



"مقاله پژوهشی"

تجزیه علیت عملکرد دانه گندم با غلبه بر چندهم خطی صفات

هادی علی پور^۱، حسین عبدی^۲ و محمدرضا بی همتا^۳

۱- استادیار گروه تولید و ژنتیک گیاهی، دانشگاه ارومیه، ارومیه، (نویسنده مسؤول: ha.alipour@urmia.ac.ir)

۲- دانشجوی دکتری گروه تولید و ژنتیک گیاهی، دانشگاه ارومیه، ارومیه

۳- استاد گروه زراعت و اصلاح نباتات، دانشگاه تهران، کرج

تاریخ دریافت: ۱۴۰۰/۲/۲ تاریخ پذیرش: ۱۴۰۰/۵/۴

صفحه: ۱۲۲ تا ۱۲۹

چکیده

تجزیه علیت از قدیمی ترین روش های آگاهی از اثرات مستقیم و غیرمستقیم صفات بر عملکرد دانه می باشد اما کاربرد گسترده این رویکرد با مشکل چند هم خطی محدود می گردد که در تحقیق حاضر سعی بر رفع آن شده است. تعداد ۲۹۸ ژنوتیپ گندم نان طی دو سال زراعی در قالب طرح لاتیس ارزیابی شدند و تجزیه علیت عملکرد دانه با روش ساده و رنج مورد مطالعه قرار گرفت. بالاترین اثر مستقیم بر عملکرد دانه در سال اول به ترتیب مربوط به وزن سنبله، تعداد دانه و وزن هزار دانه و در سال دوم مربوط به تعداد دانه، وزن هزار دانه و وزن سنبله بود. وزن سنبله و تعداد دانه در سال اول از طریق یکدیگر و نیز از طریق طول سنبله و ارتفاع بوته اثر غیرمستقیمی بر عملکرد داشتند. در سال دوم نیز تعداد دانه و وزن هزار دانه از طریق وزن سنبله بیشترین اثر غیرمستقیم را نشان دادند. همچنین نتایج نشان داد صفات روز تا آبستنی و روز تا گلدهی در هر دو سال و وزن سنبله در سال دوم از چند هم خطی بالایی برخوردار بودند که استفاده از ثابت $k=+10$ در تجزیه علیت رنج منجر به غلبه بر این مشکل شد. در نهایت برخلاف تجزیه علیت ساده، با انجام تجزیه علیت رنج مشخص شد که در سال دوم وزن سنبله از طریق تعداد دانه اثر غیرمستقیمی بر عملکرد دانه داشت. با توجه به نتایج، توصیه می گردد مسئله چند هم خطی میان صفات در مطالعات آتی تجزیه علیت مورد توجه قرار گیرد.

واژه های کلیدی: اثرات مستقیم و غیرمستقیم، اجزای عملکرد، تجزیه علیت رنج، رگرسیون گام به گام، مراحل فنولوژیک

مقدمه

گندم مهم ترین غله در سطح جهان می باشد و به نژادگران در تلاشی مداوم سعی در افزایش عملکرد دانه آن دارند (۳). صفات کمی همچون عملکرد دانه معمولاً وراثت پذیری پایینی دارند و تحت تأثیر محیط می باشند و توسط تعداد زیادی از ژن ها کنترل می شوند. بنابراین در این صفات انتخاب غیرمستقیم روشی مناسب برای دستیابی به بازده ژنتیکی است (۲۶). تجزیه علیت با تقسیم ضریب همبستگی ساده به اثرات مستقیم و غیرمستقیم، یکی از رویکردهای قدیمی به نژادگران در انتخاب مطلوب صفات می باشد. این تجزیه به طور گسترده برای بهبود عملکرد دانه گندم های ایرانی مورد استفاده قرار گرفته است (۴۶، ۱۲، ۱۴، ۱۶) و می توان از ضرایب حاصل از آن برای توسعه شاخص های گزینشی استفاده نمود (۲۰).

تجزیه علیت یک تحلیل رگرسیون جزئی و استاندارد شده است و کاربرد آن می تواند با چندهم خطی در ماتریس همبستگی صفات محدود گردد. هم خطی چندگانه یکی از مشکلات بسیار رایج تجزیه علیت در علوم گیاهی می باشد، به طوری که می تواند منجر به نتایج غیرواقعی و تفسیرهای اشتباه شود (۲۳). چندهم خطی زمانی اتفاق می افتد که متغیرهای مستقل یا ترکیبی از آن ها با یکدیگر همبسته هستند و با استفاده از شاخص های مختلفی قابل تشخیص می باشد. عامل تورم واریانس (VIF)^۱ روش رسمی تفکیک هم خطی چندگانه است و مشخص می کند که در مقایسه با حالتی که متغیرهای پیش بینی به طور خطی به هم مرتبط

نشده اند، چه مقدار از واریانس های ضرایب رگرسیون، برآورد متورم شده هستند (۱۱). شماره وضعیت (CN)^۲ و تعیین کننده (D) ماتریس همبستگی از دیگر روش های شناسایی درجه چندهم خطی میان صفات می باشد (۱۸). در مطالعه ای با هدف بررسی اثر اندازه نمونه بر دقت این شاخص ها نتیجه گیری شد که در تعداد نمونه های بالاتر از ۱۳۵ اثر تداخلی اندازه نمونه بر تشخیص چندهم خطی به صفر می رسد اما در اندازه نمونه های کوچک تر صحت تشخیص چندهم خطی صفات تضعیف می گردد (۲۲).

تاکنون تلاش های متعددی برای غلبه بر مشکل چندهم خطی در تجزیه علیت انجام گرفته است. توبه و کارگلولتی فیلو (۲۴) با مقایسه انواع تجزیه علیت اظهار داشتند که برآورد ضرایب علیت تحت چند هم خطی، بدون اهمیت بیولوژیکی هستند و نباید مورد توجه قرار گیرند. از این رو، تجزیه علیت سنتی بایستی با حذف متغیرهای بسیار همبسته انجام گیرد. اولیوتو و همکاران (۱۸) بیان داشتند که استفاده از همه مشاهدات نمونه برداری به جای میانگین داده ها در برآورد ماتریس همبستگی، یکی از رویکردهای مناسب در کاهش چندهم خطی می باشد. همچنین به منظور تجزیه علیت مناسب، اولیوتو و همکاران (۱۷) استفاده از رگرسیون گام به گام برای انتخاب یک مجموعه از صفات با چند هم خطی پایین را پیشنهاد دادند. در مطالعه ای دیگر، ویوتو دل کنته و همکاران (۲۶) حذف برخی از صفاتی که چندهم خطی بالایی دارند را روشی مناسب و ساده تر از تجزیه علیت رنج دانستند.

1- Variance Inflation Factor
4- Ridge path analysis

2- Condition Number

3- Determinant

شاخص تعیین‌کننده ماتریس همبستگی (D_{XX}) هرچه به صفر نزدیک‌تر باشد، منعکس‌کننده چند هم‌خطی شدید می‌باشد (۱۸) و از طریق رابطه زیر به دست می‌آید که در آن λ_j مقدار ویژه صفت λ م در ماتریس همبستگی XX است. $D_{XX} = \prod_{j=1}^p \lambda_j$ همچنین مقادیر بالاتر از ۱۰ برای عامل تورم واریانس (VIF) نشان‌دهنده چند هم‌خطی می‌باشد (۱۱). این شاخص از طریق تقسیم مجموع مربعات بردار ویژه (EV) بر مقدار ویژه (λ) مربوطه به صورت زیر حاصل می‌شود:

$$VIF_{\beta_k} = \left(\frac{(EV_{k_{c1}})^2}{\lambda_1} + \frac{(EV_{k_{c2}})^2}{\lambda_2} + \dots + \frac{(EV_{k_{cp}})^2}{\lambda_p} \right)$$

در تحقیق حاضر حد آستانه شاخص‌های VIF و CN به ترتیب ۵ و ۳۰ لحاظ شدند. برای غلبه بر چند هم‌خطی از ضرایب رگرسیون استاندارد شده ریح مطابق رابطه زیر استفاده شد.

$$\beta_j = (XX + k)^{-1}XY$$

که در آن عدد ثابت k نشان‌دهنده مقدار اریب در برآورد کننده‌هاست. وقتی که k برابر با صفر است، معادله به ضرایب رگرسیون مربعات حداقل معمولی در فرم استاندارد شده تبدیل می‌شود. وقتی که $k > 0$ است، ضرایب رگرسیون ریح اریب‌دار بوده ولی در مقایسه با برآوردکننده‌های مربعات حداقل معمولی پایدارتر (یعنی دارای تغییرات کمتر) هستند (۱۱). در نهایت تجزیه همبستگی میان صفات با استفاده از نرم‌افزار SPSS 25 و تجزیه رگرسیون گام‌به‌گام و تجزیه علیت عملکرد دانه با استفاده از بسته‌های biotools و metan در نرم‌افزار R 4.0.4 انجام شدند.

نتایج و بحث

ضرایب همبستگی میان صفات به تفکیک دو سال آزمایش در جدول ۱ ارائه شده است. همان‌طور که انتظار می‌رفت میان مراحل فنولوژیک به‌ویژه روز تا گلدهی و روز تا آبلستنی همبستگی مثبت و بالایی مشاهده شد. به دلیل تعداد بالای نمونه، اکثر همبستگی‌ها با وجود مقادیر عددی پایین، معنی‌دار شدند. عملکرد دانه در سال اول با صفات وزن سنبله، تعداد دانه و ، تعداد دانه و وزن هزار دانه همبستگی بالایی داشت. چنین همبستگی‌هایی قبلاً نیز در گندم گزارش شده است (۴،۱۰،۲۱).

همبستگی‌های بالا میان دو صفت می‌تواند ناشی از تأثیر یک یا مجموعه‌ای از صفات دیگر باشد، بنابراین گرچه ضرایب همبستگی پیرسون از ارتباط خطی بین متغیرها اطلاعاتی فراهم می‌آورد، اما قادر به شناسایی روابط علی و معلولی میان صفات نیستند. بنابراین تجزیه علیت ساده و تجزیه علیت ریح با مقادیر مختلف k انجام شد. صفات روز تا گلدهی و روز تا آبلستنی در هر دو سال دارای بیشترین مقادیر VIF بودند. همچنین وزن سنبله در سال دوم از VIF بالاتر از ۵ برخوردار بود. شاخص‌های CN و D نیز مؤید وجود چندهم‌خطی در ماتریس همبستگی صفات بودند (جدول ۲). همان‌طور که ملاحظه می‌شود، با افزایش مقدار k ، میانگین مربعات اشتباه

رگرسیون ریح یکی از چند روشی است که برای رفع مسئله هم‌خطی چندگانه پیشنهاد شده است. برآورد کننده‌های رگرسیون استاندارد شده ریح از طریق وارد کردن یک عدد ثابت اریبی (k)، به داخل معادلات نرمال حداقل مربعات محاسبه می‌شود (۱۱). گزارش شده است که اضافه کردن ثابت $k=0.10$ برای کاهش درجه چند هم‌خطی مؤثر می‌باشد (۲۴). سیلوا و همکاران (۲۳) نیز نشان دادند که تجزیه علیت با مدل‌سازی معادلات ساختاری^۱ (SEM) نتایج بهتری نسبت به تجزیه علیت سنتی و تجزیه علیت ریح دارد. شناسایی صفات مؤثر بر عملکرد دانه گندم و مقایسه تجزیه علیت با روش‌های مختلف از اهداف تحقیق حاضر بود.

مواد و روش‌ها

تعداد ۲۹۸ ژنوتیپ گندم نان ایرانی شامل ۹۰ رقم زراعی و ۲۰۸ توده بومی جمع‌آوری شده از نقاط مختلف کشور طی دو سال زراعی ۱۳۹۳ و ۱۳۹۴ در قالب طرح لاتیس با دو تکرار در مزرعه تحقیقاتی گروه زراعت و اصلاح نباتات دانشگاه تهران مورد ارزیابی قرار گرفتند. در تحقیق حاضر صفات مختلف فنولوژیکی و اجزای عملکرد شامل روز تا رسیدگی، روز تا گلدهی، روز تا آبلستنی، ارتفاع بوته، طول سنبله، تعداد دانه در سنبله، وزن سنبله، وزن هزار دانه و عملکرد دانه در بوته اندازه‌گیری شدند.

ضرایب رگرسیون جزئی تجزیه علیت در یک مجموعه از متغیرها، بر اساس معادلات زیر برآورد شدند.

$$XX\beta = XY \quad \beta = XX^{-1}XY$$

که β بردار ضرایب رگرسیون جزئی، XX^{-1} معکوس ماتریس همبستگی خطی بین متغیرهای پیش‌گو و XY ماتریس همبستگی متغیرهای پیش‌گو با متغیر پاسخ می‌باشد. پس از برآورد ضرایب رگرسیون می‌توان اثرات مستقیم و غیرمستقیم p متغیر پیش‌گو را به شرح زیر تشکیل داد:

$$r_{p_1:y} = \beta_1 + \beta_{2rp_1:p_2} + \dots + \beta_{k_{rp_1:p_n}}$$

که در آن $r_{p_1:y}$ همبستگی خطی بین متغیر p_1 با عملکرد (y) ، β_1 اثر مستقیم متغیر p_1 بر روی عملکرد و $\beta_{2rp_1:p_2}$ و $\beta_{k_{rp_1:p_n}}$ به ترتیب اثرات غیرمستقیم متغیر p_1 بر روی عملکرد از طریق متغیرهای p_2 و p_n می‌باشد.

در پژوهش حاضر از معیارهای مختلف برای پی بردن به درجه چندهم‌خطی استفاده شد. شماره وضعیت (CN) که در آن مقادیر کمتر از ۱۰۰ نشان‌دهنده چند هم‌خطی ضعیف، مقادیر بین ۱۰۰ تا ۱۰۰۰ بیان‌گر چند هم‌خطی متوسط و مقادیر بالاتر از ۱۰۰۰ حاکی از چند هم‌خطی بسیار شدید است (۱۵). این شاخص از طریق نسبت بیش‌ترین و کمترین مقدار ویژه ماتریس همبستگی XX به شرح زیر محاسبه می‌گردد:

$$CN = \frac{\lambda_{Max}}{\lambda_{Min}}$$

کردند. در مطالعه‌ای دیگر سیلوا و همکاران (۲۳) حداقل مقدار برای k که واریانس‌ها را ثابت می‌کند ($VIF < 10$) برابر با 0.03 دانستند. با وجود این که عدد ثابت k بر اساس اثر رنج و عامل تورم واریانس قابل تعیین است اما مقدار بهینه k تا حدودی از یک کاربرد به کاربرد دیگر تغییرپذیر و نامعلوم می‌باشد و بیشتر بر اساس نظر محقق است. شکل ۱ ضرایب علیت برای صفات مورد مطالعه را بر اساس مقادیر مختلف k نشان می‌دهد. ضرایب علیت به ویژه برای صفات روز تا گلدهی و روز تا آبلستنی که هم‌خطی بالایی داشتند، با کوچک‌ترین تغییر k از صفر به شدت نوسان می‌یابند اما با افزایش بیشتر k گرایش به تغییر کم می‌گردد.

(باقی‌مانده) افزایش یافته و در عین حال VIF صفات کوچک‌تر می‌گردد. چنین کاهش در DN و SN نیز قابل مشاهده است اما در $k=0.05$ هم‌خطی چندگانه به طور کامل رفع نشده است در حالی که در $k=0.10$ همه شاخص‌های مدنظر به زیر حد آستانه رسیدند (جدول ۲). اندازه بالای نمونه علاوه بر این که معنی‌داری ضرایب همبستگی را تحت تأثیر قرار می‌دهد، می‌تواند بر دقت تشخیص چندهم‌خطی صفات نیز اثر بگذارد (۲۲)، لذا با توجه به تعداد بالای ژنوتیپ در مطالعه حاضر می‌توان نتیجه گرفت که شاخص‌های مذکور از دقت قابل قبولی در تشخیص چندهم‌خطی برخوردارند. در راستای نتایج این پژوهش، توبه و کارگلتوتی فیلو (۲۴) برای غلبه بر چند هم‌خطی مقدار بهینه k را برابر با 0.10 عنوان

جدول ۱- همبستگی بین صفات مورد ارزیابی در سال اول (بالای قطر اصلی) و سال دوم (پایین قطر اصلی)

	DF	DB	DM	PH	SW	SL	SN	TKW	GY
DF	1/00	0/99**	0/63**	0/20**	-0/13*	0/23**	-0/14*	0/01 ^{ns}	-0/20**
DB	0/97**	1/00	0/63**	0/19**	-0/12*	0/23**	-0/14*	0/01 ^{ns}	-0/20**
DM	0/49**	0/46**	1/00	0/14*	0/03 ^{ns}	0/19**	-0/07 ^{ns}	0/05 ^{ns}	-0/07 ^{ns}
PH	0/32**	0/35**	0/16**	1/00	0/34**	0/36**	0/19**	0/22**	0/29**
SW	-0/14*	-0/16*	-0/02 ^{ns}	0/17**	1/00	0/46**	0/73**	0/24**	0/87**
SL	0/28**	0/25**	0/30**	0/35**	0/28**	1/00	0/37**	0/10 ^{ns}	0/40**
SN	-0/23**	-0/25**	-0/11 ^{ns}	0/03 ^{ns}	0/75**	0/17**	1/00	-0/15*	0/79**
TKW	-0/08 ^{ns}	-0/08 ^{ns}	0/08 ^{ns}	0/17**	0/63**	0/18**	0/19**	1/00	0/25**
GY	-0/24**	-0/24**	-0/07 ^{ns}	0/08 ^{ns}	0/89**	0/19**	0/79**	0/71**	1/00

ns, * و **: به ترتیب غیرمعنی‌داری و معنی‌داری در سطح احتمال پنج و یک درصد می‌باشد.

DF: روز تا گلدهی، DB: روز تا آبلستنی، DM: روز تا رسیدگی، PH: ارتفاع بوته، SW: وزن سنبله، SL: طول سنبله، SN: تعداد دانه، TKW: وزن هزار دانه و GY: عملکرد دانه
ns, * and **: Not significant, significant at 5% and 1% probability levels, respectively.
DB: Day to booting, DF: Day to flowering, DM: Day to maturity, PH: Plant height, SL: Spike length, SN: Seed number, SW: Spike weight, GY: Grain yield

جدول ۲- مقدار VIF صفات و سایر پارامترها برای تشخیص چندهم‌خطی در تجزیه علیت ساده ($k=0$) و تجزیه علیت رنج ($k=0.05$) و $k=0.10$ به تفکیک دو سال آزمایشTable 2. The amount of VIF traits and parameters for collinearity diagnostics in simple ($k=0$) and ridge ($k=0.05$ and $k=0.10$) path analysis during the two year

Traits	صفات	First year			Second year		
		$k=0/00$	$k=0/05$	$k=0/10$	$k=0/00$	$k=0/05$	$k=0/10$
PH	ارتفاع بوته	1/29	1/19	1/12	1/25	1/21	1/13
SL	طول سنبله	1/48	1/36	1/26	1/25	1/24	1/16
SN	تعداد دانه	2/87	2/31	1/95	3/42	2/50	2/02
SW	وزن سنبله	3/28	2/60	2/17	5/39	3/69	2/85
TKW	وزن هزار دانه	1/42	1/26	1/14	2/39	1/85	1/57
DB	روز تا آبلستنی	39/58	8/26	4/82	20/83	7/04	4/24
DF	روز تا گلدهی	40/06	8/45	4/86	21/22	7/13	4/39
DM	روز تا رسیدگی	1/72	1/54	1/41	1/41	1/29	1/20
Condition number	شماره وضعیت	212/46	43/83	24/83	113/97	37/64	22/86
Determinant	تعیین کننده	0/0027	0/024	0/073	0/004	0/024	0/073
R-squared	ضریب تبیین	0/843	0/820	0/708	0/948	0/919	0/894
Residual	باقی‌مانده	0/396	0/425	0/449	0/227	0/285	0/325

DB: Day to booting, DF: Day to flowering, DM: Day to maturity, PH: Plant height, SL: Spike length, SN: Seed number, SW: Spike weight, GY: Grain yield

بنابراین همبستگی بالای وزن سنبله با عملکرد دانه در سال اول بیشتر به اثرات مستقیم و در سال دوم عمدتاً به اثرات غیرمستقیم تقسیم گردید. این نتیجه بیانگر دو نکته بسیار مهم است؛ اول این که ناپستی با تکیه بر همبستگی‌های ساده اقدام به گزینش نمود و دوم، روابط علی و معلولی میان صفات با آزمایش در یک محیط نمی‌تواند تا حد زیادی شناسایی گردد. از این رو، توبه و کارگلتوتی فیلو (۲۴) و فراری و

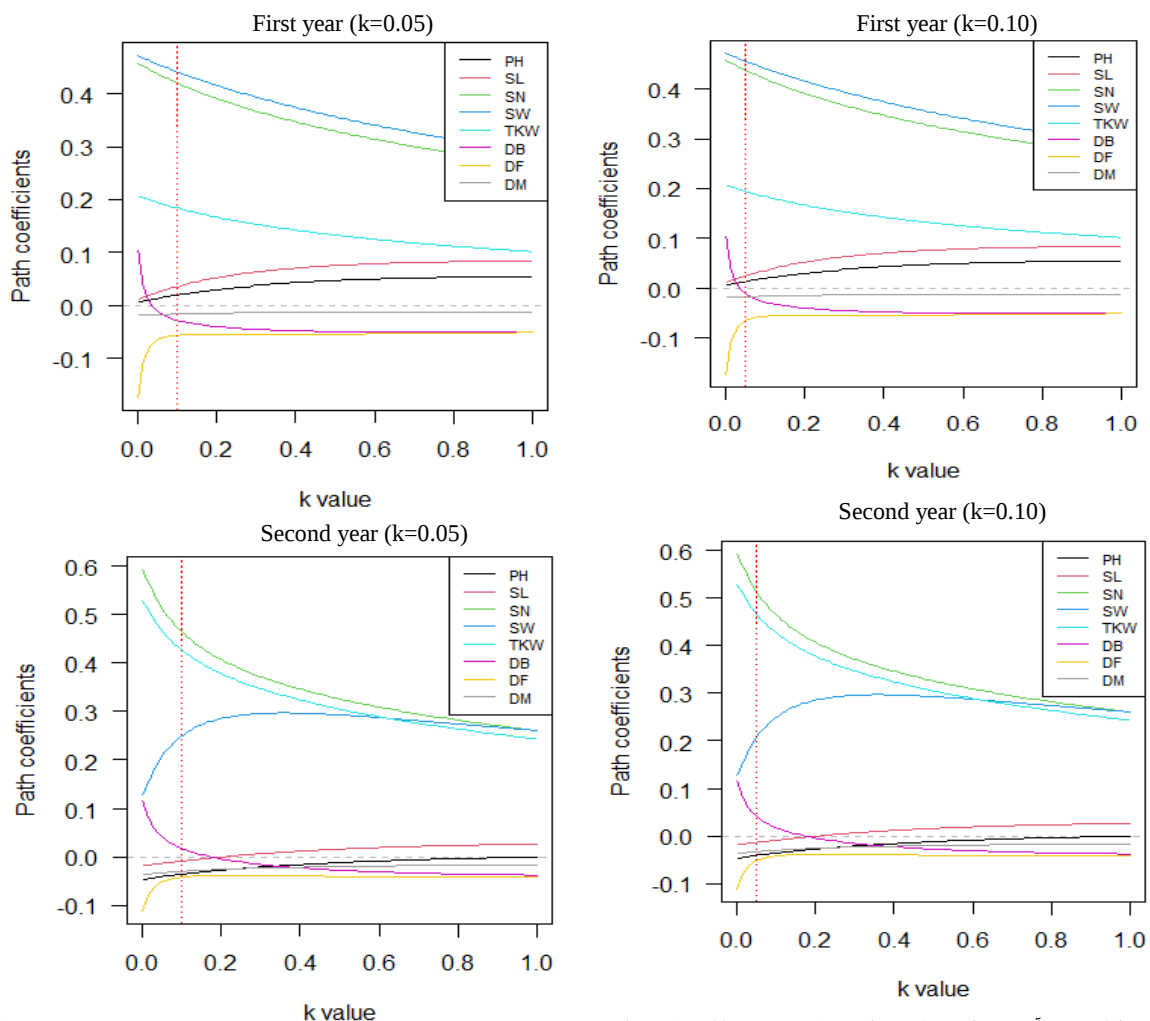
بالترین اثر مستقیم بر عملکرد دانه در سال اول به ترتیب مربوط به وزن سنبله، تعداد دانه و وزن هزار دانه و در سال دوم مربوط به صفات تعداد دانه، وزن هزار دانه و وزن سنبله بود. وزن سنبله و تعداد دانه در سال اول از طریق یک‌دیگر و نیز از طریق طول سنبله و ارتفاع بوته اثر غیرمستقیمی بر عملکرد داشتند. در سال دوم نیز تعداد دانه و وزن هزار دانه از طریق وزن سنبله بیشترین اثر غیرمستقیم را نشان دادند.

ریج توصیه شده است (۸). با انجام تجزیه علیت ریح با هر دو مقدار ثابت k مشخص شد که اثر بالای تعداد روز تا گلدهی و تعداد روز تا آبستنی به ویژه در سال اول ناشی از چند هم‌خطی بوده است و با افزوده شدن ثابت k و متعاقباً کاهش چند هم‌خطی، اثرات آن‌ها کاهش می‌یابد. در سال دوم نیز چند هم‌خطی تقریباً بالا برای وزن سنبله منجر گردید تا اثر غیرمستقیم آن از طریق تعداد دانه در تجزیه علیت سنتی پنهان بماند، حال آن‌که در تجزیه علیت ریح این صفت اثر غیرمستقیم خود را نشان داد. این نتیجه قابل انتظار بود زیرا چند هم‌خطی مشکلی متداول در تجزیه علیت می‌باشد که کمتر بدان توجه شده است (۱۹). غلبه بر چند هم‌خطی با اضافه کردن ثابت k در چندین تجزیه علیت مشاهده شده است (۵،۹).

همکاران (۷) تجزیه علیت عملکرد را در داده‌های حاصل از آزمایشات چندمحیطی اجرا کردند.

نتایج تجزیه علیت ساده در تحقیق حاضر نشان داد که صفات روز تا گلدهی به طور مستقیم و روز تا آبستنی به طور غیرمستقیم با ضرایب منفی بر عملکرد دانه تأثیر می‌گذارند. در این ارتباط محمدی و همکاران (۱۳) به اثر مستقیم و منفی تعداد روز تا رسیدگی فیزیولوژیکی بر عملکرد دانه گندم اشاره داشته‌اند و مقدم و همکاران (۱۲) از وزن هزار دانه و تعداد دانه در سنبله به عنوان صفاتی با بالاترین اثر مستقیم بر عملکرد دانه یاد کردند.

وقتی که چند هم‌خطی شدید است، ضرایب با مقادیر بالایی برآورد می‌شوند و غیرقابل اعتماد هستند و معمولاً با تفسیرهای زیستی مغایرت دارند (۱۸)، بنابراین در چنین حالتی و بدون حذف صفات با چند هم‌خطی بالا، استفاده از رگرسیون



شکل ۱- برآورد ضرایب علیت برای صفات مورد مطالعه با در نظر گرفتن مقادیر مختلف k طی دو سال زراعی (DF: روز تا گلدهی، DB: روز تا آبستنی، DM: روز تا رسیدگی، PH: ارتفاع بوته، SW: وزن سنبله، SL: طول سنبله، SN: تعداد دانه، TKW: وزن هزار دانه و GY: عملکرد دانه) Figure 1. Estimated path coefficients for studied traits obtained with different k values during the two year (DB: Day to booting, DF: Day to flowering, DM: Day to maturity, PH: Plant height, SL: Spike length, SN: Seed number, SW: Spike weight)

جدول ۳- اثرات مستقیم (قطر اصلی) و غیرمستقیم (قطر فرعی) تجزیه علیت ساده ($k=0$) و ریدج ($k=0.05$ و $k=0.10$) عملکرد دانه گندم در طی دو سال آزمایش

Table 3. Direct (diagonal) and indirect (off-diagonal) effects in simple ($k=0$) and ridge ($k=0.05$ and $k=0.10$) path analysis during the two year

Year		k	PH	SL	SN	SW	TKW	DB	DF	DM
2013	PH	./..	./..۵	./..۴	./..۸۸	./۱۶۰	./۰۴۵	./۰۲۱	./۰۳۳	./۰۰۲
		./..۵	./..۱۴	./..۰۹	./..۸۵	./۱۵۵	./۰۴۲	./۰۰۲	./۰۱۲	./۰۰۲
		./۱۰	./۰۲۱	./۰۱۳	./۰۸۱	./۱۵۰	./۰۴۰	./۰۰۶	./۰۱۱	./۰۰۲
	SL	./..	./..۲	./..۱۱	./۱۷۲	./۲۱۶	./۰۲۰	./۰۲۴	./۰۳۹	./۰۰۳
		./..۵	./..۰۵	./..۲۵	./۱۶۴	./۲۰۸	./۰۱۹	./۰۰۳	./۰۱۵	./۰۰۳
		./۱۰	./..۰۷	./۰۳۶	./۱۵۸	./۲۰۲	./۰۱۹	./۰۰۶	./۰۱۲	./۰۰۳
	SN	./..	./..۱	./..۰۴	./۴۵۸	./۳۴۴	./۰۳۲	./۰۱۵	./۰۲۵	./۰۰۱
		./..۵	./..۰۳	./..۰۹	./۴۳۸	./۳۳۱	./۰۳۰	./۰۰۲	./۰۰۹	./۰۰۱
		./۱۰	./..۰۴	./۰۱۴	./۴۲۰	./۳۲۱	./۰۲۸	./۰۰۴	./۰۰۸	./۰۰۱
	SW	./..	./..۲	./..۰۵	./۳۳۳	./۴۷۳	./۰۴۹	./۰۱۳	./۰۲۱	./۰۰۱
		./..۵	./..۰۵	./..۱۲	./۳۱۹	./۴۴۶	./۰۴۶	./۰۰۱	./۰۰۸	./۰۰۱
		./۱۰	./..۰۷	./۰۱۷	./۳۰۶	./۴۴۲	./۰۴۴	./۰۰۴	./۰۰۷	./۰۰۱
	TKW	./..	./..۱	./..۱	./۰۷۱	./۱۱۳	./۰۲۰۶	./۰۰۱	./۰۰۲	./۰۰۱
		./..۵	./..۰۳	./..۰۳	./۰۶۷	./۱۰۹	./۱۹۴	./۰۰۱	./۰۰۱	./۰۰۱
		./۱۰	./..۰۴	./..۰۴	./۰۶۵	./۱۰۵	./۱۸۴	./۰۰۱	./۰۰۱	./۰۰۱
	DB	./..	./..۱	./..۰۳	./۰۶۶	./۰۶۱	./۰۰۲	./۰۰۴	./۰۱۷	./۰۰۱
		./..۵	./..۰۳	./..۰۶	./۰۶۳	./۰۵۹	./۰۰۲	./۰۱۲	./۰۶۴	./۰۰۱
		./۱۰	./..۰۴	./..۰۸	./۰۶۱	./۰۵۷	./۰۰۲	./۰۲۸	./۰۵۵	./۰۰۹
	DF	./..	./..۱	./..۰۲	./۰۶۶	./۰۵۹	./۰۰۲	./۰۰۳	./۰۱۷	./۰۱۱
		./..۵	./..۰۳	./..۰۶	./۰۶۳	./۰۵۶	./۰۰۲	./۰۱۲	./۰۶۵	./۰۱۱
		./۱۰	./..۰۴	./..۰۸	./۰۶۱	./۰۵۵	./۰۰۲	./۰۲۷	./۰۵۶	./۰۱۰
	DM	./..	./..۱	./..۰۲	./۰۳۱	./۰۱۲	./۰۱۰	./۰۶۴	./۰۰۹	./۰۱۷
		./..۵	./..۰۲	./..۰۵	./۰۲۹	./۰۱۱	./۰۰۹	./۰۰۷	./۰۴۱	./۰۱۷
		./۱۰	./..۰۳	./..۰۷	./۰۲۸	./۰۱۱	./۰۰۹	./۰۱۷	./۰۳۵	./۰۱۵
2014	PH	./..	./..۴۷	./..۰۵	./۰۱۵	./۰۲۲	./۰۹۰	./۰۴۱	./۰۳۵	./۰۰۶
		./..۵	./..۰۴۰	./..۰۴	./۰۱۳	./۰۳۷	./۰۸۰	./۰۱۴	./۰۱۶	./۰۰۵
		./۱۰	./..۰۳۵	./..۰۳	./۰۱۲	./۰۴۴	./۰۷۳	./۰۰۶	./۰۱۳	./۰۰۴
	SL	./..	./..۱۶	./..۱۶	./۱۰۳	./۰۳۵	./۰۹۷	./۰۲۹	./۰۳۱	./۰۱۱
		./..۵	./..۰۱۴	./..۰۱۳	./۰۸۹	./۰۵۸	./۰۸۶	./۰۱۰	./۰۱۴	./۰۱۰
		./۱۰	./..۰۱۲	./..۰۸	./۰۸۱	./۰۶۹	./۰۷۸	./۰۰۴	./۰۱۱	./۰۰۹
	SN	./..	./..۰۱	./..۰۳	./۵۹۳	./۰۹۵	./۱۰۳	./۰۲۹	./۰۲۷	./۰۰۴
		./..۵	./..۰۱	./..۰۲	./۵۱۱	./۱۵۸	./۰۹۱	./۰۱۴	./۰۱۳	./۰۰۴
		./۱۰	./..۰۱	./..۰۱	./۴۶۳	./۱۸۹	./۰۸۳	./۰۰۴	./۰۱۰	./۰۰۳
	SW	./..	./..۰۸	./..۰۴	./۴۴۳	./۱۲۷	./۳۳۲	./۰۱۹	./۰۱۵	./۰۰۱
		./..۵	./..۰۷	./..۰۳	./۳۸۲	./۲۱۱	./۲۹۲	./۰۰۶	./۰۰۷	./۰۰۱
		./۱۰	./..۰۶	./..۰۲	./۳۴۷	./۲۵۲	./۲۶۷	./۰۰۳	./۰۰۶	./۰۰۱
	TKW	./..	./..۰۸	./..۰۳	./۱۱۶	./۰۷۹	./۵۲۵	./۰۱۰	./۰۰۸	./۰۰۳
		./..۵	./..۰۷	./..۰۲	./۱۰۰	./۱۳۲	./۴۶۵	./۰۰۳	./۰۰۴	./۰۰۳
		./۱۰	./..۰۶	./..۰۲	./۰۹۰	./۱۵۸	./۴۲۶	./۰۰۳	./۰۰۳	./۰۰۲
	DB	./..	./..۱۷	./..۰۴	./۱۵۰	./۰۲۰	./۰۴۳	./۱۱۶	./۰۱۰	./۰۱۷
		./..۵	./..۰۱۴	./..۰۳	./۱۲۹	./۰۳۴	./۰۳۸	./۰۴۰	./۰۵۰	./۰۱۵
		./۱۰	./..۰۱۲	./..۰۲	./۱۱۷	./۰۴۰	./۰۳۵	./۰۱۷	./۰۴۰	./۰۱۴
	DF	./..	./..۱۵	./..۰۴	./۱۴۵	./۰۱۷	./۰۴۱	./۱۱۳	./۰۱۰	./۰۱۸
		./..۵	./..۰۱۳	./..۰۴	./۱۲۵	./۰۲۹	./۰۳۶	./۰۳۹	./۰۵۱	./۰۱۶
		./۱۰	./..۰۱۱	./..۰۲	./۱۱۳	./۰۳۴	./۰۳۳	./۰۱۷	./۰۴۱	./۰۱۴
	DM	./..	./..۰۷	./..۰۵	./۰۶۶	./۰۰۲	./۰۴۳	./۰۵۴	./۰۵۵	./۰۳۶
		./..۵	./..۰۰۶	./..۰۴	./۰۵۷	./۰۰۴	./۰۳۸	./۰۱۸	./۰۲۵	./۰۳۲
		./۱۰	./..۰۰۶	./..۰۲	./۰۵۱	./۰۰۴	./۰۳۴	./۰۰۸	./۰۲۰	./۰۲۹

یک مجموعه از متغیرهای پیش گو بر اساس حداقل مقدار VIF استفاده می شود؛ سپس، یک مجموعه رگرسیون گام به گام بر اساس صفات باقی مانده تنظیم می شوند. در تحقیق حاضر، با حذف برخی از صفات از جمله تعداد روز تا گلدهی، مدل اولیه در سال نخست به ترتیب شامل صفات وزن سنبله، تعداد دانه، وزن هزار دانه، روز تا آبلستنی، روز تا رسیدگی و طول سنبله بود و از بین آن ها تنها صفات وزن

در برخی از مطالعات، ابتدا تجزیه رگرسیون گام به گام صورت می گیرد و پس از شناسایی صفاتی که بخشی از تغییرات صفت هدف را تبیین می کنند، تجزیه علیت انجام می شود (۴، ۱۰، ۱۶). در چنین مواقعی نیز توصیه می شود تجزیه رگرسیون گام به گام با در نظر گرفتن یک حد آستانه برای VIF انجام شود. به عنوان نمونه در روش ارائه شده توسط اولیویتو و همکاران (۱۷) از الگوریتم تکرار شونده برای انتخاب

بیولوژیک سهم قابل توجهی از تغییرات عملکرد دانه را تبیین می‌نماید.

پژوهش‌گر بر مبنای تجربه خود می‌تواند از مجموعه صفات وارد شده در یکی از مدل‌های مذکور برای تجزیه علیت استفاده نماید. با این که حذف برخی از صفات با چند هم‌خطی بالا، مناسب‌تر و ساده‌تر از تجزیه علیت رنج عنوان شده است (۲۴،۲۶)، اما بایستی توجه داشت که این امر دو محدودیت مهم دارد؛ اول اینکه از متغیرهای پیش‌بینی حذف شده اطلاعات مستقیمی به دست نمی‌آید و دوم این که بزرگی ضرایب رگرسیون برای متغیرهای پیش‌بینی مانده در مدل به‌وسیله متغیرهای پیش‌بینی همبسته که در مدل وارد نشده‌اند، متأثر می‌شوند (۱۱). همچنین لازم به توضیح است که در چندهم‌خطی ضعیف به‌ویژه زمانی که اندازه نمونه کوچک باشد، حذف صفات و استفاده از رگرسیون رنج می‌تواند گمراه‌کننده باشد. از این رو برای حصول نتایج دقیق‌تر در تحلیل علیت، توصیه می‌شود محققین با در نظر گرفتن تمامی جوانب از مناسب‌ترین روش‌ها استفاده نمایند.

سنبله، تعداد دانه، وزن هزار دانه، روز تا آبستنی، روز تا رسیدگی و طول سنبله بود و از بین آن‌ها تنها صفات وزن سنبله و تعداد دانه در تمامی مدل‌ها حضور داشتند. همچنین در سال دوم صفات تعداد دانه، وزن هزار دانه، روز تا رسیدگی، ارتفاع بوته و طول سنبله در مدل اولیه وارد داشتند و از میان آن‌ها صفات تعداد دانه و وزن هزار دانه در همه مدل‌ها بودند. صفت ارتفاع بوته در سال دوم با وجود عدم همبستگی با عملکرد دانه وارد مدل رگرسیونی شد، در حالی که در سال اول با وجود همبستگی معنی‌دار در مدل حضور نداشت. این نتیجه مطابق با فرضیه تراوتن‌مولر و همکاران (۲۵) بود که طی آن به عدم شباهت اثرات مستقیم و غیرمستقیم با نتایج حاصل از همبستگی‌های ساده اشاره شده است. روز تا رسیدگی در هر دو سال و روز تا آبستنی در سال نخست وارد مدل اولیه در تجزیه رگرسیون شدند. اهمیت مراحل فنولوژیکی بر عملکرد دانه به اثبات رسیده است (۲). مشابه با نتایج پژوهش حاضر، آقائی سربرز و امینی (۱) گزارش کردند که تعداد دانه در سنبله، وزن هزار دانه، روز تا ظهور سنبله‌دهی و عملکرد

جدول ۴- رگرسیون گام‌به‌گام برای انتخاب یک مجموعه از متغیرها با حداقل VIF برابر ۵

Table 4. Stepwise-based regressions to select a set of predictors with a maximum of VIF = 5

Model	2013					2014				
	CN	Det	R ²	Res	VIF	CN	Det	R ²	Res	VIF
Model 1	۱۲/۰۷	۰/۱۳۸	۰/۸۴۱	۰/۱۵۹	۳/۱۸	۳/۰۹	۰/۶۹۱	۰/۹۴۵	۰/۰۵۵	۱/۳۹
Model 2	۱۰/۶۲	۰/۱۹۸	۰/۸۴۱	۰/۱۵۹	۲/۹۷	۱/۷۸	۰/۸۹۲	۰/۹۴۵	۰/۰۵۵	۱/۰۸
Model 3	۹/۹۳	۰/۳۳۰	۰/۸۴۱	۰/۱۵۹	۲/۸۹	۱/۶۳	۰/۹۳۹	۰/۹۴۴	۰/۰۵۶	۱/۰۶
Model 4	۹/۶۷	۰/۳۳۸	۰/۸۳۶	۰/۱۶۴	۲/۸۹	۱/۴۹	۰/۹۶۲	۰/۹۴۲	۰/۰۵۸	۱/۰۴
Model 5	۶/۳۳	۰/۴۷۱	۰/۸۰۵	۰/۱۹۵	۱/۱۲					
Model 1: SW, SN, TKW, DB, DM, SL						Model 1: SN, TKW, DM, PH, SL				
Model 2: SW, SN, TKW, DB, DM						Model 2: SN, TKW, DM, PH				
Model 3: SW, SN, TKW, DB						Model 3: SN, TKW, DM				
Model 4: SW, SN, TKW						Model 4: SN, TKW				
Model 5: SW, SN										

CN: شماره وضعیت، Det: تبیین‌کننده، Res: باقی‌مانده، VIF: عامل تورم واریانس، DB: روز تا آبستنی، DM: روز تا رسیدگی، PH: ارتفاع بوته، SW: وزن سنبله، SL: طول سنبله، SN: تعداد دانه، TKW: وزن هزار دانه

منابع

1. Aghaee Sarbarze, M. and A. Amini. 2011. Genetic variability for agronomy traits in bread wheat genotype collection of Iran. Seed and Plant Journal, 27(4): 581-599 (In Persian).
2. Alipour, H. and H. Abdi. 2020. Interactive Effects of Vernalization and Photoperiod Loci on Phenological Traits and Grain Yield and Differentiation of Iranian Wheat Landraces and Cultivars. Journal of Plant Growth Regulation, 1-10. <https://doi.org/10.1007/s00344-020-10260-8>.
3. Alipour, H., H. Abdi, Y. Rahimi and M.R. Bihamta. 2019. Investigating grain yield and yield stability of wheat cultivar introduced in Iran over the last half century. Cereal Research, 9(2): 157-167 (In Persian).
4. Andaz, P., M. Moghaddam, S.S. Alavikia, M. Valizadeh, M. Valizadeh and M. Tabrizvand Taheri. 2017. Assessment of genetic diversity and its relationship with geographical factors in populations of wild diploid wheat (*Triticum boeoticum*) based on morphological characteristics. Seed and Plant Journal, 33(4): 511-533 (In Persian).
5. de Carvalho, C.G., R. Borsato, C.D. Cruz and J.M. Viana. 2001. Path analysis under multicollinearity in S0 x S0 maize hybrids. Crop Breeding and Applied Biotechnology, 1(3): 263-270.
6. Ehdaie, B. and J.G. Waines. 1989. Genetic variation, heritability and path-analysis in landraces of bread wheat from southwestern Iran. Euphytica, 41(3): 183-190.
7. Ferrari, M., I.R. Carvalho, A.J. de Pelegrin, M. Nardino, V.J. Szarecki, T. Olivoto, D.N. Follmann, C. Pegoraro, L.C. da Maia, V.Q. de Souza and T.C. da Rosa. 2018. Path analysis and phenotypic correlation among yield components of soybean using environmental stratification methods. Australian Journal of Crop Science, 12(2): 193.
8. Hoerl, A.E. and R.W. Kennard. 1970. Ridge regression: applications to nonorthogonal problems. Technometrics, 12(1): 69-82.

9. Luz, L.N.D., R.C.D. Santos and P.D.A. Melo Filho. 2011. Correlations and path analysis of peanut traits associated with the peg. *Crop Breeding and Applied Biotechnology*, 11(1): 88-95.
10. Masoudi, H., H. Sabouri, F. Taliev and J. Jaafarby. 2020. Genetic diversity of some wheat germplasm based on morpho-phenological traits. *Journal of Crop Breeding*, 12(35): 54-68 (In Persian).
11. Mesdaghi, M. 2004. Regression methods: for research in agriculture and natural resources, University of Emam Reza Publication, 290 p (In Persian).
12. Moghaddam, M., B. Ehdaie and J.G. Waines. 1997. Genetic variation and interrelationships of agronomic characters in landraces of bread wheat from southeastern Iran. *Euphytica*, 95(3): 361-369.
13. Mohammadi, M., P. Sharifi and R. Karimizadeh. 2014. Sequential Path Analysis for Determination of Relationship Between Yield and Yield Components in Bread Wheat (*Triticum aestivum* L.). *Notulae Scientia Biologicae*, 6(1): 119-124.
14. Mohammadi, M., R. Karimizadeh, M.K. Shefazadeh and B. Sadeghzadeh. 2011. Statistical analysis of durum wheat yield under semi-warm dryland condition. *Australian Journal of Crop Science*, 5(10): 1292-1297.
15. Montgomery, D.C., E.A. Peck and G. Vining. 2012. Introduction to linear regression analysis. 5th ed. John Wiley & Sons, Hoboken, NJ.
16. Noori, A., A. Mehrabi and H. Safari. 2017. Study of correlation and path coefficient analysis of agronomic traits and grain yield for *Aegilops culindrica* accessions under non-stress and drought stress conditions in Ilam. *Journal of Crop Breeding*, 9(23): 76-84 (In Persian).
17. Olivoto, T., M. Nardino, I.R. Carvalho, D.N. Follmann, M. Ferrari, V.J. Szarecki and V.D. Souza. 2017. REML/BLUP and sequential path analysis in estimating genotypic values and interrelationships among simple maize grain yield-related traits. *Genetics and Molecular Research*, 16(1): 1-19.
18. Olivoto, T., V.Q. de Souza, M. Nardino, I.R. Carvalho, M. Ferrari, A.J. de Pelegrin, V.J. Szarecki and D. Schmidt. 2017. Multicollinearity in path analysis: a simple method to reduce its effects. *Agronomy Journal*, 109(1): 131-142.
19. Petraitis, P.S., A.E. Dunham and P.H. Niewiarowski. 1996. Inferring multiple causality: the limitations of path analysis. *Functional Ecology*, 10: 421-431.
20. Rabiei, B., M. Valizadeh, B. Gharevazie and M. Moghaddam. 2004. Evaluation of selection indices for improving rice grain shape. *Field Crops Research*, 89(2-3): 359-367.
21. Rahimi, Y., M.R. Bihamta, A. Taleei, H. Alipour and P.K. Ingvarsson. 2019. Applying an artificial neural network approach for drought tolerance screening among Iranian wheat landraces and cultivars grown under well-watered and rain-fed conditions. *Acta Physiologiae Plantarum*, 41(9): 1-17.
22. Sari, B.G., A.D.C. Lúcio, T. Olivoto, D.K. Krysczun, A.L. Tischler and L. Drebes. 2018. Interference of sample size on multicollinearity diagnosis in path analysis. *Pesquisa Agropecuária Brasileira*, 53(6): 769-773.
23. Silva, F.A.D., C.C.G. Correa, B.M. Carvalho, A.P. Viana, S.D.C. Preisigke and A.T.D. Amaral Júnior. 2021. Novel approach to the selection of *Psidium quajava* genotypes using latent traits to bypass multicollinearity. *Scientia Agricola*, 78(2): e20190081.
24. Toebe, M. and A. Cargnelutti Filho. 2013. Multicollinearity in path analysis of maize (*Zea mays* L.). *Journal of Cereal Science*, 57(3): 453-462.
25. Trautenmüller, J.W., S.P. Netto, R. Balbinot, A.P. Dalla Corte and J. Borella. 2019. Path analysis applied to evaluation of biomass estimates in subtropical forests at Brazil. *Floresta*, 49(3): 587-596.
26. Viotto Del Conte, M., P.C.S. Carneiro, M.D. Vilela de Resende, F. Lopes da Silva and L.A. Peternelli. 2020. Overcoming collinearity in path analysis of soybean [*Glycine max* (L.) Merr.] grain oil content. *Plos One*, 15(5): e0233290.

Path Analysis of Wheat Grain Yield with Overcoming Multi-Collinearity of Traits**Hadi Alipour¹, Hossein Abdi² and Mohammad Reza Bihamta³**1- Assistant Professor, Department of Plant Production and Genetics, Urmia University, Urmia,
(Corresponding author: ha.alipour@urmia.ac.ir)

2- PhD Candidate, Department of Plant Production and Genetics, Urmia University, Urmia

3- Professor, Department of Agronomy and Plant Breeding, University of Tehran, Karaj

Received: April 22, 2021

Accepted July 26, 2021

Abstract

Path analysis is one of the oldest methods of knowing the direct and indirect effects of traits on yield, but the widespread application of this approach is limited by the problem of multi-collinearity, that has been addressed in the present study. A total of 298 bread wheat genotypes were evaluated during the two cropping years in the lattice design and the path analysis of grain yield was studied by simple and ridge methods. The highest direct effect on grain yield in the first year were related to spike weight, number of seeds and 1000-seed weight, respectively, and in the second year were related to number of seeds, 1000-seed weight and spike weight, respectively. Spike weight and number of seeds in the first year had an indirect effect on yield through each other as well as through spike length and plant height. In the second year, the number of seeds and 1000-seed weight through the weight of spikes showed the most indirect effect. Day to booting and day to flowering traits in both years and spike weight in the second year had a high multi-collinearity that the use of constant $k = 0.10$ in ridge causal analysis overcame this problem. Finally, in contrast to simple path analysis, ridge path analysis showed that the spike weight through the number of seeds had an indirect effect on grain yield in the second year. According to the results, it is recommended to consider the issue of multi-collinearity between traits in future path analysis studies.

Keywords: Direct and indirect effects, Phenological stages, Ridge path analysis, Stepwise regression, Yield components