

به نام یزدان پاک

همه سوال های

دکترای تخصصی (Ph.D)

بیوتکنولوژی کشاورزی

۱۴۰۴-۱۴۰۵

گردآورندگان

گروه مولفین پردازش

فهرست مطالب

پیش گفتار ۷

آزمون دکترای بیوتکنولوژی کشاورزی ۱۴۰۴

آمار و طرح آزمایش‌ها.....	۱۱
پاسخ‌نامه آمار و طرح آزمایش‌ها.....	۱۳
درسنامه مروری آمار و طرح آزمایش‌ها.....	۲۶
ژنتیک.....	۳۵
پاسخ‌نامه ژنتیک.....	۳۶
درسنامه مروری ژنتیک.....	۵۰
اصلاح نباتات.....	۵۹
پاسخ‌نامه اصلاح نباتات.....	۶۰
درسنامه مروری اصلاح نباتات.....	۷۱
بیوشیمی پیشرفته.....	۷۵
پاسخ‌نامه بیوشیمی پیشرفته.....	۷۶
درسنامه مروری بیوشیمی پیشرفته.....	۸۶
کشت سلول و بافت گیاهی.....	۸۹
پاسخ‌نامه کشت سلول و بافت گیاهی.....	۹۰
ژنتیک مولکولی.....	۹۹
پاسخ‌نامه ژنتیک مولکولی.....	۱۰۰
درسنامه مروری ژنتیک مولکولی.....	۱۰۸
مهندسی ژنتیک.....	۱۱۲
پاسخ‌نامه مهندسی ژنتیک.....	۱۱۴
درسنامه مروری مهندسی ژنتیک.....	۱۲۳

آزمون دکترای بیوتکنولوژی کشاورزی ۱۴۰۵

آمار و طرح آزمایش‌ها.....	۱۲۶
پاسخ‌نامه آمار و طرح آزمایش‌ها.....	۱۲۸
درسنامه مروری آمار و طرح آزمایش‌ها.....	۱۴۰
ژنتیک.....	۱۴۷
پاسخ‌نامه ژنتیک.....	۱۴۸
درسنامه مروری ژنتیک.....	۱۵۸
اصلاح نباتات.....	۱۶۱
پاسخ‌نامه اصلاح نباتات.....	۱۶۳
درسنامه مروری اصلاح نباتات.....	۱۷۹
بیوشیمی پیشرفته.....	۱۸۵
پاسخ‌نامه بیوشیمی پیشرفته.....	۱۸۶
درسنامه مروری بیوشیمی پیشرفته.....	۱۹۶
کشت سلول و بافت گیاهی.....	۱۹۸
پاسخ‌نامه کشت سلول و بافت گیاهی.....	۱۹۹
درسنامه مروری کشت سلول و بافت گیاهی.....	۲۰۹
ژنتیک مولکولی.....	۲۱۴

۲۱۵	پاسخ نامه ژنتیک مولکولی.....
۲۲۵	درسنامه مروری ژنتیک مولکولی.....
۲۲۸	مهندسی ژنتیک.....
۲۲۹	پاسخ نامه مهندسی ژنتیک.....
۲۴۱	درسنامه مروری مهندسی ژنتیک.....
۲۴۳	واژگان و اصطلاحات.....
۲۵۹	ضمایم.....
۲۶۹	منابع.....

جدول جامع توزیع فراوانی موضوعی سوالات آزمون دکترای بیوتکنولوژی کشاورزی (۱۴۰۰ تا ۱۴۰۵)

حوزه اصلی / مباحث فرعی	۱۴۰۰	۱۴۰۱	۱۴۰۲	۱۴۰۳	۱۴۰۴	۱۴۰۵	مجموع	درصد فراوانی
۱. آمار و طرح آزمایش‌ها	۱۳	۱۳	۱۱	۱۰	۱۰	۱۰	۶۷	۱۴.۰٪
۱.۱. آمار توصیفی، احتمالات و توزیع‌های نمونه‌برداری (نرمال، t، دو جمله‌ای و...)	۵	۵	۵	۴	۵	۴	۲۸	۵.۹٪
۱.۲. آزمون‌های مقایسه میانگین و خطاهای آزمایشی (LSD، توکی و دانکن)	۲	۲	۱	۱	۲	۲	۱۰	۲.۱٪
۱.۳. طرح‌های پایه‌ای و مرکب (CRD، بلوک‌های تصادفی، مربع لاتین و تقاطع‌ها)	۴	۴	۳	۳	۲	۲	۱۸	۳.۸٪
۱.۴. آزمایش‌های کرت‌های خردشده (Split-Plot) و ساختار درجات آزادی	۱	۱	۱	۱	۱	۱	۶	۱.۳٪
۱.۵. رگرسیون، همبستگی و تجزیه واریانس با زیرنمونه‌گیری	۱	۱	۱	۱	۰	۱	۵	۱.۰٪
۲. ژنتیک پایه و تکمیلی	۱۱	۱۱	۱۱	۱۰	۱۰	۱۰	۶۳	۱۳.۲٪
۲.۱. قوانین مندل، برهم‌کنش ژن‌ها، اپیستازی و آلل‌های کشنده	۳	۳	۳	۳	۲	۳	۱۷	۳.۶٪
۲.۲. پیوستگی ژنی، کراسینگ‌اور و نقشه‌برداری پیوستگی	۳	۲	۲	۲	۲	۲	۱۳	۲.۷٪
۲.۳. سیتوژنتیک و ناهنجاری‌های عددی و ساختاری کروموزوم‌ها (کاریوتیپ، وارونگی و پلوئیدی)	۳	۳	۲	۲	۳	۲	۱۵	۳.۱٪
۲.۴. ژنتیک جمعیت و تعادل هاردی-وینبرگ	۲	۲	۲	۱	۲	۱	۱۰	۲.۱٪
۲.۵. ژنتیک اندامکی، توارث سیتوپلاسمی و تبارزایی مولکولی	۰	۱	۲	۲	۱	۲	۸	۱.۷٪
۳. اصلاح نباتات و ژنتیک کمی	۱۶	۱۶	۱۱	۱۰	۱۰	۱۰	۷۳	۱۵.۳٪
۳.۱. روش‌های اصلاح گیاهان خودگشن (شجره‌ای، بالک، SSD و	۵	۴	۳	۳	۳	۳	۲۱	۴.۴٪

حوزه اصلی / مباحث فرعی	۱۴۰۰	۱۴۰۱	۱۴۰۲	۱۴۰۳	۱۴۰۴	۱۴۰۵	مجموع	درصد فراوانی
دابل هاپلوئیدی)								
۳.۲. روش های اصلاح گیاهان دگرگشن، واریته های سنتتیک و هیبریداسیون	۴	۵	۳	۳	۲	۳	۲۰	۴.۲٪
۳.۳. سیستم های نرغیمی ژنتیکی و سیتوپلاسمی (CMS) و خودناسازگاری	۳	۳	۲	۲	۲	۲	۱۴	۲.۹٪
۳.۴. ژنتیک کمی، برآورد واریانس ها، وراثت پذیری و پاسخ به گزینش	۲	۲	۲	۱	۲	۱	۱۰	۲.۱٪
۳.۵. تجزیه دی آلل، ترکیب پذیری عمومی و خصوصی (GCA/SCA) و پاتولوژی گیاهی	۲	۲	۱	۱	۱	۱	۸	۱.۷٪
۴. بیوشیمی پیشرفته	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۶۰	۱۲.۶٪
۴.۱. ساختار اسیدهای آمینه، پروتئین ها، زوایای رامچاندرا و اتصالات	۳	۳	۴	۳	۳	۳	۱۹	۴.۰٪
۴.۲. بیوفیزیک ساختاری، موتیف های اتصال DNA و دومین ها (Zinc Finger و...)	۲	۲	۱	۲	۲	۲	۱۱	۲.۳٪
۴.۳. چارون های مولکولی، فولدینگ پروتئین و سیستم یوبیکوئیتین-پروتئازوم	۲	۲	۲	۲	۲	۲	۱۲	۲.۵٪
۴.۴. تکنیک های آنالیز دستگاهی و خالص سازی (MS/MS، کریستالوگرافی، کروماتوگرافی)	۲	۲	۲	۲	۲	۲	۱۲	۲.۵٪
۴.۵. تغییرات پس از ترجمه (PTMs) و پیام رسانی سلولی (فسفوریلاسیون، لیپیداسیون)	۱	۱	۱	۱	۱	۱	۶	۱.۳٪
۵. کشت سلول و بافت گیاهی	۱۰	۱۰	۱۰	۱۰	۱۰	۱۰	۶۰	۱۲.۶٪
۵.۱. اصول ریزازدیادی، مراحل کشت، تمایز و اندام زایی نابجا	۲	۲	۳	۲	۲	۳	۱۴	۲.۹٪

حوزه اصلی / مباحث فرعی	۱۴۰۰	۱۴۰۱	۱۴۰۲	۱۴۰۳	۱۴۰۴	۱۴۰۵	مجموع	درصد فراوانی
۵.۲. فیزیولوژی محیط کشت، منابع کربن، آب، آگار و تعادل یونی (نیتрат/آمونوم)	۳	۲	۲	۲	۳	۲	۱۴	۲.۹٪
۵.۳. تنظیم‌کننده‌های رشد گیاهی و نقش‌های مورفوزنیک هورمون‌ها	۲	۳	۲	۲	۲	۲	۱۳	۲.۷٪
۵.۴. جنین‌زایی سوماتیکی، هاپلوئیدی و کشت گرده و تخمک	۲	۲	۱	۲	۱	۱	۹	۱.۹٪
۵.۵. ناهنجاری‌های فیزیولوژیک درون‌شیشه‌ای (شیشه‌ای شدن، قهوه‌ای شدن، تنوع سوماکلونال)	۱	۱	۲	۲	۲	۲	۱۰	۲.۱٪
۶. ژنتیک مولکولی	۱۰	۱۰	۹	۱۰	۱۰	۱۰	۵۹	۱۲.۳٪
۶.۱. سازمان‌دهی ژنوم، معماری ژن (اینترن/اگزون) و عناصر متحرک	۲	۲	۱	۲	۲	۲	۱۱	۲.۳٪
۶.۲. همانندسازی DNA، ترمیم و آنزیم‌های دخیل	۲	۲	۲	۲	۲	۲	۱۲	۲.۵٪
۶.۳. رونویسی، پروموتورها، فاکتورهای عمومی و تنظیمی در پروکاریوت و یوکاریوت	۲	۲	۲	۲	۲	۲	۱۲	۲.۵٪
۶.۴. پیرایش RNA، کلاهک‌گذاری، پلی‌آدنیلاسیون و اسپلیسوزوم	۲	۲	۲	۲	۲	۲	۱۲	۲.۵٪
۶.۵. ترجمه، کد ژنتیکی، ریبوزوم، فاکتورهای آغاز/پایان و جهش‌های نقطه‌ای	۲	۲	۲	۲	۲	۲	۱۲	۲.۵٪
۷. مهندسی ژنتیک و بیوتکنولوژی نوین	۱۰	۱۰	۸	۱۰	۱۰	۱۰	۵۸	۱۲.۱٪
۷.۱. آنزیم‌های برشی، لیگازها، اصلاح‌کننده‌های DNA و ناقلین کلونینگ/بیانی	۳	۳	۲	۲	۳	۲	۱۵	۳.۱٪
۷.۲. روش‌های انتقال ژن (اگروباکتریوم، تفنگ ژنی، ترانسفورماسیون پلاستییدی)	۲	۲	۲	۳	۲	۳	۱۴	۲.۹٪

حوزه اصلی / مباحث فرعی	۱۴۰۰	۱۴۰۱	۱۴۰۲	۱۴۰۳	۱۴۰۴	۱۴۰۵	مجموع	درصد فراوانی
۷.۳. تکنیک‌های PCR، جهش‌زایی و کتابخانه‌های ژنومی/cDNA	۲	۲	۱	۲	۲	۲	۱۱	۲.۳٪
۷.۴. خاموشی ژن، تداخل RNA (RNAi) و نشانگرهای انتخابی/گزارشگر	۲	۲	۲	۲	۲	۲	۱۲	۲.۵٪
۷.۵. فناوری‌های نوین ویرایش ژنوم (سیستم CRISPR/Cas، ZFN و TALEN)	۱	۱	۱	۱	۱	۱	۶	۱.۳٪
مجموع سوالات هر سال	۸۰	۸۰	۷۰	۷۰	۷۰	۷۰	۴۴۰	۱۰۰٪

گزارش تحلیلی و توصیه‌های راهبردی

۱. تحلیل روندها و تغییرات ساختاری آزمون

بررسی ساختار دفترچه‌های آزمون از سال ۱۴۰۰ تا ۱۴۰۵ نشان‌دهنده یک ثبات راهبردی پس از تغییرات سال ۱۴۰۲ است. از سال ۱۴۰۲ به بعد، تعداد کل سوالات دفترچه تخصصی بیوتکنولوژی کشاورزی روی ۷۰ سوال تثبیت گردیده و ساختار هر درس با ضریب مشخص ۱۰ سوالی (۷ درس تخصصی) ارائه می‌شود. گرایش طراحان از سوالات حفظی و تعریفی در سال‌های اولیه (۱۴۰۰ و ۱۴۰۱) به سمت سوالات تلفیقی، محاسبات مفهومی-سرعتی و کاربردهای صنعتی بیوتکنولوژی در سال‌های ۱۴۰۴ و ۱۴۰۵ تغییر جهت داده است.

۲. معرفی مباحث کلیدی و پرتکرار (هسته امتیازآور)

- **آمار و طرح آزمایش‌ها:** تسلط بر فرمول‌های محاسبه درجات آزادی (به‌ویژه طرح‌های کرت‌های خردشده و زیرنمونه‌گیری) و تعاریف پایه‌ای خطاهای نوع اول و دوم به همراه مقایسه آزمون‌های توکی و LSD پای ثابت آزمون‌ها است.
- **ژنتیک و اصلاح نباتات:** مباحث پیوستگی ژنی و فواصل سانتی‌مورگان، انواع اپیستازی‌ها، شناسایی لاین‌های نرعیقیم، تفکیک روش‌های خودگشن/دگرگشن و محاسبات GCA در تلاقی دی‌آلل بیشترین فراوانی را دارند.
- **بیوشیمی و کشت بافت:** نمودار رامچاندرا، موتیف Zinc Finger، چاپرون‌ها (سیستم‌های BiP و هیدرولیز ATP)، مراحل ریزازدیادی، تنظیم‌کننده‌های رشد، و تعادل نیترات/آمونیم در محیط کشت مهم‌ترین نقاط تمرکز هستند.
- **ژنتیک مولکولی و مهندسی ژنتیک:** پیرایش RNA و snRNP، سیستم‌های ویرایش ژنوم (CRISPR/Cas)، ژن‌های ناحیه بیماری‌زایی Ti-Plasmid و انتقال ژن به کلروپلاست در سال‌های اخیر مورد توجه ویژه بوده‌اند.

۳. توصیه‌های راهبردی برای داوطلبان

۱. **محاسبات سرعتی:** در دروسی مانند آمار، ژنتیک کمی و اصلاح، فرمول‌ها و روابط میانبر (مانند قوانین تفکیک درجات آزادی، ضرب احتمالات مستقل و تفکیک ژن‌ها) را جایگزین محاسبات وقت‌گیر و طولانی کنید.
۲. **دیدگاه سیستمی و تطبیقی:** ارتباط متقابل میان ساختار ژنتیکی گیاه (خودگشن، دگرگشن یا کلونال) و روش اصلاحی مربوطه را به‌صورت تطبیقی فراگیرید.
۳. **تلفیق مباحث سلولی-مولکولی با بیوتکنولوژی کاربردی:** درک سازوکار آنزیم‌ها و تکنیک‌های دستگاهی (نظیر ساترن‌بلاتینگ، انواع PCR، کروماتوگرافی تمایلی و طیف‌سنجی جرم متوالی) به داوطلب امکان می‌دهد سوالات ترکیبی جدید را در کمترین زمان پاسخ دهد.

مقدمه تحلیلی بر روند آزمون‌های دکترای تخصصی بیوتکنولوژی کشاورزی (۱۴۰۰ تا ۱۴۰۵)

بررسی ساختاری و محتوایی آزمون‌های دکترای بیوتکنولوژی کشاورزی در بازه زمانی شش ساله اخیر، نشان‌دهنده یک دگردیسی معنادار در رویکرد سنجش علمی داوطلبان است. این تحولات، از تغییر در بودجه‌بندی سوالات تا شیفت پارادایم‌های مفهومی، نقشه راه روشنی را برای تدوین منابع آموزشی جامع و هدفمند ترسیم می‌کند. تحلیل عمیق این روندها، مرز بین مطالعه کورکورانه و آمادگی استراتژیک را مشخص می‌سازد.

روند تحولات کمی و تغییرات ساختاری آزمون

مهم‌ترین تغییر ساختاری در این دوره شش ساله، افزایش چشمگیر حجم و عمق سوالات از سال ۱۴۰۳ به بعد است. آزمون‌ها از قالب ۸۰ و ۷۰ سوالی در سال‌های ۱۴۰۰ تا ۱۴۰۲، به ساختار جامع‌تر ۱۱۰ سوالی در سال‌های ۱۴۰۳ تا ۱۴۰۵ ارتقا یافته‌اند. این افزایش ظرفیت، به طراحان اجازه داده است تا از سوالات تک‌بعدی فاصله گرفته و به سمت مفاهیم ترکیبی و بین‌رشته‌ای حرکت کنند. افزایش ضریب نفوذ دروسی مانند بیوشیمی پیشرفته و مهندسی ژنتیک در این سال‌ها، نشان از اهمیت روزافزون درک مکانیسم‌های مولکولی در بیوتکنولوژی مدرن دارد.

مباحث لنگرگاهی و همواره پرتکرار

داده‌های استخراج‌شده نشان می‌دهد که برخی از سرفصل‌ها به عنوان ستون فقرات آزمون دکترای همواره مورد توجه طراحان قرار داشته‌اند:

- **ژنتیک کمی و بیومتری:** مفاهیم پایه‌ای نظیر محاسبه اجزای واریانس (افزایشی، غالبیت، اپیستازی)، برآورد وراثت‌پذیری (عمومی و خصوصی) و شاخص‌های ترکیب‌پذیری (GCA و SCA) در دی‌آلل‌ها، پای ثابت تمام آزمون‌ها بوده‌اند.
- **سیتوژنتیک گیاهی:** ناهنجاری‌های کروموزومی، به ویژه تحلیل رفتار میوزی در هتروزیگوت‌های وارونگی (پاراسنتریک و پری‌سنتریک) و پیامدهای پلوئیدی، هر ساله سهم قابل توجهی از سوالات ژنتیک را به خود اختصاص داده‌اند.
- **پروتکل‌های کشت بافت:** مراحل ریزازدیادی، نقش هورمون‌ها در اندام‌زایی و جنین‌زایی سوماتیک، و چالش‌های محیط‌های کشت (مانند شیشه‌ای شدن و نقش آگار) از مباحث کلیدی و گریزناپذیر این بخش هستند.
- **آمار و طرح آزمایش‌ها:** تحلیل درجات آزادی در طرح‌های پایه (به‌ویژه کرت‌های خردشده و مربع لاتین) و آزمون‌های مقایسه میانگین، همواره در خط مقدم سوالات آماری قرار دارند.

چرخش‌های مفهومی و تغییر گرایش طراحان

روند سوالات در سه سال اخیر (۱۴۰۳ تا ۱۴۰۵) حاکی از یک چرخش آشکار از «دانش حفظی و کلاسیک» به سمت «تکنولوژی‌های نوین و تحلیل‌های سیستمیک» است:

- **گذر از ژنتیک کلاسیک به مهندسی ژنوم:** تمرکز سوالات از روش‌های سنتی انتقال ژن، به سیستم‌های پیشرفته ویرایش ژنوم مانند CRISPR/Cas9 (عملکرد sgRNA و PAM)، تکنیک‌های خاموشی ژن (RNAi و نقش Dicer)، و روش‌های ژنتیک معکوس نظیر TILLING معطوف شده است.

- از تکثیر ساده به اپی ژنتیک و ناپایداری ژنومی: در مبحث کشت بافت، توجه طراحان از فرمولاسیون ساده محیط کشت به سمت درک عمیق مکانیسم‌های تنوع سوماکلونال، تغییرات اپی ژنتیکی (متیلاسیون DNA) ناشی از تنش‌های درون‌شیشه‌ای، و نجات جنین جلب شده است.

- تغییر فاز در به نژادی: رویکردهای کلاسیک اصلاحی جای خود را به استراتژی‌های مدرنی چون گزینش ژنومی، Speed Breeding (تسریع چرخه اصلاحی)، و نقشه یابی ارتباطی (Association Mapping) بر پایه عدم تعادل لینکاژی داده‌اند.

شخصیت‌های علمی و تئوری‌های بسامددار

ردپای برخی نظریه‌پردازان و مکاتب علمی در ادبیات سوالات و کلیدهای آزمون به شدت پررنگ است. تسلط بر دیدگاه‌ها و مدل‌های این دانشمندان برای داوطلبان امتیازآور است:

- **هاردی و واینبرگ (Weinberg & Hardy):** محاسبات فراوانی آللی و ژنوتیپی بر اساس قانون تعادل هاردی-واینبرگ، پرتکرارترین مفهوم در بخش ژنتیک جمعیت است.
- **ابرهارت و راسل (Russel & Eberhart):** در مبحث پایداری عملکرد و اثر متقابل ژنوتیپ و محیط، مدل رگرسیونی این دو دانشمند به همراه شاخص‌های وریک و شوکلا بیشترین ارجاع را دارند.
- **لارکین و اسکاوکرافت (Scowcroft & Larkin):** مفهوم تنوع سوماکلونال که توسط این محققین تبیین شد، در تحلیل چالش‌های ریزازدیادی همواره مورد پرسش قرار می‌گیرد.
- **اوون (Owen):** در اصلاح گیاهان دگرگشن، سیستم نره‌عقیمی و معرفی لاین‌های نگهدارنده (به‌ویژه O-type در چغندر قند که به افتخار وی نام‌گذاری شده) مکرراً در سوالات اصلاح نباتات دیده می‌شود.
- **فلور (Flor) و پینتر (Painter):** فرضیه «ژن در برابر ژن» فلور و دسته‌بندی مکانیسم‌های مقاومت به آفات (آنتی‌بیوز، آنتی‌زنوز) پینتر، در مباحث ژنتیک مقاومت بسیار کلیدی هستند.
- **فیشر، توکی و دانست (Fisher, Tukey, Dunnett):** در آمار زیستی، مقایسه توان و میزان خطای نوع اول و دوم در آزمون‌های مقایسه میانگین این دانشمندان همیشه یکی از چالش‌های محاسباتی داوطلبان است.

این تحلیل نشان می‌دهد که داوطلب موفق، کسی نیست که صرفاً سرفصل‌های مجزا را مطالعه کند، بلکه فردی است که بتواند ارتباط ارگانیک بین بیوشیمی مولکولی، تکنیک‌های کشت بافت، و اهداف اصلاح نباتات را درک نماید. ادغام مفاهیم (مانند ارتباط تنوع سوماکلونال با تغییرات اپی ژنتیک) و تسلط بر محاسبات سریع آماری و بیومتریک، شاه‌کلید موفقیت در آزمون‌های نوین دکتری بیوتکنولوژی کشاورزی خواهد بود.

پیش‌گفتار

به نام یزدان پاک

پویایی روزافزون مرزهای دانش زیستی و نقش بنیادین و راهبردی فناوری‌های زیستی در ارتقای امنیت غذایی، پایداری بوم‌سازگان‌های زراعی و توسعه کشاورزی دانش‌بنیان، آزمون ورودی دوره دکتری تخصصی (D.Ph) بیوتکنولوژی کشاورزی را به یکی از رقابتی‌ترین، حساس‌ترین و سرنوشت‌سازترین هماوردهای علمی در نظام آموزش عالی کشور بدل ساخته است. ماهیت چندرشته‌ای و تلفیقی این رشته که پیوندی ناگسستنی میان مبانی ریاضیاتی و بیومتریک، اصول وراثت و به‌نژادی، ظرایف بیوفیزیکی و بیوشیمیایی، فناوری‌های درون‌شیشه‌ای کشت بافت و ابزارهای پیشرفته دست‌ورزی ژنوم برقرار می‌سازد، آزمون ورودی آن را از یک سنجش حافظه‌محور و تک‌بعدی به سنجشی چندوجهی، ژرف و متکی بر تحلیل فرآیندی تبدیل کرده است. در این میان، داوطلبان متعهد و پرتلاش همواره با چالشی سه‌گانه روبه‌رو بوده‌اند: پراکندگی و گستردگی مراجع دانشگاهی، کمبود منابعی که پاسخگویی تشریحی را همراه با آموزش روش‌شناختی ارائه دهند، و دشواری مدیریت زمان در رویارویی با محاسبات پیچیده ژنتیک کمی، بیومتری و سازوکارهای چندمرحله‌ای مولکولی. کتاب حاضر، با عنوان «همه سوال‌های دکترای تخصصی (Ph.D) بیوتکنولوژی کشاورزی»، فرآورده تلاشی منسجم، علمی و نقادانه از سوی گروه مؤلفان انتشارات پردازش است که با هدف پاسخ به این ضرورت‌ها و ایجاد تحولی بنیادین در سبک آماده‌سازی داوطلبان تدوین گردیده است.

مبنای تدوین این اثر، واکاوی موشکافانه پرسش‌های آزمون دکتری در یک دوره ۶ ساله (از سال ۱۴۰۰ تا ۱۴۰۵) و تمرکز تحلیلی بر آزمون‌های سال‌های ۱۴۰۴ و ۱۴۰۵ بر اساس آخرین سرفصل‌ها و تغییرات ساختاری سازمان سنجش آموزش کشور است. گروه مؤلفان در طراحی و تنظیم این مرجع، روش کار خود را بر پایه پنج رکن استوار ساخته‌اند تا کتاب پیش رو از یک مجموعه سؤال و جواب صرف، به یک درسنامه جامع، مرجع کاربردی و نقشه راهبردی موفقیت تبدیل شود:

نخست؛ تحلیل آماری جامع و ماتریس توزیع فراوانی موضوعی (۱۴۰۰ تا ۱۴۰۵): در سرآغاز کتاب، جدول توزیع فراوانی موضوعی سوالات در قالب هفت حوزه اصلی و ۲۹ مبحث فرعی استاندارد تدوین شده است. این جدول در کنار مقدمه تحلیلی، تصویری شفاف از وزن مباحث، نقاط ثقل نمره‌آور و پویایی دیدگاه طراحان را در اختیار مخاطب می‌گذارد. تحلیل روندها آشکار می‌سازد که چگونه آزمون از سال ۱۴۰۳ با تجمیع گرایش‌های هم‌پوشان و تعریف ساختار پیشرفته، بار محاسباتی در دروسی نظیر آمار، بیومتری و ژنتیک کمی را تقویت نموده و همزمان در دروس بیوشیمی، کشت بافت، ژنتیک مولکولی و مهندسی ژنتیک، از رویکرد تعریفی به سمت واکاوی سازوکارهای دقیق سلولی، سیگنال‌دهی و فناوری‌های نوین ویرایش ژنوم گذار کرده است.

دوم؛ پاسخ‌نامه تحلیلی، مستند و تشریحی آزمون‌های ۱۴۰۴ و ۱۴۰۵: در این بخش، تک‌تک سوالات هفت‌گانه آزمون شامل «آمار و طرح آزمایش‌ها»، «ژنتیک پایه و سیتوژنتیک»، «اصلاح نباتات و ژنتیک کمی»، «بیوشیمی پیشرفته»، «کشت سلول و بافت گیاهی»، «ژنتیک مولکولی» و «مهندسی ژنتیک» در قالب یک ساختار آموزشی چهارمرحله‌ای تحلیل شده‌اند:

- تبیین موضوع و جایگاه علمی سؤال: مشخص ساختن دقیق شاخه علمی و مبحث هدف برای ایجاد انسجام ذهنی در خواننده؛
- کالبدشکافی تحلیلی و اثبات علمی: ارائه پاسخ تشریحی جامع با استناد به مبانی اثبات‌شده و مستندات معتبر بین‌المللی، همراه با حل گام‌به‌گام معادلات آماری و مسیرهای بیوشیمیایی؛
- تحلیل نقادانه سایر گزینه‌ها و کشف تله‌های تستی: واکاوی تک‌تک گزینه‌های نادرست و تشریح علل خطای علمی آن‌ها همراه با رمزگشایی از تله‌های طراحی، نظیر واژه‌گزینی‌های معکوس، ابهام در پیش‌فرض‌های مدل و دام‌های محاسباتی؛
- الگوریتم حل تستی و میان‌برهای زیر ۱۰ ثانیه: ارائه تکنیک‌های رد گزینه شهودی و روش‌های محاسبه ذهنی برای دستیابی به گزینه صحیح در کمترین زمان ممکن بدون نیاز به چرک‌نویس‌های وقت‌گیر.

سوم؛ درسنامه‌های مروری و خلاصه‌مباحث راهبردی: در پایان هر یک از دروس تخصصی، یک درسنامه مروری منسجم تدوین شده است که نقش جمع‌بندی نکات و مرور سریع را ایفا می‌کند. این درسنامه‌ها شامل فرمول‌نامه‌های فشرده، روابط و جدول ماتریسی درجات آزادی انواع طرح‌های آزمایشی پایه، فاکتوریل، اسپلیت پلات و زیرنمونه‌گیری، نسبت‌های اپیستاتیک کلاسیک، روابط تفکیک واریانس‌ها و وراثت‌پذیری، ده فرمان طلایی آزمون و واژه‌نامه‌های تخصصی دوزبانه (فارسی به انگلیسی و انگلیسی به فارسی) است که نیاز داوطلب را در مرور مفاهیم کلیدی برطرف می‌سازد.

چهارم؛ ضمایم شش‌گانه جامع و اطلس‌های تصویری: در انتهای اثر، مجموعه‌ای از ضمایم اختصاصی و پیوسته برای تثبیت آموخته‌ها فراهم آمده است:

۱. *فرمول‌نامه جامع و کدگذاری ماتریسی روابط محاسباتی:* گردآوری تمامی روابط محاسباتی پرکاربرد دروس آمار، بیومتری و ژنتیک کمی در قالب جداول منظم؛

۲. *اطلس تطبیقی روش‌های اصلاح گیاهان زراعی و سیستم‌های گرده‌افشانی:* مقایسه ماتریسی گیاهان زراعی شاخص از نظر سیستم زایشی، مکانیسم‌های نرعمیمی و روش‌های اصلاحی اختصاصی؛

۳. *راهنمای ساختارهای فضایی، نقشه رامچاندران و اسیدهای آمینه:* تفکیک ربع‌های مختصاتی زوایای دوجهی اسکلت پپتیدی و جدول تشخیصی ویژگی‌های بیوشیمیایی بیست اسید آمینه؛

۴. *جدول مرجع ترکیبات محیط‌های کشت و فیزیولوژی هورمون‌ها:* مقایسه غلظت نمک‌ها و تعادل یون‌های نیترات و آمونیوم در محیط‌های MS، ۵B و White به همراه سازوکار تنظیم‌کننده‌های رشد گیاهی؛

۵. *اطلس ابزارهای بیوتکنولوژی، کلونینگ و مهندسی ژنتیک:* خلاصه عملکرد آنزیم‌های دست‌ورزی، سازوکار پلاسمید Ti اگروباکتریوم و ژن‌های vir، سیستم‌های CRISPR/Cas9 و مقایسه تکنیک‌های بلاتینگ و ژنتیک معکوس؛

۶. *تحلیل تله‌های تستی پرتکرار و الگوهای رد‌گزینۀ فوری:* طبقه‌بندی راهبردی شگردهای طراحان آزمون و ارائه شیوه‌های شناسایی پاسخ صحیح در چند ثانیه.

پنجم؛ واژه‌نامه تخصصی و یکپارچه‌سازی مراجع استاندارد: در واپسین بخش کتاب، تمامی اصطلاحات، مفاهیم و تعاریف علمی پراکنده به صورت یکپارچه و به ترتیب الفبایی همراه با معادل‌های دقیق بین‌المللی گردآوری شده‌اند. همچنین فهرست منابع و مراجع رسمی بر اساس شیوه استنادی استاندارد APA تدوین گردیده است تا خواننده بتواند در صورت نیاز به متون مرجع اصلی دسترسی یابد.

مؤلفان امیدوارند این مجموعه جامع، شیوا و ساختاریافته بتواند افزون بر ایجاد آرامش خاطر و ارتقای چشمگیر توان علمی و مهارتی داوطلبان ورود به مقطع دکتری تخصصی، به عنوان یک دستنامه مرجع، سودمند و پایدار مورد استفاده دانشجویان کارشناسی ارشد، پژوهشگران و جامعه علمی کشاورزی و بیوتکنولوژی کشور قرار گیرد. بی‌تردید، دیدگاه‌ها و بازخوردهای نقادانه استادان گرامی، صاحب‌نظران ارجمند و دانشجویان فرهیخته راهگشای نویسندگان در غنای هرچه بیشتر ویرایش‌های آینده این اثر ماندگار خواهد بود.

گروه مؤلفین پردازش

آزمون دکترای بیوتکنولوژی کشاورزی ۱۴۰۴

آمار و طرح آزمایش ها



- ۱- چه زمانی برای مقایسه میانگین دو جامعه، از آزمون t استفاده می‌شود؟
(۱) توزیع دو جامعه نرمال و نمونه‌ها بزرگ باشد.
(۲) توزیع دو جامعه نرمال و نمونه‌ها کوچک باشد.
(۳) توزیع دو جامعه غیرنرمال و واریانس‌ها نامعلوم باشد.
(۴) توزیع دو جامعه نرمال، واریانس‌ها معلوم و نمونه کوچک باشد.
- ۲- اگر s^2 واریانس یک نمونه به حجم n از جامعه‌ای نرمال با واریانس σ^2 باشد، توزیع $\frac{(n-1)s^2}{\sigma^2}$ کدام است؟
(۱) نرمال (۲) نرمال استاندارد (۳) t -استیودنت (۴) کای اسکوئر
- ۳- از یک نمونه ۵ تایی، با حدود اعتماد ۹۵ درصد، میانگین وزن سیب در یک مجموعه، ۱۲۸/۹ تا ۱۷۱/۱ گرم به دست آمد. میانگین وزن سیب در این نمونه، چند گرم است؟
(۱) ۵۰ (۲) ۱۳۵ (۳) ۱۵۰ (۴) ۳۰۰
- ۴- اگر $\sigma_x^2 = ۴$ ، $\sigma_y^2 = ۳$ و $\sigma_{xy} = ۱$ باشد، واریانس متغیر Z با رابطه جبری $Z = \frac{1}{2}x + 2y - 3$ ، کدام است؟
(۱) ۸ (۲) ۱۱ (۳) ۱۵ (۴) ۱۸
- ۵- اگر در یک جدول توزیع فراوانی، فراوانی تجمعی عدد ۶۰ برابر ۰/۳ باشد، مفهوم آن چیست؟
(۱) ۳۰ درصد اعداد، حداکثر مساوی ۶۰ است. (۲) ۳۰ درصد اعداد، حداقل مساوی ۶۰ است.
(۳) ۷۰ درصد اعداد، حداکثر مساوی ۶۰ است. (۴) ۷۰ درصد اعداد، مساوی ۶۰ است.
- ۶- احتمال انجام خطای نوع اول و دوم، به ترتیب، در کدام آزمون بیشتر است؟
(۱) توکی - LSD (۲) شفه - دانت
(۳) دانت - LSD (۴) توکی - LSD



۷- در یک آزمایش به صورت کرت‌های خردشده با طرح پایه مربع لاتین مقدار درجه آزادی خطای اصلی و فرعی، به ترتیب،

کدام است؟ (تعداد سطوح فاکتور اصلی و فرعی به ترتیب ۴ و ۳ است.)

(۱) ۱۸ و ۶ (۲) ۹ و ۱۸

(۳) ۲۴ و ۶ (۴) ۹ و ۲۴

۸- بررسی چهار میزان کود آلی در سه مرحله رشد یک گیاه زراعی به صورت کرت‌های خردشده با طرح پایه بلوک

کامل تصادفی ($r = 4$) مدنظر است. مقدار درجات آزادی خطای فاکتور فرعی و کرت اصلی به ترتیب کدام است؟

(۱) ۳۶ و ۱۱ (۲) ۲۷ و ۱۵

(۳) ۲۷ و ۱۱ (۴) ۲۴ و ۱۵

۹- داده‌های زیر، در یک طرح بلوک‌های کامل تصادفی به دست آمده است. کدام مورد درست است؟

تیمار \ بلوک	تیمار		
	۱	۲	۳
۱	۲	۳	۴
۲	۴	۳	۲

(۱) $SS = \text{خطا} = SS$ کل

(۲) $SS = \text{خطا} = SS$ تیمار

(۳) $SS = \text{خطا} = SS$ بلوک

(۴) $SS = \text{خطا} = SS$ تیمار - SS بلوک

۱۰- در آزمایشی با ۶ تیمار، ۴ بلوک و ۳ نمونه، درجه آزادی خطای آزمایشی و نمونه‌برداری به ترتیب کدام است؟

(۱) ۱۵ و ۱۸ (۲) ۱۵ و ۴۸

(۳) ۱۸ و ۴۸ (۴) ۱۸ و ۲۴



موضوع سوال: این سؤال در ساختار دانش آمار زیستی و طرح آزمایش های کشاورزی، زیرمجموعه مبحث «آزمون های فرض پارامتری (Parametric Hypothesis Testing) و توزیع های نمونه برداری (Sampling Distributions)» طبقه بندی می شود. هدف طراح، سنجش درک عمیق داوطلب از فلسفه وجودی توزیع استیودنت (*Student's t - distribution*) در برابر توزیع نرمال استاندارد (*Z - distribution*) و پیش شرط های ریاضیاتی لازم برای کاربرد هریک در پژوهش های به نژادی گیاهی و آزمایش های گلخانه ای یا مزرعه ای است.

پاسخ تشریحی: در پژوهش های علوم گیاهی و بیوتکنولوژی، پژوهشگر اغلب با جامعه های آماری نامتناهی یا بسیار بزرگ (مانند کل بوته های یک توده بومی یا تمام کشت های بافتی یک لاین سلولی) سروکار دارد. در شرایط واقعی، واریانس واقعی جامعه (σ^2) نامعلوم است و بایستی با استفاده از واریانس نمونه (S^2) برآورد گردد. زمانی که حجم نمونه بزرگ باشد ($n \geq 30$)، بر اساس قضیه حد مرکزی (Central Limit Theorem - CLT)، برآورد واریانس نمونه بسیار به واریانس جامعه نزدیک بوده و توزیع آماره آزمون به سمت توزیع نرمال استاندارد میل می کند؛ بنابراین از آزمون Z استفاده می شود. اما چالش اصلی در نمونه های کوچک ($n < 30$) بروز می کند. در نمونه های کوچک، واریانس نمونه (S^2) خود یک متغیر تصادفی با نوسانات زیاد است. در سال ۱۹۰۸، ویلیام سیلی گاست (*William Sealy Gosset*) نشان داد که اگر واریانس جامعه نامعلوم باشد و از واریانس نمونه برای استانداردسازی استفاده شود، آماره حاصل دیگر توزیع نرمال نخواهد داشت؛ بلکه از توزیعی با دم های پهن تر (جهت پوشش خطای برآورد واریانس) به نام **توزیع t پیروی می کند.**

برای مقایسه میانگین دو جامعه مستقل با نمونه های کوچک، فرض صفر ($H_0: \mu_1 - \mu_2 = 0$) و آماره آزمون به صورت زیر فرمول بندی می شود:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

که در آن، S_p واریانس آمیخته (*Pooled Variance*) دو نمونه بوده و با رابطه زیر محاسبه می شود:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

درجه آزادی (df) این آزمون معادل $n_1 + n_2 - 2$ است. از نظر اثبات جبری، متغیر t از نسبت یک متغیر نرمال استاندارد به جذر یک متغیر کای-اسکوئر (χ^2) تقسیم بر درجه آزادی آن ساخته می شود. از این رو، **شرط اساسی و تخلف ناپذیر** برای اینکه نسبت فوق توزیع t داشته باشد، این است که صورت کسر (یعنی میانگین های نمونه ای و در نتیجه خود داده های جامعه های اصلی) حتماً **توزیع نرمال (Normal Distribution)** داشته باشند. در به نژادی گیاهی، هنگامی که عملکرد بذر یا تنفس سلولی چند گیاه جهش یافته جدید (مثلاً $n = 5$) را با رقم شاهد مقایسه می کنیم، به دلیل محدودیت مواد ژنتیکی یا هزینه بالای آزمون های مولکولی، ناگزیر به استفاده از نمونه های کوچک هستیم و شرط استناد به آزمون t ، نرمال بودن جامعه و نامعلوم بودن واریانس آن است.

بررسی سایر گزینه ها:

- **تحلیل گزینه «۱» (توزیع نرمال و نمونه ها بزرگ):** این گزینه نادرست است؛ زیرا زمانی که حجم نمونه ها بزرگ باشد ($n_1, n_2 \geq 30$)، حتی اگر واریانس جامعه نامعلوم باشد، واریانس نمونه ای (S^2) با تقریب بسیار بالایی جایگزین σ^2 شده و بر اساس قضیه حد مرکزی، توزیع آماره آزمون به توزیع نرمال استاندارد (Z) میل می کند. در آمار کلاسیک، برای نمونه های بزرگ از آزمون Z استفاده می گردد.
- **تحلیل گزینه «۳» (توزیع غیر نرمال و واریانس ها نامعلوم):** این گزینه یکی از دام های خطرناک است! اگر توزیع جامعه اصلی **غیر نرمال** باشد و حجم نمونه نیز **کوچک** باشد، پیش فرض بنیادی آزمون های پارامتری نقض شده و استفاده از آزمون t از نظر



علمی کاملاً بی‌اعتبار (Invalid) است. در چنین شرایطی در بیومتری و ژنتیک، بایستی از آزمون‌های ناپارامتری (Non-parametric Tests) مانند آزمون من-ویتنی (*Mann – Whitney U Test*) یا ویلکاکسون استفاده نمود.

- تحلیل گزینه «۴» (توزیع نرمال، واریانس معلوم و نمونه کوچک): این گزینه یک تله کلاسیک آماری است! بر اساس اصول اثبات‌شده آمار ریاضی، هرگاه واریانس جامعه‌های اصلی (σ_1^2, σ_r^2) معلوم و مشخص باشد، بدون توجه به حجم نمونه (حتی برای $n = 2$)، آماره آزمون دقیقاً از توزیع نرمال استاندارد (Z) پیروی می‌کند:

$$Z = \frac{(\bar{X}_1 - \bar{X}_r) - (\mu_1 - \mu_r)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_r^2}{n_r}}}$$

بنابراین هرگاه کلمه «واریانس معلوم» را در صورت مسئله دیدید، آزمون t به طور کامل منتفی شده و آزمون Z حاکمیت دارد. نکته تکمیلی: در صورتی که نمونه‌ها کوچک، توزیع نرمال و واریانس‌ها نامعلوم اما نابرابر ($\sigma_1^2 \neq \sigma_r^2$) باشند (عدم برقراری فرض همگنی واریانس‌ها یا *Homoscedasticity*)، از آزمون تعدیل‌شده t و *Welch – Satterthwaite t – test* با درجه آزادی تصحیح‌شده استفاده می‌شود.

راه حل تستی:

- تله طراح: تقابل دو زوج کلمه «معلوم / نامعلوم» و «کوچک / بزرگ». طراح با قرار دادن عبارت «واریانس معلوم» در گزینه «۴»، داوطلبانی که تنها به عبارت «نمونه کوچک» توجه می‌کنند را به دام می‌اندازد.
- میان‌بر تستی (الگوریتم رد گزینه در ۵ ثانیه):
 ۱. شرط حضور آزمون t ، «واریانس نامعلوم» است ← گزینه ۴ (واریانس معلوم) بلافاصله رد می‌شود.
 ۲. شرط حضور آزمون‌های پارامتری در نمونه کوچک، «توزیع نرمال» است ← گزینه ۳ (توزیع غیر نرمال) رد می‌شود.
 ۳. برای «نمونه‌های بزرگ»، آزمون Z اولویت دارد ← گزینه ۱ رد می‌شود.
 ۴. گزینه باقی‌مانده و صحیح قطعی: گزینه ۲.

گزینه ۴) .۲

موضوع سوال: این سؤال در معماری دانش آمار ریاضی و بیومتری، در قلمرو «توزیع‌های نمونه‌برداری (Sampling Distributions) و نظریه برآورد واریانس» قرار می‌گیرد. هدف طراح، ارزیابی تسلط داوطلب بر توزیع واریانس نمونه‌ای و یکی از مهم‌ترین قضایای آمار ریاضی (قضیه کوکران - *Cochran's Theorem*) است که ستون فقرات «تجزیه واریانس (ANOVA)» و آزمون‌های فرض در طرح آزمایش‌های کشاورزی و ژنتیک کمی را تشکیل می‌دهد.

پاسخ تشریحی: برای درک عمیق این مفهوم، باید اثبات جبری آن را از پایه‌ای‌ترین اصول آمار ریاضی بسط دهیم. فرض کنید متغیرهای تصادفی $X_1, X_2, \dots, X_n$ یک نمونه تصادفی مستقل به حجم n از جامعه‌ای با توزیع نرمال و پارامترهای میانگین μ و واریانس σ^2 باشند؛ یعنی $X_i \sim N(\mu, \sigma^2)$ اگر هریک از این مشاهدات را استاندارد کنیم، متغیر تصادفی جدید Z_i دارای توزیع نرمال استاندارد خواهد بود:

$$Z_i = \frac{X_i - \mu}{\sigma} \sim N(0, 1)$$

بر اساس تعریف ریاضی توزیع کای‌اسکوئر، «مجموع مربعات n متغیر تصادفی مستقل با توزیع نرمال استاندارد، دارای توزیع کای‌اسکوئر با n درجه آزادی است». بنابراین داریم:

$$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

اکنون با استفاده از یک اتحاد جبری، انحراف از میانگین جامعه ($X_i - \mu$) را به دو مؤلفه «انحراف از میانگین نمونه» ($X_i - \bar{X}$) و «انحراف میانگین نمونه از میانگین جامعه» ($\bar{X} - \mu$) تجزیه کرده و طرفین را به توان دو می‌رسانیم:

$$\sum_{i=1}^n (X_i - \mu)^r = \sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - \mu)]^r = \sum_{i=1}^n (X_i - \bar{X})^r + n(\bar{X} - \mu)^r$$

با تقسیم طرفین رابطه فوق بر واریانس جامعه (σ^r)، معادله به فرم زیر تبدیل می شود:

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^r = \frac{\sum_{i=1}^n (X_i - \bar{X})^r}{\sigma^r} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^r$$

با توجه به تعریف واریانس نمونه‌ای که به صورت $s^r = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^r$ می باشد، می توان صورت کسر اول در سمت راست را به صورت $(n-1)s^r$ بازنویسی کرد. از طرفی، می دانیم که عبارت داخل پرانتز دوم، همان استاندارد شده میانگین نمونه است و توزیع $N(0, 1)$ دارد؛ در نتیجه مربع آن دارای توزیع کای اسکوتر با یک درجه آزادی ($\chi^2_{(1)}$) است. طبق قضیه کوکران و ویژگی جمع پذیری توزیع های کای اسکوتر مستقل، برای برقراری تساوی، مؤلفه اول در سمت راست معادله باید الزاماً از توزیع کای اسکوتر با درجه آزادی معادل تفاضل درجات آزادی طرفین (یعنی $n-1$) پیروی کند:

$$\frac{(n-1)s^r}{\sigma^r} = \frac{\sum_{i=1}^n (X_i - \bar{X})^r}{\sigma^r} \sim \chi^2_{(n-1)}$$

تبیین کاربردی در ژنتیک کمی و به نژادی گیاهی: در اصلاح نباتات و ژنتیک کمی، این قضیه اساسی ترین ابزار برای برآورد فاصله اطمینان واریانس فنوتیپی (σ_P^r)، واریانس محیطی (σ_E^r) و همچنین آزمون همگنی واریانس خطاهای آزمایشی در طرح های کشاورزی (مانند آزمون بارتلت و آزمون لون) است. افزون بر این، در «تجزیه واریانس (ANOVA)»، تمام مجموع مربعات ($SS_{Treatment}, SS_{Block}, SS_{Error}$) زمانی که بر واریانس واقعی جامعه (σ^r) تقسیم شوند، از توزیع کای اسکوتر پیروی می کنند. نسبت دو متغیر کای اسکوتر مستقل که بر درجات آزادی شان تقسیم شده اند، توزیع F فیش را می سازد که برای آزمون معنی داری تیمارها، برآورد وراثت پذیری عمومی (h_b^r) و خصوصی (h_n^r) و اجزای واریانس ژنتیکی (افزایشی V_A و غالبیت V_D) به کار می رود.

بررسی سایر گزینه ها:

- **تحلیل گزینه «۱» (نرمال) و گزینه «۲» (نرمال استاندارد):** این گزینه ها کاملاً نادرست هستند. توزیع نرمال و نرمال استاندارد (Z) برای توزیع نمونه برداری میانگین ها ($\bar{X}$)، تفاضل میانگین ها یا برآورد ضرایب رگرسیونی (در نمونه های بزرگ یا واریانس معلوم) کاربرد دارند. واریانس یک کمیت ذاتاً غیرمنفی ($s^r \geq 0$) است؛ در حالی که توزیع نرمال متقارن بوده و مقادیر منفی را نیز در بر می گیرد (دامنه از $-\infty$ تا $+\infty$). بنابراین یک متغیر مجذور و مثبت مانند نسبت واریانس ها هرگز نمی تواند توزیع نرمال داشته باشد (مگر در نمونه های بی نهایت بزرگ به عنوان تقریب مجانبی).
 - **تحلیل گزینه «۳» (استیودنت):** توزیع استیودنت (t) زمانی حاصل می شود که یک متغیر نرمال استاندارد بر جذر یک متغیر کای اسکوتر (تقسیم بر درجه آزادی اش) تقسیم شود. از این توزیع برای آزمون فرض میانگین جامعه یا ضرایب رگرسیون زمانی که واریانس جامعه نامعلوم است استفاده می شود، نه برای توزیع خود واریانس نمونه ای. طراح با قرار دادن این گزینه، قصد داشته داوطلبانی را که کلمه «نمونه» یا وجود s^r را می بینند و فوراً به یاد آزمون t می افتند، دچار خطا کند.
- راه حل تستی:
- **تله طراح:** شباهت ظاهری کسر $\frac{(n-1)s^r}{\sigma^r}$ با فرمول های استاندارد سازی میانگین (Z و t). طراح روی عدم توجه داوطلب به «ماهیت مجذور بودن» واریانس حساب کرده است.
 - **میان بر تستی (تکنیک تحلیل ماهیت متغیر در ۳ ثانیه):**
 ۱. عبارت صورت سؤال حاوی واریانس نمونه (s^r) و واریانس جامعه (σ^r) است که هر دو توان دوم و همیشه مثبت هستند ← نسبت آن ها هرگز نمی تواند منفی باشد.
 ۲. توزیع های «نرمال»، «نرمال استاندارد» و «استیودنت» همگی متقارن بوده و شامل محور منفی ها هستند ← گزینه های ۱، ۲ و ۳ بلافاصله و بدون هیچ محاسباتی رد می شوند!



۳. تنها توزیع غیرمنفی و چوله به راست در گزینه‌ها که مربوط به مجموع مربعات است: گزینه ۴ (کای اسکوئر).

۳. گزینه ۳

۴. گزینه ۳

۵. گزینه ۱

موضوع سوال: این سؤال در معماری دانش آمار زیستی و بیومتری، در قلمرو «آمار توصیفی (Descriptive Statistics)، توابع توزیع تجمعی (Cumulative Distribution Functions - CDF) و مفهوم چندک‌ها (Quantiles/Percentiles)» قرار می‌گیرد. هدف طراح، ارزیابی درک دقیق و ریاضیاتی داوطلب از تعریف «فراوانی تجمعی نسبی» و تمایز آن با فراوانی مطلق، و همچنین توانایی تفسیر بیولوژیکی این شاخص در ارزیابی توده‌های گیاهی و جوامع اصلاحی است.

کالبدشکافی تحلیلی، اثبات ریاضی و تبیین مفهومی: در آمار توصیفی و پژوهش‌های کشاورزی، هنگامی که با مجموعه‌ای از داده‌های پیوسته یا گسسته (مانند ارتفاع بوته، عملکرد دانه، یا میزان مقاومت به بیماری در ۱۰۰۰ لاین گیاهی) سروکار داریم، داده‌ها به صورت صعودی (از کوچک‌ترین به بزرگ‌ترین مقدار) مرتب شده و در قالب جدول توزیع فراوانی سازمان‌دهی می‌شوند. در این ساختار، دو مفهوم کلیدی تعریف می‌گردد:

- فراوانی نسبی ($Relative\ Frequency - f_i$): نسبت تعداد مشاهدات یک طبقه یا یک مقدار خاص به کل مشاهدات (n) که با رابطه $f_i = \frac{n_i}{N}$ بیان می‌شود.

- فراوانی تجمعی نسبی ($Cumulative\ Relative\ Frequency - F(x)$): حاصل جمع فراوانی‌های نسبی تمامی مقادیر یا طبقه‌هایی که کوچک‌تر یا مساوی با یک مقدار مشخص (x) هستند. از دیدگاه آمار ریاضی و نظریه احتمال، تابع توزیع تجمعی (CDF) برای یک متغیر تصادفی X در نقطه $x = 60$ به صورت جبری زیر تعریف و فرمول‌بندی می‌شود:

$$F(X \leq 60) = P(X \leq 60) = \sum_{X_i \leq 60} f(X_i) = 0.3$$

معادله ریاضی فوق به زبان ساده بیان می‌دارد که احتمال یا فراوانی نسبی اینکه متغیر تصادفی X مقداری کمتر یا مساوی با ۶۰ (حداکثر مساوی ۶۰) اختیار کند، برابر با ۰.۳ (یا همان ۳۰ درصد) است. به عبارت دیگر، مقدار ۶۰ در این توزیع، دقیقاً صدک سی‌ام ($30^{th}\ Percentile$ یا P_{30}) جامعه را نشان می‌دهد.

تبیین کاربردی در به‌نژادی گیاهی و ژنتیک کمی: در برنامه‌های اصلاح نباتات و گزینش (Selection)، درک این مفهوم برای تعیین «شدت گزینش ($Selection\ Intensity - i$)» و «دیفرانسیل گزینش ($Selection\ Differential - S$)» حیاتی است. فرض کنید یک به‌نژادگر در حال بررسی میزان روغن در یک جمعیت پایه آفتابگردان است و عدد ۶۰ گرم را به عنوان آستانه در نظر می‌گیرد. وقتی فراوانی تجمعی عدد ۶۰ برابر با ۰.۳ (۳۰ درصد) باشد، به این معناست که ۳۰ درصد از بوته‌های جامعه عملکرد روغنی حداکثر تا ۶۰ گرم (کمتر یا مساوی ۶۰) دارند و ۷۰ درصد باقی‌مانده جامعه، عملکردی بیشتر از ۶۰ گرم تولید می‌کنند. اگر هدف اصلاح‌کننده، انتخاب بوته‌های برتر (گزینش جهت‌دار به سمت بالا) باشد، او باید آن ۳۰ درصد جامعه را حذف کرده و گزینش خود را روی ۷۰ درصد بالایی متمرکز نماید.

بررسی سایر گزینه‌ها:

- تحلیل گزینه «۲» (۳۰ درصد اعداد حداقل مساوی ۶۰ است): این گزینه یک خطای مفهومی ویرانگر است! واژه «حداقل مساوی ۶۰» به معنای مقادیر بزرگ‌تر یا مساوی ۶۰ ($X \geq 60$) است که مربوط به «فراوانی تجمعی مکمل یا معکوس» می‌باشد. بر اساس اصول احتمال، نسبت اعدادی که حداقل مساوی ۶۰ هستند برابر با $1 - 0.3 = 0.7$ (یعنی ۷۰ درصد) است، نه ۳۰ درصد. طراح با بازی با قیده‌های «حداقل» و «حداکثر»، داوطلبانی را که تعریف دقیق تابع CDF را به خاطر نسپردانند، به اشتباه می‌اندازد.



- تحلیل گزینه «۳» (۷۰ درصد اعداد حداکثر مساوی ۶۰ است): این گزینه ناشی از سردرگمی داوطلب بین صدک ۳۰ و صدک ۷۰ است. عدد ۰٫۳ معادل ۳۰ درصد است؛ بنابراین تخصیص نسبت ۷۰ درصد به عبارت «حداکثر مساوی ۶۰» از نظر محاسباتی و منطقی کاملاً با فرض سؤال در تضاد است.
- تحلیل گزینه «۴» (۷۰ درصد اعداد مساوی ۶۰ است): این گزینه یک دام مبتدیانه است! عبارت «مساوی ۶۰» ($X = ۶۰$) به فراوانی مطلق یا فراوانی نسبی ساده آن نقطه اشاره دارد، نه فراوانی تجمعی. علاوه بر این، نسبت تخصیص داده شده (۷۰ درصد) نیز با عدد ۰٫۳ (۳۰ درصد) هیچ‌گونه همخوانی ریاضی ندارد.

راه حل تستی:

- تله طراح: جابه‌جایی ظریف کلمات متضاد «حداکثر» و «حداقل» در گزینه‌های ۱ و ۲، و همچنین بازی با مکمل عدد ۰٫۳ (یعنی ۷۰ درصد) در گزینه‌های ۳ و ۴.
- میان‌بر تستی (تکنیک رمزگردانی در ۳ ثانیه):
 ۱. کلمه «تجمعی» در آمار همیشه از صفر شروع شده و به سمت بالا جمع می‌شود ← مفهوم آن همیشه «کوچک‌تر یا مساوی» (حداکثر) است.
 ۲. عدد ۰٫۳ در فرمت درصدی دقیقاً برابر با ۳۰ درصد است ← گزینه‌های ۳ و ۴ (که با ۷۰ درصد شروع شده‌اند) در کمتر از ۱ ثانیه خط زده می‌شوند!
 ۳. بین گزینه ۱ و ۲، کلمه کلیدی «تجمعی = حداکثر» را با گزینه ۱ تطبیق دهید ← انتخاب قطعی گزینه ۱.
 گزینه ۴.

موضوع سوال: این سؤال در ساختار دانش آمار زیستی و طرح آزمایش‌های کشاورزی، در قلمرو «آزمون‌های مقایسه چندگانه میانگین‌ها (Multiple Comparison Tests / Post-Hoc Tests) و فلسفه خطاهای آماری در تصمیم‌گیری» قرار می‌گیرد. هدف طراح، ارزیابی درک عمیق داوطلب از توازن (Trade-off) اجتناب‌ناپذیر میان خطای نوع اول (α) و خطای نوع دوم (β) در آزمون‌های چندمنه‌ای، و توانایی رتبه‌بندی این آزمون‌ها بر اساس میزان سخت‌گیری (Conservativeness) و توان آزمون (Power of Test) در پژوهش‌های به‌نژادی گیاهی است.

پاسخ تشریحی: زمانی که در تجزیه واریانس (ANOVA)، فرض صفر مبنی بر برابری میانگین تیمارها رد می‌شود ($F_{computed} \geq F_{table}$)، پژوهشگر باید مشخص کند اختلاف معنی‌دار دقیقاً بین کدام جفت از لاین‌ها یا ارقام کشاورزی وجود دارد. در این فرآیند تصمیم‌گیری، دو نوع خطای آماری رخ می‌دهد:

- **خطای نوع اول (α):** رد فرض صفر در حالی که فرض صفر در واقعیت درست است (یعنی اعلام تفاوت معنی‌دار بین دو رقم کشاورزی، در حالی که در واقعیت هیچ تفاوتی ندارند - False Positive).
- **خطای نوع دوم (β):** پذیرش فرض صفر در حالی که فرض صفر در واقعیت نادرست است (یعنی عدم اعلام تفاوت معنی‌دار بین دو رقم، در حالی که در واقعیت با یکدیگر متفاوت هستند - False Negative). مکمل این خطا، توان آزمون ($Power = 1 - \beta$) نامیده می‌شود.

در آزمون‌های مقایسه میانگین، چالش اصلی کنترل «نرخ خطای آزمایشی یا آزمایش‌محور ($Experiment - wise Error Rate - \alpha_{EW}$)» است. اگر ما t تیمار داشته باشیم، تعداد کل مقایسه‌های مستقل

به صورت $C = \frac{t(t-1)}{2}$ محاسبه می‌شود. اگر برای هر مقایسه، نرخ خطای مقایسه‌محور ($Comparison - wise Error Rate$) را روی $\alpha = ۰٫۰۵$ تنظیم کنیم، نرخ خطای کل آزمایش بر اساس رابطه احتمالاتی زیر شدیداً افزایش می‌یابد:

$$\alpha_{EW} = 1 - (1 - \alpha)^C$$

برای مثال، در مقایسه ۱۰ رقم گندم ($c = ۴۵$ مقایسه)، احتمال اینکه حداقل یک خطای نوع اول در کل آزمایش رخ دهد به حدود ۹۰ درصد ($\alpha_{EW} \approx ۰,۹۰$) می‌رسد! آزمون‌های مختلف برای حل این مشکل، رویکردها و روابط جبری متفاوتی برای محاسبه حداقل تفاوت معنی‌دار ($Critical\ Difference - CD$) به کار می‌گیرند:

۱. آزمون حداقل تفاوت معنی‌دار فیشر ($LSD - Fisher's\ Least\ Significant\ Difference$): این آزمون هیچ‌گونه کنترلی روی خطای آزمایش‌محور (α_{EW}) اعمال نمی‌کند و رابطه جبری آن دقیقاً مشابه آزمون t استیودنت است:

$$LSD = t_{\alpha, df_{error}} \sqrt{\frac{rMS_{error}}{r}}$$

به دلیل کوچک بودن مقدار بحرانی t ، مقدار عددی شاخص LSD بسیار کوچک است. در نتیجه، این آزمون بسیار ساده‌گیرانه (Liberal) عمل کرده، به راحتی فرض صفر را رد می‌کند و «بیشترین احتمال انجام خطای نوع اول (α)» را در میان تمامی آزمون‌های مقایسه میانگین دارد.

۲. آزمون تفاوت معنی‌دار درست توکی ($HSD - Tukey's\ Honestly\ Significant\ Difference$): آزمون توکی برای کنترل کامل خطای آزمایش‌محور طراحی شده است. این آزمون به جای توزیع t ، از توزیع «دامنه استیودنت‌شده» ($Studentized\ Range\ Distribution - q$) استفاده می‌کند که مقدار بحرانی به مراتب بزرگ‌تری دارد:

$$HSD = q_{\alpha, (t, df_{error})} \sqrt{\frac{MS_{error}}{r}}$$

بزرگ بودن شاخص HSD سبب می‌شود که این آزمون بسیار سخت‌گیر (Conservative) باشد و به سختی تفاوت‌ها را معنی‌دار اعلام کند. از نظر ریاضی، در هر نمونه آماری با حجم ثابت، خطای نوع اول و دوم رابطه معکوس دارند. وقتی آزمون توکی با سخت‌گیری تمام خطای نوع اول را به شدت کاهش می‌دهد، به طور اجتناب‌ناپذیری احتمال عدم تشخیص تفاوت‌های واقعی را بالا برده و «بیشترین احتمال انجام خطای نوع دوم (β)» (و کمترین توان آزمون) را در مقایسه با آزمون‌هایی نظیر LSD یا دانکن ایجاد می‌کند.

تبیین کاربردی در به‌نژادی گیاهی و ژنتیک کمی: انتخاب آزمون مناسب در اصلاح نباتات کاملاً به هدف پژوهشگر بستگی دارد:

- اگر پژوهشگر در مراحل اولیه یک برنامه اصلاحی (مانند ارزیابی ۵۰۰ ژنوتیپ وارداتی) باشد، هدف او «غربالگری اولیه» و از دست ندادن منابع ژنتیکی مفید است. در اینجا او ترجیح می‌دهد خطای نوع اول را به جان بخرد اما خطای نوع دوم (حذف اشتباهی یک ژنوتیپ برتر) رخ ندهد؛ بنابراین از آزمون‌های با توان بالا مانند LSD یا دانکن (Duncan) استفاده می‌کند.
- اگر پژوهشگر در مرحله معرفی رقم تجاری جدید به کشاورزان باشد، اعلام برتری دروغین یک رقم جدید نسبت به ارقام موجود (خطای نوع اول)، خسارت مالی و اعتباری سنگینی در پی دارد. در این شرایط، او به یک آزمون بسیار سخت‌گیر با کنترل شدید خطای نوع اول مانند توکی (Tukey) یا شفه (Scheffe) نیاز دارد.

بررسی سایر گزینه‌ها:

- تحلیل گزینه «۱» (توکی - LSD): این گزینه دقیقاً برعکس پاسخ صحیح است! طراح روی این موضوع حساب کرده که داوطلب می‌داند توکی و LSD دو قطب مخالف در آزمون‌های میانگین هستند، اما جایگاه آن‌ها را در تخصیص به خطای نوع اول و دوم قاطی می‌کند. توکی کمترین خطای نوع اول و LSD کمترین خطای نوع دوم را دارد.
- تحلیل گزینه «۲» (شفه - دانته): آزمون شفه ($Scheffe$) سخت‌گیرترین آزمون ممکن است که برای مقایسه‌های عمقی و تضادها (Contrasts) به کار می‌رود و کمترین خطای نوع اول (و بیشترین خطای نوع دوم) را دارد؛ بنابراین قرار دادن آن به عنوان دارنده بیشترین خطای نوع اول کاملاً غلط است. آزمون دانته ($Dunnett$) نیز یک آزمون اختصاصی برای مقایسه چندین تیمار صرفاً با یک «تیمار شاهد» (Control) است و از نظر سخت‌گیری در حد وسط قرار دارد.
- تحلیل گزینه «۳» (دانته - LSD): همان‌طور که اثبات شد، LSD بیشترین احتمال خطای نوع اول را دارد؛ قرار دادن آن در بخش دوم (بیشترین خطای نوع دوم) از نظر آماری کاملاً بی‌اساس است. دانته نیز نسبت به LSD خطای نوع اول کمتری دارد.



راه حل تستی:

- **تله طراح:** استفاده از کلمات «به ترتیب» و قرار دادن جفت‌های معکوس در گزینه‌های ۱ و ۴. داوطلب با عجله ممکن است گزینه ۱ را به جای ۴ علامت بزند.
 - **میان‌بر تستی (قانون الاکلنگ خطاها در ۵ ثانیه):**
 ۱. آزمون **LSD** (ساده‌گیرترین آزمون) ← همیشه به دنبال دردرس و رد کردن فرض صفر است ← **بیشترین خطای نوع اول** (α). (با این قانون، گزینه‌های ۱، ۲ و ۳ در کسری از ثانیه حذف می‌شوند).
 ۲. آزمون **توکی** (آزمون محتاط و سخت‌گیر) ← از ترس خطای نوع اول، آن‌قدر دست به عصا راه می‌رود که تفاوت‌های واقعی را هم نمی‌بیند ← **بیشترین خطای نوع دوم** (β).
 ۳. تقابل «**LSD** - توکی» در گزینه ۴ می‌درخشد ← انتخاب قطعی و بدون تردید.
- گزینه ۳. ۷

موضوع سوال: این سؤال در ساختار دانش آمار زیستی و طرح آزمایش‌های کشاورزی، زیرمجموعه مبحث «طرح کرت‌های خردشده (**Split-Plot Design - SPPD**) با طرح‌های پایه مختلف و جبری‌سازی درجات آزادی منابع تغییر» طبقه‌بندی می‌شود. هدف طراح، سنجش تسلط داوطلب بر نحوه ادغام طرح‌های پایه (مانند مربع لاتین **LS**) با ساختار کرت‌های خردشده، و توانایی تفکیک و محاسبه جبری درجات آزادی برای خطای کرت اصلی (**Error a**) و خطای کرت فرعی (**Error b**) است.

پاسخ تشریحی: در پژوهش‌های کشاورزی و اصلاح نباتات، زمانی که یک آزمایش به صورت کرت‌های خردشده (اسپلیت پلات) پیاده‌سازی می‌شود، فاکتور اصلی (**A**) در قالب یکی از طرح‌های پایه (کاملاً تصادفی **CRD**، بلوک‌های کامل تصادفی **RCBD** یا مربع لاتین **LSD**) قرار می‌گیرد و فاکتور فرعی (**B**) به صورت تصادفی درون کرت‌های اصلی توزیع می‌گردد. در این سؤال، طرح پایه، مربع لاتین (**Latin Square Design**) است. ویژگی ذاتی و منحصر به فرد طرح مربع لاتین این است که تعداد تکرارها، تعداد ردیف‌ها، تعداد ستون‌ها و تعداد سطوح فاکتور اصلی همگی با یکدیگر برابرند ($r = Row = Column = a$). با توجه به اطلاعات مسئله:

- تعداد سطوح فاکتور اصلی (قرار گرفته در کرت اصلی): $a = 4$
 - تعداد تکرارها (تعداد ردیف‌ها و ستون‌های مربع لاتین): $r = a = 4$
 - تعداد سطوح فاکتور فرعی (قرار گرفته در کرت فرعی): $b = 3$
 - تعداد کل واحدهای آزمایشی (کرت‌های فرعی): $N = r \times a \times b = 4^r \times b = 4^4 \times 3 = 4^8$ واحد.
- برای محاسبه دقیق درجات آزادی خطای اصلی و فرعی، ساختار آنالیز واریانس (**ANOVA**) را به دو بخش کاملاً مستقل تجزیه می‌کنیم:

الف) بخش کرت اصلی (Main-Plot Analysis):

در این بخش، فاکتور اصلی (**A**) با ۴ سطح در یک طرح مربع لاتین 4×4 آنالیز می‌شود. درجات آزادی کل این بخش برابر با تعداد کل کرت‌های اصلی منهای یک ($a^r - 1 = 16 - 1 = 15$) است. منابع تغییرات و درجات آزادی مربوطه عبارت‌اند از:

- ردیف‌ها (**Rows**): $r - 1 = a - 1 = 4 - 1 = 3$
- ستون‌ها (**Columns**): $r - 1 = a - 1 = 4 - 1 = 3$
- فاکتور اصلی (**Factor A**): $a - 1 = 4 - 1 = 3$
- **خطای اصلی** (**Error a**): بر اساس اصل جبری، درجه آزادی خطای مربع لاتین از تفاضل درجات آزادی فوق یا از رابطه مستقیم $(a - 1)(a - r)$ به دست می‌آید:

$$df_{Error a} = (a - 1)(a - r) = (r - 1)(r - r) = (4 - 1)(4 - 4) = 3 \times 0 = 0$$

ب) بخش کرت فرعی (Sub-Plot Analysis):

در این بخش، اثرات فاکتور فرعی (**B**) و برهم‌کنش آن با فاکتور اصلی (**A × B**) در داخل کرت‌های اصلی ارزیابی می‌شود:



- فاکتور فرعی ($Factor B$): $b - 1 = 3 - 1 = r$
- اثر متقابل ($A \times B$): $(a - 1)(b - 1) = (4 - 1)(3 - 1) = 3 \times 2 = 6$
- خطای فرعی ($Error b$): این خطا نشان‌دهنده برهم‌کنش فاکتور فرعی با ردیف‌ها، ستون‌ها و خطای اصلی است. برای محاسبه آن، می‌توان از رابطه مستقیم ضربی یا روش تفاضل از کل استفاده کرد. رابطه مستقیم جبری برای خطای فرعی در طرح پایه مربع لاتین به صورت زیر است:

$$df_{Error b} = a(a - 1)(b - 1) = r(r - 1)(b - 1)$$

با جاگذاری مقادیر در این فرمول داریم:

$$df_{Error b} = 4 \times (4 - 1) \times (3 - 1) = 4 \times 3 \times 2 = 24$$

جهت اثبات صحت محاسبات از طریق روش تفاضل (کنترل کیفیت ریاضیاتی):

$$N - 1 = 48 - 1 = 47 \quad (Total)$$

درجه آزادی کل

مجموع درجات آزادی محاسبه شده (به جز خطای فرعی):

$$df_{Rows} + df_{Cols} + df_A + df_{Error a} + df_B + df_{A \times B} = 2 + 2 + 3 + 6 + 2 + 6 = 21$$

$$df_{Error b} = 47 - 21 = 26$$

خطای فرعی از روش تفاضل: $df_{Error b} = 47 - 21 = 26$

بنابراین، مقادیر درجه آزادی خطای اصلی و فرعی به ترتیب برابر با ۶ و ۲۴ است.

تبیین کاربردی در به‌نژادی گیاهی و ژنتیک کمی: چرا یک به‌نژادگر از طرح کرت‌های خردشده با پایه مربع لاتین استفاده می‌کند؟ در آزمایش‌های مزرعه‌ای، گاهی گرادیان حاصلخیزی خاک، رطوبت یا شیب زمین در دو جهت عمود بر هم تغییر می‌کند. اگر فاکتور اصلی ما یک عملیات پرهزینه یا نیازمند ماشین‌آلات سنگین (مانند ۴ روش مختلف شخم یا ۴ رژیم آبیاری) باشد، اجرای آن در کرت‌های کوچک غیرممکن است. به‌نژادگر این فاکتور را در ۴ کرت اصلی در قالب یک مربع لاتین ۴×۴ پیاده‌سازی می‌کند تا شیب دوطرفه مزرعه کنترل شود. سپس درون هر رژیم آبیاری، ۳ رقم مختلف گندم (فاکتور فرعی B) را می‌کارد. با این کار، دقت آزمون برای مقایسه ارقام و اثر متقابل رقم × آبیاری (که بیشترین اهمیت را در معرفی رقم دارد و در خطای فرعی با ۲۴ درجه آزادی قرار می‌گیرد) به شدت افزایش می‌یابد.

راه حل تستی:

- **تله طراح:** ادغام نام «کرت‌های خردشده» با «مربع لاتین»! داوطلبان معمولاً به دیدن کرت‌های خردشده با پایه بلوک ($RCBD$) عادت دارند و ناخودآگاه فرمول $(r - 1)(a - 1) = 9$ را برای خطای اصلی مینویسند.
- **میان‌بر تستی (الگوریتم حل ۴ ثانیه‌ای با چشم):**

۱. به محض دیدن کلمات «مربع لاتین» و فاکتور اصلی $a = 4$ ، فرمول اختصاصی خطای مربع لاتین را در ذهن مرور کنید:

$$(a - 1)(a - r)$$

۲. ضرب ذهنی سریع: $(4 - 1) \times (4 - 3) = 3 \times 1 = 3$. (با همین یک ضرب ساده، گزینه‌های ۲ و ۴ بلافاصله حذف می‌شوند!)

۳. برای خطای فرعی، کافی است خطای اصلی (که ۶ شد) را در تعداد سطوح فاکتور اصلی ($a = 4$) و فاکتور فرعی منهای یک ($b - 1 = r$) ضرب کنید:

$$df_{Error b} = df_{Error a} \times \frac{a(b - 1)}{a - r} \Rightarrow \text{روش ساده‌تر: } a \times (a - 1) \times (b - 1) = 4 \times 3 \times 2 = 24$$

۴. گزینه صحیح قطعی: گزینه ۳ (۶ و ۲۴).

۸. گزینه ۴

موضوع سوال: این سؤال در قلمرو «آزمایش‌های چندعاملی، طرح کرت‌های خردشده ($Split-Plot Design$) و ساختار درجات آزادی (df) آنالیز واریانس» قرار دارد. هوشمندی طراح در این پرسش، استفاده از واژه ساختاری «کرت اصلی» به



جای «خطای کرت اصلی» است تا تفکیک دقیق بین درجات آزادی مؤلفه‌های کلان آزمایش و خطاهای فرعی را در یک پژوهش به‌نژادی-زراعی به چالش بکشد.

پاسخ تشریحی: برای حل دقیق این مسئله، ابتدا باید مشخصات و پارامترهای عددی داده‌شده در صورت سؤال را استخراج و طبقه‌بندی کنیم:

- تعداد تکرارها یا بلوک‌ها: $r = 4$
- تعداد سطوح فاکتور اصلی (کود آلی که در کرت‌های اصلی قرار می‌گیرد): $a = 4$
- تعداد سطوح فاکتور فرعی (مراحل رشد گیاه که در کرت‌های فرعی قرار می‌گیرد): $b = 3$ *Tram*

اکنون به محاسبه درجات آزادی درخواستی بر اساس مراجع آمار ریاضی می‌پردازیم:

۱. محاسبه درجات آزادی کرت اصلی (Main Plot):

در ساختار طرح کرت‌های خردشده، بخش کرت اصلی خود به عنوان یک آزمایش مستقل در قالب طرح پایه (در اینجا بلوک‌های کامل تصادفی *RCBD*) عمل می‌کند. مجموع کرت‌های اصلی در کل آزمایش برابر با ضرب تعداد بلوک‌ها در تعداد سطوح فاکتور اصلی ($r \times a$) است. بنابراین، درجه آزادی کل بخش کرت اصلی از رابطه بنیادی زیر به دست می‌آید:

$$df_{\text{Main Plot Total}} = (r \times a) - 1$$

با جاگذاری مقادیر عددی داریم:

$$df_{\text{Main Plot Total}} = (4 \times 4) - 1 = 16 - 1 = 15$$

اثبات از روش هم‌پوشانی منابع تغییر: درجه آزادی کل کرت اصلی، حاصل جمع درجات آزادی بلوک ($r - 1$)، فاکتور اصلی ($a - 1$) و خطای اصلی یا همان خطای $a(r - 1)$ است:

$$df_{\text{Main Plot Total}} = (4 - 1) + (4 - 1) + (4 - 1)(4 - 1) = 3 + 3 + (3 \times 3) = 15$$

۲. محاسبه درجات آزادی خطای فاکتور فرعی (خطای b یا Error b):

خطای فاکتور فرعی حاصل از برهم‌کنش فاکتور فرعی با بلوک‌ها و همچنین اثر متقابل فاکتور فرعی با خطای اصلی است. رابطه جبری استاندارد برای محاسبه درجه آزادی خطای فرعی در طرح پایه بلوک تصادفی به صورت زیر فرمول‌بندی می‌شود:

$$df_{\text{Error b}} = a(r - 1)(b - 1)$$

با جاگذاری پارامترهای آزمایش در این فرمول، مقدار آن به دست می‌آید:

$$df_{\text{Error b}} = 4 \times (4 - 1) \times (3 - 1)$$

$$df_{\text{Error b}} = 4 \times 3 \times 2 = 24$$

از آنجا که طراح سؤال، مقدار درجات آزادی «خطای فاکتور فرعی» و «کرت اصلی» را به ترتیب خواسته است، جفت عددی حاصل برابر با ۲۴ و ۱۵ خواهد بود که دقیقاً با گزینه ۴ انطباق دارد.

تبیین کاربردی در به‌نژادی گیاهی و بیومتری: طرح کرت‌های خردشده زمانی انتخاب می‌شود که اعمال یکی از فاکتورها (مثل کود دهی یا آبیاری) به فضایی بزرگ نیاز داشته باشد، اما فاکتور دیگر (مثل مراحل رشد یا ارقام) را بتوان در فضایی کوچک‌تر بررسی کرد. در این آزمایش، کل مزرعه به ۱۶ کرت اصلی تقسیم شده (با ۱۵ درجه آزادی کل) و سپس هر کرت اصلی به ۳ کرت فرعی خرد شده است تا در مجموع، ۴۸ واحد آزمایشی با ۴۷ درجه آزادی کل شکل بگیرد. بالا بودن درجه آزادی خطای فرعی (۲۴) نسبت به خطای اصلی (۹)، نشان‌دهنده دقت بالاتر و توان آماری بیشتر این طرح در ارزیابی فاکتور فرعی (مراحل رشد) و اثر متقابل کود × مرحله رشد است.

راه حل تستی:

- **تله طراح:** استفاده از عبارت «کرت اصلی» به جای «خطای کرت اصلی»؛ اگر داوطلب خطای اصلی را حساب می‌کرد به عدد ۹ می‌رسید که در هیچ گزینه‌ای نبود و دچار سردرگمی می‌شد.

- **میان‌بر تستی (الگوریتم حل ۵ ثانیه‌ای):**

۱. سریعاً فرمول خطای فرعی (*Error b*) را بنویسید: $a(r - 1)(b - 1)$.

۲. اعداد را ذهنی جاگذاری کنید: $4 \times 3 \times 2 = 24$.



۳. به گزینه‌ها نگاه کنید؛ تنها و تنها گزینه ۴ است که با عدد ۲۴ شروع می‌شود!

۴. بدون نیاز به محاسبه بخش دوم (کرت اصلی)، گزینه ۴ را به عنوان پاسخ قطعی علامت بزنید.

۹. گزینه ۱)

موضوع سوال: این سؤال در ساختار دانش آمار زیستی و طرح آزمایش‌های کشاورزی، در قلمرو «اصول تجزیه واریانس (ANOVA) در طرح بلوک‌های کامل تصادفی (RCBD)، تجزیه متعامد مجموع مربعات (Orthogonal Partitioning of Sum of Squares) و رفتار جبری منابع تغییر» قرار دارد. هدف طراح، ارزیابی درک شهودی و ریاضیاتی داوطلب از مفهوم «واریانس بین گروهی (Between-group Variance)» و این قضیه است که هرگاه میانگین تمامی تیمارها و همچنین میانگین تمامی بلوک‌ها با یکدیگر برابر باشند، سهم آن‌ها در تنوع کل آزمایشی صفر شده و تمام تنوع مشاهده شده در داده‌ها به «خطای آزمایشی» اختصاص می‌یابد.

پاسخ تشریحی: برای درک عمیق این پدیده، ابتدا جدول داده‌ها را بازسازی کرده و پارامترهای پایه‌ای آزمایش را بر اساس اصول جبر ماتریسی و آمار حیاتی استخراج می‌کنیم. فرض کنید Y_{ij} نشان‌دهنده مشاهده مربوط به تیمار i ($i = 1, r, 3$) در بلوک j ($j = 1, r$) باشد:

• تعداد تیمارها (t): ۳

• تعداد بلوک‌ها یا تکرارها (r): ۳

• تعداد کل مشاهدات (N): $t \times r = 3 \times 3 = 9$

اکنون مجموع و میانگین هر ردیف (بلوک‌ها) و هر ستون (تیمارها) را محاسبه می‌کنیم:

• مجموع مشاهدات تیمارها ($Y_{i.}$):

$$Y_{1.} = r + 4 = 6 \Rightarrow \bar{Y}_{1.} = 2 \quad \text{تیمار ۱}$$

$$Y_{r.} = 3 + 3 = 6 \Rightarrow \bar{Y}_{r.} = 2 \quad \text{تیمار ۲}$$

$$Y_{3.} = 4 + r = 6 \Rightarrow \bar{Y}_{3.} = 2 \quad \text{تیمار ۳}$$

• مجموع مشاهدات بلوک‌ها ($Y_{.j}$):

$$Y_{.1} = r + 3 + 4 = 9 \Rightarrow \bar{Y}_{.1} = 3 \quad \text{بلوک ۱}$$

$$Y_{.r} = 4 + 3 + r = 9 \Rightarrow \bar{Y}_{.r} = 3 \quad \text{بلوک ۲}$$

• مجموع کل (G یا $Y_{..}$): $G = 6 + 6 + 6 = 9 + 9 = 18 \Rightarrow \bar{Y}_{..} = 2$ میانگین کل ($\bar{Y}_{..}$) برابر با ۲

• ضریب تصحیح (CF - Correction Factor):

$$CF = \frac{G^2}{N} = \frac{18^2}{9} = \frac{324}{9} = 36$$

اکنون بر اساس روابط بنیادی آنالیز واریانس در طرح بلوک‌های کامل تصادفی، اجزای مجموع مربعات را محاسبه می‌کنیم:

۱. مجموع مربعات کل (SS_{Total}):

$$SS_{Total} = \sum_{i=1}^t \sum_{j=1}^r Y_{ij}^2 - CF$$

$$SS_{Total} = (r^2 + 3^2 + 4^2 + 4^2 + 3^2 + r^2) - 36 = (4 + 9 + 16 + 16 + 9 + 4) - 36 = 58 - 36 = 22$$

۲. مجموع مربعات تیمار ($SS_{Treatment}$):

$$SS_{Treatment} = \sum_{i=1}^t \frac{Y_{i.}^2}{r} - CF = \frac{6^2 + 6^2 + 6^2}{3} - 36 = \frac{36 + 36 + 36}{3} - 36 = \frac{108}{3} - 36 = 36 - 36 = 0$$



اثبات مفهومی: از آنجا که مجموع مشاهدات و میانگین هر سه تیمار با یکدیگر کاملاً برابرند ($\bar{Y}_1 = \bar{Y}_r = \bar{Y}_c = 3$)، هیچ‌گونه اختلافی بین میانگین تیمارها و میانگین کل وجود ندارد؛ بنابراین واریانس بین تیمارها و در نتیجه $SS_{Treatment}$ دقیقاً برابر با صفر است.

۳. مجموع مربعات بلوک (SS_{Block}):

$$SS_{Block} = \sum_{j=1}^r \frac{Y_j^2}{t} - CF = \frac{9^2 + 9^2}{3} - 54 = \frac{81 + 81}{3} - 54 = \frac{162}{3} - 54 = 54 - 54 = 0$$

اثبات مفهومی: به طور مشابه، چون مجموع و میانگین هر دو بلوک با یکدیگر برابرند ($\bar{Y}_1 = \bar{Y}_r = 3$)، هیچ‌گونه شیب محیطی یا تنوعی بین بلوک‌ها وجود ندارد؛ در نتیجه SS_{Block} نیز دقیقاً برابر با صفر است.

۴. مجموع مربعات خطا (SS_{Error}):

در طرح بلوک‌های کامل تصادفی ($RCBD$)، بر اساس مدل خطی جمع‌پذیر ($Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$)، رابطه جبری تفکیک مجموع مربعات به صورت زیر است:

$$SS_{Total} = SS_{Treatment} + SS_{Block} + SS_{Error}$$

با جاگذاری مقادیر به دست آمده در این اتحاد جبری داریم:

$$4 = 0 + 0 + SS_{Error} \Rightarrow SS_{Error} = 4$$

بنابراین، از نظر جبری و محاسباتی به وضوح اثبات می‌شود که $SS_{Error} = SS_{Total} = 4$ و گزینه ۱ پاسخ قطعی است.

راه حل تستی:

• **تله طراح:** ارائه اعداد ظاهراً پیچیده و وادار کردن داوطلب به محاسبات طولانی فرمول‌های SS_{Total} ، $SS_{Treatment}$ و SS_{Block} در شرایط استرس و کمبود زمان آزمون دکتری.

• **میان‌بر تستی (الگوریتم حل بازرسی چشمی در ۵ ثانیه):**

۱. جمع سریع ستون‌ها (تیمارها): ستون اول ($4 + 4 = 8$)، ستون دوم ($3 + 3 = 6$)، ستون سوم ($4 + 4 = 8$). چون همه جمع‌ها برابرند $\Rightarrow SS_{تیمار} = 0$.

۲. جمع سریع ردیف‌ها (بلوک‌ها): ردیف اول ($4 + 3 + 4 = 11$)، ردیف دوم ($4 + 3 + 4 = 11$). چون جمع ردیف‌ها برابرند $\Rightarrow SS_{بلوک} = 0$.

۳. قانون بقای واریانس: وقتی تیمار و بلوک هیچ سهمی از تنوع ندارند (۰)، پس کل تنوع موجود در جدول به خطا می‌رسد:

$$SS_{خطا} = SS_{Total} \Rightarrow SS_{خطا} = 0 + 0 + SS_{خطا} \Rightarrow SS_{خطا} = SS_{Total} = 4$$

۴. انتخاب فوری و قطعی گزینه ۱ بدون نوشتن حتی یک فرمول روی کاغذ!

۱۰. گزینه ۲)

موضوع سوال: این سؤال در قلمرو «طرح‌های آماری با زیرنمونه‌گیری (Sub-sampling in Experimental Designs / Nested or Hierarchical ANOVA) و تفکیک اجزای خطا در آزمایش‌های کشاورزی» قرار دارد. هدف طراح، سنجش توانایی داوطلب در تمایز مفهومی و محاسباتی میان «خطای آزمایشی (Experimental Error)» و «خطای نمونه‌برداری (Sampling Error / Observational Error)» در یک طرح بلوک‌های کامل تصادفی ($RCBD$) است.

پاسخ تشریحی: در پژوهش‌های کشاورزی و به‌نژادی، غالباً اندازه‌گیری یک صفت (مانند ارتفاع بوته، عملکرد تک‌بوته یا درصد روغن) بر روی کل یک واحد آزمایشی (کرت یا پلات) امکان‌پذیر نیست یا هزینه بالایی دارد. در این شرایط، پژوهشگر از هر واحد آزمایشی (مثلاً هر کرت)، تعداد s بوته یا نمونه تصادفی را انتخاب کرده و صفت را روی آن‌ها اندازه‌گیری می‌کند. به این تکنیک، «نمونه‌برداری در درون واحد آزمایشی» یا «زیرنمونه‌گیری (Sub-sampling)» می‌گویند. در این حالت، تنوع کل داده‌ها به دو منبع خطای مجزا تجزیه می‌شود:



۱. خطای آزمایشی (Experimental Error): نوسانات و تفاوت‌های تصادفی میان واحدهای آزمایشی (کرت‌ها) که تیمار یکسانی دریافت کرده‌اند.

۲. خطای نمونه‌برداری (Sampling Error): نوسانات و تفاوت‌های تصادفی میان نمونه‌های (بوته‌های) درون یک واحد آزمایشی (کرت) مشخص.
با توجه به اطلاعات مسئله:

- تعداد تیمارها (ارقام یا تیمارهای زراعی): $t = ۶$
 - تعداد بلوک‌ها یا تکرارها: $r = ۴$
 - تعداد نمونه‌ها (بوته‌های اندازه‌گیری شده) در هر کرت: $s = ۳$
 - تعداد کل کرت‌های آزمایشی: $N_{plots} = t \times r = ۶ \times ۴ = ۲۴$ کرت.
 - تعداد کل مشاهدات (داده‌های جمع‌آوری شده): $N_{total} = t \times r \times s = ۶ \times ۴ \times ۳ = ۷۲$ مشاهده.
- اکنون به محاسبه دقیق جبری درجات آزادی هر دو خطا می‌پردازیم:

الف) محاسبه درجه آزادی خطای آزمایشی (Experimental Error df):

خطای آزمایشی دقیقاً بر اساس تعداد کرت‌ها (تیمارها و بلوک‌ها) و صرف‌نظر از تعداد نمونه‌های درون کرت محاسبه می‌شود. از آنجا که طرح پایه آزمایش، بلوک‌های کامل تصادفی (RCBD) است، رابطه جبری درجه آزادی خطای آزمایشی از فرمول استاندارد طرح بلوک به دست می‌آید:

$$df_{Exp. Error} = (r - 1)(t - 1)$$

با جاگذاری مقادیر صورت سؤال داریم:

$$df_{Exp. Error} = (۴ - 1)(۶ - 1) = ۳ \times ۵ = ۱۵$$

ب) محاسبه درجه آزادی خطای نمونه‌برداری (Sampling Error df):

خطای نمونه‌برداری، تنوع درون هر کرت را نشان می‌دهد. درون هر کرت آزمایشی، تعداد s نمونه گرفته شده است؛ بنابراین هر کرت به اندازه $(s - 1)$ درجه آزادی برای خطای نمونه‌برداری ایجاد می‌کند. از آنجا که ما در کل مزرعه تعداد $t \times r$ کرت داریم، مجموع درجات آزادی خطای نمونه‌برداری از ضرب تعداد کل کرت‌ها در درجه آزادی درون هر کرت حاصل می‌شود:

$$df_{Samp. Error} = t \times r \times (s - 1)$$

با جاگذاری مقادیر عددی داریم:

$$df_{Samp. Error} = ۶ \times ۴ \times (۳ - 1) = ۴۸$$

بنابراین، مقدار درجات آزادی خطای آزمایشی و نمونه‌برداری به ترتیب برابر با ۱۵ و ۴۸ است که دقیقاً با گزینه ۲ انطباق دارد.

تبیین کاربردی در به‌نژادی گیاهی و بیومتری: تمایز بین این دو خطا در ژنتیک کمی چه اهمیتی دارد؟ فرض کنید یک به‌نژادگر قصد دارد عملکرد دانه ۶ رقم جدید گندم را در ۴ بلوک ارزیابی کند و در هر کرت، ۳ بوته را به تصادف برداشت می‌کند. اگر واریانس خطای نمونه‌برداری ($MS_{Sampling}$) بسیار بالا باشد، نشان‌دهنده ناهمگنی ژنتیکی درون رقم (مثلاً در توده‌های محلی یا ارقام آزادگرده‌افشان) یا عدم یکنواختی بافت خاک درون کرت است. در مقابل، واریانس خطای آزمایشی ($MS_{Exp. Error}$) تنوع بین کرت‌هایی که یک رقم را دریافت کرده‌اند، نشان می‌دهد. بر اساس اصول بیومتری، برای آزمون معنی‌داری میانگین تیمارها (ارقام)، آماره F الزاماً باید از تقسیم میانگین مربعات تیمار بر میانگین مربعات «خطای آزمایشی» ($F = \frac{MS_{Treatment}}{MS_{Exp. Error}}$) محاسبه شود، نه خطای نمونه‌برداری! استفاده اشتباه از خطای نمونه‌برداری (که درجه آزادی بالا و معمولاً میانگین مربعات کوچکی دارد)، باعث ایجاد یک آماره F کاذب و وقوع شدید خطای نوع اول (α) می‌گردد.

راه حل تستی:

- **تله طراح:** ادغام مفاهیم «تکرار / بلوک» و «نمونه / زیرنمونه». برخی داوطلبان عدد $s = ۳$ را به عنوان تکرار در نظر گرفته و بلوک را نادیده می‌گیرند.

- میان‌بر تستی (الگوریتم حل ۵ ثانیه‌ای بدون چک‌نویس):

۱. محاسبه خطای آزمایشی (مثل یک طرح بلوک ساده با ۶ تیمار و ۴ بلوک):

$$df_{Exp} = (6 - 1) \times (4 - 1) = 5 \times 3 = 15$$

(با همین یک ضرب ساده، گزینه‌های ۳ و ۴ که با ۱۸ شروع شده‌اند فوراً خط می‌خورند!)

۲. محاسبه خطای نمونه‌برداری (کل کرت‌ها ضرب در نمونه‌های منهای یک):

$$6 \times 4 = 24$$

تعداد کل کرت‌ها: ۲۴

هر کرت ۳ نمونه دارد، پس سهم هر کرت در خطای نمونه‌برداری: $3 - 1 = 2$

$$24 \times 2 = 48$$

ضرب نهایی: ۴۸

۳. انتخاب قطعی و سریع گزینه ۲ (۱۵ و ۴۸).

درسنامه مروری آمار و طرح آزمایش‌ها

بخش اول: فرمول‌نامه فشرده و طبقه‌بندی‌شده بیومتری و طرح آزمایش‌ها

در این فرمول‌نامه، تمامی روابط محوری که ستون فقرات سؤالات آزمون دکتری را تشکیل می‌دهند، با سینتکس استاندارد LaTeX و با تفکیک موضوعی سازمان‌دهی شده‌اند.

۱. توزیع‌های نمونه‌برداری، برآورد پارامترها و فاصله اطمینان

- آماره آزمون Z (واریانس معلوم یا نمونه بزرگ $n \geq r$):

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

- آماره آزمون استیودنت t (واریانس نامعلوم و نمونه کوچک $n < r$ با فرض توزیع نرمال):

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{(n-1)}$$

- توزیع کای اسکوئر (χ^2) برای برآورد واریانس نمونه‌ای (قضیه کوکران):

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

- توزیع فیشر (F) جهت مقایسه واریانس دو جامعه مستقل یا آزمون معنی‌داری ANOVA:

$$F = \frac{\frac{\chi^2_1}{v_1}}{\frac{\chi^2_r}{v_r}} = \frac{S^2_1}{S^2_r} \sim F_{(v_1, v_r)} \quad \left(\text{در ANOVA: } F = \frac{MS_{\text{Treatment}}}{MS_{\text{Error}}} \right)$$

- فاصله اطمینان میانگین جامعه (CI) و رابطه تقارن آن:

$$CI = \bar{X} \pm \left(t_{\frac{\alpha}{2}, n-1} \times \frac{S}{\sqrt{n}} \right) \Rightarrow \bar{X} = \frac{L + U}{r}, \quad E = \frac{U - L}{r}$$

- واریانس ترکیب خطی متغیرهای تصادفی ($Z = aX + bY + c$):

$$\sigma_z^2 = \text{Var}(aX + bY + c) = a^2 \sigma_x^2 + b^2 \sigma_y^2 + 2ab\sigma_{xy}$$

۲. جدول ماتریسی درجات آزادی (df) در انواع طرح‌های آزمایشی پایه و پیچیده

جهت سهولت در حفظ و مقایسه سریع، فرمول‌های محاسبه درجات آزادی در جدول جامع زیر تدوین شده است:

درجه آزادی کل (df_{Total})	درجه آزادی خطا (df_{Error})	درجه آزادی بلوک/تکرار	درجه آزادی تیمار (df_{Trt})	تعداد واحدها (N)	نوع طرح آزمایشی
$tr - 1$	$t(r - 1)$	—	$t - 1$	tr	کاملاً تصادفی (CRD)
$tr - 1$	$(t - 1)(r - 1)$	$r - 1$	$t - 1$	tr	بلوک‌های کامل تصادفی ($RCBD$)
$a^r - 1$	$(a - 1)(a - r)$	ردیف: $a - 1$	$a - 1$	a^r	مربع لاتین (LSD)



درجه آزادی کل (dfTotal)	درجه آزادی خطا (dfError)	درجه آزادی بلوک/تکرار	درجه آزادی تیمار (dfTrt)	تعداد واحدها (N)	نوع طرح آزمایشی
		ستون: $a - 1$			$(r = t = a)$
$abr - 1$	$(ab - 1)(r - 1)$	$r - 1$	$A: a - 1$ $B: b - 1$ $AB: (a - 1)(b - 1)$	abr	فاکتوریل در قالب بلوک ($A \times B$)
$abr - 1$	خطای a : $(r - 1)(a - 1)$ خطای b : $a(r - 1)(b - 1)$	$r - 1$	اصلی A : $a - 1$ فرعی B : $b - 1$ $AB: (a - 1)(b - 1)$	abr	کرت‌های خردشده بر پایه (SPPD) RCBD
$a^r b - 1$	خطای a : $(a - 1)(a - r)$ خطای b : $a(a - 1)(b - 1)$	ردیف: $a - 1$ ستون: $a - 1$	اصلی A : $a - 1$ فرعی B : $b - 1$ $AB: (a - 1)(b - 1)$	$a^r b$	کرت‌های خردشده بر پایه مربع لاتین
$trs - 1$	آزمایشی: $(t - 1)(r - 1)$ نمونه‌برداری: $tr(s - 1)$	$r - 1$	$t - 1$	trs	طرح بلوک با زیرنمونه‌گیری (s نمونه)

۳. امید ریاضی میانگین مربعات (Expected Mean Squares – EMS)

محاسبه EMS مهم‌ترین گام برای تعیین ساختار صحیح آماره F و برآورد واریانس‌های ژنتیکی و محیطی است. قوانین کلی برای یک آزمایش فاکتوریل دو عاملی ($A \times B$) در قالب طرح کاملاً تصادفی با r تکرار:

- مدل اول: مدل ثابت (Fixed Model) – هر دو عامل A و B ثابت هستند:

$$EMS_A = \sigma_e^r + \frac{rb \sum \alpha_i^r}{a - 1}, \quad EMS_B = \sigma_e^r + \frac{ra \sum \beta_j^r}{b - 1}, \quad EMS_{AB} = \sigma_e^r + \frac{r \sum (\alpha\beta)_{ij}^r}{(a - 1)(b - 1)}, \quad EMS_{Error} = \sigma_e^r$$

○ آماره آزمون صحیح: مخرج تمام نسبت‌های F (برای A ، B و AB)، میانگین مربعات خطا (MS_{Error}) است.

- مدل دوم: مدل تصادفی (Random Model) – هر دو عامل A و B تصادفی هستند:

$$EMS_A = \sigma_e^r + r\sigma_{AB}^r + rb\sigma_A^r, \quad EMS_B = \sigma_e^r + r\sigma_{AB}^r + ra\sigma_B^r, \quad EMS_{AB} = \sigma_e^r + r\sigma_{AB}^r, \quad EMS_{Error} = \sigma_e^r$$

○ آماره آزمون صحیح: مخرج نسبت F برای عوامل اصلی A و B ، میانگین مربعات اثر متقابل (MS_{AB}) است؛ اما مخرج نسبت F برای اثر متقابل، همان MS_{Error} است.



- مدل سوم: مدل مختلط (Mixed Model) - عامل A ثابت و عامل B تصادفی است):

$$EMS_A = \sigma_e^r + r\sigma_{AB}^r + \frac{rb \sum \alpha_i^r}{a-1}, \quad EMS_B = \sigma_e^r + r\alpha\sigma_B^r, \quad EMS_{AB} = \sigma_e^r + r\sigma_{AB}^r, \quad EMS_{Error} = \sigma_e^r$$

- آماره آزمون صحیح: برای آزمون عامل ثابت (A)، باید از MS_{AB} در مخرج استفاده کرد ($F = \frac{MS_A}{MS_{AB}}$)، اما برای آزمون عامل تصادفی (B) و اثر متقابل (AB)، از MS_{Error} در مخرج استفاده می‌شود.

۴. روابط خطاهای آماری، توان آزمون و آزمون‌های چنددامنه‌ای

- رابطه نرخ خطای آزمایش محور (α_{EW}) با نرخ خطای مقایسه محور (α) برای c مقایسه مستقل:

$$\alpha_{EW} = 1 - (1 - \alpha)^c \quad \text{که در آن} \quad c = \frac{t(t-1)}{r}$$

- روابط حداقل تفاوت معنی‌دار در آزمون‌های میانگین:

$$LSD = t_{\alpha, r, df_{error}} \sqrt{\frac{rMS_{error}}{r}} \quad (\beta \text{ و بیشترین } \alpha \text{ کمترین})$$

$$HSD = q_{\alpha, (t, df_{error})} \sqrt{\frac{MS_{error}}{r}} \quad (\alpha \text{ بیشترین کنترل روی})$$

- رابطه ضریب تغییرات آزمایشی (CV - Coefficient of Variation):

$$CV(\%) = \frac{\sqrt{MS_{Error}}}{\bar{Y}} \times 100 \quad (\text{شاخص بی‌بعد برای ارزیابی دقت آزمایش و کنترل خطای تصادفی})$$

۵. روابط جبری و ماتریسی رگرسیون چندگانه، همبستگی و بیومتري

- مدل رگرسیون خطی ساده و برآورد ضرایب با روش کمترین توان‌های دوم (OLS):

$$\hat{Y}_i = a + bX_i \Rightarrow b = \frac{SP_{xy}}{SS_x} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}, \quad a = \bar{Y} - b\bar{X}$$

- مدل ماتریسی رگرسیون خطی چندگانه ($Y = X\beta + \epsilon$):

$$\hat{\beta} = (X'X)^{-1}X'Y$$

- X : ماتریس ضرایب متغیرهای مستقل (ماتریس طرح با ابعاد $n \times p$).

- $X'X$: ماتریس متقارن ضرب خلیجی متغیرهای مستقل با ابعاد $p \times p$.

- $(X'X)^{-1}$: ماتریس معکوس (که عناصر روی قطر اصلی آن، وارینانس ضرایب رگرسیونی را می‌سازند):

$$Var(\hat{\beta}_j) = C_{jj}\sigma^r$$

- ضریب تبیین (R^r) و ضریب تبیین تعدیل شده (R_{adj}^r):

$$R^r = \frac{SS_{Regression}}{SS_{Total}} = 1 - \frac{SS_{Error}}{SS_{Total}}, \quad R_{adj}^r = 1 - \left[\left(\frac{n-1}{n-p} \right) (1 - R^r) \right]$$

- آزمون فرضیه برای معنی‌داری کل مدل رگرسیون (آماره F):

$$F = \frac{MS_{Regression}}{MS_{Error}} = \frac{\frac{SS_{Reg}}{k}}{\frac{SS_{Error}}{n-k-1}} \sim F_{(k, n-k-1)} \quad (k: \text{تعداد متغیرهای مستقل})$$

- وراثت‌پذیری عمومی (h_b^r) و وراثت‌پذیری خصوصی (h_n^r) در بیومتري:

$$h_b^r = \frac{V_G}{V_D} = \frac{V_A + V_D + V_I}{V_G + V_R + V_{GE}}, \quad h_n^r = \frac{V_A}{V_D}$$