



СТАТИСТИКА НЕГІЗДЕРІ

Хи-квадратқа негізделген негізделген
ассоциация өлшемдері: Φ және V
Cramer



Ассоциация және статистикалық мәнділік (significance)

Естеріңізге сала кетейік, ассоциация мен мәнділік екі түрлі нәрсе болып табылады, ал қарым-қатынас үшін күшті, бірақ елеусіз болу, елеулі, бірақ әлсіз және т.б. болуы мүмкін.

Ассоциация мен себеп-салдарлық. Еске сала кетейік, ассоциация өз бетінше себеп-салдарлы байланысты дәлелдей алмайды.

Осы дәрісті тыңдағаннан кейін

1. «Үздіксіз» реттік айнымалылар үшін қолданылатын ассоциация өлшемі – Спирменның rho корреляция коэффициентін есептейсіз және интерпретациялайсыз.

2. Бивариатив кестедегі бағанның пайыздық көрсеткіштерін есептейсіз және ассоциацияның номиналды және реттік өлшемдерін алу үшін SPSS бағдарламасын пайдаланасыз.

[1] Осы материалда, қысқа болу үшін, көбінесе тәуелсіз айнымалыны X деп және тәуелді айнымалыны Y деп сілтеме жасаймыз.

[2] Осы және келесі дәрісте енгізілетін ассоциацияның әрбір өлшемі «мінсіз ассоциация» туралы өз анықтамасын қамтиды, сондықтан бұл анықтамалар статистиканың нақты логикасына және математикасына байланысты бір балл өзгеше болады. Яғни, бірдей кестеден есептелген әртүрлі өлшемдер үшін, кейбір өлшемдер мінсіз қарым-қатынастарды көрсетуі мүмкін, ал басқалары олай жасай алмайды. Біз бұл вариацияларға мінсіз ассоциацияның математикалық анықтамаларымен тиісті уақытта назар аударамыз.

[3] Номиналды деңгейде өлшенген айнымалының баллдары сандық тәртіпке ие емес (анықтама бойынша). Демек, номиналды деңгейдегі айнымалы мәндерді қамтитын ассоциациялар, олар үлгі болуы мүмкін болғаныменде, бағытқа ие бола алмайды.

Бағанның пайыздық көрсеткіштері бивариативті ассоциациялар туралы маңызды ақпаратты береді және әрқашан есептеліп, талдануы керек. Дегенмен, олар ыңғайсыз және қолайсыз болуы мүмкін, әсіресе ірі кестелер үшін. Керісінше, ассоциация өлшемдері бивариативті ассоциациялардың жалпы күшін (және реттік деңгейдегі айнымалылардың бағытын) бір санмен жиынтықтайды, бұл түсіндіру мен талқылаудың әлдеқайда ықшамды және ыңғайлы форматы болып табылады.

Ассоциацияның көптеген өлшемдері бар, бірақ біз ең көп қолданылатындардың біріне назар аударамыз. Біз бұл статистиканы, ең алдымен, ең қолайлы өлшем деңгейімен, ең алдымен номиналды айнымалыларға сәйкес келетін өлшемдерді қамтитын, содан кейін реттік деңгейдегі айнымалыларға арналған өлшемдерді қолданамыз.

Номиналды айнымалылар үшін ассоциация өлшемдері

Номиналды деңгейдегі айнымалы мәндердің екі түрі бар: бірі хи-квадратқа негізделген, екіншісі логика бойынша «қателіктердің пропорционалды төмендеуі» деп аталатынға немесе «PRE». Бұл екі категорияны бөлек қарастырамыз.

Хи-квадратқа негізделген негізделген ассоциация өлшемдері: Phi және V Cramer.

Хи-квадратқа негізделген бірнеше ассоциация өлшемдері бар, бірақ біз тек екеуін қарастырамыз. Бірінші, phi (ϕ) деп аталатын, 2 x 2 кестелеріне сәйкес келеді (яғни екі жолмен және екі бағанмен жасалған кестелер) және ол есептеуге оңай. 11,1-ші формулада көрсетілгендей, phi қарапайым ғана - хи-квадратты N-ге бөлгендегі шаршы түбір.

11,1-ші формула

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

2 x 2 -ден асатын кестелер үшін, phi -дің жоғарғы шегі 1,00-ден асып кетуі мүмкін, сондықтан бұл статистиканы түсіндіруге қиын. Үлкенірек кестелер үшін, көбіне V Cramer деп аталатын статистиканың жалпы формасы қолданылады.



11,2-ші формула

$$V = \sqrt{\frac{\chi^2}{N(\min r - 1, c - 1)}}$$

мұнда $(\min r - 1, c - 1)$ = жолдар санының минимум мәнінен минус 1 немесе жоғарғы шегі (r) немесе бағандардың саны (c) минус 1.

V есептеу үшін, «Рет-ретімен» боксындағы нұсқауларды орындаңыз. Кез келген мөлшердегі кесте үшін V Cramer жоғарғы шегі 1,00 болады, ал кестеде екі жол немесе екі баған болса, онда phi мәнімен тең болады.

Phi және V Cramer екеуі де екі айнымалы арасындағы ассоциацияның күшін өлшейтін индекстер ретінде түсіндірілуі мүмкін. 11,8-кестеде ассоциацияның номиналды деңгейдегі өлшемдерінің күшін түсіндіру бойынша нұсқаулар берілген. Мұнда да, сипаттама терминдері тек жалпы басшылық ретінде ғана қарастырылады.

11,8-кесте Ассоциацияның өлшемдерінің номиналды деңгей үшін қарым-қатынас күшін түсіндіру бойынша нұсқаулар

Ассоциация өлшемдері	Күші
Егер мән болса 0,00 мен 0,10 арасында	Қарым-қатынастың күші болады
0,11 мен 0,30 арасында	Әлсіз
0,30 - дан көп болса	Қалыпты Күшті

11,9-кесте Клубқа мүшелігі бойынша академиялық жетістік (GPA)

Клубқа мүшелігі				
Жетістік	Грек тілі	Басқа	Ешқандай	Барлығы
Төмен	4 (17,4%)	4 (15,4%)	17 (65,4%)	25 (33,3%)
Орташа	15 (65,2%)	6 (23,1%)	4 (15,4%)	25 (33,3%)
Жоғары	4 (17,4%)	16 (61,5%)	5 (19,2%)	25 (33,3%)
Барлығы	23 (100,0%)	26 (100,0%)	26 (100,0%)	75 (99,9%)

V есептеуін көрсету үшін, әртүрлі типтегі студенттік ұйымдарына мүшелік пен колледж студенттерінің академиялық жетістігі арасындағы қарым-қатынастарды көрсететін 11,9-кестені қараңыз. Осы кесте бойынша алынған хи-квадрат 32,14, бұл мән 0,05 деңгейінде елеулі. Кестеде жолдар мен бағандардың саны бірдей, сондықтан біз бөлшек бөлімінде әрқайсысын $(r - 1)$ немесе $(c - 1)$ қолданамыз. Кез-келген жағдайда бөлшек бөлімінде мәні $(3 - 1)$ - ге немесе 2-ге көбейтілген N болады. V Cramer

$$V = \sqrt{\frac{\chi^2}{N(\min r - 1, c - 1)}}$$

$$V = \sqrt{\frac{32,14}{(75)(2)}}$$

$$V = \sqrt{\frac{32,14}{150}}$$

$$V = \sqrt{0,21}$$

$$V = 0,46$$

0,46 мәні V Cramer клубқа мүшелік пен академиялық жетістік арасындағы күшті ассоциацияны білдіреді. Бағандар пайыздарынан көріп отырғанымыздай, академиялық жетістіктері бойынша бауырластар мен сорор клубтарының мүшелері қалыпты, басқа ұйымдардың мүшелері жоғары, ал мүше еместер төмен деңгейде болады.



V Cramer-ді есептеудің орындалуын бекіту үшін келесі қадамдарды орындаймыз.

Рет-ретімен орында	Крамердің V коэффициентін есептеу
Реті	Орындалуы
1.	Кестедегі жолдардың (r) және бағандардың санын (c) табыңыз. ($\min r - 1, c - 1$) мәнін табу үшін осы екі санның ең азынан 1-ді алып тастаңыз.
2.	1-қадамда табылған мәнді N-ге көбейтіңіз.
3.	Хи-квадраттың мәнін 2-қадамда табылған мәнге бөліңіз.
4.	3-қадамда табылған санды түбір астынан шығарыңыз. Алынған мән V болады.

V Cramer-ді интерпретациялау

11,8-ші кестені жалпы басшылыққа алыңыз.

Қарым-қатынас үлгісін анықтау үшін бағанның пайыздық көрсеткішін пайдаланыңыз.

Phi мен V Cramer-дің шектеулері

Бұл өлшемдердің бір шектеуі - бұл қарым-қатынастың күші туралы жалпы анықтамалар ғана. Әрине, бұл өлшем бойынша 0,00-ге жақын қарым-қатынас әлсіз, ал 1,00-ге жақын қарым-қатынас неғұрлым күшті болады. 0,00 және 1,00 аралығындағы мәндерді 11,8-кестеде көрсетілген терминдерге сәйкес әлсіз, қалыпты немесе күшті деп сипаттауға болады, бірақ олардың тікелей немесе мазмұнды түсіндірмесі жоқ. Екінші жағынан, phi және V есептеу оңай (хи-квадраттың шамасы алынғаннан кейін) және жалпы ассоциацияның маңыздылығын көрсететін көрсеткіш болып табылады.

Әдебиетте кейде ассоциация өлшемдерінің хи-квадратқа негізделген басқа екі түрі T2 және C қарастырылады (күтпеген кездейсоқ коэффициент). Бұл екі өлшем де елеулі шектеулерге ие. T2 жоғарғы шегі 1,00 тек жолдар мен бағандар саны тең болатын кестелер үшін, ал C-ның жоғарғы шегі кестенің мөлшеріне байланысты өзгереді. Бұл сипаттамалар осы өлшемдерді түсіндіруді қиындатады және Phi немесе V Cramer-ге қарағанда аса пайдалы емес.

Қатені пропорционалды төмендету негізіндегі ассоциация өлшемі

Қатені пропорционалды төмендету (PRE) логикасына негізделген бірнеше номиналды деңгейдегі ассоциация өлшемі бар. Біз назарымызды лямбдаға (грек әріптерімен λ әрпі) аударамыз.

Бұл өлшемдер логикасы тәуелді айнымалы кейстеріндегі балдарын екі түрлі болжамдарға негізделген. Бір болжам бойынша біз тәуелсіз айнымалы туралы ақпаратты елемейміз, ал екіншісі бойынша, тәуелсіз айнымалыны ескереміз. Лямбда тәуелсіз айнымалы туралы білу қаншалықты біздің болжамдарымызды жақсартатыны туралы айтады.

Лямбда үшін біз алдымен әрбір кейс тәуелді айнымалыға (Y) түсетін категорияны болжаймыз және бұл кезде тәуелсіз айнымалы мәнді (X) елемейміз. Біз бұл жағдайда көзді жұмып болжаймыз, сондықтан біз жиі тәуелді айнымалы үшін кейстің мәнін бұрыс болжаймыз.

Екінші болжау үшін тәуелсіз айнымалыны ескереміз. Егер екі айнымалы ассоциациялы болса, онда бұл қосымша ақпарат болжаудың қателіктерін азайтады, яғни, біз аз кейстерді ғана дұрыс емес анықтауға тиіспіз. Сонымен қатар, айнымалылар арасындағы ассоциация күшті болса, қателердің азаюы да соғұрлым көп. Мінсіз ассоциациялар бар кейстерде, X баллдар есебінен Y баллдарды болжаудың барлығында біз ешқандай қатені болдырмас едік. Егер айнымалылардың арасында ассоциация жоқ болса, тәуелсіз айнымалы мәнді білу біздің болжауымыздың дәлдігін арттырмайды.

11,10-кесте Гендерге қатысты бой

Бой	Гендер		Барлығы
	Ер	Әйел	
Ұзын	44	8	52
Қысқа	6	42	48



Барлығы	50	50	100
---------	----	----	-----

Лямбданы есептеу үшін біз екі санды анықтаймыз: біріншіден, тәуелсіз айнымалыны елемей кезінде жасалған болжамдық қателер саны (E_1) және екіншіден, тәуелсіз айнымалы мәнді (E_2) ескере отырып жасалған болжау қателерінің саны. Содан кейін бұл екі қосындылар статистиканы алу үшін салыстырылады.

Мысалға, лямбданы есептеу нақтыланады. Біз 11,10-кестесі арқылы бой мен жыныс арасындағы қарым-қатынасты көрсететін 2×2 кестені келтірейік.

Алдымен, біз E_1 тәуелсіз айнымалы мәнді (гендер) елемей, жол маргиналдарымен жұмыс жасай отырып, барлық кейстерді болжау арқылы жасалған болжау қателерінің санын білеміз. Барлық субъектілердің биік екенін болжай аламыз, бұл 48 қатеге алып келеді. Яғни, бұл болжау үшін барлық 100 кейс жоғарғы жолда орналасады. Тек 52 кейс қана шынымен осы жолда жатады, сондықтан бұл болжам ($100 - 52$) немесе 48 қателіктерге әкеледі. Екінші жағынан, біз егер барлық субъектілердің қысқа болғанын болжасак, онда біз 52 қателік жасаймыз ($100 - 48 = 52$). Біз осы екі санның кішірегін қабылдаймыз және тәуелсіз айнымалыны елемей кезінде жасалған қателер санына арналған E_1 сияқты осы санға сілтеме жасаймыз. Осылайша $E_1 = 48$.

Содан кейін біз қайтадан Y баллын болжай аламыз, бірақ бұл жолы бағаннан бағанға жылжу арқылы X (гендер) есептеледі. Әрбір баған X категориясы болғандықтан, біз X -ті болжау жасауда пайдаланамыз. Әрбір баған үшін біз барлық кейстер ең үлкен жиілікті көрсететін ұяшықта түсетінін болжаймыз. Ерлерге арналған барлық 50 кейстер ұзын болады және 6 қате жаса болжанады ($50 - 44 = 6$). Әйелдер үшін біздің болжауымызша олардың барлығы қысқа және 8 қате жасалады. Бағаннан бағанға ауысып, біз болжаудың 14 қателігін жасадық, біз оны E_2 ($E_2 = 6 + 8 = 14$) деп санаймыз.

Егер айнымалы мәндер ассоциаланған болса, біз біріншіге қарағанда екінші процедурада аз қателер жасаймыз, сондықтан E_2 мәні E_1 қарағанда аздау. Біздің мысалда E_2 шынымен E_1 мәнінен аз, сондықтан гендер мен бой ассоциалы. Біздің қателіктеріміз 48-ден 14-ке дейін қысқартылды. Қатенің пропорционалды төмендеуін табу үшін 11,3-ші формуланы қолданыңыз:

11,13-формула

$$\lambda = \frac{E_1 - E_2}{E_1}$$

Үлгі мәселе үшін лямбда мәні болады

$$\lambda = \frac{E_1 - E_2}{E_1}$$

$$\lambda = \frac{48 - 14}{48}$$

$$\lambda = \frac{34}{48}$$

$$\lambda = 0.71$$

Лямбданың мәні 0,00-ден 1,00-ге дейін өзгереді. Әрине, 0,00 шамасы айнымалы мәндердің ешқайсысымен ассоциациясы емес екендігін білдіреді (E_1 мен E_2 бірдей), ал 1,00 мәні ассоциацияның мінсіз екендігін білдіреді ($E_2 = 0$ болады, ал тәуелді айнымалының баллдары тәуелсіз айнымалыдан қатесіз болжанады).

Phi немесе V-ге қарағанда, лямбда мәндерінің 0,00 мен 1,00 шектері арасында болуының дәл мағынасы бар: ол тәуелсіз айнымалының (X) тәуелді айнымалыны (Y) болжауға (немесе неғұрлым жеңілдетуге, түсінуге) қаншалықты көмектесе алатыны туралы айтады. 100-ге көбейтілгенде лямбданың мәні қателіктерді пайыздық көрсеткіштермен азайту тұрғысынан алғандағы ассоциацияның күшін білдіреді. Осылайша, 11,10-кесте үшін есептелген лямбда, гендер туралы білім бойды болжау қабілетімізді 71% -ға жақсартады деп айтуға болады. Яғни, бойды болжауға тырысқанда, біз гендер туралы 71% жақсы білеміз.

Лямбданы есептеудің орындалуын бекіту үшін келесі қадамдарды орындаймыз.



Рет-ретімен орында	Лямбданы есептеу
Реті	Орындалуы
1.	Бұл нұсқаулар жалпылама және кез келген көлемдегі кестелерге жарамды.
2.	E1 мәнін табу үшін N-нен жолдың ең үлкен аралық барлық мәнін (маргиналды) шегеріңіз.
3.	E2 мәнін табу үшін сол жақ бағаннан бастаңыз және бағанның барлық мәнінен бағандағы ең үлкен ұяшық жиілігін алып тастаңыз. Осы қадамды кестедегі барлық бағандар үшін қайталаңыз.
4.	2-қадамда тапқан барлық мәнді қосыңыз. Нәтиже E2 болады.
5.	E1 -ден E2 -ні шегеріңіз. 4-қадамда шыққан санды E1-ге бөліңіз. Шыққан сан лямбда болады.

Лямбданы интерпретациялауда

11,8-ші кестені жалпы басшылыққа алыңыз

PRE-ні түсіндіру үшін, лямбда мәнін 100-ге көбейтіңіз. Бұл пайыздық көрсеткіш тәуелсіз айнымалыны ескере отырып, тәуелді айнымалы мәнінің болжануы жақсарғанға дейінгі дәрежені білдіреді.

Қарым-қатынас үлгісін анықтау үшін бағанның пайыздық көрсеткішін пайдаланыңыз.

Лямбданың шектеулері

Лямбданың екі ерекшелігін ерекше атап өтуіміз қажет. Біріншіден, лямбда - асимметриялық. Бұл дегеніміз, статистиканың мәні қайсы айнымалы тәуелсіз ретінде қабылданатына байланысты өзгереді. Сондықтан, тәуелсіз айнымалы мәнді белгілеу кезінде кейбір сақтық шараларын назарда ұстау керек. Егер сіз бағандарда тәуелсіз айнымалыны орналастыру туралы келісімді ұдайы ұстанып отырсаңыз, онда статистиканың асимметриясы сізді шатастырмайды.

Екіншіден, егер жолдардағы барлық мәндердің біреуі басқаларына қарағанда әлдеқайда көп болса, лямбданың шатастыруы мүмкін болады. Ассоциацияның басқа өлшемдері 0,00-ден көбірек болған жағдайда да бұл 0,00 болуы мүмкін және кесте үшін шартты үлестірімдер айнымалылардың арасында ассоциация бар екенін көрсетеді. Бұл дегеніміз, маргиналдық жолдардың теңсіздігі өте үлкен болған кездегі лямбданы түсіндіруде аса бір сақтық шараларын қолдану керек екенін есте сақтаған жөн. Шын мәнінде, маргиналдық жолдардың теңсіздігі өте үлкен болған кезде ассоциация өлшемінің хи-квадраты мінсіз ассоциация өлшемі болар еді.

Күнделікті өмірден статистикалық мәселе

Конгресс туралы мәселе

Америкалықтардың Федералды үкімет пен Конгресс жұмысына наразылығына таңдануға болмайды. Адамдар кінәні кімге тағады? Олар бүкіл үкімет жүйесінің күйрегеніне сенеді ме, әлде Конгресстің қазіргі мүшелері кінәлі деп есептей ме? Зерттеушілердің жақында жүргізген америкалықтардың кездейсоқ жиынтығы бойынша сауалнама нәтижесі саяси көзқарастар бойынша таңқаларлық қайшылықты көрсетті.

Саяси партиялар				
	Республикалық	Демократ	Тәуелсіз	Барлығы
Мүшелер	234 (58,1%)	270 (57,1%)	312 (56,0%)	816 (56,9%)
Жүйе	169 (41,9%) 403(100%)	169 (42,9%) 473(100%)	245 (44,0%) 557(100%)	617 (43,1%) 1433(100%)

*Өзекті сұрақ: Сіздің көзқарасыңызда бүгінгі күні Конгреске қайсысы неғұрлым жақынырақ?

a. Саяси жүйе жақсы жұмыс істей алады, бұндағы мәселе оның мүшелерінде болып табылады.

b. Мүшелердің көбінің жақсы ниеттері бар, бұндағы мәселе бұзылған саяси жүйе.



Кітап: Статистика негіздері

Дәріс: Хи-квадратқа негізделген негізделген ассоциация өлшемдері: Phi және V Cramer

$$\chi^2 = 0,40, df = 2, p > 0,05 \quad V \text{ Cramer} = 0,02 \quad \text{Лямбда} = 0,00$$

Бұл айнымалылардың арасында ешқандай қарым-қатынас жоқ. Хи-квадрат статистикалық тұрғыдан елеулі емес және ассоциация өлшемдері іс жүзінде нөлге тең. Саяси партияға қатыстылығына қарамастан, америкалықтар жүйенің өзі емес, конгресстің қазіргі мүшелері олардың түңілуінің көзі болып табылады деген ортақ пікірде.

Түйін

Phi, V Cramer және лямбда номиналды айнымалылар үшін ассоциация өлшемдері және олардың әрқайсысы нақты бір жағдайға сәйкес келеді. Бұл статистикалар тек қарым-қатынастың күшін көрсетеді. Қарым-қатынастың үлгісі бағанның пайыздық көрсеткішін тексеру арқылы анықталуы мүмкін.