



СТАТИСТИКА НЕГІЗДЕРІ

Тәуелсіздік тесті үшін Хи-квадрат
Бағандардың пайыздық көрсеткіші





Енді тәуелсіздік үшін хи-квадрат тестін өткізуге дайынбыз. Естеріңізге сала кетейік, айнымалылар бір-бірінен тәуелсіз болған жағдайда, бір айнымалыдағы кейстің баллы басқа айнымалыдағы балға қатысы болмайды. Мәнділік (сигнификанс) тестінің бес-сатылы үлгісі әрдайым шешім қабылдауды ұйымдастыруға негіз болады. 10.3-кестеде келтірілген деректерді мысалға алайық.

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

1-қадам. Болжам жасау және тест талаптарына сәйкестендіру. Айта кетейік, тест параметрлік емес болғандықтан, жиынтық үлестірімінің формасы туралы ешқандай жорамал жасамаймыз.

Үлгі: Тәуелсіз кездейсоқ жиынтық
Өлшем деңгейі - номиналды

2-қадам. Нөлдік гипотезаны белгілеу. Нөлдік гипотеза екі айнымалының тәуелсіз екендігін көрсетеді. Егер нөлдік гипотеза шын болса, баыланатын және күтілетін жиіліктер арасындағы айырмашылықтар аз болады. Әдеттегідей, зерттеу гипотезасы нөлдік гипотезаға тікелей қайшы келеді. Осылайша, біз H_0 қабылдасак, онда зерттеу гипотезасына қолдау көрсетілетін болады.

H_0 : Екі айнымалы тәуелсіз.
 $(H_1$: Екі айнымалы тәуелді)

3-қадам. Іріктеу үлестірімін таңдау және сыни аралықты белгілеу. Жиынтық хи-квадраттың іріктеу үлестірімінің Z және t үлестірімдерінен айырмашылығы, ол үлестірімінің жоғарғы жағындағы (оң жақта) жиынтық хи-квадраттың жоғары мәндерімен позитив көлбеуленген. Осылайша, хи-квадрат тестімен сыни аралық іріктеу үлестірімінің жоғарғы жағында белгіленеді.

χ^2 (сыни) мәндер C аппендиксінде берілген. Бұл кесте t кестесіне ұқсас, альфа деңгейлері үстіңгі, ал еркіндік дәрежелері төмен жағында орналасқан. Ал хи-квадрат пен еркіндік дәрежесі (df) 10.3-формула бойынша анықталады:

10.3-формула

$$df = (r - 1)(c - 1)$$

Екі жолдан және екі бағаннан тұратын кестенде (2×2 кесте), жиынтықтегі кейстер санына қарамастан, еркіндіктің бір дәрежесі ғана болады. Екі жолдан және үш бағаннан тұратын кесте $(2 - 1)(3 - 1)$ түрінде болады немесе екі дәрежелі еркіндікке ие. Біздің жаттығуда бір дәрежелі еркіндікке ие 2×2 кесте, сондықтан егер альфаны 0,05 деңгейінде белгілесек, сыни хи-квадрат балы 3,841 болады. Осы шешімдерді қорытындылай келе:

$$\begin{aligned} \text{Жиынтық үлестірім} &= \chi^2 \text{ үлестірімі} \\ \text{Альфа} &= 0,05 \\ \text{Еркіндік деңгейі} &= 1 \\ \chi^2 (\text{сыни}) &= 3,841 \end{aligned}$$

4-қадам. Тест статистикасын есептеу. Хи-квадратты есептеу алдыңғы бөлімде таныстырылды. Естеріңізде болса, ол мынаған тең болды

$$\chi^2 (\text{алынған}) = 10,78$$



5-қадам. Шешім қабылдау және тест нәтижелерін интерпретациялау. Тест статистикасын сыни аралықпен салыстыру арқылы

$$\chi^2 (\text{алынған}) = 10,78$$

$$\chi^2 (\text{сыни}) = 3,841$$

тест статистикасы сыни аралыққа түсетінін көріп отырмыз, сондықтан еркіндіктің нөлдік гипотезасын жоққа шығарамыз. 10.3-кестеде көрсетілген бақыланған ұяшық жиіліктерінің кездейсоқ шығуы екіталай. Айнымалылар тәуелді. Атап айтқанда, осы жиынтық деректері негізінде әлеуметтік жұмыс саласында жұмыспен қамтамасыз ету ықтималдығы бағдарламаның аккредитациядан өту мәртебесіне байланысты. (Тәуелсіздік үшін хи-квадрат тестін жүргізу және түсіндіру тәжірибесі үшін тараудың соңындағы кез-келген жаттығуды орындаңыз).

Хи-квадрат тесті нақты нені анықтайды, нені жасырып қалатынын анықтап алған жөн. Мәнді (significance) хи-квадрат дегеніміз бас жиынтықтағы айнымалылар бір-бірінен тәуелді (ықтимал) екенін білдіреді. Біздің мысалда бұл аккредитациядан өту мен адамның әлеуметтік қызметкер ретінде жұмыс істеп жатқаны арасында байланыс бар дегенді білдіреді. Бірақ, айнымалылар арасындағы байланыс нақты нені білдіреді? Түлектердің қай түрі мамандық бойынша жұмыс таба алады? Бұны анықтау үшін қосымша есеп жүргізу қажет. Біз ол туралы келесі дәрісте айтамыз.

Бағандардың пайыздық көрсеткіші

Тәуелсіз айнымалы (біздің мысалдағы аккредитациядан өту жағдайы) тәуелді айнымалы мәнге (әлеуметтік қызметкер ретінде жұмыс істеу) қалай әсер ететінін анықтауды бағандардың пайыздық көрсеткіштерін есептеу арқылы немесе бивариативті кестенің әрбір бағанындағы пайыздарды есептеу арқылы табуға болады. Бұл процедура жиілік үлестірімінің пайыздық көрсеткішін есептеуге ұқсас (2-тарауды қараңыз).

10.6-кесте. 10.6-кестесі үшін бағандардың пайыздық көрсеткіші

Аккредитациялық жағдайы			
Жұмыспен қамтылу	Аккредитациядан өткен	Аккредитациядан өтпеген	Барлығы
Әлеуметтік қызметкер болып істейді	54,55%	22,22%	40,00%
Әлеуметтік қызметкер болып істемейді	45,45%	77,78%	60,00%
Барлығы	100,00% (55)	100,00% (45)	100,00%

Бағандағы пайыздық көрсеткіштерді есептеу үшін әрбір ұяшық жиілігін бағандағы кейстердің жалпы санына (бағанның маргиналы) бөліп, нәтижені 100-ге көбейтіңіз. 10.3-кесте үшін жоғарғы сол жақ ұяшықтан бастап, сол ұяшықта 30 кейс, ал бағанда 55 кейс бар екенін көреміз. Осылайша, аккредитациядан өткен бағдарламалардың 55 түлегінің 30-ы әлеуметтік қызметкер болып жұмыс істейді. Сондықтан, осы ұяшықтың бағандық пайызы $(30/55) \cdot 100 = 54,55\%$. Төменгі сол жақ ұяшық үшін баған пайызы $(25/55) \cdot 100 = 45,45\%$. Оң жақ бағандағы (аккредитациядан өтпеген бағдарлама түлектері) екі ұяшық үшін бағанның пайыздық көрсеткіші $(10/45) \cdot 100 = 22,22\%$ және $(35/45) \cdot 100 = 77,78\%$. 10.6-кесте 10.3-кесте үшін барлық бағандардың пайыздық көрсеткішін көрсетеді.

Бағанның пайыздық көрсеткіші айнымалылар арасындағы қатынасты анық көрсетеді, сондықтан 10.6-кестеден аккредитациядан өткен бағдарлама студенттерінің әлеуметтік қызметкерлер ретінде жұмыс істеуі мүмкін екенін көре аламыз. Осы түлектердің баллмен 55%-ы әлеуметтік қызметкерлер ретінде жұмыс істейді, ал аккредитациядан өтпеген бағдарлама түлектері баллмен 22%-ды құрайды. Біз бұл байланыстың мәнді екенін білеміз (кездейсоқтықтан туындауы екіталай), енді екі бағанның пайыздық көрсеткіші арқылы екі айнымалының қалай байланысты екенін білеміз. Осы нәтижелерге наар аударсақ, аккредитациядан өткен бағдарлама түлектерінің әлеуметтік жұмыспен қамтылуда сөзсіз артықшылығы бар.



Қорытындыда екі сәтті атап өтейік:

1. Хи-квадрат - статистикалық мәнділік тесті. Ол бас жиынтықтағы айнымалылар тәуелсіз болатын нөлдік гипотезаны тексереді. Егер нөлдік гипотезаны жоққа шығарсақ, онда белгілі қате ықтималдығымен (альфа деңгейімен анықталады) бас жиынтықтағы айнымалылар бір-бірінен тәуелді болады деп қорытындылаймыз. Дегенмен, хи-квадрат байланыстың сипаты туралы ақпарат бермейді.

2. Бағанның пайыздық көрсеткіші тәуелсіз айнымалы тәуелді айнымалыға қалай әсер ететінін көрсетеді. Біздің мысалда олар әлеуметтік қызметкерлер ретінде жұмыс таба алатын аккредитациядан өткен бағдарлама түлектері екенін көрсетеді. Біз 11-тарауда бивариативті ассоциацияны талқылағанда бағанның пайыздық көрсеткішін кеңінен зерттейміз.

Хи-квадрат тесті: Қосымша мысал

Осыған дейін 2 2 кестеге, яғни екі жолға және екі бағанға арналған кестелерге назар аудардық. Көрнекілік үшін есептеулер мен шешім қабылдау процесі арқылы үлкенірек кестеге қол жеткіземіз. Негізі, үлкенірек кестелер көбірек есептеуді қажет етеді (өйткені олардың ұяшықтары көп), бірақ барлық басқа аспектілері бойынша 2 2 кестесі сияқты өңделеді.

Зерттеуші колледж студенттерінің отбасылық жағдайының олардың академиялық үлгеріміне ықтимал әсерін білгісі келеді. Отбасылық міндеттері қосымша ауыртпалық салатын некедегі студенттер үйленбеген студенттерге қарағанда академиялық тұрғыдан зардап шеге ме? Академиялық көрсеткіштер отбасылық жағдайға тәуелді ме? 453 студенттің кездейсоқ іріктемесі жасалып, әрбір студент отбасылық жағдайы мен баллмен орташа баллға (GPA) негізделген жақсы, орта немесе нашар студент ретінде жіктелді. Нәтижелер 10.7-кестеде келтірілген.

10.7-кесте Отбасылық жағдайы бойынша орташа бал (GPA)

Отбасылық жағдай			
GPA	Некеде	Некеде емес	Барлығы
Жақсы	70	90	160
Орта	60	110	170
Нашар	45	78	123
Барлығы	175	278	453

Жоғарғы сол жақ ұяшық (GPA жоғары некедегі студенттер) үшін күтілетін жиілік $(160 \times 175) / 453$ немесе 61,8 болады. Осы жолдағы басқа ұяшық үшін күтілетін жиілік $(160 \times 278) / 453$ немесе 98,2. Осындай тәсілмен барлық күтілетін жиіліктер есептеледі (дұрыс жол мен бағанның маргиналдарын алғаныңызды тексеріп алыңыз), бұл 10.8-кестеде көрсетілген.

10.8-кесте 10.7-кесте үшін күтілетін жиілік

Отбасылық жағдай			
GPA	Некеде	Некеде емес	Барлығы
Жақсы	61,8	98,2	160
Орта	65,7	104,3	170
Нашар	47,5	75,5	123
Барлығы	175,0	278,0	453

Келесі қадам – әрбір ұяшық үшін тиісті f_o және f_e пайдаланатынымызға өте мұқият бола отырып, 2 (алынған) формуласын шешу. Есептеулермен жұмыс жасау үшін және одан кейін алынған статистикалық мәнділікке хи-квадраттық тестілеу үшін есептеуіш кестені (10.9-кесте) пайдаланамыз. Естеріңізге сала кетейік, алынған хи-квадрат 5-бағанның жалпы санына тең.

Алынған хи-квадраттың мәнін (2,80) енді оның мәнділігі үшін тексеруге болады.



10.9-кесте. 10.7-кесте үшін есептеуіш кестесі

(1)	(2)	(3)	(4)	(5)
f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
70	61,8	8,2	67,24	1,09
90	98,2	- 8,2	67,24	0,69
60	65,7	5,7	32,49	0,50
110	104,3	-5,7	32,49	0,31
45	47,5	-2,5	6,25	0,13
78	75,5	2,5	6,25	0,08
453	453,0	0,0	2 (алынған) = 2,80	

1-қадам. Болжам жасау және тест талаптарына сәйкестендіру. Айта кетейік, тест параметрлік емес болғандықтан, іріктеме үлестірімінің формасы туралы ешқандай жорамал жасамаймыз.

Үлгі: Тәуелсіз кездейсоқ іріктеме
Өлшем деңгейі - номиналды

2-қадам. Нөлдік гипотезаны белгілеу.

H_0 : Екі айнымалы тәуелсіз.
 $(H_1$: Екі айнымалылар тәуелді)

3-қадам. Іріктеу үлестірімін таңдау және сыни аралықты белгілеу.

Іріктемелік үлестірім = χ^2 үлестірімі
Альфа = 0,05
Еркіндік деңгейі = $(r - 1)(c - 1) = (3 - 1)(2 - 1) = 2$
 χ^2 (сыни) = 5,991

4-қадам. Тест статистикасын есептеу.

χ^2 (алынған) = 2,80

5-қадам. Шешім қабылдау және тест нәтижелерін интерпретациялау. Тест статистикасы, яғни χ^2 (алынған) = 2,80 сыни аралыққа түспейді, өйткені альфа = 0,05, $df = 2$ үшін χ^2 (сыни) 5,991 деңгейінде басталады. Сондықтан нөлдік гипотезаны жоққа шығарамыз. Бақыланатын жиіліктердің, егер айнымалы мәндер тәуелсіз болып және оларды тек кездейсоқтықпен табу күтілетін болғандағы жиіліктерден елеулі айырмашылығы жоқ. Осы іріктеме нәтижелерінің негізінде колледж студенттерінің академиялық көрсеткіштері олардың отбасылық жағдайына тәуелді емес деп қорытынды жасауға болады. Нөлдік гипотезаны жоққа шығармағандықтан, 10.3-кестеде жасағандай, бағанның пайыздық көрсеткіштерін қарастырмаймыз.

SPSS бағдарламасын пайдалану: Кросскестелер және хи-квадрат

Осы «SPSS пайдалану» бөлімінде Crosstabs деп аталатын процедураны пайдаланамыз, ол бивариативті кестелерді, бағанның пайыздық көрсеткіштерін және хи-квадратты шығарады.

Хи-квадрат тестінің шектеулері

Кез-келген басқа тест сияқты хи-квадраттың да кемшіліктері бар. Сондықтан бірнеше ықтимал қиындықтарды ескерген жөн. Біріншіден, айнымалылардың категориялары көп болған кезде хи-квадратты интерпретациялау оңай емес. Мысалы, бес категориялық екі айнымалы 25 ұяшықпен 5 5 кестесін жасайды, бұл түсіну мен қабылдауға қиын, баллдардың тым көп комбинациялары болуы мүмкін. Өте бір қиын эмпирикалық ереже ретінде хи-квадраттық тестті айнымалылардың екеуінде де төрт немесе одан да аз балл болған кезде интерпретациялауға және түсінуге оңай.

Тесттің ары қарайғы екі шектеуі іріктеме мөлшеріне қатысты. Іріктеме шағын болғанда, барлық ықтимал



іріктеме нәтижелерінде іріктемелік үлестірім хи-квадрат үлестірімі арқылы түсінікті сипатталады деп санай алмаймыз. Хи-квадрат үшін шағын іріктеме ұяшықтардағы жоғары пайыздық көрсеткіш 5 немесе одан кем күтілетін (f_e) жиіліктер болатын жерде анықталады. Зерттеушіге «ұяшықтардағы жоғары пайыздық көрсеткіштер» дегенді анықтауға көмектесу үшін түрлі эмпирикалық ережелер жасалған. Кез-келген ұяшық 5 немесе одан аз күтілетін жиіліктерге ие болғанда, түзету әрекетін жасау ең қауіпсіз әдіс болуы мүмкін.

2 x 2 кестелері үшін χ^2 (алынған) мәнін үздіксіздік үшін Йетс түзетуін қолдану арқылы нақтылауға болады, оның формуласы

10.4-формула

$$\chi_c^2 = \sum \frac{(f_o - f_e - 0,5)^2}{f_e}$$

мұнда χ_c^2 = түзетілген хи-квадрат

$|f_o - f_e|$ = әрбір ұяшық үшін бақыланатын және күтілетін жиіліктер арасындағы айырмашылықтың абсолютті мәндері. (ЕСКЕРТУ: абсолютті мәндер қосу және азайту таңбаларын елемейді).

Түзету факторы ($f_o - f_e$) шартының абсолютті мәнінен 0,5-ті шегеріп, айырмашылықты квадраттап, оны күтілетін жиілікке бөлу арқылы табылады.

2 x 2-ден үлкен кестелер үшін шағын іріктеме үшін χ^2 (алынған) есептеуге арналған түзету формуласы жоқ. Айнымалылардың кейбір категорияларын біріктіріп, осылайша ұяшық өлшемдерін ұлғайтуға болады. Мысалы, 10.7-кестеде «орташа» және «нашар» категорияларын біріктіруге болады. Бұл кестенің өлшемін 2 x 2-ге дейін азайтып, төменгі жолдағы кейстердің санын көбейтеді.

Жалпы алғанда, категорияларды мүмкін болатын кезде ғана біріктіру керек. Басқаша айтқанда, орташа және нашар студенттер арасындағы айырмашылықтарды тек қана статистикалық тест талаптарына сәйкес келмегені үшін ғана жойып жібермеу керек.

Егер категориялар ұяшық жиіліктерін құрастыру үшін біріктірілмейтінін сезсеңіз және күтілетін жиіліктермен ұяшықтардың пайыздық көрсеткіші 5 немесе одан да аз болса, онда дұрыс емес хи-квадрат тестін нәтижелер жарамды деп саналалатын сақтық көлеміне жеткенге дейін жалғастыруға болады.

Іріктеме көлеміне байланысты екінші кемшілік ірі іріктемелерде ғана кездеседі. 8-тарауда гипотезаның барлық тестілері іріктеме мөлшеріне қатысты сезімтал болатынын көрсеттім. Яғни, нөлдік гипотезаны жоққа шығару ықтималдығы көбейеді, себебі кез-келген басқа факторға қатысты кейстердің де саны артуда. Хи-квадраты іріктеме көлеміне тікелей байланысты: іріктеме көлемі артқан сайын χ^2 (алынған) мәні де артып отырады. Егер N еселенген болса, мысалы, χ^2 (алынған) мәні екі есе артады. Сондықтан, нақты байланыс тым жалпылама болған кезде үлкен іріктемелер нөлдік гипотезаны жоққа шығаруға итермелейді. (Бұл принцип туралы көрнекі мысал үшін 10.14-жаттығуды қараңыз).

Іріктеме көлемі мен хи-квадрат мәнінің арасындағы байланыс туралы білуіңіз керек, себебі ол статистикалық мәнділік пен теориялық маңыздылық арасындағы айырмашылықты тағы да жоғарылатады. Бір жағынан, мәнділік тестілері зерттеуде шешуші рөл атқарады. Кездейсоқ іріктемемен жұмыс жасаған кезде зерттеу нәтижелері қарапайым кездейсоқтықтан шыққандығы туралы білуіміз керек.

Екінші жағынан, кез-келген статистикалық әдіс сияқты, гипотезаларды тестілеу олар жауап бере алатын сұрақтар диапазонымен шектелген. Атап айтқанда, бұл тестілер біздің нәтижелеріміз статистикалық тұрғыдан мәнді немесе олай емес екенін көрсетеді. Олар кез-келген басқа мағынада нәтижелер маңызды ма, жоқ па екенін міндетті түрде айта алмайды. Маңызды сұрақтармен тікелей айналысатын болсақ, ассоциация өлшемдері деп аталатын статистикалық әдістердің қосымша жиынтығын қолдануға тиіспіз. Бұл әдістерді осы тарауда, бағанның пайыздық көрсеткіштерін есептегенде қарастырдық, ал ассоциация өлшемдері келесі дәрістерде қарастырылады.

ТҮЙІН

1. Хи-квадратты тестілеуде алдымен ұяшықтарда пайда болатын жиіліктерді анықтаймыз, егер айнымалы мәндер тәуелсіз болса (f_e), осы жиіліктерді бір-бірлеп бақыланатын ұяшықтағы жиіліктермен (f_o) салыстырамыз. Егер нөлдік гипотеза дұрыс болса, күтілетін және бақыланған жиіліктер баллас болуы керек. Бақыланған және күтілетін жиіліктер арасындағы айырмашылық неғұрлым көп болса, нөлдік



Кітап: Статистика негіздері

Дәріс: Тәуелсіздік тесті үшін Хи-квадрат. Бағандардың пайыздық көрсеткіші

гипотезаны жоққа шығару ықтималдығы артады.

2. Хи-квадрат тестінің бірнеше маңызды шектеулері бар. Кестелерде көп (төрт немесе бес) өлшем болған кезде интерпретациялау қиын. Сондай-ақ, іріктеме көлемі (N) азайған сайын, хи-квадрат тесті де сенімсіздеу болады, сондықтан қосымша түзету әрекетті қажетсінеді. Соңында, өте үлкен іріктемелермен тым жалпылама байланыстарды статистикалық тұрғыдан салыстырмалы түрде елеулі деп мәлімдей аламыз. Гипотезаның барлық тестілерінде болғандай, статистикалық мәнділік - кез-келген мағынадағы «маңыздылық» дегенді білдірмейді. Жалпы ереже ретінде, статистикалық мәнділік қажет, бірақ теориялық немесе практикалық маңыздылықтар үшін жеткіліксіз шарт.