



20-дәріс



ҚАЗАҚСТАННЫҢ  
АШЫҚ  
УНИВЕРСИТЕТІ

# ҒЫЛЫМИ ЗЕРТТЕУ ЖҮРГІЗУ ӘДІСТЕМЕСІҚАТЫНАСҚА

## Сандық деректерді талдау, 1





Кітап: Ғылыми зерттеу жүргізу әдістемесі

Дәріс: Сандық деректерді талдау, 1

Жиналған деректерді екі түрлі жолмен талдауға болады: бірі сандық тәсілдер арқылы болса, екіншісі сапалық тәсілдер арқылы.

Сандық деректер дегеніміз, аты айтып тұрғандай абстрактілі тұжырымдарды санға айналдыру арқылы, жасаған түйіндеріңізді сандар арқылы көрсету. Деректерді санға айналдыру қазір салыстырмалы түрде айтқанда әлдеқайда жеңіл болды. Өйткені қазіргі электронды дәуірде формулаларды есте сақтап, олармен қолдануды үйреніп жатудың қажеті жоқ. Өйткені оның бәрін біз үшін компьютер жасайтын болды. Дегенмен оның сіз естен шығарып алуыңыз мүмкін бір жағы бар. Компьютер ол адам емес, және ол адамның басымен жасай алатын көп функцияларды атқара алмайды. Сондықтан компьютер арқылы жасалатын дүниелердің өзі де адам тарапынан білімнің болуын талап етеді. Өздеріңіз білетіндей, сандар нақтылықты, әрбір таңбаның, әрбір символдың анық, қатесіз болуын талап етеді. Бір санда жіберілген қателік сіздің барлық жұмысыңызды қате етіп жіберуі мүмкін. Сондықтан зерттеуші статистикалық талдау жүргізуде осы жағдайлардың бәрінің қатесіз жүріп отырғанын өз бақылауында ұстай алуы тиіс. Былайша айтқанда, компьютер өз алдына бөлек, сіз өз алдына бөлек болмай, ол сіздің қолыңыздың жалғасы ретінде болғаны дұрыс.

Деректерді санға айналдырған кезде мына сұрақтарды естен шығармаңыз:

Өзіңіз күтетін нәтижелерге байланысты сұрақтар

- Мен не табуға үміттемін, яғни менің гипотезам растала ма?
- Мен нені табамын деп үміттенбеймін және мен оны қалай іздей аламын?
- Менің нәтижелерім балама жолдармен түсіндіріле ала ма? Оның салдары қандай болады?

Зерттеу сұрағына, мақсаттарына және міндеттеріне байланысты сұрақтар:

- Өзімнің зерттеу сұрақтарымды ең жақсы жолмен сұрай алуым үшін мен өз деректеріммен қалай жұмыс істегенім дұрыс?
- Менің нәтижелерім менің зерттеу сұрақтарыммен, мақсаттарыммен және міндеттеріммен қалай байланысты?

Теориямен байланысты сұрақтар:

- Менің нәтижелерім менің теорияларымды растай ма? Қалай? Неге? Неге растамайды?
- Менің теориям менің нәтижелерімді түсіндіруде хабар/көмек бере ме? Қандай жолдармен?
- Мен күтпеген нәтижелер балама теориялармен байланыса ала ма?

Әдістерге байланысты сұрақтар

- Менің деректерді жинақтау және/немесе талдау әдістерім менің нәтижелеріме бояу берді ме? Егер солай болса, қандай жолдармен?
- Менің әдістемелік кемшін тұстарым менің нәтижелеріме қалай әсер етуі мүмкін?

Енді сізде мынандай сұрақ туындауы мүмкін. Мен статистиканы түсінбеймін, оны түсініп жатуға уақытымды кетіргім келмейді, ең дұрысы маған тәжірибелі маманға тапсырыс беріп, соның көмегімен статистикалық талдауларымды жасата салу деген сияқты ой келуі мүмкін. Бірақ есіңізде болсын, сіз жинаған деректерден сізге қандай статистикалық дерек керек екенін маманға түсіндіру үшін де сізде азды-көпті статистика туралы түсінік болуы керек. Екіншіден, ол маман сізге жасап берген статистикасы қаншалықты дұрыс жасалғанын тексеру үшін де, қандай да бір қателікті көрген жағдайда оны түзетуді сұрау үшін де статистика тілінен хабардар болуыңыз тиіс. Сондай-ақ, маман жасап берген статистикалық деректермен жұмыс істеу үшін де статистикадан хабардар болуға міндеттісіз. Әйтпесе, ол маман қаншалықты керемет жұмыс істеп, сізге керемет талдау жасап бергенімен, сіз оны түсіне алмайтын болсаңыз, оны талдай да алмайсыз. Сондықтан өміріңізде ешқашан статистикамен айналыспайтыныңызға сенімді болсаңыз да, одан хабардар болғаныңыз өзіңізге жақсы.

## Статистикалық талдау жүргізу

Статистикамен айналысқанда сіз мыналарды білуіңіз керек:

- 1) өз деректеріңізбен қалай жұмыс істеу керектігін;
- 2) айнымалылардың табиғатын;
- 3) дескриптивті және инференциалды статистиканың екеуінің де рөлі мен қызметін;



- 4) статистикалық тесттерді орнымен қолдануды;
- 5) деректерді тиімді түрі көрсете алуды.

Бірақ шындығында келгенде статистикамен жұмыс істеу деген компьютермен жұмыс істеу екенін естен шығармаңыз. Статистикалық бағдарламалар бар, бірақ олармен жұмыс істеуді үйрену үшін сіз компьютерге отыруыңыз керек. Ұзақ уақыт бойы бағдарламаның ана жерін бір түртіп, мына жерін бір түртіп, ойдан шығарылған деректерді енгізіп көру арқылы онымен жұмыс істеуді үйренуіңіз керек. Бұл жүйелі түрде жаттығуды талап ететін және көбіне жеке отырып, ұзақ үйренуді талап ететін нәрсе. Кейбір студенттер оны бірден түсініп алады да, ешкімнің көмегінсіз-ақ оны игеріп кетіп жатады. Ал кейбіреулер бағдарламадан хабары бар адамдардың көмегіне жүгінуге мәжбүр болады. Сондай-ақ ютубтен де сіз ондай бағдарламамен жұмыс істеуге үйрететін сабақтарды таба аласыз.

## ДЕРЕКТЕРМЕН ЖҰМЫС ІСТЕУ ЖӘНЕ АЙНЫМАЛЫЛАРДЫ АНЫҚТАУ

Тапқан деректермен жұмыс істеу 5 сатыдан тұрады.

### 1. Сізге ыңғайлы болатын бағдарламамен танысыңыз

Статистикамен жұмыс істеуге мүмкіндік беретін ең жиі қолданылатын бағдарламалар бар. Олардың ең кең тарағандары:

- IBM SPSS statistics – күрделі және қолдану оңай ([www.spss.com](http://www.spss.com));
- SAS – көбіне институтталған стандарт, бірақ кейбіреулер оны SPSS сияқты қолдану оңай емес дейді ([www.sas.com](http://www.sas.com));
- Minitab – кіріспеге көбірек келеді, үйренушілерге және шағын деректер жиынтығы үшін жақсы ([www.minitab.com](http://www.minitab.com));
- Excel – статистикаға арналып жасалмаған бағдарлама бола тұра, ол негізгі тапсырмаларды шеше алады және барлық-дерлік ЖК-де (жеке компьютер) (Microsoft Office өнімі) бар;
- R – статистикалық есептер мен графиктер үшін тегін бағдарламалық орта ([www.r-project.org](http://www.r-project.org)).

Мүмкіндігінше осылармен танысып шығуға тырысыңыз және ішінен өзіңізге оңайын, ыңғайлысын, сіздің зерттеуіңізге сәйкес келетінін таңдап алыңыз.

### 2. Деректеріңізді жазып отырыңыз

Деректеріңізді уақтылы жазып отырмасаңыз, кейін кейбір деректерді естен шығарып алуыңыз мүмкін.

### 3. Деректеріңізді кез келген ықтимал қауіпке тексеріп отырыңыз

Сандар нақтылықты жақсы көреді. Сондықтан бір ғана санды қате енгізу нәтижесінде жұмысыңызда үлкен ауытқулар болуы мүмкін. Сондықтан сондай қателіктер кетіп қалған жоқ па екенін уақтылы тексеріп отырғаныңыз дұрыс.

### 4. Деректерді енгізіп отырыңыз

Қандай да бір дерек жинай алсаңыз, бірден сол таңдаған бағдарламасына енгізіп отырыңыз. Осылайша сіз уақытты үнемдейсіз. Кейін бәрін бір-ақ енгізе саламын деп, уақытты созып алмаңыз. Қолыңыздағы деректерден айырылып қалуыңыз да мүмкін. Бірақ егер деректерді компьютерге енгізген жағдайда да түпнұсқа дерек сізде біраз уақытқа дейін сақталуы керек. Оларды тастап жібермеңіз.

### 5. Деректерді тазартып отырыңыз

Енді статистикалық талдау жүргізуде білуге тиіс терминдерді қарастырып өтейік. Біріншісі – айнымалы шамалар. Бірнеше мәнге ие құрылымдар; айнымалы шамалар «қатаң» (мысалы, гендер, биіктік, табыс) немесе «икемді» (мысалы, өз-өзін бағалау, құн, саяси пікір) болуы мүмкін. Айнымалы шамалар, әдетте, сіздің зерттеу нысаныңызбен бірдей болып жатады. Мысалы, жаңа білім бағдарламасының білім көрсеткішіне қалай әсер еткенін зерттейтін болсаңыз, онда сіздің айнымалы шамаларыңыз білім бағдарламасы және білім көрсеткіші болады. Айнымалыларды өз ішінде екіге бөлеміз: тәуелді айнымалы шамалар және тәуелсіз айнымалы шамалар деп.

Тәуелді айнымалы шамалар – зерттеуші зерттеуге немесе бағалуға тырысып жатқан нәрселер. Мысалы, сіз созылмалы бас ауруға, мықты табыс көзіне немесе орта мектептегі жақсы үлгерімге қандай факторлар әсер ететінін білуге қызығушылық танытуыңыз мүмкін – бас ауруы, табыс көзі және жақсы үлгерім тәуелді айнымалы шамалар.



Кітап: Ғылыми зерттеу жүргізу әдістемесі

Дәріс: Сандық деректерді талдау, 1

Тәуелсіз айнымалы шамалар – сіз түсінуге тырысып жатқан құбылысқа ықпал етуі мүмкін нәрселер. Мысалы, оқу бас ауруына себеп болуы мүмкін; гендер табыс көзін анықтауда рөлге ие болуы мүмкін; ата-ананың ықпалы жетістік деңгейіне ықпал етуі мүмкін. Бұл жердегі тәуелсіз айнымалы шамалар оқу, гендер және ата-ананың ықпалы.

## Айнымалы шамаларды түсіну – өлшеу шкаласы

Айнымалы шамаларды өлшеудің төрт негізгі өлшеу шкаласы бар: номинал сандар, реттік сандар, интервал сандар және қатынас сандар. Ол шкалаларды түсіндірмес бұрын дескриптивті статистикаға тоқала өтейік. Дескриптивті статистиканы тікелей аударғанда сипаттаушы статистика деп аударылады.

## ДЕСКРИПТИВТІ СТАТИСТИКА

Ол жиынтық деректің негізгі қасиеттерін сипаттауға қолданылады. Нақтырақ айтқанда, дескриптивті статистика ортақ процесс, дисперсия және дистрибуция пішінін өлшемдермен қамтамасыз етеді. Осы дескриптивті статистиканы жүргізгенде жоғарыда келтірілген шкалалар қолданылады. Номинал сандар – сандар санаттарды білдіре алатындай етіп кез келген жолмен белгіленетін өлшем шкаласы. Олар кез келген сан болғандықтан және сандық мәні жоқ болғандықтан, оларды математикалық амалдарды орындау мақсатында қолдануға болмайды. Мысалы, гендер мысалында сіз әйел адам үшін бір санды, айталық 1 санын және басқа ер адам үшін, 2 санын қолданар едіңіз. Номиналды деректердің басты қызметі зерттеушілерге жиынтықтық дистрибуциясын түсіну үшін жауаптарды қиюластыруға мүмкіндік беру.

Реттік сандар – санаттарды мағынасы бар болатындай етіп ретке келтіріп қоятын өлшем шкаласы.

Айырмашылық магнитудасы көрсетілмейді. Мысалы, әлеуметтік-экономикалық статус төменгі, орта және жоғарғы тап ретінде топтастырылуы мүмкін. Төменгі тап өзге екі таптан төменірек статусты білдіруі мүмкін, бірақ айырмашылықтың көрсеткіші анықталмаған.

### Орталық тенденцияны бағалау

Өзіңіздің деректеріңізге қатысты сіз сұрауыңыз мүмкін ең басты сұрақтың бірі орталық тенденцияға байланысты. Мысалы: «Тесттегі орташа балл қандай болды?». Статистикада орталық үрдісті бағалаудың үш жолы бар: мән, медиана және мода – және жоғарыдағы мысалға келтіріліген сұрақтар сәйкесінше осы үш өлшеммен байланысты.

### Дисперсияны өлшеу

Ортақ үрдістердің өлшемін деректерді сипаттаудың және жеңілдетудің стандартты және аса пайдалы болғанымен, оларды жауап өзгешелігіне қатысты ақпаратпен толықтыру керек. Сізде 100, 100, 95 және 105 IQ көрсеткішіне ие студенттерден тұратын топ бар делік, және тағы бір топта 60, 140, 65 және 135 IQ көрсеткішіне ие студенттер тобы болсын, онда екі топтың ортақ үрдісі, бұл жағдайда мәні, 100 болар еді. Мәннің айналысындағы дисперсия, алайда, сізден оқу бағдарламасын жасауды және әр топқа бір-бірінен өзгеше түрде сабақ беруді талап етеді. Дисперсияны түсінудің бірнеше жолдары бар, олардың әрқайсысы әртүрлі айнымалы шамалар түріне ыңғайлы. Ортақ үрдіске қатысты статистикалық бағдарламалар автоматты түрде мына сандарды есептеп шығара алады.

Интервал айырмашылықты өлшеу мақсатында деректерді реттейтін және экуидистантты бірліктерді қолданатын өлшеу шкаласы. Бұл шкаланың, бірақ, абсолютті нөлі жоқ. Мысалы, 2016 жыл 1965 жылдан кейін 51 жылдан соң келеді, бірақ уақыт біздің заманымыздың 1 жылы басталмаған. IQ көрсеткіші де интервалды шкала болып есептеледі, бірақ шкаладағы көрсеткіштерді экуидистантты деп қарауға бола ма дегенге қатысты біршама пікірталастар бар.

Қатынас – шкаладағы әр көрсеткіш экуидистантты болатын және абсолютті нөлі бар өлшем шкаласы. Қатынас деректері жас, биіктік және арақашықтық сияқты «шынайы» сандар болғандықтан, мұнда барлық математикалық амалдарды орындауға болады.

Дескриптивті статистика – дистрибуцияның пішіні, ортақ үрдістері және дисперсиясы сияқты жиынтық сипаттамасы.

Ортақ үрдіс – дистрибуцияның орта тұсын немесе орталығын көрсететін өлшем: мән, медиана және мода.

Дисперсия – орталық өлшемнен жеке өлшемдердің қалай таралғаны.

Аралық – бұл дисперсияны есептеудің ең оңай жолы, және ол ең жоғары мәннен ең кіші мәнді азайту. Мысалы, егер сіздің респонденттеріңіздің жас аралығы 8-ден 17-ге дейін өзгешеленетін болса, аралық 9 жыл болар еді. Бұл өлшемді есептеу оңай болғанымен, ол экстремалды мәндерге тәуелді және ортаңғы



мәнді ескермейді.

Квартил – бұл аралығыңызды тең төрт бөлікке немесе «квартилге» бөлуден тұрады және әдетте ординал деректерге арналған дисперсияны өлшеуге немесе ортақ үрдісі медианамен өлшенетін деректерге қолданылады. Квартилдер зерттеушілерге әртүрлі ширектерді салыстыруға немесе дисперсия өлшем ретінде ішкі 50%-ды көрсетеді.

Айырма – бұл өлшем мәннің айналасындағы таралымды есептеу үшін барлық мәндерді қолданады және негізі ол «мәннен квадрат айырмашылықтың ортасы». Айырманы интервал мен қатынас деректерінен есептеу керек және ол дисперсияның жақсы көрсеткішін береді. Ол әлдеқайда жиі кездеседі, бірақ, зерттеушілер арасында айырманың квадрат түбінін қолдану және көрсету стандартты ауытқу ретінде белгілі.

Стандартты ауытқу (с.а.) – бұл айнымалы шаманың квадрат түбірі және ол интервал және қатынас деректерімен байланысты ең жиі қолданылатын көптеген статистикалық деректердің негізі.