

مجله ایمنی زیستی

دوره ۱۲، شماره ۳، پائیز ۱۳۹۸

ISSN 2716-9804 الکترونیکی، ISSN 2717-0632 چاپی

## نرم افزار SISSI در راستای تخمین حجم بهینه نمونه تراریخته

زهرا حاجی برات<sup>۱</sup>، مسعود توحیدفر<sup>۲\*</sup>

۱- دانشجوی دکتری دانشکده علوم حیات، دانشگاه شهید بهشتی، تهران

۲- دانشیار و عضو هیات علمی دانشکده علوم و فناوری زیستی، دانشگاه شهید بهشتی، تهران

gtohidfar@yahoo.com

تاریخ دریافت: ۹۸/۱۰/۲۹، تاریخ پذیرش: ۹۹/۰۴/۲۰

صفحه ۹۱-۱۰۶

### چکیده

از زمان ظهور علم مهندسی ژنتیک در سال ۱۹۷۰، ملاحظات دربارۀ مخاطرات احتمالی این تکنولوژی ایجاد شد همین مسئله باعث شد تا شناسایی تراریخته در اجرای قوانین برچسب‌گذاری مهم شود. بنابراین روش‌های شناسایی معتبر محصولات تراریخته در راستای کاهش و یا به حداقل رساندن این خطاها ضروری بنظر می‌رسد. معتبر بودن این روش‌ها تا حدودی تحت تاثیر نمونه‌برداری و ناشی از پراکنش غیریکنواخت تراریخته است. نمونه‌برداری در بیشتر مواقع منبع خطا بوده و بر این اساس نرم‌افزارهایی برای کاهش خطاها طراحی شده‌اند که از جمله آن‌ها می‌توان به نرم افزار SISSI اشاره کرد. این نرم‌افزار ابزاری جدید برای تعیین حجم مطلوب نمونه در مجموعه داده آزمایشی است و براساس روش جک نایف می‌باشد. این روش برای تخمین واریانس جامعه واقعی کاربرد دارد که اطلاعات چندانی از جامعه در دست نیست. این نرم‌افزار نتایج را به صورت گرافیکی با تغییرات میانگین و انحراف استاندارد نمایش می‌دهد و گزینه اتوماتیک برای تعیین حجم نمونه در این برنامه تعبیه شده است. هدف استفاده از این نرم‌افزار تعیین حجم بهینه نمونه‌برداری برای محموله‌های تراریخته‌ای است که از یکنواختی برخوردار نیستند. کاربرد دیگر آن شناسایی مواد تراریخته در نمونه‌های تجاری است که از حساسیت بالایی در تجارت جهانی برخوردار هستند.

**واژه‌های کلیدی:** SISSI، جک نایف، نمونه‌گیری مجدد، میانگین، غیرپارامتری.

## مقدمه

بیوتکنولوژی کشاورزی و بویژه مهندسی ژنتیک با تولید محصولات تراریخته فرصتی برای کاهش فقر، ایجاد امنیت غذایی، رفع سوء تغذیه، جلوگیری از تخریب منابع طبیعی و افزایش توانایی در اصلاح نباتات ایجاد کرده است. با این وجود، ورود بیوتکنولوژی در تحقیقات کشاورزی و تولید محصولات تراریخته، در بسیاری از کشورها با چالش‌هایی همراه بوده است (۱). این تکنولوژی منجر به تولید محصولات تراریخته و ایجاد محصولات مقاوم به تنش‌های غیرزیستی و زیستی شده است (۴). قوانین و مقررات در این سالها به دنبال تایید محصولات تراریخته و جلوگیری از اختلاط بیش از حد قابل قبول کشورها بوده است (۵). کنترل رسمی محصولات تراریخته در هر مرحله از تولید، فرآوری، ذخیره‌سازی و توزیع این محصولات در بعضی از کشور باید انجام شود و می‌بایست هنگام واردات محصول، مقدار اختلاط تراریخته مشخص شود و به این دلیل هنگام ورود محموله، نمونه‌برداری و کنترل محموله با

دقت و حساسیت زیادی انجام می‌شود (۶). پروتکل‌ها و روش‌های مختلف به منظور دستیابی سطح اطمینان مطلوب از شناسایی محصولات تراریخته وجود دارد (۷). مرحله تعیین نمونه‌برداری از محموله‌های بزرگی که از طریق کشتی یا کامیون‌ها وارد کشورها می‌شوند، بسیار مهم و حیاتی است که نمونه‌ها باید به میزان حداقل مقدار باشند که از نظر اقتصادی مقرون به صرفه باشد (۸). مرحله نرم‌افزاری، مرحله بسیار مهم بوده که نمونه جمع‌آوری شده نماینده مناسبی از جمعیت باشد و اطمینان کافی از نمونه‌گیری صحیح را به ما بدهد (۹). اصول کلی و روش‌های نمونه‌گیری بذر باید مطابق با قوانین انجمن بین‌المللی تست بذر (ISTA) باشد (۷). بر این اساس تست بذر برای آنالیز تراریخته، باید حاوی حداقل ۲۹۹۵ بذر برای شناسایی بذور تراریخته با ۱/۰ درصد در سطح اطمینان ۹۵ درصد انجام شود (۱۰). اطلاعات ناصحیح از آنالیز محموله وارداتی تاثیر نامطلوبی بر اقتصاد و قانون تجارت دارد، همچنین خطراتی برای

## "حاجی برات و توحیدفر، نرم‌افزار SISSI در راستای تخمین حجم بهینه نمونه تراریخته"

تراریخته مورد قبول هر کشوری در نظر گرفته شود. برآورد این مقدار نیاز به سیستم‌های غربالگری پیشرفته در شناسایی میزان اختلاط تراریخته دارد و تولیدکنندگان باید میزان اختلاط را اعلام کنند. نمونه آزمایشگاهی برای آنالیز اختلاط باید حاوی یک نمونه حداقل ۱۰۰۰۰ تایی باشد با فرض اینکه خطای نرم‌افزار تا ۲۰ درصد مورد قبول باشد (۱۵). هدف اصلی کنترل محصولات، برچسب‌گذاری و تعیین میزان اختلاط تراریخته در محصولات می‌باشد. مشکلات این بخش به فاکتورهای زیادی وابسته است که می‌توان به، نوع و اندازه محصولات مختلف اشاره کرد. محققان به کمک نرم‌افزار SISSI نمونه‌های مربوط به سویا بسته‌بندی شده را تجزیه و تحلیل کردند و میزان خطای نرم‌افزار به کمک ضریب جک نایف به شدت کاهش یافته است. ضریب جک نایف ضریبی است که برای تخمین اریبی آماره‌های نمونه و برای تخمین واریانس به کار می‌رود (۱۶). این نرم‌افزار به کمک روش جک نایف، از نمونه اصلی که توسط محققان گرفته

مدیران و تحلیلگران این بخش به همراه دارد (۱۱، ۱۲). بر این اساس، در نرم‌افزار محصولات تراریخته باید مباحثی دیگر همچون دامنه ایمنی، کیفیت غذا و علوفه در نظر گرفته شود (۹). مراحل مختلف نرم‌افزار به‌طور معمول در ارزیابی سطح محصولات تراریخته در سه مرحله (بذر، پودر بذور، مواد فرآوری‌شده) انجام می‌شود. مراحل نرم‌افزار، احتمال ایجاد خطای بزرگ در سطح نمونه‌گیری را ایجاد می‌کند که ناشی از واریانس، وابسته با هر مرحله از ارزیابی محصولات تراریخته بوده که به احتمال ۱۰۰ درصد این خطاها قابل محاسبه نیستند. برآورد بیش از حد یا برآورد کمتر از حد اندازه منجر به تصمیم غلط در هنگام نرم‌افزار می‌شود و باعث ایجاد پیامدهای اجتماعی و اقتصادی منفی می‌شود (۱۳). تشخیص محصولات تراریخته به چند دلیل مهم است. یکی از مهمترین دلایل پیروی از قوانین برچسب‌گذاری این است که به مصرف‌کنندگان اجازه انتخاب آگاهانه را می‌دهد (۱۴). براساس قوانین برچسب‌گذاری، باید حداکثر آستانه وجود

نمونه کم نیز کاربرد دارد (۷). سپس توکی پیشنهاد کرد که این ضریب جک نایف علاوه بر کاهش اریبی، می تواند برای برآورد واریانس نیز استفاده شود (۱۸).

#### ۱- نرم افزار SISSI برنامه کاربردی برای تعیین حجم نمونه از محموله های تراریخته

این نرم افزار تعیین حجم نمونه را به کمک کامپیوتر انجام می دهد (۱۶). نمونه برداری دقیق زمانی امکان پذیر است که تعادل مناسبی میان طرح تحقیق، فرضیات و قدرت آماری وجود داشته باشد. در این روش بازخورد نتایج به دست آمده به شدت به حجم و همچنین پراکنش یکنواخت نمونه ها بستگی دارد. بیشتر نرم افزارهای موجود به فرضیات آماری و همچنین به توزیع داده ها متکی هستند، درحالی که هنگام کار با نمونه های بیولوژیکی به احتمال زیاد این امکان برای کاربر فراهم نمی شود (هنگامی که مقدار زیادی نمونه وجود داشته باشد). نرم افزارهای قبلی تنها روش استنباطی متداول مبتنی بر توزیع t-student بودند.

می شود زیرنمونه هایی تشکیل می دهد و هر زیرنمونه به عنوان نمونه اصلی در نظر گرفته می شود و برای کاهش خطای نرم افزار این مراحل به صورت پی درپی تکرار می شود به طوری که به عنوان مثال اگر ۱۰ نمونه گرفته شود، نمونه به ۱۰ زیرنمونه تقسیم می شود و سپس از زیرنمونه ها به عنوان تکرار نرم افزار استفاده می شود که منجر به کاهش خطای نرم افزار می شود. SISSI نرم افزاری جدید، برای تعیین حجم نمونه در مجموعه داده آزمایشگاهی است (۱۶). این نرم افزار به کمک ضریب جک نایف، میزان بهینه نرم افزار را تخمین می زند (۱۷،۲).

نخستین بار کنولی، ضریب جک نایف را به عنوان روشی کارآمد برای کاهش خطای نرم افزار معرفی کرد و بر اساس حذف هر بار یک مشاهده از مجموعه داده های اولیه و محاسبه ی مجدد داده ها با استفاده از بقیه ی داده ها برای کاهش اریبی نمونه های گرفته شده عمل می کند (۳). در مطالعات دیگر این روش نیز تعمیم داده شده و در یک جامعه با تعداد

## "حاجی برات و توحیدفر، نرم افزار SISSI در راستای تخمین حجم بهینه نمونه تراریخته"

شرایط آزمایشگاهی داشته باشد که مرحله‌ای زمان‌بر و پرهزینه برای محققان است (۲۳).

موقعی که اطلاعات کمی از نمونه‌های گرفته شده و یا محاسبات به روش آمار سنتی، پیچیده و مشکل باشد، استفاده از روش‌های نمونه‌گیری مجدد می‌تواند مشکل خطاهای ناشی از نمونه‌برداری داده را برطرف کرده و همچنین تعیین حجم مناسب نمونه را با داده‌ها فراهم سازد. علاوه بر این، روش‌های نمونه‌گیری مجدد (مثل جک نایف) این قابلیت را دارند که به سوالاتی که نمی‌توان به آن‌ها از روش‌های پارامتری پاسخ داد، جوابی قانع کننده ارائه دهد (۱۸).

ضریب جک نایف روش مناسبی برای تخمین واریانس جامعه و محاسبه‌ی فاصله‌ی اطمینان در موقعی است که اطلاعات چندانی در مورد جامعه در دست نیست (۲۴). برآورد واریانس به روش جک نایف که ضمن سهولت در محاسبه، موجب کاهش چشم‌گیری در برآورد آریبی می‌شود.

تعیین حجم نمونه با استفاده از نرم‌افزار SeedCalc امکان‌پذیر است که به‌عنوان برنامه‌ای برای تعیین خلوص و ناخالصی بذر استفاده می‌شود (۱۹) و کاربردی در تعیین حجم مناسب برای نرم‌افزار نداشته و این در صورتی است که نرم‌افزار SISSI تعیین حجم مناسب نمونه را به کمک نمونه‌گیری مجدد (ضریب جک نایف) انجام می‌دهد (۲۰، ۲۱) و فرصت استفاده مکرر از داده‌های یک نمونه را برای کاربر فراهم می‌کند و از روش نمونه‌گیری مجدد می‌توان برای شناسایی ناخالصی موجود در بذور داخل مجموعه‌ای که دارای حجم محدود نمونه هستند استفاده کرد. در صورتی که در نرم‌افزارهای قبلی چنین امکانی وجود نداشته و این نرم‌افزار توانسته نواقص نرم‌افزارهای قبلی (KeSTE) را نیز برطرف نماید (۲۲). روش‌های نمونه‌گیری مجدد شامل نمونه‌گیری از نمونه اصلی است و از نظر اقتصادی مقرون به صرفه است به دلیل آنکه نمونه‌گیری مجدد توسط خود نرم‌افزار صورت گرفته و بدون اینکه نیازی به مرحله مجدد نمونه‌گیری در

## ۲- توصیف نرم افزار

مبانی آماری و ویژگی‌های نرم‌افزار در زیر آورده شده است. روش‌های اجرا شده در این نرم‌افزار با زمینه علمی و همچنین برخی اصول کاربردی به شکل کامل، درون فایل کمکی آورده شده است.

## ۲-۱ تعیین حجم نمونه به روش اتوماتیک

با استفاده از روش جک نایف، نمونه اصلی از  $n$  عنصر به داخل گروه‌هایی با  $k$  عنصر تقسیم بندی می‌شود.  $n_i$  برآوردی برای  $n$  بر اساس نمونه‌ای به اندازه‌ی  $n-1$  است که بعد از حذف  $i$  امین واحد نمونه به دست می‌آید.

برآورد جک نایف برای نمونه  $n$  از این فرمول به دست می‌آید.

$$\sum_{k=1}^{n-2} n! / (n-k)! k!$$

روش جک نایف به فرضیات مدل حساسیت کمتری داشته و نیازی به فرمول‌های نظری که در روش سنتی استفاده می‌شود، ندارد و در مجموع به عنوان یک روش غیرپارامتری در تحلیل‌های آماری شناخته می‌شود. روش جک

نایف نیاز به محاسبه مکرر آماره به تعداد  $n$  مرتبه دارد که در گذشته امکان‌پذیر نبود (۱۵).

میانگین و انحراف معیار برای همه نمونه‌های تشکیل شده و رسم پلات در دو نمودار محاسبه می‌شوند در حالی که حجم نمونه  $N-K$  است. نمایش تصویری از نمونه‌های به وجود آمده ناشی از تغییرات حجم نمونه اولیه، به شکل میانگین و انحراف استاندارد متفاوت از نمونه اصلی به نمایش گذاشته می‌شود. تکرار این فرآیند با استفاده از نمونه اولیه با اندازه‌های مختلف می‌تواند به عنوان راه‌حلی برای تخمین شرایط اولیه نمونه‌ها باشد. تجزیه و تحلیل میانگین نشان می‌دهد که تغییرپذیری بین میانگین‌ها با افزایش حجم نمونه به صورت قابل ملاحظه‌ای کاهش می‌یابد. در این روش رگرسیون نیز محاسبه می‌شود.

رگرسیون یک فرآیند آماری برای تخمین روابط بین متغیرها است. این مدل شامل روش‌های زیادی برای مدل‌سازی و تحلیل متغیرهای خاص و منحصر بفرد است وقتی که تمرکز روی روابط بین

می شود. برآوردگری برای پارامتر نامعلوم براساس نمونه ای به اندازه  $n-1$  است که بعد از حذف  $i$  امین واحد نمونه به دست می آید. آنگاه برآورد جک نایف این نمونه عبارت است از:

$$\theta J = (n-1) \sum_{i=1}^n [n\theta - (n-1)\theta_i]$$

برآورد واریانس جک نایف برای نمونه بالا به صورت زیر است:

$$uJ(\theta) = n(n-1) + \sum_{i=1}^n (\theta_i - \theta)^2$$

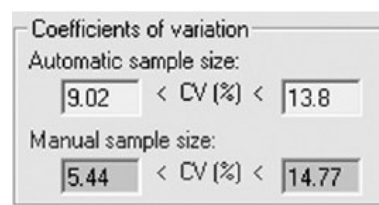
همانطور که در فرمول فوق ملاحظه می شود روش جک نایف به فرضیات مدل بستگی کمتری داشته و نیازی به فرمول های نظری که در روش سنتی به کار گرفته می شود، ندارد و در مجموع به عنوان یک روش ناپارامتری در تحلیل های آماری شناخته می شود. با این حال، جک نایف نیاز به محاسبه ی مکرر آماره به تعداد  $n$  مرتبه دارد که واقعا این کار در قدیم امکان پذیر نبوده است (۴).

متغیر وابسته و یک یا چند متغیر مستقل باشد. آنالیز رگرسیون این سوال را مطرح می کند که چگونه مقدار متغیر وابسته با تغییر هر کدام از متغیرهای مستقل و با ثابت بودن دیگر متغیرهای مستقل تغییر می کند. شاخص جهانی ( $SR^2$ ) از جمع ضرایب تعیین چهار رگرسیون محاسبه می شود. تکرار مراحل از ۱ تا ۴ برای همه نمونه ها امکان اجازه شناسایی حجم نمونه بهینه ( $N-K$ ) را می دهد. در جایی که  $SR^2$  بالاترین مقدار بوده، حجم نمونه بعدی ایجاد شده  $SR^2$  بالاتر از ۵ درصد نسبت به ارزش قبلی است، روند محاسبه رگرسیون، متوقف می شود. حذف نمونه های پرت به خارج از محدوده محاسبات، روند اصلاح نمونه های پرت را امکان پذیر می سازد. (برای مثال، ۵ درصد نمونه های اضافی).

به عنوان مثال محاسبه به این صورت است که  $X_1, \dots, X_n$  متغیرهای تصادفی مستقل و هم توزیع با مقادیر حقیقی از یک توزیع نامعلوم  $F$  باشند و آماره ی برآوردگری برای پارامتر نامعلوم بوده که از روی نمونه ی اصلی محاسبه

## ۲-۲ تعیین دستی حجم نمونه

کاربر با تنظیم دستی می‌تواند حجم بهینه نمونه به کمک روش نرم‌افزار مجدد را محاسبه کرده و این بدین معناست که به کمک روش دستی اگر حجم نمونه اولیه محاسبه شود و میزان ضریب تغییرات به دست آمده پایین باشد حجم نمونه محدود از نظر محققان قابل قبول است و اگر این میزان تغییرات بالا باشد آنقدر کاربر به کمک پیکان مربوط به تنظیمات دستی تغییر انجام می‌دهد تا ضریب تغییرات پایین آمده و مورد قبول محققان شود. شکل ۱ نمونه‌ای از مقایسه ضریب تغییرات به کمک دو روش دستی و خودکار را نشان می‌دهد.



شکل ۱- حجم نمونه دستی و اتوماتیک: مقایسه بین ضریب تغییرات

## ۲-۳ تعیین حجم نمونه براساس آزمون t-student

روش دیگر برای تعیین حجم بهینه‌ی نمونه‌برداری به کمک روش خودکار است

که با روش معمولی t-student مقایسه می‌شود. زمان تعیین تقریبی میانگین نمونه‌های برداشته شده از یک متغیر تصادفی با توزیع t-student انجام می‌شود. این توزیع اساس آزمونی به نام "تی تست" یا آزمون تی" است که مقدار اطمینان از تفاوت دو متغیر تصادفی را از روی نمونه‌هایشان اعلام می‌کند.

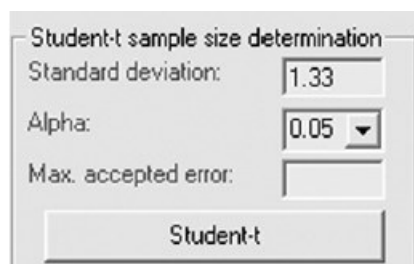
آزمون t، توزیع یا مجموعه‌ای از توزیع‌هاست که با استفاده از آن‌ها فرضیه‌هایی درباره نمونه در شرایط جامعه ناشناخته، مورد بررسی قرار می‌گیرد. اهمیت این آزمون (توزیع) زمانی است که پژوهشگر می‌تواند با نمونه‌های کوچکتر (حداقل ۲ نفر) اطلاعاتی درباره جامعه به دست آورد. در این روش کاربر حداکثر خطای قابل قبول که مورد پذیرش محققان است را برای حجم نمونه مورد نظر توسط نرم‌افزار در نظر گرفته (تفاوت بین نمونه و میانگین جمعیت) و واریانس را که وابسته با روند مورد مطالعه است، تخمین می‌زند (شکل ۲). تعیین آزمون تی با روش خودکار مقایسه شده که اعتبار آن تابع زمانی است که توزیع داده نرمال باشد.

"حاجی برات و توحیدفر، نرم‌افزار SISSI در راستای تخمین حجم بهینه نمونه تراریخته"

ضریب جک نایف تعیین شد و بهترین حجم نمونه در این روش نمونه ۵ که در قسمت مربوط به Results (Visual jackknife (auto)) است که به کمک روش جک نایف به زیرجمعیت‌هایی تقسیم‌بندی شدند.

در شکل (۴) بهترین حجم نمونه به کمک روش t-student تعیین شد. در این شکل نمونه ۷ که در قسمت مربوط به Results (Student-t based) است بهترین حجم نمونه را برآورد می‌کند.

براساس روش دستی که کاربر با کمک تغییر جهت پیکان به سمت بهترین حجم نمونه به دست می‌آورد، می‌توان ضریب تغییرات نزدیک یا کمتر از نمونه به روش خودکار را به دست آورد و در این شکل نمونه ۶ مربوط به بخش Results (Visual jackknife (manual)) بهترین حجم نمونه با استفاده از این روش تخمین زده شده است (شکل ۵).



شکل ۲- تعیین حجم نمونه مبتنی بر t-student

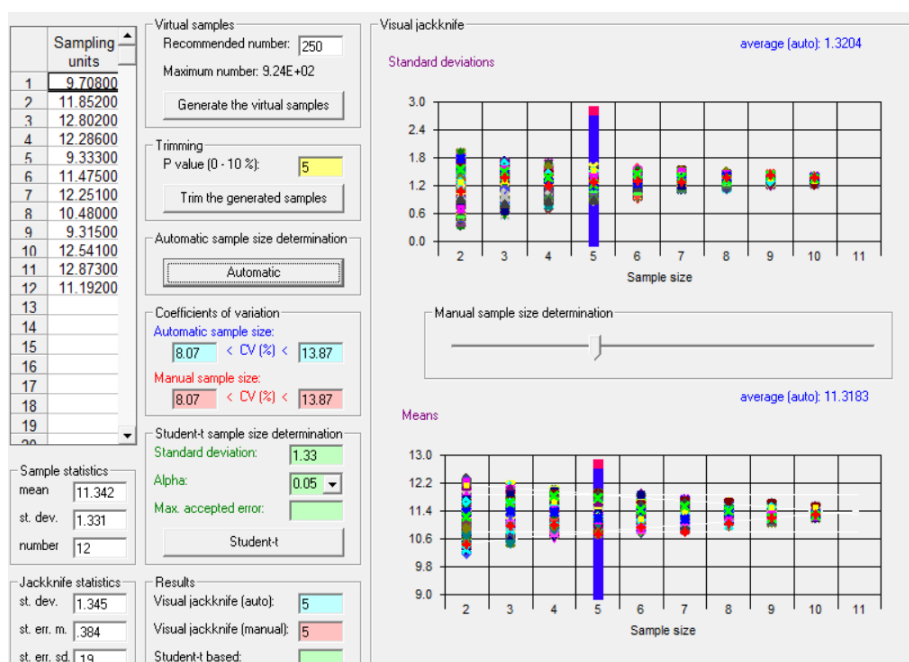
## ۲-۴ مثالی از نحوه کارکرد و مشخصات این نرم‌افزار

### ۲-۴-۱ داده ورودی برای نرم‌افزار SISSI

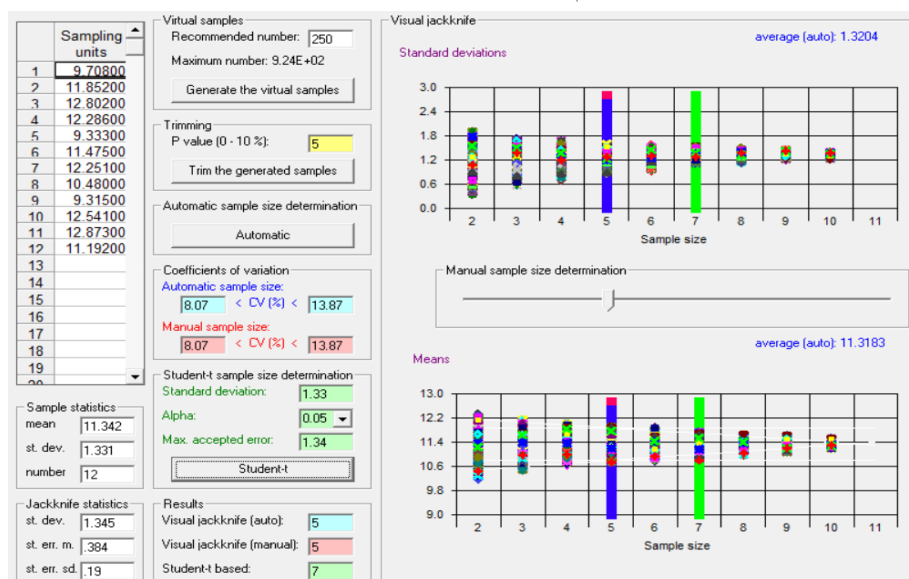
داده‌های ورودی در یک صفحه Excel آماده و ورودی عددی در یک ستون منفرد تشکیل شده است. خروجی‌های نرم‌افزار به صورت عددی و بصری نمایش داده شده است. نتایج و تنظیمات در مورد روش محاسباتی خروجی‌ها به صورت فایل‌های متنی نیز به دست می‌آید.

مثال عددی با هدف درک چگونگی پیوستگی روش جک نایف با قابلیت‌های بصری از SISSI را نشان می‌دهد.

شکل (۳) بیانگر تعیین بهترین حجم نمونه با استفاده از روش خودکار با وجود



شکل ۳- نمایش تصویری از بهترین حجم نمونه به روش خودکار و دستی (با استفاده از میانگین و واریانس جک نایف حجم نمونه ۵ با استفاده از روش خودکار تعیین شده است که در ستون اول (۹/۳۳) به عنوان حجم مناسب نرم افزار در نظر گرفته شد).



شکل ۴- تعیین حجم نمونه به روش t-student (با استفاده از میانگین و واریانس جک نایف حجم نمونه ۷ با استفاده از روش t-student تعیین شده است دستی که در ستون اول (۱۲/۲۵) به عنوان حجم مناسب نرم افزار در نظر گرفته شد).

این نرم‌افزار از بخش‌های مختلفی تشکیل شده است. Sampling units، فایل ورودی است که به صورت اکسل آماده می‌شود و از این قسمت می‌توان وارد نرم‌افزار شد. Virtual samples، در این بخش زیرنمونه‌ها به کمک روش جک نایف با توجه به فایل ورودی برآورد می‌شوند. Trimming، پیرایش داده‌ها با استفاده از دو گزینه دستی و خودکار انجام می‌شود و به این معناست که داده‌هایی که بیش از حد از میانگین داده اصلی (بالاترین و پایین‌ترین) فاصله داشتند به کمک p-value که ۰ درصد (اغماض زیاد) و ۱۰ درصد (اغماض کمتر) حذف می‌شود و ۵ درصد به عنوان p-value مناسب در نظر گرفته و بیشتر محاسبات به کمک ۵ درصد محاسبه می‌شود.

Manual Student-t، Automatic sample size determination به ترتیب گزینه خودکار، روش t-student و روش دستی است. نرم‌افزار به کمک این سه گزینه حجم بهینه نمونه برداری را محاسبه می‌کند. گراف بالایی ( Standard

deviations) براساس انحراف معیار استاندارد حجم بهینه را براساس سه روش فوق محاسبه کرده و گراف پایین که در قسمت مربوط به (Mean) مبتنی بر تغییرات میانگین است که توسط سه روش گفته شده، میانگین را محاسبه می‌کند و در قسمت Results، به کمک دو روش دستی و خودکار که هر دو مبتنی بر ضریب جک نایف است حجم بهینه نمونه را برآورد کرده و روش سوم نیز که با استفاده از t-student است در قسمت پایین شکل آورده شده و حجم نمونه را محاسبه می‌کند. براساس نتایجی که از سه روش (خودکار، t-student و دستی) به دست آمد، می‌توان نتیجه گرفت که بهترین حجم نمونه براساس این روش‌ها به ترتیب ۵، ۷ و ۵ هستند و می‌توان چنین استدلال کرد که روشی که کمترین ضریب تغییرات را داشته یعنی از نظر آماری گزینه‌ای که کمترین خطا را دارد به عنوان بهترین روش برای تعیین حجم بهینه نمونه در نظر گرفته شود و با توجه به نتایجی که به کمک این دو روش دستی و خودکار به دست می‌آید و همچنین

یا عدم وجود این محصولات تراریخته در محموله وارداتی باید از نرم‌افزارها یا روش‌های مختلف نمونه‌گیری مناسب برای تعیین این میزان اختلاط استفاده کنند. این نرم‌افزار به خوبی می‌تواند اطمینان از نمونه‌گیری را برای محققان فراهم کند. به‌عنوان مثال برای برآورد مقدار تراریخته در محموله وارداتی ابتدا باید حجم مناسب نرم‌افزار تعیین شود که نماینده مناسبی از جامعه آماری باشد و برای این کار از نرم‌افزار SISSI استفاده شده و داده ورودی در جدول ۱ آورده شده است.

کمترین ضریب تغییرات نسبت به روش  $t$ -student وجود دارد که روش  $t$ -student برای نمونه‌های با حجم کم کاربرد داشته و کارایی برای نمونه‌هایی با حجم بالا نداشته است و می‌توان این دو روش را برای آنالیز داده‌ها استفاده کرد.

## ۲-۵ مثال عملی برای سویا

کشورهای مختلف از جمله کشورهای اروپایی که حساسیت بسیار بالایی برای وجود محصولات تراریخته در محموله وارداتی و یا در بسته غذایی دارند برای تعیین میزان وجود

جدول ۱- میزان بذر برای تعیین حجم بهینه نمونه برای نرم‌افزار SISSI.

اندازه بذر (تن)	تعداد
۵۰	۱۰
۱۰۰	۲۰
۲۵۰	۵۰
۵۰۰	۱۰۰

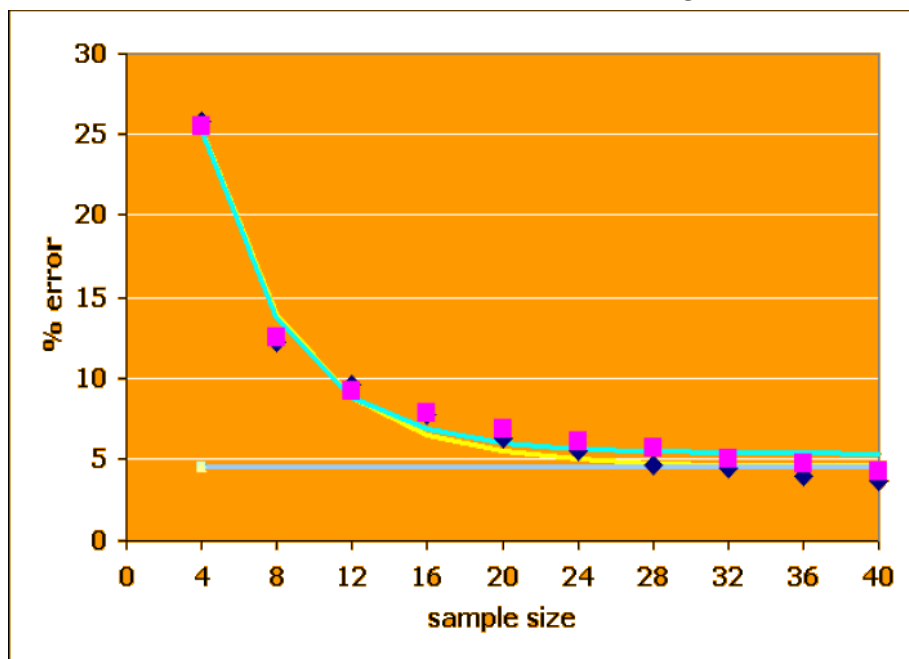
محموله وارداتی به کمک نرم‌افزار SISSI انجام شده و حجم بهینه نمونه‌برداری با استفاده از این نرم‌افزار برآورد شده است که

مثالی از محموله‌ای که حاوی بذر سویا است و برای تعیین مقدار سویای تراریخته با استفاده از تعهدنامه اروپا ۲۰۰۴/۷۸۷ در

"حاجی برات و توحیدفر، نرم افزار SISSI در راستای تخمین حجم بهینه نمونه تراریخته"

تغییرات، این تعداد بذر به عنوان حجم مناسب نمونه برداری باید در نظر گرفته شود (۱۸).

تعداد ۴۰۰ بذر در هر نمونه به عنوان مناسبترین تعداد بذر معین شد و همانطور که در شکل ۵ مشخص شده است، با کمترین میزان ضریب



شکل ۵- تعیین حجم بهینه محموله سویا

مناسب نمونه برداری، تعادلی بین حساسیت، هزینه و اطمینان را برقرار می کند (۸). نرم افزار SISSI برای تعیین حجم نمونه در پژوهش ها آزمایشی به عنوان نرم افزاری مناسب برای نمونه هایی که از پیچیدگی های مختلفی از جمله حجم متغیر و غیر یکنواخت بودن نمونه و سایر موارد برخوردار هستند، مورد استفاده قرار می گیرد. SISSI در درجه اول برای حمایت از آزمایش ها در حوزه زراعی

### بحث و نتیجه گیری

قوانین و روش های مختلفی برای وجود موجودات تراریخته در محصولات زراعی، غذاها و مواد تشکیل دهنده آن به اجرا در آمده است که استفاده از روش های قابل اعتماد و دقیق برای تشخیص این موجودات لازم و مهم تلقی می شود. نرم افزار مطلوب به عنوان روشی قابل اعتماد برای تشخیص این تراریخته در نظر گرفته شده است. استراتژی

برای نمونه‌هایی که در شرایط مزرعه‌ای با حجم بسیار متغیر هستند کارایی بسیار بالایی دارد. این نرم‌افزار روش جدیدی برای تعیین حجم بهینه نرم‌افزار به کمک روش نرم‌افزار مجدد ارائه می‌کند. این روش میانگین و انحراف استاندارد را برای هر زیرنمونه محاسبه کرده و نیازی به فرضیات آماری ندارد. همچنین تفسیر نتایج به شیوه‌ای گویا و گرافیکی برای کاربر فراهم می‌شود. تغییرات میانگین و انحراف استاندارد با افزایش نمونه به خوبی در شکل مشاهده می‌شود. نرم‌افزار فوق به خوبی چالش‌های مربوط به نمونه‌گیری و همچنین مشکلات مربوط به تهیه نمونه را نیز برطرف کرده و به‌عنوان نرم‌افزاری مناسب جهت کار برای دانشمندان در حوزه بیوتکنولوژی معرفی می‌شود.

گسترش یافته اما پس از آن این نرم‌افزار برای نمونه‌گیری محموله‌های وارداتی به کمک ضریب جک نایف تحولی بزرگ در تعیین حجم بهینه نمونه را فراهم ساخت. این نرم‌افزار از همه زمینه‌های آماری برای تعیین بهینه حجم نمونه استفاده می‌کند. نرم‌افزار مذکور دارای انعطاف پذیری و سهولت در حین انجام کار است. از ویژگی‌های خوب و قابلیت‌های این نرم‌افزار می‌توان مستقل بودن از فرض‌های آماری و نیاز به نرم‌افزار مجدد توسط کاربر که مرحله‌ای هزینه‌بر و زمان‌بر است، را کاهش داده و نمایش آماره به‌صورت بصری برای کاربر و محققان فراهم می‌کند. SISSI نرم‌افزاری جالب و مطلوب برای تعیین حجم بهینه نمونه‌های تراریخته است که اطلاع از پراکنش مواد در محموله نداشته و همچنین

## References

۱. رهنما، ح. (۱۳۸۷). اخلاق زیستی و تولید محصولات تراریخته. شماره ۲. صص ۱-۱۴.
۲. نورینی، م. (۱۳۹۱). برآورد واریانس به‌روش جک نایف در آمارگیری‌های دوچارچوبی. مجله بررسی‌های آمار رسمی ایران. شماره ۱، صص ۱۰۳-۱۲۷.
۳. عدالت سروستانی، م. ر.، شهرکی، م. ر. (۱۳۹۲). ارائه ی مدلی برای رتبه‌بندی ریسک پروژه‌ها به روش نمونه‌گیری مجدد جک نایف همراه با تحلیل بازه‌ای. دهمین کنفرانس بین‌المللی مهندسی صنایع، دانشگاه تهران، ۷ و ۸ بهمن ماه.

## فهرست منابع

4. Skinner C.J., Rao J.N.K. (1996). Estimation in dual frame survey with complex design, *Journal of the American Statistical Association*. 91, 349-356.
5. Gilbert J. (1999). Sampling of raw materials and processed foods for the presence of GMOs. *Food Control*. 10:363-365.
6. Borenstein M., Rothstein H. and Cohen J. (2001). *Power and Precision*. Biostatistics, Englewood NJ, USA.
7. Zoubir A.M. and Boashash B. (1998). The Bootstrap and Its Application in Signal Processing. *IEEE Signal Processing Magazine*. 15:56-76.
8. Hußner P., Walblinger H.U., Pietsch K. and Brodman P. (2001). Validation of PCR methods for quantitation of genetically modified plants in food. *Journal of AOAC International*. 84:1855-1864.
9. Ancel V., Bellocchi G., Berben G., Bertheau Y., Brera C., De\_Giacomo M., Janssen E., Kobilinsky A., Kozjak P., Macarthur R., Miraglia M., Onori R., Pla M., Papazova N., Rutar R., Taverniers I. and Šuštar Vozlic J. (2009). International conference. France.
10. Atkinson D. (1998). Genetically modified organisms. Available from <http://www.sac.ac.uk/info/External?Publications/Gmo.asp>., Scottish Agricultural College.
11. Zhou H.B. and Zhang, H. (2010). Dynamic risk management system for large project construction in china. *Proc. Geo. Florida Conf. Adv. Anal. Model Des*.
12. Berg P., Baltimore D., Boyer H.W., Cohen S.N., Davis R.W., Hogness D.S., Nathans D., Roblin R., Watson J.D., Weissman S. and Zinder N.D. (1974). "Letter: Potential biohazards of recombinant DNA molecules". *Science*. 185 (4148): 303.
13. Mousavi S.M., Tavakkoli-Moghaddam R., Hashemi H. and Mojtahedi S.M.H. (2011). A Novel approach based on non-parametric resampling with interval analysis for large engineering project risks. *Safety Science*. 49:340-1348.
14. Codex. (2009). Joint FAO=WHO food standards program –codex committee on methods of analysis and sampling, thirtieth session (9-13 March 2009). Proposed draft guidelines on criteria for methods for the detection and identification of foods derived from biotechnology. [ftp://ftp.fao.org/codex/xxmas30/ma30\\_08e.pdf](ftp://ftp.fao.org/codex/xxmas30/ma30_08e.pdf).
15. Council Regulation (EC). (2000). No 49/2000. *Off. J. Euro. Comm*. L6:13-14.
16. Efron B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*. 68, 589e599.
17. Bellocchi G., Confalonieri R., Acutis M. and Genovese G. (2006). SISSI: a resampling based Software for sample size determination. *Proc. 9th European Society for Agronomy Congress*, 4-6 September, Warsaw, Poland. 741-742.
18. CBD Secretariat. (2011). Central and Eastern European Regional Training of Trainers'. <http://www.nib.si/eng/index.php/novice/296-central-and-eastern-european-regional-training-of-trainers-workshop-on-the-identification-and-documentation-of-living-modified-organisms-under-the-cartagena-protocol-on-biosafety.11-15>.
19. Tsegaye T. and Hill R.L. (1998). Intensive tillage effects on spatial variability of soil physical properties. *Soil Science*. 163:143-154.

20. Efron B. and Tibshirani R. (1991). Statistical data analysis in the computer age. *Science*. 253, 390e395.
21. Kobilinsky A. and Bertheau Y. (2005). Minimum cost acceptance sampling plans for grain control, with application to GMO detection. *Chemometrics and Intelligent Laboratory Systems*. 75: 189–200.
22. Paoletti C., Donatelli M., Kay S. and Van den Eede G. (2003). Simulating kernel lot sampling: the effect of heterogeneity on the detection of GMO contaminations. *Seed Science and Technology*. 31, 629e638.
23. Ramachandran K.M. and Tsokos C.P. (2009). *Mathematical Statistics with Applications*. Amsterdam; Boston: Academic Press.
24. Remund K., Dixon D., Wright D. and Holden L. (2001). Statistical considerations in seed purity testing for transgenic traits. *Seed Science Research*. 11:101-119.

## Software SISSI to Determine the Optimum Sample Size of GM

Zahra Hajibarat<sup>1</sup>, Masoud Tohidfar<sup>2\*</sup>

1- Ph.D student of Agriculture Biotechnology Department, Faculty of Life Science, University of Shahid Beheshti, G.C Iran.

2- Associate Professor, Agriculture Biotechnology Department, Faculty of Life Science, University of Shahid Beheshti, G.C, Iran

gtohidfar@yahoo.com

### Abstract:

Since the advent of genetic engineering in the 1970s, concerns arose about risks of this technology that led to enforce labeling of transgenic crops. Therefore, it seems, it is necessary to use valid methods to identify transgenic crops to reduce or minimize relative potential errors. The validity of such methods depends on the sampling and the uniformity of distribution of transgenic crop under study. In most cases, the sampling is a source of error. Therefore, some software such as SISSI is designed to reduce sampling errors. This software which works based on Jackknife method is a new tool to determine optimum sample size for the experimental data set. The method estimates the population variance when there is little information about the population. The software graphically displays results as also changes in the mean and standard deviation of the population. Moreover, it has the option to determine the sample size automatically. One of its applications of this is to determine optimal sample size for transgenic shipments that are not uniform. The detection of transgenic material in commercial samples with high sensitivity in the international trade is another instance of its applications.

**Keywords:** SISSI, Jackknife, Re-sampling, Average, Non-Parametric.