

Bilimsel Araştırmalarda Çalışma Öncesi Örnek Büyüklüğünün Belirlenmesi

DETERMINATION OF SAMPLE SIZE BEFORE THE STUDY IN SCIENTIFIC INVESTIGATIONS

Bülent URAN

SSK Tepecik Doğumevi, İZMİR

ÖZET

Amaç: Bir çalışmayı planlarken en temel işlem olan hasta ve kontrol grubunun sayısının belirlenmesinin önemini vurgulamak.

Çalışmanın yapıldığı yer: SSK Tepecik Doğumevi-İzmir

Materyal ve Metod: Biyoistatistik konusunda yazılmış kitap ve makaleler.

Bulgular: Örnek büyüklüğünü belirleme işlemi, güç analizi ya da bir çalışmanın gücünü belirleme olarak da isimlendirilmektedir. Daha önceden belirlenmiş bir farkı gerçekte saptayabilmek için ne kadar sayıda deneğe gerek-sinimimizin olduğunu hesaplama işlemidir.

Sonuç: Bir çalışmanın gücü çalışma başlamadan önce belirlenmelidir. Yoksa gereksiz zaman ve kaynak harcanımı söz konusu olabilir

T Klin Jinekoloj Obst 1994, 4:119-122

Bilimsel araştırmalarda doğru ve yerinde istatistik kullanımı, bazı istatistiksel temel kavramları bilmek ve anlamakla olasıdır. Veriler toplandıktan sonra istatistik yöntemi seçme ve kullanma çabası yanlış değerlendirmelere ve yönlendirmelere neden olabilmektedir.

Örnek olarak Türkiye Klinikleri Jinekoloji ve Obstetrik Dergisinin 3. volümü I. sayısında yayınlanan ilk makaleye göz atalım. (Yüksek riskli gebelerde son reaktif nonstress test ile doğum arasındaki süre ve yenidoğanın apgar skorları. Büyükkören ve ark.) Bu makalede verilen mesaj büyük ölçüde istatistiksel analizlerin sonucuna dayanmaktadır. Çalışmada, son NST den 72 saatden sonra doğan ve apgar skoru 7 nin altında olan olguların oranı %14 iken, 24 saatten kısa sürede doğanlarda %6 dır. İki oran arasında İki kattan fazla fark vardır, ama ista-

Geliş Tarihi: 22.03.1993

Kabul Tarihi: 22.03.1994

Yazışma Adresi: Bülent URAN
SSK Tepecik Doğumevi
Yenişehir - İZMİR

Anatolian J Gynecol Obst 1994, 4

SUMMARY

Objective: The number of subjects to be selected for study is a fundamental consideration when planning an investigation.

Institution: SSK Tepecik Maternity Hospital-Izmir

Material and Method: Reports and books which are related to biostatistics.

Results: Determination of sample size is also referred to as power analysis or as determining the power of a study. It consists of determining how large a sample size is required to detect actual difference of some specified magnitude.

Conclusion: The power of a study should be determined before the study begins; Otherwise, the study may require more time and resources than available.

Anatolian J Gynecol Obst 1994, 4:119-122

tistik anlamlılık yoktur. Bu iki grup arasındaki farkın istatistiksel anlamlılık sınırına ulaşabilmesi için aynı sayıda olgu ile ilk oranın en az %38 olması gerekmektedir, iki grup arasındaki oranlarda 5 kattan fazla fark varken, İstatistiksel anlamlılık olmamasını nasıl açıklayacağız? Bu uyumsuzluk çalışma için belirlenen olgu sayısının yetersiz olmasından kaynaklanmaktadır. Bu çalışma için olgu sayısının en az kaç olması gerektiği bu yazının son kısmında hesaplanmıştır.

Bilimsel araştırma planlanırken, çalışmada kullanılacak olgu sayısı belirlenmelidir. Yabancı dergilerin çoğu materyal metod kısmında bu konunun açıklanmasını istemektedirler.

Örnek büyüklüğünü hesaplamak için yapılan işlemlere güç analizi ya da bir çalışmanın gücünü saptama denmektedir.

Bir çalışmayı planlarken örnek büyüklüğünü tahmin edebilmek için şu 4 soru yanıtlanmalıdır.

1. çalışma gruplarının karşılaştırılmasında hangi düzeyde anlamlılık arzu ediyoruz? (P değeri)
2. çalışmamızın, gerçek farklılığı hangi olasılık oranında saptamasını istiyoruz?

3. Çalışmamız gruplar arasında hangi büyüklükte farkı saptayacaktır?

4. Kitlenin, araştıracağımız konu ile ilgili standart sapması (SS) nedir?

GÜÇ ANALİZİ İÇİN HİPOTEZLERİ TEST ETMEK

Bir çalışmada kullanılan istatistik yöntem ne olursa olsun farksızlık(null) hipotezi denilen bir kabullenme test edilmektedir. Farksızlık hipotezi, kontrol grubu ile araştırılan grup arasında anlamlı bir fark bulunmadığını kabul eder. Alternatif hipotez denilen ikinci bir kabullenmede ise gruplar arasında fark bulunduğu kabul edilir ve farksızlık hipotezinin reddi halinde bu hipotez kabul edilir.

Hipotez testlerinde iki hata yapılabilir.

Tip I hata: Gerçekte farksızlık hipotezi doğru iken bunu reddetmek. Tip I hatanın olasılığı alfa değeridir.

Tip II hata: Alternatif hipotez doğru iken farksızlık hipotezinin rededilmemesidir. Tip II hatanın olasılığı beta ile gösterilir. 1-beta değerine güç denmektedir. Her iki olasılık birlikte çalışmanın gücünü belirler. Bir çalışmanın gücü farksızlık hipotezi yanlış iken bu hipotezin reddi olasılığı veya alternatif hipotez doğru iken bu hipotezin kabul edilme olasılığıdır. Çalışma sonucunda doğru karar verebilme gücümüzdür. Gücün yüksek olması arzu edilen bir durumdur. Çünkü tüm araştırmalar gerçekte bir fark varken bu farkı saptamak isterler.

Hipotezleri test etme kavramını testlerin tanısal ölçümleri (sensitivite/spesifite) ile bağdaştırabiliriz.

Tip I hata, yani gerçekte sonuç anlamlı değilken anlamlı fark bulmak, yanlış-negatif testle eş değerdir. (Hastalık varken testin hastalık yok diye sonuç vermesi).

Bir İstatistik testinin gücü (gerçekte fark varken bu farkı saptayabilme gücü) tanısal testlerdeki duyarlılığa (sensitivite) benzemektedir.

Tip I ve Tip II hatayı azaltmanın tek yolu örnek büyüklüğünü arttırmaktır. Eğer örnek büyüklüğünü arttırmak olası değilse o zaman araştırmacı alfa ve beta değerlerini dikkatli bir şekilde değerlendirerek son kararını vermek zorundadır.

Son yıllarda uluslararası dergi editörleri alfa ve beta değerlerinin belirtilmesini istemektedirler. Alfa değerini belirtmek zor değildir. Ancak beta değerini çalışma bittikten sonra belirlemek güçtür. Çünkü tip I hata gerçekte bir fark olmadığı zamanlarda oluşmaktadır ve bu hata sadece bir yönde oluşur. Tip iki hata ise gerçekte bir fark varken ortaya çıkmaktadır ve bu nedenle hata bir çok farklı yollardan ortaya çıkabilir.

Örneğin gece doğan bebeklerle gündüz doğanların kan glukoz düzeyleri arasında bir fark olup olmadığını araştırmak istiyelim. Geçersizlik hipotezimiz bir fark olmadığını ileri sürmektedir. Geçersizlik hipotezi doğru ise gece ve gündüz doğan bebeklerin kan düzeyi aynı sı-

nırlarda olacaktır. Ama alternatif hipotez doğru ise yani gece doğan bebeklerin glukoz düzeyleri farklı ise bu fark çok sayıda farklı değerde olabilir. Bu fark arttıkça farkı saptamamız daha kolay olacaktır. Özetle alternatif hipotezdeki değer farkı arttıkça yanlış bir geçersizlik hipotezini reddetme daha kolay olacaktır. Yani bu durumlarda tip II hata yapma şansı daha azdır. Bu nedenle tip II hata makalelerde daha az. güç analiz daha fazla tartışılmaktadır.

Yukarıdaki örnek çalışmada materyal metod kısmında şöyle bir ifade kullanmak uygun olacaktır: "Bu çalışma gece ve gündüz doğan bebeklerin serum glukoz düzeyleri arasındaki 20 mg/dl'lik farkı %90 güçle saptayacak şekilde tasarlanmıştır." Ya da "Örnek büyüklüğü 20 mg/dl'lik ya da daha fazla bir farkı %80 güçle saptayacak şekilde seçilmiştir." denebilir. İlk durumda beta değeri %10, ikinci durumda %20 olmaktadır.

O halde bir çalışmanın gücünü saptama işlemi gerçek bir farkı, belirlenmiş değerler (yukarıdaki örnekte 20 mg/dl) gözönüne alınarak, ne kadar büyüklükte bir örnek seçmeliyiz?' sorusuna yanıt aramaktır.

Bir çalışmanın gücü mutlaka çalışmaya başlamadan önce belirlenmelidir. Yoksa gereğinden çok daha fazla zaman ve kaynak harcanabilir. Eğer bir araştırmacının gücü düşükse gerçekte mevcut olan bir farkı saptamada yetersiz kalabilir. Özellikle iki tedavi ya da girişim şekli arasında fark olmadığına inanılarak başlatılan çalışmalarda güç analizi çok daha önemlidir.

ÖRNEK BÜYÜKLÜĞÜNÜN SAPTANMASI:

Belirgin bir farkı saptayamamanın tip II hata olduğunu belirtmiştik. Bu tip hata küçük örnek büyüklüklerinde daha sık yapılabilir. Negatif sonuç bildiren çoğu çalışmada örnek büyüklüğü çok küçük olduğu için gerçek etki kaçırılmaktadır.

Çalışma öncesi örnek büyüklüğünü hesaplayan değişik formüller mevcuttur. Hem tip I, hemde tip II hatayı gözönüne alan formüller yeğlenmektedir.

TEK BİR ORTALAMA İÇİN ÖRNEK BÜYÜKLÜĞÜNÜN HESAPLANMASI

Bu formül tek bir ortalamayı standart bir değerle karşılaştırmak için kullanılır.

Araştırmacı şu 4 soruyu yanıtlamalıdır:

1. Standart kitlenin ortalamasını içeren (yo) geçersizlik hipotezi için arzu edilen anlamlılık sınırı (alfa değeri) nedir ?

2. Çalışmamızın gerçek bir farkı saptama şansı (güç-1-beta) ne olmalıdır? (Eldedeceğimiz ortalama değer (y) ile ilişkili alternatif hipotezimizin beta değeri ne olmalıdır?)

3. Ortalamalar arasında hangi büyüklükteki bir farkı saptamak istiyoruz? Yani y1-y0 nedir?

4. Kitlenin standart sapmasının (SS) iyi bir tahmini nedir?

$$n = \frac{\{(z \text{ alfa} - z \text{ beta}) * SS\}^2}{y1 - y0}$$

ÖRNEK:

Kısa boylu kadınların gebeliklerinde solunum fonksiyonlarını normal boylu kadınların değerleri ile karşılaştırmak istiyoruz.

Tip I hatanın 0.05 olmasını kabul edelim. Bu değer gerçekte iki grup arasında fark yokken fark bulma olasılığını belirlemektedir. Yani bu çalışmayı 100 kere yaparsak en fazla 5 kez fark bulma sınırını benimsiyoruz.

Gerçek farkı saptama şansımızı da (1 - beta) %90 olarak kabul edelim. Normal boylu gebelerde maksimum inspirasyon basıncı 110 mm hg ve SS nı 20 mm hg olarak daha önceden belirlenmiş olsun. Klinik olarak kısa boylu hastalarda 80 mm hg ve daha aşağı ölçeceğimiz değerlerin normal kadınlardaki değerlerden anlamlı olarak farklı olduğunu kabul edersek, böyle bir farkı saptayacak örnek büyüklüğü ne olmalıdır?

Alfa 0.05 için z değeri +/- 1.96 dır. (z tablosunda %95 ile %5 i birbirinden ayıran değer.)

Beta'nın alt z değeri -1.28 dır. (alt %10 ile üst %90 nı birbirinden ayıran değer.)

$$n = \frac{\{(1.96 + 1.28) (20)\}^2}{110 - 80} = 4.67 \quad n=5^*$$

İKİ ORTALAMANIN KARŞILATIRILDIĞI ÇALIŞMALAR İÇİN ÖRNEK BÜYÜKLÜĞÜNÜ SAPTAMAK

iki bağımsız grubun ortalamasını karşılaştırmada da işlemler yukandakine benzerdir. Ancak bu tip karşılaştırmada iki kabullenme yapmak hesabı kolaylaştırır.

Bu basitleştirme veya kabullenmeler;

1. Her iki örnek grubunun SS nın aynı olması,
2. Her iki örnek büyüklüğünün aynı sayıda olmasıdır.

y1 y2 saptanacak fark ise, her bir grup için gerekli örnek büyüklüğü;

$$n = 2 \frac{\{(z \text{ alfa} - z \text{ beta}) * SS\}^2}{(y1 - y2)}$$

ÖRNEK:

Düşük ürik asitli grubun bebek ağırlığı ile yüksek ürik asitli grubun bebek ağırlıklarını karşılaştırmak ve böylece ürik asit düzeylerinin bebek gelişimine etkisini araştırmak istiyoruz.

Tip I hatamız yani yanlış olarak bebek ağırlıkları arasında fark saptama şansımız 0.05 olsun.

Gerçek farkı saptama şansımızı %80 olarak alalım.

Bebek ağırlıkları arasındaki ortalama farkın 100 gramdan fazla olmasını anlamlı kabul edelim. Bebek ağırlığı için SS nın daha önceki gözlemlerimize dayana-

(* Ondalık sonuçlar bir üst tam sayıya yuvarlanmalıdır.) Eğer 10 mm hglik farkı anlamlı kabul etseydik çok daha fazla sayıda hastaya gereksinimimiz olacaktı.

Anatolian J Gynecol Obst 1994, 4

rak 150 gram olduğunu belirlemiş olalım. Bu kabullenmelere dayanarak istediğimiz farkı saptamak için örnek büyüklüklerimiz ne olmalıdır?

Alfa için z değeri +/- 1.96 dır.

Tek yönde fark aradığımız için beta için z değeri-0.84 olarak (%20 ile %80 i birbirinden ayıran değer) z değerleri tablosundan elde edilmektedir.

$$n = 2 \frac{\{(1.96 + 0.84) (150)\}^2}{100} = 36$$

her bir grup için gerekli hasta sayısı olur.

Yani araştırmada 100 gramlık bir farkı %80 şansla yakalayabilmek için her iki grupta en az 36 hasta kullanılmak zorundayız.

Pratik olarak iki ortalama arasındaki farkı saptamak için örnek büyüklüğü şu şekilde hesaplanabilir.

Öncelikle SS nın iki ortalama arasındaki farka oranının ne olması gerektiğine karar verilir. Sonra bu oranın karesi alınır. 0.05 lik P değeri ve % 90 lik farkı saptama şansı için her iki grupta gerekli hasta sayısı, bu değerlerin yaklaşık 20 katıdır. Aynı P değeri ve %80 lik şans için oranın karesinin 15 katı kadar sayı yeterlidir.

ORAN OLARAK ELDE EDİLEN DEĞERLER İÇİN ÖRNEK BÜYÜKLÜĞÜ TEK BİR ORAN İÇİN ÖRNEK BÜYÜKLÜĞÜ

Standart bir değerle tek bir değeri karşılaştırmak için yine 4 soruya yanıt aramak gerekir.

1. istenen anlamlılık derecesi nedir? (alfa düzeyi nedir?)
2. Gerçek bir farkı saptama şansımız ne olmalıdır? Yani hangi oranda bir güç (1-beta) arzu edilmektedir.
3. Oranlar arasında hangi farklılığı saptamak istiyoruz?
4. Kitlenin SS nın iyi bir tahmini nedir?

Bir oran için geçersizlik hipotezindeki değer, p0, tahmin edilen SS dır.

$$n = \frac{\{(z \text{ alfa} * \sqrt{p0 (1-p0)} - z \text{ beta} * \sqrt{p1 (1-p2)}\}^2}{p1 - p0}$$

ÖRNEK:

Kontrolde gelmesi gereken hastalara kontrolden önce kontrollerini hatırlatan bir mektup gönderilmesinin kontrole gelen sayısını arttırmada yararlı olup olacağına araştırmak istiyoruz.

Daha önceden kontrole geliş oranının %40 ve tip I hatamızında %2 olduğunu varsayalım. Gerçek farkı da %90 olasılıkla yakalamak isteyelim. Daha önce gelenlere göre %25 artış sağlamayı (yani oranın %40 dan %50 ye çıkmasını) anlamlı fark olarak kabul edelim.

z alfa +/- 2.326 dır. Bu değer z tablosunda santal %98'i uçlardaki %2 den ayıran değerdir. Beta için alt z değeri -1.28 dır.

Formülümüze yerleştirirsek;

$$n = \frac{\{2.326 \cdot \sqrt{0.40 \cdot 0.60} - (-1.28 \cdot \sqrt{0.50 \cdot 0.50})\}^2}{0.40 - 0.50} - 317$$

Yani bu araştırmacılar 317 kişiye mektup gönderirlerse gerçekten kontrole gelenlerin oranının %50 ye artıp artmayacağına karar verebilirler.

İKİ ORANIN KARŞILAŞTIRILDIĞI ÇALIMALARDA ÖRNEKBÜYÜKLÜĞÜ

Bu tip bir çalışmada formülü basitleştirmek için yapılan bir kabullenme, her iki gruptaki örnek büyüklüğünün aynı olduğunu kabul etmektir. Pt tedavi, Pc kontrol grubundaki oranı temsil etsin.

Her bir grup için gerekli sayı:

$$n = \frac{\{z_{\alpha} \cdot \sqrt{2P_c(1-P_c)} - (z_{\beta} \cdot \sqrt{P_t(1-P_t)} + P_c(1-P_c))\}^2}{P_t - P_c}$$

ÖRNEK:

Bir antibiyotiğin postoperatif enfeksiyonu önlemedeki rolünü saptamak isteyelim. Daha önceki bulgularımıza göre %10 olan postop enfeksiyon oranımızın %5 in altına düşmesinin anlamlı olacağını kabul edelim. Tip I hata için 0.05 ve gerçek farkı saptamak için 0.90 olasılıklarını kabullenelim.

z alfa= +/- 1.96 ve z beta= 1.28

$$n = \frac{\{(1.96 \cdot \sqrt{2 \cdot 0.10 \cdot 0.90}) - 1.28 \cdot \sqrt{(0.05 \cdot 0.95)} + (0.10 \cdot 0.90)\}^2}{0.05 - 0.10}$$

- 682 her bir grupta kullanmamız gereken hasta sayısıdır.

Şimdi giriş bölümünde belirttiğimiz makalede gerekli olan olgu sayısını hesaplayalım.

Son NST yapıldıktan sonra 72 saatden daha geç doğan bebeklerin apgar skorlarının 72. saatten önce doğan bebeklerden anlamlı olarak farklı olup olmadığını araştırmak istediğimizde nasıl bir kabullenme klinik olarak uygun olacaktır? Kanımca NST gibi fetal prognozu saptama açısından hayati önemi olan bir testde, testten sonra 72 saatten fazla doğum geciktiği takdirde 7

güne Kaaar yeni or test yapmaya gerek yoktur aieoii-
mek için, 72 saatten önce ve sonra doğan düşük apgarlı bebeklerin oranları arasındaki farkın 2 katdan fazla olmaması gerekir. Diğer bir deyişle oranın 2 kat artmasını anlamlı kabul edersek, alfa - 0.05 ve beta - 0.90 lık bir power için incelememiz gereken gebe sayısı ne olmalıdır?

Çalışmaya göre 72 saatten önce doğum yapan 44 hastada 7 den düşük apgar skorlu bebek oranı %9 dur. Bu oranın %18 e çıkmasını anlamlı kabul edeceğiz.

4 nolu formülümüzde değerleri yerine koyduğumuzda her bir grup için 244 ve çalışma için 488 hasta gerektiği ortaya çıkmaktadır.

Sonuç olarak bilimsel çalışmalarımızı kullanacağımız istatistik yöntemlerine göre planlamak, çalışacağımız parametrelerin klinik anlamlılık değerlerini belirlemek ve bunlara bağlı olarak örnek büyüklüğünü saptamak, çalışma sonuçlarını değerlendirmeyi, yorumlamayı, ve belirli sonuçlar çıkarmayı çok kolaylaştıracaktır.

KAYNAKLAR

1. Saunders BT, Trapp RG: Basic and Clinical Biostatistics. Appleton & Lange, California. 1990.
2. Altman DG, Gore SM, Gardner MJ, Pocock SJ: Statistical guidelines for contributors to medical journals. Br M J 1983; 286:1489
3. Perkins RP: How to live with statistics (without having to marry them). Obstet Gynecol 1988; 72:422.
4. Altman DG: Statistics: Necessary and important. Brit J Obstet Gynaecol 1986; 93:1. '
5. Schlesselman JJ: Case-Control Studies. Oxford University Press. New York, 1982.
6. Anderson S, Auquier A, Hauck WW et al: Statistical Methods for Comparative Studies. John Wiley & Sons. New York. 1980.