



21-дәріс



ҚАЗАҚСТАННЫҢ  
АШЫҚ  
УНИВЕРСИТЕТІ

# ҒЫЛЫМИ ЗЕРТТЕУ ЖҮРГІЗУ ӘДІСТЕМЕСІҚАТЫНАСҚА

## Сандық деректерді талдау, 2





Мән – арифметикалық орта. Мәнді есептеу үшін әр жағдайдың мәндерін қосасыз және оны сол жағдайлардың санына бөлесіз. Мән математикалық есеп болғандықтан, ол интервал мен қатынасқа байланысты деректердегі ортақ үрдісті өлшеу үшін қолданылады, және сандар «кодтар» ретінде қолданылатын номиналды және реттік деректер үшін қолданыла алмайды. Мысалы, христиандарға, буддистерге және мұсылмандарға белгілуге қолданылатын 1-лер, 2-лер және 3-тердің ортақ мәнін табудың еш мағынасы жоқ.

Медиана – аралықтың ортасы. Медиананы табу үшін сіз бар-жоғы мәндерді өсу (немесе кему) реті бойынша қойып шығасыз және ортаңғы мәнді табасыз. Бұл өлшем, әдетте реттік деректерде қолданылады, және экстремалды мәнің ықпалын жоққа шығара алатынымен артықшылыққа ие (мысалы, 3 миллион \$-ға тең экстремалды жалақы табыстың мәнін жоғары етіп көрсетеді – бірақ ол ортаңғы нүктеге әсер етпейді). Сөзсіз, бұл, сондай-ақ, экстремалды деректер зерттеуге маңызды болуына байланысты шектеу де болуы мүмкін.

Мода – айнымалы шама үшін белгіленген ең жиі ұшырасатын мән немесе мәндер. Номиналды сандар санатты болатынына байланысты және математикалық тұрғыда қолдан өзгертілуі мүмкін екеніне байланысты ортақ үрдістің өлшемі ретінде ол модаға жүгінеді.

Инференциялды статистика Жиынтық арасынан алған үлгілерге негізделген жиынтық туралы тұжырым жасауға қолданылатын статистикалық өлшем.

Статистикалық маңыздылық – Әдетте ‘р-мәнін’ білдіреді. Ол сіздің тапқан жаңалығыңыз жай ғана кездейсоқтық емес екенінің ықтималдығын бағалайды.

Осы терминдерді қолдана отырып, біз статистиканы түсіндіреміз. Деректерді графиктерге салған кезде сіздің деректеріңіз қисық сияқты түседі. Қалыпты қисық болып түссе, немесе әдеттегідей болса, онда ол қоңырау сияқты болып түседі. Оны біз қалыпты қисық деп атаймыз. Ал егер ол әдеттегіден тыс жағдайлар болса, немесе қандай да бір ауытқуларды көрсететін болса, онда оның пішіні басқаша, симметриялы емес болып түседі. Симметриялы емес болғанда графикте медиана, мода және мән әртүрлі нүктелерге түседі. Осы арада эксцесс көрсеткішін көруге болады. Эксцесс көрсеткіші дистрибуцияның, яғни қандай да бір таралымның, мысалы, 10-12 жастар аралығындағы балалардың компьютерлік ойын ойнау көрсеткішін зерттейтін болсаңыз, сол олардың күніне не аптасына компьютерлік ойындарды ойнауға жұмсайтын уақыт аралығын білдірер еді. Ал бұл жағдайда әдеттегі қалыпты жағдайда балалар қанша уақыт жұмсайтынын көрсететін көрсеткіш бар және эксцесс көрсеткіші бар, яғни тым көп ойнайтын, экстремалды көп уақыт жұмсайтын уақыт көрсеткіші эксцесс көрсеткіші болады. Бұл ең жоғарғы нүкте болып түседі графикке. Сіздер қазір статистикалық бағдарламамен жұмыс істеп көрмегендіктен, бұл айтылып жатқандар сізге түсініксіз болып көрінуі мүмкін.

Графикке сандық деректеріңіз түскен кезде сіз сандық деректердің дистрибуциясын, яғни таралымын аласыз. Егер сіз дистрибуциямен жұмыс істей алмасаңыз, оны дұрыс талдай алмасаңыз, онда жасаған статистикалық талдауыңыздан еш көмек болмауы мүмкін. Сондықтан дистрибуцияны талдап үйренуге біршама уақыт жұмсауыңызға тура келеді.

## ИНФЕРЕНЦИЯЛДЫ СТАТИСТИКА

Инференциялды статистика әртүрлі айнымалылар арасындағы байланыстар туралы әртүрлі гипотезаларды тексеру үшін немесе, бәлкім одан да маңыздырағы, үлгі деректерден алынған жиынтықтың қасиеттерін есептеуге мүмкіндік беру үшін қолданыла алады. Басқаша айтқанда, инференциялды статистика сізге жалпылауға мүмкіндік береді.

### Инференциялды статистикаға ыңғайлы сұрақтар

#### Менің зерттеуімдегі қатысушылар үлкенірек жиынтықпен қалай салыстырыла алады?

Мысалы, егер сіз мұғалімдердің арасындағы күйзеліс мәселесін зерттейтін болсаңыз, онда бір немесе бірнеше мектепті зерттеп жатырсыз делік. Сіз өзіңізден менің зерттеуімнің нәтижелерін жалпы Қазақстан бойынша мұғалімдерге немесе Алматы қаласы бойынша мұғалімдерге жалпылауға бола ма деп сұрауыңыз қажет.

#### Екі немесе одан да көбірек респонденттер тобы арасында айырмашылықтар бар ма?

Егер жоғарыда келтірілген мұғалімдер арасындағы күйзеліс бойынша түсіндірсек, ауылдық мектептердегі жағдай мен қаладағы мектеп арасында айырмашылықтар бар ма? Немесе Алматы қаласындағы және Нұр-Сұлтан қаласындағы мұғалімдердің арасында айырмашылықтар бар деп сұрақ қоюыңыз қажет.



Кітап: Ғылыми зерттеу жүргізу әдістемесі

Дәріс: Сандық деректерді талдау, 2

## **Екі немесе одан да көп айнымалылар арасында байланыс бар ма?**

Дәл алдыңғы сұрақтағы сияқты, бірақ бұл жерде ұқсастықты қарастырасыз.

### **Статистикалық маңызы**

Инференциалды статистиканың статистика үшін маңызы көп. Біріншіден, ол жалпылауға көмек береді. Екіншіден, сіздің деректеріңіз жай кездейсоқтық емес, сенімді ақпарат екенін дәлелдеуге мүмкіндік береді. Ол үшін  $p$ -құны деген тәсіл қолданылады.  $P$ -құнын есептеген кезде оның көрсеткіші мүмкіндігінше төмен болуы керек, мысалы, 0,001 сияқты. Ондай көрсеткішке ие  $p$ -құны сіздің жұмысыңыздың сенім тудыру қабілеті жоғары екенін көрсетеді, ал оның көрсеткіші жоғары болған сайын оның сенімділігі күдік тудырады. Және оны тіпті қабылдамай да тастауы мүмкін.

### **Дұрыс статистикалық тестті түсіну және таңдау**

Статистикалық бағдарламаларды таңдап, оларды ноутбүгіңізге орнату үшін сіз олармен жұмыс істеп көруіңіз керек. Олардың ішінде ең кең тарағаны SPSS және SAS. Одан бөлек [www.socialresearchmethods.net/kb/index.htm](http://www.socialresearchmethods.net/kb/index.htm) мына сілтеме бойынша өтіп, сол жерден сізге қажет тестілеу бағдарламасын таңдауға көмектесін қызметтік қолдана аласыз. Статистиканы жүргізуде мынандай терминдерге жолығуыңыз мүмкін:

Бірайымалы талдау – бір мезгілде бір айнымалының статистикалық талдануы. Ол ортақ үрдіс, дисперсия және дистрибуция сияқты амалдардан тұрады. Бірайымалы талдау корреляцияға, себеп пен салдарға немесе үлгілеуге қарамайтын болғанымен, талдаудың бұл түрі дескриптивті зерттеуде басты орын алады және статистикалық талдаудың барлық түрлерінде және деңгейлерінде маңызды алғы кезең.

Қос айнымалы талдау – екі айнымалылар арасындағы қатынастарды статистикалық талдау; Қос айнымалы талдау екі айнымалы арасындағы қарым-қатынасты бағалайды; мысалы, білім деңгейі мен теледидар көру әдеті арасындағы байланыс, немесе гендер мен табыс арасындағы байланыс. Ең жиі қолданылатын тесттер кросс-табуляциялар (номиналды айнымалыларға қолданылатын чи квадратты), ANOVA (бір номиналға және бір қатынас айнымалысына қолданылады) және корреляциялар (екі қатынас айнымалы үшін қолданылады).

Көпайымалы талдау – үш немесе одан да көп айнымалылар арасындағы қатынастарды зерттейтін және зерттеушіге себеп пен салдарды іздеуге, үлгілер мен тест теорияларын құруға мүмкіндік береді. Қосайымалы талдаумен зерттеушілер тек тәуелді айнымалы екі немесе одан да көп тәуелсіз айнымалыларға тәуелді ме екенін (яғни, табыс гендерге де, оқудағы жетістікке де тәуелді) зерттеп қанай қоймай, сондай-ақ тәуелді айнымалылар арасындағы байланысты тани алады (мысалы, гендер мен оқудағы жетістік арасындағы байланыс). Көпайымалы талдауда қолданылатын кейбір әдістер факторды талдау, жіті зерттеу, құрылымдық теңдеу үлгісі, MANOVA, көп деңгейлі регрессия, канондық корреляция және соқпақ талдау. Статистика бойынша мамандарды көбіне талдау деңгейінің осы жеріне жеткенде шақыратынын айта кеткен орынды.

Оқулықта келтірілген статистикалық деректерге арналған тарауды қарап, сол жерде статистикалық деректерді іріктеуге көмек болатын кестені бір қарап шықсаңыз болады. Сол жерден қандай сандық деректерге қандай тестілер сәйкес келуі мүмкін екені қарастырылған.

### **САНДЫҚ ДЕРЕКТЕРДІ КӨРСЕТУ**

Сіз қандай бағдарламаны қолданарыңызды анықтап, статистикалық тестті жүргіздіңіз делік. Енді оны көпшіліктің алдында таныстыруыңыз керек. Оны таныстырудың бірнеше тәсілдері бар графиктер арқылы, бағандар арқылы, кестелер арқылы т.с.с. Бірақ шын мәнісінде ол барлық кезде тиімді болмауы мүмкін. Әрине, сіз сонша күшіңізді салып, осындай ауыр еңбек жасаған соң, жасаған нәрсеніздің бәрі туралы айтып бергіңіз келіп тұрады. Бірақ шындығында тыңдап отырғандар үшін оның бәрі қызық болмауы мүмкін. Оларға нақты қойылған мақсаттарға қол жеткізілді ме, қол жеткізген болсаңыз, онда қандай нәтижелерге келдіңіз сол ғана маңызды. Оны қандай жолдармен жасағаныңыз аса маңызды болмауы мүмкін. Сондықтан визуалды көмекші құралдарды таңдағанда сіздің жұмысыңыздағы ең басты нәрсе не еді деген сұрақты жақсылап ойланып, тек ең басты ақпараттарды ғана қалдырыңыз.

Сондай-ақ, тіпті басты ақпараттарды қалдырған кезде мына қағидаларды естен шығармаңыз: сіздің презентацияңыз оны оқитындар үшін түсінуі қиын болмауы керек. Графиктер мен кестелер сіздің



қаншалықты ақылды екеніңізді немесе кестелермен керемет жұмыс істей алатыныңызды көрсету үшін емес, сіздің деректеріңізді мейлінше жеңіл түрде өткізу үшін қолданылғанын естен шығармаңыз. Сондықтан визуалды құралдардағы ақпараттар тым ауыр болып кетпеуіне зер салыңыз. Оқулықтан графиктерді қалай қолдану керектігі жайлы және қандай деректерге қандай графиктер сәйкес келетіні жайлы толығырақ қарап біле аласыз.

## Кестелерді жасау

Кестелердің графиктерден басты айырмашылығы – олар бір қарағанда түсінікті болмауы мүмкін. Бағандардағы көрсеткіштерді бір-бірімен қалай байланысатынын, одан қандай тұжырымдар шығаруға болатынын өзіңіз түсіндіріп, талдап бермесеніз, оны сырттай оқыған адам дұрыс түсіне алмауы мүмкін. Сондықтан кестелерді құрастырған кезде оны таныстырып үйренуге де жаттығыңыз. Егер дұрыс таныстыра алмасаңыз, кестеңізден еш пайда болмауы мүмкін. Сондай-ақ кестелерді жасауға арналған арнайы веб-сайттар бар. Солардың бірі – [www.bagism.com/tablemaker](http://www.bagism.com/tablemaker). Ол сайтта өзіңіздің айнымалыларыңызды не басқа да ақпараттарды енгізу арқылы өзіңізге ыңғайлы кестені жасап шығара аласыз. Бұл сіздің уақытыңызды үнемдеуге көмектеседі.

Кестеге қойылатын талаптар мыналар:

- 1) сіздің оқырмандарыңыздың басын ауыртпауы керек;
- 2) ұзағынан оларды қадалтып отырғызбауы керек;
- 3) оларды одан сайын шатастырмауы керек.

## Тарау қорытындысы

Деректерді түсіндіру – кез келген зерттеуде басты қиындық. Тиімді деректерге басты ойынды көзден таса қылмаудан; деректеріңізбен жұмыс істеуден; талдаудың шынайы үдерісімен айналысудан; және өзіңіздің деректеріңізді тиімді көрсете алудан тұрады.

Деректермен жұмыс істеу сәйкес бағдарламамен таныс болудан; жүйелі түрде логиніңізбен кіріп, деректеріңізді тексеріп отырудан; деректі бағдарламаға жүктеуден; және деректерді «тазартудан» тұрады. Бұл, сондай-ақ, себеп пен салдар (тәуелді немесе тәуелсіз) айнымалыларын және олардың өлшем шкалаларын (номиналды, реттік, интервалды және қатынастық) тани алуды білдіреді.

Дескриптивті статистика ортақ үрдістің (мән, мода және медиана), дисперсияның (аралық, квантильдер, айырма және стандартты ауытқу) және дистрибуцияның (асимметриялық және эксцесс) өлшемдері арқылы деректер жиынтығының басты қасиеттерін тұжырымдау үшін қолданылады.

Инференциалды статистика зерттеушіге дереу деректерден тыс жатқан қорытындыларды шығара алуын бағалауға мүмкіндік береді. Мысалы, үлгі жиынтықты білдіре ме; екі немесе одан да көп топтар арасында айырмашылық бар ма; уақыттың өтуімен өзгерістер болды ма; немесе екі немесе одан да көп топтар арасында байланыстар бар ма.

Сандық деректерді көрсету көбіне графиктер мен кестелерді жасаудан тұрады. Олар 1) маңызды дәлел болатындай сұрыпталып жасалған болуы керек және 2) олар оқырмандарға түсінуге көмек болатындай ақпаратқа қанық әрі қарапайым болуы керек.