Математикалық статистика негіздері

Дәріс жоспары:

1. Бас және таңдама жиынтық.
2. Таңдаманың статистикалық таралуы, дискретті және интервалды вариациялық қатар.
3. Полигон және гистограмма.
4. Таңдама параметрлері.

Қандай да бір сапалық немесе сандық белгілермен сипатталатын нысандар жиыны *статистикалық жиынтық деп аталады.*

Тексерілуге жататын (ең болмағанда, теория жүзінде) барлық нысандардан тұратын статистикалық жиынтық *бас статистикалық жиынтық* деп аталады.

Бас жиынтықтан кездейсоқ түрде таңдалынып алынған қандай да бір нысандар санынан тұратын статистикалық жиынтық *таңдама жиынтық* немесе жәй *таңдама деп аталады.*

Бас жиынтық деп таңдама жүргізілетін объектілер жиының айтамыз.

Бас жиынтықтағы элементер санының оның көлемі деп аталады.

Таңдама элементтерінің саны оның көлемі деп аталады.

Таңдама жиынтықты зерттеу арқылы барлық бас жиынтық жөнінде қорытынды жасалынатын статистикалық зерттеу әдісі *таңдама әдіс* *деп аталады*.

Статистикалық әдістердің көмегімен таңдама қасиеттері бойынша бас жиынтықтың қасиеттері туралы анық тұжырым жасау үшін, таңдама ***репрезентативті*** болу керек, яғни мүмкіндігі бойынша бізге қажет бас жиынтықтың қасиеттерін бейнелеу керек.

Таңдау тәсілдері:

I. Бас жиынтықты бөлшектерге бөлуді талап етпейтін таңдау,

бұған жататындар:

а) жәй кездейсоқ қайталанбайтын таңдау;

б) жәй кездейсоқ қайталанатын таңдау;.

II. Бас жиынтық бөлшектерге бөлінетін таңдау,
бұған жататындар:

а) типтік таңдау;

б) механикалық таңдау;

в) сериялық таңдау.

Жай кездейсоқ таңдама деп барлық бас жиынтықтан обьектілерді бір-бірден алатын таңдаманы атайды. Егер алынған карточкаларды бумаға қайтармаса, онда таңдама жай кездейсоқ қайталанымсыз болады.

Типтік таңдама деп, обьектілер бас жиынтықтың барлығынан емес, оның әрбір «типтік» бөлігінен алынатын таңдаманы атайды.

Механикалық таңдама деп бас жиынтық таңдамаға қанша обьект қажет болса, сонша топқа бөлінетін таңдаманы атайды, әрбір топтан бір обьект алынады.

С**ериялық таңдама** деп бас жиынтықтан обьектілерді бір-бірден емес, жаппай зерттеуге ұшырайтын обьектілер «сериялармен» таңдап алатын таңдаманы атайды.

Таңдаманың статистикалық таралуы.

Алынған таңдамалық зерттеулерді жүйелендіруде таралудың статистикалық дискретті және интервалды қатарлар қолданылады.

1. **Дискретті статистикалық таралу. Полигон.**

Бас жиынтықтан таңдама алынсын, және x_1-n_1 рет, x_2-n_2 рет, ..., x_k-n_k рет қайталанады,

x_1 мәндерін варианталар деп, ал өсу ретімен жазылған варианталар тізбегін вариациялық қатар деп атайды.

$\sum n_i = n$ -таңдама көлемі.

Қарастырылатын мәндер санын жиіліктер, ал олардың таңдама көлеміне қатынасын салыстырмалы жиіліктер

$\frac{n_i}{n} = w_i$ - деп айтады.

Анықтама. Таңдаманың статистикалық таралуы деп варианталар x_m мен оларға сәйкес жиіліктер n_i немесе салыстырмалы жиіліктердің w_i тізбегін айтады.

x_1	x_2	...	x_i
n_1	n_2	...	n_m

түріндегі кесте дискретті статистикалық қатар деп аталады. Жиіліктердің қосындысы тандама көлеміне тең $\sum n_i = n$.

x_1	x_2	...	x_m
w_1	w_2	...	w_m

кестесі салыстырмалы жиілік статастикалық заңы деп аталады.

Салыстырмалы жиіліктердің қосындысы 1-ге тең $\sum w_i = 1$

Көрнекілік үшін статистикалық таралудың түрлі графиктер салынады, соның ішінде полигон мен гистограмма тұрғызылады.

Жиіліктер полигоны деп (x_1, n_1), (x_2, n_2), ..., (x_k, n_k) нүктелерін қосатын сынық сызықты айтады жиіліктер полигонын тұрғызу үшін абциссалар осінде x_i варианталарын, ал ординаталр осінде-оларға сәйкес n_i жиіліктерді орналастырады.

(x_i, n_i) нүктелерін түзудің кесінділерімен қосып, жиіліктер полигонын салады.

Салыстырмалы жиіліктер полигоны деп кесінділері (x_1, w_1), (x_2, w_2),..., $(x_к, w_к)$, нүктелерін қосатын сынық сызықты айтады. Салыстырмалы жиіліктер полигонын тұрғызу үшін абциссалар осінде x_i варианталарын, ал ординаталар осінде оларға сәйкес W_i салыстырмалы жиіліктер полигонын тұрғызады.

Таралудың статистикалық интервалдық қатары.

Егер бізді қызықтыратын бас жиынтықтың Х белгісі үзіліссіз болса, онда варианталар интервалдарға топтастырылады.

Статистикалық таралуды интервалдар тізбегі және оларға сәйкес жиіліктер (интервалға сәйкес жиілік ретінде осы интервалға түскен жиіліктер қосындысын қабылдайды) тізбегі түрінде беруге болады.

Ескерту. Көбіне барлық і үшін $h_i - h_{i-1} = h$, яғни топтастыру тең h қадаммен алынады. Бұл жағдайда a, k, h_i табу үшін келесі ұсынысты жетекшілікке алуға болады.

1. R =X_{max} –X_{min} табамыз, мұндағы R(размах) – ең үлкен және ең кіші варианталардың айырымы.

2. $h = \frac{R}{k}$ ***k-топтар саны, h-қадам.***

3. $k \geq 1 + 3{,}32\lg n$ **(Стерджес формуласы)**

4. a=x_{min}, b=x_{max}

5. $h_i = a + ih, \quad i = 0,1,...k$

Алынған топтастыруды жиілік кестесі түрінде ұсыну қолайлы. Бұл кестені **таралудың статистикалық интервалдық қатары** деп атайды.

Топтастыру-дың интервалы	$[h_0, h_1)$	$[h_1, h_2)$	...	$[h_{k-2}, h_{k-1})$	$[h_{k-1}, h_k]$
Жиіліктер	n_1	n_2	...	n_{k-1}	n_k

$$\sum n_k = n$$

Осы кестені n_i жиіліктерді салыстырмалы жиіліктермен алмастырып мынадай түрде жазуға болады:

Топтастыру интервалы	$\left[h_0, h_1\right)$	$\left[h_1, h_2\right)$	...	$\left[h_{k-2}, h_{k-1}\right)$	$[h_{k-1}, h_k]$
Салыстырмалы жиіліктер	w_1	w_2	...	$w_{к-1}$	$w_л$

$$\sum w_i = 1.$$

Жиіліктердің графиктік түрі - **жиіліктер гистограммасы** деп аталатын арнайы график болып табылады.

Жиіліктер гистограммасы деп табандары h-қа, биіктіктері $\frac{n_i}{h}$ (жиілік тығыздығы) қатынасына тең тіктөртбұрыштардан тұратын баспалдақты фигураны айтады.

Үзіліссіздік белгісі жағдайында гистограмма салған жөн, ол үшін белгінің барлық бақыланатын мәндер жататын интервалды ұзындығы һ-қа тең бірнеше дербес (жеке) интервалдарға бөліп, әрбір дербес n_i интервал үшін *i*-ші интервалға түскен варианталар жиіліктерінің қосындысын табады. і-ші дербес төртбұрыштың ауданы - і-ші интервалға түскен варианталар жиіліктерінің қосындысына тең, сондықтан жиіліктер гистограммасының ауданы

$$h \cdot \frac{n_i}{h} = n_i$$

барлық жиіліктердің қосындысына тең, яғни таңдаманың көлеміне тең.

Салыстырмалы жиіліктер гистограммасы деп табандары h-қа, биіктіктері $\dfrac{w_i}{h}$ (салыстырмалы жиілік тығыздығы) қатынасына тең тіктөртбұрыштардан тұратын баспалдақты фигураны айтады.

Тандама медиана – вариациалық қатардың ортасындағы варианта, тандаманың сол және он жағынан бірдей қашықтықта орналасқан.

Тандама мода – ықтималдығы көбірек, ең үлкен жиіліктері бар варианта.

Бас орта

$\overline{x}_Б$ **бас орта** деп бас жиынтық белгісінің орта арифметикалық мәнін айтады:

$$\overline{x}_Б = \frac{x_1 + x_2 + ... + x_N}{N}$$

мұнда N- жиынтықтың көлемі.

Таңдама орта

X сандық белгісіне қатысты бас жиынтықты зерттеу үшін n көлемді таңдама алынсын.

$\overline{x_T}$ **таңдама орта** деп таңдама жиынтық белгісінің орта арифметикалық мәнін айтады:

$$\overline{x_T} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n}$$

немесе

$$\overline{x_T} = \frac{\left(\sum_{i=1}^{k} n_i x_i\right)}{n}$$

Бас дисперсия

Бас жиынтықтың X сандық белгісі мәндерінің өз орта мәнінің маңайында шашырауын сипаттау үшін бас дисперсия сипаттамасы енгізіледі.

Егер N көлемді бас жиынтық белгісінің барлық $x_1, x_2, ..., x_N$ мәндері әртүрлі болса, онда

$$D_Б = \frac{\sum_{i=1}^{N}(x_i - \overline{x}_Б)^2}{N}$$

Егер белгінің барлық $x_1, x_2, \ldots, x_к$ мәндерінің сәйкес жиіліктері N_1, N_2, …, N_k бар болса, және $N_1+N_2+ \ldots,+N_k=N$, онда

$$D_Б = \frac{\sum_{i=1}^{k} N_i (x_i - \bar{x}_Б)^2}{N}$$

Бас жиынтықтың сандық белгісі мәндерінің өз орта мәнінің маңайында шашырауын сипаттау үшін дисперсиядан басқа орта квадраттық ауытқуды пайдаланады.

Бас орташа квадраттық ауытқу деп бас дисперсиядан алынған квадрат түбірді айтады:

$$\sigma_{Б} = \sqrt{D_{Б}}$$

Таңдама дисперсия $Д_т$ деп белгінің бақыланатын мәндерінің $\overline{x}_T$ орта мәнінен ауытқу квадраттарының орта арифметикалық мәнін айтады.

Егер n көлемді таңдаманың барлық $x_1, x_2, \ldots .x_n$ белгілерінің мәндері әр түрлі болса, онда

$$D_T = \frac{\sum_{i=1}^{n} (x_i - x_T)^2}{n}$$

Егер $x_1, x_2, \ldots .x_n$ мәндерінің жиіліктері бар және сәйкесінше $n_1, n_2, \ldots, n_k$ болса, мұндағы $n_1 + n_2 + \ldots + n_k = n$, онда

$$D_T = \frac{\sum_{i=1}^{n} n_i (x_i - x_T)^2}{n}$$

Теорема: Дисперсия таңдама мәндерінің квадраттарының орта мәні мен орта мәнінің квадратының айырымына тең:

$$D = \overline{x^2} - [\,\overline{x}\,]^2$$

Таңдама тексерудің қателіктері:

- Кездейсоқ,
- Кездейсоқ емес, яғни таңдау дұрыс жүргізілмейді:
  - таңдаудың араласқан әдісі қолданылады
  - таңдама негізгі бас жиынтықтан жүйелі түрде ерекшеленеді.

Қалыпты таралудың негізгі сипаттамалары:

Сандық сипаттамалардың теңдігі (орта мән, мода және медиана өз ара тең);

- орта мәннен ауытқудың симметриялылығы;
- қисық астындағы жалпы аудан 1 ге тең;
- қисықтың ұштары екі бағытта да абцисса осіне үздіксіз жақындай отырып, алайда ешқашан онымен жанаспай шексіздікке ұмтылады.
- қисықтың түрі бас жиынтықтың орта квадраттық ауытқуымен анықталады;
- орта квадраттық ауытқуы аз таралуға жіңішке, жоғары созылған қисықтар, ал орта квадраттық ауытқуы үлкен таралуға жазыңқы қисықтар сәйкес келеді.

Қалыпты таралудың негізгі сипаттамалары:

- барлық мәндердің 68,26% ±σ аралығында жатады (орта мәннен ±1 орта квадраттық ауытқу);

-барлық мәндердің 95,44% ±2σ аралығында жатады (орта мәннен ±2 орта квадраттық ауытқулар);

-барлық мәндердің 99,73% ±3σ аралығында жатады (орта мәннен ±3 орта квадраттық ауытқулар).

НАЗАРЛАРЫҢЫЗҒА РАХМЕТ