

Örnekleme ve Randomizasyon

Doç. Dr. Oktay Özdemir

Araştırma yapmanın amacı, bir gruptaki gözlemlerden yararlanarak topluma genelleme yapabilmektir. Örneğin, yeni bir antihipertansif ilacın etkinliğini tüm hipertansiflerde denemek olası değildir. Bu amaçla yapılan bir araştırmada bir grup hipertansif hasta üzerinde bu ilacın etkisi denir ve sonuçlar genellenir; yani “bu ilaç hipertansif hastalarda kan basıncını düşürmede etkilidir (ya da değildir)” denir. Çalışmaya alınan hastaların oluşturduğu gruba örneklem (*sample*), bu grubun temsil ettiği topluma evren (*population*) adı verilir. Yukarıdaki örnekte çalışmaya alınan hipertansif hastalar örnekleme oluşturmaktadırlar. Bu çalışmanın temsil ettiği evren ise tüm hipertansif hastalardır. Bir çalışmanın sonuçları, yalnız çalışma örnekleminin temsil ettiği evrene genellenebilir.

Örnekleme

Bir yıl boyunca Ankara'da hastaneye yatan hastaların tanılara göre dağılımını öğrenmek istiyorsak, en doğru sonuç tüm hastane kayıtlarının toplanmasından elde edilir. “Tüm hastane kayıtları” evreni oluşturur. Ama tüm evreni bir çalışmaya almak zaman ve para açısından altından kalkılabilecek bir şey değildir. O nedenle tüm evreni temsil eden bir grup alıp, örneklem oluşturmak ve bu örneklemdeki deneklerin verilerini evrene ekstrapole etmek daha ekonomik olur. Bunun doğru olabilmesi için örneklemin evreni doğru temsil etmesi gerekir. Bu nedenle örneklemden deneklerin evrenden rastgele (*random*) olarak seçilmesi gerekir. Örneğin, yukarıdaki çalışma için Ankara'daki tüm üniversite hastanelerinin kayıtlarının alınması yanıltıcı olabilir, çünkü diğer hastanelerle karşılaştırıldığında üniversite hastanelerine daha çok tanı ve tedavisinde sorun olan hastalar yatmış olabilir. Böyle bir örneklemden elde edilecek sonuçlar tüm evrene genellenirse, hata yapılır.

ÖRNEK 1

Behçet hastalığında bazen eklem, bazen deri, bazen göz tutulumu ön plandadır. Bu nedenle bu hastaların bazıları Romatoloji, bazıları Dermatoloji, bazıları da Oftalmoloji kliniklerinde izlenir. Dermatoloji kliniğinde Behçet hastalığında yeni bir ilacın etkinliğinin denendiği bir çalışma yapılıyor ve ilaç yararlı bulunuyor.

Örnek 1'deki çalışmadan “Bu ilaç Behçet hastalığında yararlıdır” sonucunu çıkarmadan önce, örneklemin evrenden iyi seçilip seçilmediğini araştıralım. Örneklemdaki hastalarda deri tutulumunun hakim olduğu, eklem ve göz tutulumunun ise nadir olduğu kestirilebilir. Bu nedenle bu çalışmadaki örneklemin “Behçet hastaları” evrenini iyi temsil etmediği anlaşılabilir. Bu çalışmanın sonuçlarının genellenebileceği evren “Behçet hastaları” değil “deri tutulumu olan Behçet hastaları”dır. Çünkü örnekleme girecek hastalar “tüm Behçet hastaları” evreninden rastgele seçilmemişlerdir.

Çalışmalar için örneklem oluştururken, bazı randomizasyon yöntemlerinin uygulanması gerekir. Örneklem seçimi için kullanılan başlıca dört yöntem vardır: basit tesadüfi örnekleme, sistematik örnekleme, tabakalı örnekleme ve küme (*cluster*) tipi örnekleme. Bu dört yöntemin nasıl olduğunu anlamak için aşağıdaki örnekteki çalışmayı ele alacağız.

ÖRNEK 2

Bir hastaneye son on yıl içinde yatan hastaların çeşitli özelliklerinin saptanması amacıyla bir çalışma yapılması planlanıyor. Bu süre içinde yatan hasta sayısının 12.000 olduğunu düşünelim. Bu kadar hastanın kayıtlarının incelenmesinin pratik olmadığı kesindir. Bu nedenle 12.000 hasta dosyasından (evren) 1.000 tanesinin (örneklem) seçilip, bu çalışmada kullanılması planlanmış olsun. Hangi dosyaların seçileceği aşağıdaki yöntemlerle belirlenebilir.

Basit Rastgele Örneklem

Her yatış dosyasının 1'den 12.000'e kadar numaralandırıldığını varsayalım. Basit rastgele (*random*) örneklemede çeşitli istatistik kitaplarındaki rastgele sayılar tablosundan ya da bilgisayarlardaki örnekleme modüllerinden yararlanarak 1.000 tane sayı belirlenir ve dosya numarası bu 1.000 sayının içinde olan hastalar, örnekleme alınır.

Sistematik Örneklem

Bu yöntemde evren genişliği (evrendeki denek sayısı), örneklem genişliğine (örneklemden denek

sayısı) bölünerek elde edilecek rakam kullanılarak, örnekleme alınacak denekler belirlenir. Örneğin $12.000/1.000=12$ olduğu için önce 1'den 12'ye kadar olan rakamlardan biri rastgele seçilir. Örneğin bu rakam 5 olsun. Kayıt numarası 5 olan dosya örnekleme seçilir. Daha sonra kayıt numarası $5+12=17$ olan, $17+12=29$... olan diğer dosyalar seçilerek, 1.000 denekli örneklem tamamlanır.

Tabakalı Örneklem

Tamamen rastgele seçim bazen sakıncalı olabilir. Örneğin, her yaş grubundan hastalara ait dosyaların örnekleme alınması isteniyorsa, rastgele seçim yapıldığında, bazı yaş gruplarından dosyalar seçilmemiş olabilir. Tabakalı örneklemede, örneklem seçimine başlamadan önce, evren, belirlenmiş özellikler açısından tabakalara ayrılır ve her tabakadan rastgele seçim yapılır. Yaş dışında, cinsiyet, hastalığın şiddeti ya da evresi, süresi gibi özellikler de tabakalı örnekleme için sıklıkla kullanılır. Örneklemede ölçülecek değişkenin özelliklerine uygun tabakalandırma kriterleri seçilmelidir. Aksi halde fazla tabaka kullanma nedeniyle örneklem çok genişler.

Örneğin, anemi prevalansının araştırıldığı bir kesitsel çalışmada hemoglobin değerlerini etkileyen yaş ve cinsiyet açısından tabakalandırma yapılması gereklidir. Örneklemin yaş ve cinsiyet açısından evrene benzemesi böylece sağlanabilir. Tabakalandırma yapılmadan tamamen rastgele seçilen bir örneklemden bir cinsiyetin ya da yaş grubunun daha fazla bulunması gibi istenmeyen durumlar ortaya çıkabilir. Tabakalı örneklemede göz önüne alınması gereken bir konu, örneklem seçiminden önce evrene ait bazı özelliklerin bilinmesi gerekliliğidir.

Tablo 1 incelendiğinde, anemi prevalansı çalışmasına alınan deneklerin %40'ının erkek,

ÖRNEK 3

Bir çalışmada, 200.000 erişkin kişinin yaşadığı bir şehirdeki anemi prevalansının belirlenmesi amaçlanıyor. Erkek/kadın oranının “1” olduğu şehirde yaş dağılımının Tablo 1'de “Evren %” sütununda görülen oranlara uygun olduğu daha önceden biliniyor. Çalışmaya rastgele örnekleme yöntemi ile 500 kişi alınıyor. Tablo 1'de “n” sütununda çalışmaya alınan deneklerin sayısı, “Örneklem %” sütununda örneklemden deneklerin cinsiyet ve yaş grubu açısından örneklemden ne kadarlık bir oranı oluşturduğu, “Beklenen Hb” sütununda altgrupların hemoglobin ortalamaları, “Evren %” sütununda cinsiyet ve yaş grubuna göre evrendeki oranları görülmüştür.

Tablo 1. Evrendeki yaş ve cinsiyet özellikleri dikkate alınmadan oluşturulan bir örneklemin evreni iyi temsil etmemesi nedeniyle meydana gelen hata.*

	Yaş	n	Örneklem %	Hb	Evren %	Beklenen Hb
Erkekler	20-39	50	10	16,2	20	16,2
	40-59	50	10	15,4	25	15,4
	60-79	100	20	12,3	5	12,3
	Toplam	200	40	14,0	50	15,4
Kadınlar	20-39	150	30	12,0	20	12,0
	40-59	75	15	13,6	20	13,6
	60-79	75	15	12,3	10	12,3
	Toplam	300	60	12,5	50	12,7
Toplam		500	100	13,1	100	14,1

*Seçilen örnekleme evrenin aksine kadınların ve yaşlıların çoğunlukta olması nedeniyle hemoglobin değerleri gerçekte olduğundan daha düşük bulunmuştur.

%60'ının kadın olduğu, ayrıca deneklerin %35'inin 60-79 yaş grubunda olduğu görülmektedir. Oysa evrende erkek-kadın dağılımı eşit ve 60-79 yaş grubundakilerin oranı %15'tir. O halde örnekleme evrene göre daha fazla oranda kadın ve daha fazla yaşlı denek vardır. Bu nedenle hem erkek denekler grubu, hem de tüm grup için ortalama Hb değeri olması gerekenden daha düşük olarak bulunmuştur. Bunun nedeni ya tabakalı örnekleme yönteminin kullanılmamış olması ya da her tabakaya düşecek denek sayısını belirlerken evren oranlarının dikkate alınmamış olmasıdır.

Yeniden Örnek 2'ye dönersek, hastane yatış kayıtları mevsime göre dört tabakaya ayrılıp, daha sonra her mevsimdeki yatış dosyalarından 250'şer tanesi örnekleme seçilirse, farklı mevsimlerde yatan hastaların özellikleri açısından rastgele örnekleme göre evreni daha iyi temsil eden bir örneklem oluşturulmuş olur.

Küme Tipi Örnekleme

Bu örnekleme yönteminde önce evren, küme (*cluster*) adı verilen altbirimlere bölünür. Daha sonra her kümedeki tüm denekler ya da diğer randomizasyon yöntemlerinden birine göre seçilen denekler örnekleme alınır. Örneğin çok merkezli çalışmalarda her merkez bir kümeyi oluşturur. Her merkezden hangi bireylerin örnekleme seçileceği ise başka bir yöntemle belirlenebilir.

Örnek 2'de hastalık sıklıklarının mevsimlere göre farklılık göstermesi nedeniyle, örnekleme her mevsimden hastaların alınması daha doğru olacaktır.

Bunu sağlamanın bir yolu tabakalı örnekleme yöntemidir. Alternatif olarak küme tipi örnekleme de kullanılabilir. Önce 12.000 hastane yatış kaydı aylara göre 12 kümeye ayrılıp, daha sonra bu 12 küme içinden her mevsimden birer ay olmak üzere Ocak, Nisan, Temmuz ve Ekim aylarında yatanlar örnekleme alınmaya aday olarak seçilebilir. Her aydaki yatış sayısı aşağı yukarı eşitse, 12.000 yatışın 4.000'i bu aylarda gerçekleşmiş olmalıdır. Bu sayı hala istediğimiz örneklem genişliği olan 1.000'den fazla olduğuna göre bu hastalar arasından dördte birini seçmemiz gerekir. Bu kez her aydaki yatış kayıtları haftalara göre dörder kümeye ayrılıp, Ocak ayının ilk haftası, Nisan ayının ikinci haftası, Temmuz ayının üçüncü haftası ve Ekim ayının dördüncü haftasında yatan tüm hastaların dosyaları örnekleme alınabilir. Bu şekilde seçilen örnekleme yaklaşık 1.000 denek bulunması umulabilir. Eğer çeşitli altkümelere ayrıldığı halde istenen denek sayısından daha fazla denek kalıyorsa, denek sayısı rastgele ya da tabakalı örnekleme yöntemleri ile istenen örneklem genişliğine indirilir.

Küme tipi örnekleme, kamuoyu araştırmalarında sıklıkla kullanılır. Örneğin bir şehirdeki 20 semtten beşi seçilir. Bu beş semtin her biri bir kümedir. Tüm kümelerdeki bireylerin hepsi çalışmaya alınabileceği gibi, her semtteki sokaklar ve her sokaktaki evler de altkümelere oluşturulabilir.

Tüm bu örneklem seçim yöntemleri, evreni iyi temsil edecek bir örneklemin seçilmesinden başka, klinik çalışmalarda deneklerin tedavi ya da plasebo gruplarından hangisine alınacağını belirlemek için de kullanılabilir.

Randomizasyon

Paralel düzende yapılan bir çalışmada önce hastalar çalışmaya alınma ölçütleri açısından değerlendirilir. Ölçütlere göre uygun olmayan hastalar çalışma dışı bırakılır. Uygun hastalar ise tedavi gruplarından birine alınır. Hastalara verilecek tedavinin seçilme yöntemi yapılacak çalışmanın sonuçlarının güvenilirliği açısından çok önemlidir. Tedavi seçiminin araştırmacının insiyatifinde olmasının önemli sakıncaları vardır. Örneğin, yeni ve etkinliği pek de bilinmeyen bir antidepresif ilacın plaseboyla karşılaştırıldığı bir çalışmada araştırmacılar bilinçli ya da bilinçsiz olarak, depresyon bulgularının ağır olduğu hastalarda ilacı, hafif olduğu hastalarda plaseboyu seçebilir. Böylece ilaç grubunda ağır depresyonlu, plasebo grubunda hafif depresyonlu hastalar yer alacağı için elde edilecek sonuçlar ilaç aleyhine olabilir.

Bu nedenle hastaların çalışma gruplarına seçiminin rastgele olmasının sağlanması gerekir. Bunu sağlayan yöntemlere randomizasyon adı verilir. Randomizasyon, hastaların tedavi gruplarına seçilmesinde bilinçli ya da bilinçsiz olarak yapılan “manipülasyon”ları engeller.

En kolay randomizasyon yöntemi basit randomizasyondur. Örneğin hastalara verilecek tedavinin yazı-tura ile belirlenmesi bir basit randomizasyon yöntemidir. Bu yöntemle hastaların prognostik etkenler açısından durumları dikkate alınmadığı halde, özellikle hasta sayısı fazla tutulduğunda genellikle dengeli dağılım sağlanabilir. Hasta sayısı az olduğu zaman basit randomizasyonla bile gruplar arasında dengesizlik olabilir. Hasta sayısı arttıkça, gruplar arasında dengesizlik olma olasılığı azalır.

Hasta sayısının az olduğu bir çalışmada basit randomizasyon tek başına dengesizliğin azaltılmasında yeterli olmayabilir. Bu durumda randomizasyona şu iki yöntemden biri ya da her ikisi birlikte eklenmelidir:

- Tabakalandırma
- Bloklama

Örnek 4'teki araştırmada, araştırmacı önce hastaları bir ya da daha çok prognostik etkene göre sınıflandırır (tabakalandırır), daha sonra her tabakaya baştan belirlenmiş (ve genellikle eşit) sayıda hasta girecek şekilde hastaları seçer (bloklar). Örneğin her iki erkek hastadan birini bir gruba, diğerini diğer

ÖRNEK 4

Bir ilacın etkinliğinin plasebo ile karşılaştırılacağı bir çalışmada hastaların ilaç ve plasebo gruplarına rastgele olarak seçilmesi planlanıyor. İlacın etkilerinin erkeklerde ve kadınlarda farklı boyutta olacağı beklendiği için, gruplardaki hastaların cinsiyet açısından dağılımının mutlaka benzer olması isteniyor. Bu nedenle basit randomizasyon yöntemi ile yetinmeyip, hastaların cinsiyetlerine göre gruplara dengeli olarak dağıtılmasını sağlayacak bir yöntem uygulanmak isteniyor.

gruba seçecek şekilde bir randomizasyon şeması kurar ve kadınları da aynı şekilde seçerse, çalışma grupları cinsiyete göre tabakalandırılmış ve ikili gruplar halinde bloklanmış olur. Örneğin ilk erkek hastayı yazı-tura ile A grubuna seçtikten sonra gelen ikinci erkek hasta zorunlu olarak B grubuna alınır. Üçüncü erkek hastada yine yazı-tura ile ilaç seçimi belirlenir. Aynı işlem kadınlar için de benzer şekilde yapıldığında, tabakaların birleştirilmesi ile gruplar kendiliğinden dengeli olur. Yazı-tura yerine istatistik kaynak kitaplar ya da bilgisayarlardan elde edilebilecek rastgele sayılar tablosundan yararlanarak da randomizasyon yapılabilir. Tablo 2'de cinsiyetin tabaka değişkeni olduğu ve blokların ikişer kişiden oluştuğu bir randomizasyon şeması görüyorsunuz. Blok I'de ve II'de ikişer erkek, blok III ve IV'de ikişer kadın yer almaktadır. Her bloktaki hastaların biri ilaç grubuna, diğeri ise plasebo grubuna seçilmiştir. Böylece iki grubun cinsiyet açısından benzer olması sağlanmıştır. Tablo 2'deki randomizasyon şemasına göre ilk erkek hastaya ilaç, ikinci ve üçüncü erkek hastaya plasebo, dördüncü erkek hastaya ilaç verilecektir.

Tabakalandırmaya mutlaka bloklama da eşlik etmelidir. Aksi takdirde basit randomizasyon ile bloklama yapılmadan yapılan tabakalandırma arasında, gruplar arası dengenin sağlanması açısından fark yoktur. Ama tabakalandırma olmaksızın bloklama yapılabilir. Örneğin, tek amaç gruplarda hasta sayısının eşit olması ise, tek başına bloklama yeterlidir.

Birden çok özellik açısından da tabakalandırma yapılabilir. Kaç adet tabakalandırma değişkeni kullanılacağına seçiminde tedaviye yanıtı etkileyen prognostik etkenlerin (ya da kontrol değişkenlerinin) araştırmacı tarafından, henüz araştırma başlamadan belirlenmesi önemlidir. Prognostik etkenlerin neler olduğu ya literatürdeki bilgilere dayanarak ya da araştırmacının deneyimlerine dayanarak seçilmelidir.

Tablo 2. Örnek 4'teki çalışma için cinsiyet açısından tabakalandırılmış ve her blokta iki hasta olmak üzere bloklaşma yapılmış randomizasyon şeması.

Tabaka	Blok	İlaç	Plasebo
Erkek	I	+	+
	II	+	+
Kadın	III	+	+
	IV	+	+
		4	4

Örnek 5'deki çalışmada uygulanacak tabakalandırmada özen gösterilmesi gereken konu, her tabaka değişkeninin nasıl kategorize edileceğidir. Örnek 4'te "cinsiyet" in iki adet kategoriden oluşacağı hakkında kuşku yoktur. Ancak yaş, vertebra sayısı gibi sayısal değişkenler için tabakaların kategorilerinin araştırmacı tarafından belirlenmesi gerekecektir.

ÖRNEK 5

Skolyoz cerrahisi sırasında desmopressin kullanılmasının perioperatif kanamayı azaltacağı hipoteziyle plasebo-kontrollü bir çalışma planlanıyor. İlacın etkisi dışında kanama miktarını etkileme olasılığı olan diğer özellikler açısından desmopressin ve plasebo gruplarının benzer olması isteniyor. Bu nedenle hastanın yaşı, operasyona alınacak bölgedeki vertebra sayısı ve skolyoz nedeni gibi etkenler açısından tabakalandırma yapılmasının gerekli olduğu düşünülüyor.

Gereğinden fazla kategori kullanılırsa, araştırmaya alınacak hasta sayısının çok artması ve buna rağmen gruplar arası dengenin tam olarak sağlanamaması gibi sakıncalar ortaya çıkabilir. Örneğin "yaş" tabakasının kategorilerinin "0-4, 5-9, 10-14, 15-19, ..." gibi beşer yıllık ve "vertebra sayısı" tabakasının kategorilerinin "1-3, 4-6, 7-9, ..." gibi üçer vertebra olarak belirlendiğini ve "skolyoz nedeni" nin ise "doğumsal, edinsel" olarak iki kategoriden oluştuğunu varsayalım. Operasyon bölgesindeki vertebra sayısı "7-9" grubuna giren 6 yaşında (yani 5-9 yaş grubuna giren) doğumsal skolyozu olan bir hasta desmopressin grubuna alındığında, plasebo grubuna da vertebra sayısı ve yaşı benzer olan doğumsal skolyozlu bir hastanın alınması gerekecektir. Aksi takdirde gruplar, prognostik etkenler açısından dengeli olmayacak ve tabakalandırma işe yaramamış olacaktır.

Öte yandan tabakaların kategorilerinin fazla geniş olması da sakıncalıdır. Örneğin "vertebra sayısı" tabakasının kategorileri "1-10 ve >10 vertebra" olarak belirlenirse, 2 vertebraya müdahale yapılacak bir hastayla 10 vertebraya müdahale yapılacak bir hasta aynı kategoride ele alınacak ve kanama miktarına vertebra sayısının etkisi açısından bu iki hasta arasında fark olmadığı varsayılacaktır.

Hem tabaka sayısı, hem de tabakalardaki kategori sayısı arttıkça araştırmanın yürütülmesi ve grupların dengesinin sağlanması güçleşir. Bu nedenle prognostik etkisi belirgin olmayan değişkenlere göre tabakalandırma yapılmasından kaçınılması uygun olur. Aynı şekilde tabakaların kategorilerinin alt ya da üst sınırlarının belirlenmesinde de prognostik açıdan aralarında fark olduğu varsayılan sınırlar seçilmelidir. Örneğin alt solunum yolu infeksiyonu gelişmesinde yaşın etkisine göre grupların benzerliğini sağlamak için hastaların "45 yaş altı-45 yaş üstü" olarak iki kategoriye ayrılması yeterli olursa "20-24, 25-29, 30-34, ..." gibi dar kategoriler oluşturmaya gerek yoktur.

Tabakaların kategorilerinin eşit genişliklerde olması gerekmez. Örneğin, özellikle ilk yaşta ve 60 yaşından sonra daha ağır seyreden, bu nedenle ilaca yanıt açısından bu yaş gruplarında farklı seyir gösteren bir hastalık için yaş açısından tabakalandırma yapılacağı zaman kategoriler "0-1 yaş, 2-20 yaş, 21-60 yaş ve >60 yaş" olarak belirlenebilir. Önemli olan kategorilerin alt ya da üst sınırlarının, hastaları prognoz açısından farklı olan gruplara ayırmasıdır.

Tekrar skolyoz cerrahisi örneğine dönersek, araştırmacı tarafından yaşın "<20, 20-49 ve >49" olarak üç kategoriye, vertebra sayısı tabakasının "1-5, 6-10 ve >10" olarak üç kategoriye ve "skolyoz nedeni" nin ise "doğumsal, edinsel" olarak iki kategoriye ayrıldığını düşünelim. Bu durumda bu üç prognostik etken açısından $3 \times 3 \times 2 = 18$ kombinasyon olacaktır (Tablo 3).

Görüldüğü gibi "<20 yaş, 6-10 vertebra sayısı ve doğumsal skolyoz" ve "20-49 yaş, 6-10 vertebra sayısı ve doğumsal skolyoz" tabakalarında desmopressin ve plasebo gruplarında hasta sayıları eşit değildir. Buna bağlı olarak grupların tam dengeli olmadığı söylenebilir. Bazı gruplarda hem desmopressin, hem de plasebo gruplarında hasta olmaması grupların dengesi açısından sakınca yaratmaz.

Tablo 3. Örnek 5'deki çalışmada yaş, vertebra sayısı ve skolyoz nedeni açısından tabakalandırma şemasına göre desmopressin ve plasebo grubuna alınan hasta sayıları.

Yaş	Vertebra sayısı	Skolyoz nedeni	Desmopressin	Plasebo
<20	1-5	Doğumsal	3	3
		Edinsel	-	-
	6-10	Doğumsal	4	5
		Edinsel	1	1
	>10	Doğumsal	2	2
		Edinsel	3	3
20-49	1-5	Doğumsal	1	1
		Edinsel	2	2
	6-10	Doğumsal	3	2
		Edinsel	1	1
	>10	Doğumsal	2	2
		Edinsel	4	4
>49	1-5	Doğumsal	-	-
		Edinsel	2	2
	6-10	Doğumsal	-	-
		Edinsel	1	1
	>10	Doğumsal	-	-
		Edinsel	3	3
Toplam			32	32

Çalışma süresince bu 18 kombinasyondan bazılarına uygun hastaların bulunmamasının görülmesi tabakalandırma ve bloklama için bir engel değildir. Örneğin hem desmopressin, hem de plasebo grubunda 20 yaşından küçük, 1-5 vertebraya müdahale yapılacak edinsel skolyozu olan hasta olmaması sorun oluşturmaz. Önemli olan toplam 18 altgruptaki hasta sayılarının desmopressin-plasebo gruplarında eşit olmasıdır. Ancak her şeye rağmen tüm prognostik etkenler açısından gruplar arasında denge tam olarak sağlanamayabilir.

Tabakalı Analiz

Çalışma gruplarının prognostik etkenler açısından dengeli olmasını sağlamanın en güvenilir yolu tabakalı randomizasyondur. Ancak kimi zaman araştırmaya başlarken prognostik önemi olmadığı sanılan bir etkenin araştırma sırasında ya da araştırmanın yürütülmesi bittikten sonra prognostik öneminin olduğu ve bu özellik açısından grupların den-

geli olmadığı anlaşılabilir. Örneğin, araştırmanın başlangıcında henüz bilinmeyen bir etkenin prognostik önemi olduğu, araştırma henüz devam ederken ya da bittikten sonra başka araştırma grupları tarafından gösterilebilir ya da daha sıklıkla araştırmaya başlamadan önce yapılan literatür taramasının yetersiz olmasına bağlı olarak, başlangıçta yapılan randomizasyonda bu etkene göre tabakalandırma yapılmamış olabilir.

Bu durumda prognostik etkenlerin dengesizliğinin çözümünde kullanılabilecek bir yöntem, analizi tabakalandırmaktır. Tabakalı analiz, çalışma grupları içindeki prognostik etkenler açısından homojen altgrupların analizinin ayrı ayrı yapılmasıdır. Bu yöntemle çalışmanın gücü azalır, ama ciddi dengesizlik nedeniyle işe yarayamayacak bir çalışmada yapılması gereken budur.

Tabakalı analizin nasıl yapıldığını Örnek 6'daki çalışmada inceleyelim.

ÖRNEK 6

Beta-bloker ve kalsiyum antagonistlerinin anstabil anjina pektoriste miyokard infarktüsü (Mİ) gelişmesini önleme üzerindeki etkilerinin karşılaştırılması amacıyla düzenlenen bir çalışmaya alınan 80 hasta iki gruba ayrılıyor ve bir gruba beta-bloker, diğer gruba kalsiyum antagonisti veriliyor. Tedaviye yanıtın farklı cinsiyetlerde farklı olmayacağı varsayıldığı için, hastaların ilaç gruplarına ayrılmasında cinsiyet, tabaka değişkeni olarak kullanılmıyor. Tedaviye yanıt için ilaç başlandıktan sonraki 6 ay içinde hastaların ne kadarının Mİ geçirdiği gözleniyor.

Tablo 4'ten Tablo 9'a kadar yer alan tablolarda Örnek 6'daki çalışmadaki verilere ilişkin çapraz tablolar görülmektedir. Beta-bloker grubunda 21, kalsiyum antagonisti grubunda 30 hastanın Mİ geçirdiği gözleniyor ve ki-kare analizi ile beta-bloker grubundaki Mİ oranının kalsiyum antagonisti grubundan istatistiksel olarak önemli derecede az olduğu görülüyor ($p=0,03$) (Tablo 4).

Tablo 4. İki tedavi grubunda Mİ geçiren hasta sayılarının gösterildiği çapraz tablo.

	Miyokard İnfarktüsü		Toplam
	+	-	
Beta-bloker	21	19	40
Kalsiyum antagonisti	30	10	40
Toplam	51	29	80

Ancak gruplardaki hastaların cinsiyetlere göre dağılımının farklı olduğu görülüyor. Beta-bloker grubunda 10 erkek, 30 kadın, kalsiyum antagonisti grubunda 30 erkek, 10 kadın var ($p<0,001$) (Tablo 5). Bu nedenle farklı cinslerde ilaçların etkilerinin farklı olabileceği ya da verilen ilaç ne olursa olsun farklı cinslerde hastalığın seyrinin farklı olabileceği gibi sorular akla geliyor. Gerçekten de iki farklı tedavi grubundaki hastalar cinsiyet ve prognoz arasındaki ilişki açısından incelendiği zaman her iki ilaç grubunda da erkeklerde Mİ oranının daha yüksek olduğu saptanıyor (erkeklerde $p=0,04$, kadınlarda $p<0,001$) (Tablo 6 ve 7).

Acaba beta-bloker grubunda infarktüs oranının düşük olmasının nedeni ilacın daha etkili olması mıdır, yoksa bu grupta hastalığın daha kötü seyrettiği anlaşılan cinsten, yani erkeklerden daha az hasta bulunması mıdır? Acaba her iki tedavi grubundaki hastaların cinsiyet açısından dağılımı benzer olsaydı,

sonuç yine beta-bloker lehine mi bulunacaktı? Bu sorunun çözümü için tabakalı analiz yöntemi kullanılabilir. İnfarktüs oranlarının, tedaviye yanıt ya da hastalığın seyri açısından farklı olan altgruplarda, yani erkek ve kadın tedavi gruplarında ayrı ayrı incelenmesi grupların cinsiyet açısından dengesiz olmasının neden olduğu yorumlama güçlüğünü ortadan kaldıracaktır. Gerçekten de hem erkeklerde ($p=0,41$) (Tablo 8), hem de kadınlarda ($p=0,46$) (Tablo 9) hangi ilaç kullanılırsa kullanılsın, prognoz farklı olmadığı görülmektedir.

Tablo 5. İki tedavi grubundaki hastaların cinsiyete göre dağılımının gösterildiği çapraz tablo.

	Erkek	Kadın	Toplam
Beta-bloker	10	30	40
Kalsiyum antagonisti	30	10	40
Toplam	40	40	80

Tablo 6. Beta-bloker grubunda farklı cinslerde Mİ geçiren hasta sayılarının gösterildiği çapraz tablo.

	Miyokard İnfarktüsü		Toplam
	+	-	
Erkek	8	2	10
Kadın	13	17	30
Toplam	21	19	40

Tablo 7. Kalsiyum antagonisti grubunda farklı cinslerde Mİ geçiren hasta sayılarının gösterildiği çapraz tablo.

	Miyokard İnfarktüsü		Toplam
	+	-	
Erkek	27	3	30
Kadın	3	7	10
Toplam	30	10	40

Tablo 8. İki tedavi grubunda Mİ geçiren erkek hasta sayılarının gösterildiği çapraz tablo.

	Miyokard İnfarktüsü		Toplam
	+	-	
Beta-bloker	8	2	10
Kalsiyum antagonisti	27	3	30
Toplam	35	5	80

Tablo 9. İki tedavi grubunda Mİ geçiren kadın hasta sayılarının gösterildiği çapraz tablo.

	Miyokard infarktüsü		Toplam
	+	-	
Beta-bloker	13	17	30
Kalsiyum antagonisti	3	7	10
Toplam	16	24	40

Bu örnekte olduğu gibi çalışma bittikten sonra altgrupların belirlenmesine “sonradan-tabakalandırma” (*post-stratification*) adı verilir. Çeşitli etkenlerin prognostik öneminin derecesinin belirlenmesinde sonradan-tabakalandırma yararlıdır, ancak bu yöntemde iki sorun vardır:

- Çalışmanın gücü azalır ve bazı tabakalarda hasta sayısı çok az olacağı için istatistik analiz sonuçları güvenilir olmayabilir.
- Gruplar arası karşılaştırma sayısı artacağı için “istatistiksel önemlilik enflasyonu” riski vardır.

Yeri gelmişken, çalışmanın gücü ya da istatistik analizin gücünün ne anlama geldiğini açıklamakta yarar var. Örneğin iki ilaca yanıtın incelendiği bir araştırmada, iki grupta tedaviye yanıt oranı arasında fark olduğu varsayımı araştırmada toplanan verilere dayanarak test edilir ve analiz sonucunda genellikle herkesin çok aşına olduğu bir p değeri elde edilir. P değeri, kabaca iki grup arasında gerçekte fark olmadığı halde, şans eseri olarak fark varmış gibi bir sonuç elde edilmiş olması olasılığıdır. Örneğin bir ilaca yanıt oranı %45, diğer ilaca yanıt oranı %60 ve analiz sonucunda hesaplanan p değeri 0,12 ise, iki ilaç arasında gözlenen bu farkın tamamen tesadüfi olma olasılığı %12’dir. Ya da daha doğru bir anlatımla, bu araştırma aynı sayıda hasta ile ve aynı koşullarda 100 kez yapılırsa, iki ilaç arasında 12 araştırmada fark bulunamayacak, 88’inde fark bulunacaktır.

Öte yandan gerçekte fark olduğu halde, yapılan araştırmadan elde edilen verilerin analizi ile, fark olmadığı gibi bir hata da yapılabilir. Bir çalışmanın gücü, gerçekte fark olduğu zaman bunun istatistik analizle de gösterilebilmesi, yani saptanabilmesi olasılığı olarak tanımlanabilir. Örneğin bir çalışmanın gücünün 0,85 olması, bu çalışma ile eğer

gerçekte de fark varsa bunun %85 olasılıkla saptanabileceği anlamına gelir.

Bu araştırmanın analizi 6 tane çapraz tablo ile yapılabilmektedir (Tablo 4-9). Mİ geçirmeyi önleme etkisi açısından beta-bloker ve kalsiyum antagonisti arasında fark olup olmadığı amacıyla yapılan çalışmaya 80 hasta alındığı halde erkekler (n=40) ve kadınlar (n=40) olmak üzere altgruplarda analiz yapıldığı için (yani daha az sayıda hasta ile analiz yapıldığı için) çalışmanın gücü önemli derecede azalmıştır. Eğer araştırmanın başlangıcında cinsiyetin prognoz üzerine etkisi olabileceği öngörülseydi ve buna göre hastalar tabakalandırılarak randomize edilseydi, 80 hastayla yapılmış olan bu araştırmadaki istatistik analizin gücüne, aşağı yukarı 40 hasta ile yapılacak bir çalışmayla ulaşılabilecekti.

Tabakalı analizin tek sakıncası çalışmanın gücünün azalması değildir. İstatistikte genel ilke olarak, mümkün olan en az sayıda analiz ile inceleme yapılmalıdır. Yukarıdaki örnekte hastalara verilen ilaç, cinsiyet ve prognoz olmak üzere üç değişken arasındaki ilişkiyi 6 farklı tabloda analiz etmek yalnızca zahmetli olduğu için değil, aynı zamanda istatistiksel önemlilik enflasyonuna neden olacağı için de sakıncalıdır. İstatistiksel önemlilik enflasyonu, çok sayıda analiz yapınca gerçekte fark olmadığı halde fark varmış gibi sonuç elde edilmesidir. Örneğin Tablo 4’ün analizi ile hesaplanan p değerinin 0,03 olması, iki ilaç arasında prognoz açısından tamamen şansa bağlı olarak fark olması olasılığının %3 olduğunu göstermektedir. Bilindiği gibi çoğu kez p değerinin 0,05’ten daha küçük olması halinde istatistiksel olarak önemli fark olduğu söylenir. Ancak bu araştırmada çok sayıda analiz yapıldığı için her analizle hesaplanan p değerlerinin ne kadar güvenilir olduğu tartışılır hale gelecektir. Bu durumda Tablo 4’ün analizi ile 0,03 olarak hesaplanan p değeri gerçekte olduğundan daha küçük olarak hesaplanmış olacaktır.

Sonuçların Prognostik Etkene Göre Standardizasyonu

Standardizasyon, tabakalardaki sonuçların ağırlıklandırılarak kullanılmasıdır. Ağırlıkların seçimi, analizin amacına göre değişir.

Standardizasyonun nasıl yapıldığını şu örnekte inceleyelim:

ÖRNEK 7

Hipertansiyonda A ve B ilaçlarının plasebo ile karşılaştırıldığı bir çalışmada grupların cinsiyet dağılımı açısından dengeli olmadığı görülmüştür (Tablo 10). Bu nedenle ilaca yanıt kriteri olan kan basıncındaki düşme miktarının grupların cinsiyet dağılımlarının dikkate alınmadan yapılmasının doğru olmayacağı düşünülmüştür.

Tablo 10. Örnek 7'deki çalışmada gruplardaki hastaların cinsiyete göre dağılımı.

	A ilacı	B ilacı	Plasebo	Toplam
Erkek	20	15	40	75
Kadın	20	35	20	75
Toplam	40	50	60	150

Örnek 7'deki çalışmada gruplar cinsiyet açısından dengeli olmadığı için önce kan basıncındaki düşme miktarı her grupta erkekler ve kadınlar için ayrı ayrı analiz edilmiştir (Tablo 11). Erkek ve kadınlarda kan basıncındaki azalma miktarı arasında farklılık olduğu görülmektedir. Bu nedenle erkekleri ve kadınları içeren tüm grubun ortak sonucu güvenilir olmayacaktır.

Gruplar cinsiyet açısından dengeli olmadığı için, tedavi yanıtın nasıl olacağını kestirmek için sonuçlar standardize edilmelidir. A ilacı alan hastalarda gözlenen ortalama kan basıncı düşmesi 23,9 mmHg'dir. Çalışmaya alınan 150 kişiden yarısı erkek, yarısı kadın olduğuna göre, bu 150 kişiye A ilacı verilseydi, kan basıncındaki ortalama düşme $(25,4 \times 0,50) + (22,4 \times 0,50) = 23,9$ mmHg olacaktı. Hesaplanan bu değer standardize edilmiş ortalama kan basıncı azalma miktarıdır. Bu gruptaki hastaların cinsiyete göre dağılımı tüm gruptaki dağılımın aynısı olduğu için, yani yarı yarıya erkek ve kadın olduğu için standardize ortalama, gözlenen ortalamadan farklı değildir. B ilacı verilen grupta gözlenen ortalama kan basıncı düşmesi 34,0 mmHg'dir. Benzer şekilde hesaplandığında standardize ortalama, $(22,4 \times 0,50) + (38,9 \times 0,50) = 30,6$ mmHg olarak hesaplanmıştır. Yani bu grupta 15 erkek, 35 kadın değil de; 25 erkek, 25 kadın olsaydı ortalama kan basıncı düşmesi 30,6 mmHg olacaktı. Plasebo grubunda ise gözlenen ortalama 4,5 mmHg, hesaplanan standardize ortalama ise $(1,2 \times 0,50) + (8,9 \times 0,50) = 5,1$ mmHg'dir. Bu üç ortalama, artık cinsiyetin etkisinden arıtılmıştır, çünkü her üçü de %50'si erkek, %50'si kadın olan hipotetik popülasyonlarda gözlenmesi beklenen değerlerdir.

Tablo 11. Örnek 7'deki çalışmada gruplarda kan basıncında ortalama düşme miktarı (mmHg).

	A ilacı	B ilacı	Plasebo
Erkek	25,4	22,4	1,2
Kadın	22,4	38,9	8,9
Genel ortalama	23,9	34,0	4,5
Standardize ortalama	23,9	30,6	5,1

Kontrol Değişkenlerinin Etkilerinin Artıldığı İstatistik Modeller

Gruplar arasında dengesizlik olduğu zaman, kullanılacak yöntemlerden birinin tabakalı analiz olduğundan ve bu yöntemin sakıncalarından yukarıda söz edildi. Dikkat ederseniz, tabakalı analizde prognostik etkene göre her alt grubun analizini ayrı ayrı yaparken, analize aldığımız hasta sayısı çalışmaya aldığımız hasta sayısından daha az oluyor ve buna bağlı olarak çalışmanın gücü azalıyor. Tabakalı analizin bu sakıncalarını ortadan kaldıran ve prognostik etkenlerin (kontrol değişkenlerinin) sonuç üzerine etkilerinin artıldığı ve tedavinin "saf" etkisinin belirlenebildiği matematik modeller vardır. Bu yöntemlerle sonuç üzerine etkisi olan etkenler (verilen ilaç, cinsiyet, hastalığın şiddeti, yaş vb.) parçalara ayrılarak incelenir. Buna göre yukarıdaki örnekte kan basıncındaki düşme miktarına tedavinin katkısı ve cinsiyetin katkısı ayrı ayrı hesaplanabilir.

Anstabil anjina pektoriste beta-bloker ve kalsiyum antagonistlerinin karşılaştırıldığı çalışmada tabakalı analizle iki ilaç arasında fark olmadığı, ilk bakışta gözlenen farkın iki gruptaki hastaların cinsiyete göre dağılımları arasındaki farktan kaynaklandığı, 6 farklı çapraz tablonun istatistiksel analizi ile gösterilmişti. Ancak bu sonuca ulaşabilmek için çalışmanın gücünün azalması ve istatistiksel önemlilik enflasyonu riskini göze almıştık. Bu çalışmada alternatif olarak log-lineer analiz yapıldığında iki tedavi grubundaki hastaların cinsiyet dağılımının ($p < 0,0001$) ve prognozun iki cinste benzer olmadığı ($p = 0,0001$), ancak tedavi ile prognoz arasında ilişki olmadığı ($p = 0,88$) saptanmıştır. Burada üzerinde önemle durulması gereken nokta tedavi ile prognoz arasında ilişki olmadığını söylerken, cinsiyetin etkisinin arıtılmış olmasıdır.

Prognostik etkenlerin etkilerinin arıtılmasında hangi durumlarda hangi istatistik modellerin uygulanabileceği Tablo 12'de özetlenmiştir.

Sonradan-tabakalandırma ile grup karşılaştırma sayısındaki artmaya bağlı “istatistiksel önemlilik enflasyonu”, bu istatistik modellerle önlenebilir. Bu yöntemlerle, hem tedavinin etkisi prognostik etkenlerden arıtılmış, hem de prognostik etkenlerin önemi test edilmiş olur.

Gerek tabakalı analiz, gerekse kontrol değişkenlerinin etkilerinin arıtıldığı modellerin amacı, gruplar arası dengesizliğin neden olduğu taraf tutmayı ya da yanlılığı ortadan kaldırmaktır. Ancak iki yöntem arasında önemli bir fark vardır. Tabakalı analizde, tabakalardaki hasta sayısı azalacağı için, çalışmanın gücünün azalması pahasına, taraf tutma ortadan kaldırılabilir. Kontrol değişkenlerinin etkilerinin arıtıldığı modeller kullanıldığında ise, (eğer doğru model kullanılırsa) hem taraf tutma azalır, hem de sonuçların doğruluk derecesi artar. Böylece istatistiksel önemliliği sınırda olan yargıların doğruluğu artabilir ya da bazen tersine farklı tedavi grupları arasında istatistiksel olarak önemli görünen farklılıkların tedavinin kendisinden değil, gruplar arası prognostik etken dengesizliğinden kaynaklandığı ortaya konabilir. Model kullanılmasının bir diğer yararı ise prognostik etkenlerin önemlilik derecesinin belirlenmesi ve daha sonra yapılacak çalışmaların planlanmasında yararlı olacak veriler elde edilmesidir.

Prognostik Etken Denetim Yöntemlerinin Karşılaştırılması

Çalışma grubu (örneklem) yeterince büyükse, basit randomizasyon, ölçülen ve ölçülmeyen prognostik etkenler açısından yeterli dengeyi genellikle sağlar. Randomizasyona bir-iki önemli prognostik etkene göre tabakalandırma ve bloklama eklenirse, gruplar daha da dengeli olur. Herşeye karşın, çalışmaya başlamadan önce denetlenmeyen bazı özellikler açısından gruplarda dengesizlik olabilir. Bu durumda hasta sayısının izin verdiği ölçüde sonradan-tabakalandırma yapılarak tabakalar ayrı ayrı analiz edilebilir. Grupların dengesi mükemmel olsa bile, analiz sırasında prognostik etkenlere göre standardizasyon yapılması ile gruplarda tedaviye yanıtın karşılaştırılması daha da saflaştırılmış olur.

Sonradan-tabakalandırma ve prognostik etkenlere göre standardizasyon birbirini tamamlayan iki yöntemdir. Sonradan-tabakalandırma ile prognostik etkenler açısından farklı olmalarına karşın, grupların karşılaştırılmasının ne kadar geçerli olduğu test edilmiş olur. Standardizasyon ile sonuçların doğruluğu daha da belirginleştirilip, çalışmadan daha fazla bilgi edinilmiş olur.

Araştırmanın amacı, çalışmaya alınan hastalar-daki sonuçların tüm hastalara genelleştirilebilmesidir. Bu yazıda anlatılan randomizasyon ve prognostik etken denetim çabalarının nedeni budur.

Tablo 12. Kontrol değişkenlerinin etkilerinin arıtıldığı istatistik modeller.

Grup Sayısı	İzlenen Değişken	Kontrol Değişkeni	Temel Analiz	Tabakalı Analiz	Alternatif Model
iki grup	Sayısal	Sayısal	t testi	-	Kovaryans analizi
		Kategorik	t testi	t testi	İki yönlü ANOVA
	Dikotom*	Sayısal	Ki-kare testi	-	Lojistik regresyon analizi
		Kategorik	Ki-kare testi	Ki-kare testi	Log-lineer analiz
	>2 düzeyli kategorik**	Sayısal	Ki-kare testi	-	Diskriminant analiz
		Kategorik	Ki-kare testi	Ki-kare testi	Log-lineer analiz
ikiden çok grup	Sayısal	Sayısal	Tek yönlü ANOVA	-	Kovaryans analizi
		Kategorik	Tek yönlü ANOVA	Tek yönlü ANOVA	İki yönlü ANOVA
	Dikotom	Sayısal	Ki-kare testi	-	Lojistik regresyon analizi
		Kategorik	Ki-kare testi	Ki-kare testi	Log-lineer analiz
	>2 düzeyli kategorik	Sayısal	Ki-kare testi	-	Diskriminant analiz
		Kategorik	Ki-kare testi	Ki-kare testi	Log-lineer analiz

*Yanıt kriteri “yanıt var-yok” ya da “öldü-yaşıyor” gibi iki sonuçtan ibaretse.

**Yanıt kriteri “yanıt tam-kısmi-yok” ya da “hafif-orta-ağır” gibi ikiden daha çok sonuç içeriyorsa.