فصلنامه تخصصی تحقیقات حشره‌شناسی
(علمی- پژوهشی)

جلد ۳، شماره ۱، سال ۱۳۹۰، (۵۱-۶۱)

دانشگاه آزاد اسلامی، واحد اراک
شاپا ۴۶۶۸-۲۰۰۸
www.entomologicalresearch.ir

تعیین تعداد نمونه مورد نیاز در مطالعات حشره‌شناسی با زبان محاسبه آماری R
قسمت اول- مقایسه میانگین‌ها و نسبت‌ها

*جواد کریم‌زاده اصفهانی**

بخش تحقیقات گیاه‌پزشکی، مرکز تحقیقات کشاورزی و منابع‌طبیعی استان اصفهان

چکیده

بسیاری از مطالعات حشره‌شناسی قادر به ردیابی اثرات معنی‌دار نمی‌باشند و اغلب تنها باعث اتلاف وقت و هزینه می‌گردند. در این آزمایشات تعداد نمونه گرفته‌شده اغلب از تعداد نمونه‌ای که یک آماردان توصیه می‌کند کمتر است. مقاله حاضر با بیان اهمیت قدرت آماری در تعیین حداقل تعداد نمونه لازم برای ردیابی اثرات معنی‌دار، اذعان می‌دارد که عدم موفقیت تعدادی از آزمایش‌ها تنها بخاطر عدم توجه به پیامدهای قدرت آماری در مرحله طراحی است. در مرحله طراحی آزمایش باید تعداد نمونه برای تست فرضیه مورد نظر تعیین گردد که خود بستگی به عواملی چون واریانس داده‌ها، مقدار تفاوت معنی‌دار قابل‌ردیابی در داده‌ها، ریسک خطای نوع اول و ریسک خطای نوع دوم دارد. در این نوشتار چگونگی تعیین قدرت آماری، تعداد نمونه مناسب و عوامل تاثیرگذار بر آن بیان می‌گردد. بعلاوه، روش‌های اختصاصی محاسبه تعداد نمونه مناسب برای موارد مهم در مطالعات صحرایی و آزمایشگاهی حشره‌شناسی با استفاده از زبان محاسبه آماری R معرفی می شود.

واژه‌های کلیدی: تعداد نمونه، قدرت آماری، تکرار، R

* نویسنده رابط، پست الکترونیکی: *jkisfahani@yahoo.com*

تاریخ دریافت مقاله (۸۹/۶/۲۱) - تاریخ پذیرش مقاله (۸۹/۱۱/۲۳)

مقدمه

اهداف پژوهش‌های مختلف می توانند بسیار متفاوت باشند. بطور مثال، هدف یک مطالعه می تواند اثبات تفاوت یا شباهت بین دو گروه و یا برآورد بعضی کمیت‌ها در یک جمعیت مشخص با درجه معینی از دقت باشد. صرف‌نظر از انگیزه یک تحقیق، تعداد نمونه باید بطور مناسب درنظر گرفته شود تا هدف تحقیق برآورده شود. با این حال می‌توان گفت که عمومی‌ترین هدف در پژوهش‌ها، تعیین مقداری تفاوت بین دو گروه می‌باشد. تفاوت بین دو گروه در یک مطالعه معمولا بصورت برآوردی از اثر، فاصله اطمینان مناسب و مقدار *P* جستجو می گردد. فاصله اطمینان نشان دهنده دامنه احتمالی مقادیر اثر حقیقی در جمعیت است در حالیکه مقدار *P* تعیین می کند که چقدر احتمال دارد که اثر مشاهده‌شده در نمونه ناشی از شانس باشد. یک کمیت مرتبط در اینجا قدرت آماری پژوهش است که احتمال شناسایی صحیح یک تفاوت بین دو گروه در نمونه مورد مطالعه را نشان می دهد. البته در صورتی‌که این تفاوت در جمعیتی که از آن نمونه گیری شده است وجود داشته باشد. مطالعه ایده‌آل برای یک محقق، مطالعه‌ای است که قدرت آن بالا باشد. این بدین معنی است که در چنین مطالعه‌ای، شانس بالایی وجود داشته باشد که یک تفاوت بین گروه‌ها ردیابی شود البته اگر چنین تفاوتی در واقعیت وجود داشته باشد. بالتبع اگر چنین مطالعه‌ای نشان دهد که هیچ تفاوتی بین گروه‌ها وجود ندارد محقق می تواند در نتیجه گیری اینکه "هیچ تفاوتی در واقعیت وجود ندارد" مطمئن باشد (Boitani & Fuller, 2000; Whitley & Ball, 2002).

قدرت یک مطالعه به عوامل متعددی بستگی دارد ولی به عنوان یک قانون عمومی، قدرت بالاتر با تعداد نمونه بیشتر حاصل می‌گردد. آگاهی از این مسئله بسیار مهم است زیرا که بسیاری از مطالعات انجام شده آنقدر کوچک هستند که قدرت کافی برای ردیابی اثر فرض شده را ندارند. به عبارت دیگر، حتی وقتی که یک تفاوت در واقعیت وجود دارد تعداد بسیار کمی از مطالعات می توانند به آن پی‌ببرند. در نتیجه، در چنین مطالعاتی مقدار *P* بالاتر و فاصله اطمینان بزرگتر در مقایسه با مطالعات بزرگتر بدست می‌آید و نتیجه گیری نادرست "هیچ تفاوتی بین گروه ها نیست" حاصل می‌گردد. این پدیده در این اصطلاح خلاصه می‌گردد که "عدم وجود مدرک، مدرک عدم وجود نیست". به عبارت دیگر، یک نتیجه غیر معنی‌دار بین گروه‌ها می‌تواند بطور ساده به‌علت فقدان قدرت آماری باشد که باعث می‌شود شناسایی درست یک تفاوت حقیقی بسیار نامحتمل گردد. با روشن شدن اهمیت مسئله فوق، بسیار حیرت آور است که محققین اغلب هیچ گونه محاسبه اصولی برای تعیین تعداد نمونه قبل از شروع یک مطالعه انجام نمی‌دهند. در عوض، تصمیم‌گیری در انتخاب تعداد نمونه بیشتر بر اساس راحتی، منابع موجود و یا موضوعات سهل‌الوصول انجام می‌گیرد (Crawley, 2002; Whitley & Ball, 2002; Crawley, 2005).

زبان محاسبه آماری[1] R از زبان S استخراج گردیده است. زبان S یک ابزار قوی برای مدلسازی آماری[2]، اکتشاف داده[3] و ارزیابی عبارات پیچیده حسابی[4] است و همچنین یک زبان برنامه نویسی شیءگرای[5] بسیار انعطاف‌پذیر و عمومی برای داده‌های بسیار جامع می‌باشد. وقتی که محصول تجاری S بنام S-PLUS به بازار آمد بسیار گران بود. برای حل این مشکل، دو آماردان نیوزلندی بنامهای Ross Ihaka و Robert Gentleman از دانشگاه Auckland تصمیم گرفتند که نسخه

[1] language of statistical computing
[2] statistical modelling
[3] data exploration
[4] complex arithmetic expressions
[5] object-oriented

ای کوتاه شده از S را برای اهداف آموزشی بنویسند. به‌علت اینکه حرف R قبل از S در حروف الفبا می‌آید و اسم کوچک هر دو نویسنده نیز با R شروع می‌شد نام برنامه را R گذاشتند. نسخه 1.0.0 در ۲۹ فوریه سال ۲۰۰۰ پخش گردید و سپس نسخه های دیگر آن بیرون آمد. لازم به ذکر است که R یک برنامه مجانی است و می‌تواند از تارنمای http:/cran.r-project.org گرفته شود (Crawley, 2005).

تعداد نمونه[1]

مهمترین مفاهیم اصلی که در طراحی یک آزمایش باید در نظر گرفته شوند تکرار[2] و تصادفی‌کردن[3] می‌باشند. تکرار برای افزایش اطمینان[4] به برآورد پارامترها انجام می‌گیرد در حالیکه تصادفی کردن به جهت کاهش اریب[5] (انحراف بین مقدار برآورد شده و مقدار حقیقی یک پارامتر) انجام می‌گردد. تکرار به اندازه‌گیری‌های مکرری گفته می‌شود که مستقل باشند، قسمتی از یک سری زمانی[6] نباشند (داده‌های جمع‌آوری‌شده از یک مکان در مواقع پیاپی مستقل نیستند)، با یکدیگر در یک مکان جمع نشوند (توده شدن تکرارها بدین معنی است که آنها از نظر فضایی مستقل نیستند) و دارای یک مقیاس فضایی[7] مناسب باشند. اندازه‌گیری‌های مکرر از یک فرد یا یک مکان را نمی‌توان تکرار محسوب نمود و این شاید متداول‌ترین سبب تکرار کاذب[8] در کارهای آماری باشد (Crawley, 2005). حال سئوال این است که چه تعداد تکرار (n) کافی است؟ جواب معمول این است که هرچقدر امکانات اجازه دهد. جواب دیگر تعداد ۳۰ تکرار است. یک قائده کلی[9] بسیار مفید این است که یک نمونه ۳۰ تایی یا بیشتر نمونه‌ای بزرگ است ولی یک نمونه کمتر از ۳۰، نمونه ای کوچک است (Crawley, 2002).

البته باید توجه نمود که این قانون در بسیاری موارد جواب نمی‌دهد و این تعداد نمونه ممکن است در بسیاری موارد کم باشد و یا در بعضی شرایط هزینه‌های سنگین مانع تکرار ۳۰ تایی آزمایش شود. در این حالت راه‌هایی برای محاسبه تعداد تکرار لازم برای تست یک فرضیه معین وجود دارند. در واقع، سئوال اساسی در مرحله طراحی این است که چه تعداد نمونه برای تست فرضیه مورد نظرمان نیاز داریم؟ جواب این سئوال بستگی به عواملی چون واریانس داده‌ها، مقدار تفاوت معنی‌دار قابل‌ردیابی در داده‌ها، ریسک خطای نوع اول و ریسک خطای نوع دوم دارد. در این ارتباط، هرچه واریانس بزرگتر باشد عدم قطعیت[10] مقادیر (پارامترها) برآورد شده از داده ها بیشتر است. به‌عبارت دیگر، برای بدست آوردن سطح خاصی از اطمینان، هرچه واریانس بزرگتر باشد تعداد نمونه بیشتری مورد نیاز خواهد بود. در اینجا ضروری است که مقداری اطلاعات هرچند تقریبی در مورد مقدار واریانس داده‌ها داشته باشیم. علاوه بر این، مقدار تفاوت مورد نظر در داده‌ها که قابل ردیابی بصورت معنی‌دار باشد هرچه کمتر باشد به تعداد نمونه بیشتری نیاز خواهد بود. درعمل، آزمایشات برای ردیابی ۲۵٪ یا ۵۰٪ تفاوت در میانگین داده‌ها طراحی می‌گردند زیرا تفاوت‌های کمتر اغلب نتیجه تغییرات

[1] sample size; number of samples
[2] replication
[3] randomisation
[4] reliability
[5] bias
[6] time series
[7] spatial scale
[8] pseudoreplication
[9] rule of thumb
[10] uncertainty

پیشینه‌ای، نظیر ناهمگنی محیط[1] می‌باشند و تامین تعداد تکرار لازم برای ردیابی تفاوت‌های بسیار کمتر از این مقدار مستلزم صرف هزینه‌های سنگین می‌باشد (Crawley, 2002; Whitley & Ball, 2002; Crawley, 2005; Crawley, 2007).

خطای نوع اول[2] به "رد کردن یک فرضیه صفر درست" اطلاق می‌گردد و احتمال آن را با α نشان می‌دهند. خطای نوع دوم[3] به "پذیرش یک فرضیه صفر غلط" اطلاق می‌گردد و احتمال آن را با β نشان می‌دهند. عدم موفقیت تعدادی از آزمایش‌ها تنها بخاطر عدم توجه به پیامدهای قدرت آماری[4] در مرحله طراحی است. قدرت یک آزمون[5] عبارت است از احتمال رد کردن فرضیه صفر غلط. بنابراین، قدرت آماری در ارتباط با خطای نوع دوم می‌باشد که در آن β احتمال پذیرش فرضیه صفر غلط است. هرچه β کوچکتر باشد ایده‌آل‌تر است ولی باید به این نکته توجه داشت که هرچه احتمال ارتکاب خطای نوع دوم (پذیرش فرضیه صفر غلط) را کم کنیم احتمال ارتکاب خطای نوع اول (رد کردن فرضیه صفر درست) را افزایش می‌دهیم. اکثر آماردان‌ها β را معادل ۰/۲ و α را معادل ۰/۰۵ می‌گیرند. قدرت یک آزمون بصورت $1-\beta$ تعیین می‌گردد و تحت فرض‌های استاندارد معادل ۰/۸ می‌باشد. آنالیز قدرت آماری[6] می‌تواند برای محاسبه حداقل تعداد نمونه لازم برای ردیابی اثرات معنی‌دار بکار رود. درعمل، تعداد نمونه اغلب تحت تاثیر ملاحظاتی نظیر منابع موجود، تعداد افراد انجام‌دهنده کار، مقدار فضای موجود و تعداد حیوانات آزمایشی قابل‌تهیه قرار می‌گیرد. این ملاحظات باعث می‌گردد که تعداد نمونه گرفته‌شده از تعداد نمونه‌ای که یک آماردان توصیه می‌کند کمتر باشد. متاسفانه چنین آزمایشاتی قادر به رد فرضیه موردنظر نمی‌باشند و اغلب تنها باعث اتلاف هزینه می‌گردند (Boitani & Fuller, 2000; Southwood & Henderson, 2000; Whitley & Ball, 2002; Crawley, 2005).

محاسبه تعداد نمونه

محاسبه تعداد نمونه برای ردیابی تفاوت بین میانگین‌ها

هدف در اینجا تعیین تعداد نمونه‌ای (n) است که با آن یک تفاوت معین بصورت معنی‌دار قابل ردیابی باشد مشروط به اینکه واریانس نمونه معین باشد. در مورد تک‌نمونه[7]، با واریانس s^2 و تفاوت معنی دار δ خواهیم داشت (Karandinos, 1976; Pedigo, 1991; Boitani & Fuller, 2000; Southwood & Henderson, 2000; Crawley, 2002; Whitley & Ball, 2002; Crawley, 2005):

$$n=\left(\frac{s(Z_\alpha+Z_{1-\beta})}{\delta}\right)^2 \quad \text{(معادله ۱)}$$

این فرمول نشان می‌دهد هرچه تفاوتی که می خواهیم ردیابی کنیم (δ) کوچکتر باشد تعداد نمونه بیشتری نیاز خواهیم داشت. همچنین، تعداد نمونه با بزرگ شدن قدرت آزمون ($1-\beta$) و انحراف معیار نمونه (s) افزایش می یابد. مقدار Z از جدول های پراکنش نرمال استاندارد[8] بدست می آید و در حقیقت می توان معادله (۱) را بصورت زیر خلاصه کرد:

---
[1] environmental heterogeneity
[2] Type I error
[3] Type II error
[4] statistical power
[5] power of the test
[6] statistical power analysis
[7] the single-sample case
[8] the standard Normal distribution

$$n = \left(\frac{s^2}{\delta^2}\right) C_{p,power} \quad \text{(معادله ۲)}$$

که در آن ثابت $C_{p,power}$ برابر است با:

$$C_{p,power} = (Z_\alpha + Z_{1-\beta})^2$$

این مقدار در محیط R برای $\alpha = 0.05$ و $\beta = 0.2$ بترتیب بصورت زیر محاسبه می گردد:

```
(qnorm(1-0.025)+qnorm(1-0.2))^2
[1] 7.84888
```

به علت این که $(Z_{0.05} + Z_{0.8})^2 \approx 8$، از اینرو برای مورد تک‌نمونه، تعداد نمونه مورد نیاز را می‌توان از فرمول ساده‌شده زیر محاسبه کرد:

$$n = 8 \times \frac{s^2}{\delta^2} \quad \text{(معادله ۳)}$$

در مورد دونمونه[1]، اگر نمونه‌ها دارای واریانس‌های متفاوت (s_1^2 و s_2^2) باشند و تکرارهای متفاوت در دو نمونه (n_1 و n_2) مدنظر باشد بطوریکه $n_2 = \omega n_1$ و ω نسبت تکرار در نمونه دوم در مقایسه با نمونه اول باشد اول n برای نمونه اول طبق روش فوق، با استفاده از معادله (۲)، محاسبه شده و سپس با جایگزین کردن آن در معادله زیر:

$$n_1 = n\left[1 + \frac{s_2^2}{\omega s_1^2}\right] \quad \text{(معادله ٤)}$$

تعداد نمونه در نمونه اول بصورت زیر بدست می‌آید:

$$n_1 = \left(\frac{s_1^2 + \frac{s_2^2}{\omega}}{\delta^2}\right) C_{p,power} \quad \text{(معادله ٥)}$$

تعداد نمونه در نمونه دوم نیز از فرمول $n_2 = \omega n_1$ محاسبه می گردد (Crawley, 2002; Whitley & Ball, 2002). حال اگر فرض کنیم که $n_1 = n_2$ (یعنی $\omega = 1$) و $s_1^2 = s_2^2$، داریم:

$$n_1 = n_2 = 2\left(\frac{s^2}{\delta^2}\right) C_{p,power} \quad \text{(معادله ٦)}$$

محاسبه تعداد نمونه برای ردیابی تفاوت بین نسبت‌ها (گروه‌های با اندازه یکسان)

برای محاسبه تعداد نمونه مورد نیاز برای مقایسه نسبت‌ها در دو گروه با اندازه یکسان از فرمول زیر استفاده می کنیم (Whitley & Ball, 2002):

$$n = \frac{p_1(1-p_1) + p_2(1-p_2)}{(p_1 - p_2)^2} \times C_{p,power} \quad \text{(معادله ۷)}$$

در اینجا p_1 و p_2 نسبت ها در دو گروه هستند.

[1] the two-sample case

محاسبه قدرت و تعداد نمونه در محیط R

آزمون های *t* تک‌نمونه یا دونمونه[1]

در اینجا از تابع power.t.test بصورت زیر استفاده می کنیم:

```
power.t.test(n=NULL, delta=NULL, sd=1, sig.level=0.05, power=NULL, type=c("two.sample",
"one.sample","paired"), alternative=c("two.sided","one.sided"), strict = FALSE)
```

این تابع دارای پارامترهای n (تعداد مشاهدات در گروه)، delta (تفاوت حقیقی در میانگین‌ها؛ مقدار تفاوت در میانگین که می‌خواهیم قابل‌ردیابی باشد)، sd (انحراف معیار نمونه)، sig.level (سطح معنی‌داری یا احتمال خطای نوع اول؛ اغلب مقدار پیش‌فرض[2] ۵٪ پذیرفته می‌شود)، power (قدرت آزمون؛ اغلب مقدار پیش فرض ۸۰٪ پذیرفته می‌شود)، type (نوع آزمون *t* که اجرا می‌شود ازجمله تک‌نمونه[3]، دونمونه[4] و جفتی[5])، alternative (آزمون یک‌طرفه[6] یا دوطرفه[7]؛ بطور معمول، پیش‌فرض بر اساس آزمون دوطرفه است) و strict (برای مورد دوطرفه بکار می‌رود) می‌باشد. یکی از پارامترهای n، delta، power ، sd و sig.level در تابع بصورت NULL قرار داده می‌شود (یعنی مقدار آن تعیین نمی‌گردد) که توسط تابع از پارامترهای دیگر تعیین می‌شود. توجه کنید که برای محاسبه پارامترهای n، delta و power به جای مساوی NULL قرار دادن (بطور مثال power = NULL)، می توان آنها را از تابع حذف نمود ولی پارامتر های sd و sig.level به علت این‌که دارای پیش‌فرض‌های غیر NULL هستند (یعنی اگر از تابع حذف شوند پیش فرض‌های sd = 1 و sig.level = 0.05 در محاسبه بکار می‌روند) برای محاسبه آنها حتما باید مساوی NULL قرار گیرند (Crawley, 2005; Crawley, 2007). با وجود اینکه به‌نظر می‌رسد محاسبات متعددی نیاز داریم ولی به طور نمونه با کاربرد تمام پارامترهای پیش‌فرض تنها نیاز داریم که delta و sd را مشخص کنیم تا تعداد نمونه را برای قدرت مورد نظر بدست آوریم.

بطور مثال، فرض کنید که می‌خواهیم تعداد نمونه را برای ارزیابی جمعیت پوره سن گندم در دو منطقه ورامین و اصفهان محاسبه کنیم. در یک مطالعه مقدماتی مشخص می گردد که جمعیت پوره سن گندم در اصفهان دارای میانگین ۲۰ عدد در متر مربع و انحراف معیار در حدود ۳/۵ می‌باشد. حال دستور تعیین تعداد نمونه لازم (تعداد تکرار لازم برای هریک از دو نمونه) برای ردیابی یک تفاوت ۱۰٪ (۱۰ درصد از میانگین یعنی ۲) با قدرت ۸۰٪ بصورت زیر نوشته می-شود:

```
power.t.test(delta=2,sd=3.5,power=0.8)
```

که بصورت زیر توسط R محاسبه می گردد:

```
Two-sample t test power calculation
n = 49.05349
delta = 2
sd = 3.5
sig.level = 0.05
power = 0.8
```

[1] one- and two-sample *t*-tests
[2] default
[3] one-sample
[4] two-sample
[5] paired
[6] one-sided
[7] two-sided

```
alternative = two.sided
NOTE: n is number in *each* group
```

یعنی برای ردیابی تفاوت معنی‌دار بین جمعیت سن گندم در اصفهان (با میانگین ۲۰ عدد در متر مربع) و جمعیت سن گندم در ورامین (با میانگین ۱۸ یا ۲۲ عدد در متر مربع) باید یک نمونه حداقل ۵۰ تایی در هر منطقه داشته باشیم. در اینجا می بینم که اگر از قانون "۳۰ نمونه به اندازه کافی بزرگ است" استفاده کرده بودیم به شدت مایوس می شدیم زیرا که با تعداد نمونه ۳۰ نمی توانستیم یک تفاوت ۱۰٪ را در این آزمایش ردیابی کنیم. در این مورد نیاز به ۵۰ تکرار در هر نمونه می باشد تا به یک قدرت ۸۰٪ دست پیدا کنیم. اگر بخواهیم ببینیم که چه مقدار تفاوت را با نمونه ۳۰ تایی می توان ردیابی کرد باید در تابع مورد نظر n را تعیین و delta را حذف کنیم یعنی:

```
power.t.test(n=30,sd=3.5,power=0.8)
```

که بصورت زیر توسط R محاسبه می گردد:

```
Two-sample t test power calculation
n = 30
delta = 2.574701
sd = 3.5
sig.level = 0.05
power = 0.8
alternative = two.sided
NOTE: n is number in *each* group
```

که نشان می‌دهد حدود ۱۳٪ ($2.575/20 = 0.12875$) تفاوت را می‌توان ردیابی کرد.

باید توجه نمود که قبل از آنالیز قدرت آزمون (که خود قبل از طراحی آزمایش انجام می‌شود)، مقدار انحراف معیار از مقالات یا با انجام یک آزمایش راهبر[1] مشخص شود و همچنین اندازه تفاوت قابل‌ردیابی بر اساس منابع مالی و تجربه تعیین گردد.

تعداد تکرار لازم برای مورد تک‌نمونه با همان پارامترهای فوق با دستور زیر محاسبه می‌گردد:

```
power.t.test(type="one.sample",delta=2,sd=3.5,power=0.8)
```

که خروجی R بصورت زیر می‌باشد:

```
One-sample t test power calculation
n = 26.02322
delta = 2
sd = 3.5
sig.level = 0.05
power = 0.8
alternative = two.sided
```

آزمون‌های ANOVA یک‌طرفه متوازن[2]

در اینجا از تابع power.anova.test بصورت زیر استفاده می کنیم:

[1] pilot experiment
[2] balanced one-way ANOVA tests

```
power.anova.test(groups=NULL, n=NULL, between.var=NULL, within.var=NULL, sig.level=0.05,
power=NULL)
```

این تابع دارای پارامترهای groups (تعداد گروه‌ها)، n (تعداد مشاهدات در گروه)، between.var (واریانس بین‌-گروهی)، within.var (واریانس درون‌گروهی)، sig.level (سطح معنی‌داری یا احتمال خطای نوع اول) و power (قدرت آزمون) می‌باشد. یکی از پارامترها در تابع بصورت NULL قرار داده می شود که توسط تابع از پارامترهای دیگر تعیین می‌شود. البته باید توجه نمود که پارامتر sig.level به علت اینکه دارای پیش فرض غیر NULL است (sig.level = 0.05) برای محاسبه آن حتما باید مساوی NULL قرار گیرد. قابل ذکر است که برای محاسبه واریانس‌های بین‌گروهی و درون-گروهی از جدول تجزیه واریانس استفاده می‌شود. بدین ترتیب که واریانس درون‌گروهی، همان میانگین مربعات خطا می-باشد و واریانس بین‌گروهی با تقسیم میانگین مربعات تیمار به n بدست می‌آید. البته می‌توان واریانس بین‌گروهی را با محاسبه واریانس میانگین گروه‌ها نیز بدست آورد (Crawley, 2005; Crawley, 2007).

بطور مثال، فرض کنید که می‌خواهیم تعداد نمونه را برای ارزیابی تراکم جمعیت لارو بید کلم بر روی سه گونه گیاه میزبان (کلم‌چینی، کلم‌معمولی و کلم‌گل) محاسبه کنیم. در یک مطالعه مقدماتی با شمارش لاروهای موجود بر روی ۵ بوته، تعداد ٤، ٤، ٦، ٨ و ۹ لارو برروی کلم‌چینی، تعداد ۲، ٤، ٤، ٤ و ۵ لارو برروی کلم‌معمولی و تعداد ۱، ۲، ۲، ٤ و ۵ لارو برروی کلم‌گل مشاهده می‌گردد. اگر این داده‌ها را به روش ANOVA یک‌طرفه آنالیز کنیم و جدول تجزیه واریانس را تشکیل دهیم مشخص می گردد که میانگین مربعات تیمار و خطا بترتیب ۱۵/۲٦٦۷ و ۳/۰۳۳۳ می‌باشند. بنابراین، واریانس درون‌گروهی ۳/۰۳۳۳ و واریانس بین‌گروهی ۳/۰۵۳۳ (یعنی ۱۵/۲٦٦۷ تقسیم بر ۵) می‌باشند. حال دستور تعیین تعداد نمونه لازم (تعداد تکرار لازم برای هریک از سه گروه) برای ردیابی یک تفاوت معنی‌دار (با سطح معنی‌داری ۱٪) و قدرت آزمون ۹۰٪ بصورت زیر نوشته می‌شود:

```
power.anova.test(groups=3,between.var=3.0533,within.var=3.0333,sig.level=0.01,power=0.9)
```

ویا بصورت ساده شده زیر:

```
power.anova.test(3,NULL,3.0533,3.0333,0.01,0.9)
```

که بصورت زیر توسط R محاسبه می گردد:

```
Balanced one-way analysis of variance power calculation
groups = 3
n = 10.28016
between.var = 3.0533
within.var = 3.0333
sig.level = 0.01
power = 0.9
NOTE: n is number in each group
```

در اینجا مشخص می‌گردد که حداقل ۱۱ بوته از هر گیاه میزبان باید نمونه‌گیری شود تا به هدف مورد نظر دست پیدا کنیم.

آزمون تناسبی دونمونه[1]

در اینجا از تابع power.prop.test بصورت زیر استفاده می کنیم:

```
power.prop.test(n=NULL, p1=NULL, p2=NULL, sig.level=0.05, power=NULL, alternative=
c("two.sided","one.sided"), strict=FALSE)
```

این تابع دارای پارامترهای n (تعداد مشاهدات در گروه)، p_1 (احتمال در یک گروه)، p_2 (احتمال در گروه دیگر)، sig.level (سطح معنی‌داری یا احتمال خطای نوع اول)، power (قدرت آزمون)، alternative (آزمون یک‌طرفه یا دوطرفه) و strict (برای مورد دوطرفه بکار می‌رود) می‌باشد. یکی از پارامترهای n، p_1، p_2، sig.level و power در تابع بصورت NULL قرار داده می‌شود که توسط تابع از پارامترهای دیگر تعیین می‌شود. البته باید توجه نمود که پارامتر sig.level به علت اینکه دارای پیش فرض غیر NULL است (sig.level = 0.05) برای محاسبه آن حتما باید مساوی NULL قرار گیرد (Crawley, 2005; Crawley, 2007).

بطور مثال، فرض کنید در یک مطالعه درصد خسارت کرم گلوگاه انار روی دو رقم انار (متحمل و حساس) ارزیابی می‌گردد. در مطالعه مقدماتی با بررسی میوه‌های موجود بر روی ۱۰ درخت از هر رقم، مشخص می‌شود که خسارت بر روی دو رقم مورد مطالعه در حدود ۲۰ و ۵۰ درصد می‌باشد. حال، دستور محاسبه تعداد نمونه لازم در هر رقم برای ردیابی یک تفاوت معنی دار با قدرت ۸۰٪ بصورت زیر نوشته می شود:

```
power.prop.test(p1=0.2,p2=0.5,power=0.8)
```

که بصورت زیر توسط R محاسبه می گردد:

```
Two-sample comparison of proportions power calculation
n = 38.48004
p1 = 0.2
p2 = 0.5
sig.level = 0.05
power = 0.8
alternative = two.sided
NOTE: n is number in *each* group
```

بنابراین حداقل ۳۹ درخت از هر رقم باید نمونه‌گیری شود تا تفاوت معنی‌دار در صورت وجود ردیابی گردد.

سپاسگزاری

در اینجا لازم است از زحمات اساتید بزرگوار بخصوص آقایان دکتر رحیم عبادی و دکتر حسین سیدالاسلامی (دانشگاه صنعتی اصفهان)، دکتر عزیز خرازی پاکدل (دانشگاه تهران) و دکتر غلامرضا رجبی (موسسه تحقیقات گیاه-پزشکی کشور) که راهنما و مشوق اینجانب برای مطالعه و تحقیق در علوم حشره‌شناسی و اکولوژی بوده‌اند صمیمانه قدردانی نمایم. از استاد گرانقدر دکتر میک کرولی (Mick Crawley؛ امپریال کالج لندن) که یادگیری آمار و کار با نرم‌افزار R را مدیون ایشان هستم نیز سپاس‌گزارم.

[1] two-sample test for proportions

References

Boitani ,L. and Fuller, T. K. 2000. Research Techniques in Animal Ecology: Controversies and Consequences. Columbia University Press, New York, USA. 442 pp.

Crawley ,M. J. 2002 . Statistical Computing: An Introduction to Data Analysis Using S-Plus .John Wiley and Sons, Chichester, UK. 761pp.

Crawley ,M. J. 2005. Statistics: An Introduction Using R .John Wiley and Sons, Chichester, UK. 327pp.

Crawley ,M. J. 2007. The R Book. John Wiley and Sons, Chichester, UK. 942pp.

Karandinos, M. G. 1976. Optimum sample size and comments on some published formulae. Bulletin of the Entomological Society of America, 22: 417-421.

Pedigo ,L. P. 1991. Entomology and Pest Management. Macmillan Publishing Company, New York, USA, 356 pp .

Southwood ,T. R. E. and Henderson, P. A. 2000. Ecological Methods. Blackwell Science, Oxford, UK. 575pp.

Whitley ,E. and Ball, J. 2002. Statistics review 4: sample size calculations. Critical Care, 6: 341-335.

Journal of Entomological Research

Volume 3, Issue 1, pages: 51-61

Islamic Azad University, Arak Branch
ISSN 2008-4668
www.entomologicalresearch.ir

Sample size calculation in entomological studies using R language of statistical computing
Part 1 :comparison of means and proportions

Javad Karimzadeh[*]

Department of Plant Protection, Isfahan Research Center for Agriculture and Natural Resources, Isfahan, Iran

Abstract

In many entomological studies, a smaller sample is taken than a statistician would advise. Clearly, such experiments would often fail to detect any significant effects, resulting in a complete waste of time and money. The present paper introduces statistical power of a test and its importance in determining the minimum sample size required for detecting significant effects. A number of experiments fail simply because the issues of statistical power are not confronted with at the planning stage. A fundamental question that is needed to address at the planning stage concerns the number of samples required to test the hypothesis of interest. The answer depends on the variance of response variable, the size of difference in the response variable that would be detectable as significant, the risk of a Type I error and the risk of a Type II error. Here using R language of statistical computing, the specific methods of the calculation of optimum sample size for laboratory and field studies are presented for two common scenarios.

Keywords :sample size ,statistical power, replicate, R

[*] Corresponding Author, E-mail: *jkisfahani@yahoo.com*
Received: 12 Sep. 2010 - Accepted: 12 Feb. 2011