الفصل السابع

تحليل القوى وحجم العينة
Power Analysis and Sample Size

لقد ناقشنا حتى هذه اللحظة جوانب هامة من التصميم التجريبي، وملخص البيانات، والتحليل الإحصائي الذي يسمح لنا باختبار الفروض واستخلاص النتائج بمستوى معين من الثقة.

ومع ذلك، فإننا لم نعالج بعد مسألة في غاية الأهمية. والسؤال الذي نطرحه الآن هو: "كم ينبغي أن تكون العينة كبيرة لالتقاط التغير في المجتمع الإحصائي حتى تكون معدلات خطأ النوع I والنوع II صغيرة؟" وبعبارة أخرى، ما هو كبر العينة المطلوب بحيث يكون احتمال القيام بخطأ النوع I أو النوع II في رفض أو قبول فرضية العدم مقبولاً في ظل هذه الظروف. ووجود الحالات المختلفة يستدعي معدلات خطأ مختلفة. إن خطأ مثل تشخيص التهاب الحلق الناجم عن العقديات عندما تكون البكتيريا العقدية غير موجودة هو على الأرجح ليس خطيراً كما هو عند عدم تشخيص السرطان. والطريقة الأخرى لصياغة السؤال هي: "ما مقدار قوة التحليل الإحصائي في قبول أو رفض فرضية العدم؟"

إذا كان حجم العينة صغيراً جداً، فقد تكون النتيجة فقدان أثر ما (خطأ النوع II) لأننا لا نملك ما يكفي من القدرة في الاختبار لإثبات الأثر بثقة.

ومع ذلك من السهل جداً، عند اختيار حجم العينة، القول ببساطة إن حجم العينة ينبغي أن يكون كبيراً قدر الإمكان. وحتى لو كان للباحث إمكانية الحصول على عينات كثيرة بالقدر الذي يرغب فيه، فإن هناك اعتبارات وقيود عملية تحد من حجم العينة. إذا كان حجم العينة كبيراً جداً، فإن هناك مشاكل اقتصادية وأخلاقية يجب أخذها في الاعتبار. أولاً، هناك النفقات المرتبطة بالقيام بتجربة، مثل التجربة الإكلينيكية. وهناك التكاليف المرتبطة بالأفراد الذين يقومون بهذه التجارب، والوحدات التجريبية (الحيوانات، ومستعمرات الخلية، والتعويض عن وقت الإنسان)، وربما الأدوية والإجراءات الطبية الأخرى التي يتم إعطاؤها، وغيرها. وهكذا، كلما كان عدد العينات أكبر، كلما كانت النفقات أكبر. وعادة ما يكون القيام بالتجارب الإكلينيكية مكلفاً للغاية.

إن الاعتبار الثاني للحد من حجم العينة هو الموضوع الأخلاقي. إن العديد من التجارب والاختبارات الطبية الحيوية ذات الصلة يشمل عناصر بشرية أو حيوانية. وقد تتعرض هذه العناصر للأدوية أو العلاجات التجريبية التي تنطوي على بعض المخاطر، وفي حالة الدراسات على الحيوانات، قد يتم التضحية بالحيوان في نهاية التجربة. خلاصة القول، نحن لا نرغب في استخدام العناصر البشرية أو الحيوانية لأي سبب وجيه، وخاصة إذا كنا لا نكسب شيئاً فيما يتعلق بقدرة التحليل الإحصائي من خلال زيادة حجم العينة.

ونذكر مرة أخرى أن هناك نوعين من الأخطاء المرتبطة بإجراء التحليل الإحصائي:

١- النوع I: رفض H_0 عندما تكون H_0 صحيحة (احتمال خطأ النوع I = α)؛

٢- النوع II: قبول H_0 عندما تكون H_0 خاطئة؛ فقدان أثر (احتمال خطأ النوع II = β).

(٧,١) اختبار القوة
POWER OF A TEST

عندما نشير إلى قدرة تحليل إحصائي أو اختبار معين، فإننا نحدد كمية الاحتمال، لقيمة α معينة (على سبيل المثال، 0.05)، بحيث يكشف الاختبار الإحصائي (الاختبار t، أو معامل الارتباط، أو أنوفا (ANOVA)، إلخ) أثراً حقيقياً. على سبيل المثال، إذا كان هناك حقاً اختلاف بين المتوسطات لمجتمعين، فما هو احتمال اكتشافنا لذلك الاختلاف بواسطة الاختبار t؟ إذا كانت فرصة حدوث خطأ النوع II هي β، فإن احتمال اكتشاف الاختلاف الحقيقي عندئذ هو ببساطة $1-\beta$، والذي نشير إليه بقدرة الاختبار. وفيما يلي بعض النقاط العملية حول قدرة الاختبار الإحصائي:

١- قدرة الاختبار = $1-\beta$؛

٢- صفر < القدرة < 1

٣- بشكل عام، نرغب أن تكون القدرة ≤ 0.8 في التطبيقات العملية.

عند محاولة إنشاء حجم عينة لتجربة ما، يجب علينا التحديد مسبقاً مقدار القدرة الذي نريده للتحليل الإحصائي، الذي، بدوره، يتحدد بواسطة مقدار الخطأين للنوع I والنوع II اللذين نحن على استعداد للمخاطرة بهما. يمكننا، بالنسبة لبعض تطبيقات الهندسة الطبية الحيوية، المخاطرة بخطأ أكبر من غيرها من التطبيقات. إن فقدان التشخيص قد لا يكون مشكلة في إحدى الحالات أو قد يؤدي إلى حالة تهدد الحياة. إذا فشل جهاز إزالة الرجفان القابل للزرع في كشف نُظم القلب وصدمه كهربائياً، مثل الرجفان البطيني، فقد يعني هذا فقدان الحياة. من ناحية أخرى، فإن صدم القلب عندما لا يكون هناك رجفان، وهو ما قد يحدث عندما يتحسس الجهاز لنشاط كهربائي متولد عن عضلات الهيكل العظمي القريبة، قد يؤدي إلى الألم، وربما يؤدي أيضاً إلى بدء نُظم قلب خطير.

غالباً ما يكون من الصعب تقليل كل من α و β في الوقت نفسه. عادة ما يتم تقليل واحدة على حساب الأخرى. على سبيل المثال، قد نريد الخطأ إلى حد ما من جانب جهاز إزالة الرجفان القابل للزرع فائق الكشف للرجفان بحيث لا يتم فقدان أي حدوث للرجفان. في حالة الكشف عن سرطان الثدي باستخدام التصوير الشعاعي للثدي، فإن الكشف الزائد والكشف المتدني يمكن أن يكونا مشكلة. والكشف المتدني قد يعني زيادة احتمالات الوفاة. ومع ذلك، يمكن للإيجابيات الخاطئة أن تؤدي إلى إزالة لا لزوم لها من الأنسجة السليمة. وأحد أكبر التحديات التي تواجه الباحثين في المجال الطبي الحيوي هو محاولة إيجاد الاختلافات الحقيقية بين المجتمعات الإحصائية التي من شأنها تحسين التشخيص والكشف عن المرض أو أداء الأجهزة الطبية.

(٧,٢) اختبارات القوى لتحديد حجم العينة
POWER TESTS TO DETERMINE SAMPLE SIZE

تُستخدم اختبارات القدرة لتحديد حجم العينة، والأخذ في الاعتبار التأثير الذي ينبغي الكشف عنه بالتحليل الإحصائي، ومعدلات خطأ النوع I والنوع II المسموح بها، وتغيُّر المجتمعات الإحصائية التي يتم أخذ عيناتها في التجارب.

ولإجراء اختبار القدرة، يمكن استخدام المعادلات التي تعبر عن القدرة بدلالة العوامل المذكورة سابقاً، أو يمكن استخدام منحنى القدرة وجداول القدرة التي تقدِّر بالفعل حجم العينة. توضح منحنيات وجداول القدرة العلاقة بين القدرة $1-\beta$ والأثر الذي نحاول الكشف عنه. ويمكن لهذه الآثار أن تكون مختلفة في اثنين من المتوسطات، الاختلاف في اثنين من التباينات، ومعامل الارتباط، والاختلاف في متوسطات المعالجة، والاختلافات الأخرى للمجتمعات التي نحاول الكشف عنها من خلال التجربة وتحليل البيانات.

من المهم ملاحظة أن هناك معادلات مختلفة، وبالتالي، منحنيات قدرة مختلفة للآثار المختلفة، التي بدورها يتم اكتشافها من خلال الاختبارات الإحصائية المختلفة. يتم الكشف عن اختلاف في اثنين من المتوسطات (الأثر) بواسطة الاختبار t، في حين يتم الكشف عن الاختلاف في اثنين من التباينات (الأثر) باستخدام الاختبار F. إن منحنيات القدرة للاختبار t تختلف عن منحنيات القدرة للاختبار F.

إن منحنيات القدرة الأكثر استخداماً في تقدير حجم العينة لتجربة طبية حيوية تشمل ما يلي:

١- اختبار t المزدوج أو غير المزدوج (اختلاف في واحد أو اثنين من متوسطات المجتمعات الإحصائية).

٢- معامل ارتباط بيرسون (Pearson) (الارتباط بين مجتمعين).

٣- أنوفا (ANOVA) (الاختلاف في ثلاثة أو أكثر من متوسطات المجتمعات الإحصائية).

ولإجراء اختبار القدرة باستخدام منحنيات القدرة، فإننا بحاجة إلى ما يلي:

١- حجم الأثر الذي نريد الكشف عنه (أي الاختلاف في اثنين من المتوسطات).

٢- تقدير بارامترات المجتمع الإحصائي (أي الانحراف المعياري للمجتمع (المجتمعات) استنادا إلى البيانات التجريبية).

٣- مستوى α (احتمال خطأ النوع I).

٤- مستوى القدرة = $1-\beta$ (احتمال خطأ النوع II).

يتم اختيار مستويات α و β من قبل الباحث قبل تصميم التجربة. كما يتم اختيار حجم الأثر الذي ينبغي الكشف عنه أيضاً من قبل الباحث. يحتاج الباحث مقدماً إلى تحديد ضخامة الأثر لكي يكون مهماً لاستخلاص استنتاجات. وبعبارة أخرى، كم يجب أن يكون اثنان من متوسطات المجتمعات الإحصائية مختلفين بالنسبة إلى إشارة مختلفة شيء

ما في الفسيولوجيا الأساسية، أو آثار الأدوية، أو التغيير في عملية التصنيع. أو كم يجب أن تكون قوة الارتباط لكي تعني شيئاً للباحث، من حيث البيولوجيا الأساسية أو العمليات؟ وبشكل عام، كلما كان الاختلاف الذي يجب الكشف عنه بالاختبار t أصغر أو كلما كانت قيمة الارتباط التي يجب الكشف عنها أكبر، كان حجم العينة أكبر.

قد يدَّعي المرء أن الجزء الأصعب للحصول على المعلومات هو تقدير الانحراف المعياري، أو التباين، للمجتمعات الإحصائية وهذا هو التباين المُفترض للمجتمع الذي يلعب دورًا كبيرًا في تحديد حجم العينة. بشكل عام، كلما كان التباين في المجتمع الإحصائي أكبر، كان حجم العينة المطلوب للحصول على إحصائيات المجتمع الإحصائي أو العملية أكبر، وبالتالي، كان حجم العينة المطلوب لإعطاء قدرة للتحليل الإحصائي أكبر.

يوضح الجدول (٧,١) مثالا عن جدول القدرة للاختبار t غير المزدوج. تم تقدير هذا الجدول عندما $\alpha = 0.05$. ولكن توجد جداول مماثلة لمستويات أخرى لـ α. تسمح هذه المنحنيات والجداول بتقدير حجم العينة إذا تم أولاً اختيار الاختلاف الذي تم تسويته في المتوسطين اللذين نريد الكشف عنهما وتحديد القدرة التي يجب عندها إجراء الاختبار t. يتم الحصول على الاختلاف الذي يتم تسويته عندما نأخذ الفرق المطلق الذي نود الكشف عنه وتقسيم هذا الفرق على الانحراف المعياري المُقدَّر للمجتمع. مرة أخرى، لدينا تخمين للانحراف المعياري استناداً إلى البيانات التجريبية. عند تسوية الفرق الذي يجب الكشف عنه، فلا داعي للقلق حول وحدات المقياس ويمكن استخدامها من جداول المعايرة.

إن أفضل طريقة لتوضيح استخدام اختبارات القدرة هي العمل من خلال مثال على ذلك. سوف نستخدم الجدول (٧,١) لإيجاد حجم العينة للمثال التالي:

الهدف: تحديد ما إذا كان هناك فرق في متوسط الطول بالنسبة للرجال في سن الجامعة والنساء اللواتي يعشن في مدينة المدَّعي؟

التجربة: قياس الأطوال في عينات عشوائية من الرجال في سن الجامعة والنساء في مدينة المدَّعي.

السؤال: ما عدد العينات الذي نحتاج إليه لرفض فرضية العدم التي تنص على أنه لا يوجد فرق في الطول بين متوسط الرجال في سن الجامعة والنساء في مدينة المدَّعي إذا كان هناك بالفعل فرقٌ حقيقيٌّ في المجتمعين الإحصائيين ؟

الجدول (١ , ٧). جدول لقدرة الاختبار t غير المزدوج (أحادي الذيل) للطالب ($\alpha = 0.05$).

	اختلاف في المتوسطات (تم التعبير عنه كنقاط z)									
N	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	1.0	1.2
10	0.08	0.11	0.16	0.22	0.29	0.36	0.45	0.53	0.70	0.83
13	0.08	0.13	0.18	0.26	0.34	0.44	0.54	0.63	0.80	0.91
20	0.09	0.15	0.24	0.34	0.46	0.59	0.70	0.80	0.93	0.98
30	0.10	0.19	0.31	0.46	0.61	0.74	0.85	0.92	0.99	
40	0.11	0.22	0.38	0.55	0.72	0.84	0.93	0.97		
60	0.13	0.29	0.50	0.70	0.86	0.95	0.98			
80	0.15	0.35	0.60	0.81	0.93	0.98				
100	0.17	0.41	0.68	0.88	0.97					
200	0.26	0.64	0.91	0.99						

إذا أردنا الكشف عن الاختلاف الحقيقي في المجتمعات الإحصائية فإننا بحاجة لإجراء اختبار القدرة لتحديد عدد العينات اللازم جمعها من الرجال في سن الجامعة والنساء. إننا نفترض أن المجتمعين الإحصائيين موزَّعان بشكل طبيعي. إن الأثر الذي نحاول اكتشافه هو ما إذا كان هناك اختلاف كبير في متوسطات الطول، وبالتالي، فإننا سوف نستخدم اختبار t غير مزدوج لتحليل البيانات حالما يتم جمعها.

نحن بحاجة قبل إجراء اختبار القدرة إلى اتخاذ قرار بشأن مقدار الأثر الذي نريد الكشف عنه بأنه مهم عندما نقوم بالاختبار t. وبالنسبة لهذا المثال سوف نختار فرقًا مقداره 3 بوصة في متوسطات الطول على أنه مهم. ندعي جدلاً أن فرقاً مقداره 3 بوصة في متوسطات المجتمعات الإحصائية هو ذو أهمية بيولوجية.

إن الأثر الذي يجري اختباره $(\mu_1 - \mu_2)$ في هذا المثال هو 3 أنش. لاستخدام المنحنيات الموحَّدة قياسيًّا، نحن بحاجة إلى تسوية هذا الفرق بواسطة تقدير للانحراف المعياري، σ ، للمجتمعات. وهذا هو أساساً النقاط z للفرق بين المتوسطَيْن. لاحظ أننا نفترض أن المجتمعين الإحصائيين لهما تباين متساوٍ تقريباً إذا كان علينا استخدام الاختبار t. للحصول على تقدير لـ σ ، دعونا نفترض أننا جمعنا بعض البيانات التجريبية وقدرنا الانحراف المعياري للعينة، $s = 5$ بوصة. إن الفرق المعياري الذي تم تسويته ونريد كشفه عندما تطبِّق الاختبار t يكون عندئذ 3/5.

بعد ذلك نختار $\alpha = 0.05$ وقدرة = 0.8. مرة أخرى فإن معدلات الخطأ هذه هي خيارنا.

ومن خلال مقدار الأثر (3/5)، وقيمة القدرة (0.8)، وقيمة α (0.05)، فقد استخدمنا الجداول أو المنحنيات في الجدول (٧,١) لإيجاد عدد العينات المطلوبة من جداول القدرة. على سبيل المثال، نجد أن حجم العينة يجب أن يكون 35 عينة تقريباً.

يمكن إبداء بعض الملاحظات العامة حول تأثير تباين المجتمع، معدلات خطأ النوع I والنوع II، ومقدار التأثير على حجم العينة:

١ – بالنسبة لقدرة محددة وتباين معين، فإنه كلما كان الأثر اللازم كشفه أصغر، كان حجم العينة أكبر.

٢– كلما كانت القدرة أكبر، كان حجم العينة أكبر.

٣– كلما كان تباين المجتمع الإحصائي أكبر، كان حجم العينة أكبر.

وإذا استخدمنا نفس المثال السابق والبرنامج Minitab لحساب حجم العينة للاختبار t غير المزدوج عندما نقوم بتغيير العديد من المدخلات لاختبار القدرة، فإننا نجد أحجام العينة التالية (الجدول ٧,٢):

الجدول رقم (٧,٢). العلاقات بين حجم العينة، والقدرة الإحصائية، والانحراف المعياري، ومقدار الأثر الذي يجب الكشف عنه ($\alpha = 0.05$).

حجم الاختلاف في المتوسطات (بوصة)	الانحراف المعياري (بوصة)	قدرة الاختبار	حجم العينة
1	5	0.8	394
2	5	0.8	100
3	5	0.8	45
4	5	0.8	26
3	1	0.8	4
3	2	0.8	9
3	10	0.8	176
3	5	0.9	60

وهكذا، يمكننا، قبل إجراء التجربة، تقدير حد أدنى لحجم العينة التي تسمح لنا باستخلاص استنتاجات من البيانات المتوفرة لدينا مع مستوى معين من الثقة والقدرة.