

**Дәріс 10. Биостатистикалық зерттеулердегі заманауи бағдарламалар**  
*R, SPSS, Excel, Python сияқты бағдарламаларда биостатистикалық есептерді орындау. R тілінде негізгі командалар: mean(), sd(), t.test(), cor(), lm(). Қолданбалы мысал: өсімдік биомассасының айырмашылығын t-тест арқылы бағалау және нәтижені визуализациялау.*

Қазіргі биология ғылымы мен экология, агрономия, медицина сияқты салалар деректерге негізделген зерттеу тәсілдеріне сүйенеді. Зертханалық эксперименттерден, далалық бақылаулардан және генетикалық немесе физиологиялық зерттеулерден алынған мәліметтер күн сайын артып келеді. Мұндай деректердің үлкен көлемін қолмен есептеу немесе қарапайым калькулятормен өңдеу мүмкін емес. Сондықтан қазіргі биостатистикалық зерттеулерде арнайы бағдарламалық құралдар мен бағдарламалау тілдері қолданылады. Олар зерттеушіге деректерді өңдеу, талдау, нәтижелерді графикалық түрде бейнелеу және ғылыми қорытынды жасауға мүмкіндік береді. Ең кең таралған және тиімді құралдарға Microsoft Excel, IBM SPSS Statistics, R бағдарламалау тілі және Python тілі жатады.

Microsoft Excel — бұл биологиялық және экологиялық деректерді бастапқы деңгейде өңдеуге арналған ең қолжетімді және қарапайым бағдарлама. Көптеген зерттеушілер мен студенттер дәл осы Excel бағдарламасын алғашқы статистикалық есептеулер үшін пайдаланады. Excel деректерді кесте түрінде ұйымдастыруға, арифметикалық және статистикалық формулалар қолдануға, графиктер мен диаграммалар салуға мүмкіндік береді. Мысалы, өсімдіктің биомассасын немесе өнімділігін есептеу кезінде орташа мәнді табу үшін =AVERAGE(A2:A15) формуласын, ал стандартты ауытқуды анықтау үшін =STDEV.P(A2:A15) формуласын қолдануға болады. Сонымен қатар, Excel бағдарламасында арнайы «Data Analysis» деп аталатын модуль бар, ол арқылы t-тест, ANOVA (дисперсиялық талдау), корреляция және регрессия сияқты классикалық статистикалық әдістер орындалады. Бұл модульді пайдалану үшін оны «Параметрлер» мәзірінен қосу қажет. Excel бағдарламасының басты артықшылығы — оның қарапайымдылығы мен түсініктілігі. Дегенмен, ол күрделі биологиялық модельдер мен үлкен көлемді деректер жиынтықтарын өңдеуге жеткіліксіз, сондықтан кәсіби деңгейдегі зерттеулерде басқа бағдарламалармен толықтырылады.

IBM SPSS Statistics бағдарламасы биология мен медицинада кеңінен қолданылатын кәсіби статистикалық орта болып табылады. Оның ең үлкен артықшылығы — пайдаланушыдан бағдарламалау білімін талап етпейді, себебі барлық әрекеттер мәзірлер мен батырмалар арқылы орындалады. SPSS-та студент немесе зерттеуші деректерді енгізіп, «Analyze» бөлімі арқылы қажетті талдау түрін таңдайды. Мысалы, үш түрлі тыңайтқыштың өсімдік өнімділігіне әсерін салыстыру үшін One-Way ANOVA процедурасын қолдануға болады. Бағдарлама нәтижені автоматты түрде есептеп, F-критерий, p-мән және орташа

мәндер арасындағы айырмашылықты көрсететін кестелерді шығарады. SPSS-та t-тест,  $\chi^2$ -критерий, регрессия, корреляция, факторлық және кластерлік талдаулар, сондай-ақ көпөлшемді әдістер (MANOVA, discriminant analysis) кеңінен қолданылады. Бұл бағдарлама әсіресе қолданбалы биология, экология және психофизиология салаларында танымал. SPSS-тың басты кемшілігі — оның коммерциялық (ақылы) бағдарлама болуы және код жазу арқылы бейімдеу мүмкіндігінің шектеулігі. Бірақ оның интерфейсі өте түсінікті және оқу мақсатында студенттерге ыңғайлы.

R бағдарламалау тілі қазіргі таңда биостатистика мен биоинформатика саласындағы ең қуатты құралдардың бірі болып саналады. Бұл ашық кодты, тегін және икемді бағдарлама, ол статистикалық есептеулерді, модельдеуді және визуализацияны жоғары деңгейде орындауға мүмкіндік береді. R тілі деректермен жұмыс істеу мәдениетін қалыптастырады, себебі әрбір қадамды код түрінде жазу арқылы зерттеуші өз талдауын толық қайта құра алады және нәтижелердің қайталанушылығын қамтамасыз етеді. R ортасында кез келген биостатистикалық талдау түрін орындауға арналған мыңдаған кітапханалар (packages) бар. Мысалы, «stats» пакеті негізгі функцияларды қамтиды (орташа мән, стандартты ауытқу, t-тест, корреляция және регрессия), «ggplot2» пакеті жоғары сапалы визуализация жасауға мүмкіндік береді, ал «vegan» және «phyloseq» пакеттері экологиялық және микробиологиялық деректерді талдауда кеңінен қолданылады.

R тіліндегі ең жиі қолданылатын негізгі командалар өте қарапайым. Орташа мәнді есептеу үшін `mean()`, стандартты ауытқуды анықтау үшін `sd()`, екі топ арасындағы айырмашылықты тексеру үшін `t.test()`, корреляцияны анықтау үшін `cor()`, ал регрессия моделін құру үшін `lm()` функциялары пайдаланылады. Мысалы, екі түрлі топтағы өсімдік биомассасын салыстыру үшін зерттеуші келесі кодты жаза алады:

```
group1 <- c(3.5, 3.8, 3.9, 4.0, 3.7)
group2 <- c(4.4, 4.6, 4.8, 4.7, 4.5)
t.test(group1, group2)
```

Бұл талдау нәтижесінде R бағдарламасы t-статистика мәнін, еркіндік дәрежесін және p-мәнін шығарады. Егер  $p < 0.05$  болса, айырмашылық статистикалық тұрғыдан маңызды деп есептеледі. R тілінде деректерді визуализациялау өте дамыған. Мысалы, `boxplot()` функциясы екі топтың биомассасын графикалық түрде салыстыруға мүмкіндік береді, ал `ggplot2` пакеті графиктерді кәсіби деңгейде өңдеп, ғылыми журналдарға лайық форматта жасауға жағдай жасайды.

R бағдарламасының артықшылығы — оның толық икемділігі мен ғылыми ортада стандартқа айналуы. Ол деректердің көлеміне тәуелсіз жұмыс істей алады, көпфакторлы дисперсиялық талдаулар, корреляциялық және регрессиялық модельдер, факторлық талдау, кластерлік талдау және тіпті машиналық оқыту

әдістерін де қолдайды. Сондықтан R тілі биология, экология және генетика саласында заманауи стандарт болып саналады.

Python бағдарламалау тілі де биологиялық зерттеулерде барған сайын кеңінен қолданылып келеді. Бұл тіл әмбебап және статистикалық талдау мен машиналық оқыту мүмкіндіктерін біріктіретін қуатты құрал. Python бағдарламасында биостатистикалық есептерді орындау үшін арнайы кітапханалар жиі пайдаланылады: NumPy — математикалық операциялар үшін, Pandas — деректерді кесте түрінде ұйымдастыру үшін, SciPy — статистикалық тесттер үшін, Matplotlib және Seaborn — визуализация үшін, ал Scikit-learn — регрессиялық және классификациялық модельдер үшін қолданылады. Python-ның артықшылығы оның қарапайым синтаксисінде және кез келген зерттеу қажеттілігіне бейімделу мүмкіндігінде.

Мысалы, өсімдіктің биомассасын зерттеу кезінде екі топ арасындағы айырмашылықты t-тест арқылы бағалау үшін келесі Python кодын қолдануға болады:

```
from scipy import stats
group1 = [3.5, 3.7, 3.9, 4.1, 3.8]
group2 = [4.5, 4.6, 4.7, 4.8, 4.9]
t, p = stats.ttest_ind(group1, group2)
print(t, p)
```

Егер  $p < 0.05$  болса, айырмашылық статистикалық маңызды болып есептеледі. Сонымен қатар, Python визуализация үшін Seaborn кітапханасын қолдана отырып, өте сапалы диаграммалар мен регрессиялық сызықтар жасауға мүмкіндік береді. Python тілін меңгеру болашақ биологқа машиналық оқыту және жасанды интеллект әдістерін қолдану арқылы экологиялық, генетикалық және физиологиялық процестерді болжауға мүмкіндік береді.

Заманауи биостатистикалық зерттеулерде осы төрт бағдарламаның барлығы өз орнын табады. Excel — бастапқы және оқу кезеңінде ең ыңғайлы құрал, SPSS — визуалды интерфейсі бар, дайын тесттерімен ерекшеленетін кәсіби бағдарлама, R — академиялық және ғылыми ортада кең таралған ең қуатты статистикалық орта, ал Python — ғылыми талдауды машиналық оқыту мен моделдеумен біріктіретін заманауи әмбебап тіл.

Қорытындылай келе, қазіргі биолог пен эколог тек тәжірибе жүргізіп, үлгі жинайтын зерттеуші ғана емес, сонымен қатар деректерді терең талдай алатын аналитик болуы тиіс. Биостатистикалық бағдарламалар студенттерге зерттеу нәтижелерін дәлелдеуге, заңдылықтарды анықтауға және ғылыми қорытындыны нақты сандармен көрсетуге мүмкіндік береді. Сондықтан R мен Python тілдерін меңгеру — болашақ биологтардың ғылыми әлеуетін арттыратын және халықаралық деңгейдегі зерттеу стандарттарына жақындататын маңызды кәсіби дағды болып саналады.

Пайдаланылган 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.