



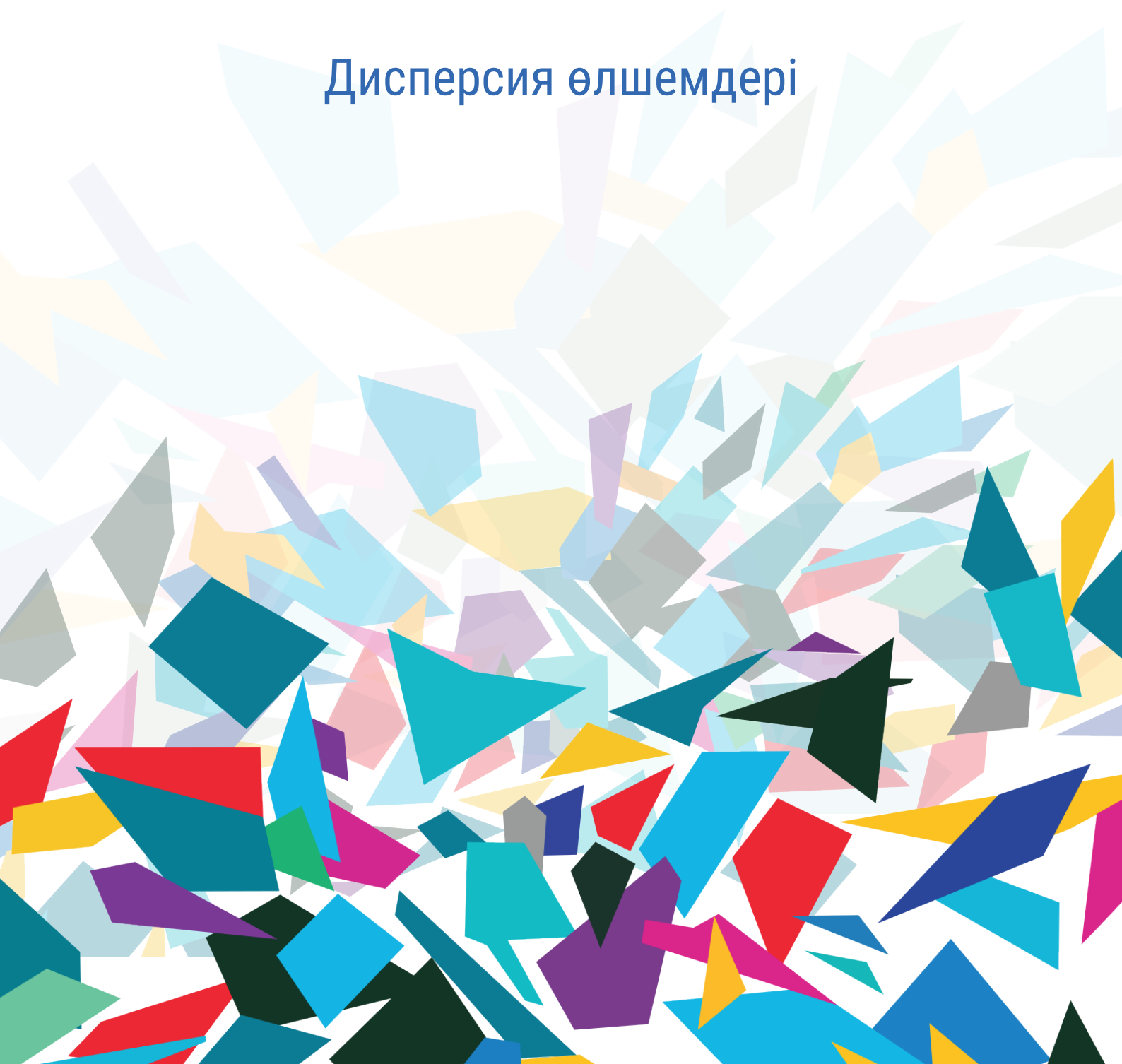
7-дәріс



ҚАЗАҚСТАННЫҢ  
АШЫҚ  
УНИВЕРСИТЕТІ

# СТАТИСТИКА НЕГІЗДЕРІ

## Дисперсия өлшемдері





### Тарауды оқығаннан кейін:

1. Дисперсия өлшемдерінің мақсаты мен олардан алынатын ақпаратты түсіндіре аласыз.
2. диапазон (R), кватиль аралық диапазон (Q), стандартты девиация (s) және вариансты ( $s^2$ ) есептеп түсіндіре аласыз.
3. Дисперсияның қолайлы өлшемін таңдай аласыз, статистиканы дұрыс есептеп, интерпретациялай аласыз.
4. Стандартты девиацияның математикалық ерекшеліктерін сипаттайсыз және түсіндіре аласыз.
5. Боксплотты талдай аласыз.
6. Стандартты девиация мен диапазонды шығару үшін SPSS бағдарламасын қолдана аласыз.

Дәрісте қолданылған статистикалық техника баллдар жиынтығындағы өзгергіштікті немесе әртүрлілікті сипаттау үшін пайдаланылады. Оны мына пункттер бойынша қолдануға болады:

- Баға вариациясы. Кейбір студенттердің бағасы әрдайым бірдей деңгейде болады, ал басқаларының бағасы жоғары деңгейден төменге өзгереді.
- Түрлі әлеуметтік орталарда өмір салты әр алуан. Шағын қалаларға қарағанда ірі қалаларда өмір сүру салтының әралуандығы кең таралған.
- Орналасқан жеріне қарай нәсілдік және этностық әралуандық. Кейбір штаттар (мысалы, Калифорния және Нью-Йорк) әртүрлі нәсілдік, мәдени және тілдік топтардың отаны болса, ал басқаларында (Айова және Мэн) өзгешелік көп емес.
- Әр елге байланысты, уақытқа байланысты табыс теңсіздігінің вариациясы. Кейбір елдерде ең бай адамдар мен кедейлер арасында байлық пен табыс бойынша орасан зор айырмашылық бар, ал басқа елдерде мұндай айырмашылық мардымсыз болады.

Дисперсия концепциясының маңызды екенін мысал арқылы түсіндіру жеңілрек болады. Айталық, Қоғамдық қауіпсіздік жөніндегі директор төтенше жағдайда медициналық көмек көрсету бойынша қала әкімдігімен келісімшарт жасасқан екі жедел жәрдем қызметін салыстырып (бғалап) көрмек. Зерттеу барысында ол көмекке келудің орташа уақыты А қызметінде 7,4 минут, ал В қызметінде 7,6 минутты көрсететіні туралы деректерді жинады.

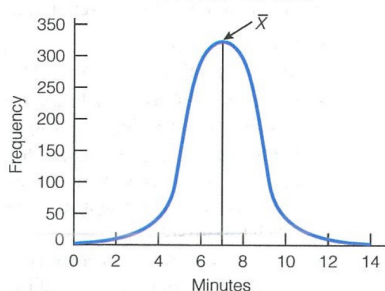
Іс жүзінде бұл орташа мәндер бірдей, олар бір қызметтің екіншісінен тиімді екенін анықтай алмайды. Алайда, дисперсия өлшемдері (орталық тенденция өлшемдері бірдей болғанда да) үлестірімдер арасындағы маңызды айырмашылықтарды анықтай алады. Мысалы, 4.1-сызбадағы екі жедел жәрдем қызметінің жауап уақыттарының үлестірімін көрсететін графиктерді қарастырайық.

А қызметіне қарағанда В Қызметінің сызықтық диаграммасы әлдеқайда жалпақ екеніне назар аударыңыз. Себебі, В қызметінің баллдары А қызметінің баллдарына қарағанда шашыраңқы орналасқан, яғни әртүрлі. Басқаша айтқанда, В қызметінің жауап беру уақыты әлдеқайда құбылмалы болды, жоғары және төмен диапазонтерде көбірек баллдарға, ал ортасында азырақ мөлшердегі баллдарға ие. А қызметі өзінің жауап беру уақытында неғұрлым тұрақты және оның балдары арифметикалық ортаның төңірегінде шоғырланған немесе топтастырылған. Екі үлестірімде де негізінен орташа жауап уақыты бірдей. Дегенмен В қызметінің жауап беру уақыттарында айтарлықтай көп вариация немесе дисперсия бар.

Егер сіз қоғамдық қауіпсіздік қызметінің директоры болсаңыз, төтенше жағдай кезінде тұрақты жедел жәрдем қызметін (А Қызметі) таңдайсыз ба? Әлде бірде баяу, бірде жылдам жауап беретін қызметті (В Қызметі) таңдайсыз ба? Осы орайда дисперсияны қарастырмасаңыз, екі жедел жәрдем қызметіндегі маңызды айырмашылықтарды байқамайтын едіңіз.

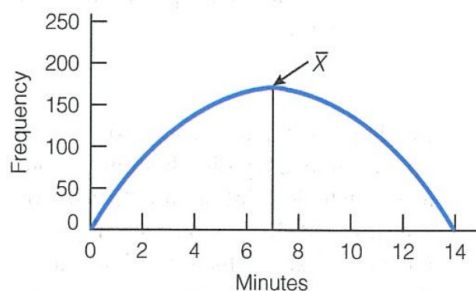
4.1-сызба. Екі жедел жәрдем қызметінің жауап беру уақыты

А қызметі





В қызметі



4.1-сызбадағы екі пішінді дисперсия концепциясының көрнекілігі ретінде есте сақтап алыңыз. А қызметінде арифметикалық орта төңірегінде баллдар көптеп шоғырлануы дисперсияның төмен екенін көрсетеді, ал В қызметіндегі үлестірімнің жалпақ қисығы әралуандықты, яғни дисперсияны көрсетеді. Баллдардың құбылмалылығы төмен болған сайын дисперсияның кез келген өлшемі төмендейді, ал үлестірім үшкірлене түседі (үлестірім А қызметіне ұқсай бастайды), баллдар әртүрлі болған сайын дисперсия ұлғаяды, ал үлестірім жалпақтана түседі (үлестірім В қызметіне ұқсай бастайды).

Бұл тұжырымдар мен 4.1-сызба дисперсия туралы жалпы түсінік береді, бірақ бұл тұжырымды сипаттап беру оңай емес: баллдар жиынтығындағы вариацияны сандық есептеуге арналған статистикалық мәліметтерді қарастырайық. Осы дәрісте дисперсияның ортақ өлшемдерімен танысамыз. Диапазон және кватиль аралық диапазоннан бастайық, ал стандартты девиацияға жеке тоқталамыз. Сонымен қатар боксплот деп аталатын дисперсияның визуалды көрінісін қарастырамыз.

### Диапазон (R) және кватиль аралық диапазон (Q)

Диапазон (R) үлестірімнің жоғары және төмен баллдары арасындағы қашықтық:

4.1-формула

$$R = \text{Жоғары балл} - \text{Төмен балл}$$

Диапазонды есептеу оңай және тез, сондай-ақ ол құбылмалылықтың жалпы индикаторы. Осы орайда статистиканы интерпретациялау да қиын емес: диапазонның мәні үлкен болған сайын жоғары балл мен төмен балл арасындағы қашықтық та үлкен болады, үлестірімдегі дисперсия өлшемі де үлкейеді.

Өкінішке қарай, диапазон тек жоғары және төмен баллдарға ғана негізделгендіктен, оның кейбір маңызды шектеулері бар. Біріншіден, көлемді үлестірімде баллдардың ішінде экстремалды жоғары не төмен баллдар (сырттағылар) болады (мысалы, 3.3-кестені қараңыз).

Осылайша, R үлестірімдегі баллдардың басым бөлігі үшін дисперсия мөлшерін асырып жіберуі мүмкін. Оған қоса, R ең жоғары және ең төмен баллдар арасындағы баллдар вариациясы туралы ешқандай ақпарат бермейді.

Кватиль аралық диапазон (Q) – диапазонның бір түрі. Ол үлестірімдегі кейстердің тек ортаңғы 50%-ын қарастыру арқылы R-ға қатысты кейбір проблемалардың алдын алады. Q-ді табу үшін, алдымен баллдарды жоғарыдан төмен орналастырыңыз, содан соң үлестірімдерді ширектерге бөліңіз (медиананы анықтауда жартыға бөлген секілді). Бірінші кватиль ( $Q_1$ ) кейстердің 25%-ы төменде қалатын, ал қалған 75%-ы жоғарыда орналасатын нүкте. Екінші кватиль ( $Q_2$ ) үлестірімді теңдей екіге бөледі (яғни  $Q_2$  медианамен тепе-тең). Үшінші кватиль ( $Q_3$ ) кейстегі үлестірімнің 75%-ы төменде орналасатын, ал қалған кейстердің 25%-ы жоғарыда орналасатын нүкте.

Кватиль аралық диапазон – үшінші мен бірінші кватиль аралығындағы қашықтық:

4.2-формула

$$Q = Q_3 - Q_1$$

Шын мәнінде, кватиль аралық диапазон үлестірімнің ортасындағы 50%-ды ғана шығарады және R секілді тек екі баллға негізделеді. Q-дің мәні R сияқты интерпретацияланады: оның мәні неғұрлым жоғары болса, дисперсия да жоғары болады. Диапазонпен салыстырғанда, Q экстремалды баллдарға негізделетін қиындықтарға әкелмейді, өзі негізделген екі баллдан басқа баллдар вариациясы туралы ешқандай ақпарат



бермейді.

### Диапазон пен квартиль аралық диапазонды есептеу

4.1-кестеде 40 елдегі туу көрсеткіші берілген (туу саны әрбір 1 000 адамға шаққанда). Осы деректердің диапазоны мен квартиль аралық диапазоны қандай? Баллдар жоғарыдан төменге қарай орналасқанын ескеру қажет. Бұл диапазонды есептеуді жеңілдетеді және квартиль аралық диапазонды табу үшін қажет. 40 елдің ішінде Нигерияның туу көрсеткіші – 50, ол жоғары қатарда, ал Германия мен Жапонияның туу көрсеткіштері – 8, олар ең төмендегілердің қатарында. диапазон 50–8, яғни 42 ( $R=42$ ).

4.1-кесте. 40 елдегі туу көрсеткіші (туу саны әрбір 1 000 адамға шаққанда), 2013 жыл

Реті	Елдер	Туу көрсеткіші	Реті	Елдер	Туу көрсеткіші
40 (Жоғары)	Нигер	50	20	Ливия	22
39	Ангола	47	19	Үндістан	22
38	Уганда	45	18	Венесуэла	21
37	Мозамбик	44	17	Мексика	19
36	Нигерия	42	16	Колумбия	19
35	Малавия	40	15	Қувейт	19
34	Танзания	40	14	Австралия	18
33	Гвинея	38	13	Вьетнам	17
32	Сенегал	38	12	Ирландия	16
31	Того	37	11	Чили	15
30	Кения	36	10	Исландия	14
29	Мавритания	35	9	АҚШ	13
28	Эфиопия	34	8	Ресей	13
27	Гана	33	7	Франция	13
26	Гватемала	32	6	Қытай	12
25	Пәкістан	30	5	Канада	11
24	Гаити	26	4	Испания	10
23	Камбоджа	25	3	Италия	9
22	Мысыр	25	2	Жапония	8
21	Сирия	25	1 Төмен	Германия	8

$Q$ -ді табу үшін бірінші және үшінші квартильдерді ( $Q_1$  және  $Q_3$ ) анықтау қажет. Медиананы тапқан секілді бұл нүктелерді де біз нақты кейстермен байланысқан баллдар арқылы анықтаймыз  $N$ -ді  $0,25$ -ке көбейту арқылы  $Q_1$  табылады.  $(40) * (0,25) = 10$  болғандықтан,  $Q_1$  ең төменнен жоғары санағанда 10-шы кейске қатысты балл. Оныншы кейс – Исландия, оның ұпайы – 14, демек  $Q_1 = 14$ .

Үшінші квартильде ( $Q_3$ ) жатқан кейс  $N$ -ді  $0,75$ -ке көбейту арқылы алынған,  $(40) * (0,75) = 30$ -кейс. Ал ең төменгі баллдан жоғары санасақ, 30-шы кейс – Кения, оның балы 36 ( $Q_3 = 36$ ). Ендеше,

$$Q = Q_3 - Q_1$$

$$Q = 36 - 14$$

$$Q = 22$$

### Стандартты девиация және вариантс

$Q$ -ға да,  $R$ -ға да тән кемшілік – екеуі де екі баллны ғана қолданады, барлық қолжетімді ақпаратты қолданбайды. Сонымен қатар статистикалық мәліметтер баллдардың бір-бірінен қаншалықты қашықта, арифметикалық орта сияқты ортаңғы нүктеден қаншалықты алыста екені туралы ақпарат бермейді. Осы



шектеулерді болдырмайтын дисперсия өлшемін қалай жасауға болады?

Біз қандай да бір ерекшеліктермен бастай аламыз. Дисперсияның жақсы жасалған өлшемі былай болуы тиіс:

1. Үлестірімнің барлық баллдарын қолданады. Статистика қолда бар барлық ақпаратты қолдануға тиіс.
2. Баллдардың орташа баллсын және типтік девиациясын сипаттайды. Статистика бізге баллдар бір-біріне немесе үлестірімнің ортасынан қаншалықты алыс орналасқаны туралы ақпарат беруі тиіс.
3. Баллдар әр алуан болған сайын мән де ұлғаяды. Бұл үлестірімді салыстыру кезінде өте пайдалы сипаттама болар еді, өйткені ол бізге қайсысы айнымалы болатынын бір қарағаннан айтуға мүмкіндік береді: статистиканың сандық мәні жоғары болған сайын, дисперсия да ауқымды болады.

Осы критерийлерге сай келетін статистиканы жасаудың бір жолы – әр балл мен арифметикалық орта арасындағы қашықтықтан немесе девиациядан ( $X_i - \bar{X}$ ) бастау. Девиацияның мәні балл мен арифметикалық ортаның айырмашылығы артқан сайын арта түседі. Егер баллдар неғұрлым арифметикалық ортаның төңірегінде кластерленген болса (4.1-сызбадағы А Қызметі), девиациялар соғұрлым аз болады. Егер баллдар неғұрлым шашыраңқы немесе әртүрлі болса (4.1-сызбадағы В Қызметі) соғұрлым девиация да үлкен болады. Тиімді статистика жасау үшін девиацияны қалай қолдануға болады?

Орындалатын әрекеттердің бір түрі – статистиканың негізі ретіндегі девиациялардың қосындысын қолдану  $\Sigma(X_i - \bar{X})$ , бірақ девиациялардың қосындысы әрдайым 0 болады.

Дегенмен девиациялардың қосындысы дисперсия өлшемінің қисынды негізі және статистиктер оң девиациялар әрдайым теріс девиацияларға тең болатын жолды қалыптастырды. Егер әрбір девиацияны квадраттайтын болсақ, онда барлық айнымалылар оң болып шығады, өйткені теріс сандарды өз-өзіне көбейткенде ол оң санға айналады. Мысалы,  $(-20) \times (-20) = +400$ . 4.2-кестедегі баллдар үшін квадратталған девиациялардың қосындысы былай болар еді:  $(400+100+0+100+400)$ , яғни 1 000 (4.3-кестені қараңыз). Демек, девиация квадраттарының қосындысына негізделген статистика дисперсияның жақсы бір өлшемін табуда ұтымды тәсіл бола алады.

4.2-кесте. Арифметикалық орта айналасындағы шамалар  
Девиациясының қосындысы әрдайым тұтастай нөл болатынын көрсету

Шамалар ( $X_i$ )	Девиациялар ( $X_i - \bar{X}$ )
10	$(10 - 30) = -20$
20	$(20 - 30) = -10$
30	$(30 - 30) = 0$
40	$(40 - 30) = 10$
50	$(50 - 30) = 20$
$\Sigma(X_i) = 150$	$\Sigma(X_i - \bar{X}) = 0$
$\bar{X} = 150/5 = 30$	

4.3-кесте. Стандартты девиацияны есептеу

Шамалар ( $X_i$ )	Девиация ( $X_i - \bar{X}$ )	Девиация квадраты ( $(X_i - \bar{X})^2$ )
10	$(10 - 30) = -20$	$(-20)^2 = 400$
20	$(20 - 30) = -10$	$(-10)^2 = 100$
30	$(30 - 30) = 0$	$(0)^2 = 0$
40	$(40 - 30) = 10$	$(10)^2 = 100$
50	$(50 - 30) = 20$	$(20)^2 = 400$
$\Sigma(X_i) = 150$	$\Sigma(X_i - \bar{X}) = 0$	$\Sigma(X_i - \bar{X})^2 = 1 000$

Дисперсия өлшемін құрастыруды аяқтамас бұрын, тағы бір мәселені қарастырып алайық. Квадратталған девиациялардың қосындысы жиынтықкөлемімен бірге артады. Баллдардың саны артқан сайын, өлшемнің мәні де үлкенірек болады. Бұл әртүрлі көлемдегі іріктемелерге негізделген үлестірім құбылмалылығын салыстыруда қиындық тудыруы мүмкін. Бұл мәселені шешу үшін девиациялар квадраттарының қосындысын N-ге (жиынтықкөлемі) бөлу керек. Осылайша, әртүрлі көлемдегі іріктемелерді стандарттаймыз.

Бұл процедуралардың негізінде вариант деп аталатын статистика анықталады, ол  $s^2$  деп таңбалады.



Варианс кейбір құрылымдарды өлшеуді жобалаудың орталық концепциясы болғанмен, ең алдымен инференциалды статистикада қолданылады. Үлестірімдегі дисперсияны сипаттау мақсатында, әдетте осыған жақын статистика – стандартты девиацияны ( $s$  деп таңбаланады) қолданады. Осы тарауда бұл ұғымға тереңірек тоқталамыз.

Варианс ( $s^2$ ) пен стандартты девиацияның ( $s$ ) формулалары:

$$4.3\text{-ФОРМУЛА} \quad s^2 = \frac{\sum (X_i - \bar{X})^2}{N}$$

$$4.4\text{-ФОРМУЛА} \quad s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}}$$

Шындығында, 4.3 және 4.4-формулар популяцияның варианты мен стандартты девиациясы. Аздаған өзгеше формулар, яғни  $N$ -нің орнына  $N-1$  қолдану бүкіл популяцияға қарағанда кездейсоқ іріктемемен жұмыс істегенде қолайлырақ. Бұл өте маңызды пункт, өйткені пайдаланатын көптеген электронды калькуляторлар мен статистикалық бағдарламалар пакеті (SPSS-ға кіреді)  $N-1$ -ді жиі қолданады, сондықтан шыққан нәтиже 4.3 және 4.4-кестеде берілген формулар бойынша есептелген нәтижеден сәл өзгеше болуы мүмкін.

4.4-кесте. Екі кампус студенттері үшін стандартты девиацияны есептеу

Жергілікті кампус		
Шамалар ( $X_i$ )	Девиациялар ( $X_i - \bar{X}$ )	Квадратталған девиациялар ( $(X_i - \bar{X})^2$ )
18	$(18 - 19) = -1$	$(-1)^2 = 1$
19	$(19 - 19) = 0$	$(0)^2 = 0$
20	$(20 - 19) = 1$	$(1)^2 = 1$
18	$(18 - 19) = -1$	$(-1)^2 = 1$
20	$(20 - 19) = 1$	$(1)^2 = 1$
$\Sigma(X_i) = 95$	$\Sigma(X_i - \bar{X}) = 0$	$\Sigma(X_i - \bar{X})^2 = 4$
$\bar{X} = \Sigma X_i / N = 95 / 5 = 19$		

Қала кампусы		
Шамалар ( $X_i$ )	Девиациялар ( $X_i - \bar{X}$ )	Квадратталған девиациялар ( $(X_i - \bar{X})^2$ )
20	$(20 - 23) = -3$	$(-3)^2 = 9$
22	$(22 - 23) = -1$	$(-1)^2 = 1$
18	$(18 - 23) = -5$	$(-5)^2 = 25$
25	$(25 - 23) = 2$	$(2)^2 = 4$
30	$(30 - 23) = 7$	$(7)^2 = 49$
$\Sigma(X_i) = 115$	$\Sigma(X_i - \bar{X}) = 0$	$\Sigma(X_i - \bar{X})^2 = 88$
$\bar{X} = \Sigma X_i / N = 95 / 5 = 23$		

Жиынтық мөлшері/көлемі артқан сайын айырмашылық кеміп отырады, бірақ тараудағы мысалдар мен жаттығуларда жиынтықаз берілген, сондықтан  $N$  мен  $N-1$  қолданудың арасындағы айырмашылық мұндай жағдайларда елеулі болуы мүмкін. Кейбір калькуляторлар стандартты девиацияны есептеуде  $N-1$  немесе  $N$ -нің біреуін таңдауға мүмкіндік береді. Егер сізде сондай калькулятор болса,  $N$ -ді таңдау арқылы жауапты осылай шығаруға болады.

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}}$$

$$s = \sqrt{\frac{1000}{5}}$$

$$s = \sqrt{200}$$

$$s = 14,14$$



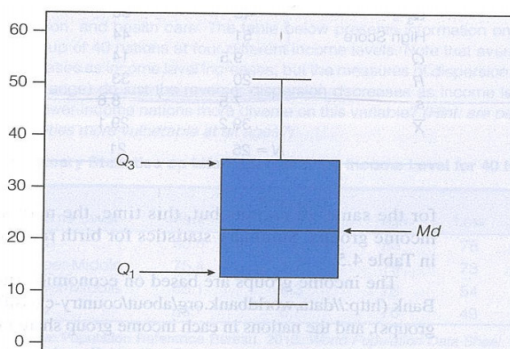


## Дисперсияны визуалдау: Боксплоттар

Боксплот немесе уискерс, бокс деп аталатын диаграмма дисперсияны визуалды көрсету мен талдаудың оңтайлы тәсілі, сондай-ақ көбейіп келе жатқан кейбір статистикалық құралдарды пайдалануға мүмкіндік береді. Боксплот орталық тенденция өлшемі мен құбылмалылықты бейнелеу үшін медиананы (алдыңғы лекцияны қара), диапазонды (R), квартиль аралық диапазонды(Q) пайдаланады. Сонымен қатар үлестірімге кіруі мүмкін кез-келген сырттағы немесе шектен тыс баллдарды да көрсетеді.

4.2-сызбаны мысалға алып, боксплотты зерттеуге кірісейік. Графикте даму және әл-ауқат деңгейі әртүрлі 99 елдегі туу көрсеткіші (әрбір 1 000 адамға шаққанда) берілген. Боксплот бір бокстан және Т-пішінді сызықтардан тұрады, олардың біреуі бокстың үстінгі жағында, екіншісі астыңғы жағында орналасқан. Бокс жоғарғы жағында 3-квартиль (Q<sub>3</sub>) мен төменгі жағында 1-квартильге (Q<sub>1</sub>) созылған. Осылайша, бокстың биіктігі-дің мәнін немесе квартиль аралық диапазонды білдіреді. Бокс арқылы өтетін горизонталь сызық медианаға созылған (Md).

4.2-сызба. 99 елдің туу көрсеткіші туралы боксплот, 2013 жыл



Әрбір 1 000 адамға шаққандағы туу саны

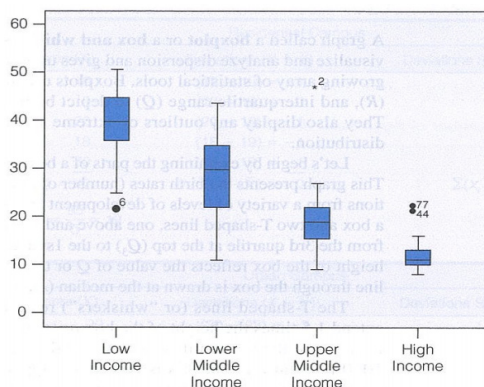
Т-пішінді сызықтар (немесе “уискерс”) баллдардың диапазонтерін білдіреді. Олар бокстың биіктігін 1,5 есе немесе қайсыбір боксқа жақын орналасқан жоғары және төмен көрсеткіштерге дейін ұзартады. Бұл жағдайда, 1 000 адамға шаққанда ең жоғарғы туу көрсеткіші – 51, бұл сызбаның жоғары жағында орналасқан уискерс. Төмен жақта орналасқан уискерс туу көрсеткіші – 8-ді көрсеткен тұста, бұл осы елдердің ішіндегі ең төменгісі.

Боксплотта “сырттағы” баллдар “уискерстен” тыс орналасқан кез-келген балл болады немесе бокстың кез-келген бағытындағы 1,5 есе биіктіктен тыс жатқан балл. Бокстың биіктігінен үш есе артық кез-келген бағыттан тыс жатқан баллдар “шектен тыс сырттағылар” деп аталады. 4.2-сызбада көрсетілген елдер ішінде шектен тыс баллдар жоқ, бірақ кейбір мысалдарды қысқаша қарастырып шығуымыз қажет.

Әртүрлі жағдайларда немесе уақыттар шеңберінде айнымалы үлестірімін салыстыру қажет болғанда боксплоттар ерекше пайдалы болады. 4.3-сурет тура сол 99 елдің туу көрсеткішін сипаттайды, бірақ енді, осы елдер табысы бойынша топтамаларға жіктелген. Табыс деңгейі бойынша туу көрсеткішін қорытындылайтын статистикалар 4.5-кестеде келтірілген.

4.3-сызба. 99 елдің табыс деңгейі бойынша туу көрсеткіші туралы боксплот, 2013 жыл

Әрбір 1 000 адамға шаққандағы туу саны





4.5-кесте. 99 елдің табыс деңгейі бойынша туу көрсеткішін қорытындылайтын статистикалық мәліметтер, 2013 жыл

Статистикалар	Төменгі табыс	Орташадан төмен табыс	Орташадан жоғары табыс	Жоғары табыс
Төмен шама	31	11	9	8
Q <sub>1</sub>	35,5	22	15,3	10
Медиана	40	30	19	11
Q <sub>3</sub>	45	36	22	13
Жоғары шама	51	44	47	22
Q	9,5	14	6,7	3
R	20	33	38	14
s	7,5	8,6	7,4	3,4
$\bar{X}$	39,2	29,1	19,7	11,8
	N =25	21	24	29

Табысы бойынша жіктелген топтамалар Дүниежүзілік Банк жасаған экономикалық критерийлерге негізделген (<http://data.worldbank.org/about/country-classifications/country-and-lending-groups>) және әр топтамадағы табыс деңгейі бойынша жіктелген елдердің ортақ әлеуметтік сипаттары бар:

Төмен табысты елдер аграрлық экономикаға жатады, бұл қоғамдар үлкен отбасыны алға тартады. Бұл категорияға Африка елдері, Гаити және кейбір Азия елдері кіреді.

Табысы орташадан төмен елдер жалпы индустриаланудың бастапқы сатысындағы елдер, оларға Мысыр, Боливия және Индия жатады.

Табысы орташадан жоғары елдердің экономикасы жетілген, бұларға Аргентина, Түркия және Қытай жатады.

Табысы жоғары елдердің барлығы дерлік озық индустриалды және урбандалған, өмір сүру деңгейі өте жоғары стандартты болады. Мұндай экономикаларда бала өсіру әлдеқайда қымбатырақ, сондықтан отбасы мүшелерінің саны да аз, соған қарай туу деңгейі де төмен. Бұл категорияға кіретін елдер: Америка Құрама Штаттары, Канада, көптеген Батыс Еуропа елдері, Жапония, Австралия және Жаңа Зеландия.

Орталық тенденция өлшемдерін бір қарап шыққанда байқайтынымыз, табыс деңгейі өскен сайын медианалық туу көрсеткіші (бокста горизонталды сызықпен белгіленген) күрт қысқарады, ал арифметикалық орта өседі (4.5-кестені қараңыз). Табысы төмен елдерде туу деңгейінің орташа көрсеткіші әлдеқайда жоғары, табысы жоғары елдерде туу көрсеткіші төмен. Ал табысы орташадан төмен және орташадан жоғары елдерде туу көрсеткіші орталық арада (intermediate).

Дисперсияға қайтып оралсақ, бірнеше үлгіні байқаймыз. Табысы орташадан жоғары елдерде диапазон өте үлкен (R = 38), сондықтан бұндағы статистика бойынша топтар әртүрлі. Бірақ бұл категориядағы елдер үшін диапазон “сырттағы” баллдарды (№2 ел – Ангола) қамтығаны үшін өсіріліп көрсетеілген. Анголаның баллы “шектен тыс” сыртта (бокстың биіктігінен 3 есе артық баллға ие) орналасқан, сондықтан ол сызбада \* белгісімен көрсетілген. Есіңізде болса, диапазон тек қана жоғары және төмен баллдарға негізделеді, сондықтан бұл кейсте дисперсияның өлшемін дұрыс көрсете алмайды.

Екіншіден, табысы орташадан төмен елдер өте жоғары кватиль аралық диапазонке (Q) және стандартты девиацияға (s) ие, сондықтан осы статистикаларға сай өте үлкен дисперсияға сәйкес келеді. табысы орташадан жоғары елдерде “сырттағы” баллдарды қамтығаны үшін өсіріліп көрсетеілгендіктен, мен бұл жағдайда дисперсияның дұрыс өлшемі. Табысы орташадан төмен елдер тууы көрсеткіші бойынша көбірек әртүрлі деп қорытындылауға болады.

Үшіншіден, табысы жоғары елдерде елеусіз ғана өзгеріс болады. Бұл қысқа боксплоттан, өте төмен кватиль аралық диапазоннан (3), диапазоннан (14) және стандартты девиациясынан (3,4) көрінеді. Бұл кейстер екі сырттағы баллдарға қарамастан (№77 Сауд Аравиясы және №44 Израиль) бір-бірімен тым ұқсас болып келеді. Неліктен бұл екі ел басқа жоғары табысты елдерге қарағанда туу көрсеткіші бойынша ерекше деп ойлайсыз?

Қорытындылай келе, боксплоттар дисперсия мен орталық тенденция өлшемдері туралы өте пайдалы визуалды ақпарат береді. Осылар сияқты боксплоттар да сан түрлі жағдайлардағы айнымалыларды салыстыруда пайдасы аса зор.





## Стандартты девиацияны интерпретациялау

Бұл жерде стандартты девиацияның мағынасы толықтай анық болмауы да мүмкін. “Стандартты девиацияны” есептеу барысында қиындықтар кездессе не істеу керек?” деген сауал туындауы мүмкін. Дисперсияны бұлай өлшеудің мағынасы үш жолмен түсіндіріледі. Біріншісі және ең маңыздысы, бұл қалыпты қисық, бұл интерпретацияны келесі тарауға қалтырамыз.

Екіншісі, стандартты девиация – үлестірімде өзгеріс көп болған сайын мәндер өсіп отыратын дисперсия индексі. Басқаша айтқанда, стандартты девиация көп түрлі мәндері бар үлестірім үшін жоғарырақ болады, ал азырақ таралған мәндер үшін төменірек болады. Стандартты девиацияның төмен мәні 0 болуы да мүмкін және бұл дисперсиясы жоқ таралымдарда пайда болады (яғни, іріктемедегі әрбір кейс тура сондай баллмен сәйкес келсе). Сондықтан, 0 стандартты девиация үшін ең төменгі болжамды мән.

Стандартты девиацияның мағынасын толық түсінудің үшінші жолы – бір таралымды екінші таралыммен салыстыру. Біз бұны 4.1-сызбада екі жедел жәрдем қызметін көрсетуде, 4.4-кестеде жергілікті және қалалық кампустарды салыстыруда қолданғанбыз. Сіз бұны бір топты екінші бір топпен (әйелдер мен ерлерді, ақ нәсілділер мен қараларды т.с.с.) немесе екі бірдей анымалыны екі түрлі уақытта салыстыру үшін қолдана аласыз. Мысалы, 4.6-кестедегі қорытынды статистикада көрсетілгендей, белгілі бір кампустағы студенттердің жастары біраз уақыт шеңберінде өзгеретінін байқадық делік.

Анығында, орта есеппен студенттердің жасы өскен, сондықтан стандартты девиация бойынша жас мөлшерінің әртүрлілігі де ұлғайған. 1990 жылғы төмен стандартты девиация жас мөлшерінің осы жылы арифметикалық орта төңірегінде көбірек шоғырланғанын білдіреді (4.1-сызбадағы А қызметі үшін үлестірімді есіңізге түсіріңіз). Керісінше, 2010 жылы үлкен стандартты девиацияда үлестірім жалпақтау әрі шашыраңқы болады, 4.1-сызбадағы В қызметінің үлестіріміне ұқсас болады. Басқа сөзбен айтқанда, 2010 жылмен салыстырғанда 1990 жылы студенттер бір-бірімен жас мөлшері жағынан ұқсас болған, яғни жас мөлшерінің диапазоны кең ауқымды емес. Стандартты девиация баллдар үлестірімін салыстырғанда аса пайдалы.

## ТҮЙІН

1. Дисперсия өлшемдері шамалар үлестіріміндегі гетерогендік немесе әралуандық туралы ақпаратты түйіндейді. Орталық тенденция өлшемдерімен қосқанда бұл статистикалар аздаған сандармен аса үлкен ақпарат көлемін жеткізеді. Орталық тенденция өлшемдері үлестірімнің ортаңғы нүктелерінде болса, дисперсия өлшемдері үлестірімдегі әрталуандық көлемін анықтайды.

2. Құлаш (R) – үлестірімдегі ең жоғары шамадан төменгі шамаға дейінгі қашықтық. Квартиль аралық (Q) – құлаш үшінші және бірінші квартиль аралығындағы қашықтық (шамалардың 50 пайыздық ортасының құлашы). Бұл екі құлаш реттік немесе интервалды-пропорционалды деңгейде өлшенетін айнымалылардың әрқайсысы үшін бірдей пайдаланылады.

3. Стандартты девиация (s) – дисперсияның ең маңызды өлшемі, өйткені көптеген озық статистикалық қолданбаларда ол негізгі рөл атқарады. Стандартты девиацияның мәні – минимум 0 (үлестірімде вариацияның жоқ екенін білдіреді) және үлестірімдегі өзгешеліктер артқан сайын оның да көлемі арта түседі. Оны интервалды-пропорционалды деңгейде өлшенетін айнымалылар үшін пайдалану ұтымдырақ, бірақ реттік деңгейдегі айнымалылар үшін де жиі есептеледі.

4. Вариа́нс ( $s^2$ ) инференциалды статистикада және кейбір ассоциация өлшемдерін құрастыруда пайдаланылады.

5. Дисперсияны визуалды көрсету және талдау үшін боксплоттар қолданылады.