

АСАБЛІВАСЦІ ДЫСПЕРСІЙНАГА АНАЛІЗУ ВЫНІКАУ ШМАТГАДОВАГА ПАЛЯВОГА ДОСЛЕДУ

Дысперсійны аналіз — адзін з асноўных метадаў статыстычнай апрацоўкі вынікаў палявога доследу — праводзіцца для ацэнкі істотнасці дзеяння вывучаемых фактараў: эфектаў варыянтаў доследу ці каэфіцыентаў рэгрэсіі. У апошнім выпадку дысперсійны аналіз з'яўляецца неабходнай часткай рэгрэсійнага аналізу.

Тэхніка разліку сум квадратаў аднагадовых доследаў выкладзена ў шматлікіх дапаможніках па біяметрыі, стандартных праграмах для ЭВМ і таму маецца ў арсенале кожнага даследчыка. Больш цяжкая задача паўстае перад эксперыментатарамі пры дысперсійным аналізе вынікаў шматгадовага доследу, бо ў існуючай метадычнай літаратуры гэта пытанне амаль не закранаецца.

Склалася парадаксальнае становішча: у значнай колькасці артыкулаў, змешчаных у навуковых выданнях, паказчык найменшай істотнай рознасці (НІР) маецца толькі для аднагадовых вынікаў, а вывады робяцца па шматгадовых сярэдніх. Бывае і так, што НІР шматгадовых сярэдніх знаходзіцца па формуле сярэдняй арыфметычнай з НІР аднагадовых значэнняў.

Адзначым, што мы разглядаем тут шматгадовы дослед як сукупнасць аднолькавых аднагадовых доследаў, праведзеных у асобныя гады на розных палях у аднолькавых глебавых умовах. Інакш кажучы, кожны з аднагадовых доследаў праводзіцца з адной і той жа культурай, па адной схеме і мае аднолькавую гісторыю. У гэтым сэнсе аднагадовыя доследы могуць уяўляць сабой і вынікі аднагадовых эксперыментаў стацыянарных доўгатэрміновых доследаў.

Разгледзім у якасці прыкладу аналіз невялікай выбаркі вынікаў палявога доследу па вывучэнні ўплыву доз азоту на ўраджайнасць ячменю (табл. 1). Памер выбаркі максімальна зменшаны з мэтай аблегчыць чытачу разлік статыстычных параметраў. Паколькі задача артыкула абмяжоўваецца разглядам метадычных аспектаў дысперсійнага аналізу, па-за нашай увагай застаецца аграхімічны бок справы.

Разглядаемы двухгадовы дослед праводзіўся на двух побач размешчаных палях, умоўна ў 1990 і 1991 гг. Штогод усе варыянты схемы доследу аб'ядноўваліся ў блокі 1, 2, 3. Такім чынам, пры дысперсійным аналізе аднагадовых доследаў можна выдзеліць дысперсію варыянтаў доследу — фактар *A*, паўтаральнасцяў — *P*, астатка — *Z* (табл. 2).

Асаблівасцю шматгадовага доследу з'яўляецца існаванне дадатковага фактара — фактара года (поля), які можна вызначыць як фактар умоў росту і развіцця раслінаў у розныя гады. Абзначым яго лацінскай літарай *B*.

Табліца 1. Ураджайнасць зерня ячменю, ц/га

Варыянт	1990 г.			1991 г.			1990—1991 гг.	
	паўтаральнасці доследу						сума	сярэдняя
	1	2	3	1	2	3		
1	25,5	32,1	21,5	17,2	20,1	16,5	132,9	22,15
2	37,7	39,8	33,5	33,7	33,0	19,1	196,8	32,80
3	39,8	42,8	40,5	37,0	36,6	31,8	228,5	38,08
4	45,2	45,1	42,0	31,4	33,7	34,5	231,9	38,65
Сума <i>P</i>	148,2	159,8	137,5	119,3	123,4	101,9	790,1	

Таблиця 2. Дисперсійний аналіз аднагадових доследаў

Від варыяцыі	Сума квадратаў	Ступень свабоды	Дысперсія	Крытэрыі F	
				фактычны	05
1990 г.					
Агульная	628,48	11			
Паўтаральнасці	62,19	2			
Варыянтаў	539,05	3	179,68	39,6	4,76
Астатак	27,24	6	4,54		
1991 г.					
Агульная	697,54	11			
Паўтаральнасці	65,16	2			
Варыянтаў	532,71	3	177,57	10,7	4,76
Астатак	99,67	6	16,61		

Сума квадратаў варыянтаў двухгадовага доследу вызначаецца сярэдняй за два гады ўраджайнасцю кожнага варыянта і разлічваецца, як і для аднагадовых доследаў, па адпаведных сумах ці сярэдніх (па табл. 1): $C_A = (132,9^2 + 196,8^2 + 228,5^2 + 231,9^2) : 6 - 26010,75 = 1053,00$, дзе лічба 26010,75 — папраўка (карэкціруючы фактар), вылічваемая па формуле: $C = (\sum x)^2 : N = 790,1^2 : 24 = 26010,75$.

Матрыца двухгадовага доследу складаецца з шасці блокаў (паўтаральнасцяў аднагадовых доследаў), сумы якіх змешчаны ў ніжнім радку табл. 1. Няцяжка заўважыць, што вар'іраванне ўраджайнасці гэтых блокаў вакол сярэдняй шматгадовага доследу абумоўлена як рознай урадлівасцю глебы блокаў, так і ўмовамі росту і развіцця ячменю ў розныя гады.

Фактар B можна статыстычна выразіць сумай квадратаў, абумоўленай ураджайнасцю ячменю ў розныя гады: $C_B = [(148,2^2 + 159,8 + 137,5)^2 + (119,3 + 123,4 + 101,9)^2] : 12 - 26010,75 = 424,20$. Гэты разлік можна больш проста выразіць формулай: $C_B = C_1 + C_2 + C$, дзе C_1 і C_2 — папраўкі, з дапамогай якіх былі разлічаны сумы квадратаў аднагадовых доследаў спосабам «ад нуля». Звычайна гэтыя папраўкі маюцца ў даследчыка пасля апрацоўкі аднагадовых доследаў, і іх не трэба разлічваць зноў.

Маючы вынікі дысперсійнага аналізу аднагадовых доследаў (мы даём іх у гадавым выглядзе без разліку ў табл. 2) і C_B , лёгка разлічыць суму квадратаў агульнай дысперсіі шматгадовага доследу па формуле $C_y = C_{y1} + C_{y2} + C_B$, дзе C_{y1} і C_{y2} — адпаведныя сумы квадратаў асобных гадоў: $C_y = 628,48 + 697,54 + 424,20 = 1750,22$. Зразумела, гэту суму можна атрымаць звычайным шляхам: $C_y = 25,5^2 + 32,1^2 + 21,5^2 + \dots + 34,5^2 - C$, але ў гэтым няма неабходнасці.

У сувязі са сказаным вышэй суму квадратаў шасці блокаў можна вызначыць як сумеснае дзеянне паўтаральнасцяў аднагадовых доследаў і гадоў: $C_{P+B} = (148,2^2 + 159,9^2 + \dots + 101,9^2) : 4 - 26010,75 = 551,55$. Важна адзначыць, што гэта сума можа быць атрымана простым складаннем адпаведных сум квадратаў: $C_{P+B} = C_{P1} + C_{P2} + C_B = 62,19 + 65,16 + 424,20 = 551,55$, дзе C_{P1} і C_{P2} — сумы квадратаў паўтарэнняў аднагадовых доследаў (табл. 2).

Аналагічна можна разлічыць суму квадратаў сумеснага дзеяння варыянтаў доследу і гадоў (мы абазначаем яе сімвалам C_{A+B} у адрозненне ад узаемадзеяння C_{AB}): $C_{A+B} = C_{A1} + C_{A2} + C_B = 539,05 + 532,71 + 424,20 = 1495,96$.

Б. А. Даспехаў [1] адзначаў, што пры аналізе шматгадовага доследу і выдзяленні дысперсіі гадоў, сума квадратаў астатка абумоўлена не толькі выпадковымі фактарамі, але і ўзаемадзеяннем варыянтаў з умовамі фарміравання ўраджаю розных гадоў.

Паколькі ўзаемадзеянне вызначаецца як мера парушэння складальнасці дысперсій асобных фактараў, мы можам яго разлічыць так: $C_{AB} = C_{A+B} - C_A - C_B$. Падставіўшы значэнне C_{A+B} з папярэдняй формулы, атрымаем: $C_{AB} = C_{A1} + C_{A2} - C_A$. Апошняя формула можна карыстацца для разліку сумы квадратаў ўзаемадзеяння любых вывучаемых у доследзе фактараў з гадамі незалежна ад іх колькасці. Для нашага прыкладу $C_{AB} = 539,05 + 532,71 - 1053,00 = 18,76$.

Заўважым, што пытанне пра мэтазгоднасць выдзялення ў якасці самастойных крыніц вар'іравання сум квадратаў гадоў і ўзаемадзеяння тыпу C_{AB} вырашае сам эксперыментатар. Але адказ ад такога аналізу звязаны са стратай каштоўнай інфармацыі пра ўмовы праяўлення эфектаў: напрыклад, у сухія ці нармальныя гады праяўляецца пэўная заканамернасць дзеяння вывучаемых фактараў. Трэба мець на ўвазе, што шматгадовыя сярэднія — гэта, па сутнасці, праяўленне «галоўных эфектаў» вывучаемых у доследзе фактараў, і з імі трэба адпаведна абыходзіцца. Напрыклад, вядома, што паўтаральнасць, на якой грунтуюцца галоўныя эфекты, павялічваецца ў столькі разоў, колькі градацый мае элімінуемы фактар. У нашым выпадку паўтаральнасць сярэдняй двухгадовай роўная 6, аднагадовай — 3. Наогул паўтаральнасць вызначаецца як колькасць дзялянак, утвараючых сярэдняю.

Са сказанага вынікае, што абавязкова праводзіць аналіз як шматгадовых, так і аднагадовых вынікаў эксперыментаў, паколькі аднагадовыя доследы з'яўляюцца не толькі сродкам атрымання вынікаў шматгадовых эксперыментаў, але і ўяўляюць самастойную каштоўнасць.

Такім чынам, у нас ёсць усе сумы квадратаў, неабходныя для пабудовы поўнай табліцы дысперсійнага аналізу шматгадовага доследу (табл. 3). Важна зразумець, што толькі пры такім аналізе, калі выдзяляецца і ўзаемадзеянне вар'янтаў з гадамі, рэалізуецца правіла: сума квадратаў выпадковага вар'іравання шматгадовага доследу роўная суме адпаведных квадратаў аднагадовых доследаў, а дысперсія (сярэдня квадрат) астатка шматгадовага доследу — сярэдня арыфметычнай дысперсій астатка аднагадовых доследаў.

У табл. 3 нязвычайна выглядае другі радок, дзе запісана дысперсія паўтаральнасцяў аднагадовых доследаў. Але яна мае дакладны фізічны сэнс і з'яўляецца састаўной часткай рассеяння ўраджайнасці шасці блокаў, увабляючых паўтаральнасць у часе і на тэрыторыі доследнага ўчастка. Калі даследчык не выдзяляе суму квадратаў гадоў, то можна аб'яднаць другі і трэці радкі, каб атрымаць дысперсію шасці блокаў, якой будучь адпавядаць 5 ступеняў свабоды. Гэта не ўплывае на велічыню дысперсіі памылкі і статыстычных крытэрыяў істотнасці эфектаў даследуемых фактараў.

Менавіта няправільная ацэнка паўтаральнасці шматгадовага доследу з'яўляецца галоўнай прычынай памылак эксперыментатараў пры статыстычным аналізе яго вынікаў. Так, В. І. Каранеўскі [2] выдзеліў як фактар трохгадовага доследу блокі (паўтаральнасці) як сукупнасць паўтаральнасцяў аднагадовых доследаў, абазначаных аднолькавым нумарам. Для нашага прыкладу (табл. 1) атрымаем: $C_P = [(148,2 + 119,3)^2 + (159,8 + 123,4)^2 + (137,5 + 101,9)^2] : 8 - 26010,75 = 123,11$.

Табліца 3. Дысперсійны аналіз шматгадовага доследу

Від варыяцыі	Сума квадратаў	Ступень свабоды	Дысперсія	Крытэрыі F	
				фактычны	05
Агульная, C_y	1750,22	23			
Паўторнасці, C_P	127,35	4			
Гадоў, C_B	424,20	1	424,20	40,1	4,75
Варыянтаў, C_A	1053,00	3	351,00	33,2	3,49
Узаемадзеянне, C_{AB}	18,76	3	6,25	0,59	3,49
Астатак, C_Z	126,91	12	10,58		

Заўважым, што абагульняючай прыкметай такога блакіравання з'яўляецца агульнасць нумарацыі і толькі. Калі б даследчык зрабіў іншую нумарацыю, напрыклад змяніў месцамі лічбы 2 і 3 блокаў 1991 г., то C_P атрымалася б зусім іншай: $C_P = [(148,2 + 119,3)^2 + (159,8 + 101,9)^2 + (137,5 + 123,4)^2] : 8 - 26010,75 = 0,74$. У выніку маем розныя ацэнкі аднаго фактару, але менавіта ад гэтага залежыць велічыня астаткавай дысперсіі, памылкі доследу, крытэрыя Фішара, НІР з усімі наступствамі.

Звернем увагу чытачоў на тое, што адна памылка В. І. Каранеўскага дапоўнілася шэрагам іншых. Тры 25-варыянтныя аднагадовыя доследы пераўтварыліся ў 75-варыянтны трохгадовы, у суму квадратаў варыянтаў трапілі сумы, азначаныя намі як C_B і C_{AB} . У выніку сярэдні квадрат (дысперсія) варыянтаў шматгадовага доследу атрымаўся ў шмат разоў большы ў параўнанні з дысперсіямі аднагадовых доследаў, што, відавочна, здарылася з-за вялікага ўплыву на ўраджайнасць умоў надвор'я розных перыядаў вегетацыі. Памылковая дысперсія варыянтаў дала завышаны крытэрыя Фішара, але не паўплывала на велічыню НІР, калі не лічыць таго, што В. І. Каранеўскі згубіў 6 ступеняў свабоды, якія належалі блокам, а трапілі ў дысперсію астатка. Таму апошняя склала 3,61, а сярэдняя арыфметычная дысперсій аднагадовых доследаў — 3,51. Менавіта гэта, а не акругленні, як мяркуе аўтар, прычына несупадзення ацэнак. У дадатак да ўсяго паўтаральнасць аднагадовых і шматгадовых сярэдніх засталася аднолькавай.

З мэтай больш поўнага разумення акалічнасцяў, звязаных са статыстычнай ацэнкай паўтаральнасці шматгадовага доследу, прывядзём вынікі дысперсійнага аналізу (табл. 1), зробленага адным з матэматыкаў, экспертаў нашага алгарытму (які не мае нічога новага, акрамя рацыяналізацыі разліку некаторых сум квадратаў), на ЭВМ па стандартнай праграме. Былі выдзелены ўсе магчымыя дысперсіі, у тым ліку паўтаральнасці як сукупнасці блокаў з агульнай нумарацыяй; узаемадзеянняў паўтаральнасці з гадамі, варыянтамі і трайнае ўзаемадзеянне APB . Паколькі астатку не засталася, у якасці дысперсіі памылкі была ўзята агульная дысперсія трох апошніх узаемадзеянняў, а паўтаральнасць засталася ацэнкай лакальнага фактару.

Усё было б добра, каб не недарэчная паўтаральнасць. Сапраўды, які сэнс мае ўзаемадзеянне PB або APB , калі толькі палова кожнага блока належыць любому з гадоў (палёў). Можна паказаць, што сума квадратаў гэтых C_P і C_{PB} заўсёды будзе роўная суме квадратаў паўтарэнняў аднагадовых доследаў, змешчанай у табл. 3, незалежна ад нумарацыі блокаў, хаця складаемая заўсёды будзе рознымі. Напрыклад, ЭВМ дала C_P 123,11 а C_{PB} — 4,24. Дададзім, што статыстычныя ацэнкі і не павінны залежаць ад такіх суб'ектыўных фактараў, як нумарацыя блокаў.

У заключэнне падкрэслім, што дысперсію памылкі шматгадовага доследу можна разлічваць па формуле сярэдняй арыфметычнай адпаведных дысперсій аднагадовых доследаў толькі ў выпадку выдзялення дысперсій усіх блокаў і ўзаемадзеяння варыянтаў з гадамі. Разлік статыстычных крытэрыяў для ацэнкі сярэдніх сярэдніх варыянтаў шматгадовага доследу: $S_x = \sqrt{10,58 : 6} = 1,33$ ц/га, $НІР_{05} = t\sqrt{2}S_x = 2,18 \cdot 1,414 \cdot 1,33 = 4,1$ ц/га.

Для прыблізнага разліку НІР шматгадовага доследу ў выпадку адсутнасці неабходных дысперсій можна падзяліць сярэдняю з НІР аднагадовых эксперыментаў на \sqrt{n} (n — колькасць гадоў доследу). Атрыманая НІР можа быць некалькі завышана з-за рознасці крытэрыяў t невялікіх эксперыментаў і дакладнай для большасці доследаў. Але лепш разлічыць НІР па даных дысперсійнага аналізу, які, як бачым, не павінен палыхаць даследчыкаў.

Summary

The insubstantiation of the definition of replications of a long-term trial as being the populations of organized replications (blocks) of annual trials marked by equal numbers as well as the necessity of grouping the variance of all blocks of a long-term trial with its decomposition into variances of replications of annual trials and years (fields) are shown. Summation of the variances of the residual of annual and long-term trials is obtained in case of grouping a sum of squares of interaction of the trial scheme variants by year.

Літаратура

1. Доспехов Б. А. Методика полевого опыта. М., 1965.
2. Короневский В. И. // Земледелие. 1985. № 11. С. 56—57.

БелНДІГА

*Паступіў у рэдакцыю
18.04.94*