

Хамгийн бага квадратын аргаар шугаман хамаарлын тооцоолол хийхэд LINEST функцийг хэрэглэх нь

П.Оюубилэг (Ph.D), МУИС-ийн, ЭЗС-ийн багш

Хамгийн бага квадратын аргаар шугаман хамаарлын коэффициентуудыг холбогдох статистик утгуудын хамт тооцоолох нь нэлээд математик бодолттой байдгийг бид мэднэ. Харин энэхүү тооцооллыг маш хялбараар хийж болох боломжийг бидэнд EXCEL системийн LINEST функц олгоно. Энд $y = m \cdot x + b$ нэг хувьсагчийн функцэн хамааралтай, эсвэл $y = m_1 \cdot x_1 + m_2 \cdot x_2 + \dots + b$ олон хувьсагчийн функцэн хамааралтай тэгшитгэлийн хувьд (энд y - нь хамааран хувьсагч, x - нь үл хамааран хувьсагч, m - нь коэффициентүүд, b - нь тогтмол хэмжигдэхүүн юм.) LINEST функцийн утга нь $\{m_n, m_{n-1}, \dots, m_1, b\}$ массив байна.

Const - нь тогтмол хэмжигдэхүүнийг тооцох, эсвэл үгүйг заасан логик утга байна. const нь TRUE утгатай, эсвэл орхигдсон байвал b сул гишүүн тооцогдоно, харин const нь FALSE бол $b = 0$ буюу сул гишүүн тооцогдохгүй.

Stats - нь нэмэлт статистик утгууд тооцох, эсвэл үгүйг заасан логик утга байна. stats нь FALSE, эсвэл орхигдсон бол зөвхөн m - коэффициентүүд, b тогтмол хэмжигдэхүүн л тооцогдоно. Харин stats нь TRUE бол доорхи хүснэгтэнд байгаа нэмэлт статистик утгууд тооцогдож, LINEST функцийн утга $\{m_n, m_{n-1}, \dots, m_1, b; se_n, se_{n-1}, \dots, se_1, se_b; r_2, se_y; F, df; ssreg, ssresid\}$ гэсэн массив болж өргөжинө.

Статистик	Бодолтын утга
se1, se2, ..., sen	m_1, m_2, \dots, m_n коэффициентүүдийн стандарт алдаа
Seb	Сул гишүүн b - ийн стандарт алдаа (const нь FALSE тохиолдолд $se_b = \#N/A$ байна)
r2	Детерминацийн коэффициент.
Se _y	Үнэлэгдсэн y - ийн стандарт алдаа.
F	F - ажиглалтын утга.
Df	Чөлөөний зэрэг.
Ssreg	Квадратуудын регресс нийлбэр.
Ssresid	Квадратуудын резидуал нийлбэр.

Функцийн ерөнхий синтакс нь :
LINEST(known_y's, known_x's, const, stats) юм.

Энд :

Known_y's - нь хамаарал дахь Y - ийн утгууд байна.

Known_x's - нь хамаарал дахь X - ийн утгууд юм. known_x's массивт нэг болон хэд хэдэн хувьсагчийн массив байж болно. known_x's нь орхигдож болох бөгөөд энэ тохиолдолд Known_x's - ийн утгыг known_y's - тэй ижил хэмжээтэй $\{1, 2, 3, \dots\}$ массив гэж тооцдог байна.

Одоо энэ функцийг ашигласан олон хэмжээст шугаман регресс бодох жишээ авч үзье.

Нэгэн аж ахуй эрхлэгч бизнесийн дүүрэгт жижиг аж ахуй эрхлэхээр тухайн дүүргийнхний худалдан авах чадварт тохирсон барилга сонгохоор зорьжээ. Энэ зорилгодоо хүрэхийн тулд олон хэмжээст шугаман регрессийн шинжилгээг дараахь хувьсагчууд дээр үндэслэн үнэлэх болжээ.

Хувьсагчууд	Агуулга
Y	Аж ахуйн барилгын үнэлэгдсэн утга
x_1	Талбайн хэмжээ
x_2	Өрөөний тоо
x_3	Хонгилын тоо
x_4	Барилгын эдэлгээний хугацаа

Ингэхдээ хамааран хувьсагч (x_1, x_2, x_3, x_4) болон үл хамааран хувьсагч (y) ийн утгууд шулуун шугаман хамааралтай гэж таамаглажээ. Тэрээр боломжтой 1500 барилгаас доорхи үзүүлэлттэй 11 барилгын загварыг сонгон авч тооцоондоо ашиглажээ.

Үүний дүнд энэ бодолтын хариу доорхи байдлаар гарчээ.

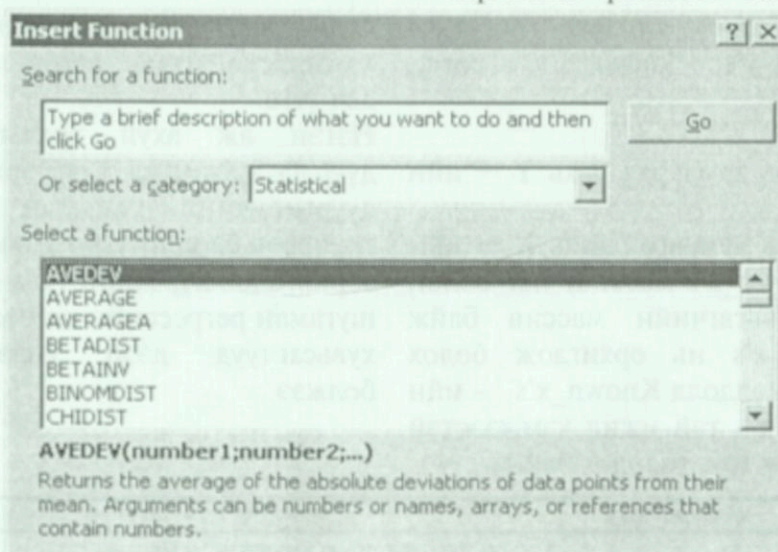
Одоо бид олон хэмжээст регрессийн тэгшитгэл $y = m_1*x_1 + m_2*x_2 + m_3*x_3 + m_4*x_4 + b$ дээ олсон дээрхи утгуудаа орлуулж бичвэл :

	A	B	C	D	E
1	Талбайн хэмжээ (x_1)	Өрөө(x_2)	хонгил (x_3)	Эдэлгээний хугацаа (x_4)	Таамаглаж буй үнэлгээ (y)
2	2310	2	2	20	142,000
3	2333	2	2	12	144,000
4	2356	3	1.5	33	151,000
5	2379	3	2	43	150,000
6	2402	2	3	53	139,000
7	2425	4	2	23	169,000
8	2448	2	1.5	99	126,000
9	2471	2	2	34	142,900
10	2494	3	3	23	163,000
11	2517	4	4	55	169,000
12	2540	2	3	22	149,000

Одоо массивын арга ашиглан дараахь бодолт хийе. A14:E18 мужид бодолтын үр дүнг гаргахыг хүсч байгаа гэж үзэн энэ мужийг идэвхижүүлээд Insert/Function команд өгнө. Командын дүнд гарах дараахь цонхны Select a category нүдэнд Statistical сонголтыг хийгээд Select a function нүднээс Linest функцээ сонгоно.

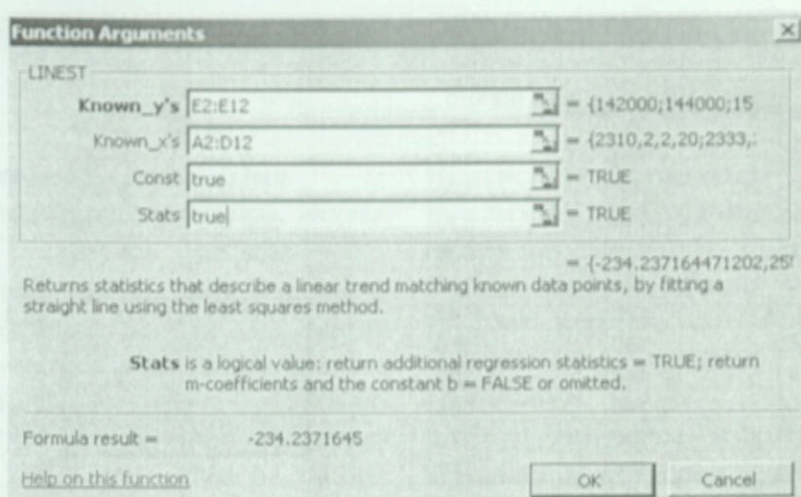
$$y = 27.64*x_1 + 12530*x_2 + 2553*x_3 - 234.24*x_4 + 52318$$

гэсэн тэгшитгэл болно. Одоо аж ахуй эрхлэгч энэ тэгшитгэлээ ашиглаад өөрийн сонирхож байгаа талбай, өрөө, хонгил, эдэлгээний хугацаатай ямар ч барилгын үнэлгээг таамаглаж болох



Энэ функцийн аргуудыг доорх байдлаар зааж өгөөд CTRL+SHIFT+ENTER товчлууруудыг зэрэг дарна.

Тухайлбал аж ахуй эрхлэгч 2500 квадрат метр талбайтай, гурван өрөөтэй, хоёр хонгилтой, 25 жилийн эдэлгээний



хугацаатай барилга сонирхсон гэвэл үнэ нь :

$$y = 27.64 \cdot 2500 + 12530 \cdot 3 + 2553 \cdot 2 - 234.24 \cdot 25 + 52318 = \$158261 \text{ болно.}$$

Одоо нэг хувьсагчийн шугаман регрессийн тэгшитгэл $y = mx + b$ ийн хувьд нэгэн жишээ авч үзье. Доорхи цонхонд 1988-1999 оны хоорондох хөгжиж буй эдийн засагтай улсуудын цэвэр орлого (X тэрбум доллар), нийт хэрэглээ (Y тэрбум

	A	B	C	D	E
14	-234.237164	2553.211	12529.77	27.64139	52317.83
15	13.26801148	530.6692	400.0668	5.429374	12237.36
16	0.996747993	970.5785	#N/A	#N/A	#N/A
17	459.7536742	6	#N/A	#N/A	#N/A
18	1732393319	5652135	#N/A	#N/A	#N/A

Энэ жишээнд детерминацийн коэффициент (A16 нүд) 0.99675 гарчээ. Энэ нь худалдах үнэ болон хамааран хувьсагчдын хооронд хүчтэй хамааралтай байгааг харуулж байна. F - статистик ашиглан өндөр хамааралтай эдгээр үр дүн тохиолдлынх уу, үгүй юу гэдгийг тодорхойлж болох юм. F – ажиглалтын утга F – критик утгаас их байвал эдгээр хувьсагчдын харилцан хамаарал сайн гэж үздэг. F – критик утгыг тооцохдоо Alpha утгыг 0.05, чөлөөний зэргийг $v1 = k = 4$ болон $v2 = n - (k + 1) = 11 - (4 + 1) = 6$ (энд k нь хувьсагчийн тоо, n нь өгөгдлийн тоо) гэж авбал F – критик утга 4.53 байгааг олж болох юм. Харин F – ажиглалтын утга 459.753674 (A17 нүд) байгаа ба энэ нь F - критик утга 4.53 – аас их юм. Иймд энэ регресс тэгшитгэлээр барилгын үнэлэгдсэн утга тун сайн илэрхийлэгдсэн гэж үзэж болно.

доллар) өгөгджээ.

Дээрх жишээнээс бодолт хийх талаар ойлгох болсон гэж үзээд үр дүн дээр ярилцая.

E2:F6 нүдний статистикуудыг дараах хүснэгтэд илэрхийлснээр ойлгоно гэдгийг та бүхэн мэдэж байгаа бизээ.

Энэ хүснэгтээс харахад үнэлэгдсэн регрессийн тэгшитгэл:

$$y = 0.862 \cdot x + 2.129 \text{ хэлбэртэй болно.}$$

Стандаргалдааны статистикуудыг ашиглан регрессийн коэффициентүүд бодитой эсэхийг шалгая. Эдгээр таамаглалуудыг шалгахдаа үнэлэгдсэн коэффициентүүдийг харгалзах стандарт алдаанд нь харьцуулсан t статистикийг ашиглая. Бидний авч үзэж буй жишээний хувьд студентийн

	A	B	C	D	E	F	G
1		Нийг хэрэглээ	Цэвэр орлого				
2	1	102	114		0.861	2.129	
3	2	106	118		0.049	7.164	
4	3	108	126		0.969	3.395	
5	4	110	130		309.602	10	
6	5	122	136		3568.732	115.268	
7	6	124	140				
8	7	126	148		=LINEST(B2:B13,C2:C13,1,1)		
9	8	130	156				
10	9	142	160				
11	10	148	164				
12	11	150	170				
13	12	154	178				
14							

t тархалтын чөлөөний зэрэг нь $12-2=10$ бөгөөд 95%-ийн магадлалд харгалзах онолын утга 2.23 байна. Сул гишүүний t статистикийн утга $2.129/7.164=0.297$, коэффициентийн t статистикийн утга $0.861/0.049=17.596$ гарчээ. Одоо бид эдгээр утгуудаа өмнөх 10 гэсэн чөлөөний зэрэгтэй t тархалтын онолын утга 2.23-тэй харьцуулбал $0.297 < 2.23$, $17.596 > 2.23$.

тэгшитгэлийн хүчин зүйлүүдийн хамаарал сайн байгааг илэрхийлнэ.

Ийнхүү эрхэм уншигч танд энэхүү функцийг хэрэглээг танилцууллаа. Цаашид өөрсдийн ажил, хичээлд өргөн хэрэглэнэ гэж итгэж байна.

Статистик	Харгалзах нүд	Статистик утга
b0-ийн коэффициент	F2	2.129
b1-ийн коэффициент	E2	0.861
b0-ийн стандарт алдаа	F3	7.164
b1-ийн стандарт алдаа	E3	0.049
детерминацийн коэффициент	E4	0.969
Y-ийн хувь дахь стандарт алдаа	F4	3.395
F-ажиглалтын утга	E5	309.602
F-тархалтын чөлөөний зэрэг	F5	10
Регрессийн квадратуудын нийлбэр	E6	3568.732
Алдааны квадратуудын нийлбэр	F6	115.268

Иймд бидний регрессийн тэгшитгэлийн сул гишүүнийг ач холбогдолгүй, харин X – ийн коэффициент ач холбогдолтой гэж үзэж болох юм. Өөрөөр хэлбэл бид дээрх шугаман регрессийг сул гишүүнгүйгээр цааш нь үнэлэх шаардлагатай боллоо. Сул гишүүнгүйгээр үнэлэхийн тулд өмнөх алхамуудыг давтан гүйцэтгээд linest функцийг хувьд const-г false(0) болгоно. Үнэлэгдсэн сул гишүүнгүй регрессийн тэгшитгэл $Y=0.876 \cdot X$ гарна.

Детерминацийн коэффициент 0.969 байгаа нь бидний олсон регрессийн