



25-дәріс



ҚАЗАҚСТАННЫҢ
АШЫҚ
УНИВЕРСИТЕТІ

СТАТИСТИКА НЕГІЗДЕРІ

Көптүрлі корреляция





Тарауды оқығаннан кейін:

1. Көптүрлі детерминация коэффициентін есептеп түсіндіресіз (R^2).
2. Жартылай және көптүрлі регрессиялық талдаудың шектеулерін түсіндіресіз.

Естеріңізге сала кетейін, алдыңғы екі және осы дәрістерде ұсынылған статистикалық әдістер интервалды-қашықтықты екі айнымалы деңгейінің арасындағы байланысты талдау үшін пайдаланылады. Бұл жағдайлар үшін зерттеу әдістерінің мысалдары:

1. Отбасын зерттеуші әлеуметтанушылар екі жылдық жалақысы бар отбасыларды зерттейді және күйеулері өздерінің үй жұмысы атқаруын балалар санының ұлғаюына қарай ұлғайтады ма деген сұрақ қояды. Ерлерінің біліміне қарамастан байланыстары бірдей болады ма?

2. АҚШ-тың 542 округі бойынша жасалған іріктелген жинақ үшін, кедейшілік пен қылмыс деңгейі арасындағы орташа және оң корреляцияны көрсететін криминолог мәліметтері берілген. Тығыз орналасқан қалалық округтерде және аз елді ауылдық мекендерде осы қатынастар бірдей ме деген сұрақ туындайды? Округтердегі жоғары және төмен білім деңгейлеріне қатысты бұл қатынастар бірдей ме?

3. Демографиялық туу деңгейі мен 76 елдегі әйелдердің білім деңгейі арасындағы өзара байланысты зерттейді. Мәліметтер айналымдар арасындағы орташа, теріс байланысты көрсетеді: білім деңгейінің өсуіне қарай туу көрсеткіші төмендейді. Қарым-қатынас барлық деңгейдегі елдер үшін өз күшін сақтайды ма?

Сонымен, біз әрбір тәуелсіз айнымалының тәуелді айнымалыға тікелей әсерлерін ажырату үшін бірнеше регрессиялық теңдеулерді қолданамыз. Көптүрлі корреляциялық техникаларды қолдану арқылы біз барлық тәуелсіз айнымалылардың тәуелді айнымалыға аралас әсерлерін анықтай аламыз. Біз мұны көптүрлі корреляция коэффициентін (R) және көптүрлі детерминация коэффициентін (R^2) есептеу арқылы жасаймыз. Соңғы статистиканың мәні Y ішіндегі вариацияның (*variance*) үлесін көрсетеді, бұл барлық тәуелсіз айнымалыларды біріктіру арқылы түсіндіріледі.

Балалардың саны (X_1) Y -тегі ($r_{y1} = (0,50)^2 = 0,25$, 100-ге көбейтілген кезде 25% болады) 25% вариацияны түсіндіреді, ал күйеудің білімі Y -тегі ($r_{y2}^2 = (-0,30)^2 = 0,09$, 00-ге көбейтілген кезде, 9%-ды құрайды) 9% вариацияны түсіндіреді. Екі нөл-тәртіптегі корреляцияны олардың Y -ке қосарланған әсерін анықтау үшін жәй біріктіре салуға болмайды, өйткені екі тәуелсіз айнымалылар өзара бір-бірімен корреляцияланған, сондықтан олар өздерінің Y -ке әсер етуінде «бір-бірін жауып тастайды» да, кейбір бірдей вариацияларды түсіндіретін болады. Бұл «бір-бірін жауып тастау» жою 13,11-ші формула бойынша жүзеге асады:

13,11-ші формула

$$R^2 = r_{y1}^2 + r_{y2.1}^2 (1 - r_{y1}^2)$$

мұнда,

R^2 = көптүрлі детерминация коэффициенті

r_{y1}^2 = Y пен X_1 арасындағы нөл-тәртіптегі корреляция, квадратталған сан.

$r_{y2.1}^2$ = X_1 бақыланатын кездегі Y пен X_2 жартылай корреляциясы, квадратталған сан.

Осы формуладағы бірінші (r_{y1}^2) - Y пен детерминацияларының коэффициенті немесе -мен түсіндірілген Y вариациясының шамасы. Бұл санға біз $(1 - r_{y1}^2)$ түріндегі Y вариациясының санын қосамыз, бұны ($r_{y2.1}^2$) болып бақыланған X_1 әсерінен кейін X_2 арқылы түсіндіруге болады. Негізінде, 13,11-формула X_1 -ге Y қаншалықты мүмкін бола алатынын түсіндіруге әкеледі, содан соң X_1 бақыланғаннан кейін X_2 әсері қосылады (осылайша X_1 мен X_2 ортақ иеленетін Y вариациясындағы «бір-бірін жауып тастауды» жойып тастайды).

R^2 есептеу және интерпретациялау

Күйеулердің үй шаруашылығындағы үлесіне (Y) балалардың саны (X_1) және күйеуінің білімінің (X_2) бірлескен әсерін бақылау үшін бізге екі мөлшер қажет. X_1 және Y ($r_{y1} = 0.50$) арасындағы корреляция табылған. 13,11-формуланы шешпес бұрын, алдымен X_1 $r_{y2.1}^2$ бақыланатын кездегі Y пен X_2 жартылай корреляциясын есептеп алу керек. Бұл санды 13,11-формуланы бойынша табуға болады, бірақ біз $r_{y2.1} = -0.08$ екенін жариялай аламыз.

13,11-формуланы енді біздің мысалға қатысты мәселеде шешуге болады:

$$R^2 = r_{y1}^2 + r_{y2.1}^2 (1 - r_{y1}^2)$$



$$R^2=(0.50)^2+(-0.08)^2(1-0.502)$$

$$R^2=0.25+(0.006)(1-0.25)$$

$$R^2=0.25+0.005$$

$$R^2=0.255$$

Бірінші тәуелсіз айнымалы (X_1) балалардың саны өз бетінше Y ішіндегі варианстың 25% -ын түсіндіреді. Осы барлық мәнге, екінші тәуелсіз айнымалы күйеуінің білімі (X_2) жарты пайыз ғана қосады, сондықтан барлық түсіндірілген вариантс мәні 25,5% болады. Комбинацияда екі тәуелсіз айнымалы тәуелді айнымалы вариацияның жалпы 25,5%-н түсіндіреді.

Көптүрлі детерминация коэффициентін (R^2) есептеудің орындалуын бекіту үшін келесі қадамдарды орындаймыз.

Рет-ретімен орында	Көптүрлі детерминация коэффициентін есептеу (R^2)
<i>Бұл процедуралар екі тәуелсіз айнымалылар мен бір тәуелді айнымалы болған кезде қолданылады. Күрделірек жағдайларды есептеу үшін SPSS сияқты компьютерленген статистикалық пакетті пайдаланыңыз.</i>	
Реті	Орындалуы
	13.11-формула арқылы көптүрлі детерминация коэффициентін есептеу үшін:
1.	r у2.1жартылай корреляция коэффициентін табыңыз.
2.	1-қадамда табылған мәнді квадраттаңыз.
3.	r у1мәнін квадраттаңыз.
4.	3-қадамда шыққан мәнді 1,00 санынан шегеріңіз.
5.	4-қадамда шыққан мәнді 2-қадамда табылған мәнге көбейтіңіз.
6.	5-қадамда шыққан мәнді 3-қадамда табылған мәнге қосыңыз. Бұл мән – көптүрлі детерминация коэффициенті.

R^2 интерпретациялау

Көптүрлі детерминация коэффициенті барлық басқа тәуелсіз айнымалылардың бірігуімен түсіндірілген Y -тегі барлық вариациялар саны.

Регрессиялық анализ жасауда SPSS-ті пайдалану

Осы тарауда көрсетілген түсінікті себептерге байланысты зерттеушілер регрессиялық талдау жүргізу үшін қолмен есептеуді емес, SPSS сияқты компьютерленген статистикалық пакеттерді пайдаланады. «SPSS қолдану» бөлімшесінде біз Мемлекеттік деректер жиынтығын пайдаланып, регрессиялық талдау жүргіземіз. Біз TrafDths11 жол апаты өлім деңгейін тәуелді айнымалы ретінде қарастырамыз. Біздің тәуелсіз айнымалыларымыз 65 жастан асқан (Older) ел тұрғындарының үлесі және халықтың тығыздығы (PopDense) немесе штаттағы шаршы миляға шаққандағы адамдардың саны болады. Егде жастағы адамдар тұрғысынан штаттардағы көлік апатынан болатын адам өлімінің көрсеткіші төмен немесе жоғары деңгейде бола ма? Халқы неғұрлым тығыз орналасқан штаттар жүргізушілер үшін қауіпті ме?

Күнделікті өмір статистикасынан мысал. Жасөспірімдер жүктілігі

Алдыңғы дәрістерде 50 штаттардағы жасөспірімдердің туу көрсеткіштері туралы деректерін пайдалан отырып кейбір корреляцияларды қарастырдық (12,5-кестені қараңыз). Біз білім деңгейі өлшемімен күшті теріс бивариативті қарым-қатынастарды (-0,78), ал кедейлік деңгейімен күшті оң қарым-қатынастарды (0,88) анықтадық. Кедейлік пен білім деңгейінің жасөспірімдердің туу көрсеткіштеріне қандай біріккен әсерлері бар?

Өз кезегінде, білім штаттағы жасөспірімдердің туу көрсеткіштерінің варианттарының 61% -ын түсіндіреді. Бұған кедейлік 11% қосады, сонда біріккен түсіндірілген вариант болдаы 72% ($R^2 = 0.72$).

Бета салмағы кедейліктің оң нәтиже беретінін көрсетеді (білім деңгейі үшін бақыланғаннан кейін) және білімнің теріс әсері (кедейлікті бақылаған соң), және кедейліктің ($b^*=0.37$) әсері ($b^* = 0.54$) білімге



қарағанда әлдеқайда күштірек екенін көрсетеді.

Осы нәтижелер туралы екі нәрсені ескеріңіз. Біріншіден, бұл әлеуметтанулық зерттеулер үшін өте жоғары корреляция. Жалпы алғанда «әлсізден қалыптыға қарай» шектегі корреляциялар әлдеқайда жиі кездеседі, ал біздің жартылай өлшемдеріміз дәл емес.

Екіншіден, осы нәтижелерге сүйенсек, онда кедей, нашар білімді жасөспірім қыздардың жүкті болу ықтималдығы көбірек екендігі туралы қорытынды жасауға тура келеді. Бұл тұжырым қисынды болса да, деректер жеке адамдарға емес, штаттарға қатысты екенін ұмытпаңыз. Бұл нәтижелер тек жоғары деңгейдегі кедей штаттар мен білім деңгейінің төмендігі жасөспірімдегі туу көрсеткішін жоғарылататынын көрсетеді. Жүкті жасөспірімдердің сипаттамалары туралы сенімді тұжырым жасамас бұрын, алдымен бізге көбірек ақпарат жинақтау қажет болады.

Көптүрлі регрессия мен корреляцияның шектеулері

Көптүрлі регрессия және корреляция үш немесе одан да көп айнымалылар арасындағы өзара әрекеттесуді талдау үшін өте мықты құрал болып табылады. Осы тарауда ұсынылған техникалар зерттеушіге екі немесе одан да көп айнымалылардан бір айнымалы мәнді болжау, тәуелді айнымалыға тікелей әсер етудің маңыздылығы тұрғысынан тәуелсіз айнымалыларды ажырату, және тәуелді айнымалыға тәуелсіз айнымалылар жинағының барлық әсерін анықтау мүмкіндігін береді. Техникалардың икемділігі мен ақпарат бере алатын көлемі бойынша, көптүрлі регрессия мен корреляция әлеуметтік ғылымдар зерттеушілерін қол жетімді кейбір аса мықты статистикалық техникалармен қамтамасыз ете алады.

Мықты құралдар арзан емес. Олар жоғары сапалы деректерді талап етеді, ал әлеуметтік ғылымдар бойынша зерттеулерде интервалды-арақашықтық деңгейде өлшеу жасау қиын болуы мүмкін. Бұдан басқа, бұл техникалар айнымалылар арасындағы өзара қарым-қатынас белгілі бір пішінге сәйкес келетінін болжайды. Біріншіден, әрбір тәуелсіз айнымалы тәуелді айнымалымен сызықтық қарым-қатынаста болады деп есептейді. Айнымалылардың осы берілген жинағы қаншалықты болжамға жақсы сәйкестенетінін скаттерплотпен тез тексеруге болады.

Екіншіден, бұл техникалар теңдеуде айнымалылар арасында ешқандай интеракцияның жоқтығын болжайды. Егер айнымалылар арасында интеракция бар болса, онда тәуелді айнымалыларды тәуелсіз айнымалылардың жай ғана қосылған әсерлері арқылы дәл бағалау немесе болжау мүмкін болмайды. Жинақтағы айнымалылар арасындағы интеракциямен жұмыс жасаудың техникалары бар, бірақ бұл техникалар осы кітаптың ауқымынан тыс.

Үшіншіден, көптүрлі регрессия мен корреляция тәуелсіз айнымалылар бір-бірімен корреляцияланбайды деп есептейді. Турасын айтқанда, бұл шарт тәуелсіз айнымалылардың барлық жұптарының нөл-тәртіптік корреляциясы нөлге тең болуға тиіс дегенді білдіреді; дегенмен, іс жүзінде біз, егер тәуелсіз айнымалылар арасындағы өзара корреляциялар төмен болатын болса, бұл жорамалар қабылданғандай әрекет етеміз.

Бұл жорамалдар бұзылатын болғандықтан, регрессиялық коэффициенттер (әсіресе жартылай және стандартталған көлбеулер) және көптүрлі детерминация (R^2) коэффициенті азая бастайды және сенімді азайта бастайды, соған қоса пайдасыздау техника болады. Биваритативті скаттерплоттарды мұқият тексеру жорамалдардың қайсысы ақылға қонымды екендігін бағалауға көмектеседі. Ақырында, біз жартылай корреляцияның, көптүрлі регрессия мен корреляцияның қарапайым ғана қосымшаларын қамтығанымызды атап өтуіміз керек. Логика мен интерпретация тұрғысынан, айнымалылардың көп мөлшерін қамтитын жағдайларға қатысты кеңірек қарастырулар салыстырмалы түрде қарапайым. Алайда, бұл жағдайларды есептеу өте күрделі. Егер сіз үш айнымалыдан астам жағдаймен жұмыс жасасаңыз, әдетте университеттерде қол жетімді болатын компьютерленген статистикалық пакеттердің (мысалы, SPSS немесе SAS) біріне жүгініңіз. Бұл бағдарламалар аса жоғары компьютерлік сауаттылықты қажет етпейді, және қас қағым сәтте күрделі есептеулерді жасай алады. Осы пакеттерді тиімді пайдалану сізді қиындықтардан аластайды, және бәр жердегі әлеуметтік ғалымдар ұнататын нәрселерді жасауға мүмкіндік береді: алынған нәтижелердің мағынасын зерделеп, әлеуметтік өмірдің сипатын туралы ойды кенейту.

Сыни көзқарас: Корреляция, себеп-салдарлық және қатерлі ісік

Танымал баспасөзде немесе күнделікті әңгімелесулеріңізде көптүрлі регрессияны кездестіруіңіз екіталай. Екінші жағынан, көптүрлі регрессиялық талдау әлеуметтік зерттеулерде кеңінен қолданылады, осы техниканы пайдаланатын мақалалармен немесе оның көп варианттарының бірі ретінде кәсіби ғылыми-зерттеу әдебиеттерімен жұмыс жасауға тура келеді. Бұл мақалалар өте күрделі болып көрінуі



мүмкін, шын мәнінде, олар арнайы мамандар үшін жазылған, оған қоса оқырманның статистикалық деңгейін жоғары деп есептейді.

Дегенмен, бар қиындықты жоққа шығармастан, «қарапайым адамдар» осы мақалалардың мәнін, біраз нұсқауларды орындай отырып, кейбір негізгі элементтерді табу арқылы түсінуі әбден ықтимал. Ең бастысы сандарға емес, сөздерге көңіл бөлу. Яғни, мәтінді оқи отырып, авторлар сандар мен символдар туралы емес, өздерінің нәтижелері туралы не нәрселер айтқанын қараңыз. Статистикалық талдаудың егжей-тегжейлері сіздің түсінігіңізден тыс болуы мүмкін, бірақ сөздерді әрдайым дерлік түсіне аласыз).

Нәтижелері, әдетте, стандартталмаған және стандартталған коэффициенттер, R^2 мүмкін, және басқа ақпарат туралы есеп беретін түйіндеме кестелерде ұсынылады. Осындай түйіндеме кестенің мысалы келесідей болады:

Тәуелсіз айнымалылар	R^2	Бета-салмақтар
X_1	0,17	0,47
X_2	0,23	0,32
X_3	0,27	0,16

Бұл кестеде бірінші тәуелсіз айнымалы X_1 тәуелді айнымалымен тікелей қарым-қатынасы бар және тәуелді айнымалы мәндегі вариацияның 17% -ын өзі ($R^2 = 0.17$) түсіндіреді. Екінші тәуелсіз айнымалы X_2 түсіндірілген вариацияға 6%-ды қосады (X_2 теңдеуге енгізілгеннен кейін $R^2 = 0.23$). Үшінші тәуелсіз айнымалы X_3 , тәуелді айнымалының түсіндірілген вариациясына 4%-ды қосады (X_3 теңдеуге енгізілгеннен кейін $R^2 = 0,27$).

Зерттеу туралы нақты есептердің мысалын қысқаша қарастырайық. Зерттеулер БАҚ-тағы зорлық-зомбылықтың жоғары деңгейі қоғамның қылмыстық виктимизациядан қауіптенуін арттыратынын анықтады, алайда әлеуметтанушы Валери Калланан бірнеше жағдайларды қарастыруды қалады: әр түрлі БАҚ-тар әр түрлі әсер етеді ме? Бұқаралық ақпарат құралдарындағы зорлық-зомбылық әртүрлі қауымдастықтарға әр түрлі әсер етеді ме?

Оның кейбір нәтижелері Калифорния тұрғындарының үлкен кездейсоқ жиынтығына негізделген, ол келесі кестеде келтірілген. Кейбір тәуелсіз айнымалылар орынды үнемдеу үшін алынып тасталды.

Білім деңгейінің өсуіне байланысты қылмыс қаупі де төмендейді. Бұқаралық ақпарат құралдарындағы зорлық-зомбылықты жақын қабылдаған адамдар (әсіресе, шындық шоуларынан) қылмыс құрбандары сияқты үркектеу болады. Бұл айнымалылардың барлығы біріге отырып қылмыс қорқынышы вариациясының 17% -ын түсіндірді.

Айнымалылар	Бета-салмақ
Білім	- 0,17***
Жергілікті ТВ жаңалықтары	0,06***
Криминалдық драмалар	0,04
Шындық шоуларындағы криминал	0,12**
Криминал құрбаны	0,43***
R^2	0,17

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Көбірек білгіңіз келсе, төменде берілген дереккөзді қараңыз.

Дереккөз: Callanan, Valerie. 2012. "Media Consumption, Perceptions of Crime Risk and Fear of Crime: Examining Race/Ethnic Differences." *Sociological Perspectives* 55: 93-115.

Көптеген регрессия процедурасында қолданбалы бағдарламалардың статистикалық пакеттерінде әдетте бірнеше корреляцияның (анықтаманың) түзетілген коэффициенті (индексі) берілген. Көптеген анықтау коэффициентінің мәні регрессиялық модельдің сапасын бағалау үшін пайдаланылады. Көптеген корреляция коэффициентінің төмен көрсеткіші (индекс) бір жағынан маңызды факторлар регрессиялық моделіне кірмейтіндігін, ал екінші жағынан, қарастырылған байланыс түрі формаға енгізілген ауыспалардың арасындағы шынайы қарым-қатынасты көрсетпейді. Модельдің сапасын арттыру және оның практикалық маңыздылығын арттыру үшін қосымша зерттеулер жүргізу қажет.

Түйін

1. Көптүрлі детерминация коэффициенті (R^2) тәуелді айнымалы барлық тәуелсіз айнымалылардың



Кітап: Статистика негіздері

Дәріс: Көптүрлі корреляция

аралас әсерлерін жалпы вариация Y үлесіне қатынасы бойынша көрсетеді.

2. Жартылай корреляция, көптүрлі регрессия және корреляция - зерттеуші үшін қол жетімді ең күшті құралдардың бірі. Бұлар жоғары сапалы өлшеу мен сызықтық және интерактивті емес айнымалылардың арасындағы қарым-қатынастарды талап етеді. Сонымен қатар, тәуелсіз айнымалылар арасындағы корреляция төмен болуы керек (дұрысы 0). Бұл техникалардың құны жоғары болғанымен, олар айнымалылар арасындағы өзара қарым-қатынас туралы нақты және егжей-тегжейлі ақпарат көлемін жасауға айтарлықтай дивидендтер беріледі.