

Дәріс 9. Дисперсиялық талдау (ANOVA)

Көп топ арасындағы айырмашылықтарды бағалау әдісі. Бір факторлы және көп факторлы ANOVA түрлері. Қолданылуы: әртүрлі топтағы өсімдіктердің өнімділігін салыстыру.

Биология ғылымында зерттеушілер жиі бірнеше топтың немесе жағдайдың орташа мәндерін салыстыруға мәжбүр болады. Мысалы, әртүрлі тыңайтқыш түрлерінің өсімдік өнімділігіне әсерін зерттеу, түрлі жарық немесе температура жағдайларындағы фотосинтез жылдамдығын бағалау, әртүрлі ортада өсірілген микроорганизмдердің өсуін салыстыру сияқты мысалдар көптеп кездеседі. Егер екі ғана топты салыстыру қажет болса, онда бұл үшін Стьюденттің t-тесті жеткілікті. Бірақ үш немесе одан да көп топ бар жағдайда әрбір топты жеке-жеке салыстыру көптеген қателіктерге (әсіресе I типті статистикалық қателікке) әкеледі. Мұндай жағдайда барлық топтарды бір мезгілде салыстыру үшін тиімді және сенімді әдіс — дисперсиялық талдау (ANOVA — Analysis of Variance) қолданылады.

Дисперсиялық талдау — бұл статистикалық әдіс, ол бірнеше топтың орташа мәндерінің арасында шынайы айырмашылық бар ма, әлде байқалған айырмашылықтар кездейсоқ па — соны анықтайды. Бұл әдіс жалпы өзгергіштікті (вариацияны) екі негізгі құрамдас бөлікке бөлуге негізделген:

Топаралық вариация (Between-group variance) — әр түрлі топтардың орташа мәндерінің айырмашылығынан туындайтын өзгергіштік. Бұл фактордың (мысалы, тыңайтқыш түрі, жарық деңгейі, температура) әсерін көрсетеді.

Топішілік вариация (Within-group variance) — бір топ ішіндегі даралар арасындағы табиғи немесе кездейсоқ айырмашылықтар.

Егер топаралық вариация топішілік вариациядан әлдеқайда жоғары болса, онда зерттеліп отырған фактордың шынайы әсері бар деген қорытынды жасалады. Ал егер айырмашылықтар аз немесе шамалас болса, онда барлық топтардың орташа мәндері іс жүзінде бірдей деп есептеледі.

Дисперсиялық талдау кезінде зерттеуші алдымен нөлдік гипотезаны (H_0) және баламалы гипотезаны (H_1) анықтайды. Нөлдік гипотеза барлық топтардың орташа мәндері бірдей дегенді білдіреді, яғни зерттеліп отырған фактордың әсері жоқ. Баламалы гипотеза кем дегенде бір топтың орташа мәні өзгеше деген пікірді білдіреді.

Мысалы, егер зерттеуші үш түрлі тыңайтқыштың өсімдік өнімділігіне әсерін зерттесе, онда:

H_0 : барлық тыңайтқыштар өнімділікке бірдей әсер етеді ($\mu_1 = \mu_2 = \mu_3$);

H_1 : кем дегенде бір тыңайтқыштың әсері өзгеше ($\mu_i \neq \mu_j$).

ANOVA талдауы осы гипотезаларды тексеру арқылы қай фактордың әсері бар екенін немесе айырмашылықтар кездейсоқ екенін анықтайды.

Дисперсиялық талдаудың бірнеше түрі бар, олардың ең жиі қолданылатындары: бір факторлы ANOVA және көп факторлы ANOVA.

Бір факторлы ANOVA (One-way ANOVA) тек бір тәуелсіз айнымалының (фактордың) әсерін зерттеу үшін қолданылады. Мысалы, үш түрлі тыңайтқыштың өсімдік өнімділігіне әсерін салыстырсақ, онда фактор — тыңайтқыш түрі, ал тәуелді айнымалы — өнімділік. Мұндай жағдайда әр топтың орташа мәндері салыстырылып, олардың айырмашылығы кездейсоқ па, әлде шынайы ма, соны бағалау қажет.

Бір факторлы ANOVA-ның мәні мынада: жалпы вариация екі бөлікке бөлінеді — топаралық вариация және топішілік вариация. Егер топаралық вариация топішілікке қарағанда әлдеқайда жоғары болса, онда зерттеліп отырған фактор топтар арасында нақты айырмашылық туғызады деп саналады.

Көп факторлы ANOVA (Two-way немесе Multi-way ANOVA) екі немесе одан да көп фактордың әсерін және олардың өзара әрекеттесуін зерттейді. Мысалы, зерттеу барысында тыңайтқыш түрі мен суару деңгейі сияқты екі фактордың өсімдік өнімділігіне әсерін бағалауға болады. Мұндай талдау әрбір фактордың жеке әсерін ғана емес, сонымен қатар олардың өзара әрекеттесуін (interaction effect) анықтауға мүмкіндік береді. Мысалы, тыңайтқыштың әсері ылғалдылық деңгейіне байланысты күшеюі немесе әлсіреуі мүмкін.

Дисперсиялық талдау нәтижелері F-критерий (F-statistic) арқылы бағаланады. Бұл критерий топаралық дисперсия мен топішілік дисперсияның қатынасын сипаттайды:

$$F = MS_{\text{between}} / MS_{\text{within}},$$

мұндағы MS (mean square) — орташа квадраттық ауытқу.

F мәні неғұрлым жоғары болса, топтардың орташа мәндері арасындағы айырмашылық соғұрлым айқын. F-критерий бойынша есептелген мән белгілі бір еркіндік дәрежесі (degrees of freedom) және таңдалған маңыздылық деңгейі ($\alpha = 0,05$ немесе $\alpha = 0,01$) үшін кестелік F мәнімен салыстырылады. Егер есептелген F мәні кестелік мәннен үлкен болса, онда нөлдік гипотеза қабылданбайды, яғни айырмашылық статистикалық тұрғыдан маңызды деп есептеледі.

Сонымен қатар, ANOVA нәтижесімен бірге әрдайым p-мәні де беріледі. Егер $p < 0,05$ болса, айырмашылық шынайы деп саналады; ал $p \geq 0,05$ болса, айырмашылық кездейсоқ сипатта деп бағаланады. Мысалы, зерттеу нәтижесінде $F = 9,32$ және $p = 0,002$ болса, бұл фактордың әсері бар екенін көрсетеді.

Дисперсиялық талдаудың негізгі нәтижесі – топтар арасындағы айырмашылық бар екенін дәлелдеу. Бірақ ANOVA нақты қандай топтар бір-бірінен ерекшеленетінін көрсетпейді. Сондықтан айырмашылық анықталғаннан кейін қосымша сынақтар — пост-хок тесттер (Post-hoc tests) жүргізіледі. Ең жиі қолданылатындары:

Tukey's HSD (Honestly Significant Difference) тесті — барлық топтардың жұптық айырмашылықтарын бірден салыстырады және қай топтардың айырмашылығы маңызды екенін көрсетеді;

Bonferroni тесті — статистикалық қатені азайту үшін түзетілген шекті мәндерді пайдаланады;

Scheffé тесті — дисперсия біркелкі емес болған жағдайларда қолданылады.

Мысалы, үш түрлі тыңайтқыштың әсерін салыстыру кезінде ANOVA айырмашылық бар екенін көрсетсе, пост-хок тест нәтижесінде белгілі болуы мүмкін: тыңайтқыш С өсімдіктің биіктігін А және В түрлеріне қарағанда айтарлықтай арттырған.

ANOVA қолдану үшін белгілі бір статистикалық шарттардың орындалуы қажет. Біріншіден, деректер қалыпты үлестірілуі тиіс, яғни әр топтағы мәндер шамамен симметриялы және орташа мәннің маңайында жинақталуы керек. Екіншіден, дисперсияның біртектілігі (homogeneity of variances) сақталуы қажет, яғни барлық топтарда өзгергіштік шамамен бірдей болуы тиіс. Бұл шартты Левен (Levene) немесе Бартлетт (Bartlett) тесттері арқылы тексеруге болады. Үшіншіден, бақылаулар тәуелсіз болуы тиіс — әрбір өлшеу бір-біріне әсер етпеуі қажет. Егер бұл талаптар орындалмаса, онда ANOVA нәтижелері сенімсіз болуы мүмкін, сондықтан кейде деректер логарифмдік, квадрат түбірлік немесе басқа трансформациялар арқылы түрлендіріледі.

Биологиялық зерттеулерде дисперсиялық талдау өте кең қолданылады. Агробиологияда ANOVA әртүрлі сорттардың, тыңайтқыштардың немесе суару режимдерінің өнімділікке әсерін салыстыру үшін қолданылады. Физиологияда ол әртүрлі температура, жарық немесе рН деңгейінің фермент белсенділігіне немесе тыныс алу қарқынына әсерін анықтау үшін пайдаланылады. Экологияда ANOVA арқылы түрлі аймақтардағы өсімдіктердің биомассасын немесе жануарлар популяцияларының тығыздығын салыстыруға болады. Генетикада бұл әдіс әртүрлі генотиптердің фенотиптік айырмашылықтарын бағалау үшін тиімді.

Мысалы, зерттеуші үш түрлі тыңайтқыштың өсімдік биіктігіне әсерін зерттегенде орташа мәндер келесідей болуы мүмкін: тыңайтқыш А — 42 см, тыңайтқыш В — 48 см, тыңайтқыш С — 55 см. Дисперсиялық талдау нәтижесінде $F = 8,7$ және $p = 0,001$ мәндері алынған. Бұл айырмашылықтың кездейсоқ емес екенін және тыңайтқыш түрі өсімдіктің өсуіне шынайы әсер ететінін көрсетеді. Tukey тесті нәтижесі бойынша С тыңайтқышы өсімдіктің биіктігін А және В тыңайтқыштарына қарағанда айтарлықтай арттырған.

Қорытындылай келгенде, дисперсиялық талдау (ANOVA) — биологиялық және экологиялық деректерді өңдеудің ең маңызды және сенімді әдістерінің бірі. Ол бірнеше топтың орташа мәндерін бір уақытта салыстырып, факторлардың әсерін сандық тұрғыда бағалауға мүмкіндік береді. ANOVA нәтижелері биологиядағы күрделі құбылыстарды түсіндіруде, эксперимент нәтижелерін объективті дәлелдеуде және болашақ зерттеу бағыттарын жоспарлауда үлкен рөл

атқарады. Бұл әдісті дұрыс қолдану зерттеу нәтижелерінің сенімділігін арттырады және биологиялық жүйелердің ішкі заңдылықтарын анықтауға 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.