

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

جواد کریم‌زاده اصفهانی*

اصفهان، مرکز تحقیقات کشاورزی و منابع طبیعی استان اصفهان، بخش تحقیقات گیاه‌پزشکی

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 می‌شود.

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

مقدمه

اهداف پژوهش‌های مختلف می‌توانند بسیار متفاوت باشند. بطور مثال، هدف یک مطالعه می‌تواند 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 در نظر 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 معمولاً بصورت برآوردی از اثر، فاصله اطمینان مناسب و مقدار P جستجو می‌گردد. فاصله اطمینان نشان دهنده دامنه احتمالی مقادیر اثر حقیقی در جمعیت است در حالیکه مقدار P تعیین می‌کند که چقدر احتمال دارد که اثر مشاهده‌شده در نمونه ناشی از شانسی باشد. یک کمیت مرتبط در اینجا قدرت آماری پژوهش است که احتمال شناسایی صحیح یک

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

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



تفاوت بین دو گروه در نمونه مورد مطالعه را نشان می‌دهد. البته در صورتی که 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 می‌تواند در نتیجه گیری اینکه "هیچ تفاوتی در واقعیت وجود ندارد" مطمئن باشد (Whitley & Boitani, 2000; Fuller & 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 می‌توانند به آن پی‌برند. در نتیجه، در چنین مطالعه‌ای مقدار P بالاتر و فاصله اطمینان بزرگتر در مقایسه با مطالعات بزرگتر به دست می‌آید و نتیجه گیری نادرست "هیچ تفاوتی بین گروه‌ها 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 به علت 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^۱ R از زبان S استخراج گردیده است. زبان S یک ابزار قوی برای مدل‌سازی آماری^۲، اکتشاف داده^۳ و ارزیابی عبارات پیچیده حسابی^۴ است و همچنین یک زبان برنامه نویسی شیء‌گرای^۵ 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 Ihaka Ross و Gentleman Robert از دانشگاه Auckland تصمیم گرفتند که نسخه ای کوتاه شده از S را برای اهداف آموزشی بنویسند. به علت 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 ۱.۰.۰ 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).

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

تعداد نمونه^۱

مهمترین مفاهیم اصلی که در طراحی یک آزمایش باید در نظر گرفته شوند تکرار^۲ و تصادفی کردن^۳ می‌باشند. تکرار برای افزایش اطمینان^۴ به برآورد پارامترها انجام می‌گیرد در حالیکه تصادفی کردن به جهت کاهش اریب^۵ (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^۶ نباشند (داده‌هایی که از یک مکان در مواقع پیاپی جمع‌آوری شده باشند 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^۷ مناسب باشند. اندازه‌گیری‌های مکرر از یک فرد یا یک مکان را نمی‌توان تکرار محسوب نمود و این شاید متداول‌ترین سبب تکرار کاذب^۸ در کارهای آماری باشد (Crawley, 2005). حال سؤال 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 فائده کلی^۹ بسیار مفید این است که یک نمونه ۳۰ تایی یا بیشتر نمونه‌ای بزرگ است ولی یک نمونه کمتر از ۳۰، نمونه‌ای 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، سؤال اساسی در مرحله طراحی این است که چه تعداد نمونه برای تست فرضیه مورد نظرمان نیاز داریم؟ جواب این سؤال 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

-
- 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

واریانس بزرگتر باشد عدم قطعیت^۱ مقادیر (پارامترها) برآورد شده از داده‌ها بیشتر است. به عبارت 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. در عمل، آزمایشات برای ردیابی ۲۵٪ یا ۵۰٪ تفاوت در میانگین داده‌ها طراحی می‌گردند زیرا تفاوت‌های کمتر اغلب نتیجه تغییرات پیشینه‌ای، نظیر ناهمگنی محیط^۲ می‌باشند و تامین تعداد تکرار لازم برای ردیابی تفاوت‌های بسیار کمتر از این مقدار مستلزم صرف هزینه‌های سنگین می‌باشد (Crawley, 2005; Ball & Whitley, 2002; Crawley, 2007).

خطای نوع اول^۳ به "رد کردن یک فرضیه صفر درست" اطلاق می‌گردد و احتمال آن را با α نشان می‌دهند. خطای نوع دوم^۴ 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^۵ در مرحله طراحی است. قدرت یک آزمون^۶ 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^۷ می‌تواند برای محاسبه حداقل تعداد نمونه لازم برای ردیابی اثرات معنی‌دار بکار رود. در عمل، تعداد نمونه اغلب به کل ملاحظات عملی نظیر منابع موجود، تعداد افراد انجام‌دهنده کار، مقدار فضای موجود و تعداد حیوانات آزمایشی قابل تهیه بستگی دارد. این ملاحظات باعث می‌گردد که تعداد نمونه گرفته‌شده از تعداد نمونه‌ای که یک آماردان توصیه می‌کند کمتر باشد. متأسفانه چنین آزمایشاتی قادر به رد فرضیه موردنظر نمی‌باشند و اغلب تنها باعث اتلاف هزینه می‌گردند (Fuller, 2000 & Boitani; Crawley, 2005; Ball, 2002 & Whitley; Henderson, 2000 & Southwood).

-
- 1- rule of thumb
 - 2- uncertainty
 - 3- environmental heterogeneity
 - 4- Type I error
 - 5- Type II error
 - 6- statistical power
 - 7- power of the test

محاسبه تعداد نمونه

محاسبه تعداد نمونه برای یک تفاوت در 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^۱، با واریانس s^2 و تفاوت معنی‌دار δ خواهیم داشت (Karandinos, ۱۹۷۶; Pedigo, 1991; Fuller, 2000 & Boitani; Henderson, 2000 & Southwood; Crawley, 2002; Whitley & Ball, 2002):

$$\beta \delta^2 - n = s^2 \alpha + Z1 \quad (\text{معادله ۱})$$

این فرمول نشان می‌دهد هرچه تفاوتی که می‌خواهیم 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 می‌یابد. مقدار Z از جدول‌های پراکنش نرمال استاندارد^۲ بدست می‌آید و در حقیقت می‌توان معادله (۱) را به صورت زیر خلاصه کرد:

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

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

$$C_{p,\text{power}} = (Z\alpha + Z1)^2$$

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

$$qnorm(1-0.025) + qnorm(1-0.2))^2$$

$$[1] \quad 7.84888$$

به علت این که $Z_{0.05} + Z_{0.2} \approx 8$ ، از اینرو برای مورد تک‌نمونه، تعداد نمونه مورد نیاز را می‌توان از فرمول ساده‌شده زیر محاسبه کرد:

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

در مورد دو نمونه^۳، اگر نمونه‌ها دارای واریانس‌های متفاوت ($s1^2$ و $s2^2$) باشند و تکرارهای متفاوت در دو نمونه ($n1$ و $n2$) مدنظر باشد بطوریکه $n2 = \omega n1$ و نسبت تکرار در نمونه دوم در مقایسه با نمونه اول باشد اول n را برای نمونه اول طبق روش فوق، با استفاده از معادله (۲)، محاسبه شده و سپس با جایگزین کردن آن در معادله زیر:

$$n1 = n1 + s2^2 \omega s1^2 \quad (\text{معادله ۴})$$

تعداد نمونه در نمونه اول بصورت زیر بدست می‌آید:

$$n1 = s1^2 + s2^2 \omega \delta^2 C_{p,\text{power}} \quad (\text{معادله ۵})$$

تعداد نمونه در نمونه دوم نیز از فرمول بطوریکه $n2 = \omega n1$ محاسبه می‌گردد (Crawley, ۲۰۰۲; Ball & Whitley, ۲۰۰۲). حال اگر فرض کنیم که $n2 = n1$ (یعنی $\omega = 1$) و $s1^2 = s2^2$ باشد داریم:

1- statistical power analysis
2- the single-sample case
3- the standard Normal distribution

$$n_1 = n_2 = 2s^2 \delta^2 C_p \text{power} \quad (\text{معادله ۶})$$

محاسبه تعداد نمونه برای یک تفاوت در نسبت‌ها (گروه‌های با اندازه یکسان)

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

$$C_p \text{power} \times 2(p_2 - p_1)^2 p_2^{-1} p_1^{-1} n = p_1 \quad (\text{معادله ۷})$$

در اینجا p_1 و p_2 نسبت‌ها در دو گروه هستند.

محاسبه قدرت و تعداد نمونه در محیط R

آزمون‌های t تک‌نمونه یا دونمونه^۱

در اینجا از تابع `power.t.test` بصورت زیر استفاده می‌کنیم:

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

این تابع دارای پارامترهای n (تعداد مشاهدات در گروه)، δ (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 ۰.۰۵ پذیرفته می‌شود)، power (قدرت آزمون؛ اغلب مقدار پیش‌فرض ۰.۸۰ پذیرفته می‌شود)، type (نوع آزمون t که اجرا می‌شود از جمله تک‌نمونه^۲، دونمونه^۳ و جفتی^۴)، alternative (آزمون یک‌طرفه^۵ یا دوطرفه^۶؛ 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`

1- the two-sample case
2- one- and two-sample t -tests
3- default
4- one-sample
5- two-sample
6- Paired

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 پیش فرض های غیر $NULL$ هستند (یعنی اگر از تابع حذف شوند پیش فرض های $sd = 1$ و $sig.level = 0.05$ در محاسبه بکار می روند) برای محاسبه آنها حتما باید مساوی $NULL$ قرار گیرند (Crawley, 2007). با وجود اینکه به نظر می رسد محاسبات متعددی نیاز داریم ولی به طور نمونه با کاربرد تمام پارامترهای پیش فرض 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 می خواهیم 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 می باشد. حال دستور تعیین تعداد نمونه لازم (تعداد تکرار لازم برای هر یک 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:

t test power calculation Two-sample

$49.05349 = n$

$2 = delta$

$3.5 = sd$

$0.05 = sig.level$

$0.8 = power$

two.sided = alternative

NOTE: n is number in *each* group

یعنی برای ردیابی تفاوت معنی دار 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) بایستی یک نمونه حداقل ۵۰ تایی در هر منطقه داشته باشیم. در اینجا می بینیم 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:

t test power calculation Two-sample

$30 = n$

$2.575701 = delta$

$3.5 = sd$

$0.05 = sig.level$

$0.8 = power$

two.sided = alternative

NOTE: n is number in *each* group

که نشان می‌دهد حدود ۱۳٪ (۰.۱۲۸۷۵=۲۰/۲.۵۷۵) 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^۱ 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:

$(\text{delta}=2, \text{sd}=3.5, \text{power}=0.8, \text{"one.sample"}=\text{power.t.test}(\text{type}$

که خروجی R به صورت زیر می‌باشد:

t test power calculation One-sample

۲۶.۰۲۳۲۲ = n

۲ = delta

۳.۵ = sd

۰.۰۵ = sig.level

۰.۸ = power

two.sided = alternative

آزمون‌های ANOVA یک‌طرفه متوازن^۲

در اینجا از تابع `power.anova.test` بصورت زیر استفاده می‌کنیم:

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

این تابع دارای پارامترهای `groups` (تعداد گروه‌ها)، `n` (تعداد مشاهدات در گروه)، `between.var` (واریانس بین‌گروهی)، `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 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 حتماً باید مساوی `NULL` قرار گیرد. قابل ذکر است که برای محاسبه واریانس‌های بین‌گروهی و درون‌گروهی از جدول تجزیه واریانس استفاده می‌شود. بدین ترتیب که واریانس درون‌گروهی، همان میانگین مربعات خطا می‌باشد 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, 2007; ۲۰۰۵).

بطور مثال، فرض کنید که می‌خواهیم تعداد نمونه را برای ارزیابی تراکم جمعیت لارو بید کلم بر روی سه گونه گیاه میزبان (کلم‌چینی، کلم معمولی و کلم گل) محاسبه کنیم. در یک مطالعه مقدماتی با شمارش لاروهای موجود بر روی ۵ بوته، تعداد ۴، ۶، ۸ و ۹ لارو بر روی کلم‌چینی، تعداد ۴، ۴، ۴ و ۵ لارو بر روی کلم معمولی و تعداد ۱، ۲، ۴ و ۵ لارو بر روی کلم گل مشاهده می‌گردد. اگر این داده‌ها را به روش ANOVA یک‌طرفه آنالیز کنیم و جدول تجزیه واریانس

1- one-sided

2- two-sided

2

را تشکیل دهیم مشخص می گردد که میانگین مربعات تیمار و خطا بترتیب ۱۵/۲۶۶۷ و ۳/۰۳۳۳ می باشند. بنابراین، واریانس درون گروهی ۳/۰۳۳۳ و واریانس بین گروهی ۳/۰۵۳۳ (یعنی ۱۵/۲۶۶۷ تقسیم بر ۵) می باشند. حال دستور تعیین تعداد نمونه لازم (تعداد تکرار لازم برای هر یک از سه گروه) برای ردیابی یک تفاوت معنی دار (با سطح معنی داری ۰/۱) و قدرت آزمون ۰/۹۰ بصورت زیر نوشته می شود:

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

و یا 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:

one-way analysis of variance power calculation Balanced

۳ = groups

۱۰.۲۸۰۱۶ = n

۳.۰۵۳۳ = between.var

۳.۰۳۳۳ = within.var

۰.۰۱ = sig.level

۰.۹ = power

n is number in each group :NOTE

در اینجا مشخص می گردد که حداقل ۱۱ بوته از هر گیاه باید نمونه گیری شود تا به هدف مورد نظر دست پیدا کنیم.

آزمون تناسبی دونمونه^۱

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

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

این تابع دارای پارامترهای n (تعداد مشاهدات در گروه)، p (احتمال در یک گروه)، p (احتمال در گروه دیگر)، sig.level (سطح معنی داری یا احتمال خطای نوع اول)، power (قدرت آزمون)، alternative (آزمون یک طرفه یا دوطرفه) و strict (برای مورد دوطرفه بکار می رود) می باشد. یکی از پارامترهای n، p، p، sig.level و power در تابع بصورت NULL قرار داده می شود که توسط تابع از پارامترهای دیگر تعیین می شود. البته باید توجه نمود که پارامتر sig.level به علت اینکه دارای پیش فرض غیر NULL است (sig.level = ۰.۰۵) برای محاسبه آن حتما باید مساوی NULL قرار گیرد (Crawley, 2007 ; ۲۰۰۵).

بطور مثال، فرض کنید در یک مطالعه درصد خسارت کرم گلوگاه انار روی دو رقم انار (حساس و متحمل) ارزیابی می گردد. در مطالعه مقدماتی با بررسی میوه های موجود بر روی ۱۰ درخت از هر رقم، مشخص می شود که خسارت بر روی دو رقم مورد مطالعه در حدود ۲۰ و ۵۰ درصد می باشد. حال، دستور محاسبه تعداد نمونه لازم در هر رقم برای ردیابی یک تفاوت معنی دار با قدرت ۰/۸۰ بصورت زیر نوشته می شود:

1- pilot experiment

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

که بصورت زیر توسط R محاسبه می‌گردد:

comparison of proportions power calculation Two-sample

$38.48004 = n$

$0.2 = p1$

$0.5 = p2$

$0.05 = sig.level$

$0.8 = power$

two.sided = alternative

n is number in *each* group :NOTE

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 (Crawley Mick؛ امپریال کالج لندن) که یادگیری آمار و کار با نرم‌افزار R را مدیون ایشان هستم نیز سپاس گزارم.

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.

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

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 issues of statistical power were not confronted 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