

#### **Дәріс 4. Деректерді графикалық көрсету**

*Гистограмма, полигон, кумулятивті кривая және қораша диаграммасы (boxplot) сияқты визуалды әдістер қарастырылады. Мақсаты – биологиялық деректердің таралу сипатын көрнекі түрде көрсету. Мысал: балық ұзындығы бойынша үлестіру графигін құрастыру.*

Биология мен экологияда жиналатын деректер әдетте көп және күрделі болады. Тек кестелер мен сандар арқылы олардың жалпы құрылымын, таралу сипатын немесе айырмашылықтарын түсіну қиын. Сондықтан статистикалық талдауда деректерді графикалық түрде көрсету – ең маңызды және тиімді кезеңдердің бірі. Графиктер мен диаграммалар деректердің мәнін көрнекі етіп, оқырманға немесе зерттеушіге табиғи құбылыстардың сандық көрінісін көзбен көруге мүмкіндік береді. Олар деректердің таралуын, асимметриясын, шеткі мәндерін (outliers) және жалпы үрдістерді тез анықтауға көмектеседі (Sokal & Rohlf, 2012; Zar, 2010).

Биологиялық жүйелерді зерттегенде деректердің көптүрлілігіне байланысты оларды ұсынудың бірнеше графикалық тәсілі бар. Ең жиі қолданылатындары — гистограмма, вариация полигоны, кумулятивті кривая және қораша диаграммасы (boxplot). Әрбір тәсіл өз ерекшелігімен деректердің құрылымын әр қырынан ашады.

Графикалық тәсілдердің мақсаты тек әдемі көрініс жасау емес, деректердің ішкі заңдылықтарын түсіндіру. Мысалы, экологиялық деректерде популяция санының өзгерісін немесе температура мен өнімділіктің байланысын көру үшін визуалды формада ұсыну деректерді талдаудың бастапқы қадамы болып саналады. Сонымен қатар графикалық бейнелер биологтарға деректердегі қателіктер мен ерекше мәндерді табуға көмектеседі, өйткені көзбен қарағанда күмәнді нүктелер немесе ауытқулар тез байқалады.

Гистограмма — деректердің таралу жиілігін көрсететін ең кең таралған графикалық әдіс. Бұл бағаналы диаграмма, онда көлденең осьте зерттелетін белгінің мәндері, ал тік осьте олардың жиілігі көрсетіледі. Гистограмма үздіксіз сандық деректерге қолданылады, мысалы, өсімдіктің биіктігі, жапырақ ауданы, фермент белсенділігі немесе жануардың массасы сияқты белгілерге.

Гистограмма жасау үшін деректер диапазоны бірнеше интервалдарға бөлінеді. Мысалы, егер балықтардың ұзындығы 10-нан 40 сантиметрге дейін өзгерсе, онда аралықтар 5 сантиметр сайын (10–15, 15–20, 20–25, 25–30, 30–35, 35–40) бөлінеді. Әр интервал бойынша даралардың саны есептеледі және соған сәйкес биіктігі әртүрлі бағаналар тұрғызылады. Бағананың биіктігі сол интервалдағы бақылаулар жиілігіне тікелей пропорционалды.

Мысалы, Каспий теңізінен ауланған 100 балықтың ұзындығын өлшегенде келесі үлестіру байқалды: 10–15 см – 5 дара, 15–20 см – 18 дара, 20–25 см – 32 дара, 25–30 см – 27 дара, 30–35 см – 13 дара, 35–40 см – 5 дара. Гистограмма тұрғызғанда 20–25 см диапазонындағы бағана ең биік болады, себебі бұл топта

даралар саны ең көп. Мұндай график популяциядағы балықтардың басым бөлігі орташа өлшемде екенін көрсетеді, ал ұзындығы 10–15 см немесе 35–40 см болатын даралар сирек кездеседі.

Гистограмма деректердің таралу формасын айқын көрсетеді. Егер бағаналар симметриялы және төбе ортасында орналасқан болса, онда таралу қалыпты, яғни деректердің басым бөлігі орташа мәнге жақын орналасқан. Ал егер бағаналар бір жағына ығысқан болса, таралу асимметриялы болады. Мысалы, жас балықтар көп болған популяцияда гистограмма солға ығысқан болуы мүмкін, себебі қысқа даралар көп кездеседі.

Вариация полигоны — гистограммаға ұқсас, бірақ бағаналардың орнына нүктелерді бір-бірімен қосу арқылы алынған сызықтық график. Ол деректердің таралу құрылымын тегіс сызық түрінде көрсетеді және екі немесе одан көп топтың таралу қисықтарын салыстыруға ыңғайлы.

Полигон тұрғызу үшін алдымен деректердің әр интервалындағы жиілік анықталады, содан кейін әр интервалдың ортасы белгіленіп, осы нүктелер арқылы сызық жүргізіледі. Мысалы, екі көлден ауланған балықтардың ұзындығын салыстырсақ, әр көл үшін жеке полигон тұрғызамыз. Бір көлдегі полигонның шыңы 20–25 см аралығында, ал екіншісінде 25–30 см аралығында болса, бұл екінші көлдегі даралар ірі екенін көрсетеді. Осылайша, полигон арқылы популяция құрылымындағы айырмашылықтар көрнекі түрде көрсетіледі.

Полигон гистограммаға қарағанда таралу контурын дәлірек бейнелейді және деректердің симметриясын немесе асимметриясын анықтауға мүмкіндік береді. Сонымен қатар полигонды пайдаланып, орташа мән мен вариация деңгейін көзбен көруге болады: сызық неғұрлым тік және тар болса, вариация аз, ал кең жайылған болса, өзгергіштік жоғары деген сөз.

Кумулятивті кривая немесе огива — деректердің жинақталған жиілігін көрсететін график. Бұл әдіс арқылы белгілі бір мәннен төмен немесе жоғары болатын бақылаулардың пайыздық үлесін анықтаймыз. Яғни, ол «қанша дара белгілі бір шектен аспайды?» деген сұраққа жауап береді.

Кумулятивті қисықты тұрғызу үшін деректер интервалдарға бөлініп, әр интервалдағы жиілік есептеледі. Содан кейін бұл жиіліктер біртіндеп қосылып отырады, яғни жиынтық жиілік анықталады. Мысалы, балықтардың ұзындығы 25 сантиметрден аспайтын даралардың үлесі 50 пайыз болса, онда 25 сантиметр шамасы медиана мәнге жуық екенін көреміз.

Кумулятивті қисықтың пішіні деректердің таралу сипатын көрсетеді. Егер сызық бірқалыпты, тегіс өссе, деректер тең таралған; егер сызық бастапқыда баяу, кейін күрт өссе, онда аз мәндер аз кездеседі де, көпшілігі жоғары мәндерге шоғырланған. Бұл график популяция құрылымын немесе экологиялық градиент бойынша көрсеткіштердің өзгерісін бағалауда кеңінен қолданылады. Мысалы, топырақтағы тұз концентрациясы немесе өсімдік жамылғысының жасы бойынша таралуын бейнелеуге болады.

Қорапша диаграммасы (boxplot) — заманауи биостатистикада ең ақпаратты графикалық тәсілдердің бірі. Ол бірден бірнеше сипаттаманы көрсетеді: медиананы, квартильдерді, таралу диапазонын және ерекше мәндерді (outliers). Boxplot графигі деректердің құрылымын қысқаша, бірақ толық түрде бейнелейді, сондықтан оны экологиялық, физиологиялық және медициналық деректерді салыстыруда жиі қолданады (Quinn & Keough, 2002; Crawley, 2013).

Қораптың ішінде көлденең сызық деректердің медианасын көрсетеді — яғни орталық мән. Қораптың төменгі және жоғарғы шеттері бірінші және үшінші квартильдерді білдіреді, бұл деректердің ортаңғы 50 пайызы қай аралықта шоғырланғанын көрсетеді. Қораптың сыртынан шығатын «мұрттар» ең кіші және ең үлкен мәндерге дейін созылады, ал осы шектен тыс жатқан нүктелер ерекше мәндер ретінде белгіленеді.

Мысалы, үш түрлі өзеннен ауланған балықтардың ұзындығын салыстырсақ, әр өзен үшін жеке boxplot салуға болады. Бір өзенде қорапша тар және медиана ортасында орналасқан болса, бұл популяцияның даралары біртектес екенін білдіреді. Екінші өзенде қорапша кең және «мұрттар» ұзын болса, бұл популяция ішіндегі вариацияның жоғары екенін, яғни даралардың өлшемдері кең диапазонда өзгеретінін көрсетеді.

Boxplot графигі топтар арасындағы айырмашылықты тез анықтауға мүмкіндік береді. Ол орташа мәндердегі айырмашылықты ғана емес, деректердің таралуындағы ерекшеліктерді де көрсетеді. Сондықтан қазіргі биологиялық зерттеулерде boxplot экологиялық факторлардың әсерін, генетикалық топтардың айырмашылығын және эксперименттік нұсқалардың нәтижелерін бейнелеуде жиі пайдаланылады.

Биологиялық мысал ретінде балықтардың ұзындығын қарастырайық. 100 дараның ұзындығы 10–40 сантиметр аралығында болды. Есеп бойынша 10–15 см – 5 дара, 15–20 см – 18 дара, 20–25 см – 32 дара, 25–30 см – 27 дара, 30–35 см – 13 дара, 35–40 см – 5 дара. Егер бұл деректер гистограмма түрінде көрсетілсе, ең биік бағана 20–25 см аралығында орналасады, яғни осы диапазондағы балықтар ең көп кездеседі. Полигон тұрғызғанда сызық дәл осы нүктеде ең жоғары болады, ал кумулятивті қисық 25 см шамасында 50 пайыздық жиынтық жиілікті көрсетеді. Ал boxplot диаграммасы бұл деректердің медианасын, квартильдерін және ең шеткі мәндерін көрнекі түрде көрсетеді. Осы төрт графигі бірге талдау популяция құрылымының толық бейнесін береді: көпшілік даралар орта өлшемде, ал аз бөлігі тым қысқа немесе тым ұзын екені анық көрінеді.

Қорыта айтқанда, деректерді графикалық көрсету биологиялық талдаудың маңызды құралы болып табылады. Ол сандық ақпаратты визуалды формаға айналдырып, табиғи құбылыстардың заңдылықтарын айқындауға мүмкіндік береді. Гистограмма және полигон белгінің таралу құрылымын көрсетсе, кумулятивті қисық жиынтық жиілікті, ал boxplot медиана мен вариацияны бір сәтте бейнелейді. Осындай тәсілдерді меңгеру студенттерге және

зерттеушілерге күрделі деректерді жеңіл түсіндіруге, нәтижелерді салыстыруға және ғылыми қорытындыларды сенімді ұсынуға мүмкіндік береді.

Пайдаланылған әдебиеттер:

Zar, J. H. Biostatistical Analysis. Pearson, 2010.

Sokal, R. R., & Rohlf, F. J. Introduction to Biostatistics. W. H. Freeman, 2012.

Quinn, G. P., & Keough, M. J. Experimental Design and Data Analysis for Biologists. Cambridge University Press, 2002.

Fowler, J., Cohen, L., & Jarvis, P. Practical Statistics for Field Biology. Wiley-Blackwell, 2013.

Crawley, M. J. The R Book. Wiley, 2013.

Magurran, A. E. Measuring Biological Diversity. Blackwell, 2004.