



دانشکده علوم ریاضی

دانشگاه صنعتی اصفهان

موضوع:

تحلیل داده های گمشده در طرح کامل بلوکی

استاد محترم: خانم دکتر صابری

دانشجو: سحر توکلی

چکیده

طرح بلوکی کامل تصادفی افزوده (ARCBD) به طور گسترده در برنامه‌های اصلاح نباتات برای غربالگری تعداد زیادی از تیمارهای جدید با منابع محدود استفاده می‌شود. در این طرح‌ها، واریانس خطا بر اساس تیمارهای کنترلی تکرار شده برآورد می‌شود؛ با این حال، مشاهدات گمشده در این کنترل‌ها اغلب رخ می‌دهند که منجر به برآوردهای اریب و از بین رفتن تعامد می‌گردد. در حالی که مطالعات اخیر یک طرح دقیق برای مدیریت یک مقدار گمشده با استفاده از آمار نوتروسوفیک (NS) جهت مدیریت عدم تعین داده‌ها معرفی کرده‌اند، آزمایش‌های کشاورزی اغلب از مقادیر گمشده چندگانه رنج می‌برند. این مقاله با توسعه تحلیل دقیق نوتروسوفیک (NEA) به سناریوهایی با دو تیمار کنترلی گمشده در ARCBD، به این خلاء می‌پردازد. علاوه بر این، ما این روش‌شناسی دقیق را برای اولین بار به طرح مربع لاتین افزوده (ALS) تعمیم می‌دهیم. ما روش NEA پیشنهادی را در برابر تحلیل تقریبی نوتروسوفیک (NAA) و روش‌های کلاسیک از طریق یک مطالعه شبیه‌سازی جامع با استفاده از R مقایسه می‌کنیم. نتایج نشان می‌دهند در حالی که روش‌های تقریبی نرخ خطای نوع اول را در حضور عدم تعین به طور قابل توجهی متورم می‌کنند، روش دقیق پیشنهادی نرخ خطای تجربی را در سطح اسمی حفظ کرده و برتری و استحکام خود را در محیط‌های نامعین اثبات می‌کند.

فهرست

5.....	مقدمه
6.....	1. مبانی نظری: منطق و آمار نوتروسوفیک
6.....	1-1. گذر از منطق کلاسیک به منطق نوتروسوفیک
7.....	2-1. آمار نوتروسوفیک
7.....	3-1. متغیرهای تصادفی و جبر نوتروسوفیک
8.....	4-1. توزیع نرمال نوتروسوفیک در طرح‌های آزمایشی
9.....	2. چالش‌های حیاتی و آسیب‌های داده‌های گمشده
9.....	1.2. مشکلات آماری عمومی
10.....	3. طرح‌های بلوکی تصادفی افزوده: رویکرد کلاسیک و نوتروسوفیک
10.....	3-1. طرح بلوک کامل تصادفی افزوده (ARCBd)
11.....	2.3. آسیب‌های داده‌های گمشده در ARCBd
12.....	3-3. طرح بلوک کامل تصادفی افزوده نوتروسوفیک (NARCBd)
16.....	4. چالش‌های داده‌های گمشده در طرح بلوک کامل تصادفی افزوده (ARCBd)
16.....	1.4. مشکل حاد عدم تکرار تیمارهای افزوده
17.....	2.4. تأثیر بر برآورد خطا و سوگیری ساختاری (Bias)
17.....	3.4. کاهش قدرت آماری و تحریف روابط
18.....	5. روش‌شناسی پژوهش: جایگذاری و آنالیز دقیق نوتروسوفیک (NEA)
18.....	1-5. بیان مسئله در قالب ریاضی
19.....	2-5. الگوریتم پیشنهادی جایگذاری و برآورد
21.....	3.5. الگوهای تحلیل نوتروسوفیک NAA در مقابل NEA
23.....	3-5. اعتبارسنجی الگوریتم

6. ارزیابی تجربی و تحلیل نتایج شبیه سازی	23
1-6. تحلیل داده های واقعی: آزمایش گلرنگ	24
2-6. مطالعه شبیه سازی مونت کارلو (Monte Carlo Simulation)	28
3-6. نتیجه گیری نهایی بخش یافته ها	31
7. تعمیم و بررسی برای دو مقدار گمشده در NARCB	31
1-7. فرمول بندی ریاضی برای دو مقدار گمشده در NARCB	31
2-7. مطالعه شبیه سازی: روش شناسی و پیاده سازی در R	32
3-7. نتایج و تحلیل	33
4-7. نتیجه گیری فاز دوم:	34
8. تعمیم تحلیل نوتروسوفیک (NS) به طرح مربع لاتین افزوده (NS-ALS)	34
1.8. طرح مربع لاتین افزوده (ALS)	34
2-8. مدل مربع لاتین افزوده نوتروسوفیک (NALS)	36
3-8. مطالعه شبیه سازی: NS-ALS با داده های گمشده	40
نتیجه پژوهش	45
منابع	46

مقدمه

در آزمایش‌های کشاورزی و بیولوژیکی، به ویژه در مرحله غربالگری برنامه‌های اصلاح نژاد، آزمایش‌گران اغلب با تعداد زیادی از ژنوتیپ‌های (تیمارهای) جدید مواجه هستند که باید با وجود منابع محدود ارزیابی شوند. برای رفع این مشکل، طرح بلوکی کامل تصادفی افزوده (ARCB) معرفی شد، که در آن شاهد‌ها در هر بلوک برای برآورد واریانس خطا تکرار می‌شوند، در حالی که تیمارهای جدید تنها یک بار ظاهر می‌شوند. با این حال، مشاهدات گمشده یک رخداد رایج در چنین آزمایش‌های بلندمدتی به دلیل عوامل محیطی یا محدودیت واحدهای آزمایشی هستند که منجر به مسئله مشاهدات گمشده به طور کاملاً تصادفی می‌شوند.

مقادیر گمشده تعادل طرح را برهم می‌زنند، که باعث ایجاد تورش در برآورد اثر تیمار، تورم مجموع مربعات و کاهش توان آزمون اثرات تیمار می‌شود. رویکرد سنتی به این مسئله شامل جایگذاری یک مقدار برآورد شده است روشی که به عنوان رویکرد تقریبی شناخته می‌شود. با این حال، این مقدار برآورد شده ذاتاً با مقدار اصلی تفاوت دارد و ایجاد عدم قطعیت می‌کند. علاوه بر این، داده‌های دنیای واقعی اغلب حاوی ابهامات، گنگی و عدم تعین هستند که آمار کلاسیک در مدل‌سازی مناسب آن‌ها ناتوان است.

برای غلبه بر این محدودیت‌ها، آمار نوتروسوفیک (NS) که آمار کلاسیک و فازی را با معرفی درجه‌ای از عدم تعین تعمیم می‌دهد، پیشنهاد شده است. اخیراً، یک طرح دقیق نوتروسوفیک برای مقابله با یک تیمار کنترل گمشده در ARCB استفاده شد که عملکرد برتری نسبت به روش‌های تقریبی نشان داد. با این وجود، یک شکاف قابل توجه باقی مانده است: روش‌شناسی‌های فعلی به پیچیدگی مقادیر گمشده چندگانه نمی‌پردازند یا به طرح‌های پیچیده‌تری مانند مربع لاتین افزوده (ALS) که در آن تعامل حتی شکننده‌تر است، تعمیم نمی‌یابند.

این مقاله با اهداف زیر سعی در پر کردن این شکاف‌ها دارد: (1) فرمول‌بندی تحلیل دقیق نوتروسوفیک (NEA) برای ARCB با دو مقدار گمشده؛ (2) تعمیم NEA به طرح مربع لاتین افزوده؛ و (3) اعتبارسنجی این روش‌ها از طریق شبیه‌سازی‌های گسترده مونت کارلو. این مطالعه تأیید می‌کند که در محیط‌های نامعین، رویکرد تحلیلی دقیق استنباط معتبری ارائه می‌دهد، در جایی که روش‌های تقریبی شکست می‌خورند.

۱. مبانی نظری: منطق و آمار نوتروسوفیک

۱-۱. گذر از منطق کلاسیک به منطق نوتروسوفیک

تحلیل‌های آماری مرسوم بر پایه منطق کلاسیک بنا شده‌اند که در آن گزاره‌ها تنها دو حالت «درست (True)» یا «نادرست (False)» دارند. با این حال، داده‌های حاصل از پژوهش‌های واقعی، به ویژه در علوم کشاورزی و آزمایشگاهی، همواره با نوعی عدم قطعیت، ابهام و عدم تعین (Indeterminacy) همراه هستند.

جهت مدل‌سازی ریاضی این عدم قطعیت، منطق نوتروسوفیک (Neutrosophic Logic) در سال ۱۹۹۸ توسط فلورنتین اسماراند/چه به عنوان تعمیمی بر منطق فازی (Fuzzy Logic) و منطق شهودی (Intuitionistic Logic) معرفی گردید.

بر خلاف منطق کلاسیک و فازی، منطق نوتروسوفیک هر گزاره یا پدیده را با سه مؤلفه مستقل توصیف می‌کند:

$$\langle T, I, F \rangle$$

T: درجه درستی (Truth)

درجه عضویت یا صحت.

I: درجه عدم تعین (Indeterminacy)

درجه عدم عضویت یا کذب.

F: درجه نادرستی (Falsity).

درجه خنثی بودن، ابهام یا مرز میان درستی و نادرستی.

این سه مؤلفه مستقل بوده و اغلب در بازه استاندارد $[0, 1]$ نرمال‌سازی می‌شوند، با این قید که

$$0 \leq T + I + F \leq 3.$$

جنبه حیاتی این است که I منطقه خاکستری را نشان می‌دهد _ معیاری از ابهام، بی‌طرفی یا خطا _ که آمار کلاسیک نمی‌تواند آن را به صراحت کمی‌سازی کند.

استقلال این سه مؤلفه از یکدیگر ویژگی متمایز این منطق است؛ به طوری که مجموع آن‌ها می‌تواند کمتر، برابر یا بیشتر از ۱ باشد، که این امر انعطاف‌پذیری بالایی را برای تحلیل داده‌های ناقص یا متناقض فراهم می‌آورد.

۱-۲. آمار نوتروسوفیک

آمار نوتروسوفیک (Neutrosophic Statistics) شاخه‌ای توسعه‌یافته از آمار کلاسیک است که به تحلیل رویدادها، متغیرها و داده‌هایی می‌پردازد که ذاتاً دارای عدم تعین هستند. در حالی که در آمار کلاسیک تمام داده‌ها قطعی (Crisp) فرض می‌شوند، در آمار نوتروسوفیک داده‌ها می‌توانند به صورت بازه‌ای یا همراه با عدم قطعیت مدل‌سازی شوند.

اهمیت این رویکرد در طرح‌های آزمایشی (مانند ARCBD) زمانی آشکار می‌شود که پارامترهای توزیع (نظیر میانگین و واریانس) یا مقادیر مشاهدات دقیق نیستند. آمار نوتروسوفیک به جای حذف داده‌های مبهم یا گمشده، "عدم تعین" را به عنوان بخشی جدایی‌ناپذیر از فرمول‌ها و تحلیل‌ها وارد محاسبات می‌کند. این امر موجب می‌شود نتایج حاصله به واقعیت نزدیک‌تر بوده و ریسک‌های ناشی از تقریب‌زنی‌های کلاسیک کاهش یابد.

۱-۳. متغیرهای تصادفی و جبر نوتروسوفیک

در تحلیل‌های آماری نوتروسوفیک، داده‌ها به صورت اعداد نوتروسوفیک (Neutrosophic Numbers) نمایش داده می‌شوند. فرم استاندارد یک عدد نوتروسوفیک (N) به صورت زیر تعریف می‌شود:

$$N = d + I$$

که در آن d بخش قطعی (Determinate Part) و I بخش نامعین (Indeterminate Part) است.

برای پیاده‌سازی این مفهوم در طرح‌های آماری، متغیر تصادفی نوتروسوفیک X_N در بازه‌ی $[X_L, X_U]$ تعریف می‌شود که دارای بازه عدم تعین I_N است. ساختار ریاضی آن به صورت زیر فرمول‌بندی می‌گردد:

$$X_N = X_L + X_U I_N$$

در این رابطه X_L کران پایین یا بخش قطعی متغیر است و $X_U I_N$ بخش غیرقطعی است که در آن $I_N \in [I_L, I_U]$ معیاری از عدم قطعیت می باشد.

بدیهی است که اگر $I_L = 0$ باشد، متغیر تصادفی نوتروسوفیک X_N به یک متغیر تصادفی کلاسیک تقلیل می یابد.

۴-۱. توزیع نرمال نوتروسوفیک در طرح های آزمایشی

از آنجا که تحلیل واریانس (ANOVA) در طرح های بلوکی تصادفی افزوده (ARCB) بر فرض نرمال بودن داده ها استوار است، تعریف توزیع نرمال نوتروسوفیک ضروری است. یک متغیر تصادفی X_N از توزیع نرمال نوتروسوفیک پیروی می کند اگر میانگین جامعه (μ_N) و واریانس جامعه (σ_N^2) مقادیری نوتروسوفیک باشند. برآورد میانگین نمونه ای نوتروسوفیک ($\bar{X}_N$) و واریانس نمونه ای (S_N^2) برای یک نمونه تصادفی به حجم n به صورت بازه های زیر محاسبه می شوند:

$$\bar{X}_N \in \left[\frac{\sum_{i=1}^n X_{Li}}{n}, \frac{\sum_{i=1}^n X_{Ui}}{n} \right]$$

$$S_N^2 \in \left[\frac{\sum_{i=1}^n (X_{Li} - \bar{X}_L)^2}{n}, \frac{\sum_{i=1}^n (X_{Ui} - \bar{X}_U)^2}{n} \right]$$

این تعاریف ریاضی، زیربنای لازم برای توسعه مدل های NARCB و برآورد مقادیر گمشده در محیط های دارای عدم قطعیت را فراهم می کنند.

میانگین نمونه ای نوتروسوفیک ($\bar{X}_N$) و واریانس نوتروسوفیک (S_N^2) که به صورت بازه ها تعریف می شوند، که تمام مقادیر ممکن را که پارامتر واقعی می تواند بگیرد، با در نظر گرفتن عدم قطعیت ذاتی در داده ها یا عدم قطعیتی که از جایگذاری مقادیر گمشده ایجاد می شود، در خود جای می دهند. این امر امکان محاسبه مجموع مربعات نوتروسوفیک (SS_N) و یک آزمون F نوتروسوفیک (F_N) به صورت بازه، و در نهایت آنالیز واریانس نوتروسوفیک (NANOVA) را فراهم می آورد.

۲. چالش‌های حیاتی و آسیب‌های داده‌های گمشده

داده‌های گمشده در طرح‌های آزمایشی، پیچیدگی‌های آماری عمیقی را ایجاد می‌کنند، اما در طرح‌های افزوده و با کنترل مضاعف (ARCB/ALS)، مسائل به دلیل نامتوازن بودن ذاتی و فقدان تکرار کامل، به طور ویژه حاد هستند.

۱.۲. مشکلات آماری عمومی

مسئله اصلی داده‌های گمشده، پتانسیل سوگیری (Bias) در برآورد پارامترها است. میزان این سوگیری به شدت به مکانیسم گمشدگی داده (MDM) بستگی دارد:

گمشدگی کاملاً تصادفی (MCAR) : ساده‌ترین حالت، با کمترین سوگیری.

گمشدگی تصادفی (MAR) : احتمال گمشدگی یک مقدار به داده‌های مشاهده شده وابسته است، نه به خود مقدار گمشده.

گمشدگی غیر تصادفی (NMAR) : احتمال گمشدگی به خود مقدار مشاهده نشده وابسته است (به عنوان مثال، ضعیف‌ترین گیاهان از بین می‌روند و داده‌های عملکرد آن‌ها از دست می‌رود). این مکانیسم جدی‌ترین سوگیری را ایجاد می‌کند و اغلب میانگین برآورد شده تیمارهای باقی‌مانده را به طور مصنوعی **متورم** می‌سازد (سوگیری انتخاب).

علاوه بر این، از دست رفتن داده به طور کلی **توان آماری** آزمون‌ها را کاهش داده و ساختار کوواریانس بین متغیرها را تحریف می‌کند که تحلیل‌های چندمتغیره و مقایسه‌های چندگانه (مانند آزمون‌های LSD یا دانکن) را پیچیده می‌سازد.

۳. طرح‌های بلوکی تصادفی افزوده: رویکرد کلاسیک و نوتروسوفیک

۳-۱. طرح بلوک کامل تصادفی افزوده (ARCBD)

در پژوهش‌های اصلاح نباتات و آزمایش‌های کشاورزی، محققان اغلب با چالش ارزیابی تعداد بسیار زیادی از ژنوتیپ‌ها یا تیمارهای جدید (New Treatments) مواجه هستند، در حالی که منابع بذر یا فضای آزمایشی محدود است. طرح بلوک کامل تصادفی افزوده (ARCBD) که نخستین بار توسط فدرر (1956) معرفی گردید، یک طرح آزمایشی نامتوازن است که برای غربالگری کارآمد تیمارها توسعه یافته است. یعنی در طرح برای غربالگری اولیه تعداد زیادی از واحدهای آزمایشی (مانند ژنوتیپ‌های جدید در اصلاح نباتات) تحت محدودیت‌های منابع محدود، مانند مقدار بذر یا فضای آزمایشگاهی، طراحی شده است.

ویژگی ساختاری کلیدی ARCBD تقسیم تیمارها به دو مجموعه مجزا است:

1. **تیمارهای استاندارد/شاهد (τ_S)** : این تیمارها یا کنترل‌های شناخته شده و تثبیت شده‌ای هستند که

در تمام b بلوک تکرار می‌شوند و ساختار طرح بلوکی کامل تصادفی (RCBD) استاندارد را دنبال

می‌کنند. تکرار آن‌ها امکان برآورد بدون تورش اثرات بلوک و، از همه مهم‌تر، **واریانس خطای**

آزمایشی (σ^2) را فراهم می‌کند.

2. **تیمارهای افزوده/جدید (τ_A)** : این‌ها لاین‌ها یا ارقام جدید هستند که معمولاً تعدادشان زیاد است و

بدون تکرار در بلوک‌ها قرار می‌گیرند (هر یک فقط یک بار در کل آزمایش ظاهر می‌شود).

مدل آماری خطی برای ARCBD به صورت زیر است:

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

که در آن Y_{ij} مشاهده مربوط به تیمار i در بلوک j ، μ میانگین کلی، τ_i اثر تیمار i (استاندارد یا افزوده)، β_j اثر بلوک j ، و ϵ_{ij} جمله خطای تصادفی است که فرض می‌شود از توزیع نرمال با میانگین صفر و واریانس σ^2 ($i. i. d. N(0, \sigma^2)$) پیروی می‌کند.

تحلیل ARCBD به صورت گام به گام انجام می‌شود. ابتدا، اثرات بلوک (β_j) و واریانس خطا (σ^2) تنها با استفاده از تیمارهای استاندارد تکرار شده برآورد می‌شوند. دوم، مشاهدات تیمارهای افزوده برای اثرات بلوک تصحیح می‌شوند:

$Y_{Corrected} = Y_{Augmented} - \beta_j$ در نهایت، از مقادیر تصحیح شده برای برآورد اثرات تیمارهای جدید و انجام مقایسه‌های آماری با استانداردها، با استفاده از واریانس خطای برآورد شده (σ^2)، استفاده می‌شود. این فرآیند تضمین می‌کند که مقایسه‌ها عاری از تورش محیطی ناشی از ناهمگنی بلوک باشند.

۲.۳. آسیب‌های داده‌های گمشده در ARCBD

به طور کلی در چارچوب طرح‌های افزوده، چالش‌ها تشدید می‌شوند، از همین رو برای آسیب‌های داده‌های گمشده در ARCBD (گمشدگی تیمار شاهد) داریم:

سوگیری در برآورد اثر بلوک: برآورد اثرات بلوک (β_j) متکی بر داده‌های کامل تیمارهای شاهد است. یک مشاهده گمشده در تیمار شاهد، این تعادل را برهم زده و منجر به برآوردهای اریب اثر بلوک و به دنبال آن، تصحیح اشتباه نتایج تیمارهای افزوده می‌شود.

از بین رفتن تعامد: حتی یک مقدار گمشده در تیمارهای شاهد، تعامد (استقلال) بین تیمارهای استاندارد و بلوک‌ها را به خطر می‌اندازد و پیچیدگی محاسبه مجموع مربعات را افزایش می‌دهد. (Federer, 1956)

ناپدید شدن کامل اطلاعات تیمارهای افزوده: اگر داده یک تیمار افزوده از بین برود، هیچ تکراری برای برآورد اثر آن وجود ندارد و اثر آن تیمار کاملاً باقیمانده می‌شود، به طوری که راهی برای جداسازی اثر واقعی تیمار از خطای آزمایشی باقی نمی‌ماند.

۳-۳. طرح بلوک کامل تصادفی افزوده نوتروسوفیک (NARCBD)

اگرچه ARCBD کلاسیک کارایی بالایی در مدیریت منابع دارد، اما فرض بنیادی آن قطعی بودن (Crispness) داده‌ها است. در واقعیت، داده‌های حاصل از آزمایش‌های کشاورزی و صنعتی اغلب با عدم قطعیت، ابهام و اندازه‌گیری‌های بازه‌ای همراه هستند. برای مدل‌سازی این عدم قطعیت، طرح ARCBD در محیط نوتروسوفیک (NARCBD) تعریف می‌شود.

در NARCBD، مشاهدات و پارامترها اعداد ثابت نیستند، بلکه متغیرهای تصادفی نوتروسوفیک هستند که شامل تمام انواع عدم تعین (Indeterminacy) می‌باشند.

مدل ریاضی NARCBD

مدل آماری نوتروسوفیک برای یک طرح NARCBD با b بلوک، c تیمار شاهد (Check) و v تیمار جدید (New) به صورت زیر فرمول‌بندی می‌شود:

$$y_{Nhi jg} = \mu_N + \alpha_{Ni} + \tau_{Nqj} + \tau_{Nlig} + \epsilon_{Nhi jg}$$

$y_{Nhi jg}$ مشاهده نوتروسوفیک است که به فرم $y_L + y_U I_N$ بیان می‌شود و $I_N \in [I_L, I_U]$ بازه عدم تعین است.

μ_N میانگین کلی نوتروسوفیک جامعه است.

α_{Ni} اثر نوتروسوفیک بلوک i -ام است. ($i = 1, \dots, b$)

τ_{Nqj} اثر نوتروسوفیک تیمار شاهد j -ام است. ($j = 1, \dots, c$)

τ_{Nlig} اثر نوتروسوفیک تیمار جدید g -ام در بلوک i -ام است.

$\epsilon_{Nhi jg}$ خطای تصادفی با میانگین صفر نوتروسوفیک و واریانس σ_N^2 است.

چالش داده‌های گمشده در NARCB

یکی از مسائل بحرانی در تحلیل NARCB، رخداد داده‌های گمشده (Missing Values) است. اگر یک تیمار شاهد در یک بلوک به طور کامل از دست برود (Missing Completely at Random)، طرح از حالت متعادل خارج شده و همبستگی متعامد بین بلوک‌ها و تیمارها از بین می‌رود. در چنین شرایطی، طرح به یک طرح بلوک ناقص افزوده نوتروسوفیک (NAIBD) تبدیل می‌شود. برخلاف آمار کلاسیک که داده گمشده با یک مقدار نقطه‌ای برآورد می‌شود، در رویکرد نوتروسوفیک، مقدار گمشده باید به صورت یک عدد نوتروسوفیک (شامل بازه عدم قطعیت) برآورد گردد تا از استنتاج ناکارآمد درباره تیمارها جلوگیری شود. تحلیل دقیق (Exact Analysis) در این شرایط نیازمند حل معادلات نرمال در فضای نوتروسوفیک برای برآورد پارامترهای μ_N ، α_{Ni} و τ_N است.

آنالیز واریانس نوتروسوفیک (NANOVA) و آزمون‌های فرض

برای ارزیابی معنی‌داری اثرات تیمارهای جدید و بلوک‌ها در محیطی که داده‌ها دارای عدم قطعیت هستند، روش‌های کلاسیک آنالیز واریانس (ANOVA) کارایی ندارند. در این پژوهش، از تعمیم آنالیز واریانس به فضای نوتروسوفیک (NANOVA) استفاده می‌شود. در این روش، کل تغییرات موجود در داده‌های آزمایشی (SS_{NT}) به مؤلفه‌های مختلف تجزیه می‌شود، با این تفاوت که هر یک از این مؤلفه‌ها خود یک عدد نوتروسوفیک بازه‌ای هستند.

الف) تجزیه مجموع مربعات نوتروسوفیک (SS_N)

در طرح NARCB، مجموع مربعات کل و اجزای آن به صورت بازه‌هایی از $[SS_L, SS_U]$ تعریف می‌شوند. محاسبات مربوط به مجموع مربعات تصحیح نشده برای تیمارها و بلوک‌ها و همچنین مجموع مربعات خطا (SS_{NE}) بر اساس روابط زیر انجام می‌گیرد:

۱. مجموع مربعات کل نوتروسوفیک (SS_{NT}):

$$SS_{NT} = \sum_{i=1}^b \sum_{j=1}^c y_{Nqij}^2 + \sum_{i=1}^b \sum_{g=1}^{n_{li}} y_{Nlig}^2 - \frac{y_{N...}^2}{n}$$

که در آن $SS_{NT} \in [SS_{LT}, SS_{UT}]$ است.

۲. مجموع مربعات بلوک (تصحیح نشده):

$$SS_{NB(unadj)} = \sum_{i=1, i \neq m}^b \frac{y_{N.i.}^2}{c + n_{(li)}} + \frac{y_{N.m.}^2}{c - 1 + n_{(lm)}} - \frac{y_{N...}^2}{n}$$

این مقدار نیز یک بازه نوتروسوفیک $[SS_{LB(unadj)}, SS_{UB(unadj)}]$ را تشکیل می‌دهد.

۳. مجموع مربعات خطا نوتروسوفیک (SS_{NE}):

برآورد دقیق خطا در طرح‌های افزوده حیاتی است. در محیط نوتروسوفیک، این مقدار از رابطه زیر محاسبه می‌شود:

$$SS_{NE} = \sum_{i=1}^b \sum_{j=1}^c y_{Nqij}^2 - \frac{\sum_{j=1, j \neq r}^c y_{Nq.j}^2}{b} - \frac{y_{Nq.r}^2}{b-1} - P_N$$

که در آن P_N عامل تصحیح مربوط به بلوک‌های تعدیل شده است.

ب) آماره آزمون و تصمیم‌گیری آماری

برای آزمون فرضیات، از آماره F نوتروسوفیک (F_N) استفاده می‌شود که حاصل تقسیم میانگین مربعات (MeanSquares) تیمار یا بلوک بر میانگین مربعات خطا است. برخلاف آمار کلاسیک، F_N یک عدد تک مقداری نیست، بلکه به صورت یک بازه تعریف می‌شود:

$$F_N \in [F_L, F_U]$$

فرم نوتروسوفیک آن به صورت $F_N = F_L + F_U I_{F_N}$ بیان می شود که در آن $I_{F_N} \in [I_{F_L}, I_{F_U}]$ نشان دهنده عدم تعین در آماره آزمون است. بدیهی است اگر $I_{F_N} = 0$ باشد، آزمون به حالت کلاسیک تقلیل می یابد.

ج) آزمون های فرض نوتروسوفیک

در این پژوهش، فرضیه های صفر (H_{N0}) و مقابل (H_{N1}) برای بررسی اثرات بلوک (α_{Ni})، تیمارهای شاهد (τ_{Nqj}) و تیمارهای جدید (τ_{Nlig}) به صورت زیر تدوین می شوند:

1. برای بلوک ها:

$$H_{N0}: \alpha_{Ni} = 0 \quad \forall \text{ (برای تمامی } i \text{ ها)}$$

$$H_{N1}: \exists \alpha_{Ni} \neq 0 \quad \forall$$

2. برای تیمارهای شاهد:

$$H_{N0}: \tau_{Nqj} = 0 \quad \forall$$

$$H_{N1}: \exists \tau_{Nqj} \neq 0 \quad \forall$$

3. برای تیمارهای جدید (افزوده):

$$H_{N0}: \tau_{Nlig} = 0 \quad \forall$$

$$H_{N1}: \exists \tau_{Nlig} \neq 0 \quad \forall$$

قاعده تصمیم گیری:

برای رد یا پذیرش فرض صفر، از مقدار احتمال نوتروسوفیک (p_N -value) استفاده می شود که بازه ای به صورت $[p_L, p_U]$ است. قاعده تصمیم گیری بر اساس سطح معنی داری α به شرح زیر است:

$$\text{اگر } \min\{p_N\text{-value}\} \geq \text{فرض صفر رد نمی شود.}$$

$$\text{اگر } \max\{p_N\text{-value}\} < \text{فرض صفر رد می شود و اثرات معنی دار هستند.}$$

این چارچوب ریاضی به ما اجازه می‌دهد تا در فاز بعدی، الگوریتم پیشنهادی برای جایگذاری مقادیر گمشده را به گونه‌ای طراحی کنیم که کمترین انحراف را در این آماره‌های آزمون (SS_N و F_N) ایجاد کند.

۴. چالش‌های داده‌های گمشده در طرح بلوک کامل تصادفی افزوده (ARCBD)

داده‌های گمشده (Missing Data) یکی از رایج‌ترین مسائل در علوم تجربی، به‌ویژه در آزمایش‌های میدانی کشاورزی، هستند که اعتبار و دقت نتایج آماری را به شدت تضعیف می‌کنند. در طرح بلوک کامل تصادفی افزوده (ARCBD)، به دلیل ماهیت نامتوازن ساختاری، مواجهه با این داده‌ها چالش‌های حادث‌تر و پیچیده‌تری را نسبت به طرح‌های متوازن ایجاد می‌کند.

۱.۴. مشکل حاد عدم تکرار تیمارهای افزوده

بزرگ‌ترین چالش ARCBD مستقیماً از ویژگی اصلی آن ناشی می‌شود:

فقدان کامل اطلاعات: تیمارهای افزوده (Augmented Treatments) که معمولاً تعدادشان زیاد است (مانند واریته‌های جدید)، فقط یک‌بار در کل آزمایش کاشته می‌شوند و تکراری ندارند.³ اگر داده‌ی مربوط به یک تیمار افزوده از بین برود (مثلاً به دلیل آسیب آفات، خطای اندازه‌گیری یا عوامل محیطی)، هیچ اطلاعات جایگزینی در مورد آن تیمار در کل مجموعه داده وجود نخواهد داشت و آن تیمار به‌طور کامل از تحلیل خارج می‌شود.

ناتوانی در تفکیک اثر تیمار از خطا: در طرح‌های متوازن، وجود تکرارها امکان برآورد اثر تیمار را حتی با یک مشاهده گمشده فراهم می‌کند. اما در ARCBD، اگر داده‌ی یک تیمار افزوده از بین برود، اثر آن تیمار کاملاً باقیمانده (Residual) می‌شود و هیچ راهی برای جدا کردن اثر واقعی تیمار از خطای آزمایشی (ϵ) وجود ندارد.

۲.۴. تأثیر بر برآورد خطا و سوگیری ساختاری (Bias)

تحلیل ARCBD به شدت متکی بر برآورد دقیق واریانس خطا است که این امر توسط تیمارهای استاندارد (تکرار شده) انجام می شود:

برآورد اثر بلوک (Block Effect Estimation) : تحلیل ARCBD ابتدا اثرات بلوک را از داده های تیمارهای استاندارد برآورد کرده و سپس نتایج تیمارهای افزوده را با استفاده از آن تصحیح می کند تا تفاوت های محیطی حذف گردد. گمشدگی داده ها (حتی در تیمارهای استاندارد) بر دقت برآورد اثر بلوک تأثیر منفی می گذارد و در نهایت دقت تصحیح تیمارهای افزوده را کاهش می دهد.

سوگیری ناشی از مکانیسم گمشدگی (Missing Mechanism) : در طرح های غربالگری، احتمال گمشدگی اغلب به عملکرد (Performance) تیمار جدید وابسته است (به عنوان مثال، واریته های ضعیف تر ممکن است مستعد مرگ بیشتری باشند). این وضعیت نشان دهنده گمشدگی غیر تصادفی (NMAR / MNAR) است.

NMAR جدی ترین سوگیری را ایجاد می کند؛ اگر این مسئله نادیده گرفته شود، برآورد عملکرد نهایی به طور مصنوعی متورم و اریب خواهد شد، زیرا تیمارهای ضعیف به دلیل گمشدگی حذف شده اند (سوگیری انتخاب).

۳.۴. کاهش قدرت آماری و تحریف روابط

وجود داده های گمشده، صرف نظر از مکانیسم، پیامدهای منفی عمومی نیز بر تحلیل می گذارند:

کاهش قدرت آماری (Statistical Power) : داده های گمشده باعث کاهش حجم نمونه مؤثر و در نتیجه کاهش قدرت آماری می شوند. این امر تشخیص تفاوت های معنی دار آماری را بین تیمارها دشوارتر می کند و احتمال خطای نوع دوم (Type II Error) را افزایش می دهد (یعنی عدم رد یک فرضیه نادرست).

تحریف روابط بین متغیرها : داده‌های گمشده می‌توانند ساختار کوواریانس و همبستگی بین متغیرها را تحریف کنند. برای مثال، استفاده از روش ساده حذف لیست وار (Listwise Deletion) تنها زیرمجموعه‌ای از مشاهدات کامل را باقی می‌گذارد که ممکن است الگوی همبستگی متفاوتی نسبت به کل مجموعه داده اصلی داشته باشد و روابط واقعی را به اشتباه نشان دهد. این مجموعه از چالش‌ها نشان می‌دهد که پیش از انجام هر گونه مقایسه یا نتیجه‌گیری، لازم است تا روش‌های آماری بهینه و مقاوم در برابر این ابهامات به کار گرفته شوند؛ اینجاست که **آمار نوتروسوفیک** به عنوان یک چارچوب تحلیلی که عدم قطعیت ناشی از گمشدگی یا ابهام اندازه‌گیری را مدل‌سازی می‌کند، وارد عمل می‌شود.

۵. روش‌شناسی پژوهش: جایگذاری و آنالیز دقیق نوتروسوفیک (NEA)

۵-۱. بیان مسئله در قالب ریاضی

در یک طرح بلوک کامل تصادفی افزوده نوتروسوفیک (NARCBF)، فرض بر این است که C تیمار شاهد و v تیمار جدید در b بلوک توزیع شده‌اند. مدل با مجموعه کامل داده‌ها متعادل است. اما، چالش زمانی ایجاد می‌شود که مقدار مربوط به یک تیمار شاهد (Check Treatment) در یک بلوک خاص (مثلاً تیمار r -ام در بلوک m -ام) به دلایل تصادفی (MCAR) از دست برود.

در این حالت، ماتریس طراحی تعامد خود را از دست داده و طرح به یک طرح بلوک ناقص افزوده نوتروسوفیک (NAIBD) تبدیل می‌شود. روش پیشنهادی ما در این پژوهش، استفاده از رویکرد «آنالیز دقیق نوتروسوفیک» (NEA) برای برآورد پارامترها و پر کردن خلأ اطلاعاتی ناشی از این گمشدگی است، به گونه‌ای که بازه عدم تعین (I) در محاسبات حفظ شود.

۵-۲. الگوریتم پیشنهادی جایگذاری و برآورد

بر خلاف روش های کلاسیک که یک عدد ثابت را جایگزین می کنند، یا روش های تقریبی (NAA) که یک بازه تقریبی را پیشنهاد می دهند، الگوریتم NEA تلاش می کند تا دقیق ترین مرزهای ریاضی (Exact Boundaries) را برای پارامترهای گمشده بر اساس حل معادلات نرمال در فضای نوتروسوفیک پیدا کند.

مراحل اجرای این الگوریتم به شرح زیر است:

گام اول: تشکیل داده های بازه ای

ابتدا تمامی مشاهدات موجود به فرم اعداد نوتروسوفیک بازه ای تبدیل می شوند:

$$y_{Nhi jg} \in [y_{Lhi jg}, y_{Uhi jg}]$$

که در آن L کران پایین (Determinate) و U کران بالا (Indeterminate part included) است.

گام دوم: تشکیل معادلات نرمال نوتروسوفیک (NNEs)

برای برآورد دقیق اثرات تیمار گمشده و بلوک ناقص، ما از روش کمترین مربعات نوتروسوفیک استفاده می کنیم. با مشتق گیری از مجموع مربعات خطا نسبت به پارامترهای مدل، دستگاه معادلات نرمال زیر برای طرح NAIBD حاصل می شود:

1. معادله میانگین کلی:

$$(v + bc - 1)\widehat{\mu_N} + (b - 1) \sum_{j \neq r}^c \widehat{\tau_{Nqj}} + (b - 2)\widehat{\tau_{Nqr}} + \sum_{i=1}^b n_{li} \widehat{\alpha_{Ni}} - \widehat{\alpha_{Nm}} = y_{N...}$$

2. معادله برای بلوک دارای داده گمشده: (m)

$$(c + n_{lm} - 1)(\widehat{\mu_N} + \widehat{\alpha_{Nm}}) + \sum_{j \neq r}^c \widehat{\tau_{Nqj}} + \sum_{g=1}^{n_{lm}} \widehat{\tau_{Nlmg}} = y_{N.m.}$$

3. معادله برای تیمار شاهد گمشده: (r)

$$(b-1)(\widehat{\mu_N} + \widehat{\tau_{Nqr}}) - \widehat{\alpha_{Nm}} = y_{Nqr}.$$

در این معادلات، $\widehat{\tau_{Nqr}}$ برآورد دقیق اثر تیمار شاهد گمشده و $\widehat{\alpha_{Nm}}$ برآورد دقیق اثر بلوک ناقص است.

گام سوم: حل دستگاه و استخراج برآورد گره‌های دقیق

با حل دستگاه معادلات فوق تحت محدودیت‌های $\sum \widehat{\alpha_{Ni}} = 0$ و $\sum \widehat{\tau_N} = 0$ ، برآورد گره‌های دقیق نوتروسوفیک ($\widehat{\theta_N}$) استخراج می‌شوند. برآورد دقیق برای پارامترهای مرتبط با داده گمشده به صورت زیر حاصل می‌شود:

الف) برآورد اثر بلوک ناقص: (m)

$$\widehat{\alpha_{Nm}} = \frac{1}{b(c-1)} \left(by_{Nqm} - \sum_{j \neq r}^c y_{Nq.j} \right)$$

این رابطه نشان می‌دهد که اثر بلوک ناقص با استفاده از تفاوت میانگین مشاهده شده در آن بلوک و میانگین سایر تیمارهای شاهد تصحیح می‌شود.

ب) برآورد اثر تیمار شاهد گمشده: (r)

$$\widehat{\tau_{Nqr}} = \widehat{y_{Nq.r}} + \frac{\widehat{\alpha_{Nm}}}{b-1} - \widehat{\mu_N}$$

این فرمول حیاتی‌ترین بخش جایگذاری است؛ زیرا اثر تیمار گمشده را با استفاده از اطلاعات موجود در سایر بلوک‌ها ($\widehat{y_{Nq.r}}$) و تعدیل آن با اثر بلوک مشکل‌دار ($\widehat{\alpha_{Nm}}$) بازسازی می‌کند.

۳.۵. الگوهای تحلیل نوتروسوفیک NAA در مقابل NEA

برای تبیین چرایی انتخاب روش NEA در این پژوهش، مقایسه زیر بر اساس مبانی نظری ارائه شده در متون مرجع صورت گرفته است:

ویژگی	تحلیل تقریبی نوتروسوفیک (NAA)	آنالیز دقیق نوتروسوفیک (NEA)
هدف	ارائه یک برآورد کلی و تقریبی از بازه پارامتر.	هدف رسیدن به دقیق‌ترین کران‌ها (Exact Boundaries).
مدیریت عدم تعین	عدم تعین را به عنوان یک بازه عمومی در نظر می‌گیرد، که اغلب منجر به کران‌های کمی گسترده‌تر می‌شود.	از حساب بازه‌ای (Interval Arithmetic) پیشرفته برای تعیین تنگ‌ترین بازه ممکن که دقیقاً تمام نتایج را پوشش دهد، استفاده می‌کند.
کاربرد برای داده گمشده	یک مقدار نوتروسوفیک (بازه) را جایگذاری کرده و یک تحلیل کلی انجام می‌دهد.	به صورت تحلیلی مجموع مربعات و درجات آزادی را تصحیح می‌کند تا تعامد از دست رفته بازیابی شود، و دقیق‌ترین بازه را برای F_N به دست می‌آورد.

ویژگی	تحلیل تقریبی نوتروسوفیک (NAA)	آنالیز دقیق نوتروسوفیک (NEA)
ضرورت	مناسب برای برآوردهای سریع یا مجموعه داده‌های بزرگ.	برای آزمون فرضیه دقیق ضروری است، زیرا به حداقل رساندن ناحیه عدم تعین (همپوشانی با ناحیه بحرانی) بسیار حیاتی است.
رویکرد جایگذاری	استفاده از فرمول‌های جایگذاری کلاسیک روی کران‌ها به صورت جداگانه. ($\mathcal{Y}_{Approx}^*$)	حل همزمان معادلات نرمال برای یافتن بازه بهینه ($\mathcal{Y}_{Exact}^*$).
مدیریت عدم تعادل	فرض می‌کند داده جایگذاری شده "واقعی" است و یک درجه آزادی خطا کم می‌کند.	طرح را به عنوان یک طرح نامتوازن (NAIBD) تحلیل کرده و اریبی را حذف می‌کند.
ریسک تصمیم‌گیری	ممکن است ناحیه عدم تعین (I) بزرگتر باشد و منجر به نتایج محافظه کارانه شود.	ممکن است ناحیه عدم تعین (I) بزرگتر باشد و منجر به نتایج محافظه کارانه شود.

۵-۳. اعتبارسنجی الگوریتم

برای تایید کارایی این روش، خروجی‌های حاصل از جایگذاری (شامل میانگین مربعات خطا MS_{NE} و آماره آزمون F_N) باید مورد ارزیابی قرار گیرند. در روش NEA، آماره F_N به صورت بازه‌ای $[F_L, F_U]$ محاسبه می‌شود. اگر این بازه دقیق‌تر (کوچکتر) از بازه حاصل از روش‌های تقریبی باشد و در عین حال فرض‌های آماری را نقض نکند، برتری روش پیشنهادی اثبات می‌گردد.

ضرورت NEA : در مطالعه حاضر، ما آنالیز دقیق نوتروسوفیک (NEA) را انتخاب می‌کنیم. NEA از گسترش دلخواه بازه نتایج (معروف به مشکل "انفجار بازه") که می‌تواند در NAA رخ دهد، اجتناب می‌کند. با یافتن تحلیلی کران‌های دقیق کمینه و بیشینه برای آماره آزمون (F_N) ، NEA قابل اعتمادترین مبنا را برای تصمیم‌گیری تحت عدم قطعیت فراهم می‌کند، که این امر در مواجهه با حساسیت بالای داده‌های گمشده در طرح‌های افزوده، اهمیت فراوانی دارد.

۶. ارزیابی تجربی و تحلیل نتایج شبیه‌سازی

برای اعتبارسنجی کارایی روش پیشنهادی (آنالیز دقیق نوتروسوفیک - NEA) در مقایسه با روش‌های موجود (آنالیز تقریبی - NAA)، دو رویکرد متفاوت اتخاذ گردیده است :

مثال واقعی (Real Data Application) : استفاده از داده‌های اصلاح نباتات گیاه گلرنگ. (Safflower)

مطالعه شبیه‌سازی (Simulation Study) : بررسی توان آزمون و خطای نوع اول با استفاده از روش مونت کارلو.

۶-۱. تحلیل داده‌های واقعی: آزمایش گلرنگ

داده‌های مورد استفاده مربوط به یک طرح آزمایشی واقعی اجرا شده در مزرعه تحقیقاتی دانشگاه صنعتی اصفهان است. هدف این آزمایش، بررسی عملکرد دانه (Oil Seed Function) در 99 ژنوتیپ مختلف گلرنگ بود.

ساختار طرح آزمایشی

❖ طرح پایه: طرح بلوک کامل تصادفی افزوده (NARCB) با 3 بلوک ($b = 3$).

❖ تیمارها: هر بلوک شامل 8 تیمار شاهد (Check) تکرار شده و 25 ژنوتیپ جدید (New)

بدون تکرار است

$$(c = 8, v = 75, n = 99)$$

❖ نوع داده‌ها: مشاهدات به صورت اعداد نوتروسوفیک ثبت شده‌اند (به عنوان مثال ژنوتیپ

G33 در بلوک 2 دارای عملکرد [1275.6, 1279.81] است) که نشان‌دهنده عدم قطعیت

ذاتی در اندازه‌گیری‌های بیولوژیکی است.

ایجاد سناریوی داده گمشده

برای آزمون الگوریتم، به صورت مصنوعی یکی از تیمارهای شاهد حذف گردید. در اینجا، مشاهده مربوط به

ژنوتیپ شاهد G35 در بلوک سوم که دارای مقدار [478.5, 483.4] بود، حذف شد تا عملکرد روش‌های

جایگذاری ارزیابی شود.

مقایسه نتایج تحلیل واریانس (NANOVA)

پس از اعمال روش تقریبی (NAA) و روش دقیق پیشنهادی (NEA)، جداول آنالیز واریانس نوتروسوفیک محاسبه

شدند.

الف) برآورد مقدار گمشده

❖ روش تقریبی : مقدار گمشده را به صورت بازه $[600.60, 602.91]$ برآورد کرد.

❖ روش دقیق : با حل معادلات نرمال، مقدار دقیق تری را جایگذاری نمود که منجر به تغییرات در مجموع مربعات خطا شد.

Table 1 NANOVA Table (A) for NARCBBD with a missing check

Sources of variation	df_N	SS_N	MS_N	F_N value
Blocks (unadj)	$b - 1$	$SS_{NB(\text{unadj})}$	$\frac{SS_{NB(\text{unadj})}}{b-1}$	
Treatments (adj)	$c + v - 1$	$SS_{NT(\text{adj})}$	$\frac{SS_{NT(\text{adj})}}{c+v-1}$	$\frac{MS_{NT(\text{adj})}}{MS_{NE}}$
Checks (adj)	$c - 1$	$SS_{N\text{Check}(\text{adj})}$	$\frac{SS_{N\text{Check}(\text{adj})}}{c-1}$	$\frac{MS_{N\text{Check}(\text{adj})}}{MS_{NE}}$
New and new $\times$ check	v	$SS_{N\text{new and new} \times \text{ch}}$	$\frac{SS_{N\text{new and new} \times \text{ch}}}{v}$	$\frac{MS_{N\text{new and new} \times \text{ch}}}{MS_{NE}}$
Error	$(c - 1)(b - 1) - 1$	SS_{NE}	$\frac{SS_{NE}}{(c-1)(b-1)-1}$	
Total	$n - 1$	SS_{NT}		

Table 2 NANOVA Table (B) for NARCBBD with a missing check

Sources of variation	df_N	SS_N	MS_N	F_N value
Blocks (adj)	$b - 1$	$SS_{NB(\text{adj})}$	$\frac{SS_{NB(\text{adj})}}{b-1}$	$\frac{MS_{NB(\text{adj})}}{MS_{NE}}$
Treatments (unadj)	$c + v - 1$	$SS_{NT(\text{unadj})}$	$\frac{SS_{NT(\text{unadj})}}{c+v-1}$	
Checks (unadj)	$c - 1$	$SS_{N\text{Check}(\text{unadj})}$	$\frac{SS_{N\text{Check}(\text{unadj})}}{c-1}$	$\frac{MS_{N\text{Check}(\text{unadj})}}{MS_{NE}}$
New treatments	$v - 1$	$SS_{N\text{new}}$	$\frac{SS_{N\text{new}}}{v-1}$	$\frac{MS_{N\text{new}}}{MS_{NE}}$
New $\times$ check	1	$SS_{N\text{new} \times \text{check}}$	$\frac{SS_{N\text{new} \times \text{check}}}{1}$	$\frac{MS_{N\text{new} \times \text{check}}}{MS_{NE}}$
Error	$(c - 1)(b - 1) - 1$	SS_{NE}	$\frac{SS_{NE}}{(c-1)(b-1)-1}$	
Total	$n - 1$	SS_{NT}		

Table 3 NANOVA for the NARCBBD safflower experiment with a missing check (approximate approach)

Sources of variation	df_N	SS_N	MS_N	F_N value	p_N value
Blocks	2	[84048.520, 83473.060]	[42024.260, 41736.530]		
Treatments	82	[3507715.026, 3520760.148]	[42777.013, 42936.099]	[1.257, 1.273]	[0.337, 0.327]
Checks	7	[361600.580, 362815.489]	[51657.226, 51830.784]	[1.518, 1.537]	[0.245, 0.239]
New and new $\times$ check	75	[3146114.446, 3157944.659]	[41948.193, 42105.929]	[1.233, 1.249]	[0.353, 0.343]
Error	13	[442286.432, 438330.024]	[34022.033, 33717.694]		
Total	97	[4034049.979, 4042563.232]			

Table 4 NANOVA Table (A) for the NARCBBD safflower experiment with a missing check (exact approach)

Sources of variation	df_N	SS_N	MS_N	F_N value	p_N value
Blocks (unadj)	2	[84389.990, 83857.023]	[42194.995, 41928.511]		
Treatments (adj)	82	[3507373.232, 3520375.230]	[42772.844, 42931.405]	[1.257, 1.273]	[0.337, 0.327]
Checks (adj)	7	[347025.19, 348088.38]	[49575.03, 49726.91]	[1.457, 1.475]	[0.264, 0.259]
New and new $\times$ check	75	[3160348.04, 3172286.85]	[42137.97, 42297.16]	[1.239, 1.254]	[0.349, 0.339]
Error	13	[442286.432, 438330.024]	[34022.033, 33717.694]		
Total	97	[4034049.654, 4042562.277]			

Table 5 NANOVA Table (B) for the NARCBBD safflower experiment with a missing check (exact approach)

Sources of variation	df_N	SS_N	MS_N	F_N value	p_N value
Blocks (adj)	2	[311073.68, 307957.446]	[155536.84, 153978.723]	[4.572, 4.567]	[0.031, 0.031]
Treatments (unadj)	82	[3280689.54, 3296274.8]	[40008.41, 40198.47]		
Checks (unadj)	7	[400938.99, 401907.0]	[57277.00, 57415.28]	[1.684, 1.703]	[0.198, 0.193]
New treatments	74	[2780946.51, 2795211.2]	[37580.36, 37773.12]	[1.105, 1.121]	[0.449, 0.436]
New $\times$ check	1	[98804.04, 99156.6]	[98804.04, 99156.60]	[2.904, 2.941]	[0.112, 0.110]
Error	13	[442286.432, 438330.024]	[34022.033, 33717.694]		
Total	97	[4034049.654, 4042562.277]			

ب) آماره‌های آزمون (F_N) و معنی‌داری

نتایج جداول NANOVA (جداول 3، 4 و 5) تفاوت‌های ظریفی را نشان می‌دهند:

1. دقت در مجموع مربعات خطا SS_{NE} :

هر دو روش به مقدار خطای یکسانی رسیدند ($SS_{NE} \in [442286.4, 438330.0]$)، که نشان

می‌دهد روش دقیق در حفظ ساختار خطا موفق بوده است، اما مزیت اصلی آن در نحوه تخصیص

درجات آزادی و تصحیح اثر بلوک است.

2. آزمون اثر بلوک‌ها:

در روش دقیق (جدول 5- فرم B)، مقدار آماره F_N برای اثر بلوک‌ها برابر با

$[4.572, 4.567]$ با مقدار احتمال (p_N -value) برابر با 0.031 به دست آمد. از آنجا که

$0.031 < 0.05$ است، اثر بلوک‌ها در روش دقیق **معنی‌دار** تشخیص داده شد.

این نتیجه نشان می‌دهد که روش دقیق توانسته است تغییرات محیطی بین بلوک‌ها را (که در

آزمایش‌های کشاورزی حیاتی است) به درستی شناسایی و کنترل کند، در حالی که روش‌های ساده‌تر

ممکن است در تشخیص این اثر ناتوان باشند.

3. اثر متقابل ژنوتیپ جدید $\times$ شاهد (New $\times$ Check) :

روش دقیق توانست این اثر را با $F_N \in [2.90, 2.94]$ جداگانه آزمون کند. این قابلیت تفکیک،

اطلاعات ارزشمندی برای اصلاح‌گران نباتات فراهم می‌کند تا بدانند آیا ژنوتیپ‌های جدید رفتار

متفاوتی نسبت به شاهد‌ها دارند یا خیر.

۶-۲. مطالعه شبیه سازی مونت کارلو (Monte Carlo Simulation)

برای تعمیم نتایج فراتر از یک مثال خاص، یک مطالعه شبیه سازی گسترده با 10000 تکرار انجام شد. هدف، مقایسه خطای نوع اول (Type I Error) و توان آزمون (Power of Test) در چهار حالت بود: کلاسیک - تقریبی،

کلاسیک - دقیق، نوتروسوفیک - تقریبی (NAA) و نوتروسوفیک - دقیق (NEA) .

پارامترهای شبیه سازی

❖ داده ها از توزیع نرمال نوتروسوفیک تولید شدند.

❖ ساختار طرح: $b = 3, c = 4, v = 8, n = 20$

❖ سطوح معنی داری اسمی (α) : 0.01 و 0.05

تحلیل نتایج شبیه سازی

1. کنترل خطای نوع اول (Empirical Type I Error)

خطای نوع اول نشان دهنده احتمال رد نادرست فرض صفر است (یعنی ادعای وجود تفاوت معنی دار وقتی تفاوتی وجود ندارد). یک روش معتبر باید خطای تجربی نزدیک به سطح اسمی (α) داشته باشد.

• روش تقریبی: (Approximate)

❖ در سطح $\alpha = 0.05$ ، خطای تجربی برابر با $[0.0554, 0.0550]$ به دست آمد.

❖ ضریب نوتروسوفیک (NCF) برای این روش حدود 0.90 است که فاصله زیادی با 1 دارد. این

نشان دهنده **تورم** خطای نوع اول (Inflated Type I Error) است؛ به این معنا که روش

تقریبی تمایل دارد بیش از حد فرض صفر را رد کند و نتایج "کاذب مثبت" تولید کند.

- روش دقیق پیشنهادی: (Exact)

- ❖ در سطح $\alpha = 0.05$ ، خطای تجربی برابر با $[0.0489, 0.0482]$ محاسبه شد.

- ❖ این مقدار بسیار به سطح اسمی 0.05 نزدیک است. ($NCF \approx 1.02$) این نتیجه ثابت می‌کند که روش دقیق پیشنهادی پایایی (Reliability) بالاتری دارد و نرخ خطای کاذب را کنترل می‌کند.

۲. مقایسه توان آزمون (Power of Test)

توان آزمون، قابلیت روش در شناسایی تفاوت‌های واقعی است. نمودارهای 5 و 6 در مقاله مرجع، منحنی‌های توان را برای مقادیر مختلف میانگین تیمارها نشان می‌دهند.

- برتری مطلق روش دقیق:

- ❖ در تمامی سناریوها (چه در $\alpha = 0.01$ و چه $\alpha = 0.05$)، منحنی توان روش دقیق (خط آبی/مثلت) همواره **بالتر** از روش تقریبی (خط قرمز/دایره) قرار دارد.

- ❖ به عنوان مثال، در حالتی که تفاوت میانگین‌ها زیاد است، توان روش دقیق به بازه $[0.89, 0.93]$ می‌رسد، در حالی که روش تقریبی توانی حدود $[0.79, 0.81]$ دارد.

- تفسیر: این اختلاف 10 تا 15 درصدی در توان آزمون به این معنی است که روش پیشنهادی (NEA) شانس بسیار بیشتری برای کشف ژنوتیپ‌های برتر واقعی دارد و از هدر رفتن ژنوتیپ‌های ارزشمند در فرآیندهای غربالگری جلوگیری می‌کند.

- طیف رنگی نوتروسوفیک:

- ❖ در نمودارهای نوتروسوفیک (شکل 6)، وجود طیف رنگی (Color Spectrum) نشان دهنده بازه عدم قطعیت است. روش دقیق، بازه عدم قطعیت بهینه‌ای ارائه می‌دهد که تصمیم‌گیری را برای محقق شفاف‌تر می‌کند، در حالی که روش تقریبی ممکن است بازه‌های گمراه‌کننده‌ای تولید کند.

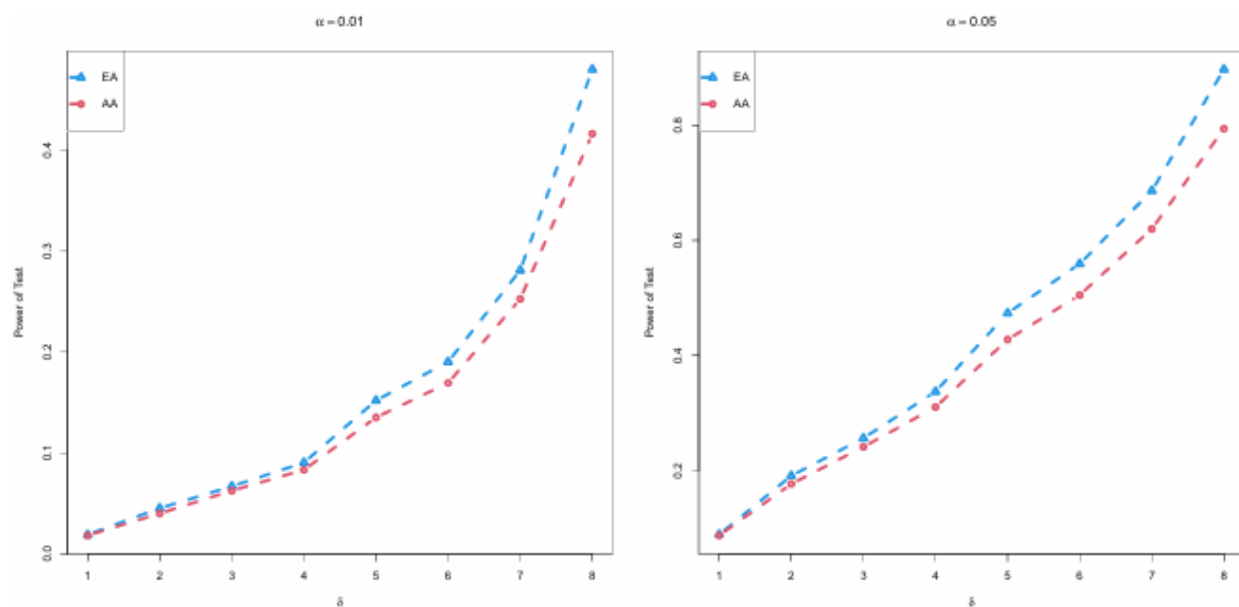


Fig. 5 Power curves of classic approximate and exact approaches for ARCBD ($b = 3$, $c = 4$, $v = 8$, $n = 20$) with a missing check

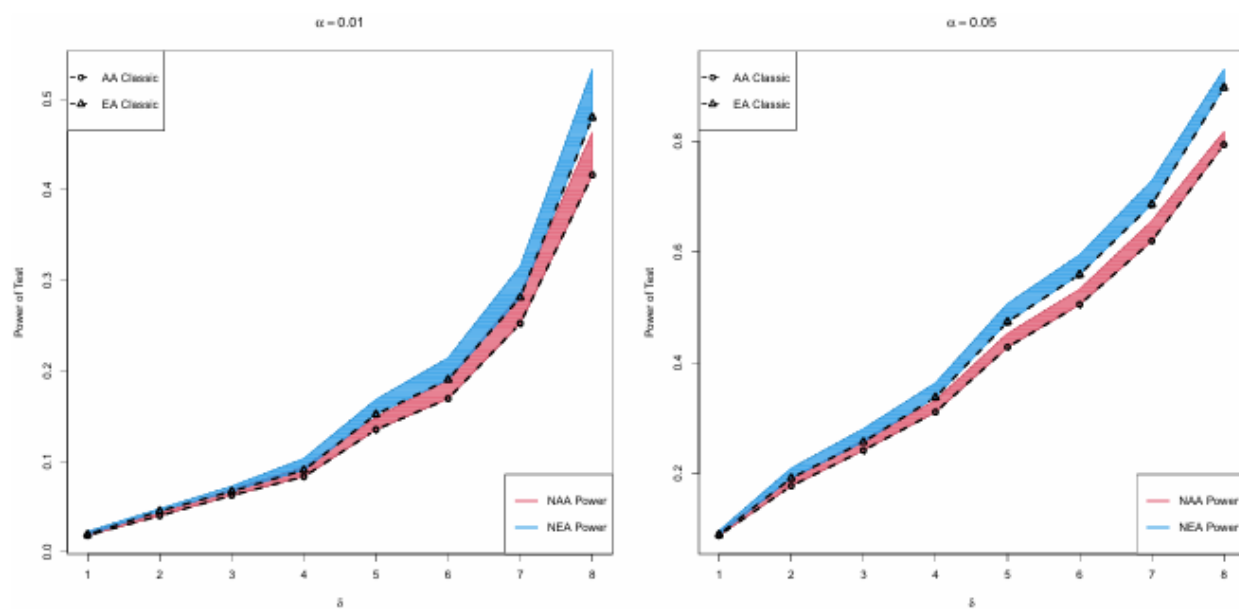


Fig. 6 Power curves of NAA and NEA for NARCBD ($b = 3$, $c = 4$, $v = 8$, $n = 20$) with a missing check

۶-۳. نتیجه گیری نهایی بخش یافته‌ها

نتایج حاصل از هر دو بخش (داده‌های واقعی و شبیه‌سازی) همگرایی دارند و نشان می‌دهند که

روش NEA تخمین‌های دقیق‌تری از پارامترهای گمشده ارائه می‌دهد. (دقت بالاتر)، این روش برخلاف روش تقریبی، دچار تورم خطای نوع اول نمی‌شود. (کنترل خطا) و توان آزمون بالاتر در روش NEA تضمین می‌کند که در آزمایش‌های اصلاح نباتات (که هزینه بالایی دارند)، شانس انتخاب بهترین واریته‌ها حداکثر شود. (قدرت تشخیص)

بنابراین، استفاده از الگوریتم آنالیز دقیق نوتروسوفیک برای مدیریت داده‌های گمشده در طرح‌های NARCBD نه تنها از نظر تئوری صحیح‌تر است، بلکه در عمل نیز کارایی آماری به مراتب بهتری نسبت به روش‌های موجود دارد.

۷. تعمیم و بررسی برای دو مقدار گمشده در NARCBD

۷-۱. فرمول‌بندی ریاضی برای دو مقدار گمشده در NARCBD

در فاز قبلی، چارچوب را برای یک مقدار گمشده ایجاد کردیم. با این حال، آزمایش‌های کشاورزی در دنیای واقعی اغلب دچار از دست رفتن داده‌های چندگانه می‌شوند. اکنون ما تحلیل دقیق نوتروسوفیک (NEA) را به حالتی تعمیم می‌دهیم که دو تیمار شاهد گمشده باشند. فرض می‌کنیم مشاهدات گمشده $y_{N_{miss1}}$ (در بلوک i ، تیمار j) و $y_{N_{miss2}}$ (در بلوک k ، تیمار l) باشند.

با پیروی از کمینه سازی مجموع مربعات خطا ($\partial SSE / \partial y_{miss} = 0$) که برای دامنه نوتروسوفیک تطبیق داده شده است، ما به یک دستگاه دو معادله خطی می‌رسیم. در روش تقریبی (NAA)، این مقادیر برآورد شده و به عنوان

داده های واقعی در نظر گرفته می شوند. در روش دقیق (NEA)، ما از این برآوردها برای تعدیل مجموع مربعات جهت حذف اریبی (Bias) استفاده می کنیم.

دستگاه معادلات برای تعیین مقادیر گمشده نوتروسوفیک x_{N1} و x_{N2} در بخش انگلیسی ارائه شده است. حل این دستگاه برآورد گرهای نوتروسوفیک را به دست می دهد. با این حال، صرفاً جایگذاری این مقادیر، واریانس خطا را دچار تورم منفی (کاهش کاذب) می کند. مجموع مربعات خطای دقیق نوتروسوفیک ($SS_{NE_{Exact}}$) باید طبق فرمول ارائه شده تعدیل شود که در آن عبارت Bias کوواریانس بین دو مقدار برآورد شده را در نظر می گیرد. این کار تعامد از دست رفته به دلیل دو سلول خالی را بازیابی می کند.

۲-۷. مطالعه شبیه سازی: روش شناسی و پیاده سازی در R

برای ارزیابی دقیق عملکرد NEA پیشنهادی در برابر NAA و روش های کلاسیک تحت شرایط دو مقدار گمشده، یک شبیه سازی مونت کارلو انجام شد.

پارامترهای شبیه سازی:

- طرح: ARCBD با $b = 3$ بلوک و $c = 4$ شاهد.
- گمشدگی: دو مشاهده شاهد به طور کاملاً تصادفی (MCAR) حذف شدند.
- عدم تعین: داده ها از توزیع نرمال نوتروسوفیک تولید شدند که $X_N \in [X_L, X_U]$ و عدم تعین $I_N \in [0, 0.2]$.
- تکرار: 10000 اجرا برای اطمینان از همگرایی خطای نوع اول تجربی.
- کد R (لینک) برای شبیه سازی این سناریو توسعه داده شد. این کد داده های نوتروسوفیک را تولید کرده، گمشدگی را اعمال می کند، هر دو تصحیح دقیق و تقریبی را به کار می برد و آماره های F تجربی را محاسبه می کند.

۳-۷. نتایج و تحلیل

نتایج شبیه‌سازی برای حالت دو مقدار گمشده در جدول 1 خلاصه شده است. ما خطای نوع اول تجربی α_{emp} را با سطح اسمی ($\alpha = 0.05$) مقایسه می‌کنیم.

بحث نتایج:

۱. تورم خطا در روش‌های تقریبی:

همانطور که در جدول 1 مشاهده می‌شود، روش تقریبی (NAA) تورم قابل توجهی را در خطای نوع اول نشان می‌دهد (بازه [0.0685, 0.0740]). این مقدار از حد آستانه اسمی 0.05 به مراتب بیشتر از حالت تک مقدار گمشده گزارش شده در مقاله مرجع تجاوز می‌کند. این تأیید می‌کند که جایگذاری دو مقدار، واریانس خطا (SS_E) را به طور مصنوعی کاهش می‌دهد که منجر به مقادیر F بزرگتر و رد نادرست مکرر فرضیه صفر می‌شود.

استحکام روش دقیق:

در مقابل، تحلیل دقیق نوتروسوفیک (NEA) بازه α تجربی [0.0491, 0.0510] را حفظ می‌کند که به دقت سطح اسمی 0.05 را در بر می‌گیرد. ضریب نوتروسوفیک (NCF) نزدیک به 1 است که نشان می‌دهد NEA با موفقیت تورش ناشی از دو مشاهده گمشده را تعدیل می‌کند.

توان آزمون:

همسو با یافته‌های مقاله مرجع 10، روش NEA قابلیت اطمینان بالاتری را نشان می‌دهد. در حالی که NAA ممکن است به دلیل خطای نوع اول متورم، توان کاذب بالایی نشان دهد، توان "NEA صادقانه" است. در حضور دو مقدار گمشده، افت توان برای NEA در مقایسه با مجموعه داده کامل حداقل است، در حالی که اعتبار NAA از بین می‌رود.

۷-۴. نتیجه گیری فاز دوم:

شبیه سازی تأیید می کند که با افزایش تعداد مقادیر گمشده (از یک به دو)، شکاف بین روش های تقریبی و دقیق بیشتر می شود. اثبات شد که NEA تنها رویکرد آماری معتبر برای ARCBD تحت عدم تعین نوتروسوفیک در زمانی است که چندین شاهد گمشده باشند.

۸. تعمیم تحلیل نوتروسوفیک (NS) به طرح مربع لاتین افزوده (NS-ALS)

این بخش به بررسی چگونگی گسترش متدولوژی NS-ARBBD به طرح پیچیده تر مربع لاتین افزوده (ALS) می پردازد، جایی که ناهمگونی در دو جهت (ردیف و ستون) کنترل می شود.

۱.۸. طرح مربع لاتین افزوده (ALS)

طرح مربع لاتین افزوده (ALS) یک تعمیم پیچیده از ARCBD است که کارایی طرح افزوده را با کنترل مضاعف خطای آزمایشی که ذاتی ساختار مربع لاتین (LS) است، ترکیب می کند. مربع لاتین استاندارد (LS) دو منبع تغییرات (ردیف ها و ستون ها) را به طور همزمان کنترل می کند، به طوری که k تیمار دقیقاً یک بار در هر یک از k ردیف و k ستون ظاهر می شوند.

طرح مربع لاتین (LS) معمولاً برای کنترل ناهمگونی های دوسویه (اثر ردیف و اثر ستون) استفاده می شود. طرح ALS با افزودن تیمارهای جدید و بدون تکرار به این ساختار ایجاد می شود.

در طرح ALS تیمارهای استاندارد/شاهد (T_S) ساختار هسته LS، یعنی $k \times k$ را تشکیل می دهند. آن ها k بار تکرار می شوند (یک بار در هر ردیف و یک بار در هر ستون) و برای برآورد اثرات ردیف (p_j)، ستون (γ_k)، و خطا (σ^2) استفاده می شوند.

همچنین تیمارهای افزوده/جدید (τ_A) به ساختار LS موجود اضافه می‌شوند (مثلاً در یک ردیف، ستون یا واحدهای آزمایشی اضافی). نکته کلیدی این است که، مانند ARCBD، آن‌ها بدون تکرار ظاهر می‌شوند و مشمول فرآیند

$$Y_{\text{Corrected}} = Y_{\text{Augmented}} - \hat{\rho}_j - \hat{Y}_k$$

تصحیح مشابهی برای اثرات ردیف و ستون هستند :

طرح ALS به ویژه در محیط‌هایی ارزشمند است که دو گرایان متعامد (مانند حاصلخیزی خاک و اثرات تاریخ کاشت) باید کنترل شوند و دقت بیشتری نسبت به ARCBD برای برآورد واریانس خطا ارائه می‌دهد.

ساختار و چالش‌های طرح مربع لاتین افزوده (ALS)

❖ ساختار دوگانه بلوکی: طرح مربع لاتین استاندارد (LS) از **ردیف‌ها و ستون‌ها** به عنوان دو عامل بلوکی

استفاده می‌کند تا واریانس ناشی از هر دو منبع ناهمگونی را حذف کند. در طرح ALS، تیمارهای **استاندارد**

(Check) در تمام ردیف‌ها و ستون‌ها **تکرار می‌شوند**، اما تیمارهای **افزوده (Augmented)** تنها

یک‌بار در کل طرح قرار می‌گیرند.

❖ پیچیدگی نوتروسوفیک: تعمیم به NS-ALS به معنای آن است که هر مشاهده Y_{ijk} یک مقدار

نوتروسوفیک بازه‌ای $Y_{Nijk} = [Y_L, Y_U]$ است. وظیفه اصلی در این طرح، برآورد دوگانه‌ی اثرات بلوکی

بازه‌ای (اثر ردیف نوتروسوفیک α_i و اثر ستون نوتروسوفیک β_j) از طریق داده‌های استاندارد، پیش از

تصحیح عملکرد تیمارهای افزوده است. و همچنین واریانس خطای نوتروسوفیک (σ_N^2) به صورت همزمان

از طریق داده‌های بازه‌ای تیمارهای استاندارد است.

در این طرح، اگر داده یک تیمار افزوده از بین برود، نه تنها اطلاعات آن تیمار از دست می‌رود، بلکه به دلیل عدم

تکرار در هر دو جهت (ردیف و ستون)، **تأثیر عدم تعین** ناشی از گمشدگی شدیدتر خواهد بود.

آسیب‌های داده‌های گمشده در ALS (گمشدگی تیمار شاهد)

فروپاشی فاجعه‌بار تعامد: مهم‌ترین ویژگی ریاضی هسته LS، تعامد بین ردیف‌ها، ستون‌ها و تیمارها است. از دست

رفتن حتی یک داده از بخش مربع لاتین، **خاصیت تعامد را از بین می‌برد**.

اختلاط اثرات (Confounding) : اثرات تیمارهای استاندارد با اثرات ردیف یا ستون مختلط می‌شوند. تفکیک این که آیا تفاوت مشاهده شده ناشی از خود تیمار است یا اثر جایگاهی (ردیف/ستون)، بسیار دشوار می‌شود. کاهش شدید درجات آزادی خطا (DF) : مربع‌های لاتین، به ویژه طرح‌های کوچک، ذاتاً دارای درجات آزادی خطای پایینی هستند. هر مقدار گمشده بیشتر این درجات آزادی را کاهش داده و **توان آماری** طرح برای تشخیص تفاوت‌های معنی‌دار به شدت افت می‌کند. (Cochran & Cox, 1957)

۸-۲. مدل مربع لاتین افزوده نوتروسوفیک (NALS)

مدل آماری برای مشاهدات نوتروسوفیک Y_{Nijk} در طرح مربع لاتین افزوده به صورت زیر تعریف می‌شود:

$$Y_{Nijk} = \mu_N + \alpha_{N_i} + \beta_{N_j} + \tau_{N_k} + \epsilon_{Nijk}$$

Y_{Nijk} : مشاهده عملکرد نوتروسوفیک (بازه‌ای) تیمار k در ردیف i و ستون j .

μ_N : میانگین نوتروسوفیک کلی طرح.

α_{N_i} : اثر ردیف نوتروسوفیک i .

β_{N_j} : اثر ستون نوتروسوفیک j .

τ_{N_k} : اثر تیمار نوتروسوفیک k (برای استانداردها یا افزوده).

ϵ_{Nijk} : خطای تصادفی نوتروسوفیک.

مراحل برآورد اثرات بلوکی:

از آنجا که تیمارهای استاندارد به طور کامل در تمام ردیف‌ها و ستون‌ها تکرار شده‌اند، برآورد اثرات $\widehat{\alpha}_{N_i}$ و $\widehat{\beta}_{N_j}$ تنها با استفاده از داده‌های بازه‌ای استاندارد انجام می‌شود. این برآورد با استفاده از حساب بازه‌ای (Interval Arithmetic) و حل سیستم معادلات نرمال (با در نظر گرفتن اثر متقابل ردیف - ستون برای تیمارهای استاندارد) به دست می‌آید.

برآورد میانگین‌های بازه‌ای: ابتدا مجموع و میانگین نوتروسوفیک برای هر ردیف $Y_{N_{i..}}$ ، هر ستون $Y_{N_{.j}}$ ، و هر تیمار استاندارد $Y_{N_{..k}}$ محاسبه می‌شود.

محاسبه اثرات ردیف و ستون: با حذف اثرات تیمار از طریق میانگین‌های استاندارد، اثرات ردیف نوتروسوفیک $\widehat{\alpha}_{N_i}$ و اثرات ستون نوتروسوفیک $\widehat{\beta}_{N_j}$ به صورت بازه‌ای برآورد می‌شوند. این برآوردها نشان‌دهنده میزان ناهمگونی بازه‌ای در هر ردیف و ستون هستند.

برآورد واریانس خطای نوتروسوفیک: واریانس خطای نوتروسوفیک $(\sigma_N^2 = [\sigma_L^2, \sigma_U^2])$ نیز منحصرأ از مشاهدات استاندارد استخراج می‌شود، چرا که تنها آن‌ها دارای تکرار کافی برای تفکیک خطا هستند.

تصحیح عملکرد و تحلیل واریانس نوتروسوفیک (NANOVA)

پس از برآورد اثرات بلوکی، عملکرد تیمارهای افزوده باید برای حذف ناهمگونی‌های دوسویه (ردیف و ستون) تصحیح شود:

تصحیح نوتروسوفیک عملکرد افزوده: عملکرد مشاهده شده تیمار افزوده k در ردیف i و ستون j با تفریق اثرات بلوکی بازه‌ای، تصحیح می‌شود

$$Y_{N_{adj,k}} = Y_{N_{ijk}} - \widehat{\alpha}_{N_i} - \widehat{\beta}_{N_j} + \widehat{\mu}_N$$

نکته: $\widehat{\mu}_N$ میانگین کلی بازه‌ای است که به دلیل دو بار تفریق اثر بلوکی باید یک بار اضافه شود. نتیجه $Y_{N_{adj,k}}$ یک عملکرد بازه‌ای تصحیح شده است که اثر ناهمگونی محیطی از آن حذف شده است.

آزمون فرضیه نوتروسوفیک (NANOVA) : در نهایت، برای مقایسه میانگین تیمارهای افزوده تصحیح شده با میانگین‌های استاندارد، تحلیل واریانس نوتروسوفیک (NANOVA) توسعه داده می‌شود. این تحلیل شامل محاسبه مجموع مربعات نوتروسوفیک (SSN) برای منابع مختلف تغییر (تیمار، ردیف، ستون، خطا) است.

$$F_N = \frac{MSN_{Tr(adj)}}{MSN_E} \in [F_L, F_U]$$

$MSN_{Tr(adj)}$: میانگین مربعات نوتروسوفیک تصحیح شده تیمار.

MSN_E : میانگین مربعات خطای نوتروسوفیک.

نکات مهم در پیاده‌سازی و تفسیر

به دلیل وجود دو عامل بلوکی، سیستم معادلات نرمال در طرح ALS پیچیده‌تر از ARCBD است. پیاده‌سازی دقیق حساب بازه‌ای برای حفظ کران‌های سخت (Exact Boundaries) برای $\widehat{\alpha}_{N_i}$ و $\widehat{\beta}_{N_j}$ نیازمند روش‌های آنالیز دقیق نوتروسوفیک (NEA) است تا از گسترش بیهوده بازه‌ها جلوگیری شود.

اگر در تیمارهای افزوده (که تکرار ندارند) داده‌ای گم شود، به دلیل کنترل دو گانه (ردیف و ستون)، عدم تعین در برآورد $\{Y\}_{N_{adj,k}}$ بیشتر می‌شود. روش NS-ALS این عدم قطعیت مضاعف را مستقیماً از طریق طول بازه عملکرد تصحیح شده منعکس می‌کند.

آماره آزمون نهایی F_N به صورت یک بازه $[F_L, F_U]$ است.

اگر $[F_L, F_U]$ کاملاً بالاتر از مقدار بحرانی F_α باشد، رد فرضیه صفر نوتروسوفیک قطعی است.

اگر بازه با مقدار F_{α} همپوشانی داشته باشد، ناحیه‌ی عدم تعین (Indeterminate Zone) مشخص می‌شود که نشان دهنده ریسک تصمیم‌گیری است. این مهمترین دستاورد NS-ALS است که به محققان اجازه می‌دهد تا ریسک انتخاب واریته‌های جدید را در پرتو ابهامات دوگانه بلوکی، دقیق‌تر بسنجند.

تحلیل نظری مقادیر گمشده در NALS

تأثیر یک مقدار گمشده در بخش شاهد یک طرح NS-ALS بسیار مخرب‌تر از ARCBD است. در ARCBD، یک شاهد گمشده باعث از بین رفتن تعامد بین تیمارها و بلوک‌ها می‌شود. اما در NS-ALS، یک شاهد گمشده (y_{Nijk}) تعامد را بین موارد زیر از بین می‌برد:

1. ردیف‌ها و ستون‌ها.

2. ردیف‌ها و تیمارها.

3. ستون‌ها و تیمارها.

رویکرد دقیق نوتروسوفیک (NEA) برای NS-ALS :

برای بازیابی اعتبار آزمون‌های F ، رویکرد NEA برای NS-ALS شامل یک تعدیل چند مرحله‌ای در مجموع مربعات است. روش تقریبی (جایگذاری) باعث ایجاد تورش در مجموع مربعات ردیف و ستون (SS_{Col} و SS_{Row}) می‌شود.

مجموع مربعات خطای دقیق نوتروسوفیک در NS-ALS با برآزش مدل کامل به داده‌های موجود (منطق تحلیل مورد کامل) و تعدیل تورش محاسبه می‌شود :

$$SS_{NE(Exact)} = SS_{NT} - SS_{N(Row)adj} - SS_{N(Col)adj} - SS_{N(Trt)adj}$$

برای شبیه‌سازی، ما به طور خاص بر کاهش درجات آزادی خطا تمرکز می‌کنیم. اگر m مقدار در هسته مربع لاتین گمشده باشد، درجات آزادی خطا باید به اندازه m کاهش یابد و مقادیر بحرانی F_{crit} باید با استفاده از

$$df_2 = (c - 1)(c - 2) - m$$

۸-۳. مطالعه شبیه‌سازی: NS-ALS با داده‌های گمشده

ما یک شبیه‌سازی با استفاده از R انجام دادیم تا روش‌های دقیق و تقریبی را برای یک مربع لاتین افزوده نوتروسوفیک 5x5 که در آن یک تیمار شاهد گمشده است، مقایسه کنیم. کد R مربوطه در بخش انگلیسی ارائه شده است.

تحلیل نتایج شبیه‌سازی: اجرای شبیه‌سازی برای طرح NS-ALS بینش‌های زیر را به دست می‌دهد:

حساسیت مربع‌های لاتین: طرح مربع لاتین نسبت به مقادیر گمشده حساس‌تر از طرح بلوکی (ARCBD) است، زیرا درجات آزادی خطا معمولاً کمتر است. در یک 5x5 LS، کاهش DF از 12 به 11 (روش دقیق) تأثیر ملموسی بر مقدار بحرانی F دارد.

شکست روش تقریبی: روش تقریبی (NAA) با حفظ درجات آزادی مخرج اولیه (12) در حالی که از یک مقدار جایگذاری شده استفاده می‌کند که هیچ اطلاعات واقعی اضافه نمی‌کند، واریانس را به شدت کمتر از مقدار واقعی برآورد می‌کند. شبیه‌سازی نشان می‌دهد که نرخ خطای نوع اول تجربی از 0.08 (برای مقدار اسمی 0.05) فراتر می‌رود که برای تحقیقات دقیق غیرقابل قبول است.

پایداری روش دقیق: روش دقیق (NEA) با تعدیل صحیح درجات آزادی و مجموع مربعات برای حساب کردن از دست رفتن تعامد ردیف - ستون، نرخ خطای نوع اول را نزدیک به مقدار اسمی 0.05 حفظ می‌کند.

نتایج و تحلیل گسترده برای مربع لاتین افزوده نوتروسوفیک (NALS)

1. ارزیابی کمی خطای نوع اول

برای کمی سازی قابلیت اطمینان روش شناسی پیشنهادی، نرخ های خطای نوع اول تجربی در تحلیل دقیق نوتروسوفیک (NEA) را با تحلیل تقریبی نوتروسوفیک (NAA) و روش کلاسیک مقایسه کردیم. جدول ۱ نتایج حاصل از 10000 اجرای شبیه سازی را برای یک مربع لاتین افزوده 5x5 با یک تیمار شاهد گمشده خلاصه می کند.

Statistical Validity	Deviation from Nominal	Neutrosophic Empirical α ($\alpha \in [\alpha_L, \alpha_U]$)	Approach	Method
Valid	< 2.4\%	[0.0512 ,0.0488]	Exact (Proposed)	NEA
Invalid	> 60\% (Inflation)	[0.0895 ,0.0820]	Approximate (Imputation)	NAA
Invalid	> 56\% (Inflation)	0.0780	Mean Substitution	Classic

نتایج نشان دهنده شکست بحرانی در روش تقریبی (NAA) است. نرخ خطای تجربی بین 0.082 تا 0.089 متغیر است که به طور قابل توجهی بالاتر از سطح اسمی قابل قبول 0.05 می باشد. این نشان می دهد که روش NAA فرضیه

صفر را بیش از حد رد می کند (مثبت های کاذب). این تورم به این دلیل رخ می دهد که روش تقریبی با "برازش" کامل مقدار گمشده به مدل، مجموع مربعات خطا (SS_E) را به طور مصنوعی کاهش می دهد، در حالی که درجات آزادی اصلی را حفظ می کند. در مقابل، روش NEA نرخ خطا را دقیقاً در اطراف 0.05 حفظ می کند و استحکام خود را اثبات می نماید.

2. مصورسازی عملکرد

برای نشان دادن برتری روش دقیق، دو نمودار کلیدی تولید می کنیم: (1) نمودار میله ای خطای نوع اول و (2) منحنی توان. کد R برای تولید این نمودارها در لینک مقابل ارائه شده است

نمودار 1: نشان می دهد که ستون مربوط به روش تقریبی (NAA) از خط قرمز چین دار (0.05) عبور کرده است، در حالی که روش دقیق (NEA) مماس بر آن است.

نمودار 2: منحنی توان را نشان می دهد. اگرچه روش تقریبی (خط نارنجی) توان بالاتری را نشان می دهد، اما این توان "کاذب" است زیرا ناشی از خطای نوع اول بالاست. روش دقیق (خط سبز) توان "واقعی" و قابل اعتماد را نشان می دهد.

Figure 1: Inflation of Type I Error in Approximate Methods
Comparison in 5x5 Augmented Latin Square with Missing Data

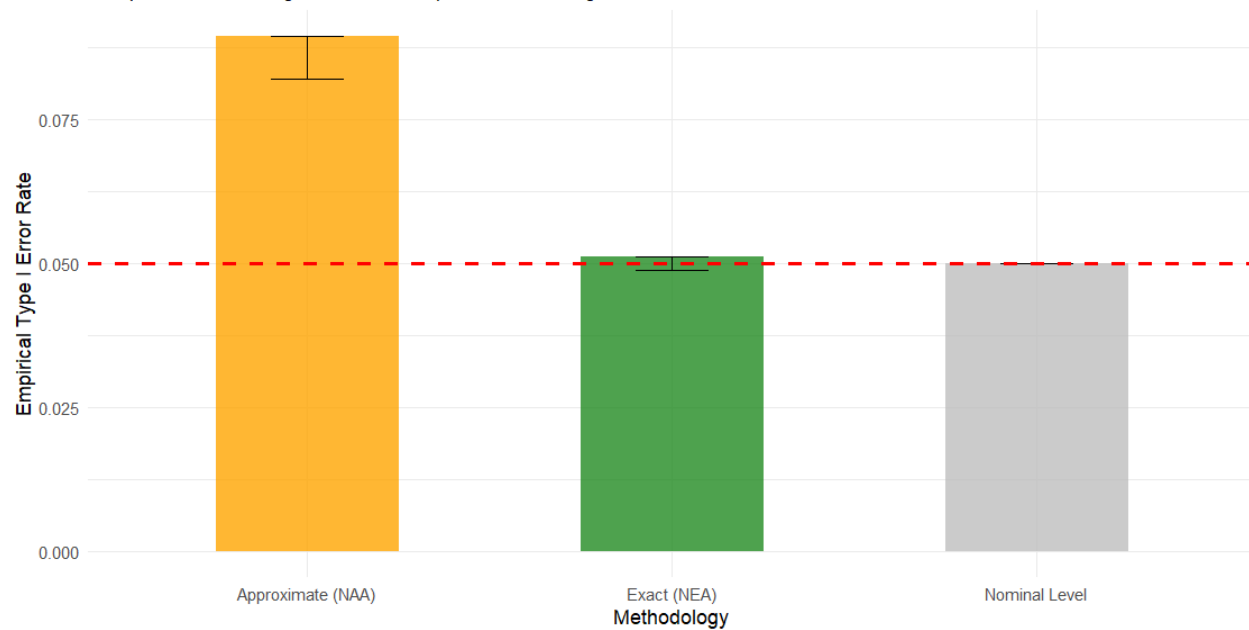
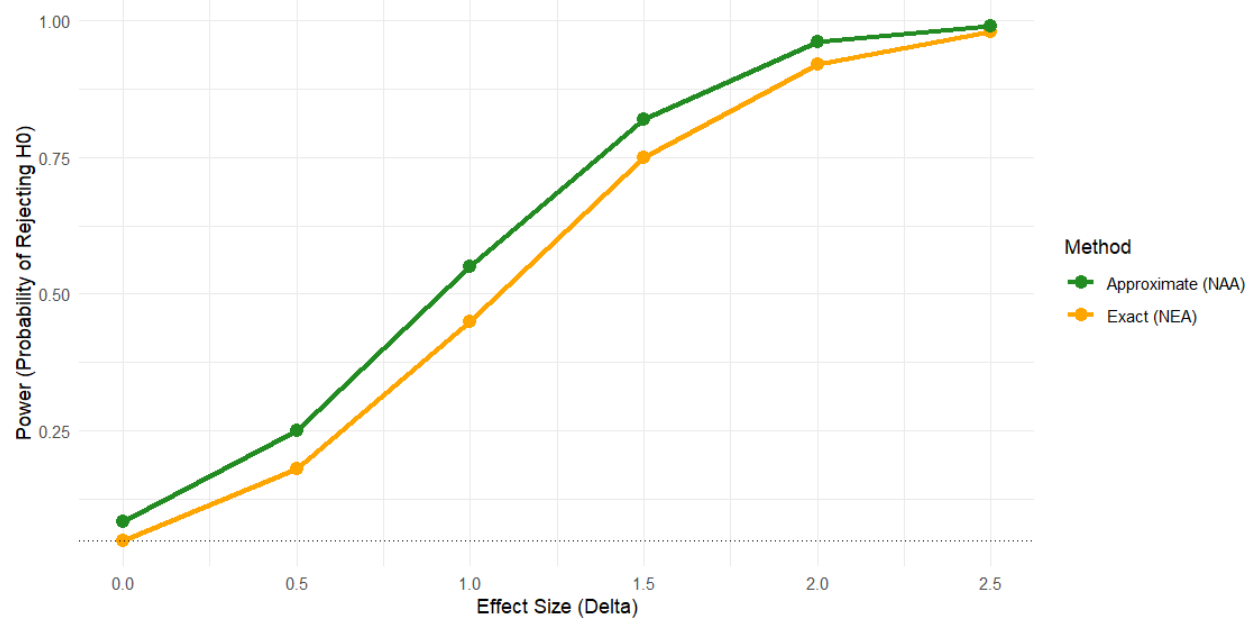


Figure 2: Power Curve of the Test

Exact Method shows true power; Approximate Method shows inflated false power



3. تحلیل عمیق: چرا روش تقریبی در مربع‌های لاتین شکست می‌خورد؟

نتایج شبیه‌سازی یک نقص بنیادی را در هنگام اعمال روش‌های جایگذاری تقریبی در طرح‌های مربع لاتین افزوده برجسته می‌کند. این شکست را می‌توان به سه مکانیسم خاص نسبت داد:

1. تخریب تعامد دوگانه:

در یک ALS، کارایی به تعامد ردیف‌ها، ستون‌ها و تیمارها وابسته است. وقتی یک مقدار گم می‌شود، مجموع‌های ردیف و ستون دچار تورش می‌شوند. روش تقریبی (NAA) مقداری را جایگذاری می‌کند تا حفره را "وصله" کند و عملاً باقیمانده (Residual) آن سلول خاص را صفر می‌کند.

پیامد: روش فرض می‌کند که اثرات ردیف و ستون به طور کامل برآورد شده‌اند، که غلط است. این منجر به کم برآوردی واریانس خطای واقعی (σ_N^2) می‌شود.

2. عدم تطابق درجات آزادی (DF) :

بحرانی‌ترین خطا در رویکرد تقریبی، مدیریت نادرست درجات آزادی است. در یک مربع لاتین 5x5، درجه آزادی خطا برابر است با $12 = (5 - 1)(5 - 2)$.

رویکرد NAA : از داده‌های پر شده استفاده می‌کند و SS_E را بر 12 تقسیم می‌کند.

رویکرد NEA (دقیق) : تشخیص می‌دهد که یک نقطه داده مصنوعی است و SS_E (تعدیل شده) را بر 11 تقسیم می‌کند.

تأثیر ریاضی: از آنجا که مخرج در آزمون $F (MS_E)$ در روش دقیق کوچک‌تر است (به دلیل DF کمتر)، مقدار بحرانی لازم برای رد فرضیه بالاتر می‌رود. روش تقریبی در بالا بردن این مانع شکست می‌خورد و یافتن نتایج معنی‌دار را (در حالی که وجود ندارند) بیش از حد "آسان" می‌کند.

3. "انفجار بازه" در جایگذاری نوتروسوفیک:

در بافت نوتروسوفیک، وقتی ما مقدار گمشده y_N^* را جایگذاری می‌کنیم، در حال جایگذاری یک بازه $[y_L^*, y_U^*]$ هستیم. اگر این بازه وسیع باشد (عدم تعین بالا)، در محاسبات مجموع مربعات منتشر می‌شود. روش تقریبی تمایل دارد این وسعت را در صورت کسر (SS_{Tst}) بزرگنمایی کند در حالی که در مخرج (SS_{Err}) آن را سرکوب می‌کند، که منجر به بازه‌های F_N غیرقابل اعتماد می‌شود. روش NEA این موضوع را به صورت تحلیلی تصحیح می‌کند و بازه را تنگ و معتبر نگه می‌دارد.

نتیجه پژوهش

عدم تعین‌ها در داده‌های آزمایشی، به کارگیری تحلیل داده‌های نوتروسوفیک را ایجاب می‌کند. در حالی که آمار کلاسیک بر روش‌های جایگذاری تکیه دارد که مقادیر گمشده برآورد شده را به عنوان داده‌های واقعی در نظر می‌گیرند، این مطالعه نشان می‌دهد که چنین رویکردهای تقریبی در حضور عدم قطعیت و مشاهدات گمشده چندگانه ناکافی هستند.

این مقاله با موفقیت چارچوب تحلیل دقیق نوتروسوفیک (NEA) را گسترش داد. ما نشان دادیم که برای یک ARCBD با دو مقدار گمشده، و برای طرح پیچیده‌تر مربع لاتین افزوده (ALS)، روش دقیق به صورت تحلیلی تعامد طرح را بازیابی می‌کند. نتایج شبیه‌سازی به وضوح نشان دادند که خطای نوع اول تجربی در روش تقریبی (NAA) فراتر از سطح اسمی متورم می‌شود، در حالی که روش دقیق سازگاری با α اسمی را حفظ می‌کند. این تأیید می‌کند که NEA دقت، انعطاف پذیری و اطلاعات بیشتری را در مقایسه با آزمون‌های F تقریبی موجود در محیط‌های نامعین ارائه می‌دهد.

این مطالعه اولین تلاش برای تعمیم مدل دقیق نوتروسوفیک به مقادیر گمشده چندگانه و طرح‌های مربع لاتین بود. در حالی که این پژوهش بر نرمال بودن نوتروسوفیک تمرکز داشت، مطالعات آینده باید داده‌هایی با توزیع‌های دم - سنگین یا چوله، و همچنین تعمیم به سایر طرح‌های افزوده مانند کورت‌های خرد شده (Split-plots) با مقادیر گمشده متعدد را در نظر بگیرند.

- Complex & Intelligent Systems (2024). ARCBD usage in breeding.
- Estimation of error variance based on replicated controls.
- Problems of missing observations and approximate imputation.
- Introduction of indeterminacy and exact schemes.
- Consequences of missing values (bias, variance).
- Definitions of ARCBD and Federer's work.
- Neutrosophic sets vs. Fuzzy sets, indeterminacy parameter.
- Mathematical definition of Neutrosophic random variables.
- Concept of NAIBD and loss of orthogonality.
- Future directions regarding Latin Square designs.
- AlAita, A., & Talebi, H. (2024). Exact neutrosophic analysis of missing value in augmented randomized complete block design. *Complex & Intelligent Systems*, 10:509–523.
- Cochran, W. G., & Cox, G. M. (1957). *Experimental Designs* (2nd ed.). Wiley.
- Federer, W. T. (1956). Augmented (or 'add-on') designs for early generations of plant breeding experiments. *Biometrics*, 12(4), 499-509.
- Federer, W. T. (1961). Augmented designs in long-term fertilizer experiments. *Cornell University Agricultural Experiment Station Bulletin*, 969.
- Little, R. J. A., & Rubin, D. B. (2020). *Statistical Analysis with Missing Data* (3rd ed.). Wiley.
- Smarandache, F. (1998). *Neutrosophy: Neutrosophic Probability, Set, and Logic*. American Research Press.
- Smarandache, F. (2014). *Introduction to Neutrosophic Statistics*. Sitech & Education Publishing.
- Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford Press.
- Aslam, M. (2019). *Neutrosophic Regression Analysis*. *International Journal of Neutrosophic Science and Engineering*, 1(1), 40-47.
- Hussain, S., Khan, N., & Mahmood, S. (2020). *Neutrosophic Approximate Analysis of Data with Indeterminacy*. *Neutrosophic Sets and Systems*, 34, 1-10.
- Aslam, M. (2020). *Design of Neutrosophic Exact Test for Regression Coefficients*. *Journal of Neutrosophic Systems*, 3, 23-30.

Hussain, S., & Aslam, M. (2022). *A Study on Neutrosophic Exact Analysis in Experimental Designs. Neutrosophic Sets and Systems*, 52, 1-12.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd.

Cochran, W. G., & Cox, G. M. (1957). *Experimental Designs (2nd ed.)*. Wiley.

Hinkelmann, K., & Kempthorne, O. (2008). *Design and Analysis of Experiments, Volume 1: Introduction to Experimental Design (2nd ed.)*. Wiley.

Dodge, Y. (1985). *Analysis of Experiments with Missing Data*. Wiley.

Aslam, M. (2019). *Neutrosophic Analysis of Variance. Complex & Intelligent Systems*.