

## Бірнеше тестілеу

Жоғарыда айтқанымыздай, статистикада «Деректерді ұзақ қинасаң, ерте ме, кеш пе мойындайды» деген сөз бар. Бұл дегеніміз, егер сіз деректерді өте көп санды әртүрлі көзқарастардан қарасаңыз және тым көп сұрақтар қойсаңыз, сіз статистикалық маңызды әсер табасыз.

### *Негізгі терминдер*

*1-ші қателік (1-ші қате) эффект статистикалық маңызды деген қате тұжырым.*

*Жалған ашу коэффициенті (false discovery rate) бірнеше тестілеу нәтижесінде 1-ші қатені жасау үлесі.*

*P-мәндерін түзету (p-мәндерін түзету) бір деректер бойынша бірнеше тестілеуді орындау кезінде мәнді нақтылау.*

*Қайта орнату (overfitting) шуға сәйкес келеді.*

Мысалы, егер 20 болжамды айнымалылар және бір алынған  $p$  - белдік болса және олардың барлығы кездейсоқ түрде жасалса, онда  $\alpha$ -деңгейінде 0,05-ке тең 20 маңыздылық тексерулерінің сериясын орындау арқылы кем дегенде бір болжаушының (жалған түрде) статистикалық маңызды болып шығуының жеткілікті жақсы мүмкіндігі бар. Бұрын талқыланғандай, бұл жағдай 1 - ші қателік деп аталады. Бұл ықтималдылықты алдымен барлық айнымалылардың 0,05 деңгейінде маңыздылығын дұрыс көрсете отырып, тексеруден өту ықтималдығын табу арқылы есептеуге болады. Айнымалылардың біреуінің маңыздылығын дұрыс көрсету арқылы тексеруден өту ықтималдығы 0,95, сондықтан барлық 20 болжаушының маңыздылығын дұрыс көрсету арқылы тексеруден өту ықтималдығы 0,95 0,95 0,95 болады...  $\times \times$  немесе  $20 \cdot 0,95 \cdot 0,36 = 1$ . Кем дегенде бір болжаушының (жалған) маңыздылығын көрсету ықтималдығы осы ықтималдыққа кері немесе  $1 - (\text{барлығының шамалы болу ықтималдығы}) = 0,64$ . Бұл сұрақ деректерді терең талдаудағы немесе "модельді шуға сәйкестендірудегі" қайта құру мәселесімен байланысты. Неғұрлым көп айнымалыларды қоссаңыз немесе көбірек модельдер жасасаңыз, соғұрлым таза кездейсоқтықпен "мағыналы - менікі" ретінде көрінетін нәрсе пайда болады. Мұғаліммен оқыту тапсырмаларында модельдер бұрын көрмеген мәліметтерге диагноз қойылған кешіктірілген деректерді бақылау жиынтығы бұл тәуекелді азайтады. Белгіленген бақылау жиынтығымен ұштаспайтын статистикалық

және машиналық оқыту міндеттерінде статистикалық шуға негізделген қорытындыларға келу қаупі сақталады.

Статистикада бұл мәселені өте нақты жағдайларда жеңуге арналған бірнеше процедуралар бар. Мысалы, көптеген бақылау топтарындағы нәтижелерді салыстыру кезінде көптеген сандық сұрақтар қоюға болады. Сонымен, А-С нұсқалары үшін сұрауға болады:

- А В-ден өзгеше ме?
- В С-дан өзгеше ме?
- А С-дан өзгеше ме?

Немесе клиникалық сынақта сіз терапия нәтижелерін бірнеше сатыда көргіңіз келуі мүмкін. Әр жағдайда сіз көптеген сұрақтар қоясыз және әр сұраққа сіз кездейсоқтыққа алданып қалу мүмкіндігін арттырасыз. Бірақ бұл алдауды статистикадағы түзету процедуралары арқылы өтеуге болады, гипотезаны бір рет тексеру үшін белгіленгеннен гөрі статистикалық маңыздылыққа қатаң жолақ орнатылады. Мұндай түзету процедуралары, әдетте, тексерулер санына сәйкес " $\alpha$ -деңгейді бөлумен" байланысты. Бұл әдіс әрбір тексеру үшін  $\alpha$ -деңгейінің төмендеуіне әкеледі (яғни статистикалық маңыздылық үшін қатаң жолақ). Осындай процедуралардың бірі, Бонферрониі түзету, жай ғана  $\alpha$ -ны бақылаулар санына бөледі  $n$ . дегенмен, бірнеше салыстыру мәселесі осы жоғары құрылымдалған жағдайлардан асып түседі және деректерді қорлау туралы мәлімдеме тудырған деректерді бірнеше рет "тарау" құбылысымен байланысты. Басқаша айтқанда, егер сізде жеткілікті күрделі деректер болса және олардан қызықты ештеңе таппасаңыз, онда сіз оларға ұзақ және мұқият қарамадыңыз. Бүгінгі күні бұрынғыдан да көп деректер қол жетімді. Сонымен, 2002 және 2010 жылдар аралығында жарияланған журнал мақалаларының саны екі есеге жуық өсті. Біз деректерде қызықты нәрсені, соның ішінде көптік сұрақтарын табуға көптеген мүмкіндіктер аламыз, атап айтқанда:

- барлық топтар бойынша көптеген жұптық айырмашылықтарды салыстыру;
- көптеген кіші топтардың нәтижелерін қарастыру ("тұтастай алғанда, біз опция шарттарының маңызды әсерін таппадық, бірақ үйленбеген әйелдерге 30 - дан кіші әсер таптық");
- статистикалық модельдердің массасын сынау;
- модельдердегі айнымалылар массасын тарту;

- әр түрлі сұрақтардың көп санын қою (яғни әр түрлі мүмкін нәтижелерді алу).

Бірқатар себептерге байланысты, әсіресе жалпы "көптік" мәселесін қоса алғанда, ең жақсы зерттеу міндетті түрде жақсы зерттеуді білдірмейді. Мысалы, Bayer фармацевтикалық компаниясы 2011 жылы 67 ғылыми зерттеулерді қайталауға тырысқанда, олардың тек 14-ін толығымен қайталай алатынын анықтады. 2/3 бөлігін мүлдем қайталау мүмкін болмады. Қалай болғанда да, егжей-тегжейлі анықталған және құрылымдалған статистикалық тексерулерге арналған түзету процедуралары деректерді талдаушылар арасында кеңінен қолдану үшін тым нақты және икемсіз. Құрғақ қалдықтағы көптікке қатысты деректер талдаушылары үшін келесілер болады:

- болжалды модельдеу жағдайында айқын тиімділігі негізінен кездейсоқ мүмкіндіктің өнімі болып табылатын иллюзиялық модельді алу қаупі кросс-тексеру арқылы азаяды (бөлімді қараңыз. 4 - тараудың "қайта тексеру") және кейінге қалдырылған деректермен бақылау үлгісін пайдалану;
- модельді тексеру үшін белгіленген бақылау жиынтығы жоқ басқа процедураларға келетін болсақ, сену керек:
  1. деректерді неғұрлым көп сұрасаңыз және оны басқарсаңыз, жағдайдың пайда болу мүмкіндігі соғұрлым жоғары болатынын түсіну;
  2. кездейсоқ эталондарды (эталондарды) қамтамасыз ету үшін қайта іріктеу мен модельдеуге байланысты эвристика.

### ***Бірнеше тестілеудің негізгі идеялары***

- *Зерттеу жұмысындағы немесе терең талдау жобасындағы көптік - деректер үшін (бірнеше салыстыру, көптеген айнымалылар, көптеген модельдер және т.б.) кездейсоқ мүмкіндіктің арқасында маңызды нәрсе туралы қорытынды жасау қаупін арттырады.*
- *Бірнеше статистикалық салыстырулармен байланысты жағдайлар үшін (яғни, маңыздылығын бірнеше рет тексеру) статистикалық түзету процедуралары бар.*
- *Деректерді терең талдау жағдайында таңбаланған нәтиже айнымалылары бар бақылау үлгісін пайдалану тауарлық емес нәтижелердің алдын алады.*

## Бостандық дәрежелері

Құжаттамада және көптеген статистикалық гипотезаларды тексеруге арналған параметрлерде "еркіндік дәрежелеріне" сілтемелерді көруге болады. Бұл тұжырымдама таңдамалы деректерден есептелген статистикаға қолданылады және еркін өзгеруі мүмкін мәндер санына жатады. Мысалы, егер 10 мәннен алынған үлгінің орташа мәні белгілі болса, сонымен қатар 9 мән белгілі болса, онда 10-шы мән де белгілі. Тек 9 мән еркін өзгеруі мүмкін.

### *Негізгі терминдер:*

*n* немесе үлгі өлшемі (үлгі өлшемі)

*Деректердің бақылаулар саны (жолдар немесе жазбалар).*

*d. f. еркіндік дәрежелері.*

Еркіндік дәрежелерінің саны көптеген статистикалық тексерулердің кірісі болып табылады. Мысалы, еркіндік дәрежесі-дисперсия мен стандартты ауытқуды есептеу кезінде пайда болған 1 n-бөлгішке берілген атау. Бұл ұғымның маңыздылығы неде? Популяцияға қатысты дисперсияны бағалау үшін үлгіні пайдаланған кезде, Сіз N бөлгіште пайдалансаңыз, сәл төмен қарай жылжитын бағамен аяқтайсыз. Егер сіз 1 n-бөлгіште қолдансаңыз, онда бағалау бұл орын ауыстырудан бос болады.

Дәстүрлі Статистика курсының немесе тиісті әдістемелік материалдың айтарлықтай үлесі әртүрлі стандартты гипотеза тексерулеріне жұмсалады (t-Статистика, F-статистика және т. б. негізделген тексерулер). Таңдамалы Статистика дәстүрлі статистикалық пішіндерде пайдалану үшін стандартталған кезде, еркіндік дәрежелері стандартталған деректеріңіздің сәйкес тірек үлестіріміне (t үлестірімі, F үлестірімі және т.б.) сәйкес келуін қамтамасыз етуге арналған стандарттау есептеулерінің бөлігі болып табылады. Еркіндік дәрежелері деректер ғылымы үшін шынымен маңызды ма? Дәл емес. Кем дегенде маңыздылығын тексеру контекстінде. Бір жағынан, ресми жүздік тексерулер деректер ғылымында өте үнемді қолданылады. Бірге дос-гой, деректер мөлшері әдетте соншалықты үлкен, сондықтан деректерді талдаушы үшін, мысалы, N немесе 1 n-бөлгіші бар-жоғы сирек маңызды рөл атқарады .

Дегенмен, еркіндік дәрежелері сәйкес келетін бір контекст бар: регрессияда факторизацияланған айнымалыларды қолдану (логистикалық регрессияны қоса). Егер қатаң саятшылықтар болса, регрессиялық Алгоритмдер тоқтайды-дәл болжамды айнымалылар. Бұл көбінесе категориялық айнымалылар екілік

индикаторларға (жалған айнымалылар) факторизацияланған кезде пайда болады. Аптаның бір күнін алайық. Аптаның 7 күн болғанына қарамастан, аптаның күнін көрсеткен кезде тек 6 еркіндік дәрежесі болады. Мысалы, сіз аптаның күні сенбіге сәйкес келмейтінін білсеңіз, онда бұл күн жексенбі болуы керек екеніне сенімдісіз. Сондықтан дүйсенбі-сенбі индикаторларын қосу мультиколлинеарлық қатеге байланысты регрессия үшін сәтсіздікке себеп болады дегенді білдіреді.

### ***Еркіндік дәрежелеріне арналған негізгі идеялар***

- *Еркіндік дәрежелерінің саны (D. f.) тексеру статистикасын стандарттау мақсатында есептеулердің құрамына кіреді, нәтижесінде оларды анықтамалық үлестірулермен (t - үлестіру, F - үлестіру және т. б.) салыстыруға болады.*
- *Еркіндік дәрежелері ұғымы регрессияны орындау кезінде (мультиколлинеарлықты болдырмау үшін)  $1 \text{ } n$  – индикаторлық немесе жалған айнымалылардағы категориялық айнымалыларды факторизациялаудың негізінде жатыр.*

## **ANOVA**

A / B тестінің орнына біз көптеген топтарды салыстырдық делік, айталық, A-B-C-D, олардың әрқайсысында сандық деректер бар. Топтар арасындағы статистикалық маңызды айырмашылықты тексеруді жүзеге асыратын статистикалық процедура дисперсиялық талдау немесе ANOVA деп аталады (ағылш. analysis of variance).

### ***Негізгі терминдер:***

*Жұптық салыстыру (pairwise comparison) көптеген топтардың екі тобы арасындағы гипотезаны (мысалы, орташа мәндер туралы) тексеру.*

*Әмбебап тест (omnibus test) көптеген топтардың орташа мәндері арасындағы жалпы дисперсия гипотезасын бір рет тексеру.*

*Синонимдер: omnibus тесті, omnibus критерийі.*

*Дисперсияның ыдырауы (variance декомпозициясы) жеке мәнге ықпал ететін компоненттерді бөлектеу (мысалы, жалпы орташа мәннен, опция үшін орташа мәннен және дәл қатенің қалдығынан).*

*F-статистика (F-statistic) кездейсоқ модельде күткеннен жоғары топтық орташа айырмашылықтардың дәрежесін өлшейтін стандартталған статистика.*

*SS "квадраттардың қосындысы" кез-келген орташа мәннен ауытқуды білдіреді.*

Кестеде. 3.3 төрт веб-беттің жабысқақтығын (stickiness, яғни келушілердің назарын аудару дәрежесі) бетте өткізілген секундтар санында көрсетеді. Веб-сайтқа кірушілердің әрқайсысы бір-бірінен алшақтап кетуі үшін төрт бет кездейсоқ ретпен өшіріледі. Әр бетке және кестеге барлығы 5 келуші бар. 3.3 әрбір баған тәуелсіз деректерді білдіреді. 1-беттің бірінші келушісі 2 - беттің бірінші келушісімен ешқандай байланысы жоқ. Айта кетейік, мұндай веб-тестте біз классикалық рандомизацияланған іріктеу жоспарын толығымен жүзеге асыра алмаймыз, онда әрбір келуші белгілі бір көптеген популяциялардан белгілі бір тәртіппен таңдалады. Біз келушілерді келген кезде қабылдауымыз керек. Келушілер сіз - тәуліктің уақытына, аптаның уақытына, жылдың уақытына, Интернеттің күйіне, олар қолданатын құрылғының сипатына және т.б. байланысты тақырыптық түрде ерекшеленуі мүмкін.

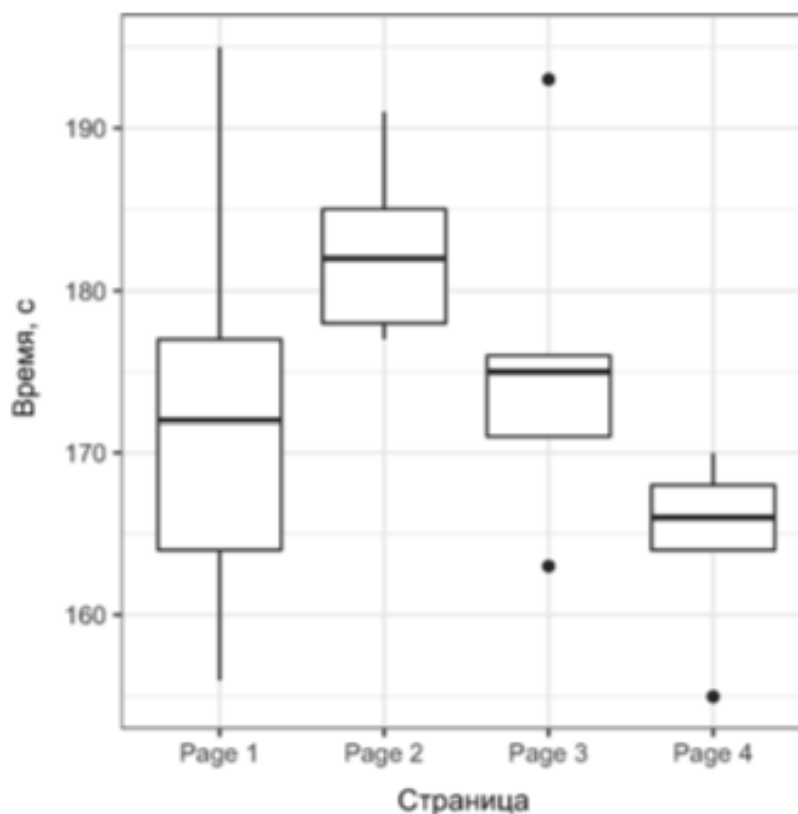
Кесте 3.3. Төрт веб-бетке жабысқақтық (секундпен)

	<b>Страница 1</b>	<b>Страница 2</b>	<b>Страница 3</b>	<b>Страница 4</b>
	164	178	175	155
	172	191	193	166
	177	182	171	164
	156	185	163	170
	195	177	176	168
<b>Среднее</b>	172	185	176	162
<b>Общее среднее</b>				173,75

Енді қиын міндет туындайды (сурет. 3.6). Біз тек екі топты салыстырған кезде, бәрі қарапайым болды: біз әр топтың орташа мәндерінің арасындағы айырмашылықты қарастырдық. Төрт орта жағдайда топтар арасында алты мүмкін салыстыру бар:

- 1 бет және 2 бет;

- 1 бет және 3 бет;
- 1 бет және 4 бет;



3.6. - сурет. Төрт топқа арналған қорап диаграммалары беттер арасында айтарлықтай сәйкессіздіктерді көрсетеді

- 2 бет 3 бетпен салыстырғанда;
- 2 бет және 4 бет;
- 3-бет және 4-бет.

Мұндай жұптық салыстыруларды неғұрлым көп жасасақ, кездейсоқ мүмкіндікке алданып қалу мүмкіндігі соғұрлым жоғары болады (бөлімді қараңыз. Осы тараудың басында "бірнеше тестілеу"). Біз жасай алатын жеке беттер арасындағы барлық мүмкін салыстырулар туралы аландамай, "барлық беттер бірдей жабысқақтыққа ие бола ала ма және бұл беттер арасындағы айырмашылықтар олардың арасындағы шектеудің кездейсоқ сипатынан туындауы мүмкін бе" деген сұраққа жауап беретін бір ғана әмбебап тестті орындауға болады сеанс уақыттарының жалпы жиынтығы?"Бұл сұраққа жауап беру үшін ANOVA процедурасы қолданылады. Оны қолданудың

негізгі принциптерін келесі қайта іріктеу процедурасынан көруге болады (мұнда А-В-С-D-веб-беттердің жабысқақтық сынағы үшін берілген): 1. Барлық деректерді бір қорапта біріктіріңіз. 2. Әрқайсысы бес мәннен тұратын төрт қайталанатын үлгіні араластырыңыз және алыңыз. 3. Төрт топтың әрқайсысының орташа мәнін жазыңыз. 4. Төрт топтың орташа мәндерінің арасында дисперсияны жазыңыз. 5. 2-4 қадамдарды бірнеше рет қайталаңыз (айталық, 1000). Қайта тексерілген дисперсия жағдайлардың қанша үлесін бақыланатын дисперсиядан асып түсті? Бұл р мәні.

Ауыстыру сынағының бұл түрі А-ға қарағанда біршама күрделі тест туралы бөлім. "Осы тараудың басында "ауыстыру тесті". Бақытымызға орай, бұл жағдайда `lmPerm` бағдарламалық жасақтамасында ауыстыру тестін есептейтін `avf` функциясы бар:

```
> library(lmPerm)
```

```
> summary(aovp(Time ~ Page, data=four_sessions))
```

```
[1] "Settings: unique SS "
```

```
Component 1 :
```

```
Df R Sum Sq R Mean Sq Iter Pr(Prob)
```

```
Page      3   831.4   277.13 3104 0.09278 .
```

```
Residuals 16 1618.4   101.15
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

`Pr(Prob)` - де берілген Р мәні 0,09278. `Ite` бағанында ауыстыру тестінде талап етілетін қайталанулар саны көрсетіледі. Басқа бағандар дәстүрлі ANOVA кестесіне сәйкес келеді және төменде сипатталған.

### **F-статистика**

Екі топтың орташаларын салыстыру үшін ауыстыру тестінің орнына қолдануға болатын t-статистикаға негізделген гипотезаны тексеруге ұқсас, ANOVA үшін F-статистикаға негізделген статистикалық тексеру бар. Бұл F-статистика барлық топтық орташа (яғни нұсқа шарттарының әсері) бойынша дисперсияның қалдық қатеге байланысты дисперсияға қатынасына сүйенеді. Бұл қатынас неғұрлым жоғары болса, нәтиже соғұрлым статистикалық маңызды болады. Егер деректер қалыпты үлестірімге бағынатын болса, онда



статистикалық теория статистиканың белгілі бір үлестірімге ие болуын талап етеді. Осы негізде  $p$  мәнін есептеу мүмкіндігі бар. R-де AOV функциясын пайдаланып ANOVA кестесін есептеуге болады:

```
> summary(aov(Time ~ Page, data=four_sessions))
```

Df	Sum Sq	Mean Sq	F value	Pr(>F)
Page	3	831.4	277.1	2.74 0.0776 .
Residuals	16	1618.4	101.2	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Df - "еркіндік дәрежелері", Sum Sq — "квадраттардың қосындысы", Mean Sq — "орташа квадраттар" (орташа квадраттық ауытқулар үшін аббревиатура), F value - F — Статистика. Жалпы орташа, яғни барлық орташа мәндердің орташа мәні үшін квадрат қосындысы-бұл жалпы орташа мәннің 0-ден ауытқуы, квадратталған және 20-ға көбейтілген (бақылаулар саны бойынша). Анықтама бойынша жалпы орта үшін еркіндік дәрежелері 1-ге тең. Варианттың орташа мәндері үшін еркіндік дәрежесі 3-ке тең (үш мән анықталғаннан кейін және одан әрі жалпы орташа мән анықталғаннан кейін, варианттың орташа мәні өзгермейді). Варианттың орташа мәндері үшін квадраттардың қосындысы-варианттың орташа мәндері мен жалпы орташа арасындағы квадраттық ауытқулардың қосындысы. Қалдықтар үшін еркіндік дәрежесі 20 - ға тең (барлық бақылаулар әр түрлі болуы мүмкін) және SS инди — визуалды бақылаулар мен нұсқаның орташа мәндері арасындағы квадраттық айырмашылықтардың қосындысы болып табылады. Орташа квадраттар (MS) — еркіндік дәрежесіне бөлінген квадраттардың қосындысы. F-статистика  $MS(нұсқа)/MS(қате)$ . Осылайша, F мәні тек берілген қатынасқа байланысты және кездейсоқ мүмкіндіктің әсерінен болатын вариацияда күткеннен үлкен нұсқалардың орташа мәндеріндегі айырмашылықтар бар-жоғын анықтау үшін стандартты F үлестірімімен салыстыруға болады.

### ANOVA екі жақты процедурасы

Жаңадан сипатталған A-B-C-D сынағы-бұл бір өзгермелі факторы бар (топ) ANOVA-ның "бір жақты" процедурасы. Бірақ екінші фактор қатысуы мүмкін — мысалы, "жұмыс күнімен салыстырғанда демалыс", мұнда деректер әр комбинация бойынша жиналады (A тобы-демалыс, A тобы-жұмыс, B тобы-демалыс және т.б.). Бұл "екі жақты ANOVA процедурасы" және біз онымен

"өзара әрекеттесу әсерін" анықтау арқылы бір жақты ANOVA процедурасы сияқты жұмыс істейтін боламыз. Жалпы орташа әсерді және нұсқа шарттарының әсерін анықтағаннан кейін біз әр топ үшін демалыс және жұмыс күндері өткізілген бақылауларды бөлеміз және осы ішкі жиындар үшін орташа және опция орташа арасындағы айырмашылықты табамыз. ANOVA процедурасы және ANOVA екі жақты процедурасы көптеген факторлар мен олардың әсерлерін модельдеуге болатын регрессия және логистикалық регрессия сияқты толық статистикалық модельге алғашқы қадамдар екенін көруге болады.

### ***ANOVA процедурасының негізгі идеялары***

- *ANOVA - көптеген топтармен тәжірибе нәтижелерін талдауға арналған статистикалық процедура.*
- *Бұл процедура A / B сынағы үшін ұқсас процедуралардың кеңеюі болып табылады және жалпы вариация кездейсоқ вариация диапазонындағы топтар арасында екенін бағалау үшін қолданылады.*
- *ANOVA-ның пайдалы нәтижесі-топтық опциялармен, өзара әрекеттесу әсерлерімен және қателіктермен байланысты дисперсия компоненттерін анықтау.*