

CFA LEVEL I · 2026 · VOLUME 1

LEARNING MODULE 10

V.2 · САЙЖРУУЛСАН

ЭНГИЙН ШУГАМАН РЕГРЕСС

Simple Linear Regression

Зургаан Сурах Үр Дүн · Бодит дасгал бодлого ·
Бүрэн томъёоны лавлах · Нийт 18 жишээ

CFA AI CURRICULUM

Агуулга

Оршил	Шугаман регресс яагаад хэрэгтэй вэ	3
LOS 1	SLR загвар, OLS, коэффициент	5
	· Хамаарлын ба үл хамаарлын хувьсагч	5
	· OLS — Slope ба intercept	7
	· ROA × CAPEX жишээ + дасгал бодлого	9
LOS 2	Дөрвөн таамаглал ба residual plot	13
	· Linearity, Homoskedasticity, Independence, Normality	13
	· Зохицуулалтын режимийн жишээ + дасгал	17
LOS 3	Тааруулалтын хэмжүүр ба тестүүд	20
	· SST, SSR, SSE, R^2	20
	· F-test, t-test ба дасгал бодлого	22
LOS 4	ANOVA ба standard error of estimate	26
	· ANOVA хүснэгт + McCoin restaurant жишээ	26
LOS 5	Прогноз ба prediction interval	30
	· Standard error of forecast + Vasileva Amtex жишээ	30
LOS 6	Log-lin, lin-log, log-log хэлбэрүүд	34
Cheat	Cheat Sheet — Бүх томъёо нэг хуудсанд	38
Term	Нэр томъёоны жагсаалт (En ↔ Mn)	40

Шугаман регресс яагаад хэрэгтэй вэ

Хөрөнгө оруулалтын шинжээч өдөр тутамдаа нэг хувьсагчийн өөрчлөлтийг нөгөө хувьсагчийн утгаар тайлбарлахыг хүсдэг. Орлогын өсөлт компанийн зах зээлийн үнэлгээг тайлбарлаж чадах уу? Хүүний түвшин инфляцитай ямар холбоотой вэ? Эдгээр асуултад тоон үнэн зөв хариулт өгөх арга нь **шугаман регрессийн загвар** юм.

Шугаман регресс нь зөвхөн хоёр хувьсагчдын хооронд статистик хамаарал бий эсэхийг тогтоох төдийгүй уг хамаарлын хүчийг тоон утгаар илэрхийлэх ба нэг хувьсагчийн утгад үндэслэн нөгөөгийн утгыг урьдчилан таамаглах боломжийг олгодог. Зөвхөн нэг үл хамаарлын хувьсагчтай тохиолдлыг **simple linear regression** буюу SLR гэдэг.

Энэ модуль танд юу заах вэ

- ▶ **OLS** буюу ordinary least squares — хамгийн сайн тааруулсан шугамыг олох арга.
- ▶ **Регрессийн коэффициентүүд** — slope болон intercept-ийн утга.
- ▶ **Дөрвөн үндсэн таамаглал** — linearity, homoskedasticity, independence, normality.
- ▶ **Тааруулалтын хэмжүүрүүд** — R^2 , F-statistic, standard error of estimate.
- ▶ **Гипотезийн тестүүд** — slope, intercept, correlation = 0 гэдэг таамаглалуудад.
- ▶ **Функциональ хэлбэрүүд** — log-lin, lin-log, log-log.

ГОЛ САНАА

- **Dependent variable (Y)** — бид тайлбарлахыг хүсэж буй хувьсагч.
- **Independent variable (X)** — Y-г тайлбарлахад ашиглах хувьсагч.
- Загвар: $Y_i = b_0 + b_1 X_i + \epsilon_i$ — бидний оношлох ёстой математик хэлбэр.
- **OLS** гэдэг бол residual-уудын квадратын нийлбэрийг хамгийн бага байлгах байдлаар b_0, b_1 -ийг олох арга.
- $R^2 = SSR/SST$ — Y-ийн хэлбэлзлийн хэдэн хувийг X тайлбарлав гэдгийг харуулна.

Энэ модулийн дотор бид

6

СУРАХ ҮР ДҮН

18

БОДИТ ЖИШЭЭ

18

ДАСГАЛ + ШИЙДЭЛ

- ▶ Зургаан компанийн ROA болон CAPEX-ийн өгөгдлөөр $\hat{b}_1 = 1.25$, $\hat{b}_0 = 4.875$ гэсэн загварыг бүтээж бодит интерпретац хийнэ.
- ▶ Богино хугацааны хүүний түвшин ба инфляцийн жишээгээр heteroskedasticity-ийн зөрчлийг харуулна.
- ▶ ROA загварт $R^2 = 80.01\%$, $F = 16.0104$ — статистикийн ач холбогдолтой загвар бий гэдгийг нотлоно.
- ▶ $X_f = 6\%$ -ийн үед ROA-ийн 95% prediction interval [2.00%, 22.75%] гэдгийг тооцоолно.
- ▶ Log-lin загвараар компанийн орлогын \ln -г цагаар регресс хийж $R^2 = 98.08\%$ хүртэл сайжруулна.

1 SLR загвар, OLS, коэффициент

describe a simple linear regression model, how the least squares criterion is used to estimate regression coefficients, and the interpretation of these coefficients

1А. Хамаарлын ба үл хамаарлын хувьсагч

Регрессийн шинжилгээ нь хоёр хувьсагчийн хоорондын дундаж шугаман хамаарлыг тоон загвараар илэрхийлэх процесс юм. Загварт оролцох хувьсагчдын үүрэг өөр өөр:

- ▶ **Dependent variable (Y)** — бид яагаад нэг компани өндөр ROA-тай, нөгөө нь бага байгааг тайлбарлахыг хүсэж буй хувьсагч. "Тайлбарлагдсан" хувьсагч гэж бас нэрлэдэг.
- ▶ **Independent variable (X)** — Y-ийн хэлбэлзлийг тайлбарлахад ашиглах хувьсагч. "Тайлбарлагч" хувьсагч.

Жишээлбэл, шинжээч компанийн EPS-ийн өөрчлөлт хувьцааны өгөөжийг тайлбарлах эсэхийг судалж байна гэвэл — Y нь хувьцааны өгөөж, X нь EPS-ийн өөрчлөлт болно. Хэрэв шинжээч ДНБ-ийн өсөлт инфляцид хэрхэн нөлөөлж байгааг үзэхийг хүсвэл — Y нь инфляц, X нь ДНБ-ийн өсөлт.

ЖИШЭЭ 1 · ХУВЬСАГЧДЫГ ТАНИХ

Нөхцөл: Шинжээч 100 компанийн 5 жилийн сар бүрийн EPS-ийн revision болон хувьцааны өгөөжийн өгөгдлийг цуглуулсан. EPS-ийн revision нь хувьцааны өгөөжид хэрхэн нөлөөлж байгааг шинжлэх зорилготой.

Y юу вэ, X юу вэ?

ШИЙДЭЛ

Y = сар бүрийн хувьцааны өгөөж (тайлбарлахыг хүсэж буй); X = EPS revision (тайлбарлагч). Шинжээч хувьцааны өгөөжийн хэлбэлзлийг EPS-ийн revision-аар тайлбарлахыг хүсэж байна.

ЧУХАЛ ЯЛГАХ ХЭЛЛЭГ

- Англиар "Y is regressed on X" гэдэг — Y-г X дээр регресс хийнэ. Y тайлбарлагдсан, X тайлбарлагч.
- X-ийг "random биш" гэж таамагладаг — ажиглагдсан утга нь тогтмол байх ёстой.
- X нэг байх — simple linear regression. Хоёр ба түүнээс олон X — multiple regression.
- Y-ийн нэгж нь b_0 , b_1 -ийн нэгжийг тодорхойлно — Y нь % бол b_0 ч мөн % болно.

1Б. Шугаман загвар

Шугаман регрессийн загвар нь популяцийн дотор Y ба X-ийн хооронд дундаж шугаман хамаарал байдаг гэж таамагладаг. Энэхүү хамаарлыг дараах байдлаар бичнэ:

ШУГАМАН РЕГРЕССИЙН ПОПУЛЯЦИЙН ЗАГВАР

$$Y_i = b_0 + b_1 X_i + \epsilon_i, \quad i = 1, 2, \dots, n$$

Y_i — i -р ажиглалтын dependent variable-ийн утга

X_i — i -р ажиглалтын independent variable-ийн утга

b_0 — Intercept (популяцийн), $X = 0$ үед Y -ийн утга

b_1 — Slope (популяцийн), X -ийн нэг нэгж өөрчлөгдөхөд Y -д гарах өөрчлөлт

ϵ_i — Error term, ажиглалт ба үнэн шугамаас зөрөх хазайлт

Анхаар: b_0 ба b_1 нь популяцийн параметрууд, тэдгээрийг бид зөвхөн түүврээс үнэлдэг. Регрессийн загвар нь " $\hat{Y} = a + bX$ " гэсэн математикийн шугамтай адил харагддаг ч нэг чухал ялгаатай: error term ϵ_i . Энэ нь үнэн утга популяцийн шугам дээр яг тогтохгүй, харин бага зэрэг хазайдаг гэдгийг илэрхийлдэг.

ϵ нь "систематик алдаа" гэдэг утга **биш** — энэ нь зүгээр л X -ээр тайлбарлагдаагүй санамсаргүй хэлбэлзэл.

1B. OLS — Хамгийн бага квадратын арга

Бид популяцийн b_0, b_1 -ийг харж чадахгүй учир тэдгээрийг түүврээс үнэлэх ёстой. Үнэлэгдсэн коэффициентүүдийг \hat{b}_0, \hat{b}_1 ("hat" гэж нэрлэдэг) гэж тэмдэглэдэг. Эдгээр үнэлгээг олох хамгийн түгээмэл арга нь **ordinary least squares (OLS)** юм.

OLS-ийн санаа маш энгийн: ажиглалт бүрээс шугам хүртэлх босоо зайны квадратын нийлбэрийг хамгийн бага байх шугамыг сонго. Энэхүү шугамыг "best-fit line" буюу хамгийн сайн тааруулсан шугам гэдэг.

SUM OF SQUARES ERROR — OLS-ИЙН ШАЛГУУР

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n [Y_i - (\hat{b}_0 + \hat{b}_1 X_i)]^2 = \sum_{i=1}^n e_i^2$$

OLS нь \hat{b}_0, \hat{b}_1 -ийг дээрх SSE хамгийн бага байхаар сонгодог.

$e_i = Y_i - \hat{Y}_i$ — i -р ажиглалтын residual.

Анхаар: Error term (ε_i) — популяцийн; residual (e_i) — түүврийн.

OLS-ийн шалгуурын дагуу \hat{b}_1 -ийг олбол slope нь Y, X хоёрын **covariance**-ийг X -ийн **variance**-д харьцуулсан утгатай тэнцүү:

OLS — SLOPE БА INTERCEPT-ИЙН ТОМЬЁО

$$\hat{b}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\text{Cov}(Y, X)}{\text{Var}(X)}$$

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}$$

\bar{Y} — Y -ийн түүврийн дундаж; \bar{X} — X -ийн түүврийн дундаж.

Slope \hat{b}_1 — Y, X -ийн covariance / X -ийн variance.

Intercept \hat{b}_0 — регрессийн шугам Y -тэнхлэгийг $X = 0$ цэг дээр огтлох өндөр.

OLS-ИЙН МАТЕМАТИК ЧУХАЛ ШИНЖ

- Residual-уудын нийлбэр үргэлж яг тэг — $\sum e_i = 0$.
- Residual-уудын дундаж ч тэг — $E(\varepsilon) = 0$.
- Регрессийн шугам үргэлж (\bar{X}, \bar{Y}) цэгийг дайран өнгөрнө.
- OLS-ийн "best-fit" гэдэг нь огтгоор хазайлтын нэмэх/хасах ялгаа биш — *квадратын* нийлбэрийг бууруулж буй гэсэн утга.

1Г. Slope ба correlation-ийн ялгаа

Slope \hat{b}_1 ба Pearson correlation r -ийн томьёо маш ойролцоо:

SLOPE VS CORRELATION — DENOMINATOR-ЫН ЯЛГАА

$$\hat{b}_1 = \frac{\text{Cov}(Y, X)}{\text{Var}(X)}, \quad r = \frac{\text{Cov}(Y, X)}{s_Y \cdot s_X}$$

Slope-ын denominator нь зөвхөн X-ийн variance.

Correlation-ийн denominator нь Y болон X-ийн стандарт хазайлтын үржвэр.

Үр дүн: Slope ба correlation үргэлж нэг ижил тэмдэгтэй (covariance тэмдэгээс хамаарна).

1Д. Жишээ: ROA × CAPEX

Зургаан компанийн ROA (хөрөнгийн өгөөж, %) болон CAPEX (өмнөх үеийн scaled капитал зардал, %)-ийн өгөгдлөөр SLR загвар бүтээе. Шинжээч CAPEX нь ROA-г тайлбарлах боломжтой эсэхийг шалгахыг хүсэж байна.

Компани	ROA (Y, %)	CAPEX (X, %)	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})(X_i - \bar{X})$
A	6.0	0.7	42.25	29.16	35.10
B	4.0	0.4	72.25	32.49	48.45
C	15.0	5.0	6.25	1.21	-2.75
D	20.0	10.0	56.25	15.21	29.25
E	10.0	8.0	6.25	3.61	-4.75
F	20.0	12.5	56.25	40.96	48.00
Нийт	75.0	36.6	239.50	122.64	153.30
<i>Дундаж</i>	<i>12.5</i>	<i>6.1</i>	—	—	—

Эдгээр нийлбэрийг ашиглан \hat{b}_1 ба \hat{b}_0 -ийг тооцъё:

ROA × CAPEX ТООЦОО

$$\hat{b}_1 = \frac{153.30}{122.64} = 1.25$$

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X} = 12.5 - (1.25 \times 6.1) = 12.5 - 7.625 = 4.875$$

$$\hat{Y}_i = 4.875 + 1.25X_i$$

Slope: CAPEX 1%-аар нэмэгдвэл ROA дунджаар 1.25%-аар нэмэгдэнэ.

Intercept: CAPEX = 0 үед загвараар ROA нь 4.875% болно.

Pearson correlation тооцох:

$$s_Y = \sqrt{\frac{239.50}{5}} = 6.9210, \quad s_X = \sqrt{\frac{122.64}{5}} = 4.9526$$

$$\text{Cov}(Y, X) = \frac{153.30}{5} = 30.66$$

$$r = \frac{30.66}{6.9210 \times 4.9526} = 0.8945$$

- **Slope-ийн нэгж = Y-ийн нэгж / X-ийн нэгж.** ROA × CAPEX-д хоёулаа %-аар хэмжигдсэн учир slope = 1.25 нь "X 1%-аар нэмэгдвэл Y 1.25%-аар нэмэгдэнэ" гэсэн утгатай.
- Slope **эерэг** → X ба Y нэг чигт хөдөлдөг ($X \uparrow \rightarrow Y \uparrow$).
- Slope **сөрөг** → эсрэг чигт ($X \uparrow \rightarrow Y \downarrow$).
- **Intercept-ийн утгатай эсэхийг шалгах:** X = 0 бодит ертөнцөд боломжтой юу? CAPEX = 0 боломжтой учир $\hat{b}_0 = 4.875\%$ бодит утгатай.

ЖИШЭЭ 2 · ЭЭРЭГ БА СӨРӨГ SLOPE

Нөхцөл: Y нь сая евроор, X нь сая ам.доллороор хэмжигдэнэ.

- Хэрэв slope = **+1.2** бол:
 - ↑ USD 1 сая → ↑ EUR 1.2 сая
 - ↓ USD 1 сая → ↓ EUR 1.2 сая
- Хэрэв slope = **-1.2** бол:
 - ↑ USD 1 сая → ↓ EUR 1.2 сая
 - ↓ USD 1 сая → ↑ EUR 1.2 сая

ЖИШЭЭ 3 · ROA × CAPEX ИНТЕРПРЕТАЦ

Дээрх ROA загвар $\hat{Y} = 4.875 + 1.25X$:

- Капитал зардлаа огт хийгээгүй компанийн ROA нь дунджаар 4.875%.
- CAPEX нэг нэгжээр (жишээ нь 4%-аас 5% болж) нэмэгдвэл ROA нь 1.25%-аар нэмэгдэнэ.
- Хэрэв CAPEX = 6% бол урьдчилан таамагласан ROA = $4.875 + 1.25 \times 6 = 12.375\%$.

ДАСГАЛ БОДЛОГО · LOS 1

Бодлого 1–3 дараах мэдээлэлд хамаарна: Шинжээч компанийн ROE-ийг (return on equity) GO-той (growth opportunities — 3 жилийн compound annual growth rate of sales)-той 20 жилийн өгөгдөл дээр регресс хийсэн:

$$ROE_i = 4 + 1.8 \cdot GO_i + \varepsilon_i$$

Хоёр хувьсагч хоёулаа %-аар илэрхийлэгдсэн (GO = 5% бол утга нь 5).

Бодлого 1. Хэрэв компанийн GO = 10% бол урьдчилан таамагласан ROE хамгийн ойр нь:

A. 1.8% B. 15.8% C. 22.0%

ШИЙДЭЛ — ХАРИУЛТ С

$$\hat{Y} = 4 + 1.8 \times 10 = 4 + 18 = 22.0\%$$

Бодлого 2. GO 5%-аас 6% болж нэмэгдэхэд ROE дахь өөрчлөлт хамгийн ойр нь:

A. 1.8% B. 4.0% C. 5.8%

ШИЙДЭЛ — ХАРИУЛТ А

Slope нь X-ийн нэг нэгжийн өөрчлөлтөд Y-д гарах өөрчлөлтийг харуулна. Slope = 1.8 учир GO 1%-аар нэмэгдвэл ROE 1.8%-аар нэмэгдэнэ.

Бодлого 3. GO = 8% үед ажиглагдсан ROE = 21% бол residual хамгийн ойр нь:

A. -1.8% B. 2.6% C. 12.0%

ШИЙДЭЛ — ХАРИУЛТ В

$$\hat{Y} = 4 + 1.8 \times 8 = 18.4\%. \text{ Residual} = Y - \hat{Y} = 21 - 18.4 = 2.6\%.$$

2

Дөрвөн таамаглал ба residual plot

explain the assumptions underlying the simple linear regression model, and describe how residuals and residual plots indicate if these assumptions may have been violated

2A. Дөрвөн үндсэн таамаглал

OLS-аар үнэлсэн регрессийн загвар *statistically valid* дүгнэлтэд хүрэхийн тулд дөрвөн таамаглал биелэх ёстой. Эдгээр таамаглал зөрчигдвөл хэдий бид яг ижил тооцоог хийсэн ч үр дүн андуурч найдваргүй байж болзошгүй.

№	Таамаглал	Юу шаарддаг вэ	Зөрчигдсөн үед яах вэ
1	Linearity (шугаман шинж)	Y ба X-ийн жинхэнэ хамаарал шугаман	Загвар bias-тай → log/exp хэлбэрт шилжих
2	Homoskedasticity	Residual-ийн variance бүх ажиглалтад тэнцүү	Heteroskedasticity → robust SE ашиглах
3	Independence	Ажиглалтууд хоорондоо хамааралгүй	Autocorrelation → time-series аргууд
4	Normality	Residual-ууд нормал тархалттай	n том бол CLT-ээр хүлцэгддэг; n бага бол асуудалтай

2B. Linearity — Шугаман шинж

Эхний таамаглал нь Y ба X-ийн жинхэнэ хамаарал шугаман байх явдал. Хэрэв жинхэнэ хамаарал нь bend-тэй буюу exponential хэлбэртэй (жишээ нь $Y = b_0 \cdot e^{b_1 X}$ хэлбэртэй) бол шугамыг тэр гажил хамаарал дээр тааруулах нь эхний болон сүүлийн X-ийн утганд underestimate, дунд утгад overestimate гарах юм.

LINEARITY ЗӨРЧИГДСӨНИЙГ ХЭРХЭН ИЛРҮҮЛЭХ

- Residual plot-д X-ийн нэмэгдэлтэй холбоотой системчилсэн загвар (U-хэлбэр, n-хэлбэр) харагдаж байна уу?
- Хэрэв тийм бол residual-ууд санамсаргүй биш, харин X-тэй холбоотой — шугаман загвар тохирохгүй гэсэн дохио.
- Энэ үед Y-г лог хэлбэрт хувиргах эсвэл X^2 -г нэмэх гэх мэт **functional form**-ыг өөрчлөх хэрэгтэй.
- X нь random биш байх ёстой гэдэг нэмэлт нөхцөлийг анхаарах хэрэгтэй — random X нь шугаман хамаарлыг алдагдуулдаг.

2B. Homoskedasticity — Variance тэнцүү байх

Хоёр дахь таамаглал нь residual-ийн variance бүх ажиглалтад ижил байх явдал. Хэрэв variance ажиглалт хооронд өөр өөр бол загварт **heteroskedasticity** байгаа бөгөөд энэ нь стандарт алдааг буруу үнэлж, гипотезийн тестүүдийг найдваргүй болгодог.

НOMOSKEDASTICITY — МАТЕМАТИК ИЛЭРХИЙЛЭЛ

$$E(\varepsilon_i^2) = \sigma_\varepsilon^2, \quad i = 1, 2, \dots, n$$

Бүх i ажиглалтын хувьд error term-ийн variance ижил тогтмол σ_ε^2 утгатай байх ёстой.

Хэрэв энэ зөрчигдвөл — variance ажиглалтаас хамаарч өөрчлөгдвөл — heteroskedasticity байна.

ЖИШЭЭ 4 · ЗОХИЦУУЛАЛТЫН РЕЖИМИЙН ЖИШЭЭ

Богино хугацааны хүүний түвшин (Y) ба инфляцийн түвшин (X)-ийг 16 жилийн хугацаанд регресс хийе.

- Эхний 8 жил хэвийн ханш ("Regime 1"), residual нь ± 0.4 хүрээнд багтсан.
- Сүүлийн 8 жил төв банк хүүг хиймлээр доош нийлүүлэв ("Regime 2"), residual нь ± 1.0 хүрээнд өргөжсөн.

Дүгнэлт: Хэдий нэг л шугамыг 16 жилд тааруулсан ч residual-ууд хоёр өөр кластерт хуваагдан, Regime 2-ийн хэлбэлзэл Regime 1-ээс хамаагүй өргөн харагдана. Энэ бол тодорхой **heteroskedasticity**.

НОМОСКЕДАСТИЦИ ЗӨРЧИГДСӨНИЙГ ХЭРХЭН ИЛРҮҮЛЭХ

- Residual-уудыг X (эсвэл \hat{Y})-ийн эсрэг scatter plot-оор зурах.
- Туузны өргөн нь X -ийн дагуу нэгэн жигд биш — эхэндээ нарийн, төгсгөлд нь өргөн (эсвэл эсрэгээр) байвал — heteroskedasticity байна.
- **Statistical тест:** Breusch-Pagan, White тестүүд.
- **Засварлах:** heteroskedasticity-consistent (HC) standard errors ашиглах.

2Г. Independence — Хамаарагүй ажиглалтууд

Гурав дахь таамаглал нь Y , X хосуудын ажиглалтууд хоорондоо **uncorrelated** байх явдал. Хэрэв ажиглалт хоорондоо хамааралтай байвал residual-ууд ч хамааралтай болж — **autocorrelation** үүснэ. Энэ нь регрессийн коэффициентийн стандарт алдааг буруу болгож гипотезийн тестийг найдваргүй болгодог.

ЖИШЭЭ 5 · УЛИРЛЫН AUTOCORRELATION

Компанийн улирлын орлогыг 40 улиралд цагийн дагуу регресс хийсэн. Орлого нь улирлын мөчлөгтэй — 4-р улирал бүрд огцом өсч, дараагийн улиралд буурдаг. Residual plot-д энэ давтамж тод харагдана: 4-р улиралд эерэг spike, 1-р улиралд сөрөг — энэ нь autocorrelation-ийн тодорхой жишээ.

INDEPENDENCE ЗӨРЧИГДСӨНИЙГ ХЭРХЭН ИЛРҮҮЛЭХ

- Residual-уудыг цаг хугацаа эсвэл ажиглалтын дарааллын дагуу зурах.
- Системчилсэн дараалал — sin/cos долгион, улирлын spike, нэг чиглэлд аажмаар өсөх/буурах хандлага — independence зөрчигдсөн гэдгийг харуулна.
- **Statistical тест:** Durbin-Watson statistic. Утра 1.5-2.5 хооронд бол autocorrelation бараг үгүй.
- Зөвхөн time-series өгөгдөлд асуудал — cross-sectional өгөгдөлд (өөр өөр компани, улс) ийм асуудал гарах магадлал бага.

2Д. Normality — Residual-уудын нормал тархалт

Дөрөв дэх таамаглал нь residual-ууд нормал тархалттай байх явдал. Энэ нь Y болон X нормал тархалттай байх ёстой гэсэн утга **биш** — зөвхөн residual-ууд. Шинжээч ихэвчлэн residual-уудын histogram эсвэл Q-Q plot-оор шалгадаг.

NORMALITY-ИЙН ТУХАЙ ПРАКТИК ЗӨВЛӨМЖ

- $n \geq 30$ бол CLT-ээр residual-ууд approximately normal болдог — асуудалгүй.
- $n < 20$ ба residual-ууд хүчтэй хазайлттай эсвэл fat-tailed бол t-test, F-test үр дүн найдваргүй болно.
- Outlier нь residual-ыг нормалаас хол байдалтай болгож загварыг гажуудуулдаг — outlier-уудыг шалгах нь сайн дадлага.
- **Тест:** Shapiro-Wilk, Jarque-Bera. Q-Q plot нь visual үнэлгээний хамгийн хүчирхэг хэрэгсэл — нормал бол цэгүүд шулуун шугам дээр унана.

2E. Residual plot — Загварыг оношлох гол хэрэгсэл

Дөрвөн таамаглал биелэх эсэхийг шалгахад хамгийн хүчирхэг хэрэгсэл нь **residual plot**. Шинжээч e_i -уудыг (1) X-ийн эсрэг, (2) \hat{Y} -ийн эсрэг, (3) цаг хугацааны дагуу зурж шалгана.

Plot-д юу харагдвал	Аль таамаглал зөрчигдсөн	Юу хийх вэ
U-хэлбэрийн дараалал	Linearity	Y эсвэл X-ийг log хэлбэрт шилжүүлэх
Туузны өргөн өсөж байна	Homoskedasticity	Robust standard errors ашиглах
Цаг хугацааны дагуу мөчлөг	Independence	Time-series загвар (AR, ARIMA)
Outlier — туйл утга	Normality	Outlier-ыг шалгах, шаардлагатай бол хасах
Санамсаргүй чимээ	Бүх таамаглал ОК	Үргэлжлүүлэх — загвар сайн

ДАСГАЛ БОДЛОГО · LOS 2

Бодлого 4. Homoskedasticity-ийн хамгийн зөв тодорхойлолт нь:

- A. Регрессийн residual-ийн variance тэгтэй тэнцүү.
- B. Residual-ийн variance нь нормал тархалттай.
- C. Residual-ийн variance бүх ажиглалтад тогтмол.

ШИЙДЭЛ — ХАРИУЛТ С

Homoskedasticity нь $E(\epsilon_i^2) = \sigma_\epsilon^2$ — error term-ийн variance бүх ажиглалтад тогтмол байх явдал. А буруу: variance тэг биш, тогтмол. В буруу: residual нормал байх таамаглал нь тусдаа (normality) шаардлага.

Бодлого 5. Шинжээч компанийн улирлын орлогыг цаг хугацааны дагуу регресс хийхэд residual plot-д 4 улирал тутамд эерэг spike, 1-р улиралд сөрөг spike тогтмол давтагдаж байгааг ажиглав. Энэ нь аль таамаглалын зөрчлийг харуулж байна вэ?

- A. Linearity
- B. Homoskedasticity
- C. Independence

ШИЙДЭЛ — ХАРИУЛТ С

Residual-ийн системчилсэн улирлын дараалал нь autocorrelation байгааг харуулна — энэ нь independence таамаглалын зөрчил. Шийдэл: time-series загвар (AR, ARIMA, эсвэл seasonal dummy variable) ашиглах.

Бодлого 6. Хөрөнгө оруулагч 16 жилийн хугацаанд хүү ба инфляцийн регресс хийгээд эхний 8 жилийн residual нь ± 0.4 , сүүлийн 8 жилд ± 1.0 болж нэмэгдсэнийг анзаарлаа. Энэ нь юу гэдгийг харуулна вэ, мөн шийдэл нь юу вэ?

- A. Autocorrelation байна — AR загвар ашиглах.
- B. Heteroskedasticity байна — robust (HC) standard errors ашиглах.
- C. Linearity зөрчигдсөн — лог хэлбэрт шилжих.

ШИЙДЭЛ — ХАРИУЛТ В

Residual-ийн variance цаг хугацаанд нэмэгдэж байна — энэ нь homoskedasticity-ийн зөрчил буюу heteroskedasticity. Шийдэл: heteroskedasticity-consistent (HC) standard errors-ыг ашиглах эсвэл хоёр режим хооронд dummy variable ашиглах.

3

Тааруулалтын хэмжүүр ба тестүүд

calculate and interpret measures of fit and formulate and evaluate tests of fit and of regression coefficients in a simple linear regression

3А. Variance-ийн задаргаа: $SST = SSR + SSE$

Регрессийн загвар хэр сайн тааруулсан болохыг үнэлэхийн тулд эхлээд Y -ийн нийт хэлбэлзлийг хоёр хэсэгт хуваана. Y -ийн нийт хэлбэлзэл нь "тайлбарлагдсан" ба "тайлбарлагдаагүй" хоёр хэсгээс бүрдэнэ.

VARIANCE-ИЙН ЗАДАРГАА

$$\underbrace{\sum_{i=1}^n (Y_i - \bar{Y})^2}_{SST} = \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{SSR} + \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{SSE}$$

$$\boxed{SST = SSR + SSE}$$

SST — Total Sum of Squares: Y -ийн нийт хэлбэлзэл (бид тайлбарлахыг хүсэж байна).

SSR — Sum of Squares Regression: регрессийн шугам тайлбарласан хэсэг.

SSE — Sum of Squares Error: residual-уудын квадратын нийлбэр (тайлбарлагдаагүй хэсэг).

ROA × CAPEX жишээнд: $SST = 239.50$, $SSE = 47.875$, $SSR = 191.625$. Өөрөөр хэлбэл Y -ийн нийт хэлбэлзлийн $191.625/239.50 = 80.01\%$ -ийг CAPEX тайлбарлаж чадсан байна.

3Б. R^2 — Determination-ы коэффициент

Coefficient of determination буюу R^2 нь Y -ийн хэлбэлзлийн хэдэн хувийг X тайлбарласныг харуулдаг. Утга нь 0%-аас 100% хооронд.

R^2 — COEFFICIENT OF DETERMINATION

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$R^2 = 0$ → загвар Y -г огт тайлбарлаагүй (residual-ууд дунджаас өөр зүйл биш).

$R^2 = 1$ → загвар Y -г төгс тайлбарласан (бүх ажиглалт шугам дээр).

Simple linear regression-д: $R^2 = r^2$ (Pearson r -ийн квадрат).

ROA жишээнд: $r = 0.8945 \rightarrow r^2 = 0.8001 = R^2$.

R^2 -ЫГ ХЭТ ӨНДӨР ШҮТЭХГҮЙ БАЙХ

- R^2 өндөр гэдэг "загвар үнэн" гэсэн утга биш. Time-series-д trend-trend регрессийн R^2 хиймлээр өндөр гарч магадгүй.

- R^2 бага гэдэг "загвар муу" гэсэн утга биш. Хүний зан төлвийг тайлбарлах загварт $R^2 = 0.30$ ч өндөр гэж тооцогддог.
- **Бизнесийн ач холбогдол \neq statistical significance.** $R^2 = 0.05$ ч n том бол significant болж магадгүй — гэхдээ практикт хэрэглэхэд бараг боломжгүй.
- Effect size, residual plot, theoretical justification — гурвыг хамтад нь үнэлэх хэрэгтэй.

ЗВ. F-test — Загварын тааруулалтын тест

R^2 нь descriptive хэмжүүр — статистикийн тест биш. Загвар нийтдээ ач холбогдолтой эсэхийг тогтоохын тулд **F-test** ашиглана. Simple linear regression-д F-test-ийн null hypothesis нь slope тэгтэй тэнцүү гэсэн санаа.

F-STATISTIC — ТААРУУЛАЛТЫН ТЕСТ

$$F = \frac{MSR}{MSE}$$

$$MSR = \frac{SSR}{k}, \quad MSE = \frac{SSE}{n - k - 1}$$

k — independent variable-ийн тоо. SLR-д $k = 1$, иймд $MSR = SSR$ ба $MSE = SSE/(n - 2)$.

Null hypothesis: $H_0 : b_1 = 0$ vs $H_a : b_1 \neq 0$.

F-distribution: $df_1 = k = 1$, $df_2 = n - 2$.

F-test нь үргэлж **right-tailed** — F нь үргэлж эерэг учраас.

F том → SSR харьцангуй том → загвар тайлбарлаж байна.

ROA × CAPEX жишээнд:

$$MSR = 191.625, \quad MSE = \frac{47.875}{4} = 11.96875$$

$$F = \frac{191.625}{11.96875} = 16.0104$$

Critical F (df = 1, 4; $\alpha = 5\%$) = 7.71. $16.0104 > 7.71$ учир H_0 -ыг татгалзана: загвар **statistically significant**.

ЗГ. t-test — Slope-ийн коэффициентийн тест

F-test нь "slope тэг үү?" гэсэн ерөнхий асуултанд хариулдаг. Гэхдээ бид заримдаа "slope нь 1.0 уу?" эсвэл "slope эерэг үү?" гэсэн илүү тодорхой асуулт асуухыг хүсдэг — энэ үед **t-test** ашиглана.

SLOPE-ИЙН T-TEST

$$t = \frac{\hat{b}_1 - B_1}{s_{\hat{b}_1}}$$

$$s_{\hat{b}_1} = \frac{s_e}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} \quad df = n - 2$$

B_1 — null hypothesis-ийн тогтоосон утга (ихэвчлэн 0).

$s_{\hat{b}_1}$ — slope-ийн стандарт алдаа.

$s_e = \sqrt{MSE}$ — стандарт алдааны үнэлгээ.

ROA жишээнд:

$$s_e = \sqrt{11.96875} = 3.4596$$

$$s_{\hat{b}_1} = \frac{3.4596}{\sqrt{122.64}} = 0.3124$$

$$t = \frac{1.25 - 0}{0.3124} = 4.0013$$

$df = 4$, critical $t = \pm 2.776$. $|4.0013| > 2.776$ учир H_0 -ыг татгалзана. Slope нь статистикийн ач холбогдолтой ялгаатай тэгээс.

F-TEST БА T-TEST ХООРОНДЫН ХОЛБОО

- Simple linear regression-д $t^2 = F$. ROA жишээнд: $4.0013^2 = 16.0104$.
- Хоёр тест ижил мэдээлэл өгдөг — $H_0 : b_1 = 0$ -ийн тухай.
- Илүү ерөнхий тестүүдэд (slope = 1.0 уу, slope > 0 уу гэх мэт) **t-test л** хэрэглэдэг.
- Multiple regression-д F-test нь "бүх slope-ууд тэг үү" гэсэн ерөнхий асуултанд, t-test нь нэг бүрчлэн хариулдаг.

3Д. Correlation-ийн тест: $r = 0$ уу?

Сонирхолтой шинж: SLR-д $b_1 = 0$ эсэхийг туршихад ашигладаг t-statistic нь Pearson $r = 0$ эсэхийг туршихад ашигладаг t-statistic-тэй яг ижил утга өгдөг.

CORRELATION-ИЙН T-TEST

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \quad df = n - 2$$

$H_0 : r = 0$ vs $H_a : r \neq 0$.

ROA жишээ: $r = 0.8945$, $n = 6$.

$$t = \frac{0.8945 \times \sqrt{4}}{\sqrt{1 - 0.8001}} = \frac{1.789}{0.4471} = 4.0013$$

Slope-ийн t-statistic-тэй яг ижил → ижил дүгнэлт.

3Е. Intercept-ийн тест

Бид зарим тохиолдолд intercept нь тогтоосон утгаас ялгаатай эсэхийг шалгахыг хүсдэг. Жишээ нь ROA загварт: "CAPEX = 0 байсан ч ROA нь 3%-аас илүү байх уу?"

INTERCEPT-ИЙН T-TEST

$$t = \frac{\hat{b}_0 - B_0}{s_{\hat{b}_0}}$$

$$s_{\hat{b}_0} = s_e \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}, \quad df = n - 2$$

B_0 — null hypothesis-ийн тогтоосон intercept-ийн утга.

$s_{\hat{b}_0}$ нь slope-ийн стандарт алдаанаас өөр томъёотой — учир нь intercept нь \bar{X} -аас хамаарна.

ДАСГАЛ БОДЛОГО · LOS 3

Бодлого 7–9 — McCoin restaurant жишээ: X нь market cap >

100Мрестораныцэвэршиг/орлогынхарьцаа, Y ньүйлажиллагаанымөнгөнгүйлгээ/орлогынхарьцаа. $n = 24$, $R^2 = 0.7436$, slope = 0.826, intercept = 0.077, $F = 63.81$, p-value (F ба slope-ийн t-test) < 0.05.

Бодлого 7. X ба Y -ийн correlation хамгийн ойр нь:

- A. -0.7436 B. 0.7436 C. 0.8623

ШИЙДЭЛ — ХАРИУЛТ С

SLR-д $R^2 = r^2$. Slope эерэг (+0.826) учир r ч мөн эерэг. $r = \sqrt{0.7436} = 0.8623$.

Бодлого 8. Хэрэв ресторан цэвэр ашиг/орлогын харьцаа 5% бол урьдчилан таамагласан CFO/sales нь хамгийн ойр нь:

- A. -4.054 B. 0.524 C. 4.207

ШИЙДЭЛ — ХАРИУЛТ С

$$\hat{Y} = 0.077 + 0.826 \times 5 = 0.077 + 4.130 = 4.207.$$

Бодлого 9. CFO/sales болон цэвэр ашиг/орлогын хамаарал 0.05 түвшинд statistically significant эсэх нь:

- A. Үгүй, учир нь $R^2 > 0.05$.
B. Үгүй, учир нь intercept ба slope-ийн p-values 0.05-аас бага.
C. Тийм, учир нь F ба slope-ийн t-ийн p-values 0.05-аас бага.

ШИЙДЭЛ — ХАРИУЛТ С

Хамаарал significant эсэхийг F-test эсвэл slope-ийн t-test-ээр шалгана. $F = 63.81$ -ийн p-value < 0.05 , slope t-statistic = 7.988-ийн p-value < 0.05 учир хоёр тест хоёулаа хамаарал significant байгааг батлана. А буруу: R^2 нь significance тестийн утга биш. В буруу: p-value < 0.05 нь significance байгааг харуулна (татгалзах).

4

ANOVA ба standard error of estimate

describe the use of analysis of variance (ANOVA) in regression analysis, interpret ANOVA results, and calculate and interpret the standard error of estimate in a simple linear regression

4A. ANOVA хүснэгт

Регрессийн загварын variance-ийн задаргааг ANOVA (Analysis of Variance) хүснэгтэд цэгцтэйгээр харуулдаг. Энэ хүснэгт нь sum of squares, degrees of freedom, mean squares, ба F-statistic-ийг нэгтгэдэг. Excel, R, Python зэрэг бүх стандарт пакет регрессийн гаралтын нэг хэсэг болгон ANOVA хүснэгтийг гаргадаг.

Source	Sum of Squares	df	Mean Square	F-statistic
Regression	SSR	1 (= k)	$MSR = SSR/1$	$F = MSR/MSE$
Error	SSE	$n - 2$	$MSE = SSE/(n - 2)$	—
Total	SST	$n - 1$	—	—

ROA × CAPEX жишээ дээр ANOVA хүснэгтийг бөглөө:

Source	Sum of Squares	df	Mean Square	F-statistic
Regression	191.625	1	191.625	16.0104
Error	47.875	4	11.96875	—
Total	239.50	5	—	—

4Б. Standard error of estimate (s_e)

Standard error of estimate (s_e) буюу регрессийн стандарт алдаа нь Y-ийн ажиглагдсан утгууд ба регрессийн шугамаар таамагласан утгуудын хоорондох "дундаж" зайг хэмждэг. Бага утгатай байх тусам загвар сайн тааруулсан гэсэн утга.

STANDARD ERROR OF ESTIMATE

$$s_e = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n - 2}} = \sqrt{\frac{SSE}{n - 2}}$$

Y-ийн нэгжтэй ижил нэгжтэй (ROA загварт %).

R^2 -аас ялгаатай нь — s_e нь absolute хэмжүүр; R^2 нь relative.

Prediction interval тооцоход хэрэглэдэг гол утга.

ROA жишээнд: $s_e = \sqrt{11.96875} = 3.4596 \rightarrow$ дундаж prediction error нь $\pm 3.46\%$ орчим ROA.

ГУРВАН ХЭМЖҮҮРИЙГ ХЭРХЭН НИЙЦҮҮЛЭХ

- R^2 (relative) — "загвар Y -ийн хэдэн хувийг тайлбарласан бэ?"
- F-statistic (тест) — "загвар statistically significant юу?"
- s_e (absolute) — "prediction-ийн хэдэн % алдаатай байх вэ?"
- Сайн загвар: R^2 өндөр, F significant, s_e бага. Гурвыг хамт үнэлэх ёстой.
- ROA жишээнд: $R^2 = 80.01\%$ (өндөр), $F = 16.01 > 7.71$ (significant), $s_e = 3.46\%$ (Y -ийн дундаж 12.5%-тай харьцуулахад харьцангуй бага) — нийтдээ сайн загвар.

ЖИШЭЭ 6 · LIU SHORT INTEREST × DEBT RATIO

Шинжээч 50 компанийн short interest ratio (Y) ба debt ratio (X) хоорондын регресс хийсэн. ANOVA хүснэгт:

Source	df	SS	MS
Regression	1	38.4404	38.4404
Residual	48	373.7638	7.7867
Total	49	412.2042	—

F-statistic: $F = 38.4404/7.7867 = 4.937$.

R^2 : $R^2 = 38.4404/412.2042 = 0.0933$ — debt ratio short interest-ийн зөвхөн 9.3%-ийг тайлбарлаж байна.

s_e : $s_e = \sqrt{7.7867} = 2.7905$.

ДАСГАЛ БОДЛОГО · LOS 4

Бодлого 10–12 нь Liu-ийн short interest жишээнд хамаарна ($n = 50$, slope = -4.1589 , intercept = 5.4975 , critical t ($df=48$, $\alpha=0.05$, two-sided) = ± 2.011).

Бодлого 10. Slope-ийн t-test-д ашиглах degrees of freedom:

- A. 48 B. 49 C. 50

ШИЙДЭЛ — ХАРИУЛТ А

Slope-ийн t-test-д $df = n - 2 = 50 - 2 = 48$.

Бодлого 11. Slope коэффициент 0.05 түвшинд тэгээс ялгаатай эсэхийг шалгахад дүгнэлт нь:

- A. Slope тэгээс ялгаатай (татгалзана), учир нь $|t| = 2.222 > 2.011$.
 B. Slope тэгээс ялгаагүй (татгалзахгүй), учир нь $|t| = 2.222 > 2.011$.
 C. Тэс хийх боломжгүй — мэдээлэл дутуу.

ШИЙДЭЛ — ХАРИУЛТ А

$t = \hat{b}_1 / s_{\hat{b}_1} = -4.1589 / 1.8718 = -2.222$. $|-2.222| = 2.222 > 2.011$ учир H_0 -ыг 0.05 түвшинд татгалзана. Slope тэгээс ялгаатай.

Бодлого 12. ANOVA хүснэгтээс F-statistic хамгийн ойр нь:

A. -2.222 B. 3.504 C. 4.937

ШИЙДЭЛ — ХАРИУЛТ С

$F = MSR/MSE = 38.4404/7.7867 = 4.937$. Мөн $t^2 = (-2.222)^2 = 4.937$ нь F-тэй тэнцүү — SLR-д $t^2 = F$. (A нь t-statistic; B-ийн утгад зүй ёсны үндэс байхгүй.)

5 Прогноз ба prediction interval

calculate and interpret the predicted value for the dependent variable, and a prediction interval for it, given an estimated linear regression model and a value for the independent variable

5А. Точкийн прогноз

Регрессийн загварыг бүтээсний дараа бид түүгээр шинэ X -ийн утганд тааруулан Y -ийг таамаглах боломжтой. Энэ нь зүгээр л үнэлэгдсэн шугаман загварт X_f -ийг оруулах юм.

ТОЧКИЙН ПРОГНОЗ

$$\hat{Y}_f = \hat{b}_0 + \hat{b}_1 X_f$$

X_f — таамаглахыг хүсэж буй future эсвэл hypothetical X -ийн утга.

\hat{Y}_f — Y -ийн таамагласан утга.

Жишээ: ROA загвар $\hat{Y} = 4.875 + 1.25X$. CAPEX = 6%-ийн үед: $\hat{Y}_f = 4.875 + 1.25 \times 6 = 12.375\%$.

5Б. Standard error of forecast (s_f)

Точкийн прогноз нь зөвхөн нэг тоо — гэхдээ загвар нь төгс тааруулсангүй учир бодит Y нь \hat{Y}_f -ээс зөрж болно. Энэ хэлбэлзлийг илэрхийлэхэд бид prediction interval хэрэглэдэг — гэхдээ эхлээд **standard error of forecast (s_f)**-ийг тооцох ёстой.

STANDARD ERROR OF FORECAST

$$s_f^2 = s_e^2 \left[1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right]$$

$$s_f = s_e \sqrt{1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

s_e — стандарт алдааны үнэлгээ (regression-аас).

n — түүврийн хэмжээ.

$(X_f - \bar{X})^2$ — X_f нь түүврийн дунджаас хэр хол байгааг харуулна.

$\sum (X_i - \bar{X})^2$ — X -ийн нийт хэлбэлзэл.

s_f -ИЙН ГУРВАН ЧУХАЛ ШИНЖ

- Загвар сайн тааруулсан байх тусам s_e бага → s_f бага.
- Түүврийн хэмжээ n том байх тусам s_f бага — $1/n$ гишүүн багасна.
- X_f нь түүврийн дунджид ойртох тусам s_f бага — $(X_f - \bar{X})^2$ гишүүн багасна.

- Энэ нь яагаад "extrapolation" — өгөгдлийн хүрээнээс гадуур таамаглах нь — эрсдэлтэй болохыг тайлбарладаг.

5B. Prediction interval

Standard error of forecast-ийг ашиглан prediction interval-ийг бүтээнэ. Энэ нь t-distribution дээр тулгуурласан confidence interval-ын хэлбэртэй.

$(1 - \alpha)$ PREDICTION INTERVAL

$$\hat{Y}_f \pm t_{\text{critical}, \alpha/2} \cdot s_f$$

α — significance level (ихэвчлэн 5%, 95% PI).

t_{critical} — $df = n - 2$ -той t-distribution-аас critical value.

Энэ интервал нь "бодит Y_f нь энэ хүрээнд тогтох магадлал $(1 - \alpha)$ " гэсэн утгатай.

ЖИШЭЭ 7 · ROA ЗАГВАР, $X_f = 6\%$ — 95% PI

ROA загвар: $\hat{Y} = 4.875 + 1.25X$. $s_e = 3.4596$, $n = 6$, $\bar{X} = 6.1$, $\sum(X_i - \bar{X})^2 = 122.64$.
CAPEX = 6%-ийн үед 95% prediction interval-ийг тооцъё.

Алхам	Тооцоо
1. Точкийн прогноз	$\hat{Y}_f = 4.875 + 1.25 \times 6 = 12.375$
2. $(X_f - \bar{X})^2$	$(6 - 6.1)^2 = 0.01$
3. s_f^2	$3.4596^2 \cdot \left[1 + \frac{1}{6} + \frac{0.01}{122.64}\right] = 11.969 \times 1.16675 = 13.964$
4. s_f	$\sqrt{13.964} = 3.7369$
5. Critical t ($df = 4$, $\alpha = 5\%$)	$t = \pm 2.776$
6. Prediction interval	$12.375 \pm 2.776 \times 3.7369 = 12.375 \pm 10.374$

Үр дүн: 95% PI = [2.001%, 22.749%].

ҮР ДҮНГИЙН ТАЙЛБАР

- $X_f = 6$ нь $\bar{X} = 6.1$ -тэй маш ойр учир s_f нь s_e -тэй ойролцоо ($3.74 \approx 3.46$).
- PI нь маш өргөн ($\pm 10\%$) — учир нь $n = 6$ жижиг ба critical $t = 2.776$ том.
- Энэ нь регрессийн жинхэнэ practical хязгаар: жижиг түүврээс гарсан прогноз маш бага нарийвчлалтай байж магадгүй.
- **Бизнесийн хэрэглээнд:** загварыг ашиглахаасаа өмнө prediction interval-ийг үргэлж тооцож, бодит шийдвэр гаргахад хангалттай нарийвчлалтай эсэхийг шалгах хэрэгтэй.

ЖИШЭЭ 8 · NPM × RDR — EXTRAPOLATION ЭРСДЭЛ

8 компанийн NPM (Y) ба RDR (X) — slope = -1.3, intercept = 16.5, $s_e = 1.86190$, $\bar{X} = 7.5$,
 $\sum(X_i - \bar{X})^2 = 30$. RDR = 5 болон RDR = 15-ийн үед 95% PI-ийг харьцуулъя ($df = 6$, critical $t = \pm 2.447$).

Тооцоо	$X_f = 5$	$X_f = 15$
\hat{Y}_f	$16.5 - 1.3 \times 5 = 10.0$	$16.5 - 1.3 \times 15 = -3.0$
$(X_f - \bar{X})^2$	$(5 - 7.5)^2 = 6.25$	$(15 - 7.5)^2 = 56.25$
s_f	$1.8619 \sqrt{1 + \frac{1}{8} + \frac{6.25}{30}} = 2.150$	$1.8619 \sqrt{1 + \frac{1}{8} + \frac{56.25}{30}} = 3.225$
95% PI	[4.74, 15.26] — өргөн $\approx \pm 5.26$	[-10.89, 4.89] — өргөн $\approx \pm 7.89$

Дүгнэлт: $X_f = 15$ нь дунджаас хол байгаа учир PI 50%-аар өргөжсөн. PI нь тэгийг давсан — "Y эерэг үү, сөрөг үү?" гэдэг асуултанд найдвартай хариулж чадахгүй.

ЖИШЭЭ 9 · VASILEVA AMTEX × CRUDE OIL

Шинжээч Amtex (нефтийн компанийн) хувьцааны өгөөж (Y) ба crude oil-ийн өгөөжийг (X) сар бүрийн өгөгдлөөр (n = 36) регресс хийж дараах үр дүн авав:

- Intercept = 0.0095, Slope = 0.2354.
- $\sum(Y_i - \hat{Y}_i)^2 = 0.071475$, $\bar{X} = -0.018056$.
- Standard error of estimate: $s_e = \sqrt{0.071475/(36 - 2)} = \sqrt{0.002102} = 0.04585$.
- Сарын 37-д crude oil өгөөж $X_f = -0.01$ гэж таамаглаж байгаа.

Точкийн прогноз: $\hat{Y}_f = 0.0095 + 0.2354 \times (-0.01) = 0.00715$ буюу 0.715%.

s_f өгөгдсөн: 0.0469.

99% PI: $0.00715 \pm 2.728 \times 0.0469 = 0.00715 \pm 0.12796 = [-0.121, 0.135]$.

ДАСГАЛ БОДЛОГО · LOS 5

Бодлого 13. Шинжээч $ROA = 4.875 + 1.25 \cdot CAPEX$ загварыг ашиглан $CAPEX = 8\%$ үед ROA-ийн point estimate-ыг тооцов. Хариулт хамгийн ойр нь:

- A. 10.0% B. 14.875% C. 17.625%

ШИЙДЭЛ — ХАРИУЛТ В

$$\hat{Y}_f = 4.875 + 1.25 \times 8 = 4.875 + 10.0 = 14.875\%$$

Бодлого 14. Vasileva Amtex загварт slope = 0.2354, $s_{\hat{b}_1} = 0.0760$. Slope нь 0.15-аас илүү гэдэг таамаглалыг 1% түвшинд (one-sided right, critical $t = +2.441$) шалгахад дүгнэлт нь:

- A. Татгалзана — slope 0.15-аас илүү байна.
B. Татгалзахгүй — мэдээлэл хангалтгүй.
C. Татгалзана — slope 0-ээс ялгаатай.

ШИЙДЭЛ — ХАРИУЛТ В

$H_0 : b_1 \leq 0.15$ vs $H_a : b_1 > 0.15$. $t = (0.2354 - 0.15)/0.0760 = 0.0854/0.0760 = 1.124$.
 $1.124 < 2.441$ — татгалзахгүй. Бид slope нь 0.15-аас илүү гэж дүгнэх хангалттай нотолгоо байхгүй.

Бодлого 15. Дараах хоёр ажиглалтын аль нь зөв вэ?

Ажиглалт 1: Тогтсон confidence level-д prediction interval нь X_f -ээс үл хамаарна.

Ажиглалт 2: s_e том бол confidence interval илүү өргөн болно.

- A. Зөвхөн Ажиглалт 1.
B. Зөвхөн Ажиглалт 2.
C. Хоёулаа.

ШИЙДЭЛ — ХАРИУЛТ В

Ажиглалт 1 буруу: X_f нь \bar{X} -аас холдох тусам s_f нэмэгдэж PI өргөн болно — иймд PI-ийн өргөн нь X_f -ээс шууд хамаардаг. Ажиглалт 2 зөв: s_e -ийн томрол шууд s_f -ийг томруулж өргөн PI үүсгэдэг.

6

Log-lin, lin-log, log-log хэлбэрүүд

describe different functional forms of simple linear regressions

6А. Яагаад функциональ хэлбэр өөрчлөх вэ

Бодит ертөнцийн өгөгдөл нь ихэвчлэн шугаман шинжтэй биш — компанийн орлого цаг хугацаандаа exponential байдлаар өсдөг, амьжиргааны түвшин ба аз жаргал ойролцоох logarithmic хамаарал, гэх мэт. Иймд бид Y эсвэл X эсвэл хоёуланг logarithm руу хувиргаснаар шугаман загварыг шугаман бус өгөгдөлд тааруулж болдог.

Хэлбэр	Y	X	Загвар	Slope-ийн утга
Lin-lin (стандарт)	Y	X	$Y = b_0 + b_1 X$	X нэг нэгж → Y absolute өөрчлөлт
Log-lin	$\ln Y$	X	$\ln Y = b_0 + b_1 X$	X нэг нэгж → Y % өөрчлөлт
Lin-log	Y	$\ln X$	$Y = b_0 + b_1 \ln X$	X 1% → Y absolute өөрчлөлт
Log-log	$\ln Y$	$\ln X$	$\ln Y = b_0 + b_1 \ln X$	Elasticity — X 1% → Y %

6Б. Log-lin загвар

Log-lin загварт зөвхөн хамаарлын хувьсагч Y -г logarithm хэлбэрт оруулна. Энэ нь тогтмол хувиар өсдөг (exponential) өгөгдөлд маш сайн тохирдог.

LOG-LIN ЗАГВАР

$$\ln Y_i = b_0 + b_1 X_i$$

b_1 — X -ийн нэг нэгж өөрчлөгдөхөд Y -ийн \ln -д гарах өөрчлөлт.

Практик утгаар: "X нэг нэгж нэмэгдвэл Y нь $b_1 \times 100\%$ орчмоор өсдөг".

Прогноз: $Y_f = e^{\hat{b}_0 + \hat{b}_1 X_f}$ — antilog авах ёстой.

Жишээ: $\ln Y = 4.7932 + 0.1087X$ — X нэг нэгж нэмэгдвэл Y дунджаар 10.87% орчим өсдөг.

LOG-LIN ХЭРЭГТЭЙ ЖИШЭЭНҮҮД

- Компанийн орлого/ашиг цаг хугацаагаар exponential өсөж байгаа.
- Хүн амын тоо, зээлийн үлдэгдэл, хадгаламжийн өсөлт — compound growth.
- Технологийн дамжуулагч (broadband, internet penetration) хадгалагдах үед logistic-д ойрхон.
- Compound interest-тэй холбоотой бүх явц.

6B. Lin-log загвар

Lin-log загварт зөвхөн үл хамаарлын хувьсагч X -ийг logarithm хэлбэрт оруулна. Y нь хэвээрээ үлдэнэ.

LIN-LOG ЗАГВАР

$$Y_i = b_0 + b_1 \ln X_i$$

b_1 — $\ln X$ -ийн нэг нэгж өөрчлөгдөхөд (X -ийн e дахин нэмэгдвэл) Y -д гарах absolute өөрчлөлт.

Практик утгаар: "X 1%-аар нэмэгдвэл Y нь $b_1/100$ нэгжээр өөрчлөгдөнө".

Жишээ: $Y = 2.6286 + 1.1797 \ln X$ — X 1%-аар нэмэгдэхэд Y нь $1.1797/100 = 0.0118$ нэгжээр өсдөг.

LIN-LOG ХЭРЭГТЭЙ ЖИШЭЭНҮҮД

- Орлого ба эдийн засгийн чанар — орлого 2 дахин нэмэгдвэл аз жаргал тогтмол утгаар нэмэгддэг.
- Operating profit margin ба unit sales — энд жинхэнэ жишээ $R^2 = 55\%$ -аас 97%-д үсэрсэн.
- Ширхэгийн зардлын муруй — үйлдвэрлэлийн хэмжээ хэдэн дахин нэмэгдэхэд зардал нэгж өсдөг.

6Г. Log-log загвар (double-log)

Log-log загварт хоёр хувьсагчийг хоёуланг logarithm хэлбэрт оруулдаг. Энэ нь **elasticity**-ыг шууд өгдөг учир эдийн засагт маш түгээмэл.

LOG-LOG ЗАГВАР (DOUBLE-LOG)

$$\ln Y_i = b_0 + b_1 \ln X_i$$

b_1 нь **elasticity** — X 1%-аар нэмэгдэхэд Y хэдэн % нэмэгдэхийг шууд харуулдаг.

Энэ загвар Cobb-Douglas функц, demand curve, supply curve-д стандарт хэрэглэгддэг.

Жишээ: $\ln \text{Revenue} = 5.31 + 1.37 \ln \text{ADVERT}$ — реклам зардал 1%-аар нэмэгдэхэд орлого 1.37%-аар өсдөг.

ЖИШЭЭ 10 · ГУРВАН ЗАГВАРЫГ ХАРЬЦУУЛАХ

Жишээ загвар	Lin-lin R^2	Log-хэлбэр R^2	Сонгосон шалтгаан
Орлого vs цаг (28 улирал)	—	0.9808 (log-lin)	Exponential өсөлт
Operating margin vs unit sales	0.5510	0.9717 (lin-log)	Diminishing return
Revenue vs advertising	0.2089	0.8491 (log-log)	Elasticity-ийн шинж

Гурван жишээнд хувиргалт хийсний дараа R^2 маш ихээр сайжирсан — энэ нь анхны өгөгдлийн чанар нь шугаман бус хамаарал болохыг харуулна.

ФУНКЦИОНАЛЬ ХЭЛБЭРИЙГ ХЭРХЭН СОНГОХ

- Эхлээд lin-lin (хэвийн) регресс хийж R^2 , F-statistic, residual plot-ыг харах.
- Хэрэв residual plot-д systematic дараалал (U , ρ) харагдвал — функциональ хэлбэр өөрчилж үзэх хэрэгтэй.
- Хувиргалтын дараа R^2 өсөж, residual-ууд илүү санамсаргүй болж байгааг шалгах.
- **Анхаар:** log-lin ба lin-lin загваруудыг R^2 шууд харьцуулах боломжгүй — Y -ийн хэлбэр өөр учраас. s_e -аар харьцуулахад илүү найдвартай.
- **Theory важна:** эдийн засгийн хувьд elasticity (log-log) утгатай юу? Compound growth (log-lin) утгатай юу? Сонголт нь зөвхөн statistical биш — theoretical justification-той байх ёстой.

ДАСГАЛ БОДЛОГО · LOS 6

Бодлого 16. Шинжээч компанийн орлогыг сараар exponential байдлаар өсч байгааг ажигласан. Аль функциональ хэлбэрийг хэрэглэх нь хамгийн тохиромжтой вэ?

- A. Lin-lin: $Y = b_0 + b_1X$
- B. Log-lin: $\ln Y = b_0 + b_1X$
- C. Log-log: $\ln Y = b_0 + b_1 \ln X$

ШИЙДЭЛ — ХАРИУЛТ В

Compound (тогтмол хувиар) өсөлт нь exponential шинжтэй. Үүнийг шугаман загварт шилжүүлэхэд Y-г лог хэлбэрт оруулна → log-lin загвар. Slope нь "X нэг сар нэмэгдэхэд Y хэдэн %-аар өсөх вэ" гэдгийг харуулна. (C нь elasticity, lin-log төстэй боловч X-ийг лог хэлбэрт хувиргах нь cdf орлогод утгагүй — учир нь сар бол үргэлжилсэн утга биш индекс.)

Бодлого 17. Эдийн засагч компанийн advertising зардал ба revenue-ийн хооронд elasticity-ыг хэмжихийг хүсэж байна. Загвар нь $\ln Rev = 5.31 + 1.37 \ln Adv$. Slope коэффициентийн утгыг хэрхэн тайлбарлах вэ?

- A. Adv 1 нэмэгдэхэд Rev 1.37 нэмэгдэнэ.
- B. Adv 1%-аар нэмэгдэхэд Rev 1.37%-аар нэмэгдэнэ.
- C. Adv 1 нэмэгдэхэд Rev 1.37%-аар нэмэгдэнэ.

ШИЙДЭЛ — ХАРИУЛТ В

Log-log загварт slope нь шууд elasticity. Adv 1%-аар нэмэгдэхэд Rev нь slope × 1% = 1.37% нэмэгдэнэ. Энэ загвар нь Cobb-Douglas функцэд харгалзана.

Бодлого 18. Шинжээч хоёр загвар бүтээсэн. Lin-lin загварт $R^2 = 0.55$, $s_e = 1.2$. Log-lin загварт $R^2 = 0.97$, $s_e = 0.04$. Энэ хоёр загварыг шууд R^2 -аар харьцуулж сонголт хийх боломжтой юу?

- A. Тийм, log-lin илүү тохиромжтой — R^2 өндөр.
- B. Үгүй, R^2 шууд харьцуулах боломжгүй — Y-ийн хэлбэр өөр.
- C. Тийм, lin-lin илүү тохиромжтой — s_e их.

ШИЙДЭЛ — ХАРИУЛТ В

Lin-lin-д Y нь анхны нэгжээр, log-lin-д $\ln Y$ хэлбэрээр явсан учир R^2 нь өөр өөр scale-тай хэлбэлзлийг тайлбарлаж буй. Шууд харьцуулахад буруу дүгнэлтэд хүрч магадгүй. **Зөв арга:** residual plot-ыг харьцуулж, theory-той тохиромжтой эсэхийг шалгах. Хэрэв тоон хэмжүүр хэрэгтэй бол — s_e -г Y-ийн анхны нэгжид буцаасны дараа харьцуулна.

1. **Хувьсагчдыг тогтоо:** Y юу вэ (тайлбарлахыг хүсэж буй)? X юу вэ (тайлбарлагч)? Cross-sectional уу, time-series юу?
2. **Slope ба intercept-ийг тооц:** $\sum(Y_i - \bar{Y})(X_i - \bar{X})$ ба $\sum(X_i - \bar{X})^2$ -ыг гарга. $\hat{b}_1 = \text{ratio}$, $\hat{b}_0 = \bar{Y} - \hat{b}_1\bar{X}$.
3. **Дөрвөн таамаглалыг шалга:** Residual plot-ыг X-ийн эсрэг, цаг хугацааны эсрэг, histogram-аар. Зөрчилтэй бол функциональ хэлбэр өөрчлөх эсвэл өөр загвар сонгох.
4. **Тааруулалтыг үнэл:** R^2 , F-statistic, s_e — гурвыг хамт. Statistically significant үү? Effect size утгатай юу?
5. **Хэрэгтэй бол прогноз гаргаж PI-ийг тооц:** X_f -ийг загварт оруулж \hat{Y}_f . s_f -ийг тооц. Critical t -аар PI-ийг бүтээ. Бизнесийн хэлээр тайлбарла.

Cheat Sheet

Шалгалтын өмнө 5 минут — энэ хуудас уншвал бүх томьёо санагдана. Хэвлээд кофены аяганы дэргэд тавиарай.

📌 Үндсэн томьёонууд

POPULATION ЗАГВАР

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

ҮНЭЛЭГДСЭН ЗАГВАР

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i$$

SLOPE (OLS)

$$\hat{b}_1 = \frac{\sum(Y_i - \bar{Y})(X_i - \bar{X})}{\sum(X_i - \bar{X})^2}$$

INTERCEPT (OLS)

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}$$

RESIDUAL

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

CORRELATION

$$r = \frac{\text{Cov}(Y, X)}{s_Y s_X} \quad | \quad R^2 = r^2$$

📊 Variance задаргаа ба тааруулалт

SST

$$\sum(Y_i - \bar{Y})^2$$

SSR

$$\sum(\hat{Y}_i - \bar{Y})^2$$

SSE

$$\sum(Y_i - \hat{Y}_i)^2$$

R^2

$$\frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

F-STATISTIC

$$F = \frac{MSR}{MSE}, \text{ df} = (1, n-2)$$

s_e (ST.ERROR OF ESTIMATE)

$$s_e = \sqrt{MSE} = \sqrt{\frac{SSE}{n-2}}$$

📌 Hypothesis testing

SLOPE T-TEST

$$t = \frac{\hat{b}_1 - B_1}{s_{\hat{b}_1}}, \text{ df} = n-2$$

SLOPE SE

$$s_{\hat{b}_1} = \frac{s_e}{\sqrt{\sum(X_i - \bar{X})^2}}$$

INTERCEPT T-TEST

$$t = \frac{\hat{b}_0 - B_0}{s_{\hat{b}_0}}$$

INTERCEPT SE

$$s_{\hat{b}_0} = s_e \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i - \bar{X})^2}}$$

CORRELATION T-TEST

SLR ОНЦГОЙ ШИНЖ

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$t^2 = F \text{ (slope-ийн хувьд)}$$

Прогноз ба PI

ТОЧКИЙН ПРОГНОЗ

$$\hat{Y}_f = \hat{b}_0 + \hat{b}_1 X_f$$

s_f (ST.ERROR OF FORECAST)

$$s_f = s_e \sqrt{1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum(X_i - \bar{X})^2}}$$

PREDICTION INTERVAL

$$\hat{Y}_f \pm t_{\text{critical}, \alpha/2} \cdot s_f \quad | \quad df = n - 2$$

Функциональ хэлбэрүүд

LIN-LIN

$$Y = b_0 + b_1 X \text{ — absolute}$$

LOG-LIN

$$\ln Y = b_0 + b_1 X \text{ — \% өсөлт}$$

LIN-LOG

$$Y = b_0 + b_1 \ln X \text{ — diminishing}$$

LOG-LOG

$$\ln Y = b_0 + b_1 \ln X \text{ — elasticity}$$

Регрессийн шийдвэрийн мод

1. Slope ба intercept хэр их вэ?

$$\hat{b}_1 = \frac{\sum(Y_i - \bar{Y})(X_i - \bar{X})}{\sum(X_i - \bar{X})^2}, \quad \hat{b}_0 = \bar{Y} - \hat{b}_1\bar{X}$$

2. Residual plot шалга →

- ▶ U-хэлбэр? → Linearity зөрчигдсэн → log хэлбэрт шилжих
- ▶ Туузны өргөн нэмэгдэж байна? → Heteroskedasticity → robust SE
- ▶ Цаг хугацааны мөчлөг? → Autocorrelation → time-series загвар
- ▶ Outlier? → Normality асуудалтай → шалгах, шаардлагатай бол хасах
- ▶ Санамсаргүй чимээ? → Бүх таамаглал ОК ✓

3. Загвар significant юу?

$F = MSR/MSE$ → critical F (df = 1, n-2)-тэй харьцуул. F том → significant.

4. Slope-ийн талаар тодорхой асуулт юу вэ?

- ▶ " $b_1 = 0$ уу?" → t-test эсвэл F-test (хоёулаа ижил)
- ▶ " $b_1 =$ тогтоосон утга уу?" → t-test л
- ▶ " $b_1 > 0$ уу?" → one-sided t-test

5. Прогноз хэрэгтэй юу?

$$\hat{Y}_f = \hat{b}_0 + \hat{b}_1 X_f \rightarrow s_f \text{ тооц} \rightarrow PI = \hat{Y}_f \pm t_{\text{critical}} s_f$$

6 LOS-ийн нэг өгүүлбэрийн хураангуй

LOS	Гол санаа
1	SLR нь $Y = b_0 + b_1 X + \varepsilon$ загвараар хоёр хувьсагчийн дундаж шугаман хамаарлыг илэрхийлдэг. OLS-аар residual-уудын квадратын нийлбэрийг хамгийн бага байлгах байдлаар \hat{b}_0, \hat{b}_1 -ийг олно.
2	Дөрвөн таамаглал: linearity, homoskedasticity, independence, normality . Residual plot нь зөрчлийг илрүүлэх гол хэрэгсэл.
3	$SST = SSR + SSE$. $R^2 = SSR/SST$. F-test нь "slope тэг үү?" гэсэн ерөнхий тест. Slope-ийн t-test нь илүү ерөнхий — SLR-д $t^2 = F$.
4	ANOVA хүснэгт нь sum of squares, df, mean squares, F-statistic-ийг нэгтгэдэг. $s_e = \sqrt{MSE}$ нь absolute хэмжүүр.
5	$\hat{Y}_f = \hat{b}_0 + \hat{b}_1 X_f$. s_f -нь X_f дунджаас холдох тусам нэмэгдэнэ. $PI = \hat{Y}_f \pm t_{\text{critical}} s_f$. Extrapolation эрсдэлтэй.
6	Log-lin (Y лог) — exponential өсөлт; lin-log (X лог) — diminishing return; log-log (хоёулаа лог) — elasticity. Сонголт нь R^2, s_e , residual plot, theory-той хослуулсан байх ёстой.

ШАЛГАЛТЫН ӨМНӨ 3 ЭЦСИЙН ЗӨВЛӨМЖ

- **Slope-ийн нэгжийг шалга:** Y-ийн нэгж / X-ийн нэгж. % дотор % бол slope нь "X 1% → Y slope%" гэсэн утгатай.
- **Intercept-ийн утгатай эсэхийг шалга:** X = 0 бодит ертөнцөд боломжтой юу? Боломжгүй бол intercept-ыг тайлбарлах нь утгагүй.
- **R^2 ба significance хоёрыг хольж бүү буруу:** R^2 нь fit-ийг хэмждэг, F/t-test нь significance-ийг шалгадаг — хоёр өөр асуулт.

Нэр томьёо

Англи нэр томьёо ↔ монгол орчуулга, тайлбар.

English	Монгол	Тайлбар
Simple linear regression (SLR)	энгийн шугаман регресс	Нэг үл хамаарлын хувьсагчтай регрессийн загвар
Dependent variable (Y)	хамаарлын хувьсагч	Тайлбарлахыг хүсэж буй хувьсагч
Independent variable (X)	үл хамаарлын хувьсагч	Y-ийг тайлбарлахад ашиглах хувьсагч
Intercept (b_0)	intercept / y-тэнхлэгийн огтлол	X = 0 үед Y-ийн утга
Slope (b_1)	налуу / slope	X-ийн нэг нэгж өөрчлөгдөхөд Y-д гарах өөрчлөлт
Error term (ϵ)	алдааны гишүүн	Ажиглалт ба үнэн шугамаас зөрөх популяцийн хазайлт
Residual (e)	residual / үлдэгдэл	Ажиглалт ба үнэлэгдсэн шугамаас зөрөх түүврийн хазайлт
Ordinary least squares (OLS)	ердийн хамгийн бага квадратын арга	Residual-уудын квадратын нийлбэрийг минимумлах арга
Sum of squares total (SST)	квадратын нийт нийлбэр	$\sum(Y_i - \bar{Y})^2$ — Y-ийн нийт хэлбэлзэл
Sum of squares regression (SSR)	регрессийн квадратын нийлбэр	$\sum(\hat{Y}_i - \bar{Y})^2$ — тайлбарлагдсан хэлбэлзэл
Sum of squares error (SSE)	алдааны квадратын нийлбэр	$\sum(Y_i - \hat{Y}_i)^2$ — тайлбарлагдаагүй хэлбэлзэл
Coefficient of determination (R^2)	тодорхойлолтын коэффициент	Y-ийн хэлбэлзлийн хэдэн хувийг X тайлбарласныг харуулна
F-statistic	F-статистик	Загварын тааруулалтын тест: MSR/MSE
Mean square regression (MSR)	регрессийн дундаж квадрат	SSR/k — SLR-д SSR-тэй адил
Mean square error (MSE)	алдааны дундаж квадрат	$SSE/(n - 2)$ — error variance-ийн үнэлгээ
Standard error of estimate (s_e)	стандарт алдааны үнэлгээ	\sqrt{MSE} — Y-ийн нэгжтэй ижил absolute хэмжүүр

English	Монгол	Тайлбар
Standard error of forecast (s_f)	прогнозын стандарт алдаа	Future \hat{Y}_f -ийн нарийвчлалыг хэмждэг
Prediction interval	прогнозын интервал	\hat{Y}_f -ийн орчмын $(1 - \alpha)$ confidence interval
Linearity	шугаман шинж	Y ба X-ийн жинхэнэ хамаарал шугаман
Homoskedasticity	гомоскедастик чанар	Residual-уудын variance бүх ажиглалтад тэнцүү
Heteroskedasticity	гетероскедастик чанар	Variance ажиглалтын дагуу өөрчлөгддөг
Independence	хараат бус байдал	Ажиглалтууд хоорондоо хамаарагүй
Autocorrelation	автокорреляц	Residual-ууд цаг хугацаандаа хамаарал илэрнэ
Normality	нормал тархалт	Residual-ууд нормал тархалттай
Residual plot	residual диаграм	Загварыг оношлох гол визуал хэрэгсэл
ANOVA table	ANOVA хүснэгт	Sum of squares, df, mean squares, F-ийг нэгтгэсэн хүснэгт
Indicator variable	индикатор хувьсагч	0/1 утгатай dummy variable
Functional form	функциональ хэлбэр	Хувьсагчдын хувиргалт (log, exp, square)
Log-lin model	log-lin загвар	$\ln Y = b_0 + b_1 X$
Lin-log model	lin-log загвар	$Y = b_0 + b_1 \ln X$
Log-log model	log-log загвар	$\ln Y = b_0 + b_1 \ln X$ — elasticity загвар
Elasticity	уян хатан чанар	X 1%-аар өөрчлөгдөхөд Y хэдэн % өөрчлөгдөх вэ
Extrapolation	экстраполяци	Өгөгдлийн хүрээнээс гадуурх X_f -д прогноз гаргах
Goodness of fit	тааруулалтын чанар	R^2 , F-statistic, s_e — гурван үндсэн хэмжүүр

LEARNING MODULE 10 ТӨГСӨВ

Дараагийн модуль: Introduction to Big Data Techniques