22-дәріс

ҚАЗАҚСТАННЫҢ
АШЫҚ
УНИВЕРСИТЕТІ

СТАТИСТИКА НЕГІЗДЕРІ

Шашыранды диаграмма және регрессиялық анализ

Каузалдық айнымалылардың бір-біріне қалай әсер ететіні туралы күнделікті өмірде де, сондай-ақ ғылымда да маңызды мәселе. Мысалға, сіз жаңалықтың комментаторы экономиканың төмендеуі қылмыстың жоғары деңгейіне әкелгенін немесе жанармай бағасының жоғарылауы адамдардың көлік жүргізу дағдыларының өзгеруіне себеп болды естіген боларсыз. Бір айнымалының басқасының себебі болатыны туралы дәлелдердің сенімділігін қалай бағалай аламыз? Бұл сауаққа жауапты біз бүгінгі тақырыбымыз аясында іздейміз. Ендеше бастайық

Тақырыпты меңгергеннен кейін:

1. Пирсонның корреляциялық коэффициентін есептей аласыз
2. Детерминация коэффициентін интерпретациялай аласыз
3. Корреляциялық матрицаны талдай аласыз
4. Корреляция, каузалдық, шартты айнымалы ұғымдарымен жұмыс жасай аласыз.

Пирсонның корреляциясы X және Y арасындағы жалпы сызықтық байланысты өлшейтін статистика. Ең кіші-квадраттар регрессиялық қисығының көлбеуі (b) X-ң Y-ға әсерінің өлшемі, сәйкесінше X артқан сайын ол да өсіп отырады. Дегенмен 0 мен 1,00 арасында b өзгермейді, сондықтан да оны байланыс өлшемі ретінде пайдалану орынды емес. Оның орнына зерттеушілер интервалдық-пропорционалды айнымалылыар арасындағы байланысты өлшеуде Пирсонның r-і немесе корреляциялық коэффициент деп аталатын статистиканы кең түрде (әрдайым тек соны) пайдаланады. Пирсонның r-і 0,00 мен 1,00 арасында өзгеріп отырады, мұндағы 0,00 ешқандай байланыс жоқ екенін, ал +1,00 мен -1,00, сәйкесінше, идеалды оң және идеалды теріс байланыс бар екенін көрсетеді. Пирсонның r-ң формуласы келесідей

12.4-формула

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\left[\sum(X-\bar{X})^2\right]\left[\sum(Y-\bar{Y})^2\right]}}$$

Бұл формуланың алымы X пен Y-ң ковариациясы екенін ескеріңіз.

Пирсонның r-н есептеу үшін

1. Көлбеуді (b) есептеуде қолданған кестеге бір баған қосыңыз. (Y-Y) -ң мәнін квадраттап, нәтижесін 7-бағанға жазыңыз.
2. 7-бағандағы жалпы нәтижені жазыңыз.
3. 6-бағандағы нәтижені 7-бағандағы нәтижеге көбейтіңіз.
4. 3-қадамда тапқан мәннің түбірін табыңыз.
5. 4-қадамда тапқан санды 5-бағандағы санға немесе тура көбейтіндісіне бөліңіз. Нәтиже Пирсонның r-і болады.

Пирсонның r-і - екі айнымалы арасындағы сызықтық байланыстың мықтылығының индексі. 0,00 мәні ешқандай байланыстың жоқ екенін, ал 1,00 идеалды сызықтық байланысты көрсетсе, осы аралықтар арасындағы мәндер ешқандай тікелей интерпретацияға ие емес. Әрине, біз байланыстардың аралықтарған қаншалықты жақын болатыны тұрғысынан сипаттай аламыз (мысалы, 0,00-ге жақын коэффициент "әлсіз", ал 1,00-ге жақыны "мықты" деп сиппауға болады), бірақ бұл сипаттама белгілі бір деңгейде субьективті болады. Сонымен қатар, реттік-деңгейдегі өлшемдер байланысының мықтылығын әлсіз. Әрине, бұл белгілер тек нұсқаулықтар ғана, сондықтан барлық ықтимал зерттеу жағдайларына қатысты тиімді және пайдалы болмайды.

Бақытымызға орай, біз детерминация коэффициенті деп аталатын қосымша есептеу статистикасын қамтитын r-ң барынша тікелей интерпретациясын пайдалана аламыз. Бұл статистиканы, қарапайым түрде Пирсонның r-ң квадраты (r2), қатедегі пропорционалды қысқаруға ұқсас логиканы пайдалана отырып, интерпретациялауға болады. Есіңізде болса, байланыстың өлшемдерінің логикасы - екі түрлі жағдайдағы тәуелді айнымалының мәнін болжау. Бірінші, X арқылы алынып отырған ақпаратты елемеу барысында Y болжанады; екіншісі, тәуілсіз айнымалы есепке алынады. r2 жағдайында қорытынды статистиканы болжау және құрастыру әдісі кейбір жаңа ұғымдарды таныстыруды талап етеді.

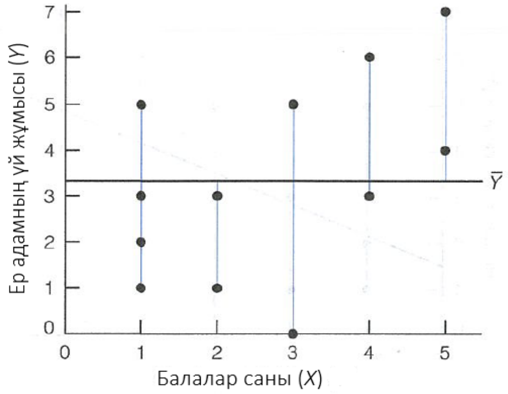
Интервалдық-пропорционалды деңгейде өлшенетін айнымалылармен жұмыс жасағанда бірінші шарт аясында Y көрсеткіштерінің болжамдары Y көрсеткіштерінің арифметикалық орташасы болады.

Бұл болжау стратегиясы X бойынша еш ақпаратсыз тиімді болады, өйткені біз кез-келген үлестірімнің арифметикалық орташасы кез-келген басқа нүктелерге қарағанда үлестірімдегі барлық нүктелерге жақын болатынын білеміз. Өткен дәрістердегі берілген минималды вариация принциптері мен оның шығару жолын еске түсірейік

Кез-келген айнымалының көрсеткіштері басқа нүктелерге қарағанда арифметикалық орташа айналасында азырақ өзгереді. Егер біз Y үшін кез-келген басқа мәнді болжаудан гөрі, әр кейс үшін Y-ң арифметикалық орташасын болжасақ, болжауда азырақ қателесеміз.

Әрине, тіпті, біз бұл стратегияны бұлжытпай орындасақ та, Y-ді болжауда көп қате жібереміз. Қате көлемі, белгіленген Y-ң арифметикалық орташасымен бірге, балалар саны мен ер адамның үй жұмысы арасындағы байланысты көрсететін экрандағы 12.6-сызбада берілген. Нақты көрсеткіштерден болған көрсеткіштерге қарай жүргізілген вертикалды сызықтар X-ті елемей Y-ді болжауда жіберуміз мүмкін қате санын көрсетеді.

12.6-сызба. X-сіз Y-ді болжау (ерлі-зайыптының екеуі де жұмыс жасайтын отбасылар)



Әрбір нақты Y көрсеткіштен Y-ң арифметикалық орташасын азайтып, оны квадраттап және бұл девиацияларды қосу арқылы, бірші шарт аясында (X-ті елемеу арқылы) біз өзіміздің болжаудағы қатеміздің таралуын анықтай аламыз.   деп белгіленген нәтижесі Y-ғы жалпы вариация деп аталады. Енді бізде X туралы ақпаратсыз Y-ді болжауда пайдаланатын визуалды бейне (12.6-сызба) мен қатені есептеу әдісі бар. Алда көретініміздей, іс жүзінде, r2, детерминация коэффициентінің мәнін анықтауда жалпы вариацияны есептеу қажеттілігі жоқ.

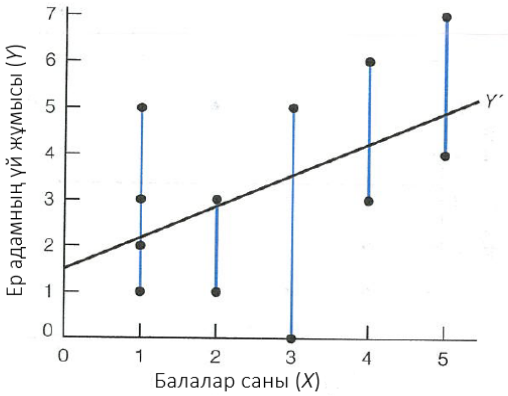
X арқылы Y-ді болжау

Біздің келесі қадамымыз қаншалықты X туралы мағлұмат Y-ді болжауда біздің дағдыларымызды жақсартатынын анықтайды. Егер екі айнымалы арасында сызықтық байланыс болса, онда ең кіші квадраттар теңдеуінен Y көрсеткіштерін болжау X туралы мағлұматты біріктіріп, біздің болжамдағы қателерімізді азайтады. Сондықтан екінші шарт аясында X-ң әр мәні үшін болжанған Y көрсеткіші келесідей болады.

$$Y' = a + bX$$

Экрандағы 12.7-сызба ерлі-зайыптының екеуі де жұмыс жасайтын отбасылардан алынған деректер мен олардың үстінен жүргізілген регрессиялық қисықты бейнелейді. Әрбір дерек бірлігінен/нүктесінен регрессиялық қисыққа дейінгі вертикалды сызықтар, X ескерілген жағдайда да қалып/сақталып отыратын, Y-ді болжаудағы қателер санын көрсетеді.

12.7-сызба. X арқылы Y-ді болжау (ерлі-зайыптының екеуі де жұмыс жасайтын отбасылар)



Біз X-ті ескеру нәтижесінде нақты түрде қатенің азаюын анықтай аламыз. Нақтырақ айтқанда, болжамның дұрысталғанын көрсететін статистиканы құрастыру үшін екі түрлі қосынды анықталып, әрі қарай Y-ң жалпы вариациясымен салыстырылады.

Түсіндірілген вариация деп аталатын, бірінші нәтиже X-ті ескергенде Y-ді болжауда біздің дағдыларымыздың дұрысталғанын көрсетеді. Бұл нәтиже әр кейстің регрессиялық теңдеулерімен болжанған көрсеткіштен Y-ді азайту, әрі қарай оны квадраттап, бұл айырмаларды қосу арқылы анықталады. Бұл амалдар реті экранда бар. ал қаншалықты X туралы мағлұмат Y-ді болжауда біздің дағдыларымызды жақсартатынын көрсету/дәлелдеу үшін қорытынды сан Y-ғы жалпы вариациямен салыстырылуы мүмкін. Нақтырақ айтқанда, ол математикалық жолмен келесідей болады

$$r^2 = \frac{\sum(Y' - \overline{Y})^2}{\sum(Y - \overline{Y})^2} = \frac{\text{Түсіндірілген вариация}}{\text{Жалпы вариация}}$$

Осылайша детерминация коэффициенті немесе r^2- X-пен байланысты немесе түсіндірілетін Y-ғы жалпы вариацияның үлесі. Басқа да алдын-ала (PRE) өлшемдер сынды r^2 Y-ді болжау, түсіну не түсіндіруде қаншалықты X-ң көмектесетінін нақты түрде көрсетеді.

Түсіндірілген вариация сынды біз X арқылы Y-ді болжаған кезде дұрыстау туралы айтамыз. Бұл терминді пайдалану Y-дағы кейбір вариациялар «түсіндірілмеген» немесе X-ң әсеріне байланысты болмайтынын меңзейді. Шындығында, 12.7-сызбадағы вертикалды сызықтар түсіндірілмеген вариацияларды немесе біздің X арқылы Y-дің ең жақсы болжамы және шынайы көрсеткіштер арасындағы айырмашылықты білдіреді. Осылайша, түсіндірілмеген вариация дегеніміз - регрессиялық қисық айналасында нақты көрсеткіштердің шашырауы; оны әрбір кейс үшін шынайы Y көрсеткішінен болжанған Y көрсеткішін азайту, әрі қарай алынған мәнді квадраттап, айырмашылықтардың нәтижесін шығару арқылы анықталады. Бұл амалдарды $\sum(Y-Y')^2$ ретінде сипаттап, ал нәтиже сан, X ескерілгеннен кейін де қалып отыратын, Y-ді болжағанда қате мөлшерін өлшейді. X-пен түсіндірілмеген Y-дағы жалпы вариацияның үлесі 1,00-нен r^2 мәнін азайту арқылы анықталады. Түсіндірілмеген вариация, әдетте, басқа айнымалылардың кейбір комбинацияларының ықпалына байланысты болады.

Осы уақытқа дейін өзіңіз көргендей, түсіндірілген және түсіндірілмеген вариациялар бір-бірімен өзара қарым-қатынаста болады. Бірінің мәні артқан сайын, екіншісінің мәні азайып отырады. Сонымен қатар X және Y арасындағы сызықтық байланыс неғұрлым мықты/қарқынды болған сайын, түсіндірілген вариацияның мәні жоғары, түсіндірілмеген вариацияның мәні төмен болады. Идеалды байланыс жағдайында ($r = \pm1{,}00$), түсіндірілмеген вариация 0 болар еді де, ал r^2 1,00 болар еді. Бұл Y-дағы барлық вариацияларды X түсіндіретінін немесе олардың себебі болатынын, сонымен қатар X-тен Y-ды қатесіз болжауымызға болатынын көрсетеді.

Басқа жағынан, X және Y сызықтық байланысты болмаған ($r = 0{,}00$) кезде, түсіндірілген байланыс 0 болар еді, ал r^2 0,00 болар еді. Мұндай жағдайда, Y вариацияларының ешқайсысын X түсіндірмейді, сондай-ақ біздің Y-ді түзету дағдымызды жақсартпайды деген қорытынды жасауымызға болады.

Осы екі шектер арасындағы байланыстар қаншалықты X біздің Y-ді болжау немесе түсіндіру дағдымызды арттырады деген тұрғыда интерпретациялануы тиіс. Ерлі-зайыптылардың екеуі де жұмыс жасайтын отбасылар үшін $r = 0{,}50$ деп есептедік. Бұл мәнді квадраттау $r^2 = 0{,}25$ детерминация коэффициентін береді, ол ер адамның үй жұмысы сағатының

Пирсонның корреляциясының қарқындығын интерпретациялау үшін

Сіз қарқындық/мықтылықты жалпылама интерпретациялау үшін орынды/тиімді әдісі - r мәнін квадраттап, оны 100-ге көбейту болар еді. Бұл мән X арқылы түсіндірілетін Y-ғы вариацияның пайыздық көрсеткішін көрсетеді.

Пирсонның корреляциясының бағытын интерпретациялау үшін

Егер r қосу белгісімен (немесе ешқандай белгісіз) болса, онда бағыты оң, ал айнымалылар бірдей бағытта өзгеріп отырады. Егер r азайту белгісімен болса, онда бағыты теріс, ал айнымалылар қарама-қайшы бағытта өзгеріп отырады. Біздің мысалдағы тапсырмада байланыс оң болды: балалар саны неғұрлым артқан сайын, ер адамның үй жұмысына араласу сағаты да артып отырды.

Әлеуметтік ғылымдардағы жобалар бірнеше айнымалылардан тұрады, сондықтан жобаның деректерді талдау кезеңі көбінесе корреляциялық матрицаларды тексеруден басталады - айнымалылардың барлық ықтимал жұптары арасындағы байланысты көрсететін кесте. Корреляциялық матрицалар бізге деректер жиынындағы өзара байланыстардың жедел, оқуға ыңғайлы шолуын береді, сонымен қатар кейінгі талдау үшін стратегиялар мен "бағыттар" ұсынады. Мұндай кестелер көбінесе кәсіби зерттеу әдебиеттеріне енгізіледі, сәйкесінше оларды оқуда белгілі бір тәжірибенің болғаны дұрыс.

12.5-кесте. Төрт айнымалы арасындағы байланысты көрсететін корреляциялық матрица

	1	2	3	4
Туу көрсеткіші	Туу көрсеткіші	Білімі	Кедейлік	Жасөспірімдердің тууы
Білімі	1,00	-0,19	0,10	0,40
Кедейлік	-0,19	1,00	-0,71	-0,76
Жасөспірімдердің тууы	0,10	-0,71	1,00	0,81
	0,40	-0,76	0,81	1,00

Экрандағы кестеде Төрт айнымалы арасындағы байланысты көрсететін корреляциялық матрица берілген

мұндағы "Туу көрсеткіші" - 1000 адамға шаққандағы туу саны/көлемі/мөлшері
"Білімі" - арнайы орта немесе одан жоғары білімі бар тұрғындардың пайыздық көрсеткіші
"Кедейлік" - 2009 жылы кедейлік сызығынан төмен отбасылардың пайыздық көрсеткіші
"Жасөспірімдердің тууы" - 15-19 жастағы 1000 отбасына шаққандағы туу көрсеткіші

Корреляциялық матрицаны талдап көрелік

Қатарлар мен бағандар сынды матрица айнымалы атауларын пайдаланады, ал кестедегі ұяшықтар айнымалылардың әр комбинациясы үшін екі-айнымалы корреляцияны көрсетеді. Қатарлардың атауы бағандар атауларын қайталап отырғанын ескеріңіз. Кестені оқу үшін, сол жақтағы бағандағы (1-баған) бірінші қатарда (1-қатар) орналасқан туу көрсеткішінен бастаңыз. Бұл айнымалы мен басқа айнымалы, оның ішінде сол жақтағы ұяшықтағы туу көрсеткішінің өз-өзімен (1,00) корреляциясын көру үшін 1-баған бойымен төмен қарай немесе 1-қатар бойымен оқыңыз. Басқа айнымалылар арасындағы байланысты көру үшін бағаннан бағанға немесе қатардан қатарға жылжып отырыңыз.

Матрицаның сол жағында жоғарыдан оң жағында төменге қарай диагональ әр айнымалының өз-өзімен корреляциясынан тұратынын айта кету керек. Бұл диагональ бойындағы мәндер әрдайым 1,00 болады, ал бұл ақпарат пайдалы емес болғандықтан, оны нәтижелерден оңай алып тастауға болады.

Сондай-ақ, диагональдің төмендегі оңнан солға қарай орналасқан ұяшықтары жоғарыдағы солдан оңға қарай орналасқан ұяшықтарын қайталайды. Мысалы, 1-бағандағы екінші ұяшыққа (2-қатар) қараңыз. Бұл ұяшық, 2-бағанның жоғарғы қатарындағы ұяшық (1-қатар) сынды, туу көрсеткіші/коэффициенті мен арнайы орта білімі бар тұрғындардың пайыздық көрсеткіші арасындағы корреляцияны бейнелейді. Басқаша айтқанда, төмендегі оңнан солға қарай орналасқан ұяшықтар жоғарыдағы солдан оңға қарай орналасқан ұяшықтардың айнасы болады. Әдетте, кәсіби әдебиеттегі мақалаларда кестені оқуға ыңғайлы ету үшін артық ұяшықтар кетіріп тасталады.

Бұл матрица бізге не айтады? Кестенің жоғарғы сол жағынан бастасақ (1-баған), біз туу көрсеткіші/коэффициенті мен арнайы орта білім деңгейі арасында әлсіз және орташа теріс қатынас бар екенін көре аламыз. Бұл білім деңгейі өскен сайын туу көрсеткіші төмендейтінін білдіреді. Тұрғындардың арнайы орта білім деңгейі жоғары мемлекеттерде туу деңгейі төмен болады. 1-бағанды жалғастырсақ, туу көрсеткіші/коэффициенті мен кедейлік арасында әлсіз оң байланыс бар екенін (кедейлік артқан сайын туу коэффициенті артады) және жасөспірімдердің туу коэффициенті орташа оң байланыс бар екенін (туу коэффициенті мемлекеттерде жасөспірімдердің туу деңгейінің жоғарырақ болады) көреміз.

Деректер жинағындағы басқа байланыстарды бағалау үшін рет-ретімен бір айнымалы бойынша бағаннан бағанға немесе қатардан қатарға жылжып отырыңыз. Әрбір келесі айнымалы үшін жаңа ақпарат ұяшығы болады. Мысалы, 2-баған мен 2-қатардағы білімді (арнайы орта білімі ең аз болған мемлекет тұрғындарының пайыздық көрсеткіші) қарастырайық. Біз жаңа ғана оның туу деңгейімен әлсіз және орташа теріс байланысын көрдік, сондай-ақ әрине, айнымалының өз-өзімен байланысын елемеуге болады. Бұл 2-баған бойымен төмен қарай немесе 2-қатар бойымен оқуға болатын екі жаңа байланысты ғана қалдырады. Білім беру кедейлікпен мықты теріс байланыста (арнайы орта білімі бар тұрғындардың пайыздық көрсеткіші жоғары болған сайын кедейлік көрсеткіші төмен болады) және жасөспірімдердің туу көрсеткішімен де мықты теріс байланыста (білім деңгейі жоғары мемлекеттерде жасөспрімдердің тууы аз).

3-бағандағы айнымалы, кедейлік үшін бір ғана жаңа байланыс бар: жасөспірімдердің тууымен мықты оң корреляция (кедейлік неғұрлым жоғары болса, жасөспірімдердің туу соғұрлым жоғарырақ болады). 4-баған (жасөспірімдердің тууы) үшін қарастыратын ешқандай жаңа байланыс жоқ: біз бұл айнымалы мен барлық басқа айнымалылар арасындағы барлық корреляцияларды қарастырып қойдық.

Корреляция мен каузалдық ұғымдарына талдау жасайық

Әртүрлі жағдайларда байқағанымыздай, корреляция мен каузалдылық екі түрлі нәрсе. Аталған мысалда біз бірқатар мықты байланыстарды көруіміз мүмкін және бұл айнымалылардың бірі екіншісінің себебі деп қорытыдылауға сұранып отыр. Мысалы, білім беру мен кедейлік арасындағы тығыз байланысты көру таңқалардырмас, алайда, себеп-салдар/каузалдық мәселесі анық емес. Білім берудің төмен деңгейі кедейлікті тудырады ма, әлде кедейлік білім деңгейінің төмен болуына әкеледі ме? Қандай айнымалы себеп, ал қайсысы салдары болады?

Кедейлік пен білім беру сынды айнымалылар арасындағы себеп-салдарлық/каузалды байланыс туралы кейсті қалай жасауға болады? Қандай дәлелдер қажет? Бірінші шарт - байланыс өлшемінің мәні. Неғұрлым өлшем әлсіз/нашар болса, соғұрлым каузалдық байланыс туралы айту қиынырақ болады. Бұл критерий бойынша кедейліктің және білім беру арасында каузалдық байланыс бар деген дәлелмен мықты корреляция анық болады.

Екінші шарт - уақыт тәртібі: каузалдықты дәлелдеу үшін біз тәуелсіз айнымалы тәуелді айнымалыға дейін орын алатынын көрсетуіміз керек. Мысалы, көптеген жылдар бойына кедейлік деңгейінің артуы колледжге түсетін адамдардың санының азаюына алып келгенін көрсете алуымыз мүмкін: бұл да каузалдық байланыс бар екеніне дәлел болады.

каузалдық байланыс үшін кейс жасаудағы соңғы қадам - екі-айнымалы байланысқа ешқандай үшінші айнмалы әсер етпейтінін көрсету. Біздің мысалымызды жалғастыру үшін, біз білім деңгейлері мен кедейлік басқа да экономикалық немесе мәдени өзгерістерге қарамастан (мысалы, жұмыссыздық деңгейі, отбасы көлемінің өзгеруі, қылмыс көрсеткіштері немесе көптеген басқа да айнымалылар) өзара байланысты екенін көрсете алуымыз қажет. Бұл критерийді қанағаттандыру көп айнымалы талдауларды қажет етеді.

Қорытындылайтын болсақ, есіңізде болсын, корреляция қажетті, бірақ каузалдық үшін жеткілікті не қанағаттандырарлық шарт емес. Егер корреляция нөлге тең немесе өте әлсіз болса, онда бұл байланыс каузалды емес деуге болады. Тіпті өте мықты корреляциялармен де, бір айнымалы екіншісінің себебі деген тұжырым жасамас бұрын, біз бірнеше принциптерді/шарттарды қарастыруымыз керек. Сонымен қатар, корреляция мен статистикалық мәнділік те екі түрлі нәрсе екенін есте сақтаңыз.

Корреляция, Регрессия, Өлшем деңгейі және шартты айнымалы

Корреляция және регрессия - бұл мүмкіншілігі өте көп және тиімді әдістер екені соншалық олар реттік деңгейдегі айнымалылар арасындағы байланыстарды талдауда жиі пайдаланылады. Бұл тәжірибе “үзіліссіз” болып табылатын әсіресе, реттік айнымалылармен мәселе туғызбайды. Дегенмен бұл икемділік діни деноминация немесе гендер сынды номиналды деңгейдегі айнымалыларға қолданылмайды. Номиналды айнымалылардың көрсеткіші/мәні/шамасы сандар емес және олардың математикалық белгісі жоқ. Біз Протестантты “2” деген көрсеткішпен, ал Католикті “1” деген көрсеткішпен алуымызға болады, дегенмен соңғы көрсеткіш алдыңғысына қарағанда “екі есе көп” болмайды. Номиналды деңгейдегі айнымалылар - сандар емес, олар - белгілер, сондықтан мұндай номиналды деңгейдегі айнымалылар үшін “көлбеуді” есептеу немесе оң не теріс байланыстар туралы айтудың мәні жоқ.

Бұл - келеңсіз жағдай. Біздің күнделікті әлеуметтік өміріміздегі, гендер, отбасылық статус, нәсіл мен ұлт, ең маңызды айнымалылар - номиналдық деңгейдегі өлшемдер, сәйкесінше олар үшін - әлеуметтік ғылымдардағы зерттеулер үшін ең мүмкіндігі мол және күрделі екі құрал - регрессиялық теңдеулер немесе корреляциялық анализ қарастырылмайды.

Алайда зерттеушілер шартты айнымалыларды құрастыру арқылы номиналды деңгейдегі айнымалыларды өңдеуді өзгертудің жолын жасап шығарды. шартты айнымалылар кез-келген өлшем деңгейінде, оның ішінде номиналды және бірі 0, екіншісі 1 деп кодталған екі категориясы да болуы мүмкін. Бұл тұрғыда гендер (мысалы, ер адамдар үшін 0 деген кодпен, ал әйел адамдар үшін 1 деген кодпен), нәсіл (ақ нәсілділер үшін 0 деген кодпен, ал қара нәсілділер үшін 1 деген кодпен) және діни деноминация (Католиктер үшін 0 деген кодпен, ал Протестанттар үшін 1 деген кодпен) сынды номиналды деңгейдегі айнымалылар көбінесе регрессиялық теңдеулерге енгізіледі.

Біз каузалды байланыстың бар екенін көрсету үшін қолданылатын үш өлшемді атап өттік. Біз оларды күнделікті өмірдегі ұсынылған байланыстар үшін қалай қолдануға болады?

Мысал ретінде Шылым шегу мен қатерлі ісік арасындағы байланысты қарастырайық. Бүгінде іс жүзінде бәрі де бірі екіншісінің себебі екенін біледі, бірақ бірнеше ұрпақ бұрын бұл ақпарат қабылданған саналы ойдың бөлігі болмады. 1950 жылдарда ғана, барлық адамзаттың жартысы шылым шекті, ал шылы шегу сырқат пен аурудың емес, ілгері не алдыңғы қатарлы ересекке теңестірілді. Содан бері медициналық зерттеу каузалды жағдайды анықтап, бұл болжам кеңінен таратылды. Әсері анық болды: қазір ересектердің 20%-н азы шылым шегуде.

Пирсонның корреляциясы сынды байланыс өлшемі каузалдықты құрастыруда маңызды рөл атқарды. Статистиканың күші мен бағыты темекі шегу мен аурудың арасындағы мәнді корреляцияны тұрақты түрде көрсетті.

Айнымалылар арасындағы уақыт тәртібін көрсету үшін ең жақсы, ең күшті зерттеулер ұзақ уақыт бойы адамдар тобын бақылады. Зерттеулер денсаулығы сыр бермеген/жақсы саны жағынан көп шылым шегушілерден басталды. Егер зерттеуде шылым шегуші кейін қатерлі ісікке ұшыраған болса, қатерлі ісік тәуелді айнымалы болғаны. Яғни, уақыт тәртібін ескере отырып, темекі шегу қатерлі ісікке әкелуі мүмкін, бірақ кері әсер мүмкін болмайды: қатерлі ісікпен ауру адамның темекі шегуіне себеп болуы мүмкін емес.

Уақыт тәртібі өз бетімен байланыстың каузалды екенін дәлелдей алмайды. Адамдардың шылым шегуі және олардың қатерлі ісікке шалдығуы басқа да айнымалылардың болуы және түсіндіре алуы керек. Мысалы, тым мазасыз адамдар “жүйке жүйесін тыныштандыру үшін» шылым шегуі мүмкін, сәйкесінше шылым шегуден гөрі, өздерінің мазасыздығы салдарынан қатерлі ісікке шалдығатын болар. Егер бұл кейс болса, шылым шегу мен қатерлі ісік арасындағы байланыс жасанды немесе жалған болады: екі-айнымалы сынды каузалды байланысты, шын мәнінде, мазасыздық тудырады.

Зерттеушілер мұндай мүмкіндікті қалай қабылдады? Ең сенімді зерттеулер мысалы, мазасыздық, отбасы тарихы, гендер, нәсіл және басқа да үшінші айнымалылардың әсерін бақылау мен зерттеуде әртүрлі құралдарды пайдаланады, сонымен қатар екі-айнымалы байланыстарға бұл факторлардың әсер ету ықтималдығын статистикалық түрде жойды.

Әрқашан бір айнымалы екіншісінің себебі болады дегенде мына сұрақтар арқылы ойланыңыз: байланыс өлшемі қаншалықты мықты? Алдымен тәуелсіз айнымалы пайда болды ма? Байланысқа қандай басқа айнымалылар әсер етуі мүмкін? Олар ескерілген бе? Байланыстың бұл критерийлерді қанағанттандыруы неғұрлым жоғары болса, каузалды байланыс соғұрлым мықты болады.

Осымен статистика курсы бойынша тағы бір дәрісіміз аяқталды, назар қойып тыңдағаныңыз үшін алғыс білдіреміз.