



ҮНДЭСНИЙ СТАТИСТИКИЙН ХОРОО

Орхигдсон утгыг орлуулан тооцох арга зүйн гарын авлага

УЛААНБААТАР ХОТ
2019 ОН

Арга зүйн байнгын зөвлөлд танилцуулсан.

Хянасан: ҮТШСГ-ын дарга А.Дэмбэрэл
 ҮТШСГ-ын ахлах статистикч Б.Анхзаяа
 ҮТШСГ-ын ахлах статистикч Н.Өлзийханд

Боловсруулсан: ҮТШСГ-ын статистикч О.Норовсамбуу

АГУУЛГА

| | |
|--|----|
| УДИРТГАЛ..... | 2 |
| 1. ОРЛУУЛГА ХИЙХ АРГА ЗҮЙН ҮНДСЭН ОЙЛГОЛТ | 3 |
| 1.1. Ойлголт, тодорхойлолт | 3 |
| 1.2. Гарч болох асуудал болон тэдгээрийн шийдэл..... | 5 |
| 1.2.1. Тоон болон чанарын мэдээлэлд орлуулах хувьсагч..... | 6 |
| 1.2.2. Детерминистик(Deterministic) болон стохастик(Stochastic) орлуулга | 7 |
| 1.2.3. Регресс болон хот-дек орлуулгыг сонгох..... | 7 |
| 1.2.4. Жинг засварлах | 8 |
| 1.3. Баримтжуулалт | 9 |
| 2. ДЕДУКТИВ ОРЛУУЛГЫН АРГА..... | 10 |
| 4. ХАРЬЦААНЫ ОРЛУУЛГЫН АРГА..... | 17 |
| 5. РЕГРЕССИЙН ОРЛУУЛГЫН АРГА | 19 |
| 6. ХОТ-ДЕК ОРЛУУЛГЫН АРГА | 21 |
| 6.1. Санамсаргүй болон дараалсан хот-дек орлуулга | 22 |
| 6.2. Хамгийн ойрын хөршийн орлуулга | 24 |
| 7. ОЛОН ШАТАТ ОРЛУУЛГЫН АРГА | 25 |
| 8.1 Интерполяци | 33 |
| 8.2 Сүүлийн ажиглалтын утгыг урагшлуулах болон хойшлуулах..... | 35 |
| 8.3 Литтл болон Су арга..... | 38 |

УДИРТГАЛ

Статистик бол үр ашигтай шийдвэр гаргахад шаардлагатай тоон мэдээллийг цуглуулах, боловсруулах, нэгтгэн дүгнэх, дүн шинжилгээ хийх, тайлбарлах, танилцуулах арга зүйн тухай шинжлэх ухаан юм. Эдгээр үе шаттай үйл ажиллагаанаас хамгийн их цаг хугацаа, хөрөнгө шаарддаг хэсэг нь мэдээлэл цуглуулах үйл ажиллагаа юм.

Мэдээллийг тооллого, түүвэр судалгаа, захиргааны статистикийн мэдээлэл гэсэн гурван үндсэн хэлбэрээр цуглуулдаг. Статистикийн мэдээллийг цуглуулах явцад судалгаанд оролцогч иргэд, аж ахуйн нэгж, байгууллагууд судалгааны зарим асуултад янз бүрийн шалтгаанаар хариулахаас татгалзах, асуултыг ойлгохгүй, мэдэхгүй байгаагаас болж буруу хариулах зэрэг олон асуудлаас болж мэдээллийн орхигдол бий болдог.

Энэхүү орхигдсон мэдээллийг боловсруулахдаа статистикчид тухайн мэдээллийг дахиж асуух, тухайн оролцогчийн мэдээллийг орхиж боловсруулалт хийх, эсвэл орлуулан тооцох зэрэг шийдлийг ашигладаг. Орхигдсон, дутуу мэдээллийг нөхөж авахад цаг хугацаа хөрөнгө шаардах учраас төдийлөн ашиглах боломжгүй байдаг. Харин тухайн оролцогчийн мэдээллийг орхиж боловсруулалт хийх тохиолдол байдаг боловч энэ нь судалгааны үр дүнг гажуудуулах сөрөг талтай. Сүүлийн арга нь орхигдсон утгыг шинжлэх ухааны үндэслэлтэйгээр орлуулан тооцох явдал юм.

Орхигдсон утгыг орлуулан тооцох олон арга байдаг бөгөөд судалгааны мэдээлэл нь эдийн засгийн, нийгмийн, тоон, чанарын, нэг болон олон тооны мэдээлэл орхигдсон эсэхээс хамаарч тохирох аргуудыг ашигладаг. Бидний мэдэх хамгийн ойрын жишээ нь логик уялдааг шалгаж засварлах мөн балансын шалгалтуудыг хийж орхигдлыг засварлах хэлбэр юм.

Энэхүү гарын авлагад статистикийн мэдээлэл боловсруулах шатанд түгээмэл ашиглагддаг орхигдсон утгыг орлуулан тооцох аргуудын талаар авч үзсэн. Өмнө дурдсанчлан орлуулан тооцох олон төрлийн аргууд байдаг ба нийгэм, эдийн засгийн судалгааны онцлогт таарсан аргуудыг ашиглах нь чухал юм. Мөн аливаа тооллого, судалгааг зохион байгуулахын өмнөх шатанд буюу үзэл баримтлал, судалгааны аргачлал, арга зүйг боловсруулах үед тухайн судалгааныхаа онцлогт тохирох орлуулан тооцох аргуудыг боловсруулах шаардлагатай. Ингэснээр цуглуулагдсан мэдээлэл орхигдсон утгуудтай байвал тэдгээрийг орхих эсвэл орлуулан тооцох шийдвэр гаргахад хялбар болох юм.

Уг гарын авлагыг боловсруулахад Нидерландын статистикийн газраас эрхлэн гаргасан орлуулгын арга зүйн гарыг авлагыг тулгуур болгон ашиглаж, олон улсын байгууллагаас гаргасан зарим ойлголтуудаар сайжруулж, баяжуулсан.

1. ОРЛУУЛГА ХИЙХ АРГА ЗҮЙН ҮНДСЭН ОЙЛГОЛТ

1.1. Ойлголт, тодорхойлолт

Аливаа тооллого, судалгаа зохион байгуулж мэдээлэл цуглуулах явцад судалгаанд оролцогчид нэг болон хэд хэдэн асуултад хариулахаас татгалзах тохиолдол гардаг. Хариулт өгөөгүй шалтгаан нь ихэвчлэн хариулагчид тухайн үед хариулахад бэлэн биш юм уу хариулт өгөх чадваргүй байсан эсвэл асуултад хариулахаас эрс татгалзсан байдаг. Тухайлбал хүмүүс ойлгомжгүй, ойлгоход хэцүү асуултад хариулт өгдөггүй ба ихэнх тохиолдолд шууд хариулахад хүнд, эмзэг асуултад хариулт өгөхөөс татгалздаг. Энэ тохиолдолд хариу өгөөгүй, нэг болон хэсэгчлэн хариулаагүй асуултуудыг зааглаж өгөх шаардлага гардаг. Үүний дараа ямар арга ашиглан орхигдсон утгыг орлуулахыг шийдвэрлэнэ.

Орхигдсон утгыг орлуулах хэд хэдэн арга бий. Эхнийх нь цуглуулсан мэдээллээс орхигдсон утгын хувьд хамгийн ойр хүлээн зөвшөөрөгдөх утгыг тооцох явдал юм. Үүнийг боловсруулалтын үеийн орлуулга (Imputation) хийх гэж нэрлэнэ.

Өөр нэг арга нь үл мэдэгдэх утгыг орхиж үр дүнг боловсруулах юм. Хөдөлмөр эрхэлдэггүй хүмүүсээс ажлын орчны талаар асуулт тавих шаардлагагүй ба хөдөлмөр эрхлэлтийн талаар гүнзгийрүүлсэн асуултууд нь зөвхөн хөдөлмөр эрхэлж буй хүмүүс рүү чиглэгддэг. "Мэдэхгүй", "Үгүй", "Татгалзсан" зэрэг хариулт нь хариулагчийн мэдлэг, үзэл бодлын талаар тодруулсан асуултад ашиглагддаг. Гэсэн хэдий ч орхигдсон утгын хувьд ч гэсэн орлуулга хийхгүй байх шийдвэр гаргаж болдог ба цуглуулсан мэдээллийн боловсруулалтын үед орлуулахгүйгээр шинжилгээ, үнэлгээний явцад орлуулга хийдэг. Ялангуяа чанарын үзүүлэлтийн хувьд "үл мэдэгдэх" гэсэн утгыг ашиглана. Орлуулга нь тоон үзүүлэлтийн хувьд илүү үр дүнтэй байх учир нийгмийн статистикийн үзүүлэлтээс илүүтэй бизнес, эдийн засгийн статистикт илүү ашиглагддаг.

Бүрэн мэдээлэл бүхий цуглуулсан мэдээллийг нэгтгэн дүгнэх явцад хүснэгтүүдийг боловсруулах нь хялбар, тоон мэдээлэл нь зөрүүгүй гарна. Жишээлбэл: Боловсролын түвшинг насаар, хүйсийг насаар авч үзэхэд үл мэдэгдэх утга байхгүй бол нийлбэр нь хоорондоо тэнцүү буюу нийт судалгаанд оролцогчдын тоотой адил байна. Харин боловсрол, хүйс аль нэгэнд нь хариулаагүй, үл мэдэгдэх утгатай байвал хүснэгтүүдийн нийт дүн зөрүүтэй гарна.

Хэрэв судалгаанд орлогын мэдээлэл нөхөгдөөгүй байвал нийт судалгаанд оролцогчдын дундаж орлого эсвэл орлогын задаргаанд дахь мэдээллүүдийн дунджаар тооцож орлуулга хийж болох боловч энэ нь төдийлөн ач холбогдолтой биш юм. Үүнээс үзэхэд орлуулга хийх арга нь чухал боловч зарим илэрхий үзүүлэлтийн хувьд хэрэглэх нь нөлөө үзүүлэхгүй болох нь харагдаж байна.

Орлуулга хийх арга зүйд доорх ойлголт, тодорхойлолтыг ашиглагддаг. Үүнд:

| Ойлголт | Тодорхойлолт |
|--|--|
| Хариулаагүй утга, үл мэдэгдэх утга, орхигдсон утга | Судалгаанд оролцогчоос үл мэдэгдэх, алдаатай хариулт ирэхийг хэлнэ. |
| Орлуулга, орлуулах | Үл мэдэгдэх эсвэл хариулаагүй гэж тэмдэглэсэн утгад шинэ утгыг тодорхойлох, орлуулахыг хэлнэ. |
| Орлуулагч утга | Үл мэдэгдэх утгад орлуулсан утгыг ойлгоно. |
| Орлуулах бүлэг | Тус тусад нь орлуулах алгоритмыг ашигласан дэд бүлгүүдийг хэлнэ. |
| Дедуктив орлуулга (логик орлуулга) –Deductive imputation | Тухайн утга нь 100 хувь тодорхой биш тохиолдолд ямар нэг магадлалын механизм ашиглалгүйгээр логикт үндэслэн үл мэдэгдэх утгыг орлуулах арга зүйг хэлнэ. |
| Дундаж орлуулга | Судалгаанд хариулсан оролцогчдын хувьсагчдын дундаж утгыг тооцож үл мэдэгдэх утгад орлуулахыг ойлгоно. |
| Бүлгийн дундаж орлуулга | Бодит утгатай ба ижил дэд бүлэгт байдаг хариултгүй оролцогчдын утгад үл мэдэгдэх утгыг дундаж дүнгээр орлуулахыг хэлнэ. |
| Хот-дек орлуулга – Hot-deck imputation | Үл мэдэгдэх утгыг хүлээн авагчийн аль болох ижил шинж тэмдэгтэй донор оролцогчоос авч орлуулах арга зүй. |
| Олон хувьсагчийн орлуулга - Multivariate imputation | Нэг оролцогчийн олон үл мэдэгдэх утгыг орлуулах |
| Масс орлуулга - Mass imputation | Тодорхой үзүүлэлтийн хувьд бүх дутуу орхигдсон үл мэдэгдэх утгуудыг орлуулах арга зүй. |
| Өөр хугацааны ижил мэдээлэлд үндэслэсэн орлуулга - Longitudinal imputation | Ижил эсвэл бусад мэдээллийн өөр хугацаа/үе-ийн ижил үзүүлэлтийг ашиглан үл мэдэгдэх утгыг орлуулах арга зүй. Энэ орлуулга нь мөн олон хувьсагчийн орлуулга байж болно. |

Ерөнхий тэмдэглэгээ

Энэ гарын авлагад дараах ерөнхий тэмдэглэгээнүүдийг ашиглах болно.

i – үзүүлэлт, оролцогчийн индекс (дугаар)

y – зорилтот үзүүлэлт, орлуулах үзүүлэлт

y_i – зорилтот/орлуулах y үзүүлэлтийн i дугаар үзүүлэлт

obs – y_i -ийн ажиглагдсан утгуудын багц

mis - y_i – ийн үл мэдэгдэх утгуудын багц

\tilde{y}_i - y_i -ийн үл мэдэгдэх утгад орлуулсан утга

1.2. Гарч болох асуудал болон тэдгээрийн шийдэл

Заримдаа үл мэдэгдэх утга байсан ч үзүүлэлтийн шинж чанараас хамааран бодит үр дүнг 100 хувийн баталгаатай гаргах боломжтой байдаг. Энэ тохиолдолд тухайн бодит утгыг тооцохын тулд *дедуктив орлуулгын* (Бүлэг 2) аргыг ашиглаж болно. Засварлахдаа хоорондоо уялдаа холбоо бүхий ойролцоох шалгуур үзүүлэлтүүдийг ашиглаж засварлана. Хэрэв энэ аргыг ашиглах боломжтой байвал бусад орлуулгын аргуудаас өмнө давуу эрхтэйгээр хэрэглэдэг.

Хэдийгээр утгыг нь тодорхойлох боломжгүй боловч үл мэдэгдэх утгыг (y хувьсагч) нэмэлт мэдээллийн (x хувьсагчийн) тусламжтайгаар нарийвчлан тооцох боломжтой юм. Тохиромжтой, үр дүнтэй тайлбарлах загварыг тодорхойлох замаар орлуулга хийх арга зүйг оновчтой сонгож параметрууд эсвэл судалгааны үндсэн мэдээллийн чанарыг сайжруулна. Сонгосон загварын тусламжтайгаар үл мэдэгдэх утгыг тооцож нөхдөг. Гэвч бусад эх үүсвэр эсвэл түүвэр судалгааны мэдээлэлтэй харьцуулах боломжгүй бол тооцож гаргасан утгын чанарыг нарийн үнэлэх боломжгүй байдаг. Загварын тооцоог зөвхөн тухайн судалгаанд оролцогчдын мэдээлэлд үндэслэн тооцох боломжтой.

Хэрэв \hat{y}_i утгыг орлуулахад тодорхой бус зүйл байгаа бол түүнийг загвар ашиглан тооцож орлуулж болох юм. Ингэхийн тулд y үл мэдэгдэх утгыг аль болох нарийвчлалтай тодорхойлох y_i загварыг тодорхойлно. Ихэнхдээ энэ тохиолдол регрессийн загварыг ашигладаг бөгөөд үүнийг *Регрессийн орлуулгын арга зүй* (Бүлэг 5) гэж нэрлэдэг. Энэ арга зүй нь ихэвчлэн тоон y хувьсагчийн хувьд хэрэглэгддэг.

Дунджаар орлуулах арга зүй болон харьцааны орлуулах арга зүйн талаар Бүлэг 3 болон 4 авч үзэх ба энэ нь регрессийн орлуулах арга зүйн тусгайлсан хувилбарууд юм. Дунджаар орлуулах арга зүй нь нэмэлт мэдээлэлгүй үед ашиглагдаг бол харьцааны орлуулах арга зүйн зөвхөн нэг нэмэлт тоон хувьсагч байгаа тохиолдолд ашиглагддаг. Эдгээр аргууд нь түгээмэл ашиглагддаг учраас тусгайлан авч үзнэ.

Мөн хот-дек орлуулгын арга байдаг бөгөөд санамсаргүй хот-дек, дараалсан хот-дек, хамгийн ойрын хөрш гэсэн төрлүүдтэй (Бүлэг 6). Зорилгынхоо хувьд эдгээр орлуулгын аргууд нь регрессийн орлуулгын арга зүйтэй адил юм. Гэвч эдгээр хот-дек орлуулах аргуудыг нэг оролцогчийн олон мэдээлэл дутуу үед ашиглахад илүү хялбар байх бөгөөд хувьсагчдын хоорондын хамаарлыг илүү нарийвчлалтай тооцоолдог. Хариулаагүй i оролцогч бүрийн хувьд орлуулах i үзүүлэлттэй ижил шинж чанартай, шинж чанар нь зорилтот y хувьсагчийг өөрчлөхөд нөлөөлдөг гэж үзвэл аль болох олон d донор хувьсагчийг хайна. Үүний дараа донор утгыг y_d гэж тэмдэглэн $\hat{y}_i = y_d$ орлуулна.

Дараа нь 7-р бүлэгт нэг оролцогчийн олон тооны үл мэдэгдэх утгуудыг орлуулах олон шатат орлуулгын арга зүйг авч үзэх бөгөөд үүний хэд хэдэн арга байдаг. Бүлэг 8-д урт хугацааны туршид үргэлжилсэн судалгааны өгөгдөлд орлуулга хийх арга зүйг авч үзнэ. Тодруулбал энэ арга нь ялгаатай хугацааны ижил мэдээлэл ашиглаж орлуулах арга зүй ба энд ижил хугацааны өөр мэдээлэл ашиглаж орлуулах арга орохгүй. Энэ хэсэгт орлуулга хийх арга зүйг хэрхэн тодорхойлох, сонгох талаар авч үзнэ.

1.2.1. Тоон болон чанарын мэдээлэлд орлуулах хувьсагч

Хот-дек орлуулгын арга зүйг (Бүлэг 6) y хувьсагчийн аль ч хэлбэрт (тоон болон чанарын) ашиглаж болдог.

Регрессийн орлуулгын арга зүйг y хувьсагч нь ихэвчлэн тоон хувьсагч байгаа тохиолдолд ашигладаг. Ингэхдээ регрессийн шугаман хэлбэрийн тэгшитгэлийг ашигладаг боловч y хувьсагчтай уялдуулж бусад хэлбэрийн регрессийн тэгшитгэлээр тооцож болдог. Тухайн мэдээллийн онцлогоос хамаарч аль ч хэлбэрийн регрессийн тэгшитгэл ашиглаж орлуулга хийж болно. y нь чанарын хувьсагч тохиолдолд хоёр эсвэл олон хувьсагчийн ложистик регрессийн аль тохирох загварыг ашиглана.

Хэрэв y тоон хувьсагчийн хувьд тооцоолол хийхэд нэмэлт мэдээлэл байхгүй эсвэл тэдгээр нь үр дүнгүй байгаа бол нэмэлт мэдээлэл (x хувьсагч) ашиглагдахгүй тул регрессийн орлуулга хийх боломжгүй болж дунджийн орлуулга (Бүлэг 3)-г хийдэг. Тус тохиолдолд түгээмэл байдаг учраас энэ аргын талаар тусад нь авч үзэх болно.

Хэрэв чанарын y хувьсагчийн хувьд нэмэлт мэдээлэл байхгүй нь бол төдийлөн зөвлөдөггүй арга болох хамгийн их тохиолддог утгаар эсвэл пропорциональ магадлалтай ангиллуудын ажиглагдсан категорийн давтамжаас санамсаргүйгээр сонгож орлуулж болно. Сүүлийн арга нь нийт судалгаанд оролцогчдоос тооцоолсон параметрууд нь үр дүнд бага нөлөө үзүүлсэн тохиолдолд үндэслэлтэй болно. Санамсаргүй хот-дек (Бүлэг 6) оролцогчийг сонгож орлуулгад ашиглаж болно. Нэмэлт мэдээлэлгүйгээр орлуулах нь зөвхөн цөөн хэдэн оролцогч хариулаагүй бөгөөд y үл мэдэгдэх хувьсагчийн хувьд регрессийн загварт нийцүүлэн хамгийн сайн тохирох таамаглал дэвшүүлэх боломжтой. Хэрэв энэ нь бүх үл мэдэгдэх утгын хувьд тооцогдсон бол орлуулга нь "хэтэрхий төгс" болно. Бүх орлуулсан үзүүлэлтүүд нь оруулгын загварыг бүрэн хангаж өгдөг. Орлуулсан утгуудыг судалгааны микро өгөгдлийн цаашдын боловсруулалтад ихэнхдээ дурддаггүй эсвэл зарим энгийн хүснэгтэд "тэмдэглэгээ" хийж орлуулсан утгатай гэдгийг тайлбар бичиж өгдөг (Бүлэг 9). Хүн ам зүйн статистикийн хамгийн ойрхон жишээ бол орлуулга хийхдээ нөхөр эсвэл эхнэрийн нас нь мэдэгдэхгүй байвал нөхөр нь эхнэрээсээ хоёр насаар ах байдаг гэсэн тооцооллыг баримталж болно. Ийм орлуулгын загвар нь эрэгтэй, эмэгтэй насны ангилал, тархалтад сайнаар нөлөөлж болох талтай.

Ерөнхийдөө, хамгийн сайн таамаглал бүхий регрессийн загвараар орлуулга хийсэн утгын хазайлтыг дутуу үнэлэх талтай байдаг. Энэ нь хэт өндөр болон бага утгыг тархсан хэсэгт хуваарилахад хүргэх бөгөөд y үл мэдэгдэх хувьсагч нь хэт олон байвал y -ын дисперсийн тайлбарлах чадвар хангалтгүй байдаг. Энэ нөлөө нь дунджийн орлуулгад хамгийн их илэрдэг. Энэ нь нийт үр дүнг эсвэл дундаж утгуудын тооцоонд ямар нэг нөлөө үзүүлэхгүй боловч хуваарилалтын тооцоолол (орлогын хуваарилалт gm) болон тархалтыг хэмжихэд илэрдэг. Хуваарилалтыг бодитой тодорхойлохын тулд санамсаргүй хазайлтыг хамгийн сайнаар урьдчилан таамаглаж орлуулга хийх хэрэгтэй.

Регрессийн шинжилгээнд:

1. Энгийн магадлалт тархалтаас түүвэрлэх,
2. Санамсаргүй байдлаар түүвэрлэсэн донорын үлдэгдлийг нэмэх гэсэн хоёр аргаас сонгож болно. 5-р бүлэгт ийм үлдэгдэлтэй болон үлдэгдэлгүй регрессийн орлуулгын ялгааг авч үзнэ.

1.2.2. Детерминистик(Deterministic) болон стохастик(Stochastic) орлуулга

Хэрэв үлдэгдлийн тархалтаас эсвэл донороос санамсаргүйгээр сонгож орлуулга хийсэн бол үүнийг стохастик орлуулга гэж нэрлэнэ. Учир нь санамсаргүй сонголтоос болж орлуулгыг дахин хийх боломжгүй байдаг.

Детерминистик орлуулга нь орлуулгын загварыг сонгох бөгөөд дахин орлуулга хийж болдог. Ихэнх тохиолдолд стохастик болон детерминистик орлуулгын ялгаа нь өмнө дурдсанчлан үлдэгдлийг орлуулгад ашиглах болон ашиглахгүй байх юм. Хамгийн ойрын хөршийн орлуулга нь тохирох урьдчилсан таамаглалыг багтаасан ч энэ нь детерминистик орлуулга юм. Учир нь донор утга нь тодорхой функц ашиглан засварлагдсан байдаг.

1.2.3. Регресс болон хот-дек орлуулгыг сонгох

Регресс болон хот-дек орлуулгыг алийг сонгох нь шууд тодорхой харагддаггүй. Энэ нь ихэвчлэн бодит болон үл мэдэгдэх утга нь тодорхойгүй байгаатай холбоотой ба аль загварыг нь илүү сайн болохыг үнэлэх боломжгүй байдаг. Гэхдээ эдгээр аргуудын сонголтод нөлөөлж болох хэд хэдэн асуудлуудыг авч үзье.

- Регрессийн болон хамгийн ойрын хөршийн орлуулгын хувьд чанарын болон тоон хувьсагчийн x хувьсагчийг орлуулж болно. Хот-дек орлуулгын хувьд зөвхөн чанарын хувьсагчийг орлуулах ба тоон хувьсагчийг урьдчилан засварлаагүй тохиолдолд оруулах боломжгүй байдаг. Гэхдээ энэ тохиолдолд хувьсагчийн тоон тал нь хэсэгчлэн алдагддаг.
- Хот-дек орлуулгын хувьд регрессийн орлуулгатай харьцуулахад загварт x хувьсагчдыг орлуулах хязгаарлалт байдаг. Ингэхдээ чанарын хувьсагчдын хоорондын бүх харилцан үйлчлэлийг багтаасан байх шаардлагатай ба энэ нь параметруудийн тоог түүврийн хэмжээтэй харьцуулахад их байж болно гэсэн

үг юм. Харин регрессийн загварт түүврийн хэмжээнээс цөөн тооны параметруудийг ашигладаг.

- Чанарын хувьсагчдыг ангилж тэдгээрийг хуурмаг хувьсагчдаар (ангилал бүрд нэг хуурмаг хувьсагч) орлуулбал мэдээлэл орхигдуулах болно. Хэрэв тэдгээр хуурмаг хувьсагчид нь у хувьсагчтай хүчтэй шугаман бус хамааралтай бол энэ ангилал илүү тайлбарлах чадвартай дисперстэй хийгдэх хэрэгтэй.
- Хот-дек орлуулгын хувьд орлуулсан донор утга нь үргэлж төлөөлөхүйц байдаг. Жишээлбэл, у нь бүхэл тоо бол регрессийн тооцоонд бараг хэзээ ч бүхэл тоо гардаггүй, харин хот-дек орлуулгын хувьд зөвхөн бүхэл тоо орлуулах боломжтой байдаг. Хот-дек орлуулгад донор оролцогчийн донор хувьсагч нь хүлээн авагч х хувьсагчтай тохирч байх ёстой ба засварлах зорилготой нийцсэн тохиолдолд хүлээн авагч оролцогчийн үл мэдэгдэх х утгыг автоматаар орлуулж өгнө.
- Хэрэв нэг оролцогчийн олон мэдээлэл үл мэдэгдэх байвал хот-дек орлуулгыг ашиглахад тохиромжтой байдаг.

1.2.4. Жинг засварлах

Өмнө дурдсан ихэнх орлуулгын аргуудад судалгаанд оролцогчдыг тэгш бус байдлаар жинлэх орлуулгын сонголттой, жишээлбэл тэдгээрийг орлуулах магадлалтай урвуу пропорционалиар жинг хуваарилах эсвэл сонгогдсон нэгжийн хариулаагүй оролцогчийн хувьд дахин жинлэлт хийж жинг хуваарилж болно. Шугаман регрессийн орлуулгын хувьд жинлэгдсэн үзүүлэлтээр хамгийн бага квадратын аргаар үнэлгээ хийгддэг бол хот-дек орлуулгын хувьд бага орлуулах магадлалтай боломжит доноруудаас өндөр орлуулах магадлалтай боломжит донорууд нь их орлуулгын жинтэй байдаг учраас донор болох магадлал их байдаг. Дедуктив болон хамгийн ойрын хөршийн орлуулгын хувьд жин нөлөөгүй байдаг.

Загварын хувьд, хэрэв хариулах магадлал эсвэл орлуулах магадлалаас үл хамааран алдааг жигд тархаасан гэж таамаглавал бүх үр дүн нь тэгш найдвартай хэмжигддэг. Тиймээс орлуулгын загварт жинг ашиглах шаардлагагүй ба ашиглахгүй байх нь илүү үр дүнтэй, учир нь жинлэх нь стандарт алдааг илүү их болгодог. Хэрэв жинтэй хувьсагчийг эсвэл жинлэх суурь хувьсагчийг загварт тайлбарлагч хувьсагчаар авсан бол жин тооцох шаардлагагүй. Иймээс х хувьсагчийг сонгохдоо жинтэй эсэх болон жигнэх суурь болох эсэхийг анхаарах хэрэгтэй. Гэсэн хэдий ч түүврийн онолын үүднээс авч үзвэл түүврийн нэгжийн хариулт нь эх олонлогийн сонгогдоогүй элементүүдийн хариултыг яг ижил хариулт өгнө гэж "төлөөлөгч" болдог. Энэ зарчимд үндэслэн (эсвэл нэгжийг санамсаргүйгээр хариулт өгөхгүй гэж таамаглавал) түүврийг хазайлтгүй үнэлэхийн тулд жинлэх шаардлагатай болно. Хот-дек орлуулгад донор байх магадлал нь орлуулах хувьсагчийн жинтэй пропорциональ байх хэд хэдэн аргыг Грахам Калтон (Graham Kalton-1983) санал

болгосон. Түүний дэвшүүлсэн санаа нь донор болон хүлээн авагч хувьсагчид ижил жинтэй бол шууд ашиглах боломжтой ба харин донор хувьсагчийн жин хүлээн авагч хувьсагчийн жингээс их болон бага байх тохиолдолд түүний жингийн хазайлт нь үр дүнгийн тархалтад зөрүү үүсгэх эрсдэлтэй. Ангилсан х хувьсагчийг жинд нь үндэслэн жингийн суурь хувьсагч эсвэл нэмэлт хувьсагч гэж ялгах замаар энэхүү зөрүүтэй тооцоолол үүсэхээс урьдчилан сэргийлж болно.

1.3. Баримтжуулалт

Тооллого, судалгааны орхигдсон утгыг орлуулах явцад ямар утгыг орлуулсан, орлуулахдаа ямар нэмэлт хувьсагч, параметр, загвар ашигласан болон загварын зорилго ямар байсан зэргийг нарийн баримтжуулах шаардлагатай. Баримтжуулалт нь дахиж тухайн аргыг ашиглах болон засварлах тохиолдолд гол чиглүүлэгч, зааварчилгаа болдог. Судалгааны мэдээллийн санд орлуулсан утгыг тэмдэглэх, ялгах хэд хэдэн сонголтууд байдаг. Үүнд:

- “Flagging” гэж орлуулсан утгыг тэмдэглэх
- Орлуулсан болон орлуулаагүй файльтай ажиллах
- Орлуулга хийсний дараа болон өмнөх хувьсагчдыг хооронд нь ялгах

Мөн баримтжуулалт нь орлуулга хийсэн мэдээлэл ашиглах сонирхолгүй зарим судлаачдад микро буюу анхан шатны судалгааны мэдээлэлд үндэслэн шинжилгээ судалгаа хийхэд шаардлагатай мэдээлэл болдог. Стандарт алдааг тодорхойлохдоо ямар аргыг ашиглаж орлуулга хийсэн, аль утга нь бодит, аль утга нь орлуулсан болохыг мэдэж байх шаардлагатай байдаг учир зайлшгүй баримтжуулалт шаардлагатай юм.

2. ДЕДУКТИВ ОРЛУУЛГЫН АРГА

Дедуктив орлуулга нь загварт үндэслэн орхигдсон утгыг урьдчилан таамаглах арга юм. Гэхдээ зарим тохиолдолд загвартай адил тооцох ямар нэг параметр байхгүй тохиолдолд ойролцоолох дүрмийг баримтлах ба орлуулгыг ойролцоо логик асуултууд үндэслэн эсвэл ижил оролцогчийн мэдээлэлд ажиглагдсан утгуудаас шууд орлуулдаг.

Жишээ 1: Гэрлэлтийн статус нь тодорхойгүй, гэвч насны асуултад 10 настай гэж хариулсан бол энэ хүн гэрлээгүй гэдгийг баттай хэлж болно.

Жишээ 2: Аж ахуйн нэгж байгууллагын судалгаанд нийт бараа эргэлт (O), үндсэн үйл ажиллагааны бараа эргэлт (O1), туслах үйл ажиллагааны бараа эргэлт (O2) талаар асуусан. Хэрэв эдгээр 3 мэдээллийн нэг нь байхгүй бол $O1+O2=O$ тэнцэтгэлийг ашиглаж тооцоолж болно.

Дээрх орлуулгын аргууд нь дедуктив болон логик орлуулгын аргуудын жишээ юм. Энэхүү орлуулгын аргын хувьд хувьсагчдын хоорондох логик эсвэл математикийн уялдаа холбоонд үндэслэн ажиглагдсан утгуудаас нэг буюу түүнээс дээш үл мэдэгдэх хувьсагчуудыг гаргаж авах боломжтой эсэхийг шалгаж үздэг. Энэ орлуулах боломжтой үл мэдэгдэх хувьсагчдын хувьд орлуулах утга нь дедуктив орлуулга юм.

Дедуктив орлуулгын хувьд загваруудыг тодорхойлох эсвэл хэмжих шаардлагагүй байдаг. Орлуулгыг зөвхөн засварлах дүрмээр хийх бөгөөд орлуулах үйл явцыг автоматаар гүйцэтгэх боломжтой байдаг. Цаашилбал дедуктив орлуулгын арга нь нэг талаар хамгийн сайн орлуулгын арга юм. Хэрэв судалгаанд оролцогчийн бусад мэдээлэл зөв, бодитой цуглуулагдсан бол орлуулсан утга нь бодит утгатай яг ижил байна. Эндээс үндэслэн энэ аргыг аль болох судалгааны мэдээллийг цэвэрлэж, “үл мэдэгдэх” гэж тэмдэглээ хийсний дараа орлуулах, үр дүн боловсруулахад бэлэн болсон мэдээлэлд тооцох нь чухал юм. Дедуктив орлуулгын арга нь хамгийн логик дэс дараалалтай алхмууд юм. Энэхүү аргын дараа загварт үндэслэсэн болон хот-дек аргуудыг ашигладаг. Параметруудийг үнэлэхийн тулд эдгээр аргууд нь аль хэдийн ойролцоолсон утгуудаас авч үнэлэхэд тооцооны хувьд илүү хялбар байх болно.

Энгийн орлуулгын дүрмүүд

Ихэнх дедуктив орлуулгын аргууд нь “хэрэв тийм бол, тийм байна” гэсэн энгийн дүрэмтэй. Жишээлбэл: Хэрэв гэрэлтийн статус нь үл мэдэгдэх болон нас нь 18 наснаас доош байвал гэрлэлтийн статус нь огт гэрлээгүй байна.

Хэрэв нийт ажиллагчдын зардал нь үл мэдэгдэх болон цалинтай ажиллагчид 0 байвал нийт ажиллагчдын зардал нь 0 байна.

Тэгшитгэлийн хязгаарлалт (балансын шалгалт)

Дедуктив орлуулгын хувьд хамгийн түгээмэл практик эх үүсвэр нь эдийн засгийн статистикийн тэгшитгэлийн системүүд юм. Энд 30 тэгшитгэлийн хязгаарлалттай 100 орчим хувьсагчид байдаг. Хамгийн түгээмэл тэнцэтгэлийн хэлбэр нь "хувьсагчдын нийт дүн нь дэд бүлгүүдийн дүнгийн нийлбэртэй тэнцүү" юм. Хэрэв ийм тохиолдолд эдгээрийн дунд нэг тоо эсвэл нийт дүн үл мэдэгдэж байвал үл мэдэгдэх хувьсагчийг хялбархан тооцох боломжтой. Энэ нь нэг тодорхойгүй утгатай нэг тэгшитгэл юм. Практикт олон тодорхойгүй хувьсагчидтай олон тэгшитгэлүүд тохиолддог. Эдгээр нь олон хувьсагчидтай олон тэгшитгэлүүд байж болох бөгөөд тэдгээр тэгшитгэлд олон үл мэдэгдэх утга байвал системт тэгшитгэлээр тооцож үл мэдэгдэх утгыг тодорхойлдог зарим тохиолдолд тодорхойлох боломжгүй байдаг. Ийм тэгшитгэлийн системийг зохиож дедуктив орлуулгын автоматаар үүсгэх аргыг доор тайлбарлав.

Мэдээлэл нь p хувьсагчдаас бүрдсэн бөгөөд p хувьсагчид нь q шугаман тэгш хязгаарлалттай гэж үзье. Эдгээр хязгаарлалтыг дараах хэлбэрээр илэрхийлье.

$$Ry = 0 \quad (2.1)$$

Энд y нь p хувьсагчтай вектор, R нь мөр бүр нь нэг хязгаарлалттай $q * p$ матриц байна. Жишээлбэл үйл ажиллагааны орлого нь дараах таван хэсгээс бүрддэг

Үйл ажиллагааны орлогын таван бүлэг

| | |
|---|-------|
| Үндсэн үйл ажиллагааны цэвэр бараа эргэлт | y_1 |
| Бусад үйл ажиллагааны цэвэр бараа эргэлт | y_2 |
| Нийт цэвэр бараа эргэлт | y_3 |
| Бусад үйл ажиллагааны орлого | y_4 |
| Нийт үйл ажиллагааны орлого | y_5 |

Эдгээрийн хувьд хоёр үндсэн хязгаарлалт байна $y_3 = y_1 + y_2$ болон $y_5 = y_3 + y_4$. Эдгээр хязгаарлалтыг 2.1 хэлбэрт оруулбал.

$$R = \begin{bmatrix} 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$

Хэрэв векторын y хувьсагчид нь o мэдэгдэж буй утгатай, m үл мэдэгдэх утгатай гэж үзвэл үндсэн векторыг хоёр векторт хувааж болно $y = (y_o, y_m)'$ энд y_o нь y -ийн мэдэгдэж буй утгуудаас бүрдсэн o вектор, y_m нь үл мэдэгдэх утгуудаас бүрдсэн m вектор. Хэрэв y хувьсагчийг хуваасны дагуу R -ийг хуваавал дараах байдлаар бичиж болно.

$$[R_o \quad R_m] \begin{bmatrix} y_o \\ y_m \end{bmatrix} = 0 \quad (2.2) \text{ энэ нь}$$

$$R_m y_m = -R_o y_o = a \quad (2.3) \text{ болно.}$$

Сүүлийн тэгшитгэл нь y_m үл мэдэгдэх утгуудын шугаман тэгшитгэлийн систем юм. Дедуктив орлуулгын зорилго нь энэ системээс үл мэдэгдэх утгуудыг аль болох тооцож гаргахад оршино.

Шугаман тэгшитгэлийн системийн хувьд гурван тохиолдлоор ялгах нь түгээмэл байдаг:

1. Шийдгүй (систем тохиромжгүй)
2. Нэг шийдтэй
3. Хязгааргүй шийдтэй.

Нэгдүгээр тохиолдол нь $[R_m a]$ -ын зэрэг нь R_m -ын зэрэгтэй тэнцэхгүй байхад үүсдэг. Хэрэв хязгаарлалтыг ямар нэг алдаагүйгээр тооцож томъёолсон бол тоон өгөгдөл дээр алдаа гарсан тохиолдолд бий болдог. Хязгаарлалтын зөрүүтэй эдгээр төрлийн алдааг эдийн засгийн статистикт хамгийн түрүүнд илрүүлдэг. Дараа нь “үл мэдэгдэх” гэж тэмдэглэхдээ буруу тэмдэглэсэн утгуудын тоог илрүүлдэг. Шинээр тоологдсон үл мэдэгдэх утгууд болон хязгаарлалтаар дутуу үнэлэгдсэн утгыг орлуулах байдлаар засварлана. Хэрэв эдгээр алдаануудыг засварласан тохиолдолд нэгдүгээр хариу дахиж гарахгүй.

Хоёрдугаар тохиолдол нь R_m -ын зэрэг нь үл мэдэгдэх утгын тоотой тэнцүү байхад гардаг. Энд бүх үл мэдэгдэх утга орлуулагдах боломжтой ба y_m хязгаарлалтын хувьд зөвхөн нэг утгатай байна.

Гэхдээ ихэнх тохиолдолд y_m -ын хязгааргүй тооны шийдлүүд буюу **Гуравдугаар тохиолдол**той тулгардаг. Энд y_m -ын зарим элементүүд бүх боломжит шийдүүдийн дотор ижил утгатай байж болдог. Тэдгээр элементүүдийг орлуулах боломжтой юм.

y_m -ын шийдлийн багцыг \widetilde{y}_m гэвэл

$$\widetilde{y}_m = R_m^{-1} a + (R_m^{-1} R_m - I)z = b + Cz \quad (2.4)$$

R_m^{-1} нь R_m -ын өргөтгөсөн урвуу матриц ($R_m R_m^{-1} R_m = R_m$ үүний хувьд $m \times q$ матриц), z нь m векторын дурын тохиолдол. Учир нь z -г дурын тооноос сонгосон, (2.4) нь ерөнхийдөө y_m -ын хувьд хязгааргүй тооны шийдлийг гаргадаг, R_m нь бүрэн зэрэгтэй байх тохиолдолд зөвхөн нэг өвөрмөц шийдэлтэй байх ба R_m^{-1} нь ердийн урвуу матриц болдог. Хэрэв y_m -ын зарим элементүүд нь бүх боломжит шийдлүүдийн хувьд ижил байвал z -ийн дурын утга бүрд C -ийн харгалзах мөрөнд зөвхөн 0 байх ёстой. Эдгээр элементүүдийг илрүүлснээр b -ийн харгалзах утгуудын хамт хялбархан орлуулж чадна.

Дедуктив орлуулгыг хийх өөр нэг боломж нь олон хувьсагчдын эерэг байдлыг ашиглах юм. Жишээлбэл найман утга ажиглагдсан ба нэмэлтээр хоёр дэд утга

ажиглагдсан гэж үзье гэхдээ эдгээр нь тайланд тусгасан нийт дүн хүртэл нэмэгдэнэ. Хэрэв үл мэдэгдэх дэд утгууд нь эерэг бол тэдгээр үл мэдэгдэх дэд утгуудыг тэг гэж орлуулж болно, учир нь тэдгээрийн нийлбэр нь ихэвчлэн тэг байх ёстой.

Энэхүү шийдлийн хэлбэрүүдийг олохын тулд $R_m y_m = a$ тэгшитгэлийг дахин авч ашиглая. a -ийн хувьд тэгтэй тэнцэх a_j элемент байдаг гэж үзье. Тэгвэл R_m -ын $r_{m,j}$ харгалзах мөрийн хувьд $r_{m,j}' y_m = 0$ байна. Хэрэв y_m -ын бүх элементүүдийн хувьд $r_{m,j}$ -ын харгалзах элементүүд тэг утгагүй бол:

- i. Эдгээр элементүүд нь эерэг
- ii. $r_{m,j}$ -ын харгалзах элементүүд нь бүгд сөрөг эсвэл бүгд эерэг байна. Тэгвэл y_m -ын бүх элементүүд тэгтэй тэнцүү байна.

Жишээ:

Эдийн засгийн статистикт дедуктив орлуулгыг Паннекоек, Темпелмэн (2005) нар ашигласан байна. Энэ жишээ нь эдийн засгийн статистикаас Бөөний болон жижиглэнгийн худалдааны салбарын бүтцийн тооцоо юм. Бөөний худалдааны салбарын тоон өгөгдөл нь 875 компанийн 102 хувьсагчидтай. Эдгээр хувьсагчуудад 30 тэгшитгэлийн хязгаарлалт байдаг бөгөөд “хэрэв $y_1 = 0$ бол $y_2 = 0$ ” гэсэн хэлбэрийг ашигласан харилцан холбоо бүхий 26 энгийн орлуулгын дүрэмтэй ба эдгээр хувьсагчид бараг бүгд эерэг. Жижиглэнгийн худалдааны салбарын тоон өгөгдөл нь 1242 компанийн 54 хувьсагчидтай, 15 тэгшитгэлийн хязгаарлалттай ба бөөний худалдааны салбартай ижил хэлбэрийн 21 энгийн орлуулгын дүрэмтэй байв.

Энэхүү эдийн засгийн статистикийн бүтцийн тоон өгөгдөл нь маш тодорхой алдааг засварлах хэд хэдэн алхмуудыг аль хэдийн гүйцэтгэсэн байна. Үүнд сөрөг утгатай ажиглалтыг тодорхойлох эсвэл маш олон хэлбэрийн алдаанууд засварлах ордог. Цаашилбал энэ алхмыг хийх явцад дэд дүнг бөглөсөн бол хоосон нийт дүн болон дэд дүн бөглөгдсөн байна. Энэ алхам нь дедуктив орлуулга хийх эхний алхам юм. Бидний жишээний тоон өгөгдөлд тэгшитгэлийн хязгаарлалт, энгийн орлуулгын дүрмүүд зэрэг бүх боломжтой дедуктив орлуулгыг гүйцэтгэсэн. Үр дүнг дараах хүснэгтэд харуулав.

| <i>Бөөний болон жижиглэн худалдааны салбар дах ойролцоолох орлуулгын тоо</i> | | | |
|--|--------------------------|----------------|--------------------------------|
| | Бөөний худалдааны салбар | | Жижиглэнгийн худалдааны салбар |
| Үл мэдэгдэх утгын тоо | 35068 | | 27693 |
| Ойролцоолох орлуулга хийсэн тоо | 24048 (69%) | | 12927 (47%) |
| Тэгтэй тэнцүү утгын тоо | | 22647 (94%) | 11708 (91%) |
| Тэгтэй тэнцэхгүй утгын тоо | | 1401 (6%) | 1219 (9%) |
| Үлдсэн үл мэдэгдэх утгын тоо | 11020 | | 14766 |

Энэ хүснэгтэд дедуктив орлуулгыг ашиглах нь өндөр үр дүнтэй байгааг харуулж байна. Ийм байдлаар үл мэдэгдэх утгын ихээхэн хэсгийн (67%, 47%) оруулгын загваргүйгээр зөвхөн боломжит утгууд, энгийн засварлах дүрмийг ашиглан нөхдөг. Хүснэгтийн дедуктив орлуулгаас харахад орлуулсан утгын ихэнх нь (90-ээс дээш хувь) тэгтэй тэнцүү байна. Эдгээр нь зөвхөн дедуктив орлуулга биш гэдгийг зааж өгөх хэрэгтэй. Учир нь үл мэдэгдэх буюу зарим үзүүлэлт хоосон байгаа нь 0 утгыг бичилгүй орхисон байж болох юм. Дедуктив орлуулгыг ашигласнаар эдгээр олон тооны 0 утгыг нөхөх боломжтой болно.

3. ДУНДАЖ ОРЛУУЛГА БА БҮЛГИЙН ДУНДАЖ ОРЛУУЛГЫН АРГА

Дундаж орлуулгын арга нь судалгаанд хариулсан оролцогчдын хувьсагчдын дундаж утгыг тооцож үл мэдэгдэх утгад орлуулдаг.

Бүлгийн дундаж орлуулга, бодит утгатай ба ижил дэд бүлэгт байдаг хариултгүй оролцогчдын хувьсагчдад үл мэдэгдэх утгыг дундаж дүнгээр орлуулах арга юм. Үл мэдэгдэх утга болгонд ижил дундаж дүн орлуулсан учир дундаж орлуулга нь тархалтыг хэвээр хадгалдаг.

Дундаж орлуулгад нэмэлт мэдээлэл ашигладаггүй. Тиймээс нэмэлт мэдээлэл байхгүй эсвэл байгаа нэмэлт хувьсагч нь орлуулах у хувьсагчтай бага зэргийн хамааралтай нөхцөлд, дундаж орлуулгын арга зүй нь зохимжтой байдаг. Хэрвээ хувьсагчийн үл мэдэгдэх утгын хэмжээ маш бага ба үр дүнд бага нөлөөлөхөөр бол дундаж орлуулгын аргыг ашиглахыг зөвшөөрдөг. Гэхдээ дээр дурдагдсан тохиолдолд л энэхүү аргыг ашиглах хэрэгтэй.

Нэмэлт мэдээллийг бүлгийн дундаж орлуулгад ашигладаг, ба үүнд нэг болон хэд хэдэн чанарын хувьсагчдад үндэслэн бүлгүүдийг (дэд зүйлүүд, орлуулгын бүлэг) –д хуваадаг. Судалгаанд оролцогчид дундаас үр дүнгээр ялгасан дэд бүлгийн ангилалд үндэслэн тооцсон дундаж нь орлуулах хувьсагчийн дэд бүлгүүдтэй илүү төстэй байх тусмаа орлуулга нь сайн болдог.

Дундаж орлуулга нь тархалтыг хэвээр хадгалсан үр дүнг гаргадаг. Тиймээс үр дүн нь эх олонлогийн дундаж болон нийт дүнгийн үнэлгээгээр хязгаарлагдвал энэхүү арга зүйг ашиглах нь тохиромжтой. Орлуулга нь судалгааны мэдээллийн санг бүрэн мэдээлэл болгож үр дүнг бодитой тооцоолох баталгаа болдог. Харин дундаж орлуулга нь орцын (эсвэл бусад зүйлсийн) тархалт эсвэл стандарт хазайлт гэх мэт тархалтын хэмжигдэхүүнийг үнэлэхэд тохиромжгүй юм. Энэ нь ерөнхийдөө бусад орлуулгаас чанарын ялгаа багатай боловч бусад ямар ч орлуулга ийм төрлийн баталгааг олгохгүй.

Бүлгийн дундаж орлуулгын хувьд, орлуулгад бүлэг хоорондын хувьсагчийн дунджийг оруулдаг учраас бүлэг доторх хувьсагчтай хамааралгүйгээр тархалтын өөрчлөлт нь ихэнхдээ маш бага байдаг. Хэрэв энэхүү бүлэг хоорондын болон

бүлгийн дисперсийн харьцаа их байвал орлуулгын загварын оновчтой байдлыг харгалзан тархалтын үнэлгээг үндэслэлтэй тооцоолоход ашиглаж болно.

Нэгдүгээр бүлэг дэх ерөнхий тэмдэглэгээний дагуу y_i үл мэдэгдэх утгыг \tilde{y}_i утгаар орлуулна гэвэл дундаж орлууга нь орлуулах \tilde{y}_i утга нь ажиглагдсан утгуудын дундажтай тэнцүү байна.

$$\tilde{y}_i = \bar{y}_{obs} = \frac{\sum_{k \in obs} y_k}{n_{obs}} \quad (3.1)$$

y_k нь k^{th} оролцогчийн ажиглагдсан утга ба y хувьсагчийн хувьд хариулсан утгуудын тоо нь n_{obs} .

Оролцогчдыг тэгш бусаар жинлэж болно жишээ нь орлуулах магадлалаар нь жинлэж болно; 1 дүгээр бүлэгт энэ талаар дурдсан. Энэ тохиолдолд, эх олонлогийн нийт тоо хүртэл өсгөхөд нь тогтмол өсгөх хүчин зүйл N/n (N нь эх олонлогийн хэмжээ, n нь хариулагчдын тоо эсвэл түүврийн хэмжээ) ашигладаггүй, харин өөр хоорондоо ялгаатай w_i жингүүдийг ашигладаг. Энэ оруулгын үр дүн

$$\tilde{y}_i = \bar{y}_{obs}^{(w)} = \frac{\sum_{obs} w_k y_k}{\sum_{obs} w_k} \quad (3.2)$$

нь ихэнхдээ илүү дээр байдаг ба эх олонлогийн дунджийн хазайлтгүй үнэлэлт юм.

Дундаж орлуулгыг түүвэрт судалгааны хариулаагүй оролцогчид эсвэл эх олонлогийн үл мэдэгдэх утгад ашиглан орлуулж болно. Үл мэдэгдэх утга бүрд ижил дундаж орлуулдаг. Ихэнх тохиолдолд, эхлээд орлуулгын бүлгүүдийг тодорхойлсны дараа энэ аргыг ашиглах нь үр дүнтэй байдаг. Энэ бүлгийн дундаж орлуулгыг, (3.1)-д сольж бичвэл

$$\tilde{y}_{hi} = \bar{y}_{h;obs} = \frac{\sum_{h \in obs} y_{hk}}{n_{h;obs}} \quad (3.3)$$

h бүлгийн k^{th} оролцогчийн ажиглагдсан утга нь y_{hk} бөгөөд h бүлгийн хувьсагчийн хувьд хариулсан утгын тоо нь $n_{h;obs}$.

Дундаж болон бүлгийн дунджийг орлуулахад ямар нэгэн иж бүрэн програм хангамж хэрэг болохгүй. Дундаж эсвэл бүлгийн дундаж орлуулгыг SPSS програмын Transform\ Replace Missing values\Method Series mean командыг ашиглан амархан орлуулж болдог. Replace Missing values гэсэн цэс нь хугацааны цувааны мэдээлэлд зориулагдсан ба өөр хугацааны ижил мэдээллээр орлуулах (longitudinal imputation) аргад мөн ашиглаж болно (Бүлэг 8).

Тиймээс, дундаж орлуулгыг дараах зүйлсээр орлуулж болно:

- Эх олонлогийн, түүврийн эсвэл орлуулгын бүлэг бүрийн дундаж
- w_i жингээр жинлэсэн болон жинлээгүй

Энэ загварын стандарт алдааг ашиглах хувилбарыг 5-р бүлгийн регрессийн орлуулгад авч үзнэ.

Дундаж орлуулгын шинж чанар:

1. Дундаж орлуулгыг (3.1)-н дагуу бүх хариултгүй утгуудад орлуулсны дараа жинлээгүй түүврийн дундаж нь жинлээгүй хариултын дундажтай ижил байна. Боломжит хариултгүй утгуудын хувьд судалгаанд оролцогчдын хариултын дунджаар их хэмжээний орлуулга хийвэл, энэ аргаар тооцоолсон эх олонлогийн дундаж нь хариултын дундажтай тэнцүү бөгөөд эх олонлогийн дундаж шууд тооцоолсон утгатай тэнцүү байна. (Нэмэгдүүлсэн жинтэй хамт N/n)
2. Үүнтэй нэгэн адил, давхардлыг орлуулгын ангилал шиг хийвэл, бүлгийн дундаж орлуулга нь давхардал эсвэл давхардлын дараах тооцоологчийн ерөнхий дүн болон дундажтай ижил үр дүн гаргана.
3. (3.2)-д заасны дагуу бүх хариултгүй объектуудад жинлэсэн дундаж тооцоог хэрэглэсний дараа хариултгүй объектуудын жингээс үл хамааран магадлалын хамт жинлэсэн түүврийн дундаж нь магадлалын хамт жинлэсэн хариултын дундажтай тэнцүү байна. Энд, орлуулгыг эх олонлогийн тооцоололд нөлөөлөхгүй байлгах бололцоог жинлэлт (нэмэлт) нь хангаж өгдөг. Үүнтэй адил, их хэмжээний орлуулгын дараа хүн амын дундаж нь жинлэсэн хариултын дундажтай тэнцүү байна.

Дундаж орлуулгын үр дүн нь y орлуулах хувьсагчийн S_y^2 дисперсийг дутуу үнэлдэг,

$$\hat{S}_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

Учир нь, оролцогчийн хариулаагүй утгын хувьд хүртвэр нь тэг байдаг. \bar{y} түүврийн дунджийн дисперс $V(\bar{y})$ -ийн хувьд хялбаршуулсан үнэлгээг ашиглаж болно

$$\hat{V}(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{1}{n} \hat{S}_y^2 = \left(1 - \frac{n}{N}\right) \frac{1}{n(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2$$

Мэдэгдэж буй y_i -үүд эсвэл орлуулсан \bar{y} , тэгвэл энэ дисперсийг (стандарт алдааны квадрат) мөн дутуу үнэлсэн байх тул итгэх интервалыг авч үзэх хэрэгтэй болно. Буруу үнэлгээг ашиглавал зөвхөн судалгаанд хариулсан y оролцогчдын хувьд биш, бүх n оролцогчийн хувьд өгөгдөл идэвхтэй байна. Зөв үнэлгээг $V(\bar{y})$ -ын хувьд дээрх томъёонд түүврийн хэмжээ n сольсноор, n_{obs} судалгаанд хариулсан оролцогчдын тоогоор болон оролцогчдын утгад зөвхөн тодорхойлсон S_y^2 -аар тооцоолсон. Мэдээжийн хэрэг түүврийн дундаж y нь хариулсан утгын дундажтай тэнцүү байна.

4. ХАРЬЦААНЫ ОРЛУУЛГЫН АРГА

Харьцааны орлуулга нь y хувьсагчийн хувьд y хувьсагчтай ойролцоо пропорциональ, хүчтэй хамааралтай нэг нэмэлт мэдээлэл x хувьсагч ашигладаг. Хэрэв R нь x болон y хоорондын хамаарлыг илэрхийлдэг бол, y_i үл мэдэгдэх утгыг нь дараах байдлаар орлуулна.

$$\tilde{y}_i = Rx_i \quad (4.1)$$

Жишээ нь, компанийн тодорхойгүй байгаа бараа эргэлтийг (y) ажилчдын тооноос (x) тодорхойлж болно. R -ийн хувьд компанийн дундаж бараа эргэлтийг ажилтан бүрд ашиглана. Хамгийн түгээмэл тохиолдол нь x нь y -тэй ижил хэлбэрээр хэмждэг боловч өмнөх ажиглалтын үеийг ашигладаг. Тэгээд, y болон x хувьсагчуудыг y^t болон y^{t-1} гэж тэмдэглэдэг. Томъёо (4.1) нь дараах

$$\tilde{y}_i^t = Ry_i^{t-1} \quad (4.2)$$

болж өөрчлөгддөг, R нь $t-1$ хугацаанаас t хугацаан дахь хувьсагчийн харьцангуй өсөлт. Ерөнхийдөө R нь судалгааны мэдээллээс тооцоологдсон.

Харьцааны орлуулгад хэрэв x тоон хувьсагч нь зорилтот y хувьсагчаас илүү эсвэл бага харьцаатай бол тоон хувьсагч y -ийн үл мэдэгдэх утгыг орлуулахад ашиглаж болно. (4.1) томъёог регрессийн шугам нь тайлбарлах утгаас хамаарсан энгийн регресс тэгшитгэл гэж харж болно. Энэ нь ямар ч тогтмол утга загварт ашигладаггүй гэсэн үг. Тиймээс харьцааны орлуулга нь регресс шинжилгээний (жинлэсэн хамгийн бага квадратын үнэлсэн) онцгой тохиолдол юм. Хэрэв загвар нь тогтмол утга оруулж илүү сайжруулсан эсвэл (4.1) томъёонд илүү хувьсагч нэммээр байвал, ерөнхий регрессийн орлуулгын арга зүйг ашиглах нь илүү тохиромжтой байх.

Харьцааны орлуулгын арга зүй ашигладаг ихэнх статистикийн дундаж болон нийт утгууд нь гол үр дүн болдог. Өмнөх оноос бараа эргэлт нь өссөн эсвэл буурсан компаниудын тоотой хүснэгтэд зарим бараа эргэлтэд тэдгээр компаниуд нөлөө үзүүлдэггүй. Хэрэв (4.2)-н дагуу орлуулга хийж R нь 1.01 байхаар тооцвол, бүх хариултгүй утгуудын орлого нь $t - 1$ -эс $t - n$ хоорондох хугацаанд өссөн гэж авч үзэх нь энэ нөхцөлд юу л бол. Тийм учраас энэ хүснэгтийн хувьд, (4.2)-д үлдэгдэл нэмэх нь чухал юм. Энэ үлдэгдлийг цаашид үргэлжлүүлэн 5-р бүлэгт авч үзнэ, мөн 1-р бүлгээс харна уу.

Дундаж орлуулга шиг, дэд олонлог (орлуулгын бүлэг) тус бүрд нь харьцааны орлуулгыг ашиглаж болно. Дэд олонлог хоорондын харьцаа их тохиолдолд ихэвчлэн ашигладаг.

Ихэнх тохиолдолд y -ээс их эсвэл бага пропорциональ байх нэмэлт x хувьсагч байдаг. Хэрэв y_i нь үл мэдэгдэх, x_i нь мэдэгдэж байвал 4.1 томъёоллоор орлуулга хийж болох ба R нь пропорциональ тогтмол байна. Ерөнхийдөө R нь тодорхойгүй байдаг ба x болон y нь мэдэгдэж байгаа оролцогчдын мэдээллээс тооцдог.

$$\hat{R} = \sum_{obs} y_i / \sum_{obs} x_i \quad (4.3)$$

Үүнийг 4.1-д орлуулбал $\tilde{y}_i = \hat{R}x_i = \frac{\sum_{obs} y_i}{\sum_{obs} x_i} x_i \quad (4.4)$

Тэгэхээр пропорциональ тогтмол гэдэг нь судалгаанд хариулсан у хувьсагчийн утгын хувьд x болон y -ийн дунджуудын харьцаатай тэнцүү байна.

Хэрэв x болон y нь тухайн хугацааны ялгаа байвал 4.2 томъёоны дагуу өөрчлөгдөнө

$$\tilde{y}_i^t = \hat{R}y_i^{t-1} = \frac{\sum_{obs} y_i^t}{\sum_{obs} y_i^{t-1}} y_i^{t-1} \quad (4.5)$$

Тооцсон параметр R нь энэ тохиолдолд хувьсагчийн $t - 1$ -ээс t хүртэлх хугацааны харьцангуй өсөлт юм.

Загвар 4.4 –ийг өөр өөр дэд олонлогт тусад нь тооцож болно. h дэд олонлогуудын хувийн харьцаа нь R_h байна. Үүнийг бүлгийн харьцааны орлуулга гэж хэлж болох юм. Энэ аргыг x болон y –ын хооронд хүчтэй шугаман хамааралтай, мөн дэд олонлогуудын хооронд эс ялгаатай байвал ашиглахад тохиромжтой. Дэд олонлогууд нь хэтэрхий жижиг байж болохгүйг, учир нь нийт үнэлгээний хувьд стандарт алдааг ихэсгэж болзошгүй юм. Бүлэг, дэд бүлгүүдэд тооцоолол хийхэд бүлгийн харьцааны нь орлуулга нь бүлгийн дундаж орлуулгаас харьцангуй бага ашигтай байдаг. Бүлгийн харьцаа нь бүлгийн дунджаас илүү нэгэн төрлийн байдаг.

R –ийн харьцааг тодорхойлохдоо оролцогчдын жинг харгалзан үзэж жинтэй утгаас тооцоолол хийх сонголт байж болно.

Харьцааны орлуулгыг тооцоход нарийн төвөгтэй программ шаардлагагүй юм. R харьцааг үнэлсний дараа 4.4 болон 4.5 томъёоллыг хялбархан тооцох боломжтой.

Шинж чанарууд

- Харьцааны орлуулгын нэг тохиолдолд нь $R = 1$ гэж ашиглах гүйцэтгэх юм. Үүний утга нь \tilde{y}_i -ийг x_i -тэй тэнцүү гэж үзэн орлуулах юм. x хувьсагч нь y хувьсагчийн хувьд “итгэмжлэгдсэн төлөөлөгч хувьсагч” нь болно. Хэрэв x -ийг гадаад эх үүсвэрээс орлуулбал үүнийг “cold deck орлуулга” гэж нэрлэдэг (Бүлэг 6). Тухайлбал y_i^t үл мэдэгдэх утгад өмнөх хугацааны y_i^{t-1} утгыг оруулах нь гадаад эх үүсвэрээс оруулах юм. Хугацаа өнгөрөх тусам тогтвортой байдаг хувьсагчдын энэ аргыг ашиглах боломжтой боловч ихэнхдээ R -ийг 1-тэй тэнцүү гэж үзэхийн оронд тооцоолохыг илүүд үздэг.
- Харьцааны орлуулгад $\sum y_i / \sum x_i$ харьцаа өөрчлөгддөггүй. Хэрэв харьцааны үнэлгээг y хувьсагчийн хувьд x нэмэлт хувьсагчийг ашиглан түүврээс эх олонлог хүртэл өсгөсөн бол эх олонлогийн үнэлгээ нь орлуулсан утгаас хамаарч өөрчлөгддөггүй.

Хэрэв загварт хазайлтыг оруулж тооцоогүй бол харьцааны орлуулга нь $y_i - Rx_i$ –ийн утгуудын тархалтыг дутуу үнэлдэг. Хэрэв эх олонлогийн дунджийн үнэлгээний дисперсийн хувьд харьцааны үнэлгээг ашигласан бол дараах байдлаар тооцно.

$$\hat{V}(\hat{Y}_R) \equiv \hat{V}(\hat{R}\bar{X}) = (1 - \frac{n}{N}) \frac{1}{n(n-1)} \sum_{i=1}^n (y_i - \hat{R}x_i)^2$$

5. РЕГРЕССИЙН ОРЛУУЛГЫН АРГА

Регрессийн орлуулга нь y_i үл мэдэгдэх утгын хувьд оновчтой урьдчилан таамагласан регрессийн загвараар нэг болон түүнээс дээш x хувьсагчийн тусламжтайгаар y хувьсагчийг тодорхойлдог. Загварийн параметрууд нь y болон x хувьсагчуудын хариулсан утгуудыг ашиглан үнэлэгддэг.

Регрессийн орлуулгад зорилтот y хувьсагч нь тоон хувьсагч байна. Регрессийн загварын тайлбарлагч нэмэлт хувьсагчид нь тоон хувьсагч байх боловч хуурмаг хувьсагч, чанарын хувьсагчийг загварт оруулж болно. Энэ нь шугаман регрессийн шинжилгээ гэх ба мөн “дисперсийн шинжилгээ” гэж нэрлэдэг. Ийм регресс нь онолын хувьд боломжгүй зүйлийг бий болгодог боловч хэрэв y хувьсагч нь зөвхөн бүхэл тоо бол регресс нь бутархай тоо гардаг. Хот-дек орлуулга нь тодорхой хэмжээгээр регрессийн шинжилгээний нэг хэлбэр гэж ойлгож болох бөгөөд ийм (зөвхөн бүхэл тоо үед бутархай тоо гарах асуудал) асуудал үүсдэггүй.

Өмнөх загваруудад регрессийн шинжилгээний онолын талаар авч үзээгүй боловч ерөнхий мэдээллүүдийг дурдсан. Шугаман болон бусад хэлбэрийн регрессийн талаар хангалттай мэдээлэл байдаг.

Регрессийн орлуулга нь x_1, x_2, \dots, x_p нэмэлт хувьсагчдын тусламжтайгаар регрессийн загвар ашиглаж y хувьсагчийг таамаглах юм. Регрессийн загвар нь

$$y = \alpha + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon = \alpha + \beta'x + \varepsilon \quad (5.1)$$

X нь x_1, x_2, \dots, x_p хувьсагчид бүхий p вектор, α скаляр параметр, β нь p векторын параметрууд, $\varepsilon \sim N(0, \sigma^2 I)$ нь бие даасан, σ^2 дисперстэй хэвийн тархалттай вектор. Мөн α тогтмол орхих замаар тогтмол утгагүйгээр регресс тооцох боломжтой.

α болон β_1, \dots, β_p параметруудыг y болон нэмэлт хувьсагчдын ажиглагдсан (хариулт өгсөн) утгуудыг ашиглаж үнэлдэг. Параметруудийг үнэлдэг хамгийн түгээмэл арга нь хамгийн бага квадратын арга ба үнэлэгдсэн параметруудийг $\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_p$ гэж тэмдэглэдэг. Үнэлэгдсэн тэгшитгэлийг дараах байдлаар бичнэ.

$$\hat{y} = \hat{\alpha} + \hat{\beta}'x \quad (5.2)$$

Энэхүү таамаглах хувьсагчид нь хариулсан болон хариулаагүй утгуудын аль алинд тодорхойлогддог.

Хариулаагүй утгын хувьд \tilde{y}_i орлуулах хоёр арга байдаг

1. Үлдэгдэлгүй

$$\tilde{y}_i = \hat{y} = \hat{\alpha} + \hat{\beta}'x \quad (5.3)$$

2. Үлдэгдэлтэй

$$\tilde{y}_i = \hat{y} + e_i = \hat{\alpha} + \hat{\beta}'x + e_i \quad (5.4)$$

Үлдэгтэй эсэхийг тодорхойлох хоёр хэлбэр байдаг

а. $e_i = e_d$, e_d нь дурын эсвэл тусгайлан сонгосон донорын үлдэгдэл

б. e_i нь σ^2 дисперстэй, 0 дундажтай хэвийн тархалтаас сонгосон үлдэгдэл

Аль тохиолдолд үлдэгдлийг регрессийн загвар ашиглаж тодорхойлно.

Шугаман бус регрессийн ерөнхий загвар

$$y = f(\beta'x) \quad (5.5)$$

Мөн энэ загварт ε үлдэгдлийг нэмэх боломжтой эсвэл түүнд агуулагдсан байж болно.

y нь 0 болон 1 гэсэн хоёрдмол утгатай байвал логистик регрессийг ашигладаг.

$$\text{Ln} \frac{p}{1-p} = \alpha + \beta_1 x_1 + \dots + \beta_p x_p \equiv \alpha + \beta'x \quad (5.6)$$

p нь өгөгдсөн загвар болон x хувьсагчийн хувьд y хувьсагчийн утга 1 байх магадлал.

SPSS программын SPSS \ Analyze \ Regression цэсийг ашиглаж шугаман болон шугаман бус регрессийн урьдчилан таамагласан утгуудыг тооцож SAVE \ Unstandardized predicted values-аар хадгалах боломжтой . Энэ тохиолдолд PRE_1 гэдэг ч юм уу товч нэр бүхий \hat{y}_i шинэ хувьсагч үүсгэх бөгөөд үүнд хариулсан болон хариулаагүй утгуудын таамагласан утгууд байна. y хувьсагчид орлуулга хийсний дараа хариулт өгсөн оролцогчдын хувьд \hat{y} утгад y_i бодит утгуудыг сольж тавина.

Шинж чанар

- Хэрэв 5.1 тэгшитгэлд x нэмэлт хувьсагч ашиглаагүй бол μ y хувьсагчийн хүлээгдэж буй үнэ цэнэ бүхий $y = \mu + \varepsilon$ тэгшитгэл болно. Мөн 5.3 тэгшитгэл нь $\tilde{y}_i = \hat{\mu} = \bar{y}$ болно. Энэ нь дунжийн орлуулга юм (Бүлэг 3)
- Хэрэв зөвхөн тоон нэмэлт x_1 хувьсагч болон тогтмол утга ашиглаагүй бол 5.1 тэгшитгэл нь $y = Rx + \varepsilon$ болж, 5.3 нь 4.1 болж өөрчлөгдөнө. Гетерогенетий таамаглалын дагуу жинлэгдсэн хамгийн бага квадратын аргаар тооцсон регрессийн нь 4.4 тэгшитгэлийн харьцааны орлуулга болно.

Чанарын үзүүлэлтүүд

Орлуулгын чанарын талаар мэдэж байх нь чухал байдаг. Үүнтэй холбоотой асуудал нь мэдээллийн бодит үнэ цэнэ нь ихэвчлэн тодорхойгүй байдаг. Ихэнх тохиолдолд орлуулгын дараа болон өмнөх дунджууд нь ялгаатай байна. Энэ нь санаа зовох шалтгаан биш юм, учир нь хариулаагүй утгаас шалтгаалсан. Хэрэв бусад хийгдсэн судалгаа болон харьцуулах мэдээлэл байвал орлуулгын чанарыг баталгаажуулж авах боломжтой. Гэсэн хэдий ч судалгаануудын тодорхойлолт болон эх олонлогууд нь харилцан адилгүй байдаг учир эдгээр баталгаажуулалтыг хийх боломж хязгаарлагдмал байдаг.

Ерөнхийдөө орлуулгын чанарыг бодитой үнэлэх боломжгүй байдаг тул регрессийн орлуулгын чанар нь зөвхөн судалгаанд оролцогчдод тохирсон загварыг ашигласан эсэхэд үндэслэнэ.

- Хамгийн бага квадратын аргаар үнэлсэн шугаман регрессийн шинжилгээний хувьд R^2 буюу детерминацийн коэффициентээр загварын оролцогчдын тайлбарлах чадварыг тодорхойлдог ба өөр өөр орлуулгын загваруудыг хооронд харьцуулж үзэх боломжтой юм.
- Өөр нэг орлуулгын загварын чанарыг шалгах арга нь симуляцын туршилт хийх юм. Хариулсан утгуудыг түр хасаж, хасагдсан утгуудын оронд шинж тооцсон утгуудаар орлуулна. Бүх судалгаанд оролцогчдын утгыг нэг нэгээр нь эсвэл жижиг бүлгээр нь сольж болох боловч судалгаанд оролцогчдын тодорхой хэсгийг солихоор хязгаарлалт хийх боломжтой. Хэрэв тоон y хувьсагчдын хувьд сольж тавьсан \tilde{y}_i утгууд нь анхны y_i -ийн утгуудтай тэнцүү байвал орлуулгын загварт итгэх үндэслэлтэй болно. Тохиромжтой зайны функцийг тодорхойлсноор хамгийн тохиромжтой аргыг эсвэл загварыг сонгох боломжтой болдог. Зайн функцийг жишээ нь орлуулсан утгаас, бодит утгын хазайх абсолют дундаж хазайлт $\frac{1}{I} \sum_{i=1}^I |\tilde{y}_i - y_i|$, I нь орлуулсан утгуудын тоо.
- Орлуулсан утга болон нийт үр дүн аль алины хувьд бусад эх сурвалжийн мэдээлэлтэй тохируулж үзэх боломжгүй байдаг. Судалгаанд хариулаагүй оролцогчдтой хандаж дутуу мэдээллийг олж авах нь мөн хялбар биш.

6. ХОТ-ДЕК ОРЛУУЛГЫН АРГА

Хот-дек орлуулга нь судалгаанд оролцогчийн i хариулаагүй утга бүрийн хувьд боломжит ижил шинж чанартай, y орлуулах хувьсагчид нөлөөлөх боломжтой d донор утгыг судалгааны мэдээллээс хайдаг. Энэ донор утгыг орлуулгад y_d гэж тэмдэглэдэг

$$\tilde{y}_i = y_d$$

Судалгаанд оролцогчийн хариулаагүй утгыг “хүлээн авагч” гэж нэрлэдэг.

Донор утгыг олох янз бүрийн аргууд байдаг. Эдгээрийг дотор нь дараах байдлаар хувааж болно.

1. Орлуулгын бүлгүүдийг ашигладаг аргууд
2. Зайны функцийг ашиглаж зайг багасгах замаар донорыг хайх (хамгийн ойрын хөрш)

Эхний бүлгийн жишээ бол санамсаргүй хот-дек болон дараалсан хот-дек орлуулга юм. Санамсаргүй хот-дек орлуулгын хувьд орлуулгын бүлэг нь ангилсан нэмэлт хувьсагчдаас (суурь шинж чанарууд) бүрддэг. Энэ арга нь судалгааны оролцогчийн хариулаагүй утгатай ижил шинж чанар бүхий потенциалт донор хувьсагчдын бүлгээс донорыг санамсаргүйгээр нэгийг сонгодог. Дараалсан хот-дек орлуулгын арга нь бүлгүүдийг тодорхой үүсгэдэггүй боловч судалгаанд оролцогчийн хариулаагүй утга бүрийн хувьд судалгааны мэдээллээс ижил шинж чанартай утгуудаас зорилтот хувьсагчтай хамгийн ойролцоо утгыг хайж орлуулдаг.

Хоёрдугаар бүлгийн онцлох жишээ нь сонгосон регрессийн загвараар у-ийн утгыг таамаглаж түүний хамгийн ойрын хөршийн донор утгын урьдчилан таамагласан дунджийг тулгах юм.

Хот-дек орлуулгын аргаас гадна колд-дек (Cold desk) орлуулгын арга гэж байдаг. Энэ арга нь бусад судалгааны мэдээллээс орлуулдаг. Жишээлбэл хугацааны өмнөх үеийн ижил хувьсагчаар ижил утгыг сонгож орлуулах юм. Энэ утгаараа колд дек нь донор орлуулгыг алдаатай хийх сул талтай байдаг. Судлаачид колд дек орлуулгын аргыг баталгаатай арга гэж үздэггүй ба энэ аргыг бараг хэрэглэдэггүй. Хэрэв бусад судалгааны мэдээллийн орлуулга нь ойролцоо утгатай бол үүнийг дедуктив эсвэл логик орлуулга (Бүлэг 2) гэж үзэж болно.

Нэг хувьсагчийн хувьд олон утга нь үл мэдэгдэх байгаа тохиолдолд мөн хот-дек орлуулгыг ашигладаг. Энд нэг донор утгыг орлуулснаар орлуулалтын зөрүүтэй байдлаас урьдчилан сэргийлдэг. Үүнийг олон хувьсагчийн орлуулгын асуудлыг шийдвэрлэх нэг арга хэрэгсэл гэж үзэж болно (Бүлэг 7).

6.1. Санамсаргүй болон дараалсан хот-дек орлуулга

Хот-дек орлуулгын арга нь ижил төрлийн суурь шинж чанар бүхий ижил төрлийн мэдээллээс орлуулах утгыг хайж олж оруулах арга зүй юм. Жишээлбэл ижил хүйстэй, ижил насны бүлэгт багтах, ижил орон нутгийн, ижил салбарт хөдөлмөр эрхэлдэг нэг оролцогчийн мэдээллээр орлуулна. Энэ санааг дахин авч үзвэл хэрэв хоёр судалгаанд оролцогчийн олон тооны онцлог шинж чанарууд тохирч байвал хувьсах хувьсагчийн утга хоорондоо илүү сайн нийцэн гэсэн үг юм.

Санамсаргүй болон дараалсан хот-дек орлуулгын аргад орлуулах донорууд нь тухайн үзүүлэлтийн суурь шинж чанартай ижил утгатай байх ёстой, өөрөөр хэлбэл

тэдгээр донорууд нь орлуулгын ижил бүлэгт байх ёстой. Хамгийн ойрын хөршийн орлуулга нь орлуулгын бүлэг үүсгэдэггүй ба хүлээн авагч болон донор хувьсагчийн хоорондох х хувьсагчдын хоорондын зөрүүгээр тооцогддог.

Тиймээс санамсаргүй болон дараалсан хот-дек орлуулгад суурь шинж тэмдгүүд нь ижил байх шаардлагатай байдаг. Дээрх жишээнээс харахад судалгаанд хариулаагүй оролцогчтой ижил дөрвөн шинж тэмдэг бүхий хариулагч олж чадахгүй бол орлуулгын бүлэг нь хязгаарлагдмал болох нь ойлгомжтой юм. Ийм судалгаанд оролцогчийн утгыг орлуулах утгыг олохын тулд дөрвөн шинж тэмдгийн дор хаяж нэгийг хасаж эсвэл орлуулгын бүлгүүдийг нэгтгэх шаардлагатай болно.

Хэрэв холбогдох орлуулгын бүлэгт нэгээс олон боломжит донор байвал нэг донорыг санамсаргүйгээр сонгоно. Санамсаргүйгээр сонгохын оронд нэг донор үлдэх хүртэл шинж тэмдгүүдийг нэмж болдог. Гэхдээ нэг судалгаанд хариулагч олон хүлээн авагчийн донор болох нөхцөл байдлаас урьдчилан сэргийлэх хэрэгтэй. Олон хүлээн авагчийн энэ хэлбэр нь у-ийн дундаж болон нийт дүнгүүдийн стандарт алдааг нэмэгдсэний улмаас хэт зөрүүтэй утга бий болох эрсдэлийг ихэсгэдэг. Жишээлбэл Судалгааны мэдээллийг нэг бүрэн шалгасны дараа орлуулгын бүлэгд олон доноруудыг нэмж бүлгийг томруулах замаар үүнээс урьдчилан сэргийлэх боломжтой.

Дараалсан хот-дек орлуулгын арга нь судалгаанд хариулаагүй утга бүрийн хувьд ижил суурь шинж тэмдэгтэй оролцогчдыг судалгааны мэдээллээс дараалан хайж у утгад оруулдаг тухай тайлбарласан. Мэдээжийн хэрэг тэдгээрийн суурь шинж тэмдгүүдтэй ижил судалгаанд оролцогчийн мэдээллийг дахин ашиглах боломжтой байдаг.

Хэрэв судалгааны мэдээллийн орлуулгын бүлгүүд нь өөр хоорондоо ойролцоо, тэдгээрээс судалгаанд оролцогчийн хариулаагүй утгыг орлуулсан бол бүлгүүд нь ижил донор өгөх эрсдэлтэй байдаг. Үүнээс урьдчилан сэргийлэхийн тулд дараалсан хот-дек орлуулгын нь нэг судалгаанд оролцогчийг донор утгаар сонгодоггүй ба эхний m судалгаанд оролцогчдын мэдээллээс нэг оролцогчийг санамсаргүйгээр сонгодог. Дараалсан хот-дек орлуулгыг судалгаанд оролцогчдын мэдээллийг санамсаргүйгээр эрэмбэлсний дараа ашигладаг ба үүнийг “санамсаргүй дараалсан хот-дек орлуулгын арга” гэж нэрлэдэг. Дараалсан хот-дек орлуулгын аргыг урьдчилан эрэмбэлэлгүйгээр эсвэл зөвхөн сонгосон шинж тэмдгээр эрэмбэлсний дараа гүйцэтгэж болдог.

Энэ аргын хувьд нэмэлт хувьсагчдыг сонгох нь хэцүү үйл явц байдаг. Энэхүү үйл явцад агуулгад үндэслэн болон статистикийн аргууд чухал үүрэг гүйцэтгэдэг.

6.2. Хамгийн ойрын хөршийн орлуулга

Хамгийн ойрын хөршийн орлуулгын хувьд $d(i; j)$ зайны функц нь судалгаанд хариулаагүй i дугаар утга болон дурын оролцогчийн j дугаар утга хооронд тодорхойлсон зай юм. Зайны функцийг олон янзаар тодорхойлж болдог. Түгээмэл ашигладаг функц нь Минковскийн зайны функц юм. x нь тоон хувьсагчид байна.

$$d(i; j) = \left(\sum_k |x_{ki} - x_{kj}|^z \right)^{1/z}$$

$d(i; j)$ -ийн хамгийн бага утгатай j дугаар оролцогч нь i дугаар оролцогчийн хариулаагүй утгын хамгийн ойрын хөрш бөгөөд донор болно. $z=2$ байвал Минковскийн зай нь Эүклидийн зай болно, $z=1$ байвал байшин хоорондын зай гэж нэрлэдэг. z их байх тусам x_{ki} , x_{kj} хоорондын зай ихэсдэг.

Дээрх функцийг илүү сайжруулсан хэлбэр нь жигнэсэн зайны функц юм

$$d_v(i; j) = \left(\sum_k v_k |x_{ki} - x_{kj}|^z \right)^{1/z} \quad (6.1)$$

v_k нэмэлт хэсэг нь x_k хувьсагчийн жин юм. Учир нь зөвхөн бодит жинтэй хамааралтай тооцооны алдаа гарахгүй ба $\sum_k v_k = 1$ гэж тооцож болно. x хувьсагч бүрийн жинг урьдчилан тодорхойлох зайлшгүй шаардлагатай байдаг. Үнэндээ энэ жинг x хувьсагчийн тархалт эсвэл утгын далайцаас тусад нь харах боломжгүй юм. Жинг x хувьсагчуудын дисперсийг 1-тэй тэнцүү диспертэй болгож нормальчлан тооцвол тодорхойлоход илүү хялбар байдаг.

Мөн $d(i; j)$ функцийг тодорхойлохдоо хувьсагчдын хоорондын ковариацийг харгалзан тооцох боломжтой байдаг боловч энэ нь жинг тодорхойлоход илүү төвөгтэй болгодог. Өөр боломжит зайны функц нь $\max_k v_k |x_{ki} - x_{kj}|$ эсвэл илүү сайжруулсан $\max_k v_k d(x_{ki}, x_{kj})$ юм. Үүнд x -хувьсагчдаас хүлээн авагчтай эрс ялгаатай биш донор хайх хэрэгтэй юм. Энэхүү зайны функц (6.1) нь z хязгааргүй утгад үүсдэг.

Хамгийн ойрын хөршийн орлуулгын нэг хэлбэр нь Литтлийн (1988) тодорхойлсон урьдчилан таамагласан дунджийг тохируулах юм. Энэ орлуулгын арга нь шугаман регрессийг у орлуулах хувьсагчийг өөр тоон тайлбарлах x хувьсагчдаас хамааруулан хийдэг ба регрессийн хувьсагчдад хариулаагүй утгатай оролцогчид мэдээлэл ордоггүй. Дараа нь регрессийн тэгшитгэлийн үр дүнг бүх судалгаанд оролцогчдын хувьд у орлуулах хувьсагчийн утгыг томъёоны дагуу таамаглахад ашигладаг. i дугаар оролцогчийн хариулаагүй утгад j дугаар оролцогчийн утгыг донор болгож өгөхөд тооцсон \hat{y}_j утга нь таамагласан \hat{y}_i хариулаагүй утгатай аль болох ойрхон байх болно. Ингэснээр нь j дугаар y_j утга донор утгаар орлуулагдаж ажиглагдсан утга болох ба $\tilde{y}_i = y_{d=y_j}$ гэж тооцож болно. Хамгийн ойрын хөршийн орлуулгын нэг

хэлбэр болох урьдчилан таамагласан дунджийг тохиоруулах орлуулгын зайны функцийг дараах байдлаар тодорхойлдог.

$$d(i, j) = |\hat{y}(x_i) - \hat{y}(x_j)| \quad (6.2)$$

Хамгийн ойрын хөршийн орлуулгад урьдчилан таамагласан дунджийг тохируулах ордог ба судалгаанд оролцогчдоос хамгийн ойрхон m болон тэдгээрийн нэгийг санамсаргүй сонгох, дараалсан хот-дек-д зайны функцэд бага оноо авсан донор сонгогдох магадлал өндөр байдаг.

7. ОЛОН ШАТАТ ОРЛУУЛГЫН АРГА

Энэ хүртэл нэг хариулаагүй утгатай хувьсагчид хэрхэн орлуулга хийх талаар авч үзсэн. Нэг судалгаанд оролцогчийн хувьд хариулаагүй утга ихэвчлэн олон хувьсагчид байх бөгөөд эдгээр хувьсагчууд нь хоорондоо хамааралтай байдаг. Ийм тохиолдолд бүх үл мэдэгдэх хувьсагчуудыг орлуулах орлуулгыг олон шатат орлуулга гэнэ. Энэ бүлэгт олон талт орлуулгыг бодох хэд хэдэн аргуудын талаар авч үзнэ .

Бүлэг 6-т авч үзсэн хот-дек орлуулга нь үл мэдэгдэх олон хувьсагчийн үед хэрэглэхэд хялбар байдаг. Нэг донор оролцогч нь хүлээн авагчийн бүх үл мэдэгдэх утгуудыг бүрдүүлдэг. Ийм тохиолдолд олон зорилтот хувьсагчуудад ижил төрлийн орлуулгын бүлэг үүсгэх эсвэл, хамгийн ойрын хөршийн хувьсагчийн орлуулгын хувьд зорилтот хувьсагчуудтай хамааралтай нэмэлт хувьсагчууд зайн функцэд байгааг нягтлах хэрэгтэй. Ижил донор оролцогчоос бүх үл мэдэгдэгч зорилтот хувьсагчуудыг авах нь, орлуулсан утгууд өөр хоорондоо нийцтэй эсэхийг баталж байгаа юм. Ерөнхийдөө хүлээн авагчийн жинхэнэ утга болон орлуулсан утгын хооронд нийцэл байх нь баттай зүйл биш хэдий ч донорын сонголтод харгалзсан бодит утга болон орлуулсан утгын хоорондын уялдааг гаргаж авах боломжтой. Ийм хэлбэрийн донор орлуулгыг 6-р бүлэгт дурдсан.

Хэрэв регрессийн орлуулгын үед (болон харьцааны орлуулгын тусгай тохиолдолд) хариулаагүй утгатай олон хувьсагчууд байвал ихэвчлэн таамаглал нь үл мэдэгдэх хувьсагчуудыг агуулж байдаг. Энэхүү асуудлыг хоёр аргаар шийдвэрлэдэг. Нэг шийдэл нь зорилтот хувьсагчдын урьдчилан тогтоосон дараалал дээр суурилсан байдаг. Эхний зорилтот хувьсагчийг зөвхөн үнэлэгдээгүй утга агуулаагүй таамаглалын загварыг ашиглан орлуулна. Дараагийн зорилтот хувьсагчийн хувьд таамаглал нь өмнөх алхам дээрх сонгогдсон орлуулсан хувьсагч болон үл мэдэгдэх утгагүй хувьсагчуудаас гэх мэтчилэн сонгон авч болно. Хоёр дах шийдэл нь таамаглалд орлуулсан хувьсагчийг авч ашигладаггүй, харин зорилтот хувьсагч бүрийн хувьд өөр өөр таамаглал бүхий хэд хэдэн нэмэлт загварыг ашигладаг. Тодорхой оролцогчийн тодорхой зорилтот хувьсагчид хэрэглэх загварыг сонгоход урьдчилж дараалуулсан загваруудаас сонгоно. Хэрэв эхний моделийн таамаглал

ямар нэг үл мэдэгдэх утга агуулаагүй бол тухайн моделийг ашиглана, үгүй бол уг шаардлагыг хангасан дараагийн модел гэх мэтчилэн сонгоно.

Эдийн засгийн статистикт өөр өөр зорилтот хувьсагчуудад хамааралтай хязгаарлалтууд байх тохиолдлууд байдаг. Жишээ нь, нийт бараа эргэлт болон дэд салбарын бараа эргэлтийн тоо мэдэгдэж болох боловч бусад дэд хэсгүүдийн мэдээлэл байхгүй байж болно.

Энэ бүлэгт зорилтот хувьсагчуудыг орлуулахад зориулагдсан нэмэлт хувьсагчууд нь үл мэдэгдэгчийг агуулж байх боломжтой учраас мөн адил зорилтот хувьсагч байх боломжтой гэж үзэж байгаа. Нэмэлт (х-хувьсагчууд) хувьсагчууд болон зорилтот (у-хувьсагчууд) хувьсагчуудын ялгаа үйлчлэхгүй учир энэ бүлэгт бүх хувьсагчийн хувьд у-ийг ашиглана.

Олон хувьсагчийн орлуулгыг шийдвэрлэхтэй холбоотой хот-дек болон регрессийн тооцооллын техникийг хувьсагчийн үнэлгээний түвшинд мөн адил ашиглана. Бүлэг 5, 6-д эдгээр арга техникийг хоёрдмол байдлаар ашиглахыг харуулсан болно.

Бүлэг 5-д ганц хувьсагчийн регрессийн орлуулгыг авч үзсэн. Цаашид бид олон хувьсагчийг регрессийн орлуулгын аргыг ашиглан тооцох болно. Уг тооцооллыг бодох хамгийн хялбар арга бол нэг хувьсагчийн орлуулах аргыг давтан ашиглах явдал юм. Хэрэв хувьсагч бүрийн нэмэлт хувьсагчид хариулаагүй утгыг агуулаагүй бол давтан ашиглах арга нь хоёрдмол утгагүй арга юм. Харин нэмэлт хувьсагчид хариулагдаагүй утгыг агуулсан тохиолдолд оновчтой шийдэлд хүрэхийн тулд янз бүрийн сонголтуудыг авч үзэх хэрэгтэй болно.

Нэг сонголт нь хувьсагч бүрийг эхлээд урьдчилан таамаглаж, зорилтод хувьсагчдыг тодорхой дарааллаар орлуулахыг хэлнэ. Энэ тохиолдолд урьдчилан таамаглах утга үргэлж өгөгдсөн байдаг бөгөөд уг аргыг бүтцийн статистикт ашигладаг.

(эсвэл

Нэг сонголт нь зорилтод хувьсагчдыг тодорхой дарааллаар орлуулах явдал юм. Ингэснээр урьдчилан таамаглагчид хариулаагүй утга агуулж буй зорилтод хувьсагч бүрийг өөрт өгөгдсөн утгуудаар эхлээд орлуулах боломжтой юм. Энэхүү аргыг жишээлбэл Бүтцийн бизнесийн статистикт ашигладаг.)

Өөр нэг сонголт нь зорилтот хувьсагч бүрийн хувьд өөр өөр таамаглалтай загваруудыг тодорхойлох явдал юм. Холбогдох өгөгдлийн оролцогчид ажиглагдсан таамаглалын нэг загварыг орлуулгад сонгож болно. Энэ тохиолдолд урьдчилан таамагласан ямар ч тооцоологдсон утгыг ашигладаггүй.

Урьдчилан таамаглалын эхэлж орлуулсан аргыг (Structural Business Statistics) бүтцийн бизнесийн статистикт хэрэглэдэг орлуулах аргачлалын хялбаршуулсан

тайлбарлая. Харьцааны орлуулгыг бүтцийн бизнесийн статистикт ашигладагтай адилаар бусад олон эдийн засгийн статистикт ашигладаг. Хүснэгт 3 нь харьцааны орлуулгыг тусламжтайгаар хариулагдаагүй утгыг орлуулж буй нэмэлт хувьсагчийг ашигладаг хэд хэдэн зорилтот хувьсагчийн дарааллыг харуулна.

Хүснэгт 3. Бүтцийн бизнесийн статистикийн хувьсагчдын орлуулах диаграмм

| Хувьсагч | Нэмэлт хувьсагч |
|---------------------------------|-----------------------------|
| у1: Бараа гүйлгээ | - |
| у2: Үйл ажиллагааны нийт зардал | Бараа гүйлгээ |
| у3: Нийт ажилчдын зардал | Үйл ажиллагааны нийт зардал |
| у4: Орон байрны зардал | Үйл ажиллагааны нийт зардал |
| у5: Эрчим хүчний зардал | Үйл ажиллагааны нийт зардал |
| у6: Бусад зардал | Үйл ажиллагааны нийт зардал |
| у7: Үндсэн ажилчдын зардал | Ажилтнуудын нийт зардал |
| у8: Бусад ажилчдын зардал | Ажилтнуудын нийт зардал |

Бараа гүйлгээний хувьсагчийг орлуулдаггүй. Энэхүү гол хувьсагч дутагдаж байгаа бичлэгүүдийг (record) хариу өгөхгүй гэж үздэг. Түүнчлэн бүртгэлийг орлуулахдаа бараа гүйлгээг үргэлж авч үзнэ. Бусад хувьсагчдыг 4-р бүлэгт тодорхойлсон харьцааны аргыг ашиглан орлуулна. Бичлэг (i) дэх зорилтот хувьсагчийн (y_i) орлуулгын утгыг (\tilde{y}_{ij}) дараах байдлаар илэрхийлж болно.

$$\tilde{y}_{ij} = y_{ik} * \hat{R}_{jk}$$

Хэрэв зорилтод хувьсагч (y_j) ажиглагдсан бол, (y_{ik}^*) нь зорилтод хувьсагчийн нэмэлт хувьсагч болох (y_k) -ийн утга бөгөөд өөрөөр хэлбэл орлуулагдсан (\tilde{y}_{ik}) утга юм. (\hat{R}_{jk}) нь (y_j) болон (y_k) хувьсагчидтай тогтмол пропорциональ хамааралтай (R_{jk}) -ийн үнэлгээ юм.

Зорилтот хувьсагчдыг тооцох дараалал нь дараах байдалтай байна: нэгдүгээрт, y_1 -ийг ашиглан y_2 -ийг ашиглан y_3 -ийг ашиглан y_7 ба y_8 -ийг тус тус орлуулна. Нэмэлт хувьсагч болгон ашигладаг хувьсагч бүрийг нэмэлт хувьсагч болгон ашиглахаасаа өмнө эхлээд орлуулна. Ийм байдлаар нэмэлт хувьсагчдад ажиглагдсан утга эсвэл орлуулсан утга бүхий утгууд үргэлж байдаг.

Нэмэлт хувьсагчийн орлуулагдсан утгын тусламжаар харьцааны орлуулгыг янз бүрийн орлуулгад ашигласан загваруудын хувьд харьцуулах боломжтой байдаг. Гэхдээ орлуулагдсан утгыг нэмэлт хувьсагчид дээр ашигладаггүй. Энэхүү хамаарал нь доор илэрхийлэгдэв. Хэрэв нэмэлт хувьсагч орлуулагдсан бол зорилтот хувьсагчийн орлуулгыг дараах байдлаар илэрхийлэгдэнэ.

$$\tilde{y}_{ij} = y_{ik} * \hat{R}_{jk} = \tilde{y}_{ij} \hat{R}_{jk} = y_{il} \hat{R}_{kl} \hat{R}_{jk}$$

Энд (y_l) нь (y_k) -ийн нэмэлт хувьсагч бөгөөд дээрх илэрхийлэлд (y_{il}) ажиглагдсан гэж авч үзсэн болно. Энэ нь (y_k) орлуулагдсан бичлэгийн хувьд (y_k) -ийн зөрүүгээр бус (y_l) - зөрүүгээр орлуулдгийг баталж байна. $(\hat{R}_{kl} \hat{R}_{jk})$ -ийн үржвэрийг (R_{jl}) харьцааны үнэлэгч гэж ойлгож болно. Хэрэв (R_{jl}) , (R_{kl}) болон (R_{jk}) харьцааны үнэлэгч нь ижил бүртгэл дээр үндэслэгдсэн бол $(\hat{R}_{jl} = \hat{R}_{kl} \hat{R}_{jk})$ илэрхийлэл хүчин төгөлдөр байх бөгөөд (y_l) нэмэлт хувьсагч байх үеийн харьцааны орлуулга нь орлуулагдсан утгатай тэнцүү юм. Дээр тайлбарласан аргыг дараах байдлаар харьцуулж болно: Хэрэв (y_k) нэмэлт хувьсагч ажиглагдсан бол түүнийг ашиглан (y_j) -г орлуулах эсвэл ажиглагдаагүй тохиолдолд (y_l) нэмэлт хувьсагчийг ашиглан (y_j) -г орлуулах. Энэ нь нэг зорилтот хувьсагчийн хувьд өөр өөр загварыг тодорхойлсон жишээ юм.

Зорилтот хувьсагч тус бүрийн хувьд өөр өөр тодорхойлсон загвар болон таамаглагчдын сонгосон загвар нь ерөнхийдөө регрессийн орлуулгад хэрэглэгдэх боломжтой. Уг аргын сул тал нь урьдчилан таамаглагчдын орлуулгаас илүү олон загварыг авч үзэх шаардлагатай байдагт юм. Гэсэн хэдий ч хамгийн сайн боломжит урьдчилсан загварыг тодорхойлох олон сонголтууд байдгаараа давуу талтай. Жишээлбэл, тодорхой нэг салбарын бүтцийн бизнесийн статистикт нийт ажилчдын зардлын хувьсагч нь нийт үйл ажиллагааны зардлын хувьсагчтай нягт уялдаатай байгаа бол нийт үйл ажиллагааны зардал дээр үндэслэсэн нийт ажилчдын зардлын хариулагдаагүй утга нь нэмэлт хувьсагч болох бөгөөд нийт үйл ажиллагааны зардал дээр үндэслэсэн нийт үйл ажиллагааны зардлын хариулагдаагүй утга нь нэмэлт хувьсагч гэж үзэж орлуулж болно. Хэрэв хоёр хувьсагч орхигдсон бол бараа гүйлгээг эдгээр хувьсагч бүрийн нэмэлт хувьсагч болгон ашиглах боломжтой.

Өмнөх хэсгийн жишээнд харьцааны орлуулгыг нэмэлт хувьсагчдын нийлбэр болох дэд хувьсагчийн хүрээнд ашигласан. Энэ нөхцөл байдал нь эдийн засгийн статистикт байнга гардаг.

Эдгээр нь гол төлөв (y_j) хувьсагчдын $j = 0, \dots, J$ хязгаарлалттай үед хамаарна: (өөрөөр хэлбэл засварлах дүрэм = or edit rule)

$$y_0 = \sum_{j=1}^J y_j$$

Хэрэв нэг дэд хувьсагч (y_j) үнэлэгдээгүй тохиолдолд, уг хариулагдаагүй утгыг (deductive method) дедуктив орлуулгын аргаар хялбархан тооцоолж болно (2-р бүлгийг харна уу).

Үүнээс гадна, хэрэв ажиглагдсан хувьсагчуудын нийлбэр нь "нийт хувьсагч" -тай тэнцүү төдийгүй хариулагдаагүй утга тэг тоон утгыг агуулж байгаа тохиолдолд deductive орлуулгыг хэрэглэх боломжтой. Гэсэн хэдий ч, хэрэв ажиглагдсан дэд хувьсагчдын нийлбэр нь нийт хувьсагчийн утгаас бага бөгөөд хариулагдаагүй утга агуулсан хэд хэдэн дэд хувьсагчууд байгаа тохиолдолд хариулагдаагүй утгуудаас бүрдсэн нийлбэрийг нийт нийлбэрээс салгаж авч үзэх хэрэгтэй.

Энэхүү тархалтыг тодорхойлох нэг арга бол дэд хувьсагчдын харьцааг нийлбэрт ашиглан орлуулсан утгуудын нийлбэр нь ажиглагдсан дэд хувьсагчдын нийлбэрийн зөрүүтэй тэнцэж байхаар шинээр хэмжих явдал юм.

Хэрэв бид (i) бичлэгт ажиглагдаж буй дэд хувьсагчийг ($j = 1, \dots, J_{i,obs}$), мөн үнэлэгдээгүй дэд хувьсагчийг ($j = J_{i,obs} + 1, \dots, J$)-аар индексжүүлвэл, (i) бичлэгт ажиглагдсан дэд хувьсагчийн болон үнэлэгдээгүй дэд хувьсагчийн нийлбэр дараах байдлаар тус тус тооцогдоно.

| | |
|---|---|
| $S_{i,obs} = \sum_{j=1}^{J_{i,obs}} y_{ij}$ | $S_{i,mis} = y_{i0} - S_{i,obs}$ |
| (а. ажиглагдсан дэд хувьсагчийн нийлбэр) | (б. үнэлэгдээгүй дэд хувьсагчийн нийлбэр) |

Шинээр хэмжигдсэн харьцааг ашиглан дэд хувьсагчдын орлуулгын нийлбэрийг дараах байдлаар тэмдэглэв

$$y_{ij} = S_{i,mis} \frac{\hat{R}_j}{\sum_{j=J_{i,obs}+1}^J \hat{R}_j}$$

Шинээр хэмжигдсэн харьцаа нь 1 гардаг тул орлуулсан утгуудын нийлбэр нь ($S_{i,mis}$)-тэй тэнцүү бөгөөд орлуулсан бичлэг нь засварлах дүрмийг хангана.

Үнэлэгдээгүй утгуудын нийлбэр нь мэдэгдэж байгаа гэсэн нэмэлт мэдээллийг ашигладаг харьцааны орлуулгын энэ хэлбэр нь үнэлэгдээгүй нийт утгуудыг ашигладаггүй харьцааны орлуулгаас илүү сайн үр дүнд хүргэдэг.

Регрессийн нэгэн зэрэг орлуулга

Орлуулах аргачлалын талаар олон эх үүсвэр дээр тодорхойлогдсон ерөнхий олон талт регрессийн арга нь зорилтот болон нэмэлт хувьсагчийн нэгэн зэрэг тархаах нь олон талт нормал гэсэн таамаглал дээр үндэслэсэн арга юм. Энэ аргын

тусламжтайгаар орлуулагдсан хувьсагчдын вариацийг төдийгүй бүх хувьсагчдын хоорондын хамаарлыг аль болох нарийвчлалтай байлгах боломжтой стохастик тооцооллыг үүсгэх боломжтой.

Уг аргын үндсэн зарчим бол хариулаагүй хувьсагч бүрийг бүх хувьсагч ажиглагдсан регрессийн загварыг ашиглан таамаглал байдлаар орлуулдаг. Жишээ нь. Судалгаанд оролцогчийн эхний гурван хувьсагч нь хариулагдаагүй утгатай бол бид гурван регрессийн загварыг ашиглан орлуулгын гүйцэтгэнэ (5.3.1-томьёотой төстэй).

$$y_{i1} = \alpha_1 + \beta_1^{y_{i,obs}} + \varepsilon_{i1}$$

$$y_{i2} = \alpha_2 + \beta_2^{y_{i,obs}} + \varepsilon_{i2}$$

$$y_{i3} = \alpha_3 + \beta_3^{y_{i,obs}} + \varepsilon_{i3}$$

энд $(y_{i,obs})$ бол бичлэг (i) –д ажиглагдсан хувьсагчуудын вектор утга юм.

Ерөнхийдөө (i) судалгаанд оролцогчийн хариулагдаагүй утгын регрессийн тэгшитгэлүүдийг дараах хэлбэрээр нэгтгэж илэрхийлж болно

$$y_{i,mis} = \alpha_{i,mis} + \beta_{m.o.(i)} y_{i,obs} + \varepsilon_{i,mis}$$

$(y_{i,mis})$ нь (i) – бичлэгийн хариулаагүй утгын вектор; $(\alpha_{i,mis})$ нь регрессийн тогтмол утгатай вектор; $(\beta_{m.o.(i)})$ нь (i) - бичлэгт үнэлэгдээгүй (q_i) хувьсагчдыг (i) - бичлэгт үнэлэгдсэн (p_i) (буюу Таамаглагч) хувьсагчдын регрессийн коэффициент бүхий $(q_i \times p_i)$ матриц бөгөөд $(\varepsilon_{i,mis})$ бол (q_i) регрессийн алдааны вектор юм.

Регрессийн коэффициентүүдийн матриц нь үнэлэгдээгүй хувьсагчууд бичлэг бүрд өөр өөр байх цорын ганц шалтгааны үед (i) -ээс хамааралтай байна. Ижил хувьсагч үнэлэгдээгүй бичлэгийн хувьд $(\beta_{m.o.(i)})$ матриц нь адилхан байна. Алдаа нь хоородоо хамааралтай байдаг гэдгээс бид алдаа нь 0 хүлээлтийн утга болон диагональ бус коварианцын матрицтай нормал буюу хэвийн тархалттай гэж бид үздэг.

$$\varepsilon_{i,mis} \sim N(0, \sum_{\varepsilon_{i,mis}})$$

Хэрэв үнэлэгдээгүй утга агуулаагүй бол олон талт регрессийн загварын параметруудийг хувьсах хэмжигдэхүүнтэй (univariate) регрессийн загваруудын орлуулах аргачлалын адил хамгийн бага квадратын аргаар олж авах боломжтой. Хэрэв үнэлэгдээгүй утга агуулсан бол бүх хувьсагч ажиглагдсан бичлэгээс параметруудийг орлуулж болно. Гэсэн хэдий ч, ялангуяа олон хувьсагч байгаа тохиолдолд бүрэн бичлэгийн тоо хязгаарлагдмал байж болох юм. Ийм тохиолдолд EM алгоритм гэж нэрлэгддэг аргыг ашиглан таамаглан тооцоолно. EM алгоритм нь бүрэн бус бичлэгт суурилсан параметруудийг таамаглах боломжтой давталтын процесс юм. Уг процессд бүх өгөгдлийг ашигладаг.

$(\alpha_{i,mis})$ болон $(\beta_{m.o.(i)})$ параметрын $(\alpha_{i,mis}), (b_{m.o.(i)})$ үнэлэгчийг ашиглан (i) бичлэн хариулагдаагүй утгыг дараах илэрхийллийн дагуу орлуулж болно.

$$y_{i,mis} = a_{i,mis} + b_{m.o.(i)}y_{i,obs}$$

Энэ нь вариаци эсвэл ковариациудыг бус харин дунджийг илэрхийлэхийн тулд алдааг авч үзээгүй орлуулга юм. Хэрэв бид орлуулгын дараа хувьсагчдын вариаци болон коварианцуудыг аль болох нарийвчлалтай хадгалахыг хүсвэл хүлээлт нь 0 ба коварианц $\Sigma_{\varepsilon_{i,mis}}$ -тай олон талт нормал тархалхалтаас сонгогдсон алдааны $\varepsilon_{i,mis}$ векторыг ашиглаж болно. EM алгоритм нь дээрх коварианц матрицын үнэлгээг мөн илэрхийлдэг.

8. ӨӨР ХУГАЦААНЫ ИЖИЛ МЭДЭЭЛЭЛД ҮНДЭСЛЭН ОРЛУУЛАХ АРГА

Ижил оролцогчдын хувьд ижил хувьсагчдыг ашиглан олон удаа хэмжилт хийх үед уг аргыг ашигладаг. Түүврээр сонгогдсон нэгжүүд нь дараагийн хугацааны туршид дахин сонгогдох нь уг аргын онцгой шинж юм. Энэ бүлэгт хугацаанд үндэслэсэн орлуулгын аргыг тодорхой түвшинд ашиглах боломжтой байдаг бүртгэлийн мэдээ гэх мэт өөр өөр хугацааны өгөгдөл ашиглан тайлбарлана. Хүн амын бүртгэл, шилжих хөдөлгөөний төдийгүй гэр бүлийн байдал гэх мэт цаг хугацаанд өөрчлөгдөх мэдээллийг жил тутамд шинэчилдэг. Ихэнх тохиолдолд ажил, мэргэжилтэй холбоотой мэдээлэл болон орлого, тэтгэмж гэх мэт өөр өөр цаг үеийн бүртгэлийн мэдээллийг холбож хугацааны давтамжит мэдээллийг бүрдүүлдэг. Ялангуяа нийгмийн статистикийн мэдээллээс хугацааны давтамжит өгөгдлүүдийг нэгтгэж болох хэдий ч эдгээр нь ихэвчлэн хугацаанаас хамаарсан тооцооллыг шаардах үе байдаг.

Хугацааны давтамжид үндэслэсэн орлуулгын арга нь энэхүү гарын авлагад тусгагдсан бусад аргуудаас ялгагдах онцлог нь ялгаатай объектуудын өгөгдлийг ашиглах бус харин ижил оролцогчийн өөр өөр хугацааны өгөгдөлд орлуулгын аргыг ашигладаг. Оролцогч бүрийн хувьд үл мэдэгдэх утга байх үед тооцоолол хийх замаар утга орлуулдаг.

Цаг хугацааны давтамжид үндэслэсэн өгөгдөлд үл мэдэгдэх утга дараах 2 хэлбэртэй илэрнэ.

1. Үл мэдэгдэх утга тархсан үе: Нэг болон хэд хэдэн хугацааны үед объектууд нь ажиглагдаагүй, эсвэл объектуудаас ямар ч хувьсагч ажиглагдаагүй үед
2. Хугацааны давтамжит байдал илрээгүй (панел шинж завсардсан үе): тодорхой цаг үед ажиглалт хийгдээгүй буюу тухайн үеийн мэдээлэл байхгүй үед

Нас баралт болон шилжилт хөдөлгөөн нь үл мэдэгдэх утгыг бий болгодоггүй гэдгийг цохон тэмдэглэх нь зүйтэй. Эдгээр нь нэмэлт тооцоолол шаардах зорилтот бүлгийн

хэсэг биш, учир орлуулах тооцоолол хийх ёсгүй. Лепковски (1989) нь урт хугацааны мэдээлэл дэх хоосон утгыг олон талаас нь тайлбарлахыг санал болгодог.

Хэрэв урт хугацааны мэдээлэлд үл мэдэгдэх утга байх тохиолдолд хугацаанд үндэслэсэн орлуулгын аргыг ашигладаг. y_{it} -г i -р объектын t хугацааны ухувьсагчийг үл мэдэгдэх гэе. Үүнийг дараагаар y хувьсагчийн утгыг i -р объектын өмнөх болон дараагийн хугацаанаас хамааруулан \tilde{y}_{it} утгыг тооцоолон гаргаж болно. Ихэнхдээ y -ийн мэдээлэл нь хугацааны өмнөх үеүд буюу $y_{it-1}, y_{it-2}, \dots$ утгуудаар хязгаарлагддаг. Үүнийг өмнө нь дурдсан үл мэдэгдэх утга тарсан үе болон панел шинж завсардсан үеүдэд ашиглаж болно. Хэрэв хүлээгдэж үр дүнг тодорхойлох эсвэл ижил хугацааны мэдээллийг хэд хэдэн удаа орлуулан тооцоолох тохиолдолд энэхүү хожуу үеийн мэдээлэл нь ашиглагдах бөгөөд энэ нь хамгийн тохиромжтой хувилбар байх боломжтой.

Урт хугацааны өгөгдөлд үндэслэсэн орлуулгын арга, техникийг ашиглах үндсэн 2 шалтгаан байдаг талаар өмнөх бүлэгт авч үзсэн.

1. Нэгдүгээрт: ижил объектуудын өмнөх үеийн болон хожуу үеийн ажиглалтын утгууд нь үл мэдэгдэх утгыг сайн таамаглагч байдаг. Энэ нь орлуулгын арга, тооцооллын чанарыг сайжруулсан гэсэн үг юм. Энэ шинжийг хадгалахын тулд туслах хувьсагчаар ашиглагдах өмнөх болон ирээдүйн ажиглалтыг утгуудыг бодитоор ашиглаж болно.
2. Хоёрдугаарт: зөвхөн кросс-секшон (жишээ нь тодорхой цаг хугацаанд хамт амьдарсан хүмүүсийн тоо) өгөгдлийн хувьд бус урт хугацааны өгөгдлийг анхаардаг боловч цаг хугацааны явц дахь өөрчлөлт (жишээ нь хамт амьдарч эхэлсэн хүмүүсийн тоо) нь илүү сонирхолтой.

Панел шинж завсардсан үед ихэвчлэн тухайн түүврийн жинг ашигладаг. Хэрэв бид эх олонлогийн параметрийг тодорхойлохыг хүсвэл, хариултын утга хоосон буюу сүүлийн завсардалтын өмнөх үеийг ашиглах боломжтой. Нас барсан болон шилжсэн тохиолдлуудад бүртгэлийн шижтэй панел өгөгдлийн завсардах байдал үүсэх үндэслэлтэй.

Урт хугацааны өгөгдлийн үл мэдэгдэх утгыг орлуулах олон төрлийн арга байдаг. Жишээлбэл, Van der Laan and Kuijvenhoven (2008) -д урт хугацааны өгөгдөлд анализ хийх хэд хэдэн аргын талаар авч үзсэн байдаг боловч эдгээр нь үл мэдэгдэх утгыг орлуулахад төдийлөн ашиглагддаггүй. Шинжилгээний зорилгоос хамааран зарим тохиолдолд тооцоо хийхгүй байхыг илүүд үздэг.

Хугацаанд үндэслэсэн орлуулгын арга нь хэд хэдэн бүрдэл хэсгүүдээс бүрддэг. Тиймээс уг аргатай хослуулан ашигладаг цөөн хэдэн аргын шинж чанарыг энд авч үзье.

Өөр хоорондоо ялгаатай аргууд нь өөрсдийн онцлог шинж чанаруудтай.

- Бусад объектуудын мэдээллийг ашиглах боломжтой. Ашиглагдаж буй арга тус бүр орлуулгын тооцоололд өнгөрсөн болон ирээдүйн ажиглалтыг утгуудын

мэдээллийг ашигладаг. Томоохон мэдээллийн санд үндэслэсэн орлуулгын тооцоолол хийх үед ашиглахад энгийн бөгөөд хялбар байдаг нь үүний давуу тал юм. Харин эсрэгээрээ бусад объектуудаас нэмэлт мэдээлэл авалгүйгээр мэдээллийн алдагдал үүсгэх тохиолдолд энэ нь сул талыг бий болгоно. Жишээлбэл, орлогын талаарх нэмэлт мэдээллийг өмнөх үеэс авснаараа дундаж өсөлтийг засварлах боломжтой болно. Энэ мэтээр өмнөх үеийн мэдээллийг ашиглах нь илүү сайн тооцооллыг хийх боломжийг бүрдүүлдэг.

- Тасралтгүй болон ангилагдсан өгөгдөлд ашиглахад тохиромжтой. Бүх аргыг тасралтгүй шинжтэй өгөгдөлд ашиглах боломжтой боловч тэдгээрээс тодорхой хэдэн аргыг ангилагдсан өгөгдөлд ашигладаг.
- Цаг хугацааны өгөгдлийн нэг объект дахь ухувьсагчид олон үл мэдэгдэх утга байх нь үүний тухайн тохиолдол байдаг. Зарим арга нь ижил хугацааны үл мэдэгдэх утгуудыг тооцоолдог бөгөөд энэ ажиглалтын утга хоорондын хамаарлыг өөр өөр хугацаанд хадгалах боломжтой. Бусад аргууд нь тухайн үеийн хоосон утгыг тооцдог талтай. Хэрэв олон тооны үл мэдэгдэх утга байх тохиолдолд эдгээр аргуудыг хэд хэдэн удаа ашиглах боломжтой хэдий ч ажиглалтын утга хоорондын хамаарлыг өөр өөр хугацаанд төгс хадгална гэсэн баталгааг өгдөггүй.

Хүснэгтэд дурдсан шинж чанаруудыг арга тус бүрийн хувьд харуулав. Эдгээр аргуудыг дараа дараагийн хэсгүүдэд илүү нарийвчлан тайлбарласан болно.

Орлуулах аргуудын шинж чанар.

| Аргачлал | Тасралтгүй чанар | Ангилагдсан чанар | Бусад объектоос мэдээлэл авч ашигладаг эсэх | Олон хүчин зүйлийн нөлөөтэй эсэх |
|---|------------------|-------------------|---|----------------------------------|
| Интрополяци | + | - | - | - |
| Өмнөх ажиглалтаас сүүлийн ажиглалт хамаарах | + | + | - | - |
| Харьцаагаар орлуулах | + | - | + | - |
| Регрессийн орлуулга | + | + | + | + |
| Колд-дек орлуулга | + | + | - | +/- |
| Хот-дек орлуулга | + | + | + | + |
| Литтл болон Су орлуулга | + | - | + | + |

8.1 Интерполяци

Интрополяци гэдэг нь үл мэдэгдэх утгуудыг өмнөх үеийн болон ирээдүйн утгаас хамааруулан үнэлэх арга юм. Энэ үед туслах хувьсагчдын эсвэл тус тусын ялгаатай мэдээллийг ашиглахгүй. i, \tilde{y}_{it} нь дараах тэгшитгэлээр тодорхойлогдоно

$$\tilde{y}_{it} = f(y_{it-1}, y_{it-2}, \dots, y_{it-K}, y_{it+1}, y_{it+2}, \dots, y_{it+L}) \quad (8.1)$$

Энд: K нь өнгөрсөн үеийн ажиглалтын утгууд, L нь ирээдүйн утгаар ашиглагдана.

Тооцоолол хийх үед бусад объектууд ямар нэг мэдээллээр хангахгүй бөгөөд загварт таамаглал дэвшүүлэхэд хэцүү тохиолдолд тоон хувьсагчийн тусламжтайгаар ашиглаж болно. Бусад объектууд тооцоололд шаардлагатай мэдээллийг агуулдаг

байх үед энэ аргыг (регрессээр орлуулах, харьцаагаар орлуулах болон Little and Su аргачлал) ашиглах нь санал болгож байна.

Тоон y хувьсагчийн ерөнхий интерполяцийн томьёо \tilde{y}_t -ийн ажиглалтын $y_{t-k}, \dots, y_{t-1}, y_{t+1}, \dots, y_{t+l}$ утгуудын хувьд дараах хэлбэртэй байна

$$\tilde{y}_t = \frac{\sum_{k=1}^K w_{-k} y_{t-k} + \sum_{l=1}^L w_l y_{t+l}}{\sum_{k=1}^K w_{-k} + \sum_{l=1}^L w_l} \quad (8.2)$$

$w_{-1} \geq w_{-2} \geq \dots \geq w_{-k}$ болон $w_1 \geq w_2 \geq \dots \geq w_L$ нь жин байх бөгөөд энэ 2 үеүүдэд y_T нь t -ээс T хүртэлх хугацаанаас хамаарсан байна. Жинг чөлөөтэй сонгох боломжтой. Жишээ нь: $w_k = w_{-k} = \frac{1}{k}$ байж болно.

Хэрэв олон тооны хоосон утгатай байх тохиолдолд тэгшитгэл 8.2-ыг ашиглаж болно. Жишээ нь: y_{t+k} нь үл мэдэгдэх бөгөөд \tilde{y}_t -ийг тодорхойлох шаардлагатай үед $w_k = 0$ гэе. Тэгшитгэлийн дагуу өнгөрсөн үеийн мэдээлэл ($w_1 = w_2 = \dots = w_L = 0$) бол энэ үед панел алдагдал үүснэ.

Тэгшитгэл 8.2-ын онцгой тохиолдол:

1. Өмнөх болон дараагийн ажиглалт хоорондох шугаман интерполяци.

Хэрэв y_{t-1} болон y_{t+1} нь хоёулаа боломжит утгатай гэвэл тэгшитгэл 8.2 дараах байдлаар хувирна.

$$\tilde{y}_t = \frac{w_1(y_{t-1} + y_{t+1})}{2w_1} = \frac{y_{t-1} + y_{t+1}}{2} \quad (8.3)$$

Хэрэв $w_{-1} = w_1$ эсвэл y_{t-1} болон y_{t+1} нь хоосон бол y_{t-k} болон y_{t+l} ажиглалтын утгуудаар t хугацаатай хамгийн ойр байгааг сонгох бөгөөд жин нь тус бүр $w_{-k} = \frac{1}{k}$, $w_l = \frac{1}{l}$ байна. Энэ үед тэгшитгэл 8.2 дараах хэлбэртэй болно.

$$\tilde{y}_t = \frac{l y_{t-k} + k y_{t+l}}{k+l} \quad (8.4)$$

Жишээ нь: y_t болон y_{t+1} нь мэдэгдэхгүй байг, $y_{t-1} = 3$ болон $y_{t+2} = 4$ гэе. Тэгвэл энэ тохиолдолд тэгшитгэл 8.4-ийн дагуу $\tilde{y}_t = \left(\frac{(2*3) + (1*4)}{(1+2)} \right) = \frac{10}{3} = 3.333$ гарахаар байна. Интерполяцийн аргаар y_{t+1} -ийг олвол: $\tilde{y}_{t+1} = \left(\frac{(1*3) + (2*4)}{(2+1)} \right) = \frac{11}{3} = 3.667$ байна. \tilde{y}_{t+1} -ийн хувьд өмнө нь орлуулан тооцоологдсон \tilde{y}_t -ээс мэдэгдэж буй утгыг авна. Ингэснээр: $\tilde{y}_{t+1} = \left(\frac{(1*\tilde{y}_t) + (1*y_{t+2})}{(1+1)} \right) = \frac{\frac{10}{3} + 4}{2} = 3.667$ болно.

2. Өмнөх ба дараагийн ажиглалтын утгын дундаж.

Өмнөх үеийн болон дараагийн үеийн ажиглалтын утгуудаас эсвэл 8.2 тэгшитгэлийн дагуу жинлэгдээгүй дундаж утгыг бид тодорхойлж чадна. Шугаман интерполяцийн онцгой тохиолдол нь $p = 1$, $w_{-k} = \frac{1}{k}$ болон $w_l = \frac{1}{l}$ байна.

3. Шугаман тренд (y болон T хоорондын регресс)

Регрессийн үнэлгээ нь y хувьсагчийг агуулсан $y = \alpha + \beta T + \varepsilon$ тэгшитгэлээр илэрхийлэгдэнэ. Энэ үед $T=t$ байвал регрессийн үнэлгээд орлуулга хийгдэж дараах хэлбэртэй болно.

$$\tilde{y}_t = \hat{y}_t = a + bt \quad (8.5)$$

Энд a болон b нь хамгийн бага квадратын аргын үнэлэгчид бөгөөд тэгшитгэл 5.3 үйлчилнэ. Регрессийн үнэлгээний зарчмаар \hat{y}_t нь y_t ажиглалтын утгуудаар үнэлэгдсэн утга юм. Хэрэв туслах мэдээллийг өмнөх болон дараагийн үеийн мэдээлэлд үндэслэвэл шугаман хандлага нь шугаман интерполяци болж өөрчлөгдөнө. Энэ үед ажиглалтын утгуудыг жинлэхгүй.

Тэгшитгэл 8.5-ын параметруудийг бусад функцүүдээр эсвэл шугаман бус регрессийн аргаар үнэлэх боломжтой. Харин тэгшитгэл 8.2-т орлуулгын тооцоолол хийх шаардлагагүй.

Дээр дурдсан гурван аргуудаас энгийн шугаман интерполяцийг ерөнхийд нь илүүд үздэг. Хэрэв өгөгдөл тухайн процессын үнэн илэрхийлсэн бол боломжтой гэж үзнэ. Өмнөх болон дараагийн үеийн ажиглалтын утгуудаас бүхий л мэдээллийг авах боломжтой. Хэрэв энэ төрлийн өгөгдөлд тооцооллын алдаа гаргавал бусад аргачлалыг чухалчлах шаардлагатай болно.

SPSS программ нь цаг хугацаанаас хамаарсан хоосон утгыг орлуулах RMV функцтэй. Тэрхүү модуль нь дараах аргачлалуудыг агуулсан байдаг бөгөөд ямар томъёоноос гаргаж болохыг хаалтад оруулав.

- Шугаман интерполяци (тэгшитгэл 8.4)
- Өмнөх ба дараагийн ажиглалтын утгын дундаж (дэд хэсэг 2)
- Өмнөх ба дараагийн ажиглалтын утгын медиан (дэд хэсэг 2-ын хувилбар)
- Хугацааны дундаж буюу цаг хугацааны өгөгдөл дэх бүх утгуудын дундаж (тэгшитгэл 8.2-ын онцгой тохиолдол)
- Шугаман хандлага (дэд хэсэг 3)

Шинж чанарууд

- Интерполяцыг их хэмжээний өгөгдлийн санд ашиглахад хялбар байдаг. Учир нь интерполяци нь зөвхөн нэг объектын мэдээллийг ашигладаг. Тиймээс объектыг нэг нэгээр нь боловсруулж болно гэсэн үг.
- Бусад объектуудын ашиглаж байгаа ямар ч мэдээллийг ашиглаагүй тул энэ арга нь бусад объектын мэдээллийг ашигладаг аргуудаас бага нарийвчлалтай тооцоог гаргаж магадгүй юм.
- Орлуулгын арга зүйд ямар ч хазайлтыг ашигладаггүй. Цуваа нь хэтэрхий төгс байдаг. Өөр хугацаануудын хоорондын корреляцийн ач холбогдол өндрөөр тооцогддог. Ингэснээр хазайлт нэмэгдэхээс сэргийлж байдаг.

8.2 Сүүлийн ажиглалтын утгыг урагшлуулах болон хойшлуулах

Сүүлийн ажиглалтын утгыг урагшлуулах (CAУУ) арга нь практикт ихэвчлэн ашигладаг. Энэ арга нь ямар ч асуудалгүй, ашиглахад маш хялбар учраас өргөн хэрэглэгддэг. Тус арга нь сүүлийн ажиглалтын хувийн утгыг бүх дараагийн үеийн утгуудад орлуулж хэрэглэдэг. Энэ аргын вариацийг илүү дэлгэрэнгүй тайлбарлах болно.

Энэ арга нь голдуу ангилагдсан хувьсагчуудад хэрэглэгддэг бөгөөд тэдгээрийн хугацаа өнгөрөх тусам маш бага эсвэл огт өөрчлөгддөггүй нь мэдэгддэг байх ёстой. Ийм жишээ хувьсагч гэвэл жендер юм. Бусад категорийн болон тоон хувьсагчийн хувьд энэ арга нь ихэвчлэн бодит нөхцөл байдлын талаар тогтвортой зураглалыг алдаатай гаргадаг. Тухайлбал, тоон үзүүлэлтийн графикийн хувьд энэ арга нь бодит бус тогтвортой үнийн ажиглалтын утгатай болдог.

Сүүлийн ажиглалтын утгыг урагшлуулах аргад сүүлийн ажиглалтын утгын y_{it-1} -г үл мэдэгдэх утга y_{it} -д орлуулахад ашиглана. Өөр вариант бол сүүлийн ажиглалтын утгыг хойшлуулах буюу дараагийн ажиглалтын утга y_{it+1} -ийг y_{it} -ийн утгад сольж орлуулна. Сүүлийн ажиглалтын утгыг урагшлуулах аргын хувьд энэ утгыг олон тооны дараалсан үл мэдэгдэх утгуудад ашиглах боломжтой.

Санамсаргүй шилжүүлдэггүй (Виллиам болон Байлей, 1996), үл мэдэгдэх завсрын утга y_{it} -г y_{it-1} эсвэл y_{it+1} -ээр орлуулдаг. Энэ нь тохиолдлоор хэрэв хоёр буюу түүнээс дээш дараалсан үеийн туршид утга байхгүй бол энэ аргыг ашиглах боломжгүй гэсэн үг. Үүнээс гадна энэ арга нь хэрэв эхний ажиглалтын утга эсвэл сүүлийн ажиглалтын утга үл мэдэгдэх бол хэрэглэх боломжгүй гэсэн үг юм. Энэ нөхцөлд бусад орлуулгын аргыг ашиглах ёстой.

Сүүлийн ажиглалтын утгыг урагшлуулах аргын асуудал бол сүүлчийн утга нь цаг хугацааны явцад өөрчлөгдөхгүй гэж таамаглах нь ихэнхдээ бодитой бус байдаг. Ийм таамаглалыг заавал судлах ёстой. Энгийн үед санамсаргүй хэлбэлзэл(эсвэл тооцооллын алдаа)-ээс шалтгаалан бие даасан өгөгдөл цаг хугацааны явцад зарим хазайлттай байдаг. Сүүлийн ажиглалтын утгыг урагшлуулах аргад энэ хазайлт нь хүлээн зөвшөөрөгддөггүй. Тийм учраас орлуулгын тодорхой бус байдлыг зохих хэмжээгээр тооцоогүй буюу энэ нь буруу статистикийн үр дүнд хүргэдэг. Энгийн шийдэх арга бол хазайлтыг нэмэх юм. Үүнтэй адилаар сүүлийн ажиглалтын утгыг хойшлуулах аргын хувьд хэрэглэхдээ тогтвортой хугацаан цуваа нь бодит байдлаас хазайсан эсэхээс үл хамааран судлах шаардлагатай.

Харьцааны орлуулга

Энэ аргыг 4-р бүлэгт үзсэн дээр нэмэлтээр энэ арга нь байнга урт хугацааны өгөгдлийн үед ашигладаг буюу t үе дэх ажиглалтын утга нь $t-1$ үеийн ажиглалтын утгатай пропорциональ байна гэж үзэх нь ихэвчлэн үндэслэлтэй таамаглал байдаг. Энэ арга сүүлийн ажиглалтын утгыг урагшлуулж сайжруулалт хийсэн гэж үзэж болох ба тодорхой хугацааны туршид ерөнхий өөрчлөлтийг бас залруулах болно. Цаг хугацаа бүрд өөр өөр хазайлт бүрийг сонгож болно гэдгийг тэмдэглэх нь зүйтэй. Ялгаатай хазайлтын үе бүрд үргэлж сонгогдож байх юм. Энэ аргыг үргэлжлүүлж судлахыг хүсвэл 4-р бүлгээс харна уу. Энэхүү орлуулгын хэлбэрийг ихэвчлэн эдийн засгийн статистикт ашигладаг.

Регрессийн орлуулга

Регрессийн орлуулгыг 5-р бүлэгт бид үзсэн. Энэ бүлэгт тайлбарласнаар урт хугацааны нөхцөлийн үед үнэн байдаг. Энэ хэсэгт бид өгөгдлийн шинж чанар нь урт хугацааны байх асуудлыг авч үзнэ. Урт хугацааны өгөгдөл нь олон талт байдаг учраас шинжилгээ нь ихэвчлэн илүү төвөгтэй байдаг. Гэсэн хэдий ч урт хугацааны өгөгдлийн давуу тал нь хувьсагчийн өнгөрсөн ба/эсвэл ирээдүйн ажиглагдсан утгууд нь байхгүй утгуудыг маш сайн таамагладаг.

5-р бүлэгт x_j хувьсагчийн тоог ашиглан ганц y хувьсагчийн утгыг таамаглахыг оролдсон билээ. Энэ нөхцөлд бид y хувьсагчийг сонирхож байгаа юм. Урт хугацааны өгөгдөлтэй нөхцөлд бидэнд y_{ij} гэсэн олон ажиглалтын утга байгаа ба i бүрийн хувьд бие даасан байх ба энд t нь 1-ээс M хүртэл гүйнэ. Олон хэмжээст y_{ij} -д ганц нэг утга нь орхигдож болно. Урт хугацааны өгөгдлийн шинжилгээнд ерөнхийдөө өөр хугацааны үеийн ажиглалтын утгуудын хоорондын корреляцийг сонирхдог. Жишээлбэл, Бид судалгааг өөрчилмөөр байна. Тиймээс орлуулгын ажиглалтын хоорондын корреляцийг авч үлдэх нь чухал юм. Энэ нь орлуулга нь олон хэмжээст гэсэн үг. Олон хэмжээст орлуулгыг 7-р бүлэгт үзсэн. Хэрэв олон хэмжээст биш бол y_{it} тус бүрт салангид нэг хэмжээст загварууд суулгаж өгнө. Энд y_{it} нь x_{ij} ковариацийн болон y_{it} -ийн хэтийн ажиглалтын утгаас хамаарна.

$$E[y_{it}] = f(x_{i1}, \dots, x_{ip}, y_{it-1}, y_{it-2}, \dots, y_{it+1}, y_{it+2}, \dots) \quad (8.6)$$

Тухайн загвар нь байхгүй ажиглалтын утга бүрд үүсгэх ёстой ба Олон тооны байхгүй ажиглалтын утга болон ковариатуудын нөхцөлд загварыг салгаж байгуулах ёстой. Энэ нь маш төвөгтэй бөгөөд ажиглалтын утгуудын хоорондын корреляцийг хэвээр үлдээхэд маш хэцүү байдаг.

Өөр нэг сонголт бол олон хэмжээст загварыг ашиглах юм. Энд ганц загварыг суулгаснаар бүхий л ажиглалтын утгыг тодорхойлно. Бие даасан i -ийн ялгаатай ажиглалтын утгуудыг вектор y_i гэж бичиж, тухайн загвар нь энэ векторыг тодорхойлохоор үүсдэг. Тухайлбал, шугаман загвар

$$y_i = X_i \beta + \varepsilon_i \quad (8.7)$$

Энд вектор ε_i нь олон хэмжээст нормал тархалттай байна. Урт хугацааны өгөгдлийн нөхцөлд өөр өөр ажиглалтын хоорондох хамаарлыг загварчлах нь чухал юм (жишээ нь тодорхой ажиглалт хийхэд өндөр орлоготой хэн нэгэн дараагийн ажиглалтад өндөр орлоготой байх магадлалтай).

Хэрэглээний тухай

- Регрессийн орлуулгыг тоон болон категорийн хувьсагчуудад хоёуланд ашиглаж болдог. Хоёрдугаарт, олон хэмжээст шугаман регресс, логистик регрессээр ашиглалт хийх боломжгүй юм. Жишээлбэл, орлуулж ашиглах хэрэгтэй.
- Дээр дурдсан олон талт регрессийн загварууд нь янз бүрийн бие даасан ажиглалтын утгын хугацааг ихэвчлэн зохицуулж байдаг. Энэ бүлэгт дурдсан

ихэнх бусад аргууд нь бүх бие даасан утгуудыг тогтмол ажиглалтын утгуудын хугацаанд таамаглана.(жишээлбэл жил бүр эсвэл улирал бүр)

Шинж чанар

Урт хугацааны өгөгдлийн шинжилгээнд бид ерөнхийд нь хугацааны өөрчлөлтийг судалдаг. Орлуулгад хазайлтыг ашиглах эсэхээс үл хамаарч шийдвэр гаргах юм. Хэрэв урт хугацааны өгөгдлийн нөхцөлд хазайлтыг ашиглаагүй бол өөрчлөлтийн ач холбогдол хүчтэйгээр үнэлэгдэх юм.

Колд дек

Колд дек орлуулгын аргын тухай бүлэг 6-д үзсэн билээ. Энэхүү арга нь баталгаажсан арга биш юм. Энэ аргын талаар бид цааш нь үзэхгүй. Бид ажиглалтын утгыг урагшлуулах/хойшлуулах аргад колд-дек аргыг тооцож үзэхгүй байгааг анхаарах хэрэгтэй. Колд-дек орлуулгын үед гаднын эх сурвалжаас авсан мэдээлэл ашиглан хэрэглэдэг. Сүүлийн ажиглалтын утгыг урагшлуулах/хойшлуулах аргад ч гэсэн эртний эсвэл дараагийн цаг хугацааны мэдээлэл тус бүр ашиглаж хэрэглэдэг. Энэ мэдээллийг гадаад эх сурвалж гэж үзэхгүй.

Хот-дек

Хот-дек орлуулгын аргыг 6-р бүлэгт үзсэн. Хэрэв судалгаанд оролцогчийн олон утгууд үл мэдэгдэх байвал хот-дек орлуулгын аргыг ашигладаг гэж өмнө нь авч үзсэн. Энэ нь урт хугацааны өгөгдөлд хот-дек орлуулгыг ашиглахад тохирсон арга болж байгаа юм. Хот-дек аргад ганц бие даасан утгыг олон утгын орлуулгад ашиглаж болно. Дүрмээр бол орлуулгын хоорондох уялдаа холбоог хангах зорилгоор нэг донорыг орлуулдаг.

8.3 Литтл болон Су арга

Литтл болон Су арга нь орлуулгад цаг хугацааны хувьсах түвшин болон тооцооллын дундаж хандлагыг хоёуланг нь агуулдаг. Дараах загвар нь энэ сэдэвт ашиглагдана:

$$\text{Орлуулга} = \text{мөрийн нөлөөлөл} * \text{баганын нөлөөлөл} * \text{үлдэгдэл} \quad (8.8)$$

Баганын нөлөөлөл нь хугацааны дундаж өөрчлөлтийг тодорхойлдог ба хугацааны нөлөөлөл гэнэ. Мөрийн нөлөөл нь хугацааны нөлөөллийн хувьд залруулсан хувьсах түвшнийг тодорхойлдог. Литтл болон Су аргад үлдэгдлийг өөр хувьсагчаар авдаг буюу мөрийн нөлөөний хувьд энэ нь орлуулсан хувьсагчтай маш адилхан юм. Мөрийн нөлөөлөл болон үлдэгдлийн хувьсагч нь адилхан гэсэн таамаглал байдаг. Хэрэв хувьсагч нь олон хамааралтай утгуудаар дутаж байвал (тухайлбал, бохир болон цэвэр цалин;) эдгээр утгуудыг бүгдийг нь нэгэн зэрэг орлуулдаг ба ганц л донорыг үлдэгдэлд ашигладаг.

Литтл болон Су аргыг у эерэг тоон хувьсагчийн үл мэдэгдэх утгуудад ашиглаж болдог буюу хувьсагчийн нөлөөг хугацааны нөлөөнд үржүүлсэн болгоныг загварчлах боломжтой бөгөөд стохастик орлуулга хийдэг. Энэ аргыг ашиглахад хялбар бөгөөд байхгүй өгөгдлийн янз бүрийн хэлбэр, түүнчлэн хувьсагчид ногдох олон тооны утгыг ашиглах боломжтой.

Энэ арга нь хувьсагчийн хувьд ажиглагдсан утгууд бүгд тэгтэй тэнцэх асуудалтай тулгардаг. Эдгээр хувьсагчуудыг Литтл болон Су аргыг ашиглан орлуулж болдоггүй.

Баганын нөлөөлөл C_t нь хугацааны явцад объектын өөрчлөлтийн дунджийг харуулах ба ингэж тооцно.

$$c_t = \frac{\bar{y}^t}{\frac{1}{M} \sum_{t=1}^M \bar{y}^t}$$

Энд \bar{y}^t нь t хугацааны y_i^t утгын дундаж, M бол хугацааны тоо. Мөрийн нөлөөллийг i хувьсагчтай r_i -г доор томъёолвол

$$r_i = \frac{1}{m_i} \sum_t \frac{y_i^t}{c_t}$$

Энд i хувьсагчийн хувьд m_i -г мэдэгдэж байгаа y_i^t -ийн нийлбэрээр тооцов.

Үлдэгдлийг өөр нэг j хувьсагчаас гарган авч, хувьсагч i -ийн хувьд хугацааны алдагдал нь ажиглагдана. Хувьсагч j -ийг эхлээд бүх нөлөөнд үндэслэн бүх хувьсагчийг эрэмбэлж, дараа нь мөрийн нөлөө i -тэй хамгийн ойр байх хувьсагчийг сонгоно. J хувьсагчийн үлдэгдлийг томъёолбол

$$e_j^t = \frac{y_j^t}{r_j c_t}$$

(8.8) тэгшитгэлтэй нэгтгэвэл дараах томъёог гаргаж авна.

$$\tilde{y}_i^t = r_i c_t e_j^t = r_i c_t \frac{y_j^t}{r_j c_t} = \frac{r_i}{r_j} y_j^t$$

Хийсвэр нөхцөлд донор (үлдэгдлийн) нь хүлээн авагчтай олон төрлийн ижил шинж чанартай байдаг. Дээр дурдсан стандарт арга нь мөрийн нөлөөлөл бүрийг ашиглан донор болон хүлээн авагчийг хооронд нь тохируулах замаар гүйцэтгэдэг. Гэхдээ стратад стандарт аргыг хэрэглэснээр энэ аргыг өргүүлэх боломжтой юм. Үүний үр дүнд баганын нөлөөллүүд страта хооронд ялгаатай байх боломжтой, тиймээс цаг хугацааны дундаж өсөлт страта хооронд ялгаатай байж болно.