



СТАТИСТИКА НЕГІЗДЕРІ

Жартылай корреляция және көптүрлі
регрессия және корреляция





Тарауды оқығаннан кейін:

1. Корреляцияның жартылай коэффициенттерін есептей аласыз және түсіндіресіз.
2. Жартылай көлбеу ең кіші квадраттардың көптүрлі регрессия теңдеуін табасыз және түсіндіресіз.
3. Стандартты жартылай көлбеулерді немесе бета-салмағын табасыз және түсіндіресіз (b^*).

Статистиканы пайдалан

Осы және келесі дәрістерде ұсынылған статистикалық әдістер интервалды- қашықтықты екі айнымалы деңгейінің арасындағы байланысты талдау үшін пайдаланылады. Бұл жағдайлар үшін зерттеу әдістерінің мысалдары:

1. Отбасын зерттеуші әлеуметтанушылар ерлі-зайыптылардың екеуі де жұмыс істейтін (қос жалақылы) отбасыларды зерттейді. Отбасында ерінің үй шаруасына араласуы балалар санының көбеюіне қарай ұлғая ма деген сұрақ қызықтырады. Ерлерінің білім деңгейіне қатысты байланыс бірдей бола ма?

2. Криминологтың қолында АҚШ-тың 542 округі бойынша жасалған жинақтан кедейлік пен қылмыс деңгейі арасындағы орташа және оң корреляцияны көрсететін деректер бар. Тығыз орналасқан қалалық округтер мен халқы аз ауылдық жерлерде осы байланыстар бірдей ме деген сұрақ қызықтырып отыр. Округтегі халықтың жоғары және төмен білім деңгейіне қатысты байланыс бірдей ме?

3. Демограф туу көрсеткіші мен 76 елдегі әйелдердің білім деңгейі арасындағы байланысты зерттеді. Деректер айнымалылар арасындағы қалыпты, теріс байланысты көрсетеді: білім деңгейінің артуына қарай туу көрсеткіші төмендейді. Христиан және христиан емес елдерде байланыс бірдей ме? Байланыс әлауқаты бойыша барлық деңгейдегі елдерде бірдей ме?

Статистика негіздері кітабының 12 бөлімінде және алдыңғы дәрістерде атап өткендей, әлеуметтік зерттеулер оның табиғаты бойынша көп өлшемді және көп айнымалыларды бір мезгілде талдауды қамтиды. Көп өлшемді талдау үшін ең қуатты және кең қолданылатын статистикалық құралдардың кейбірі осы және келесі дәрістерде көрсетілген. Біз айнымалылардың арасындағы себеп-салдарлық байланыстарды талдау үшін пайдаланылатын әдістерді қарастырамыз және кез келген ғылымдағы маңызды бастама ретінде болжам жасаймыз.

Бұл әдістер Пирсонның r корреляциясына негізделген (12-тарауды қараңыз) және жоғары сапалы, нақты өлшенген интервалды-пропорционалды айнымалылар үшін қолданылады. Бұған дейін атап өткендей, мұндай деректер әлеуметтік зерттеулерде сирек кездеседі, сондықтан осы тарауда ұсынылған техникалар әдетте реттік айнымалыларға және жалған айнымалылардағы номиналды деңгейге қолданылады (12-тарау).

Алдымен, үшінші айнымалыны бақылай отырып, бивариатив қатынастарды зерттеуге мүмкіндік беретін жартылай корреляциялық талдауды зерттеу нысанына алайық. Бұл бірнеше себеп бойынша маңызды әдіс болып саналады, мысалы, үшінші айнымалы бивариатив қатынасқа ықпал етеді. 12-тарауда көрсетілгендей, себеп-салдарлық байланысты дәлелдеу үшін қажет тестілердің бірі.

Екіншіден, зерттеушілерге бірден көп тәуелсіз айнымалылардың тәуелді айнымалыға жеке және бірігіп әсер етуін бағалауға мүмкіндік беретін әдістерді, көптүрлі регрессиямен корреляцияны қарастырамыз. Сондай-ақ бұл әдістер бивариатив байланысқа үшінші айнымалылардың әсерін зерттеу үшін маңызды статистикалық құралдарды ұсынады.

Осы және келесі дәрістерде көрсетілгендей біз математикалық рәсімдер мен негізгі логиканы бейнелеу үшін үш айнымалы және аз көлемді жиынтықтарға байланысты жағдайларды зерттеуге назар аударамыз. Төрт немесе одан көп айнымалыларды қамтитын жағдайлардың көбеюі салыстырмалы түрде қарапайым, бірақ олардың есептеулері өте күрделі болады. Осылайша, оңайлатылған мысалдардың жұмыс істегеннен кейін, біз r Пирсон сияқты есептеулерді орындау үшін SPSS бағдарламасын пайдаланамыз. Сонымен қатар, сіз математикалық процедураларды түсінесіз, бірақ көптүрлі мәселелер SPSS пайдалануды талап етеді.

Жартылай корреляция

12-тарауда Пирсонның r корреляциясын бивариатив байланыстың күші мен бағытын өлшеу үшін пайдаландық. Мәселен, 12 отбасынан құралған іріктемеде күйеуінің үй шаруасына қосқан үлесі (тәуелді айнымалы немесе Y) мен балалар саны (тәуелсіз айнымалы немесе X) арасындағы өзара байланысты зерттедік. Қалыпты күштің оң байланысын ($r = 0,50$) анықтадық және ері балалар санының ұлғаюына



байланысты үй шаруасына көбірек араласатыны туралы қорытындыға келдік.

Әдетте зерттеушілер оны қалай жасайды деген заңды сұрақ туады. Бұл байланысты отбасының барлық түріне қолдану әділ ме? Мысалы, әртүрлі діндерде ер адамдар (күйеулері) әртүрлі әрекет ете ме? Саяси тұрғыдан консервативті ұстанымдағы ерлер либералды ұстанымдағы еркектерге қарағанда басқаша әрекет ете ме? Әлде жоғары білімді ерлер білімсіз еркектерден басқаша әрекет ете ме? Мұндай мәселелерді жартылай корреляция деп аталатын техника көмегімен шеше аламыз, онда дін, саяси идеология немесе білім сияқты үшінші айнымалы енгізілгенде, бивариатив байланыс қалай өзгеретінін байқаймыз. Үшінші айнымалылар көбіне-көп Z айнымалысы немесе бақылау айнымалысы деп аталады.

Жартылай корреляцияны табу үшін алдымен бивариатив байланысты корреляцияның r корреляциясы арқылы есептейміз (кейде нөлдік тәртіптегі корреляция деп те аталады), содан кейін жартылай корреляция коэффициенті (немесе бірінші тәртіптегі) есептеледі. Егер жартылай корреляция коэффициенті нөлдік тәртіптегі корреляция коэффициентінен өзгеше болса, үшінші айнымалы бивариатив байланысқа әсер етеді деген қорытынды жасаймыз. Мысалы, жоғары білімді еркектер білімі төмен еркектерге қарағанда балалар саны артқан сайын басқаша әрекет етсе, онда жартылай корреляция коэффициенті күші бойынша (мүмкін, бағыт бойынша) бивариатив корреляция коэффициентінен өзгеше болады.

Есептеулерді қарастырмас бұрын, жартылай және бивариатив коэффициенттер арасындағы байланысты және олар нені білдіретінін талдап алайық.

Байланыс түрлері

Тікелей байланыс. Мүмкіндіктердің бірден-бірі – жартылай корреляция коэффициенті мәні бойынша бивариатив коэффициентке тең. Мысалы, Пирсонның r корреляциясының $+0,50$ нөлдік тәртібімен салыстырғанда ерінің білімін бақылағаннан кейін $+0,49$ жартылай корреляция коэффициентін тапты деп алайық. Бұл дегеніміз үшінші айнымалы (ерінің білімі) балалардың саны мен ерінің үй шаруасымен айналысатын сағат саны арасындағы байланысқа әсер етпейді дегенді білдіреді. Қысқасы, ерлер біліміне қарамастан балалар саны артқан сайын жауапты болады. Бұл нәтиже X пен Y арасындағы тікелей немесе себеп-салдарлық байланыс бар (13.1-сызба) және үшінші айнымалының (Z) қатысы жоқ, сондықтан бұдан былай қарастырмауы тиіс деген қорытындыға әкеледі. Зерттеу жобасындағы келесі қадам: басқа да ықтимал бақылау айнымалыларымен қосымша тест жүргізілуі мүмкін (мысалы, зерттеуші отбасының дінін немесе этносын бақылай алады).

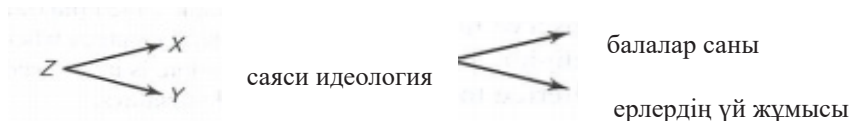
13.1-сызба. X және Y арасындағы тікелей байланыс.

$$X \rightarrow Y$$

Жалған және аралас байланыс. Екінші ықтимал нәтиже жартылай корреляция коэффициенті бивариатив корреляциядан әлдеқайда әлсіз болғанда, тіпті 0-ге дейін түсіп қалады. Бұл нәтиже екі түрлі байланысқа сәйкес келеді. Біріншісі жалған байланыс деп аталады: бақылау айнымалысы (Z) тәуелсіз (X) және тәуелді (Y) айнымалылардың себебі болады (13.2-сызбаны қараңыз). Бұл нәтиже X және Y мүлде байланысты емес дегенді білдіреді. Олар жалпы (Z) айнымалы болғандықтан ғана өзара байланысты болады. Z есепке алынған кезде, X және Y арасындағы айқын байланыс жоғалып кетеді.

Жалған байланыс дегеніміз не? Ата-аналарыңыздың саяси идеологиясын бақылап жүрсіз, осы орайда жартылай корреляция коэффициенті бивариатив Пирсонның r корреляциясына қарағанда әлсіз екені анықталды делік. Бұл балалардың саны ерінің үй шаруасына қосқан үлесін өзгертпейтінін көрсетеді (яғни X және Y арасындағы байланыс тікелей емес). Керісінше, саяси идеология басқа да айнымалылардың екеуінде де пайда болуы мүмкін: консервативті отбасылар дәстүрлі гендерлік рөлдік үлгілерді ұстанып (ерлер үй шаруасына аз үлес қосады), балалары көп болуы мүмкін.

13.2-сызба. X және Y арасында бөлінбейтін байланыс

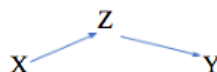




<p>Күнделікті өмірдегі статистика</p>	<p>Жалғандық</p> <p>Жалған байланыс тек ғылыми зерттеулердің ғана емес, күнделікті өмірдің де айнымас бөлігі. Кейде жаңалықтар туралы сөз болғанда немесе достармен әңгімелескенде «не нәрсенің неге себеп болғаны туралы» бұрыс ой түйеміз. Жалған байланысты білу қоғамдық өмірдің күрделілігі және себеп-салдардың өте бір нәзік ұғым екенін бағалауды арттырады. Жалғандық туралы ой қозғағанда кейбір жеңіл мысалдарға сүйенген жөн. Мұнда үш нұсқа бар, олардың әрқайсысы бір айнымалы (X) басқа айнымалыға (Y) себепші екенін көрсетеді. Әр нұсқадағы X пен Y ортақ себепшісін анықтай аласыз ба? Мысалы:</p> <ol style="list-style-type: none"> 1. Ауруханада адамдар жиі көз жұмады. Сондықтан ауруханалар (X) өлімнің себепшісі (Y). 2. Өрт сөндірушілердің саны неғұрлым көп болса, өрт кезінде соғұрлым көп жарақат алады. Сондықтан өрт сөндірушілер (X) өрт жарақатына себепші (Y) болады. 3. 1960 жылдардан бері халық әндерін көп тыңдайтын адамдар жыныстық импотенциядан зардап шегеді. Сондықтан халық музыкасы (X) импотенцияны тудырады (Y). <p>Жауаптар:</p> <ol style="list-style-type: none"> 1. Z=ауру және ауыр жарақат. Ауруы мен жарақаты асқынған адамдар ауруханаға түседі және сол аурулардан көз жұмады. 2. Z=өрттің дәрежесі. Алапат өртке көп өрт сөндіруші жұмылдырылады, сондықтан зардабы да көп. 3. Z=жасы. Ересек адамдар осы жанрдағы музыканы көбірек тыңдайды, сондықтан импотент болуы мүмкін.
--	---

Бұл паттерн (жартылай корреляция бивариатив корреляцияға қарағанда әлдеқайда әлсіз) айнымалылар арасындағы араласатын (intervening) байланысқа сәйкес келеді (13.3-сызбаны қараңыз). Бұл жағдайда X пен Y тікелей байланысты емес, бақылау айнымалысы арқылы қосылады. Z қайтадан бақыланғаннан кейін, ендеше X және Y арасындағы айқын байланыс жоғалып кетеді.

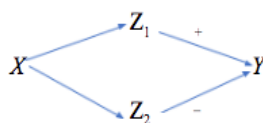
13.3-сызба. X және Y арасындағы өзара байланыс



Жалған және өзара қарым-қатынастардың арасындағы айырмашылықты қалай айта аламыз? Бұл ерекшелік статистикалық негізде жасалмайды: жалған және аралас қатынастар статистика тұрғысынан бірдей. Зерттеуші осы екі қатынас арасындағы айырмашылықтарды статистикалық негізде емес, уақыт тәртібіне (яғни алдымен пайда болған) немесе теориялық негізге байланысты ажырата алады.

Интеракция. Айнымалылар арасындағы үшінші ықтимал байланысты жартылай корреляциялық талдау арқылы анықтау мүмкін болмаса да атап өту керек. Интеракция деп аталатын бұл қатынас X және Y арасындағы байланыс Z-тің әртүрлі мәндері бойынша айтарлықтай өзгерген кезде пайда болады. Мысалы, әлеуметтік тапты бақылайтын болсақ, орта таптағы отбасылардың ерлері үй шаруасындағы үлесін арттырса, балалардың саны өседі, ал жұмысшы таптағы отбасыларда керісінше болады және осы үш айнымалы арасында интеракция болатыны туралы қорытынды жасауға болады. Басқаша айтқанда, 13.4-сызбада көрсетілгендей, X пен Y арасында Z категориясы үшін бір оң байланыс, ал басқа категориялар үшін теріс байланыс болады.

13.4-сызба. X, Y және Z арасындағы байланыс





Интеракция мысалы.

Мысалы, көше қылмыстарының себептерін зерттеуде кедейлік («кедейлеу аудандарда көшедегі қылмыс деңгейі жоғарырақ») пен жас мөлшерінің («жасөспірімдер мен жастардың пайыздық үлесі жоғары аудандарда көше қылмысы көбірек болады») арасында орташа күшті ассоциация бар екені байқалды, бірақ өте күшті байланыс айнымалылар біріктірілген кезде (кедейшілік пен жастары көп аудандарда көше қылмысы әлдеқайда жоғары деңгейде болады») пайда болады. Бұл интеракцияның айқын мысалы. Олардың аралас әсерлері тәуелді айнымалыға өте күшті ықпал етеді.

Жартылай корреляция коэффициентін есептеу және интерпретациялау

Жартылай корреляция коэффициентінің логикасы мен есептеуі кейбір жаңа ұғымдар мен терминологияны талап етеді. Алдымен терминологияны енгізіп, кейін формулаға көшеміз.

Терминология. Жартылай корреляция біреуден көп бивариативті қарым-қатынастармен жұмыс жасауды талап етеді және олардың арасындағы төменгі индекстерді ажыратып алуымыз қажет. Осылайша, r_{yx} Y және X айнымалылары арасындағы корреляция коэффициентіне сілтеме жасайды; r_{yz} Y және Z арасындағы корреляция коэффициентіне сілтеме жасайды; ал r_{xz} X пен Z арасындағы корреляция коэффициентіне сілтеме жасайды. Бивариативті қарым-қатынастар үшін есептелген корреляциялық коэффициенттер көбінесе нөл-тәртіптегі корреляциялар деп аталады.

Жартылай корреляция коэффициенттері немесе бірінші реттік жартылар - $r_{yx.z}$ символы болып таңбалады. Осылайша, $r_{yx.z}$ Төменгі бөліктегі нүктенің оң жағындағы айнымалы бақыланатын айнымалы болып табылады. Осылайша, $r_{yx.z}$ белгісі Z айнымалыны бақылаудағы X және Y айнымалылары арасындағы қарым-қатынастарды өлшейтін жартылай корреляция коэффициентіне қатысты болады. Бірінші реттік жартының формуласы

13,1-ші формула

$$r_{yx.z} = \frac{r_{yx} - (r_{yz})(r_{xz})}{\sqrt{1-r_{yz}^2} \sqrt{1-r_{xz}^2}}$$

Бұл формуланы шешпестен бұрын, алдымен айнымалылардың барлық мүмкін болатын жұптары (X және Y , X және Z және Y және X айнымалылары) арасындағы нөл-тәртіптегі коэффициенттерін есептеу алу керек.

Есептеу. Бірінші реттік жартыны есептеуді көрсету үшін 12 қос жалақылы отбасыларға оралайық, мұнда балалар саны (X) және ерінің үй шаруасына қосқан үлесі (Y) арасындағы байланысты қарастырамыз. Осы екі айнымалы арасындағы нөлдік тәртіптегі r қалыпты, яғни ($r_{yx} = 0,50$), оң байланысты (балалардың саны көбейген сайын ерлер үй шаруасына көбірек үлес қосуға бейім) екенін білдіреді. Айталық, зерттеуші ерлердің білім деңгейінің бивариатив байланысқа ықтимал әсерін зерттеуді қалайды. Бастапқы деректер (12.1-кестеде) және жаңа айнымалылардағы 12 отбасының балдары 13.1-кестеде келтірілген.

13,1-ші кесте Қос жалақы түсетін 12 отбасыдағы үш айнымалының балдары

Отбасы	Ерінің үй шаруасына араласуы (Y)	Балалар саны (X)	Ерінің білім алған жылдары (Z)
A	1	1	12
B	2	1	14
C	3	1	16
D	5	1	16
E	3	2	18
F	1	2	16
G	5	3	12
H	0	3	12
I	6	4	10
J	3	4	12
K	7	5	10



L	4	5	16
---	---	---	----

13,2-ші Нөл-тәртіптегі корреляция

	Ерінің үй шаруасына араласуы (Y)	Балалар саны (X)	Ерінің білім алған жылдары (Z)
Ерінің үй шаруасына араласуы (Y)	1,00	0,50	-0,30
Балалар саны (X)		1,00	-0,47
Ерінің білім алған жылдары (Z)			1,00

13.2-кестеде көрсетілген корреляциялық матрицада ұсынылған нөлдік тәртіптегі корреляциялар ерінің үй шаруасына қосқан үлесі балалардың санына ($r_{yx} = 0,50$) қатысты оң екенін білдіреді, ал жоғары білімді ерлердің үй шаруасына қосқан үлесі аз ($r_{yz} = -0,30$) және олардың отбасында балалар саны аз ($r_{xz} = -0,47$).

Ерінің білім алған жылдары (Z) балалар саны (X) мен оның үй шаруасына араласуы (Y) арасындағы байланысқа ықпал етті ме? 13.1-формулаға нөлдік тәртіптегі корреляцияларды саламыз

$$r_{yx.z} = \frac{r_{yx} - (r_{yz})(r_{xz})}{\sqrt{1-r_{yz}^2} \sqrt{1-r_{xz}^2}}$$

$$r_{yx.z} = \frac{(0,50) - (-0,30)(-0,47)}{\sqrt{1-(-0,30)^2} \sqrt{1-(-0,47)^2}}$$

$$r_{yx.z} = \frac{(0,50) - (0,14)}{\sqrt{1-0,09} \sqrt{1-0,22}}$$

$$r_{yx.z} = \frac{0,36}{\sqrt{0,91} \sqrt{0,78}}$$

$$r_{yx.z} = \frac{0,36}{\sqrt{0,95} \sqrt{0,88}}$$

$$r_{yx.z} = \frac{0,36}{0,84}$$

$$r_{yx.z} = 0,43$$

Интерпретация. Бірінші реттік жарты ($r_{yx.z} = 0,43$) ерінің білім алған жылдарын бақылаудағы (Z) ерінің үй шаруасына араласуы (Y) мен балалар саны (X) арасындағы байланыстың күшін өлшейді. Бұл нөлдік тәртіптегі коэффициентке ($r = 0,50$) қарағанда төменірек, бірақ екі мәннің айырмашылығы аса үлкен емес. Бұл нәтиже X және Y айнымалылары арасындағы тікелей байланысқа нұсқайды. Яғни, ерінің біліміне бақылау жасағанда, ерінің үй шаруасы мен балалардың саны арасындағы статистикалық байланыс айтарлықтай өзгермейді. Білім деңгейіне қарамастан, ерінің үй шаруасымен айналысуының уақыты балалардың санының көбеюімен қоса артуда.

Статистикалық талдаудағы келесі қадам – басқа бақылау айнымалысын таңдау. Бивариатив байланыс өз күшін үшінші бір айнымалылар үшін бақылау элементтері бойынша сақтаған сайын (Z-тер), X және Y арасындағы тікелей байланыстың бар екенінің дәлелдері де күшейеді.

Қорыта келе, қосымша нәтиже туралы айта кеткен жөн, онда жартылай корреляция коэффициенті нөлдік тәртіптегі коэффициентке қарағанда үлкенірек. Бұл нәтиже тәуелсіз ретінде қабылданған айнымалы мен бақылаушы айнымалы мәндердің әрқайсысы тәуелді айнымалыға жеке-жеке ықпал ететін және бір-бірімен корреляцияланбайтын себеп-салдарлық модельге сәйкес келеді. Бұл байланыс 13.5-сызбада бейнеленген. X пен Z арасында нұсқардың болмауы оларда өзара байланыстың жоқ екенін білдіреді.

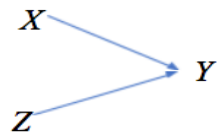
Бұл үлгі X және Z тәуелсіз айнымалылар ретінде қарастырылуы керек, ал статистикалық талдаудың келесі қадамы осы дәрісте ұсынылатын көптүрлі корреляция мен регрессияның болуы мүмкін дегенді білдіреді. (Жартылай корреляция коэффициенттерін есептеу және интерпретациялауға машықтану үшін 13.1a, b, 13.4a және 13.5 жаттығуларын қараңыз).

13.5-сызба. Үш айнымалы арасында мүмкін болатын себеп-салдарлы байланыс



Кітап: Статистика негіздері

Дәріс: Жартылай корреляция және көптүрлі регрессия және корреляция



Түйін

Жартылай корреляцияда біз, үшінші айнымалы мәнді бақылайтын кезде бивариативті қарым-қатынастармен не болатынын байқаймыз. Бұл әдіс арқылы біз X және Y арасындағы тікелей, жалған немесе интервенциялы байланыстарды табуға болады.