Модуль I.

№5 дәріс:Деректерді графикалық кескіндеудің мәні мен принциптері. Сапалық деректердің графикалық түрде кескінделуі.

Дәріс мақсаты: Биологиялық деректерді графикалық түрде кескіндеудің мәнін, негізгі түрлерін және таңдау ережелерін түсіндіру. Сапалық деректерге сәйкес диаграмма мен график құру дағдыларын қалыптастыру.

Күтілетін нәтиже: студент деректер түріне сай графикті таңдайды, гистограмма, жиіліктер полигоны мен шашырау диаграммасын құрады, графиктегі жиі кездесетін қателерді анықтайды.

Әдістері: дәріс.

Дәрістің топтама-тәсімі (тірек конспектісі немесе тезистер).

1. Деректерді графикалық кескіндеудің мәні мен принциптері.
2. Сапалық деректерді кескіндеу.
3. Деректерді диаграмма түрінде бейнелеу.

1. Деректерді графикалық кескіндеудің мәні мен принциптері.

1. Деректердің графиктік бейнеленуінің түрлері.

Графикалық кескіндеу деректің құрылымы мен заңдылығын көзбен көруге мүмкіндік беретін әдіс. Кесте нақты мәнді береді, ал график деректің таралуын, ауытқуын, топтар арасындағы айырмашылықты және ерекше мәндерді бірден көрсетеді. Сондықтан графикті есептеуден бұрын құрған жөн: деректегі қателерді сол кезде-ақ байқауға болады. Тарихта бағанды және дөңгелек диаграмманы Уильям Плейфэйр (William Playfair, 1759–1823) енгізді. Флоренс Найтингейл (Florence Nightingale, 1820–1910) Қырым соғысы кезіндегі өлім себептерін «полярлық диаграммамен» көрсетті. Эдвард Тафти (Edward Tufte) жақсы графиктің басты талабы деректі айқын әрі бұрмалаусыз беру екенін негіздеді. Графиктің құрылымына тақырып, осьтер, өлшем бірлігі, масштаб және шартты белгілер кіреді. Әр осьте атау мен өлшем бірлігі көрсетілуі тиіс.

Терең статистикалық талдауға кіріспес бұрын деректерге алдын ала талдау жасап алу немесе визуализациялау өте маңызды. Биологиялық зерттеулерде алынған сандық деректерді талдау және интерпретациялау үшін оларды графикалық түрде бейнелеу маңызды рөл атқарады. Оның негізгі мақсаты – күрделі деректерді, сандар және кестелермен көрсетілген деректерді адам миы тез қабылдай алатын пішінге түрлендіру.

Графиктер мен диаграммалар деректер арасындағы заңдылықтарды, үрдістерді және айырмашылықтарды көрнекі түрде көрсетуге мүмкіндік береді, нәтижелерді түсінуді жеңілдетеді және ғылыми қорытындылардың дәлдігін арттырады.

Сонымен қатар зерттеу нәтижелерін диаграмма түрінде де көрсету оны интерпретациялауды жеңілдетеді. Визуализацияның психологиялық аспектісі бар: адамдар визуал ақпаратты басқа модальдылықтарға қарағанда тезірек және анығырақ түсінеді.

Биологиялық деректерді графиктік бейнелеудің бірнеше негізгі түрлері қолданылады. Оларды 3 үлкен топқа бөлінеді.

1) Дискретті және номинальды (категориялық) деректердің графикалық бейнеленуінің түрлері (бағаналы, дөңгелек (секторлық), сегменттелген бағаналы, нүктелік график (диаграмма), жиіліктер полигоны).

Деректерді диаграмма түрінде бейнелеу.

Диаграммалар деректерді салыстыру және құрылымын көрсету үшін қолданылады. Ол пішініне қарай жазықтық және көлемдік болады.

Бағанды диаграмма – әртүрлі топтар немесе категориялар арасындағы сандық айырмашылықтарды көрсетуге ыңғайлы.

Дөңгелек немесе секторлық диаграмма – бүтіннің құрамындағы үлестерді пайыздық қатынаста бейнелейді немесе дөңгелекті категориялардағы салыстырмалы жиілікке пропорционал сегменттерге бөлінеді. Айнымалылардың саны 4-5-тен аспағаны жөн. Өйткені көп болса, дөңгелек диаграмманы толтырып жібереді.

Бағанды (бағаналы) диаграммада дөңгелек айнымалылар абсцисса, ал жиіліктер немесе пайыздық көрсеткіштер ордината бойынша орналасады. Көбінесе категорияларды салыстыру үшін қолданылады.

Жиіліктер полигоны – гистограмма интервалдарының орта нүктелерін қосу арқылы алынатын сызықтық график. Ол таралымды салыстыруда және динамиканы талдауда қолданылады.

2. Сапалық деректерді кескіндеу.

Сапалық (атаулы, реттік) деректер санаттарға бөлінеді, сондықтан олар жиілік немесе үлес арқылы кескінделеді. Бағанды диаграммада әр санат бір бағанмен беріледі, бағандардың арасында бос орын қалады, биіктігі жиілікке тең. Дөңгелек диаграмма бөлікті бүтінмен салыстыруға қолайлы, бірақ санат саны 5–6-дан асса, оқу қиындайды. Парето диаграммасында санаттар жиілігінің кему ретімен орналасады және жинақталған үлес сызығы қосылады. Экологияда бұл ластану көздерін ретке келтіруге ыңғайлы. Мысалы, Грегор Мендельдің екі белгі бойынша тәжірибесінде ұрпақ төрт фенотип класына бөлінді: 315 дөңгелек сары, 101 әжімделген сары, 108 дөңгелек жасыл және 32 әжімделген жасыл тұқым. Бағанды диаграммада шамамен 9:3:3:1 қатынасы анық көрінеді. Бағанды диаграмманың осі міндетті түрде нөлден басталуы керек.

3. Деректерді диаграмма түрінде бейнелеу.

Деректерді график түрінде бейнелеу.

Графиктер айнымалылар арасындағы функционалдық немесе корреляциялық байланысты көрсетуге арналған.

Сызықтық график – уақыт бойынша өзгерістерді (өсім динамикасы, физиологиялық көрсеткіштердің өзгеруі) бейнелейді.

Нүктелік график (scatter plot) – екі айнымалы арасындағы өзара байланысты, корреляцияны анықтауға мүмкіндік береді.

Графиктерді дұрыс құрастыру (осьтерді белгілеу, өлшем бірліктерін көрсету, атау беру) ғылыми деректерді түсіндірудің маңызды шарты болып табылады.

Биологияда диаграммалар популяция құрылымын, түрлік құрамды, тәжірибе нәтижелерін салыстыруда кеңінен қолданылады.

2) Үздіксіз деректердің графикалық бейнеленуінің **түрлері (гистограмма, жапырақты сабақша, мұртты жәшік (боксплот) диаграммалары)**.

Гистограмма – сандық деректердің интервалдар бойынша таралуын көрсететін бағанды график. Ол белгінің вариациясын, деректердің таралу сипатын (қалыпты, асимметриялы және т.б.) анықтауға мүмкіндік береді. Гистограмма – бағандық диаграммаға ұқсас, бірақ бағандар ең төмен шамадан ең жоғары шамаға дейін бірін бірі үздіксіз жалғап отырады. Бұл диаграмма кең диапозонды қамтитын шамалары көп интервалдық айнымалылар үшін тиімді.

Боксплот немесе бокс + уискерс (мұртты жәшік немесе мұртты қорап) деп аталатын диаграмма дисперсияны визуальды көрсету мен талдау тәсілі. Боксплотта медиана, құлаш (R), квартиль аралық құлаш (Q) пайдаланылады. Сондай ақ үлестірімге кіруі мүмкін кез келген немесе экстремальды белгілерді көрсетеді. Боксплот бір бокстан және Т-пішінді сызықтардан (мұрттан) тұрады, олардың біреуі бокстың үстіңгі жағында, ал екіншісі астыңғы жағында орналасқан.

Жапырақты сабақша диаграммасы (boxplot) – сандық деректердің таралуын, орталық тенденциясын және шашырауын қысқаша әрі көрнекі түрде көрсететін графикалық әдіс. Бұл диаграмма деректер жиынтығын **квартильдер** негізінде сипаттайды және биологиялық көрсеткіштердің вариациясын талдауда кеңінен қолданылады.

Boxplot келесі статистикалық шамаларға негізделеді:

– ең кіші мән (minimum);

– бірінші квартиль (Q1);

– медиана (Q2);
– үшінші квартиль (Q3);
– ең үлкен мән (maximum).

Мұртты жәшік диаграммасы төмендегі негізгі элементтерден тұрады:

Жәшік (box)

– Q1 мен Q3 аралығын қамтиды;
– жәшіктің ішінде медиана сызық түрінде көрсетіледі;
– жәшіктің ұзындығы деректердің негізгі шашырауын сипаттайды (квартильаралық аралық).

Мұрттар (whiskers)

– жәшіктің төменгі және жоғарғы жағынан ең кіші және ең үлкен мәндерге дейін созылады;
– деректердің жалпы диапазонын көрсетеді.

Шектен тыс мәндер (outliers)

– мұрттардан тыс орналасқан жеке нүктелер;
– биологияда бұл сирек кездесетін, ерекше немесе қателікке байланысты өлшемдер болуы мүмкін.

Бокстың жоғары жағында 3-квартиль (Q3) мен төменгі жағында 1-квартильге (Q1) жатады. Осылайша бокстың биіктігі Q –дың мәнін немесе квартиль аралық құлашты білдіреді. Бокс арқылы медиана сызық (Md) өтеді.Т-пішінді сызықтар немесе мұрттар зерттелетін жерек немесе көрсеткіштер, белгілердің құлашын біолдіреді. Олар сызбадағы ең жоғары және төменгі көрсеткіштерді білдіреді.

Алма ағашының көшеттерінің биіктігі, м

Көрсеткіш	3 жылдық көшет	5-жылдық көшет
Төмен шама –	10,9	27,5
Q1 –	11,6	29,7
Медиана –	12,5	31,4
Орташа шама	12,8	31,8
Q3 –	14,1	34,2
R –	2,5	4,5
Жоғары шама –	14,8	35,5

Boxplot және мұртты жәшік диаграммалары биологияда келесі жағдайларда тиімді қолданылады:

– өсімдіктер биіктігі, жапырақ ұзындығы, тұқым массасы сияқты морфометриялық көрсеткіштерді талдау;
– популяциялар немесе тәжірибелік топтар арасындағы айырмашылықтарды салыстыру;
– деректердің симметриясын немесе асимметриясын анықтау;
– шектен тыс мәндерді табу және биологиялық тұрғыдан түсіндіру.

Артықшылықтары

– деректердің таралуын қысқаша әрі нақты көрсетеді;
– бірнеше топты бір графикте салыстыруға мүмкіндік береді;
– қалыпты таралуға тәуелді емес (параметрлік емес әдіс);
– биометрияда, экологияда, физиологияда кеңінен қолданылады.

Шектеулері

– нақты мәндер емес, тек жалпыланған статистикалық сипаттама береді;
– деректер саны өте аз болған жағдайда ақпараттылығы төмендейді;
– таралудың нақты пішінін (мысалы, екі төбелі) көрсете алмайды.

Қорытынды

Жапырақты сабақша немесе мұртты жәшік диаграммасы биологиялық деректерді талдаудың тиімді графикалық құралы болып табылады. Ол деректердің орталық мәнін, шашырауын және ерекше жағдайларын анықтауға мүмкіндік беріп, зерттеу нәтижелерін ғылыми тұрғыдан дәл әрі көрнекі түрде ұсынуға жағдай жасайды.

Бақылау сұрақтары:

1. Биометрияда деректерді графикалық бейнелеудің қандай әдістері қолданылады?
2. Гистограмма мен бағанды диаграмманың айырмашылығы неде?
3. Деректердің графиктік бейнелеу (визуализациялау) оларды талдауға қалай көмектеседі?
4. Графикалық кескіндеудің кестеден артықшылығы неде?
5. Мәліметтерді график немесе диаграмма арқылы көосету қандай мүмкіндіктер береді?
6. Сапалық деректер үшін қандай диаграммалар қолданылады?

Әдебиеттер:

1) Раманқұлова А.А. Биологиялық статистика: оқу құралы. – Алматы: Эверо, 2023. – 14-21 бет.

2) Аубакиров Х.Ә. Биометрия: оқулық / Х.Ә.Аубакиров. – А.: Дәуір, 2011. – 354-359 б.

3) Биологиядағы бағдарламамен қамтамасыз етілген статистикалық әдістер: оқу-әдістемелік құралы / А.М. Калимағамбетов, З.М.Бияшева, А.В. Ловинская, С.Б. Дәулетбаева. – Алматы: Қазақ университеті, 2019. – 28-29 б.