

ECONOMETRIE

Curs 3

Modelul unifactorial de regresie liniară (II)

Inferenta statistica pentru parametrii modelului de regresie

Inferența statistică pentru parametrii modelului se poate realiza prin:

- ▶ *Testarea ipotezei* statistice referitoare la *semnificația parametrilor* modelului de regresie;
- ▶ *Estimarea pe interval de încredere* a parametrilor modelului din colectivitatea totală.

În cazul regresiei liniare coeficienții ecuației de regresie în eșantion, \hat{a} și \hat{b} , sunt estimații ale coeficienților ecuației de regresie în populația generală a și b.

Testarea ipotezelor statistice – Concepte

- ▶ **Ipoteză statistică** = se intelege “presupunerea” care se face cu privire la parametrul unei repartitii sau a legii de repartitie pe care o urmeaza anumite variabile aleatoare.
- ▶ **Ipoteză nulă (H_0)** = este ipoteza care urmeaza a fi testata. Aceasta presupune ca nu exista deosebiri esentiale sau ca eventualele deosebiri au un caracter intamplator; constă întotdeauna în admiterea caracterului întâmplător al deosebirilor.
- ▶ **Ipoteză alternativă (H_1)** = reprezinta negarea ipotezei nule. Ea va fi acceptată doar când există suficiente dovezi, evidențe, pentru a se stabili că este adevărată.

Testarea ipotezelor statistice – Concepte

- ▶ Ipoteza alternativa poate lua una dintre urmatoarele forme (care vor fi exemplificate pentru testarea ipotezei privind media unei colectivitati)

- **Test bilateral:**

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0 (\mu < \mu_0 \text{ sau } \mu > \mu_0)$$

- **Test unilateral la dreapta:**

$$H_0: \mu = \mu_0$$

$$H_1: \mu > \mu_0$$

- **Test unilateral la stanga:**

$$H_0: \mu = \mu_0$$

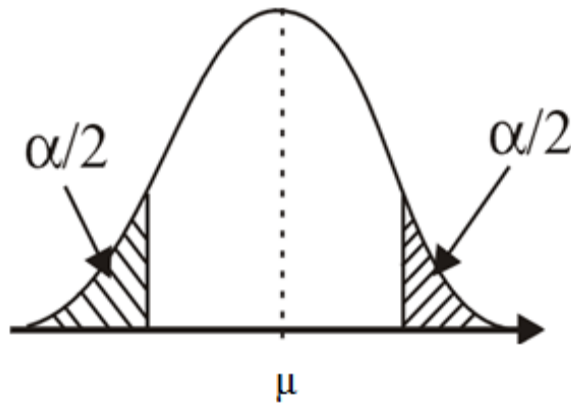
$$H_1: \mu < \mu_0$$

Testarea ipotezelor statistice – Concepte

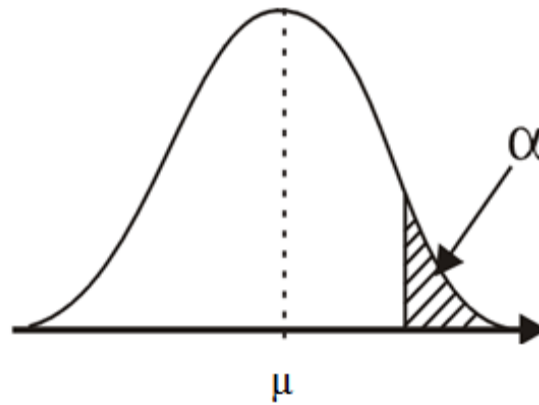
- ▶ **Testul statistic** (criteriu de semnificație) este procedeul de verificare a unei ipoteze statistice, utilizat drept criteriu de acceptare sau de respingere a ipotezei nule.
- ▶ **Regiunea critică**, R_c reunește valorile numerice ale testului statistic pentru care ipoteza nulă va fi respinsă.
- ▶ Regiunea critică este delimitată de **valoarea critică** – punctul de tăietură în stabilirea acesteia.

Testarea ipotezelor statistice – Concepte

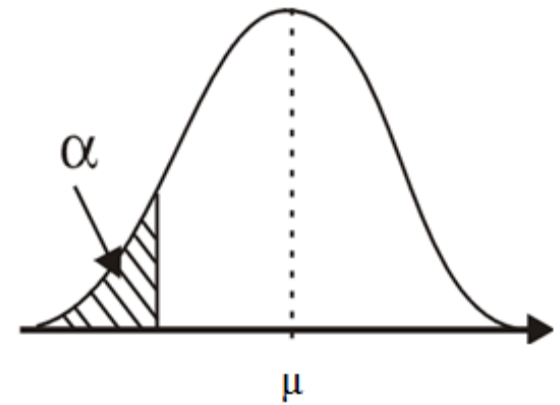
► Regiunea critica



a) Test bilateral



b) Test unilateral dreapta



c) Test unilateral stanga

Testarea ipotezelor statistice – Concepte

- ▶ **Eroare de genul întâi** = eroarea pe care o facem eliminând o ipoteză nulă, deși ea este adevărată.
- ▶ **Riscul de genul întâi (α)** = probabilitatea comiterii unei erori de genul întâi.
 - α se numește **nivel** sau **prag de semnificație**.
- ▶ **Nivelul de încredere** al unui test statistic este $(1-\alpha)$ iar în expresie procentuală, $(1-\alpha)100$ reprezintă probabilitatea de garantare a rezultatelor.
- ▶ **Eroare de genul al doilea** = eroarea pe care o facem acceptând o ipoteză nulă, deși este falsă
 - Probabilitatea (riscul) comiterii unei erori de genul al doilea este β .
 - Puterea testului statistic este $(1-\beta)$.

Exemplu de eroare de tip 1 si 2

Type I Error



Fals pozitiv

Type II Error



Fals negativ

Testarea semnificației parametrilor

– Testul t –

Pentru coeficientul de regresie:

Coeficientul b măsoară schimbarea variabilei rezultative determinată de modificarea cu o unitate a variabilei cauzale.

► *Pasul 1: Definirea ipotezelor:*

$$H_0: b = 0$$

$$H_1: b \neq 0$$

► *Pasul 2: Calculul statisticii t :*

$$t_{calc} = \frac{\hat{b} - 0}{\sigma_{\hat{b}}} \approx \frac{\hat{b}}{s_{\hat{b}}}$$

Testarea semnificației parametrilor

– Testul t –

► *Pasul 3: Valoarea critică:*

$$t_{\frac{\alpha}{2}; n-2}$$

unde α este pragul de semnificație,
 $n-2$ grade de libertate (n = nr de observații, 2 =
numărul de parametri)

► *Pasul 4: Decizia:*

Dacă $|t_{calc}| > t_{\frac{\alpha}{2}; n-2}$ atunci se respinge H_0 iar
parametrul diferă semnificativ de zero (este
semnificativ statistic).

Valori critice pentru testul t

		t Table												
Test unilateral	→	one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	
Test bilateral	→	two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001	
Grade de libertate	df													
	1	0.000	1.000	1.378	1.983	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	1.880	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	2.680	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			

Testarea semnificației parametrilor

– Testul t –

Pentru termenul liber:

- ▶ *Pasul 1:* Definirea ipotezelor:

$$H_o: a = 0$$

$$H_1: a \neq 0$$

- ▶ *Pasul 2:* Calculul statisticii t:

$$t_{calc} = \frac{\hat{a} - 0}{\sigma_{\hat{a}}} \approx \frac{\hat{a}}{s_{\hat{a}}}$$

- ▶ *Pasul 3:* Valoarea critică:

$$t_{\frac{\alpha}{2}; n-2}$$

- ▶ *Pasul 4:* Decizia:

Dacă $|t_{calc}| > t_{\frac{\alpha}{2}; n-2}$ atunci se respinge H_o iar parametrul diferă semnificativ de zero.

Abaterile medii pătraticale ale estimatorilor

Pentru a putea calcula *statistica t*, avem nevoie de abaterile medii pătraticale ale estimatorilor:

Pentru \hat{a}	Pentru \hat{b}
$s_{\hat{a}} = s_{\hat{\varepsilon}} \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2}}$	$s_{\hat{b}} = s_{\hat{\varepsilon}} \sqrt{\frac{1}{\sum (X_i - \bar{X})^2}}$

unde:

$$s_{\hat{\varepsilon}}^2 = \frac{\sum \hat{\varepsilon}_i^2}{n - 2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n - 2}$$

este estimatorul dispersiei variabilei reziduale

Exemplu

Investigarea legăturii dintre corupție și nivelul bunăstării unei țări.

- ▶ Variabila dependentă (Y) este corupția, pe care o cuantificăm pe baza *Indicelui de Percepție a Corupției* (CPI). Acest indice ia valori între 0 și 100 și cu cât valoarea lui este mai mică, cu atât nivelul corupției din țara respectivă este mai mare.
- ▶ Variabila independentă (X) este Produsul Intern Brut pe cap de locuitor (*PIB/loc*, în mii euro) – cea mai frecventă măsură a bunăstării unei țări.

Exemplu

Datele colectate sunt prezentate in tabelul alaturat.
Valorile inregistrate sunt pentru anul 2019.

Tara	CPI	PIB/loc (mii euro)
Austria	77	44,78
Belgia	75	41,20
Bulgaria	43	8,68
Croatia	47	13,26
Cipru	58	25,31
Cehia	56	20,99
Danemarca	87	53,27
Estonia	74	21,22
Finlanda	86	43,57
Franta	69	35,96
Germania	80	41,51
Grecia	48	17,50
Ungaria	44	14,72
Irelanda	74	72,26
Italia	53	29,66
Letonia	56	15,92
Lituania	60	17,34
Luxembourg	80	102,20
Malta	54	26,53
Olanda	82	46,71
Polonia	58	13,78
Portugalia	62	20,74
Romania	44	11,50
Slovacia	50	17,21
Slovenia	60	22,98
Spania	62	26,43
Suedia	85	46,13
UK	77	37,78

Testarea semnificației parametrilor

– exemplu –

Țara	X_i	Y_i	\hat{Y}_i	$\hat{\varepsilon}_i$	$\hat{\varepsilon}_i^2$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
Austria	44,78	77	71,00253	6,00	35,97	13,03	169,65
Belgia	41,20	75	69,16619	5,83	34,03	9,45	89,21
Bulgaria	8,68	43	52,48524	-9,49	89,97	-23,08	532,46
Croatia	13,26	47	54,83453	-7,83	61,38	-18,50	342,07
Cipru	25,31	58	61,0155	-3,02	9,09	-6,45	41,54
Cehia	20,99	56	58,79958	-2,80	7,84	-10,77	115,89
Danemarca	53,27	87	75,35742	11,64	135,55	21,52	462,90
Estonia	21,22	74	58,91756	15,08	227,48	-10,54	110,99
Finlanda	43,57	86	70,38186	15,62	243,93	11,82	139,59
Franta	35,96	69	66,47836	2,52	6,36	4,21	17,68
Germania	41,51	80	69,3252	10,67	113,95	9,76	95,16
Grecia	17,50	48	57,00941	-9,01	81,17	-14,26	203,21
Ungaria	14,72	44	55,58342	-11,58	134,18	-17,04	290,19
Irlanda	72,26	74	85,09823	-11,10	123,17	40,51	1.640,66
Italia	29,66	53	63,24681	-10,25	105,00	-2,10	4,39
Letonia	15,92	56	56,19896	-0,20	0,04	-15,84	250,75
Lituania	17,34	60	56,92734	3,07	9,44	-14,42	207,79
Luxembourg	102,20	80	100,4558	-20,46	418,44	70,45	4.962,50
Malta	26,53	54	61,6413	-7,64	58,39	-5,23	27,30
Olanda	46,71	82	71,99251	10,01	100,15	14,96	223,65
Polonia	13,78	58	55,10126	2,90	8,40	-17,98	323,10
Portugalia	20,74	62	58,67135	3,33	11,08	-11,02	121,33
Romania	11,50	44	53,93174	-9,93	98,64	-20,26	410,27
Slovenia	17,21	50	56,86066	-6,86	47,07	-14,55	211,56
Slovenia	22,98	60	59,82034	0,18	0,03	-8,78	77,00
Spania	26,43	62	61,59	0,41	0,17	-5,33	28,36
Suedia	46,13	85	71,695	13,30	177,02	14,38	206,64
UK	37,78	77	67,41192	9,59	91,93	6,03	36,30
Total	889,14	1.801,00	-	0,00	2.429,87	-	11.342,10

► Calcule intermediare

$$\hat{Y}_i = \hat{a} + \hat{b}X_i = 48,0328 + 0,5129 \cdot X_i$$

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$$

$$\sum \hat{\varepsilon}_i^2 = 2429,87$$

$$\bar{X} = \frac{\sum X_i}{n} = 31,76$$

$$\sum (X_i - \bar{X})^2 = 11342,10$$

$$s_{\hat{\varepsilon}}^2 = \frac{\sum \hat{\varepsilon}_i^2}{n-2} = 93,456$$

$$s_{\hat{a}} = s_{\hat{\varepsilon}} \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum (X_i - \bar{X})^2}} = 3,41$$

$$s_{\hat{b}} = s_{\hat{\varepsilon}} \sqrt{\frac{1}{\sum (X_i - \bar{X})^2}} = 0,09$$

Testarea semnificației parametrilor

- exemplu -

Pentru termenul liber:

- Pasul 1: Definirea ipotezelor:

$$H_0: a = 0$$

$$H_1: a \neq 0$$

- Pasul 2: Calculul statisticii t:

$$t_{calc} = \frac{\hat{a}}{s_{\hat{a}}} = \frac{48,03}{3,41} = 14,07$$

- Pasul 3: Valoarea critică:

$$t_{\frac{\alpha}{2}; n-2} = 2,055$$

- Pasul 4: Decizia:

$$|t_{calc}| > t_{\frac{\alpha}{2}; n-2} \Leftrightarrow 14,07 > 2,055 \text{ (A)}$$

adică respingem H_0 , parametrul a este semnificativ diferit de zero.

Pentru coeficientul de regresie:

- Pasul 1: Definirea ipotezelor:

$$H_0: b = 0$$

$$H_1: b \neq 0$$

- Pasul 2: Calculul statisticii t:

$$t_{calc} = \frac{\hat{b}}{s_{\hat{b}}} = \frac{0,51}{0,09} = 5,65$$

- Pasul 3: Valoarea critică:

$$t_{\frac{\alpha}{2}; n-2} = 2,055$$

- Pasul 4: Decizia:

$$|t_{calc}| > t_{\frac{\alpha}{2}; n-2} \Leftrightarrow 5,65 > 2,055 \text{ (A)}$$

adică respingem H_0 , parametrul b este semnificativ statistic

Testarea semnificației parametrilor

- exemplu -

Outputul de regresie din Excel ne oferă următoarele informații:

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	48,03288543	3,412705603	14,074723	1,13242E-13	41,0179686	55,04780226
PIB/loc (mii euro)	0,512944202	0,090773237	5,6508308	6,088E-06	0,326357141	0,699531263

- ▶ *Standard Error* - reprezintă eroarea standard a estimatorilor:

$$s_{\hat{a}} = 3,4127$$

$$s_{\hat{b}} = 0,0907$$

- ▶ *t Stat* - reprezintă valoarea calculată a statisticii t (t_{calc})

$$t_{calc} \hat{a} = 14,0747$$

$$t_{calc} \hat{b} = 5,6508$$

- ▶ Decizia se ia pe baza valorii lui *P-value*

Dacă $P\text{-value} < \alpha$ atunci acel parametru este semnificativ statistic (α = pragul de semnificație, de obicei $\alpha = 0,05$).

Intervale de încredere pentru parametri

Pentru o probabilitate de garantare a rezultatelor $P = 100(1-\alpha)\%$, se poate afirma că:

- ▶ **Parametrul a** se găsește în următorul interval

$$\hat{a} - t_{\frac{\alpha}{2}; n-2} \cdot s_{\hat{a}} \leq a \leq \hat{a} + t_{\frac{\alpha}{2}; n-2} \cdot s_{\hat{a}}$$

- ▶ **Parametrul b** se găsește în intervalul:

$$\hat{b} - t_{\frac{\alpha}{2}; n-2} \cdot s_{\hat{b}} \leq b \leq \hat{b} + t_{\frac{\alpha}{2}; n-2} \cdot s_{\hat{b}}$$

Exemplu:

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	48,03288543	3,412705603	14,074723	1,13242E-13	41,0179686	55,04780226
PIB/loc (mii euro)	0,512944202	0,090773237	5,6508308	6,088E-06	0,326357141	0,699531263

Lower 95% - limita inferioară a intervalului de încredere pt $P=95\%$

Upper 95% - limita superioară a intervalului de încredere pt $P=95\%$

Intervale de încredere pentru parametri

- exemplu -

- ▶ *Parametrul a* se găsește în următorul interval

$$P(\hat{a} - t_{\frac{\alpha}{2}; n-2} \cdot s_{\hat{a}} \leq a \leq \hat{a} + t_{\frac{\alpha}{2}; n-2} \cdot s_{\hat{a}}) = (1 - \alpha) 100$$
$$48,032 - 2,055 \cdot 3,412 \leq a \leq 48,032 + 2,055 \cdot 3,412$$
$$P(41,02 \leq a \leq 55,05) = 95\%$$

- ▶ *Parametrul b* se găsește în intervalul:

$$P(\hat{b} - t_{\frac{\alpha}{2}; n-2} \cdot s_{\hat{b}} \leq b \leq \hat{b} + t_{\frac{\alpha}{2}; n-2} \cdot s_{\hat{b}}) = (1 - \alpha) 100$$
$$0,513 - 2,055 \cdot 0,091 \leq b \leq 0,513 + 2,055 \cdot 0,091$$
$$P(0,326 \leq b \leq 0,699) = 95\%$$

Testarea validității modelului

– metoda analizei varianței (ANOVA) –

În literatura de specialitate, precum și în pachetele de programe informatice specializate se utilizează notațiile:

- ▶ **SST = varianța totală**, suma pătratelor abaterilor totale
- ▶ **SSE = varianța neexplicată (reziduală)**, suma pătratelor erorilor
- ▶ **SSR = varianța explicată**, suma pătratelor abaterilor datorate regresiei
- ▶ Deci, **$SST = SSR + SSE$** .

ANOVA pentru regresie

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

SST = **SSE** + **SSR**

SST – Total Sum of Squares

Măsoară variația valorilor observate Y_i în jurul mediei \bar{Y}

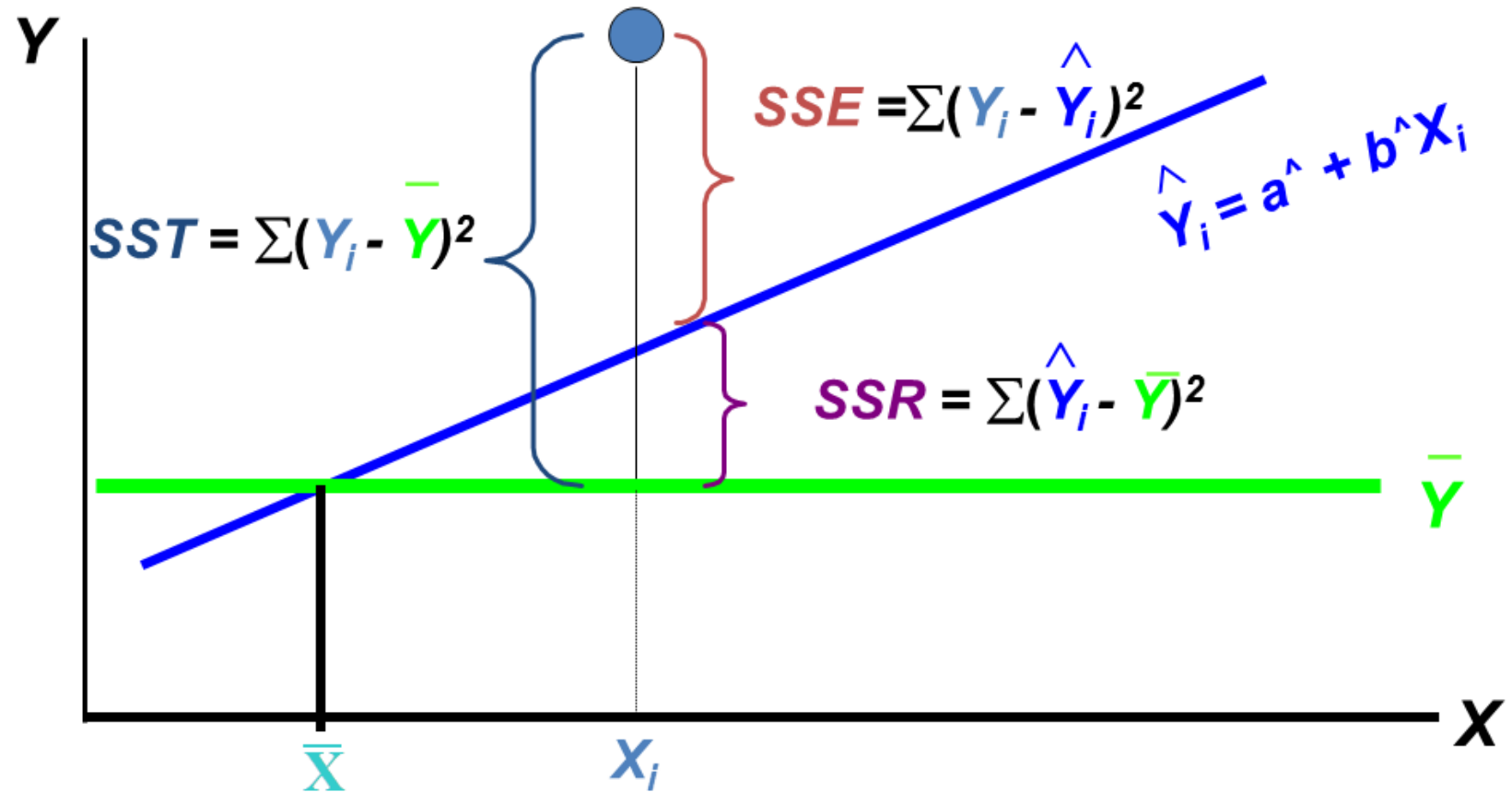
SSE – Error Sum of Squares

Măsoară variația ce poate fi atribuită altor factori, diferiți de variabila explicativă X

SSR – Regression Sum of Squares

Măsoară variația explicată de modelul de regresie

Descompunerea variației



Testarea validității modelului

Ipoteze:

H_0 : $MSR = MSE$ (sau $b = 0$) *model nevalid (nesemnificativ) statistic*

H_1 : $MSR > MSE$ (sau $b \neq 0$) *model valid (semnificativ) statistic*

Statistica utilizată este:

$$F_{calc} = \frac{MSR}{MSE}$$

Decizia: Se compară valoarea calculată a testului F cu valoarea teoretică pentru un prag de semnificație α , și $k-1$, respectiv $n-k$ grade de libertate, preluată din tabelul repartiției Fisher: $F_{\alpha, k-1, n-k}$.

- ▶ Dacă $F_{calc} > F_{\alpha, k-1, n-k}$ se respinge H_0 , adică se concluzionează că modelul este valid.

Tabelul F

Pentru $\alpha = 0,05$

Puteti gasi valorile critice pentru alte praguri de semnificatie la adresa:

<http://www.stat.ufl.edu/~a/sta6126/tables.pdf>

$\alpha = .05$										
df_2	df_1									
	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18	1.98	1.73
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16	1.96	1.71
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12	1.91	1.65
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00	1.79	1.51
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92	1.70	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83	1.61	1.25
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.75	1.52	1.00

Calculul statisticii F

Sursa variației	Suma pătratelor (<i>SS-Sum of Squares</i>)	Grade de libertate (<i>df- degree of freedom</i>)	Media pătratelor (<i>MS- Mean of Squares</i>)	Testul Fisher (<i>testul F</i>)
Datorată regresiei	$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$k-1$	$MSR = \frac{SSR}{k-1}$	$F_{calc} = \frac{MSR}{MSE}$
Reziduală	$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$n - k$	$MSE = \frac{SSE}{n-k}$	
Totală	$SST = \sum_{i=1}^n (y_i - \bar{y})^2$	$n - 1$		

unde k este numărul de parametri (pentru regresia unifactorială $k = 2$)

Validitatea modelului – exemplu

SUMMARY OUTPUT						
<i>Regression Statistics</i>						
Multiple R	0,74242649					
R Square	0,551197093					
Adjusted R Square	0,533935443					
Standard Error	9,667286275					
Observations	28					
ANOVA						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>	
Regression	1	2984,240121	2984,2401	31,93188864	6,088E-06	
Residual	26	2429,867022	93,456424			
Total	27	5414,107143				
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	48,03288543	3,412705603	14,074723	1,13242E-13	41,0179686	55,04780226
PIB/loc (mii euro)	0,512944202	0,090773237	5,6508308	6,088E-06	0,326357141	0,699531263

Decizia: Dacă $Significance F < \alpha$ atunci modelul este valid.

Aprecierea calității modelului de regresie

Pentru a măsura calitatea ajustării în cazul regresiei liniare unifactoriale se calculează:

1. Raportul de corelație
2. Coeficientul de determinație
3. Coeficientul de determinație ajustat
4. Abaterea medie pătratică (eroarea standard) a reziduurilor – măsură absolută a calității ajustării pe baza regresiei în eșantion

Coeficientul de determinație

- *Coeficientul de determinație* (R^2) arată proporția din variația totală a variabilei dependente explicată de variația variabilei independente

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{cu } R^2 \in [0; 1]$$

R^2 este afectat de creșterea numărului de parametri; de aceea pentru modele cu multi parametri se calculează R^2 ajustat, care are aceeași interpretare.

$$R^2_{adj} = 1 - (1 - R^2) \frac{n-1}{n-k}$$

Raportul de corelație

Raportul de corelație este un indicator relativ utilizat pentru:

- măsurarea intensității legăturii dintre variabile
- validarea modelelor de regresie.

Raportul de corelație se calculează ca:

$$R = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{sau} \quad R = \sqrt{\frac{SSR}{SST}} = \sqrt{1 - \frac{SSE}{SST}} \quad \text{sau} \quad R = \sqrt{R^2}$$

- Dacă $R \rightarrow 1$ legatura dintre X și Y este puternică
- Dacă $R \rightarrow 0$ legatura dintre X și Y este slabă
- În cazul legăturilor liniare: $R = |r_{xy}|$

Abaterea medie pătratică a erorilor

- ▶ Abaterea medie pătratică a erorilor în eșantion este:

$$s_e = \sqrt{s_e^2} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}}$$

unde s_e^2 este un estimator nedeplasat al dispersiei reziduurilor σ_e^2

- ▶ Acest indicator s_e este important în determinarea intervalului de încredere pentru coeficientul de regresie b și pentru termenul liber a .
- ▶ s_e este util în compararea modelelor. Dacă avem la dispoziție câteva modele dintre care trebuie să alegem, cel mai potrivit a fi utilizat este cel pentru care s_e este mai scăzut.

Aprecierarea calității modelului de regresie – exemplu

► Rezultatele:

Raportul de corelatie
Coeficientul de determinatie
Coeficientul de determinatie ajustat
Abaterea medie patratica a erorilor
Numarul de observatii

<i>Regression Statistics</i>	
Multiple R	0,74243
R Square	0,55120
Adjusted R Square	0,53394
Standard Error	9,66729
Observations	28

Previziunea valorii variabilei Y in ipoteza modificarii variabilei factoriale

- ▶ Previziune punctuala

$$\hat{Y}_i = \hat{a} + \hat{b} \cdot X_i$$

- ▶ Previziune pe baza de interval de incredere:

$$\hat{Y}_i - t_{\alpha/2, n-k} \cdot s_{\hat{e}} \cdot \sqrt{h_i} \leq \mu_{Y/X=X_i} \leq \hat{Y}_i + t_{\alpha/2, n-k} \cdot s_{\hat{e}} \cdot \sqrt{h_i}$$

unde

$$h_i = 1 + \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$