінок параметрів регресії методом найменших квадратів

- Якщо знайти оцінки коефіцієнтів, то рівняння регресії має вигляд:

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$$

- Помилки моделі:

$$\varepsilon_t = y_t - \hat{y}_t$$

- Оптимізаційна функція:

$$Q = \sum_{t=1}^{n} \varepsilon_t^2 = \sum_{t=1}^{n} \left(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t \right)^2 \to \min$$

МНК

$$\begin{cases} \dfrac{\partial Q}{\partial \hat{\beta}_0} = \sum\limits_{t=1}^{n} 2(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)(-1) = 0, \\ \dfrac{\partial Q}{\partial \hat{\beta}_1} = \sum\limits_{t=1}^{n} 2(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)(-x_t) = 0. \end{cases}$$

$$\begin{cases} \sum\limits_{t=1}^{n} y_t = n\hat{\beta}_0 + \hat{\beta}_1 \sum\limits_{t=1}^{n} x_t, \\ \sum\limits_{t=1}^{n} y_t x_t = \hat{\beta}_0 \sum\limits_{t=1}^{n} x_t + \hat{\beta}_1 \sum\limits_{t=1}^{n} x_t^2. \end{cases}$$

- За методом Крамера

$$\hat{\beta}_1 = \frac{n\sum\limits_{t=1}^{n} x_t y_t - \sum\limits_{t=1}^{n} x_t \sum\limits_{t=1}^{n} y_t}{n\sum\limits_{t=1}^{n} x_t^2 - \left(\sum\limits_{t=1}^{n} x_t\right)^2}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

- **Графік регресійної прямої проходить через точку середніх значень залежної та незалежної змінних.**

Незміщеність оцінок регресії

$$M\hat{\beta}_1 = M\frac{n\sum x_t y_t - \left(\sum x_t\right)\left(\sum y_t\right)}{n\sum x_t^2 - \left(\sum x_t\right)^2} = M\frac{\sum\left(x_t - \overline{x}\right)\left(y_t - \overline{y}\right)}{\sum\left(x_t - \overline{x}\right)^2}$$

$$= \frac{\sum\left(x_t - \overline{x}\right)M\left(y_t - \overline{y}\right)}{\sum\left(x_t - \overline{x}\right)^2} = \frac{\sum\left(x_t - \overline{x}\right)\beta_1\left(x_t - \overline{x}\right)}{\sum\left(x_t - \overline{x}\right)^2} = \beta_1$$

$$M\hat{\beta}_0 = M\left(\frac{1}{n}\sum y_t - \frac{1}{n}\sum x_t\hat{\beta}_1\right) = M\overline{y} - \overline{x}M\hat{\beta}_1 = \left(\beta_0 + \beta_1\overline{x}\right) - \beta_1\overline{x} = \beta_0$$

Теорема Гауса-Маркова

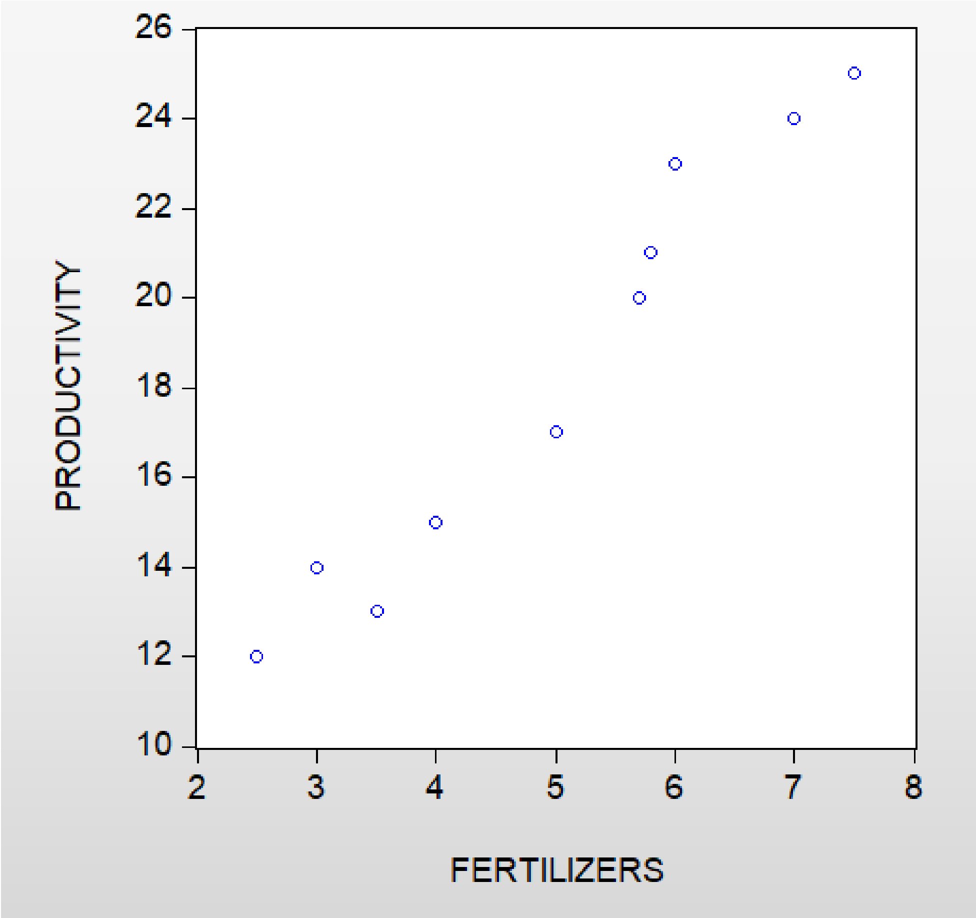
- Для простої лінійної регресії з гомоскедастичними, некорельованими збуреннями оцінки МНК мають найменшу дисперсію в класі всіх лінійних незміщених оцінок.

Приклад – 1

Group: UNTITLED Workfile...

View | Proc | Object | Print | Name | Freeze | Default

	FERTILIZERS	PRODUCTI...
1	4.0	15
2	2.5	12
3	5.0	17
4	5.8	21
5	7.5	25
6	5.7	20
7	7.0	24
8	3.0	14
9	6.0	23
10	3.5	13



Приклад – 2

Equation: EQ01   Workfile: EXAMPLE_01::Example_01\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

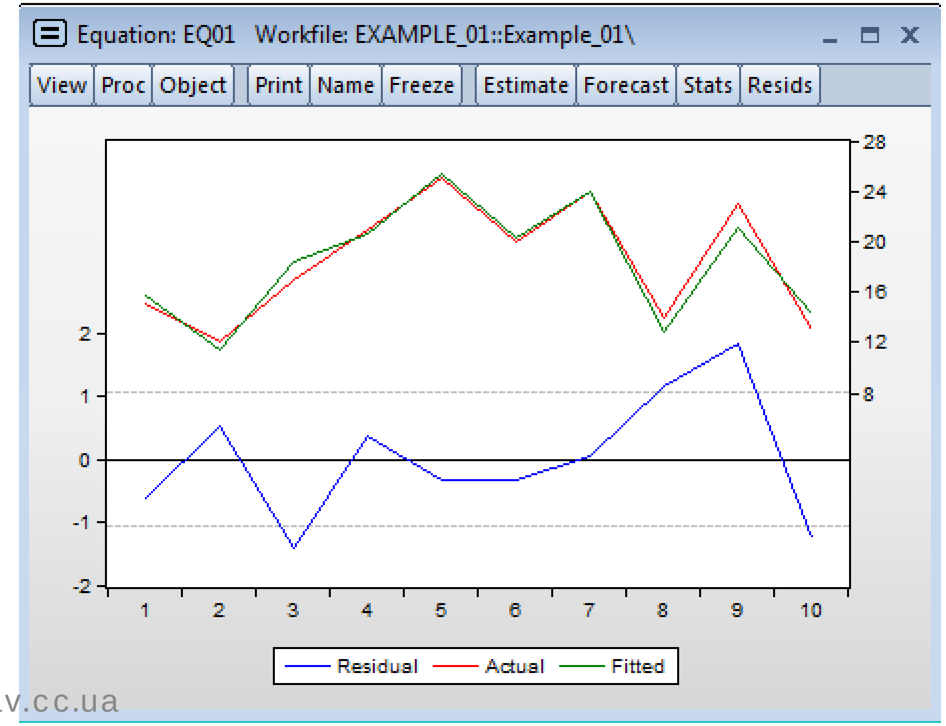
Dependent Variable: PRODUCTIVITY
Method: Least Squares
Date: 09/06/12   Time: 12:29
Sample: 1 10
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.528284	1.107332	4.089364	0.0035
FERTILIZERS	2.774343	0.210823	13.15959	0.0000

R-squared	0.[illegible]5584[illegible]	Mean dependent var	18.40000
Adjusted R-squared	0.950324	S.D. dependent var	4.812022
S.E. of regression	1.072507	Akaike info criterion	3.154731
Sum squared resid	9.202164	Schwarz criterion	3.215248
Log likelihood	-13.77365	Hannan-Quinn criter.	3.088344
F-statistic	173.1748	Durbin-Watson stat	2.171922
Prob(F-statistic)	0.000001		

$$\hat{y} = 4,528 + 2,774x$$

збільшення x на 1 призводить до збільшення y на 2,77.



Формула розкладу дисперсії

$$TSS = ESS + RSS$$

- $TSS = S_{yy} = \sum_{t=1}^{n} (y_t - \bar{y})^2$ – загальна сума квадратів,
- $ESS = \hat{\beta}^2 S_{xx} = \sum_{t=1}^{n} (\hat{y}_t - \bar{\hat{y}})^2$ – пояснена сума квадратів,
- $RSS = \sum_{t=1}^{n} \hat{\varepsilon}_t^2$ – сума квадратів залишків.

Коефіцієнт детермінації

- Коефіцієнт детермінації є частиною дисперсії залежної змінної, яка пояснюється за рахунок моделі, або, іншими словами, завдяки мінливості незалежної змінної. Коефіцієнт детермінації є мірою тісноти **саме лінійного зв'язку** між x та y.

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$R^2 = \frac{\hat{\beta}^2 S_{xx}}{S_{yy}} = \frac{\hat{\beta} S_{xy}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

- Чим ближчий R^2 до 1, тим точніше x пояснює y.
- $0 \le R^2 \le 1$

Приклад

Equation: EQ01 Workfile: EXAMPLE_01::Example_01\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PRODUCTIVITY
Method: Least Squares
Date: 09/06/12 Time: 12:29
Sample: 1 10
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.528284	1.107332	4.089364	0.0035
FERTILIZERS	2.774343	0.210823	13.15959	0.0000

R-squared	0.955844	Mean dependent var	18.40000
Adjusted R-squared	0.950324	S.D. dependent var	4.812022
S.E. of regression	1.072507	Akaike info criterion	3.154731
Sum squared resid	9.202164	Schwarz criterion	3.215248
Log likelihood	-13.77365	Hannan-Quinn criter.	3.088344
F-statistic	173.1748	Durbin-Watson stat	2.171922
Prob(F-statistic)	0.000001		

Хоча наша модель матиме високий рівень якості, проте перевірити її адекватність можна лише за допомогою спеціального тесту.

Перевірка адекватності регресії

$$H_0 : R^2 = 0 \text{ (модель неадекватна)},$$

$$H_1 : R^2 \neq 0 \text{ (модель адекватна)}.$$

$$F_{pr} = \frac{\dfrac{R^2}{1}}{\dfrac{1-R^2}{n-2}} \sim F(1-\alpha, 1, n-2)$$

Equation: EQ01 Workfile: EXAMPLE_01::Example_01\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PRODUCTIVITY
Method: Least Squares
Date: 09/06/12 Time: 12:29
Sample: 1 10
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.528284	1.107332	4.089364	0.0035
FERTILIZERS	2.774343	0.210823	13.15959	0.0000

R-squared	0.955844	Mean dependent var	18.40000
Adjusted R-squared	0.950324	S.D. dependent var	4.812022
S.E. of regression	1.072507	Akaike info criterion	3.154731
Sum squared resid	9.202164	Schwarz criterion	3.215248
Log likelihood	-13.77365	Hannan-Quinn criter.	3.088344
F-statistic	173.1748	Durbin-Watson stat	2.171922
Prob(F-statistic)	0.000001		

Table entry for p is the critical value F^* with probability p lying to its right.

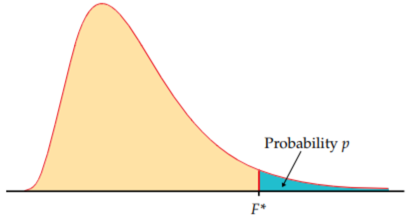


TABLE E

F critical values

Degrees of freedom in the denominator	p	Degrees of freedom in the numerator: 1	2	3	4	5	6	7	8	9
1	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
	.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
	.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
	.001	405284	500000	540379	562500	576405	585937	592873	598144	602284
2	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
3	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
4	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
5	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24
6	.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69
7	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33
8	.100	[illegible]	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	.025	[illegible]	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
	.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77

Перевірка гіпотез про значимість коефіцієнтів регресії

$$H_0:\ \beta = 0$$
$$H_1:\ \beta \neq 0$$

$$t_{pr} = \left| \frac{\hat{\beta}}{\mathbf{s.e.}(\hat{\beta})} \right| \sim t(1-\alpha, n-2)$$

Equation: EQ01 Workfile: EXAMPLE_01::Example_01\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PRODUCTIVITY
Method: Least Squares
Date: 09/06/12 Time: 12:29
Sample: 1 10
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.528284	1.107332	4.089364	0.0035
FERTILIZERS	2.774343	0.210823	13.15959	0.0000

R-squared	0.955844	Mean dependent var	18.40000
Adjusted R-squared	0.950324	S.D. dependent var	4.812022
S.E. of regression	1.072507	Akaike info criterion	3.154731
Sum squared resid	9.202164	Schwarz criterion	3.215248
Log likelihood	-13.77365	Hannan-Quinn criter.	3.088344
F-statistic	173.1748	Durbin-Watson stat	2.171922
Prob(F-statistic)	0.000001		

***t* Table**

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
					Confidence Level						

Перевірка гіпотези про значення коефіцієнта регресії

$$H_0 : \beta = m$$

$$H_1 : \beta \neq m$$

$$t_{pr} = \left| \frac{\hat{\beta} - m}{\mathbf{s.e.}(\hat{\beta})} \right| \sim t\left(1 - \alpha, n - 2\right)$$

Надійні інтервали для коефіцієнтів

$$[\hat{\beta}_0 - \mathbf{s.e.}(\hat{\beta}_0)\cdot t_{\mathrm{teor}};\ \hat{\beta}_0 + \mathbf{s.e.}(\hat{\beta}_0)\cdot t_{\mathrm{teor}}]$$

$$[\hat{\beta}_1 - \mathbf{s.e.}(\hat{\beta}_1)\cdot t_{\mathrm{teor}};\ \hat{\beta}_1 + \mathbf{s.e.}(\hat{\beta}_1)\cdot t_{\mathrm{teor}}]$$

Equation: EQ01 Workfile: EXAMPLE_01::Example_01\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Coefficient Confidence Intervals
Date: 08/29/20 Time: 13:03
Sample: 1 10
Included observations: 10

		90% CI		95% CI		99% CI	
Variable	Coefficient	Low	High	Low	High	Low	High
FERTILIZERS	2.774343	2.382308	3.166378	2.288185	3.260502	2.066951	3.481736
C	4.528284	2.469147	6.587422	1.974772	7.081797	0.812756	8.243813

Коефіцієнт еластичності

- Середній коефіцієнт еластичності показує, на скільки відсотків в середньому зміниться значення y при зміні незалежної змінної x на 1% від свого середнього значення:

$$\bar{E} = f'(x)\frac{\bar{x}}{\bar{y}} = \hat{\beta}\frac{\bar{x}}{\bar{y}}$$

Equation: EQ01 Workfile: EXAMPLE_01::Example_01\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Scaled Coefficients
Date: 08/29/20 Time: 13:08
Sample: 1 10
Included observations: 10

Variable	Coefficient	Standardized Coefficient	Elasticity at Means
FERTILIZERS	2.774343	0.977673	0.753898
C	4.528284	NA	0.246102

Прогнозування за допомогою простої лінійної регресії

- Прогноз $$\hat{y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}$$

- Середня стандартна помилка прогнозу дорівнює:

$$\mathbf{s.e.}(\hat{y}_{n+1}) = \sqrt{\frac{RSS}{n-2}\left(1+\frac{1}{n}+\frac{(x_{n+1}-\overline{x})^2}{\sum_{t=1}^{n}\left(x_t-\overline{x}\right)^2}\right)}$$

- Надійний інтервал для прогнозу:

$$[\hat{y}_{n+1} - \mathbf{s.e.}(\hat{y}_{n+1}) \cdot t_{\text{teor}} ; \hat{y}_{n+1} + \mathbf{s.e.}(\hat{y}_{n+1}) \cdot t_{\text{teor}}]$$

$$t_{teor} = t\left(1-\alpha, n-2\right)$$

Абсолютні критерії точності прогнозів

$$MSE = \frac{1}{p}\sum_{t}\left(y_t - \hat{y}_t\right)^2$$

Середньоквадратична похибка прогнозу за p періодів.

$$RMSE = \sqrt{\frac{1}{p}\sum_{t}\left(y_t - \hat{y}_t\right)^2}$$

Корінь з середньоквадратичної похибки прогнозу за p періодів.

$$MAD = \frac{1}{p}\sum_{t}\left|y_t - \hat{y}_t\right|$$

Середня абсолютна похибка за p періодів.

Відносні критерії точності прогнозів

$$RMSPE = 100\sqrt{\frac{1}{p}\sum_t \left(\frac{y_t - \hat{y}_t}{y_t}\right)^2}$$

корінь з середньоквадратичної похибки у відсотках від фактичних значень за p періодів.

$$MAPE = \frac{100}{p}\sum_t \left|\frac{y_t - \hat{y}_t}{y_t}\right|$$

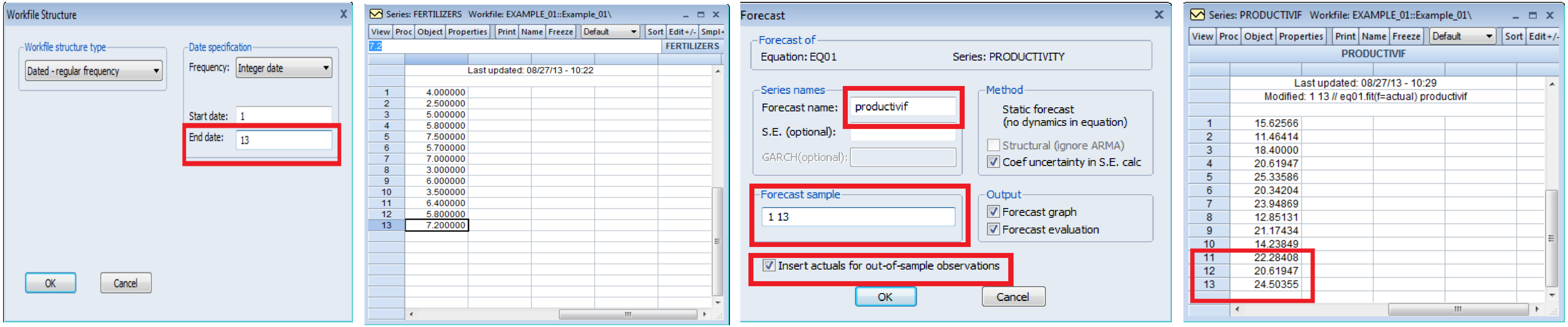
середня абсолютна похибка у відсотках за p періодів.

Оцінка точності мікроекономічного прогнозу

MAPE, RMSPE	**Точність прогнозу**
менше 5%	Висока
5% – 10%	Добра
10% – 15%	Задовільна
15% – 20%	Погана
більше 20%	Незадовільна

Приклад

- Нехай відомі значення залежної змінної на наступні 3 періоди (fertilizers): $x_{11}=6{,}4$, $x_{12}=5{,}8$, $x_{13}=7{,}2$. Необхідно знайти прогноз врожайності на ці періоди.



$$RMSPE = 100\sqrt{\frac{1}{p}\sum_t\left(\frac{y_t-\hat{y}_t}{y_t}\right)^2}$$

$$=100\sqrt{\frac{1}{3}\left(\left(\frac{20{,}9-22{,}28}{20{,}9}\right)^2+\left(\frac{21{,}5-20{,}62}{20{,}62}\right)^2+\left(\frac{24{,}2-24{,}50}{24{,}2}\right)^2\right)}=4{,}595$$

Моделі, що зводяться до простої лінійної регресії – 1

- Для моделювання залежності індивідуального споживання C від наявного прибутку У Кейнс запропонував лінійне рівняння

$$C = c_0 + bY$$

де c_0 - величина автономного споживання;

- *b* - гранична схильність до споживання ($0<b\leq 1$).

Моделі, що зводяться до простої лінійної регресії – 2

- Залежність між рівнем безробіття x і рівнем інфляції y відображається так званою кривою Філіпса:

$$y = \frac{a}{x - b}$$

де $a > 0$, $b > 0$ - параметри моделі, а змінні x і y вимірюються у процентах.

Моделі, що зводяться до простої лінійної регресії – 3

- При маркетингових і ринкових дослідженнях, при дослідженні збуту продукції та в демографії застосовують так звану криву Гомперця:

$$y = e^{ab^{x}+c}$$

де параметри *a* та *c* можуть набувати будь-яких значень, а *b* перебуває в таких межах: $0 < b < 1$.

Приклад

$$y = \beta_0 e^{\beta_1 x} + \varepsilon$$

- Для приведення цієї моделі до простої лінійної регресії необхідно провести наступні заміни:

1. Прологарифмувати обидві частини рівняння:

$$\ln y = \ln \beta_0 + \beta_1 x$$

2. Перепозначити отримані величини:

$$y^* = \ln y;$$

$$\beta_0^* = \ln \beta_0$$

3. Записати рівняння простої лінійної регресії:

$$y^* = \beta_0^* + \beta_1 x + \varepsilon$$

Приклад розв’язання задачі: умова

На основі статистичних даних доходу підприємства y (у млн. грн.) та кількості працюючих x (у тис. чол.):

1. Знайти оцінки параметрів простої лінійної регресії.
2. Перевірити модель на адекватність з рівнем надійності 95%.
3. Визначити значимість коефіцієнту нахилу регресії з рівнем надійності 95%.
4. Визначити надійні інтервали для коефіцієнтів регресії з рівнем надійності 95%.
5. Обчислити середній коефіцієнт еластичності.

Y	X
10,8	2,53
11,9	3,54
12,4	3,84
13,2	3,84
14,1	4,22
15,2	4,81
16,0	6,53
17,4	5,82
18,6	6,43
19,4	7,73
20,5	8,19
21,3	7,65
22,5	9,31
23,7	9,26
25,0	9,86

Приклад розв'язання задачі: розв'язок – 1

- Знаходимо оцінки регресії

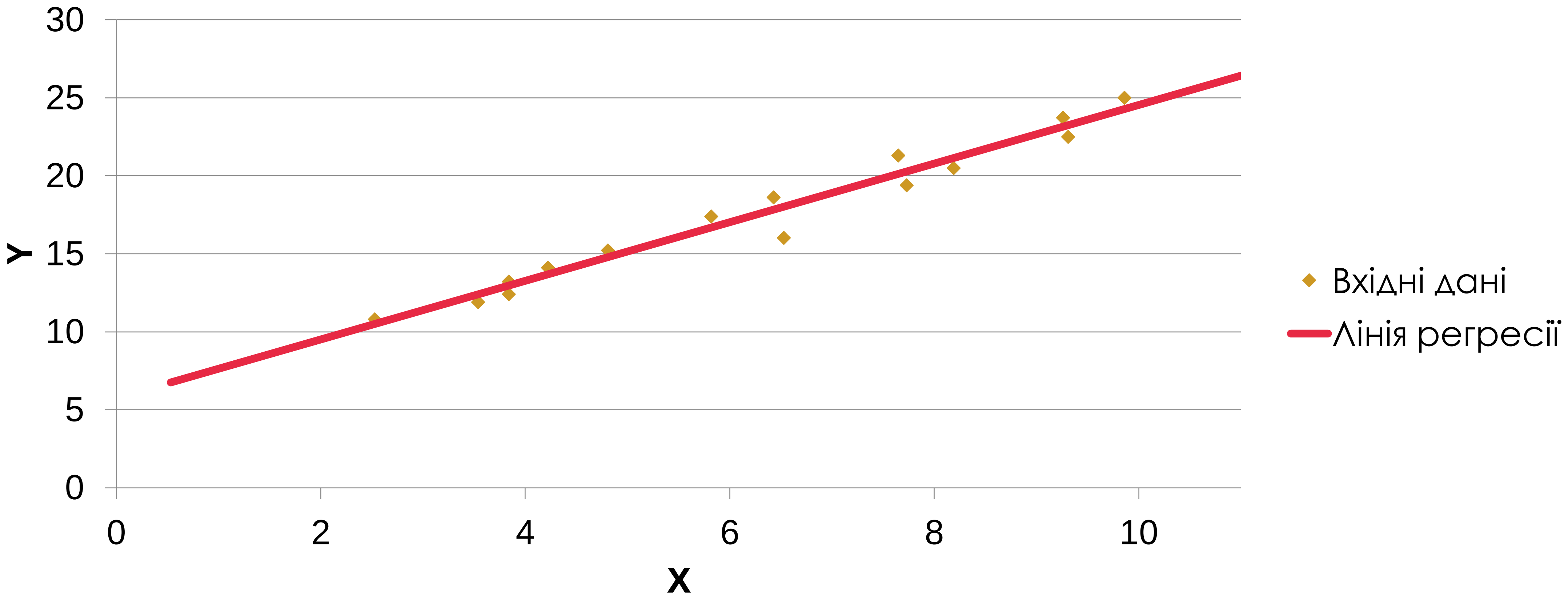
	Y	X	X²	Y²	XY
	10,8	2,53	6,40	116,64	27,32
	11,9	3,54	12,53	141,61	42,13
	12,4	3,84	14,75	153,76	47,62
	13,2	3,84	14,75	174,24	50,69
	14,1	4,22	17,81	198,81	59,50
	15,2	4,81	23,14	231,04	73,11
	16,0	6,53	42,64	256,00	104,48
	17,4	5,82	33,87	302,76	101,27
	18,6	6,43	41,34	345,96	119,60
	19,4	7,73	59,75	376,36	149,96
	20,5	8,19	67,08	420,25	167,90
	21,3	7,65	58,52	453,69	162,95
	22,5	9,31	86,68	506,25	209,48
	23,7	9,26	85,75	561,69	219,46
	25,0	9,86	97,22	625,00	246,50
Сума	**262,00**	**93,56**	**662,22**	**4864,06**	**1781,95**

$$\hat{\beta}_1 = \frac{n\sum_{t=1}^{n} x_t y_t - \sum_{t=1}^{n} x_t \sum_{t=1}^{n} y_t}{n\sum_{t=1}^{n} x_t^2 - (\sum_{t=1}^{n} x_t)^2} = \frac{15 \cdot 1781,95 - 93,56 \cdot 262}{15 \cdot 662,22 - (93,56)^2} = 1,8787$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{\sum y_t}{n} - \hat{\beta}_1 \frac{\sum x_t}{n} = \frac{262}{15} - 1,8787\frac{93,56}{15} = 5,7486$$

Побудована регресія

$$\hat{y} = 5,7486 + 1,8787x$$



Приклад розв'язання задачі: розв'язок – 2

- Заповнюємо таблицю

$\hat{y}_t$	$(y_t - \hat{y}_t)^2$	$(\hat{y}_t - \bar{y})^2$	$(y_t - \bar{y})^2$
10,502	0,089	48,511	44,444
12,399	0,249	25,679	30,988
12,963	0,317	20,285	25,671
12,963	0,056	20,285	18,204
13,677	0,179	14,364	11,334
14,785	0,172	7,191	5,138
18,017	4,066	0,302	2,151
16,683	0,515	0,615	0,004
17,829	0,595	0,131	1,284
20,271	0,759	7,864	3,738
21,135	0,403	13,458	9,201
20,121	1,391	7,044	14,694
23,239	0,547	33,323	25,334
23,145	0,308	32,248	38,854
24,273	0,529	46,320	56,751
Сума	**10,175**	**277,618**	**287,793**

- Таким чином

$$RSS = 10{,}175 \qquad ESS = 277{,}618 \qquad TSS = 287{,}793$$

Приклад розв'язання задачі: розв'язок – 2

- Коефіцієнт детермінації моделі становить

$$R^2 = \frac{ESS}{TSS} = \frac{277,618}{287,793} = 0,965$$

- Перевіримо модель на адекватність. Практичне значення статистики Фішера дорівнює

$$F_{pr} = \frac{\frac{R^2}{1}}{\frac{1-R^2}{n-2}} = 354,71,$$

- Теоретичне

$$F_{teor} = F\left(0,95;1;13\right) = 4,7$$

- Таким чином модель є адекватною.

Приклад розв'язання задачі: розв'язок – 3

$$H_0 : \beta_1 = 0$$

$$s.e.\left(\hat{\beta}_1\right) = \sqrt{\sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{\sum\left(x_t - \overline{x}\right)^2}\right)} = \sqrt{\frac{RSS}{n-2}\left(\frac{1}{n} + \frac{\overline{x}^2}{\sum\left(x_t - \overline{x}\right)^2}\right)} = 0,0998$$

$$t_{pr} = \left|\frac{\hat{\beta}_1}{s.e.\left(\hat{\beta}_1\right)}\right| = \frac{1,8787}{0,0998} = 18,8338$$

$$t_{teor} = t\left(0,95;13\right) = 2,16$$

- гіпотеза повинна бути відхилена, а значить коефіцієнт є значимим.

Приклад розв'язання задачі: розв'язок – 4

- Визначимо надійні інтервали для коефіцієнтів з рівнем надійності 0,95 за формулою

$$[\hat{\beta} - \mathbf{s.e.}(\hat{\beta}) \cdot t_{\text{teor}}; \hat{\beta} + \mathbf{s.e.}(\hat{\beta}) \cdot t_{\text{teor}}]$$

$$t_{teor} = t_{0,95;n-2} = t_{0,95;13} = 2,16$$

$$4,316715 < \beta_0 < 7,180448;$$

$$1,663201 < \beta_1 < 2,094201.$$

Приклад розв'язання задачі: розв'язок – 5

- Середній коефіцієнт еластичності

$$\varepsilon = \hat{\beta}_1 \frac{\overline{x}}{\overline{y}} = 1,8787 \frac{93,56}{262} = 0,67$$

- При збільшенні кількості працюючих на 1% дохід підприємства зросте на 0,67%.

Питання?