



15-дәріс



ҚАЗАҚСТАННЫҢ  
АШЫҚ  
УНИВЕРСИТЕТІ

# СТАТИСТИКА НЕГІЗДЕРІ

## Хи квадрат





### Осы дәрісті тыңдағаннан кейін:

1. Хи-квадрат тесті қолданылатын жағдайлардан мысал келтіріп, түсіндіре аласыз.
2. бивариативті кестенің құрылымын, кестеде күтілетін және бақыланатын жиіліктерге қатысты тәуелсіздік тұжырымдамасын түсіндіре аласыз.
3. Хи-квадрат бойынша гипотезаны тестілеудің логикасын түсіндіре аласыз.
4. Бес-сатылы үлгіні пайдалана отырып, хи квадрат тесті арқылы алынған нәтижелерді дұрыс түсіндіре аласыз.
5. Хи-квадрат тестінің шектеулерін, әсіресе, статистикалық мәнділік пен маңыздылық арасындағы айырмашылықты түсіндіре аласыз.

Дәрісте кестедегі айнымалылардың арасындағы байланыстың статистикалық мәнділігін тексеру үшін қолданылатын статистикалық әдіс-тәсілдер қарастырылады. Егер кестедегі жиіліктер арасында мәнді (significance) айырмашылық болса және кездейсоқтықтан нені күтуімізге болатынына қатысты бас жиынтықтағы айнымалылар байланысқан деген қорытынды жасауға болады. Бұл тұжырымда белгілі қате ықтималдығы бар, ол 0,05 деңгейінде орнатылады.

Хи-квадрат тесті қолданылатын жағдайларға мысал төменде көрсетілген:

1. Зерттеуші ерікті қауымдастықтарға мүшелікті зерттеуде некеге тұрмаған адамдардың отбасылық міндеттемелерінің аз және бос уақыттарының молырақ болатындығына байланысты волонтерлік жұмысқа көбірек араласатынын болжайды. Ол өз қауымдастығынан кездейсоқ жиынтық жинап, осы жиынтықтегі мүшелік деңгейі мен отбасылық жағдайдың арасындағы мәнді (significance) байланысты табады. Ол бас жиынтықтағы айнымалылар байланысты деп қорытынды жасайды.

2. Медициналық зерттеушілер салмақ тастаудың нақты жоспары барлық жастағы адамдар үшін бірдей орындала ма деген сауал қояды. Жоспарды тексеру үшін еріктілер жиынтықсі жиналды. Жиынтық жас топтарына қарай бөлінді, содан соң 2 айдан кейін жас топтары салмақ тастау бойынша салыстырылады. Нәтижелер қауымдастық үшін генерализацияланады.

3. Зерттеуші шектеулі иммиграциялық саясатты қолдау көздеріне қатысты сауалға алаңдайды. Сондықтан АҚШ-тың ұлттық репрезентативті жиынтықсін әртүрлі айнымалылар (гендер, мамандық, білім беру) арасындағы байланысты тексеру үшін және иммиграцияны қысқарту туралы бастаманы қолдау үшін пайдаланды. Егер зерттеуші жиынтықтегі маңызды байланысты тапса, ол бас жиынтықта осы айнымалылар (барлық ересек америкалық) байланысқан деп қорытындылайды.

**Хи-квадрат ( $\chi^2$ ) тесті** – әлеуметтік ғылымдардағы гипотезаларда ең жиі қолданылатын тестілердің бірі, бес-сатылы үлгінің 1-қадамындағы болжамдар мен талаптарды орындау оңай болғандықтан танымал. Тест номиналды айнымалыларға (өлшеудің ең төменгі деңгейі) жүргізіледі, онда ол параметрлік емес болады, яғни іріктеу үлестірімі немесе бас жиынтықтың формасы туралы ешқандай болжамды қажет етпейді.

Оңай орындалатын болжамдар мен талаптардың қандай артықшылығы бар? Нөлдік гипотезаны қабылдамау туралы шешімді (5-қадам) іске асыру мүмкін емес: бұл дегеніміз үлгідегі бір ғана мәлімдеме (1-қадам) немесе нөлдік гипотеза (2-қадам) дұрыс емес деген сөз. Әрине, әдетте, қабылдамауға болатын нөлдік гипотезаны бөліп шығарамыз. 1-қадамда айтылған үлгіге неғұрлым сенімді болсақ, соғұрлым нөлдік гипотеза қате болатынына сеніміміз кәміл. «Нашар» немесе оңай орындалатын модель нөлдік гипотезаны қабылдамау туралы шешім үлкенірек сенімділікпен жасалуы мүмкін дегенді білдіреді.

Хи-квадрат өз икемділігі үшін де жиі қолданылады: ол көптеген категориялар немесе баллдары бар айнымалылар үшін қолданылуы мүмкін. Мысалы, 8-дәрісте еріктілер бірлестігінде «өте жоғары қатысушылықтарымен» ерекшеленетін қара және ақнәсілді азаматтардың пропорциясындағы айырмашылықтың мәнділігін тексердік. Осы тестте зерттеуші латынамерикалық және азиялық америкалықтарды қосып тексергісі келсе, қалай болар еді? Мұндайда екі жиынтық тест қолданылмайды, есесіне хи-квадрат неғұрлым күрделі айнымалыларды оңай өңдейді. Сондай-ақ, 9-дәрісте ұсынылған ANOVA тестінен айырмашылығы хи-квадрат тесті кез-келген деңгейдегі айнымалыларға жүргізіледі.

### Бивариативті кестелер

Хи-квадраты кейтегі екі түрлі айнымалының баллдарын бір уақытта көрсетеді, сондықтан бивариативті кестемен есептеледі. Бивариативті кестелер мәнді (significance) қарым-қатынасты тестілеу үшін және



кейінгі тақырыптарда олар басқа мақсаттарды зерттеу үшін пайдаланылады. Шын мәнінде, бұл кестелер зерттеуде өте жиі пайдаланылады, ары қарай егжей-тегжейлі қарастырайық.

Ең алдымен, бивариативті кестелердің екі өлшемі бар (сөзсіз). Горизонталь (көлденең) өлшемдерді жолдар және вертикаль өлшемдерді (жоғары және төмен) бағандар деп атаймыз. Әрбір баған немесе жол айнымалыны көрсетеді, жолдар мен бағандардың қиылысы (ұяшықтар) екі айнымалы бойынша біріккен балдарды білдіреді.

Түсіндіру үшін мысал келтірейік. Мысалы, зерттеуші егде жастағы азаматтардың өміріне қызығушылық танытып, олардың ерікті топтарға қатысуы отбасылық жағдайына байланысты екенін білгісі келеді. Зерттеуші талдауды жеңілдету үшін жиынтықке үйленген немесе үйленбеген (ерлі-зайыптылар мен ажырасқан адамдарды қосқанда) адамдарды ғана кіргізді, ал ерікті қауымдастықтарға қатысуды қарапайым дихотомия ретінде өлшеді: адамдар жоғары немесе төмен деңгейде қатысуы бойынша жіктелді.

10.1-кесте. 100 егде жастағы азаматтың отбасылық жағдайы бойынша ерікті қауымдастықтарға қатысу реті

Отбасылық жағдайы			
Қатысу деңгейі	Отбасылы	Отбасы жоқ	Барлығы
Жоғары			50
Төмен	--	--	50
Барлығы	50	50	--
			50

Шартты түрде, бағандарға тәуелсіз айнымалыны (себеп болып табылатын айнымалы) және жолдарға тәуелді айнымалыны орналастырамыз. Бұл мысалда бір жағынан, отбасылық жағдай себепші айнымалы («Қатысуыңызға отбасылық жағдайыңыз әсер етті ме?» деген сауал болды) және әрбір баған осы айнымалы мәнге ие болады. Екінші жағынан, әрбір жол үлестірім деңгейі (жоғары немесе төмен) бойынша баллды көрсетеді. 10.1-кесте 100 егде азаматтан тұратын жиынтықке арналған бивариативті кестені көрсетеді.

Кестенің қосымша мәліметтеріне назар аударыңыз. Біріншіден, әрбір баған мен жол үшін аралық жинақтар енгізілген. Бұлар жол немесе бағанның маргиналдары деп аталады, бұл кейстер жиынтық бойынша 50 адам үйленген, ал 50 адам үйленбеген (бағана маргиналдары), 50-і қатысуы бойынша жоғары деңгейде, ал 50-і төмен деңгейде (жол маргиналдары) екенін білдіреді.

Екіншіден, жиынтықтағы ( $N = 100$ ) кейстердің жалпы саны жол мен баған маргиналдарының қиылысында көрсетіледі. Соңында кестенің таңбалануына назар аударыңыз. Әрбір жол және баған анықталған, кестеде айнымалылардың атаулары бар сипаттамалық тақырып бар, мұндағы тізімде алдымен тәуелді айнымалылар келтірілген. Анық, толық таңбалар мен қысқа тақырыптар барлық кестеде, графикте және диаграммада болуы тиіс.

Сіз байқағандай, 10.1-кестеде кейбір маңызды ақпарат түсіп қалған: ол кесте бойындағы сандар. Кестені аяқтау үшін жиынтықтың әрбір мүшесінің отбасылық жағдайы мен қатысу деңгейін жіктеуіміз қажет, баллдардың әрқайсысының комбинациясы қаншалықты жиі кездесетінін түртіп алып, олардың кестелердің тиісті ұяшықтарына жазып отыруымыз керек. Әрбір айнымалыда екі ықтимал бал бар, сондықтан кестедегі ұяшыққа сәйкес келетін төрт балл комбинациясы болуы мүмкін. Мысалы, жоғары деңгейде қатысатын үйленген адамдар жоғарғы сол жақтағы ұяшықта есептеледі, төменгі деңгейде қатысатын үйленбеген адамдар төменгі оң жақ ұяшықта есептеледі және т.с.с. Есептеу аяқталған кезде, әр ұяшық баллдардың әр комбинациясының қанша рет кездескенін көрсетеді.

Ақыр соңында, екі балдан асатын айнымалыларды орналастыру үшін кестені кеңейте аламыз. Егер басқа отбасылық жағдайы бар адамдарды да қосқымыз келсе (жесір, бөлек тұратын және т.с.с.), жай ғана баған қоса саламыз. Күрделірек тәуелді айнымалыларды да оңай бейімдеуге болады. Егер үш категория бойынша қатысу дәрежесін өлшеген болсақ (мысалы, жоғары, орташа және төмен), онда жай ғана кестеге жол қосамыз.

## Хи-квадраттың логикасы

Хи-квадраттың тесттің бірнеше қолданысы бар, бірақ тек тәуелсіздікті есептеу үшін хи-квадрат тестін



қарастырамыз. Тәуелсіздік термині екі-жиынтық гипотезалық тесті (8-тарау) және ANOVA тестінің (9-тарау) талаптарында кездесті. Хи-квадрат тұрғысынан алғанда, тәуелсіздік жиынтықтар арасында емес, айнымалылар арасындағы байланысқа қатысты. Егер бір айнымалының белгілі бір категорияға жіктелуі екінші айнымалының нақты бір категориясына түсіп кету ықтималдығына әсер етпейтін болса, екі айнымалы бір-бірінен тәуелсіз болады. Мысалы, некеде және некеде тұрмайтындар кейсінің жіктелуі жоғары не төмен деңгейде қатысуын жіктеу кейсіне әсер етпесе, 10.1- кестедегі айнымалылар тәуелсіз болады. Басқаша айтқанда, қатысу деңгейі мен отбасылық жағдайының бір-біріне мүлде қатысы жоқ болса, айнымалылар тәуелсіз болғаны.

10.1-кестені қайта қарастырайық. Егер айнымалылар тәуелсіз болса, ұяшықтардың жиіліктері тек қана кездейсоқ шанспен анықталаушы еді, сондықтан некедегі респонденттердің жартысына жуығы жоғары деңгейде қатысатынын, ал қалған жартысы төменгі деңгейде қатысатынын байқайтын едік. Бұл шыншыл монетаның бір жағы баллмен 50% түсетінін көрсететін еді. Осыған ұқсас схема 50 тұрмыс құрмаған респонденттер үшін де болады, 10.2-кестеде көрсетілгендей, ондағы төрт ұяшықтың әрқайсысында баллмен 25 кейс бар. Ұяшық жиілігінің орналасуы отбасылық жағдайдың адамның қатысу деңгейіне әсер етпейтінін көрсетеді. Жоғары немесе төмен деп жіктелу ықтималдығы екі отбасылық жағдай үшін де 0,50 болады, сондықтан айнымалылар тәуелсіз бола алады.

Тәуелсіздікке арналған хи-квадрат тесті үшін нөлдік гипотеза бұл айнымалылар тәуелсіз дегенді білдіреді.

Нөлдік гипотезаны шынайы деп қарастыратын болсақ, онда кездейсоқ шанс қана мүмкін болады деп күтілген ұяшық жиілігі есептеледі. Бұл жиіліктер күтілетін жиіліктер деп аталады ( $f_e$  болып рәсімделеді), оларды кестеде нақты бақыланған жиіліктермен (бақыланған жиіліктер,  $f_o$  болып рәсімделеді), яғни ұяшықты ұяшықпен салыстырамыз. Нөлдік гипотеза дұрыс, ал айнымалылар тәуелсіз болса, күтілетін және бақыланған жиіліктер арасында сәл ғана айырмашылық болады. Алайда нөлдік гипотеза жалған болса, екеуінің арасында керісінше айырмашылық үлкен болады. Күтілетін ( $f_e$ ) және бақыланған жиіліктер ( $f_o$ ) арасындағы айырмашылықтар неғұрлым көп болса, соғұрлым айнымалылар тәуелсіздігі туралы ықтималдық та төмен болады, нөлдік гипотезаны қабылдай алмауымыз әбден мүмкін.

10.2-кесте. 100 егде жастағы азаматтың отбасылық жағдайы бойынша ерікті қауымдастықтарға қатысу реті

Қатысу деңгейі	Отбасылық жағдайы		
	Отбасылы	Отбасы жоқ	Барлығы
Жоғары	25	25	50
Төмен	25	25	50
Барлығы	50	50	100

Хи-квадратты есептеу

Хи-квадрат тестін жүргізу – барлық гипотезаларды тестілеудегі сияқты - біз деректер жиынтығынан тест статистикасын,  $\chi^2$  (алынған) есептеп, содан кейін барлық ықтимал жиынтықтардың шығысына (outcomes) жиынтық үлестіріміндегі осы мәнді (significance)(significance)салыңыз. Атап айтқанда,  $\chi^2$  (алынған) мәнін  $\chi^2$ (сыни) мәнімен салыстыруға болады, ол белгілі бір альфа деңгейіне және еркіндік дәрежесіне қатысты хи-квадраттық кестеге (C аппендиксі) жүгіну арқылы анықталады.

Хи-квадратты есептеу тәртібі 10.1-формулада келтірілген:

$$10.1\text{-формула } \chi^2 (\text{алынған}) = \sum \frac{(f_o - f_e)^2}{f_e}$$

мұнда  $f_o$  = Бивариативті кестеде бақыланған ұяшық жиіліктері

$f_e$  = айнымалылар тәуелсіз болған жағдайда күтілетін ұяшық жиілігі

Әрбір ұяшық үшін күтілетін жиілікті бақыланған жиіліктен шегеріп, нәтижесін квадраттаңыз, содан соң сол ұяшық үшін күтілетін жиілікке бөліңіз. Барлық ұяшықтар үшін алынған мәндердің нәтижелерін қосыңыз.



Бұл формула кестедегі әрбір ұяшық үшін күтілетін жиілікті қажет етеді. 10.2-кестеде маргиналдар барлық жолдар мен бағандар үшін бірдей мәнді, ал күтілетін жиіліктер интуиция арқылы анықталады: барлық төрт ұяшық үшін  $f_e = 25$ . Кәдімгі кейсте күтілетін жиіліктер анық болмайды, ал маргиналдар тең емес, сондықтан әрбір ұяшықтағы күтілетін жиілікті табу үшін 10.2-формуланы пайдалануға міндеттіміз:

$$10.2\text{-формула } f_e = \frac{\text{Маргиналдық жол} \times \text{Маргиналдық баған}}{N}$$

Яғни, кез-келген ұяшық үшін күтілетін жиілік, жол бойынша (маргиналдық жол) келтірілген кейстердің жалпы санын бағандағы (маргиналдық баған) кейстердің жалпы санына көбейтіп, алынған нәтижені кестедегі кейстердің жалпы санына (N) бөлгенге тең.

### Есептеу үлгісі

10.3-кесте мысалы осы процедураларды түсіндіреді. 100 әлеуметтік қызметкерден тұратын кездейсоқ жиынтық Әлеуметтік қызмет саласы бойынша білім беру кеңесінің бакалавриат (баған мен тәуелсіз айнымалы) бағдарламасына аккредитациясы бойынша және оқу бітірген соң үш ай ішінде әлеуметтік жұмысқа тұру (жол немесе тәуелді айнымалы) бойынша жіктелді.

10.3-кесте. Бакалавриат бағдарламасының аккредитациясы бойынша 100 әлеуметтік маманның жұмысқа орналасуы

Аккредитациялық статус			
Жұмыспен қамтылуы	Аккредитациядан өткен	Аккредитациядан өтпеген	Барлығы
Әлеуметтік қызметкер болып істейді	30	10	40
Әлеуметтік қызметкер болып істемейді	25	35	60
Барлығы	55	45	100

10.4-кесте. 10.3-кесте үшін күтілетін жиіліктер

Жұмыспен қамтылуы	Аккредитациядан өткен	Аккредитациядан өтпеген	Барлығы
Әлеуметтік қызметкер болып істейді	22	18	40
Әлеуметтік қызметкер болып істемейді	33	27	60
Барлығы	55	45	100

10.5-кесте. 10.3-кесте үшін есеп кестесі

(1)	(2)	(3)	(4)	(5)
$f_o$	$f_e$	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
30	22	8	64	2,91
10	18	-8	64	3,56
25	33	-8	64	1,94
35	27	8	64	2,37
100	100	0		10,78

Жоғарғы сол жақ ұяшықтан (әлеуметтік қызметкерлер болып жұмыс істейтін аккредитациядан өткен бағдарлама түлектері) бастасақ, осы ұяшық үшін күтілетін жиілік 10.2-формула бойынша –  $(40 \times 55) / 100$





немесе 22 болады. Осы жолдағы басқа ұяшықтар үшін (әлеуметтік қызметкерлер болып жұмыс істейтін аккредитациядан өтпеген бағдарлама түлектері) күтілетін жиілік  $(40 \times 45) / 100$  немесе 18-ге тең. Төменгі жолдағы екі ұяшық үшін күтілетін жиілік  $(60 \times 55) / 100$ , яғни 33 немесе  $(60 \times 45) / 100$ , яғни 27 болады. Барлық төрт ұяшықтар үшін күтілетін жиілік 10.4-кестеде көрсетілген.

Жолдар мен бағанның маргиналдары, сондай-ақ 10.4-кестеде келтірілген кейстердің жалпы саны 10.3-кестеде көрсетілген кейстермен дәлме-дәл екенін ескеріңіз. Күтілетін жиіліктерге арналған жолдар мен бағанның маргиналдары әрқашан бақыланатын жиіліктерге тең болуы қажет, бұл осыған дейінгі арифметикалық есептеуіңізді тексерудің ыңғайлы әдісі.

Осы деректер үшін хи-квадраттың мәні 10.1-формула арқылы анықталады. Хи-квадратты есептеуге қажет бірнеше кадамды орындау үшін 10.5-кесте сияқты есептеу кестесін қолданған жөн. Кестеде жоғарғы сол жақ ұяшықтан төменгі оң жақ ұяшыққа дейін, кесте бойынша солдан оңға қарай және жоғарыдан төменге қарай жылжығанда 1-бағанда бақыланатын жиіліктер ( $f_o$ ) көрсетілген. 2-бағанда күтілетін жиіліктер ( $f_e$ ) дәл осы тәртіппен жазылған. Ұяшық жиілігін осы бағандардың екеуінде де бірдей тәртіпте тізімдегеніңізге сенімді болу үшін қайталап тексеріп шығыңыз.

Келесі кадам - әрбір ұяшық үшін бақыланатын жиіліктен күтілетін жиілікті азайтып, осы мәндерді 3-бағанға жазу. 4-бағанды толықтыру үшін 3-бағандағы мәнді (significance)(significance)квадраттаңыз, одан кейін 5-бағанда, 4-бағанды осы ұяшық үшін күтілетін жиіліктің мәніне бөлесіз. Ақырында, 5-бағанды қосыңыз. Бұл бағанның қорытындысы хи-квадрат болады. 10.3-кесте үшін  $\chi^2(\text{алынған}) = 10,78$ .

1-баған ( $f_o$ ) мен 2 баған ( $f_e$ ) мәндері тең. Бұл әрдайым осылай болатын кейс, барлық мәндер сәйкес келмесе, есепте қате жібердіңіз. Бәлкім, күтілетін жиіліктерді есептеу кезінде қате кеткен болар. Сондай-ақ, 3-бағандағы барлық қосынды әрдайым 0 болады, бұл математикалық есептеуді тексеріп алудың тағы бір ыңғайлы жолы.

Хи-квадрат үшін осы жиынтықтың мәнін мәнділік үшін де тексеру қажет.

10.2-формулананы пайдалана отырып күтілетін жиіліктерді ( $f_e$ ) есептеудің орындалуын бекіту үшін келесі қадамдарды орындаймыз.

1. Бивариативті кестенің жоғарғы сол жақ ұяшығынан бастаңыз. Осы ұяшық үшін маргиналды жолға маргиналды бағанды көбейтіңіз.

2. 1-қадамда тапқан санды N-ге бөліңіз. Нәтиже - осы ұяшық үшін күтілетін жиілік.

3. Осы  $f_e$  белгісін бөлек кестеге жазыңыз (10.4-кестені қараңыз) немесе есептеуіш кестенің екінші бағанына (10.5-кестені қараңыз) жазыңыз.

4. Әр ұяшық үшін 1-3 қадамдарды қайталаңыз.

5. Күтілетін жиіліктерді есептегенде дұрыс маргиналды жол мен бағанды таңдағаныңызға сенімді болу үшін қайталап тексеріп шығыңыз.

10.1-формулананы және 10.5-кесте сияқты есептеу кестесін пайдалана отырып Хи-квадратты есептеу

1. 1-бағанда бақыланатын жиіліктерді ( $f_o$ ) тізіп шығыңыз. Бұл бағанның жалпы саны (N).

2. 2-бағанда күтілетін жиіліктерді ( $f_e$ ) тізіп шығыңыз. Осы бағанның жалпы саны 1-баған санына (N) тең болуы тиіс.

3. Әрбір ұяшық үшін бақыланған жиіліктен ( $f_o$ ) күтілетін жиілікті ( $f_e$ ) азайтып, шыққан мәндерді есептеуіш кестенің  $f_o - f_e$  деп белгіленген үшінші бағанына тізімдеңіз. Осы баған үшін барлық мәндерді табыңыз. Егер барлық мән нөлге тең болмаса, есептеуде қате жібердіңіз, сондықтан есептеулерді қайта тексеруіңіз керек.

4. Кестенің үшінші бағанындағы әрбір мәнді квадраттаңыз және нәтижені  $(f_o - f_e)^2$  деп белгіленген төртінші бағанға жазыңыз.

5. 4-бағандағы әрбір мәнді сол ұяшық үшін күтілетін жиілікке бөліп, нәтижені бесінші бағанға жазыңыз  $(f_o - f_e)^2 / f_e$ .

6. Бесінші бағанның барлық мәндерін табыңыз. Сол мән -  $\chi^2(\text{алынған})$  болады.

## ТҮЙІН

1. Тәуелсіздіктің хи-квадрат тесті айнымалылар кесте форматына салынған кезде қолданылады. Нөлдік гипотеза айнымалылардың тәуелсіздігі немесе кейстің бір айнымалы бойынша нақты категорияға жіктелуі екінші айнымалының кез-келген нақты бір критерийіне жатқызылатынына ықпалын тигізбеуі болып



табылады.

2. Хи-квадрат параметрлік емес болғандықтан және тек номиналды айнымалы мәндерді талап ететіндіктен, оның үлгілік болжамы жеңіл орындалады. Одан әрі, ол бивариативті кестелерден есептеледі, онда жолдар мен бағандардың саны оңай кеңейтіледі. Сондықтан, басқа тесттерді қолдану жарамсыз немесе қиындық туғызатын көптеген жағдайларда хи-квадрат тестін пайдалануға болады.