Қазақстан Республикасы Ғылым және жоғары білім министрлігі

«Ғұмарбек Дәукеев атындағы Алматы энергетика және байланыс университеті» коммерциялық емес акционерлік қоғамы

Л.К.Ибраева

ИНЖЕНЕРЛІК ТӘЖІРИБЕ ТЕОРИЯСЫ ЖӘНЕ ТЕХНИКАСЫ

Оқу құралы

Алматы 2022

ӘОЖ 62.(075.8)
КБЖ 30.я73
И15

Пікір берушілер:
техника ғылымдарының докторы, АҚ «ХАТУ» профессоры
Б.К. Синчев,
физика-математика ғылымдарының кандидаты, Қ.И.Сатпаев атындағы ҚазҰТЗУ профессоры
А.Н.Молдагулова
Алматы энергетика және байланыс университетінің ММ кафедрасының аға оқытушысы
Г.А. Рысбекова

Алматы энергетика және байланыс университетінің Ғылыми кеңесі баспаға ұсынды (12.11.22ж. №7 хаттама). АЭжБУ 2022ж. әдебиеттер басылымдарын шығарудың қосымша тақырыптық жоспары бойынша басылады, реті 39

Ибраева Л.К.
И15 Инженерлік тәжірибе теориясы және техникасы: Оқу құралы (7М07105 - «Автоматтандыру және басқару» білім беру бағдарламасының магистранттарына арналған) Л.К.Ибраева – Алматы: АЭжБУ, 2022 –115 б.: кесте - 9, без. - 10, әдеб. көрсеткіші – 18 атау.

ISBN 978-601-358-056-2

Оқу құралында тәжірибелерді жоспарлау әдістерін қолдана басқару объектілерінің регрессиялық модельдерін құрудың жалпы мәселелері қарастырылады. Регрессиялық талдаудың негізгі алғышарттары , тәжірибелік модель коэффициенттерін есептеу және нәтижелерді статистикалық талдау көрсетілген. Регрессиялық талдау жоспарларының оңтайлығының негізгі критерийлері, толық факторлық және бөлшек факторлық тәжірибелер әдістерімен тәжірибелерді жүргізу, екінші ретті жоспарлар, сондай-ақ, ізденіс оңтайландыру әдістері тұжырымдалған.

Оқу құралы 7М07105 – Автоматтандыру және басқару білім беру бағдарламасы бойынша оқитын магистранттарға арналған.

ӘОЖ62.(075.8)
КБЖ 30.я73

ISBN 978-601-358-056-2
©АЭжБУ, 2022
Ибраева Л.К.,2022

Кіріспе

Күрделі жүйелерді (осындай жүйелерге басқару объекттер да жатады) жобалағанда және жасағанда қарастырып отырған объекттердің сандық және сапалық қасиеттерін білу қажет. Көп деген жағдайлар себебінен әдетте күрделі объекттердегі осындай заңдылықтарын практикалық тексеру мүмкін емес. Сонымен бірге осындай тексеру көп материалдар шығындарын және көп уақытты талап етеді. Осы себептерден қарастырып отырған күрделі объекттердің қасиеттері мен заңдылықтарын *модельдеу әдістері* негізінде зерттеудің өте көп мағынасы бар. Берілген объекттің барлық немесе кейбір қасиеттерімен сәйкес келетін қасиеттері бар кез келген объектті берілген объекттің *моделі* деп атаймыз. Модельдеудің маңызды түрі *математикалық* модельдеу болып табылады. Математикалық өрнектермен немесе алгоритмдермен формалданған жүйе бейнеленуі *математикалық модельдеу* деп аталады. Математикалық модель оригиналдың тек қана математикалық бейнелеуі бар және математикалық өрнектермен байланысқан параметрлерін зерттеуге мүмкіндік береді. Сонымен бірге зерттелетін процестің физикасы сақталмайды, математикалық модельдер табиғаты жағынан әртүрлі құбылыстарды бірдей теңдеулермен бейнелеп, объекттің жүріс-тұрысын толық бейнелемей, оның бөлек функционалды байланыстарын табуға мүмкіндік береді. Сонымен, нақты объекттің математикалық моделі деген берілген физикалық объектке сәйкес қойылған математикалық объект деуге болады..

Басқару объектінің математикалық моделін құруда қолданылатын ақпарат көзіне байланысты аналитикалық және статистикалық (эмпирикалық) модельдер ажыратылады.

Аналитикалық модельдер күрделі теңдеулер жүйесі түрінде (алгебралық, дифференциалдық, интегралдық немесе дифференциалды- интегралдық) ұсынылады, бұл объектідегі процестерді дәл сипаттауға және осы процестерді тікелей бақылау мүмкін емес факторлық кеңістік нүктелеріне экстраполяцилауяға мүмкіндік береді.

Модельденген объект туралы теориялық ақпарат болмаған немесе шектеулі болған кезде, оның қасиеттерін сипаттайтын қатынастардың жуықтаған түрі белгісіз болса, математикалық сипаттамалау теңдеулері қолданыстағы объектіні зерттеу нәтижесінде алынған эмпирикалық тәуелділіктер жүйесі бола алады. Яғни, статистикалық модельдер зерттелетін объектіде жиналған тәжірибелік ақпаратты статистикалық өңдеу нәтижесінде алынады. Бұл модельдер салыстырмалы түрде қарапайым құрылымға ие және көбінесе полиномдар түрінде ұсынылады. Олар тәжірибелер жүргізілетін жұмыс нүктелерінің жақын аймағында қолданылады. Көптеген жағдайларда мұндай модельдерді құру уақыт пен шығындардың кіші көлемін талап етеді.

Статистикалық модельдерді алу үшін тәжірибелерді жоспарлау әдістері қолданылады. Тәжірибелерді жоспарлау теориясы әртүрлі физикалық сипаттағы объектілерді зерттеуде тәжірибелерді оңтайлы ұйымдастырудың амалдары мен әдістерін тұжырымдайды. Әртүрлі эмпирикалық

(статистикалық) модельдерді құрудың негізі болатын математикалық теория – жаппай кездейсоқ құбылыстардың заңдылықтарын зерттейтін ықтималдық теориясы болып табылады.

Тәжірибелік модельдің негізгі ерекшелігі - мұндай модель кез келген белгілі тәжірибеде объектінің әрекетін дәл сипаттай алмайды. Статистикалық модель объектінің мінез-құлқын орта есеппен сипаттайды, бірдей жағдайда тәжірибелер бірнеше рет қайталаған кезде ғана көрінетін объектінің кездейсоқ емес қасиеттерін көрсете алады, .

1 Басқару объектілерді статистикалық модельдеу

1.1 Бақылау қатарларының арасындағы байланыстар түрлерін сипаттау

Өлшеу нәтижелері негізінде объект, құбылыс туралы ақпарат алуға арналған объектіге қолданылатын операциялар мен әсерлер жүйесі тәжірибе деп аталады. Тәжірибе барысында алынған ақпаратты өңдеу әдістерінің маңызды міндеті зерттелетін құбылыстың, процестің, объектінің математикалық моделін құру мәселесі, осы модель белгілі бір оператор түрінде көрсетіледі: $y = f(\overline{x})$. Оны процестерді талдауда және объектілерді жобалауда қолдануға болады. Тәжірибе барысында алынған ақпаратты өңдеудің тағы бір міндеті - оңтайландыру мәселесі, яғни әсер ететін тәуелсіз айнымалылардың таңдалған оптималдылық көрсеткіші экстремумды мәнді қабылдайтын комбинациясын табу.

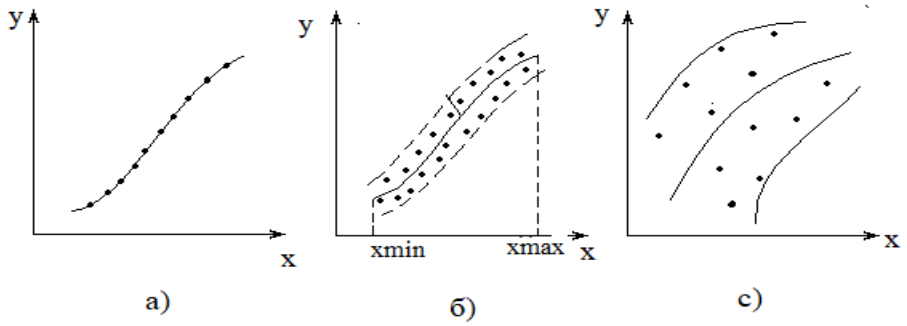
Тәжірибелер қателіктерін талдау және нәтижелерді өңдеу кезінде математикалық статистика және ықтималдық теориясы аппараты кеңінен қолданылады. Ықтималдықтар теориясы мен математикалық статистиканың негізгі ұғымдарын қарастырғанда, таралу заңдар параметрлерін анықтағанда сынақтарының жеткілікті, шегінде шексіз саны *N,* жүзеге асырылады деп есептеледі, бұл іс жүзінде мүмкін емес.

Сондықтан, әдетте, шектеулі көлемді таңдама арқылы *бас жиынның* параметрлерін бағалауға мүмкіндік беретін *таңдамалы әдіс* қолданылады. Көптеген өлшеулерді орындау қажеттілігі объектіде басқа шамалардың функциялары болып табылатын кездейсоқ шамалардың болуымен байланысты. Тәжірибелерді жүргізудің мақсаты $y = f(\overline{x})$ функционалдық тәуелділік түрін анықтау болып табылады. Ол үшін $\overline{x}$ векторының мәндерімен бірге, оларға сәйкес келетін y мәндері де бір уақытта анықталуы керек, ал тәжірибелердің міндеті зерттелетін тәуелділіктің математикалық моделін құру болып табылады. Бақылаулардың (өлшеулердің) екі қатары арасындағы *байланысты* орнату негізгі мақсат болып табылады.

Жалпы жағдайда байланыстар әртүрлі және күрделі болады. Әдетте келесі байланыстар түрлері қарастыруға болады.

Функционалдық байланыстар (тәуелділіктер). Осындай байланыстарда $\overline{x}$ мәні өзгерген кезде басқа y шаманың мәні өзгеріп, x_i -дің әр мәніне y_i -тің

нақты анықталған мәні сәйкес келеді (1.1, а сурет). Осылайша, тәжірибелердің барлық жағдайлары бірдей таңдалса, онда сынақтарды қайталай отырып, бірдей тәуелділікке ие боламыз, яғни қисықтар барлық сынақтар үшін өте жақсы сәйкес келеді. Өкінішке орай, мұндай жағдайлар нақтылықта кездеспейді. Сонымен қатар, бөлек әрбір кездейсоқ фактордың әсері кіші болуы мүмкін, бірақ, барлығы бірге олар тәжірибе нәтижесіне маңызды әсер етуі мүмкін. Бұл кезде айнымалылар арасындағы стохастикалық (ықтималдық) байланыстар туралы айтылады.



а) функционалдық байланыс, барлық нүктелер қисықта орналасады; б) байланыс жеткілікті тығыз, нүктелер регрессия қисығының аймағында топтасады, бірақ онда орналаспайды; в) байланыс әлсіз

1.1 сурет – Бақылау қатарлары арасындағы байланыс түрлері

Байланыстың *стохастикалық* болуы келесіде: бір кездейсоқ шама y басқа $\overline{x}$ шаманың өзгеруіне өзінің *таралу заңының* өзгеруімен жауап береді (4.1, б сурет). Осылайша, тәуелді айнымалы бір нақты мәнді емес, бірнеше мәндерді қабылдайды. Сынақтарды қайталау кезінде жауап беру функциясының басқа мәндері алынады және $[x_{min};\ x_{max}]$ интервалындағы бірдей $\overline{x}$ мәніне әртүрлі сынақтарда y-ің әртүрлі мәндері сәйкес келеді. Ізделінетін тәуелділікті $\hat{y} = f(x)$ алынған $\overline{x}$ және y мәндерін бірлесіп өңдеу нәтижесінде табуға болады.

4.1, б - суретте тәжірибелік нүктелер жолағының ортасынан өтетін тәуелділік қисығы (математикалық күтім), ізделінетін $\hat{y} = f(x)$ қисығында орналаспауы мүмкін, бірақ оның айналасында белгілі бір жолақ алады. Бұл ауытқулар өлшеу қателіктерінен, модельдің толық еместігінен және ескерілетін факторлардан, зерттелетін процестердің кездейсоқ сипатынан және басқа себептерден пайда болады.

Стохастикалық байланыстарды талдау кезінде айнымалылар арасындағы тәуелділіктің келесі негізгі түрлерін ажыратуға болады.

1. Бір кездейсоқ $\overline{x}$ айнымалы мен басқа y кездейсоқ айнымалы және олардың шартты орташа мәндері арасындағы байланыс *корреляциялық* деп аталады; $\bar{y}_i$ шартты мәні – бұл кездейсоқ $\overline{x}$ мәні $\overline{x}_i$ мәнін қабылдаған жағдайда y кездейсоқ шамасын жүзеге асыруларының арифметикалық ортасы.

2. y кездейсоқ айнымалысының $\overline{x}$ кездейсоқ емес айнымалыға тәуелділігі немесе y кездейсоқ шамасының M_y математикалық күтімінің *детерминді* $\overline{x}$

мәніне тәуелділігі *регрессиялық* деп аталады. Келтірілген тәуелділік $\overline{x}$ шамасының өзгеруінің *y* шамасының орташа мәніне әсерін сипаттайды.

Стохастикалық тәуелділіктер *формасы, байланыс тығыздығы* және *регрессия теңдеуінің коэффициенттерінің сандық мәндерімен* сипатталады.

Байланыс формасы $\hat{y} = f(x)$ функционалды тәуелділік түрін белгілейді және *регрессия теңдеуімен* сипатталады. Функционалдық тәуелділіктің түрін таңдау есебі - формалданбаған (бейресми) есеп, өйткені белгілі аралықта бір қисықты бірдей дәлдікпен әртүрлі аналитикалық өрнектермен сипаттауға болады. Осыдан маңызды практикалық қорытынды шығады: белгілі бір математикалық модельді таңдау туралы шешім қабылдау зерттеушіде қалады. Бұл модель болашақта не үшін қолданылатынын, оның параметрлері қандай тұжырымдамалар негізінде түсіндірілетінін тек тәжірибе орындаушы біледі.

Байланыстың *тығыздығы* деп стохастикалық тәуелділіктің функционалды тәуелділікке жақындық дәрежесін түсінеміз, яғни қабылданған модель теңдеуіне қатысты тәжірибелік нәтижелерінің топтасу тығыздығының көрсеткіші (4.1, б сурет).

Тәжірибелерді жоспарлау теориясы әртүрлі физикалық сипаттағы объектілерді зерттеуде тәжірибелерді оңтайлы ұйымдастырудың амалдары мен әдістерін тұжырымдайды.

Тәжірибе - зерттеу сынақтары кезінде объект туралы ақпарат алуға бағытталған операциялар, әсерлер және бақылаулар жүйесі. *Сынақ* – нәтижелерді тіркеу мүмкіндігі болған кезде зерттелетін объектіні белгіленген жағдайларда жаңғырту (тәжірибелердің элементарлы бөлігі).

Тәжірибелерді жоспарлау - өлшеулер мен бақылауларды ұтымды ұйымдастыру. Тәжірибелерді жоспарлау объектіні белгілі бір дәлдікпен зерттеу үшін қажетті және жеткілікті сынақтар саны мен шарттарын таңдау процедурасынан тұрады.

Тәжірибені жоспарлау мыналарды *қамтамасыз етеді*:

- арнайы ережелер бойынша барлық айнымалыларды бір мезгілде өзгерту;
- тәжірибе орындаушының көптеген әрекеттерін рәсімдейтін математикалық аппаратты қолдану;
- тәжірибенің әрбір сериясынан кейін негізделген шешімдер қабылдауға мүмкіндік беретін нақты стратегияны таңдау;
- тәжірибелер санын, ресурстарды (қаржылық, уақытша, материалдық, адами) минимумдау.

Тәжірибені жоспарлаудың мақсаты - аз еңбек шығындарда объект туралы сенімді және ақиқат ақпарат алуға болатын сынақтарды жүргізудің шарттары мен ережелерін табу, сонымен қатар бұл ақпаратты дәлдікті сандық бағалаумен ықшам және ыңғайлы түрде ұсыну.

1.2 Тәжірибелік зерттеулер түрлерін классификациялау

Нәтижелерді ұсыну түрі бойынша тәжірибелерді келесідей түрлерге бөледі:

- *сапалық тәжірибе.* Кейбір құбылыстардың болу фактісі анықталады, бірақ сандық сипаттамалар берілмейді. Қаншалықты қиын көрінсе де кез келген тәжірибе, нәтижелерді ұсынумен, қорытындыларды тұжырымдаумен және ұсыныстар берумен аяқталады. Бұл ақпарат графиктер, сызбалар, кестелер, формулалар, статистикалық мәліметтер немесе ауызша сипаттамалар түрінде ұсынылуы мүмкін. Сапалы тәжірибе, әдетте, ауызша сипаттаманы қамтамасыз етеді. Бірақ, ауызша сипаттау нәтижелерді ұсынудың ең тиімсіз әдісі болып табылады, өйткені ол сандық ұсыныстар беруге, объектінің қасиеттерін басқа жағдайларда талдауға және оны басқару мәселелерін шешуге мүмкіндік бермейді. Инженерлік практикасында тәжірибенің негізгі мазмұны сан немесе сандық тәуелділікпен ұсынылуы керек.

- *сандық тәжірибе* - белгілі бір құбылыстың болуын тіркеп қана қоймай, процестің барысын анықтайтын факторлар арасында сандық қатынастарды орнатуға, сондай-ақ, осы факторлардың белгілі бір құбылысқа әсер етуінің математикалық моделін құруға мүмкіндік береді.

Өткізу шарттары бойынша келесідей айырылады:

- *зертханалық тәжірибе* – зертханалық жағдайда кездейсоқ қателіктердің әсері аз, сынақтарды жүргізу жағдайларының «стерильділігі» қамтамасыз етіледі, көп жағдайда мұқият дайындық жүргізіледі, бір сөзбен айтқанда, «тәжірибе мәдениеті» жоғары. Әдетте, өнеркәсіп жағдайына қарағанда зертханада зерттеуші сынақты әлдеқайда жақсы өткізе алады. Басқа сөзбен айтқанда, бірдей шарттарда, бір фактіні анықтау үшін зертханаға қарағанда зауытта әлдеқайда көп тәжірибе қажет болады. Тағы бір маңызды айырмашылық – факторлардың өзгеру мүмкіндігіне әртүрлі шектеулер. Мысалы, зертханада химиялық реакция зерттелгенде, температураны кең ауқымда өзгертуге болады, ал металлургиялық пештерде, керісінше, оны өзгерту болатын болсада, әлдеқайда тар диапазонда және сақтықпен;

- *өнеркәсіптік тәжірибелерде* бұл жағдайларды қамтамасыз ету әлдеқайда қиын. Өлшеу және ақпарат жинау күрделене түседі, тәжірибені, өлшеу құралдарын ұйымдастыруға және өткізуге әртүрлі кедергілердің әсері едәуір көп, сондықтан арнайы әдістерді қолдану қажет. Өлшемдердің аз саны бойынша мүмкіндігінше ең сенімді нәтижелерді алу қажет. Қазіргі математикалық статистикада өлшемдердің бірдей санымен дәлдікті арттыруға немесе тіпті азайтылған кезде көбірек ақпарат алуға мүмкіндік беретін арнайы әдістер бар екенін ескеріңіз.

Объектіге әсер ету түрі бойынша:

- *активті тәжірибе* - процестің жұмыс істеу шарттарын зерттеуші мақсатты түрде өзгертеді;

- *пассивті тәж*ірибе - процестің жұмыс істеу шарттарын зерттеуші әр тәжірибеде тіркейді, бірақ оның мәнін орнатпайды.

Соңынан осындай тіркеудің нәтижелері өңделеді және ұсыныстар беріледі.

1.3. Зерттеу нәтижелерінің қателіктері

Тәжірибе нәтижелері әдетте дәл болмайды. Әртүрлі себептерге байланысты кез келген екі параллель тәжірибенің нәтижелері бір-бірінен ерекшеленеді, кездейсоқ сәйкестіктерден басқа. Тәжірибенің дәлдігі деп оның сапасы түсініледі, ол алынған нәтижелердің қажетті шаманың шынайы мәніне жақындығын көрсетеді. Тәжірибенің қателігі неғұрлым аз болса, оның дәлдігі соғұрлым жоғары болады. Қателіктер түрлері төменде келтірілген.

Абсолютті қателік – x тәжірибе нәтижесі мен ізделінетін шаманың x^* шынайы мәні арасындағы Δx айырымы:

$$\Delta = |x - x^*|.$$

Салыстырмалы қателік:

$$\Delta x^* = \frac{|x - x^*|}{x^*} \cdot 100\% = \frac{\Delta}{x^*} \cdot 100\%.$$

Айта кету керек, тәжірибе нәтижесінде анықталған шаманың шын мәні әрқашан белгісіз болып қалады, сондықтан тәжірибе қателіктерін тек қана шамамен бағалауға болады.

Δx абсолютті қателіктің $x_н$ нормаланған мәніне пайызбен көрсетілген қатынасы келтірілген қателік болып табылады:

$$\gamma = \frac{\Delta x}{x_н} \cdot 100\%.$$

Нормалау мәні ретінде өлшенетін шаманың қабылданған мәні алынады; бірліктер ретінде, әдетте, шкаланың жоғарғы және төменгі шектерінің айырмашылығы қолданылады.

Тәжірибелер өткізгенде пайда болу себебінен оның қателіктері шартты түрде жүйелі, кездейсоқ және өрескел (қателіктер) болып бөлінеді.

Қайталанған тәжірибелерде тұрақты болып қалатын немесе бір заңдылық бойынша өзгеретін қателік *жүйелік* деп аталады. Жүйелік қателіктердің болуын өлшенетін шаманың бірдей мәнін әртүрлі әдістермен немесе құрылғылармен өлшеу жағдайларын талдау арқылы анықтауға болады. Айнымалы жүйелік қатенің мысалы ретінде қуат көзінің кернеуін өлшеу қателігі болуы мүмкін, егер өлшеу нәтижесі кернеуге (мысалы, потенциометрге) байланысты болса. Жүйелік қателіктерді параллель тәжірибелер санының көбеюімен азайтуға болмайды. Оларды тудыратын себептер жойылуы керек. Жүйелік қателіктердің себептерін анықтаудың жалпы

әдісі калибрлеу (тексеру) болып табылады, ол белгілі эталонның көмегімен өлшенетін шаманың барлық диапазонында құрылғыны тексеру болып табылады. Құрылғы көрсеткіштердің өте аз шашырауын бере алады, бірақ жүйелі қатенің болуына байланысты нәтиже дұрыс болмайды.

Жүйелік қателіктердің келесі көздерін ажыратуға болады:

- өлшеу құралдарының *инструменталды* (аспаптық немесе аппаратуралық) қателіктері депосы өлшеу құралына жататындар аталады, оларды сынау кезінде анықтауға және оның паспортына енгізуге болады. Өлшеу құралдарының *негізгі қателігін*, яғни қалыпты жағдай деп қабылданған жағдайдағы қателігін және әсер етуші параметрлердің қалыпты мәндер аймағынан (діріл, орта ылғалдылығы, инерция және т. б.) ауытқуынан туындаған *қосымша* қатені ажырату әдетке айналған;

- *әдістемелік* қателіктер дегенімі: бұл құрылғыға жатқызуға болмайтын, оның паспортында көрсетілмейтін, яғни құрылғының өзімен емес, өлшеу әдісімен байланысты қателер. Көбінесе әдістемелік қателіктің пайда болу себебі - өлшеулерді ұйымдастыру кезінде, өлшеу қажет емес шаманы өлшеуге немесе өлшеуге мәжбүр болу, ол шама жақын, бірақ оған тең емес.

Әдістемелік қателіктердің ерекшелігі: оларды зерттелетін объектінің математикалық моделін құру арқылы ғана анықтауға болады және оларды тек өлшеу құралының өзін мұқият зерттеу арқылы табуға болмайды:

- зерттеушінің ерекшеліктеріне байланысты *субъективті* қателер.

Жүйелік қателіктерді толығымен жою мүмкін емес екенін есте ұстаған жөн, өйткені жүйелі қателерді анықтайтын және бағалайтын әдістер мен құралдардың өз қателіктері бар.

Тәжірибеден тәжірибеге кездейсоқ өзгеретін, бірқатар себептердің әсерінен болатын қателік *кездейсоқ* қателік деп аталады. Бұл қатенің мәнін әр тәжірибеде анықтау және оған әсер ету мүмкін емес. Сонымен қатар, көптеген тәжірибелердің нәтижесінде қателіктердің осы түріне тән кейбір заңдылықтар анықталуы мүмкін. Кездейсоқ қателіктерге пайда болу себептері белгісіз тұрақты емес қателер жатады. Бұл қателіктер, әдетте, тәжірибе орындаушыға белгісіз және талдауға қиын болатын өзгеретін факторлардың күрделі жиынтығынан туындайды. Осылайша, кездейсоқ қателер өлшенетін шаманың нақты мәніне қатысты құрылғы көрсеткіштерінің кездейсоқ ауытқуы. Тәжірибелер кезінде пайда болатын кездейсоқ қателіктерді зерттеу үшін математикалық статистика мен ықтималдық теориясы кеңінен қолданылады.

Өрескел қателіктер тәжірибе шарттарының, өлшеу сапасының күтпеген өзгеруінен, құрылғының сынуы, жұмыс журналдарындағы дұрыс емес жазба, құрылғының механикалық соққысы, құрылғы көрсеткіштерінің дұрыс емес есебі, қуат көзін өшіру және т.б. нәтижесінде пайда болады. Мұндай нәтижелер тәжірибе нәтижелерін математикалық өңдеуден бұрын алып тасталуы керек.

1.4 Бақылау сұрақтары

1.4.1 Кездейсоқ шаманы анықтамасын беріңіз.

1.4.2 Келесі ұғымдарды анықтаңыз: ықтималдылықтардың таралу заңы, ықтималдылықтар таралуының интегралды функциясы, ықтималдылықтар таралуының дифференциалды функциясы:

1.4.3 Бас жиын деген не? Таңдама анықтамасын беріңіз.?

1.4.4 Таңдамалы әдістің міндеті?

1.4.5 Корреляциялық байланысты анықтаңыз.

1.4.6 Регрессиялық тәуелділік дегеніміз не?

1.4.7 Тәжірибелік зерттеулер түрлерінің классификацяисын келтіріңіз.

1.4.8 Кездейсоқ және жүйелік қателіктерді анықтаңыз.

1.4.9 Жүйелік қателіктердің көздерін атаңыз.

1.4.10 Кездейсоқ өлшеу қателіктерінің пайда болуының себебі неде?

2 Тәжірибелерді жоспарлау теориясының негізгі түсініктемелері

2.1 Анықтамалар

Зерттеу мақсатын тұжырымдау кезеңінде «қара жәшік» идеясына негізделген кибернетикалық тәсілді қолданамыз. «Қара жәшік» термині нақты объектілердің кіріс тәуелсіз айнымалылары шығыс тәуелді айнымалылардың әрқайсысына әсер ететін аз немесе нашар зерттелген қасиеттерінің моделін білдіреді. Бұл жағдайда кіріс айнымалыларының мәндері мен объектінің шығыс айнымалыларының мәндері арасында белгілі байланыс болады (2.1 сурет). Зерттеуші «қара жәшіктің» кірістері мен шығуларын бақылай алады және бақылау нәтижелері бойынша кірістер мен шығулар арасындағы байланысты анықтай алады деп саналады. Кірістер детерминді (бақыланатын, басқарылатын) немесе кездейсоқ (бақыланатын, бірақ басқарылмайтын) өзгеруі мүмкін.

$\bar{x} = (x_1, \dots, x_n)$ параметрлер кешені негізгі деп аталады, ол тәжірибе шарттарын анықтайды. Кездейсоқ параметрлер ретінде, әдетте, белгілі бір себептермен ескерілмейтін (немесе ескеруге өте қиын) параметрлер қарастырылады. Мұндай негізгі және кездейсоқ параметрлерге бөлу шартты болады. Негізгі кешенге кірмеген кез келген параметр, егер ол жақсы зерттелген болса да, кездейсоқ болады.

$x_1 \dots x_n$ кіріс айнымалылары *факторлар* деп аталады.

Тәуелді y шығыс айнымалысы – *объект жауабы* (реакция) деп аталады. Яғни, *жауап* (факторларға реакция) - бұл бақыланатын кездейсоқ айнымалы, болжам бойынша факторларға тәуелді.

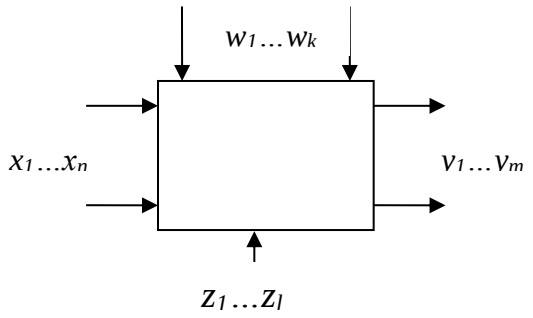
2.1 суреттегі сұлбада келесідей белгілеулер:

$\bar{x} = (x_1, \dots, x_n)$ – объектінің кірістегі бақыланатын және басқарылатын параметрлері;

$\bar{z} = z_1, \dots, z_l)$ - бақыланатын, бірақ басқарылмайтын кіріс параметрлері;

$\bar{w} = (w_1, \dots, w_k)$ – бақыланбайтын, кездейсоқ ретімен өзгеретін параметрлер, объектінің «шуы";

$\bar{y} = (y_1, \dots, y_m)$ – объектінің бақыланатын шығысы.

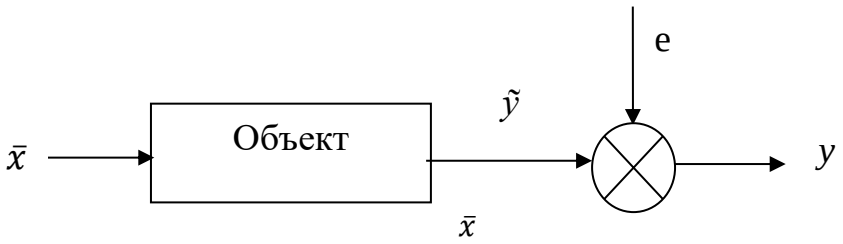


2.1 сурет – Зерттелетін объектінің құрылымдық сұлбасы

Әдетте $\bar{z}$ және $\bar{w}$ айнымалылары бір $\bar{e}$ векторына біріктіріледі, оның әсерінен объектінің шығуы жеткілікті жүйелі түрде («объектінің дрейфі» болған жағдай) немесе кездейсоқ өзгеруі мүмкін – бұл жағдайда кездейсоқ шулар өрісінің болуы туралы айтады. Сонымен бірге, әдетте, шу өрісінің ықтималдық қасиеттері уақыт өте келе өзгермейді деп саналады.

Физикалық табиғаты бойынша $\bar{e}$ айнымалыларына ең алдымен өлшеу құралдарының немесе талдау әдістерінің қателері, сондай-ақ, өнеркәсіптік объектілердегі шикізаттың бақылаусыз өзгеруі немесе қондырғылардың қартаюы мен тозуына, сыртқы ортаның әсеріне және т. б. байланысты қасиеттердің өзгеруі, соның ішінде бақыланатын, бірақ зерттелетін айнымалылар санына кірмейтін (оларды өлшеу қиындықтары себептерінен) айнымалылардың әсері жатады. $\bar{e}$ айнымалылар тобына зерттеушінің квалификациясы, пайдаланылатын тәжірибелік жабдықтың ерекшеліктері және өлшеу әдіснамасы, метрологиялық қамтамасыз ету деңгейі, яғни осы нақты тәжірибеде әрекеті көрінбеуі мүмкін, бірақ әртүрлі тәжірибелердің, зертханалардың және т. б. нәтижелерін салыстыру сатысына әсер етуі мүмкін айнымалылар кіруі керек.

Тәжірибелерді жоспарлау әдістерінің көпшілігі $\bar{y}$ векторының өлшемі 1-ге тең болған жағдайда және $\bar{e}$ айнымалыларының (шу өрісінің) әсерін y өлшенетін шаманың аддитивті компоненті ретінде ұсынуға болады, яғни объектінің сұлбасы эквивалентті сұлба түрінде ұсынылады, мұндағы $\tilde{y}$ - объектінің қатесіз шығуы (2.2 сурет).



2.2 сурет- Зерттелетін объектінің *e* аддитивті шулары бар эквивалентті сұлбасы

Тәжірибелерді жоспарлау практикасындағы зерттеу объектілері мыналар болуы мүмкін: нақты физикалық объектілер, нақты объектілердің физикалық модельдері, нақты объектілердің математикалық модельдері (соңғы екі жағдайда $\bar{e}$ шуына объектіні физикалық немесе математикалық модельдеудің дәл еместігіне байланысты компоненттер кіреді).

Объект жауабы сандық түрде анықталуы керек. Алайда, *y*-тің сапалық белгілері де болуы мүмкін. Бұл жағдайда рангтік тәсілін қолдануға болады. Рангтік тәсілінің мысалы - студенттің білімі туралы алынған ақпараттың күрделі кешенін бір санмен бағалаған кездегі емтихан бағасы.

Тәуелсіз айнымалылары $x_1, x_2, \ldots, x_n$ – факторлардың да сандық бағалары болуы керек. Егер сапалы факторлар қолданылса, онда олардың әр деңгейіне қандай да бір сан меншіктелу керек. Факторлар ретінде тек тәуелсіз айнымалыларды таңдау маңызды, яғни басқа факторларды өзгертпей өзгертуге болатын айнымалыларды. Тиімді математикалық модельді құру үшін факторлардың маңыздылығына (функцияға әсер ету дәрежесіне), олардың ранжирлеуіне алдын-ала талдау жүргізіп, маңыздылығы төмен факторларды алып тастаған жөн.

Фактор – уақыттың әр сәтінде озінің анықтау аймағынан белгілі бір мән алатын және объектіге сыртқы әсерді көрсететін кейбір айнымалы мән.

Тәжірибеде факторлар қабылдайтын мәндер факторлар деңгейлері деп аталады.

Фактор деңгейі – объектіні тәжірибелік зерттеу кезінде өзінің анықтау саласынан қабылданатын фактордың белгілі мәні.

Тәжірибелер барысында кіріс факторларының нақты мәндерінің белгілі бір жиыны объектінің бір күйіне сәйкес келеді. Зерттеу есебін және зерттелетін объект туралы априорлық ақпаратты талдау негізінде әр кіріс факторы үшін тәжірибе кезінде оның өзгеретін аймағы анықталады. Барлық кіріс факторлары үшін осындай аймақтардың комбинациясы тәжірибенің факторлық кеңістігі деп аталады.

Факторлық кеңістік – бұл тәжірибені дайындау және жүргізу кезінде зерттеуші мәндерін басқара алатын факторлар жиыны.

Тәжірибелер мақсаты - *y* функциясын табу болады, сонымен бірге объект жауабының мәні екі компоненттен тұрады деп болжанады:

$$y = f(x_1, x_2, \ldots, x_n) + e(x_1, x_2, \ldots, x_n), \qquad (2.1)$$

мұндағы $f(x_1 \ldots x_n)$ – жауап функциясы (факторларға тәуелді кездейсоқ емес функциясы);

$e\ (x_1 \ldots x_n)$ – тәжірибе қатесі (кездейсоқ шама);

$x_1 \ldots x_n$ – факторлық кеңістіктегі факторлар деңгейлерінің белгілі бір комбинациясы.

y кездейсоқ айнымалы болатыны анық, өйткені ол $e\ (x_1 \ldots x_n)$ кездейсоқ шамасына тәуелді.

Өлшеу дәлдігін сипаттайтын *D[y]*дисперсиясы D[y]=D[e] тәжірибелер қателігінің дисперсиясына тең *D[y]=D[e]*.

Жауап беру функциясы - жауаптың математикалық күтімінің факторларға тәуелділігі.

Жауап беті -жауап беру функциясының геометриялық көрінісі.

(2.1) модель құрылымын таңдау бастапқы кезеңде, тәжірибелерді жүргізуге дейін таңдалынады. Бұл кезеңде *факторларды априорлық рангтеу* орындалады - эксперттік бағалауда негізделген маңызды факторларды таңдау әдісі.

Тәжірибелерді жоспарлау теориясында активті тәжірибелермен байланысты мәселелер қарастырылады.

Тәжірибелер жоспары дегеніміз - сынақтар санын, шарттары мен тәртібін анықтайтын мәліметтер жиынтығы.

Тәжірибелерді жоспарлау - берілген талаптарды қанағаттандыратын тәжірибе жоспарын таңдау. Яғни, бұл зерттелетін құбылыстың механизмін толық білмеу жағдайында жүзеге асырылатын мақсаттық басқару.

Әрекет аймағы - тәжірибе кезінде $\bar{x}$ факторларының мүмкін мәндерінің аймағы.

Жоспарлау аймағы - қолданылатын тәжірибе жоспарының жүргізу шарттарына сәйкес келетін нүктелер орналасқан $\bar{x}$ мәндерінің аймығы (бұл әрекет аймағының бөлігі немесе оған сәйкес келеді).

2.2 Факторлардың деңгейлері мен өзгеру аралықтарын таңдау. Тәжірибені қайтадан өндіру туралы түсінік

Тәжірибені жоспарлау әр кіріс факторының негізгі деңгейін таңдаудан басталады.

Негізгі деңгей - қазіргі уақытта үлкен қызығушылық тудыратын деңгей - $\bar{x}_0$ векторы. Жоспардың барлық нүктелері оның маңында орналасқан. $\bar{x}_0$ векторы жоспарлау аймағының ортасы (тәжірибелер ортас) болып табылады.

Тәжірибе жоспарын құру деген – негізгі (кей кезде нөлдік деп аталатын) деңгейге қатысты нүктелерді (кіріс факторларының деңгейлерін) таңдау. Кіріс факторларының басқа деңгейлерін анықтау үшін әр кіріс факторының *вариациалану аралығы* енгізіледі. Кіріс факторының жоғарғы деңгейін анықтау үшін өзгеру аралығын берілген фактордың нөлдік деңгейіне қосу керек, ал төменгі деңгейді анықтау үшін өзгеру аралығын нөлдік деңгейден алып тастау керек. Өзгеру интервалына төменнен және жоғарыдан шектеулер қойылады.

Кіріс факторының өзгеру аралығына келесі талаптар қойылады:

- бұл фактор өлшенетін қатеден кем болмауы керек, әйтпесе фактор деңгейлері ажыратылмайтын болады;

- ол тым үлкен болмауы керек, яғни төменгі және жоғарғы деңгейлер факторды анықтау аймағынан және тәжірибе аймағынан шықпауы керек.

Факторлардың деңгейлер саны әртүрлі болуы мүмкін. Факторлардың өзгеру диапазоны *y*-тың анықтау аймағын береді. Факторлар деңгейлерінің өзгеру диапазоны тәжірибенің нақты жағдайларына сүйене отырып анықталады.

Егер әр факторға координаталық ось сәйкес келеді деп қабылданса, онда алынған кеңістік факторлық кеңістік болады, ол *n=2* кезінде тіктөртбұрыш, *n=3* кезінде – куб, *n >3* кезінде - гиперкуб.

Жоспарлау аймағын факторлардың мүмкін болатын өзгеру *интервалдары* $x_{imin} \leq x_i \leq x_{imax}$,*i=1,…,n* арқылы анықтауға болады. Жоспарлау аймағының сипаттамалық нүктелері - x_{imin}, x_{imax}, негізгі деңгей - x_{i0}.

x_i фактордың өзгеру (варияциялау) аралығын (қадамын) Δx_i белгілейміз.

Сонда $x_{imax} = x_{i0} + \Delta x_i$; $x_{imin} = x_{i0} - \Delta x_i$;

Әдетте

$$x_{i0} = \frac{x_{imin} + x_{imax}}{2}.$$

x_i фактордың өзгеру аралығы (қадамы):

$$\Delta x_{i0} = \frac{x_{imax} - x_{imin}}{2}.$$

Фактордың өзгеру ауқымы - осы жоспардағы фактордың максималды және минималды мәндерінің арасындағы айырмашылық. Сонда фактордың өзгеру аралығы фактордың өзгеру ауқымының жартысына тең болады.

Әдетте, тәжірибені бастапқы жоспарлау кезінде барлық кіріс факторлары үшін деңгейлер саны бірдей таңдалады. Сонда тәжірибедегі сынақтар санын (N_{exp}) келесі формула бойынша анықтауға болады:

$$N_{exp} = p_{exp}^{k},$$

мұндағы p_{exp} – әр кіріс фактордың деңгейлер саны;

k – тәжірибеде зерттелетін кіріс факторлар саны.

Егер белгілі бір фактордың бақыланатын айнымалыға әсері басқа фактордың мәні өзгерген кезде өзгеретін болса, онда факторлар арасында *өзара әрекеттесу* бар деп айтылады. Факторлардың *өзара әрекеттесу әсері* - бір фактор әсерінің өзгеруінің басқа факторлар деңгейіне тәуелділігінің көрсетеді.

Талдаудың мағынасы кездейсоқ шаманың жалпы вариациясын тәуелсіз қосындыларына - *эффектілерге* жіктеу болып табылады; әр эффект белгілі бір фактордың (негізгі эффект) немесе олардың өзара әрекеттесуінің (*өзара әрекеттесу эффектісі*) әсерін сипаттайды. Бөлек фактордың әсерін сипаттайтын мүше *негізгі эффект* деп аталады ал факторлардың өзара әрекеттесуін сипаттайтын эффект - *өзара әрекеттесу* әсері.

Факторлар - өлшемді шамалар. Өлшемдер әртүрлі болуы мүмкін және олардың мәндерін білдіретін сандар әртүрлі дәрежеге ие болуы мүмкін. Сондықтан тәжірибе жоспарының матрицасын құруды және нәтижелерді өңдеуді жеңілдету үшін факторлық кеңістікті алдын-ала *нормаландыру* қажет, ол үшін сызықтық түрлендіру қолданылады. Мұндай түрлендіру координаттардың жаңа ортасын таңдаудан (орталықтандыру) және координаттық осьтер бойынша масштабтың өзгеруінен (масштабтау) тұрады. Осы мақсатта факторлар координаттарының басын түрлендіріп, нормаланған, (стандартты) масштабқа көшеді:

$$x_i = \frac{x_i - \tilde{x}_{i0}}{\Delta\tilde{x}_i}, i = 1,2,\ldots,n.$$

мұндағы x_i– нормаланған мән;

$\tilde{x}_i$ – бастапқы мән;

$\tilde{x}_{i0}$ – негізгі деңгей;

$\Delta\tilde{x}_i$ – вариациялау интервал.

Мұнда координаттар басы - тәжірибе ортасы, өлшем бірлігі - факторлардың вариациялау қадамы. Мұндай түрлендіру *өлшемсіз* шамаларға әкеледі және математикалық есептеулерді жеңілдетеді.

Тәжірибені жоспарлау теориясының негізгі мақсаты - тәжірибеде зерттелетін процестің математикалық моделін құру. Тәжірибелерді жоспарлау теориясының әдістерін қолдану үшін алдымен тәжірибе бірдей қатемен қайтадан өндірілетініне көз жеткізу керек.

Қайтадан өндірілетін (бірдей қатемен) тәжірибе деп келесідей тәжірибелерді түсінеміз: кез келген уақытта зерттеу объектісі мен өлшеу жабдықтарын бастапқы күйіне қайтаруға болатын және тәжірибені кездейсоқ қателіктердің бірдей деңгейінде қайталауға болатын тәжірибе.

Кездейсоқ қателіктердің әсерін өтеу үшін әр тәжірибені *n* рет қайталау ұсынылады. Факторлардың бірдей мәндерінде бірнеше рет қайталанған тәжірибелер *параллель* тәжірибелер деп аталады. *Дубльдеу* (қайталау) деп параллельді тәжірибелерді орындауын айтады. Әдетте *n* параллель тәжірибелер саны 2-3, кейде 4-5 болады. Зерттеулер жүргізу кезінде тәжірибелерді дубльдеудің үш нұсқасын қолдануға болады: 1. тәжірибелер біркелкі қайталау; 2. тәжірибелер әркелкі қайталанады; 3. тәжірибелер қайталанбайды. *Біркелкі қайталану* кезінде жоспарлау матрицасының барлық жолдарында параллель тәжірибелердің бірдей саны болады. Параллель тәжірибелердің саны бірдей болмаса - *біркелкі емес қайталану*. Қайталану болмаған кезде параллель тәжірибелер жүргізілмейді. Қайталанудың үш нұсқасының ішіндегі ең қолайлысы - біріншісі. Бұл нұсқада тәжірибе жоғары дәлдікпен сипатталады, ал тәжірибелік мәліметтерді математикалық өңдеу қарапайым болады. Қайталану сипаты бақылау нәтижелерін математикалық өңдеудің мазмұнына әсер етеді.

Тәжірибенің бірдей қатемен қайтадан өндірілетінін тексеру үшін параллель тәжірибелердің бірнеше сериясын өткізеді.

Егер тәжірибелер бірдей қатемен қайтадан өндірілмейтін болса, онда тәжірибенің тұрақсыздық көздерін анықтау және жою, сондай-ақ дәл өлшеу әдістері мен құралдарын қолдану арқылы қайтадан өндірілетінге қол жеткізуге тырысуға болады. Егер қайтадан өндірілуді ешқандай жолмен қол жеткізу мүмкін болмаса, онда мұндай тәжіибелерге жоспарлаудың математикалық әдістерін қолдануға болмайды.

Тәжірибелердің бірдей жағдайларында параллель тәжірибелердің саны неғұрлым көп болса, қайтадан өндірілудің алынған дисперсиясының бағалары соғұрлым сенімді болады. Алайда, параллельді тәжірибелер санының артуымен жоспардың сынақтар саны артады. Сондықтан, егер қайтадан өндірілу дисперсиясын бағалаудың қажетті сенімділігі қамтамасыз етілсе, онда параллель тәжірибелердің ең аз саны таңдалады, ол әдетте 2 немесе 3-ке тең.

Тәжірибелер жоспарын жоспар матрицасы немесе қайталанатын матрицамен бірге жоспар спектрінің матрицасы түрінде орнатуға болады.

Жоспар матрицасы - тәжірибе шарттарын матрица түрінде жазудың стандартты түрі; матрицаның *(i,j)* –ші элементі – *i* -ші тәжірибедегі *j*-ші фактордың деңгейі. Матрицаның бірдей жолдары болуы мүмкін.

Жоспар спектрінің матрицасы - жоспар матрицасының кем дегенде бір фактордың деңгейімен ерекшеленетін барлық жолдарынан тұратын матрица. Бұл матрицаның барлық жолдары әртүрлі.

Қайталау (дубльдеу) матрицасы - диагоналды элементтері жоспар спектрінің тиісті нүктелеріндегі параллель тәжірибелер санына тең квадратты матрица.

Қайталау (дубльдеу) дегеніміз - бір сынақта өлшеу сериясы емес, әр тәжірибенің толық қайталануы.

Сонымен, *тәжірибе жоспары* –$\bar{x}_1, \bar{x}_{2,\dots,}\bar{x}_N$ нүктелердің жиыны, мұндағы $\bar{x}$– кіріс векторы, N – бақылау саны, $m_1, m_2, \dots, m_N$ – әр нүктедегі бақылау саны, яғни келесідей спектрдің матрицасы:

$$X = \begin{pmatrix} x_{11}x_{12}\dots x_{1n} \\ \dots\dots\dots\dots\dots\dots \\ x_{N1}x_{N2}\dots x_{Nn} \end{pmatrix},$$

тәжірибелерді қайталау матрицасымен бірге:

$$\begin{pmatrix} m_1 0\dots\dots\dots 0 \\ \dots\dots\dots\dots\dots\dots \\ 0\dots\dots\dots\dots m_N \end{pmatrix},$$

сонымен қатар

$$\sum_{g=1}^{N} m_g = N.$$

Тәжірибелердің жүйелік қателіктерін белгілі бір дәрежеде өтеу үшін *рандомизация* деп аталатын әдіс қолданылады. Бұның мағынасы: кездейсоқ сандар кестесін қолдана отырып, тәжірибелерді кездейсоқ ретімен жүргізу.

Параллель тәжірибелер - барлық факторлардың деңгейлері өзгеріссіз сақталатын уақыт бойынша рандомизацияланған тәжірибелер.

Мысалы, I, II,..., VI сандармен белгіленген 6 тәжірибені уақыт бойынша рандомизациялау қажет болсын. Оларға кездейсоқ сандар кестесінің кез келген жолында немесе кез келген бағанында алынған кез келген 6 тізбектелген сандарды сәйкестіреміз. Егер қайталанатын сандар кездессе, онда оларды алып тастау керек. Мысалы, келесі жұптарды алуға болады:
I–60; II–12; III–05; IV–15; V–34; VI–30. Кездейсоқ сандарды өсу (немесе кему) ретімен орналастыра отырып, біз сынақтарды орындау ретін аламыз: III, II, IV, VI, V, I (немесе I, V, VI, IV, II, III).

Тәжірибелер мен оларды орындаушылар бір-бірінен өзгеше болуы мүмкін, бірақ практикада барлық тәжірибелерді жоспарлау, жүргізу және талдау бірдей ретімен орындалады. Зерттеу объектілер әртүрлі болғанымен, тәжірибелік зерттеу әдістері көптеген ұқсастықтарға ие:

- тәжірибе қаншалықты қарапайым болса да, алдымен оны жүргізу жоспары жасалады;
- айнымалылар санын азайтуға тырысады, өйткені бұл жұмысты жеңілдетеді және оны үнемді етеді;
- сыртқы кездейсоқ факторлардың әсерін болдырмауға тырысады;
- тәжірибенің барысын бақылау қажет;
- өлшеу құралдарының дәлдігіне және деректерді алу дәлдігіне назар аудару керек;
- кез келген тәжірибе барысында алынған нәтижелерді талдап, оларды түсіндіру қажет, өйткені осы шешуші кезеңсіз бүкіл процестің мағынасы жоқ.

Тәжірибелерді жоспарлаудың екі негізгі мақсаты бар:

- нәтижелердің сенімділігі мен дәлдігіне қойылатын талаптарды сақтау кезінде сынақтардың жалпы көлемін қысқарту;
- әр бөлек тәжірибенің ақпараттылығын арттыру.

2.3 Бақылау сұрақтары

2.3.1 Сандық және сапалық тәжірибелердің айырмашылығы неде?

2.3.2 Активті және пассивті тәжірибелер арасындағы айырмашылықтар қандай?

2.3.3 Факторлық кеңістік дегеніміз не?

2.3.4 Басқарылатын факторларға қойылатын талаптарды атаңыз.

2.3.5 Жауап беру функциясы дегеніміз не? Неліктен бұл функция кездейсоқ?

2.3.6 Фактордың деңгейін анықтаңыз.

2.3.7 Фактордың өзгеру аралығына қандай талаптар қойылады?

2.3.8 Жоспар матрицасы, жоспар спектрінің матрицасы және қайталау матрицасы анықтамаларын беріңіз.

2.3.9 Параллельді тәжірибелер дегеніміз не?

2.3.10 Қайтадан өндірілетін тәжірибе анықтамасын беріңіз.

3. Статистикалық бағалау

3.1 Кездейсоқ шамалардың статистикалық бағаларының қасиеттері

Тәжірибелік мәліметтерді өңдеумен байланысты көптеген қолданбалы есептерді шешкенде зерттеуші белгілі ықтималдық сипаты бар нәтижелерді өңдеуі керек. Алайда, тиісті кездейсоқ шамалардың қажетті сипаттамалары көбінесе белгісіз және тәжірибелік мәліметтер негізінде табылады. Бақылау нәтижелерін статистикалық сипаттау, ықтималдық ұғымын қолданатын әртүрлі математикалық модельдерді құру және тексеру математикалық статистиканың негізгі мазмұнын құрайды. Тәжірибелік мәліметтерді өңдеу әдістері математикалық статистиканың негізгі ұғымдарына негізделген. Оларға *бас жиын, таңдама* және *эмпирикалық таралу функция* ұғымдары жатады.

Бас жиын – берілген жағдайларда жүргізілуі мүмкін, кездейсоқ шаманы бақылаудың барлық мүмкін болатын нәтижелерінің жиынтығы. Яғни, *бас жиын* дегеніміз -объектіні шектеусіз уақыт бойынша бақылау кезінде тіркеуге болатын параметрдің барлық мүмкін мәндері. Мұндай жиынтық элементтердің *шексіз жиынтығынан тұрады*. Объектіні бақылау нәтижесінде параметр мәндерінің шектеулі көлемі бар $x_1, x_2, \dots, x_n$жиынтығы қалыптасады.

Бақылау кезінде алынған кездейсоқ шаманың шектелген жиынтығын *таңдама* деп аталады. Таңдама элементтерінің саны оның *көлемі* болып табылады.

Бақылау нәтижелері бойынша қандай да бір қорытынды жасау үшін таңдама *репрезентантты* болуы, яғни бас жиынның пропорцияларын дұрыс көрсету керек. Егер таңдаманың көлемі жеткілікті үлкен болса және бас жиынның әр элементінің таңдамаға ену ықтималдығы бірдей болса бұл талап орындалады.

Егер таңдама бас жиынның ерекшеліктері туралы жеткілікті түсінік берсе, ол *репрезентантты* деп аталады.

Егер бас жиын туралы ешқандай ақпарат болмаса, онда репрезентанттықтың жалғыз кепілі бола алады - *рандомизация* деп аталаты кездейсоқ таңдау. Осы таңдау арқылы бекітілген көлемнің әр мүмкін болатын таңдамасы бірдей таңдау ықтималдығына ие, ал тізбектелген бақылаулар тәуелсіз.

Статистикалық әдістердің міндеті: шектеулі N көлемі бар таңдама негізінде, яғни бас жиынның белгілі бір бөлігі бойынша, оның қасиеттері туралы негізделген пікір ұсыну болып табылады. Мұндай пікір қандай да бір қолайлы *бақылаулардан тәуелді функциялардың* көмегімен бас жиынның таңдамалы параметрлерін (сипаттамаларын) *бағалау* арқылы қалыптастырылуы мүмкін.

Қарастырылып отырған бас жиыннан алынған таңдама элементтерінен тәуелді $G = g(x_1, x_2, \dots, x_n)$ функциясы *статистика* деп аталады. Бас жиынның белгілі бір параметрін бағалауға жарамды анықталған статистика осы параметрдің *статистикалық бағасы* (таңдамалы мәні) деп аталады.

Еркін статистиканың өзі кездейсоқ шама болады және оны кез келген кездейсоқ шама сияқты ықтималдық тығыздығы немесе сандық сипаттамалары арқылы сипаттауға болады. Жалпы жағдайда, бұл бас жиынның параметрлеріне және таңдаманың N көлеміне тәуелді. Бас жиынның параметрінің кез келген бағасы кездейсоқ шама болып табылады. Параметрдің қасиеттері (ықтималдық) таңдамалы таралу функциясы мен оның сипаттамалары арқылы да анықталуы мүмкін.

Бас жиынның параметрлері мен олардың бағалары үшін әртүрлі белгілер енгізіледі: θ параметрдің бағасы $\hat{\theta}$ деп белгіленеді. Егер бас жиынның параметрлері грек әріптерімен көрсетілсе (мысалы, дисперсия σ_x^2, корреляция коэффициенті ρ және т.б.), тиісті таңдама мәндері латын әріптерімен көрсетіледі (S_x^2, r және т. б.). Математикалық күтімнің m_x бағасы үшін арнайы белгілеу $\bar{x}$ қабылданады.

Тәжірибелік мәліметтердің шектеулі көлемі бойынша есептелген параметрдің мәні кездейсоқ шама болып табылады, яғни мұндай шаманың бір таңдамадан басқа таңдамаға дейінгі мәні алдын ала болжанбаған түрде өзгеруі мүмкін. Сондықтан, тәжірибелік мәліметтерді өңдеу нәтижесінде θ параметрдің мәні анықталмайды, тек қана оның $\hat{\theta}$ жуықтаған мәні - параметрдің статистикалық *бағасы* анықталады.

Теориялық таралудың параметрінің статистикалық бағасын алу дегеніміз қажетті параметрдің жуықталған мәнін беретін бақылаудың қол жетімді нәтижелеріне тәуелді функцияны табу.

Бағалаудың екі түрі бар – *нүктелік* және *аралық*.

Бір санмен сипатталатын бағалар *нүктелік* деп аталады. Таңдаманың кіші көлемінде нүктелік бағалалар параметрлердің нақты мәндерінен айтарлықтай ерекшеленуі мүмкін, сондықтан олар таңдаманың үлкен көлемінде қолданылады.

Аралық бағалар мүмкін болатын параметр мәнінің ықтимал диапазонын анықтайтын екі санмен беріледі. Бұл бағалау кішігірім және үлкен таңдамалар үшін қолданылады.

Алдымен нүктелік бағалауды қарастырамыз.

Бір параметрді бағалау үшін әртүрлі бағаларды (статистикаларды) қолдануға болады. Олардың ең жақсысын таңдау үшін тәжірибе тұрғысынан қажет бағалау қасиеттеріне қойылатын кейбір талаптарды тұжырымдау қажет.

Бағалардың сапасы келесі қасиеттермен сипатталады: *орнықтылық, ығыспаушылық, тиімділік және жеткілікті.*

Егер де таңдама көлемі N шексіз өсетін жағдайда параметрдің $\hat{\theta}$ бағасының мәні ықтималдық бойынша өзінің теориялық мәніне ұмтылатын болса, яғни

$$N \to \infty \text{ болғанда}, \hat{\theta} \to \theta.$$

онда параметр бағасы *орнықты* деп аталады.

Сонда N мәнінің өсуімен таңдамалы таралу мәні θ-ның аймағында жинақталады және баға дәлдігі шексіз өседі.

Орнықты баға үшін келесі шарт орындалады:

$$\lim_{N\to\infty} lim\, D\,\{\hat{\theta}(X_1, X_2, \ldots, X_N)\} \to 0.$$

Егер де таңдаманың кез келген көлемі N үшін θ параметрдің $\hat{\theta}$ бағасы бағаланатын параметрдің өзіне тең болса:

$$m_{\hat{\theta}} = M\{\hat{\theta}\} = \theta,$$

онда параметр бағасы $\hat{\theta}$ *ығыспаған* деп аталады.

Ығыспаушылық талабын қанағаттандыру N таңдама көлеміне байланысты параметрді бағалаудың жүйелік қателігін жояды және орнықты жағдайда $N\to\infty$ болған кезде нөлге ұмтылады. Таңдаманың кез келген шекті, соның ішінде шағын көлемінде бағалардың параметрлердің өзінен жүйелік ауытқуларының болмауын ығыспаушылық сипаттайды. Математикалық күтімі бағаланатын параметрге тең болмайтын статистикалық бағаларды қолдану жүйелік қателіктерге әкеледі. Бірақ ығыспаған бағада тіпті шынайы мәннен алыс болуы мүмкін.

Параметрдің барлық ығыспаған бағаларының ішінде, дисперсиясы минималды болатын, яғни

$$D\{\hat{\theta}_{эф}\} = \sigma^2_{\hat{\theta}_{эф}} = M\{(\hat{\theta} - \theta)^2\} = min,$$

$\hat{\theta}_{эф}$ бағасы *тиімді (эффективті)* деп аталады.

Тиімділік бағалардың кездейсоқ мәндерінің параметрдің нақты мәніне жақын аймақта таралуын сипаттайды. Барлық бағалар ішінен мәндері бағаланатын параметрдің жанында шоғырланғанын таңдау керек.

Жоғарыда айтылғандардың бәрі тек қана қалыпты тарату заңына бағынатын кездейсоқ қателікті қамтитын өлшеулерге қатысты.

Параметрлерді нүктелік бағалау есебінің мәні. Нүктелік бағалау параметрдің мәні ретінде қабылданатын жалғыз сандық шаманы табуды

қамтиды. Мұндай бағалауды тәжірибелік мәліметтердің көлемі жеткілікті үлкен болған жағдайларда анықтаған жөн. Сонымен қатар, тәжірибелік мәліметтердің жеткілікті көлемі туралы бірыңғай түсінік жоқ, оның мәні бағаланатын параметрдің түріне байланысты (біз осы мәселеге параметрлерді аралық бағалау әдістерін қарастырғанда ораламыз, әзірше кемінде 10 мәні бар таңдаманы жеткілікті деп қарастырамыз). Тәжірибелік мәліметтердің аз көлемінде нүктелік бағалау параметрлердің шынайы мәндерінен айтарлықтай ерекшеленуі мүмкін, бұл оларды қолдануға жарамсыз етеді.

Есептің қойылымның типтік нұсқасындағы параметрлерді нүктелік бағалау міндеті келесідей: кездейсоқ X мәні үшін бақылаулардың $(x_1, x_2, \dots, x_n)$ таңдамасы алынған. Таңдама репрезентантты болуы керек.Таңдама көлемі N белгілі. X шамасының таралу заңы белгілі, мысалы, $f(x,\theta)$ таралу тығыздығы түрінде, мұндағы θ – таралудың белгісіз (жалпы жағдайда векторлық) параметрі. Параметр кездейсоқ емес шама. Таралу заңының θ параметрінің $\hat{\theta}$ бағасын табу қажет.

Параметрлерді нүктелік бағалау мәселесін шешудің бірнеше әдістері бар, олардың ішінде ең көп тарағандары: *максималды (ең үлкен) ықтималдық, моменттер* және *квантильдер* әдістері болып табылады.

3. 2 Максималды ықтималдық әдісі

Әртүрлі θ параметрлер бағаларының қасиеттері көбінесе зерттелетін бас жиынның таралу заңының түрімен, сондай-ақ, бағалау әдісімен анықталады. Көптеген жағдайларда белгісіз параметрлердің ең жақсы бағаларын тәжірибелік мәліметтері бойынша табу мәселесін шешуге мүмкіндік беретін жалпы әдістер белгілі. Тәжірибелік мәліметтер бойынша белгісіз параметрлерді бағалаудың маңызды жалпы әдісі - *максималды ықтималдық әдісі* болып табылады.

Бұл әдісті 1912 жылы Р.Фишер ұсынған. Әдіс X шаманың кейбір сандық көрсеткішін $(x_1, x_2, \dots x_N)$ бақылаулардың таңдамасын алу ықтималдығын зерттеуде негізделген. Таңдама мәндері кездейсоқ, сондықтан олар бірдей таралуы бар тәуелсіз кездейсоқ шамалар ретінде қарастырылады. Кездейсоқ шаманың тәжірибеде ықтималдығы максималды болатын мәндерін іске асыру қасиеті максималды ықтималдық әдісінің мәні болып табылады. Бұл жағдайда N кездейсоқ шамалардың ықтималдықтарының бірлескен таралу тығыздығы максималды болуы керек.

Тұрақты белгісіз θ параметріне тәуелді ықтималдық тығыздығы $f_x(x;\theta)$ болатын бас жиыннан $x_1, x_2, \dots, x_N$ таңдама алынған болсын. Таңдама элементтері статистикалық тәуелсіз, сондықтан кездейсоқ N көлемі бар таңдама үшін таңдамалы ықтималдық тығыздылығын жаза аламыз:

$$f_{x_1,x_2,\dots,x_N}(x_1, x_2, \dots, x_N; \theta) = f_{x_1}(x_1;\theta) f_{x_2}(x_2;\theta) \cdot \ldots \cdot f_{x_N}(x_N;\theta). \qquad (3.1)$$

Онда $x_1, x_2, \ldots, x_N$ таңдама мәндерін (3.1) ықтималдық тығыздығына қойғаннан кейін алынған ықтималдық тығыздығы θ параметрі үшін $L(\theta)$ *шындыққа ұқсас функциясы* (*функция правдоподобия*, келешекте біз бұл функцияны *максималды ықтималдық функциясы* деп атаймыз) деп аталады:

$$L(\theta) = f_{x_1}(x_1;\theta) f_{x_2}(x_2;\theta) \cdot \ldots \cdot f_{x_N}(x_N;\theta).$$

Тәжірибелерге дейін шындыққа ұқсас функциясы белгілі таңдамада θ парметрінің әртүрлі мәндерінің басым болуын білдіреді. Тәжірибе жүргізілгеннен кейін таңдама мәндері белгілі болады – бұл кейбір сандар, ал θ параметрі белгісіз.

Максималды ықтималдық әдісінің мағнасы: белгісіз θ параметрін $\hat{\theta}$ бағасы ретінде оның шындыққа ұқсас функцияны максимумға жеткізетін мәні таңдалады. Көптеген жағдайларда бұл функцияның максимумы келесі теңдеуден табылады:

$$\frac{dL(\theta)}{d\theta} = 0.$$

Кейде $L(\theta)$ функцияның максимумын емес, оның логарифмінің $l(\theta) = ln\, L\,(\theta)$ максимумын табу ыңғайлы. Сонда келесі теңдеуді аламыз:

$$\frac{dl(\theta)}{d\theta} = \frac{d}{d\theta}[\sum_{i=1}^{N} lnf_{x_i}(\,x_i;\theta)] = 0.$$

Соңғы екі теңдеу *ықтималдық теңдеулері* деп аталады. Бағалауды анықтау үшін ықтималдық функцияның максимумын табу процедурасы қолданылады, өйткені бұл функция белгілі бір таңдамада әртүрлі θ мәндерінің басым болуын көрсетеді. Мұндай баға басымды мәндерді береді, өйткені таңдаманы алу ықтималдығы максималды болады.

Белгісіз параметрлердің $\theta_1, \theta_2, \ldots, \theta_n$_саны n болса, олардың бағалары $L(\theta_1, \theta_2, \ldots, \theta_n)$ максималды ықтималдық функциясының немесе оның логарифмін максимумын барлық n параметрлері бойынша бір уақытта анықтау жолымен табылады. Сәйкесінші бағалар соңғы екі теңдеулер сияқты n теңдеулер жүйесін шешу жолымен алынады (туындылар дербес болады).

Максималды ықтималдық әдісінің басты артықшылығы - бұл жақсы қасиеттері бар бағаларды оңай табуға мүмкіндік береді:

- егер тиімді баға болатын болса, максималды ықтималдық әдісі дәл осындай бағалана береді және одан дәлірек бағаны табу мүмкін емес;

- өте әлсіз шектеулермен максималды ықтималдылық әдісін бағалау бай, сонымен қатар асимптотикалық қалыпты.

Сонымен, максималды ықтималдық бағаларын табу келесі кезеңдерден тұрады: ықтималдық функцияны құру (оның натуралды логарифмін); осы

функцияны ізделінетін параметрлері бойынша дифференциалдау және бағаларды табу үшін теңдеулер жүйесін анықтау; алынған теңдеулер жүйесін шешу.

Мысал 1. Аспаптың істен шығуына дейінгі уақыт көрстекіш заңымен анықталады:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0; \\ 0, \quad x < 0. \end{cases}$$

2,3,5,6,8 элементтері бар таңдама бойынша λ параметрінің бағасын анықтау керек.

Шешім: максималды ықтималдық функцияның түрі:

$$L(2,3,5,6,8,\lambda) = \lambda e^{-\lambda 2} \cdot \lambda e^{-\lambda 3} \cdot \lambda e^{-\lambda 5} \cdot \lambda e^{-\lambda 6} \cdot \lambda e^{-\lambda 8} = \lambda^5 e^{-\lambda 24} \cdot$$

Онда

$$lnL(2.3.5.6.8) \ = \ 5\, ln\, \lambda - 24\, \lambda.$$

Ықтималдық теңдеуі келесі түрде жазылады

$$\frac{\partial lnL}{\partial \lambda} = \frac{5}{\lambda} - 24 = 0.$$

Осыдан параметр бағасын аламыз:

$$\hat{\lambda} = \frac{5}{24}.$$

Максималды ықтималдық әдісі орнықты, тиімді (егер бар болса, алынған шешім тиімді баға береді), асимптотикалық қалыпты таралған бағаларды алуға мүмкіндік береді. Бұл әдіс ығысқан және ығыспаған бағаларды бере алады. Ығысуды жою үшін түзетулер енгізіледі. Бұл әдіс әсіресе таңдамалар көлемі кіші болғанда пайдалы. Егер максималды ықтималдық функциясы бірнеше максимумға ие болса, онда олар арасында глобалдысы таңдалады.

3.3 Моменттер әдісі

Бұл әдісті 1894 жылы К.Пирсон ұсынған. Зерттелетін сандық белгінің таралу функциясының түрі белгілі деп санаймыз, бірақ бұл таралудың параметрлері белгісіз. Бұл параметрлерді таңдама арқылы бағалау керек.

Моменттер әдісінің мәні келесіде: таралуы $\theta_1, \ldots, \theta_m$ параметрлер дәлдігіне дейін белгілі бас жиынның белгілі $x_1, x_2, \ldots, x_N$ таңдамасы бойынша, оның таралуының бағаланатын белгісіз параметрлерінің санына сәйкес эмпирикалық моменттер таңдалады. Әдетте, төменгі ретті моменттерді қолданған жөн, өйткені моменттер реті жоғарылан сайын бағалар қателіктері

өседі. Моменттердің бағалары есептелген соң олар теориялық моменттерге теңестіріледі; таралу параметрлері моменттер арқылы анықталады және параметрлердің моменттерге тәуелділігін білдіретін теңдеулер құрылады, нәтижесінде теңдеулер жүйесі пайда болады. Бұл жүйенің шешімі бас жиынның таралу параметрлерінің бағаларын береді.

Мысал 2. Аспаптың істен шығуына дейінгі уақыт көрстекіш заңымен анықталады:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0; \\ 0, \qquad x < 0. \end{cases}$$

Таңдамалы мәндері 2, 3, 5, 6, 8 бойынша белгісіз λ параметрінің нүктелік бағасын табамыз. Көрсеткіш таралу үшін тек қана жалғыз λ параметрді бағалау керек, сондықтан оны бағалауға жалғыз теңдеу жеткілікті. Мысалы, бірінші бастапқы моменттерін – теориялық және таңдамалы теңестіреміз.

Бірінші бастапқы теориялық момент тең болады:

$$m_1 = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \lambda.$$

Бірінші бастапқы таңдамалы момент келесі формуламен есептеледі:

$$\hat{m}_x = \sum_{i=1}^{N} \frac{x_i}{N}.$$

Осыларды теңестіріп, аламыз:

$$\frac{1}{\lambda} = \sum_{i=1}^{N} \frac{x_i}{N}.$$

Осыдан:

$$\vec{\lambda} = \frac{1}{\sum_{i=1}^{N} \frac{x_i}{N}}.$$

3.4 Қалыпты таралудың параметрлерін нүктелік бағалау

Кездейсоқ шамалардың сандық сипаттамалары математикалық статистикада маңызды рөл атқарады. Бірақ кездейсоқ шама ықтималдыықтардың таралу заңын қолдану арқылы толығымен анықталады.

Көбінесе әртүрлі кездейсоқ шамалардың қасиеттерін сипаттау үшін қалыпты таралу заңы қолданылады, оның өрнегі:

$$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{\frac{(x-m_x)^2}{2\sigma_x^2}},$$

мұндағы m_x- кездейсоқ шаманың математикалық күтімі;

σ_x^2 – кездейсоқ шаманың дисперсиясы.

Қалыпты таралудың рөлінің теориялық негіздемесі *орталық шекті теорема* болып табылады. Кездейсоқ шаманы тәуелсіз кездейсоқ әсерлердің үлкен санының қосындысы ретінде қарастыруға негіз болған кезде, олардың әрқайсысының әсері шамалы болса да, әсердің құрамдас бөліктерінің таралуы ерікті болса да, орталық шекті теоремаға сәйкес зерттелетін кездейсоқ шаманың қалыпты заңына сәйкес таралуын күтуге болады. Көптеген есептерде қалыпты таралудың артықшылығы - ол ыңғайлы математикалық қасиеттермен сипатталады. Сондықтан математикалық статистика әдістерінің көпшілігі зерттелетін шама қалыпты таралуғы бағынады деген болжамға негізделген, дегенмен іс жүзінде бұл болжам әрқашан арнайы тексеруді қажет етеді.

Әрине, кез келген кездейсоқ шама қалыпты заңға сәйкес таралады деуге болмайды. Қалыпты таралу - бұл табиғатта кездесетін көптеген құбылыстарды жақсы сипаттайтын және үлкен практикалық қолданысы бар таралудың түрі.

Орнықтылық, ығыспаушылық, тиімділік ұғымдары әдетте нүктелік бағалар үшін қолданылады. Бірқатар маңызды сипаттамалар үшін бағалар мен олардың қасиеттерін қарастырайық.

Қалыпты таралу заңымен таралған белгілі σ_x^2 дисперсиясы бар X кездейсоқ шама берілген және оның дисперсиясы белгілі болсын.

1. N бақылаулар бойынша кездейсоқ шаманың m_x математикалық күтімін бағалау қажет.

Ықтималдық функциясы келесі түрге ие болады:

$$L(m_x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x_1-m_x)^2}{2\sigma_x^2}} \cdot \ldots \cdot \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x_N-m_x)^2}{2\sigma_x^2}} = \frac{1}{(\sqrt{2\pi}\sigma_x)^N} e^{-\frac{\sum_{i=1}^{N}(x_i-m_x)^2}{2\sigma_x^2}}.$$

Логарифмдік ықтималдық функция тең болады:

$$l(m_x) = -\frac{N}{2} ln\, 2\, \sigma_x - N\, ln\, \sigma_x - \frac{1}{2\sigma_x^2}\sum_{i=1}^{N}(x_i - m_x)^2 = 0.$$

Бұл функцияның максимумы келесі теңдеуден табылады:

$$\frac{dl(m_x)}{dm_x} = -\frac{2}{2\sigma_x^2}\sum_{i=1}^{N}(x_i - m_x) = 0.$$

Осыдан математикалық күтімнің бағасын табамыз:

$$\widehat{m}_x = \bar{x} = \sum_{i=1}^{N} \frac{x_i}{N}.$$

Бұл функцияның белгілі N-де таңдамадан таңдамаға өзгергіштігін сипаттау үшін $\bar{X}$ кездейсоқ шамны қарастыамыз:

$$\bar{X} = \frac{1}{N}(X_1 + X_2 + \ldots + X_N) = \frac{1}{N}\sum_{i=1}^{N} X_i.$$

Қалыпты таралған кездейсоқ шамалардың сызықты комбинациясы да қалыпты түрде таралатыны белгілі. Сондықтан келесіні жазуға болады:

$$M\{\bar{X}\} = \frac{1}{N}\sum_{i=1}^{N} M\{X_i\} = \frac{Nm_x}{N} = m_x;$$

$$D\{\bar{X}\} = \frac{1}{N^2}\sum_{i=1}^{N} \sigma_x^2 = \frac{N\sigma_x^2}{N^2} = \frac{\sigma_x^2}{N}.$$

Осыдан келесіні айтуға болады: $\bar{X}$ таңдамалы орта m_x үшін орнықты, ығыспаған бағасы болып табылады. Қалыпты бас жиынның таңдамасы үшін үшін бұл баға тиімді да болады.

Бағалардың қасиеттерін анықтаған кезде зерттелетін бас жиынның таралу заңының түрі маңызды болатынын ескеріңіз. Мысалы, бірқалыпты таралу үшін m_x шамасының тиімді баға ретінде келесі статистика қолданылады:

$$\widehat{m}_x = \frac{x_{max} + x_{min}}{2}.$$

2. Математикалық күтім m_x белгілі болғанда қалыпты жиын үшін оның σ_x^2 дисперсиясының бағасын табамыз.

σ_x^2 үшін логарифмдік ықтималдық функциясын жазамыз:

$$l(\sigma_x^2) = -\frac{N}{2} ln\, 2\,\pi - N\, ln\, \sigma_x^2 - \frac{1}{2\sigma_x^2}\sum_{i=1}^{N}(x_i - m_x)^2 = 0.$$

Сонда ықтималдық теңдеудің түрі келесідей болады:

$$\frac{dl(\sigma_x^2)}{d\sigma_x^2} = -\frac{N}{2\sigma_x^2} + \frac{1}{2\sigma_x^4}\sum_{i=1}^{N}(x_i - m_x)^2 = 0,$$

Осыдан келесіні айтуға болады: *X* кездейсоқ шаманың математикалық күтімі белгілі болғанда, оның σ_x^2 дисперсиясының бағасы келесі формуламен табылады:

$$\hat{\sigma}_x^2 = S_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - m_x)^2. \qquad (3.2)$$

S_x^2 баға σ_x^2 дисперсия үшін орнықты, ығыспағын және қалыпты таралған жиын үшін тиімді бағасы болып табылады.

3. Математикалық күтім m_x белгісіз болғанда дисперсия бағасын табамыз. Яғни екі - m_x и σ_x^2 параметрді бағалау керек. Максималды ықтималдық функция екі параметрге тәуелді:

$$l(m_x, \sigma_x^2) = -\frac{N}{2} ln\, 2\,\pi - N\, ln\, \sigma_x^2 - \frac{1}{2\sigma_x^2}\sum_{i=1}^{N}(x_i - m_x)^2 = 0.$$

Максималды ықтималдық әдісі бойынша бағалар келесі теңдеулер жүйесін шешу жолымен анықталады:

$$\begin{cases} \dfrac{\partial l(m_x, \sigma_x^2)}{\partial m_x} = -\dfrac{1}{2\sigma_x^2}\displaystyle\sum_{i=1}^{N}(x_i - m_x) = 0\,; \\ \dfrac{\partial l(m_x, \sigma_x^2)}{\partial \sigma_x^2} = -\dfrac{N}{2\sigma_x^2} - \dfrac{1}{2\sigma_x^4}\displaystyle\sum_{i=1}^{N}(x_i - m_x) = 0\,. \end{cases}$$

Осыдан келесіні аламыз:

$$\widehat{m}_x = \bar{x} = \sum_{i=1}^{N}\frac{x_i}{N};$$

$$\hat{\sigma}_x^2 = S_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \widehat{m}_x)^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2.$$

Дисперсияның осындай бағасы орнықты болады. Бірақ бұл баға ығысқан, себебі:

$$M\{\hat{\sigma}_x^2\} = \frac{N-1}{N}\sigma_x^2 = \sigma_x^2\left(1 - \frac{1}{N}\right).$$

$N \to \infty$ болғанда $M\{\hat{\sigma}_x^2\} \to \sigma_x^2$ шарты орындалғанмен яғни баға асимптоталық ығыспаған, *N* шамасы кіші болған жағдайда бағаның ығысуы өте үлкен болуы мүмкін. Практикада σ_x^2 үшін келесі ығыспаған бағаны қолданады:

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2. \qquad (3.3)$$

Бұл баға орнықты, қалыпты таралған жиын үшін асмптоталық тиімді.

Сонымен, өлшеулерді жүргізумен байланысты кез келген есепте σ_x^2 мәнін бағалаудың екі әдісі мүмкін.

Бірінші әдісті қолданған кезде құрылғы көрсеткіштерінің тізбегі алынады және алынған нәтижелерді өлшенетін шаманың белгілі немесе калибрленген мәнімен салыстыру арқылы ауытқулар тізбегі табылады. Содан кейін алынған ауытқулар тізбегі (3.2) формуласы бойынша орташа квадраттық ауытқуды есептеу үшін қолданылады.

σ_x^2 мәнін бағалаудың екінші әдісі: $\overline{x}$ арифметикалық орташа мәнін анықтау болып табылады, өйткені бұл жағдайда өлшенетін шаманың нақты мәні белгісіз. Бұл жағдайда дисперсияны бағалауды анықтау үшін басқа (3.3) формуланы қолданған жөн. Бұл жерде *(N-1)*-ге бөлу келесі себептермен байланысты: егер бас жиын толығымен қарастырылмай, таңдама қаралатын болса, *X* массивін орталақтандыру жолымен алынатын ең жақсы баға нақты мәннен белгілі бір шамаға ерекшеленеді. Бұл жағдайда ауытқу квадраттарының қосындысы:

$$\sum_{i=1}^{N}(x_i - \bar{x})^2$$

шынайы орташа мәнді

$$\sum_{i=1}^{N}(x_i - m_x)^2$$

қолданғаннан гөрі аз болады.

N орнына (*N-1*)-ге бөлген кезде, бұл қате ішінара түзетіледі. Математикалық статистика бойынша кейбір нұсқаулықтарда таңдамалы орташа квадраттық ауытқуларды есептеу кезінде әрқашан (*N-1*)-ге бөлуге кеңес беріледі, бірақ кейде бұл жасалмауы керек. Шынайы мән тәуелсіз түрде алынбаған жағдайларда ғана (*N-1*)-ге бөлу керек

Бағалардың қасиеттерін анықтаған кезде зерттелетін бас жиынның таралу заңының түрі маңызды болып табылады. Атап айтқанда, таралудың басқа жиі қолданылатын түрі үшін - бірқалыпты, ол келесідей анықталады:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \textit{егер } a \le x \le b; \\ 0, & \textit{егер } x < a, x > b, \end{cases}$$

ал параметрлердің (математикалық күтім мен дисперсия) тиімді бағалары келесі формулалармен анықталады:

$$m_x = \frac{a+b}{2},$$
$$\sigma_x^2 = \frac{(b-a)^2}{12}.$$

Келесі анықтамаларға тоқталайық.

Таңдама элементтеріне тәуелді және таңдамалы дисперсия формуласына кіретін әрбір мән *байланыс* деп аталады.

Таңдама көлемі мен байланыстар саны арасындағы айырмашылық *еркіндік дәрежелерінің саны* деп аталады. Яғни, статистикалық жағдайларда еркіндік дәрежелерінің саны кездейсоқ шамалардың өзгеру еркіндігін шектейтін байланыстарын есепке алады. Мысалы, (3.3) таңдамалы дисперсиясын есептеу кезінде бір байланыс бар, ол $\overline{x}$ орташа арифметикалық мәнімен анықталады, сондықтан еркіндік дәрежелерінің саны $m=N$-1 болады, ал (3.2) дисперсия үшін еркіндік дәрежелерінің саны $m=N$ сынақ санына тең болады, өйткені m_x тәуелсіз түрде анықталады.

Еркіндік дәрежесі туралы ұғымды сызықтық алгебралық теңдеулер жүйесін шешу мысалында түсіндіреміз. Бізде n белгісіз x_1, x_2, ..., x_n айнымалылары бар сызықтық алгебралық теңдеулер жүйесі бар делік. Мұндай жүйенің шешімі (теңдеулердің сызықтық тәуелсіздігі орындалатын болса) шектеулі болатыны анық, яғни мұндай жүйеде ешқандай еркіндік болмайды. Бірақ егер n белгісіз айнымалылар үшін бізде k теңдеулер болса, онда мұндай теңдеулер жүйесінде $m=n-k$ еркіндік дәрежелерінің саны болады.

Неғұрлым таңдама көлемі үлкен болса, бас жиынның математикалық күтімі мен дисперсиясы таңдамалы орташа және таңдамалы дисперсиямен соғұрлым дәлірек бағаланады.. Бұл жағдайда орта шама өлшеу нәтижесін, ал дисперсия - бұл нәтиженің дәлдігін (*бірдей қателікпен қайта өндірілудің дисперсия*) сипаттайды. Егер m параллельді тәжірибелері орындап, $x_1, x_2, \ldots, x_N$ таңдаманы алсақ, онда бірдей қателікпен қайта орындау дисперсиясы:

$$S^2_{\textit{өндірілуі}} = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{m-1},$$

ал тәжірибе қатесі (қайта орындау қатесі):

$$S_{өндірілу} = \sqrt{S^2_{өндірілу}}.$$

3.5 Интервалды бағалар

Нүктелік бағалру $\hat{\theta}$-нің сәйкес теориялық θ параметріне жақындық дәрежесі туралы ақпарат бермейді. Сондықтан белгісіз параметрлерді бағалаудың неғұрлым ақпараттық әдісі жалғыз нүктелік мәнді анықтау емес, белгілі бір сенімділік деңгейімен бағаланатын параметр орналасатын аралықты құру, яғни θ параметрін аралық бағалауды құру болып табылады.

Кездейсоқ шамалардың таралу параметрлерін бағалаудың аралық әдісі берілген сенімділік дәрежесімен бағаланатын параметрдің мәні жататын аралықты (жалғыз мәнді емес) анықтаудан тұрады. Аралық бағалау екі санмен сипатталады – оның ішінде параметрдің нақты мәні болады деп болжанатын интервалдың шектерімен. Басқаша айтқанда, бағаланатын параметр үшін бөлек нүктенің орнына мәндер аралығын орнатуға болады, осы аралық нүктелерінің біреуі «ең жақсы» бағасы болып табылады.

Аралық бағалар нүктелік бағаларға қарағанда неғұрлым толық және сенімді, олар үлкен және кіші таңдамалар үшін қолданылады. θ параметрдің мәні жататын аралықты анықтау әдістерінің жиынтығы аралық бағалау әдістері деп аталады.

Параметрдің *аралық бағасы* дегеніміз – интервалдың l_1 және l_2 шекаралары $x_1, x_2, \ldots, x_N$ таңдамалы мәндерінің функциялары болып табылатын және p берілген ықтималдықпен бағаланатын параметрді қамтитын аралық:

$$P\{l_1 \leq \theta \leq l_2\} = p.$$

(l_1, l_2) аралық - *сенімділік интервалы* деп аталады, оның кездейсоқ шамалар болып табылатын l_1 және l_2 шекаралары (олар таңдамалы мәліметтер негізінде анықталады) – сәйкесінше *төменгі* және *жоғарғы сенімділік шектері*, ықтималдық p – *сенімділік ықтималдығы*, ал $q = 1-p$ мәні - сенімділік аралығын құру кезінде қолданылатын *маңыздылық деңгей*.

Сонымен, *сенімділік интервал* дегеніміз - сандық сипаттаманы нүктелік бағасы орталығы болып табылатын кесінді, ол белгілі бір *сенімділік* ықтималдықпен берілген сандық сипаттаманың шын мәнін қамтиды. Яғни, сенімділік интервалы бағаның дәлдігінің өлшемі болып табылады, ал сенімділік ықтималдық бағаның сенімділігін сипаттайды. Тәжірибе өткізуші сенімділік ықтималдықтың қандай мәнін таңдағанына байланысты сенімділік интервалдың өлшемі орнатылады. Сенімділік ықтималдығы неғұрлым жоғары болса, берілген ықтималдықпен сандық сипаттаманың шынайы мәні орналасатын аралық соғұрлым кең болуы керек. Көбінесе сенімділік ықтималдығының мәні $p = 0.95$ таңдалады, бұл мәнді сенімділік интервалы «практикалық әрдайым» шынайы мәнді жабады деп болжауға жеткілікті үлкен

деп санайды. Тек кейде өте жауапты зерттеулер жағдайында p = *0,99* немесе *0,999* таңдалынады.

Егер сенімділік интервалы төмен сенімділік ықтималдығына сәйкес келсе, сенімділік интервалдың ені $\hat{\theta}$ бағаның жоғары сапасын сипаттамайды. Сенімділік ықтималдығы мен сенімділік аралығын әрқашан бірге қарастырылуы керек. Орнықты және ығыспаған бағалар үшін сенімділік аралықтың ені бекітілген кезінде, таңдама көлемінің ұлғаюына қарай *p* сенімділік ықтималдығы артады. Яғни, таңдаманың белгілі бір көлемінде сенімділік аралығының енін арттырмай *p*-ны көбейту мүмкін емес немесе сенімділік ықтималдығын төмендетпестен осы аралықтың енін азайту мүмкін емес. *p*-ны таңдауды зерттеуші жүргізеді; осы зерттеуге қолайлы сенімділік деңгейі жалпы жағдайда еркін түрде таңдалады, дегенмен ол нәтижелердің қажетті сенімділігіне байланысты. Көбінесе $p = 0.95$ мәні қолданылады (яғни, маңыздылық деңгейі 0.5).

Аралық бағаны табудың жалпы процедурасы:

1. Кейбір кездейсоқ функция туралы келесі түрдегі ықтималдылық өрнек жазылады:

$$P\{\delta_1 \leq g(\theta, \hat{\theta}) \leq \delta_2\} = \int_{l_1}^{l_2} f(g)dg,$$

мұндағы *f(g)* - кейбір жарамды g статистиканың ықтималдық тығыздығының функциясы. Бұнда δ_1 және δ_2 мәндері қосымша шарттарды есепке ала таңдалынады:

$$P\{g(\theta, \hat{\theta}) < \delta_1\} = P\{g(\theta, \hat{\theta}) > \delta_2\} = (1 - p)/2 = q/2.$$

Мұндай функция шамалардың жақындық дәрежесі туралы ақпарат береді. Аталған функцияның таралу заңы белгілі болуы керек;

2. Ықтималдылық өрнегі түрлендіріледі; онда сандық сипаттаманың сенімділік интервалының шекаралары айқын түрде көрсетілуі керек, яғни соңғы түрде бағаланатын параметр таңдама бойынша табылған шамалар арасында орнатылуы керек. Бұл сенімділік аралығының шекаралары болады.

Статистика $g(\theta, \hat{\theta})$ осындай түрлендіру мүмкін және белгілі *f(g)* ықтималдық тығыздығының таралу функциясы (кестелік мәндері бар болғаны жақсы) бар болатындай таңдалынады. Соңғы талап δ_1 және δ_2. мәндерін анықтауды айтарлықтай жеңілдетеді.

Сенімділік ықтималдықтарды анықтау үшін нормаланған Лаплас функциясы пайдаланылады; есептеуді жеңілдету үшін бұл функция кестелермен ұсынылған. Кездейсоқ шаманың шынайы мәнінің математикалық күтімнен ауытқуы 0,997 ықтималдығымен үш есе орташа квадраттық ауытқудан аспайды. Математикалық статистикадағы бұл қасиет «үш сигма ережесі» деп аталады.

Белгілі бір θ параметрдің мәні белгілі деген болжамда осы параметрдің $\hat{\theta}$ бағасының ықтималдық қасиеттерін талдау үшін *таңдамалы таралулар* қолданылады. Атап айтқанда, мұндай жағдайда N көлемді таңдаманы алмай тұрып, берілген p ықтималдықпен $\hat{\theta}$ мәндерінің таңдамалы таралуы орналасатын шектерді көрсету мүмкін болады. Басқаша айтқанда, бас жиынның жалпы қасиеттерін білу негізінде таңдама және оның негізінде алынған бағалар туралы пікір ұсынуға таңдамалы таралулар мүмкіндік береді. Алайда, бас жиынның параметрлері іс жүзінде белгісіз және кері мәселенің шешімі үлкен практикалық қызығушылық тудырады: белгілі таңдама бойынша бас жиынның қасиеттері туралы негізделген шешім шығару, яғни θ параметрдің қасиеттері туралы ықтималдық мәлімдемелер алу үшін $\hat{\theta}$ таңдамалы бағаны қолдану.

Келесі параграфта қарастырылатын таңдамалы таралулар деп аталатын таралулар арқылы табылатын сенімділік интервалдары бірқатар маңызды жағдайлар үшін әдебиетте келтірілген. Естеріңізге сала кетейік, бұл аралықтар тиісті кездейсоқ шамалар қалыпты таралған кезде және бақылаулар тәуелсіз болған кезде ғана орын алады.

3.6 Таңдамалы таралулар

Математикалық статистикада таралудың бірқатар теориялық заңдары кеңінен қолданылады. Жоғарыда айтылғандай, олардың ең маңыздысы *қалыпты таралу* болып табылады. Нүктелік бағалаулардың мәндері туралы әртүрлі ықтималдық мәлімдемелерді *таңдамалы таралулар* көмегімен тұжырымдауға болады. Бұл *хи-квадрат*, *Стьюдент*, *Фишер* таралулары, және *ықтималдықтар интегралы*. Бұл заңдар үшін таралу функциялары аналитикалық түрде ұсынылмайды. Функция мәндері кестелерден немесе қолданбалы пакеттердің стандартты процедуралары арқылы анықталады. Бұл кестелер келтіретін статистикалық гипотезаларды сынауға ыңғайлы болу үшін құрастырылады - оларда таралу функцияларының мәндері емес, аргументтің *критикалық* (*сыни*) мәндері орнатылған.

Дисперсия белгілі болғанда таңдамалы орта мәнінің таралуы.

Математикалық статистикада дәлелденген: егер X кездейсоқ шама m_x, σ_x^2 параметрлері бар қалыпты таралуға бағынатын болса, онда кездейсоқ келесі шама:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$m_x, \frac{\sigma_x^2}{N}$ параметрлері бар қалыпты таралуға ие.

Сонда келесідей

$$G = \frac{\overline{x} - m_x}{\sigma_x \sqrt{N}}$$

статистиканың таралуы *нормаланған қалыпты таралу* (*U-таралу*) болып табылады. Осы таралудың кестесін бастапқы таралудың m_x және σ_x^2 параметрлері белгілі болған кезде, таңдамалы орта мәні $\bar{x}$ шектері белгілі бір интервалда болу ықтималдығын таңдама алынбай-ақ болжау үшін

Басқа мәселені шешуге де болады: берілген ықтималдықпен таңдамалы орташа мәнді қамтитын интервалды анықтау.

3.1 мысал. Кездейсоқ қалыпты таралған шама келесі параметрлерге ие: $m_x = 5, \sigma_x^2 = 25$. Көлемі *N=16* таңдаманы алу жоспарлансын. *p*=0,95 ықтималдығымен $\bar{x}$ шаманың түсетін интервалын табу керек.

Келесідей шарты қанағаттандыратын l_1және l_2 шектерін анықтау керек:

$$P\{l_1 \leq \frac{\bar{x} - m_x}{\frac{\sigma_x}{\sqrt{N}}} \leq l_2\} = p.$$

Бұл шартты интервалдардың әртүрлі ендеріне сәйкес l_1 және l_2, шекараларының әртүрлі мәндерінің шексіз саны қанағаттандырады. Олардың ішіндегі ең кішісі 0-ге қатысты симметриялы, яғни $l_2 = -l_1 - l$. Сонда:

$$P\left\{-U_{\alpha=q/2} \leq \frac{\bar{x} - m_x}{\frac{\sigma_x}{\sqrt{N}}} \leq U_{\alpha=\frac{q}{2}}\right\} = 0{,}95.$$

l шекарасын табу үшін нормаланған қалыпты таралудың кестесін қолдану керек: $l = U_{\alpha=0.025} = 1{,}96$.

Сонымен,

$$P\{-1{,}96 \leq \frac{\bar{x} - m_x}{\sigma_x/\sqrt{N}} \leq 1{,}96\} = 0{,}95,$$

ал m_x, σ_x^2 және *N* мәндерін қойып, қарапайым түрлендірулерді орындап, аламыз:

$$P\{2{,}55 \leq \bar{x} \leq 7{,}45_2\} = 0{,}95.$$

Бұл қатынастың интерпретациясы: *N=16* көлемдегі таңдамаларды көп рет қайталап алсақ, орта есеппен берілген бас жиынның таңдамаларының 95%-ның $\bar{x}$ орта мәні 2,55 пен 7,45 аралығында болады.

Дисперсия бағасының таңдамалы таралуы ($(\chi^2$-таралу).

Егер X_i^*, $i = 1,2, \dots$, ν - математикалық күтімі $m_X = 0$ және дисперсиясы $\sigma_X^2 = 1$ болатын тәуелсіз нормаланған қалыпты таралған кездейсоқ шамалар болса, онда бұл айнымалылардың квадраттарының қосындысы $\chi^2 = (X_1^*)^2 + (X_2^*)^2 + \ldots + (X_\nu^*)^2$ - χ_ν^2 *таралуға бағынады («еркіндік дәрежесі ν, хи-квадрат» таралуы).

ν мәні өскен сайын χ_ν^2 таралу $M\{\chi_\nu^2\} = \nu, D\{\chi_\nu^2\} = 2\nu$ болатын қалыпты таралуға жақындайды. χ^2-таралуды пайдалану ыңғайлылы болу үшін көптеген кестелер бар.

ν=1÷30 және α ықтималдықтарының кейбір мәндері үшін кестеде ықтималдығы $P\chi^2 > \chi_{\nu,\alpha}^2 = \alpha$ болатын $\chi_{\nu,\alpha}^2$ мәндері көрсетілген. χ_ν^2-таралуының асимметриясына байланысты кестеге α-ның кіші мәндері (0,01, 0,05, 0,10), және 1-ге жақын мәндері де (0,99, 0,95, 0,90) енгізілген. Егер υ>30 болса, $\chi_{\nu,\alpha}^2$ мәндерін анықтау үшін қалыпты таралу кестелері негізінде алынған қатынасты пайдаланады:

$$\chi_{\nu,\alpha}^2 \cong \frac{1}{2}(\sqrt{2\nu - 1} + U_\alpha)^2.$$

$\chi_{\nu,\alpha}^2$-таралу таңдамалы дисперсияларының ықтималдық қасиеттерін сипаттау үшін қолданылады: егер σ_x^2 белгілі болса, берілген p ықтималдықпен S_x^2 таңдамалы мәнін қамтитын интервалын табу.

Сәйкесінші кездейсоқ шама S_X^2 келесі түрде болады:

- m_x математикалық күтім белгілі болған кезде:

$$S_X^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - m_X)^2; \tag{3.4}$$

- m_x белгісіз болғанда:

$$S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2. \tag{3.5}$$

Егер X_i параметрлері m_x, σ_x^2 болатын қалыпты таралуға бағынатын тәуелсіз кездейсоқ шамалар болса, онда келесі статистика

$$G = N\frac{S_X^2}{\sigma_X^2}$$

χ_N^2-таралуына бағынады. Шынында да

$$G = N\frac{S_X^2}{\sigma_X^2} = \sum_{i=1}^{N}\left(\frac{X_i - m_x}{\sigma_X}\right)^2 = \sum_{i=1}^{N}(X_i^2)^2 = \chi_N^2. \tag{3.6}$$

Егер m_X белгісіз болса, онда келесі статистиканы қарастырған жөн:

$$G = \frac{(N-1)S_X^2}{\sigma_X^2} = \sum_{i=1}^{N}(\frac{X_i-\bar{X}}{\sigma_X})^2. \quad (3.7)$$

m_x белгісіз болғанда, осы қатынас арқылы берілген *G* статистикасы *v=N-1* еркіндік дәрежесінің саны бар χ_{N-1}^2 -таралуына бағынатынын көрсетуге болады.

Жалпы жағдайда белгілі бір дисперсияның бағасына (немесе кейбір квадраттар қосындысына) сәйкес келетін еркіндік дәрежелерінің саны осы дисперсияның бағасы есептелетін тәуелсіз бақылаулар саны минус ретінде байқалатын мәндердің арасындағы сызықты байланыстар саны ретінде анықталатынын ескереміз. Соңғы өрнектегі $\sum_{i=1}^{N}(X_i - \bar{X})^2$квадраттарының қосындысы үшін X_i арасында жалғыз сызықты қатынас бар, атап айтқанда:

$$\sum_{i=1}^{N}(X_i - \bar{X}) = \sum_{i=1}^{N}(X_i - \sum_{i=1}^{N}\frac{X_i}{N}) = 0,$$

сондықтан бұл жағдайда *v=N-1*.

Егер σ_x^2 дисперсиясы априори белгілі болса, (3.6) және (3.7) түріндегі *G* статистикаларды берілген *p* ықтималдығымен S_x^2 нақты мәнін қамтитын интервалын табу үшін пайдаланылуы мүмкін.

3.2 мысал. Белгісіз m_x математикалық күтімі және $\sigma_x^2 = 18$ дисперсиясы болатын қалыпты жиыннан *N=10* көлемді таңдаманы алу керек. *p=0,9* ықтималдығымен S_x^2 таңдамалы мәні орналасатын аралықты табу керек.

$G = \frac{(N-1)S_X^2}{\sigma_X^2}$ статистикасы χ_{N-1}^2 ретінде таралатынын пайдаланып, l_1 және l_2 шекараларын келесі қатынастан

$$P\{l_1 \leq \frac{(N-1)S_X^2}{\sigma_X^2} \leq l_2\} = p$$

v=N-1 үшін χ_v^2-таралу арқылы табуға болады:

$$P\{\chi_{v,\alpha=0,95} \leq \frac{(N-1)S_X^2}{\sigma_X^2} \leq \chi_{v,\alpha=0,05} = 0,9.$$

Әрине, мұндай шекараларды шексіз көп жолмен табуға болады. Бірақ практикада оларды $\chi_{v,\alpha=0,95} < l_1$ және $\chi_{v,\alpha=0,05} > l_2$ аймақтарына түсу ықтималдықтар бірдей болатындай етіп таңдайды (бұл жағдайда әрқайсысы 0,05). Соңғы өрнектен түрлендірулерден кейін біз аламыз:

$$P\left\{\frac{\chi_{v,\alpha=0,95}}{N-1}\sigma_X^2 \leq S_X^2 \leq \frac{\chi_{v,\alpha=0,05}}{N-1}\sigma_X^2\}\right\} = 0,9.$$

$\chi_{v,\alpha}^2$-таралу кестесін пайдаланып, аламыз:

$$\chi_{9,\alpha=0,95}^2 = 3,325; \rightleftarrows \chi_{9,\alpha=0,05}^2 = 16,916;$$

$$P\{6,65 \leq S_X^2 \leq 33,84\} - 0,9.$$

Осылайша, қарастырылып отырған бас жиыннан *N=10* көлемді таңдамалардың үлкен санын алу кезінде орташа алғанда 10 дисперсияның 9 таңдамалы мәні (6,65; 33,84) интервалға түседі деп күтуге болады.

Белгісіз дисперсия жағдайында таңдамалы орта мәнінің таралуы (Стьюденттің t-таралуы).

Егер *U*-нормаланған қалыпты кездейсоқ шама болса, χ_v^2- *v* еркіндік дәрежелер саны бар χ^2-таралуы бойынша таралған кездейсоқ шама болса, сонымен бірге *U* және χ_v^2 - тәуелсіз, онда келесі қатынас:

$$t_v = \frac{U}{\sqrt{\chi_v^2/v}}$$

v еркіндік дәрежесі бар Стьюденттің *t* таралуына бағынады. *t*-таралуының ықтималдық тығыздығының қисығы 0-ге қатысты симметриялы және *v* өскен сайын нормаланған қалыпты таралу қисығына жақындайды. Кіші *v* үшін Стьюдент таралу қалыптыдан айтарлықтай ерекшеленеді.

Әртүрлі v үшін *t*-тарату кестесінде $P\{t_v > t_{v,\alpha}\} = \alpha$ болатындай $t_{v,\alpha}$ мәндері көрсетілген. Симметриялық қасиеті себебінен $P\{t_v < -t_{v,\alpha}\} = \alpha$ болады, сондықтан кестеде мәндер тек қана $t_{v,\alpha} > 0$ үшін берілген. Кестенің *v*=∞ жолы қалыпты таралуға сәйкес келеді және v>60 үшін *t*-таралу нормаланған қалыпты таралу арқылы өте жоғары дәлдікпен жуықталады.

Стьюденттің таралуы σ_x^2 бас жиынның дисперсиясы белгісіз болған кезде $\bar{x}$ таңдамалы математикалық күтімнің ықтималдық қасиеттерін анықтау үшін қолданылады.

Шынында да, кездейсоқ шама $G = \frac{\bar{x}-m_X}{\sigma_X/\sqrt{N}}$ нормаланған қалыпты таралуға ие және $\frac{(N-1)S_X^2}{\sigma_X^2}$ - χ_{N-1}^2таралуға бағынады; бұл кездейсоқ шамалардың тәуелсіз екенін де көрсетуге болады. Сонда келесі қатынас

$$t = \frac{\bar{X}-m_X}{\sigma_X\sqrt{N}} : \sqrt{\frac{(N-1)S_X^2}{\sigma_X^2(N-1)}} = \frac{\bar{X}-m_X}{S_X\sqrt{N}}$$

анықтамасы бойынша *v=N-1* еркіндік дәрежесі бар *t*-таралуына ие.

$G = \frac{\bar{x}-m_X}{S_X\sqrt{N}}$ статистикасын таңдаманы алмай тұрып таңдамалы $\bar{x}$ орта мәндері туралы әртүрлі ықтималдық мәлімдемелер жасау үшін пайдалануға болады. Мысалы, *p* ықтималдығымен $\bar{x}$ түсетін интервалды құру үшін.

3.3 мысал. $m_x = 5$ және дисперсиясы σ_X^2 белгісіз қалыпты жиыннан *N=16* көлемді таңдаманы алу жоспарлануда. *p=0,95* ықтималдығымен $\bar{x}$ шамасы түсетін интервалды анықтау керек.

Интервалының l_1 және l_2 шекаралары келесі шартпен беріледі:

$$P\left\{l_1 \leq \frac{\bar{x}-m_X}{S_X\sqrt{N}} \leq l_2\right\} = p.$$

0-ге қатысты симметриялы болатын және барлық ұқсас интервалдардың ең кіші ені бар интервал үшін келесіні аламыз: $l = l_2 = -l_1 = t_{\nu,\alpha=1-\frac{p}{2}}$. *t*-таралу кестесін қолданып, $\nu = N - 1$=15 және α=0,025 үшін $t_{15,\alpha=0,025} = 2,13$ табамыз.

Сандық мәндерді соңғы формулаға қойып, түрлендіруден кейін аламыз:

$$P\{5\text{-}0,5335S_X \leq \bar{x} \leq 5+0,5335S_X\}=0,95.$$

Осылайша, көлемдері *N=16* таңдамалар жиынтықтарын алу кезінде орта есеппен 95% жағдайда $\bar{x}$ мәні $(5 - 0,5335 \cdot S_X, 5 + 0,5335 \cdot S_X)$ аралықта болады. Бұл нұсқада таңдаманы алудан бұрын $\bar{x}$ үшін интервал шекараларының нақты мәнін анықтау мүмкін емес, өйткені S_X. таңдамалы мәні белгісіз болып қалады.

Екі таңдамалы дисперсиялар қатынасының таралуы (Фишердің F-таратуы).

Еркіндік дәрежелер саны сәйкесінше ν_1, ν_2 болатын $\chi^2_{\nu_1}$ және $\chi^2_{\nu_2}$-таралулары бар екі тәуелсіз кездейсоқ шама болсын. Онда келесі қатынасқа

$$F_{\nu_1,\nu_2} = \frac{\chi^2_{\nu_1}}{\nu_1} : \frac{\chi^2_{\nu_2}}{\nu_2}$$

бағынатын таралу ν_1, ν_2 еркіндік дәрежелері бар *Фишердің F -таралуы* деп аталады, бұндағы ν_1 - алымның еркіндік дәрежелерінің саны, ν_2 - бөлімнің еркіндік дәрежелерінің саны. *F*-таралудың тығыздығы - бір максимумы бар, $\nu_2 > 2$ үшін асимметриялы функция. Ол үлкен ν_1 және ν_2 үшін 1 шамасында шоғырланады. Фишер таралудың кестесінде ν_1 және ν_2 кейбір қатарлары үшін $F_{\nu_1,\nu_2,\alpha}$ кездейсоқ шамасының $P\{F > F_{\nu_1,\nu_2,\alpha}\} = \alpha$ болатын мәндері көрсетілген.

Кесте тек $F \geq 1$ үшін берілген. Бірақ оны $F<1$ үшін де қолдануға болады, өйткені $F_{\nu_1,\nu_2,1-\alpha} = \frac{1}{F_{\nu_1,\nu_2,\alpha}}$.

Ерекше жағдайлар ретінде келесіні жазуға болады:

- $\nu = \infty$ болса, $F^2_{\nu_1,\infty} = \frac{1}{\nu_1}\chi^2_{\nu_1}$.

$-\ \nu = 1$ болса, $F_{1,\infty} = t^2_{\nu_2}$, мұндағы t_{ν_2}- еркіндік дәрежелер саны ν_2 болатын t-таралуға бағынатын кездейсоқ шама.

F-таралу N_1 және N_2 көлемдері бар параметрлері екі тәуелсіз таңдамалар бойынша алынған дисперсиялардың $S^2_{X_1}$ және $S^2_{X_2}$ бағаларын салыстыру үшін пайдаланылады. Параметрлері $m_{X_1}\sigma^2_{X_1}$ және $m_{X_2}, \sigma^2_{X_2}$ болатын екі қалыпты бас жиындардан осындай таңдамаларды алғанда $\nu_1 \cdot \frac{S^2_{X_1}}{\sigma^2_{X_1}}$ және $\nu_2 \cdot \frac{S^2_{X_2}}{\sigma^2_{X_2}}$ статистикалар еркіндік дәрежелері $\nu_1 = N_1 - 1$ және $\nu_2 = N_2 - 1$ болатын χ^2-таралураға бағынады (әрине, егер m_{X_1} және m_{X_2} белгісіз болса; m_{X_1} және m_{X_2} белгілі болғанда $\nu_1 = N_1, \nu_2 = N_2$). Осының себебінен келесі қатынас:

$$F_{\nu_1,\nu_2} = \frac{S^2_{X_1}}{S^2_{X_2}} : \frac{\sigma^2_{X_2}}{\sigma^2_{X_1}} \tag{3.8}$$

F_{ν_1,ν_2}-таралуына ие. Практикада $\sigma^2_{X_1}$ және $\sigma^2_{X_2}$ теориялық дисперсиялары белгісіз, бірақ $\sigma^2_{X_1} = \sigma^2_{X_2}$ деп болжасақ, онда статистика $G = \frac{S^2_{X_1}}{S^2_{X_2}}$ келесідей ν_1 және ν_2 еркіндік дәрежелері бар F-таралуына бағынады. Ал, егерде $\sigma^2_{X_1} \neq \sigma^2_{X_2}$ болса, G статистиканың таралуы $\frac{\sigma^2_{X_1}}{\sigma^2_{X_2}} F_{\nu_1,\nu_2}$ болады, сондықтан мұндай таралудың тығыздылық функциясы 1 айналасында емес, $\frac{\sigma^2_{X_1}}{\sigma^2_{X_2}}$ мәнінің айналасында шоғырланған.

F-таралу таңдаманы алудан бұрын $\frac{S^2_{X_1}}{S^2_{X_2}}$ таңдамалы дисперсияларының қатынасы туралы белгілі бір ықтималдық мәлімдемелер жасау үшін пайдаланылуы мүмкін, мысалы, берілген p ықтималдықпен бұл қатынас орналасатын интервалды анықтау үшін.

3.4 мысал. m_{X_1} белгісіз және $\sigma^2_{X_1} = 5$ болатын бір қалыпты бас жиыннан көлемі $N_1 = 21$ таңдама алынған, ал көлемі $N_2 = 10$ болатын таңдама параметрлері $\sigma^2_{X_2} = 15$, m_{X_2}- белгісіз болатын басқа қалыпты бас жиыннан алынған. Таңдамалы дисперсиялардың қатынасын *p=0,9* ықтималдығы қамтитын интервалды табыңыз.

(3.8) өрнегінен $3 \cdot \frac{S^2_{X_1}}{S^2_{X_2}}$ шамасының $\nu_1 = 20$ және $\nu_1 = 9$ еркіндік дәрежелерімен F-таралуы бар екендігі шығады. Сонда келесідей шарт орындалатын

$$P\left\{l_1 \leq 3\frac{S_{X_1}^2}{S_{X_2}^2} \leq l_2\right\} = p = 0.9$$

интервалдың l_1 және l_2 шекараларын F таралу кестесінің көмегімен табуға болады.

Егер әдеттегідей $F < l_1$ және $F > l_2$ оқиғаларының ықтималдықтары бірдей деп есептесек (мұнда әрқайсысы 0,05), онда

$$l_2 = F_{20;9;\alpha=0.05} = 2.94; l_1 = \frac{1}{F_{20;9;\alpha=0,95}} F_{20;9;\alpha=0.05} = 0.41.$$

Сонда $P\left\{0.14 \leq \frac{S_{X_1}^2}{S_{X_2}^2} \leq 0.98\right\} = 0.9$. Сонымен, p=0.9 ықтималдығымен $\frac{S_{X_1}^2}{S_{X_2}^2}$ қатынасы 0,14-тен 0,98-ге дейінгі аралықта болады деп күтуге болады.

Осылайша, таңдамалы таралулар бас жиынтықтың бұл параметрі белгілі деп есептей отырып, белгілі бір параметрдің әртүрлі бағалау түрлерінің ықтималдық қасиеттерін талдау үшін қолданылады. Кері есепті шешуге болады: жеке таңдама үшін жалпы жиынтық параметрлерінің қасиеттері туралы негізді пайымдаулар жасаңыз, яғни параметр туралы ықтималдық мәлімдемелерді алу үшін таңдамалы бағалауды пайдаланыңыз. Атап айтқанда, популяция параметрінің интервалдық бағасын алуға болады.

Бірқатар маңызды жағдайлар үшін қарастырылған іріктеме үлестірімдерін пайдалану арқылы табылған сенімділік бағалаулары 3.1-кестеде көрсетілген. Бұл интервалдар сәйкес кездейсоқ шамалар қалыпты және бақылаулар тәуелсіз болғанда ғана орын алатынын есте ұстаған жөн.

Бірдей көлемдегі таңдамалар бойынша қалыпты бас жиынның бірнеше дисперсияларын салыстыр (Кохрен критерий).

$X_1, X_2, X_3, \ldots, X_m$ бас жиындар қалыпты таралған болсын. Бұл жиындардан бірдей N көлемдері бар m таңдамалар алынды және олар бойынша еркіндік дәрежелері бірдей $f = n - 1$ болатын S_1^2, S_2^2, S_3^3, ..., S_m^2 таңдамалы дисперсиялары табылды. Егер f мәні барлық таңдамалар үшін бірдей болса, дисперсияларды салыстыру үшін Кохрен критерийін қолдануға болады. Бұл критерий максималды эмпирикалық дисперсияның барлық дисперсиялардың қосындысына қатынасының таралу статистикасына негізделген:

$$K = \frac{S_{max}^2}{S_1^2 + S_2^2 + . S_3^2 + \cdots + S_m^2}.$$

Келтірілген K кездейсоқ шаманың таралуы тек еркіндік дәрежелерінің ν санына және m таңдамалар санына ғана тәуелді.

Нөл- гипотеза ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \cdots = \sigma_m^2$) орындалады деген болжамда, K-критерийдің мәнінің осы аймақта болатынының ықтималдығы қабылданған α

маңыздылық деңгейіне тең деп есептей отырып, критикалық аймақты оң жақты ретінде құрастырады:

$$P[K > K_\alpha(\nu; m)] = \alpha.$$

$K_\alpha(\nu; m)$ критикалық нүктесі Кохрен таралуының кестелерінен табылады. Тәжірибелер нәтижелері бойынша есептелген мән $K_{есеп} > K_\alpha(\nu; m)$ болса, α маңыздылық деңгейінде нөлдік гипотеза қабылданбайды.

3.7 Интервалдық бағаларды құру кезінде тәжірибені жоспарлау

Тәжірибені жоспарлау белгіленген сенімділік ықтималдығы p кезінде параметрді бағалаудың көрсетілген дәлдігіне қол жеткізу үшін қажетті таңдама N көлемін анықтауға дейін қысқарады. Бұл дәлдіктің сипаттамасы ретінде $\varepsilon = \frac{L}{2\sigma_x}$ салыстырмалы мән пайдаланылады.

Б.1 кестедегі (1) формула негізінде математикалық күтімді анықтау кезінде келесіні аламыз:

$$L = (\hat{l}_2 - \hat{l}_1) = \frac{2U_{\alpha=q/2}\sigma_x}{\sqrt{N}}.$$

Осыдан, $\varepsilon = \frac{U_{\alpha=q/2}\sigma_x}{\sqrt{N}}$ және $\varepsilon_{рұқ}$ рұқсат етілген ең үлкен салыстырмалы қатені ескере отырып, аламыз

$$N \geq \frac{U^2{}_{\alpha=\frac{q}{2}}}{\varepsilon^2_{рұқ}}. \qquad (3.9)$$

Яғни, таңдаманың N көлемі p ықтималдықпен және сенімділік интервалының салыстырмалы мәнімен (стандартты ауытқуға қатысты), яғни өлшенетін мәнді анықтаудың қажетті дәлдігімен анықталады.

3.5 мысал. m_x мәнін рұқсат етілген $\varepsilon_{рұқ} = 0{,}5$ қателікпен бағалау талап етілсін; $q = 0.01$. Сонда $U_{\alpha=0.005} = 2.58$.

(3.9) формулаға сандық мәндерді қойып, $N \geq 26.62$ табамыз.

Сонымен, 27 тәжірибе жасау керек.

Стандартты ауытқуды бағалау кезінде $\nu=N-1$ анықтау үшін теңдеу қолданылады (σ_x үшін сенімділік интервалы салыстырмалы қателігі 100ε% аспайтын стандартты ауытқуды бағалайды):

$$\chi^2_{\upsilon,\alpha=q/2}/\chi^2_{\upsilon,\alpha=1-q/2} \leq (1 + \varepsilon_{рұқ})^2.$$

Шешімі χ^2-таралу кестесін қолдана отырып таңдау әдісімен ізделеді: берілген теңсіздікті қанағаттандыратын ν мәні, содан кейін $N = \nu + 1$ мәндері табылады.

3.8 Бақылау сұрақтары

3.8.1 Бас жиын мен таңдама анықтамаларын беріңіз.

3.8.2 Реперезантты таңдама дегеніміз не?

3.8.3 Кездейсоқ шама таралуының параметрінің статистикалық бағасының анықтамасын беріңіз.

3.8.4 Параметрлердің бағаларына сапа көз қарасы жағынан талаптары қандай?

3.8.5 Максималды ықтималдық әдісін түсіндіріңіз.

3.8.6 Моменттер әдісін түсіндіріңіз.

3.8.7 Кездейсоқ шамалардың қасиеттерін сипаттаудағы қалыпты таралудың рөлі.

3.8.8 Бас жиын параметрлерін анықтаудың теориялық формулалары.

3.8.9 Параметрлердің нүктелік және интервалдық бағалары.

3.8.10 Математикалық күтім мен дисперсияны бағалау кезінде таңдаманың қажетті көлемін анықтау үшін тәжірибелерді жоспарлау.

4 Статистикалық гипотезалар

4.1 Статистикалық гипотезаларды тексеру

Бас жиының кейбір ерекшеліктерін зерттеу кезінде алдымен одан репрезентантты таңдама алынатынын естеріңізге сала кетейік; осы таңдаманы зерттеу негізінде бүкіл жиын туралы қорытынды жасалады (*таңдамалы* әдіс). Бас жиынның параметрлерін статистикалық бағалау кезінде белгісіз таралу параметрінің бағаларын таңдауды негіздеу үшін белгілі бір жорамал жасалуы керек. Мысалы, зерттелетін шаманың таралуы қалыпты таралу заңына қайшы келмейтіні; параметрлер бағаларын есептеу кезінде олардың арасында өрескел қателер (өте айырмашылықтары үлкен көрсеткіштер) табылмағаны туралы.

Кездейсоқ шамаларды анықтау үшін ғана емес, сонымен қатар оларды бір-бірімен салыстыру қажет болғанда, тіпті тәжірибе нәтижелері бойынша жауап беру функциясын құрастыру кезінде одан да көп әртүрлі болжамдармен бетпе-бет келуге тура келеді. Бұл есептерді тексеру және шешу статистикалық гипотезалардың көмегімен жүзеге асырылады.

Статистикалық гипотеза – таңдама негізінде тұжырымдалған таңдама алынатын бас жиынның қасиеттері туралы белгілі бір болжам (белгісіз таралу туралы немесе белгілі таралулардың параметрлері туралы болжам).

Нөлдік гипотеза (нөл-гипотеза) деп алға қойылған гипотеза (тексерілетін гипотеза) аталады және H_0 арқылы белгіленеді. Көбінесе әртүрлі таңдамалар үшін алынған бағалар бір бас жиынға жататыны тұжырымдалады, яғни бағалар

арасындағы айырмашылық шын мәнінде нөлге тең және элементтердің кездейсоқ таңдалуы мен таңдаманың шектеулі көлеміне байланысты пайда болды. Егер бас жиынның қандай да бір сапалық қасиеті сыналатын болса, онда гипотезаны тексеруді сандық түрге келтіруге тырысады.

Нөлдік гипотезаға қарама-қарсы қойылатын гипотеза *альтернативті гипотеза* деп аталады және H_1 деп белгіленеді.

Егер нөл гипотеза қабылданбаса, онда альтернативті гипотеза орын алады.

Бұл гипотезалардың дұрыстығы бақылау (өлшеу) мәліметтері бойынша кейбір сандық сипаттамаларды есептеу жолымен және тексерілетін гипотеза ақиқат болу шартында осы сипаттамалардың мүмкін мәндерімен салыстыру арқылы тексеріледі, ал байқалған ауытқулар таңдамалардағы кездейсоқ өзгерулері мен олардың шектеулерінен пайда болады. Мұндай сипаттамалар статистикалық гипотезаларды тексеру критерийлері деп аталады.

Статистикалық гипотеза критерийі – таңдама негізінде гипотезаны қабылдамауға немесе қабылдауға мүмкіндік беретін ереже. Таңдама бойынша есептелген критерийдің мәні эмпирикалық (есептелген) мән деп аталады. Мұндай ережені құру кезінде бақылау нәтижелерінің кейбір функциялары қолданылады, оларды *гипотезаларды тексеру үшін статистикалар* деп аталады.

Белгілі бір гипотезаны тексеру үшін мұндай статистиканың барлық мүмкін мәндері екі топқа бөлінеді: гипотезаны қабылдау аймағы және сыни аймақ.

Гипотезаны қабылдау аймағы (рұқсат етілген мәндер аймағы) нөлдік гипотеза қабылданатын критерий мәндерінің жиынтығы болып табылады.

Сыни (критикалық) аймақ - гипотезаны жалған деп қабылдамау шешімді ұсынуға мүмкіндік беретін статистиканың барлық мәндерінен құрылады.

Сыни нүктелер (шекаралар) деп гипотезаның қабылдау аймағынан сыни аймағын бөлетін нүктелер аталады.

Гипотезаны тексеру деген қолданатын статистиканың мәні сыни аймағына түсетінін немесе түспейтінін анықтау болады: егер ол оған түспеген болса, гипотеза бақылау нәтижелеріне қайшы келмейтін болып қабылданады, егер түссе - гипотеза қабылданбайды.

Осылайша, статистикалық гипотезалар ықтималдық сипатта болады. Бұл кейбір жағдайларда қателесу мүмкін екенін көрсетеді. Қателік дәрежесін сандық анықтау үшін *маңыздылық деңгей* деп аталатын q көрсеткіші қолданылады. $q \cdot 100\%$ көбейтіндісі 100 жағдайдың қаншасы қате болуы мүмкін екенін, яғни гипотеза қанша пайыз қабылданбағанын көрсетеді. Сонда $p=1-q$ мәні *сенімділік деңгейі* болады. Маңыздылық деңгей – жауапқа біздің талаптарымыздың өлшемі болып табылады: неғұрлым үлкен кепілдіктер талап етілсе, жауап соншалықты аз анықталған болады. Жауап өте дәл және өте сенімді, әдетте өте қымбат. Бірнеше ондаған жылдар бойы әртүрлі жағдайларда статистиканы қолдану тәжірибесі әдетте практикалық жағдайларда q шаманың анықтаушы мәні 0,05 болатынын көрсетті. Кейде 5% тәуекел деңгейі деп

аталатын бұл мән дұрыс жауаптың ықтималдылығына сәйкес келеді, яғни гипотезаларды тексеру кезінде оның сенімділігі $p=1-0,05=0,95$ немесе 95%. Яғни орта есеппен 100 жағдайдың 5-і ғана ықтимал қателер екенін айтады. Әрине, ешқашан 100% кепілдік беру мүмкін емес. Мұндай тілек шексіз көп тәжірибелерді жүргізу қажеттілігіне әкеледі, бұл абсурд.

Гипотезаны тексеру шектеулі ақпаратпен (таңдама бойынша) жүзеге асырылатындықтан, екі түрлі қателер болуы мүмкін.

Егер дұрыс гипотеза қабылданбайтын болса, *бірінші түрдегі қате* жіберіледі. Бірінші түрдегі қатенің ықтималдығы *маңыздылық деңгейі q* болып табылады.

Если будет отвергнута правильная гипотеза, то совершается *ошибка первого рода*

Егер дұрыс емес гипотеза қабылданса, *екінші түрдегі қате* жіберіледі. Екінші түрдегі қатенің ықтималдығы β арқылы белгіленеді. Ол қуат ұғымымен байланысты.

Критерийдің қуаты дегеніміз - қате гипотезаны қабылдамаудың ықтималдығы болып табылады және ($1-\beta$-ге тең. Критерийдің қуаты неғұрлым көп болса, оның екінші түрдегі қатені жіберу ықтималдығы соғұрлым аз болады. Мысалы, өнімнің сапасын бақылау кезінде q мәні дұрыс гипотезаны жоққа шығара отырып, өндіріс тәуекелін көрсетеді, бұл мән жақсы өнімді жоққа шығарады. β мәні жалған гипотезаға жол бере отырып, тұтынушының тәуекелін сипаттайды, бұл мән іс жүзінде жарамсыз өнімді қабылдайды және пайдаланады.

Гипотезаларды статистикалық тексеру кезінде туындайтын жағдайлар 4.1 - 4.2 кестелерде берілген.

4.1 кесте – Гипотезалардың дұрыстығы туралы шешім қабылдаудың мүмкін нұсқалары

Гипотеза	Объективті дұрыс	Объективті жалған
Қабылданады	Дұрыс шешім	2-ші түрдегі қате
Қабылданбайды	1-ші түрдегі қате	Дұрыс шешім

4.2 кесте – Гипотезаларды тексерудегі қателер

	Шешім	
	H_0 қабылдау	H_1 қабылдау
H_0 дұрыс	Ықтималдығы 1-α болатын дұрыс	Ықтималдығы α болатын қате
H_1 дұрыс	Ықтималдығы β болатын қате	Ықтималдық 1-β болатын дұрыс

Бірінші және екінші түрдегі қателердің ең кіші болатын жалпы ықтималдығын алуға мүмкіндік беретін критерийлерді табу қиын мәселе. Сондықтан маңыздылық деңгейінің белгілі бір мәні таңдалып, β-ны минимумдайтын критерийді табуға әрекет жасалады. Іс жүзінде келесідей q мәндері қолданылады: 0,01, 0,05 (ең таралған), 0,1.

Таңдама көлемі мен маңыздылық деңгей белгілі болса, бір гипотезаны тексеруге арналған екі критерийдің ішінен қандай да бір мағынада көбірек қуатқа ие болатынын (мысалы, осы критерийдің қуат функциясының қисығы салыстырылған критерийдің ұқсас қисығына қарағанда жоғарырек орнатылады) таңдау дұрыс болады

Таңдама өлшемі берілген болса, I түрдегі қатені жасау ықтималдығын q маңыздылық деңгейін төмендету арқылы азайтуға болады. Бірақ бұл жағдайда екінші түрдегі қателік ықтималдығы β артады (критерийдің ұуаты төмендейді). Екі ықтималдықты да азайтудың жалғыз жолы таңдама көлемін ұлғайту болып табылады (сонда параметр бағасының таралу тығыздығы «тар» болады). Сыни аймақты таңдау кезінде Нейман-Пирсон ережесін басшылыққа алады: гипотеза ақиқат болса q ықтималдығы аз, ал басқа жағдайда үлкен болатындай сыни аймақты таңдау керек. Дегенмен, q-ның белгілі бір мәнін таңдау салыстырмалы түрде ерікті. Көбінесе қолданатын мәндер 0,001-ден 0,2-ге дейін аралықта болады.

Маңыздылық деңгейді таңдау кезінде альтернативті гипотеза бойынша критерийдің қуатын ескеру қажет. Кейде критерийдің үлкен қуаты кіші маңыздылық деңгейінен маңыздырақ болып шығады және оның мәні салыстырмалы түрде үлкен таңдалады, мысалы, 0,2. Егер екінші түрдегі қателердің салдары бірінші түрдегі қателерге қарағанда маңыздырақ болса, мұндай таңдау, ақталады. Мысалы, егер «пайдаланушылар жұмысын ағымдағы құпия сөздермен жалғастыру» дұрыс шешімі қабылданбаған болса, бірінші түрдегі қате құпия сөздерді өзгертуге байланысты жүйенің қалыпты жұмысында біршама кідірістерге әкеледі. Рұқсат етілмеген тұлғалардың ақпаратқа рұқсатсыз кіру қаупіне қарамастан, құпия сөздерді өзгертпеу туралы шешім қабылданса, бұл қате одан да ауыр зардаптарға әкеледі.

4.2 Маңыздылық критерийлері

Белгілі (параметрлерге дейін) ықтималдық таралу функциясы бар бас жиыны үшін параметрлердің абсолютті мәндері немесе олардың арасындағы байланыстар туралы гипотезалар тексерілетін критерийлер *маңыздылық критерийлері* деп аталады. Мұндай гипотезаларды тексеру кезінде бас жиыны таралу параметрлердің тұтас тобымен анықталса, ал тексерілетін гипотеза олардың бір бөлігіне ғана қатысты болса, қалған параметрлер белгілі деп есептеледі немесе таңдама мәліметтерінен табылады.

Маңыздылық критерийлері көмегімен анықталады:

а) бақылаулар нәтижелері бас жиынның кейбір параметрлері белгілі бір сандарға тең деген болжамға қайшы келмей ме;

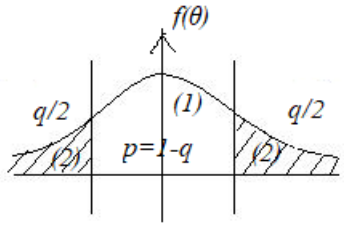
б) егер екі немесе одан да көп таңдамалар болса, оларды бірдей бас жиынның өкілдері деп санауға объективті негіздер бар ма немесе тәжірибелік мәліметтер таңдамалар арасында кездейсоқ емес айырмашылықтардың болуын көрсетеді, бұл олардың әртүрлі бас жиындардан алынғанын көрсетеді.

Кейбір баға $\hat{\theta}$ көлемі N таңдама бойынша есептелсін. Бағаланатын θ параметрінің шын мәні, яғни оның бас жиындағы мәні θ_0 тең деуге негіздер болсын. Бұл тексерілетін болжам, *нөлдік гипотеза*, келесідей белгіленеді H_0: $\theta = \theta_0$.

Нөлдік гипотеза ақиқат болса да, параметрдің таңдамалы мәні θ_0-ға дәл сәйкес келмейді, өйткені $\hat{\theta}$ дегеніміз көлемдері N болатын барлық ықтимал таңдамалармен құрылған $\hat{\theta}$ кездейсоқ шамасының ерекше мәндерінің біреуі ғана. Нөлдік гипотезаны жеткілікті негізде қабылдамау үшін $\hat{\theta}$ шамасының θ_0-ден қандай айырмашылығы болуы керек? Нөлдік гипотеза ақиқат болса, $\hat{\theta}$ таңдамалы шамалардың мүмкін болатын мәндер диапазонынан әртүрлі аймақтарына түсудің ықтималдығын білу қажет. Яғни, нөлдік гипотеза ақиқат деген болжамымен теориялық тұрғыда құрылған $f_{\hat{\theta}}(\hat{\theta})$ ықтималдықтың тығыздық функциясын білу керек. Бұл функция $f_{\hat{\theta}}(\hat{\theta})$ статистиканың таңдамалы таралуы болып табылады. Егер ол алынған болса, нөлдік гипотеза ақиқат болған кезде оның көмегімен таңдама $\hat{\theta}$ мәнінің түсуінің ықтималдығы өте төмен болатын аймағын табу оңай.

Осындай түсу ықтималдығы жеткілікті аз q санына тең болсын. Сонда $\hat{\theta}$ бәрібір осы аймаққа түссе, нөлдік гипотезаның дұрыстығына күмән келтіруге елеулі себептер бар. Мұндай төмен ықтималдығы бар аймағы критикалық (сыни) аймақ ретінде пайдаланылуы мүмкін, егер $\hat{\theta}$ оның ішіне түссе, нөлдік гипотезаны қабылдамауға болады. Әрине, бұл жағдайда бірінші түрдегі қатені жіберу мүмкіндігі сақталады, дегенмен бұл қатенің ықтималдығы q-ға тең, шарт бойынша кіші шама және маңыздылық деңгейін таңдау арқылы анықталады.

Сыни аймақтың түрі H_1 *альтернативті гипотезаның* сипатына байланысты, яғни нөлдік гипотезаға қарама-қарсы қойылатын және зерттелетін нөлдік гипотезаны қабылдамай отырып, оның пайдасына зерттеуші бейім гипотеза. Нөлдік гипотезаға H_0: $\theta=\theta_0$ альтернативті гипотезаға H_1: $\theta\neq\theta_0$ қарсы қойылса, онда H_0 -ді тексеру критерийі екі жақты деп аталады, ал оның сыни аймағы екі бөліктен тұрады (4.1 сурет).



4.1 сурет – Екі жақты критерийдің гипотезаны қабылдау (1) және сыни (2) аймақтары

Әдетте сыни (критикалық) аймақтың шекаралары сыни аймақтың сол және оң бөліктеріне түсу ықтималдығы бірдей болатындай етіп таңдалады ($q/2$-е тең). Сонда екі жақты сыни аймағы $|k|>k_{kp}$ теңсіздігімен анықталады.

Егер H_0:$\theta=\theta_0$ нөлдік гипотезамен альтернативті гипотезалар H_1: $\theta>\theta_0$ немесе H_1: $\theta<\theta_0$ түрінде тұжырымдалса, онда сәйкес критерийлер *бір жақты* (оң жақты және сол жақты) деп аталады және олардың сыни аймақтары тек бір ауданды қамтиды (4.2 сурет).

$k < k_{kp}$ $\qquad$ $k > k_{kp}$

а) $\qquad$ б)

4.2 сурет – Бір жақты критерий үшін гипотезаны қабылдау (1) және сыни (2) аймақтары

а) H_1: $\theta<\theta_0$я, б) H_1: $\theta>\theta_0$

Маңыздылық критерийін құрудың негізгі кезеңдері:

1. H_0 нөлдік гипотезаны ұсыну;
2. H_1 альтернативті гипотезаны анықтау;
3. маңыздылықтың жарамды q деңгейін орнату (бірінші түрдегі қателік ықтималдығы);
4. Гипотезаны тексеру үшін жарамды g статистиканы таңдау. Бұл статистика, мүмкіндігінше, келесі талаптарды қанағаттандыруы керек: а) есептеуі оңай; б) белгілі (кестеленген) таңдамалы таралуы болуы; в) ең жақсы болу (мысалы, ең қуатты критерийді алу мағынасында). Көптеген гипотезаларды тексеру үшін мұндай статистикалар белгілі және жақсы зерттелген (3.6 б.қараңыз);
5. H_0 гипотезасы ақиқат деп есептей отырып, пайдаланылған статистикалық мәліметтердің таңдамалы таралуын анықтау;
6. H_1 альтернативті гипотезаны және q маңыздылық деңгейін ескере отырып, гипотезаны тексеру үшін сыни аймақты орнату; ол үшін біз q маңыздылық деңгейін таңдап, келесі қатынастар негізінде сыни (критикалық) нүктелерді іздейміз:

- оң жақты сыни аймағы үшін $P(k>k_{kp})=q(k_{kp}>0)$;
- сол жақты сыни аймағы үшін $P(k<k_{kp})=q(k_{kp}<0)$;
- екі жақты симметриялы аймақ үшін $P(k > k_{kp}) = \frac{q}{2}(k_{kp} > 0)$; $P(k < -k_{kp}) = \frac{q}{2}$;

7. *N* көлемі бар кездейсоқ таңдаманы алу, статистиканың мәнін есептеу және шешім қабылдау: егер статистика шешім қабылдау аймағына түссе, онда H_0 гипотезасы дұрыс болады; егер статистика сыни аймақта болса, онда H_0 қабылданбайды (H_0 ақиқат болсда *100q%* жағдайда қабылданбайды).

4.1-мысал. Кейбір құрылғының статикалық ток беру коэффициентінің мәндері $\sigma_x^2 = 25$ параметрі бар кездейсоқ, қалыпты таралған мәндер болып табылады. Параметрдің номиналды мәні 50-ге тең.

20 құрылғы үшін коэффициент мәндерін өлшей отырып, коэффициенттердің номиналды мәнге тең екендігін анықтау қажет, егер оның төмен ауытқуы қажет болмаса.

Тексеру реті:

1) нөлдік гипотезаны тұжырымдаңыз

$$H_0: m_x = 50;$$

2) альтернативті гипотеза ретінде таңдалады:

$$H_1: m_x < 50$$

(өйткені коэффициенттің кіші жағына ауытқуы қажет емес);

3) маңыздылық деңгейін таңдаймыз *q=0,05*;

4) зерттелетін мәннің дисперсиясы белгілі болғандықтан, пайдалану үшін ең жақсы статистика:

$$g = \frac{\overline{x} - m_x}{\sigma_x/\sqrt{N}};$$

5) g статистикасы нормаланған қалыпты таралу заңына ие;

6) біржақты критерийдің сыни аймағының шекарасы нормаланған қалыпты таралу кестесі бойынша табылады және тең

$$-U_{\alpha=q} = -U_{\alpha=0.05} = -1{,}64$$

(минус белгісімен аламыз, өйткені сыни аймақ сол жақта);

7) коэффиценттің 20 тәуелсіз өлшеуі жүргізілді; $\bar{x}$ орташа мәні алынады; статистиканың таңдамалы мәні *g=-1,43* есептелді.

Статистиканың таңдамалы мәні сыни аймаққа түспейтіндіктен, тәжірибе нәтижелері айтылған болжамға қайшы келмейді деп есептей отырып, нөлдік гипотезаны қабылдау керек.

Бас жиын параметрлерінің мәндеріне қатысты әртүрлі гипотезаларды тексеру үшін жиі қолданылатын критерийлер математикалық статистика бойынша әдебиеттерде келтірілген. Оларды келесі шарттар орындалған жағдайда пайдалануға болады:

- таңдамалар кездейсоқ, олар тәуелсіз бақылаулар бойынша алынған;
- зерттелетін жиындар қалыпты таралуға ие.

Бұл критерийлер олардың барлық ықтимал түрлерін бермейді.

4.3. Өрескел қателерді жою

Ықтималдық теориясынан белгілі: қалыпты заңы бойынша таралған кездейсоқ шама өзінің математикалық күтімінен келесідей ауытқымауы тиіс: 90% болса 1,64σ_x шамасына, 95% - 1,96 σ_x шамасына, т.с. Бұл m_x математикалық күтім. σ_x ортаквадратты ауытқу белгілі және өлшеу нәтижелері саны N болғанда қалыпты таралу заңына бағынады деген болжамдағы теориялық есептеулер. Сонымен, қалыпты таралу заңы жағдайында, бақылаулардың жеткілікті үлкен саны болса (практикалық тәжірибе көрсеткендей, 30-дан астам), егер кейбір өлшенген мәні өзінің математикалық күтімінен 1,96·σ_x шамадан астам ерекшеленсе, оны 95% ықтималдығымен алып тастауға болады.

Практикада, әдетте, өлшеулер саны шектеулі және көп жағдайда 15-30 аспайды. Бақылаулардың осындай аз санымен біз тек m_x математикалық күтім мен σ_x стандартты ауытқудың бағаларын анықтай аламыз.

Өлшеу өрескел қателерді жою процедурасы келесідей:

1. Бақылаулардың (өлшеулердің) нәтижелері және таңдама көлемі N негізінде математикалық күтім $\overline{x}$ және S_x^2 дисперсияның бағалалары есептеледі.

2. Бақылаулардың барлық қатарынан x_{max} орташа арифметикалық мәнінен ең үлкен ауытқуы бар бақылау (өлшеу) таңдалады.

3. Нөлдік гипотезасы H_0 тұжырымдалады: x_{max}-тың $\overline{x}$-тен ауытқуы p сенімділік ықтималдығымен (q маңыздылық деңгейімен) мәнсіз.

4. Бұл гипотезаны бағалау үшін ең үлкен салыстырмалы (орташа квадратқа қатысты) ауытқу есептеледі:

$$t_{\text{эксп}} = \frac{|x_{max} - \bar{x}|}{S_x}, \tag{4.1}$$

мұндағы x_{max} – «секіріс» (ауыту) мәні.

5. Статистикалық нөл-гипотезаны тексеру критерийі ретінде q маңыздылық деңгейіне немесе $p=1-q$ сенімділік ықтималдығына тәуелді t Стьюдент критерийінің $t_{q,m=N-1}$ теориялық мәні пайдаланылады, бұл кездейсоқ шаманың рұқсат етілген ауытқуы, стандартты ауытқу бағаларының бөліктерімен өрнектеледі және таңдаманың шектеулі көлемін (N) мен көрсетілген сенімділік ықтималдықты (p) есепке алады.

6. Егер $t_{эксп} > t_{\alpha,m}$, болса, онда p ықтималдығымен «секіріс» мәнін өрескел қате ретінде алып тастауға және нөл-гипотезаны қабылдамауға жеткілікті себеп бар. Кері жағдайда, H_0 нөлдік гипотезасы қабылданады және p ықтималдығымен «секіріс» мәнін алып тастау қажет емес.

4.1 мысал. Пирометр қыздырылған дененің бетінің температурасы өлшенеді. Қыздырылған дененің көрінетін бетінің температурасы барлық

нүктелерде бірдей деп есептейміз. Әртүрлі зерттеушілер алты температураны өлшеп, келесі мәндерді алды: ^{0}C: 925, 950, 975, 1000, 1025, 1050 (N=6).

Осы өлшемдердің арасында өрескел қателер бар ма? Алдын-ала $\overline{x}$ и S_x бағаларын есептейміз:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 987.5^0 C; S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{N-1}} = 46.8^0 C.$$

Өлшенетін температураның шын мәнін білмегендіктен, S_x бағаны анықтау формуласында (*N-1*) қолдандық. Бұл маңызды екенін ескеріңіз, өйткені саны аз өлшемдер жүргізілді (тек қана *N=6*).

Орташа арифметикалық мәннен ең үлкен ауытқуы бар өлшемдерді таңдаймыз. Мұндай екі мән болды: 925 ^{0}C және 1050 ^{0}C.

Алдын ала есептейміз:

$$t_{\text{тәж}} = \frac{1050 - 987{,}5}{46{,}8} = 1{,}34.$$

q=0,05 және *m=N-1=5* кезінде $t_{0,05;5}=2,57$ анықтаймыз.

$t_{эксп}<t_{\alpha,m}$, болғандықтан, асып тұрған бақылаулардан алып тастаудан бас тартқан дұрыс.

Статистикалық қорытындының ерекшеліктері. Статистикалық қорытынды кез келген статистикалық әдіс сияқты белгілі бір қарама-қайшылықтарды қамтиды. Статистикалық қорытындының ерекшелігі. Гипотеза қабылданбаған кезде (дұрыс болжам жоққа шығарылды) мүмкін болатын қатенің ықтималдығын алдын ала бағалауға болады. Бірақ егер гипотеза қабылданса, бұл оның берілген ықтималдықпен расталғанын білдірмейді. Бұл статистикалық тәжірибенің нәтижелерімен келіседі, дегенмен, мүмкін, қосымша тәжірибелік материал алынған кезде, одан бас тарту керек екені белгілі болуы мүмкін. Басқаша айтқанда, статистикалық гипотезаларды тексеру әдісі сол немесе басқа гипотезаны дұрыс емес деп жоққа шығаруға мүмкіндік береді, бірақ оның ақиқаттығын дәлелдеуге мүмкіндік бермейді. Яғни, статистикалық критерийлер бұл тәжірибелер арқылы теріске шығарудың жоқтығын ғана көрсетеді.

Практикада бақылаулардың нәтижелерімен бірдей сәйкес келетін көптеген әртүрлі гипотезаларды ұсынуға болады және барлық мүмкін гипотезалардың қарастырылғанына ешқашан кепілдік беруге болмайды. Статистикалық гипотезаларды тексерудің формальды процедуралары гипотезаларды қабылдау немесе қабылдамау туралы шешім қабылдау ережелерін ғана қамтамасыз етеді, сонымен бірге, статистикалық талдау нәтижелеріне тым көп объективтілік жатқызуға болмайды, өйткені әрқашан априорлық, субъективті ақпарат қолданылады.

4.4 Бақылау сұрақтары

4.4.1 Статистикалық гипотеза анықтамасын беріңіз.
4.4.2 Нолдік гипотеза деген не? Ол қалай тұжырымдалады?
4.4.3 Альтернативті гипотеза қалай тұжырымдалады?
4.4.4 Гипотезаларды тексерудегі қателер түрлерін түсіндіріңіз.
4.4.5 Маңыздылық деңгей ұғымын түсіндіріңіз.
4.6.6 Мағыздылық критерий деген не?
4.4.7 Бас параметрлер мәндерін тексеруге қолданылатын стандартты критерийлерді атаңыз.
4.4.8 Әртүрлі альтернативті гипотезлар үшін гипотезаларды қабылдау аймағы қалай анықталады?
4.4.9 Сыни аймақ дене не? Оны қалай анықтауға болады?
4.4.10 Гипотезаларды тексерудің негізгі қадамдарын келтіріңіз.

5. Объектілердің регрессиялық модельдерін құру

5.1 Регрессиялық талдаудың элементтері

Активті тәжірибені жоспарлаудың статистикалық әдістері күрделі зерттеу объектілерінің статикасының математикалық бейнелеуін алудың эмпирикалық әдістерінің бірі болып табылады, яғни *y* объектісінің жауабы мен тәуелсіз бақыланатын нормаланған кіріс айнымалылары (факторлары) $\bar{x} = (x_1; x_2; \dots, x_n)$ арасындағы байланыс теңдеуі.

Сонымен бірге математикалық бейнелеу белгілі бір көпмүше – Тейлор қатарының үзіндісі түрінде беріледі, оған белгісіз тәуелділік негізгі $\bar{x}_0$: нүктесіне жақын жерде жіктеледі:

$$y = M[y\} = \varphi(x_1, x, \dots,\ x_n) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{\substack{i=1,l \\ i<l}}^{n} \beta_{il} x_i\, x_l + \sum_{i=1}^{n} \beta_{ii} x_i^2,$$

мұндағы

$\beta_i = \left.\frac{\partial \varphi}{\partial x_i}\right|_{\bar{x}=\bar{x}_0}$, $\beta_{il} = \left.\frac{\partial^2 \varphi}{\partial x_i \partial x_l}\right|_{\bar{x}=\bar{x}_0}$, $\beta_{ii} = \left.\frac{\partial^2 \varphi}{\partial^2 x_i}\right|_{\bar{x}=\bar{x}_0}$ - теориялық коэффициенттер.

Басқарылынбайтын және тіпті бақыланбайтын факторлардың болуына байланысты *y* шаманың өзгеруі кездейсоқ сипатқа ие, сондықтан $\varphi(\bar{x})$ функционалдық тәуелділік бақыланатын факторлар мен әрбір g-ші тәжірибеде объектінің y_q реакциясы арасында нақты байланысты бермейді. Тек қана басқарылатын факторлар мен кездейсоқ *y* шамасының математикалық күтімі арасындағы байланысты:

$$M\{y_g\} = \varphi(\bar{x}_g),$$

мұнда $\bar{x}_g = (x_{1g}; x_{2g}; \dots, x_{ng})$ - тәуелсіз басқарылатын факторлар кеңістігінің нүктесі (факторлық кеңістіктің) g – ші нүктесі.

Зерттеу объектілерінің мұндай модельдерін құрудың классикалық статистикалық әдісі регрессиялық талдау болып табылады. Бұл жағдайда тәжірибе нәтижелері бойынша белгілі бір көпмүше түріндегі регрессия теңдеуін табуға болады:

$$\hat{y} = b_0 + \sum_{i=1}^{n} b_i x_i + \sum_{\substack{i;l=1 \\ i<l}}^{n} b_{il} x_i x_l + \sum_{i=1}^{n} b_{ii} x_i^2 + \cdots,$$

мұнда b_0, b_i, b_{il},b_{ii}, … таңдамалы регрессия коэффициенттері сәйкесінше $\beta_0, \beta_1, \dots, \beta_{ii}$,... теориялық коэффициенттер үшін тек бағалау болып табылады, ал $\hat{y}$- $M\{y\}$ үшін бағалау.

Регрессиялық талдаудың негізгі мақсаты – тәжірибелік мәліметтерді қолдана отырып зерттеу объект моделінің келесі түрдегі регрессиялық тәуелділігін алу болып табылады:

$$y = \varphi(\bar{x}) + e,$$

яғни аддитивті кездейсоқ сипаты бар e шулардың болуы кезінде шығыс айнымалының мәнін $\bar{x}$ факторларымен байланыстыратын формулалық тәуелділікті алу.

Ең көп таралған модельдер параметрлері бойынша сызықтық болып табылады, яғни:

$$y = \varphi(\bar{x}) + e = \sum_{j=0}^{d} \beta_j f_j(\bar{x}) dx = \bar{f}^T(\bar{x})\bar{\beta} + e,$$

мұндағы $\bar{x} = \{x_1, \dots, x_n\}$ – факторлар векторы

$\bar{\beta} = \{\beta_1, \dots, \beta_n\}$ – белгісіз регрессия коэффициенттерінің векторы;

$f(\bar{x}) = \{f_0(\bar{x}), f_1(\bar{x}), \dots, f_d(\bar{x})\}$- базистік функциялар векторы.

Базистік функциялар жүйесін тәжірибелер және одан кейінгі есептеулер алдында таңдау керек. Бұл жүйені таңдау, әдетте, $\varphi(\bar{x})$ тәуелділіктің сипаты туралы априорлық ақпаратқа және есептеулер үшін максималды қарапайымдылықты қамтамасыз етуге негізделген

Теориялық тұрғыдан модельдің түрі белгілі деп есептелінеді және N тәжірибені пайдаланып тәжірибелер мәліметтерден белгісіз $\bar{\beta}$ табу қажет. Нақтылықта адекватты (зерттеу объектілеріне сәйкес) модельді сынап көру және қателесу арқылы таңдау қажет болады. Көбінесе ортогональды көпмүшеліктер жүйелері (Лагер полиномдары, Лежандр полиномдары), тригонометриялық функциялар жүйесі және көпмүшелік айнымалылар жүйесі қолданылады. Базистік функциялардың мәндерін әрқашан факторлардың белгілі өлшемдерінен есептеуге болатынын ескеру маңызды. Практикада көпмүшелік регрессия модельдері жиі қолданылады. Онда адекватты модельді таңдау мәселесі ең қарапайымнан (бірінші ретті көпмүшеден) бастап, ретін біртіндеп арттыра отырып, көпмүшенің ретін таңдауға дейін қысқарады. Әдетте, факторлардың саны неғұрлым көп болса, орнатылған модельдің максималды дәрежесі соғұрлым төмен болады.

Мұндай модельдерді құру үшін *e* кездейсоқ құрастырушының статистикалық қасиеттеріне қатысты кейбір алғышарттар енгізіледі:

1. Факторлар кеңістігінің *N* нүктелеріндегі жауап функцияның $y_1,...y_N$ бақылауларының нәтижелері қалыпты таралған кездейсоқ шама болып табылады, яғни оларға нөлдік математикалық күтімі $M\{e\}=0$ және тұрақты дисперсиясы бар $\sigma_e^2=const$ қалыпты таралған кездейсоқ шу әсер етеді. Дисперсияның тұрақтылығы шудың интенсивтілігі бақылаулар арасында өзгермейтінін білдіреді. Бұл талап регрессиялық талдаудағы барлық статистикалық процедураларды дұрыс орындау үшін қажет.

2. Әртүрлі бақылаулардағы шу мәндері корреляцияланбайды. Бұл талапты қамтамасыз ету үшін тәжірибелерді рандомизациялау, яғни оларды орындау реті кездейсоқ таңдалады.

3. $\sigma^2\{y_g\}$ $g=1,..., N$ дисперсиялары тең. Басқаша айтқанда, x_g нүктесіндегі y_g мәнін бірнеше рет қайталап бақылау нәтижесінде алынған $S_g^2\{y\}$ таңдамалы бағалаулары біртекті, ал $\sigma^2\{y_g\}$ дисперсиясы $M\{y_g\}$ күтімге тәуелді емес, яғни, оның факторлық кеңістіктің кез келген басқа нүктелерінде қайталап бақылау нәтижесінде алынған $\sigma^2\{y_g\}$ дисперсиясынан айырмашылығы жоқ (бірдей дәлдікпен қайталану). Бұл тәжірибелердің қайта орындалуының дисперсиясына тең болатын шулар дисперсиясының тұрақтылығын $\sigma_e^2=const$ білдіреді. Бұл факторлар өзгерген кезде шудың қарқындылығы өзгермейтінін білдіреді.

Әртүрлі жағдайларда өлшеулерді орындау кезінде өлшемдердің дәлдігін салыстыру мәселесі туындайды. Бұл жерде тәжірибелік мәліметтерді олардың дисперсиялары біртекті болғанда ғана салыстыруға болатынын атап өткен жөн. Бұл, жоғарыда айтылғандай, тәжірибелік мәліметтердің бірдей бас жиынға жататынын білдіреді. Еске салайық, дисперсиялардың біртектілігі салыстырылған дисперсиялар арасында берілген сенімділікпен барлық басқалардан асып түсетіндер, яғни үлкен қателікпен, жоқ екенін көрсетеді.

4. Тәуелсіз $x_1,...x_n$ басқарылатын факторларын өлшеу қателері y -ті анықтау қателерімен салыстырғанда кішігірім қателіктер болады. Яғни, $\bar{x}$ факторларының кездейсоқ өлшеу қателерінің тәжірибені қайта жасау дисперсиясына қосқан үлесі, e шуларды құрайтын басқа бақыланбайтын факторлардың әрекетімен салыстырғанда елеусіз болуы керек.

Бұл болжамды регрессия моделін алғаннан кейін ғана тексеруге болады, өйткені белгілі бір фактордың өлшеу қателігінің әсер ету дәрежесі қатенің шамасына ғана емес, сонымен бірге фактордың өзі y-ке қаншалықты әсер ететініне де байланысты. Оның әсері неғұрлым күшті болса, оны өлшеудегі қателік соғұрлым көп болады. Сондықтан, бұл алғышарт жай ғана болжамдалады, содан кейін ғана, егер табылған регрессия моделі қолдануға жарамаса, бұл x өлшеуіндегі елеулі қателердің болуына байланысты екенін талдауға кіріседі.

5. Базистік функциялардың баған векторлары сызықтық тәуелсіз болуы керек, яғни:

$$\sum_{g=1}^{N}[af_j(\bar{x}_g) + bf_i(\bar{x}_g)] \neq 0 \text{ при} \forall\ a \text{ и } b \neq 0.$$

Бұл талаптың орындалуы регрессия моделінің құрамындағы барлық β регрессия коэффициенттерінің бағаларын бөлек алу үшін қажет. Активті тәжірибеде бұл талап тәжірибе жоспарын сәйкесінші таңдау жолымен қамтамасыз етіледі. Пассивті тәжірибеде оның орындалуы есептеулер барысында тексерілуі керек. Қатаң сызықтық қатынастың орын алуы екіталай, бірақ бағандардың қатты корреляциясы болуы мүмкін. Бұл жағдай барлық регрессия коэффициенттерін бағалауға кедергі болмаса да, үлкен есептеу қателерінің пайда болуына әкелуі мүмкін. Қарастырылып отырған шарт регрессиялық модельге кіретін коэффициенттердің жалпы санын да шектейді: ол сынақтар санынан көп болмауы керек, яғни $d+1 \leq N$.

Регрессиялық талдау мыналарды қамтиды:

а) регрессиялық модельдің белгісіз β коэффициенттерін, сонымен қатар σ_e^2 тәжірибелерді қайталау дисперсиясын бағалау.

б) нәтижелерді статистикалық талдау (бөлек регрессия коэффициенттер бағаларының және регрессия моделдін өзінің дәлдігін, модельдің адекваттылығын анықтау, жұмыс қабілетін тексеру).

5.2 Регрессиялық модельдің параметрлерін бағалау

Көп өлшемді регрессиялық зерттеулерде көп факторлық кеңістіктің графикалық интерпретациясы жоқ. Мұндай жағдайда тәжірибелерді жүргізу кезінде зерттеуші барлық факторлар туралы және оларға тәуелді жауап беру функциясының күйі туралы аспаптар көрсеткіштерін жазып алады. Яғни, регрессиялық модель параметрлерінің нүктелік бағасын алудың бастапқы

ақпарат $x_1,...x_n$ факторларының және y жауап функцияның N бақылаулар мәндерінің матрицасы болып табылады:

$$\bar{x} = \left\| \begin{matrix} \bar{x}_1 \\ \bar{x}_2 \\ \dots \\ x_N \end{matrix} \right\| = \left\| \begin{matrix} x_{11} \dots x_{1n} \\ x_{21} \dots x_{2n} \\ \dots\dots\dots\dots \\ x_{N1} \dots x_{Nn} \end{matrix} \right\|; \quad y = \left\| \begin{matrix} y_1 \\ y_2 \\ \dots \\ y_N \end{matrix} \right\|,$$

мұндағы x_{ij} – i-ші сынақтағы j-ші фактордың мәні;

x_{ij} – шығыс параметрінің i-ші сынақтағы мәні.

Әрбір бақылаудағы барлық базистік функциялардың сандық мәндері де матрицалық түрде жазылады:

$$\bar{F} = \left\| \begin{matrix} \bar{f}_1(\bar{x}) \\ \bar{f}_2(\bar{x}) \\ \dots \\ \bar{f}_g(\bar{x}) \\ \dots \\ \bar{f}_N(\bar{x}) \end{matrix} \right\| = \left\| \begin{matrix} f_{10}\ f_{11}(\bar{x}) \dots f_{1j}\ (\bar{x}) \dots f_{1d}(\bar{x}) \\ f_{20}\ f_{21}(\bar{x}) \dots f_{2j}\ (\bar{x}) \dots f_{2d}(\bar{x}) \\ \dots\ \dots\ \dots\ \dots\ \dots\ \dots \\ f_{g0}\ f_{g1}(\bar{x}) \dots f_{gj}\ (\bar{x}) \dots f_{gd}(\bar{x}) \\ \dots\ \dots\ \dots\ \dots\ \dots\ \dots \\ f_{N0}\ f_{N1}(\bar{x}) \dots f_{Nj}\ (\bar{x}) \dots f_{Nd}(\bar{x}) \end{matrix} \right\|.$$

Осы ақпаратты пайдалана отырып, белгісіз $\beta_0, \dots\ \beta_d$ коэффициенттерін бағалау қажет, яғни осы коэффициенттердің *бағаларының* сәйкес мәндерін $\bar{b}=\{b_0, \dots\ b_d\}$ табу. Бұл бағалар - кездейсоқ шамалар, өйткені өлшеу нәтижелері кездейсоқ қателермен бірге жүреді. Критерийді таңдау бөгеттердің сипатына, яғни олардың статистикалық қасиеттеріне байланысты. Бөгеттердің қалыпты таралуы кезінде орташа квадраттық критерий ең үлкен дәлдікті береді, ал мәселе жай шешіледі (әдіс *ең кіші квадраттар әдісі* деп аталады). Бұл әдіске сәйкес, $\bar{b}$ мәндері модельдің көмегімен алынған жауап мәндерінің оның өлшенген мәндерінен квадраттық ауытқуларының қосындысын яғни келесі сумманы минимумдау шартынан табылады:

$$S = \sum_{g=1}^{N} [y_g - \varphi(x_i, \bar{\beta})]^2 \to min$$

немесе

$$S = \sum_{g=1}^{N} [y_g - (b_0 f_{g0} + b_1 f_{g1} + \dots + b_d f_{gd}]^2 \to min.$$

Соңғы өрнекте b_i коэффициенттері олардың нақты мәндері ретінде емес, берілген тәжірибелік материалға сәйкес келетін айнымалылар ретінде

қарастырылады. Көрсетілген квадраттар қосындысы минимумға жететін b_i мәндері регрессия коэффициенттерінің бағалары ретінде қабылданады. Квадраттардың қосындысын минимумдау кәдімгі әдіспен дербес туындыларды барлық b_0, b_1, ..., b_d: бойынша нөлге теңестіру арқылы жүзеге асырылады:

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = -2\sum_{g=1}^{N}[y_g - (b_0 f_{g0} + b_1 f_{g1} + \ldots + b_d f_{gd})] \cdot f_{g0} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \dfrac{\partial S}{\partial bd} = -2\sum_{g=1}^{N}[y_g - (b_0 f_{g0} + b_1 f_{g1} + \ldots + b_d f_{gd})] \cdot f_{gd} = 0. \end{cases}$$

Түрлендірулерден кейін сызықтық алгебралық теңдеулер жүйесі алынады:

$$\begin{cases} b_0\sum_{g=1}^{N} f_{g0}^2 \quad + b_1\sum_{g=1}^{N} f_{g0}\, f_{g1} + \cdots + b_d\sum_{g=1}^{N} f_{g0}\, f_{gd} = \sum_{g=1}^{N} y_g f_{g0}\,, \\ b_0\sum_{g=1}^{N} f_{g0} f_{g1} + b_1\sum_{g=1}^{N} f_{g1}^2 \quad + \cdots + b_d\sum_{g=1}^{N} f_{g1}\, f_{gd} = \sum_{g=1}^{N} y_g f_{g1}\,, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ b_0\sum_{g=1}^{N} f_{g0} f_{gd} + b_1\sum_{g=1}^{N} f_{g1}\, f_{gd} + \cdots + b_d\sum_{g=1}^{N} f_{gd}^2 \quad = \sum_{g=1}^{N} y_g f_{gd}. \end{cases}$$

оны шешу жолымен ізделінетін b_0, b_1,...b_d бағаларын табуға болады. Бұл жүйе *қалыпты теңдеулер жүйесі* деп аталады.

Қалыпты теңдеулер жүйесі матрицалық түрде келесідей болады:

$$(\bar{F}^T \cdot \bar{F}) \cdot \bar{b} = (\bar{F}^T \cdot y)$$

$\bar{F}^T \cdot \bar{F} = \overline{\Phi}$ осы жүйенің белгілі коэффициенттерінің матрицасы:

$$\bar{F}^T \cdot \bar{F} = \bar{\Phi} = \left\| \begin{matrix} \sum_{g=1}^{N} f_{g0}^2 & \sum_{g=1}^{N} f_{g0} f_{g1} \cdots \sum_{g=1}^{N} f_{g0} f_{gd} \\ \sum_{g=1}^{N} f_{g0} f_{g1} & \sum_{g=1}^{N} f_{g1}^2 \cdots \sum_{g=1}^{N} f_{g1} f_{gd} \\ \cdots\cdots\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots\cdots \\ \sum_{g=1}^{N} f_{g0} f_{gd} & \sum_{g=1}^{N} f_{g0} f_{gd} \cdots \sum_{g=1}^{N} f_{gd}^2 \end{matrix} \right\|.$$

Бұл матрица *Фишердің ақпараттық матрицасы* деп аталады және (d+1) жолдар мен *(d+1)* бағандардан тұрады. Осы матрицаның *j*-ші жолы мен *k*-ші бағанының қиылысында келесі қосынды орнатылған:

$$\sum_{g=1}^{N} f_{gj} f_{gk}.$$

Бұл матрица бас диагональға қатысты симметриялы (бұл белгілі бір регрессия моделі үшін қалыпты теңдеулер жүйесін құрастыруды жеңілдетеді).

Қалыпты теңдеулер жүйесінің жалғыз шешімі болуы үшін келесі $det\|\Phi\| \neq 0$ шарты орындалуы қажет және жеткілікті. Бұл шарт $\bar{F}$ матрицасының баған векторларының сызықтық тәуелсіздік шартын қанағат етеді (регрессиялық талдаудың алғы шарттарының бірі.

Қалыпты теңдеулер жүйесін кез келген белгілі әдіспен шешуге болады. Жалпы матрицалық түрде шешімді келесідей жазуға болады:

$$\bar{b} = \varphi^{-1} \cdot (\bar{F}^T \cdot y),$$

мұндағы φ^{-1} - φ матрицасына кері матрица.

C_{jk}– φ^{-1} матрицасының элементі болсын: $C_{jk} = \varphi_{kj}/|\varphi|$, мұндағы φ_{kj}- φ матрицасындағы f_{kj} элементіне алгебралық толықтаушысы. Сонда:

$$b_j = \sum_{k=0}^{d} C_{jk} (\sum_{g=1}^{N} y_g \cdot f_{gk}), j = 0, .., d.$$

Бұл өрнектен коэффициенттердің мәндері регрессия теңдеуіндегі мүшелер санына байланысты екенін көруге болады. Яғни, теңдеудегі мүшелер санының көбеюі немесе азаюы теңдеудегі барлық мүшелердің коэффициенттерінің мәндеріне әсер етеді. Регрессия коэффициенттерін

бағалаудағы мұндай анықталмағандық олардың физикалық интерпретациясын өте қиындатады.

Егер Фишер матрицасы диагоналды болса ($\bar{F}$ матрицасының бағандары жұптап ортогоналды), онда

$$C_{jl} = \sum_{g=1}^{N} f_{gj} \cdot f_{gl} = 0, j \neq l,$$

онда қалыпты теңдеулер жүйесі тәуелсіз теңдеулерге ыдырайды, ал коэффициенттер тәуелсіз түрде алынады:

$$b_j = \sum_{k=0}^{d} C_{jj} (\sum_{g=1}^{N} y_g f_{gk}), j = \overline{0, d}.$$

Осылайша, $\bar{b}$ коэффициенттерінің тәуелсіз бағаларын алу үшін $\bar{F}$ атрицасының бағандары сызықты тәуелсіз және ортогональды болуы керек. Сызықтық тәуелсіздік регрессиялық модельге кіретін коэффициенттердің жалпы санын шектейді: $d+1 \leq N$ (мысалы, үш коэффициент, екі фактор – кемінде төрт тәжірибе); ортогоналдық коэффициенттердің тәуелсіз бағаларын алуға мүмкіндік береді (регрессия теңдеуіндегі мүшелер санына қарамастан).

Мысалдар. Қарапайым регрессиялық модельдер үшін регрессия коэффициенттерін бағалау формулаларын жазайық.

1. Модель $y = \varphi(\bar{x}, \beta) = \beta x$ түрінде таңдалынған.

Ол үшін $f_i(x)$ базистік функциялар жүйесі және Осылайша, $\bar{b}$ коэффициенттерінің тәуелсіз бағаларын алу үшін $\bar{F}$ базистік функциялардың сандық мәндерінің матрицасы келесідей болады:

$$f_0(x) = 0; f_1(x) = x; N = 2,$$
$$F = \begin{Vmatrix} f_{10} & f_{11} \\ f_{20} & f_{21} \end{Vmatrix} = \begin{Vmatrix} 0 & x_1 \\ 0 & x_2 \end{Vmatrix}.$$

Сонда қалыпты теңдеулер жүйесі келесі түрде жазылады:

$$\begin{cases} b_0 \cdot 0 + b_1 \cdot 0 = y_1 \cdot 0 + y_2 \cdot 0 = 0, \\ b_0 \cdot 0 + b_1 \cdot (x_1^2 + x_2^2) = y_1 \cdot x_1 + y_2 \cdot x_2. \end{cases}$$

Осы жүйеден модельдің белгісіз коэффициентінің бағасын табамыз:

$$b = (\sum_{i=1}^{2} x_i y_i) / (\sum_{i=1}^{2} x_i^2).$$

Жалпы жағдайда, N сынақтарда аламыз:

$$b = (\sum_{i=1}^{N} x_i y_i)/(\sum_{i=1}^{N} x_i^2).$$

2. Келесі түрдегі модель үшін $y = \varphi(\bar{x}, \beta) = \beta_0 + \beta_1 x$

базистік функциялар жүйесі: $f_0(x) = 1; f_1(x) = x;$
матрица $F = \begin{Vmatrix} f_{10} & f_{11} \\ f_{20} & f_{21} \end{Vmatrix} = \begin{Vmatrix} 1 & x_1 \\ 1 & x_2 \end{Vmatrix},$
ал екі тәжірибе жағдайында қалыпты теңдеулер жүйесі:

$$\begin{cases} b_0 \cdot 2 + b_1 \cdot (x_1 + x_2) = y_1 + y_2, \\ b_0 \cdot (x_1 + x_2) + b_1 \cdot (x_1^2 + x_2^2) = y_1 \cdot x_1 + y_2 \cdot x_2. \end{cases}$$

Жалпы алғанда:

$$\begin{cases} b_0 \cdot N + b_1 \cdot \sum_{g=1}^{N} x_g = \sum_{g=1}^{N} y_g, \\ b_0 \cdot \sum_{g=1}^{N} x_g + b_1 \cdot \sum_{g=1}^{N} x_g^2 = \sum_{g=1}^{N} x_g y_g. \end{cases}$$

Бұл жүйені шешіп, b_0 және b_1 үшін өрнектерді жаза аламыз (коэффициенттерді есептеу өрнектері өте көлемді болғандықтан, бұл жерде көрсетілмеді).

3. $(\bar{x}, \beta) = \beta_0 + \beta_1(x_1 - \bar{x})$ модельді қарастырайық,
мұндағы $\bar{x} = \frac{1}{N}\sum_{g=1}^{N} x_g,$
базистік функциялар келесідей болады:

$$f_0(x) = 1; f_1(x) = x_1 - \bar{x},$$

Бұл сызықтық модель жоғарыдағы ұқсас, бірақ бұл жағдайда коэффициенттерді бағаларын алу мәселесі оңай шешіледі, ал модель коэффициенттері үшін келесі формулалар анықталады:

$$b_0 = (\sum_{g=1}^{N} y_g)/N = \bar{y};$$

$$b_1 = (\sum_{g=1}^{N}(x_g - \bar{x}) \cdot y_g) = (\sum_{g=1}^{N}(x_g - \bar{x})^2);$$

Осылайша, базистік функциялардың әртүрлі модификацияларын қолдана отырып және қажетті регрессиялық модельдің түрін өзгертпестен есептеулерді жеңілдетуге болады. Практика көрсеткендей, әрқашан орталықтандырылған $X - \bar{X}$ айнымалыларды қолдану қалаулы..

Коэффициенттер бағаларының қасиеттері.

Регрессиялық модельдің параметрлер бағаларының негізгі қасиеттерін келтірейік:

1. Бағалардың математикалық күтімі $M\{b_j\} = \beta_j, j = \overline{0, d},$ яғни b_j бағалары ығыспаған.

2. Коэффициенттер бағаларының дисперсиялары келесіге өрнекпен есептеледі $\sigma_{b_j}^2 = \sigma_e^2 \cdot C_{jj}$ және сызықты ығыспаған бағалар класында минималды болады (осыдан коэффиценттік бағаларынан басқа, тәжірибелер қайталанудағы дисперсиясын бағалау қажет болатыны анық).

3. $b_0,...,b_d$ бағалары $M\{\bar{b}\} = \bar{\beta}$ математикалық күтім векторы бар бірлескен *(d+1)* өлшемді таралуға бағынады.

Жоғарыда қарастырылған модельдер үшін коэффициенттерді анықтаудың дәлдігін бағалайық.

$\varphi(\bar{x}, \beta) = \beta x$ моделі үшін:

$$\sigma_b^2 = \sigma_e^2 / (\sum_{g=1}^{N} x_g^2).$$

Берілген *x* бойынша бұл формуланы σ_e^2 бөлшектеріндегі коэффициенттер бағаларының дисперсиясын анықтау үшін пайдалануға болады. Бірақ бұл қатынасты тәжірибені жоспарлау кезеңінде қолдануға болады, мысалы, σ_b^2 шамасын азайтатын *x* мәндерін таңдау үшін пайдалануға болады.

Осы формула бойынша *x* факторының мәндерінің берілген жиынында σ_e^2 бөлшектеріндегі коэффициенттер бағасының дисперсиясын анықтауға болады. Бірақ бұл қатынасты тәжірибені жоспарлау кезеңінде *x* факторының мәндерін таңдау үшін қолдануға болады, таңдалынған мәндер белгілі бір бақылау санында регрессияның *β* коэффициентін бағалаудың ең жақсы дәлдігін қамтамасыз ететіндей, мысалы, σ_b^2минималды дисперсиясы болатын бағалау.

Келесі модель

$$\varphi(\bar{x}, \beta) = \beta_0 + \beta_1(x_1 - \bar{x}), \bar{x} = \frac{1}{N}\sum_{g=1}^{N} x_g$$

үшін табуға болады:

$$\sigma_{b_0}^2 = \frac{\sigma_e^2}{N}; \sigma_{b_1}^2 = \sigma_e^2 / [\sum_{g=1}^{N} (x_g - \bar{x})^2)].$$

Бұл қатынастарды барлық N сынақтарда x факторының сандық мәндері берілген кезде коэффициенттерді бағалау дисперсиясын есептеу үшін тәжірибелер нәтижелерін өңдеуде де, бағалаудың дәлдігін қамтамасыз ету үшін осы мәндерді таңдау үшін тәжірибені жоспарлау кезеңінде де қолдануға болады:

- $\sigma_{b_0}^2$ шамасы от x-ке тәуелді емес, тек тәжірибе санына тәуелді;

- $\sigma_{b_1}^2$ шамасы от x-ке тәуелді - x-тің өзгеру диапазоны неғұрлым кең болса, соғұрлым кішірек $\sigma_{b_1}^2$ қол жеткізуге болады; онда нүктелерді жоспардың диапазонның шекарасына қойған дұрыс.

5.3 Регрессиялық модельді құру нәтижелерін статистикалық талдау

Статистикалық талдау нәтижесінде алынған регрессия моделінің сапасын анықтауға мүмкіндік береді: дәлдігін, адекваттылығын және жұмысқа қабілеттілігін.

5.3.1 Тәжірибелер қайтадан оңдірілуінің дисперсияны бағалау.

b_j регрессия коэффициенттерін бағалаудан басқа, жалпы жағдайда σ_e^2 тәжірибелер қайтадан өндірілуінің дисперсиясын (дисперсия воспроизводимости) бағалау қажет.

Егер регрессиялық модель зерттеу объектісіне адекватты екені алдын ала белгілі болса, онда σ_e^2 дисперсияның бағасы ретінде қалдық дисперсия деп аталатын дисперсияны пайдалануға болады:

$$S_{\text{қалдық}}^2 = \frac{1}{N-(d+1)} \sum_{g=1}^{N} (y_g - \hat{y}_g)^2,$$

мұндағы $\hat{y}_g$ – нөмірі g тәжірибедегі регрессия теңдеуімен болжанған жауап мәні:

$$\hat{y}_g = b_0 f_0(\bar{x}_g) + b_1 f_1(\bar{x}_g) + \ldots + b_d f_d(\bar{x}_g).$$

Шынында да, адекватты модель үшін g-ші сынақтағы y_g нақты мен $\hat{y}_g$ теңдеуімен болжанған мәндері арасындағы айырмашылықтардың жалғыз себебі - шудың әсері болуы мүмкін. Сондықтан σ_e^2 шаманың бағасы $S_{\text{қалдық}}^2$ болып табылады:

$$S_e^2 = S_{ост}^2; \nu = N - (d+1), \nu - \text{ еркіндік дәрежелерінің саны.}$$

Таңдаманың элементтеріне тәуелді және дисперсия формуласына кіретін шама *байланыс* деп аталатынын еске салайық. Таңдаманың N көлемі мен байланыстар саны арасындағы айырмашылық *еркіндік дәрежесінің саны* деп аталады.

Егер адекватты модель түрі белгісіз болса, онда біз тәжірибе қайтадан өндірілуін тексереміз, яғни регрессиялық талдаудың алғышарттарының ішіндегі таңдамалы S_g^2 дисперсияларының біртектілігін тексереміз. Басқаша айтқанда, $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N$ нүктелердегі сынақтарда бас жиын дисперсияларының теңдігі $\sigma^2\{y_1\} = \sigma^2\{y_2\} = \dots = \sigma^2\{y_N\}$ туралы гипотезаны тексергіміз келеді.

Дисперсиялардың бағалары белгілі формула бойынша табылады:

$$S_g^2 = \frac{1}{m-1}\sum_{k=1}^{m}(y_{gk} - \bar{y}_g)^2, \nu_{вос} = m - 1.$$

S_g^2 дисперсияларының біртектілігі туралы гипотезаны тексеру үшін Кохран критерийі қолданылады (кестеленген функция); бұл критерий максималды дисперсия бағасының барлық салыстырылатын дисперсиялар бағаларының қосындысына қатынасының таралу заңына негізделген:

$$G_{\text{кр}} = \max\{(y)\}/(\sum_{g=1}^{N} S_g^2\,(y)), \nu_{өндірілу} = m - 1, \nu_2 = N;$$

Егер тәжірибелік мәліметтерден есептелген G критерийінің мәні кестеден алынған сыни (критикалық) мәннен аз болса (алымның еркіндік дәрежелер саны ν_1, бөлімнің - ν_2 және мысалы, q=5% маңыздылық деңгейімен), онда біртектілік гипотезасы тәжірибелер нәтижелеріне сәйкес келеді және бас жиын дисперсияның $\sigma^2\{y\}$ең жақсы бағасы болады:

$$S_{вос}^2\{y\} = \frac{1}{N}\sum_{g=1}^{N} S_g^2\,\{y\} = \frac{1}{N(m-1)}\sum_{g=1}^{N}\sum_{k=1}^{m}(y_{gk} - \bar{y}_g)^2$$

$\nu = N(m-1)$ еркіндік дәрежелерінің санымен.

Егер S_e^2 бағаларын алу үшін арнайы тәжірибелер (L сынақтар) қолданылса, оның нәтижелері модельді алу үшін пайдаланылмаған болса, онда:

$$S_e^2\{y\} = \frac{1}{L-1}\sum_{l=1}^{L}(y_l - \bar{y})^2, \bar{y} = \frac{1}{L}\sum_{l=1}^{L} y_l.$$

Егер тәжірибелер қайталану дисперсиясы теріс нәтиже берсе, онда басқарылмайтын және бақыланбайтын факторларға қатысты үлкен шудың болуына байланысты тәжірибелер басқарылатын факторларға қатысты бұрынғы қателік деңгейінде қайталанбайтындықтан, параллельді тәжірибелер санын көбейту қажет.

Дәлдіктің интегралдық сипаттамалары азырақ, мысалы, регрессия коэффициенттер бағаларының $\sigma_{b_j}^2$дисперсиялары неғұрлым кіші болса, соғұрлым есепті шешу нәтижесінде алынған $\bar{b}$ коэффициенттерінің (және $\hat{y}$ жауап беру функцияның болжамдық мәндерінің) бағалары қолайлы болады. Өңдеу әдісі (ең кіші квадраттар әдісі) берілген, базистік функциялар жүйесі алдын-ала таңдалған, сондықтан бұл сипаттамалар барлық сынақтардағы кіріс мәндеріне ғана тәуелді. Пассивті тәжірибелерде бұл тәуелділіктен пайда табу мүмкін емес. Ал, егерде тәжірибе активті болған кезде әрбір жеке сынақта $\bar{x}$ мәндерінің қандай комбинациясында тәжірибелерді жүргізу керектігін болжауға болады, яғни тәжірибе жоспарын алдын ала құрастыруға болады.

Параметрлері бойынша сызықтық болып табылатын қарастырылып отырған регрессия модельдері үшін жоспарлар жауап функцияның өлшеулер нәтижелеріне тәуелді емес, сондықтан оларды алдын ала табуға болады.

5.3.2 Регрессия коэффициенттерінің маңыздылығын тексеру.

Регрессия коэффициенттерінің $\bar{b}$ бағаларын анықтағаннан кейін олардың маңыздылығы туралы гипотезаларды тексеру қажет. Тексеру нәтижесінде регрессия коэффициенттерін бағаларының нөлден айырмашылығы статистикалық маңызды ма немесе маңызды емес пе анықталады. Яғни, b_i коэффициентінің нөлден айырмашылығы тек қана кездейсоқ жағдайларға байланысты ма, бөгеттер әсерінен бе немесе бұл айырмашылық кездейсоқ емес пе және теориялық модельде сәйкес $\beta_j \neq 0$ регрессия коэффициенті болуымен туындады ма. Регрессия коэффициенттерінің маңыздылығы көбінесе әрбір b_i баға үшін кезекпен бөлек тексеріледі.

Сонымен, әрбір коэффициент үшін $\beta=0$ нөл- гипотеза тексеріледі.

Маңыздылықты тексеру келесі түрдегі гипотезаларын тексеруден тұрады:

$$\begin{cases} H_0: \beta_j = 0, \\ H_1: \beta_j \neq 0, j = 0, .., d. \end{cases}$$

Мұндай гипотезалар Стьюдент критерийі арқылы тексеріледі, оның эмпирикалық мәні:

$$t_j = \frac{|b_j|}{S_{b_j}},$$

мұнда

$S_{b_j}^2 = S_e^2 \cdot C_{jj} = \frac{1}{Nm} S_e^2, j = 0,..,d$ - b_j коэффициентінің бағасының дисперсиясы;

N - тәжірибе жүргізілетін факторлық кеңістіктегі нүктелер саны.

m- осы нүктелердегі параллель тәжірибелер саны.

Егер t_j параметрінің табылған мәні берілген q маңыздылық деңгейінде (әдетте 5%), $\nu_{зн} = N$ еркіндік дәрежелерінің саны үшін t-таралу кестесінен анықталған $t_{кр}$ мәнінен асып кетсе, онда тексерілетін нөлдік гипотеза қабылданбайды және b_j коэффициентінің сәйкес бағасы статистикалық маңызды деп танылады; коэффициент сақталуы керек. Бұл бағаның нөлден айырмашылығы кездейсоқ емес екенін білдіреді; коэффициент регрессия теңдеуінде сақталуы керек.

Кері жағдайда нөл-гипотеза жоққа шығарылмайды және b_j бағасы статистикалық маңызды емес болып саналады. Оның нөлден айырмашылығын таза кездейсоқ жағдайлармен түсіндіруге болады. Бұл жағдайда шешім қабылдау қажет: маңызды коэффициентті қалдыру немесе алып тастау керек пе. Формальды түрде бұл коэффициентті теңдеуден алып тастау керек және ол теңдеуге кірмеуі керек, бірақ практикада шешім талдаудан кейін қабылданады, әсіресе t_j критикалық мәнге жақын болса. Өйткені, бір немесе басқа коэффициенттің маңызды еместігі себептері түбегейлі әртүрлі болуы мүмкін. Олар регрессия теңдеуінде сәйкес коэффициенттің жоқтығын, яғни оның бағасы маңызды емес, эмпирикалық теңдеуден алып тастау заңды екенін көрсетуі мүмкін. Дегенмен, коэффициенттің маңызды еместігі осы нақты тәжірибе барысында осы коэффициенттің әсері бөгеттер себептерінен көрінбеуімен байланысты болуы мүмкін.

Сондықтан шешім қабылдау үшін оны модельде сақтаудың пайдасына қосымша аргументтер бар-жоғын талдау керек, әсіресе критерий мәні кестелік критикалық мәніне жақын болса. Мамандардың бұл фактор қажет деген сенімі қалдыруға дәлел болуы мүмкін. Егер базистік функциялардың бастапқы жүйесі дұрыс таңдалса, онда t-критерийі бойынша коэффициенттері нөлден аздап ерекшеленетіндерді алып тастау қалған регрессия коэффициенттерін бағаларының ығысуына әкеледі.

Регрессия коэффициентінің b_j бағасының статистикалық маңызды еместігі келесі себептерге байланысты болуы мүмкін:

1) белгілі j-ші фактордың y жауап функциямен функционалдық байланысы жоқ, яғни

2) $\bar{x}_0$-дің x_{j0} базалық режимінің деңгейі жауап функциясының x_j факторы бойынша дербес экстремум нүктесінде болады, сонда $\beta_j = \frac{\partial y}{\partial x_j} = 0$;

3) Δx_j вариация аралығы шағын таңдалған;

4) басқарылмайтын және бақыланбайтын факторлардың әсерінен тәжірибенің қайталануындағы қателігі үлкен болады.

Коэффициенттер бағаларының маңыздылығын кезекпен тексеру процедурасының кемшіліктері бар. Егер φ матрицасы диагоналды болмаса,

онда, біріншіден, t-критерийдің қуаты төмендейді және ол басқа жағдайда қалдырылатын коэффициенттерді маңызды емес ретінде алып тастай алады. Барлық коэффициенттер маңызды емес болып қалуы мүмкін. Екіншіден, егер коэффиценттердің маңыздылығын тексеру реті өзгертілсе, регрессия моделінің соңғы түрінде екіұштылық пайда болуы мүмкін.

Бұл жағдайды жою үшін келесі әрекеттер орындалады:

- барлық b_j үшін t_j статистикалар есептеледі;

- оларды өсу бойынша реттейді $t_{(0)} = \min t_j \leq t_{(1)} \leq t_{(2)} \leq t_{(d)} = max\ t_j$;

- $t_{(0)}$ шамасын критикалық мәнімен салыстырады: егер $t_{(0)} > t_{\text{кр}}$ болса, онда барлық коэффициенттердің маңыздылығы туралы шешім қабылданады; егер $t_{(0)} < t_{\text{кр}}$ болса, онда сәйкесінші b_j коэффициенті маңызды емес.

Егер коэффициенттің маңызды емес болуына қарамастан, ол теңдеуде қалдырылса, онда кестелік мәнімен келесі статистика $t_{(1)}$ салыстырылады; егер коэффициент алып тасталынатын болса, онда процедура аяқталды.

5.3.3 Регрессиялық модельдің адекваттылығын тексеру.

Модельді мұндай тексеру модель теңдеуінің түрі дұрыс таңдалғанын немесе таңдалмағанын анықтау үшін жүргізіледі. Тәжірибелік мәліметтер негізінде құрастырылған математикалық сипаттаудың адекваттылығы туралы гипотезаны тексеру үшін факторлардың бірдей g-ші нүктелеріндегі бақылаулар нәтижелерінен алынған $\hat{y}_g$ жауап беру функцияның мәндерінің регрессия теңдеуімен болжанған $\hat{y}_g$ жауап мәндерінен ауытқуын бағалау жеткілікті. Сынауда маңызды коэффициенттері бар модель қолданылады. Бұл тексеру S_e^2 шулар дисперсиясының бағасы алынса және $N>d^*(d$ – модельде қалдырылған коэффициенттер саны) шарты орындалса мүмкін болады. Шынында да, егер $N=d^*=d+1$ болса, онда адекваттылықты тексерудің мәні жоқ, өйткені бұл жағдайда $y_g = \bar{y}_g, g = 1,..,N,$ яғни модельге сәйкес келетін жауап функцияның беті дәл барлық тәжірибелік нүктелер арқылы өтеді. Бұл жағдайда математикалық сипаттаманың адекваттылығы туралы гипотезаны тексеру үшін еркіндік дәрежелері қалмайды.

Нақты жауап функциясын бағалайтын регрессия теңдеуіне жақын болатын бақылау нәтижелерінің шашырауын адекваттылық дисперсиясының көмегімен сипаттауға болады:

$$S_{ад}^2 = \frac{m}{N-d}\sum_{g=1}^{N}(\bar{y}_g - \hat{y}_g)^2,$$

мұндағы d – жуықтаушы көпмүшенің мүшелерінің саны. Адекваттылықтың дисперсиясы $\nu_{зн} = N - d$ еркіндік дәрежесінің санымен анықталады.

Адекваттылықты тексеру идеясы келесідей. Адекватты модель үшін қалдық дисперсиясы (адекваттылық дисперсиясы) шу дисперсиясының бағалауы болып табылады σ_e^2 ($S_{\text{қалдық}}^2 = S_{ад}^2 = S_e^2$). Егер модель адекватты болмаса, онда $S_{қалдық}^2 = S_{ад}^2$ шамасы σ_e^2 бірге тағы адекватты еместікке байланысты кейбір құрамдас компонентасының бағасы болады. Яғни $S_{қалдық}^2$ шамасын шу дисперсиясының σ_e^2 кейбір бағалауымен S_e^2 салыстыру қажет. Адекватты модельде олардың арасындағы айырмашылық кездейсоқ сан болуы керек; адекватты емес модельде $S_{қалдық}^2$ шамасы $S_e^2 2$ мәнінен айтарлықтай үлкен болуы керек. Бұл байланыс *Фишердің F-критерийі* арқылы тексеріледі:

$$F = \frac{S_{қалдық}^2}{S_e^2}.$$

Фишер критерийі екі $S_{ад}^2$ және S_e^2 таңдамалы дисперсияларының біртектілігі туралы гипотезаны тексеруге мүмкіндік береді. Егер бақылаулар нәтижелері бойынша есептелген F критерийінің эмпирикалық мәні берілген маңыздылық деңгейде $q_{ад,}$ сәйкес еркіндік дәрежелері үшін $\nu_1 = N - d, \nu_2 = \nu_{зн} = N(m-1)$ Фишер критерийінің кестелерінен табылған критикалық мәнінен кіші болса ($F \leq F_{кр}$), онда адекваттылық туралы гипотезасы жоққа шығарылмайды. Кері жағдайда гипотеза жоққа шығарылады, ал математикалық сипаттама адекватты емес деп саналады.

Егер адекваттылық туралы гипотеза жоққа шығарылса, математикалық сипаттаудың күрделі түріне көшу немесе мүмкін болса, Δx_j вариациялау интервалын азайтып, тәжірибелерді жүргізу қажет. Вариациялау аралығының максималды мәні вариациялау аймағындағы объектінің адекватты сипатталуының шартымен анықталады. Егер математикалық модель үлкен вариациялау интервалдары үшін адекватты болмаса, онда коэффициенттерді анықтауда жүйелі қателер орын алады, оларды азайту үшін вариациялау диапазонын тарылту қажет. Дегенмен, вариациялау интервалының азаюымен бірқатар жаңа қиындықтар туындайды: бөгеттің пайдалы сигналға қатынасы артады, бұл шудың фонында пайдалы сигналды оқшаулау үшін параллель тәжірибелер санын көбейту қажеттілігіне әкеледі, яғни Δx_j шамаларына тікелей тәуелді болатын коэффициенттердің b_j бағалаларының абсолютті мәндері төмендейді, ал коэффициенттердің бағалары статистикалық маңызды емес болуы мүмкін.

Вариациялау аралығын таңдау үшін алдын ала тәжірибелер жүргізіледі. Вариациялау аралығын факторлардың рұқсат етілген вариациялау диапазонының 0,05-0,3 тең таңдауға болады, яғни вариациялау диапазоны барлық диапазонның шамамен 10-60% құрайды. Вариацияның бастапқы нүктесі (базалық нүктесі) мүмкіндігінше объектінің математикалық сипаттамасы ізделетін факторлық кеңістік аймағының ортасында жақын таңдалады.

5.3.4 Модельдің жұмысқа қабілеттігін талдау.

Осындай талдаудың көмегімен нақты мәселені шешу үшін регрессия моделін қолданудың практикалық мүмкіндігі анықталады (жауап функциясының мәнін болжау, объектіні басқару және т.б.). Кейде төмен дәлдікті болғандықтан адекватты модельдер пайдасыз болады, сондықтан осындай талдау қажет,.

Талдау бірнеше процедураларды қамтиды:

а) қалдықтарды зерттеу – регрессия теңдеуімен болжанған жауап функцияның мәндерінің байқалған айырмашылығын l_g=y_g-y_g;

Егер регрессия моделі дұрыс таңдалса, яғни ол жауап беру функциясына адекватты болса, онда қалдық шу әрекетінің көрінісі болып табылады. Шудың қасиеттері туралы белгілі бір болжамдар жасалғандықтан, енді қалдық қасиеттерінің бұл болжамдарға қайшы келетін-келмейтінін білуге болады. Мұндай талдау графикалық түрде жүзеге асырылады (l_g шаманың уақытқа, y_g, x_g және т.б. тәуелділігі). Қалдықтарды қарастыра отырып, өлшеулер жүргізілген жерде болжау қателігінің қандай екенін түсінуге болады, қате өлшемі қанағаттандыра ма, факторлардың қандай комбинацияларында қателер үлкенірек және т.б.

Қалдықтарды зерттеу бақылаулар саны коэффициенттер санынан көп болған жағдайда жемісті болады. Егер бұлай болмаса, онда модельдің сапасын тексеру үшін қосымша тәжірибелер жүргізген жөн;

б) *детерминация коэффициентін* есептеу. Бұл коэффициент тәжірибелік жауап мәндерінің жалпы шашырауының оның орташа мәніне қатысты регрессия тәуелділігіне байланысты қандай үлесте екенін көрсетеді:

$$R^2 = \left[\sum_{g=1}^{N}(y_g - \bar{y}_g)^2\right] / [\sum_{g=1}^{N}(y_g - \hat{y}_g)^2].$$

Бұл коэффициент келесідей өзгереді: $0 \leq R^2 \leq 1$.

$R^2 = 0$ болғанда, жауап беті y толығымен кездейсоқ әсерлермен анықталады, факторлардың әсері табылған жоқ; $R^2=1$ кезінде регрессия қисығы барлық тәжірибе нүктелері арқылы өтеді.

R не ғұрлым бірге жақын болған сайын, регрессия теңдеуі y өзгерісін соғұрлым жақсы сипаттайды, ал кіші R әрқашан теңдеудің төмен дәлдігін көрсетеді. Бұл қорытынды $N>d+1$ шартында жарамды, яғни. сынақтар саны модель параметрлерінің санымен салыстырғанда жеткілікті үлкен болса. Модель параметрлерінің саны бақылаулар санына жақындаған сайын квадраттардың қалдық қосындысы азаяды және нөлге жақын болуы мүмкін. Детерминация коэффициентінің мәндерінің кейбір төменгі шегін орнатуға болады, одан жоғары мәндері үшін регрессия теңдеуі болжау мақсаттары немесе басқа қолданбалар үшін жеткілікті практикалық қызығушылық тудырады. Егер детерминация коэффициентінің шекаралық мәндері келесі

шарттан табылатын болса: регрессия теңдеуі бойынша болжау қатесі орташа жауап мәні бойынша болжаудың қатесінен (факторларды есепке алмағанда), екі есе аз болса, онда $R^2_{min}\approx 0,75$. $R^2<0,75$ үшін регрессия теңдеуінің жұмысқа қабілеттілігі жоқ деуге болады.

5.4 Бақылау сұрақтары

5.4.1 Регрессия моделін анықтаңыз.

5.4.2 Базисті функциялар жүйесі дегеніміз не?

5.4.3 Кездейсоқ e құрастырушылардың статистикалық қасиеттеріне қатысты регрессиялық талдаудың алғышарттарын атаңыз.

5.4.5 Регрессиялық модельдің параметрлік бағаларын анықтау үшін бастапқы ақпарат қалай ұсынылады?

5.4.6 F^{-}матрицасының компоненттерін түсіндіріңіз.

5.4.7 Қалыпты теңдеулер жүйесі қалай құрылады?

5.4.8 Фишер ақпарат матрицасын анықтаңыз.

5.4.9 Регрессиялық модель параметрлеріні бағалары қалай анықталады?

5.4.10 Модельдің коэффициенттер бағаларының қасиеттерін беріңіз.

5.4.11 Тәжірибелерді қателердің бұрынғы деңгейінде қайталану мүмкіндігін қалай тексеруге болады?

5.4.12 Қандай жағдайларда тәжірибенің қайталану талабы сақталмайды және бұл жағдайда не істеу керек?

5.4.13 Коэффициенттер «маңыздылығы» түсінігін түсіндіріңіз. Регрессия коэффициенттерінін бағалар маңыздылығы қалай тексеріледі?

5.4.14 Қандай жағдайларда регрессия коэффициенттерінің бағалары маңызды емес және бұл шарттарды қалай жоюға болады?

5.4.15 Адекватты тәжірибелік модель анықтамасын беріңіз. Математикалық модельдің адекваттылығын қалай тексеруге болады?

5.4.16 Қандай жағдайларда математикалық модельдің адекваттылық талабы орындалмайды және бұл жағдайда не істеу керек?

5.4.17 Қай модель жұмысқа жарамды? Модельдің жұмысқа қабілеттілігін қалай тексеруге болады?

6 Регрессиялық талдаудың жоспарлары

6.1 Тәжірибелерді жоспарлаудың негізгі принциптері

Дәлдіктің интегралдық сипаттамалары неғұрлым кішірек болса, мысалы, регрессия коэффициенттері бағаларының дисперсиясы $\sigma^2_{b_j}$, соғұрлым регрессиялық модельді құрудың есебін шешу нәтижесінде алынған коэффициенттерінің $\bar{b}$ бағалары (және $\hat{y}$ жауап функцияның болжамдық мәндері) қолайлы болады. Регрессиялық талдаудың негізгі алғышарттары орындалған кезде қарастырылып отырған параметрлері бойынша сызықты модельдер үшін осындай сипаттамалары базистік функциялардың векторына,

барлық сынақтардағы факторлардың мәндеріне, сондай-ақ мәліметтерді өңдеу әдісіне тәуелді. Өңдеу әдісі (ең кіші квадраттар әдісі) берілген, базистік функциялар жүйесі таңдалған болғандықтан, бұл сипаттамалар барлық сынақтардағы байқалатын кіріс мәндерге ғана тәуелді. Пасивті тәжірибеде бұл тәуелділіктен пайда табу мүмкін емес. Тәжірибелер активті болған кезде әрбір жеке тәжірибені $\bar{x}$ мәндерінің қандай комбинациясында жүргізу керектігін болжауға, яғни тәжірибе жоспарын алдын ала құрастыруға болады. Онда әртүрлі жоспарлар олардың негізінде құрылған регрессиялық модельдердің дәлдігінің сандық сипаттамалары тұрғысынан айтарлықтай ерекшеленуі мүмкін екендігі бірден анықталады. Басқа сөзбен айтқанда, белгілі бір мағынада тәжірибенің оңтайлы жоспарын таңдау қажеттілігі туындайды.

Тәжірибелерді жоспарлаудың негізгі принциптері тәжірибелерді орындаудың тиімділігін арттыруға, яғни сынақтардың минимумы санында максималды ақпарат алуға бағытталған және келесілер болады:

1. *Ықтимал күйлерді толық санаудан бас тарту*. Жауап беру функциясының қасиеттері туралы толық ақпаратты алу үшін жоспарлау аймағының барлық нүктелерінде шексіз көп тәжірибе қажет. Әйтпесе, жауап функция бетінің кейбір ерекшеліктерін әрқашан жоғалту мүмкіндігі бар. Бұл жағдай практикалық орындалмайды. Сондықтан факторлар мәндерінің дискретті торы орнатылады және әрбір фактор үшін деңгейлердің белгіленген саны таңдалады. Нәтижесінде жауап беру функциясының кейбір ерекшеліктері анықталмайды. Тәжірибеде мұндай қатенің қаупі әрқашан болады. Іс жүзінде деңгейлер мен факторлардың санын көбейту қажет, бұл тәжірибелердің жалпы санының өсуіне әкеледі. Тәжірибелер жоспарлау теориясында кіріс күйлерін толық санаудан бас тартады. Практикада әдетте жауап функциясының түрін таңдау арқылы және әрбір фактор үшін вариация деңгейлерінің санын осы функциямен байланысты таңдау орнатылады. Бұл функция келесі принцип арқылы таңдалады.

2. *Математикалық модельдің біртіндеп күрделену принципі* (тізбекті жоспарлау принципі). Ақпарат болмаған жағдайда күрделі модельді бірден құрудың мағынасы жоқ, ең қарапайымынан бастау керек. Егер модель жарамсыз болса, келесі кезең қажет: күрделі модельді алуға мүмкіндік беретін жаңа қосымша тәжірибелерді орындау және т.б., жеткілікті жақсы деп саналатын модель алынғанша. Кеңінен қолданылатын көпмүшелік модельдерге жүгінсек, сызықтық модельден бастау керек; сәтсіздікке ұшыраған жағдайда квадраттық модельдерді құруға көшу және т.б.

Жоғарыда аталған әдіс тізбектелген тәжірибелердің жалпы принципінің бөлігі болып табылады: әрбір қадамнан кейін нәтижелер талданады (нәтижелердің сапасын тексеру), содан кейін одан әрі іс-әрекеттер туралы шешім қабылданады. Тәжірибелерді жоспарлау теориясында тексеру үшін шудың ықтималдық қасиеттерінен туындайтын әртүрлі статистикалық процедуралар қолданылады.

3. *Шумен салыстыру принципі*. Кездейсоқ кедергі маңызды болса, күрделі модельді алуға күш жұмсау орынды ма? Мысалы, үлкен кездейсоқ

қателікпен өлшенген айнымалыны өлшеу үшін жоғары дәлдікті аспаптарды қосу ысырапшылдық. Яғни, алынған модельдің дәлдігі кездейсоқ шудың қарқындылығымен салыстырмалы болуы керек. Шу деңгейі неғұрлым төмен болса, модель соғұлым дәл (демек, күрделі) болуы керек; және шу деңгейі неғұрлым жоғары болса, соғұрлым қарапайым және дәлдігі төмендеу модель жұмыс істейді деп күтуге болады. Көптеген нақты объектілер шудың жоғары деңгейімен сипатталатындықтан, көпмүшелік регрессия модельдері кеңінен таралған. Теориялық тұрғыдан алғанда бұл модельдер дифференциалдық теңдеулер сияқты теориялық модельдерге қарағанда мағынасы төмен. Бірақ практикалық тұрғыдан алғанда, олар өте тиімді, кейде күрделі объектілерді зерттеудің жалғыз құралы болып табылады.

4. *Рандомизация принципі* (кездейсоқтыққа келтіру принципі). Бұл принцип бойынша жүйелік, есепке алынбайтын бөгеттердің әсерін есепке алудың мүмкіншілігі болмағандықтан осы бөгеттер әсерін кездейсоқ (рандомды) жасауға мүмкіндік беретіндей тәжірибені ұйымдастырады. Яғни кездейсоқ емес айнымалылардың әсерін есепке ала алмайтын болғандықтан, тәжірибелерде кездейсоқ жағдайды жасанды түрде орнатамыз. Мысалы, тәжірибелік қондырғыны орнатудағы кейбір қателіктерге байланысты жүйелі қателер болуы мүмкін. Сонда егер сегіз тәжірибе сериясында алғашқы төрт тәжірибе бірінші фактордың төменгі деңгейінде жүргізілсе, онда қателік барлық төрт тәжірибеде бірдей болады (егер тәжірибелер қатарынан жүргізілсе).

5. *Тәжірибені жоспарлаудың оңтайлы принципі* - Тәжірибелерді жоспарлау теориясында орталық болып табылады. Жоспар белгілі бір алдын ала таңдалған оңтайлылық критерийі бойынша кейбір оңтайлы қасиеттерге ие болуы керек. Критерийлер шешілетін мәселенің түріне байланысты әртүрлі тәсілдермен тұжырымдалуы мүмкін, бірақ «*аз тәжірибе – көп ақпарат*» принципі сақталады.

Осылайша, жоспардың тиімділігінің таңдалған көрсеткішінің экстремалды мәндеріне сәйкес келетін жоспар қажет. Бұл көрсеткіштер регрессия жоспарларының *оңтайлылық критерийлері* деп аталады.

6.2 Жоспарлардың оңтайлық критерийлері

Оңтайлық критерийлерінің екі тобы бар: бірінші топқа регрессия коэффициенттерін бағалаудың дәлдігіне байланысты критерийлер жатады; екінші топ эмпирикалық регрессия теңдеуінің болжамдық қасиеттеріне қатысты критерийлерден тұрады. Екі топтың бірнеше критерийлерін қарастыайық:

1-ші топ) Ең жиі қолданылатын критерий *D-оңтайлылық критерийі*: ε* жоспары *D-оптималды* деп аталады, егер

$$det\,\Phi\,(\varepsilon^*) = \max_{\varepsilon\in\{\varepsilon\}} det\,\Phi\,(\varepsilon),$$

мұндағы $\{\varepsilon\}$ - берілген жоспарлау аймағынан алынған жоспарлар жиынтығы.

D - оңтайлы жоспар Φ матрицасының анықтауышының мәнін максималдайды, яғни регрессия коэффициенттерінің бағаларының жалпыланған дисперсиясын азайтады. Геометриялық тұрғыдан алғанда, бұл осы бағалардың шашырау эллипсоидының көлемін азайтуды білдіреді.

Ортогоналдылық критерийі. Егер Φ (және, демек, Φ^{-1}) матрицасы диагональ болса, яғни $\bar{F}$матрицаның бағандары жұптық ортогональ векторлар болса, ε^* жоспары *ортогонал* деп аталады. Φ матрицасының диагональдық элементтерінің берілген мәндері үшін регрессия коэффициенттерінің бағаларының дисперсиялары ε^* ортогональды жоспары үшін минималды болатыны дәлелденген. Регрессия коэффициенттерінің бағалары тәуелсіз табылады, бұл олардың талдауын жеңілдетеді және барлық есептеулерді күрт қысқартады.

2- ші топ) Егер де келесі шарт орындалса

$$\max_{x\in H}\sigma^2\{\hat{y}(\bar{x},\varepsilon^*)\} = \min_{\varepsilon\in\{\varepsilon\}}\max_{x\in H}\sigma^2\{\hat{y}(\bar{x},\varepsilon)\},$$

жоспар *G – оңтайлы* деп аталады.

Яғни, *G*-оңтайлы жоспар *H* аймағында регрессия теңдеуі бойынша модель шығысын болжаудың максималды дисперсиясын минимумдайды.

Ротатабельдік критерийі. Егер осы жоспарды қолдану арқылы алынған регрессия теңдеуі бойынша жауап функцияның болжамының дисперсиясы тәжірибелер ортасынан белгіленген қашықтықта тұрақты болса, ε^* жоспары ротатабельді деп аталады, яғни

$$\sigma^2\{y(\bar{x},\varepsilon^*)\} = h(\rho),$$

мұндағы $h(\rho)$ – жоспарлау орталығынан ρ арақашықтықтың кейбір функциясы: $\rho = \sqrt{x_i^2}$.

Ротатабельді жоспарлау үшін тәжірибелер ортасынан бірдей қашықтықта болжау дәлдігі бірдей болады және $y(\bar{x})$ есептелетін $\bar{x}$ нүктелері қандай қашықтықта болса да.

Осы критерийлерден басқа да жоспарлардың кейбір қасиеттері (қанықтығы және композициялық) де қажетті.

Егер бақылаулар саны регрессия моделінің белгісіз параметрлерінің санына тең болса, жоспар *қаныққан* деп аталады. Кез келген жоспардың қаныққанға жақын болғаны жөн, әсіресе бастапқы кезеңде, аз шығынмен объект туралы жуықтаған түсінік алу қажет болғанда..

Егер жоспардың спектрі құрамдас бөлігі ретінде қарапайым модельді құру кезінде жүзеге асырылған жоспар спектрінің нүктелерін орнатылса, жоспар *композициялық* деп аталады, яғни, егер алдыңғы кезеңде алынған

модель қанағаттырмаса және оны жақсарту қажет болса, мысалы, k-ші дәрежелі көпмүшеден ($k + 1$)-ші дәрежелі көпмүшеге өтсек, онда бірінші кезеңнің спектрінің нүктелері әрі қарай қолданылады. Жоспардың композициялық қасиеті – модельдің біртіндеп күрделену принципін практикалық жүзеге асыру тиімділігінің маңызды шарты болып табылады.

Егер жоспардың центрі координаталар басында орналасса, жоспар *орталық* деп аталады.

Модельдің коэффициенттерін есептеудің қарапайымдылығы, жоспар нүктелерін орналастырудағы симметрия және айнымалылардың әрқайсысы үшін вариациялау деңгейлерінің ең аз саны да қолайлы.

Параметрлері бойынша сызықты болып табылатын қарастырылып отырған регрессия модельдері үшін жоспарлар жауап функциясының өлшеулер нәтижелеріне тәуелді емес, сондықтан оларды алдын ала табуға болады.

6.3 Бірінші ретті жоспарлар

Құрылымы көз қарасы жағынан объект модельдері сызықты-параметрленген және бейсызықты-параметрленген, яғни параметрлері бойынша сызықты және бейсызықты болуы мүмкін. Сызықты-параметрленген модельдерді қарастырамыз. Мұндай модельдер факторлар бойынша сызықты, факторлар бойынша толық емес квадраттық және факторлар бойынша квадраттық болуы мүмкін. Параметрлері бойынша сызықты болып табылатын қарастырылатын регрессия модельдері үшін жоспарлар жауап функциясын өлшеу нәтижелеріне тәуелді емес, сондықтан оларды алдын ала табуға болады. Практикада көпмүшелік регрессия модельдері жиі қолданылады.

Бірінші ретті жоспарлар келесі түрдегі сызықты регрессия модельдерін тәжірибелік түрде алуға арналған:

$$\varphi(X_1, \ldots, X_n) = B_0 + B_1 X_1 + \ldots + B_n X_n.$$

Мұндай модельдердің белгісіз регрессия коэффициенттерінің бағаларын анықтау үшін факторлардың әрқайсысы кем дегенде екі деңгейде өзгеруі керек. X_i^0- базалық (негізгі) деңгейді, ΔX_i -X_i факторының вариация қадамын белгілейік. Тәжірибелік нүктелері келесі гиперпараллелепипедтің ішінде немесе шекарасында орналасуы керек:

$$X_i^0 - \Delta X_i \leq X_i \leq X_i^0 + \Delta X_i, i = \overline{1, n}.$$

Есептеулерді жеңілдету үшін әдеттегі масштабтан нормаланған (стандартталған) масштабына көшейік:

$$x_i = \frac{X_i - X_i^0}{\Delta X_i}, i = \overline{1, n}.$$

Нормаланған айнымалылар үшін аламыз:

$$\varphi(\bar{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n;$$

$$\beta_i = B_i \Delta X_i, i = \overline{1,n}; \beta_0 = B_0 + \sum_{i=1}^{n} B_i X_{i;}$$

$$-1 \leq x_i \leq +1, i = \overline{1,n}.$$

Бірінші ретті жоспарлардың бірнеше түрлері бар – бір факторлық тәжірибе, толық факторлық тәжірибе, бөлшек факторлық тәжірибе. Әрине, регрессиялық талдаудың барлық негізгі алғышарттары қанағаттандырылады деп есептеледі.

6.3.1 Бір факторлық тәжірибе.

Бір факторлық тәжірибе факторлардың әрқайсысын кезекпен өзгертуді қамтиды, ал қалған барлық факторлар сол уақытта белгілі бір деңгейде тұрақталады. Сызықты модельді алу үшін әрбір фактор үшін екі деңгей жеткілікті. Әрбір фактор $x_{iж} = x_i^0 + \Delta x_i; x_{iт} = x_i^0 - \Delta x_i$, мәндерін қабылдайды деп есептейміз және осы сәтте барлық басқа факторлар $x_j^0, j \neq i$ негізгі деңгейінде тұрақтанды. Нормаланған айнымалылар үшін $x_{iж} = +1, x_{iт} = -1, x_j = 0$.

Сонда спектр матрицасы келесідей көрсетіледі (6.1, а сурет) немесе кесте түрінде (6.1, б сурет) көрінеді.

Геометриялық тұрғыдан жоспардың нүктелері шектейтін шаршының қабырғаларының ортасында орналасқан.

$$\left\| \begin{matrix} -1 & 0 & 0 \ldots\ldots\ldots 0 \\ +1 & 0 & 0 \ldots\ldots\ldots 0 \\ 0 & -1 & 0 \ldots\ldots\ldots 0 \\ 0 & +1 & 0 \ldots\ldots\ldots 0 \\ \ldots & \ldots & \ldots\ldots\ldots\ldots \\ 0 & 0 & 0 \ldots\ldots -1 \\ 0 & 0 & 0 \ldots\ldots +1 \end{matrix} \right\|$$

а)

g	x_1	x_2	x_3	…	x_n
1	-1	0	…	…	0
2	+1	0	…	…	0
3	0	-1	0	…	0
4	0	+1	0	…	0
…	….	…	…	…	…
N-1	…	…	…	…	-1
N	…	…	…	…	+1

б)

6.1 сурет – Бір факторлық тәжірибе жоспарының спектрінің матрицалық (а) және кестелік (б) көрінісі

Жоспардың нүктелері геометриялық гиперкуб қырларының ортасында орналасқан ($n=2$ үшін шектеуші шаршы жақтарының ортасында).

Базистік функцияларының баған-векторы– $\bar{f}^T(\bar{x}) = \|1, x_1, \dots, x_n\|$.

Базистік функциялардың сандық мәндерінің матрицасының баған векторлары жұптық ортогональды болады. Осыны ескере отырып, Фишер ақпараттық матрицасы Φ және оның кері Φ^{-1} матрицасы 6.2 суреттегідей.

Осыдан коэффициенттер бағаларының дисперсиясының формулаларын алуға болады (жалпы формуланың түрі келесідей болатыныне еске түсірейік: $\sigma_{b_i}^2 = \sigma_e^2 \cdot C_{ii}, i = \overline{1,n},$ мұндағы C_{ii} - Φ^{-1} матрицасының диагональ элементі:

$$\sigma_{b_0}^2 = \frac{\sigma_e^2}{2n}, \sigma_{b_i}^2 = \frac{\sigma_e^2}{2}, i = \overline{1,n}.$$

$$\Phi = \begin{Vmatrix} 2n & 0 & 0 \dots 0 \\ 0 & 2 & 0 \dots 0 \\ \dots & \dots & \dots \dots \dots \\ 0 & 0 & 0 \dots 2 \end{Vmatrix}, \Phi^{-1} = \begin{Vmatrix} \frac{1}{2n} & 0 & 0 \dots 0 \\ 0 & \frac{1}{2} & 0 \dots 0 \\ \dots & \dots & \dots \dots \dots \\ 0 & 0 & 0 \dots \frac{1}{2} \end{Vmatrix}$$

а) б)

6.2 сурет – Қалыпты теңдеулер жүйесінің матрицасы (а) және оған кері (b) матрица

Регрессиялық модель бойынша болжау дисперсиясы келесі өрнектен табылады:

$$\sigma^2\{\hat{y}(\bar{x}, \bar{b})\} = D\{b_0 + b_1 x_1 + \dots + b_n x_n\} = \sigma_{b_0}^2 + \sum_{i=1}^{n} x_i \, \sigma_{b_i}^2 = \sigma_e^2 \left[\frac{1}{2n} + \frac{1}{2}\sum_{i=1}^{n} x_i^2\right].$$

Тәжірибе орталығынан бекітілген қашықтық үші $\rho = \sqrt{\sum_{i=1}^{n} x_i^2}$ бұл дисперсия $\sigma_e^2 \left[\frac{1}{2n} + \frac{1}{2}\rho^2\right] = const.$ Яғни, қарастырылып отырған жоспардың ротатабельдік қасиеті бар.

Басқа оңтайлылық критерийлері тұрғысынан бір факторлық тәжірибе қанағаттанарлық емес. Сондықтан ол кеңінен қолданылмайды.

6.3.2 Толық факторлық тәжірибе.

n тәуелсіз бақыланатын факторлардың барлық мүмкін болатын қайталанбайтын комбинацияларын олардың өзгеруінің барлық деңгейінде жүзеге асыратын тәжірибе *толық факторлық тәжірибе* (ТФТ) деп аталады.

Егер әрбір фактор екі деңгейде өзгерсе (бұл сызықты регрессиялық модельдерді құру үшін жеткілікті), онда мұндай комбинациялардың жалпы саны $N = 2^n$. Мұндай жоспарлар ТФТ 2^n деп белгіленеді. 2^n – жоспардың спектрінің элементтерінің саны.

n=2 үшін ТФТ 2^2 жоспар матрицасының спектрі нормалданған айнымалылары үшін 6.3 суретте көрсетілген түрі болады. ТФТ 2^3 жоспар матрицасы 6.4 суретте көрсетілген.

$$\begin{Vmatrix} -1 & -1 \\ +1 & -1 \\ -1 & +1 \\ +1 & +1 \end{Vmatrix}$$

g	x_1	x_2
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1

а) б)

6.3 сурет - ТФТ 2^2 жоспар спектрінің матрицалық (а) және кестелік (б) көрінісі

$$\begin{Vmatrix} -1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ +1 & +1 & -1 \\ -1 & -1 & +1 \\ +1 & -1 & +1 \\ -1 & +1 & +1 \\ +1 & +1 & +1 \end{Vmatrix}$$

g	x_1	x_2	x_3
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1

а) б)

6.4 сурет - ТФТ 2^3 жоспар спектрінің матрицалық (а) және кестелік көрінісі (б)

Еркін ТФТ 2^n жоспарының спектрін жазу үшін қарапайым ереже қолданылады: біріншіден, x_1 бағаны толтырылады, бірінші жолда -1-ден басталып, келесі жолға өткенде таңбаларын кезекпен өзгертеді; әрбір келесі баған үшін белгілердің өзгеру жиілігі алдыңғы бағандағыдан 2 есе аз болады. Регрессиялық модельдің бос коэффициенті үшін барлық сынақтарда мәні +1 болатын фиктивті x_0 айнымалысы енгізіледі.

Геометриялық түрде ТФТ 2^n жоспарының нүктелері *n*-өлшемді кубтың төбелерінде орналасқан (*n=2* үшін нүктелер шектейтін квадраттың төбелерінде, *n=3* үшін, кубтың төбелерінде орналасқан).

ТФТ 2^n негізінен регрессиялық модельдің базистік функцияларының 2^n коэффициенттерін бағалауға мүмкіндік береді. Бірінші ($n+1$) базистік функциялар анық: $f_0(\bar{x}) = 1, f_1(\bar{x}) = x_1, \dots, f_n(\bar{x}) = x_n$. Олар сызықты модельді құрайды. Регрессия моделіне тағы қандай базистік функцияларды қосуға болады? Еске салайық, регрессиялық талдаудың алғышарттары осы базистік функциялардың сандық мәндерінің бағандарының сызықтық тәуелсіздік шартын қамтиды ($\bar{F}$матрицасының бағандары). Ерікті дәрежелі көпмүшеге x_i дәрежелері немесе олардың көбейтінділерінің әртүрлі комбинациялары қосылуы мүмкін болғандықтан және ТФТ 2^n-дегі бұл факторлардың өзі +1 немесе -1 мәндерін қабылдайды, онда кез келген базистік функцияға сәйкес келетін баған +1 және -1 мәндерден тұрады. Сонда сызықтық тәуелсіздік шартын былай тұжырымдауға болады: $\bar{F}$ матрицаның толық сәйкес келетін немесе мүлдем қарама-қарсы (белгі бойынша) бағандары болмауы керек. Бұдан шығатыны: базистік функциялардың ішінде x_i^k (мұндағы $k>1$, бүтін сан) түріндегі функция болуы мүмкін емес, өйткені жұп k үшін баған +1, ал тақ k үшін $x_i^k = x_i$ мәндерінен тұрады. Сонымен қатар факторлардың көбейтінділерінің кез келген комбинациясы базистік функциялардың қатарында болуы мүмкін.

Мысалы, $n=2$ үшін базистік функциялар жиыны:

$$f_0 = 1, f_1 = x_1, f_2 = x_2,$$

және $f_3 = x_1 \cdot x_2$ функцияны қосуға болады;

$n=3$ үшін базистік функциялар жиыны келесі түрде болады:

$$f_0 = 1, f_1 = x_1, f_2 = x_2, f_3 = x_3, f_4 = x_1x_2, f_5 = x_1x_3, f_6 = x_2x_3, f_7 = x_1x_2x_3.$$

Бұл жиындардағы функциялардың әрқайсысының басқалардан ерекшеленетін өз бағанасы бар (6.4, 6.5-суреттер). Бұл суреттерде жоспардың спектрі тікбұрышты жақтаумен белгіленген.

g	$f_0=1$	$f_1=x_1$	$f_2=x_2$	$f_3=x_1x_2$
1	+1	-1	-1	+1
2	+1	+1	-1	-1
3	+1	-1	+1	-1
4	+1	+1	+1	+1

6.4 сурет – ТФТ 2^2 жоспар матрицасы

Осылайша, ТФТ 2^n көмегімен келесі түрдегі регрессия модельдерін алуға болады:

$$\varphi(\bar{x}) = \sum_{j=0}^{2^n-1} \beta_j f_j(\bar{x}) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_1<i_2}}^{n} \beta_{i_1 i_2} x_{i_1} x_{i_2}$$

$$+ \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{\substack{i_3=1 \\ i_1<i_2<i_3}}^{n} \beta_{i_1 i_2 i_3} x_{i_1} x_{i_2} x_{i_3} + + \ldots + \beta_{12\ldots n} x_1 x_2 \ldots x_n$$

g	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7
1	+1	-1	-1	-1	+1	+1	+1	-1
2	+1	+1	-1	-1	-1	-1	+1	+1
3	+1	-1	+1	-1	-1	+1	-1	+1
4	+1	+1	+1	-1	+1	-1	-1	-1
5	+1	-1	-1	+1	+1	-1	-1	+1
6	+1	+1	-1	+1	-1	+1	-1	-1
7	+1	-1	+1	+1	-1	-1	+1	-1
8	+1	+1	+1	+1	+1	+1	+1	+1

6.5 сурет - ТФТ 2^3 жоспар матрицасы

Коэффициенттердің жалпы саны $1 + n + C_n^2 + C_n^3 + \cdots + C_n^n = 2^n$, яғни ТФТ жоспары қаныққан болып табылады.

Базистік функциялардың сандық мәндерінің $\bar{F}$ матрицасы келесі қасиеттерге ие:

1) симметрия қасиеті: базистік функциялардың бағандарының векторлар элементтерінің алгебралық қосындысы нөлге тең (бос мүшеге сәйкес келетін бағанды қоспағанда):

$$\sum_{g=1}^{N} f_j\left(x_g\right) = 0, j = 1, \ldots, d-1;$$

$$\sum_{g=1}^{N} f_0\left(x_g\right) = N;$$

2) базистік функциялар мәндер матрицасының бағандарының ортогональдық қасиеті: барлық баған-векторларының скалярлық көбейтіндісі (матрицаның кез келген екі баған-векторының мүшелер бойынша көбейтінділерінің қосындысы) нөлге тең:

$$\sum_{g=1}^{N} f_j(\bar{x}_g) f_k(\bar{x}_g) = 0, j \neq k, k = 0, \dots, N-1.$$

Осы қасиетінің арқасында регрессия теңдеуінің коэффициенттерін есептеуге байланысты қиындықтар азаяды.

3) нормалау қасиеті: әрбір бағанның элементтер квадраттарының қосындысы тәжірибелер санына тең:

$$\sum_{g=1}^{N} f_j^2\left(\bar{x}_g\right) = N.$$

Бұл өрнектер тәжірибелердің қайталануы жоқ деген болжаммен жазылған. Осы жерден табамыз:

$$\Phi = \left\| \begin{matrix} N & 0 & 0 \dots 0 \\ 0 & N & 0 \dots 0 \\ \dots & \dots & \dots \dots \dots \\ 0 & 0 & 0 \dots N \end{matrix} \right\| \Phi^{-1} = \left\| \begin{matrix} \frac{1}{N} & 0 & 0 \dots 0 \\ 0 & \frac{1}{N} & 0 \dots 0 \\ \dots & \dots & \dots \dots \dots \\ 0 & 0 & 0 \dots \frac{1}{N} \end{matrix} \right\|$$

яғни ТФТ 2^n ортогональды жоспар болып табылады.

4) $\sigma_{b_j}^2 = \frac{\sigma_e^2}{N}$, яғни барлық коэффициенттер бірдей дәлдікпен есептеледі.

5) ТФТ 2^n жоспарлау көмегімен алынған сызықты регрессиялық модель бойынша жауап функцияны болжаудың дисперсиясы:

$$\sigma^2 \hat{y}\left(\bar{x}, \bar{b}\right) = D b_0 + \sum_{i=1}^{n} b_i x_i =$$

$$= \sigma_{b_0}^2 + \sigma_{b_i}^2 \sum_{i=1}^{n} x_i^2 = \frac{\sigma_e^2}{N}\left(1 + \sum_{i=1}^{n} x_i^2\right) = \frac{\sigma_e^2}{N}(1 + \rho^2).$$

Яғни, бекітілген ρ үшін дисперсия тұрақты, сондықтан ТФТ 2^n жоспары ротатабельді болады. Ол сондай-ақ *A*-, *E*-, *D*-, *G*-оңтайлы. Демек, құрылысының қарапайымдылығымен бірге ТФТ 2^n жоспары жақсы қасиеттерге ие.

6.3.3 Бейсызықты регрессиялық модель.

Сызықты модельдер есептердің шектеулі ауқымын шешкенде ғана адекватты болып шығады. Егер сызықты модель адекватты емес болып шықса, онда бейсызықты модельге көшу керек. Математикалық модельдердің сызықты

еместігінің ең көп тараған себептерінің бірі - бір фактордың әсері (шығыс параметріне әсер етуі) басқа факторлардың қандай мәндер деңгейлерінде орналасқанына байланысты болады. Бұл жағдайда факторлар арасында *өзара әрекеттесу* бар деп айтады.

Мысал: Режим параметрлерін өзгеруінің жұмыс аймағында бұрғылау сұйықтығының Q шығынының және қашаудағы $G_қ$ осьтік жүктеменің жоғарылауымен бұрғылаудың механикалық жылдамдығы $V_м$ өсетіні бұрғылау тәжірибесінен белгілі. $G_қ$ 15%-ға артқанда $V_м$ шамасы 20-ға өседі дейік, ал тек қана Q 20%-ға артқанда $V_м$ 10%-ға артады. $G_қ$ және Q бір мезгілде бірдей мәндерге ұлғайған кезде $V_м$ шамасы 20 + 10 = 30% -ға артады деп күтуге болады. Дегенмен, шын мәнінде, $V_м$, мысалы, 40% -ға артуы мүмкін.

Оның себебі физикалық процестің ерекшелігінде – осьтік жүктеменің қашауға әсер ету механизмінде және бұрғылау ұңғысының кенжарында тау жыныстарының бұзылу процесіне бұрғылау ерітіндісінің жұмсалуында. Бұл құбылыстың механизмі келесідей. Қашаудағы осьтік жүктеменің ұлғаюымен ұңғыма кенжарын бұзу процесі күшейеді, ал үңгілеудің механикалық жылдамдығы артады. Бұл белгілі бір шекке дейін орын алады - нәтижесінде алынған шлам сұйықтық ағынымен уақтылы алынып тасталса, ал қашау таза ұңғымада жұмыс істейтін болса. Қашаудағы осьтік жүктеменің одан әрі ұлғаюымен ұңғыма түбінде шлам көлемі өседі, бұрынғы ерітінді ағыны жеткіліксіз, бұрғылау жылдамдығының өсуі төмендейді. Егер осы сәтте бұрғылау сұйықтығының шығыны айтарлықтай жоғарыласа, онда бұрғылау жылдамдығы әлдеқайда жоғары жылдамдықпен өсе бастайды. Яғни, бұрғылау режимінің бұл екі параметрі бұл жағдайда бір-бірін күшейтеді. Факторлар бір-бірінің әрекетін әлсірету бағытында да әсер етуі мүмкін.

Факторлардың өзара әрекеттесу ретін анықтау.

Екі x_1, x_2 фактор жағдайында өзара әрекеттесу әдетте осы факторлар арасында ғана орын алады, ол $x_1 \cdot x_2$ ретінде жазылады. Онда жұптық әрекеттесу немесе 2-ші ретті өзара әрекеттесу бар дейді. Үш факторлардың x_1, x_2, x_3 арасында өзара әрекеттесудің саны төртеу болуы мүмкін: үш $x_1 \cdot x_2, x_1 \cdot x_3, x_2 \cdot x_3$ 2-ші ретті өзара әрекеттесулер және тағы біреуі – 3-ші ретті өзара әрекеттесу – $x_1 \cdot x_2, x_1 \cdot x_3, x_2 \cdot x_3$. Факторлар саны көбейген сайын ықтимал өзара әрекеттесулердің саны да артады. Бұл сан комбинаторика формуласы бойынша табылады: n элементтерден k бойынша топтастыру комбинациялар саны:

$$C_n^k = \frac{n!}{k!(n-k)!},$$

мұндағы n – факторлардың саны;

k - өзара әрекеттесудегі элементтер саны.

Факторлар арасындағы мұндай өзара әрекеттесудің механизмдері белгілі болса, онда оларды техникадағы да, адам өмірінің де әртүрлі салаларындағы

көптеген мәселелерді шешуде сәтті пайдалануға болады. ТФТ факторлардың өзара әрекеттесуінің әсерін оның механизмі белгілі немесе белгісіз екеніне қарамастан сандық бағалауға мүмкіндік береді. Егер өзара әрекеттесу механизмі белгілі болса, онда бұл мақсатты түрде зерттеулер жүргізуге, факторлардың деңгейлерін және вариациялау интервалдарын дәлірек таңдауға мүмкіндік береді.

Факторлардың өзара әрекеттесуінің әсерлерін есепке алу үшін бағандарды көбейту ережесін қолдана отырып, математикалық модельге кіретін факторлардың сәйкес көбейтіндісінің бағаны алынады. Өзара әрекеттесу әсеріне сәйкес коэффициентті есептеу кезінде жаңа баған векторы жаңа фактор сияқты қарастырылады. Өзара әрекеттесу әсерлерінің бағандарын қосқанда жоспарлау матрицасының барлық қарастырылатын оңтайлы қасиеттерінің сақталуы өте маңызды. Тәуелсіз факторлардың бағандары тәжірибенің жоспарын белгілейді, олар тәжірибелерді жүргізу шарттарын тікелей анықтайды, ал есептеу үшін барлық басқа бағандар, соның ішінде жалған айнымалының бағаны және тәжірибелер нәтижелерінің *y* бағанының векторы қолданылады.

Мұндай жоспарды пайдалана отырып, толық емес дәрежелі моделінің теңдеуінің регрессия коэффициенттерін есептеуге болады. Барлық ықтимал эффекттердің жалпы саны, соның ішінде b_0, сызықты эффекттер және барлық әртүрлі ретті өзара әрекеттесу эффекттері, ТФТ-гі сынақтарының санына тең.

Жоспарлау матрицаның ортогональдық қасиеті бір-біріне тәуелсіз коэффициенттердің бағаларын алуға мүмкіндік береді. Бірақ бұл мәлімдеме модель тек сызықты әсерлер мен өзара әсер ету эффекттерін қамтитын жағдайлар үшін ғана дұрыс. Бірақ, факторлардың квадраттарындағы коэффициенттер, олардың кубтары және т.б. маңызды болып шығуы мүмкін.

Практикалық тұрғыдан алғанда, n факторларының санының ұлғаюымен тәжірибелер санының тез өсуімен байланысты ТФТ жоспарының айтарлықтай кемшілігі бар (мысалы, $n=10$ кезінде тәжірибелер саны $N=1024$). Бірақ, жауап функцияларында x_j^2 квадраттық мүшелері елеусіз, ал сонымен бірге жоғары ретті ($x_1 x_2 \dots x_n$) өзара әрекеттесулерінің рөлі жоғары болады деген де дұрыс емес. Яғни, $n>5$ үшін осындай модельдердің практикалық қолдануды табуы екіталай. Әдетте, алдымен зерттеуші қосымша жұптық (төтенше жағдайларда, үштік) өзара әрекеттесуі бар сызықтық модельмен шектеледі. Мұндай жағдайда ТФТ жоспары артық болады. Тәжірибелердің саны бойынша неғұрлым үнемді әртүрлі жоспарларды құру қажет. Сондықтан бөлшек факторлық эксперименттер (БФТ) өңделген.

6.3.4 Регрессия теңдеуінің коэффициенттерін анықтау.

ТФТ жоспар матрицасының қасиеттерін ескере отырып, регрессия моделінің бос және сызықты мүшелерінің коэффициенттер бағалары тәуелсіз алынады және келесі формулалар арқылы табылады:

$$b_0 = \frac{1}{N}\sum_{g=1}^{N} x_{0g}\bar{y}_g\,;$$

$$b_i = \frac{1}{N}\sum_{g=1}^{N} x_{ig}\bar{y}_g\ (i = 1,2,\ \dots,\ n).$$

Яғни, регрессия теңдеуінің кез келген коэффициенттері $\bar{y}$ бағанының сәйкес фактордың бағанымен скаляр көбейтіндісі ретінде анықталады.

Сызықты өзара әрекеттесулердің коэффициенттері (2-ші, 3-ші ретті) ұқсас жолмен анықталады:

$$b_{il} = \frac{1}{N}\sum_{g=1}^{N} x_{ig}x_{lg}\bar{y}_g\ (i, l = 1,2,\ \dots,\ n; i \neq l).$$

$$b_{ilk} = \frac{1}{N}\sum_{g=1}^{N} x_{ig}x_{lg}x_{kg}\bar{y}_g\ (i, l, k = 1,2,\ \dots,\ n; i \neq l \neq k).$$

Коэффициенттер тәуелсіз болғандықтан, әрбір коэффициент процестегі сәйкес айнымалының рөлін немесе факторлардың әсер ету күшін сипаттайды. Коэффициенттің сандық мәні неғұрлым үлкен болса, соғұрлым бұл фактордың әсері күшейеді. Коэффициенттің плюс таңбасы болса, онда фактор мәнінің жоғарылауымен жауап функцияның мәніде жоғарылайды, ал егер минус белгісі болса, ол төмендейді. Регрессия теңдеуінің статистикалық талдауынан кейін кейбір коэффициенттерді маңызды емес деп қабылдануы мүмкін және коэффициенттері нөлден айтарлықтай айырмашылығы жоқ факторлар теңдеудің құрамынан шығарылады, өйткені олардың жауап функцияға әсері тәжірибелік қатеге жатқызылады. Бұл жағдайда жоспардың ортогональ қасиетін ескере отырып, регрессия теңдеуінің қалған коэффициенттерін қайтадан есептеу қажет емес. Егер тәжірибе жоспары ортогональды болмаса, барлық коэффициенттерді қайта есептеу керек.

6.3.5 ТФТ әдісімен регрессия моделін құру мысалы.

t уақытына және T^0C температурасына тәуелді кейбір химиялық реакцияның *y* шығымын зерттейміз. Бұл тәуелділікті *t=4* сағат, $T=220^0C$ нүктесінің аймағында сызықты деп санауға болады.

Осы процестің регрессиялық моделін құрастырып, оны талдап көрейік. Ыңғайлы болу үшін реакцияның ұзақтығын t - x_1, температураны T^0 - x_2, шығыс айнымалысын - *y* деп белгілейміз.

Сызықты модельдің құрылымы келесі түрде болады:

$$y = b_0 + b_1x_1 + b_2x_2,$$

модель мүшелерінің саны $d=3$.

Факторлары мен олардың өзгеру деңгейлері кестеде келтірілген:

Факторлар	x_1	x_2
Негізгі деңгей	4	220
Вариациялау интерва	1	10
Жағарғы деңгей	5	230
Төменгі деңгей	3	210

Тәжірибені жоспарлау ТФТ 2^2 әдісі бойынша модельді табу келесі қадамдардан тұрады:

а) тәжірибені жоспарлау.

Әдетте басқару факторлары тәжірибе жоспарын құрайды, F матрицаның қалған бағандары көбейту арқылы алынады; біздің жағдайда нормаланған масштабтағы жоспар матрицасы келесідей болады:

x_1	x_2
-1	-1
+1	-1
-1	+1
+1	+1

б) тәжірибені жүргізу.

Шығудағы шаманың өзгеруі кездейсоқ болғандықтан, жоспардың әрбір нүктесінде m параллель тәжірибе жүргізіп, бақылау нәтижелері бойынша орташа мәнді есептеу қажет:

$$\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i.$$

Әрбір серияда жоспарлау матрицасы жүзеге асырылады. Жоспарды жүзеге асырмас бұрын вариациялау нұсқаларын рандомизациялау қажет, яғни тәжірибелердің әрбір сериясында жоспарлау матрицасының нұсқаларын іске асырудың кездейсоқ реттілігін орнату.

$N=4$, $m=2$ үшін тәжірибелер нәтижелері төменде келтірілген (y^I, y^{II} – тәжірибелердің бірінші және екінші серияларындағы жауап функцияның мәндері, $\bar{y}$ – жауаптың орташа мәні):

g	x_1	x_2	y^I	y^{II}	$\bar{y}$
1	-1	-1	44,9	45,0	44,95
2	+1	-1	55,0	55,2	55,1
3	-1	+1	55,0	54,8	54,9
4	+1	+1	65,5	65,6	65,55

в) регрессия коэффициенттерін есептеу.

Базистік функциялардың сандық мәндер матрицасының ортогональдылығына байланысты Фишер теңдеулер жүйесінің матрицасы диагональды болып табылады, бұл модельдің коэффициенттерін бірден есептеуге мүмкіндік береді. Нәтижесінде келесі модель табылды:

$$y = 55.125 + 5.2x_1 + 5.1x_2$$

Нәтижелерді статистикалық талдау.

г) σ_i^2 таңдамалы дисперсияларының қайтадан өндірілуін тексеру.

Яғни регрессиялық талдаудың алғышарттарының бірін - σ_i^2 таңдамалы дисперсияларының біртекті болатынын тексеру. Дисперсияны әр нүктедегі бағалары мына формула бойынша жүргізіледі:

$$S_g^2 = \frac{1}{m-1}\sum_{k=1}^{m}(y_{gk} - \bar{y}_g)^2, \nu_{\text{өндіру}} = m - 1.$$

Тексеру үшін максималды дисперсия бағасының барлық салыстырылатын дисперсиялар бағаларының қосындысына қатынасының таралу заңына негізделген Кокрен критерийі қолданылады:

$$G = \frac{max\{S_g^2\}}{\sum_{g=1}^{N} S_g^2\{y\}}; \nu_{1\text{өндіру}} = m - 1, \nu_{2\text{өндіру}} = N;$$

Егер $\nu_1 = m - 1$, $\nu_2 = N$ және таңдалған маңыздылық q деңгейі үшін $G_{есеп}$<$G_{кесте}$, болса, онда біртектілік гипотезасы қабылданады. Ал бас жиынның қайтадан өндірілуі дисперсиясының $\sigma^2\{y\}$ бағасы тең болады:

$$S^2_{\text{өндіру}}\{y\} = \frac{1}{N}\sum_{g=1}^{N} S_g^2\{y\}$$

еркіндік дәрежелер санымен $\nu = N(m-1)$.

Егер тәжірибенің қайтадан өндірілуі туралы сынақ теріс нәтиже берсе, онда не параллельді тәжірибелер санын көбейту керек, немесе объект

шығысында «шу» деңгейін көтеретін басқарылмайтын және бақыланбайтын айнымалылардағы тербелістердің болуына байланысты басқарылатын айнымалыларға қатысты тәжірибелер қателерінің бірдей деңгейінде қайталанбайтын деп мойындау керек.

Есептеулер кестеде көрсетілген:

g	x_1	x_2	y^I	y^{II}	y_{cp}	S_g^2
1	-1	-1	44,9	45,0	44,95	0,0025
2	+1	-1	55,0	55,2	55,1	0,02
3	-1	+1	55,0	54,8	54,9	0,02
4	+1	+1	65,5	65,6	65,55	0,005
						∑=0,0475

$$\mathrm{G}_{\text{есеп.}} = \frac{max\{S_g^2\}}{\sum_{g=1}^{N} S_g^2\{y\}};$$

$G_{\text{есеп}} = \dfrac{0{,}02}{0{,}0475} = 0{,}421.$

$G_{\text{кесте}} = 0{,}9\ (\nu_1 = m - 1 = 1, \nu_2 = N = 4).$

$G_{\text{есеп}} < G_{\text{кесте}}$ – тәжірибе қайтадан өндіріледі.

Қайтадан өндірілу дисперсияны бағасы:

$$S_e^2\{y\} = \frac{1}{N}\sum_{g=1}^{N} S_g^2\{y\} = 0{,}105.$$

Егер тәжірибенің қайтадан өндірілуін тексеру теріс нәтиже берсе, онда не параллельді тәжірибелер санын көбейту керек, не басқарылмайтын және бақыланбайтын айнымалылардағы тербелістердің болуына байланысты басқарылатын айнымалыларға қатысты тәжірибе қайтадан өндірілмейді деп тану керек, бұл шығуда үлкен "шу" деңгейін тудырады.

д) коэффициенттердің маңыздылығын зерттеу Стьюдент критерийі бойынша жүргізіледі. Бізде бар:

$$C_{jj} = 4,\ S_a^2 = \frac{S_e^2}{4} = 0{,}026,$$

мұндағы C_{jj} – Фишер матрицасының диагональ элементі.

Барлық коэффициенттер үшін критерийдің есептелген мәндері:

$$g_i = \frac{b_i}{S_b^2};\ g_1 = 2094, g_2 = 197.6, g_1 = 193{,}8.$$

Критерийдің кестелік мәні: $t_{\text{кесте}} = 2{,}13$.

$t_j > t_{\text{кесте}}$ болғандықтан, нөл-гипотеза қабылданбайды, барлық коэффициенттер маңызды болып табылады;

е) модельдің адекваттылығын тексеру.

Адекваттылық дисперсиясы мына формула бойынша есептеледі:

$$S_{\text{ад}}^2 = \frac{1}{N-d}\sum_{g=1}^{N}(\bar{y}_g - \hat{y}_g)^2,$$

мұндағы *N=4, d=3*

$\hat{y}_g$- регрессия теңдеуі бойынша есептелген жауап мәні.

Есептеулер төменде кестеде келтірілген:

$y_{орта}$	$\hat{y}_g$	$(y_{орта}-\hat{y}_g)^2$
44,95	44,825	0.0156
55,1	55,225	0.000625
54,9	55,025	0,0156
65,55	65,425	0.000625
		∑=0,0312

Адекваттылық дисперсияның мәні: $S_{\text{ад}}^2 = 0.0312$.

Фишер критерийі бойынша адекваттылықты тексеру:

$$F_{rasch} = \frac{S_{ad}^2}{S_e^2} = \frac{0{,}0312}{0{,}105} = 0{,}297.$$

$\nu_1 = N - d^* = 1, \nu_2 = N(m-1) = 4$ болғанда $F_{\text{кесте}} = 7{,}71$,

мұндағы d^*- маңызды коэффициенттері бар модель мүшелерінің саны (мысалда барлық коэффициенттер маңызды).

$F_{\text{есеп}} > F_{\text{кесте}}$, сондықтан модель адекватты

ж) модельдің жұмысқа қабілеттігін тексеру. Детерминация коэффициентін келесі формула бойынша есептейміз:

$$R^2 = \left[\sum_{g=1}^{N}(y_g - \bar{y}_g)^2\right] / [\sum_{g=1}^{N}(y_g - \hat{y}_g)^2].$$

Есептеулер төмендегі кестеде келтірілген:

y^I	y^{II}	y_{cp}	$\hat{y}_g$	$\sum_{g=1}^{N}(y_g-\bar{y}_g)^2$	$\sum_{g=1}^{N}(y_g-\hat{y}_g)^2$
44,9	45,0	44,95	44,825	0,005	0,03625
55,0	55,2	55,1	55,225	0,02	0,05125
55,0	54,8	54,9	55,025	0,02	0,05125
65,5	65,6	65,55	65,425	0,005	0,03625
				∑=0,05	∑=0,175

$R^2 = 0.2857 < 1.$

R шаманың аздығы теңдеудің төмен дәлдігін көрсетеді, модельдің жұмысқа қабілеті төмен. Бірақ бұл қорытынды тәжірибелер саны модель параметрлерінің санымен салыстырғанда жеткілікті, үлкен болған кезде жарамды. Біздің жағдайда модель параметрлерінің саны бақылаулар санына жақындайды. Модельдің сапасын тексеру үшін қосымша тәжірибелер жүргізген жөн.

6.3.6 Бөлшек факторлық тәжірибе.

Көптеген практикалық идентификациялау мәселелерінде мәселелерінде екінші және жоғары ретті өзара әрекеттесулердің (факторлар көбейтінділулерінің) әсері жоқ немесе шамалы. Мысалы, негізгі режимде энергоблоктарды автоматты басқару жүйелерін реттеу кезінде, реттелетін параметрлердің статистикалық сипаттамаларын тек сызықтық жуықтауда білу жеткілікті. Сонымен қатар, көбінесе бастапқы кезеңдерде бірінші жуықтауда зерттелетін теңдеуінің тек қана сызықты жуықтауын тәжірибелердің ең аз санында алу қажет. Басқа есептерде, негізгі факторлардан басқа, ТФТ жоспарларына кіретін көптеген ықтимал өзара әрекеттесулерден тек кейбір өзара әрекеттесулерді ескеру қажет. Сондықтан, тек сызықтық мүшелері мен кейбір екінші ретті өзара әрекеттесулер үшін коэффиценттерді бағалауға ТФТ-ні пайдалану тиімсіз, себебі өте көп вариациялау нұсқаларының санын жүзеге асыру керек (2^n), әсіресе факторлар саны n үлкен болса. Тәуелсіз факторлар санының сызықтық ұлғаюымен вариация нұсқаларының саны экспоненциалды заңға сәйкес өседі, соның нәтижесінде адекваттылық гипотезасын тексеру үшін тым көп еркіндік дәрежесі қалады.

Жоғарыда айтылғандарға байланысты тәжірибелер санын азайту мәселесі туындайды. Бұл жағдайда тәжірибелер санын қысқарту ғана емес, сонымен қатар зерттеу объектісі туралы ақпараттың жеткілікті көлемін алу міндеті тұр. Тәжірибелердің санын ТФТ жоспарының артықшылығын пайдалану арқылы азайтуға болады. Әдетте, ТФТ-дегі тәжірибелер саны маңызды (немесе зерттеушіні қызықтыратын) коэффициенттер санынан көп. Мысалы, толық емес дәрежелі екі факторлы моделі келесі түрде болады:

$$.\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2.$$

Таңдалған вариация интервалдары үшін объектінің әрекеті сызықтық модельмен қажетті дәлдікпен сипатталатыны априорлық ақпараттан белгілі болсын делік:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3.$$

Мұндай жағдайларда объектінің статистикалық моделін құру үшін бөлшек факторлық тәжірибе қолданылады, оның барысында *ТФТ жоспарының бөлігі* жүзеге асырылады (*бөлшек реплика*). Толық факторлық тәжірибенің бір бөлігін жүзеге асыратын тәжірибе *бөлшек факторлық тәжірибе* (БФТ) деп аталады. БФТ минималды сынақтар санымен базалық режим нүктесінің белгілі бір шағын аймағында ізделінетін регрессия тәуелділігінің сызықтық жуықтауын алуға мүмкіндік береді.

Жоғарыда келтірілген мысал үшін факторлардың өзара әрекеттесуін сипаттайтын $b_{12}x_1x_2$ сызықты емес мүшесі ескерілмеуі мүмкін. Сондықтан, x_1x_2 баған-векторын тәжірибелер санын көбейтпей, ТФТ жоспарына кейбір жаңа факторды енгізу үшін пайдалануға болады, мысалы, x_3 факторын. Бұл матрица үлкенірек жоспардың жартысы болады. Яғни, үш фактордың әсерін бағалау үшін ТФТ 2^2 жоспарының жартысын пайдалануға болады.

Егер алты факторлы модель қарастырылса және x_1, x_2, x_3 факторлардан құрастыылған маңызы төмен өзара әрекеттесулері орнына қосымша үш жаңа факторлар, мысалы, x_4, x_5, x_6 енгізілсе, онда алынған жаңа жоспар көрсететіні ТФТ 2^6 жоспарының 1/8 бөлігін құрайды. Нақты қандай 1/8 бөлімін бұл жоспар көрсететінін x_4, x_5, x_6 бағандарының белгілеріне тәуелді, бұл белгілер x_4, x_5, x_6 орнына енгізілетін өзара әрекеттесулер бағандарының белгілеріне тәуелді.

Факторларының саны *n* көп сызықтық жуықтауды алу үшін бөлшектіктің жоғары дәрежесіндегі бөлшек репликаларын құруға болады. БFТ 2^{n-p} жоспарлар *бөлшектеу дәрежесі p болатын репликалар* деп аталады; *n* – факторлар саны. БФТ белгілеуінде, нақты *n* және *p* үшін азайту орындалмайды, екі сан да сақталады, мұнда біріншісі факторлардың санын, ал екіншісі БФТ бөлшектену дәрежесін сипаттайды. Жоғарыда келтірілген мысалдар үшін үш факторлы модель үшін БФТ 2^{3-1} жоспарын іске асыруға болады – ТФТ 2^3 жоспарының 1/2 репликасын. Алты факторлы модель үшін - ТФТ 2^6 түрі жоспардың 1/8 репликасын – БФТ 2^{6-3} жоспарын іске асыруға болады.

БФТ жоспарын алу тәртібі: тәжірибелер санын көбейтпей бастапқы ТФТ жоспарына жаңа факторды енгізу үшін, бұл факторға бастапқы факторлардан құрастырылған және әсерін елемеу мүмкін болатын өзара әрекеттесу бағанының векторы меншіктелуі керек. Бастапқы жоспарға енгізілген жаңа фактордың мәні осы бағанның белгілеріне сәйкес өзгеруі керек. Яғни, негізгі идея – тәжірибелер санын азайту. Мұндай үнемдеу зерттелетін объектінің қасиеттері туралы белгілі бір априорлық ақпараттың арқасында қол жеткізіледі.

БФТ жоспары бойынша тәжірибеде факторлар деңгейлерінің кейбір комбинациялары ескерілмейді. Деңгейлердің санын азайту ақпараттың жоғалуына әкеледі. Сондықтан бөлшек факторлық тәжірибеде берілген есептің

қойылуына сәйкес маңыздылығы төмен ақпарат жоғалатындай жоспарды құру керек. Көбінесе зерттелетін факторлардың өзара әрекеттесуі туралы ақпарат жоғалатындай бөлшек факторлық тәжірибелер қолданылады. Бұл өзара әрекеттесу әсерлері болмаған немесе оларды елемеу мүмкін болатындай аз болған жағдайда негізделеді.

Яғни, бізде зерттелетін объектінің қасиеттері туралы қосымша априорлық ақпарат болуы керек. Бұл ақпарат *маңызды айнымалылар* тізбесі түрінде құрылуы керек - одан әрі зерттеу үшін қызығушылық тудыратын факторлар мен олардың өзара әрекеттесуілері. Яғни, «маңызды айнымалылар» модельге орнату үшін таңдалған айнымалылар болып табылады. Әрине, одан әрі зерттеу барысында бұл айнымалылардың кейбірі жауап функциясына әсер етпейтіні белгілі болуы мүмкін. Маңызды айнымалылар тізімінде әрқашан b_0 коэффициентіне сәйкес келетін x_0 фиктивті айнымалы болады. Маңызды айнымалылар тізбесі немесе бастапқы модельдің түрі зерттелетін құбылыстың физикалық мәнін сапалы зерттеу нәтижесінде, мамандардың пікірін алу жолымен, объектінің белгілі бір қасиеттерін қабылдау нәтижесінде алуға болады.

БФТ көмегімен алынған нәтижелердің сапасы «маңызды айнымалылар» тізімінің дұрыстығына байланысты. Табылған модельдің сапасын талдау кезінде күрделірек және дәл модельді құру туралы шешім қабылдауға болады.

БФТ 2^{n-p} жоспарының спектрлік матрицасы ТФТ 2^n матрицасының бөлігі болып табылады. БФТ 2^{n-p} спектрінің матрицасын құрайтын 2^{n-p} жолдарын таңдау *F* базистік функциялардың сандық мәндерінің матрицасында бағандардың сызықтық тәуелсіздігінің шарты сақталатындай етіп жасалуы керек, яғни мүлдем сәйкес келетін немесе мүлдем қарама-қарсы бағандар жоқ болуы керек. Бұл талап орындалғанда ғана регрессия теңдеуінің барлық коэффициенттерін бөлек бағалауға болады.

Егер маңызды айнымалылар тізімінде *(d+1)* элементтер болса, регрессия коэффициенттерін бөлек бағалау үшін келесі шарт орындалуы керек:

$$N = 2^{n-p} \geq d + 1.$$

Бұл шарт *n* факторлардың берілген саны мен маңызды айнымалылардың белгілі тізімі *(d+1)* үшін *p* бөлшектің минималды дәрежесін табуға мүмкіндік береді.

Қай ТФТ жоспары бастапқы ретінде таңдалуы керек деген сұрақ туындайды. Бөлшек реплика ортогональды жоспар болуы үшін реплика ретінде ең жақын толық факторлық эксперимент алынуы керек. Көп жағдайда сызықты модельдер инженерлік талдау үшін қолайлы. *n* факторы бар сызықты модельдің коэффициенттерін бағалауға арналған сынақтар саны кемінде *n + 1* болуы керек. Сонымен қатар, сызықты модельдің адекваттылығын тексеру үшін кем дегенде бір қосымша тәжірибе (бір еркіндік дәрежесі) болуы керек. Сондықтан БФТ үшін сынақтардың минималды саны *(n+2)*-ден кем болмауы керек. Осы (n+2) мәнін анықтап, 8,16,32,... қатарынан ең жақын үлкенін таңдайды. Ол

бастапқы ретінде таңдалатын ТФТ жоспарының минималды сынақтар санын, демек, БФТ жоспарының сынақтар санын көрсетеді.

Мысалы, технологиялық процестің шығыс параметріне бес фактордың әсерін зерттеу үшін БФТ жоспарын құру талап етіледі. Факторлардың өзара әрекеттесуінің әсерінің күші туралы мәліметтер жоқ. Сызықты модельді алу үшін БФТ жоспарын құрайық. $n+1=7$ болғандықтан, ТФТ жоспардың сынақтарының минималды саны - 8. Сондықтан бастапқы жоспар ТФТ 2^3 болады, бастапқы факторлар ретінде кез келген үш факторды таңдаймыз. Осы факторлар үшін ТФТ 2^3 жоспарын құрамыз. Қажетті БФТ 2^{5-2} құру үшін маңызы төмен әрекеттесулердің орнына тек қана бастапқы факторлардан құрылған екі факторды енгіземіз. Өзара әрекеттесу әсерлерінің күші туралы ақпарат болмағандықтан, x_4 факторы үшін бастапқы ТФТ 2^5 -ден жоғары ретті өзара әрекеттесуді таңдаған дұрыс, мысалы, $x_1x_2x_3$. x_5 факторы үшін x_1x_2, x_1x_3 немесе x_2x_3 көбейтінділердің біреуін пайдалануға болады. Құрылған БФТ 2^{5-2} жоспары ТФТ 2^5 жоспарының ¼ бөлігі болып табылады. Егер БФТ 2^{5-2} сынақтарының нәтижелері негізінде құрастырылған сызықты модель адекватты емес болып шықса, онда модельге әсері ең үлкен болатын факторлардың әрекеттесулерін енгізу қажет.

БФТ жоспарлау кезінде маңызды мәселе оның орнына жаңа факторларды енгізу үшін өзара әрекеттесулердің қайсысын таңдау керек. Егер өзара әрекеттесу әсерлерінің күші туралы ақпарат болмаса, онда бастапқы жоспарға жаңа фактор енгізілгенде өзара әрекеттесу реті жоғары баған векторы таңдалады, өйткені жоғары дәрежелі өзара әрекеттесу әсерлерінің әдетте маңыздылығы төмен.

Мысалы, үш факторлы есепті сызықтық жуықтауда шешу керек, яғни $\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_33$ моделін алу керек. Мәселені шешу үшін біз вариациялаудың тек қана төрт нұсқасын қолдануға болады, егер де 2^2 типті жоспарлауда x_1x_2 көбейтіндіні үшінші тәуелсіз факторға теңестірілген болса. Мұндай жоспарлау b_0 еркін мүшесін және регрессия коэффициенттерінің сызықты мүшелерінің b_1, b_2, b_3 үш бағасын табуға мүмкіндік береді (төрт тәжірибеден регрессия коэффициенттерінің төрттен артық бағасын алу мүмкін емес). Сонымен, ТФТ 2^2 -де біз x_3 үшін жоспар ретінде хx_1x_2 бағанын қолданамыз. Осындай қысқартылған жоспар төмендегі кестеде көрсетілген.

g	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	+1	-1	-1	+1	+1	-1	-1	+1
2	+1	+1	-1	-1	-1	-1	+1	+1
3	+1	-1	+1	-1	-1	+1	-1	+1
4	+1	+	+1	+	+1	+	+1	+1

БФТ-ні пайдалану әрқашан араластырумен, яғни математикалық модельдің бірнеше теориялық коэффициенттерін бірлесіп бағалаумен байланысты. Қарастырылған мысалда, егер жұптық өзара әрекеттесу үшін

регрессия коэффициенттері нөлге тең болмаса, әр табылған b_j коэффициенттердің әрқайсысы екі теориялық регрессия коэффициентінің бағасы ретінде қызмет етеді:

$$b_0 \to \beta_0 + \beta_{123}; b_1 \to \beta_1 + \beta_{23}; b_2 \to \beta_2 + \beta_{13}; b_3 \to \beta_3 + \beta_{12}.$$

Шынында да, мұндай жоспарлаудағы көрсетілген теориялық коэффициенттерді бөлек бағалануы мүмкін емес, өйткені сызықтық мүшелер мен жұп көбейтінділер үшін жоспарлау матрицасының бағандары толығымен сәйкес келеді (толық корреляцияланған).

Өзара әрекеттесу әсерін дәл қалай таңдау керек? Жоғарыда келтірілген формальды алгоритм (*F* матрицасының бағандарын тексеру) күрделілігіне байланысты қабылданбайды. Практикалық қолдану үшін келесі процедура қолданылады:

1. $x_1, x_2, \ldots, x_n$ факторлардың санынан жетекші *k* фактор таңдалады, *k*=*n*-*p*. Мысалы, бірінші $x_1, x_2, \ldots, x_k$ болсын. Олар үшін ТФТ 2^k, жоспары жазылады, яғни тәжірибелер кезінде жетекші факторлардың әрқайсысына өзгеру бағдарлама орнатылады.

2. Қалған *p*=*n*-*k* факторлары үшін тәжірибе кезінде өзгеру бағдарламасы ретінде жетекші факторлардың сол немесе басқа көбейтісіне сәйкес келетін бағандар таңдалады.

$x_{k+1}, x_{k+2}, \ldots, x_n$ бағандарының бірін $x_1, x_2, \ldots, x_k$ факторларының белгілі бір көбейтіндісінің бағанына келтіретін қатынас *генерациялаау қатынас* деп аталады.

Барлығы *p* генерациялау қатынас болуы керек. Олардың таңдауы ерікті, бірақ маңызды айнымалыларды – факторларды және маңызды айнымалылар тізіміндегі жетекші факторлардың көбейтінділерін пайдалана алмайсыз (әйтпесе *F* матрицасында факторлардың бірі және өзара әрекеттесулердің бірі үшін сәйкес келетін бағандар болады). Қарастырылып отырған мысалда ТФТ 2^3 жоспарының жартылай репликасы $x_3 = x_1 x_2$ генерациялау қатынасымен арқылы орнатылды. Генерациялау қатынасты «-» белгісімен пайдалануға болады.

3. Енді жоспардың табылған спектрінің жарамдылығын тексеру қажет. Егер *F* матрицасында толық сәйкес келетін немесе мүлдем қарама-қарсы (белгі мағынасында) бағандар болмаса, онда жоспар спектрі қолайлы. Егер мұндай бағандар болса, онда басқа генерациялау қатынасты таңдау немесе жетекші факторлардың жиынын өзгерту қажет.

Минималды нүктелер санын қамтитын және қажетті регрессия моделін қамтамасыз ететін БФТ жоспарының спектрін құру процедурасы іске асыру кезінде айтарлықтай күш салуды қажет етеді. Сонымен қатар, шешімнің типті бар-жоғын алдын ала болжау қиын. *F* матрицасын әр уақытта жазу қажеттілігін болдырмайтын және бірлескен бағалау жүйелерін алуға мүмкіндік беретін формальды алгоритм бар. Ол үшін *анықтаушы қатынас* ұғымы қолданылады.

Анықтауыш қатынастың өрнегі (АҚ) генерациялау қатынастан алынады және келесі түрде болады:

сол жақта – 1, оң жақта – генерациялау қатынастың сол және оң бөліктерінің көбейтіндісі.

Мысалы, $x_3 = -x_1x_2, АҚ: 1 = x_1x_2x_3$.

Анықтаушы қатынастардың саны генерациялау қатынастардың санына тең.

Егер бірнеше анықтаушы қатынастар болса, онда барлық анықтаушы қатынастардың мүшелері бойынша көбейтіндісіне тең *жалпыланған анықтаушы қатынас* (ЖБҚ) жазылады.

Анықтаушы қатынасты білу БФТ жоспарлау матрицасын зерттемей-ақ бірлескен бағалардың бүкіл жүйесін табуға мүмкіндік береді. Бұл бағаларды анықтайтын арақатынастарды маңызды айнымалылар тізімінен әрбір айнымалыны анықтаушы қатынасқа ретімен көбейту арқылы табуға болады ($x_f^{2k} = 1, x_f^{2k+1} = x_f$ екенін ескере отырып) *d+1* теңдігі алынады, оның әрқайсысында 2^p айнымалы бар (факторлар және олардың өзара әрекеттесулері). Белгілі бір теңдікке кіретін барлық айнымалылардың әрекеттері бір-бірімен араласады. Оларға *F* матрицасының бірдей бағандарына сәйкес келеді.

Осылайша, табылған БФТ жоспарының жарамдылығын анықтау үшін көрсетілген теңдіктердің кез келгенінде маңызды айнымалылар тізімінен кемінде екі айнымалы бар-жоғын анықтау қажет. Егер бұл жоспар регрессия моделінің коэффициенттерін бағалауға мүмкіндік бермесе, оны өзгерту қажет.

Сонымен, бұл алгоритм ТФТ 2^{n-p} жоспарының спектрінің жарамдылығын тексеруге ғана емес, сонымен қатар маңыздылар тізіміне кірмейтін айнымалылардың қайсысы белгілі бір регрессия коэффициенттерінің мәндеріне әсер етуі мүмкін екенін алдын ала білуге мүмкіндік береді. Егер зерттеу барысында маңызды айнымалылар тізімін түзету қажет екені анықталса, онда бұл ақпарат қандай регрессия коэффициенттері дәл анықталмағанын және бірнеше айнымалылардың әрекетінің бағалауын білдіретінін анықтауға мүмкіндік береді.

БФТ 2^{n-p} жоспарларының қасиеттерін атап өтейік. Бұл жоспарлар ортогональды жоспарлар тобына жатады. Егер тәжірибені жоспарлау аймағы төбелері ±1 мәндерін қабылдайтын гиперкуб болса сызықты модельге қатысты олар *A*-, *D*- және *G*-оңтайлы болып табылады. Бұл қасиеттер, әрине, модельдің барлық коэффициенттерінің араласпаған бағаларын алуға болатын жағдайларда ғана орын алады.

БФТ пайдаланған кезде бөлшек репликаның рұқсат қасиеті туралы нақты түсінік болуы керек, яғни сәйкесінші бас коэффициенттердің бағалары болатын қандай коэффициенттердің араласпаған бағалар екенін алдын ала анықтау керек. Содан кейін тапсырмаға байланысты бөлшек репликалар алынады.

1 мысал: Маңызды айнымалылар тізімі берілген

$$x_0, x_1, x_2, x_3, x_4, x_1x_2, x_2x_3, x_2x_4.$$

Факторлар саны - 4, модель 8 базистік функцияны қамтиды. Сондықтан біз БФТ 2^{4-1} қолдануға тырысамыз.

Жетекші факторлар ретінде $x_0, x_1, x_2. x_3$ таңдаймыз. x_4 факторы үшін генерациялау қатынасын көрсету керек. Генерациялау қатынас ретінде келесі көбейтіндіні $x_4 = x_1x_2x_3$ таңдаймыз. Яғни, $x_1x_2x_3$ әрекеттесу бағанының өзгеретініне сәйкес келетін x_4 фактордын өзгеруге арналған бағдарламасын орнатамыз.

Анықтаушы қатынас келесідей болады:

$$ОҚ: 1 = x_1x_2x_3x_4.$$

Осы жерден араластыру жүйесін аламыз:

$$x_0 = x_0x_1x_2x_3x_4,\ x_1 = x_2x_3x_4, x_2 = x_1x_3x_4, x_3 = x_1x_2x_4, x_4 = x_1x_2x_3,$$
$$x_1x_2 = x_3x_4, x_2x_3 = x_1x_4, x_2x_4 = x_1x_3.$$

Жоспар жарамды. Қарама-қарсы таңбалы жоспар генераторын да пайдалануға болады: $x_4 = -x_1x_2x_3$.

Осы модель үшін $x_4 = x_1x_3$ генерациялау қатынасы қолайлы ма, соны анықтайық?

Анықтаушы қатынас мынадай пішінге ие: *АҚ*: $1 = x_1x_3x_4$.

Араластыру жүйесі:

$$x_0 = x_0x_1x_3x_4, x_1 = x_3x_4, x_2 = x_1x_2x_3x_4, x_3 = x_1x_4, x_4 = x_1x_3,$$
$$x_1x_2 = x_2x_3x_4, x_2x_3 = x_1x_2x_4, x_2x_4 = x_1x_2x_3.$$

Қарастырылып отырған жоспарды қолдануға болады.

x_1x_2, x_2x_3 немесе x_2x_4 генерациялау қатынастарын пайдалану мүмкін емес, өйткені олар маңызды айнымалылар тізімінде бар.

2 мысал. Алдыңғы мысалдағы маңызды айнымалылар тізімін сәл өзгертейік:

$$x_0, x_1, x_2, x_3, x_4, x_1x_2, x_2x_3, x_3x_4.$$

(Соңғы әрекеттесу өзгертілді).

Біз қайтадан ДФЭ 2^{4-1} қолдануға тырысамыз. Араластыру жүйесінде $x_1x_2 = x_3x_4$ өрнегі пайда болады. Алынған жоспар қажетті регрессия моделін табу үшін жарамсыз.

Мысал 3. Маңызды айнымалылар тізімі бар модель үшін

$$x_0, x_1, x_2, x_3, x_4, x_5, x_1x_2$$

БФТ 2^{5-2} қолдануға тырысайық.

Қажетті шарт $N= 2^{5-2}=8>d=1=7$ орындалды,

Жоспар генераторларын таңдаймыз:

$$x_4 = x_1x_2x_3, x_5 = x_1x_3.$$
$$ЖАҚ: 1 = x_1x_2x_3x_4 \cdot x_1x_3x_5 = x_2x_4x_5.$$

Араластыру жүйесі:

$$x_0 = x_2x_4x_5, x_1 = x_1x_2x_4x_5, x_2 = x_4x_5, x_3 = x_2x_3x_4x_5, x_4 = x_2x_5,$$
$$x_5 = x_2x_4, x_1x_2 = x_1x_4x_5.$$

Жоспар жарамды.

Егер біз осы мысалдағы маңыздылар тізіміне x_4x_5 мүшесін қосатын болсақ, араластыру жүйесі келесідей болады:

$$x_0 = x_2x_4x_5, x_1 = x_1x_2x_4x_5, x_2 = x_4x_5, x_3 = x_2x_3x_4x_5, x_4 = x_2x_5,$$
$$x_5 = x_2x_4, x_1x_2 = x_1x_4x_5, x_4x_5 = x_2.$$

Демек, базистік функциялардың сандық мәндерінің матрицасында сәйкес келетін x_2 және x_4x_5 бағандары табылды. Жоспар жарамсыз, өйткені x_2 және x_4x_5 коэффициенттерінің бағалары бір-бірімен араласады.

Осылайша, бұл алгоритм маңызды айнымалылар тізіміне кірмейтін айнымалылардың қайсысы белгілі бір регрессия коэффициенттерінің мәндеріне әсер етуі мүмкін екенін анықтауға мүмкіндік береді.

6.3.7 Тәжірибелер жоспарларын іске асыру.

Бірінші реттік жоспарларды практикалық іске асыру келесі қадамдарды қамтиды:

а) жоспардың ортасын, яғни әрбір фактор үшін базалық деңгейлердің мәндерін таңдау. Алдында айтылғандай, жоспардың ортасы қазіргі уақытта ең көп қызығушылық тудыратын фактордың деңгейін білдіреді. Осы нүктеге жақын жерде регрессиялық модель құрастырылады;

б) тәжірибелерге дайындықтың маңызды кезеңдерінің бірі - әрбір фактор үшін вариациялау интервалдарын анықтау болып табылады. Бұл аралықтарды таңдауға мыналар әсер етеді:

- бөгеттердің қарқындылығы туралы ақпарат: басқа шарттар бірдей болған жағдайда, шудың әсері неғұрлым қуатты болса, сигнал/шу жеткілікті қатынасын қамтамасыз ету үшін вариациялау қадамын соғұрлым кең таңдау керек;

- жауап функциясының сызықты еместігінің дәрежесі туралы ақпарат: жауап функциясының сызықтықтан айырмашылығы неғұрлым көп болса, фактордың өзгеру интервалының мәні соғұрлым аз болуы керек;

- шешілетін мәселенің түріне байланысты вариациялау аралығының мәні тәуелді: регрессия моделі болжау мақсаттары үшін құрастырыладыма немесе вариация қадамы болжау мәселелеріне қарағанда әлдеқайда аз болуы керек

болатын экстремумдерді іздеу процедураларында қолданыладыма. Іздеу процедураларында тәжірибе жауап функция бетінің жергілікті қасиеттерін зерттеу үшін қолданылады яғни тәжірибе ортасына жақын жерде, вариациялық интервалдың кіші болатыны осыған байланысты. Болжамдық мақсатында құрылатын регрессиялық модельдердің тәжірибелері зерттеуге қызығушылық танытатын кең аймағында орындалады.

Вариация аралықтарын таңдау үшін пайдаланылатын бастапқы ақпарат шектеулі болғандықтан, бастапқы нұсқаны түзету қажет болуы мүмкін;

в) сынақтардың жалпы санын, яғни тәжірибелер жоспарын орнату; жоспардың спектрі жоғарыда аталған ережелерге сәйкес құрастырылғандықтан, бұл жағдайда мәселе қайталанатын матрицаның элементтерін таңдауға дейін қысқарады. Бұның мағынасы келесіде: жауап функцияның бақылау нәтижелері кездейсоқ сипатта болатындығына байланысты, жоспардың әрбір нүктесінде бір емес, m параллель тәжірибелерді жүргізу қажет, оның нәтижелерін орташалау процедурасы жауаптың шынайы мәні бағасының қатесін азайтуға мүмкіндік береді.

Бірінші ретті жоспарлауда әдетте тәжірибелердің біркелкі қайталануы қолданылады, яғни қайталау матрицасында барлық диагональ элементтері бір-біріне тең және m-ге тең. Яғни, жоспардың спектрлік матрицасының белгілі бір жолында фактор мәндерінің әрбір комбинациясы тәжірибе кезінде m рет қайталанады. Бұл жағдайда тәжірибелердің жалпы саны Nm-ге тең.

Тәжірибе жеке серияларда жүзеге асырылады. Әрбір серия жоспар матрицасының барлық жолдарына сәйкес келетін N тәжірибені қамтиды. Сериялар саны – m. Әрбір серияда сынақтарды орындау реті кездейсоқ болуы керек, ол үшін рандомизация қолданылады: кездейсоқ сандар кестелерін пайдалана отырып, тәжірибелерді жүзеге асырудың кездейсоқ реттілігі анықталады. Бұл процедураны m-реттік орындау сынақтардың әрбір сериясында жоспар спектрінің матрицасының жолдарының жүзеге асырудың кездейсоқ реттілігін қамтамасыз етеді. Бұл белгілі бір дәрежеде зерттеушінің біржақтылығы мен субъективтілігін, сонымен қатар, мысалы, агрегаттарды қыздыру немесе салқындату және процесті зерттеу кезіндегі басқа факторлармен байланысты жүйелі қателерді жоюға мүмкіндік береді.

Тәжірибелерді орындаудың түпкілікті нәтижесі - шығыс көрсеткішінің мәндерінің кестесі болып табылады; бұл одан әрі өңдеу үшін бастапқы материал болып табылатын объектінің жауабы.

Әртүрлі бірінші ретті жоспарларды бір-бірімен салыстыруға мүмкіндік беретін негізгі деректер 6.1-кестеде көрсетілген. Бұл ақпарат практикалық қажеттіліктеріңізге сәйкес келетін нақты жоспарды таңдауда пайдалы болуы мүмкін.

6.4 Бақылау сұрақтары

6.4.1 Тәжірибелерді жоспарлаудың негізгі принциптерін келтіріңіз.

6.4.2 Жоспарлардың оңтайлық критерийлерін анықтаңыз.

6.4.3 Бірінші ретті жоспарлардың анықтамаларын беріңіз: бір факторлық тәжірибе, толық факторлық тәжірибе, бөлшек факторлық тәжірибе.

6.4.4 Бірінші ретті жоспарлардың қасиеттері.

6.4.5 Бірінші ретті жоспарларын пайдалана отырып, қандай регрессия моделдерді алуға болады

6.1 кесте – Бірінші ретті жоспарлардың негізгі сипаттамалары

Жоспар	Вариация-лау деңгей-лерінің саны	Жоспардың спектріндегі нүктелер саны; қаныққанға жақындығы	Оңтайлылық критерийлері, басқа қасиеттері	Ықтимал модель көрінісі
Бір факторлық	3	2n; n>1 (2n>n+1) болғанда қанықпаған	Ротатабелді, ортогоналды	Тек сызықты
ТФТ 2^n	2	2^n; n>1 (2^n>n+1) болғанда қанықпаған	*A*-, *E*-, *D*-, *G*-оңтайлы, ротатабелді, ортогональды	Сызықтық және толық емес дәрежелі
БФТ 2^{n-p}	2	2^{n-p}; 2^{n-p} = n+1; болғанда қаныққан; 2n>n+1болғанда қанықпаған (ТФТ қарағанда қаныққанға жақындау)		Маңызды айнымалылар тізіміне сәйкес

6.4.6 ТФТ 2^n жоспарлау матрицасы қалай құрастырылады? ТФТ әдісі арқылы тәжірибе жүргізу тәртібін түсіндіріңіз.

6.4.7 Бөлшек реплика дегеніміз не? Бөлшек репликалардың рұқсат ету қасиеті немен анықталады?

6.4.8 Генерациялау қатынас дегеніміз не және ол қалай таңдалады?

6.4.9 БФТ жоспарлау матрицасы қалай құрастырылады?

6.4.10 Анықтаушы қатынас дегеніміз не және оның көмегімен бірлескен бағалар жүйесі қалай жасалады?

7. Екінші ретті жоспарлар

7.1 Жалпы ақпарат

Екінші ретті жоспарлар толық квадраттық көпмүше (екінші дәрежелі көпмүше) түріндегі регрессия моделін алуға арналған:

$$y(x) = b_0 + \sum_{i=1}^{n} b_i x_i + \sum_{i=1}^{n} b_{ii} x_i^2 + \sum_{\substack{i,l=1 \\ i<1}}^{n} b_{il} x_i x_l.$$

Мұндай жоспарлар объектінің айтарлықтай сызықтық емес қасиеттері бар екендігі алдын ала белгілі болған кезде немесе бірінші ретті жоспарлауды қолдану адекватты регрессия моделін алуға мүмкіндік бермеген және оны күрделендіру қажет болған кезде қолданылады.

Екінші ретті жоспарлар құрылымы жағынан күрделірек және оларды жүзеге асыруда тәжірибелердің санын көбейтуді талап етеді. Модель (7.1) $k = 1 + 2n + C_n^2 = \frac{(n+1)(n+2)}{2}$ мүшені қамтиды, бұл сызықты модельге қарағанда $\frac{n+2}{2}$ есе көп. Демек, жоспар спектріндегі нүктелерінің қажетті минималды саны артады. Квадраттық қатынасты алу үшін әрбір фактор кем дегенде үш деңгейде өзгеруі керек.

Барлық екінші ретті жоспарлар екі үлкен топқа бөлінеді: симметриялы және асимметриялқ.

Егер төмендегі шарттар орындалса жоспар *симметриялы* деп аталады:

$$\sum_{g=1}^{N} x_{ig} = \Lambda_2; \sum_{g=1}^{N} x_{ig}^2\, x_{jg} = \Lambda_{22}, i \neq j; \sum_{g=1}^{N} x_{ig}^4 = \Lambda_4;$$

$$\sum_{g=1}^{N} x_{ig} = \sum_{g=1}^{N} x_{ig}^3 = \sum_{g=1}^{N} x_{ig}\, x_{jg} = \sum_{g=1}^{N} x_{ig}^2\, x_{jg} \sum_{\substack{g=1 \\ i \neq j \neq k}}^{N} x_{ig}\, x_{jg} x_{kg} = 0$$

кез келген *i,j,k=1,2, ..., N, i≠j* үшін; *N* - жоспар нүктелерінің саны.

Яғни, барлық x_i факторлары немесе бірнеше факторлардың көрсетілген комбинациялары үшін жазылған қосындылар белгілі бір тұрақтыларға тең және бұл тұрақтылар фактор сандарының нақты мәндеріне қарамастан бірдей болады. Қарапайымдылық үшін параллельді тәжірибелер жүргізілмейді деп болжанады. Асимметриялы жоспарларда мұндай қасиеттер болмайды.

Симметриялық жоспарлар нүктелерді орналастыруда көбірек тәртіпке ие. Олар факторлардың әрқайсысының өзгерістер бағдарламасының ұқсастығымен, ал кейбір жағдайларда толық симметриялықпен сипатталады. Симметриялық жоспарлар үшін регрессия коэффициенттерін, олардың таңдамалы дисперсияларын және ковариацияларын бағалауға өте қарапайым қатынастар жарамды. Бұл Фишердің ақпараттық матрицасының блоктық құрылымы бар және оның элементтерінің көпшілігі нөлге, ал қалғандары Λ_2, Λ_{22} −ге тең болуына байланысты.

Асимметриялы жоспарлар симметриялыларға қарағанда көп жағынан ыңғайлы емес. Бірақ екінші жағынан, олар тәжірибелердің қажетті саны бойынша үнемді.

Екінші ретті жоспарларды пайдалану кезінде тәжірибені жоспарлау аймағы, яғни жоспардың нүктелері орналасатын факторлық кеңістіктің бөлігі туралы мәселе өте маңызды. Оның орнатудың үш опцияны нұсқасын атауға болады:

1) жоспарлау аймағы табиғи, бірінші кезектегі жоспарларды толтыру арқылы алынған. Бұл кезде координаталары белгілі бір оңтайлылық критерийлерін ескере отырып анықталатын жаңа, қосымша нүктелер бірінші кезеңде жоспарлау аймағынан, яғни гиперкубтың *-1<=xi<=+1, i=1,2,...,n*

шегінен шығуы мүмкін. Басқа қажетті нүктелерді қосу арқылы алынған жоспарлау аймағының өлшемдері оңтайлылық критерийінің түріне және факторларының *n* санына байланысты.

2) жоспарлау аймағы - гиперкуб $|x_{ig}| \leq 1, i = 1, \ldots, .n, g = 1, \ldots, N.$

3) жоспарлау аймағы – бірлік гипершар $x_1^2 + x_2^2 + \ldots + x_n^2 \leq 1$..

Соңғы екі нұсқа, біріншіден айырмашылығы, шешілетін мәселенің сипаттамаларына негізделген жоспарлау аймағының алдын ала таңдалуын қамтамасыз етеді. Гиперкуб түріндегі аймақ практикалық тұрғыдан ыңғайлы.

Қажет болса, жоспарды әрқашан басқа жоспарлау аймағы үшін қайта түрлендіруге болады. Мысалы, егер табиғи аймақта берілген жоспар болса, гиперкубқа орнатылған жоспарды алу үшін, әрбір x_i факторы үшін абсолютті мәндегі максималды мәнді тауып, x_i орнына $\frac{x_i}{\max x_{il}}$ қолдануға болады.

7.2 Екінші ретті симметриялық жоспарлар

Орталық композициялық (ОКЖ), яғни тізбектеліп құрылатын жоспарлар үш блоктан тұрады:

1) Жоспардың негізі немесе өзегі ТФТ 2^n немесе БФТ 2^{n-p} нүктелері болып табылады;

2) Тәжірибе центрінен ±α қашықтықта координата осьтерінде орналасқан «жұлдыздық» деп аталатын нүктелер; мұндай нүктелердің жалпы саны $N_\alpha = 2n$;

3) Жоспардың ортасындағы тәжірибелер, яғни нөлдік (орталық) нүктелер; мұндай нүктелердің саны $N_0 >= 1$.

ОКЖ нүктелерінің жалпы саны: $N = N_ф + N_á + N_0$.

Ерікті симметриялы ОКТ 7.1-кестеде көрсетілген.

7.1 кесте – ОКЖ-ның ерікті симметриялы жоспары

ОКЖ құрамдас бөліктері	g	x_1	x_2	…	x_n	Нүктелер саны
Жоспардың өзегі	1	-1	-1	..	-1	2n-p
	2	+1	-1	…	-1	
	…	-1	+1	…	-1	
	2^{n-p}	…	…	…	…	
«Жұлдыздық» нүктелері	2n-p +1	-α				2n
	2^{n-p} +2	+ α				
	…		- α			
	,,,		+ α			
	…				- α	
	2^{n-p} +2n				+ α	
Орталық нүктелері	2n-p +2n+1	0			0	N0
	…	…				
	2^{n-p} +2n+N_0	0			0	

Мысал ретінде 7.2 кестеде 5 деңгейде: -α, -1, 0, +1, + α өзгеретін үш (нормаланған) факторға екінші ретті жоспары келтірілген. α және N_0 нақты мәндері белгілі мәндері регрессиялық тәжірибелер оңтайлылы ейбір критерийлері негізінде таңдалады. Осыған байланысты ортогональды (ООКЖ) және ротатабелді (РОКЖ) орталық композиция жоспарлары қарастырылады.

7.2 кесте – Үш факторлар үшін бойынша екінші ретті ОКЖ

g	x_0	x_1	x_2	x_3	x_1^2	x_2^2	x_3^3
1	+1	-1	-1	-1	+1	+1	+1
2	+1	+1	-1	-1			
…	…	…	…	…			
8	+1	+1	+1	+1	+1	+1	+1
9	+1	-α	0	0	α^2	0	0
10	+1	+ α	0	0	α^2	0	0
11	+1	0	- α	0	0	α^2	0
12	+1	0	+ α	0	0	α^2	0
13	+1	0	0	- α	0	0	α^2
14	+1	0	0	+ α	0	0	α^2
15	+1	0	0	0	0	0	0

Ортогональды орталық композициялық жоспар. ООКЖ жағдайында жоспардың оңтайлылық критерийі жоспарлау матрицасының бағандарының ортогональдылығы болып табылады. Ортогональдыққа байланысты барлық коэффициенттердің бағалары бір-бірінен тәуелсіз анықталады. Әдетте, ООКЖ-де $N_0=1$. Ортогональдылық критерийі бойынша Фишер матрицасы диагональды болуы керек, ол үшін β_0 бос мүшесіне және $\beta_{ii}, i=1,\ldots,n$ квадраттық коэффициенттеріне сәйкес келетін бағандардың жұптық ортогональдылығын қамтамасыз ету шараларын қабылдау қажет, сондай-ақ өзара квадрат мүшелерге сәйкес келетін бағандары үшін де.

x_0 және x_i^2 бағандарын ортогональдау үшін келесі түрлендіру қолданылады:

$$\tilde{x}_i^2 = x_i^2 - \frac{1}{N}\sum_{g=1}^{N} x_{ig}^2 = x_i^2 - \bar{x}_i^2,$$

мұндағы N – жоспарлау матрицасының жолдарының жалпы саны.

Бұл жағдайда ортогональдылық шарты орындалады, яғни:

$$\sum_{g=1}^{N} x_{0g}\,\tilde{x}_{ig}^2 = \sum_{g=1}^{N} x_{0g}[x_{ig}^2 - \frac{1}{N}\sum_{i=1}^{N} x_{ig}^2] = \sum_{i=1}^{N} x_{ig}^2 - \frac{1}{N}N\sum_{i=1}^{N} x_{ig}^2 = 0. \quad (7.2)$$

Басқаша айтқанда, регрессия моделі келесі түрде ізделеді:

$$\varphi(\bar{x}) = \beta_0^{'} + \sum_{i=1}^{n} \beta_i\, x_{i-} \sum_{i=1}^{n-1} \sum_{j=1}^{n} \beta_{ij}\, x_i x_j + \sum_{i=1}^{n} \beta_{ii}\, (x_i^2 - \bar{x}_i^2); \beta_0^{'} = \beta_0 - \sum_{i=1}^{n} \beta_{ij} \bar{x}_i^2$$

(7.2) қатынас кез келген α үшін жарамды. Бірақ ерікті α үшін әртүрлі квадраттық айнымалыларға сәйкес келетін *F* матрицасының бағандары ортогональды емес болып қалады. Сондықтан ООКЖ-да α мәні дәл осы бағандардың ортогональдылық шартынан, яғни келесі шарттан таңдалады:

$$\sum_{g=1}^{N} \tilde{x}_{ig}^2\, \tilde{x}_{jg}^2 = \sum_{\substack{g=1 \\ i<j}}^{N} (x_{ig}^2 - \bar{x}_i^2)(x_{jg}^2 - \bar{x}_j^2) = 0.$$

Осыдан жұлдыздық нүкте енінің шамасын есептеу үшін теңдеу шығады:

$$4\alpha^4 + 4N_{\text{ф}}\alpha^2 - N_{\text{ф}}(N_\alpha - N_0) = 0,$$

мұндағы

α – жұлдыздық нүктенің ені;

N_ϕ – ТФТ (БФТ) нүктелерінің саны;

N_0 – орталық нүктелердің саны;

N_α – «жұлдыздық» нүктелерінің саны.

n=2,3,4 үшін ООКЖ параметрлері 7.3-кестеде берілген, 7.4-кестеде екі факторлы модель үшін базистік функциялар көрсетілген.

7.3 кесте - n=2,3,4 үшін ООКЖ параметрлері

n	α	$N_{\acute{a}}$	N_0	N_ϕ	N
2	1,0	4	1	4	9
3	1,215	6	1	8	15
4	1,414	8	1	16	25

7.4 кесте – n=2 үшін ООКЖ-ң базистік функциялары

g	x_0	x_1	x_2	x_1x_2	$(x_1^2 - \bar{x}_1^2)$	$(x_2^2 - \bar{x}_2^2)$
1	+1	-1	-1	+1	1/3	1/3
2	+1	+1	-1	-1	1/3	1/3
3	+1	-1	+1	-1	1/3	1/3
4	+1	+1	+1	+1	1/3	1/3
5	+1	-1	0	0	1/3	-2/3
6	+1	+1	0	0	1/3	-2/3
7	+1	0	-1	0	-2/3	1/3
8	+1	0	+1	0	-2/3	1/3
9	+1	0	0	0	-2/3	-2/3

ООКЖ-ң Фишер матрицасы келесі түрде болады:

$$\Phi = \left\| \begin{matrix} N & 0 & \dots & 0 & 0 & \dots & 0 & 0 \dots 0 \\ 0 & 2^{n-p} + 2\alpha^2 & \dots & 0 & 0 & \dots & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & \dots 2^{n-p} + 2\alpha^2 & & 0 & \dots & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & \dots & 0 & 2^{n-p} & \dots & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & \dots & 0 & 0 \dots & 2^{n-p} & 0 & \dots 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 2\alpha^4 \dots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & \dots & 0 & 0 & 0 \dots & 0 & 0 \dots 2\alpha^4 \end{matrix} \right\|$$

Кері матрица:

$$\Phi = \left\| \begin{matrix} \frac{1}{N} & 0 & \dots & 0 & 0 & \dots & 0 & 0 \dots 0 \\ 0 & \frac{1}{2^{n-p} + 2\alpha^2} & \dots & 0 & 0 & \dots & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & \dots \frac{1}{2^{n-p} + 2\alpha^2} & & 0 & \dots & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & \dots & 0 & \frac{1}{2^{n-p}} & \dots & 0 & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & 0 & 0 & \frac{1}{2^{n-p}} & & & 0 \dots 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & \frac{1}{2\alpha^4} \dots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \dots \frac{1}{2\alpha^4} \end{matrix} \right\|$$

Коэффициенттер бағаларының дисперсиялары келесі формулалар арқылы есептеледі:

$$\sigma^2_{b'_0} = \frac{\sigma^2_e}{N};\ \sigma^2_{b_i} = \frac{\sigma^2_e}{2^{n-p} + 2\alpha^2}\ ;$$

$$\sigma_{b_{ij}}^2 = \frac{\sigma_e^2}{2^{n-p}};\ \sigma_{b_{ii}}^2 = \frac{\sigma_e^2}{2\alpha^4}\ ; i,j = 1,\dots,n.$$

Регрессия коэффициенттерінің әртүрлі топтарын бағалау дәлдігі біркелкі емес (ТФТ-дегідей емес): еркін мүшенің β_0', коэффициенті ең дәл бағаланады, сызықты коэффициенттер - біршама дәлдігі төменірек, жұптық өзара әрекеттесудің β_{ij} коэффициенттері одан да нашар және квадрат мүшелерінің β_{ii} коэффициенттері ең аз дәлдікпен бағаланады.

n=2 үшін бізде:

$$\sigma_{b_0'}^2 = \frac{\sigma_e^2}{9};\ \sigma_{b_1}^2 = \sigma_{b_2}^2 = \frac{\sigma_e^2}{6}\ ; \sigma_{b_{12}}^2 = \frac{\sigma_e^2}{4};\ \sigma_{b_{11}}^2 = \sigma_{b_{22}}^2 = \frac{\sigma_e^2}{2}.$$

Регрессия коэффициенттерінің бағалары әдеттегідей ең кіші квадраттар әдісімен жүзеге асырылады және жоспардың ортогональдық қасиетіне байланысты олар үшін бірден өрнектерді жазуға болады:

$$b_0' = \frac{\sum_{g=1}^N \bar{y}_g}{N};\ b_i = \frac{\sum_{g=1}^N x_{ig}\bar{y}_g}{2^{n-p} + 2\alpha^2};\ b_{ik} = \frac{\sum_{g=1}^N x_{ig}x_{kg}\bar{y}_g}{2^{n-p}}, i,k = 1,2,\dots,n, i < k;$$

$$b_{ii} = \frac{\sum_{g=1}^N \tilde{x}_{ig}^2\bar{y}_g}{\sum_{g=1}^N (\tilde{x}_{ig}^2)^2} = \frac{\sum_{g=1}^N [\tilde{x}_i^2 - \frac{1}{N}(2^{n-p} + 2\alpha^2]\bar{y}_g}{2\alpha^4};$$

$$b_0 = b_0' - N\sum_{i=1}^n b_{ii}\left(\bar{x}_i^2\right) = b_0' - \frac{2^{n-p}+2\alpha^2}{N}\sum_{i=1}^n b_{ii}.$$

ООКЖ жоспарлау ротатабелді критерийін қанағаттандырмайды және жоспарлау орталығынан белгіленген қашықтықта әртүрлі бағыттар бойынша жауап болжамының ауытқулары айтарлықтай ерекшеленуі мүмкін - екі немесе одан да көп есе.

Ротатабелдің орталық композициялық жоспарлау. РОКЖ ротатабелдік критерийінің талаптары негізінде құрастырылады. Бұл жағдайда оңтайлылық критерийі *R*=const үшін $\sigma^2\{\hat{y}\} = const$. Жұлдыздық енінің α мәні ротатабелдің мүмкіндігін қарастыру арқылы келесі формуласы $\alpha = 2^{\frac{n-p}{4}}$ арқылы есептеледі.

Ротатабелді жоспарлар зерттеу нәтижелерін екінші ретті көпмүшелермен көрсетудің адекваттылығымен байланысты жүйелі қателерді азайтуға мүмкіндік беретін мағынада да оңтайлы болып табылады.

Орталық нүктелердің саны N_0 ротатабелді қасиетіне әсер етпейді және жоспарлау аймағының ішінде $\hat{y}$ шығыс мәнін болжау дәлдігін қамтамасыз ететіндей етіп таңдалады.

α_0 және N_0 қажетті мәндері 7.5-кестеде келтірілген, 7.6-кестеде екі факторлы модель үшін базистік функциялар көрсетілген.

Ротатабелді жоспар ортогоналдауды талап етпейді, сондықтан жоспарлау матрицасын құрастыру кезінде айнымалылардың түрлендірулері орындалмайды. Сондықтан Фишер матрицасы мен оған кері матрицаның жазбалары, коэффициенттерді бағалаудың дисперсиясы, жалпы жағдайда, өте күрделі өрнектермен көрсетіледі.

7.5 кесте - n=2,3,4 үшін РОКЖ параметрлері

n	α	$N_{\acute{a}}$	N_0	N_{ϕ}	N
2	1,414	4	5	4	13
3	1,682	6	6	8	20
4	2,000	8	7	16	31

7.6 кесте – Екі факторлы есептің жоспарлау матрицасы (*n=2, α=1,414; N_0=5)*

g	x_1	x_2	x_1x_2		x_1^2	x_2^2
1	+1	-1	+1		1	1
2	+1	+1	-1		1	1
3	+1	-1	-1		1	1
4	+1	+1	+1		1	1
5	-1,414	0	0		$(1,414)^2$	0
6	+1,414	0	0		$(1,414)^2$	0
7	0	-1,414	0		0	$(1,414)^2$
8	0	+1,414	0		0	$(1,414)^2$
9	0	0	0		0	0
10	0	0	0		0	0
11	0	0	0		0	0
12	0	0	0		0	0
13	0	0	0		0	0

Екінші ретті көпмүшелерді пайдаланып экстремум нүктені табу. Адекватты екінші ретті көпмүшелер оңтайлы нүктені (экстремум нүктесін) болжау үшін қолданылады, өйткені градиент шамасын анықтаудың шектеулі дәлдігіне байланысты оңтайлы іздеудің градиенттік әдістері қолайлы емес (сынақ қадамы шектеулі, шығысты өлшеу мәнінің дәлдігі шектеулі). Оңтайлы нүктені табу үшін градиент векторының құрамдас бөліктерінің мәндерін нөлге теңестіреді:

$$\frac{\partial y}{\partial x_i} = b_i + 2b_{ii}x_i = \sum_{k=1}^{n} b_{ik}x_k = 0, \;\; i,k = 1,2,\dots,n, i < k.$$

n фактордың санына сәйкес *n* сызықты теңдеулер жүйесін шешу арқылы салыстырмалы бірліктерде оңтайлы нүктенің *n* координатасын аламыз (бұл жағдайда координаталар басы тәжірибенің ортасында екенін ұмытпау керек, ал

факторлардың деңгейлері нормаланған). Содан кейін $\sigma^2\{\hat{y}\}$ дисперсиясының формуласы бойынша $\hat{y}_g$ мақсат функциясының болжау дәлдігі бағаланады.

Алынған көпмүшені пайдалана отырып, факторлық кеңістікте берілген x_{ij} нүктедегі шығыс шамасының мәнін болжау дәлдігінің формуласы:

$$\sigma^2\{\hat{y}_g\} = \sigma^2\{b_0\} + \sum_{i=1}^{n} x_{ig}^2\, \sigma^2\{b_i\} + \sum_{i=1}^{n} x_{ig}^2 x_{kig}^2\, \sigma^2\{b_{ik}\}, i < k.$$

7.3 Бақылау сұрақтары

7.3.1 Екінші ретті жоспарлау қандай жағдайларда қолданылады?

7.3.2 Орталық композициялық жоспарлар мен ТФТ және БФТ жоспарлауының айырмашылығы неде?

7.3.3 Екінші ретті симметриялы жоспарлардың артықшылығы неде?

7.3.4 ОКЖ нүктелерінің жалпы саны қалай анықталады?

7.3.5 ООКЖ және РОКЖ жоспарларында оңтайлылық критерийілері қандай?

7.3.6 ООКЖ -да жоспарлау матрицасының ортогональдылығына қалай қол жеткізіледі?

7.3.7 Алынған математикалық сипаттамасы бойынша экстремум нүктесі қалай табылады?

7.3.8 Екі факторлы және үш факторлы тәжірибе үшін ООКЖ жоспарлау матрицасын жазыңыз.

7.3.9 Екі факторлы және үш факторлы тәжірибе үшін РОКЖ жоспарлау матрицасын жазыңыз.

7.3.10 Факторлық кеңістіктің бірдей нүктелерінде параллель тәжірибелер жүргізу не үшін қажет?

8 Экстремалды тәжірибелерді жоспарлау

8.1 Тәжірибелік зерттеулердегі оңтайландыру мәселесі

Оңтайландыру мәселесі – теориялық және қолданбалы сипаттамасы жағынан ғылыми және инженерлік зерттеулер тәжірибесінде кездесетін ең маңызды және кең таралған мәселелердің бірі. Көптеген жағдайларда негізгі мақсат белгілі бір мағынада ең жақсы жағдайларды, шешімдерді, параметр мәндерін, фактор деңгейлерін анықтау болып табылады. Мұндай оңтайландыру мәселелері мысалы, келесі жағдайларда туындайды:

1) Ең жақсы сапамен және ең аз шығындармен максималды өнімге қол жеткізу қажет болатын әртүрлі технологиялық процестерді, агрегаттарды басқару кезінде;

2) Әртүрлі инженерлік құрылғыларды, аспаптарды жобалау кезінде ең жақсы өнімділік сипаттамаларын алатындай параметрлерді таңдау қажет болғанда;

3) Ең жақсы қасиеттері бар (мысалы, максималды беріктігі бар қорытпа) жаңа өнім үлгілерін жасау кезінде;

4) Регрессия моделін ең кіші квадраттар әдісімен анықтау немесе таңдалған критерийге сәйкес тәжірибелер жоспарын сандық тұрғыда құру кезінде таза есептік сипаттағы мәселелер туындайды.

Оңтайландыру мәселесі келесідей тұжырымдалады: жауап функциясы y (мақсат функциясы, оңтайландыру критерийі) өзінің экстремалды мәніне (минимумына немесе максимумына) жететін

$$y^* = extr\ y(\overline{x}),$$

зерттелетін объектінің $x_1 = x_1^*, x_2 = x_2^*, \dots, x_n = x_n^*$ басқарылатын факторларының мәндерін табу керек. Мұндағы $\overline{x} = (x_1, x_2, \dots, x_n)$.

Көптеген жағдайларда экстремалды нүктелер белгілі бір $\bar{x}$ шамаларын қойылатын шектеулермен анықталады.

Объектілердің көпшілігінде шектеулер бар болады және бұл жағдайларда ізденіс әдістерін қолданудың ерекшеліктері бар. Шектеудің екі түрі бар: факторлық және функционалдық.

Бөлек факторлардың рұқсат етілген мәндеріне қойылған шектеулер *факторлық шектеулер* деп аталады:

$$x_{min} \le x_i \le x_{max}, i = 1, \dots, n. \qquad (8.1)$$

Бұл түрдегі шектеулер кез келген оңтайландыру мәселесін шешу кезінде әрқашан болады.

Бірнеше факторлардың жиындар мәндеріне шектеулер

$$A_{jmin} \le \varphi_i(x_1, \dots, x_n \le A_{jmax},\ j = 1,2, \dots, k. \qquad (8.2)$$

Егер x_i факторлар мәндерінің кейбір комбинациялары рұқсат етілмейтін болса осындай шектеулер орын алады. Мысалы,

$$x_1 + x_2 + x_3 \le A_{max}, x_1 \cdot x_2 \le A_{min}, \sqrt{\sum x_i^2 \le A},$$ және т.б.

Келтірілген шектеу түрлері толығымен белгілі және аналитикалық түрде берілген.

Функционалдық шектеулер – объектінің сапалық немесе сандық аспектілерін сипаттайтын (зерттелетін объектің көп жауап функциялары болғанда) мақсатты функцияларға қойылатын шектеулер:

$$y_{pmin} \leq y_p \leq y_{pmax,} \quad (8.3)$$

мұндағы p – жауап функцияның нөмірі;

Яғни, басқа жауап функцияларына қойылған шектеулер бойынша бір жауап функциясының экстремалды мәнін алу қажет (мысалы, өнім көлемі жоспардан төмен болмауы керек, ал дайын өнімдегі қоспалардың пайызы рұқсат етілген шамадан жоғары болмауы керек).

Жалпы жағдайда (8.1-8.3) теңсіздіктер факторлық кеңістігіндегі рұқсат етілген Ω мәндер ауқымын анықтайды. Егер экстремум нүктесі Ω аймақтан тыс болса, онда оңтайландыру мәселесі шартты экстремумды іздеу ретінде тұжырымдалады, бұнда y максимумын $x \in \Omega$ ішкі жиында, яғни шектеу аймағының шекарасындағы ең жақсы мәнді табу қажет болғанда.

Шектеу жағдайында ізденіс оңтайландыру мәселесі келесідей тұжырымдалады: мақсатты іздеуді пайдалана отырып, шектеулер аймағындағы, яғни $\max\{y(x)|x \in \Omega$, мақсат функциясының максимум (минимум) нүктесінің координаттарын табу керек.

Барлық оңтайландыру әдістерін екі класқа бөлуге болады:

1) оңтайландыру мәселесі математикалық тұрғыдан анықталған және аналитикалық оңтайландыру әдістерін – дифференциалдық немесе вариациялық есептеулерді, сызықтық, динамикалық, бүтін бағдарламалауды қолдануға мүмкіндік беретін теориялық әдістер;

2) $y(\bar{x})$ жауап функциясы белгісіз және $x_1, x_2, \ldots, x_n$ шамалардың әртүрлі комбинациялары үшін y мәндерін өлшеуге болатын жағдайларда қолданылатын тәжірибелік әдістер. Мұндай жағдай физикалық объектілерді зерттегенде ғана емес, аналитикалық әдістер жарамсыз және сандық әдістер қолданылғанда (мысалы, тәжірибелер компьютерде жүргізіледі) теориялық есептерде де кездесуі мүмкін.

Нақты объектілерді зерттеу мәселелеріндегі оңтайландыру әдістерінің бірыңғай есептеу процедураларынан айырмашылығы басқарылмайтын факторлардың - кездейсоқ шулардың болуы.

Қарастырылып отырған оңтайландыру әдістері екі топқа бөлінеді: ізденіс әдістері; объектінің жүріс-тұрысын оптимум аймағында сипаттайтын эмпирикалық моделін алдын ала алуда негізделген әдістер.

Ізденіс әдістерінде жауап функцияның бетін тізбектеп жергілікті зерттеу жүргізіледі. Бұл жағдайда ізденіс алгоритміне байланысты тізбектелген процедуралар арқылы экстремалды мәнге қол жеткізіледі. Әдістердің екінші тобы оптимумы болжамды түрде орналасқан факторлық кеңістіктің белгілі бір аймағындағы жауап функцияның бетінің жергілікті емес, жалпы қасиеттерін зерттеуді қарастырады.

Осындай модель ретінде толық дәрежелі (квадраттық) көпмүше түріндегі регрессия модельдерін қолдануға болады. Экстремум нүктесін анықтау белгілі теориялық әдістерді (дифференциалдық есептеулер, сызықтық бағдарламалау және т.б.) қолдану арқылы жүзеге асырылады. Бұл топтың оңтайландыру әдістері үшін оптимумның координаталарын анықтауда ең жоғары дәлдікті

қамтамасыз ететін тәжірибені жобалау мәселесі үлкен қызығушылық тудырады.

Болашақта қарастырылған әдістер үшін мыналар болжанады:

а) зерттелетін объектінің жалғыз стационарлы жауап функциясы бар, (уақыттың ауытқуы жоқ);

б) жауап функциясының зерттелетін аймағында жалғыз экстремумі (унимодалдық қасиеті) бар (яғни жергілікті экстремум аралығы белгілі);

в) регрессиялық модельдерді қолдану арқылы оңтайландыру әдістерінде регрессиялық талдаудың алғышарттары орындалады.

8.2 Бір өлшемді ізденіс

Бір өлшемді іздеу әдістері бір факторлы оңтайландырудың қарапайым есептерін шешу үшін қолданылады және көп өлшемді іздеу процедураларына қосуға болады. Оларды қарастыру болашақта көп өлшемді ізденіс әдістерін жақсы түсіну үшін пайдалы. Бұл әдістер екі топқа бөлінеді: *шығарып тастау* әдістері және қадамдық *іздеу* әдістері.

Шығарып тастау әдістерінің мәні келесідей. Алдын ала белгілі (x_{min}, x_{max}) интервалдан x факторының кейбір мәнінде экстремум нүктесі орнатылады деп болжанады; бұл интервал *анықталмағандық аралығы* деп аталады. Тәжірибелердің ең аз санының көмегімен бұл интервалдың экстремумның орналасуы мүмкін емес бөліктерін кезекпен қараудан алып тастап, итервал ұзындығын максималды тарылту қажет. Жауап беру функциясы *y(x)* унимодалды және тұрақтылық сегменттері жоқ деп есептеледі (монотонды); бөгеттер жоқ. Бұл шарттарда $L = |x_{min} - x_{max}|$ бастапқы анықталмағандық интервалының ұзындығын азайту үшін $x_{min} < x_{(1)} < x_{(2)} < x_{max}$ орындалатындай кемінде екі тәжірибені кейбір $x_{(1)}, x_{(2)}$ нүктелерінде орындау қажет. Мұндай тәжірибенің келесі нәтижелері болуы мүмкін:

1) $y(x_{(1)}) < y(x_{(2)})$; максимум $x^* > x_{(1)}$ нүктесінде болады, ал бастапқы анықталмағандық аралығы жаңаға $(x_{(1)}, x_{max})$ айналады;

2) $y(x_{(1)}) > y(x_{(2)})$; максимум $x^* < x_{(2)}$нүктесінде болады, ал бастапқы аралықты жаңаға $(x_{(1)}, x_{max})$ ауыстыру керек;

3) $y(x_{(1)}) = y(x_{(2)})$; онда максимум $(\mathrm{x}_{(1)}, \mathrm{x}_{(2)})$ аралықта болады.

Тәжірибенің әрбір кезеңінде аталған нүктелерді орналастырудың әртүрлі әдістері бар. Оларды салыстыру үшін сәйкес тәжірибелік жоспардың тиімділік көрсеткіші енгізіледі: L бастапқы анықталмағандық интервалының ұзындығы мен N тәжірибеден кейін алынған L_N интервалының ұзындығының қатынасы: $E = \frac{L}{L_N}$. .

Бір өлшемді іздеудің кейбір әдістерін қарастырайық.

Тізбекті дихотомия әдісі. Бұл әдіс тәжірибенің әрбір кезеңінде бірден екі жаңа нүктені орналастыруды қарастырады, олар бір-бірінен ε қашықтықта анықталмағандық аралығының ортасына қатысты симметриялы орналасқан,

мұндағы ε кіші шама. Осы ε шаманың рұқсат етілген мәні – өлшеу құралдарының көмегімен аспаптық жолмен анықталуы мүмкін көрші бақылаулар арасындағы минималды айырмашылық. Алғашқы екі нүктенің координаталары:

$$x_{(1)} = \frac{(x_{max}+x_{min}-\varepsilon)}{2}; x_{(1)} = \frac{(x_{max}+x_{min}+\varepsilon)}{2}.$$

Келесі қадамдардағы тәжірибе өткізілетін нүктелерінің координаталары алынған анықталмағандық интервалдарының жаңа шекараларын ескере отырып, ұқсас формулалармен анықталады. k қадамнан кейінгі белгісіздік интервалының ұзындығы:

$$L_N = \frac{L}{2^k} + \left(1 - \frac{1}{2^k}\right) \cdot \varepsilon; N = 2k.$$

Осы жерден әдіс тиімділігі көрсеткішінің мәнін табуға болады.

Алтын қима әдісі. Бұл әдісте бірінші нүктенің координаталары мына формула бойынша табылады:

$$x_{(1)} = x_{min} + qL,$$

мұндағы *q=0,382*.

Бұл әдісте әрбір жаңа нүкте келесі интервалды екі бөлікке бөледі, сонымен бірге үлкен және кіші бөліктердің қатынасы бүкіл интервалдың оның үлкен бөлігіне қатынасына тең болады. Кесіндіні осылай бөлу «алтын қима» деп аталады. Бұл әдістің тиімділігі:

$$E = 1/(1-q)^{N-1} = 1/(1-0{,}618)^{N-1}.$$

Қарастырылып отырған әдістердің кемшілігі кездейсоқ бөгеттер кезінде олардың жұмыс қабілеттілігі жоғалады. Бұл жағдайларда қадамдық іздеу процедураларын қолдану қажет.

Тұрақты қадамы бар алгоритм. Бастапқы x_0 нүктесінің кіші аймағында

$$x_{(1)} = x_0 - \frac{\lambda}{2}, x_{(2)} = x_0 + \frac{\lambda}{2}$$

екі нүктелерінде екі тәжірибе жүргізіледі, $y(x_{(1)}), y(x_{(2)})$ анықталады.

Бұл мәндерді салыстыру арқылы қозғалыс бағытын анықтауға және осы бағытта қадамдарды орындауға болады, қадамдар саны алдын ала орнатылмайды. Үш дәйекті қадам үшін

$$y\left(x_{(m-1)}\right) < y\left(x_{(m)}\right), y\left(x_{(m+1)}\right) < y\left(x_{(m)}\right)$$

қатынастары орындалғанда, процедура аяқталады. Сонда x_m нүктесі ізделінген максималды мәнге сәйкес келеді.

Бұл алгоритм қарапайым, бірақ максимумды анықтау дәлдігі λ-мен байланысты. Бұл шама тым үлкен болуы мүмкін емес, бірақ оның шамадан тыс төмендеуі көптеген тәжірибелерді қажет етуді талап етеді.

Айнымалы қадамы бар алгоритм. Бұл әдістің идеясы: алдымен үлкен қадамдарды пайдалана отырып, экстремум нүктесі бар интервал туралы шамамен ақпарат алады, содан кейін қадамды азайту арқылы оны нақтылайды.

8.3 Көп өлшемді ізденіс әдістері

Көпөлшемді жауап функцияның экстремумын табу мәселесі тәжірибеде бірөлшемді есептерге қарағанда жиі кездеседі. Дегенмен, бұл мәселелерді шешу қиын:

- есептің өлшемі сандық түрде ұлғаяды, сондықтан есептеулер күрделене түседі;
- көпөлшемді жағдайда жауап функциясының унимодалдығы талап қасиетінің ықтималдығы төмен;
- жауап беру функциясы унимодалды болса да, іздеуге жауап беру бетінің бір өлшемді жағдайда аналогы болмайтын жергілікті қасиеттері әсер етуі мүмкін;
- әртүрлі іздеу алгоритмдерін салыстырудың қиындығы, өйткені кез келген алгоритм үшін басқаларға қарағанда қолайлырақ болатын жауап функциясын таңдауға болады, сонымен қатар бұл алгоритм жұмыс істемейтін болып қалатын басқа жауап функциясын құруға болады;
- экстремум нүктесінің локализациялау дәлдігіне қойылатын талаптардың жоғарылауы.

Ізденіс әдістерінде жауап бетін дәйекті жергілікті зерттеу жүргізіледі. Бұл жағдайда экстремалды мәнге дәйекті процедураларды қолдану арқылы қол жеткізіледі, соның ішінде:

а) арнайы жасалған тәжірибелердің нәтижелеріне сүйене отырып, оған жақын жерде тәжірибе жүргізілетін белгілі бір бастапқы нүктеден қозғалыс бағытын анықтау. Бұл бағыт берілген нүктенің жанындағы жауап бетінің жергілікті қасиеттеріне байланысты және табылған бағыттағы қозғалыс алдыңғы нүктеге қарағанда оңтайлық мәнге жақын жауап мәніне әкелетіндей анықталады;

б) табылған бағытта қозғалысты ұйымдастыру;

в) оңтайлы нүктеге жеткенше көрсетілген кезеңдерді қайталау.

Экстремумды іздеу әдістері екі үлкен топқа бөлінеді:

- градиенттік әдістері: (қарапайым градиент әдісі, тік көтерілу әдісі, жұптасқан градиент әдісі;
- градиенттік емес әдістер: Гаусс-Зайдель әдісі, кездейсоқ іздеу әдісі, симплекс әдісі.

Әр топтан ең көп тараған бір әдісті қарастырамыз. Бұл қарапайым, алгоритмдер.

8.3.1 Гаусс-Зайдель әдісі.

Бұл әдіс қарапайым градиенттік емес әдісі болып табылады. Гадиенттің емес іздеу әдістері олардың негізінде жасалған идеялардың алуан түрлілігімен ерекшеленеді. Бұл әдістерде сынақ қадамдарын жасау қажеттілігі бар, олардан кейін сынама нәтижелері қолайлы болған бағытта қозғалыстар орнydалады. Гаусс-Зейдель әдісінде экстремумды іздеу кезінде факторлардың әрқайсысы кезекпен өзгереді.

Гаусс-Зайдель әдісі әрбір x_i *(i=1,2,..,n).* факторы бойынша мақсат функциясының дербес экстремумдарын кезекпен табуды қамтиды. Сонымен бірге әрбір *i*-ші кезеңде *n-1* факторлары тұрақтанады және тек жалғыз *i*-ші фактор өзгереді.

Екі факторлы мысал үшін әдіс процедурасын қарастырмыз.

Гаусс-Зайдель әдісі арқылы максимумды табу мәселесі циклдерге біріктірілген бірнеше кезеңдермен шешіледі:

1-ші кезең.

а) жоғарыда сипатталған ережелер бойынша негізгі (бастапқы, базалық) нүкте таңдалады;

б) x_1 фактор бойынша Δx_1 вариациялау интервал таңдалынады. Әрине, вариация қадамы тым аз болмауы керек, әйтпесе экстремумға қарай қозғалыс баяу болады. Сонымен қатар, $\Delta x_i (i = 1,2,\ldots,n)$ вариация интервалында мақсаттық функцияның өзгеруі δу өлшеу қателігінен айтарлықтай үлкен болуы керек (кемінде 5-10 есе):

в) M_1 және M_2 сынақ нүктелерінің координаталары анықталады:

$$\bar{x}(M_1), = (x_{10} + \Delta x_1; x_{20} \ldots; x_{n0}), \bar{x}(M_2), = (x_{10} - \Delta x_1; x_{20} \ldots; x_{n0});$$

г) M_1 және M_2 нүктелерінде сынақ тәжірибелер орындалады (нәтижелердің дәлдігін жақсарту үшін параллелді тәжірибелер жүргізуге болады), $y(M_1), y(M_2)$ жауап функцияның мәндері өлшенеді;

д) алынған жауаптар салыстырылады және егер

$$y(M_2) > y(M_1),$$

онда $\overrightarrow{M_0 M_2}$ бағытында M_3 нүктесіне бір жұмыс қадамға Δx_1 жұмыс қозғалысын орындайды;

ж) осындай қадамдарды осы бағытта қандай да бір *k*-қадамда

$$y(M_k) < y(M_{k-1}),$$

шарт орындалғанша жалғастырады; яғни кезекті k-шы жұмыс нүктесінде жауаптың мәні төмендегенше - бұл дербес экстремумға жетудің белгісі. Дербес экстремум ретінде $y(M_{k-1})$ жауабы бар (k-1)--ші нүктені қабылдайды.

2-ші кезең. Ол 1-кезеңдегідей ретпен жүзеге асырылады, жалғыз айырмашылығы x_2 факторынан басқа барлық факторлардың тұрақтанады. Жаңа базалық нүкте үшін келесідей координаттары бар нүкте алынады:

$$\bar{x}(M_{k-1}), = (x_{10} \pm \Delta x_1 \cdot (k-2); x_{20} \ldots ; x_{n0}),$$

ал x_2 ұқсас шарттарда таңдалған Δx_2 вариациялау аралығының мәнімен өзгереді. x_2 факторы бойынша дербес экстремумға жеткенде, жаңа дербес экстремумның нүктесі жаңа базалық нүкте ретінде қабылданады.

Экстремумға қарай қозғалудың *бірінші циклы* x_n факторынан басқа барлық факторлар тұрақтанатын n-ші кезеңімен аяқталады. Ол үшін Δx_n вариация қадамы таңдалып, сынақ және содан кейін x_n факторы бойынша дербес экстремумға жеткенше жұмыс қозғалысы орындалады. Егер экстремум табылмаған болса, екінші іздеу циклі орындалады.

Екінші цикл, біріншісі сияқты, 1-ші кезеңнен басталады, онда x_1 факторы өзгереді, ал қалған $x_i (i \neq 1)$ тұрақталады, содан кейін әрбір n фактор үшін n кезең дәйекті түрде орындалады.

Экстремумды іздеу қозғалысы факторлық кеңістіктегі осындай нүктеге жеткенде аяқталады: бұл нүктеден кез келген бағытта барлық n факторлық x_i осьтер бойынша оң және теріс бағытта қозғалған кезде жауап функцияның мәндері кіші болады. Мұндай нүкте экстремум (максимум) ретінде қабылданады.

Гаусс-Зейдел әдісі практикалық қолданысқа қарапайым, бірақ іздеу траекториясы ең қысқа болмайтыны анық. Сонымен қатар, ізделінетін нүктесі жауап беру функциясының тар «тарағында» болса, іздеу процедурасының жалған тоқтатылуы мүмкін. Бұл жағдай осы әдістің қолданылу аясын айтарлықтай төмендетеді.

Басқа градиенттік әдістерден *кездейсоқ іздеу әдісі*, *симплекс әдісі* туралы айтуға болады.

8.3.2 Қарапайым градиент әдісі.

Барлық градиенттік әдістері жауап беру $y = y(x_1, .., x_n)$ функциясының градиентін алдын-ала анықтауға негізделген. Үздіксіз $y(\bar{x})$функцияның градианті деп келесідей координаталары бар вектор:

$$\frac{\partial y}{\partial x_i} |_{\bar{x}_0}, i = 1,2, \ldots, n,$$

аталады, мұндағы $\bar{x}^0 = (x_1^0, x_2^0, \ldots, x_n^0)$ – градиент мәні алынатын нүкте

Градиенттік әдістердің экстремумға дейінгі қозғалыстың әрбір кезеңінде вариация қадамдарын және жұмыс қадамдарын таңдау ережелерімен ерекшеленетін бірнеше түрлері бар. Барлық осы түрлерінің стратегиясының мәні мынада: әрбір кезеңде кезекті базалық нүктенің айналасында сынақ тәжірибелер ұйымдастырылады, оның нәтижелері бойынша градиенттің жаңа бағыты бағаланады, содан кейін осы бағытта бір жұмыс қадамы жасалады.

n-факторлық кеңістіктегі градиент векторы келесі қатынас арқылы анықталады

$$grady = \frac{\partial y}{\partial x_1}\bar{x}_1^0 + \frac{\partial y}{\partial x_2}\bar{x}_2^0 + \ldots + \frac{\partial y}{\partial x_n}\bar{x}_n^0,$$

мұндағы $\bar{x}_i^0 (i = 1,2,\ldots,n)$ - факторлық осьтердің бойында орналасқан бірлік бағыт векторлары;

$\frac{\partial y}{\partial x_i}$ - i-ші фактор бойынша мақсат функциясының дербес туындысы.

Сынақ тәжірибелер (әрбір факторлық оське параллель орналасқан және базистік нүкте арқылы өтетін түзулерде әрқайсысында екі нүкте) дербес туындылардың жуық бағасын алу үшін жүргізіледі.

Қарапайым градиент әдісі келесі процедура бойынша жүзеге асырылады:

а) бастапқы (негізгі) нүкте таңдалынады $\bar{x}_0 = (x_{10}, x_{20}, \ldots, x_{n0})$ - L_0 нүктесі;

б) жоғарыдағы ережелерді қолдана отырып, $(x_i, i = 1,2,\ldots,n)$ факторлардың әрқайсысы үшін Δx_i вариация аралығы таңдалынады;

в) сынақ нүктелерінің координаталары анықталады:

$$\bar{x}(L_1), = (x_{10} - \Delta x_1; x_{20}\ldots; x_{n0}), \bar{x}(L2), = (x_{10} + \Delta x_1; x_{20}\ldots; x_{n0}),$$

яғни бір фактор x_1 өзгереді, ал басқа факторлар базалық деңгейде тұрақтанады. Сол сияқты сынақ нүктелерінің координаталары $x_2, x_3, \ldots, x_n$. басқа факторлық осьтерге параллель бағыттар бойынша есептеледі. Тәжірибелер сынақ нүктелерінде қойылып, y мақсат функциясының мәндері алынады;

г) сынақ тәжірибелер нәтижелері бойынша L_0 нүктесіндегі градиент векторының құрамдас бөліктерінің бағалары әрбір i-ші фактор үшін есептеледі:

$$\mathrm{grad}y(L_0)|_{x_i} = \frac{\partial y}{\partial x_i} \approx \frac{\Delta y|_{x_i}}{2\Delta x_i} = \hat{a}_i.$$

Мысалы, x_1 факторы үшін L_1 және L_2 нүктелеріндегі тәжірибелер нәтижелері бойынша есептеу келесі формула бойынша орындалады:

$$\mathrm{grad}y(L_0)|_{x_1} \approx \frac{\Delta y|_{x_i}}{2\Delta x_i} = \frac{y(L_2) - y(L_1)}{x_1(L_2) - x_1(L_1)} = \hat{a}_i;$$

д) градиент бағыты бойынша жұмыс нүктесінің координаталарын табылады. Ол үшін $\rho_{гр}$ жұмыс қадамының параметрін таңдап, барлық $x_i, i = 1,2,\ldots,n$ факторлық осьтер бойынша бірінші жұмыс нүктесінің координаталары есептеледі:

$$x_{i1} = x_{i0} + \rho_{гр}\hat{a}_{i0};$$

е) бірінші жұмыс нүктесі жаңа базалық нүкте ретінде қабылданады және оның айналасында градиенттің жаңа бағытын бағалау үшін жаңа сынақ тәжірибелері ұйымдастырылады, содан кейін жаңа жұмыс қадамы орындалады.

Жалпы жағдайда әрбір k-ші жұмыс нүктесінде оның айналасындағы сынақ тәжірибелерінің нәтижелері бойынша $\hat{a}_{ik}$ градиентінің құрамдас бөліктерінің бағалары алынады және *(k+1)*-ші жұмыс қадамы (*k=0,1,2…,*) келесідей координаттары бар нүктеге орындалады:

$$x_{i,k+1} = x_{ik} + \rho_{гр}\hat{a}_{ik};$$

ж) келесі қадамда градиенттің барлық құрамдас бөліктері елеусіз аз болғанша жұмыс қозғалысы орындала береді, яғни барлық $\hat{a}_{i,k+1} \approx 0, (i = 1,2,\ldots,n)$ болғанша. Ол үшін келесі шарттың орындалуы жеткілікті:

$$\rho_{гр}\hat{a}_{i,k+1} < 1.$$

Егер сынақ тәжірибелерінің нәтижесінде (*k+1*)-ші жұмыс нүктесінде бұл шарт орындалса, онда экстремумға қозғалыс тоқтатылады және экстремум нүктесі ретінде осы жұмыс нүктесі алынады.

Жалпа қарапайым градиент әдісі тиімді және экстремумға қысқа жолмен өтуге мүмкіндік береді. Бірақ көп жағдайда ол сынақтардың айтарлықтай көп санын өткізуді қажет етеді.

Басқа градиенттік әдістер: *тік көтерілу әдісі, жұптасқан градиенттер әдісі.*

Айта кету керек, әртүрлі жағдайларда әртүрлі іздеу әдістері әртүрлі шуға қарсы қасиеттеріне ие. Әдістің шуға қарсы қасиеті оның жұмыс қозғалысының бағытын дұрыс бағалау қабілеті, сонымен қатар бөгеттердің болуына қарамастан жұмыс нүктесін экстремумдық аймағына тез және дәл жеткізу мүмкіндігі ретінде түсініледі. Гаусс-Зейдель әдісі өте шуылға төзімді. Қарапайым градиент әдісінің шуылға төзімділігі тым жоғары емес және оны арттыруға тырысқанда, тәжірибелер санын көбейту қажет болады. Алайда, ол практикада жиі қолданылады, әсіресе бөгеттер деңгейі төмен болған кезде.

Шартты экстремумды анықтау мәселесіне келетін болсақ, қарастырылып отырған іздеу әдістерін формальды қолдану табысқа әкелмейді. Әдістерді өзгерту керек. Шартты экстремумды табу мәселесін шешу үшін әртүрлі әдістер әртүрлі дәрежеде қолайлы. Мысалы, Гаусс-Зайдель әдісі (8.1)

сияқты шектеулермен жұмыс істеуге жақсы бейімделген; симплекс әдісі тіпті шектеулер кезінде де айтарлықтай тиімді және іздеу алгоритмі өзгермейді, өйткені шектеулерді қанағаттандырмайтын кез келген жаңа нүкте ең нашар ретінде жойылады. Қазіргі уақытта шектеулер бар болғанда экстремумды іздеу кезінде градиенттік әдістер кеңінен қолданылады.

Іздеу нүктесі Ω аймағының ішінде болғанша, осы әдістерге тән барлық процедуралар өзгеріссіз қалады. Жалғыз айырмашылық – әрбір сынақтың әрбір қадамында көрсетілген шектеулердің орындалуын тексеру қажеттілігі: аналитикалық түрде (8.1), (8.2) типті шектеулер үшін, тәжірибелік түрде – (8.3) типті шектеулер үшін. Бір немесе бірнеше шектеулердің бұзылғаны анықталса ғана, келесі іздеу қадамының қозғалыс бағытын анықтау алгоритмін өзгерту қажет.

8.4 Бақылау сұрақтары

8.4.1 Оңтайландыру есебін құрастырыңыз.

8.4.2 Оңтайландырудың ізденіс әдістері мен теориялық әдістердің айырмашылығы.

8.4.3 Тәжірибелік барлық ізденіс әдістеріне ортақ не бар?

8.4.4 Ізденіс әдістерінде сынақ тәжірибелер қандай мақсатпен орындалады?

8.4.5 Экстремумды іздеу әдістерінің базалық және жұмыс нүктелерінің айырмашылығы неде?

8.4.6 Бір өлшемді экстремумді іздеу әдістерін сипаттаңыз.

8.4.7 Экстремумге жетудің критерийі қандай?

8.4.8 Гаусс-Зайдель әдісі арқылы экстремумды табу алгоритмін сипаттаңыз.

8.4.9 Қарапайым градиент әдісінің негізгі идеясы мен процедурасы қандай?

8.4.10 Экстремумды табу есептерінде қандай шектеулер болуы мүмкін?

Әдебиеттер тізімі

1 Берикашвили В. Ш. Статистическая обработка данных, планирование эксперимента и случайные процессы: учебное пособие для бакалавриата и магистратуры – Москва: Издательство Юрайт, 2019.

2 Третьяк Л. Н. Основы теории и практики обработки экспериментальных данных: учебное пособие для бакалавриата и магистратуры – Москва: Издательство Юрайт, 2017.

3 Ибраева Л.К. Теория и техника инженерного эксперимента. Конспект лекций для магистрантов специальности 6М070200 – Автоматизация и управление . – Алматы: АУЭС, 2015.

4 Сидняев Н. И. Теория планирования эксперимента и анализ статистических данных: учебное пособие для магистров. – Москва: Издательство Юрайт, 2014.

5 Ибраева Л.К., Тохтасынова Н.Р. Теория и техника инженерного эксперимента. Методические указания к выполнению лабораторных работ для магистрантов специальности 6М070200 – Автоматизация и управление. – Алматы: АУЭС, 2017.

6 Карпушкин С., Глебов А.О. Теория инженерного эксперимента. Учебное пособие. – Тамбов: ТГТУ, 2017.

7 Реброва, И. А. Теория планирования эксперимента: учебное пособие. – Омск: СибАДИ, 2016.

8 Адамбаев М.Д. Теория и практика технического эксперимента в электроэнергетике. 6М070200 – Автоматизация и управление. – Алматы: КазНТУ, 2013.

9 Семенов В. А. Теория вероятностей и математическая статистика: Учебное пособие — СПб.: Питер, 2013.

10 Григорьев-Голубев В.В. Теория вероятностей и математическая статистика. Руководство по решению задач. – СПб.: БХВ-Петербург, 2014.

12 Астраханцева Л.Н., Байсалова М.Ж. Ықтималдық теориясы және математикалық статистика. Дәрістер жинағы. Алматы: АЭжБУ, 2013.

13 Ақанбай Н. Математиалық статистика.-А.: «Қазақ университеті», 2011.

14 Красовский Г.И., Филаретов Г.Ф. Планирование эксперимента. Изд-во БГУ им. В. И. Ленина, 1982. - https://www.studmed.ru/krasovskiy-gi-filaretov-gf-planirovanie-eksperimenta_1c0562109a1.html

15 https://www.matematicus.ru/tablitsy-po-teorii-veroyatnosti-i-statistiki

16 http://bek.sibadi.org/fulltext/esd104.pdf

17 https://www.chitai-gorod.ru/catalog/book/1053786/

18 http://elar.urfu.ru/bitstream/10995/39965/1/978-5-9904848-4-9_2015.pdf

Мазмұны

Кіріспе 3
1 Басқару объектілерді статистикалық модельдеу 4
1.1 Бақылау қатарларының арасындағы байланыстар түрлерін сипаттау 4
1.2 Тәжірибелік зерттеудер түрлерін классификациялау 7
1.3 Зерттеу нәтижелерінің қателіктері 1 8
1.4 Бақылау сұрақтары 10
2 Тәжірибелерді жоспарлау теориясының негізгі түсініктемелері 10
2.1 Анықтамалар 10
2.2 Факторлардың деңгейлері мен өзгеру аралықтарын таңдау. Тәжірибені қайтадан өндіру туралы түсінік 13
2.3 Бақылау сұрақтары 17
3 Статистикалық бағалау 18
3.1 Кездейсоқ шамалардың статистикалық бағаларының қасиеттері 18
3. 2 Максималды ықтималдық әдісі 21
3.3 Моменттер әдісі 23
3.4 Қалыпты таралудың параметрлерін нүктелік бағалау 24
3.5 Интервалды бағалар 30
3.6 Таңдамалы таралулар 32
3.7 Интервалдық бағаларды құру кезінде тәжірибені жоспарлау 40
3.8 Бақылау сұрақтары 41
4 Статистикалық гипотезалар 41
4.1 Статистикалық гипотезаларды тексеру 41
4.2 Маңыздылық критерийлері 44
4.3. Өрескел қателерді жою 48
4.4 Бақылау сұрақтары 50
5 Объектілердің регрессиялық модельдерін құру 50
5.1 Регрессиялық талдаудың элементтері 50
5.2 Регрессиялық модельдің параметрлерін бағалау 53
5.3 Регрессиялық модельді құру нәтижелерін статистикалық талдау 60
5.4 Бақылау сұрақтары 66
6 Регрессиялық талдаудың жоспарлары 66
6.1 Тәжірибелерді жоспарлаудың негізгі принциптері 66
6.2 Жоспарлардың оңтайлық критерийлері 69
6.3 Бірінші ретті жоспарлар 71
6.4 Бақылау сұрақтары 93
7 Екінші ретті жоспарлар 94
7.1 Жалпы ақпарат 94
7.2 Екінші ретті симметриялық жоспарлар 96
7.3 Бақылау сұрақтары 102

8 Экстремалды тәжірибелерді жоспарлау 102
8.1 Тәжірибелік зерттеулердегі оңтайландыру мәселесі 102
8.2 Бір өлшемді ізденіс 105
8.3 Көп өлшемді ізденіс әдістері 107
8.4 Бақылау сұрақтары 112
Әдебиеттер тізімі 113
Мазмұны 114

Ибраева Лида Куандыковна

ИНЖЕНЕРЛІК ТӘЖІРИБЕ ТЕОРИЯСЫ ЖӘНЕ ТЕХНИКАСЫ

Оқу құралы

Редактор Изтелеуова Ж.Н.

Басылымға қол қойылды ___.2022
Таралымы 100 дана. Пішімі 60×84 1/16

Типографиялық қағаз №2
Оқу- бас.ә. 7,0. Тапсырыс № _____
Бағасы 3500 теңге.

«Ғұмарбек Дәукеев атындағы АЭжБУ» КЕАҚ
Алматы қ., Байтұрсынов к., 126/1

«Ғұмарбек Дәукеев атындағы Алматы энергетика және байланыс университеті» коммерциялық емес акционерлік қоғамының көшірме-көбейту бюросы
Алматы қ., Байтұрсынов к., 126/1