Модуль III.

№21 дәріс: Регрессиялық талдаудың мәні.

Әдістері: дәріс.

Дәрістің топтама-тәсімі (тірек конспектісі немесе тезистер).

1. Регрессиялық талдаудың мақсаты.
2. Регрессиялық талдау түрлері.
3. Сызықтық регрессиялық талдау.
4. Регрессиялық коэффициентінің мәнділігін және регрессия теңдеуінің сапасын бағалау.

Регрессиялық талдау бір белгінің (тәуелді айнымалы, у) екінші белгіге (тәуелсіз айнымалы, х) қалай тәуелді екенін теңдеу түрінде сипаттайды. «Регрессия» терминін Фрэнсис Гальтон (Francis Galton, 1886) енгізді: ол биік әкелердің ұлдары орта есеппен әкелерінен аласа болатынын, яғни өсімнің орташа мәнге қарай «қайтатынын» байқады.

Корреляция байланыстың күші мен бағытын көрсетсе, регрессия оның нақты пішінін береді және у мәнін х арқылы болжауға мүмкіндік береді. Корреляцияда екі белгі тең дәрежелі, ал регрессияда бағыты бар. Талдаудың мақсаттары: тәуелділіктің пішінін анықтау; х бір бірлікке өзгергенде у қанша өзгеретінін бағалау; болжам жасау; бір белгіні екіншісі арқылы жанама өлшеу. Мысалы, балықтың ұзындығын өлшеп, салмағын тірі күйінде өлшемей-ақ болжауға болады; орман экологиясында діңнің диаметрі бойынша ағаштың биомассасын бағалайды.

Корреляциялық талдау белгілер арасындағы байланыстың бар екендігін, оның бағыты мен күшін анықтайды, бірақ бір көрсеткіштің екінші көрсеткішке қаншалықты әсер ететінін көрсетпейді. Ғылыми зерттеулерде бір көрсеткіштің орташа мәні екенші көрсеткішке қаншалық ты әсер ететінін білу қажеттігі туындайды. Бұл мақсатқа жеті үшін регрессиялық талдау жүргізу қажет. Тәуелді айнымалының орта мәні мен предикторлардың орта мәндері арасындағы өзара байланыс регрессия теңдеуі түрінде өрнектеледі.

Регрессия теңдеуі тәуелді айнымалы мен предикторлық айнымалылардың бастапқы статистикалық деректері негізінде таңдалынып алынатын математикалық функция, көбінесе сызықтық функция қолданылады. Бұл жағдайда регрессиялық талдау сызықты *регрессиялық талдау* деп аталады.

Регрессиялық талдаудың мақсаты: регрессия теңдеуі арқылы тәуелсіз айнымалының күтілетін орта мәнін алдын ала болжау.

Регрессиялық талдаудың негізгі міндеттері:

– Тәуелділіктің түрі мен моделін анықтау;
– Регрессия теңдеукінің модельдің параметрлерін анықтау.
– Регрессия теңдеуінің мәнділігін тексеру.
– Регрессия теңдеуінің коэффициентерінің мәнділігін тексеру.
– Коэффициенттердің аралық бағаларын тұрғызу.
– Модельдің популяциядағы ақиқат регрессия теңдеуімен сәйкестігін бағалау.
– Тәуелді айнымалының нүктелік және аралық болжамдарын тұрғызу.

Регрессиялық талдау түрлері:

– Жай сызықты регрессия;
– Көптік сызықты регрессия;
– Жай логистикалық регрессия;
– Көптік логистикалық регрессия.

2. Регрессиялық талдау түрлері

Регрессия тәуелсіз айнымалылар санына қарай жұптық (бір х) және көпше (бірнеше х) болады. Байланыс пішініне қарай сызықтық және сызықтық емес түрлерге бөлінеді.

Сызықтық емес түрлерге полиномдық, дәрежелік және көрсеткіштік регрессия жатады. Мысалы, экологияда түр санының аумаққа тәуелділігі $S = c \cdot A^z$ дәрежелік

функциясымен сипатталады; оны логарифмдеп, сызықтық түрге келтіруге болады. Ферменттер кинетикасында Леонор Михаэлис (Leonor Michaelis) пен Мод Ментен (Maud Menten, 1913) теңдеуі қолданылады. Тәуелді айнымалы екі мәнді болса (түр кездеседі немесе кездеспейді), логистикалық регрессия (David Cox, 1958) қолданылады.

Қазіргі зерттеулерде аралас әсерлі модельдер, жалпыланған аддитивті модельдер (GAM) және LASSO, ridge сияқты регуляризация әдістері кеңінен қолданылады, олар көп белгі болғанда артық бейімделуден сақтайды.

Регрессиялық талдау – бұл бір тәуелді айнымалы мен бір немесе бірнеше тәуелсіз айнымалылар арасындағы статистикалық өзара байланысты зерттеу әдісі. Регрессиялық талдауда тәуелсіз айнымалыны – «предиктор», ал тәуелді айнымалыны «жауап айнымалы» деп атайды.

Регрессия коэффициенті салыстырмалы шама емес. Оның көмегімен екі белгінің де өзгеру дәрежелерін анықтауға болады. Атап айтқанда x-белгінің мәнінің y-белгіге байланысты өзгеруі немесе керісінше y-белгінің x-белгі шамасының өзгеруіне байланысын көрсете алады.

Белгілер арасындағы регрессиялар эмпирикалық және теориялық регрессиялық қатарлар, графикалық және регрессиялық теңдеулер арқылы көрсетіледі.

Биологияда көбінесе жай сызықтық регрессиялық талдау қолданылады.

3. Сызықтық регрессиялық талдау

Қарапайым сызықтық регрессия теңдеуі: $\hat{y} = a + b \cdot x$, мұндағы a — бос мүше, b — регрессия коэффициенті. Параметрлерді ең кіші квадраттар әдісімен табады: онда қалдықтар квадратының қосындысы $\Sigma(y - \hat{y})^2$ ең аз болады.

Әдісті Адриен-Мари Лежандр (Adrien-Marie Legendre, 1805) жариялады, ал Карл Фридрих Гаусс (Carl Friedrich Gauss) оны 1809 жылы негіздеді. Формулалар: $b = \Sigma(x - \bar{x})(y - \bar{y}) / \Sigma(x - \bar{x})^2 = r \cdot s_y / s_x$ және $a = \bar{y} - b \cdot \bar{x}$. Коэффициент b x бір бірлікке артқанда y орташа қанша өзгеретінін көрсетеді. Мысалы, су температурасына (x, °C) еріген оттегі мөлшерінің (y, мг/л) тәуелділігі шамамен $\hat{y} = 13{,}9 - 0{,}22 \cdot x$ түрінде болуы мүмкін: температура 1 °C көтерілсе, оттегі орта есеппен 0,22 мг/л азаяды.

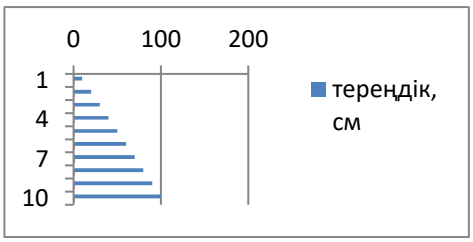
Болжамды зерттелген x диапазонының ішінде ғана жасаған жөн; одан тыс экстраполяция сенімсіз.

4. Регрессия коэффициентінің мәнділігін және теңдеу сапасын бағалау

Коэффициенттің мәнділігін Стьюдент критерийімен тексереді: $t = b / m_b$, мұндағы $m_b = s_e / \sqrt{\Sigma(x - \bar{x})^2}$, $s_e = \sqrt{[\Sigma(y - \hat{y})^2 / (n - 2)]}$, еркіндік дәрежесі $n - 2$. Сенімділік аралығы $b \pm t \cdot m_b$. Мысалы, $n = 20$, $b = -0{,}22$, $m_b = 0{,}03$ болса, $t \approx 7{,}3$; 18 еркіндік дәрежесінде $t_{0{,}001} = 3{,}92$ болғандықтан, $p < 0{,}001$. Теңдеу сапасын детерминация коэффициенті R^2 көрсетеді: $R^2 = 1 - \Sigma(y - \hat{y})^2 / \Sigma(y - \bar{y})^2$, бұл мысалда $R^2 \approx 0{,}75$, яғни оттегі өзгерісінің 75%-ы температурамен түсіндіріледі. Модельді бүтін түрде Фишер критерийі (F) бағалайды; жұптық регрессияда $F = t^2$. Сонымен қатар қалдықтарды талдау қажет: қалдықтар нормал таралып, дисперсиясы тұрақты, бір-біріне тәуелсіз болуы керек; ерекше мәндердің әсерін Кук қашықтығы (R. Dennis Cook, 1977) арқылы тексереді.

Тереңдік, см	10	20	30	40	50	60	70	80	90	100
Топырақ тығыздығы, г/см3	1,08	1,12	1,15	1,18	1,23	1,28	1,28	1,29	1,34	1,4

Бұл жағдайда топыһақ тығыздығы топырақ тереңдеген сайын артатынын көреміз. Яғни олардың арасында функциональдық тауелділік бар.

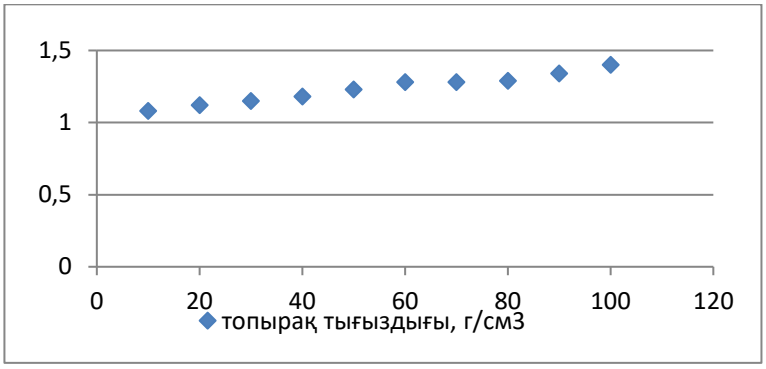


Енді корреляция коэффициентін есептейміз.

	тереңдік, см	*топырақ тығыздығы, г/см3*
тереңдік, см	1	
топырақ тығыздығы, г/см3	0,988188	1

Есептелген корреляция коэффициенті 0,98-ге тең.
Бірақ та бұл нақты талдаулар үшін жеткіліксіз.

Қаншалықты байланыс бар екенін білу үшін скаттерплот тұрғызамыз.



Скаттерплот бойынша сызықтық байланыс бар екені анықталады.
Осы байланысты формула бойынша өрнектейміз.
y = 0,0033x + 1,0527
немесе басқаша өрнектейміз

$$\rho = 1{,}05 + 0{,}0033h; \qquad (1)$$

мұндағы ρ – топырақ тығыздығы, г/см3; 1,05 бс мүше, 0,033 – регрессия коэффициенті, h – топырақ тереңдігі.

Осы формула бойынша біз өзімізге қажетті тереңдіктегі топырақ тығыздығын шамамен болжап таба аламыз.

Мысал 70 см тереңдікте

$\rho_{70} = 1{,}05 + 0{,}0033*70 = 1{,}28$

50 см тереңдікте – 1,22 г/см3 болады.

Яғни есептелген мәндер нақты мәндерге сайкес келеді, айырмашылығы көп емес. әмен теориялық мәандер арасындағы айырмашылық елеулі болмайды.

Осы мақсатқа жету үшін экспериментальды жұмыстарәжүруі керек.

Факторлы белгінің натижелік немесе результативті белгіге қаншалық асер ететінін корсетеді.

$$R^2 = 0{,}9765 \qquad (2)$$

Бақылау сұрақтары:

1. Регрессия коэффициентін есептеу арқылы нені анықтауға болады?
2. Сызықтық регрессия не үшін қолданылады?
3. Регрессия коэффициенті қанша вариантта (белгіде) анықталады?
4. Үлкен жиынтықтарда регрессия коэффициентін есептеу жұмыстарында қажетті аралық мән ретінде қолданылатын көрсеткішті атаңыз.
5. Регрессиялық талдаудың шектеулері.

Әдебиеттер:

1) Аубакиров Х.Ә. Биометрия: оқулық / Х.Ә.Аубакиров. – А.: Дәуір, 2011. – 177-183 б.; 342-347 беттер.

2) Раманқұлова А.А. Биологиялық статистика: оқу құралы. – Алматы: Эверо, 2023. – 188-201 бет.