Таңдама. Таңдаманың сандық сипаттамалары.

1.Бас жиынтық және таңдама.

2. Вариациялық қатар.

3. Полигон және гистограмма.

4. Таңдаманың сандық сипаттамалары.

5. Таңдама орта және дисперсияны есептеу көбейту әдісі.

1.Бас жиынтық және таңдама.

Көптеген кездейсоқ құбылыстардың бағынатын заңдылығын анықтау тәселесі, бақылау нәтижесінің статистикасын зерттейтін ықтималдық теориясының әдістіне негізделгегн.

Математикалық статистииканың мынадайесептерін қарастырамыз:

1. Жүргізілген тәжірибенің немесе бақылаудың нәтижесінде, алынған статистикалық мағлұматтарға жинау жән оларды топқа бөлу әдістерін көрсету.
2. Зерттеудің мақсатына байланысты статистикалық мағлұматтарға анализ жасау әдістерін іздеп табу.

Математикалық статистиканың міндеті ғылыми және трактикалық тұжырымдар жасау үшін сиатистикалық мағлұматтарды жинау.

Кейде қажетті белгісі бойынша , заттар жиынтығындағы әрбір затты түгелдей тексеруге тура ккеледі. Іс жүзінде бұлай түглдей тексеру өте сирек зерттейді. **Таңдап алынған жиынтық** немесе жай ғана **таңдамалы** деп, кездейсоқ таңдап алынған заттардың жиынын айтады.

Бас жиынтық деп, таңдамалы жасалатын заттардың жиынтығын айтады.
Көлем жиынтығы деп, осы жинақтың ішіндегі заттардың санын айтады. Мысалы, егер 1000 бөлшектен тексеру ғғшін 100 бөлшек бөліп алынса, онда бас жиыниықтың көлемі N=1000, ал таңдамалы жиынтығының көлемі n=100.
2. Вариациялық қатар.
Бақылау нәтижесінде қарастырылып отырған белгінің жиынтықтағы, әрдір бірлікке қатысты сандық немесе сапалық өзгерісі туралы мәлімет аламыз. Статистикалық бақылаудың мақсаты сол жиынтықта белгінің өзгеруін (**вариасиясын**) шешу. Ал белгінің мүмкін мәндерін статистикада **варианта** деп атайды. Варианталар сандық (дискретті немесе үздіксіз) болатын мүмкіндігін көрдік.

Тәжірибе жүргізілгенде белгі мәндері қалай болса солай орналасуы мүмкін. Мысалы, тексерілген 100 вал диаметрі см-мен 15, 12, 16, 12, 13, 14, 16, 13, 14, 12 болып шықты. Мұны реттеп жазсақ 12, 12, 12, 13, 13, 14, 14, 15, 16, 16 болар еді. Мұны ықшамдап кесте түрінде жазсақ, мынадайболады:

x	12	13	14	15	16	Σ
n_i	3	2	2	1	2	10

Бұл кестенің жоғарғы жолын белгі мәндері (варианталары), ал төменгі жолында әрбір мәннің неше рет кездескені келтірілген. Осылай реттелген кестені вариацмялық қатар деп атайды.

Әдетте белгіні (вариантаны) кездейсоқ шамалар сияқты x, y, z, ..., y_k, z_1, z_2, …., z_k арқылы белгілейміз. Варианта қайталап отыруы мүмкін. Ол қайталаулардың абсалютті санын (жиілігін) n_1, n_2… n_k деп белгілесек, онда вариациялық қатардың жалпы түрін мына кесте көрсетеді.

2 – кесте

x	x_1	x_2	...	x_k	Σ
n_i	n_1	n_2	…	n_k	n

Мұнда, x_i – варианталары сәйкес;

n_i – жиіліктер; $n = \sum_{n=1}^{k} n_i$ - вариация қатардың көлемі.

Іс жүзінде варианта абсалютті жиілікпен қатар салыстырмалы жиілік түрінде де беріледі. Бұл жағдайда 2 кесте былай жазылады:

3 – кесте

x_i	x_1	x_2	...	x_k	Σ
w_i	w_1	w_2	…	w_k	1

Мұндағы, $wi = \frac{n_i}{n}$ - салыстырмалы жиілік

ал, $n = \sum n_i \ \sum w_i \ = 1$ - салыстырмалы жиіліктердің қосындысына тең бірге

Егер вариация үздіксіз өзгеретін болса, онда вариациалық қатарды интервалдар бойынша құруға тура келеді.

Жалпы түрде интервалдық қатар мынадай болады:

4 – кесте. Жиіліктің интервалдық түрі.

x	$(x_1;x_2)$	$(x_2;x_3)$	$(x_3;x_4)$	…	$(x_m;x_{m+1})$
n_i	n_1	n_2	n_3	…	n_m

немесе

5 – кесте. Салыстырмалы жиіліктің интервалдық түрі.

x $(x_1;x_2)$ $(x_2;x_3)$ $(x_3;x_4)$ … $(x_m;x_{m+1})$

w_i	w_1	w_2	w_3	…	w_m

Мұндағы $(x_1;x_2)$, $(x_2;x_3)$, $(x_3;x_4)$ ..., $(x_m;x_{m+1})$ аралықтары белгінің мүмкін мәндері жататын интервалды $к_1= x_2-x_1$; $к_2= x_3-x_2$, $к_m= x_{m+1}-x_m$ – айырымдары интервалды сипаттайды.

3. Полигон және гистограмма.

Жиілік полигоны деп $(x_i; n_i)$, $(x_2; n_2)$, ..., $(x_к; n_к)$, нүктелерін кесінділермен қосқанда шыққан сынық сызықтарды айтады. Полигон салу үшін абцисса осіне x_i – варианттары, ал оларға сәйкес жиіліктері ордината осіне салынады. $(x_i; n_i)$ нүктелерін кесінділермен қоссақ жиілік полигонын аламыз (1 сурет). Салыстырмалы жиілік полигоны деп $(x_1; w_1)$, $(x_2; w_2)$, ..., $(x_к; w_к)$ нүктелерін кесінділермен қосқанда шыққан сынық сызықтарды айтады. Салыстырмалы жиілік полигонын салу үшін, абцисса осіне x_i – варианттары, ал оларға сәйкес w_i – салыстырмалы жиіліктер ордината осіне салынады. $(x_i; w_i)$ – нүктелерін кесінділермен қоссақ **салыстырмалы жиілік** полигонын аламыз.

Жиілік гистограммасы деп табандарының ұзындықтары h – қа тең дербес интервалдан, ал биіктері $\frac{n_i}{h}$ қатынасындай болатын тік төртбұрыштардан құрылған сатылы фигураны айтады.

Жиілік гистограммасын салу үшін абцисса осіне дербес интервалды, ал олардың жоғарғы жағынан абцисса осіне паралелль және одан ара қашықтықтары $\frac{n_i}{h}$ - қа тең кесінділер жүргіземіз. Мұндағы $\frac{n_i}{h}$ - жиілік тығыздығы. і – ші тік төртбұрыштың $h \bullet \frac{n_i}{h} = n_i$ - ге тең. Сонымен жиілік гистограммасының ауданы барлық жиіліктердің қосындысына тең, яғни**таңдау көлеміне тең**.

Салыстырмалы жиілік гистограммасы деп, табндарының ұзындықтары h – қа тең интервалдар, ал биіктіктері $\frac{w_i}{h}$ (салыстырмалы жиілік тығыздығы) қатынвсына тең тік төртбұрыштардан құрылған фигураны айтады.

Салыстырмалы жиілік гистограммасын салу үшін абцисса осіне интервалдарды, ал оның жоғарғы жағынан абцисса осіне паралелль және одан қашықтықтары $\frac{w_i}{h}$ қатынасына тең кесінділерді саламыз.

і – ші тік төртбұрыштың ауданы: $h \bullet \frac{w_i}{h} = w_i$ - ге тең.

Сонымен салыстырмалы жиілік гистограммасының ауданы, барлық салыстырмалы жиіліктер қосындысына, яғни бірге тең.

Мысал. Берілген 3 – ші суретке сәйкес көлемі n-100 жиілік үлестірімінің гистограммасы бейнеленген.

Ұзындығы h=5	Интервалдардың вариант жиіліктерінің қосындысы	жиілік

интервалдар

Статистикалық үлестірімнің сандық сипаттамаларын және оларды есептеу формулаларын қарастырайық.

1. **<u>Арифметикалық таңдамалы орта</u>**. Белгінің (X) арифметикалық ортасы $\overline{x}_T$ деп варианталардың жалпы санына (таңдаманың көлеміне) қатынасын айтады, яғни (егер барлық варианталар әртүрлі болса):

$$\overline{x}_m = \frac{\sum_{i=1}^{\kappa} x_i}{n}$$

x_i – варианталар (белгі мәндері): $n = \sum_{i=1}^{\kappa} n_i$ - таңдама көлемі.

Егер таңдама вариациялық қатармен беоілсе, онда

x	x_1	x_2	…	$x_к$
n_i	n_1	n_2	…	$n_к$

Мұнда,

n_i – варианталар салмағы (жиілігі); $n = \sum n_i$ - таңдамалы көлемі;

x_i – варианталар.

2. **Мода (M_0)**. Берілген вариациялық қатардың ең жиі кездесетін вариантасын <u>мода</u>деп атайды. Басқаша айтқанда, ең жоғары жиілікке сәйкес варианта мәні мода болады.
3. **Медиана (M_e).** Жиынтықты тең етіп екіге бөлетін белгі мәнін <u>медиана </u>деп атаймыз. Егер белгінің өзгеруші мәндері тақ болып, ұлғаю ретімен орналасса x_1, x_2, ..., x_{m-1}, x_m, x_{m+1}, ..., x_{2n-1}, онда бұл үйлестіру үшін M_e медианасы x_m вариантасына тең, яғни $M_e = x_m$, өйткені $M_e = x_m$ – нен төмен де жоғары да белгінің саны бірдей m-1 мәндері орналасқан.

Ал варианта саны жұп болса, x_1, x_2, ..., x_{m-1}, x_m, x_{m+1}, ..., x_{2m}, онда бұл жиынтықты тең екіге бөлетін медиана мәні (x_m, x_{m+1}) аралығында болады. Бұл жағдайда медиана M_e – нің мәні осы екі вариантаның арифметикалық ортасы болады, яғни

$$M_e = \frac{x_m + x_{m+1}}{2}$$

4. **Вариация құрамы**. Ауытқу сипаттамаларының ішіндегі ең қарапайым – вариация құрамы. Мұның мәні R белгінің максимум және минимум мәнінің айырымына тең:

$$\mathrm{R} = x_{\max} - x_{\min}$$

Дисперсия және орташа квадраттық ауытқу.

Анықтама.

Кездейсоқ шамамен оның математикалық күтімі айырымының квадратының математикалық күтімін дисперсия (шашырау дейді).

X кездейсоқ шамасының дисперсиясын Д(х) арқылы белгілесек, онда анықтама бойынша:

$$D(x) = M\left[X - M(X)\right]^2$$

Белгі мәндерінің арифметикалық ортадан ауытқу квадраттары қосындысының арифметикалық ортасын таңдамалы дисперсия немесе дисперсия дейміз.

$$D_T = \sigma^2 = \frac{\sum n_i (x_i - x_T)^2}{n}$$ - өлшенген түрі, немесе

$$D_T = \sigma^2 = \frac{\sum (x_i - x_T)^2}{n}$$ - жай түрі

Түзетілген дисперсия

$$\sigma^2 = \frac{n}{n-1} \cdot Д_T$$

Орташа квадраттық ауытқу

$$\sigma_T = \sqrt{Д_T}$$

Вариация коэффициенті

$$V = \frac{\sigma_m}{x_m} \cdot 100\%$$

5. **Тандама орта және дисперсияны есептеу көбейту әдісі.**

а) Шарттық варианталар.

Делік, таңдама варианттарлары өспелі ретінде орналасқан, яғни вариациялық қатар түрінде.

Айырымы h – қа тең арифметикалық прогрессияны құрастыратын варианталарды бірдей қашықтықтағы варианталар деп атайды.

$$u_i = \frac{x_i - C}{h}$$ формуламен анықталатын вариантталарды **шарттық варианта** деп атайды.

Мұндағы С – жалған ноль (санақтың жаңа бастамасы).

Һ – қадам (екі көрші варианталардың айырмасы).

Алғашқы варианталарды шарттық варианталарға ауыстыруы, ықшамдалған әдістер таңдамының жинақты сипаттамаларын есептеу үшін негізделген.

б) **Таңдамалы орта мен дисперсияны есептеу көбейтінді әдісін** қарастырайық.

Делік, таңдама бірдей қашықтықтағы варианталар және оларға сәйкес жиіліктер түрінде берілсін. Бұл жағдайда таңдамалы ортаны және дисперсияны көбейту әдісі формулаларымен табу қолайлы.

Яғни, $\bar{x}_T = M_1 \cdot h + C$ - таңдамалы орта

$D_T = \lfloor M_2^{\ *} - (M_1^{\ *})^2 \rfloor \cdot h^2$ - таңдамалы дисперсия

Мұндағы h – қадам

C – жалған ноль (ең үлкен жиілігі бар варианта).

$M_1^* = \dfrac{\sum n_i \cdot u_i}{n}$ - бірінші ретті шарттық сәт.

$M_2^* = \dfrac{\sum n_i \cdot u_i^{\ 2}}{n}$ - екінші ретті шарттық сәт.

Көбейтінді әдісін қолдауын бір мысалда қарастырайық.

Мысал. Көлемі n=100 берілген үлестірімнің таңдамалы орта мен дисперсиясын табыңыз.

Варианта x_i	12	14	16	18	20	22
Жиілік n_i	5	15	50	16	10	4

Шешуі. Бірінші есептеу кестені құрамыз; Ол үшін:

1. Варианталарды бірінші бағанға жазамыз.
2. Жиіліктерді екінші бағанға жазамыз, жиілік қосындысын (n=100) екінші бағанның төменгі торшасына жазамыз.
3. Жалған ноль (C) ретінде C=16 вариантаны аламыз, оның ең үлкен жиілігі бар (C ретінде бағаның ортасында тұрған әлде қандай вариантаны алуға болады). Жалған ноль тұрған жолдың үшінші бағанның торшасына 0 жазамыз, оның үстінен тізбектеп – 1, -2 – жазамыз, ал о – дың астына 1,2,3 жазамыз.
4. Жиіліктердің шарттық варианталарға $n_i \cdot u_i$ көбейтінділерін төртінші бағанға жазамыз, ал олардың қосындысын $\sum n_i \cdot u_i$ бағанның төменгі торшасына орналастырамыз.

x_i	n_i	u_i	$n_i \cdot u_i$	$n_i \cdot u_i^{\ 2}$	$n_i(\cdot u_i + 1)^2$	$(\cdot u_i + 1)^2$
12	5	-2	-10	20	5	1
14	15	-1	-15	15	0	0

16	50	0	0	0	50	1
18	16	1	16	16	64	4
20	10	2	20	40	90	9
22	4	3	12	36	64	16
$\sum$	$\sum n_i = 100$		$\sum n_i \cdot u_i$ =23	$\sum n_i \cdot u_i^2 = 127$	$\sum n_i (u_i + 1)^2 = 273$	

Бақылау: $\sum n_i (u_i + 1)^2 = \sum n_i \cdot u_i^2 + 2 \sum n_i \cdot u_i + n$

273=127+2*23+100

273=127+146

273=273

5) Жиілікті шарттық варианталардың квадраттарына көбейтіп шыққан көбейтінділерді $n_i \cdot u_i^2$ бесінші бағанға жазамыз, шыққан сандарды қосып, олардың қосындысын $\sum$ $n_i \cdot u_i^2$ бағанның төменгі торшасына орналастыр-

ады.

6)Шарттық ықтималдылықтарды 1 санына үлкейтіп және олардың квадраттарын сәйкес жиеліктерге көбейтіп $n_i (u_i + 1)^2$ көбейтіндіні алтыншы бақылау бағанға жазамыз; барлық шыққан сандарды қосып, олардың қосындысын $\sum$ $n_i (u_i + 1)^2$ бағанның төменгі торшасына орналастырамыз.

Қорытындыда 1-ші кесте шығады.

Бақылау үшін мына теңбе-теңдікті қолданамыз:

$$\sum n_i (u_i + 1)^2 = \sum n_i \cdot u_i^2 + 2 \sum n_i \cdot u_i + n$$

Бақылау: $\sum n_i (u_i + 1)^2 = 273$; $\sum n_i \cdot u_i^2 + 2 \sum n_i \cdot u_i + n$=127+2*23+100

Бақылау қосындылардың дәл келу есептеуінің дұрыс болу куәлігі. Бірінші және екінші ретті шарттық сәттерді есептейміз.

$$M_1^* = \frac{\sum n_i \cdot u_i}{n} = \frac{23}{100} = 0,23$$

$$M_2^* = \frac{\sum n_i \cdot u_i^2}{n} = \frac{127}{100} = 1,27$$

Қадамын табайық h=14 – 12=2

Жалған ноль C=16 еске алып, таңдамалы орта мен дисперсияны есептейміз.

$$\bar{x}_t = M_1 \cdot h + C = 0,23 \cdot 2 + 16 = 16,46 \quad \bar{x}_T = 16,46$$

$$D_T = \left\lfloor M_2^* - (M_1^*)^2 \right\rfloor h^2 = \left[1.27 - 0.23^2\right] \cdot 2^2 = 4.87$$

$$\sigma_T = \sqrt{D_T} = \sqrt{4,87} \approx 2,21$$

Сенімділік интервалды анықтайық ($\bar{x}_T$ - δ ; $\bar{x}_T + \delta$); мұндағы δ – бағаны мына формуламен анықтаймыз δ = $t \cdot \frac{\sigma_T}{\sqrt{n}}$, ал t параметрді $2\phi(t) = \gamma$ қатынасын табамыз. $\gamma = 0.95$ сенімділігі бар жиынтықтың белгісіз математикалық күтімін a бағалау үшін _ сенімділік интервалды табамыз. Сонымен $2\phi(t) = 0.95$; $\phi(t) = \frac{0.95}{2} = 0.475$;

(екінші кестеден t=1.96) енді баға $\delta = t \cdot \frac{\sigma_T}{\sqrt{n}} = \frac{1,96 \cdot 2,21}{\sqrt{100}} = \frac{4,3316}{10} \approx 0,43$

$\bar{x}_T - \delta < a < \bar{x}_T + \delta$**;** 16,46-0,43<a<a

$\bar{x}_T - t \cdot \frac{\sigma_T}{\sqrt{n}} < a < \bar{x}_T + t \cdot \frac{\sigma_T}{\sqrt{n}}$ немесе $\bar{x}_T - \delta < a < \bar{x}_T + \delta$ </a