Модуль I.

№6 дәріс: **Сандық деректерді графикалық кескіндеудің мәні мен принциптері**.

Дәріс мақсаты:Сандық деректерге сәйкес диаграмма мен график құру дағдыларын қалыптастыру. Заманауи бағдарламалармен (Excel, R, Python) таныстырып, нәтижені дұрыс әрі бұрмалаусыз көрсетуге үйрету.

Күтілетін нәтиже: студент деректер түріне сай графикті таңдайдысандық деректерді графикалық бейнелейді, графиктегі жиі кездесетін қателерді анықтайды.

Әдістері: дәріс.

Дәрістің топтама-тәсімі (тірек конспектісі немесе тезистер).

1. Сандық деректерді кескіндеу.
2. Аралас деректердің графикалық бейнеленуінің түрлері.
3. Графиктерді құру құралдары және жиі кездесетін қателер

1. Сандық деректерді кескіндеу.

Сандық деректер дискретті және үздіксіз болады. Үздіксіз шаманың таралуын гистограммамен көрсетеді: x осінде интервалдар (кластар), y осінде жиілік беріледі; шама үздіксіз болғандықтан бағандар бір-біріне тиіп тұрады. «Гистограмма» терминін Карл Пирсон (Karl Pearson) енгізген.

Интервал санын Стерджес формуласымен (Herbert Sturges, 1926) $k = 1 + 3{,}322 \cdot \lg n$ есептеуге болады. Мысалы, 50 бидай сабағының биіктігі үшін $k \approx 7$.

Жиілік полигоны интервал ортасын түзу сызықпен қосады, куммулята жинақталған жиілікті көрсетеді. Джон Тьюки (John Wilder Tukey, 1915–2000) «жәшік пен мұртша» диаграммасын (box plot) ұсынды: онда медиана, квартильдер және ерекше мәндер көрінеді, сондықтан топтарды салыстыруға қолайлы.

Екі өлшемді деректер, мысалы, балықтың ұзындығы мен салмағы, шашырау диаграммасымен кескінделеді. Нүктелердің жайылуына қарап байланыстың бағыты мен тығыздығын бағалауға болады.

2. Аралас деректердің графикалық бейнеленуінің түрлері (екі өлшемді шашырау диаграммасы, нүктелік диаграмма немесе (скаттерплот).

Айнымалылар арасындағы байланысты анықтау үшін шашыранды диаграммалар пайдаланылады. Шашыранды диаграмманың екі өлшемі болады. Тәуелсіз X айнымалы мәндер горизонталь ось бойымен, ал тәуелді Y айнымалы мәндер вертикаль ось бойымен орналасады. Шашыранды диаграммадағы әрбір нүкте іріктемедегі бір белгіні білдіреді, ал нүктелердің орналасуы екі айнымалы бойынша мәндерді көрсетеді.

Скаттерплот (нүктелік диаграмма) — екі сандық айнымалы арасындағы өзара байланысты графикалық түрде көрсетуге арналған диаграмма. Әрбір нүкте деректер жұбын (x; y) білдіреді. Бұл әдіс биологиялық зерттеулерде корреляцияны, тәуелділікті және деректердің таралу сипатын анықтау үшін кеңінен қолданылады.

Скаттерплот келесі элементтерден тұрады:

Көлденең ось (X) – тәуелсіз айнымалы (мысалы, уақыт, температура, жарық мөлшері);

Тік ось (Y) – тәуелді айнымалы (мысалы, өсімдік биіктігі, фотосинтез қарқындылығы);

Нүктелер – бақылау немесе өлшеу нәтижелері;

Тренд сызығы (қажет болғанда) — жалпы бағытты көрсететін сызық.

Мысал. Орташа гүл шоғырының саны есептеу алаңқайындағы өсімдік санына байланысты ма?

Егер нүктелер регрессия сызығының айналасында орналасса, байланыс соғұрлым мықты болады. Шашыранды диаграмманы болжам үшін пайдалануға болады. Жоғарыдағы мысалда егер есептеу алаңқайында 30 болса, гүл шоғырының саны қалай өзгереді?

График түрін таңдау деректердің сипатына, зерттеу мақсатына және айнымалылар арасындағы байланысқа тәуелді.

Скаттерплот арқылы анықталатын байланыс түрлері:

– Нүктелердің орналасуына қарай айнымалылар арасындағы байланыс сипаты анықталады:

– Оң корреляция: Бір айнымалы артқан сайын екіншісі де артады: (мысалы, тыңайтқыш мөлшері мен өсім биіктігі)

Теріс корреляция: Бір айнымалы артқан сайын екіншісі кемиді (мысалы, ластану деңгейі мен түрлер саны)

Корреляцияның болмауы: Нүктелер ретсіз орналасады, айқын байланыс жоқ

Сызықтық емес байланыс: Қисық немесе күрделі тәуелділік байқалады: (физиологиялық процестерге тән)

Скаттерплот келесі биологиялық бағыттарда жиі қолданылады:

– морфометрия (жапырақ ұзындығы мен ені);

– экология (температура мен организмдер саны);

– физиология (жарық қарқындылығы мен фотосинтез);

– генетика (ген экспрессиясы деңгейлерін салыстыру);

– биометрия және статистикалық талдау.

Артықшылықтары:

– екі айнымалы арасындағы байланысты көрнекі көрсетеді;

– корреляцияның бағыты мен күшін алдын ала бағалауға мүмкіндік береді;

– шектен тыс мәндерді (outliers) анықтауға қолайлы;

– сызықтық және сызықтық емес байланыстарды ажыратуға болады.

Шектеулері:

– тек екі айнымалыны бір уақытта көрсетеді;

– сандық деректер үшін ғана қолданылады;

– себеп-салдар байланысын дәлелдемейді, тек тәуелділікті көрсетеді.

Скаттерплотты құру кезінде:

– осьтер міндетті түрде белгіленуі тиіс;

– өлшем бірліктері көрсетілуі қажет;

– график атауы нақты және мазмұнды болуы керек;

– қажет жағдайда корреляция коэффициенті (r) мен тренд теңдеуі беріледі.

Скаттерплот – биологиялық деректерді талдауда ең тиімді графикалық құралдардың бірі. Ол айнымалылар арасындағы байланыстың сипатын анықтауға, деректерді визуалды түрде интерпретациялауға және ғылыми қорытынды жасауға негіз болады.

3. Графиктерді құру құралдары және жиі кездесетін қателер

Excel қарапайым графиктер үшін ыңғайлы. R (ggplot2 пакеті, авторы Хэдли Уикхэм, Hadley Wickham; негізінде Леланд Уилкинсонның «графика грамматикасы» идеясы жатыр) мен Python (matplotlib, seaborn) қайталанатын және жариялауға дайын графиктер үшін қолайлы. Код сақталатындықтан, басқа зерттеуші нәтижені қайталай алады.

Жиі кездесетін қателер: бағанды диаграммада ось нөлден басталмайды, нәтижесінде айырмашылық асыра көрсетіледі; артық 3D безендіру; өлшем бірлігі мен шартты белгінің болмауы; қателік жолағының нені білдіретіні (стандартты ауытқу немесе орташаның қатесі $m = s/\sqrt{n}$) көрсетілмейді; аз іріктемеде тек орташа беріліп, жеке нүктелер жасырылады.

Түстерді таңдағанда түсті ажырата алмайтын адамдарды ескеру қажет: viridis және ColorBrewer палитралары қолайлы. Графикті ақ-қара басып шығарғанда да оқылатынын тексерген жөн.

Биологиялық деректерді графикалық түрде бейнелеу зерттеу нәтижелерін талдауды жеңілдетіп, ақпараттың көрнекілігін арттырады. Диаграммалар, гистограммалар және

графиктер деректердің құрылымын, таралуын және өзара байланысын ашуға мүмкіндік беріп, ғылыми зерттеулердің сапасын арттыруға ықпал етеді.

Бақылау сұрақтары:

1. Гистограмма мен жиілік полигондарының айырмашылығы неде?
2. Деректердің графиктік бейнелеу (визуализациялау) оларды талдауға қалай көмектеседі?
3. Деректердің графиктік бейнеленудің типтік қателіктері.
4. Мәліметтерді график немесе диаграмма арқылы көосету қандай мүмкіндіктер береді?
5. Графиктегі қандай қателер нәтижені бұрмалайды?

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.

4) Gotelli, Nicholas James; Ellison, Aaron Michael. A Primer of Ecological Statistics. 2nd ed. Sunderland: Sinauer Associates, 2013.

5) Sokal, Robert Reuven; Rohlf, F. James. Biometry: The Principles and Practice of Statistics in Biological Research. 4th ed. New York: W. H. Freeman, 2012.