24-дәріс

ҚАЗАҚСТАННЫҢ
АШЫҚ
УНИВЕРСИТЕТІ

СТАТИСТИКА НЕГІЗДЕРІ

Көптүрлі регрессия: тәуелді айнымалыны болжау

Дәрісті тыңдағаннан кейін

1. Жартылай көлбеумен ең кіші квадраттардың көптүрлі регрессия теңдеуін табасыз және түсіндіресіз.

Естеріңізге сала кетейін, алдыңғы, 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 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 берлген. Тығыз орналасқан қалалық округтерде және аз елді ауылдық мекендерде осы қатынастар бірдей ме деген сұрақ туындайды? Округтердегі жоғары және төмен білім деңгейлерімен байланысты 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?

Алдыңғы дәрістерде ең кіші квадраттар регрессия сызығы екі интервалды-арақашықтық айнымалар арасындағы жалпы сызықтық қарым-қатынасты сипаттау әдісі ретінде енгізілді және Х-дегі баллдардан Y-те болжанатын балдарды белгіледі. Бұл сызық бивариативті қарым-қатынасты түйіндеудің ең ұтымдысы болды және келесі формула бойынша анықталды:

13.2 формула

$$Y = a + bX$$

Ең кіші квадраттар регрессия сызығы (теориялық тұрғыдан) тәуелсіз айнымалылардың кез келген санын қосу үшін өзгертілуі мүмкін. Бұл әдіс көптүрлі регрессия деп аталады. Түсіндіруді жеңілдету үшін біз екі тәуелсіз айнымалыларға қатысты кейстерге назар аударамыз. Ең кіші квадраттардың көптүрлі регрессиялық теңдеуі болады:

13.3 формула

$$Y = a + b_1X_1 + b_2X_2$$

онда b1 = бірінші тәуелсіз айнымалы (X1) және Y арасындағы сызықтық қатынастың жартылай көлбеуі
b2 = екінші тәуелсіз айнымалы (X2) және Y арасындағы сызықтық қатынастың жартылай көлбеуі

Осы формулаға кейбір жаңа белгілер мен кейбір жаңа ұғымдар енгізілген. Біріншіден, тәуелді айнымалы мән әліде Y ретінде белгіленеді, ал тәуелсіз айнымалылар төменгі индекстермен ерекшеленіп жазылады. Осылайша, X1 бірінші тәуелсіз айнымалы және X2 екіншіні анықтайды. Көлбеудің белгісі (b) өзі ассоциацияланған тәуелсіз айнымалы мәнді анықтау үшін төменгі индекспен жазылады.

Жартылай көлбеулер

Көптүрлі және бивариативті регрессия теңдеулерінің арасындағы негізгі айырмашылық көлбеулерге (b-ларға) қатысты. Көптүрлі регрессия жағдайында b сандары жартылай көлбеулер деп аталады және олар теңдеуде басқа тәуелсіз айнымалылардың әсерлерін бақылай отырып, тәуелсіз айнымалыдағы бірліктің өзгеруіне арналған Y-тегі өзгеріс шамасын көрсетеді. Жартылай көлбеулер былайша жартылай корреляция коэффициенттеріне ұқсас болдаы және Y бойынша тәуелсіз айнымалымен ассоциацияланған тікелей әсерді көрсетеді.

Жартылай көлбеулерді есептеу. Тәуелсіз айнымалылар үшін жартылай көлбеулер 13,4 және 13,5: формулалары бойынша анықталады.

13,4-ші формула

$$b_1 = \left(\frac{s_y}{s_1}\right)\left(\frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2}\right)$$

13,5-ші формула

$$b_2 = \left(\frac{s_y}{s_2}\right)\left(\frac{r_{y2} - r_{y1} r_{12}}{1 - r_{12}^2}\right)$$

мұнда, b_1 = Y бойынша жартылай көлбеу X_1
b_2 = Y бойынша жартылай көлбеу X_2
s_1 = Y бойынша стандартты девиация (X_1)
s_2 = Y бойынша стандартты девиация (X_2)
r_{y1} = Y пен X_1 арасындағы бивариативті корреляция
r_{y2} = Y пен X_2 арасындағы бивариативті корреляция
r_{12} = X_1 мен X_2 арасындағы бивариативті корреляция

Жартылай көлбеулерді есептеуді көрсету үшін, біз балалардың саны (X1) мен күйеуінің білімінің (X2) күйеуінің үй жұмысына қосқан үлесіне біріккен әсерін бағалаймыз. Барлық тиісті ақпаратты 13,1-кесте бойынша есептеуге болады және мұнда:

Күйеудің үй шаруасын жасауы	**Балалар саны**	**Күйеудің білім алған жылдары**
$\bar{Y}$ = 3,3	$\bar{X}$ = 3,3	$\bar{X}_2$= 13,7
s_y = 2,1	s_1 = 2,1	s_2 = 2,1
Нөл-тәртіптегі корреляциялар		
	r_{y1} = 0,50 r_{y2} = - 0,30 r_{12} = - 0,47	

Бірінші тәуелсіз айнымалының (X_1) жартылай көлбеу болады:

$$b_1 = \left(\frac{s_y}{s_1}\right)\left(\frac{r_{y1} - r_{y2} r_{12}}{1 - r_{12}^2}\right)$$

$$b_1 = \left(\frac{2.1}{1.5}\right)\left(\frac{0.50 - (-0.30)(-0.47)}{1 - (-0.47)^2}\right)$$

$$b_1 = (1.4)\left(\frac{0.50 - (0.14)}{1 - 0.22}\right)$$

$$b_1 = (1.4)\left(\frac{0.36}{0.78}\right)$$

$$b_1 = (1.4)(0.46)$$

$$b_1 = 0.64$$

Екінші тәуелсіз айнымалының (X2) жартылай көлбеу болады:

$$b_1 = \left(\frac{s_y}{s_2}\right)\left(\frac{r_{y2} - r_{y1} r_{12}}{1 - r_{12}^2}\right)$$

$$b_1 = \left(\frac{2.1}{2.6}\right)\left(\frac{0.30 - (+0.50)(-0.47)}{1 - (-0.47)^2}\right)$$

$$b_1 = (0.81)\left(\frac{-0.30 - (0.24)}{1 - 0.22}\right)$$

$$b_1 = (0.81)\left(\frac{-0.06}{0.78}\right)$$

$$b_1 = (0.81)(-0.08)$$

$$b_2 = -0.07$$

Жартылай көлбеуді есептеудің орындалуын бекіту үшін келесі қадамдарды орындаймыз.

Рет-ретімен орында	Жартылай көлбеуді есептеу
Бұл процедуралар екі тәуелсіз айнымалы мен бір тәуелді айнымалы болған кезде қолданылады. Күрделірек жағдайларды есептеу үшін SPSS сияқты компьютерленген статистикалық бағдарламаны пайдаланыңыз.	
Реті	**Орындалуы** 13.4-формула арқылы бірінші тәуелсіз айнымалымен байланысты жартылай көлбеуді шешу үшін:
1.	s_y мәнді s_1 мәнге бөліңіз.
2.	r_{y2} мәнді r_{12} мәніне көбейтіңіз.
3.	2 қадамда табылған мәнді r_{y1} мәнінен шегеріңіз.
4.	r_{12} мәнін квадраттаңыз.
5.	4-қадамда табылған санды (r^2_{12}) 1 санынан шегеріңіз.
6.	3-қадамда шыққан мәнді 5-қадамда табылған мәнге бөліңіз.
7.	6-қадамда шыққан мәнді 1-қадамда есептелген мәнге көбейтіңіз. Бұл мән – бірінші тәуелсіз айнымалымен өзара байланысты жартылай көлбеу.
	13.5-формула арқылы екінші тәуелсіз айнымалымен байланысты жартылай көлбеуді шешу үшін:
1.	sy мәнді s_2 мәнге бөліңіз.
2.	r_{y1} мәнді r_{12} мәнге көбейтіңіз
3.	2 қадамда табылған мәнді r_{y2} мәнінен шегеріңіз.
4.	r_{12} мәнін квадраттаңыз.
5.	4-қадамда шыққан санды 1 санынан шегеріңіз.
6.	3-қадамда есептелген мәнді 5-қадамда табылған мәнге бөліңіз.
7.	6-қадамда шыққан мәнді 1-қадамда есептелген мәнге көбейтіңіз. Бұл мән – екінші тәуелсіз айнымалымен өзара байланысты жартылай көлбеу.
	Жартылай көлбеуді интерпретациялау Жартылай көлбеу мәні басқа тәуелсіз айнымалылардың әсерлерін бақылай отырып, тәуелсіз айнымалыдағы бірлік өзгергенде Y мәні өзгереді.

Y интерсепциясын табу. Енді жартылай көлбеулер тәуелсіз айнымалылардың екеуі үшін де анықталған, Y интерсепцияларын (a) табуға болады. Айта кету керек, a тәуелді айнымалы мәнінің арифметикалық орта мәнінен ($\bar{Y}$ ретінде бейнеленген) және екі тәуелсіз айнымалылардан $\bar{X}_1$ және $\bar{X}_2$) есептеледі.

13.6 формула

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$$

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$$

$$a = 3{,}3 - (0{,}64)(6{,}7) - (-0{,}07)(13{,}7)$$

$$a = 3{,}3 - 1{,}7 - (-1{,}0)$$

$$a = 2{,}6$$

Y интерсепциясын есептеудің орындалуын бекіту үшін келесі қадамдарды орындаймыз.

Рет-ретімен орында	Y интерсептін есептеу
Реті	**Орындалуы** 13.6-формула арқылы Y интерсептін табу:
1.	X_2 мәнінің арифметикалық ортасын b_2 мәнге көбейтіңіз.
2.	X_1 мәнінің арифметикалық ортасын b_1 мәнге көбейтіңіз.
3.	1-қадамда табылған мәнді 2-қадамда табылған мәннен шегеріңіз.
4.	3-қадамда табылған мәнді Y мәнінің арифметикалық ортасынан шегеріңіз. Алынған нәтиже – а мәні Y интерсепті.

Ең кіші квадраттар көптүрлі регрессия сызығына болжам жасау Y'.

Біздің мысалға қатысты мәселе үшін ең кіші квадраттар көптүрлі регрессия теңдеуінің толық түрі болады:

$Y = a + b_1X_1 + b_2X_2$

$Y = 2,6 + (0,64)X_1 + (– 0,07)X_2$

Бивариатив регрессия сызығы сияқты бұл формула тәуелді айнымалыдағы балдарды тәуелсіз айнымалылардағы балдармен болжау үшін пайдаланылады. Мысалы, күйеуі 11 жылдық мектепті аяқтаған (X2 = 11) төрт балалы отбасында (X1 = 4) күйеуінің үй шаруасына қатысты неғұрлым жақсы болжау (Y') қалай болар еді? Бұл мәндерді ең кіші квадрат теңдеуіне салсақ

Y' = 2,6 + (0,64)(4)+(–0,07)(11)

Y' = 2,6 + 2,6 – 0,8

Y' = 4,4

Біздің болжауымызша, ері үй шаруасына аптасына 4,4 сағат жұмсайды. Әрине, бұл болжам «дәлелденген шамалау» тәрізді дәлме-дәл солай болуы мүмкін емес. Алайда болжаудың кез келген басқа әдісін қолданғанға қарағанда, ең кіші квадраттар сызығын (оған қоса осылайша, тәуелсіз айнымалы мәндерден алынған ақпаратты) пайдалану арқылы болжауда қателерді азайтамыз (сөзсіз, мұнда тәуелсіз және тәуелді айнымалылар арасында сызықтық ассоциация бар деп қарастырамыз). (Тарау соңындағы барлық жаттығулар стандартты емес көлбеулер мен Y интерсептін табуға мүмкіндік береді, сондай-ақ Y балдарын болжау мүмкіндігін ұсынады.

Ең кіші квадраттар көптүрлі регрессия сызығын Y балдарына болжам жасау үшін пайдалануды бекіту үшін келесі қадамдарды орындаймыз.

Рет-ретімен орында	Ең кіші квадраттар көптүрлі регрессия сызығын Y балдарына болжам жасау үшін пайдалану
Реті	**Орындалуы**
1.	X1 мәнін таңдаңыз. Осы мәнді b1 мәнге көбейтіңіз
2.	X2 мәнін таңдаңыз. Осы мәнді b2 мәнге көбейтіңіз.
3.	1 және 2 қадамдарда табылған мәнді а мәніне, Y интерсептке қосыңыз. Алынған нәтиже
4.	– Y бойынша болжанған мән.

Көптүрлі регрессия: тәуелсіз айнымалылардың әсерін бағалау

Ең кіші квадраттардың көптүрлі регрессиялық теңдеуі (13.3-формула) тәуелсіз айнымалылардың жекешеленген әсерлерін оқшаулау үшін және тәуелді айнымалының балдарын болжау үшін пайдаланылады. Алайда көп жағдайда осы формуланы пайдалану тәуелсіз айнымалылардың салыстырмалы түрдегі маңыздылығын анықтау үшін ыңғайсыз, әсіресе тәуелсіз айнымалы өлшем бірліктері тұрғысынан (мысалы, балалардың саны және білім алу жылдары) ерекшеленсе қолайсыз болады. Өлшем бірліктері әртүрлі болған кезде жартылай көлбеулерді салыстыру міндетті түрде қандай тәуелсіз айнымалының ең күшті әсері бар екенін, сондықтан ол соншалықты маңызды болып табылады дегенді білдірмейді. Өлшем бірліктерінде айырмашылығы бар айнымалылардың жартылай көлбеулерін салыстыруға келмейді, айырмашылығы жер мен көктей.

Біз тәуелсіз айнымалылардың әсерлерін теңдеулердегі барлық айнымалыларды жалпы шкалаға айналдыру арқылы салыстыра аламыз. Сондықтан жартылай көлбеулердің мәндерінде, өлшем бірліктерінде тек қана айырмашылықтардың функциясы болып табылатын өзгерістерді жоя аламыз. Мысалы, барлық айнымалылардың балдарын Z шамаға ауыстыру арқылы барлық үлестірімдерді стандарттай аламыз. Әрбір үлестірімнің арифметикалық ортасы 0 болады, стандартты девиациясы 1-ге тең (5-тарауды қараңыз), сондықтан тәуелсіз айнымалылар арасындағы салыстыру әлдеқайда маңыздырақ.

Стандартталған регрессия коэффициенттерін есептеу

Айнымалыларды қалыпты қисыққа стандарттау үшін біз барлық баллдарды баламалы Z балдарға түрлендіріп, содан кейін көлбеулер мен Y интерсепцияны қайта есептей аламыз. Бұл біраз жұмысты талап етеді; бірақ бақытымызға қарай, стандартталған баллдардың көлбеулерін есептеу тікелей қол жетімді. Бета-салмақ деп аталатын бұл стандартталған жартылай көлбеулердің белгіленуі b* болып табылады.

Бета-салмақтар. Бета салмақтары барлық басқа тәуелсіз айнымалылардың әсерін бақылау кезінде, әрбір тәуелсіз айнымалының стандартталған баллдарында бір бірліктік өзгеру үшін Y стандартталған баллдарындағы өзгерістер мөлшерін көрсетеді.

Бета-салмақтар үшін формулалар және есептеу. Бізде екі тәуелсіз айнымалы болған кезде, олардың әрқайсысына арналған бета-салмақ 13,7-ші және 13,8-ші формулалар бойынша анықталады:

13.7 формула

$$b_1^* = b_1\left(\frac{s_1}{s_y}\right)$$

13.7 формула

$$b_1^* = b_2\left(\frac{s_2}{s_y}\right)$$

Енді біз жиынтықтағы мәселе үшін екі тәуелсіз айнымалылардың қайсысы тәуелді айнымалылардың қайсысына күшті әсер ететінін көру үшін бета-салмақты есептей аламыз. Бірінші тәуелсіз айнымалы үшін балалардың саны (X_1):

$$b_1^* = b_1\left(\frac{s_1}{s_y}\right)$$

$$b_1^* = (0.64)\left(\frac{1.5}{2.1}\right)$$

$$b_1^* = (0.64)(0.71)$$

$$b_1^* = 0.45$$

Екінші тәуелсіз айнымалы үшін балалардың саны (X2):

$$b_2^* = b_2\left(\frac{s_2}{s_y}\right)$$

$$b_1^* = (-0.07)\left(\frac{2.6}{2.1}\right)$$

$$b_1^* = (-0.07)(1.24)$$

$$b_1^* = -0.09$$

Бета-салмақтардың мәндерін салыстыру арқылы біз балалардың саны (X_1) күйеудің үй шаруасын жасауына (Y) күйеуінің білім деңгейіне (X_2) қарағанда әлдеқайда күштірек әсер ететінін көреміз. Сонымен қатар, бірінші тәуелсіз айнымалы мәндегі таза әсер (ерлердің білім деңгейінің әсерін бақылағаннан кейін) оң болады, ал екінші тәуелсіз айнымалының таза әсері теріс болады.

Стандартталған ең кіші квадраттардың регрессия сызығы. Стандартталған баллдарды пайдалана отырып, ең кіші квадраттардың регрессия теңдеуінің жазылуы

13.9- формула

$$Z_y = a_z + b_1^* Z_1 + b_2^* Z_2$$

Мұнда: Z барлық баллдар қалыпты қисыққа стандартталған дегенді білдіреді.

Стандартталған регрессия теңдеуін Y интерсепциясы үшін терминді алып тастау арқылы одан әрі жеңілдетуге болады, себебі бұл термин баллдар стандартталған кезде әрдайым 0 болады. Бұл мән регрессия сызығы Y осін қиып, барлық тәуелсіз айнымалы мәндер 0-ге тең болғанда Y мәнінің арифметикалық ортасына тең болған нүкте болып табылады. Бұл қарым-қатынасты 13,6-шы формуладағы барлық тәуелсіз айнымалылар үшін 0-ді қойғанда көруге болады:

$$a = -b_1 \bar{\bar{X}}_1 - b_2 \bar{\bar{X}}_2$$

$$a = -b_1(0) - b_2(0)$$

$$a = \bar{Y}$$

Кез-келген стандартты үлестірімнің арифметикалық орта мәні 0 болғандықтан, стандартталған Y баллдарының арифметикалық орта мәні 0 болады, Y интерсепцияда 0-ге тең болады (a = = 0). Осылайша, 13.9-шы формула келесідей жеңілдетіледі

13.10- формула

$$Z_y = b_1^* Z_1 + b_2^* Z_2$$

Біздің жиынтықтағы мәселе үшін регрессия теңдеуі бета-салмақпен көрсету бойынша болады

$$Z_y = (0{,}45)\ Z_1 + (-0{,}09)\ Z_2$$

және бірінші тәуелсіз айнымалының, екінші тәуелсіз айнымалыға қарағанда, Y бойынша әлдеқайда күшті тікелей әсер етуі анық көрінеді.

Байқағанымыздай, көптүрлі регрессиялық талдау зерттеушіге екі немесе одан да көп тәуелсіз айнымалы мен тәуелді айнымалы арасындағы сызықтық байланысты түйіндеуге мүмкіндік береді. Стандартталмаған регрессиялық теңдеу (13.2-формула) айнымалы мәндердің түпнұсқа бірліктеріндегі тәуелсіз айнымалы мәндерден болжанатын Y мәндерін береді. Стандартталған регрессиялық теңдеу (13.10-формула) зерттеушіге бета-салмақты салыстыру арқылы түрлі тәуелсіз айнымалылардың салыстырмалы маңыздылығын оңай бағалауға жол ашады. (Тарау соңындағы барлық жаттығулар бета-салмақты есептеу және түсіндіру мүмкіндігін ұсынады.)

Бета-салмақтарды (b*) есептеудің орындалуын бекіту үшін келесі қадамдарды орындаймыз.

Рет-ретімен орында	**Бета-салмақты есептеу (b*)**
Бұл процедуралар екі тәуелсіз айнымалы мен бір тәуелді айнымалы болған кезде қолданылады. Күрделірек жағдайларды есептеу үшін SPSS сияқты компьютерленген статистикалық бағдарламасын пайдаланыңыз.	
Реті	**Орындалуы**
	13.7-формула арқылы бірінші тәуелсіз айнымалымен байланысты бета-салмақты есептеу үшін:
1.	s_1 мәнді s_y мәнге бөліңіз.
2.	1 қадамда табылған мәнді бірінші тәуелсіз айнымалының жартылай көлбеуіне (b_1) көбейтіңіз. Бұл мән – бірінші тәуелсіз айнымалымен өзара байланысты бета-салмақ.
	13.8-формула арқылы екінші тәуелсіз айнымалымен өзара байланысты бета-салмақты есептеу үшін:
1.	s_2 мәнді s_y мәнге бөліңіз.
2.	1 қадамда табылған мәнді екінші тәуелсіз айнымалының жартылай көлбеуіне (b_2) көбейтіңіз. Бұл мән – екінші тәуелсіз айнымалымен өзара байланысты бета-салмақ.

Бета-салмақтарды интерпретациялау (b*)

Бета салмақ (немесе стандартталған жартылай көлбеу) барлық басқа тәуелсіз айнымалылардың әсерін бақылау кезінде, әрбір тәуелсіз айнымалының стандартталған баллдарында бір бірліктік өзгеру үшін Y стандартталған баллдарындағы өзгерістер мөлшерін көрсетеді.

Түйін

Бивариативті регрессиялық теңдеуді жартылай көлбеулермен қоса пайдалана отырып, тәуелді айнымалы мәндерді бірнеше тәуелсіз айнымалы бойынша баллдардан болжай аламыз. Тәуелсіз айнымалылардың салыстырмалы түрдегі маңыздылығын стандартталған жартылай көлбеумен немесе бета-салмақпен ажырата аламыз.