



СТАТИСТИКА НЕГІЗДЕРІ

Ықтимал іріктеменің негізгі әдістері мен басты ұғымдары: бас жиынтық, іріктеме, параметрлер, статистика, репрезентативтілік, EPSEM



Кітап: Статистика негіздері

Дәріс: Ықтимал іріктеменің негізгі әдістері мен басты ұғымдары: бас жиынтық, іріктеме, параметрлер, статистика, репрезентативтілік, EPSEM

Біздің күнделіктілігіміз бен тұрмысымыз статистикадан алыс және оның бізге еш қатысы жоқ деп ойлауымыз мүмкін, алайда қарап отырсақ статистика мен телефон қоңырауларының да байланысы бар екен.

Кейде сізге телефон арқылы қоңырау шалып, мазаңызды кетіріп, әр-түрлі тауарлар мен қызметтерді ұсынғаны есіңізде болар. Ендеше бұны - сіздің телефон нөміріңіз кездейсоқ іріктеменің күрделі, кешенді әдістері арқылы таңдалды деп біліңіз. Мұндай іріктеу әдістері коммерциялық мақсаттар мен бұқаралық маркетинг үшін де қолданылады. Бұл сіздің көңіліңізді көншіте қоймас, десек те бұл - іріктеу технологиясының күнделікті өміріміздің құрамдас бөлігіне айналғанын көрсетіп отыр. Тіпті әлеуметтік ғылымдардан тыс салаларда да осы іріктеу технологиясы қолданылатынының анық белгісі.

Сонда ол қалай жүзеге асады? Ендеше тақырыпқа үңілерлік

Өздеріңізбен бүгін жаңа модуль «Инференциалды статистиканы» меңгереміз. Алдағы лекцияларда инференциалды статистика ұғымдары мен әдістері қарастырылады. Жалпы, бұл әдіс мұқият іріктеліп алынған шағын топ (іріктеме) арқылы ірі топтар (бас жиынтық) туралы ақпарат алуға мүмкіндік береді. Бұл статистикалық әдістер – маңызды әрі аса тиімді. Олар қоғамдық пікірді зерттеу, жаңа тұтыну тауарлары үшін нарықтың әлеуетін зерттеу, сайлау нәтижелерін болжау сауалнамалары, ғылыми гипотезалар мен теорияларды тексеруде, сонымен қатар әлеуметтік ғылымдардан тыс басқа да жағдайларда пайдаланылады.

Бүгінгі тақырып аясында іріктеу қысқаша сипатталады. Тарауда қарастырылатын маңызды тақырыптардың бірі – инференциалды статистикадағы жалғыз маңызды концепция – іріктеу үлестірімі. Іріктеу үлестірімі формасы жағынан қалыпты, сондай-ақ бас жиынтық пен іріктеме арасындағы негізгі байланыс.

Тақырыптың мақсаты:

1. Іріктемені бас жиынтыққа генерализациялау тұрғысынан инференциалды статистиканың мақсатын түсіндіресіз.
2. Ықтимал іріктеменің негізгі әдістері мен басты ұғымдарын түсіндіресіз: олар бас жиынтық, іріктелген жиынтық (біз оны іріктеме деп қолданамыз), параметрлер, статистика, репрезентативтілік, EPSEM.
3. Іріктеме үлестірімі, іріктеме және бас жиынтықты ажырата аласыз.
4. Ұсынылған екі теореманы түсіндіре аласыз.

Алдымен дәріс аясында пайдаланатын негізгі ұғымдарды ажыратып кетейік

Орталық шектеу теоремасы. Іріктеменің үлкен болуын ескере отырып, іріктеу үлестірімінің арифметикалық ортасы, іріктеу үлестірімі мен формасын нақтылайтын теорема.

EPSEM. Іріктемелерді таңдау үшін тең ықтималды таңдау әдісі. Бас жиынтықтағы әр элемент пен кейстің іріктемеге таңдалу ықтималдығы тең болуы қажет.

μ . Бас жиынтықтың арифметикалық ортасы.

p . Іріктеме пропорцияларының іріктеу үлестірімінің арифметикалық ортасы.

X . Іріктемелік арифметикалық орта пропорцияларының іріктеу үлестірімінің орташа мәні.

Ықтимал емес іріктеме. EPSEM критерийіне сәйкес келмейтін кез келген іріктеме. Мұндай іріктемелер әлеуметтік зерттеулерде әртүрлі қолданысқа ие, бірақ бас жиынтыққа генерализациялау үшін қолдануға болмайды.

P_s Кез-келген іріктеме пропорциясы.

P_u Кез-келген бас жиынтық пропорциясы.

Параметр. Бас жиынтықтың сипаттамасы.

Репрезентативті іріктеме. Іріктеліп алынған бас жиынтықтың негізгі сипаттамаларын көрсететін іріктеме.

Іріктеу үлестірімі. Нақты көлемдегі ықтимал іріктеме нәтижелерінің барлығы үшін статистиканың үлестірімі. Екі теоремада нақтыланған шарттарға сәйкес, іріктеу үлестірімінің формасы қалыпты болады, оның арифметикалық ортасы бас жиынтықтың арифметикалық ортасына, ал оның стандартты девиациясы (немесе стандартты қатесі) бас жиынтықтың стандартты девиациясын квадрат түбір астындағы N -ге бөлгенге тең.

Қарапайым кездейсоқ іріктеме. Бас жиынтықтан кейстерді іріктеу әдісі, мұндағы барлық кейс пен кейстер жиынтығының іріктелу мүмкіндігі тең болады.

Әлеуметтік ғылымдардағы зерттеу мақсаттарының бірі – түрлі адамдар, топтар, қоғамдар мен тарихи



Кітап: Статистика негіздері

Дәріс: Ықтимал іріктеменің негізгі әдістері мен басты ұғымдары: бас жиынтық, іріктеме, параметрлер, статистика, репрезентативтілік, EPSEM

дәуірлерді пайдалана отырып теорияларды тестілеу. Әрине, теория сан алуан кейстер мен әлеуметтік жағдайлардың тестілеуінен өтсе, оған жоғары сенімділікпен қарауға болады. Дегенмен, әлеуметтік ғылыми зерттеудің басты қиындығы – тестілеуге қажетті халық санының көп болуы. Мысалы, Қазақстандық азаматтардың саяси қалауына қатысты теорияны тестілеу үшін бүкіл электоратты пайдалану дұрыс болар еді. Алайда 11 миллионнан астам барлық ересек қазақстандықтармен сұхбаттасу мүмкін емес. Шынында, қала немесе университет сияқты шағын бас жиынтықтар арқылы сыналатын теориялар үшін де бас жиынтықтағы әрбір кейстен жиналатын ақпаратты жеткізу оңай емес.

Егер бас жиынтыққа қол жеткізу мүмкін болмаса, теорияларды қалай тексере аламыз? Бұл мәселені шешу үшін әлеуметтанушылар зерттегелі отырған бас жиынтықтан іріктеме жасайды немесе кейстердің шағын тобын таңдап алады. Инференциалды статистиканы тарқатудағы басты мақсат – іріктемеден алынатын ақпарат (олар параметрлер деп аталады) негізінде бас жиынтық сипаттамаларын анықтау болады.

Ықтимал іріктеу: Негізгі ұғымдар

Әлеуметтік сала ғалымдары әр түрлі іріктеу әдістерін жетілдірді. Қазір ықтимал іріктемелерді таңдаудың негізгі процедураларына шолу жасаймыз. Ықтимал іріктемелер – бас жиынтыққа генерализациялауда инференциалды статистиканы пайдалануды толық қолдайтын жалғыз іріктеме. Мұндай іріктемелер көбіне кездейсоқ деп сипатталады, аталмыш терминология сізге таңсық емес. Ол кеңінен танымал болғандықтан, келесі тарауларда кездейсоқ іріктеме ұғымын жиі қолданамыз.

Алайда ықтимал іріктеме деп атаған дұрыс болар, себебі күнделікті қолданыста “кездейсоқ” сөзі көбіне «аяқастынан» немесе болжанбайтын деген мағынада қолданылады.

Көріп отырғанымыздай, ықтимал іріктеме ешқандай ретсіздікке жол бермейтіндей, мұқият әрі жүйелі түрде құрастырылады. Сауда орталығында кездесетін адамдармен бір күнде сұхбат жүргізу бір қарағанда «кездейсоқ» болуы мүмкін, бірақ бұл әдіс инференциалды статистика талап ететін нәтиже емес.

Ықтимал іріктемені қарастырудан бұрын айта кетейік: әлеуметтік ғылымдар көбінесе ықтимал емес іріктемені пайдаланады. Мысалы, кішігірім топтардың динамикасы, көзқарастары мен жеке құндылықтарының құрылымын зерттейтін әлеуметтанушы ғалымдар аудиториядағы студенттерді зерттеу пәні ретінде таңдауына болады. Бұл «ыңғайлы» іріктеме көптеген мақсат үшін тиімді (мысалы, ірі жобаға кіріспес бұрын, идеяларды зерделеу немесе сауалнама парақтарын алдын ала тексеру мақсатында), сонымен қатар арзан әрі жеңіл жиналады. Мұндағы басты кемшілік – нәтижелер сыналған топтан басқаларға генерализациялауға жарамайды. Мысалы, жаңсақ пікір теориясы нақты бір университеттегі әлеуметтануға кіріспе кафедрасының белгілі бір бөлімінде оқыған студенттер арасында ғана сыналса, нәтиже басқа адамдарға да қатысты деп айта алмаймыз. Сондықтан, тестілеу нәтижесінде қатаң дәлелдер айтылса да, ықтимал емес іріктеме негізінде сыналған теорияларға үлкен сенімділікпен қарай алмаймыз.

Енді біз іріктемелерді таңдаудағы репрезентативтілік мәселесіне тоқталып өтеміз

Ықтимал іріктемені құрудағы мақсат – кейстерді соңғы іріктеме алынған бас жиынтық үшін репрезентативті болатындай етіп таңдау. Репрезентативті іріктеме дегеніміз – іріктемеде бас жиынтықтың маңызды сипаттамаларының қамтылуы. Мысалы, бас жиынтықтың 60%-ы әйелдер, ал 40%-ы ерлер болса, іріктеменің де құрамы соған пропорционал болуы керек. Негізі, репрезентативті іріктеме – бас жиынтыққа өте ұқсас, бірақ ықшам.

Инференциалды статистика үшін іріктеменің репрезентативті болуы өте маңызды.

Ал іріктеменің репрезентативті екенін қалай білуге болады?

Өкінішке орай, іріктеменің репрезентативті болатынына кепілдік жоқ. Дегенмен, ықтимал іріктеменің түпкі принципін – EPSEM (Таңдау әдісінің тең дәрежедегі ықтымалдығы) қағидасын сақтай отырып, репрезентативті іріктемеге қол жеткізу мүмкіндігін арттыра аламыз. EPSEM-ді ұстану және ықтималдықты мейлінше арттыру үшін іріктеме репрезентативті болуы керек. Яғни бас жиынтықтағы әрбір элемент немесе кейстің таңдау ықтымалдығы тең дәрежеде болатындай етіп іріктемені таңдаймыз. Ойын тасын лақтырғанда 1-ден 6-ға дейін мәнді көрсету ықтималдығы тең болғандай бас жиынтықтың да әр кейсінің іріктемеге іліну мүмкіндігі бірдей болуы тиіс.



Кітап: Статистика негіздері

Дәріс: Ықтимал іріктеменің негізгі әдістері мен басты ұғымдары: бас жиынтық, іріктеме, параметрлер, статистика, репрезентативтілік, EPSEM

EPSEM мен репрезентативтік екі түрлі нәрсе екенін есте сақтаңыз.

Іріктеменің EPSEM-ге сәйкес жасалуы оның популяцияны дәлме-дәл репрезентациялауына кепілдік бермейді. Ықтималдығы өте жоғары EPSEM іріктемесі репрезентативті болады; бірақ, шындығына келгенде, тиынды тастағанда 10 рет қатарынан бір жағымен түскені сияқты, EPSEM іріктемесі де бас жиынтықты сипаттауда қателесуі мүмкін. Инференциалды статистиканың артықшылықтарының бірі – ол зерттеушіге мұндай қатенің ықтималдығын бағалауға және нәтижелерді тиісті деңгейде түсіндіруге мүмкіндік береді.

Енді Қарапайым кездейсоқ іріктеме туралы айтып өтелік

Қарапайым кездейсоқ іріктеме бастапқы EPSEM іріктеу әдісі арқылы жасалады. Бұл әдістің бірқатар нұсқалары мен толықтырулары бар, бірақ біз ең қарапайым қолданыстарына ғана тоқталамыз.

Қарапайым кездейсоқ іріктемені жасау үшін бас жиынтықтағы барлық кейстің, яғни әрқайсысының өзіндік ерекшелігі (идентификатор) бар кейстердің тізімі қажет. Әдетте ол – сан, сонымен қатар әр кейстің іріктемеге ілінуіне кепілдік беретін іріктеу жүйесі болуы шарт. Таңдау процесі тиынды айналдыру, ойын тасын лақтыру немесе қалпақтардан сандарды алып шығудан тұруы мүмкін, ал әлеуметтік зерттеулерде кездейсоқ сандардың тізімін жасау үшін компьютерлік бағдарламаларды қолданамыз.

Тізімде кейстің идентификациялық нөмері шықса, ол іріктемеге таңдалады. Егер номерлер тізімі шынымен кездейсоқ болса, кез-келген нақты ID нөмірі басқа ID нөмірге ұқсас болады, сәйкесінше іріктеу процесі EPSEM іріктемесін береді. Сізге қажет іріктеме өлшеміне қол жеткізген кезде кейстерді таңдауды тоқтатыңыз; егер идентификациялық нөмірі бірнеше рет таңдалса, қайталанғандары таңдалмайды.

Инференциалды статистиканың мақсаты – іріктеме арқылы жинақталған ақпаратқа сүйене отырып, бас жиынтық туралы мәлімет алу. Инференциалды статистиканың әр әдісі іріктемелердің EPSEM қағидасына сүйене отырып таңдалуын талап етеді. Тіпті, ең ұқыпты және жетілдірілген іріктеу әдісінің өзі репрезентативтікке кепілдік бермейді, тек ықтималдығы жоғары EPSEM іріктемелері таңдалынып алынған бас жиынтық үшін репрезентативті болады.

Іріктеу үлестірімі

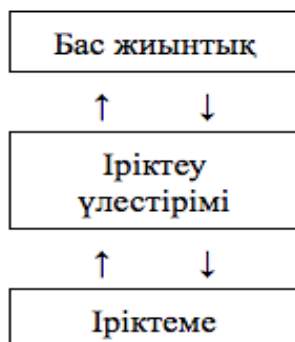
EPSEM процедураларына сәйкес ықтимал іріктемені таңдаудан не ұтамыз? Бір жағынан, іріктемедегі кейстерден мол мәлімет жинай аламыз. Екінші жағынан, бас жиынтық туралы ештеңе білмейміз. Негізгі бас жиынтық туралы ақпарат болса, бәлкім, іріктемеге мұқтаж болмас едік. Бас жиынтық туралы көбірек ақпарат алу үшін инференциалды статистиканы қолданамыз, ал іріктемеден алынған ақпарат, ең алдымен, бас жиынтыққа генерализациялауға мүмкіндік беретіндіктен маңызды.

Негізгі, инференциалды статистиканы пайдаланған кезде іріктемедегі бір айнымалыны өлшейміз (ол, мысалы, жасы, саяси көзарасы немесе жасанды түсік тастау туралы пікірі болуы мүмкін), содан кейін ғана бас жиынтықтан сол айнымалы туралы білу үшін іріктемедегі мәліметті пайдаланамыз. Өткен дәрістерден айнымалыны дұрыс сипаттау үшін үш түрлі ақпарат қажет екенін білдіңіз. Олар:

1. Оның үлестірімінің формасы
2. Орталық тенденция өлшемі
3. Дисперсия өлшемі.

Әрине, ақпараттың үш түрінің барлығы іріктемедегі кейстер айнымалысы бойынша жинақталуы (немесе есептелуі) мүмкін. Нақтырақ айтқанда, бас жиынтық үшін ешқандай ақпарат жоқ. Сирек жағдайларды қоспағанда (мысалы, IQ тесттері бас жиынтық шамаларының үлестірімі шамамен қалыпты болатындай жасалған, ал кіріс үлестірімі әрқашан ассиметриялы) бас жиынтықтағы айнымалы үлестірімінің нақты формасы туралы ешқандай ақпарат жоқ. Сондай-ақ, бас жиынтықтағы айнымалылардың арифметикалық ортасы мен стандартты девиациясы да белгісіз. Бас жиынтық үшін осы ақпарат болған жағдайда, инференциалды статистика қажет болмас еді.

6.1-сызба. Іріктеме, іріктеу үлестірімі мен бас жиынтық арасындағы байланыс



Статистикада іріктемеден алынған ақпаратты іріктеу үлестіріміндегі бас жиынтыққа байланыстырамыз: Белгілі бір көлемдегі (N) ықтимал іріктемелер үшін статистиканы теориялық, ықтимал үлестіру. Яғни, іріктеу үлестірімі – бас жиынтықтағы кейстердің барлық мүмкін комбинацияларына негізделген статистика үлестірімі (мыс., арифметикалық орта немесе пропорция). Іріктеу үлестірімі туралы маңызды мәселе – оның сипаттамалары эмпирикалық ақпаратқа емес, ықтималдық заңдарына негізделеді, бұл кеңінен танымал. Негізінен, іріктеу үлестірімі – инференциалды статистикадағы басты ұғым, ендеше оның сипаттамаларын ұзақ зерделеу міндетті.

Инференциалды статистикада пайдаланылатын үш үлестірім. 6.1-сызбада көрсетілгендей, инференциалды статистиканың барлық тәжірибелік қолданысы іріктеу үлестірімі арқылы іріктеме мен бас жиынтық арасында қозғалады/өзгереді. Осылайша, айнымалының бір-бірінен айырмашылығы бар үш бөлек үлестірімі инференциалды статистиканың барлық қолданысына кірістіріледі:

1. Айнымалының бас жиынтығының үлестірімі – эмпирикалық болғанда белгісіз. Бас жиынтыққа қатысты ақпарат жинау немесе тұжырым жасау – инференциалды статистиканың жалғыз мақсаты.
2. Айнымалының іріктеу үлестірімі – эмпирикалық емес немесе теориялық. Ықтималдық заңдарына байланысты осы үлестірім туралы көп нәрсе белгілі. Атап айтқанда, үлестірімнің формасын, орташа мәні мен дисперсиясын есептеуге болады, сондықтан үлестірім тиісінше/дұрыс сипатталады.
3. Айнымалы іріктемесінің үлестірімі – эмпирикалық (яғни, шынайы өмірде бар) және белгілі. Іріктеме үшін айнымалының формасын, орталық тенденция өлшемі мен дисперсиясын есептеуге болады. Дегенмен іріктемеден алынған ақпарат ең алдымен зерттеушіге бас жиынтық туралы мәлімет алуға мүмкіндік бергенде ғана маңызды.

Іріктеу үлестірімінің тиімділігі оның анықтамасымен түсіндіріледі.

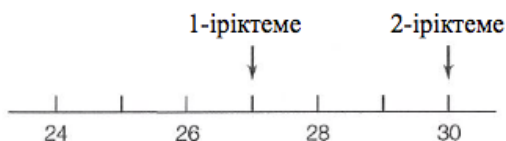
Ол барлық ықтимал үлгі нәтижелерінен алынған статистиканы қамтиды. Сондықтан, іріктеу үлестірімі кез-келген нақты іріктеме нәтижесінің ықтималдығын бағалауға мүмкіндік береді. Бұл осы және келесі тақырыптарда назар аударатын процесс.

Іріктеу үлестірімін құрастыру.

Іріктеу үлестірімі – теориялық, яғни ол ешқашан нақты құрастырылмайды. Дегенмен, үлестірімнің құрылымы мен функциясын жақсырақ түсіну үшін оны қалай құруға болатынын талдайық. Мысалы, 10 000 адамнан тұратын нақты аумақтың жасы туралы ақпарат жинағымыз келеді. 100 тұрғыннан құралған EPSEM іріктемесін жасап, 100 респонденттің барлығынан олардың жасын сұраймыз, әрі қарай орташа жас – 27-ні есептеу үшін сол адамдардың жауабын (шама) пайдаланамыз. Осы шама 6.2-суретте келтірілген. Бұл іріктеме – 10 000 адамнан тұратын бас жиынтықтан алынған 100 адамның сансыз ықтимал комбинацияларының бірі, ал статистика (арифметикалық орта – 27) ықтимал іріктеме нәтижелерінің миллионнан бірі екеніне назар аударыңыз.

Енді алғашқы 100 респонденттің орнына тағы сол іріктеменің көлеміндей ($N = 100$) басқа іріктеме жасап, қайтадан орташа жасын есептеңіз. Мысалы, екінші іріктеменің арифметикалық ортасы – 30, осы іріктеме нәтижесін өз беттеріңізбен есептеп көріңіз.

6.2-сызба. Іріктеу үлестірімін құрастыру



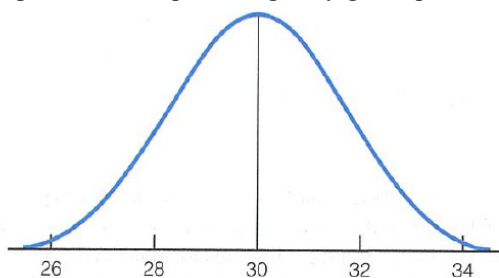
Екінші іріктеме 10 000 адамдық бас жиынтықтың кездейсоқ комбинациясынан алынған 100 адамнан тұрады, ал 30-ға тең арифметикалық орта – іріктеменің статистикалық деректерінің миллионнан бірі. Осы респонденттердің орнына тағы бір үлгі жасаңыз, арифметикалық ортасын есептеп, жазып алыңыз, оны үшінші іріктемемен ауыстырыңыз, әрі қарай осы операцияларды шексіз қайталай отырып, әрбір іріктеменің арифметикалық ортасын есептеп, жазып алып, төртінші іріктемені жасаңыз. Енді іріктемелердің мыңынан оны жинақталып, әрбір іріктеменің арифметикалық ортасы есептелген соң, 6.2-сурет қандай болатынын елестетіп көріңіз. 10 000 популяциядан 100 респонденттің барлық ықтимал комбинациясын жинағаннан кейін іріктемелік арифметикалық орта үлестірімінің формасы, арифметикалық ортасы және стандартты девиациясы қандай болады?

Біріншіден, әр іріктеменің басқасынан, кем дегенде болар-болмас айырмашылығы бар екені айқын. Өйткені екі рет дәл сол 100 адамды таңдау мүмкін емес. Әрбір іріктеме бірегей болатындықтан, әр іріктемелік арифметикалық орта да әр түрлі болады.

Сонымен қатар EPSEM-ді қатаң ұстанғанымызға қарамастан, барлық іріктеменің репрезентативті болмайтыны да шындық. Мысалы, 100 адамнан тұратын іріктемені алуды жалғастыра берсек, ең соңында ең жас тұрғындардан тұратын іріктемені таңдап аламыз. Мұндай іріктемелік арифметикалық орта бас жиынтыққа қарағанда әлдеқайда төмен болады. Сол сияқты кейбір іріктемелерімізде қарт адамдар ғана болады және олар бас жиынтықтың арифметикалық ортасынан әлдеқайда жоғары. Дегенмен, мұндай репрезентативті емес іріктемелер сирек кездеседі, ал іріктемелік арифметикалық ортаның көбі бас жиынтықтың шынайы мәні төңірегінде топтасады.

Әрі қарай мысал келтіретін болсақ, 10 000 адамның шынайы орташа жасы 30 екенін білдік. Себебі, бұған дейін көргеніміздей, іріктемелік арифметикалық ортаның көбі шамамен 30-ға тең, ендеше осы іріктемелік арифметикалық орталардың іріктеу үлестірімінің шыңы 30 болуы керек. Іріктемелердің кейбірі репрезентативті емес және олардың арифметикалық ортасы «көздеген максатына жетпейді», алайда 30-ға тең негізгі арифметикалық ортадан алыстаған, мұндай қателіктер жиілігі азаюы тиіс. Яғни, бас жиынтық мәнінен алыстаған сайын үлестірім негізіне қарай қисаюу тиіс – 29-дан 31-ге дейінгі аралықтағы іріктемелік арифметикалық орта ортақ болуы тиіс; 20 немесе 40-қа тең орташа мәндер сирек болуы керек. Іріктемелер кездейсоқ болғандықтан, олардың арифметикалық ортасы бас жиынтық мәнінің екі жағынан бірдей болуы керек, соған байланысты үлестірімнің өзі де симметриялы болуы шарт. Қысқасы, барлық ықтимал іріктемелік арифметикалық орталардың іріктеу үлестірімі шамамен қалыпты болуы керек және 6.3-сызбада ұсынылған үлестірімге ұқсас болады. Кез-келген қалыпты қисық үшін арифметикалық ортаға жақын кейстер (мысалы, ± 1 стандартты девиация ішінде) жиі кездеседі, ал арифметикалық ортадан алыс кейстер (мысалы, 3 стандартты девиациядан тыс) сирек кездесетіні 5-тарауда айтылған. (Іріктеу үлестірімдерін құрастыру үшін 6.20-жаттығуды қараңыз).

6.3-сызба. Іріктемелік арифметикалық ортаның іріктеу үлестірімі



Екі теорема.

Іріктеу үлестірімінің формасы туралы жүйелі түсінік, сонымен қатар орталық тенденция мен дисперсия



өлшемдері туралы маңызды ақпарат екі теоремада берілген. Теореманы қарастырмас бұрын, пайдаланатын символдарды шолып шығайық. Еске түсірейік, іріктемелік арифметикалық орта – X , ал бас арифметикалық орта – арқылы белгіленеді. Дисперсияны өлшеу үшін іріктеменің стандартты девиациясы – s символын пайдаланамыз, ал бас жиынтықтың стандартты девиациясын σ (Грек әрпі кіші сигма) символымен белгілейміз.

Бұл теоремалардың бірі былай тұжырымдалады:

N көлемді қайталанған кездейсоқ іріктеменің арифметикалық ортасы – μ (мю), ал стандартты девиациясы – σ (сигма) қалыпты популяциядан алынса, іріктемелік арифметикалық ортаның іріктеу үлестірімі қалыпты болады, яғни мұндағы арифметикалық орта – μ және стандартты девиация – σ / \sqrt{n} тең.

Түсіндірмесі: Егер бас жиынтыққа қалыпты үлестірілген белгімен бастап, (мысалы, IQ көрсеткіштерін алуға болады), әрі қарай сол бас жиынтықтан тең көлемді кездейсоқ іріктемелердің шексіз санын алсақ, іріктемелік арифметикалық ортаның іріктеу үлестірімі қалыпты болады. Егер біз айналымы бас жиынтыққа қалыпты үлестірілгенін білсек, іріктеу үлестірімі қалыпты болады дей аламыз.

Дегенмен іріктеу үлестірімінің формасына қарағанда теорема көп нәрседен хабардар етеді. Сонымен қатар, ол арифметикалық орта мен стандартты девиацияны анықтайды. Шындығында, іріктеу үлестірімінің арифметикалық ортасы бас жиынтықтың арифметикалық ортасының мәнімен бірдей. Яғни егер тұтас бас жиынтықтың орташа IQ-і – 100 болса, іріктеу үлестіріндегі іріктелген кез-келген IQ-дың арифметикалық ортасы да 100-ге тең болады.

Неліктен бұл екі арифметикалық орта бірдей мәнге ие екені дәл осы деңгейде толық түсіндірілмейді. Алайда, іріктемелік арифметикалық ортаның көбі ұзақмерзімді уақытқа бас жиынтықтың мәні айналасында топтасады. Олай болса, осы екі мәндің тең болғаны туралы факттің өзі интуициялық болады. Ал дисперсияға келер болсақ, теорема бойынша іріктеу үлестіріндегі стандартты девиация, яғни арифметикалық ортаның стандартты қателігі N -нің квадрат түбіріне бөлінген бас жиынтықтың стандартты девиациясына тең (символдық түрде ол былай беріледі: σ / \sqrt{n}).

Егер бас жиынтықтағы қалыпты таралған айналымының арифметикалық ортасы мен стандартты девиациясы белгілі болса, теорема арқылы іріктеу үлестірімінің арифметикалық ортасы және стандартты девиациясын есептеуге жол ашылады. Осылайша, кез-келген эмпирикалық үлестірім туралы білетініміздей, енді іріктеу үлестіріміне де қатысты соншалықты нақты білетін боламыз (мысалы, формасы, орталық тенденциясы және дисперсиясы болуы мүмкін).

Бірінші теоремаға сәйкес, айналымы бас жиынтыққа қалыпты үлестірілуі тиіс. Қарастырып отырған айналымы белгісіз немесе формасы жағынан қалыпты емес болса ше (мысалы, әрдайым оң асимметриялы кіріс)? Мұндай (шын мәнінде, өте кең тараған) ықтималдықтар орталық шектік теорема деп аталатын екінші теорема аясында қарастырылады:

Егер N көлемді қайталанатын, бір-біріне ұқсас кездейсоқ іріктемелердің арифметикалық ортасы – μ , стандартты девиациясы – σ қалыпты бас жиынтықтан алынса, онда N үлкен болған сайын, іріктемелік арифметикалық ортаның іріктеу үлестірімі қалыптыға жақын болады, мұндағы арифметикалық орта – μ және стандартты девиация – σ / \sqrt{n} тең.

Түсіндірмесі: Тіпті айналымы бас жиынтық бойынша қалыпты үлестірілмесе де, кез-келген айналымы үшін іріктеме көлемі артқан сайын, іріктемелік арифметикалық ортаның іріктеу үлестірімі қалыптыға жақын бола бастайды. Егер N үлкен болса, онда іріктеу үлестірімінің арифметикалық ортасы бас жиынтықтың арифметикалық ортасына тең болады, ал оның стандартты девиациясы (немесе арифметикалық ортаның қалыпты қателігі) σ / \sqrt{n} өрнегіне тең болады.

Орталық шектік теорема – маңызды, себебі мұнда айналымы бас жиынтықта қалыпты үлестірілуі керек деген шарт жоқ. Іріктеме көлемі үлкен болған сайын, үлестірімнің арифметикалық ортасы бас жиынтықтың арифметикалық ортасына тең болады, стандартты девиациясы σ / \sqrt{n} өрнегіне тең болады. Ендеше іріктеу үлестірімі формасы жағынан қалыпты деп ұйғара аламыз. Олай болса тіпті егер (кіріс сияқты) симметриялы емес үлестірімді айналымымен жұмыс жасасақта, біз әлі де қалыпты іріктеу үлестірімі деп болжауымызға болады.

Ең соңғысы, үлкен іріктеме нені білдіретінін анықтап алайық.

Басты ереже – іріктеме көлемі (N) 100-ге тең немесе одан артық болса, Орталық шектік теорема қолданылады. Яғни, іріктемелік статистиканың іріктеу үлестірімінің формасы қалыпты деп болжауға болады. Егер N 100-ден кем болса, іріктеу үлестірімін қалыпты деп қабылдас бұрын, бас жиынтықтың қалыпты үлестірімінің жақсы дәлелі болуы керек. Ендеше, іріктемелер 100-ден артық болған сайын іріктеу үлестірімі қалыпты болады деп ұйғаруымызға болады.



Іріктеу үлестірімі ұғымын (яғни, ол не, неге маңызды екенін) жеткізу – оңай шаруа емес. Осы орайда іріктеу үлестіріміне қатысты бірқатар маңызды нүктелерін қойып шығайық:

1. Анықтамасы: Іріктеу үлестірімі дегеніміз белгілі бір көлемдегі ықтимал іріктеме нәтижесіне арналған статистиканың үлестірімі (мысалы, арифметикалық орта немесе пропорция). Іріктеме мен бас жиынтыққа қарағанда, іріктеу үлестірімі теориялық болады, осы лекцияда ұсынылған теориялардың арқасында оның формасы, орталық тенденция өлшемін және дисперсиясы белгілі.
2. Формасы: қалыпты (5-тарау мен А аппендиксін қараңыз). · Іріктеме «үлкен» ($N > 100$ -ден әлдеқайда артық), сондықтан іріктеу үлестірімінің формасы қалыпты болады. Бұл бас жиынтықтағы айнымалының формасына қарамастан шын/әділетті болады.
3. Орталық тенденция және дисперсия өлшемі: Іріктеу үлестірімінің арифметикалық ортасымен бас жиынтықтың арифметикалық ортасының мәні бірдей болады. Егер барлық ересек қазақстандықтар орта есеппен 11 жылда ($=11$) мектеп бітірсе, іріктеудің де арифметикалық ортасы 11 болады.
4. Іріктеу үлестірімінің стандартты девиациясы (немесе орташа мәнінің қалыпты қателігі) бас жиынтықтың стандартты девиациясын (кіші сигма) квадрат түбір астындағы N -ге бөлгенге тең.
5. Инференциалды статистикадағы іріктеу үлестірімінің рөлі: Іріктемені бас жиынтықпен байланыстырады (6.1-суретті қараңыз).

Осылайша, теоремалар іріктеу үлестірімінің (форма, орталық тенденция өлшемі және дисперсия) статистикалық сипаттамалары туралы мәлімет береді, ал бұл ақпарат бас жиынтық пен іріктемені байланыстыруға мүмкіндік береді.

Қорытындылай келе, мақсат – бас жиынтық туралы ақпарат беру (GSS жағдайында, барлық ересек америкалық). Бас жиынтық тестілеу үшін тым үлкен болған кезде, ақпаратты бас жиынтықтың сипаттамаларын бағалау үшін мұқият жасалған кездейсоқ іріктемеден аламыз.

Осы тұжырымдарды бекіту үшін оқулықтағы SPSS жаттығуларына арналған дерекқорлардың бірі – Жалпы Әлеуметтік сауалнаманы (General Social Survey) пайдалана отырып, қосымша мысалды қарастырулыраңызға болады. Жалпы әлеуметтік сауалнама (GSS) 1972 жылдан бері ересек америкалықтардың кездейсоқ таңдалған іріктемесін қолданады. Ол әртүрлі мәселелердің кең ауқымын зерттейді. GSS-тің кемшіліктері де бар, дегенмен ол теорияны сынау мен сол қоғамды көбірек білудегі өте бағалы ресурс.

Бас жиынтықтан немесе зерттеу нысанына алынған топтан бастайық. GSS жағдайында, бас жиынтық (18 жастан жоғары) 235 миллион америкалықтан тұрады. Әрине, олардың барлығымен сұхбат жүргізіп, олардың қандай екені немесе аборт жасау, қару-жарақты бақылау, орта мектептердегі жыныстық тәрбие немесе кез-келген басқа мәселе туралы ойын білу мүмкін емес. Дегенмен, осы ақпараттың маңызы бар екенін ескеруіміз керек. Бұл қоғамдық талқылауларды ақпараттандыруға көмектесіп, көптеген даулы мәселелерді талқылаудың негізі болады (мысалы, сауалнамалар тұрақты түрде америкалықтардың көпшілігі қаруды бақылаудың кейбір түрлеріне қолдау көрсететінін анықтайды) және адамдарға жеке сенімдерін түсіндіруге көмектеседі. Егер ақпарат құнды, қажет болса, үлкен бас жиынтық туралы қалайша көбірек ақпарат алуға болады?

Бұл бас жиынтықтың мұқият таңдалған топтамасына, яғни іріктемеге алып келеді.

GSS бірнеше мың адам үшін ұйымдастырылған, олардың әрқайсысы EPSEM қағидасына негізделген күрделі технологиямен таңдалады. Белгілі бір бас жиынтықтан таңдалып алынған іріктеме сол жиынтық үшін репрезентативті болады. Іріктемеге қатыстының барлығы бас жиынтыққа да қатысты болады (әлбетте, кейбір кемшіліктер және ескертпелер де баршылық).

Респонденттермен сұхбат жүргізіп, олардың шығу-тегі (ұлты, діні, жасы, жынысы, білім алу жылдары және т.б.) туралы ақпарат, сондай-ақ пікірлері мен көзқарастары сұралады. Осындай ақпарат толық жинақталғанда, GSS дерекқоры іріктемедегі адамдарға қатысты жүздеген айнымалы бойынша (жасы, жаңсақ пікір деңгейі, отбасылық жағдайы) мәліметті (формасы, орталық тенденция өлшемі, дисперсия өлшемі) қамтиды. Сонымен, іріктеме үшін айнымалылар туралы мол ақпарат қолымызда бар дерлік (олар шынымен сауалнамаға жауап берген адамдар), бірақ бас жиынтық үшін бұл айнымалылар туралы ешқандай мәлімет жоқ (яғни бас жиынтық деп отырғанымыз бұл жағдайда 235 миллион ересек америкалықтар).

Белгілі іріктеме арқылы белгісіз бас жиынтық туралы қалай ақпарат ала аламыз? Бұл – инференциалды статистикадағы басты мәселе; ал оның жауабы – «іріктеу үлестірімін қолдану».



Күнделікті өмірдегі СТАТИСТИКА	<i>АҚШ-ты басқа елдермен салыстыру үшін сауалнамаға жүзінү</i> <i>Әлемдік құндылықтар сауалнамасы (http://www.worldvaluessurvey.org/) 1981 жылдан бері көптеген елдің кездейсоқ іріктемелерін қолданады. Соңғы зерттеу нәтижелері АҚШ-тың гомосексуализмге қатысты пікірінің бір елге қарағанда толерантты, басқа елге қарағанда толерантты емес екенін көрсетті.</i>	
	«Сіздің ойыңызша, гомосексуализм – әрқашан да негізі бар, ешқашан да негізі жоқ немесе осы екеуінің ортасы ма?»	
	Елдер	“Ешқашан да негізі жоқ» жауабының пайыздық көрсеткіші
	Ресей	54,1%
	Қытай	49,4%
	АҚШ	24%
	Германия	17,8%
Жапония	17,6%	
<i>Дереккөз: World value Survey. http://www.worldvaluessurvey.org/. Results are for wave 6 of the survey, 2010-2014.</i>		

Символдар және терминология

Келесі лекцияларда мүлде бөлек үш үлестірімдермен жұмыс істейміз. Сонымен қатар, іріктеу үлестірімінің бірнеше түрін негізге аламыз: іріктеменің арифметикалық ортасының іріктеу үлестірімі және іріктеме пропорцияларының іріктеу үлестірімі.

Осындай әртүрлі үлестірімдерді ажырату үшін көбіне-көп символдарды пайдаланамыз. 6.1-кестеде іріктеу үлестірімі үшін қолданылатын символдар берілген. Негізінен іріктеу үлестірімі іріктемелік статистикаға сәйкес грек әріптерімен белгіленеді.

6.1-кесте. Үш үлестірімнің арифметикалық орталары және стандартты девиацияларының символдары

	Арифметикалық орта	Стандартты девиация	Пропорция
1. Іріктемелер	\bar{X}	s	P_s
2. Бас жиынтықтар	μ	σ	P_μ
3. Арифметикалық ортаның іріктеу үлестірімі Пропорциялардың іріктеу үлестірімі	μ_x	σ_x	
	μ_p	σ_p	

Іріктеменің арифметикалық ортасы мен стандартты девиациясы ағылшын әріптерімен (\bar{X} және s деп) белгіленеді, ал бас жиынтықтың арифметикалық ортасы мен стандартты девиациясы грек әріптерімен (μ және σ) таңбаланады. Іріктеме бойынша есептелетін пропорциялар P_s , (мысалы, P -төменгі- s ; іріктеме үшін s), ал бас жиынтықтың пропорциялары P_μ ретінде (мысалы, P -төменгі- μ ; популяция немесе “жер әлем” үшін μ) бейнеленген. Іріктеу үлестірімінің символдары – ағылшын бас әріптерімен бірге грек әріптері арқылы таңбаланады. Іріктеменің арифметикалық ортасының іріктеу үлестірімінің арифметикалық ортасы мен стандартты девиациясы – μ_x («мю-төменгі- x -сызықты») және σ_x («сигма-төменгі- x -сызықты»). Іріктемелік пропорциялардың іріктеу үлестірімінің арифметикалық ортасы мен стандартты девиациясы – μ_p («мю-төменгі- p ») және σ_p («сигма-төменгі- p ») болады.

Түйін

1. Бас жиынтық тестілеу үшін әрдайым тым үлкен болғандықтан, әлеуметтік зерттеулердің фундаменталды стратегиясы – нақты бас жиынтықтан іріктемені таңдау, содан кейін іріктелген жиынтық арқылы жинақталған деректерді бас жиынтыққа генерализациялау. Бұл бағалау (есептеу) арқылы немесе гипотезаны сынау арқылы жүзеге асырылады.



Кітап: Статистика негіздері

Дәріс: Ықтимал іріктеменің негізгі әдістері мен басты ұғымдары: бас жиынтық, іріктеме, параметрлер, статистика, репрезентативтілік, EPSEM

2. EPSEM қағидасына сүйене отырып, бас жиынтық тізімінен кейстерді іріктеу арқылы қарапайым кездейсоқ іріктемелер жасалады (әр кейстің іріктелуықтималдығы бар). EPSEM қағидасы негізінде таңдалған іріктемелердің репрезентативті болуы ықтималдығы өте жоғары.

3. Іріктеу үлестірімі – инференциалды статистикадағы негізгі ұғым, барлық ықтимал іріктеме нәтижелерінің теориялық үлестірімі. Оның жалпы формасы, арифметикалық ортасы мен стандартты девиациялары белгілі болғандықтан (екі теоремада көрсетілген шарттар негізінде), зерттеу іріктеу үлестірімін тиісті түрде сипаттап және қолдана алады.

4. Осы тарауда берілген екі теорема бойынша, қызығушылық танытқан айнымалы бас жиынтықта қалыпты үлестірілген немесе іріктеме көлемі үлкен болса, іріктеу үлестірімінің формасы қалыпты болады, оның арифметикалық ортасы бас жиынтықтың арифметикалық ортасына, ал оның стандартты девиациясы (немесе стандартты қатесі) бас жиынтықтың стандартты девиациясын квадрат түбір астындағы N -ге бөлгенге тең болады.